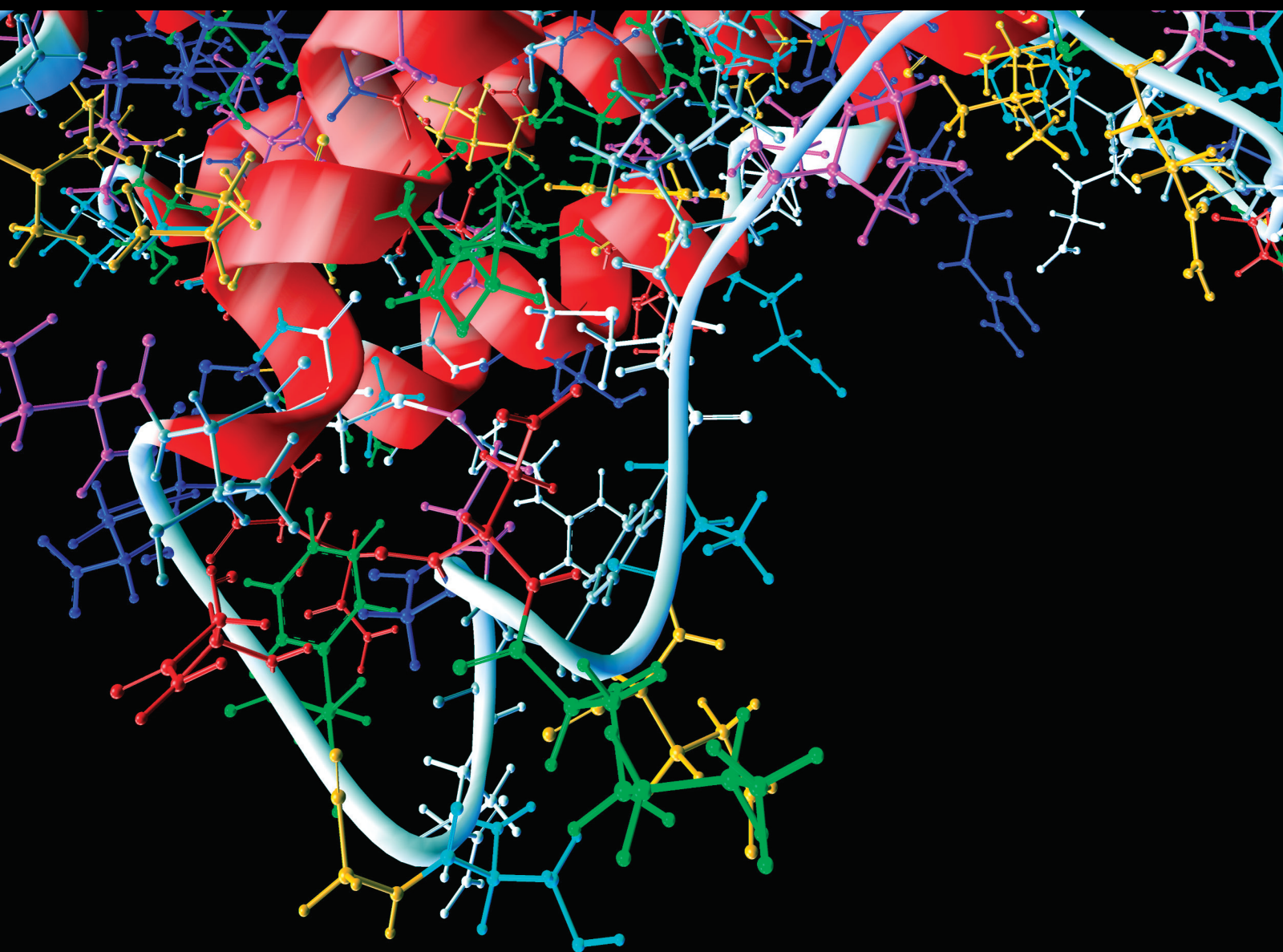


Machine Learning and Network Methods for Biology and Medicine 2022

Lead Guest Editor: Lei Chen

Guest Editors: Tao Huang, Chuan Lu, Lin Lu, and Dandan Li





**Machine Learning and Network Methods for
Biology and Medicine 2022**

Computational and Mathematical Methods in Medicine

**Machine Learning and Network
Methods for Biology and Medicine 2022**

Lead Guest Editor: Lei Chen




Guest Editors: Tao Huang, Chuan Lu, Lin Lu, and
Dandan Li



Copyright © 2023 Hindawi Limited. All rights reserved.

This is a special issue published in “Computational and Mathematical Methods in Medicine.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Associate Editors

Ahmed Albahri, Iraq
Konstantin Blyuss , United Kingdom
Chuangyin Dang, Hong Kong
Farai Nyabadza , South Africa
Kathiravan Srinivasan , India

Academic Editors

Laith Abualigah , Jordan
Yaser Ahangari Nanekaran , China
Mubashir Ahmad, Pakistan
Sultan Ahmad , Saudi Arabia
Akif Akgul , Turkey
Karthick Alagar, India
Shadab Alam, Saudi Arabia
Raul Alcaraz , Spain
Emil Alexov, USA
Enrique Baca-Garcia , Spain
Sweta Bhattacharya , India
Junguo Bian, USA
Elia Biganzoli , Italy
Antonio Boccaccio, Italy
Hans A. Braun , Germany
Zhicheng Cao, China
Guy Carrault, France
Sadaruddin Chachar , Pakistan
Prem Chapagain , USA
Huiling Chen , China
Mengxin Chen , China
Haruna Chiroma, Saudi Arabia
Watcharaporn Cholamjiak , Thailand
Maria N. D.S. Cordeiro , Portugal
Cristiana Corsi , Italy
Qi Dai , China
Nagarajan Deivanayagam Pillai, India
Didier Delignières , France
Thomas Desaive , Belgium
David Diller , USA
Qamar Din, Pakistan
Irimi Doytchinova, Bulgaria
Sheng Du , China
D. Easwaramoorthy , India

Esmaeil Ebrahimie , Australia
Issam El Naqa , USA
Ilias Elmouki , Morocco
Angelo Facchiano , Italy
Luca Faes , Italy
Maria E. Fantacci , Italy
Giancarlo Ferrigno , Italy
Marc Thilo Figge , Germany
Giulia Fiscon , Italy
Bapan Ghosh , India
Igor I. Goryanin, Japan
Marko Gosak , Slovenia
Damien Hall, Australia
Abdulsattar Hamad, Iraq
Khalid Hattaf , Morocco
Tingjun Hou , China
Seiya Imoto , Japan
Martti Juhola , Finland
Rajesh Kaluri , India
Karthick Kanagarathinam, India
Rafik Karaman , Palestinian Authority
Chandan Karmakar , Australia
Kwang Gi Kim , Republic of Korea
Andrzej Kloczkowski, USA
Andrei Korobeinikov , China
Sakthidasan Sankaran Krishnan, India
Rajesh Kumar, India
Kuruva Lakshmana , India
Peng Li , USA
Chung-Min Liao , Taiwan
Pinyi Lu , USA
Reinoud Maex, United Kingdom
Valeri Makarov , Spain
Juan Pablo Martínez , Spain
Richard J. Maude, Thailand
Zahid Mehmood , Pakistan
John Mitchell , United Kingdom
Fazal Ijaz Muhammad , Republic of Korea
Vishal Nayak , USA
Tongguang Ni, China
Michele Nichelatti, Italy
Kazuhisa Nishizawa , Japan
Bing Niu , China

Hyuntae Park , Japan
Jovana Paunovic , Serbia
Manuel F. G. Penedo , Spain
Riccardo Pernice , Italy
Kemal Polat , Turkey
Alberto Policriti, Italy
Giuseppe Pontrelli , Italy
Jesús Poza , Spain
Maciej Przybyłek , Poland
Bhanwar Lal Puniya , USA
Mihai V. Putz , Romania
Suresh Rasappan, Oman
Jose Joaquin Rieta , Spain
Fathalla Rihan , United Arab Emirates
Sidheswar Routray, India
Sudipta Roy , India
Jan Rychtar , USA
Mario Sansone , Italy
Murat Sari , Turkey
Shahzad Sarwar, Saudi Arabia
Kamal Shah, Saudi Arabia
Bhisham Sharma , India
Simon A. Sherman, USA
Mingsong Shi, China
Mohammed Shuaib , Malaysia
Prabhishek Singh , India
Neelakandan Subramani, India
Junwei Sun, China
Yung-Shin Sun , Taiwan
Min Tang , China
Hongxun Tao, China
Alireza Tavakkoli , USA
João M. Tavares , Portugal
Jlenia Toppi , Italy
Anna Tsantili-Kakoulidou , Greece
Markos G. Tsipouras, North Macedonia
Po-Hsiang Tsui , Taiwan
Sathishkumar V E , Republic of Korea
Durai Raj Vincent P M , India
Gajendra Kumar Vishwakarma, India
Liangjiang Wang, USA
Ruisheng Wang , USA
Zhouchao Wei, China
Gabriel Wittum, Germany
Xiang Wu, China

KI Yanover , Israel
Xiaojun Yao , China
Kaan Yetilmezsoy, Turkey
Hiro Yoshida, USA
Yuhai Zhao , China




Contents



Retracted: Clinical Characteristics and Treatment of Adult Idiopathic Carpal Tunnel Syndrome Accompanied with Trigger Digit
Computational and Mathematical Methods in Medicine
Retraction (1 page), Article ID 9837426, Volume 2023 (2023)


Retracted: Functions of Heparin Sodium Injection in the Prevention of Peripherally Inserted Central Catheter-Related Venous Thrombosis in NSCLC Patients during Postoperative Chemotherapy
Computational and Mathematical Methods in Medicine
Retraction (1 page), Article ID 9827825, Volume 2023 (2023)





Retracted: HMGB3 Targeted by miR-145-5p Impacts Proliferation, Migration, Invasion, and Apoptosis of Breast Cancer Cells
Computational and Mathematical Methods in Medicine
Retraction (1 page), Article ID 9783693, Volume 2023 (2023)



Retracted: Clinicopathological Features and Prognostic-Related Risk Factors of Gastric Signet Ring Cell Carcinoma: A Meta-Analysis
Computational and Mathematical Methods in Medicine
Retraction (1 page), Article ID 9764041, Volume 2023 (2023)

Combined Network Pharmacology and Molecular Docking to Verify the Treatment of Type 2 Diabetes with *Pueraria Lobata Radix* and *Salviae Miltiorrhizae Radix*
Jingxin Mao , Guowei Wang, Lin Yang, Lihong Tan, Cheng Tian, Lijing Tang, Ling Fang, Zhenqiang Mu, Zhaojing Zhu , and Yan Li 
Research Article (15 pages), Article ID 9150324, Volume 2023 (2023)

Thirdhand Smoke May Promote Lung Adenocarcinoma Development through HN1
Jiamin Peng, Jia Shuopeng, Huiju Wang, Xin Wang , and Ligang Wang 
Research Article (11 pages), Article ID 3407313, Volume 2023 (2023)

Analysis of the Clinical Efficacy and Molecular Mechanism of Xuefu Zhuyu Decoction in the Treatment of COPD Based on Meta-Analysis and Network Pharmacology
Yu Hu, Yunxi Lan, Qiqi Ran, Qianrong Gan, and Wei Huang 
Research Article (24 pages), Article ID 2615580, Volume 2022 (2022)

[Retracted] HMGB3 Targeted by miR-145-5p Impacts Proliferation, Migration, Invasion, and Apoptosis of Breast Cancer Cells
Yangying Hu , Deqi Wu , Rong Huang , and Zhijie Shi 
Research Article (12 pages), Article ID 1954099, Volume 2022 (2022)



[Retracted] Functions of Heparin Sodium Injection in the Prevention of Peripherally Inserted Central Catheter-Related Venous Thrombosis in NSCLC Patients during Postoperative Chemotherapy
Shanquan Li  and Hong Lu 
Research Article (7 pages), Article ID 1239058, Volume 2022 (2022)

The Clinical Value of Blood miR-654-5p, miR-126, miR-10b, and miR-144 in the Diagnosis of Colorectal Cancer

Gang Du , Chongren Ren , Ju Wang , and Jun Ma 



Research Article (8 pages), Article ID 8225966, Volume 2022 (2022)

[Retracted] Clinical Characteristics and Treatment of Adult Idiopathic Carpal Tunnel Syndrome Accompanied with Trigger Digit

Jinjiong Hong , Xiaofeng Wang, Jianbo Xue, Jimin Li, Minghua Zhang, and Weisheng Mao 



Research Article (7 pages), Article ID 8104345, Volume 2022 (2022)

Construction and Validation of a Prognostic Model Based on mRNasi-Related Genes in Breast Cancer

Xugui Zhao  and Jianqing Lin 

Research Article (21 pages), Article ID 6532591, Volume 2022 (2022)

Comparative Study on Feature Selection in Protein Structure and Function Prediction

Wenjing Yi, Ao Sun, Manman Liu, Xiaoqing Liu, Wei Zhang , and Qi Dai 


Research Article (13 pages), Article ID 1650693, Volume 2022 (2022)

miR-302a-3p Promotes Radiotherapy Sensitivity of Hepatocellular Carcinoma by Regulating Cell Cycle via MCL1

Zifeng Yang , Menglong Zhang , Jian Zhang , Cunkun Chu , Bijuan Hu , and Liyin Huang 












Research Article (12 pages), Article ID 1450098, Volume 2022 (2022)

Identification of Drug-Disease Associations Using a Random Walk with Restart Method and Supervised Learning

Xiaoqing Liu, Wenjing Yi, Baohang Xi, and Qi Dai 







Research Article (10 pages), Article ID 7035634, Volume 2022 (2022)

Identification of the Novel Gene Markers Based on the Gene Profile among Different Severity of Obstructive Sleep Apnea

Yi Ren , Yanyan Li , Xuemei Sui , Jinhe Yuan , Jun Lan , Xiayu Li , Yue Deng , Zhiping Xu , Xiu Cheng , Changjing Zhao , and Junyu Lu 







Research Article (13 pages), Article ID 6517965, Volume 2022 (2022)

Predicting Conserved Water Molecules in Binding Sites of Proteins Using Machine Learning Methods and Combining Features

Wei Xiao , Juhui Ren , Jutao Hao , Haoyu Wang , Yuhao Li , and Liangzhao Lin 

Research Article (11 pages), Article ID 5104464, Volume 2022 (2022)





Construction of Prognostic Risk Model for Small Cell Lung Cancer Based on Immune-Related Genes

Feng Deng , Feng Tao , Zhili Xu , Jun Zhou , Xiaowei Gong , and Ruhui Zhang 

Research Article (20 pages), Article ID 7116080, Volume 2022 (2022)

Contents

A Prognostic Model for the Respiratory Function of Patients with Nonsevere Pulmonary Infection Based on Breathing Exercises and Acupuncture Therapy: Development and Validation

Yulin Shi , Yong Hu , Guomeng Xu , and Yaoqi Ke 



Research Article (6 pages), Article ID 9057575, Volume 2022 (2022)

A Risk-Assessing Signature Based on Hypoxia- and Immune-Related Genes for Prognosis of Lung Adenocarcinoma Patients

Yue Wang , Jian Feng , Yang Liu , and Tingyue Han 





Research Article (13 pages), Article ID 7165851, Volume 2022 (2022)

Correlation of Pollen Concentration and Meteorological Factors with Medical Condition of Allergic Rhinitis in Shenyang Area

Feifei Jiang  and Aihui Yan 



Research Article (10 pages), Article ID 4619693, Volume 2022 (2022)

Prognosis and Efficacy of Laparoscopic Surgery on Patients with Endometrial Carcinoma: Systematic Evaluation and Meta-Analysis

Jiong Ma , Chunxia Zhou , Jinyan Chen , and Xuejun Chen 






Research Article (11 pages), Article ID 9384134, Volume 2022 (2022)

GPC1 Is Associated with Poor Prognosis and Treg Infiltration in Colon Adenocarcinoma

Ying Liu , Hui Ren, Mu-qing Yang, and Ji-yu Li 


Research Article (15 pages), Article ID 8209700, Volume 2022 (2022)

Comprehensive Analysis of the Immune Implication of AKAP12 in Stomach Adenocarcinoma

Zhiqian Xu , Ling Xiang , Linglong Peng , Haitao Gu , and Yaxu Wang 



Research Article (20 pages), Article ID 3445230, Volume 2022 (2022)

Integrative Bioinformatics Analysis Reveals That miR-524-5p/MEF2C Regulates Bone Metastasis in Prostate Cancer and Breast Cancer

QingHua Tian, YingYing Lu, BiCong Yan, and ChunGen Wu 





Research Article (13 pages), Article ID 5211329, Volume 2022 (2022)

Identification and Verification of an Alternative Polyadenylation-Related lncRNA Prognostic Signature for Glioma

Hui Wang  and ZhiJun Jiang 



Research Article (15 pages), Article ID 2164229, Volume 2022 (2022)

Expression and the Prognostic Value of Biglycan in Gastric Cancer

Sizhe Hu , Peipei Li , Chenying Wang , and Xiyong Liu 








Research Article (9 pages), Article ID 2656480, Volume 2022 (2022)

Endoscopic Surgical Treatment of Osteoarthritis and Prognostic Model Construction

Qi Su  and Guokang Xu 







Research Article (7 pages), Article ID 1799177, Volume 2022 (2022)

LINC00887 Fosters Development of Clear Cell Renal Cell Carcinoma via Inhibiting CD8+ T Cell Immune Infiltration

Jinfeng Wu , Rongcheng Lin , Lei Zhang , Yongbao Wei , Ruochen Zhang , Wanghai Cai , and Weilie Hu 






Research Article (17 pages), Article ID 2582474, Volume 2022 (2022)

Construction of Prognostic Risk Model of Patients with Skin Cutaneous Melanoma Based on TCGA-SKCM Methylation Cohort

Xiaoming Yu , Ping Cong , Wei Wei , Yong Zhou , Zhengqiang Bao , and Huaying Hou 





Research Article (13 pages), Article ID 4261329, Volume 2022 (2022)

A COVID-19 Auxiliary Diagnosis Based on Federated Learning and Blockchain

Ziyu Wang , Lei Cai , Xuewu Zhang , Chang Choi , and Xin Su 



Research Article (12 pages), Article ID 7078764, Volume 2022 (2022)

MiR-96-5p Facilitates Lung Adenocarcinoma Cell Phenotypes by Inhibiting FHL1

Feng Zhou , Chaojie Qian , Tingting Chen , and Xiaoliang Zang 

Research Article (10 pages), Article ID 7891222, Volume 2022 (2022)

Effectiveness and Safety of High-Power Radiofrequency Ablation Guided by Ablation Index for the Treatment of Atrial Fibrillation

Xuefeng Zhu , Chunxiao Wang , Hongxia Chu , Wenjing Li , Huihui Zhou , Lin Zhong , and Jianping Li 

Research Article (11 pages), Article ID 5609764, Volume 2022 (2022)

Comprehensive Multiomics Analysis Reveals Potential Diagnostic and Prognostic Biomarkers in Adrenal Cortical Carcinoma

Xiunan Li, Jiayi Li , Leizuo Zhao, Zicheng Wang, Peizhi Zhang , Yingkun Xu , and Guangzhen Wu 

Research Article (33 pages), Article ID 2465598, Volume 2022 (2022)

Identification of Human Retinal Organoid Cell Differentiation-Related Genes via Single-Cell Sequencing Data Analysis

He Dong , Liang Yu , Jian Song , Lili Ji , Xiaoxia Yu , and Lijun Zhang 

Research Article (19 pages), Article ID 9717599, Volume 2022 (2022)

Double-Balanced Loss for Imbalanced Colorectal Lesion Classification

Chang Yu , Wei Sun , Qilin Xiong , Junbo Gao , and Guoqiang Qu 

Research Article (14 pages), Article ID 1691075, Volume 2022 (2022)

[Retracted] Clinicopathological Features and Prognostic-Related Risk Factors of Gastric Signet Ring Cell Carcinoma: A Meta-Analysis

Ying Guo , Qian Wang , Qing Tian , Changwen Bo , Na Li , Sujing Zhang , and Peishun Li 




Review Article (11 pages), Article ID 3473445, Volume 2022 (2022)

Contents




Network Pharmacology and Molecular Docking Analysis on the Pharmacological Mechanisms of Modified Sanmiaosan in Treating Ulcerative Colitis

Yong Wang , Ying Sun , Ruoran Wang , Jisha Du , and Qingqing Wang 
Research Article (14 pages), Article ID 2556521, Volume 2022 (2022)

Research on Optimization Scheme for Blocking Artifacts after Patch-Based Medical Image Reconstruction

Yan Xu , Shunbo Hu , and Yuyue Du 
Research Article (17 pages), Article ID 2177159, Volume 2022 (2022)







Effects of Roflumilast on Patients with Chronic Obstructive Pulmonary Disease Treated with Inhaled Corticosteroid/Long-Acting β 2 Agonist: A Meta-analysis

Shasha Zeng , Haibing Bai , and Mi Zou 
Research Article (9 pages), Article ID 8101099, Volume 2022 (2022)

Mathematical Modeling and Computational Prediction of High-Risk Types of Human Papillomaviruses

Junchao Zhang  and Kechao Wang
Research Article (11 pages), Article ID 1515810, Volume 2022 (2022)

TRIM66 Promotes Malignant Progression of Non-Small-Cell Lung Cancer Cells via Targeting MMP9

Yufen Xu , Qi Yang , Zhixian Fang , Xiaoli Tan , Ming Zhang , and Wenyu Chen 
Research Article (8 pages), Article ID 6058720, Volume 2022 (2022)

Dysregulation of Circadian Clock Genes Associated with Tumor Immunity and Prognosis in Patients with Colon Cancer

Yongshan He, Yuanyuan Chen, Xuan Dai, and Shiyong Huang 
Research Article (19 pages), Article ID 4957996, Volume 2022 (2022)

Screening of Prognostic Markers for Hepatocellular Carcinoma Patients Based on Multichip Combined Analysis

Yong Dong , Qian Miao , and Da Li 
Research Article (12 pages), Article ID 6881600, Volume 2022 (2022)





Effect of Enhanced Recovery after Surgery with Integrated Traditional Chinese and Western Medicine on Postoperative Stress Response of Patients with Gastrointestinal Tumors

Haiping Zhao  and Wenhui Sun 
Research Article (8 pages), Article ID 3663246, Volume 2022 (2022)

Influences of Oral Administration of Probiotics on Posthepatectomy Recovery in Patients in Child-Pugh Grade



Hao Huang , Fang Fang , Zheng Jia , Wen Peng , and Yi Wu 
Research Article (7 pages), Article ID 2942982, Volume 2022 (2022)

Construction of Lymph Node Metastasis-Related Prognostic Model and Analysis of Immune Infiltration Mode in Lung Adenocarcinoma

Wujin Li , Debin Ou , Jiguang Zhang , and Mingfan Ye 


Research Article (10 pages), Article ID 3887857, Volume 2022 (2022)



Combination of the NRF2 Inhibitor and Autophagy Inhibitor Significantly Inhibited Tumorigenicity of Castration-Resistant Prostate Cancer

Yong Zhang, Zhixiang Xin, Baijun Dong , and Wei Xue 

Research Article (11 pages), Article ID 4182401, Volume 2022 (2022)

Effect of Probiotics Therapy on Nonalcoholic Fatty Liver Disease

Yuanshe Huang, Xiaodong Wang, Lai Zhang, Ke Zheng, Jie Xiong, Jing Li, Chunlei Cong , Zhaomiao

Gong , and Jingxin Mao 

Research Article (15 pages), Article ID 7888076, Volume 2022 (2022)

Retraction

Retracted: Clinical Characteristics and Treatment of Adult Idiopathic Carpal Tunnel Syndrome Accompanied with Trigger Digit

Computational and Mathematical Methods in Medicine

Received 26 September 2023; Accepted 26 September 2023; Published 27 September 2023

Copyright © 2023 Computational and Mathematical Methods in Medicine. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] J. Hong, X. Wang, J. Xue, J. Li, M. Zhang, and W. Mao, "Clinical Characteristics and Treatment of Adult Idiopathic Carpal Tunnel Syndrome Accompanied with Trigger Digit," *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 8104345, 7 pages, 2022.

Retraction

Retracted: Functions of Heparin Sodium Injection in the Prevention of Peripherally Inserted Central Catheter-Related Venous Thrombosis in NSCLC Patients during Postoperative Chemotherapy

Computational and Mathematical Methods in Medicine

Received 26 September 2023; Accepted 26 September 2023; Published 27 September 2023

Copyright © 2023 Computational and Mathematical Methods in Medicine. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] S. Li and H. Lu, "Functions of Heparin Sodium Injection in the Prevention of Peripherally Inserted Central Catheter-Related Venous Thrombosis in NSCLC Patients during Postoperative Chemotherapy," *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 1239058, 7 pages, 2022.

Retraction

Retracted: HMGB3 Targeted by miR-145-5p Impacts Proliferation, Migration, Invasion, and Apoptosis of Breast Cancer Cells

Computational and Mathematical Methods in Medicine

Received 26 September 2023; Accepted 26 September 2023; Published 27 September 2023

Copyright © 2023 Computational and Mathematical Methods in Medicine. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Y. Hu, D. Wu, R. Huang, and Z. Shi, "HMGB3 Targeted by miR-145-5p Impacts Proliferation, Migration, Invasion, and Apoptosis of Breast Cancer Cells," *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 1954099, 12 pages, 2022.

Retraction

Retracted: Clinicopathological Features and Prognostic-Related Risk Factors of Gastric Signet Ring Cell Carcinoma: A Meta-Analysis

Computational and Mathematical Methods in Medicine

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

Copyright © 2023 Computational and Mathematical Methods in Medicine. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Y. Guo, Q. Wang, Q. Tian et al., "Clinicopathological Features and Prognostic-Related Risk Factors of Gastric Signet Ring Cell Carcinoma: A Meta-Analysis," *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 3473445, 11 pages, 2022.

Research Article

Combined Network Pharmacology and Molecular Docking to Verify the Treatment of Type 2 Diabetes with *Pueraria Lobata Radix* and *Salviae Miltiorrhizae Radix*

Jingxin Mao ^{1,2}, Guowei Wang,² Lin Yang,^{1,3} Lihong Tan,^{1,3} Cheng Tian,^{1,3} Lijing Tang,^{1,3} Ling Fang,^{1,3} Zhenqiang Mu,^{1,3} Zhaojing Zhu ^{1,3} and Yan Li ^{1,3}

¹Chongqing Medical and Pharmaceutical College, Chongqing 400030, China

²College of Pharmaceutical Sciences, Southwest University, Chongqing 400715, China

³Chongqing Key Laboratory of High Active Traditional Chinese Drug Delivery System, Chongqing 400030, China

Correspondence should be addressed to Zhaojing Zhu; zhaojing6271@126.com and Yan Li; ly20031079@163.com

Received 5 October 2022; Revised 27 October 2022; Accepted 24 November 2022; Published 11 February 2023

Academic Editor: Lei Chen

Copyright © 2023 Jingxin Mao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. To explore the potential molecular mechanism of *Pueraria Lobata Radix* (RP) and *Salviae Miltiorrhizae Radix* (RS) in the treatment of type 2 diabetes mellitus (T2DM) based on network pharmacology and molecular docking. **Methods.** The chemical constituents and core targets of RP and RS were searched by Traditional Chinese Medicine System Pharmacology (TCMSP); target genes related to T2DM were obtained through GeneCards database, component target network diagram was constructed, intersection genes of active compounds and T2DM were synthesized, protein-protein interaction (PPI) relationship was obtained, and core targets were screened by using Cytoscape 3.7.2. Gene Ontology (GO) biological process and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were analyzed utilizing R studio 4.0.4 according to David database. Based on molecular docking, the screened active components of RP and RS were verified by molecular docking with the core target using Discovery Studio 2019. **Results.** There were totally 92 components and 29 corresponding targets in the component target network of RP and RS drug pair, of which 6 were the core targets of RP and RS in the treatment of T2DM. Molecular docking results showed that the active compounds of puerarin, formononetin, tanshinone iia, and luteolin had better binding activity with AKT1, VEGFA, NOS3, PPARG, MMP9, and VCAM1, respectively. Among them, puerarin showed significant effects in activating NOS3 pathway and luteolin exhibited significant effects in activating MMP9 pathway, respectively. The main biological processes mainly including xenobiotic stimulus, response to peptide, gland development, response to radiation, cellular response to chemical stress, response to oxygen levels, and the main signal pathways include response to xenobiotic stimulus, cellular response to chemical stress, response to peptide, gland development, and response to oxygen levels. **Conclusion.** Network pharmacology is an effective tool to explain the action mechanism of Traditional Chinese Medicine (TCM) from the overall perspective. RP and RS pair could alleviate T2DM via the molecular mechanism predicted by the network pharmacology, which provided new ideas and further research on the molecular mechanism of T2DM.

1. Introduction

Diabetes mellitus (DM) is a group of metabolic diseases characterized by increased blood glucose due to genetic and environmental effects, resulting in defective insulin secretion and reduced sensitivity of target cells to insulin. Type 2 diabetes mellitus (T2DM) is a chronic metabolic disease, mainly associated with the accumulation of lipids in the insulin β -cells

can lead to elevated blood glucose levels and abnormal glucose tolerance due to the dysfunction of the pancreatic β -cells, and the clinical symptoms are persistent hyperglycemia. In recent decades, the prevalence of T2DM has gradually increased worldwide. The prevalence of DM in the world has increased by 102.9% from 1990 to 2010 and is the country with the largest number of diabetic patients in the world [1]. T2DM is a common clinical endocrine and metabolic disease, and the

TABLE 1: A total of possible active ingredient information of RP and RS. (a) Intersection genes between T2DM and RS. (b) Intersection genes between T2DM and RP. (c) Intersection genes among T2DM, RP, and RS.

MOL_ID	Classification	Molecule_name	Ob	Mw	DI
MOL000358	RP	Beta-sitosterol	36.9139058327	414.790	0.75123
MOL012297	RP	Puerarin	24.0308979964	416.410	0.69099
MOL002959	RP	3'-Methoxydaidzein	48.5690937439	284.280	0.24261
MOL004631	RP	7,8,4'-Trihydroxyisoflavone	20.6680736276	270.250	0.21583
MOL000392	RP	Formononetin	69.6738806088	268.280	0.21202
MOL000390	RP	Daidzein	19.4410626585	254.250	0.18694
MOL007140	RS	(Z)-3-[2-[(E)-2-(3,4-dihydroxyphenyl)vinyl]-3,4-dihydroxy-phenyl]acrylic acid	88.5360210103	314.310	0.25869
MOL007150	RS	(6S)-6-hydroxy-1-methyl-6-methylol-8,9-dihydro-7H-naphtho[8,7-g]benzofuran-10,11-quinone	75.3858784662	312.340	0.4551
MOL007058	RS	Formyltanshinone	73.4446220045	290.280	0.41736
MOL007120	RS	Miltionone II	71.0297032063	312.390	0.43711
MOL007105	RS	Epidanshenspiroketallactone	68.2731592943	284.380	0.30549
MOL007155	RS	(6S)-6-(hydroxymethyl)-1,6-dimethyl-8,9-dihydro-7H-naphtho[8,7-g]benzofuran-10,11-dione	65.25893771	310.370	0.44871
MOL007130	RS	Prolithospermic acid	64.3709620672	314.310	0.31017
MOL007050	RS	2-(4-hydroxy-3-methoxyphenyl)-5-(3-hydroxypropyl)-7-methoxy-3-benzofurancarboxaldehyde	62.7841472598	356.400	0.39628
MOL007068	RS	Przewaquinone B	62.2400596208	292.300	0.41374
MOL000569	RS	Digallate	61.8486180263	322.240	0.25635
MOL007081	RS	Danshenol B	57.950875299	354.480	0.55764
MOL007082	RS	Danshenol A	56.9652489923	336.410	0.52172
MOL007069	RS	Przewaquinone c	55.7416730964	296.340	0.40408
MOL007108	RS	Isocryptotanshinone	54.9819324596	296.390	0.39449
MOL007125	RS	Neocryptotanshinone	52.4879970089	314.410	0.32306
MOL007079	RS	Tanshinaldehyde	52.4747043036	308.350	0.45196
MOL007088	RS	Cryptotanshinone	52.3419622629	296.390	0.39555
MOL007094	RS	Danshenspiroketallactone	50.4312810266	282.360	0.3067
MOL007111	RS	Isotanshinone II	49.9160257407	294.370	0.39674
MOL007154	RS	Tanshinone iia	49.8873000352	294.370	0.39781
MOL007119	RS	Miltionone I	49.6843943314	312.390	0.32125
MOL007098	RS	Deoxyneocryptotanshinone	49.4003470541	298.410	0.28555
MOL007048	RS	(E)-3-[2-(3,4-dihydroxyphenyl)-7-hydroxy-benzofuran-4-yl]acrylic acid	48.2436324414	312.290	0.31229
MOL007156	RS	Tanshinone VI	45.6373060194	296.340	0.29549
MOL007141	RS	Salvianolic acid g	45.5648557799	340.300	0.60602
MOL001942	RS	Isoimperatorin	45.4642467387	270.300	0.22524
MOL007115	RS	Manool	45.0443163606	304.570	0.20208
MOL007101	RS	Dihydrotanshinone I	45.0432791888	278.320	0.36015
MOL007123	RS	Miltirone II	44.9510664818	272.320	0.23537
MOL007045	RS	3 α -hydroxytanshinone II A	44.92933597	310.370	0.44272
MOL001659	RS	Poriferasterol	43.8298515785	412.770	0.75596
MOL002651	RS	Dehydrotanshinone II A	43.7622859945	292.350	0.40019
MOL007077	RS	Sclareol	43.6706845842	308.560	0.2058
MOL007152	RS	Przewaquinone E	42.8548520397	312.340	0.45301
MOL007151	RS	Tanshindiol B	42.6658104891	312.340	0.45303
MOL007070	RS	(6S,7R)-6,7-dihydroxy-1,6-dimethyl-8,9-dihydro-7H-naphtho[8,7-g]benzofuran-10,11-dione	41.3104570553	312.340	0.453
MOL007041	RS	2-isopropyl-8-methylphenanthrene-3,4-dione	40.8601540814	264.340	0.22897

TABLE 1: Continued.

MOL_ID	Classification	Molecule_name	Ob	Mw	DI
MOL007071	RS	Przewaquinone F	40.3078839947	312.340	0.45925
MOL002776	RS	Baicalin	40.1236099599	446.390	0.75264
MOL007118	RS	Microstegiol	39.6122945749	298.460	0.27734
MOL006824	RS	α -Amyrin	39.5120897831	426.800	0.76221
MOL007124	RS	Neocryptotanshinone II	39.4629911416	270.350	0.23157
MOL007093	RS	Dan-shexinkum D	38.8830210096	336.410	0.55453
MOL007122	RS	Miltirone	38.7569863502	282.410	0.25418
MOL001601	RS	1,2,5,6-Tetrahydrotanshinone	38.7453867227	280.340	0.35791
MOL007100	RS	Dihydrotanshinolactone	38.6847682981	266.310	0.32227
MOL007063	RS	Przewalskin A	37.1065006596	398.490	0.64901
MOL007061	RS	Methylenetanshinquinone	37.07319368	278.320	0.36017
MOL001771	RS	Poriferast-5-en-3beta-ol	36.9139058327	414.790	0.75034
MOL007121	RS	Miltipolone	36.5561120639	300.430	0.36803
MOL000006	RS	Luteolin	36.1626293429	286.250	0.24552
MOL002222	RS	Sugiol	36.113534855	300.480	0.27648
MOL007107	RS	C09092	36.069489858	286.500	0.2474
MOL007127	RS	1-methyl-8,9-dihydro-7H-naphtho[5,6-g]benzofuran-6,10,11-trione	34.7208221315	280.290	0.36634
MOL007149	RS	NSC 122421	34.4929230859	300.480	0.27645
MOL007049	RS	4-Methylenemiltirone	34.348675885	266.360	0.22726
MOL007036	RS	5,6-dihydroxy-7-isopropyl-1,1-dimethyl-2,3-dihydrophenanthren-4-one	33.7652523626	298.410	0.28585
MOL007143	RS	Salvilenone I	32.4347085631	270.400	0.22895
MOL007059	RS	3-beta-Hydroxymethylenetanshiquinone	32.1610337648	294.320	0.40894
MOL007145	RS	Salviolone	31.7241503895	268.380	0.23568
MOL007085	RS	Salvilenone	30.3836538737	292.400	0.37639

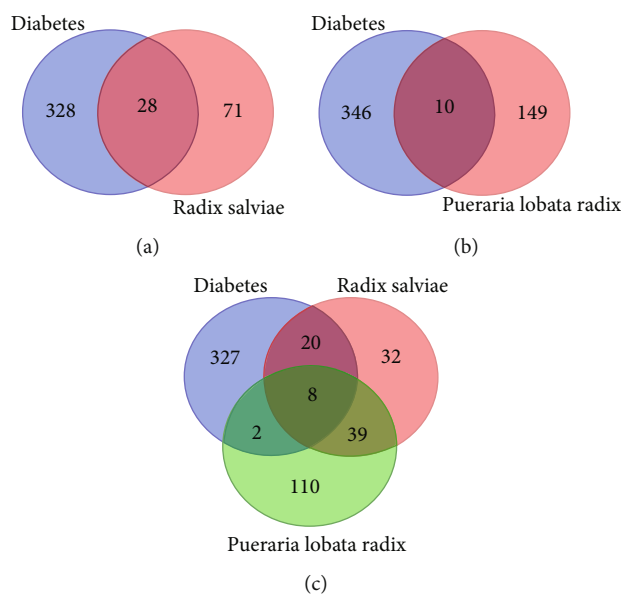


FIGURE 1: Venn diagram of intersection genes of active ingredient related targets of RP and RS with anti-T2DM targets.

prevalence of DM in Chinese population is about 11.6%, and the prevalence of pre-DM is 50.1% [2], which is a chronic metabolic disease that seriously affects human health.

The important cause of T2DM is the disorder of lipid metabolism, therefore, it is especially important to control blood sugar and regulate blood lipid in the process of

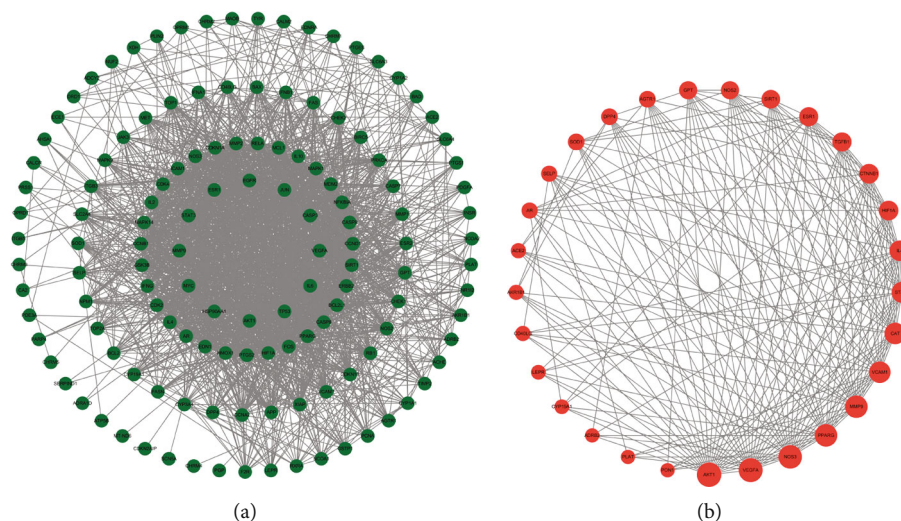


FIGURE 2: Active ingredients of RP and RS with anti-T2DM target network.

TABLE 2: Main ingredients from RP and RS based on number of corresponding targets. (a) PPI visualization network diagram of 120 targets; (b) PPI visualization network diagram of 29 core targets.

MOL_ID	Molecule_name	Classification	Ob	Mw	DI
MOL012297	Puerarin	RP	24.0308979964	416.410	0.69099
MOL000392	Formononetin	RP	69.6738806088	268.280	0.21202
MOL007154	Tanshinone iia	RS	49.8873000352	294.370	0.39781
MOL000006	Luteolin	RS	36.1626293429	286.250	0.24552

T2DM treatment in time [3]. At present, the main methods of T2DM drug treatment are subcutaneous insulin injection and oral hypoglycemic drugs (such as double pulse, sulfonyl artery, and a glyoxylase inhibitors); but with the development of the diseases

e, single-target development of drugs is difficult to treat T2DM and prevent the occurrence of complications such as glucose disorders and gastrointestinal tract, thus reducing the quality of life of patients [4].

Traditional Chinese Medicine (TCM) treatment of T2DM is based on single-target superposition and, multi-component and multitarget synergistic toxicity dispersion effect to achieve better efficacy and lower toxicity group, which is a complex theoretical system to achieve multidimensional regulation from a whole [5]. DM belongs to the category of “thirst disorders” in TCM, and the main treatment of DM in TCM is to clear heat and moisten dryness, and to nourish yin and produce fluid [6]. After the TCM, doctors have formulated prescriptions according to the above treatment rules, the overall regulating advantages of TCM, which can play a role in treating both the symptoms and the root cause, and TCM is milder and more durable than western medicine, with fewer side effects [7].

Pueraria Lobata Radix (RP) and *Salviae Miltiorrhizae Radix* (RS) are a common pair of medicine contained in the famous Chinese medicine book “Shi Jinmo on Medicine.” RP is the dried root of *Pueraria lobata* (Willd.) Ohwi, which has the functions of quenching thirst, raising yang,

relieving diarrhea, relieving muscle and fever, and activating meridians [8]. RS is also called “DanShen” that derived from the dried root and rhizome of *Salvia miltiorrhiza* Bunge, which has the effects of activating blood circulation, cooling the blood and clearing the heart, and removing irritation and calming the mind [9]. The combination of RP and RS exhibits the effect of promoting “qi,” resolving blood stasis, promoting blood circulation, and relieving pain, and the two are used as a pair to treat diabetes [10, 11]. Modern medicinal chemistry and pharmacology research shows that the main chemical components of RP are isoflavones, triterpenoids, flavonoids, coumarins, etc., which have pharmacological effects such as hypoglycemic and hypolipidemic, anti-inflammatory and antioxidant, and hepatoprotective [12]. The main chemical components of RS mainly including tanshinones, tanshin acids, and volatile oils, which have antioxidant, anti-inflammatory, and antithrombotic pharmacological effects [13].

Previous studies have been demonstrated that *Pueraria Lobata Radix* and *Salviae Miltiorrhizae Radix* (RP-RS) paired exhibits varies effects on T2DM or diabetes related diseases [10, 14–17]. However, the research on effective compounds of RP-RS is not in-depth, and the analytical methods are not comprehensive in these studies. Therefore, based on the idea of multicomponent and multitarget research, the present study was conducted to predict the mechanism of action and targets of RP-RS drug for the treatment of T2DM through network pharmacology and molecular docking methods, to find

TABLE 3: Corresponding core targets genes of main ingredient based on the degree value.

Gene name	Degree	Betweenness centrality	Closeness centrality
AKT1	26	0.09722044	0.93333333
VEGFA	25	0.08297252	0.90322581
NOS3	24	0.0748313	0.875
PPARG	22	0.06089927	0.82352941
MMP9	22	0.04616644	0.82352941
VCAM1	20	0.04275445	0.77777778
CAT	20	0.0317474	0.77777778
STAT3	19	0.0224157	0.75675676
IL4	18	0.02462019	0.73684211
HIF1A	16	0.00700457	0.7
CTNNA1	16	0.01164871	0.7
ESR1	15	0.01381435	0.68292683
TGFB1	15	0.0043485	0.68292683
SIRT1	14	0.00583073	0.66666667
GPT	13	0.00790826	0.65116279
NOS2	13	0.00240188	0.65116279
SOD1	10	0.00254587	0.60869565
AGTR1	10	0.00472411	0.60869565
DPP4	10	0.00459492	0.60869565
AR	9	0.00176367	0.58333333
SELP	9	0.00262661	0.59574468
CD40LG	7	0	0.57142857
AKR1B1	7	7.94E-04	0.57142857
ACE2	7	6.42E-04	0.57142857
CYP19A1	6	5.93E-04	0.56
LEPR	6	1.89E-04	0.56
PLAT	5	0	0.54901961
PON1	5	2.04E-04	0.53846154
ADRB2	5	2.94E-04	0.54901961

the active chemical components, disease therapeutic targets and signaling pathways, and to provide a scientific basis for the experimental research and clinical application of RP-RS drug for the treatment of T2DM.

2. Materials and Methods

2.1. Screening of RP and RS for Active Components and Potential Targets. The chemical components of RP and RS were obtained with the help of the Traditional Chinese Medicine System Pharmacology Analysis Platform (TCMSP, <http://temspw.com/temsp.php>), according to the set the bio-availability (OB) > 20 and drug-like properties (DL) > 0.18, screening out eligible active ingredients and related targets.

2.2. T2DM-Related Target Prediction. Through the GeneCards (<http://www.genecards.org/>) database, with “Type 2 diabetes mellitus or T2DM” as the search term, the targets related to T2DM were obtained, and the correlation value (relevance score > 18) was used as the screening condition,

and the screening results were used as candidate target genes of T2DM.

2.3. Construction of Protein Interaction Network (PPI). The effect of drugs on disease treatment is ultimately manifested in protein interactions. The Venn diagram (<https://bioinfo.gp.cnb.csic.es/tools/venny/index.html>) was used to obtain the intersection genes of the active ingredient target and the T2DM target, and imported into STRING11.0 (<http://www.string-db.org/>) to obtain the protein interaction relationship. Using Cytoscape 3.7.2 draws the interaction network, analyzes the key targets by degree value, and constructs the interaction network of RP-RS in the treatment of T2DM.

2.4. Construction of Chemical Components-Target Network of TCM. The screened candidate compounds of RP-RS that predicted component target genes were imported into Cytoscape 3.7.2 software to construct a compound-target network, and the main active ingredients/components of RP and RS were analyzed, respectively, according to the degree value in the network.

2.5. Analysis of Biological Function and Pathway Enrichment of RP-RS Pair in the Treatment of T2DM. GO enrichment is a commonly used method for analysis of omic data, which is usually used to discover the enrichment degree of GO term in differential genes. Through GO enrichment analysis, the genes in the difference table can be classified according to their functions to achieve the purpose of annotation and classification of genes [18]. KEGG is a practical program database for understanding advanced functions and biological systems (such as cells, organisms, and ecosystems) from molecular level information, especially genome sequencing and other high-throughput experimental technologies generated from large molecular data sets. It can predict the role of PPI networks in various cell activities [19]. Based on the David database (<https://david.ncifcrf.gov/>), the targets of RP-RS for the treatment of T2DM were collected and imported into R studio 4.0.4 (The species was limited to “Human”). GO functional annotation and KEGG pathway enrichment analysis were performed on the intersection target genes of RP-RS and T2DM, respectively ($P < 0.05$). Potential targets were analyzed for biological process (BP), cellular component (CC), and molecular function (MF), and KEGG was used for pathway enrichment analysis of potential targets. To validate the anti-T2DM mechanism of RP-RS across the key targets and multiple pathways, the KEGG mapper functional analysis was used to mark the target genes on the pathway associated with T2DM.

2.6. Molecular Docking Prediction of Key Targets of Active Ingredient Intervention in the Treatment of T2DM by RP-RS. According to the analysis of network pharmacology results, molecular docking software was used to predict the key targets of RP-RS intervention on the main active components. The 3D structure of the compound was constructed using Chem 3D of ChemOffice software and saved in.mol2 format. Download the protein structure of the target from the PDB (<https://www.rcsb.org/>) database, use PyMOL 2.5 to perform protein and ligand separation, dehydrogenation, water addition, and other operations on the original PDB

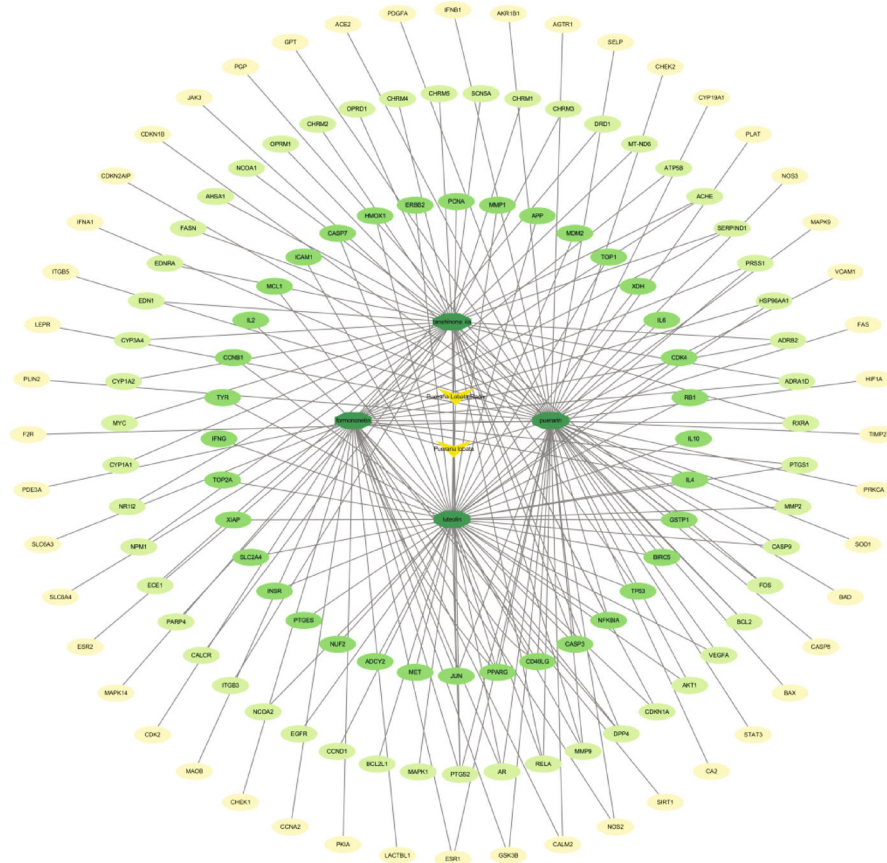


FIGURE 3: Protein interaction network of RP and RS with anti-T2DM.

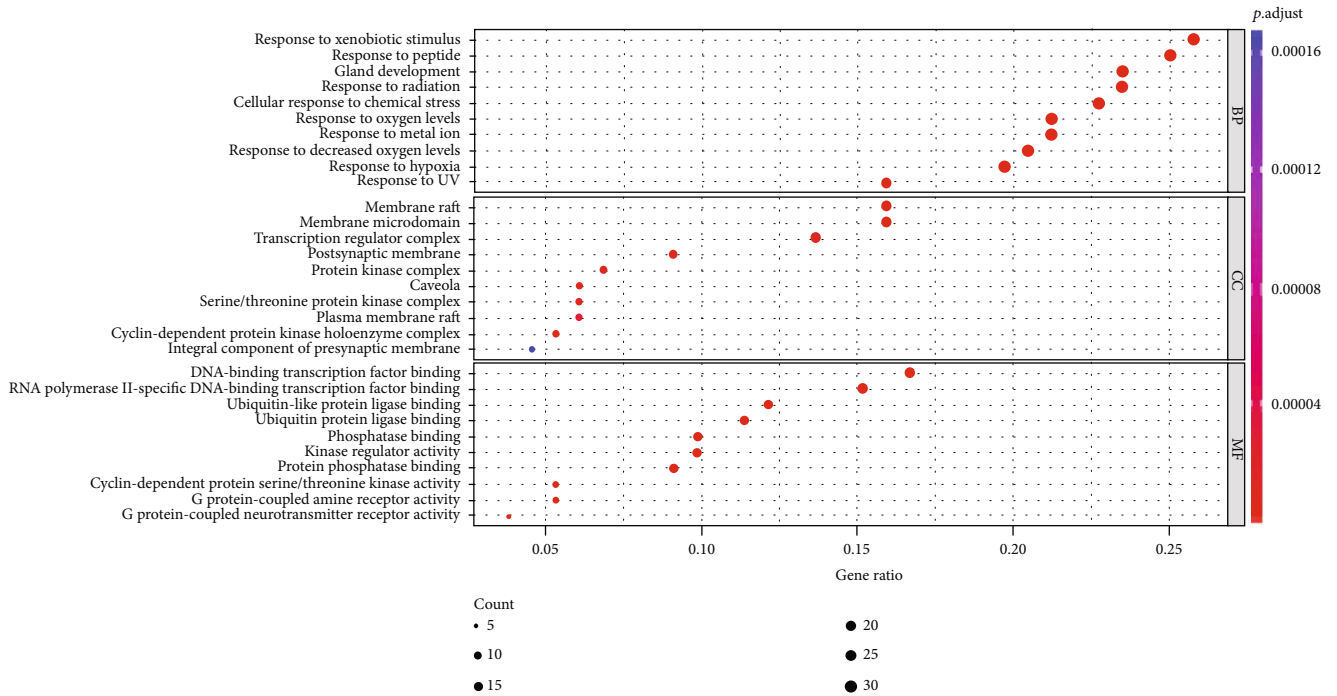


FIGURE 4: RP and RS prevention and treatment of T2DM gene GO enrichment analysis diagram.

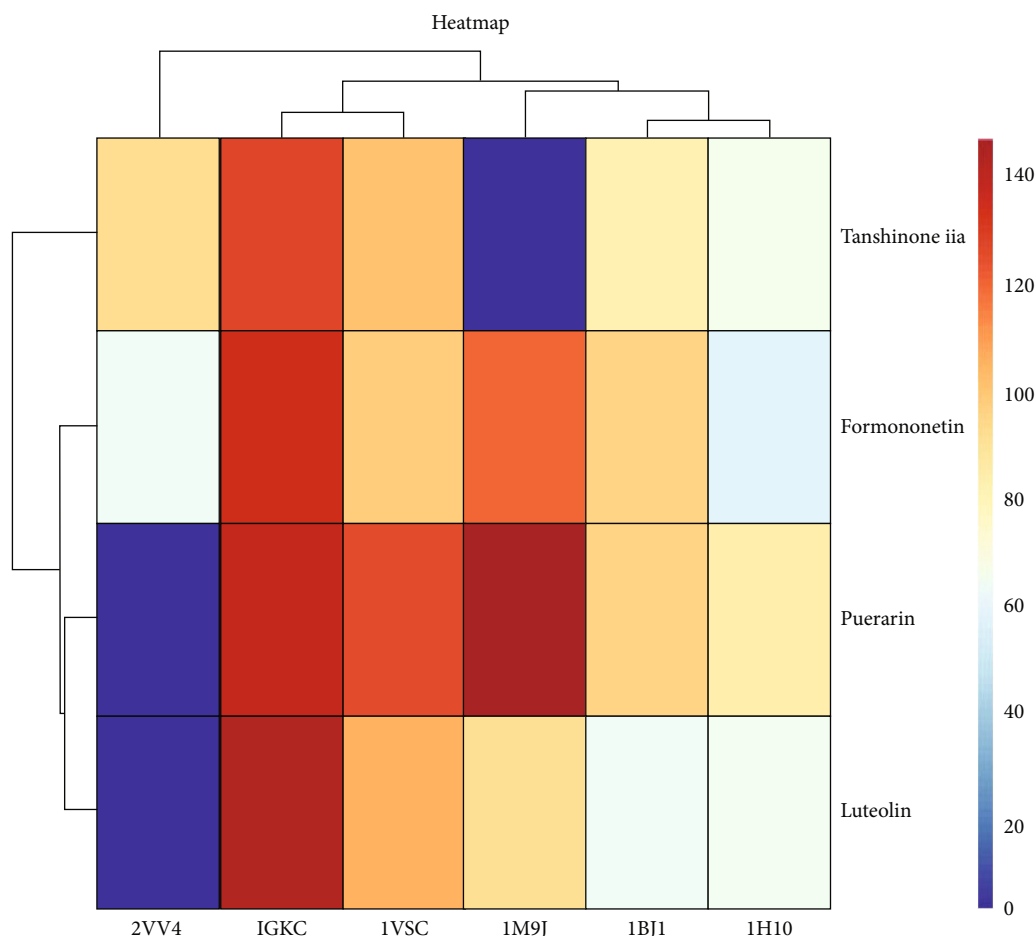


FIGURE 7: Heatmap of main active ingredients of RP and RS with core target genes.

the screening conditions, 7 and 85 chemical components of RP and RS were obtained, respectively (Table 1).

3.2. Intersection Genes of RP-RS with T2DM. By database analysis, 356, 159, and 99 targets were identified for T2DM, RP, and RS, respectively. The online Venn diagram analysis showed that there were 28 intersecting genes between the active ingredient target of RS and T2DM (Figure 1(a)), 10 intersecting genes between the active ingredient target of RP and T2DM (Figure 1(b)), and 8 intersecting genes among RP, RS, and T2DM (Figure 1(c)), which were the main potential targets of RP-RS for the treatment of T2DM.

3.3. Protein Interaction PPI Network Construction. In this study, we analyzed the interaction between the targets of RP and RS for the treatment of T2DM based on the STRING database, and constructed the target interaction network (Figures 2(a) and 2(b)) by importing compound-disease shared genes into STRING with the interaction confidence setting, medium confidence (0.4), and performed the network topology analysis. The top 6 nodes in the interaction network were AKT1 (26), VEGFA (25), NOS3 (24), PPARG (22), MMP9 (22), and VCAM1 (20), (Table 2, Figure 2(b)), which may play an important role in the treatment of

T2DM with RS. These 6 key targets may play an important role in the treatment of T2DM with RP.

3.4. Compound-Target Interaction Network. The network diagram visually reflects the interaction between compounds and targets, the nodes represent compounds, and the edges represent the relationship between compounds and targets, which can reflect the synergistic superposition of multicomponent and multitarget in Chinese medicine. The top 4 compounds in terms of degree value were MOL000006-luteolin (degree value = 54), MOL012297-puerarin (degree value = 53), MOL07154-tanshinone iia (degree value = 40), and MOL000392-formononetin (degree value = 32), respectively (Table 3, Figure 3).

3.5. Core Target Pathway Analysis. A total of 2015 BPs, 68 CCs, and 171 MFs of key targets were obtained by GO functional annotation, respectively ($P < 0.05$) (Figure 4). BP mainly involved response to xenobiotic stimulus, response to peptide, gland development, response to radiation, cellular response to chemical stress, response to oxygen levels, response to metal ion, response to decreased oxygen levels, response to hypoxia, and response to UV. CC is mainly membrane raft, membrane microdomain, transcription regulator complex, postsynaptic membrane, protein kinase

TABLE 4: The results of molecular docking.

Compound	Target	PDB	LibDock score
Puerarin	VEGFA	1BJ1	117.1350
Puerarin	MMP9	1GKC	133.3540
Puerarin	AKT1	1H10	79.3393
Puerarin	NOS3	1M9J	141.9620
Puerarin	VCAM1	1VSC	121.3260
Puerarin	PPARG	2VV4	0
Formononetin	VEGFA	1BJ1	90.4544
Formononetin	MMP9	1GKC	130.3760
Formononetin	AKT1	1H10	53.2793
Formononetin	NOS3	1M9J	115.5420
Formononetin	VCAM1	1VSC	93.8531
Formononetin	PPARG	2VV4	60.7435
Tanshinone iia	VEGFA	1BJ1	77.7759
Tanshinone iia	MMP9	1GKC	123.0630
Tanshinone iia	AKT1	1H10	61.7867
Tanshinone iia	NOS3	1M9J	0
Tanshinone iia	VCAM1	1VSC	96.6147
Tanshinone iia	PPARG	2VV4	87.5627
Luteolin	VEGFA	1BJ1	59.7402
Luteolin	MMP9	1GKC	138.5800
Luteolin	AKT1	1H10	61.3038
Luteolin	NOS3	1M9J	87.0491
Luteolin	VCAM1	1VSC	100.627
Luteolin	PPARG	2VV4	0

complex, caveola, serine/threonine protein kinase complex, plasma membrane raft, cyclin-dependent protein kinase holoenzyme complex, and integral component of presynaptic membrane. MF mainly involves DNA-binding transcription factor binding, RNA polymerase I-specific DNA-binding transcription factor binding, ubiquitin-like protein ligase binding, ubiquitin protein ligase binding, phosphatase binding, kinase regulator activity, protein phosphatase binding, cyclin-dependent protein serine/threonine kinase regulator activity, G protein-coupled amine receptor activity, and G protein-coupled neurotransmitter receptor activity. The KEGG analysis results showed that a total of 136 signaling pathways were enriched (Figure 5), mainly related to response to xenobiotic stimulus, cellular response to chemical stress, response to peptide, gland development, response to oxygen levels, response to radiation, response to decreased oxygen levels, response to UV, response to metal ion, response to hypoxia, response to reactive oxygen species, response to estradiol, response to oxidative stress, response to light stimulus, regulation of apoptotic signaling pathway, reproductive structure development, neuron death, reproductive system development, cellular response to oxidative stress, and cellular response to peptide. Based on the GO functional annotation and KEGG pathway enrichment analysis of these two main aspects, the RP-RS drug pair mainly treats T2DM through the synergistic effect of multiple pathways and multiple targets. Annotated map of the key

target genes locations of RP-RS in T2DM-related pathways was presented in Figure 6. It was found that most of the key target genes are closely with AGE-RAGE signaling pathway in T2DM.

3.6. Molecular Docking Prediction of Key Targets for Active Ingredient Intervention in the Treatment of T2DM by RP-RS. The key targets predicted by network pharmacology were AKT1 (1H10), VEGFA (1BJ1), NOS3 (1M9J), PPARG (2VV4), MMP9 (1GKC), and VCAM1 (1VSC), respectively. Molecular docking results of the top 4 main active ingredients and 6 core targets are shown in Figure 7 and Table 4, respectively. The key targets were individually docked to the active compounds puerarin (Figure 8), formononetin (Figure 9), tanshinone iia (Figure 10), and luteolin (Figure 11). Using Discovery Studio 2019 software, the selected compounds were finally docked into the binding site by utilizing the LibDock modules. In addition, the docked pose with the highest LibDock Score was retained for each compound for the LibDock results. The results of molecular docking showed that MMP9 had the best binding energy (LibDock Score) to puerarin, formononetin, tanshinone iia, and luteolin (133.3540, 130.3760, 123.063, and 138.5800), VEGFA to puerarin, formononetin, tanshinone iia, and luteolin (117.1350, 90.4544, 77.7759, and 59.7402), AKT1 to puerarin, formononetin, tanshinone iia, and luteolin (79.3393, 53.2793, 61.7867, and 61.3038), NOS3 to puerarin, formononetin, and luteolin (141.9620, 115.5420, and 87.0491), VCAM1 to puerarin, formononetin, tanshinone iia, and luteolin (121.3260, 93.8531, 96.6147, and 100.627), and PPARG to formononetin and tanshinone iia (60.7435 and 87.5627), respectively, and the key targets were molecularly docked to the active compounds.

4. Discussion

Current research on T2DM has focused on several pathological alterations such as obesity-related insulin resistance and defective insulin secretion as well as decreased B-cell mass through B-cell apoptosis, and has a close association with inflammatory alterations, immune gene defects, and impaired mitochondrial function [20, 21]. In this way, the results of the network pharmacological analysis, i.e., the results of the potential mechanism of RP-RS for the treatment of T2DM and the results of the molecular docking of potential pharmacodynamic substances and key targets will be discussed.

In this study, a holistic analysis was performed using network pharmacology from multigene, multipassage, and multitarget, and 92 active ingredients of RP-RS for the treatment of T2DM were identified; the most important of which were puerarin, formononetin, tanshinone iia, and luteolin. The phytochemicals are estrogenically active polyphenolic nonsteroidal phytochemicals, which are commonly found in various plants. It has been shown that formononetin has therapeutic effects on diabetes, such as formononetin exhibited hypoglycemic effects in mice with tetraoxacillin-induced type 1 diabetes by inhibiting pancreatic β -cell apoptosis [22]. It was shown that formononetin significantly increased

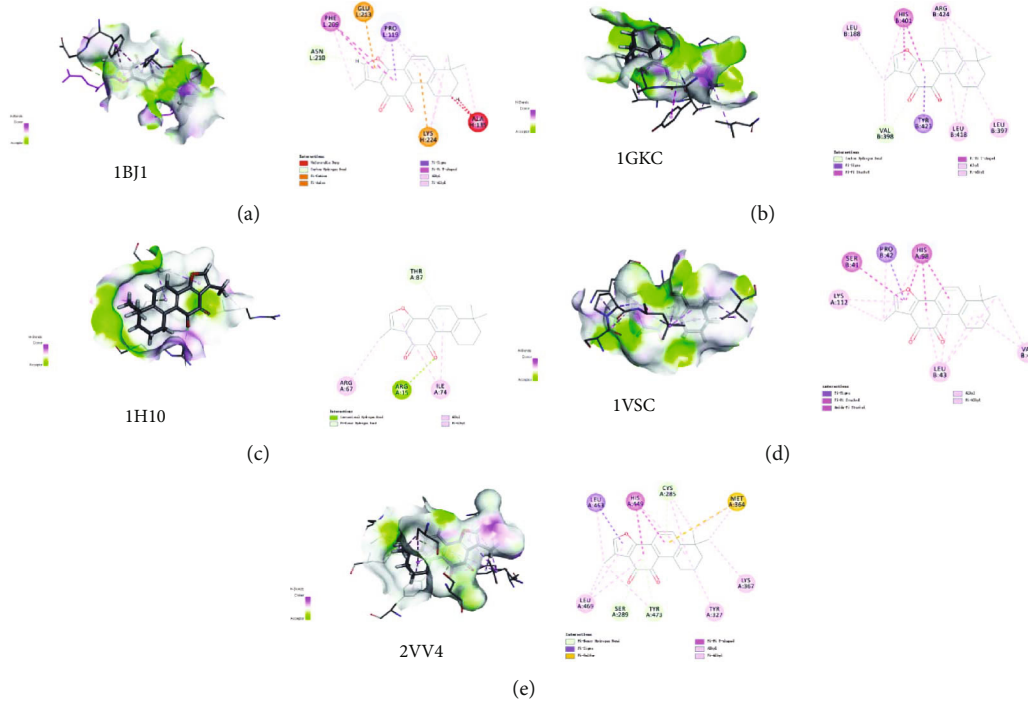


FIGURE 10: Molecular docking of tanshinone iia with core target genes. (a) VEGFA, (b) MMP9, (c) AKT1, (d) VCAM1, and (e) PPARG, respectively.

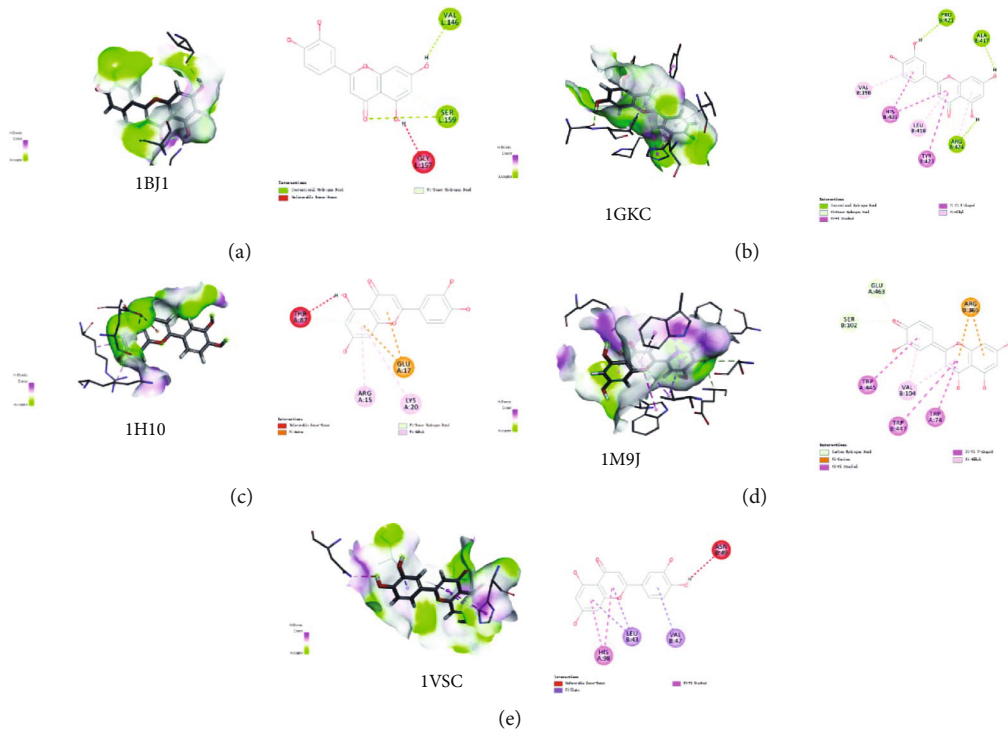


FIGURE 11: Molecular docking of luteolin with core target genes. (a) VEGFA, (b) MMP9, (c) AKT1, (d) NOS3, and (e) VCAM1, respectively.

insulin sensitivity index, decreased HOMA-IR, and improved insulin resistance [23]. Puerarin, as its main active ingredient in the treatment of T2DM may participate in the whole process of inflammatory factor expression in type 2

diabetes patients, reduce the inflammatory response, regulate the body’s internal environment, and improve the state of insulin resistance and disorders of glucose and lipid metabolism, while the symptoms of type 2 diabetes can be

improved [24]. It was reported that puerarin may also ameliorated streptozotocin (STZ) pancreatic injury in mice, upregulated insulin receptor substrate 1 and insulin-like growth factor protein expression in the pancreas, inhibited STZ-induced apoptosis of islet β -cells in diabetic mice, and increased serum insulin levels; and its protective effects on β -cells may be mediated by modulating the phosphatidylinositol 3-kinase/protein kinase B pathway, thereby exerting hypoglycemic effects and improving glucose tolerance [25]. In previous study, puerarin may also upregulated the gene expression of retinal vascular endothelial growth factor and hypoxia-inducible factor-1, which had a significant protective effect against STZ-induced diabetic retinopathy in rats, while VEGF, an angiogenic and vascular permeability factor, was significantly increased in the vitreous and aqueous fluids of eyes with proliferative diabetic retinopathy [26]. In addition, tanshinone iia has been reported to antagonize endothelial damage and antioxidant effects to effectively reduce diabetic nephropathy [27] and diabetic neuropathy [28]. Previous studies have shown that luteolin inhibits high-glucose-induced activation of NF- κ B in human monocytes and the release of the proinflammatory factor TNF- α , with potential preventive and therapeutic activity in diabetes mellitus [29]. It was revealed that have a regulatory effect on endothelial cell function, and luteolin enhance insulin action in adipocytes by activating the PPAR pathway [30]. Recent studies have revealed the potential protective effect of luteolin on diabetes-related hypertension, which may significantly reduce diabetes-induced vascular complications and hypertension [31]. Therefore, these components of RP-RS may play an important role in the prevention and treatment of T2DM.

Core targets including AKT1, VEGFA, NOS3, PPARG, MMP9, and VCAM1 have been identified as important by network pharmacological analysis. AKT1 is involved in several regulatory processes, including glucose metabolism, and AKT mediates insulin signaling and interacts with the transcription factors PGC-1 α and FoxO1 to stimulate gluconeogenic gene expression [32]. AKT is part of the insulin-signaling pathway that directly inhibits the expression of PGC-1 α protein in hepatocytes [33]. This pathway is able to regulate lipid secretion in type 2 diabetes. Because of AKT-induced inhibition of PGC-1 α leads to inhibition of fatty acid oxidation in the liver, promoting PGC-1 α activity in insulin-resistant hepatocytes may be able to eliminate lipid imbalance in T2DM patients [34]. It was reported that VEGFA promotes angiogenesis, and abnormal levels of VEGFA expression can exacerbate pathological angiogenesis and the development of diabetic retinopathy [35]. It was shown that reduced local production of VEGFA in the glomeruli of diabetic mouse models promotes endothelial injury and accelerates the progression of glomerular injury. It suggests that upregulation of VEGFA in diabetic kidneys protects microvasculature from injury, while decreased VEGFA in diabetes may be detrimental [36]. The NOS3 gene is an important 21-22 kb long gene located in vascular endothelial cells and contains 26 exons and 25 introns [37]. It has been reported that endothelial cells produce more oxygen radicals and release large amounts of Ca²⁺ in

the early stages of diabetes, which leads to increased intracellular flow and activation of NOS, causing more endothelial cells to be produced. The end effector molecule of B-cell damage is NO, and NO can cause apoptosis by inducing endoplasmic reticulum stress-related apoptotic factors, indirectly damaging islet cells, and affecting insulin synthesis, and the activity of NOS will be reduced in diabetic patients in the later stages, while plasma NO levels will be reduced [38]. PPARG, a member of the nuclear hormone receptor superfamily, which has an important role in controlling lipid and glucose metabolism as well as in T2DM development [39]. Previous studies have shown that high expression of PPARG levels in adipose tissue reduces plasma lipid levels, has a beneficial role in long-term glucolipid homeostasis, reduces the incidence of visceral adipose IR, and is involved in regulating the pathological process of T2DM in obese populations [40]. DM is not only a metabolic disorder, activation of innate immunity and inflammatory response play an important role in the development of diabetes and its onset. MMP9 is a vital effector molecule of inflammatory cells; it will act as a switch in acute and chronic inflammation and is presumptively concerned within the initial part of inflammation and later tissue remodeling [41]. Hyperglycemia promotes the production of proinflammatory cytokines TNF- α and IL-6 production, and the expression of these proinflammatory cytokines induces an increase in MMP9 expression in an autocrine and paracrine manner [42]. Studies have demonstrated that the level of MMP9 directly affects the development of diabetic nephropathy, a complication of diabetes, and that high glucose downregulates the expression of MMP9 protein, thereby affecting its proportional imbalance and also exacerbating the development of diabetic nephropathy [43]. VCAM1 is a cell adhesion molecule that is a member of the immunoglobulin superfamily, and Guillén-Gómez et al. found that urinary VCAM1 levels were significantly higher in DN patients compared to diabetic patients, which could be a marker of renal pathology in diabetic patients [44]. In addition, VCAM1 protein level was significantly correlated with T2DM complications. The enhanced induction of VCAM1 expression in endothelial cells by circulating factors may play a role in the development of atherosclerosis in diabetes [45].

Through GO functional annotation analysis of the intersection target genes, it was found that RP-RS may regulate T2DM through various BP, cellular composition CC and MF. The GO enrichment analysis showed that RP-RS exhibits efficacy mainly via xenobiotic stimulus, response to peptide, response to oxygen levels, membrane raft, membrane microdomain, transcription regulator complex, DNA-binding transcription factor binding, RNA polymerase I-specific DNA-binding transcription factor binding, ubiquitin-like protein ligase binding, etc. The KEGG pathway enrichment results revealed that a total of 136 signaling pathways were enriched, mainly related to response to xenobiotic stimulus, cellular response to chemical stress, response to peptide, gland development, response to oxygen levels, and so on. The results of molecular docking validation showed that the key components screened by network

pharmacology showed stable binding to the core targets NOS3 and MMP9, among which, the docking scores of puerarin and NOS3 were the highest, followed by those of luteolin and MMP9, and the stronger the interaction, the more stable the conformation of the compounds, indicating that puerarin showed good effects in activating NOS3 and luteolin exhibited good effects in activating MMP9 pathways, respectively.

5. Conclusion

In conclusion, a total of 92 active ingredients of RP-RS were obtained. The active components of RP-RS mainly including puerarin, formononetin, tanshinone iia, and luteolin which are closely associated with T2DM. A total of 29 intersecting target genes was acquired between drugs (RP-RS) and diseases (T2DM). Among them, the core target genes in treatment of T2DM mainly including VEGFA, MMP9, AKT1, NOS3, VCAM1, and PPARG, respectively. In addition, puerarin showed significant anti-T2DM effects in activating NOS3 and luteolin exhibited significant anti-T2DM effects in activating MMP9 pathways, respectively. The biological function exhibits efficacy mainly via positive regulation of xenobiotic stimulus, response to peptide, response to oxygen levels, membrane raft, and membrane microdomain. The signaling pathways that exerting their therapeutic effect on T2DM mainly including response to xenobiotic stimulus, cellular response to chemical stress, response to peptide, gland development, and response to oxygen levels. From the perspective of modern molecular biology, it was confirmed that RP-RS has good therapeutic effect on T2DM, the multitarget and multipath effects of TCM may rely on its therapeutic characteristics.

Abbreviations

RP:	<i>Pueraria Lobata Radix</i>
RS:	<i>Salviae Miltiorrhizae Radix</i>
RP-RS:	<i>Pueraria Lobata Radix</i> (RP) and <i>Salviae Miltiorrhizae Radix</i>
T2DM:	Type 2 diabetes mellitus
TCMSP:	Traditional Chinese medicine system pharmacology
GO:	Gene Ontology
KEGG:	Kyoto Encyclopedia of Genes and Genomes
PPI:	Protein-protein interaction
TCM:	Traditional Chinese medicine
DM:	Diabetes mellitus
OB:	Bioavailability
DL:	Drug-like properties
BP:	Biological process
CC:	Cellular component
MF:	Molecular function
STZ:	Streptozotocin.

Data Availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Conflicts of Interest

All authors declare that there have no any commercial or associative interest that represents competing interests in connection with the work submitted.

Authors' Contributions

YanLi and Zhaojing Zhu conceived and designed the research. Jingxin Mao, Guowei Wang, Lin Yang, Lihong Tan, and Cheng Tian carried out the data analysis and wrote the paper in the study. Jingxin Mao, Yan Li, Lijing Tang, and Zhenqiang Mu finished the drawing and manuscript revision work. Yan Li and Zhaojing Zhu contributed equally to the manuscript. Jingxin Mao and Guowei Wang contributed equally to this work.

Acknowledgments

This work was supported by the General Program of Chongqing Natural Science Foundation (No. cstc2021jcyjmsxmX0452), the Scientific Research and Seedling Breeding Project of Chongqing Medical Biotechnology Association (cmba2022kyym-zkxmQ0003), the Project of Science and Technology Research Program of Chongqing Education Commission of China (No. KJQN202002801), the Talent introduction project of Chongqing Medical and Pharmaceutical College (No. ygz2022101), and the 2020 Ministerial Project of China (No. 2020YYCXCQSJ050), respectively.

References

- [1] J. Liu, Z. H. Ren, H. Qiang et al., "Trends in the incidence of diabetes mellitus: results from the global burden of disease study 2017 and implications for diabetes mellitus prevention," *BMC Public Health*, vol. 20, no. 1, pp. 1–12, 2020.
- [2] G. Ning, W. Zhao, and W. Wang, "Prevalence and control of diabetes in Chinese adults," *JAMA*, vol. 310, no. 9, pp. 948–959, 2013.
- [3] R. A. DeFronzo, E. Ferrannini, L. Groop et al., "Type 2 diabetes mellitus," *Nature Reviews Disease primers*, vol. 1, no. 1, pp. 1–22, 2015.
- [4] F. Petrak, S. Herpertz, E. Stridde, and A. Pfützner, "Psychological insulin resistance in type 2 diabetes patients regarding oral antidiabetes treatment, subcutaneous insulin injections, or inhaled insulin," *Diabetes Technology & Therapeutics*, vol. 15, no. 8, pp. 702–710, 2013.
- [5] E. Xiao and L. Luo, "Alternative therapies for diabetes: a comparison of western and traditional Chinese medicine (TCM) approaches," *Current Diabetes Reviews*, vol. 14, no. 6, pp. 487–496, 2018.
- [6] H. Zhang, C. Tan, H. Wang, S. Xue, and M. Wang, "Study on the history of traditional Chinese medicine to treat diabetes," *European Journal of Integrative Medicine*, vol. 2, no. 1, pp. 41–46, 2010.
- [7] W.. Feng, H. Ao, C. Peng, and D. Yan, "Gut microbiota, a new frontier to understand traditional Chinese medicines," *Pharmacological Research*, vol. 142, pp. 176–191, 2019.

- [8] L. Yang, J. Chen, H. Lu et al., "Pueraria lobata for diabetes mellitus: past, present and future," *The American Journal of Chinese Medicine*, vol. 47, no. 7, pp. 1419–1444, 2019.
- [9] S. U. Chun-Yan, M. I. N. G. Qian-Liang, K. Rahman, H. A. N. Ting, and Q. I. N. Lu-Ping, "Salvia miltiorrhiza : traditional medicinal uses, chemistry, and pharmacology," *Chinese Journal of Natural Medicines*, vol. 13, no. 3, pp. 163–182, 2015.
- [10] W. Niu, J. Miao, X. Li, Q. Guo, Z. Deng, and L. Wu, "Metabonomics combined with systematic pharmacology reveals the therapeutic effects of Salvia miltiorrhiza and Radix Pueraria lobata herb pair on type 2 diabetes rats," *Journal of Functional Foods*, vol. 89, article 104950, 2022.
- [11] W. Zhao, Y. Yuan, H. Zhao, Y. Han, and X. Chen, "Aqueous extract of Salvia miltiorrhiza Bunge-Radix Puerariae herb pair ameliorates diabetic vascular injury by inhibiting oxidative stress in streptozotocin-induced diabetic rats," *Food and Chemical Toxicology*, vol. 129, pp. 97–107, 2019.
- [12] Z. Zhang, T. N. Lam, and Z. Zuo, "Radix Puerariae: an overview of its chemistry, pharmacology, pharmacokinetics, and clinical use," *The Journal of Clinical Pharmacology*, vol. 53, no. 8, pp. 787–811, 2013.
- [13] H. Pang, L. Wu, Y. Tang, G. Zhou, C. Qu, and J. A. Duan, "Chemical analysis of the herbal medicine Salviae miltiorrhizae radix et Rhizoma (Danshen)," *Molecules*, vol. 21, no. 1, p. 51, 2016.
- [14] L. Ni, X. Chen, X. Gong et al., "Patent information analysis of TCM prescription for the treatment of diabetes based on patent analysis and SWOT model," *Phytomedicine Plus*, vol. 2, no. 3, article 100307, 2022.
- [15] D. W. S. Cheung, C. M. Koon, C. F. Ng et al., "The roots of Salvia miltiorrhiza (Danshen) and Pueraria lobata (Gegen) inhibit atherogenic events: A study of the combination effects of the 2-herb formula," *Journal of Ethnopharmacology*, vol. 143, no. 3, pp. 859–866, 2012.
- [16] P. Hao, F. Jiang, J. Cheng, L. Ma, Y. Zhang, and Y. Zhao, "Traditional Chinese medicine for cardiovascular disease: evidence and potential mechanisms," *Journal of the American College of Cardiology*, vol. 69, no. 24, pp. 2952–2966, 2017.
- [17] J. O. Kim, K. S. Kim, G. D. Lee, and J. H. Kwon, "Antihyperglycemic and antioxidative effects of new herbal formula in streptozotocin-induced diabetic rats," *Journal of Medicinal Food*, vol. 12, no. 4, pp. 728–735, 2009.
- [18] J. Zhang, Z. Xing, M. Ma et al., "Gene ontology and KEGG enrichment analyses of genes related to age-related macular degeneration," *BioMed Research International*, vol. 2014, Article ID 450386, 10 pages, 2014.
- [19] L. Chen, C. Chu, J. Lu, X. Kong, T. Huang, and Y. D. Cai, "Gene ontology and KEGG pathway enrichment analysis of a drug target-based classification system," *PLoS One*, vol. 10, no. 5, article e0126492, 2015.
- [20] M. S. H. Akash, K. Rehman, and S. Chen, "Role of inflammatory mechanisms in pathogenesis of type 2 diabetes mellitus," *Journal of Cellular Biochemistry*, vol. 114, no. 3, pp. 525–531, 2013.
- [21] F. H. J. Van Tienen, S. F. Praet, H. M. De Feyter et al., "Physical activity is the key determinant of skeletal muscle mitochondrial function in type 2 diabetes," *The Journal of Clinical Endocrinology & Metabolism*, vol. 97, no. 9, pp. 3261–3269, 2012.
- [22] G. Qiu, W. Tian, M. Huan, J. Chen, and H. Fu, "Formononetin exhibits anti-hyperglycemic activity in alloxan-induced type 1 diabetic mice," *Experimental Biology and Medicine*, vol. 242, no. 2, pp. 223–230, 2017.
- [23] M. J. Oza and Y. A. Kulkarni, "Formononetin treatment in type 2 diabetic rats reduces insulin resistance and hyperglycemia," *Frontiers in Pharmacology*, vol. 9, p. 739, 2018.
- [24] X. Pan, J. Wang, Y. Pu, J. Yao, and H. Wang, "Effect of puerarin on expression of ICAM-1 and TNF- α in kidneys of diabetic rats," *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, vol. 21, pp. 2134–2140, 2015.
- [25] Z. Li, Z. Shangguan, Y. Liu et al., "Puerarin protects pancreatic β -cell survival via PI3K/Akt signaling pathway," *Journal of Molecular Endocrinology*, vol. 53, no. 1, pp. 71–79, 2014.
- [26] L. N. Hao, M. Wang, J. L. Ma, and T. Yang, "Puerarin decreases apoptosis of retinal pigment epithelial cells in diabetic rats by reducing peroxynitrite level and iNOS expression," *Acta Physiologica Sinica*, vol. 64, no. 2, pp. 199–206, 2012.
- [27] G. Chen, X. Zhang, C. Li, Y. Lin, Y. Meng, and S. Tang, "Role of the TGF β /p65 pathway in tanshinone IIA-treated HBZY-1 cells," *Molecular Medicine Reports*, vol. 10, no. 5, pp. 2471–2476, 2014.
- [28] Y. Liu, L. Wang, X. Li, C. Lv, D. Feng, and Z. Luo, "Tanshinone IIA improves impaired nerve functions in experimental diabetic rats," *Biochemical and Biophysical Research Communications*, vol. 399, no. 1, pp. 49–54, 2010.
- [29] H. J. Kim, W. Lee, and J. M. Yun, "Luteolin inhibits hyperglycemia-induced proinflammatory cytokine production and its epigenetic mechanism in human monocytes," *Phytotherapy Research*, vol. 28, no. 9, pp. 1383–1391, 2014.
- [30] A. C. Puhl, A. Bernardes, R. L. Silveira et al., "Mode of peroxisome proliferator-activated receptor γ activation by luteolin," *Molecular Pharmacology*, vol. 81, no. 6, pp. 788–799, 2012.
- [31] R. Sangeetha, "Luteolin in the management of type 2 diabetes mellitus," *Current Research in Nutrition and Food Science Journal*, vol. 7, no. 2, pp. 393–398, 2019.
- [32] F. S. Eshaghi, H. Ghazizadeh, S. Kazami-Nooreini et al., "Association of a genetic variant in AKT1 gene with features of the metabolic syndrome," *Genes & Diseases*, vol. 6, no. 3, pp. 290–295, 2019.
- [33] R. L. Tuttle, N. S. Gill, W. Pugh et al., "Regulation of pancreatic β -cell growth and survival by the serine/threonine protein kinase Akt1/PKB α ," *Nature Medicine*, vol. 7, no. 10, pp. 1133–1137, 2001.
- [34] H. Wang, Y. Bei, Y. Lu et al., "Exercise prevents cardiac injury and improves mitochondrial biogenesis in advanced diabetic cardiomyopathy with PGC-1 α and Akt activation," *Cellular Physiology and Biochemistry*, vol. 35, no. 6, pp. 2159–2168, 2015.
- [35] N. Sellami, L. B. Lamine, A. Turki, S. Sarray, M. Jailani, and A. K. Al-Ansari, "Association of VEGFA variants with altered VEGF secretion and type 2 diabetes: a case-control study," *Cytokine*, vol. 106, pp. 29–34, 2018.
- [36] G. A. Sivaskandarajah, M. Jeansson, Y. Maezawa, V. Eremina, H. J. Baelde, and S. E. Quaggin, "Vegfa protects the glomerular microvasculature in diabetes," *Diabetes*, vol. 61, no. 11, pp. 2958–2966, 2012.
- [37] F. Chen, Y. M. Li, L. Q. Yang, C. G. Zhong, and Z. X. Zhuang, "Association of NOS2 and NOS3 gene polymorphisms with susceptibility to type 2 diabetes mellitus and diabetic

- nephropathy in the Chinese Han population,” *IUBMB Life*, vol. 68, no. 7, pp. 516–525, 2016.
- [38] M. J. Yahya, P. B. Ismail, N. B. Nordin et al., “CNDP1, NOS3, and MnSOD polymorphisms as risk factors for diabetic nephropathy among type 2 diabetic patients in Malaysia,” *Journal of Nutrition and Metabolism*, vol. 2019, Article ID 8736215, 13 pages, 2019.
- [39] N. Sarhangi, F. Sharifi, L. Hashemian et al., “PPARG (Pro12Ala) genetic variant and risk of T2DM: a systematic review and meta-analysis,” *Scientific Reports*, vol. 10, no. 1, article 12764, 2020.
- [40] G. Abreu-Vieira, A. W. Fischer, C. Mattsson et al., “Cidea improves the metabolic profile through expansion of adipose tissue,” *Nature Communications*, vol. 6, no. 1, p. 7433, 2015.
- [41] M. Chadzinska, P. Baginski, E. Kolaczowska, H. F. Savelkoul, and B. M. Lidy Verburg-van Kemenade, “Expression profiles of matrix metalloproteinase 9 in teleost fish provide evidence for its active role in initiation and resolution of inflammation,” *Immunology*, vol. 125, no. 4, pp. 601–610, 2008.
- [42] D. Min, J. G. Lyons, J. Bonner, S. M. Twigg, D. K. Yue, and S. V. McLennan, “Mesangial cell-derived factors alter monocyte activation and function through inflammatory pathways: possible pathogenic role in diabetic nephropathy,” *American Journal of Physiology-Renal Physiology*, vol. 297, no. 5, pp. F1229–F1237, 2009.
- [43] A. Kaminari, E. C. Tsilibary, and A. Tzinia, “A new perspective in utilizing MMP-9 as a therapeutic target for Alzheimer’s disease and type 2 diabetes mellitus,” *Journal of Alzheimer’s Disease*, vol. 64, no. 1, pp. 1–16, 2018.
- [44] E. Guillén-Gómez, B. Bardaji-de-Quixano, S. Ferrer et al., “Urinary proteome analysis identified neprilysin and VCAM as proteins involved in diabetic nephropathy,” *Journal of Diabetes Research*, vol. 2018, Article ID 6165303, 12 pages, 2018.
- [45] M. M. Fadel, F. R. A. Ghaffar, S. K. Zwain, and H. M. Ibrahim, “Serum netrin and VCAM-1 as biomarker for Egyptian patients with type II diabetes mellitus,” *Biochemistry and Biophysics Reports*, vol. 27, article 101045, 2021.

Research Article

Thirdhand Smoke May Promote Lung Adenocarcinoma Development through HN1

Jiamin Peng,¹ Jia Shuopeng,² Huiju Wang,³ Xin Wang ,⁴ and Ligang Wang ⁵

¹Department of Clinical Laboratory, Zhejiang Tongde Hospital, Hangzhou, Zhejiang 310014, China

²Department of Clinical Trials Center, National Cancer Center/National Clinical Research Center for Cancer/Hebei Cancer Hospital, Chinese Academy of Medical Sciences, Langfang 065001, China

³Key Laboratory of Gastroenterology of Zhejiang Province, Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College, Hangzhou, China

⁴Department of Clinical Trials Center, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100041, China

⁵Cancer Center, Department of Ultrasound Medicine, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, China

Correspondence should be addressed to Xin Wang; xiaowuxian2006@126.com and Ligang Wang; wangligang@hmc.edu.cn

Received 18 October 2022; Revised 31 October 2022; Accepted 24 November 2022; Published 30 January 2023

Academic Editor: Tao Huang

Copyright © 2023 Jiamin Peng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Thirdhand smoke (THS) refers to residual tobacco smoking pollutants that can be adsorbed to indoor surfaces and dust and persist for years after active smoking. THS-related chemicals such as N-nitrosornicotine (NNN) and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) are tobacco-specific lung carcinogens that involved in lung cancer development and progression. In this study, we computed the differentially expressed genes (DEGs) between THS and paired control samples. THS-related overexpressed genes (OEs) were overlapped with OEs of lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). Survival analyses of these overlapped genes were performed using LUAD and LUSC data. 6 genes were selected for validation based on their expression levels and prognostic value. Hematological and neurological expressed 1 (HN1) was further selected due to its novelty in LUAD research. The potential roles of HN1 in LUAD were explored in several ways. In summary, HN1 is overexpressed in THS samples and is associated with the prognosis of patients with LUAD. It may promote cancer progression through several pathways and could serve as a potential therapeutic target especially for THS-related LUAD. In-depth mechanistic studies and clinical trials are warranted.

1. Introduction

Thirdhand smoke (THS) refers to residual tobacco smoking pollutants that can adsorbed to indoor surfaces, clothes, and dust and persist for years after active smoking [1]. These deposited chemicals such nicotine could reemit into the air and react with nitrous acid to form carcinogenic tobacco-specific nitrosamines (TSNAs) including N-nitrosornicotine (NNN) and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), which are tobacco-specific lung carcinogens that are involved in the development of lung cancer [2]. The presence of THS in indoor environment

is widespread, and it is a severely underestimated health hazard. Currently, there are no effective ways to eliminate THS.

A randomized clinical trial identified 389 genes that were differentially expressed in human respiratory epithelium in response to acute THS exposure or clean air for 3 hours. Gene ontology analysis indicated that these genes were enriched in cell stress and survival-related signaling pathways such as respiratory electron transport chain, DNA repair, and activation of cell viability [3]. It showed that human respiratory epithelium could respond rapidly to THS. Importantly, Hang et al. reported that THS could induce lung cancer development and increase lung cancer

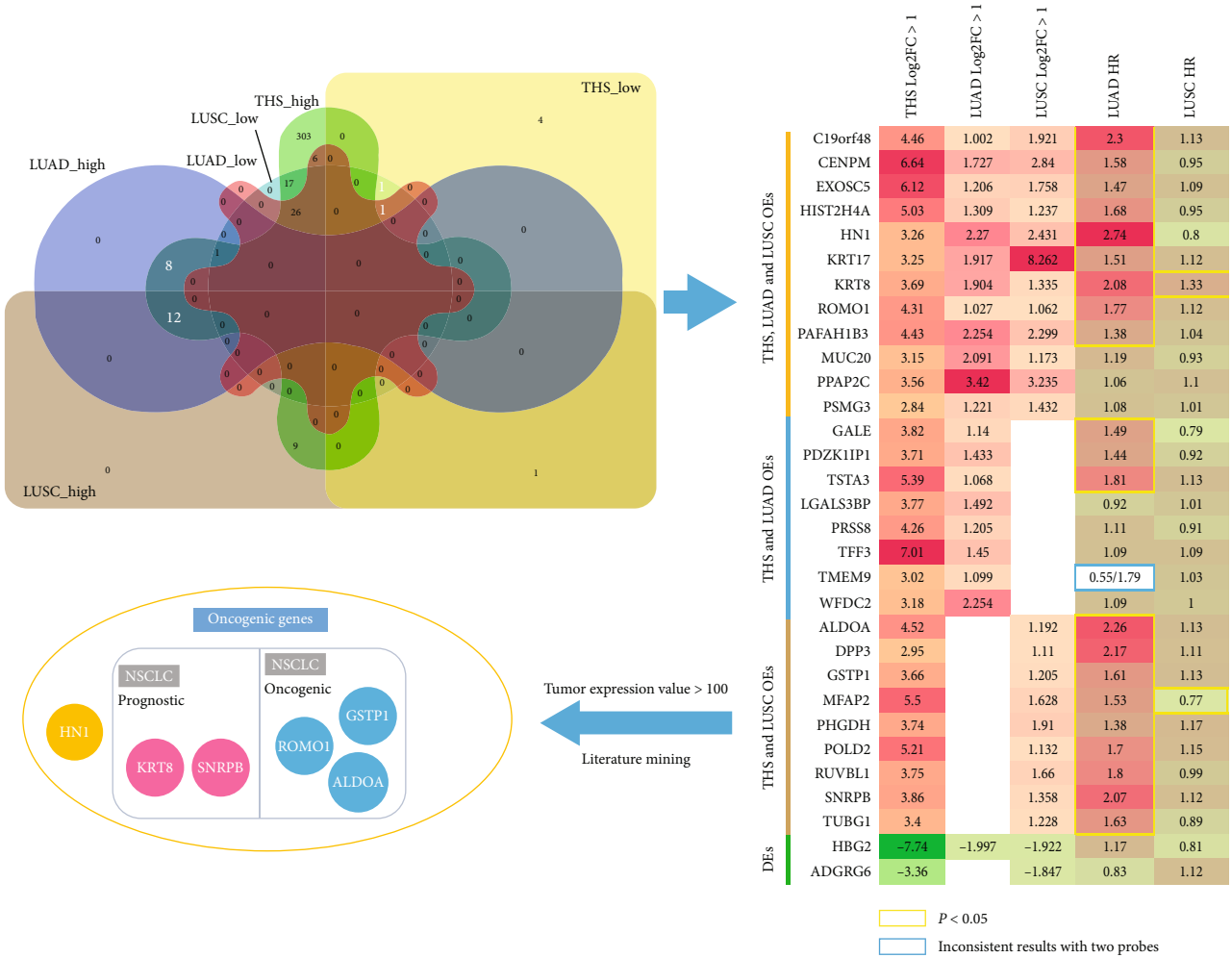


FIGURE 1: Screening for potential therapeutic targets of THS-related lung cancer.

incidence in mice [4, 5]. However, the association between THS and lung cancer is still poorly understood due to limited numbers of studies to date.

In this study, we explored the potential carcinogenesis roles of THS in human lung cancer by investigating overlapped-overexpressed genes in THS and lung cancer. This research will provide a mechanistic link between THS and human lung cancer, which would pave the way for future further investigations.

2. Results

2.1. THS-Related Genes Are more Associated with Lung Adenocarcinoma (LUAD) Not Lung Squamous Carcinoma (LUSC). Of 382 THS-related overexpressed genes (OEs), 12 genes are overexpressed both in LUAD and LUSC, 8 genes are overexpressed in LUAD, and 9 genes are overexpressed in LUSC. Of 7 THS-related downexpressed genes (DEs), one gene is downexpressed both in LUAD and LUSC while another gene is only downexpressed in LUSC (Figure 1, upper left). As shown in the heat map (Figure 1, upper right), 21 genes are associated with poor overall survival (OS) of LUAD. Even 18/21 genes are overexpressed in LUSC

and 9/21 genes are exclusively overexpressed in LUSC based on filter of log 2FC > 1, and only KRT8 correlated with poor OS of LUSC while MFAP2 associated with better OS. Detailed survival analysis results of 31 common DEGs in THS and lung cancer are shown in Table 1. This is very interesting since THS-related DEGs are mainly related to prognosis of LUAD.

Gene expression levels are very important for their biological functions, the 21 OS-related genes were further filtered based on their expression in LUAD, and 6 genes with expression value > 100 were obtained (middle-left). These 6 genes might play critical roles in THS-associated LUAD. Literature mining confirmed that all these 6 genes are associated with cancer. KRT8 and SNRBP are overexpressed and correlated with poor prognosis of NSCLC; ROMO1, GSTP1, and ALDOA are oncogenic in NSCLC. HN1 can promote cancer progression in other cancer types, but there is no report on its association with NSCLC.

2.2. Validation of the Prognostic Value of the 6 Genes in LUAD and Other Cancer Types. We validated the prognostic value of the 6 genes using a merged LUAD dataset from KmPlot. Survival analyses results confirmed that these 6

TABLE 1: Survival analyses of common DEGs in THS and lung cancer.

Gene symbol	LUAD		LUSC	
	Logrank P	Hazard ratio	Logrank P	Hazard ratio
C19orf48	6.60E-11	2.3 (1.78-2.97)	0.44	1.13 (0.83-1.54)
CENPM	1.30E-04	1.58 (1.25-1.99)	0.68	0.95 (0.75-1.21)
EXOSC5	1.20E-03	1.47 (1.16-1.85)	0.5	1.09 (0.86-1.37)
HIST2H4A	1.20E-05	1.68 (1.33-2.13)	0.7	0.95 (0.75-1.21)
HN1	1.00E-15	2.74 (2.12-3.54)	0.17	0.8 (0.59-1.1)
KRT17	5.70E-04	1.51 (1.19-1.9)	0.33	1.12 (0.89-1.42)
KRT8	1.60E-09	2.08 (1.63-2.66)	0.019	1.33 (1.05-1.68)
MUC20	1.60E-01	1.19 (0.93-1.51)	0.65	0.93 (0.68-1.27)
PAFAH1B3	0.0066	1.38 (1.09-1.75)	0.72	1.04 (0.82-1.32)
PPAP2C	6.30E-01	1.06 (0.84-1.33)	0.44	1.1 (0.87-1.39)
PSMG3	0.52	1.08 (0.85-1.38)	0.93	1.01 (0.74-1.38)
ROMO1	9.20E-06	1.77 (1.37-2.28)	0.49	1.12 (0.82-1.52)
GALE	0.0013	1.49 (1.17-1.9)	0.15	0.79 (0.58-1.09)
LGALS3BP	0.46	0.92 (0.73-1.15)	0.96	1.01 (0.79-1.27)
PDZK1IP1	0.0032	1.44 (1.13-1.84)	0.62	0.92 (0.68-1.26)
PRSS8	4.00E-01	1.11 (0.88-1.39)	0.45	0.91 (0.72-1.16)
TFF3	0.46	1.09 (0.87-1.38)	0.46	1.09 (0.86-1.39)
TMEM9	1.40E-06	0.55 (0.43-0.7)	0.87	1.03 (0.75-1.4)
TSTA3	8.40E-07	1.81 (1.43-2.3)	0.32	1.13 (0.89-1.43)
WFDC2	4.70E-01	1.09 (0.86-1.38)	0.98	1 (0.79-1.27)
ALDOA	2.50E-11	2.26 (1.77-2.89)	0.31	1.13 (0.89-1.43)
DPP3	1.60E-09	2.17 (1.68-2.81)	0.52	1.11 (0.81-1.51)
GSTP1	1.00E-04	1.61 (1.26-2.05)	0.31	1.13 (0.89-1.43)
MFAP2	3.90E-04	1.53 (1.21-1.94)	0.031	0.77 (0.61-0.98)
PHGDH	0.0059	1.38 (1.1-1.75)	0.2	1.17 (0.92-1.48)
POLD2	1.10E-05	1.7 (1.34-2.15)	0.24	1.15 (0.91-1.46)
RUVBL1	1.20E-06	1.8 (1.42-2.29)	0.9	0.99 (0.78-1.25)
SNRPB	2.30E-09	2.07 (1.62-2.64)	0.33	1.12 (0.89-1.42)
TUBG1	4.70E-05	1.63 (1.29-2.08)	0.33	0.89 (0.7-1.13)
HBG2	0.19	1.17 (0.93-1.47)	0.08	0.81 (0.64-1.03)
ADGRG6	0.14	0.83 (0.65-1.06)	0.49	1.12 (0.82-1.52)

genes are all associated with poor prognosis of patients with LUAD (Figure 2, red line indicates high expression; black line indicates low expression). Specifically, HN1 (HR = 2.74, logrank $p = 1E - 15$); KRT8 (HR = 2.27, logrank $p = 1.1E - 10$); ROMO1 (HR = 1.77, logrank $p = 9.2E - 06$); GSTP1 (HR = 1.61, logrank $p = 1E - 04$); ALDOA (HR = 2.26, logrank $p = 2.5E - 11$); SNRPB (HR = 2.07, logrank $p = 2.3E - 09$).

We also explored the expression and prognostic value of these 6 genes in other types of cancer using GEPIA 2.0 webtool. As shown in Supplementary Figure 1, 6 genes were overexpressed in most cancer types, such as pancreatic cancer and colorectal cancer (heat map of log₂ transformed fold changes, red box: high expression in cancer tissue against normal control, $p < 0.05$; blue box: low expression in cancer tissue, $p < 0.05$). Supplementary Figure 2 is the hazard ratio heat map of 6 genes in 29 cancer types. The upper panel is overall survival (OS), and

the lower panel is recurrent free survival (RFS). Red box indicates association with poor prognosis while blue box indicates association with better prognosis ($p < 0.05$). As we can see in Supplementary Figure 2, these 6 genes are associated with poor prognosis of several cancer types. These results demonstrated that THS-related genes not only important in LUAD but may also play critical roles in many other cancer types.

2.3. Expression and Function Analyses of HN1 in LUAD. Literature mining showed that HN1 is the only one in the 6 genes that has not been reported its association with lung cancer. Thus, we selected HN1 to explore its potential function and oncogenic roles in LUAD. Further analyses show that HN1 expressions in different stages of LUAD were all significantly higher than in normal control (Supplementary Figure 3), and its high expression was associated with poor prognosis of patients with LUAD (Supplementary Figure 4). GSEA

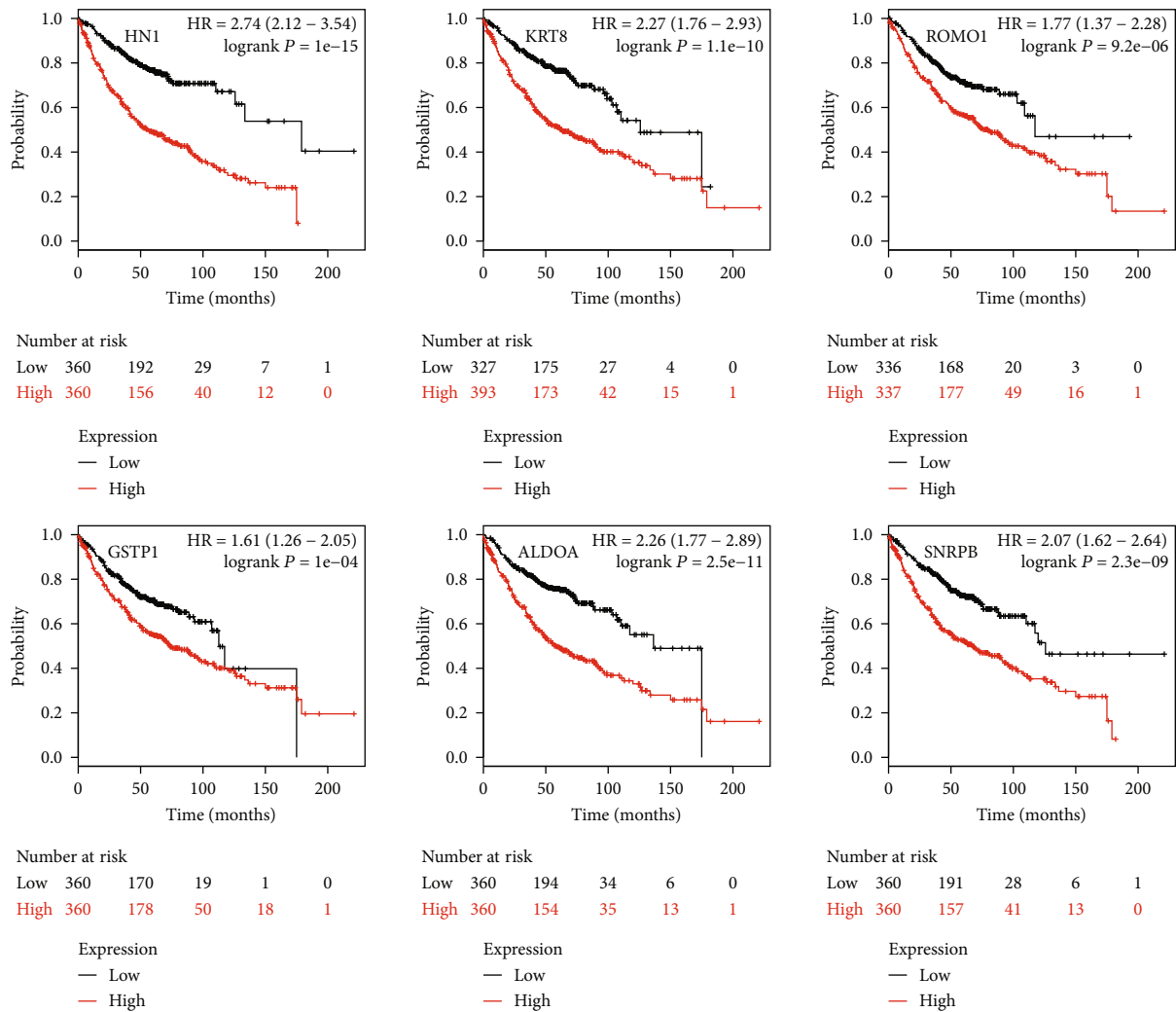


FIGURE 2: Kaplan-Meier survival analysis of HN1, KRT8, ROMO1, GSTP1, ALDOA, and SNRPB in lung adenocarcinoma. Expression levels of the 6 genes are all associated with poor prognosis of patients with lung adenocarcinoma.

analyses were performed using DEGs between HN1 high and low groups. Result shows that HN1 expression is positively correlated with embryonic stem cell signatures, metastasis, invasiveness, and hypoxia pathways (Figure 3, $p < 0.0001$, $p < 0.0001$, $p = 0.015$, and $p = 0.009$, respectively). Pathway enrichment analysis using clusterProfiler shows that HN1-related DEGs were enriched in cell adhesion molecules (CAMs), Hedgehog signaling pathway, proteoglycans in cancer, cellular senescence, PI3K-Akt signaling pathway, etc. (Supplementary Table 1). Visualization of enrichment pathway network is presented in Figure 4.

Figure 5 shows the correlation between HN1 and RPPA protein abundances. Figure 5(a) is the box plot of spearman r computed by using LUAD-HN1 expression data and RPPA data. HN1 is positively correlated with cell cycle and metastasis-related proteins such as CCNB1, CCNE1, FN1, and FOXM1, while it is negatively correlated with PI3K-AKT pathway-related proteins such as AKT pS473 and PRAS40 pT246. Figure 5(b) is the correlation map of HN1 and RPPA proteins (blue represents positive correlation

while red represents negative correlation, correlations with p value < 0.05 were presented in the map). These protein level analysis results are in accordance with the above pathway enrichment analyses.

2.4. Association Analyses of HN1 and Cancer Immunotherapy-Related Factors. The possible association between HN1 and cancer immunotherapy-related factors such as immune checkpoint blockade (ICB) therapy efficacy-related genes and abundance of cancer microenvironment cells were also investigated. Figure 6 shows the correlation between HN1 and ICB efficacy-related genes (Figure 6(a) is box plot of spearman r value; Figure 6(b) is correlation map). Figure 6(a) shows that HN1 expression is positively correlated with LDHA, LDHB, TNFSF9, TNFRSF18, IFNG, etc., while negatively correlated with JAK1, JAK2, CD40LG, etc., indicating the possibility that HN1 might serve as a prognostic biomarker or potential therapeutic target for cancer immunotherapy. Figure 6(b) is the correlation map of HN1 and ICB efficacy-related molecules. It shows the correlation clusters of these molecules in LUAD.

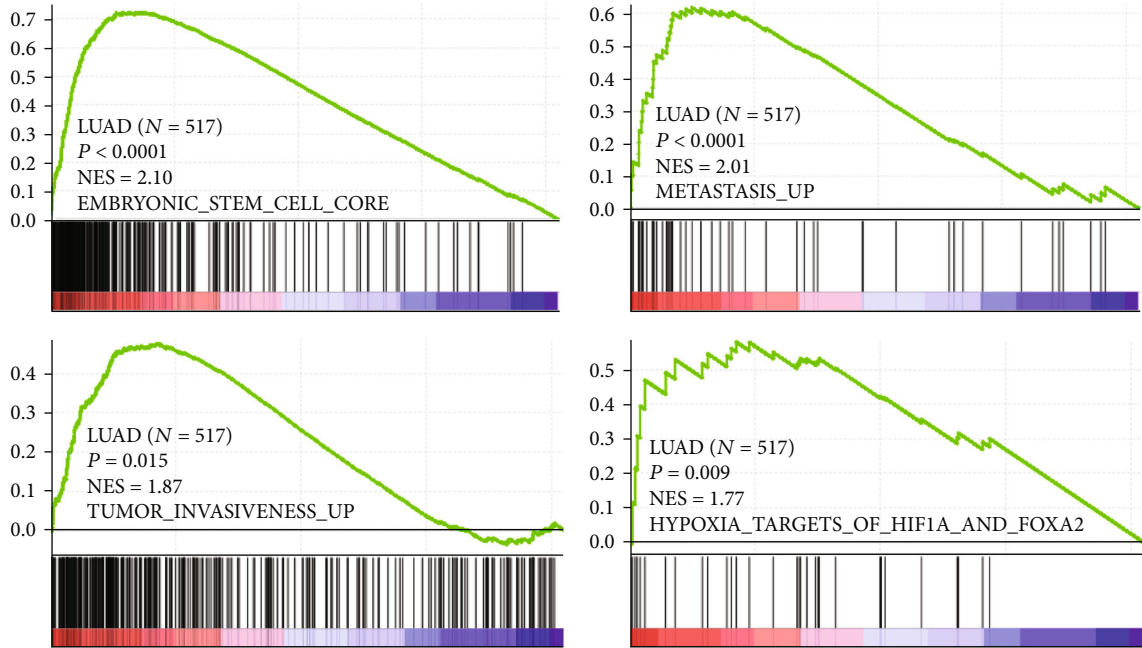


FIGURE 3: GSEA results indicate that HN1 expression is associated with stem cell, metastasis, invasiveness, and hypoxia-related pathways.

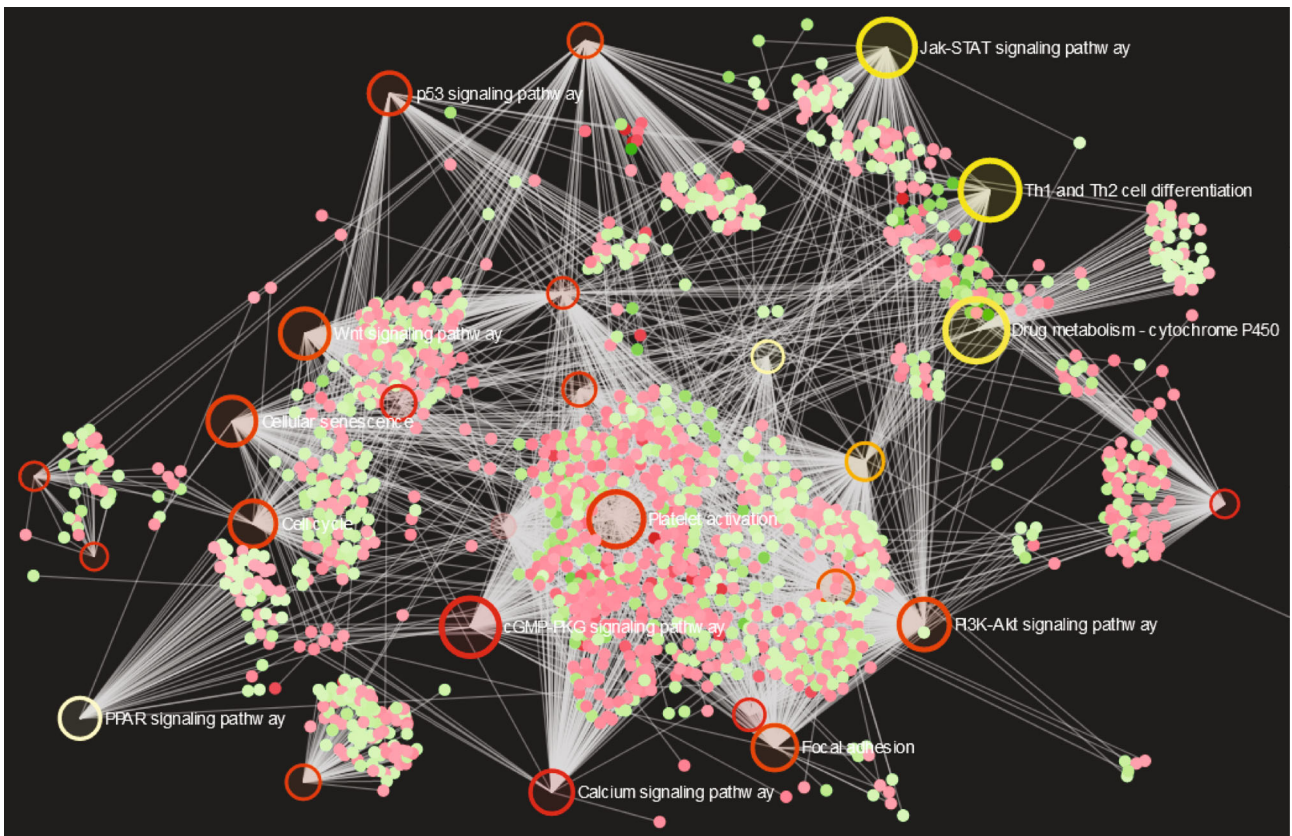
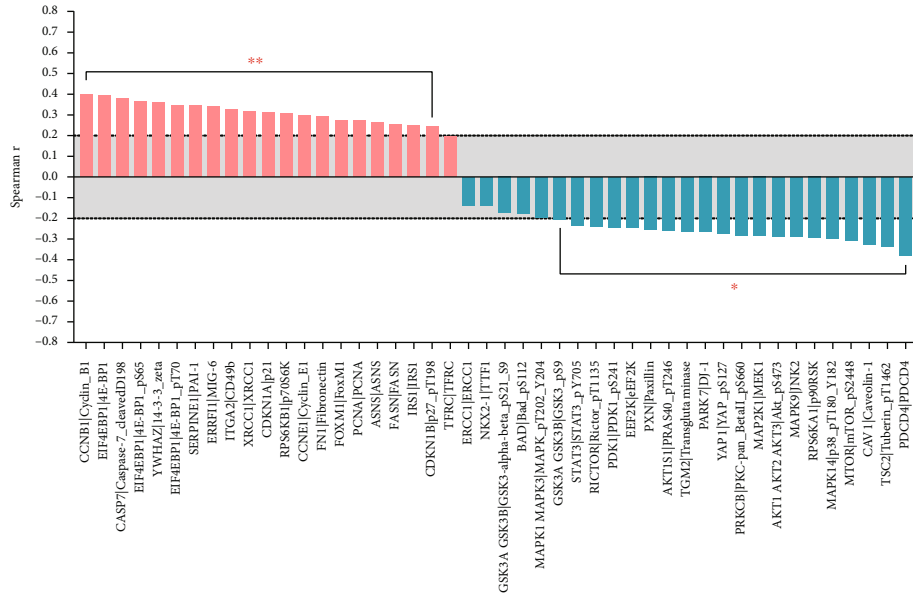


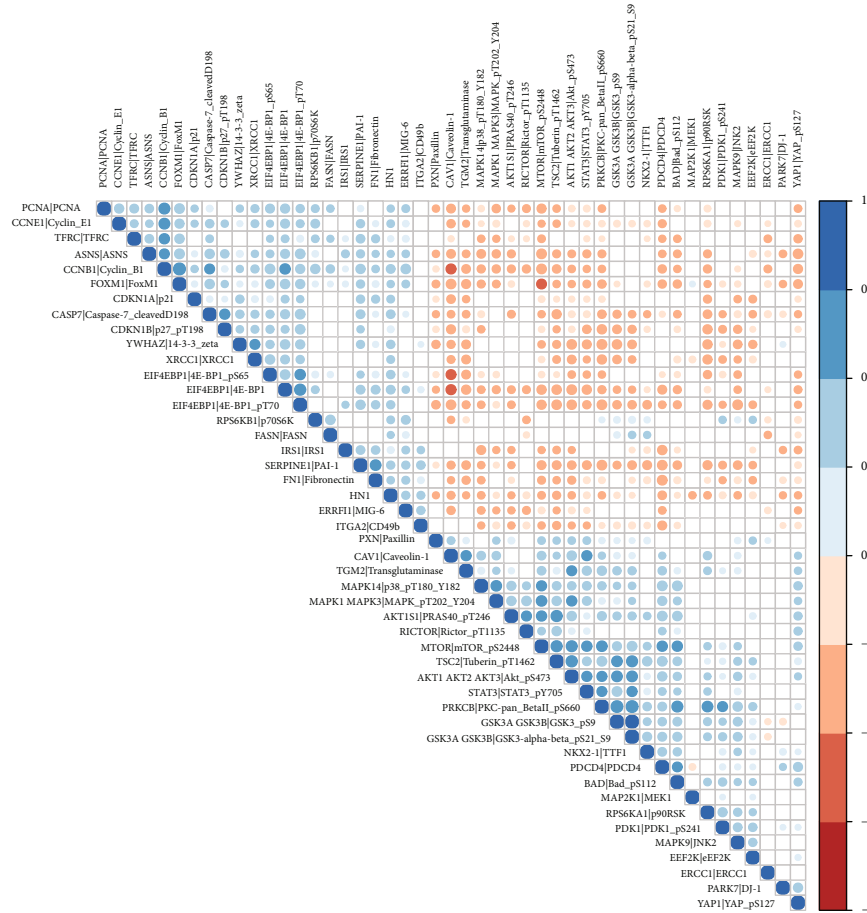
FIGURE 4: Pathway enrichment visualization of differentially expressed genes between HN1 high/low groups using NetworkAnalyst.

The correlation between HN1 and cancer microenvironment cells is presented in Figure 7 (Figure 7(a) is box plot of spearman r value; Figure 7(b) is correlation map). As we can

see from Figure 7(a), HN1 is most correlated with Th1 cells, Th2 cells, CD8 naive T cells, etc., while negatively correlated with CD4 naive T cells, fibroblasts, mast cells, etc.



(a)

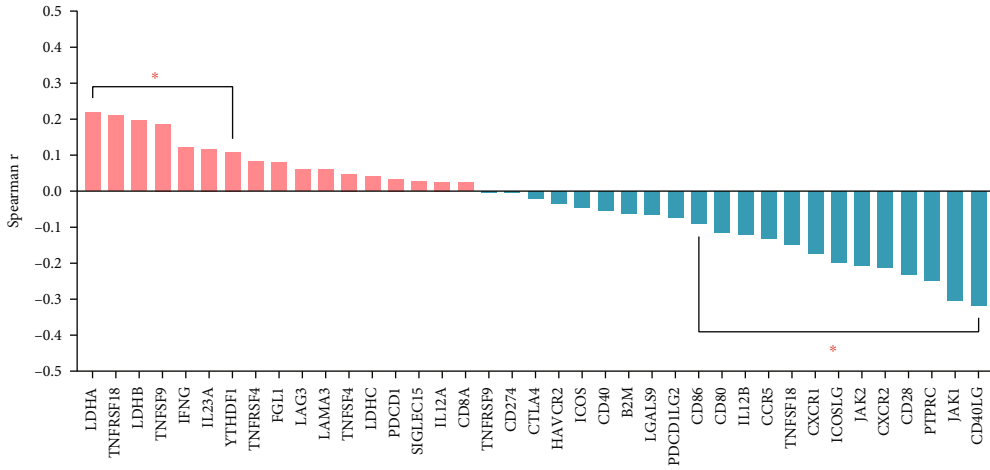


(b)

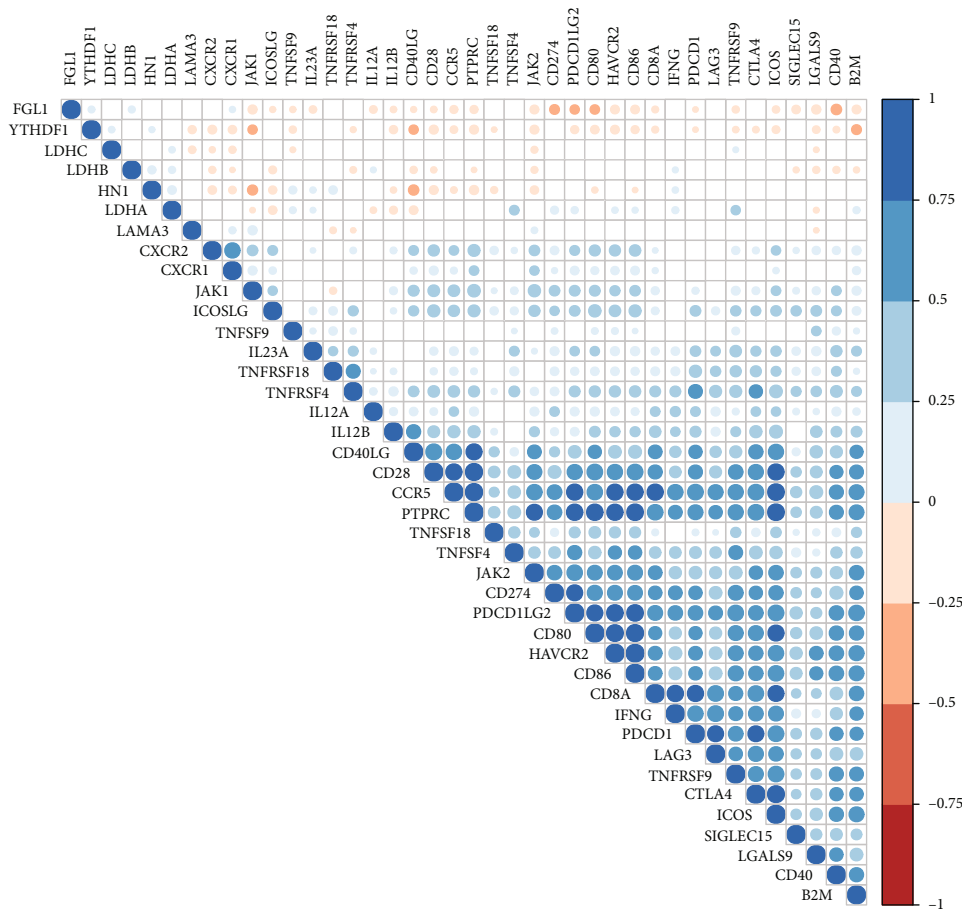
FIGURE 5: Correlation between HN1 and proteins in key pathways. (a) Box plot of spearman r value from correlation analysis between HN1 and proteins in key pathways. (b) Correlation map of HN1 with proteins in key pathways.

Figure 7(b) is the correlation map of HN1 expression and abundance of stroma cells. It shows some intensity modules or clusters of different cell types. Figure 8 is the box plot of

cancer infiltrating immune cells in HN1 high and low groups. The abundance of CD4 T cells, iTreg cells, B cells, etc. were lower in HN1 high samples, while Th1 cells, nTreg



(a)



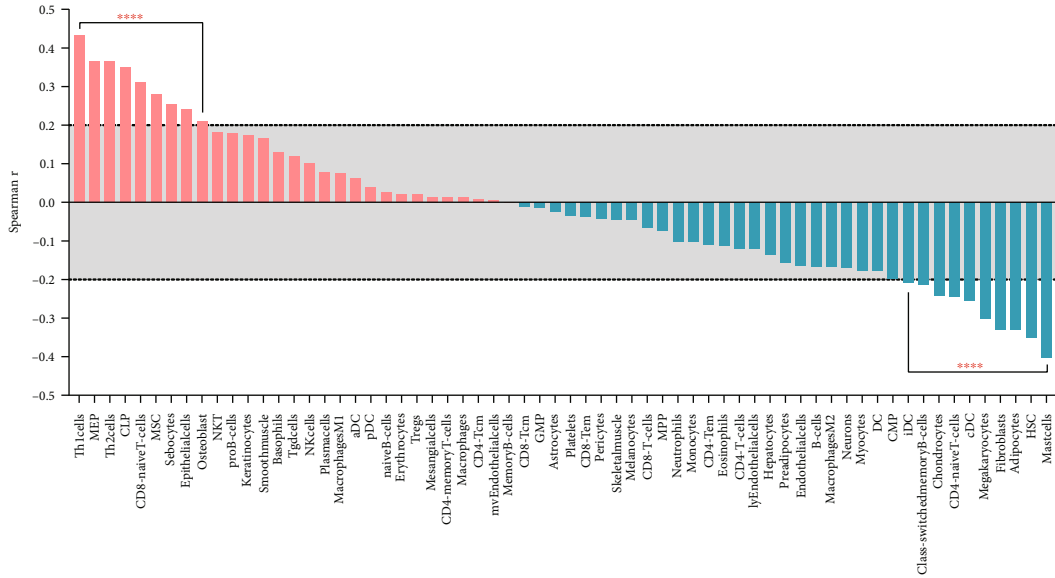
(b)

FIGURE 6: Correlation between HN1 and ICB efficacy-related genes. (a) Box plot of spearman r value from correlation analysis between HN1 and ICB efficacy-related genes. (b) Correlation map of HN1 with ICB efficacy-related genes.

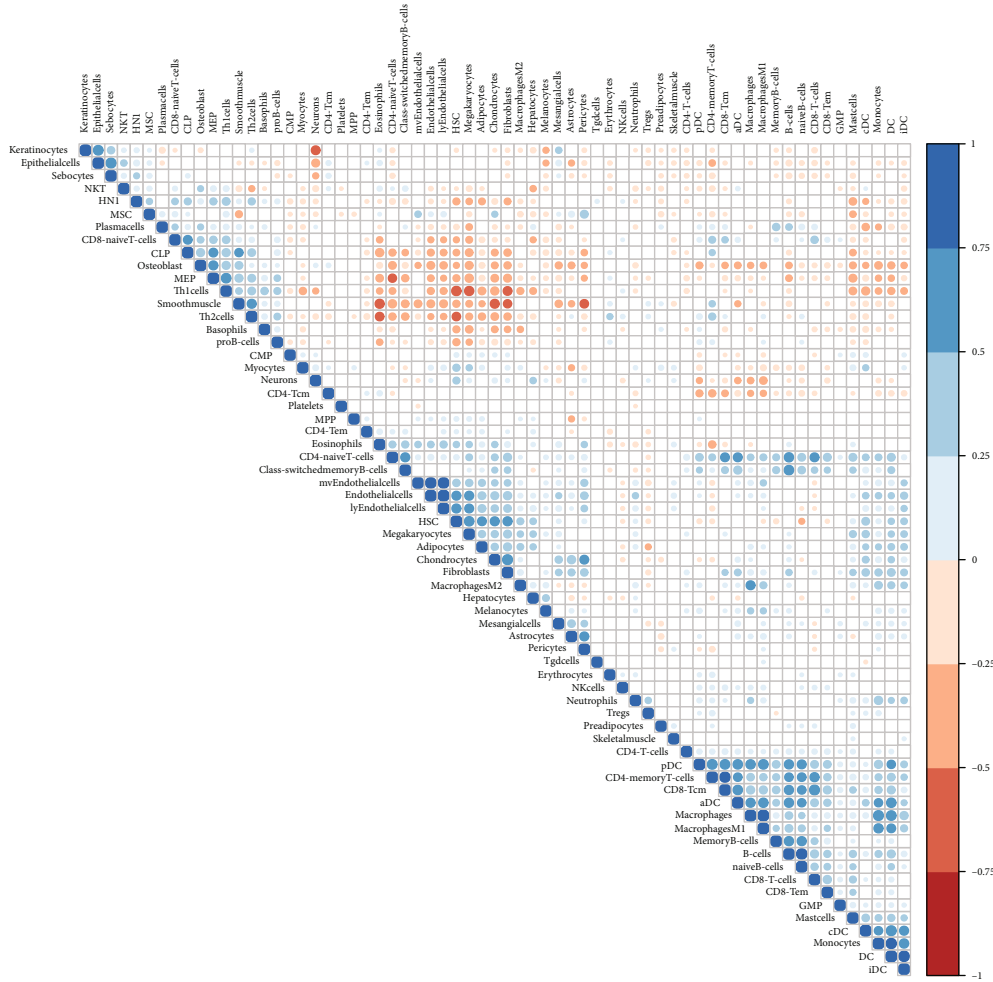
cells, etc. were higher in HN1 high samples. Correlation map and box plot of spearman r are shown in Supplementary Figure 5. These results indicated that HN1 might play important roles in cancer progression by interacting with ICB-related factors and regulating biological behaviors of cancer microenvironment cells.

3. Discussion

Lung cancer is the leading cause of cancer death and the second most commonly diagnosed cancer, with an estimated 2.2 million new cancer cases and 1.8 million deaths worldwide in 2020 [6]. Big progresses have been achieved in



(a)



(b)

FIGURE 7: Correlation between HN1 and the abundance of microenvironment cells from xCell database. (a) Box plot of spearman r value from correlation analysis between HN1 and abundance of microenvironment cells. (b) Correlation heat map of HN1 and abundance of microenvironment cells.

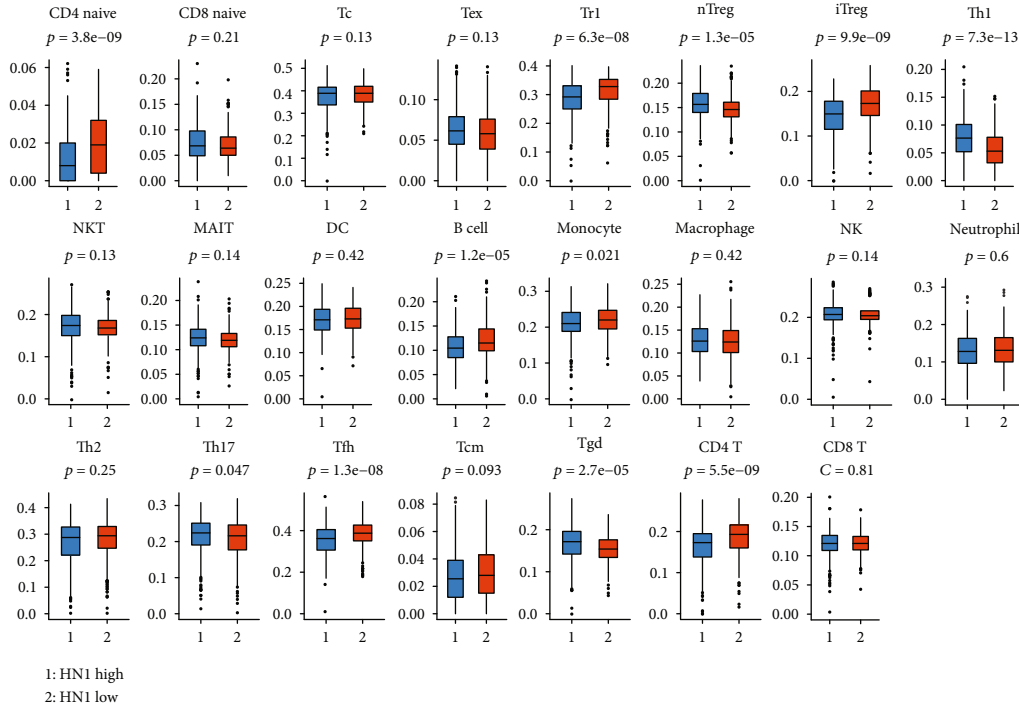


FIGURE 8: Comparison of the abundance of immune cells in HN1 high and HN1 low groups. The data are from ImmuCellAI database.

clinical practices in recent years. KRAS and EGFR-targeted drugs were applied for lung cancer patients with specific mutations [7, 8]. Immune checkpoint blockade (ICB) therapies are also promising in the management of lung cancer [9]. However, therapeutic resistance emerges rapidly after treatment posing formidable obstacles to cancer therapeutics. Clarify the molecular mechanism lung cancer carcinogenesis and searching for novel prognostic biomarkers and therapeutic targets are of crucial importance.

THS is the residual of tobacco smoke that remains in the environments after active smoking. Several reports show that THS is a great public health hazard in indoor environment [10, 11]. For instance, THS could induce damage in human DNA and stimulate high levels of inflammatory cytokines and may involve in chronic obstructive pulmonary disease and asthma [5, 12]. Previously, Hang et al. reported that early exposure to THS was associated with increased lung cancer incidences in mouse model [4].

High-dimensional biological data resource is very important in screening new biomarkers and potential therapeutic targets [13]. In this study, the potential molecular mechanisms underlying the association between THS and LUAD were also explored using public available databases. Interestingly, we found THS-related DEGs were mainly associated the prognosis of patients with LUAD not LUSC. These results were in accordance with previous reports that THS was associated with increased lung adenocarcinoma incidences in mouse model.

Based on analysis results and literature mining, we choose HN1 for further investigation. HN1 was found to be overexpressed in LUAD and could serve as a prognostic biomarker for poor prognosis of patients with LUAD. Fur-

ther analysis indicated that HN1 may promote LUAD progression by modulating tumor microenvironment and immune-related pathways. HN1, also named as JPT1, plays critical roles in the regulation of cell cycle and cell adhesion [14] and could negatively modulate AKT-mediated GSK3B signaling [15]. It has been reported that HN1 can promote tumorigenesis and metastasis in several cancer types such as breast cancer [16], prostate cancer [14], liver cancer [17], cervical cancer [18], and thyroid cancer [19]. However, this is the first report to show the association between HN1 and LUAD.

In summary, this is the first report using human data to show that THS-related transcriptional responses of respiratory epithelium are mainly associated with the development and progression of LUAD not LUSC. Exposure to THS is a significant health threat for nonsmokers, especially for children. The public should pay more attention to the potential risks of THS exposure, and potential buyers or renters of houses should be notified if there were THS risks in order to avoid unnecessary exposure to a potent lung carcinogen. Policy makers should also take THS into consideration when developing future environmental and health policies.

4. Materials and Methods

4.1. Ethics Statement. All the data were obtained from public datasets. The Research Ethics Committee of Zhejiang People’s Hospital waived the requirement for ethical approval.

4.2. Data Sources. Gene expression data and protein data of LUAD and LUSC were downloaded from The Cancer Genome Atlas (TCGA: <http://cancergenome.nih.gov/>).

THS-related gene expression data were obtained from Gene Expression Omnibus (GEO, accession no. GSE129959) [3, 20]. This dataset contains 4 participants that receive the clean air exposure first and THS exposure second. Abundance data of stroma and immune cells were obtained from xCell [21] and ImmuCellAI [22] databases.

4.3. Bioinformatics and Statistical Analyses. Survival analyses of 6 genes were analyzed online and downloaded from KmPlot [23] using a merged LUAD dataset. Differential expression and survival analyses of LUAD and LUSC TCGA data was performed using GEPIA version 2.0 [24]. UALCAN webtool was used for HN1 expression analysis in different stages of LUAD and normal control [25]. R 3.4.4 (R Foundation for statistical computing [http://www.r-project.org/]) was used for computing differentially expressed genes. Venn diagram analysis was performed using an online tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). Pathway enrichment visualization was performed using NetworkAnalyst version 3.0 [26]. Gene set enrichment analysis (GSEA) was performed using clusterProfiler version 4.0 [27] and GSEA v4.0.3 (<https://www.gsea-msigdb.org/gsea/downloads.jsp>). Correlation analysis and visualization were performed using “corrplot” package. Other statistical analyses were performed using GraphPad Prism 5.01 (GraphPad Software, Inc. (<http://www.graphpad.com>)). Adjust p value was corrected for multiple comparisons using the Benjamini and Hochberg’s false discovery rate [28]. Statistical significance was defined as a p value < 0.05 .

Data Availability

Data used in this study are all from publicly available datasets, and all data sources are noted in the manuscript.

Conflicts of Interest

The authors declare no competing financial interests.

Authors’ Contributions

Jiamin Peng and Jia Shuopeng contribute equally to this work.

Supplementary Materials

Supplementary Figure 1: expression of HN1, KRT8, ROMO1, GSTP1, ALDOA, and SNRPB in other types of tumors in comparison to normal control. Red: high expression; blue: low expression. Supplementary Figure 2: prognostic value of HN1, KRT8, ROMO1, GSTP1, ALDOA, and SNRPB in other types of tumors. Red: good prognosis; blue: poor prognosis. Supplementary Figure 3: HN1 expression in different stages of LUAD is significantly higher in comparison to normal control. Supplementary Figure 4: high expression of HN1 is associated with poor prognosis of patients with LUAD. Supplementary Figure 5: correlation between HN1 and the abundance of immune cells from ImmuCellAI database. (A) Box plot of spearman r value from correlation

analysis between HN1 and abundance of immune cells. (B) Correlation heat map of HN1 and abundance of immune cells. Supplementary Table 1: GSEA pathway enrichment of HN1-correlated genes. (*Supplementary Materials*)

References

- [1] P. F. DeCarlo, A. M. Avery, and M. S. Waring, “Thirdhand smoke uptake to aerosol particles in the indoor environment,” *Science Advances*, vol. 4, no. 5, p. eaap8368, 2018.
- [2] J. L. Thomas, S. S. Hecht, X. Luo, X. Ming, J. S. Ahluwalia, and S. G. Carmella, “Thirdhand tobacco smoke: a tobacco-specific lung carcinogen on surfaces in smokers’ homes,” *Nicotine & Tobacco Research*, vol. 16, no. 1, pp. 26–32, 2014.
- [3] G. L. Pozuelos, M. S. Kagda, S. Schick, T. Girke, D. C. Volz, and P. Talbot, “Experimental acute exposure to thirdhand smoke and changes in the human nasal epithelial transcriptome: a randomized clinical trial,” *JAMA Network Open*, vol. 2, no. 6, article e196362, 2019.
- [4] B. Hang, Y. Wang, Y. Huang et al., “Short-term early exposure to thirdhand cigarette smoke increases lung cancer incidence in mice,” *Clinical Science*, vol. 132, no. 4, pp. 475–488, 2018.
- [5] B. Hang, J. H. Mao, and A. M. Snijders, “Genetic susceptibility to thirdhand-smoke-induced lung cancer development,” *Nicotine & Tobacco Research*, vol. 21, no. 9, pp. 1294–1296, 2019.
- [6] H. Sung, J. Ferlay, R. L. Siegel et al., “Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA: a Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [7] X. Du, B. Yang, Q. An, Y. G. Assaraf, X. Cao, and J. Xia, “Acquired resistance to third-generation EGFR-TKIs and emerging next-generation EGFR inhibitors,” *The Innovation*, vol. 2, no. 2, article 100103, 2021.
- [8] D. Jiao and S. Yang, “Overcoming resistance to drugs targeting KRAS^{G12C} mutation,” *The Innovation*, vol. 1, no. 2, 2020.
- [9] Y. Ye, Y. Zhang, N. Yang et al., “Profiling of immune features to predict immunotherapy efficacy,” *The Innovation*, vol. 3, no. 1, article 100194, 2022.
- [10] J. X. Wu, A. T. Y. Lau, and Y. M. Xu, “Indoor secondary pollutants cannot be ignored: third-hand smoke,” *Toxics*, vol. 10, no. 7, p. 363, 2022.
- [11] B. Hang, A. H. Sarker, C. Havel et al., “Thirdhand smoke causes DNA damage in human cells,” *Mutagenesis*, vol. 28, no. 4, pp. 381–391, 2013.
- [12] M. Martins-Green, N. Adhami, M. Frankos et al., “Cigarette smoke toxins deposited on surfaces: implications for human health,” *PLoS One*, vol. 9, no. 1, article e86391, 2014.
- [13] Y. Xu, X. Liu, X. Cao et al., “Artificial intelligence: a powerful paradigm for scientific research,” *The Innovation*, vol. 2, no. 4, article 100179, 2021.
- [14] L. Varisli, B. E. Ozturk, G. K. Akyuz, and K. S. Korkmaz, “HN1 negatively influences the β -catenin/E-cadherin interaction, and contributes to migration in prostate cells,” *Journal of Cellular Biochemistry*, vol. 116, no. 1, pp. 170–178, 2015.
- [15] L. Varisli, C. Gonen-Korkmaz, B. Debelec-Butuner et al., “Ubiquitously expressed hematological and neurological expressed 1 downregulates Akt-mediated GSK3 β signaling, and its knockdown results in deregulated G2/M transition in prostate cells,” *DNA and Cell Biology*, vol. 30, no. 6, pp. 419–429, 2011.

- [16] Z. G. Zhang, W. X. Chen, Y. H. Wu, H. F. Liang, and B. X. Zhang, “miR-132 prohibits proliferation, invasion, migration, and metastasis in breast cancer by targeting HN1,” *Biochemical and Biophysical Research Communications*, vol. 454, no. 1, pp. 109–114, 2014.
- [17] J. J. Chen, X. Sun, Q. Q. Mao et al., “Increased expression of hematological and neurological expressed 1 (HN1) is associated with a poor prognosis of hepatocellular carcinoma and its knockdown inhibits cell growth and migration partly by down-regulation of c-Met,” *The Kaohsiung Journal of Medical Sciences*, vol. 36, no. 3, pp. 196–205, 2020.
- [18] J. Chen, J. Qiu, F. Li et al., “HN1 promotes tumor associated lymphangiogenesis and lymph node metastasis via NF- κ B signaling activation in cervical carcinoma,” *Biochemical and Biophysical Research Communications*, vol. 530, no. 1, pp. 87–94, 2020.
- [19] Z. Pan, Q. Fang, L. Li et al., “HN1 promotes tumor growth and metastasis of anaplastic thyroid carcinoma by interacting with STMN1,” *Cancer Letters*, vol. 501, pp. 31–42, 2021.
- [20] T. Barrett, S. E. Wilhite, P. Ledoux et al., “NCBI GEO: archive for functional genomics data sets—update,” *Nucleic Acids Research*, vol. 41, no. Database issue, pp. D991–D995, 2013.
- [21] D. Aran, Z. Hu, and A. J. Butte, “xCell: digitally portraying the tissue cellular heterogeneity landscape,” *Genome Biology*, vol. 18, no. 1, p. 220, 2017.
- [22] Y. R. Miao, Q. Zhang, Q. Lei et al., “ImmuCellAI: a unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy,” *Advanced Science*, vol. 7, no. 7, article 1902880, 2020.
- [23] B. Györfy, P. Surowiak, J. Budczies, and A. Lánczky, “Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer,” *PLoS One*, vol. 8, no. 12, article e82241, 2013.
- [24] Z. Tang, B. Kang, C. Li, T. Chen, and Z. Zhang, “GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis,” *Nucleic Acids Research*, vol. 47, no. W1, pp. W556–W560, 2019.
- [25] D. S. Chandrashekar, B. Bashel, S. A. H. Balasubramanya et al., “UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses,” *Neoplasia*, vol. 19, no. 8, pp. 649–658, 2017.
- [26] G. Zhou, O. Soufan, J. Ewald, R. E. W. Hancock, N. Basu, and J. Xia, “NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis,” *Nucleic Acids Research*, vol. 47, no. W1, pp. W234–W241, 2019.
- [27] T. Wu, E. Hu, S. Xu et al., “clusterProfiler 4.0: a universal enrichment tool for interpreting omics data,” *The Innovation*, vol. 2, no. 3, article 100141, 2021.
- [28] Y. Benjamini and Y. Hochberg, “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

Research Article

Analysis of the Clinical Efficacy and Molecular Mechanism of Xuefu Zhuyu Decoction in the Treatment of COPD Based on Meta-Analysis and Network Pharmacology

Yu Hu, Yunxi Lan, Qiqi Ran, Qianrong Gan, and Wei Huang 

School of Basic Medical Sciences, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

Correspondence should be addressed to Wei Huang; gracehw@126.com

Received 4 August 2022; Accepted 10 October 2022; Published 26 November 2022

Academic Editor: Lei Chen

Copyright © 2022 Yu Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Chronic obstructive pulmonary disease (COPD) is becoming a major public health burden worldwide. It is urgent to explore more effective and safer treatment strategy for COPD. Notably, Xuefu Zhuyu Decoction (XFZYD) is widely used to treat respiratory system diseases, including COPD, in China. **Objective.** This study is aimed at comprehensively evaluating the therapeutic effects and molecular mechanism of XFZYD on COPD. **Methods.** Original clinical studies were searched from eight literature databases. Meta-analysis was conducted using the Review Manager software (version 5.4.1). Network pharmacology and molecular docking experiments were utilized to explore the mechanisms of action of XFZYD. **Results.** XFZYD significantly enhanced the efficacy of clinical treatment and improved the pulmonary function and hypoventilation of COPD patients. In addition, XFZYD significantly improved the hypercoagulability of COPD patients. The subgroup analysis suggested that XFZYD exhibited therapeutic effects on both stable and acute exacerbation of COPD. XFZYD exerted its therapeutic effects on COPD through multicomponent, multitarget, and multipathway characteristics. The intervention of the PI3K-AKT pathway may be the critical mechanism. **Conclusion.** The application of XFZYD based on symptomatic relief and supportive treatment is a promising clinical decision. More preclinical and clinical studies are still needed to evaluate the safety and therapeutic effects of long-term use of XFZYD on COPD.

1. Introduction

Chronic obstructive pulmonary disease (COPD) is a chronic inflammatory disease of the respiratory system induced by long-term exposure to risk factors such as tobacco smoke, toxic gases, or polluted air [1]. COPD is mainly characterized by progressive involvement of the airways, alveoli, and microvasculature, leading to irreversible airflow limitation [2]. Emphysema and chronic bronchitis are two typical phenotypes of COPD, but are usually present in one patient simultaneously [3, 4]. Currently, COPD has a global prevalence of approximately 328 million and has become the third leading death cause after ischemic heart disease and tumors [5, 6]. Smoking is considered as a vital risk factor of COPD. Although the proportion of population that smokes has declined globally, environmental particulate matter pollution and occupational second-hand smoke exposure have

increased dramatically over the past 25 years, which may become major risk factors for COPD in the future [7, 8]. As a result, the prevalence of COPD is still increasing and is forecasted to become the fifth largest global disease burden by 2030 [9]. Additionally, age is an important risk factor for COPD [10]. The incidence risk of COPD tends to increase exponentially with age, with the middle-aged and elderly population being the main groups affected [11]. Obviously, the prevalence and mortality of COPD are gradually increasing and the burden of disease continues to grow with the aging of global society [12, 13]. There is no doubt that COPD is increasingly becoming a serious global public health concern [14].

Although the treatment of COPD has evolved and improved significantly over the past 20 years, there is no known cure for COPD but only options to improve symptoms and support health of patients [3, 15]. Moreover, it is

still dependent on bronchodilators and corticosteroids, which have significant limitations. The long-term use of corticosteroids increases the risk of adverse events in patients [16]. At the same time, the progression of emphysema cannot be inhibited by such drugs [17]. In addition, COPD patients who suffer from viral or bacterial infections or increased bacterial load in the airways, resulting in acute worsening of respiratory symptoms, are referred to as acute exacerbation of COPD (AECOPD) [9, 18]. It is noticed that efficacy of corticosteroids in AECOPD is very limited although they are the main anti-infective drugs used in COPD [19]. Therefore, there is an urgent need for alternative, more effective, and safer treatments for COPD to alleviate clinical symptoms, stop disease progression, and reverse the disease process [20].

As an important component of complementary alternative medicine, traditional Chinese medicine is believed to exert its pharmacological effects through multicomponent, multitarget, and multipathway characteristics, thus offering significant advantages in treating many chronic diseases with fewer adverse events [21]. Xuefu Zhuyu Decoction (XFZYD) is a formula from *Corrections on the Errors of Medical Works (Yi Lin Gai Cuo)* of Wang Qingren in the Qing Dynasty, comprising the dried ripe seed of *Prunus persica* (L.) Batsch (*Tao Ren*), the dried flower of *Carthamus tinctorius* L. (*Hong Hua*), the dried root of *Angelica sinensis* (Oliv.) Diels (*Dang Gui*), the fresh or dried tuberous root of *Rehmannia glutinosa* (Gaertn.) DC. (*Sheng Di*), the dried rhizome of *Conioselinum anthriscoides* “*Chuanxiong*” (*Chuan Xiong*), the dried root of *Paeonia lactiflora* Pall. (*Chi Shao*), the dried root of *Bupleurum chinense* DC. (*Chai Hu*), the dried unripe fruit of *Citrus × aurantium* L. (*Zhi Qiao*), the dried root of *Platycodon grandiflorus* (Jacq.) A.DC. (*Jie Geng*), the dried root of *Achyranthes bidentata* Blume (*Niu Xi*), and the dried root and rhizome of *Glycyrrhiza glabra* L. (*Gan Cao*), which has been used to treat cardiovascular and microvascular diseases in clinical practice [22, 23]. Several systematic evaluations have assessed the efficacy in XFZYD in chronic diseases such as uterine fibroids, pulmonary fibrosis, and coronary heart disease with positive results [24–26]. Meanwhile, a large number of clinical observational studies have focused on its therapeutic effects on COPD [27]. However, to date, there is still a lack of systematic evaluation of the therapeutic effects of XFZYD on COPD. Moreover, the specific molecular mechanisms of XFZYD in the treatment of COPD are still unclear. In the present study, the efficacy and safety of XFZYD in COPD were investigated by meta-analysis, and the therapeutic advantages of XFZYD in the stable COPD and AECOPD were further measured by subgroup analysis. Then, the network pharmacology and molecular docking experiments were used for exploring the mechanism of action of XFZYD in the treatment of COPD (Figure 1).

2. Materials and Methods

2.1. Assessment of Efficacy and Safety of XFZYD Based on Meta-Analysis. This study followed the Cochrane criteria and the Preferred Reporting Items for Systematic Review

and Meta-Analysis (PRISMA) statement guidelines [28]. Moreover, this meta-analysis protocol was registered with the PROSPERO (CRD42022304855).

2.1.1. Study Selection. The databases, including Cochrane library, PubMed, Embase, Web of Science (WOS), China Academic Journals (CNKI), Wan-fang database, Chinese biomedical literature service system (CBM), and Chinese Science and Technology Journals (CQVIP), were searched systematically before January 2022. No language and region restrictions were considered for original studies. The search terms were “Chronic obstructive pulmonary disease OR COPD” and “Xuefu Zhuyu OR Xuefuzhuyu OR XueFuZhuYu.” The detailed searching strategies are accessible in Supplementary Table 1.

2.1.2. Inclusion and Exclusion Criteria. The PICOS criteria were used for the inclusion and exclusion of original studies. The inclusion criteria are as follows: (1) participants: patients diagnosed with COPD using definite diagnostic criteria, including stable COPD and AECOPD; (2) intervention: XFZYD combined with symptomatic supportive treatments (only the whole formula of XFZYD used could be included); (3) control: symptomatic relief and supportive treatment alone; (4) outcomes: the pulmonary function indicators and clinical effective rate are the primary outcomes, the blood gas analysis results, coagulation indexes, and adverse events are the secondary outcomes; (5) study: randomized controlled trials (RCTs).

The exclusion criteria are as follows: (1) original study without any diagnostic criteria; (2) patients with COPD combined with other critical illnesses; (3) other TCM treatments, such as acupuncture and Tai Chi, were used in the treatment and control groups; (4) case reports, retrospective studies, reviews, systematic reviews, and other non-RCT studies.

2.1.3. Data Extraction. The original studies were reviewed independently by two authors (Y.L. and Q.R.). Then, the specific data were extracted, including first author, publication year, numbers of patients, average age, disease stage, outcomes, and adverse events. Differences were resolved by the discussion between two authors; differences which could not be resolved were settled by the third author (Q.G.).

2.1.4. Risk of Bias Assessment. The revised Cochrane risk-of-bias tool for randomized trials (RoB 2) was used to assess the risk of bias [29]. Specifically, five domains including randomization process, deviations from the intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result were judged as low risk of bias, some concerns, or high risk of bias based on the summaries of answers to signaling questions. Finally, an overall bias, which corresponds to the worst risk of bias in any of the domains, for every included study was judged.

2.1.5. Outcomes. Impaired pulmonary function is a critical diagnostic standard of COPD as well as an important indication of the prognosis of patients [30]. As a result, the pulmonary function indexes, including the forced expiratory volume in one second/forced vital capacity (FEV1/FVC (%)),

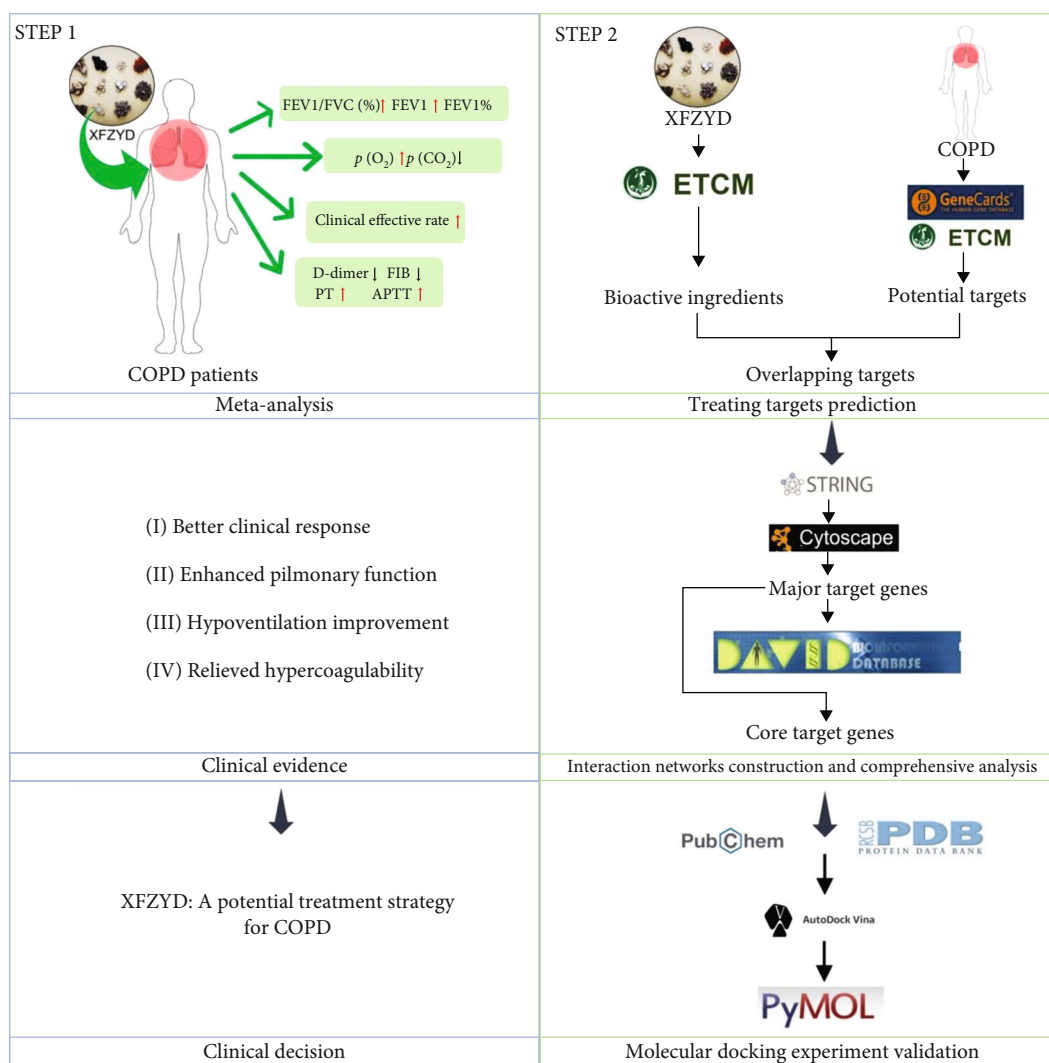


FIGURE 1: The theme of the present study.

the volume of forced expiratory volume in one second (FEV1), and the percentage of forced expiratory volume in one second (FEV1%), were settled as the primary outcomes. Similarly, the clinical effective rate was also identified as primary outcomes since it tends to be comprehensive for the evaluation of the therapeutic effects of treatment strategies. Then, the blood gas analysis results, including arterial partial pressure of oxygen ($p(O_2)$) and the arterial partial pressure of carbon dioxide ($p(CO_2)$), coagulation indexes, comprising D-dimer, fibrinogen (FIB), prothrombin time (PT), and activated partial thromboplastin time (APTT), and adverse events were the secondary outcomes.

In order to clarify the therapeutic advantage of XFZYD of different stages of COPD, we set up the subgroup analysis of these outcomes of stable COPD and AECOPD.

2.1.6. Statistical Analysis. The Review Manager (RevMan) software (version 5.4.1) was used to perform the meta-analysis. The weighted mean difference (WMD) or standardized mean difference (SMD) with a 95% confidence interval (CI) was used to assess the continuous variables.

The risk ratio (RR) or risk difference (RD) with a 95% CI were applied to analyze the dichotomous variables. Heterogeneity was estimated by Cochran's Q test and assessed using I^2 . When I^2 was less than 50%, it was considered homogeneous and is evaluated by a fixed-effects model. When I^2 was greater than 50%, it is considered heterogeneous and is assessed using a random-effects model, and the sensitivity analysis would be conducted to assess the sources for heterogeneity in primary outcomes. Finally, the funnel plots were created to check for publication bias. The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) profiler software (version 3.6) was applied to evaluate the quality of the clinical evidence, which mainly have four grades, high, moderate, low, and very low [31].

2.2. Exploration of the Molecular Mechanisms of XFZYD for COPD Based on Network Pharmacology

2.2.1. Collection of the Bioactive Ingredients and Potential Targets of XFZYD. The bioactive ingredients of XFZYD were searched from the Encyclopedia of Traditional Chinese

Medicine (ETCM, <http://www.tcmip.cn/ETCM/>) database [32]. Ingredients in ETCM database were mainly assessed by the quantitative estimate of drug-likeness (QED) score based on models in the Pipeline Pilot ADMET collection. Therefore, the ingredients with moderate to good-likeness ($QED > 0.49$) were considered as bioactive ingredients. Then, the targets of bioactive components, which were predicted by the MedChem Studio (version 3.0; Simulations Plus, Lancaster, CA, 2012) according to the structural and functional similarities of drugs, were also collected from ETCM database [33]. The targets with high similarity (structural similarity score > 0.80) were obtained for further analysis.

2.2.2. Prediction of the Targets of COPD. Firstly, the targets of COPD were searched from the ETCM database, including the results of Therapeutic Target Database (TTD, <http://db.idrblab.net/ttd/>) and Human Phenotype Ontology (HPO, <https://hpo.jax.org/app/>) databases. Then, the targets were searched from GeneCards (<https://www.genecards.org/>) database. The targets in GeneCards that exceed the median of all target relevance score were retained as potential targets for COPD. Finally, the targets from ETCM and GeneCards were combined to further investigation.

2.2.3. Construction of the Protein-Protein Interaction (PPI) Network and Target Analysis. The overlapping targets of XFZYD and COPD were input to STRING (<https://cn.string-db.org/>) database to perform the PPI network with organism limited to *Homo sapiens*. The minimum required interaction score was set at the highest confidence (> 0.900). Moreover, the independent unconnected nodes are hidden. Then, the network from STRING was exported to Cytoscape (version 3.9.1) for visualization and analysis [34]. The node targets whose three topological parameters (betweenness centrality, closeness centrality, and degree) exceeding median were considered as the major targets of XFZYD on COPD. Finally, major node targets with all three topology parameters exceeding median were identified as core node targets.

2.2.4. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Enrichment Analysis. The GO and KEGG enrichment analysis were carried out to explore the molecular mechanisms of XFZYD on COPD by the Database for Annotation, Visualization, and Integrated Discovery (DAVID, <https://david.ncifcrf.gov/>) [35, 36]. The top 10 GO and KEGG enrichment results with p values less than 0.05 were retained and input to bioinformatics (<https://www.bioinformatics.com.cn/>), a free online platform for data analysis and visualization, for visualization.

2.2.5. Construction of Related Networks. Two networks were constructed and visualized by Cytoscape: (1) the PPI network and (2) herb-bioactive ingredient-major target-pathway network.

2.2.6. Molecular Docking. According to the analysis results of the herb-bioactive ingredient-major target-pathway network, the top 3 ingredients of XFZYD were selected for molecular docking. The structures of ingredients were downloaded from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). Then, the tar-

get genes for molecular docking need to meet two requirements: (1) targets must be core node targets in PPI network and (2) the top 3 targets in herb-bioactive ingredient-major target-pathway network sorted by degree value. The structures of target proteins were obtained from Protein Data Bank (PDB, <https://www.rcsb.org/>) database. The molecular docking was conducted by AutoDock Vina (version 1.2.0) and visualized by PyMOL (version 2.5) [37].

3. Results

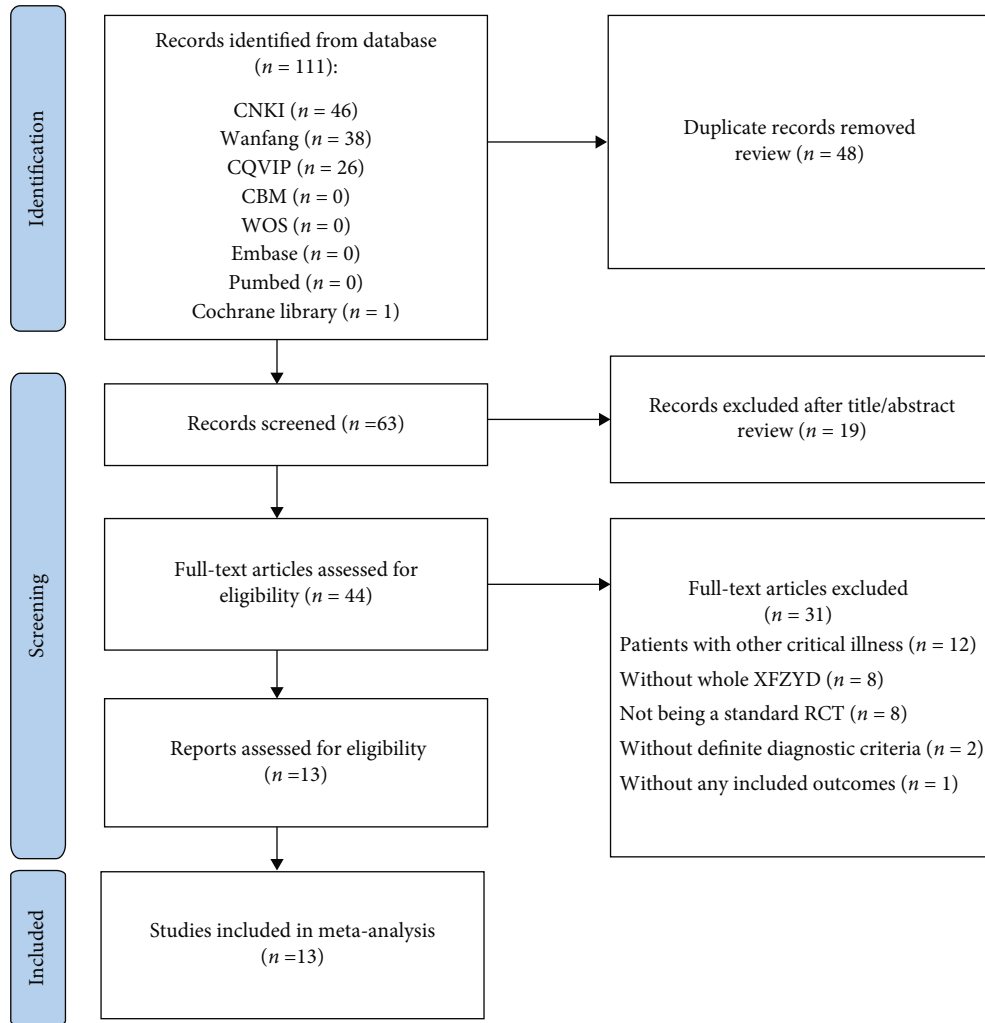
3.1. The Efficacy and Safety of XFZYD on COPD

3.1.1. Study Research and Study Characteristics. A total of 111 relevant studies were searched from 8 databases. After removal of duplicate articles, 63 studies were remained. Then, 44 articles were retained after the title/abstract reviewing. Finally, 31 articles were excluded after the full-text reviewing, and the 13 remaining articles were included for this meta-analysis (Figure 2(a)) [27, 38–49].

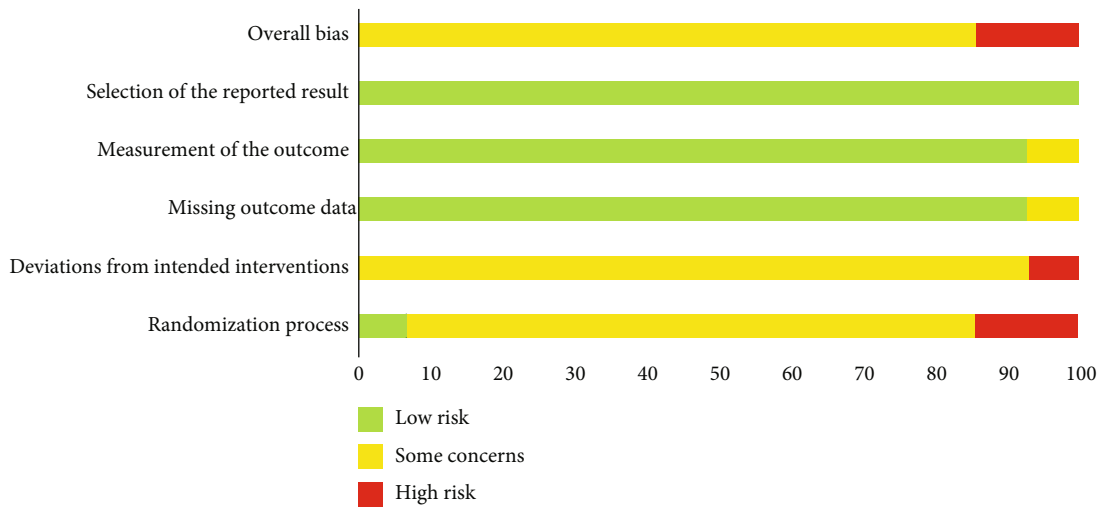
The general characteristics of the included studies are shown in Table 1. The specific dosages of XFZYD and the treatment strategies of the control group in each study are shown in Supplementary Table 2. The published articles ranged from 2011 to 2021. 893 patients were included, with 448 in the treatment group and 445 in the control group. All of the studies were conducted in China. Notably, 1 included article was divided into two separate studies (Li et al. 2021a and Li et al. 2021b) for meta-analysis because both stable and acute exacerbation studies and different control groups were established [27].

3.1.2. Risk of Bias Evaluation. The RoB 2 was applied to assess the risk of bias of included studies (Figures 2(b) and 2(c)). For overall bias, a total of 12 included studies were assessed as some concerns, and another 2 trials were judged as high risk. The poor results of the overall risk of bias assessment were mainly due to the absence of random assignment and allocation concealment details in the original studies, especially in the clinical trials conducted by Li et al. [42] and Liu et al. [45], which suggested the low quality of original studies.

3.1.3. Meta-Analysis of Pulmonary Function Indexes. Ten studies reported on the FEV1/FVC (%) of XFZYD treatment. The results showed that the FEV1/FVC (%) in the experimental group was significantly higher than that in the control group (WMD = 4.12; 95% CI: 2.09, 6.15; $p < 0.0001$; $I^2 = 71\%$) (Figure 3(a)). Due to the appreciable heterogeneity, the sensitivity analysis was conducted on ten studies by excluding them one by one. The result showed that the trials conducted by Li et al. [47], Li et al. [27], and Mao [38] are the source of heterogeneity; I^2 decreased from 71% to 29% after their exclusion, with $p < 0.00001$ (Supplementary Figure 1). Then, the subgroup analysis was performed based on the different disease stages of COPD. The results showed that the XFZYD significantly improved FEV1/FVC (%) in stable COPD (WMD = 4.37; 95% CI: 0.16, 8.57; $p = 0.04$; $I^2 = 83\%$) (Figure 3(a)) and AECOPD (WMD = 4.19; 95% CI: 1.16, 7.23; $p = 0.007$; $I^2 = 62\%$) (Figure 3(a)).



(a)



(b)

FIGURE 2: Continued.

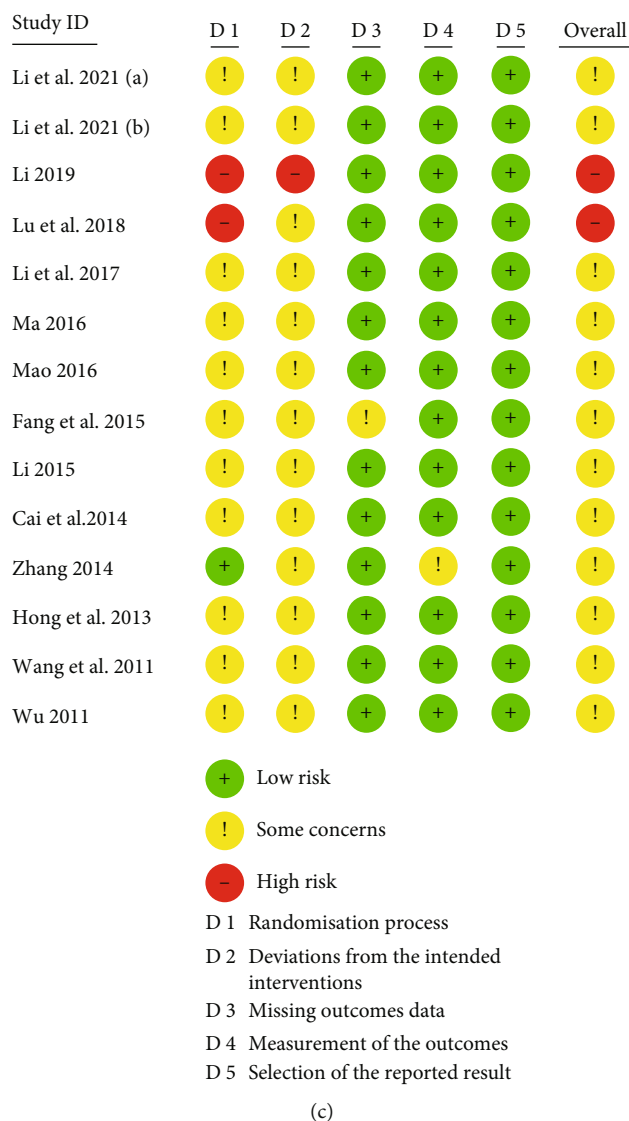


FIGURE 2: (a) The study selection process, (b) results of the summary of risk of bias, and (c) the risk of bias of each original study.

Six studies evaluated the FEV1% after the XFZYD treatment. The meta-analysis results suggested that compared with that in the control group, the FEV1% was significantly increased in the experimental group (WMD = 7.33, 95% CI: 2.38, 12.27; $p = 0.004$; $I^2 = 91\%$) (Figure 3(b)). The heterogeneity remained high under the investigation of sensitivity analysis by excluding the trials one by one, resulting in the inability to identify the source of heterogeneity. Regarding the subgroup analysis, the FEV1% was significantly enhanced in the experimental group in the stable COPD (WMD = 9.70, 95% CI: 1.35, 18.04; $p = 0.02$; $I^2 = 92\%$) (Figure 3(b)). However, there was no significant difference in the AECOPD after the XFZYD treatment (WMD = 5.29, 95% CI: -0.97, 11.54; $p = 0.10$; $I^2 = 88\%$) (Figure 3(b)).

FEV1 was reported in nine included studies. The results showed that XFZYD could significantly improve the FEV1 level (WMD = 0.16, 95% CI: 0.06, 0.26; $p = 0.002$; $I^2 = 83\%$) (Figure 3(c)). In light of the obvious heterogeneity, the sensitivity analysis was conducted. However, the results implied

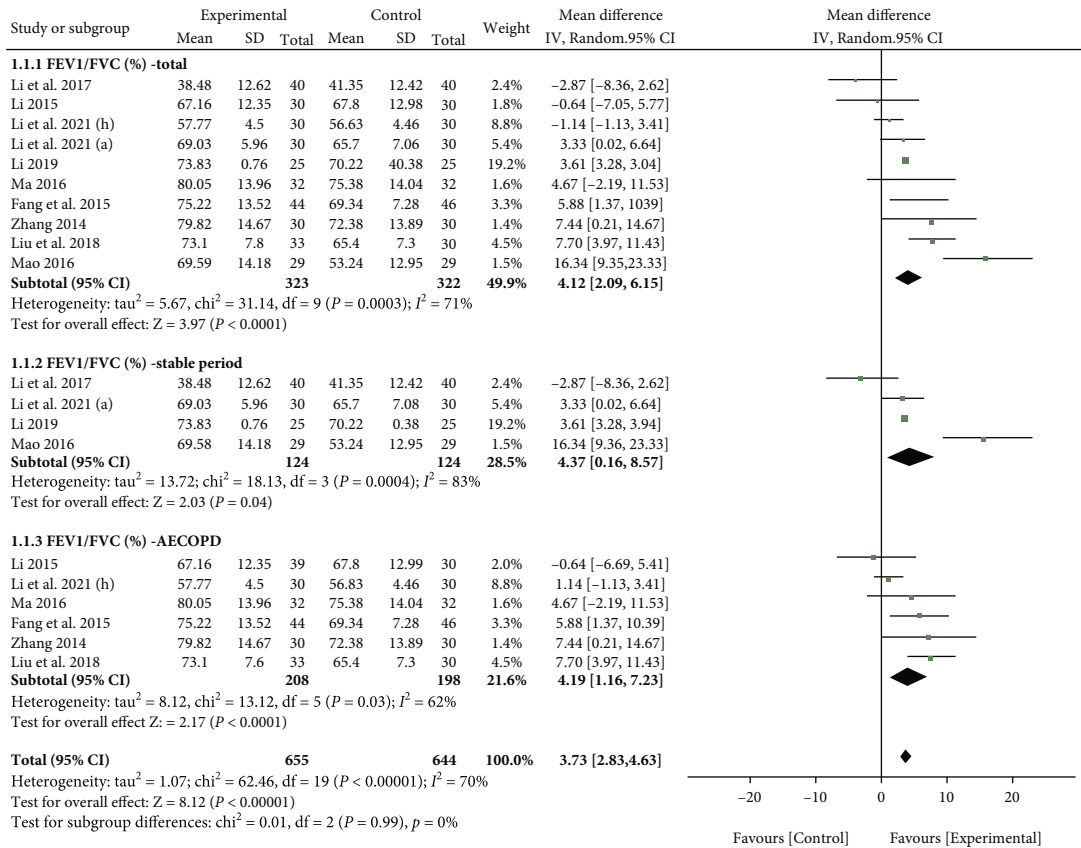
that the heterogeneity was still large. By setting the subgroup analysis, we found that the FEV1 level in the experimental group was higher than that in the control group at stable COPD (WMD = 0.17, 95% CI: 0.03, 0.31; $p = 0.02$; $I^2 = 88\%$) (Figure 3(c)). Nevertheless, there was no difference in FEV1 level between the experimental group and the control group on AECOPD (WMD = 0.14, 95% CI: -0.03, 0.31; $p = 0.10$; $I^2 = 79\%$) (Figure 3(c)).

3.1.4. Meta-Analysis of Clinical Effective Rate. As another primary outcome, the clinical effective rate was reported in eight included studies. Assessed by a fixed-effects model, the meta-analysis results showed that the clinical effective rate of the experimental group was significantly higher than that of the control group (RR = 1.26, 95% CI: 1.17, 1.37; $p < 0.00001$; $I^2 = 0\%$) (Figure 3(d)). Moreover, the subgroup analysis implied that XFZYD treatment could significantly enhance the clinical effective rate in both stable COPD (RR = 1.17, 95% CI: 1.02, 1.35; $p = 0.02$; $I^2 = 45\%$)

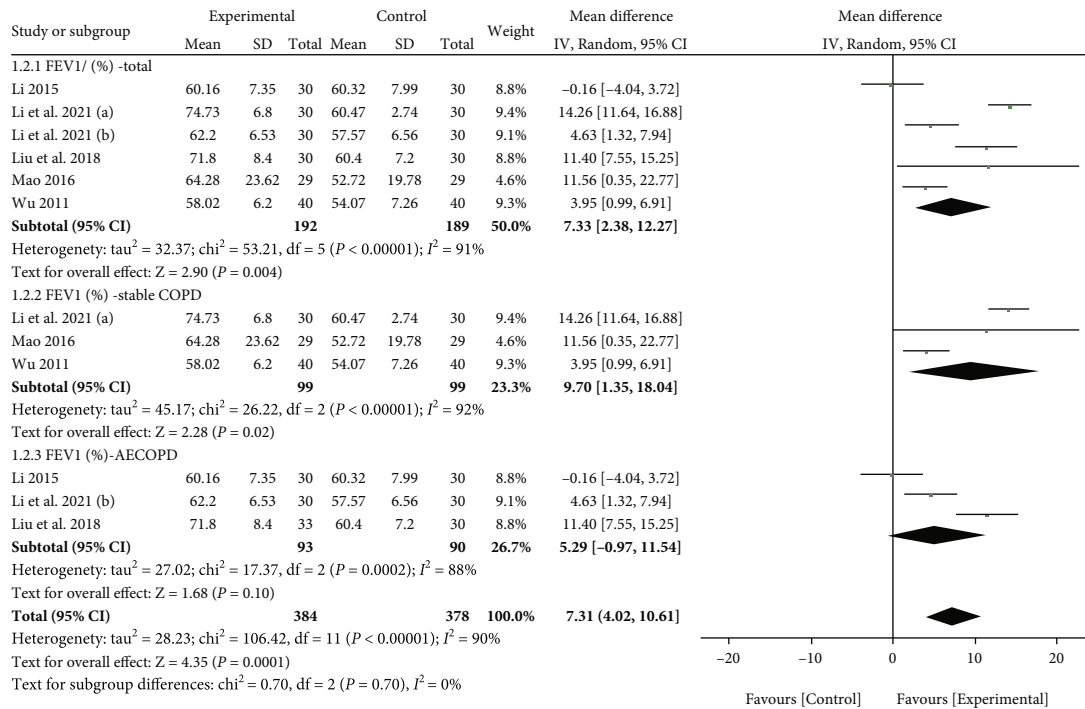
TABLE 1: The characteristics of the original studies included in the meta-analysis.

Authors	No. of patients		Age of patients		Disease course		Stage	Outcomes		Adverse events		References
	Treatment	Control	Treatment	Control	Treatment	Control		Treatment	Control	Treatment	Control	
Li et al. 2021 (a)	30	30	74.80 ± 9.49	76.50 ± 7.10	15.63 ± 7.51	17.30 ± 4.52	Stable COPD	(1)(2)(3)(5)(6)(7)(8)(9)(10)	0	0	[27]	
Li et al. 2021 (b)	30	30	75.77 ± 9.27	76.37 ± 5.81	14.90 ± 6.64	17.47 ± 3.28	AECOPD	(1)(2)(3)(5)(6)(7)(8)(9)(10)	0	2	[27]	
Li 2019	25	25	63.7 ± 4.5	63.7 ± 4.3	9.0 ± 2.3	8.5 ± 2.0	Stable COPD	(1)(3)(4)	NR	NR	[42]	
Liu et al. 2018	33	30	63 ± 6.2	64 ± 5.5	16 ± 2.4	15 ± 2.7	AECOPD	(2)(3)(4)(5)(6)(7)(9)(10)	NR	NR	[45]	
Li et al. 2017	40	40	61.63 ± 5.14	62.39 ± 4.32	13.73 ± 6.83	13.23 ± 7.54	Stable COPD	(1)(3)(4)	0	0	[47]	
Ma 2016	32	32	63.18 ± 5.02	62.5 ± 4.49	10.3 ± 2.45	9.42 ± 3.72	AECOPD	(1)(3)(4)	NR	NR	[41]	
Mao 2016	29	29	67.17	17.47	NR	NR	Stable COPD	(2)(3)	NR	NR	[38]	
Fang et al. 2015	44	46	65.0 ± 4.8	66.0 ± 5.9	10.2 ± 3.3	12.5 ± 4.0	AECOPD	(1)(3)(5)(6)(7)(8)(10)	0	0	[43]	
Li 2015	30	30	59.5 ± 4.42	58.6 ± 4.63	9.55 ± 3.53	9.17 ± 3.48	AECOPD	(1)(2)(3)(4)(5)(6)	NR	NR	[49]	
Cai et al. 2014	24	22	50~76	50~75	7~26	7~26	NR	(4)	NR	NR	[44]	
Zhang 2014	30	30	61.9 ± 5.7	62.1 ± 5.3	14.2 ± 3.9	13.4 ± 4.6	AECOPD	(1)(3)(8)	NR	NR	[40]	
Hong et al. 2013	30	30	72.1 ± 9.3	74.5 ± 8.8	NR	NR	AECOPD	(4)(7)	NR	NR	[48]	
Wang et al. 2011	31	31	NR	NR	NR	NR	AECOPD	(4)	NR	NR	[46]	
Wu 2011	40	40	67.2	66.8	NR	NR	Stable COPD	(1)(2)(5)(6)	NR	NR	[39]	

(1): FEV1; (2): FEV1%; (3): FEV1/FVC (%); (4): clinical effective rate; (5): p(O₂); (6): p(CO₂); (7): D-dimer; (8): FIB; (9): PT; (10): APTT; AECOPD: acute exacerbation of chronic obstructive pulmonary disease; NR: not reported; COPD: chronic obstructive pulmonary disease.

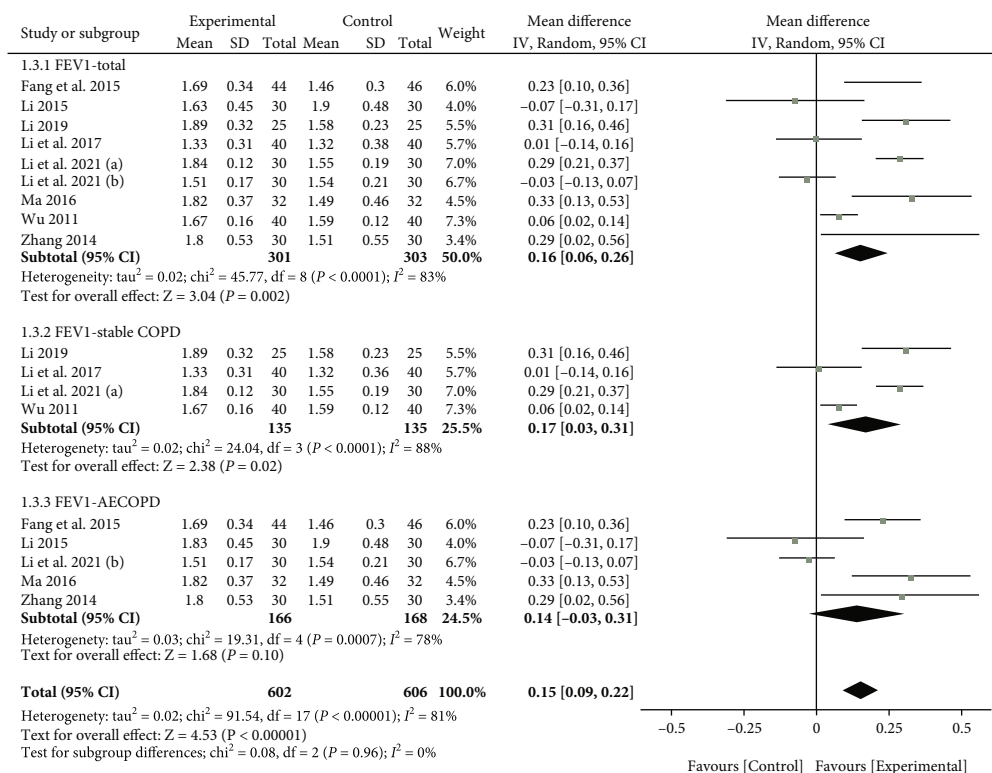


(a)

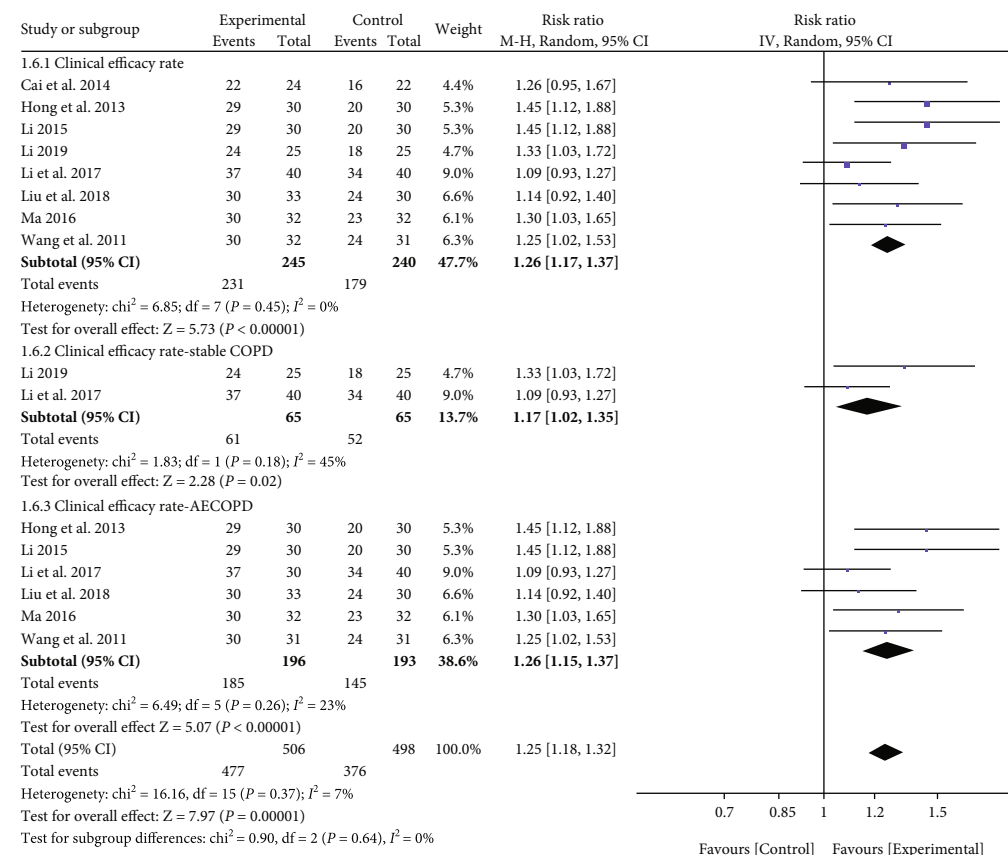


(b)

FIGURE 3: Continued.



(c)



(d)

FIGURE 3: The meta-analysis results of primary outcomes. (a) The forest plot of FEV1/FVC (%). (b) The forest plot of FEV1%. (c) The forest plot of FEV1. (d) The forest plot of clinical effective rate.

(Figure 3(d)) and AECOPD (RR = 1.26, 95% CI: 1.15, 1.37; $p = 0.00001$; $I^2 = 23\%$) (Figure 3(d)).

3.1.5. Meta-Analysis of Blood Gas Indexes. A total of six studies reported on the level of $p(O_2)$. Meta-analysis suggested that the level of $p(O_2)$ was higher in the experimental group compared to the control group (WMD = 8.38, 95% CI: 4.88, 11.88; $p < 0.00001$; $I^2 = 81\%$) (Figure 4(a)). Then, the subgroup analysis also showed that the $p(O_2)$ level could be restored after XFZYD treatment both in stable COPD (WMD = 6.37, 95% CI: 0.21, 12.52; $p = 0.04$; $I^2 = 88\%$) (Figure 4(a)) and AECOPD (WMD = 9.57, 95% CI: 5.30, 13.84; $p < 0.001$; $I^2 = 73\%$) (Figure 4(a)).

Six studies reviewed $p(CO_2)$ level. Our meta-analysis results suggested that XFZYD could exhibit a decreasing effect on $p(CO_2)$ level (WMD = -3.43, 95% CI: -5.57, -1.28; $p = 0.002$; $I^2 = 74\%$) (Figure 4(b)). Similarly, XFZYD was effective in reducing the level of $p(CO_2)$ in both stable COPD (WMD = -2.63, 95% CI: -4.44, -0.82; $p = 0.004$; $I^2 = 0\%$) (Figure 4(b)) and AECOPD (WMD = -3.98, 95% CI: -7.60, -0.35; $p = 0.03$; $I^2 = 84\%$) (Figure 4(b)).

3.1.6. Meta-Analysis of Coagulation Indexes. Four coagulation indexes including FIB, D-dimer, PT, and APTT were evaluated in this meta-analysis. The results showed that after XFZYD treatment, the D-dimer (SMD = -2.41, 95% CI: -3.98, -0.84; $p = 0.003$; $I^2 = 97\%$) (Figure 4(c)), FIB (WMD = -1.45, 95% CI: -2.16, -0.74; $p < 0.0001$; $I^2 = 98\%$) (Figure 4(d)), and PT (WMD = 3.01, 95% CI: 2.51, 3.52; $p < 0.00001$; $I^2 = 33\%$) (Figure 4(e)) levels were significantly improved. However, there was no difference on APTT level after the XFZYD treatment (WMD = 1.46, 95% CI: -1.15, 4.07; $p = 0.27$; $I^2 = 89\%$) (Figure 4(f)).

3.1.7. Meta-Analysis of Adverse Events. Among the included 14 studies, only 4 studies reported the adverse events. Notably, three of these studies reported no adverse events in either the experimental group or the control group, and the results showed a nonsignificant difference between the experimental group and the control group (RD = -0.01, 95% CI: -0.05, 0.02; $p = 0.40$; $I^2 = 0\%$) (Figure 5).

3.1.8. Publication Bias. The publication bias of primary outcomes was assessed by funnel plots. The funnel plots were asymmetrical, which showed a possible publication bias (Figure 6).

3.1.9. Quality of Evidence. The quality of the evidence was generally low, with only two indicators of clinical effective rate and PT judged to be of moderate quality (Supplementary Figure 2). The reason for the downgrading of the evidence quality is primarily from the worrisome methodology of included studies, which is consistent with the factors influencing the risk of bias assessment.

3.2. The Molecular Mechanisms of XFZYD on COPD Based on Network Pharmacology

3.2.1. The Bioactive Ingredients and Potential Targets of XFZYD and Treating Targets of COPD. A total of 499 ingre-

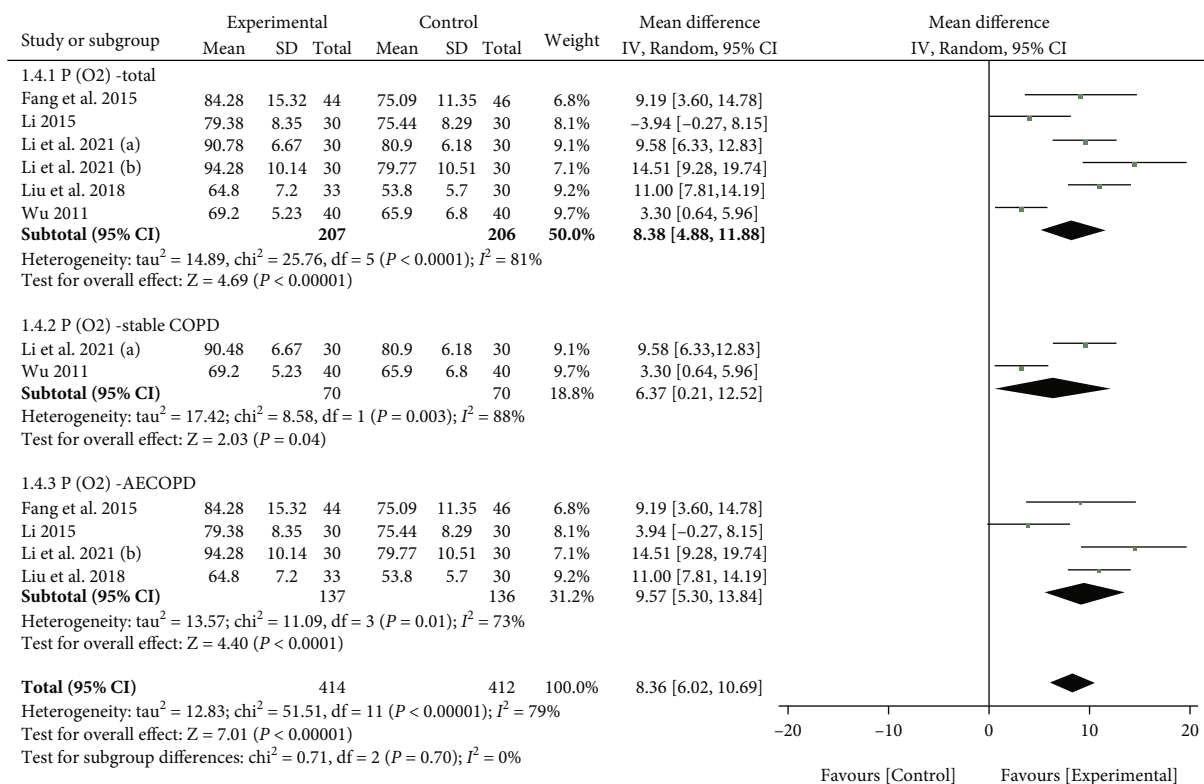
dients were obtained from the ETCM database. After the assessment of QED score, 248 components were retained and identified as bioactive ingredients of XFZYD. Then, 115 ingredients which had high similarity targets were chosen. Finally, 115 bioactive ingredients with 578 potential targets were left.

3247 treating targets were searched and screened from ETCM and GeneCards databases after removing the duplicates. After matching the potential targets of XFZYD with the COPD-related targets, 239 overlapping targets were acquired and considered as the molecular basis of XFZYD on COPD (Figure 7).

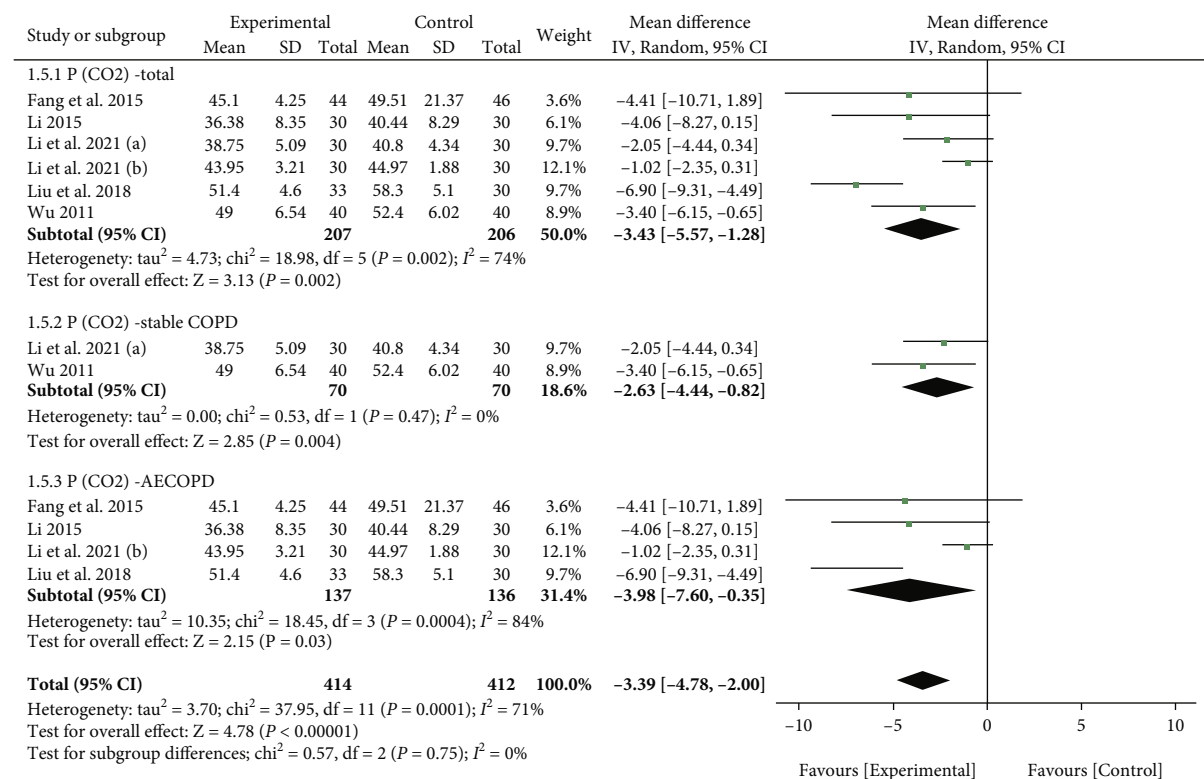
3.2.2. The Construction of PPI Network by STRING. 239 targets were uploaded to STRING database and analyzed by setting the highest confidence (Figure 8(a)). The PPI network, containing 183 nodes and 495 edges, was further downloaded and imported to Cytoscape (Figure 8(b)). By calculating three topological parameters, 49 node targets were identified as major targets of XFZYD on COPD (Figure 8(c)). Based on the results of further evaluation, 18 node targets were finally recognized as the core target genes for XFZYD in the treatment of COPD (Figure 8(d)).

3.2.3. Results of GO and KEGG Pathway Enrichment Analysis. 49 major targets were uploaded to the DAVID database for GO and KEGG enrichment analysis. The top 10 terms in biological process were positive regulation of transcription from RNA polymerase II promoter, positive regulation of transcription, DNA-templated, response to drug, protein phosphorylation, signal transduction, positive regulation of nitric oxide biosynthetic process, transcription initiation from RNA polymerase II promoter, innate immune response, negative regulation of apoptotic process, and cellular response to lipopolysaccharide (Figure 9(a)). In terms of cell components, the targets of XFZYD on COPD were mainly concentrated in the nucleus, cytosol, nucleoplasm, cytoplasm, plasma membrane, membrane, perinuclear region of cytoplasm, protein complex, cell surface, and nuclear chromatin (Figure 9(a)). In terms of molecular functions, the major target genes were focused on protein binding, DNA binding, ATP binding, enzyme binding, identical protein binding, transcription factor activity, sequence-specific DNA binding, transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding, RNA polymerase II core promoter proximal region sequence-specific DNA binding, and zinc ion binding (Figure 9(a)).

The KEGG pathway enrichment results indicated that the major target genes of XFZYD for COPD are mainly enriched in pathways in cancer, PI3K-Akt signaling pathway, osteoclast differentiation, hepatitis B, HTLV-I infection, nonalcoholic fatty liver disease (NAFLD), influenza A, Chagas disease (American trypanosomiasis), T cell receptor signaling pathway, and TNF signaling pathway (Figure 9(b)). Based on the gene counts enriched in these pathways and the pathogenesis of COPD, the PI3K-Akt signaling pathway was regarded as the critical pathway of XFZYD in the treatment of COPD.



(a)



(b)

FIGURE 4: Continued.

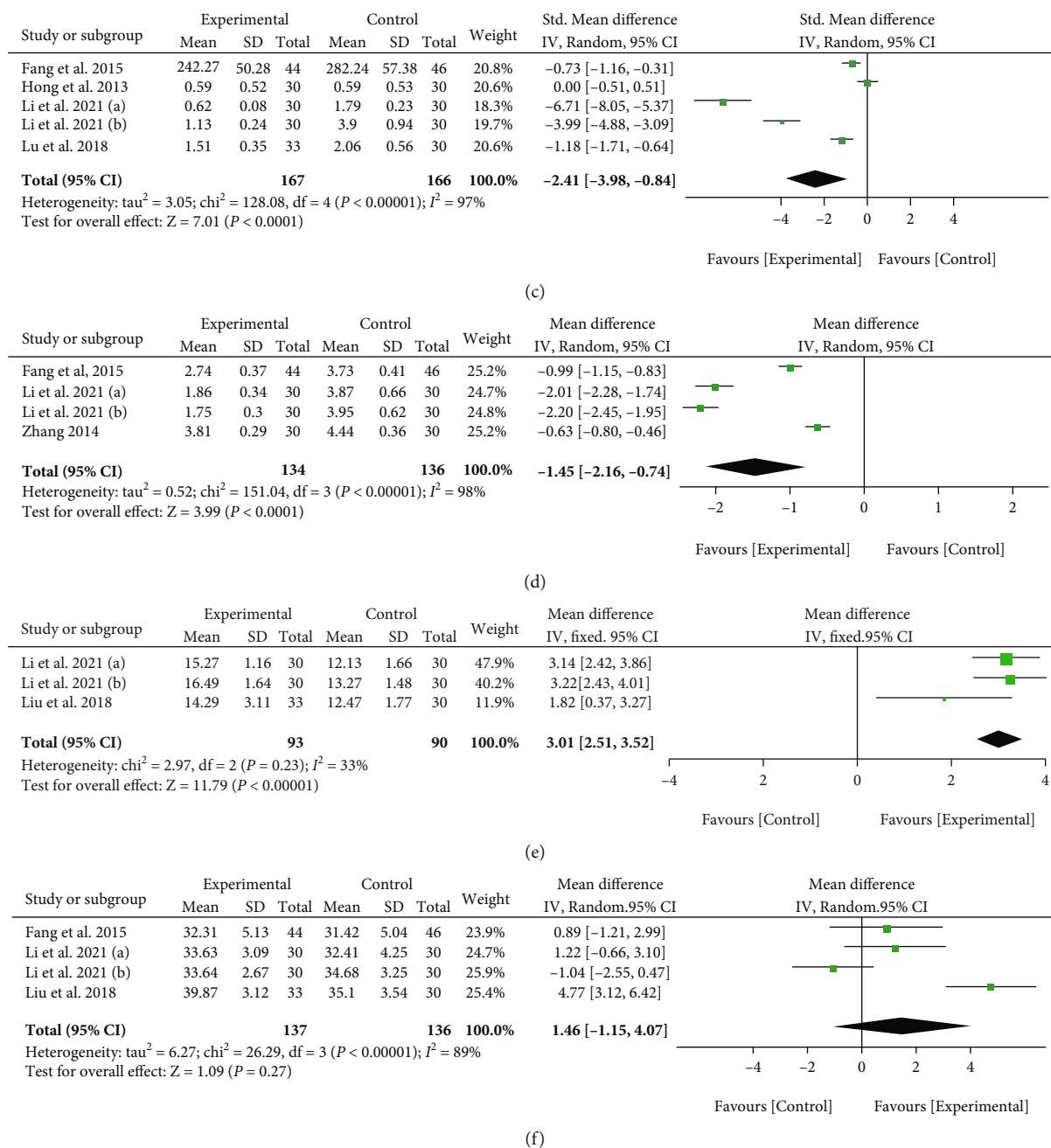


FIGURE 4: The meta-analysis results of secondary outcomes. (a) The forest plot of $p(O_2)$. (b) The forest plot of $p(CO_2)$. (c) The forest plot of D-dimer. (d) The forest plot of FIB. (e) The forest plot of PT. (f) The forest plot of APTT.

3.2.4. Molecular Docking Results. The herb-bioactive ingredient-major target-pathway network was constructed by Cytoscape to directly represent the main molecular mechanisms of XFZYD in the treatment of COPD (Figure 10). Based on the results of herb-bioactive ingredient-major target-pathway network and PPI network, the top 3 core genes (ESR1, AKT1, and NFKB1) and top 3 ingredients, containing adenosine, quercetol, and caprylic acid, were chosen to perform the molecular docking experiments. Generally, the binding energy which is less than -4.25 kcal/mol, -5.0 kcal/mol, or -7.0 kcal/mol indicates a certain, good, or

strong binding activity between ligands and receptors [50]. The docking results showed that 3 ligands and 3 receptors had a certain binding activity (Figure 11), suggesting the treating potential of XFZYD on COPD. The docking results were visualized by PyMOL (Figure 12).

4. Discussion

4.1. Clinical Evidence Summary. The 5-year survival rate for COPD patients ranges from 56% to 92% depending on the severity of the disease, which suggests that patients need

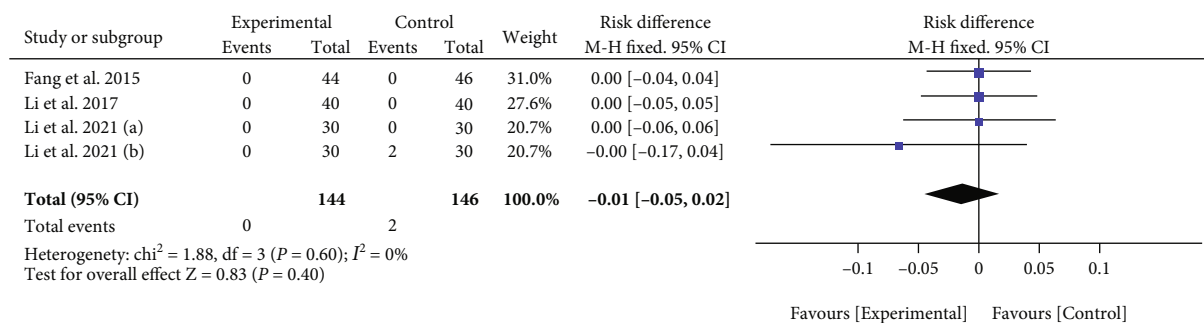


FIGURE 5: The meta-analysis results of adverse events.

early clinical interventions to control disease progression [51]. However, there is a considerable gap in the armamentarium available for treating COPD. For example, drugs that effectively control the emphysema is still scarce, but COPD patients are usually accompanied by variable degrees of emphysema [17]. Similarly, the overreliance on bronchodilators and corticosteroids leads to inevitable adverse effects, which will be detrimental to the long-term treatment of patients [52]. Notably, numerous clinical trials have provided evidence for the efficacy of XFZYD in the treatment of COPD. Therefore, the aim of the present study is to conduct a comprehensive assessment of the efficacy and safety of XFZYD on COPD.

13 original studies comprising 893 patients were included in this meta-analysis. FEV1/FVC (%), a direct indicator of pulmonary function, has been the primary measure used to diagnose COPD and determine disease severity [53]. Likewise, both FEV1 and FEV1 (%) values are important indicators for diagnosing COPD [54]. Therefore, the levels of FEV1/FVC (%), FEV1, and FEV1 (%) were used as the primary outcomes to evaluate the therapeutic effect of XFZYD on COPD. The results of our meta-analysis showed that the treatment with XFZYD significantly restored the FEV1/FVC (%), FEV1, and FEV1 (%) levels in COPD patients compared with the symptomatic supportive treatments alone, suggesting that XFZYD contributed to the recovery of pulmonary function in COPD patients. Another primary outcome, the clinical effective rate, mainly reflects whether COPD patients respond to the treatment with XFZYD. All of the included original literature which evaluated the clinical effective rate are based on the guidelines for clinical studies of new Chinese medicines. Similarly, the results demonstrated the therapeutic effect of XFZYD on COPD. Overall, the use of XFZYD on the basis of symptomatic supportive treatments significantly improved the response of COPD patients and their pulmonary function, which suggested better clinical outcomes.

Decreased $p(O_2)$ and increased $p(CO_2)$ are frequent clinical events in patients with COPD, which is strongly linked to hypoventilation, with imbalance of ventilation and perfusion being secondary factors [55]. Moreover, the presence of such hypoxemia and hypercapnia may lead to arrhythmias and pulmonary hypertension and even increase the mortality of COPD patients. Remarkably, XFZYD contributed to

the recovery of $p(O_2)$ and the reduction of $p(CO_2)$ in COPD patients. As a result, the early intervention with XFZYD may help to reduce numerous adverse outcomes associated with hypoxemia and hypercapnia in COPD.

In addition to the known serious respiratory consequences, COPD is thought to significantly increase the risk of venous embolism and cardiovascular diseases, which is associated with the development of a hypercoagulable state in COPD patients and may further lead to atherosclerotic thrombosis and venous thrombosis [56, 57]. As a dissolution product of cross-linked fibrin, elevated D-dimer levels predict an increased propensity for intravascular coagulation [58]. FIB is a recognized important risk factor for cardiovascular diseases such as thrombosis and atherosclerosis. Therefore, D-dimer and FIB are considered as critical biomarkers for the assessment of coagulation and fibrinolytic activity [59]. The results of the present meta-analysis suggested that XFZYD was capable of significantly reducing D-dimer and FIB levels in COPD patients.

Generally, the cascade of coagulation consists of two pathways. The intrinsic pathway includes coagulation factors I (fibrinogen), II (prothrombin), IX, X, XI, and XII. Another pathway, the exogenous pathway, mainly involves coagulation factors I, II, VII, and X. PT responds to the activation time of the extrinsic pathway, and APTT responds to the intrinsic-pathway activation time [57]. The results of meta-analysis showed that XFZYD prolonged PT levels in COPD patients, but the effect on APTT levels was not significant, implying that XFZYD may further regulate the hypercoagulable state in COPD patients by alleviating the intrinsic pathway of coagulation to reduce the occurrence of cardiovascular diseases.

AECOPD are usually triggered by bacterial or viral infections of the respiratory tract. Significant airway inflammation, mucus secretion and gas trapping, and even systemic inflammation are present during AECOPD compared to the stable phase [60, 61]. Frequent AECOPD events can significantly reduce pulmonary function, affect quality of life, and increase the risk of death of patients [62]. With AECOPD, the medications used in the stable phase are no longer applicable for the clinical management of patients, and therefore, the 2017 Global Initiative for Chronic Obstructive Lung Disease (GOLD) document defines AECOPD as “an acute worsening of respiratory symptoms

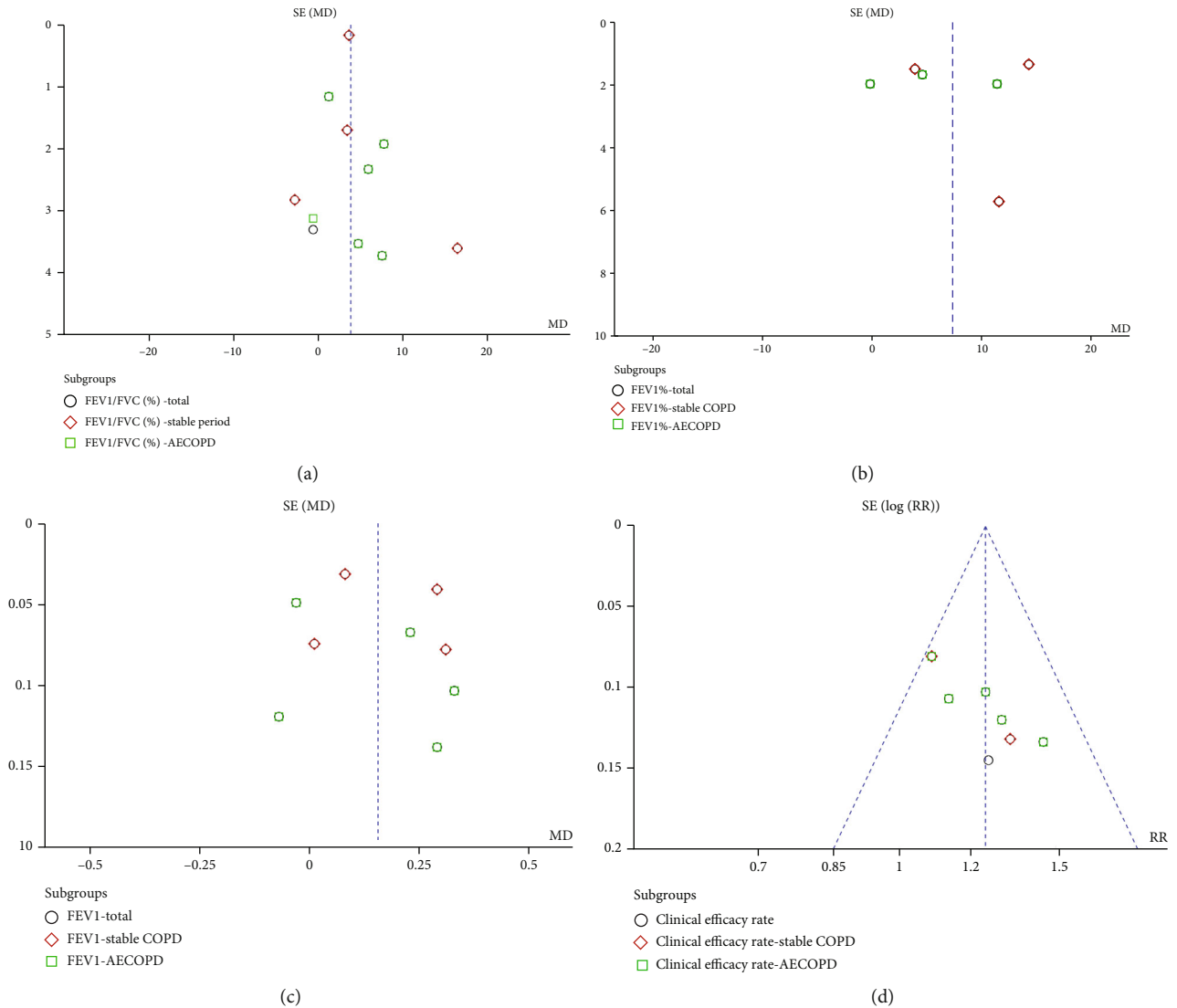


FIGURE 6: The publication bias assessment of primary outcomes. (a) The publication bias of FEV1/FVC (%). (b) The publication bias of FEV1%. (c) The publication bias of FEV1. (d) The publication bias of clinical effective rate.

that leads to additional treatment” [63]. As a consequence, research on AECOPD treatment strategies is also exigent. The results of our subgroup analysis showed that XFZYD helped improve pulmonary function and gas retention (hypoxemia and hypercapnia) in patients, suggesting the potential of XFZYD as an adjunctive treatment for AECOPD.

The incidence of adverse events is an important indicator to evaluate the safety of treatment approaches. In included original studies, only four trials reported adverse events. For specific adverse events, only one study mentioned transient gastrointestinal symptoms in the control group, and no original study mentioned adverse events after the use of XFZYD.

Although our meta-analysis provided some clinical evidence for the treatment of COPD with XFZYD, the significant heterogeneity among studies suggested that the strength of the evidence should be viewed with caution. Basically, the clinical and methodological aspects are important factors that contribute to the heterogeneity between studies

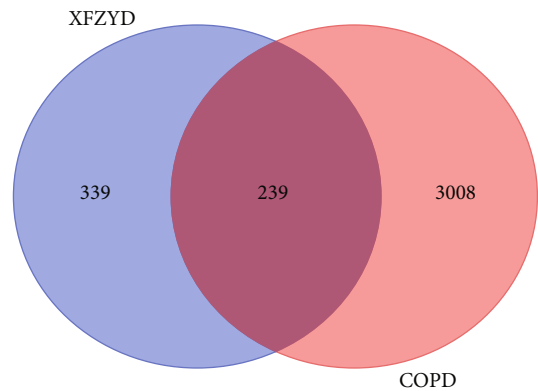
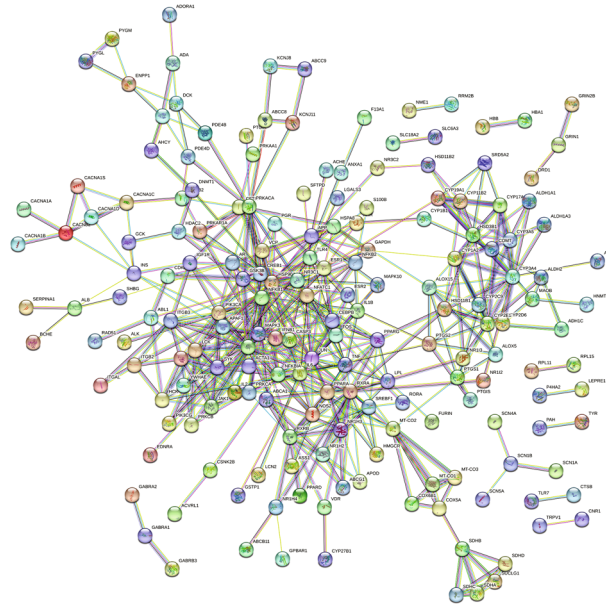
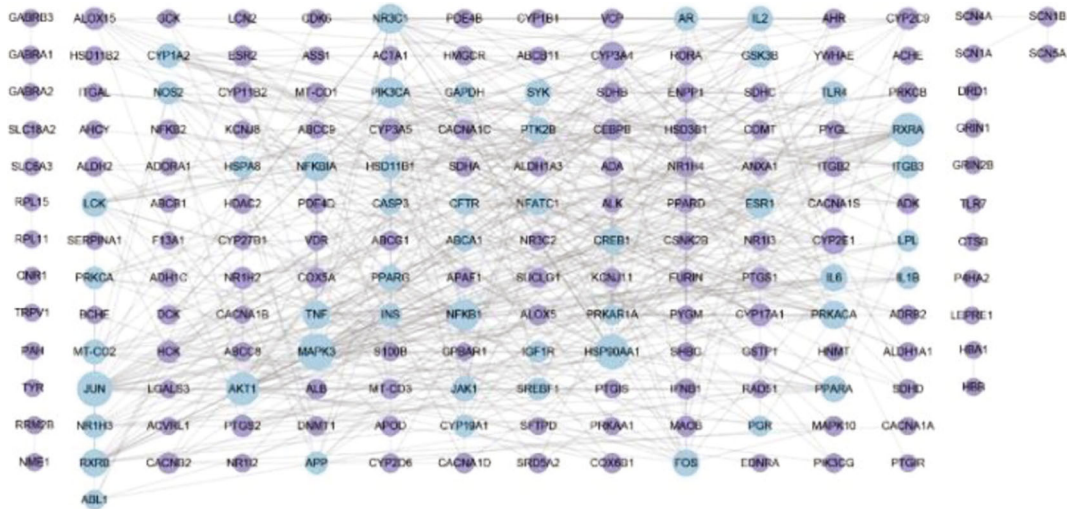


FIGURE 7: The Venn diagram of XFZYD and COPD targets.

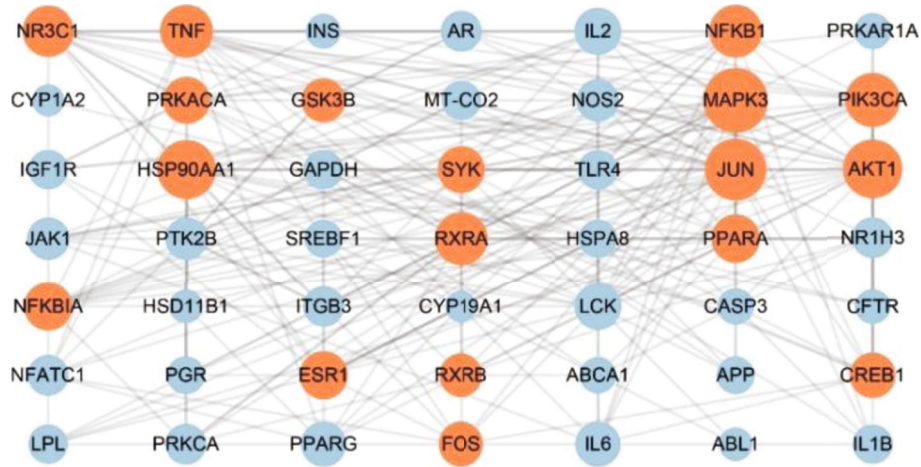
[64]. The strict inclusion and exclusion criteria have been adopted to avoid the clinical heterogeneity among the included literature in our study; however, there are still some



(a)



(b)



(c)

FIGURE 8: Continued.

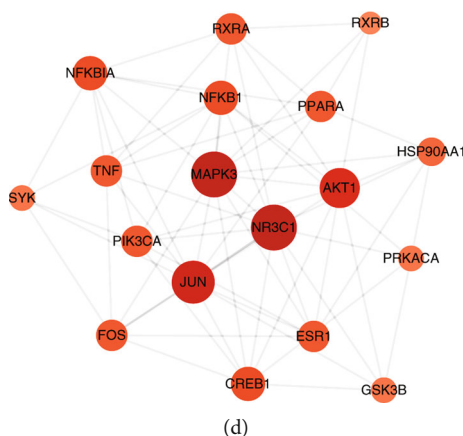


FIGURE 8: The PPI network of XFZYD on COPD. (a) The PPI network obtained from STRING database. (b) The PPT network analyzed by Cytoscape. The node genes with blue color were major genes. (c) The PPI network including the 49 major node genes. The node genes with blue color were core genes. (d) The PPI network including the 19 core node genes. The deeper the color of the node, the larger its degree value.

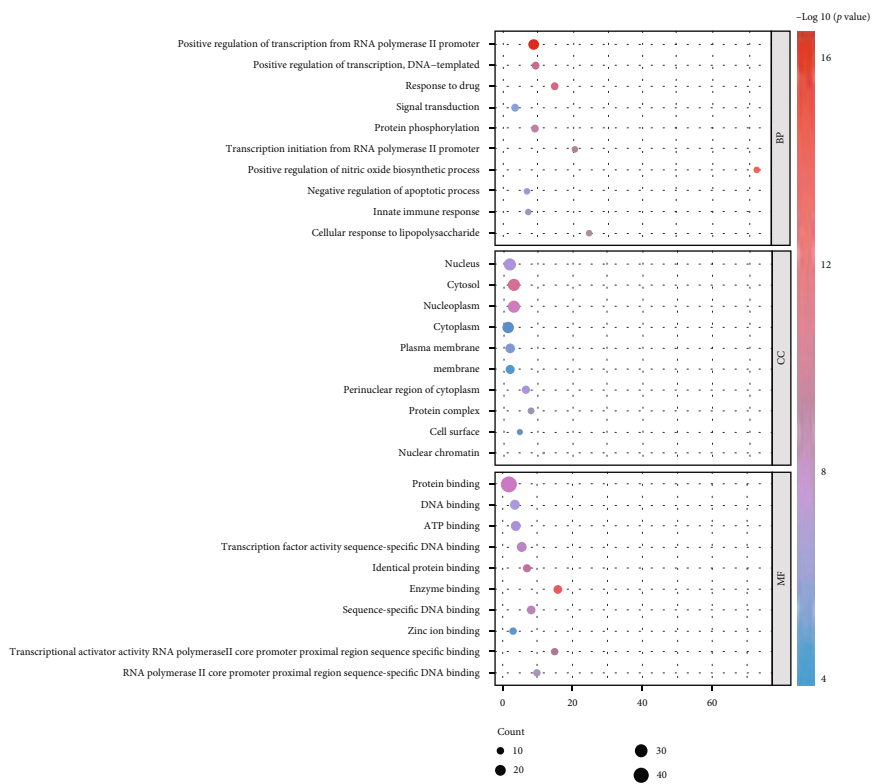
factors which may have an influence on clinical heterogeneity that cannot be prevented in TCM-based meta-analysis. For example, the syndrome differentiation and treatment are the basic principle of TCM application in clinical practice, which means that the physicians will still add or subtract some TCMs depending on the symptoms and signs of patients possibly even if a uniform treatment protocol is prescribed in advance [65]. Moreover, the same TCMs may be used in different dosages in the treatment periods. As a consequence, these phenomena may contribute to potential differences in treatment measures between included clinical trials. In addition, the different herbal dosage forms in separate studies are also an influential factor that cannot be ignored [66]. Notably, the studies conducted by Li et al. [27] and Wu [39] applied the capsule formulation of XFZYD, whereas the remaining included studies used decoction tonics, which may also be a source of clinical heterogeneity. The methodological heterogeneity is particularly evident in the meta-analysis of TCM, which is consistent with the risk of bias assessment and GRADE evaluation in our study. A large number of TCM clinical trials are characterized by low methodological quality and low reporting quality, especially reflected in the lack of attention to randomization, allocation concealment, and blinding [67]. As a result, the methodology improvement and harmonization of clinical measures may be an important means to improve the clinical evidence in TCM and to make it more convincing.

4.2. Molecular Mechanisms Based on Network Pharmacology.

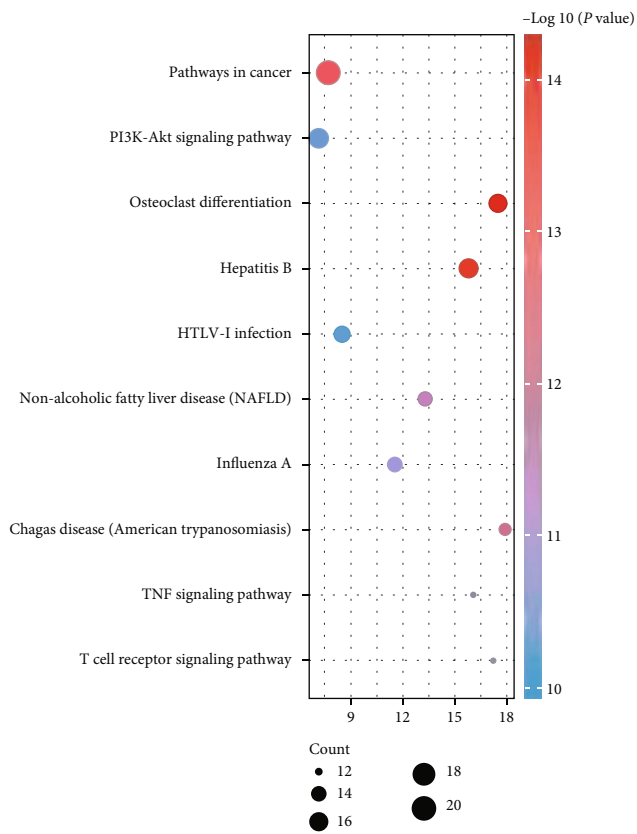
The pharmacological activities of XFZYD have been studied to some extent, including but not limited to anti-inflammatory, tissue protection, and antiapoptotic [22, 68, 69]. Notably, numerous studies have focused on exploring the treatment potential and mechanisms of XFZYD on cardiovascular diseases, such as traumatic brain injury and coronary heart disease but with little concentration on COPD. In the current study, 293 common target genes were found between XFZYD and COPD, indicating the molecular basis of treating potential of XFZYD. According to the analysis

results of the herb-bioactive ingredient-major target-pathway network, *Gan Cao*, *Dang Gui*, and *Chai Hu* showed the highest degree values, which suggested that they may play necessary roles in XFZYD on COPD. *Gan Cao* has been shown to exert expectorant and pulmonary function relief effects [70]. Several components in *Gan Cao*, including isoliquiritigenin, glycyrrhizic acid, and flavonoids, have been proven for their therapeutic potentials on COPD with significant anti-inflammatory and pulmonary tissue protective effects in pre-clinical studies [71, 72]. *Dang Gui* is receiving attention due to its anti-inflammatory, antifibrotic, and antioxidant activities [73]. *Dang Gui* and its active ingredients have been demonstrated to have therapeutic effects for some respiratory diseases, such as pulmonary fibrosis, asthma, and lung injury in preclinical researches, suggesting that it may have therapeutic potential on COPD [74–76]. Another medicinal, *Chai Hu*, has also been shown to have anti-inflammatory effects and its polysaccharides can improve the lung injury in vivo [77, 78].

Adenosine, quercetol, and caprylic acid were identified as the key ingredients of XFZYD and were used for molecular docking experiments. Adenosine is a nucleoside signaling molecule produced by cells when differing stress or damage [79]. After being produced, adenosine interacts with its receptors to activate adenosine-related signal, further regulating the cell functions [80]. The adenosine signal is believed to participate in the occurrence and development of COPD [81]. Moreover, experts suggest that the adenosine receptors may be the treating targets of COPD [82]. However, signals induced by different adenosine receptors will exert proinflammatory or anti-inflammatory properties in pulmonary inflammatory diseases, which means that whether adenosine exogenously intake via XFZYD has a therapeutic effect on COPD still needs to be further evaluated [80]. Quercetol, commonly named quercetin, its therapeutic potential on COPD has been proven by preclinical and clinical studies [83, 84]. The pharmacological mechanisms are related to its anti-inflammatory and antioxidant effects [85]. Furthermore, an in vitro study based on mononuclear cells from COPD patients showed that quercetin can



(a)



(b)

FIGURE 9: The GO and KEGG pathway enrichment analysis results of 49 major node genes. (a) The top 10 GO enrichment analysis results. (b) The top 10 KEGG pathway enrichment analysis results.

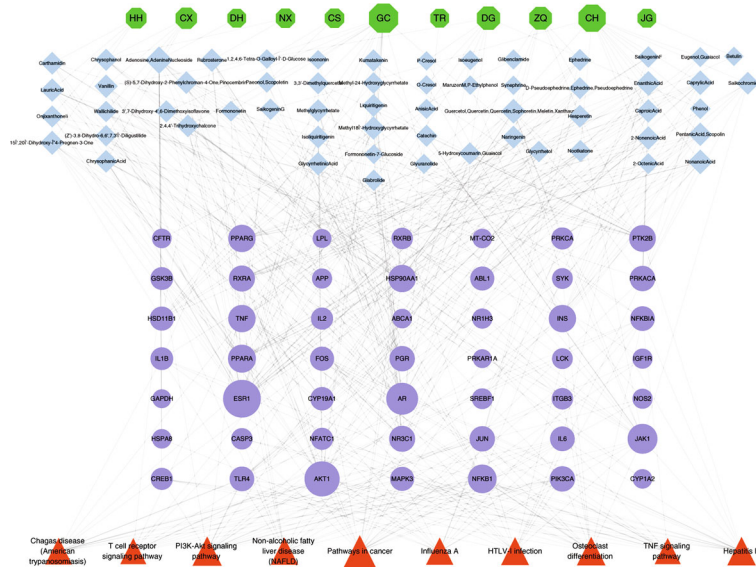


FIGURE 10: The medicine-bioactive ingredient-major target-pathway network. Green ellipse nodes represent medicines of XFZYD. Nodes with blue stand for ingredients of XFZYD. Purple ellipse nodes were target genes. Orange nodes indicate the top 10 pathways. The larger the node, the larger its degree value.

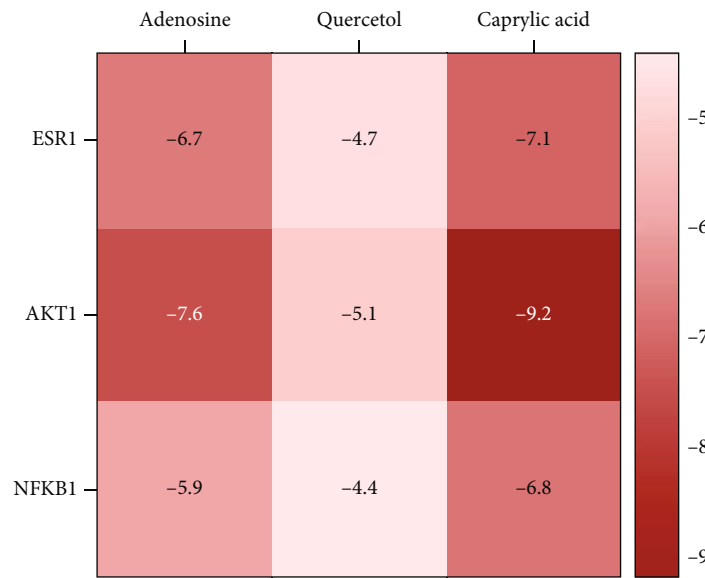


FIGURE 11: The binding energy of 3 compounds and 3 protein receptors for molecular docking.

increase the cellular sensitivity to steroids, which is clinically important for steroid-resistant COPD patients [86]. Caprylic acid is a medium chain fatty acid derived mainly from coconut oil and goat milk and is also applied as an antimicrobial agent in industry production [87]. Current researches have revealed its advantages in terms of energy homeostasis, neuroprotection, hypoglycemia, and even antitumor, suggesting its clinical application capability [88]. However, its therapeutic advantages on COPD and specific mechanisms apparently remain uncovered.

The pathogenesis of COPD is a complex progress containing various inflammatory cells, inflammatory mediators, and cell signaling pathways [89]. Excessive oxidative stress and inflammatory responses are necessary factors in COPD

pathogenesis [90, 91]. It is worth noting that the dysregulation of the PI3K/AKT pathway plays an important role in the activation of inflammatory cells and release of inflammatory mediators in COPD [92]. In addition, the activation of Nrf2, mainly induced by the PI3K/AKT pathway, also regulates the oxidative stress and chronic inflammation in COPD [93]. In preclinical investigations, the intervention of the PI3K/AKT pathway has shown the advantages in improving chronic inflammation in COPD, suggesting that the PI3K/AKT pathway may be the hopeful treating target [94, 95]. In our study, the influence on the PI3K/AKT signaling pathway was recognized as key mechanism of XFZYD based on the KEGG pathway enrichment analysis, which provides a vision to fully reveal the mechanisms of XFZYD in the

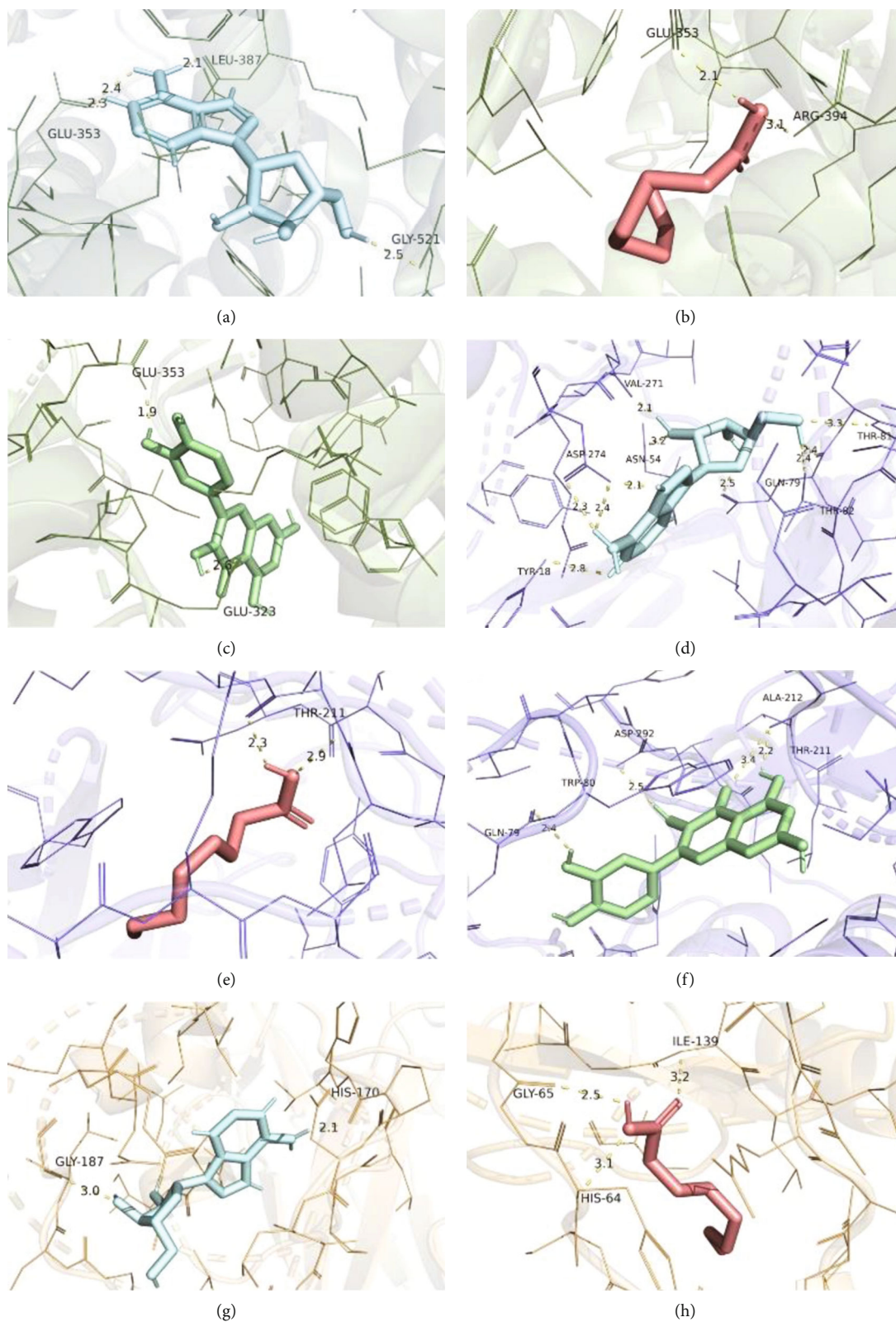


FIGURE 12: Continued.

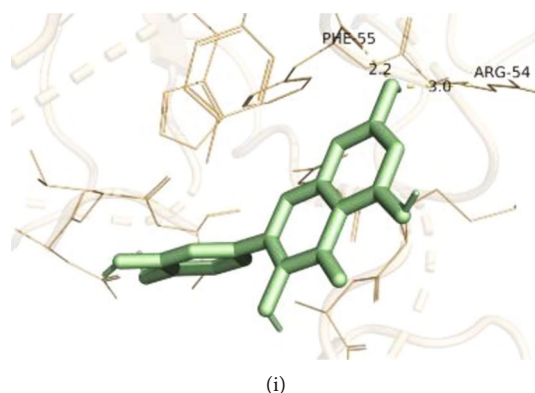


FIGURE 12: The molecular docking results visualized by PyMOL. (a) Adenosine with ESR1. (b) Quercetol with ESR1. (c) Caprylic acid with ESR1. (d) Adenosine with AKT1. (e) Quercetol with AKT1. (f) Caprylic acid with AKT1. (g) Adenosine with NFKB1. (h) Quercetol with NFKB1. (i) Caprylic acid with NFKB1.

treatment of COPD. In our future research, chemical analysis and animal studies can be combined with network pharmacology to comprehensively identify the key active ingredients and validate the complex mechanisms of XFZYD in the treatment of COPD, which will provide stronger evidence of its clinical application.

4.3. The Communication between Clinical Evidence and Network Pharmacology Results. The application of XFZYD is contributed to the improved pulmonary function, relieved ventilatory disturbances, and corrected hypercoagulability of COPD patients in an evidence-based medicine view. Moreover, the network pharmacology results suggested that XFZYD is capable of intervening with the PI3K/AKT pathway to improve COPD. Thus, in reality, the oxidative stress and inflammatory damage are critical aspects of the numerous pathogenic mechanisms of COPD. The risk factors, such as cigarette smoke, lead to the production of numerous reactive oxygen species (ROS), which further leads to lung cell damage, alveolar epithelial damage, and excessive mucus secretion [96]. At the same time, ROS can promote the expression of inflammatory mediators and aggravate lung tissue damage, leading to impaired lung function in patients [89]. Furthermore, the chronic immune cell infiltration and inflammatory mediator release promote the secretion of proteases and growth factors, which promote the deleterious changes in epithelial, stromal, and parenchymal cells and induce airway remodeling, leading to airway obstruction and impaired ventilation in COPD patients [97]. In addition, ROS and inflammatory mediators can also increase the production of coagulation factors, activate the coagulation system, and lead to the hypercoagulable state of patients [57].

It has been shown that activation of the PI3K/AKT pathway is associated with oxidative stress, release of inflammatory mediators, and airway remodeling in COPD [92]. Several PI3K protein inhibitors have been shown to reduce the ROS production and inflammatory cell infiltration and to modulate protease dysregulation to mitigate COPD progression in clinical or preclinical studies [98–100]. XFZYD effectively treats COPD by intervening in the PI3K/AKT pathway, ameliorating oxidative stress and inflammatory responses, relieving the airway remodeling and ventilation

disorders, correcting hypercapnia and hypoxemia, and alleviating hypercoagulability of patients, suggesting the consistency between the results of meta-analysis and network pharmacology predictions and revealing an important mechanism for the treatment of COPD with XFZYD.

4.4. Strengths and Limitations. As far as we know, this is the first systematic and comprehensive evaluation and exploration of the efficacy and molecular mechanisms of XFZYD in the treatment of COPD. Some alternative medicine approaches usually are not taken into consideration in clinical practice because of the lack of evidence of efficacy and unclear mechanisms of action. Notably, the meta-analysis and network pharmacology techniques have become well-established tools for investigating the disease treatment strategies. Such researches can help to reveal the therapeutic potential and advantages of complementary and alternative medicine, especially TCM, for various diseases. In our study, the combination of meta-analysis and network pharmacology may provide more specific information about the results of clinical trials, that is, the molecular mechanism of XFZYD for COPD, which will more rationally explain why XFZYD is a potential strategy for COPD. Moreover, it will also facilitate the subsequent research of XFZYD and even the development work of potent drugs for COPD. As a consequence, a research model that uses meta-analysis and network pharmacology in combination is a worthy advocating tool to explore the clinical potential and mechanisms of action of treatment options, especially in alternative and complementary medicine fields.

Our current study still has limitations. Firstly, the lack of high-quality, multicenter, randomized controlled studies and the fact that some of the included original studies did not mention the details of random assignment principles and blinding may reduce the quality of the present meta-analysis. Secondly, the absence of original studies using XFZYD alone as the treatment group and placebo as the control group resulted in the inability to assess the clinical efficacy of XFZYD alone in the treatment of COPD. Finally, due to the inadequate reporting of adverse events, we were unable to conduct a comprehensive evaluation of the safety of XFZYD on COPD. More high-quality, multicenter, rigorous RCTs are demanded to investigate the efficacy and safety of XFZYD in the treatment of COPD

in further research. Finally, the exploration of molecular mechanisms based on network pharmacology is contributed to reveal the multicomponent, multitarget, and multipathway basis of XFZYD in the treatment of COPD. However, the experiments with animal models are still needed to provide stronger evidence for analysis results of network pharmacology, which is our future direction.

5. Conclusion

XFZYD may be a potential adjuvant therapeutic strategy for the treatment of COPD. Intervention with XFZYD is beneficial in restoring pulmonary function and improving gas retention and coagulation disorders in COPD patients in clinical practice. The early use of XFZYD during the stable phase and AECOPD may help patients to reduce the appearance of dangerous clinical outcomes. XFZYD, which has many active ingredients, determines its multitarget and multipathway mechanism in the treatment of COPD, and the intervention of the PI3K/AKT pathway may be the critical mechanism. In conclusion, this study provides some evidence for the clinical application of XFZYD.

Data Availability

The original contributions presented in the study are included in the article/supplementary materials; further inquiries can be directed to the corresponding author.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors' Contributions

Y.H. was responsible for conceptualization, methodology, and writing—original draft; Y.L. was responsible for methodology, software, and visualization; Q.R. was responsible for software and validation; Q.G. was responsible for investigation; W.H. was responsible for writing—review and editing and project administration.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (82074329).

Supplementary Materials

Supplementary 1. Supplementary Table 1: the searching details in each database. Supplementary Table 2: the specific dosages of XFZYD and the treatment strategies of the control group in each study.

Supplementary 2. Supplementary Figure 1: the sensitivity analysis results of the FEV1/FVC (%).

Supplementary 3. Supplementary Figure 2: the quality of evidence assessment by GRADE profiler.

References

- [1] S. Marsh, S. Aldington, P. Shirtcliffe, M. Weatherall, and R. Beasley, "Smoking and COPD: what really are the risks?," *The European Respiratory Journal*, vol. 28, no. 4, pp. 883–884, 2006.
- [2] K. F. Rabe and H. Watz, "Chronic obstructive pulmonary disease," *Lancet*, vol. 389, no. 10082, pp. 1931–1940, 2017.
- [3] K. Szalontai, N. Gémes, J. Furák et al., "Chronic obstructive pulmonary disease: epidemiology, biomarkers, and paving the way to lung cancer," *Journal of Clinical Medicine*, vol. 10, no. 13, p. 2889, 2021.
- [4] D. C. Flenley, "Chronic obstructive pulmonary disease," *Disease-a-Month*, vol. 34, no. 9, pp. 543–599, 1988.
- [5] J. B. Soriano, P. J. Kendrick, K. R. Paulson et al., "Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," *Respiratory Medicine*, vol. 8, no. 6, pp. 585–596, 2020.
- [6] T. Vos, A. D. Flaxman, M. Naghavi et al., "Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the global burden of disease study 2010," *Lancet*, vol. 380, no. 9859, pp. 2163–2196, 2012.
- [7] GBD, "Risk factors collaborators global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the global burden of disease study 2016," *The Lancet*, vol. 390, pp. 1345–1422, 2017.
- [8] S. S. Salvi and P. J. Barnes, "Chronic obstructive pulmonary disease in non-smokers," *The Lancet*, vol. 374, no. 9691, pp. 733–743, 2009.
- [9] Z. Quan, G. Yan, Z. Wang et al., "Current status and preventive strategies of chronic obstructive pulmonary disease in China: a literature review," *Journal of Thoracic Disease*, vol. 13, no. 6, pp. 3865–3877, 2021.
- [10] T. R. L. Lima, V. P. Almeida, A. S. Ferreira, F. S. Guimarães, and A. J. Lopes, "Handgrip strength and pulmonary disease in the elderly: what is the link?," *Aging and Disease*, vol. 10, no. 5, pp. 1109–1129, 2019.
- [11] H. C. Do-Umehara, C. Chen, Q. Zhang et al., "Epithelial cell-specific loss of function of *Miz1* causes a spontaneous COPD-like phenotype and up-regulates *Ace2* expression in mice," *Science Advances*, vol. 6, no. 33, 2020.
- [12] J. Qin, Y. Qin, Y. Wu et al., "Application of albumin/globulin ratio in elderly patients with acute exacerbation of chronic obstructive pulmonary disease," *Journal of Thoracic Disease*, vol. 10, no. 8, pp. 4923–4930, 2018.
- [13] P. Y. Tiew, F. W. S. Ko, J. K. Narayana et al., "High-risk clinical and inflammatory clusters in COPD of Chinese descent," *Chest*, vol. 158, no. 1, pp. 145–156, 2020.
- [14] E. S. Ford, J. B. Croft, D. M. Mannino, A. G. Wheaton, X. Zhang, and W. H. Giles, "COPD surveillance—United States, 1999–2011," *Chest*, vol. 144, no. 1, pp. 284–305, 2013.
- [15] C. F. Vogelmeier, G. J. Criner, F. J. Martinez et al., "Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary," *American Journal of Respiratory and Critical Care Medicine*, vol. 195, no. 5, pp. 557–582, 2017.
- [16] M. Cazzola, P. Rogliani, D. Stolz, and M. G. Matera, "Pharmacological treatment and current controversies in COPD," *F1000Research*, vol. 8, p. 1533, 2019.

- [17] D. Singh, "Pharmacological treatment of stable chronic obstructive pulmonary disease," *Respirology*, vol. 26, no. 7, pp. 643–651, 2021.
- [18] S. Sethi, "Infectious etiology of acute exacerbations of chronic bronchitis," *Chest*, vol. 117, no. 5, pp. 380S–385S, 2000.
- [19] P. J. Barnes, "Identifying molecular targets for new drug development for chronic obstructive pulmonary disease: what does the future hold?," *Seminars in Respiratory and Critical Care Medicine*, vol. 36, no. 4, pp. 508–522, 2015.
- [20] M. G. Matera, M. Cazzola, and C. Page, "Prospects for COPD treatment," *Current Opinion in Pharmacology*, vol. 56, pp. 74–84, 2021.
- [21] L. Ren, X.-Y. Guo, F. Gao, M.-L. Jin, and X.-N. Song, "Identification of the perturbed metabolic pathways associating with renal fibrosis and evaluating metabolome changes of pre-treatment with Astragalus polysaccharide through liquid chromatography quadrupole time-of-flight mass spectrometry," *Frontiers in Pharmacology*, vol. 10, p. 1623, 2019.
- [22] Z. Xing, Z. Xia, W. Peng et al., "Xuefu Zhuyu decoction, a traditional Chinese medicine, provides neuroprotection in a rat model of traumatic brain injury via an anti-inflammatory pathway," *Scientific Reports*, vol. 6, no. 1, p. 20040, 2016.
- [23] J. Wang, X. Yang, F. Chu et al., "The effects of Xuefu Zhuyu and Shengmai on the evolution of syndromes and inflammatory markers in patients with unstable angina pectoris after percutaneous coronary intervention: a randomised controlled clinical trial," *Evidence-based Complementary and Alternative Medicine*, vol. 2013, Article ID 896467, 9 pages, 2013.
- [24] F. Kui, W. Gu, F. Gao et al., "Research on effect and mechanism of Xuefu Zhuyu decoction on CHD based on meta-analysis and network pharmacology," *Evidence-based Complementary and Alternative Medicine*, vol. 2021, Article ID 9473531, 15 pages, 2021.
- [25] S. Zhang, Z.-L. Chen, Y.-P. Tang, J.-L. Duan, and K.-W. Yao, "Efficacy and safety of Xue-Fu-Zhu-Yu decoction for patients with coronary heart disease: a systematic review and meta-analysis," *Evidence-based Complementary and Alternative Medicine*, vol. 2021, Article ID 9931826, 19 pages, 2021.
- [26] J. Guo, B. Li, W. Wu, Z. Wang, F. Wang, and T. Guo, "Chinese herbal medicines compared with N-acetylcysteine for the treatment of idiopathic pulmonary fibrosis: a systematic review of randomized controlled trials," *Evidence-based Complementary and Alternative Medicine*, vol. 2019, Article ID 5170638, 18 pages, 2019.
- [27] Q. Li, C. Zhai, Z. Shi, J. Ju, and G. Wang, "Clinical efficacy and correlation analysis in elderly patients of chronic obstructive pulmonary disease with Xuefu Zhuyu capsule," *Chinese Traditional and Herbal Drugs (Zhong Cao Yao)*, vol. 42, pp. 4268–4276, 2021.
- [28] M. J. Page, J. E. McKenzie, P. M. Bossuyt et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, article n71, 2021.
- [29] J. A. C. Sterne, J. Savović, M. J. Page et al., "RoB 2: a revised tool for assessing risk of bias in randomised trials," *BMJ*, vol. 366, article 14898, 2019.
- [30] N. Shrine, A. L. Guyatt, A. M. Erzurumluoglu et al., "New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries," *Nature Genetics*, vol. 51, no. 3, pp. 481–493, 2019.
- [31] J.-P. Liu, "GRADE methods in traditional medicine," *Integrative Medicine Research*, vol. 11, no. 2, article 100836, 2022.
- [32] H.-Y. Xu, Y.-Q. Zhang, Z.-M. Liu et al., "ETCM: an encyclopaedia of traditional Chinese medicine," *Nucleic Acids Research*, vol. 47, no. D1, pp. D976–D982, 2019.
- [33] Z. Ma, Y. Liu, C. Li, Y. Zhang, and N. Lin, "Repurposing a clinically approved prescription Colquhounia root tablet to treat diabetic kidney disease via suppressing PI3K/AKT/NF-KB activation," *Chinese Medicine*, vol. 17, no. 1, p. 2, 2022.
- [34] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [35] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.
- [36] B. T. Sherman, M. Hao, J. Qiu et al., "DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)," *Nucleic Acids Research*, vol. 50, no. W1, pp. W216–W221, 2022.
- [37] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of Computational Chemistry*, vol. 31, no. 2, pp. 455–461, 2010.
- [38] J. Mao, "Clinical observation on treating COPD with the Mahuang Zhuyu decoction," *Clinical Journal of Chinese Medicine (Zhong Yi Lin Chuang Yan Jiu)*, vol. 8, pp. 69–71, 2016.
- [39] C. Wu, "Clinical observation of Xuefu Zhuyu caps on treating elderly patients stable chronic obstructive pulmonary disease," *Journal of Medical Information (Yi Xue Xin Xi)*, vol. 24, pp. 215–215, 2011.
- [40] J. Zhang, "The effect of Xuefu Zhuyu decoction on blood rheology and pulmonary function in patients with acute exacerbation of chronic obstructive pulmonary disease," *Henan Traditional Chinese Medicine (He Nan Zhong Yi)*, vol. 34, pp. 469–470, 2014.
- [41] Y. Ma, "The effect of Xuefu Zhuyu decoction on hemorheology and pulmonary function in patients with acute exacerbation of chronic obstructive pulmonary disease," *Chinese Journal of Trauma and Disability Medicine (Zhong Guo Shang Can Yi Xue)*, vol. 24, pp. 112–113, 2016.
- [42] Y. Li, "A study on the therapeutic effect of Xuefu Zhuyu decoction on the stable stage of chronic obstructive pulmonary disease," *Practical Clinical Journal of Integrated Traditional Chinese and Western Medicine (Shi Yong Zhong Xi Yi Jie He Lin Chuang)*, vol. 19, pp. 134–136, 2019.
- [43] L. Fang and Z. Wang, "Effect of Xuefu Zhuyu decoction on improving lung function and pre-thrombosis state of chronic obstructive pulmonary disease patients with acute exacerbation," *Zhejiang Journal of Integrated Traditional Chinese and Western Medicine (Zhe Jiang Zhong Xi Yi Jie He Za Zhi)*, vol. 25, pp. 833–836, 2015.
- [44] J. Cai, M. Chen, and X. Zheng, "Clinical study on the treatment of exacerbation of chronic obstructive pulmonary disease with the addition and subtraction of Xuefu Zhuyu decoction," *Clinical Journal of Traditional Chinese Medicine (Zhong Yi Yao Lin Chuang Za Zhi)*, vol. 26, pp. 231–232, 2014.
- [45] Q. Liu, Q. Gu, Q. Xia, J. Cheng, and S. Liu, "Efficacy of the Xuefu Zhuyu decoction on coagulation index in patients with

- acute exacerbation chronic obstructive pulmonary disease,” *Clinical Journal of Chinese Medicine (Zhong Yi Lin Chuang Yan Jiu)*, vol. 10, pp. 12–14, 2018.
- [46] Z. Wang and Y. Jin, “The effect of Xuefu Zhuyu decoction additive on hemorheology in patients with acute exacerbation of chronic obstructive pulmonary disease,” *Journal of GuiZhou University of Traditional Chinese Medicine (Gui Zhou Zhong Yi Yao Da Xue Xue Bao)*, vol. 33, pp. 145–147, 2011.
- [47] X. Li and P. Xiao, “Treatment of 40 cases of chronic obstructive pulmonary disease in remission with the additive treatment of Xuefu Zhuyu decoction,” *Global Traditional Chinese Medicine (Huan Qiu Zhong Yi Yao)*, vol. 10, pp. 963–965, 2017.
- [48] X. Hong and G. Wang, “Clinical observation on the treatment of acute exacerbation of blood stasis type chronic obstructive pulmonary disease by combining Xuefu Zhuyu decoction with Western medicine,” *Journal of Guangxi University of Chinese Medicine (Guang Xi Zhong Yi Yao Da Xue Xue Bao)*, vol. 16, pp. 10–12, 2013.
- [49] Z. Li, “The efficacy of combined Chinese and Western medicine in treating 30 cases of acute exacerbation of chronic obstructive pulmonary disease,” *Hunan Journal of Traditional Chinese Medicine (Hu Nan Zhong Yi Za Zhi)*, vol. 31, pp. 42–43, 2015.
- [50] J. Liu, J. Liu, X. Tong et al., “Network pharmacology prediction and molecular docking-based strategy to discover the potential pharmacological mechanism of Huai Hua San against ulcerative colitis,” *Drug Design, Development and Therapy*, vol. 15, pp. 3255–3276, 2021.
- [51] P. Almagro, P. Martinez-Cambor, J. B. Soriano et al., “Finding the best thresholds of FEV1 and dyspnea to predict 5-year survival in COPD patients: the COCOMICS study,” *PLoS One*, vol. 9, no. 2, article e89866, 2014.
- [52] M. Cazzola, C. P. Page, L. Calzetta, and M. G. Matera, “Emerging anti-inflammatory strategies for COPD,” *The European Respiratory Journal*, vol. 40, no. 3, pp. 724–741, 2012.
- [53] J. Schrijver, A. Lenferink, M. Brusse-Keizer et al., “Self-management interventions for people with chronic obstructive pulmonary disease,” *Cochrane Database of Systematic Reviews*, vol. 2022, no. 1, article CD002990, 2022.
- [54] L. Yang, D. Zhong, Y. Zhang et al., “Tai chi for chronic obstructive pulmonary disease (COPD): an overview of systematic reviews,” *Int J Gen Med*, vol. 14, pp. 3017–3033, 2021.
- [55] O. Marrone, A. Salvaggio, and G. Insalaco, “Respiratory disorders during sleep in chronic obstructive pulmonary disease,” *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 1, no. 4, pp. 363–372, 2006.
- [56] C. Schneider, U. Bothner, S. S. Jick, and C. R. Meier, “Chronic obstructive pulmonary disease and the risk of cardiovascular diseases,” *European Journal of Epidemiology*, vol. 25, no. 4, pp. 253–260, 2010.
- [57] C. Kyriakopoulos, A. Gogali, K. Kostikas, and A. Konstantinidis, “Hypercoagulable state in COPD—a comprehensive literature review,” *Diagnostics*, vol. 11, no. 8, 2021.
- [58] J. D. Olson, “D-dimer: an overview of hemostasis and fibrinolysis, assays, and clinical applications,” *Advances in Clinical Chemistry*, vol. 69, pp. 1–46, 2015.
- [59] H. Pang, L. Wang, J. Liu et al., “The prevalence and risk factors of venous thromboembolism in hospitalized patients with acute exacerbation of chronic obstructive pulmonary disease,” *The Clinical Respiratory Journal*, vol. 12, no. 11, pp. 2573–2580, 2018.
- [60] M. Liu, R. Hu, X. Jiang, and X. Mei, “Coagulation dysfunction in patients with AECOPD and its relation to infection and hypercapnia,” *Journal of Clinical Laboratory Analysis*, vol. 35, no. 4, article e23733, 2021.
- [61] C. Xiao, S. Cheng, R. Li et al., “Isoforskolin alleviates AECOPD by improving pulmonary function and attenuating inflammation which involves downregulation of Th17/IL-17A and NF-KB/NLRP3,” *Frontiers in Pharmacology*, vol. 12, article 721273, 2021.
- [62] W. H. Meeraus, B. M. DeBarmore, H. Mullerova, W. A. Fahy, and V. S. Benson, “Terms and definitions used to describe recurrence, treatment failure and recovery of acute exacerbations of COPD: a systematic review of observational studies,” *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 16, pp. 3487–3502, 2021.
- [63] A. I. Ritchie and J. A. Wedzicha, “Definition, causes, pathogenesis, and consequences of chronic obstructive pulmonary disease exacerbations,” *Clinics in Chest Medicine*, vol. 41, no. 3, pp. 421–438, 2020.
- [64] W. G. Melsen, M. C. J. Bootsma, M. M. Rovers, and M. J. M. Bonten, “The effects of clinical and statistical heterogeneity on the predictive values of results from meta-analyses,” *Clinical Microbiology and Infection*, vol. 20, no. 2, pp. 123–129, 2014.
- [65] P. Wang, S. Ding, L. Sun et al., “Characteristics and differences of gut microbiota in patients with different traditional Chinese medicine syndromes of colorectal cancer and normal population,” *Journal of Cancer*, vol. 11, no. 24, pp. 7357–7367, 2020.
- [66] F. Liu, F. Liu, J. Li, Y. Du, and X. Han, “Effectiveness and safety of treating negative emotions after PCI from the perspective of qi and blood: a systematic review and meta-analysis,” *Evidence-based Complementary and Alternative Medicine*, vol. 2022, Article ID 8604472, 13 pages, 2022.
- [67] Y. Zhang, J. Jiang, H. Shen, Y. Chai, X. Wei, and Y. Xie, “Total flavonoids from *Rhizoma Drynariae* (Gusuibu) for treating osteoporotic fractures: implication in clinical practice,” *Drug Design, Development and Therapy*, vol. 11, pp. 1881–1890, 2017.
- [68] C.-W. Yeh, H.-K. Liu, L.-C. Lin et al., “Xuefu Zhuyu decoction ameliorates obesity, hepatic steatosis, neuroinflammation, amyloid deposition and cognition impairment in metabolically stressed APPswe/PS1dE9 mice,” *Journal of Ethnopharmacology*, vol. 209, pp. 50–61, 2017.
- [69] D. Zhou, J. Liu, Y. Hang et al., “TMT-based proteomics analysis reveals the protective effects of Xuefu Zhuyu decoction in a rat model of traumatic brain injury,” *Journal of Ethnopharmacology*, vol. 258, article 112826, 2020.
- [70] S. Wahab, S. Annadurai, S. S. Abullais et al., “*Glycyrrhiza glabra* (licorice): a comprehensive review on its phytochemistry, biological activities, clinical evidence and toxicology,” *Plants*, vol. 10, no. 12, 2021.
- [71] D. Yu, X. Liu, G. Zhang, Z. Ming, and T. Wang, “Isoliquiritigenin inhibits cigarette smoke-induced COPD by attenuating inflammation and oxidative stress via the regulation of the Nrf2 and NF-KB signaling pathways,” *Frontiers in Pharmacology*, vol. 9, p. 1001, 2018.
- [72] S.-H. Kim, J.-H. Hong, W.-K. Yang et al., “Herbal combination medication of *Glycyrrhiza glabra*, *gastache rugosa* containing glycyrrhizic acid, tilianin inhibits neutrophilic lung inflammation by affecting CXCL2, interleukin-17/STAT3

- signal pathways in a murine model of COPD,” *Nutrients*, vol. 12, no. 4, 2020.
- [73] W.-L. Wei, R. Zeng, C.-M. Gu, Y. Qu, and L.-F. Huang, “*Angelica sinensis* in China—a review of botanical profile, ethnopharmacology, phytochemistry and chemical analysis,” *Journal of Ethnopharmacology*, vol. 190, pp. 116–141, 2016.
- [74] C.-C. Lin, C.-J. Liou, C.-Y. Chiang, W.-Y. Huang, and W.-C. Huang, “Danggui Buxue Tang attenuates eosinophil infiltration and airway hyperresponsiveness in asthmatic mice,” *Annals of Allergy, Asthma & Immunology*, vol. 107, no. 6, pp. 501–509, 2011.
- [75] C.-H. Xie, M.-S. Zhang, Y.-F. Zhou et al., “Chinese medicine *Angelica sinensis* suppresses radiation-induced expression of TNF- α and TGF- β 1 in mice,” *Oncology Reports*, vol. 15, no. 6, pp. 1429–1436, 2006.
- [76] W. Qian, X. Cai, Q. Qian, D. Wang, and L. Zhang, “*Angelica sinensis* polysaccharide suppresses epithelial-mesenchymal transition and pulmonary fibrosis via a DANCR/AUF-1/FOXO3 regulatory axis,” *Aging and Disease*, vol. 11, no. 1, pp. 17–30, 2020.
- [77] F. Yang, X. Dong, X. Yin, W. Wang, L. You, and J. Ni, “*Radix Bupleuri*: a review of traditional uses, botany, phytochemistry, pharmacology, and toxicology,” *BioMed Research International*, vol. 2017, Article ID 7597596, 22 pages, 2017.
- [78] J. Xie, H. Di, H. Li, X. Cheng, Y. Zhang, and D. Chen, “*Bupleurum chinense* DC polysaccharides attenuates lipopolysaccharide-induced acute lung injury in mice,” *Phytomedicine*, vol. 19, no. 2, pp. 130–137, 2012.
- [79] Y. Zhou, D. J. Schneider, and M. R. Blackburn, “Adenosine signaling and the regulation of chronic lung disease,” *Pharmacology & Therapeutics*, vol. 123, no. 1, pp. 105–116, 2009.
- [80] R. Polosa and M. R. Blackburn, “Adenosine receptors as targets for therapeutic intervention in asthma and chronic obstructive pulmonary disease,” *Trends in Pharmacological Sciences*, vol. 30, no. 10, pp. 528–535, 2009.
- [81] M. Versluis, N. ten Hacken, D. Postma et al., “Adenosine receptors in COPD and asymptomatic smokers: effects of smoking cessation,” *Virchows Archiv*, vol. 454, no. 3, pp. 273–281, 2009.
- [82] R. Polosa, “Adenosine-receptor subtypes: their relevance to adenosine-mediated responses in asthma and chronic obstructive pulmonary disease,” *The European Respiratory Journal*, vol. 20, no. 2, pp. 488–496, 2002.
- [83] M. K. Han, T. A. Barreto, F. J. Martinez, A. T. Comstock, and U. S. Sajjan, “Randomised clinical trial to determine the safety of quercetin supplementation in patients with chronic obstructive pulmonary disease,” *Respiratory Research*, vol. 7, no. 1, 2020.
- [84] M. Farazuddin, R. Mishra, Y. Jing, V. Srivastava, A. T. Comstock, and U. S. Sajjan, “Quercetin prevents rhinovirus-induced progression of lung disease in mice with COPD phenotype,” *PLoS One*, vol. 13, no. 7, article e0199612, 2018.
- [85] S. Ganesan, A. N. Faris, A. T. Comstock et al., “Quercetin prevents progression of disease in elastase/LPS-exposed mice by negatively regulating MMP expression,” *Respiratory Research*, vol. 11, no. 1, p. 131, 2010.
- [86] A. Mitani, A. Azam, C. Vuppusetty, K. Ito, N. Mercado, and P. J. Barnes, “Quercetin restores corticosteroid sensitivity in cells from patients with chronic obstructive pulmonary disease,” *Experimental Lung Research*, vol. 43, pp. 417–425.
- [87] F. Lemarié, E. Beauchamp, G. Drouin, P. Legrand, and V. Rioux, “Dietary caprylic acid and ghrelin O-acyltransferase activity to modulate octanoylated ghrelin functions: what is new in this nutritional field?,” *Prostaglandins, Leukotrienes, and Essential Fatty Acids*, vol. 135, pp. 121–127, 2018.
- [88] M. A. Altinoz, A. Ozpinar, and T. N. Seyfried, “Caprylic (octanoic) acid as a potential fatty acid chemotherapeutic for glioblastoma,” *Prostaglandins, Leukotrienes, and Essential Fatty Acids*, vol. 159, article 102142, 2020.
- [89] C. Wang, J. Zhou, J. Wang et al., “Progress in the mechanism and targeted drug therapy for COPD,” *Signal Transduction and Targeted Therapy*, vol. 5, no. 1, p. 248, 2020.
- [90] J. E. Repine, A. Bast, and I. Lankhorst, “Oxidative stress in chronic obstructive pulmonary disease,” *American Journal of Respiratory and Critical Care Medicine*, vol. 156, no. 2, pp. 341–357, 1997.
- [91] M. Saetta, G. Turato, P. Maestrelli, C. E. Mapp, and L. M. Fabbri, “Cellular and structural bases of chronic obstructive pulmonary disease,” *American Journal of Respiratory and Critical Care Medicine*, vol. 163, no. 6, pp. 1304–1309, 2001.
- [92] H. Jiang, P. W. Abel, M. L. Toews et al., “Phosphoinositide 3-kinase gamma regulates airway smooth muscle contraction by modulating calcium oscillations,” *The Journal of Pharmacology and Experimental Therapeutics*, vol. 334, no. 3, pp. 703–709, 2010.
- [93] X. Sun, L. Chen, and Z. He, “PI3K/Akt-Nrf2 and anti-inflammation effect of macrolides in chronic obstructive pulmonary disease,” *Current Drug Metabolism*, vol. 20, no. 4, pp. 301–304, 2019.
- [94] H.-X. Zhang, J.-J. Yang, S.-A. Zhang et al., “HIF-1 α promotes inflammatory response of chronic obstructive pulmonary disease by activating EGFR/PI3K/AKT pathway,” *European Review for Medical and Pharmacological Sciences*, vol. 22, no. 18, pp. 6077–6084, 2018.
- [95] Q. Li, G. Wang, S. H. Xiong et al., “Bu-Shen-Fang-Chuan formula attenuates cigarette smoke-induced inflammation by modulating the PI3K/Akt-Nrf2 and NF- κ B signalling pathways,” *Journal of Ethnopharmacology*, vol. 261, article 113095, 2020.
- [96] S. F. van Eeden and D. D. Sin, “Oxidative stress in chronic obstructive pulmonary disease: a lung and systemic process,” *Canadian Respiratory Journal*, vol. 20, Article ID 509130, 3 pages, 2013.
- [97] Y. Wang, J. Xu, Y. Meng, I. M. Adcock, and X. Yao, “Role of inflammatory cells in airway remodeling in COPD,” *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 13, pp. 3341–3348.
- [98] H. Watz, D. Bock, M. Meyer et al., “Inhaled pan-selectin antagonist bimosiamose attenuates airway inflammation in COPD,” *Pulmonary Pharmacology & Therapeutics*, vol. 26, no. 2, pp. 265–270, 2013.
- [99] J. Doukas, L. Eide, K. Stebbins et al., “Aerosolized phosphoinositide 3-kinase gamma/delta inhibitor TG100-115 [3-[2,4-diamino-6-(3-hydroxyphenyl)pteridin-7-Yl]phenol] as a therapeutic candidate for asthma and chronic obstructive pulmonary disease,” *The Journal of Pharmacology and Experimental Therapeutics*, vol. 328, no. 3, pp. 758–765, 2009.
- [100] R. Vlahos, P. A. B. Wark, G. P. Anderson, and S. Bozinovski, “Glucocorticosteroids differentially regulate MMP-9 and neutrophil elastase in COPD,” *PLoS One*, vol. 7, no. 3, article e33277, 2012.

Retraction

Retracted: HMGB3 Targeted by miR-145-5p Impacts Proliferation, Migration, Invasion, and Apoptosis of Breast Cancer Cells

Computational and Mathematical Methods in Medicine

Received 26 September 2023; Accepted 26 September 2023; Published 27 September 2023

Copyright © 2023 Computational and Mathematical Methods in Medicine. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Y. Hu, D. Wu, R. Huang, and Z. Shi, "HMGB3 Targeted by miR-145-5p Impacts Proliferation, Migration, Invasion, and Apoptosis of Breast Cancer Cells," *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 1954099, 12 pages, 2022.

Research Article

HMGB3 Targeted by miR-145-5p Impacts Proliferation, Migration, Invasion, and Apoptosis of Breast Cancer Cells

Yangying Hu ¹, Deqi Wu ¹, Rong Huang ², and Zhijie Shi ¹

¹Department of Thyroid and Breast Diagnosis and Treatment Center, Shulan (Hangzhou) Hospital Affiliated to Zhejiang Shuren University Shulan International Medical College, Hangzhou, Zhejiang 310000, China

²Pathology Department, Shulan (Hangzhou) Hospital Affiliated to Zhejiang Shuren University Shulan International Medical College, Hangzhou, Zhejiang 310000, China

Correspondence should be addressed to Zhijie Shi; zhijieshi@163.com

Received 22 July 2022; Accepted 30 August 2022; Published 10 November 2022

Academic Editor: Tao Huang

Copyright © 2022 Yangying Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study focused on the investigation into how HMGB3 works in breast cancer (BC) progression. Firstly, we analyzed the relationship between HMGB3 and BC patients through the TCGA database. We performed qRT-PCR for determining the HMGB3 mRNA level and Western blot for detecting the protein level of HMGB3 in BC cell lines. CCK-8, flow cytometry, transwell, and wound healing assays were utilized to detect the effect of HMGB3 on BC cell phenotypes. Next, the prediction of the binding site shared by miR-145-5p and HMGB3 was performed by the bioinformatics method. The targeting relationship between miR-145-5p and HMGB3 was validated by using dual-luciferase assay. Finally, rescue experiments were employed for assessing the effect of the miR-145-5p/HMGB3 axis on BC cells. HMGB3 was demonstrated to have a high-level expression in BC cell lines and facilitated BC progression. On the contrary, miR-145-5p was shown a low-level expression in BC cell lines, which could target HMGB3. miR-145-5p restrained the proliferation, migration, and invasion of BC cells via inhibiting HMGB3.

1. Introduction

As a commonly diagnosed cancer in women, breast cancer (BC) is a big threat to the global women's health [1]. Recently, common clinical therapies of BC include surgery treatment, hormone therapy, cytotoxic chemotherapy, immunotherapy, and target therapy [2–4]. The patient's survival rate is increased by those therapies, but mortality remains relatively high. The main cause of the death of patients with BC is the metastasis of tumor cells [5]. Hence, continuous exploration into the molecular mechanism of BC progression is indispensable, thus to formulate a more effective treatment.

High mobility group box 3 (HMGB3) is first known as an expressed sequence tag during embryonic development [6, 7]. HMGB3 exerts an indispensable part in DNA recombination, repair, and replication. HMGB3 can modulate gene expression, induce innate immune activity, and regulate the balance between differentiation of hematopoietic

stem cells and self-renewal [8, 9]. Besides, HMGB3 is widely involved in regulating pathological states including various cancers, like gastric cancer [10], esophageal cancer [11], and non-small cell lung cancer [12]. Studies demonstrated a high expression level of HMGB3 in BC, and HMGB3 was regarded as a cancer promoter to modulate the development of BC [13, 14]. In BC cells, HMGB3 is upregulated and its silencing impairs cell proliferation *in vitro* and *in vivo* [15]. HDAC3 can increase the expression of HMGB3 by down-regulating miR-130a-3p and promote the immune escape of BC cells [16]. However, there are scarce studies on the molecular mechanism of HMGB3 involvement in BC and whether HMGB3 exerts a regulatory role in BC progression remains an unanswered question.

MicroRNAs (miRNAs) are pivotal for modulating gene expression. The mechanism of miRNAs is to modulate expression of protein-coding genes by causing mRNA cleavage or translational repression [17]. The role of miRNAs in BC has been investigated [18]. Ashirbekov et al. [19]

unveiled the essential part of miR-145-5p in BC which was able to restrain some oncogene expression. miR-145-5p, by targeting SOX2, is able to impede the proliferation of BC cells [20]. But whether there is a connection between the suppressive effect of miR-145-5p and HMGB3 on BC cells has not been revealed.

In this research, the roles of HMGB3 and its upstream regulatory gene miR-145-5p in BC cells were analyzed by bioinformatics analysis and cellular molecular experiments. This study underpins the exploitation of the potential therapeutic targets for BC treatment.

2. Materials and Methods

2.1. Bioinformatics Analysis. Mature miRNA data (normal: 104, tumor: 1,103) and count data of mRNA (normal: 113, tumor: 1,109) of BC were downloaded from The Cancer Genome Atlas (TCGA) database (<https://tcgadata.nci.nih.gov/tcga/tcgaHome2.jsp>) (searching keywords were “breast cancer,” “BRCA,” and “Homo sapiens”). After preprocessing them according to PAM50 [21], we obtained Luminal A, Luminal B, Basal, Her2, and nonsubtype samples. Differential expression of HMGB3 in normal breast tissue and BC tissue was determined by *t*-test. Tumor samples were divided on the basis of the median value of HMGB3 expression in all samples or different subtypes of BC, and the “survival” package was used for the conduction of the survival analysis. Afterwards, the Kruskal-Wallis test was employed for analyzing the association between HMGB3 and clinical features and differences of HMGB3 in different BC subtypes.

With “EdgeR” package ($|\log_{FC}| > 2.0$, $P_{adj} < 0.01$), differential analysis was undertaken on miRNAs. To determine the upstream regulator miRNA of HMGB3, starBase (<http://starbase.sysu.edu.cn/index.php>) and TargetScan (http://www.targetscan.org/vert_72/) databases were applied for prediction. The predicted miRNAs were overlapped with differentially downregulated miRNAs. Pearson correlation analysis was performed on miRNA and HMGB3, and differential expression of miRNA in BC and normal breast tissues was further defined by *t*-test.

2.2. Cell Culture and Transfection. The human mammary epithelial cell line MCF-10A (BNCC102049) and BC cell lines MDA-MB-231 (BNCC337894), MDA-MB-468 (BNCC100687), T47D (BNCC339607), and ZR-75-30 (BNCC100125) were all provided by BeNa Culture Collection (China). MCF-10A, MDA-MB-231, T47D, and ZR-75-30 were all treated with 10% FBS-contained (HyClone, USA) DMEM complete medium (Sigma, USA). MDA-MB-468 was maintained in Leibovitz Medium (L15) (Sigma, USA) with 10% FBS. The above cells were placed in a constant temperature incubator at 37°C with 5% CO₂. miR-mimic, miR-inhibitor, and corresponding negative control (miR-NC) were acquired from GeneChem (China). The following were the sequences: miR-145-5p mimic: 5'-GUCCAGUUUCCAGGAAUCCCU-3', 3'-CAGGUCAAAAGGGUCCUAGGGA-5'; miR-NC: 5'-UUUGUACUACACAAAAGUACUG-3', 3'-AAACAUGAUGUGUUUU

TABLE 1: Primer sequences used in the assay.

Gene	Primer sequences (5'→3')
miR-145-5p	F: GCCGAGGTCCAGTTTCCCA
	R: CTCAACTGGTGTCTGCGTGA
U6 snRNA	F: CTCGCTTCGGCAGCACACA
	R: AACGCTTACGAATTTGCGT
HMGB3	F: GACCAGCTAAGGGAGGCAA
	R: ACAGGAAGAATCCAGACGGT
β -Actin	F: CACCATTGCAATGAGCGGTTCC
	R: AGGTCCTTTCGCGGATGTCCACGT

CAUGAC-5'; miR-inhibitor: 5'-AGGGAUCCUGGGA AACUGGAC-3'; inhibitor-NC: 5'-CAGUACUUUUG UGUAGUACAAA-3'; and HMGB3-specific siRNA (si-HMGB3: 5'-GGUCUUCGCCUUGAUUCAUTT-3', 5'-AUGAUUAAGGCGAAGACCTT-3' were procured from GenePharma (China). pcDNA3.1-HMGB3 plasmid that encodes HMGB3 (oe-HMGB3) and empty pcDNA3.1 plasmid (oe-NC) were constructed by GenePharma (China). The 6-well plates were utilized for BC cell (2×10^5 cells/well) seeding. When cells proliferated into 40–50% confluency, the target plasmids (10 μ g) or synthesized sequences (50 nM) were transiently transfected for 48 h using Lipofectamine 3000 (Thermo Fisher Scientific, USA). Stable cell lines were selected for culture using 3 μ g/mL puromycin (Sigma, USA) for one week.

2.3. Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR). Total RNA extraction was accomplished by TRIzol reagent (Invitrogen, USA), and NanoDrop 2000 system (Thermo Fisher Scientific, USA) was utilized for the detection of RNA concentration. cDNA transcription of miRNA and mRNA was done, miRNA with the miScript IIRT Kit (QIAGEN, Germany), and mRNA with the ipsogen RT Kit (QIAGEN, Germany) and PrimeScript RT Master Mix (TaKaRa, Japan). miRNA expression and mRNA expression were measured with the miScript SYBR Green PCR Kit (QIAGEN, Germany) and SYBR® Premix Ex Taq TM II (TaKaRa, Japan). Applied Biosystems®7500 Real-Time PCR Systems (Thermo Fisher Scientific, USA) was applied for qRT-PCR analysis. U6 spliceosomal small nuclear RNA (U6 snRNA) and β -actin were taken as internal controls. Table 1 displayed the primer sequences. Differences in the relative level of miR-145-5p and HMGB3 mRNA were compared by $2^{-\Delta\Delta Ct}$.

2.4. Cell Counting Kit-8 (CCK-8) Assay. 2×10^3 cells/well were seeded into 96-well plates with 100 μ L DMEM containing 10 μ L FBS per well. The medium was changed with DMEM (100 μ L) plus CCK-8 (10 μ L, Beyotime, China) at a designated time point on indicated days, and cells were cultured for 2 h. Next, absorbance was read at 450 nm (optical density (OD)), and cell viability in each group was estimated using an average OD value.

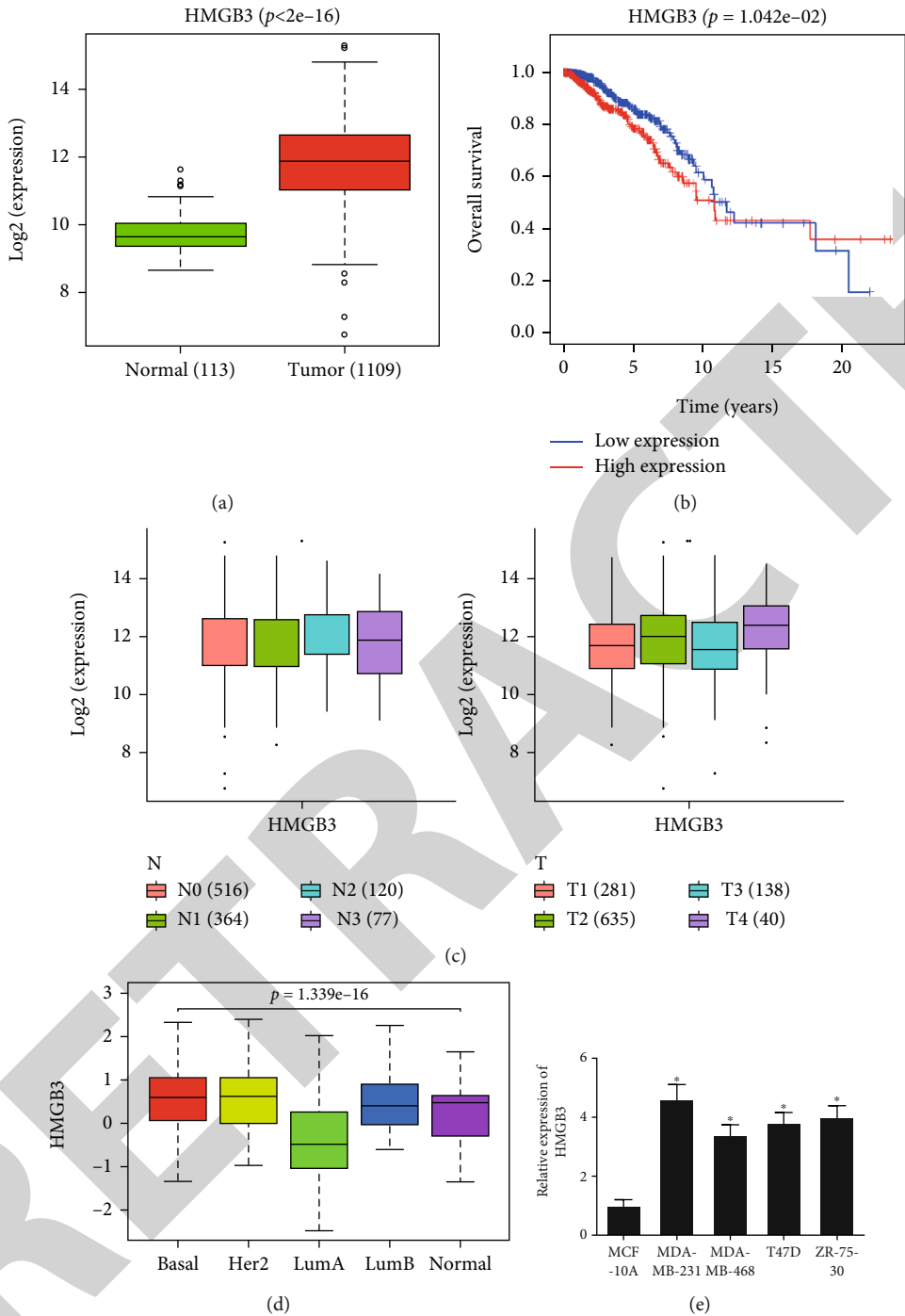


FIGURE 1: Continued.

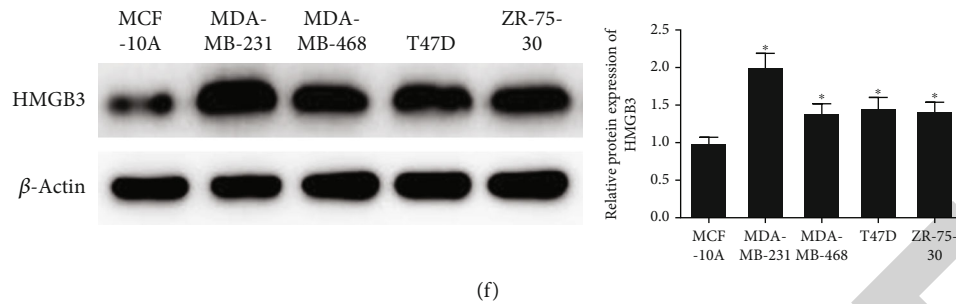


FIGURE 1: HMGB3 expression is substantially high in BC cells. (a) Boxplot of HMGB3 expression in healthy breast tissues and BC tissues (comparison between two groups: $P < 2e - 16$). (b) Survival analysis of HMGB3 expression on BC patient's prognosis (comparison between two groups: $P = 1.042e - 02$). (c) Correlation between HMGB3 and BC clinical features (T stage, N stage, asterisk indicates statistically significant differences, and $P < 0.05$). (d) Relationship between HMGB3 expression and BC subtypes (asterisk indicates statistically significant differences and $P < 0.05$). (e, f) HMGB3 mRNA and protein levels in human breast epithelial cell line MCF-10A and BC cell lines MDA-MB-468, MDA-MB-231, T47D, and ZR-75-30. Results were presented as mean \pm SD. Log-rank test was utilized for survival analysis. Correlation between HMGB3 and clinical features with different subtypes of BC was analyzed by the Kruskal-Wallis test. The others were statistically analyzed by Student's t -test. Samples in each group were plated in triplicate. * $P < 0.05$.

2.5. Wound Healing Assay. Transfected BC cell line MDA-MB-231 and ZR-75-30 were plated into 6-well plates (1×10^6 cells/well) with DMEM complete medium and 10% FBS. When reaching 80% of cell coverage, a 200 μ L sterile plastic tip was applied to scratch the surface of cells to create a line. Cell fractions were rinsed off with PBS. Cells were continuously cultured in medium without serum. At 0h and 24h, wounds were photographed by a phase-contrast microscope. Wound distance was measured to calculate the rate of wound closure at 24h according to the formula hereinafter: wound healing rate = (the wound width of 0h – the wound width of 24h)/0h wound width. This experiment was repeated independently 3 times.

2.6. Transwell Invasion Assay. Matrigel (356234, BD Company, USA) was dissolved at 4°C overnight, diluted with serum-free medium at 1:3, supplemented to the transwell chamber (8 μ m; Corning Inc., USA) with 50 μ L per well, and incubated for 30 min. The transfected cells (5×10^4 cells/well) in FBS-free DMEM were plated in the upper chamber, and the lower chamber was added with DMEM to complete medium plus 10% FBS. Under standard culture conditions, cells were cultured for 24h. Subsequently, a cotton swab was employed to remove cells on the membrane of the transwell chamber. Cells under the membrane were treated with 3 PBS washes, fixed with 4% paraformaldehyde solution for 30 min, and dyed with 0.1% crystal violet for 30 min. 5 fields were randomly chosen and the average number of invading cells in these fields were counted under a microscope.

2.7. Western Blot. RIPA buffer (Sigma, USA) plus phenylmethanesulfonylfluoride was applied to lyse cells and the BCA assay kit (Sigma, USA) to determine the concentration of total proteins. 10% SDS-PAGE was utilized for the separation of equivalent proteins (20 μ g) which were then transferred onto a PVDF membrane (Sigma, USA). 5% skim milk was utilized to seal the membrane for 1h at room temperature. Next, the membrane was maintained with primary antibodies (rabbit anti-HMGB3 (ab72544, Abcam, UK), rab-

bit anti-E-cadherin (ab40772, Abcam, UK), rabbit anti-N-cadherin (ab76011, Abcam, UK), rabbit anti-vimentin (ab92547, Abcam, UK), and rabbit anti- β -actin (ab115777, Abcam, UK)) overnight at 4°C. Afterwards, the membrane was rinsed with PBST thrice each time for 10 min and was maintained with secondary goat anti-rabbit IgG H&L (ab205718, Abcam, UK) for 1h under room temperature. The ECL kit (Solarbio, China) was applied for development. The ChemiDoc XRS Plus imaging analyzer (Bio-Rad, USA) was used for imaging. ImageJ software (NIH, USA) was utilized to analyze protein band gray values.

2.8. Dual-Luciferase Assay. miR-145-5p binding sites into the wild-type (WT)/mutant (MUT) 3' untranslated region (3'UTR) of HMGB3 were synthesized and were then cloned into the pmirGLO vector (Generay, China) to obtain HMGB3-WT and HMGB3-MUT vectors, respectively. Cells were transfected with synthesized plasmids and miR-145-5p mimics/miR-145-5p inhibitor. Luciferase analysis was undertaken after 48h transfection on the dual-luciferase reporter gene analysis system (Promega, USA) on the Fluoroskan Ascent Type 379 fluorescence plate reader (Thermo Fisher Scientific, USA). The standardized luciferase activity was reported as firefly luciferase intensity/Renilla luciferase intensity.

2.9. Flow Cytometry. 1×10^6 transfected cells were gathered for flow cytometry. After two washes with PBS, cells were collected, stained with the Annexin V-FITC/PI cell apoptosis detection kit (BD Biosciences, USA), detected by BD Accuri C6 Plus (BD Biosciences, USA), and analyzed by FlowJo software.

2.10. Analysis of Statistics. Data were processed using SPSS 22.0 and expressed as mean \pm SD. Intragroup differences were analyzed by Student's t -test, while analysis of variance was applied for comparison among groups. The Kruskal Wallis test was utilized to analyze differences in HMGB3 expression in BC with different clinical features and subtypes. Kaplan-Meier was applied for survival analysis.

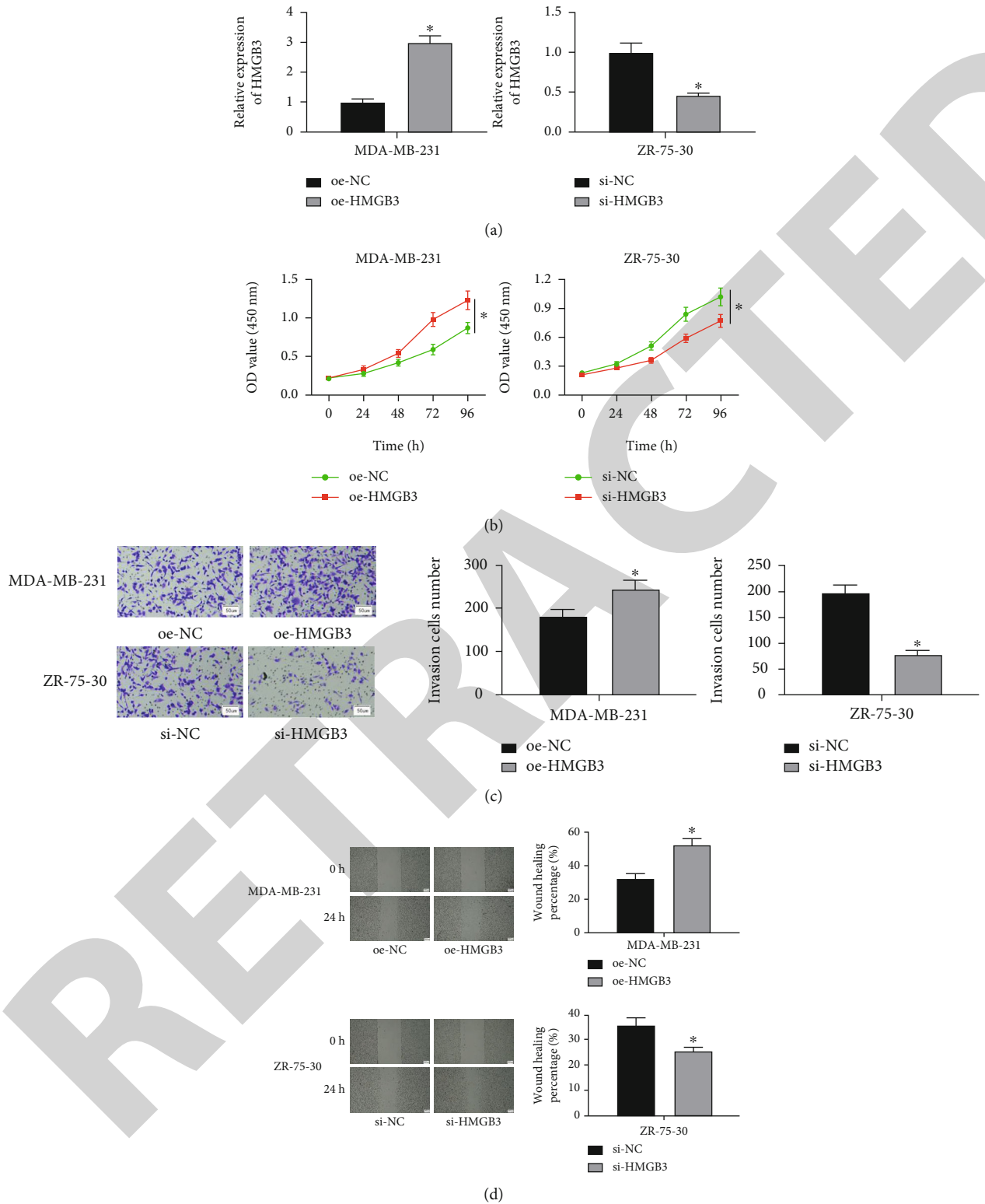


FIGURE 2: Continued.

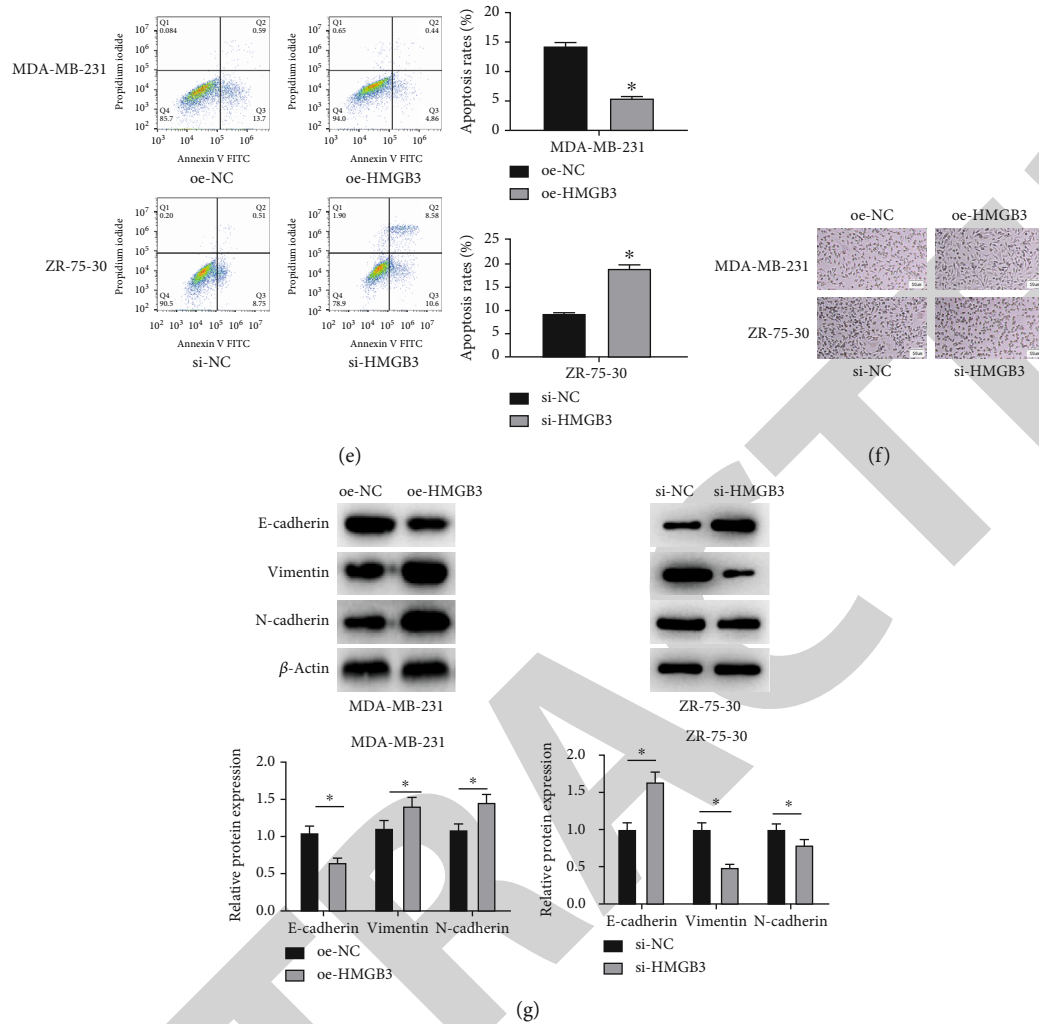


FIGURE 2: Altered HMGB3 expression affects BC cell proliferation, invasion, migration, apoptosis, and EMT. (a) Transfection efficiency of cells in each group. (b) BC cell proliferation. (c) BC cell invasion (magnification: 100x; migrating cell numbers were the average value of 5 random fields). (d) Migratory ability of BC cells (magnification: 40x; wound healing rate = (0 h wound width – 24 h wound width)/0 h wound width). (e) Apoptotic rate of BC cells. (f) Morphology changes of cells were detected by an inverted microscope. (g) Protein expression of E-cadherin, vimentin, and N-cadherin in BC cells. Student's *t*-test was employed for statistics. Samples in each group were plated in triplicate. The experiments were repeated thrice. * $P < 0.05$.

Differences in the survival rate were tested via log-rank. $P \leq 0.05$ was statistically significant. Samples in each group were plated in triplicate and each experiment was repeated 3 times.

3. Results

3.1. HMGB3 Expression Is Substantially High in BC Cells. Differential analysis was performed on mRNA expression data in TCGA. Then, we noted that the HMGB3 level was remarkably higher in BC tissues (Figure 1(a)). According to survival analysis, the HMGB3 high level was implicated in poor survival of BC patients (Figure 1(b)), especially Luminal A- (LumA-) type BC (Supplementary Figure (available here)). Besides, HMGB3 was significantly correlated with clinical features (T stage, N stage) of BC (Figure 1(c)). In addition, the HMGB3 level varied significantly in different subtypes of BC (LumA, LumB, HER2 enriched, basal

like, and normal like) (Figure 1(d)). Next, the HMGB3 level in the human breast epithelial cell line and 4 BC cell lines was detected and it was displayed that HMGB3 mRNA and protein levels were markedly elevated in BC cells (Figures 1(e) and 1(f)). The above results indicated that in BC cells, HMGB3 expression remained at a notably high level. To dive deeper into the functional mechanism of HMGB3 in BC, MDA-MB-231 and ZR-75-30 with the highest HMGB3 expression were utilized for the following experiments.

3.2. Altered HMGB3 Expression Affects BC Cell Proliferation, Invasion, Migration, Apoptosis, and EMT. To figure out the effect of HMGB3 in BC cells, HMGB3 was overexpressed in MDA-MB-231 cells and knocked down in ZR-75-30 cells and the corresponding control groups were constructed. qRT-PCR suggested that HMGB3 was substantially upregulated in the oe-HMGB3 group and downregulated in the

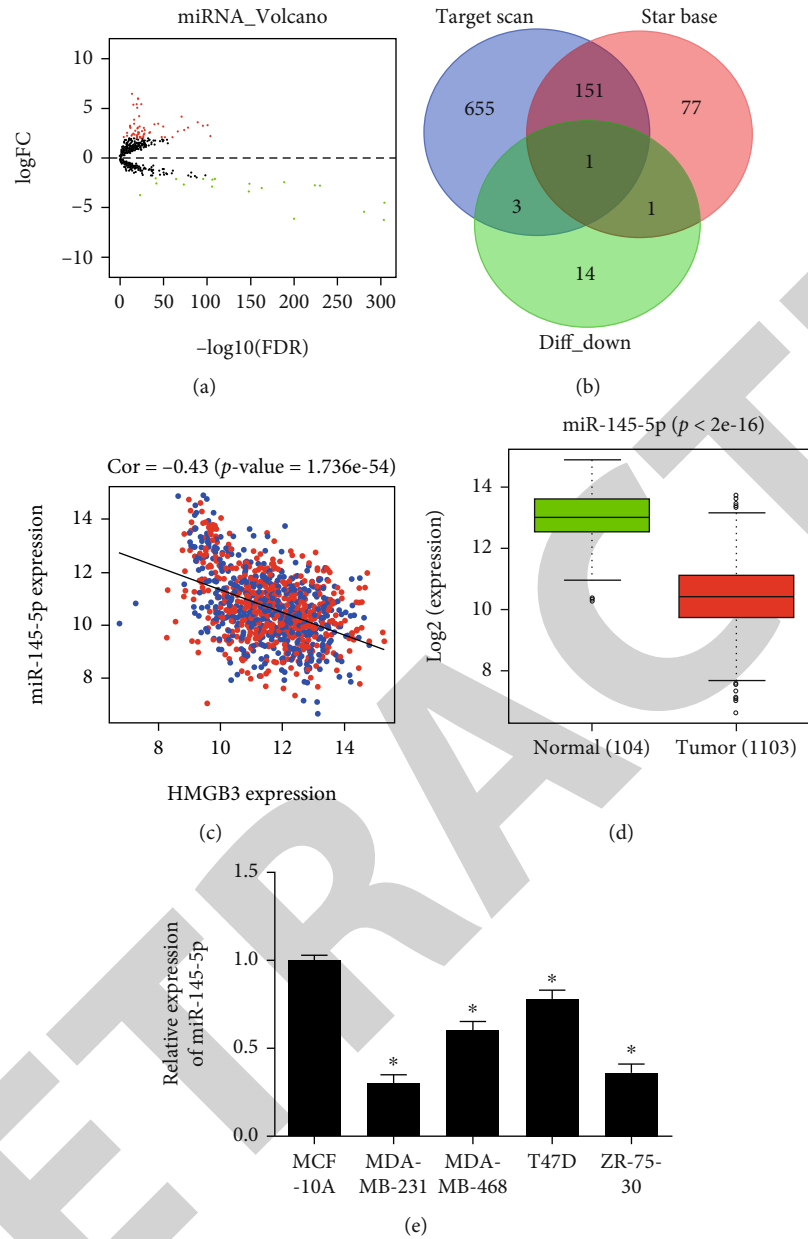


FIGURE 3: Low level of miR-145-5p in BC cells. (a) Volcano plot of differential miRNAs in the TCGA-BRCA dataset (red: 55 upregulated miRNAs; green: 19 downregulated miRNAs). (b) Intersection between predicted upstream regulatory genes of HMGB3 and differentially downregulated miRNAs. (c) Correlation between HMGB3 and miR-145-5p (comparison between two groups: $P = 1.736e-54$). (d) Boxplot of miR-145-5p expression in normal and tumor groups (comparison between two groups: $p < 2e-16$). (e) miR-145-5p level in normal and BC cell lines. Results were shown as mean \pm SD. Data analysis was completed by Student's t -test. Samples in each group were plated in triplicate. $*P < 0.05$, compared with the corresponding NC group.

si-HMGB3 group (Figure 2(a)). Additionally, overexpressing HMGB3 substantially promoted proliferative, invasive, and migratory properties of MDA-MB-231 cells. Knock-down of HMGB3 suppressed those properties of ZR-75-30 cells (Figures 2(b)–2(d)). Moreover, flow cytometry showed that overexpressing HMGB3 markedly downregulated the apoptotic level of MDA-MB-231 cells whereas silencing HMGB3 upregulated ZR-75-30 cell apoptosis (Figure 2(e)). The inverted microscope was utilized for observing the cell morphology of MDA-MB-231 cells. The elongated spindle mesenchymal cells increased after

HMGB3 overexpression. The proportion of paving stone-like cells in ZR-75-30 cells increased after HMGB3 knock-down (Figure 2(f)). With the development of cancer, the elevated expression of mesenchymal cell markers like vimentin and N-cadherin indicates that cancer cells are undergoing epithelial-mesenchymal transition (EMT). The Western blot result displayed that protein expression of E-cadherin (epithelial marker) could be downregulated but protein expression of N-cadherin and vimentin could be upregulated by overexpressed HMGB3. HMGB3 overexpression significantly promoted the EMT process of

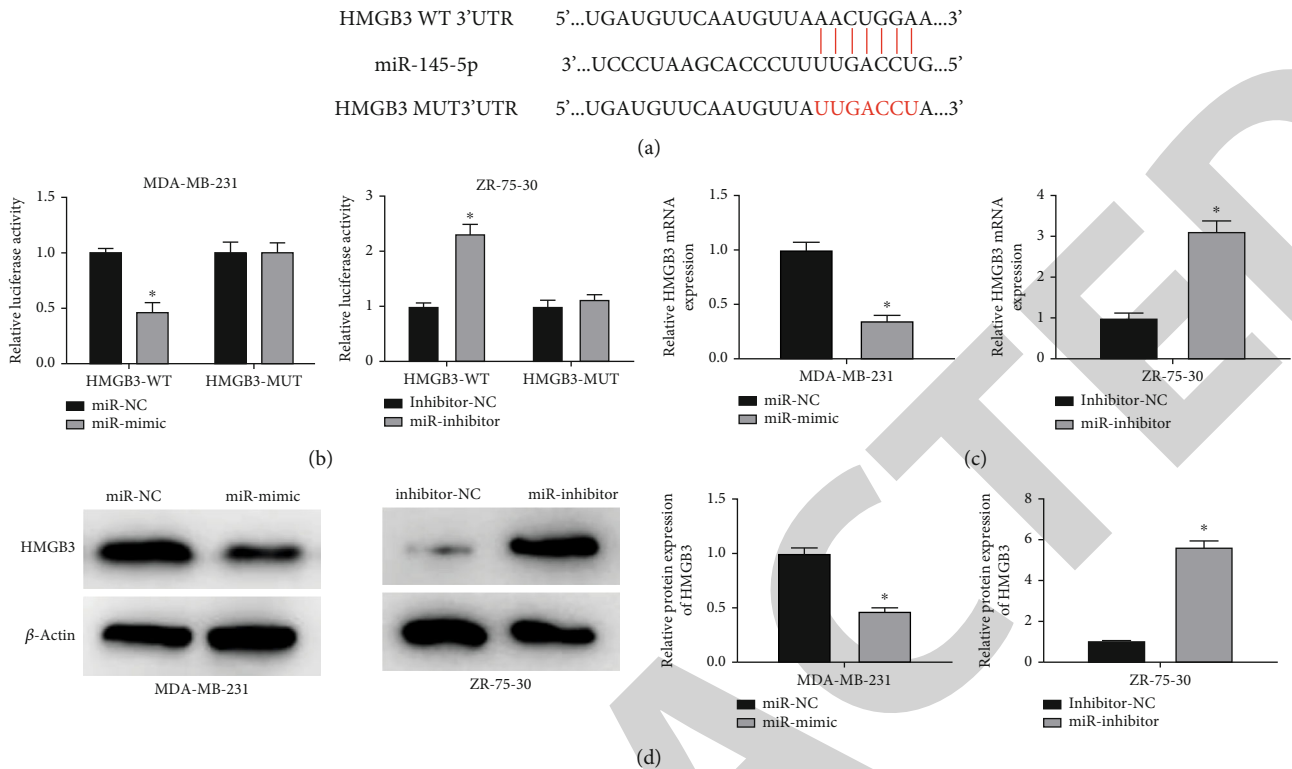


FIGURE 4: miR-145-5p targets HMGB3 in BC cells. (a) Predicted binding sites between miR-145-5p and HMGB3. (b) Luciferase activity of cells in the transfection groups. (c) Level of HMGB3 mRNA of cells in treatment groups. (d) Level of HMGB3 protein of cells in treatment groups. Results were shown as mean \pm SD. Data were analyzed via Student's *t*-test. Samples in each group were plated in triplicate. 3 repetitions were carried out for each experiment. * $P < 0.05$, compared with the corresponding NC group.

MDA-MB-231 cells (Figure 2(g)). In conclusion, overexpressed HMGB3 facilitated malignant phenotype progression of BC cells but the knockdown of HMGB3 had an opposite effect.

3.3. MiR-145-5p Is Predicted as an Upstream Regulatory miRNA of HMGB3 in BC Cells. To study upstream functional mechanism of HMGB3 in BC cells, EdgeR was applied for differential analysis and 74 differential miRNAs were acquired (55 upregulated miRNAs and 19 downregulated miRNAs) (Figure 3(a)). Next, differentially downregulated miRNAs were overlapped with target miRNAs predicted by databases to obtain miR-145-5p (Figure 3(b)). Notable negative correlation was represented between HMGB3 and miR-145-5p as well (Figure 3(c)). In comparison to normal breast tissue, BC tissue had a remarkably lower miR-145-5p level (Figure 3(d)). Then, miR-145-5p in 4 BC cell lines was also shown to have a significant downregulation by qRT-PCR (Figure 3(e)). In summary, miR-145-5p showed a low-level expression in BC cells, which may restrain BC cell progression.

3.4. miR-145-5p Targets HMGB3 in BC. By using bioinformatics, it was discovered that miR-145-5p and HMGB3 had a binding relationship (Figure 4(a)). Thereafter, their binding was proved via dual-luciferase analysis and it was elucidated that overexpressing miR-145-5p or inhibiting miR-145-5p restrained or stimulated luciferase activity of

cells with WT HMGB3 3'UTR, respectively, but for MUT HMGB3 3'UTR, there was no effect on it (Figure 4(b)). Next, the HMGB3 mRNA level in MDA-MB-231 and ZR-75-30 cell lines in the transfection groups was measured by qRT-PCR. The result demonstrated that HMGB3 mRNA level was notably decreased in miR-145-5p overexpressed cell lines, but it showed a marked increase in miR-145-5p inhibitor cells (Figure 4(c)). As revealed via Western blot, upregulating miR-145-5p remarkably reduced HMGB3 protein expression whereas inhibiting miR-145-5p noticeably increased the HMGB3 protein level (Figure 4(d)). All the above results presented that miR-145-5p targeted HMGB3 and suppressed its level in BC cells.

3.5. miR-145-5p Inhibits BC Cell Proliferation, Migration, Invasion, and EMT and Promotes Apoptosis via Negatively Regulating HMGB3. The MDA-MB-231 cell line with overexpressed HMGB3 or miR-145-5p alone and simultaneously overexpressed HMGB3 and miR-145-5p were established. The miR-145-5p inhibitor and si-HMGB3 were transfected to ZR-75-30 cells. qRT-PCR analysis demonstrated that HMGB3 expression was substantially lower in the miR-mimic + oe-NC group when with a contrast to that in the NC group. Compared with those of miR-145-5p mimic + oe-HMGB3 and miR-mimic + oe-NC cells, the HMGB3 mRNA level was elevated. But it was hindered in the miR-145-5p mimic + oe-HMGB3 group compared with that in the miR-NC + oe-HMGB3 group. Likewise, simultaneously

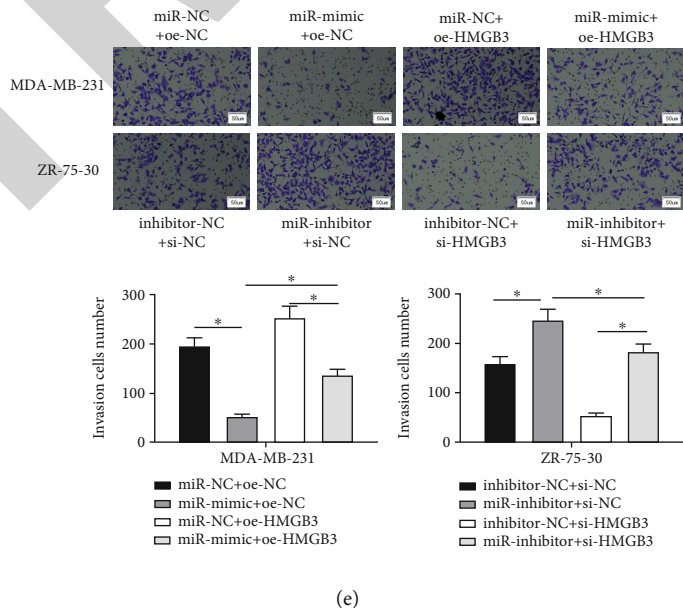
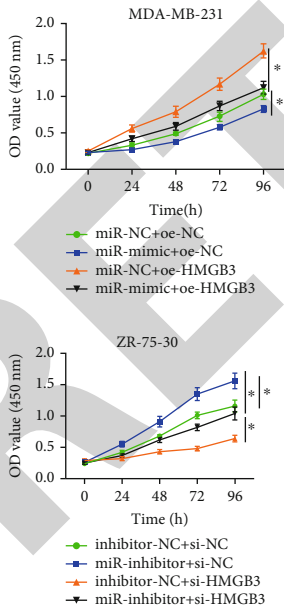
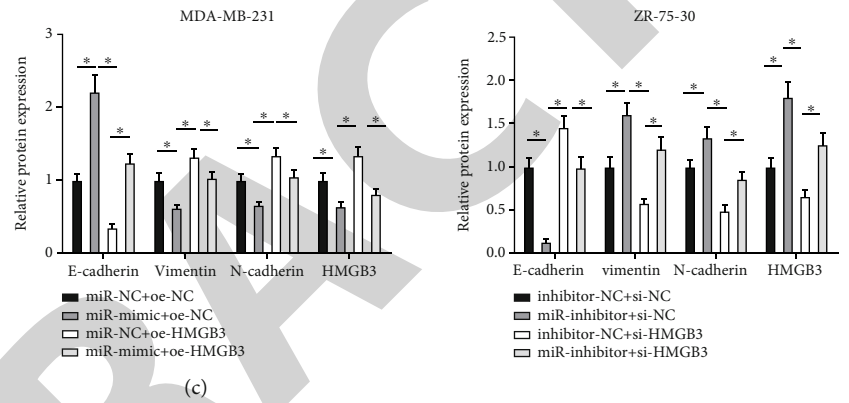
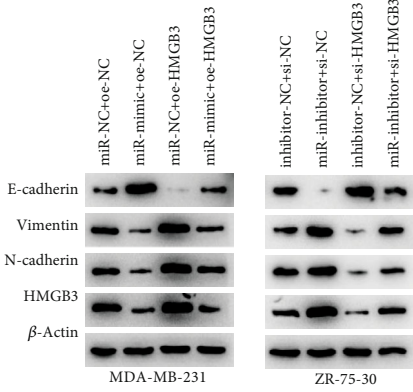
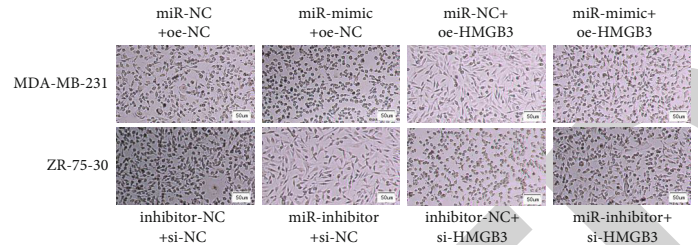
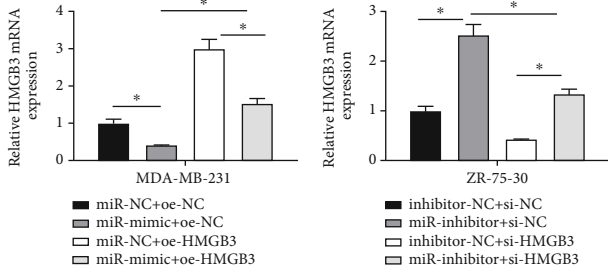


FIGURE 5: Continued.

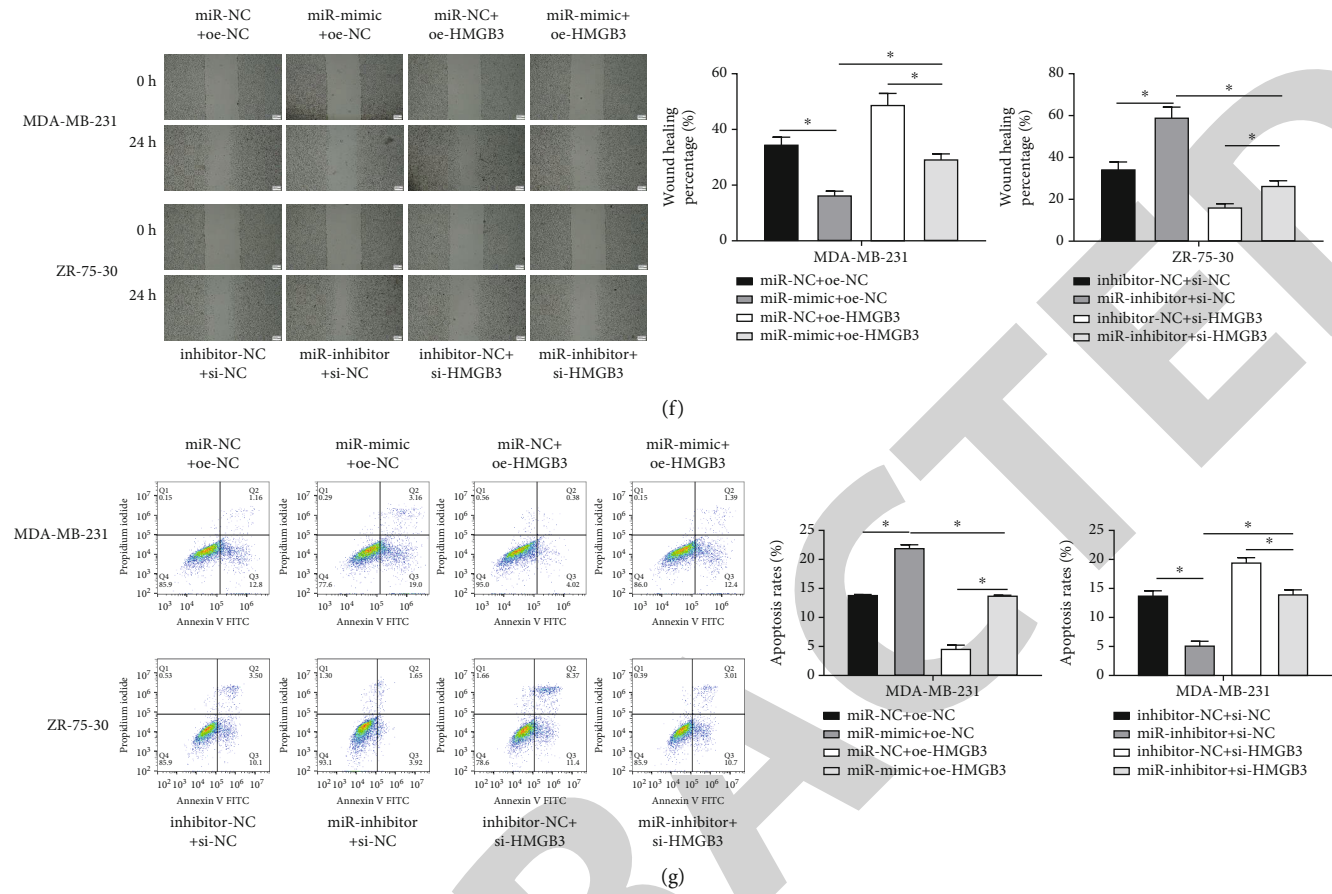


FIGURE 5: miR-145-5p suppresses BC cell phenotype progression through targeting HMGB3. (a) Level of HMGB3 mRNA of cells in the groups. (b) Changes of cell morphology as detected by an inverted microscope. (c) Protein levels of E-cadherin, vimentin, N-cadherin, and HMGB3 cells in treatment groups. (d) Cell proliferation in treatment groups. (e) Cell invasion in treatment groups (magnification: 100x; migrating cell numbers are the average value of 5 random fields). (f) Cell migration in transfection groups (magnification: 40x; wound healing rate = (0 h wound width - 24 h wound width)/0 h wound width). (g) Apoptotic level of cells in transfection groups. Data were analyzed via Student's *t*-test. Samples in each group were plated in triplicate. Each assay was performed 3 times. * $P < 0.05$.

inhibiting miR-145-5p and HMGB3 recovered the promoting effect of the miR-145-5p inhibitor on HMGB3 expression (Figure 5(a)). By observing cell morphology, overexpressed miR-145-5p elevated the proportion of paving stone-like cells (Figure 5(b)). As suggested by Western blot, overexpressing miR-145-5p alone substantially suppressed HMGB3, N-cadherin, and vimentin protein levels, while elevating E-cadherin protein expression, while overexpressing HMGB3 as well could recover these effects. Inhibition of miR-145-5p and HMGB3 simultaneously restored the promoting effect of the miR-145-5p inhibitor on EMT of BC cells (Figure 5(c)), which further denoted that overexpressing miR-145-5p counteracted the facilitating impact of HMGB3 on protein expression and the EMT process of cancer cells. Additionally, overexpressing miR-145-5p remarkably restrained the proliferative, invasive, and migratory properties of BC cells, while overexpressing HMGB3 simultaneously could offset such effect. Simultaneous inhibition of miR-145-5p and HMGB3 expression reversed the promoting effect of knockdown of miR-145-5p on malignant progression of BC cells (Figures 5(d)–5(f)). Additionally, it was also identified by flow cytometry that transfecting miR-

145-5p could noticeably stimulate BC cell apoptosis while simultaneously overexpressing miR-145-5p, and HMGB3 recovered such effect. Simultaneously inhibiting miR-145-5p and HMGB3 rescued the repressive effects of the miR-145-5p inhibitor on BC cell apoptosis (Figure 5(g)). These results concluded that miR-145-5p downregulated HMGB3 expression to suppress malignant progression of BC cells.

4. Discussion

In recent years, major progress had been made in researching BC in biology with the improved diagnostic technology and newly developed targets. This study was also made in researching BC and manifested that miR-145-5p downregulated HMGB3 expression to suppress malignant progression of BC cells.

Currently, HMGB3 protein has shown substantial effect on the development of BC [13, 14]. High-level HMGB3 is usually discovered in BC and it exerts a cancer-promotive role in cancer [22, 23]. Much literature so far verifies that HMGB3 may be a biomarker of several cancers. It has already been shown that HMGB3 had a

targeting relationship with miR-145 in non-small cell lung cancer [24]. In BC cells, HMGB3 is upregulated and its silencing impairs cell proliferation *in vitro* and *in vivo* [15]. It is worth noting that the study of Li et al. [25] manifested that enforced HMGB3 expression accelerates bladder cancer cell proliferation and pertains to poor survival outcomes, which is consistent with our study results. This study demonstrated the highly expressed HMGB3 in BC cells, and it could enhance proliferation, migration, invasion, and the EMT process of tumor cells, while inhibiting cell apoptosis.

This study predicted by bioinformatics analysis showed that miR-145-5p had a targeting relationship with HMGB3. Several studies have discovered that various cancers have remarkably low-level miR-145-5p expression which can inhibit the advancement of bladder cancer [26], non-small cell lung cancer [27], and gastric cancer cells [28]. Similarly, we found the significantly low-level miR-145-5p expression in BC cells and the dual-luciferase assay demonstrated that there was indeed a targeting relationship between HMGB3 and miR-145-5p. We also discovered that miR-145-5p was capable of negatively regulating the expression of HMGB3. Cell function experiments showed that overexpressed miR-145-5p could reverse the promoting effect of overexpression of HMGB3 on proliferation, invasion, and migration of tumor cells.

In summary, these findings demonstrated the high HMGB3 in BC, which was closely bound up to the poor prognoses of BC patients. miR-145-5p, by targeting HMGB3, could inhibit the proliferative, invasive, and migratory abilities and facilitated apoptosis of BC cells. This study reveals that the miR-145-5p/HMGB3 signaling axis is a new network that regulates the advancement of BC cells, which provides a prospective strategy for the diagnosis and treatment of BC.

Data Availability

The data used to support the findings of this study are included within the article.

Consent

All authors gave consent to submit the manuscript for publication.

Conflicts of Interest

The authors declare that they have no potential conflicts of interest.

Authors' Contributions

All authors contributed to data analysis and drafting and revising of the article, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work.

Supplementary Materials

Supplementary Figure: the relationship between HMGB3 expression and the prognosis of different subtypes of BC.

A–E Analysis of HMGB3 expression and the survival of basal-like-type, Her2E-type, LumA-type, LumB-type, and normal-like-type BC patients, respectively. Survival analysis was performed using the log-rank test. (*Supplementary Materials*)

References

- [1] C. E. DeSantis, J. Ma, A. Goding Sauer, L. A. Newman, and A. Jemal, "Breast cancer statistics, 2017, racial disparity in mortality by state," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 6, pp. 439–448, 2017.
- [2] F. Magnoni, V. Galimberti, G. Corso, M. Intra, V. Sacchini, and P. Veronesi, "Axillary surgery in breast cancer: an updated historical perspective," *Seminars in Oncology*, vol. 47, no. 6, pp. 341–352, 2020.
- [3] Y. Yuan, J. S. Lee, S. E. Yost et al., "A phase II clinical trial of pembrolizumab and enobosarm in patients with androgen receptor-positive metastatic triple-negative breast cancer," *The Oncologist*, vol. 26, no. 2, pp. 99–e217, 2021.
- [4] R. S. Finn, H. S. Rugo, K. A. Gelmon et al., "Long-term pooled safety analysis of palbociclib in combination with endocrine therapy for hormone receptor-positive/human epidermal growth factor receptor 2-negative advanced breast cancer: updated analysis with up to 5 years of follow-up," *The Oncologist*, vol. 26, no. 5, pp. e749–e755, 2021.
- [5] W. Zhao, M. Wang, M. Cai et al., "Transcriptional co-activators YAP/TAZ: potential therapeutic targets for metastatic breast cancer," *Biomedicine & Pharmacotherapy*, vol. 133, p. 110956, 2021.
- [6] T. Vaccari, M. Beltrame, S. Ferrari, and M. E. Bianchi, "Hmg4, a new member of the Hmg1/2 gene family," *Genomics*, vol. 49, no. 2, pp. 247–252, 1998.
- [7] R. Hock, T. Furusawa, T. Ueda, and M. Bustin, "HMG chromosomal proteins in development and disease," *Trends in Cell Biology*, vol. 17, no. 2, pp. 72–79, 2007.
- [8] M. J. Nemeth, A. P. Cline, S. M. Anderson, L. J. Garrett-Beal, and D. M. Bodine, "Hmg3 deficiency deregulates proliferation and differentiation of common lymphoid and myeloid progenitors," *Blood*, vol. 105, no. 2, pp. 627–634, 2005.
- [9] M. J. Nemeth, M. R. Kirby, and D. M. Bodine, "Hmg3 regulates the balance between hematopoietic stem cell self-renewal and differentiation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 37, pp. 13783–13788, 2006.
- [10] J. Fang, X. Ge, W. Xu et al., "Bioinformatics analysis of the prognosis and biological significance of HMGB1, HMGB2, and HMGB3 in gastric cancer," *Journal of Cellular Physiology*, vol. 235, no. 4, pp. 3438–3446, 2020.
- [11] J. Gao, Z. Zou, J. Gao et al., "Increased expression of HMGB3: a novel independent prognostic marker of worse outcome in patients with esophageal squamous cell carcinoma," *International Journal of Clinical and Experimental Pathology*, vol. 8, no. 1, pp. 345–352, 2015.
- [12] J. Wang, Z. Sheng, and Y. Cai, "Effects of microRNA-513b on cell proliferation, apoptosis, invasion, and migration by targeting HMGB3 through regulation of mTOR signaling pathway in non-small-cell lung cancer," *Journal of Cellular Physiology*, vol. 234, no. 7, pp. 10934–10941, 2019.
- [13] X. Li, Y. Wu, A. Liu, and X. Tang, "MiR-27b is epigenetically downregulated in tamoxifen resistant breast cancer cells due

Retraction

Retracted: Functions of Heparin Sodium Injection in the Prevention of Peripherally Inserted Central Catheter-Related Venous Thrombosis in NSCLC Patients during Postoperative Chemotherapy

Computational and Mathematical Methods in Medicine

Received 26 September 2023; Accepted 26 September 2023; Published 27 September 2023

Copyright © 2023 Computational and Mathematical Methods in Medicine. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] S. Li and H. Lu, "Functions of Heparin Sodium Injection in the Prevention of Peripherally Inserted Central Catheter-Related Venous Thrombosis in NSCLC Patients during Postoperative Chemotherapy," *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 1239058, 7 pages, 2022.

Research Article

Functions of Heparin Sodium Injection in the Prevention of Peripherally Inserted Central Catheter-Related Venous Thrombosis in NSCLC Patients during Postoperative Chemotherapy

Shanquan Li ¹ and Hong Lu ²

¹Venous Catheterization Nursing Clinic, Cancer Hospital Chinese Academy of Medical Sciences, Shenzhen Hospital, Shenzhen, Guangdong Province, China 518100

²Nursing Department, Cancer Hospital Chinese Academy of Medical Sciences, Shenzhen Hospital, Shenzhen, Guangdong Province, China 518100

Correspondence should be addressed to Hong Lu; hongluuu@163.com

Received 14 July 2022; Revised 21 September 2022; Accepted 27 September 2022; Published 17 October 2022

Academic Editor: Lin Lu

Copyright © 2022 Shanquan Li and Hong Lu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. This study intended to analyze hazardous factors of venous thrombosis by comparing the effect of different doses of heparin sodium injection on the incidence rate of peripherally inserted central catheter (PICC)-related venous thrombosis in non-small cell lung carcinoma (NSCLC) patients during postoperative chemotherapy. **Methods.** 425 NSCLC patients who received PICC catheterization in Cancer Hospital Chinese Academy of Medical Sciences, Shenzhen Hospital from July 2019 to July 2021 were collected. Based on their different pathological types, patients were given two different chemotherapy regimens: pemetrexed+cisplatin or paclitaxel+cisplatin. Patients were grouped according to the different doses of heparin sodium injection adopted. Control group ($n = 140$). Catheters were sealed with 10 mL saline only. Group I ($n = 142$). In addition to routine maintenance with normal saline, 2 mL of 10 IU/mL heparin sodium injection was sealed in the catheters under positive pressure every time after catheterization. Group II ($n = 143$). In addition to routine maintenance with normal saline, 5 mL of 10 IU/mL heparin sodium injection was sealed in the same manner as Group I. The baseline characteristics of the three groups of patients were compared by statistical means. Doppler ultrasonography was applied to check the venous thrombosis. The hazardous factors of venous thrombosis were analyzed through correlation analysis and binary logistic regression method. **Results.** The incidence rates of thrombosis in the control group, Group I, and Group II were 20.00%, 7.04%, and 2.09%, respectively, with statistically significant differences ($P < 0.01$). Additionally, through the collinear correlation analysis of baseline characteristics, a significant correlation between the dosage of heparin sodium injection and the incidence of thrombosis was observed ($P < 0.05$), but there were no significant differences between other baseline data and the incidence of thrombosis ($P > 0.05$). Binary logistic regression analysis revealed that postoperative use of heparin sodium injection (Group I: OR = 0.312; $P = 0.003$; Group II: OR = 0.082, $P < 0.001$) was a protective factor for preventing thrombosis. In addition, the thromboprophylaxis effect of Group II was better than that of Group I. No serious adverse reactions were found in safety analysis. **Conclusion.** Heparin sodium could significantly lower the incidence rate of PICC-related venous thrombosis in NSCLC patients during postoperative chemotherapy. Heparin sodium injection is safe enough to be promoted among PICC patients with a high risk of venous thrombosis.

1. Introduction

Primary bronchial lung cancer is one of the malignancies with high morbidity and mortality in China and the world [1]. With the boosting lung cancer incidence rate, more and more patients are developed with lung cancer, and most of them are diagnosed in the advanced stage with chemotherapy as the main treatment method [2]. Peripherally inserted central catheter (PICC) is a kind of catheterization that send the catheter tip to the bottom third of the superior vena cava, or the junction of the superior vena cava and right atrium, by puncturing the peripheral vein with the catheter. This method allows longer indwelling time, fewer puncture times, and less drug-caused irritation to veins. It provides a long-term effective pathway in veins for the patients who need tumor chemotherapy, long-term intravenous infusion, and inoculation of irritation drugs [3, 4]. Besides, due to easy indwelling and long indwelling time [5, 6], PICC is widely adopted in clinical practice, especially in patients with malignancies who are undergoing chemotherapy [7]. However, what comes next is to prevent catheter-related complications.

PICC-related complications mainly include venous thrombosis, phlebitis, catheter-related bloodstream infection (CRBSI), and catheter ectopy (including abnormal course and abnormal position of blocked catheter tip) [8–12]. Among them, PICC-related venous thromboembolism refers to the process that the blood clots are formed on the inner wall of the vessel where the catheter is located and the adherent wall of the catheter after the placement of PICC, due to factors such as direct damage to the vascular intima by puncturing or catheter and the state of patients themselves. PICC-related venous thrombosis is a common complication of PICC, which is mainly manifested as pain at the involved site, increased body surface temperature, superficial vein exposure, erythema, numbness in the extremities, and impaired neck and limb movements [13]. Multiple publications have exhibited that the incidence of symptomatic PICC-related venous thrombosis is 2%-75% [14, 15], while that of asymptomatic one is as high as 50% [16, 17]. PICC-related venous thrombotic events are also common in patients with lung cancer undergoing chemotherapy. Domestic researchers have reported that the incidence of PICC-related venous thrombosis in lung cancer patients undergoing chemotherapy is 5%-20%, seriously affecting the therapeutic effect [18]. Another work indicated that the incidence of PICC-related venous thrombosis ranged from 3%-30% in lung cancer patients undergoing chemotherapy [19]. Accordingly, preventing the development of PICC-related venous thrombosis and improving the quality of life of patients are urgent. Preventative anticoagulants are clinically applied as a key method to prevent PICC-associated thrombosis in cancer patients [20]. According to research reports, low molecular weight heparin (LMWH), unfractionated heparin (UFH), and warfarin are anticoagulants currently applied in clinical practice, among which, saline-diluted UFH or heparin sodium injection-sealed catheters under positive pressure can effectively prevent microthrombus formation in indwelling catheter [21].

Effective sealing can effectively prevent blockage, exudation, and catheter-related thrombosis, of which heparin sodium is one of the conventional sealing solutions [22]. Heparin sodium can bind to antithrombin III, enhance the inhibitory effect of antithrombin III on the activation of coagulation factors II, IX, X, XI, and XII, and repress functions of coagulation substances such as thrombin and fibrin, thus exerting anticoagulant effect. Recently, studies have revealed that heparin sodium injection can lower the incidence of venous thrombosis, which is safe and effective. However, a consensus has not been reached on the medication population, medication time, medication dosage, etc.

This study adopted statistical methods to compare the baseline characteristics and the incidence of venous thrombosis of patients treated with different doses of heparin sodium injection. At the same time, the collinearity relationship between the baseline characteristics was analyzed. Then a binary logistic regression analysis was performed to analyze the hazardous factors of thrombosis. In addition, the incidence of adverse reactions of heparin sodium injection was counted, thereby analyzing the safety of the treatment method.

2. Data and Methods

2.1. General Data. 425 NSCLC patients who received PICC catheterization in Cancer Hospital Chinese Academy of Medical Sciences, Shenzhen Hospital from July 2019 to July 2021 were collected. Based on their different pathological types, NSCLC patients were given two different chemotherapy regimens: pemetrexed+cisplatin or paclitaxel+cisplatin. Patients were grouped according to the different heparin sodium injection doses adopted. Control group: 140 patients aged from 28 to 84, including 96 males and 44 females. Their catheters were sealed with only 10 mL saline. Group I: 142 patients aged from 38 to 82, including 114 males and 28 females. In addition to routine maintenance with normal saline, 2 mL of 10 IU/mL heparin sodium injection was sealed in the catheters under positive pressure every time after catheterization. Group II: 143 patients aged from 30 to 84, including 101 males and 42 females. In addition to routine maintenance with normal saline, 5 mL of 10 IU/mL heparin flush injection was sealed in the same manner as Group I. The baseline characteristics of the three groups of patients were compared by statistical methods and were not significantly different ($P > 0.05$) (Table 1). The three groups underwent Doppler ultrasonography on the 7th day after catheterization to check whether venous thrombosis was formed.

2.2. Inclusion Criteria. NSCLC patients who need PICC catheterization for chemotherapy. Their blood biochemistry and coagulation tests should be normal before catheterization. The patient's physique was fair, and The Eastern Cooperative Oncology Group (ECOG) scores were between 0 and 2. All patients had cognitive ability and could respond actively to the treatment. Informed consent was signed.

TABLE 1: Baseline characteristics of patients.

Baseline characteristics	Control group ($n = 140$)	Group I ($n = 142$)	Group II ($n = 143$)	P value
Age (years old)	61.88 ± 10.37	61.30 ± 9.10	59.94 ± 10.25	0.243
Sex				0.379
Male	96	114	101	
Female	44	28	42	
Weight (kg)	61.85 ± 7.65	60.63 ± 8.10	60.81 ± 7.68	0.370
Smoking history	70	66	67	0.320
ECOG score	1.06 ± 0.78	1.04 ± 0.78	1.06 ± 0.77	0.973
Pathology type				0.238
Squamous carcinoma	69	68	77	
Adenocarcinoma	71	74	66	
Operation type				0.498
Lung resection	72	72	73	
Lobectomy	68	70	70	

2.3. Exclusion Criteria. Patients with preoperative coagulation disorder, heparin allergy, and active bleeding; patients who took anticoagulant drugs before surgery or had a history of thrombosis; and patients with concomitant heart failure were excluded.

2.4. Treatments. Three groups were maintained with normal saline. In the control group, 10 mL normal saline was given each time to seal the catheter 1 to 7 d after catheterization. In Group I, 2 mL of heparin sodium injection (Brand name: Heparin Sodium Injection for Lock Flush, Huabicheng; Specification: 5 mL, 50 units) was sealed in the catheters under positive pressure 1 to 7 d after catheterization. While in Group II, 5 mL of heparin sodium injection was sealed in the same manner as Group I.

2.5. Indicators for Observation. All patients included in the study underwent color Doppler ultrasound examination within 6 days, $2w \pm 6d$, $5w \pm 6d$, $8w \pm 6d$ since PICC catheterization. Color Doppler ultrasound diagnostic apparatus (German PHILIPS ie3.3, L11-3 probe, 12 MHz) was utilized, to observe whether there were substances attached to the vein that gave out solid mass echo and venous thrombosis at the catheterization site after the catheterization. Thereby the incidence of thrombosis in each group was compared. The incidence of thrombosis refers to the proportion of patients with catheter-related thrombosis in all PICC catheterized patients.

Thrombosis rate = the number of PICC-related thrombosis within a specified time/the total number of patients with PICC catheterization within the specified time $\times 100\%$ [23]. It was observed whether there were adverse reactions, hemorrhage or partial bleeding, PICC infections, and other local adverse reactions [24].

2.6. Statistical Analysis. SPSS 26.0 software was employed for statistical analysis. Measurement data were analyzed by one-way analysis of variance and examined by LSD- t test. Enumeration data underwent chi-square test among

groups. $P < 0.05$ (two-sided) implied a statistical significance. The correlation coefficient matrix was obtained by collinearity analysis of the baseline indicators. The relationship between the dosage and venous thrombosis risk was obtained through binary logistic regression.

3. Results

3.1. Clinical Efficacy. Compared with the control group using normal saline (Table 2), heparin sodium injection significantly reduced the incidence of PICC-related venous thrombosis ($P < 0.01$).

3.2. Diagnosis of Collinearity Correlation of Baseline Characteristics. Correlation analysis was performed on patient's group, age, sex, weight, smoking history, ECOG score, pathological type, and operation types (Table 3). The results indicated that different doses of heparin sodium injection were significantly associated with the incidence of venous thrombosis ($P < 0.05$).

3.3. Multivariate Binary Logistic Regression. The risk factors or protective factors related to thrombosis were analyzed by multivariate binary logistic regression (Table 4). The results indicated that heparin sodium injection at different doses was a protective factor against thrombosis. Additionally, Group II had the best efficacy (OR = 0.082, 95% CI 0.024-0.277, $P < 0.001$).

3.4. Safety. Serious adverse reactions such as deep vein thrombosis (DVT) disease, purpura, skin or systemic allergies, and severe thrombocytopenia were not observed in the three groups of patients.

4. Discussion

In recent years, the age of lung cancer patients in China tends to be younger. More than 1.6 million people are diagnosed with lung cancer every year. Venous thromboembolism (VTE) is a common lung cancer complication, mainly

TABLE 2: Comparison of venous thrombosis incidence among groups.

	Control group	Group I	Group II	P value
Incidence rates of venous thrombosis	20.00%	7.04%	2.09%	<0.001

TABLE 3: Correlation between incidence of venous thrombosis and baseline characteristics.

Baseline characteristics	Venous thrombosis (n = 42)	No venous thrombosis (n = 383)	P value
Group			<0.01
Control group	28	112	
Group I	11	131	
Group II	3	140	
Age			0.373
≤63	20	210	
>63	22	173	
Sex			0.231
Male	34	277	
Female	8	106	
Weight (kg)			0.356
≤62 kg	20		
>62 kg	22	172	
Smoking history			0.528
Yes	22	181	
No	20	202	
ECOG score			0.752
0	10	106	
1	19	151	
2	13	126	
Pathological types			0.354
Squamous carcinoma	24	190	
Adenocarcinoma	18	193	
Operation type			0.613
Lung resection	23	194	
Lobectomy	19	189	

TABLE 4: Results of multivariate binary logistic regression analysis.

Variate	OR	95% CI		P
		Lower	Upper	
Group I (group I vs control group)	0.312	0.146	0.665	0.003
Group II (group II vs group I)	0.082	0.024	0.277	<0.001
Group I vs group II	3.824	1.039	14.078	0.044
Age (>63 vs ≤63)	1.335	0.680	2.619	0.401
Sex (female vs male)	0.550	0.234	1.291	0.170
Weight (>62 vs ≤62)	1.303	0.669	2.538	0.436
Smoking history (yes vs no)	1.003	0.505	1.992	0.993
Pathological types (adenocarcinoma vs squamous carcinoma)	0.707	0.356	1.403	0.321
ECOGS = 1 (1 vs 0)	0.739	0.367	1.487	0.396
ECOGS = 2 (2 vs 0)	0.601	0.277	1.302	0.197
Operation type (lobectomy vs lung resection)	0.842	0.428	1.655	0.618

including pulmonary embolism (PE) and deep vein thrombosis (DVT) [25]. Connolly et al. [26] indicated that the incidence of VTE in outpatient lung cancer patients reaches 14%. Recent studies have presented that the pathophysiological mechanism of VTE in lung cancer patients is mainly due to the direct activation of coagulation by malignant cells through producing tissue factor (TF), cancer procoagulant (CP), cytokines, and inflammatory factors [27]. By contacting with endothelial cells, platelets, and leukocytes, tumor cells can activate local coagulation, thus promoting platelet activation and aggregation, and stimulating leukocytes to release cytokines [25]. The main risk factors for VTE in patients with lung cancer may be different types and stages of lung cancer, patient factors (history of VTE, elevated platelets, complications (infection, heart failure, etc.)), and tumor treatment measures (radiotherapy, chemotherapy, surgery, and PICC catheterization) [25].

PICC is commonly used in patients who have malignant tumors and are critically ill because it is beneficial to reduce the risk of drug infiltration, relieving vascular inflammation, avoiding pain caused by repeated punctures, and improving the quality of life of patients. It is suitable for intravenous treatment of home-care patients, thus becoming more popular with medical staff and patients [28, 29]. But recent studies revealed that PICC markedly increased the risks of venous thrombosis [30, 31]. The causes of PICC-related venous thrombosis may be as follows: (1) the diameter of the PICC affects the central flow in the lumen, thereby increasing the risk of turbulent flow and venous thrombosis; (2) the stiffness of the PICC and the direct venous injury and inflammation caused by the insertion, resulting in thrombosis; (3) the introduction of PICC prompts the human body to form a surrounding biofilm to isolate it, coupled with low flow and venous blood stasis, which is conducive to the thrombosis [10]. Hence, both the selection of PICC materials and antithrombotic therapy after surgery are important.

In clinical nursing, heparin is a common drug for effective prevention and treatment of thrombosis [32]. Earlier studies have shown that intravenous heparin is safe and effective for the prevention of recurrent VTE [33]. Heparin is administered by continuous intravenous infusion and subcutaneous injection [34]. Previous randomized trials have shown that there is a correlation between heparin dose and efficacy and safety, although the intensity and duration of anticoagulant effect of heparin is nonlinear with heparin dose [34]. In addition, an increase in heparin dose also increases the risk of bleeding. Therefore, heparin anticoagulation response is commonly monitored clinically by activated partial thromboplastin time (APTT) to adjust continuous intravenous dose [34].

Nevertheless, the clinical dosage of heparin sodium injection is not yet clear. This study confirmed that the 5 mL of 10 IU/mL heparin sodium injection could significantly reduce the incidence of venous thrombosis in NSCLC patients. Additionally, the results of binary logistic regression analysis showed that heparin sodium injection

was a protective factor against thrombosis. A study illustrated that old age, obesity, and lung resection are independent predictors of VTE [35]. Through OR analysis, our results showed that age > 63 years (OR = 1.335), weight > 62 kg (OR = 1.303), and smoking history (OR = 1.003) may be risk factors for PICC-related thrombosis, but there was no statistical difference.

5. Conclusion

In summary, the use of heparin sodium injection can improve the coagulation function of patients, facilitate normal blood flow, significantly reduce the occurrence of venous thrombosis, facilitate the surgical treatment of patients, and the survival benefits. This retrospective study still has many shortcomings, including the limited sample size and diversity. In addition, no blood samples were collected for analysis of coagulation parameters and hematology data in our study. Our conclusion calls for further verification by large-sample clinical trials and blood samples in the future.

Data Availability

The data used to support the findings of this study are included within the article. The data and materials in the current study are available from the corresponding author on reasonable request.

Ethical Approval

The study was approved by the ethics committee of Cancer Hospital Chinese Academy of Medical Sciences, Shenzhen Hospital. The methods were carried out in accordance with the approved guidelines. Written informed consent was obtained prior to the study.

Consent

All authors consent to submit the manuscript for publication.

Conflicts of Interest

The authors declared that they have no potential conflicts of interest.

Authors' Contributions

All authors contributed to data analysis, drafting, and revising the article, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work.

References

- [1] H. Sung, J. Ferlay, R. L. Siegel et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

- [2] I. Garcia-Olive, E. Monsó, F. Andreo et al., "Sensitivity of linear endobronchial ultrasonography and guided transbronchial needle aspiration for the identification of nodal metastasis in lung cancer staging," *Ultrasound in Medicine & Biology*, vol. 35, no. 8, pp. 1271–1277, 2009.
- [3] S. Bertoglio, B. Faccini, L. Lalli, F. Cafiero, and P. Bruzzi, "Peripherally inserted central catheters (PICCs) in cancer patients under chemotherapy: a prospective study on the incidence of complications and overall failures," *Journal of Surgical Oncology*, vol. 113, no. 6, pp. 708–714, 2016.
- [4] A. Simon, R. A. Ammann, G. Wiszniewsky, U. Bode, G. Fleischhack, and M. M. Besuden, "Taurolidine-citrate lock solution (TauroLock) significantly reduces CVAD-associated grampositive infections in pediatric cancer patients," *BMC Infectious Diseases*, vol. 8, no. 1, 2008.
- [5] E. Johansson, F. Hammarskjöld, D. Lundberg, and M. H. Arnlin, "Advantages and disadvantages of peripherally inserted central venous catheters (PICC) compared to other central venous lines: a systematic review of the literature," *Acta Oncologica*, vol. 52, no. 5, pp. 886–892, 2013.
- [6] X. Chen, L. Zhou, Y. Tan, and Z. Tao, "Selection of PICC catheter location in neonates via evidence-based ACE star model," *Zhong Nan Da Xue Xue Bao. Yi Xue Ban*, vol. 45, pp. 1082–1088, 2020.
- [7] Z. Mei, "Application of ultrasound-guided PICC catheterization combined with modified Sedinger technique in cancer chemotherapy," *Journal of Clinical Medicine in Practice*, vol. 20, pp. 188–189, 2016.
- [8] S. Duwadi, Q. Zhao, and B. S. Budal, "Peripherally inserted central catheters in critically ill patients - complications and its prevention: a review," *International Journal of Nursing Sciences*, vol. 6, no. 1, pp. 99–105, 2019.
- [9] V. Chopra, S. Anand, S. L. Krein, C. Chenoweth, and S. Saint, "Bloodstream infection, venous thrombosis, and peripherally inserted central catheters: reappraising the evidence," *The American Journal of Medicine*, vol. 125, no. 8, pp. 733–741, 2012.
- [10] V. Zochios, I. Umar, N. Simpson, and N. Jones, "Peripherally inserted central catheter (PICC)-related thrombosis in critically ill patients," *The Journal of Vascular Access*, vol. 15, no. 5, pp. 329–337, 2014.
- [11] B. Westergaard, V. Classen, and S. Walther-Larsen, "Peripherally inserted central catheters in infants and children – indications, techniques, complications and clinical recommendations," *Acta Anaesthesiologica Scandinavica*, vol. 57, no. 3, pp. 278–287, 2013.
- [12] O. Al-Asadi, M. Almusarhed, and H. Eldeeb, "Predictive risk factors of venous thromboembolism (VTE) associated with peripherally inserted central catheters (PICC) in ambulant solid cancer patients: retrospective single centre cohort study," *Thrombosis Journal*, vol. 17, no. 1, p. 2, 2019.
- [13] Professional Committee of Intravenous Infusion Therapy, C N A, "Practice guidelines for clinical nursing of common complications of venous catheter," *Chinese Journal of Modern Nursing*, vol. 28, pp. 2381–2395, 2022.
- [14] A. Aw, M. Carrier, J. Koczerginski, S. McDiarmid, and J. Tay, "Incidence and predictive factors of symptomatic thrombosis related to peripherally inserted central catheters in chemotherapy patients," *Thrombosis Research*, vol. 130, no. 3, pp. 323–326, 2012.
- [15] M. Kanin and G. Young, "Incidence of thrombosis in children with tunneled central venous access devices versus peripherally inserted central catheters (PICCs)," *Thrombosis Research*, vol. 132, no. 5, pp. 527–530, 2013.
- [16] M. Hammes, A. Desai, S. Pasupneti et al., "Central venous catheters: incidence and predictive factors of venous thrombosis," *Clinical Nephrology*, vol. 84, no. 7, pp. 21–28, 2015.
- [17] J. J. Menendez, C. Verdú, B. Calderón et al., "Incidence and risk factors of superficial and deep vein thrombosis associated with peripherally inserted central catheters in children," *Journal of Thrombosis and Haemostasis*, vol. 14, no. 11, pp. 2158–2168, 2016.
- [18] F. Huaying and F. Yi, "Analysis of risk factors affecting PICC catheter-related venous thrombosis in patients with lung cancer undergoing chemotherapy," *Clinical Misdiagnosis & Mistherapy*, vol. 32, no. 9, pp. 54–58, 2019.
- [19] G. C. Connolly, L. Menapace, S. Safadjou, C. W. Francis, and A. A. Khorana, "Prevalence and clinical significance of incidental and clinically suspected venous thromboembolism in lung cancer patients," *Clinical Lung Cancer*, vol. 14, no. 6, pp. 713–718, 2013.
- [20] J. Liping and Z. Zheng, "Research progress on risk factors and nursing intervention of PICC catheter-related venous thrombosis in cancer patients," *Nursing Practice and Research*, vol. 17, pp. 54–57, 2020.
- [21] D. Ling, "Clinical research progress of PICC sealing technique," *Smart Healthcare*, vol. 6, no. 29, pp. 31–33, 2020.
- [22] C. Ying, W. Xuehua, W. Caijuan, and W. Meihua, "Effect of different sealing solutions on right subclavian deep venous catheter sealing in patients with cervical cancer," *Journal of Qilu Nursing*, vol. 27, pp. 170–172, 2021.
- [23] L. Luo, G. Wang, and Y. Qin, "A prospective study on the incidence and time of PICC-related upper extremity venous thrombosis in tumor patients," *Chinese Journal of Practical Nursing*, vol. 32, pp. 576–580, 2016.
- [24] X. Bai, J. Song, X. Lin, C. B. Fan, J. H. Chang, and L. L. Han, "Observation on the effect of rivaroxaban in preventing breast cancer related thrombosis through peripherally inserted central venous catheter," *People's Military Surgeon*, vol. 63, no. 4, 2020.
- [25] H. Du and J. Chen, "Occurrence of venous thromboembolism in patients with lung cancer and its anticoagulant therapy," *Zhongguo Fei Ai Za Zhi*, vol. 21, no. 10, pp. 784–789, 2018.
- [26] G. C. Connolly, M. Dalal, J. Lin, and A. A. Khorana, "Incidence and predictors of venous thromboembolism (VTE) among ambulatory patients with lung cancer," *Lung Cancer*, vol. 78, no. 3, pp. 253–258, 2012.
- [27] S. R. Deitcher, "Cancer and thrombosis: mechanisms and treatment," *Journal of Thrombosis and Thrombolysis*, vol. 16, no. 1/2, pp. 21–31, 2003.
- [28] F. K. Y. Santos, R. L. G. Flumignan, L. L. Areias et al., "Peripherally inserted central catheter versus central venous catheter for intravenous access: a protocol for systematic review and meta-analysis," *Medicine (Baltimore)*, vol. 99, article e20352, 2020.
- [29] G. D. Wang, H. Z. Wang, Y. F. Shen et al., "The influence of venous characteristics on peripherally inserted central catheter-related symptomatic venous thrombosis in cancer patients," *Cancer Management and Research*, vol. 12, pp. 11909–11920, 2020.
- [30] V. Chopra, S. Anand, A. Hickner et al., "Risk of venous thromboembolism associated with peripherally inserted central catheters: a systematic review and meta-analysis," *The Lancet*, vol. 382, no. 9889, pp. 311–325, 2013.

Research Article

The Clinical Value of Blood miR-654-5p, miR-126, miR-10b, and miR-144 in the Diagnosis of Colorectal Cancer

Gang Du ^{1,2}, Chongren Ren ^{1,2}, Ju Wang ^{1,2} and Jun Ma ^{1,2}

¹General Surgery Department, Shanxi Bethune Hospital/General Surgery Department, Third Hospital of Shanxi Medical University, Taiyuan, Shanxi 030032, China

²Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430030, China

Correspondence should be addressed to Jun Ma; majun430030@163.com

Received 12 May 2022; Accepted 10 August 2022; Published 13 October 2022

Academic Editor: Lin Lu

Copyright © 2022 Gang Du et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Colorectal cancer (CRC) is the third cause of cancer-related death and the fourth most frequently diagnosed cancer across the globe. The objective of this study is to obtain novel and effective diagnostic markers to enrich CRC diagnosis methods. Herein, exosomal miRNA expression data of CRC and normal blood were subjected to XGBoost algorithm, and 5 miRNAs related to CRC diagnosis were primarily confirmed. Then multilayer perceptron (MLP) classifiers were constructed based on different subsets. Via integrated feature selection (IFS), we noticed that the MLP classifier constructed by the first four miRNAs (miR-654-5p, miR-126, miR-10b, and miR-144) had the highest Matthews correlation coefficient (MCC). Subsequently, principal component analysis (PCA) for dimensionality reduction was performed on samples based on the miR-654-5p, miR-126, miR-10b, and miR-144 expression data. The signature based on these four feature miRNAs, as the analysis indicated, could effectively distinguish CRC samples from normal samples. Further, we extracted the exosomes from clinical blood samples and applied qRT-PCR analysis, which revealed that the expression of these four feature miRNAs was in the trend of that in the test set. Collectively, these four feature miRNAs might be tumor biomarkers in the serum, and our study offers innovative thinking on early-stage CRC diagnosis.

1. Introduction

As the third cause of cancer-related death and the fourth most frequently diagnosed cancer [1], colorectal cancer (CRC) presents a growing morbidity and death rate, making it a public health burden [2]. According to population and disease statistics, nearly 2.2 million new cases would be developed by 2030 [3]. CRC is a genotypically and phenotypically heterogeneous disease characterized by different molecular characteristics [4]. Accurate early diagnosis enables CRC patients to receive timely and precise treatment, thereby reducing CRC mortality. Although colonoscopy screening is the gold standard for CRC screening, its participation rate in population screening programs is still poor due to the invasive nature of the test and the need for adequate bowel preparation [5–8]. In addition, some studies have implied that carcinoembryonic antigen and calprotectin can be used as diagnostic markers for CRC, but their

specificity and sensitivity are low, and they cannot be effectively applied to the early diagnosis of clinical CRC at present [9, 10]. Hence, it is necessary to develop effective biomarkers for CRC to improve the early diagnosis rate for CRC and offer effective biomarkers for CRC treatment.

Recently, exosome biomarkers containing multiple RNA and proteins have become the focus of research in cancer diagnosis and treatment [11]. Exosomes are tiny goblet vesicles with 30–140 nm in diameter that are secreted by cells including immune cells, neural cells, stem cells, and tumor cells [12–14]. Increasing research manifested that exosomes relate to tumorigenesis. Tumor-derived exosomes are involved in the exchange of genetic information between tumor cells and basal cells, thereby regulating angiogenesis and promoting tumor growth and invasion [15]. Currently, useful biomarkers have been identified from exosomes for the application in CRC diagnosis. It has been demonstrated that in blood exosomes, miR-125a-3p and miR-638 are

TABLE 1: Primers of feature miRNAs for qRT-PCR.

Gene ID	Forward sequence (5'-3')	Reverse sequence (5'-3')
miR-654-5p	GGGTGGTGGGCCGAGAAC	CTCAACTGGTGTTCGTGGA
miR-126	UCGUACCGUGAGUAAUAAUGCG	CAUUAUUACUUUUGGUACGCG
miR-10b	TACCCTGTAGAACCGAATTTGTG	CAGTGCGTGTTCGTGGAGT
miR-144	TACAGTATAGATGAT	GTGCAGGGTCCGAGG
U6	ATTGGAACGATACAGAGAAGATT	GGAACGCTTCACGAATTTG

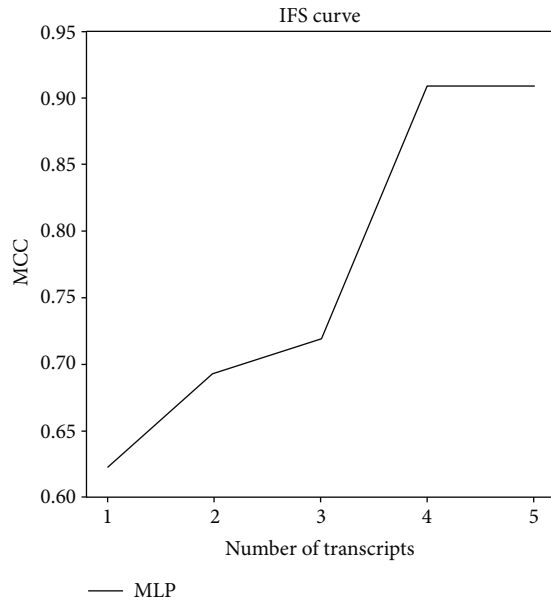


FIGURE 1: IFS curves of the five feature miRNAs based on MLP classifiers. Abscissa: the number of transcripts; ordinate: the MCC value.

helpful for early diagnosis of CRC in clinical practice [16, 17]. These all demonstrated the importance of exosomal miRNAs in screening early-stage CRC. Therefore, we further identified potentially effective exosomal miRNAs that may work for CRC diagnosis, so as their regulatory networks, which are beneficial for comprehensively understanding the molecular mechanisms underlying CRC development.

The rapid development of biotechnology in the age of big data stimulated the application of bioinformatics in medical research; bioinformatics technology based on high-throughput sequencing data is an effective and promising analytical tool for analyzing and identifying biomarkers for cancer diagnosis [18, 19]. Machine learning is a new artificial intelligence technique that has been gradually applied to medical research in recent years. Lian et al. [20] trained medulloblastoma stemness index based on a machine learning method of one-class logistic regression to obtain gene expression-based stemness index and methylation-based stemness index and further identified their corresponding potential drugs, which provides new ideas for the survival of medulloblastoma patients or targeting stem cells. Koppad et al. [21] screened diagnostic candidate genes for CRC based on six methods of machine learning classification including Adaboost, ExtraTrees, logistic regression, Naive Bayes classifier, random forest, and XGBoost. Thus, there is potential for

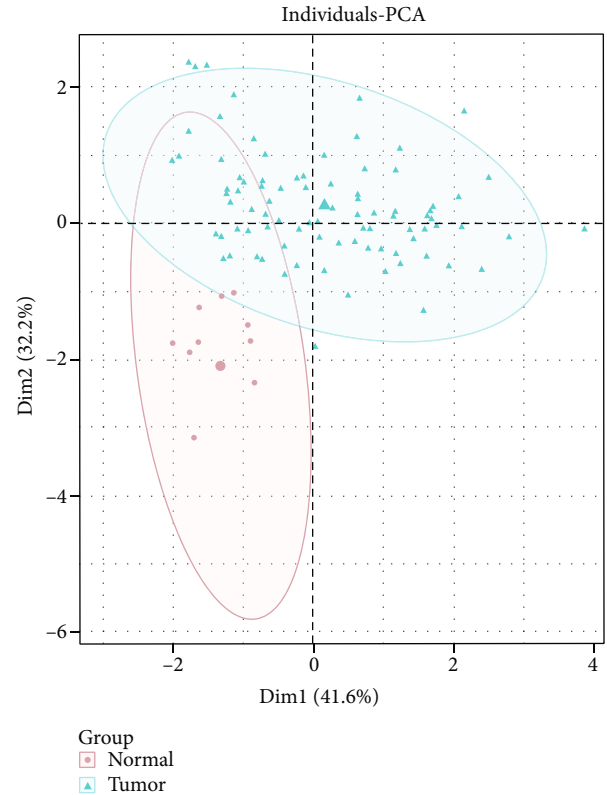


FIGURE 2: The PCA analysis for CRC and normal samples based on four feature miRNAs.

wider application of novel bioinformatics methods to identify novel diagnostic biomarkers based on public databases.

In this study, by analyzing the miRNA expression data of CRC patients and normal people in the Gene Expression Omnibus (GEO) database, we preliminarily screened miRNAs with potential diagnostic value based on XGBoost and established a multilayer perceptron (MLP) classifier to determine the optimal miRNA combination by taking integrated feature selection (IFS). Thereafter, the clinical value of diagnostic markers in CRC was dissected by testing their levels in the blood exosomes of clinical patients with CRC. To conclude, our study provided potential biomarkers which are supposed to be effective to CRC clinical diagnosis.

2. Materials and Methods

2.1. Data Source and Preprocessing. Exosomal miRNA data of CRC patients and normal people were downloaded as GSE39833 (tumor: 88 and normal: 11) from Gene

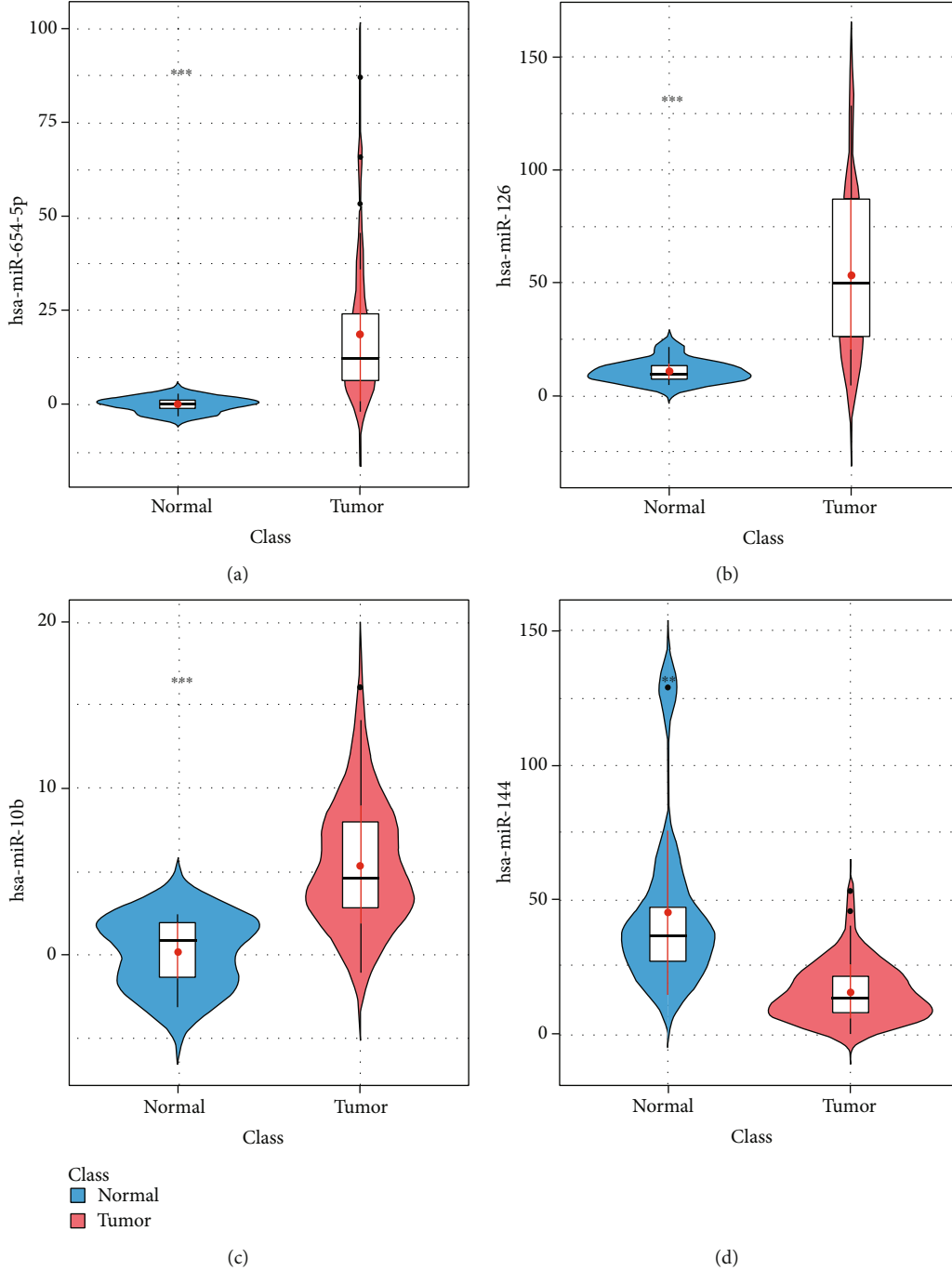


FIGURE 3: Violin plots of four feature miRNAs in the optimal MLP classifier. The violin plots of differentially expressed (a) miR-654-5p, (b) miR-126, (c) miR-10b, and (d) miR-144 in CRC and normal samples $***p < 0.001$.

Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>), annotated by the platform of Agilent-021827 Human miRNA Microarray G4470C GPL14767. Differential analysis was performed by R package “limma” [22] on the standardized miRNA expression data ($|\log\text{FC}| > 1.5$, $\text{adjPvalue} < 0.05$).

2.2. XGBoost Feature Selection. XGBoost is a tree boosting scalable machine learning system, which generates a single strong learner by combining multiple weak learners.

XGBoost estimates the value of the loss function through a second-order Taylor series and further reduces the likelihood of overfitting by applying regularization [23]. The objective function of XGBoost is a gradient advancing decision tree approach defined as

$$\mathcal{L}(\phi) = \sum_{i=1}^n \text{loss}(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k). \quad (1)$$

Loss means training loss, $\Omega(f)$ represents the complexity of trees, and k stands for the amount of trees. The model can be optimized by minimizing the objective function. Hence, we adopted the addition training method to calculate the training loss and rapidly optimized the prediction of the n^{th} round of addition training by taking the Taylor expansion method. The optimal complexity of the tree was determined via the greedy algorithm.

In order to find miRNAs that could distinguish CRC from normal samples in GSE39833, we utilized XGBoost to rank the importance of feature miRNAs. Five characteristic miRNAs associated with CRC diagnosis were filtered for subsequent analysis. Then, based on SMOTE method, we applied python package “imblearn” and Bayesian optimization to resample the training set to reduce the effect caused by data disequilibrium.

2.3. Construction of the MLP Classifier. To construct a diagnostic classifier that was more precise, we constructed MLP classifiers of different subsets based on these five characteristic miRNAs by python package “sklearn” [24] after XGBoost feature selection. For the MLP classifier, hidden layers were set as 2, and all possible combinations were scanned in the first layer (the number of nodes from 1 to 5) and in the second layer (the number of nodes from 1 to 5) by sklearn.neural_network. Other parameters included (1) solver = “adam”, (2) alpha = 0.001, (3) random_state = 1, and (4) max_iter = 1000.

2.4. Screen of Optimal Feature Genes. The MCC of the above classifiers was obtained using IFS. MCC is the correlation coefficient of binary classification between the observation and prediction, with its value between -1 and +1. +1 stands for a perfect prediction, while -1 for a total inconsistency between observation and prediction. The MCC value is a single score that is the most informative for the prediction quality of binary classifiers built in a confusion matrix environment [25]. The IFS curves were plotted, with abscissa for MLP classifiers based on different subsets and ordinate for MCC of subsets. The classifier with the highest MCC was selected as the optimal classifier for CRC diagnosis.

2.5. Principal Component Analysis (PCA). PCA is a dimensionality reduction algorithm that is most widely adopted. Its main idea is to map the n-dimensional data in space onto the k-dimension, a novel orthogonal feature that is the principal component [26]. We performed PCA by the R package “FactoMineR” [27] based on the characteristic miRNA expression data in the optimal MLP classifier to explore the sample discriminatory capability of this classifier (<https://www.rdocumentation.org/packages/FactoMineR/versions/2.4>).

2.6. Clinical Collection of Blood Sample. Between 10-2018 and 10-2021, 100 patients with CRC and 120 healthy participants were recruited from Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Third Hospital of Shanxi Medical University in Taiyuan city, Shanxi province, with their clinical information and serum samples collected (Supplementary Table 1). None of the CRC patients received any treatment, while their cancer stages were judged on the basis of the American

TABLE 2: Clinical information summary of 220 subjects.

Variables	External validation cohort	
	CRC ($n = 100$)	Normal ($n = 120$)
Gender		
Male	49 (49%)	51 (42.5%)
Female	51 (51%)	69 (57.5%)
Age		
<60	43 (43%)	47 (39.2%)
≥60	57 (57%)	73 (60.8%)
Tumor location		
Colon	44 (44%)	
Rectum	56 (56%)	
Stage		
1	24 (24%)	
2	17 (17%)	
3	30 (30%)	
4	29 (29%)	

Joint Committee on Cancer (AJCC) Cancer Staging Manual (7th Edition) [28]. Peripheral blood (5 ml) from all participants was collected in 5 ml clotting tubes (Greiner Bio-One, Austria). Serum was separated by centrifugation and stored at -80°C for subsequent miRNA extraction.

This research is approved by the Ethics Committee of Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Third Hospital of Shanxi Medical University. Besides, all participants were well-informed about the necessary information of this study and signed the written informed consent.

2.7. Exosome Separation. The exosome separation followed the steps described by Han et al. [29]. And the exosomes acquired were resuspended in phosphate-buffered saline (PBS). The suspension was placed on a chloroform-coated copper grid with 0.125% Formvar and negatively stained with uranyl acetate. Morphological identification of the exosomes was by a transmission electron microscopy (TEM).

2.8. RNA Extraction and qRT-PCR. Total RNA from the obtained exosomes was extracted following the miRNeasy Micro Kit (QIAGEN, Germany), and RNA quantity and quality were tested via Agilent Bioanalyzer 2100 (Agilent, USA). cDNA was synthesized by reverse transcription from total RNA using SuperScript III Reverse Transcriptase kit (Invitrogen, USA), and qPCR was performed using SYBR Premix Ex Taq II (Takara, Japan). qRT-PCR was performed using ABI7500 (7500, ABI, USA), and the relative expression of all miRNAs was calculated using the $2^{-\Delta\Delta CT}$ method. U6 was the internal reference. Table 1 shows primer sequences for feature miRNAs.

2.9. Statistical Analysis. Based on analysis performed by GraphPad 8.0, box plots were drawn. Differences in the relative expression of miRNAs between tumor and normal samples were analyzed using the t -test, and $p < 0.05$ indicated a difference that was statistically significant.

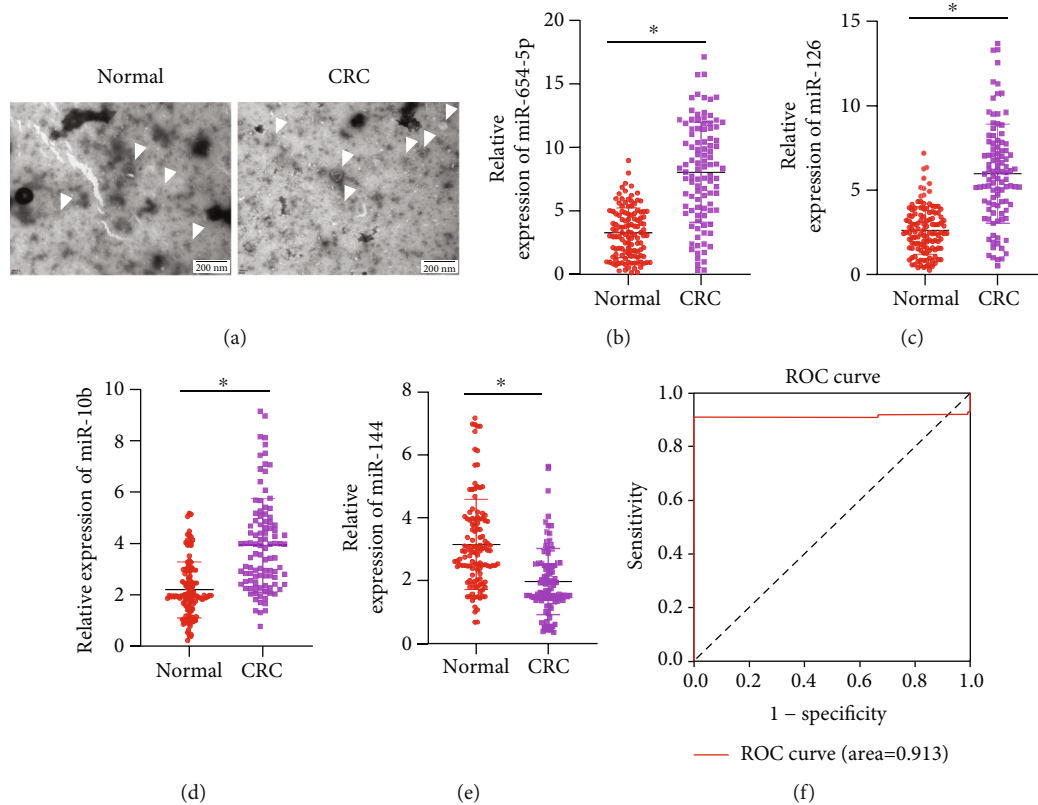


FIGURE 4: qRT-PCR and ROC curve analysis of miRNAs in clinical samples. (a) TEM observation of exosomes. Scale bar: 200 μm . Box plots of differentially expressed (b) miR-654-5p, (c) miR-126, (d) miR-10b, and (e) miR-144 in CRC and normal samples. (f) ROC curves of 4-miRNA in clinical diagnosis $*p < 0.05$.

3. Results

3.1. Constructing the Diagnostic Model of CRC. 56 differentially expressed miRNAs (DEmiRNAs) were obtained by normalization and differentially analyzing miRNAs data derived from CRC and normal exosomes. Subsequent XGBoost feature selection indicated the top five miRNAs with the best ability to distinguish sample types. To determine the optimal diagnostic classifier for CRC, we constructed different MLP classifiers and plotted IFS curves to visually select miRNA combinations. Through the IFS curve, it was found that the classification effect of the MLP classifier composed of the first four miRNAs (miR-654-5p, miR-126, miR-10b, and miR-144) was good, and the 10-fold cross-validation results showed that its MCC value was high (Figure 1), and the sensitivity of this model was 0.977, the specificity was 1.000, the accuracy was 0.980, and the MCC was 0.909.

3.2. Validation of the Performance of the Diagnostic Model. The expression data of four miRNAs in MLP classifiers in CRC and normal samples were subjected to PCA dimensionality reduction. Shown in Figure 2 were that PCA could significantly distinguish CRC and normal samples. Dim1 contributed 41.6% and Dim2 contributed 32.2%. From the violin plots, we could see that levels of blood exosomal miR-654-5p, miR-126, and miR-10b from CRC patients

were markedly higher, but miR-144 was markedly lower than normal participants (Figures 3(a)–3(d)). The above results exhibited that the MLP formed by the former four miRNAs showed the value to assist CRC diagnosis.

3.3. qRT-PCR of miRNAs from Clinical Samples and Receiver Operator Characteristic (ROC) Analysis. To validate the performance of this model in clinical CRC diagnosis, we recruited 100 CRC and 120 healthy participants (Table 2), collected their blood samples, and extracted exosomes for qRT-PCR. Exosomes were first extracted from the blood of CRC patients as well as healthy participants, and the isolated exosomes were subsequently validated for size and morphology. Under a TEM, we could observe that the extracted exosomes were oval membrane-bound vesicles, which were about 50 nm–150 nm in diameter (Figure 4(a)). Thereafter, the qRT-PCR revealed that levels of blood exosomal miR-654-5p, miR-126, and miR-10b from CRC patients were markedly higher (Figures 4(b)–4(d)), but miR-144 was markedly lower than normal participants (Figure 4(e)). Data from qRT-PCR were collected for validation of the performance of the diagnostic model in CRC diagnosis. As results suggested, the ROC of the 4-miRNA diagnostic model was 0.913 (Figure 4(f)), and the recall of the model was 0.91, specificity was 0.34, accuracy was 0.6, and f1 was 0.67. Collectively, qRT-PCR on clinical samples validated that this 4-miRNA model could distinguish CRC

and normal samples precisely, enabling these miRNAs to be biomarkers for CRC diagnosis.

4. Discussion

As key regulators in a variety of biological and physiological processes, miRNA dysregulation may be tightly linked to changes in the pathological environment of disease [30–32]. Colonoscopy is the gold standard for the pathological diagnosis of CRC, but it causes a large physical as well as psychological burden to patients due to its high invasiveness [5–8]. Owing to patients' avoidance of colonoscopy, CRC cannot be diagnosed promptly at the early stage and is only diagnosed at advanced stages when tumor metastasizes to other tissue [33]. The advantage of miRNA detection relative to invasive colonoscopy is that samples are more accessible in clinical practice both in body fluids and blood. At the same time, this noninvasive examination greatly alleviates the physical burden on patients [32, 34, 35]. Given its non-invasive and easily accessible properties, miRNAs are promising biomarkers in CRC diagnosis.

We here utilized XGBoost to determine the key features by ranking feature importance and recursive elimination. We determined the top 5 miRNAs that could accurately distinguish CRC cancer patients from healthy individuals and subsequently found via IFS method that the MLP classifier composed of the top four miRNAs was the best for CRC diagnosis. MLP is a dynamic classifier based on neural network, which could directly determine the separating hyperplanes between the two types of events, with high accuracy of classification and strong ability of parallel distribution processing. At present, there are also some studies on constructing CRC diagnostic classifiers based on machine learning algorithms. Koppad et al. [21] screened CRC diagnosis-related genes by the random forest algorithm, which has the advantage of avoiding data overfitting and reducing the computational load of the model. We aimed to filter biomarkers that could diagnose cancer. While MLP is to classify two types of events, therefore, it was our tool for identifying miRNAs that could assist CRC diagnosis.

The top four miRNAs selected by IFS (miR-654-5p, miR-126, miR-10b, and miR-144) could accurately diagnose CRC. These four miRNAs have all been reported in CRC. Reported by Li et al. [36], the decreased level of miR-654-5p is markedly correlated with the clinical stage of colon cancer by analyzing miR-654-5p level in tissue from CRC patients and normal participants, indicating that its level might be closely related to the CRC progression. As stated by Ebrahimi et al. [37], low miR-126 level in CRC is linked to CRC histological subtype, perineural tumor invasion, microsatellite instability pathological analysis, and lymph node distal metastasis. One study indicated that upregulated miR-10b is discovered in CRC patients with liver metastases, positively linked to advanced TNM stage, and able to predict advanced clinicopathological features and liver metastasis in CRC [38]. Research by Choi et al. [39] indicated that stool from CRC patients is a novel screening biomarker, and the miR-144 level in the stool has good sensitivity and specificity

for CRC detection. Finally, we collected blood samples from CRC patients and normal participants for qRT-PCR, and the expression trends of miRNAs were consistent with those reported in the literature, which also validated the accuracy of our study. Further, PCA revealed that the MLP diagnostic classifier composed of miR-654-5p, miR-126, miR-10b, and miR-144 could well distinguish samples from CRC patients and normal individuals. Hence, these four miRNAs could be unique biomarkers for noninvasive examination of CRC.

However, limitations still exist. Our study utilized the limited numbers of public datasets and did not take into account factors like age, gender, ethnicity, and tumor TNM stages, which may affect miRNA expression. Hence, the construction of a more precise diagnosis model can be achieved by carrying a more detailed analysis on these factors, providing science-based evidence for the clinical noninvasive diagnosis of CRC. Overall, we performed XGBoost and constructed an MLP classifier to identify four miRNAs with the highest diagnostic value. PCA and ROC curves suggested favorable performance of the 4-miRNA classifier to distinguish CRC patients from normal individuals. This study sheds light on science-based theory for the noninvasive diagnosis of CRC.

Data Availability

The data and materials in the current study are available from the corresponding author on reasonable request.

Ethical Approval

This research is approved by the Ethics Committee of Shanxi Bethune Hospital.

Consent

All participants were well-informed about the necessary information of this study and signed the written informed consent.

Disclosure

The funders did not participate in designing, performing, or reporting in the current study.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

GD and CR conceived and designed the study. GD, CR, and JW performed the experiments. JW provided the mutants. GD and JM wrote the paper. GD, CR, and JW reviewed and edited the manuscript. All authors read and approved the manuscript.

Acknowledgments

This study was supported in part by grants from the Health Commission of Shanxi Province (No. 2021150) and Shanxi Province “136 Revitalization Medical Project Construction Funds”.

Supplementary Materials

Supplementary Table 1: clinical information of 220 participants. (*Supplementary Materials*)

References

- [1] E. Goodarzi, R. Beiranvand, H. Naemi, V. Momenabadi, and Z. Khazaei, “Worldwide incidence and mortality of colorectal cancer and human development index (HDI): an ecological study,” *World Cancer Research Journal*, vol. 6, 2019.
- [2] M. Arnold, M. S. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, “Global patterns and trends in colorectal cancer incidence and mortality,” *Gut*, vol. 66, no. 4, pp. 683–691, 2017.
- [3] P. Rawla, T. Sunkara, and A. Barsouk, “Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors,” *Gastroenterology Review*, vol. 14, no. 2, pp. 89–103, 2019.
- [4] J. Bogaert and H. Prenen, “Molecular genetics of colorectal cancer,” *Annals of Gastroenterology*, vol. 27, no. 1, pp. 9–14, 2014.
- [5] H. Strul and N. Arber, “Screening techniques for prevention and early detection of colorectal cancer in the average-risk population,” *Gastrointestinal Cancer Research*, vol. 1, no. 3, pp. 98–106, 2007.
- [6] K. K. Ciombor, C. Wu, and R. M. Goldberg, “Recent therapeutic advances in the treatment of colorectal cancer,” *Annual Review of Medicine*, vol. 66, no. 1, pp. 83–95, 2015.
- [7] K. Simon, “Colorectal cancer development and advances in screening,” *Clinical Interventions in Aging*, vol. Volume 11, pp. 967–976, 2016.
- [8] U.S. Preventive Services Task Force, “Screening for colorectal cancer: recommendation and rationale,” *Annals of Internal Medicine*, vol. 137, no. 2, pp. 129–131, 2002.
- [9] P. Vega, F. Valentin, and J. Cubiella, “Colorectal cancer diagnosis: pitfalls and opportunities,” *World Journal of Gastrointestinal Oncology*, vol. 7, no. 12, pp. 422–433, 2015.
- [10] A. C. von Roon, L. Karamountzos, S. Purkayastha et al., “Diagnostic precision of fecal calprotectin for inflammatory bowel disease and colorectal malignancy,” *The American Journal of Gastroenterology*, vol. 102, no. 4, pp. 803–813, 2007.
- [11] R. J. Simpson, J. W. Lim, R. L. Moritz, and S. Mathivanan, “Exosomes: proteomic insights and diagnostic potential,” *Expert Review of Proteomics*, vol. 6, no. 3, pp. 267–283, 2009.
- [12] X. Li, A. L. Corbett, E. Taatizadeh et al., “Challenges and opportunities in exosome research-perspectives from biology, engineering, and cancer therapy,” *APL Bioengineering*, vol. 3, no. 1, article 011503, 2019.
- [13] N. P. Hessvik and A. Llorente, “Current knowledge on exosome biogenesis and release,” *Cellular and Molecular Life Sciences*, vol. 75, no. 2, pp. 193–208, 2018.
- [14] A. E. Sedgwick and C. D’Souza-Schorey, “The biology of extracellular microvesicles,” *Traffic*, vol. 19, no. 5, pp. 319–327, 2018.
- [15] B. N. Hannafon and W. Q. Ding, “Intercellular communication by exosome-derived microRNAs in cancer,” *International Journal of Molecular Sciences*, vol. 14, no. 7, pp. 14240–14269, 2013.
- [16] J. Wang, F. Yan, Q. Zhao et al., “Circulating exosomal miR-125a-3p as a novel biomarker for early-stage colon cancer,” *Scientific Reports*, vol. 7, no. 1, p. 4150, 2017.
- [17] S. Yan, G. Dang, X. Zhang et al., “Downregulation of circulating exosomal miR-638 predicts poor prognosis in colon cancer patients,” *Oncotarget*, vol. 8, no. 42, pp. 72220–72226, 2017.
- [18] C. Gao, C. Zhou, J. Zhuang et al., “MicroRNA expression in cervical cancer: novel diagnostic and prognostic biomarkers,” *Journal of Cellular Biochemistry*, vol. 119, no. 8, pp. 7080–7090, 2018.
- [19] Z. Han, Y. Li, J. Zhang et al., “Tumor-derived circulating exosomal miR-342-5p and miR-574-5p as promising diagnostic biomarkers for early-stage lung adenocarcinoma,” *International Journal of Medical Sciences*, vol. 17, no. 10, pp. 1428–1438, 2020.
- [20] H. Lian, Y. P. Han, Y. C. Zhang et al., “Integrative analysis of gene expression and DNA methylation through one-class logistic regression machine learning identifies stemness features in medulloblastoma,” *Molecular Oncology*, vol. 13, no. 10, pp. 2227–2245, 2019.
- [21] S. Koppad, A. Basava, K. Nash, G. V. Gkoutos, and A. Acharjee, “Machine learning-based identification of colon cancer candidate diagnostics genes,” *Biology*, vol. 11, no. 3, 2022.
- [22] M. E. Ritchie, B. Phipson, D. Wu et al., “limma powers differential expression analyses for RNA-sequencing and microarray studies,” *Nucleic Acids Research*, vol. 43, no. 7, article e47, 2015.
- [23] T. Chen and C. Guestrin, “XGBoost: a scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, San Francisco, California, USA, 2016.
- [24] F. Yang, X. Wang, H. Ma, and J. Li, “Transformers-sklearn: a toolkit for medical language understanding with transformer-based models,” *BMC Medical Informatics and Decision Making*, vol. 21, no. S2, p. 90, 2021.
- [25] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” *BMC Genomics*, vol. 21, no. 1, p. 6, 2020.
- [26] M. Ringner, “What is principal component analysis?,” *Nature Biotechnology*, vol. 26, no. 3, pp. 303–304, 2008.
- [27] A. Garcia-Rudolph, A. Garcia-Molina, E. Opisso, and J. Tormos Munoz, “Personalized web-based cognitive rehabilitation treatments for patients with traumatic brain injury: cluster analysis,” *JMIR Medical Informatics*, vol. 8, no. 10, article e16077, 2020.
- [28] S. B. Edge and C. C. Compton, “The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM,” *Annals of Surgical Oncology*, vol. 17, no. 6, pp. 1471–1474, 2010.
- [29] S. Han, G. Li, M. Jia et al., “Delivery of anti-miRNA-221 for colorectal carcinoma therapy using modified cord blood mesenchymal stem cells-derived exosomes,” *Frontiers in Molecular Biosciences*, vol. 8, article 743013, 2021.
- [30] T. Treiber, N. Treiber, and G. Meister, “Regulation of microRNA biogenesis and its crosstalk with other cellular pathways,” *Nature Reviews. Molecular Cell Biology*, vol. 20, no. 1, pp. 5–20, 2019.

- [31] D. P. Bartel, "Metazoan microRNAs," *Cell*, vol. 173, no. 1, pp. 20–51, 2018.
- [32] C. E. Condrat, D. C. Thompson, M. G. Barbu et al., "miRNAs as biomarkers in disease: latest findings regarding their role in diagnosis and prognosis," *Cell*, vol. 9, no. 2, 2020.
- [33] J. S. Levine and D. J. Ahnen, "Adenomatous polyps of the colon," *The New England Journal of Medicine*, vol. 355, no. 24, pp. 2551–2557, 2006.
- [34] R. Rupaimoole and F. J. Slack, "MicroRNA therapeutics: towards a new era for the management of cancer and other diseases," *Nature Reviews. Drug Discovery*, vol. 16, no. 3, pp. 203–222, 2017.
- [35] A. Fesler, J. Jiang, H. Zhai, and J. Ju, "Circulating microRNA testing for the early diagnosis and follow-up of colorectal cancer patients," *Molecular Diagnosis & Therapy*, vol. 18, no. 3, pp. 303–308, 2014.
- [36] P. Li, J. X. Cai, F. Han et al., "Expression and significance of miR-654-5p and miR-376b-3p in patients with colon cancer," *World Journal of Gastrointestinal Oncology*, vol. 12, no. 4, pp. 492–502, 2020.
- [37] F. Ebrahimi, V. Gopalan, R. Wahab, C. T. Lu, R. Anthony Smith, and A. K. Y. Lam, "Deregulation of miR-126 expression in colorectal cancer pathogenesis and its clinical significance," *Experimental Cell Research*, vol. 339, no. 2, pp. 333–341, 2015.
- [38] H. Jiang, J. Liu, Y. Chen, C. Ma, B. Li, and T. Hao, "Up-regulation of mir-10b predicate advanced clinicopathological features and liver metastasis in colorectal cancer," *Cancer Medicine*, vol. 5, no. 10, pp. 2932–2941, 2016.
- [39] H. H. Choi, Y. S. Cho, J. H. Choi, H. K. Kim, S. S. Kim, and H. S. Chae, "Stool-based miR-92a and miR-144* as noninvasive biomarkers for colorectal cancer screening," *Oncology*, vol. 97, no. 3, pp. 173–179, 2019.

Retraction

Retracted: Clinical Characteristics and Treatment of Adult Idiopathic Carpal Tunnel Syndrome Accompanied with Trigger Digit

Computational and Mathematical Methods in Medicine

Received 26 September 2023; Accepted 26 September 2023; Published 27 September 2023

Copyright © 2023 Computational and Mathematical Methods in Medicine. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] J. Hong, X. Wang, J. Xue, J. Li, M. Zhang, and W. Mao, "Clinical Characteristics and Treatment of Adult Idiopathic Carpal Tunnel Syndrome Accompanied with Trigger Digit," *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 8104345, 7 pages, 2022.

Research Article

Clinical Characteristics and Treatment of Adult Idiopathic Carpal Tunnel Syndrome Accompanied with Trigger Digit

Jinjiong Hong , Xiaofeng Wang, Jianbo Xue, Jimin Li, Minghua Zhang, and Weisheng Mao 

Department of Hand Surgery, Department of Plastic Reconstructive Surgery, Ningbo No. 6 Hospital, China

Correspondence should be addressed to Weisheng Mao; s4150651@126.com

Received 18 August 2022; Revised 13 September 2022; Accepted 23 September 2022; Published 11 October 2022

Academic Editor: Lei Chen

Copyright © 2022 Jinjiong Hong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose. To investigate the clinical characteristics and treatment of adult idiopathic carpal tunnel syndrome (CTS) accompanied with trigger digit. **Materials and Methods.** A retrospective analysis was performed on a total of 74 patients with adult idiopathic CTS accompanied with trigger digit admitted to and treated at the Hand Surgery Department of Ningbo No. 6 Hospital from January 1, 2017 to December 31, 2019. Data on patients' gender, age, occupation, course of the disease, menstruation, surgeries, examination-related information, complications, treatment methods, and prognoses during follow-up were recorded and subsequently used to analyze the pathogeneses, clinical characteristics, and treatment. **Results.** A total of 74 patients (72 females and 2 males) were included. Among female patients, 51 were postmenopausal and 18 were non-postmenopausal. There were 101 fingers with trigger digit, including 14 patients with trigger digit in both hands, and 115 wrists affected by the CTS. The average course of CTS was 34.5 ± 49.3 months, and that of trigger digit was 10.5 ± 22.4 months. Seventy had both trigger digit and CTS in one hand, while among patients with both hands involved, only 4 had trigger digit or CTS in one hand. Eighty-nine fingers underwent A1 pulley release, and 104 hands underwent carpal tunnel surgery, with steroids being injected under the adventitia of the median nerve during the surgery. All patients who underwent surgeries had I/A-healed incisions, and 14 of them had obvious synovial hyperplasia observed in the carpal tunnel and flexor tendon sheath during surgeries. Follow-up visits, which lasted 3 to 35 months, had an average duration of 1.34 years and included 72 patients. In 63 patients (63/72), the syndrome of tenosynovitis and numbness disappeared and normal hand functions were restored; in 6 patients, the numbness in hands greatly improved and normal hand functions were almost completely restored, while no improvement in numbness of hands and limited hand functions were still observed in 3 patients. **Conclusion.** CTS accompanied with trigger digit was more common in postmenopausal females, and the course of CTS was longer than that of trigger digit. CTS and trigger digit were more likely to simultaneously occur in the same hand, while some patients might not have obvious synovial hyperplasia in the carpal tunnel. Surgeries were effective in severe cases.

1. Introduction

Idiopathic carpal tunnel syndrome (CTS) and stenosing tenosynovitis, also known as trigger digit (TD), are common diseases treated at hand surgery departments [1–3]. Existing studies have suggested that multiple factors, such as age, gender, diabetes, perimenopause, pregnancy, rheumatoid arthritis, and other inflammatory arthritis and engagement in assembly line works, may increase the risk of CTS [4]. Some studies suggested that CTS accompanied with TD may occur in the same hand, successively or simultaneously [5, 6]. Yet, CTS with TD simultaneously is less common [7],

and factors related to the risk of the onset of CTS with TD and its treatment have not yet been reported. The aim of this study was to investigate patients' clinical characteristics and treatment.

2. Materials and Methods

A retrospective analysis was performed on patients with CTS admitted to and treated at the Hand Surgery Department of Ningbo No. 6 Hospital from January 1, 2017 to December 31, 2019. In this study, adults with symptoms of CTS and

TD were enrolled. The criteria for exclusion were malignancy, less than 20 years old, pregnancy, a history of forearm and hand trauma, CTS induced by tophus, tumor lesion, or wrist infection (e.g., nonmycobacterium tuberculosis infection), traumatic peripheral nerve injury and central nervous system diseases, and recurrent CTS. Patients with gout-induced stenosing tenosynovitis, tuberculous tenosynovitis, infective tenosynovitis, and other secondary tenosynovitis were also excluded. All of the patients were admitted to the hospital after invalid conservative treatment since the onset of the disease. The study was approved by the ethics committee of the Ningbo No.6 Hospital in Zhejiang (China). All patients signed informed consents at admission.

2.1. Statistical Analysis. SPSS software for Windows 13.0 (SPSS Inc., Chicago) was used to analyze all data. A P value of <0.05 was considered to be statistically significant. Measurement data were expressed as mean \pm standard deviation (SD), and the two-independent-samples t -test was used for data analysis.

3. Results

Eventually, 74 patients were included in the study, including 72 females aged 40 to 71 years old (average age of 53.9 ± 6.7 years) and only 2 males aged 32 and 68 years old, respectively (Figure 1).

Among female participants, there were 51 postmenopausal, 18 non-postmenopausal, and 3 women who underwent hysterectomy before menopause. In terms of occupation, there were 29 operating workers, 25 housekeeping workers, 18 farmers, 1 office worker, and 1 freelancer. The course of the disease ranged from 6 months to 15 years, with an average CTS course of 34.5 months and an average TD course of 10.5 months. One of the patients suffered from chronic gastric ulcer, 3 had hysterectomy, 1 had thyroidectomy, 1 had arrhythmia, 1 had oophorocystectomy, 1 had hyperthyroidism, 5 had diabetes, and 14 had hypertension.

3.1. Disease Status and Examinations. In total, there were 101 fingers (including 46 on the left hand and 55 on the right hand) with TD, including 52 thumbs, 7 index fingers, 29 middle fingers, 11 ring fingers, and 2 little fingers. Based on the classification described by Lim et al. [8], the study included 3 Grade I (3%), 9 Grade II (9%), 78 Grade III (77%), and 11 Grade IV (11%) trigger digits.

Moreover, a total of 115 hands involved carpal tunnel presentation, with 41 patients having CTS in both wrists, 11 patients having CTS in the left wrist, and 22 patients having CTS in the right wrist. According to the electrophysiological grading of severity of CTS proposed by Padua et al. [9], 22 carpal tunnels were graded as mild (MILD), 43 as moderate (MOD), 47 as severe (SEV), and 3 as extremely severe (EXT) (Table 1 and Figure 2). If a patient had both TD and CTS in one hand, he/she was classified as an ipsilateral patient, regardless if he/she had TD and/or CTS in the other hand. If a patient had only TD or CTS in either hand, he/she was classified as a nonipsilateral patient. Finally, there were 70 ipsilateral patients and 4 nonipsilateral patients.

Specifically, there were 12 ipsilateral patients with both hands involved and 58 ipsilateral patients with only one hand involved (including 24 with the left hand involved and 34 with the right hand involved).

All of the 74 patients underwent preoperative X-ray examination and electrophysiological examination of the median nerve, while some were selectively given cervical MRI, wrist MRI, and diagnostic ultrasonography. The X-ray examination, which was used to rule out the possibility of median nerve compression induced by malunion of an old fracture of the distal radius with lunate dislocation, confirmed that none of the patients had signs of fracture-induced dislocation, bone tumor, or bone erosion.

Patients with stenosing tenosynovitis and idiopathic CTS accounted for 8.4% of all patients with CTS who were admitted to the hospital (74/884). Among female patients, 51 were postmenopausal aged 44 to 71 years old (average age of 56.6 ± 5.8 years), with the average age of menopause being 49.5 ± 3.0 years old, including 1 patient who underwent a hysterectomy after menopause. The menopause in these women started at the average age of 49.5 ± 3.0 years, with the duration of menopause ranging from 0.5 to 20 years (average of 7.1 ± 5.7 years). There were 18 non-postmenopausal female patients aged 40 to 53 years old (average age 47.1 ± 3.7 years). The ages of non-postmenopausal female patients and postmenopausal female patients approximately conformed to the normal distribution; thus, two-independent-samples t -test was used to analyze them with P value of <0.05 .

Among the 74 patients, there were 101 fingers with TD, including 52 thumbs, 7 index fingers, 29 middle fingers, 11 ring fingers, and 2 little fingers, with the thumbs accounting for 51.5% (52/101). Of all patients, 52 patients had one finger with TD, accounting for 70.3% (52/74), and 22 had multiple fingers with TD, accounting for 29.7% (22/74); 14 patients had both hands with TD, accounting for 19% (14/74), including 9 patients with both thumbs involved, accounting for 64% (9/14), and 2 patients having multiple (two or more) fingers on both hands involved. Among the 74 patients, CTS was found in 115 hands, with those severely and extremely severely involved, accounting for 43.5% of the total (50/115), and those with CTS in both hands, accounting for 55.4% of all patients (41/74), including 13 patients diagnosed with severe and 1 patient diagnosed with extremely severely CTS, accounting for 34.1% of all patients having CTS in both hands (14/41). Of all patients, 94.6% (70/74) had both TD and CTS in one hand.

The average course of CTS was significantly longer compared to the TD course, lasting 34.5 ± 49.3 months and 10.5 ± 22.4 months, respectively ($P < 0.01$). Forty-six patients had a longer course of CTS than TD, including 2 patients who underwent blocking of the diseased wrist and 2 patients who underwent blocking of the diseased finger during the course of the disease. The average course of CTS in these 4 patients was 50.0 ± 56.0 months, and that of TD in patients was 6.5 ± 11.9 months ($P < 0.01$). Comparatively, 11 patients had a longer course of TD than CTS, including 2 patients who underwent intrathecal injection of steroids during the course. Furthermore, 17 patients

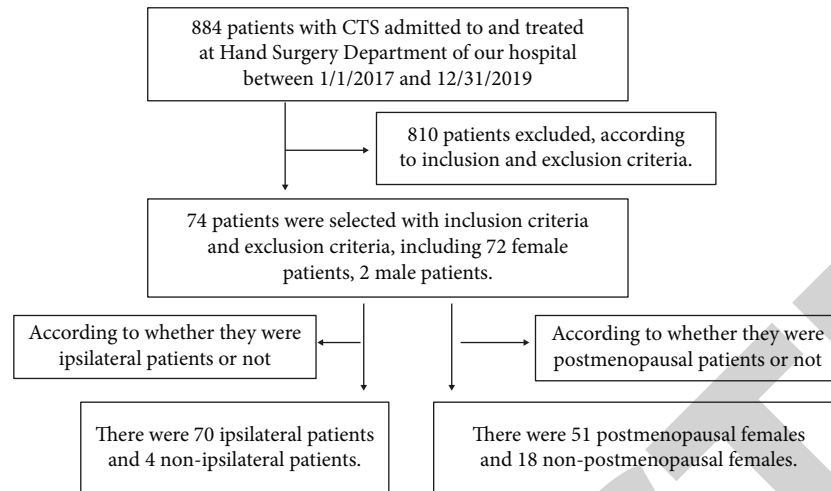


FIGURE 1: Flow chart of patients in this study.

TABLE 1: Characteristics and electrophysiological grading of severity of CTS of the patients.

Characteristics	Type	Number of cases (%)
Gender	Female/male	72/2 (97.3/2.7)
Menstruation	Postmenopausal females/non-postmenopausal females/others	51/18/3 (70.8/25/4.2)
Average age (year)	Postmenopausal females	56.6 ± 5.8
The average course of CTS (month)		34.5 ± 49.3
The average course of TD (month)		10.5 ± 22.4
Electrophysiological grading of severity of CTS	Severe (SEV)	47 (40.9)
	Moderate (MOD)	43 (37.4)
	Mild (MILD)	22 (19.1)
	Extremely severe (EXT)	3 (2.6)

simultaneously had CTS and TD, with an average course of disease of 7.2 ± 9.2 months.

3.2. Treatment Methods. All surgeries were performed by senior hand surgeons. According to the severity of stenosing tenosynovitis, 12 fingers were given conservative treatment: 6 fingers of grade I and grade II patients who were given external application of Voltaren emulsion and local immobilization, and 6 fingers that were treated with intrathecal injection of steroids.

Eighty-nine fingers of grade III and grade IV patients were treated with A1 pulley release. In one patient with right thumb TD and moderate right CTS with the course of TD lasting 6 months who was treated with A1 pulley release of the right thumb and conventional open carpal tunnel release surgery, a long segment of large partial-rupture and adhesion in the flexor longus tendon was observed during the surgery, after which the patient was given a graft of palmaris longus.

Three wrists were treated with simplex intrajunction of steroids, 8 wrists with conservative immobilization treatment for the carpal tunnel, and surgery and injection of steroids under the neural adventitia were used to treat the carpal tunnel in 104 wrists, including endoscopic carpal tunnel release (ECTR) performed in 42 wrists, miniopen carpal tunnel release (MOCRT) performed in 39 wrists, mini-

incision OCTR performed in 15 wrists, and conventional large OCTR performed in 8 wrists [10]. One patient who had severe CTS in the right wrist and tenosynovitis in the 2nd to 4th fingers on the right hand with the course of the disease lasting over about 15 years received A1 pulley release, OCTR, and palmaris longus tendon transfer for improving thumb function. Fourteen patients in the study had obvious synovial hyperplasia observed in the carpal tunnel with or without flexor tendon sheath during surgery, in whom the synovium was removed and sent for pathological test, which suggested fibrous synovial tissue hyperplasia. Clinical results at the final follow-up were evaluated according to the criteria of Kelly et al. [11]

TD symptoms significantly improved in all patients; numbness was significantly alleviated in 66 patients and 107 hands, while 8 patients with 8 hands, of which 3 hands were graded extremely severe and 5 were graded severe according to electrophysiological grading of severity before surgery, showed no significant alleviation of numbness. On Day 1 after surgery, TD symptoms and numbness did not aggravate in any of the patients. All incisions in the patients' given surgery were I/A-healed. Follow-up visits were completed in 72 patients and lasted 3 to 35 months, with an average duration of 1.34 years. According to the results at the final follow-up, sixty-three patients (63/72) experienced

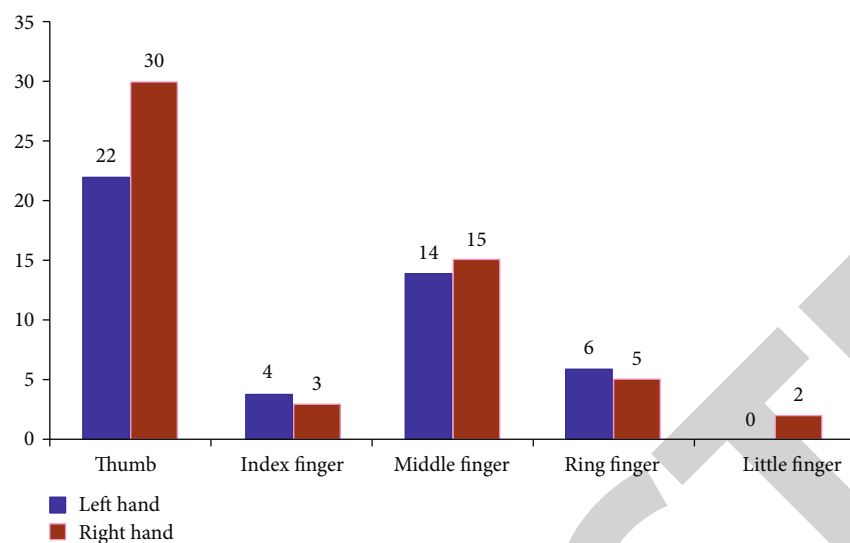


FIGURE 2: Distribution of TDs.

tenosynovitis; the numbness disappeared, and normal hand functions were restored; numbness was greatly improved, and normal hand function was almost completely restored in 6 patients, while 3 patients showed no improvement in numbness of hands and continued to have limited hand function. Using Kelly's criteria was excellent in 63 patients (87.5%), good in 6 patients (8.3%), and poor in 3 patients (4.2%).

4. Discussion

4.1. Factors Related to the Pathogenesis of Adult Stenosing Tenosynovitis with CTS. Previous studies have shown that women are more prone to CTS and that idiopathic CTS is more often seen in women than in men, while TD is the most often seen in women aged 40 to 60 years old [1]. Idiopathic CTS shows a female predominance and epidemiologic studies have shown that postmenopausal status or oophorectomy, which both represent fluctuations in estrogen level, is associated with idiopathic CTS [11]. According to McKinlay et al., women on average go through perimenopause at 47 years of age and, with the continuous depletion of ovarian follicles in middle age, they eventually experience changes in menstrual patterns and fluctuations in hormonal levels. Postmenopausal women are more prone to idiopathic CTS, which is related to steroid hormones, especially estrogens [13]. According to Kim et al. [12], marked increases in the expressions of estrogen receptor- α and estrogen receptor- β were found in the tenosynovial tissues of postmenopausal idiopathic CTS patients. Kaplan et al. believed that hypoestrogenic might affect the median nerve or other carpal tunnel structures, thus resulting in CTS [15]. In our study, the fact that there were more postmenopausal women suggested that postmenopause was the pathogenesis of stenosing tenosynovitis with CTS, although there was a lack of convincing evidence for the correlation between the onset of tenosynovitis and estrogen fluctuation in postmenopausal women. It might also be intriguing to investigate into the

incidence of CST and TD in other complicated conditions [16–21].

According to Vance et al., around 1.5% to 20% of diabetic patients have secondary TD [22]. Ashour et al. [23] pointed out that the lifetime incidence of TD in the healthy population was 2.6%, and that in the diabetes population, it was as high as 16.5%. In a prospective study that included 160 patients with stenosing tenosynovitis, 38 had diabetes, accounting for 23.7% [24], while in this retrospective study, 5 patients with CTS and TD also had diabetes, accounting for only 7% of patients. There still lacks evidence for the significant correlation between diabetes and the onset of CTS with TD.

Adult TD may occur in one or more fingers on one or both hands and more often in the thumbs, middle fingers, and ring fingers. A previous study found that patients having TD in both hands accounted for 19% (14/74) of the total, where 64% (9/14) had both thumbs involved. Also, those having CTS in both hands accounted for 55.4% (41/74), where those having SEV and EXT accounted for 34.1% (14/41). Some observational reports suggested that TD was related to occupation or frequent hand activities. Nonetheless, this is still being disputed as some scholars believe that repeated hand working, especially repeated pinching and gripping at the flexor position and repeated flexing and stretching of the tendons of hands, may result in CTS induced by synovial hyperplasia of flexor tendons [25]. Previous studies using MRI provided strong arguments that idiopathic CTS development was closely related to the abnormality of synovial tissues in the carpal tunnel [26]. Ten et al. [3] suggested that CTS might result from tenosynovitis without significant clinical symptoms, which could be detected by ultrasonography. They also believed that synovitis symptoms observed in surgeries and ultrasonography were usually reactive changes responding to mechanical stress or pressure in the carpal tunnel. Our results showed that operating workers and long-term housekeeping workers accounted for 73.0% of all patients (54/74), and the

involvement of both hands was more common in such patients. This suggested that TD with CTS was greatly related to occupation or frequent hand activities. Especially after the disease involved the wrist or thumb on one hand, the activities of the other hand increased as compensation, eventually resulting in the involvement of both hands.

4.2. Analysis of Characteristics and Causes of Adult Stenosing Tenosynovitis with CTS. TD is one of the most common causes of hand pain in adults and often the leading cause of frequent visits to the hospital. CTS patients develop numbness and tingling in fingers at the early stage of the disease, which is aggravated as the disease progresses, especially at night, thus affecting sleep. Some patients may visit the hospital only for numbness of the middle finger. With the disease's progression, patients develop atrophy of the thenar muscle and lack the strength of the thumb against the palm. Some patients may also visit the hospital only for atrophy of the thenar muscle and feel no sensory disturbance. Our results showed the course of TD in such patients was similar to that in patients having CTS and TD at the same time. Therefore, it is possible that patients with a longer course of CTS than TD did not visit the hospital in time for initial symptoms of CTS, while patients having CTS and TD at the same time visited the hospital more timely. The tolerance and repeatability of the initial symptoms of CTS might explain the delayed visit to the hospital in the majority of patients at the early stage of the disease. Patients with early CTS visited the hospital only after they had TD that caused hand pain.

The flexor tendon moves from the forearm to the far side through the carpal tunnel and enters the flexor sheath pulley system of fingers so that it can glide freely in this bone-fiber channel. In most patients with TD, the compression occurs at the A1 pulley. According to a previous study, the main histopathological change in patients with TD was the metaplasia of fibrous cartilage in the ligamentum layer of tendon sheath at A1 pulley, accompanied by secondary reduction of the cross-sectional area of the fibrous bone tunnel [27]. In the present study, 14 patients had obvious synovial hyperplasia in the carpal tunnel, and TD with CTS more often occurred in one hand, accounting for 94.6% of the total (70/74). This could be explained by the cooccurrence of tenosynovitis in the tendon sheath and carpal tunnel, which might be the result of the anatomical features of the flexor tendon gliding in the two-way (carpal tunnel-tendon sheath) tandem channel. CTS and TD occurred in different hands in 4 patients in the present study, which might be related to systemic factors such as gender, age, and hormonal levels. Also, CTS and TD might involve both hands in some cases due to being independent of each other.

4.3. Treatment for CTS with TD. The treatment for idiopathic CTS with stenosing tenosynovitis is basically the same as that used for individual treatment of these conditions, and it includes conservative treatment and surgical treatment. It is currently believed that surgical treatment for CTS is more effective, especially in the long run than splintage and other conservative treatments [28]. Carpal tunnel release (CTR) that cuts the transverse carpal ligament (TCL) is considered

the most reliable method to improve the symptoms. Surgical treatment of CTS, which depends on surgical vision and how surgeons cut the TCL, is divided into the conventional open carpal tunnel release surgery, mini-incision carpal tunnel incision, MOCTR, double-incision carpal tunnel incision, and endoscopic surgery [29]. The use of the surgical method is mostly at the personal discretion of the surgeon. Patients with obvious synovial hyperplasia may be treated with synovial dissection around the flexor tendon after CTR. Our data showed that surgical treatment had satisfactory results, although some patients had commemorative irreversible neurological dysfunction due to the long course of CTS and delayed treatment. However, there was no evidence for the necessity of synovial dissection after incision of the A1 pulley and the transverse carpal ligament. Moreover, 14 patients with CTS and TD had obvious synovial hyperplasia observed in the carpal tunnel during surgeries, accounting for 19.0% of all patients. According to Ten et al., the clinical efficacy of carpal tunnel incision might be better when tenosynovitis is suggested in ultrasonography, and the carpal tunnel is subject to excessive internal pressure [3]. However, cutting the transverse carpal ligament may cause surgical risks and complications, such as injuries to the palmar cutaneous branch, the thenar branch of the median nerve, the superficial palmar arch or flexor tendon, and secondary stenosing tenosynovitis, in either the conventional CTR or endoscopic CTR [6]. Studies on synovial specimens from patients with idiopathic CTS showed that only 10% had inflammatory changes [30], and the majority showed edema or fibrosis. Different from patients with CTS induced by nonmycobacterium tuberculosis infection, some patients with CTS and TD did not have obvious abnormal synovial hyperplasia in the tendon sheath or carpal canal, which might be due to other possible unknown diseases or pathogenic factors in such patients that should be addressed by further studies. The treatment for idiopathic CTS with TD with obvious synovial hyperplasia in the carpal tunnel by removing proliferated synovial tissues and retaining the transverse carpal ligament after complete laboratory examinations needs to be further explored and studied.

5. Conclusion

In summary, we found that CTS accompanied with trigger digit was more common in postmenopausal females, and the course of CTS was longer than that of trigger digit. CTS and trigger digit were more likely to simultaneously occur in the same hand, while some patients might not have obvious synovial hyperplasia in the carpal tunnel. Surgeries were effective in severe cases.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was funded by the Natural Science Foundation of Ningbo, Grant number: 2019A610279; the Public Welfare Project of Ningbo, Grant number: 2019C50044; and the Agricultural and Social Science and Technology Project of Yinzhou District, Grant number: 2020AS0030.

References

- [1] J. S. Moore, "Flexor tendon entrapment of the digits (trigger finger and trigger thumb)," *Journal of Occupational and Environmental Medicine*, vol. 42, no. 5, pp. 526–545, 2000.
- [2] T. Civi Karaaslan, O. Berköz, and E. Tarakci, "The effect of mirror therapy after carpal tunnel syndrome surgery: a randomised controlled study," *Hand Surgery & Rehabilitation*, vol. 39, no. 5, pp. 406–412, 2020.
- [3] D. F. Ten Cate, N. Glaser, J. J. Luime et al., "A comparison between ultrasonographic, surgical and histological assessment of tenosynovitis in a cohort of idiopathic carpal tunnel syndrome patients," *Clinical Rheumatology*, vol. 35, no. 3, pp. 775–780, 2016.
- [4] B. Graham, A. E. Peljovich, R. Afra et al., "The American Academy of Orthopaedic Surgeons evidence-based clinical practice guideline on," *The Journal of Bone and Joint Surgery. American Volume*, vol. 98, no. 20, pp. 1750–1754, 2016.
- [5] M. A. Acar, H. Kutahya, A. Gulec, M. Elmadag, N. Karalezli, and T. C. Ogun, "Triggering of the digits after carpal tunnel surgery," *Annals of Plastic Surgery*, vol. 75, no. 4, pp. 393–397, 2015.
- [6] D. Zhang, J. Collins, B. E. Earp, and P. Blazar, "Relationship of carpal tunnel release and new onset trigger finger," *The Journal of Hand Surgery*, vol. 44, no. 1, pp. 28–34, 2019.
- [7] S. Mandiroglu and E. Alemdaroglu, "Idiopathic carpal tunnel syndrome and de Quervain's tenosynovitis: is there an association?," *Somatosensory & Motor Research*, vol. 38, no. 4, pp. 353–356, 2021.
- [8] M. H. Lim, K. K. Lim, M. Z. Rasheed, S. Narayanan, A. Beng-Hoi Tan, and P. Tonali, "Outcome of open trigger digit release," *Journal of Hand Surgery (European Volume)*, vol. 32, no. 4, pp. 457–459, 2007.
- [9] L. Padua, M. Lo Monaco, B. Gregori, E. M. Valente, R. Padua, and P. Tonali, "Neurophysiological classification and sensitivity in 500 carpal tunnel syndrome hands," *Acta Neurologica Scandinavica*, vol. 96, no. 4, pp. 211–217, 1997.
- [10] A. M. Schwarz, G. Lipnik, G. M. Hohenberger, A. Krauss, and M. Plecko, "Mini-open carpal tunnel release: technique, feasibility and clinical outcome compared to the conventional procedure in a long-term follow-up," *Scientific Reports*, vol. 12, no. 1, p. 9122, 2022.
- [11] C. P. Kelly, D. Pulisetti, and A. M. Jamieson, "Early experience with endoscopic carpal tunnel release," *Journal of Hand Surgery (British)*, vol. 19, no. 1, pp. 18–21, 1994.
- [12] J. K. Kim, H. J. Hann, M. J. Kim, and J. S. Kim, "The expression of estrogen receptors in the tenosynovium of postmenopausal women with idiopathic carpal tunnel syndrome," *Journal of Orthopaedic Research*, vol. 28, no. 11, pp. 1469–1474, 2010.
- [13] S. M. McKinlay, D. J. Brambilla, and J. G. Posner, "The normal menopause transition," *American Journal of Human Biology*, vol. 4, no. 1, pp. 37–46, 1992.
- [14] Y. Yamanaka, K. Menuki, T. Tajima et al., "Effect of estradiol on fibroblasts from postmenopausal idiopathic carpal tunnel syndrome patients," *Journal of Cellular Physiology*, vol. 233, no. 11, pp. 8723–8730, 2018.
- [15] Y. Kaplan, S. G. Kurt, and H. Karaer, "Carpal tunnel syndrome in postmenopausal women," *Journal of the Neurological Sciences*, vol. 270, no. 1–2, pp. 77–81, 2008.
- [16] Z. Shi, L. Ma, H. Wang et al., "Insulin and hypertonic glucose in the management of aseptic fat liquefaction of post-surgical incision: a meta-analysis and systematic review," *International Wound Journal*, vol. 10, no. 1, pp. 91–97, 2013.
- [17] L. Deng, X. Li, Z. Shi, P. Jiang, D. Chen, and L. Ma, "Maternal and perinatal outcome in cases of fulminant viral hepatitis in late pregnancy," *International Journal of Gynecology & Obstetrics*, vol. 119, no. 2, pp. 145–148, 2012.
- [18] X. M. Li, L. Ma, Y. B. Yang, Z. J. Shi, and S. S. Zhou, "Prognostic factors of fulminant hepatitis in pregnancy," *Chinese Medical Journal*, vol. 118, no. 20, pp. 1754–1757, 2005.
- [19] J. Vasquez-Vivar, Z. Shi, K. Luo, K. Thirugnanam, and S. Tan, "Tetrahydrobiopterin in antenatal brain hypoxia-ischemia-induced motor impairments and cerebral palsy," *Redox Biology*, vol. 13, pp. 594–599, 2017.
- [20] Z. Shi, J. Vasquez-Vivar, K. Luo et al., "Ascending lipopolysaccharide-induced intrauterine inflammation in near-term rabbits leading to newborn neurobehavioral deficits," *Developmental Neuroscience*, vol. 40, no. 5–6, pp. 534–546, 2019.
- [21] J. Vasquez-Vivar, Z. Shi, J. W. Jeong et al., "Neuronal vulnerability to fetal hypoxia-reoxygenation injury and motor deficit development relies on regional brain tetrahydrobiopterin levels," *Redox Biology*, vol. 29, p. 101407, 2020.
- [22] M. C. Vance, J. J. Tucker, and N. G. Harness, "The association of hemoglobin A1c with the prevalence of stenosing flexor tenosynovitis," *The Journal of Hand Surgery*, vol. 37, no. 9, pp. 1765–1769, 2012.
- [23] A. Ashour, A. Alfattni, and A. Hamdi, "Functional outcome of open surgical A1 pulley release in diabetic and nondiabetic patients," *Journal of Orthopaedic Surgery*, vol. 26, no. 1, p. 2309499018758069, 2018.
- [24] A. D. Sobel, A. E. M. Eltorai, B. Weiss, P. K. Mansuripur, and A. C. Weiss, "What patient-related factors are associated with an increased risk of surgery in patients with stenosing tenosynovitis? A prospective study," *Clinical Orthopaedics and Related Research*, vol. 477, no. 8, pp. 1879–1888, 2019.
- [25] A. M. Ettema, P. C. Amadio, C. Zhao, L. E. Wold, and K. N. An, "A histological and immunohistochemical study of the subsynovial connective tissue in idiopathic carpal tunnel syndrome," *The Journal of Bone and Joint Surgery. American Volume*, vol. 86, no. 7, pp. 1458–1466, 2004.
- [26] S. Uchiyama, T. Itsubo, K. Nakamura, H. Kato, T. Yasutomi, and T. Momose, "Current concepts of carpal tunnel syndrome: pathophysiology, treatment, and evaluation," *Journal of Orthopaedic Science*, vol. 15, no. 1, pp. 1–13, 2010.
- [27] M. C. Sbernardori and P. Bandiera, "Histopathology of the A1 pulley in adult trigger fingers," *The Journal of Hand Surgery, European Volume*, vol. 32, no. 5, pp. 556–559, 2007.
- [28] R. J. Verdugo, R. A. Salinas, J. L. Castillo, J. G. Cea, and Cochrane Neuromuscular Group, "Surgical versus non-surgical treatment for carpal tunnel syndrome," *Cochrane Database of Systematic Reviews*, vol. 4, article CD001552, 2008.

Research Article

Construction and Validation of a Prognostic Model Based on mRNAsi-Related Genes in Breast Cancer

Xugui Zhao  and Jianqing Lin 

Department of Thyroid and Breast Surgery, The Second Affiliated Hospital of Fujian Medical University, Quanzhou 362000, China

Correspondence should be addressed to Jianqing Lin; ljq13905977336@163.com

Received 20 July 2022; Accepted 12 September 2022; Published 11 October 2022

Academic Editor: Tao Huang

Copyright © 2022 Xugui Zhao and Jianqing Lin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Breast cancer is a big threat to the women across the world with substantial morbidity and mortality. The pressing matter of our study is to establish a prognostic gene model for breast cancer based on mRNAsi for predicting patient's prognostic survival. **Methods.** From The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases, we downloaded the expression profiles of genes in breast cancer. On the basis of one-class logistic regression (OCLR) machine learning algorithm, mRNAsi of samples was calculated. Kaplan-Meier (K-M) and Kruskal-Wallis (K-W) tests were utilized for the assessment of the connection between mRNAsi and clinicopathological variables of the samples. As for the analysis on the correlation between mRNAsi and immune infiltration, ESTIMATE combined with Spearman test was employed. The weighted gene coexpression network analysis (WGCNA) network was established by utilizing the differentially expressed genes in breast cancer, and the target module with the most significant correlation with mRNAsi was screened. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were conducted to figure out the biological functions of the target module. As for the construction of the prognostic model, univariate, least absolute shrinkage and selection operator (LASSO) and multivariate Cox regression analyses were performed on genes in the module. The single sample gene set enrichment analysis (ssGSEA) and tumor mutational burden were employed for the analysis on immune infiltration and gene mutations in the high- and low-risk groups. As for the analysis on whether this model had the prognostic value, the nomogram and calibration curves of risk scores and clinical characteristics were drawn. **Results.** Nine mRNAsi-related genes (CFB, MAL2, PSME2, MRPL13, HMGB3, DCTPP1, SHCBP1, SLC35A2, and EVA1B) comprised the prognostic model. According to the results of ssGSEA and gene mutation analysis, differences were shown in immune cell infiltration and gene mutation frequency between the high- and low-risk groups. **Conclusion.** Nine mRNAsi-related genes screened in our research can be considered as the biomarkers to predict breast cancer patients' prognoses, and this model has a potential relationship with individual somatic gene mutations and immune regulation. This study can offer new insight into the development of diagnostic and clinical treatment strategies for breast cancer.

1. Introduction

Breast cancer is a common threat to the women with increased annual incidences, and it has surpassed lung cancer ranking 1st on the global cancer-statistics list in 2020 [1]. Usually, the factors to assess the conditions of breast cancer patients lie in tumor stage, histological grade, and molecular subtype. However, when it comes to the prediction of patients' prognoses, they just do little help regarding the accuracy [2]. Prediction based solely on pathological fea-

tures is likely to cause inaccurate diagnosis of patient's prognosis. For one thing, low low-risk patients are likely to undergo unnecessary or excessive treatment. For another, the improper treatment tends to put high-risk patients at risk of cancer recurrence or metastasis [3]. For example, He et al. [4] explored SNP-related genes as novel prognostic markers for breast cancer, whose predictive performance for either disease-free survival or prognostic risk of patients is difficult to be realized by other clinicopathological characteristics. As a result, to explore novel biomarkers capable of

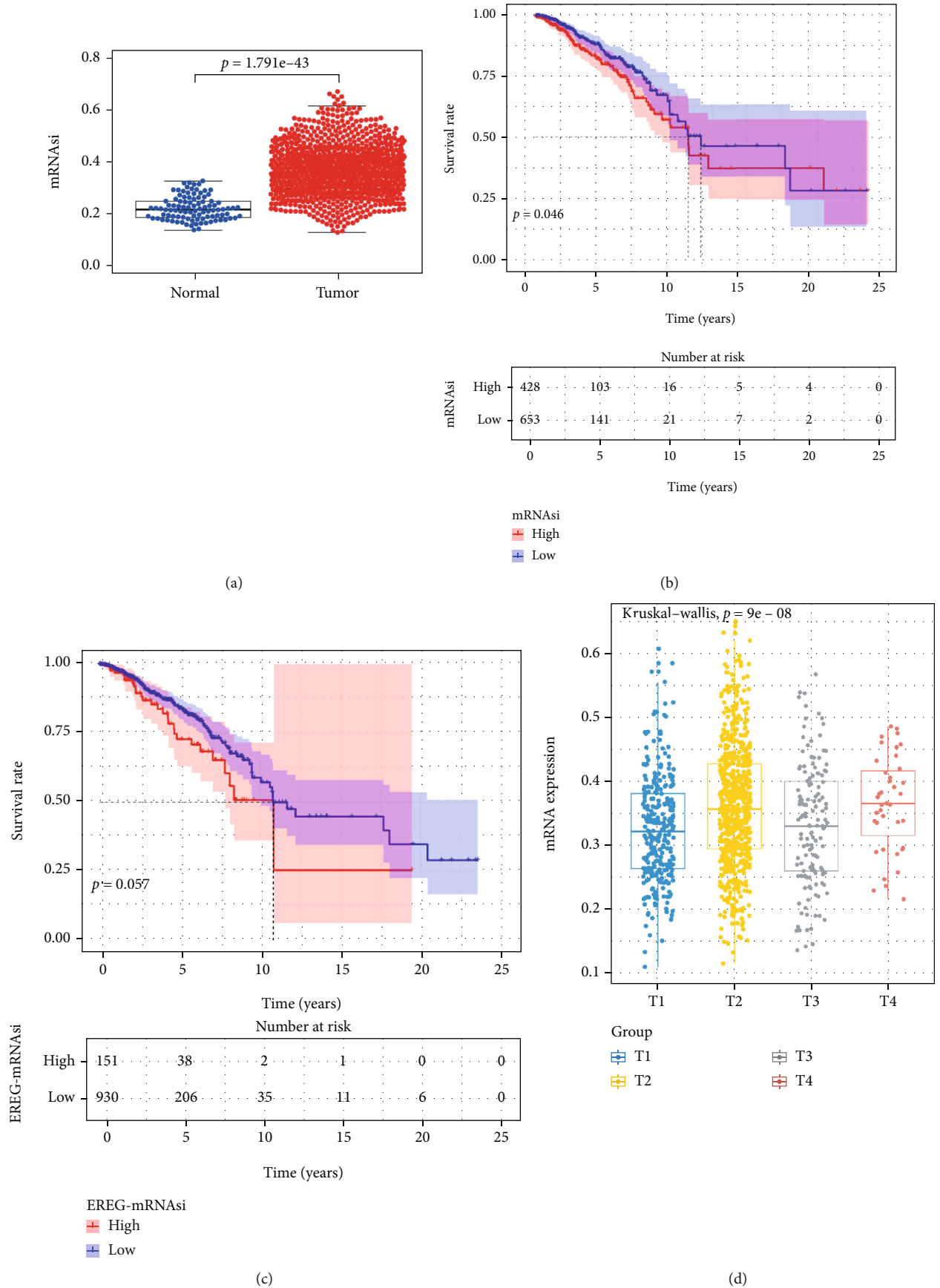
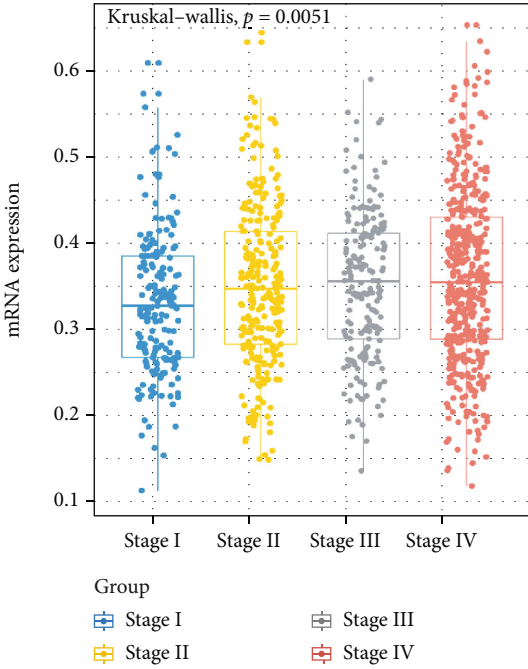
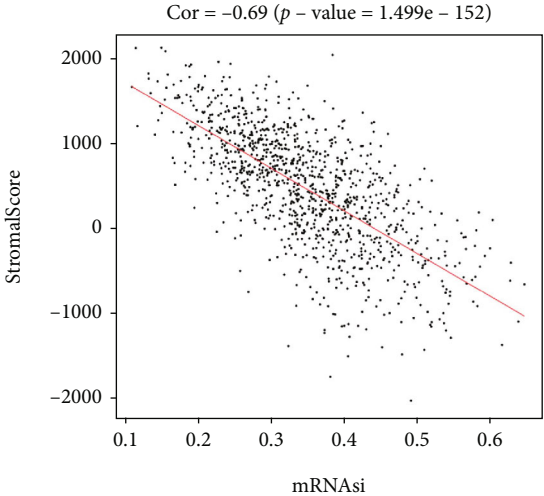


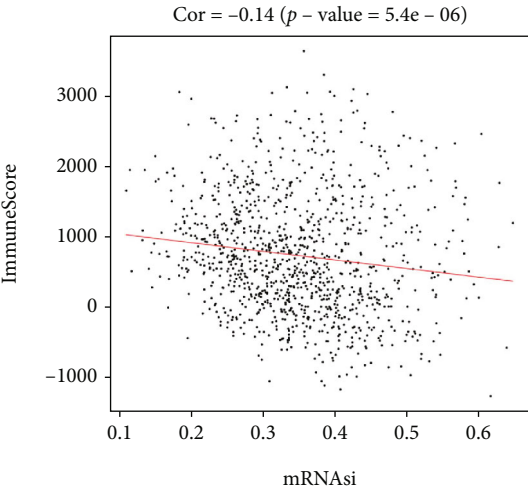
FIGURE 1: Continued.



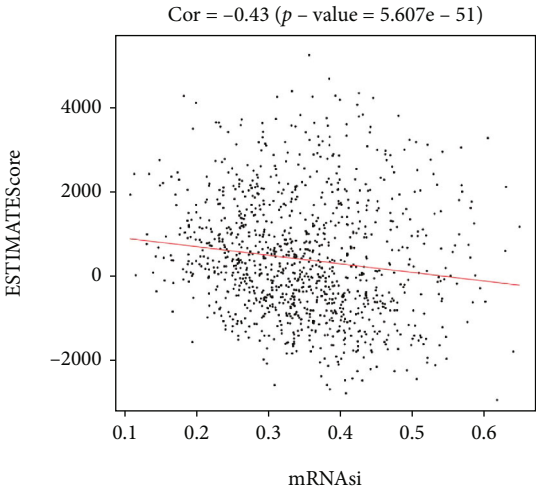
(e)



(f)



(g)



(h)

FIGURE 1: Continued.

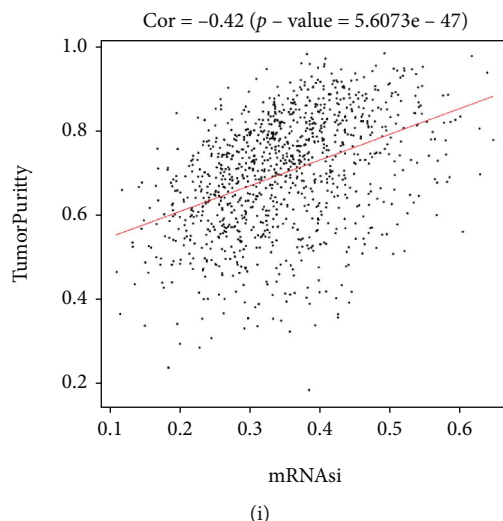


FIGURE 1: Association between mRNAsi of TCGA-BRCA and clinical features and tumor immune microenvironment. (a) Differences in mRNAsi between TCGA-BRCA tissues and healthy samples. (b) K-M survival curve for mRNAsi. (c) K-M survival curve of EREG-mRNAsi. Tumor growth sizes in TCGA-BRCA tissue samples (d) and difference of mRNAsi expression in different clinical stages (e). (f-i) Correlation analysis between mRNAsi and stromal score (f), immune score (g), ESTIMATE score (h), and tumor purity (i) assessed by ESTIMATE algorithm.

predicting breast cancer patient's prognosis is of great significance for treating patients with more precise therapeutic strategies.

The complexity and diversity in tumor microenvironment are beyond our imagination. Some cells are responsible for tumor initiation, metastasis, and recurrence in tumor microenvironment, but others are not. For a few cells with stemness features and invasiveness, they can trigger the development of tumor and invade human immune system, inducing innate resistance to external killing [5, 6]. These cells are named as cancer stem cells (CSCs), which have the features of continuous proliferation, self-renewal, and multidirectional differentiation [4]. Breast cancer stem cells can induce various primary tumors, facilitating the development and metastasis of tumors, resulting in a poor prognostic response in breast cancer patients [7]. Notably, multiple CSC-associated breast cancer molecular markers have been identified, such as CD44, CD24, ALDH1, PROCR, and MUC1 [8, 9]. Among them, CD44, CD24, and ALDH1 are capable of predicting the prognoses of triple-negative breast cancer patients, which can be predictive markers for cancer recurrence, distant metastasis, disease-free survival, and overall survival [10, 11]. Revealing breast cancer prognostic markers from the perspective of CSCs may be an important entry point.

Given the important regulation of CSC properties for tumor progression, existing studies have established a new method to describe CSCs through machine learning algorithms capable of quantifying the differentiation phenotype during cancer progression and the development characteristics of stem cell populations in tumor tissues [12]. For the identification of various stem cells and tumor cells, the one-class logistic regression (OCLR) machine learning algorithm is a great choice utilized to extract the expression profiles of these cells [13]. The algorithm has been applied to the

genome-wide expression data of enormous TCGA samples and successfully quantified the differentiation degree of various cancers of the breast cancer, lung cancer, glioma, and so on, as well as the stemness features and tumorigenicity of paired healthy tissues. Finally, a new stemness feature mRNAsi was proposed [13]. mRNAsi is a cancer stem cell index describing the similarity degree between tumor and stem cells, which can be considered a quantification of cancer stem cells [13]. The values of mRNAsi range from 0 to 1, and it has a close connection with the tumor dedifferentiation level and biological processes of CSCs [14, 15]. mRNAsi has been verified as an indicator of survival, classification, and disease progression in cancer patients [15–17]. Above-mentioned studies have paved the way for us to dive deeper into the mechanisms of breast cancer stem cells and the mining of prognostic molecular markers. The epigenetically regulated mRNA expression-based stemness index (EREG-mRNAsi) is obtained by training the expression level of genes associated with the epigenetically regulated stem cells. The index ranges from 0 to 1. The closer the index value is to 1, the lower the degree of cell differentiation and the stronger the stemness features, reflecting the degree of dedifferentiation of cancer cells [18, 19].

Our study initially determined the mRNAsi of TCGA-BRCA dataset samples, predicted tumor purity, and abundance of stromal cells and immune cells within the tumor and analyzed the correlation of mRNAsi with immune infiltration. The target gene module associated with mRNAsi was screened by weighted gene coexpression network (WGCNA). Next, a bioinformatics analysis on the target module revealed 9 mRNAsi-related genes that were capable of predicting breast cancer patient's prognosis, and a prognosis-assessing model was hence established. Subsequently, the study revealed the complex role of prognostic signature genes with somatic gene mutations and immune

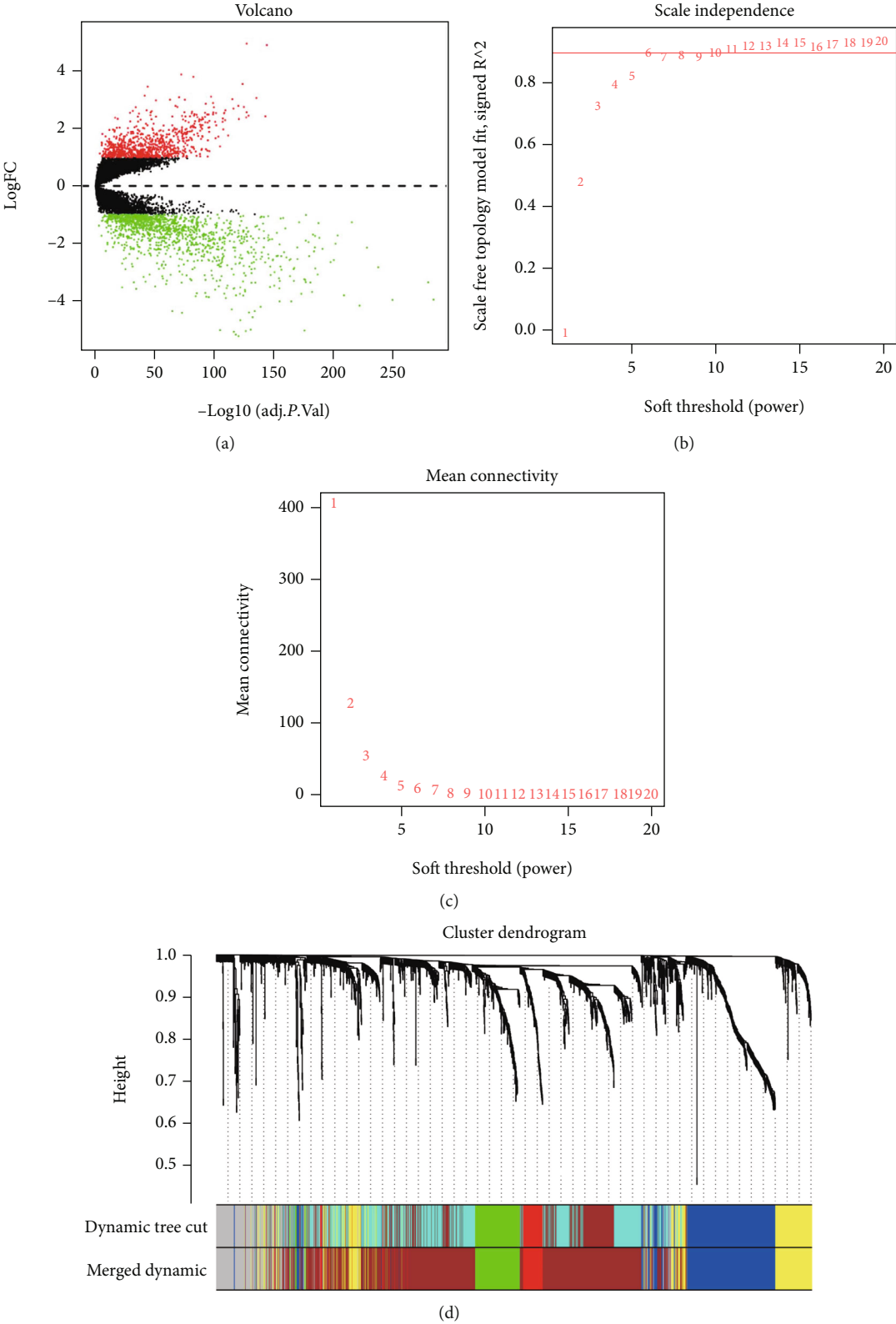


FIGURE 2: Continued.

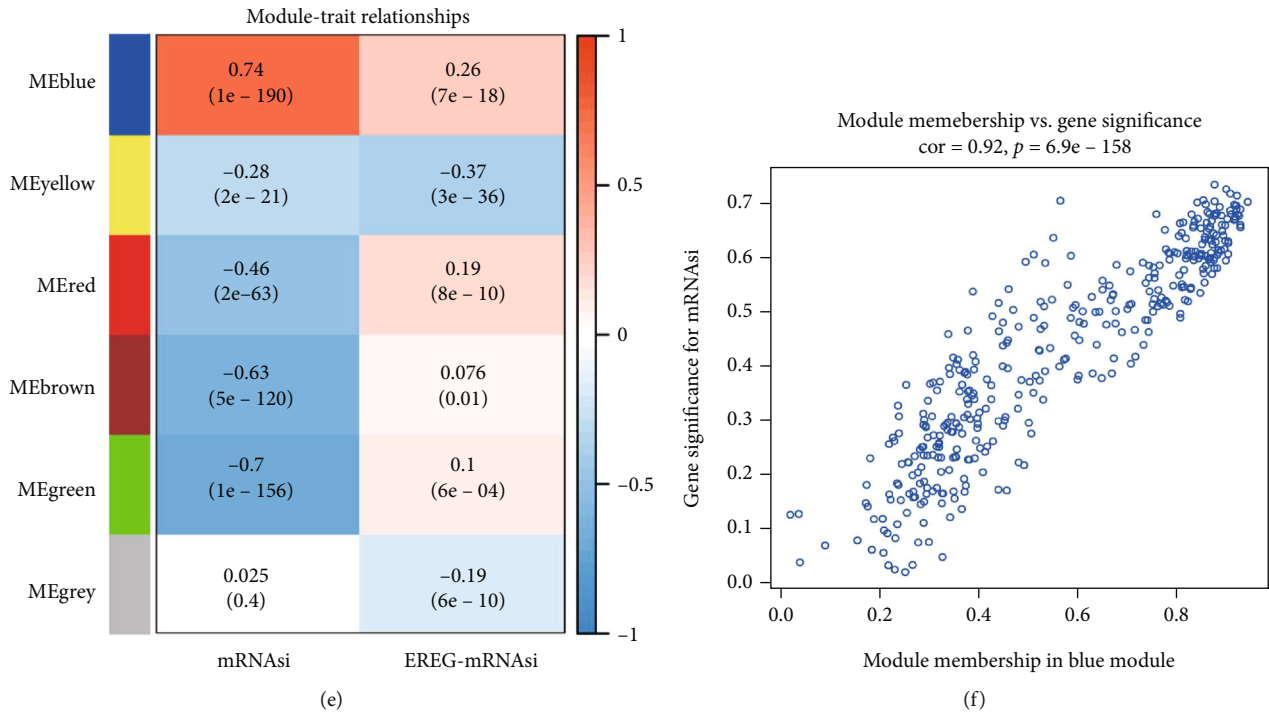


FIGURE 2: Construction of weighted gene coexpression network of TCGA-BRCA samples. (a) Volcano plot showed the distribution of DEGs in breast cancer tumor tissue relative to normal breast tissue; red indicated upregulated genes, green indicated downregulated genes, and black indicated genes excluded by DEG screening criteria. (b) Scale-free topological model fit index screening. (c) Average connectivity of soft threshold of adjacency matrix. (d) Identification of breast cancer coexpressed gene modules; different colors represent different gene modules. (e) Heat map of correlation between gene modules and mRNA_{asi} score or EREG-mRNA_{asi}. (f) Scatter plot of blue gene modules; each circle represents a gene.

cell infiltration, providing a reference for the expansion of the prediction field of prognostic models. To sum up, the risk assessment model constructed in our study was able to effectively predict the prognosis of patients with breast cancer. Besides, the connection between 9 mRNA_{asi}-related genes and somatic gene mutations as well as immune regulation was revealed in our study. These mRNA_{asi}-related genes can be applied as biomarkers with great value in clinical practice like predicting prognoses of breast cancer patients.

2. Materials and Methods

2.1. Breast Cancer Sample Data Collection. From TCGA database (<https://portal.gdc.cancer.gov/>), breast cancer RNA expression data, gene mutation data, and corresponding clinical data were obtained as training sets, involving 1109 breast cancer samples and 113 healthy breast samples. From the EGB (<http://asia.ensembl.org/index.html>), the GTF annotation file was acquired. From the GEO library (<https://www.ncbi.nlm.nih.gov/geo/>), the breast cancer sample expression profile GSE42568 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42568>) was downloaded as the validation set. The mRNA_{asi} of samples was calculated by OCLR [13] for the comparison of the mRNA_{asi} differences between the normal and tumor groups.

2.2. Correlation between Stemness Index of mRNA_{asi} and Clinicopathological Variables and Immune Infiltration.

Overall survival was compared between different mRNA_{asi} samples by Kaplan-Meier (K-M) analysis according to the optimal threshold. The R package ggpubr (<https://cran.r-project.org/web/packages/ggpubr/index.html>) was employed for comparing mRNA_{asi} in the context of clinical characteristics. The Kruskal-Wallis (K-W) test was employed for assessing the connection between mRNA_{asi} and clinical characteristics. Based on the gene expression profiles of breast cancer samples, ESTIMATE was utilized to generate immune, stromal, and ESTIMATE scores, as well as tumor purity. The correlation analysis on mRNA_{asi} and these scores and tumor purity were achieved by Spearman's test, and p values were calculated.

2.3. WGCNA. FPKM data from TCGA-BRCA were identified for differentially expressed genes (DEGs) utilizing the R package limma [20] ($|\log_2FC| > 1, FDR < 0.05$). On the basis of these DEGs, the R package WGCNA was utilized for the analysis of the Gene modules [21], and the specific processes were as follows: genes with missing values were removed using the goodSamplesGenes function, tumor samples were clustered, outliers were removed, and 100 was set as a cut line. The coexpression network was constructed by setting 6 as the optimal soft threshold. Then, by transforming the adjacency matrix into a TOM matrix, the genetic connectivity of the network was detected. Next, the average linkage hierarchical clustering was performed on the basis of the differences in TOM. By employing a dynamic shearing

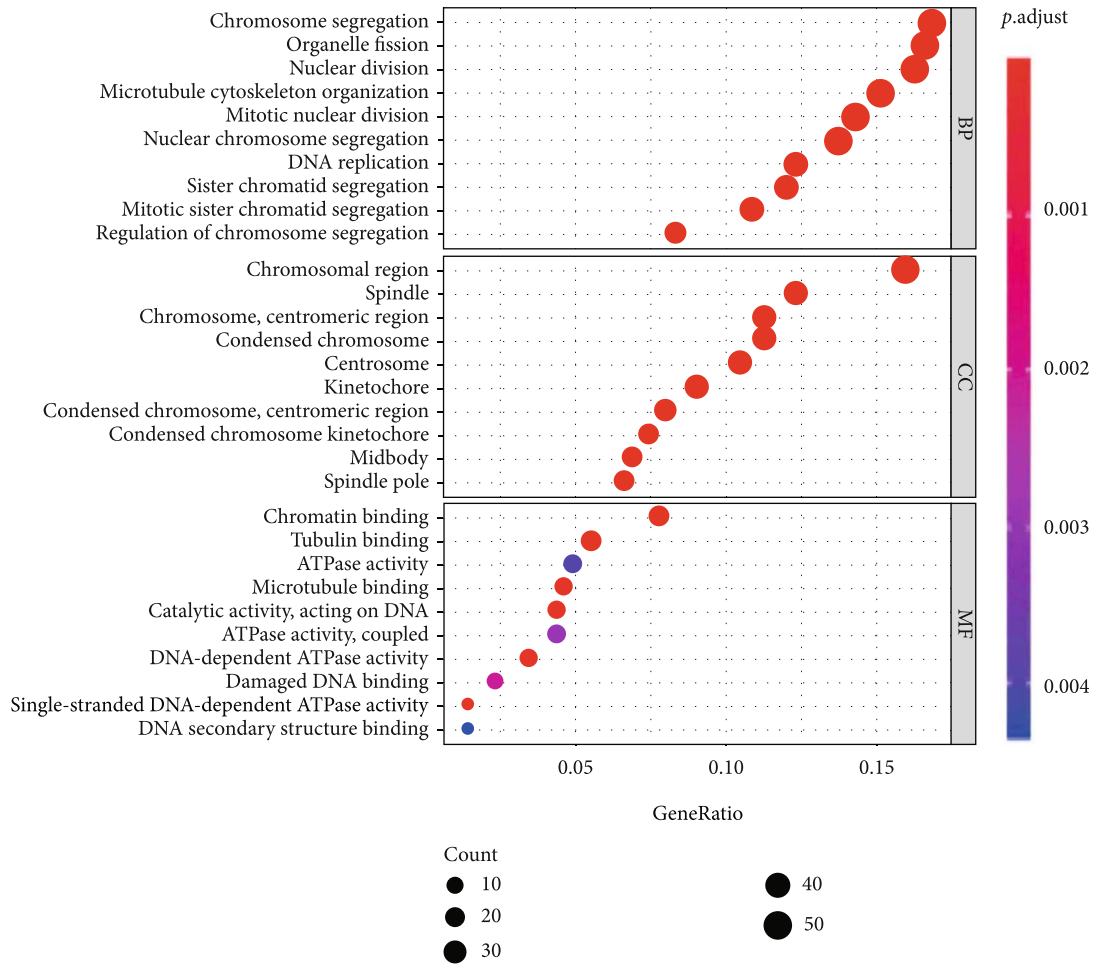


FIGURE 3: Continued.

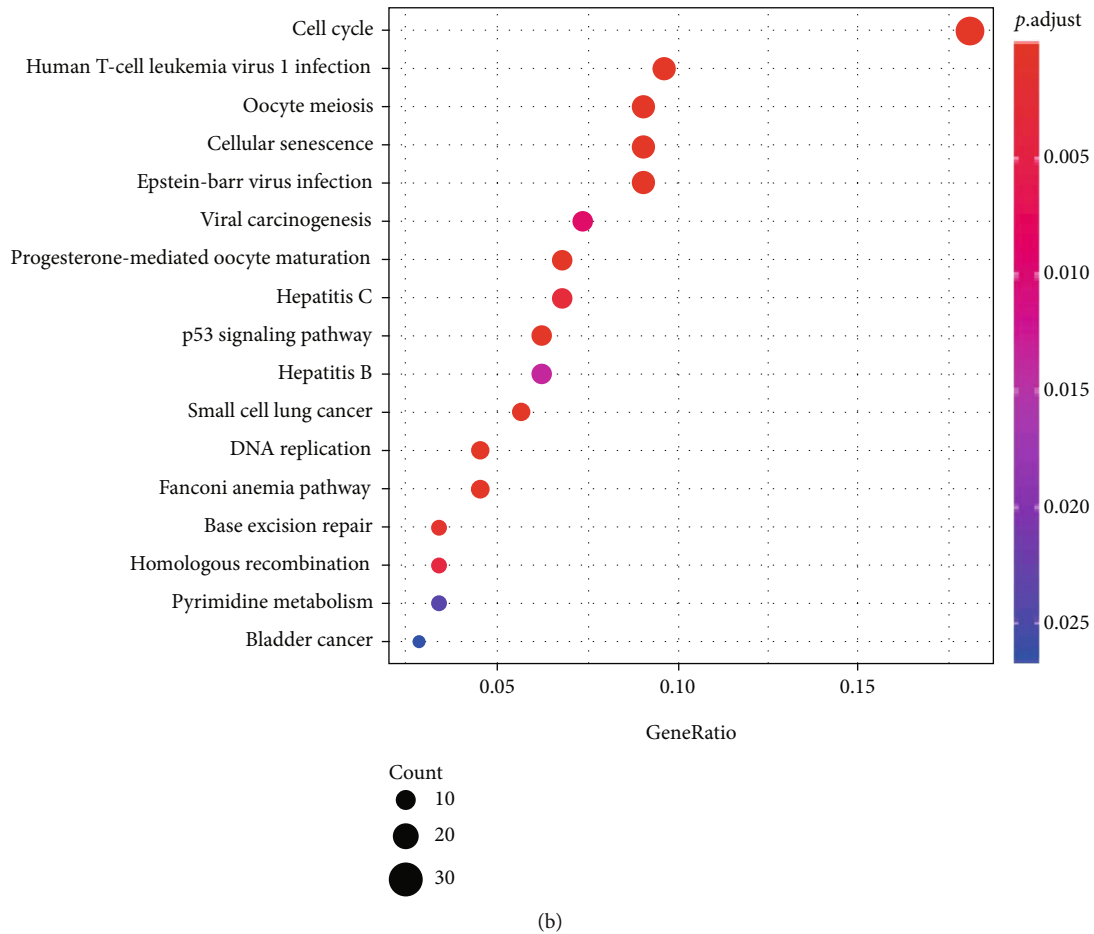


FIGURE 3: Blue module gene function annotation. (a) GO analysis of genes in the blue module. (b) KEGG pathway enrichment analysis of genes in the blue module; the color of the bubble represents the p value, and the size of the bubble represents the number of genes involved in the pathway.

approach, the gene tree was then divided into different modules. And the minimum number of genes in each module was set to 50, and MEDisThres was set at 0.25 to cluster and merge similar modules.

2.4. Gene Function Annotation of Gene Module. In each gene module, there was a primary component, namely, module eigengenes (MEs) which could represent all genes within the module. To mine the gene modules associated with mRNAsi of tumor samples, the ME of each module was calculated separately with the mRNAsi of the samples for correlation coefficient, and the gene modules highly associated with mRNAsi were retained as the target modules. The R package clusterProfiler [22], enrichplot (<https://bioconductor.org/packages/release/bioc/html/enrichplot.html>), and ggplot2 (<https://ggplot2-book.org/>) were utilized for the annotation and visualization of KEGG and GO pathways.

2.5. Construction and Validation of the Prognostic Model. The R package survival (<https://cran.r-project.org/web/packages/survival/index.html>) was used to perform univariate Cox regression analysis on genes in the target module to identify genes that have a close connection with patient's overall survival rate ($p < 0.01$). The R package glmnet [23]

and survival were utilized for the conduction of LASSO analysis, which was combined with multivariate Cox analysis to further screen genes and risk coefficients remarkably linked to prognosis, thus to construct a risk model. Data from TCGA-BRCA was classified as high- and low-risk groups taking the median risk score as a demarcation. Differences in mRNAsi between the two groups were analyzed employing the R package ggpubr, and K-M curves and ROC curves were plotted employing the R package survival. At last, the risk score, survival state plots, and gene expression heat map of the two risk groups was plotted.

2.6. Analysis of the Correlation between Prognostic Models and Tumor Immunity and Gene Mutations. The R package GSEABase with 29 immune-related features [24] was employed for the conduction of ssGSEA analysis of genes in the prognostic risk assessment model. By utilizing the R package heat map, the antitumor immune-enrichment results of the high- and low-risk groups were visualized. In gene mutation analysis, R package maftools [25] was utilized for analyzing the tumor mutational burden, and R package GenVisR [26] was utilized for analyzing the differences in gene mutation types and mutation numbers of the samples.

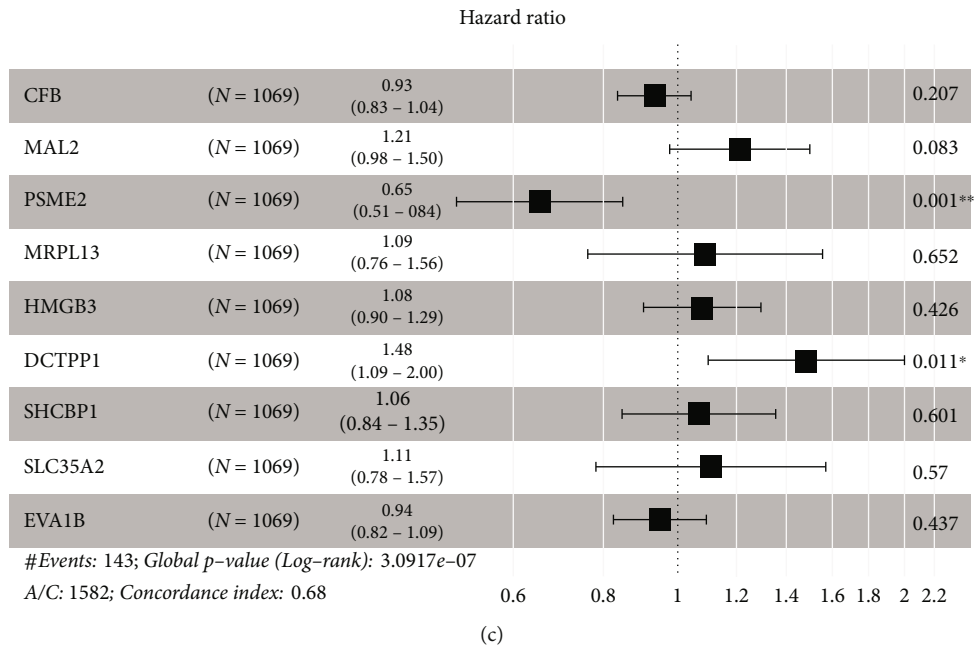
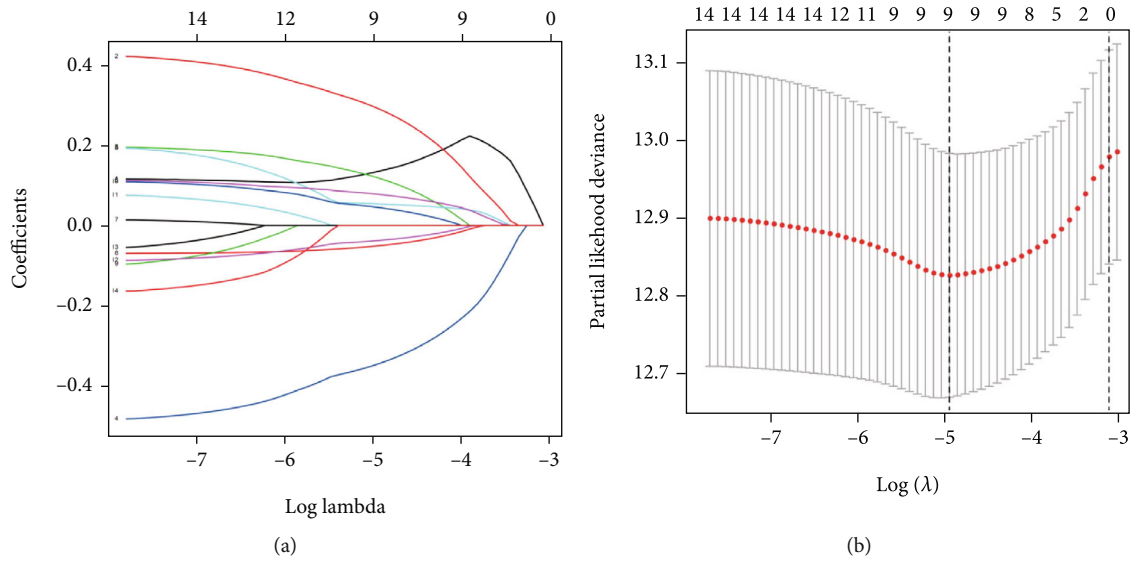
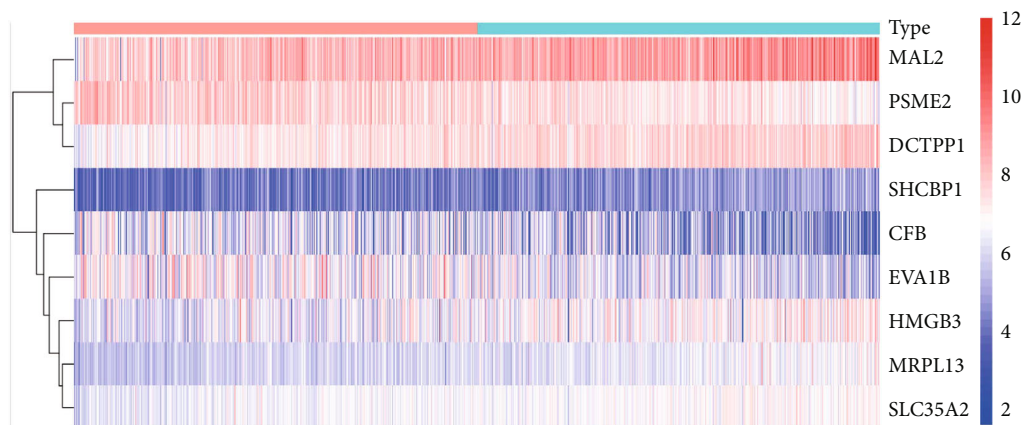
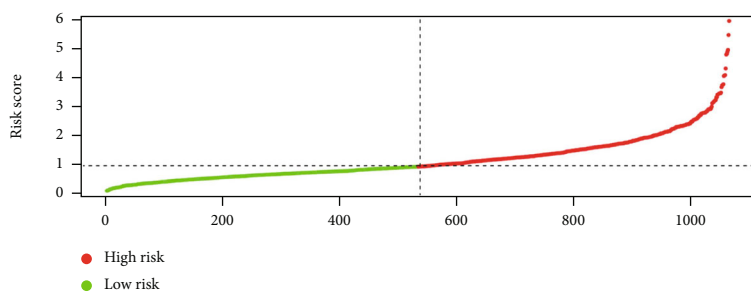


FIGURE 4: Continued.

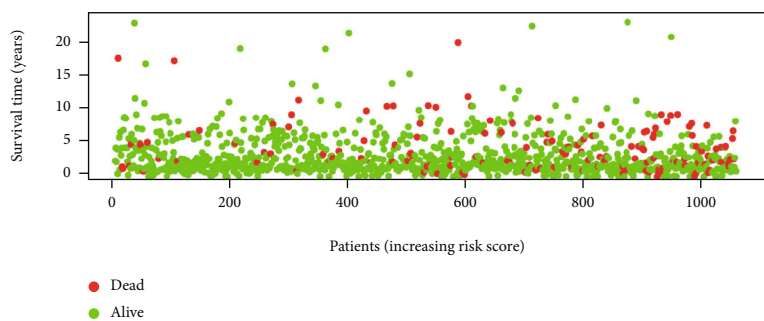


Type
 High
 Low

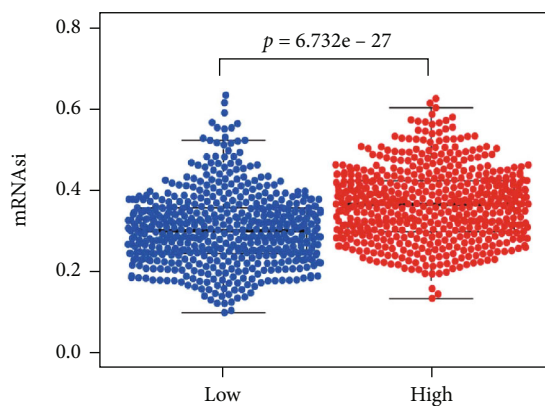
(d)



(e)



(f)



(g)

FIGURE 4: Continued.

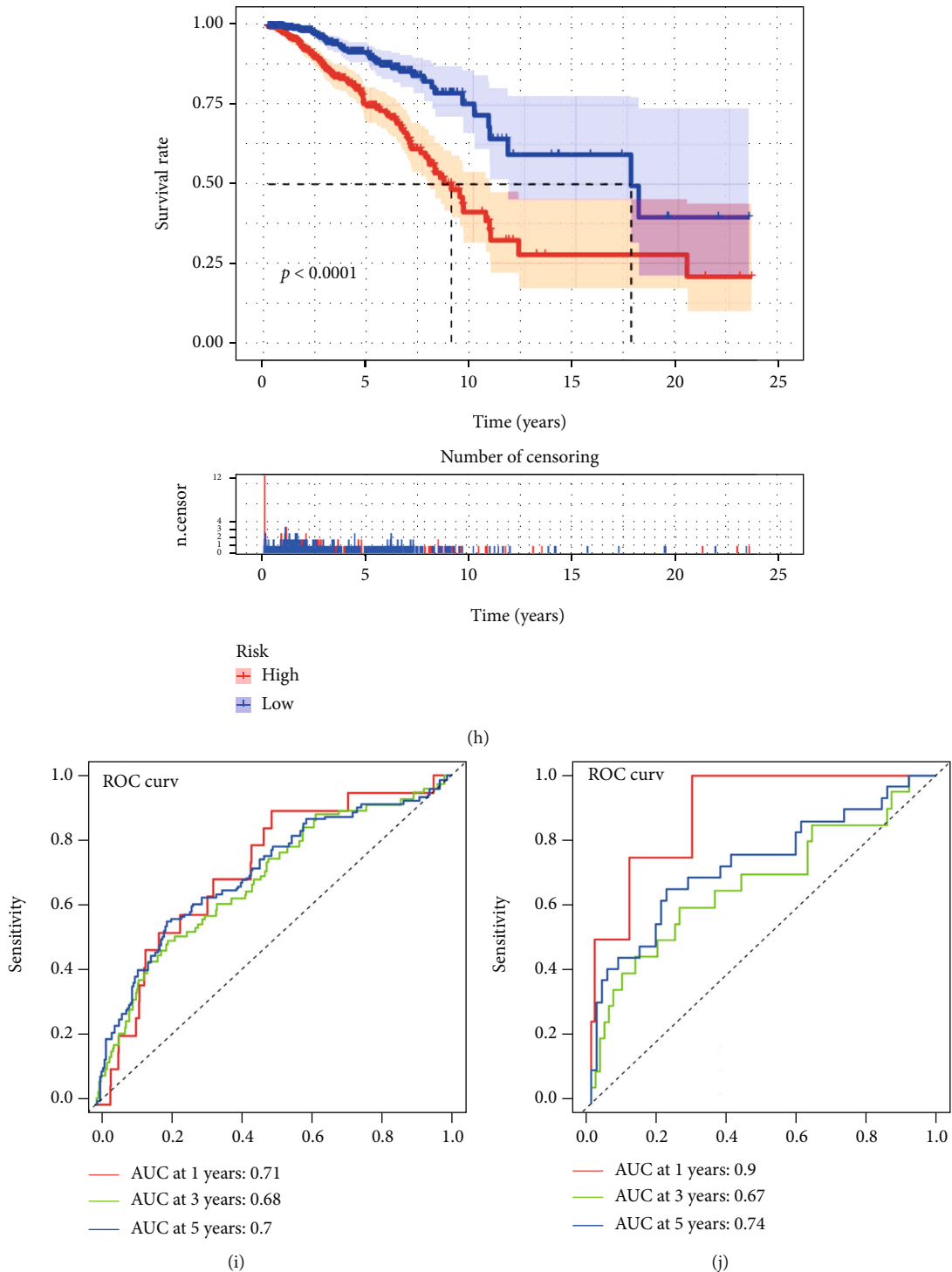


FIGURE 4: Establishment of the mRNAi-based prognostic model. (a) LASSO coefficient distribution of 14 prognosis-related genes. (b) Partial likelihood deviation calculated from LASSO regression cross-validation plotted as a function of $\log(\lambda)$. (c) Multivariate Cox analysis of 9 mRNAi-related genes. (d) Heat map of gene expression for patients in high- and low-risk groups. (e) Risk score plot for patients in high- and low-risk groups. (f) Survival state diagram of patients in high- and low-risk groups. (g) Differential analysis of mRNAi in patients in high- and low-risk groups. (h) K-M survival curves for patients in the high- and low-risk groups. (i) ROC curve of TCGA-BRCA sample to assess the predictive performance of the risk signature for 1-, 3-, and 5-year overall survival in the training set. (j) ROC curve of the GEO database GSE42568 dataset sample, used to assess the predictive performance of the 1-, 3-, and 5-year overall survival risk signature in the validation set.

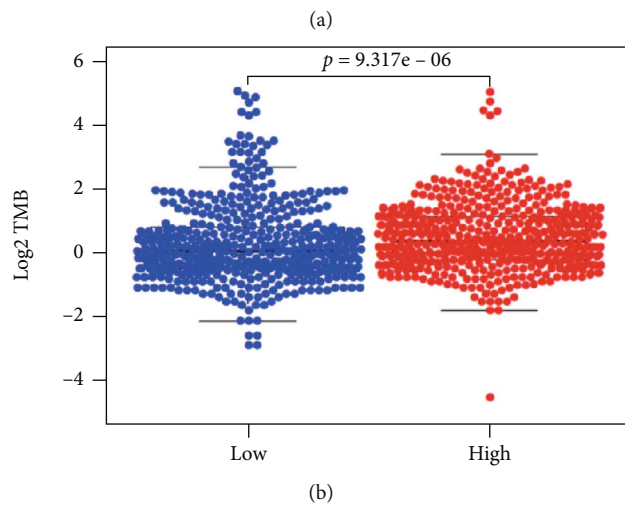
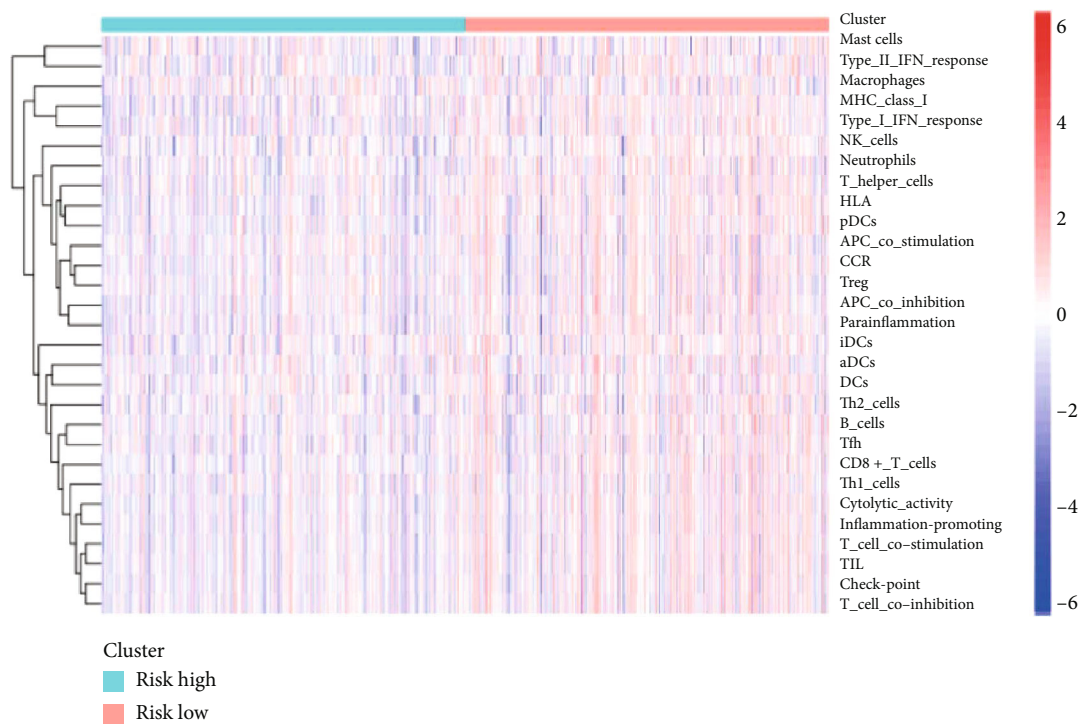
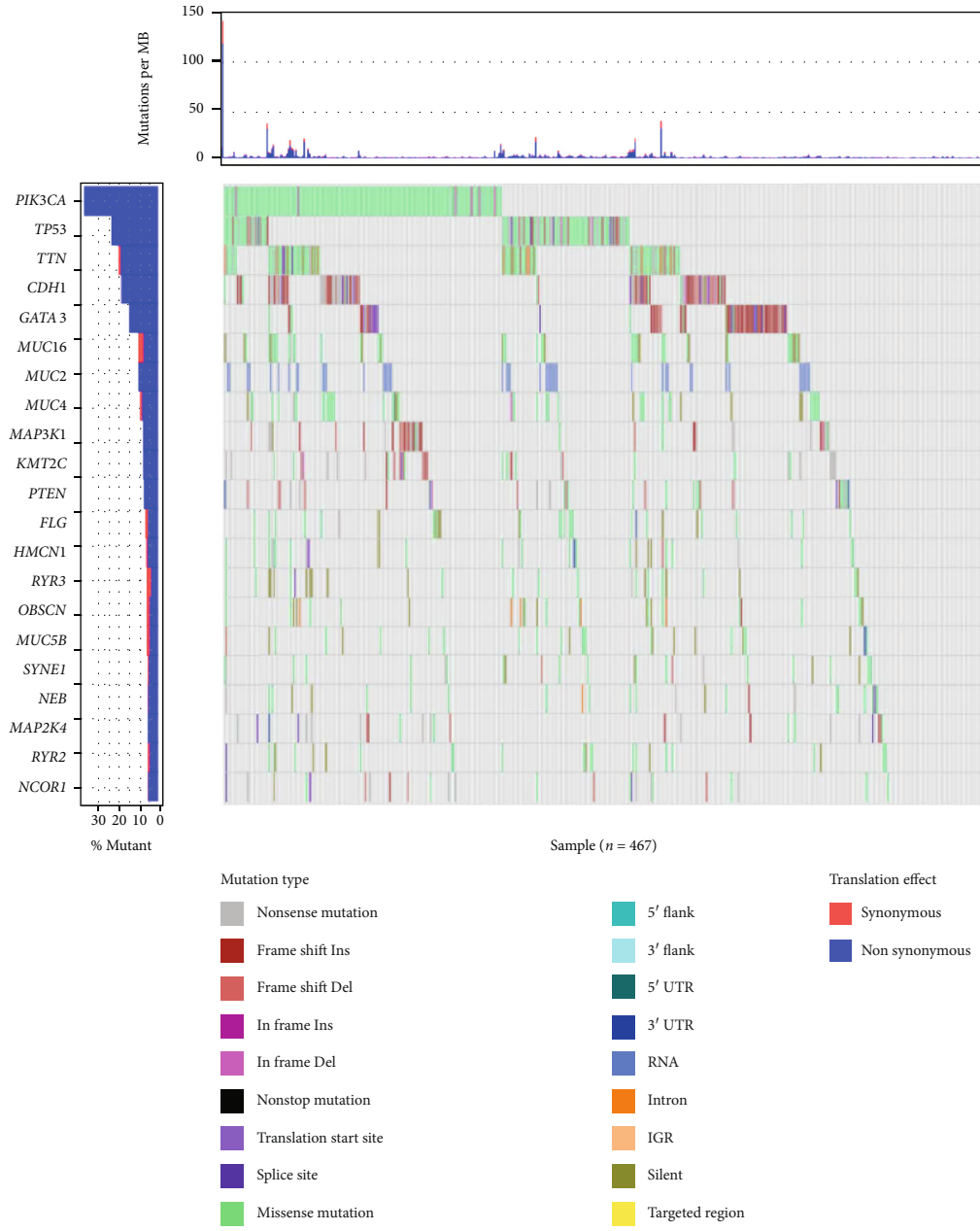


FIGURE 5: Continued.



(c)

FIGURE 5: Continued.



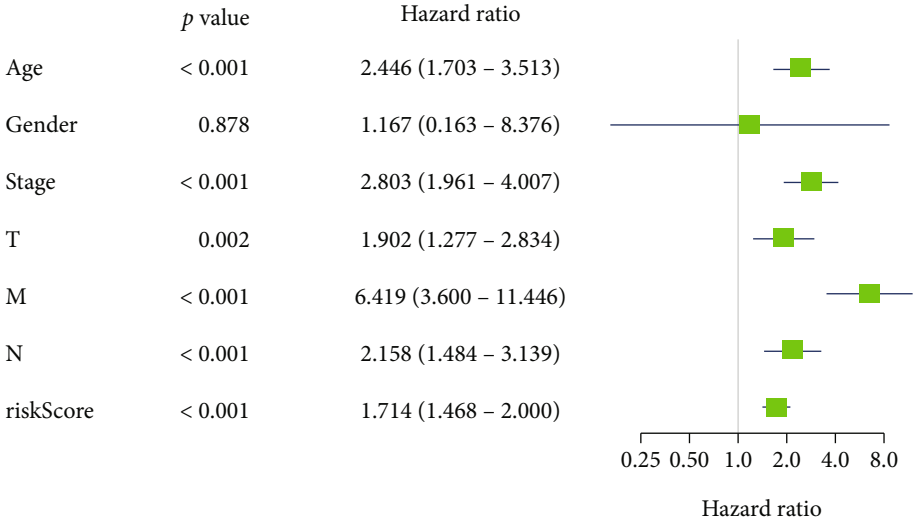
(d)

FIGURE 5: Immune infiltration and gene mutation revelation in samples with different risk scores. (a) ssGSEA was used to infer the level of immune-infiltrating cells in the gene set of breast cancer samples from the high-risk and low-risk groups. (b) TMB differences between patients from the high-risk and low-risk groups. (c) Distribution of significantly mutated genes in breast cancer samples with mutations in the high-risk group. (d) Distribution of significantly mutated genes in breast cancer samples with mutations in the low-risk group, with different colors on the right side as different mutation types.

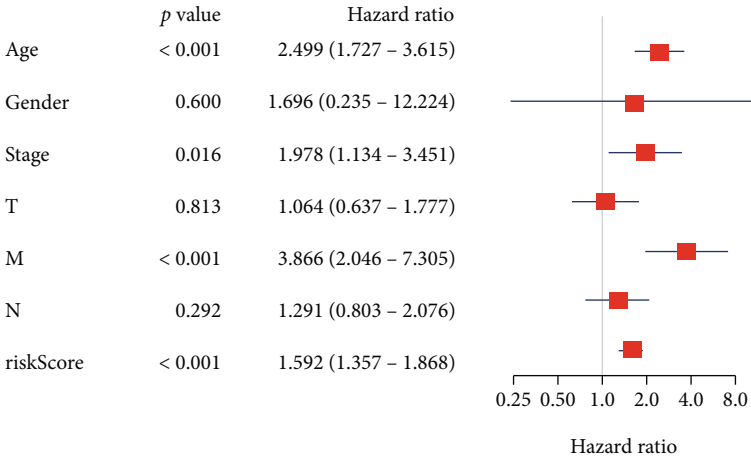
2.7. *Prognostic Model Validity Assessment.* In this study, univariate and multivariate Cox regression analyses of risk score, age, gender, pathological stage, and other clinical characteristic parameters were conducted utilizing the R package survival in TCGA-BRCA or GSE42568 datasets. The R packages rms, regplot, tibble, and survival were used to draw nomograms according to the risk score and 6 clinicopathological factors. Calibration curve was drawn to predict the consistency between nomogram-predicted 1-, 3-, and 5-year survival and actual survival of patients.

3. Results

3.1. *Breast Cancer mRNasi Is Closely Bound Up with the Clinical Characteristics of Patients and Immune Regulation of Cancer Tissues.* mRNasi can reflect the similarity between tumor cells and stem cells. In this study, differential analysis of breast cancer samples with mRNasi data and corresponding healthy breast samples demonstrated that mRNasi was substantially upregulated in breast cancer tissues (Figure 1(a)). Samples were classified into



(a)



(b)

FIGURE 6: Continued.

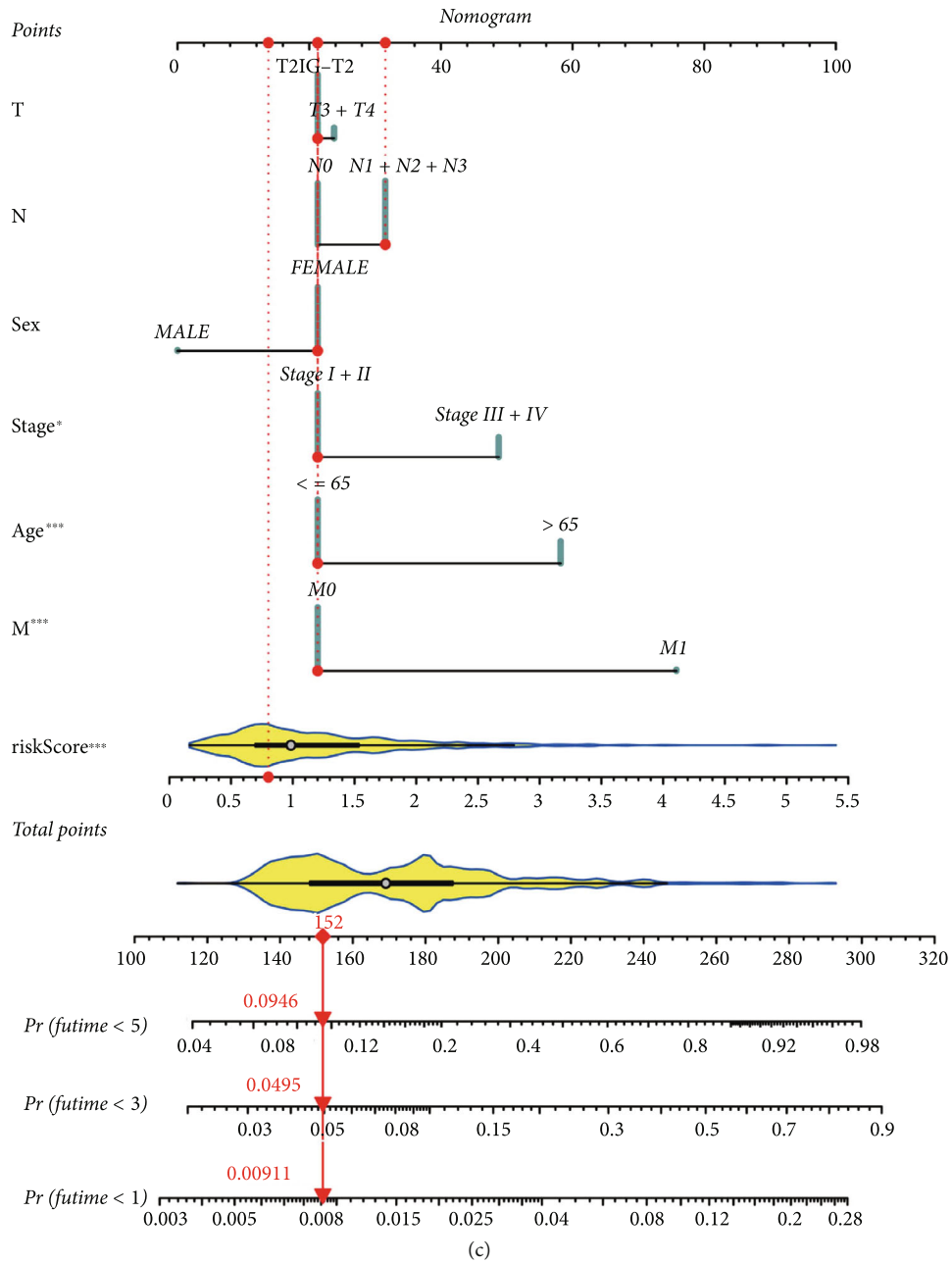


FIGURE 6: Continued.

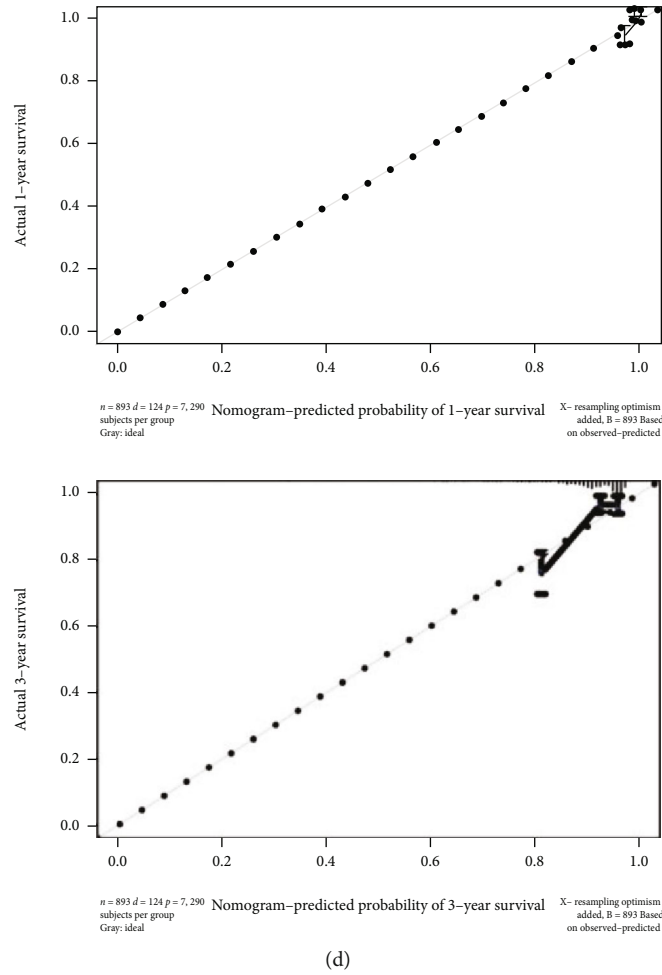


FIGURE 6: Construction and evaluation of nomogram. (a) Univariate Cox regression analysis of risk score and clinicopathological characteristics. (b) Multivariate Cox regression analysis of risk score and clinicopathological characteristics. (c) Nomogram constructed from risk score and clinicopathological characteristics to predict 1-, 3-, and 5-year survival rates of patients in the training cohort. (d) Calibration curve depicted the agreement between nomogram predicted 1-, 3-, and 5-year survival rates of patients and actual survival rates.

high- and low-risk groups according to the optimal threshold of mRNasi. Survival analysis illustrated that a higher mRNasi index in patients led to a poorer prognostic survival rate with a contrast to that of patients with a lower mRNasi index (Figure 1(b)). Patients with higher EREG-mRNasi had poorer overall survival compared with patients with lower EREG-mRNasi (Figure 1(c)). Correlation analysis between mRNasi and clinicopathological characteristic variables of breast cancer showed that mRNasi expression did not significantly change with tumor growth (Figure 1(d)) but was significantly increased with the progression of pathological stage (Figure 1(e)). Given the potential role of mRNasi in the antitumor immune process, this study speculated that mRNasi added much diversity to the tumor immune micro-environment. Therefore, the construction of correlation between the ESTIMATE assessment results of TCGA-BRCA tissues with the mRNasi of the samples told us that mRNasi was negatively correlated with the stromal score, immune score, and ESTIMATE score of tumor tissues. But a positive correlation was found in mRNasi with tumor purity (Figures 1(f)–1(i)). The finding revealed a close con-

nection between breast cancer mRNasi and the clinical characteristics of patients and the immune regulation of cancer tissues. This index was worthy of inclusion in subsequent studies to reveal its biological function.

3.2. Identification of mRNasi-Related Modules. In view of the significant difference in mRNasi between normal and tumor tissues, we first screened DEGs from the mRNA level to elucidate differences in mRNasi, which were visualized in a volcano plot (Figure 2(a)). To dig out key mRNasi-related genes, WGCNA was conducted for the construction of a coexpression network of mRNAs for TCGA-BRCA. The index of the scale-free topology was taken to reach 0.90 (Figures 2(b) and 2(c)). By using a dynamic tree pruning algorithm (module size = 50), genes with similar expression patterns were introduced into the same module to form a hierarchical clustering tree with modules. According to the weighted correlation and the set criteria, hierarchical clustering analysis was performed, and clustering results were segmented (Figure 2(d)). Six gene modules were finally identified, and correlation analysis of MEs with mRNasi

and EREG-mRNasi in each module revealed that the blue module presented the highest correlation with cell stemness index mRNasi ($r = 0.74$, $p = 1e - 190$) (Figure 2(e)). As shown in Figure 2(f), among the blue module genes, the closer module membership value is to 1 indicates that the gene is more strongly correlated with this module. The higher value of gene significance for mRNasi indicates that mRNasi is more correlated with the gene in the module. As a result, we adopted the blue module with 385 genes as the target module for subsequent studies.

3.3. Gene Function Annotation of Target Module. To investigate how mRNasi-related genes and pathways functioned biologically, GO functional annotation and KEGG pathway enrichment analyses were performed on 385 genes from the target blue module. GO functional annotation results indicated that these genes were primarily bound up with functions including chromosome segregation, organelle fission, and nuclear division (Figure 3(a)). KEGG pathway enrichment analysis suggested that the involvement of these genes was found in cell cycle, human T cell leukemia virus 1 infection, and oocyte meiosis (Figure 3(b)). GO and KEGG results showed that genes of the blue module were mainly enriched in signaling pathways associated with cell cycle, T cell leukemia virus, and oocyte division, which are closely related to cancer development.

3.4. Establishment of the mRNasi-Based Prognostic Model. Firstly, the prognostic effect of the genes in the blue module was assessed by the univariate Cox regression analysis. Then, 9 candidate feature genes were screened by the LASSO Cox regression analysis under the optimal value of λ (Figures 4(a) and 4(b)). The risk assessment model was finally constructed based on 9 genes, through multivariate Cox regression analysis (Figure 4(c)). Risk score = $-0.0725 * \text{CFB} + 0.1894 * \text{MAL2} - 0.4245 * \text{PSME2} + 0.0826 * \text{MRPL13} + 0.0736 * \text{HMGB3} + 0.3917 * \text{DCTPP1} + 0.0628 * \text{SHCBP1} + 0.1012 * \text{SLC35A2} - 0.0566 * \text{EVA1B}$.

TCGA-BRCA samples were divided into high- and low-risk groups by setting the median risk score as the cutoff value, and the heat map showed the expression levels of 9 mRNasi-related genes (Figure 4(d)). The distribution of risk score and survival time among samples in TCGA dataset showed that as risk score increased, the mortalities from cancer also mounted and the survival time decreased (Figures 4(e) and 4(f)). Differential analysis of mRNasi demonstrated that patients in the high-risk group had markedly higher mRNasi than those in the low-risk group ($p = 6.732e - 27$) (Figure 4(g)). Survival analysis demonstrated that high-risk group patients had a remarkably lower overall survival rate than low-risk group patients ($p < 0.001$) (Figure 4(h)). ROC curves demonstrated that the AUC values of the risk assessment model for predicting 1-year, 3-year, and 5-year survival of TCGA dataset samples were 0.71, 0.68, and 0.70, respectively (Figure 4(i)). The AUC values of the model for predicting 1-year, 3-year, and 5-year survival of GSE42568 dataset samples were 0.9, 0.67, and 0.74, respectively (Figure 4(j)). It was shown that the risk score for constructing a risk assessment model based on the 9 mRNasi-

related genes obtained from TCGA-BRCA dataset had predictive potential for breast cancer patients.

3.5. Immunological Infiltration and Gene Mutation Revelation in High- and Low-Risk Groups. We inferred the immune cell infiltration level in the breast cancer gene set by ssGSEA, and the expression level of immune gene set in the low-risk group was higher compared with the high-risk group (Figure 5(a)). Simultaneous tumor mutation burden (TMB) analysis showed that TMB values demonstrated higher in high-risk patients ($p = 9.3e - 06$) (Figure 5(b)). Subsequently, further mutation gene analysis demonstrated that the high-risk group samples had a much higher gene mutation frequency than the low-risk group samples (Figures 5(c) and 5(d)). There are differences in genetic variants between high- and low-risk groups, contributing to the difference in patient prognosis or immune cell infiltration.

3.6. Construction and Evaluation of the Nomogram. Univariate Cox analysis of risk score and other pathological features in TCGA dataset showed that age, pathological stage, distant tumor metastasis (M), lymph nodes metastasis (N), and risk score were all bound up closely with the prognosis of breast cancer patients, with a HR of 1.714 ($p < 0.001$) for risk score (Figure 6(a)). Multivariate analysis demonstrated that the HR of risk score was 1.592 ($p < 0.001$) (Figure 6(b)), indicating that risk score could be used as a prognostic factor independent of clinical characteristics. The nomogram plotted in combination with risk score, T, N, M stage, sex, age, and stage was used to predict the overall survival rate at 1, 3, and 5 years in patients with breast cancer (Figure 6(c)), corresponding to a better fit of the calibration curve (Figure 6(d)), demonstrating that this nomogram had a favorable predictive ability.

4. Discussion

CSCs have gained much attention in the cancer-related research. The intensive findings about CSCs have enriched our understandings of cancer development, thus propelling us to explore novel effective therapeutic strategies for combating cancer [10, 27]. mRNasi can reflect stemness in cancer patients. With the help of computational biology and bioinformatics, mRNasi can be used efficiently for mining genes related to tumor stemness [12, 13]. Since then, there have been a number of studies applying mRNasi to cancer prognosis. For example, it has been shown that mRNasi expression in hepatocellular carcinoma increases with tumor pathological grade, and mRNasi established from gene expression data has a deep connection with poor overall survival of hepatocellular carcinoma patients [28]. In glioblastoma, the mRNasi index of cancer tissue can be used to distinguish glioblastoma subtypes, and there is a marked difference in the prognostic overall survival rate of patients with each subtype [29]. The above reports all provide an important reference for the construction of predictive prognostic model for breast cancer based on mRNasi.

Our study first established a correlation between TCGA-BRCA tissue assessment results and sample mRNasi, and

differentially expressed mRNAs were then obtained. Based on WGCNA mining the target modules closely related to mRNAsi, GO functional annotation and KEGG analyses of the genes in this module showed that they were mainly associated with functions such as chromosome segregation, organelle fission, and mitosis and were involved in cell cycle, human T cell leukemia virus 1 infection, and oocyte division pathways. It has been found in breast cancer, colon cancer, and ovarian cancer that the cell cycle mainly regulates specific transcription dependent on cell cycle genes in cancer [30–32]. Human T cell leukemia virus (HTLV-1) is a retrovirus isolated from human T cell tumors and induces cancer development through multiple mechanisms [33]. Oocyte division is also strongly associated with ovarian carcinogenesis [34]. Therefore, the above pathways are closely related to cancer. Nine feature genes were then selected by Cox regression analysis, and a prognostic model for breast cancer consisting of nine mRNAsi-related genes was constructed. The model involved CFB, MAL2, PSME2, MRPL13, HMGB3, DCTPP1, SHCBP1, SLC35A2, and EVA1B, of which CFB, PSME2, and EVA1B were used as cancer prognostic protective factors, and the remaining genes were used as prognostic risk factors. CFB is stably upregulated in various cancer tissues, and in studies of adenocarcinoma, this gene has been shown to alleviate cancer progression by activating cellular immune responses, consistent with the trend of this study in predicting progression of breast cancer [35]. PSME2 has been less studied in cancer, and reports indicate that this gene is a typical poor prognostic marker in renal cell carcinoma and promotes malignant tumor progression by inhibiting autophagy [36]. High expression of EVA1B is bound up closely with high infiltration levels of T cells, macrophages, and neutrophils in cancer tissues, and high expression of this gene implies poor prognosis in glioma patients [37]. This contrasts with our finding, perhaps PSME2 and EVA1B possess cancer specific, and whether the regulation of these two genes also involves autophagy and tumor immune regulation in this study remains to be further explored. The remaining genes exist as risk factors for cancer prognosis, and most of the genes have confirmed this in existing studies. For example, MAL2 and MRPL13 can inhibit tumor antigen presentation to drive breast cancer immune escape, and upregulation of two genes in breast cancer has been demonstrated to drive malignant progression of cancer [38, 39]. Similarly, HMGB3 is also a prototypical marker of breast cancer progression but worsens cancer progression primarily by promoting formation of breast layers of breast cancer cells [40, 41]. DCTPP1 is an oncogene regulated by the oncogenic factor miR-378a-3p, and this gene facilitates breast cancer cell proliferation through the interference of DNA repair signaling pathway [42, 43]. The phenomenon of overexpression of SHCBP1 in breast cancer has been studied, and cellular experiments have demonstrated that this gene directly regulates breast cancer cell proliferation and promotes the cell cycle [44]. SLC35A2 is associated with hypoxia-inducible factors, heat shock proteins, transcription factors, and DNA damage-associated signaling and is involved in the regulation of neutrophil and macrophage polarization in breast cancer [45]. In summary,

the majority of the genes associated with mRNAsi of breast cancer in this study are closely related to cancer development or immune regulation of breast cancer, and it is reasonable to use this constructed prognostic model for clinical prognostic guidance.

In addition to uncovering the corresponding key genes, the results of ssGSEA analysis based on 9 mRNAsi genes in this study demonstrated that the difference regarding survival rate from the high-risk and low-risk group may originate from differences in immunoinfiltrating cells (e.g., Th2, CD8+ T cells, and NK cells). Th2 cells can secrete interleukins to participate in the body's humoral response and assist in the activation of human B cells and participate in antitumor immune responses. Downregulated infiltration of this cell in high-risk group predicts an immunosuppressive response, consistent with the results of this study. Similarly, this study revealed that CD8+ T cells downregulated in the TME act as the cells of choice for targeting cancer, activating cytotoxic T lymphocytes in the tumor immune circulation and mediating antitumor immune responses [46]. In clinical studies, NK cells often synergize with CD8+ T cells in antitumor immune processes, and both have similar cytotoxic mechanisms [47, 48]. This study revealed that the downregulation of multiple immune cell infiltration levels in the high-risk group was an indicator of an immunosuppressive microenvironment in this group, which might be the reason of the unsatisfactory prognosis discovered in high-risk group patients.

In summary, this study revealed the association between mRNAsi and clinical variables in breast cancer samples by K-M curve plotting and K-W test analysis. The gene modules associated with mRNAsi in breast cancer samples were constructed by WGCNA, which was used as a basis to screen and construct a 9-gene risk assessment model. The assessing performance this model on breast cancer patient's prognosis was also validated by WGCNA. ssGSEA analysis revealed the potential association of this risk model with individual somatic mutations and immune cell infiltration, which opens up new possibility for the development of diagnostic and clinical therapeutic strategies for treating breast cancer. However, this study is a bioinformatics analysis for model construction which is lack of clinical trials. Therefore, in future studies, we will collect more clinical sample data and incorporate some clinical information to increase the reliability of the model when constructing the model. At the same time, we did not use wet experiments to verify the constructed model, so we will perform relevant cellular experiments and molecular experiments to verify the model in subsequent experiments.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

XG Z contributed to conceptualization and data curation. JQ L contributed to methodology and formal analysis. XG Z contributed to writing. All authors have reviewed and approved the final manuscript.

References

- [1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2021," *CA: a Cancer Journal for Clinicians*, vol. 71, no. 1, pp. 7–33, 2021.
- [2] J. Y. S. Tsang and G. M. Tse, "Molecular classification of breast cancer," *Advances in Anatomic Pathology*, vol. 27, no. 1, pp. 27–35, 2020.
- [3] X. Wang, C. Sun, X. Huang et al., "The advancing roles of exosomes in breast cancer," *Frontiers in Cell and Development Biology*, vol. 9, article 731062, 2021.
- [4] Y. He, H. Liu, Q. Chen, Y. Shao, and S. Luo, "Relationships between SNPs and prognosis of breast cancer and pathogenic mechanism," *Molecular Genetics & Genomic Medicine*, vol. 7, no. 9, article e871, 2019.
- [5] C. A. Lyssiotis and A. C. Kimmelman, "Metabolic interactions in the tumor microenvironment," *Trends in Cell Biology*, vol. 27, no. 11, pp. 863–875, 2017.
- [6] M. Najafi, K. Mortezaee, and R. Ahadi, "Cancer stem cell (a)symmetry & plasticity: tumorigenesis and therapy relevance," *Life Sciences*, vol. 231, article 116520, 2019.
- [7] M. D. Brooks, M. L. Burness, and M. S. Wicha, "Therapeutic implications of cellular heterogeneity and plasticity in breast cancer," *Cell Stem Cell*, vol. 17, no. 3, pp. 260–271, 2015.
- [8] F. Yang, J. Xu, L. Tang, and X. Guan, "Breast cancer stem cell: the roles and therapeutic implications," *Cellular and Molecular Life Sciences*, vol. 74, no. 6, pp. 951–966, 2017.
- [9] W. Li, H. Ma, J. Zhang, L. Zhu, C. Wang, and Y. Yang, "Unraveling the roles of CD44/CD24 and ALDH1 as cancer stem cell markers in tumorigenesis and metastasis," *Scientific Reports*, vol. 7, no. 1, article 13856, 2017.
- [10] C. J. O'Connor, T. Chen, I. Gonzalez, D. Cao, and Y. Peng, "Cancer stem cells in triple-negative breast cancer: a potential target and prognostic marker," *Biomarkers in Medicine*, vol. 12, no. 7, pp. 813–820, 2018.
- [11] H. Wang, L. Wang, Y. Song et al., "CD44⁺/CD24⁻ phenotype predicts a poor prognosis in triple-negative breast cancer," *Oncology Letters*, vol. 14, no. 5, pp. 5890–5898, 2017.
- [12] H. Zheng, K. Song, Y. Fu et al., "An absolute human stemness index associated with oncogenic dedifferentiation," *Briefings in Bioinformatics*, vol. 22, no. 2, pp. 2151–2160, 2021.
- [13] T. M. Malta, A. Sokolov, A. J. Gentles et al., "Machine learning identifies stemness features associated with oncogenic dedifferentiation," *Cell*, vol. 173, no. 2, pp. 338–354.e15, 2018.
- [14] M. Zhang, H. Chen, B. Liang et al., "Prognostic value of mRNAs_i/corrected mRNAs_i calculated by the one-class logistic regression machine-learning algorithm in glioblastoma within multiple datasets," *Frontiers in Molecular Biosciences*, vol. 8, article 777921, 2021.
- [15] Y. Zhang, J. T. Tseng, I. C. Lien, F. Li, W. Wu, and H. Li, "mRNAs_i index: machine learning in mining lung adenocarcinoma stem cell biomarkers," *Genes (Basel)*, vol. 11, no. 3, 2020.
- [16] M. Zhang, X. Wang, X. Chen, F. Guo, and J. Hong, "Prognostic value of a stemness index-associated signature in primary lower-grade glioma," *Frontiers in Genetics*, vol. 11, p. 441, 2020.
- [17] J. Pei, Y. Wang, and Y. Li, "Identification of key genes controlling breast cancer stem cell characteristics via stemness indices analysis," *Journal of Translational Medicine*, vol. 18, no. 1, p. 74, 2020.
- [18] X. Y. Huang, W. T. Qin, Q. S. Su et al., "A new stemness-related prognostic model for predicting the prognosis in pancreatic ductal adenocarcinoma," *BioMed Research International*, vol. 2021, Article ID 6669570, 13 pages, 2021.
- [19] J. Du, X. Yan, S. Mi et al., "Identification of prognostic model and biomarkers for cancer stem cell characteristics in glioblastoma by network analysis of multi-omics data and stemness indices," *Frontiers in Cell and Development Biology*, vol. 8, article 558961, 2020.
- [20] M. E. Ritchie, B. Phipson, D. Wu et al., "Limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, article e47, 2015.
- [21] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, no. 1, p. 559, 2008.
- [22] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, "clusterProfiler: an R package for comparing biological themes among gene clusters," *Omics: a journal of integrative biology*, vol. 16, no. 5, pp. 284–287, 2012.
- [23] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
- [24] M. F. S. Morgan and R. Gentleman, "GSEABase: gene set enrichment data structures and methods," 2022, R Package Version 1.58.0.
- [25] A. Mayakonda, D. C. Lin, Y. Assenov, C. Plass, and H. P. Koefler, "Maftools: efficient and comprehensive analysis of somatic variants in cancer," *Genome Research*, vol. 28, no. 11, pp. 1747–1756, 2018.
- [26] Z. L. Skidmore, A. H. Wagner, R. Lesurf et al., "GenVisR: genomic visualizations in R," *Bioinformatics (Oxford, England)*, vol. 32, no. 19, pp. 3012–3014, 2016.
- [27] J. C. Chang, "Cancer stem cells: role in tumor growth, recurrence, metastasis, and treatment resistance," *Medicine (Baltimore)*, vol. 95, no. 1S, pp. S20–S25, 2016.
- [28] K. H. Bai, S. Y. He, L. L. Shu et al., "Identification of cancer stem cell characteristics in liver hepatocellular carcinoma by WGCNA analysis of transcriptome stemness index," *Cancer Medicine*, vol. 9, no. 12, pp. 4290–4298, 2020.
- [29] Z. Wang, Y. Wang, T. Yang et al., "Machine learning revealed stemness features and a novel stemness-based classification with appealing implications in discriminating the prognosis, immunotherapy and temozolomide responses of 906 glioblastoma patients," *Briefings in Bioinformatics*, vol. 22, no. 5, 2021.
- [30] K. Keyomarsi and A. B. Pardee, "Redundant cyclin overexpression and gene amplification in breast cancer cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 3, pp. 1112–1116, 1993.
- [31] K. Kitahara, W. Yasui, H. Kuniyasu et al., "Concurrent amplification of cyclin E and CDK2 genes in colorectal carcinomas," *International Journal of Cancer*, vol. 62, no. 1, pp. 25–28, 1995.
- [32] M. Marone, G. Scambia, C. Giannitelli et al., "Analysis of cyclin E and CDK2 in ovarian cancer: gene amplification and

- RNA overexpression,” *International Journal of Cancer*, vol. 75, no. 1, pp. 34–39, 1998.
- [33] J. T. Schiller and D. R. Lowy, “An introduction to virus infections and human cancer,” *Recent Results in Cancer Research*, vol. 217, pp. 1–11, 2021.
- [34] P. E. Bouet, T. Boueilh, J. M. C. de la Barca et al., “The cytokine profile of follicular fluid changes during ovarian ageing,” *Journal of gynecology obstetrics and human reproduction*, vol. 49, no. 4, article 101704, 2020.
- [35] P. Wu, J. Shi, W. Sun, and H. Zhang, “The prognostic value of plasma complement factor B (CFB) in thyroid carcinoma,” *Bioengineered*, vol. 12, no. 2, pp. 12854–12866, 2021.
- [36] X. Wang, F. Wu, Y. Deng et al., “Increased expression of PSME2 is associated with clear cell renal cell carcinoma invasion by regulating BNIP3-mediated autophagy,” *International Journal of Oncology*, vol. 59, no. 6, 2021.
- [37] S. Qu, J. Liu, and H. Wang, “EVA1B to evaluate the tumor immune microenvironment and clinical prognosis in glioma,” *Frontiers in Immunology*, vol. 12, article 648416, 2021.
- [38] Y. Fang, L. Wang, C. Wan et al., “MAL2 drives immune evasion in breast cancer by suppressing tumor antigen presentation,” *The Journal of Clinical Investigation*, vol. 131, no. 1, 2021.
- [39] Z. Tao, H. Suo, L. Zhang et al., “MRPL13 is a prognostic cancer biomarker and correlates with immune infiltrates in breast cancer,” *Oncotargets and Therapy*, vol. 13, pp. 12255–12268, 2020.
- [40] J. Gu, T. Xu, C. M. Zhang, H. Y. Chen, Q. H. Huang, and Q. Zhang, “HMGB3 small interfere RNA suppresses mammosphere formation of MDA-MB-231 cells by down-regulating expression of HIF1 α ,” *European Review for Medical and Pharmacological Sciences*, vol. 23, no. 21, pp. 9506–9516, 2019.
- [41] J. Gu, T. Xu, Q. H. Huang, C. M. Zhang, and H. Y. Chen, “HMGB3 silencing inhibits breast cancer cell proliferation and tumor growth by interacting with hypoxia-inducible factor 1 α ,” *Cancer Management and Research*, vol. 11, pp. 5075–5089, 2019.
- [42] M. Niu, M. Shan, Y. Liu et al., “DCTPP1, an oncogene regulated by miR-378a-3p, promotes proliferation of breast cancer via DNA repair signaling pathway,” *Frontiers in Oncology*, vol. 11, article 641931, 2021.
- [43] Y. Wang, P. Chen, X. Chen et al., “ROS-induced DCTPP1 upregulation contributes to cisplatin resistance in ovarian cancer,” *Frontiers in Molecular Biosciences*, vol. 9, article 838006, 2022.
- [44] W. Feng, H. C. Li, K. Xu et al., “SHCBP1 is over-expressed in breast cancer and is important in the proliferation and apoptosis of the human malignant breast cancer cell line,” *Gene*, vol. 587, no. 1, pp. 91–97, 2016.
- [45] H. D. K. Ta, D. T. Minh Xuan, W. C. Tang et al., “Novel insights into the prognosis and immunological value of the SLC35A (solute carrier 35A) family genes in human breast cancer,” *Biomedicine*, vol. 9, no. 12, 2021.
- [46] B. Farhood, M. Najafi, and K. Mortezaee, “CD8⁺ cytotoxic T lymphocytes in cancer immunotherapy: a review,” *Journal of Cellular Physiology*, vol. 234, no. 6, pp. 8509–8521, 2019.
- [47] G. Xie, H. Dong, Y. Liang, J. D. Ham, R. Rizwan, and J. Chen, “CAR-NK cells: a promising cellular immunotherapy for cancer,” *eBioMedicine*, vol. 59, article 102975, 2020.
- [48] P. André, C. Denis, C. Soulas et al., “Anti-NKG2A mAb is a checkpoint inhibitor that promotes anti-tumor immunity by unleashing both T and NK cells,” *Cell*, vol. 175, no. 7, pp. 1731–1743.e13, 2018.

Research Article

Comparative Study on Feature Selection in Protein Structure and Function Prediction

Wenjing Yi,¹ Ao Sun,² Manman Liu,² Xiaoqing Liu,³ Wei Zhang ,² and Qi Dai ¹

¹College of Life Sciences, Zhejiang Sci-Tech University, Hangzhou 310018, China

²College of Informatics Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China

³College of Sciences, Hangzhou Dianzi University, Hangzhou 310018, China

Correspondence should be addressed to Wei Zhang; zhangweicse@zstu.edu.cn and Qi Dai; daiailiu04@yahoo.com

Received 27 July 2022; Accepted 14 September 2022; Published 11 October 2022

Academic Editor: Lin Lu

Copyright © 2022 Wenjing Yi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Many effective methods extract and fuse different protein features to study the relationship between protein sequence, structure, and function, but different methods have preferences in solving the research of protein structure and function, which requires selecting valuable and contributing features to design more effective prediction methods. This work mainly focused on the feature selection methods in the study of protein structure and function, and systematically compared and analyzed the efficiency of different feature selection methods in the prediction of protein structures, protein disorders, protein molecular chaperones, and protein solubility. The results show that the feature selection method based on nonlinear SVM performs best in protein structure prediction, protein solubility prediction, protein molecular chaperone prediction, and protein solubility prediction. After selection, the accuracy of features is improved by 13.16%~71%, especially the Kmer features and PSSM features of proteins.

1. Introduction

Protein structure and function is the basic research field of protein research, which is of great significance for the study of protein folding rate, DNA binding sites, and protein folding recognition [1–7]. In recent years, the gap between protein sequence and protein structure is becoming larger and larger with the development of sequencing technology, and the speed of identifying protein structure and function through experimental methods is relatively slow. Therefore, it is necessary to develop computational methods to quickly and accurately determine protein structure and function.

The function of a protein is determined by its spatial structure, which is determined by its sequence. Therefore, sequence information can be used to predict protein structure and function directly, so as to further guide biological experiments and reduce experimental costs. After the concept of protein structure class was put forward, several protein structure and function prediction methods were proposed [3–5, 7–11]. Some methods use protein composition information to predict protein structure and function [1, 12, 13]. For example, short pep-

tide composition [14–16], pseudo amino acid composition [17–20], and functional domain composition match [21]. The sequence characteristic information is expressed as amino acid composition (AAC) by calculating the ratio of 20 amino acid residues in the sequence [14–16], but it does not take into account the physicochemical properties and interaction of amino acids. In order to overcome the above problems, pseudo amino acid composition (PseACC) calculates the composition of amino acid residues based on the hydrophobicity and other physical and chemical properties of amino acid residues [17–23].

The above methods are outstanding in high similarity data, but for low similarity data, their performance is ordinary, with prediction accuracy 50%. Therefore, we need to design more effective prediction algorithm. Kurgan et al. predicted protein secondary structures and designed SCPRED method on this basis [24]. Zhang et al. calculated the TPM matrix and took it as the characteristic representation of the protein secondary structures [25]. Dai et al. statistically analyzed the characteristic distribution of protein secondary structures and applied them to protein structure prediction [26]. Ding et al. constructed a

multidimensional representation vector of protein secondary structure features, and fused it with existing features to achieve protein structure prediction [27]. Chen et al. and Kumar et al. combined structural information with physical and chemical characteristics to design a protein structure prediction method [28, 29]. Nanni et al. calculated the primary sequence characteristics and secondary structure characteristics of protein, respectively, for protein structure and function prediction [30]. Wang et al. simplified PSSM features and combined them with protein secondary structure features for protein structure prediction [31].

Through the fusion of the above features, the prediction accuracy of some methods on low similarity data sets has been improved to more than 80%, but there are still some problems in the development of protein structure and function prediction. In order to improve the prediction accuracy and efficiency of the model, the existing research is mainly achieved by fusing different types of protein features. However, it is worth noting that a simple combination of different features does not necessarily improve the prediction performance. If the combination is not appropriate, it may even offset the information contained in each other, which will not only lead to information redundancy but also increase the complexity and calculation of the model. This requires selecting valuable and contributing features, and then effective fusion, in order to design more effective prediction methods of protein structure and function.

With the above problems in mind, we introduced 16 feature selection methods based on mutual information, feature selection based on support vector machine, feature selection based on genetic algorithm, feature selection based on kurtosis and skewness, ReliefF, and sequentialfs information selection, and systematically compared their performance in protein structure class prediction, protein disorder prediction, protein molecular chaperone prediction, and protein solubility prediction. Through a comprehensive comparison and discussion, some novel valuable guidelines for use of the feature selection method for protein structure and function prediction are obtained.

2. Materials and Methods

2.1. Datasets. Four standard data sets for protein structure and function prediction were used in this work, which are protein structural class data set, molecular chaperone data set [32], solubility data set [33], and protein disorder data set [34]. The structure data set consists of 278 α structural proteins and 192 β proteins composition of structure. The molecular chaperone data set [35] is composed of 109 proteins that need Dnak/GroEL molecular chaperones to fold correctly, and 39 proteins that can fold autonomously. The solubility data set [36] is composed of 1000 proteins with high solubility and 1000 proteins with low solubility. The protein disorder data set is composed of 630 disordered proteins from DisProt and 3347 structural proteins from SCOP [37]. The detailed information of the data set is shown in Table 1.

TABLE 1: Detailed information of protein structure and function data.

Dataset	Positive	Negative	Total
Protein structural class	278	192	470
Molecular chaperone	109	39	148
Solubility	1000	1000	2000
Protein disorder	630	3347	3977

2.2. Sequence Feature. Six kinds of different characteristic information of proteins are extracted [26]. They are Kmer, Pseudo Amino Acid Composition (PseAAC), Correlation-based features (correlation), composition-transition-distribution (RCTD), order-based features (order), position-based features (position), GO, and position-specific score matrix (PSSM).

2.3. Feature Selection

2.3.1. Feature Selection Based on Mutual Information. Feature selection based on mutual information has become more and more popular in data mining, especially because of their ease of use, effectiveness, and strong theoretical foundation rooted in information theory. We adopted nine feature selection algorithms based on mutual information [38], which are maxRelFS, MRMRFS, minRedFS, MIQFS, QPFSS, SPECCMI_Fs, MRMRFS, CMIMFSand, and CIFEFS. The common point of these methods is that they all focus on the concepts of redundancy and correlation, and use greedy schemes to build the selected feature sets incrementally. Given a sample, the column is the characteristic matrix X , and the corresponding category is C . The calculation formula of mutual information is

$$Rel(X_i) = I(X_i; C) = \sum_{X_i, C} P(X_i, C) \log \frac{P(X_i, C)}{P(X_i)P(C)}. \quad (1)$$

If the selected set is S , the calculation formula of redundancy is as follows:

$$Red(X_i|S) = \frac{1}{S} \sum_{X_j \in S} I(X_i; X_j). \quad (2)$$

The above nine feature selection algorithms calculate the mutual information value of each feature and category C , and select the feature with the largest mutual information as the optimal feature. Then, according to the feature selection method of quadratic programming, the features with minimum redundancy and maximum correlation are selected one by one. Finally, we can get a feature vector sorted according to the importance of features.

2.3.2. Support Vector Machine Recursive Feature Extraction (SVM-RFE). Support vector machine recursive feature extraction (SVM-RFE) is divided into linear SVM-RFE and nonlinear SVM-RFE. The details are as follows:

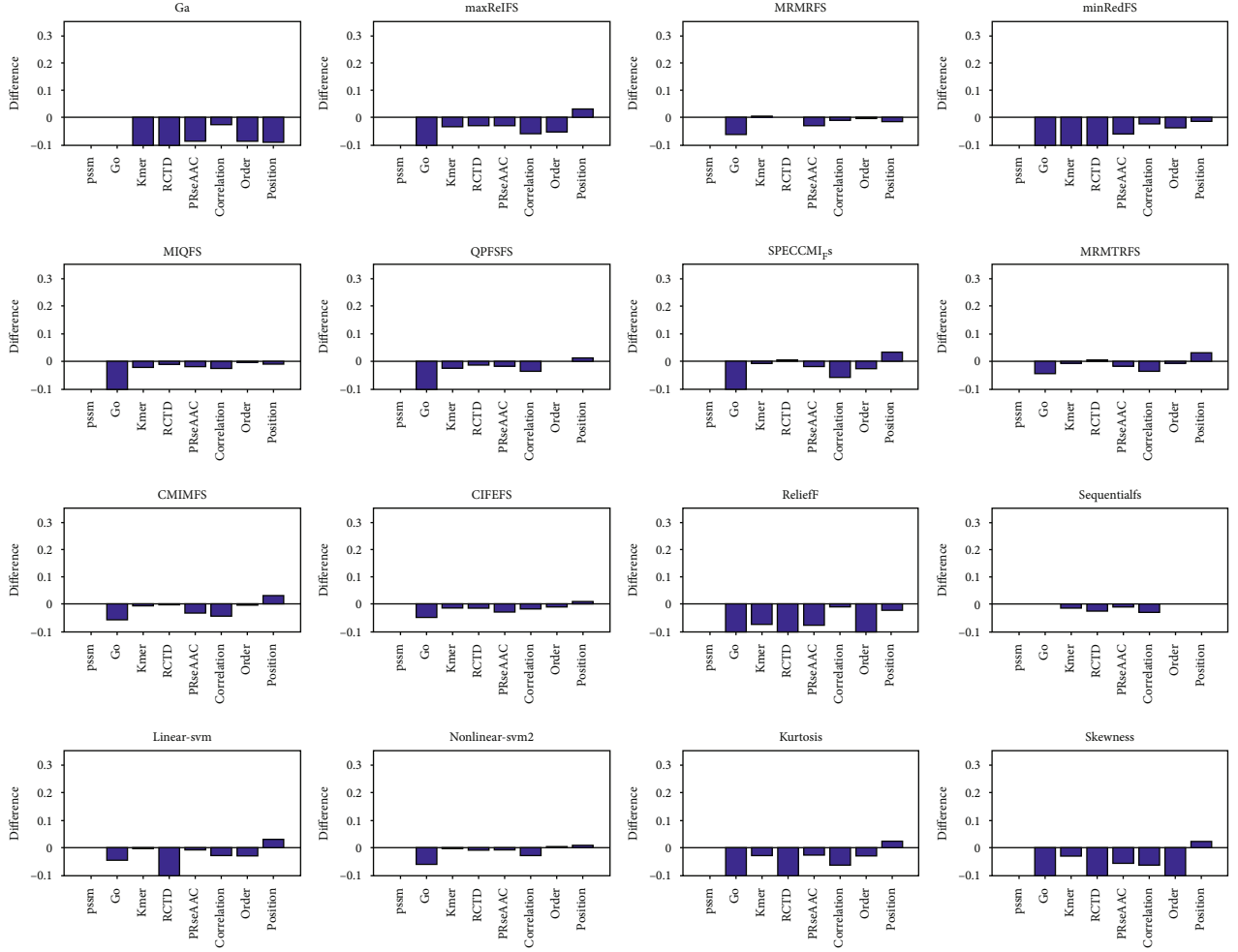


FIGURE 1: The comparison between the accuracy of support vector machine prediction and that of single class feature prediction after selecting the top 10 features. For each graph, the selection method is arranged from left to right and from top to bottom. They are GA, and there are nine selection methods of mutual information, relief, sequentialfs, linear SVM, nonlinear SVM, kurtosis, and skewness. The horizontal axis represents sequence features, which are PSSM, go, Kmer, RCTD, PRseAAC, correlation, order, and position, respectively.

(1) *Linear SVM-RFE*. For a samples $\{x_i, y_i\}$, the objective function of linear SVM-RFE is

$$f(x) = a \cdot x + b, \quad (3)$$

where a is weight factor and b is deviation. Thus, the Lagrangian version of this problem can be expressed as

$$L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j, \quad (4)$$

where α_i is Lagrange factor. α_i can be calculated by LD maximum under the condition of $\alpha_i \geq 0$ and $\sum_{i=1}^n \alpha_i y_i = 0$. Weighting factors can be calculated by the following formula:

$$a = \sum_{i=1}^n \alpha_i y_i x_i. \quad (5)$$

k -th feature sorting criteria is the square of the k -th weighting factor.

$$J(k) = w_k^2. \quad (6)$$

In the training process, the feature with the smallest influence factor will be deleted every time, and so on, until all the features are deleted. Then, the importance of features is sorted according to the order in which they are deleted [39].

(2) *Nonlinear SVM-RFE*. In many cases, the number of features of the sample will be more than the number of samples. At this time, using linear SVM-RFE can avoid the phenomenon of over fitting [40]. However, when the number of samples is greater than the number of features, the selection result of nonlinear SVM-RFE will be better than that of linear SVM-RFE.

Nonlinear SVM-RFE will map features to new spaces with higher dimensions as follows:

$$x \in R^d \mapsto \varphi(x) \in R^h. \quad (7)$$

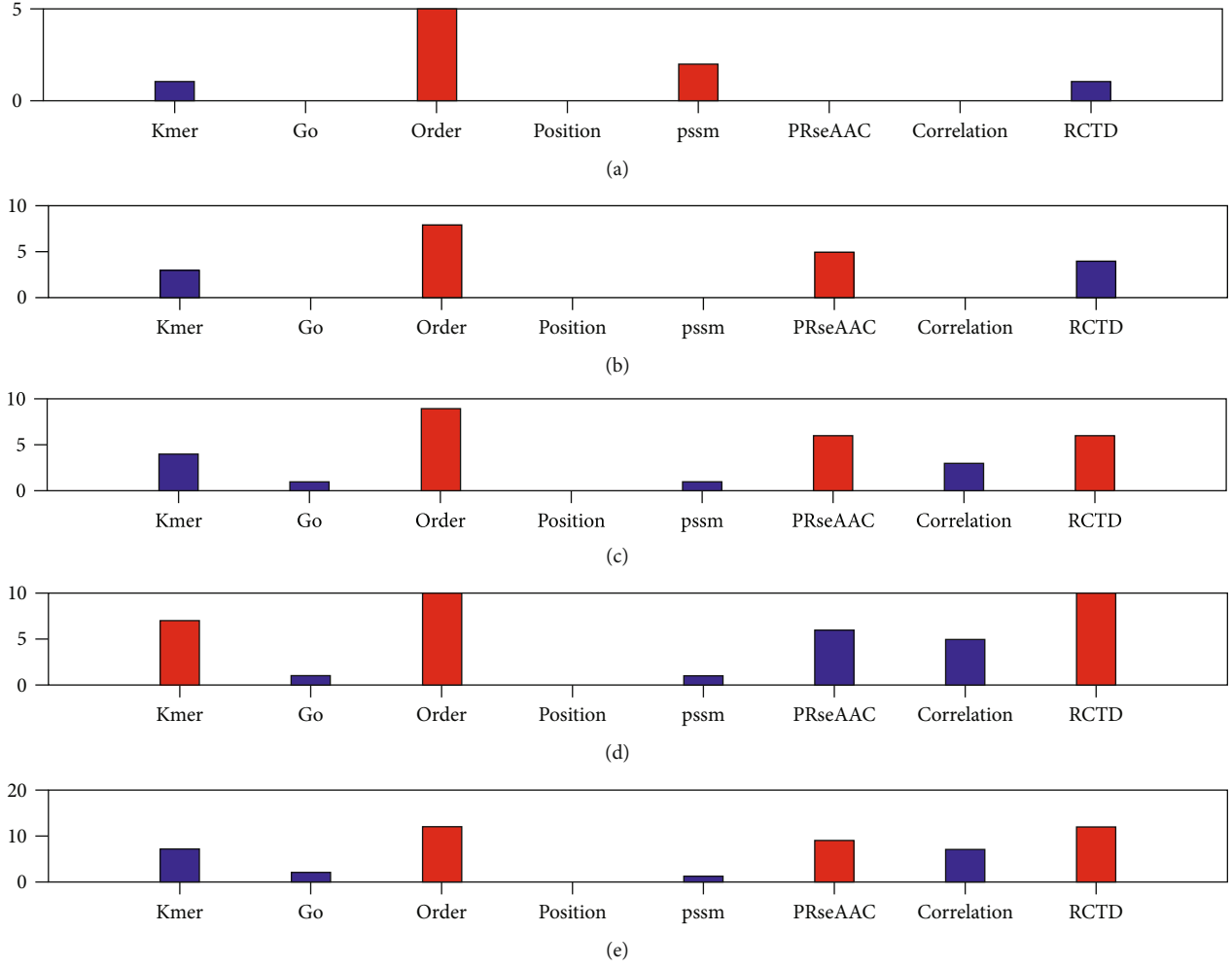


FIGURE 2: The number of 8 types of features in the top selected features in the protein structural data. From (a) to (e), it means that the number of selected features is 10 to 50, respectively.

In the new space, the samples are expected to be linearly separable. Its Lagrangian form can be expressed as

$$L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \varphi(x_i) \varphi(x_j). \quad (8)$$

Thus, we could transform inner product $\varphi(x_i) \varphi(x_j)$ into a Gaussian kernel $K(x_i, x_j)$ as follows:

$$K(x_i, x_j) = e^{-\lambda \|x_i - x_j\|^2}. \quad (9)$$

Thus, k -th feature sorting criteria could be expressed as

$$J(k) = \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i^{(-k)}, x_j^{(-k)}). \quad (10)$$

$x_i^{(-k)}$ represents that feature k has been removed.

2.3.3. Feature Selection Based on Genetic Algorithm. We adopted the assembled neural network (ASNN) algorithm. This method carries out combinatorial optimization by using the idea of genetic algorithm. For a given data set, a behavior sample can be constructed and listed as the matrix X of features [41], and finally a feature vector will be obtained, which is the optimal feature set, but the ranking of each feature is not related to its importance.

2.3.4. Feature Selection Based on Kurtosis and Skewness. For a vector of length n $\{x_1, x_2, \dots, x_n\}$, its kurtosis and skewness are calculated as follows:

$$\begin{aligned} \text{Kurtosis} &= \frac{\sum_{i=1}^n (x_i - x)^4}{(n-1)SD^4} - 3, \\ \text{skewness} &= \frac{\sum_{i=1}^n (x_i - x)^3}{(n-1)SD^3}. \end{aligned} \quad (11)$$

Kurtosis and skewness are statistics used to measure the distribution of data. In this work, we calculated the skewness

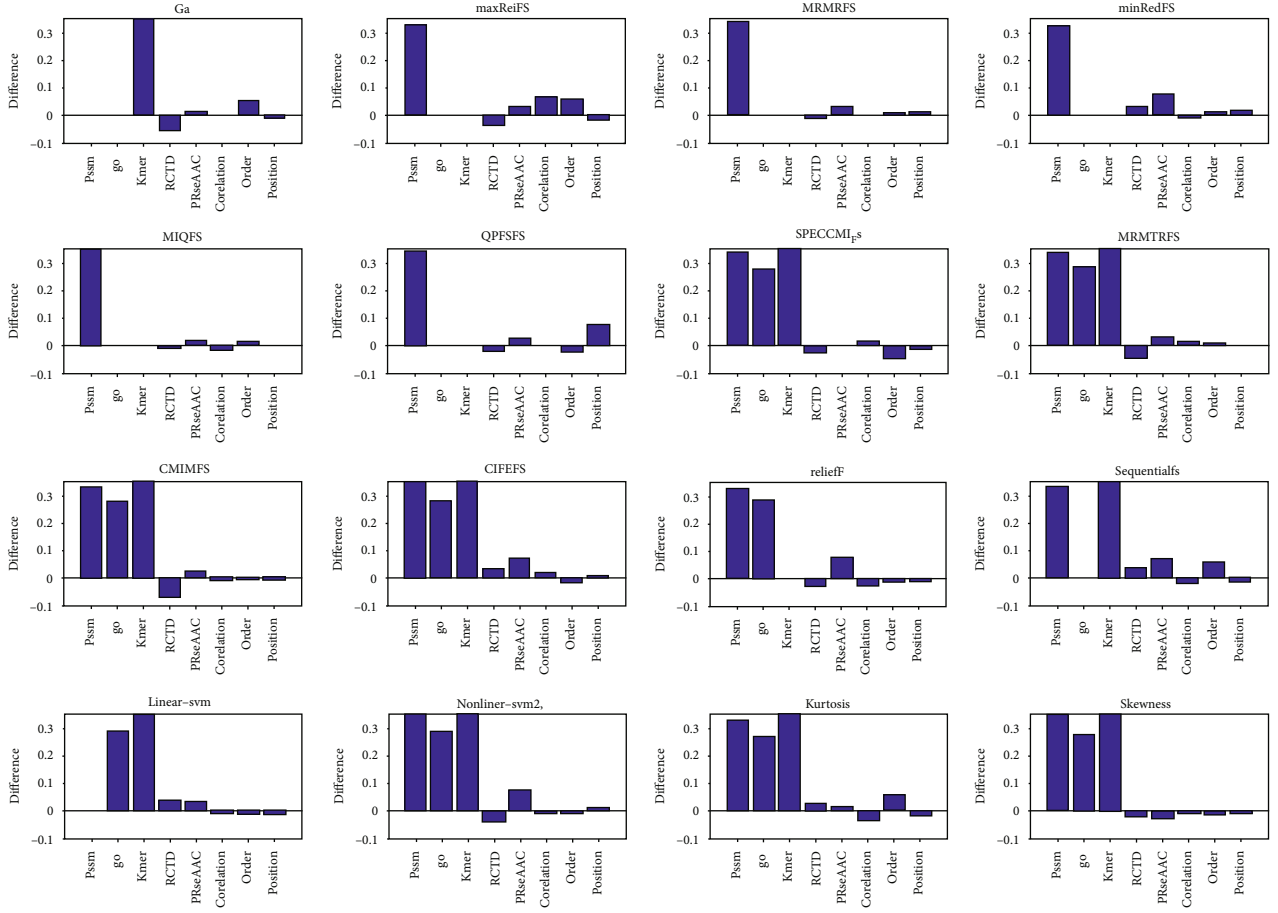


FIGURE 3: The comparison between the accuracy of support vector machine prediction and that of single class feature prediction after selecting the top 10 features. For each graph, the selection method is arranged from left to right and from top to bottom. They are GA, and there are nine selection methods of mutual information, relief, sequentialfs, linear SVM, nonlinear SVM, kurtosis and skewness. The horizontal axis represents sequence features, which are PSSM, go, Kmer, RCTD, PRseAAC, correlation, order, and position, respectively.

and kurtosis of each feature, and then sort them according to their values as a method to measure the importance of features.

2.3.5. Relief Algorithm. Relief algorithm randomly takes a sample R from the training sample set every time, then finds k nearest neighbor samples of R from the sample set of the same kind as R , and finds k nearest neighbor samples from the sample set of different classes of each R , and then updates the weight of each feature. The formula is as follows:

$$W(A) = W(A) - \sum_{j=1}^k \frac{\text{diff}(A, R, H_j)}{mk} + \sum_{C \neq \text{class}(R)} \frac{p(C)/1 - p(\text{Class}(R)) \sum_{j=1}^k \text{diff}(A, R, M_j(C))}{mk}, \quad (12)$$

where $\text{diff}(A, R1, R2)$ means the difference between feature $R1$ and $R2$ in feature A . $M_j(C)$ means the j -th nearest neighbor sample in class C . Formula is as follows:

$$\text{diff}(A, R1, R2) = \begin{cases} \frac{|R1[A] - R2[A]|}{\max(A) - \min(A)} & \text{if } A \text{ is consequent,} \\ 0 & \text{if } A \text{ is unconsequent and } R1[A] = R2[A], \\ 1 & \text{if } A \text{ is unconsequent and } R1[A] \neq R2[A]. \end{cases} \quad (13)$$

2.3.6. Sequentialfs. We adopted the forward feature selection algorithm of sequence in this work. For a training set $\{x_{\text{train}}, y_{\text{train}}\}$ and validation set $\{x_{\text{validation}}, y_{\text{validation}}\}$, the evaluation criteria can be expressed as

$$\left\| y_{\text{validation}} - x_{\text{validation}} \frac{x_{\text{train}}}{y_{\text{train}}} \right\|_2. \quad (14)$$

2.4. Classification Algorithm. Support vector machine is a large-scale edge classifier based on statistical learning theory [42]. It uses the optimal separation hyperplane to separate two kinds of data. For binary support vector machines, the decision function is

$$f(x) = \sum_{i=1}^N \alpha_i y_i K(x_i, x) + b. \quad (15)$$

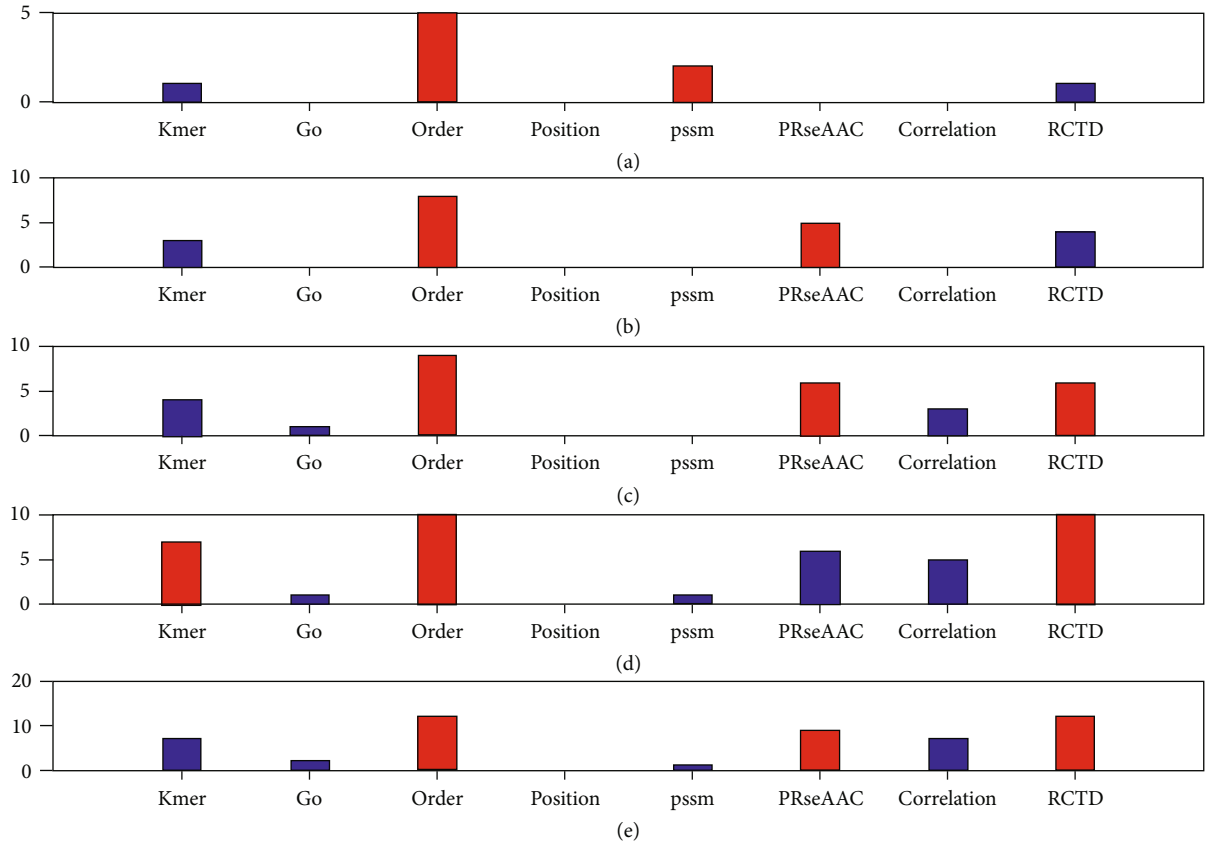


FIGURE 4: The number of 8 types of features in the top selected features in the protein structural data. From (a) to (e), it means that the number of selected features is 10 to 50, respectively.

where b is a constant, C is a cost parameter controlling the trade-off between allowing training errors and forcing rigid margins, $y_i \in \{-1, +1\}$, x_i is the support vector, $0 \leq \alpha_i \leq C$, and $K(x_i, x)$ is the kernel function. This work chooses the Gauss kernel function of support vector machine because of its superiority in solving nonlinear problems [34, 37]. Furthermore, a simple grid search strategy is used to select the parameters C and gamma with the highest overall prediction. It is designed based on 10 times cross validation of each dataset, and the values of C and gamma are taken from the 2^{-10} to 2^{10} .

2.5. Performance Evaluation. This work adopted different feature selection methods for different data sets, and used the leave one method for evaluation. Finally, the prediction results are compared by calculating accuracy.

For each data set, we compared the efficiency of different feature selection methods through the following steps. The following takes the feature selection method based on genetic algorithm (GA) and PSSM features as examples to introduce the evaluation process:

- (1) PSSM information is selected by GA feature selection method
- (2) Select the top 10, 20, 30, 40, and 50 features using GA (if the number of features is insufficient, all the information will be taken out), input them into

SVM classifier for classification prediction, and calculate the accuracy of prediction ACC1, ACC2, ACC3, ACC4, and ACC5

- (3) Subtract the accuracy of the whole PSSM information from ACC1, ACC2, ACC 3, ACC 4, and ACC 5
- (4) Compare the changes in accuracy of various special products after 16 selection methods

We also compared and analyzed the characteristics of selection, and the main steps are as follows:

- (1) Use the above 16 selection methods to select each type of feature
- (2) According to the selection results of 16 feature selection methods, the importance of each type of feature is ranked
- (3) Take out the first 10, 20, 30, 40, and 50 features of each type of feature, respectively, (if the number of features is insufficient, all the information will be taken out) and mix them together as five new mixed features ($I_{10}, I_{20}, I_{30}, I_{40}, I_{50}$);
- (4) Then, 16 feature selection methods are used to select the mixed features

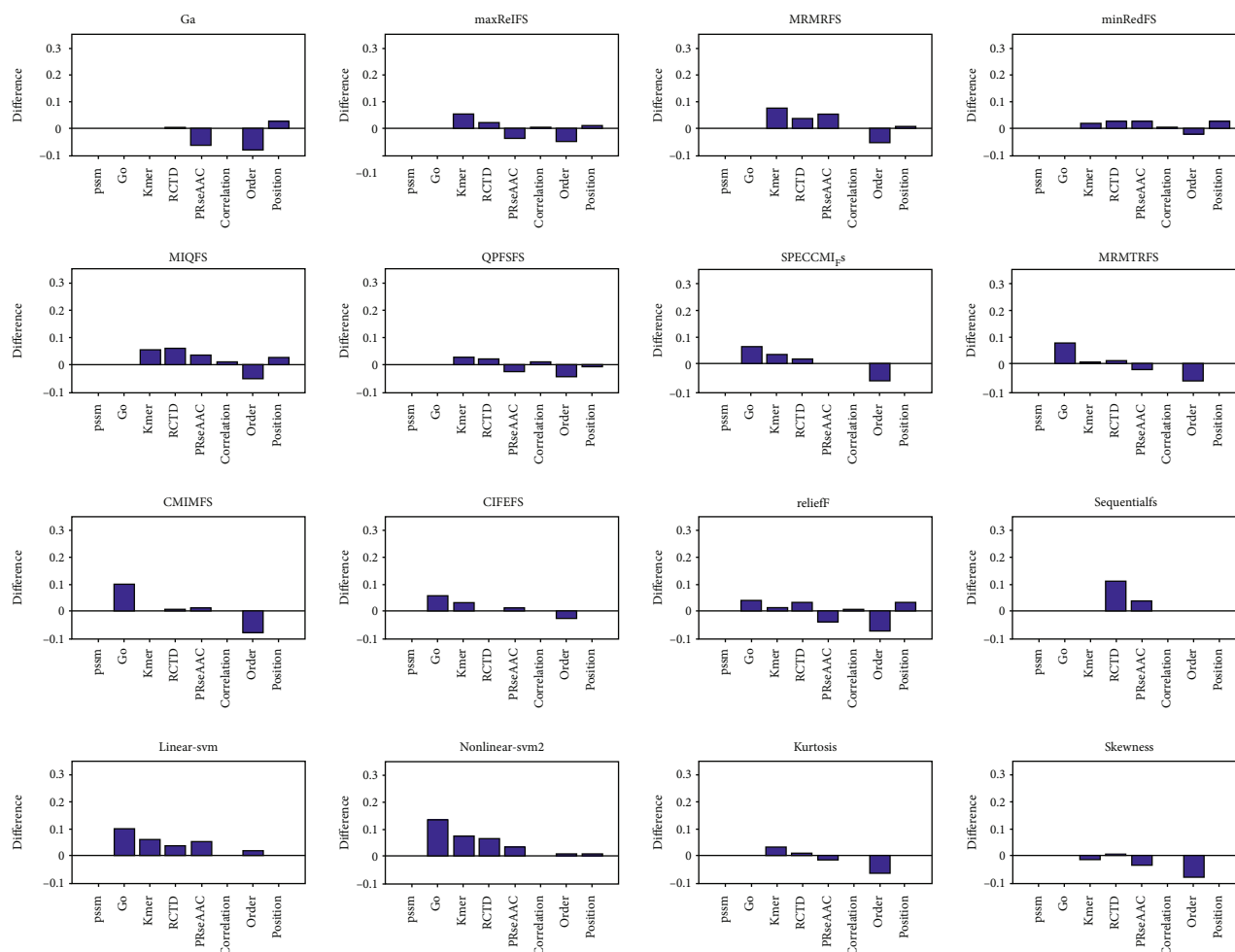


FIGURE 5: The comparison between the accuracy of support vector machine prediction and that of single class feature prediction after selecting the top 10 features. For each graph, the selection method is arranged from left to right and from top to bottom. They are GA, and there are nine selection methods of mutual information, relief, sequentialfs, linear SVM, nonlinear SVM, kurtosis, and skewness. The horizontal axis represents sequence features, which are PSSM, go, Kmer, RCTD, PRseAAC, correlation, order, and position, respectively.

- (5) According to the results of the fourth step, the importance of the fused features is ranked
- (6) Take out the top 10 features and count the type of features from which these 10 features come. Take out the top 20 features and count the categories of features
- (7) In five cases, if there are a large number of certain features (or observed), it means that such features are more important

3. Results and Discussion

3.1. Comparison of Feature Selection in Protein Structure Prediction. We first discussed the efficiency of different feature selection methods in protein structure prediction. We adopted the structural data set, which contains 278 items α structural proteins and 192 β structural proteins. In this work, eight kinds of features are selected through 16 feature

selection methods, and the selected features are input into the support vector machine to predict the structural class of protein. The quality of feature selection methods is evaluated based on the accuracy of prediction, which are represented in Figure 1 and Supplementary Figures 1–4.

From Figure 1 and Supplementary Figures 1–4, it is easy to note that the accuracy of MRMRFs, MRMTRFS, CMIMFS, CIFEFS, and nonlinear SVM feature selection methods changes the most, and the change range is 3.19% for the position feature. By comparing the accuracy of the first 20-50 features selected with that of the unselected features, it can be seen that the biggest change in accuracy is the GO features selected by nonlinear SVM, with changes of 2.13%, 6.39%, 6.17%, and 4.68%, respectively. Therefore, nonlinear SVM feature selection method performs best in protein structure prediction.

For structural data sets [43], we further compared and analyzed the types of selected features. First, eight types of features are fused, and the fused features are selected through 16 feature selection methods, and the top 10-50 features are

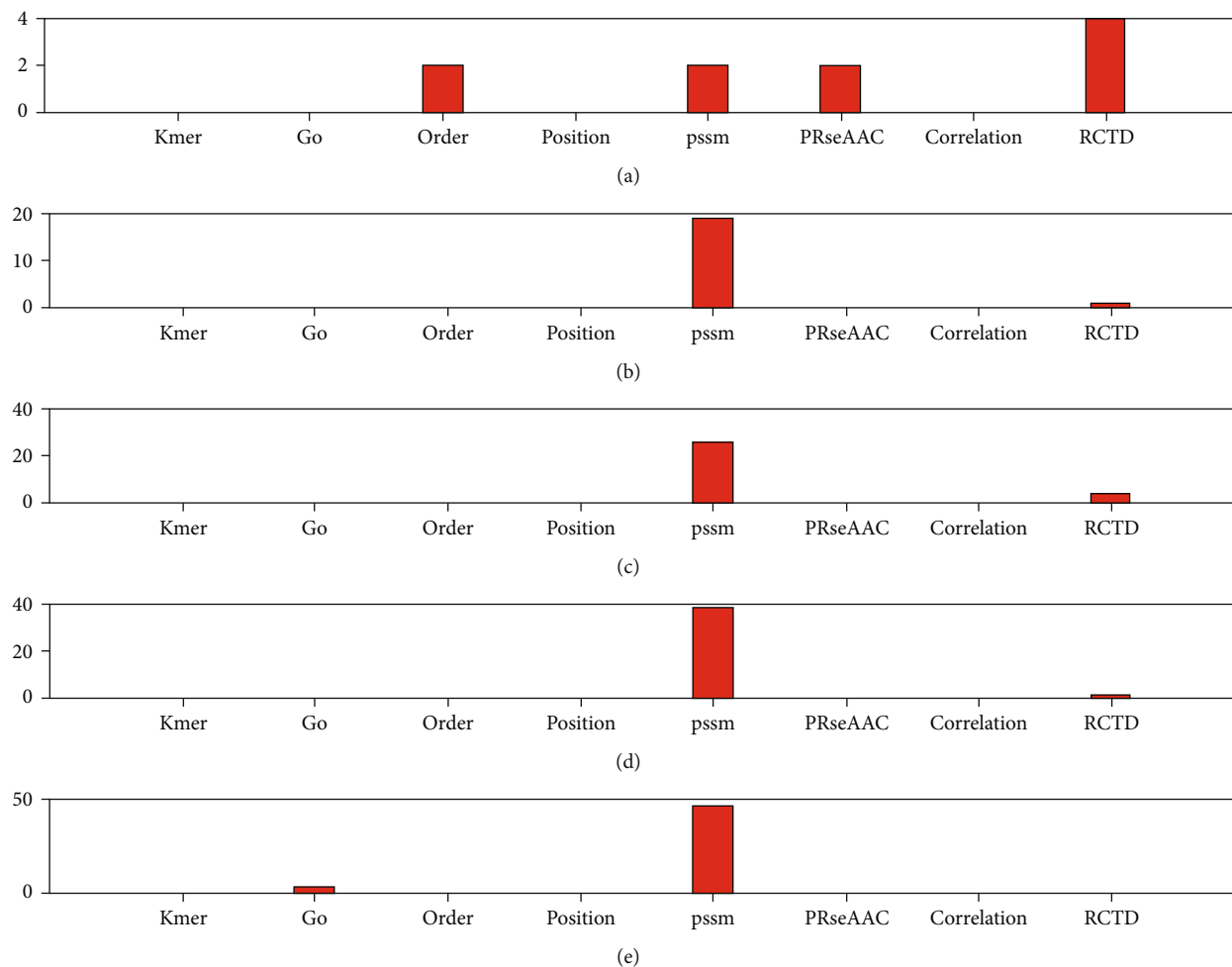


FIGURE 6: The number of 8 types of features in the top selected features in the protein structural data. From (a) to (e), it means that the number of selected features is 10 to 50, respectively.

selected. Then, the number of eight types of features in the top 10-50 total selected features is counted, and the preference of eight types of features is evaluated by proportion. Figure 2 shows the number of 8 types of features in the top 10-50 total selected features in the protein structural data.

Figure 2 show that when the total number of selections is 10, there are 5 order features, accounting for 50%. When the total selection number is 20, there are 8 order features, accounting for 40%. When the selection number is 40, both order and RCTD have 10, accounting for 25% of the top 40 features. When the total selection number is 50, there are 12 orders and RCTD features, respectively, accounting for 24% of the total. The above results show that order feature is the first choice for protein structure prediction, followed by RCTD feature.

3.2. Comparison of Feature Selection in Protein Disorder Prediction. We then discussed the efficiency of different feature selection methods in protein disorder prediction. The protein disorder data set [44] used in this chapter is from two protein databases related to structural classes, including 630 disordered proteins from disProt and 3347 structural proteins from SCOP. In this work, eight kinds of features

are selected through 16 feature selection methods, and the selected features are input into KNN to predict protein disorder. The quality of feature selection methods is evaluated based on the accuracy of prediction, which are represented in Figure 3 and Supplementary Figures 5–8.

It can be seen from Figure 3 and Supplementary Figures 5–8 that when PSSM feature, go feature and Kmer feature are input into KNN algorithm for prediction, the change values of their accuracy are 51.28%, 55.11% and 26.95%, respectively. It can be seen that after feature selection, the accuracy of protein disorder prediction is significantly improved. When selecting 10 features, SPECCMI_FS performs best based on Kmer feature, and its accuracy by 71%. When selecting the first 20 and 30 features, the nonlinear SVM feature selection method is particularly prominent in Kmer features, and its accuracy has increased by 64.19%. Among the top 40 features selected, CIFEFS selection method performs best in Kmer features, and the accuracy is improved to 65.21%. Among the top 50 features selected, CIFEFS and linear SVM selection methods are outstanding, and the accuracy has increased by 59.61%. The above results show that for protein disorder data sets, SPECCMI_FS, CIFEFS, nonlinear SVM, and linear SVM

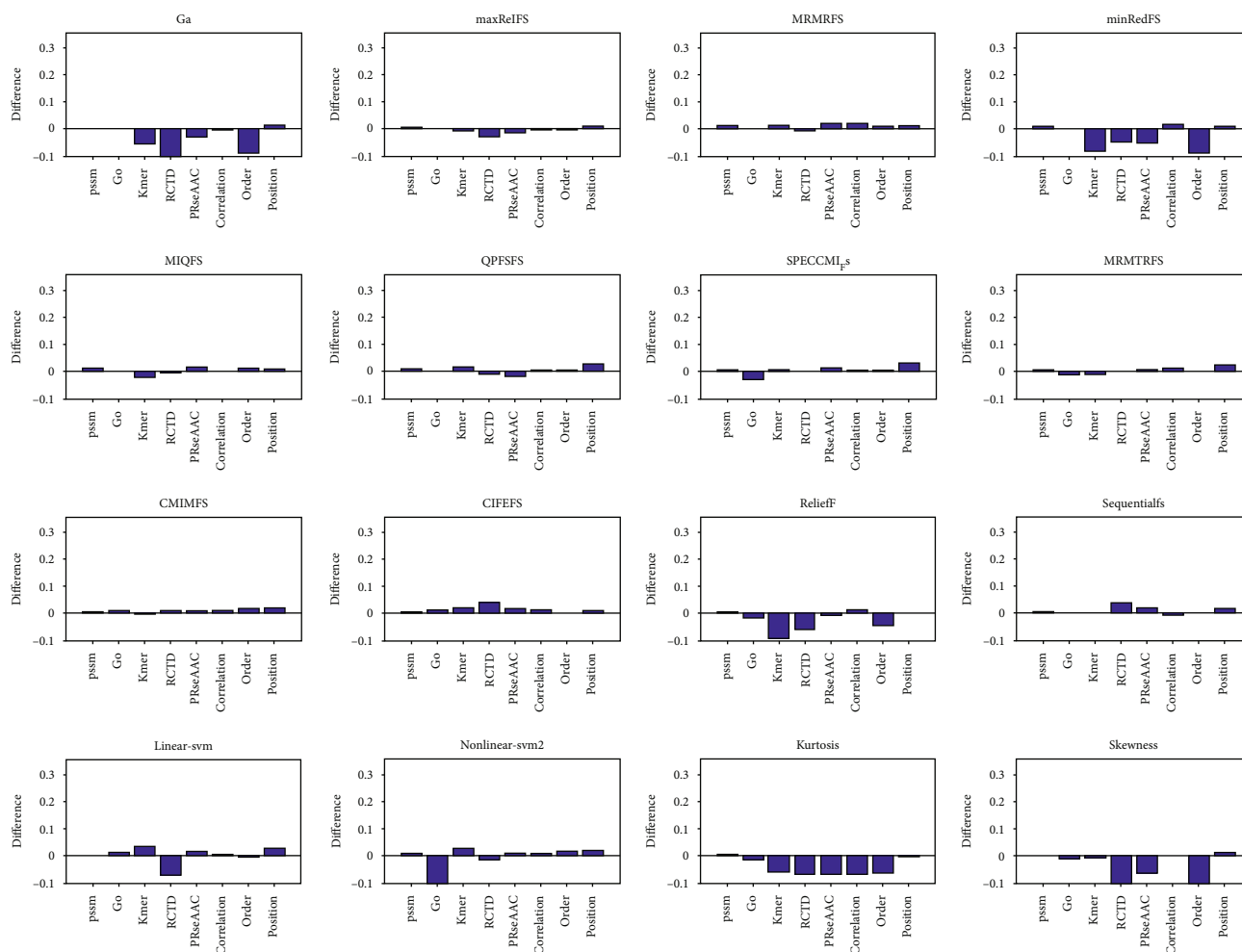


FIGURE 7: The comparison between the accuracy of support vector machine prediction and that of single class feature prediction after selecting the top 10 features. For each graph, the selection method is arranged from left to right and from top to bottom. They are GA, and there are nine selection methods of mutual information, relief, sequentialfs, linear SVM, nonlinear SVM, kurtosis, and skewness. The horizontal axis represents sequence features, which are PSSM, go, Kmer, RCTD, PRseAAC, correlation, order, and position, respectively.

feature selection methods can select core features from Kmer features, which improve its accuracy by 59.61% ~71%.

We also further compared the types of selected features. First, eight types of features are fused, and the fused features are selected through 16 feature selection methods, and the top 10-50 features are selected. Then, the number of eight types of features in the top 10-50 total selected features is counted, and the preference of eight types of features is evaluated by proportion. Figure 4 shows the number of 8 types of features in the top 10-50 total selected features in the protein disorder data set.

Figure 4 shows the number of features selected at five levels from top to bottom. If the top 10 fusion features are selected, 5 of them are from order features. If the first 20 fusion features are selected, 8 of them are from order features. If the first 30 fusion features are selected, 9 of them are from order features. If the first 40 fusion features are selected, there are 10 features from order and RCTD, respectively. If the top 50 fusion features are selected, 12 of them are from order and RCTD features, respectively. Therefore,

the order and RCTD feature will help to improve the accuracy of the protein disorder prediction.

3.3. Comparison of Feature Selection in Protein Molecular Chaperone Prediction. We then discussed the efficiency of different feature selection methods in protein molecular chaperone prediction. In the data set used in this work, there are 109 proteins that need Dnak/GroEL molecular chaperones to fold correctly, and the remaining 39 proteins that can fold autonomously. In this work, eight kinds of features are selected through 16 feature selection methods, and the selected features are input into KNN to predict protein disorder. The quality of feature selection methods is evaluated based on the accuracy of prediction, which are represented in Figure 5 and Supplementary Figures 9–12.

Figure 5 and Supplementary Figures 9–12 show that when selecting the top 10 and 20 features, the accuracy of GO feature selection using nonlinear SVM is improved by 13.16% and 14.48%. When selecting the first 30 and 50 features, the accuracy of using sequentialfs to select RCTD

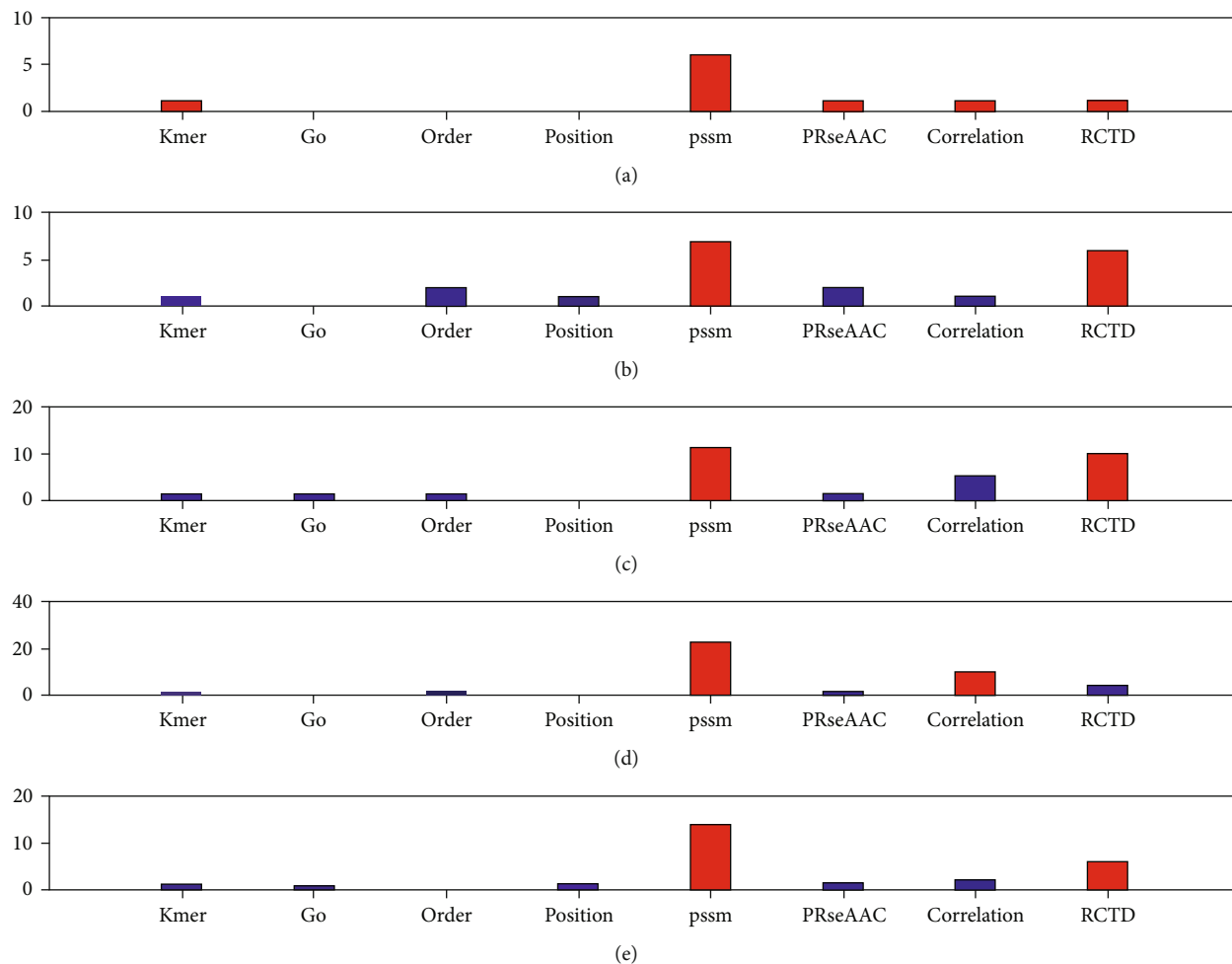


FIGURE 8: The number of 8 types of features in the top selected features in the protein structural data. From (a) to (e), it means that the number of selected features is 10 to 50, respectively.

TABLE 2: Time consumption of feature selection methods.

	Mutual Information (/S)	Sequentialfs (/S)	Linear-svm (/S)	Nonlinear-svm (/S)
PSSM	5.8	14074	23.8	2082.4
Go	360.33	—	42.4	0.75
RCTD	4.2	5571	11.7	1.3
Kmer	6.3	7423	18.9	1.7
PRseAAC	0.67	5.83	4.32	0.36
Order	1	35.2	22.8	270.5
Position	0.75	2.09	2.04	0.17
Correlation	0.62	3.87	1.39	0.33

features is improved by 13.16% and 17.17%. When selecting the first 40 features, linear SVM is used to select Kmer features, which improves its accuracy by 14.48%. Therefore, nonlinear SVM, sequentialfs and linear SVM are used to select features in the molecular chaperone prediction, which improves its accuracy by 13.16%~17.17%.

We also further compared the types of selected features. First, eight types of features are fused, and the fused features are selected through 16 feature selection methods, and the top 10-50 features are selected. Then, the number of eight types of features in the top 10-50 total selected features is counted, and the preference of eight types of features is evaluated by proportion. Figure 6 shows the number of 8 types of features in the top 10-50 total selected features in the protein disorder data set.

When selecting 10 comprehensive features, there are 5 RCTD features, accounting for 50%. When selecting 20-30 comprehensive features, PSSM features have an absolute advantage, with 19, 26, 39, and 47 selected, respectively. It can be seen that PSSM is the preferred feature if you want to check whether a protein sequence is self-folding or molecular chaperone to help complete the correct folding.

3.4. Comparison of Feature Selection in Protein Solubility Prediction. Finally, the efficiency of different feature selection methods in protein solubility prediction is discussed. In this work, more than 7000 proteins from *E. coli* were selected and sorted according to their solubility. The first

1000 protein sequences with higher solubility and the last 1000 protein sequences with the lowest solubility were taken out to form a protein sequence data set. Through 16 feature selection methods, 8 kinds of features are selected, respectively, and the selected features are input into KNN to predict the solubility of protein. The quality of feature selection methods is evaluated based on the accuracy of prediction, which are represented in Figure 7 and Supplementary Figures 13–16.

When selecting 10 and 20 features, using CIFEFS based on mutual information to select RCTD features, the accuracy is improved the most, which is 3.93% and 3.88%, respectively. When selecting 30 features, using sequentialfs to select RCTD features, the accuracy is improved by 3.12%. When 40 and 50 features are selected, the accuracy of nonlinear SVM is improved by 3.12% and 4.76%, respectively. The above results show that CIFEFS, sequentialfs and nonlinear SVM feature selection methods perform well in protein solubility prediction.

We also further compared the types of selected features. First, eight types of features are fused, and the fused features are selected through 16 feature selection methods, and the top 10-50 features are selected. Then, the number of eight types of features in the top 10-50 total selected features is counted, and the preference of eight types of features is evaluated by proportion. Figure 8 shows the number of 8 types of features in the top 10-50 total selected features in the protein disorder data set.

When selecting 10-50 comprehensive features, PSSM features always account for the most, with 6, 7, 11, 23 and 28 PSSM features, accounting for 60%, 35%, 36.67%, 50.75% and 56% of the total. Therefore, using PSSM characteristics as input features to predict the solubility of new protein sequences is more reliable [45].

3.5. Comparison of Calculation Efficiency of Various Methods. The above analysis shows that the nonlinear SVM feature selection method based on support vector machine performs well in the prediction of various protein structures and functions. In order to further study the computational efficiency of feature selection methods, we calculated the time-consuming of various feature selection methods to select 8 types of features, as shown in Table 2. Mutual information represents the average time of the nine selection methods. It is not difficult to find that the nonlinear SVM selection method is related to the size of matrix elements. The larger the data elements, the longer the time required. Therefore, the matrix is normalized before feature selection. Sequentialfs consumes the most time, and the time-consuming ratio of nonlinear SVM, linear SVM, and single mutual information selection method is 2.5: 27.5:1. Therefore, the nonlinear SVM selection method is the preferred feature selection method in the prediction of protein structure and function.

4. Conclusion

Feature selection can reduce the problem of over fitting, improves the performance of the model, and reduces the

time and space cost of the learning algorithm. 16 feature selection methods used in this work are feature selection method based on mutual information, feature selection method based on support vector machine, feature selection method based on genetic algorithm, feature selection method based on kurtosis and skewness, ReliefF, and sequentialfs information selection methods. Different feature selection methods were compared and analyzed in protein structure class prediction, protein disorder prediction, protein molecular chaperone prediction, and protein solubility prediction.

Through a comprehensive comparison and discussion, we found that nonlinear SVM feature selection method performs best in protein structure prediction, the first choice is order feature, followed by RCTD feature. In protein disorder prediction, SPECCMI_FS, CIFEFS, nonlinear SVM, and linear SVM feature selection methods can select core features from Kmer features, which improves its accuracy by 59.61%~71%. At the same time, order or RCTD features as input information will help to improve the accuracy of prediction. In protein molecular chaperone prediction, nonlinear SVM, sequentialfs, and linear SVM are used to select features, which improves the accuracy by 13.16%~17.17%, and the preferred feature is PSSM feature. In protein solubility prediction, CIFEFS, sequentialfs, and nonlinear SVM feature selection methods perform well, and PSSM is the preferred feature. These results can be regarded as some novel valuable guidelines for use of the feature selection method for protein structure and function prediction.

Data Availability

The data are available in <https://github.com/bioinfo0706/RaaMLab>.

Conflicts of Interest

The authors declare that there are no conflicts of interest, financial, or otherwise.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (62172369); Key Research and Development Plan of Zhejiang Province (2021C02039); and Zhejiang Provincial Natural Science Foundation of China (LY20F020016). The authors thank all the anonymous referees for their valuable suggestions and support.

Supplementary Materials

Supplementary Figures 1-16 are the precision comparison between support vector machine prediction and single class feature prediction based on the selected 20, 30, 40 and 50 features. (*Supplementary Materials*)

References

- [1] P. Klein and C. Delisi, "Prediction of protein structural class from the amino acid sequence," *Biopolymers*, vol. 25, no. 9, pp. 1659–1672, 1986.
- [2] K. C. Chou, "Structural bioinformatics and its impact to biomedical science and drug discovery," *Frontiers in Medicinal Chemistry*, vol. 3, no. 1, pp. 455–502, 2006.
- [3] C. Chothia and M. Levitt, "Structural patterns in globular proteins," *Nature*, vol. 261, no. 5561, pp. 552–558, 1976.
- [4] A. Andreeva, D. Howorth, S. E. Brenner, T. J. P. Hubbard, C. Chothia, and A. G. Murzin, "SCOP database in 2004: refinements integrate structure and sequence family data," *Nucleic Acids Research*, vol. 32, no. 90001, pp. 226D–2269, 2004.
- [5] A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia, "SCOP: a structural classification of proteins database for the investigation of sequences and structures," *Journal of Molecular Biology*, vol. 247, no. 4, pp. 536–540, 1995.
- [6] P. Ferragina, R. Giancarlo, V. Greco, G. Manzini, and G. Valiente, "Compression-based classification of biological sequences and structures via the universal similarity metric: experimental assessment," *BMC Bioinformatics*, vol. 8, no. 1, p. 252, 2007.
- [7] D. Qi and T. Wang, "Comparison study on k-word statistical measures for protein: from sequence to 'sequence space'," *BMC Bioinformatics*, vol. 9, no. 1, pp. 394–394, 2008.
- [8] C. Chen, Y. X. Tian, X. Y. Zou, P. X. Cai, and J. Y. Mo, "Using pseudo-amino acid composition and support vector machine to predict protein structural class," *Journal of Theoretical Biology*, vol. 243, no. 3, pp. 444–448, 2006.
- [9] K. C. Chou, "Prediction of protein structural classes and subcellular locations," *Current Protein & Peptide Science*, vol. 1, no. 2, pp. 171–208, 2000.
- [10] K. D. Kedarisetti, L. Kurgan, and S. Dick, "Classifier ensembles for protein structural class prediction with varying homology," *Biochemical and Biophysical Research Communications*, vol. 348, no. 3, pp. 981–988, 2006.
- [11] D. Qi, W. Li, and L. Li, "Improving protein structural class prediction using novel combined sequence information and predicted secondary structural features," *Journal of Computational Chemistry*, vol. 32, no. 16, pp. 3393–3398, 2011.
- [12] K. C. Chou, "A key driving force in determination of protein structural classes," *Biochemical & Biophysical Research Communications*, vol. 264, no. 1, pp. 216–224, 1999.
- [13] K. C. Chou and H. B. Shen, "Recent progress in protein subcellular location prediction," *Analytical Biochemistry*, vol. 370, no. 1, pp. 1–16, 2007.
- [14] R. Y. Luo, Z. P. Feng, and J. K. Liu, "Prediction of protein structural class by amino acid and polypeptide composition," *European Journal of Biochemistry*, vol. 269, no. 17, pp. 4219–4225, 2002.
- [15] X. D. Sun and R. B. Huang, "Prediction of protein structural classes using support vector machines," *Amino Acids*, vol. 30, no. 4, pp. 469–475, 2006.
- [16] S. Zhang, Y. Liang, and X. Yuan, "Improving the prediction accuracy of protein structural class: approached with alternating word frequency and normalized Lempel-Ziv complexity," *Journal of Theoretical Biology*, vol. 341, no. 1, pp. 71–77, 2014.
- [17] Y. S. Ding, T. L. Zhang, and K. C. Chou, "Prediction of protein structure classes with pseudo amino acid composition and fuzzy support vector machine network," *Protein and Peptide Letters*, vol. 14, no. 8, pp. 811–815, 2007.
- [18] L. Wu, Q. Dai, B. Han, L. Zhu, and L. Li, "Combining sequence information and predicted secondary structural feature to predict protein structural classes," in *2011 5th International Conference on Bioinformatics and Biomedical Engineering*, pp. 1–4, Wuhan, China, 2011 May 10.
- [19] B. Liao, Q. Xiang, and D. Li, "Incorporating secondary features into the general form of Chou's PseAAC for predicting protein structural class," *Protein & Peptide Letters*, vol. 19, no. 11, pp. 1133–1138, 2012.
- [20] L. Kong, L. Zhang, and J. Lv, "Accurate prediction of protein structural classes by incorporating predicted secondary structure information into the general form of Chou's pseudo amino acid composition," *Journal of Theoretical Biology*, vol. 344, no. 1, pp. 12–18, 2014.
- [21] M. S. Rahman, S. Shatabda, S. Saha, M. Kaykobad, and M. S. Rahman, "DPP-PseAAC: a DNA-binding protein prediction model using Chou's general PseAAC," *Journal of Theoretical Biology*, vol. 452, pp. 22–34, 2018.
- [22] Y. Zuo, Y. Li, Y. Chen, G. Li, Z. Yan, and L. Yang, "PseKRAAC: a flexible web server for generating pseudo K-tuple reduced amino acids composition," *Bioinformatics*, vol. 33, no. 1, pp. 122–124, 2017.
- [23] K. C. Chou and Y. D. Cai, "Prediction of protein subcellular locations by GO-FunD-PseAA predictor," *Biochemical and Biophysical Research Communications*, vol. 320, no. 4, pp. 1236–1239, 2004.
- [24] L. Kurgan, K. Cios, and K. Chen, "SCPRED: accurate prediction of protein structural class for sequences of twilight-zone similarity with predicting sequences," *BMC Bioinformatics*, vol. 9, no. 1, p. 226, 2008.
- [25] S. Zhang, S. Ding, and T. Wang, "High-accuracy prediction of protein structural class for low-similarity sequences based on predicted secondary structure," *Biochimie*, vol. 93, no. 4, pp. 710–714, 2011.
- [26] Q. Dai, Y. Li, X. Liu, Y. Yao, Y. Cao, and P. He, "Comparison study on statistical features of predicted secondary structures for protein structural class prediction: From content to position," *BMC Bioinformatics*, vol. 14, no. 1, pp. 1–4, 2013.
- [27] H. Ding, H. Lin, W. Chen et al., "Prediction of protein structural classes based on feature selection technique," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 6, no. 3, pp. 235–240, 2014.
- [28] C. Chen, L. X. Chen, X. Y. Zou, and P. X. Cai, "Predicting protein structural class based on multi-features fusion," *Journal of Theoretical Biology*, vol. 253, no. 2, pp. 388–392, 2008.
- [29] A. V. Kumar, R. F. M. Ali, Y. Cao, and V. V. Krishnan, "Application of data mining tools for classification of protein structural class from residue based averaged NMR chemical shifts," *Biochimica et Biophysica Acta*, vol. 1854, no. 10, pp. 1545–1552, 2015.
- [30] L. Nanni, S. Brahnam, and A. Lumini, "Prediction of protein structure classes by incorporating different protein descriptors into general Chou's pseudo amino acid composition," *Journal of Theoretical Biology*, vol. 360, pp. 109–116, 2014.
- [31] J. Wang, C. Wang, J. Cao, X. Liu, Y. Yao, and Q. Dai, "Prediction of protein structural classes for low-similarity sequences using reduced PSSM and position-based secondary structural features," *Gene*, vol. 554, no. 2, pp. 241–248, 2015.
- [32] I. Antes, S. W. I. Siu, and T. Lengauer, "DynaPred: a structure and sequence based method for the prediction of MHC class I binding peptide sequences and conformations," *Bioinformatics*, vol. 22, no. 14, pp. e16–e24, 2006.

- [33] P. Klus, B. Bolognesi, F. Agostini, D. Marchese, A. Zanzoni, and G. G. Tartaglia, "The cleverSuite approach for protein characterization: predictions of structural properties, solubility, chaperone requirements and RNA-binding abilities," *Bioinformatics*, vol. 30, no. 11, pp. 1601–1608, 2014.
- [34] G. Colonna, M. Costantini, and S. Costantini, "Frequencies of specific peptides in intrinsic disordered protein domains," *Protein & Peptide Letters*, vol. 17, no. 11, pp. 1398–1402, 2010.
- [35] B. Boeckmann, A. Bairoch, R. Apweiler et al., "The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003," *Nucleic Acids Research*, vol. 31, no. 1, pp. 365–370, 2003.
- [36] C. C. H. Chang, J. Song, B. T. Tey, and R. N. Ramanan, "Bioinformatics approaches for improved recombinant protein production in *Escherichia coli*: protein solubility prediction," *Briefings in Bioinformatics*, vol. 15, no. 6, pp. 953–962, 2014.
- [37] S. Idicula-Thomas and P. V. Balaji, "Understanding the relationship between the primary structure of proteins and their amyloidogenic propensity: clues from inclusion body formation," *Protein Engineering Design & Selection*, vol. 18, no. 4, pp. 175–180, 2005.
- [38] X. V. Nguyen, J. Chan, S. Romano, and J. Bailey, "Effective global approaches for mutual information based feature selection," in *Acm Sigkdd International Conference on Knowledge Discovery & Data Mining*, pp. 512–521, New York, New York, USA, 2014.
- [39] Y. Yang and L. Chen, "Identification of drug-disease associations by using multiple drug and disease networks," *Current Bioinformatics*, vol. 17, no. 1, pp. 48–59, 2022.
- [40] X. Li, L. Lu, and L. Chen, "Identification of protein functions in mouse with a label space partition method," *Mathematical Biosciences and Engineering*, vol. 19, no. 4, pp. 3820–3824, 2021.
- [41] X. Pan, L. Chen, Liu, Z. Niu, T. Huang, and Y. D. Cai, "Identifying protein subcellular locations with embeddings-based node2loc," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 1, pp. 1–675, 2021.
- [42] J. P. Zhou, L. Chen, and Z. H. Guo, "iATC-NRAKEL: an efficient multi-label classifier for recognizing anatomical therapeutic chemical classes of drugs," *Bioinformatics*, vol. 36, no. 5, pp. 1391–1396, 2020.
- [43] J. Van Durme, S. Maurer-Stroh, R. Gallardo, H. Wilkinson, F. Rousseau, and J. Schymkowitz, "Accurate prediction of DnaK-peptide binding via homology modelling and experimental data," *PLoS Computational Biology*, vol. 5, no. 8, article e1000475, 2009.
- [44] A. M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, and L. Serrano, "Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins," *Nature Biotechnology*, vol. 22, no. 10, pp. 1302–1306, 2004.
- [45] J. Winkelmann, G. Calloni, S. Campioni, B. Mannini, N. Taddei, and F. Chiti, "Low-level expression of a folding-incompetent protein in *Escherichia coli*: search for the molecular determinants of protein aggregation *in vivo*," *Journal of Molecular Biology*, vol. 398, no. 4, pp. 600–613, 2010.

Research Article

miR-302a-3p Promotes Radiotherapy Sensitivity of Hepatocellular Carcinoma by Regulating Cell Cycle via MCL1

Zifeng Yang ¹, Menglong Zhang ², Jian Zhang ³, Cunkun Chu ⁴, Bijuan Hu ⁵,
and Liyin Huang ⁵

¹Department of Interventional Radiology, The Fifth People's Hospital of Jinan, Jinan, 250000 Shandong, China

²Department of Minimally Invasive Intervention, Ganzhou People's Hospital, Ganzhou, Jiangxi 341000, China

³Department of Pathology, Ganzhou People's Hospital, Ganzhou, Jiangxi 341000, China

⁴Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan, Shandong 250000, China

⁵Department of Ultrasonography, Ganzhou People's Hospital, Ganzhou, Jiangxi 341000, China

Correspondence should be addressed to Menglong Zhang; gzinterv@163.com

Received 18 April 2022; Accepted 9 September 2022; Published 10 October 2022

Academic Editor: Tao Huang

Copyright © 2022 Zifeng Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. The relationship between tumor suppressor gene miR-302a-3p and radiotherapy for hepatocellular carcinoma (HCC) remains unclear. This study intended to illustrate the molecular mechanism how miR-302a-3p regulated radiotherapy sensitivity of HCC. **Methods.** miR-302a-3p expression in HCC tissues and cells was examined by qRT-PCR. The effect of miR-302a-3p on HCC radiotherapy sensitivity were detected by CCK-8, colony formation, and flow cytometry assays. The expression levels of cell cycle-related proteins were detected by Western blot. The influence of miR-302a-3p on radiotherapy sensitivity of HCC was further investigated via cell cycle inhibitor (Caudatin) treatment. The target gene (MCL1) of miR-302a-3p was obtained by bioinformatics analysis, and their binding relationship was confirmed by RNA-binding protein immunoprecipitation assay. The mechanisms of miR-302a-3p regulating cell cycle and affecting radiotherapy sensitivity of HCC cells through MCL1 were further explored through the rescue experiments. **Results.** miR-302a-3p expression was remarkably reduced in radiotherapy-resistant tissues and cells of HCC. miR-302a-3p overexpression restored sensitivity of radiotherapy-resistant HCC cells to radiotherapy. Treatment with cell cycle inhibitor Caudatin could reverse suppressive effect of miR-302a-3p downregulation on sensitivity of HCC to radiotherapy. Additionally, miR-302a-3p could restrain MCL1 expression. **In vitro** cell assays further revealed that miR-302a-3p/MCL1 axis could enhance radiotherapy sensitivity of HCC cells by inducing G0/G1 arrest. **Conclusions.** miR-302a-3p facilitated radiotherapy sensitivity of HCC cells by regulating cell cycle via MCL1, which provided a new underlying target for radiotherapy resistance of HCC patients.

1. Introduction

Hepatocellular carcinoma (HCC) ranks fifth in prevalence of all cancers and second in cancer-related deaths worldwide [1]. The prognosis of HCC was poor with a 5-year survival rate of less than 30% [2]. However, due to unapparent symptoms at the early stage, most HCC patients are detected at an advanced stage, at which time patients cannot tolerate a surgical resection [3]. Radiation therapy is especially important for patients who are not eligible for a resection. The mainstream radiotherapy currently is the radiation therapy with

a particle accelerator [4]. In addition, recent studies have shown that radioactive iodine-125 particle implantation therapy is an effective nonsurgical treatment for HCC patients who are unable to undergo resection. Iodine-125 particles are synthetic radionuclides that emit X- and γ -rays, which can damage tumor DNA and thus cause free radical production in tissues, which in turn kill tumor cells [5]. Currently, the feasible and well-tolerated radiotherapy has gradually become a noninvasive treatment for local ablation of HCC patients. However, radiotherapy-induced radiation resistance has seriously affected the control effect of

radiotherapy on tumor [6]. The generation of radiotherapy resistance is related to many biological factors; the specific mechanism remains unclear. Therefore, elucidating the molecular mechanisms involved in radiotherapy resistance may help to explore therapeutic targets to improve the effectiveness of radiotherapy, thus achieving better therapeutic outcomes in HCC patients.

MicroRNAs (miRNAs) play imperative regulatory roles in multiple biological and pathological processes of cancer [7]. Accumulating evidence has found that miRNAs affect radiotherapy resistance of various types of malignant tumors. For example, miR-612 was bound by TRPM2-AS in gastric cancer to increase FOXM1 expression and enhance radiotherapy resistance of gastric cancer [8]. MiR-208a could promote proliferation and radiotherapy resistance of lung cancer cells through targeting p21 [9]. In addition, miR-302a-3p can be a tumor suppressor in endometrial cancer, non-small-cell lung cancer, and melanoma [10–12], suggesting a tight relationship between miR-302a-3p and antitumor radiotherapy resistance. However, no studies have been reported on regulation of radiotherapy resistance of HCC via miR-302a-3p. Therefore, the present study preliminarily explored the potential impact and mechanism of miR-302a-3p on HCC radiotherapy resistance.

DNA damage response induced by ionizing radiation is a highly complex and coordinated system. Ionizing radiation can affect cell cycle progression by activating DNA damage checkpoints, which are specific points that prevent or slow down the cell from entering the next stage in the cell cycle [13]. The G2/M checkpoint is modulated by many proteins in eukaryotic cells, like cell division cycle 2 protein (Cdc2) and cyclin B protein (cyclin B), and reduced expression of Cdc2 and cyclin B can trigger G2/M arrest [14]. It is noteworthy that the radiosensitivity of cells shows varying characteristics in different cell cycle phases [15]. Specifically, G2/M phase is the most sensitive stage to radiation, while cells in S phase are the most resistant to radiation. Therefore, drugs that alter the course of the cell cycle are usually effective radiotherapy modulators [16, 17]. In addition, cyclin-dependent kinases (CDKs) have been shown to drive normal cells from G1 phase into the cell cycle process. Storch and Cordes [18], for example, found that CDK9 loss delayed cell transition from G1 to S phase, thereby enhancing radiotherapy sensitivity of head and neck squamous cell carcinoma cells. It seems that cell cycle condition correlates much to the efficacy of radiotherapy. Therefore, we assumed that miR-302a-3p enhances radiotherapy sensitivity by modulating cell cycle in HCC.

Herein, we revealed that miR-302a-3p expression is related to tumor radiotherapy resistance, and further cell experiments found that miR-302a-3p could induce G0/G1 arrest and enhance the radiotherapy sensitivity of HCC. Our study experimentally proved that miR-302a-3p regulated cell cycle progression and facilitated radiotherapy sensitivity of HCC by downregulating MCL1 expression. These findings shed new insights into the mechanism of radiotherapy resistance of HCC and provided theoretical basis for miR-302a-3p as a promising radiotherapy sensitization target.

2. Materials and Methods

2.1. Bioinformatics Analysis. The mRNA expression profile of HCC was obtained from TCGA database, and then, the TargetScan (http://www.targetscan.org/vert_72/), miRDB (<http://mirdb.org/>), and starBase (<http://starbase.sysu.edu.cn/>) databases were used to predict downstream target genes of miR-302a-3p to obtain the differentially expressed mRNAs with binding site to miR-302a-3p. The predicted results were intersected with differentially downregulated mRNAs. Finally, the target gene was determined through correlation analysis.

2.2. Patient Tissues. With the approval of the Ethics Committee of Ganzhou People's Hospital, we collected 30 HCC tissue samples from patients who were sensitive or resistant to radiotherapy. The mean age of the patients included in this study was 60.26 years, including 13 females and 17 males. All the patients were in intermediate and advanced clinical stages. Cancer tissue sample collection was performed after the patients received the first radiotherapy. All samples were from patients receiving radical radiotherapy without a distant metastasis in Ganzhou People's Hospital. Samples were taken during surgery and then quickly frozen and stored at -80°C . Radiotherapy was performed according to standard treatment regimens, and efficacy was assessed at the end of radiotherapy to determine whether patients (and corresponding specimens) are resistant or sensitive to radiotherapy. All patients had signed the informed consents.

2.3. Cell Culture and Radiation Treatment. Human HCC cell line HepG2 (BNCC338070) was provided by BeNa Culture Collection (BNCC, China). HepG2 cells were cultured in DMEM +10% fetal bovine serum (FBS) medium and placed in an incubator at 37°C with 5% CO_2 . The establishment of radiation-resistant HCC cells was performed as per the method described by Chen et al. [19]. Specifically, the HepG2 cells were treated with 6-MV X-rays generated by a linear accelerator (Varian 2300EX, dose rate of 2 Gy/min; Varian, USA). The cells were cultured after the initial radiation of 2 Gy and subcultured twice. The surviving cells were then exposed to a series of gradually increasing radiation doses (4, 6, 8, and 10 Gy) twice. The total radiation dose was 50 Gy, and the whole selection process was finished within 6 months. The final surviving cells were defined as HepG2/RR (radiation-resistant). We used parental HepG2 cells as controls and were defined as HepG2/RS (radiation-sensitive). HepG2/RR and HepG2/RS were cultured at 37°C for 24 h before further analyses.

2.4. Cell Transfection. miR-302a-3p mimic (miR-mimic), miR-302a-3p inhibitor (miR-inhibitor), and their negative controls (mimic-NC or NC-inhibitor) were all provided by Ribobio (China). The plasmids of miR-mimic/miR-NC, miR-inhibitor/NC-inhibitor, and oe-NC/oe-MCL1 were transfected into HCC cell line HepG2/RR by using Lipofectamine 2000 kit (Invitrogen, USA). The cells were then treated with X-rays and cultured for 24 h for following experiments.

TABLE 1: Primer sequence for qRT-PCR analysis.

Gene	Sequence	
miR-302a-3p	Forward primer	5'-ACACUCCAGCUGGGAGUGGUUUUGUACCUUC-3'
	Reverse primer	5'-CUCAACUGGUGUCGUGGAGUCGGCAAUUCAGUUGAGUCGUGAAU-3'
U6	Forward primer	5'-CTCGCTTCGGCAGCAC-3'
	Reverse primer	5'-AACGCTTCACGAATTTGCGT-3'
MCL1	Reverse primer	5'-GGGCAGGATTGTGACTCTCATT-3'
	Forward primer	5'-GATGCAGCTTTCTTGTTTATGG-3'
GAPDH	Reverse primer	5'-AAGGTGAAGGTCGGAGTCAA-3'
	Forward primer	5'-AATGAAGGGTCATTGATGG-3'

2.5. Cell Cycle Inhibitors Treat Cells. Caudatin (Medherb Biotechnology, China) was dissolved in dimethyl sulfoxide, and HepG2/RR cells were treated with 100 μ M Caudatin for 24 h for subsequent experiments.

2.6. qRT-PCR. Total RNA was separated with RNAiso Plus (Takara, Japan), and Prime Script TM RT Master Mix (Takara, Japan) was recommended for reverse transcription. qPCR was completed on SYBR Premix Ex Taq II (Takara, Japan). The operation for RNA extraction and PCR analysis were carried out as per the methods described by predecessors [20]. U6 snRNA and GAPDH were served as internal references to quantify miRNA and mRNA in cells. qPCR was performed with StepOnePlus Real-Time PCR system (AB, USA). The PCR results were quantified by $2^{-\Delta\Delta C_t}$, and the experiment was repeated 3 times. Table 1 exhibits the primer sequences.

2.7. Western Blot. The required proteins were extracted from cells, and the protein concentration was determined by BCA kit (Beyotime, China). The same amount of proteins (50 μ g) was added to each well of SDS-PAGE gel. After electrophoresis for 1.5 h, the protein was transferred to PVDF membrane. The PVDF membrane was placed in 0.5% skim milk powder and blocked at room temperature for 1.5 h. The PVDF membrane was then probed with the primary antibodies and incubated overnight (4°C). After washing three times with TBST, the membrane was placed in the second antibody and incubated for 2 h (room temperature). ECL kit (Thermo Scientific, USA) was used for color developing. Detailed steps of Western blot were performed according to previously described methods [21], and the experiment was repeated 3 times. The primary antibodies included rabbit anti-cyclin D1 (ab134175), rabbit anti-cyclin E1 (ab224819), rabbit anti-P27 (ab137736), and rabbit anti-GAPDH (ab181602), and the secondary antibody was goat anti-rabbit IgG (ab205718), all purchased from Abcam (Abcam, UK).

2.8. CCK-8 Assay. To investigate sensitivity of HepG2/RS and HepG2/RR cells to radiotherapy, cells that had received different ionizing radiation doses (0, 2, 4, 6, 8, and 10 Gy) were inoculated into a 96-well plate (2×10^4 cells/well). After 24 h of culture, CCK-8 reagent was added as per the instruc-

tions of CCK-8 kit (Beyotime, China), and absorbance of cells was determined at 450 nm.

To detect sensitivity of miR-302a-3p-transfected HepG2/RR cells to radiotherapy, cell viability was detected on d1, d2, d3, d4, d5, d6, d7, and d8 according to kit instructions after 8 Gy radiation treatment, and the experiment was repeated 3 times.

2.9. Colony Formation Assay. Transfected or untransfected HepG2/RR cells were plated into a 6-well plate (1×10^3 cells per well). The cells received different ionizing radiation doses (0, 2, 4, 6, 8, and 10 Gy). Culture medium was replaced after 24 h, and then, the cells were maintained for 10-14 days. When visible cell colonies appeared, the medium was discarded. The cells were rinsed 3 times with cold PBS, followed by fixing with paraformaldehyde for 15 min and staining with 0.5% crystal violet solution for 15 min. After washing with PBS, cells were dried, and colonies were photographed and counted using camera.

2.10. Flow Cytometry. For cell cycle detection, transfected cells in each group were cultured to logarithmic growth phase. After receiving 8 Gy ionizing radiation and being cultured for 24 h, the cells were digested with trypsin, washed with PBS, and fixed in 70% ethanol overnight. Then, the cells were dyed with propidium iodide. The distribution of cell cycle was analyzed by flow cytometry.

2.11. RIP Assay. RIP assay was carried out according to the steps of Zhang et al. [22]. In brief, HepG2 cells were treated with miR-NC or miR-mimic. 48 h later, the transfected cells were subjected to RIP assay with the Magna RIP™ RNA Binding Protein Immunoprecipitation Kit (Millipore, USA). Subsequently, the cells were probed with anti-Ago2 antibody or negative control IgG. The relative enrichment degree of MCL1 was determined by qRT-PCR.

2.12. Data Analysis. Data graphs and data analysis in this paper were plotted and analyzed using GraphPad Prism 8 (GraphPad Software, USA), and intergroup comparison was performed using ANOVA or *t*-test. The data were presented as the mean \pm standard deviation (SD) from at least three independent experiments. $p < 0.05$ meant that the difference was significant.

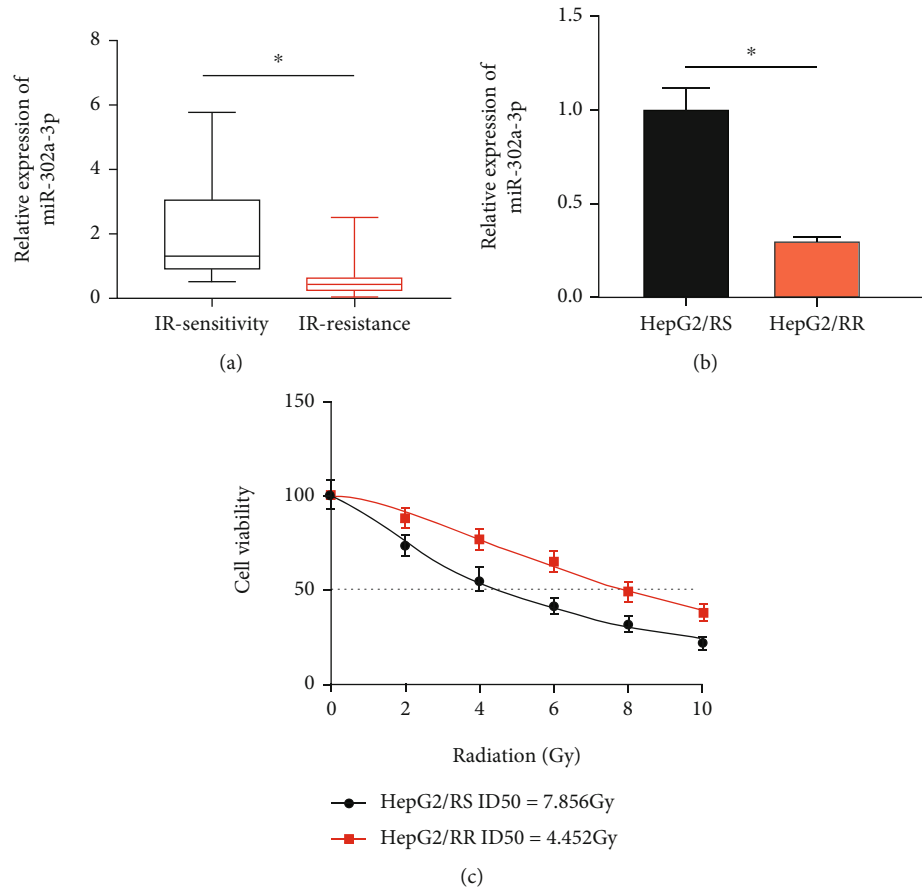


FIGURE 1: miR-302a-3p was underexpressed in radiotherapy-resistant HCC tissues and cells. (a) miR-302a-3p expression level in radiation-sensitive and radiation-resistant HCC tissues; (b) miR-302a-3p expression level in HepG2/RS and HepG2/RR cells; (c) cell viability of HepG2/RS and HepG2/RR cells under different radiation doses. * represents $p < 0.05$.

3. Results

3.1. miR-302a-3p Was Lowly Expressed in Radiotherapy-Resistant HCC Tissues and Cells. To study the role of miR-302a-3p in regulating HCC resistance to radiotherapy, we first measured the expression of miR-302a-3p expression in tumor tissues of HCC patients with radiotherapy sensitivity or radiotherapy resistance. The result displayed that miR-302a-3p was dramatically underexpressed in tumor tissues with radiotherapy resistance (Figure 1(a)). Then, HepG2 cells were used to establish radiotherapy-sensitive HepG2/RS and radiotherapy-resistant HepG2/RR. qRT-PCR was utilized to detect miR-302a-3p level in HepG2/RS and HepG2/RR cells. The result displayed a notable reduction in miR-302a-3p expression in radiotherapy-resistant HepG2/RR cells compared to its parental radiotherapy-sensitive HepG2/RS cells (Figure 1(b)). Subsequent CCK-8 result displayed that radiotherapy resistance of HepG2/RR cells was significantly enhanced compared with HepG2/RS, indicating that HepG2/RR cells obtained higher radiotherapy resistance, with a median lethal radiation dose of 8 Gy (Figure 1(c)). In conclusion, we successfully constructed radiotherapy-sensitive HepG2/RS and radiotherapy-resistant HepG2/RR cells and found that miR-302a-3p was

dramatically underexpressed in HCC tissues and cells with radiotherapy resistance.

3.2. miR-302a-3p Overexpression Enhanced Radiotherapy Sensitivity of HCC. To preliminarily explore effects of miR-302a-3p on cell viability, survival rate, and cycle of radiotherapy-resistant HCC cells, we constructed HepG2/RR cells with overexpression miR-302a-3p (miR-mimic) and its negative control (mimic-NC). First, the transfection efficiency of miR-302a-3p was evaluated by qRT-PCR, and the result indicated a prominent increase in miR-302a-3p expression in miR-mimic group (Figure 2(a)). Then, the cell viability was detected by CCK-8. It was demonstrated that miR-302a-3p overexpression remarkably reduced viability of HepG2/RR cells compared with control group (Figure 2(b)). Besides, the result of colony formation assay demonstrated that survival rate of HepG2/RR cells with miR-302a-3p mimic was notably reduced (Figure 2(c)) after cells were exposed to radiation, indicating that radiotherapy sensitivity of HepG2/RR cells was improved. Then, flow cytometry was adopted to detect cell cycle distribution in each treatment group. It was found that cells in G0/G1 phase were prominently increased and those in S phase were remarkably decreased in miR-mimic group (Figure 2(d)).

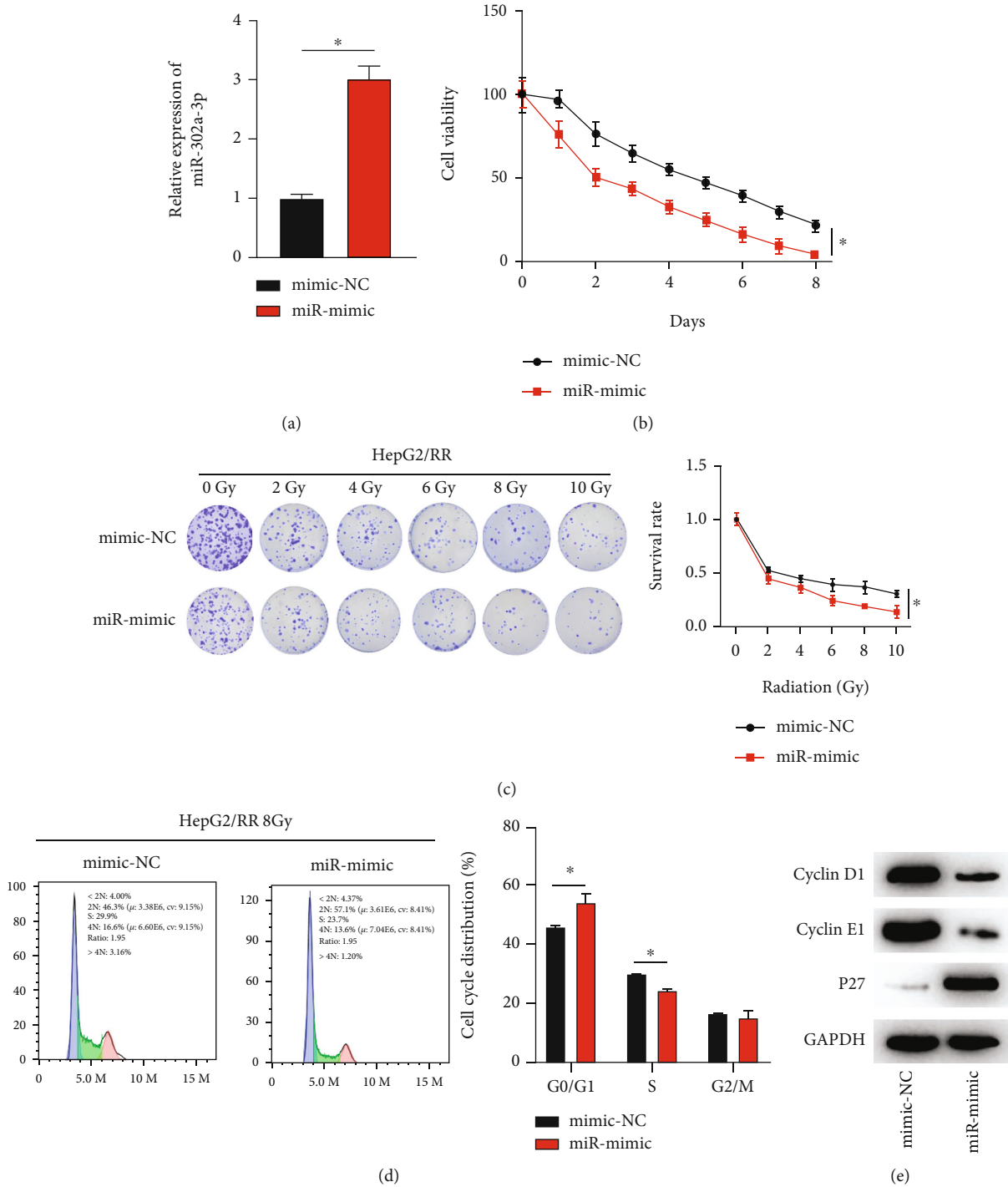
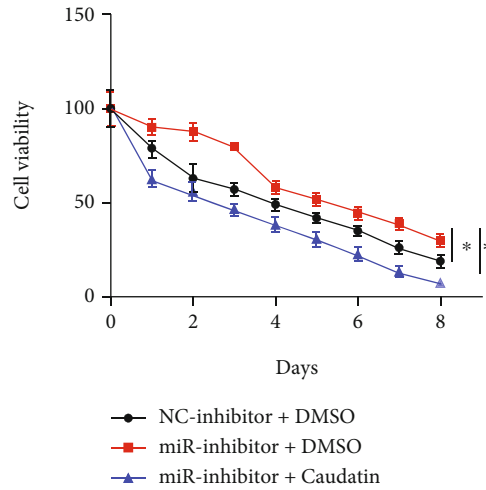


FIGURE 2: miR-302a-3p overexpression enhanced the radiotherapy sensitivity of HCC cells. (a) miR-302a-3p expression in HepG2/RR cells with miR-mimic/mimic-NC; (b) cell viability of HepG2/RR cells treated for different days with 8 Gy radiation dose; (c) survival rate of HepG2/RR cells treated with different radiation doses; (d) cell cycle distribution of HepG2/RR cells treated with 8 Gy radiation doses; (e) expression of cyclins in HepG2/RR cells. * represents $p < 0.05$.

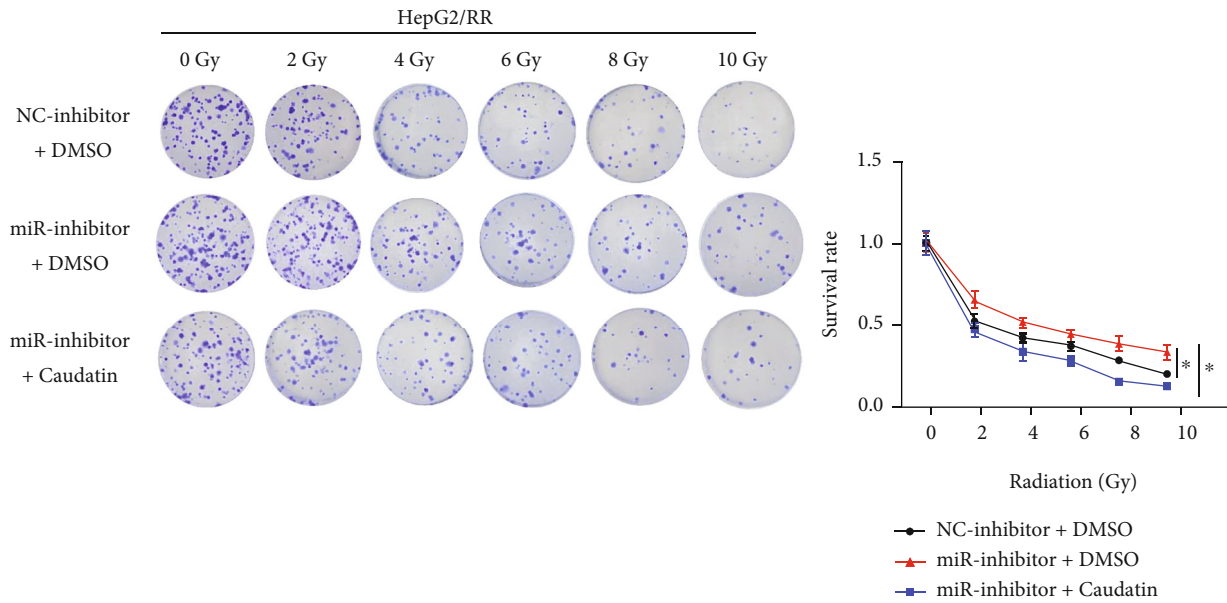
Western blot analysis showed that expression levels of cyclin D1 and cyclin E1 were significantly downregulated, while expression of cell cycle regulator P27 was remarkably upregulated after miR-302a-3p overexpression (Figure 2(e)). In conclusion, these results indicated that overexpression of

miR-302a-3p enhanced radiotherapy sensitivity, inhibited cell viability, and led to cell cycle arrest at G0/G1 phase.

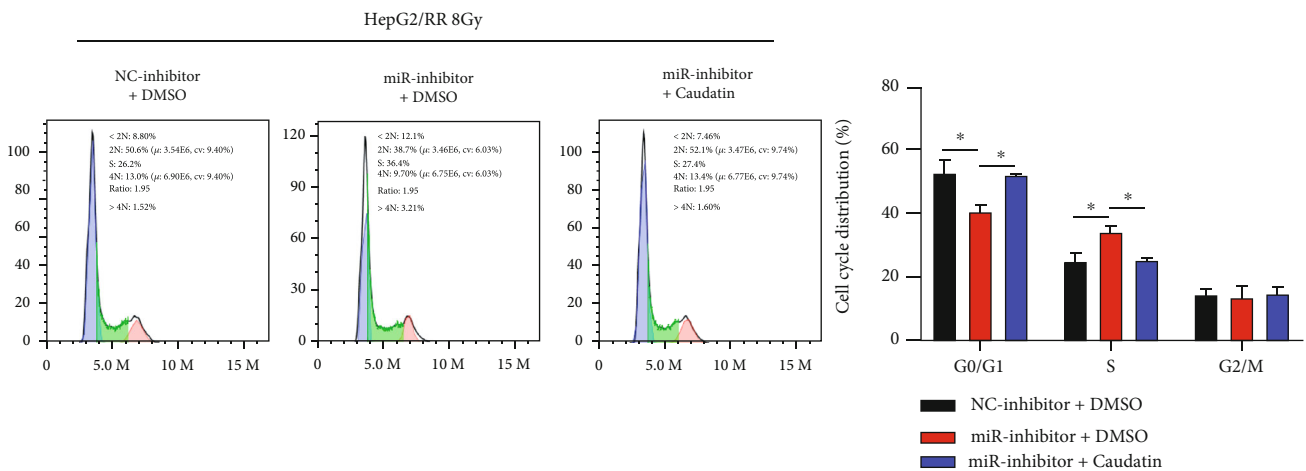
3.3. miR-302a-3p Enhanced Radiotherapy Sensitivity of HCC by Modulating Cell Cycle. To further analyze the impact of



(a)



(b)



(c)

FIGURE 3: Continued.

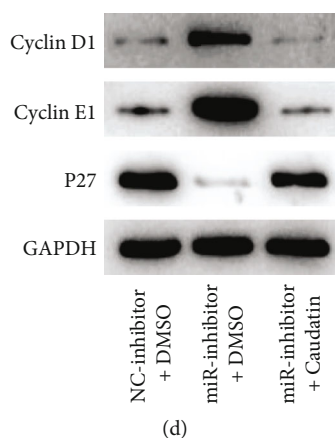


FIGURE 3: miR-302a-3p enhanced radiotherapy sensitivity of HCC cells by regulating cell cycle. (a) Cell viability of HepG2/RR cells treated for different days with 8 Gy radiation dose; (b) survival rate of HepG2/RR cells treated with different radiation doses; (c) cell cycle phase distribution of HepG2/RR cells treated with 8 Gy radiation doses; (d) expression of cyclins in HepG2/RR cells. * represents $p < 0.05$.

miR-302a-3p on sensitivity of HCC cells by affecting cell cycle, we set the following cell groups: negative control group (NC-inhibitor+DMSO), miR-302a-3p low expression group (miR-inhibitor+DMSO), and miR-302a-3p low expression with 100 μ M cell cycle inhibitor Caudatin group (miR-inhibitor+Caudatin). First, CCK-8 was utilized to detect effects of different treatments on cell viability. Compared to control group, the HepG2/RR cells with downregulated miR-302a-3p expression showed enhanced viability, which was reversed by the simultaneous addition of cell cycle inhibitor (Figure 3(a)). Then, colony formation assay was introduced to measure survival rate of HepG2/RR cells in each treatment group. The experimental result demonstrated that survival rate of HepG2/RR cells with downregulated miR-302a-3p was significantly increased after radiation exposure, while simultaneous addition of cell cycle inhibitor reversed effect of miR-302a-3p on survival rate of HepG2/RR cells (Figure 3(b)). Subsequently, flow cytometry was applied to analyze cell cycle distribution of each treatment group. According to the results, the number of G0/G1 phase cells was notably reduced in miR-inhibitor+DMSO group and prominently increased in miR-inhibitor+Caudatin group compared to NC-inhibitor+DMSO group (Figure 3(c)). Additionally, Western blot was applied to evaluate expression of cell cycle-related proteins. According to the experimental results, cyclin D1 and cyclin E1 expression was upregulated, and P27 expression was downregulated in miR-inhibitor+DMSO group compared to NC-inhibitor+DMSO group, while cyclin D1 and cyclin E1 expression was downregulated, and P27 expression was upregulated in miR-inhibitor+Caudatin group compared to miR-inhibitor+DMSO group (Figure 3(d)). These results implied that miR-302a-3p could make HCC cells sensitive to radiotherapy by regulating cell cycle.

3.4. miR-302a-3p Enhanced Radiotherapy Sensitivity of HCC Cells by Regulating Cell Cycle via MCL1. Studies have shown that miRNA mainly participates in the molecular regulatory pathway by regulating expression of target genes. We further explored downstream target mRNAs of miR-302a-3p

through bioinformatics methods. First, differential analysis on the mRNAs in TCGA-HCC database were performed. Then, the obtained differentially downregulated mRNAs were intersected with the target genes of miR-302a-3p predicted by TargetScan, starBase, and miRDB databases. By this way, differential mRNAs with binding site to miR-302a-3p were obtained (Figure 4(a)). After reviewing the literature, we found that MCL1 was a key protein related to cell cycle [23] and that MCL1 was significantly overexpressed in tumor tissues with radiotherapy resistance (Figure 4(b)). Therefore, we selected MCL1 as the research object to explore its influence on the radiotherapy sensitivity of HCC cells. Subsequently, the binding site between miR-302a-3p and MCL1 was predicted by bioinformatics analysis (Figure 4(c)). The binding relationship between miR-302a-3p and MCL1 was further verified by RIP experiment, and the result showed that MCL1 was significantly enriched in miR-mimic cells (Figure 4(d)). To investigate whether miR-302a-3p can affect cell cycle progression by targeting MCL1, we first constructed miR-302a-3p overexpression cells (miR-mimic+oe-NC), MCL1 overexpression cells (mimic-NC+oe-MCL1), and miR-302a-3p and MCL1 simultaneous overexpression cells (miR-mimic+oe-MCL1) using HepG2/RR cells. According to qRT-PCR result, the mRNA expression of MCL1 was significantly reduced in miR-mimic+oe-NC group in comparison with control group, but it was recovered in miR-mimic+oe-MCL1 cotransfection group (Figure 4(e)). With regard to CCK-8 result, the viability of HepG2/RR cells in mimic-NC+oe-MCL1 group was enhanced compared to the mimic-NC+oe-NC, but it was restored in miR-mimic+oe-MCL1 group (Figure 4(f)). Then, flow cytometry result indicated the number of G0/G1 cells in mimic-NC+oe-MCL1 was significantly reduced compared to mimic-NC+oe-NC group, but the number in miR-mimic+oe-MCL1 group was recovered (Figure 4(g)). Western bolt was used to evaluate the expression of cell cycle-related proteins. According to the result, cyclin D1 and cyclin E1 expression was upregulated, and P27 expression was downregulated in mimic-NC+oe-MCL1 group compared to mimic-NC+oe-NC group, while

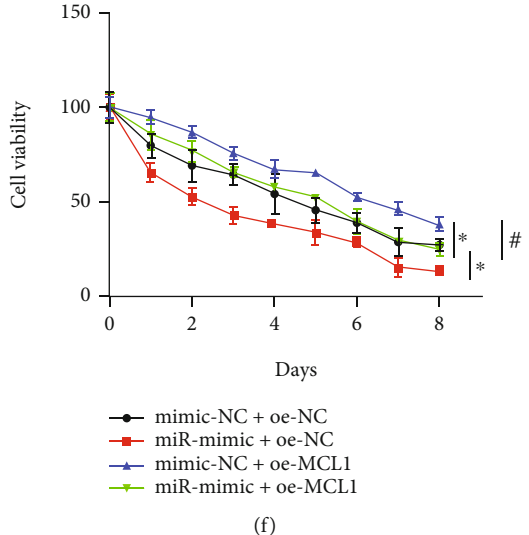
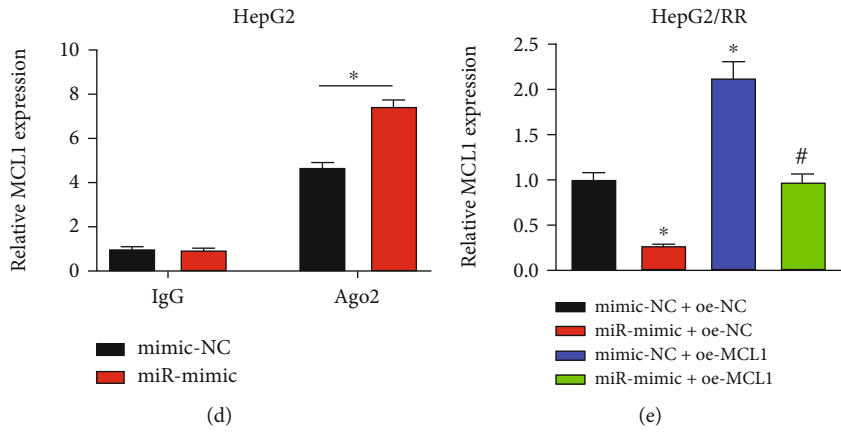
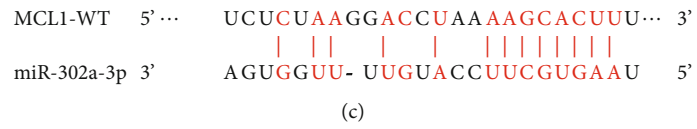
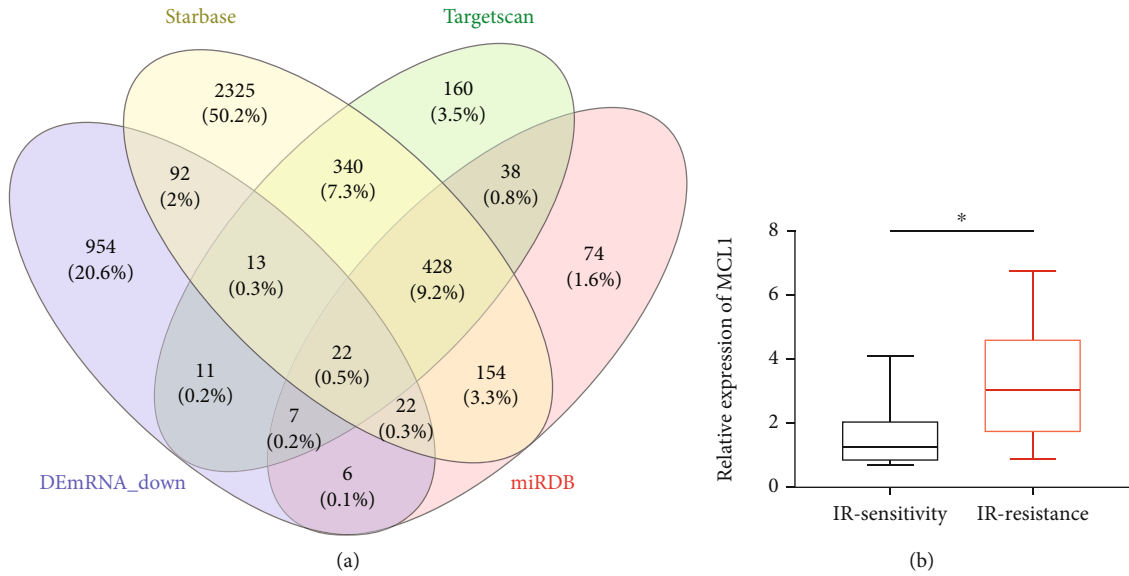


FIGURE 4: Continued.

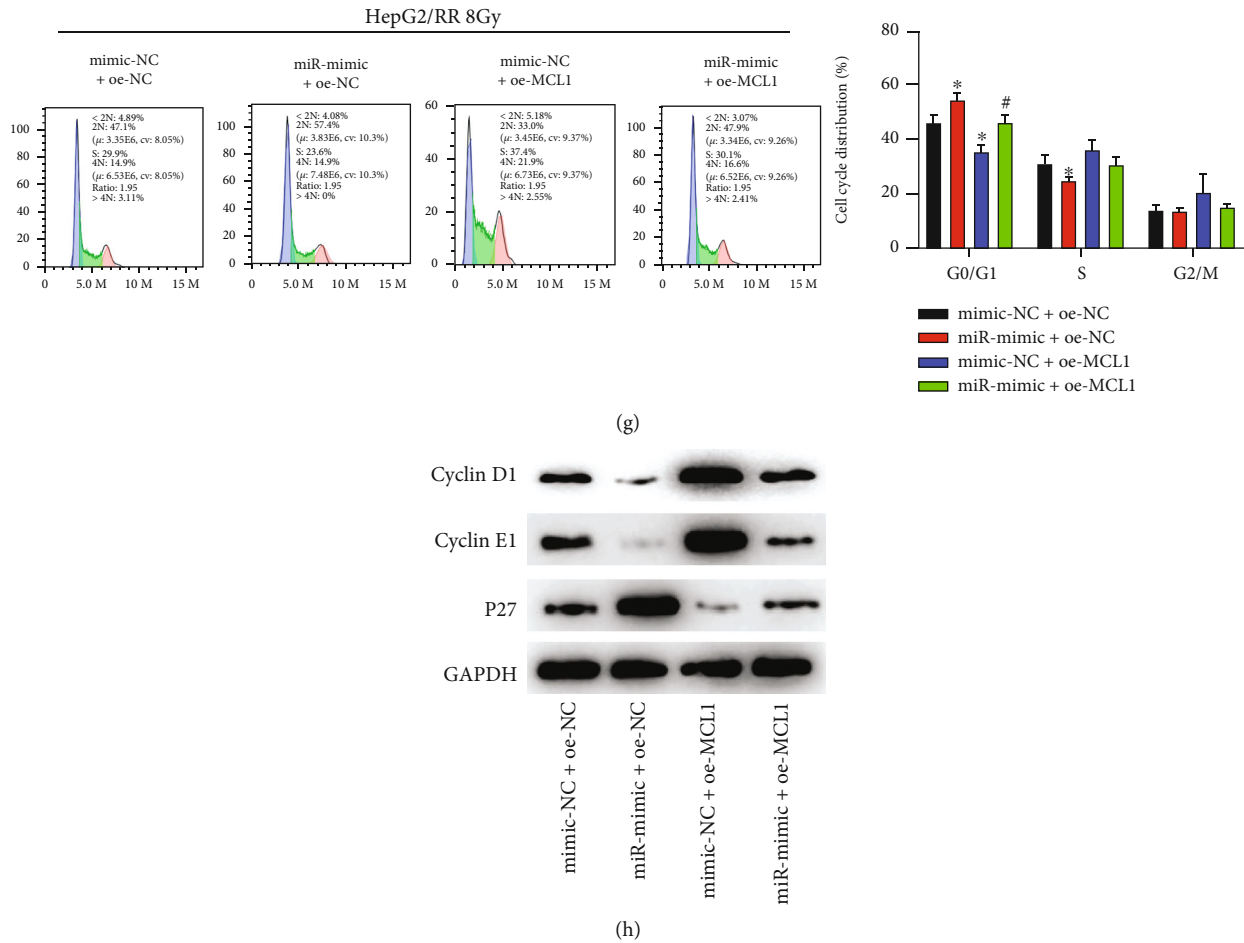


FIGURE 4: miR-302a-3p enhanced radiotherapy sensitivity of HCC cells by regulating cell cycle via MCL1. (a) Venn diagram of predicted mRNAs of miR-302a-3p and downregulated differential mRNAs; (b) MCL1 expression in radiation-sensitive and radiation-resistant tissues of HCC patients; (c) schematic diagram of the binding between MCL1 and miR-302a-3p sequences predicted by starBase; (d) binding relationship between miR-302a-3p and MCL1 verified by RIP experiment; (e) MCL1 mRNA expression of HepG2/RR cells in different treatment groups; (f) cell viability of HepG2/RR cells treated for different days with 8 Gy radiation dose; (g) cell cycle distribution of HepG2/R cells treated with 8 Gy radiation dose; (h) the expression of cyclins in HepG2/RR cells. * $p < 0.05$ vs. (mimic-NC + oe-NC), # $p < 0.05$ vs. (mimic-NC+oe-MCL1).

expression levels of these proteins were recovered in miR-mimic+oe-MCL1 group (Figure 4(h)). These results indicated that miR-302a-3p could promote radiotherapy sensitivity of HCC cells by regulating HepG2/RR cell cycle progression via downregulation of MCL1.

4. Discussion

Radiotherapy resistance is an important factor leading to clinical radiotherapy failure in HCC patients. Radiotherapy resistance can be attributed to the inherent radioresistance of tumor cells in hypoxic microenvironment or to the resistance acquired during hyperfraction radiotherapy [24]. Studies have shown that radiation exposure can increase the levels of intracellular free radical species, cause DNA strand breaks, and lead to subsequent dysfunction of some organelles such as mitochondria and endoplasmic reticulum [25]. These radiation-induced cellular events facilitate proapoptotic signal activation and ultimately cause tumor cell

killing [26]. Nevertheless, the key molecules involved in radiation-induced radiotherapy resistance are still poorly understood. Hence, it is an urgent task to study the molecular mechanism of radiotherapy resistance and find new therapeutic targets, which are of great importance to overcome cancer radiotherapy resistance.

The regulatory mechanisms of miRNAs in the progression of many complex diseases have been extensively studied. miR-302a-3p plays a potential cancer-promoting role. For example, Zhang et al. [27] found that miR-302a-3p directly targets SOCS5 to promote STAT3 phosphorylation and induce transcription of STAT3 target genes, thereby promoting metastasis of pancreatic cancer cells. However, miR-302a-3p acts as an important tumor suppressor in most cases, where it inhibits the biological processes like proliferation, invasion, and migration of various human cancers, including HCC, colon cancer, and gastric cancer [28–30]. It is worth noting that miR-302a, as the precursor of miR-302a-3p, has been proved to participate in the inhibition of

tumor chemotherapy resistance development [31, 32]. In addition, Liang et al. [33] discovered that miR-302a sensitizes radiation-resistant breast cancer cells to radiotherapy both *in vivo* and *in vitro*. Yu et al. [34] obtained similar results and found that miR-302a overexpression hampers proliferation of non-small-cell lung cancer, promotes cell apoptosis, and reduces cell radioresistance. Therefore, it is reasonable to speculate that miR-302a-3p is associated with HCC radiotherapy resistance. In this study, miR-302a-3p was remarkably downregulated in radiotherapy-resistant cells (HepG2/RR) compared with radiotherapy-sensitive cells (HepG2/RS). However, miR-302a-3p overexpression could enhance the radiotherapy sensitivity of HCC cells, confirming that miR-302a-3p mediated radiotherapy resistance in HCC.

Radiation resistance is a complex cellular response involving many signaling pathways and genes. Ionizing radiation can induce DNA damage, including DNA single-strand break, DNA base damage, and DNA double-strand break [35]. In addition, DNA damage induced by ionizing radiation can activate a series of cell cycle checkpoints [36]. Hence, cell cycle checkpoint block is involved in the regulation of tumor radiotherapy resistance. For example, Wang et al. [37] showed that cyclin D1 and cyclin E1 are specific targets of miR-16-5p and that miR-16-5p overexpression can downregulate expression of cyclin D1 and cyclin E1 and induce cell cycle arrest in G0/G1 phase, which enhances radiosensitivity of prostate cancer cells. Although numerous studies have revealed the role of periodic checkpoint block in tumor radiotherapy resistance, there is little research on the regulatory role of miR-302a-3p. In the present study, *in vitro* cell experiments confirmed that miR-302a-3p induced cell arrest in G0/G1 phase by regulating expression of key cell cycle proteins cyclin D1, cyclin E1, and P27 and promoted the sensitivity of radiation-resistant HCC cells to radiotherapy, which was consistent with previous studies. In addition, this study predicted a new downstream target of miR-302a-3p, MCL1, through bioinformatics analysis, and verified the targeted relationship between the two. MCL1 has been proved to be overexpressed in many human cancers and contributes to cancer occurrence and inhibits apoptosis [38]. Specific targeting of MCL1 may overcome the antiapoptotic ability of malignant tumor cells. For example, BAG3 can upregulate MCL1 by downregulating miR-29b, thereby inducing chemotherapy resistance to paclitaxel in ovarian cancer [39]. In addition, Yu et al. found that MCL1 overexpression significantly inhibited mulanin-induced autophagy and cell cycle arrest in colorectal cancer cells [40]. In the present study, rescue experiments demonstrated that miR-302a-3p induced cell cycle arrest and promoted sensitivity of HCC cells to radiotherapy by targeting MCL1 expression, which was in accordance with the results of previous reports.

Our results confirmed that miR-302a-3p could enhance radiotherapy sensitivity of radiation-resistant HCC cells by regulating the cell cycle. In addition, our study revealed influences of miR-302a-3p/MCL1 axis on the radiotherapy sensitivity of HCC cells for the first time. However, the mechanism of miR-302a-3p enhancing the radiotherapy

sensitivity of HCC was analyzed at the cellular level, which has not been verified *in vivo* at the animal level. Meanwhile, it also lacked the exploration of relevant signal pathways, which was the deficiency of this study. In summary, this study for the first time clarified the modulatory mechanism of miR-302a-3p in HCC radiotherapy resistance, suggesting that miR-302a-3p may be a potential sensitizer of radiotherapy. Our results provided new evidence for the radiotherapy resistance mechanism of HCC cells and can help overcome the difficulties in cancer radiotherapy in the future.

Data Availability

The data that support the findings of this study are available on request from the corresponding author.

Ethical Approval

The study was approved by the Ethics Committee of Ganzhou People's Hospital. The methods were carried out in accordance with the approved guidelines.

Consent

Written informed consent was obtained prior to the study.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors' Contributions

Zifeng Yang, Menglong Zhang is responsible for the project development, data analysis, and manuscript writing; Jian Zhang for the data analysis and collection and manuscript writing; Cunkun Chu for the data analysis and collection and manuscript editing; and Bijuan Hu and Liyin Huang for the data analysis and manuscript editing.

Acknowledgments

This study was supported by the Science and Technology Project of Jiangxi Provincial Health Commission (project no. 202212514), the Key Research and Development Project of Ganzhou Science and Technology Bureau (project no. 2020-275), and the Doctoral Research Initiation Project of Ganzhou People's Hospital (project no. Bsqd2020002).

References


- [1] K. Ito, N. Takemura, F. Inagaki, F. Mihara, and N. Kokudo, "Difference in treatment algorithms for hepatocellular carcinoma between world's principal guidelines," *Global Health & Medicine*, vol. 2, pp. 282–291, 2020.
- [2] Y. Dong, F. Li, J. Wang et al., "miR-369 inhibits liver cancer progression by targeting ZEB1 pathway and predicts the prognosis of HCC patients," *Journal of Cancer*, vol. 12, no. 10, pp. 3067–3076, 2021.

- [3] H. C. Yu, M. H. Hung, Y. L. Chen et al., “Erlotinib derivative inhibits hepatocellular carcinoma by targeting CIP2A to reactivate protein phosphatase 2A,” *Cell Death & Disease*, vol. 5, no. 7, article e1359, 2014.
- [4] D. Citrin and E. Recent, “Recent developments in radiotherapy,” *The New England Journal of Medicine*, vol. 377, no. 11, pp. 1065–1075, 2017.
- [5] Q. H. Qin, B. S. Huang, Q. X. Tan, W. P. Yang, B. Lian, and C. Y. Wei, “Radiobiological effect induced by different activities of (125)I seed brachytherapy in a hepatocellular carcinoma model,” *International Journal of Clinical and Experimental Medicine*, vol. 7, no. 12, pp. 5260–5267, 2014.
- [6] J. Sun, Z. Zhu, W. Li et al., “UBE2T-regulated H2AX monoubiquitination induces hepatocellular carcinoma radioresistance by facilitating CHK1 activation,” *Journal of Experimental & Clinical Cancer Research*, vol. 39, no. 1, p. 222, 2020.
- [7] J. Kim, F. Yao, Z. Xiao, Y. Sun, and L. Ma, “MicroRNAs and metastasis: small RNAs play big roles,” *Cancer Metastasis Reviews*, vol. 37, no. 1, pp. 5–15, 2018.
- [8] J. Xiao, L. Lin, D. Luo et al., “Long noncoding RNA TRPM2-AS acts as a microRNA sponge of miR-612 to promote gastric cancer progression and radioresistance,” *Oncogene*, vol. 9, no. 3, p. 29, 2020.
- [9] Y. Tang, Y. Cui, Z. Li et al., “Erratum to: radiation-induced miR-208a increases the proliferation and radioresistance by targeting p21 in human lung cancer cells,” *Journal of Experimental & Clinical Cancer Research*, vol. 35, no. 1, p. 20, 2016.
- [10] X. Pan, D. Li, J. Huo, F. Kong, H. Yang, and X. Ma, “LINC01016 promotes the malignant phenotype of endometrial cancer cells by regulating the miR-302a-3p/miR-3130-3p/NFYA/SATB1 axis,” *Cell Death & Disease*, vol. 9, no. 3, p. 303, 2018.
- [11] D. Wei, Y. Q. Ke, P. Duan, L. Zhou, C. Y. Wang, and P. Cao, “MicroRNA-302a-3p induces ferroptosis of non-small cell lung cancer cells via targeting ferroportin,” *Free Radical Research*, vol. 55, no. 7, pp. 722–731, 2021.
- [12] X. Chang, Y. Y. Lin, L. N. Bai, and W. Zhu, “miR-302a-3p suppresses melanoma cell progression via targeting METTL3,” *Journal of Chemotherapy*, vol. 34, no. 1, pp. 55–66, 2022.
- [13] B. Pauwels, A. Wouters, M. Peeters, J. B. Vermorken, and F. Lardon, “Role of cell cycle perturbations in the combination therapy of chemotherapeutic agents and radiation,” *Future Oncology*, vol. 6, no. 9, pp. 1485–1496, 2010.
- [14] B. P. Soule, N. L. Simone, W. G. DeGraff, R. Choudhuri, J. A. Cook, and J. B. Mitchell, “Loratadine dysregulates cell cycle progression and enhances the effect of radiation in human tumor cell lines,” *Radiation Oncology*, vol. 5, no. 1, p. 8, 2010.
- [15] X. Qian, C. Tan, B. Yang et al., “Astaxanthin increases radiosensitivity in esophageal squamous cell carcinoma through inducing apoptosis and G2/M arrest,” *Diseases of the Esophagus*, vol. 30, no. 6, pp. 1–7, 2017.
- [16] J. Liebmman, J. A. Cook, J. Fisher, D. Teague, and J. B. Mitchell, “In vitro studies of Taxol as a radiation sensitizer in human tumor cells,” *Journal of the National Cancer Institute*, vol. 86, no. 6, pp. 441–446, 1994.
- [17] Y. H. Cui, H. J. Liang, Q. Q. Zhang et al., “Radiosensitivity enhancement by arsenic trioxide in conjunction with hyperthermia in the EC-1 esophageal carcinoma cell line,” *Asian Pacific Journal of Cancer Prevention*, vol. 13, no. 4, pp. 1693–1697, 2012.
- [18] K. Storch and N. Cordes, “The impact of CDK9 on radiosensitivity, DNA damage repair and cell cycling of HNSCC cancer cells,” *International Journal of Oncology*, vol. 48, no. 1, pp. 191–198, 2016.
- [19] Y. Chen, Z. Shen, Y. Zhi et al., “Long non-coding RNA ROR promotes radioresistance in hepatocellular carcinoma cells by acting as a ceRNA for microRNA-145 to regulate RAD18 expression,” *Archives of Biochemistry and Biophysics*, vol. 645, pp. 117–125, 2018.
- [20] T. Li, Y. Xing, F. Yang et al., “lncRNA SNHG3 sponges miR-577 to up-regulate SMURF1 expression in prostate cancer,” *Cancer Medicine*, vol. 9, no. 11, pp. 3852–3862, 2020.
- [21] W. Chen, L. Zhai, H. Liu et al., “Downregulation of lncRNA ZFAS1 inhibits the hallmarks of thyroid carcinoma via the regulation of miR-302-3p on cyclin D1,” *Molecular Medicine Reports*, vol. 23, no. 1, 2021.
- [22] Y. Zhang, F. Wang, G. Chen, R. He, and L. Yang, “lncRNA MALAT1 promotes osteoarthritis by modulating miR-150-5p/AKT3 axis,” *Cell & Bioscience*, vol. 9, no. 1, p. 54, 2019.
- [23] L. Yi, X. X. Ji, H. Tan et al., “Involvement of Mcl1 in diallyl disulfide-induced G2/M cell cycle arrest in HL-60 cells,” *Oncology Reports*, vol. 27, no. 6, pp. 1911–1917, 2012.
- [24] Y. Tang, Y. Cui, Z. Li et al., “Radiation-induced miR-208a increases the proliferation and radioresistance by targeting p21 in human lung cancer cells,” *Journal of Experimental & Clinical Cancer Research*, vol. 35, no. 1, p. 7, 2016.
- [25] M. Toulany and H. P. Rodemann, “Phosphatidylinositol 3-kinase/Akt signaling as a key mediator of tumor cell responsiveness to radiation,” *Seminars in Cancer Biology*, vol. 35, pp. 180–190, 2015.
- [26] M. I. Koukourakis, “Radiation damage and radioprotectants: new concepts in the era of molecular medicine,” *The British Journal of Radiology*, vol. 85, no. 1012, pp. 313–330, 2012.
- [27] Z. Zhang, J. Li, H. Guo et al., “BRM transcriptionally regulates miR-302a-3p to target SOCS5/STAT3 signaling axis to potentiate pancreatic cancer metastasis,” *Cancer Letters*, vol. 449, pp. 215–225, 2019.
- [28] W. Li, W. Xu, J. S. Song, T. Wu, and W. X. Wang, “lncRNA SNHG16 promotes cell proliferation through miR-302a-3p/FGF19 axis in hepatocellular carcinoma,” *Neoplasma*, vol. 66, no. 3, pp. 397–404, 2019.
- [29] D. Ke, Q. Wang, S. Ke, L. Zou, and Q. Wang, “Long-non coding RNA SNHG16 supports colon cancer cell growth by modulating miR-302a-3p/AKT axis,” *Pathology Oncology Research*, vol. 26, no. 3, pp. 1605–1613, 2020.
- [30] C. Yang and S. Deng, “Mechanism of hsa-miR-302a-3p-targeted VEGFA in the inhibition of proliferation of gastric cancer cell,” *Sichuan Da Xue Xue Bao. Yi Xue Ban*, vol. 50, no. 1, pp. 13–19, 2019.
- [31] L. Sun, Y. Fang, X. Wang et al., “miR-302a inhibits metastasis and cetuximab resistance in colorectal cancer by targeting NFIB and CD44,” *Theranostics*, vol. 9, no. 26, pp. 8409–8425, 2019.
- [32] N. Liu, J. Li, Z. Zhao et al., “MicroRNA-302a enhances 5-fluorouracil-induced cell death in human colon cancer cells,” *Oncology Reports*, vol. 37, no. 1, pp. 631–639, 2017.
- [33] Z. Liang, J. Ahn, D. Guo, J. R. Votaw, and H. Shim, “MicroRNA-302 replacement therapy sensitizes breast cancer cells to ionizing radiation,” *Pharmaceutical Research*, vol. 30, no. 4, pp. 1008–1016, 2013.
- [34] Z. Yu, G. Wang, C. Zhang et al., “lncRNA SBF2-AS1 affects the radiosensitivity of non-small cell lung cancer via modulating

- microRNA-302a/MBNL3 axis,” *Cell Cycle*, vol. 19, no. 3, pp. 300–316, 2020.
- [35] T. Shimura, M. Fukumoto, and N. Kunugita, “The role of cyclin D1 in response to long-term exposure to ionizing radiation,” *Cell Cycle*, vol. 12, no. 17, pp. 2738–2743, 2013.
- [36] B. Wang, “Analyzing cell cycle checkpoints in response to ionizing radiation in mammalian cells,” *Methods in Molecular Biology*, vol. 1170, pp. 313–320, 2014.
- [37] F. Wang, A. Mao, J. Tang et al., “microRNA-16-5p enhances radiosensitivity through modulating cyclin D1/E1-pRb-E2F1 pathway in prostate cancer cells,” *Journal of Cellular Physiology*, vol. 234, no. 8, pp. 13182–13190, 2019.
- [38] C. Akgul, “Mcl-1 is a potential therapeutic target in multiple types of cancer,” *Cellular and Molecular Life Sciences*, vol. 66, no. 8, pp. 1326–1336, 2009.
- [39] A. Sugio, M. Iwasaki, S. Habata et al., “BAG3 upregulates Mcl-1 through downregulation of miR-29b to induce anticancer drug resistance in ovarian cancer,” *Gynecologic Oncology*, vol. 134, no. 3, pp. 615–623, 2014.
- [40] H. Yu, S. Yin, S. Zhou et al., “Magnolin promotes autophagy and cell cycle arrest via blocking LIF/Stat3/Mcl-1 axis in human colorectal cancers,” *Cell Death & Disease*, vol. 9, no. 6, p. 702, 2018.

Research Article

Identification of Drug-Disease Associations Using a Random Walk with Restart Method and Supervised Learning

Xiaoqing Liu,¹ Wenjing Yi,² Baohang Xi,² and Qi Dai ²

¹College of Sciences, Hangzhou Dianzi University, Hangzhou 310018, China

²College of Life Sciences, Zhejiang Sci-Tech University, Hangzhou 310018, China

Correspondence should be addressed to Qi Dai; daiailiu04@yahoo.com

Received 24 July 2022; Accepted 23 September 2022; Published 10 October 2022

Academic Editor: Lin Lu

Copyright © 2022 Xiaoqing Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Drug-disease correlations play an important role in revealing the mechanism of disease, finding new indications of available drugs, or drug repositioning. A variety of computational approaches were proposed to find drug-disease correlations and achieve good performances. However, these methods used a variety of network information, but integrated networks were rarely used. In addition, the role of known drug-disease association data has not been fully played. In this work, we designed a combination algorithm of random walk and supervised learning to find the drug-disease correlations. We used an integrated network to update the model and selected a gene set as the start of random walk based on the known drug-disease correlations data. The experimental results show that the proposed method can effectively find the correlation between drugs and diseases, and the prediction accuracy is 82.7%. We found that there are 8 pairs of drug-disease relationships that have not yet been reported, and 5 of them have pharmacodynamic effects on Parkinson's disease. We also found that a key linkage between Parkinson's disease and phenylhexol, a drug for the treatment of Parkinson's disease α -synuclein and tau protein, provides a useful exploration for the effectiveness of the treatment of Parkinson's disease.

1. Introduction

With the prevalence of complex diseases, the existing drugs are far from meeting the needs of human beings to fight against diseases. At the same time, due to the rising cost of drug research and development, long research and development cycle, large difference in research and development success rate, and high loss rate of new drugs, the research and development of innovative drugs has become a major challenge in the medical field.

At present, reusing compounds that have reduced risk to treat common or rare diseases has become a popular means of drug research and development. This strategy is called drug repositioning or drug reuse. This method not only reduces the overall development cost but also shortens a large amount of research and development time [1–3]. Through drug repositioning, pharmaceutical companies have achieved many successes, such as Pfizer's Viagra for erectile dysfunction [4] and Celgene's thalidomide for severe nodular leprosy erythema [5].

With the rapid expansion of large-scale genome, transcriptome and proteome data, computational drug repositioning study has emerged as one of the leading methods. Huang et al. developed a new drug repositioning pipeline to analyze four lung cancer microarray datasets, enrich biological processes, potential therapeutic drugs, and target genes for the treatment of non-small-cell lung cancer (NSCLC) [6]. They integrated two methods: machine learning algorithm and classification based on topological parameters. Zheng et al. designed a weighted ensemble similarity (WES) algorithm which provides a new perspective for drug repositioning and discovery [7]. Wang et al. integrated two drug transfer methods and proposed a new method for drug repositioning [8]. Cheng et al. [9] integrated the integration of chemical, gene, and disease networks, inferred the chemical hazard profile, studied the exposure data gap, and fully considered the gene and disease network in the chemical safety assessment [10]. A large number of genetic and molecular biology studies have shown that diseases reflect the interaction of multiple molecular components on a

certain level [11–14]. Therefore, drug repositioning study should consider the interaction between different disease-related genes [15–18]. Luo et al. found the potential indications of a given drug based on some comprehensive similarity measures and Bi-Random walk (BiRW) [19]. Yu et al. inferred the correlation between drugs and diseases by studying the characteristics of known protein complexes [20]. PREDICT (PREdicting Drug IndiCaTions) considers that similar drugs are suitable for similar diseases; the prediction task is achieved by designing similarity measures between multiple drugs and diseases [21].

The above method was successfully applied to drug-disease association study and achieved good performance. However, these methods have used a variety of network information, but the integrated network is still less used. With the increase of the related data of known drug diseases, a supervised learning method should be designed to further improve the drug-disease association research by using the related data of known drug diseases. In this paper, we used an integrated network consisting of HPRD, BioGRID, STRING, and other databases. Unlike previous network-based studies, which used the random walk method with restart on the network, we updated the model using the known data of the relationship between drugs and diseases and selected a gene set as the starting point of random walk, thus realizing the supervised learning of random walk with restart method. We also evaluated the performance of the proposed methods in various diseases and analyzed their GO and KEGG function enrichment.

2. Datasets and Methods

2.1. Protein–Protein Interaction (PPI) Network. Human protein–protein interaction (PPI) network is selected, which has been compiled by Jörg et al. that contained experimentally documented human physical interactions from TRANSFAC, IntAct39, MINT40, BioGRID41, HPRD42, KEGG43, BIGG44, CORUM45, PhosphoSitePlus46, and a large scale signaling network47. We used the largest connected component of the interaction in our analysis, consisting of 141,150 interactions between 13,329 proteins. Entrez Gene IDs were used to map disease-associated genes to the corresponding proteins in the interaction. The interaction and disease-gene association data is provided as a supplementary data set in Menche et al. [22]

2.2. Disease and Disease-Genes Data. Medical Subject Headings (MeSH) is an authoritative thesaurus compiled by the National Medical Library of the United States [23]. The disease subject words in the vocabulary provided by MeSH have perfect vocabulary classification. Our disease data and drug data are derived from Menche et al. [22], which integrate some genetic disease-related genes from the human Mendelian inheritance in man (OMIM: Online Mendelian Inheritance in Man) and trait gene association data from GWAS central. Through the medical topic title Ontology (MeSH) [24], the disease names of different disease nomenclature are combined into one name.

We screened diseases containing at least 20 disease-related genes from 1489 diseases in MeSH. In this paper, we considered at least 20 disease-related genes in order to understand the role of related genes in the interaction network, rather than the occurrence of disease due to the mutation of a gene. Finally, 299 diseases and their 3173 genes were obtained. In the process of disease screening, we required at least one drug for each disease. By searching the DrugBank database, the drug information that can treat 79 diseases corresponding to FDA approval is obtained, and Metab2Mesh is used for text mining [25]. If the text mining results indicate that there is a strong correlation between disease and drug, we added the relationship between the drug and disease into the known data set.

2.3. Drug and Drug-Target Data. DrugBank is a comprehensive drug information database, which not only includes the information of drug structure, drug target, and drug action mechanism but also integrates the information of drug experiment and clinical research. DrugBank has strong retrievable ability, coupled with its convenient web visualization function, which provides researchers with powerful convenience in drug research and development, drug mechanism exploration, and so on. DrugBank 5.0 contains information about 10971 drugs and 4900 protein targets, including 2391 FDA approved small molecule drugs, 934 approved biotechnology drugs, 109 nutritional drugs, and more than 5090 experimental drugs. We collected the drug and drug-target information certified by the FDA from DrugBank, and then searched for the strong literature evidence of drug-early-warning-disease association through Metab2Mesh, and finally obtained 238 drugs that can treat corresponding diseases.

2.4. Random Walk with Restart Method. PPI network can be expressed as $G = (V, E)$, where V denotes protein and E stands for protein–protein interaction. The $n * n$ adjacency matrix A is used to represent the PPI network, where n is the total number of the proteins. If there is interaction between protein i and protein j , $A_{[i,j]}$ is 1, otherwise it is 0. We then normalized the adjacency matrix A :

$$A'_{[i,j]} = \frac{A_{[i,j]}}{\sum_{k=1}^n A_{[k,j]}}. \quad (1)$$

Random walk is used to find potential gene association data of diseases or drugs. When the random walk converges, the probability of a disease or drug at each point of the PPI network can be obtained. The relationship between drugs and diseases can be calculated based on the correlation between the probability distribution of diseases and drugs.

Random walk starts with a set of seed genes. The initial vector of seed genes is defined as follows:

$$P_0 = [\psi_1, \psi_2, \dots, \psi_n]^T. \quad (2)$$

For a disease, we listed all the drugs that can treat it, incorporate all the genes of these drugs into the relevant genes of the disease, and took the combined gene set as the

seed gene of the disease. Among them, the genes directly related to the disease are defined as

$$P_{\text{dis}_{\text{dir}}} = \left[\psi_{\text{dis}_{\text{dir}_1}}, \psi_{\text{dis}_{\text{dir}_2}}, \dots, \psi_{\text{dis}_{\text{dir}_n}} \right]^T, \quad (3)$$

where the disease-related genes $\psi_{\text{dis}_{\text{dir}_i}}$ will be set to 1, otherwise it will be set to 0. Then $P_{\text{dis}_{\text{dir}}}$ is normalized as

$$P'_{\text{dis}_{\text{dir}_k}} = \frac{P_{\text{dis}_{\text{dir}_k}}}{\sum_{k=1}^n P_{\text{dis}_{\text{dir}_k}}}. \quad (4)$$

Suppose there are m drugs that can treat the same disease, they are represented as $P_{\text{dis}_{\text{drug}_1}}, P_{\text{dis}_{\text{drug}_2}}, \dots$, and $P_{\text{dis}_{\text{drug}_m}}$

$$P_{\text{dis}_{\text{drug}_m}} = \left[\psi_{\text{dis}_{\text{drug}_{m_1}}}, \psi_{\text{dis}_{\text{drug}_{m_2}}}, \dots, \psi_{\text{dis}_{\text{drug}_{m_n}}} \right]^T. \quad (5)$$

Sum all drugs for a disease:

$$P_{\text{dis}_{\text{drug}}} = \sum_{k=1}^m P_{\text{dis}_{\text{drug}_k}}. \quad (6)$$

Then we normalized $P_{\text{dis}_{\text{drug}}}$ as

$$P'_{\text{dis}_{\text{drug}_k}} = \frac{P_{\text{dis}_{\text{drug}_k}}}{\sum_{k=1}^m P_{\text{dis}_{\text{drug}_k}}}. \quad (7)$$

Finally we got its seed gene for a given disease,

$$P_{\text{disease}} = tP'_{\text{dis}_{\text{dir}}} + (1-t)P'_{\text{dis}_{\text{drug}}}, \quad (8)$$

where t is 0.5.

We also got the seed gene P_{drug} of a given drug following the same method. Start random walk and randomly access adjacent genes in each time scale ($t \rightarrow t+1$), the state probability P_{t+1} at time $t+1$ is

$$P_{t+1} = (1-r)A'P_t + rP_0, \quad (9)$$

where P_0 is the initial vector, P_t is the probabilities at time t , and r is the restart probability. If the difference between P_t and P_{t+1} is less than $1e-6$, it is considered that the process will reach a stable state. After reaching the stable state, the correlation between drugs and diseases, drugs and drugs, and diseases and diseases is calculated according to the probability of drugs and diseases accessing each node on the network.

2.5. Supervised Learning. Cross-validation is a frequently used model validation technology. It divides the known data into two subsets, adds the data of one subset to the model training, and verifies the model with the remaining subset to evaluate the performance of the model in unknown data. For example, when using k -fold cross validation, the known

data set needs to be randomly divided into k parts. In each cross-validation, $k-1$ data is selected to be added to the model training, and the remaining data is used for validation. Repeat for k times and select one piece of data for verification each time until each piece of data is tested.

The goal of cross-validation is to test the prediction ability of the model in new data, and it can also reflect the problem of overfitting or selection bias. In this paper, the idea of this method is used for supervised learning of random walk. For a certain disease, all drugs that can treat the disease in the data set are listed, and the genes associated with these drugs are incorporated into the relevant genes of the disease, and the combined gene set is used as the start of random walk. Needles are treated in the same way as drugs. In this paper, 403 known drug-disease associations between 78 diseases and 238 drugs were randomly divided into 10 parts. Nine of the disease and drug association data were selected to update the model, and the updated model was used to process the other data, so as to achieve the effect of supervised learning.

2.6. Evaluation Method. Receiver operating characteristic (ROC) curve is a curve based on the true positive rate (TPR) and false positive ratio (FPR) under various threshold settings. Area under the curve of ROC, also known as AUC value, can well reflect the performance of the classifier. The value of AUC varies between 0 and 1. When the AUC value is equal to 0.5, it means that the classifier cannot work. The larger the AUC value, the better the performance of the classifier. When the AUC value is 1, the classifier can produce perfect results.

3. Results

3.1. Performance Evaluation of the Random Walk with Restart Method Based on Supervised Learning. In order to evaluate the effectiveness of the proposed method, we first took the known drug-disease association as an independent validation data set. According to the relevant genes of 78 diseases and the drug targets of 238 drugs, the correlation information between diseases and drugs was obtained through restart random walk on PPI network. According to the ranking of drug-disease information pair by correlation, the AUC value was calculated. Three PPI networks BioGrid, HPRD, and STRING were independently verified, and their AUC results were 0.64, 0.52, and 0.66.

In order to further explore the efficiency of methods in different diseases, MeSH was used to classify all diseases. There are also some diseases in the classification that belong to a variety of disease types, such as colorectal tumors, which belong to C04 tumor diseases and C06 digestive system diseases. For the above case, we only calculated the average AUC value according to one of them. The AUC value was calculated on the basis of PPI network and optimal parameters. The classification results of various diseases are shown in Table 1.

From Table 1, it is easy to note that the performance of the random walk with the restart method is different among various diseases. It achieves good performance in the diseases of blood and lymphatic system C15, endocrine system diseases C19, eye diseases C11, and Male genitourinary

TABLE 1: Average AUC value of various diseases based on the random walk with restart method and three PPI networks BioGrid, HPRD, and STRING.

Disease classification based on MeSH	Number of diseases	Average AUC
Viral diseases C02	2	0.70467
Tumor C04	16	0.764632
Musculoskeletal diseases C05	5	0.748488
Digestive system diseases C06	9	0.760698
Respiratory diseases C08	2	0.68288
Nervous system diseases C10	9	0.62232
Eye diseases C11	2	0.83772
Male genitourinary system C12	1	0.81407
Cardiovascular disease C14	11	0.66685
Blood and lymphatic system C15	4	0.876637
Skin and connective tissue diseases C17	5	0.675556
Nutritional and metabolic diseases C18	6	0.734407
Endocrine system diseases C19	2	0.87446
Immune system diseases C20	4	0.62869

system C12, with AUC values above 0.8. The highest is blood and lymphatic system C15, with an AUC value of 0.877. The AUC value of nervous system diseases is low, only 0.62.

In order to further verify the efficiency of the random walk with the restart method and supervised learning, we randomly divided all known drug-disease relationships into ten parts, nine pieces of data are used as the training set and the other is used to calculate AUC. For a certain disease, we listed all the drugs that can treat the disease in the known training set, and then integrated all the related genes of these drugs into the related genes of the disease, and took the combined gene set as the start of random walk [26]. For drugs, the same method is used; that is, the relevant genes of diseases that can be treated by a drug in the training set were combined into the target information of the drug. Ten AUC values were obtained for each experiment. In order to reduce random interference, the above experiment was repeated 10 times, and a total of 100 AUC values were obtained, as shown in Figure 1.

The results show that the average value of 100 AUC values is 0.827, indicating that the proposed method found the relationship between drugs and diseases. With the help of the training data of the known network relationship between drugs and diseases, the prediction sensitivity of drugs and diseases was further improved. Adding the target information of drugs that can treat a disease to the disease-gene information will indirectly add some potential disease-gene information, making the disease-gene information more abundant. Similarly, adding the genes of all diseases that can be treated by a drug to the target information of drugs can also enrich the information of drug-action targets and make the relationship between drugs and dis-

eases more discovered, thus improving the prediction of drug sensitivity.

3.2. Analysis of the Relationship between Drugs and Diseases.

In this work, disease-related genes were taken as the starting point of random walk on one side, and the target genes of drugs were taken as the starting point of random walk on the other side. Through the restart random walk on the whole PPI network, the relationship between each disease and each drug on the PPI network was obtained, and their correlation coefficient was further calculated. We got 18564 group correlations of 78 diseases and 238 drugs. According to their correlation coefficients, 61 pairs of disease drugs with a correlation degree of more than 0.8 are found, of which 53 diseases and drugs have been confirmed by research, and 8 pairs belong to unknown drug-disease relationship. The relevant information of 8 pairs of diseases and drugs is shown in Table 2.

Methylprednisolone (DB00959) can treat autoimmune diseases, but we found that methylprednisolone is strongly associated with hematological diseases. According to the definition of MeSH, blood diseases include blood tumors, bone marrow diseases, and other diseases. Methylprednisolone is a biological macromolecular drug, a steroid derivative, and also a glucocorticoid. It can affect the expression of some genes through the cell membrane, thus interfering with the inflammatory response, inhibiting humoral immune response, and has a strong anti-inflammatory effect. Bowen et al. found that high-dose methylprednisolone has a certain effect on patients with recurrent chronic lymphoblastic leukemia [27]. Yao et al. found that methylprednisolone inhibited Wnt signaling pathway by downregulating the expression of LEF-1 protein, and Wnt signaling pathway is highly related to recurrent chronic lymphoblastic leukemia [28].

Mitoxantrone (DB01204) is associated with non-Hodgkin's lymphoma (NHL) and multiple sclerosis (MS). We found that it is also strongly correlated with lymphoid leukemia [29, 30]. Mitoxantrone has significant benefits for tumor control and overall survival in patients with recurrent acute lymphoblastic leukemia.

Prednisolone (DB00860) is a typical steroid drug, which can treat a variety of diseases, including rheumatoid arthritis, asthma, allergies, psoriasis, and multiple sclerosis [31]. However, these diseases are all autoimmune diseases. Therefore, we also found that prednisolone has a strong connection with autoimmune diseases.

We also found that apomorphine (DB00714), cabergoline (DB00248), bromocriptine (DB01200), and rotigotine (DB05271) are related to Parkinson's disease. After querying DrugBank, we knew that these four drugs have therapeutic effects on Parkinson's disease, but they are not included in the known data set.

3.3. GO Function Enrichment Analysis. Eight reusable drugs were found in this work, five of which have pharmacodynamic effects on Parkinson's diseases. We further performed GO function enrichment analysis on disease-related genes of the disease before drug action. The results are shown in Figure 2(a). It is easy to note that genes are mainly enriched

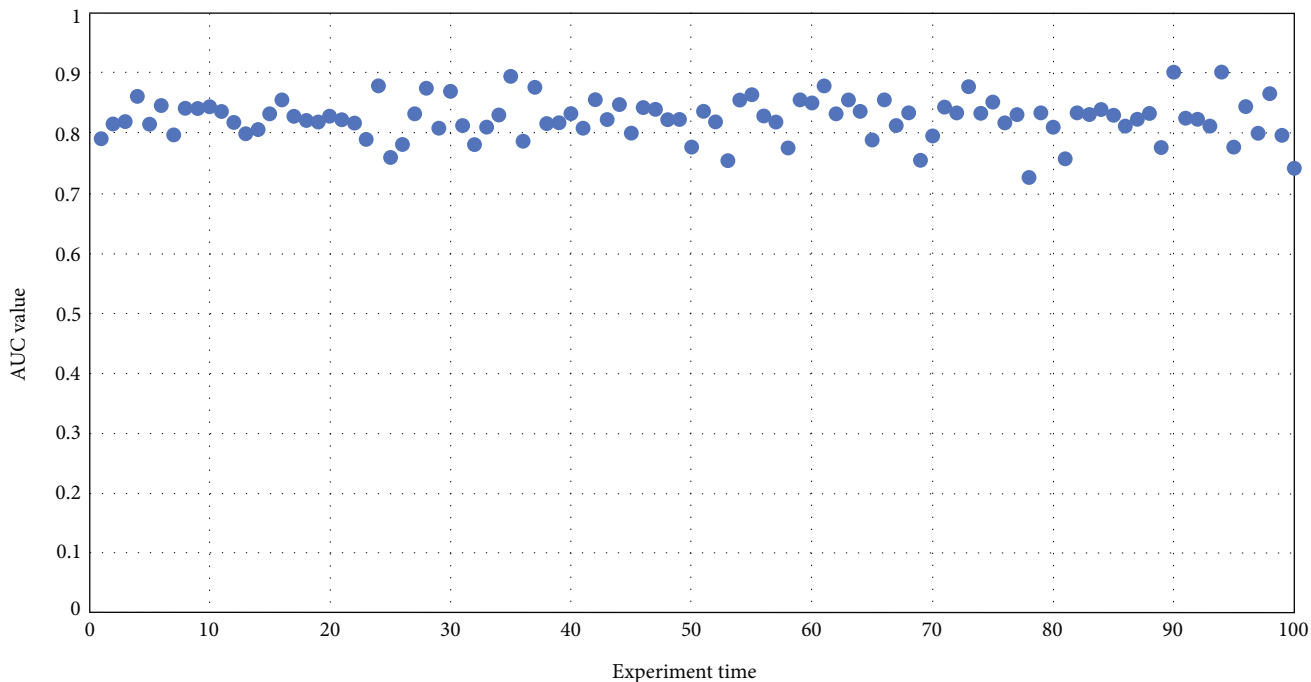


FIGURE 1: The AUC distribution of the random walk with restart method and supervised learning.

TABLE 2: The relevant information of eight diseases and eight drugs.

Disease	Drug	Pearson
Parkinsonian disorders	Apomorphine	0.876
Parkinsonian disorders	Cabergoline	0.876
Bone diseases metabolic	Calcitriol	0.841
Parkinsonian disorders	Bromocriptine	0.840
Leukemia lymphoid	Mitoxantrone	0.834
Hematologic diseases	Methylprednisolone	0.811
Parkinsonian disorders	Rotigotine	0.806
Autoimmune diseases	Prednisolone	0.806

in functional modules such as chromosome breakage (GO:0031052), upregulated cell migration (GO:0030335), and chain replacement (GO:0000732).

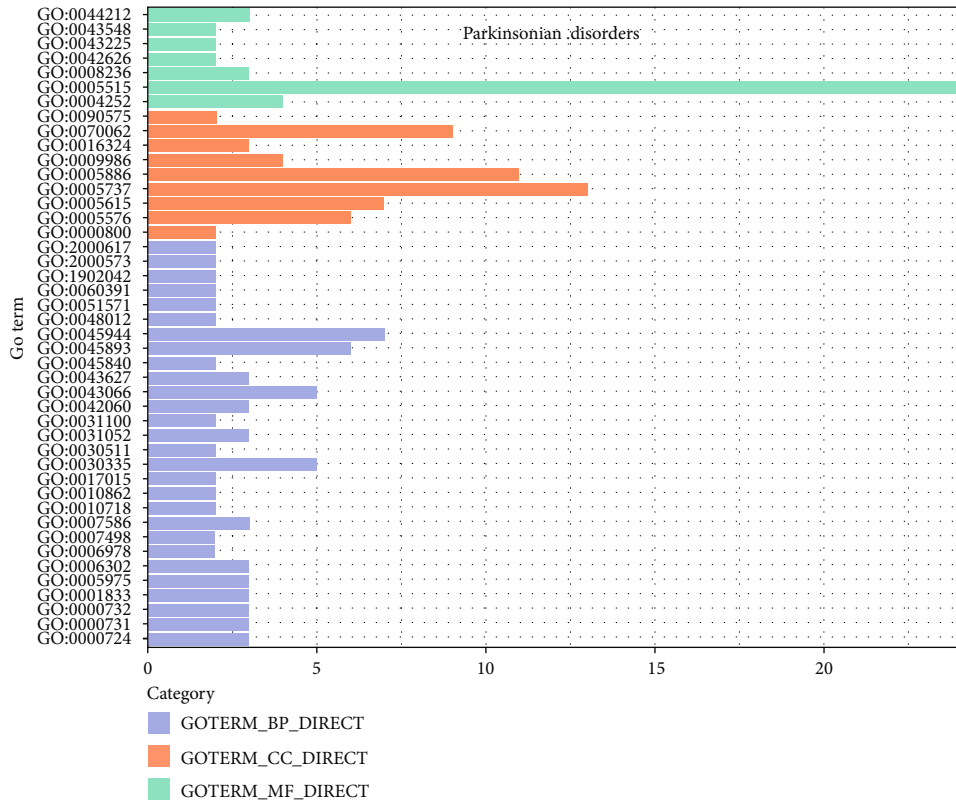
We then analyzed the related genes of Parkinson’s diseases after drug action. The results are shown in Figure 2(b). The results show that the gene is enriched in the following functional modules, such as the regulation of exercise (GO:0040012), dopamine binding (GO:0035240), and serotonin binding (GO:0051378).

Before and after random walk, the GO enrichment module of Parkinson’s disease has changed significantly. Before random walk, the main enrichment module of Parkinson’s syndrome is related to gene expression and cell movement in cells, which may be related to the pathogenesis of Parkinson’s disease. After random walk, the relevant genes of Parkinson’s syndrome are mainly enriched in some neural transmission modules, which are closely related to the treatment of Parkinson’s syndrome.

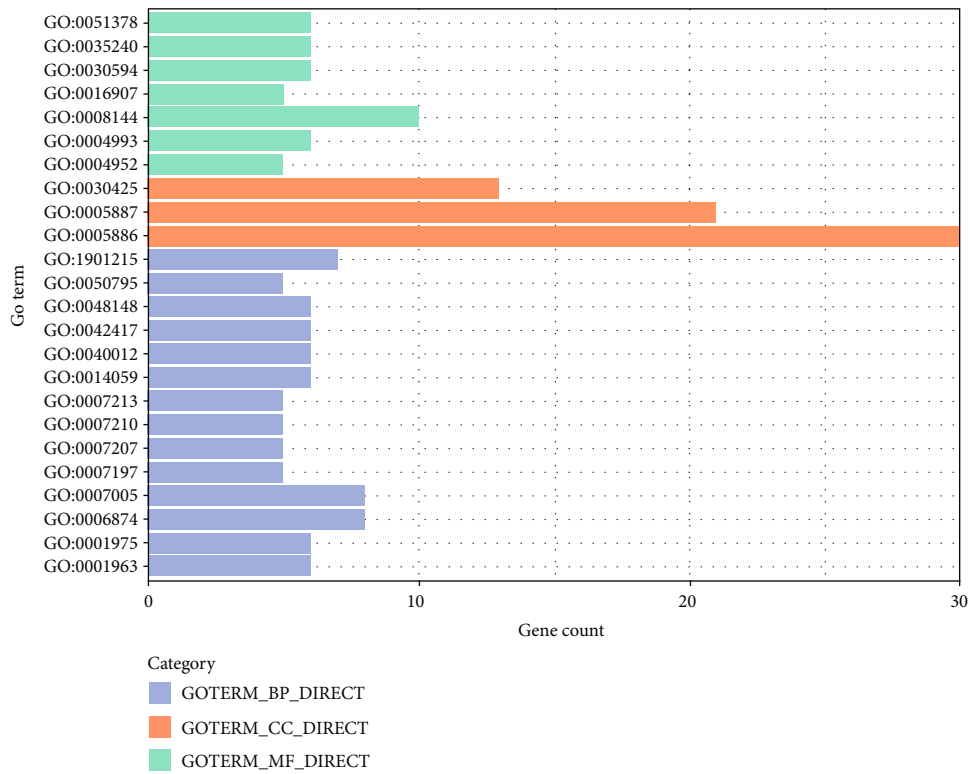
3.4. *KEGG Pathway Analysis.* We further analyzed the genes related to Parkinson’s disease by KEGG pathway. The results are shown in Figure 3. Figure 3(a) shows that genes are mainly enriched in pancreatic secretion (hsa04972), PI3K Akt signaling pathway (hsa04151), and other pathways. After adding drug information and random walk, we conducted KEGG pathway analysis on relevant genes. The results show that the genes are mainly enriched in neural active ligand receptor interaction (hsa04080), calcium signaling pathway (hsa04020), serotonin receptor synapse (hsa04726), and dopamine receptor synapse (hsa04728).

Before and after random walk, the KEGG pathway enrichment module of Parkinson’s disease has changed significantly. The approximate change is similar to the result of GO enrichment analysis. Before random walk, the main enrichment pathways of Parkinson’s syndrome are related to intracellular signaling pathways. After random walk, the relevant genes of Parkinson’s syndrome are mainly enriched in some neural transmission pathways, which are closely related to the treatment of Parkinson’s syndrome.

3.5. *Key Gene Analysis.* In order to further study key genes of Parkinson’s disease, we studied the local relationship between Parkinson’s disease and trihexyphenidyl, a drug that can treat Parkinson’s disease and their related genes on the network (Figure 4). It can be seen from Figure 4 that the key genes of Parkinson’s disease are α -synuclein (Gene ID: 6622) and tau protein (Gene ID:4137). α -Synuclein mainly exists at the synapse of the nerve cells and plays a key role in the transmission of neurotransmitters. Tau protein is a microtubule-associated protein that mainly exists in nerve cells. These two proteins are closely related to the pathogenesis of Parkinson’s disease.



(a)



(b)

FIGURE 2: The distribution of the related genes of Parkinson's diseases. (a) GO function enrichment on the disease-related genes of Parkinson's disease before drug action; (b) GO function enrichment on disease-related genes of Parkinson's disease after drug action.

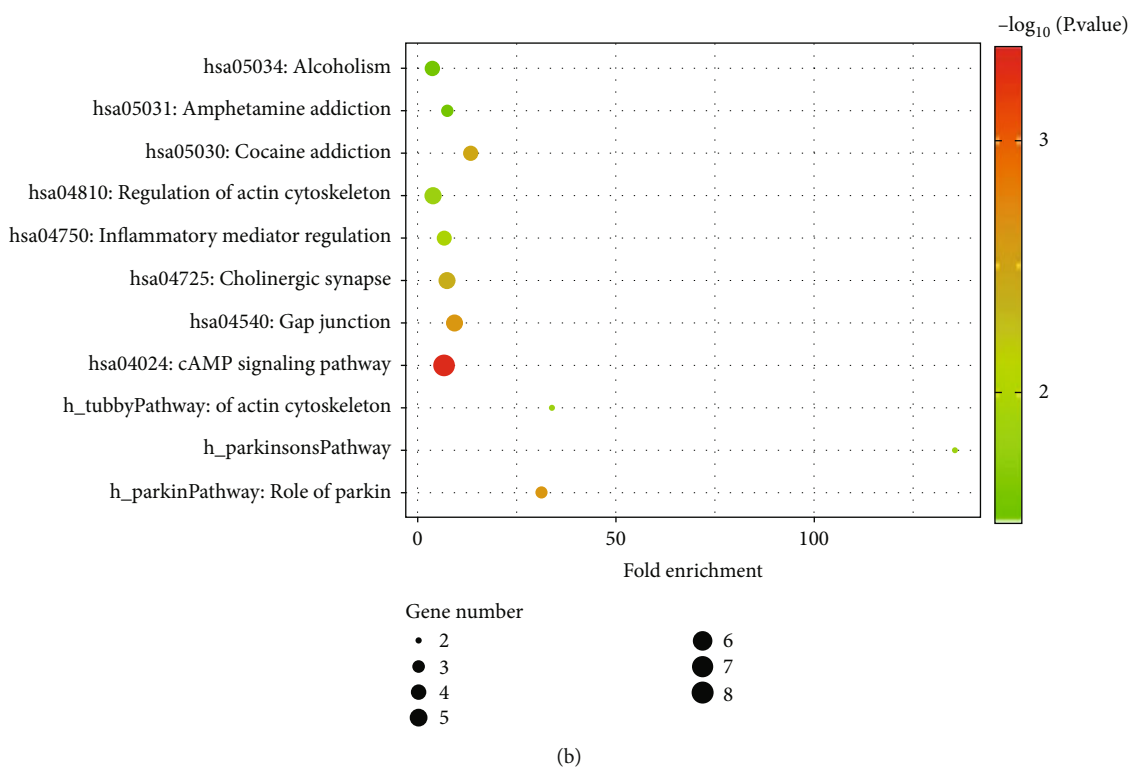
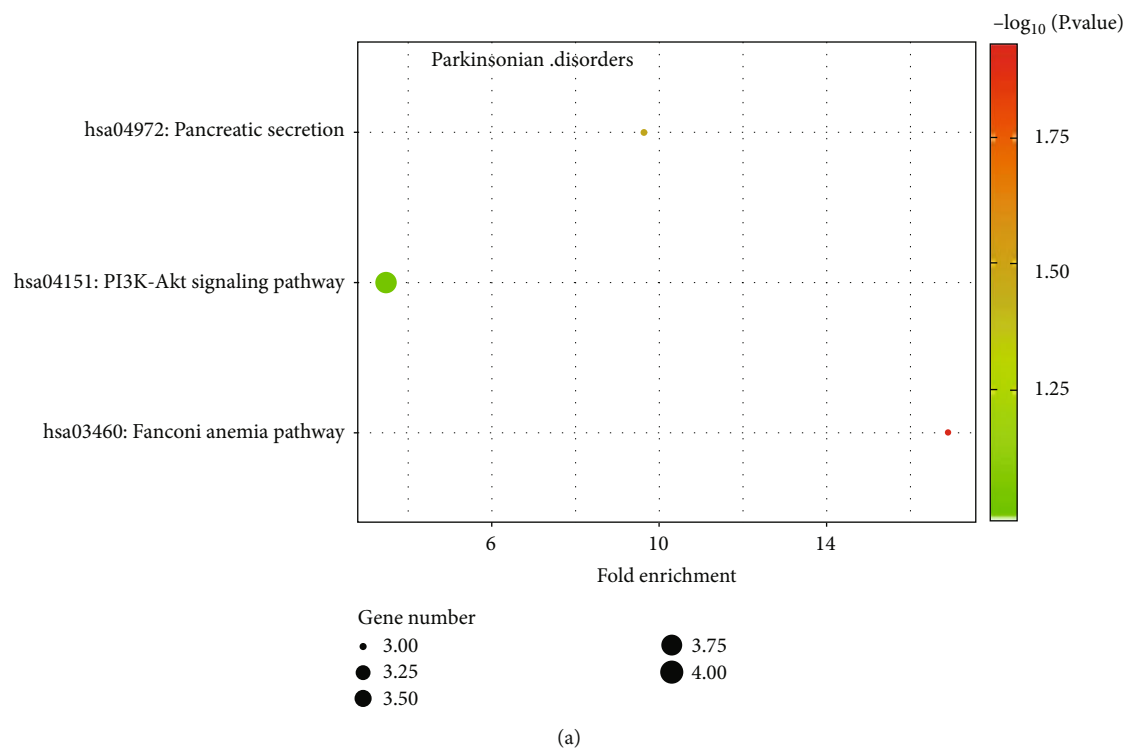


FIGURE 3: KEGG pathway of the related genes of Parkinson’s diseases. (a) KEGG pathway on the disease-related genes of Parkinson’s disease before drug action; (b) KEGG pathway on disease-related genes of Parkinson’s disease after drug action.

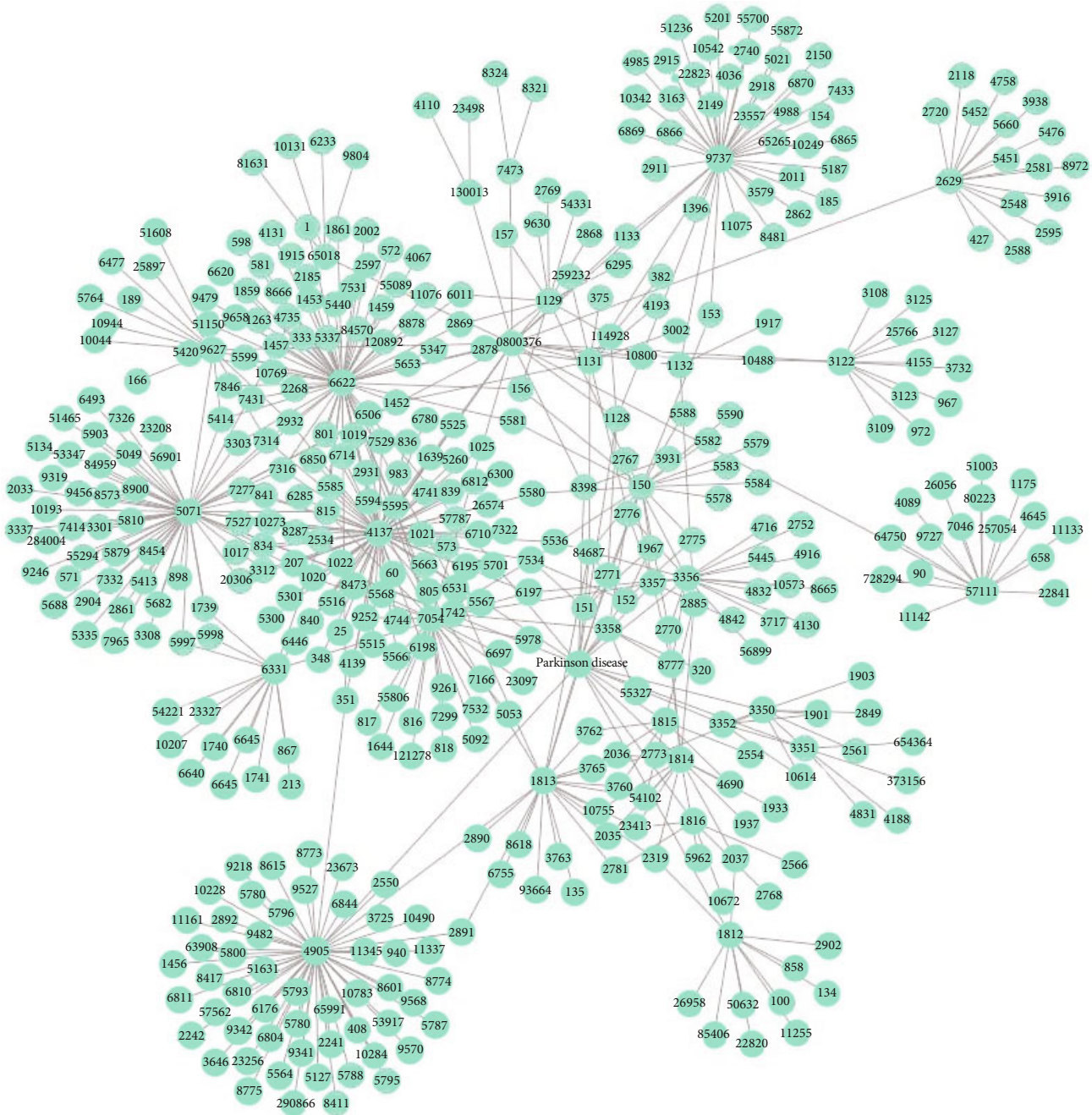


FIGURE 4: The gene network between Parkinson’s disease and trihexyphenidyl.

4. Conclusion

With the prevalence of complex diseases, the existing drugs are far from meeting the needs of human beings to fight diseases. At the same time, due to the rising cost of drug research and development and the long research and development cycle, the research and development means of innovative drugs have become a major challenge in the medical field. In recent years, with the continuous enrichment of disease and drug databases, researchers have realized drug reuse through the correlation analysis of disease-related genes, drugs, and drug-target data. This is a new research

and development idea in the field of pharmaceutical research and development, which reduces the research and development cost of innovative drugs and saves resources. Because most diseases are not single gene defects, they often involve the destruction of the coordination function between genes [32]. Therefore, we explored the relationship between drugs and diseases based on the biological function network. Using HPRD, BioGRID, STRING, and other databases; the protein–protein interaction (PPI) network was constructed. We designed a combination algorithm of random walk and supervised learning to predict the sensitivity of drugs. The accuracy of sensitivity prediction is 82.7%.

With the help of the proposed method, we found that 8 pairs of drug-disease relationships have not been reported, and 5 of them have pharmacodynamic effects on Parkinson's diseases. For Parkinson's disease, we found the changes of its functional modules by adding drug information and comparing before and after random walk, combined with the results of GO and KEGG function enrichment analysis. Using the network diagram of disease and drug-related genes after random walk, we found the key linkage between Parkinson's disease and phenylhexol, a drug for the treatment of Parkinson's disease α -synuclein and tau protein, which provide a useful exploration for the effectiveness of the treatment of Parkinson's disease.

Data Availability

The data used to support the findings of this study are available from the Protein-Protein Interaction (PPI) Network: <https://ppi-net.org>

Conflicts of Interest

The authors declare no conflict of interest, financial or otherwise.

Acknowledgments

The authors thank all the anonymous referees for their valuable suggestions and support. This work is supported by the National Natural Science Foundation of China (62172369); the Key Research and Development Plan of Zhejiang Province (2021C02039); and the Natural Science Foundation of Zhejiang Province (LY20F020016).

References

- [1] A. Joseph, "New drug development in the United States from 1963 to 1999," *Clinical Pharmacology & Therapeutics*, vol. 69, no. 5, pp. 286–296, 2001.
- [2] C. P. Adams and V. V. Brantner, "Estimating the cost of new drug development: is it really \$802 million?," *Health Affairs*, vol. 25, no. 2, pp. 420–428, 2006.
- [3] S. H. Sleigh and C. L. Barton, "Repurposing strategies for therapeutics," *Pharmaceutical Medicine*, vol. 24, no. 3, pp. 151–159, 2010.
- [4] N. Novac, "Challenges and opportunities of drug repositioning," *Trends in Pharmacological Sciences*, vol. 34, no. 5, pp. 267–272, 2013.
- [5] S. L. Walker, M. F. R. Waters, and D. N. J. Lockwood, "The role of thalidomide in the management of erythema nodosum leprosum," *Leprosy Review*, vol. 78, no. 3, pp. 197–215, 2007.
- [6] C. H. Huang, P. M. Chang, C. W. Hsu, C. Y. Huang, and K. L. Ng, "Drug repositioning for non-small cell lung cancer by using machine learning algorithms and topological graph theory," *BMC Bioinformatics*, vol. 32, no. 17, pp. 2664–2671, 2016.
- [7] C. Zheng, Z. Guo, C. Huang et al., "Large-scale direct targeting for drug repositioning and discovery," *Scientific Reports*, vol. 5, article 11970, 2015.
- [8] H. Wang, Q. Gu, J. Wei, Z. Cao, and Q. Liu, "Mining drug-disease relationships as a complement to medical genetics-based drug repositioning: where a recommendation system meets genome-wide association studies," *Clinical Pharmacology and Therapeutics*, vol. 97, no. 5, pp. 451–454, 2015.
- [9] F. Cheng, W. Li, Y. Zhou et al., "Prediction of human genes and diseases targeted by xenobiotics using predictive toxicogenomic-derived models (PTDMs)," *Molecular BioSystems*, vol. 9, no. 6, pp. 1316–1325, 2013.
- [10] J. P. Zhou, L. Chen, and Z. H. Guo, "iATC-NRAKEL: an efficient multi-label classifier for recognizing anatomical therapeutic chemical classes of drugs," *Bioinformatics*, vol. 36, no. 5, pp. 1391–1396, 2020.
- [11] E. E. Schadt, "Molecular networks as sensors and drivers of common human diseases," *Nature*, vol. 461, no. 7261, pp. 218–223, 2018.
- [12] A. Califano, A. J. Butte, S. Friend, T. Ideker, and E. Schadt, "Leveraging models of cell regulation and GWAS data in integrative network-based association studies," *Nature Genetics*, vol. 44, no. 8, pp. 841–847, 2012.
- [13] A. Zanzoni, M. Soler-López, and P. Aloy, "A network medicine approach to human disease," *22nd IUBMB Congress/37th FEBS Congress*, vol. 583, no. 11, pp. 1759–1765, 2012.
- [14] S. D. Ghiassian, "Network medicine: a network-based approach to human diseases," *Dissertations & Theses-Gradworks*, vol. 12, no. 1, pp. 56–68, 2015.
- [15] K. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A. L. Barabási, "The human disease network," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 21, pp. 8685–8690, 2007.
- [16] K. Lage, K. Møllgård, S. Greenway et al., "Dissecting spatio-temporal protein networks driving human heart development and related disorders," *Molecular Systems Biology*, vol. 6, no. 1, pp. 381–381, 2010.
- [17] H. Y. Chuang, E. Lee, Y. T. Liu, D. Lee, and T. Ideker, "Network-based classification of breast cancer metastasis," *Molecular Systems Biology*, vol. 3, p. 140, 2007.
- [18] T. Rolland, M. Taşan, B. Charlotiaux et al., "A proteome-scale map of the human interactome network," *Cell Cambridge Ma*, vol. 159, no. 5, pp. 1212–1226, 2014.
- [19] H. Luo, J. Wang, M. Li et al., "Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm," *Bioinformatics*, vol. 32, no. 17, pp. 2664–2671, 2016.
- [20] L. Yu, J. Huang, Z. Ma, J. Zhang, Y. Zou, and L. Gao, "Inferring drug-disease associations based on known protein complexes," *BMC Medical Genomics*, vol. 8, no. S2, p. 2, 2015.
- [21] A. Gottlieb, G. Y. Stein, E. Ruppim, and R. Sharan, "PREDICT: a method for inferring novel drug indications with application to personalized medicine," *Molecular Systems Biology*, vol. 7, no. 1, p. 496, 2011.
- [22] J. Menche, A. Sharma, M. Kitsak et al., "Disease networks. Uncovering disease-disease relationships through the incomplete interactome," *Science*, vol. 347, no. 6224, article 1257601, 2015.
- [23] S. Peng, R. You, H. Wang, C. Zhai, H. Mamitsuka, and S. Zhu, "DeepMeSH: deep semantic representation for improving large-scale MeSH indexing," *Bioinformatics*, vol. 32, no. 12, pp. i70–i79, 2016.
- [24] S. D. Jani, G. L. Argraves, J. L. Barth, and W. S. Argraves, "GeneMesh: a web-based microarray analysis tool for relating differentially expressed genes to MeSH terms," *BMC Bioinformatics*, vol. 11, no. 1, pp. 166–166, 2010.

- [25] M. A. Sartor, A. Ade, Z. Wright et al., “Metab2MeSH: annotating compounds with medical subject headings,” *Bioinformatics*, vol. 28, no. 10, pp. 1408–1410, 2012.
- [26] X. Pan, L. Chen, Liu, Z. Niu, T. Huang, and Y. D. Cai, “Identifying protein subcellular locations with embeddings-based node2loc,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 1, pp. 1–675, 2021.
- [27] D. A. Bowen, T. G. Call, G. D. Jenkins et al., “Methylprednisolone-rituximab is an effective salvage therapy for patients with relapsed chronic lymphocytic leukemia including those with unfavorable cytogenetic features,” *Leukemia & Lymphoma*, vol. 48, no. 12, pp. 2412–2417, 2007.
- [28] Q. M. Yao, P. P. Li, S. M. Liang et al., “Methylprednisolone suppresses the Wnt signaling pathway in chronic lymphocytic leukemia cell line MEC-1 regulated by LEF-1 expression,” *International Journal of Clinical and Experimental Pathology*, vol. 8, no. 7, pp. 7921–7928, 2015.
- [29] P. J. Rosen, C. Rankin, D. R. Head et al., “A phase II study of high dose ARA-C and mitoxantrone for treatment of relapsed or refractory adult acute lymphoblastic leukemia,” *Leukemia Research*, vol. 24, no. 3, pp. 183–187, 2000.
- [30] C. Parker, R. Waters, C. Leighton et al., “Effect of mitoxantrone on outcome of children with first relapse of acute lymphoblastic leukaemia (ALL R3): an open-label randomised trial,” *The Lancet*, vol. 376, no. 9757, pp. 2009–2017, 2010.
- [31] Y. Yang and L. Chen, “Identification of drug-disease associations by using multiple drug and disease networks,” *Current Bioinformatics*, vol. 17, no. 1, pp. 48–59, 2022.
- [32] X. Li, L. Lu, and L. Chen, “Lei Chen, identification of protein functions in mouse with a label space partition method,” *Mathematical Biosciences and Engineering*, vol. 19, no. 4, pp. 3820–3824, 2021.

Research Article

Identification of the Novel Gene Markers Based on the Gene Profile among Different Severity of Obstructive Sleep Apnea

Yi Ren , Yanyan Li , Xuemei Sui , Jinhe Yuan , Jun Lan , Xiayu Li , Yue Deng ,
Zhiping Xu , Xiu Cheng , Changjing Zhao , and Junyu Lu 

Department of Respiratory and Critical Care Medicine, The Fifth People's Hospital of Chongqing, Chongqing 400062, China

Correspondence should be addressed to Junyu Lu; junyulu218@163.com

Received 24 May 2022; Revised 24 August 2022; Accepted 4 September 2022; Published 4 October 2022

Academic Editor: Tao Huang

Copyright © 2022 Yi Ren et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Obstructive sleep apnea (OSA) is caused by repeated blockage of the upper respiratory airways during sleep. The traditional evaluation methods for OSA severity are yet limited. This study aimed to screen gene signatures to effectively evaluate OSA severity. Expression profiles of peripheral blood mononuclear cells in the different severities of OSA patients were accessed from Gene Expression Omnibus (GEO) database. A total of 446 differentially expressed genes (DEGs) were screened among the varying severities of OSA samples by analysis of variance (ANOVA) test. A total of 1,152 DEGs were screened between the pre- and post-treatment OSA samples by using *t* test. Overlap of the two groups of DEGs was selected (88 DEGs) for Metascape enrichment analysis. Afterwards, Mfuzz package was used to perform soft clustering analysis on these 88 genes, by which 6 clusters were obtained. It was observed that the gene expression condition of the cluster 3 was positively associated with OSA severity degree; also, the gene expression condition in cluster 4 was negatively correlated with OSA severity. A total of 10 gene markers related to OSA progression were selected from cluster 3 and cluster 4. Their expression levels and correlation were analyzed. The marker genes in cluster 3 and cluster 4 were examined, finding that most genes were significantly correlated with apnea hypopnea index (AHI). An accurate and objective assessment of the severity of OSA is of great significance for formulating follow-up treatment strategies for patients with OSA. In this paper, a set of marker genes that can detect the severity of OSA were screened by bioinformatics methods, which could be jointly used with the traditional OSA diagnostic index to achieve a more reliable OSA severity evaluation.

1. Introduction

Obstructive sleep apnea (OSA) is a common and easily overlooked sleep breathing disorder caused by the upper respiratory tract collapse during sleep. Its clinical symptoms are loud snoring, daytime hypersomnia, and fatigue [1]. Therapeutic strategy largely depends on the severity of OSA, which is evaluated by apnea hypopnea index (AHI). AHI is defined as the times of episodes of apnea and hypopnea per hour during sleep [2]. According to evaluation criteria for severity of OSA issued by American Academy of Sleep Medicine (AASM; <https://aasm.org/>), severity of OSA is mild when AHI ranges from 5 to 15, moderate as AHI ranges from 15 to 30, and severe as AHI is larger than 30 [3]. Relevant studies show that AHI has certain limitations in assessing the severity of disease, which may be affected

by factors such as gender and age [4], and the complications such as hypoxia caused by OSA cannot be fully reflected by the level of AHI [5]. Although AHI is still the gold standard for diagnosing the severity of OSA, it may not be comprehensive to evaluate the severity of OSA by using AHI alone. Therefore, it is urgent to search for powerful gene markers to assist in jointly examining the severity of OSA.

With the development of RNA sequencing, gene chip technology, and microarray tools, there is a growing body of literature that studies OSA-related genes, which provides several theoretic evidence for diagnosis, evaluation, and treatment of OSA [6]. For instance, Kun Li et al. [7] predicted OSA-related miRNAs and mRNA by a series of microarray analyses in 2017. Yuanyuan Cao et al. [8] screened and identified potential biomarkers of OSA and its related diseases by microarray analysis. Peng Lu et al.

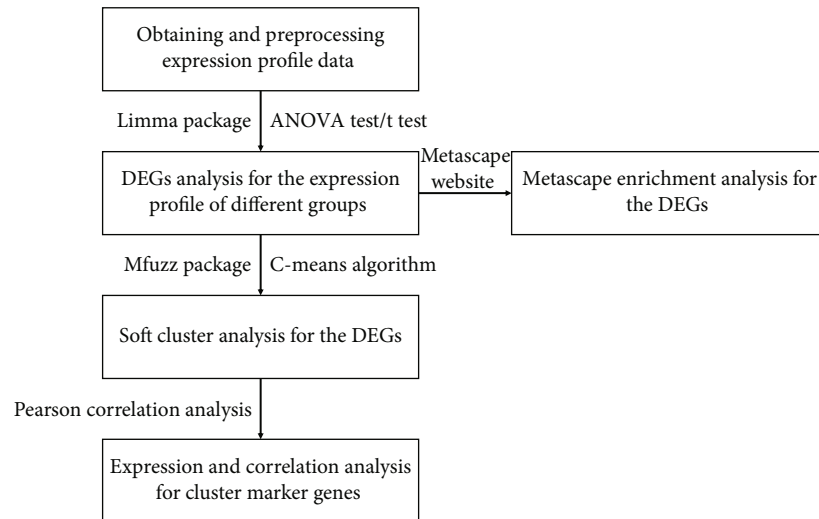


FIGURE 1: Flow chart of microarray analysis.

[9] established and validated the prognostic model of OSA based on microarray datasets and relevant bioinformatics analysis, providing a basis for further research on the molecular mechanism and potential targets of lipid metabolism regulation of OSA. However, few studies have involved gene markers related to different severity of OSA. Therefore, this study aimed to mine gene markers among the different severity OSA via microarray analysis.

This study downloaded the expression profile data of peripheral blood mononuclear cells in the patients with different severity OSA from public databases. Subsequently, differential expression analysis was performed to obtain DEGs. Functional enrichment analysis and soft clustering analysis were conducted to investigate the biological functions related to the DEGs and the expression states among different OSA severity, respectively. Finally, the correlation between marker genes expression and AHI was analyzed. Taken together, this study accessed expression profile data in different severity of OSA from GEO public database and screened relevant gene markers via a series of bioinformatics analysis, which could be jointly used with the traditional OSA diagnostic index to achieve a more reliable OSA severity evaluation and is of great significance for formulating appropriate treatment strategies for OSA patients.

2. Methods

2.1. Data Acquisition and Microarray Analysis. OSA-related expression matrix (GSE75097) was downloaded from Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) [10]. The severity of OSA was graded according to the boundary value of AHI. The dataset in this study included 6 primary snoring (PS) (AHI = 5 – 15), 16 moderate to severe OSA (MSO) (AHI = 15 – 30), 12 very severe OSA (VSO) (AHI > 30), and 14 very severe OSA with long-term continuous positive airway pressure treatment (VSOC) patients' whole genome expression profiles of peripheral blood mononuclear cells and related clinical data. Gene expression data were acquired by GPL 10904 detection

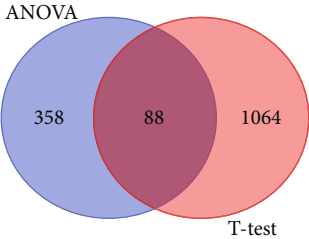
platform. Based on the above dataset, we performed the following microarray analyses (Figure 1).

2.2. Data Preprocessing and Differential Expression Analysis. Limma package [11] was used to read the expression matrix. Background correction and quantile normalization were undertaken [12]. Next, probes that could not match gene symbol were removed. In the case of multiple probes corresponding to one gene, the probe with the highest average expression was retained to filter genes.

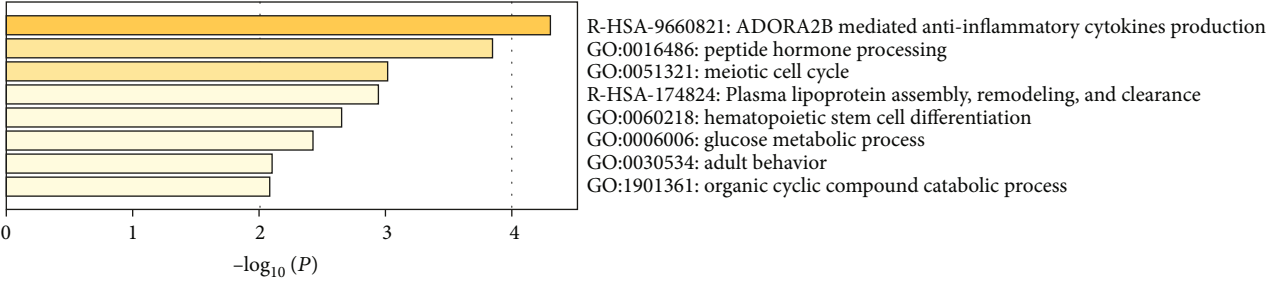
Significantly DEGs ($p < 0.05$) were determined from PS, MSO, and VSO groups using ANOVA test based on preprocessed expression profile data. Significantly DEGs ($p < 0.05$) were determined from VSO and VSOC groups by using t test. Lastly, the interaction of the above groups was taken as DEGs related to different severity of OSA.

2.3. Metascape Enrichment Analysis. To analyze the biological functions and signaling pathways involved in DEGs related to different OSA severity, we performed Metascape enrichment analysis. Metascape, a website tool (<https://metascape.org/gp/index.html#/main/step1>), integrates more than 40 microarray databases and can conduct multiple microarray analyses on gene lists. Metascape enrichment analysis can simplify redundant analysis results and analyze less correlated terms [13] and display the most important experimental results with concise bargraph. Moreover, enriched biological pathways can be expressed in a network manner, which is more conducive to understanding the relationship between pathways or biological processes.

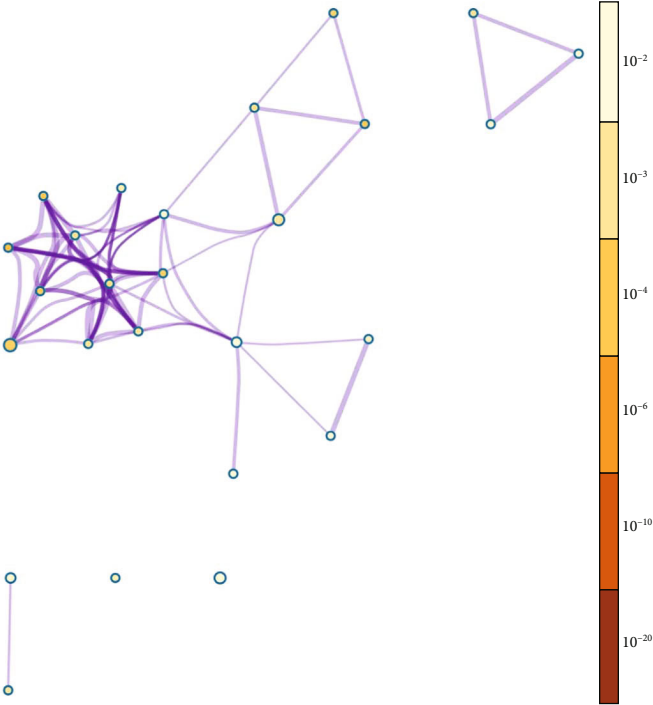
2.4. Soft Clustering Analysis. Mfuzz package [14] (<http://www.bioconductor.org/packages/release/bioc/html/Mfuzz.html>, version 2.6.1) was applied in this study for soft clustering analysis on DEGs related to different OSA severity. Acore threshold value was set as 0.6. Mfuzz package applied fuzzy c-means algorithm. Several time-course experiments found that it is difficult to distinguish DEGs in different time courses by a clear boundary [15–17]. Therefore, these genes



(a)



(b)



created by <http://metascape.org>

(c)

FIGURE 2: Continued.

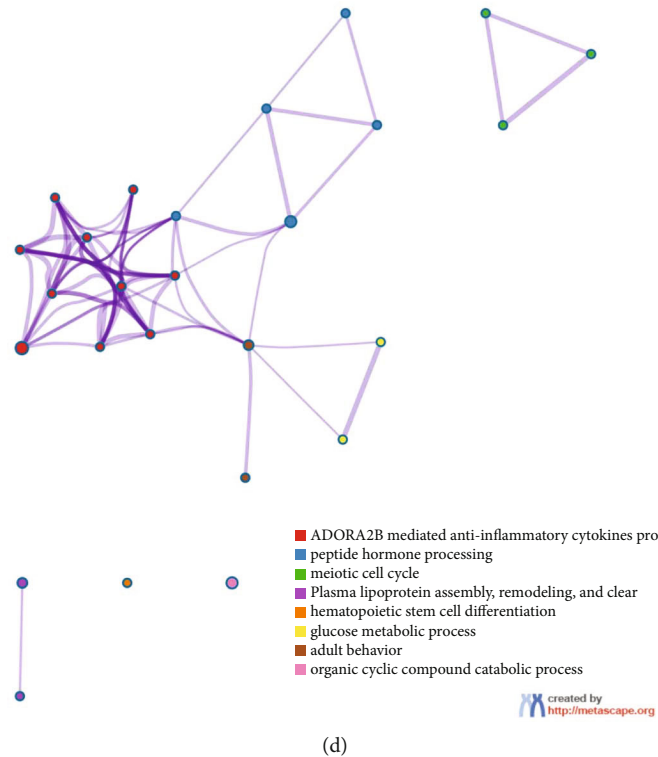


FIGURE 2: Differential expression analysis and functional enrichment analysis. (a). Venn diagram of DEGs of PS, MSO, and VSO patients (ANOVA test) and DEGs of VSO and VSOC (t test). (b) Bar chart of Metascape enrichment analysis (enriched terms are ranked by p value; deeper color of the bar chart denotes smaller p value; smaller p value denotes higher ranking). (c) Metascape enrichment network displayed by p value. Smaller p value presents deeper nodes color. (d) Metascape enrichment network displayed by terms (different colors denote different terms; nodes with the same color belong to the same term; large nodes include more genes; and the thicker the line between nodes, the higher the correlation).

are more suitable to be classified by soft clustering analysis. In hard clustering analysis such as k-means, each gene can only be divided into one cluster. Soft clustering algorithm, also known as fuzzy clustering, is different from hard clustering in that it can divide the samples in the datasets into one or more clusters with a certain probability and is more applicable for datasets that are hard to be distinguished and can decrease noise [18].

3. Results

3.1. Differential Expression Analysis and Functional Enrichment Analysis. The expression profiles of patients with varying OSA severities were analyzed by ANOVA test ($p < 0.05$), and a total of 446 DEGs were screened. (Supplementary Table 1). DEGs of VSO and VSOC patients were analyzed using t test. A total of 1,152 DEGs were obtained ($p < 0.05$) (Supplementary Table 2). By overlapping the two groups of DEGs, a total of 88 DEGs reflecting OSA severity were finally determined (Figure 2(a)) (Supplementary Table 3). To analyze the biological functions and signaling pathways these DEGs involved, Metascape enrichment analysis was conducted. It was shown that these genes were mainly enriched in ADORA2B mediated anti-inflammatory cytokines production, plasma lipoprotein assembly, remodeling, clearance, and glucose metabolic process (Figures 2(b)–2(d)).

3.2. Soft Clustering for the DEGs Related to OSA Severity. Mfuzz package was utilized to cluster the samples based on the 88 DEGs, by which obtained 6 clusters (Supplementary Table 4). The expression of genes in cluster 3 elevated as severity increased while dropped by treatment. The expression of genes in cluster 4 dropped as disease progressed, but the expression was elevated after treatment (Figures 3(a)–3(b)). The expression trend of the other clusters showed no relevant trend with disease severity. Therefore, we focused on genes in cluster 3 and cluster 4 in the next step.

3.3. Identification of Cluster Marker Genes. Cluster marker genes from both of the above clusters were selected (score > 0.6), by which *KCNN2*, *MIR942*, *PLXNB3*, *SERPINA12*, *TDRD3*, and *UPK1B* were screened from cluster 3, also *AASS*, *ASIP*, *LINC00494*, and *SAXO2* were screened from cluster 4. Subsequently, the expression of screened genes in different severity of patients (PS, MSO, VSO, and VSOC) was determined based on GEO-derived dataset, followed by Pearson correlation analysis with AHI. AHI is considered the most objective index for measurement of OSA severity. The expression of *KCNN2*, *MIR942*, *PLXNB3*, *SERPINA12*, *TDRD3*, and *UPK1B* significantly elevated as OSA progressed, while significantly dropped after treatment. Except for *MIR942*, the expression of above 5 genes was

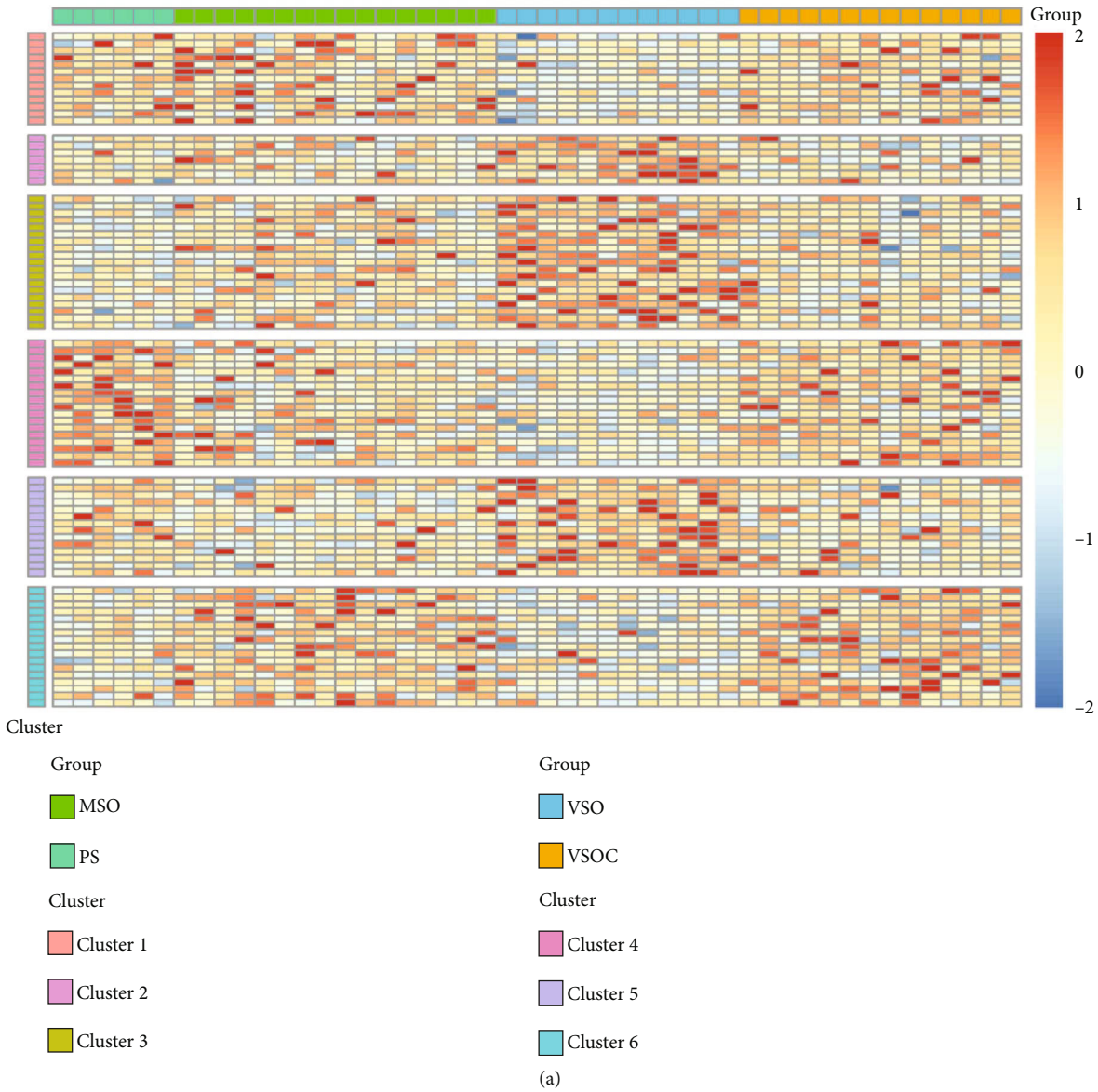


FIGURE 3: Continued.

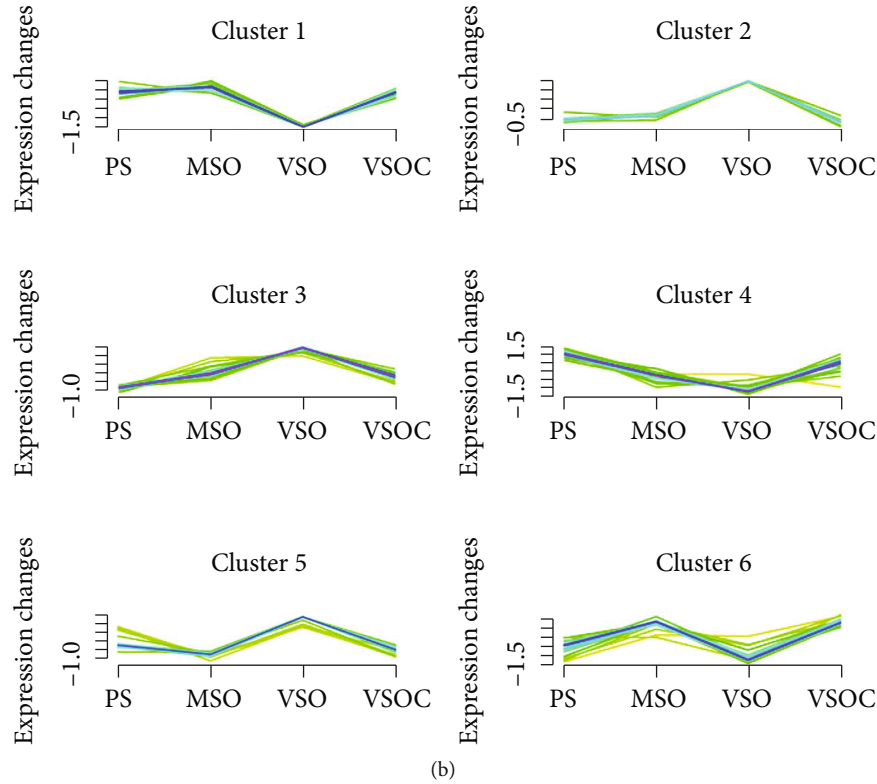


FIGURE 3: Soft clustering analysis. (a). Heatmap of gene expression in soft clustering analysis (x -axis: different severity OSA patients; y -axis: 6 clusters). (b) The expression changes of genes in 6 clusters of patients in different stages (green and yellow curves: gene expression mode with low score; red and purple: gene expression mode with high score; higher score presents closer gene expression mode is to that of CLUSTER).

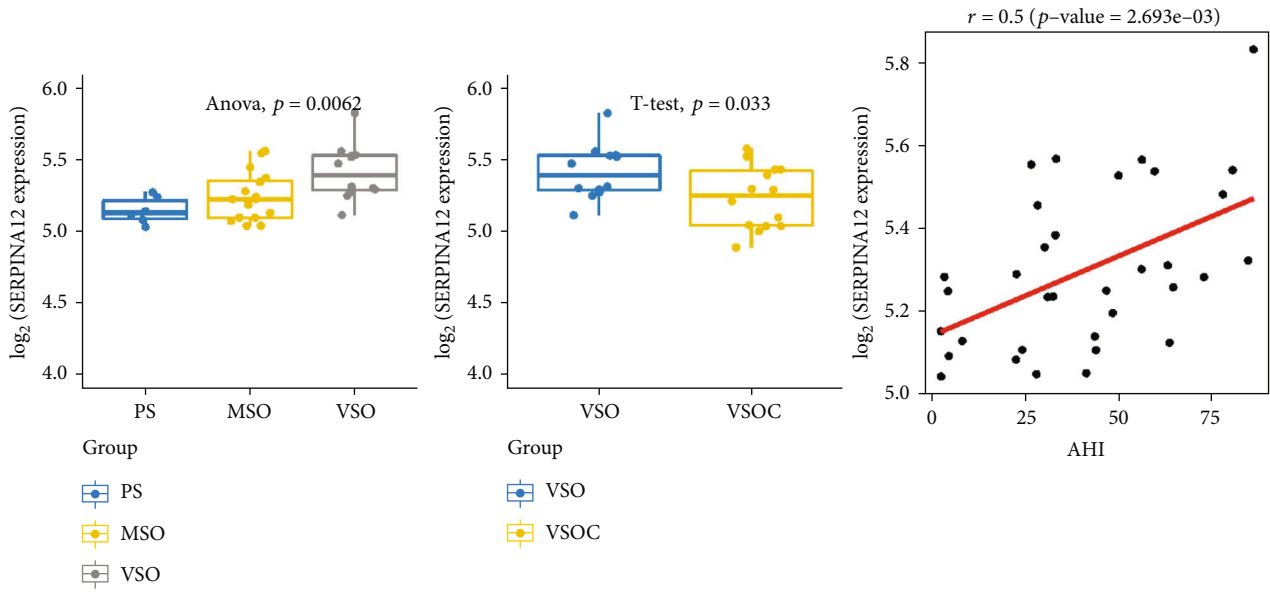
significantly positively correlated with AHI (Figures 4(a)–4(f)). The expression of *AASS*, *ASIP*, *LINC00494*, and *SAXO2* significantly declined as OSA progressed while significantly elevated after treatment. These gene expression levels were markedly negatively correlated with AHI (Figures 4(g)–4(j)). Taken together, cluster cores in cluster 3 could indicate an increasing OSA severity, while those in cluster 4 were potentially related to the reduced severity of OSA.

4. Discussion

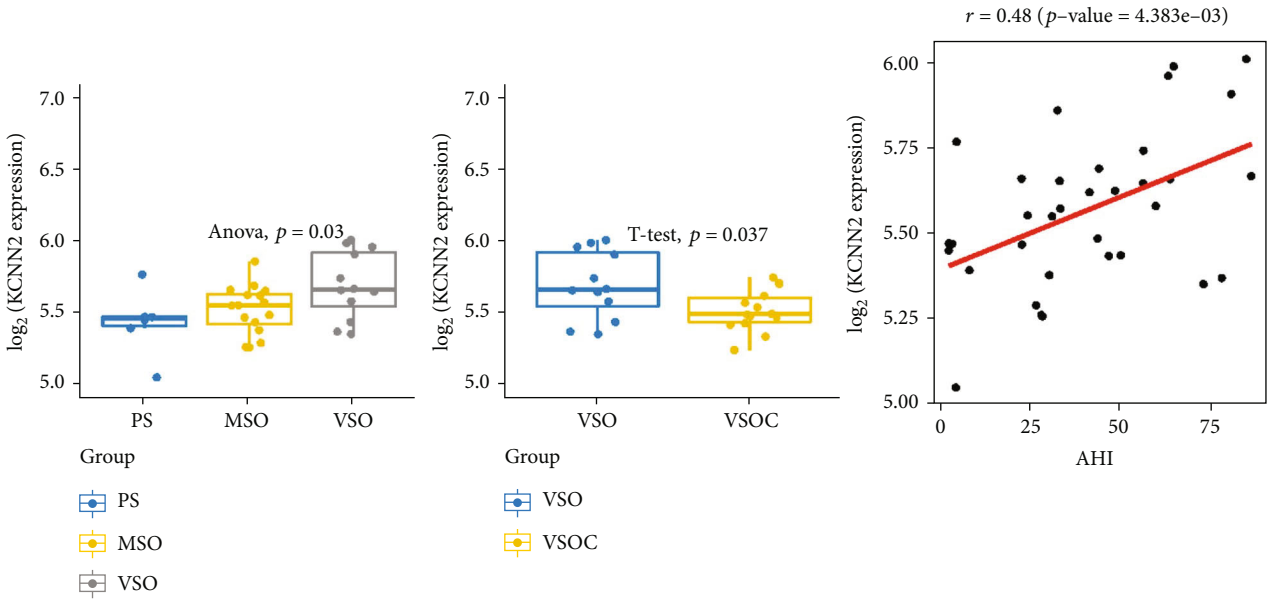
Some disease-related gene studies conducted a series of microarray analyses based on DEGs in different severity of specific diseases. For instance, in a study on bipolar disorder (BD), Ya-Chin Lee et al. [19] undertook differential expression analysis on RNA expression profile from blood samples of manic and remission patients. Lastly, encoding genes *TAS2R5* and *TASER3* and some lncRNAs and miRNAs (*MIR181B1* and *MIR103A1*) were found to be gene markers of BD manic episode via differential expression analysis [19]. Similarly, this study further analyzed DEGs related to different severity of OSA and finally determined gene markers to reflect OSA severity. Compared to the study mentioned above, this study adopted similar approaches to dig DEGs reflecting OSA progression despite different classification methods.

It is well known that OSA elevates risk of cardiovascular disease in adults [20]. A theory pointed out that OSA-derived systemic inflammatory response triggers cardiovascular disease [21, 22]. In this theory, OSA first causes intermittent hypoxia hypercapnia (IH) in patients, and then IH activates patient's systemic inflammatory response via HIF-1 or $\text{NF}\kappa\text{B}$ [23, 24]. All the above, inflammatory response-related biological functions or signaling pathways play a vital role in OSA course. In this study, Metascape enrichment analysis displayed that different severity of OSA-related DEGs was mainly enriched in ADORA2B mediated anti-inflammatory cytokines production. Therefore, these DEGs were believed to participate in biological functions related to inflammatory response.

In gene expression profile analysis, researchers are often concerned about expression profile characteristics changing with time sequences. For example, the expression mode of iron death-related genes was analyzed during hepatocellular carcinoma progression by Zuo Fei et al. [25]. A Chinese research team investigated survival-related genes in primary glioma via microarray analysis [26]. These studies all used Mfuzz package to perform soft clustering analysis on expression profile data. Soft clustering analysis is not only suitable for gene dynamic mode changing with time. This method can also be applied to analyze gene expression characteristics of diseases of varying severity. For example, Hongyun Qin et al. [27] used Mfuzz package to undertake soft clustering

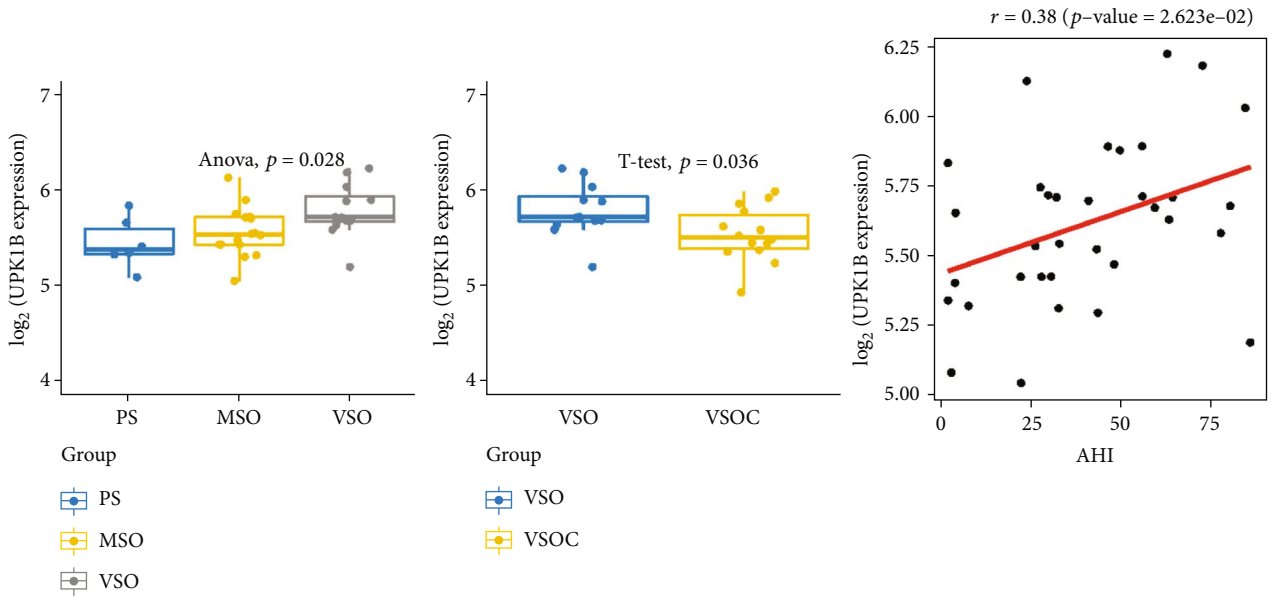


(a)

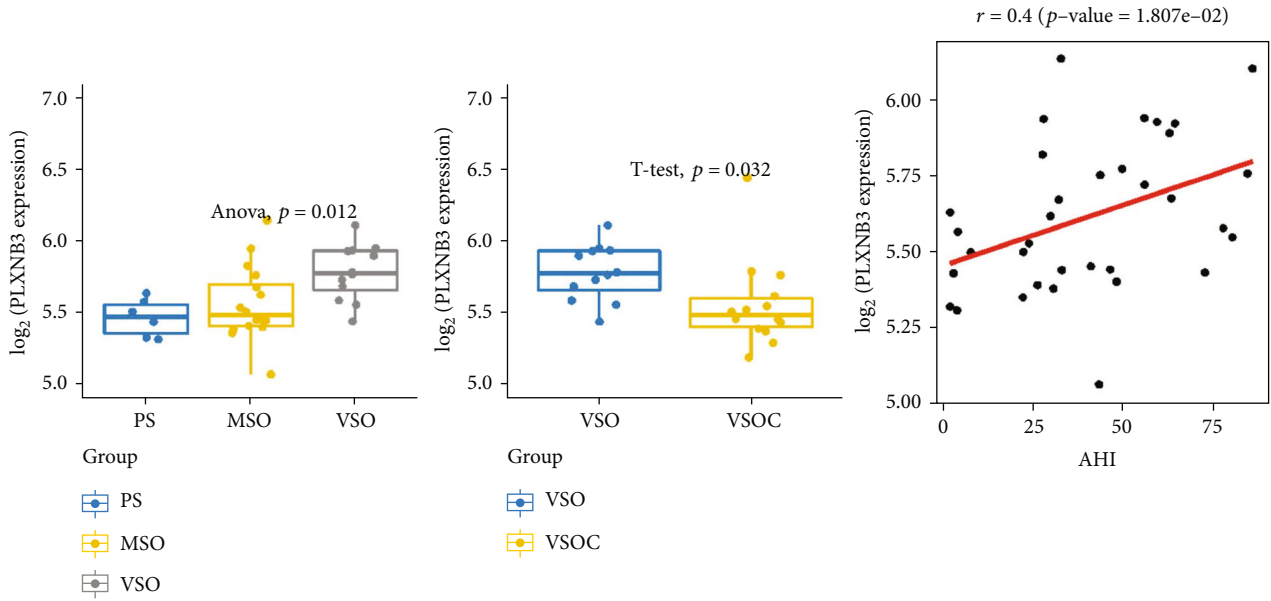


(b)

FIGURE 4: Continued.

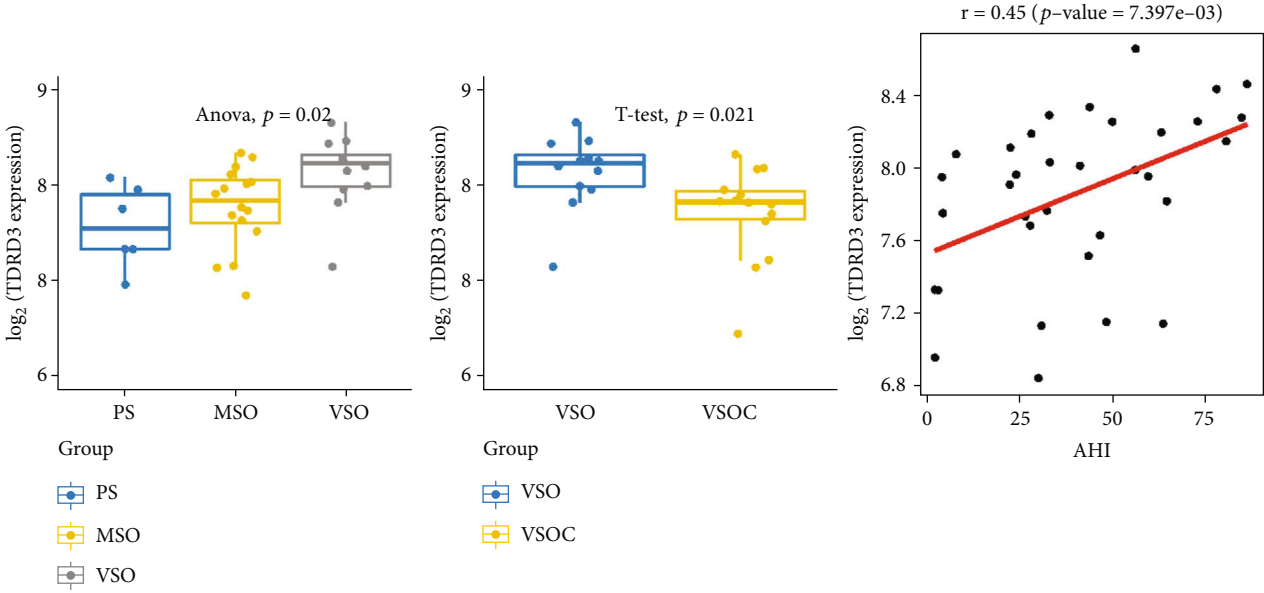


(c)

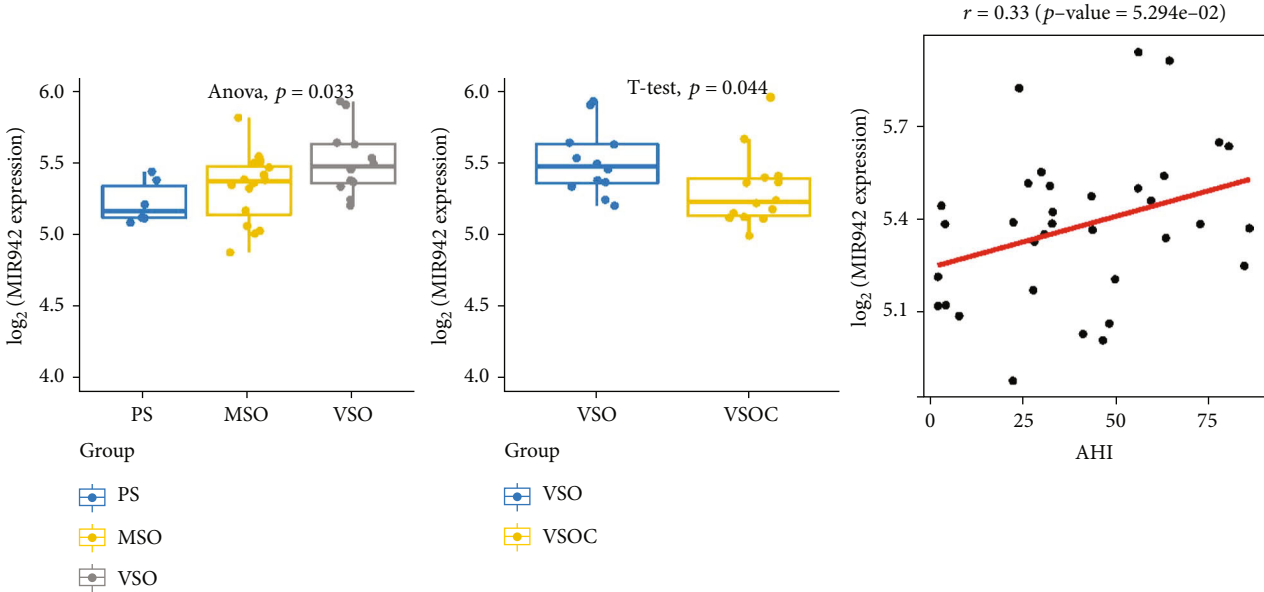


(d)

FIGURE 4: Continued.

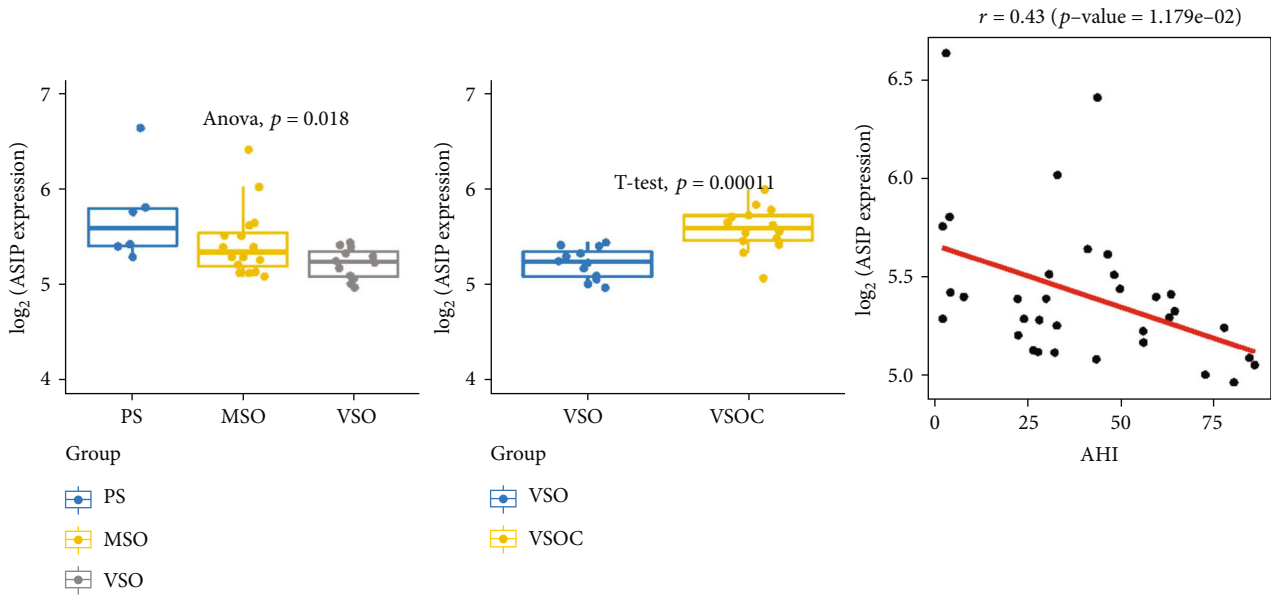


(e)

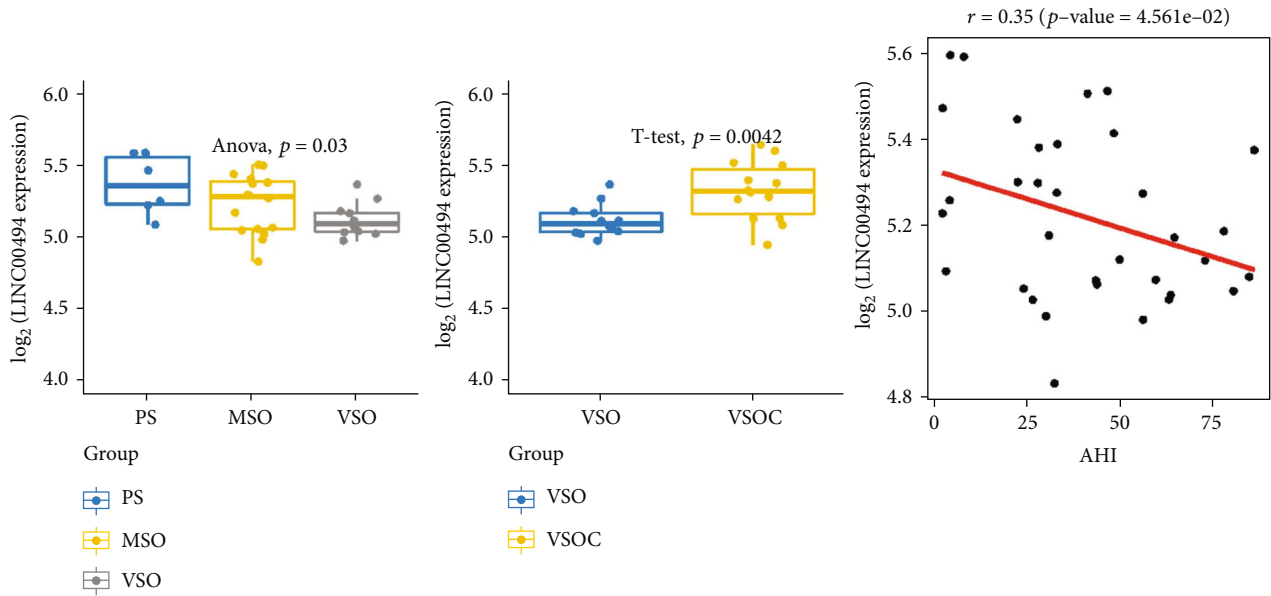


(f)

FIGURE 4: Continued.



(g)



(h)

FIGURE 4: Continued.

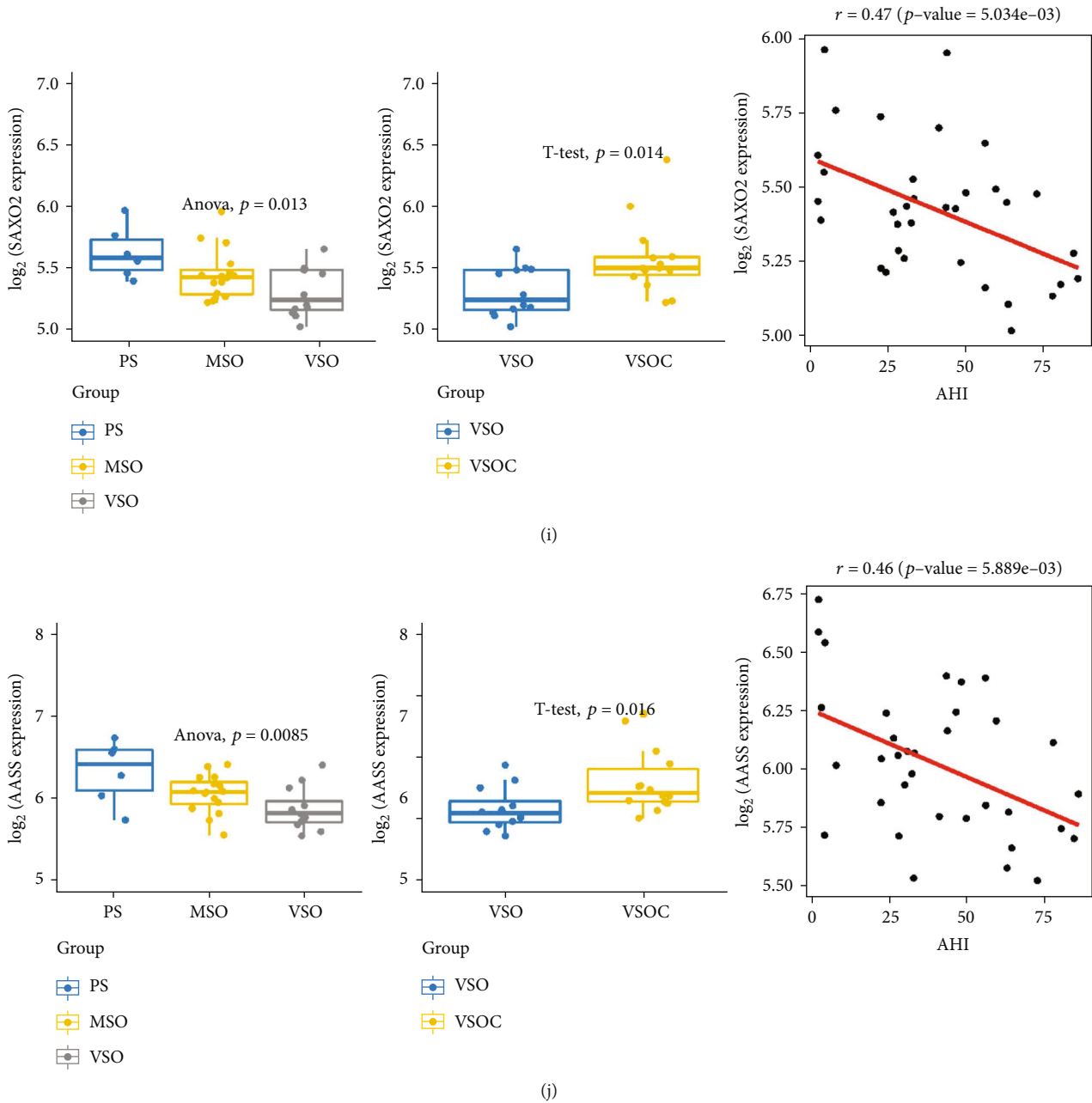


FIGURE 4: Expression of cluster core genes and correlation analysis. (a–j) The expression of all cluster marker genes in PS, MSO, VSO, and VSOC groups was analyzed. Pearson correlation analysis was undertaken on these genes and AHI.

analysis on DEGs of mild cognitive impairment and Alzheimer’s disease to explore corresponding biomarkers. However, there is a lack of studies on OSA-related expression genes at different severity. Although Xiandong Gu et al. [28] conducted functional enrichment analysis and protein-protein interaction (PPI) analysis based on DEGs of gene expression profile of visceral adipose tissue OSA patients and normal people, they also determined main functional subset in PPI network during OSA pathogenetic process. However, they only focused on genes related to OSA onset. Differently, our study conducted soft clustering analysis on DEGs based on the gene expression profile of

OSA at different severity, and finally screened out the marker genes reflecting OSA disease changes.

A recent study concluded that OSA-induced chronic intermittent hypoxemia leads to cognitive impairment through inflammatory responses [21]. Cluster marker gene *KCNN2*, which encodes potassium calcium to activate pathway protein, was screened out in this study and played an important role in neural network function [29]. A study published in Nature Neuroscience in 2020 showed that *KCNN2* expression in mouse nerve cells is associated with motor learning disabilities [30]. In addition, Fumiya Tatsuki et al. [31] reported in a study published in Neuron that

impaired *KCNN2* leads to shorter sleep duration. On the whole, it is posited that *KCNN2* is crucial in the regulation of neural networks in sleep and may also be altered during OSA-induced cognitive impairment and other neurological injuries.

Taken together, this study accessed expression profile data in different severity of OSA from GEO public database and screened relevant gene markers via differential expression analysis and soft clustering analysis. Next, we detected the expression of these genes in OSA patients at varying severity and those before and after treatment. Correlation between these gene expressions and AHI was tested. Altogether, *KCNN2*, *PLXNB3*, *SERPINA12*, *TDRD3*, *UPK1B*, *AASS*, *ASIP*, *LINC00494*, and *SAXO2* (significantly correlated with AHI) are expected to be gene markers for OSA severity. Despite comprehensive microarray analysis used in this paper, limitations still existed. There was a lack of mechanism studies; thus, we could not explain the specific role of these genes in the progression of OSA. Therefore, we plan to construct OSA mouse model and conduct molecular experiments and cellular experiments for further analysis of major action mechanism of these genes.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Ethical Approval

An ethics statement was not required for this study type; no human or animal subjects or materials were used.

Consent

Consent for publication was not required for this study type.

Conflicts of Interest

All authors confirm no conflicts of interest in this work.

Authors' Contributions

All authors read and approved the final manuscript. Yi Ren, Yanyan Li, and Xuemei Sui contributed equally to this study.

Supplementary Materials

Supplementary 1. Supplementary Table 1. DEGs of OSA samples with of varying severity.

Supplementary 2. Supplementary Table 2. DEGs of OSA samples before and after treatment.

Supplementary 3. Supplementary Table 3. Genes in the overlap of two groups of DEGs.

Supplementary 4. Supplementary Table 4. Six clusters obtained by soft clustering analysis.

References

- [1] S. Patel and R. Obstructive, "Obstructive sleep apnea," *Annals of Internal Medicine*, vol. 171, no. 11, pp. ITC81–ITC96, 2019.
- [2] V. K. Kapur, "Obstructive sleep apnea: diagnosis, epidemiology, and economics," *Respiratory Care*, vol. 55, no. 9, pp. 1155–1167, 2010.
- [3] L. J. Epstein, D. Kristo, Strollo PJ Jr et al., "Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults," *Journal of Clinical Sleep Medicine*, vol. 5, no. 3, pp. 263–276, 2009.
- [4] D. L. Bliwise, N. G. Bliwise, M. Partinen, A. M. Pursley, and W. C. Dement, "Sleep apnea and mortality in an aged cohort," *American Journal of Public Health*, vol. 78, no. 5, pp. 544–547, 1988.
- [5] T. Kumano-Go, A. Mikami, N. Suganuma et al., "Three components of obstructive sleep apnea/hypopnea syndrome," *Psychiatry and Clinical Neurosciences*, vol. 57, no. 2, pp. 197–203, 2003.
- [6] S. Mukherjee, R. Saxena, and L. J. Palmer, "The genetics of obstructive sleep apnoea," *Respirology*, vol. 23, no. 1, pp. 18–27, 2018.
- [7] K. Li, P. Wei, Y. Qin, and Y. Wei, "MicroRNA expression profiling and bioinformatics analysis of dysregulated microRNAs in obstructive sleep apnea patients," *Medicine (Baltimore)*, vol. 96, p. e7917, 2017.
- [8] Y. Cao, X. Cai, Q. Zhu, and N. Li, "Screening and identification of potential biomarkers for obstructive sleep apnea via microarray analysis," *Medicine (Baltimore)*, vol. 100, p. e24435, 2021.
- [9] L. Peng, X. Wang, and D. Bing, "Identification and validation of prognostic factors of lipid metabolism in obstructive sleep apnea," *Frontiers in Genetics*, vol. 12, p. 747576, 2021.
- [10] T. Barrett, S. E. Wilhite, P. Ledoux et al., "NCBI GEO: archive for functional genomics data sets—update," *Nucleic Acids Research*, vol. 41, no. Database issue, pp. D991–D995, 2013.
- [11] M. E. Ritchie, B. Phipson, D. I. Wu et al., "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, p. e47, 2015.
- [12] Y. Xie, X. Wang, and M. Story, "Statistical methods of background correction for Illumina BeadArray data," *Bioinformatics*, vol. 25, no. 6, pp. 751–757, 2009.
- [13] Y. Zhou, B. Zhou, L. Pache et al., "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets," *Nature Communications*, vol. 10, no. 1, p. 1523, 2019.
- [14] L. Kumar and M. E. Futschik, "Mfuzz: a software package for soft clustering of microarray data," *Bioinformatics*, vol. 2, no. 1, pp. 5–7, 2007.
- [15] P. T. Spellman, G. Sherlock, M. Q. Zhang et al., "Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization," *Molecular Biology of the Cell*, vol. 9, no. 12, pp. 3273–3297, 1998.
- [16] R. J. Cho, M. J. Campbell, E. A. Winzler et al., "A genome-wide transcriptional analysis of the mitotic cell cycle," *Molecular Cell*, vol. 2, no. 1, pp. 65–73, 1998.
- [17] S. Chu, J. DeRisi, M. Eisen et al., "The transcriptional program of sporulation in budding yeast," *Science*, vol. 282, no. 5389, pp. 699–705, 1998.

- [18] M. E. Futschik and B. Carlisle, "Noise-robust soft clustering of gene expression time-course data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 4, pp. 965–988, 2005.
- [19] Y. C. Lee, Y. L. Chao, C. E. Chang et al., "Transcriptome changes in relation to manic episode," *Frontiers in Psychiatry*, vol. 10, p. 280, 2019.
- [20] D. F. Smith and R. S. Amin, "OSA and cardiovascular risk in pediatrics," *Chest*, vol. 156, no. 2, pp. 402–413, 2019.
- [21] X. Liu, Y. Ma, R. Ouyang et al., "The relationship between inflammation and neurocognitive dysfunction in obstructive sleep apnea syndrome," *Journal of Neuroinflammation*, vol. 17, no. 1, p. 229, 2020.
- [22] D. Unnikrishnan, J. Jun, and V. Polotsky, "Inflammation in sleep apnea: an update," *Reviews in Endocrine & Metabolic Disorders*, vol. 16, no. 1, pp. 25–34, 2015.
- [23] S. Ryan, C. T. Taylor, and W. T. McNicholas, "Predictors of elevated nuclear factor-kappaB-dependent genes in obstructive sleep apnea syndrome," *American Journal of Respiratory and Critical Care Medicine*, vol. 174, no. 7, pp. 824–830, 2006.
- [24] S. Ryan, C. T. Taylor, and W. T. McNicholas, "Selective activation of inflammatory pathways by intermittent hypoxia in obstructive sleep apnea syndrome," *Circulation*, vol. 112, pp. 2660–2667, 2005.
- [25] Z. Fei, Y. Lijuan, Z. Jing, Y. Xi, P. Yuefen, and H. Shuwen, "Molecular characteristics associated with ferroptosis in hepatocellular carcinoma progression," *Human Cell*, vol. 34, no. 1, pp. 177–186, 2021.
- [26] Y. Yang, R. Yan, L. Zhang, X. Meng, and W. Sun, "Primary glioblastoma transcriptome data analysis for screening survival-related genes," *Journal of Cellular Biochemistry*, vol. 121, no. 2, pp. 1901–1910, 2020.
- [27] H. Qin, C. Hu, X. Zhao, M. Tian, and B. Zhu, "Usefulness of candidate mRNAs and miRNAs as biomarkers for mild cognitive impairment and Alzheimer's disease," *The International Journal of Neuroscience*, pp. 1–14, 2021.
- [28] X. Gu, W. Yang, X. Luo, X. Wang, J. Tang, and Z. Cai, "Bioinformatics analysis to reveal the key genes related to obstructive sleep apnea," *Sleep & Breathing*, vol. 23, no. 1, pp. 259–267, 2019.
- [29] J. F. Liegeois, F. Mercier, A. Graulich, F. Graulich-Lorge, J. Scuvee-Moreau, and V. Seutin, "Modulation of small conductance calcium-activated potassium (SK) channels: a new challenge in medicinal chemistry," *Current Medicinal Chemistry*, vol. 10, no. 8, pp. 625–647, 2003.
- [30] S. Mohammad, S. J. Page, L. Wang et al., "Kcnn2 blockade reverses learning deficits in a mouse model of fetal alcohol spectrum disorders," *Nature Neuroscience*, vol. 23, no. 4, pp. 533–543, 2020.
- [31] F. Tatsuki, G. A. Sunagawa, S. Shi et al., "Involvement of Ca^{2+} -dependent hyperpolarization in sleep duration in mammals," *Neuron*, vol. 90, no. 1, pp. 70–85, 2016.

Research Article

Predicting Conserved Water Molecules in Binding Sites of Proteins Using Machine Learning Methods and Combining Features

Wei Xiao , Juhui Ren , Jutao Hao , Haoyu Wang , Yuhao Li , and Liangzhao Lin 

School of Electronic and Information, Shanghai Dianji University, Shanghai 201306, China

Correspondence should be addressed to Wei Xiao; xiaow@sdju.edu.cn and Liangzhao Lin; linlz@sdju.edu.cn

Received 9 August 2022; Accepted 15 September 2022; Published 3 October 2022

Academic Editor: Lei Chen

Copyright © 2022 Wei Xiao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Water molecules play an important role in many biological processes in terms of stabilizing protein structures, assisting protein folding, and improving binding affinity. It is well known that, due to the impacts of various environmental factors, it is difficult to identify the conserved water molecules (CWMs) from free water molecules (FWMs) directly as CWMs are normally deeply embedded in proteins and form strong hydrogen bonds with surrounding polar groups. To circumvent this difficulty, in this work, the abundance of spatial structure information and physicochemical properties of water molecules in proteins inspires us to adopt machine learning methods for identifying the CWMs. Therefore, in this study, a machine learning framework to identify the CWMs in the binding sites of the proteins was presented. First, by analyzing water molecules' physicochemical properties and spatial structure information, six features (i.e., atom density, hydrophilicity, hydrophobicity, solvent-accessible surface area, temperature B-factors, and mobility) were extracted. Those features were further analyzed and combined to reach a higher CWM identification rate. As a result, an optimal feature combination was determined. Based on this optimal combination, seven different machine learning models (including support vector machine (SVM), *K*-nearest neighbor (KNN), decision tree (DT), logistic regression (LR), discriminant analysis (DA), naïve Bayes (NB), and ensemble learning (EL)) were evaluated for their abilities in identifying two categories of water molecules, i.e., CWMs and FWMs. It showed that the EL model was the desired prediction model due to its comprehensive advantages. Furthermore, the presented methodology was validated through a case study of crystal 3skh and extensively compared with Dowser++. The prediction performance showed that the optimal feature combination and the desired EL model in our method could achieve satisfactory prediction accuracy in identifying CWMs from FWMs in the proteins' binding sites.

1. Introduction

The research on water molecules in the proteins' binding sites has attracted increasing attention during the past decade [1–6]. The water molecules usually interact with the surrounding atoms by forming the bridging hydrogen bonds, which are important in stabilizing protein structures, and assisting protein folding [2, 3]. Besides, it has been shown that water molecules can improve the binding affinity by increasing the binding energy [5]. In a typical crystal structure, water molecules are normally randomly distributed in the structure. To study the solvent effects of the

water molecules, one often adopts the implicit solvent models, which mainly include the Poisson-Boltzmann solvent accessible surface model [7, 8] and the generalized Born solvent accessible surface model [9]. This category can accurately predict and evaluate the binding energy between ligands and targets by calculating the corresponding solvent entropy. However, they cannot reflect the mediating interactions of water molecules between the ligands and the targets, thus affecting the prediction accuracy of the binding modes [10, 11]. The other category is the explicit water models which involve the free energy calculation methods (such as *free energy perturbations* [12] and *thermodynamic*

integration [13]) to evaluate the solvent entropy. Although those models can accurately calculate the solvent entropy, they cannot be applied to large-scale drug design due to their computationally demanding nature [3].

Generally speaking, the water molecules in the binding sites of the crystal structures can be divided into two groups, i.e., free water molecules (FWMs) and conserved water molecules (CWMs). The FWMs mean the water molecules that are easily displaced by ligands (often coined as the *displaced water molecules*) and those that are not displaced by ligands but are highly variable in crystal structures [14]. The FWMs not only occupy a certain space in the binding sites but also play an important role in molecular recognition and drug screening. Differently, the CWMs are not displaced by ligands; however, they exist in the overwhelming majority of the crystal structures [14]. In some studies, e.g., [15], the CWMs are determined if the distance between waters in the ligand-free and bound structures is less than 1.2 Å. Moreover, the CWMs that can be deeply buried in proteins and form strong hydrogen bonds with the polar groups of the surrounding proteins are regarded as the *structural water molecules* [16, 17], which have important effects on the structure and function of biomacromolecules (e.g., the catalytic activity of enzymes, the folding and unfolding of proteins, and the conformation of biomacromolecules) [16, 17]. Furthermore, if the CWM is located within 1 Å of another water molecule lying in at least one other homologous protein, then this CWM often refers to as the *consensus water molecule* [18]. Effective identification of the conserved (consensus) water molecules can facilitate ligand designs. For example, if the conserved (consensus) water molecules are known a priori in a protein's binding site, then the ligand design can be improved by including polar atoms at appropriate locations in the ligand to form the hydrogen bonds with the water molecules or to displace them from the binding site [19]. Also, the conserved (consensus) water molecules generally have more neighboring protein atoms, which lead to a more hydrophilic environment, and more hydrogen bonds to the proteins, making the protein atoms less mobile [20]. Additionally, the conserved (consensus) water molecules also play a key role in maintaining and stabilizing the alanine racemase dimer [21] and reducing the flexibility of the Ω -loop in class A β -lactamases [20]. However, if the influence of the two categories of water molecules on the crystal structures is taken into account, the computational complexity will be greatly increased. Previous studies have found that the CWMs not only stay in a certain space in the binding sites but also directly participate in protein-ligand interactions. Hence, to provide necessary insights for the conformational stability of the macromolecules and to refine the protein-ligand binding and the structural optimization of the ligands, it is necessary to effectively identify the CWMs from the FWMs in the binding sites.

Mainly due to the limitations in X-ray crystallography technology, neutron diffraction, or nuclear magnetic resonance, the position information of water molecules is often inaccurate or not accessible [22]. Therefore, it is difficult to identify the CWMs in the binding sites directly. Currently, four categories of computational methods are mainly used

to determine their potential sites in practice [22]. The first category is the simulation-based methods which adopt the *molecular dynamics* (MD) or *Monte Carlo* (MC) simulations to predict the most possible transition status of water molecules in the binding sites. Typical methods include Water-Map [23], Dowser++ [24], and JAWS [25]. For example, JAWS [25] performs with a Metropolis MC scheme to locate the water molecules in the binding sites of a protein or protein-ligand complex. The simulation-based methods can accurately determine the water molecules' sites and obtain their conformation structures. However, this method comes at a cost of high computational complexity; the second category is based on empirical methods [26, 27], which mainly discriminate the water molecules by extracting their significant features (such as the temperature B-factor, solvent-contact surface area, and numbers of protein-water interactions). Hence, the extraction and selection of certain specific features can greatly affect the prediction and migration ability of the models. Differently, the third group, i.e., the knowledge-based methods [28–30], extracts the large-scale experimental data information and summarizes them into “knowledge” which can be used to aid the model prediction. However, to fit the models with high reliability, this category has special requirements on the experimental data's quantities and types. Methods in the fourth category (such as 3D-RISM [31], GIST [32], and GRID [33]) are the grid-based interaction methods in which an array of the grid points are generated first throughout and around the protein, then utilized to calculate the interaction potential [33]. The methods allow many thermodynamic quantities to be calculated in a fraction of the time. However, it is difficult to extract the physical information from the atomic-site density distributions [34]. Over the past few decades, machine learning techniques have been widely applied in solving the problem, such as analysis, classification, and prediction in big data; thus, it is developing rapidly in bioinformatics research [35–37].

Motivated by the above discussions, in this study, a machine learning-based method was presented to predict the CWMs in proteins' binding sites. First, the homologous protein structures of the training dataset were collected and overlapped, and the protein structure pairs with a large root-mean-square deviation (RMSD) value were filtered out. Then, the nearest Euclidean distance (NED) between the water molecule in the binding site and the nearest water molecule in the overlapping protein was calculated. Following the definition in [15], a water molecule with a distance less than and equal to 1.2 Å was defined as the CWM; otherwise, it was defined as the FWM. Next, by analyzing the physicochemical properties and the spatial structure information of each water molecule, six important features (i.e., atom density, hydrophilicity, hydrophobicity, solvent-accessible surface area, temperature B-factors, and mobility) were extracted. Based on this, a feature selection method was adopted to evaluate different feature combinations. As a result, the optimal combination with the best prediction performance was determined. Furthermore, seven machine learning models (i.e., support vector machine (SVM) [38, 39], *K*-nearest neighbor (KNN) [26], decision tree (DT)

[40], logistic regression (LR) [41], discriminant analysis (DA) [42], naïve Bayes (NB) [43], and ensemble learning (EL) [44] were adopted to evaluate their discriminating performance based on the optimal feature combination. Finally, the EL model was investigated as the desired model to identify the CWMs. At last, the performance of the proposed model was evaluated against a test set and further compared with Dowser++. The results revealed that the CWMs could be accurately identified by the proposed feature combination and the machine learning model.

2. Methods

2.1. Data Collection and Processing. Based on the previous work [22], 2003 pairs of protein-ligand crystal structures with a resolution less than 2.0 Å were collected as the training set. Since the conformational and chemical differences between the homologous protein pairs may affect the position comparison of the water molecules, the overlapping was performed using the Pymol software [45]. Only the homologous protein pairs with the RMSD less than or equal to 2.0 Å were retained.

Taking 1D7R (i.e., the crystal structure of the complex of 2,2-dialkylglycine decarboxylase with 5PA [46]) as an example (see Figure 1), the detailed training procedure was shown.

- (I) *Align and overlap* the homologous crystal structure 1M0Q on 1D7R such that the homologous protein pair was in the same coordinate system;
- (II) *Form* the binding pocket by the protein atoms within a distance of 7.0 Å of any ligand atoms [27, 47] centered on the center point of the ligand in 1D7R;
- (III) In the binding site of 1D7R, there were seven water molecules (magenta spheres). For each water molecule, *calculate* the corresponding NED to the water molecules (green spheres) in 1M0Q;
- (IV) *Determine* the CWMs using 1.2 Å [15] as a threshold for the NEDs between the oxygen atoms of the two water molecules in the homologous protein pair. When the NEDs were less than and equal to 1.2 Å, the water molecules in the original crystal structures were regarded as the CWMs (yellow spheres). Otherwise, they were referred to as the FWMs (cyan spheres).

Based on the above data processing steps, the proportion of the number of the FWMs against that of the CWMs in the training set was around 1 : 1.25.

2.2. Feature Extraction of Water Molecules. After the training dataset was processed, the extraction of effective features was important for the prediction accuracy of the training model. In this work, by analyzing the physicochemical properties and the spatial structure information of the water molecules in the binding sites, the following six features were extracted to characterize their microenvironments.

- (I) *Atom Density.* It was defined as the number of protein atoms within a distance of 3.6 Å of each water molecule [22]. Due to the influence of the morphology of the protein surface, the atom density in the concave groove was normally higher than that in the convex. As a result, the water molecules in the concave grooves tend to interact more with the surrounding polar atoms; thus, they were considered to be highly conservative.
- (II) *Atomic Hydrophilicity.* By analyzing the surface-bounded water molecules in 56 high-resolution crystal structures, the individual hydration propensities for each type of amino acid atoms, h_i , could be determined by dividing the total number of the water molecules that hydrates an atom by the number of the surface-exposed occurrences [48]. Based on this, the atomic hydrophilicity [18] (Equation (1)) could be calculated by the weighted summation of the propensities from all the atoms (denoted by N) within 4 Å of the water molecule, i.e.,

$$\sum_{i=1}^N h_i e^{r_i/d_0}, \quad (1)$$

where r_i was the distance between the atom i and a water molecule, and d_0 was the distance scale of the interaction.

- (III) *Atomic Hydrophobicity.* The hydrophobicity properties of the protein-ligand interfaces varied with proteins, and they reflected the local chemical environment of the water molecule. For the lipophilic score as considered in this work, the corresponding atomic hydrophobicity [18], i.e.,

$$\sum_{i=1}^N l_i e^{-r_i/d_0}, \quad (2)$$

where l_i was the carbon propensity of the atom i . The other variables were defined the same as in Equation (1).

- (IV) *Solvent-Accessible Surface Area (SASA).* SASA was a measure of the accessibility of water molecules to the outer bulk aqueous environment. As mentioned earlier, the water molecules in the concave grooves on the surface of proteins had fewer contacts with the surrounding aqueous environment as compared to that in the convex. Normally, the NACCESS program [49] was adopted to calculate the SASA of both CWMs and FWMs in the area of concave groove and convex.
- (V) *Temperature B-Factors (BFs).* The BF [27] was often used to measure the atomic stability level in a crystal structure, which was obtained through the square of the average displacement, \bar{U} , of an atom, as shown below:

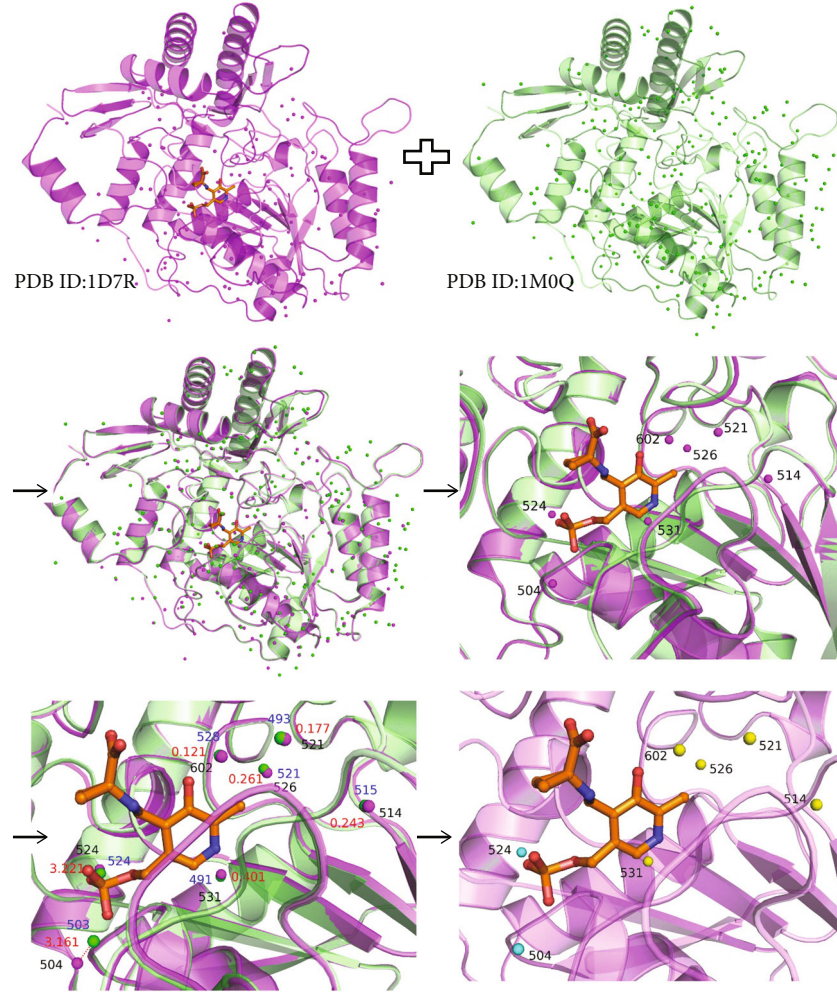


FIGURE 1: The training process of the dataset. The crystal structure 1D7R is marked in magenta. The crystal structure 1M0Q (i.e., the crystal structure of dialkylglycine decarboxylase complexed with S-1-aminoethanephosphonate [43]) marked in green is the homologous protein of the crystal structure 1D7R. The ligand in the conformation of the crystal structure 1D7R is shown as an orange ball-and-sticks. The magenta and green spheres represent the water molecules in the crystal structures 1D7R and 1M0Q, respectively, while the yellow and cyan ones represent the CWMs and FWMs in the crystal structure 1D7R, respectively. The distances between each of the two water molecules are indicated in red.

$$BF = 8\pi^2 \bar{U}^2. \quad (3)$$

$$M = \frac{BF_i / (\sum_{i=1}^m BF_i / m)}{O_i / (\sum_{i=1}^m O_i / m)}, \quad (4)$$

The BF value reflected the trend of position changing of water molecules in the structure. Generally, the more flexible an atom was, the greater its displacement from its average position. Therefore, water molecules with higher BF values had stronger fluidity than those with lower ones.

(VI) *Mobility*. Instead of staying at a fixed position in a protein, water molecules tend to move around within a certain range. To measure the mobility, M , differences of the water molecules, the following equation was adopted to calculate the displacement degree of an atom from its average position:

where BF_i and O_i were the average values of temperature B-factors and the occupancy rates of the i th atom, respectively.

The combinations of these features were further evaluated in the following to choose the optimal combination in terms of the CWM identification performance.

2.3. *Prediction Models*. Based on the above six features, the seven most sophisticated machine learning models were adopted to evaluate their performance in terms of CWM identification in the binding sites of the proteins. These models included the SVM, KNN, DT, LR, DA, NB, and EL.

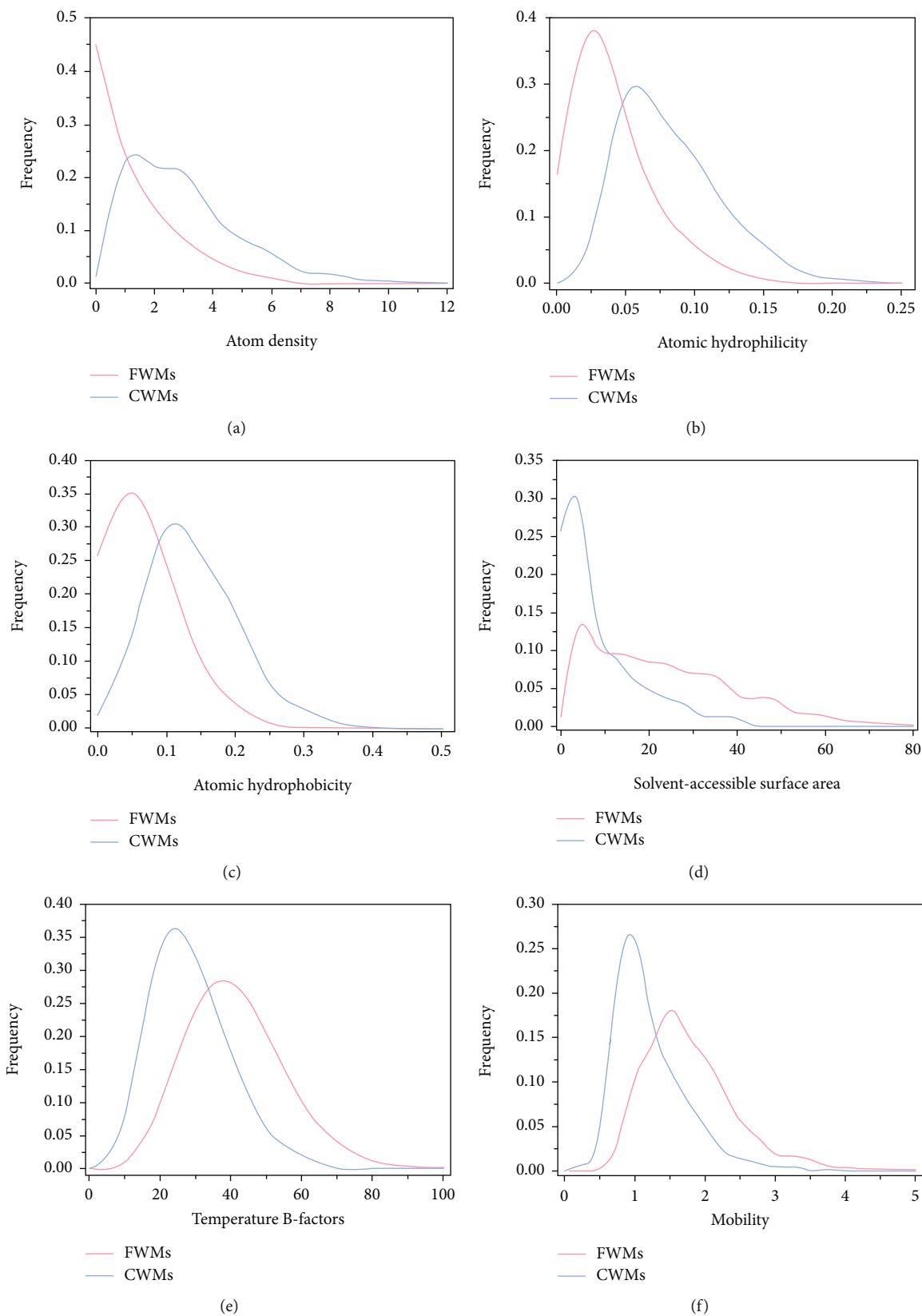


FIGURE 2: Distributions of different features: (a) atom density; (b) atomic hydrophilicity; (c) atomic hydrophobicity; (d) solvent-accessible surface area; (e) temperature B-factors; (f) mobility. The blue and red curves represent the distributions of the features for the CWMs and FWMs, respectively.

TABLE 1: The minimum, maximum, and average values of the features.

Categories of water molecules	Values	Features					
		Atom density	Atomic hydrophilicity	Atomic hydrophobicity	SASA (\AA^2)	BFs	Mobility
FWMs	Min	0.000	0.000	0.000	0.000	0.000	0.000
	Max	7.000	0.147	0.249	84.949	99.930	8.992
	Mean	1.146	0.029	0.049	21.877	36.371	1.673
CWMs	Min	0.000	0.005	0.000	0.000	0.000	0.027
	Max	12.000	0.243	0.505	40.877	94.67	12.289
	Mean	3.033	0.071	0.116	6.683	23.740	1.099

2.4. Performance Assessment. In order to quantify the performance of different prediction models, a quality measure was required in order to evaluate the validity of the different feature combinations selected. The performance of different feature combinations and prediction models was further evaluated by considering the following aspects: accuracy (ACC), sensitivity (SN), positive predictive value (PPV), and F -score. Mathematically, these parameters were defined in Equations (5)–(8), respectively:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (5)$$

$$\text{SN} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (6)$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (7)$$

$$F\text{-score} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}}, \quad (8)$$

where TP and TN meant the numbers of the *true positive* and the *true negative*, respectively, while FP and FN indicated the numbers of the *false positive* and the *false negative*, respectively. More specifically, in this study, they were defined, respectively, as follows:

True positive: CWMs that were correctly identified as CWMs.

False positive: FWMs that were incorrectly identified as CWMs.

True negative: FWMs that were correctly identified as FWMs.

False negative: CWMs that were incorrectly identified as FWMs.

Besides, as a useful tool to assess the ability of the prediction model, the *area under the receiver operating characteristic curve* (AUC) was also considered to evaluate their performance. Note that, for all the prediction models, a five-fold cross-validation procedure was adopted to avoid overfitting issues.

3. Results and Discussions

3.1. Analysis of Features. The positions of water molecules in the binding sites of protein were influenced by various factors. To identify the CWMs among them in a more effective

way, the distributions of all six features were analyzed in Figure 2. Moreover, the minimum, maximum, and average values of these features for both CWMs and FWMs were listed in Table 1.

From Figure 2, it was obvious that the distributions of all six features of the CWMs and FWMs did not overlap completely. Take the atom density (Figure 2(a)), hydrophilicity (Figure 2(b)), and hydrophobicity (Figure 2(c)) of the CWMs for examples, their *modes* (i.e., the most frequent value in a dataset) were around 1.00, 0.06, and 0.12, respectively, which were all larger than those of the FWMs. This was due to the fact that the CWMs were generally located in the concave grooves of the binding sites, while the FWMs tended to be on the convex surface. In this sense, the atoms around the CWMs were more densely packed than those around the FWMs. As for the B-factors (Figure 2(e)) and the mobility (Figure 2(f)), the corresponding *modes* for the CWMs were around 25 and 1, respectively, which were smaller than those for the FWMs. It was mainly because the CWMs were relatively stable and had less displacement from their average positions. Accordingly, their temperature B-factors and mobility values were relatively small. However, their *modes* of distributions of SASA (Figure 2(d)) for the two categories of water molecules were about 5, which were roughly the same despite the significant frequency differences.

In a summary, the distributions of these six features for the two categories of water molecules overlapped greatly. This made it challenging to identify the CWMs from the FWMs in the proteins' binding sites using one feature alone. Therefore, it motivated us to explore the benefits of the combined features.

3.2. Evaluation of Feature Combinations. As shown in Figure 2, it was difficult to identify the key water molecules directly using a single feature alone; we instead considered the combined features to discriminate their final performance. In order to find the desired feature combination in a reasonable way, we evaluated their averaged performance of ACCs, SNs, PPVs, F -scores, and AUCs under the seven most commonly used machine learning models (i.e., SVM, KNN, DT, LR, DA, NB, and EL). For example, the averaged ACC, i.e., $\overline{\text{ACC}}$, was defined as the averaged value of all ACC values from all the models, that is,

$$\overline{\text{ACC}} = \frac{\sum_i^n \text{ACC}_i}{n}, \quad (9)$$

TABLE 2: Averaged performance indices under different feature combinations.

No.	Combination	\overline{ACC}	\overline{SN}	\overline{PPV}	$\overline{F\text{-score}}$	\overline{AUC}
1	ABCDEF*	0.725**	0.809	0.753	0.779	0.790
2	ABCDE	0.717	0.799	0.749	0.773	0.774
3	ABCDF	0.723	0.811	0.751	0.779	0.790
4	ABCEF	0.723	0.810	0.751	0.778	0.787
5	ABDEF	0.710	0.814	0.733	0.771	0.770
6	ACDEF	0.720	0.815	0.744	0.778	0.780
7	BCDEF	0.724	0.807	0.753	0.778	0.786
8	ABCD	0.719	0.804	0.749	0.774	0.771
9	ABCE	0.718	0.798	0.751	0.773	0.771
10	ABCF	0.724	0.808	0.754	0.780	0.786
11	ABDE	0.698	0.810	0.722	0.763	0.753
12	ABDF	0.711	0.817	0.734	0.773	0.771
13	ABEF	0.708	0.807	0.735	0.769	0.771
14	ACDE	0.711	0.813	0.735	0.771	0.766
15	ACDF	0.722	0.821	0.744	0.780	0.782
16	ACEF	0.719	0.814	0.743	0.777	0.780
17	ADEF	0.703	0.847	0.714	0.775	0.745
18	BCDE	0.718	0.797	0.751	0.773	0.771
19	BCDF	0.723	0.806	0.753	0.778	0.787
20	BCEF	0.719	0.799	0.752	0.774	0.783
21	BDEF	0.710	0.811	0.735	0.771	0.773
22	CDEF	0.720	0.812	0.746	0.777	0.780
23	ABC	0.718	0.802	0.749	0.774	0.771
24	ABD	0.699	0.818	0.720	0.765	0.753
25	ABE	0.697	0.810	0.721	0.763	0.754
26	ABF	0.709	0.810	0.735	0.770	0.773
27	ACD	0.712	0.816	0.735	0.773	0.762
28	ACE	0.707	0.807	0.734	0.768	0.760
29	ACF	0.720	0.819	0.743	0.779	0.783
30	ADE	0.679	0.827	0.696	0.755	0.705
31	ADF	0.703	0.844	0.715	0.774	0.745
32	AEF	0.701	0.836	0.716	0.771	0.739
33	BCD	0.718	0.799	0.750	0.773	0.773
34	BCE	0.713	0.785	0.751	0.767	0.767
35	BCF	0.716	0.796	0.749	0.771	0.777
36	BDE	0.699	0.811	0.723	0.764	0.754
37	BDF	0.712	0.816	0.735	0.773	0.773
38	BEF	0.705	0.798	0.735	0.765	0.771
39	CDE	0.712	0.808	0.738	0.771	0.767
40	CDF	0.722	0.819	0.745	0.780	0.783
41	CEF	0.721	0.817	0.745	0.779	0.780
42	DEF	0.703	0.835	0.718	0.772	0.748
43	AB	0.695	0.815	0.717	0.763	0.740
44	AC	0.708	0.820	0.729	0.772	0.760
45	AD	0.673	0.848	0.684	0.757	0.691
46	AE	0.676	0.846	0.688	0.759	0.693
47	AF	0.700	0.847	0.711	0.773	0.739
48	BC	0.707	0.789	0.742	0.764	0.755

TABLE 2: Continued.

No.	Combination	\overline{ACC}	\overline{SN}	\overline{PPV}	$\overline{F-score}$	\overline{AUC}
49	BD	0.699	0.815	0.721	0.765	0.756
50	BE	0.694	0.802	0.722	0.759	0.750
51	BF	0.702	0.793	0.734	0.761	0.766
52	CD	0.713	0.817	0.736	0.774	0.764
53	CE	0.704	0.795	0.736	0.764	0.759
54	CF	0.713	0.813	0.737	0.772	0.773
55	DE	0.677	0.830	0.693	0.755	0.703
56	DF	0.702	0.840	0.716	0.772	0.746
57	EF	0.697	0.828	0.715	0.767	0.740
58	A	0.664	0.867	0.667	0.759	0.662
59	B	0.674	0.796	0.703	0.745	0.729
60	C	0.691	0.811	0.714	0.759	0.738
61	D	0.670	0.842	0.684	0.754	0.693
62	E	0.670	0.838	0.685	0.754	0.690
63	F	0.685	0.842	0.697	0.763	0.726

*A, B, C, D, E, and F represent the features of the atom density, mobility, temperature B-factors, atomic hydrophilicity, atomic hydrophobicity, and SASA, respectively. **In each category, we highlight the values of the best performance in bold. Note that we allow ± 0.001 deviations for the values. For example, the \overline{ACC} values of the best performance are 0.725 and 0.724, respectively.

where i indicated the i th machine learning model, $i = 1, \dots, n$. In this study, n was chosen as 7. \overline{SN} , \overline{PPV} , $\overline{F-score}$, and \overline{AUC} were defined in a similar way. The results were shown in Table 2 with the detailed table attached in Table S1.

As can be seen from Table 2, in terms of \overline{ACC} , the highest rates of predicting CWMs and FWMs correctly to the total predictions could be achieved by feature combinations No. 1, No. 7, and No. 10 with respective values: 0.725, 0.724, 0.724 (± 0.001). These results indicated that combining features in a reasonable way could lead to a better identification ability of the water molecules in the binding sites of proteins. However, a single criterion may cause the loss of the generality. Hence, a comprehensive evaluation of the varied performance resulting from different feature combinations was necessary. To this end, in the following, other commonly used criteria such as \overline{SN} , \overline{PPV} , $\overline{F-score}$, and \overline{AUC} were considered as well. \overline{SN} was a measure of how effective the prediction model could identify the actual positives (CWMs). It turned out the feature combination No. 58 gave the highest \overline{SN} value of 0.867. This indicated that the feature of the atom density was important in correctly identifying CWMs. As for \overline{PPV} , which reflected the precision of identifying the CWM, as a result, the combinations No. 1, No. 7, No. 10, and No. 19 achieved the highest values of 0.754 (± 0.001). When it came to $\overline{F-score}$, which was determined by both PPV and SN (see Equation (8)), the feature combinations No. 1 and No. 10 performed better than other combinations. As for \overline{AUC} , the feature combinations No. 1 and No. 3 gave the best CWM prediction performance with a value of 0.790. Given the above analyses, it was easy to conclude that feature combination No. 1 achieved the best performance in four (i.e., \overline{ACC} , \overline{PPV} , $\overline{F-score}$, and \overline{AUC}) out of the five criteria. Naturally, feature combination No. 1 was chosen as the optimal feature combination for the

following analysis, which indicated that the water molecules in the binding sites of proteins could be identified more accurately by combining all six features.

3.3. Comparison of Prediction Models. Based on the chosen optimal feature combination, the performance of seven commonly used machine learning models in identifying water molecules in the binding sites of proteins was evaluated with results shown in Table 3.

It could be seen from Table 3 that different models were accompanied by their respective performances. Among them, the EL model performed best in four (i.e., ACC, SN, $\overline{F-score}$, and AUC) out of five criteria, and also, its average performance value was 0.853, which was the highest among all the models. However, in terms of PPV, the DT model posed advantages over other models. After comprehensively considering their performance in terms of different kinds of criteria, it was not hard to conclude that the EL model performed better in identifying the water molecules in the binding sites of the proteins. Therefore, the EL model was selected as the desired prediction model.

3.4. Case Study. In the following, we took 3skh (i.e., the crystal structure of I. Novel HCV NS5B polymerase inhibitors: discovery of indole 2-carboxylic acids with C3-heterocycles [46]) as a case study, where eleven water molecules were distributed in the binding site of Chain B of the crystal structure 3skh (Figure 3(a)). Among them, the W788 was an FWM (cyan sphere), and the others were CWMs (yellow spheres). By employing the EL model (Figure 3(b)), it could successfully identify all the CWMs but failed on the FWM W788 (magenta sphere). It showed that the prediction model could achieve satisfactory accuracies in predicting the CWMs in the binding site of Chain B of the crystal structure 3skh.

TABLE 3: Performance comparison of seven machine learning models in identifying water molecules in the binding sites of proteins using the optimal feature combination.

Prediction models	ACC	SN	PPV	<i>F</i> -score	AUC	Average performance**
SVM	0.809	0.889	0.793	0.838	0.880	0.842
KNN	0.805	0.873	0.797	0.833	0.890	0.840
DT	0.805	0.838	0.817	0.827	0.900	0.837
LR	0.795	0.831	0.807	0.819	0.870	0.824
DA	0.793	0.836	0.801	0.818	0.870	0.824
NB	0.798	0.828	0.812	0.820	0.890	0.830
EL	0.817*	0.890	0.803	0.844	0.910	0.853

*Bold values indicate the highest performance values. **For each model, the average performance is defined by averaging out all the values from five criteria.

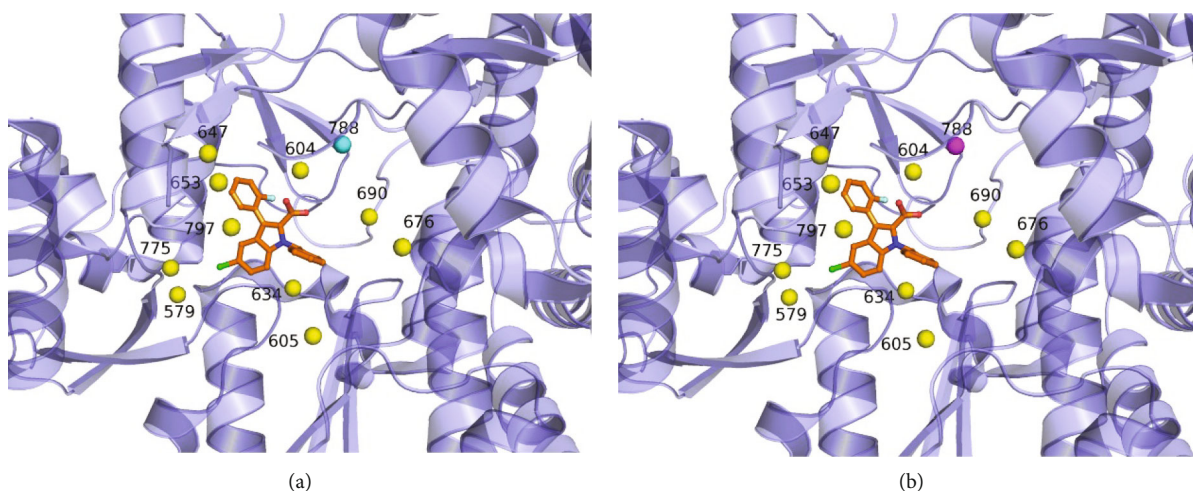


FIGURE 3: (a) All water molecules in the binding site of Chain B of the crystal structure 3skh, where the yellow and cyan spheres represent the CWMs and FWM, respectively. (b) The predicted results using the EL model, where the yellow spheres represent the correctly identified CWMs, and the magenta sphere represents the mispredicted FWM. Note that the ligands of Chain B in these conformations are shown as the orange ball-and-stick models.

3.5. Comparison with Other Methods. In this section, the performance of our method in identifying the CWMs in the proteins' binding sites had been compared with Dowser++ [24] using the same test set [22]. The Dowser++ was based on a semiempirical modification of a program for protein hydration Dowser [50], AutoDock Vina [51], and WaterDock [18]. The six features and the categories of water molecules in the test set were collected in Table S2. Encouragingly, the accuracies of the proposed EL model in predicting the CWMs could reach 77.0% (the detailed predicted results were attached in Table S3), as compared with 59.3% by using Dowser++ (the detailed predicted results were attached in Table S3). These results demonstrated that our method was performing better in predicting the CWMs in the proteins' binding sites.

4. Conclusion

In this study, a machine learning-based approach was proposed to identify the CWMs in proteins' binding sites. By analyzing the physicochemical properties and the spatial structure information of the water molecules, six features

were extracted to characterize their surrounding microenvironment. A feature selection method was used to train and evaluate different feature combinations, and the optimal combination with better performance was determined. On this basis, seven machine learning models were introduced to evaluate their abilities in identifying the two categories of water molecules. As a result, the EL model with better performance was selected according to various evaluations. A test set was used to verify the effectiveness of the optimal feature combination and the chosen prediction models in our method and compared to Dowser++. The results indicated that our method demonstrated strong performance, which further showed that the desired feature combination and prediction model proposed in this study could effectively identify the CWMs in proteins' binding sites.

Abbreviations

CWMs: Conserved water molecules
 SVM: Support vector machine
 KNN: *K*-nearest neighbor
 DT: Decision tree

LR: Logistic regression
 DA: Discriminant analysis
 NB: Naïve Bayes
 EL: Ensemble learning
 FWMs: Free water molecules
 RMSD: Root-mean-square deviation
 NED: Nearest Euclidean distance
 SASA: Solvent-accessible surface area
 BFs: Temperature B-factors
 ACC: Accuracy
 SN: Sensitivity
 PPV: Positive predictive value
 AUC: Area under the receiver operating characteristic curve.

Data Availability

The datasets supporting the conclusions of this article are included in the additional files.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' Contributions

Liangzhao Lin conceived the study and revised the manuscript. Wei Xiao developed the methods, performed the analysis, and drafted the manuscript. Juhui Ren, Haoyu Wang, Yuhao Li, and Jutao Hao contributed to developing the program and participated in manuscript preparation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China through Grants (Nos. 61703267 and 61802247).

Supplementary Materials

Additional file 1: average values of the performance parameters of different feature combinations obtained by using seven models. Additional file 2: the results of the six features and the categories of water molecules using the test set. Additional file 3: prediction results obtained using the optimal feature combination and the chosen prediction models in our method. Additional file 4: prediction results obtained using the program Dowser++. (*Supplementary Materials*)

References

- [1] F. Spyrakis, M. H. Ahmed, A. S. Bayden, P. Cozzini, A. Mozzarelli, and G. E. Kellogg, "The roles of water in the protein matrix: a largely untapped resource for drug discovery," *Journal of Medicinal Chemistry*, vol. 60, no. 16, pp. 6781–6827, 2017.
- [2] Y. Li, Y. D. Gao, M. K. Holloway, and R. X. Wang, "Prediction of the favorable hydration sites in a protein binding pocket and its application to scoring function formulation," *Journal of Chemical Information and Modeling*, vol. 60, no. 9, pp. 4359–4375, 2020.
- [3] E. Nittinger, N. Schneider, G. Lange, and M. Rarey, "Evidence of water molecules—a statistical evaluation of water molecules based on electron density," *Journal of Chemical Information and Modeling*, vol. 55, no. 4, pp. 771–783, 2015.
- [4] R. E. Skyner, J. L. McDonagh, C. R. Groom, T. Van Mourik, and J. B. O. Mitchell, "A review of methods for the calculation of solution free energies and the modelling of systems in solution," *Physical Chemistry Chemical Physics*, vol. 17, no. 9, pp. 6174–6191, 2015.
- [5] J. Liu, X. He, and J. Z. Zhang, "Improving the scoring of protein-ligand binding affinity by including the effects of structural water and electronic polarization," *Journal of Chemical Information and Modeling*, vol. 53, no. 6, pp. 1306–1314, 2013.
- [6] D. Cappel, R. Wahlström, R. Brenk, and C. A. Sotriffer, "Probing the dynamic nature of water molecules and their influences on ligand binding in a model binding site," *Journal of Chemical Information and Modeling*, vol. 51, no. 10, pp. 2581–2594, 2011.
- [7] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, "Pairwise solute descreening of solute charges from a dielectric medium," *Chemical Physics Letters*, vol. 246, no. 1–2, pp. 122–129, 1995.
- [8] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, "Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium," *The Journal of Physical Chemistry. B*, vol. 100, no. 51, pp. 19824–19839, 1996.
- [9] H. Y. Liu, I. D. Kuntz, and X. Zou, "Pairwise GB/SA scoring function for structure-based drug design," *The Journal of Physical Chemistry. B*, vol. 108, no. 17, pp. 5453–5462, 2004.
- [10] J. E. Ladbury, "Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design," *Chemistry & Biology*, vol. 3, no. 12, pp. 973–980, 1996.
- [11] A. Biela, M. Khayat, H. Tan et al., "Impact of ligand and protein desolvation on ligand binding to the S1 pocket of thrombin," *Journal of Molecular Biology*, vol. 418, no. 5, pp. 350–366, 2012.
- [12] W. Wang, O. Donini, C. M. Reyes, and P. A. Kollman, "biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions," *Annual Review of Biophysics and Biomolecular Structure*, vol. 30, no. 1, pp. 211–243, 2001.
- [13] S. A. Adcock and J. A. McCammon, "Molecular dynamics: survey of methods for simulating the activity of proteins," *Chemical Reviews*, vol. 106, no. 5, pp. 1589–1615, 2006.
- [14] S. E. Graham, R. D. Smith, and H. A. Carlson, "Predicting displaceable water sites using mixed-solvent molecular dynamics," *Journal of Chemical Information and Modeling*, vol. 58, no. 2, pp. 305–314, 2018.
- [15] E. Nittinger, F. Flachsenberg, S. Bietz, G. Lange, R. Klein, and M. Rarey, "Placement of water molecules in protein structures: from large-scale evaluations to single-case examples," *Journal of Chemical Information and Modeling*, vol. 58, no. 8, pp. 1625–1637, 2018.
- [16] V. A. Likić, N. Juranić, S. Macura, and F. G. Prendergast, "A "structural" water molecule in the family of fatty acid binding proteins," *Protein Science*, vol. 9, no. 3, pp. 497–504, 2000.

- [17] S. Fischer and C. S. Verma, "Binding of buried structural water increases the flexibility of proteins," *Proceedings of the National Academy of Sciences*, vol. 96, no. 17, pp. 9613–9615, 1999.
- [18] G. A. Ross, G. M. Morris, and P. C. Biggin, "Rapid and accurate prediction and scoring of water molecules in protein binding sites," *PLoS One*, vol. 7, no. 3, article e32036, 2012.
- [19] P. C. Sanschagrin and L. A. Kuhn, "Cluster analysis of consensus water sites in thrombin and trypsin shows conservation between serine proteases and contributions to ligand specificity," *Protein Science*, vol. 7, no. 10, pp. 2054–2064, 1998.
- [20] L. Craig, P. C. Sanschagrin, A. Rozek, S. Lackie, L. A. Kuhn, and J. K. Scott, "The role of structure in antibody cross-reactivity between peptides and folded proteins¹," *Journal of Molecular Biology*, vol. 281, no. 1, pp. 183–201, 1998.
- [21] G. Mustata and J. M. Briggs, "Cluster analysis of water molecules in alanine racemase and their putative structural role," *Protein Engineering, Design & Selection*, vol. 17, no. 3, pp. 223–234, 2004.
- [22] W. Xiao, Z. H. He, M. J. Sun, S. L. Li, and H. L. Li, "Statistical analysis, investigation and prediction of the water positions in the binding sites of proteins," *Journal of Chemical Information and Modeling*, vol. 57, no. 7, pp. 1517–1528, 2017.
- [23] L. Wang, B. Berne, and R. Friesner, "Ligand binding to protein-binding pockets with wet and dry regions," *Proceedings of the National Academy of Sciences*, vol. 108, no. 4, pp. 1326–1330, 2011.
- [24] A. Morozenko and A. A. Stuchebrukhov, "Dowser++, a new method of hydrating protein structures," *Proteins*, vol. 84, no. 10, pp. 1347–1357, 2016.
- [25] J. Michel, J. Tirado-Rives, and W. L. Jorgensen, "Prediction of the water content in protein binding sites," *The Journal of Physical Chemistry. B*, vol. 113, no. 40, pp. 13337–13346, 2009.
- [26] M. L. Raymer, P. C. Sanschagrin, W. F. Punch, S. Venkataraman, E. D. Goodman, and L. A. Kuhn, "Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm¹," *Journal of Molecular Biology*, vol. 265, no. 4, pp. 445–464, 1997.
- [27] A. T. Garcia-Sosa, R. L. Mancera, and P. M. Dean, "WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes," *Journal of Molecular Modeling*, vol. 9, no. 3, pp. 172–182, 2003.
- [28] G. Rossato, B. Ernst, A. Vedani, and M. Smiesko, "AcquaAlta: a directional approach to the solvation of ligand-protein complexes," *Journal of Chemical Information and Modeling*, vol. 51, no. 8, pp. 1867–1881, 2011.
- [29] M. Y. Zheng, Y. L. Li, B. Xiong, H. L. Jiang, and J. K. Shen, "Water PMF for predicting the properties of water molecules in protein binding site," *Journal of Computational Chemistry*, vol. 34, no. 7, pp. 583–592, 2013.
- [30] A. S. Bayden, D. T. Moustakas, D. Joseph-McCarthy, and M. L. Lamb, "Evaluating free energies of binding and conservation of crystallographic waters using SZMAP," *Journal of Chemical Information and Modeling*, vol. 55, no. 8, pp. 1552–1565, 2015.
- [31] T. Imai, R. Hiraoka, A. Kovalenko, and F. Hirata, "Locating missing water molecules in protein cavities by the three-dimensional reference interaction site model theory of molecular solvation," *Proteins*, vol. 66, no. 4, pp. 804–813, 2007.
- [32] T. Hufner-Wulsdorf and G. Klebe, "Advancing GIST-based solvent functionals through multiobjective optimization of solvent enthalpy and entropy scoring terms," *Journal of Chemical Information and Modeling*, vol. 60, no. 12, pp. 6654–6665, 2020.
- [33] P. J. Goodford, "A computational procedure for determining energetically favorable binding sites on biologically important macromolecules," *Journal of Medicinal Chemistry*, vol. 28, no. 7, pp. 849–857, 1985.
- [34] C. Nguyen, T. Yamazaki, A. Kovalenko et al., "A molecular reconstruction approach to site-based 3D-RISM and comparison to GIST hydration thermodynamic maps in an enzyme active site," *PLoS One*, vol. 14, no. 7, article e0219473, 2019.
- [35] W. Yue, Z. Wang, H. Che, A. Payne, and X. Liu, "Machine learning with applications in breast cancer diagnosis and prognosis," *Designs*, vol. 2, no. 2, pp. 13–29, 2018.
- [36] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, 2017.
- [37] S. M. Yusuf, F. Zhang, M. Zeng, and M. Li, "DeepPPF: a deep learning framework for predicting protein family," *Neurocomputing*, vol. 428, pp. 19–29, 2021.
- [38] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer Science & Business Media, 1999.
- [39] W. Li, Y. Zhuo, J. Bao, and Y. Shen, "A data-based soft-sensor approach to estimating raceway depth in ironmaking blast furnaces," *Powder Technology*, vol. 390, pp. 529–538, 2021.
- [40] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Routledge, 2017.
- [41] S. Z. Mousavi, A. Kaviani, K. Soleimani, S. R. Mousavi, and A. Shirzadi, "GIS-based spatial prediction of landslide susceptibility using logistic regression model," *Geomatics, Natural Hazards and Risk*, vol. 2, no. 1, pp. 33–50, 2011.
- [42] Y. Guo, T. Hastie, and R. Tibshirani, "Regularized linear discriminant analysis and its application in microarrays," *Biostatistics*, vol. 8, no. 1, pp. 86–100, 2007.
- [43] I. Rish, "An empirical study of the naive Bayes classifier," *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, pp. 41–46, 2001.
- [44] C. Zhang and Y. Ma, *Ensemble Machine Learning: Methods and Applications*, Springer Science & Business Media, 2012.
- [45] W. L. DeLano, "Pymol: an open-source molecular graphics tool," *CCP4 Newsletter on Protein Crystallography*, vol. 40, pp. 82–92, 2002.
- [46] H. M. Berman, J. Westbrook, Z. Feng et al., "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.
- [47] M. L. Verdonk, G. Chessari, J. C. Cole et al., "Modeling water molecules in protein-ligand docking using GOLD," *Journal of Medicinal Chemistry*, vol. 48, no. 20, pp. 6504–6515, 2005.
- [48] L. A. Kuhn, C. A. Swanson, M. E. Pique, J. A. Tainer, and E. D. Getzoff, "Atomic and residue hydrophilicity in the context of folded protein structures," *Proteins*, vol. 23, no. 4, pp. 536–547, 1995.
- [49] S. J. Hubbard and P. Argos, "Detection of internal cavities in globular proteins," *Protein Engineering, Design & Selection*, vol. 8, no. 10, pp. 1011–1015, 1995.
- [50] A. Morozenko, I. V. Leontyev, and A. A. Stuchebrukhov, "Dipole moment and binding energy of water in proteins from crystallographic analysis," *Journal of Chemical Theory and Computation*, vol. 10, no. 10, pp. 4618–4623, 2014.
- [51] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of Computational Chemistry*, vol. 31, no. 2, pp. 455–461, 2010.

Research Article

Construction of Prognostic Risk Model for Small Cell Lung Cancer Based on Immune-Related Genes

Feng Deng , Feng Tao , Zhili Xu , Jun Zhou , Xiaowei Gong , and Ruhui Zhang 

Thoracic Surgery Jiangyin Hospital Affiliated to Nanjing University of Chinese Medicine, Jiangsu, China

Correspondence should be addressed to Xiaowei Gong; xw5_gong@163.com and Ruhui Zhang; zrhys7@163.com

Received 21 July 2022; Revised 30 August 2022; Accepted 7 September 2022; Published 30 September 2022

Academic Editor: Tao Huang

Copyright © 2022 Feng Deng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Small cell lung cancer (SCLC) is a highly invasive and fatal malignancy. Research at the present stage implied that the expression of immune-related genes is associated with the prognosis in SCLC. Accordingly, it is essential to explore effective immune-related molecular markers to judge prognosis and treat SCLC. Our research obtained SCLC dataset from Gene Expression Omnibus (GEO) and subjected mRNAs in it to differential expression analysis. Differentially expressed mRNAs (DEmRNAs) were intersected with immune-related genes to yield immune-related differentially expressed genes (DEGs). The functions of these DEGs were revealed by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses. Thereafter, we categorized 3 subtypes of immune-related DEGs via K-means clustering. Kaplan-Meier curves analyzed the effects of 3 subtypes on SCLC patients' survival. Single-sample gene set enrichment analysis (ssGSEA) and ESTIMATE validated that the activation of different immune gene subtypes differed significantly. Finally, an immune-related-7-gene assessment model was constructed by univariate-Lasso-multiple Cox regression analyses. Riskscores, Kaplan-Meier curves, receiver operating characteristic (ROC) curves, and independent prognostic analyses validated the prognostic value of the immune-related-7-gene assessment model. As suggested by GSEA, there was a prominent difference in cytokine-related pathways between high- and low-risk groups. As the analysis went further, we discovered a statistically significant difference in the expression of human leukocyte antigen (HLA) proteins and costimulatory molecules expressed on the surface of CD274, CD152, and T lymphocytes in different groups. In a word, we started with immune-related genes to construct the prognostic model for SCLC, which could effectively evaluate the clinical outcomes and offer guidance for the treatment and prognosis of SCLC patients.

1. Introduction

Small cell lung cancer (SCLC) represents 15% of all lung cancer (LC) cases. 60%-65% of SCLC patients had metastasis at diagnosis [1, 2]. SCLC grows fast with high invasiveness, which results in the poor prognosis of patients [3]. Presently, the combination of radiotherapy and chemotherapy is the standard to treat SCLC patients [3]. Nevertheless, over recent years, works have revealed that only some patients with limited-stage SCLC benefit from such a treatment pattern (41% and 16% for the 2- and 5-year survival, respectively), but this pattern does not work well for patients at extensive-stage (distant metastasis of cancer), with 6.4% of overall response rate after treatment [4].

The emergence of targeted therapy has improved the prognoses of lung cancer patients but has only limited effects on patients with SCLC. To take an example, the epidermal growth factor receptor (EGFR) inhibitor erlotinib is successful in treating non-small cell lung cancer (NSCLC) but is less effective in treating SCLC [5]. The reason is that EGFR is not an effective target due to the extremely low mutation frequency in SCLC. Imatinib, a KIT gene-targeted drug, can effectively improve the prognoses of NSCLC patients but is not effective on SCLC patients [6, 7]. Imatinib cannot be an effective drug for SCLC because KIT mutation rarely exists in SCLC [6, 7]. Accordingly, the results are less satisfactory in SCLC patients' prognoses by using present treatment methods. Now, it is essential to seek

prognostic targets at the molecular level for clinicians to precisely assess the prognosis and guide the treatment of SCLC patients, thereby improving the poor prognoses of SCLC patients.

As reported, the malignancy of cancer is associated with the immune-infiltrating microenvironment apart from the regulation of tumor cells [8]. A report analyzed the connection between T-cells and clinical outcomes of tumor patients, revealing that the infiltration of lymph cells, mainly cytotoxicity T cells and memory T cells, are positively linked to favorable clinical outcomes in patients with malignancy [9]. At the same time, there are accumulating works proving that immune-related genes can either facilitate or suppress cancer by affecting the immune microenvironment in cancer. It is of great significance to find out immune-related markers at the molecular level. For instance, lncRNA KCNQ1OT1/CD155 axis facilitates antitumor immune response by regulating T-cell depletion status in colorectal cancer; and lncRNA KCNQ1OT1 level is an indicator of patients' prognoses by reflecting the immune status of tumor tissue [10]. Immune-related gene PDK1 is able to manipulate PD-L1 level in tumor tissue by mTOR signaling, and further affects the immune escape of tumor [11]. The tumor progression can effectively slow down by downregulating tumor immune escape via PDK1 [11]. Besides, traditional immune markers like PD-1/PD-L1/CTLA4 play a marked role in indicating cancer progression and drug guidance [12]. Given that, we posited that there was a connection between immune-related genes, immune infiltrating microenvironment in tumor tissue, and prognoses of patients with malignancies. And potential targets can be revealed for improving the clinical outcomes of these patients by investigating such a connection. To investigate, the immune-related regulation can effectively help with the development of clinical drugs and medication guidance.

We constructed a 7-gene model for assessment of the prognoses of SCLC patients based on immune-related genes in the GSE60052 dataset and ImmPort Shared Data (ImmPort). We divided samples into high- and low-risk groups based on their Riskscores. Differences in the expression of the immune-related proteins, human leukocyte antigen (HLA), and the immune checkpoint gene between SCLC high- and low-risk groups were revealed by Wilcoxon test to reveal the underlying function of the prognostic model in predicting the immune mode of samples. The biomarkers we herein revealed and the risk assessment model can be positive tools for SCLC patients' prognoses.

2. Materials and Methods

2.1. Data Resource. mRNA and clinical data from GSE60052 (platform: GPL11154) of SCLC were downloaded from the GEO database (Home-GEO-NCBI (<http://nih.gov>)), containing 7 healthy samples and 79 tumor samples. Gene data closely correlating cellular immune function were from the ImmPort.

2.2. Screening of Differentially Expressed Genes (DEGs) Related to Immune Regulatory Functions in Cells. The mRNA data of healthy and tumor samples in GSE60052 were differentially analyzed using the R package "limma" [13], and DEGs with $|\log_{2}FC| > 1.0$ and $FDR < 0.05$ were retained. Immune-related DEGs were sieved by taking an intersection of DEGs and genes

associated with immune regulation function in ImmPort. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed on immune-related DEGs using the R package "clusterProfiler" [14] (p value < 0.05).

2.3. Consensus Clustering and Subtype Assessment Based on Immune-Related DEGs. To identify the immune subtype of SCLC, K-means clustering algorithm [15] was performed on immune-related DEGs using the R package "ConsensusClusterPlus". The overall level of immune infiltration in subtype groups was assessed using the R package "ESTIMATE" (<https://bioinformatics.mdanderson.org/estimate/rpackage.html>). The immune score, matrix score, ESTIMATE score, and tumor purity calculated were subject to Wilcoxon analysis to plot violin plots in different subtypes. The activation of immune-related gene sets in different immune subtypes was assessed using R package "GSVA" [16] based on single-sample gene set enrichment analysis (ssGSEA) method.

2.4. Construction and Validation of Prognostic Models Related to Immune Regulatory Functions in Cells. Using the R package "survival" (survival: Survival Analysis (<http://r-project.org>)), univariate Cox regression analysis was performed on immune-related DEGs and those with p value < 0.05 were selected as the candidate genes. To prevent model overfitting, Lasso regression analysis of candidate genes was performed using the R package "glmnet" [17], and cross-validation was used to choose penalty parameter λ to remove genes with strong correlation to reduce model complexity. R package "survival" (survival: Survival Analysis (<http://r-project.org>)) was utilized to construct a multivariate Cox regression model for candidate genes screened by Lasso regression analysis. Based on the following formula, the obtained characteristic genes and Riskscore were subject to cumulative weighting to generate the prognostic model:

$$\text{Riskscore} = \sum_{i=1}^n \exp_i * \beta_i, \quad (1)$$

Where n is the number of genes related to immune characteristics and patient's prognosis, \exp_i is the expression level of each gene, and β_i is the risk coefficient calculated by multivariate Cox analysis.

The risk score of the patients was calculated according to the expression and risk coefficient of each gene in the prognostic model, and the samples were divided into high- and low-risk groups using the median Riskscore. By using R package "survival" (survival: Survival Analysis (<http://r-project.org>)), we plotted the survival curves of the high- and low-risk groups based on the Riskscore. R package "timeROC" plotted the ROC curves [18].

2.5. GSEA. KEGG pathway enrichment analysis was performed on high- and low-risk groups using GSEA software.

2.6. Analysis of Levels of Immune-Regulating Molecules in Tumor. Box plots were prepared by Wilcoxon analysis in each

sample to count the expression of HLA proteins, immune checkpoints PD-L1, PD-1, CDLA4, and CD28.

2.7. Assessment of the Independence of the Prognostic Model.

Univariate and multivariate regression analyses were performed on GSE60052 samples by combining patients' clinical information such as age, gender, stage, pathological stages (T and N), and prognostic Riskscores to draw the corresponding forest plot [19].

3. Results

3.1. Screening and Enrichment Analysis of Immune-Related DEGs. Differentially expression analysis between normal and tumor groups in GEO data set yielded 3,022 DEGs in SCLC, including 1,301 highly expressed genes and 1,721 lowly expressed ones in tumor tissue (Figure 1(a)). Immune-related genes were searched from the ImmPort database. DEGs and immune-related genes were intersected to obtain 228 immune-related DEGs (Figure 1(b), Table S1). GO analysis suggested that these immune-related DEGs were mainly enriched in biological functions of positive regulation of cytokine production, positive regulation of response to external stimulus, and leukocyte migration (Figure 1(c)). KEGG analysis suggested that these DEGs were mainly enriched in PI3K-Akt signaling pathway, MAPK signaling pathway, and human T-cell leukemia virus 1 infection (Figure 1(d)).

3.2. Construction and Evaluation of SCLC Subtypes Based on Immune-Related DEGs. Consensus clustering analysis was conducted on 228 immune-related DEG loci to identify molecular subsets of different immune-related DEGs. Clustering results were visualized using cumulative distribution function (CDF) plots and CDF incremental area plots, where k represents the number of groups (Figures 2(a) and 2(b)). The results of undifferentiated clustering analysis showed that the internal consistency of the clusters was high, and the clustering effect was best when $k=3$ (Figure 2(c)), so the samples could be categorized into three subtypes. Survival curves were plotted by grouping the subtypes and combining clinical information for survival analysis, showing that subtype 3 had a higher survival rate than subtypes 2 and 1 (Figure 2(d)). Among them, the activation levels of immune function-related gene sets of subtype 2 and subtype 3 were higher than those of subtype 1 (Figure 2(e)). Further evaluation of the immune microenvironment of these three subtypes indicated that there were differences in the tumor microenvironment scores among the three subtypes. Of which, the subtype 3 with the best overall survival status had the highest stromal score, immune score, and ESTIMATE score and the lowest tumor purity, while the subtype 1 presented opposite results to that of subtype 3 (Figures 2(f)–2(i)). The above results indicated that the immune-related DEGs selected in this study can distinguish SCLC samples based on immune status and that there were marked differences in the prognosis among the three immune subtypes, indicating the feasibility of constructing prognostic model with these genes.

3.3. Prognostic Modeling and Validation of DEGs Associated with Immune Regulatory Functions in Tumor. Univariate Cox regression analysis was performed on DEGs associated with immune function to obtain 20 candidate genes with prominent association with survival (Table S2). Lasso regression analysis of these 20 candidate genes filtered out high-fit immune function-related DEGs. As depicted in Figure 3(b), the smaller the partial-likelihood deviance value, the more stable the model was. When the $\log(\lambda)$ was 2.5, the model was relatively more stable, and thus, the $\log(\lambda)$ value was considered to be the optimal penalty coefficient. As plotted in Figure 3(a), the regression coefficients varied with $\log(\lambda)$. A multivariate regression model was established for these 7 signature genes associated with prognosis, and the final modeling was completed based on the risk score coefficient of each gene: $\text{Riskscore} = -0.24923 \times \text{CXCL2} - 0.43238 \times \text{ENG} - 0.41932 \times \text{ARRB1} - 0.06037 \times \text{BMP1} - 0.94496 \times \text{IRF1} - 0.33634 \times \text{CCL5} + 0.14949 \times \text{LCP2}$ (Figure 3(c)).

ROC curves for assessing the 1-, 3-, and 5-year overall survival of patients' prognosis were plotted based on the Riskscores obtained from this model, and their AUC values were 0.82, 0.9, and 0.95, respectively (Figure 3(d)). The survival curves were plotted by grouping the samples into high- and low-risk groups according to the median value of the Riskscore, and patients in the high-risk group were found to have lower survival (Figure 3(e)). Combined with Riskscore distribution and survival time, it was found that the increase of Riskscore was associated with the increased number of deaths and the decreased survival time of patients (Figures 3(f) and 3(g)). And there was a difference in the level of the 7 genes (Figure 3(h)). In sum, our work generated a 7-signature gene risk assessment model to predict the prognosis of SCLC patients, with good predictive ability.

3.4. KEGG Analysis of High- and Low-Risk Groups. KEGG analysis of both groups using GSEA software revealed that the low-risk group was markedly enriched in immune-related pathways, including chemokine signaling pathway, FC gamma R mediated phagocytosis, nod-like receptor signaling pathway, T cell receptor signaling pathway, systemic lupus erythematosus, and B cell receptor signaling pathway (Figures 4(a)–4(f)). Most of the above pathways belong to cytokine-related pathways, which are closely related to tumor immunity and antitumor killing effect, supporting the difference in immune pattern between groups, and also implying that the difference in prognoses of patients in high- and low-risk groups in this study may be associated with these signaling pathways, but it still needs subsequent validation.

3.5. Analysis of Immune Characteristics in Tumor Based on the Risk Assessment Model. Given that the GSEA elucidated the differences in immune signaling-related pathways between the high- and low-risk groups, we worked to further elucidate the differences in immune patterns between groups, HLA expression in the high- and low-risk groups. The results indicated significantly higher expression of HLA series antigen molecules in the low-risk group (Figure 5(a)). Immune-related checkpoint molecules or targeted inhibitory molecules of immunotherapy, PD-L1 (Figure 5(b)), CTLA4 (Figure 5(d)),

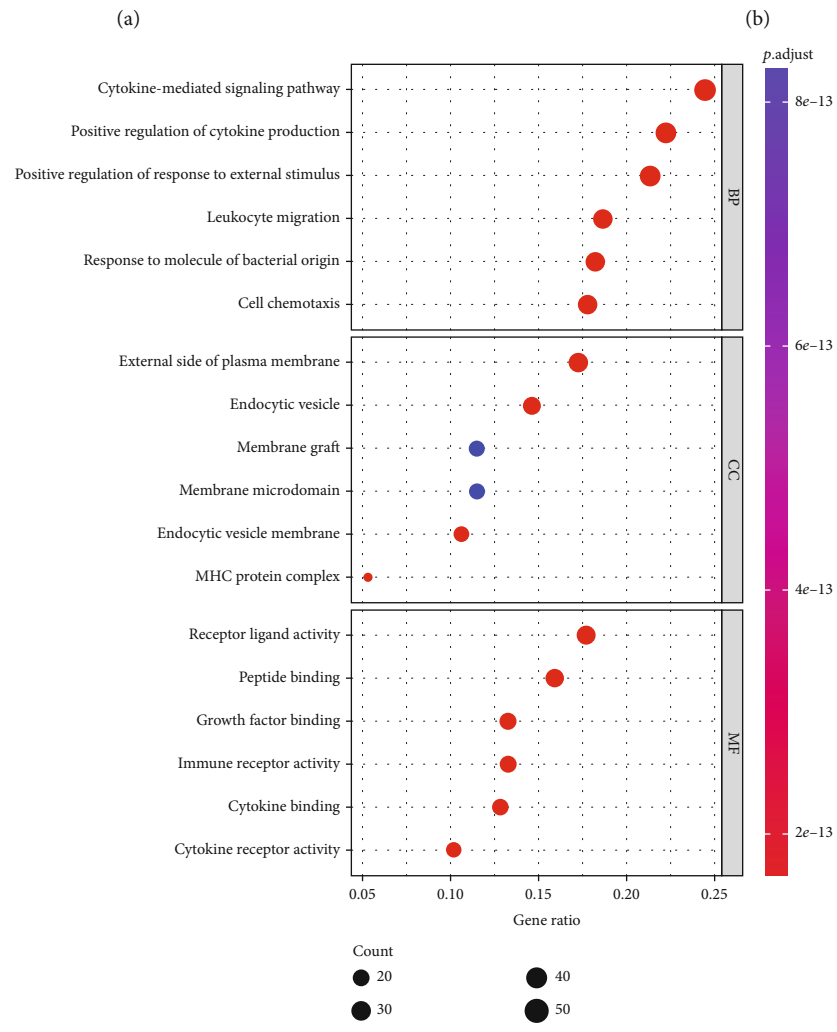
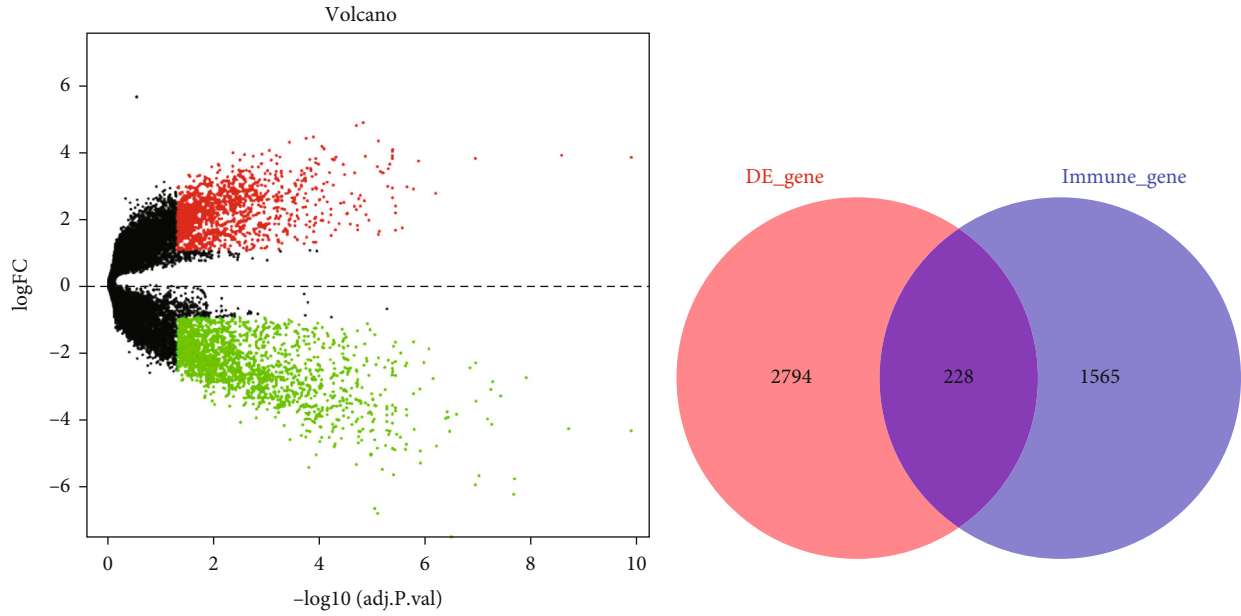
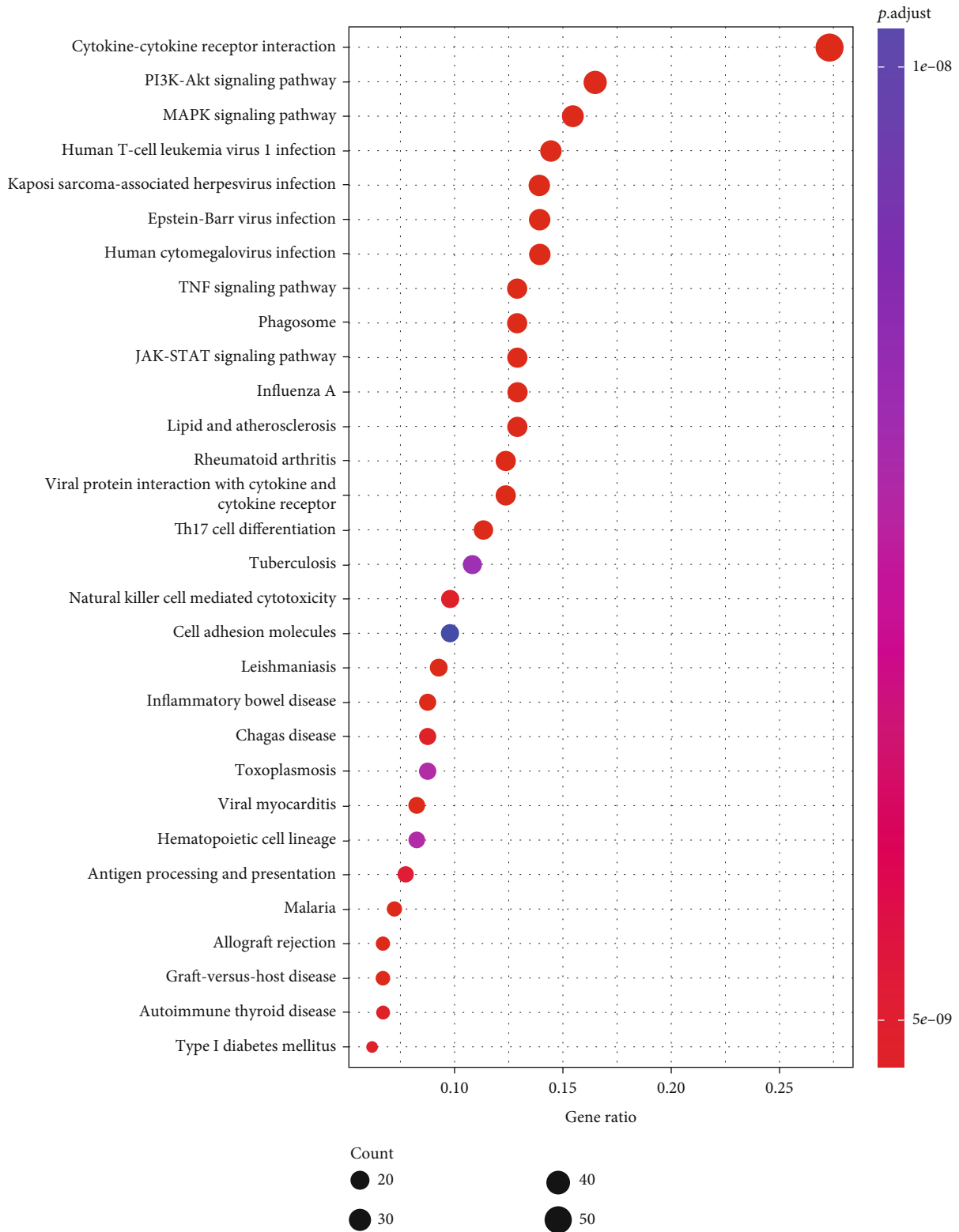
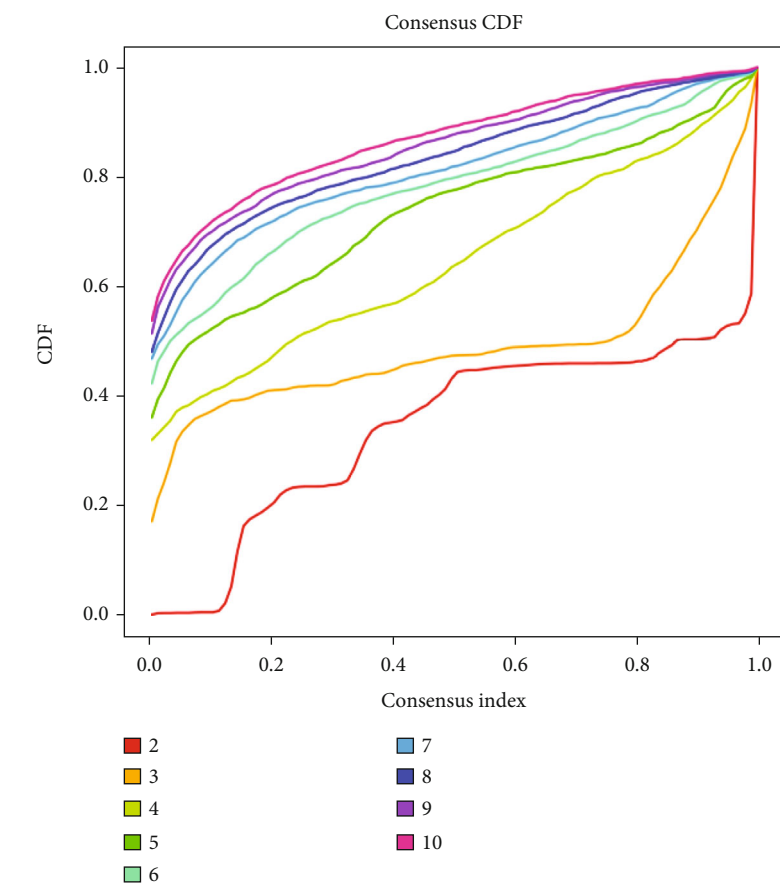


FIGURE 1: Continued.



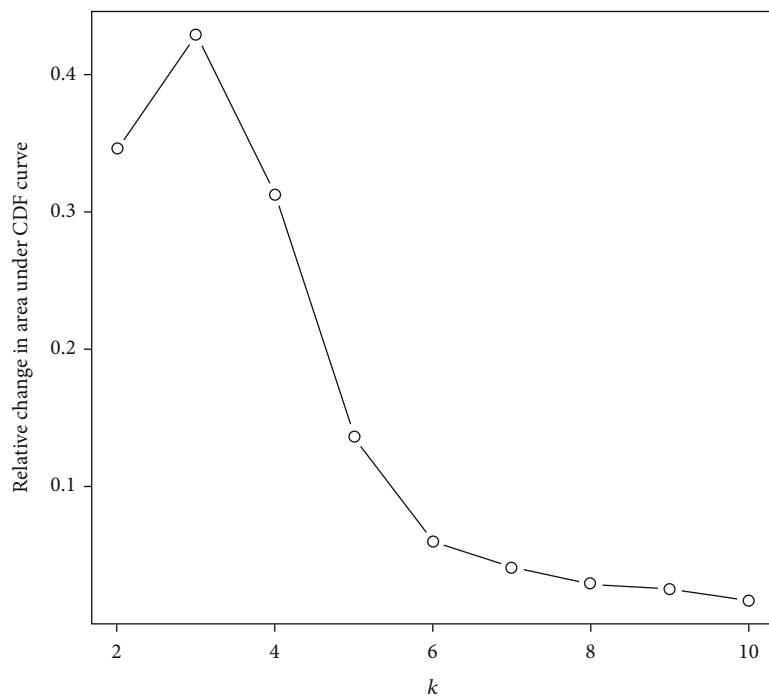
(d)

FIGURE 1: The screening and functional enrichment of immune-related DEGs. (a) Volcano plot of DEGs in SCLC samples and healthy samples in GEO dataset. Red represents significantly upregulated genes and green represents significantly downregulated genes. (b) Venn diagram of DEGs and immune-related genes. (c) GO and (d) KEGG analyses of immune-related DEGs.



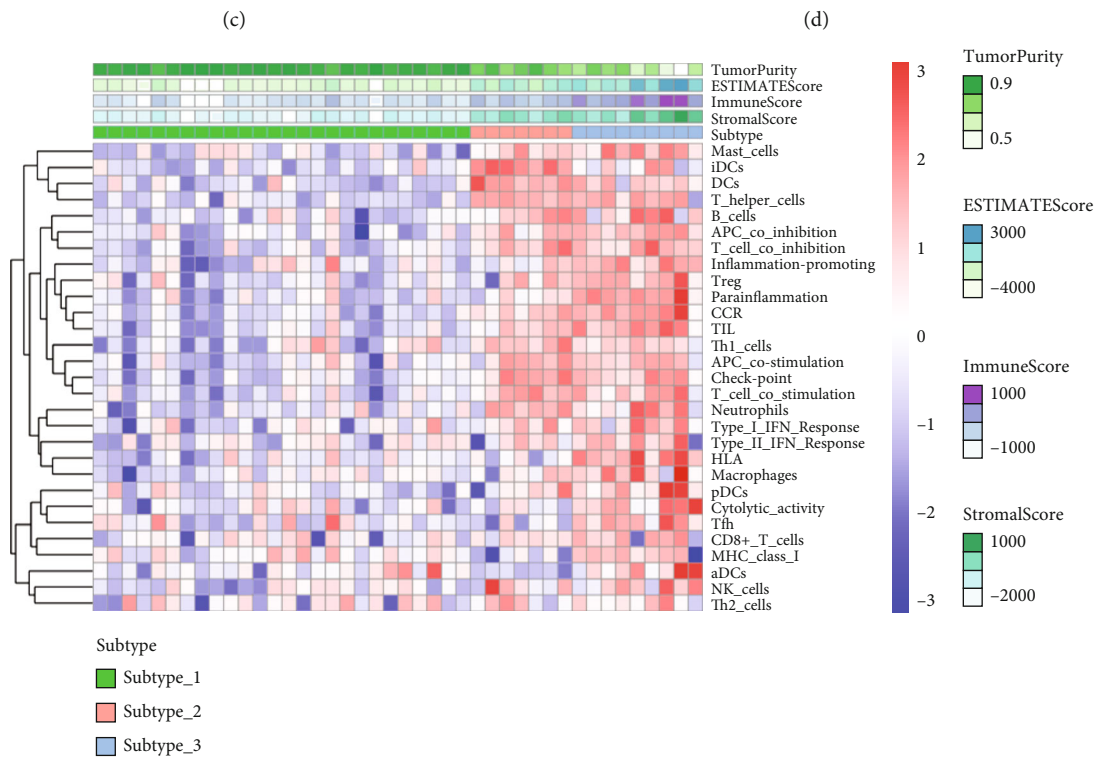
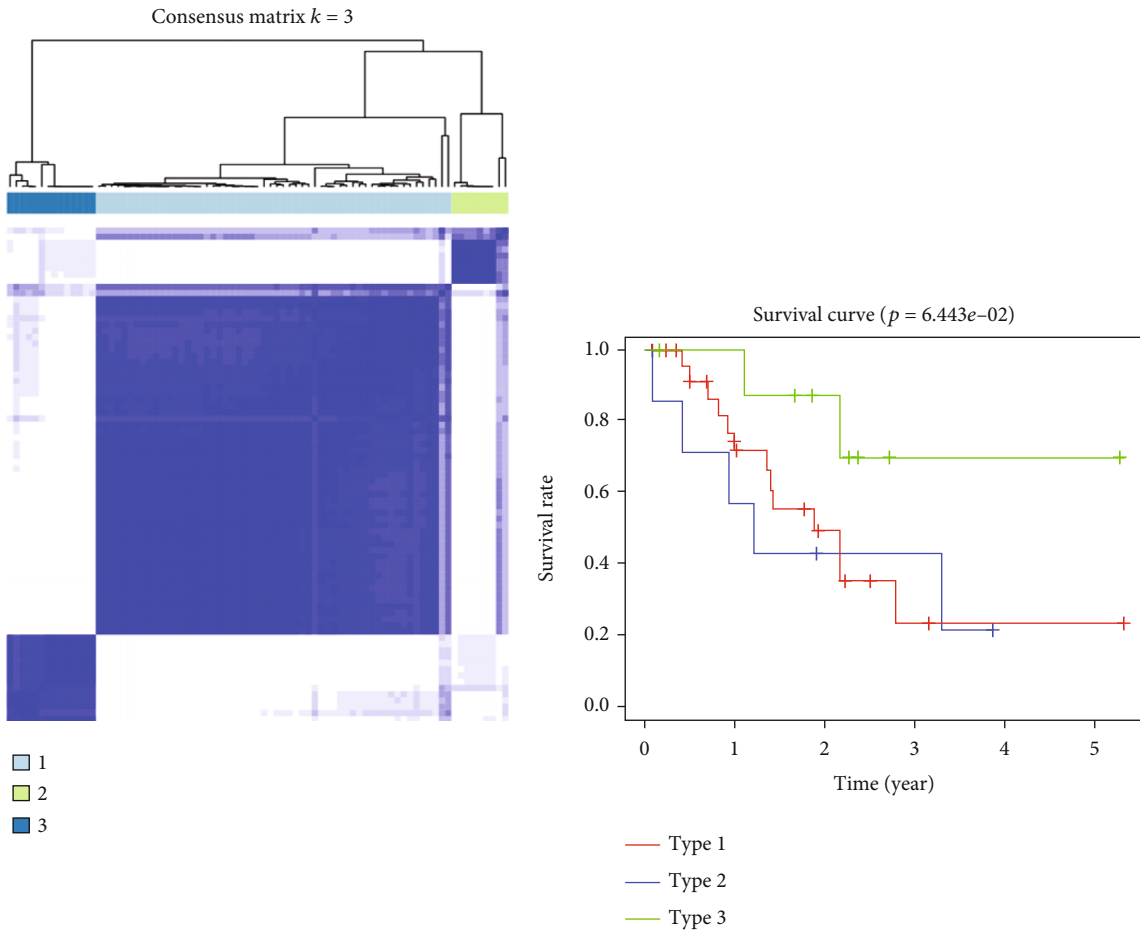
(a)

Delta area



(b)

FIGURE 2: Continued.



(e)

FIGURE 2: Continued.

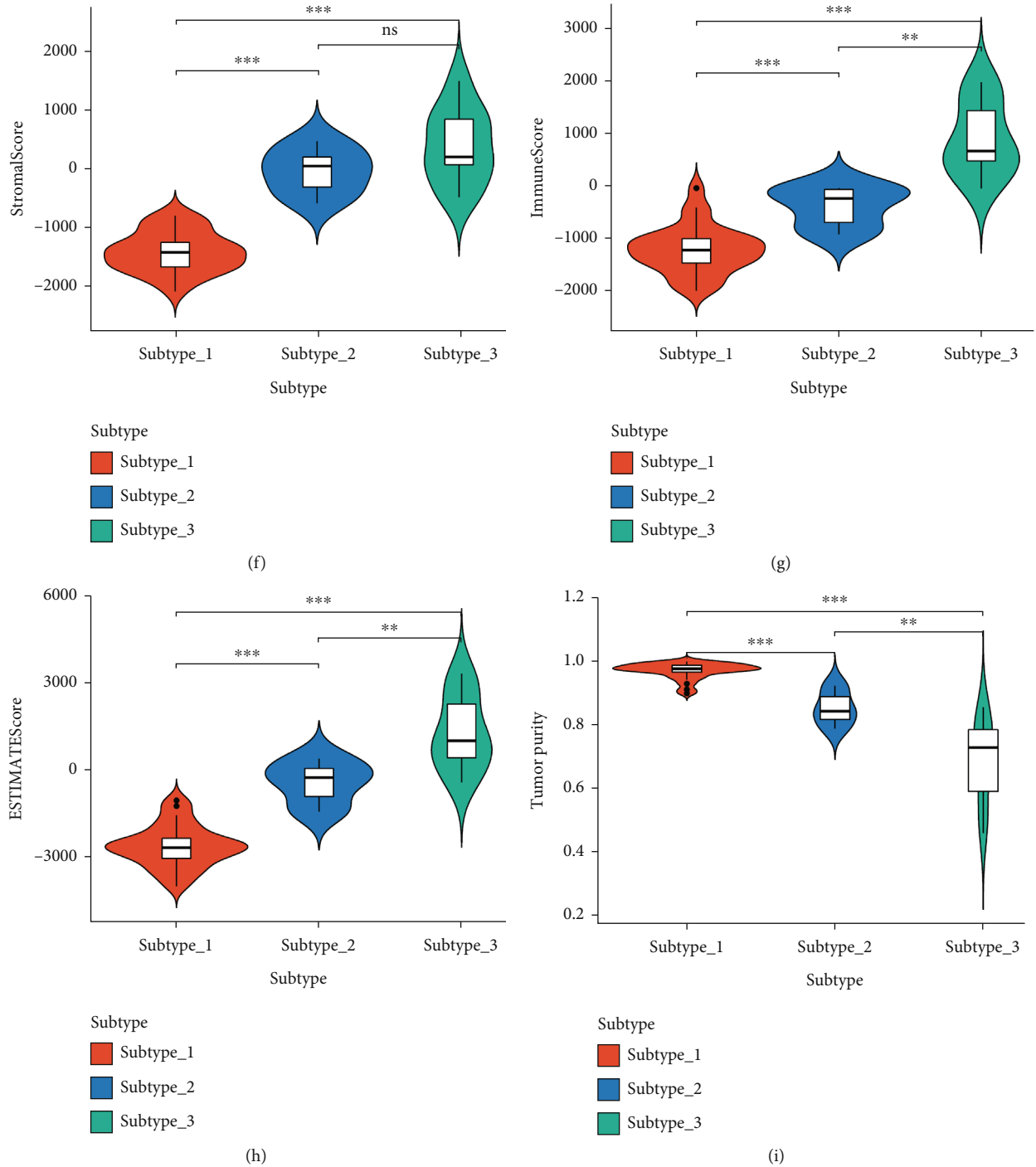
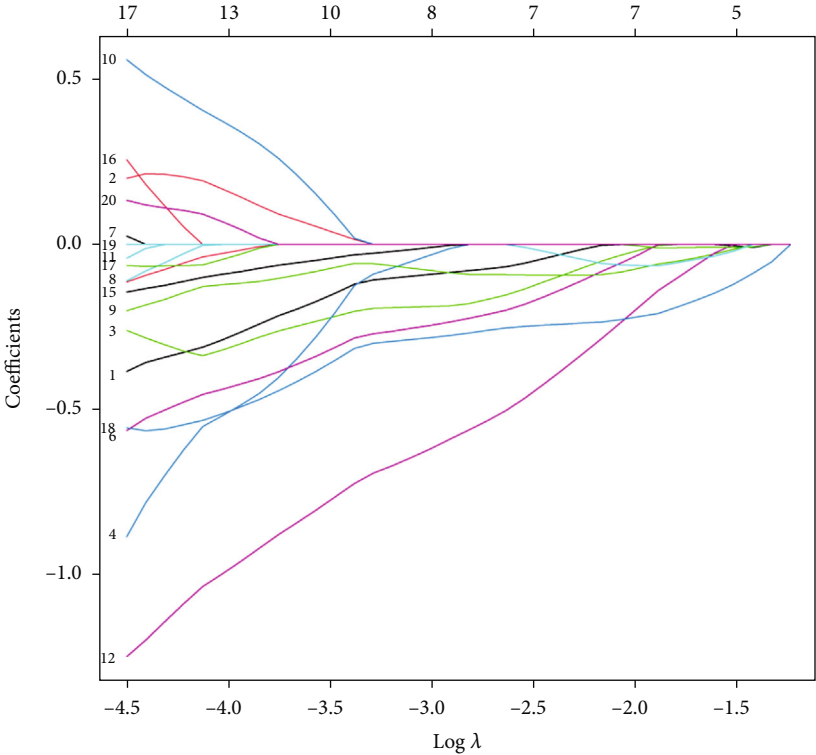


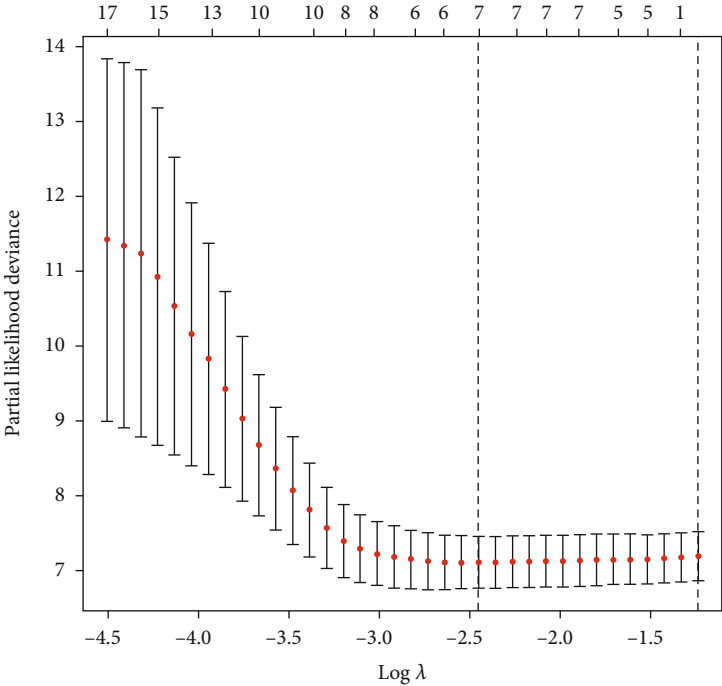
FIGURE 2: Construction and assessment of SCLC subtypes based on immune-related DEGs. (a) Diagram of consensus cumulative distribution function (CDF). (b) Plot of the relative area under the CDF curve. (c) K-means cluster analysis of SCLC samples. (d) Survival analysis of different immune subtypes. (e) Heatmap of tumor microenvironment in different subtypes. (f) Stromal score (red = subtype-1, blue = subtype-2, and green = subtype-3), (g) immune scores (red = subtype-1, blue = subtype-2, and green = subtype-3), (h) ESTIMATE scores (red = subtype-1, blue = subtype-2, and green = subtype-3), and (i) Tumor purity scores for different subtypes (red = subtype-1, blue = subtype-2, and green = subtype-3). ns: not significant; ** $P < 0.01$; *** $P < 0.001$.

and CD28 (Figure 5(e)) manifested statistically significant differences of high levels in the low-risk group. While the targeted inhibitory molecule PD-1 showed no significant difference in the low-risk group (Figure 5(c)). The above models validated

the differences in immune patterns between high- and low-risk groups, and the differences in patient outcomes between both groups may be driven by differences in these immune proteins or checkpoints.

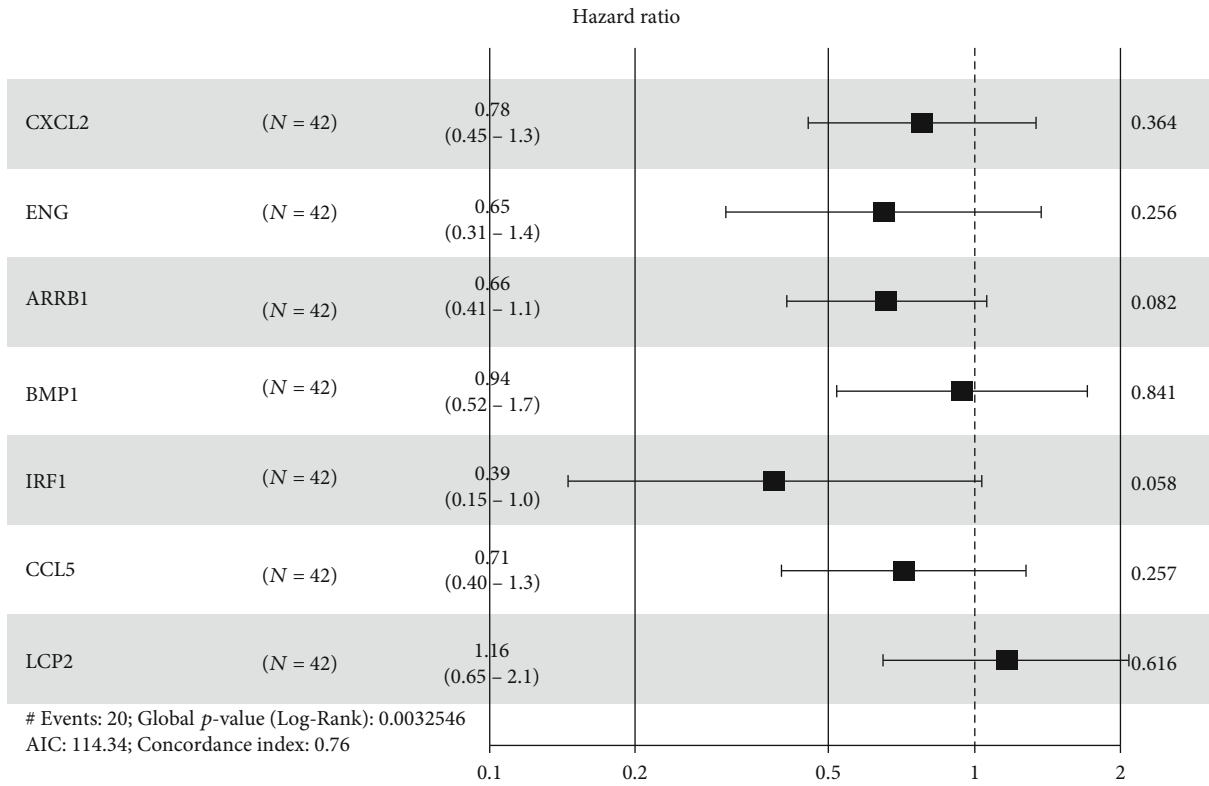


(a)

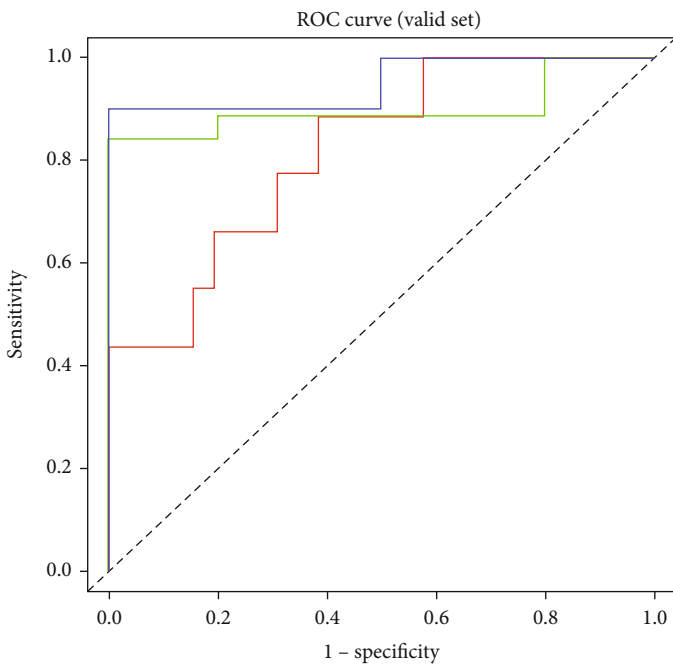


(b)

FIGURE 3: Continued.

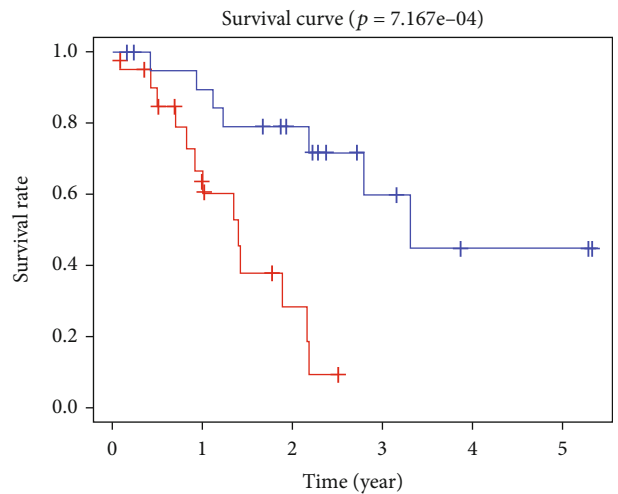


(c)



— AUC at 1 year: 0.82
— AUC at 3 years: 0.9
— AUC at 5 years: 0.95

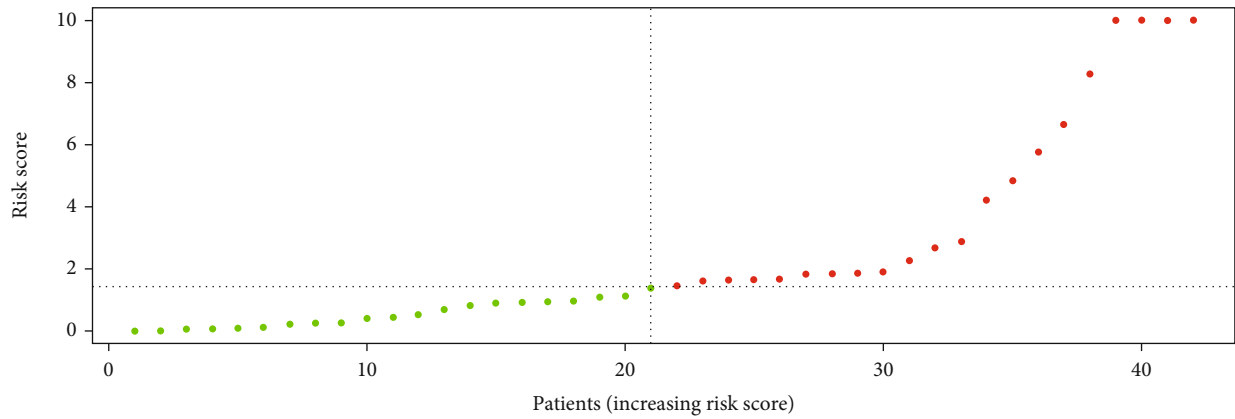
(d)



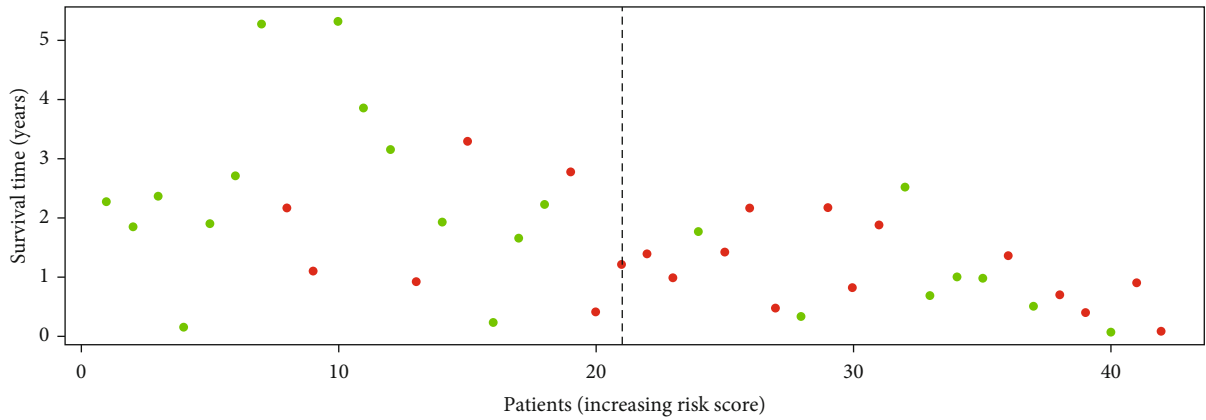
— High
— Low

(e)

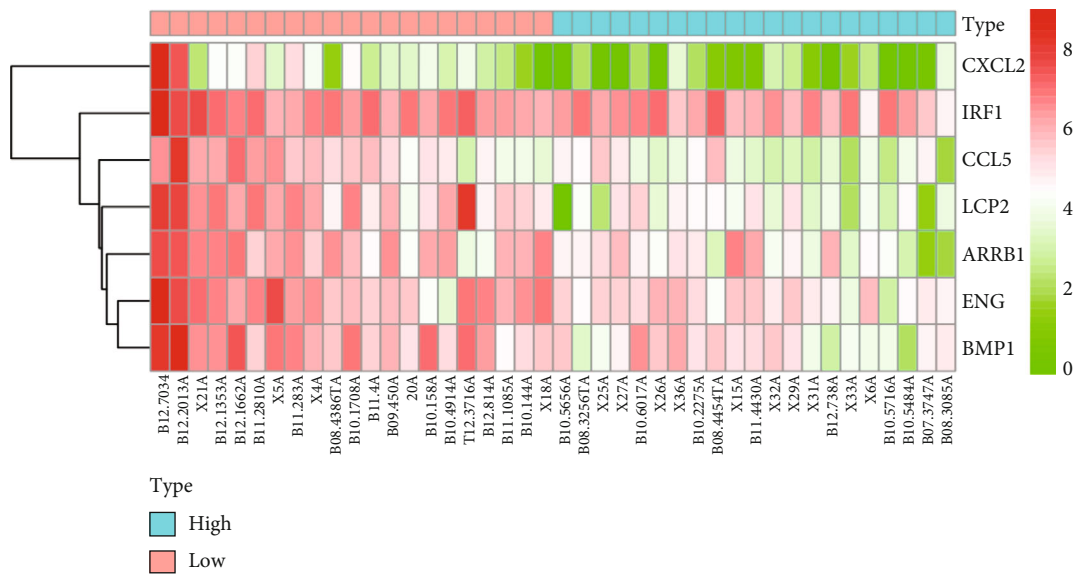
FIGURE 3: Continued.



(f)

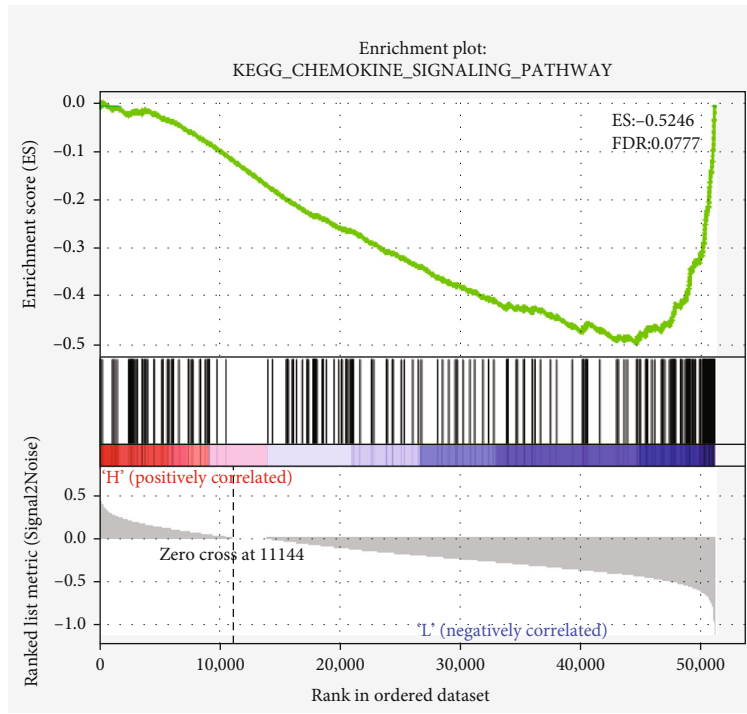


(g)



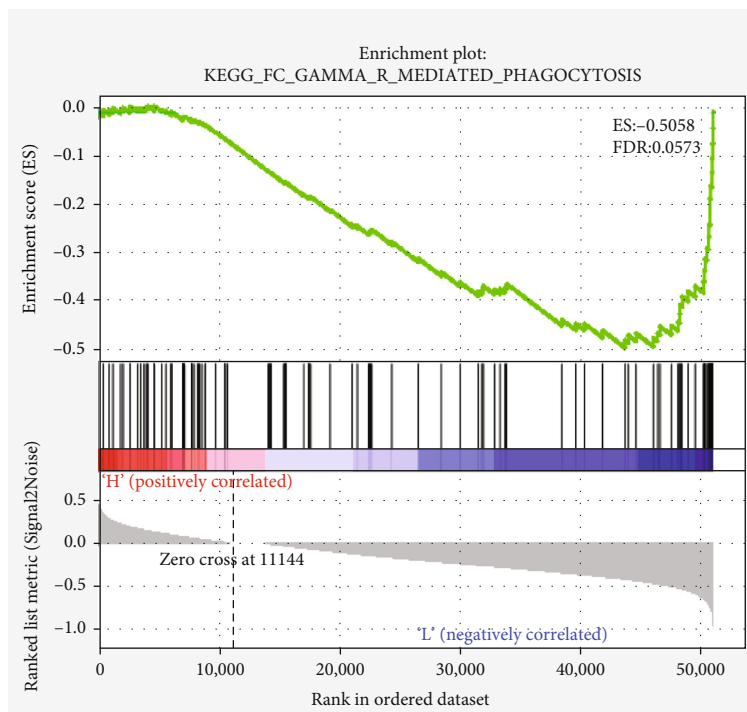
(h)

FIGURE 3: Construction of a prognostic risk assessment model for SCLC patient. (a) Trajectory plot of the gene coefficients with the increasing $\log(\lambda)$ value in the Lasso model. (b) The optimal $\log(\lambda)$ was selected according to the partial likelihood deviance in the Lasso model. (c) Forest plot of 7 genes obtained from multivariate Cox regression analysis. (d) ROC curves of the GSE60052 data sample risk score in both groups. (e) Survival analysis of GSE60052 dataset by the Riskscore model. (f) Distribution of scores, (g) survival status, and (h) the heatmap of 7 genes in both groups.



— Enrichment profile
— Hits
— Ranking metric scores

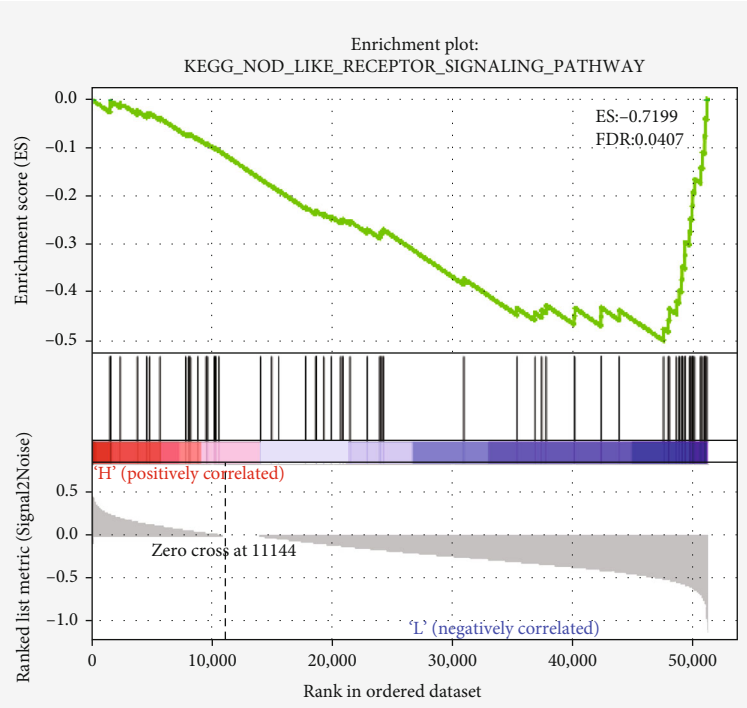
(a)



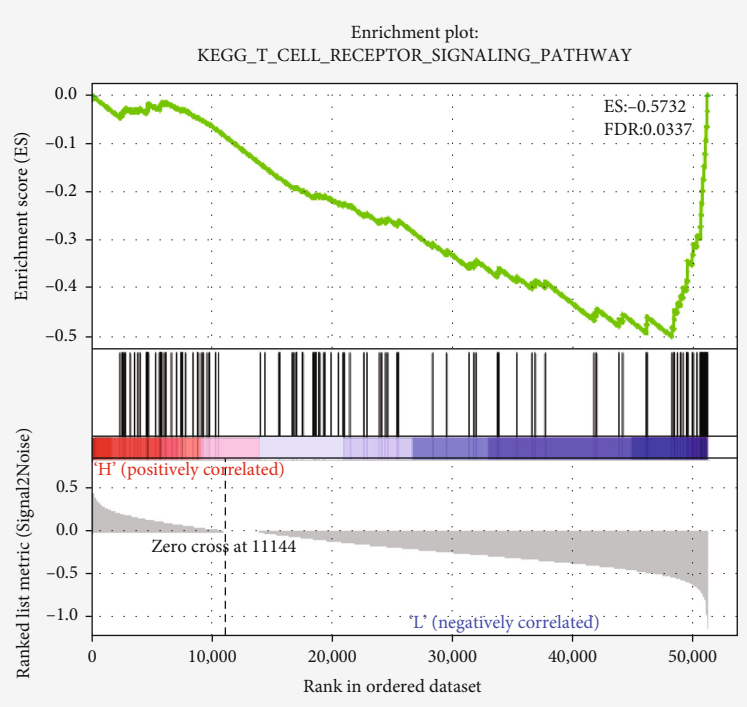
— Enrichment profile
— Hits
— Ranking metric scores

(b)

FIGURE 4: Continued.

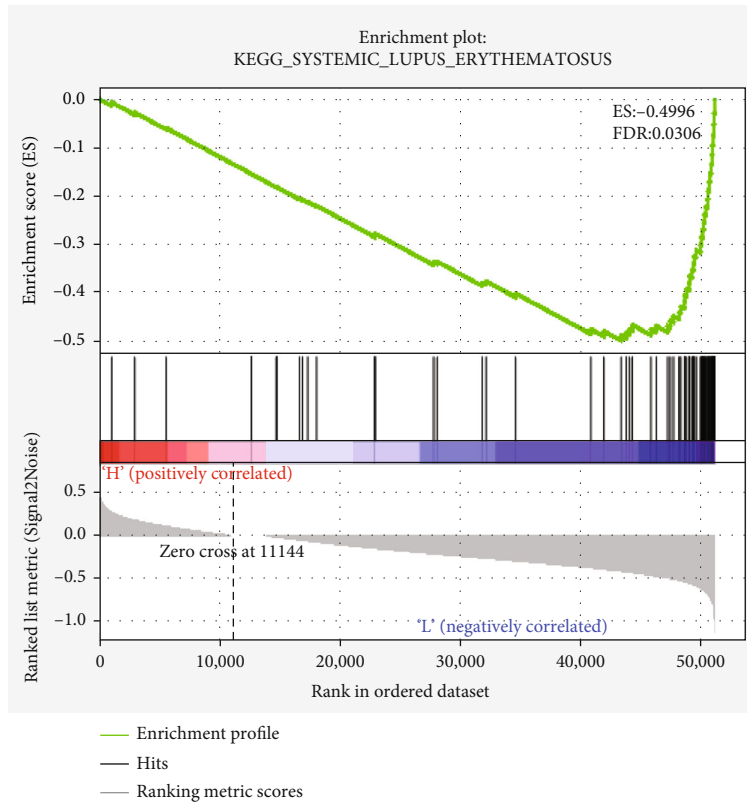


(c)

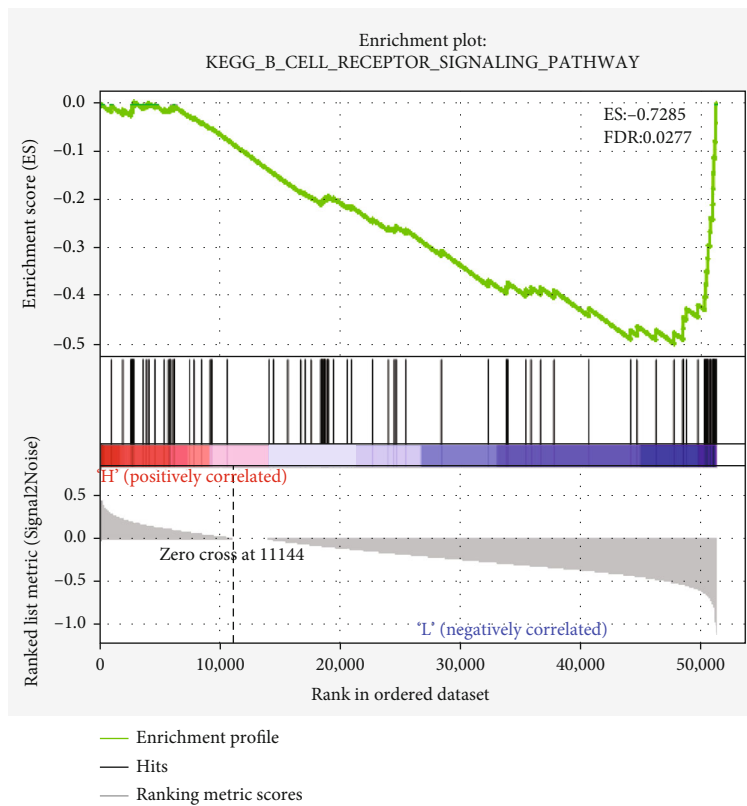


(d)

FIGURE 4: Continued.

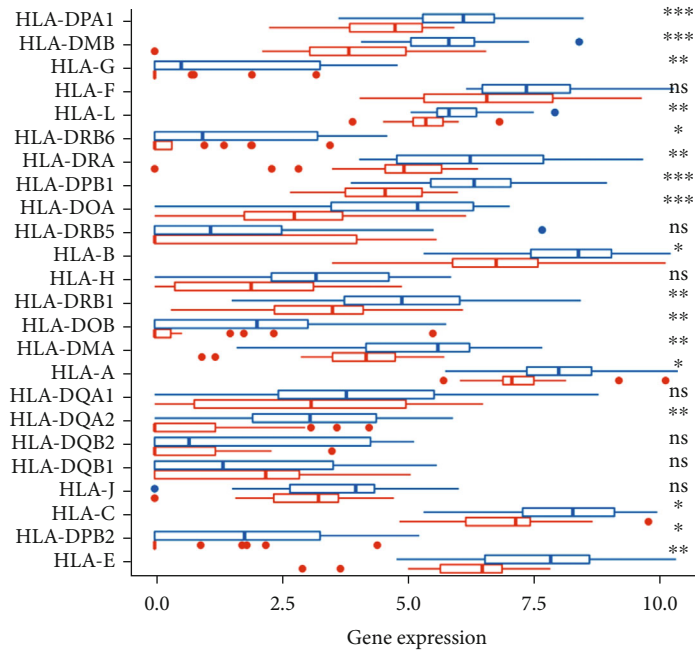


(e)



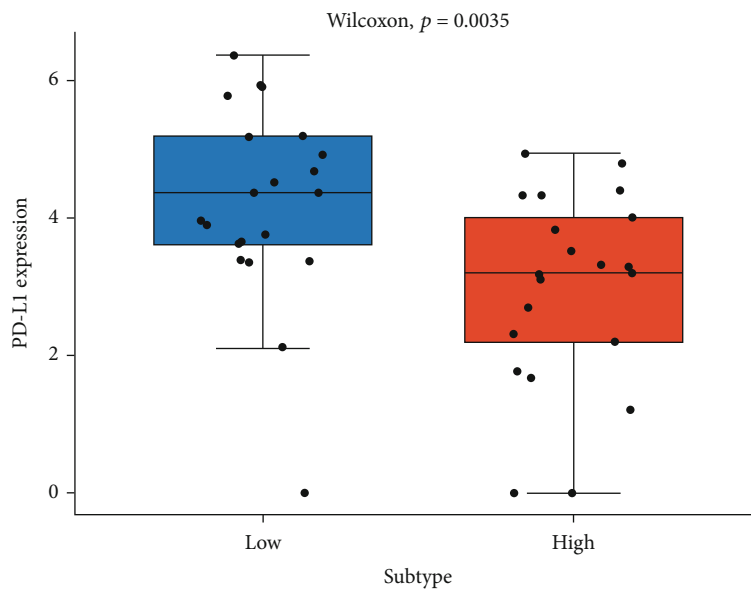
(f)

FIGURE 4: GSEA for high- and low-risk groups.



Subtype
▢ High
▢ Low

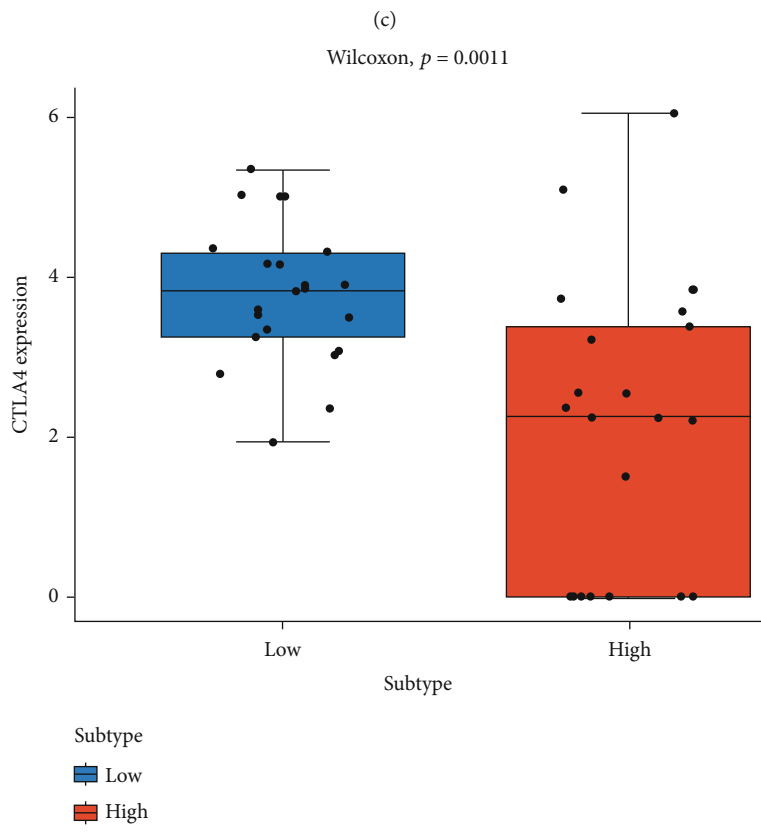
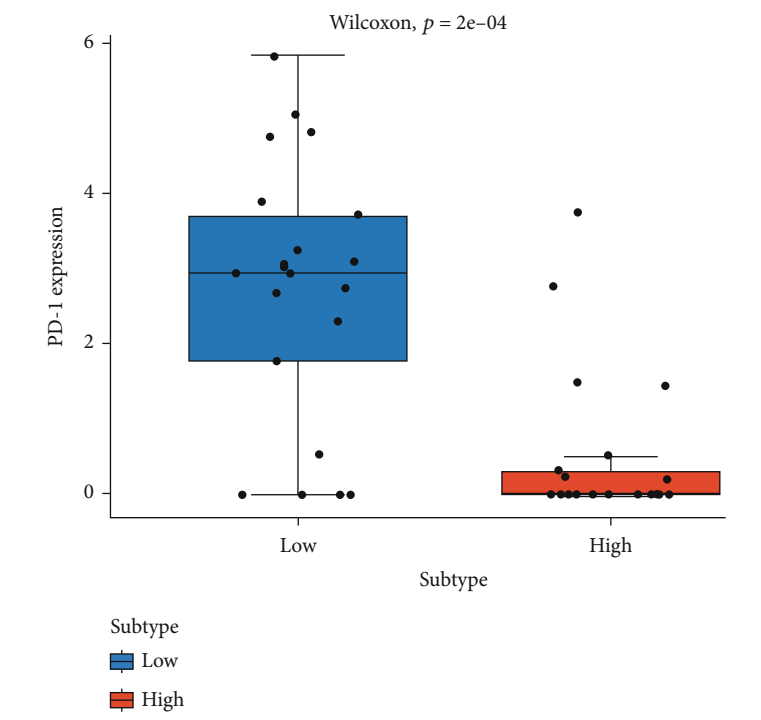
(a)



Subtype
▢ Low
▢ High

(b)

FIGURE 5: Continued.



(d)

FIGURE 5: Continued.

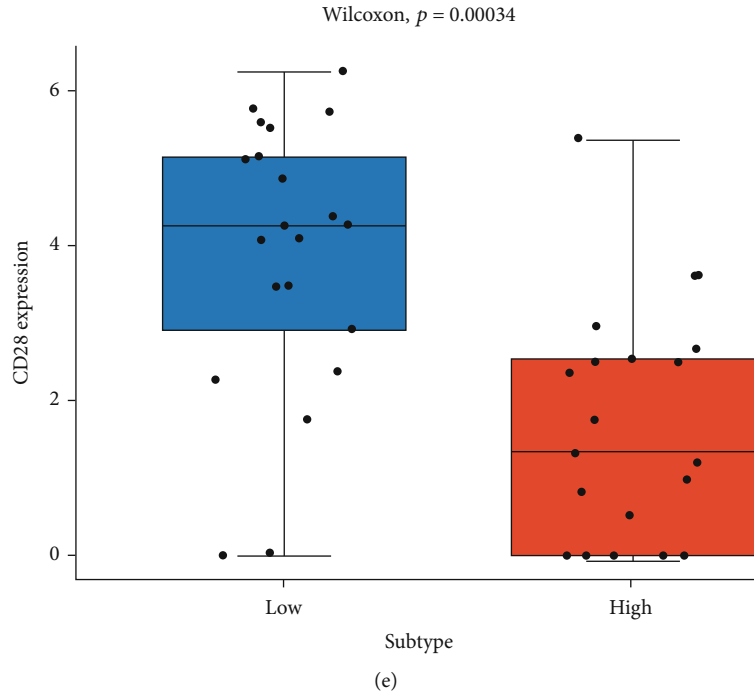


FIGURE 5: Analysis of tumor immune characteristics in high- and low-risk groups. Box plots of (a) HLA, (b) PD-L1, (c) PD-1, (d) CTLA4, and (e) CD28 in high- (red) and low-risk (blue) groups. ns: not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

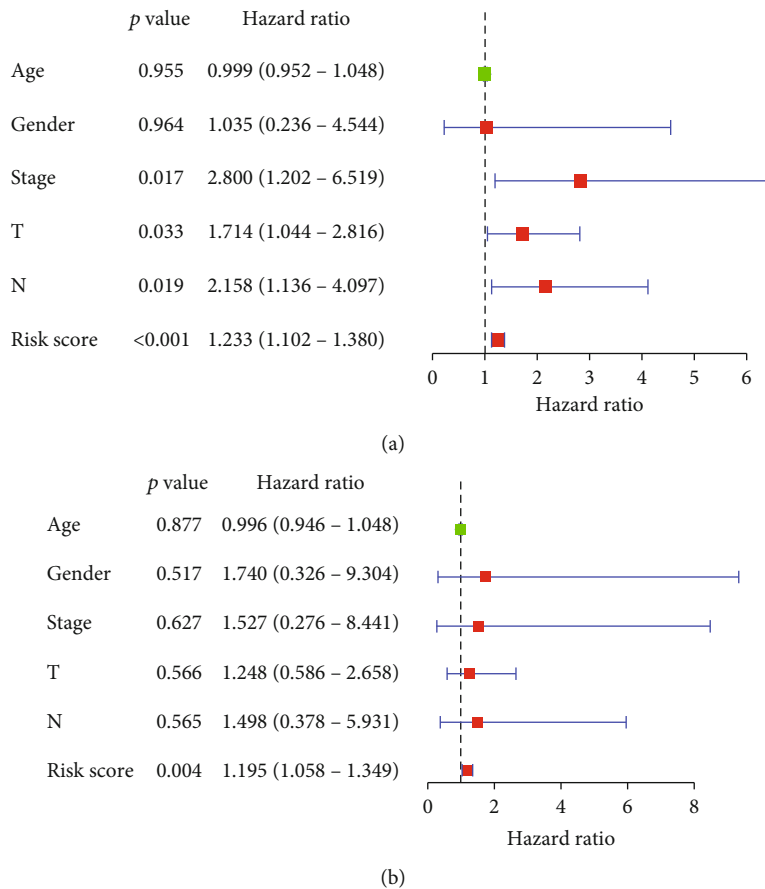


FIGURE 6: Correlation analysis between Riskscore and clinical factors in SCLC patients. (a) Forest plot of univariate and (b) multivariate Cox regression analyses based on clinical information, Riskscore, and overall survival.

3.6. Evaluation of Clinical Value of the Prognostic Model. Univariate regression analysis on Riskscore and clinical information revealed significance between tumor clinical stages, pathological stages (T and N) and Riskscore (p value < 0.05) (Figure 6(a)). Subsequent multivariate regression analysis revealed that significance only lay in Riskscore (Figure 6(b)). The prognostic score model constructed from these 7 signature genes could therefore be used as an independent prognostic factor.

4. Discussion

Tumor immune-related genes are important for revealing the prognosis of LC patients. CD133 is a key biomarker closely related to the prognosis of patients, and the use of this marker for tumor screening in SCLC can detect the presence of tumors earlier and reduce the risk of disease in patients [20]. At present, the use of bioinformatics approach to mine public databases and select effective therapeutic targets or biomarkers is currently the most advanced research method. As proposed by Wu et al. [21], a TME risk model constructed with the tumor immune infiltration-related genes SERPINE1, CX3CR1, CD200R1, GBP1, IRF1, STAP1, LOX, and OR7E47P based on public databases can be used to predict the prognostic survival of LC patients. Our study focused on exploring the correlation between SCLC immune-related genes and prognosis while distinguishing SCLC immune subtypes, revealing differences in immune-related molecular indicators between these subtypes, and validating the effectiveness of related models. Most importantly, this study constructed a risk assessment model based on immune-related DEGs that can be used to assess the prognoses of SCLC patients, which provided a basis for SCLC diagnosis and treatment.

In this study, we identified 7 genes that can be used to construct a prognostic model for SCLC patients via bioinformatics approach, of which CXCL2, ENG, ARRB1, BMP1, IRF1, and CCL5 were prognostic protective factors and LCP2 was the only prognostic risk factor. Among them, CXCL2 is a traditional inflammatory chemokine, which can mediate the recruitment of neutrophils to lung tissue [22]. This may provide a reference for CXCL2 as a prognostic risk factor in this paper; the recruitment of neutrophils by CXCL2 may facilitate antitumor immune responses. ENG (CD105) is a coreceptor of TGF- β , which is essential for angiogenesis/vascular development. The expression of this gene can impel angiogenesis in tumor tissue and further cause malignant progression of tumors, which has been always considered as an excellent therapeutic target [23]. There was a mice experiment indicating that anti-ENG monoclonal antibody in treatment can suppress tumor progression [24]. The possible effect of ARRB1 on tumor development is still controversial, and ARRB1 is associated with the prognosis of patients with tumors in a variety of cancers [25–27]. In our study, data models suggested that ARRB1 may improve the prognosis of patients with SCLC. BMP1 is considered to be a key factor in promoting tumor growth and metastasis of LC, and BMP1 facilitates NSCLC metastasis by inhibiting TGF- β activity in NSCLC [28]. Interestingly, it was found in the present study that BMP1 may improve the prognoses of SCLC patients, and the reasons for this need to be further elucidated. The expression of IRF1 in

NSCLC tissue was generally lower than that in normal lung tissue, which is cancer-suppressive by regulating KPNA2 [29]. We found that IRF1 could be used as a prognostic protective factor in SCLC, consistent with the previous study. CCL5 acts as a chemokine ligand, which is supposed to propel the cytotoxicity of tissue-resident T and NK cells and strengthen antitumor immune responses [30]. LCP2, the only prognostic risk factor among the 7 genes of our prognostic model, is thought to be related to the infiltrating level of toxic lymphocytes and plays a regulatory role in antitumor immunity. Unfortunately, its role in SCLC remains unknown [31, 32]. In addition, in the study of LC, it has been exhibited that this gene is positively correlated with PD-L1 level in lung adenocarcinoma tissue, and we speculated that it may further affect the disease progression of SCLC by affecting the immune escape mechanism in cancer [33]. In short, the prognosis-related marker molecules associated with immune regulatory function obtained in this paper were generally involved in the regulation of tumor growth and immunity. These molecules can serve as biomarkers to predict the prognosis of patients and potential targets for SCLC treatment.

After establishing a prognostic model, we divided SCLC samples into high- and low-risk groups based on the median value of the risk score of the risk assessment model, and the results implied that the HLAs, especially HLA-DPA1, HLA-DPB1, HLA-DMB, and HLA-DOA, were remarkably upregulated in the low-risk group. HLA is the expression product of the major histocompatibility complex (MHC) in humans, and the HLA system is the most complex polymorphic system in the human body known so far. Given that this family is often dysregulated in the tissue of patients receiving immunotherapy, recent studies put their focus on the immune-regulatory effects of this family [34]. To take an example, HLA-DPA1 and HLA-DPB1, dysregulated in our low-risk group, have been manifested to be associated with the maintenance of long-term immune efficacy after HBV vaccination [35]. In the field of cancer, HLA-DPA1 activates chemokines and toll-like receptor signaling pathways to regulate hepatocellular carcinoma progression [36]. HLA-DPB1 can be an antitumor factor to recruit NK cells, CD8⁺ T cells, and tumor-infiltrating lymphocytes such as Th1 and Tfh into breast cancer [37]. Both HLA-DPA1 and HLA-DPB1 play an antitumor role by activating immune cell infiltration, which well explained their downregulation in the high-risk group of this study. HLA-DMB is an essential component of MHC complex synthesis, and the expression of this gene is prominently positively correlated with the level of infiltration of tumor-infiltrating CD8⁺ T cells [38]. Similarly, HLA-DOA regulates the level of B cell infiltration in tissue and ensures the stable expression of MHC in cancer tissue, balancing their biological functions [39]. It can be seen that, similar to HLA-DPA1 and HLA-DPB1, HLA-DMB, and HLA-DOA are also associated with the upregulation of tumor-infiltrating immune cells, consistent with the results predicted by our study. The immune-regulatory effect of HLA family validated the effectiveness of the construction based on the prognostic model of immune-related genes and also confirmed the rationality of this risk assessment model in predicting tumor immune patterns. In addition to leukocyte antigens, we observed notable

differential expression of immune checkpoint genes PD-L1, CTLA4, and CD28 in high- and low-risk groups. Activated T cells often express PD-1 on their surface and can act as immune checkpoint receptors, while PD-L1 produced on the surface of many cancer cells acts as a PD-1 ligand and the combination of the two leads to tumor immunosuppression [40]. CTLA4 is an immune checkpoint protein expressed on activated T cells to downregulate the activation of T cells [41]. CD28 is a cell surface glycoprotein receptor expressed primarily on activated T cells and belongs to the immunoglobulin (Ig) superfamily [42]. This molecule has been manifested to negatively regulate T cell antitumor responses and is widely involved in tumor immune escape [42].

In summary, in this study, based on bioinformatics analysis, immune-related DEGs in the GSE60052 dataset were screened and classified into three subtypes that could represent different immune patterns to assess the rationality of immune-related genes. At the same time, the 7-gene prognostic model established by bioinformatics analysis based on immune-related DEGs could evaluate the prognoses of SCLC patients more accurately. The GSE60052 dataset was divided into high- and low-risk groups according to the median Riskscore, revealing the differences in the expression of immune checkpoint genes and antigen molecules, and confirming the rationality of the model in predicting tumor immune patterns. This study is conducive to deepening the understanding of immune-related genes in SCLC while providing a powerful tool for prognostic evaluation and immunotherapy of SCLC patients. Of course, this study has some limitations; the data of this study were derived from open databases, there were certain systematic errors, and the accuracy of model prediction needs to be subsequently verified in more clinical samples. Additionally, wet experiments were not conducted to validate the constructed model. Relevant cellular experiments and molecular experiments are therefore warranted to verify the model.

Data Availability

All data generated or analyzed during this study are included in this article. Further enquiries can be directed to the corresponding author.

Ethical Approval

No animal or human cell was used.

Conflicts of Interest

The authors report no conflict of interest.

Authors' Contributions

Feng Deng was assigned in conception and design. Feng Tao and Zhili Xu were assigned to the collection and assembly of data. Jun Zhou contributed on data analysis and interpretation. Feng Deng worked on manuscript drafting. Ruhu Zhang and Xiaowei Gong were assigned in reviewing and editing. All authors approved the final manuscript.

Acknowledgments

This research received the grant from National Science Foundation of China (No. 8888888).

Supplementary Materials

Supplementary 1. Table S1 Screening of 228 immune-related differential genes.

Supplementary 2. Table S2 Univariate Cox regression analysis of immune-related DEGs.

References

- [1] C. M. Rudin, E. Brambilla, C. Faivre-Finn, and J. Sage, "Small-cell lung cancer," *Nature Reviews. Disease Primers*, vol. 7, no. 1, p. 3, 2021.
- [2] E. B. Bernhardt and S. I. Jalal, "Small cell lung cancer," *Cancer Treatment and Research*, vol. 170, pp. 301–322, 2016.
- [3] K. N. Chen, "Small cell lung cancer and TNM staging," *Zhongguo Fei Ai Za Zhi*, vol. 19, no. 6, pp. 409–412, 2016.
- [4] Y. Wang, S. Zou, Z. Zhao, P. Liu, C. Ke, and S. Xu, "New insights into small-cell lung cancer development and therapy," *Cell Biology International*, vol. 44, no. 8, pp. 1564–1576, 2020.
- [5] H. Mamdani, R. Induru, and S. I. Jalal, "Novel therapies in small cell lung cancer," *Translational Lung Cancer Research*, vol. 4, no. 5, pp. 533–544, 2015.
- [6] R. Roskoski Jr., "Properties of FDA-approved small molecule protein kinase inhibitors: a 2022 update," *Pharmacological Research*, vol. 175, article 106037, 2022.
- [7] O. Altundag, K. Altundag, C. Boruban, Y. S. Silay, and S. Turen, "Imatinib mesylate lacks activity in small cell lung carcinoma expressing c-kit protein: a phase II clinical trial," *Cancer*, vol. 104, no. 9, pp. 2033–2034, 2005.
- [8] D. Hanahan and L. M. Coussens, "Accessories to the crime: functions of cells recruited to the tumor microenvironment," *Cancer Cell*, vol. 21, no. 3, pp. 309–322, 2012.
- [9] W. H. Fridman, F. Pages, C. Sautès-Fridman, and J. Galon, "The immune contexture in human tumours: impact on clinical outcome," *Nature Reviews. Cancer*, vol. 12, no. 4, pp. 298–306, 2012.
- [10] Z. B. Lin, P. Long, Z. Zhao et al., "Long noncoding RNA KCNQ1OT1 is a prognostic biomarker and mediates CD8(+) T cell exhaustion by regulating CD155 expression in colorectal cancer," *International Journal of Biological Sciences*, vol. 17, no. 7, pp. 1757–1768, 2021.
- [11] Z. Chen, Z. Chen, S. Xu, and Q. Zhang, "LncRNA SOX2-OT/miR-30d-5p/PDK1 regulates PD-L1 checkpoint through the mTOR signaling pathway to promote non-small cell lung cancer progression and immune escape," *Frontiers in Genetics*, vol. 12, article 674856, 2021.
- [12] L. M. Sholl, "Biomarkers of response to checkpoint inhibitors beyond PD-L1 in lung cancer," *Modern Pathology*, vol. 35, no. S1, pp. 66–74, 2022.
- [13] M. E. Ritchie, B. Phipson, D. Wu et al., "Limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, article e47, 2015.
- [14] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, "clusterProfiler: an R package for comparing biological themes among gene clusters," *OMICS*, vol. 16, no. 5, pp. 284–287, 2012.

- [15] M. D. Wilkerson and D. N. Hayes, "ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking," *Bioinformatics*, vol. 26, no. 12, pp. 1572–1573, 2010.
- [16] S. Hanzelmann, R. Castelo, and J. Guinney, "GSVA: gene set variation analysis for microarray and RNA-seq data," *BMC Bioinformatics*, vol. 14, no. 1, p. 7, 2013.
- [17] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
- [18] P. Blanche, J. F. Dartigues, and H. Jacqmin-Gadda, "Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks," *Statistics in Medicine*, vol. 32, no. 30, pp. 5381–5397, 2013.
- [19] C. Huang, Z. Liu, L. Xiao et al., "Clinical significance of serum CA125, CA19-9, CA72-4, and fibrinogen-to-lymphocyte ratio in gastric cancer with peritoneal dissemination," *Frontiers in Oncology*, vol. 9, p. 1159, 2019.
- [20] A. G. Kunihiro, S. M. Sarrett, K. J. Lastwika et al., "CD133 as a biomarker for an autoantibody-to-immunoPET paradigm for the early detection of small cell lung cancer," *Journal of Nuclear Medicine*, 2022.
- [21] J. Wu, L. Li, H. Zhang et al., "A risk model developed based on tumor microenvironment predicts overall survival and associates with tumor immunity of patients with lung adenocarcinoma," *Oncogene*, vol. 40, no. 26, pp. 4413–4424, 2021.
- [22] S. Liu, J. Liu, X. Yang et al., "Cis-acting Inc-Cxcl2 restrains neutrophil-mediated lung inflammation by inhibiting epithelial cell CXCL2 expression in virus infection," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 41, 2021.
- [23] C. Ollauri-Ibanez, B. Ayuso-Inigo, and M. Pericacho, "Hot and cold tumors: is endoglin (CD105) a potential target for vessel normalization?," *Cancers (Basel)*, vol. 13, no. 7, p. 1552, 2021.
- [24] B. K. Seon, A. Haba, F. Matsuno et al., "Endoglin-targeted cancer therapy," *Current Drug Delivery*, vol. 8, no. 1, pp. 135–143, 2011.
- [25] X. Zhang, Z. Kong, X. Xu et al., "ARRB1 drives gallbladder cancer progression by facilitating TAK1/MAPK signaling activation," *Journal of Cancer*, vol. 12, no. 7, pp. 1926–1935, 2021.
- [26] Y. Lei, X. Xu, H. Liu et al., "HBx induces hepatocellular carcinogenesis through ARRB1-mediated autophagy to drive the G1/S cycle," *Autophagy*, vol. 17, no. 12, pp. 4423–4441, 2021.
- [27] K. Mamouni, J. Kim, B. L. Lokeshwar, and G. Kallifatidis, "ARRB1 regulates metabolic reprogramming to promote glycolysis in stem cell-like bladder cancer cells," *Cancers (Basel)*, vol. 13, no. 8, p. 1809, 2021.
- [28] X. Wu, T. Liu, O. Fang, L. J. Leach, X. Hu, and Z. Luo, "miR-194 suppresses metastasis of non-small cell lung cancer through regulating expression of BMP1 and p27^{kip1}," *Oncogene*, vol. 33, no. 12, pp. 1506–1514, 2014.
- [29] J. X. Huang, Y. C. Wu, Y. Y. Cheng, C. L. Wang, and C. J. Yu, "IRF1 negatively regulates oncogenic KPNA2 expression under growth stimulation and hypoxia in lung cancer cells," *Oncotargets and Therapy*, vol. Volume 12, pp. 11475–11486, 2019.
- [30] W. Seo, K. Shimizu, S. Kojo et al., "Runx-mediated regulation of CCL5 via antagonizing two enhancers influences immune cell function and anti-tumor immunity," *Nature Communications*, vol. 11, no. 1, p. 1562, 2020.
- [31] Z. Wang and M. Peng, "A novel prognostic biomarker LCP2 correlates with metastatic melanoma-infiltrating CD8⁺ T cells," *Scientific Reports*, vol. 11, no. 1, p. 9164, 2021.
- [32] Z. Wang, Y. Wang, M. Peng, and L. Yi, "UBASH3B is a novel prognostic biomarker and correlated with immune infiltrates in prostate cancer," *Frontiers in Oncology*, vol. 9, p. 1517, 2019.
- [33] Y. Huo, K. Zhang, S. Han, Y. Feng, and Y. Bao, "Lymphocyte cytosolic protein 2 is a novel prognostic marker in lung adenocarcinoma," *The Journal of International Medical Research*, vol. 49, no. 11, article 3000605211059681, 2021.
- [34] B. Seliger, "Molecular mechanisms of HLA class I-mediated immune evasion of human tumors and their role in resistance to immunotherapies," *HLA*, vol. 88, no. 5, pp. 213–220, 2016.
- [35] W. C. Wang, Y. S. Lin, Y. F. Chang et al., "Association of HLA-DPA1, HLA-DPB1, and HLA-DQB1 alleles with the long-term and booster immune responses of young adults vaccinated against the hepatitis B virus as neonates," *Frontiers in Immunology*, vol. 12, article 710414, 2021.
- [36] X. K. Wang, X. W. Liao, C. K. Yang et al., "Diagnostic and prognostic biomarkers of human leukocyte antigen complex for hepatitis B virus-related hepatocellular carcinoma," *Journal of Cancer*, vol. 10, no. 21, pp. 5173–5190, 2019.
- [37] L. Lyu, J. Yao, M. Wang et al., "Overexpressed pseudogene HLA-DPB2 promotes tumor immune infiltrates by regulating HLA-DPB1 and indicates a better prognosis in breast cancer," *Frontiers in Oncology*, vol. 10, p. 1245, 2020.
- [38] M. J. Callahan, Z. Nagymanyoki, T. Bonome et al., "Increased HLA-DMB expression in the tumor epithelium is associated with increased CTL infiltration and improved prognosis in advanced-stage serous ovarian cancer," *Clinical Cancer Research*, vol. 14, no. 23, pp. 7667–7673, 2008.
- [39] M. Ningappa, C. Ashokkumar, B. W. Higgs et al., "Enhanced B cell alloantigen presentation and its epigenetic dysregulation in liver transplant rejection," *American Journal of Transplantation*, vol. 16, no. 2, pp. 497–508, 2016.
- [40] Y. Han, D. Liu, and L. Li, "PD-1/PD-L1 pathway: current researches in cancer," *American Journal of Cancer Research*, vol. 10, no. 3, pp. 727–742, 2020.
- [41] B. Rowshanravan, N. Halliday, and D. M. Sansom, "CTLA-4: a moving target in immunotherapy," *Blood*, vol. 131, no. 1, pp. 58–67, 2018.
- [42] M. Khan, S. Arooj, and H. Wang, "Soluble B7-CD28 family inhibitory immune checkpoint proteins and anti-cancer immunotherapy," *Frontiers in Immunology*, vol. 12, article 651634, 2021.

Research Article

A Prognostic Model for the Respiratory Function of Patients with Nonsevere Pulmonary Infection Based on Breathing Exercises and Acupuncture Therapy: Development and Validation

Yulin Shi ¹, Yong Hu ¹, Guomeng Xu ¹, and Yaoqi Ke ²

¹Department of Rehabilitation Medicine, Xiangyang Central Hospital,
Affiliated Hospital of Hubei University of Arts and Science, China

²Department of Respiratory Medicine, Xiangyang Central Hospital,
Affiliated Hospital of Hubei University of Arts and Science, China

Correspondence should be addressed to Guomeng Xu; drxuguomeng@163.com and Yaoqi Ke; keqiqiyyy@163.com

Received 28 July 2022; Revised 5 September 2022; Accepted 9 September 2022; Published 29 September 2022

Academic Editor: Tao Huang

Copyright © 2022 Yulin Shi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. In this study, a prognostic model for the respiratory function was constructed based on the treatment methods of patients with nonsevere pulmonary infection, aiming to provide a reference for clinical decision-making. **Method.** A total of 500 patients with nonsevere pulmonary infection were included in this study. The patients were randomized into training set ($n = 350$) and validation set ($n = 150$), and the baseline characteristics were collected. All patients received breathing exercises or breathing exercises combined with acupuncture therapy for 3 months, and then the forced expiratory volume in one second (FEV1)/forced vital capacity (FVC) was assessed. Next, an ordinal multinomial logistic regression model was used to analyze prognostic factors affecting respiratory function of patients with nonsevere pulmonary infection. The Test of Parallel Lines was used to determine the accuracy (ACC) of the model and screen the influencing factors. The confusion matrix was drawn, and the ACC and harmonic mean (F1 score) were calculated to evaluate the feasibility of the model results. **Results.** Results of the ordinal multinomial logistic regression model showed that age ($P = 0.000$), treatment method ($P = 0.000$), underlying diseases ($P < 0.001$), and sex ($P = 0.389$) were independent factors affecting the respiratory function of patients in the training set. The ACC value of the training set was 88.86%, and that of the validation set was 91.33%, indicating a high accuracy and favorable predictive ability of the model. Besides, the F1 score was 62.38%, indicating a high reliability of the model. **Conclusion.** The prognostic model for respiratory function of patients with nonsevere pulmonary infection constructed in this study had favorable predictive performance, which is of great significance in the clinical nursing and treatment of patients with pulmonary infection.

1. Introduction

Pulmonary infection generally refers to bacterial, fungal, or viral infection in the lungs or pulmonary parenchyma, resulting in a series of inflammatory reactions [1]. Pneumonia patients show a variety of clinical symptoms, among which the most typical one is dyspnea. Dyspnea patients cannot smoothly breathe with asthma of varying severity, shortness of breath, and even gasp [1]. It was demonstrated that patients with pulmonary infection generally have a favorable prognosis when they receive timely treatment [2].

Breathing exercises are often used to help patients with pulmonary infection recover the respiratory function and improve their quality of life.

Breathing exercises can mend ventilation, improve diaphragmatic function, endurance, and coordination, and maintain or improve thoracic motion, thereby helping to establish effective breathing patterns and relieve respiratory diseases [3]. As a common clinical nursing method, breathing exercises are widely used. For example, the incidence of postoperative pulmonary complications after laparoscopic colorectal surgery can be reduced by perioperative breathing

TABLE 1: Comparison of baseline characteristics of patients.

Characteristics	Training set ($n = 350$)	Validation set ($n = 150$)	P value
Age (years old)	54 (45, 65)	56 (45, 67)	0.684
Sex			0.938
Male (%)	174 (49.7)	74 (49.3)	
Female (%)	176 (50.3)	76 (50.7)	
Smoking history			0.342
Never (%)	113 (32.3)	42 (28.0)	
Smoking (%)	237 (67.7)	108 (72.0)	
PaO ₂ ^a	85 (81, 93)	88.5 (82, 95)	0.784
PaCO ₂ ^b	40 (38, 43)	40 (38, 43)	0.600
PH	7.39 (7.37, 7.42)	7.40 (7.37, 7.43)	0.476
Including a least one underlying disease (%)	90 (25.7)	43 (28.7)	0.254
Hypertension (%)	36 (10.3)	23 (15.3)	
Diabetes (%)	34 (9.7)	25 (16.7)	
Hyperlipoidemia (%)	22 (6.3)	19 (12.7)	
COPD (%)	16 (4.6)	14 (9.3)	
Treatment methods			0.520
Breathing exercises only (%)	171 (48.9)	78 (52.0)	
Breathing exercises combined with acupuncture therapy (%)	179 (51.1)	72 (48.0)	

Note: a: partial pressure of oxygen; b: partial pressure of carbon dioxide.

exercise [4]. Charusisin et al. [5] demonstrated that breathing exercise can improve the respiratory function of patients with chronic obstructive pulmonary disease (COPD). However, a randomized trial shows no significant improvement in the respiratory function of patients with COPD after exercise-training plus breathing-retraining [6]. Furthermore, acupuncture, as a unique Chinese therapy, is a kind of medical skill of “internal diseases treated by external application” which is also widely used in clinical treatment of lung diseases to decrease adverse symptoms such as dyspnea. A study indicated that warm acupuncture plays an effective role in the improvement of symptoms of patients with acute exacerbation COPD with phlegm-turbid obstruction of the lung [7]. A systematic review illustrated that acupuncture may remarkably improve dyspnea and quality of life of patients [8]. Hence, we constructed a prognostic model for the respiratory function based on breathing exercises and acupuncture therapy, thereby investigating the prognostic factors affecting the respiratory function of patients with nonsevere pulmonary infection.

The study of pneumonia prognosis is of notable value. Ahn et al. [9] constructed a prognostic model for patients with pneumonia, confirming that Eastern Cooperative Oncology Group (ECOG) scores, oxygen saturation (SpO₂), and lactic acid are independent prognostic factors of mortality. Another research constructed a COX proportional-hazards model to analyze the mechanical ventilation prognosis of patients with hypoxemic acute respiratory failure treated with high-flow nasal cannula (HFNC) [10]. In this study, we constructed an ordinal multinomial logistic regression model with the baseline characteristics of patients as input variables, used parallel line hypothesis to determine the accuracy of the model, screened the influenc-

ing factors, drew confusion matrices, and calculated the prediction accuracy of the model.

2. Materials and Methods

2.1. Subject. Clinical data of 500 patients with nonsevere pulmonary infection treated in our hospital were collected in this study. The inclusion criteria for patients were CURB (confusion, uremia, elevated respiratory rate, and hypotension) score ≤ 2 , consciousness disorders, urinary nitrogen ≥ 7 mmol/L, respiratory rate ≥ 30 times/min, systolic pressure ≤ 90 mmHg or diastolic blood pressure ≥ 60 mmHg, and age ≥ 65 . Patients with serious underlying diseases were excluded.

All patients were divided into training set ($n = 350$) and validation set ($n = 150$) by random sampling method. The training set was utilized to construct the model and the validation set was employed to verify the accuracy of the prognostic model. Breathing exercise mainly referred to two maximal continuous ventilation maneuvers per day for 15 min with a 15-min interval between the two maneuvers. Breathing exercise combined with acupuncture therapy was performed twice a week for 10 min each on the basis of breathing training. The cycles for both treatments were 3 months. Manifested in Table 1 are clinical characteristics of the two groups. The study received ethical approval from the Ethics Committee of Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science. Since this was a retrospective study, the informed consent of patients was waived.

2.2. Prediction of Results. The analysis of qi and blood was performed on the patients before treatment, and partial arterial oxygen pressure (PaO₂), partial pressure of carbon

TABLE 2: Parameter estimates of each variable in the ordinal multinomial logistic regression model.

		Estimate	Std. Error	Sig.
Threshold	Stage = 1	-9.501	1.043	0.000
	Stage = 2	15.539	0.817	0.000
	Age	-0.047	0.014	0.000
	[treatment = 0]	-4.754	0.445	0.000
	[treatment = 1]	0 ^a	.	.
Location	[underlying disease = 0]	19.980	0.000	< 0.001
	[underlying disease = 1]	0 ^a	.	.
	[sex = 0]	0.262	0.304	0.389
	[sex = 1]	0 ^a	.	.

treatment = 0: breathing exercise alone; treatment = 1: breathing exercise combined with acupuncture therapy; a: this parameter is redundant and set to 0.

dioxide (PaCO₂), and pH of the patients were measured and included as variables affecting respiratory function. Forced expiratory volume in one second (FEV1), forced vital capacity (FVC), and FEV1/FVC ratio were measured after three months of treatment. FEV1/FVC was used as the evaluation index and was divided into three categories according to the diagnosis and classification criteria of GOLD2013 edition. FEV1/FVC < 0.7 was classified as category 1, and 14 patients in this category still had chronic obstructive pulmonary diseases. 0.7 ≤ FEV1/FVC < 0.8 was classified as category 2, and there were 266 patients in total who recovered better. 0.8 ≤ FEV1/FVC < 8 was classified as category 3, and there were 220 patients in total. The respiratory function of these patients had even reached a good level.

2.3. Statistical and Analysis. SPSS 26.0 software was used to test the statistical differences of baseline characteristics between two groups. Median (P25, P75) was used for statistical description of measurement data that did not conform to normal distribution. *U*-test was applied for intergroup comparison, while chi-square test was used for intergroup comparison of categorical data. *P* value less than 0.05 was considered statistically significant. Patients were randomly divided into validation set and test set by ‘sample’ package of R software, with the ratio of 7:3.

The principle of ordinal multinomial logistic regression is that the multiple categories of dependent variable are successively divided into multiple binary logistic regressions. In this study, the FEV1/FVC index was selected, and the dependent variable was three categories with an order relation between the categories. In the analysis, it was divided into two binary logistic regressions, namely (1 vs. 2 + 3) and (1 + 2 vs. 3), both of which contrasted lower level with higher level, and its mathematical Formula (1) was

$$\ln \left[\sum_{i=1}^j p_i \setminus \left(1 - \sum_{i=1}^j p_i \right) \right] = a_j + \sum_{i=1}^m b_i x_i, j = 1, 2, \quad (1)$$

where $X = (x_1, x_2, \dots, x_m)$ is the independent variable vector,

p_j is the probability estimation of the dependent variable taking j level, a_j is the estimate of the intercept parameter, and b_i is the estimate of the partial regression coefficient β_i .

In the ordinal multinomial logistic regression model, the hypothesis of equal coefficients of independent variables should be tested by parallel lines. The Test of Parallel Lines is also the process of factor screening. First, the factor selected should make the test *P* (Sig.) value greater than 0.05. In this case, the null hypothesis was not rejected, so the hypothesis could be considered as valid and ordinal multinomial logistic regression can be used. Next, the chi-square (likelihood ratio) test was used to make an overall evaluation of the model, $\chi^2 = 487.665, P = 0.000$ indicating that the model was statistically significant. At the same time, by referring to the Pearson’s chi-squared test $P = 1.000 > 0.05$ and deviation test $P = 1.000 > 0.05$ for the goodness of fit, model was suggested to be statistically significant.

To screen the remarkable influencing factors of evaluation index, the ordinal multinomial logistic regression model was gradually established in the direction of the increase of *P* (Sig.) value assumed by parallel lines. The *P* value of the full variable model before screening was 0.000, and the *P* value of the optimal model after screening was 0.816, with $P > 0.05$ in the results, so the null hypothesis was accepted. The results of optimal multinomial ordered logistic regression are shown in Table 2. Age ($P = 0.000$), treatment method ($P = 0.000$), underlying diseases ($P < 0.001$), and sex ($P = 0.389$) were independent factors affecting the respiratory function of patients in the training set.

An ordinal multinomial logistic regression was constructed. The Test of Parallel Lines was used to determine the accuracy [11] of the model and screen the influencing factors. The confusion matrix was drawn, and the accuracy (ACC) of predicted result by the model (ACC, Formula (2)) as well as the harmonic mean (F1 score, Formula (3)) was calculated.

$$ACC = \frac{TP + TN}{TP + FN + FP + TN}, \quad (2)$$

$$F_1 = \frac{2 \times P \times R}{P + R}, \quad (3)$$

where TP is the true positive, TN is the true negative, FP is the false positive, FN is the false negative, *P* is the Precision, and *R* is the recall.

3. Results

3.1. Clinical Characteristics of Patients. In the training set, the age distribution of patients was 54 (45, 65) years old, including 174 males (49.7%) and 176 females (50.3%), and 171 patients (48.9%) with only breathing exercises and 179 patients (51.1%) with combined breathing exercise and acupuncture therapy. In the validation set, the age distribution of patients was 56 (45, 67) years old, including 74 males (49.3%) and 76 females (50.7%), and 78 patients (50.7%) with only breathing exercises and 72 patients (48%) with

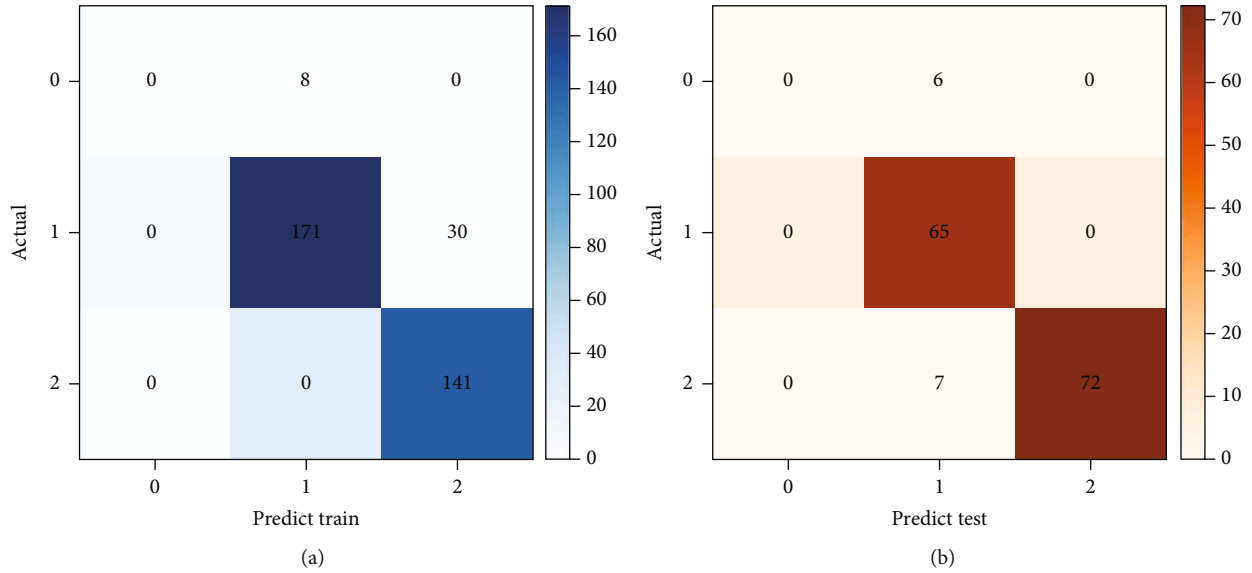


FIGURE 1: Confusion matrix diagram of predicted-actual values.

TABLE 3: Confusion matrix of the results predicted by optimal multinomial ordered logistic regression model in the validation set.

Confusion matrix	Predicted category			Total number
	0	1	2	
Actual category 0	0	6	0	6
Actual category 1	0	65	0	65
Actual category 2	0	7	72	79
Total number	0	78	72	150

combined breathing exercise and acupuncture therapy. Other baseline characteristics are shown in Table 1.

3.2. Construction of an Ordinal Multinomial Logistic Regression Model. The Estimates of Stage = 1 and 2 corresponding to threshold were the constant terms of the two binary logistic regressions split in this analysis. The parameter Estimate (-0.047) corresponding to the age was the Estimate of the independent variable.

-4.754 of the treatment = 0 coefficient (Estimate) indicated that patients treated with breathing exercise alone were at least 12.922 times more likely to recover at least one grade higher respiratory function than those treated with breathing exercise combined with acupuncture under certain variables [$\exp(-4.754)$], indicating that respiratory training combined with acupuncture therapy had a better recovery effect. The other coefficients were interpreted the same way.

The sex variable coefficient presented no statistical significance ($P = 0.389$). However, this study considered age affecting respiratory function. Even in this study, its coefficient was not significant and it should be retained in the model (because no statistical significance might be caused by small sample size, and it did not mean the variable had no effect). Therefore, it was retained in this study.

3.3. Assessment and Validation of the Model. Confusion matrix diagram (Figure 1) and confusion matrix table for

predicted results (Table 3) were drawn based on the constructed optimal multinomial ordered logistic regression model, and then the ACC value of the model was calculated. The ACC value of the training set was 88.86%, and that of the validation set was 91.33%, indicating that the prognostic model of respiratory function in patients with nonsevere pulmonary infection constructed in this study had high ACC. Subsequently, by calculating the recall ratio (63.71%) and precision ratio (61.11%), the F1 score was obtained to be 62.38%, further indicating that the performance of the model was reliable.

4. Discussion

In this study, a prognostic model of respiratory function in patients with nonsevere pulmonary infection was constructed based on the treatment methods, which indicated the effects of the age, treatment methods, underlying diseases, and sex on the recovery of respiratory function. The results showed that ACC value of the training set was 88.86%, and that of the validation set was 91.33%, indicating the favorable predictive ability and high credibility of the model, which provided certain references for the treatment of patients with nonsevere pulmonary infection.

A study on the risk of COVID-19 progression showed that old age and complications are the independent risk factors of the progression of pulmonary infection [12]. In recent years, studies on COVID-19 patients have shown that advanced age and comorbidities are susceptible to COVID-19 and increase the proportion of severe cases, which in turn increases the risk of death [13–16]. Singh et al. [17] pointed out that diabetes may worsen the conditions of pulmonary infection. The common underlying diseases of elderly patients mainly include hypertension, coronary heart disease, COPD, diabetes, and cardiovascular and cerebrovascular diseases [14], which may be related to the low immune function in elderly patients. In this study, the results of

optimal multinomial ordered logistic regression indicated that the prognosis of respiratory function in elderly patients with pulmonary infection and in patients with underlying diseases is poor. In addition, studies have reported that men are more likely to suffer from COVID-19 [13, 18] and severe acute respiratory distress syndrome [18–20], which may be related to the innate and adaptive immune effects of female sex hormones, although our results showed no statistically significant difference in the effect of sex on respiratory function in patients with nonsevere pulmonary infection, which may be caused by small sample size. Therefore, the clinical management of nonsevere pulmonary infection should comprehensively consider all influencing factors and pay special attention to elderly male patients with underlying diseases.

Breathing exercise was found to help improve COVID-19 symptoms in a randomized controlled trial [21]. Pu et al. [22] demonstrated through meta-analysis that breathing exercise before curative pneumonectomy in lung cancer patients can help reduce hospital length of stay, postoperative pulmonary complications, and pneumonia. In addition, breathing exercise has been confirmed to improve respiratory function and quality of life in patients with COPD and asthma [11, 23, 24]. Acupuncture can also be used as an adjunctive treatment for COPD, asthma, lung cancer, and COVID-19, demonstrating its significant efficacy and safety for respiratory diseases [25–29]. Acupuncture may assist in alleviating symptoms of respiratory diseases by reducing bronchial immune-mediated inflammation and promoting the release of vascular and immune regulatory factors [30]. In this study, a multivariate logistic regression model was constructed to predict respiratory function using FEV1/FVC of patients after treatment as the dependent variable. All patients included in the study underwent breathing exercise or breathing exercise combined with acupuncture for 3 months. The results showed that the respiratory function prognosis of patients with lung infection treated with combined therapy was better than that of patients who underwent breathing exercise alone. However, there were some defects because of the small sample size and the fact that only FEV1/FVC was used as evaluation index of respiratory function.

In conclusion, a prognostic model for respiratory function in patients with nonsevere pulmonary infection was constructed based on the treatment methods, thus obtaining the prognostic impact factors. However, the safety of both treatments was not analyzed in this study. Therefore, more samples should be included in the future, and evaluation indicators should be added based on the optimization of prediction model. Then the impact of breathing exercises and acupuncture therapy on the respiratory function of patients with nonsevere pulmonary infection can be further explored.

Data Availability

The datasets generated and analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

Ethical Approval

The study received ethical approval from the Ethics Committee of Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science. Since this was a retrospective study, the informed consent of patients was waived.

Conflicts of Interest

All authors declare that they have no potential conflicts of interest.

Authors' Contributions

YL S contributed to conceptualization and design. Y H contributed to data collection and formal analysis. GM X contributed to original drafting. YQ K contributed to the revising of the article. All authors read and approved the final manuscript. Dr. Yulin Shi and Dr. Yong Hu contributed equally to this work.

References

- [1] L. A. Mandell, "Community-acquired pneumonia: an overview," *Postgraduate Medicine*, vol. 127, no. 6, pp. 607–615, 2015.
- [2] A. Fernández-Cruz, L. Ortega, G. García et al., "Etiology and prognosis of pneumonia in patients with solid tumors: a prospective cohort of hospitalized cases," *The Oncologist*, vol. 25, no. 5, pp. e861–e869, 2020.
- [3] A. Vatwani, "Pursed lip breathing exercise to reduce shortness of breath," *Archives of Physical Medicine and Rehabilitation*, vol. 100, no. 1, pp. 189–190, 2019.
- [4] P. P. Qin, J. Y. Jin, W. J. Wang, and S. Min, "Perioperative breathing training to prevent postoperative pulmonary complications in patients undergoing laparoscopic colorectal surgery: a randomized controlled trial," *Clinical Rehabilitation*, vol. 35, no. 5, pp. 692–702, 2021.
- [5] N. Charususin, R. Gosselink, A. McConnell et al., "Inspiratory muscle training improves breathing pattern during exercise in COPD patients," *The European Respiratory Journal*, vol. 47, no. 4, pp. 1261–1264, 2016.
- [6] E. G. Collins, C. Jelinek, S. O'Connell, J. Butler, D. Reda, and F. Laghi, "The effect of breathing retraining using metronome-based acoustic feedback on exercise endurance in COPD: a randomized trial," *Lung*, vol. 197, no. 2, pp. 181–188, 2019.
- [7] F. Xie, Y. P. Wu, L. Liu, J. G. Ren, and B. D. Zhang, "Warm acupuncture on chronic obstructive pulmonary disease with phlegm-turbid obstructing of the lung: a randomized controlled trial," *Zhongguo Zhen Jiu*, vol. 39, no. 9, pp. 918–922, 2019.
- [8] M. E. Coyle, J. L. Shergis, E. T. Huang et al., "Acupuncture therapies for chronic obstructive pulmonary disease: a systematic review of randomized, controlled trials," *Alternative Therapies in Health and Medicine*, vol. 20, no. 6, pp. 10–23, 2014.
- [9] B. K. Ahn, Y. S. Lee, Y. J. Kim et al., "Prediction model for mortality in cancer patients with pneumonia: comparison with CURB-65 and PSI," *The Clinical Respiratory Journal*, vol. 12, no. 2, pp. 538–546, 2018.

- [10] O. Roca, J. Messika, B. Caralt et al., "Predicting success of high-flow nasal cannula in pneumonia patients with hypoxemic respiratory failure: the utility of the ROX index," *Journal of Critical Care*, vol. 35, pp. 200–205, 2016.
- [11] K. B. Evaristo, F. A. R. Mendes, M. G. Saccomani et al., "Effects of aerobic training versus breathing exercises on asthma control: a randomized trial," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 8, no. 9, pp. 2989–2996.e4, 2020.
- [12] D. Ji, D. Zhang, J. Xu et al., "Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score," *Clinical Infectious Diseases*, vol. 71, no. 6, pp. 1393–1399, 2020.
- [13] N. Chen, M. Zhou, X. Dong et al., "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," *The Lancet*, vol. 395, no. 10223, pp. 507–513, 2020.
- [14] S. Niu, S. Tian, J. Lou et al., "Clinical characteristics of older patients infected with COVID-19: a descriptive study," *Archives of Gerontology and Geriatrics*, vol. 89, article 104058, 2020.
- [15] N. Lovell, M. Maddocks, S. N. Etkind et al., "Characteristics, symptom management, and outcomes of 101 patients with COVID-19 referred for hospital palliative care," *Journal of Pain and Symptom Management*, vol. 60, no. 1, pp. e77–e81, 2020.
- [16] Y. D. Peng, K. Meng, H. Q. Guan et al., "Clinical characteristics and outcomes of 112 cardiovascular disease patients infected by 2019-nCoV," *Zhonghua Xin Xue Guan Bing Za Zhi*, vol. 48, no. 6, pp. 450–455, 2020.
- [17] A. K. Singh, R. Gupta, A. Ghosh, and A. Misra, "Diabetes in COVID-19: prevalence, pathophysiology, prognosis and practical considerations," *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, vol. 14, no. 4, pp. 303–310, 2020.
- [18] X. Yang, Y. Yu, J. Xu et al., "Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study," *The Lancet Respiratory Medicine*, vol. 8, no. 5, pp. 475–481, 2020.
- [19] R. Channappanavar, C. Fett, M. Mack, P. P. ten Eyck, D. K. Meyerholz, and S. Perlman, "Sex-based differences in susceptibility to severe acute respiratory syndrome coronavirus infection," *Journal of Immunology (Baltimore, Md. : 1950)*, vol. 198, no. 10, pp. 4046–4053, 2017.
- [20] A. Badawi and S. G. Ryoo, "Prevalence of comorbidities in the Middle East respiratory syndrome coronavirus (MERS-CoV): a systematic review and meta-analysis," *International Journal of Infectious Diseases*, vol. 49, pp. 129–133, 2016.
- [21] C. Rodríguez-Blanco, C. Bernal-Utrera, E. Anarte-Lazo et al., "Breathing exercises versus strength exercises through telerehabilitation in coronavirus disease 2019 patients in the acute phase: a randomized controlled trial," *Clinical Rehabilitation*, vol. 36, no. 4, pp. 486–497, 2022.
- [22] C. Y. Pu, H. Batarseh, M. L. Zafron, M. J. Mador, S. Yendamuri, and A. D. Ray, "Effects of preoperative breathing exercise on postoperative outcomes for patients with lung cancer undergoing curative intent lung resection: a meta-analysis," *Archives of Physical Medicine and Rehabilitation*, vol. 102, no. 12, pp. 2416–2427.e4, 2021.
- [23] R. Yun, Y. Bai, Y. Lu, X. Wu, and S. D. Lee, "How breathing exercises influence on respiratory muscles and quality of life among patients with COPD? A systematic review and meta-analysis," *Canadian Respiratory Journal*, vol. 2021, Article ID 1904231, 11 pages, 2021.
- [24] A. Hasanpour Dehkordi, S. Ebrahimi-Dehkordi, F. Banitalebi-Dehkordi, S. Salehi Tali, S. Kheiri, and A. Soleimani Babadi, "The effect of teach-back training intervention of breathing exercise on the level of dyspnea, six-minutes walking test and FEV1/FVC ratio in patients with chronic obstructive pulmonary disease; a randomized controlled trial," *Expert Review of Respiratory Medicine*, vol. 15, no. 1, pp. 161–169, 2021.
- [25] C. S. Cheng, L. Y. Chen, Z. Y. Ning et al., "Acupuncture for cancer-related fatigue in lung cancer patients: a randomized, double blind, placebo-controlled pilot trial," *Supportive Care in Cancer: Official Journal of the Multinational Association of Supportive Care in Cancer*, vol. 25, no. 12, pp. 3807–3814, 2017.
- [26] C. F. Liu and L. W. Chien, "Efficacy of acupuncture in children with asthma: a systematic review," *Italian Journal of Pediatrics*, vol. 41, no. 1, p. 48, 2015.
- [27] T. Reinhold, B. Brinkhaus, S. N. Willich, and C. Witt, "Acupuncture in patients suffering from allergic asthma: is it worth additional costs?," *Journal of Alternative and Complementary Medicine (New York, N.Y.)*, vol. 20, no. 3, pp. 169–177, 2014.
- [28] M. Yu, L. Gao, Y. Kong et al., "Safety and efficacy of acupuncture for the treatment of chronic obstructive pulmonary disease: a systematic review protocol," *Medicine (Baltimore)*, vol. 98, no. 37, article e17112, 2019.
- [29] B. Zhang, K. Zhang, Q. Tang, K. Sun, and Z. Han, "Acupuncture for breathlessness in COVID-19: a protocol for systematic review and meta-analysis," *Medicine (Baltimore)*, vol. 99, no. 27, article e20701, 2020.
- [30] E. R. Carneiro, C. R. Carneiro, M. A. Castro, Y. Yamamura, and V. L. Silveira, "Effect of electroacupuncture on bronchial asthma induced by ovalbumin in rats," *Journal of Alternative & Complementary Medicine*, vol. 11, no. 1, pp. 127–134, 2005.

Research Article

A Risk-Assessing Signature Based on Hypoxia- and Immune-Related Genes for Prognosis of Lung Adenocarcinoma Patients

Yue Wang , Jian Feng , Yang Liu , and Tingyue Han 

Cardio-Thoracic Surgery, Funing People's Hospital, Yancheng 224400, China

Correspondence should be addressed to Tingyue Han; hty13851333996@163.com

Received 7 July 2022; Revised 22 August 2022; Accepted 26 August 2022; Published 28 September 2022

Academic Editor: Tao Huang

Copyright © 2022 Yue Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lung Adenocarcinoma (LUAD) drastically influences human health. Tumor hypoxia and immunity impact hugely on the immunotherapeutic effect of LUAD patients. This study is aimed at exploring the prognostic markers associated with hypoxia and immunity in LUAD patients and evaluates their reliability. The relationship between hypoxia and immune-related genes and prognoses of LUAD patients was investigated by the univariate regression analysis. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) methods were used to reveal the enriched pathways and biological processes of prognosis-related genes. Univariate, LASSO, and multivariate Cox regression analyses were used to construct a prognostic signature and verify its independence. The reliability of the signature was evaluated by the Principal Component Analysis (PCA), the Kaplan-Meier (K-M) curve, and the receiver operating characteristic (ROC) curve. Gene set enrichment analysis (GSEA), tumor mutational burden (TMB), and single-sample GSEA (ssGSEA) further verified the performance of the signature. Finally, a prognostic signature for LUAD was constructed based on 7 hypoxia- and immune-related genes. According to riskScores acquired from the signature, the test set was divided into groups, where the prognosis of high-risk patients was poor. The feature genes had good reliability, and the riskScore could be used as an independent prognostic factor for LUAD patients. Meanwhile, high TMB scores and low immune scores were found in high-risk patients, and feature genes were enriched in signaling pathways such as cell cycle and p53 signaling pathway. In sum, a prognostic signature based on 7 hypoxia- and immune-related genes was constructed.

1. Introduction

The morbidity and mortality of lung cancer (LC) are high worldwide. Statistic by International Agency for Research on Cancer (IARC) showed that 11.4% new cases of cancers are LC, only followed by breast cancer (11.7%) [1]. In addition, the mortality of lung cancer accounts for 18%, ranking the first globally [1]. Lung adenocarcinoma (LUAD) is the most frequent histological subtype of LC [2]. Progress has been made in LUAD-related research in the past decades, but the survival rate is not noticeably improved, with 70% of patients developing local progression or metastasis when they were first diagnosed [3–5]. In the above conditions, the impact of traditional interventions on patients is only revealed after treatment, which adds patients' pain. Hence, the clinician needs to take the treatment sensitivity and

prognosis of patients into consideration for making further treatment plans.

Hypoxia is a typical feature of the tumor microenvironment. Cells were driven and became invasive in the hypoxia condition in hepatocellular carcinoma, colorectal cancer, and esophageal squamous cell carcinoma [6–8]. Evidence has shown that with hypoxia conditions, tumor cells can activate multiple transcription factors to further induce cell proliferation, invasion, and apoptosis [8, 9]. This enhances the drug resistance of tumor cells under a hypoxia environment, and hypoxia-induced cell function changes will affect the prognoses of patients [10]. Therefore, hypoxia is deemed to relate to drug resistance and poor treatment efficacy [11]. Research proved that a patient's prognosis can be predicted using hypoxia-related genes. Jiang et al. [12] confirmed that LBH is lowly expressed in glioma under hypoxia conditions

and causes poor prognosis. NLUCAT1 and TRB3 impact hugely on LUAD patient's prognosis in the hypoxia condition [13, 14]. It can be concluded that hypoxia is bound up with the prognosis of LUAD patients.

With deep investigation being carried out, immunotherapy has drawn more and more attention in treating lung cancer. In 2013, Science listed tumor immunotherapy as the top out of 10 breakthroughs in the science field. With the development of immunotherapy (like immune checkpoint inhibitors (ICIs)), methods for tumor therapy are upgraded [15]. The tumor microenvironment comprises surrounding immune and inflammatory cells, tumor cells, tumor-related fibroblasts, and nearby interstitial tissues, capillaries and various cytokines and chemokines. Many immune cells are highly sensitive to hypoxia, with their viabilities and antitumor capabilities lowered in such condition [16]. Lu et al. [17] revealed that proinflammatory cytokines (IL-13 and IL-23) and costimulatory molecules (CD80 and CD60) are downregulated in non-small-cell lung cancer cells, but anti-inflammatory cytokine (IL-10) secreted from the CD1c⁺ DCs is promoted; thereby, blocking the body's antitumor immunity. Ganesan et al. [18] analyzed purified CD8⁺ T cells in untreated, early NSCLC samples by RNA sequencing. The higher the density of tissue-resident memory T cells, the better the performance of LUAD patients' prognoses, but this fact is irrelevant with the density of cytotoxic T lymphocytes.

Meanwhile, immunotherapy was also affected by hypoxia. Presently, the goal of immunotherapy is to induce and enhance cytotoxic T lymphocyte (CTL) effect [19]. In spite of that, hypoxia induces resistance towards CTL. The nuclear translocation of hypoxic cancer cells HIF-1 α , STAT3 phosphorylation, and VEGF secretion inhibit specific CTL-mediated cell lysis [20]. Additionally, Noman et al.'s study [21] reported hypoxia-induced autophagy as a crucial factor in CTL-mediated innate and adaptive antitumor immunity. Johnson et al. [22] discovered that MHCII can regulate the infiltration of T cells and respond to PD-1 sensitivity. Brooks et al. [23] developed a prognostic classifier for the head and neck cancer through analyzing hypoxia- and immune-related genes. Accordingly, the development of biomarkers based on hypoxia- and immune-related genes is meaningful to clinicians to judge the prognosis of patients.

In recent years, risk signatures based on genes have been widely investigated and applied to predict the prognosis of patients with colon cancer, breast cancer, and hepatocellular carcinoma [24, 25]. In some cancers, their prognostic performance is even better than histopathological diagnosis and tumor staging [24, 25]. However, few studies combined hypoxia and immune characteristics to investigate their relationship with the prognosis of LUAD. This study performed bioinformatics analysis on the prognostic value of hypoxia- and immune-related genes in LUAD, and finally constructed and validated a prognostic risk-assessing signature based on these 7 genes. Finally, tumor mutation burden (TMB) analysis and immune infiltration assessment were used to evaluate the impact of these genes on patients at different risk levels. This signature provides a reference to screening prognostic genes for LUAD patients and to optimizing LUAD immunotherapy.

2. Materials and Methods

2.1. Data Acquisition. We downloaded the gene expression profiles of 594 LUAD patients from The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov/>) database, along with the related clinical information and single nucleotide variants (SNV) mutation data (VarScan2 Annotation, including 561 samples). 243 hypoxia-related genes were downloaded from the Molecular Signatures Database (MSigDB; <https://www.gsea-msigdb.org/gsea/msigdb/>) [26]. At the same time, the expression data of 1,811 immune-related genes were downloaded from The Immunology Database and Analysis Portal (ImmPort, <https://www.immport.org>) [27]. GSE31210 dataset downloaded from the Gene Expression database (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) worked as a validation set. The expression of mRNAs and clinical data of 266 LUAD samples were downloaded as well (raw data were required from GPL570). All datasets used in this study were acquired from public databases, and thus ethical approval was not needed.

2.2. Screening and Enrichment Analysis of the Prognostic Hypoxia- and Immune-Related Genes. To screen prognostic hypoxia- and immune-related genes, 490 tumor samples with survival time exceeding 30 d were selected from TCGA-LUAD dataset as the test set. The test set was intersected with hypoxia- and immune-related datasets. Thereafter, univariate regression analysis ($p < 0.05$) was performed with package "survival" [28] (<https://cran.r-project.org/web/packages/survival/index.html>) on the dataset required. All the gene sets were aggregated to obtain hypoxia- and immune-related DEGs. To investigate the molecular mechanism of the above-acquired genes, GO (Gene ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) were adopted for enrichment analyses (q value < 0.05) using "clusterprofile" package (CRAN-Package shadowtext (<http://r-project.org>)). The results were visualized.

2.3. Construction and Evaluation of a Prognostic Signature. LASSO regression was performed using "glmnet" package [29] to reduce the overfitting of genes in building the signature. Multivariate Cox regression analysis was performed with "survival" package. The genes whose LASSO regression coefficient is not 0 were analyzed again, and finally a risk-assessing signature was obtained. The riskScore calculation formula was

$$\text{riskScore} = \sum_{i=1}^n \exp_i * \beta_i \quad (1)$$

In this formula, n represented the number of hypoxia- and immune-related prognostic genes and \exp_i represented the expression value of the hypoxia- and immune-related prognostic genes. β_i was the coefficient in the multivariate Cox regression analysis. Patient's riskScore was calculated accordingly and patients were divided into the high- and low-risk groups with the median riskScore as the critical value.

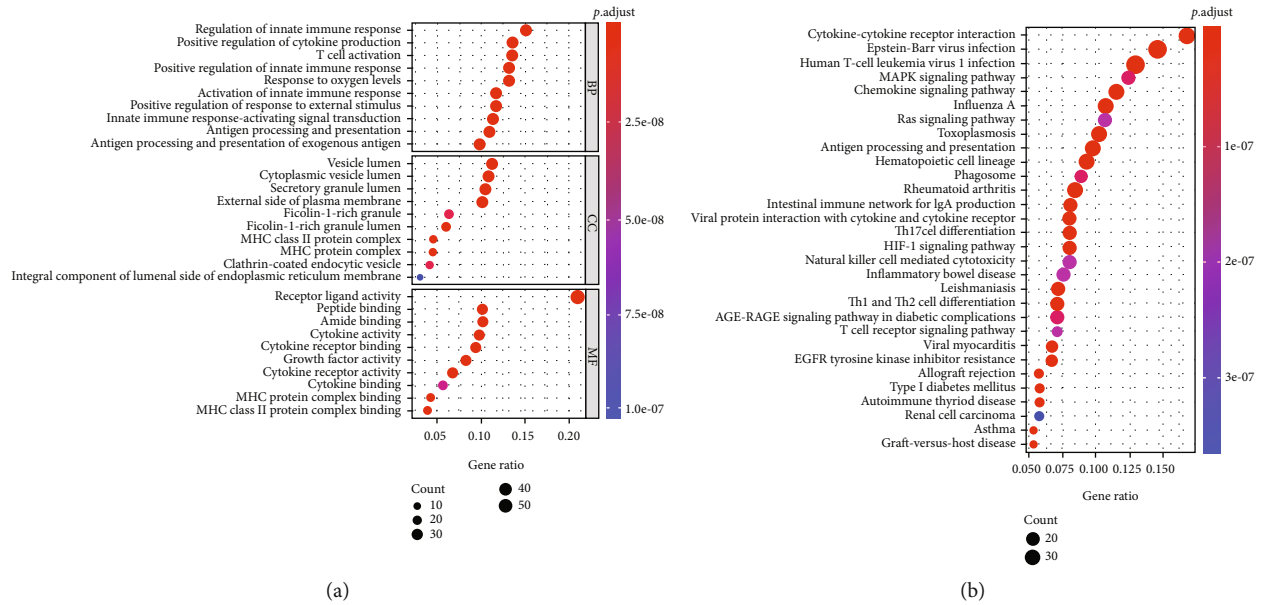


FIGURE 1: The results of functional enrichment analyses. (a, b) GO and KEGG enrichment analyses.

“FactoMineR” package was adopted to perform Principal Component Analysis (PCA) [30] on characteristic genes. “Survival” and the “timeROC” packages [31] were applied to draw the Kaplan-Meier (K-M) survival curve and receiver operating characteristic (ROC) curve. The overall survival (OS) of the two groups of patients was compared and analyzed. Then the AUC value of 1-year, 3-year, and 5-year OS was calculated based on the ROC curve to evaluate the predictive ability of the signature. GSE31210 data set was used for verification. Finally, in order to explore whether the riskScore has the value of independently predicting the prognosis of LUAD patients, the riskScore was combined with clinical information to conduct the univariate and multivariate Cox regression analyses.

2.4. Enrichment Analysis. To investigate reasons for riskScore divergence, gene set enrichment analysis (GSEA) was carried out on the high- and low-risk groups. We introduced differentially expressed genes (DEGs) to the GSEA software and calculated the Enrichment Score (ES) by comparing the DEGs with gene sets in KEGG pathways. In the line graph for ES, $ES > 0$ indicated an upregulation of the pathway in the high-risk group. Gene sets with $FDR < 0.25$ were considered statistically significant.

2.5. TMB Analysis. To discuss the way gene mutation frequency and patient’s riskScore interplayed, “GenVisR” package [32] was used to calculate the TMB values of the high- and low- risk groups. These values were then subject to the Wilcoxon test using “GenVisR” package. The mutation landscape waterfall chart of different risk groups was drawn.

2.6. Immune Infiltration Evaluation. The immune infiltration level and patient’s riskScore were evaluated; hereby, “estimate” package [33] and “GSVA” package [34] were uti-

lized to assess the matrix score and immune score of the LUAD samples in the test set. And each LUAD tumor sample was evaluated through single-sample GSEA (ssGSEA).

3. Results

3.1. Obtaining Prognostic-Related Genes from Hypoxia- and Immune-Related Genes. First, the 243 hypoxia-related genes from the MSigDB and the 1,811 immune-related genes from the ImmPort database were intersected with the genes with mRNA profiles of LUAD patients in TCGA. Then, the univariate Cox regression analysis was adopted to screen 65 prognostic-related hypoxia genes and 213 prognostic-related immune genes. After merging the prognostic-related hypoxia genes and prognostic-related immune genes, totally, 268 prognostic-related genes were obtained (Table S1). The GO enrichment analysis revealed gene enrichment in regulation of innate immune response, positive regulation of cytokine production, and T cell activation (Figure 1(a)). KEGG results suggested that most of these genes were enriched in cytokine-cytokine receptor interaction, MAPK signaling pathway, chemokine signaling pathway, Ras signaling pathway, and antigen processing and presentation (Figure 1(b)). Integrative analyses implicated that the 268 DEGs were mainly associated with the body’s innate immune regulatory response.

3.2. The Construction of Risk-Assessing Hypoxia- and Immune-Related Prognostic Signature. The feature genes were screened from 268 DEGs by the LASSO Cox regression analysis. The genes in the range marked within the dotted line were the best range for the identification of signature. Seven important genes were obtained when the LASSO regression coefficient was not 0 (Figures 2(a) and 2(b)). Multivariate Cox regression analysis was performed on these 7 genes and the final 7 hypoxia- and immune-

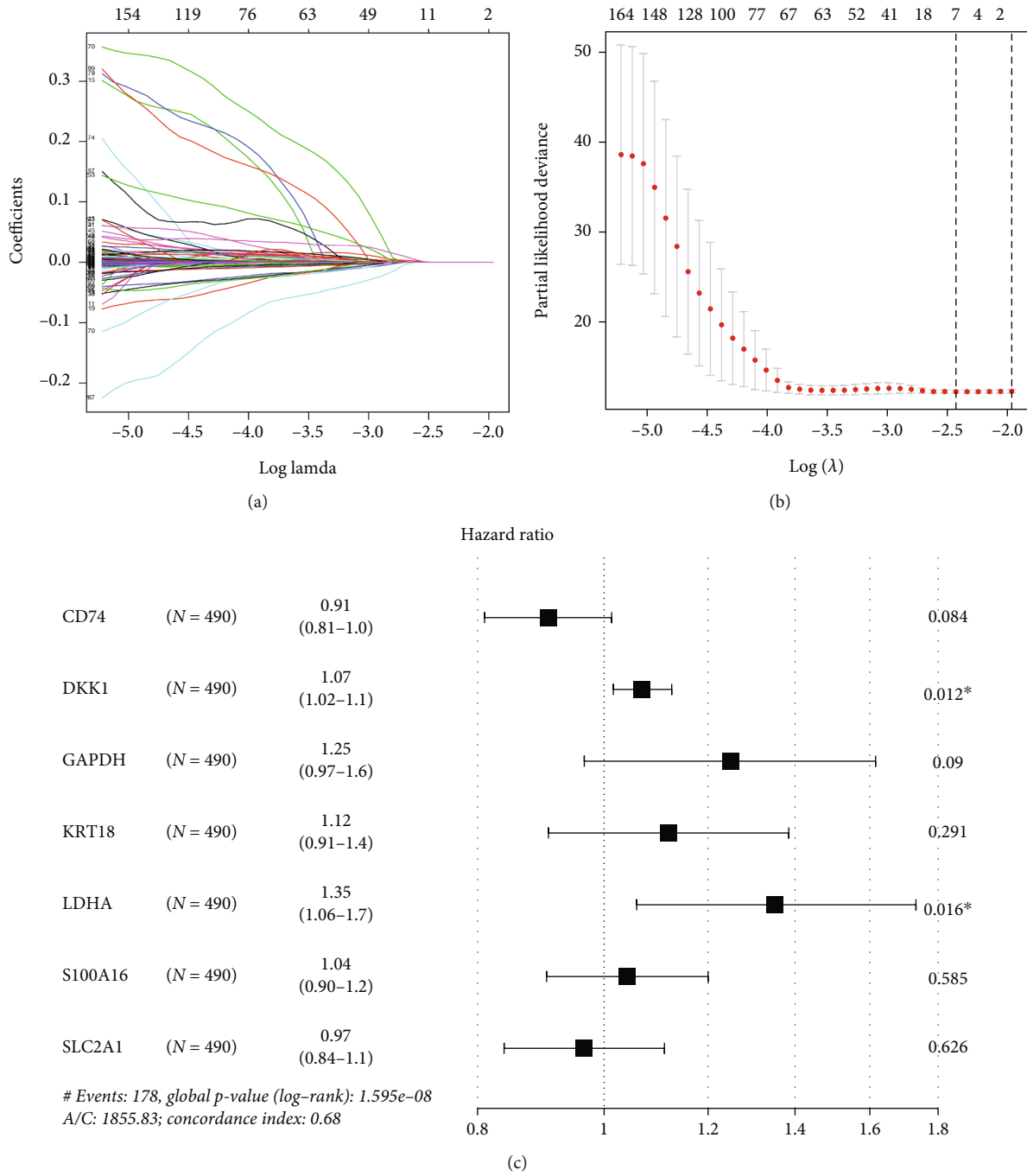


FIGURE 2: Construction of hypoxia- and immune-related signature. (a) The changing trajectory of coefficients of 268 hypoxia- and immune-related genes with the penalty function λ in the LASSO analysis. (b) The selection interval of the best penalty function λ . (c) Forest plot manifesting the multivariate Cox regression analysis ($*p < 0.05$).

related genes were CD74, DKK1, GAPDH, KRT18, LDHA, S100A16, and SLC2A1. The hazard ratio (HR) values of risk factors (DKK1, GAPDH, KRT18, LDHA, and S100A16) were greater than 1, and the HR values of protective factors (CD74 and SLC2A1) were less than 1 (Figure 2(c)). The final LUAD signature was $\text{riskScore} = -0.09897 * \text{CD74} + 0.06688 * \text{DKK1} + 0.2221 * \text{GAPDH} + 0.1131 * \text{KRT18} + 0.3026 * \text{LDHA} + 0.03997 * \text{S100A16} - 0.03507 * \text{SLC2A1}$.

3.3. Validation of the Prediction Ability of the 7-Genes Based Risk-Assessing Signature. The riskScore of the LUAD samples was calculated according to the established risk-assessing signature, and samples were divided into high- and low-risk groups according to the median value of riskScore (Figure 3(a)). A scatter plot of patient's survival time was drawn based on the grouping. The results revealed that as riskScore increased, patient's survival time was shortened and their mortality increased (Figure 3(b)). At the same

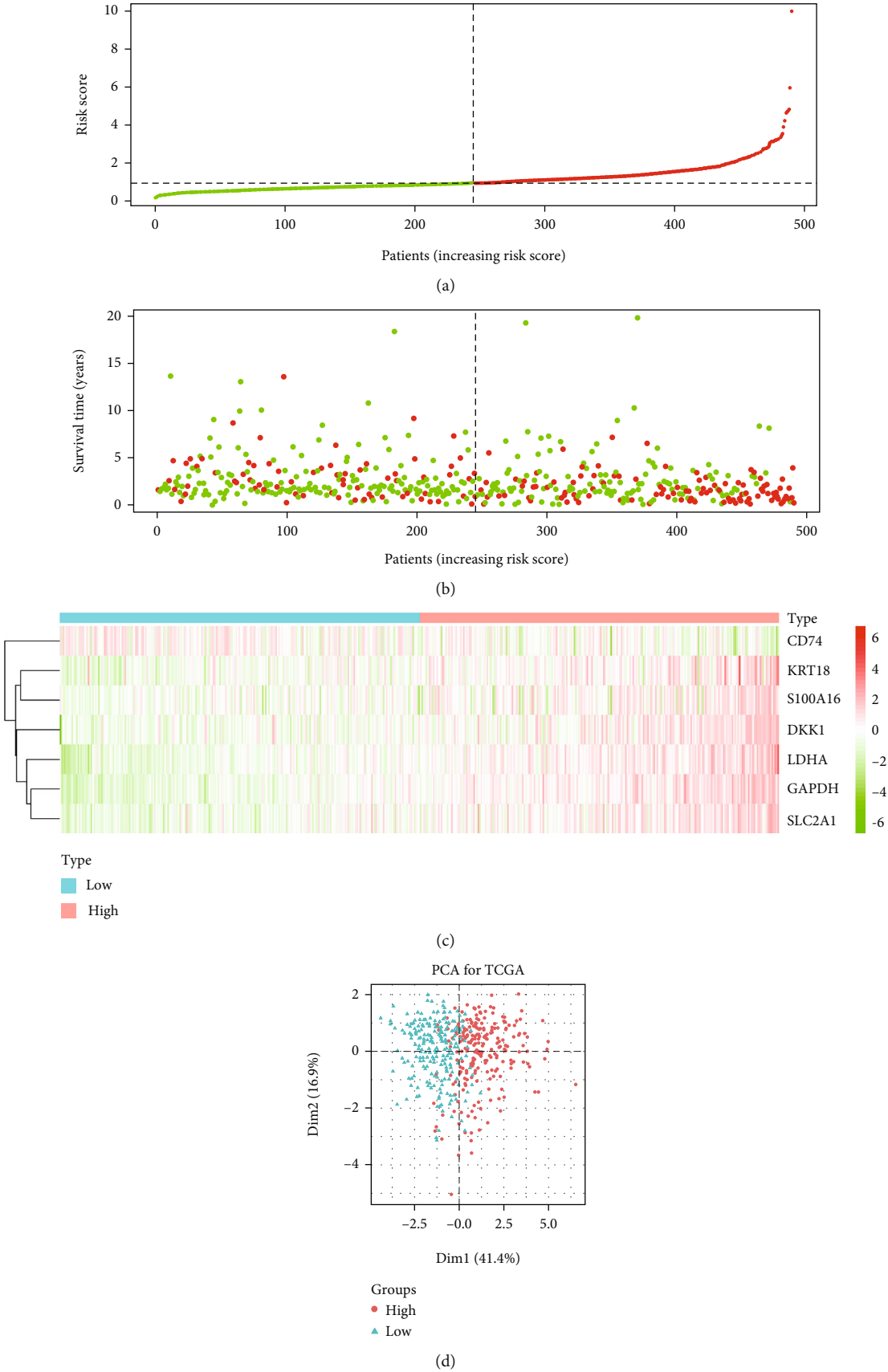


FIGURE 3: Continued.

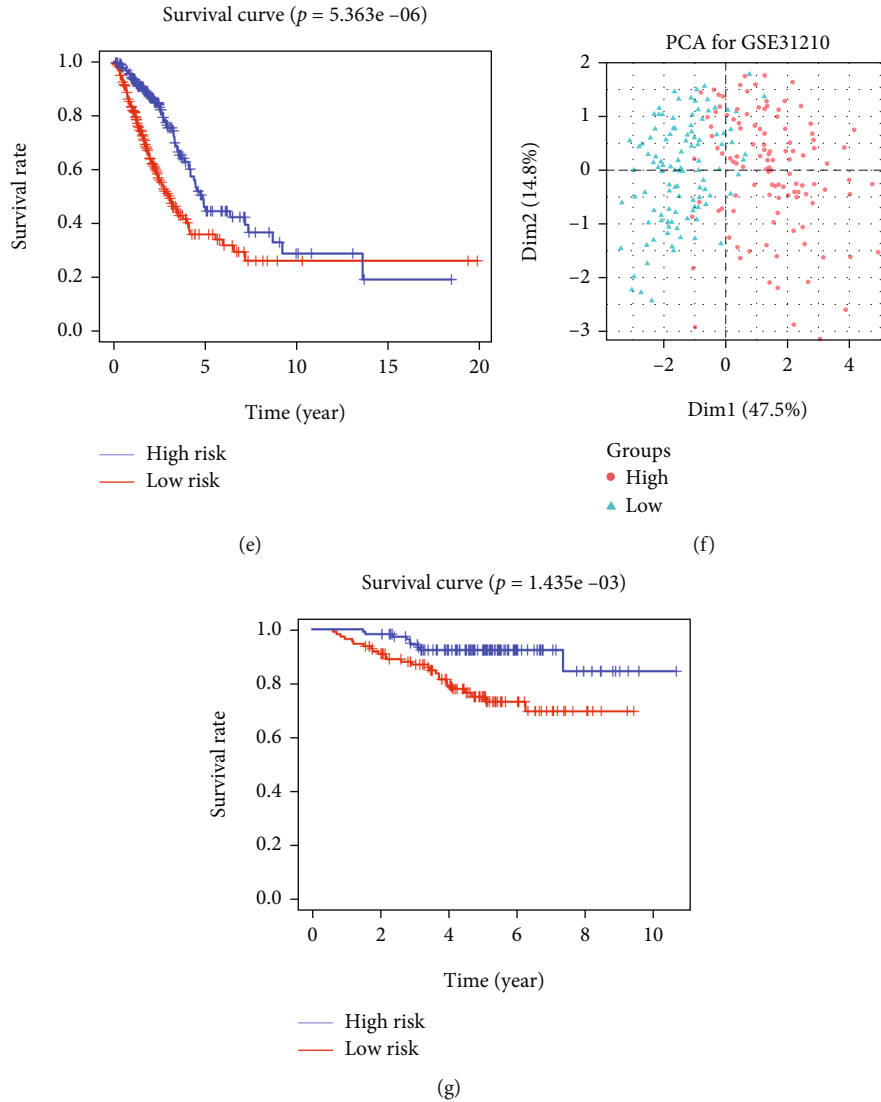


FIGURE 3: The evaluation of the prediction performance of the 7-gene based risk-assessing signature. (a) RiskScore distribution map of LUAD samples, where green represents low risk, and red represents high risk. (b) Scatter plot manifests the distribution of survival status in the high- and low-risk groups, where green represents survival samples, and red represents dead samples. (c) Heat map showing the expression of 7 characteristic genes, where green means low expression, and red means high expression. (d) PCA cluster map of test set and (f) validation set in the high- (red) and low-risk (cyan) groups. (e) The test set and (g) the K-M survival curve of the high- (red) and low-risk (blue) groups in the validation set.

time, a heat map indicated that with the increase of the riskScore, the expression of DKK1, GAPDH, KRT18, LDHA, S100A16, and SLC2A1 increased, while the expression of CD74 decreased (Figure 3(c)). The PCA results suggested that patients in different groups could be clustered clearly according to the defaults (Figure 3(d)). K-M survival curve unveiled that the prognosis of patients with low riskScore was often better than those with high one (Figure 3(e)). PCA and survival analysis were performed on both groups with GSE31210 as the validation set. The results of the validation set were similar to the test set (Figures 3(f) and 3(g)).

Based on ROC curves, the AUC values of 1-, 3-, and 5-year OS were 0.74, 0.68, and 0.62, respectively (Figure 4(a)). The AUC values of 1-, 3-, and 5-year OS in the validation set was 0.81, 0.67, and 0.7, respectively (Figure 4(b)). The above results

revealed that the signature-based riskScore predicted the prognosis of LUAD patients robustly.

In order to explore whether the riskScore could independently assess the prognosis of LUAD patients, we combined the clinical data of the patients, including age, sex, and tumor stage, to perform the univariate Cox regression analysis. The results revealed that the riskScore and the traditional clinical prognostic factors T stage, N stage, and clinical stage were prominently correlated with patients' OS (Figure 5(a)). Multivariate Cox regression analysis of the above results suggested that the riskScore and clinical stage were notably correlated with patients' OS (Figure 5(b)). The riskScore according to the signature were independent enough to be the prognostic factor of LUAD.

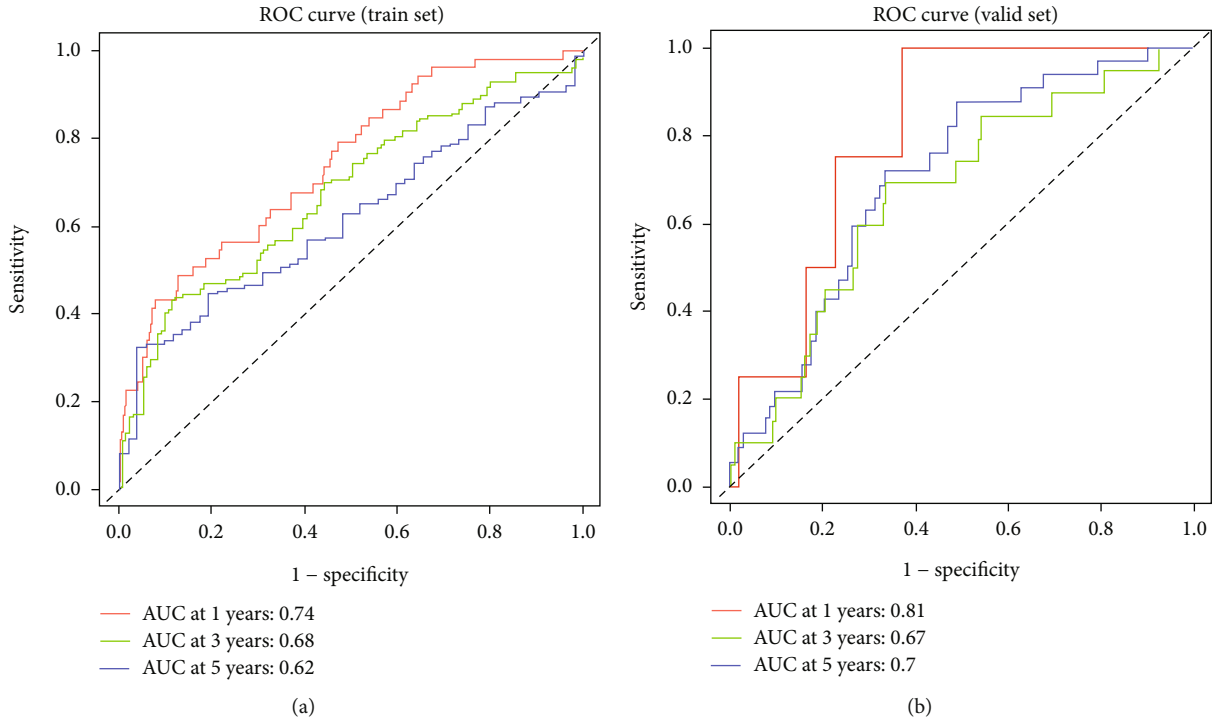


FIGURE 4: The evaluation of 7-gene based risk-assessing signature. (a, b) ROC curves of the test and validation sets.

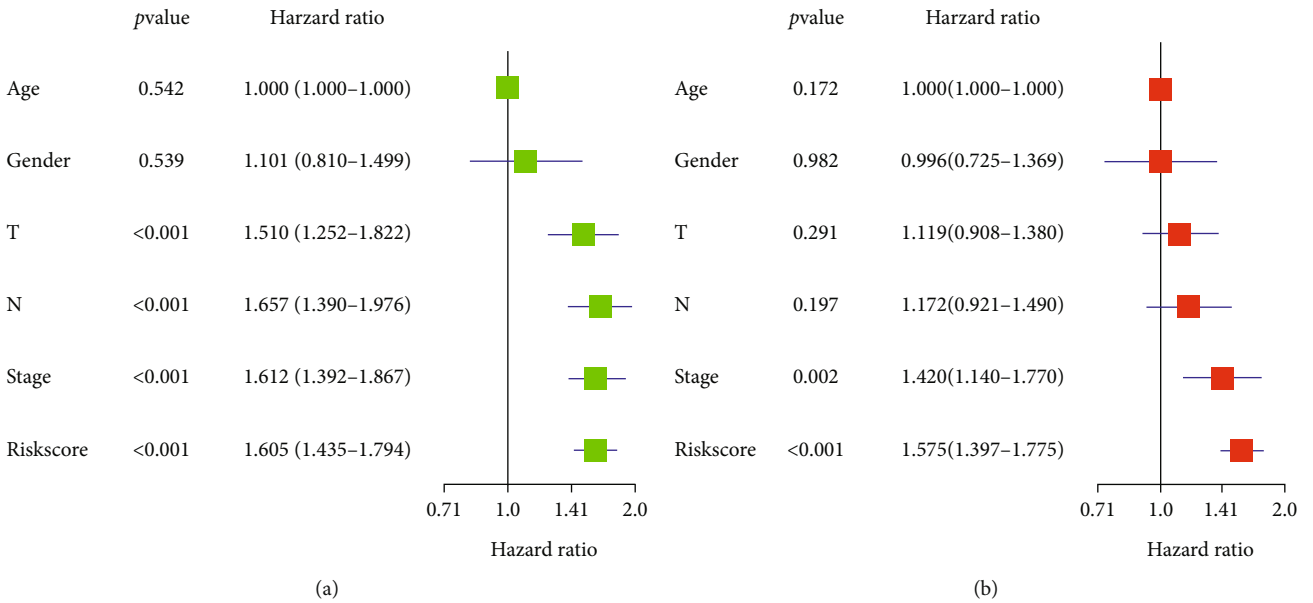


FIGURE 5: The Cox regression analysis of factors relating to patient’s prognosis. (a, b) The results of the univariate and multivariate Cox regression analyses.

3.4. *GSEA of the High- and Low-Risk Groups.* GSEA software was utilized to explore the reason for riskScore divergency. The results suggested that the expression of these 7 genes in the high-risk group was remarkably enriched in p53 signaling pathway, cell cycle, DNA replication, pyrimidine metabolism, glycolysis and gluconeogenesis, and glyoxylate and dicarboxylate metabolism (Figure 6). These pathways were all related to cancer progression, indicating a reliable performance of the 7-gene prognostic risk model.

3.5. *Analysis of Mutation Characteristics of Genes in LUAD.* With the increasing application of immunotherapy in treating LUAD, improving the efficiency of immunotherapy has drawn considerable attention. At present, studies have confirmed that TMB works as a biomarker to verify the effectiveness of immunotherapy. To explore the relationship between riskScore and TMB in LUAD, we analyzed the genes in different risk groups and found that there were differences in TMB between the high- and low-risk groups (Figures 7(a) and 7(b)).

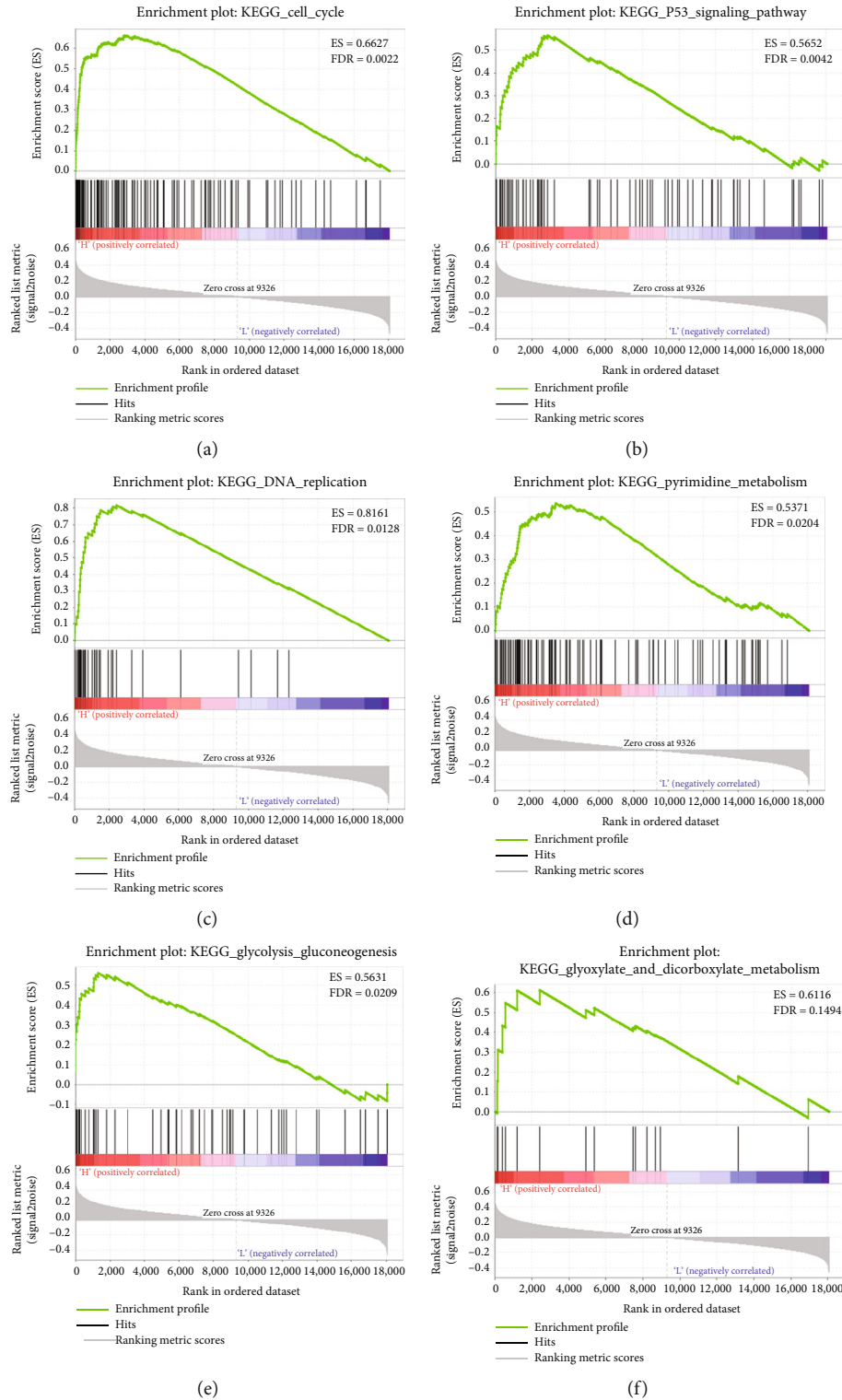


FIGURE 6: The 7 genes in the high- and low-risk groups were enriched in (a) cell cycle; (b) p53 signaling pathway; (c) DNA replication; (d) pyrimidine metabolism; (e) glycolysis and gluconeogenesis; and (f) glyoxylate and dicarboxylate metabolism.

Subsequently, the Wilcoxon test was performed on the TMB values, and the results indicated a remarkable high TMB in the high-risk group (Figure 7(c)). Hence, we speculated that changes in gene hypoxia and immune characteristics might affect the frequency of gene mutations.

3.6. Evaluation of Immune Infiltration. The “estimate” package was utilized to score the stromal cell components and immune cell components of LUAD tumor samples. Tumor purity was calculated and differentially analyzed. The results indicated that the immune scores and estimate scores of the

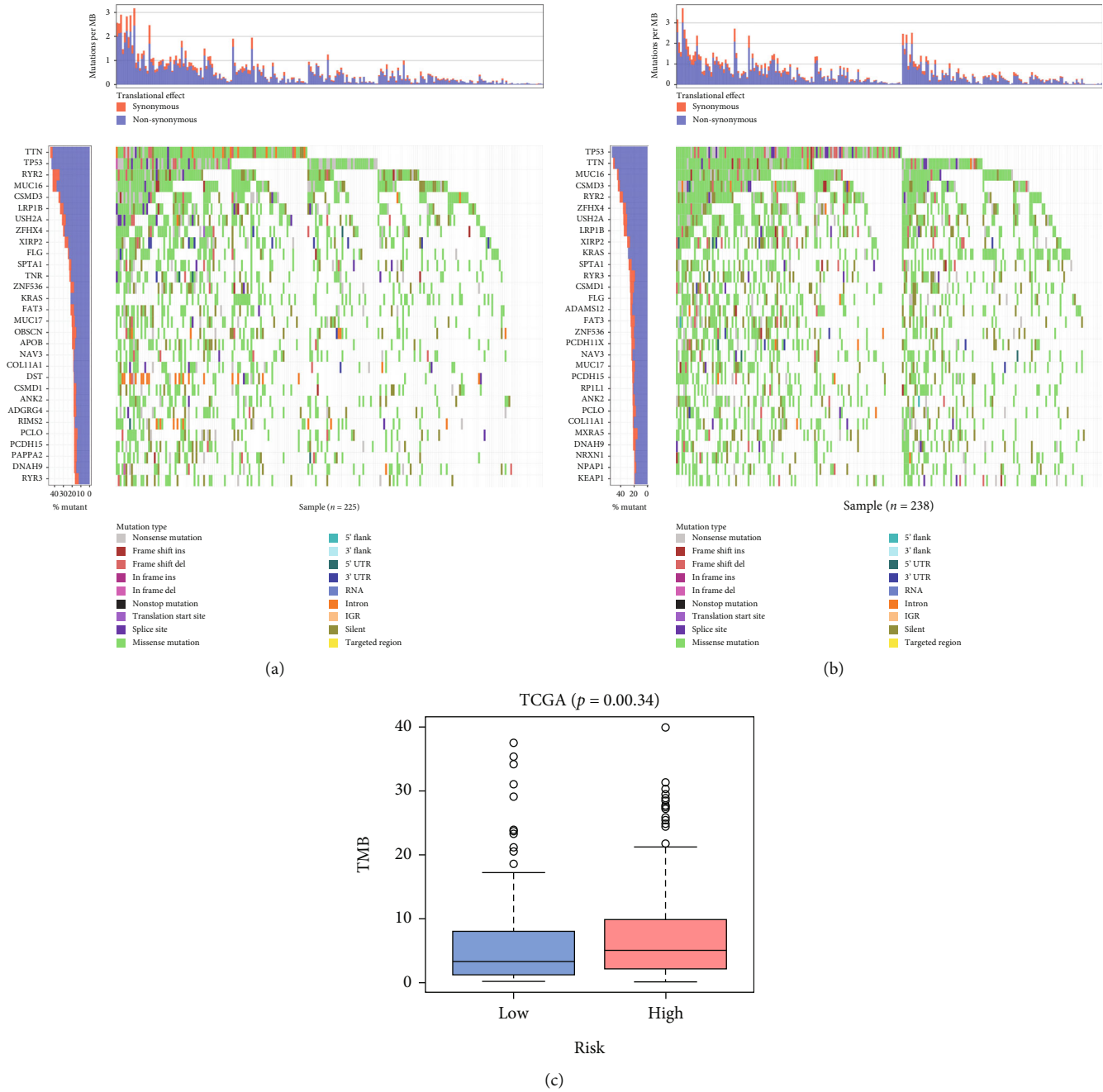


FIGURE 7: Mutation characteristics of genes in LUAD patients. (a, b) Waterfall chart of top30 genes in the low- and high-risk groups, with the horizontal axis represents different samples, while the vertical axis indicates the gene. The filled means that the gene is mutated. Different colors represent different mutation types. (c) Differences in TMB in the high- and low-risk groups.

high-risk group were significantly lower than those of the low-risk group (Figure 8(a)). The results of ssGSEA analysis revealed that compared with the high-risk group, the low-risk group had higher levels of immune infiltration in aDCs, iDCs, mast cells, B cells, neutrophils, TIL, and T helper cells (Figure 8(b)). While the immune activities in HLA, type II IFN response, and T cell costimulation were higher in the low-risk group (Figure 8(c)). On the above, we concluded relatively low immune scores of patients in the high-risk group, which may impact the immunotherapy of LUAD.

4. Discussion

With the increasing development of high-throughput technology and research in LUAD, the treatment of LUAD has improved, especially in immunotherapy. However, the therapeutic effect and patient’s prognosis are still not optimistic. In recent years, biomarkers have been proven to guide patients’ diagnosis, treatment, and prognosis prediction. Based on the TCGA database, this study analyzed the hypoxia and immune characteristics of LUAD and combined

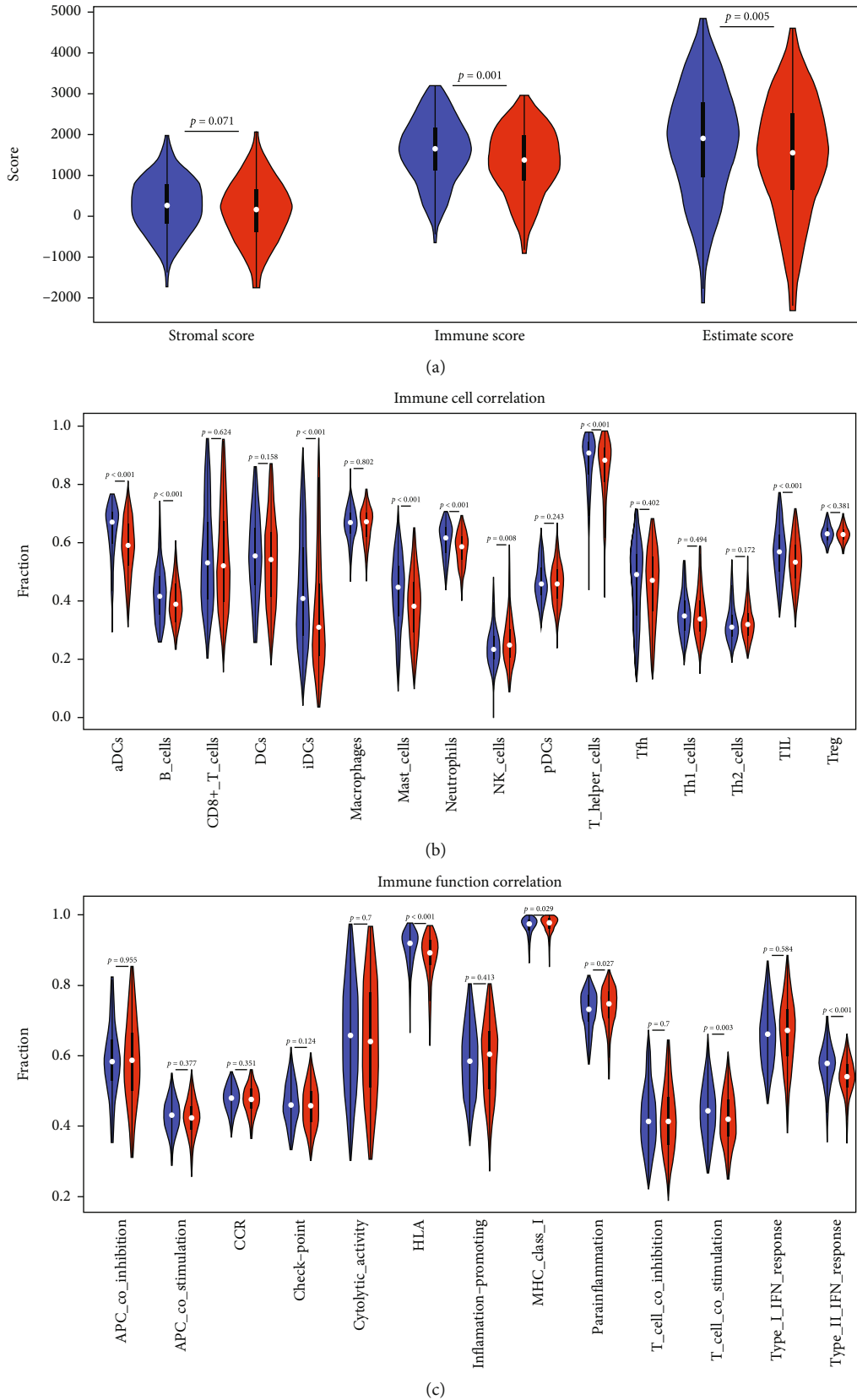


FIGURE 8: Immune infiltration assessment results. (a) Differential analysis of stromal scores, immune scores, and estimate scores. (b) Differential analysis of immune cell components. (c) Differential analysis of immune functions. Blue represents the low-risk group, and red represents the high-risk group.

the two to build a prognostic-assessment signature based on 7 genes for LUAD patients.

The seven genes for constructing a risk-assessing signature for prognosis were as follows: CD74, DKK1, GAPDH, KRT18, LDHA, S100A16, and SLC2A1, among which the four genes CD74, DKK1, GAPDH, and S100A16 were related to immunity. S100A16 is a member of the S100A family. Studies have found that S100A16 can inhibit the immune infiltration of CD8⁺ T cells through the focal adhesion-Ras-stimulated signaling pathway in pancreatic cancer [35]. The study of Ou et al. [36] found that S100A16 inhibits activities of CRC cells through the JNK/p38 MAPK signaling pathway and subsequent EMT. DKK1 encodes secreted proteins and regulates protein-protein interactions [37]. Especially in myeloma cells, DKK1 was secreted to inhibit function of osteocytes [37]. Some scholars found that multiple myeloma cells can phosphorylate cAMP-responsive element-binding protein (CREB) through p38 kinase under hypoxic conditions and drive CREB into the nucleus to activate DKK1 transcription [38]. CD74 is a type II transmembrane glycoprotein and becomes the proinflammatory cytokine macrophage migration inhibitor in an inflammatory environment [39]. Studies have found that in pancreatic cancer, CD74 expression is related to perineural infiltration and the poor prognosis of patients after surgical resection [40]. Although GAPDH can be stably expressed as a housekeeping protein most of the time, there have been studies taking GAPDH as a prognostic gene and a potential therapeutic target for certain cancers [41, 42]. Thus, the immune-related genes tested in this study were closely related to the prognosis and immunotherapy.

KRT18, LDHA, and SLC2A1 were three characteristic genes relating to hypoxia. Keratin 18 is one of cytoskeletal proteins and functions in various cancers. Works found that KRT18 facilitated the progression of gastric cancer and is relating to the prognosis of GC patients [43]. LDHA-encoded proteins are involved in the last step of anaerobic glycolysis and catalyze the transformation of L-lactic acid and NAD. Studies found that the accumulation of LDHA-related lactic acid in melanoma can repress the function and activity of T cells and NK cells, leading to immune escape of tumors [44]. SLC2A1 encodes the unidirectional protein GLUT-1 and involves in glucose transport. Studies have uncovered that GLUT-1 is related to the cell proliferation of pancreatic cancer cells and is an important regulator in the prognosis of patients with pancreatic cancer [45, 46]. On the above, the selected 7 genes in this study were associated with hypoxia, the tumor microenvironment, and patient's prognosis.

With the development of immunotherapy, people have paid more and more attention to what characteristics of patients could benefit from immunotherapy. Current studies have confirmed that TMB can be a biomarker for immunotherapy, and immunotherapy is quite effective in patients with high TMB [47]. The evaluation of TMB and tumor immune infiltration is important in evaluating whether the patient receives immunotherapy and the efficiency of immunotherapy. Zhang et al. [48] analyzed the head and neck

squamous cell carcinoma and found that patient's immunotherapy is related to the patient's TMB and tumor immune infiltration, and the ICI score constructed based on tumor immune infiltration is a predictor of immunotherapy free from TMB. The study by Kang et al. [49] confirmed that in melanoma, the relationship between TMB and immune infiltration, especially the abundance of macrophages and Tregs, could lead the prediction signature of immunotherapy response. This study combined the riskScore with the TMB and immune infiltration assessment of LUAD patients and discovered high TMB scores and low immune scores in the high-risk group. Combined with existing studies, this study inferred that patients with a high riskScore may reflect a relatively high TMB value and a relatively low immune score, which is a proper condition for immunotherapy, thereby producing a better prognosis.

However, this study also has certain limitations. First, genetic data used in this study were based only on public databases. The risk-assessing signature for the prognosis of LUAD patients was constructed on the basis of the hypoxia- or immune-related DEGs in LUAD patients. Secondly, although TMB and immune infiltration assessment analysis were carried out on LUAD patients, subsequent clinical trials were required to verify the results of bioinformatics analysis.

In summary, we generated a risk-assessing signature for the prognosis of LUAD patients based on 7 hypoxia- and immune-related genes via the LASSO and the Cox regression analysis. It predicted patient's prognosis robustly, and the riskScore could be regarded as an important prognostic assessment factor independent of clinical characteristics. At the same time, the hypoxia- and immune-related genes of the constructed signature were likely to be potential targets of LUAD treatment, which provided reference to determining the prognosis and making clinical treatment plans for LUAD patients.

Data Availability

All data generated or analyzed during this study are included in this article.

Conflicts of Interest

The authors declare no conflict of interest that may be relevant to the submitted manuscript.

Authors' Contributions

Yue Wang was responsible for the conception and design and for the provision of study materials or patients. Jian Feng was responsible for the administrative support. Yang Liu was responsible for the collection and assembly of data. Tingyue Han was responsible for the data analysis and interpretation. All authors contributed to the manuscript writing. All authors gave their final approval of the manuscript.

Supplementary Materials

Table 1: Details of 268 prognostic-related DEGs. (*Supplementary Materials*)

References

- [1] H. Sung, J. Ferlay, R. L. Siegel et al., “Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [2] H. Asamura, T. Goya, Y. Koshiishi et al., “A Japanese lung cancer registry study: prognosis of 13,010 resected lung cancers,” *Journal of Thoracic Oncology*, vol. 3, no. 1, pp. 46–52, 2008.
- [3] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [4] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2017,” *CA: A Cancer Journal for Clinicians*, vol. 67, no. 1, pp. 7–30, 2017.
- [5] J. R. Molina, P. Yang, S. D. Cassivi, S. E. Schild, and A. A. Adjei, “Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship,” *Mayo Clinic Proceedings*, vol. 83, no. 5, pp. 584–594, 2008.
- [6] D. Kung-Chun Chiu, A. Pui-Wah Tse, C. T. Law et al., “Hypoxia regulates the mitochondrial activity of hepatocellular carcinoma cells through HIF/HEY1/PINK1 pathway,” *Cell Death & Disease*, vol. 10, no. 12, p. 934, 2019.
- [7] K. Qureshi-Baig, D. Kuhn, E. Viry et al., “Hypoxia-induced autophagy drives colorectal cancer initiation and progression by activating the PRKC/PKC-EZR (ezrin) pathway,” *Autophagy*, vol. 16, no. 8, pp. 1436–1452, 2020.
- [8] Q. Zhang, J. Zhang, Z. Fu et al., “Hypoxia-induced microRNA-10b-3p promotes esophageal squamous cell carcinoma growth and metastasis by targeting TSGA10,” *Aging (Albany NY)*, vol. 11, no. 22, pp. 10374–10384, 2019.
- [9] J. C. Walsh, A. Lebedev, E. Aten, K. Madsen, L. Marciano, and H. C. Kolb, “The clinical importance of assessing tumor hypoxia: relationship of tumor hypoxia to prognosis and therapeutic opportunities,” *Antioxidants & Redox Signaling*, vol. 21, no. 10, pp. 1516–1554, 2014.
- [10] X. Jing, F. Yang, C. Shao et al., “Role of hypoxia in cancer therapy by regulating the tumor microenvironment,” *Molecular Cancer*, vol. 18, no. 1, p. 157, 2019.
- [11] P. Vaupel, “Hypoxia and aggressive tumor phenotype: implications for therapy and prognosis,” *The Oncologist*, vol. 13, Suppl 3, pp. 21–26, 2008.
- [12] Y. Jiang, J. Zhou, D. Zou et al., “Overexpression of limb-bud and heart (LBH) promotes angiogenesis in human glioma via VEGFA-mediated ERK signalling under hypoxia,” *eBioMedicine*, vol. 48, pp. 36–48, 2019.
- [13] L. Moreno Leon, M. Gautier, R. Allan et al., “The nuclear hypoxia-regulated NLUCAT1 long non-coding RNA contributes to an aggressive phenotype in lung adenocarcinoma through regulation of oxidative stress,” *Oncogene*, vol. 38, no. 46, pp. 7146–7165, 2019.
- [14] X. Cao, X. Fang, W. S. Malik et al., “TRB3 interacts with ERK and JNK and contributes to the proliferation, apoptosis, and migration of lung adenocarcinoma cells,” *Journal of Cellular Physiology*, vol. 235, no. 1, pp. 538–547, 2020.
- [15] J. Couzin-Frankel, “Breakthrough of the year 2013. Cancer immunotherapy,” *Science*, vol. 342, no. 6165, pp. 1432–1433, 2013.
- [16] G. Multhoff and P. Vaupel, “Hypoxia compromises anti-cancer immune responses,” *Advances in Experimental Medicine and Biology*, vol. 1232, pp. 131–143, 2020.
- [17] Y. Lu, W. Xu, Y. Gu et al., “Non-small cell lung cancer cells modulate the development of human CD1c(+) conventional dendritic cell subsets mediated by CD103 and CD205,” *Frontiers in Immunology*, vol. 10, p. 2829, 2019.
- [18] A. P. Ganesan, M. Johansson, B. Ruffell et al., “Tumor-infiltrating regulatory T cells inhibit endogenous cytotoxic T cell responses to lung adenocarcinoma,” *Journal of Immunology*, vol. 191, no. 4, pp. 2009–2017, 2013.
- [19] L. Song, M. C. Yang, J. Knoff, Z. Y. Sun, T. C. Wu, and C. F. Hung, “Cancer immunotherapy using a potent immunodominant CTL epitope,” *Vaccine*, vol. 32, no. 46, pp. 6039–6048, 2014.
- [20] M. Z. Noman, S. Buart, J. van Pelt et al., “The cooperative induction of hypoxia-inducible factor-1 alpha and STAT3 during hypoxia induced an impairment of tumor susceptibility to CTL-mediated cell lysis,” *Journal of Immunology*, vol. 182, no. 6, pp. 3510–3521, 2009.
- [21] M. Z. Noman, B. Janji, B. Kaminska et al., “Blocking hypoxia-induced autophagy in tumors restores cytotoxic T-cell activity and promotes regression,” *Cancer Research*, vol. 71, no. 18, pp. 5976–5986, 2011.
- [22] A. M. Johnson, B. L. Bullock, A. J. Neuwelt et al., “Cancer cell-intrinsic expression of MHC class II regulates the immune microenvironment and response to anti-PD-1 therapy in lung adenocarcinoma,” *Journal of Immunology*, vol. 204, no. 8, pp. 2295–2307, 2020.
- [23] J. M. Brooks, A. N. Menezes, M. Ibrahim et al., “Development and validation of a combined hypoxia and immune prognostic classifier for head and neck cancer,” *Clinical Cancer Research*, vol. 25, no. 17, pp. 5315–5328, 2019.
- [24] Y. Gu, P. Li, F. Peng et al., “Autophagy-related prognostic signature for breast cancer,” *Molecular Carcinogenesis*, vol. 55, no. 3, pp. 292–299, 2016.
- [25] W. Chen, M. Ou, D. Tang, Y. Dai, and W. Du, “Identification and validation of immune-related gene prognostic signature for hepatocellular carcinoma,” *Journal of Immunology Research*, vol. 2020, Article ID 5494858, 14 pages, 2020.
- [26] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, and P. Tamayo, “The molecular signatures database hallmark gene set collection,” *Cell Systems*, vol. 1, no. 6, pp. 417–425, 2015.
- [27] S. Bhattacharya, S. Andorf, L. Gomes et al., “ImmPort: disseminating data to the public for the future of immunology,” *Immunologic Research*, vol. 58, no. 2-3, pp. 234–239, 2014.
- [28] Ø. Borgan, *Modeling Survival Data: Extending the Cox Model*, T. M. Therneau and M. Patricia, Eds., vol. 20, Grambsch, Springer-Verlag, New York, 2001.
- [29] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
- [30] I. Lambert, C. Paysant-Le Roux, S. Colella, and M. L. Martin-Magniette, “DiCoExpress: a tool to process multifactorial RNAseq experiments from quality controls to co-expression analysis through differential analysis based on contrasts inside GLM models,” *Plant Methods*, vol. 16, no. 1, p. 68, 2020.

- [31] P. Blanche, J. F. Dartigues, and H. Jacqmin-Gadda, "Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks," *Statistics in Medicine*, vol. 32, no. 30, pp. 5381–5397, 2013.
- [32] Z. L. Skidmore, A. H. Wagner, R. Lesurf et al., "GenVisR: genomic visualizations in R," *Bioinformatics*, vol. 32, no. 19, pp. 3012–3014, 2016.
- [33] K. Yoshihara, M. Shahmoradgoli, E. Martínez et al., "Inferring tumour purity and stromal and immune cell admixture from expression data," *Nature Communications*, vol. 4, no. 1, p. 2612, 2013.
- [34] S. Hanzelmann, R. Castelo, and J. Guinney, "GSVA: gene set variation analysis for microarray and RNA-seq data," *BMC Bioinformatics*, vol. 14, no. 1, p. 7, 2013.
- [35] H. Zhuang, X. Chen, F. Dong et al., "Prognostic values and immune suppression of the S100A family in pancreatic cancer," *Journal of Cellular and Molecular Medicine*, vol. 25, no. 6, pp. 3006–3018, 2021.
- [36] S. Ou, Y. Liao, J. Shi et al., "S100A16 suppresses the proliferation, migration and invasion of colorectal cancer cells in part via the JNK/p38 MAPK pathway," *Molecular Medicine Reports*, vol. 23, no. 2, 2020.
- [37] O. Sezer, "Myeloma bone disease," *Hematology*, vol. 10, no. -sup1, pp. 19–24, 2005.
- [38] Y. Xu, J. Guo, J. Liu et al., "Hypoxia-induced CREB cooperates MMSET to modify chromatin and promote DKK1 expression in multiple myeloma," *Oncogene*, vol. 40, no. 7, pp. 1231–1241, 2021.
- [39] F. Borghese and F. I. Clanchy, "CD74: an emerging opportunity as a therapeutic target in cancer and autoimmune disease," *Expert Opinion on Therapeutic Targets*, vol. 15, no. 3, pp. 237–251, 2011.
- [40] J. F. Zhang, R. Hua, D. J. Liu, W. Liu, Y. M. Huo, and Y. W. Sun, "Effect of CD74 on the prognosis of patients with resectable pancreatic cancer," *Hepatobiliary & Pancreatic Diseases International*, vol. 13, no. 1, pp. 81–86, 2014.
- [41] S. H. Kim, W. S. Park, S. J. Lee et al., "The quantified level of circulating prostate stem cell antigen mRNA relative to GAPDH level is a clinically significant indicator for predicting biochemical recurrence in prostate cancer patients after radical prostatectomy," *BioMed Research International*, vol. 2015, Article ID 292454, 2015.
- [42] G. S. Krasnov, A. A. Dmitriev, A. V. Snezhkina, and A. V. Kudryavtseva, "Deregulation of glycolysis in cancer: glyceraldehyde-3-phosphate dehydrogenase as a therapeutic target," *Expert Opinion on Therapeutic Targets*, vol. 17, no. 6, pp. 681–693, 2013.
- [43] P. B. Wang, Y. Chen, G. R. Ding, H. W. Du, and H. Y. Fan, "Keratin 18 induces proliferation, migration, and invasion in gastric cancer via the MAPK signalling pathway," *Clinical and Experimental Pharmacology & Physiology*, vol. 48, no. 1, pp. 147–156, 2021.
- [44] A. Brand, K. Singer, G. E. Koehl et al., "LDHA-associated lactic acid production blunts tumor immunosurveillance by T and NK cells," *Cell Metabolism*, vol. 24, no. 5, pp. 657–671, 2016.
- [45] Y. Cheng, K. Wang, L. Geng et al., "Identification of candidate diagnostic and prognostic biomarkers for pancreatic carcinoma," *eBioMedicine*, vol. 40, pp. 382–393, 2019.
- [46] M. Yu, Q. Zhou, Y. Zhou et al., "Metabolic phenotypes in pancreatic cancer," *PLoS One*, vol. 10, no. 2, article e0115153, 2015.
- [47] T. A. Chan, M. Yarchoan, E. Jaffee et al., "Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic," *Annals of Oncology*, vol. 30, no. 1, pp. 44–56, 2019.
- [48] X. Zhang, M. Shi, T. Chen, and B. Zhang, "Characterization of the immune cell infiltration landscape in head and neck squamous cell carcinoma to aid immunotherapy," *Molecular Therapy-Nucleic Acids*, vol. 22, pp. 298–309, 2020.
- [49] K. Kang, F. Xie, J. Mao, Y. Bai, and X. Wang, "Significance of tumor mutation burden in immune infiltration and prognosis in cutaneous melanoma," *Frontiers in Oncology*, vol. 10, article 573141, 2020.

Research Article

Correlation of Pollen Concentration and Meteorological Factors with Medical Condition of Allergic Rhinitis in Shenyang Area

Feifei Jiang  and Aihui Yan 

Department of Otorhinolaryngology, The First Hospital of China Medical University, Shenyang, 110001 Liaoning, China

Correspondence should be addressed to Aihui Yan; yanaihui123@126.com

Received 14 July 2022; Revised 24 August 2022; Accepted 30 August 2022; Published 27 September 2022

Academic Editor: Tao Huang

Copyright © 2022 Feifei Jiang and Aihui Yan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. The pathogenesis of allergic rhinitis (AR) was affected by meteorological and environmental factors. This study investigated the association between clinical symptoms of AR patients with pollen dispersal and meteorological conditions. **Methods.** The clinical features of 10,838 AR patients who were treated in the Department of Otorhinolaryngology, The First Hospital of China Medical University, Shenyang, from March 2021 to October 2021 were retrospectively analyzed. We collected pollen by a pollen collector, read and counted it under a microscope, identified the species of the pollen particles, and recorded meteorological data (average daily temperature, maximum and minimum temperature, average daily wind, average daily precipitation, average daily humidity, average pressure, air quality index, PM_{2.5}, PM₁₀, SO₂, NO₂, CO, and O₃), to analyze the correlation among meteorological conditions, pollen dispersal, and number of AR visits. Finally, pollen allergen-positive and symptoms were scored. **Results.** Among the AR visits, patients >41 years old accounted for the highest proportion (64.15%). 43.67% of the patients were complicated with bronchial asthma, and the disease incidence peaked in September. During the period of the study, a total of 27,512 pollen grains were collected, and 17 species were identified. The pollens of Compositae and Moraceae were the main allergenic sources leading to the increase in AR visits from August to September. The peak of pollen dispersal was in spring, summer, and autumn. The total amount of pollen was not only related to the average daily minimum temperature, average daily precipitation, and average daily humidity but also had a significant correlation with air quality index and air pollutants (PM_{2.5} and PM₁₀, SO₂, NO₂, and CO). In addition, there was a significant correlation between the number of daily pollen allergen-positive patients and the pollen concentration of Compositae and Moraceae as well as air pollution components. The clinical symptoms of pollen allergen-positive patients were mainly nasal congestion, red/itchy eyes, and epiphora. **Conclusion.** The peak seasons of pollen dispersal in Shenyang were in spring, summer, and autumn, and the allergenic pollens were mainly Compositae and Moraceae. In addition, AR was substantially correlated with pollen concentration and meteorological factors. This study may help provide early warning information and prevention for AR patients.

1. Introduction

Allergic rhinitis (AR) is a chronic inflammatory disease of the nasal mucosa caused by reactions in subjects sensitized to allergens mediated by immunoglobulin E (IgE) [1]. AR is also recognized as a global health issue. Recently, the prevalence and incidence of AR have been increasing worldwide, especially in developing countries [2, 3]. AR is commonly classified into seasonal AR (SAR) and perennial AR (PAR),

depending on the presence or absence of seasonality and the source of allergens causing its symptoms. SAR has a high incidence in the general population and is an important disease affecting the health of the population. Common clinical features of AR include massive runny nose, sneezing, itchy and congested nose, and occasional itching of the conjunctiva, ears, and throat [4]. Although the symptoms of AR are not life-threatening, they are often disturbing, which affect the quality of work and life and place a heavy burden on

individual health and social economy [5, 6]. In addition, it is worth noting that AR often coexists with other allergic diseases, such as asthma and atopic dermatitis [7, 8]. Epidemiological studies have shown that most patients with asthma are associated with rhinitis, and the presence of rhinitis increases the risk factors for asthma [9]. Recently, the number of AR patients in China has increased year by year due to environmental improvement and greening restoration. Therefore, in-depth research is needed to evaluate the epidemiological characteristics and risk factors of AR.

Complex interactions among genes, epigenetics, environmental factors, and lifestyle lead to the occurrence and progression of AR [10, 11]. Airborne allergens such as dust mites, pollen, fungi, and animal dander are widely studied environmental factors affecting allergic reactions [12]. At present, a significant number of studies has reported the risk factors of AR in different regions of the world. For example, Takahashi et al. [13] found that high levels of stress, asthma, atopic dermatitis, tuberculosis, depression, and thyroid disease all increased the risk of AR in the Korean population. In China, An et al. [14] reported that family history of AR, current place of residence, place of residence in infancy, smoking, home decoration, and pet ownership are important risk factors associated with AR prevalence in the general population of Guangzhou. Yang et al. [15] found that self-reported AR prevalence is associated with sinusitis and asthma in the Kazakh population of Fukang, Xinjiang, China, and was mainly positively correlated with carpet use as well as meat and fruit consumption. In addition, air pollutants such as O_3 , NO_2 , and PM are also closely related to the development and deterioration of asthma and AR [12, 16]. For example, high PM10 and PM2.5 concentration aggravates the subjective symptoms of AR in children in Shanghai [17]. It can be seen from the above studies that the causes of AR are various, and it is of great importance to study the etiology of AR for the formulation of effective intervention measures. The purpose of this study was to analyze the risk factors influencing the incidence of AR in Shenyang area and to explore the correlation of pollen concentration and meteorological conditions with the number of AR patients, so as to provide theoretical basis for clinical prevention and treatment of AR.

2. Materials and Methods

2.1. Subjects. According to the diagnostic criteria, a total of 10,838 AR patients who visited the otolaryngology department of The First Hospital of China Medical University (hereinafter referred to as “our hospital”) from March 1, 2021 to October 18, 2021 were retrospectively analyzed, and the clinical data of the patients were collected from the Electronic Information Center of Hospital. Clinical data of the patients were collected. The inclusion criteria are as follows: patients living in Shenyang for a long time and were diagnosed with AR and patients with no skin scratches, nasal polyps, and chronic sinusitis. The exclusion criteria are as follows: patients with severe eczema and skin abrasions, patients with a history of antihistamine use 20 days prior to enrollment, and patients complicated with mental disor-

ders or autoimmune defects [18]. This study was approved by the Review Committee of our hospital, and the written informed consent was obtained from all participants.

The diagnosis of AR conformed to the clinical diagnosis basis of the “Chinese Guideline for Diagnosis and Treatment of Allergic Rhinitis (2022, Revision)” [19] formulated by the Subspecialty Group of Rhinology, Society of Otorhinolaryngology Head and Neck Surgery, Chinese Medical Association which are as follows:

- (1) Two or more symptoms of paroxysmal sneezing, clear water-like nasal discharge, nasal itching, and nasal congestion, which persist or accumulate for more than 1 h
- (2) The nasal mucosa is pale and edematous, and the nasal cavity secretes watery secretions
- (3) Skin prick test (SPT) detects at least one allergen and/or serum specific IgE-positive, or nasal challenge test-positive

2.2. Meteorological and Pollen Data. Meteorological data was daily maximum and minimum temperatures, wind scale, air quality index, PM2.5, PM10, etc. All data are recorded daily by Shenyang Meteorological Station and reported by Shenyang Meteorological Bureau. The pollen data were collected by the pollen monitoring station near the meteorological monitoring station. The pollen collection condition was kept consistent with the climate condition as far as possible. The pollen station was managed by Department of Otorhinolaryngology, The First Hospital of China Medical University. Pollen grains were collected daily using a Durham sampler, which is commonly used in China and counted on a slide under a microscope. Average daily pollen count was used for analysis [20]. Pollen grains were identified based on the *Color Atlas of Airborne Pollen and Plant in China*, the *Pollen Morphology of Tropical and Subtropical Angiosperms in China* (Pollen Morphology Office of South China Institute of Botany), and the *Pollen Morphology of Plants in China* (second edition) [21].

2.3. Serum Allergen-Specific IgE (SIgE). The ImmunoCAP system (Thermo Fisher Scientific, USA) were adopted for serum allergen-specific IgE (SIgE) for all patients. The antigens included dust mites, animal fur, mold, willow, elm, cypress, Oleaceae, walnut, birch, pine, mulberry, ragweed, mugwort, etc. 5 ml of venous blood was first taken from all patients, and the content of serum SIgE was detected with CAP System RAST FEIA enzyme immunofluorescence detector. According to the WHO 75/502 standard, the measurement range was 0.35~100 ku/L, and the CAP test results were divided into 0-6 grades, of which grade 0 was <0.35 ku/L, level 1 was ≥ 0.35 to <0.70 ku/L, level 2 was ≥ 0.7 to <3.5 ku/L, level 3 was ≥ 3.5 to <17.5 ku/L, level 4 was ≥ 17.5 to <50.0 ku/L, level 5 was ≥ 50.0 to <100 ku/L, and level 6 was ≥ 100 ku/L. It was judged as positive by level 1 (≥ 0.35 ku/L) or above.

Subjective clinical symptoms (nasal itching, sneezing, rhinorrhea, nasal congestion, eye redness/itching, and epiphora) were scored for AR patients as follows: nasal itching

was 0 for no symptoms, 1 for intermittent nasal itching, 2 for tolerable creeper sensation, and 3 for unbearable creeper sensation; sneezing was 0 for the number of sneezes in one consecutive time <3, 1 for 3~5 sneezes, 2 for 6~10 sneezes, and 3 for ≥ 11 sneezes. Rhinorrhea was 0 for no symptoms, 1 for ≤ 4 times, 2 for 5~9 times, and 3 for ≥ 10 times (number of nasal blows per day); nasal congestion was 0 for no symptoms, 1 for conscious inhalation, 2 for intermittent or interactive congestion, and 3 for breathing through mouth almost all day; eye redness/itching was 0 for no symptoms, 1 for intermittent eye redness/itching, 2 for obvious but tolerable eye redness/itching, and 3 for persistent and unbearable eye redness/itching; and epiphora was 0 for no epiphora, 1 for a sense of tear filling in the conjunctiva sac but no tear overflow at the eyelid margin, 2 for occasional tear overflow at the eyelid margin, and 3 for frequent or almost continuous tear overflow at the eyelid margin accompanied by nose blowing.

2.4. Statistical Analysis. GraphPad software and Excel statistical software were used for all statistical analysis. Counting data were subjected to Pearson correlation analysis to examine the correlation between AR and potential risk factors, and to evaluate the correlation between the daily case number of pollen allergen-positive patients and pollen concentration and air pollutants. $P < 0.05$ was considered to be statistically significant.

3. Results

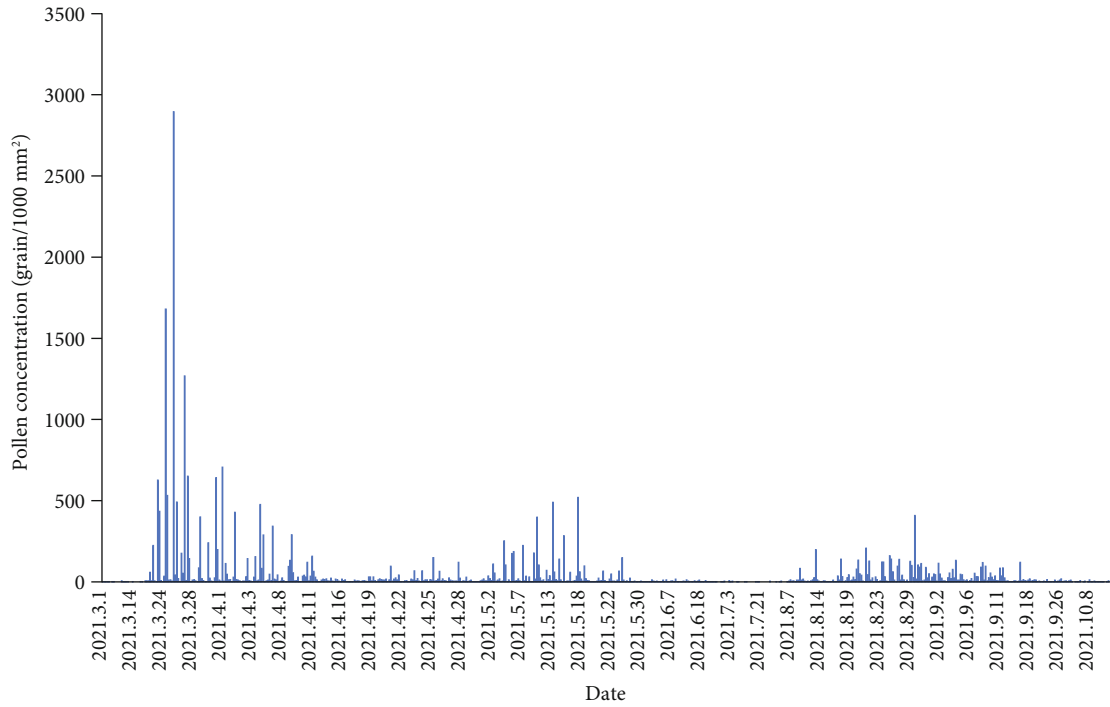
3.1. Epidemiological Characteristics of AR. From March 1, 2021 to October 18, 2021, a total of 10,838 AR patients were collected (Table 1). The average age of patients in the cohort was 42 years old (range: 1–99 years old). Among them, 1,734 patients (16.00%) were younger than 18 years old, 2,151 patients (19.85%) were between 18 and 40 years old, 3,489 patients (32.19%) were between 41 and 60 years old, and the remaining 3,464 patients (31.96%) were older than 60 years old. The number of visits was 5,080 (46.87%) for males and 5,758 (53.13%) for females. In addition, 47.91% of patients smoked or had a smoking history, 68.53% of patients kept pets at home, 73.11% of patients grew flowers indoors, and 55.21% of patients had a family history of AR. The number of AR patients complicated with bronchial asthma was 4,733 (43.67%). Meanwhile, the monthly number of cases was counted, and it was found that September had the most visits (20.09%), followed by April (15.62%), March (14.16%), July (14.13%), August (12.16%), May (7.79%), and October (5.53%). The minimum number of visits in October may be related to the fact that the number of visits was only counted up to the 18th day of the month. On the whole, minors accounted for the least proportion of all AR visits, and the proportion of female visits was slightly higher than that of male visits, and nearly half of AR patients had bronchial asthma.

3.2. Correlation between Pollen Dispersal and AR. During the study period, a total of 230 films were exposed and 27,512 pollen grains were collected, with an average of 3,439 pollen

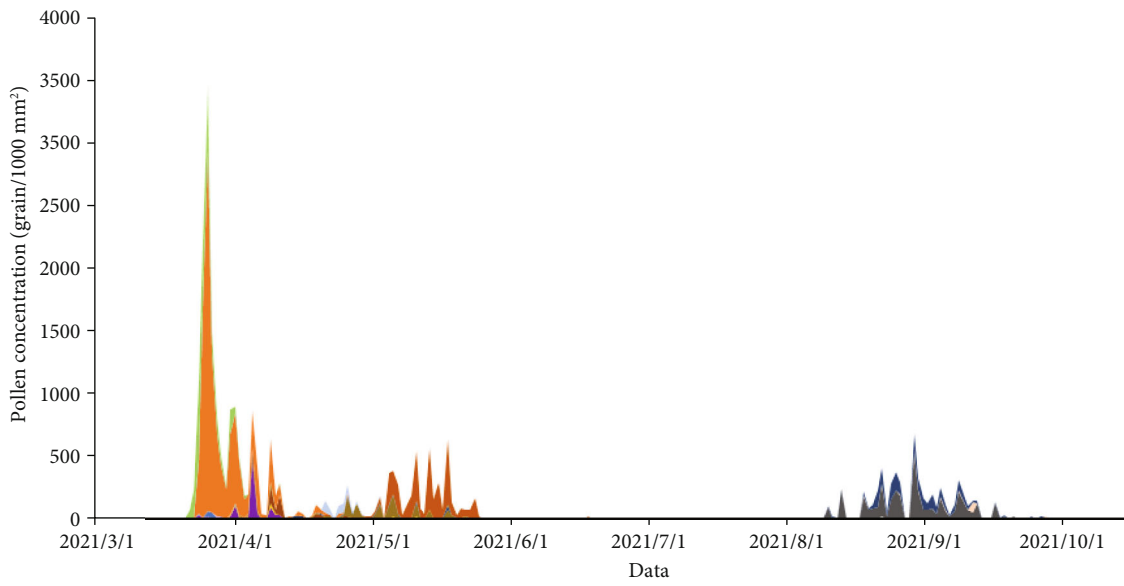
TABLE 1: Clinical features of AR in Shenyang area.

Characteristics	Number (n)	Percentage (%)
Age (years)		
<18	1734	16.00
18–40	2151	19.85
41–60	3489	32.19
>60	3464	31.96
Gender		
Male	5080	46.87
Female	5758	53.13
Smoking habit		
Never smoking	5645	52.08
Now smoking	3550	32.75
Ever smoking	1643	15.16
Keeping pets		
Yes	7427	68.53
No	3411	31.47
Family history of AR		
Yes	5983	55.21
No	4855	44.79
Indoor flower cultivation		
Yes	7923	73.11
No	2915	26.89
Combined diseases		
Bronchial asthma	4733	43.67
Nonbronchial asthma	6105	56.33
Month		
March	1535	14.16
April	1693	15.62
May	844	7.79
June	1140	10.52
July	1531	14.13
August	1318	12.16
September	2177	20.09
October	600	5.53

grains per month. There were three peaks of airborne pollens from March to October, which were from March to April, May, and from August to September. A total of 21,478 pollen grains were collected in spring (March–May), and 6,034 pollen grains in summer and autumn (June–October). The pollen peak period in spring was late March and mid-May, and the pollen peak period in summer and autumn was August (Figure 1(a)). The known pollen species recorded in the exposed film included 17 species, which were Salicaceae, Ulmaceae, Gramineae, Cupressaceae, Ginkgoaceae, Oleaceae, Juglandaceae, Betulaceae, Pinaceae, Leguminosae, Moraceae, Rosaceae, Urticaceae, Cyperaceae, Compositae, Amaranthaceae, and Chenopodiaceae. The fragmented and unidentifiable pollens were classified as others. Among the 17 pollen species, 9 types of dominant pollens were found, among which the dominant pollens in spring were Salicaceae, Pinaceae, Ulmaceae, Cupressaceae,



(a)



- | | |
|-----------------|------------------|
| ■ Paco | ■ Ulmaceae |
| ■ Betulaceae | ■ Gramineae |
| ■ Other | ■ Chenopodiaceae |
| ■ Pinaceae | ■ Sangko |
| ■ Ginkgoceae | ■ Urticaceae |
| ■ Leguminous | ■ Juglandaceae |
| ■ Composite | ■ Oleaceae |
| ■ Rosaceae | ■ Cyperaceae |
| ■ Amaranthaceae | ■ Salicaceae |

(b)

FIGURE 1: Continued.

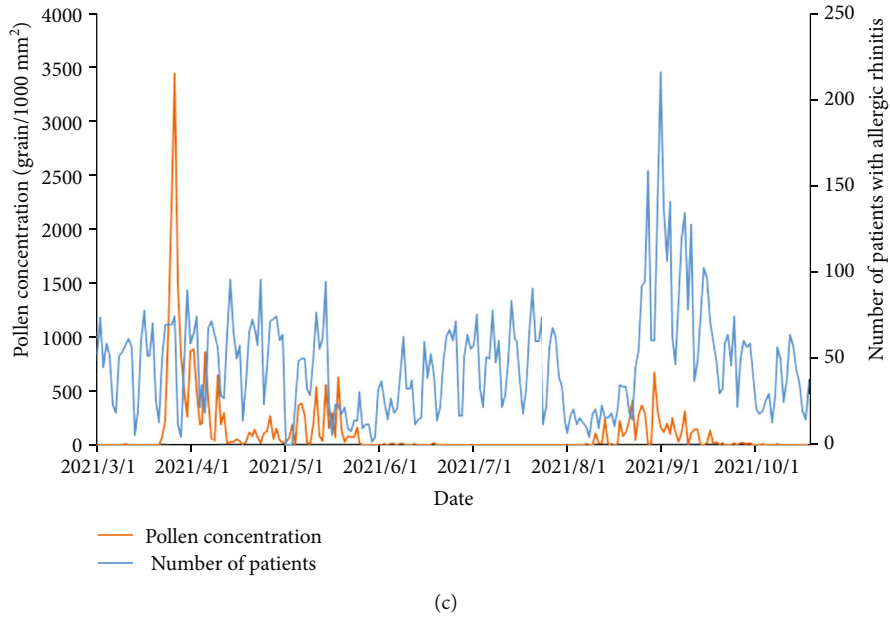


FIGURE 1: Correlation between pollen dispersal and AR. (a) Variation trend of airborne pollen concentration during pollen dispersal period. (b) Distribution of dispersal of dominant pollens. (c) Number of visits for AR and trend of pollen concentration variation.

Rosaceae, Oleaceae, and Ginkgoaceae, and the dominant pollens in summer and autumn were Compositae and Moraceae (Figure 1(b)). Then, the pollen concentration and outpatient visits of AR patients from March to October 2021 were analyzed. It was found that the trend of the number change of AR visits and the trend of pollen concentration change were inconsistent in time, and the number of cases was obviously divided into three periods: from March to mid-May, from June to July, and from September to October. However, we found that the number of visits of AR patients in summer and autumn was significantly higher than that in spring, and the number of cases peaked in September (Figure 1(c)). Due to the significant differences in pollen dispersal in different seasons found above, we conjectured that AR prevalence might be associated with pollen species. Combined with the analysis of pollen species distribution characteristics, it was found that although the pollen concentration of Salicaceae accounted for the highest proportion of all pollens (40.85%), the number of visits of AR patients was not high in the peak period of dissemination of Salicaceae. However, with the increase of pollen concentration of Compositae and Moraceae, the number of visits to AR patients increased significantly in summer and autumn. Subsequently, we conducted correlation analysis on the daily total pollen concentration of Compositae, Moraceae, and Salicaceae and AR outpatient visits (Table 2). We found that both the daily total pollen concentration and Salicaceae pollen concentration had no correlation with AR outpatient visits ($P > 0.05$), while Compositae and Moraceae pollen concentration had a positive correlation with AR outpatient visits ($P < 0.05$, Table 2). Based on the above analysis, Compositae and Moraceae were the main allergenic herbaceous pollens for AR in summer and autumn.

TABLE 2: Correlation analysis of pollen dispersal and allergic rhinitis.

Pollen concentration	Number of patients	
	R ²	P value
Total concentration	0.01388	0.0733
Compositae	0.05165	0.0005
Moraceae	0.1029	<0.0001
Salicaceae	0.003098	0.3987

3.3. Relationship of Meteorological Conditions and Pollen Dispersal with AR. In addition to the internal factors such as the condition of regional vegetation, the law of plant growth and development, and the characteristics of pollen itself, the external meteorological factors also affect the law of airborne pollen dispersal. Meteorological elements (including daily average temperature, daily average minimum or maximum temperature, daily average wind, daily average precipitation, daily average relative humidity, and daily average atmospheric pressure) and air pollution status (including air quality index, PM2.5, PM10, SO₂, NO₂, CO, and O₃) in Shenyang from March 2021 to October 2021 as well as total pollen count and number of AR patients in the same period were enrolled for statistical analysis (Table 3). It could be seen that total pollen count was significantly correlated with daily minimum temperature, average daily precipitation, and average daily relative humidity ($P < 0.05$), but not with daily average temperature, maximum temperature, average atmospheric pressure, and wind scale ($P > 0.05$). In addition, total pollen count had a significant correlation with air quality index, PM2.5, PM10, SO₂, NO₂, and CO ($P < 0.05$), indicating that pollen dispersal

TABLE 3: Correlation analysis of meteorological conditions with pollen dispersal and allergic rhinitis.

Meteorological elements	Pollen concentration		Number of patients	
	R ²	P value	R ²	P value
Average temperature	0.01087	0.1133	0.001782	0.5223
Minimum temperature	0.01789	0.0418	3.605e-005	0.9275
Maximum temperature	0.004501	0.3089	0.006541	0.2197
Average wind	0.007776	0.1807	0.0001071	0.8754
Average daily precipitation	0.244036	<0.0001	0.01344	0.09
Average daily humidity	0.106276	<0.0001	0.014641	0.066
Average pressure	0.000676	0.695	0.013689	0.075
Air quality	0.06817	<0.0001	0.0008591	0.6570
PM 2.5	0.05973	0.0002	0.0006355	0.7025
PM 10	0.1211	<0.0001	0.002178	0.4793
SO ₂	0.1654	<0.0001	0.008125	0.1712
NO ₂	0.08412	<0.0001	0.01568	0.0569
CO	0.05527	0.0003	0.005179	0.2750
O ₃	0.01140	0.1048	0.003550	0.3663

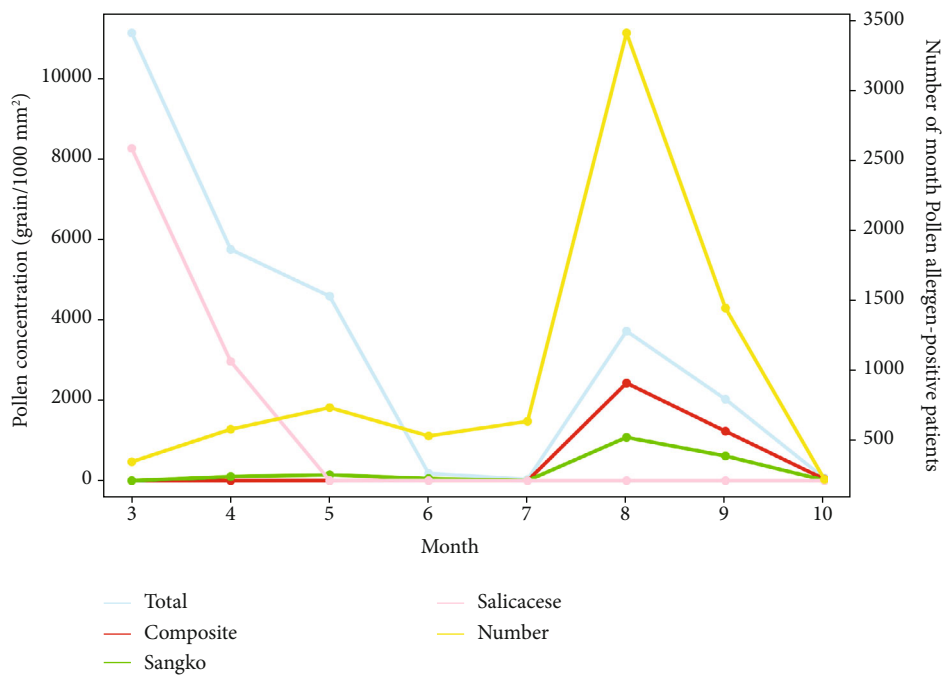


FIGURE 2: Correlation between the number of pollen allergen-positive patients per month and pollen concentration.

was one of the factors contributing to the increase of the concentration of these adverse pollutants in the air. It was worth noting that there was no correlation between the number of AR cases and meteorological factors ($P > 0.05$). Therefore, we concluded that the change of meteorological conditions was one of the factors that make pollen become the main allergen of AR.

3.4. Correlation of the Number of Daily Pollen Allergen-Positive Patients with Air Pollution Components and Pollen Concentration. A total of 7,931 (73.18%) positive patients were detected from March 1, 2021, to October 18, 2021,

including 4,364 males (55.02%) and 3,567 females (44.98%). The number of positive cases per month from March to October were 349, 582, 737, 534, 639, 3418, 1449, and 223, respectively. We conducted statistical analysis on the daily total pollen concentration, pollen concentration of Compositae, Moraceae, and Salicaceae, and the daily number of pollen allergen-positive patients, and the results were shown in Figure 2. The yellow line presents the number of allergen-positive patients per month, which was consistent with the curve of pollen concentration of Compositae and Moraceae but significantly different from the curve trend of total pollen concentration and pollen concentration of Salicaceae. This

further confirmed that Compositae and Moraceae were the main allergenic pollen of AR in Shenyang area. In addition, there was a weak correlation between the daily number of pollen allergen-positive patients and the total daily pollen concentration ($P = 0.043$), while the daily case number was significantly correlated with the pollen concentration of Compositae and Moraceae ($P < 0.001$) but not with the pollen concentration of Salicaceae (Table 4). Meanwhile, correlation analysis was conducted between the concentration of 6 kinds of air pollutants and the daily number of pollen allergen-positive patients, and significant correlations were found between the daily number of patients and the daily concentration of pollutants PM2.5, PM10, SO₂, and NO₂ ($P < 0.05$) (Table 4).

3.5. Pollen Allergen-Positive and Symptom Score. According to the results of pollen allergen test, AR patients were divided into the positive group and the negative group, and the clinical symptoms of the two groups were compared. The results were shown in Table 5. There were statistically significant differences between the positive group and the negative group in nasal congestion, eye red ness/itching, and epiphora symptom scores ($P < 0.05$), while no significant differences were seen in nasal itching, sneezing, and rhinorrhea between the two groups ($P > 0.05$). Additionally, the three symptoms of nasal congestion, eye redness/itching, and epiphora were more pronounced in the pollen allergen-positive group than in the negative group.

4. Discussion

In the past decade, the incidence of AR in China has been on the rise, especially in densely populated cities [22]. However, the reasons for the rapid increase of the incidence of AR are not completely clear at present, and it is urgent to find the reasons for AR increase and make appropriate countermeasures. Although there have been several epidemiological studies on AR in major cities in China, the allergen characteristics of AR patients in different regions and ages in China differ greatly due to the vast territory and large population [23, 24]. Improved sanitation, increased allergen exposure, and changes in dietary habits and lifestyle may increase the prevalence of AR [13, 25]. Therefore, it is of great importance for the prevention and treatment of AR to understand the distribution characteristics of allergens and related risk factors in each region. The present study analyzed the factors influencing the number of AR outpatient visits among the general population in Shenyang, China from March 1 to October 18, 2021. Our study showed that the AR consultation rate of middle-aged people aged 41–60 in Shenyang was the highest, and the gender difference was not significant, but the female rate was slightly higher than the male rate. Similarly, Hong et al. [26] analyzed the prevalence of AR in Guangzhou and Zhuhai in Southern China, and they found that there was no significant gender difference, but the incidence of AR in males was slightly higher than that in females. Therefore, we guessed that gender was not the primary risk factor for AR. In addition, it is worth noting that nearly half of the AR patients in the present study were com-

TABLE 4: Correlation of the number of daily pollen allergen-positive patients with air pollution components and pollen concentration.

Elements	Number of daily pollen allergen-positive patients	
	R ₂	P value
Total concentration	0.017689	0.043* (<0.05)
Compositae	0.429025	<0.001
Moraceae	0.370811	<0.001
Salicaceae	0	0.994
PM2.5	0.0441	0.001* (<0.05)
PM10	0.057121	<0.001
SO ₂	0.26244	0.014* (<0.05)
NO ₂	0.016641	0.04* (<0.05)
CO	0.014641	0.066
O ₃	0.007225	0.199

*Represents $P < 0.05$.

TABLE 5: Pollen allergen-positive grouping and symptom score.

Symptom	Pollen allergen-positive	Pollen allergen-negative
Nasal itching	1.86 ± 0.95	1.69 ± 1.03
Sneezing	2.32 ± 0.93	2.14 ± 0.93
Rhinorrhea	2.15 ± 1.01	1.97 ± 0.97
Nasal obstruction	2.24 ± 0.89	1.9 ± 0.92 ^a
Red eye/itching	1.98 ± 0.94	1.23 ± 0.89 ^a
Epiphora	1.36 ± 0.97	0.86 ± 0.91 ^a

Compared with the pollen allergen-positive group, ^a represents $P < 0.05$.

plicated with bronchial asthma. This is consistent with the results reported in previous studies that asthma and AR are usually comorbidities and coexist in the same patient in China [27]. Previous studies have shown that AR is an independent risk factor for the development of asthma or other dangerous respiratory diseases in individuals [28]. Therefore, knowing more about the risk factors associated with AR can help improve symptoms and quality of life and reduce asthma exacerbation in patients with coexisting asthma and persistent AR.

Pollen has been reported to be an important cause of seasonal AR, and typical pollens involved include wind-pollinated trees and grasses [29]. Recently, with the improvement of the environment, pollen concentration has been increasing year by year in major cities in China, and pollen-related allergens have attracted the attention of many scholars in China. A study investigated the prevalence of AR caused by high pollen exposure in the grasslands of northern China and found that the incidence of AR caused by pollen was very high in the investigated area due to the influence of local environmental and climatic conditions [30]. Zhang et al. [31] found that with the increase of pollen concentration in ambient air, the relative risks on daily number of outpatients of AR were increased. Thus, high pollen level was a risk factor for the development of AR symptoms.

However, there was no report on the effect of airborne pollen on AR visit frequency of Shenyang residents. In this study, it was found that there were two obvious peak periods of pollen distribution in Shenyang area, one in spring (from March to May) and the other in summer and autumn (from June to October). The actual temperature in these two periods could increase the yield and concentration of pollen and enhance the antigenicity of pollen [17, 32]. Interestingly, in this study, there was no linear correlation between the daily total pollen concentration and the daily number of AR patients, and the number of AR patients in spring was significantly lower than that in summer and autumn, which may be due to the fact that the dominant Salicaceae pollen disseminated in spring is not an important source of allergens [33]. On the contrary, the number of AR visits peaked in August and September. During the same period, the pollen concentration of Compositae and Moraceae increased significantly. This is consistent with the results of Bishan et al. [21] that Moraceae, Artemisia in Compositae, and Gramineae are the main allergenic pollen types, with peak concentrations from April to May, from August to September, and from October to December, respectively [21]. Additionally, the positive result of pollen SIgE showed that the number of pollen-allergen positive AR patients during pollen seeding period was significantly positively correlated with the concentration of Compositae and Moraceae pollen in the air, which might be because AR patients allergic to these two kinds of pollen had higher sensitivity to SIgE during pollen seeding period. Therefore, we believed that Compositae and Moraceae were the most important herbaceous plants with allergenic pollen in Shenyang area. It has been reported that pollen allergy is significantly related to the symptom severity of rhinitis patients [34]. By analyzing the clinical symptoms and characteristics of AR patients, it was found that patients with positive pollen allergen had more evident symptoms of nasal congestion, eye redness/itching, and epiphora. It is easy to cause allergic reactions in the mucous membrane when pollen falls down into the eyes or is inhaled through the nose and pharynx. Therefore, patients with positive pollen allergen had more obvious accompanying ocular symptoms and nasal congestion than those in negative patients. The subsequent study will further perfect the symptoms and objective evaluation indicators, and the related causes and pathogenesis will be analyzed.

Since the nature and quantity of pollen varies with vegetation, geography, temperature, and climate, the sensitivity of people in different areas to pollen species is different [35]. Studies have shown that changes in weather conditions such as rainfall, atmospheric temperature, humidity, wind speed, and wind direction can change the seasonality and concentration of plant pollen, thus inducing allergic symptoms and leading to the occurrence and development of AR [36–38]. Temperature plays a very important role in increasing the pollen concentration in the air and is the main factor controlling the start of grass pollen season and the peak duration of pollen [21, 39, 40]. In the present study, we found that total pollen amount in Shenyang significantly correlated with daily minimum temperature, daily average precipitation, and daily average relative humidity but not with daily average temperature, maximum temperature,

average atmospheric pressure, and wind scale. In fact, it is difficult to explain the correlation between pollen concentration and meteorological parameters, because the relationship between pollen and meteorological conditions can be nonlinear. Pollen concentration can be affected by multiple meteorological conditions at the same time, so it is difficult to make correlation tests. We assumed that the meteorological conditions shown in the present paper were not correlated with pollen concentration based on the above reasons [21, 41]. In addition, the concentration of air pollutants (such as SO₂, NO₂, CO, and O₃) was consistent and significantly correlated with the concentration of grass pollen in the air [42]. In this study, the results show that total pollen amount had a significant correlation with the concentrations of PM_{2.5}, PM₁₀, SO₂, and NO₂ in Shenyang area, which is also confirmed at this point. Moreover, there were also significant correlations between the number of pollen allergen-positive patients and the daily concentrations of PM_{2.5}, PM₁₀, SO₂, and NO₂. Based on the above studies, it can be concluded that allergenic pollen could mediate AR consultation rate through local meteorological conditions.

In summary, this study illustrated how pollen dispersal, climate change, and outdoor air pollution acted as the environmental risk factors for AR development in Shenyang area. To be specific, the lower temperature from August to September was conducive to the pollen dispersal of Compositae and Moraceae, leading to the increase of particulate pollutants in the air and the deterioration of air quality, which leads to a significant increase in outpatient visits of AR patients in the same period. This study had a large sample size, which could reflect the whole AR consultation laws and the relevant air environmental factors in Shenyang area. However, clinical data collection of subjects in the cohort was incomplete, and the impact of other risk factors, such as residential habits and occupational dust exposure and allergy history, on AR is still unclear. Subsequently, clinical data of patients should be improved to comprehensively analyze multiple factors affecting AR consultation rate, providing reference materials for clinical prevention and treatment of AR.

Data Availability

All data generated or analyzed during this study are included in this article.

Ethical Approval

The study was approved by the Ethical Committee of The First Hospital of China Medical University.

Conflicts of Interest

The authors report no conflict of interest.

Authors' Contributions

Conception and design were done by Feifei Jiang. Administrative support was done by Aihui Yan. Provision of study

materials or patients was done by Aihui Yan. Collection and assembly of data were done by Feifei Jiang. Data analysis and interpretation were done by Feifei Jiang. Manuscript writing was done by all authors. Final approval of manuscript was done by all authors

Acknowledgments

This work was supported by Science and Technology Planning Project of Liaoning Province [grant numbers: 20170058]. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- [1] Y. Zhang and L. Zhang, "Prevalence of allergic rhinitis in China," *Allergy, Asthma & Immunology Research*, vol. 6, no. 2, pp. 105–113, 2014.
- [2] C. Mastroianni, D. Posa, F. Cipriani, and C. Caffarelli, "Asthma and allergic rhinitis in childhood: what's new," *Pediatric Allergy and Immunology*, vol. 27, no. 8, pp. 795–803, 2016.
- [3] Q. Deng, C. Lu, Y. Yu, Y. Li, J. Sundell, and D. Norbäck, "Early life exposure to traffic-related air pollution and allergic rhinitis in preschool children," *Respiratory Medicine*, vol. 121, pp. 67–73, 2016.
- [4] N. D. Dave, L. Xiang, K. E. Rehm, and G. D. Marshall Jr., "Stress and allergic diseases," *Immunology and Allergy Clinics of North America*, vol. 31, no. 1, pp. 55–68, 2011.
- [5] G. W. Canonica, J. Bousquet, J. Mullol, G. K. Scadding, and J. C. Virchow, "A survey of the burden of allergic rhinitis in Europe," *Allergy*, vol. 62, Suppl 85, pp. 17–25, 2007.
- [6] S. D. Reed, T. A. Lee, and D. C. McCrory, "The economic burden of allergic Rhinitis," *PharmacoEconomics*, vol. 22, no. 6, pp. 345–361, 2004.
- [7] S. Hong, D. K. Son, W. R. Lim et al., "The prevalence of atopic dermatitis, asthma, and allergic rhinitis and the comorbidity of allergic diseases in children," *Environmental Health and Toxicology*, vol. 27, article e2012006, 2012.
- [8] W. C. Chiang, Y. M. Chen, H. K. K. Tan et al., "Allergic rhinitis and non-allergic rhinitis in children in the tropics: prevalence and risk associations," *Pediatric Pulmonology*, vol. 47, no. 10, pp. 1026–1033, 2012.
- [9] D. A. Khan, "Allergic rhinitis and asthma: epidemiology and common pathophysiology," *Allergy and Asthma Proceedings*, vol. 35, no. 5, pp. 357–361, 2014.
- [10] I. Davila, J. Mullol, M. Ferrer et al., "Genetic aspects of allergic rhinitis," *Journal of Investigational Allergology & Clinical Immunology*, vol. 19, Suppl 1, pp. 25–31, 2009.
- [11] D. Passali, C. Cingi, P. Staffa, F. Passali, N. B. Muluk, and M. L. Bellussi, "The international study of the allergic rhinitis survey: outcomes from 4 geographical regions," *Asia Pacific Allergy*, vol. 8, no. 1, article e7, 2018.
- [12] D. Jenerowicz, W. Silny, A. Dańczak-Pazdrowska, A. Polańska, A. Osmola-Mańkowska, and K. Olek-Hrab, "Environmental factors and allergic diseases," *Annals of Agricultural and Environmental Medicine*, vol. 19, no. 3, pp. 475–481, 2012.
- [13] Y. Takahashi, P. Sheridan, A. Niida et al., "The AURKA/TPX2 axis drives colon tumorigenesis cooperatively with MYC," *Annals of Oncology*, vol. 26, no. 5, pp. 935–942, 2015.
- [14] S. Y. An, H. G. Choi, S. W. Kim et al., "Analysis of various risk factors predisposing subjects to allergic rhinitis," *Asian Pacific Journal of Allergy and Immunology*, vol. 33, no. 2, pp. 143–151, 2015.
- [15] Y. Yang, Y. Wang, L. Lv et al., "The prevalence and associated lifestyle risk factors of self-reported allergic rhinitis in Kazakh population of Fukang City," *Medicine*, vol. 96, no. 39, article e8032, 2017.
- [16] T. S. Higgins and D. D. Reh, "Environmental pollutants and allergic rhinitis," *Current Opinion in Otolaryngology & Head and Neck Surgery*, vol. 20, no. 3, pp. 209–214, 2012.
- [17] S. He, Z. Mou, L. Peng, and J. Chen, "Impacts of meteorological and environmental factors on allergic rhinitis in children," *International Journal of Biometeorology*, vol. 61, no. 5, pp. 797–806, 2017.
- [18] B. Song, D. Xu, Z. Ma, Y. Sun, and Z. Fengjiu, "Pollen observation and prevention and treatment of allergic rhinitis in Inner Mongolia," *Chinese ENT News and Reviews*, vol. 4, 2019.
- [19] Specialty Group of Rhinology, "Chinese guideline for diagnosis and treatment of allergic rhinitis (2022, revision)," *Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi*, vol. 57, no. 106–129, 2022.
- [20] Y. Ouyang, J. Li, D. Zhang, E. Fan, Y. Li, and L. Zhang, "A model to predict the incidence of allergic rhinitis based on meteorological factors," *Scientific Reports*, vol. 7, no. 1, article 10006, 2017.
- [21] C. Bishan, L. Bing, C. Chixin et al., "Relationship between airborne pollen assemblages and major meteorological parameters in Zhanjiang, South China," *PLoS One*, vol. 15, no. 10, article e0240160, 2020.
- [22] H. Tong, L. Gao, Y. Deng et al., "Prevalence of allergic rhinitis and associated risk factors in 6 to 12 years schoolchildren from Wuhan in Central China: a cross-sectional study," *American Journal of Rhinology & Allergy*, vol. 34, no. 5, pp. 632–641, 2020.
- [23] J. Wang, Y. Zhang, B. Li et al., "Asthma and allergic rhinitis among young parents in China in relation to outdoor air pollution, climate and home environment," *Science of the Total Environment*, vol. 751, article 141734, 2021.
- [24] M. Zheng, X. Wang, M. Wang et al., "Clinical characteristics of allergic rhinitis patients in 13 metropolitan cities of China," *Allergy*, vol. 76, no. 2, pp. 577–581, 2021.
- [25] S. K. Steinsvaag, "Allergic rhinitis: an updated overview," *Current Allergy and Asthma Reports*, vol. 12, no. 2, pp. 99–103, 2012.
- [26] H. Hong, Q. Yang, K. Zuo et al., "A hospital-based survey on the prevalence of bronchial asthma in patients with allergic rhinitis in southern China," *American Journal of Rhinology & Allergy*, vol. 27, no. 6, pp. 502–505, 2013.
- [27] Y. Shen, J. H. Zeng, S. L. Hong, and H. Y. Kang, "Prevalence of allergic rhinitis comorbidity with asthma and asthma with allergic rhinitis in China: a meta-analysis," *Asian Pacific Journal of Allergy and Immunology*, vol. 37, no. 4, pp. 220–225, 2019.
- [28] M. Thomas, "Allergic rhinitis: evidence for impact on asthma," *BMC Pulmonary Medicine*, vol. 6, Suppl 1, p. S4, 2006.
- [29] C. F. Schuler IV and J. M. Montejo, "Allergic rhinitis in children and adolescents," *Pediatric Clinics of North America*, vol. 66, no. 5, pp. 981–993, 2019.
- [30] X. Y. Wang, T. T. Ma, X. Y. Wang et al., "Prevalence of pollen-induced allergic rhinitis with high pollen exposure in

- grasslands of northern China,” *Allergy*, vol. 73, no. 6, pp. 1232–1243, 2018.
- [31] F. Zhang, T. Krafft, D. Zhang, J. Xu, and W. Wang, “The association between daily outpatient visits for allergic rhinitis and pollen levels in Beijing,” *Science of the Total Environment*, vol. 417–418, pp. 39–44, 2012.
- [32] P. J. Beggs, “Adaptation to impacts of climate change on aero-allergens and allergic respiratory diseases,” *International Journal of Environmental Research and Public Health*, vol. 7, no. 8, pp. 3006–3021, 2010.
- [33] A. Costache, O. Berghi, R. Cergan et al., “Respiratory allergies: Salicaceae sensitization (review),” *Experimental and Therapeutic Medicine*, vol. 21, no. 6, p. 609, 2021.
- [34] J. Li, Y. Huang, X. Lin et al., “Influence of degree of specific allergic sensitivity on severity of rhinitis and asthma in Chinese allergic patients,” *Respiratory Research*, vol. 12, no. 1, p. 95, 2011.
- [35] V. Mahillon, S. Saussez, and O. Michel, “High incidence of sensitization to ornamental plants in allergic rhinitis,” *Allergy*, vol. 61, no. 9, pp. 1138–1140, 2006.
- [36] I. R. Lake, N. R. Jones, M. Agnew et al., “Climate change and future pollen allergy in Europe,” *Environmental Health Perspectives*, vol. 125, no. 3, pp. 385–391, 2017.
- [37] J. I. Silverberg, M. Braunstein, and M. Lee-Wong, “Association between climate factors, pollen counts, and childhood hay fever prevalence in the United States,” *The Journal of Allergy and Clinical Immunology*, vol. 135, no. 2, pp. 463–469.e5, 2015.
- [38] G. D’Amato and L. Cecchi, “Effects of climate change on environmental factors in respiratory allergic diseases,” *Clinical and Experimental Allergy*, vol. 38, no. 8, pp. 1264–1274, 2008.
- [39] H. Garcia-Mozo, J. A. Oteros, and C. Galan, “Impact of land cover changes and climate on the main airborne pollen types in southern Spain,” *Science of the Total Environment*, vol. 548–549, pp. 221–228, 2016.
- [40] G. Sofia, T. Emma, T. Veronica, and F. Giuseppe, “Climate change: consequences on the pollination of grasses in Perugia (Central Italy). A 33-year-long study,” *International Journal of Biometeorology*, vol. 61, no. 1, pp. 149–158, 2017.
- [41] A. M. Jones and R. M. Harrison, “The effects of meteorological factors on atmospheric bioaerosol concentrations –a review,” *Science of the Total Environment*, vol. 326, pp. 151–180, 2004.
- [42] Y. Ouyang, Z. Yin, Y. Li, E. Fan, and L. Zhang, “Associations among air pollutants, grass pollens, and daily number of grass pollen allergen-positive patients: a longitudinal study from 2012 to 2016,” *International Forum of Allergy & Rhinology*, vol. 9, no. 11, pp. 1297–1303, 2019.

Research Article

Prognosis and Efficacy of Laparoscopic Surgery on Patients with Endometrial Carcinoma: Systematic Evaluation and Meta-Analysis

Jiong Ma , Chunxia Zhou , Jinyan Chen , and Xuejun Chen 

Department of Gynecology, Second Affiliated Hospital, Zhejiang University, School of Medicine, Hangzhou, Zhejiang 310009, China

Correspondence should be addressed to Xuejun Chen; 2303011@zju.edu.cn

Received 18 July 2022; Revised 26 August 2022; Accepted 31 August 2022; Published 22 September 2022

Academic Editor: Tao Huang

Copyright © 2022 Jiong Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. The prognosis and efficacy of laparoscopic surgery (LPS) and open surgery or robotic surgery (RS) on endometrial carcinoma (EC) patients were compared. **Methods.** Data as of May 2021 were retrieved from databases like PubMed, Embase, Cochrane Library, and Web of Science. The study involved randomized controlled trials (RCTs), cohort studies, or case-control studies for comparing the effects of LPS and open surgery or robotic surgery (RS) on EC treatment. The primary outcomes included duration of operation, blood loss, length of stay (LOS), postoperative complications, and recurrence rate. Secondary outcomes included 3-year progression-free survival (PFS) rate/disease-free survival (DFS) rate and 3-year overall survival (OS) rate. **Results.** A total of 24 studies were involved, and all of them were cohort studies except 1 RCT and 1 case-control study. There was no significant difference in duration of operation between LPS and open surgery (MD = -0.06, 95% CI: -0.37 to 0.25) or RS (MD = -0.15, 95% CI: -1.27 to 0.96). In comparison with the open surgery, LPS remarkably reduced blood loss (MD = -0.43, 95% CI: -0.58 to -0.29), LOS (MD = -0.71, 95% CI: -0.92 to -0.50), and the complication occurrence rate (RR = 0.83, 95% CI: 0.73 to 0.95). However, LPS and RS saw no difference in blood loss (MD = 0.01, 95% CI: -0.77 to 0.79). Besides, in comparison with RS, LPS prominently shortened the LOS (MD = 0.26, 95% CI: 0.12 to 0.40) but increased the complication occurrence rate (RR = 1.74, 95% CI: 1.57 to 1.92). In contrast to open surgery or RS, LPS saw no difference in occurrence rate (RR = 0.75, 95% CI: 0.56 to 1.01; RR = 0.97, 95% CI: 0.62 to 1.53), 3-year PFS/DFS (RR = 0.99, 95% CI: 0.90 to 1.09; RR = 1.30, 95% CI: 0.87 to 1.96), and 3-year OS (RR = 0.97, 95% CI: 0.91 to 1.04; RR = 1.21, 95% CI: 0.91 to 1.60). **Conclusion.** In sum, LPS was better than open surgery, which manifested in the aspects of less blood loss, shorter LOS, and fewer complications. LPS, therefore, was the most suitable option for EC patients. Nevertheless, LPS had no advantage over RS, and sufficient prospective RCTs are needed to further confirm its strengths.

1. Introduction

Endometrial carcinoma (EC) is the most commonly diagnosed gynecologic malignant tumor, especially in some developed countries [1], whose 5-year survival rate was 34.7% (445805 cases) [2]. The risk factors of EC include early menarche, delay menopause, diabetes, polycystic ovarian syndrome (PCOS), metabolic syndrome, current treatments with tamoxifen, and obesity [3–5]. With the increase of risk factors such as aging of population and obesity, the morbidity of EC will continue to rise. Most EC patients who are diagnosed in the early stages (Federation Internatio-

nale of Gynecologie and Obstetrigue (FIGO) stage I or II) have better prognoses [6].

Surgeries remain the major treatment for early EC, which mainly include vaginal surgery, laparotomy (LT), or open surgery, laparoscopic surgery (LPS), and robotic surgery (RS). Clinical practice guideline and multiple clinical trials indicate that minimally invasive surgery (MIS) is recommended as the preferred surgical approach for EC patients [7, 8]. In the past, LT has always been the first choice for early EC patients. Since the first report of LPS on EC in 1993 [9], LPS, as a MIS, has become increasingly popular in the treatment of EC [10, 11]. In contrast to open

surgery, LPS is characterized by less blood loss, less renaescent adhesion and lower morbidity [12]. RS is a new MIS developed on the basis of LPS. Previous studies show that RS for EC treatment results in a shorter length of stay (LOS), less blood loss, lower conversion rate of open surgery, and lower occurrence rate of intraoperative damage to surrounding organs, compared to LPS [13, 14]. However, in contrast to LPS, RS prolongs operation and recovery time [15, 16]. Recently, a multicenter retrospective study compared the therapeutic efficacy of LPS and radical abdominal hysterectomy on early EC patients. The result indicated a decrease in the disease-free survival (DFS) of patients who underwent LPS [17]. However, a prospective study validated that LPS could dramatically improve the short- and long-term quality of life (QOL) of EC patients [18]. Hence, this study systematically evaluated the prognosis and efficacy of LPS.

Consequently, we conducted a meta-analysis to elucidate the prognosis and effect of LPS on EC by comparison with open surgery or RS from the perspective of perioperative results, postoperative complications, recurrence rate, and survival time. This effort will bring insight into the surgical treatment of EC patients.

2. Methods

2.1. Literature Retrieval. The study was performed in adherence to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement [19]. All relevant literature included in the databases like PubMed, Embase, Cochrane Library, and Web of Science were retrieved from the construction of the databases to May 2021. Keywords used for searching included “endometrial carcinoma,” “hysteroscopic surgery,” “minimally invasive surgery,” “laparoscopic surgery,” “robotic surgery,” and “open surgery.” The detailed strategy of literature retrieval was as follows: (((Endometrial Neoplasm*[MeSH Terms]) OR (Endometrial Carcinoma*[MeSH Terms])) OR (Endometrial Cancer*[MeSH Terms])) AND (((Laparoscopy[Title/Abstract]) OR (Hysteroscope[Title/Abstract])) OR (Minimally invasive[Title/Abstract])) OR (Open[Title/Abstract])) OR (robotic[Title/Abstract])) AND ((operation[Title/Abstract]) OR (surgery[Title/Abstract])).

2.2. Selection of Studies. Inclusion criteria of the literature were as follows: (1) patients diagnosed with EC; (2) comparison between effects of LPS and open surgery or RS on EC treatment; (3) at least one of the results such as duration of operation, blood loss, LOS, postoperative complications, recurrence rate, 3-year progression-free survival (PFS) rate/DFS rate, and 3-year overall survival (OS) rate was reported; and (4) study was designed as RCT, cohort study, or case-control study. The following were the exclusion criteria of the literature: (1) repeated publication, case series, case report, comments, meeting abstract, review, editorial, letter, and so on; (2) data were not sufficient to obtain the result of our study; (3) article replications; and (4) studies lack of efficacy-related data.

2.3. Data Extraction and Quality Assessment. The information obtained from the literature included the information of the authors, publication year, country, study design, the year the samples were collected, the number of samples, and intervening measures. Data of the patients included age, body mass index (BMI), FIGO stage, pathological grading, and outcome indicator. Primary outcomes included duration of operation, blood loss, LOS, and postoperative complications. Secondary outcomes involved postoperative recurrence rate, 3-year PFS rate/DFS rate, and 3-year OS rate.

Cochrane risk of bias assessment tool was employed to assess the quality of RCT which was graded as “low risk,” “high risk,” and “uncertain risk.” Besides, the Newcastle-Ottawa Scale (NOS) was utilized to evaluate the risk of publication bias in observational studies. Aggregate points of NOS were 9, and literature with the points greater than or equal to 6 was considered of good quality.

The search and selection of articles and the extraction and quality evaluation of data were independently finished by two investigators. Disputes were solved by the third investigator through consultation.

2.4. Statistical Analysis. Meta-analysis was performed using the Stata 16.0 software. Continuous data were expressed as mean \pm standard deviation (SD). Mean difference (MD) was measured via continuous results, and 95% confidence intervals (CIs) were used to assess the concrete therapeutic effect. If CI included 0, it denoted no statistical difference between two groups. Besides, two-category data were merged and analyzed utilizing relative risk (RR) and their 95% CI. If CI included 1, it indicated no statistical difference between two groups. I^2 statistic was applied to assess the statistical heterogeneity of studies involved. A random effect model was used if $p < 0.1$ or $I^2 > 50\%$, suggesting a remarkable heterogeneity. Otherwise, a fixed effect model was used.

3. Results

3.1. Screening and Selection of Reports. A total of 1666 reports were retrieved based on the established searching strategies, among which 224 reports were excluded. Further, 1400 reports were excluded by scanning their title and abstract. After reading the full text of the remaining 42 reports, 7 reports of them reported unrelated data while 11 of them lacked sufficient data for obtaining the result of our study. Finally, 24 reports were selected for our study [13, 20–42]. The procedures of report screening are shown in Figure 1.

3.2. Characteristics of the Studies and Quality Evaluation. A total of 24 reports were involved in the study, in which 6814 patients underwent LPS and 5315 patients underwent open surgery. Besides, 6121 patients underwent RS. Except for 1 RCT and 1 case-control study, the other reports were cohort studies. All characteristics of reports involved and the results of quality assessment are displayed in Table 1. Papers with the points greater than or equal to 6 were of high quality.

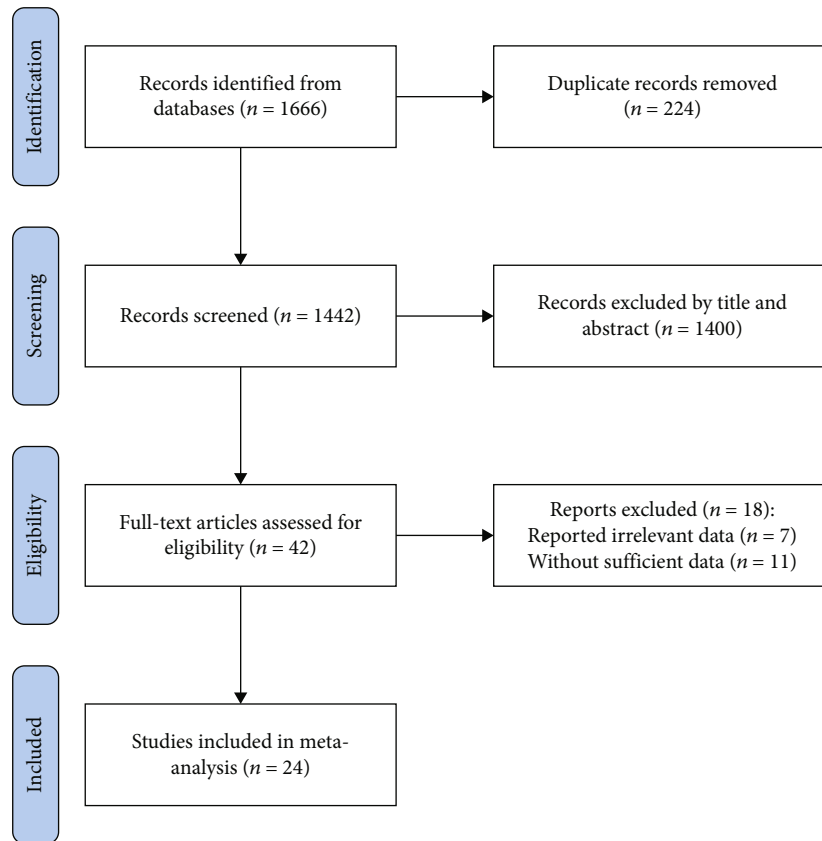


FIGURE 1: Flow chart about literature screening.

3.3. Results of Meta-Analysis

3.3.1. Duration of Operation and Blood Loss. In respect of the operation time, 9 studies compared that in LPS and open surgery ($I^2 = 95.2\%$), and 2 studies compared that in LPS and RS ($I^2 = 98.0\%$). Due to huge heterogeneity, a random effect model was introduced for analysis. The outcome of meta-analysis suggested that there was no significant difference in duration of operation between LPS and open surgery (MD = -0.06 , 95% CI: -0.37 to 0.25) or RS (MD = -0.15 , 95% CI: -1.27 to 0.96) (Figures 2(a) and 2(b)).

In respect of blood loss, 6 studies compared that in LPS and open surgery ($I^2 = 47.1\%$), and 2 studies compared that in LPS and RS ($I^2 = 96.0\%$). Based on the results of heterogeneity analysis, a fixed effect model and a random effect model were employed, respectively. The result of meta-analysis demonstrated that the blood loss of LPS was dramatically decreased in contrast to open surgery (MD = -0.43 , 95% CI: -0.58 to -0.29), but it had no remarkable difference when compared with that of RS (MD = 0.01 , 95% CI: -0.77 to 0.79) (Figures 3(a) and 3(b)).

3.3.2. Postoperative LOS, Complications, and Recurrence Rate. In respect of LOS, 5 studies compared that in LPS and open surgery ($I^2 = 72.1\%$), and 2 studies compared that in LPS and RS ($I^2 = 0.0\%$). Based on the results of heterogeneity analysis, a random effect model and a fixed effect model were employed, respectively. It was exhibited in

meta-analysis that the LOS of LPS was shorter than that of open surgery (MD = -0.71 , 95% CI: -0.92 to -0.50) (Figure 4(a)). Meanwhile, the LOS of RS was shorter by comparison with that of LPS (Figure 4(b)).

In respect of postoperative complications, 15 studies compared that in LPS and open surgery ($I^2 = 83.3\%$), and 8 studies compared that in LPS and RS ($I^2 = 76.6\%$). Due to huge heterogeneity, a random effect model was employed. As demonstrated in the result of meta-analysis, LPS resulted in a decrease in the occurrence rate of complications relative to open surgery (RR = 0.83 , 95% CI: 0.73 to 0.95) (Figure 5(a)) but an increase in that compared to RS (RR = 1.74 , 95% CI: 1.57 to 1.92) (Figure 5(b)).

In respect of recurrence rate, 7 studies compared that in LPS and open surgery ($I^2 = 0.0\%$), and 3 studies compared that in LPS and RS ($I^2 = 5.3\%$). Because of small heterogeneity, a fixed effect model was employed. According to the result of meta-analysis, there was no notable difference in recurrence rate between LPS and open surgery (RR = 0.75 , 95% CI: 0.56 to 1.01) or RS (RR = 0.97 , 95% CI: 0.62 to 1.53) (Figures 6(a) and 6(b)).

3.3.3. The 3-Year PFS/DFS and OS. In respect of 3-year PFS/DFS, 2 studies compared that in LPS and open surgery ($I^2 = 0.0\%$), and 3 studies compared that in LPS and RS ($I^2 = 90.2\%$). Based on the heterogeneity analysis, a fixed effect model and a random effect model were employed,

TABLE 1: Characteristics of literature involved.

Author	Year	Country	Center	Study design	Year of sample collection	FIGO stage	Grade	Age, mean (SD), y	BMI, mean (SD), kg/m ²	Sample size (n)	Interventions	NOS
Escobar et al.	2012	USA	Multi	Retrospective cohort study	2009.4-2010.9	I-II	I-III	60.9 (12.1)	31.2 (6.7)	30	Laparoscopy	7
Wright et al.	2012	USA	Multi	Retrospective cohort study	2008.10-2010.3			59.7 (9.2)	31.4 (6.6)	30	Robotic	7
Obermair et al.	2012	Australia, New Zealand	Multi	RCT	2005.10-2010.6	IA-IVB	I-III	61	29.3	1027	Open	8
Goicoechea et al.	2013		Multi	Retrospective cohort study	2003.1-2010.1	I-IV	I-III	62	29.2	232	Robotic	9
Bogani et al.	2014	Italy	Single	Retrospective cohort study	2002.5-2012.10	I-IV	I-III	78	26.6	59	Open	9
Bogani et al.	2014	Italy	Single	Retrospective cohort study	1992.1-2013.5	I-III	I-III	83	25.4 (3.6)	22	Open	8
Boosz et al.	2014		Single	Retrospective cohort study	2002.1-2009.12	I-III	I-III	66.7 (11.3)	29.7 (6.2)	107	Open	9
Chiou et al.	2014	China	Single	Retrospective cohort study	2005-2013	IA-IIIIC		53.6 (11.1)	26.0 (5.2)	150	Robotic	9
Park et al.	2015	USA	Single	Retrospective cohort study	2001.1-2012.7	IA-IVB	I-III	60	30.4	586	Robotic	6
Corrado et al.	2015	Italy	Single	Retrospective cohort study	2010.8-2013.12	IA-IVB	I-III	62	29	277	Robotic	6
Stefano et al.	2015	Italy	Multi	Retrospective case-control study	2000.1-2013.3	IV		63.2 (11.3)	27	764	Open	6
Ling-hui Chu et al.	2015	China	Single	Retrospective cohort study	2002.1-2012.6	I-III	I-III	55.3	25.4	70	Open	8
Barrie et al.	2016	USA	Single	Retrospective cohort study	2009.1-2014.1	0-IV	0-III	62	30.6	688	Robotic	9
Monterossi et al.	2016	Italian	Multi	Retrospective cohort study	2000.5-2015.6	I-II	I-III	67	27	141	Open	9
Beck et al.	2017	USA	Single	Retrospective cohort study	2008-2013			63 (25-96)	23.5 (18.0-44.6)	400	Robotic	8
Ruan et al.	2018	Singapore	Single	Retrospective cohort study	2008-2014	I	I-III	55.6 (9.7)	28.1 ± 5.7	145	Open	8
Corrado et al.	2018	Italian	Multi	Retrospective cohort study	2010-2012	I-IV	I-III	63.43	35.4 (5.8)	406	Robotic	9
Deura et al.	2018	Japan	Single	Retrospective cohort study	2005-2016	I	I-II	57	23.6 (15.9-48.8)	40	Open	8
Jørgensen et al.	2018	Denmark	Single	Prospective cohort study	2005.1.1-2015.6.30	I-II	I-III	67 (33-94)	361 (29.4)	569	Robotic	9
Vardar et al.	2018		Single	Retrospective cohort study	2005.1-2016	I-III	I-III	61 (11.8)	35.8 (7.4)	286	Open	9
Ghazali et al.	2018	Malaysia	Single	Retrospective cohort study	2010.1-2014.12			55.62 (12.75)	29.24 (3.71)	26	Open	8
Dieterich et al.	2019	Germany	Single	Retrospective cohort study	2005.1-2014.12	I-III	I-III	64.00 (11.04)	33.71 (6.26)	108	Open	9
Papadia et al.	2019	Swiss confederation	Single	Retrospective cohort study	2001.10-2015.11	III		65 (11)	26.6 (7.4)	51	Open	9
Tanaka et al.	2020	Japan	Single	Retrospective cohort study	2004.1-2019.12	IA-IV	I-III	55.3 (10.6)	24.4 (4.9)	226	Open	8

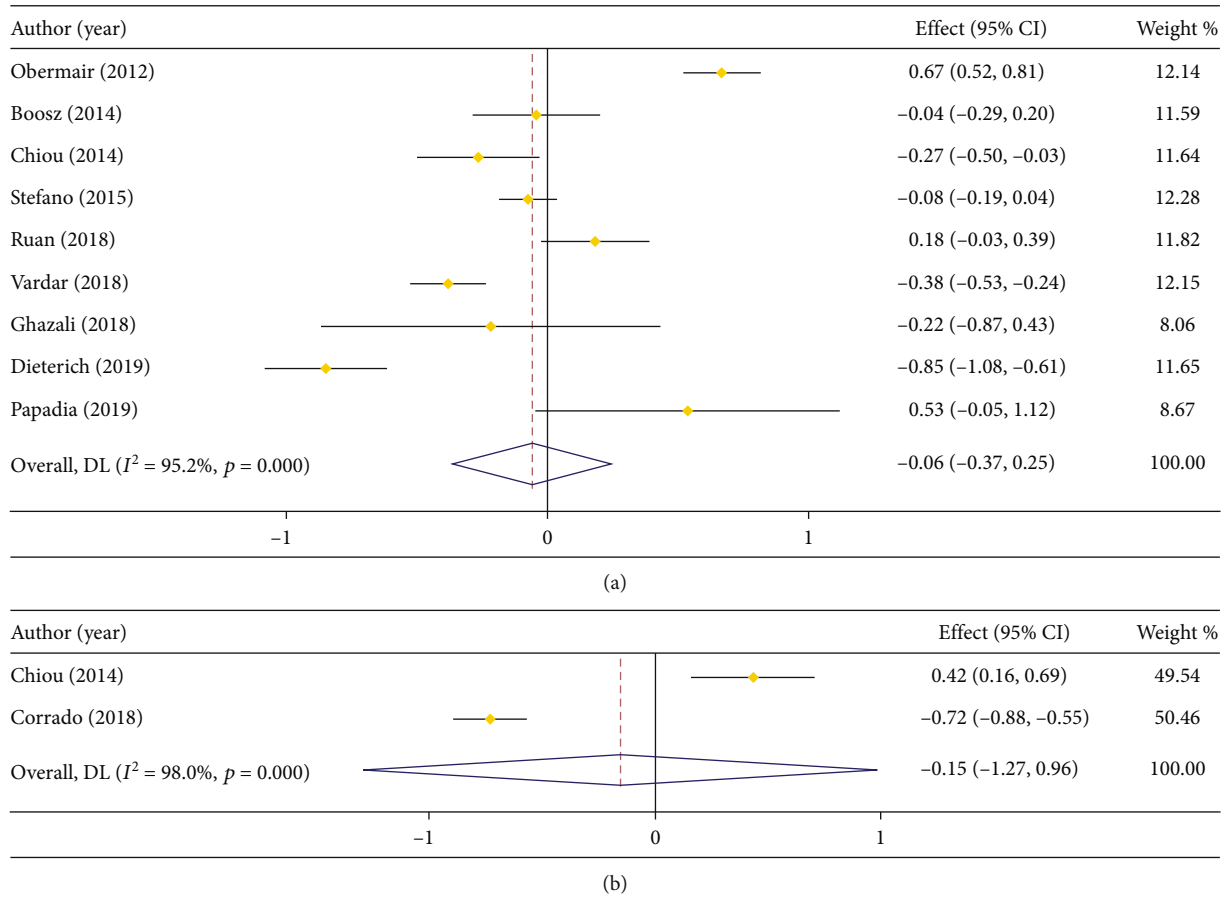


FIGURE 2: Forest plot comparing duration of operation. (a) LPS vs. open surgery; (b) LPS vs. RS.

respectively. The outcome of meta-analysis indicated an insignificant difference in 3-year PFS/DFS between LPS and open surgery (RR = 0.99, 95% CI: 0.90 to 1.09) or RS (RR = 1.30, 95% CI: 0.87 to 1.96) (Figures 7(a) and 7(b)).

In respect of 3-year OS, 3 studies compared that in LPS and open surgery ($I^2 = 0.0\%$), and 4 studies compared that in LPS and RS ($I^2 = 91.2\%$). Based on the results of heterogeneity analysis, a fixed effect model and a random effect model were employed, respectively. As shown in the result of meta-analysis, there was no prominent difference in 3-year OS between LPS and open surgery (RR = 0.97, 95% CI: 0.91 to 1.04) or RS (RR = 1.21, 95% CI: 0.91 to 1.60) (Figures 8(a) and 8(b)).

4. Discussion

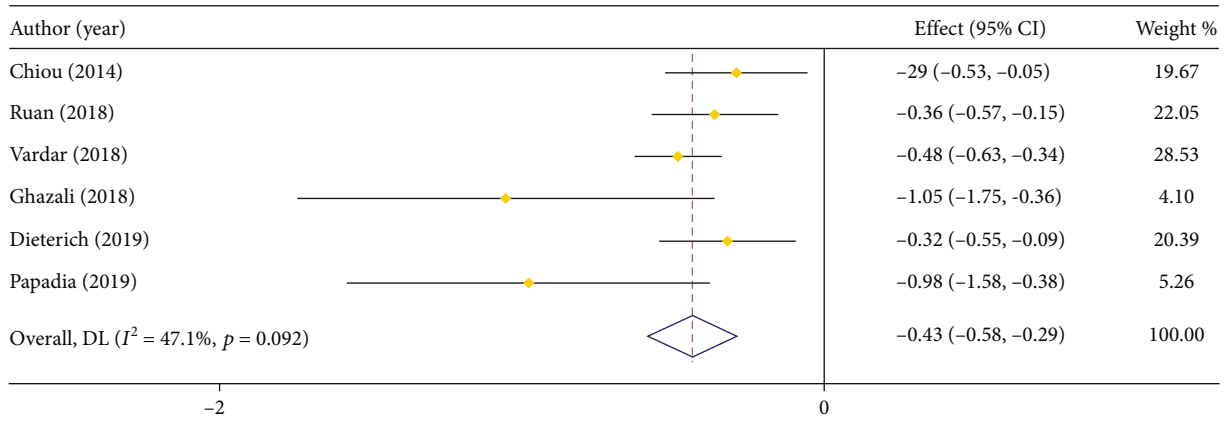
This study found that LPS did not improve the survival time of patients. A retrospective study compared the clinical effect of LPS and open surgery on the treatment of low risk EC patients (grade 1 or 2 EC and mesometrium invasion < 1/2). The result suggested that the 5-year recurrence-free survival (RFS) and OS of LPS were similar to those of open surgery [43]. Besides, another study reported the 5-year survival rate of EC patients who underwent LPS, open surgery, or RS, suggesting that there was no significant difference in 5-year DFS and OS of patients [44]. In addition, a multicenter database

study verified that the long-term prognosis of MIS on treatment of high-risk EC was no worse than that of LT [45]. The above results were in agreement with the outcome of our study, which indicated that LPS did not dramatically improve the 3-year PFS/DFS and OS of patients.

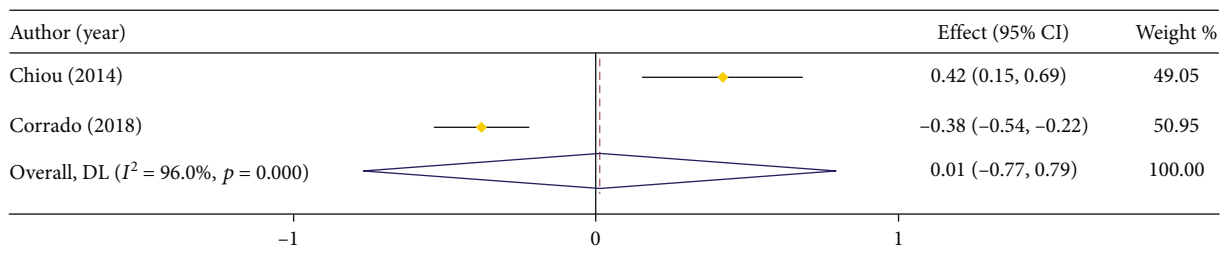
Generally, LPS takes longer time on operation [36]. However, our study manifested that there was no significant difference in duration of operation between LPS and open surgery or RS. Importantly, since the duration of operation would be subjected to the skill of the operator, we could not figure out which method was the most potential to reduce the duration of operation among these surgeries. Meanwhile, the blood loss of LPS was obviously lower than that of open surgery, but it had no remarkable difference in comparison with that of RS.

The LOS of LPS was notably shorter than that of open surgery and LPS also resulted in fewer postoperative complications. However, RS was overwhelmingly better than LPS in the aspect of duration of time and postoperative complications. This study suggested that no obvious difference in recurrence rate was found between LPS and open surgery or RS.

A previous meta-analysis has proved that uterine manipulator is irrelevant to an increase in occurrence rate of positive peritoneal cytology, lymphovascular space invasion, or recurrence in EC patients [46]. Our study only investigated



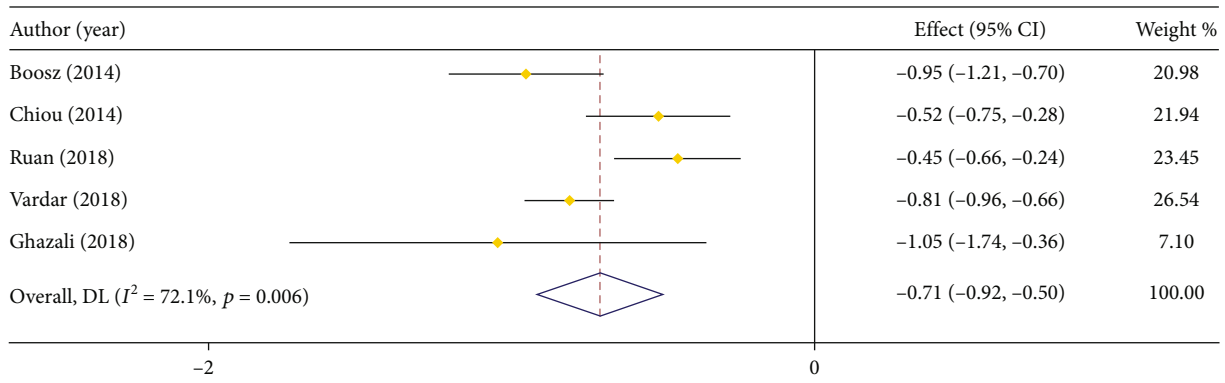
(a)



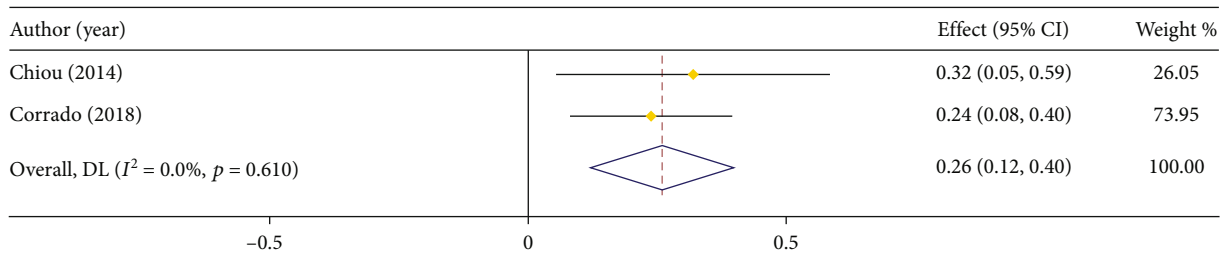
(b)

Note: Weights are from random-effects model

FIGURE 3: Forest plot about the comparison on blood loss. (a) LPS vs. open surgery; (b) LPS vs. RS.



(a)



(b)

FIGURE 4: Forest plot about comparison on LOS. (a) LPS vs. open surgery; (b) LPS vs. RS.

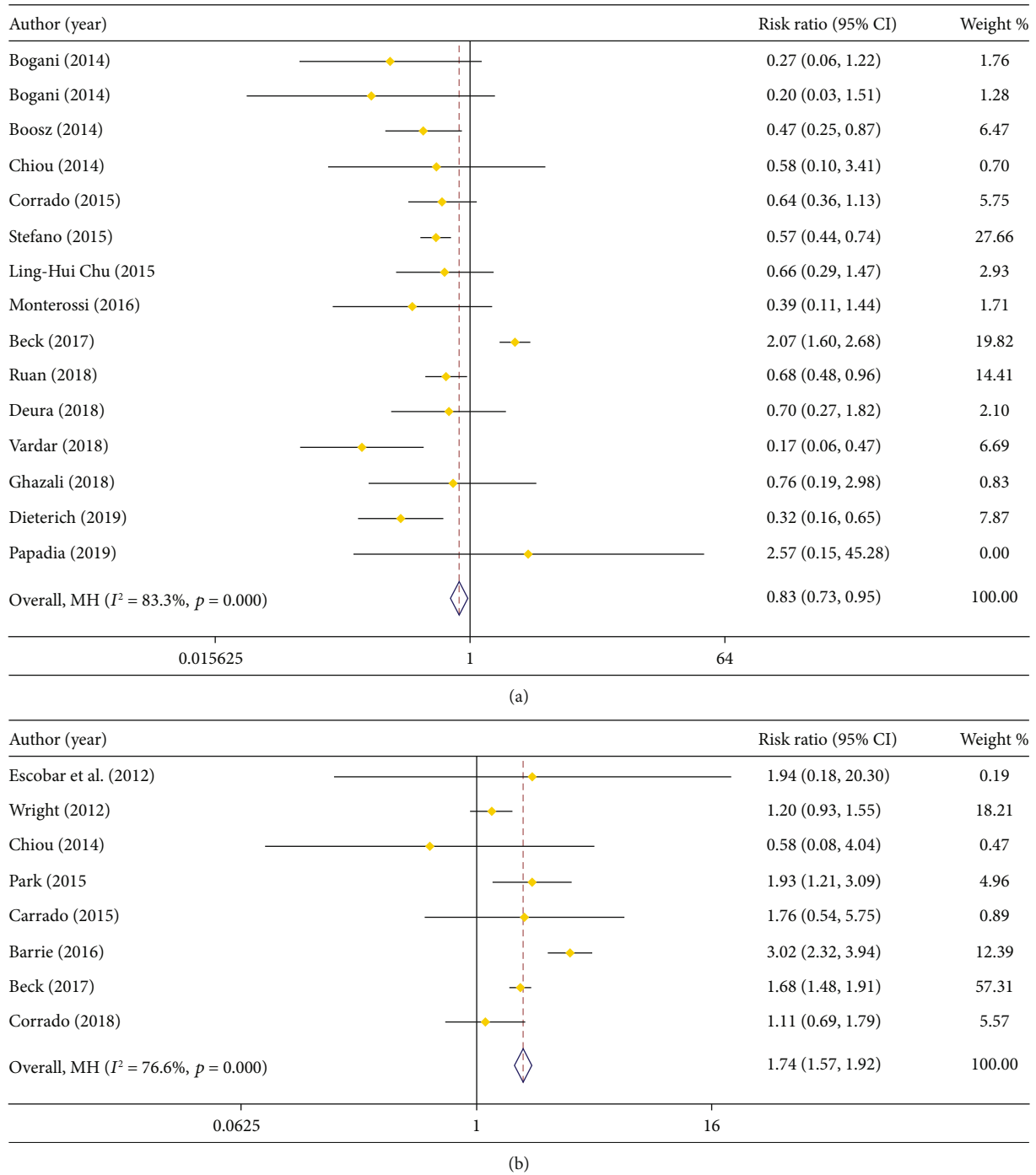
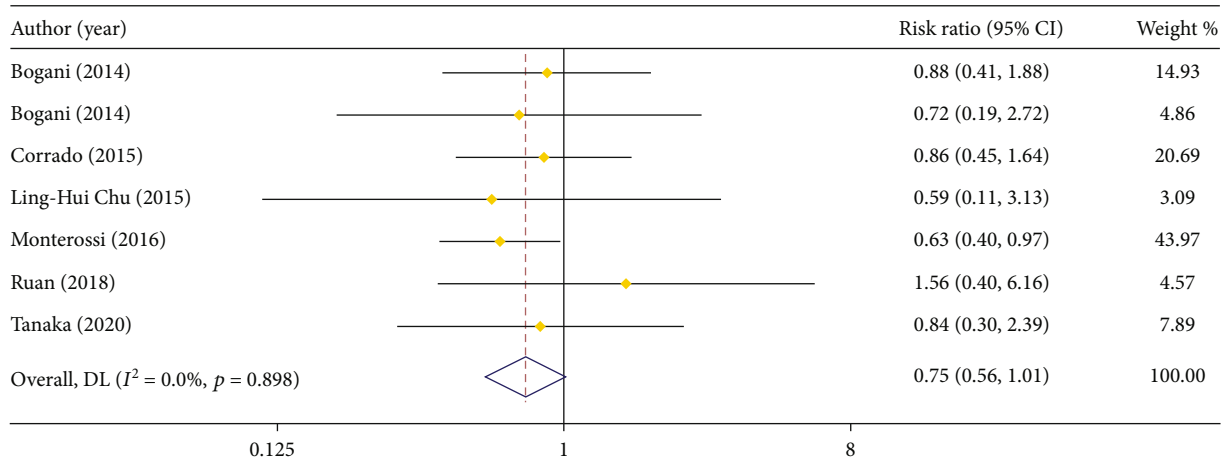


FIGURE 5: Forest plot comparing complications. (a) LPS vs. open surgery; (b) LPS vs. RS.

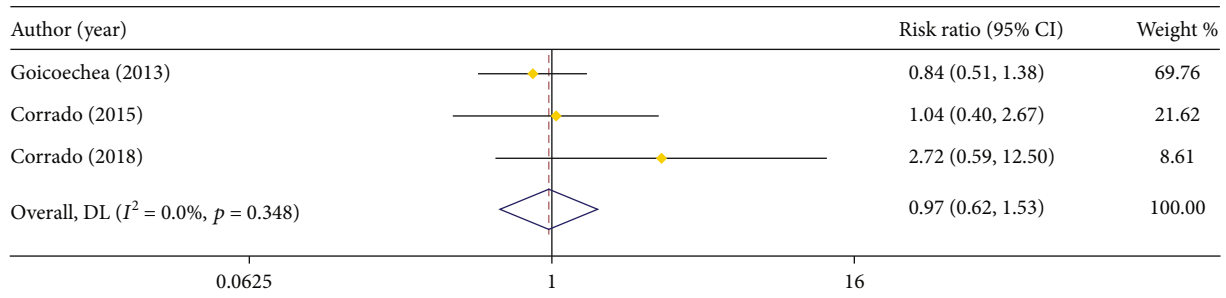
conventional LPS. A recent meta-analysis involved 4 RCTs that compared the clinical effect of laparoendoscopic single-site surgery (LESS) and conventional LPS on the treatment of EC patients, which suggested an insignificant difference between the two surgeries. Meanwhile, LESS only has advantage on reducing trauma [47]. Additionally, another meta-analysis that was similar to our study compared the differences between RS and LPS or open surgery. The analysis suggested that RS was characterized by less

blood loss and blood transfusion, fewer postoperative complications, and less conversion to LT plus shorter LOS compared with the other two surgeries. However, RS took a longer time on operation in surgical staging of EC [48]. Interestingly, these results were consistent with the outcomes of our study. Moreover, we also analyzed the oncological outcome of EC patients.

This study presented some advantages. Firstly, our meta-analysis involved some recent clinical trials with a vast

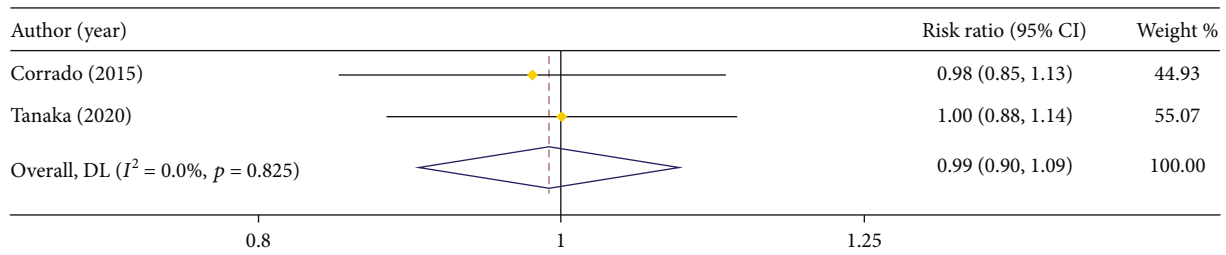


(a)

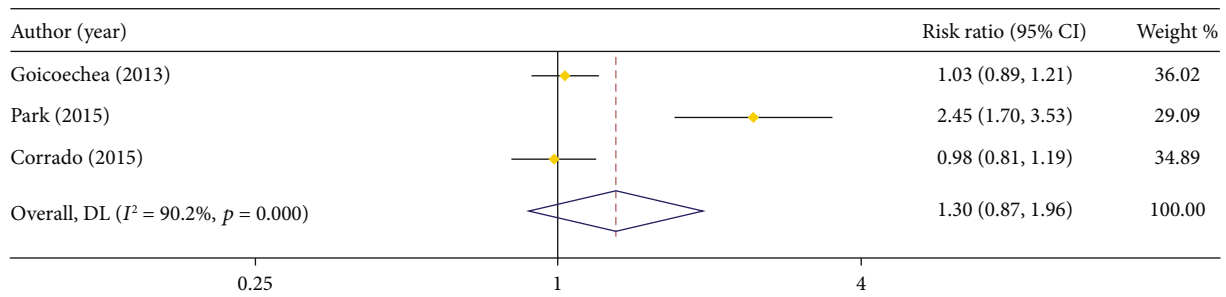


(b)

FIGURE 6: Forest plot of comparison on recurrence rate. (a) LPS vs. open surgery; (b) LPS vs. RS.



(a)



(b)

FIGURE 7: Forest plot about comparison on 3-year PFS/DFS. (a) LPS vs. open surgery; (b) LPS vs. RS.

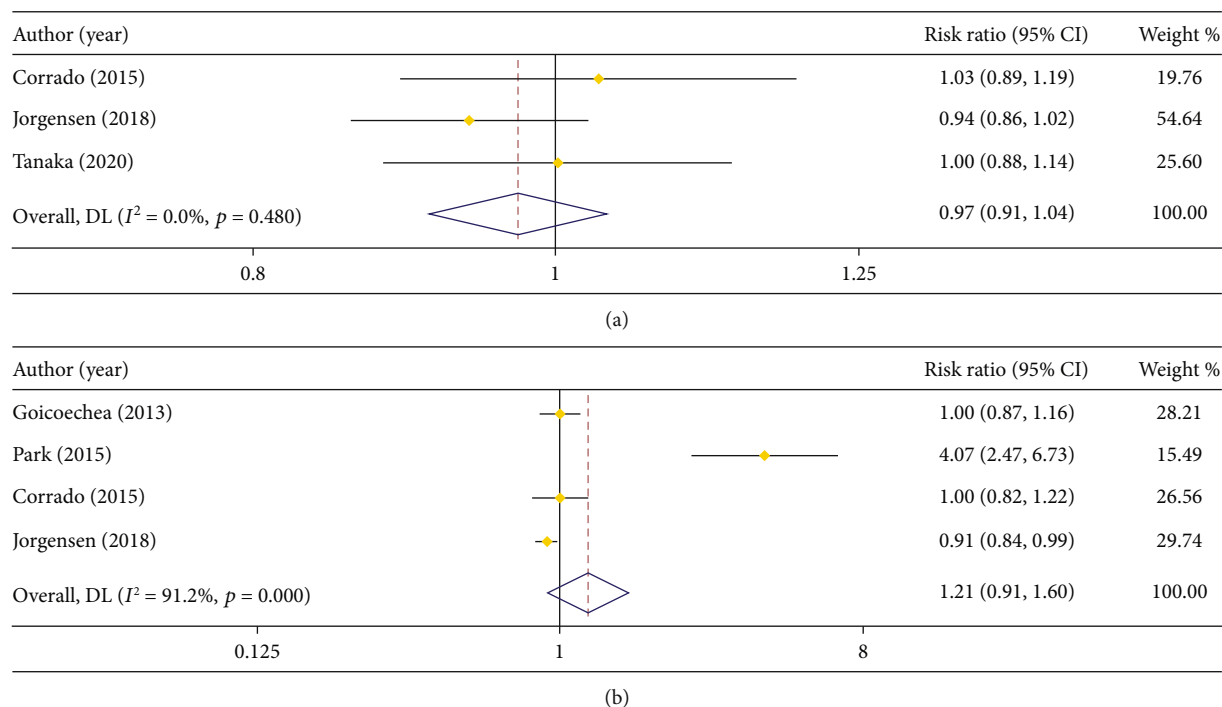


FIGURE 8: Forest plot comparing 3-year OS. (a) LPS vs. open surgery; (b) LPS vs. RS.

number of samples. Furthermore, we also analyzed the survival time of patients. But few samples were involved in the analysis, which was a limitation of our study.

Inevitably, there were limitations in this study. First of all, most of the studies involved in our study were retrospective cohort studies which are inherently subjected to the risk of selection bias. Secondly, the speculation about whether various risk factors affect the prognosis of EC patients who underwent LPS is needed to be further verified. In addition, Cusimano et al. [49] performed a meta-analysis on EC patients with obesity who underwent LPS or robotic hysterectomy. The result of the analysis revealed that LPS and robotic hysterectomy had similar incidence of perioperative complications. However, robotic hysterectomy may reduce conversions due to the positional intolerance of patients suffering from morbid obesity. Finally, because of the lack of related reports on 5-year survival time of patients, we only analyzed the 3-year PFS/DFS and OS of the patients.

In summary, our study revealed that LPS was a safe and effective treatment for EC patients, which was better than open surgery. Nevertheless, LPS was at a disadvantage in the comparison with RS on duration of operation and post-operative complications. In the future, more randomized trials with complete data are needed to verify our conclusion.

Data Availability

The data used to support the findings of this study are included within the article. The data and materials in the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Jiong Ma and Xuejun Chen wrote the main manuscript text, and Chunxia Zhou and Jinyan Chen prepared Figures 1–8. All authors reviewed the manuscript.

Acknowledgments

This study was financially supported by the Zhejiang Medical and Health Science and Technology Plan Project (2014KYA244) and the National Natural Science Foundation of China Youth Science Fund Project (81202064).

References

- [1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2021," *CA: a Cancer Journal for Clinicians*, vol. 71, no. 1, pp. 7–33, 2021.
- [2] GLOBOCAN, WHO, "Estimated cancer incidence, mortality and prevalence worldwide in 2018," 2018, <https://gco.iarc.fr/today/data/factsheets/cancers/24-Corpus-uteri-fact-sheet.pdf>.
- [3] L. Saed, F. Varse, H. R. Baradaran et al., "The effect of diabetes on the risk of endometrial cancer: an updated a systematic review and meta-analysis," *BMC Cancer*, vol. 19, no. 1, p. 527, 2019.
- [4] M. N. Shafiee, N. Razak, M. F. Ahmad, N. Abd Aziz, and N. Adeeb, "A single centre experience of metabolic syndrome and endometrial carcinoma: 5 years review," *Journal of Obstetrics and Gynaecology*, vol. 41, no. 2, pp. 285–289, 2021.

- [5] D. A. Yela, T. A. Ikejiri, C. R. Machado, D. Mutta, and C. L. Benetti-Pinto, "Tamoxifen use as a malignancy risk factor in postmenopausal women with endometrial polyps," *Meno-pause*, vol. 26, no. 8, pp. 863–866, 2019.
- [6] J. Stubert and B. Gerber, "Current issues in the diagnosis and treatment of endometrial carcinoma," *Geburtshilfe und Frauenheilkunde*, vol. 76, no. 2, pp. 170–175, 2016.
- [7] N. Concin, C. L. Creutzberg, I. Vergote et al., "ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma," *Virchows Archiv*, vol. 478, no. 2, pp. 153–190, 2021.
- [8] S. Salehi, Y. Brandberg, E. Åvall-Lundqvist et al., "Long-term quality of life after comprehensive surgical staging of high-risk endometrial cancer - results from the RASHEC trial," *Acta Oncologica*, vol. 57, no. 12, pp. 1671–1676, 2018.
- [9] J. M. Childers, P. R. Brzechffa, K. D. Hatch, and E. A. Surwit, "Laparoscopically assisted surgical staging (LASS) of endometrial cancer," *Gynecologic Oncology*, vol. 51, no. 1, pp. 33–38, 1993.
- [10] R. Asher, A. Obermair, M. Janda, and V. Gebiski, "Disease-free and survival outcomes for Total laparoscopic hysterectomy compared with total abdominal hysterectomy in early-stage endometrial carcinoma: a meta-analysis," *International Journal of Gynecological Cancer*, vol. 28, no. 3, pp. 529–538, 2018.
- [11] M. Janda, V. Gebiski, L. C. Davies et al., "Effect of total laparoscopic hysterectomy vs Total abdominal hysterectomy on disease-free survival among women with stage I endometrial cancer: a randomized clinical trial," *JAMA*, vol. 317, no. 12, pp. 1224–1233, 2017.
- [12] S. Salehi, E. Avall-Lundqvist, B. Legerstam, J. W. Carlson, and H. Falconer, "Robot-assisted laparoscopy versus laparotomy for infrarenal paraaortic lymphadenectomy in women with high-risk endometrial cancer: a randomised controlled trial," *European Journal of Cancer*, vol. 79, pp. 81–89, 2017.
- [13] G. Corrado, E. Vizza, V. Cela et al., "Laparoscopic versus robotic hysterectomy in obese and extremely obese patients with endometrial cancer: A multi-institutional analysis," *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology*, vol. 44, pp. 1935–1941, 2018.
- [14] A. Lindfors, H. Heshar, C. Adok, K. Sundfeldt, and P. Dahm-Kahler, "Long-term survival in obese patients after robotic or open surgery for endometrial cancer," *Gynecologic Oncology*, vol. 158, no. 3, pp. 673–680, 2020.
- [15] E. K. Crane, J. Brown, A. Lehman, D. L. Tait, and R. W. Naumann, "Perioperative recovery and narcotic use in laparoscopic versus robotic surgery for endometrial cancer," *Journal of Minimally Invasive Gynecology*, vol. 28, no. 11, pp. 1898–1902, 2021.
- [16] F. Narducci, E. Bogart, T. Hebert et al., "Severe perioperative morbidity after robot-assisted versus conventional laparoscopy in gynecologic oncology: results of the randomized ROBOGYN-1004 trial," *Gynecologic Oncology*, vol. 158, no. 2, pp. 382–389, 2020.
- [17] J. Rodriguez, J. A. Rauh-Hain, J. Saenz et al., "Oncological outcomes of laparoscopic radical hysterectomy versus radical abdominal hysterectomy in patients with early-stage cervical cancer: a multicenter analysis," *International Journal of Gynecological Cancer*, vol. 31, no. 4, pp. 504–511, 2021.
- [18] S. E. Ferguson, T. Panzarella, S. Lau et al., "Prospective cohort study comparing quality of life and sexual health outcomes between women undergoing robotic, laparoscopic and open surgery for endometrial cancer," *Gynecologic Oncology*, vol. 149, no. 3, pp. 476–483, 2018.
- [19] M. J. Page, J. E. McKenzie, P. M. Bossuyt et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, article n71, 2021.
- [20] W. S. Ko, T. T. D. Tran, I. Bhattacharya, K. W. Ng, H. Sun, and C. Chang-Hasnain, "Illumination angle insensitive single indium phosphide tapered nanopillar solar cell," *Nano Letters*, vol. 15, no. 8, pp. 4961–4967, 2015.
- [21] J. D. Wright, W. M. Burke, E. T. Wilde et al., "Comparative effectiveness of robotic versus laparoscopic hysterectomy for endometrial cancer," *Journal of Clinical Oncology*, vol. 30, no. 8, pp. 783–791, 2012.
- [22] A. Obermair, M. Janda, J. Baker et al., "Improved surgical safety after laparoscopic compared to open surgery for apparent early stage endometrial cancer: results from a randomised controlled trial," *European Journal of Cancer*, vol. 48, no. 8, pp. 1147–1153, 2012.
- [23] J. Cardenas-Goicoechea, A. Shepherd, M. Momeni et al., "Survival analysis of robotic versus traditional laparoscopic surgical staging for endometrial cancer," *American Journal of Obstetrics and Gynecology*, vol. 210, no. 2, pp. 160.e1–160.e11, 2014.
- [24] G. Bogani, A. Cromi, S. Uccella et al., "Laparoscopic staging in women older than 75 years with early-stage endometrial cancer: comparison with open surgical operation," *Menopause*, vol. 21, no. 9, pp. 945–951, 2014.
- [25] G. Bogani, A. Cromi, S. Uccella et al., "Perioperative and long-term outcomes of laparoscopic, open abdominal, and vaginal surgery for endometrial cancer in patients aged 80 years or older," *International Journal of Gynecological Cancer*, vol. 24, no. 5, pp. 894–900, 2014.
- [26] A. Boosz, L. Haeberle, S. P. Renner et al., "Comparison of reoperation rates, perioperative outcomes in women with endometrial cancer when the standard of care shifts from open surgery to laparoscopy," *Archives of Gynecology and Obstetrics*, vol. 290, no. 6, pp. 1215–1220, 2014.
- [27] H. Y. Chiou, L. H. Chiu, C. H. Chen, Y. K. Yen, C. W. Chang, and W. M. Liu, "Comparing robotic surgery with laparoscopy and laparotomy for endometrial cancer management: a cohort study," *International Journal of Surgery*, vol. 13, pp. 17–22, 2015.
- [28] H. K. Park, I. B. Helenowski, E. Berry, J. R. Lurain, and N. L. Neubauer, "A comparison of survival and recurrence outcomes in patients with endometrial cancer undergoing robotic versus open surgery," *Journal of Minimally Invasive Gynecology*, vol. 22, no. 6, pp. 961–967, 2015.
- [29] G. Corrado, G. Cutillo, G. Pomati et al., "Surgical and oncological outcome of robotic surgery compared to laparoscopic and abdominal surgery in the management of endometrial cancer," *European Journal of Surgical Oncology*, vol. 41, no. 8, pp. 1074–1081, 2015.
- [30] S. Uccella, M. Bonzini, S. Palomba et al., "Impact of obesity on surgical treatment for endometrial cancer: a multicenter study comparing laparoscopy vs. open surgery, with propensity-matched analysis," *Journal of Minimally Invasive Gynecology*, vol. 23, no. 1, pp. 53–61, 2016.
- [31] L. H. Chu, W. C. Chang, and B. C. Sheu, "Comparison of the laparoscopic versus conventional open method for surgical staging of endometrial carcinoma," *Taiwanese Journal of Obstetrics & Gynecology*, vol. 55, no. 2, pp. 188–192, 2016.

- [32] A. Barrie, A. H. Freeman, L. Lyon et al., "Classification of post-operative complications in robotic-assisted compared with laparoscopic hysterectomy for endometrial cancer," *Journal of Minimally Invasive Gynecology*, vol. 23, no. 7, pp. 1181–1188, 2016.
- [33] G. Monterossi, F. Ghezzi, E. Vizza et al., "Minimally Invasive Approach in Type II Endometrial Cancer: Is It Wise and Safe," *Journal of minimally invasive gynecology*, vol. 24, pp. 438–445, 2017.
- [34] T. L. Beck, M. A. Schiff, B. A. Goff, and R. R. Urban, "Robotic, Laparoscopic, or Open Hysterectomy: Surgical Outcomes by Approach in Endometrial Cancer," *Journal of minimally invasive gynecology*, vol. 25, pp. 986–993, 2018.
- [35] X. C. Ruan, W. L. Wong, H. Q. Yeong, and Y. K. T. Lim, "Comparison of outcomes following laparoscopic and open hysterectomy with pelvic lymphadenectomy for early stage endometrial carcinoma," *Singapore medical journal*, vol. 59, pp. 366–369, 2018.
- [36] I. Deura, M. Shimada, Y. Azuma et al., "Comparison of laparoscopic surgery and conventional laparotomy for surgical staging of patients with presumed low-risk endometrial cancer: The current state of Japan," *Taiwanese journal of obstetrics & gynecology*, vol. 58, pp. 99–104, 2019.
- [37] S. L. Jørgensen, O. Mogensen, C. S. Wu, M. Korsholm, K. Lund, and P. T. Jensen, "Survival after a nationwide introduction of robotic surgery in women with early-stage endometrial cancer: a population-based prospective cohort study," *European journal of cancer (Oxford, England : 1990)*, vol. 109, pp. 1–11, 2019.
- [38] M. A. Vardar, U. K. Gulec, A. B. Guzel, D. Gumurdulu, G. Khatib, and G. Seydaoglu, "Laparoscopic surgery for low, intermediate and high-risk endometrial cancer," *Journal of gynecologic oncology*, vol. 30, no. e24, 2019.
- [39] W. Ghazali, S. A. Jamil, and I. A. Sharin, "Laparoscopic versus Laparotomy: Staging Surgery for Endometrial Cancer - Malaysia's Early Experience," *Gynecology and minimally invasive therapy*, vol. 8, pp. 25–29, 2019.
- [40] M. Dieterich, V. Schröter, J. Stubert, T. Reimer, B. Gerber, and A. Stachs, "Oncologic Outcome of Patients with (Low-Risk) Endometrial Carcinoma Undergoing Laparotomy versus Minimally Invasive Hysterectomy: A Retrospective Analysis," *Oncology research and treatment*, vol. 42, pp. 636–649, 2019.
- [41] A. Papadia, A. Garbade, M. L. Gasparri, J. Wang, and A. P. Radan, "Mueller MD Minimally invasive surgery does not impair overall survival in stage IIIC endometrial cancer patients," *Archives of gynecology and obstetrics*, vol. 301, pp. 585–590, 2020.
- [42] T. Tanaka, S. Ueda, S. Miyamoto et al., "Oncologic outcomes for patients with endometrial cancer who received minimally invasive surgery: a retrospective observational study," *International journal of clinical oncology*, vol. 25, pp. 1985–1994, 2020.
- [43] S. Togami, T. Kawamura, S. Yanazume, M. Kamio, and H. Kobayashi, "Comparison of survival outcomes between laparoscopic and open surgery in patients with low-risk endometrial cancer," *Japanese Journal of Clinical Oncology*, vol. 50, no. 11, pp. 1261–1264, 2020.
- [44] G. Siesto, F. Romano, N. P. Ieda, and D. Vitobello, "Survival outcomes after surgical management of endometrial cancer: Analysis after the first 10-year experience of robotic surgery in a single center," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 16, no. 6, pp. 1–9, 2020.
- [45] M. Koskas, M. Jozwiak, M. Fournier et al., "Long-term oncological safety of minimally invasive surgery in high-risk endometrial cancer," *European Journal of Cancer*, vol. 65, pp. 185–191, 2016.
- [46] Y. Meng, Y. Liu, S. Lin et al., "The effects of uterine manipulators in minimally invasive hysterectomy for endometrial cancer: A systematic review and meta-analysis," *European Journal of Surgical Oncology*, vol. 46, no. 7, pp. 1225–1232, 2020.
- [47] Z. Tuoheti, L. Han, and G. Mulati, "Laparo-endoscopic single-site surgery vs conventional laparoscopic surgery for endometrial cancer: A systematic review and meta-analysis," *Medicine*, vol. 100, no. 12, article e24908, 2021.
- [48] J. Wang, X. Li, H. Wu, Y. Zhang, and F. Wang, "A Meta-Analysis of Robotic Surgery in Endometrial Cancer: Comparison with Laparoscopy and Laparotomy," *Disease Markers*, vol. 2020, Article ID 2503753, 10 pages, 2020.
- [49] M. C. Cusimano, A. N. Simpson, F. Dossa et al., "Laparoscopic and robotic hysterectomy in endometrial cancer patients with obesity: a systematic review and meta-analysis of conversions and complications," *American Journal of Obstetrics and Gynecology*, vol. 221, no. 5, pp. 410–428.e19, 2019.

Research Article

GPC1 Is Associated with Poor Prognosis and Treg Infiltration in Colon Adenocarcinoma

Ying Liu ¹, Hui Ren,² Mu-qing Yang,¹ and Ji-yu Li ^{1,3}

¹Department of General Surgery, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai 200092, China

²School of Pharmacy, East China University of Science and Technology, Shanghai, China

³Geriatric Cancer Center, Huadong Hospital Affiliated to Fudan University, Shanghai 200040, China

Correspondence should be addressed to Ji-yu Li; lijiyu@fudan.edu.cn

Received 17 May 2022; Accepted 8 August 2022; Published 14 September 2022

Academic Editor: Lei Chen

Copyright © 2022 Ying Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

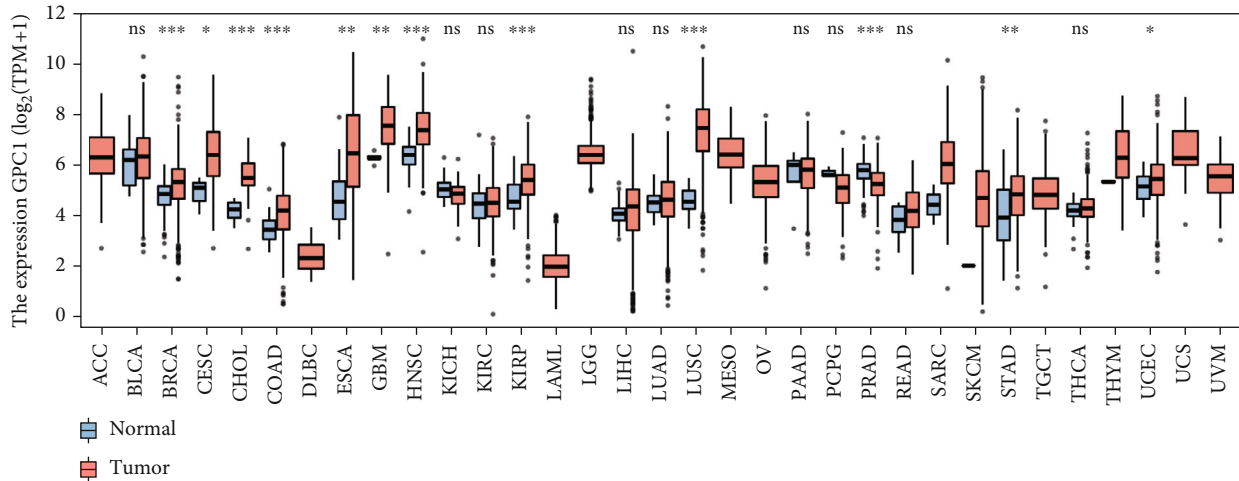
Glypican-1 (GPC1) is a glycosylated protein recognized as a promising biomarker for cancer. Nonetheless, there have been few systematic studies on GPC1 in colon adenocarcinoma (COAD). We conducted bioinformatic analysis based on The Cancer Genome Atlas (TCGA) and used clinical samples to verify that GPC1 is overexpressed in colon adenocarcinoma. Kaplan-Meier analysis showed that higher GPC1 expression was associated with poor overall survival (OS). The Cox regression model further showed that GPC1 expression is an independent negative prognostic factor for COAD. Gene set enrichment analysis demonstrated that multiple oncogenic signaling pathways were differentially enriched in GPC1 high- versus low-expressing COAD tumors, including DNA methylation, G2/M damage checkpoint, and telomere dysfunction. We observed a positive correlation between GPC1 expression and immune cell infiltration, such as regulatory T cells (Tregs), macrophages, and mast cells, and immunohistochemistry of 50 COAD tissues revealed that GPC1 expression was positively associated with Treg enrichment. Our results provide a promising candidate gene to predict the prognosis of COAD and new insights into tumor immunity. Further research is required to validate these results.

1. Introduction

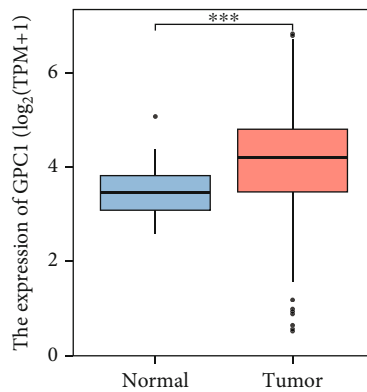
Colorectal cancer is the third most common cancer worldwide, accounting for 1.4 million new cases annually [1]. Most colon cancers are sporadic in nature. Western diets, chronic inflammation, environmental factors, and colonic polyps have been implicated in their pathogenesis. The most common type of colon cancer is colon adenocarcinoma (COAD), which accounts for more than 90% of colon cancers. In recent years, with the development of living standards and changes in diet, the incidence of colorectal cancer has increased. Despite advances in screening, surgery, and treatment, colorectal cancer remains the third leading cause of cancer-related deaths worldwide, with nearly 900,000 deaths in 2018 [2]. These observations underscore the need to identify new biological markers.

Glypican-1 (GPC1) is one of six members of the glycosaminoglycan-anchored cell surface glypican protein family [3, 4]. Glypicans are predominantly expressed during

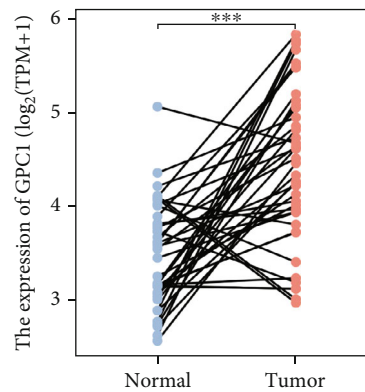
embryonic development and have been reported to play an important role in organ morphological development by influencing signaling pathways including Wnt, hedgehog, transforming growth factor- β , and fibroblast growth factor [5, 6]. Glypicans participate in many important processes, including cellular proliferation, migration, differentiation, extracellular matrix, and tumor microenvironment remodeling. Studies have suggested that aberrant expression of GPC1 is detected in multiple cancers and that disturbance of GPC1 influences cancer progression. GPC1 has been proven to be a useful biomarker for multiple cancerous tissues, such as prostate, hepatocellular, and pancreatic carcinomas [7–9]. A mouse model of sporadic colorectal cancer showed that mouse adenocarcinoma tissues contained much higher levels of GPC1 than normal tissues [10]. Papiewska-Pajak et al. demonstrated that exosomes released from the mouse colon adenocarcinoma cell line MC38, after stimulation by the transcription factor Snail, presented higher levels of GPC1 [11]. Nonetheless, there are few systematic studies



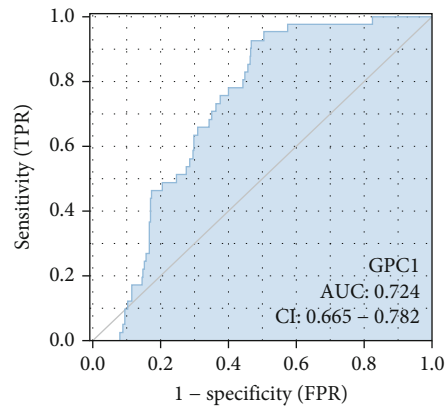
(a)



(b)



(c)



(d)

FIGURE 1: GPC1 expression is increased in COAD: (a) GPC1 mRNA expression of different cancers compared with corresponding normal tissues in TCGA database; (b, c) GPC1 mRNA expression in COAD and normal tissue (unpaired tissues (b) and paired tissues (c)) in TCGA database; (d) ROC curve to test the value of GPC1 to identify COAD tissues.

of GPC1 on the progression, prognosis, and immune response in COAD using bioinformatic approaches, especially in human colorectal cancer samples. This study is aimed at exploring the potential biological value of GPC1 in prognosis and immunity, thereby providing new insights into COAD.

2. Materials and Methods

2.1. Data Source and Comparison of GPC1 Expression. The Fragments Per Kilobase per Million (FPKM) RNAseq dataset from TCGA, containing 480 tumor samples, 41 normal samples, and their general and clinical data, was downloaded

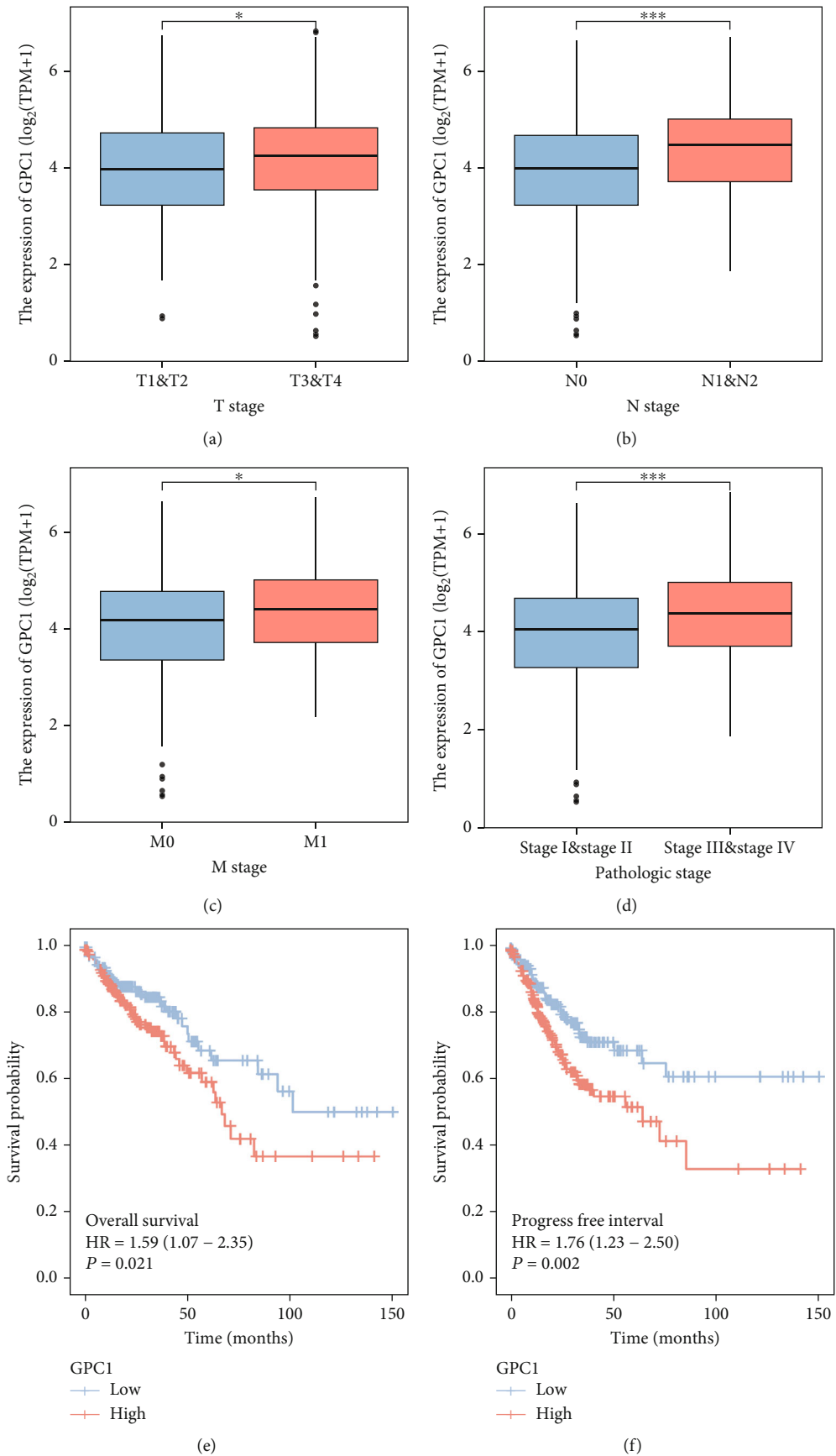


FIGURE 2: Continued.

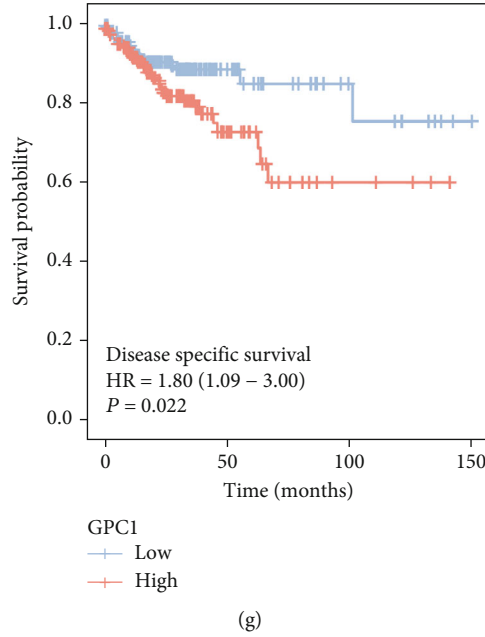


FIGURE 2: Associations between GPC1 expression and clinicopathological features and prognosis: (a–d) associations between GPC1 expression and T stage (a), N stage (b), M stage (c), and pathologic stage (d); (e–g) Kaplan-Meier analysis of OS (e), progress-free interval (f), and disease-specific survival (g) in COAD in TCGA database.

TABLE 1: Univariate and multivariate Cox hazard regression analyses of COAD in TCGA.

Variables	Total (n vs. n)	Univariate analysis		Multivariate analysis	
		HR (95% CI)	p value	HR (95% CI)	p value
Age (>60 vs. \leq 60)	477 (194 vs. 283)	1.610 (1.052-2.463)	0.028	2.553 (1.538-4.241)	<0.001
Gender (female vs. male)	477 (226 vs. 251)	1.101 (0.746-1.625)	0.627		
Race (Asian vs. Black or African-American vs. White)	306 (11 vs. 63 vs. 232)	0.927 (0.208-4.133)	0.921		
T stage (T1&T2 vs. T3&T4)	476 (94 vs. 382)	3.072 (1.423-6.631)	0.004	3.471 (1.066-11.305)	0.039
N stage (N0 vs. N1&N2)	477 (283 vs. 194)	2.592 (1.743-3.855)	<0.001	0.409 (0.152-1.097)	0.076
M stage (M0 vs. M1)	414 (348 vs. 66)	4.193 (2.683-6.554)	<0.001	2.264 (1.309-3.915)	0.003
Pathologic stage (I&II vs. III&IV)	466 (267 vs. 199)	2.947 (1.942-4.471)	<0.001	4.772 (1.528-14.900)	0.007
GPC1 ($\log_2(\text{TPM} + 1)$)	477	1.339 (1.114-1.609)	0.002	1.225 (1.002-1.497)	0.048

for expression analysis (<https://portal.gdc.cancer.gov/>). The tumor samples included 41 matched COAD and normal adjacent tissues. The transcript data were converted to transcripts per million (TPM) format data before analysis. The R pROC package was used for receiver operating characteristic (ROC) curve analysis.

The Ethics Committee of Shanghai Tenth People’s Hospital (no. 20KT95) approved the use of tissue microarray slides (Biotech, D100Co01, Xi’an, China) with 50 matched pairs of COAD samples and adjacent normal tissues for immunohistochemical (IHC) analysis. All patients were

pathologically diagnosed with COAD, and none of them had received preoperative therapy or had a history of other malignant tumors. Basic clinical information is summarized in Table S1.

2.2. Correlation between GPC1 Expression and Clinicopathological Parameters and Prognosis. Correlation analyses were performed with R statistical software using the “ggplot2” package. The patients were divided into two groups based on the median: high (above the median) and low (below the median). Disparities in survival by GPC1

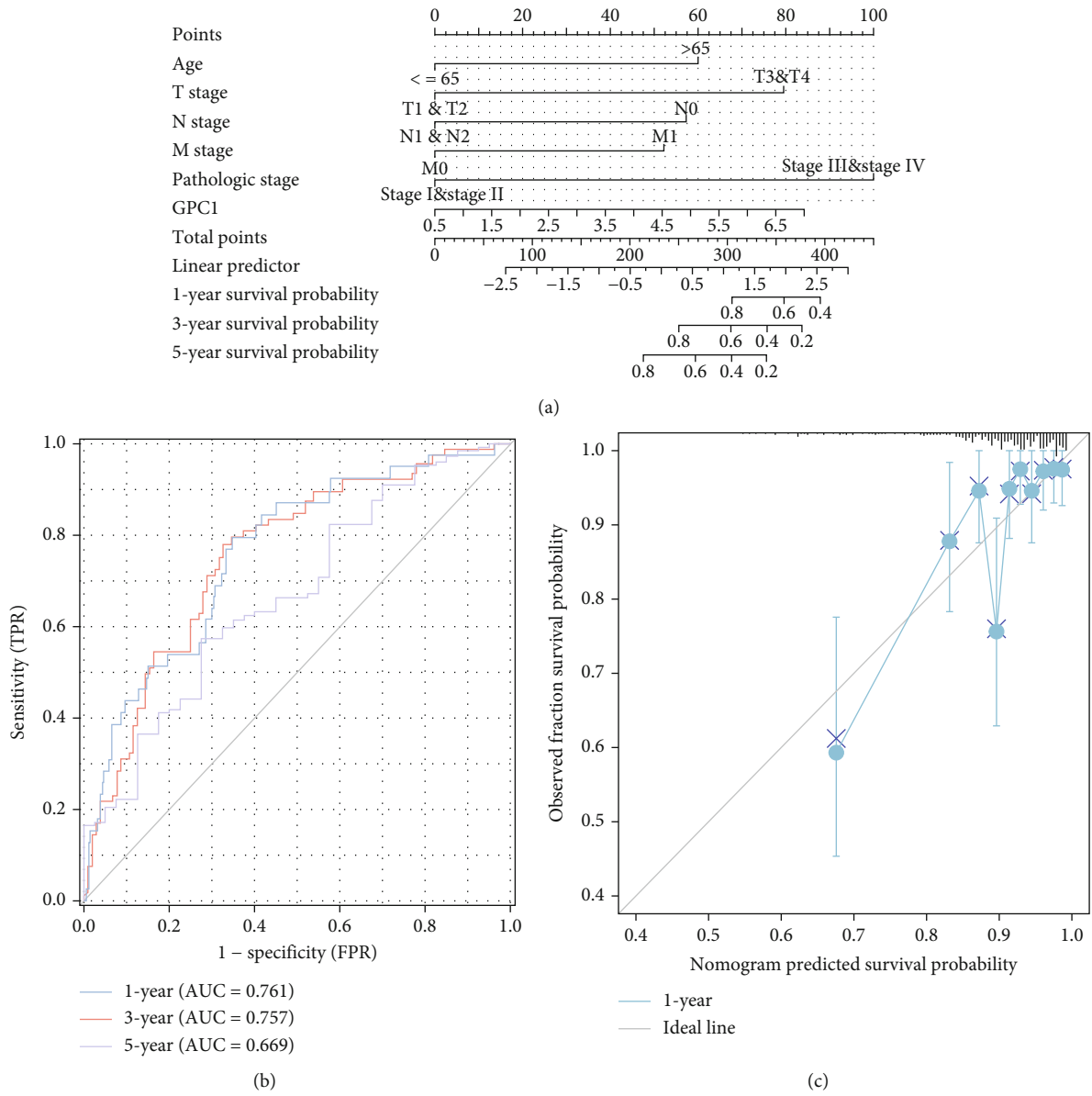


FIGURE 3: Continued.

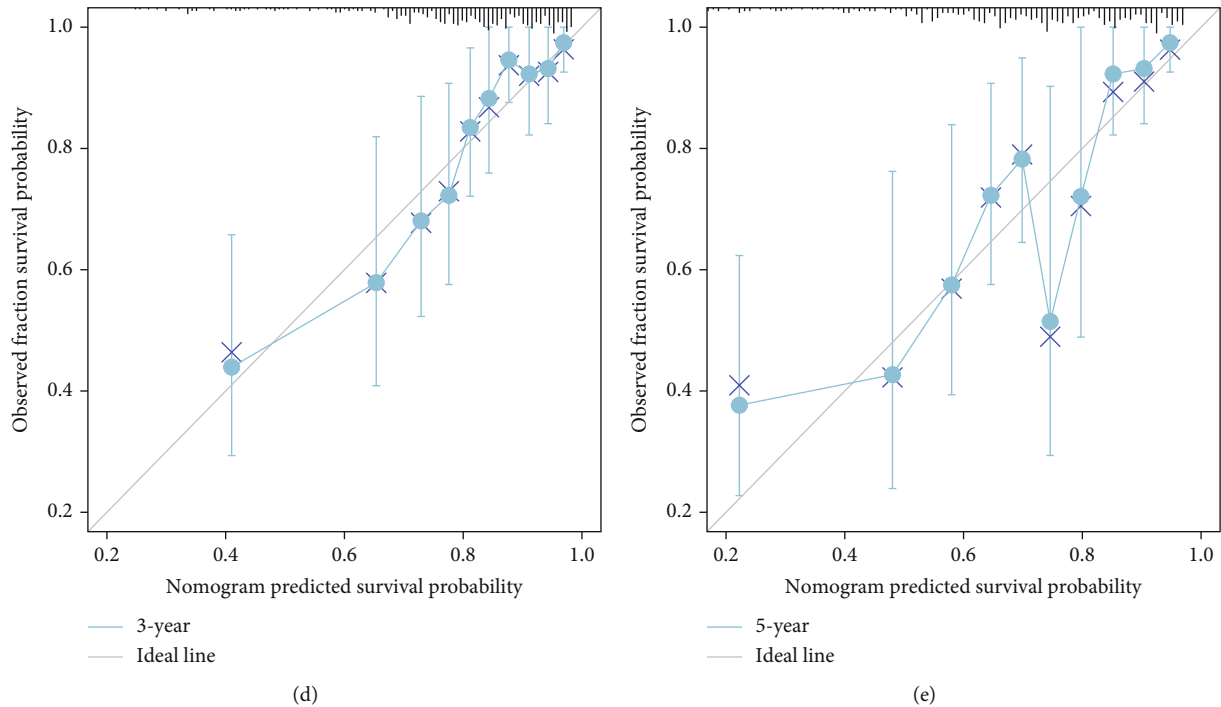


FIGURE 3: Prognostic value of the mRNA level of GPC1 in patients with COAD: (a) nomogram based on variables including gender; T, N, and M stage; and GPC1 expression in COAD from TCGA datasets; (b) 1-, 3-, and 5-year ROC curve; (c–e) 1-, 3-, and 5-year calibration curve.

expression and Kaplan-Meier (KM) curves were performed using the R package “survival” and “survminer.”

2.3. Univariate and Multivariate Cox Hazard Regression Analyses. Cox proportional hazard model analyses were generated using the R packages “survminer” and “survival.” In the univariate Cox hazard regression analysis, age, sex, race, T stage, N stage, M stage, pathologic stage, and GPC1 expression were set as independent variables, and overall survival (OS) was the dependent variable. A univariate Cox regression analysis was performed for each prognostic variable. To identify independent prognostic factors, all significant variables ($p < 0.1$) on univariate Cox regression analysis were further evaluated using multivariate Cox regression analysis.

2.4. Construction and Assessment of GPC1-Based Nomogram. Based on the results of the Cox hazard regression analysis, a nomogram was established using the rms package in R to analyze and visualize the multiple risk factors for predicting OS. Calibration plots were generated using the “rms” R package to show the predicted probabilities versus actual probabilities at multiple time points.

2.5. GPC1-Related Signaling Pathways according to Gene Set Enrichment Analysis (GSEA). Differential expression between the high and low groups based on the median was determined using the R package DESeq2. Using this procedure, we obtained the logFC and p value for each gene. GSEA was performed using the Cluster Profiler R package and the MSigDB collections. Only pathways with FDR q

values < 0.25 , $p_{\text{adjust}} < 0.05$, and $|\text{NES}| > 1$ were considered significantly enriched.

2.6. Correlation between GPC1 and Immunity. Single-sample GSEA was performed in the R package GSVA to analyze immune cell enrichment in the two groups. The R package was used to calculate the immune, stromal, and estimate scores.

2.7. IHC. Primary antibodies included anti-GPC1 (Proteintech, Rosemont, IL, USA) and anti-Foxp3 (Proteintech, Rosemont, IL). Three microscopic fields (magnification, $\times 40$) were randomly selected for analysis. The IHC score of GPC1 was calculated by multiplying the intensity of IHC staining (intensity score: 1-4) by the percentage of positive cells (percentage score: 1-4). The number of Foxp3-positive cells was counted under a high-power microscope. The mean values were used as the final results.

2.8. Statistical Analysis. Statistical analyses were performed using R statistical software (version 3.6.3). When normality and homogeneity of variance were met, Student’s t -test was used to analyze the differences between the two groups. The Wilcoxon rank-sum test was used when the variance did not satisfy the normality or homogeneity. Differences between two matched datasets were analyzed using paired Student’s t -test. Analysis of multiple groups was performed using the Kruskal-Wallis test. The influence of prognostic factors on survival was evaluated using the Cox proportional hazard regression model. Correlations were tested using Pearson’s method. If normality assumptions were not met,

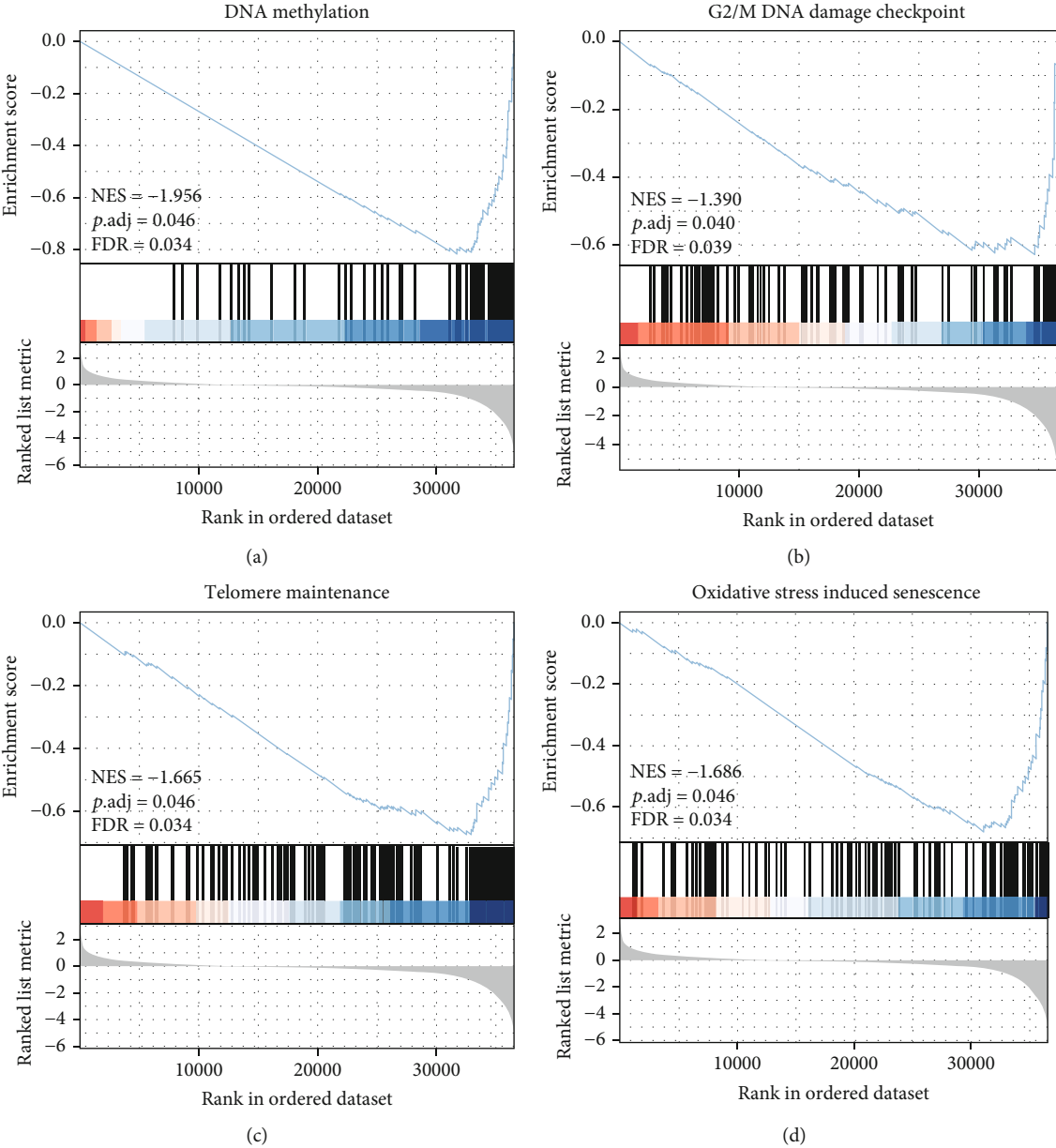


FIGURE 4: Continued.

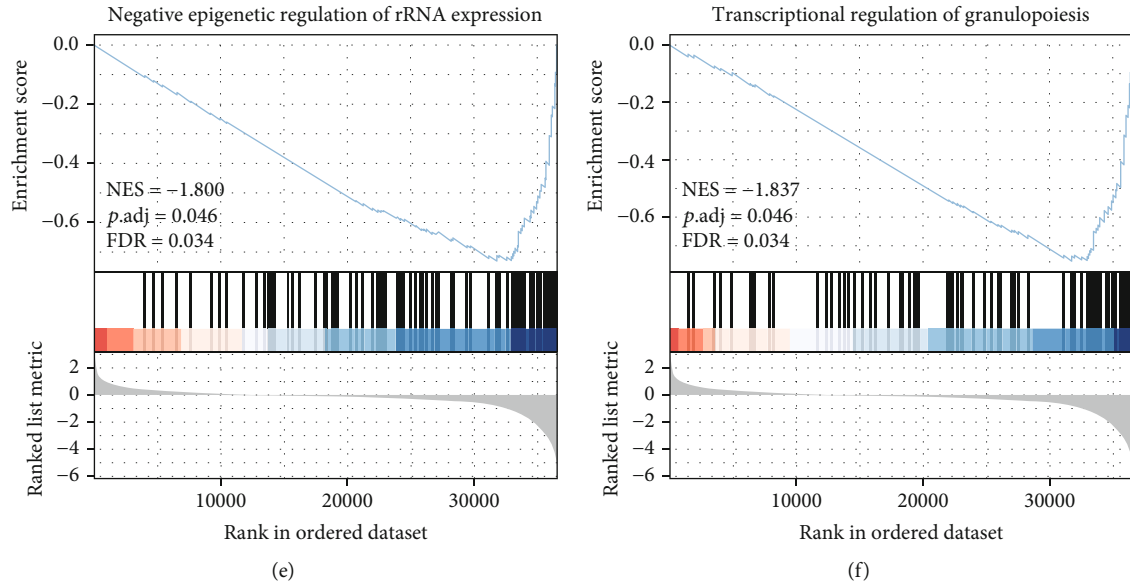


FIGURE 4: GSEA results based on GPC1 mRNA expression in COAD from TCGA datasets. (a–f) Significantly enriched pathway of DNA methylation, G2/M DNA damage checkpoint, telomere maintenance, oxidative stress-induced senescence, negative epigenetic regulation of rRNA expression, and transcriptional regulation of granulopoiesis.

Spearman's method was used. Correlations between GPC1 expression and immune cell infiltration, immune score, stromal score, and estimated score were assessed using the Spearman analysis. For all data, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$; ns means not significant.

3. Results

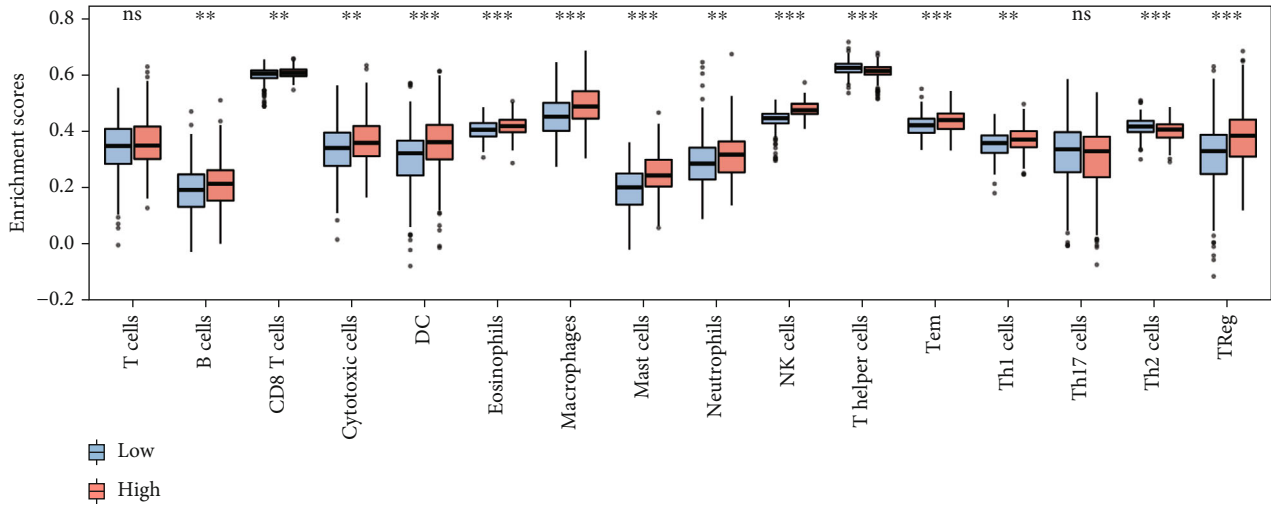
3.1. GPC1 Expression Is Elevated in COAD. Based on TCGA data, we found that GPC1 expression was upregulated in multiple cancers, including breast cancer, cervical cancer, bile duct cancer, colon cancer, glioblastoma, head and neck cancer, kidney papillary cell carcinoma, lung adenocarcinoma, and pancreatic cancer (Figure 1(a)). Next, we focused on COAD data from TCGA for further analysis. Compared to normal tissues, GPC1 mRNA was expressed at higher levels in COAD tissues (Figure 1(b), $N = 41$, $T = 41$, $p < 0.001$). We screened 41 paired COAD and normal tissues from TCGA data. Similarly, GPC1 expression was upregulated in tumor tissues (Figure 1(c), $p < 0.001$). The area under the ROC curve (AUC) was 0.724 (Figure 1(d)), indicating that GPC1 may potentially be able to distinguish COAD from normal tissue.

3.2. GPC1 Expression Correlates with Clinicopathological Characteristics and Poor Prognosis. Next, we evaluated whether GPC1 expression is associated with clinicopathological characteristics. We found that high GPC1 expression was associated with advanced T stage (Figure 2(a), $T1\&T2 = 94$, $T3\&T4 = 383$; $p = 0.038$), N stage (Figure 2(b), $N0 = 284$, $N1\&N2 = 194$, $p < 0.001$), M stage (Figure 2(c), $M0 = 349$; $M1 = 66$; $p = 0.04$), and pathologic stage (Figure 2(d), Stage I and Stage II = 268, Stage III and Stage IV = 199, $p < 0.001$), indicating that patients with more advanced tumor stage or higher grade tumors were more likely to have

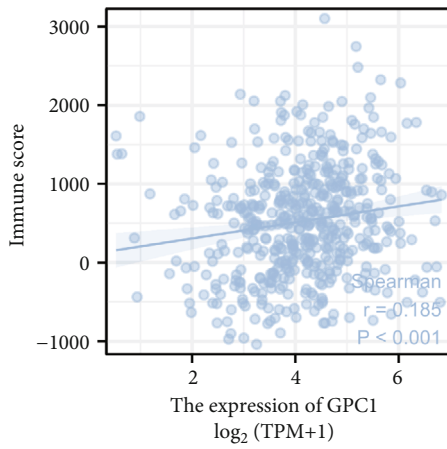
higher levels of GPC1. Next, we analyzed the correlation between GPC1 expression and prognosis using TCGA data. KM analyses revealed that patients with low GPC1 expression had longer OS (Figure 2(e), $p = 0.021$, HR = 1.59), progression-free interval (Figure 2(g), $p = 0.002$, HR = 1.76), and disease-specific survival (Figure 2(f), $p = 0.022$, HR = 1.80) than those with high expression. Two groups were divided based on the median (high = 239, low = 238).

Univariable and multivariable Cox proportional hazard regression analyses were performed to identify potential factors associated with the OS of patients with COAD. In univariable Cox proportional hazard regression analysis, OS was associated with age ($\leq 65 = 194$, $> 65 = 283$, $p = 0.028$, HR = 1.610), T stage ($T1\&T2 = 94$, $T3\&T4 = 382$, $p = 0.004$, HR = 3.072), N stage ($N0 = 283$, $N1\&N2 = 194$, $p < 0.001$, HR = 2.592), M stage ($M0 = 348$, $M1 = 66$, $p < 0.001$, HR = 4.193), pathologic stage (Stage I and Stage II = 267, Stage III & Stage IV = 199, $p < 0.001$, HR = 2.947), and GPC1 ($\log_2(\text{TPM} + 1)$, $p = 0.002$, HR = 1.339). Variables significant in the univariate Cox regression analysis ($p < 0.1$) were subjected to further multivariate Cox regression analysis. The results showed that OS was associated with age ($p < 0.001$, HR = 2.553), T stage ($p = 0.039$, HR = 3.471), M stage ($p = 0.003$, HR = 2.264), pathological stage ($p = 0.007$, HR = 1.225), and GPC1 ($p = 0.048$, HR = 1.225). Overall, age, T stage, M stage, pathological stage, and GPC1 expression were independent prognostic factors for patients with COAD (Table 1).

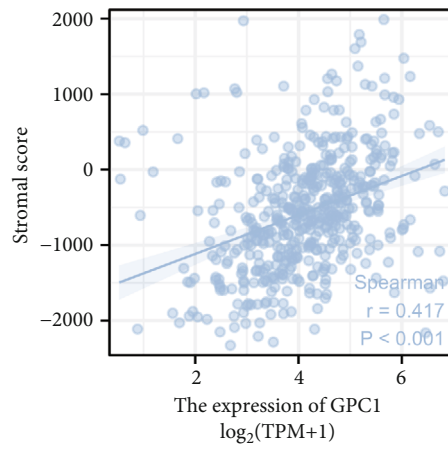
3.3. Establish and Evaluate the Prognostic Models Based on GPC1. Based on the independent factors identified in the Cox regression analyses, we created a nomogram for COAD patients to show the associations with OS of 1, 3, and 5 years (Figure 3(a)). ROC analysis was used to quantify the accuracy of the prognostic model. The AUCs at 1 year, 3 years,



(a)



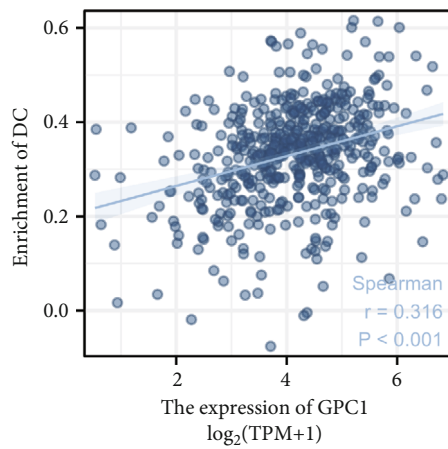
(b)



(c)

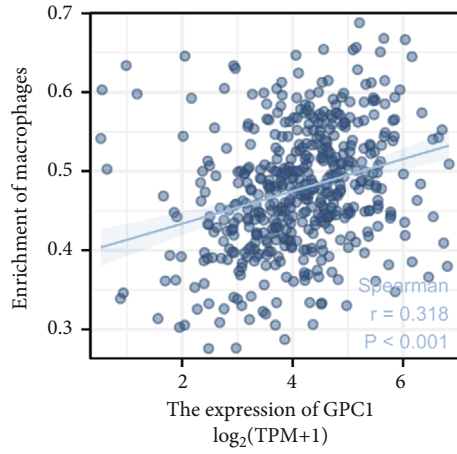


(d)

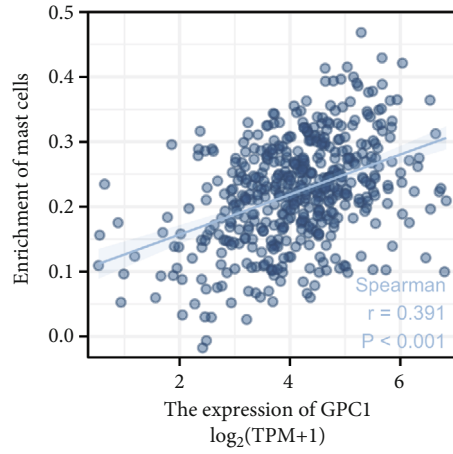


(e)

FIGURE 5: Continued.



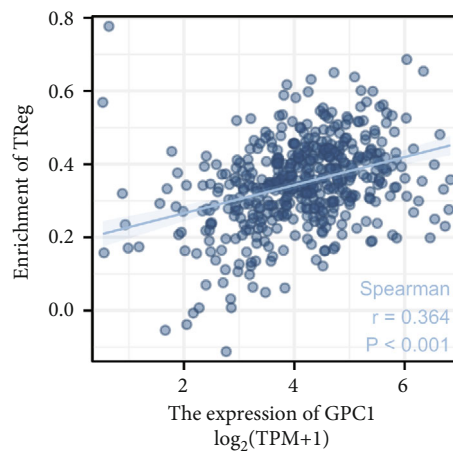
(f)



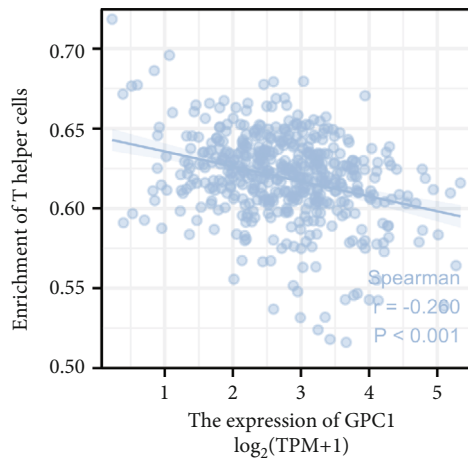
(g)



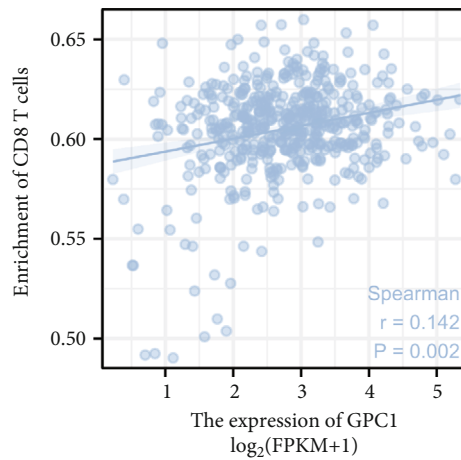
(h)



(i)



(j)



(k)

FIGURE 5: Continued.

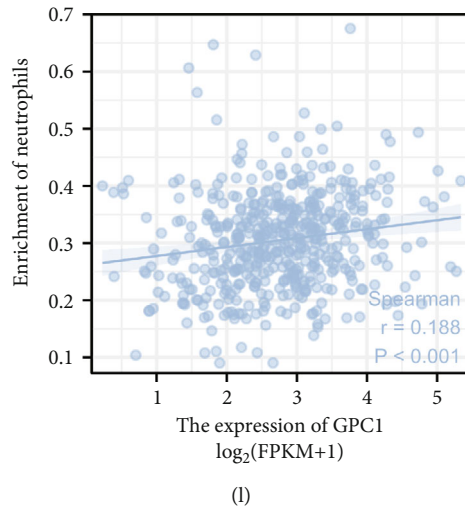


FIGURE 5: GPC1 is associated with proportions of immune cells: (a) immune cell distribution in patients with high GPC1 expression level and low GPC1 expression level; (b–d) associations between the expression of GPC1 and the immune score, stromal score, and estimate score; (e–l) associations between the expression of GPC1 and enrichment of immune cell (DC, macrophage, mast cells, NK cells, Tregs, T helper cells, CD8 T cells, and neutrophils).

and 5 years were 0.761, 0.757, and 0.669, respectively (Figure 3(b)), indicating moderate predictive power for 1 year and 3 years. The c-index for the multivariable model was 0.758, which also implies moderate predictive power. Moreover, the calibration plot was close to the diagonal, confirming that they were close to the ideal (Figures 3(c)–3(e)).

3.4. GPC1-Related Signaling Pathways according to GSEA. To explore the biological functions of GPC1, we investigate GPC1-associated signaling pathways enriched differently between GPC1-high and GPC1-low patients by GSEA. We found that several pathways showed significant enrichment, such as DNA methylation, G2/M DNA damage checkpoint, telomere maintenance, oxidative stress-induced senescence, negative epigenetic regulation of rRNA expression, and transcriptional regulation of granulopoiesis signaling pathways (Figures 4(a)–4(f)).

3.5. GPC1 Is Related to Immune Cell Infiltration in COAD. The tumor immune microenvironment and its various cellular components play a crucial role in cancer progression. Thus, we explored the correlation between GPC1 expression and tumor immunity based on TCGA data. COAD with high GPC1 expression presented higher infiltration levels of B cells, CD8 +T cells, cytotoxic cells, dendritic cells (DC), eosinophils, macrophages, mast cells, neutrophils, natural killer (NK) cells, effective memory T (Tem) cells, TH1 cells, and Treg cells and lower infiltration levels of T helper cells and Th2 cells (Figure 5(a)). Spearman correlation revealed that the expression of GPC1 was positively correlated with the immune score (Figure 5(b), $r = 0.185, p < 0.001$), stromal score (Figure 5(c), $r = 0.417, p < 0.001$), and estimated score (Figure 5(d), $r = 0.328, p < 0.001$). In other words, the higher the level of GPC1, the higher the degree of immune and stromal infiltration. Moreover, GPC1 expression was positively associated with the enrichment of DC (Figure 5(e), $r = 0.316, p < 0.001$), macrophages

(Figure 5(f), $r = 0.318, p < 0.001$), mast cells (Figure 5(g), $r = 0.391, p < 0.001$), NK cells (Figure 5(h), $r = 0.659, p < 0.001$), and Treg cells (Figure 5(i), $r = 0.364, p < 0.001$). However, no significant relationship was found between GPC1 expression and the enrichment of T helper cells (Figure 5(j), $r = 0.260, p < 0.001$), CD8 T cells (Figure 5(k), $r = 0.142, p = 0.002$), and neutrophils (Figure 5(l), $r = 0.188, p < 0.001$).

3.6. Validation Using Clinical Specimens. IHC was performed on 50 matched pairs of COAD and normal tissue specimens. Representative IHC staining images of GPC1 are shown in Figure 6(a). COAD exhibited an elevated IHC score for GPC1 compared to matched normal tissues (Figure 6(b), $n = 50, p < 0.001$). Next, we analyzed the correlation between the clinical characteristics of patients and the IHC score of GPC1. Unfortunately, no correlation was observed in our specimens (Figures 6(c)–6(f)). We also performed IHC staining for Foxp3 in all tumor samples to label Tregs. GPC1 and Foxp3 staining was conducted on serial sections, and a schematic diagram is depicted in Figure 6(g). The patients were divided into the low- and high-expression groups based on the median IHC score. The comparison indicated that higher GPC1 expression tended to be associated with a higher frequency of Foxp3+ cells (Figure 6(h), $n = 25, p = 0.001$). Spearman's correlation analysis showed that the IHC score of GPC1 staining was positively associated with Foxp3+ cell frequency in COAD (Figure 6(i), $r = 0.439, p = 0.001$).

4. Discussion

Previous studies have demonstrated abnormal GPC1 expression in tumor tissues of pancreatic ductal adenocarcinoma and prostate cancer, but GPC1 was nearly absent in normal and adjacent tissues [12, 13]. In cancer tissues of esophageal squamous cell carcinomas, cervical cancers, and glioma, GPC1 expression levels were higher than those in normal

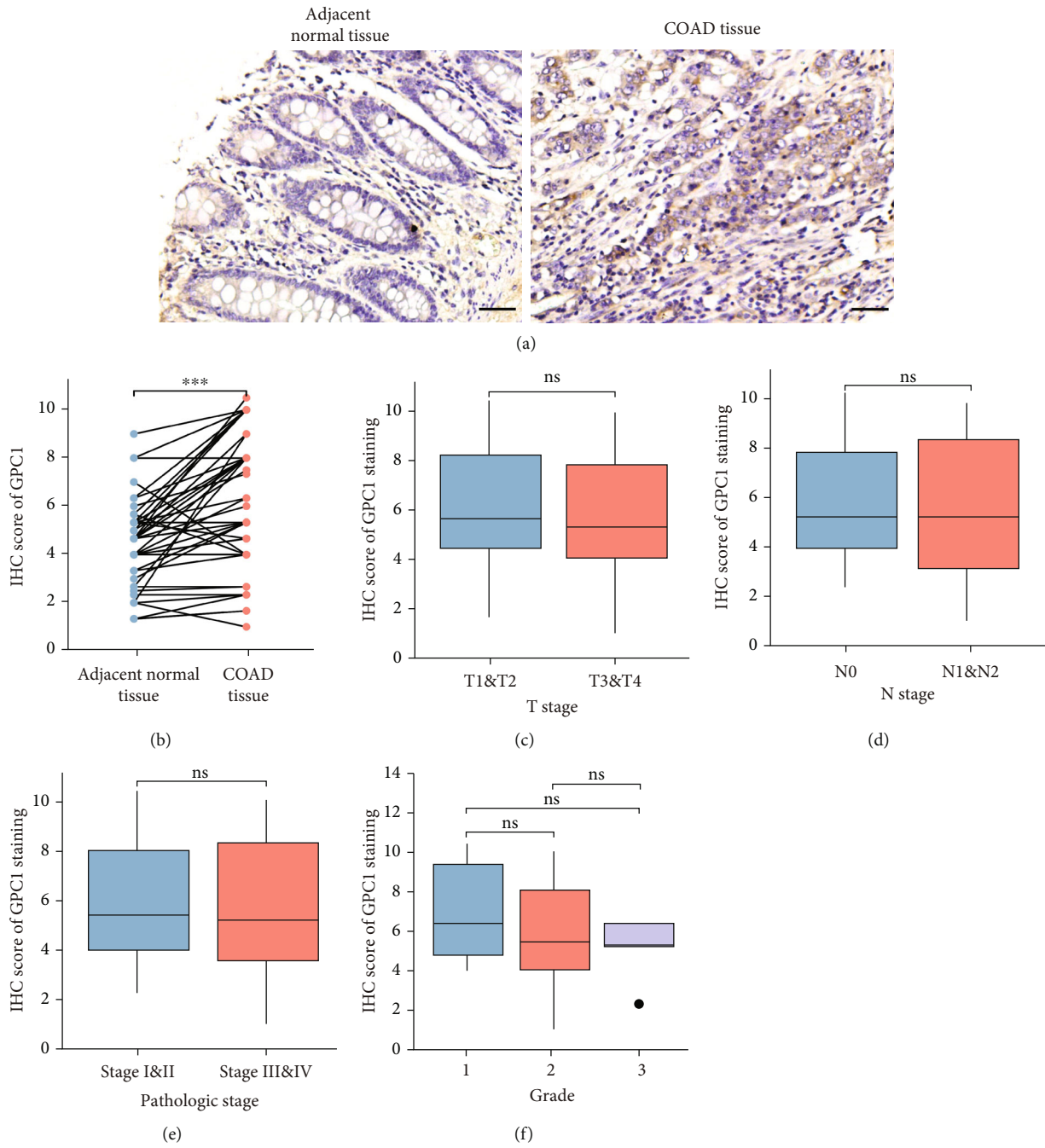


FIGURE 6: Continued.

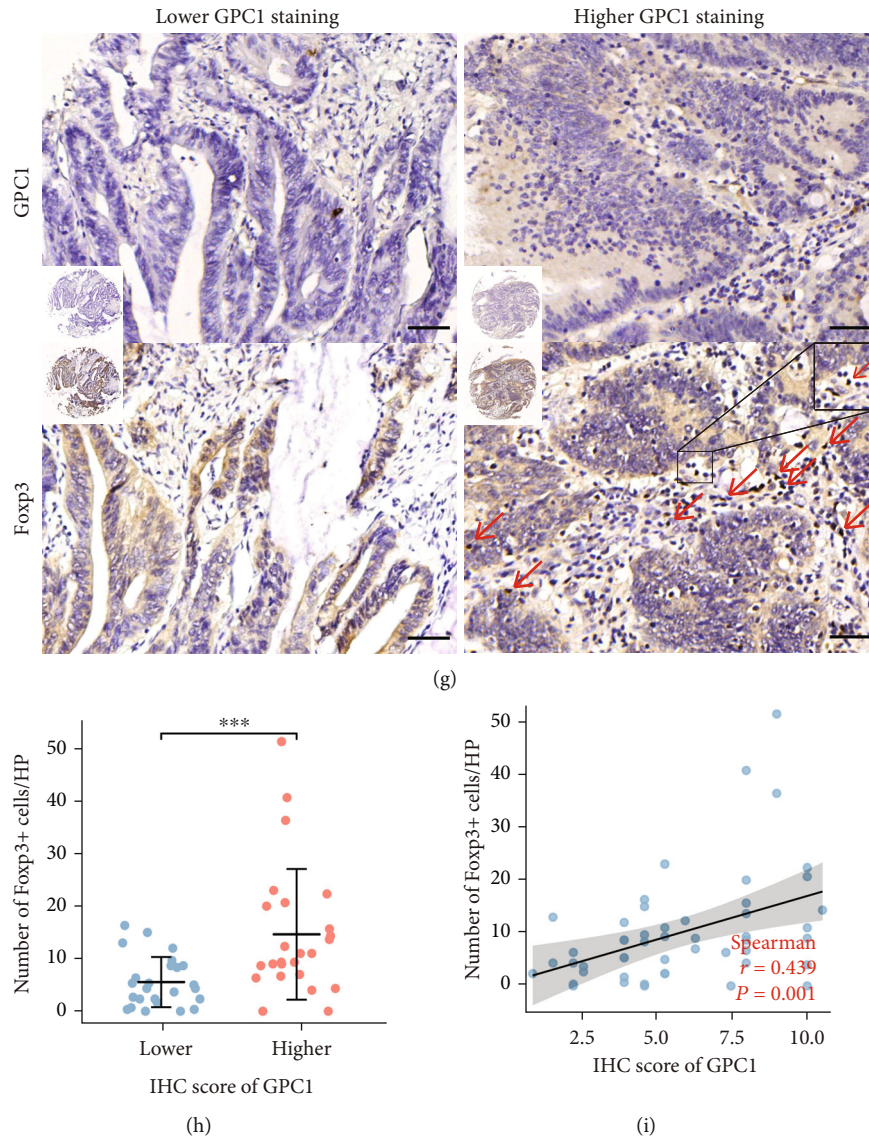


FIGURE 6: Validation using clinical specimens: (a) representative images of IHC staining of GPC1 in COAD and paired adjacent normal tissues; (b) IHC score of GPC1 of COAD compared with corresponding adjacent normal tissues; (c–f) associations between GPC1 IHC score and T stage ((c) T1&T2 = 12, T3&T4 = 38; $p = 0.638$), N stage ((d) N0 = 35, N1&N2 = 15, $p = 0.782$), pathologic stage ((e) Stage I & Stage II = 37, Stage III & Stage IV = 16; $p = 0.961$), and pathologic grade ((f) Grade 1 = 6, Grade 2 = 39, and Grade 3 = 5; $p = 0.892$); (g) representative images of IHC staining of GPC1 and Foxp3 in COAD; (h) number of Foxp3+ cells per high power field (HP) in GPC1 high- versus low-expressing groups; (i) associations between GPC1 IHC score and number of Foxp3+ cells per HP. Black scale bar = 50 μm . Red arrow points to Foxp3+ cells.

tissues [14–16]. Other investigations have recently suggested that GPC1 expression is a prognostic biomarker for pancreatic ductal adenocarcinoma [12, 17]. As GPC1 also has a secreted soluble form, it has been reported to be detected in serum exosomes and is promising for the early detection of pancreatic cancer [18]. Because its plasma, serum, and urine levels are elevated in patients with prostate carcinoma, GPC1 might serve as a reliable diagnostic marker for prostate cancer [19, 20]. These results suggest that GPC1 has a potential clinical value as a predictive and prognostic biomarker for several cancers. However, the role of GPC1 in patients has rarely been studied. A previous study showed that plasma GPC1+ exosomes were significantly increased

in colorectal cancer patients compared to healthy patients and reduced after surgery [21]. Notably, one disadvantage of serum is that its composition is source-dependent, and serum GPC1 may be confounded by other sources, as GPC1 is widely expressed in mammalian tissues. In addition, the release of GPC1 relies on protease notum [4], which has not been fully studied in normal and cancerous tissues. Concurrently, IHC analysis of 50 paired human tissues confirmed higher GPC1 protein expression in COAD tissues. These results suggest that GPC1 may be a potential clinical biomarker and exert active biological functions in COAD.

Based on TCGA data, we found that GPC1 was associated with the tumor stage. However, our external validation

did not support this view. In univariate and multivariate Cox regression analyses, the GPC1 expression level was an independent negative prognostic factor. To validate our results, we established a nomogram and calibration models to predict the survival probabilities of patients with COAD. These results, including the ROC curve, showed that our model had a moderate ability to predict the prognosis.

GSEA was performed to explore differentially expressed genes and pathway enrichment in GPC1 high- versus low-expressing COAD tumors in TCGA dataset. Results showed that multiple oncogenic signaling pathways were found to be aberrantly expressed. DNA methylation is an epigenetic process that modulates gene expression. Aberrant DNA methylation is one of the earliest and most universal signatures of cancer [22, 23]. Multiple biomarker-based DNA methylations have been used in the prediction, diagnosis, and prediction of therapeutic approaches for colorectal cancer [24]. The G2/M damage checkpoint prevents cells from entering M phase and promotes DNA damage repair processes, which is important for malignant transformation [25]. Chromosome instability triggered by telomere dysfunction is considered a promoter of tumorigenesis. Studies have revealed that telomeres in human colon cancer are often shorter than those in normal tissues, which is a key driving event in colorectal carcinomas [26, 27].

The composition of infiltrating immune cells influences the immune status of the tumor microenvironment. Tumors evade the immune system by establishing an immunosuppressive tumor microenvironment. Tregs are a predominant subset of suppressor T cells that play a vital role in tumor tolerance. Tregs exert immunosuppressive functions through the secretion of inflammatory cytokines such as IL-10, IL-35, and TGF- β , inducing target immune cell death by granzyme B and negatively regulating DC maturation and self-antigen presentation via negative costimulatory molecules [28–30]. Although the prognostic value of Foxp3+ Tregs in COAD is controversial [31], some studies have shown that the role of Treg cells in the tumor microenvironment is unclear and that Tregs may play a role in promoting antitumor immunity. Research on colorectal cancer has demonstrated that increased frequencies of Treg cells are often correlated with tumor immune evasion and poor prognosis of patients [32]. Because we found that higher GPC1 expression indicated a worse prognosis in COAD patients, we speculated that the frequency of intratumoral Treg cells correlated positively with GPC1 expression. However, we have only demonstrated the correlation between the two and have not demonstrated that GPC1 has a chemotactic effect on Treg cells or could regulate their proliferation. Previous research has shown that TGF- β signaling is fundamental for both Th17 and Treg cell differentiation. GPC1 modulates various signaling pathways, including TGF- β . In the future, we hope to explore whether GPC1 directly affects the infiltration of Treg cells and its mechanism. Our findings on the potential association between GPC1 and Tregs provide direction for future COAD studies.

Some of the data analyzed in this study were derived from public databases, and the present study has some limitations. Further research is needed to validate our findings and investigate the biological functions of GPC1 in COAD.

5. Conclusions

In this study, we addressed the role of GPC1 in prognosis and immune cell infiltration in colon cancer for the first time based on TCGA datasets. GPC1 is overexpressed in colon adenocarcinoma and is associated with poor OS. GPC1 expression is an independent negative prognostic factor for COAD, and DNA methylation, the G2/M damage checkpoint, and telomere dysfunction were differentially enriched in GPC1 high- versus low-expressing COAD tumors. In addition, we observed a positive correlation between GPC1 expression and immune cell infiltration based on TCGA dataset. The IHC results of clinical samples revealed that GPC1 expression was positively associated with the enrichment of Tregs. However, further in vivo and in vitro evidence is needed to elucidate that GPC1 may participate in immune cell infiltration and influence prognosis.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the following grant: National Natural Science Foundation of China (No. 82072647 and No. 81970565).

Supplementary Materials

See Table S1 in the Supplementary Materials for comprehensive analysis. (*Supplementary Materials*)

References

- [1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA: a Cancer Journal for Clinicians*, vol. 65, no. 2, pp. 87–108, 2015.
- [2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [3] J. Filmus, M. Capurro, and J. Rast, "Glypicans," *Genome Biology*, vol. 9, no. 5, p. 224, 2008.
- [4] J. Filmus and S. B. Selleck, "Glypicans: proteoglycans with a surprise," *The Journal of Clinical Investigation*, vol. 108, no. 4, pp. 497–501, 2001.
- [5] X. Lin, "Functions of heparan sulfate proteoglycans in cell signaling during development," *Development*, vol. 131, no. 24, pp. 6009–6021, 2004.
- [6] S. Sarrazin, W. C. Lamanna, and J. D. Esko, "Heparan sulfate proteoglycans," *Cold Spring Harbor Perspectives in Biology*, vol. 3, no. 7, 2011.

- [7] L. Duan, X. Q. Hu, D. Y. Feng, S. Y. Lei, and G. H. Hu, "GPC-1 may serve as a predictor of perineural invasion and a prognosticator of survival in pancreatic cancer," *Asian Journal of Surgery*, vol. 36, no. 1, pp. 7–12, 2013.
- [8] J. Y. Wang, X. K. Wang, G. Z. Zhu et al., "Distinct diagnostic and prognostic values of glypicans gene expression in patients with hepatocellular carcinoma," *BMC Cancer*, vol. 21, no. 1, p. 462, 2021.
- [9] Q. Truong, I. O. Justiniano, A. L. Nocon et al., "Glypican-1 as a biomarker for prostate cancer: isolation and characterization," *Journal of Cancer*, vol. 7, no. 8, pp. 1002–1009, 2016.
- [10] M. De Robertis, M. Arigoni, L. Loiacono et al., "Novel insights into notum and glypicans regulation in colorectal cancer," *Oncotarget*, vol. 6, no. 38, pp. 41237–41257, 2015.
- [11] I. Papiewska-Pajak, D. Krzyzanowski, M. Katela et al., "Glypican-1 level is elevated in extracellular vesicles released from MC38 colon adenocarcinoma cells overexpressing Snail," *Cells*, vol. 9, no. 7, 2020.
- [12] H. Lu, F. Niu, F. Liu, J. Gao, Y. Sun, and X. Zhao, "Elevated glypican-1 expression is associated with an unfavorable prognosis in pancreatic ductal adenocarcinoma," *Cancer Medicine*, vol. 6, no. 6, pp. 1181–1191, 2017.
- [13] A. V. Suhovskih, L. A. Mostovich, I. S. Kunin et al., "Proteoglycan expression in normal human prostate tissue and prostate cancer," *ISRN Oncology*, vol. 2013, Article ID 680136, 9 pages, 2013.
- [14] G. Su, K. Meyer, C. D. Nandini, D. Qiao, S. Salamat, and A. Friedl, "Glypican-1 is frequently overexpressed in human gliomas and enhances FGF-2 signaling in glioma cells," *The American Journal of Pathology*, vol. 168, no. 6, pp. 2014–2026, 2006.
- [15] H. Hara, T. Takahashi, S. Serada et al., "Overexpression of glypican-1 implicates poor prognosis and their chemoresistance in oesophageal squamous cell carcinoma," *British Journal of Cancer*, vol. 115, no. 1, pp. 66–75, 2016.
- [16] S. Matsuzaki, S. Serada, K. Hiramatsu et al., "Anti-glypican-1 antibody-drug conjugate exhibits potent preclinical antitumor activity against glypican-1 positive uterine cervical cancer," *International Journal of Cancer*, vol. 142, no. 5, pp. 1056–1066, 2018.
- [17] C. Y. Zhou, Y. P. Dong, X. Sun et al., "High levels of serum glypican-1 indicate poor prognosis in pancreatic ductal adenocarcinoma," *Cancer Medicine*, vol. 7, no. 11, pp. 5525–5533, 2018.
- [18] J. Y. Qian, Y. L. Tan, Y. Zhang, Y. F. Yang, and X. Q. Li, "Prognostic value of glypican-1 for patients with advanced pancreatic cancer following regional intra-arterial chemotherapy," *Oncology Letters*, vol. 16, no. 1, pp. 1253–1258, 2018.
- [19] D. H. Campbell, M. E. Lund, A. L. Nocon et al., "Detection of glypican-1 (GPC-1) expression in urine cell sediments in prostate cancer," *PLoS One*, vol. 13, no. 4, article e0196017, 2018.
- [20] R. A. Levin, M. E. Lund, Q. Truong et al., "Development of a reliable assay to measure glypican-1 in plasma and serum reveals circulating glypican-1 as a novel prostate cancer biomarker," *Oncotarget*, vol. 9, no. 32, pp. 22359–22367, 2018.
- [21] J. Li, Y. Chen, X. Guo et al., "GPC1 exosome and its regulatory miRNAs are specific markers for the detection and target therapy of colorectal cancer," *Journal of Cellular and Molecular Medicine*, vol. 21, no. 5, pp. 838–847, 2017.
- [22] C. A. Demetriou, K. Van Veldhoven, C. Relton, S. Stringhini, K. Kyriacou, and P. Vineis, "Biological embedding of early-life exposures and disease risk in humans: a role for DNA methylation," *European Journal of Clinical Investigation*, vol. 45, no. 3, pp. 303–332, 2015.
- [23] J. Smith, S. Sen, R. J. Weeks, M. R. Eccles, and A. Chatterjee, "Promoter DNA hypermethylation and paradoxical gene activation," *Trends in Cancer*, vol. 6, no. 5, pp. 392–406, 2020.
- [24] J. M. Yi, "DNA methylation change profiling of colorectal disease: screening towards clinical use," *Life (Basel)*, vol. 11, no. 5, p. 412, 2021.
- [25] M. De Marco Zompit and M. Stucki, "Mechanisms of genome stability maintenance during cell division," *DNA Repair (Amst)*, vol. 108, article 103215, 2021.
- [26] C. Garcia-Aranda, C. De Juan, A. Diaz-Lopez et al., "Correlations of telomere length, telomerase activity, and telomeric-repeat binding factor 1 expression in colorectal carcinoma," *Cancer*, vol. 106, no. 3, pp. 541–551, 2006.
- [27] E. Rampazzo, R. Bertorelle, L. Serra et al., "Relationship between telomere shortening, genetic instability, and site of tumour origin in colorectal cancers," *British Journal of Cancer*, vol. 102, no. 8, pp. 1300–1305, 2010.
- [28] S. Sakaguchi, M. Miyara, C. M. Costantino, and D. A. Hafler, "FOXP3⁺ regulatory T cells in the human immune system," *Nature Reviews. Immunology*, vol. 10, no. 7, pp. 490–500, 2010.
- [29] C. M. Paluskiewicz, X. Cao, R. Abdi, P. Zheng, Y. Liu, and J. S. Bromberg, "T regulatory cells and priming the suppressive tumor microenvironment," *Frontiers in Immunology*, vol. 10, p. 2453, 2019.
- [30] W. J. Grossman, J. W. Verbsky, W. Barchet, M. Colonna, J. P. Atkinson, and T. J. Ley, "Human T regulatory cells can use the perforin pathway to cause autologous target cell death," *Immunity*, vol. 21, no. 4, pp. 589–601, 2004.
- [31] S. Yaqub, K. Henjum, M. Mahic et al., "Regulatory T cells in colorectal cancer patients suppress anti-tumor immune activity in a COX-2 dependent manner," *Cancer Immunology, Immunotherapy*, vol. 57, no. 6, pp. 813–821, 2008.
- [32] S. Ladoire, F. Martin, and F. Ghiringhelli, "Prognostic role of FOXP3⁺ regulatory T cells infiltrating human carcinomas: the paradox of colorectal cancer," *Cancer Immunology, Immunotherapy*, vol. 60, no. 7, pp. 909–918, 2011.

Research Article

Comprehensive Analysis of the Immune Implication of AKAP12 in Stomach Adenocarcinoma

Zhiqian Xu ¹, Ling Xiang ², Linglong Peng ¹, Haitao Gu ¹, and Yaxu Wang ¹

¹Department of Gastrointestinal Surgery, The Second Affiliated Hospital of Chongqing Medical University, Chongqing 400010, China

²Department of Clinical Nutrition, The Second Affiliated Hospital of Chongqing Medical University, Chongqing 400010, China

Correspondence should be addressed to Yaxu Wang; 300897@hospital.cqmu.edu.cn

Received 12 July 2022; Revised 16 August 2022; Accepted 20 August 2022; Published 13 September 2022

Academic Editor: Tao Huang

Copyright © 2022 Zhiqian Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A kinase anchor protein 12 (AKAP12) as a tumor suppressor in various cancers has been extensively studied and confirmed. However, its immune implication in stomach adenocarcinoma (STAD) remains uncertain. Here, using The Cancer Genome Atlas (TCGA), Human Protein Atlas (HPA), Tumor Immune Estimation Resource (TIMER), Cancer Cell Line Encyclopedia (CCLE), integrated repository portal for tumor-immune system interactions (TISIDB), and Search Tool for the Retrieval of Interaction Gene/Proteins (STRING) database, we systematically analyzed the immune correlation of AKAP12 from three aspects including immune infiltration cells, immune-related pathways, and immunomodulators and developed a AKAP12-related 4-gene signature for prognosis prediction. Our results showed that AKAP12 mRNA and protein levels were downregulated in STAD patients, and its expression was positively related to CD4+ T cells and macrophages. In addition, the immune cell infiltration levels were associated with AKAP12 gene copy number deletion in STAD. Based on CCLE database, we found that AKAP12 coexpressed genes were enriched in several immune- and cancer-related pathways, which was further validated by Gene Set Enrichment Analysis (GSEA). Moreover, we identified 46 immunomodulators that were significantly related to AKAP12 expression using TISIDB database, and these immunomodulators were involved in immune-related pathways including Th17 cell differentiation and natural killer cell-mediated cytotoxicity. Additionally, based on the 46 AKAP12-related immunomodulators, a 4-gene risk prediction signature was developed using the Cox regression model. The risk signature was identified as an independent prognostic factor, which can accurately predict the prognosis of patients with STAD, showing good predictive performance. Furthermore, we constructed a prognostic nomogram and calibration to predict and assess patient survival probabilities by integrating the risk score and other clinical factors. In conclusion, our study provides strong evidence that AKAP12 is closely related to tumor immunity in STAD from three aspects: immune infiltration cells, immune pathways, and immunomodulators. More importantly, the AKAP12-related prognostic signature may have a good application prospect for clinical practice.

1. Introduction

Stomach adenocarcinoma (STAD) is the fifth most common malignant tumor and one of the main causes of cancer deaths, accounting for 7.7% of all cancer deaths [1]. Its occurrence is a progressive process involving multiple steps and multiple factors, such as environmental diet, helicobacter pylori infection, heredity, and precancerous state. In the early stage of STAD, there are only some unclear upper gastrointestinal symptoms, such as abdominal pain, anorexia, and

mild anemia. With the progress of the disease, pyloric or cardiac obstruction may occur, and complications such as upper gastrointestinal hemorrhage and perforation may also occur. Local masses, ascites, and supraclavicular lymph node enlargement can be seen in advanced STAD. At present, surgery is the main treatment for gastric adenocarcinoma. Chemotherapy, radiotherapy, and traditional Chinese medicine can be used as auxiliary treatments. Clinically, for resectable STAD patients, the comprehensive treatment model based on surgical resection combined with chemotherapy has

certain curative effects, but the tumor is prone to progress or relapse, and the 5-year survival rate of advanced STAD patients is as low as 10% to 15% [2]. For unresectable and metastatic STAD, patients can obtain limited benefit from typical therapies, including systemic chemotherapy and targeted drugs, but they are facing a larger treatment bottleneck, and the survival time of patients usually does not exceed 1 year [3]. Although the individualized comprehensive treatment strategy based on surgery has been widely used in STAD, tumor recurrence and metastasis are still the main reasons for its high mortality. Moreover, the early diagnosis rate of STAD is low. Most patients with STAD have been in local progression stage when they visit the hospital, and the direct surgical effect is not ideal. Therefore, novel treatment strategies and biomarkers are urgently needed.

Recently, the emergence of immune checkpoint inhibitors (ICI) has officially kicked off the era of tumor immunotherapy. As a new type of treatment, tumor immunotherapy is based on the human immune system to exert antitumor effects by inhibiting negative feedback immune regulation mechanisms [4]. Immune checkpoint regulation is one of the most important immune regulation mechanisms for antitumor treatment, which regulates the intensity or breadth of the immune response through ligand-receptor binding [5]. Programmed cell death 1 (PD-1) and programmed cell death ligand 1 (PD-L1) are the most well-known target molecules for immune checkpoint blockade [6]. PD-1/PD-L1 pathway inhibitors have achieved significant clinical effects in the treatment of a variety of malignant tumors such as melanoma, bladder cancer, lung cancer, and renal cell carcinoma [7]. Especially, PD-L1 expression is upregulated in gastric cancer and is closely related to tumor progression and patient prognosis [8]. Moreover, in several clinical trials, ICI alone or in combination with other treatment options have achieved good positive results and less toxic side effects for STAD [9–11]. Thus, some ICI were used in clinical. For instance, trastuzumab is used for human EGFR-2 positive gastric cancer, and pembrolizumab is used for second-line and later-line treatment of gastric cancer with highly microsatellite instability/different mismatch repair (MSI-H/dMMR) [10, 11]. However, the effective response rate for immunotherapy is limited to fraction of gastric cancer patients, and the commonly used markers, such as PD-L1, tumor mutation burden and tumor infiltrating leukocytes, also have certain defects in assessing the response to immunotherapy [12]. Therefore, it is essential to identify novel biomarker and to explore its immune implication, thus enhancing individualized treatment and improving the success rate of immunotherapy. *In silico* analysis is promising to screen out useful biomarkers for immune therapy. For example, Wang and his colleagues screened immune-associated DEGs between esophageal cancer and normal samples using TCGA and immPORT databases, based on which a 6-immune gene prognostic model was constructed by regression algorithm and examined both in survival and immunity terms [13]. Similarly, the current study was processed bioinformatically on the role of AKAP12 in the immunity of stomach carcinoma.

A kinase anchor protein 12 (AKAP12), as an intracellular macromolecular scaffold protein and a member of kinase anchoring proteins family, participates in the integrated regulation of multiple signals in biological cells and exert an important role in maintaining the integrity of normal tissue structure and function by anchoring protein kinase C, protein kinase A, F-actin, etc. [14]. In recent years, a series of studies have reported that AKAP12 expression is downregulated in many types of tumors, such as liver cancer, colorectal cancer, prostate cancer, and lung cancer [15]. Its decreased expression and low activity in tumor cells may be due to hypermethylation in the promoter region of AKAP12 gene and are closely related to the malignancy degree [16, 17]. Moreover, the *in vitro* and *in vivo* experiments have confirmed that overexpression of AKAP12 significantly inhibits tumor cell proliferation and metastasis by controlling oncogenic signaling pathways in a spatiotemporal manner [18]. In contrast, tumor malignant biological behaviors were enhanced by interfering with the expression of AKAP12 [19, 20]. The occurrence of tumor is determined by the interaction between tumor cells and the tumor microenvironment including immune infiltrating cell microenvironment around tumor cells [21]. However, as a well-known tumor suppressor, the immune implication of AKAP12 in STAD has not been explored.

In the present study, comprehensive bioinformatics analysis was performed to evaluate the association between AKAP12 expression and immune infiltration cells, immunomodulators, and the signaling pathways regulating the AKAP12-mediated immune response. Moreover, based on AKAP12-associated immunomodulators, we developed a 4-immunomodulator prognostic signature for survival prediction in STAD. In this step, we used univariate and multivariate Cox regression analyses to construct and validate prognostic characteristics and evaluate the predictive performance of risk scores, etc. Cox regression analysis takes survival outcome and survival time as dependent variables and can simultaneously analyze the impact of many factors on survival. Because of its excellent nature, it has been widely used in medical follow-up studies.

2. Materials and Methods

2.1. Data Collection. The transcriptome profiling (type: HTSeq-FPKM) and clinical information (type: bcr xml) of STAD were both downloaded from The Cancer Genome Atlas (TCGA) project (<https://portal.gdc.cancer.gov/>). A total of 407 RNA expression profile samples were obtained, which included 375 tumor samples and 32 normal samples. The clinical data of 443 STAD cases were downloaded. After excluding patients with missing survival data and crucial clinical factors, the information of 371 samples was retained for further analysis.

2.2. AKAP12 Expression Analysis. Difference analysis of AKAP12 mRNA expression in normal and STAD samples was performed by R package “limma.” The immunohistochemical protein expression of AKAP12 gene in normal and STAD samples was obtained from the Human Protein Atlas (HPA) database (<https://www.proteinatlas.org/>) [22].

2.3. Correlation Analysis between AKAP12 and Immune Cell Based on Tumor Immune Estimation Resource (TIMER) and CIBERSORT. TIMER database contains three modules: immune association, cancer exploration, and immune estimation, and it is a reliable tool to investigate the relationship between immune infiltration and target gene (<https://cistrome.shinyapps.io/timer/>) [23]. In this study, the correlations between AKAP12 expression levels/copy number alterations (CNA) and six types of immune cells were estimated using the TIMER “immune association” modules. CIBERSORT is an R/web version tool that deconvolves the expression matrix of human immune cell subtypes based on the principle of linear support vector regression [24]. This method is based on a known reference gene set and provides a set of gene expression characteristics of 22 immune cell subtypes, namely, LM22 [25]. In our study, we used CIBERSORT to calculate the abundance of immune cells in TCGA-STAD samples based on LM22, and samples were assigned into AKAP12^{high} and AKAP12^{low} groups according to AKAP12 expression medium. Difference analysis of 22 types of immune cells between the two groups was performed by R package “limma” and visualized by R package “vioplot.”

2.4. Analysis and Validation of Immune-Related Pathways of AKAP12 Coexpressed Genes. To investigate AKAP12-related pathways, we first downloaded the RNA-seq data of 37 stomach cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE) website (<https://sites.broadinstitute.org/ccle>) and processed them by perl [26]. Based on the expression matrix data of stomach cell lines, we further screened out AKAP12 coexpressed genes through the “limma” R package with the filtering conditions: correlation coefficient > 0.5 and p value < 0.05. Finally, these coexpressed genes were subjected to Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses with the filtering condition p value < 0.05, which was visualized by the “ggplot2” R package. To further clarify the gene set related to AKAP12, the medium expression value of AKAP12 divided 37 stomach cancer cell lines into AKAP12^{high} and AKAP12^{low} groups, and Gene Set Enrichment Analysis (GSEA) was performed by setting the predefined gene sets “c2.cp.kegg.v7.4.symbols.gmt” with false discovery rate (FDR) p value < 0.05 and permutation value 1000 [27].

2.5. Analysis of AKAP12-Associated Immunomodulators. Integrated repository portal for tumor-immune system interactions (TISIDB) is a powerful website containing a large amount of tumor immunity related data, in which the association between genes and immune functions (such as lymphocytes, immunomodulators, and chemokines) for 30 types of TCGA cancers was precalculated (<http://cis.hku.hk/TISIDB/>) [28]. In this study, we used TISIDB to identify AKAP12 expression-associated immunostimulators and immunoinhibitors under the condition p value < 0.05 through Spearman correlation test. Then, AKAP12-associated immunomodulators were used to construct a protein-protein interaction (PPI) network using Search Tool for the Retrieval of Interaction Gene/Proteins (STRING) database (<https://string-db>

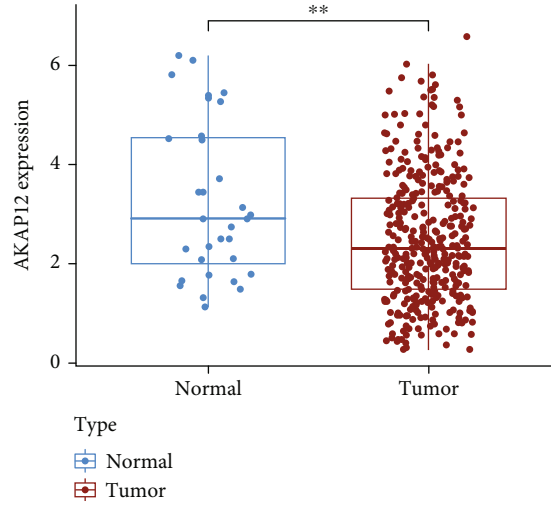
<http://www.webgestalt.org/>) with setting high confidence at 0.7 [29]. Afterward, AKAP12-associated immunomodulators were subjected to GO and KEGG analysis using WebGestalt online tool (<http://www.webgestalt.org/>) with FDR < 0.05 [30].

2.6. Construction of the Prognostic Signature and Nomogram. The association of AKAP12-associated immunomodulators and prognosis was analyzed by stepwise Cox regression analysis using R package “survmine” and “glmnet.” A prognostic signature was established based on the AKAP12-associated immunomodulators. According to the multivariate Cox result, the risk score for each patient was calculated through the formula: risk score = $\beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n$, where β_i represents the coefficient of each prognostic gene, and x_i represents the expression level of prognostic gene [31]. The association between risk score and survival time was determined by Kaplan–Meier survival curve and log-rank test. Multivariate Cox regression analysis was used to assess the independence of risk score from other clinical factors. The prediction performance of risk score was evaluated using R package “survivalROC.” The nomogram was constructed to predict the impact of important clinical parameters including risk score on the overall survival of patients using R package “rms,” and the consistency index (C-index) is calculated to evaluate the prediction accuracy of the nomogram [32]. Calibration curves were plotted to compare the predicted survival rate with actual survival rate by R package “rms.”

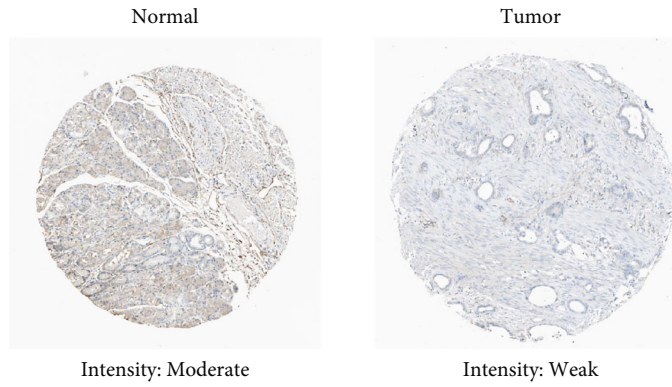
2.7. Statistical Analysis. Statistical analyses were conducted by R (v4.0.5). Student’s t -test was used to compare differences between groups. Survival curves were generated using the Kaplan–Meier method. Correction analysis of gene expression was performed by Spearman method. Univariate and multivariate analyses were conducted using Cox regression models to determine the independent prognostic factors. p value < 0.05 indicated statistical significance.

3. Results

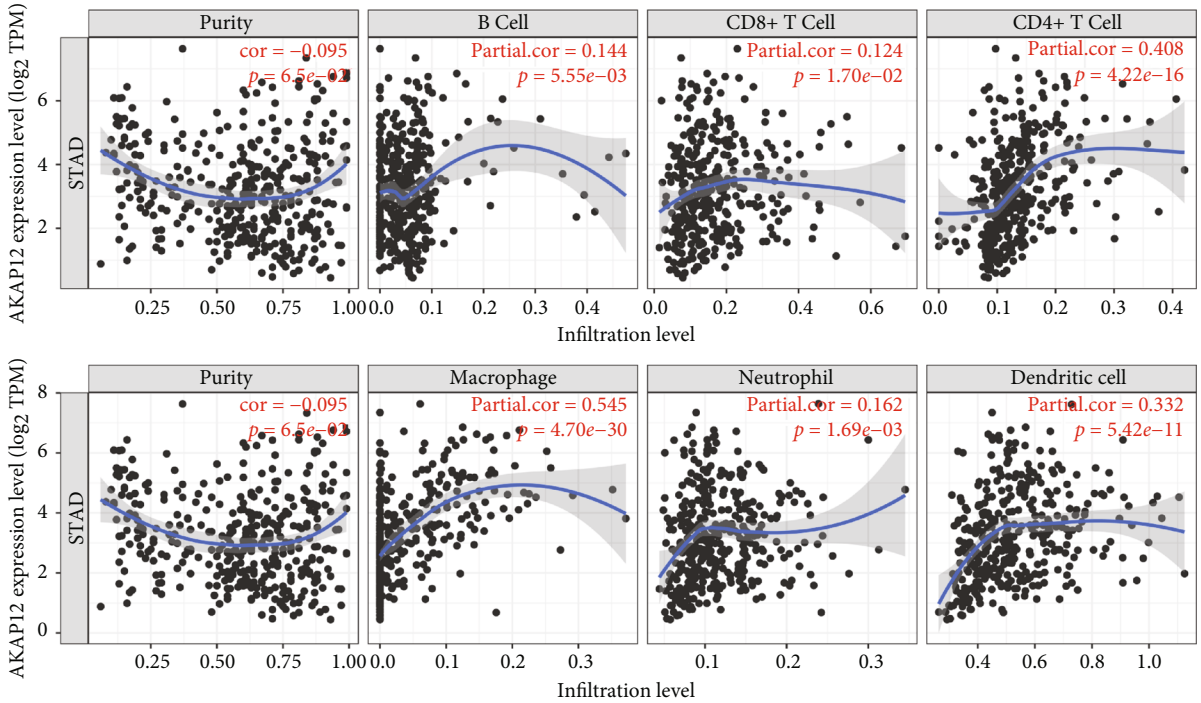
3.1. Association between AKAP12 and Immune Infiltration Cells in STAD. As a tumor suppressor, AKAP12 mRNA and protein levels were downregulated in STAD patients (Figures 1(a) and 1(b)). We next investigate whether AKAP12 expression and copy number alterations were associated with tumor immune infiltration based on TIMER database. As shown in Figure 1(c), AKAP12 expression was positively related with infiltration levels of CD4+ T cell ($\text{cor} = 0.408$) and macrophage ($\text{cor} = 0.545$). Besides that, with chromosome arm-level deletion of AKAP12, infiltration levels of B cells, CD8+ T cells, CD4+ T cells, and dendritic cells were significantly decreased (Figure 1(d)). To further clarify the relationship between AKAP12 expression and specific immune infiltration cells, CIBERSORT method was used to determine the infiltrating fraction of 22 types of immune cells in TCGA-STAD samples, and all samples were divided into AKAP12^{high} and AKAP12^{low} expression groups based on AKAP12 expression median. Our results showed that the infiltrating proportions of plasma cells,



(a)



(b)



(c)

FIGURE 1: Continued.

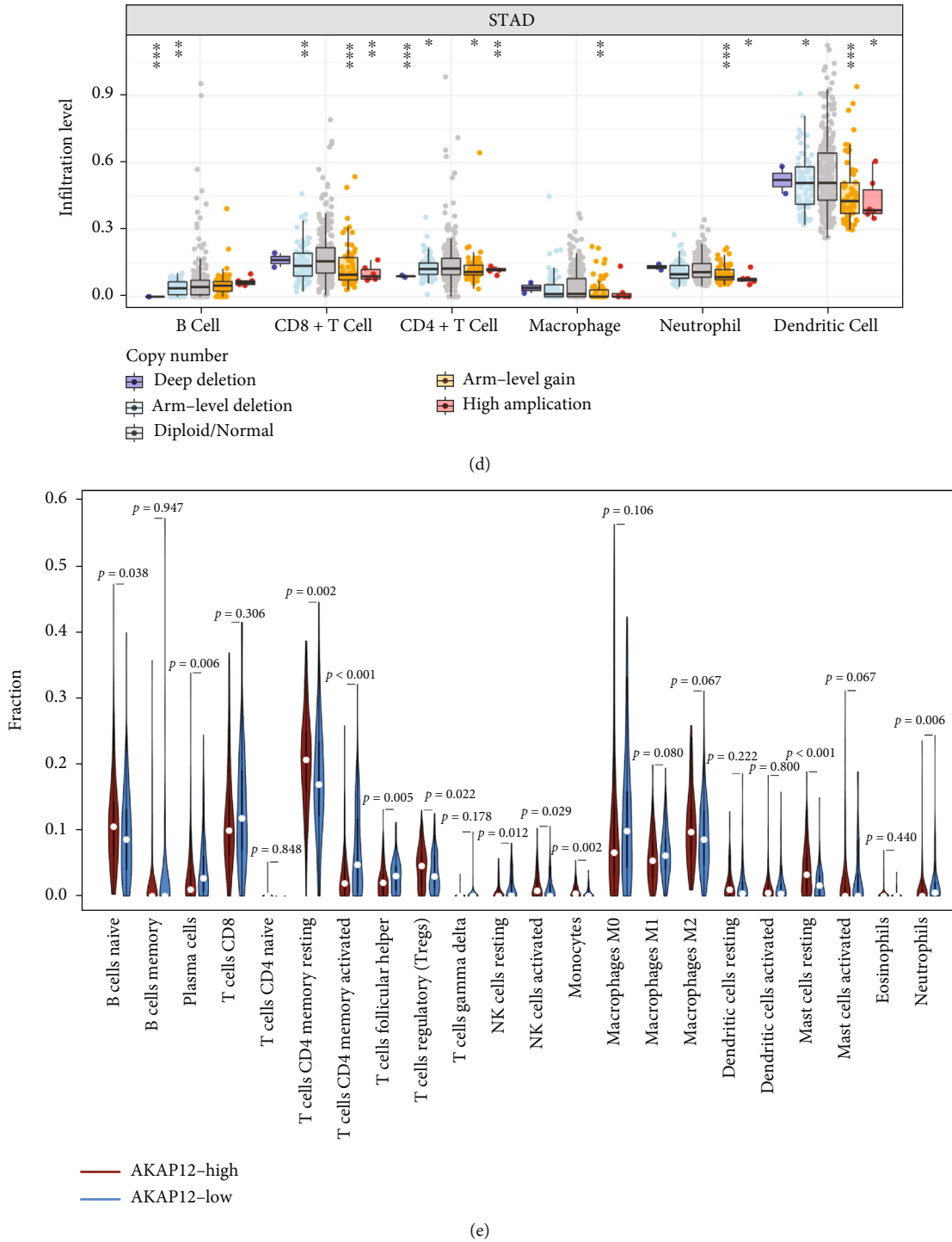
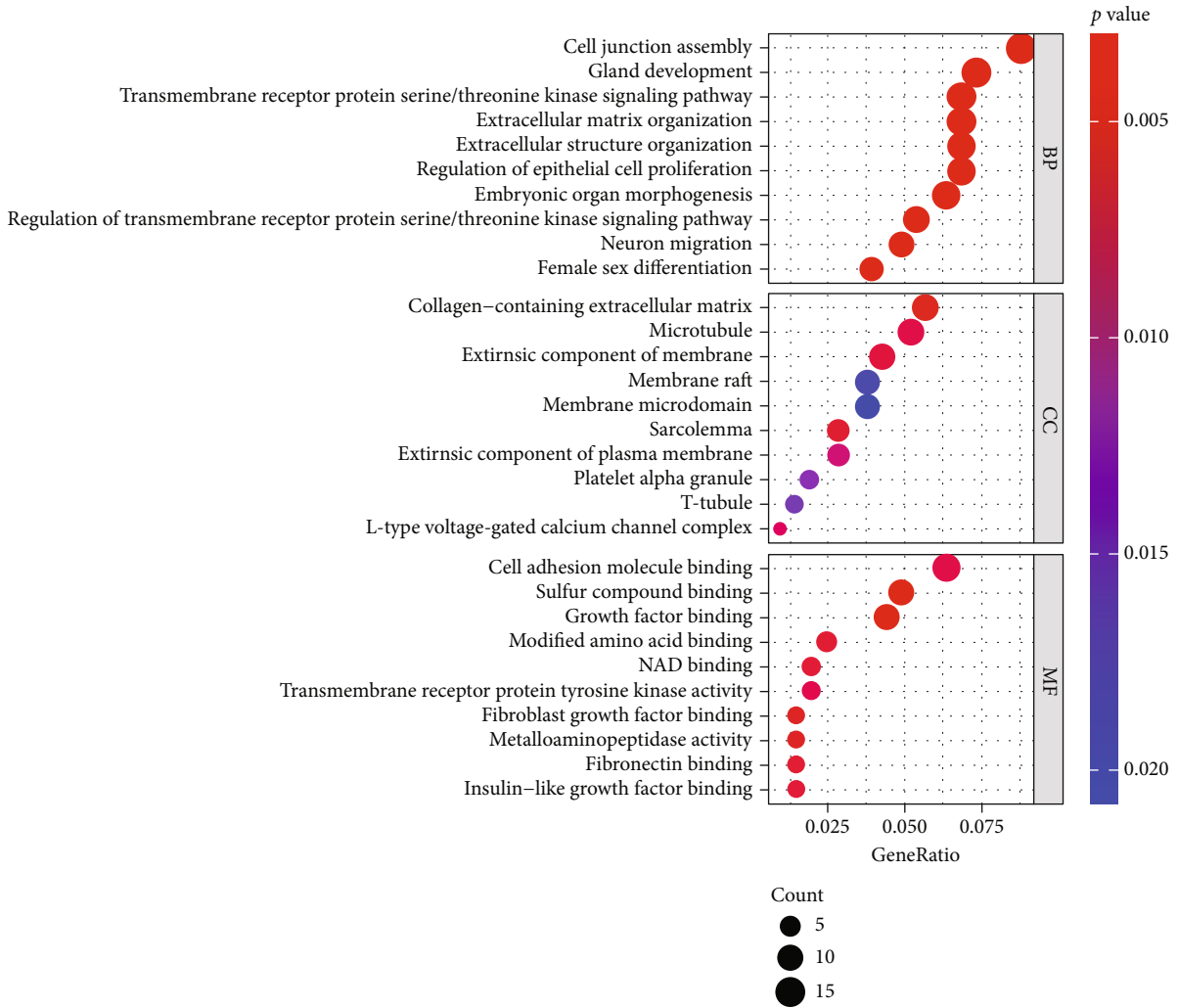


FIGURE 1: The expression of AKAP12 in STAD and its relationship with immune infiltrating cells. (a) The mRNA expression levels of AKAP12 in TCGA-STAD tissues and normal samples. $**p < 0.01$. (b) Immunohistochemical analysis of AKAP12 protein expression in STAD and normal tissues. Data were obtained from HPA database. The brown areas represent positive expression and the blue negative. (c) Correlation analysis between AKAP12 mRNA expression and six types of immune infiltrating cells based on TIMER database. The relation coefficient and p value are shown in each plot. (d) Associations between AKAP12 gene copy numbers and immune cell infiltration levels based on TIMER database. The p value is shown in each plot. (e) Violin plot displaying the infiltrating levels of 22 types of immune cells between the AKAP12^{high} expression group and the AKAP12^{low} expression group (group cutoff = AKAP12 expression median). The p value is shown in the figure.



(a)

FIGURE 2: Continued.

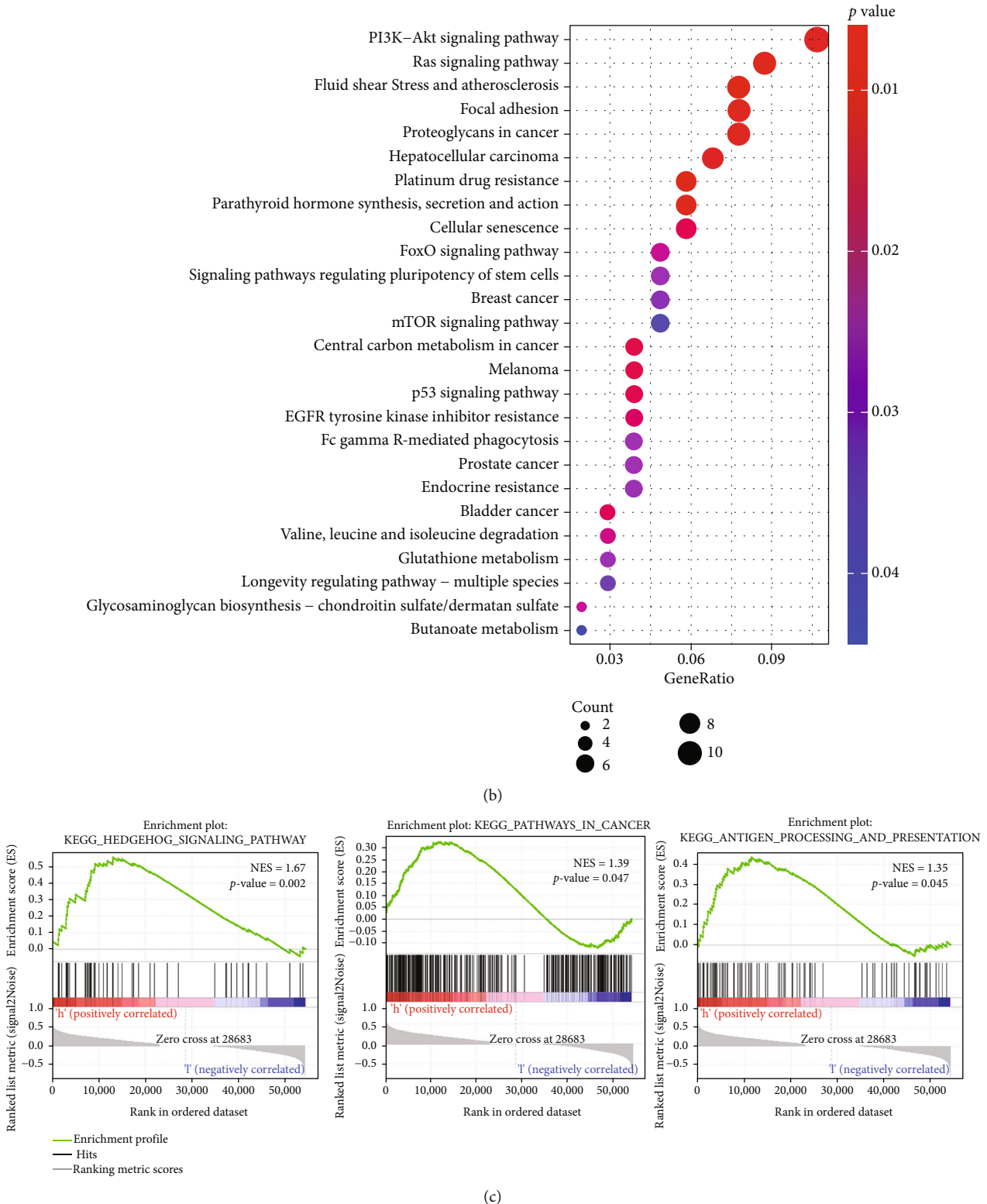
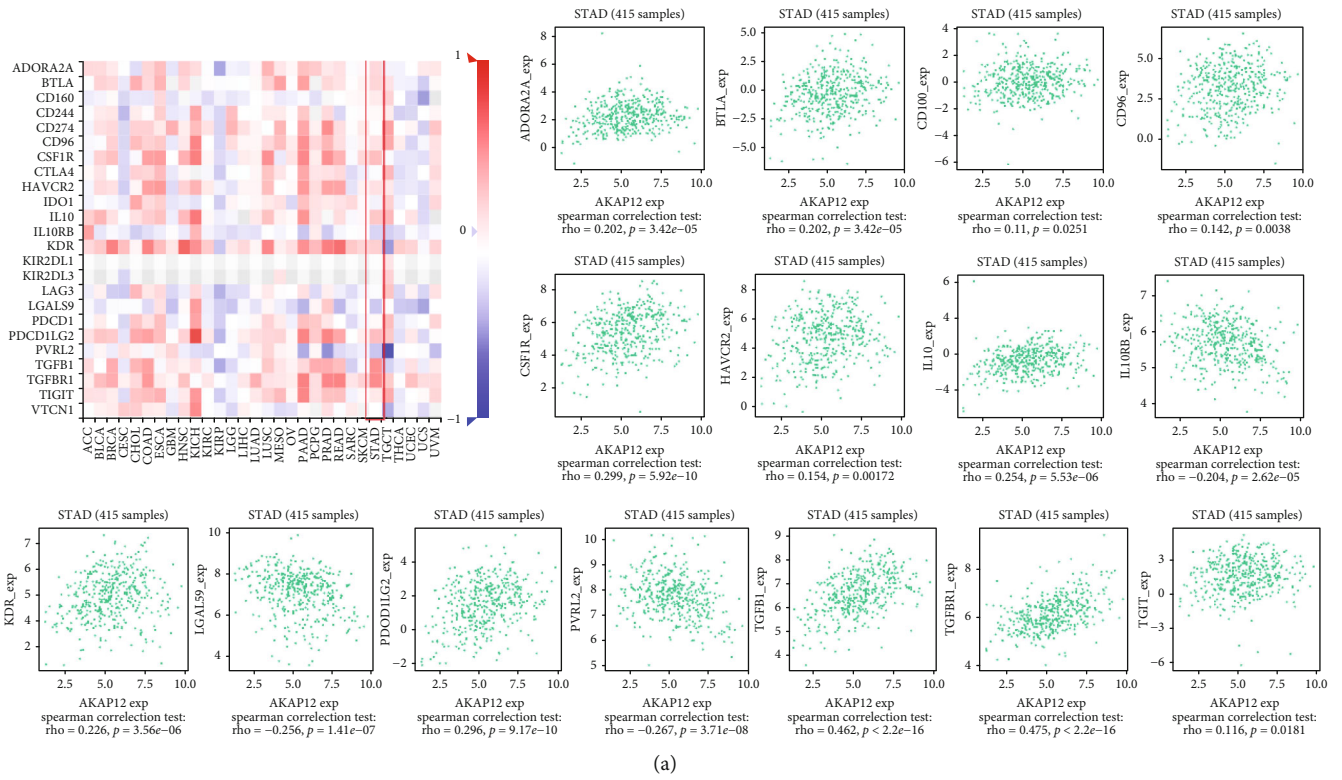


FIGURE 2: AKAP12 is associated with cancer- and immune-related pathways based on CCLE project. (a) GO function annotation based on AKAP12 coexpressed genes. (b) KEGG pathway analysis based on AKAP12 coexpressed genes. (c) GSEA enrichment analysis of AKAP12^{high} and AKAP12^{low} expression groups.



(a)

FIGURE 3: Continued.

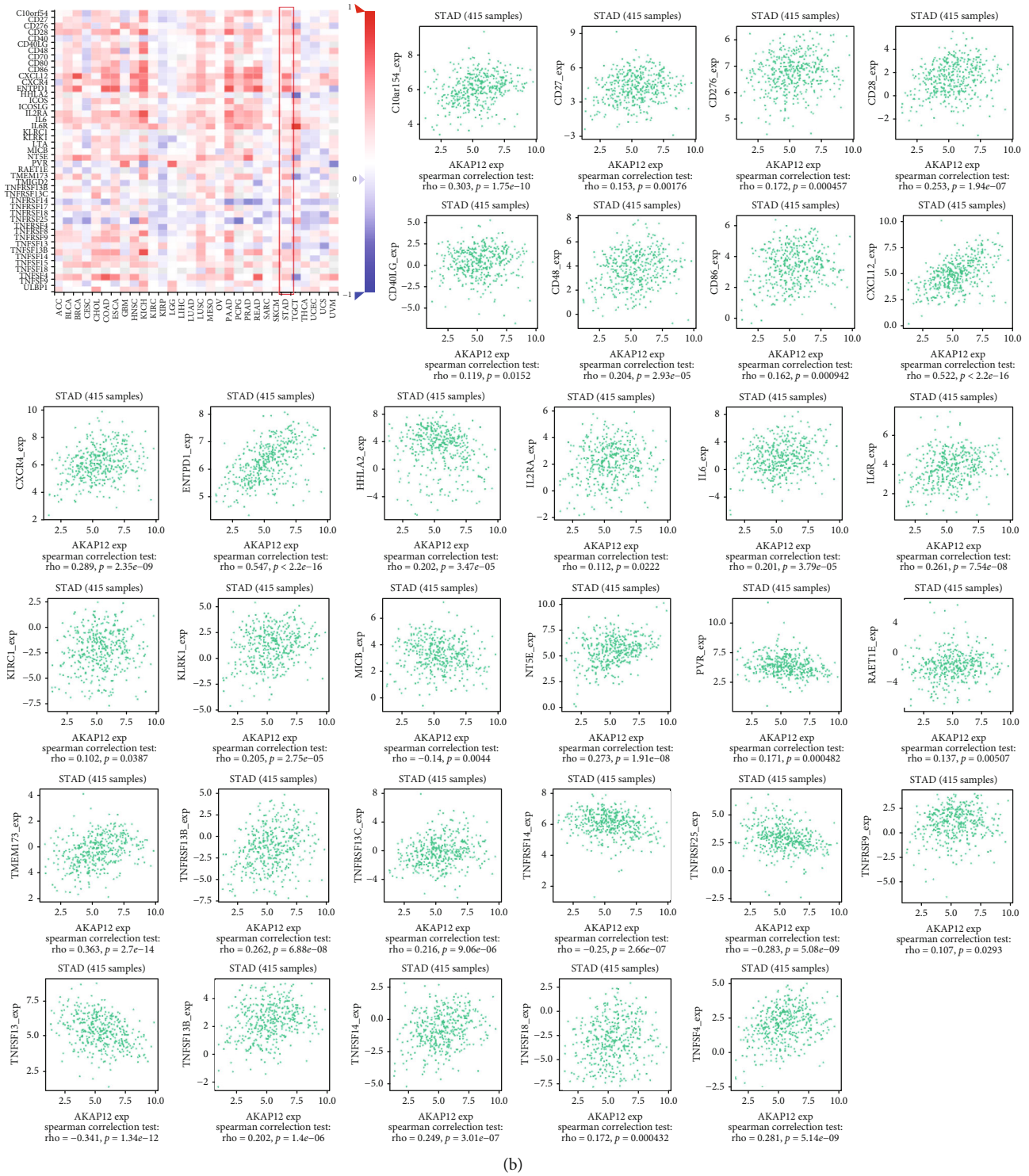
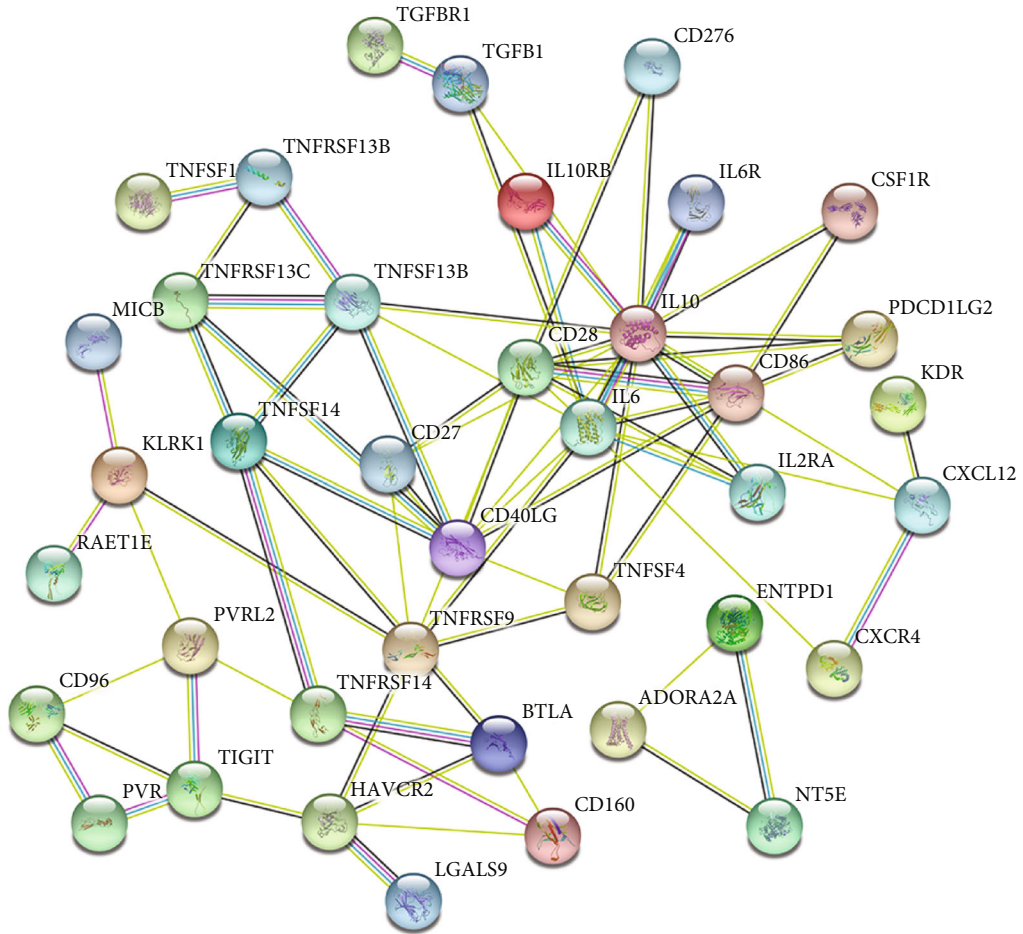


FIGURE 3: Identification of AKAP12-associated immunomodulators in STAD. (a) Correlation heat map and dot plots showed 15 immunoinhibitors significantly associated with AKAP12 expression in STAD. (b) Correlation heat map and dot plots showed 31 immunostimulators significantly associated with AKAP12 expression in STAD.

memory activated CD4 T cells, follicular helper T cells, resting NK cells, and neutrophils were significantly higher in the AKAP12^{low} expression group (Figure 1(e)). On the contrary, patients with high AKAP12 expression displayed higher infiltrating proportions of native B cells, memory resting

CD4 T cells, regulatory T cells, activated NK cells, monocytes, and resting mast cells (Figure 1(e)). The above results indicate that the expression of AKAP12 and the alteration of copy number may affect the infiltration level of tumor immune cells.



(a)

FIGURE 4: Continued.

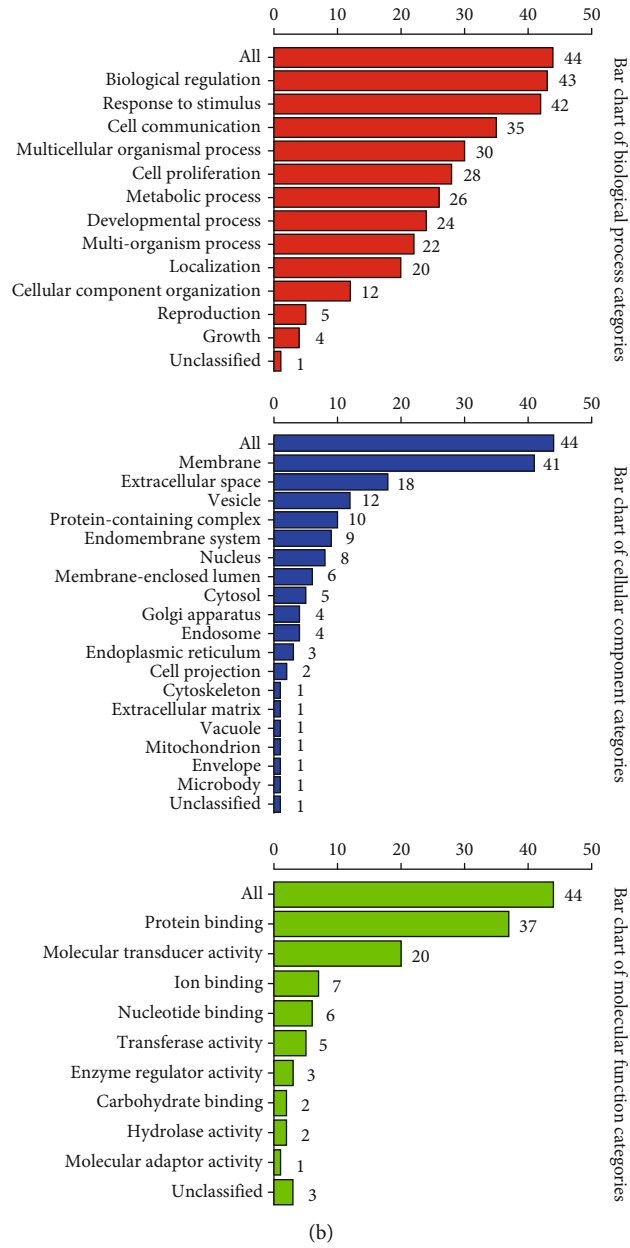


FIGURE 4: Continued.

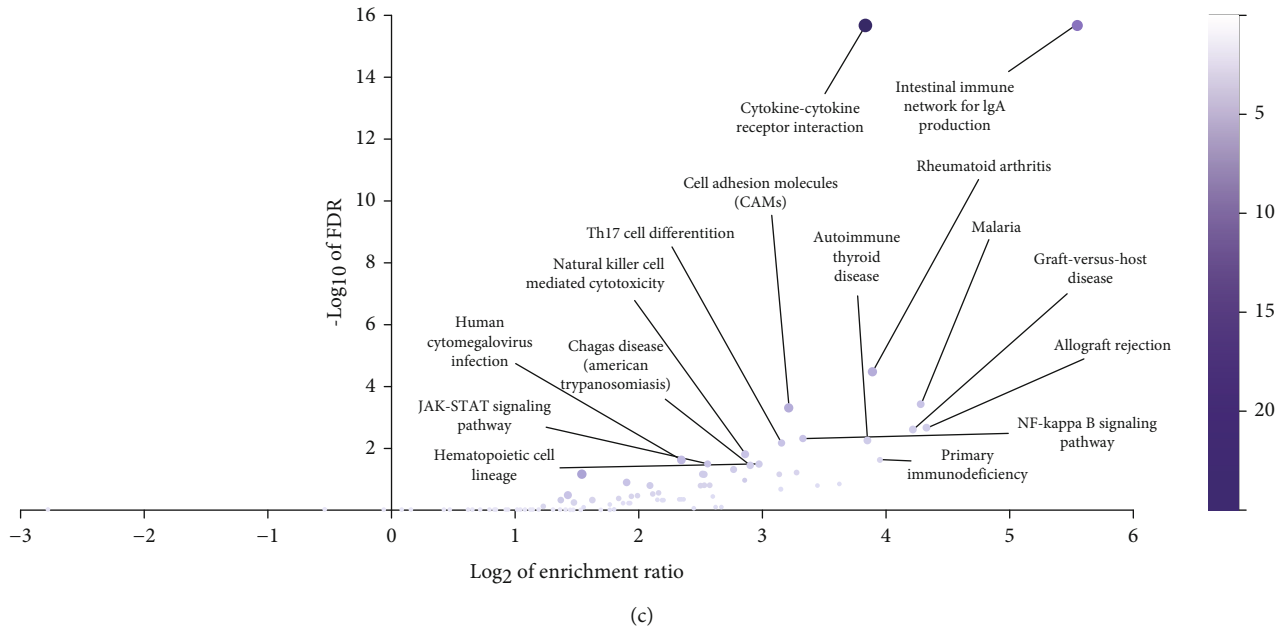


FIGURE 4: Analysis of the pathways involved in AKAP12-associated immunomodulators. (a) Constructing a PPI network based on the 46 AKAP12-associated immunomodulators using STRING tool. (b) GO function annotation based on the 46 AKAP12-associated immunomodulators using WebGestalt online tool. (c) KEGG enrichment analysis based on the 46 AKAP12-associated immunomodulators using WebGestalt online tool.

3.2. AKAP12 Is Associated with Immune-Related Pathways in Gastric Cancer. To investigate the functional and pathway enrichment of the AKAP12-associated genes, we performed the GO and KEGG analyses of the AKAP12 coexpressed genes. We first analyzed the RNA-seq data containing 37 stomach cancer cell lines from the CCLE datasets. Based on the expression matrix data of stomach cell lines, we further filtered out a series of AKAP12 coexpressed genes by setting the filtering conditions: correlation coefficient > 0.5 and p value < 0.05 . GO results of these coexpressed genes are shown in Figure 2(a), and KEGG analysis showed that AKAP12-related coexpressed genes were enriched in cancer- and immune-related pathways, including EGFR tyrosine kinase inhibitor resistance, Ras signaling pathway, FoxO signaling pathway, p53 signaling pathway, Fc gamma R-mediated phagocytosis, PI3K-Akt signaling pathway, and mTOR signaling pathway (Figure 2(b)). Furthermore, based on the expression medium value of AKAP12 in CCLE datasets, stomach cell lines were assigned into AKAP12^{high} and AKAP12^{low} groups. As shown in Figure 2(c), GSEA enrichment analysis demonstrated that several cancer- and immune-related pathways, such as hedgehog signaling pathway, pathway in cancer, and antigen processing and presentation, were activated in the AKAP12^{high} group, which further validated that AKAP12 may affect tumor progression through immune-related pathways in gastric cancer.

3.3. AKAP12 Is Associated with Immunomodulators in Gastric Cancer. We next aimed to explore the relationship between AKAP12 and immunomodulators. For this purpose, TISIDB, an integrated repository portal for tumor-immune system interactions, was used to analyze the relations between immunomodulators and AKAP12 expression. Here,

15 immunoinhibitors (ADORA2A, BTLA, CD96, CD160, CSF1R, HAVCR2, IL10, IL10RB, KDR, LGALS9, PDCD1LG2, PVRL2, TGFBI, TGFBR1, and TIGIT) (Figure 3(a)) and 31 immunostimulators (C10orf54, CD27, CD28, CD40LG, CD48, CD86, CD276, CXCL12, CXCR4, ENTPD1, HHLA2, IL2RA, IL6, IL6R, KLRC1, KLRK1, MICB, NT5E, PVR, RAET1E, TMEM173, TNFRSF9, TNFRSF13B, TNFRSF13C, TNFRSF14, TNFRSF25, TNFSF4, TNFSF13, TNFSF13B, TNFSF14, and TNFSF18) (Figure 3(b)) were identified as either positively or negatively correlated with AKAP12 expression. By constructing a PPI network of the 46 immunomodulators, we obtain a close interaction relationship between AKAP12-related immunoregulatory genes (Figure 4(a)). GO function annotations showed that AKAP12-related immunomodulators were involved in important functions related to tumorigenesis and development (Figure 4(b)). Pathway enrichment volcano map showed that significantly enriched pathways of AKAP12-related immunomodulators include some immune and tumor-related pathways (Figure 4(c)).

3.4. Construction of an AKAP12-Related Immunomodulator Prognostic Signature. To investigate the prognostic significance of the AKAP12-related immunomodulators, the expression profile of 46 candidate AKAP12-related immunomodulators combined with survival data was analyzed in TCGA-STAD dataset ($n = 371$) using a stepwise multivariate Cox regression analysis. Through filtration with p value < 0.05 as the cutoff in univariate Cox regression analysis, 7 of 46 AKAP12-related immunomodulators were identified to be associated with the overall survival of patients (Figure 5(a)). Then, a prognostic risk signature containing four AKAP12-related immunomodulators (CXCR4, IL6, NT5E, and TNFSF18) was constructed following stepwise

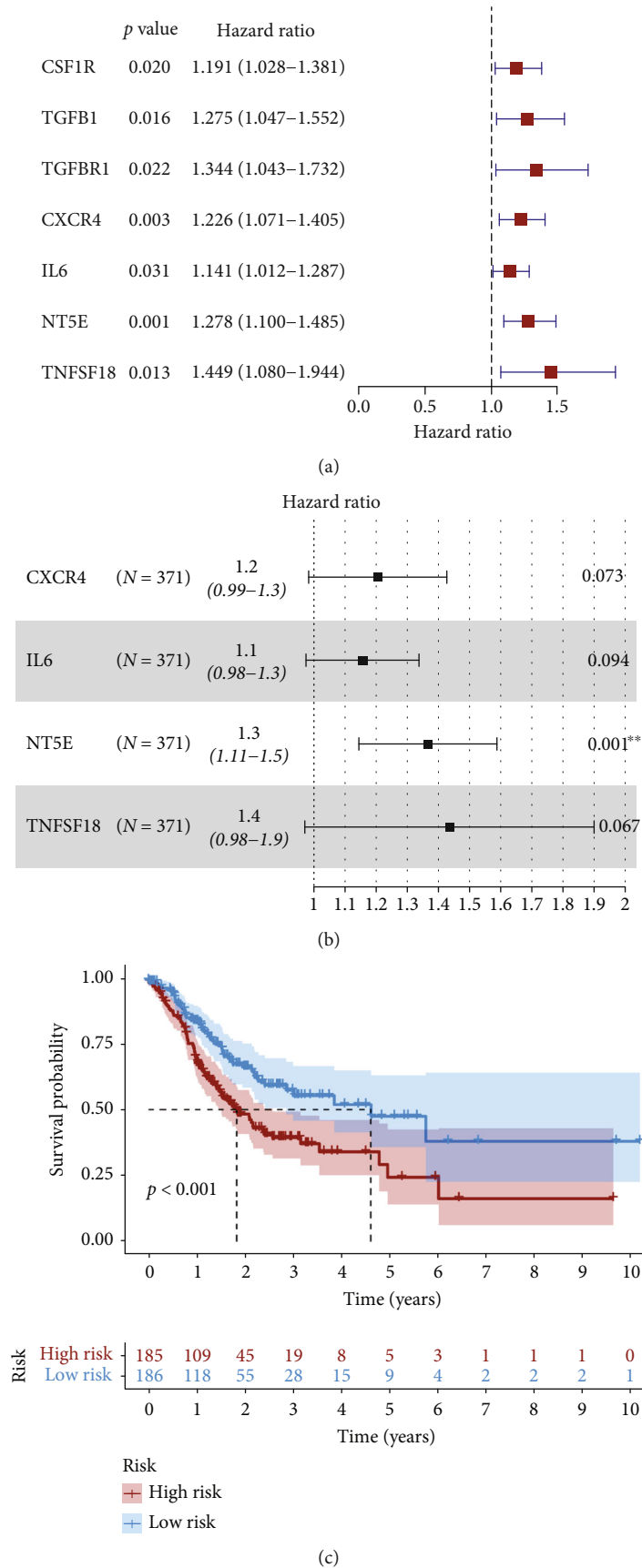
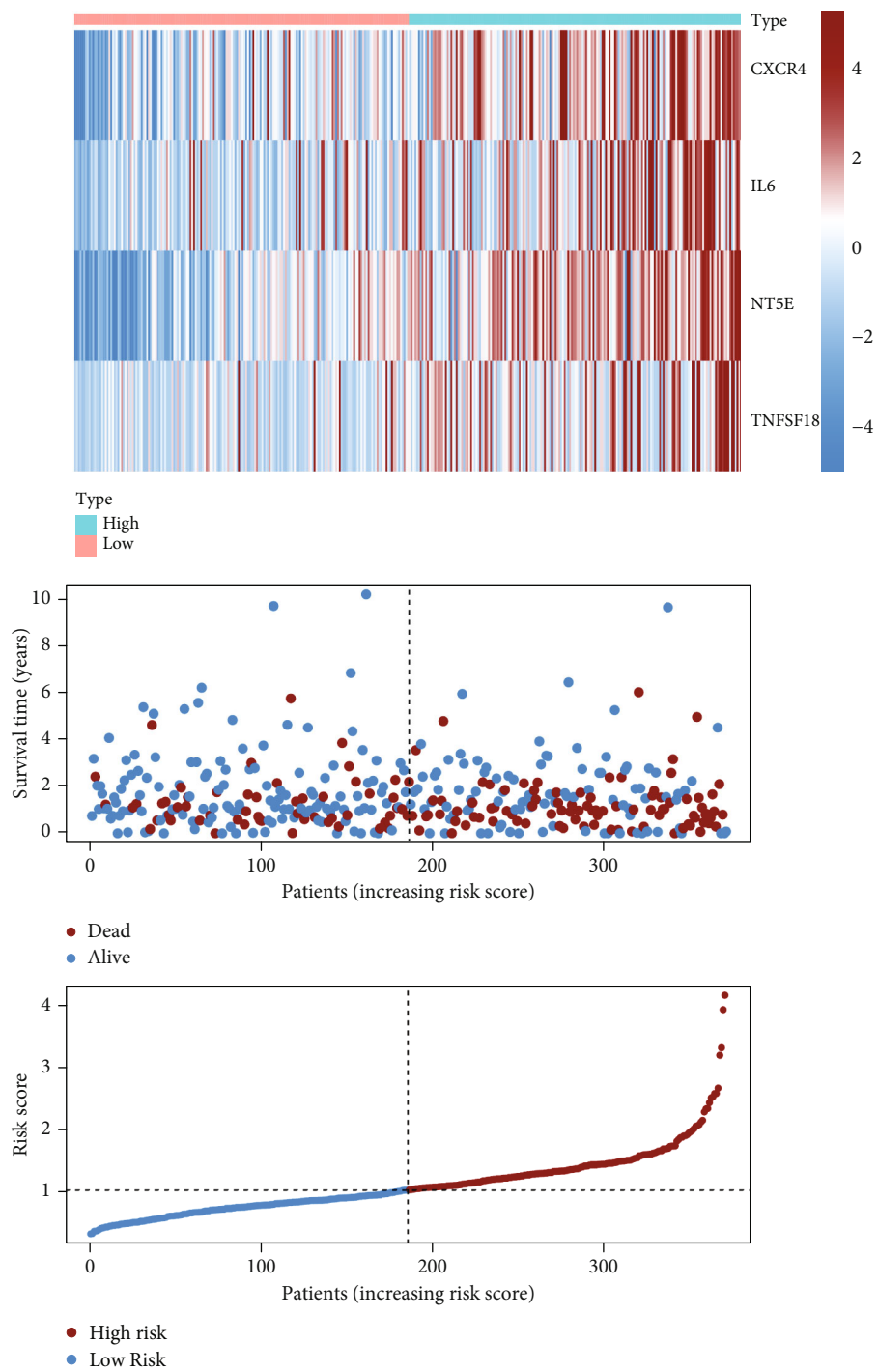
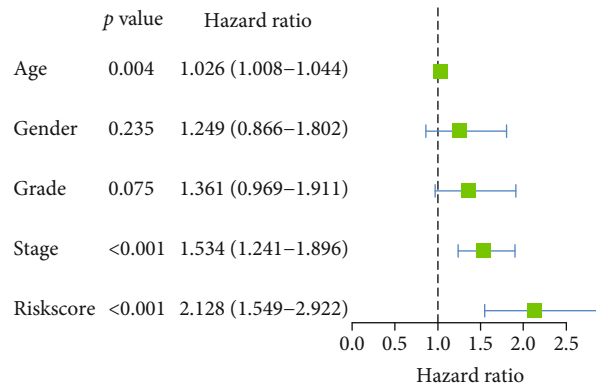


FIGURE 5: Continued.

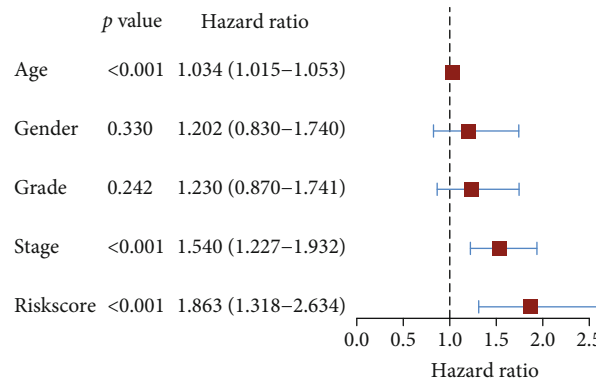


(d)

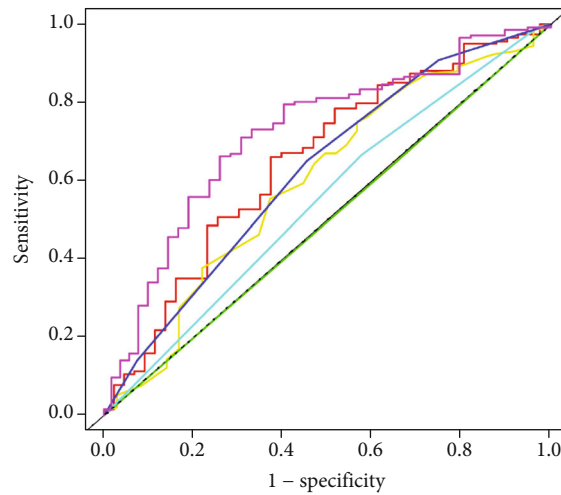
FIGURE 5: Continued.



(e)



(f)



— Risk, AUC = 0.652
 — Age, AUC = 0.601
 — Gender, AUC = 0.497
 — Grade, AUC = 0.548
 — Atage, AUC = 0.625
 — Risk+clinical, AUC = 0.723

(g)

FIGURE 5: Construction of a prognostic gene signature based on AKAP12-associated immunomodulators. (a) The associations between AKAP12-associated immunomodulators and overall survival in STAD using univariate Cox regression analysis. (b) The hazard ratios of AKAP12-associated immunomodulators integrated into the prognostic signatures are shown in the forest plots for TCGA-STAD patients. (c) Survival curve for risk score based on Kaplan–Meier analysis in TCGA-STAD patients. (d) Distribution of risk scores, survival status, and gene expression profiles for the TCGA-STAD patients. (e) Univariate and (f) multivariate Cox regression analyses of the risk score and overall survival in TCGA-STAD patients. (g) Time-dependent ROC curve for the risk score in TCGA-STAD patients.

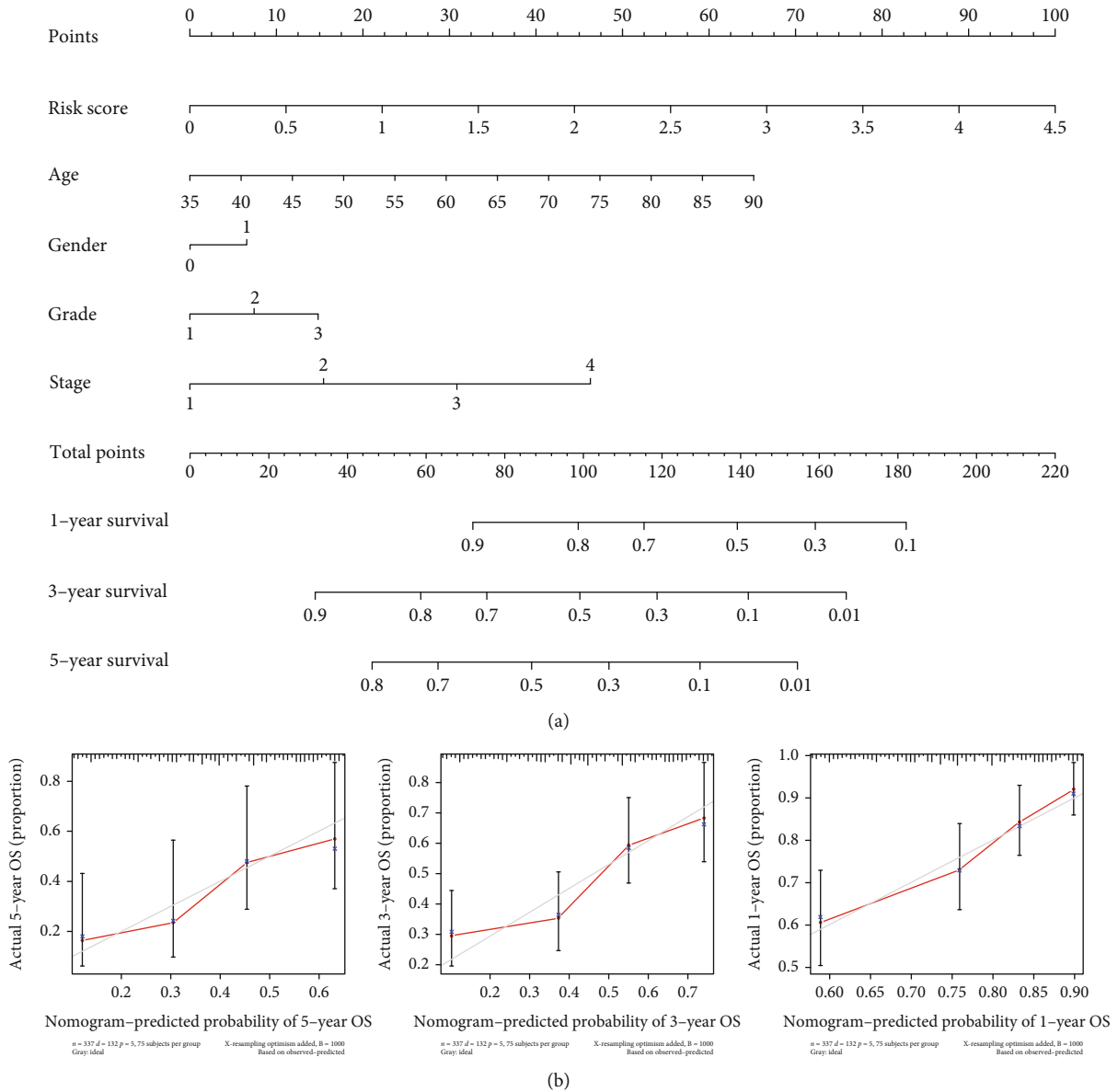


FIGURE 6: Prognostic nomogram for TCGA-STAD dataset. (a) Nomogram for predicting the overall survival of TCGA-STAD patients. (b) Calibration curves of 1-, 3-, and 5-year overall survival of TCGA-STAD patients. Red line: nomogram-predicted survival curve. Gray line: ideal survival reference curve.

multivariate Cox regression analysis (Figure 5(b)). In this prognostic signature, each patient risk score was calculated through the proposed formula: risk score = sum of the expression level of the four immunomodulators \times their respective coefficient. Patients were then divided into high- and low-risk groups according to the optimal risk score cutoff. Survival curve exhibited that the high-risk group patients had a worse prognosis than the low-risk group patients, with p value < 0.001 (Figure 5(c)). The risk heat map displayed the expression of the four risky immunomodulators was upregulated as the patient's risk score increased, and the high-risk group also had higher deaths than the low-risk group (Figure 5(d)). Then, we aimed to test whether the risk signature was independent of other clinicopathological characteris-

tics; our results showed that the risk score was significantly associated with overall survival following the univariate Cox regression analysis (HR = 2.128, 95%CI = 1.549 – 2.922, and p value < 0.001) (Figure 5(e)). Moreover, with correction for age, gender, grade, and stage in the multivariate Cox regression model, we found that the risk score still retained its prognostic significance as an independent prognostic factor for STAD patients (HR = 1.863, 95%CI = 1.318 – 2.634, and p value < 0.001) (Figure 5(f)). Furthermore, the receiver operating characteristic (ROC) curve showed that risk signature had an area under the curve (AUC) value of 0.652 (Figure 5(g)). Finally, we generated a nomogram to evaluate the predicting performance of the risk score combined with other clinical factors in TCGA-STAD patients (Figure 6(a)). The concordance

index (C-index) that reflects the accuracy of prediction and actual survival situation was 0.608. The calibration curve displayed that the predicted probability of the nomogram (red line) closely matched the ideal reference line (gray line) for the 5-year survival prediction (Figure 6(b)).

4. Discussion

STAD is one of the malignant tumors that endanger human health. For early STAD, surgery combined with chemotherapy has a significant effect, but for advanced STAD, its effect is not ideal [33]. Immunotherapy provides hope for the treatment of patients with advanced STAD, but there is a defect in drug resistance [12]. Thus, it is essential and urgently needed to identify novel biomarkers of immunotherapy in STAD. In the present study, we systematically analyzed the immune significance of the tumor suppressor AKAP12 from various aspects including immune infiltration cells, immune-related pathways, and immunomodulators and constructed an AKAP12-related immunomodulator prognostic signature. Our results suggest that AKAP12 may be a potential immunotherapeutic target for STAD.

The regulatory effect of AKAP12 as a tumor suppressor gene in tumors has been extensively studied and confirmed. For instance, KAP12 expression is downregulated in colorectal tumor tissues, and AKAP12 methylation level was positively associated with tumor grade [16]. Exogenous overexpression of AKAP12 in Lovo colon cancer cell line inhibits tumor cell proliferation, migration, and invasion ability [18]; in contrast, AKAP12 silencing or AKAP12/HDAC3 cosilencing promotes tumor cell proliferation, colony-forming ability, and cell cycle progression [20]. These studies suggest AKAP12 plays a protective role in preventing the occurrence and progression of colorectal cancer. In addition to its inhibitory effect on tumor growth and metastasis by scaffolding key regulatory proteins such as protein kinase C, F-actin, and cyclins, AKAP12 also exerts important role in the drug resistance mechanism of tumors [14, 15, 34]. In our present study, we also detected a significantly lower expression of AKAP12 in STAD. Moreover, CpG island hypermethylation was frequently observed in the promoter region of AKAP12 in gastric cancer, and the expression of AKAP12 can be restored through methyltransferase inhibitor [17], suggesting DNA methylation is directly involved in the silencing of the AKAP12 in gastric cancer. Furthermore, the reexpression of AKAP12 in gastric cancer cell line induced apoptotic cell death and reduced colony formation [17], indicating AKAP12 is a potential tumor suppressor of gastric cancer.

Tumor immunotherapy reflects the complex cellular and molecular interactions between tumor cells and the surrounding immune microenvironment [21]. An important content of our research is to clarify the association between AKAP12 and immune infiltration cells in STAD. Based on TIMIR database, our results demonstrated that AKAP12 expression was positively related to CD4⁺ T cells and macrophages. Besides that, the deletion of AKAP12 somatic copy number causes a reduction in the infiltrating proportion of B cell, CD8⁺ T cell, CD4⁺ T cell, and dendritic cell. It should

be noted that the abovementioned immune cells are a fuzzy classification concept. It is essential to investigate the precise tumor-promoting or -suppressing immune cells to clarify the specific role of AKAP12-related immune cells. For this purpose, CIBERSORT algorithm containing 22 types of infiltrating immune cells was performed, and the results showed the infiltrating abundance of tumor-suppressing immune cells, such as native B cells, memory resting CD4 T cells, and activated NK cells, was significantly higher in patients with AKAP12-high expression. These results suggest that AKAP12 may play an important effect on immune cell infiltration of STAD.

Furthermore, our results confirmed that AKAP12 is involved in the regulation of immune-related signaling pathways from at least three perspectives. Firstly, KEGG pathway analysis in our study revealed that AKAP12 coexpressed genes were significantly enriched in PI3K-Akt signaling pathway and mTOR signaling pathway. It is well established that the PI3K-AKT-mTOR signaling network is dysregulated in human cancer, and PI3K-AKT-mTOR inhibitors can not only target cancer cell biology but also weaken immune cell effector functions [35, 36]. Secondly, GSEA enrichment analysis in our study showed that antigen processing and presentation are activated in patients with AKAP12 high expression, which has been proved to be an important and essential link in the process of immune cells killing cancer antigens in tumor immunity [37]. Thirdly, KEGG pathway analysis based on AKAP12-related immunomodulators demonstrated that several important immune pathways including Th17 cell differentiation and natural killer cell-mediated cytotoxicity were involved in AKAP12-mediated immune events. Sun et al. reported that tumor exosomes promote Th17 cell differentiation and inhibit tumor growth by delivering lncRNA CRNDE-h in colorectal cancer [38]. Shen et al. provide evidence that natural killer cell-mediated cytotoxicity can be enhanced to kill breast cancer cells by silencing NKG2D ligand-targeting miRNA [39]. Taken together, we speculate that AKAP12 may also play an important role in gastric cancer tumor immunity through these immune-related pathways.

In recent years, as ICI have made breakthroughs in the treatment of gastric cancer, and more and more immune-related gene signatures have been identified to evaluate the prognosis of gastric cancer patients. For example, Wang et al. identified a stromal-immune score-based 4-gene prognostic signature by estimating stromal and immune scores from TCGA and Gene Expression Omnibus (GEO) gastric cancer data [40]. Liu et al. developed an immune-related gene pair signature based on ImmPort database, and the signature is associated with overall survival and immune checkpoint expression in gastric cancer patients [41]. In our study, we also identified an AKAP12-related immunomodulator signature for STAD prognosis prediction, in which four immunomodulators (CXCR4, IL6, NT5E, and TNFSF18) were included. CXCR4 encodes a CXC chemokine receptor specific to stromal cell-derived factor-1, which has seven transmembrane regions and is located on the cell surface. The cases studied by Kamihara et al. [42] indicated that CXCL12 (SDF-1)/CXCR4 axis is involved in the observed

metastasis of diffuse large B cell lymphoma to primary STAD and considered that the interaction between chemokines and their receptors may be the potential mechanism of the observed metastasis between tumors. IL6 encodes a cytokine that plays a role in inflammation and B cell maturation. The study of Ju et al. [43] showed that tumor-associated macrophages through IL-6 and TNF- α signal induce the expression of PD-L1 in gastric cancer cells and help tumor cells escape cytotoxic T cell killing. NT5E encodes a plasma membrane protein that catalyzes the conversion of extracellular nucleotides to membrane permeable nucleosides and performs many homeostatic functions in healthy organs and tissues. Since the free adenosine produced by NT5E can inhibit the cellular immune response and promote the immune escape of tumor cells, it can be used as an inhibitory immune checkpoint molecule [44]. TNFSF18 is a member of the tumor necrosis factor (ligand) superfamily 18 and has the activity of binding to tumor necrosis factor receptors. The results of Chen et al. showed that five differently expressed genes, including TNFSF18, can be used as a single biomarker to predict the efficacy of anti-PD-1 in patients with metastatic non-small-cell lung cancer [45]. In addition, this risk signature is not only an independent prognostic factor for gastric cancer but also presents good prediction performance. Moreover, we provide a reference for the individualized prognosis of patients by integrating the risk score and other clinical factors to construct a nomogram.

In conclusion, despite the flaws in our study, for example, pure bioinformatics analysis without experimental verification, our study provides strong evidence that AKAP12 is closely related to tumor immunity in STAD from three aspects: immune infiltration cells, immune pathways, and immunomodulators. More importantly, the AKAP12-related prognostic signature may provide good application prospects for clinical practice.

Abbreviations

AKAP12:	A kinase anchor protein 12
AUC:	Area under the curve
CCL:	Cancer Cell Line Encyclopedia
CIBERSORT:	Cell-Type Identification by Estimating Relative Subsets of RNA Transcripts
C-index:	Consistency index
CNA:	Copy number alterations
dMMR:	Different mismatch repair
FDR:	False discovery rate
GEO:	Gene Expression Omnibus
GO:	Gene Ontology
GSEA:	Gene Set Enrichment Analysis
HPA:	Human Protein Atlas
ICI:	Immune checkpoint inhibitors
KEGG:	Kyoto Encyclopedia of Genes and Genomes
MSI-H:	Highly microsatellite instability
PD-1:	Programmed cell death 1
PD-L1:	Programmed cell death ligand 1
PPI:	Protein-protein interaction
ROC:	Receiver operating characteristic

STAD:	Stomach adenocarcinoma
TCGA:	The Cancer Genome Atlas
TIMER:	Tumor Immune Estimation Resource
WebGestalt:	Web-Based Gene Set Analysis Toolkit.

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request. Data used included The Cancer Genome Atlas (TCGA, <http://portal.gdc.cancer.gov/projects>).

Conflicts of Interest

The authors declare no competing interests exist.

Authors' Contributions

Y. Wang designed the research; Z. Xu and L. Peng prepared the figures and drafted the manuscript; Z. Xu, L. Peng, L. Xiang, H. Gu and Y. Wang analyzed the data; Z. Xu and L. Peng contributed analytic tools and finalized the manuscript. All authors have read and approved the final manuscript.

Acknowledgments

This work was supported by The General Project of Chongqing Natural Science Foundation, Chongqing Science and Technology Commission, China (grant number cstc2021jcyj-msxmX0153 (Linglong Peng)), and The General Project of Chongqing Natural Science Foundation, Chongqing Science and Technology Commission, China (grant number cstc2021jcyj-msxmX0112 (Yaxu Wang)).

References

- [1] H. Sung, J. Ferlay, R. L. Siegel et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [2] Z. Zhai, Z. Y. Zhu, X. L. Cong et al., "Changing trends of clinicopathologic features and survival duration after surgery for gastric cancer in Northeast China," *World journal of gastrointestinal oncology*, vol. 12, no. 10, pp. 1119–1132, 2020.
- [3] N. Boku, M. H. Ryu, K. Kato et al., "Safety and efficacy of nivolumab in combination with S-1/capecitabine plus oxaliplatin in patients with previously untreated, unresectable, advanced, or recurrent gastric/gastroesophageal junction cancer: interim results of a randomized, phase II trial (ATTRACTION-4)," *Annals of oncology: official journal of the European Society for Medical Oncology*, vol. 30, no. 2, pp. 250–258, 2019.
- [4] Q. Zhao, L. Cao, L. Guan et al., "Immunotherapy for gastric cancer: dilemmas and prospect," *Briefings in Functional Genomics*, vol. 18, no. 2, pp. 107–112, 2019.
- [5] L. Figueroa-Protti, R. Soto-Molinari, M. Calderón-Osorno, J. Mora, and W. Alpizar-Alpizar, "Gastric cancer in the era of immune checkpoint blockade," *Journal of Oncology*, vol. 2019, Article ID 1079710, 2019.
- [6] A. Constantinidou, C. Aliferis, and D. T. Trafalis, "Targeting programmed cell death -1 (PD-1) and ligand (PD-L1): a new

- era in cancer active immunotherapy,” *Pharmacology & Therapeutics*, vol. 194, pp. 84–106, 2019.
- [7] Y. Han, D. Liu, and L. Li, “PD-1/PD-L1 pathway: current researches in cancer,” *American Journal of Cancer Research*, vol. 10, no. 3, pp. 727–742, 2020.
- [8] E. D. Thompson, M. Zahurak, A. Murphy et al., “Patterns of PD-L1 expression and CD8 T cell infiltration in gastric adenocarcinomas and associated immune stroma,” *Gut*, vol. 66, no. 5, pp. 794–801, 2017.
- [9] M. Chénard-Poirier and E. C. Smyth, “Immune checkpoint inhibitors in the treatment of gastroesophageal cancer,” *Drugs*, vol. 79, no. 1, pp. 1–10, 2019.
- [10] Y. Y. Janjigian, S. B. Maron, W. K. Chatila et al., “First-line pembrolizumab and trastuzumab in HER2-positive oesophageal, gastric, or gastro-oesophageal junction cancer: an open-label, single-arm, phase 2 trial,” *The Lancet Oncology*, vol. 21, no. 6, pp. 821–831, 2020.
- [11] A. Marabelle, A. P. A. Le DT, A. M. Di Giacomo et al., “Efficacy of pembrolizumab in patients with noncolorectal high microsatellite instability/mismatch repair-deficient cancer: results from the phase II KEYNOTE-158 study,” *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, vol. 38, no. 1, pp. 1–10, 2020.
- [12] J. Xie, L. Fu, and L. Jin, “Immunotherapy of gastric cancer: past, future perspective and challenges,” *Pathology, Research and Practice*, vol. 218, article 153322, 2021.
- [13] L. Wang, Q. Wei, M. Zhang et al., “Identification of the prognostic value of immune gene signature and infiltrating immune cells for esophageal cancer patients,” *International Immunopharmacology*, vol. 87, article 106795, 2020.
- [14] I. H. Gelman, “Suppression of tumor and metastasis progression through the scaffolding functions of SSeCKs/Gravin/AKAP12,” *Cancer Metastasis Reviews*, vol. 31, no. 3-4, pp. 493–500, 2012.
- [15] X. Wu, T. Wu, K. Li et al., “The mechanism and influence of AKAP12 in different cancers,” *Biomedical and environmental sciences: BES*, vol. 31, no. 12, pp. 927–932, 2018.
- [16] W. Liu, M. Guan, B. Su et al., “Quantitative assessment of AKAP12 promoter methylation in colorectal cancer using methylation-sensitive high resolution melting: correlation with Duke’s stage,” *Cancer Biology & Therapy*, vol. 9, no. 11, pp. 862–871, 2010.
- [17] M. C. Choi, H. S. Jong, T. Y. Kim et al., “AKAP12/Gravin is inactivated by epigenetic mechanism in human gastric carcinoma and shows growth suppressor activity,” *Oncogene*, vol. 23, no. 42, pp. 7095–7103, 2004.
- [18] W. Liu, M. Guan, T. Hu, X. Gu, and Y. Lu, “Re-expression of AKAP12 inhibits progression and metastasis potential of colorectal carcinoma in vivo and in vitro,” *Plo S one*, vol. 6, no. 8, article e24015, 2011.
- [19] R. Y. Z. Soh, J. P. Lim, R. P. Samy, P. J. Chua, and B. H. Bay, “A-kinase anchor protein 12 (AKAP12) inhibits cell migration in breast cancer,” *Experimental and Molecular Pathology*, vol. 105, no. 3, pp. 364–370, 2018.
- [20] P. He, K. Li, S. B. Li et al., “Upregulation of AKAP12 with HDAC3 depletion suppresses the progression and migration of colorectal cancer,” *International Journal of Oncology*, vol. 52, no. 4, pp. 1305–1316, 2018.
- [21] I. Kaymak, K. S. Williams, J. R. Cantor, and R. G. Jones, “Immunometabolic interplay in the tumor microenvironment,” *Cancer Cell*, vol. 39, no. 1, pp. 28–37, 2021.
- [22] F. Pontén, K. Jirström, and M. Uhlen, “The Human Protein Atlas—a tool for pathology,” *The Journal of Pathology*, vol. 216, no. 4, pp. 387–393, 2008.
- [23] T. Li, J. Fan, B. Wang et al., “TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells,” *Cancer Research*, vol. 77, no. 21, pp. e108–e110, 2017.
- [24] A. M. Newman, C. L. Liu, M. R. Green et al., “Robust enumeration of cell subsets from tissue expression profiles,” *Nature Methods*, vol. 12, no. 5, pp. 453–457, 2015.
- [25] B. Chen, M. S. Khodadoust, C. L. Liu, A. M. Newman, and A. A. Alizadeh, “Profiling tumor infiltrating immune cells with CIBERSORT,” *Methods in molecular biology (Clifton, NJ)*, vol. 1711, pp. 243–259, 2018.
- [26] M. Ghandi, F. W. Huang, J. Jané-Valbuena et al., “Next-generation characterization of the Cancer Cell Line Encyclopedia,” *Nature*, vol. 569, no. 7757, pp. 503–508, 2019.
- [27] A. Subramanian, P. Tamayo, V. K. Mootha et al., “Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.
- [28] B. Ru, C. N. Wong, Y. Tong et al., “TISIDB: an integrated repository portal for tumor-immune system interactions,” *Bioinformatics (Oxford, England)*, vol. 35, no. 20, pp. 4200–4202, 2019.
- [29] D. Szklarczyk, A. Franceschini, S. Wyder et al., “STRING v10: protein-protein interaction networks, integrated over the tree of life,” *Nucleic Acids Research*, vol. 43, no. D1, pp. D447–D452, 2015.
- [30] Y. Liao, J. Wang, E. J. Jaehnig, Z. Shi, and B. Zhang, “WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs,” *Nucleic Acids Research*, vol. 47, no. W1, pp. W199–w205, 2019.
- [31] I. Choi, B. J. Wells, C. Yu, and M. W. Kattan, “An empirical approach to model selection through validation for censored survival data,” *Journal of Biomedical Informatics*, vol. 44, no. 4, pp. 595–606, 2011.
- [32] A. Iasonos, D. Schrag, G. V. Raj, and K. S. Panageas, “How to build and interpret a nomogram for cancer prognosis,” *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, vol. 26, no. 8, pp. 1364–1370, 2008.
- [33] S. Rausei and G. D. Lianos, “Treatment of gastric cancer,” *Cancers*, vol. 12, no. 9, p. 2627, 2020.
- [34] N. W. Bateman, E. Jaworski, W. Ao et al., “Elevated AKAP12 in paclitaxel-resistant serous ovarian cancer cells is prognostic and predictive of poor survival in patients,” *Journal of Proteome Research*, vol. 14, no. 4, pp. 1900–1910, 2015.
- [35] J. S. O’Donnell, D. Massi, M. W. L. Teng, and M. Mandala, “PI3K-AKT-mTOR inhibition in cancer immunotherapy, redux,” *Seminars in Cancer Biology*, vol. 48, pp. 91–103, 2018.
- [36] D. C. Collins, M. Chenard-Poirier, and J. S. Lopez, “The PI3K pathway at the crossroads of cancer and the immune system: strategies for next generation immunotherapy combinations,” *Current Cancer Drug Targets*, vol. 18, no. 4, pp. 355–364, 2018.
- [37] M. Y. Lee, J. W. Jeon, C. Sievers, and C. T. Allen, “Antigen processing and presentation in cancer immunotherapy,” *Journal for Immunotherapy of Cancer*, vol. 8, no. 2, 2020.
- [38] J. Sun, H. Jia, X. Bao et al., “Tumor exosome promotes Th17 cell differentiation by transmitting the lncRNA CRNDE-h in colorectal cancer,” *Cell Death & Disease*, vol. 12, no. 1, p. 123, 2021.

- [39] J. Shen, J. Pan, C. Du et al., "Silencing NKG2D ligand-targeting miRNAs enhances natural killer cell-mediated cytotoxicity in breast cancer," *Cell Death & Disease*, vol. 8, no. 4, article e2740, 2017.
- [40] H. Wang, X. Wu, and Y. Chen, "Stromal-immune score-based gene signature: a prognosis stratification tool in gastric cancer," *Frontiers in Oncology*, vol. 9, p. 1212, 2019.
- [41] C. Liu, B. Chen, Z. Huang, C. Hu, L. Jiang, and C. Zhao, "Comprehensive analysis of a 14 immune-related gene pair signature to predict the prognosis and immune features of gastric cancer," *International Immunopharmacology*, vol. 89, no. Part B, article 107074, 2020.
- [42] Y. Kamihara, S. Murai, S. Kikuchi et al., "Tumor-to-tumor metastasis of diffuse large B cell lymphoma to gastric adenocarcinoma via CXCL12 (SDF-1)/CXCR4 axis: a case report," *BMC Gastroenterology*, vol. 21, no. 1, p. 270, 2021.
- [43] X. Ju, H. Zhang, Z. Zhou, M. Chen, and Q. Wang, "Tumor-associated macrophages induce PD-L1 expression in gastric cancer cells through IL-6 and TNF- α signaling," *Experimental Cell Research*, vol. 396, no. 2, article 112315, 2020.
- [44] T. Kordass, W. Osen, and S. B. Eichmuller, "Controlling the immune suppressor: transcription factors and microRNAs regulating CD73/NT5E," *Frontiers in Immunology*, vol. 9, p. 813, 2018.
- [45] B. Chen, M. Yang, K. Li et al., "Immune-related genes and gene sets for predicting the response to anti-programmed death 1 therapy in patients with primary or metastatic non-small cell lung cancer," *Oncology Letters*, vol. 22, no. 1, p. 540, 2021.

Research Article

Integrative Bioinformatics Analysis Reveals That miR-524-5p/MEF2C Regulates Bone Metastasis in Prostate Cancer and Breast Cancer

QingHua Tian, YingYing Lu, BiCong Yan, and ChunGen Wu 

Department of Diagnostic and Interventional Radiology, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China

Correspondence should be addressed to ChunGen Wu; wucgsh@163.com

Received 6 June 2022; Accepted 28 July 2022; Published 10 September 2022

Academic Editor: Lei Chen

Copyright © 2022 QingHua Tian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Bone metastases are highly prevalent in patients with advanced prostate cancer and breast cancer and have a serious impact on the survival time and quality of life of these patients. It has been reported that microRNAs (miRNAs) are expressed abnormally in different types of cancer and metastases. However, it remains unknown whether the underlying miRNAs are associated with prostate and breast cancer bone metastasis. Differentially expressed miRNAs (DE-miRNAs) and their potential targets in the metastatic process were identified by bioinformatics analysis. Additionally, qPCR confirmed that the miR-524-5p expression was downregulated in prostate and breast cancer cells. The overexpression of miR-524-5p restrained cell proliferation, invasion, and metastasis in prostate and breast cancer cells. Meanwhile, miR-524-5p could target and inhibit the expression of MEF2C, which was verified by a luciferase assay. In conclusion, our data strongly suggest that downregulation of miR-524-5p appears to be a precocious event in prostate and breast cancer, and the miR-524-5p/MEF2C axis plays a novel role in bone metastases from prostate and breast cancers.

1. Introduction

Prostate cancer and breast cancer are the two most common invasive cancers in men and women, respectively. Although these two cancers arise from different organs, they are typically hormone-dependent and have an underlying biological mechanism in common [1]. As cancer is a metastatic disease, hormone therapy plays an essential role in receptor-positive breast and prostate cancer. However, although hormone therapy initially has an effect on the inhibition of breast and prostate tumours, bone metastasis is still a problem for 70% of metastatic prostate and breast cancer patients [2], showing a poor prognosis. In addition, prostate cancer and breast cancer are malignancies that are destined to become metastatic if screening has been unable to identify them at an early stage before symptoms appear [3], causing a serious threat to patients' lives.

Bone is the tissue most susceptible to metastasis in prostate cancer [4] and breast cancer [5]. Once cancer has

metastasized to the bones, and numerous skeletal-related events (SREs), such as fracture, intractable pain, bone marrow aplasia, nerve compression syndrome, and spinal cord compression, occur [6], which can rarely be cured and result in significant morbidity in patients with prostate cancer and breast cancer. Despite its morbidity, bone metastasis is one of the most intriguing and complex biological processes of all oncogenic processes and consists of the three key steps of seeding, dormancy, and outgrowth [7]. All of the processes involved in metastasis and interaction with host cells can be targeted to treat bone metastasis and tumor progression in prostate cancer and breast cancer.

MicroRNAs (miRNAs or miRs-), which are 21-25 nucleotides long, are single-stranded noncoding RNAs that are evolutionarily conserved and endogenously produced. miRNAs play an essential role in targeting the 3' untranslated region (3' UTR) of mRNAs, mainly to repress their expression [8]. Over the past several years, many miRNAs have been found and characterized in the pathogenesis of

many human malignancies, including critical chemokines and cytokines in the bone metastasis microenvironment [9, 10]. For example, it was reported that miR-124 could inhibit bone metastasis by inhibiting interleukin-11 in breast cancer [11]. miR-133a-3p represses bone metastasis of prostate cancer [12]. In addition, the occurrence and development of many kinds of cancer, such as colon cancer [13], melanoma [14], gastric cancer [15], osteosarcoma [16], ameloblastoma [17], and particularly breast cancer [18], are highly associated with the abnormal expression of miRNAs. Jin et al. identified miR-524-5p as a tumor suppressor. Moreover, miR-524-5p inhibited cell migration, invasion, and epithelial-mesenchymal transition and progression in breast cancer [18]. However, little is known about its function in bone metastases of prostate and breast cancer.

Here, for the first time, we revealed the role of cancer cell-derived miR-524-5p in bone metastases of prostate cancer and breast cancer. Perturbation of the miR-524-5p/MEF2C regulatory axis contributes to bone metastasis in prostate cancer and breast cancer. These results might provide novel therapeutic and diagnostic targets for bone metastases of breast cancer and prostate cancer.

2. Materials and Methods

2.1. Download of mRNA and miRNA Expression Profiles. The expression of genes and miRNAs between patients with primary prostate and breast cancer with bone metastatic prostate and breast cancer was compared. The Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) was employed to obtain the gene expression profiles and the miRNA expression profile. One dataset, GSE32269, deposited by Cai et al. [19], contains 22 primary prostate cancer (hormone-dependent) versus 29 metastatic prostate cancer samples. Another dataset, GSE137842, was submitted by Lefley et al. [20] and contains 3 primary breast cancer and 3 breast cancer bone metastasis samples. These two datasets were acquired from the Affymetrix Human Genome U133 Plus 2.0 Array. The miRNA microarray dataset GSE26964 [21] was composed of 6 primary prostate cancer samples and 7 prostate cancer bone metastatic samples (platform: Capitalbio mammal microRNA V3.0).

2.2. Identification of Differentially Expressed miRNAs and mRNAs. Herein, the mRNA and miRNA expression profile data preprocessing mainly consists of background correction, quantile normalization, and probe summarization [22]. Then, limma package in bioconductor was used to extract differentially expressed miRNAs (DE-miRNAs) and differentially expressed genes (DEGs) [23] following criterion P value < 0.05 and $|\log_2 \text{foldchange}| > 1$.

2.3. GO and KEGG Pathway Annotation. The online tool of Database for Annotation and Visualization and Integrated Discovery (DAVID) was used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment of the identified common DEGs and DE-miRNAs [24, 25]. GO analysis was conducted

for the cellular component (CC), biological process (BP), molecular function (MF) categories [26], and KEGG pathway enrichment analysis for the selected genes [27]. All parameters were set as default, and P value < 0.01 was considered significant.

2.4. Construction of the mRNA-miRNA Regulation Network. miRBase, TargetScan (http://www.targetscan.org/vert_71/), miTarBase, and miRWalk databases were used to identify the number of miRNA-regulated target gene pairs. The threshold of the correlation coefficient was set as -0.3 , and the significance P value was set as 0.05 . The pairs supported by two or more databases were further processed and retained. Regulatory network visualization for the regulatory relationship between miRNA-mRNA was conducted using Cytoscape [28].

2.5. PPI Network and Hub Gene Analysis. The interaction of the DEGs was detected using the Search Tool for the Retrieval of Interacting Genes (STRING) database, with confidence scores > 0.7 . Cytoscape was used to visualize the PPI network. Hub genes and screen modules of the PPI network were identified using the CytoHubba plug-in and Molecular Complex Detection (MCODE) plug-in, of which all of the parameters were left as the defaults. Metascape was used to analyze the genes in modules.

2.6. Cell Culture. The DU145 and LNCAP cell lines of human prostate cancer and MCF7 of breast cancer were obtained from COWELDGEN SCIENTIFIC (Coweldgen Scientific Co., Ltd., Shanghai, China). DU145 and MCF-7 cells were cultured in MEM (Invitrogen, USA) supplemented with 10% FBS (Invitrogen, USA). LNCAP was cultured in RPMI 1640 medium (Invitrogen, USA) with 10% FBS. All medium were supplemented with penicillin (100 U/ml) and streptomycin (100 mg/ml) (Biosharp, China). All cells were cultured at 5% CO_2 and 37°C humidified atmosphere.

2.7. Cell Transfection. Transfection was conducted when the cell density reached 70-80% in six-well plates. miR-524-5p mimics and negative control (NC) were designed and synthesized by GenePharma (Shanghai, China). miR-524-5p mimics or NC (50 nM) was transfected into cells using Lipofectamine 3000 (Invitrogen, USA) following the manufacturer's instructions.

2.8. RNA Extraction and qPCR Analyses. Total RNA was extracted from cell lines with TRIzol Reagent (Invitrogen, USA) following the manufacturer's instructions. In details, total RNA (2.0 μg) was reverse transcribed using the PrimeScript TM RT reagent Kit (Takara, Japan) following the manufacturer's instructions. cDNA was amplified and quantified on the LightCycler® 96 system (Roche, Switzerland) using SYBR® Premix Ex Taq™ II (Takara, Japan). miRcute Plus miRNA First-Strand cDNA Kit (TianGen Biotech, Beijing, China) and SYBR Green (Takara, Japan) were used to quantify mature miRNA levels. $2^{-\Delta\Delta\text{Ct}}$ was calculated to represent the relative fold expressions. The primers are shown in Table 1.

TABLE 1: List of primers.

Gene name	Primer sequence
β -Actin-F	GGCTGTGCTATCCCTGTACG
β -Actin-R	GGCTGTGCTATCCCTGTACG
MEF2C-F	GCACCAACAAGCTGTTCCAG
MEF2C-R	TGTCTGAGTTTGTCCGGCTC
miR-U6-F	CTCGCTTCGGCAGCAC
miR-U6-R	AACGCTTCACGAATTTGCGT
miR-524-5p	CTACAAAGGGAAGCACTTTCTC

2.9. Western Blot. Total protein was extracted from cell lysate using PMSF. Then, protein sample was quantified by bicinchoninic acid (BCA), separated by SDS-PAGE gel electrophoresis, and blocked with 5% skim milk. After that, membranes were incubated with the specific primary antibody at 4°C overnight. Membranes were washed three times using TBST and incubated with the corresponding HRP-conjugated secondary antibody for 2 h at room temperature. Enhanced chemiluminescence (PerkinElmer) was used to expose membranes after washing three times. Western blots were quantified by densitometry using ImageJ.

2.10. Dual-Luciferase Reporter Gene Assay. The 3' UTR sequence of MEF2C containing wild-type or mutant binding sites was cloned into the pmirGLO luciferase vector. Then, the MEF2C 3' UTR WT or MUT and miR-524-5p mimics or negative control were cotransfected using Lipofectamine 3000. After 72 h, the cell lysate was centrifuged at 12,000 rpm for 5 min to perform the Dual-Lumi Luciferase Assay (Beyotime Biotechnology, Shanghai, China). Luciferase activity was measured by an EnVision Multifunctional Microplate Reader (PerkinElmer, Germany).

2.11. Cell Counting Kit-8 (CCK-8) Assay. A 100 μ L cell suspension (1×10^4 cells) was placed in each well of a 96-well plate. Ten microliters of CCK-8 reagent (Beyotime Biotechnology, Shanghai, China) was added at 6 h, 24 h, 48 h, 72 h, and 96 h. Then, the cells were cultured for 2 h, and the absorbance value was detected at a wavelength of 450 nm by a microplate reader (SpectraMax M5; Molecular Devices, USA).

2.12. Transwell Assay. A cell suspension (1×10^5 cells/ml) was prepared with serum-free medium. Then, 100 μ L suspension was added to the chamber, and 600 μ L of complete medium was supplied to the basolateral chamber, which was incubated overnight. Subsequently, unpenetrating cells above the chamber were removed, the chamber was fixed in 4% paraformaldehyde for 30 min and then dyed with 1% crystal violet for 10-15 min, and five randomly selected fields were captured using an inverted microscope (Olympus, Tokyo, Japan).

2.13. In Vitro 3D Model of Prostate and Breast Cancer Metastasis. The animal experiments performed in this study were approved by the Institutional Animal Care and Use Committee of ShangHai Sixth People's Hospital. To imitate

the bone metastasis microenvironment in vivo, neonatal CD-1 mice were used to build an in vitro 3D model. In detail, after CD-1 mice were sacrificed, their calvarial bone was separated under sterile conditions and cut in the occipital lobe to produce an arch structure [29]. Then, calvarial bones were washed with PBS and cocultured with DU145 cells or MCF7 cells and DU145 cells or MCF7 cells transfected with miR-524-5p mimics in a 48-well plate (5×10^5 cells). Cranium bones without culturing with cells were used as a negative control. The crystal violet staining experiment was performed after incubation at 37°C for 4 days. In brief, the bone fragments were removed, fixed with 95% alcohol for 10 min, washed with PBS 3 times, and stained with 5 mg/ml crystal violet for 15 min. The stained bone slices were washed with PBS 3 times and observed under an inverted microscope. Each bone slice was randomly taken from 3 different fields to count the adhered cells [30, 31].

2.14. Statistical Processing. SPSS software (22.0, US) was utilized for statistical analysis. All data were represented as mean \pm SD. Statistical differences were determined by a one or two-sided Student's *t*-tests or two-way ANOVAs. A *P* value <0.05 was considered statistically significant.

3. Results

3.1. Identification of DE-miRNAs and DEGs. A total of 1555 DEGs were obtained from breast cancer gene expression profiles of GSE137842, of which 625 DEGs were overexpressed and 930 DEGs were downregulated in breast cancer bone metastasis samples compared to primary breast cancer samples. Volcano plots were generated to visualize all the DEGs (Figures 1(a) and 1(b)). A total of 833 DEGs were identified from prostate cancer gene expression profiles of GSE32269, of which 522 were upregulated and 311 were downregulated in prostate cancer bone metastasis samples compared to primary prostate cancer samples. Volcano plots were generated to visualize all the DEGs (Figures 1(c) and 1(d)). A total of 88 DE-miRNAs were obtained from prostate cancer miRNA expression profiles of GSE26964, of which 11 were upregulated and 77 were downregulated in prostate cancer bone metastatic samples compared to primary prostate cancer samples. Volcano plots were generated to visualize all DE-miRNAs (Figures 1(e) and 1(f)). We overlapped the DEGs (mRNA) screened in the prostate cancer bone metastasis dataset and breast cancer bone metastasis and identified 17 genes that were upregulated and 20 that were downregulated (Figures 1(g) and 1(h)).

3.2. Pathway and Process Enrichment Analysis of DEGs. For the selected common DEGs, pathway and process enrichment analyses were conducted through GO processes and KEGG pathways [32]. The representative top 15 clusters from the 3 categories are shown in Figure 2(a). The terms enriched in the biological process (BP) category included cardiac tissue and animal organ development, embryonic organ morphogenesis, cardiac muscle cell differentiation, bone development, and bone morphogenesis. The molecular

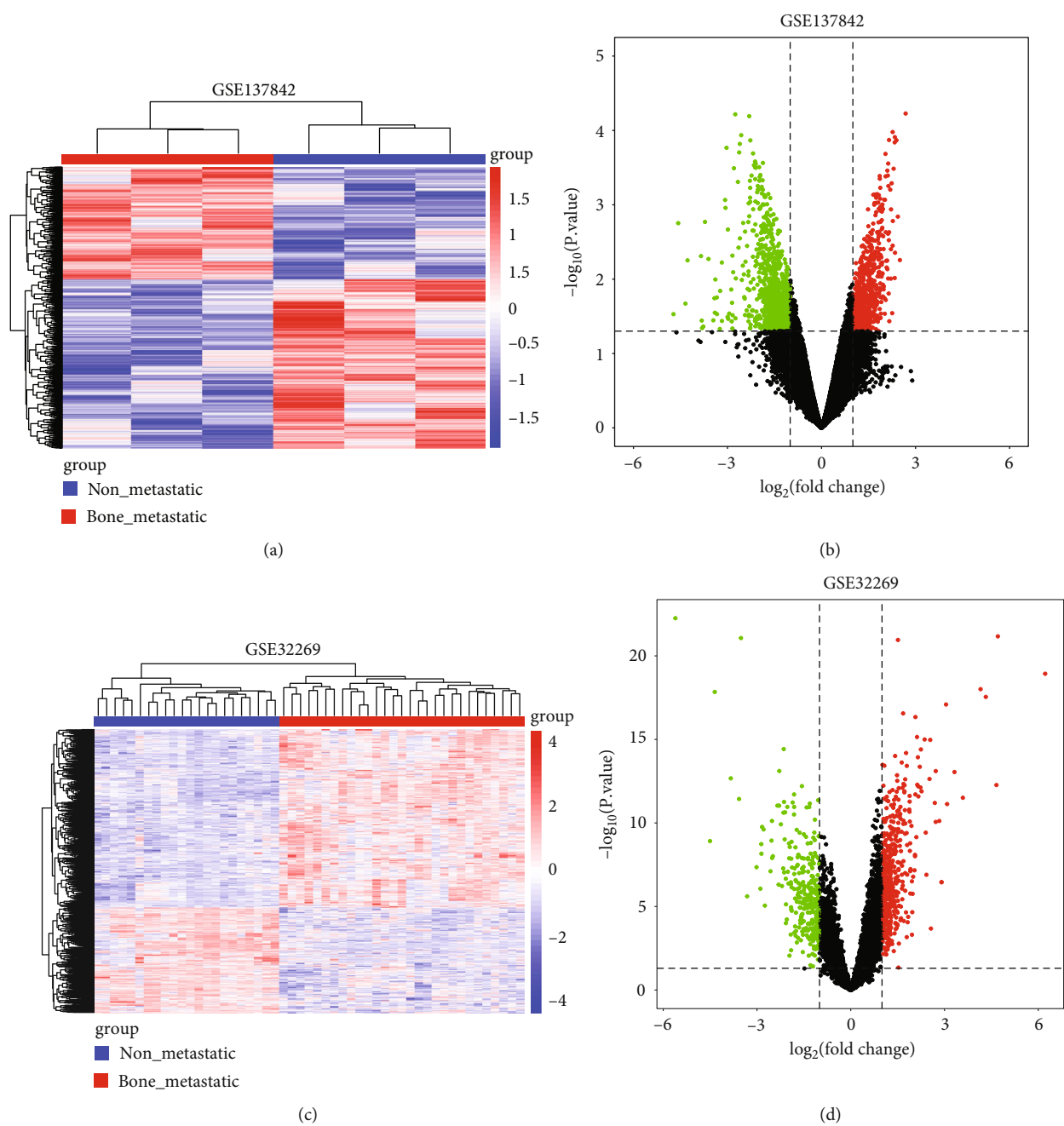


FIGURE 1: Continued.

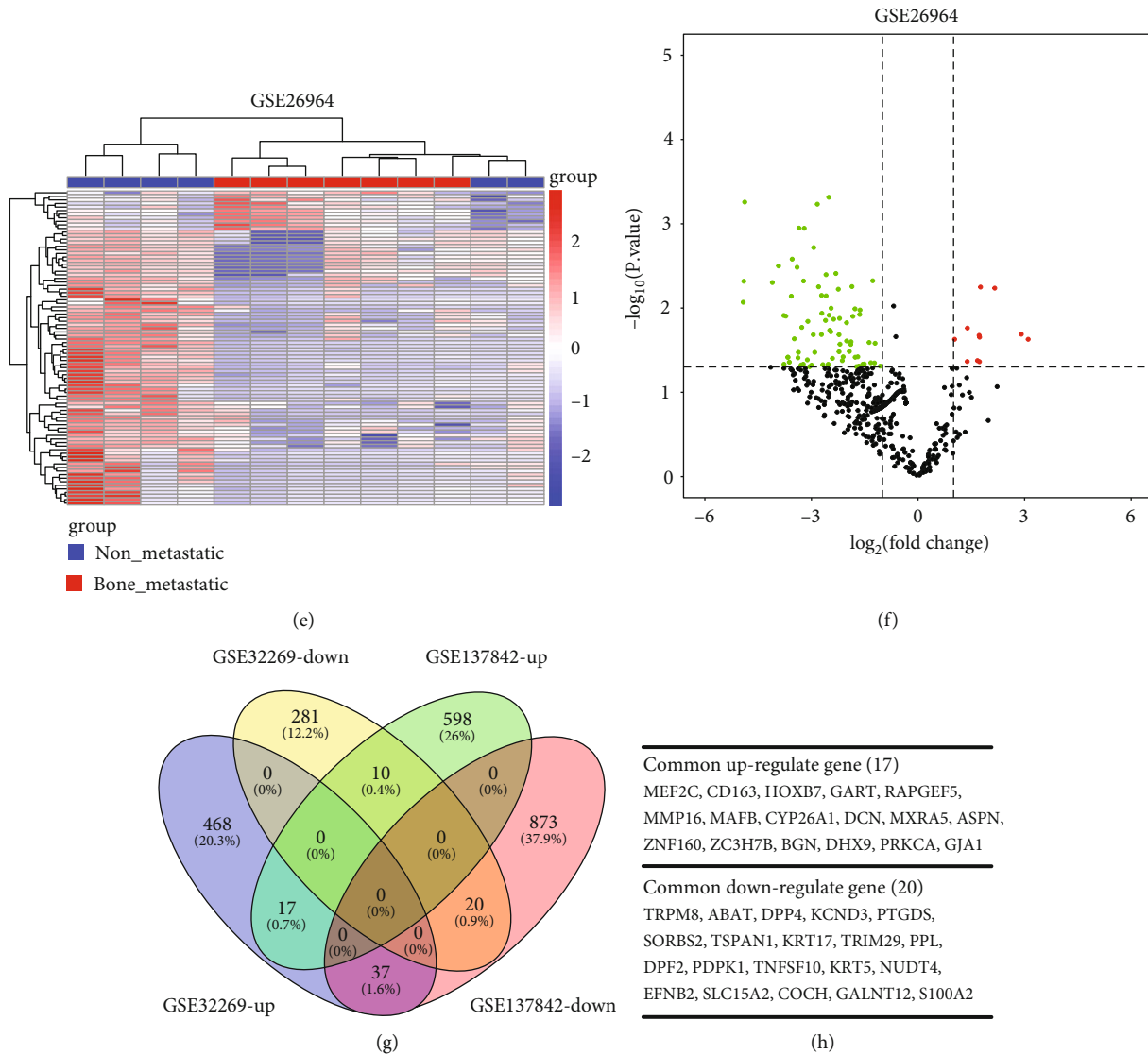


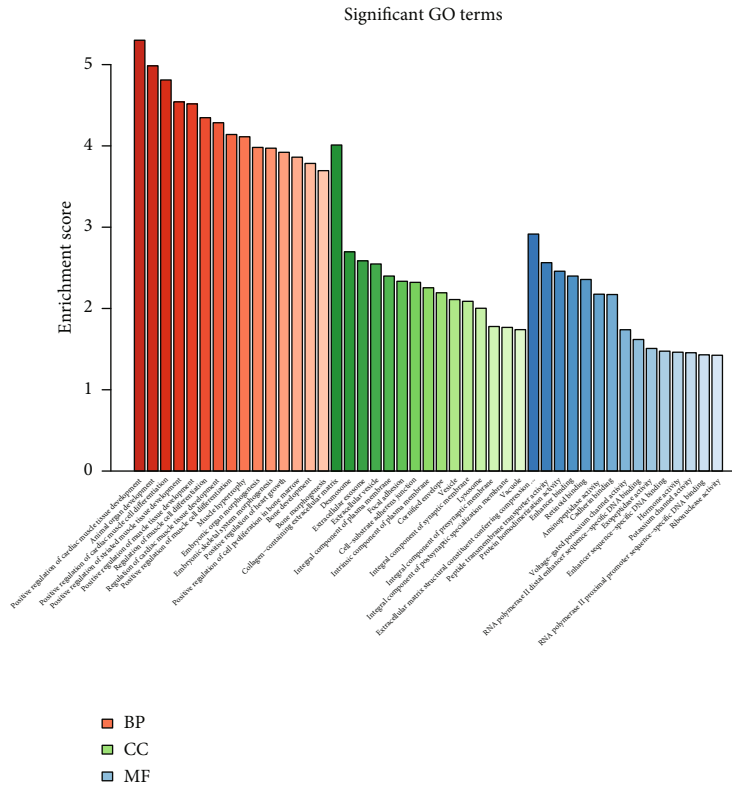
FIGURE 1: Identification of DE-miRNAs and DEGs related to cancer bone metastasis. (a) Clustered heat map of 1555 DEGs in GSE137842. Red: significantly upregulated genes; blue: significantly downregulated genes. (b) Differentially expressed volcano plots in GSE137842. Red dots: significantly upregulated; green dots: significantly downregulated; and black dots: no significant differences. (c) Clustered heat map of 833 DEGs in GSE32269. Red: significantly upregulated genes; blue: significantly downregulated genes. (d) Differentially expressed volcano plots in GSE32269. Red dots: significantly upregulated; green dots: significantly downregulated; and black dots: no significant differences. (e) Clustered heat map of 88 DE-miRNAs in GSE26964. Red: upregulated genes; blue: downregulated genes. (f) Differentially expressed volcano plots in GSE26964. Red dots: significantly upregulated; green dots: significantly downregulated; and black dots: no significant differences. (g) The common DEGs from GSE32269 and GSE137842 involved in prostate cancer bone metastasis and breast cancer bone metastasis by Venny 2.1.0. (h) List of the common DEGs from GSE32269 and GSE137842.

function category demonstrated enrichment in factors involved in the extracellular exosome, collagen-containing extracellular matrix, cell-substrate adherens junction, and focal adhesion. In addition, the GO cell component category showed enrichment in cadherin binding, ribonuclease activity, retinoid binding, DNA binding, enhancer binding, and aminopeptidase activity.

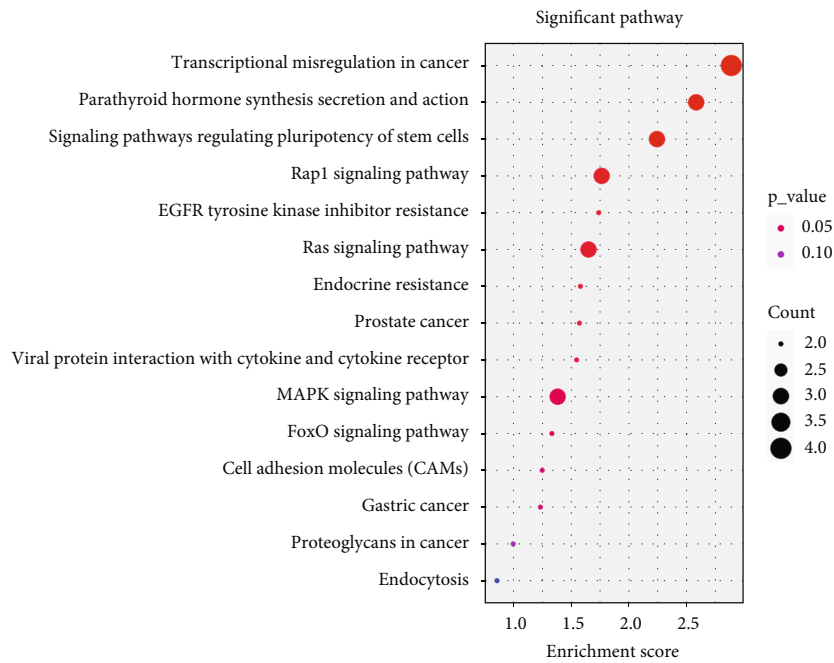
The results from KEGG pathway enrichment analysis demonstrated that common DEGs were significantly involved in transcriptional misregulation in cancer, parathyroid hormone synthesis secretion and action, regulating pluripotency of stem cells, MAPK signaling, Ras signaling pathways, EGFR

tyrosine kinase inhibitor resistance, prostate cancer, viral protein interaction with cytokine and cytokine receptor, FoxO adhesion molecules, gastric cancer, proteoglycans in cancer, and endocytosis (Figure 2(b)).

3.3. Construction of the mRNA-miRNA Regulation Network and PPI Network. Three target prediction databases were employed to identify the target genes of selected DE-miRNAs. In total, 37 miRNAs were identified for 88 DE-miRNAs, which had different expression levels between patients with primary prostate cancer and patients with prostate cancer bone metastasis. In addition, 13 of the target



(a)



(b)

FIGURE 2: Visualizations of pathway and process enrichment analysis results of DEGs. (a) GO term analysis of DEGs. The vertical axis represents the enrichment score value enriched on the GO term item, and the horizontal axis represents the name of the corresponding GO term item in the GO database. The figure only shows the top 15 terms of each category. BP: biological process; CC: cellular component; MF: molecular function. (b) KEGG enrichment analysis of DEGs. The top 15 clusters with their representative KEGG pathways. The color changes gradually from blue to red, which represents the significant level of enrichment to the pathway, and the redder the color, the greater the significance. The horizontal axis represents the enrichment score of the pathway, the vertical axis represents the corresponding pathway name in the KEGG database, and the size of the circle represents the number of genes enriched in the pathway.

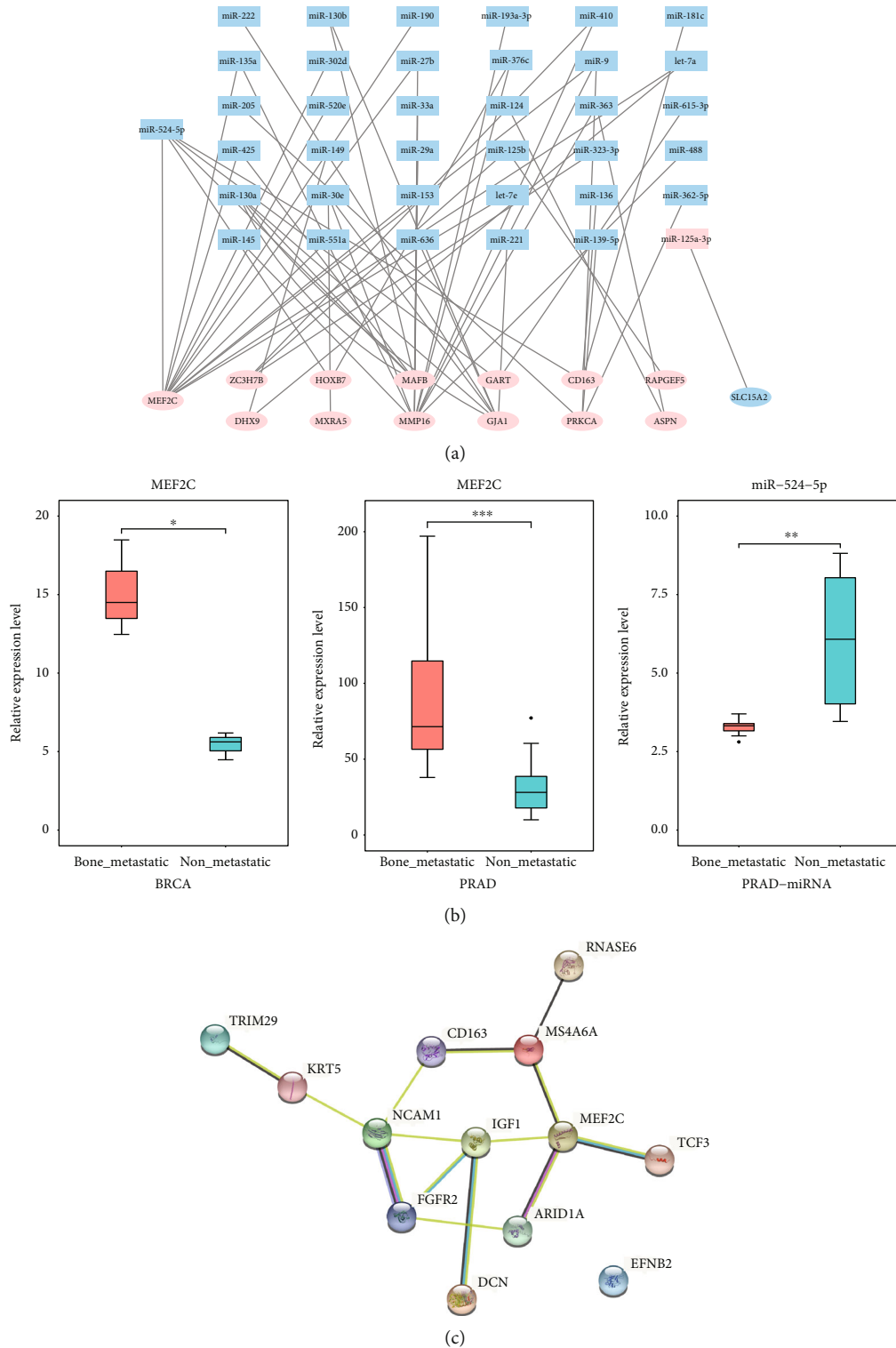


FIGURE 3: The landscape of mRNA-miRNA interactions and PPI networks. (a) mRNA-miRNA interaction network. The graphic visualization uses different shapes representing different gene types: oval represents mRNA, and square represents miRNA. Different colors represent different gene expressions: blue represents downregulated genes, and red represents upregulated genes. (b) Differential expression of MEF2C in bone metastatic and nonbone metastatic prostate and breast cancers. The expression of miR-524-5p in bone metastatic and nonbone metastatic prostate cancer. (c) Protein-protein interaction (PPI) network of DEGs. The globules represent proteins, and the line between the globules represents the interaction between the two proteins.

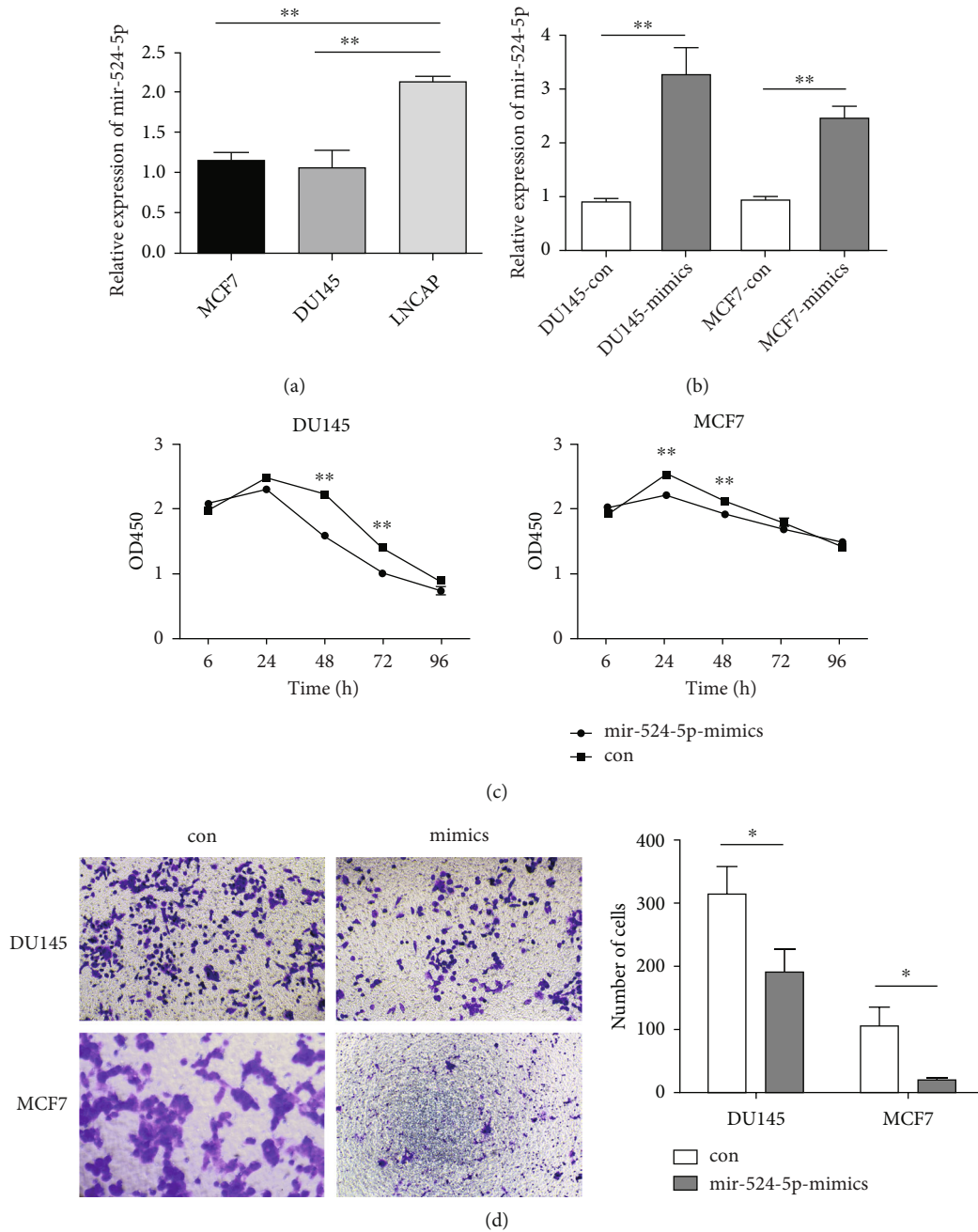


FIGURE 4: miR-524-5p repressed the proliferation and invasion of DU145 and MCF7 cells. (a) The expression of miR-524-5p in MCF7, DU145, and LNCAP cells. (b) The expression of miR-524-5p was increased in DU145 and MCF7 cells after transfection with miR-524-5p mimics. (c) Inhibition of proliferation of DU145 or MCF7 cells after transfection with miR-524-5p mimics by CCK-8 assay. (d) Inhibition of migration of DU145 or MCF7 cells after transfection with miR-524-5p mimics by transwell assay.

genes (MEF2C, ZC3H7B, HOXB7, MAFB, GART, CD163, RAPGEF5, DHX9, MXRA5, MMP16, GJA1, PRKCA, ASPN) were upregulated, and one was downregulated. Thirty-seven DE-miRNAs and 14 target genes were identified in the miRNA-mRNA regulatory network, as shown in Figure 3(a). Among those DEGs, MEF2C showed the extremely high expression in patients with bone metastatic prostate and breast cancer (Figure 3(b)). Notably, MEF2C was one of the targets of miR-524-5p.

PPI networks were constructed through the string database from 13 proteins (confidence level of 0.4) consisting of 13 nodes and 14 edges for the DEGs in bone metastasis. The PPI network analysis showed that MEF2C, FGFR2, IGF1, and NCAM1 were hub genes (Figure 3(c)).

3.4. miR-524-5p Overexpression Restored Cell Proliferation and Invasion in Prostate and Breast Cancer. miR-524-5p was found to be one of the most significantly downregulated

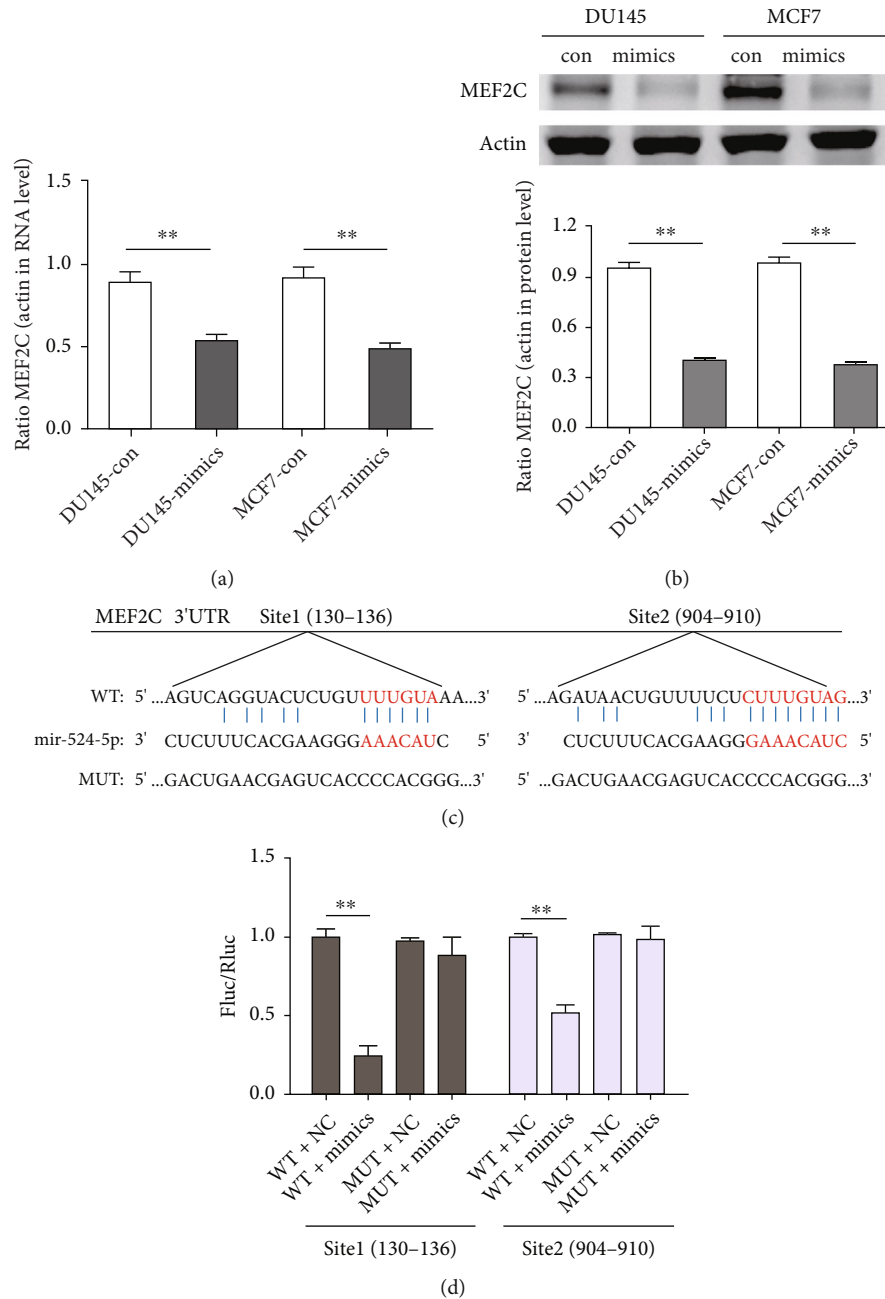


FIGURE 5: miR-524-5p specifically targets MEF2C. (a) The expression of MEF2C in DU145 or MCF7 cells after transfection with miR-524-5p mimics. (b) The protein expression of MEF2C in DU145 or MCF7 cells after transfection with miR-524-5p mimics. (c) Schematic representation of the predicted target site for miR-524-5p in MEF2C. (d) Luciferase activity assay in the wild-type or mutant reporter containing MEF2C 3' UTR (two sites) when transfected with miR-524-5p mimics or NC in 293 T cells.

miRNAs in prostate and breast cancer with bone metastasis (Figures 3(a) and 3(b)). To further validate this result, the expression of miR-524-5p in the human prostate cancer cell lines DU145 and LNCAP and breast cancer cell line MCF7 was measured. qPCR showed that the expression of miR-524-5p was lower in MCF7 and DU145 cell lines than in LNCAP (Figure 4(a)).

To explore the biological function of miR-524-5p, miR-524-5p mimics were transfected into DU145 and MCF7 cell lines. The expression of miR-524-5p in DU145 and MCF7

cells was significantly increased in transfected cells (Figure 4(b)). In the CCK-8 assay, the cell proliferation ability of DU145 and MCF7 cells was inhibited by miR-524-5p mimics (Figure 4(c)). In the transwell assay, miR-524-5p mimics inhibited the migration ability of DU145 and MCF7 cells (Figure 4(d)).

3.5. miR-524-5p Specifically Targets MEF2C. MEF2C, a transcription factor, has been proposed as a new player in breast cancer brain metastasis development [33]. In contrast to

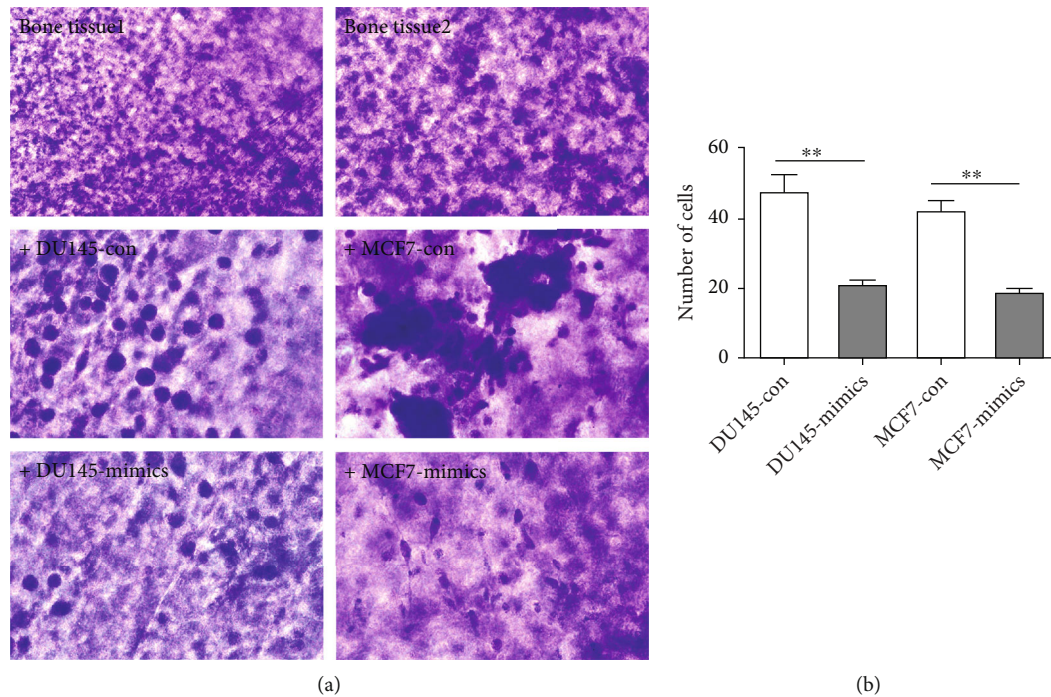


FIGURE 6: miR-524-5p promoted the metastasis of DU145 or MCF7 cells in an in vitro 3D model. (a) Bone tissue was cocultured with MCF7 or DU145 cells with or without miR-524-5p mimics. (b) The number of MCF7 or DU145 cells adhering to calvarial bone tissue was counted.

miR-524-5p, MEF2C was upregulated in prostate and breast cancer with bone metastasis, and the MEF2C gene was predicted to be a target for miR-631, miR-524-5p, miR-330-3p, and miR-346 (Figure 3(a)). qPCR results showed that MEF2C mRNA significantly decreased in cells transfected with miR-524-5p mimics compared with the NC group (Figure 5(a)), and Western blotting demonstrated that the MEF2C protein was also dramatically decreased (Figure 5(b)). The data indicated that MEF2C is most likely the target of miR-524-5p.

TargetScan predicted that there were binding sites of miR-524-5p in the 3' UTR of MEF2C at bases 130-136 (site 1) and 904-910 (site 2) (Figure 5(c)). To further confirm the inference that miR-524-5p targeted MEF2C, a luciferase reporter assay was conducted. For site 1, luciferase activities decreased in the cells cotransfected with wild-type MEF2C and miR-524-5p mimics compared with those in the NC group. However, there were no differences in the luciferase activities compared with their control group after cotransfection with MEF2C mutation (mut) and miR-524-5p mimics. For site 2, there was less change in the luciferase activities compared with those in their control group after cells were cotransfected with wild-type MEF2C and miR-524-5p mimics (Figure 5(d)). Therefore, MEF2C could be targeted by miR-524-5p through binding to both sites.

3.6. Effect of miR-524-5p on an In Vitro 3D Model of Prostate or Breast Cancer Bone Metastasis. Generally, the calvarial bone tissue was smooth and flat. However, in the model group cultured with DU145 or MCF7 cells, many DU145 or MCF7 cells obviously grew on the surface of calvarial bone. The number of DU145 or MCF7 cells on the calvarial

bone was measured through crystal violet staining. The miR-524-5p mimics inhibited the growth of DU145 or MCF7 cells on the surface of the calvarial bone tissue (Figures 6(a) and 6(b)), which proved that the calvarial bone had strong adhesion to tumor cells. The adhered tumor cells caused damage to the calvarial bone in the coculture group. However, the overexpression of miR-524-5p repressed the adhesion and metastasis of tumor cells to bone tissue.

4. Discussion

As we described previously, bone is the most preferential metastatic site for prostate and breast cancer. Bone metastasis for prostate and breast cancer is a multistep process including tumor cell dissemination into the circulation, homing to the bone, and proliferation in bone tissue. A complicated network of molecular events plays an essential role in the development of bone metastasis. However, the underlying mechanism is not fully understood. In this study, a comprehensive genetic interaction network was established to explore the role of potential miRNAs, particularly miR-524-5p and its target MEF2C. We found that MEF2C was upregulated in bone metastasis development and regulated by miR-524-5p. These findings provide new insights into the altered miRNAs serving as potential biomarkers and MEF2C as a potential target for preventing or breaking prostate and breast cancer metastasis.

Several lines of literature have reported that miRNAs play a key role in cancer progression and metastasis. We found 11 downregulated miRNAs, including miR-524-5p, miR-330-3p, and miR-346, and 77 upregulated miRNAs,

such as miR-564, miR-602, and miR-129-5p, when searching for miRNAs with aberrant expression in prostate metastasis. Overexpressed miRNAs in various tumors have been found to be oncogenic miRNAs [34–37]. Other miRNAs with decreased expression are predicted to be tumor suppressors [38–40]. Among those miRNAs, as revealed by target prediction, miR-524-5p emerged as the most promising molecule and was reported to be involved in various cancers through different mechanisms. It has been predicted that miR-524-5p plays a tumor suppressor role in multiple types of cancers. Several studies have reported that miR-524-5p was downregulated in tumors, including glioma [41], colon cancer cells [13], gastric cancer [15], and papillary thyroid carcinoma [42], which can inhibit tumor proliferation and metastasis. In breast cancer, Jin et al. found that miR-524-5p inhibited the progression of migration, invasion, and epithelial–mesenchymal transition by targeting FSTL1 [18]. Consistent with Jin et al.’s result, our study confirmed that miR-524-5p was downregulated in prostate cancer bone metastasis compared to its expression in primary prostate cancer, and that the expression of miR-524-5p inhibited cell proliferation and invasion ability. The contradictions in previous reports may have arisen because a single miRNA can regulate various genes and functions or influence the expression of multiple factors in different cellular contexts [43].

The myocyte enhancer factor 2 (MEF2) protein family includes MEF2A, MEF2B, MEF2C, and MEF2D. MEF2C is widely expressed in muscle, neuronal, chondroid, immune, and endothelial cells [44]. In addition, MEF2C has close connections with uncontrolled cancer cell proliferation and enhanced invasion [45]. Regarding cancer metastasis, a recent study reported that MEF2C was consistently expressed in breast cancer brain metastases, and that its nuclear translocation was related to brain metastatic disease severity via VEGFR-2 and β -catenin signaling [46]. MEF2C was predicted to be regulated by miR-802-5p and miR-194-5p in brain metastases of breast cancer, which indicates that MEF2C plays a role in tumor metastasis. In this study, we found that MEF2C is upregulated in prostate and breast cancer bone metastasis. MEF2C was regulated by miR-524-5p, which was confirmed by luciferase assays and in vitro experiments. The absence of miR-524-5p in primary cancer tissue may promote bone metastasis through the upregulation of MEF2C. However, the specific downstream mechanism involved needs to be further investigated. Collectively, downregulation of miR-524-5p appears to be a precocious event in prostate and breast cancer, and MEF2C serves as a new player in prostate and cancer bone metastasis development.

Data Availability

The miRNA microarray data (GSE26964) and mRNA microarray data (GSE137842, GSE32269) used to support the findings of this study have been deposited in the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) in NCBI.

Disclosure

A preprint has previously been published [47].

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was sponsored by grant National Natural Scientific Fund of China (grant numbers 81701798, 81703751), Natural Fund from Shanghai Science And Technology Commission (grant numbers 18ZR1429400, 19411971800), and Natural Fund from Shanghai Municipal Health Commission (grant number 202040340), China.

Supplementary Materials

Supplementary 1. Table S1: list of upregulated expressed mRNAs in breast cancer bone metastasis samples from the GSE137842 dataset.

Supplementary 2. Table S2: list of downregulated expressed mRNAs in breast cancer bone metastasis samples from the GSE137842 dataset.

Supplementary 3. Table S3: list of upregulated expressed mRNAs in prostate cancer bone metastasis samples from the GSE32269 dataset.

Supplementary 4. Table S4: list of downregulated expressed mRNAs in prostate cancer bone metastasis samples from the GSE32269 dataset.

Supplementary 5. Table S5: list of upregulated expressed miRNAs in prostate cancer bone metastatic samples from the GSE26964 dataset.

Supplementary 6. Table S6: list of downregulated expressed miRNAs in prostate cancer bone metastatic samples from the GSE26964 dataset.

References

- [1] G. P. Risbridger, I. D. Davis, S. N. Birrell, and W. D. Tilley, “Breast and prostate cancer: more similar than different,” *Nature Reviews Cancer*, vol. 10, no. 3, pp. 205–212, 2010.
- [2] R. K. Hernandez, S. W. Wade, A. Reich, M. Pirolli, A. Liede, and G. H. Lyman, “Incidence of bone metastases in patients with solid tumors: analysis of oncology electronic medical records in the United States,” *BMC Cancer*, vol. 18, no. 1, p. 44, 2018.
- [3] H. G. Welch, D. H. Gorski, and P. C. Albertsen, “Trends in metastatic breast and prostate cancer—lessons in cancer dynamics,” *The New England Journal of Medicine*, vol. 373, no. 18, pp. 1685–1687, 2015.
- [4] R. R. Langley and I. J. Fidler, “The seed and soil hypothesis revisited—the role of tumor-stroma interactions in metastasis to different organs,” *International Journal of Cancer*, vol. 128, no. 11, pp. 2527–2535, 2011.
- [5] C. Liu, L. Wang, J. Zhuang et al., “Should de-escalation of bone-targeting agents be standard of care for patients with

- bone metastases from breast cancer? A systematic review and meta-analysis," *Annals of Oncology*, vol. 29, no. 5, pp. 1329–1330, 2018.
- [6] F. Saad, A. Lipton, R. Cook, Y. M. Chen, M. Smith, and R. Coleman, "Pathologic fractures correlate with reduced survival in patients with malignant bone disease," *Cancer*, vol. 110, no. 8, pp. 1860–1867, 2007.
- [7] M. Esposito, T. Guise, and Y. Kang, "The Biology of Bone Metastasis," *Cold Spring Harbor Perspectives in Medicine*, vol. 8, no. 6, 2018.
- [8] F. Wahid, A. Shehzad, T. Khan, and Y. Y. Kim, "MicroRNAs: synthesis, mechanism, function, and recent clinical trials," *Biochimica et Biophysica Acta*, vol. 1803, no. 11, pp. 1231–1243, 2010.
- [9] Y. S. Lee and A. Dutta, "MicroRNAs in cancer," *Annual Review of Pathology*, vol. 4, no. 1, pp. 199–227, 2009.
- [10] E. Hesse and H. Taipaleenmäki, "MicroRNAs in bone metastasis," *Current Osteoporosis Reports*, vol. 17, no. 3, pp. 122–128, 2019.
- [11] W. L. Cai, W. D. Huang, B. Li et al., "microRNA-124 inhibits bone metastasis of breast cancer by repressing Interleukin-11," *Molecular cancer*, vol. 17, no. 1, p. 9, 2018.
- [12] Y. Tang, J. Pan, S. Huang et al., "Downregulation of miR-133a-3p promotes prostate cancer bone metastasis via activating PI3K/AKT signaling," *Journal of Experimental & Clinical Cancer Research*, vol. 37, no. 1, p. 160, 2018.
- [13] X. Li, Z. Li, Y. Zhu et al., "miR-524-5p inhibits angiogenesis through targeting WNK1 in colon cancer cells," *American Journal of Physiology. Gastrointestinal and Liver Physiology*, vol. 318, no. 4, pp. G827–g839, 2020.
- [14] M. T. Nguyen, C. H. Lin, S. M. Liu et al., "miR-524-5p reduces the progression of the BRAF inhibitor-resistant melanoma," *Neoplasia*, vol. 22, no. 12, pp. 789–799, 2020.
- [15] C. Y. Zhu, F. Q. Meng, and J. Liu, "MicroRNA-524-5p suppresses cell proliferation and promotes cell apoptosis in gastric cancer by regulating CASP3," *European Review for Medical and Pharmacological Sciences*, vol. 23, no. 18, pp. 7968–7977, 2019.
- [16] H. Yang, C. Cao, and L. J. Wang, "LncRNA LINC-PINT regulating proliferation and apoptosis of osteosarcoma cells by targeting miR-524-5p," *Zhonghua Zhong Liu Za Zhi*, vol. 42, no. 4, pp. 325–330, 2020.
- [17] L. Chen, G. Wang, X. Qiao et al., "Downregulated miR-524-5p participates in the tumor microenvironment of Ameloblastoma by targeting the Interleukin-33 (IL-33)/suppression of Tumorigenicity 2 (ST2) Axis," *Medical Science Monitor*, vol. 26, article e921863, 2020.
- [18] T. Jin, Y. Zhang, and T. Zhang, "MiR-524-5p suppresses migration, invasion, and EMT progression in breast cancer cells through targeting FSTL1," *Cancer Biotherapy & Radiopharmaceuticals*, vol. 35, no. 10, pp. 789–801, 2020.
- [19] C. Cai, H. Wang, H. H. He et al., "ERG induces androgen receptor-mediated regulation of SOX9 in prostate cancer," *The Journal of Clinical Investigation*, vol. 123, no. 3, pp. 1109–1122, 2013.
- [20] D. Lefley, F. Howard, F. Arshad et al., "Development of clinically relevant in vivo metastasis models using human bone discs and breast cancer patient-derived xenografts," *Breast Cancer Research*, vol. 21, no. 1, 2019.
- [21] X. S. Peng, W. Guo, T. Liu et al., "Identification of miRs-143 and -145 that is associated with bone metastasis of prostate cancer and involved in the regulation of EMT," *PLoS One*, vol. 6, no. 5, 2011.
- [22] R. A. Irizarry, B. Hobbs, F. Collin et al., "Exploration, normalization, and summaries of high density oligonucleotide array probe level data," *Biostatistics*, vol. 4, no. 2, pp. 249–264, 2003.
- [23] M. E. Ritchie, B. Phipson, D. Wu et al., "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic acids research*, vol. 43, no. 7, p. e47, 2015.
- [24] G. Dennis, B. T. Sherman, D. A. Hosack et al., "DAVID: Database for Annotation, Visualization, And Integrated Discovery," *Genome Biology*, vol. 4, no. 5, p. P3, 2003.
- [25] D. Lai, L. Tan, X. Zuo et al., "Prognostic ferroptosis-related lncRNA signatures associated with immunotherapy and chemotherapy responses in patients with stomach cancer," *Frontiers in Genetics*, vol. 12, article 798612, 2021.
- [26] Gene Ontology Consortium, "The Gene Ontology (GO) project in 2006," *Nucleic Acids Research*, vol. 34, no. 90001, pp. D322–D326, 2006.
- [27] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, "KEGG: Kyoto Encyclopedia Of Genes And Genomes," *Nucleic Acids Research*, vol. 27, no. 1, pp. 29–34, 1999.
- [28] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [29] P. Curtin, H. Youm, and E. Salih, "Three-dimensional cancer-bone metastasis model using ex-vivo co-cultures of live calvarial bones and cancer cells," *Biomaterials*, vol. 33, no. 4, pp. 1065–1078, 2012.
- [30] R. Alfred, J. T. Taiani, R. J. Krawetz, A. Yamashita, D. E. Rancourt, and M. S. Kallos, "Large-scale production of murine embryonic stem cell-derived osteoblasts and chondrocytes on microcarriers in serum-free media," *Biomaterials*, vol. 32, no. 26, pp. 6006–6016, 2011.
- [31] Y. P. Zhao, W. L. Ye, D. Z. Liu et al., "Redox and pH dual sensitive bone targeting nanoparticles to treat breast cancer bone metastases and inhibit bone resorption," *Nanoscale*, vol. 9, no. 19, pp. 6264–6277, 2017.
- [32] K. Chen, D. Shen, L. Tan et al., "A pan-cancer analysis reveals the prognostic and immunotherapeutic value of ALKBH7," *Frontiers in Genetics*, vol. 13, p. 822261, 2022.
- [33] M. Sereno, J. Haskó, K. Molnár et al., "Downregulation of circulating miR 802-5p and miR 194-5p and upregulation of brain MEF2C along breast cancer brain metastasization," *Molecular Oncology*, vol. 14, no. 3, pp. 520–538, 2020.
- [34] L. He, J. M. Thomson, M. T. Hemann et al., "A microRNA polycistron as a potential human oncogene," *Nature*, vol. 435, no. 7043, pp. 828–833, 2005.
- [35] S. Costinean, N. Zaneni, Y. Pekarsky et al., "Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 18, pp. 7024–7029, 2006.
- [36] A. Esquela-Kerscher and F. J. Slack, "Oncomirs - microRNAs with a role in cancer," *Nature Reviews Cancer*, vol. 6, no. 4, pp. 259–269, 2006.
- [37] P. M. Voorhoeve, C. le Sage, M. Schrier et al., "A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors," *Cell*, vol. 124, no. 6, pp. 1169–1181, 2006.

- [38] G. A. Calin, C. D. Dumitru, M. Shimizu et al., “Frequent deletions and down-regulation of micro- RNA genes miR15- and miR16 at 13q14 in chronic lymphocytic leukemia,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 24, pp. 15524–15529, 2002.
- [39] N. Yanaihara, N. Caplen, E. Bowman et al., “Unique micro-RNA molecular profiles in lung cancer diagnosis and prognosis,” *Cancer Cell*, vol. 9, no. 3, pp. 189–198, 2006.
- [40] C. Welch, Y. Chen, and R. L. Stallings, “MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells,” *Oncogene*, vol. 26, no. 34, pp. 5017–5022, 2007.
- [41] L. C. Chen, W. Zhang, W. Yan et al., “The putative tumor suppressor miR-524-5p directly targets Jagged-1 and Hes-1 in glioma,” *Carcinogenesis*, vol. 33, no. 11, pp. 2276–2282, 2012.
- [42] H. Liu, X. Chen, T. Lin, X. Chen, J. Yan, and S. Jiang, “MicroRNA-524-5p suppresses the progression of papillary thyroid carcinoma cells via targeting on FOXE1 and ITGA3 in cell autophagy and cycling pathways,” *Journal of Cellular Physiology*, vol. 234, no. 10, pp. 18382–18391, 2019.
- [43] D. M. Dykxhoorn, “MicroRNAs and metastasis: little RNAs go a long way,” *Cancer Research*, vol. 70, no. 16, pp. 6401–6406, 2010.
- [44] C. Dong, X. Z. Yang, C. Y. Zhang et al., “Myocyte enhancer factor 2C and its directly-interacting proteins: a review,” *Progress in Biophysics and Molecular Biology*, vol. 126, pp. 22–30, 2017.
- [45] J. H. Ostrander, A. R. Daniel, K. Lofgren, C. G. Kleer, and C. A. Lange, “Breast tumor kinase (protein tyrosine kinase 6) regulates heregulin-induced activation of ERK5 and p38 MAP kinases in breast cancer cells,” *Cancer Research*, vol. 67, no. 9, pp. 4199–4209, 2007.
- [46] S. Takashima, S. Usui, O. Inoue et al., “Myocyte-specific enhancer factor 2c triggers transdifferentiation of adipose tissue-derived stromal cells into spontaneously beating cardiomyocyte-like cells,” *Scientific Reports*, vol. 11, no. 1, p. 1520, 2021.
- [47] Q. Tian, Y. Lu, B. Yan, and C. G. Wu, *Integrative Bioinformatics Analysis Reveals miR-524-5p/MEF2C Regulates Bone Metastasis in Prostate Cancer and Breast Cancer*, 2021.

Research Article

Identification and Verification of an Alternative Polyadenylation-Related lncRNA Prognostic Signature for Glioma

Hui Wang  and ZhiJun Jiang 

Department of Pathology, The First People's Hospital of Fuyang, Hangzhou City, Zhejiang Province 31400, China

Correspondence should be addressed to Hui Wang; dhtnjin@163.com

Received 20 May 2022; Revised 8 August 2022; Accepted 22 August 2022; Published 7 September 2022

Academic Editor: Chuan Lu

Copyright © 2022 Hui Wang and ZhiJun Jiang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the high mortality and modality of glioma, it was urgently needed to develop a glioma prognostic assessment system. Previous studies demonstrated that alternative polyadenylation- (APA-) related genes are important in immune response and oncogenesis. mRNA and lncRNA expression information of glioma samples were acquired from CGGA and TCGA databases, and lncRNAs associated with APA were selected through correlation analysis. The prognosis model of APA-related lncRNAs was built by the univariate Cox, random forest, and univariate Cox regression analyses. Glioma samples were assigned into high- and low-risk groups. Independence and effectiveness of the prognostic model were evaluated by Kaplan-Meier analysis, ROC curve, and Cox regression analyses. GO, KEGG enrichment, and GSEA analyses showed that the mainly involved signaling pathways were enriched in cellular immunity and immune signal transduction. We further analyzed expression differences of negative immune regulatory genes and immune cell infiltration degree between two groups. Immune checkpoints CTLA4 and LAG3 and immune suppressors TGFB, IL10, NOS3, and IDO1 and immune cell infiltration were notably upregulated in the high-risk group. The PD1/PDL1 expression was significantly correlated with risk score, showing that the prognostic model of APA-related lncRNA could effectively assess the tumor immune suppression. In conclusion, we established a risk assessment model of APA-related lncRNA in glioma, which could effectively evaluate prognosis of patients with glioma and tumor immune suppression and could provide guidance for clinical treatment.

1. Introduction

Glioma is a brain tumor produced by glial cells within the brain and spinal cord. It can be classified into astrocytomas, oligodendrogliomas, and ependymomas according to the morphological appearance [1]. Neurosurgical resection is a traditional treatment strategy, followed by concurrent radiation therapy and chemotherapy for a 4-8 weeks recovery phase [2-4], while the tumor recurrence near the resection area happens to the majority of patients during the recovery phase [5-7], which indicates a poor prognosis and survival rate in glioma patients.

Targeted therapy and immunotherapy of cancer cells are investigated to improve survival rate of glioma patients. Wang et al. found that anlotinib could effectively treat

patients with FGFR3-TACC3 fusion and recurrent glioblastoma (GBM) [8]. Other studies showed that IDO, CTLA-4, and PD-L1 are several immune checkpoints, which can negatively modulate T-cell activation [9, 10]. Inhibiting the immune checkpoints is able to significantly improve the efficacy of tumor therapy. Zhang et al. [11] discussed a set of promising targets for glioma treatment, particularly for IDH-wild-type GBM, which can provide support to the development of biomarkers for guiding cancer targeted therapies. Despite advanced development in glioma therapy, the overall survival rate of glioma patients was still at a relatively low level, and the prognosis was not ideal. For the clinical aspect, histopathology was often used to predict prognosis of glioma patients, but due to individual differences and complexity of glioma classification, patients with the same

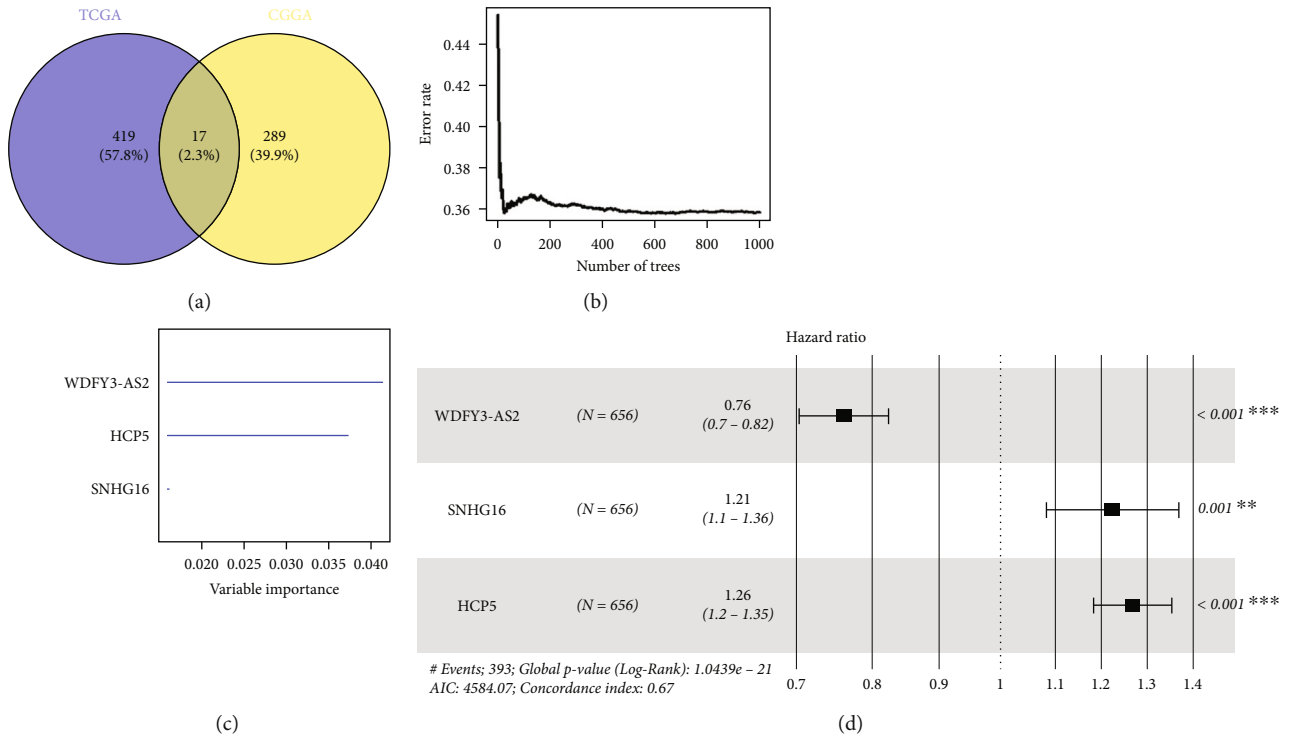


FIGURE 1: The construction of a prognostic model of APA-related lncRNAs. (a) The Venn diagram concerning the prognosis- and APA-related lncRNAs in TCGA-GBM_LGG and CGGA datasets. (b) The diagram of error rate and classification tree. (c) The importance ranking of the 3 feature lncRNAs. (d) Forest diagram of 3 feature lncRNAs, ** $p < 0.01$, *** $p < 0.001$.

pathology might experience different results [12, 13]. As a result, more biomarkers are needed for prognosis and treatment of glioma.

With the further analysis of clinical information and expression profiles in the public database by researchers, more and more biomarkers of glioma were excavated. For example, some research showed that the lncRNA HOXA11-AS was excavated on the basis of CGGA database mRNA microarray, which could be used as a biomarker for recognizing molecular subtypes of gliomas and a molecular target for glioma treatment [14]. Other research showed that ATP metabolism-related signatures constructed on the basis of the ATP metabolism-related gene expression could effectively evaluate the prognosis and immune microenvironment of glioma patients [15]. There is a great potential to develop prognostic risk models by using public databases; so, it is necessary to continue to excavate the genes associated with glioma prognosis. APA might control stability of mRNA, translation, and intracellular localization. The shortening of the APA structure at the end of mRNA has differential effects on oncogenes and tumor suppressors [16]. Recently, disorder of APA events in tumor cells was widely recognized as a driving factor of tumor genesis. For example, downregulation of APA-related CFLm25 gene in GBM significantly enhanced the size of tumors [17]. Increased CSTF2 facilitates cell progression in urothelial carcinoma of bladder (UCB) and could be a predictive indicator for unfavorable prognosis of UCB [18]. The enrichment of mRNA APA events in pancreatic ductal adenocarcinoma

was related to the expression of oncogenes and an independent indicator for prognostic prediction of pancreatic ductal adenocarcinoma [19]. As a result, combining the data analysis to the public databases and associating APA-related genes with the prognosis of glioma was of great importance for excavating potential glioma targets and establishing effective prognostic models.

Herein, a prognostic model was established on the basis of APA-related lncRNAs in glioma from public databases. Further analysis of the correlation between risk score and negative immune regulatory genes revealed potential connection between prognostic models and immune suppression of tumors. The feature lncRNAs selected in this study were expected to become potential biomarkers for glioma patients, and the prognostic model of APA-related lncRNAs could offer guidance for glioma immune treatment.

2. Materials and Methods

2.1. Data Source. In this study, 12 APA-related genes (CPSF1, CPSF2, CPSF3, CPSF4, CSTF1, CSTF2, CSTF3, CFI, PCF11, CLP1, NUDT21, PABPN1) were obtained from the latest literature [20]. The glioma transcriptome data and clinical information were downloaded from the Chinese Glioma Genome Atlas database (CGGA; <http://www.cgga.org.cn/>) as a training set. The glioma transcriptome data and clinical information (including 5 normal samples and 698 tumor samples) as a certification set (TCGA-GBM_LGG) were accessed from The Cancer Genome Atlas database

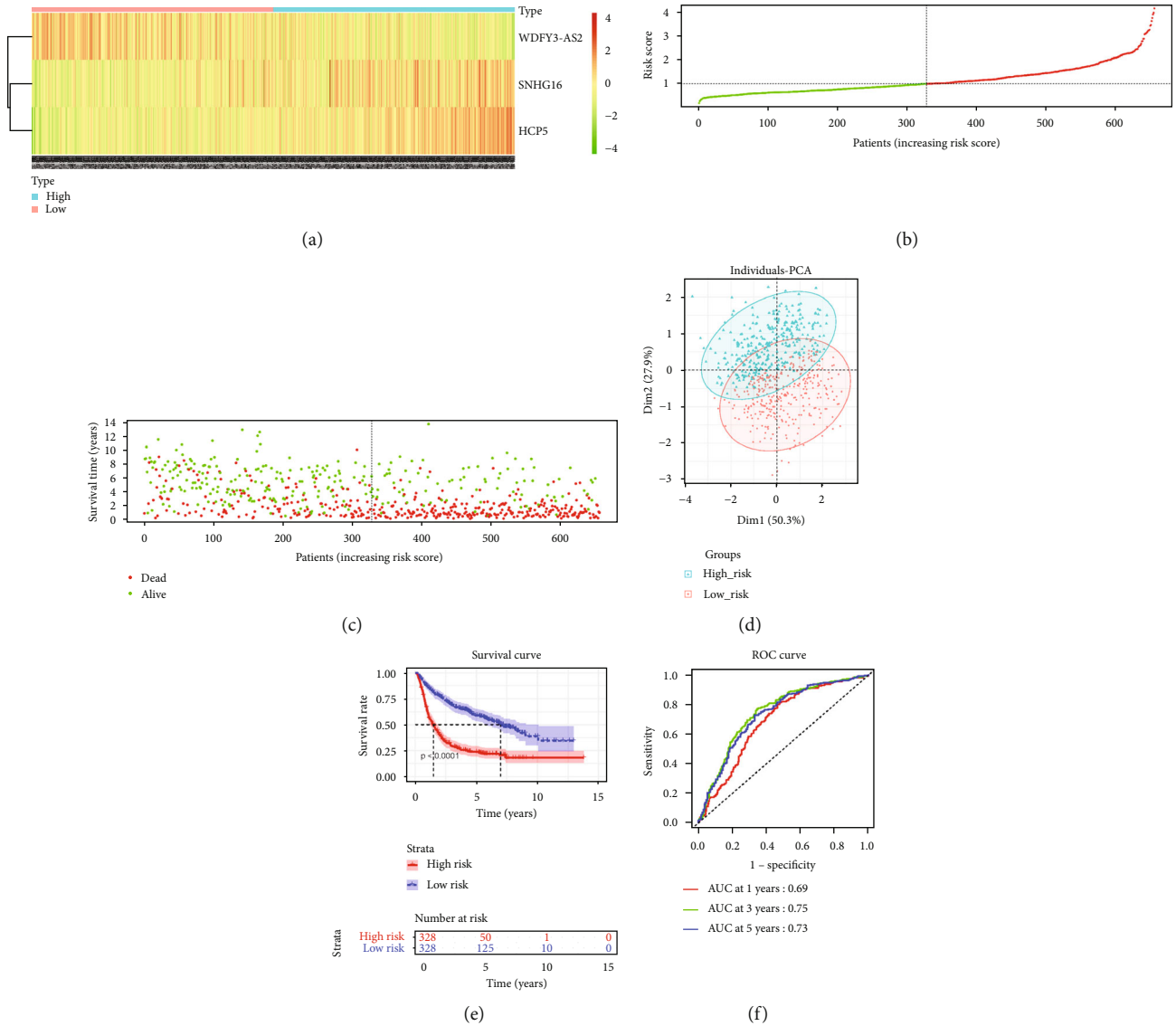


FIGURE 2: CGGA training collection evaluates the effectiveness of APA-related lncRNA prognostic models. (a) Heat map of expression levels of 3 lncRNAs in the high- and low-risk groups. (b) The risk score distribution map of patients with glioma, the green dots representing the samples in the low-risk group, and the red points representing the samples in the high-risk group. (c) Scatter plot of the survival status of patients with glioma. The green and red dots represent survival and death, respectively. (d) The PCA dimension reduction analysis is conducted on the samples of high- and low-risk groups. (e) *K-M* survival curve for high- and low-risk groups. (f) The ROC curve of the prognostic model to predict the overall survival rate of patients at 1, 3, and 5 years.

(TCGA; <https://portal.gdc.cancer.gov/>). The transcriptome data of the two datasets were annotated through Ensembl (<http://uswest.ensembl.org/index.html>), and finally, 4348 lncRNAs and 16,689 mRNAs were obtained from the CGGA dataset, and 8252 lncRNAs and 17,887 mRNAs were obtained from the TCGA-GBM_LGG dataset.

2.2. Screen for lncRNAs Related to APA. The lncRNA expression data were gained from TCGA-GBM_LGG and CGGA datasets, and the mRNA expression data of 12 known APA-related genes were standardized by log2. The lncRNAs and 12 APA-related genes in two datasets were analyzed by Pearson correlation analysis. The correlation coefficient

between the lncRNAs and APA-related genes in the datasets was determined, with $|r| > 0.5$ and $p < 0.001$ as the screening criteria to gain the lncRNAs that were related to APA.

2.3. The Establishment and Verification of the Prognostic Model. The samples that were followed up for less than 30 days were removed, and the R package survival (<https://CRAN.R-project.org/package=survival>) was utilized to perform univariate Cox regression analysis on APA-related lncRNAs in CGGA and TCGA-GBM_LGG datasets. $p < 0.001$ was used as the criteria to screen lncRNAs that were related to prognosis, and the screening results of the two datasets were overlapped to obtain lncRNAs related to

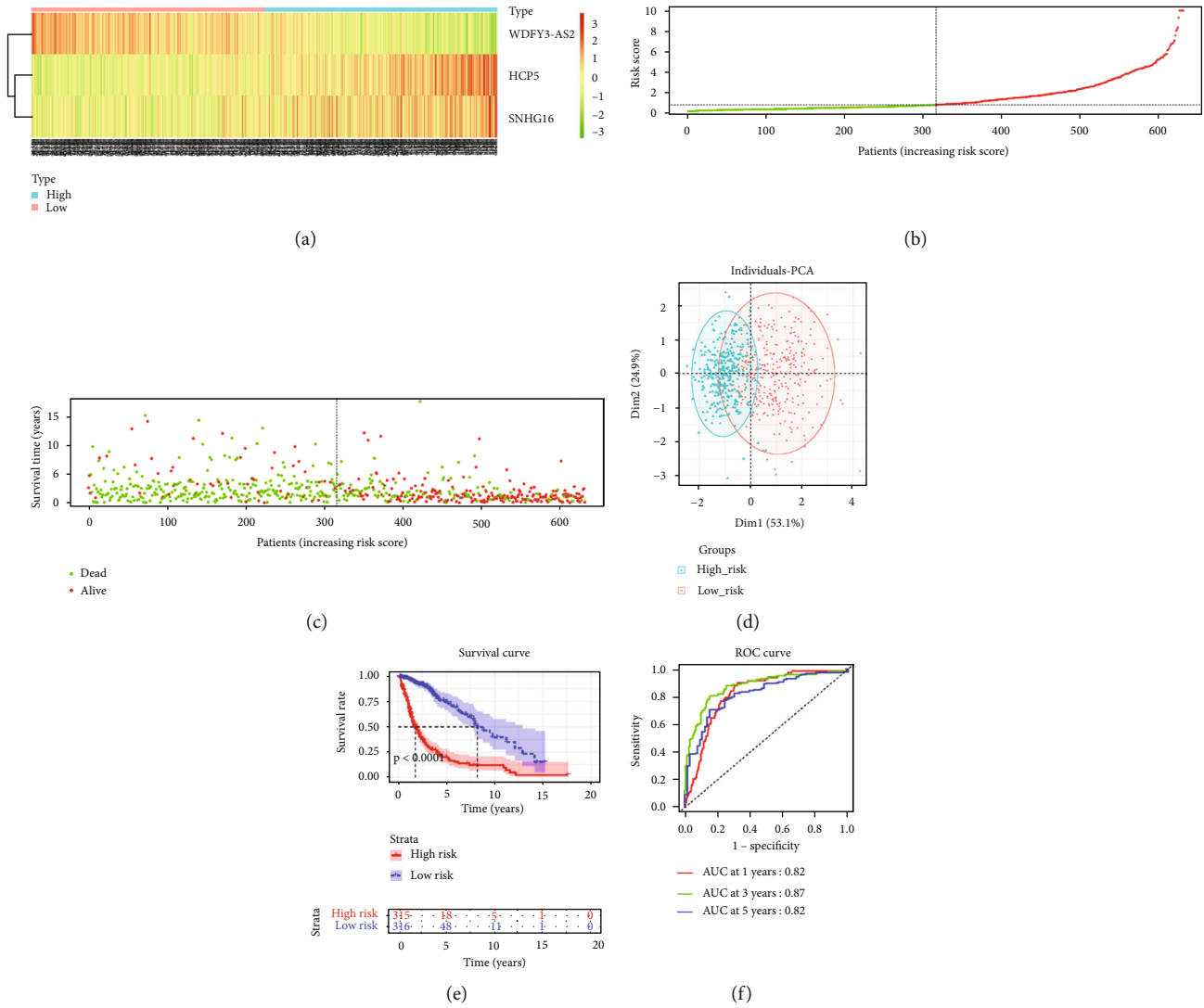


FIGURE 3: The TCGA-GBM_LGG certification set evaluates the effectiveness of the prognostic model of APA-related lncRNAs. (a) The heat map of the expression levels of 3 lncRNAs in the high- and low-risk groups. (b) The risk score distribution map of patients with glioma, the green representing the samples in the low-risk group, and the red points representing the samples in the high-risk group. (c) The scatter plot of the survival status of patients with glioma. The green and red dots refer to survival and death, respectively. (d) The PCA dimension reduction analysis of samples in the high- and low-risk groups. (e) The K-M survival curve for high- and low-risk groups. (f) The ROC curve of the prognostic model to predict the overall survival rate of patients at 1, 3, and 5 years.

APA and prognosis. To prevent the overfitting of the model, the R package randomForestSRC (<https://cran.r-project.org/web/packages/randomForestSRC/index.html>) was used to screen APA and prognosis-related lncRNAs in the CGGA dataset ($n_{tree} = 1000$, $n_{rep} = 20$) to obtain the candidate feature lncRNAs. R package survival was used to perform univariate Cox regression analysis on the selected candidate feature lncRNAs, and finally, the prognosis model of APA-related lncRNAs was obtained.

According to the expression level and risk parameter of each feature lncRNA, the risk score of the glioma samples in CGGA training set and TCGA-GBM_LGG certification set were calculated, and the samples were divided into two groups (high- and low-risk) according to the median of risk score. R package factoextra (<https://cloud.r-project.org/>

package=factoextra/) was used to perform principal component analysis (PCA) dimensionality reduction on these two groups in the training set on the basis of the expression of 3 lncRNAs. R package survival was utilized to portray the survival curve, and the difference in survival between the two groups was measured by log rank. R package timeROC [21] was used to draw ROC curve. The AUC value of the predicted 1-year, 3-year, and 5-year OS of glioma patients was calculated [21].

2.4. Functional Enrichment. The R package limma [22] was used to analyze the differential expression of the high- and low-risk groups of the CGGA dataset, and the differentially expressed genes of the high-risk group compared with the low-risk group were excavated under the condition of

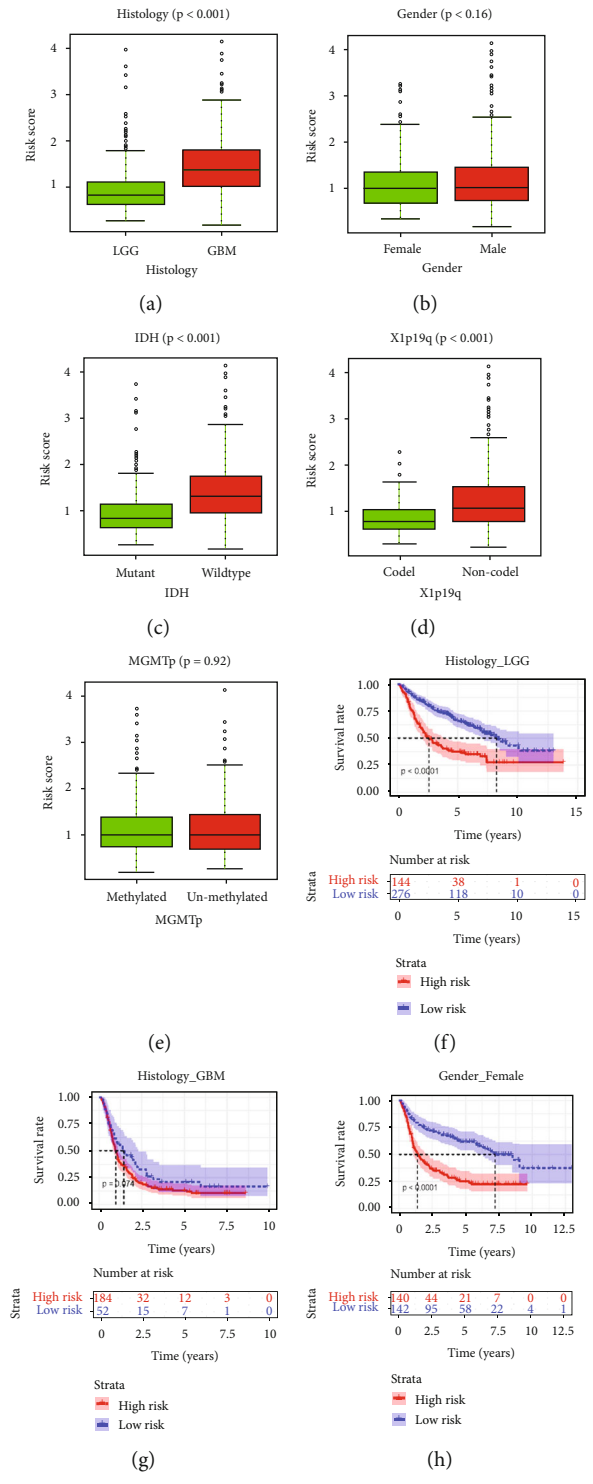


FIGURE 4: Continued.

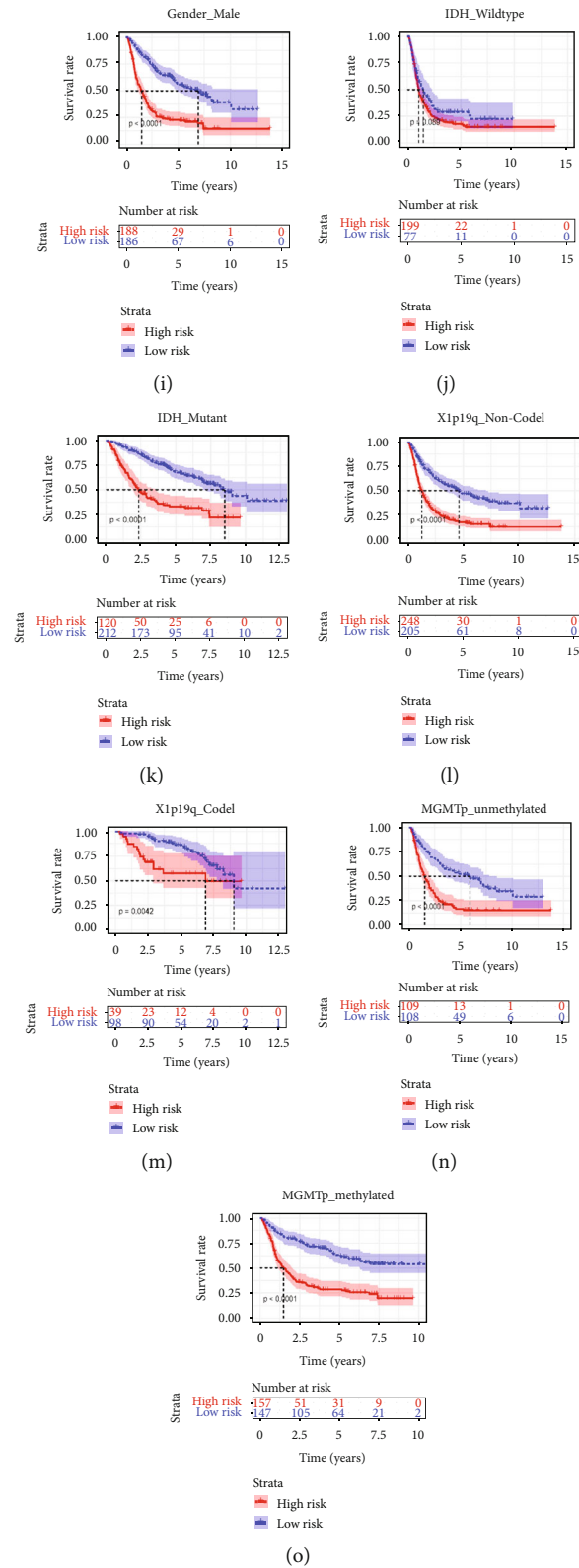


FIGURE 4: Correlation analysis between risk score and clinic pathologic features for patients with glioma. (a)–(e) Risk score box plots of clinical subgroups for various pathological features in the CGGA dataset. (f)–(o) The survival analysis of high- and low-risk groups in various clinical subgroups of the CGGA dataset.

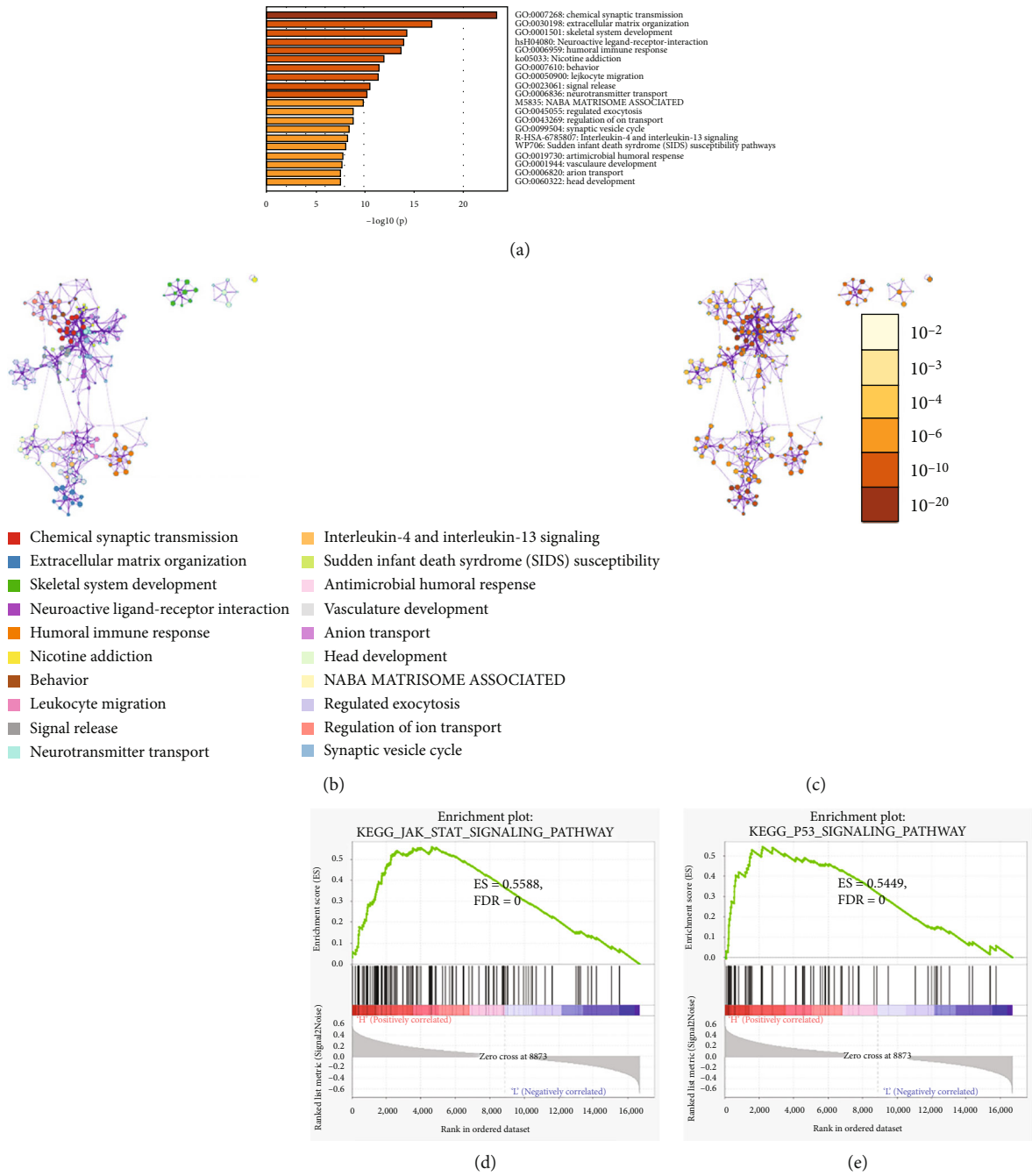
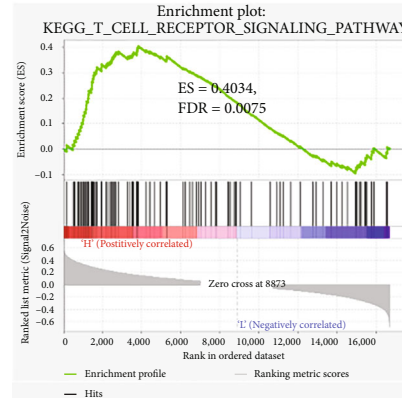


FIGURE 5: Continued.



(f)

FIGURE 5: Enrichment analysis of functions in the high- and low-risk groups in CGGA dataset. (a) The top 20 pathways and p value distribution of the biological process enriched with differentially expressed genes in high- and low-risk groups. (b) Enrichment term network of differentially expressed genes is analyzed. The color of the node shows the function or path cluster of the node. (c) p value of differentially expressed genes is clustered as a network. The color of the node indicates the degree of significance. (d)–(f) Differential activation pathways gained by GSEA enrichment in the high- and low-risk group. (d) JAK-STAT signaling pathway: ES = 0.5588, FDR = 0. (e) P53 signaling pathway: ES = 0.5449, FDR = 0. (f) T cell receptor signaling pathway: ES = 0.4034, FDR = 0.0075.

$\log_{FC} > 2$ and $p_{adj} < 0.05$ [22]. In order to investigate the biological functions and signaling pathways involved in the differentially expressed genes in the high-risk group, Metascape (<https://metascape.org/gp/index.html#/main/step1>) was used to conduct GO and KEGG enrichment analyses. GSEA software [23] was applied to analyze differentially activated pathways between the two groups.

2.5. Correlation Analysis between Prognostic Model and Tumor Immunity. It was known that tumor immune suppression was closely related to tumor progression. The excavation of immune suppressive factors would provide guidance to find suitable immune checkpoint inhibitors in immune therapies. In order to predict the correlation between the prognostic model and tumor immunity in this study, genes related to tumor immune suppression were acquired from the Tracking Tumor Immunophenotype database (TIP; <http://biocc.hrbmu.edu.cn/TIP/index.jsp>). They were intersected with the mRNA data of the CGGA brain glioma dataset, and the immune suppressive factor genes related to brain glioma were finally obtained. The expression of these immune suppressive factors in the high- and low-risk groups of glioma was analyzed. The expression differences of immune checkpoint and immune suppression-related genes between the two groups were measured by R package anova (<https://www.scribbr.com/statistics/anova-in-r>). The ssGSEA algorithm was used to explore the immune infiltration of cells in the high- and low-risk groups of glioma by R package GSEA [24].

2.6. Evaluation of the Effectiveness of the Prognostic Model. The univariate and multivariate Cox regression analyses were conducted on risk score, gender, histology, IDH_mutation, X1p19q_deletion, and MGMTp_methy in the CGGA dataset, and nomogram was generated by using R package rms [25]. The calibration curve was plotted to measure the

ability of the nomogram to predict the survival of patients with glioma.

2.7. The Statistical Analyses. All statistical analyses were performed by using R software (<https://www.r-project.org/>). The t -test was used to compare the significance between two groups. The log-rank test was introduced to test the significance for K - M analysis, and omnibus test was used in Cox regression analysis.

3. Results

3.1. Construction of a Prognostic Model of APA-Related lncRNAs. The mRNA expression of 12 APA-related genes was obtained from the CGGA and TCGA-GBM_LGG datasets, and Pearson correlation analysis was performed on the lncRNAs in the corresponding datasets. Finally, 1118 APA-related lncRNAs (Table S1) were obtained from the CGGA dataset. 500 APA-related lncRNAs were obtained from TCGA-GBM_LGG dataset (Table S2). Univariate Cox regression analysis was performed, 306 lncRNAs that were related to the prognosis were obtained from the CGGA dataset (Table S3), and 436 prognosis-related lncRNAs were obtained from TCGA-GBM_LGG dataset (Table S4). The results of the two datasets were intersected to identify 17 lncRNAs related to APA and prognosis (Figure 1(a)). In the CGGA dataset, random forest was used for feature selection of these 17 lncRNAs. In the random forest algorithm, 1000 decision trees were constructed, followed by ranking the diverse models in order of the number of decision trees involved (Figure 1(b)). Considering the relationship between the error rate and the number of trees, 3 lncRNAs were screened based on variable importance. The importance of the prognostic effect of 3 feature lncRNAs in glioma patients ranged from high to low: WDFY3-AS2, HCP5, and SNHG16 (Figure 1(c)). Afterwards, in the CGGA dataset, univariate Cox

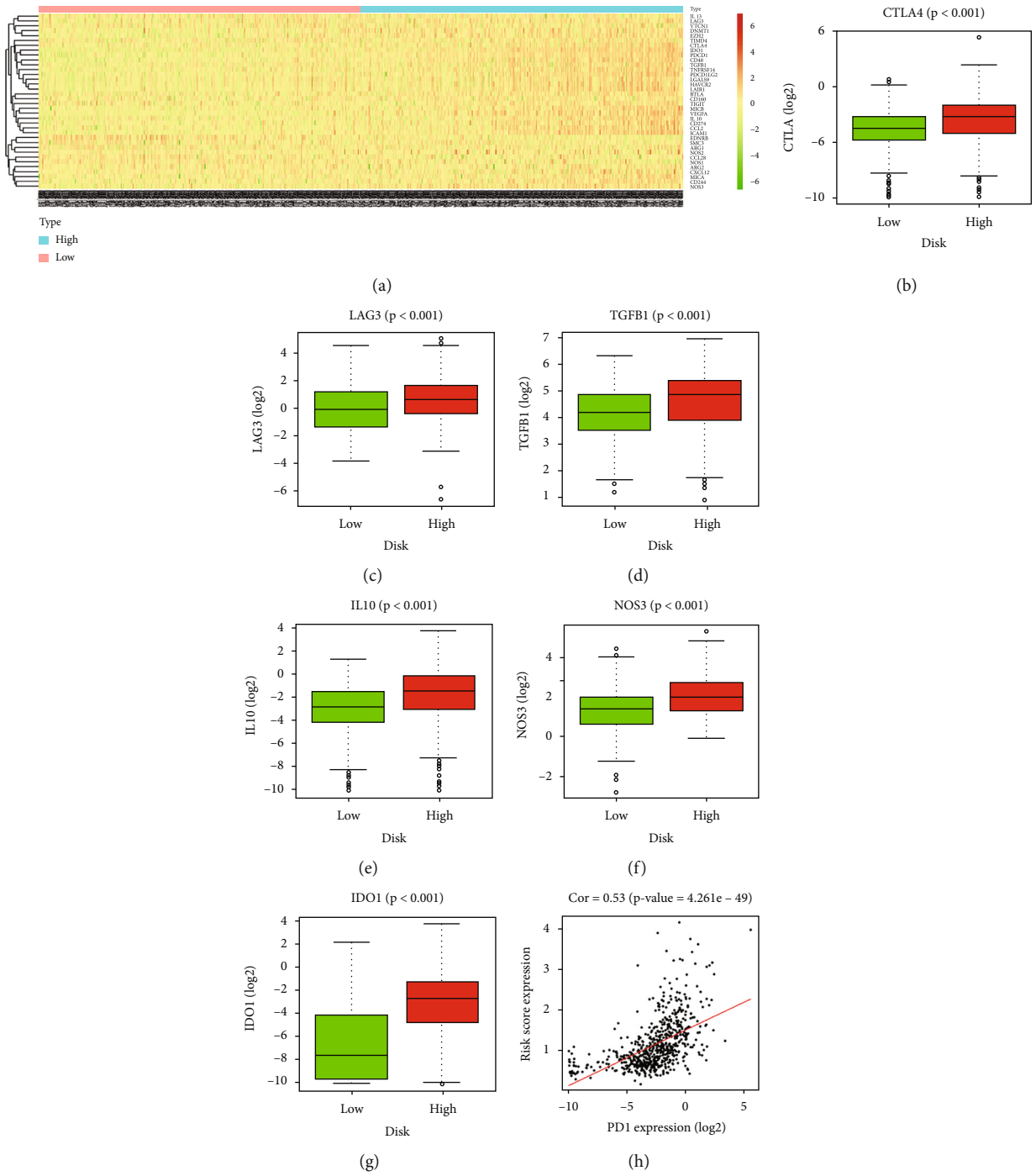


FIGURE 6: Continued.

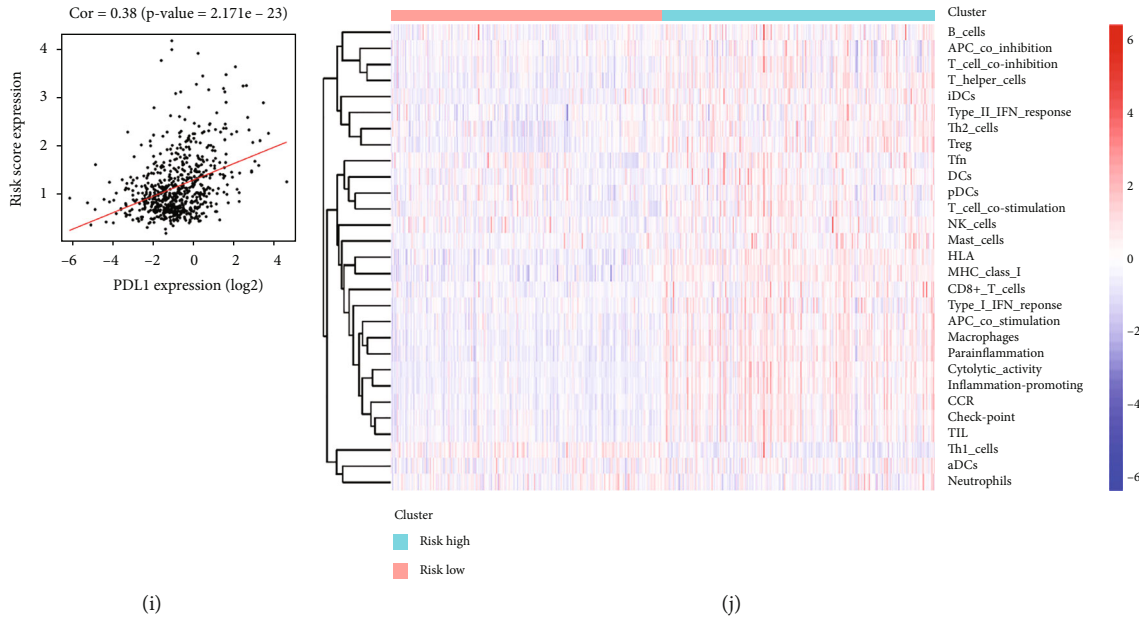


FIGURE 6: The correlation between prognosis model of APA-related lncRNAs and tumor immune suppression. (a) Heat map of negative immune regulatory gene expression in high- and low-risk groups. (b, c) The differences in the expression levels of CTLA4 (b) and LAG3 (c) in samples of high- and low-risk groups. (d)–(g) The differences in the expression levels of immune suppressive cytokines TGFβ1 (d), IL10 (e), NOS3 (f), and IDO1 (g) in the high- and low-risk groups. (h, i) The correlation diagram between the risk score and PD1 (h) as well as PDL1 (i) concerning CGGA data collection samples. (j) Heat map of immune cell infiltration levels in high- and low-risk groups.

regression analysis was done on screened 3 lncRNAs, and all 3 lncRNAs were identified to be independent prognostic indicators for glioma (Figure 1(d)). An APA-related lncRNA prognostic model was finally obtained: $risk\ score = -0.272 * WDFY3 - AS2 + 0.195 * SNHG16 + 0.231 * HCP5$.

3.2. APA-Related lncRNA Prognostic Model Evaluation. In the CGGA training set, levels of 3 feature lncRNAs were remarkably different between high- and low-risk groups. The WDFY3-AS2 expression in the high-risk group was substantially downregulated, while SNHG16 and HCP5 levels in the high-risk group were notably upregulated (Figure 2(a)). Combining with risk score distribution and survival time of the high- and low-risk groups in the CGGA dataset, we noticed that, as the risk score increased, the number of deaths of patients increased and survival time declined (Figures 2(b) and 2(c)). According to the expression of the three feature lncRNAs, PCA dimensionality reduction was conducted on samples in the high- and low-risk groups, and samples in the two groups could be distinctly distinguished (Figure 2(d)). In addition, the survival curves of the CGGA dataset presented that patients with high-risk scores had a worse prognosis (Figure 2(e)). The results in TCGA-GBM_LGG certification collection were reflected in the same way (Figures 3(a)–3(e)). Finally, the ROC curve illustrated that the risk assessment model predicted the AUC values of the 1-year, 3-year, and 5-year survival time of samples in the CGGA dataset, which were 0.69, 0.75, and 0.73, respectively (Figure 2(f)). The AUC values pre-

dicted by the model for 1-year, 3-year and 5-year survival time of the samples in the TCGA-GBM_LGG dataset were 0.82, 0.87, and 0.82, respectively (Figure 3(f)). Thus, the risk assessment model on the basis of three features could predict prognosis of glioma patients.

3.3. Correlation Analysis between the Risk Score and Clinical Data. To investigate the relationship between risk score of the prognostic model and clinical characteristics, samples in the CGGA dataset were classified by clinical factors and pathological characteristics to gain different clinical subgroups (Figures 4(a)–4(e)). GBM is the most malignant form of astrocytoma [26]. The analysis results presented that risk score of GBM group was notably higher than that of the LGG group (Figure 4(a)), while that in the male group was slightly higher than that in the female group (Figure 4(b)). Analysis results showed that the risk score of the IDH_Wild type group was clearly higher than that of the IDH_Mutant group (Figure 4(c)), and the risk score of the X1P19Q_Non-Codel group was clearly higher than that of the X1p19q_Codel group (Figure 4(d)). These results showed that mutations in isocitrate dehydrogenase and deletion of chromosomal X1p19q were closely associated with risk score, and that the risk score of the prognostic model could be utilized to evaluate pathological features illustrated above.

After the samples in each clinical subgroup were assigned into high- and low-risk groups according to the risk score, the results of KM survival analysis showed that, except for the prognostic survival rate of patients in the two groups of Histology_GBM and IDH_Wild type, the other 8 clinical

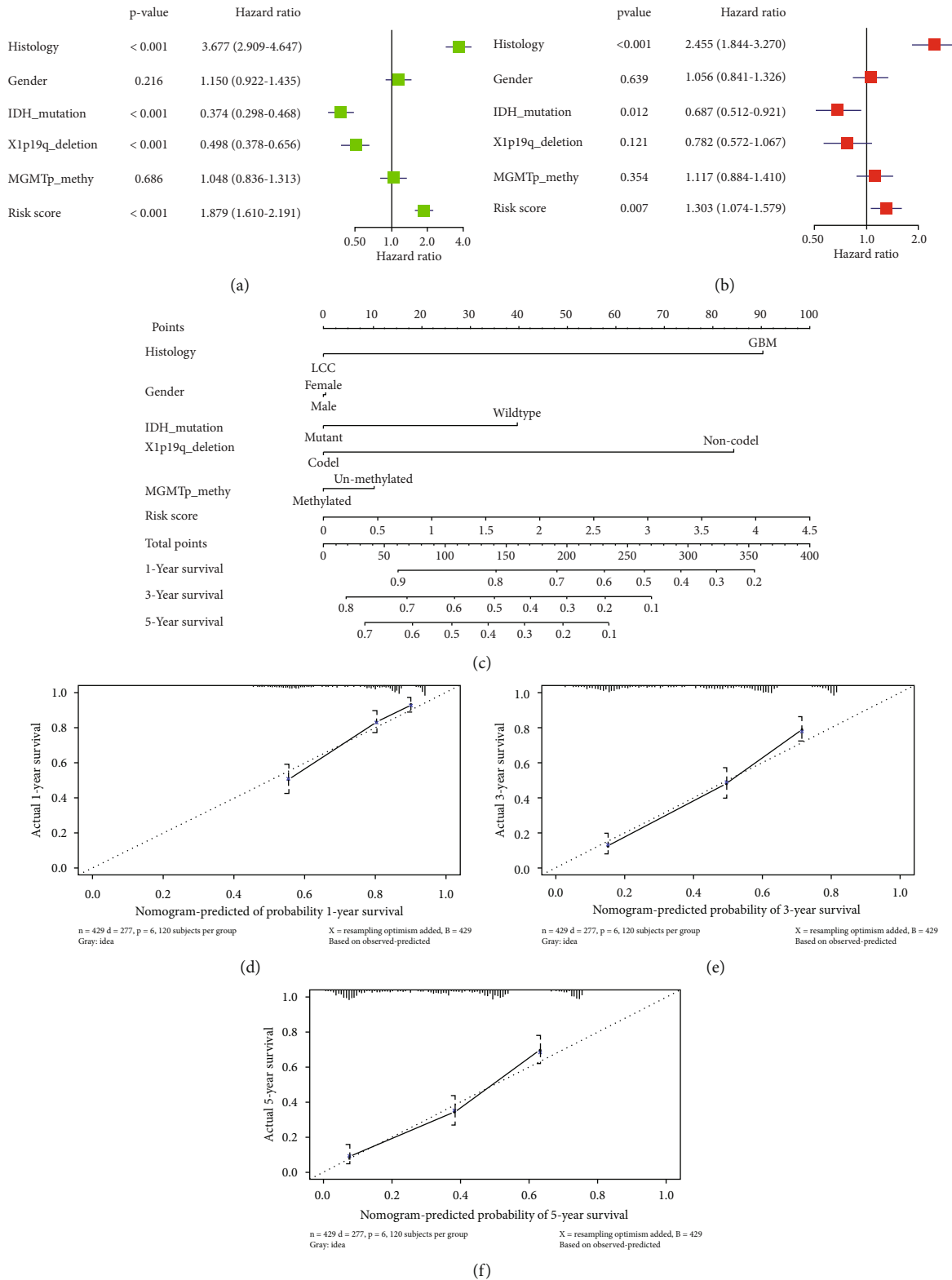


FIGURE 7: Construction and assessment of the nomogram. (a) Forest plot that combines the univariate Cox analysis of risk score and other pathological features. (b) Multivariate Cox analysis forest plot combining risk score and other pathological features. (c) A nomogram that combines the risk score of three feature lncRNAs and other clinical factors and pathological characteristics to predict the overall survival rate of patients diagnosed with glioma for 1, 3, and 5 years. (d)–(f) Calibration curve of the nomogram predicting the survival rate of patients at 1 year (d), 3 years (e), and 5 years (f).

subgroups had worse prognosis performance than the low-risk group (Figures 4(f)–4(o)). These results showed that the prognostic model could be utilized to measure survival rate of patients with different pathological characteristics.

3.4. The GSEA Enrichment Analysis. Intergroup analysis was conducted on samples in the high- and low-risk groups of the CGGA dataset, and 335 differential expression genes in the high-risk group were obtained (Table S5). GO biological function enrichment, as well as KEGG analysis, was conducted on these genes, and results showed that these genes were mainly enriched in humoral immune response, leukocyte migration, interleukin-4 and interleukin-13 signaling, and biological functions and pathways related to the immune signal transduction (Figures 5(a)–5(c)). The GSEA enrichment analysis results exhibited that the two groups had significant differences in the activation of JAK-STAT signaling pathway, P53 signaling pathway, and T cell receptor signaling pathway (Figures 5(d)–5(f)). The differentially expressed genes were principally enriched in biological functions and pathways including activation, migration of immune cells, and immune signal transduction. These functions and pathways could be one of the main reasons contributing to the different prognosis in the high- and low-risk groups.

3.5. Correlation between Prognostic Model and Immune Suppression in Glioma. To investigate whether the prognostic model of APA-related lncRNAs could measure immune suppression in glioma, this study visualized the expression of immune suppressive genes in the high- and low-risk groups of the CGGA dataset. The result displayed that majority of the negative immune regulatory genes glioma were upregulated in the high-risk group (Figure 6(a)), and the expression of immune checkpoint CTLA4 and immune checkpoint receptor LAG3 was notably higher in the high-risk group (Figures 6(b) and 6(c)). Immune suppressive cytokines TGFB, IL10, NOS3, and IDO1 in the high-risk group were also notably upregulated (Figures 6(d)–6(g)). Pearson correlation analysis was carried out on the typical immune suppressive molecules PD1/PDL1 and the prognostic model. The results showed that risk score had a positive correlation with the expression of PD1 and PDL1 (Figures 6(h) and 6(i)). Also, ssGSEA was employed between high- and low- risk groups, where the results indicated that the infiltration of the immune cells performed a higher level in the high-risk group (Figure 6(j)). Interestingly, it seemed to conflict with the findings in Figures 6(h)–6(j), which showed a positive relationship between risk score and immunity suppression. The results above indicated that the 3-lncRNA prognostic model could be used to examine the expression levels of immune checkpoints including PD1 and PDL-1, revealing the tumor immunity status to some extent.

3.6. The Construction and Evaluation of Nomogram. Univariate Cox analysis of risk score and other pathological characteristics were performed in the CGGA dataset. The results presented that both histology and risk score had a substan-

tial impact on prognosis. Among them, HR of risk score was 1.879 (p value=1.03E-15) (Figure 7(a)). The analysis results of multifactors presented that the HR of risk score was 1.303 (p value=0.007165) (Figure 7(b)), indicating that risk score could be used as a prognostic factor independent of clinical features. The nomogram drawn by combination of risk score, gender, histology, IDH_mutation, X1p19q_deletion, and MGMTp_methy was used to predict overall survival rate of patients with glioma at 1, 3, and 5 years (Figure 7(c)), and the corresponding calibration curve had relatively good fitting degree (Figures 7(d)–7(f)), which proved that the nomogram had relatively good predictive ability.

4. Discussion

RNA modifications, including 3'-terminal APA, editing of A-I, N1-methyladenosine modification (m1A), and N6-methyladenosine modification (m6A), are demonstrated to be relevant to tumor genesis and immune modulation. The studies of Zhang et al. demonstrated that the decrease of m6A RNA methylation would activate the oncogenic Wnt/PI3K-Akt pathway and thus lead to deterioration of stomach cancer [27]. Zhao et al. discovered that m1A regulatory gene in digestive tract cancer was relevant to the occurrence of cancer by modulating ErbB2 and mTOR [28]. Other studies revealed that APA was relevant to NSCLC, and cleavage stimulation factor subunit 2 (CSTF2) may regulate 3'UTR length to serve as an oncogene in NSCLC [29]. Therefore, by conducting bioinformatic analyses on APA-related lncRNAs, a risk assessment model composed of 3 feature lncRNAs was finally developed. It was found by the stratified analysis of clinicopathological characteristics that the model was able to predict prognosis of patients with different pathological features. Functional enrichment and immune suppressor analysis manifested that the 3 feature lncRNAs could effectively evaluate the suppressive microenvironment of tumor immunity. Finally, Cox regression analysis of risk score and other pathological characteristics was conducted from perspective of single and multiple factors, and it was found that risk score could be utilized as a prognostic factor deemed to be independent and effective.

More and more researches show that lncRNA is an important prognostic biomarker and epigenetic regulator, playing a crucial role in progression of cancer [30–32]. In this study, three APA-related feature lncRNAs (WDFY3-AS2, HCP5, SNHG16) were dug up. WDFY3-AS2 was a prognostic factor beneficial for glioma, while remaining lncRNAs were regarded as unfavorable prognostic factors. These three lncRNAs were confirmed to be correlated with cancer development. Wu et al. found that WDFY3-AS2 could be an independent prognostic factor for glioma, and its expression was involved in synaptic transmission, glutamate receptor, and tumor necrosis factor (TNF) signaling [33]. Kong et al. pointed out that the low WDFY3-AS2 expression in the samples of esophageal tumor predicted an unfavorable prognosis, and the reexpression of this gene could inhibit the development of the esophageal tumor by

inhibiting proliferation and migration [34]. However, in ovarian cancer, the function of WDFY3-AS2 was contrary to our results. Its expression was increased in tumor tissues and silencing of which could dramatically suppress progression of ovarian cancer cells [35]. So, we inferred that this gene might play different roles in different types of tumors. The HCP5 gene was principally expressed in immune system cells [36], which affects susceptibility to hepatitis C virus-associated hepatocellular carcinoma [37]. HCP5 could act as a tumor suppressor by modulating malignant phenotypes of glioma cells through binding to microRNA-139. Inhibiting the expression of HCP5 could induce the cell apoptosis and reduce the proliferation, migration, and invasion of glioma cells [38]. SNHG16, as a potential oncogenic factor in many cancers, was usually associated with the modulation of miRNA to influence the occurrence and development of tumor cells. For example, SNHG16 represses miR-216-5p level in the samples of the cervical cancer [39]. SNHG16 could inhibit the expression of miR-542-3p and upregulate the autophagy related gene 5 (ATG5) to foster proliferation, migration, invasion, and autophagy of neuroblastoma [40]. Therefore, the three distinctive lncRNAs found in this study were related to progression of tumors and might be potential biomarkers of glioma.

This study investigated the relationship between immune suppression and prognosis model of three distinctive lncRNAs and revealed that several immune-associated genes (CTLA4, LAG3, TGFB, IL10, NOS3, and IDO1) were remarkably upregulated in the high-risk group, and that PDL1/PD1 had positive correlations with risk score. Among them, CTLA4, PD1, and PDL1 were all typical immune checkpoint molecules, closely related to tumor immune suppression [41–43]. Studies showed that CTLA4 could regulate activation of T cells, and the high expression of this gene was notably associated with high levels of tumor-infiltrating lymphocytes. Also, CTLA4 can serve as a prognostic biomarker for clear cell renal cell carcinoma [44]. The mechanisms that influence infiltration and activation of immune cells in the immune microenvironment may vary from patient to patient. Referring to our findings in Figure 6(h)–6(j), the PD1/PDL1 expression showed positive associations with the risk score, while the risk score positively associated with immune cell infiltration. It was assumed as T-cell exhaustion [45], characterized as a significant upregulation of immune checkpoints and a minor change in cell infiltration. PD-1 was a membrane protein expressed by immune cells. Blocking PD-1 signal could enhance the anticancer effect of T cells [46]. Liu and other researchers found that PD-1/CTLA4 levels were relevant to the progression of various cancers and infiltration level of immune cells, and their expression levels were notably relevant to the survival rate of patients diagnosed with different cancers [42]. Thus, the prognostic model of APA-related lncRNAs could further predict immune suppression of tumor tissues by measuring the expression level of immune suppressive factors.

To sum up, the prognostic model of APA-related lncRNAs established in this research could be an alternative independent prognostic factor for glioma. The three feature lncRNAs had the potential to be prognostic predictors and

therapeutic targets for glioma patients. For clinical aspect, the prognostic model of APA-related lncRNAs was suitable for samples of different pathological types and could be used to evaluate the expression of immune suppression factors such as CTLA4, PD-1, and IL1, thus predicting the immune suppression of tumor tissues. In conclusion, this prognostic feature could provide guidance for clinical treatment of glioma patients. However, certain limitations existed in this study. The prognostic model was gained on the basis of bioinformatic analysis of retrospective data; so, further clinical data are needed to support the results of this study.

Data Availability

The datasets generated and analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors report no conflict of interest.

Authors' Contributions

Hui Wang contributed to the conceptualization, methodology, data curation, and writing-original draft preparation, reviewing, and editing. ZhiJun Jiang contributed to the visualization, investigation, supervision, software, and validation.

Acknowledgments

This study was supported in part by grants from the Hangzhou Warehousing Science and Technology Program (20181228Y144) and Zhejiang Medical and Health Science and Technology Plan (2020PY022).

Supplementary Materials

Supplementary 1. Table S1: screening APA-related lncRNAs from the CGGA dataset based on Pearson analysis.

Supplementary 2. Table S2: screening APA-related lncRNAs from the TCGA dataset based on Pearson analysis.

Supplementary 3. Table S3: univariate Cox regression analysis of prognosis-related lncRNAs from the CGGA dataset.

Supplementary 4. Table S4: univariate Cox regression analysis of prognosis-related lncRNAs from the TCGA dataset.

Supplementary 5. Table S5: differentially expressed genes in the high- and low-risk groups of the CGGA dataset.

References

- [1] D. N. Louis, H. Ohgaki, O. D. Wiestler et al., "The 2007 WHO classification of tumours of the central nervous system," *Acta Neuropathologica*, vol. 114, no. 2, pp. 97–109, 2007.
- [2] R. Stupp, W. P. Mason, M. J. van den Bent et al., "Radiotherapy plus concomitant and adjuvant temozolomide for

- glioblastoma,” *The New England Journal of Medicine*, vol. 352, no. 10, pp. 987–996, 2005.
- [3] U. Linz, “Commentary on effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-Year analysis of the EORTC-NCIC trial (Lancet Oncol. 2009;10:459-466),” *Cancer*, vol. 116, no. 8, pp. 1844–1846, 2010.
 - [4] R. Stupp, S. Taillibert, A. Kanner et al., “Effect of tumor-treating fields plus maintenance Temozolomide vs maintenance Temozolomide alone on survival in patients with glioblastoma: a randomized clinical trial,” *JAMA*, vol. 318, no. 23, pp. 2306–2316, 2017.
 - [5] A. Pirzkall, C. McGue, S. Saraswathy et al., “Tumor regrowth between surgery and initiation of adjuvant therapy in patients with newly diagnosed glioblastoma,” *Neuro-Oncology*, vol. 11, no. 6, pp. 842–852, 2009.
 - [6] C. Pennington, L. Kilbride, R. Grant, and J. M. Wardlaw, “A pilot study of brain tumour growth between radiotherapy planning and delivery,” *Clinical Oncology (Royal College of Radiologists)*, vol. 18, no. 2, pp. 104–108, 2006.
 - [7] J. D. Waters, B. Rose, D. D. Gonda et al., “Immediate post-operative brachytherapy prior to irradiation and temozolomide for newly diagnosed glioblastoma,” *Journal of Neuro-Oncology*, vol. 113, no. 3, pp. 467–477, 2013.
 - [8] Y. Wang, D. Liang, J. Chen et al., “Targeted therapy with anlotinib for a patient with an oncogenic FGFR3-TACC3 fusion and recurrent glioblastoma,” *The Oncologist*, vol. 26, no. 3, pp. 173–177, 2021.
 - [9] D. A. Wainwright, A. L. Chang, M. Dey et al., “Durable therapeutic efficacy utilizing combinatorial blockade against IDO, CTLA-4, and PD-L1 in mice with brain tumors,” *Clinical Cancer Research*, vol. 20, no. 20, pp. 5290–5301, 2014.
 - [10] H. Zhang, Z. Dai, W. Wu et al., “Regulatory mechanisms of immune checkpoints PD-L1 and CTLA-4 in cancer,” *Journal of Experimental & Clinical Cancer Research*, vol. 40, no. 1, p. 184, 2021.
 - [11] K. Yang, Z. Wu, H. Zhang et al., “Glioma targeted therapy: insight into future of molecular approaches,” *Molecular Cancer*, vol. 21, no. 1, p. 39, 2022.
 - [12] S. W. Coons, P. C. Johnson, B. W. Scheithauer, A. J. Yates, and D. K. Pearl, “Improving diagnostic accuracy and interobserver concordance in the classification and grading of primary gliomas,” *Cancer*, vol. 79, no. 7, pp. 1381–1393, 1997.
 - [13] D. P. Cahill, A. E. Sloan, B. V. Nahed et al., “The role of neuropathology in the management of patients with diffuse low grade glioma: a systematic review and evidence-based clinical practice guideline,” *Journal of Neuro-Oncology*, vol. 125, no. 3, pp. 531–549, 2015.
 - [14] Q. Wang, J. Zhang, Y. Liu et al., “A novel cell cycle-associated lncRNA, HOXA11-AS, is transcribed from the 5-prime end of the HOXA transcript and is a biomarker of progression in glioma,” *Cancer Letters*, vol. 373, no. 2, pp. 251–259, 2016.
 - [15] R. Huang, G. Li, Z. Wang et al., “Identification of an ATP metabolism-related signature associated with prognosis and immune microenvironment in gliomas,” *Cancer Science*, vol. 111, no. 7, pp. 2325–2335, 2020.
 - [16] C. P. Masamha and E. J. Wagner, “The contribution of alternative polyadenylation to the cancer phenotype,” *Carcinogenesis*, vol. 39, no. 1, pp. 2–10, 2018.
 - [17] C. P. Masamha, Z. Xia, J. Yang et al., “CFIm25 links alternative polyadenylation to glioblastoma tumour suppression,” *Nature*, vol. 510, no. 7505, pp. 412–416, 2014.
 - [18] X. Chen, J. X. Zhang, J. H. Luo et al., “CSTF2-induced shortening of the RAC1 3’UTR promotes the pathogenesis of urothelial carcinoma of the bladder,” *Cancer Research*, vol. 78, no. 20, pp. 5848–5862, 2018.
 - [19] S. Venkat, A. A. Tisdale, J. R. Schwarz et al., “Alternative polyadenylation drives oncogenic gene expression in pancreatic ductal adenocarcinoma,” *Genome Research*, vol. 30, no. 3, pp. 347–360, 2020.
 - [20] H. Chen, J. Yao, R. Bao et al., “Cross-talk of four types of RNA modification writers defines tumor microenvironment and pharmacogenomic landscape in colorectal cancer,” *Molecular Cancer*, vol. 20, no. 1, p. 29, 2021.
 - [21] P. Blanche, J. F. Dartigues, and H. Jacqmin-Gadda, “Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks,” *Statistics in Medicine*, vol. 32, no. 30, pp. 5381–5397, 2013.
 - [22] M. E. Ritchie, B. Phipson, D. Wu et al., “Limma powers differential expression analyses for RNA-sequencing and microarray studies,” *Nucleic Acids Research*, vol. 43, no. 7, article e47, 2015.
 - [23] A. Subramanian, P. Tamayo, V. K. Mootha et al., “Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.
 - [24] S. Hanzelmann, R. Castelo, and J. Guinney, “GSVA: gene set variation analysis for microarray and RNA-seq data,” *BMC Bioinformatics*, vol. 14, no. 1, p. 7, 2013.
 - [25] C. Huang, Z. Liu, L. Xiao et al., “Clinical significance of serum CA125, CA19-9, CA72-4, and fibrinogen-to-lymphocyte ratio in gastric cancer with peritoneal dissemination,” *Frontiers in Oncology*, vol. 9, p. 1159, 2019.
 - [26] R. Chen, M. Smith-Cohn, A. L. Cohen, and H. Colman, “Glioma subclassifications and their clinical significance,” *Neurotherapeutics*, vol. 14, no. 2, pp. 284–297, 2017.
 - [27] C. Zhang, M. Zhang, S. Ge et al., “Reduced m6A modification predicts malignant phenotypes and augmented Wnt/PI3K-Akt signaling in gastric cancer,” *Cancer Medicine*, vol. 8, no. 10, pp. 4766–4781, 2019.
 - [28] Y. Zhao, Q. Zhao, P. J. Kaboli et al., “m1A regulated genes modulate PI3K/AKT/mTOR and ErbB pathways in gastrointestinal cancer,” *Translational Oncology*, vol. 12, no. 10, pp. 1323–1333, 2019.
 - [29] S. Zhang, X. Zhang, W. Lei et al., “Genome-wide profiling reveals alternative polyadenylation of mRNA in human non-small cell lung cancer,” *Journal of Translational Medicine*, vol. 17, no. 1, p. 257, 2019.
 - [30] X. D. Zhang, G. W. Huang, Y. H. Xie et al., “The interaction of lncRNA EZR-AS1 with SMYD3 maintains overexpression of EZR in ESCC cells,” *Nucleic Acids Research*, vol. 46, no. 4, pp. 1793–1809, 2018.
 - [31] W. Li, M. Sun, C. Zang et al., “Upregulated long non-coding RNA AGAP2-AS1 represses LATS2 and KLF2 expression through interacting with EZH2 and LSD1 in non-small-cell lung cancer cells,” *Cell Death & Disease*, vol. 7, no. 5, article e2225, 2016.
 - [32] Y. Wang, W. Zhang, Y. Wang, and S. Wang, “HOXD-AS1 promotes cell proliferation, migration and invasion through

- miR-608/FZD4 axis in ovarian cancer,” *American Journal of Cancer Research*, vol. 8, pp. 170–182, 2018.
- [33] F. Wu, Z. Zhao, R. Chai et al., “Expression profile analysis of antisense long non-coding RNA identifies WDFY3-AS2 as a prognostic biomarker in diffuse glioma,” *Cancer Cell International*, vol. 18, no. 1, p. 107, 2018.
- [34] Q. Kong, G. Li, G. Yin, K. Li, D. Zhang, and W. Xu, “Long noncoding RNA WDFY3-AS2 represses the progression of esophageal cancer through miR-18a/PTEN Axis,” *Journal of Oncology*, vol. 2021, Article ID 9951010, 12 pages, 2021.
- [35] Y. Wu, T. Wang, L. Xia, and M. Zhang, “LncRNA WDFY3-AS2 promotes cisplatin resistance and the cancer stem cell in ovarian cancer by regulating hsa-miR-139-5p/SDC4 axis,” *Cancer Cell International*, vol. 21, no. 1, p. 284, 2021.
- [36] Y. Liu, C. Helms, W. Liao et al., “A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci,” *PLoS Genetics*, vol. 4, no. 4, article e1000041, 2008.
- [37] C. M. Lange, S. Bibert, J. F. Dufour et al., “Comparative genetic analyses point to *_HCP5_* as susceptibility locus for HCV-associated hepatocellular carcinoma,” *Journal of Hepatology*, vol. 59, no. 3, pp. 504–509, 2013.
- [38] H. Teng, P. Wang, Y. Xue et al., “Retracted: role of HCP5-miR-139-RUNX1 feedback loop in regulating malignant behavior of glioma cells,” *Molecular Therapy*, vol. 24, no. 10, pp. 1806–1822, 2016.
- [39] H. Zhu, Y. Zeng, C. C. Zhou, and W. Ye, “SNHG16/miR-216-5p/ZEB1 signal pathway contributes to the tumorigenesis of cervical cancer cells,” *Archives of Biochemistry and Biophysics*, vol. 637, pp. 1–8, 2018.
- [40] Y. Wen, X. Gong, Y. Dong, and C. Tang, “Long non coding RNA SNHG16 facilitates proliferation, migration, invasion and autophagy of neuroblastoma cells via sponging miR-542-3p and upregulating ATG5 expression,” *Oncotargets and Therapy*, vol. 13, pp. 263–275, 2020.
- [41] C. S. Pai, D. M. Simons, X. Lu et al., “Tumor-conditional anti-CTLA4 uncouples antitumor efficacy from immunotherapy-related toxicity,” *The Journal of Clinical Investigation*, vol. 129, no. 1, pp. 349–363, 2019.
- [42] J. N. Liu, X. S. Kong, T. Huang, R. Wang, W. Li, and Q. F. Chen, “Clinical implications of aberrant PD-1 and CTLA4 expression for cancer immunity and prognosis: a pan-cancer study,” *Frontiers in Immunology*, vol. 11, p. 2048, 2020.
- [43] K. Yang, J. Xu, Q. Liu, J. Li, and Y. Xi, “Expression and significance of CD47, PD1 and PDL1 in T-cell acute lymphoblastic lymphoma/leukemia,” *Pathology, Research and Practice*, vol. 215, no. 2, pp. 265–271, 2019.
- [44] S. Liu, F. Wang, W. Tan et al., “CTLA4 has a profound impact on the landscape of tumor-infiltrating lymphocytes with a high prognosis value in clear cell renal cell carcinoma (ccRCC),” *Cancer Cell International*, vol. 20, no. 1, p. 519, 2020.
- [45] E. J. Wherry and M. Kurachi, “Molecular and cellular insights into T cell exhaustion,” *Nature Reviews Immunology*, vol. 15, no. 8, pp. 486–499, 2015.
- [46] Y. Yu, “Molecular classification and precision therapy of cancer: immune checkpoint inhibitors,” *Frontiers in Medicine*, vol. 12, no. 2, pp. 229–235, 2018.

Research Article

Expression and the Prognostic Value of Biglycan in Gastric Cancer

Sizhe Hu ¹, Peipei Li ², Chenying Wang ², and Xiyong Liu ³

¹Department of Gastrointestinal Surgery, Affiliated Dongyang People's Hospital of Wenzhou Medical University, Dongyang, Zhejiang 322100, China

²School of Medicine, Zhejiang University City College, Hangzhou, Zhejiang 310015, China

³Sino-American Cancer Foundation, Covina, CA 91722, USA

Correspondence should be addressed to Sizhe Hu; husizhe@zju.edu.cn and Xiyong Liu; xiyongliu@sacfamerica.org

Received 13 June 2022; Accepted 23 August 2022; Published 6 September 2022

Academic Editor: Chuan Lu

Copyright © 2022 Sizhe Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Biglycan (BGN) is a family member of small leucine-rich repeat proteoglycans. High expression of BGN might enhance the invasion and metastasis in some types of tumors. Here, the prognostic significance of BGN was evaluated in gastric cancer. **Material and Methods.** Two independent Gene Expression Omnibus (GEO) gastric cancer microarray datasets ($n = 64$ and $n = 432$) were collected for this study. Kaplan-Meier analysis was applied to evaluate if BGN impacts the outcomes of gastric cancer. Protein-protein interaction (PPI) analysis was performed on gastric cancer-related genes and BGN targets, and those interactions with confidence interval (CI) ≥ 0.7 were chosen to construct a PPI network. The gene set enrichment analysis (GSEA) was used to explore BGN and cancer-related gene signatures. Gene Transcription Regulation Database (GTRD) and ALGGEN-PROMO predicted the transcription factor binding sites (TFBSs) of the BGN promoter. BGN protein level in gastric cancer tissue was determined by immunohistochemistry (IHC). Bioinformatic analysis predicted the putative TFs of BGN. **Results.** For gastric cancer, the mRNA expression level of BGN in tumor tissue was significantly higher than that in normal tissue. Kaplan-Meier analysis showed that higher expression of BGN mRNA was significantly associated with more reduced recurrence-free survival (RFS). GSEA results suggested that BGN was significantly enriched in gene signatures related to metastasis and poor prognosis, revealing that BGN might be associated with cell proliferation, poor differentiation, and high invasiveness of gastric cancer. Meanwhile, the putative TFs, including AR, E2F1, and TCF4, were predicted by bioinformatic analysis and also significantly correlated with expression of BGN in mRNA levels. **Conclusion.** High expression of BGN mRNA was significantly related to poor prognosis, which suggested that BGN was a potential prognostic biomarker and therapeutic target of gastric cancer.

1. Introduction

Gastric cancer is the sixth most common malignant tumor and is the second leading cause of cancer-induced death in the world [1]. In East Asia (China, Japan, and Korea), the incidence of gastric cancer is higher than in other areas over the world [2]. It was estimated that about one million new cases of gastric cancer were diagnosed globally in 2018, and about half of the new cases occurred in China [3]. The 5-year overall survival of gastric cancer is only 20% to 30% due to cancer progression [4], although numerous new treatments have been utilized, including but not limited to

chemotherapy, targeted therapy, and immunotherapy. However, for early gastric cancer, the 5-year overall survival is more than 90% [5]. Unfortunately, early-stage gastric cancer usually has no or only nonspecific symptoms. Thus, the appearance of symptoms usually suggests the advanced gastric cancer. Gastroscopy is a routine screening method for gastric cancer, but it is not widely accepted because it is invasive [6]. Currently, several tumor markers are used in the clinic for early detection of gastric cancer. These markers include carcinoembryonic antigen (CEA), pepsinogen, α -fetoprotein (AFP), carbohydrate antigens (CA), CA72-4, CA125, and CA24-2. However, the sensitivity and specificity

of these serum indicators are poor [7]. Thus, it is urgently needed to explore novel biomarkers for early diagnosis and prognosis prediction for gastric cancer patients.

Biglycan (BGN) is a family member of small leucine-rich repeat proteoglycans (SLRPs) characterized by a core protein with leucine-rich repeats [8]. Initially, BGN was only considered as a component maintaining the structural integrity of the extracellular matrix (ECM) and involved in the regulation of inflammatory response, skeletal muscle development, and regeneration [9, 10]. In a decade, it was found that BGN is a signal molecule, playing an essential role in angiogenesis, cell proliferation, differentiation, and migration [11–13]. In recent years, it has been gradually found that BGN is highly expressed in various malignant tumors, such as endometrial cancer [14], ovary cancer [15], pancreatic adenocarcinoma [16], esophageal squamous cell carcinoma [17], colorectal cancer [18], and gastric cancer [19], suggesting an essential role of BGN in the pathogenesis and progression of cancer. In some types of these cancers, high expression of BGN enhances the ability of invasion and metastasis of tumor cells [18–20] or contributes to poor prognosis [16, 17, 21, 22].

Therefore, BGN is closely related to the occurrence and development of a variety of tumors and is a potential target molecule for tumor treatment. The purpose of the present study is to verify the BGN expression and the prognostic value of BGN in gastric cancer. In this study, we investigated the prognostic value of BGN in gastric cancer by involving an external transcriptome data set from the TCGA database. To understand the role of BGN in gastric cancer, we analyzed our tissue microarray, including 125 cases of gastric cancers, for immunohistochemical BGN expression.

2. Materials and Methods

2.1. Microarray Gene Expression Datasets. Two independent Gene Expression Omnibus (GEO) gastric cancer microarray datasets (total $n = 496$) were collected for this study. There were 432 cases of gastric cancer patients from South Korea in GSE26253 dataset [23], and all participants had clinical and follow-up annotations. GSE65801 [24] contained 32 Chinese patients but had no follow-up annotations. Detailed information about the two downloaded datasets is listed in Table 1. To normalize the mRNA expression levels in the GSE26253 dataset, we restratified BGN scores into four grades (Q1, Q2, Q3, and Q4) based on the percentile. Low-BGN score grades (Q1+Q2) and high-BGN score grades (Q3+Q4) were also divided by the median value of gene expression.

The recurrence-free survival (RFS) period was defined as the time from initial surgery until tumor recurrence. Kaplan-Meier survival plot was used to display the proportion of the population’s RFS by the length of follow-up.

2.2. Gene Set Enrichment Analysis (GSEA). The GSEA software v3.0 was downloaded from <http://www.broad.mit.edu/gsea> and run on the Java 8.0 platform [25]. All dataset (.gct) and phenotype label (.cls) files were created and loaded into the GSEA software, and gene sets were updated from the above website. The detailed protocol could see in our

TABLE 1: Summary of gene expression datasets of gastric cancer.

Accession no.	GSE26253	GSE65801
No. of patients	432	32
No. of healthy	0	32
Platforms*	GPL8432	GPL14550
Country	South Korea	China
Date of study	2010-2019	2015-2016
Sex	N/A	Y
Clinical stage	Y	N/A
RFS months (range)	1.9-167.6	N/A

Platforms: GPL8432: Illumina HumanRef-8 WG-DASL v3.0. GPL14550: Agilent-028004 SurePrint G3 Human GE 8x60K Microarray (Probe Name Version).

previous publications [26]. Here, the permutation number was 1,000, and the phenotype label was ILMN_2206746 (BGN).

2.3. Data Management and Statistical Methods. Student’s t -test, one-way analysis of variance (ANOVA), and nonparametric tests were used to test differences among subgroups for continuous data. The Pearson Chi-square and likelihood test was used for categorical data analyses. Kaplan-Meier analysis was used to estimate the proportion of the population’s RFS by the length of follow-up in months. Hazard ratios (HRs) [21] with 95% confidence intervals (CI) were calculated using Cox proportional hazard regression analysis. Two-sided p values less than 0.05 were considered statistically significant. The R and JMP statistical software were used for the above analysis unless otherwise noted.

2.4. Eligible Transcription Factor (TF) Prediction. The promoter region of the BGN gene was visualized on <http://genome.ucsc.edu/cgi-bin/hgGateway>. The signal of H3K4Me3 was used to localize the promoter region. The TF binding sites (TFBSs) of BGN promoter were predicted by Gene Transcription Regulation Database (GTRD) (<http://gtrd.biouml.org/>) and ALGGEN-PROMO (http://algggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3).

2.5. Immunohistochemistry (IHC) Assays of Tissue Microarray. The protocol for the use of human tissue was approved by the Institutional Review Board [13] of the Affiliated Dongyang People’s Hospital of Wenzhou Medical University (Zhejiang, China). Before the study, all patients gave their written informed consent to allow us to use left tissue samples for scientific research. All eligible participants had received radical gastrectomy or palliative gastrectomy. The primary tumor samples were obtained from surgical specimens. The exclusion criteria of participants were those with (1) no informed consent signed and (2) multiple cancers. A total of 125 pairs of gastric cancer specimens, including cancerous tissue and adjacent normal tissue, that underwent surgery in 2018, were eventually enrolled. The above-mentioned tissue specimens were fixed in ethanol at 4°C for 1 h, followed by paraffin embedding. Thereafter, specimens were sliced with a

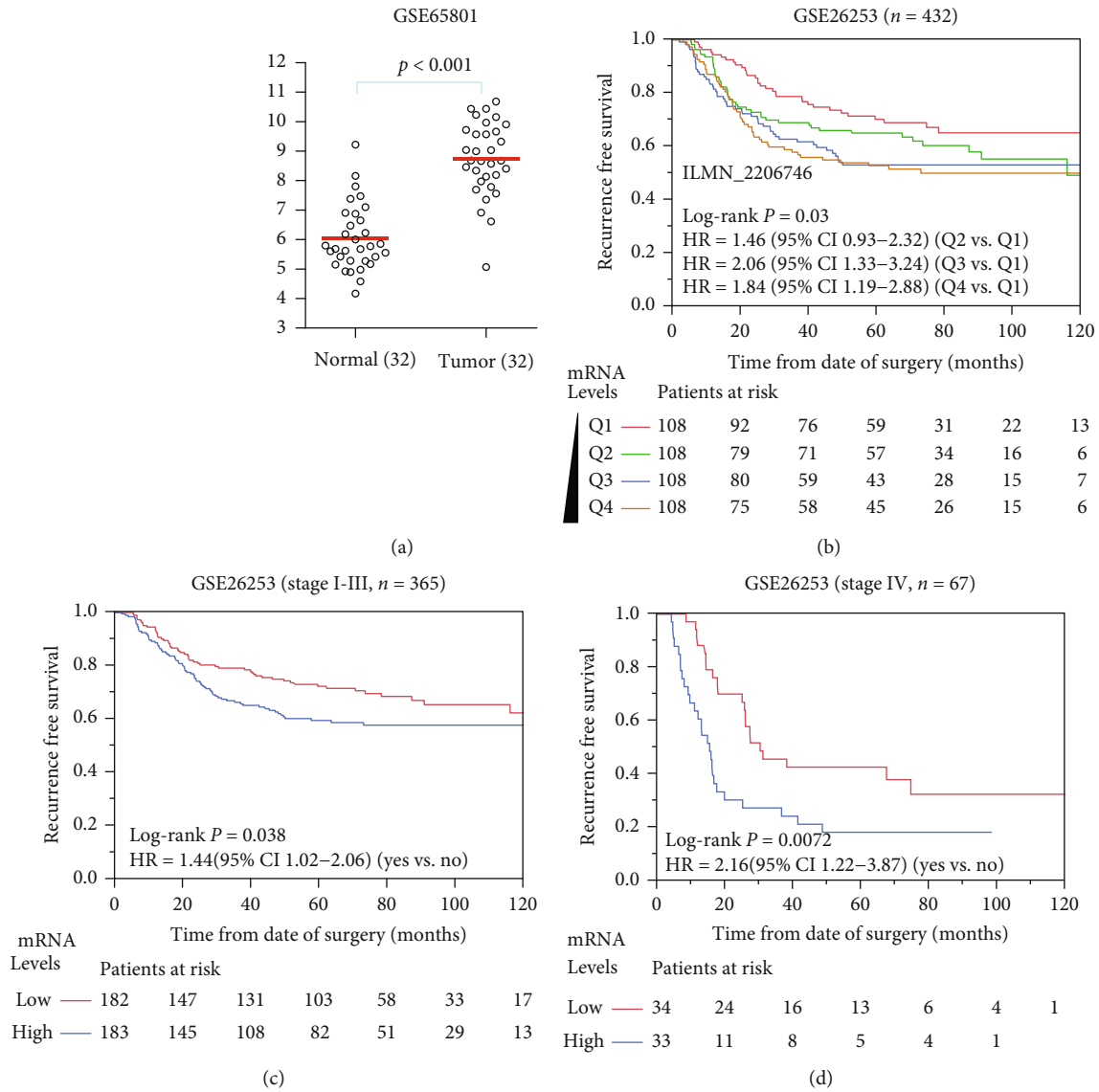


FIGURE 1: BGN expression in gastric cancer tissue and prognostic value of BGN. (a) Differential expression of BGN between normal and gastric cancer tissue in the GSE65801 dataset. (b) Kaplan-Meier analysis of BGN and RFS in the GSE26253 dataset. The curves of red, green, blue, and brown represented Q1, Q2, Q3, and Q4 subgroups, respectively. Q1: 0 to 25% percentile; Q2: 25% to the median; Q3: the median to 75% percentile; Q4: 75% percentile to the maximum. (c) BGN impacts poor RFS on stage I-III and (d) stage IV gastric cancer patients from GSE26253 dataset.

microtome into 4 μm sections. These sections were cultivated with 3% H_2O_2 at room temperature for 5-10 min to eradicate the activity of endogenous peroxidase, followed by 10 min of block with bovine serum albumin (BSA). The primary antibody working solution was then added dropwise followed by a 1-2 h cultivation at 37°C or an overnight one at 4°C. Phosphate-buffered saline (PBS) was introduced to rinse sections 3 times. Subsequently, the secondary antibody working solution was also introduced dropwise, and the system was incubated at 37°C for 10-30 min. Antibodies involved were as follows: anti-rabbit BGN (1:2000, ab209234, Abcam, UK) and goat anti-rabbit IgG (1:500, ab150077, Abcam, UK). The previously described protocols of deparaffinization and

immunohistochemistry (IHC) staining were used to apply to the multiple-tissue array [27].

3. Results

3.1. The Prognostic Significance of BGN for Gastric Cancer. In the GSE65801 dataset, we came to the same conclusion that BGN mRNA level was higher in tumor tissue than the normal tissue (Figure 1(a)). Kaplan-Meier analysis showed that higher expression of BGN was significantly associated with poorer RFS in gastric cancer patients. In the GSE26253 dataset, samples were divided into four subgroups, Q1, Q2, Q3, and Q4, according to the expression level of BGN. BGN mRNA levels

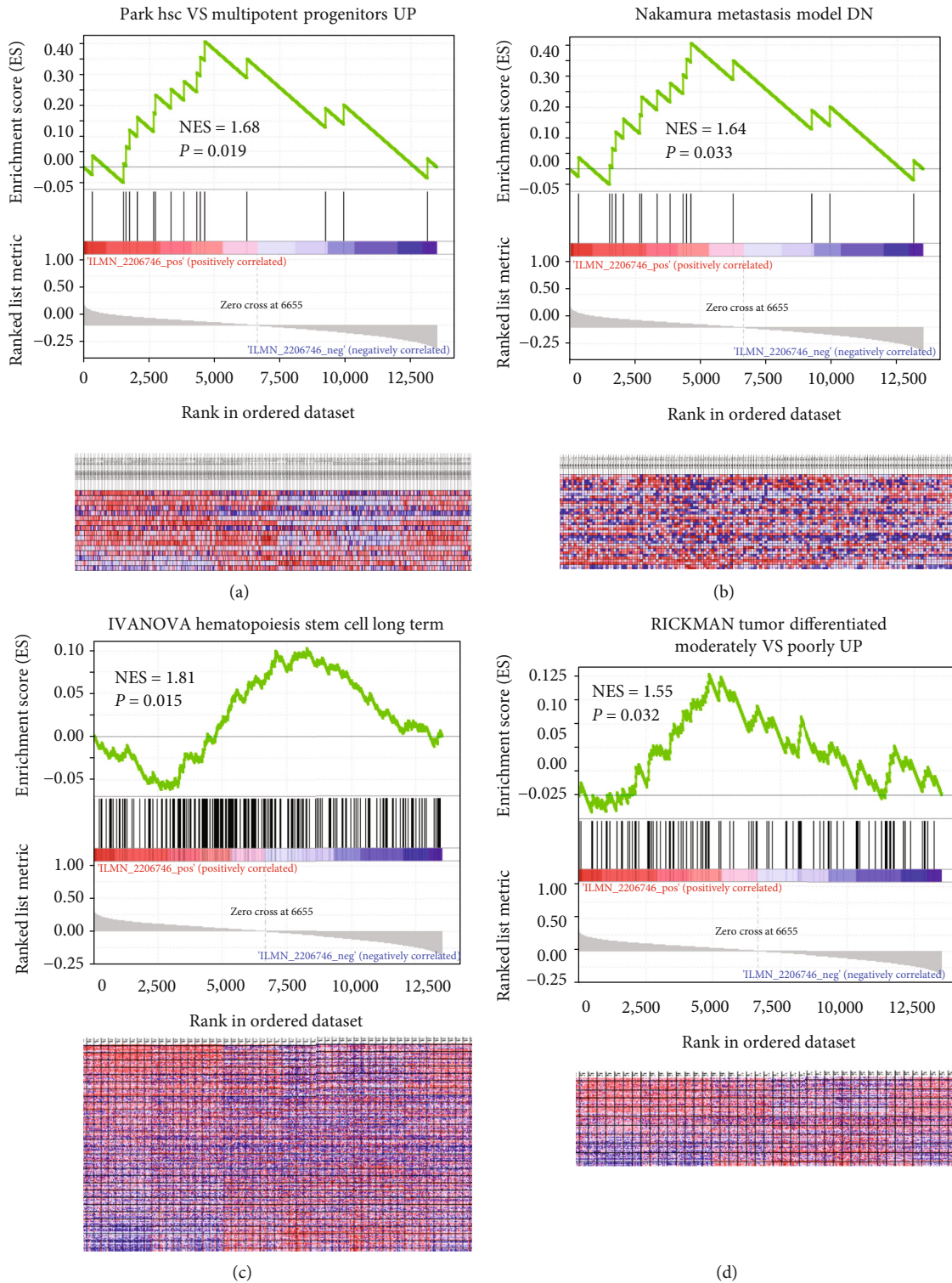


FIGURE 3: GSEA analysis of BGN. The expression of BGN was enriched in signatures of (a) Park hsc VS multipotent progenitors UP, (b) Nakamura metastasis model DN, (c) IVANOVA Hematopoiesis Stem Cell Long Term, and (d) RICKMAN Tumor Differentiated Moderately VS Poorly UP in GSE26253 dataset. As for heatmap of GSEA, columns are cases ranked by BGN expression, and rows are genes in the signature. Red represents the upregulated genes, and blue is the downregulated ones.

were negatively correlated with RFS of gastric cancer patients (Figure 1(b)). Therefore, the BGN expression level was negatively correlated with the prognosis of gastric cancer patients

in a dose-dependent manner. In a stratified survival analysis according to the pathological stage, samples were re-stratified as BGN-high (equal or greater BGN levels than the median)

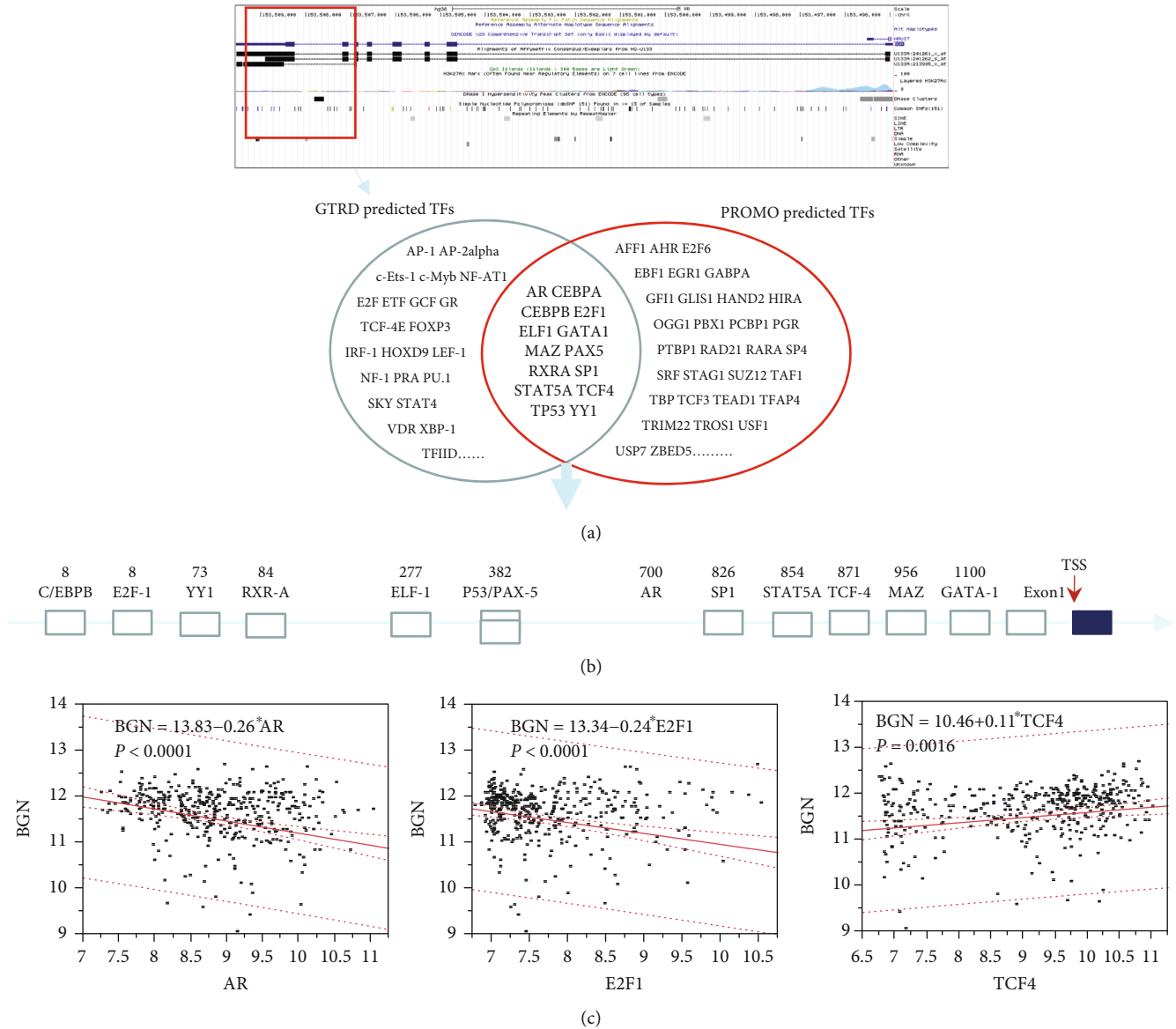


FIGURE 4: Prediction of putative TFs regulating BGN in gastric cancer. The prediction of BGN promoter region was processed using GTRD (<http://gtrd.biouml.org/>) and ALGGEN-PROMO (http://algggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3) databases. The potential TFBSs were predicted by GTRD and ALGGEN-PROMO. (a) Overlapped TFs with sequence alignment and correlation significance were considered as eligible TFs. (b) The location of eligible TFBSs on the promoter of BGN. (c) Scatter plot of correlation between AR, E2F1, TCF4, and BGN expression levels.

and BGN-low (less BGN levels than the median), according to the expression levels of BGN mRNA. The HRs were 1.44 (95% CI 1.02-2.06, $p = 0.038$) and 2.16 (95% CI 1.22-3.87, $p = 0.007$) for high BGN expression in stages I-III ($n = 365$) and stage IV gastric cancer patients, respectively (Figures 1(c) and 1(d)). These results suggested that high BGN mRNA levels were significantly related to poor prognosis of gastric cancer patients.

3.2. Bioinformatics Analysis for the Gene and Protein Interaction Network of BGN. To understand the biological functions of BGN, we conducted bioinformatics analysis for genes coexpressed BGN on Oncomine. The analysis of

genes coexpressed with BGN was conducted on Chen Gastric dataset [28]. We screened more than 10 genes with a strong correlation with BGN, such as THBS2, ARHGAP5, FN1, INHBA, and CDH11 (Figure 2(a)). Meanwhile, the bioinformatics analysis for the protein-protein interaction (PPI) network was conducted using STRING database (<http://www.string-db.org>). Figure 2(b) shows the PPI network of BGN; more than a dozen of genes were reported interacting with BGN through text mining, including genes like VCAN, TLR4, HSPG2, TGFB1, and GPC1. Most of the above genes were involved in cell growth, cell communication, signal transduction, and cell adhesion (Figure 2(c)), which was closely related to tumorigenesis.

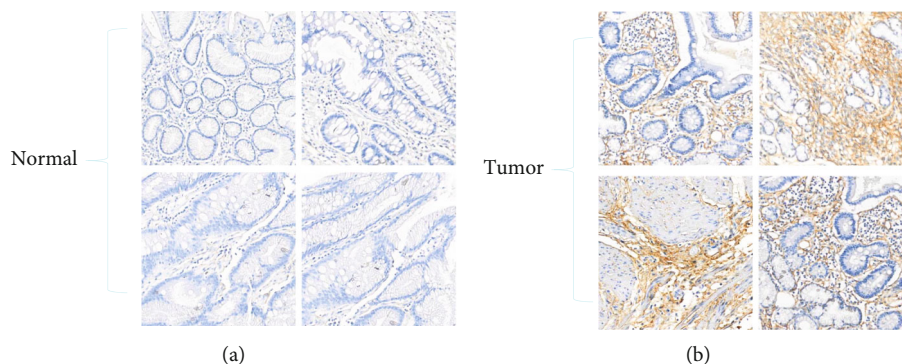


FIGURE 5: Protein expression level of BGN in gastric cancer tissue was determined by immunohistochemical (IHC) staining. The representative images are (a) normal tissue and (b) cancerous tissue of gastric cancer.

3.3. GSEA of BGN in Gastric Cancer. To explore the cancer-related gene signatures of BGN, we performed a GSEA on the GSE26253 dataset, a downloaded microarray dataset of 432 gastric cancer cases. The expression of BGN was significantly associated with the following gene sets: Park hsc VS multipotent progenitors UP (Figure 3(a)), Nakamura metastasis model DN (Figure 3(b)), IVANOVA Hematopoiesis Stem Cell Long Term (Figure 3(c)), and RICKMAN Tumor Differentiated Moderately VS Poorly UP (Figure 3(d)) in GSE26253 dataset. GSEA results suggested that BGN was significantly enriched in gene signatures related to metastasis and poor prognosis, revealing that BGN might be associated with proliferation, poor differentiation, and high invasiveness of gastric cancer.

3.4. Prediction of Putative TFs of BGN by Bioinformatic Analysis. In order to further understand the carcinogenic mechanism, it is essential to explore the upstream regulation of BGN in gastric cancer. The TF prediction for BGN promoter region was processed using GTRD (<http://gtrd.biouml.org/>) and ALGGEN-PROMO databases (http://algggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3). In Figure 4(a), the promoter region of BGN would be around the signal of H3K27Ac, which was located around the 1st exon and partially overlapped with CpG island. Meanwhile, the potential TFBSs were screened through GTRD and PROMO databases. By intersecting the two groups of gene sets, the eligible TFs were identified, including AR, CEBPA, CEBPB, E2F1, ELF1, GATA1, MAZ, PAX5, RXRA, SP1, STAT5A, TCF4, TP53, and YY1. Figure 4(b) shows the location of these eligible TFBSs on the promoter of BGN. A linear regression analysis indicated that the expression of BGN was significantly and positively associated with TCF4 level, while negatively associated with AR or E2F1 (Figure 4(c)).

3.5. BGN Protein Level in Gastric Cancer Tissue. For gastric cancer, the protein expression level of BGN in tumor tissue was significantly higher than that in normal tissue (Figure 5). Unfortunately, BGN was mainly expressed in the extracellular matrix rather than in the intracellular matrix, which made quantitative analysis difficult.

4. Discussion

BGN, a member of the family of small leucine-rich repeat proteoglycans (SLRPs), is only considered as a component maintaining the structural integrity of extracellular matrix, involved in the regulation of inflammatory response, skeletal muscle development, and regeneration [9, 10]. In recent years, it has been gradually found that BGN is closely related to the occurrence and development of various malignant tumors, such as endometrial cancer [14], ovary cancer [15], pancreatic adenocarcinoma [16], esophageal squamous cell carcinoma [17], colorectal cancer [18], and prostate cancer [21]. In some malignant tumors, higher expression of BGN predicts more considerable invasiveness and worse prognosis [16, 17, 21, 22]. Therefore, it is valuable to reevaluate the prognostic significance and clinical meaning of BGN on other cancers.

A previous study has shown that BGN promotes tumor invasion and metastasis in gastric cancer both *in vitro* and *in vivo* and is associated with TNM stage. BGN plays an oncogenic role by activating the FAK signaling pathway in gastric cancer [19]. In this study, through analysis of public datasets (Figure 1(a)) and immunohistochemical analysis of tissue arrays (Figure 5), we confirmed that BGN expression was higher in tumor tissue than that in normal tissue. Unfortunately, since BGN was mainly distributed in the extracellular matrix, it cannot be quantified. Besides, we acquired a public microarray dataset, the GSE26253 dataset, containing 432 gastric cancer cases. Kaplan-Meier analysis of BGN for the RFS revealed that higher BGN expression level portended poorer prognosis in gastric cancer patients ($p = 0.03$). Stratification analysis showed that BGN was significantly associated with RFS of both stage I-III ($p = 0.038$) and stage IV ($p = 0.007$) patients with gastric cancer (Figure 1). Meanwhile, to explore the cancer-related gene signatures of BGN, we performed a GSEA on the GSE26253 dataset, revealing that BGN might be associated with poor proliferation, poor differentiation, and high invasiveness of gastric cancer. Also, we analyzed and predicted the potential TFs of BGN by bioinformatic analysis.

Limitations of this study included the following: (1) the protein expression levels of BGN could not be evaluated by

immunohistochemistry (IHC) analysis. In gastric cancer tissue samples, the signal of BGN protein could only be seen in the extracellular matrix rather than in the intracellular matrix (Figure 5), which made it difficult for quantification. Meanwhile, (2) the mechanisms of BGN-associated aggressiveness and poor outcome of gastric cancer were still not clarified. (3) It needs to be further validated if BGN was a therapeutic target by experimental study.

Taken together, high BGN level could be enriched in gene signatures related to poor proliferation, poor differentiation, and high invasiveness. Kaplan-Meier analysis revealed that overexpression of BGN was significantly associated with poorer RFS in a dose-dependent manner in both stage I-III and stage IV gastric cancer patients. Therefore, BGN may be a potential prognostic and therapeutic biomarker for gastric cancer.

Abbreviations

BGN:	Biglycan
GSEA:	Gene set enrichment analysis
RFS:	Recurrence-free survival
HR:	Proportional hazard ratios
CEA:	Carcinoembryonic antigen
AFP:	α -Fetoprotein
CA:	Carbohydrate antigens
SLRPs:	Small leucine-rich repeat proteoglycans
ECM:	Extracellular matrix
95% CI:	95% confidence interval
TFBSs:	Transcription factor binding sites
IHC:	Immunohistochemistry
IRB:	Institutional Review Board.

Data Availability

All the datasets used and analyzed during the current study are downloaded from NIH Gene Expression Omnibus (GEO) and ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>) which are available from the corresponding authors on reasonable request.

Ethical Approval

The protocol for the use of human tissue was approved (batch number: 2019-YX-076) by the Institutional Review Board [13] of the Affiliated Dongyang People's Hospital of Wenzhou Medical University (Zhejiang, China).

Consent

Before the study, all patients gave their written informed consent to allow us to use left tissue samples for scientific research.

Disclosure

A preprint has previously been published [29]. The link is as follows: <https://www.researchsquare.com/article/rs-103093/v1>.

Conflicts of Interest

The authors declare that they have no competing interests, and all authors should confirm its accuracy.

Authors' Contributions

SH carried out the collection of gastric cancer patients' specimens, the production and analysis of tissue chips, data collection, and manuscript writing. PL performed statistics and bioinformatics analysis for gastric cancer datasets, GSEA, and English editing for the manuscript. CW involved in study design and performed bioinformatics analysis of BGN. XL was charged for study design, data analysis, and manuscript writing. All authors read and approved the final manuscript.

Acknowledgments

We gratefully acknowledge NIH Gene Expression Omnibus (GEO) and ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>) which made the genomic data and clinical data of gastric cancer available. This project was supported by the Jinhua Science and Technology Development Plans (Grant number 2020-04-134) (Zhejiang, China) and the Science and Education Institute of Dongyang People's Hospital (Grant number 2018-Y-002) (Zhejiang, China).

References

- [1] S. S. Kupfer, "Gaining ground in the genetics of gastric cancer," *Gastroenterology*, vol. 152, no. 5, pp. 926–928, 2017.
- [2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 1, p. 7, 2018.
- [3] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [4] D. G. Power, D. P. Kelsen, and M. A. Shah, "Advanced gastric cancer – slow but steady progress," *Cancer Treatment Reviews*, vol. 36, no. 5, pp. 384–392, 2010.
- [5] H. H. Hartgrink, E. P. M. Jansen, N. C. T. van Grieken, and C. J. H. van de Velde, "Gastric cancer," *The Lancet*, vol. 374, no. 9688, pp. 477–490, 2009.
- [6] R. Sitarz, M. Skierucha, and J. Mielko, "Gastric cancer: epidemiology, prevention, classification, and treatment," *Cancer Management & Research*, vol. 10, pp. 239–248, 2018.
- [7] W. Tong, F. Ye, L. He et al., "Serum biomarker panels for diagnosis of gastric cancer," *Oncotargets & Therapy*, vol. 9, pp. 2455–2463, 2016.
- [8] L. Schaefer and R. V. Iozzo, "Biological functions of the small leucine-rich proteoglycans: from genetics to signal transduction," *Journal of Biological Chemistry*, vol. 283, no. 31, pp. 21305–21309, 2008.
- [9] L. Schaefer, A. Babelova, E. Kiss et al., "The matrix component biglycan is proinflammatory and signals through Toll-like receptors 4 and 2 in macrophages," *The Journal of clinical investigation*, vol. 115, no. 8, pp. 2223–2233, 2005.
- [10] M. L. Mercado, A. R. Amenta, H. Hagiwara et al., "Biglycan regulates the expression and sarcolemmal localization of

- dystrobrevin, syntrophin, and nNOS,” *Faseb Journal Official Publication of the Federation of American Societies for Experimental Biology*, vol. 20, no. 10, pp. 1724–1726, 2006.
- [11] M. V. Nastase, M. F. Young, and L. Schaefer, “Biglycan: a multivalent proteoglycan providing structure and signals,” *Journal of Histochemistry & Cytochemistry Official Journal of the Histochemistry Society*, vol. 60, no. 12, pp. 963–975, 2012.
- [12] A. D. Berendsen, E. L. Pinnow, A. Maeda et al., “Biglycan modulates angiogenesis and bone formation during fracture healing,” *Matrix Biology Journal of the International Society for Matrix Biology*, vol. 35, pp. 223–231, 2014.
- [13] M. Myren, D. J. Kirby, M. L. Noonan et al., “Biglycan potentially regulates angiogenesis during fracture repair by altering expression and function of endostatin,” *Matrix Biology*, vol. 52–54, pp. 141–150, 2016.
- [14] Y. Liu, W. Li, X. Li et al., “Expression and significance of biglycan in endometrial cancer,” *Archives of Gynecology & Obstetrics*, vol. 289, no. 3, pp. 649–655, 2014.
- [15] S. Pan, L. Cheng, J. T. White et al., “Quantitative proteomics analysis integrated with microarray data reveals that extracellular matrix proteins, catenins, and P53 binding protein 1 are important for chemotherapy response in ovarian cancers,” *Omics A Journal of Integrative Biology*, vol. 13, no. 4, pp. 345–354, 2009.
- [16] G. Aprile, C. Avellini, M. Reni et al., “Biglycan expression and clinical outcome in patients with pancreatic adenocarcinoma,” *Tumor Biology*, vol. 34, no. 1, pp. 131–137, 2013.
- [17] Y. H. Zhu, F. Yang, S. S. Zhang, T. T. Zeng, X. Xie, and X. Y. Guan, “High expression of biglycan is associated with poor prognosis in patients with esophageal squamous cell carcinoma,” *International Journal of Clinical & Experimental Pathology*, vol. 6, no. 11, pp. 2497–2505, 2013.
- [18] X. Gu, Y. Ma, J. Xiao et al., “Up-regulated biglycan expression correlates with the malignancy in human colorectal cancers,” *Clinical & Experimental Medicine*, vol. 12, no. 3, pp. 195–199, 2012.
- [19] L. Hu, Y. T. Duan, J. F. Li et al., “Biglycan enhances gastric cancer invasion by activating FAK signaling pathway,” *Oncotarget*, vol. 5, no. 7, pp. 1885–1896, 2014.
- [20] H. Sun, X. Wang, Y. Zhang et al., “Biglycan enhances the ability of migration and invasion in endometrial cancer,” *Archives of Gynecology and Obstetrics*, vol. 293, no. 2, pp. 429–438, 2016.
- [21] F. Jacobsen, J. Kraft, C. Schroeder et al., “Up-regulation of biglycan is associated with poor prognosis and PTEN deletion in patients with prostate cancer,” *Neoplasia*, vol. 19, no. 9, pp. 707–715, 2017.
- [22] G. B. Schulz, T. Grimm, C. Sers et al., “Prognostic value and association with epithelial-mesenchymal transition and molecular subtypes of the proteoglycan biglycan in advanced bladder cancer,” *Urologic Oncology*, vol. 37, no. 8, pp. 530–530, 2019.
- [23] J. Lee, I. Sohn, I. G. do et al., “Nanostring-based multigene assay to predict recurrence for gastric cancer patients after surgery,” *PLoS One*, vol. 9, no. 3, article e90133, 2014.
- [24] H. Li, B. Yu, J. Li et al., “Characterization of differentially expressed genes involved in pathways associated with gastric cancer,” *PLoS One*, vol. 10, no. 4, 2015.
- [25] A. Subramanian, P. Tamayo, V. K. Mootha et al., “Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.
- [26] J. Ding, M. L. Kuo, L. Su et al., “Human mitochondrial pyrroline-5-carboxylate reductase 1 promotes invasiveness and impacts survival in breast cancers,” *Carcinogenesis*, vol. 38, no. 5, pp. 519–531, 2017.
- [27] T. S. Halstensen, H. Scott, and P. Brandtzaeg, “Human CD8+ intraepithelial T lymphocytes are mainly CD45RA–RB+ and show increased co-expression of CD45R0 in celiac disease,” *European Journal of Immunology*, vol. 20, no. 8, pp. 1825–1830, 1990.
- [28] X. Chen, S. Y. Leung, S. T. Yuen et al., “Variation in gene expression patterns in human gastric cancers,” *Molecular Biology of the Cell*, vol. 14, no. 8, pp. 3208–3215, 2003.
- [29] S. Hu, P. Li, C. Wang, and X. Liu, *Prognostic Value of Biglycan in Gastric Cancer*, 2020.

Research Article

Endoscopic Surgical Treatment of Osteoarthritis and Prognostic Model Construction

Qi Su  and Guokang Xu 

Department of Orthopedics, The First People's Hospital of Fuyang District, Hangzhou, China

Correspondence should be addressed to Qi Su; susususuqiiii@163.com

Received 13 June 2022; Revised 15 August 2022; Accepted 20 August 2022; Published 5 September 2022

Academic Editor: Tao Huang

Copyright © 2022 Qi Su and Guokang Xu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose. Osteoarthritis (OA) is a degenerative disease of joints. Currently, there is still a lack of effective tools to predict the long-term efficacy of surgical treatment of OA. The purpose of this study was to explore the prognostic factors of endoscopic surgery for OA and to predict the long-term efficacy of this type of surgery for OA by establishing a prognostic model. **Methods.** Baseline and follow-up data on 236 OA patients who underwent surgery in our hospital from January 2017 to December 2021 were selected and patients were randomly assigned to a training set ($n = 165$) and a test set ($n = 71$). The Pearson correlation coefficient was used to analyze the correlation between features. Feature selection was performed by recursive feature elimination (RFE) and linear regression. K -means clustering analysis was performed on the selected features to obtain the number of output layers. Finally, a single hidden layer error backpropagation (BP)-artificial neural network (ANN) model was established on the training set, and receiver operating characteristic (ROC) curve was drawn on the test set for verification. **Results.** Correlation analysis revealed no redundancy among features. RFE and linear regression screened out the features associated with postoperative prognosis under endoscopic surgery: sex, age, BMI, region, morning stiffness time, step count, and osteophyte area. K -means clustering yielded that the optimal number of categories was three, the same as the number of categories for the outcome variable. Therefore, a 7-1-3 BP neural network model was established based on these 7 features, and this model could predict the postoperative situation within one year to a relatively accurate extent: area under curve values (AUC) were 0.814, 0.700, and 0.761 in patients with worse, unchanged, and improved conditions one year after surgery, respectively, higher than the multiclass AUC value (0.646). **Conclusion.** The prognostic model of endoscopic surgery for OA constructed in this study can well predict the disease progression of patients within one year after surgery.

1. Introduction

Osteoarthritis (OA) is a kind of degenerative joint disease that usually occurs in the elderly, and age over 65 years is a clinical risk factor for OA [1]. In addition to age, obesity and joint injury are also common causes of OA [2]. The rising number of years lived with disability (YLD) of OA patients is adding more burdens on the health care and economy in many countries [3]. For OA patients who do not respond to basic treatment or drug therapy, surgery is an alternative to relieve their symptoms [4]. However,

according to survey data, only 12.5% of OA patients have effective symptom improvement one to two years after surgery [5]. In-depth study on how to effectively predict the long-term efficacy of surgery for OA patients is still deficient.

Arthroscopic surgery is a routine choice for alleviating OA symptoms and is considered to be effective in improving the clinical symptoms of OA patients [6]. In one study on the effect of arthroscopic repair surgery, Stafford et al. [7] found that the condition of the hip joint of patients undergoing surgical treatment is significantly improved. However, the efficacy of arthroscopic surgery for knee lesions is still

controversial. In a placebo-controlled study, the clinical outcome of arthroscopic surgery for knee OA did not significantly improve patients' clinical symptoms compared with drug therapy or placebo [8, 9]. As the efficacy of surgical treatment for OA is affected by many clinicopathological factors, how to select OA patients that are appropriate for surgical intervention needs to be studied.

With the continuous development of machine learning algorithms, artificial neural network (ANN) has been increasingly applied in medical field. Janjic et al. [10] predicted the neural development of premature infants using feed-forward neural networks (fNNs), and their predictive sensitivity on cognitive delay and motor delay in preterm infants reached 85.7% and 76.9%, respectively. Error back-propagation (BP) algorithm-based ANN also plays an important role in the study of immune-mediated diseases [11]. However, few studies have been conducted to predict the long-term outcomes of OA patients undergoing surgical treatment.

Based on the above background, this study intended to analyze the prognostic influencing factors by recursive feature elimination (RFE) and linear regression, based on which we constructed an ANN model to predict the prognosis of hip or knee OA patients after arthroscopic surgery. The purpose of this study was to develop a model that can effectively predict the long-term outcomes of OA patients undergoing surgical treatment and to provide a high-confidence decision-making tool for OA patients' clinical treatment.

2. Materials and Methods

2.1. Subjects. In this study, 236 OA patients who received surgical treatment in The First People's Hospital of Fuyang District from January 2017 to December 2021 were enrolled. The inclusion criteria were as follows: [1] in line with the clinical manifestations described in the *Guidelines for Diagnosis and Treatment of Osteoarthritis (2018 Edition)* [4]: varying degrees of joint pain and tenderness, joint swelling, limitation of motion, deformity, bone friction, and muscle atrophy; [2] all were diagnosed as knee or hip osteoarthritis by X-ray examination; [3] the patient received endoscopic surgery on the joint for the first time. The exclusion criteria were as follows: [1] coagulation dysfunction; [2] combined with heart, lung, brain, and kidney dysfunction or serious mental illness. The research program was approved by the Medical Ethics Committee of The First People's Hospital of Fuyang District. The informed consent of patients was abandoned since this study was a retrospective study and involved no clear personal information about patients.

2.2. Data Collection. Preoperative baseline data of patients were collected. The preoperative baseline data included age, sex, body mass index (BMI), morning stiffness time, Kellgren and Lawrence classification, osteophyte area, step count, site of onset, type of surgery, and the Western Ontario and McMaster Universities Arthritis Index (WOMAC) scores before surgery and one year after surgery. WOMAC is a disease-specific instrument for assessing pain, stiffness, and physical function in patients with hip or knee

osteoarthritis [12]. Outcome variables were divided into 3 categories. "Better": postoperative WOMAC score - preoperative WOMAC score <0 ; "Maintain": postoperative WOMAC score - preoperative WOMAC score = 0; "Worse": postoperative WOMAC score - preoperative WOMAC score >0 .

2.3. Feature Selection. In the pre-processing stage of data, the correlation between features was calculated first. Then, features with a correlation less than 0.5 were subjected to RFE for secondary screening. RFE is a wrapping feature selection method. Starting from including all features, RFE ranks the importance of all features by establishing a model; the modeling process is repeated after removing the least important features in sequence until the optimal feature subset is screened out [13]. Subsequently, the selected features were used as the input layer to construct the BP-ANN model.

The features selected by RFE were analyzed by k -means clustering. The optimal number of classifications obtained by the k -means cluster analysis was three categories, which were the same as the outcome variables set in this study. Therefore, the output layer was 3 in the establishment of the BP neural network.

2.4. Machine Learning Model Construction and Solving Algorithms. To predict the long-term efficacy of endoscopic surgery for OA, a feed-forward neural network was used for modeling. The patients were randomly assigned to a training set ($n=165$) and a test set ($n=71$), and a single hidden layer BP-ANN model was established with the training set.

In order to predict the long-term efficacy of endoscopic surgery for OA, we adopted fNN for modeling. fNN is a one-way multi-layer neural network, whose structure includes input layer, hidden layer, and output layer (Figure 1). Parameters of the model mainly include weights of connections between layers and thresholds of both hidden layer and output layer (ω , ν , θ , and γ in Figure 1). In this study, the features obtained from RFE were brought into the neurons in the input layer and solved by BP algorithm. During training, we set the learning rate to 0.001 and the BP algorithm adjusts parameters of models based on the idea of gradient descent method until the accuracy of the result reaches the required level [14].

2.5. Model Verification. The receiver operating characteristic (ROC) curve was used to evaluate the established model in the test set. Since the purpose of the model was to predict a result of three-way classification, we plotted the ROC curve of each category and the overall ROC curve to reflect the accuracy of prediction with the area under curve (AUC).

2.6. Statistical Analysis. The selection of features in this study was performed using "caret" package, and the establishment and solving of the neural network model depended on the "nnet" package in R package. In the modeling process, the number of neurons in the hidden layer was set in turn from small to large. After ten-fold cross verification, the number of neurons corresponding to the minimum mean squared error (MSE) was taken. ROC curve plotting

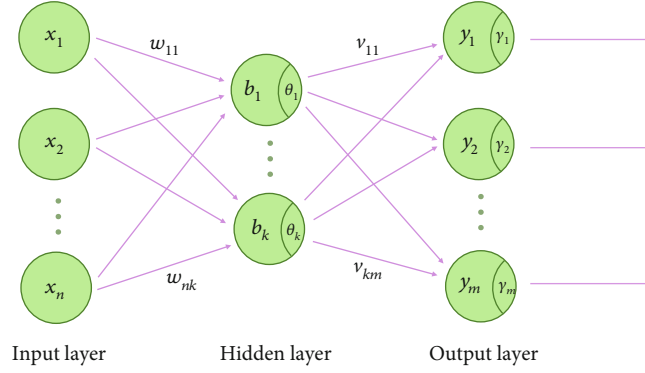


FIGURE 1: Structure of fNN. Note: $x_1, \dots, x_n, b_1, \dots, b_k$, and y_1, \dots, y_m are neurons of input layer, hidden layer, and output layer, respectively; w_{11}, \dots, w_{nk} and v_{11}, \dots, v_{km} are the connection weights between input layer and hidden layer and between hidden layer and output layer; $\theta_1, \dots, \theta_k$ and $\gamma_1, \dots, \gamma_m$ are the thresholds of the hidden layer and the output layer, respectively.

TABLE 1: Baseline data and pathological features of patients.

Feature	Description
Sample size, n	236
Age, years	67.00 (60.00, 70.00)
Sex, %	
Female	146 (61.9)
Male	90 (38.1)
BMI	24.51 (21.26, 26.60)
Morning stiffness time, %	
≤10 minutes	142 (60.2)
> 10 minutes and ≤ 20 minutes	43 (18.2)
>20 minutes	51 (21.6)
Kellgren and Lawrence classification, %	
I	18 (7.6)
II	84 (35.6)
III	83 (35.2)
IV	51 (21.6)
Osteophyte area, mm^2	5.483 (4.030, 6.830)
Daily average step count, %	
≤3000	57 (24.2)
>3000 and ≤8000	101 (42.8)
>8000	78 (33.1)
Involved parts, %	
Hip joint	128 (54.2)
Knee joint	108 (45.8)
Number of joints involved, %	
Unilateral	132 (55.9)
Bilateral	104 (44.1)
Surgical type, %	
Articular cartilage repair	84 (35.6)
Arthroscopic debridement	117 (49.6)
Osteotomy	35 (14.8)

was completed with the “pROC” package. Continuous variables conforming to normal distribution were represented as mean \pm standard deviation, while continuous variables not conforming to normal distribution were represented with median value (interquartile range). Counting variables were represented as n , %.

3. Results

3.1. Patient Features. A total of 236 patients were included in the study, with an age of 67.00 (60.00, 70.00) years, 38.1% male and 61.9% female, median BMI of 24.51 (21.26, 26.60). The number of patients with morning stiffness time less than or equal to 10 min, within 10-20 min, and more than 20 min was 142, 43, and 51, respectively. Other articular-related information, including Kellgren and Lawrence classification, osteophyte area, mean daily step count, lesion status, and surgical type, are shown in Table 1.

3.2. Feature Selection. To test the redundancy between features, we calculated the Pearson correlation coefficient between features and obtained the correlation matrix (Table 2). The absolute values of the correlation coefficients between patient features were all <0.5 , so there was no redundancy between features.

Linear regression was used to calculate the importance of features, and the results showed that the importance of sex, age, BMI, and other features decreased in descending order (Figure 2(a)). The results of RFE screening showed that the root mean square error (RMSE) of the model reached the minimum value when seven features were retained (Figure 2(b)). These features included sex, age, BMI, region, morning stiffness time, step count, and osteophyte area, and the minimum RMSE was 0.687. Therefore, these 7 features were used as the input layer for building the BP-ANN model.

Furthermore, the factors screened by RFE method were subjected to cluster analysis. The optimal number of clusters of k -means was determined by the total cluster sum of squares. The optimal results were obtained when the number of clusters was 3 (Figures 3(a) and 3(b)), which was the

TABLE 2: Correlation matrix between features.

	Age	Sex	BMI	Morning stiffness time	Region	KL score	Osteophyte area	Step count	Joint	Involved joint	Surgical type	Evaluation
Age	1.000	-0.062	0.008	-0.106	-0.010	0.003	0.021	-0.048	0.021	0.044	0.068	-0.293
Sex	-0.062	1.000	-0.036	0.029	0.152	-0.056	-0.083	0.023	-0.056	0.006	0.073	0.319
BMI	0.008	-0.036	1.000	0.022	0.087	0.007	-0.010	-0.182	0.002	0.122	0.089	-0.235
Morning stiffness time	-0.106	0.029	0.022	1.000	-0.012	0.031	0.009	-0.027	0.069	0.011	0.062	0.120
Region	-0.010	0.152	0.087	-0.012	1.000	-0.043	0.025	0.225	0.147	-0.179	-0.063	0.172
KL score	0.003	-0.056	0.007	0.031	-0.043	1.000	0.059	0.033	-0.081	-0.073	-0.262	-0.040
Osteophyte area	0.021	-0.083	-0.010	0.009	0.025	0.059	1.000	0.080	-0.074	-0.025	-0.039	-0.051
Step count	-0.048	0.023	-0.182	-0.027	0.225	0.033	0.080	1.000	0.072	0.054	-0.105	0.152
Joint	0.021	-0.056	0.002	0.069	0.147	-0.081	-0.074	0.072	1.000	0.041	0.005	0.024
Involved joint	0.044	0.006	0.122	0.011	-0.179	-0.073	-0.025	0.054	0.041	1.000	0.083	-0.060
Surgical type	0.068	0.073	0.089	0.062	-0.063	-0.262	-0.039	-0.105	0.005	0.083	1.000	-0.016
Evaluation	-0.293	0.319	-0.235	0.120	0.172	-0.040	-0.051	0.152	0.024	-0.060	-0.016	1.000

Note: BMI: body mass index; KL score: Kellgren-Lawrence classification.

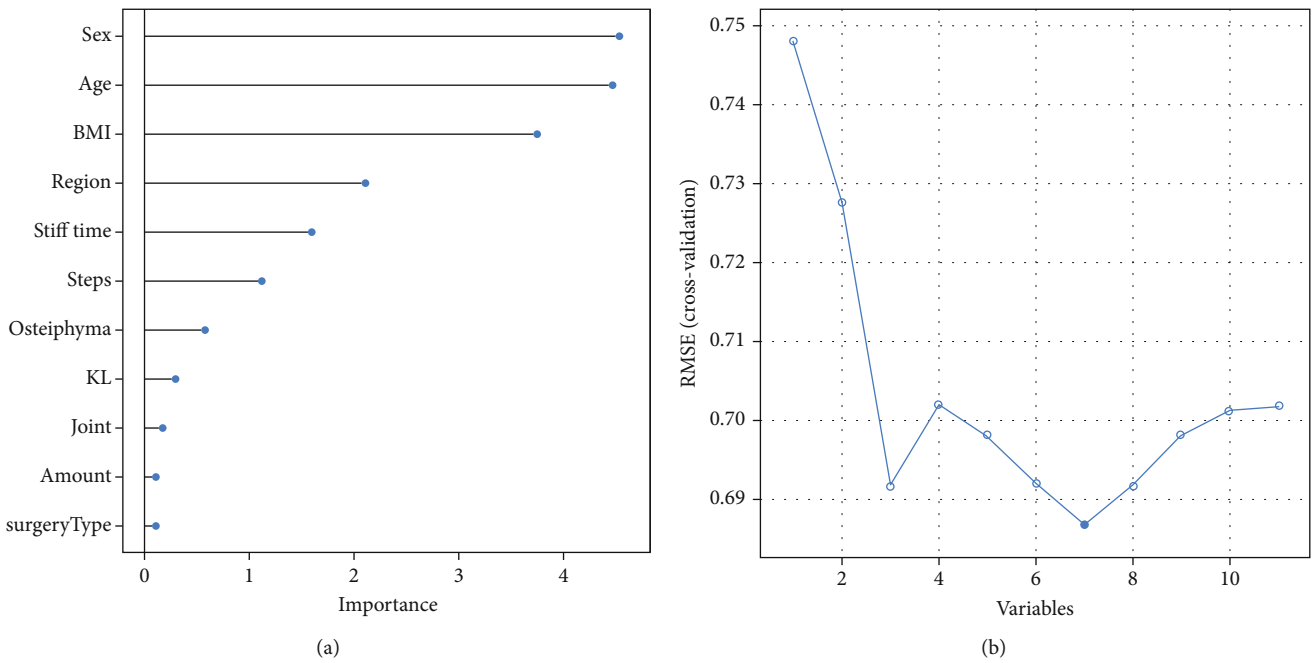


FIGURE 2: Feature selection by linear regression. (a) Order of importance of all features in linear regression; (b) Selection of optimal number of retained features by RFE. Note: RMSE: root mean square error.

same as the cluster number of outcome variables in this study.

3.3. *BP-ANN Model.* Seven features (sex, age, BMI, region, morning stiffness time, step count, and osteophyte area) screened out by the RFE method in the training set were used as the input layer, and three types of outcome variables as the output layer to build a single hidden layer neural network model. When the maximum number of iterations was set to 200, a neural network model was established by BP

algorithm (Figure 4). The maximum number of iterations was not achieved when the computation stopped.

3.4. *Model Evaluation.* After the BP neural network was trained on training set, the predicted classification of the test set was obtained, and then the accuracy of the model was verified with the test set and the ROC curve of the model validation results in the training set was drawn (Figure 5). The ROC curve illustrated that the AUC values of the model in patients with OA worse (0.814), maintain (0.700), and better

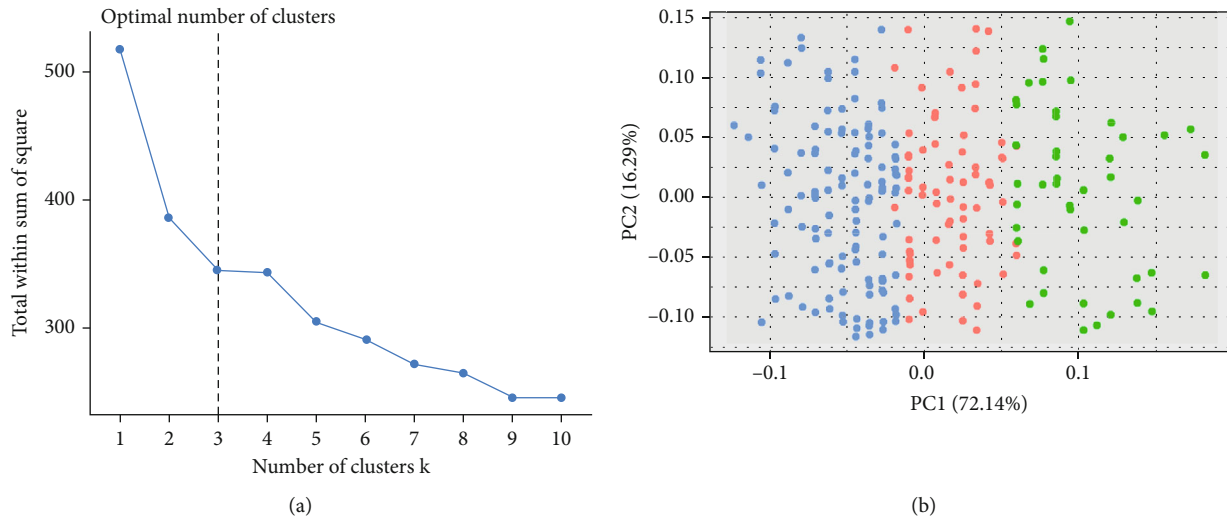


FIGURE 3: K-means clustering. (a) Total cluster sum of squares at different number of clusters (k). (b) Clustering results when the number of clusters $k = 3$. Note: PC: primary component.

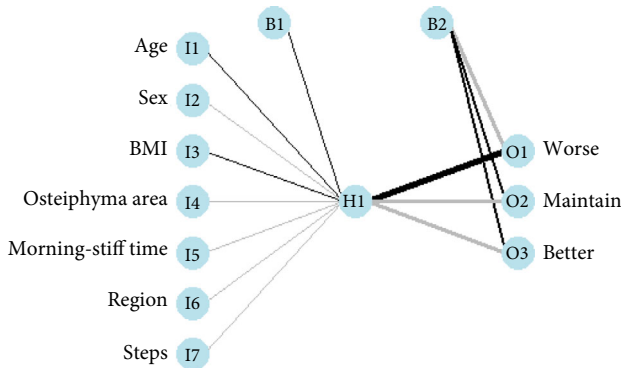


FIGURE 4: Neural network model.

(0.761) were higher than the multiclass AUC value (0.646), indicating high accuracy of the model and its potential for prediction of one-year postoperative outcome in patients with OA.

4. Discussion

In this study, patients with hip or knee OA after arthroscopic surgery were analyzed, and the prognostic influencing factors were screened by RFE. The RMSE value of the linear model reached the minimum value of 0.687 when 7 features were selected, and these 7 clinical features included sex, age, BMI, region, morning stiffness time, average step count, and osteophyte area. Subsequently, a 7-1-3 type BP-ANN model was established by using these 7 influencing factors as input variables to predict the prognosis one year after surgery.

Seven characteristics including sex, age, BMI, region, morning stiffness time, step count, and osteophyte area were found to have an impact on the prognosis of surgery through linear regression and RFE method, and their importance decreased in order. Abramoff et al. [15] reported that the incidence of OA increases with age and is the greatest risk

factor for OA. Additionally, BMI (overweight/obesity) is also associated with the progression of hip/knee OA in clinical practice [16–18]. Most studies indicated that women are more likely to develop OA than men [19–21]. Other risk factors for OA are race, morning stiffness, osteophytes, and limited motion [15, 22–24]. In this study, morning stiffness time, step count, and osteophyte area ranked relatively low, and their influence on OA needs further analysis. Knee joint and hip joint are commonly involved joints in OA [2]. Patients with this type of OA often have difficulties in walking when the condition is serious, which greatly reduced their quality of life [2]. As the number of involved joints increases, the walking ability of patients becomes worse [2]. Through one-way ANOVA or t -test, Kwon et al. [25] selected knee OA patients’ clinical features for constructing an artificial intelligence network model. This is different from the strategy of our study. During feature selection, the importance of each feature was first calculated by the linear regression model, and then the RFE algorithm was used to remove the least important features to obtain 7 clinical features, which were selected as the input layer to build the BP-ANN model. When 7 clinical features were screened out, the RMSE value reached the minimum. A related study believes that arthroscopic hip surgery is relatively safe because of its low incidence of postoperative complications [26], while the benefits of arthroscopic knee surgery remain controversial [27]. Unfortunately, we did not compare the hazard ratio of hip and knee surgeries in the present study, and therefore, we cannot discuss the clinical benefits of either procedure.

In the present study, BP-ANN network was used to predict the disease progression of OA patients one year after surgery. Firstly, the correlation matrix analysis was carried out on the clinical characteristics of patients by the Pearson correlation coefficient, and the results presented no redundancy among characteristics. Then, combined with linear regression and RFE to screen their importance, 7 feature

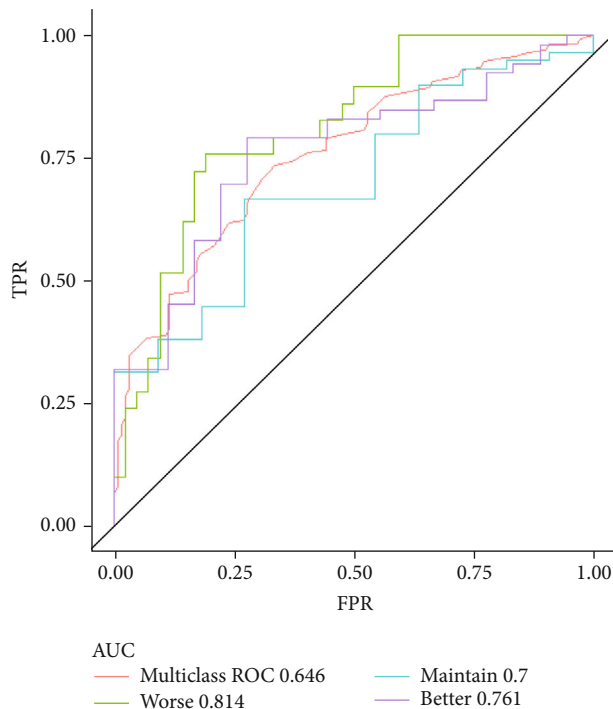


FIGURE 5: ROC curves of model clustering results. Note: FTR: false positive rate; TPR: true positive rate.

factors related to prognosis were obtained. K -means clustering revealed the number of classifications was 3. Subsequently, a “7-1-3” BP-ANN model was constructed based on these 7 features on the training set. Finally, the ROC curve verified high prediction accuracy of the model. Halilaj et al. [28] also carried out research related to the progress of OA using ANN model. Their study conducted cluster analysis on the clinical features of OA patients, and based on the clustering results, a Least Absolute Shrinkage and Selection (LASSO) model was developed to predict changes in joint spaces and WOMAC scores, with predictive AUC of 0.86 and 0.95, respectively. This is similar to the results of our study. Our findings may generate new insights into the construction of prognostic models for OA patients. In addition, by using machine learning algorithm, Sniderman et al. [29] predicted the outcomes (including hip disability and osteoarthritis outcome score (HOOS) 3 months after surgery) of patients undergoing hip replacement. Their study constructed a model for prediction by LASSO algorithm. Different from their study, our study integrated patients with OA of hip joint and knee joint and included three surgical methods: articular cartilage repair, arthroscopic debridement, and osteotomy. Based on the analysis results of our study, surgical method was not a significant clinical feature that affected the prognosis of OA patients. Currently, various researchers are attempting to apply artificial intelligence algorithms to multiple OA-related fields, such as the screening of OA clinical pathologic factors, prediction of postoperative efficacy, and OA pathological classification. However, developing more efficient and accurate algorithms and

models that can guide clinicians to make better decisions is still needed.

To conclude, this study found that sex, age, BMI, region, morning stiffness time, step count, and osteophyte area were prognostic influencing factors for arthroscopic surgery and established a neural network structured prognostic model. However, due to insufficient follow-up data, we only conducted a one-year postoperative study, and the features included in the study were limited. In the future, the follow-up data at multiple time points can be used to establish a model to predict the outcome of surgery at different times. In addition, the inclusion of more factors will improve the screening of prognostic factors. The achievements of these improvements will provide more favorable evidence for the treatment of OA.

Data Availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Ethical Approval

The research program was approved by the Medical Ethics Committee of The First People’s Hospital of Fuyang District.

Consent

No written consent has been obtained from the patients as there is no patient identifiable data included.

Conflicts of Interest

The authors report no proprietary or commercial interest in any product mentioned or concept discussed in this article.

References

- [1] M. Kloppenburg and F. Berenbaum, “Osteoarthritis year in review 2019: epidemiology and therapy,” *Osteoarthritis and Cartilage*, vol. 28, no. 3, pp. 242–248, 2020.
- [2] G. A. Hawker, “Osteoarthritis is a serious disease,” *Clinical and Experimental Rheumatology*, vol. 37, no. 120, pp. 3–6, 2019.
- [3] G. Li, G. Wang, M. Li et al., “Morin inhibits *Listeria monocytogenes* virulence in vivo and in vitro by targeting listeriolysin O and inflammation,” *BMC Microbiology*, vol. 20, no. 1, p. 112, 2020.
- [4] Joint Surgery Group B. o. O., “Guidelines for diagnosis and treatment of osteoarthritis,” *Chinese Journal of Orthopaedics*, vol. 38, pp. 705–715, 2018.
- [5] J. L. Kemp, N. J. Collins, M. Makdissi, A. G. Schache, Z. Machotka, and K. Crossley, “Hip arthroscopy for intra-articular pathology: a systematic review of outcomes with and without femoral osteoplasty,” *British Journal of Sports Medicine*, vol. 46, no. 9, pp. 632–643, 2012.
- [6] S. M. Tahami and S. M. Rad, “Outcome of ACL reconstruction and concomitant articular injury treatment,” *Arch Bone Jt Surg*, vol. 3, no. 4, pp. 260–263, 2015.
- [7] G. H. Stafford, J. R. Bunn, and R. N. Villar, “Arthroscopic repair of delaminated acetabular articular cartilage using fibrin

- adhesive,” *Results at one to three years*, vol. 21, pp. 744–750, 2011.
- [8] J. B. Moseley, K. O’Malley, N. J. Petersen et al., “A controlled trial of arthroscopic surgery for osteoarthritis of the knee,” *The New England Journal of Medicine*, vol. 347, no. 2, pp. 81–88, 2002.
- [9] A. Kirkley, T. B. Birmingham, R. B. Litchfield et al., “A randomized trial of arthroscopic surgery for osteoarthritis of the knee,” *The New England Journal of Medicine*, vol. 359, no. 11, pp. 1097–1107, 2008.
- [10] T. Janjic, S. Pereverzyev Jr., M. Hammerl et al., “Feed-forward neural networks using cerebral MR spectroscopy and DTI might predict neurodevelopmental outcome in preterm neonates,” *European Radiology*, vol. 30, no. 12, pp. 6441–6451, 2020.
- [11] Y. Shao, Z. Wang, N. Cao et al., “Prediction of 3-month treatment outcome of IgG4-DS based on BP artificial neural network,” *Oral Diseases*, vol. 27, no. 4, pp. 934–941, 2021.
- [12] N. Bellamy, W. W. Buchanan, C. H. Goldsmith, J. Campbell, and L. W. Stitt, “Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee,” *The Journal of Rheumatology*, vol. 15, pp. 1833–1840, 1988.
- [13] A. Filali, C. Jlassi, and N. Arous, “Recursive feature elimination with ensemble learning using SOM variants,” *International Journal of Computational Intelligence and Applications*, vol. 16, article 1750004, 2017.
- [14] R. Hecht-Nielsen, “Theory of the backpropagation neural network,” in *International 1989 Joint Conference on Neural Networks*, Washington, pp. 593–605, USA, 1992.
- [15] B. Abramoff and F. E. Caldera, “Osteoarthritis: Pathology, Diagnosis, and Treatment Options,” *Medical Clinics of North America*, vol. 104, no. 2, pp. 293–311, 2020.
- [16] S. M. Bierma-Zeinstra and B. W. Koes, “Risk factors and prognostic factors of hip and knee osteoarthritis,” *Nature Clinical Practice. Rheumatology*, vol. 3, no. 2, pp. 78–85, 2007.
- [17] A. N. Bastick, J. Runhaar, J. N. Belo, and S. M. Bierma-Zeinstra, “Prognostic factors for progression of clinical osteoarthritis of the knee: a systematic review of observational studies,” *Arthritis Research & Therapy*, vol. 17, no. 1, p. 152, 2015.
- [18] J. Sellam and F. Berenbaum, “Is osteoarthritis a metabolic disease?,” *Joint, Bone, Spine*, vol. 80, no. 6, pp. 568–573, 2013.
- [19] S. Holden, J. Kasza, M. Winters, M. van Middelkoop, Adolescent Knee Health Group, and M. S. Rathleff, “Prognostic factors for adolescent knee pain: an individual participant data meta-analysis of 1281 patients,” *Pain*, vol. 162, no. 6, pp. 1597–1607, 2021.
- [20] J. E. Collins, J. N. Katz, E. E. Dervan, and E. Losina, “Trajectories and risk profiles of pain in persons with radiographic, symptomatic knee osteoarthritis: data from the osteoarthritis initiative,” *Osteoarthritis and Cartilage*, vol. 22, no. 5, pp. 622–630, 2014.
- [21] S. Muraki, T. Akune, H. Oka et al., “Incidence and risk factors for radiographic knee osteoarthritis and knee pain in Japanese men and women: a longitudinal population-based cohort study,” *Arthritis and Rheumatism*, vol. 64, no. 5, pp. 1447–1456, 2012.
- [22] J. F. Holla, M. P. M. Steultjens, L. D. Roorda, M. W. Heymans, S. ten Wolde, and J. Dekker, “Prognostic factors for the two-year course of activity limitations in early osteoarthritis of the hip and/or knee,” *Arthritis Care Res (Hoboken)*, vol. 62, no. 10, pp. 1415–1425, 2010.
- [23] J. Dekker, G. M. van Dijk, and C. Veenhof, “Risk factors for functional decline in osteoarthritis of the hip or knee,” *Current Opinion in Rheumatology*, vol. 21, no. 5, pp. 520–524, 2009.
- [24] Y. Liem, A. Judge, Y. Li, and M. Sharif, “Biochemical, clinical, demographic and imaging biomarkers for disease progression in knee osteoarthritis,” *Biomarkers in Medicine*, vol. 16, no. 8, pp. 633–645, 2022.
- [25] S. B. Kwon, Y. Ku, H. S. Han, M. C. Lee, H. C. Kim, and D. H. Ro, “A machine learning-based diagnostic model associated with knee osteoarthritis severity,” *Scientific Reports*, vol. 10, no. 1, p. 15743, 2020.
- [26] M. Kowalczyk, M. Bhandari, F. Farrokhyar et al., “Complications following hip arthroscopy: a systematic review and meta-analysis,” *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 21, no. 7, pp. 1669–1675, 2013.
- [27] S. M. Hussain, D. W. Neilly, S. Baliga, S. Patil, and R. Meek, “Knee osteoarthritis: a review of management options,” *Scottish Medical Journal*, vol. 61, no. 1, pp. 7–16, 2016.
- [28] E. Halilaj, Y. Le, J. L. Hicks, T. J. Hastie, and S. L. Delp, “Modeling and predicting osteoarthritis progression: data from the osteoarthritis initiative,” *Osteoarthritis and Cartilage*, vol. 26, no. 12, pp. 1643–1650, 2018.
- [29] J. Sniderman, R. B. Stark, C. E. Schwartz, H. Imam, J. A. Finkelstein, and M. T. Nousiainen, “Patient factors that matter in predicting hip arthroplasty outcomes: a machine-learning approach,” *The Journal of Arthroplasty*, vol. 36, no. 6, pp. 2024–2032, 2021.

Research Article

LINC00887 Fosters Development of Clear Cell Renal Cell Carcinoma via Inhibiting CD8+ T Cell Immune Infiltration

Jinfeng Wu ^{1,2}, Rongcheng Lin ², Lei Zhang ¹, Yongbao Wei ², Ruochen Zhang ²,
Wanghai Cai ² and Weilie Hu ^{1,3}

¹The First School of Clinical Medicine, Southern Medical University, China

²Department of Urology, Shengli Clinical Medical College of Fujian Medical University, Fujian Provincial Hospital, China

³Department of Urology, Guangdong Hydropower Hospital, China

Correspondence should be addressed to Weilie Hu; huweilie_hwl@163.com

Received 10 May 2022; Accepted 27 June 2022; Published 25 August 2022

Academic Editor: Tao Huang

Copyright © 2022 Jinfeng Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. lncRNAs affect adaptive and innate immunity of cancer via mediating functional states of immune cells, genes, and pathways. Nonetheless, little is known about the molecular mechanism of lncRNA-mediated CD8+ T cell immune infiltration in progression of clear cell renal cell carcinoma (ccRCC). We designed this work to investigate the role of LINC00887 in regulating CD8+ T cell immune infiltration in ccRCC. **Methods.** Correlation between LINC00887 and immune factors and the expression level of LINC00887 in ccRCC were analyzed by bioinformatics methods (TCGA-KIRC database, “edgeR” package, “clusterProfiler” package, and “CIBERSORT” package). LINC00887 expression in ccRCC was examined via RT-qPCR. The cytotoxic capacity of CD8+ T cells was evaluated by the lactate dehydrogenase assay. The apoptotic ability of CD8+ T cells was measured by flow cytometry. The chemotactic ability of CD8+ T cells was revealed by chemotaxis assay. CXCR3, CXCL9, and CXCL10 levels were assessed by RT-qPCR. **Results.** As suggested by bioinformatics analysis, LINC00887 was markedly upregulated in ccRCC patients and associated with expression of immune-suppression molecule, thereby abating the immune infiltration level of CD8+ cells in tumor tissue. As revealed by cellular assay, LINC00887 was upregulated in ccRCC cells, and knockdown of LINC00887 resulted in a decreased PD-L1 expression, increased CD8+ T cell toxicity, decreased apoptotic levels, and enhanced chemotaxis. Moreover, we found that LINC00887 exhibited inhibitory effect on immune infiltration of CD8+ cells in clinical tissues. **Conclusions.** The results of this study suggested that LINC00887 promoted ccRCC progression by inhibiting immune infiltration of CD8+ T cells, providing new insights into pathogenesis of ccRCC and suggesting LINC00887 being a promising immunotherapy target for ccRCC.

1. Introduction

Renal cell carcinoma (RCC) ranks third in common cancers of urinary system [1]. According to histopathological classification, RCC has three main subtypes: papillary renal cell carcinoma, clear cell renal cell carcinoma (ccRCC), and chromophobe cell renal carcinoma. Among them, ccRCC is the most frequent histological subtype, accounting for 80-90% [2]. Clinically, due to no evident symptoms at early stage and insufficient specific diagnostic markers, 20-30% of

ccRCC patients have been in advanced-stage and developed metastasis when diagnosed [3]. Metastatic ccRCC is not sensitive to radiotherapy and chemotherapy and that new targeted drugs have poor effects on patients with this type of ccRCC. Therefore, it is urgent to develop appropriate therapies to improve the clinical efficacy [4]. It is encouraging to note that recent use of immune checkpoint therapy is effective in patients with metastatic ccRCC [4]. For example, McDermott et al. [5] found pembrolizumab as a first-line therapy for advanced-stage ccRCC patients with an objective

response rate of 36.4%, a median PFS of 7.1 months, and an OS rate of 88.2% and 70.8% at 12 and 24 months. Results from CheckMate 214 further established that the combination of ipilimumab and nivolumab was associated with higher response rates (RR) (39% in the ITT population), complete response rates (8%) and duration of response compared to sunitinib [6]. All above indicated that immunotherapy has a good effect on RCC patients. Although the above studies showed that immunotherapy has a good therapeutic effect on patients, the therapeutic effect of immunotherapy varies in different patients. Currently, it is believed that the effect of tumor immunotherapy is greatly affected by immune escape of tumor cells [7]. CD8+T cells infiltrated by tumor are the main effect or cells mediating anti-tumor immune response [8]. However, the immune infiltration of CD8+T cells is usually negatively regulated by various conditions in tumor microenvironment, so that tumor cells can evade immune surveillance [9]. In conclusion, finding the key factors affecting CD8+ T cell infiltration is of great value for optimizing treatment of ccRCC patients.

Long noncoding RNAs (lncRNAs) are involved in regulating cellular biological functions like energy metabolism changes, angiogenesis, proliferation, and migration in tumors [10]. It has been previously reported that lncRNAs participate in innate and adaptive immune responses through a complex regulatory network [11]. Huang et al.'s study [12] reported that lncRNA NKILA sensitized T cells to activation-induced cell death, thus promoting tumor immune escape. lncRNA-MM2P promotes macrophage M2 polarization, which in turn promotes immune escape of tumor cells [13]. In lung cancer cells, overexpressed lncRNA NEAT1 inhibits T cell infiltration in tumor tissues by downregulating chemokines such as CXCL10 and CCL5 [14]. These results indicated the potential of lncRNAs in tumor immune regulation, but there are still many tumor immune-related lncRNAs that have not been explored. Therefore, further exploration of immune-related genes regulating CD8+ T cells in ccRCC can provide new targets for the immunotherapy in ccRCC. We found that LINC00887 is a key factor affecting tumor immune escape, and it may play a role as a target of immunotherapy.

LINC00887 (NR_024480.1), with a length of 3,021 nucleotides, is composed of three exons [15]. Previous reports have shown that LINC00887 can affect the development of different cancers. For example, LINC00887 expression is upregulated in glioma, and knocking out LINC00887 can inhibit the proliferation of cancer cells [16]. LINC00887 expression is also upregulated in nasopharyngeal carcinoma, and overexpression of LINC00887 can significantly accelerate proliferative ability of nasopharyngeal carcinoma cells [17]. It was also found that LINC00887 was remarkably overexpressed in RCC and significantly facilitated RCC cell proliferation in vitro [18]. Although these studies indicate the key role of LINC00887 in cancer development, no literature has explored LINC00887 and its role in ccRCC immunity. Thus, we aimed to elucidate the impacts of LINC00887 on ccRCC via regulating CD8+ T cell immune infiltration.

Herein, bioinformatics analysis identified that LINC00887 had a potential in inhibiting immune invasion of CD8+ T cells in tumor tissues, and experiments confirmed that LINC00887

knockdown in ccRCC cells could significantly increase chemotaxis and cytotoxicity of T cells to tumor cells. This study initially revealed the crucial role of LINC00887 in immunoregulation of CD8+ T cells, providing a new reference for the study of immune escape mechanism of ccRCC.

2. Materials and Methods

2.1. Collection of Patient Samples. 10 pairs of ccRCC tissues and corresponding normal paracancer tissues were collected from patients receiving ccRCC surgery in our hospital. Inclusion criteria are as follows: ccRCC was confirmed by histopathological section, and no other treatment was received before surgery. Patients were excluded if they complicated with other tumors or other organic lesions [16]. Immediately after excision, the tissues frozen in liquid nitrogen were prepared for subsequent usage. The protocol involving human sample collection was approved by the Shengli Clinical Medical College of Fujian Medical University, Fujian Provincial Hospital Ethics Committee. Written informed consent was acquired from patients.

2.2. Bioinformatics Analysis. The lncRNA differential analysis was carried out using the “edgeR” package [19] ($\log_{2}FC > 2$, $FDR < 0.05$) (normal: 72, tumor: 539). Target genes were identified through literature review. Then, the target genes were grouped and subjected to differential analysis with “edgeR” package ($\log_{2}FC > 2$, $FDR < 0.05$). The “clusterProfiler” package for KEGG and GO analyses were carried out for differentially expressed genes (DEGs). Based on mRNA (FPKM) data, the infiltration of 22 kinds of immune cells in tumor samples were scored using the “CIBERSORT” package. The cells with all infiltration scores being 0 were filtered out, and samples with $P > 0.05$ were also filtered out. The correlation between target genes and immune cell infiltration scores was analyzed. The difference of immune cell infiltration score was analyzed by grouping target genes. Pearson's analysis was performed to determine correlation between LINC00887 expression and mRNA expression of immunosuppressive molecules PD-L1 (CD274) and PD-1 (PDCD1). Then, according to SNV data of samples, the mutation landscape oncoplot of LINC00887 high- and low-expression groups in ccRCC tumor samples was drawn using the R package GenVisR [20].

2.3. Cell Culture. Human normal renal cell line HK-2 and ccRCC cell lines Caki-1, 786-O, and A498 were provided by the American Type Culture Collection (ATCC, USA). A498 and Caki-1 were cultivated in Dulbecco's modified Eagle medium (DMEM) (Sigma, USA). 786-O was cultured in RPMI-1640 medium (Sigma, USA). HK-2 was incubated in DMEM/F12 medium (Sigma, USA). All the mediums contained 1% penicillin/streptomycin and 10% fetal bovine serum (FBS) at 37°C with 5% CO₂.

2.4. Cell Transfection. The sh-NC and sh-LINC00887 were designed by GenePharma (China). A498/786-O experimental cells (1.2×10^5 cells/well) were harvested and seeded onto 6-well plates. And then the cells were randomly divided into two groups which transfected with sh-NC or sh-LINC00887,

respectively. The cells were transfected using Lipofectamine 2000™ (Invitrogen, USA) following the manufacturer's suggestions, with transfection efficiency ensured by qRT-PCR after transfection for 48 h.

2.5. RT-qPCR. Total RNA was separated with TRIzol reagent (Invitrogen, USA). Reverse transcription was carried out by PrimeScript RT Reagent Kit, and RT-qPCR analysis was performed with SYBR Green PCR Master Mix (Takara, Japan) on the CFX96 assay system. The reaction conditions were 40 cycles of 96°C for 1 min, 96°C for 10 s, 60°C for 5 s, and 72°C for 30 s. Primers are as follows: LINC00887 forward: 5'-GAGGCTGAAATTGTCTGAAGTC-3' and reverse: 5'-ATTCGCAAGAGGGTGACAG-3'; CXCR3 forward: 5'-CCACCTAGCTGTAGCAGACAC-3' and reverse: 5'-AGGGCTCCTGCGTAGAAGTT-3'; CXCL9 forward: 5'-CCAGTAGTGAGAAAGGGTCCG-3' and reverse: 5'-AGGGCTTGGGGCAAATTGTT-3'; CXCL10 forward: 5'-GTGGCATTCAAGGAGTACCTC-3' and reverse: 5'-TGATGGCCTTCGATTCTGGATT-3'; and GAPDH forward: 5'-ACATCGCTCAGACACCATG-3' and reverse: 5'-TGTAGTTGAGGTCAATGAAGGG-3'.

2.6. Immunohistochemistry (IHC) Assay. Paraffin sections (4 μm) were prepared from human tissues fixed in 10% neutral-buffered formalin. To detect CD8, tissue sections were dewaxed in xylene, rehydrated in diluted ethanol, blocked with 0.3% H₂O₂, then blocked with normal goat serum, and incubated with primary rabbit anti-CD8 (ab4055, Abcam, UK) at 4°C overnight. The tissues were incubated at room temperature with biotinylated goat anti-rabbit IgG (BP-9100, Vector Labs, USA) for 30 min and then treated with Vectastain ABC-HRP solution (Vector Labs, USA) for 30 min. Diaminobenzidine (DAB) (Sigma, USA)-was utilized for development.

2.7. Lactate Dehydrogenase (LDH) Cytotoxicity Test. CD8+ T cells were tested for cytotoxicity using CytoTox LDH cytotoxicity Assay Kit (Promega, USA). In short, the isolated effector CD8+ T cells were cocultivated with ccRCC cells in different proportions at 37°C for 24 h. The conditioned medium was transferred to 96-well plates three times and cultured in reaction mixture for 30 min at room temperature [21]. After adding stop buffer, we measured OD490 value. The formula for cytotoxicity calculation of CD8+ T cells was as follows: cytotoxicity (%) = [(LDH activity of experimental group (OD₄₉₀))/(maximum LDH activity (OD₄₉₀))] × 100%.

2.8. Apoptosis Detected by Flow Cytometry. T cells were first cocultured with tumor cells in a 5:1 ratio, followed by the detection of apoptotic cells using the Annexin V-FITC apoptosis detection Kit (Thermo Fisher Scientific, USA). Briefly, the treated cells were plated onto 24-well plates and cultivated for about 24 h. They were then stained for 15 min with Annexin V-FITC and PI solution. FACSCalibur analysis was used to determine the percentage of apoptotic cells.

2.9. Transwell Assay. Chemotactic migration of CD8+ T cells was assessed in a 24-well plate with a 5 μm aperture polycar-

bonate filter (Corning, USA). First, ccRCC tumor tissue was incubated in DMEM containing 10% FBS for 48 h. Next, 600 μL of tumor supernatant was introduced to the lower chamber. Anti-CCL5 (1 μg/mL) and/or anti-CXCL10 (5 μg/mL) was introduced to neutralize antibodies. DMEM containing 10% FBS was the control. Purified CD8+ T cells (purity > 90%) from TILs were counted. Next, 5 × 10⁵ CD8+ T cells were supplemented to the upper chamber and cultivated for 2 h at 37°C with 5% CO₂. Cells in the lower chamber were calculated by an automatic cell counter [22].

2.10. Western Blot. Total proteins were extracted from cells and subjected to a bicinchoninic acid (BCA) Protein Assay Kit (Thermo Fisher Scientific, USA). Protein samples (40 μg/lane) were subsequently analyzed by Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis. Thereafter, the proteins were transferred to a polyvinylidene fluoride (Millipore, USA) membrane. The membrane was subsequently blocked in 5% skimmed milk for 2 h at room temperature and then kept in specific primary antibodies overnight (4°C). Main antibodies included rabbit anti-PD-L1 antibody (ab205921, Abcam, UK), rabbit anti-GAPDH antibody (ab9485, Abcam, UK). The following day, it was incubated with secondary antibody conjugated to HRP (HRP-goat anti-rabbit IgG, ab205718, Abcam, UK) for 1 h. Finally, immunoreactivity was examined using an ECL kit (Thermo Fisher Scientific, USA).

2.11. Data Analysis. Statistical analysis was processed with the SPSS 13.0 software (SPSS Inc., USA). Quantitative data from 3 independent assays were expressed as mean ± standard deviation. Student's *t* test compared the difference between two groups. One-way analysis of variance compared differences among three or more groups. *P* value less than 0.05 represents significant difference.

3. Results

3.1. LINC00887 Is Remarkably Highly Expressed in ccRCC. First, LINC00887 expression was analyzed in ccRCC using TCGA-KIRC database. LINC00887 level was dramatically upregulated in ccRCC tissues versus normal tissues (Figure 1(a)). Subsequently, LINC00887 expression in ccRCC cell lines (A498, Caki-1, and 786-O) and HK-2 cell line was revealed by RT-qPCR, indicating that LINC00887 was significantly upregulated in ccRCC cells (Figure 1(b)). The above experiments demonstrated a high level of LINC00887 in ccRCC. LINC00887 level was relatively high in A498 and 786-O cells, so the two cell lines were selected for subsequent assays.

3.2. LINC00887 Affects CD8+ T Cell-Mediated Immune Responses. Through differential expression analysis, 577 DEGs were obtained. GO and KEGG analyses demonstrated that these genes were mainly enriched in complement and coagulation cascades, acute inflammatory response, and other signaling pathways (Figures 2(a) and 2(b)). This suggested that LINC00887 may have a potential link with immunity. Then, CIBERSORT analyzed immune cell infiltration of tumor tissues from the samples. After the unqualified samples and immune cells were filtered out, the relative infiltration levels

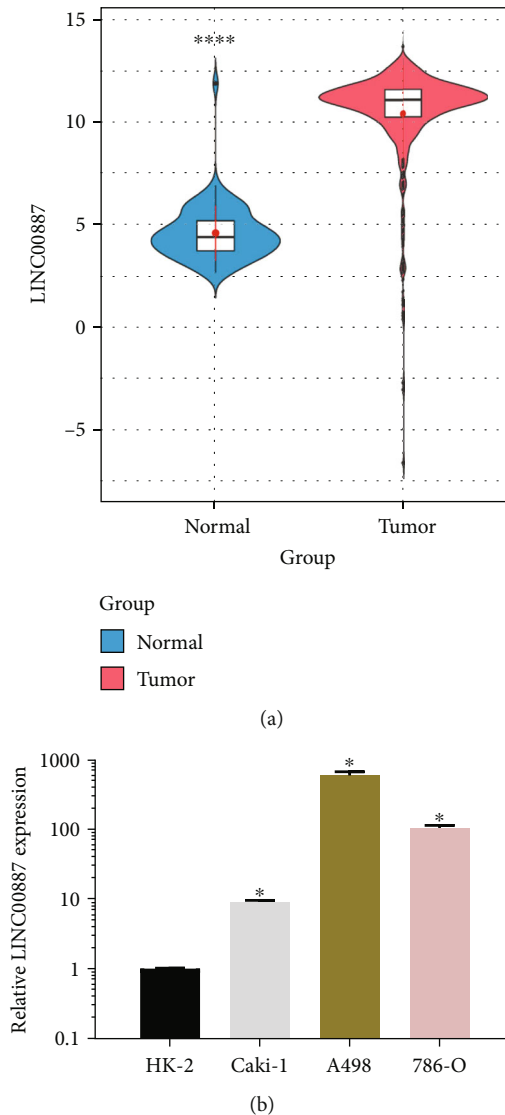
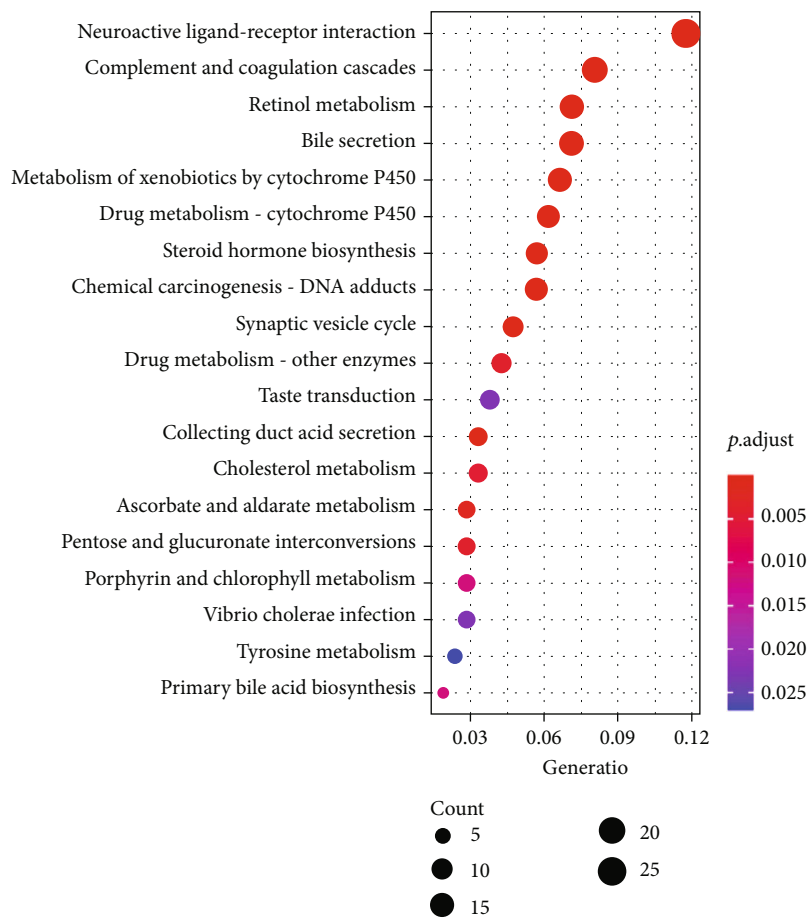


FIGURE 1: LINC00887 is highly expressed in ccRCC. (a) Violin plot of LINC00887 expression in ccRCC (red) and normal (blue) tissues. (b) Expression of LINC00887 in human renal tubular epithelial cell line (HK-2) and ccRCC cell lines (A-498, 786-O, and Caki-1), * $P < 0.05$.

of 21 kinds of immune cells in 414 ccRCC tumor tissues were obtained (Figure 2(c)). Then, the difference in the abundance of immune cell infiltration was analyzed between LINC00887 high- and low-expression groups. The results stated that abundance of CD8+ T cell immune infiltration was markedly reduced in LINC00887 high-expression group (Figure 2(d)). We then analyzed the correlation between the LINC00887 level and infiltration abundance of 21 kinds of immune cells. As revealed above, LINC00887 level was notably negatively correlated with abundance of CD8+ T cell infiltration (Figure 2(e)). Therefore, we hypothesized that LINC00887 may inhibit immune infiltration of CD8+ T cells in tumor tissues.

3.3. LINC00887 Is Associated with the Expression of Immunosuppressive Molecules. Since the above analysis revealed the close relation between LINC00887 and immune cell infiltration in ccRCC, we further explored the correla-

tion between LINC00887 and immunosuppressive molecules. LINC00887 was prominently positively correlated with PD-1 expression (Figure 3(a)) while negatively correlated with PD-L1 expression (Figure 3(b)). Subsequently, expression differences of checkpoint molecules such as CTLA4, LAG3, TGFBI, NOS3, and IDO1 between LINC00887 high- and low-expression groups were analyzed. As a result, CTLA4 and LAG3 were highly expressed in the LINC00887 low-expression group (Figures 3(c) and 3(d)), while TGFBI, NOS3, and IDO1 immunosuppressive genes were highly expressed in the LINC00887 high-expression group (Figures 3(e)–3(g)). SNP analysis results showed that mutations in immune checkpoint molecules such as PBRM1 were more frequent in the group with high LINC00887 expression (Figures 3(h) and 3(i)). In summary, bioinformatics studies revealed the association between LINC00887 and the immunosuppression of ccRCC.



(a)

FIGURE 2: Continued.

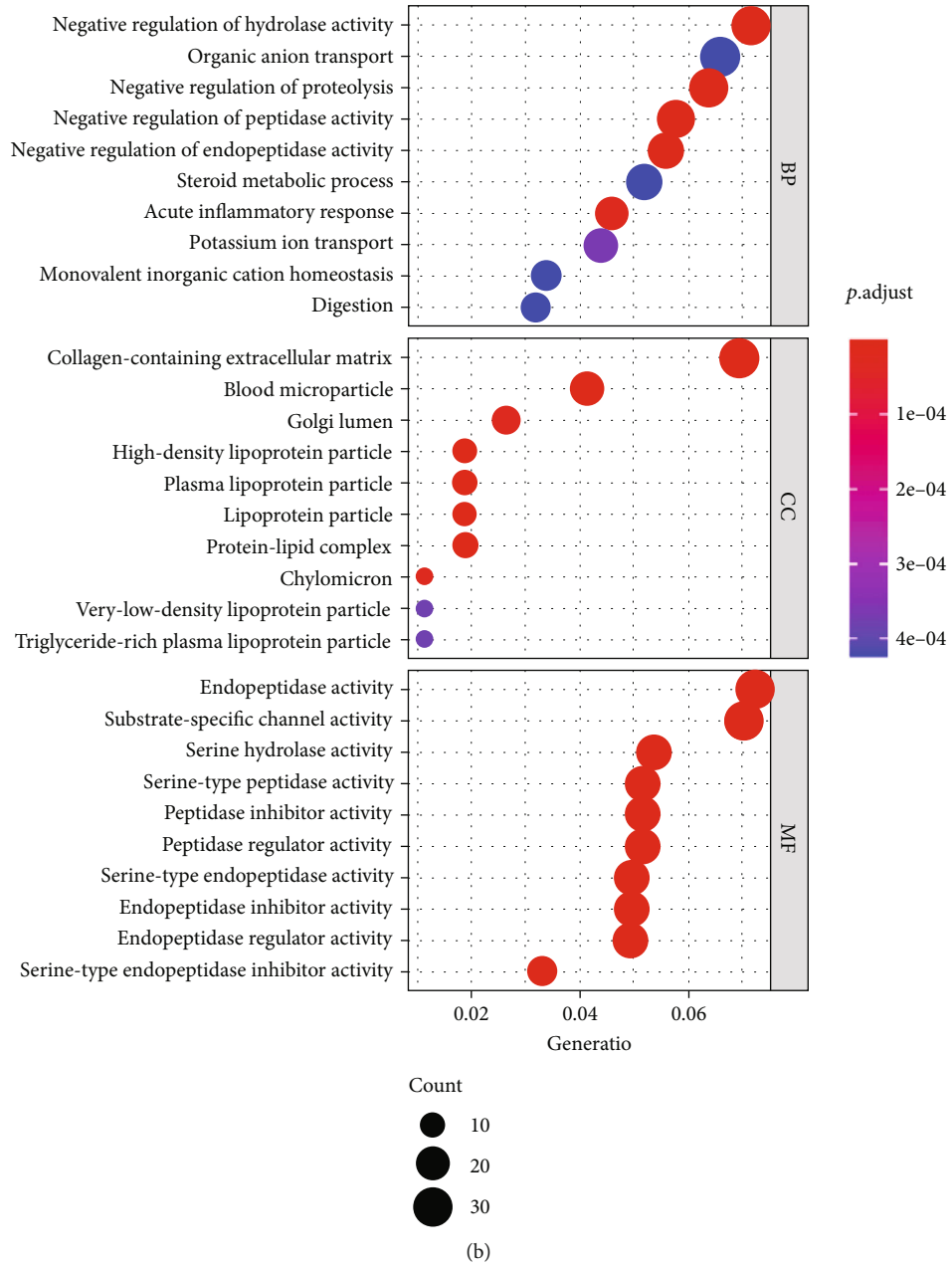
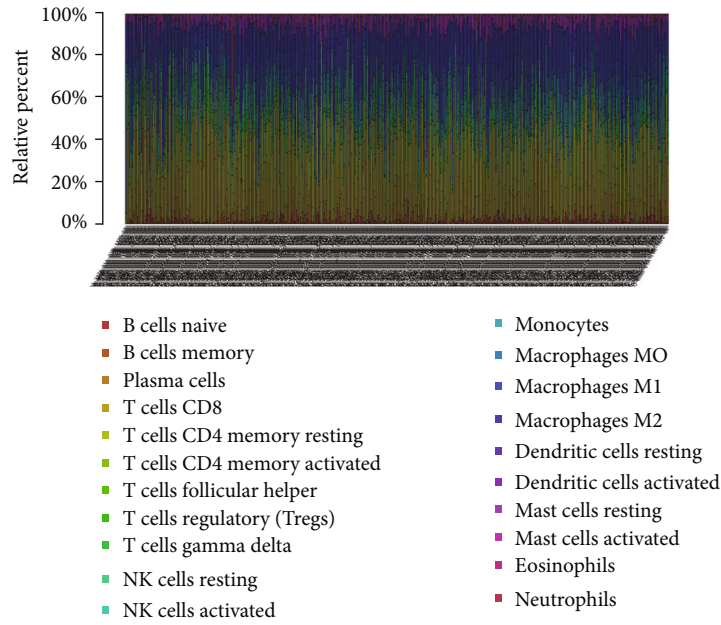
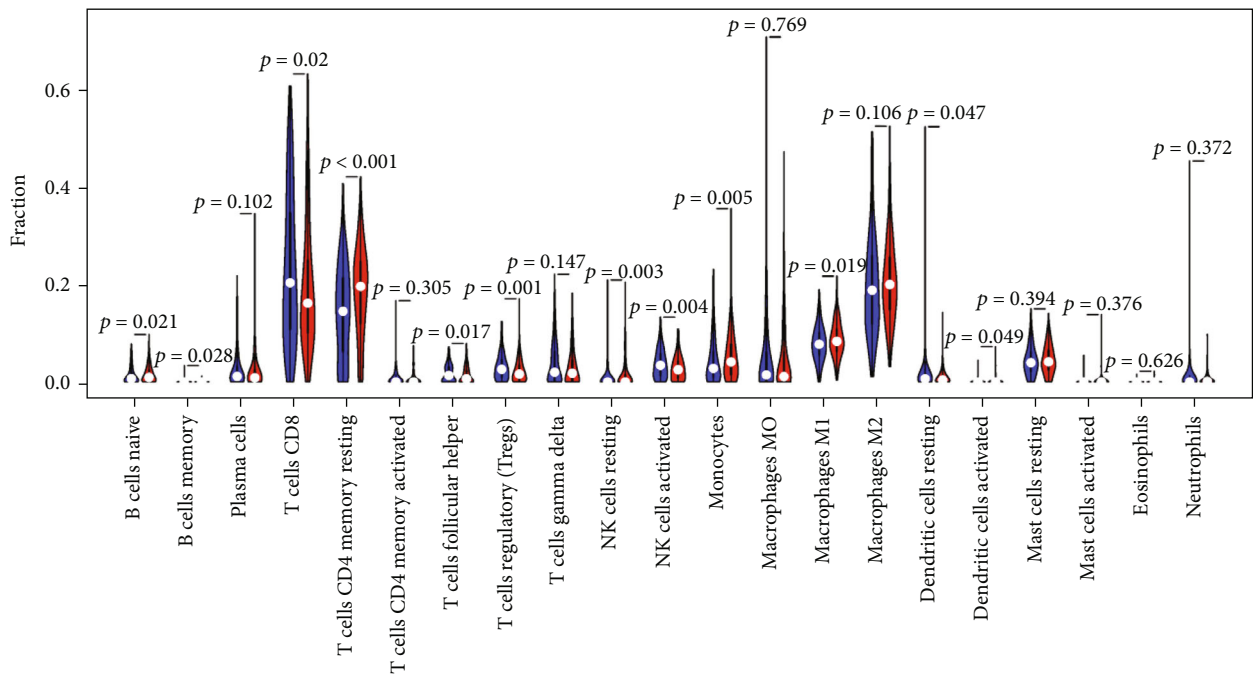


FIGURE 2: Continued.



(c)



(d)

FIGURE 2: Continued.

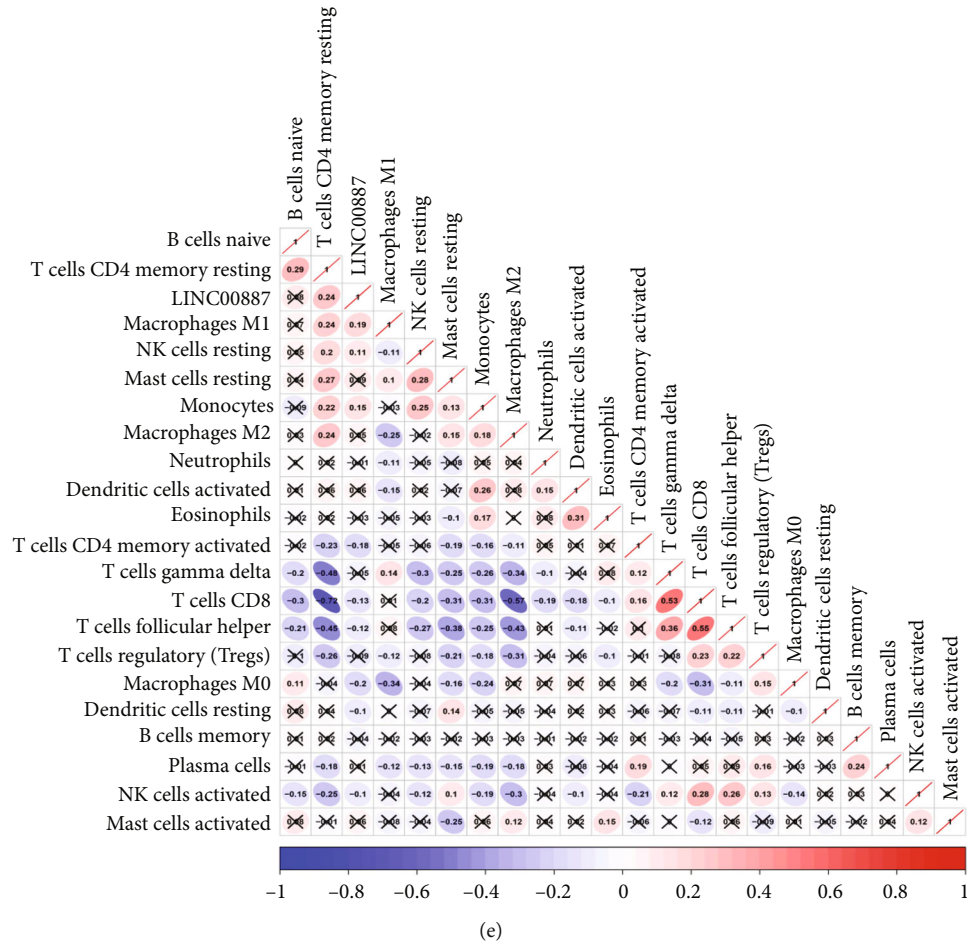


FIGURE 2: Bioinformatics prediction of correlation between LINC00887 expression and CD8+ T cell immune infiltration. (a) KEGG enrichment analysis of DEGs between LINC00887 high- and low-expression groups. (b) GO enrichment analysis of DEGs between LINC00887 high- and low-expression groups. (c) Relative infiltration abundance of 21 kinds of immune cells in ccRCC samples. (d) Differences in immune cell infiltration between LINC00887 high- and low-expression groups. (e) Correlation analysis between LINC00887 expression and immune cell infiltration abundance, * $P < 0.05$.

3.4. LINC00887 Inhibits the Cell Killing and Chemotactic Abilities of Immune Cells. Bioinformatics analysis exhibited that LINC00887 level was associated with PD-L1 level, combined with the report that the expression of PD-L1 may affect the functions of CD8+ T cells [23], we detected the protein level of PD-L1 in tumor cells, manifesting that knockdown LINC00887 could inhibit the expression of PD-L1 (Figure 4(a)). To further verify the bioinformatics results, CD8+ T cells were isolated using EasySep™ Direct Human CD8+ T Cell Isolation Kit. 786-O and A498 cells were stably transfected with sh-NC and sh-LINC00887, and tumor cells were cocultured with CD8+ T cells. LDH assay was used to measure cytotoxicity of CD8+ T cells to the above cells. As a result, sh-LINC00887 could significantly increase the cell killing ability of CD8+ T cells to ccRCC cells (Figure 4(b)). Then, apoptosis of CD8+ T cells was detected. The sh-LINC00887 treatment could significantly reduce apoptosis of CD8+ T cells (Figure 4(c)). Chemotactic experiments showed that the chemotactic ability of CD8+ T cells was enhanced after sh-LINC00887 treatment (Figure 4(d)). RT-qPCR demonstrated that sh-

LINC00887 could promote expression of chemokines CXCL9, CXCL10, and CXCR3 (Figure 4(e)). In summary, cell experiments confirmed that LINC00887 could indeed promote the development of ccRCC by inhibiting immune response mediated by CD8+ T cells.

3.5. LINC00887 Can Repress CD8+ T Cell Immune Infiltration in Clinical Tissues. To verify the impact of LINC00887 on immune infiltration of CD8+ T cells, 10 pairs of ccRCC tissues and corresponding adjacent tissues were gathered and subjected to RT-qPCR for LINC00887 expression quantitation. The results suggested that evident upregulation of LINC00887 in ccRCC tissues compared with paracancerous tissues (Figure 5(a)). Next, patients were assigned into high and low LINC00887 expression groups. Then, IHC assay determined CD8+ cell immune infiltration level in ccRCC tissues. CD8+ cells had a lower infiltration level in ccRCC tissues with high LINC00887 expression (Figure 5(b)). This suggested that LINC00887 showed an inhibitory effect on CD8+ cell immune infiltration in clinical tissues.

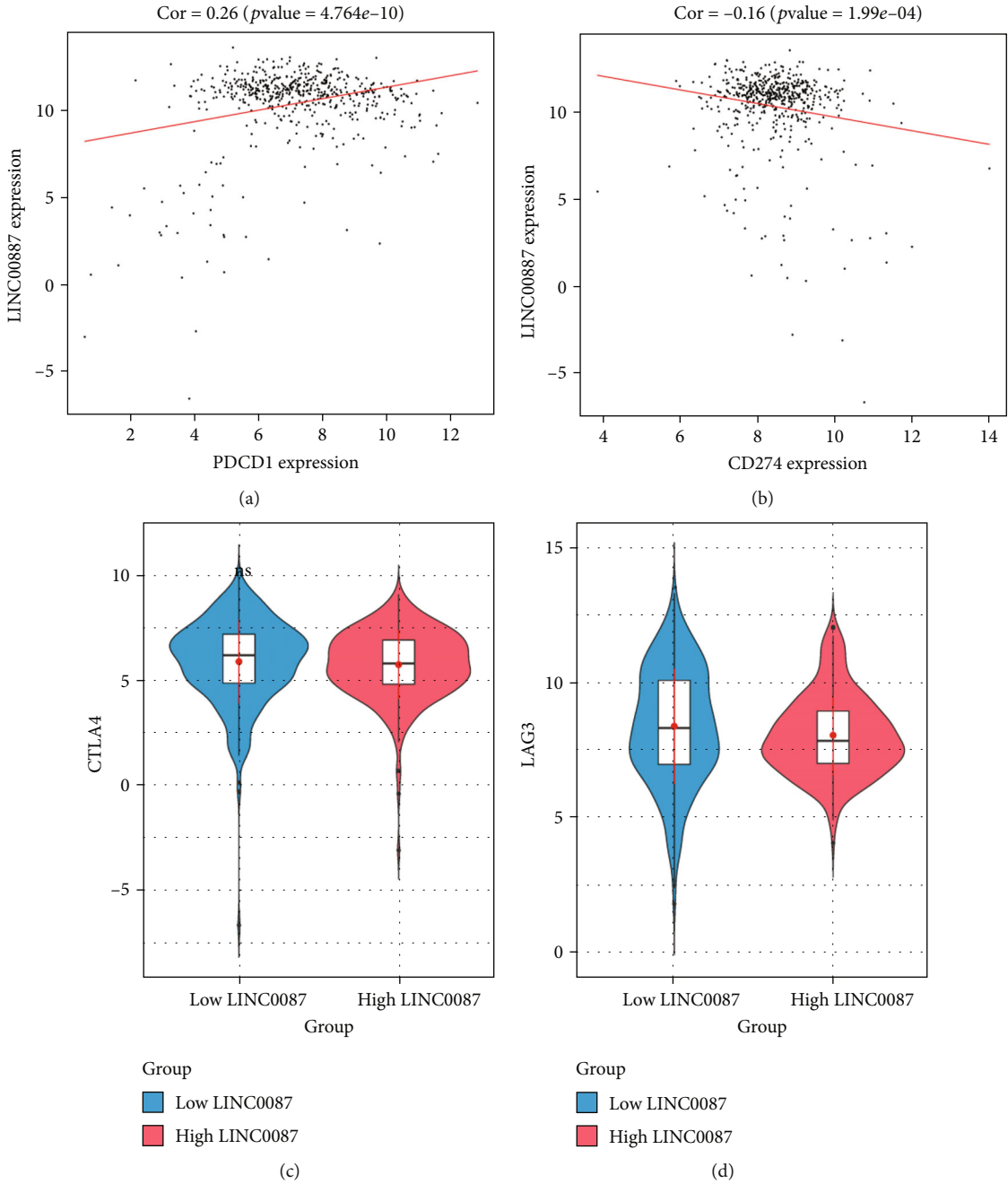


FIGURE 3: Continued.

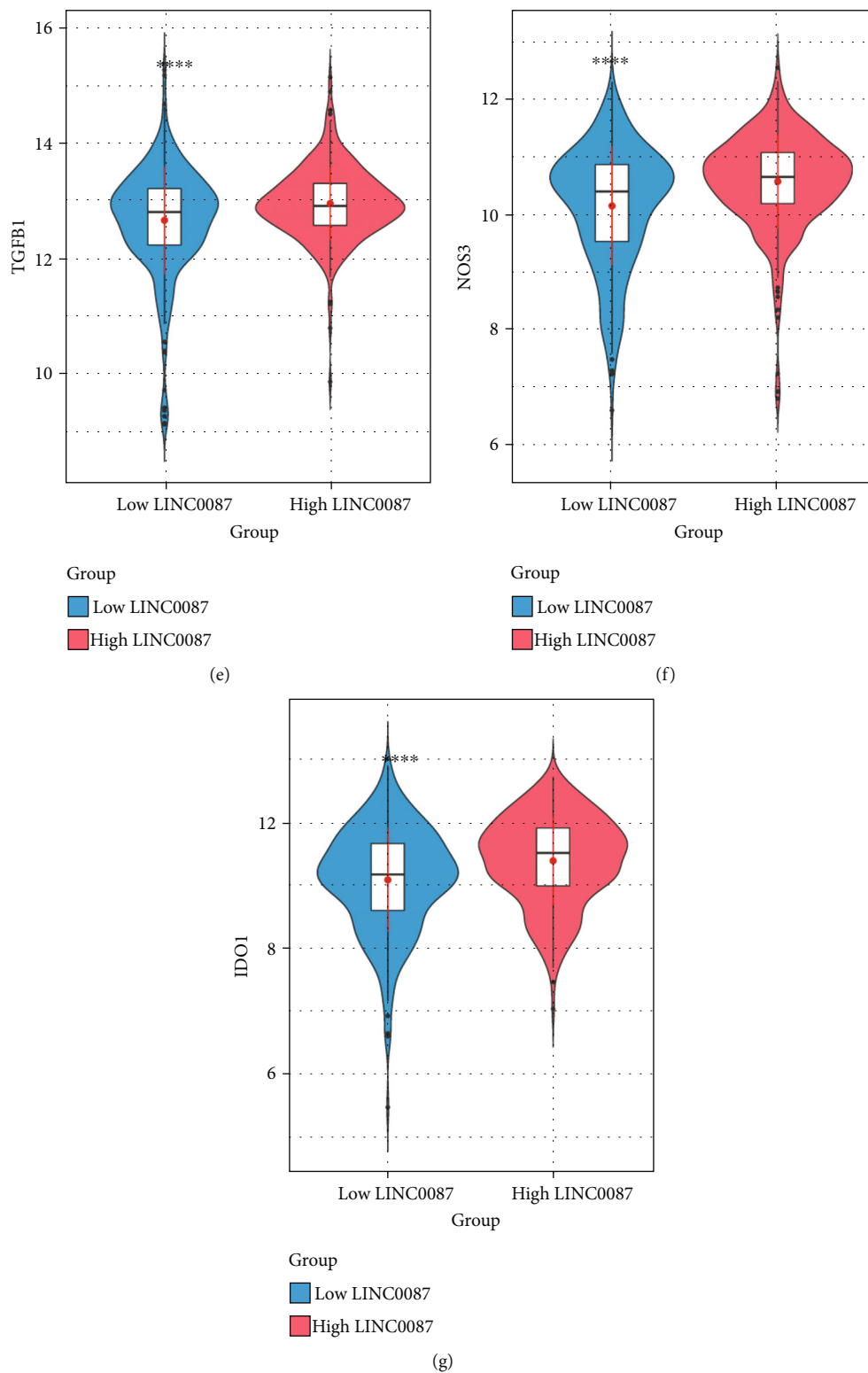
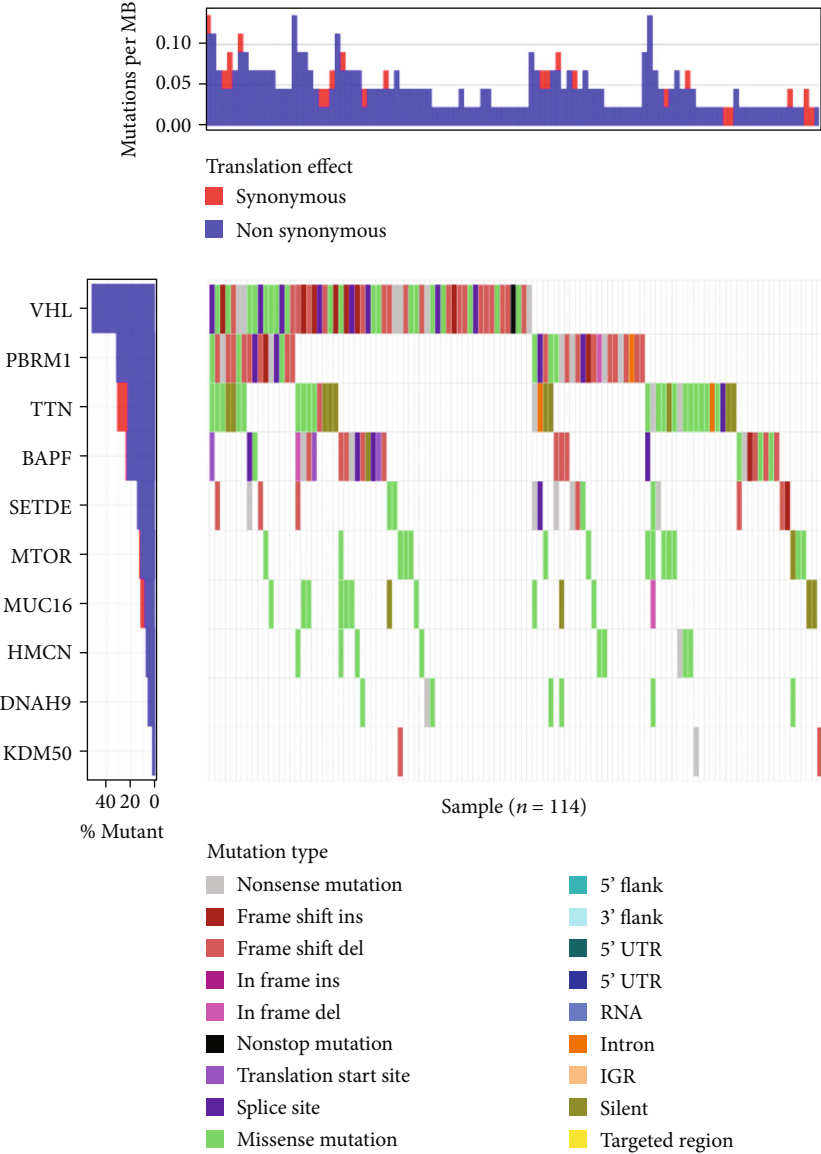
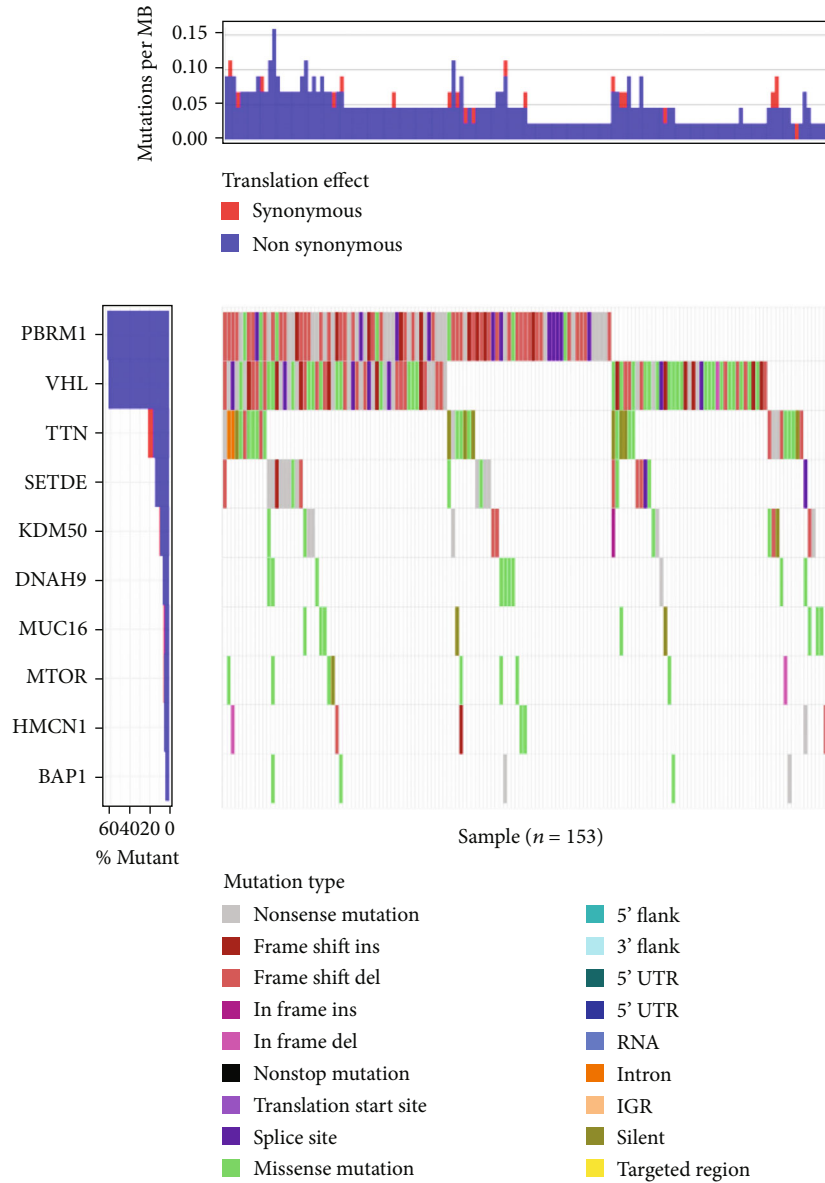


FIGURE 3: Continued.



(h)

FIGURE 3: Continued.



(i)

FIGURE 3: The correlation of LINC00887 with immunosuppressive molecules and SNP. (a, b) The correlation of LINC00887 with the expression of PD-1 and PD-L1. (c-g) Differences in the expression levels of immunosuppressive molecules CTLA4, LAG3, TGFB1, NOS3, and IDO1 between LINC00887 high- (red) and low-expression (blue) groups. (h) SNV waterfall plot of the LINC00887 low-expression group. (i) SNV waterfall plot of the LINC00887 high-expression group.

4. Discussion

lncRNAs participate in multiple processes of human cancer, including tumor growth, angiogenesis, metastasis, drug resistance, self-renewal of cancer cells, and tumor microenvironment regulation [24]. Recent studies have uncovered important functions of various aberrantly expressed lncRNAs in ccRCC. For instance, DMDRMR-mediated manipulation of m6A receptor IGF2BP3 to m6A modified CDK4 drives ccRCC progression [25]. lncRNA-LET suppresses ccRCC cell growth by modulating miR-373-3p [26]. LINC00973 is involved in cancer immunosuppression

via positive modulation of Siglec-15 in ccRCC [27]. Bioinformatics analysis in this study revealed significant overexpression of LINC00887 in ccRCC and its potential link with immune infiltration.

CD8+ T cell infiltration degree can be used as an indicator of tumor prognosis [28]. The immune infiltration of CD8 +T cells is negatively regulated by the tumor microenvironment, leading to tumor cells evading immune surveillance and thus promoting tumor progression [9]. Zhang et al. [29] found that TCL6 correlates with CD8+ T cell infiltration and indicates poor survival in breast cancer. Chen et al. [30] found that GOLM1 is positively connected to

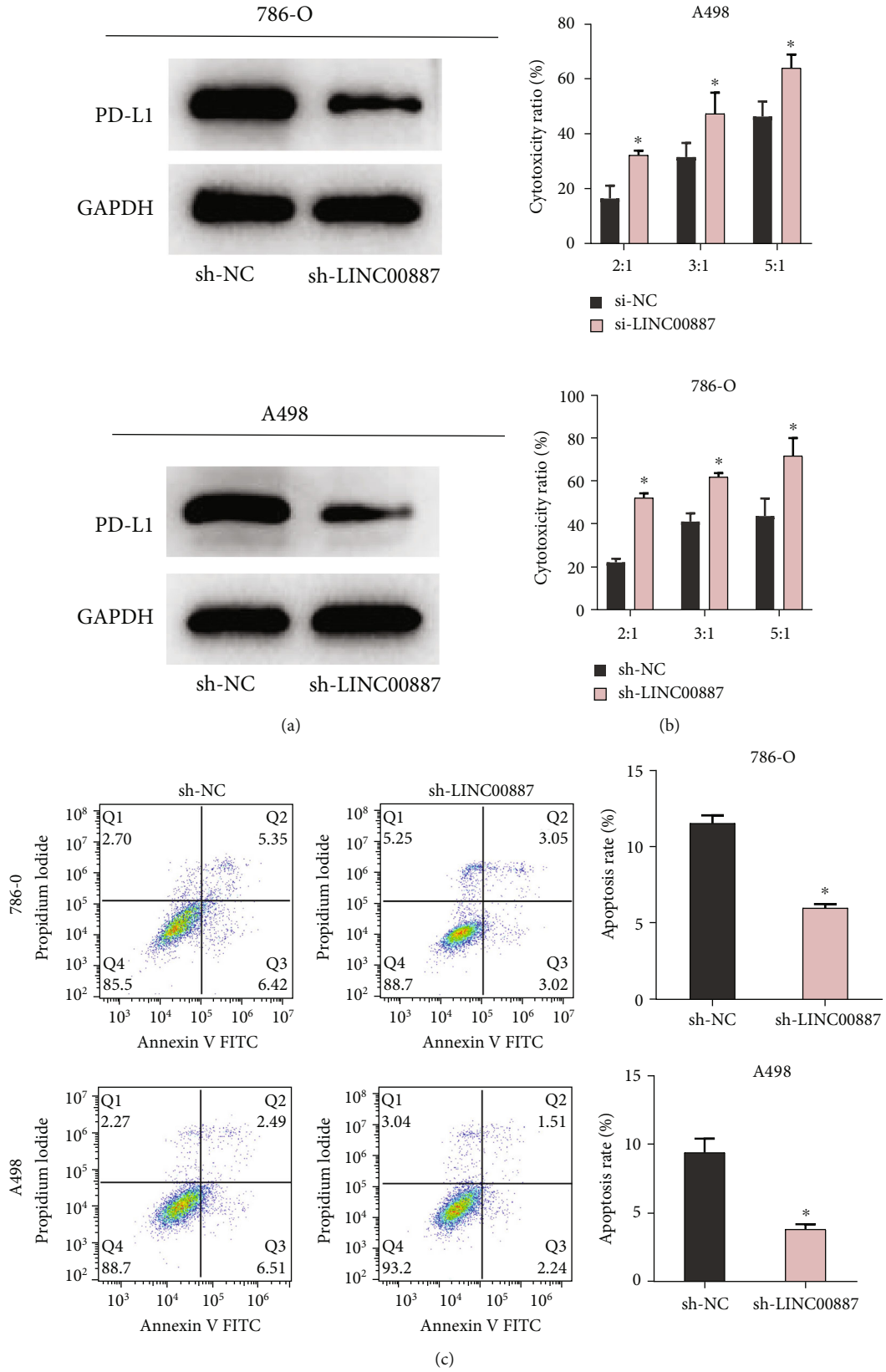


FIGURE 4: Continued.

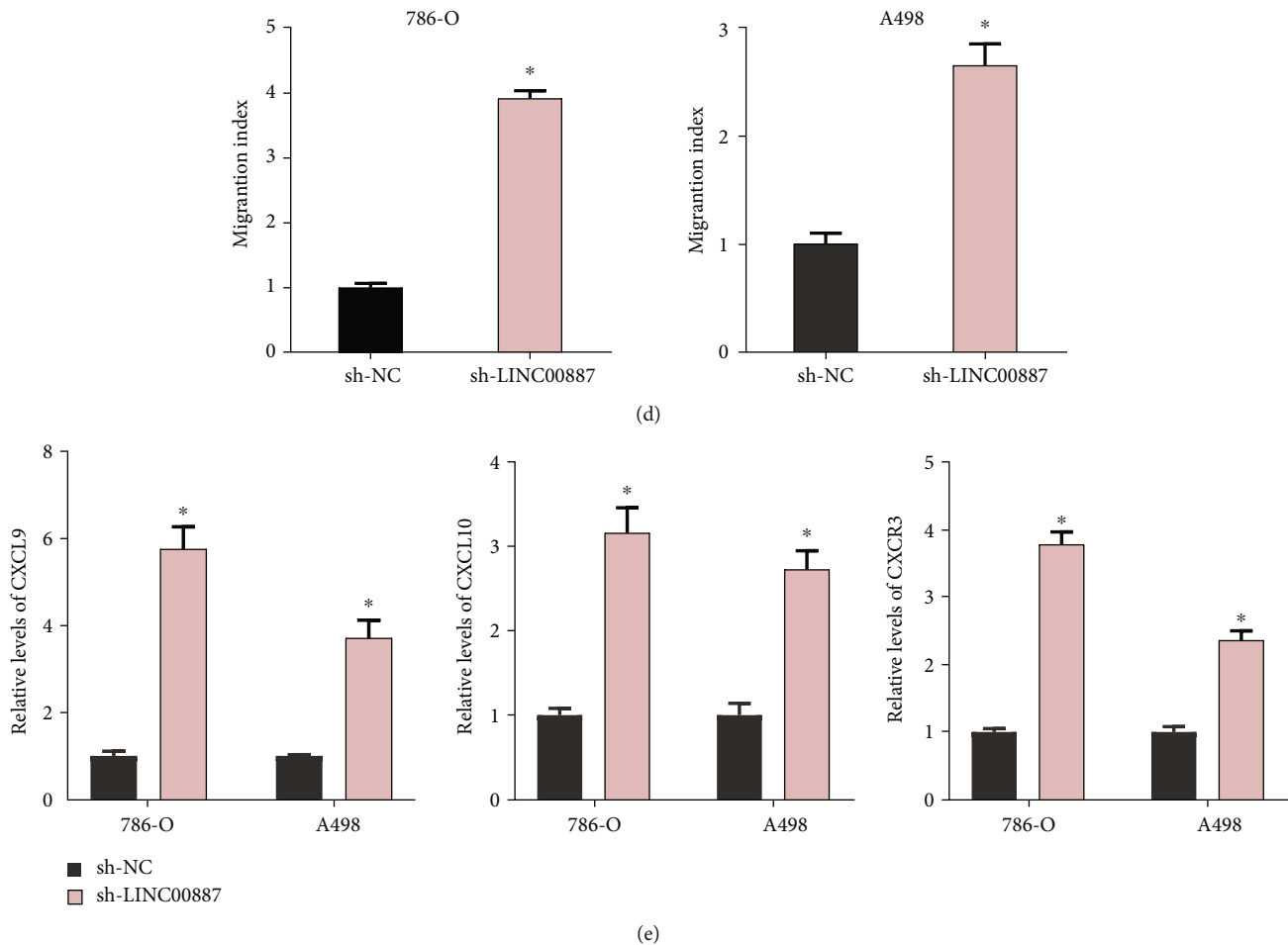


FIGURE 4: LINC00887 promotes ccRCC progression through CD8+ T cell-mediated immune response. (a) Effects of knocking down LINC00887 on ccRCC cells and PD-L1 were revealed by western blot assay. (b) The cytotoxicity of CD8+ T cells was evaluated by LDH assay. (c) Apoptosis of CD8+ T cells was detected by flow cytometry. (d) The chemotactic ability of CD8+ T cells was evaluated by chemotactic assay. (e) Expression of CXCL9, CXCL10, and CXCR3 in CD8+ T cells was evaluated by RT-qPCR, $*P < 0.05$.

invasive tumor-associated macrophages inhibited by PD-L1 and CD8+ T cells in colorectal cancer tissues, and zoledronic acid combined with anti-PD-L1 therapy reduces CD8+ T cell inhibition and pD-L1 + TAMs infiltration, resulting in inhibition of tumor growth in mice HCC model. Our work found that LINC00887 knockdown in ccRCC could improve the toxicity and chemotactic ability of CD8+ T cells to ccRCC cells. This suggested that LINC00887 promoted ccRCC progression through a CD8+ T cell-mediated immune response.

Herein, LINC00887 was significantly positively correlated with immunosuppressive molecule PD-1, and immunosuppressive molecules TGFB1, NOS3, and IDO1 also showed higher expression levels in LINC00887 high-expression group. Considering that PD-1 is the main inhibitory receptor of T cells, the higher its expression level, the weaker the effector capacity of T cells [31]. LINC00887 was positively correlated with PD-1 expression, further suggesting that LINC00887 mediated stronger T cell inhibition. TGFB1 is also a key gene in T cell regulation, and blocking TGFB can effectively amplify CD8+ T cells, thus enhancing the efficacy of immune checkpoints [32]. In this study, the expression trend of TGFB1

and LINC00887 was consistent, which may suggest that the two exert a synergistic role in mediating CD8+ T cell immunosuppression. IDO1 is a potential target for tumor immunotherapy. Targeted inhibition of IDO1 expression reverses tumor suppressive immune microenvironment [33]. In our results, IDO1 expression was also relatively high in the LINC00887 group. In the present study, SNP mutation landscape of LINC00887 high- and low-expression groups was analyzed. The results demonstrated that PBRM1 mutation was more frequent in the LINC00887 high-expression group. PBRM1 mutation is known to be closely related to the high angiopoiesis level in renal cancer [34], and PBRM1 mutations reduce the clinical benefit of immune checkpoint blocking therapy in RCC [35]. The results suggested that the oncogenic effect of LINC00887 may be associated with a higher PBRM1 mutation level. Also, NOS3 was reported to serve as an oncogene in gastric cancer [36], which supports the predicted result on the expression relationship between LINC00887 and NOS3 in this study. In a word, the expression trend of LINC00887 is relatively consistent with that of most immunosuppressive genes. This further confirmed that LINC00887 was involved

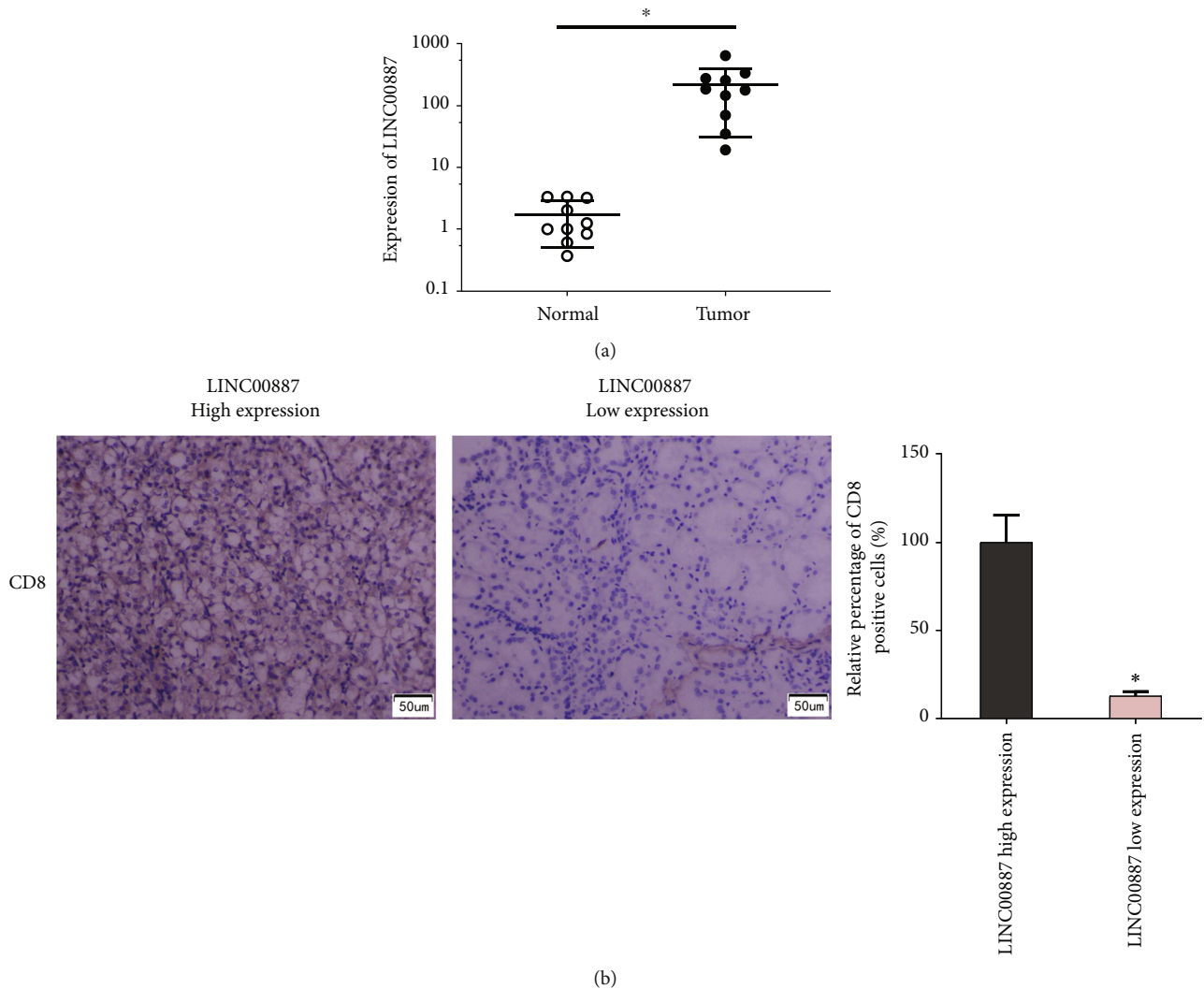


FIGURE 5: High expression of LINC00887 significantly inhibits immune infiltration level of CD8+ T cells. (a) Expression LINC00887 in 10 pairs of tissue samples. (b) The positive cell rate of CD8 cells, * $P < 0.05$.

in the immunosuppressive regulation of ccRCC and promoted malignant progression of ccRCC through synergistic expression of immunosuppressive molecules.

In conclusion, this study preliminarily illustrated the high level of LINC00887 in ccRCC, and it was negatively connected to the abundance of CD8+ T cell immune infiltration; LINC00887 knockdown led to increased cytotoxicity, weakened apoptosis ability, and enhanced chemotactic ability of CD8+ T cells. However, limitations still exist. First, the results have not been verified in vivo at the animal level. Second, the detailed regulatory mechanism of LINC00887 in regulating the immune response in ccRCC remains to be determined. In the future, we will further verify the results and explore the role of LINC00887 in ccRCC antitumor drug development through animal experiments. Based on above studies, our team believes that LINC00887 can be a biomarker for ccRCC diagnosis and can play a role in future drug development as a target for immunotherapy of ccRCC.

Data Availability

The data used to support the findings of this study are included within the article. The data and materials in the current study are available from the corresponding author on reasonable request.

Ethical Approval

The protocol involving human sample collection was approved by the Shengli Clinical Medical College of Fujian Medical University, Fujian Provincial Hospital Ethics Committee.

Consent

All authors consent to submit the manuscript for publication.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Jinfeng Wu and Rongcheng Lin contributed equally to this work.

Acknowledgments

The project was supported by the Natural Science Foundation of Fujian Province (2020J011085).

References

- [1] J. Schödel, S. Grampp, E. R. Maher et al., "Hypoxia, hypoxia-inducible transcription factors, and renal cancer," *European Urology*, vol. 69, no. 4, pp. 646–657, 2016.
- [2] Z. Fang, Q. Sun, H. Yang, and J. Zheng, "SDHB suppresses the tumorigenesis and development of ccRCC by inhibiting glycolysis," *Frontiers in Oncology*, vol. 11, p. 639408, 2021.
- [3] J. Ma, M. Li, J. Chai et al., "Expression of RSK4, CD44 and MMP-9 is upregulated and positively correlated in metastatic ccRCC," *Diagnostic Pathology*, vol. 15, no. 1, p. 28, 2020.
- [4] X. Meng, K. Liu, Z. Xiang, X. Yu, P. Wang, and Q. Ma, "MiR-125b-2-3p associates with prognosis of ccRCC through promoting tumor metastasis via targeting EGR1," *American Journal of Translational Research*, vol. 12, no. 9, pp. 5575–5585, 2020.
- [5] D. F. McDermott, J. L. Lee, G. A. Bjarnason et al., "Open-label, single-arm phase II study of pembrolizumab monotherapy as first-line therapy in patients with advanced clear cell renal cell carcinoma," *Journal of Clinical Oncology*, vol. 39, pp. 1020–1028, 2021.
- [6] B. Ljungberg, L. Albiges, Y. Abu-Ghanem et al., "European association of urology guidelines on renal cell carcinoma: the 2019 update," *European Urology*, vol. 75, no. 5, pp. 799–810, 2021.
- [7] J. Ou, P. Lei, Z. Yang et al., "LINC00152 mediates CD8(+) T-cell infiltration in gastric cancer through binding to EZH2 and regulating the CXCL9, 10/CXCR3 axis," *Journal of Molecular Histology*, vol. 52, no. 3, pp. 611–620, 2021.
- [8] C. Fu and A. Jiang, "Dendritic cells and CD8 T cell immunity in tumor microenvironment," *Frontiers in Immunology*, vol. 9, p. 3059, 2018.
- [9] J. S. Dolina, N. Van Braeckel-Budimir, G. D. Thomas, and S. Salek-Ardakani, "CD8(+) T cell exhaustion in cancer," *Frontiers in Immunology*, vol. 12, p. 715234, 2021.
- [10] Y. T. Tan, J. F. Lin, T. Li, J. J. Li, R. H. Xu, and H. Q. Ju, "LncRNA-mediated posttranslational modifications and reprogramming of energy metabolism in cancer," *Cancer Communications*, vol. 41, no. 2, pp. 109–120, 2021.
- [11] J. A. Heward and M. A. Lindsay, "Long non-coding RNAs in the regulation of the immune response," *Trends in Immunology*, vol. 36, no. 9, pp. 408–419, 2014.
- [12] D. Huang, J. Chen, L. Yang et al., "NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death," *Nature Immunology*, vol. 19, no. 10, pp. 1112–1125, 2018.
- [13] J. Cao, R. Dong, L. Jiang et al., "lncRNA-MM2P identified as a modulator of macrophage M2 polarization," *Cancer Immunology Research*, vol. 7, no. 2, pp. 292–305, 2019.
- [14] F. Ma, Y. Y. Lei, M. G. Ding, L. H. Luo, Y. C. Xie, and X. L. Liu, "lncRNA NEAT1 interacted with DNMT1 to regulate malignant phenotype of cancer cell and cytotoxic T cell infiltration via epigenetic inhibition of p53, cGAS, and STING in lung cancer," *Frontiers in Genetics*, vol. 11, p. 250, 2020.
- [15] P. Li, J. Wang, L. Zhi, and F. Cai, "Linc00887 suppresses tumorigenesis of cervical cancer through regulating the miR-454-3p/FRMD6-Hippo axis," *Cancer Cell International*, vol. 21, no. 1, p. 33, 2021.
- [16] X. M. Shen, S. Han, N. Liu, H. Q. Xu, C. X. Yan, and C. J. Yu, "LINC00887 aggravates the malignant progression of glioma via upregulating CCND1," *European Review for Medical and Pharmacological Sciences*, vol. 25, pp. 1928–1935, 2021.
- [17] W. J. Yue, Y. Wang, W. Y. Li, and Z. D. Wang, "LINC00887 regulates the proliferation of nasopharyngeal carcinoma via targeting miRNA-203b-3p to upregulate NUP205," *European Review for Medical and Pharmacological Sciences*, vol. 24, pp. 8863–8870, 2020.
- [18] J. Xie, Y. Zhong, R. Chen et al., "Serum long non-coding RNA LINC00887 as a potential biomarker for diagnosis of renal cell carcinoma," *FEBS Open Bio*, vol. 10, no. 9, pp. 1802–1809, 2020.
- [19] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: a bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, vol. 26, no. 1, pp. 139–140, 2010.
- [20] Z. L. Skidmore, A. H. Wagner, R. Lesurf et al., "GenVisR: genomic visualizations in R," *Bioinformatics (Oxford, England)*, vol. 32, pp. 3012–3014, 2016.
- [21] Q. H. Chen, B. Li, D. G. Liu, B. Zhang, X. Yang, and Y. L. Tu, "lncRNA KCNQ1OT1 sponges miR-15a to promote immune evasion and malignant progression of prostate cancer via up-regulating PD-L1," *Cancer Cell International*, vol. 20, no. 1, p. 394, 2020.
- [22] J. Liu, F. Li, Y. Ping et al., "Local production of the chemokines CCL5 and CXCL10 attracts CD8+ T lymphocytes into esophageal squamous cell carcinoma," *Oncotarget*, vol. 6, pp. 24978–24989, 2015.
- [23] R. Zhao, Y. Song, Y. Wang et al., "PD-1/PD-L1 blockade rescue exhausted CD8+ T cells in gastrointestinal stromal tumours via the PI3K/Akt/mTOR signalling pathway," *Cell Proliferation*, vol. 52, no. 3, article e12571, 2019.
- [24] M. Zhang, N. Wang, P. Song et al., "lncRNA GATA3-AS1 facilitates tumour progression and immune escape in triple-negative breast cancer through destabilization of GATA3 but stabilization of PD-L1," *Cell Proliferation*, vol. 53, no. 9, article e12855, 2020.
- [25] Y. Gu, S. Niu, Y. Wang et al., "DMDRMR-mediated regulation of m6A-modified CDK4 by m6A reader IGF2BP3 drives ccRCC progression," *Cancer Research*, vol. 81, no. 4, pp. 923–934, 2021.
- [26] Z. Ye, J. Duan, L. Wang, Y. Ji, and B. Qiao, "lncRNA-LET inhibits cell growth of clear cell renal cell carcinoma by regulating miR-373-3p," *Cancer Cell International*, vol. 19, no. 1, p. 311, 2019.
- [27] Y. Liu, X. Li, C. Zhang, H. Zhang, and Y. Huang, "LINC00973 is involved in cancer immune suppression through positive regulation of Siglec-15 in clear-cell renal cell carcinoma," *Cancer Science*, vol. 111, no. 10, pp. 3693–3704, 2020.

- [28] L. T. Buck, G. Burness, K. L. Campbell et al., “50 years of comparative biochemistry: The legacy of Peter Hochachka,” *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology*, vol. 224, pp. 1–11, 2018.
- [29] Y. Zhang, Z. Li, M. Chen et al., “lncRNA TCL6 correlates with immune cell infiltration and indicates worse survival in breast cancer,” *Breast Cancer*, vol. 27, no. 4, pp. 573–585, 2020.
- [30] J. Chen, Z. Lin, L. Liu et al., “GOLM1 exacerbates CD8+ T cell suppression in hepatocellular carcinoma by promoting exosomal PD-L1 transport into tumor-associated macrophages,” *Signal Transduction and Targeted Therapy*, vol. 6, no. 1, p. 397, 2021.
- [31] J. H. Cha, L. C. Chan, C. W. Li, J. L. Hsu, and M. C. Hung, “Mechanisms controlling PD-L1 expression in cancer,” *Molecular Cell*, vol. 76, no. 3, pp. 359–370, 2019.
- [32] X. Bai, M. Yi, Y. Jiao, Q. Chu, and K. Wu, “Blocking TGF-beta signaling to enhance the efficacy of immune checkpoint inhibitor,” *OncoTargets and Therapy*, vol. 12, pp. 9527–9538, 2019.
- [33] L. Xie, K. Hu, Y. Duo et al., “Off-tumor IDO1 target engagements determine the cancer-immune set point and predict the immunotherapeutic efficacy,” *Journal for Immunotherapy of Cancer*, vol. 9, 2021.
- [34] R. J. Motzer, R. Banchereau, H. Hamidi et al., “Molecular subsets in renal cancer determine outcome to checkpoint and angiogenesis blockade,” *Cancer Cell*, vol. 38, pp. 803–817, 2020.
- [35] X. D. Liu, W. Kong, C. B. Peterson et al., “PBRM1 loss defines a nonimmunogenic tumor phenotype associated with checkpoint inhibitor resistance in renal carcinoma,” *Nature Communications*, vol. 11, no. 1, p. 2135, 2020.
- [36] D. Zou, Z. Li, F. Lv et al., “Pan-cancer analysis of NOS3 identifies its expression and clinical relevance in gastric cancer,” *Frontiers in Oncology*, vol. 11, p. 592761, 2021.

Research Article

Construction of Prognostic Risk Model of Patients with Skin Cutaneous Melanoma Based on TCGA-SKCM Methylation Cohort

Xiaoming Yu , Ping Cong , Wei Wei , Yong Zhou , Zhengqiang Bao ,
and Huaying Hou 

Department of Cancer Center, The Second Hospital of Shandong University, Jinan, Shandong 250033, China

Correspondence should be addressed to Huaying Hou; huayingh33@163.com

Received 6 June 2022; Revised 13 July 2022; Accepted 23 July 2022; Published 25 August 2022

Academic Editor: Tao Huang

Copyright © 2022 Xiaoming Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Skin cutaneous melanoma (SKCM) is a common malignant skin cancer. Early diagnosis could effectively reduce SKCM patient's mortality to a large extent. We managed to construct a model to examine the prognosis of SKCM patients. The methylation-related data and clinical data of The Cancer Gene Atlas- (TCGA-) SKCM were downloaded from TCGA database. After preprocessing the methylation data, 21,861 prognosis-related methylated sites potentially associated with prognosis were obtained using the univariate Cox regression analysis and multivariate Cox regression analysis. Afterward, unsupervised clustering was used to divide the patients into 4 clusters, and weighted correlation network analysis (WGCNA) was applied to construct coexpression modules. By overlapping the CpG sites between the clusters and turquoise model, a prognostic model was established by LASSO Cox regression and multivariate Cox regression. It was found that 9 methylated sites included cg01447831, cg14845689, cg20895058, cg06506470, cg09558315, cg06373660, cg17737409, cg21577036, and cg22337438. After constructing the prognostic model, the performance of the model was validated by survival analysis and receiver operating characteristic (ROC) curve, and the independence of the model was verified by univariate and multivariate regression. It was represented that the prognostic model was reliable, and riskscore could be used as an independent prognostic factor in SKCM patients. At last, we combined clinical data and patient's riskscore to establish and testify the nomogram that could determine patient's prognosis. The results found that the reliability of the nomogram was relatively good. All in all, we constructed a prognostic model that could determine the prognosis of SKCM patients and screened 9 key methylated sites through analyzing data in TCGA-SKCM dataset. Finally, a prognostic nomogram was established combined with clinical diagnosed information and riskscore. The results are significant for improving the prognosis of SKCM patients in the future.

1. Introduction

Skin cutaneous melanoma (SKCM) is a malignant skin cancer formed by the canceration of melanocytes below the epidermis. Pathogenic factors for SKCM are mainly classified into external and internal factors [1]. The most common external pathogenic factor for SKCM is ultraviolet radiation, and studies found that when the underlying cells of the skin are exposed to ultraviolet radiation, DNA damage can be caused to further induce SKCM [2–5]. Moreover, environmental factors are also important external factors for SKCM. For example, di-(2-ethylhexyl) phthalate (DEHP) in cosmetics

and PM2.5 in the air can increase the incidence rate of SKCM [6, 7]. Besides common environmental carcinogenic factors, genetic and epigenetic modifications are also common carcinogenic factors, and studies considered that people from families with CDKN2A gene mutation have higher incidence rate of melanoma [8, 9]. Telomerase reverse transcriptase (TERT) can also increase the risk of SKCM [10]. In addition, epigenetic modification is a crucial factor for melanoma. Studies disclosed that methylation of genes like APC, PYCARD, and COL11A1 is relevant to the incidence of melanoma, and histone H3K27me3 upregulation can promote melanoma progression [11, 12]. Additionally, current studies uncovered that

immunosuppression and pigment characteristics are risk factors for melanoma [13–15].

DNA methylation is one of the most important epigenetic modifications. A considerable number of studies considered that DNA methylation modification is related to the incidence of melanoma. For example, de Unamuno Bustos et al. [16] found that aberrant methylation of genes like RARB and PTEN is associated with clinical melanoma progression. DNA methylation is correlated with the metastasis and drug resistance of melanoma as well. For example, Venza et al. [17] discovered that DNA methylation can promote melanoma metastasis by silencing E-cadherin. MGMT gene promoter methylation can promote the tolerance of temozolomide in melanoma [18]. In conclusion, DNA methylation can be widely involved in the occurrence and progression of melanoma. We believed that it had a good potential value to diagnose the prognosis of SKCM patients by detecting DNA methylation level.

The Cancer Genome Atlas (TCGA) is a database commonly applied in tumor biomarker screening, which includes abundant clinical experimental data of tumor patients, wherein it contains SKCM methylation data of 470 cases that can be used for analyzation and have significant clinical value [19]. Presently, massive studies have applied TCGA database for screening of tumor biomarkers. For example, Zhu et al. [20] obtain methylation sites related to prognosis of lung cancer via digging methylation data of lung cancer patients in TCGA. Olkhov-Mitsel et al. [21] acquired methylation biomarkers that can identify bladder cancer stage differences through digging the methylation data of patients with bladder cancer. In this study, the methylation data as well as clinical data in TCGA-SKCM database were used to screen methylation sites that resulted in poor prognosis of SKCM patients.

In this study, TCGA database was applied to screen prognosis-related methylation sites via univariate regression analysis, multivariate regression analysis, the weighted correlation network analysis (WGCNA), and unsupervised clustering analysis. After constructing a risk model by LASSO and multivariate Cox regression, we identified the independence and accuracy of the model. Lastly, a prognostic nomogram was constructed combined with relevant clinical data. The nomogram has great guiding significance for the clinical diagnosis and treatment of SKCM.

2. Materials and Methods

2.1. Data Processing. Methylation sequencing data and corresponding clinical information in TCGA-SKCM dataset from TCGA database (<https://portal.gdc.cancer.gov/>) were included in this study. Clinical data like patient’s age, gender, and tumor stage in TCGA-SKCM dataset were first downloaded from TCGA website (Table S1). Afterward, DNA methylation (450K) data were accessed from TCGA-SKCM dataset on 20 July 2020, including DNA methylation sequencing results of 2 healthy tissue samples and 473 tumor tissue samples.

Downloaded DNA methylation data were filtrated according to the following criteria: (1) removing methylation sites that missed more than 70% of data, (2) filtrating all

non-GpG methylation sites, (3) filtrating all SNP-related methylation sites, (4) filtrating all methylation sites that mapped to multiple locations, and (5) filtrating all methylation sites in the X and Y chromosomes. Afterward, 361,126 methylation sites were obtained. R package “KNN” [22] was applied to complete the missing values in the methylation expression profile, and R package “ChAMP” [23] was used to standardize the data. After obtaining standardized methylation data, tumor tissue samples with the follow-up time more than 0d were selected ($n = 455$). The samples were randomly divided into the training set ($n = 318$) and validation set ($n = 137$) in a proportion of 7:3.

2.2. Preliminary Screening of Methylation Sites Related to the Prognosis. Data in the training set were analyzed by univariate regression to calculate methylation sites that were significantly correlated with patient’s overall survival (OS) ($p < 0.05$). Thereafter, relevant clinicopathological data were combined to perform multivariate regression analysis on screened prognosis-related methylation sites, thereby screening methylation sites correlated with patient’s OS ($p < 0.05$). All analyses were finished with R package “survival” [24].

2.3. Unsupervised Clustering Analysis. According to the methylation level of methylation sites related to patient’s prognosis, unsupervised clustering analysis was performed on patients with R package “ConsensusClusterPlus” [25]. The “pam” method was selected for clustering, and the “Euclidean” method was used to calculate the sample distance. The optimal cluster number was assessed by the cumulative distribution function (CDF) and its area under curve (AUC). After clustering, OS of patients with different disease subtypes was analyzed with R package “survival” [24].

2.4. WGCNA. WGCNA was performed on data in the training set with the R package “WGCNA” [26]. Firstly, we evaluated prognosis-related methylation sites and screened methylation sites with the variance ranked in the top 5000 to construct a coexpression network. Thereafter, the Pearson correlation index was used to establish adjacent matrix with the soft threshold $\beta = 8$. Afterward, the adjacent matrix was transferred into the topological matrix (TOM). Based on TOM, average-linkage hierarchical clustering was applied to cluster methylation sites. Lastly, a dynamic tree cut algorithm was exerted to identify coexpression modules with the size of minimum module of 30.

After clustering coexpression modules, module subtype (MS) correlation was used to identify the correlation between coexpression modules and patients’ subtypes. Following this, methylation sites correlated with patients’ subtypes were screened by GpG-site significance (CS) and module membership (MM). CS represented the correlation between the methylation level of methylation sites and patients’ subtypes, and MM represented the correlation between methylation level of methylation sites corresponding to the patients’ subtypes and module eigenvalue.

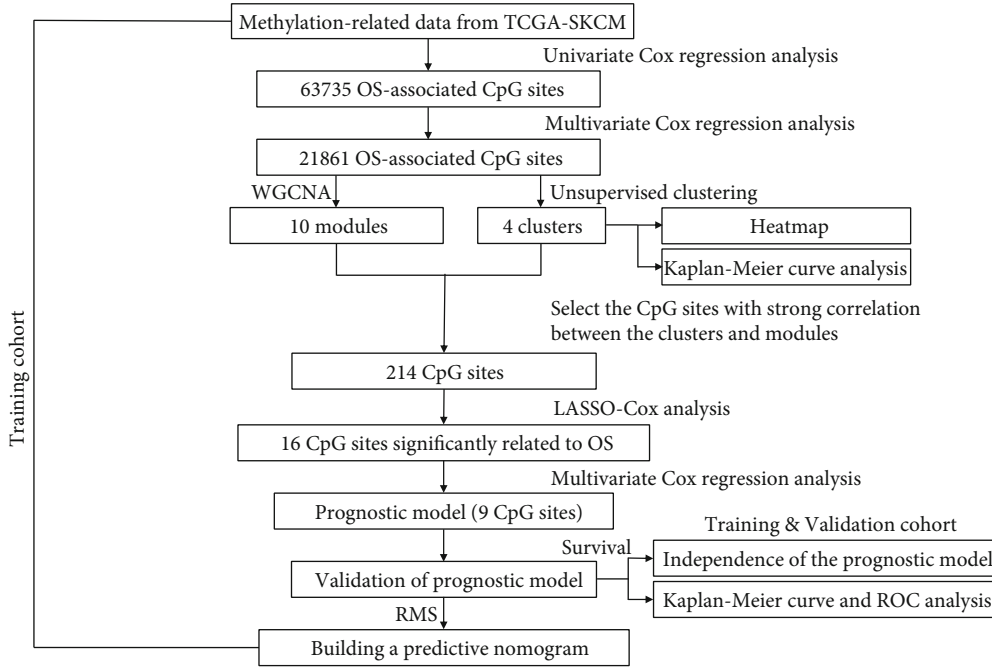


FIGURE 1: Overall flowchart of this study.

Methylation sites ($CS > 0.6$ and $MM > 0.8$) were screened to establish a prognostic risk model.

2.5. Construction and Validation of Methylation-Related Prognostic Model. After acquiring key methylation sites, LASSO Cox regression analysis was performed on data in the training set with R package “glmnet” [27] to reduce the complexity of the module and screen key methylation sites with prognostic value. Afterward, multivariate regression analysis was undertaken with R package “survminer” [28] on key methylation sites screened by LASSO Cox. Finally, a risk model was established. Riskscore in the risk model was calculated by the formula as follows:

$$\text{Riskscore} = \sum_{i=1}^n (\text{Coef}_i \times x_i). \quad (1)$$

In the formula, Coef_i represents the risk index of each methylation site, x_i represents methylation level of each methylation site, and Riskscore represents the ultimate riskscore. After screening prognosis-related methylation sites, related information of methylation sites was searched according to Ensembl database (version GRCh38.p13) (<http://asia.ensembl.org>).

Patients were divided into two groups according to the median of the score based on the formula: the high-risk group and the low-risk group. Survival analysis was undertaken on two groups with R package “survival,” and the accuracy of the model was identified by receiver operating characteristic (ROC) curve. Finally, the score distribution map, survival status distribution map, and methylation level heatmap of samples were drawn.

2.6. Assessment of Clinical Characteristics and Prognostic Independence of Riskscore. Univariate regression and multivariate regression analyses were performed on riskscore combined with clinical information including age, gender, pathological_T stage, pathological_N stage, pathological M stage, and tumor stage to analyze the correlation between the indexes and patient’s OS, respectively. The indexes that were significantly correlated with patient’s OS both in the results of univariate and multivariate regression analyses were considered to have the independent prognostic value.

2.7. Construction of a Prognostic Nomogram. Combined with clinically related indexes and riskscore, a nomogram that could predict 1-, 3-, and 5-year survival rates of patients was established with R package “rms” [29]. Correction curves of 1, 3, and 5 years were generated with R package “foreign” [30] after establishing the nomogram to identify the predictive effect of the nomogram.

3. Results

3.1. Screening of Prognosis-Related Methylation Sites. The flow chart of this study is shown in Figure 1. After downloading and preprocessing methylation-related data, we screened 63,735 prognosis-related methylation sites by univariate regression analysis and then followed by multivariate regression analysis obtaining 21,861 methylation sites which were remarkably correlated with SKCM patients’ OS (Table S2).

3.2. Four Groups of Patients with Different Subtypes Found by Unsupervised Clustering Analysis. After screening methylation sites that were significantly correlated with SKCM patient’s OS, we clustered patients by unsupervised

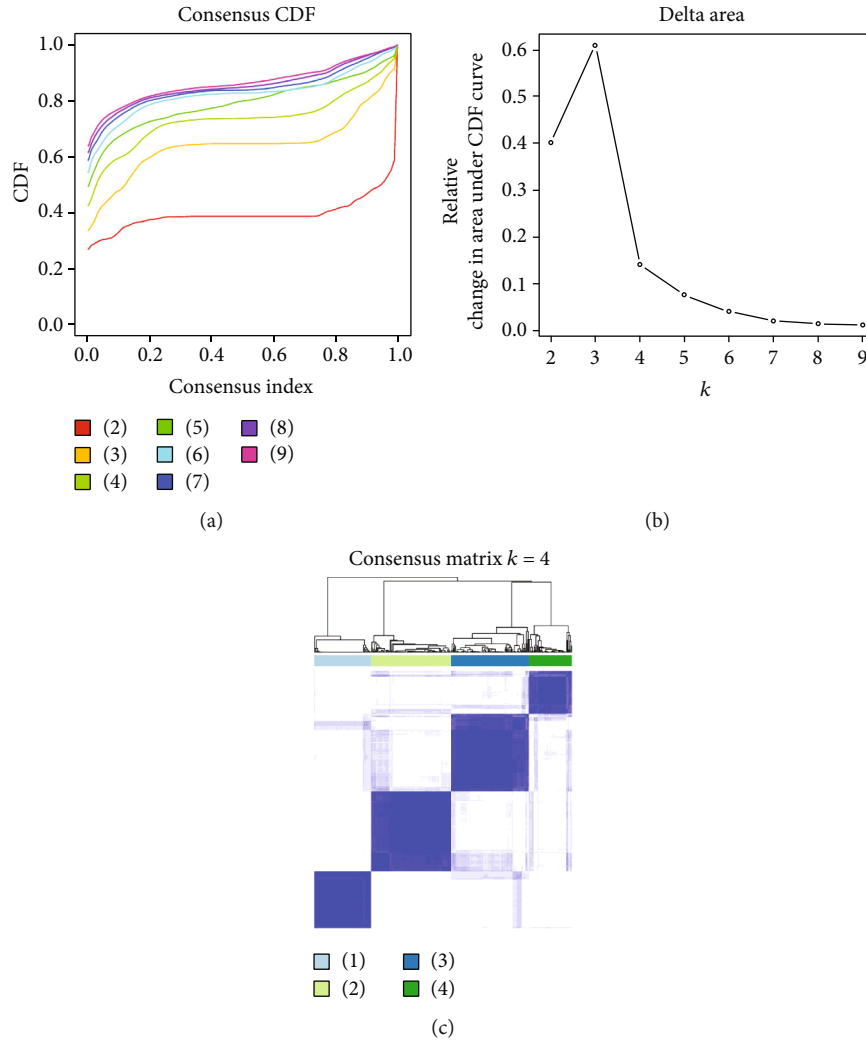


FIGURE 2: Unsupervised clustering of SKCM patients. (a) Cumulative distribution function curve of unsupervised clustering. (b) Relative change in area under cumulative distribution function (CDF) curve. (c) Clustering heatmap of 4 types of SKCM patients. Each cluster represents a subgroup of patients.

clustering method (Figures 2(a) and 2(b)). Clustering results showed that patients were mainly divided into 4 subtypes: cluster 1, cluster 2, cluster 3, and cluster 4 (Figure 2(c)). Heatmap of 5,000 methylation sites with the highest variance was drawn combined with patient's clinical information. The result represented that the methylation levels differed in patients with 4 subtypes (Figure 3). Finally, survival analysis was performed on patients in the 4 groups. It was shown that OS of patients with 4 different subtypes had differences; the prognosis of cluster 2 patients was the poorest, and the prognosis of cluster 1 patients was the best (Figure 4). The above results exhibited that the unsupervised clustering analysis could reliably cluster the patients into 4 subtypes, and there were differences in methylation level and OS of patients with different subtypes had differences.

3.3. WGCNA. After grouping patients and screening 5,000 methylation sites with highest variance, we further performed WGCNA on the methylation sites (Figures 5(a) and 5(b)) and ultimately obtained 10 different modules

(Figure 5(c)). Correlation analysis was undertaken on the 10 methylation modules and 4 different patient's subtypes. Then, modules related to patient's subtypes were screened by MS. The results showed that patients with cluster 1 subtype and cluster 2 subtype were significantly correlated with most modules and represented opposite trends (Figure 5(d)). The further analysis discovered that the turquoise module had the highest correlation with cluster 1 and cluster 2. Hence, we chose the turquoise module for further analysis. Afterward, correlation analysis was performed on the 3,031 methylation sites in the turquoise module and cluster 1/cluster 2, respectively, to screen methylation sites. A total of 502 methylation sites both related to the turquoise module and cluster 1 were obtained, and 219 methylation sites both related to the turquoise module and cluster 2 were obtained (Figures 6(a) and 6(b)). At last, to further screen for methylation sites that are prominently associated with SKCM patients, Venn plot was used to intersect the sites and 214 key methylation sites were obtained (Figure 6(c)).

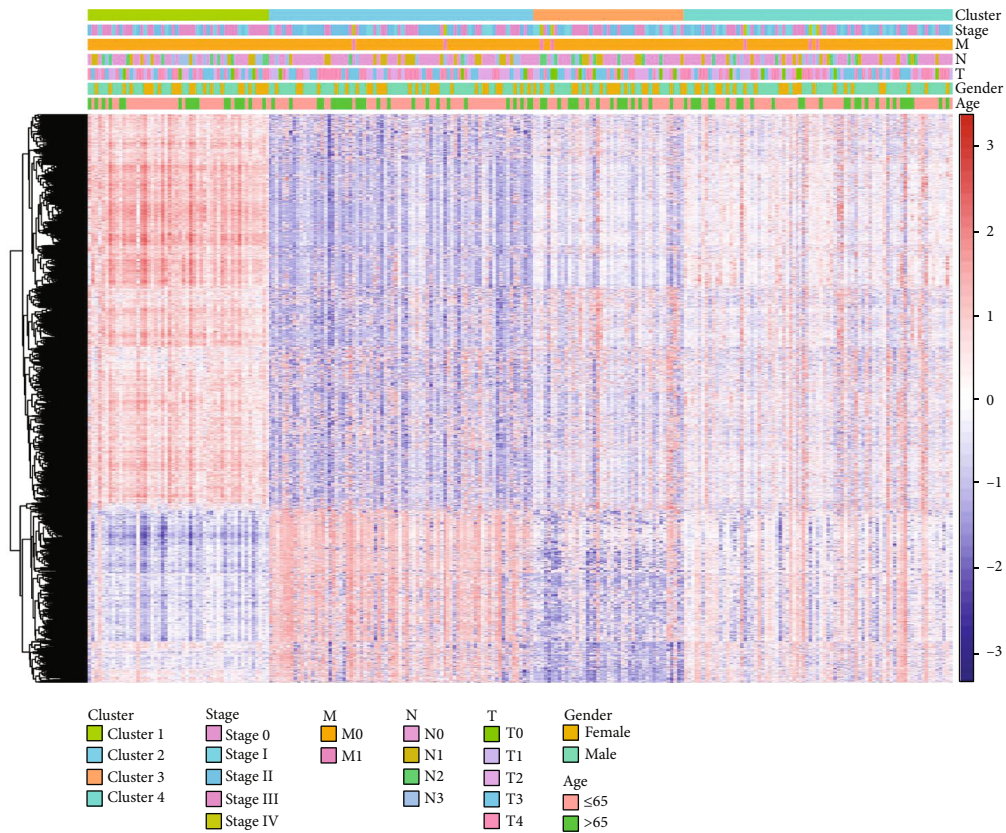


FIGURE 3: Heatmap of top 5000 variance CpG sites, with red representing high level of DNA methylation and blue representing low level of DNA methylation.

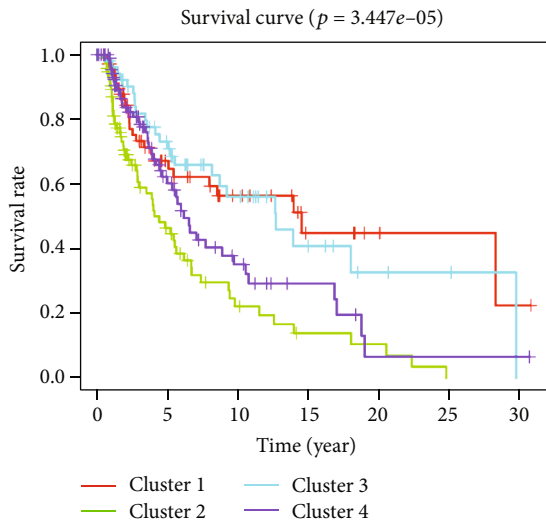


FIGURE 4: Survival curve of SKCM patients' OS, with each line representing a subgroup of SKCM patients.

3.4. Construction and Validation of Prognosis-Related Methylation Model. To avoid module overfitting, we screened key methylation sites with LASSO Cox regression and obtained 16 key methylation sites (Figures 7(a) and 7(b)). Multivariate regression analysis was performed on the 16 key methylation sites to construct multivariate regres-

sion model, and finally, 9 prognosis-related methylation sites (cg01447831, cg14845689, cg20895058, cg06506470, cg09558315, cg06373660, cg17737409, cg21577036, and cg22337438) were screened in SKCM (Figure 7(c) and Table 1). Meanwhile, risk model was obtained: $Riskscore = -1.0719 * cg01447831 + 0.6563 * cg14845689 + 0.7445 * cg20895058 + 0.6031 * cg06506470 - 0.6421 * cg09558315 - 0.5512 * cg06373660 - 0.4686 * cg17737409 - 0.9057 * cg21577036 + 0.8385 * cg22337438$.

After establishing the risk model, we detected the distribution and survival status of high- and low-risk patients in the training set with the score distribution map and survival status distribution map (Figure 7(d)). The result showed that patients with high risk commonly were more likely to die and the survival time of high-risk patients was relatively lower. Changes of the methylation levels of 9 methylation sites of high- and low-risk patients were analyzed with heatmap. The result was consistent with the risk index of model (Figure 7(e)). To identify the reliability of risk model, survival analysis was performed to compare the OS differences between high- and low-risk patients in the training set and the validation set. It was represented that OS in the low-risk group was significantly higher than that in the high-risk group (Figures 8(a) and 8(b)). ROC curve was further used to analyze the 1-, 3-, and 5-year survival of patients in the training set and the validation set. It was exhibited that AUC of ROC curve was 0.73, 0.71, and 0.73 in the training set, respectively, and AUC

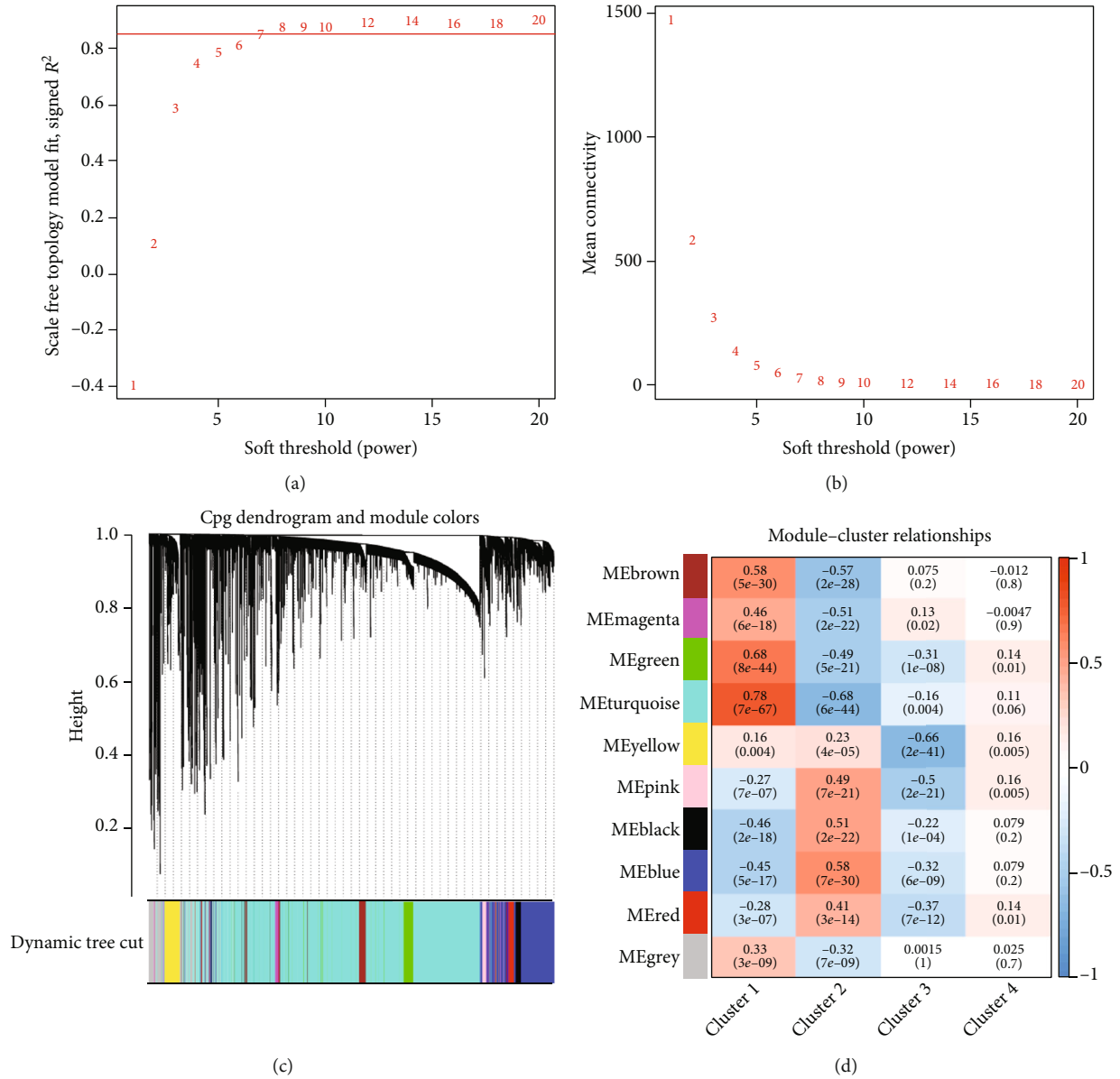


FIGURE 5: WGCNA analysis of the top 5000 variance CpG sites. (a) Analysis of the scale-free fit index for various soft-thresholding powers (β). (b) Analysis of the mean connectivity for various soft-thresholding powers. (c) Dendrogram of prognosis-related CpG sites clustered based on a dissimilarity measure (1-TOM). (d) Heatmap of the correlation between module and subgroups of patients.

of ROC curve was 0.74, 0.67, and 0.71 in the validation set, respectively, indicating that the model was reliable (Figures 8(c) and 8(d)). The results showed that the risk model could accurately determine 1-, 3-, and 5-year survival of patients. The above findings represented that the results of the risk model were accurate and the model could be used for predicting the prognosis of SKCM patients.

3.5. Identifying Model's Independence and Establishing Prognostic Nomogram. After establishing the risk model and validating accuracy of the model, univariate regression and multivariate regression analyses were applied combining with clinical data (age, gender, T, N, M, and tumor stage)

and riskscore to validate whether riskscore could independently determine patient's prognosis. It was shown that riskscore in the risk model was significantly correlated with patient's OS and could independently determine patient's prognosis in univariate and multivariate regression analyses (Figures 9(a) and 9(b)). After validating the independence of the risk model, we combined clinically related data to establish a nomogram that could be used to determine 1-, 3-, and 5-year survival rate of patients (Figure 10(a)). Afterward, fitting curve was used to validate the accuracy of the nomogram. The result showed that fitting results of 1-,3-, and 5-year were good (Figures 10(b)–10(d)). The nomogram can assist clinical doctors to diagnose patient's prognostic risk and help doctors to arrange therapeutic plans more accurately.

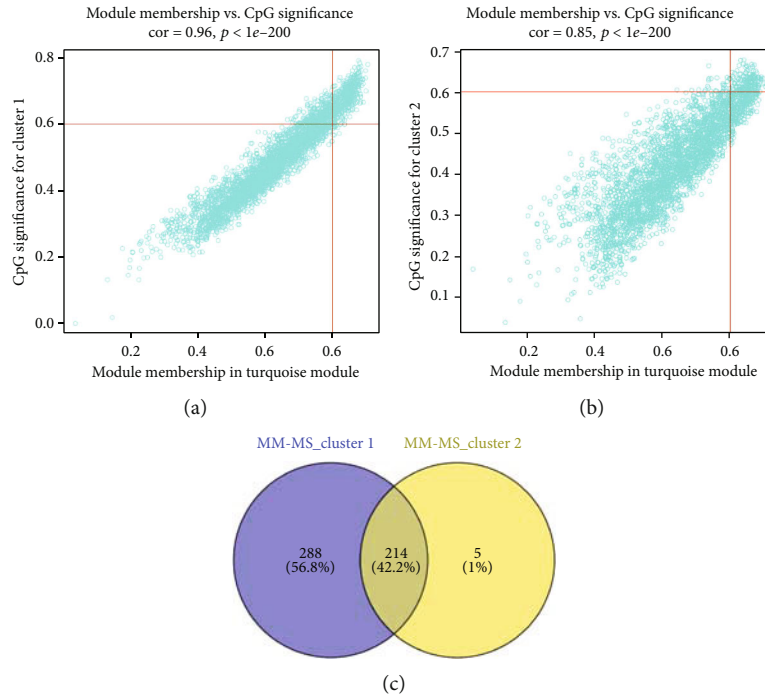


FIGURE 6: Filtering of OS-associated CpG sites. (a) Scatter plot of turquoise module member CpG sites related to cluster 1. Each dot represents a CpG site. (b) Scatter plot of turquoise module member CpG sites related to cluster 2. (c) Venn plot of important CpG sites. The blue circle represents cluster 1-associated CpG sites, and the yellow circle represents cluster 2-associated CpG sites.

4. Discussion

Melanoma is a common modern disease, and SKCM is a melanoma that occurs in the epidermis. Despite the high incidence rate of SKCM, its mortality can be relatively low if it is diagnosed in time in the early stage. We screened 214 methylation (CpG) sites remarkably associated with OS in SKCM patients from the TCGA-SKCM dataset by unsupervised clustering and WGCNA and finally constructed a prognostic model of 9 signature CpG sites using LASSO Cox regression analysis and demonstrated that the model had a better prognostic effect.

With the development of high-throughput sequencing technology, high-throughput sequencing on tumor patients has become an important method for tumor research, from which biomarkers with prognostic values are filtered [31]. Currently, the mainstream research method is evaluating patient's risk of cancer by analyzing mRNA expression data and combining with the expression of many genes [32]. The advantage of the method is that the combination of the expression of multiple genes to assess patient's prognosis is more accurate than using each gene alone. However, it is still not comprehensive enough for the determination of patient's prognosis. Presently, researchers managed to increase the accuracy for determining patient's prognosis by screening biomarkers via analyzing miRNAs and lncRNAs, and even patient's metabolic data [33]. Besides predicting patient's prognosis based on RNA expression data, a study tried to predict patient's prognosis by the methylation level of sites through digging relevant data of gene methylation [34], which can obtain more stable prediction

result and requires lower sample preserve conditions compared with using RNA expression level to predict prognosis. Hence, methylation sites that affected SKCM patient's prognosis were explored in this study by digging DNA methylation relevant data in TCGA-SKCM dataset combined with patient's clinical survival time. A total of 21,861 prognosis-related methylation sites were found through univariate and multivariate regression analyses.

Mining TCGA database by bioinformatics methods is a common study method, in which unsupervised cluster is an effective means to classify patients with cancer in the present [34–36]. For example, Wu et al. [37] found 3 subtypes with different molecular characters in lung adenocarcinoma by unsupervised clustering, and patients with each subtype are relevant to abnormal specific molecular pathways. Patients were divided into 4 groups in this study by unsupervised clustering. The OS of patients in different groups was different, wherein cluster 2 patients had the poorest prognosis, while cluster 1 patients had the best prognosis. WGCNA is also a common study method in biomarker screening, and a number of studies screened biomarkers with value for patient's prognosis by WGCNA [34]. Ten methylation modules related to patient's prognosis were obtained in this study by WGCNA. At last, 9 methylation sites and risk model relevant to OS were further screened combined with the results of WGCNA and unsupervised clustering.

The 9 prognosis-related methylation sites were screened in this study based on the previous research. Among the 9 sites, high methylation of cg01447831, cg09558315, cg06373660, cg17737409, and cg21577036 can reduce prognostic risk, while

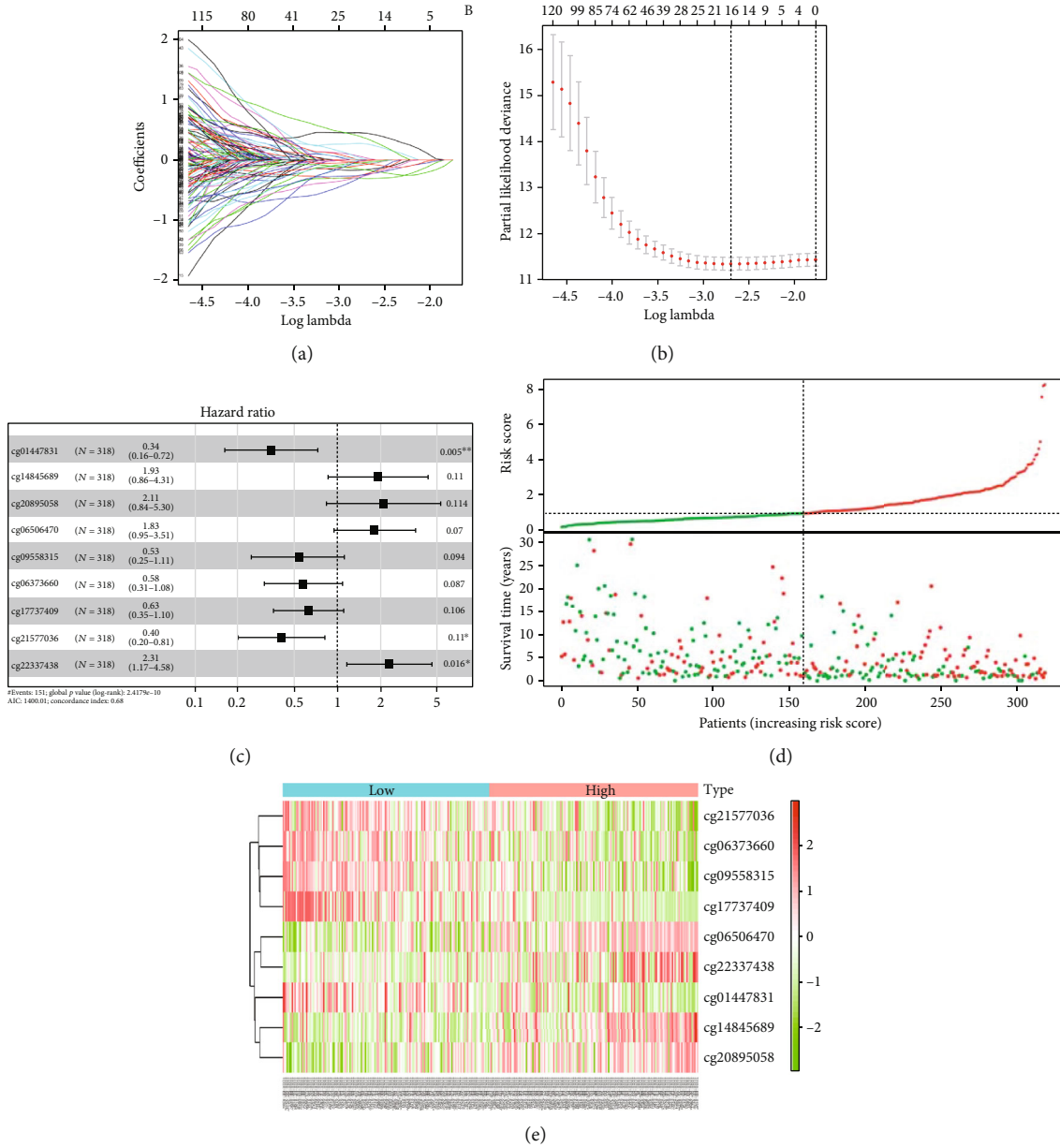


FIGURE 7: Construction of SKCM risk model. (a) LASSO coefficient profiles of key CpG sites. (b) Selection of the optimal parameter (lambda) in the LASSO model for TCGA-LUAD. (c) Key CpG sites filtered by multivariate Cox regression analysis. (d) Risk score and survival status of patients in training cohort. (e) Heatmap of each CpG site in risk model.

TABLE 1: Detailed information of filtered CpG sites.

CpG site	Chrom	Position	Gene symbol
cg01447831	chr7	155,456,390-155,456,439	EN2
cg14845689	chr10	117,546,682-117,546,731	EMX2
cg20895058	chr14	60,516,569-60,516,618	AL049874.3
cg06506470	chr7	3,990,973-3,991,022	SDK1
cg09558315	chr4	5,709,574-5,709,623	EVC
cg06373660	chr13	80,840,851-80,840,900	AL590807.1
cg17737409	chr1	230,114,215-230,114,264	GALNT2
cg21577036	chr15	90,301,567-90,301,616	AC091167.6
cg22337438	chr19	53,824,135-53,824,184	NLRP12

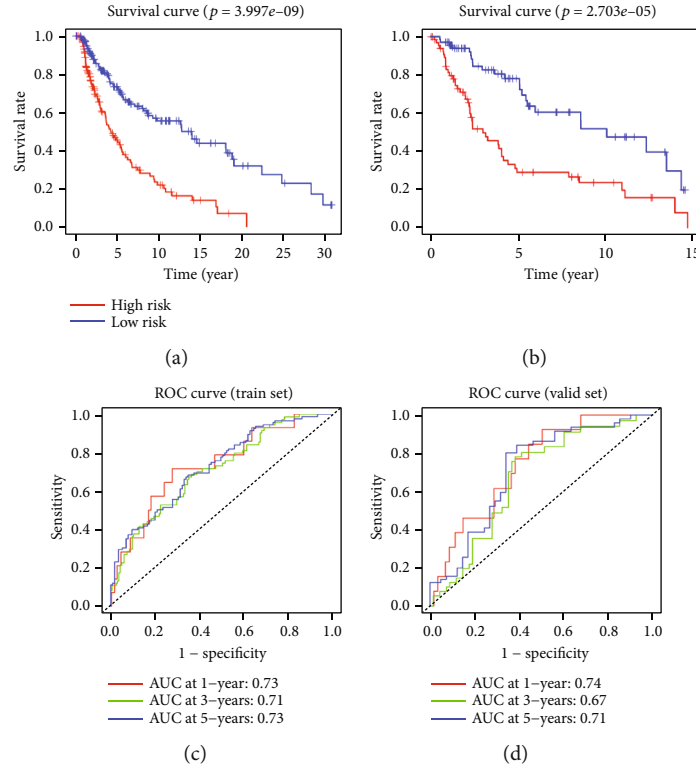


FIGURE 8: Validation of SKCM risk model. (a and b) Survival curve analysis of training cohort and validation cohort. (c and d) ROC curve analysis of training cohort and validation cohort.

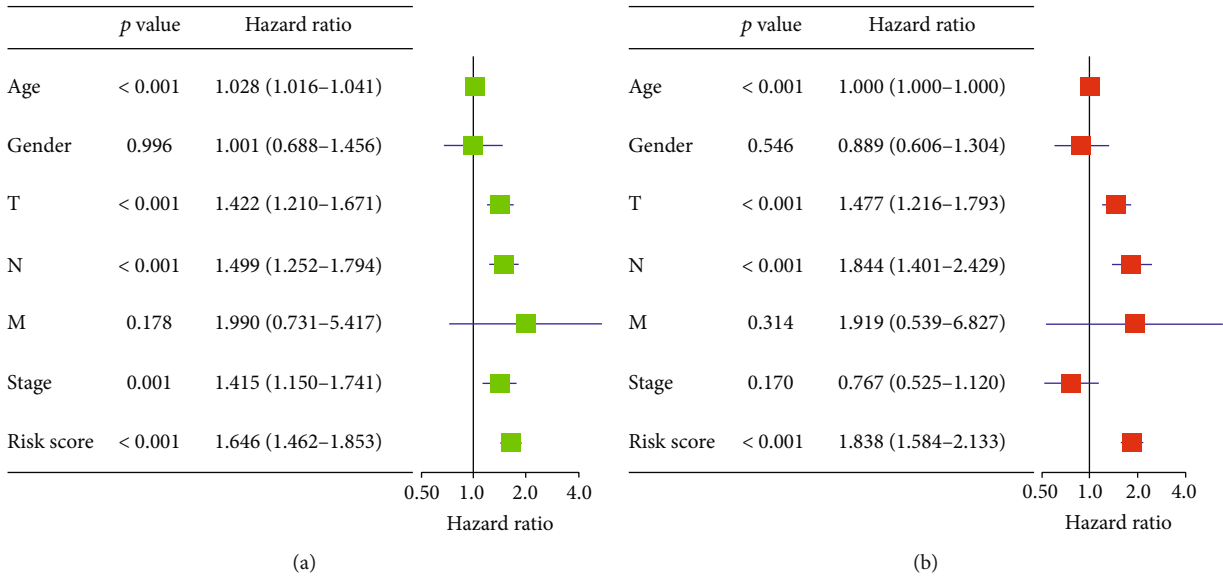


FIGURE 9: Univariate and multivariate validation of risk model’s independence. (a) Univariate analysis validates independence of SKCM risk model. (b) Multivariate analysis validates independence of SKCM risk model.

high methylation of cg14845689, cg20895058, cg06506470, and cg22337437 can increase prognostic risk. After searching regulatory genes corresponding to these methylation sites, we found that these sites were located in EN2, EVC, AL590807.1, GALNT2, AC091167.6, EMX2, AL049874.3, SDK1, and NLRP12 genes, respectively. EN2 as a transcription factor was

found to be a biomarker for prostate cancer and breast cancer in relevant studies [38, 39]. Current research considered that highly expressed EN2 can promote the proliferation and drug resistance of cancer cells [40, 41]. In this study, it was found that cg01447831 high methylation could reduce the prognostic risk of patients, which may be because cg01447831 can inhibit

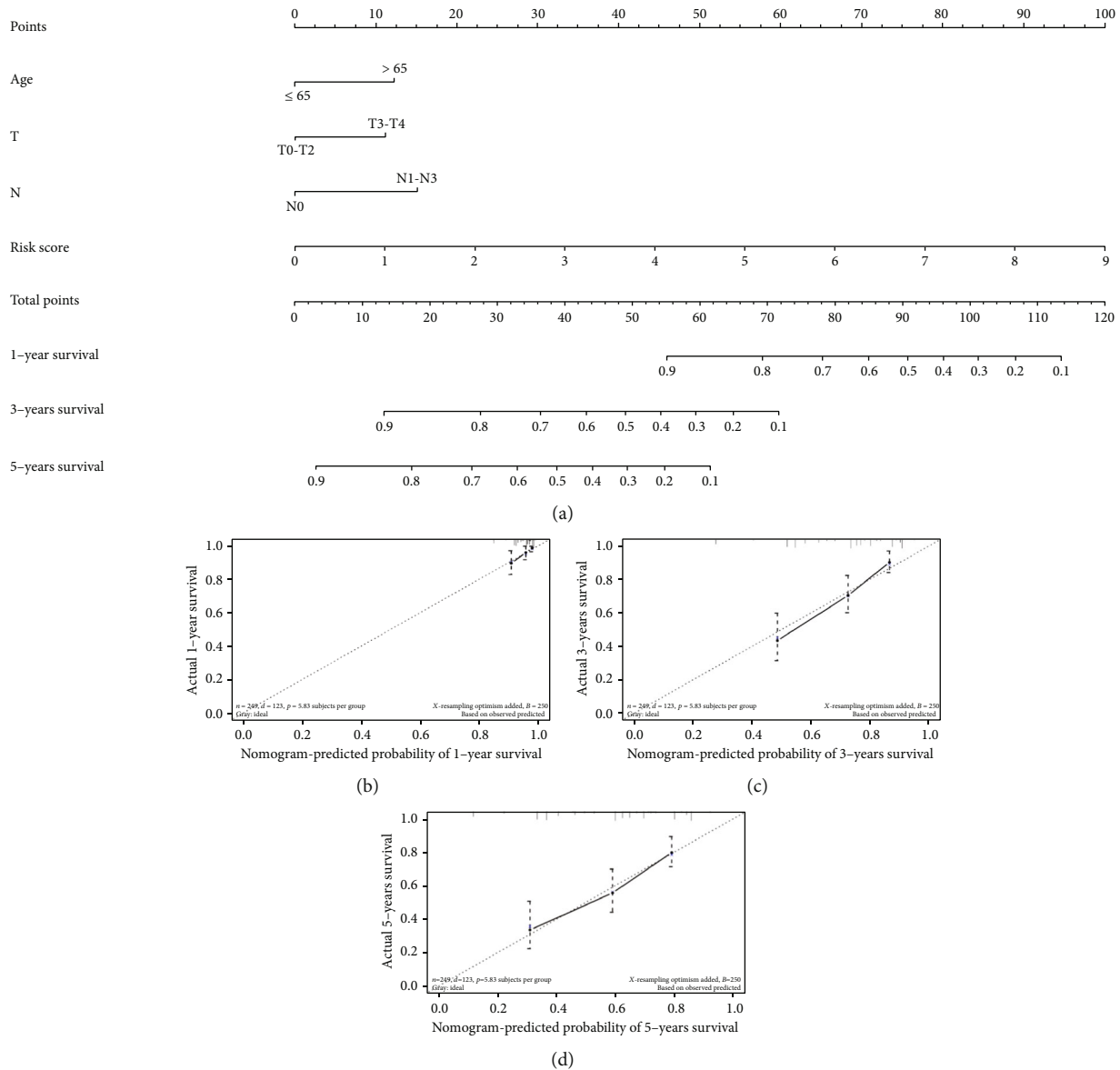


FIGURE 10: Construction and validation of prognostic nomogram. (a) Nomogram to predict 1-, 3-, and 5-year survival rate of SKCM patients. (b-d) Fitting curves used to validate prognostic nomogram.

the expression of EN2. It was discovered that methylation site cg09558315 was in the driver zone of EVC, which indicated that EVC high expression may promote the occurrence of cancers. At present, few studies were undertaken on the carcinogenic mechanism of EVC, and studies considered that EVC mutation may be an inducement for Ellis-van Creveld syndrome and EVC may be a prognostic marker of colon cancer in the early stage [42, 43]. GALNT2 gene is a Mucin O-glycosylase which is disputed for its effect in cancers. A study thought that GALNT2 may promote cancer progression via activating EGFR/PI3K/Akt/mTOR pathway [44]. Furthermore, it was also considered that GALNT2 can inhibit cancer development by reducing MET phosphorylation in some conditions [45], and it was found that cg17737409 site methylation in GALNT2 was helpful for patient's prognosis in this study. Moreover, we found that high methylation

silencing of cg14845689 in EXM2 could increase patient's prognostic risk. EXM2 is thought as a tumor suppressor gene, and high expression of EXM2 can induce cell cycle stagnation thereby inhibiting the proliferation of cancer cells [46, 47]. SDK1 is a kind of immune-related protein, and studies found that SDK1 mutation causes cancers, but the effect of SDK1-related site methylation on the incidence of cancer has not been fully studied [48, 49]. Our study found that high methylation of cg06506470 site in SDK1 gene could cause the poor prognosis of patients. NLRP12 is an immune-sensor element that triggers an inflammatory response, causing the release of IL-1B and IL-18, and the cleavage and activation of Caspase-1 [50]. After screening methylation sites, we identified prognostic model and established a prognostic nomogram that could assist clinical doctors to determine patient's prognosis.

All in all, we analyzed related data in TCGA-SKCM dataset, screened 9 prognosis-related methylation sites, and established a prognostic model by univariate, multivariate, and LASSO Cox regression analyses, unsupervised clustering analysis, and WGCNA. The model was accurate for determining patient's prognosis. The model screened by experiments is reliable validated by data in the validation set, but this paper is a bioinformatics essay solely without experimental data support. Therefore, more animal and clinical experiments are needed to support the conclusion in this paper to be clinically used.

Data Availability

All data generated or analyzed during this study are included in this article.

Conflicts of Interest

The authors declare that they have no conflicts of interest with the contents of this article.

Authors' Contributions

Xiaoming Yu was responsible for the conceptualization, formal analysis, and project administration. Ping Cong carried out data curation and investigation. Wei Wei was responsible for the methodology and software. Yong Zhou wrote the original draft and validated the study. Zhengqiang Bao wrote, reviewed, and edited the manuscript. Huaying Hou wrote the original draft and provided resources.

Acknowledgments

This work was supported by the Science and Technology Development Plan of Jinan City (No. 201805087) and the Key Research and Development Plan of Shandong Province (No. 2016GSF201065).

Supplementary Materials

Table S1: the clinical data was accessed from TCGA-SKCM dataset. Table S2: the result of multivariate Cox regression for the methylation sites based on SKCM patients. (*Supplementary Materials*)

References

- [1] W. H. Ward, F. Lamberton, N. Goel, J. Q. Yu, and J. M. Farma, *Cutaneous Melanoma: Etiology and Therapy*, W. H. Ward and J. M. Farma, Eds., Brisbane (AU): Codon Publications, Australia, 2017.
- [2] S. Craig, C. H. Earnshaw, and A. Viro, "Ultraviolet light and melanoma," *The Journal of Pathology*, vol. 244, no. 5, pp. 578–585, 2018.
- [3] Y. H. You, D. H. Lee, J. H. Yoon, S. Nakajima, A. Yasui, and G. P. Pfeifer, "Cyclobutane pyrimidine dimers are responsible for the vast majority of mutations induced by UVB irradiation in mammalian cells," *The Journal of Biological Chemistry*, vol. 276, no. 48, pp. 44688–44694, 2001.
- [4] T. Budden and N. A. Bowden, "The role of altered nucleotide excision repair and UVB-induced DNA damage in melanoma-genesis," *International Journal of Molecular Sciences*, vol. 14, no. 1, pp. 1132–1151, 2013.
- [5] R. Greinert, B. Volkmer, S. Henning et al., "UVA-induced DNA double-strand breaks result from the repair of clustered oxidative DNA damages," *Nucleic Acids Research*, vol. 40, no. 20, pp. 10263–10273, 2012.
- [6] J. W. Lee, S. Park, H. K. Han, M. C. Gye, and E. Y. Moon, "Di-(2-ethylhexyl) phthalate enhances melanoma tumor growth via differential effect on M1-and M2-polarized macrophages in mouse model," *Environmental Pollution*, vol. 233, pp. 833–843, 2018.
- [7] A. Kiss, C. Wei, Z. Aligabi et al., "693 p38 signaling regulates human cutaneous metastatic melanoma (MM) invasion and MM-dependent disruption of keratinocyte differentiation," *Journal of Investigative Dermatology*, vol. 140, no. 7, p. S93, 2020.
- [8] C. B. Begg, I. Orlov, A. J. Hummer et al., "Lifetime risk of melanoma in CDKN2A mutation carriers in a population-based sample," *Journal of the National Cancer Institute*, vol. 97, no. 20, pp. 1507–1515, 2005.
- [9] D. T. Bishop, F. Demenais, A. M. Goldstein et al., "Geographical variation in the penetrance of CDKN2A mutations for melanoma," *Journal of the National Cancer Institute*, vol. 94, no. 12, pp. 894–903, 2002.
- [10] N. E. Thomas, S. N. Edmiston, Y. S. Tsai et al., "Utility of TERT promoter mutations for cutaneous primary melanoma diagnosis," *The American Journal of Dermatopathology*, vol. 41, no. 4, pp. 264–272, 2019.
- [11] D. Sarkar, E. Y. Leung, B. C. Baguley, G. J. Finlay, and M. E. Askarian-Amiri, "Epigenetic regulation in human melanoma: past and future," *Epigenetics*, vol. 10, no. 2, pp. 103–121, 2015.
- [12] F. Hoffmann, D. Niebel, P. Aymans, S. Ferring-Schmitt, D. Dietrich, and J. Landsberg, "H3K27me3 and EZH2 expression in melanoma: relevance for melanoma progression and response to immune checkpoint blockade," *Clinical Epigenetics*, vol. 12, no. 1, p. 24, 2020.
- [13] J. R. Fredricks and P. A. Bejarano, "Primary malignant melanoma of the esophagus with separate foci of melanoma in situ and atypical melanocytic hyperplasia in a patient positive for human immunodeficiency virus: a case report and review of the literature," *Archives of Pathology & Laboratory Medicine*, vol. 132, no. 10, pp. 1675–1678, 2008.
- [14] J. Nguyen, T. Alexander, H. Jiang et al., "Melanoma in patients with GATA2 deficiency," *Pigment Cell & Melanoma Research*, vol. 31, no. 2, pp. 337–340, 2018.
- [15] C. M. Olsen, H. J. Carroll, and D. C. Whiteman, "Estimating the attributable fraction for melanoma: a meta-analysis of pigmented characteristics and freckling," *International Journal of Cancer*, vol. 127, no. 10, pp. 2430–2445, 2010.
- [16] B. de Unamuno Bustos, R. Murria Estal, G. Pérez Simó et al., "Aberrant DNA methylation is associated with aggressive clinicopathological features and poor survival in cutaneous melanoma," *The British Journal of Dermatology*, vol. 179, no. 2, pp. 394–404, 2018.
- [17] M. Venza, M. Visalli, T. Catalano et al., "DNA methylation-induced E-cadherin silencing is correlated with the clinicopathological features of melanoma," *Oncology Reports*, vol. 35, no. 4, pp. 2451–2460, 2016.

- [18] J. C. Hassel, A. Sucker, L. Edler et al., “MGMT gene promoter methylation correlates with tolerance of temozolomide treatment in melanoma but not with clinical outcome,” *British Journal of Cancer*, vol. 103, no. 6, pp. 820–826, 2010.
- [19] K. Tomczak, P. Czerwinska, and M. Wiznerowicz, “The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge,” *Contemporary Oncology*, vol. 19, no. 1A, pp. A68–A77, 2015.
- [20] X. F. Zhu, B. S. Zhu, F. M. Wu, and H. B. Hu, “DNA methylation biomarkers for the occurrence of lung adenocarcinoma from TCGA data mining,” *Journal of Cellular Physiology*, vol. 233, no. 10, pp. 6777–6784, 2018.
- [21] E. Olkhov-Mitsel, A. J. Savio, K. J. Kron et al., “Epigenome-wide DNA methylation profiling identifies differential methylation biomarkers in high-grade bladder cancer,” *Translational Oncology*, vol. 10, no. 2, pp. 168–177, 2017.
- [22] M. Kuhn, “Building predictive models in R using the caret package,” *Journal of Statistical Software*, vol. 28, no. 5, pp. 1–26, 2008.
- [23] Y. Tian, T. J. Morris, A. P. Webster et al., “ChAMP: updated methylation analysis pipeline for Illumina BeadChips,” *Bioinformatics*, vol. 33, no. 24, pp. 3982–3984, 2017.
- [24] R. S. Zhou, E. X. Zhang, Q. F. Sun et al., “Integrated analysis of lncRNA-miRNA-mRNA ceRNA network in squamous cell carcinoma of tongue,” *BMC Cancer*, vol. 19, no. 1, p. 779, 2019.
- [25] J. Zheng, T. Zhang, W. Guo et al., “Integrative analysis of multi-omics identified the prognostic biomarkers in acute myelogenous leukemia,” *Frontiers in Oncology*, vol. 10, article 591937, 2020.
- [26] W. Liang, F. Sun, Y. Zhao, L. Shan, and H. Lou, “Identification of susceptibility modules and genes for cardiovascular disease in diabetic patients using WGCNA analysis,” *Journal Diabetes Research*, vol. 2020, article 4178639, 2020.
- [27] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for Cox’s proportional hazards model via coordinate descent,” *Journal of Statistical Software*, vol. 39, no. 5, pp. 1–13, 2011.
- [28] S. Li, S. Chen, B. Wang, L. Zhang, Y. Su, and X. Zhang, “A robust 6-lncRNA prognostic signature for predicting the prognosis of patients with colorectal cancer metastasis,” *Frontiers in Medicine*, vol. 7, p. 56, 2020.
- [29] C. Huang, Z. Liu, L. Xiao et al., “Clinical significance of serum CA125, CA19-9, CA72-4, and fibrinogen-to-lymphocyte ratio in gastric cancer with peritoneal dissemination,” *Frontiers in Oncology*, vol. 9, p. 1159, 2019.
- [30] Z. Jia, Y. Yan, J. Wang et al., “Development and validation of prognostic nomogram in ependymoma: a retrospective analysis of the SEER database,” *Cancer Medicine*, vol. 10, no. 17, pp. 6140–6148, 2021.
- [31] J. Gagan and E. M. Van Allen, “Next-generation sequencing to guide cancer therapy,” *Genome Medicine*, vol. 7, no. 1, p. 80, 2015.
- [32] X. R. Tang, Y. Q. Li, S. B. Liang et al., “Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study,” *The Lancet Oncology*, vol. 19, no. 3, pp. 382–393, 2018.
- [33] S. Müller, S. Raulefs, P. Bruns et al., “Next-generation sequencing reveals novel differentially regulated mRNAs, lncRNAs, miRNAs, sdrRNAs and a piRNA in pancreatic cancer,” *Molecular Cancer*, vol. 14, no. 1, p. 94, 2015.
- [34] Y. Wang, Y. Wang, Y. Wang, and Y. Zhang, “Identification of prognostic signature of non-small cell lung cancer based on TCGA methylation data,” *Scientific Reports*, vol. 10, no. 1, p. 8575, 2020.
- [35] S. H. Patel, L. M. Poisson, D. J. Brat et al., “T2-FLAIR mismatch, an imaging biomarker for IDH and 1p/19q status in lower-grade gliomas: a TCGA/TCIA project,” *Clinical Cancer Research*, vol. 23, no. 20, pp. 6078–6085, 2017.
- [36] P. Danaher, S. Warren, R. Lu et al., “Pan-cancer adaptive immune resistance as defined by the tumor inflammation signature (TIS): results from The Cancer Genome Atlas (TCGA),” *Journal for Immunotherapy of Cancer*, vol. 6, no. 1, p. 63, 2018.
- [37] J. Wu, Y. Cui, X. Sun et al., “Unsupervised clustering of quantitative image phenotypes reveals breast cancer subtypes with distinct prognoses and molecular pathways,” *Clinical Cancer Research*, vol. 23, no. 13, pp. 3334–3342, 2017.
- [38] R. Morgan, A. Boxall, A. Bhatt et al., “Engrailed-2 (EN2): a tumor specific urinary biomarker for the early diagnosis of prostate cancer,” *Clinical Cancer Research*, vol. 17, no. 5, pp. 1090–1098, 2011.
- [39] N. L. Martin, M. K. Saba-El-Leil, S. Sadekova, S. Meloche, and G. Sauvageau, “EN2 is a candidate oncogene in human breast cancer,” *Oncogene*, vol. 24, no. 46, pp. 6890–6901, 2005.
- [40] E. Gómez-Gómez, J. M. Jiménez-Vacas, S. Pedraza-Arévalo et al., “Oncogenic role of secreted engrailed homeobox 2 (EN2) in prostate cancer,” *Journal of Clinical Medicine*, vol. 8, no. 9, 2019.
- [41] Z. Zhao, W. Sun, Z. Guo, J. Zhang, H. Yu, and B. Liu, “Mechanisms of lncRNA/microRNA interactions in angiogenesis,” *Life Sciences*, vol. 254, article 116900, 2020.
- [42] T. Q. Nguyen, M. Saitoh, H. T. Trinh et al., “Truncation and microdeletion of EVC/EVC2 with missense mutation of EFCAB7 in Ellis-van Creveld syndrome,” *Congenital Anomalies*, vol. 56, no. 5, pp. 209–216, 2016.
- [43] X. Zhang, S. Wan, Y. Yu et al., “Identifying potential DNA methylation markers in early-stage colorectal cancer,” *Genomics*, vol. 112, no. 5, pp. 3365–3373, 2020.
- [44] Z. Sun, H. Xue, Y. Wei et al., “Mucin O-glycosylating enzyme GALNT2 facilitates the malignant character of glioma by activating the EGFR/PI3K/Akt/mTOR axis,” *Clinical Science (London, England)*, vol. 133, no. 10, pp. 1167–1184, 2019.
- [45] S. Y. Liu, C. T. Shun, K. Y. Hung et al., “Mucin glycosylating enzyme GALNT2 suppresses malignancy in gastric adenocarcinoma by reducing MET phosphorylation,” *Oncotarget*, vol. 7, no. 10, pp. 11251–11262, 2016.
- [46] A. Monnier, R. Boniface, R. Bouvet et al., “The expression of EMX2 lead to cell cycle arrest in glioblastoma cell line,” *BMC Cancer*, vol. 18, no. 1, p. 1213, 2018.
- [47] L. Wang, J. Jin, Y. Zhou et al., “EMX2 is epigenetically silenced and suppresses epithelial-mesenchymal transition in human esophageal adenocarcinoma,” *Oncology Reports*, vol. 42, pp. 2169–2178, 2019.
- [48] Y. Zhang, X. Y. Mao, X. Liu et al., “High frequency of the SDK1:AMACR fusion transcript in Chinese prostate cancer,” *International Journal of Clinical and Experimental Medicine*, vol. 8, no. 9, pp. 15127–15136, 2015.

- [49] S. Ren, Z. Peng, J. H. Mao et al., “RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings,” *Cell Research*, vol. 22, no. 5, pp. 806–821, 2012.
- [50] J. R. Lukens, P. Gurung, P. J. Shaw et al., “The NLRP12 sensor negatively regulates autoinflammatory disease by modulating interleukin-4 production in T cells,” *Immunity*, vol. 42, no. 4, pp. 654–664, 2015.

Research Article

A COVID-19 Auxiliary Diagnosis Based on Federated Learning and Blockchain

Ziyu Wang ¹, Lei Cai ¹, Xuewu Zhang ¹, Chang Choi ², and Xin Su ¹

¹College of IoT Engineering, Hohai University, Changzhou 213022, China

²Department of Computer Engineering, Gachon University, Seongnam-si 13120, Republic of Korea

Correspondence should be addressed to Chang Choi; changchoi@gachon.ac.kr

Received 1 June 2022; Revised 8 July 2022; Accepted 19 July 2022; Published 23 August 2022

Academic Editor: Lei Chen

Copyright © 2022 Ziyu Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the high transmission rate and high pathogenicity of the novel coronavirus (COVID-19), there is an urgent need for the diagnosis and treatment of outbreaks around the world. In order to diagnose quickly and accurately, an auxiliary diagnosis method is proposed for COVID-19 based on federated learning and blockchain, which can quickly and effectively enable collaborative model training among multiple medical institutions. It is beneficial to address data sharing difficulties and issues of privacy and security. This research mainly includes the following sectors: in order to address insufficient medical data and the data silos, this paper applies federated learning to COVID-19's medical diagnosis to achieve the transformation and refinement of big data values. With regard to third-party dependence, blockchain technology is introduced to protect sensitive information and safeguard the data rights of medical institutions. To ensure the model's validity and applicability, this paper simulates realistic situations based on a real COVID-19 dataset and analyses problems such as model iteration delays. Experimental results demonstrate that this method achieves a multiparty participation in training and a better data protection and would help medical personnel diagnose coronavirus disease more effectively.

1. Introduction

Epidemic novel coronavirus (COVID-19) [1] has been spreading since 2019, with high morbidity, transmission, and mortality rate. There is an urgent worldwide need for a prompt and accurate diagnosis, together with an effective treatment. Meanwhile, biomedical imaging has become an indispensable diagnosis tool in clinical trials [2, 3]. Currently, human-induced analysis to numerous images alone in medical institutions consumes a lot of human, material, and financial resources. Deep learning, as a forefront technology in the smart healthcare system, is bound to find a wide application in this field [4].

Implementing smart healthcare with traditional artificial intelligence approaches requires cloud servers or data centers to share local datasets. Such cloud servers have a computing power to provide efficient data training and analysis, whose abuse and malfunction may result in leakage of health information or compromise of sensitive data [5]. Even without authorization, attackers have a potential access

to AI centers or third parties, such as the cloud server, for data. Similarly, they may control and modify data without a client's awareness or permission [6], while an efficient AI model inevitably needs numerous data to support training [7, 8]. Data exchanges among medical institutions partially support model training, but diversified institutional policies, privacy issues, and high costs hinder dataset sharing between institutions for model training purposes [9]. Against this background, current issues in smart healthcare manifest as how to achieve a secure sharing of medical data, overcoming data silos, and protecting sensitive private information.

Federated learning (FL), first put forward in 2016 [10], enables data sharing as well as model coconstruction without leaking respective data privacy. After that, smart healthcare undergoes rapid development in the federated learning ecosystem. Xu et al. [11] provide federated learning application scenarios in biomedicine and confirmed its feasibility in smart healthcare. In a federated learning-based healthcare system, AI only needs to train local models to upload parameters, while a central server updates the global

picture, thus reducing the sensitive information leakage risk [12, 13]. Being immutable and traceable [14], blockchain is an effective tool to help decentralize federated learning. Timely updates of local model parameters are uploaded by clients to the blockchain's distributed ledgers, and these model updates are audited. All transaction records are tamper-evident and repudiation-resistant, with or without an authoritative inspection. In addition, differences in communicating and computing power of client devices cause model iteration delays [15, 16]. Blockchain helps trace back and correlate with individual participants, analyze block generation times, data block propagation times, and forking rates. Blockchain applied in the healthcare sector also enables timely electronic healthcare data processing. It protects sensitive patient information privacy and ensures secure data sharing among multiple organizations. Federated learning and blockchain jointly applied in the novel coronavirus diagnosis are likely to achieve a globally secure epidemic data sharing without sensitive information leakage, significantly contributing to epidemic control together with life and property safeguarding [17].

At present, the bottleneck concerning intelligent healthcare lies in the collection of multiple global datasets with diverse features, which are inherently difficult to integrate [18]. Healthcare organizations are also highly sensitive to data privacy and security. It becomes a sharp issue to guarantee large-scale model training accuracy and secure data use. This paper accordingly designs a new federated learning (FL) and Blockchain-Based COVID-19 Auxiliary Diagnostic Model (FB-COVID-19 AD) with the following research details and innovations.

- (1) Federated learning is applied to COVID-19 diagnosis and partially breaks the bottleneck of low model performance due to insufficient medical data. The framework proposed here allows medical staff to reap the benefits of the feature learning without privacy worries
- (2) In response to federated learning's reliance on third parties, a blockchain-distributed ledger is introduced to protect sensitive information. Moreover, a public chain is adopted to ensure users' rights and interests, and the proof of work mechanism (PoW) is adopted for transaction consensus. At the same time, the model iteration delay is analyzed, and blockchain forking is taken into account to maintain model performance stability
- (3) Based on real datasets to simulate actual situations, an independently and identically distributed (IID) experiment and nonindependently and nonidentically distributed (non-IID) experiment are designed in this paper. It verifies the system model's effectiveness and applicability. Private data from multiple medical institutions are soundly available and invisible in the training process, with the training efficiency significantly improved

The remaining paper follows this structure. Section 2 presents and compares relevant medical works in terms of

federated learning and other technologies. Section 3 introduces the FB-COVID-19 AD model and algorithms for the system operation in detail. Its effectiveness and applicability are compared and analyzed with other research results in Section 4. Section 5 summarizes the paper.

2. Related Works

2.1. Federated Learning-Related Medical Image Researches. The healthcare industry has been profoundly undergoing a digital transformation based on artificial intelligence and big data. It triggers higher requirements for healthcare data collection and sharing. With information silos being a primary challenge for smart healthcare, secure data sharing becomes a real challenge as they are from worldwide hospitals and other medical institutions. Federated learning shows some promise for smart healthcare [19]. As multiple hospitals store their datasets locally, the model's updated parameters uploaded to the central server alone complete a global model update [20].

Federated learning expands dataset sources and results in a significant improvement in medical data training quality. A horizontal federated learning (HFL) on medical images confirms the framework's outperformance over individually trained models by 6.3% on average [21]. A vertical federated learning (VFL) combined with deep learning used external data to validate the model's generalization ability [22]. So, federated learning is liable for medical image processing. Data privacy issues are tackled with a federated learning framework to address and verify its good performance on different neural network classification models [23]. However, a federated learning framework alone is adopted as a solution to data sharing and data security here. Attention is not attached to the protection of model parameters in federated learning yet. Given the sensitivity and privacy of healthcare data, this model alone fails to meet current and immediate needs. Huge system losses are also possible if an attacker abuses model parameters to derive data.

Privacy data from medical institutions may also get leaked through the gradient information. Similarly, label and membership information may be launched by an attacker while issuing the uploaded model parameters. Differential privacy techniques are commonly applied to federated learning with regard to various privacy securities at the user's end. Federated learning on medical devices and differential privacy techniques (DP) in the system are adopted to protect information during model exchanges [24]. On this basis, a trade-off between privacy cost and model accuracy is implemented with Gaussian noise [25]. A Chinese study applies model parameters with Gaussian noise on the server side and adopts an adaptive gradient cropping strategy to enhance model performance [26]. A multisource diffusion image dataset accounts for variation between clients [27]. Each medical institution in these studies employs a generative adversarial network (GAN) to transform the original images into the target space to address the cross-client privacy protection. These articles all deal with data privacy protection through federated learning models and break data

TABLE 1: Comparison of healthcare systems programs based on federated learning.

Ref.	FL type	Technology	Clients	Aggregator	Dataset type	Application
[19]	HFL	/	Hospital	Data center	Acute neurological disorders	Object detection
[21]	HFL	/	Medical sites	Federated sever	Breast density classification	Image classification
[22]	VFL	/	Hospital	Federated sever	COVID-19	Object detection
[23]	HFL	/	Hospital	Federated sever	Pneumonia	Image classification
[24]	HFL	DP	MRI machines	Federated sever	Brain tumour	Image segmentation
[25]	HFL	DP	Hospital	Data center	Diabetic retinopathy	Image classification
[26]	HFL	DP	Medical sites	Federated sever	Multitype lesion map	Object detection
[27]	VFL	GAN	Hospital	Cloud sever	Prostate cancer	Image classification
[31]	VFL/HFL	Blockchain	Smart service	Blockchain	/	/
[33]	HFL	Blockchain	Hospital	Blockchain	COVID-19	Image segmentation/class
[34]	HFL	Blockchain	Medical sites	Blockchain	COVID-19	Image segmentation
[35]	HFL	Blockchain	Hospital	Blockchain	MNIST	Image classification

barriers between different medical institutions. Some consider algorithm optimization with differential privacy techniques to achieve either client-side or server-side data protection. Unfortunately, the central server itself remains untouched once it fails or gets paralyzed by attacks. The overall system model optimization also lacks, unable to ensure a system model’s optimal performance.

2.2. Blockchain Involved Medical Image Researches. Section 2.1 discusses federated learning used to enable multiple parties to share data, only which it has to address dependency on third parties. Blockchain, as the underlying technology of Bitcoin, is essentially a decentralized database [28, 29]. Also, it is a technical solution to collectively maintain a reliable database using decentralization and distrust. Blockchain has brought together new technologies such as cryptography, distributed storage, smart contracts, and consensus algorithms. Blockchain combined in the healthcare sector is in line with the technological needs to support complex application scenarios [30]. The use of blockchain in a distributed federated learning healthcare system has already been studied to beat untrustworthy servers and external attacks, coordinating global model update computations in a peer-to-peer manner through institutions’ block consensus [31].

Concerns in healthcare data mainly point to patient information sensitivity and dataset finiteness [32]. Blockchain has essential characteristics like “decentralization,” immutability, and anonymity, which effectively make up for the gap that federated learning relies on central nodes. In the long run, combining federated learning and blockchain helps artificial intelligence to develop in the medical field. Consortium chain and block authorization methods are used in federated learning, where each institution uses the local COVID-19 dataset to train the model and only uploads model parameters such as gradient updates [33]. In this method, a miner’s failure to complete learning affects its connected clients in the blockchain. To deal with the issue, the miner unites other miners’ clients to confirm the transaction [34]. In addition, a reputation value calculation

method based on dual subjective logic is used in client screening to incentivize clients to participate in training [35].

Concentrating information on one server potentially brings about potential risks like attacks and unfairness to the system. As an indispensable tool for doctors’ diagnosis, medical images have high data privacy protection requirements. In view of these, federated learning and blockchain combined well to meet diagnosis needs with medical images [36]. FL opens new opportunities for smart healthcare, but relevant literature is insufficient in terms of the system model latency and block forking, etc. Besides an adverse impact on the transaction confirmation time, they also have negative impacts on the system stability and global model accuracy.

This paper summarizes system solutions in medical imaging utilizing the federated learning framework, as shown in Table 1. A comprehensive comparison of existing medical diagnostic researches based on federated learning is presented in terms of technological coordination, client side, aggregation side, and application scope, respectively.

As summarized in Table 1, most healthcare systems use horizontal federated learning as the model framework to combine multiple medical institutions with the same characteristics for multisample learning. Most system models in the literature use common medical image datasets to evaluate the training effect and hopefully improve farsighted FL healthcare systems. Besides, regarding data privacy protection, some researches compensate for the shortcomings of FL in sensitive data protection by combining differentiated privacy protection and blockchain technology, etc. In brief, FL combined with other technologies is a very effective learning approach to accelerate the AI model training accuracy rates.

3. Materials and Methods

3.1. FB-COVID-19 AD Model

3.1.1. Introduction. As shown in Figure 1, FB-COVID-19 AD sets multiple medical institutions as the client, the blockchain composed of public chains as the server. Apart from

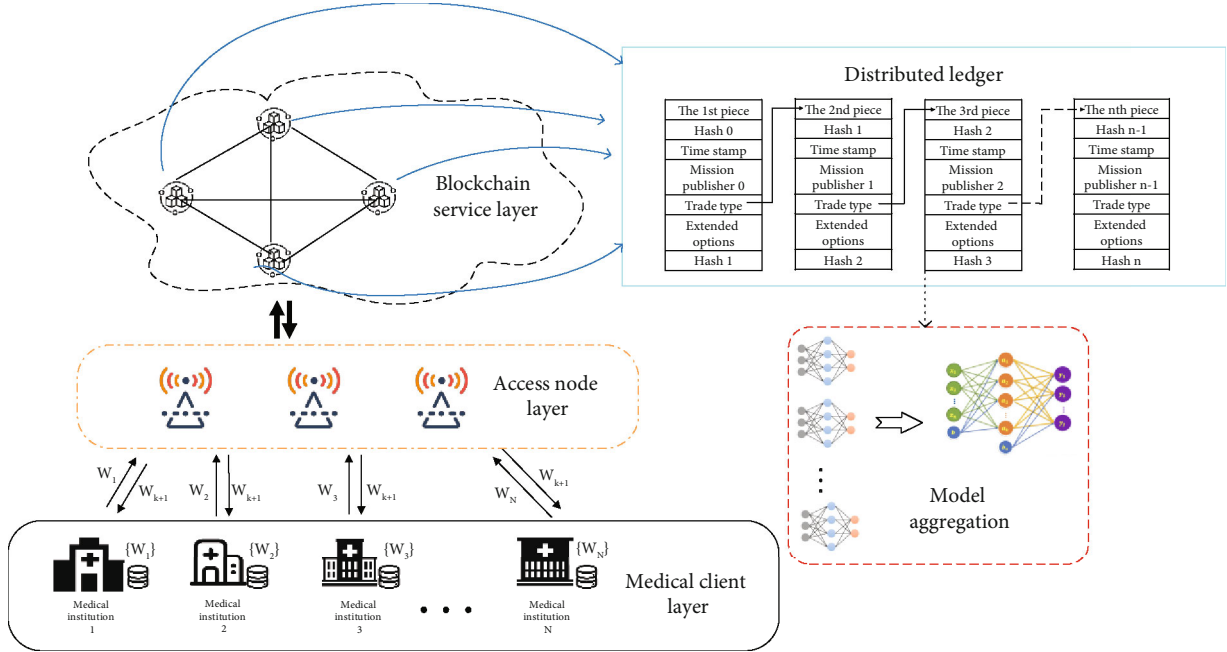


FIGURE 1: FB-COVID-19 AD.

that, data exchanges between the client and the server are realized at an access node layer. Different medical institutions are successfully united to train the pneumonia image classification model. Each component is described in detail as follows:

- (1) **Medical client layer:** any owner of local medical image datasets. Based on the local dataset, each client trains models and uploads model parameters. At the same time, clients in this layer also accept global update parameters to export global model updates
- (2) **Access node layer:** a communication infrastructure with computing power. It receives model parameters uploaded by individual medical institutions, as well as downlinks block information from the blockchain service layer
- (3) **Blockchain service layer:** a complete transaction data owner and verifier. Using distributed data storage, several complete transaction data sets are stored in each block, and parameters are authenticated using consensus protocols

3.1.2. Training in Medical Institutions. In FB-COVID-19 AD, since multiple medical institutions jointly participate in model training, they only need to provide model parameters to data requestors. Given the practicalities, $\text{data}_{n,i} = \{x_{n,i}, y_{n,i}\}$ is the local dataset belonging to each medical institution, where $x_{n,i}$ is the input sample vector of the n th hospital added to the current federated training. Similarly, $y_{n,i}$ is its corresponding label vector. $\text{Data} = \sum_{n \in N} \sum_{i \in I} \text{data}_{n,i}$ is the total dataset participating in the current federated training, where $i \in I$ denotes sample i , $n \in N$ denotes medical institution n , and N denotes the set of all

medical institutions. At the same time, W_n is the model parameters of the n th medical institution participating in the local training, and the local model update is uploaded to the blockchain node miner M_i . Likewise, the set of input samples and the corresponding set of tags are set to X_n, Y_n . In the model training process, F_n is the objective function of the n th medical institution, and $f_n(w)$ is the corresponding local loss function, which is defined as follows:

$$F_n \triangleq \min \left\{ \frac{1}{N} \cdot \sum_{n \in N} f_n(w) \right\}. \quad (1)$$

Using optimizers to find optimal values help solve the above equation minimization and improve model accuracy. In a traditional FL setup, it is the local updates that are uploaded to the federated server. For the FB-COVID-19 AD system model, it is the updates that are saved in the candidate blocks to be mined. Setting the learning rate to e , the model parameters for the t th iteration update in the k th round of the healthcare facility n are formulated as

$$w_n^{t,k} = w_n^{t-1,k} - e \cdot \nabla f(w_n^{k-1}, x_{n,i}, y_{n,i}). \quad (2)$$

At the same time, the k round updated global weights are formulated as

$$w^k = \frac{1}{N} \cdot \sum_{n=1}^N \sum_{i=1}^I \frac{w_n^k \cdot \text{data}_n}{\text{Data}}, \quad (3)$$

$$\nabla f(w^k) = \frac{1}{N} \cdot \sum_{n=1}^N \sum_{i=1}^I \nabla f(w_n^k, x_{n,i}, y_{n,i}).$$

The updated local model gradient parameter $\{w_n^k, \nabla f(w^k)\}$ is then uploaded to the server-side blockchain. The federated averaging algorithm (FedAvg) [9] is used to compute the global model gradient. When a new block is generated, the medical institution downloads the data about the new block from the miner and uses its aggregated results to compute w^k to update the model. In FL setup, the medical institution iterates the process until it satisfies the inequality: $|f(w^k) - f(w^{k-1})| \leq \delta$, where $\delta \in R$ is the loss of precision.

3.1.3. Blockchain Aggregation. By using the value representation and transfer function from blockchain, the value created by a local medical institution is tagged while the benefits are distributed. Each medical institution has its corresponding miners, all of whom are used as nodes to form the blockchain. The device uploads updated model parameters $\{w_n^k, \nabla f(w^k)\}$ and the computing time used $T_{local,n}^k$ to its corresponding miners. To prevent unauthorized block copying, miners store the defined data structure Merkle trees and create new blocks at specific times, which contain unique hash values, block generation rate λ , Nonce values, timestamps, etc. To achieve a consistent state of the chain, all miners apply the consensus protocol PoW to implement mining to verify the new blocks. When mining is complete, the verified blocks are placed in candidate blocks and broadcast across the network. At this moment, parameters in the local model update are safely stored in the blockchain. The information from those accepted blocks is then used to derive the parameters of the global model update $\{w^k, \nabla f(w^k)\}$ for calculation. Clients in the model exchange local information via the blockchain, which keeps track of individual model updates and does not implement model aggregation via third parties. Through decentralized learning, the global model is constructed by local medical institutions or miners with sufficient computational power. The global model calculation is completed with individual medical institutions in this paper.

The consensus mechanism is the blockchain's core technology. It allows transactions to be conducted safely and reliably in a centralized network [37]. A public chain is used to achieve complete decentralization, and PoW is used as the consensus protocol to achieve distribution in correspondence to its workload. Miners compete for arithmetic power, constantly searching for the random number Nonce through their own arithmetic power [38]. Only the licensed miner has the right for bookkeeping, who first finds the Nonce value and broadcasts it to the whole network for recognition. The packaged candidate block is broadcasted to the network as a new block. Then, other miners stop computing and add the new block to the local ledger once they receive the message.

PoW provides a high degree of decentralization and robustness for large-scale blockchain. But in the process, different nodes tend to have different perspectives towards the blockchain due to differences in their arrival times. Hence, the process is termed as "Fork" (Figure 2). A chain fork may cause some clients to receive incorrect global model updates. Inevitably, it brings a negative effect to the next

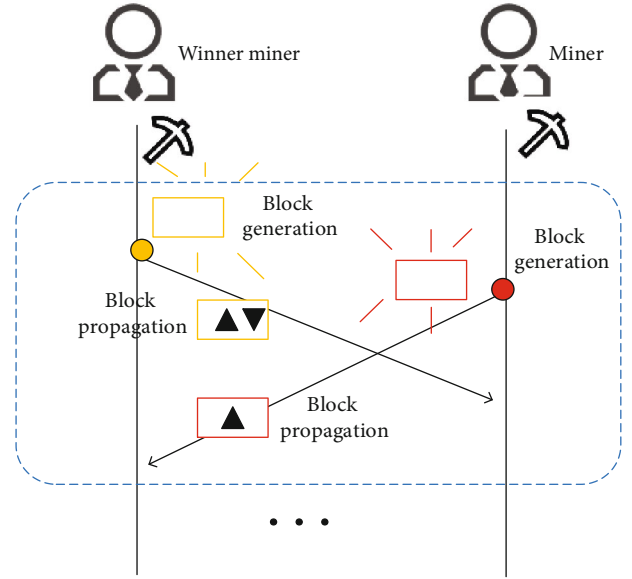


FIGURE 2: Blockchain fork.

round of local model updates. Assuming that all miners work in parallel, the probability of the fork is defined as

$$p_f = 1 - e^{-\lambda(M-1)\beta_{BP}}, \quad (4)$$

where λ is the block generation rate and $1/\lambda$ is the expected time for a block generation. Besides, M is the number of miners, and β_{BP} is the block body propagation delay. So, the block mining time is expressed as a Poisson process.

In the FB-COVID-19 AD process, the server-side blockchain provides a credible incentive for contribution to encourage more medical institutions to join and improve the data processing accuracy. All medical institutions are timely rewarded with a reference to their data quantity and quality. Simultaneously, miners M_i are rewarded according to the quantity and quality of data $_{n,i}$ from the local medical institutions they are associated with. By writing the reward resources into the blockchain, the blockchain openness and transparency will attract more clients to join and improve the collaboration efficiency among medical institutions.

3.2. FB-COVID-19 AD Algorithm

3.2.1. FB-COVID-19 AD Single Iteration Process. As shown in Figure 1, decentralization is realized through federated learning and blockchain combination. In FB-COVID-19 AD, FL clients' devices interact with the local model update parameters through the blockchain to achieve model aggregation without a third-party server. A single iteration of the local healthcare facility is implemented applying the following steps.

Local calculation: using the local dataset data $_{n,i}$, each medical institution calculates w_n^k through a local optimization mechanism.

Data upload: a miner is randomly connected to medical facilities which uploads the calculated parameters $\{w_n^k, \nabla f(w^k)\}$ and time $T_{\text{local},n}^k$ to the corresponding miner M_i .

Block generation: while waiting for enough transactions to be received (determined by the maximum block size) or exceeding the maximum waiting time τ , a candidate block is generated for mining.

Block mining: with the consensus mechanism PoW, a random number Nonce is calculated to determine the next block to be added to the blockchain, or to receive a block broadcast by another miner.

Block propagation: $M_i \in M$ is the first miner to complete the calculation. Propagate M_i the candidate block upload message through the P2P network.

Block download: the medical institution downloads the latest block via the corresponding miner M_i .

Global model aggregation: local medical institutions use the information in the accepted blocks (local model updates) to calculate and derive global model updates.

The above operation steps continue until inequality $|f(w^k) - f(w^{k-1})| \leq \delta$ is satisfied. In step 7 of the above iteration, a global model update is done by the client, based on the differences in the miners' computing power. In fact, a failure is liable to affect the global model's update. For the IID dataset, the global model is initially downloaded by each medical institution itself and then updated using local data. Once the local model is updated, values of the model update parameters are uploaded to the blockchain to generate a new block to mine. The latest block will contain the global model parameters updated using all the local update parameters, so the block size is adapted to fit all update values. For the non-IID dataset, the medical institution submits the model update parameter values asynchronously after completing the local model update. A new block is generated when enough transactions have been received or when the maximum wait time τ has been exceeded. The specific algorithms for the FB-COVID-19 AD system model for the IID dataset and the non-IID dataset are shown in Algorithm 1 and Algorithm 2.

In the above steps, a block generation is determined by block size S_B and waiting time τ . Block size S_B is an important parameter in FB-COVID-19 AD. The larger the block size, the higher the information quality provided by a single iteration in federated learning, but the higher the fork rate and iteration latency time. Setting the block size correctly will affect the blockchain throughput and thus the completion time of federated learning. Block size optimization does not fall within the focus of this paper and is ignored for the time being.

3.2.2. FB-COVID-19 AD Iterative Latency Analysis. Due to equipment performance differences, the network, and other factors, this section will focus on analyzing the running time of the FB-COVID-19 AD single iteration process in Section 3.2.1 and defining following delay time parameters:

- (1) Block fill delay β_{BF} : a block is filled with data from step 2 to upload the local update. This contains the

```

Initialize:  $n \in N, e, \text{time} = 0, w^k = w^0, \delta$ 
1. while  $|f(w^k) - f(w^{k-1})| > \delta$ :
2.   for  $k \in K$ :
3.     for  $\forall n \in N$ :
4.        $w_n^{k+1} \leftarrow \text{LocalUpdate}(w^k, e, x_n, y_n)$ 
5.       UploadLocalUpdate( $S_n, C_n$ )
6.     end
7.     MineBlock( $\lambda, \gamma_i^M$ )
8.     TransmitBlock( $S_B, C_{Cap}$ )
9.      $w^{k+1} \leftarrow \text{GlobalUpdate}(w^k, w_n^{k+1}, e)$ 
10.     $k = k + 1$ 
11.  end
12.end

```

ALGORITHM 1: FB-COVID-19 AD (IID).

```

Initialize:  $n \in N, e, \text{time} = 0, w^k = w^0, \delta$ 
Medical Institutions(Asynchronous):
1. while  $|f(w^k) - f(w^{k-1})| > \delta$ :
2.    $w_n^{k+1} \leftarrow \text{LocalUpdate}(w^k, e, x_n, y_n)$ 
3.    $u_n^k \leftarrow \text{LocalUpdate}(S_n, C_n)$ 
BlockChain(Synchronous):
4.   if  $S_B \leq |u^t|$  or time  $> \tau$ :
5.     MineBlock( $\lambda, \gamma_i^M$ )
6.     TransmitBlock( $\min(|u^t|, S_B), C_{Cap}$ )
7.      $w^{k+1} \leftarrow \text{GlobalUpdate}(w^k, w_n^{k+1}, e)$ 
8.      $k = k + 1$ 
9.     Restart time
7.   else:
8.     Wait
9.   end
10.end

```

ALGORITHM 2: FB-COVID-19 AD (non-IID).

local model computation time (i.e., the time it takes for the client to update the local model gradient). Also, this includes the local model upload time (i.e., the time it takes for the client to hand over the model to the miner)

- (2) Block generation delay β_{BG} : the time miners spend looking up random values of Nonce, or waiting to receive a new valid block, as calculated by the PoW consensus mechanism
- (3) Block propagation delay β_{BP} : the full network propagation of the mined blocks is achieved through a P2P network, assuming that all miners receive the propagated blocks at the same time
- (4) Global model aggregation delay β_{AG} : the block downloading process in step 6 is executed first, and then, the client aggregates the local model information from the newly mined blocks to calculate the global model update

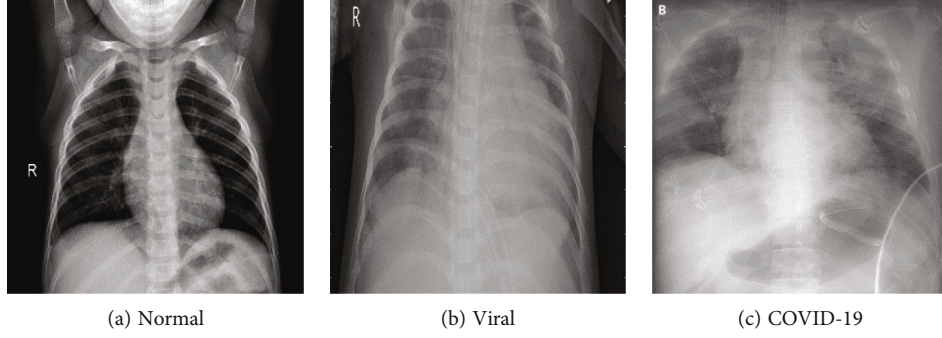


FIGURE 3: Images in the dataset.

- (5) Block download latency β_{BD} : the client downloads the latest block propagated from the miner, which contains either the local model updated in the last iteration or the global model. It depends on the model aggregation method

By calculating all delay times in the FB-COVID-19 AD process, and by considering the effect of bifurcation p_f , the expected iteration delay time was defined as:

$$T_{\text{iter}}^k = \frac{(\beta_{BF} + \beta_{BG} + \beta_{BP})}{1 - p_f} + \beta_{AG} + \beta_{BD}. \quad (5)$$

As block filling, block generation, and block propagation are all affected by the fork, these steps will be repeated once a conflict occurs.

4. Results and Discussion

4.1. The Dataset and Evaluating Indicators. A CT map dataset of COVID-19 is used (Figure 3) from the reference [39], created by collating from several public databases as well as recently published articles. A typical COVID-19 patient's chest image appears white on CT, and the shade depth correlates to its lesion gravity. At an earlier stage of the symptom, there may be multiple small patchy shadows as well as interstitial changes. Then, it would develop into multiple ground-glass shadows, infiltrative shadows, and lung consolidation in severe cases [3]. Accordingly, this dataset is filtered which contains 1313 CT maps of normal chests, 1316 CT maps of novel coronavirus diseases and 1171 CT maps of COVID-19 diseases. The entire dataset is divided into a training set, a test set, and a validation set. Accordingly, the basic information on the division of the dataset is shown in Table 2.

In this paper, the accuracy ACC is set as the evaluation index. The larger the ACC value, the higher the accuracy of the model's classification. As long as the samples are correctly predicted, the formula numerator will automatically increase by one, and the denominator will still represent all samples.

TABLE 2: Dataset distribution.

	Normal	Viral	COVID-19	Total
Train	1113	1116	971	3200
Test	180	180	180	540
Valid	20	20	20	60

TABLE 3: Confusion matrix.

Prediction Labe	Real label	
	Positive	Negative
Positive	True positive	False positive
Negative	False negative	True negative

TABLE 4: Software and hardware environments.

Term	Total
CPU	INTEL I9-12900K
GPU	NVIDIA RTX3090
Video memory	48 G
Internal memory	128 G
Operating system	Ubuntu 20.04
Development language	Python 3.7

$$\text{ACC} = \frac{\text{TP}_n + \text{TN}_n}{\text{TP}_n + \text{FP}_n + \text{FN}_n + \text{TN}_n}, \quad (6)$$

where TP_n , TN_n , FP_n , and FN_n represent true positive, true negative, false positive, and false negative for the corresponding healthcare facility clients, respectively, where n is the n th hospital. The confusion matrix is shown in Table 3.

4.2. Numerical Results and Discussion. The information about the implementation of our software and hardware platforms is shown in Table 4. The local medical institution uses residual networks to train model. Some training hyper-parameters were set as: the initial learning rate is 0.005, momentum is 0.5, and a ReLu function serves as the activation function. To ensure the comparison accuracy of experimental results, the time for each block generation in the model blockchain should remain as consistent as possible with the time for the parameter aggregation using the central server in FL based on FedAvg.

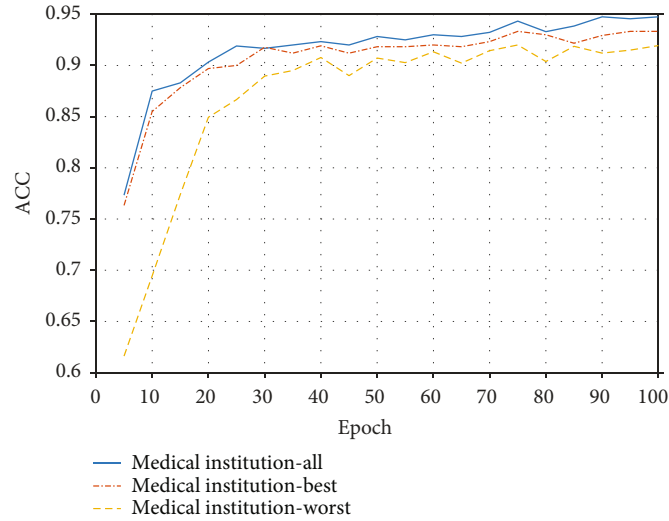


FIGURE 4: Differences in training accuracy of medical institutions.

To verify the effectiveness of FL to address the data silos, an experiment of training 20 clients with different sized IID samples is conducted. As is shown in Figure 4, we compare the classification accuracy of the proposed model using all distributed samples with the best model and the worst model of the local training clients. It is apparent that the model obtained from the joint multiparty participation training is better than the one trained by medical institutions alone. In terms of accuracy, the model training accuracy in this paper has a 1.01% margin higher than that of the medical institution with the best training effect on average. At the same time, the model training enjoys a 4.61% effectiveness margin on average in this paper over that of the worst institution's training effectiveness. It soundly confirms multiparty participation training with apparent improvements of the model training effect.

As is shown in Figure 5, an experiment is conducted to compare the centralized training model, the proposed model, and the FL model. Clients of FL models upload model parameters through parallel training and carry out aggregated computation through the FedAvg algorithm. It is clear that a blockchain introduction to FL for decentralization does not cause a negative impact on the classification performance from these accuracy curves.

In a practical scenario, samples of the dataset are distributed at different medical institutions with distinct quality, quantity, and nonindependent identically distribution. The difference in the training samples makes it difficult to ensure that the model update operation is consistent for each participant. In Table 5, centralized training using all samples has higher accuracy than both the proposed and FL model based on FedAvg using distributed samples. However, there is a potential for the asynchronous method in the FL process since it achieves similar performance when using non-IID datasets.

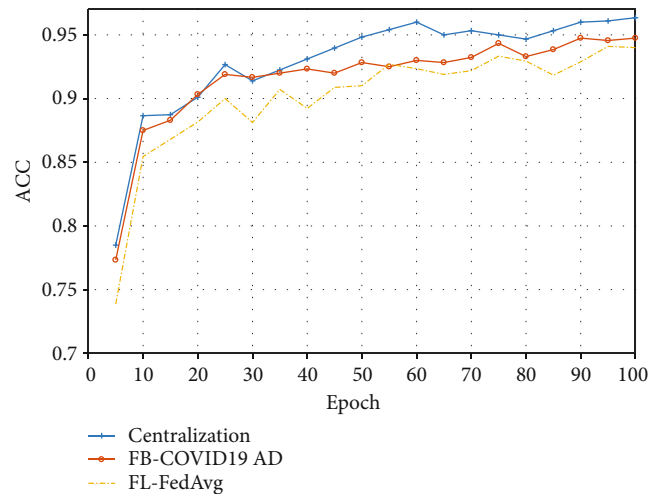


FIGURE 5: Three methods to train curves.

Due to PoW operation, the performance of the FB-COVID-19 AD model can be hindered by forks. The fork splits the BC into different states and potentially undermines the overall agreement between participants. In the experiment, different numbers of miners are set as $M = \{10, 100, 500\}$, and consider different link capacities $\text{Cap} = \{1\text{Mps}, 5\text{Mps}, 20\text{Mps}\}$. As is shown in Figure 6, the fork rate increases with the number of federated learning participants, which requires a change in the block size to contain more parallel transactions. At the same time, with the same number of miners, the fork rate decreases as the link capacity increases. This phenomenon becomes more pronounced when the number of miners $M = 100$ and $M = 500$, respectively. These results confirm that in larger model networks, the link capacity needs to be increased for the exchange between miners to reduce the fork rate. Thus, it helps to reduce exchange time during the process.

TABLE 5: The accuracy of three methods on COVID-19 dataset.

Method	Centralized training	FB-COVID-19 AD IID	FL-FedAvg IID	FB-COVID-19 AD non-IID	FL-FedAvg non-IID
Epoch = 20	0.901	0.903	0.881	0.728	0.738
Epoch = 40	0.931	0.923	0.892	0.847	0.835
Epoch = 60	0.960	0.931	0.923	0.917	0.903
Epoch = 80	0.946	0.933	0.929	0.932	0.923
Epoch = 100	0.963	0.947	0.940	0.933	0.923

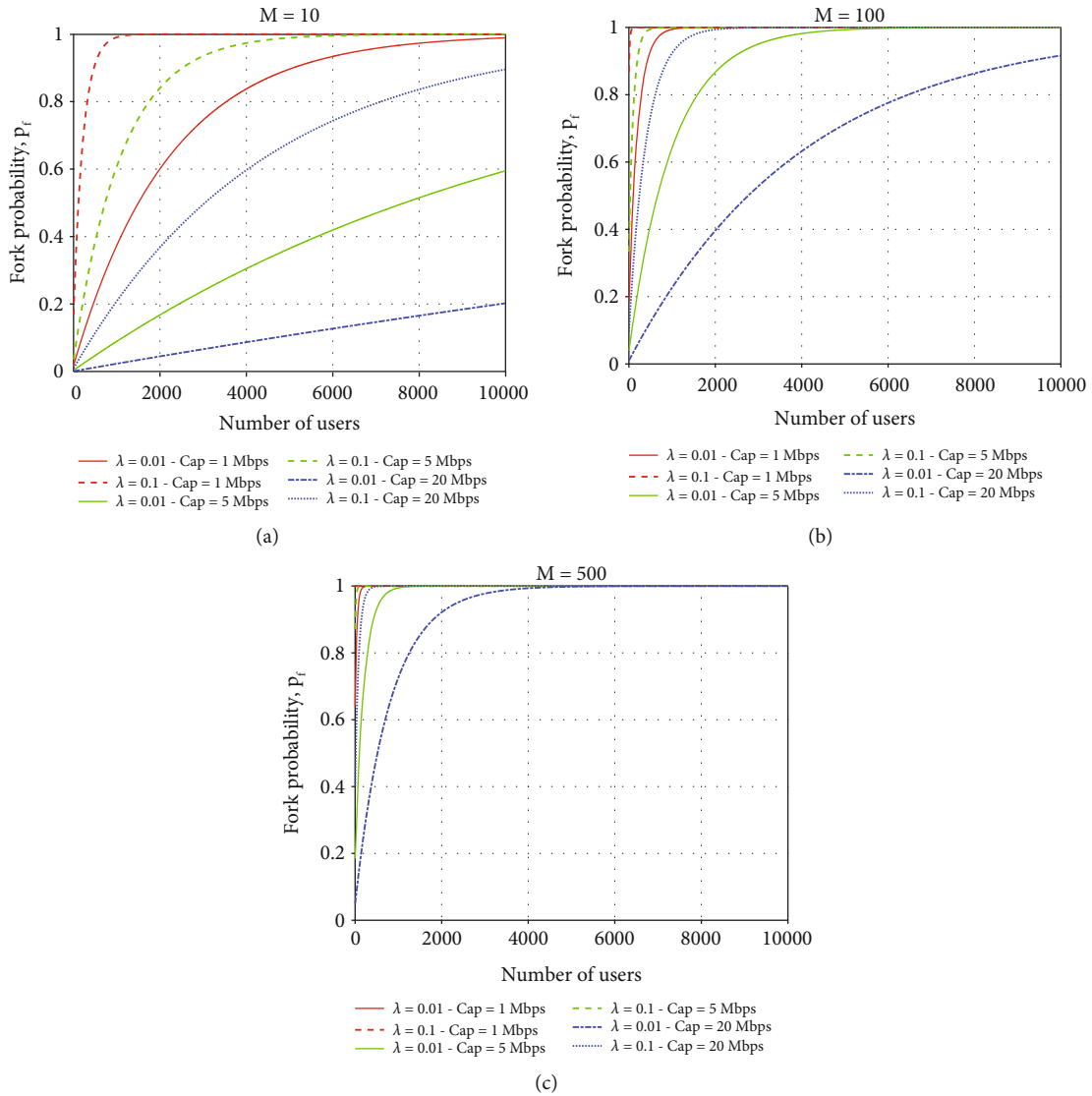


FIGURE 6: Fork rate versus number of clients and number of miners.

The relationship between iteration delay time T_{iter} , fork rate p_f , and block generation rate λ was analyzed for different link capacities. As is shown in Figure 7, $Cap = \{1\text{Mps}, 20\text{Mps}, 50\text{Mps}\}$ have three cases with different link capacities. In these cases, the fork rate is rising with the block generation

rate increase. Meanwhile, the iteration delay time decreases, and the trend manifests a concave shape. The comparison of the three link capacities illustrates that the link capacity increase leads to a reduced impact of the fork rate. The iteration delay time decreases as the link capacity increases.

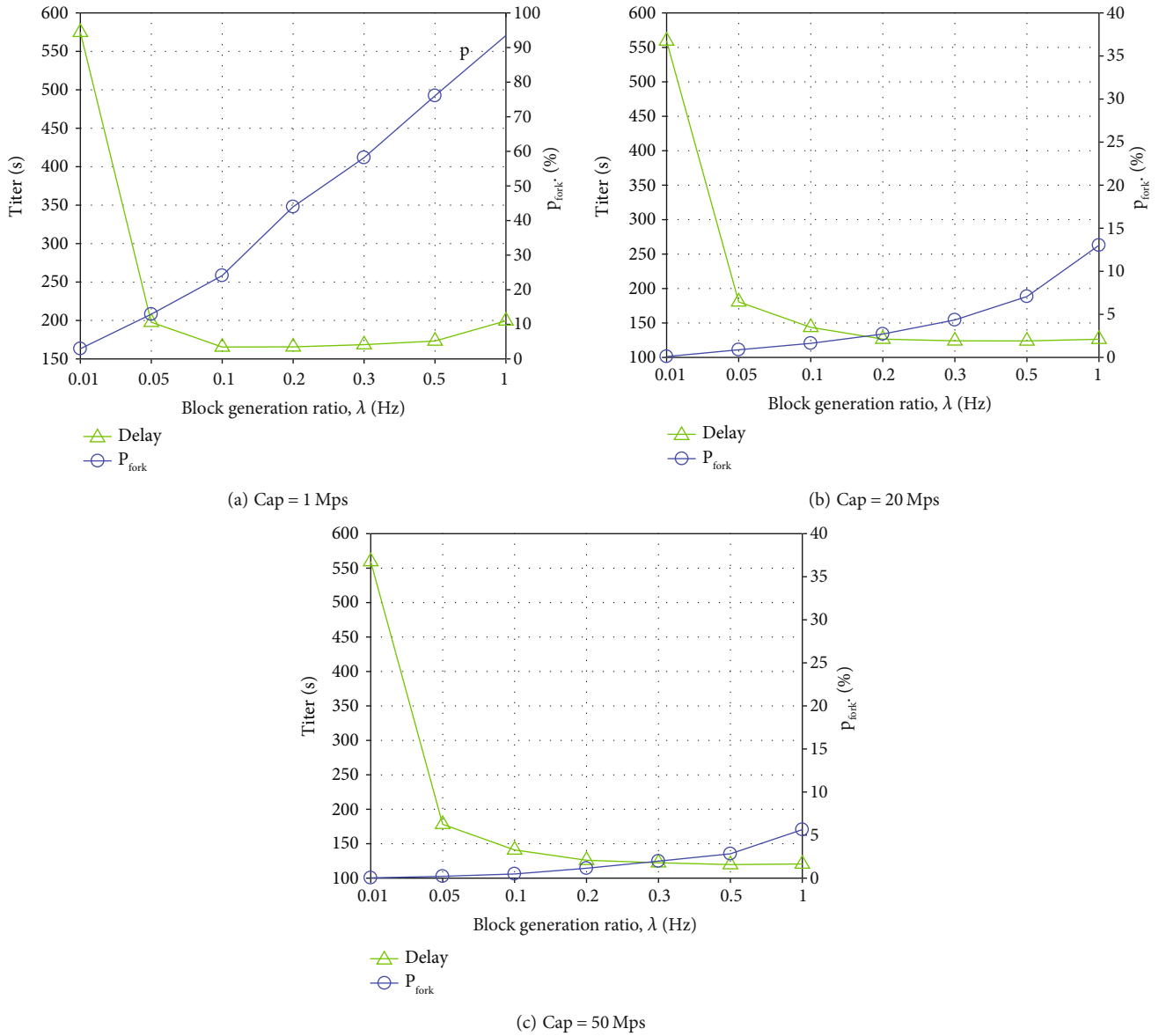


FIGURE 7: Delay time versus fork rate and block generation rate.

5. Conclusion

This paper mainly designs a method for auxiliary COVID-19 diagnosis through a combination of federated learning and blockchain technologies. Federated learning is used to address data barriers as well as insecure data sharing. Blockchain achieves decentralization, protects medically sensitive information, and encourages more medical institutions to participate in collaborative training. The proposed method is tested on a COVID-19 image dataset and achieves the expected training results. Compared with the centralized training, FB-COVID-19 AD model achieves collaborative cooperation among multiple participants without significant sacrifice of diagnostic accuracy. At the same time, experiments also investigate the single iteration latency of the system model and analyze the relationship between parameters such as fork rate, number of miners, block generation rate, and block capacity. In the future of smart healthcare, feder-

ated learning combined with other privacy and security technologies such as blockchain and differential privacy are expected to provide more innovative ideas for a secure sharing of healthcare data. Amidst the current global epidemic situations, federated learning and blockchain combined points to an important breakthrough for algorithmic models designed to assist epidemic diagnosis.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Authors' Contributions

Each author has made significant contributions to the conceptualization and development of this paper. Ziyu Wang, Lei Cai, and Chang Choi designed the research process. Xuewu Zhang gave a lot of guidance to the paper in the field of artificial intelligence and pointed out the key innovation direction of the paper. Xin Su used software for analysis and polished this article. All authors have read and approved the final version of the manuscript.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2021R1A2B5B02087169) and National Key R&D Program of China (2018YFC0407101).

References

- [1] J. Bedford, D. Enria, J. Giesecke et al., "COVID-19: towards controlling of a pandemic," *The Lancet*, vol. 395, no. 10229, pp. 1015–1018, 2020.
- [2] P. Wei, "Diagnosis and treatment protocol for novel coronavirus pneumonia (trial version 7)," *Chinese Medical Journal*, vol. 133, no. 9, pp. 1087–1095, 2020.
- [3] G. Hanxiong, X. Ying, S. Nanxi et al., "Clinical imaging features of novel coronavirus pneumonia (COVID-19)," *Radiology Practice*, vol. 35, no. 2, pp. 125–130, 2020.
- [4] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*, pp. 1273–1282, PMLR, 2017.
- [5] Y. Zhao, J. Zhao, L. Jiang et al., "Privacy-preserving blockchain-based federated learning for IoT devices," *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1817–1829, 2021.
- [6] J. Yang, J. Li, and Y. Niu, "A hybrid solution for privacy preserving medical data sharing in the cloud environment," *Future Generation Computer Systems*, vol. 43–44, pp. 74–86, 2015.
- [7] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Gan-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018.
- [8] S. Xia, Z. Yao, Y. Li, and S. Mao, "Online distributed offloading and computing resource management with energy harvesting for heterogeneous MEC-enabled IoT," *IEEE Transactions on Wireless Communications*, vol. 20, no. 10, pp. 6743–6757, 2021.
- [9] C. Xu, N. Wang, L. Zhu, K. Sharif, and C. Zhang, "Achieving searchable and privacy-preserving data sharing for cloud-assisted e-healthcare system," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8345–8356, 2019.
- [10] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, "Federated learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 13, no. 3, pp. 1–207, 2019.
- [11] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated learning for healthcare informatics," *Journal of Healthcare Informatics Research*, vol. 5, no. 1, pp. 1–19, 2021.
- [12] N. Rieke, J. Hancox, W. Li et al., "The future of digital health with federated learning," *NPJ Digital Medicine*, vol. 3, no. 1, pp. 1–7, 2020.
- [13] D. C. Nguyen, P. N. Pathirana, M. Ding, and A. Seneviratne, "Bedgehealth: a decentralized architecture for edge-based IoMT networks using blockchain," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11743–11757, 2021.
- [14] Y. Li, H. Ma, L. Wang, S. Mao, and G. Wang, "Optimized content caching and user association for edge computing in densely deployed heterogeneous networks," *IEEE Transactions on Mobile Computing*, vol. 21, no. 6, pp. 2130–2142, 2022.
- [15] Y. Li, C. Liao, Y. Wang, and C. Wang, "Energy-efficient optimal relay selection in cooperative cellular networks based on double auction," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4093–4104, 2015.
- [16] D. Preuveneers, V. Rimmer, I. Tsingenopoulos, J. Spooren, W. Joosen, and E. Ilie-Zudor, "Chained anomaly detection models for federated learning: an intrusion detection case study," *Applied Sciences*, vol. 8, no. 12, p. 2663, 2018.
- [17] B. A. Moustapha, "The effect of propagation delay on the dynamic evolution of the bitcoin blockchain," *Digital Communications and Networks*, vol. 6, no. 2, pp. 157–166, 2020.
- [18] Z. Che, Y. Cheng, S. Zhai, Z. Sun, and Y. Liu, "Boosting deep learning risk prediction with generative adversarial networks for electronic health records," in *Proc. of 2017 IEEE International Conference on Data Mining (ICDM)*, pp. 787–792, New Orleans, LA, USA, 2017.
- [19] U. C. Srivastava, D. Upadhyay, and V. Sharma, "Intracranial hemorrhage detection using neural network based methods with federated learning," *Computer Science*, 2020, <http://arxiv.org/abs/2005.08644>.
- [20] Z. Sun, J. Wan, L. Yin, Z. Cao, T. Luo, and B. Wang, "A blockchain-based audit approach for encrypted data in federated learning," *Digital Communications and Networks*, vol. 5, 2022.
- [21] H. R. Roth, K. Chang, P. Singh et al., "Federated learning for breast density classification: a real-world implementation," in *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning*, pp. 181–191, Springer, 2020.
- [22] Q. Dou, T. Y. So, M. Jiang et al., "Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study," *NPJ Digital Medicine*, vol. 4, no. 1, p. 60, 2021.
- [23] Y. Su, H. Zhang, and J. Liu, "Research on the classification method of pneumonia based on federated learning and deep learning," *Journal of Hefei Normal University*, vol. 39, no. 6, pp. 35–39, 2021.
- [24] W. Li, F. Milletari, D. Xu et al., "Privacy-preserving federated brain tumour segmentation," in *Machine Learning in Medical Imaging*, pp. 133–141, Springer, Cham, 2019.
- [25] M. Malekzadeh, B. Hasircioglu, N. Mital, K. Katarya, M. E. Ozfatura, and D. Gündüz, "Dopamine: differentially private federated learning on medical data," p. 2101, 2021, <http://arxiv.org/abs/2101.11693>.
- [26] S. Wang, S. Lu, and B. Cao, "A medical imaging target detection algorithm for privacy-preserving federated learning," *Journal of Computer-Aided Design & Computer Grap*, vol. 33, no. 10, pp. 1553–1562, 2021.
- [27] Z. Yan, J. Wicaksana, Z. Wang, X. Yang, and K.-T. Cheng, "Variation-aware federated learning with multi-source decentralized medical image data," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, pp. 2615–2628, 2020.

- [28] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, 2019.
- [29] P. J. Taylor, T. Dargahi, A. Dehghantanha, R. M. Parizi, and K. K. R. Choo, "A systematic literature review of blockchain cyber security," *Digital Communications and Networks*, vol. 6, no. 2, pp. 147–156, 2020.
- [30] Y. Yuan and F. Wang, "Current situation and prospect of blockchain technology development," *Acta Automatica Sinica*, vol. 42, no. 4, pp. 481–494, 2016.
- [31] J. Li, Y. Shao, K. Wei et al., "Blockchain assisted decentralized federated learning (BLADE-FL): performance analysis and resource allocation," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 10, pp. 2401–2415, 2022.
- [32] Y. Wu, W. Meng, Z. Yan, and V. Varadharajan, "Editorial: special issue on blockchain and communication networks," *Digital Communications and Networks*, vol. 6, no. 2, pp. 145–146, 2020.
- [33] R. Kumar, A. A. Khan, J. Kumar et al., "Blockchain-federated-learning and deep learning models for COVID-19 detection using CT imaging," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 16301–16314, 2021.
- [34] S. Wang, J. Chen, and Y. Lu, "New coronary pneumonia chest CT image segmentation based on federated learning and blockchain," *Journal of Jilin University (Engineering and Technology Edition)*, vol. 51, no. 6, pp. 2164–2173, 2021.
- [35] Y. Wen and M. Chen, "Medical data sharing based on reputation screening federated learning data sources," *Computer Engineering*, vol. 61, pp. 1–15, 2022.
- [36] Q. Yang, "AI and data privacy protection: the solution to federated learning," *Journal of Information Security Research*, vol. 5, no. 11, pp. 961–965, 2019.
- [37] L. Qiao, S. Dang, B. Shihada, M.-S. Alouini, R. Nowak, and Z. Lv, "Can blockchain link the future?," *Digital Communications and Networks*, vol. 7, 2021.
- [38] S. Nakamoto, "Bitcoin: a peer-to-peer electronic cash system," *Decentralized Business Review*, p. 21260, 2008.
- [39] M. Chowdhury, T. Rahman, A. Khandakar et al., "Can AI help in screening viral and COVID-19 pneumonia?," *IEEE Access*, vol. 8, pp. 132665–132676, 2020.

Research Article

MiR-96-5p Facilitates Lung Adenocarcinoma Cell Phenotypes by Inhibiting FHL1

Feng Zhou ¹, Chaojie Qian ¹, Tingting Chen ¹ and Xiaoliang Zang ²

¹Department of Respiratory Medicine, Changxing People's Hospital, Huzhou, Zhejiang Province, China 313100

²Department of Respiratory Medicine, The First People's Hospital of Huzhou, Huzhou, Zhejiang Province, China 313000.

Correspondence should be addressed to Xiaoliang Zang; zangxxxliang@163.com

Received 16 May 2022; Revised 22 July 2022; Accepted 27 July 2022; Published 16 August 2022

Academic Editor: Tao Huang

Copyright © 2022 Feng Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. FHL1 is understood as a tumor repressor gene in various cancers and a possible target for cancer treatment. We investigated the influences of FHL1 on cell functions as well as its molecular mechanisms in lung adenocarcinoma (LUAD) cells. **Methods.** The miRNA-mRNA modulatory axis was predicted by bioinformatics. The expression levels of FHL1 mRNA and protein in LUAD cells were, respectively, analyzed by qRT-PCR and western blot. Dual luciferase analysis was introduced to verify the interaction between miR-96-5p and FHL1. CCK-8, cell colony formation, and Transwell assays were utilized to analyze proliferation, colony formation, migration, and invasion of A549 cells. **Results.** Expression of FHL1 mRNA and protein in LUAD tissue and cells was downregulated, which was linked with poor prognoses of patients. In addition, FHL1 overexpression could hamper colony formation, proliferation, invasion, and migration of LUAD cells. In addition, dual-luciferase analysis verified miR-96-5p as an upstream regulator of FHL1. Overexpression of miR-96-5p suppressed FHL1 expression in LUAD cells and promoted proliferation, invasion, and migration of LUAD cells, while overexpression of FHL1 could simultaneously restore the above-mentioned promoting effect. **Conclusion.** MiR-96-5p fostered cell malignant behaviors by targeting FHL1. This research uncovered the regulatory mechanism of FHL1 in LUAD and offered optional therapeutic targets for LUAD patients.

1. Introduction

Lung cancer is a type of prevalent malignancy derived from tunica bronchial mucosa or glands. Globally, lung cancer is dominant in cancer morbidity and mortality with its new cases and deaths ranking first among all cancers [1]. The incidence and mortality of lung cancer, one major cause of cancer-related deaths in China, have increased significantly in recent years, with about 7.87 million cases and 631,000 deaths in 2015 [2]. As the predominant type of lung cancer, non-small-cell lung cancer (NSCLC) patients involve 60% of lung adenocarcinoma (LUAD) [3]. LUAD patients in advanced stage are always accompanied with disease metastasis to nearby and distant organs, suggesting their poor prognosis [4]. Hence, an intensive study of LUAD pathogenesis becomes a central issue for seeking out a new target in molecular therapy.

FHL1 is mainly expressed in skeleton and myocardium. FHL1 plays a pivotal role in sarcomere assembly as a scaffold [5], as well as cell cycle and metastasis [6], whose dysfunction is relevant to muscular disorders. FHL1 is considered as a suppressor gene in several tumors. For instance, FHL1 arrests G1/S cell cycle to hamper tongue squamous cell carcinoma cell growth [7]. In glioma, decreased FHL1 protein represses tumor growth via PI3K/AKT signaling transduction [8]. Nevertheless, few investigations have probed into the function of FHL1 in LUAD, which is valuable and ought to be invoked.

MicroRNAs (miRNAs) are capable of affecting tumor formation, cell proliferation, cell invasion, and metastasis via targeting gene expression [9]. Recent investigations have gone deep into aberrantly expressed miRNAs in LUAD. For example, miR-210 targets LOXL4 to facilitate phenotype progression of LUAD [10]. MiR-363-3p constrains LUAD growth via targeting PCNA [11]. MiR-186-5p fosters cell

TABLE 1: Primer sequences used in qRT-PCR.

Gene	Primer sequences (5' → 3')
miR-96-5p	F: 5'-TCAACTGGTGTCTCGTGGAGTCGGCAATTCAGTTGAGAGCAAAAA-3' R: 5'-ACACTCCAGCTGGGTTTGGCACTAGCACATT-3'
U6	F: 5'-CTCGCTTCGGCAGCACATA-3' R: 5'-CGAATTTGCGTGTCTATCCT-3'
FHL1	F: 5'-GAAGTGTGCTGGATGCAAGA-3' R: 5'-GGGGGCTTCCTAGCTTTAGA-3'
β -Actin	F: 5'-ATCACCATTGGCAATGAGCG-3' R: 5'-TTGAAGGTAGTTTCGTGGAT-3'

malignant phenotypes of LUAD via PTEN [12]. MiR-96-5p has become a research hotspot in recent years, and it was found to play an essential role in hepatocellular carcinoma [13], gastric adenocarcinoma [14], and head and neck squamous cell carcinoma [15]. Nonetheless, its regulatory role in LUAD has been less well defined.

Herein, we discovered that FHL1 may exert an imperative function in LUAD progression and was a possible target of miR-96-5p. The expression of miR-96-5p was measured by quantitative real-time polymerase chain reaction (qRT-PCR). The mRNA and protein levels of FHL1 in LUAD cell lines were, respectively, analyzed by qRT-PCR and western blot. Through a trail of biological experiments, the effect and modulatory role of miR-96-5p/FHL1 axis in LUAD were authenticated. Together our findings shed novel insights into LUAD development and potential therapeutic approaches.

2. Materials and Methods

2.1. Bioinformatics Analysis. Mature miRNAs (normal sample: $n = 46$, cancer sample: $n = 521$) and mRNAs (normal sample: $n = 59$, cancer sample: $n = 535$) expression profiles were accessed from The Cancer Genome Atlas (TCGA)-LUAD dataset. Comparison of FHL1 expression between normal and cancer tissue was analyzed by t -test. With the medium FHL1 level as threshold, 535 cancer tissue samples were sorted into high- ($n = 267$) and low-expression ($n = 268$) groups. Survival analysis was conducted between the two groups using “Survival” package.

Differential expression analysis was conducted based on miRNAs expression data by “EdgeR” package with $|\log_{2}FC| > 2.0$ and $\text{padj} < 0.01$ as thresholds. miRDB, mirDIP, TargetScan, miRTarBase, and starBase databases were utilized to screen the upstream regulatory miRNAs of FHL1. The overlap miRNAs between the predicted miRNAs and the differentially upregulated miRNAs were considered as the candidates. Pearson correlation analysis was applied to select the most optimal candidate. t -test was used to unravel miRNA expression.

2.2. Cell Cultivation and Transfection. LUAD cell lines A549 (BNCC100441), H1650 (BNCC232963), H441 (BNCC292357), and H1299 (BNCC334400) and human

bronchial epithelial cell line BEAS-2B (BNCC254518) were all procured from BeNa Culture Collection (Beijing, China). Cells were prepared in DMEM or RPMI-1640 medium (Gibco, USA) containing 10% FBS and 100 mg/mL streptomycin/penicillin under routine conditions. The culture conditions were 5% CO₂ and 37°C. Synthesized miR-96-5p-mimic (miR-mimic), miR-NC, pcDNA3-FHL1 (oe-FHL1), and empty vector pcDNA3 (oe-NC) were all accessed from GenePharma (Shanghai, China). Lipofectamine 2000 Reagent (Invitrogen, USA) was applied for cell transfection. The transfected cells were maintained for 2 d with 5% CO₂ at 37°C for subsequent experiments.

2.3. qRT-PCR. TRIZOL reagent (Thermo Fisher Science, USA) was recommended for total RNA extraction and NanoDrop 2000 spectrophotometer (Thermo Fisher Science, USA) for RNA evaluation and quantification. miScript II RT kit (Qiagen, Germany) as well as PrimeScript RT Master Mix (Takara, Japan) was applied to synthesize complementary DNA (cDNA). Then, miScript SYBR Green PCR Kit (Qiagen, Germany) with SYBR® Premix Ex Taq™ II (Takara, Japan) was applied to evaluate miRNA and mRNA expression. qRT-PCR was done on Applied Biosystems® 7500 Real-Time PCR Systems (Thermo Fisher Scientific, USA) to assess miR-96-5p and FHL1 mRNA levels, respectively. U6 and β -actin were taken as endogenous references for miR-96-5p and FHL1, respectively. The $2^{-\Delta\Delta C_t}$ value was utilized to compare relative expression differences. The primer sequences are shown in Table 1.

2.4. Western Blot Assay. Following the instructions, protein extraction kit (Beyotime, China) was employed to isolate total proteins. The proteins underwent 10% SDS-PAGE, followed by being transferred onto a PVDF membrane. Then the membrane was blocked with 5% BSA for 1 h and maintained at 4°C with primary antibodies rabbit anti-FHL1 (1:1000, ab255828, Abcam, UK) and mouse anti- β -actin (1 μ g/mL, ab8226, Abcam, UK) overnight. Afterwards, the membrane was incubated for 2 h at room temperature with secondary antibody goat antirabbit IgG H&L (HRP) (1:2000, ab6721, Abcam, UK). Thereafter, the membrane was rinsed with PBST three times, and protein level was assayed with ECL kit (GE Healthcare, USA).

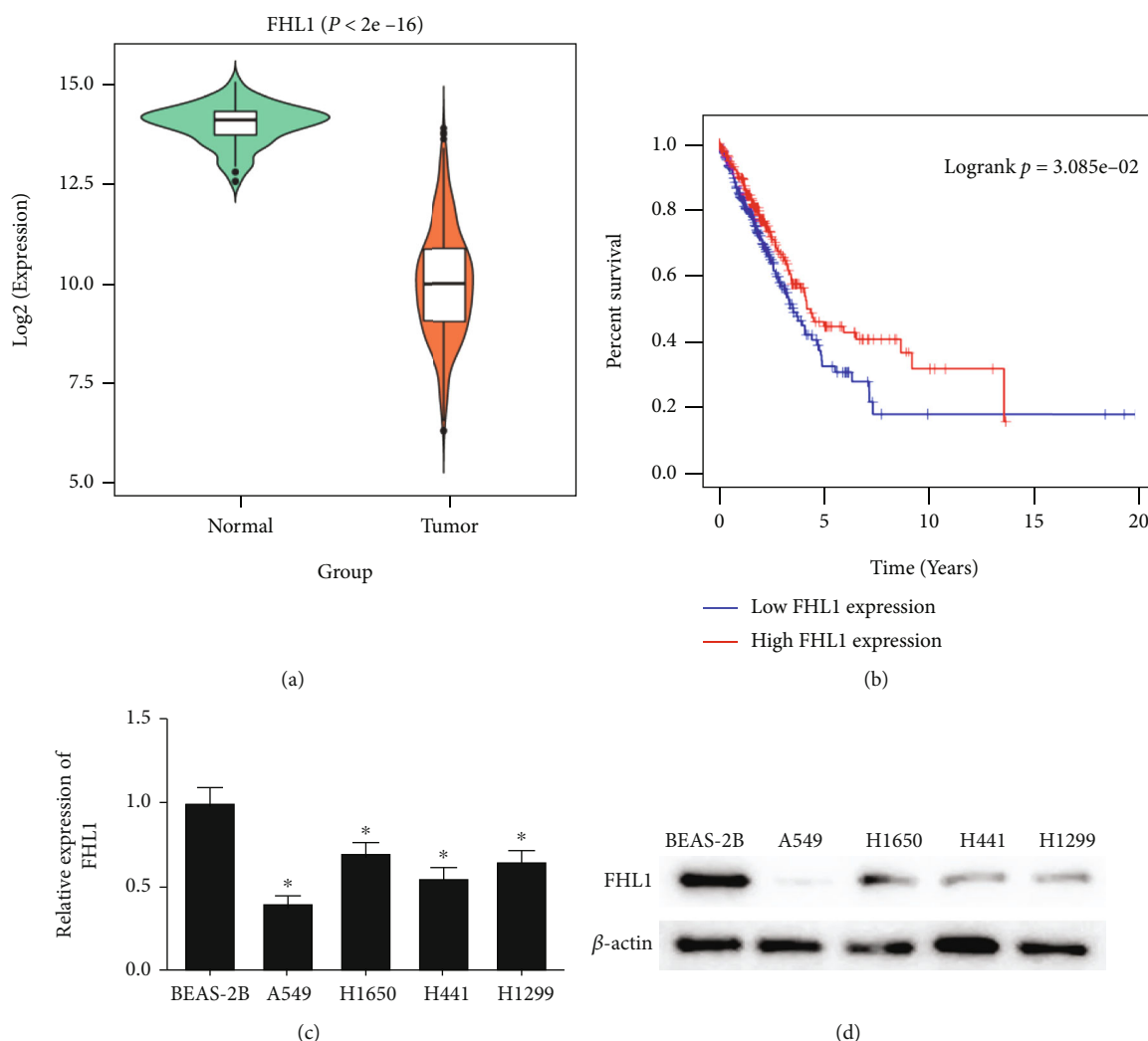


FIGURE 1: FHL1 expression is notably decreased in LUAD cells. (a) Violin plot of FHL1 level in normal group (green; $n = 59$) and tumor group (orange; $n = 535$); (b) survival curves of high-FHL1 (red) and low-FHL1 (blue) groups. Abscissa represents time (year), and ordinate represents survival rate; (c) FHL1 mRNA expression in A549, H1650, H441, H1299, and BEAS-2B cells; (d) FHL1 protein level in cells; * $p < 0.05$.

2.5. Cell Counting Kit-8 (CCK-8) Assay. The transfected cells were placed into 96-well plates (3×10^4 cells/well) for culture. After $10 \mu\text{L}$ of CCK-8 solution was supplemented at 1, 2, 3, and 4 days of incubation, cells were maintained for 2 h under routine conditions. A microplate reader was employed to assess optical density (OD) value at 450 nm. Based on the OD value, relative growth rate of cancer cells was acquired.

2.6. Cell Colony Formation Assay. 2×10^2 cells/well were planted on 6-well plates at 37°C for 2 weeks. Following being rinsed with PBS twice, cells were fixed for 5 min with methanol. Then 3 min of cell staining at 37°C was completed with crystal violet. To remove the residual crystal violet, each well was cleaned with sterile water. A microscope was introduced for counting number of colonies (≥ 50 cells).

2.7. Transwell Invasion and Migration Assays. Matrigel (BD Biosciences, USA) was supplemented to upper chamber of

Transwell and placed in a humidified incubator for 60 min of gelation at 37°C . LUAD A549 cells were dispersed with trypsin/EDTA (0.25%/0.04%), centrifuged, and washed twice with serum-free culture medium (580 g, 3 min). Then, 2×10^5 transfected cells that suspended in serum-free culture medium were placed on the upper chamber. RPMI-1640 plus 10% FBS was placed below the cell membrane to facilitate cell invasion. 48 h later, cells on the upper layer of the insert with membrane were swabbed with a cotton bud, and cells that invaded through membrane were fixed with 4% paraformaldehyde for 15 min. Next, 25 min of staining at 37°C was done with 0.1% crystal violet. A microscope was used to observe stained cells. The number of stained cells in 5 random fields was counted ($100\times$) for statistical analysis. The experiment was performed in triplicate.

Steps of cell migration assay were the same as cell invasion assay except for addition of Matrigel in invasion assay. The experiment was performed 3 times.

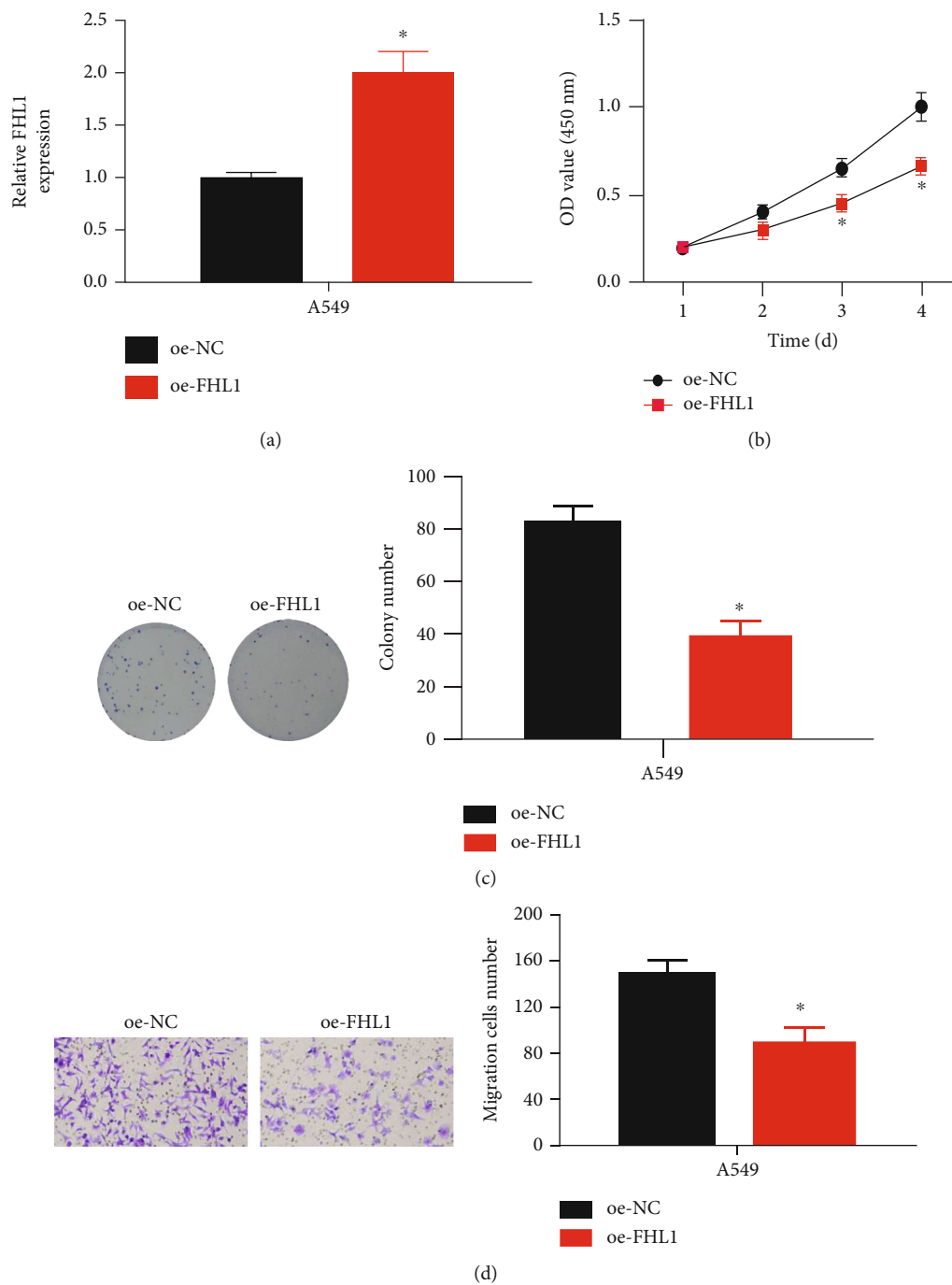


FIGURE 2: Continued.

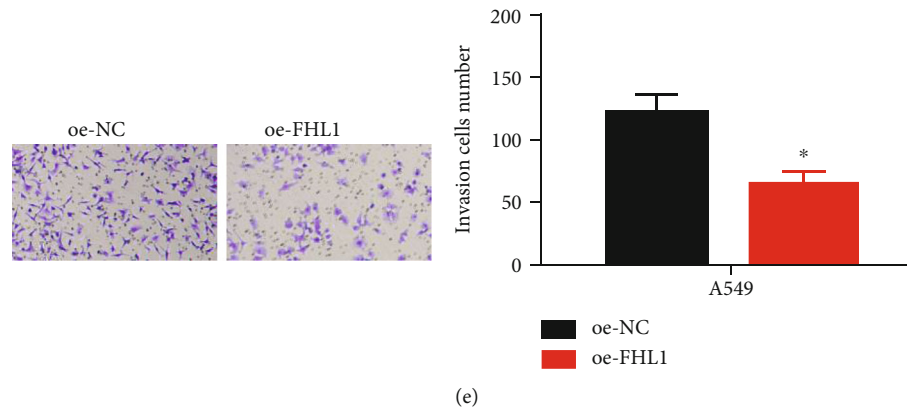


FIGURE 2: Overexpression of FHL1 hinders cell phenotype progression of LUAD. (a) FHL1 level in LUAD A549 cells in oe-NC and oe-FHL1; (b) proliferative efficiency of A549 cells in various transfection groups; (c) colony formative capability of A549 cells in various transfection groups; (d and e) migratory and invasive properties of two groups of A549 cells (100 \times); * $p < 0.05$.

2.8. Dual-Luciferase Reporter Gene Assay. Putative targeted relationship between FHL1 and miR-96-5p was predicted through bioinformatics databases. Wild-type FHL1-3'-untranslated region (UTR) sequences (FHL1-WT) and mutant sequences (FHL1-MUT) corresponding to the predicted miR-96-5p binding sites were synthesized and then cloned into dual-luciferase reporter gene vector psiCHECK, respectively. Cells were plated to 24-well plates and prepared in complete culture medium for over 24 h before transfection. PBS (pH 7.4) was employed to rinse cells for transient transfection. The psiCHECK-FHL1 WT/MUT and miR-96-5p mimic/mimic NC were cotransfected with A549 cells with Lipofectamine 2000 (Invitrogen, USA) reagent. Forty-eight h later, cells were harvested, and luciferase activity was assessed and quantified with dual-luciferase reporter assay kit (Promega, USA).

2.9. Data Analysis. All data analyses were carried out on GraphPad Prism 6.0 (La Jolla, CA). Each experiment was repeated three times. Data were showed as mean \pm SD. t -test was utilized for comparison between two groups. $p < 0.05$ suggested statistically significant difference.

3. Results

3.1. FHL1 Level Is Notably Decreased in LUAD Cells. TCGA-LUAD dataset exhibited that mRNA expression of FHL1 was evidently downregulated in LUAD (Figure 1(a)). Kaplan-Meier analysis revealed that LUAD patients with low FHL1 expression had reduced overall survival (OS) (Figure 1(b)). FHL1 mRNA and protein levels in LUAD cell lines were measured via qRT-PCR and western blot, respectively. As illustrated in Figures 1(c) and 1(d), FHL1 was decreased in A549, H1650, H441, and H1299 cell lines compared with BEAS-2B cell line. Based on previous investigations and experiments, it was authenticated that FHL1 was decreased in LUAD cells. To make a thorough inquiry of the impact of FHL1 on LUAD cells, A549 cell line with the lowest FHL1 expression among all LUAD cell lines was chosen for the subsequent cellular functional experiments.

3.2. Overexpression of FHL1 Hinders Cell Malignant Phenotypes of LUAD. To investigate the biological functions of FHL1, we established FHL1 overexpression cell line. Transfection efficiency of FHL1 was assayed through qRT-PCR. FHL1 was significantly overexpressed in oe-FHL1 group (Figure 2(a)), thereby being well prepared for following experiments. CCK-8 and cell colony formation experiments exhibited that FHL1 overexpression restrained proliferation and colony formation of A549 cells (Figures 2(b) and 2(c)). Transwell assays showed that FHL1 overexpression in A549 cells repressed the migratory and invasive abilities (Figures 2(d)–2(e)). Taken together, these findings demonstrated that enforced FHL1 level could repress malignant phenotype of LUAD cells.

3.3. MiR-96-5p Downregulates FHL1 Level in LUAD Cells. A differential expression analysis was carried out by EdgeR. A total of 112 differential miRNAs were obtained: 92 differentially upregulated miRNAs and 20 differentially downregulated miRNAs (Figure 3(a)). Then, differentially upregulated miRNAs were intersected with the predicted target miRNAs in the database, by which two targets were obtained: miR-105-5p and miR-96-5p (Figure 3(b)). Pearson correlation analysis was performed subsequently to select the miRNA with high negative correlation with FHL1 at expression manner, where miR-96-5p show the highest negative correlation with FHL1 (Figure 3(c)). In addition, bioinformatics analysis showed significantly highly expressed miR-96-5p in tumor tissues (Figure 3(d)). Later, miR-96-5p expression in normal cells and LUAD cells was assayed through qRT-PCR. Result displayed that miR-96-5p was prominently upregulated in LUAD cells (Figure 3(e)), which was in accordance with previous investigations. Thus, miR-96-5p was dramatically inversely associated with FHL1 and may be an upstream regulatory miRNA of FHL1.

Also, starBase database was employed to identify targeted relationship of miR-96-5p and FHL1 (Figure 3(f)). The binding of FHL1 3'-UTR and miR-96-5p was observed. Dual-luciferase analysis indicated that forced expression of miR-96-5p weakened luciferase activity of FHL1-WT, but did not affect that of FHL1-MUT. Hence, miR-96-5p could

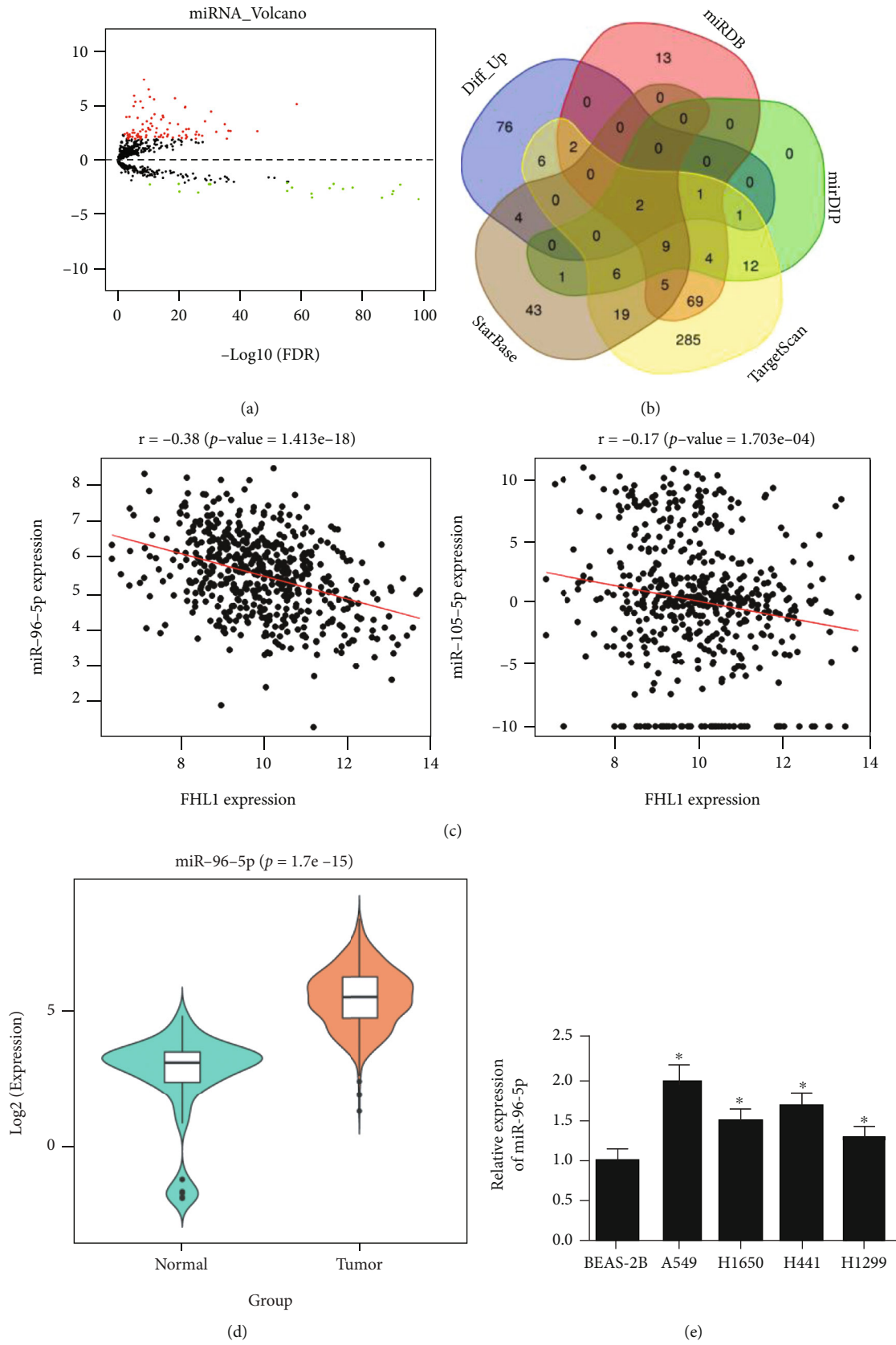


FIGURE 3: Continued.

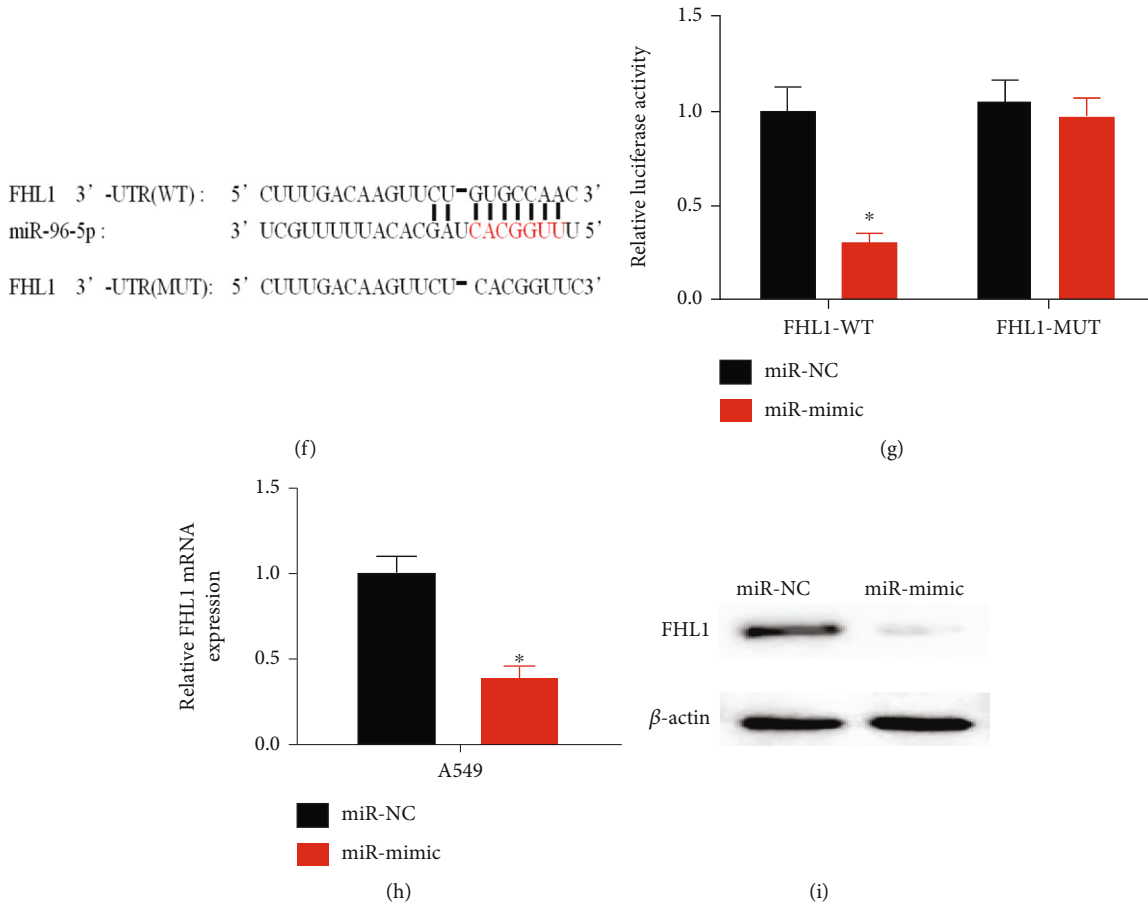


FIGURE 3: MiR-96-5p downregulates FHL1 expression in LUAD cells. (a) Volcano map of differential miRNAs in normal and tumor groups in TCGA database. Red indicates differentially upregulated miRNAs, and green indicates differentially downregulated miRNAs; (b) Venn diagram of predicted upstream miRNAs of FHL1 and differential miRNAs; (c) Pearson correlation analysis of FHL1 and its predicted upstream miRNAs; (d) violin plot of miR-96-5p expression in normal (green; $n = 46$) and tumor (orange; $n = 521$) tissue; (e) MiR-96-5p level in BEAS-2B and A549, H1650, H441, and H1299 cell lines; (f) diagram of binding of miR-96-5p and FHL1-WT and FHL1-MUT sequences; (g) luciferase activity of A549 cells in treatment groups (miR-NC and miR-mimic); (h) FHL1 mRNA level in A549 cells; (i) FHL1 protein expression in A549 cells; $*p < 0.05$.

target FHL1 (Figure 3(g)). Next, qRT-PCR and western blot were performed to assay mRNA and protein levels of FHL1 in transfected cells (Figures 3(h) and 3(i)), respectively. Thus, enforced miR-96-5p level constrained FHL1 level. The findings proved that miR-96-5p could downregulate FHL1 level in LUAD.

3.4. MiR-96-5p Facilitates LUAD Cell Phenotype Progression via Suppressing FHL1. To further examine miR-96-5p/FHL1 regulatory axis in LUAD cells, rescue experiments were designed for verification. MiR-NC+oe-NC, miR-mimic+oe-NC, and miR-mimic+oe-FHL1 groups were established to investigate whether overexpression of FHL1 can rescue impact of miR-96-5p overexpression. qRT-PCR and western blot outcomes displayed that compared with miR-NC-oe-NC group, the levels of FHL1 mRNA and protein in miR-mimic+oe-NC group were prominently decreased. Additionally, the levels of FHL1 mRNA and protein in miR-mimic+oe-FHL1 group were significantly higher than those in miR-mimic+oe-NC group (Figure 4(a)). CCK-8 and cell colony formation assays illustrated that enforced expression

of miR-96-5p hastened LUAD cell proliferation. Transwell assays exhibited that transfection of overexpression of miR-96-5p increased migratory and invasive capabilities of LUAD cells. Overexpression of FHL1 and miR-96-5p simultaneously weakened facilitating impact of miR-96-5p on cell phenotype progression of LUAD (Figures 4(b)–4(e)). Thus, miR-96-5p facilitated LUAD cell malignant phenotypes via suppressing FHL1.

4. Discussion

Accumulating evidence existed that the expression of various mRNAs is decreased in LUAD, such as PRDM16 [16], MFAP4 [17], and VIPR1 [18]. FHL1 as a tumor repressor gene is authenticated to be poorly expressed in varying cancers, containing liver cancer [19], oral cancer [20], head and neck squamous cell carcinoma [21], and breast cancer [22]. We selected FHL1 as the research object to unravel its specific expression. We revealed that FHL1 expression was decreased in LUAD tissue and cells. Besides, decreased FHL1 was associated with dismal outcomes of patients.

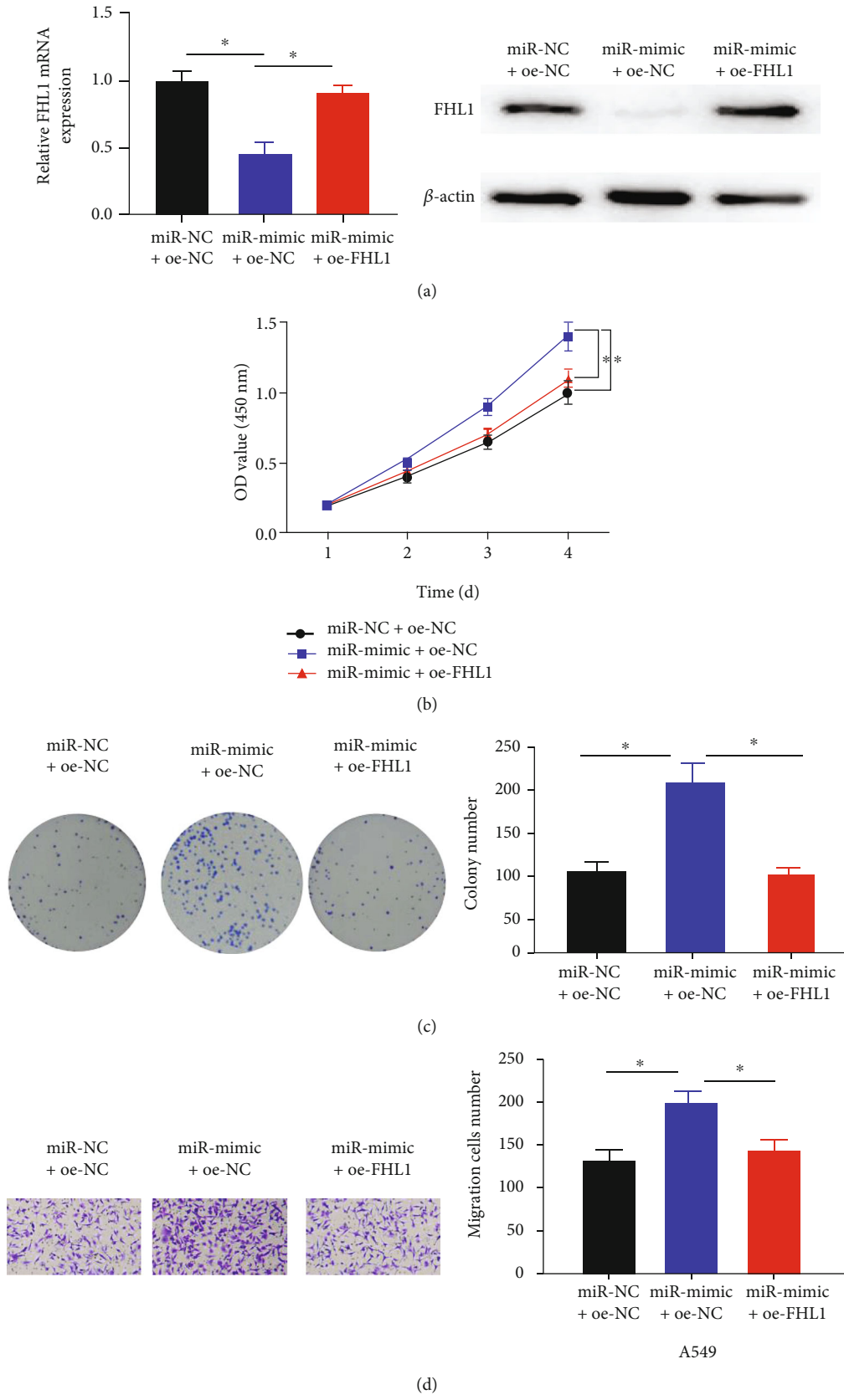
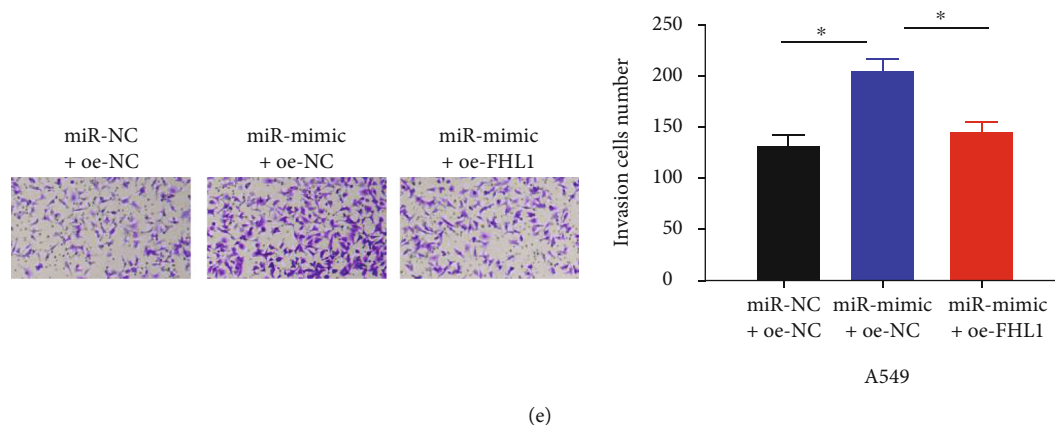


FIGURE 4: Continued.



(e)

FIGURE 4: MiR-96-5p facilitates cell malignant behaviors via suppressing FHL1. (a) FHL1 mRNA and protein levels in LUAD A549 cells in varying groups; (b) A549 cell proliferative properties in varying groups; (c) cell colony formation of A549 cells in varying groups; (d and e) migratory and invasive abilities of A549 cells in varying groups (100 \times); * $p < 0.05$.

Studies existed that silencing FHL1 noticeably fosters the growth of cervical cancer cells [23], and decreased FHL1 also hastens the growth of breast cancer cells [22], indicating that FHL1 exerted an inhibitory effect in several cancers. To validate the impact of FHL1 level on the progression of LUAD cells, FHL1 expression level was upregulated in LUAD cells. A trail of biological experiments displayed that overexpressed FHL1 repressed cell malignant phenotypes of LUAD. Niu et al. [24] found that overexpression of FHL1 restrains LUAD A549 cell growth, which is congruous with our experimental results. The above findings authenticated that FHL1 serves as a tumor repressor in LUAD.

FHL1 has upstream regulatory genes in multiple cancers. MiR-103 targets FHL1 and inversely modulates its expression in NSCLC [25]. In colorectal and liver tumors, miR-410 targets FHL1, and silencing miR-410 increases FHL1 expression [26]. Dual-luciferase analysis validated targeted relationship of FHL1 3'-UTR and miR-96-5p. Existing investigations unraveled that miR-96-5p serves as an oncogene in varying cancer cells. MiR-96-5p targets CAV1 to restrain AKT phosphorylation and its down-stream Cyclin D1 and P70 protein levels, thereby facilitating cancer cell proliferation and migration [27]. Zhao *et al.* [28] unearthed that miR-96-5p enhances cell proliferative ability while hampers cell apoptosis of LUAD via targeting CYLD. As such, overexpression of miR-96-5p exerts a similar promotive impact on papillary thyroid carcinoma cells [29]. This work proved that miR-96-5p was markedly upregulated in LUAD, and upregulation of miR-96-5p resulted in increased cell proliferation, migration, and invasion. Rescue experiments elucidated that enforced expression of FHL1 could reverse promotion impact of miR-96-5p overexpression on LUAD malignant progression. Together the above findings enunciated that miR-96-5p was increased in LUAD cells and decreased FHL1 expression, which also exacerbated LUAD malignant progression.

Overall, this study verified the decreased expression of FHL1 and increased expression of miR-96-5p in LUAD. FHL1 was the target of miR-96-5p. Overexpression of

FHL1 restrained proliferation, invasion, and migration of LUAD cells. This study first proposed that miR-96-5p facilitated malignant phenotypes of LUAD cells by targeting FHL1. Besides, the experimental data suggested that FHL1 may be a potential biomarker for drug development. Nevertheless, the interaction between the expression of these two genes in LUAD tissue has not been proved. In the future, we will focus on this direction.

Data Availability

The data used to support the findings of this study are included within the article. The data and materials in the current study are available from the corresponding author on reasonable request.

Consent

All authors consent to submit the manuscript for publication.

Conflicts of Interest

No potential conflicts of interest in the article.

Authors' Contributions

All authors contributed to data analysis, drafting, and revising the article; gave final approval of the version to be published; and agreed to be accountable for all aspects of the work.

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

- [2] PA for Medical, "Clinical practice guideline for stage IV primary lung cancer in China (2021 version)," *Zhonghua zhong liu za zhi [Chinese journal of oncology]*, vol. 43, no. 39-59, 2021.
- [3] J. Wang, K. Zou, X. Feng et al., "Downregulation of NMI promotes tumor growth and predicts poor prognosis in human lung adenocarcinomas," *Molecular Cancer*, vol. 16, no. 1, p. 158, 2017.
- [4] E. J. Greenawalt, M. D. Edmonds, N. Jain, C. M. Adams, R. Mitra, and C. M. Eischen, "Targeting of SGK1 by miR-576-3p inhibits lung adenocarcinoma migration and invasion," *Molecular Cancer Research*, vol. 17, no. 1, pp. 289–298, 2019.
- [5] S. Han, C. Cui, H. He et al., "FHL1 regulates myoblast differentiation and autophagy through its interaction with LC3," *Journal of Cellular Physiology*, vol. 235, no. 5, pp. 4667–4678, 2020.
- [6] J. M. Matthews, K. Lester, S. Joseph, and D. J. Curtis, "LIM-domain-only proteins in cancer," *Nature Reviews. Cancer*, vol. 13, no. 2, pp. 111–122, 2013.
- [7] W. Ren, P. Lian, L. Cheng et al., "FHL1 inhibits the growth of tongue squamous cell carcinoma cells via G1/S cell cycle arrest," *Molecular Medicine Reports*, vol. 12, no. 3, pp. 3958–3964, 2015.
- [8] S. Z. Li, Y.-Y. Hu, J.-L. Zhao et al., "Downregulation of FHL1 protein in glioma inhibits tumor growth through PI3K/AKT signaling," *Oncology Letters*, vol. 19, pp. 3781–3788, 2020.
- [9] Y. Peng and C. M. Croce, "The role of MicroRNAs in human cancer," *Signal Transduction and Targeted Therapy*, vol. 1, no. 1, p. 15004, 2016.
- [10] S. Xie, G. Liu, J. Huang, H. B. Hu, and W. Jiang, "miR-210 promotes lung adenocarcinoma proliferation, migration, and invasion by targeting lysyl oxidase-like 4," *Journal of Cellular Physiology*, vol. 234, no. 8, pp. 14050–14057, 2019.
- [11] Y. Wang, T. Chen, H. Huang et al., "miR-363-3p inhibits tumor growth by targeting PCNA in lung adenocarcinoma," *Oncotarget*, vol. 8, no. 12, pp. 20133–20144, 2017.
- [12] H. Feng, Z. Zhang, X. Qing, S. W. French, and D. Liu, "miR-186-5p promotes cell growth, migration and invasion of lung adenocarcinoma by targeting PTEN," *Experimental and Molecular Pathology*, vol. 108, pp. 105–113, 2019.
- [13] N. Iwai, K. Yasui, A. Tomie et al., "Oncogenic miR-96-5p inhibits apoptosis by targeting the caspase-9 gene in hepatocellular carcinoma," *International Journal of Oncology*, vol. 53, no. 1, pp. 237–245, 2018.
- [14] H. Y. Zhou, C. Q. Wu, and E. X. Bi, "MiR-96-5p inhibition induces cell apoptosis in gastric adenocarcinoma," *World Journal of Gastroenterology*, vol. 25, no. 47, pp. 6823–6834, 2019.
- [15] M. Vahabi, C. Pulito, A. Sacconi et al., "miR-96-5p targets PTEN expression affecting radio-chemosensitivity of HNSCC cells," *Journal of Experimental & Clinical Cancer Research*, vol. 38, no. 1, p. 141, 2019.
- [16] L. R. Fei, W. J. Huang, Y. Wang et al., "PRDM16 functions as a suppressor of lung adenocarcinoma metastasis," *Journal of Experimental & Clinical Cancer Research*, vol. 38, no. 1, p. 35, 2019.
- [17] Y. Y. Feng, C. H. Liu, Y. Xue, Y. Y. Chen, Y. L. Wang, and X. Z. Wu, "MicroRNA-147b promotes lung adenocarcinoma cell aggressiveness through negatively regulating microfibril-associated glycoprotein 4 (MFAP4) and affects prognosis of lung adenocarcinoma patients," *Gene*, vol. 730, article 144316, 2020.
- [18] L. Zhao, Z. Yu, and B. Zhao, "Mechanism of VIPR1 gene regulating human lung adenocarcinoma H1299 cells," *Medical Oncology*, vol. 36, no. 11, p. 91, 2019.
- [19] J. Wang, F. Huang, J. Huang, J. Kong, S. Liu, and J. Jin, "Epigenetic analysis of FHL1 tumor suppressor gene in human liver cancer," *Oncology Letters*, vol. 14, no. 5, pp. 6109–6116, 2017.
- [20] C. M. Pereira, A. C. de Carvalho, F. R. da Silva et al., "In vitro and in silico validation of CA3 and FHL1 downregulation in oral cancer," *BMC Cancer*, vol. 18, no. 1, p. 193, 2018.
- [21] W. Cao, J. Liu, R. Xia et al., "X-linked FHL1 as a novel therapeutic target for head and neck squamous cell carcinoma," *Oncotarget*, vol. 7, no. 12, pp. 14537–14550, 2016.
- [22] L. Ding, C. Niu, Y. Zheng et al., "FHL1 interacts with oestrogen receptors and regulates breast cancer cell growth," *Journal of Cellular and Molecular Medicine*, vol. 15, no. 1, pp. 72–85, 2011.
- [23] Z. Fan, X. Xu, X. Zhang et al., "FHL1 knockdown mediated by lentiviral shRNA promotes the growth of HeLa and HepG2 cells," *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi*, vol. 31, no. 7, pp. 879–883, 2015.
- [24] C. Niu, C. Liang, J. Guo et al., "Downregulation and growth inhibitory role of FHL1 in lung cancer," *International Journal of Cancer*, vol. 130, no. 11, pp. 2549–2556, 2012.
- [25] Z. Wang, J. Zhang, B. Yang et al., "Long intergenic noncoding RNA 00261 acts as a tumor suppressor in non-small cell lung cancer via regulating miR-105/FHL1 Axis," *Journal of Cancer*, vol. 10, no. 25, pp. 6414–6421, 2019.
- [26] Y. Wang, J. Fu, M. Jiang et al., "MiR-410 is overexpressed in liver and colorectal tumors and enhances tumor cell growth by silencing FHL1 via a direct/indirect mechanism," *PLoS One*, vol. 9, no. 10, article e108708, 2014.
- [27] B. Liu, J. Zhang, and D. Yang, "miR-96-5p promotes the proliferation and migration of ovarian cancer cells by suppressing Caveolae1," *Journal of Ovarian Research*, vol. 12, no. 1, p. 57, 2019.
- [28] M. Zhao, X. F. Xin, J. Y. Zhang, W. Dai, T. F. Lv, and Y. Song, "LncRNA GMDS-AS1 inhibits lung adenocarcinoma development by regulating miR-96-5p/CYLD signaling," *Cancer Medicine*, vol. 9, no. 3, pp. 1196–1208, 2020.
- [29] Z. M. Liu, Z. Y. Wu, W. H. Li, L. Q. Wang, J. N. Wan, and Y. Zhong, "MiR-96-5p promotes the proliferation, invasion and metastasis of papillary thyroid carcinoma through down-regulating CCDC67," *European Review for Medical and Pharmacological Sciences*, vol. 23, no. 8, pp. 3421–3430, 2019.

Research Article

Effectiveness and Safety of High-Power Radiofrequency Ablation Guided by Ablation Index for the Treatment of Atrial Fibrillation

Xuefeng Zhu ¹, Chunxiao Wang ¹, Hongxia Chu ¹, Wenjing Li ², Huihui Zhou ³,
Lin Zhong ¹ and Jianping Li ¹

¹Department of Cardiology, The Affiliated Yantai Yuhuangding Hospital of Qingdao University, China

²Department of Ultrasound Medicine, The Affiliated Yantai Yuhuangding Hospital of Qingdao University, China

³Department of Pathology, The Affiliated Yantai Yuhuangding Hospital of Qingdao University, China

Correspondence should be addressed to Jianping Li; lijianpingtyhd@163.com

Received 15 April 2022; Revised 9 June 2022; Accepted 21 July 2022; Published 12 August 2022

Academic Editor: Tao Huang

Copyright © 2022 Xuefeng Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. To investigate the efficacy and safety of ablation index- (AI-) guided high-power radiofrequency ablation in the treatment of atrial fibrillation (AF). **Methods.** Outcomes of radiofrequency (RF) applications were compared in a swine ventricular endocardial model ($n = 10$ each for 50 W, 40 W, and 30 W; AI = 500). And a total of 100 consecutive patients with paroxysmal AF undergoing pulmonary vein isolation (PVI) were included. The patients were divided into two groups ($n = 50$ for each) as follows: control group, treated with conventional power (30 W) ablation mode; and study group, treated with high power (40 W) radiofrequency ablation mode. All groups were treated with the same AI value guided the ablation (target AI = 400/500 on posterior/anterior wall, respectively). Acute pulmonary vein (PV) reconnection was assessed post adenosine administration 20 minutes after ablation. Subsequently, pathological observation of porcine heart lesions and necrotic tissue was performed. Additionally, statistical analyses were carried out on patients' baseline clinical characteristics, surgical data, and total RF energy. **Results.** In swine ventricular endocardial RF applications, compared with 40 W and 30 W, the use of 50 W was associated with shallower tissue lesion depth ($p < 0.001$) and greater lesion maximum diameter ($p < 0.001$). Compared with 40 W and 30 W, tissue necrosis caused by 50 W was the deepest and largest ($p < 0.001$). In pulmonary vein isolation (PVI), there was no significant difference in baseline data between the study group and control group ($p > 0.05$). In patients with paroxysmal atrial fibrillation, the procedure time in the high-power group was significantly shortened ($p < 0.001$). The ablation time was significantly shorter ($p < 0.001$). Compared with control group, RF energy per point and acute pulmonary vein (PV) reconnection were lower ($p < 0.001$), and first-pass PVI was higher ($p < 0.01$) in study group. There were no significant differences in complications and sinus rhythm maintenance at 12 months between the two groups ($p > 0.05$). **Conclusions.** Compared with conventional (30 W) PVI, AI-guided high-power (40 W) was safe and associated with shorter procedure time and reduced acute PV reconnection.

1. Introduction

Radiofrequency (RF) ablation is widely used in the treatment of patients with atrial fibrillation (AF) via transmural, continuous, and lasting lesion formation with pulmonary vein isolation (PVI). Despite optimization of RF current transfer, RF energy transfer duration, catheter contact forces (CF), and catheter stability [1], pulmonary vein (PV) reconnection occurs acutely and at 3 months after PVI at a frequency of 22% and 15%, respectively [2, 3], mainly secondary to

reversible damage, partial thickness, and/or incomplete ablation [4–6]. In recent years, the ablation index (AI) has been widely applied to treat drug-resistant paroxysmal AF [7]. AI is known as a novel quantitative ablation lesion marker that combines contact force (CF), time, and power in a weighted formula [8, 9]. In recent studies, patients undergoing AI-guided ablation of AF have good clinical outcomes, and the procedure time and ablation time are relatively short; meanwhile, it is considered a feasible and safe technique [7, 10, 11].

Traditional RF ablation typically employs low power long duration (LPLD) ablation (25-35 W), which is intended to produce mature ablated lesions and minimize complications [12, 13]. However, high-power ablation consistently resulted in shorter procedural times, reduced fluoroscopy dose, and decreased total RF energy delivery [14, 15]. Furthermore, high-power short-duration (HPSD) RF ablation is resistive heating and is more suitable for wide, continuous lesions while reducing complications associated with posterior wall isolation (PWI) [13, 16]. For example, Bhaskaran et al. [17] indicated that HPSD (50 W/5 seconds and 60 W/5 seconds) is as safe and effective as LPLD (40 W/30 s), with increased transmission and fewer complications. Winkle et al. [18] indicated that shorter duration 50 W lesions has been associated with better long-term outcomes without an increase in rates of complications. Additionally, high-power (40-50 W) RF ablation under the guidance of unipolar signal modification displays higher first-pass PVI and guarantees a robust one-year efficacy [19]. In recent animal studies, 50 W/5 seconds was superior to LPLD ablation for lesion creation with lower complication rates [17, 20]. In AI-guided high-power (50 W) ablation, the time to complete PVI is shorter, with higher first-pass isolation rate, lower left atrial-(LA-) PV reconnection rate, and no increase in complications, so it is considered a safe and feasible ablation technique [9, 21–23]. But reports on the safety and efficacy of AI-guided high-power RF ablation for paroxysmal atrial fibrillation are warranted.

In the present 2-tiered study, we first tested in a swine model the effect of different power outputs on lesion creation at the ventricular level and compared lesion dimensions for the same AI value. A clinical study then assessed whether AI-guided high-power ablation relative to AI-guided standard power ablation is relevant for PVI in cases of paroxysmal atrial fibrillation.

2. Methods

The present study was divided into 2 successive parts. An animal study provided efficacy and safety data, which oriented the clinical study for the choice of RF duration. The 2 parts followed the same technical principles for clinical research.

2.1. Animal Study. Swine hearts were obtained from an approved vendor (Shandong Animal Experiment Center) and studied at a laboratory of Yantai Yuhuangding Hospital. After pericardium excision, hearts were placed in a tissue bath with circulating arterial blood at a rate of 200 ml/min and temperature of 37°C within the thigh preparation chamber. Arterial blood was extracted (400 ml), and unfractionated heparin was added to maintain an activated clotting time of 350 to 400 s. Individual RF applications were performed at the endocardial aspect of the ventricles. RF ablation was performed with a 3.5 mm-tip open-irrigated contact force-sensing catheter (ThermoCool Smart Touch (ST), Biosense Webster, South Diamond Bar, USA). The catheter was positioned perpendicular to the left ventricular tissue and adjusted by the operator maintaining a constantly

desired contact-force (CF) of 10 g throughout energy delivery. For the same AI of 500, power output was 30 W for the control lesions and 40 W and 50 W for the study lesions, respectively, at ten ablation points. Ablation sites were > 10 mm apart to avoid any lesion overlap and to facilitate histopathologic evaluation. Lesions and necrotic areas were then bisected in two perpendicular planes to measure depth and maximal internal width. After proper fixation in 10% natural buffered formalin, hearts were trimmed to isolate all ablation sites, which were processed and embedded in paraffin. All paraffin blocks were subjected to microtome dissection twice serially at 5 µm, before being stained with hematoxylin and eosin for histopathologic analysis.

2.2. Clinical Study

2.2.1. Patient Population. The subjects were consecutive symptomatic patients with paroxysmal AF undergoing initial point-by-point RF ablation at Yantai Yuhuangding Hospital (Yantai, China) from April 2018 to February 2019. A total of 100 patients were included and randomly divided into the study group ($n = 50$) and the control group ($n = 50$). The control group was treated with standard power (30 W), whereas the study group was treated with high-power (40 W). Study exclusion criteria were left atrial thrombus, age < 18 years, prior AF ablation, left ventricular ejection fraction < 35%, severe valvular or coronary artery disease, thyrotoxicosis, and left atrial diameter > 60 mm. All patients provided written informed consent for the ablation procedure. The Ethics Committee of Yantai Yuhuangding Hospital (Yantai, China) approved the protocol of the present study.

2.3. Ablation Protocol. Our periprocedural anticoagulation protocols have been previously described [24]. The procedure was performed under local anesthesia consisting of lidocaine administration in the groin and left subclavian region and fentanyl as an analgesic. The CARTO3 system (Biosense Webster, Diamond Bar, USA) was used in all cases for 3D mapping. First, one 6F decapolar catheter was positioned in the coronary sinus, followed by double transseptal punctures, 2 SL1 (8F, Abbott, USA) sheaths were inserted into the LA, one for the multipolar mapping catheter to map the anatomy of the LA and the other for the 3.5 mm tip open-irrigated contact force-sensing catheter. After creating an anatomical CARTO map of the left atrium, the borders of all four pulmonary veins (PVs) were marked on the CARTO map. The patient underwent systemic anticoagulation with intravenous heparin during surgery to maintain an activated clotting time > 300 seconds.

Next, all 4 PVs underwent encirclement by point-by-point RF applications (target AI was 400 for the posterior segment, and 500 elsewhere) and other ablations as clinically indicated (Figure 1). RF applications were performed with power-control mode, temperature limited to 45°C, and saline irrigation (17 to 30 mL/min). Ablation was delivered aiming for a CF of 10-20 g. The procedural endpoint was PVI. The entrance block was confirmed by the absence of PV potentials recorded with the tissue proximity instruction (TPI) of pentaray catheter [25]. Exit block was defined as the

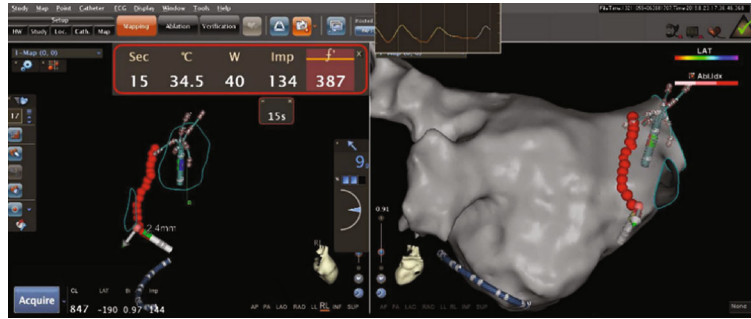


FIGURE 1: AI-guided ablation injury. VISITAGs placement with distance measure tool.

TABLE 1: Sources of drugs and equipment.

Name	Scientific name	Specification	Manufacturer	Approval number	Country
Lidocaine		2 ml/4 mg	Linyi Chenghui Pharmaceutical Co., Ltd.	SFDA approval number H37023396	China
Fentanyl		2 ml:0.1 mg	Jiangsu Nhwa Pharmaceutical Co., Ltd.	SFDA approval number H20113509	China
6F decapolar catheter	Multipolar diagnostic catheter	—	Biosense Webster	—	Diamond Bar, USA
Multipolar mapping catheter	Pentaray, Biosense Webster	—	Biosense Webster	—	Diamond Bar, USA
3.5 mm tip open-irrigated contact force-sensing catheter	ThermoCool Smart Touch (ST)	—	Biosense Webster	—	South Diamond Bar, USA

failure to capture the left atrium while pacing with the TPI of pentaray catheter within the antrum carina included [26]. In both groups, a total of 20 minutes waiting period was observed following PVI, after which 18 mg of adenosine was administered for each PV with the pentaray positioned to record any PV reconnection. The AI Software Module (Biosense Webster, Diamond Bar, CA) was used for determination of Automatic lesion tagging (VISITAG). Ablation was delivered with ST catheters in power-controlled mode with 40 W in study group and 30 W in the control group. The “distance ruler” function was used on CARTO; it shows a continuous measure of the distance from the center of the last VISITAG which changes in real-time (Figure 1). This distance was kept <6 mm when commencing an ablation lesion. VISITAGs were displayed showing ablation lesions with 3 mm radius so that VISITAGs < 6 mm apart could be seen to overlap (Figure 1). After the circumferential line of ablation was completed, the geometry was made transparent to look for any gaps between VISITAGs (which would indicate a gap > 6 mm). If any gaps were seen, they were filled with an ablation lesion. If any VISITAGs were not red (indicating that they had not met their regional AI target), then, it was left at the operator’s discretion as to whether to repeat the lesion. Only after the circumferential line of ablation was completed, if the PV was not isolated, further ablation was guided by the pentaray catheter delivered either on the circumferential line of ablation where there was signal or just within it. Ablation was not delivered within the PVs or on the intervenous ridge between PVs unless PV isolation could not be achieved without doing so. First-pass antral isolation designated exit block obtained after initial anatomic encir-

lement. The information on drugs and equipment used in this study are shown in Table 1.

2.4. Data Collection and Analysis. For each patient, we recorded age, gender, CHADS₂ ((congestive heart failure, hypertension, age ≥ 75 years), and diabetes, previous stroke (double weight)) and CHA₂DS₂-VASc (congestive heart failure, hypertension, age ≥ 75 years (double weight), diabetes, previous stroke (double weight), vascular disease, age 65–74 years, and female sex category) scores, body mass index (BMI), LA size, prior strokes/transient ischemic attacks (TIAs), and the presence of hypertension, diabetes, coronary artery disease (CAD), cardiomyopathy, and obstructive sleep apnea. For each ablation, we determined procedural, total RF times, and CF. Complications were reported on a per-procedure basis. Complications examined were death, incidence of pericardial tamponade, strokes occurring within 48 hours, strokes occurring from 48 hours–30 days, and PV stenosis requiring intervention, phrenic nerve paralysis, atrioesophageal fistulas, steam pops, and catheter char. Procedure time was defined as time from the first ablation point to PVI, not including additional ablation beyond PVI. The RF time was defined as the time for which RF energy was applied. Primary outcome measures were defined as no AF, atrial flutter, or atrial tachycardia (AT) lasting more than 30 seconds off antiarrhythmic drugs after a 3-month blanking period.

2.5. Follow-Up. All patients were discharged home within 48 to 72 hours. Vitamin K antagonists or direct oral anticoagulants were prescribed for at least 2 months (subsequent strategy depending on the CHA₂DS₂-VASc score). Patients were

TABLE 2: Swine ventricular lesion and necrosis characteristics.

Variable	50 w	40 w	30 w
Ablation time per point, s*	20.00 ± 1.10	27.78 ± 9.72	44.83 ± 9.83
Ventricular lesion impedance drop, Ω *	13.5 ± 1.87	9.11 ± 1.54	6.17 ± 0.98
Energy delivery per point, J*	992.30 ± 54.33	1085.10 ± 45.02	1337.0 ± 24.42
Ventricular tissue lesion depth, mm*	3.95 ± 0.16	4.38 ± 0.13	5.06 ± 0.16
Ventricular tissue necrosis depth, mm*	3.15 ± 0.18	2.71 ± 0.17	2.42 ± 0.13
Ventricular tissue lesion width, mm*	9.08 ± 0.15	8.42 ± 0.18	7.81 ± 0.15
Ventricular tissue necrosis width, mm*	5.58 ± 0.18	5.18 ± 0.16	3.94 ± 0.17

*All p values < 0.001.

followed up in clinic with an ECG at 3 and 9 months with 24 hours. Holter monitoring was performed at 6 and 12 months and beyond as dictated by symptoms.

2.6. Statistical Analysis. For each part of the study, variables are presented as mean ± SD or percentages, as appropriate. Continuous and categorical data were compared with the Student t test (two-tailed) and the χ^2 test (or Fisher exact test in case of small sample), respectively. $p < 0.05$ was considered as statistically significant.

3. Results

3.1. Animal Study. Table 2 shows a comparison of lesions and necrosis in control (30 W) and study (40 W, 50 W) swine hearts for a total of 30 RF applications. By visual inspection, 50 W use was associated with shallower lesions; and increasing RF power and shortening RF duration yielded greater maximum diameter (green line) and larger and deeper tissue necrosis area (red line) and nuclear pyknosis by histopathology (Figure 2).

3.2. Clinical Study. Since there was a small amount of scab on the catheter tip during ablation at 50 W in animal experiments, we chose 40 W as the high-power ablation in the clinical study. A total of 100 consecutive patients with paroxysmal AF were evenly distributed to control group (first 50 patients) and study group (last 50 patients); their baseline clinical characteristics are summarized in Table 3. A total of 8726 RF lesions were delivered in these 100 patients, and procedural data are summarized in Table 4. Figure 3 shows the distribution of lesions by ranges of CF. The study vs. control group had shorter total procedure time ($p < 0.001$), left and right encirclement procedure time ($p < 0.001$), total ablation time required for PVI ($p < 0.001$), ablation time per point in different parts of PV, and total RF energy delivered per procedure ($p < 0.001$; Figure 4). Figure 5 shows the energy delivery per point throughout PV. First-pass isolation was more frequent in the study group ($p < 0.01$), occurring in both left and right in 70% PVs (35/50) of study patients vs. 42% (21/50) of controls ($p < 0.001$). The study vs. control group required fewer additional ablations on the intervenous ridge between the PVs to isolate them ($p < 0.001$), and had fewer acute PV reconnections ($p < 0.001$).

At mean 1.04 ± 0.62 years follow-up, the percentage of patients free from AF/AT after a single procedure was 92% (46/50) in the study group and 84% (42/50) in the control group ($p = 0.22$). There were no cases of stroke, TIA, pericardial tamponade, atrial-esophageal fistulae, PV stenosis, or death.

4. Discussion

In the present study, two-tiered animal and clinical study yielded four major findings. First, in swine hearts, higher vs. lower power RF applications were more effective in creating larger and deeper necrosis yet shallower lesions, while lower power was associated with deeper lesions without basophilic changes in connective tissue. Second, in patients with paroxysmal AF, high-power ablation was associated with shorter procedure times and less total RF energy delivery, especially on the left anterior wall. Third, AI, the novel marker incorporating contact force-time power, reliably predicted the degree of necrosis in RF delivery. Fourth, higher power combined with AI increased PVI effectiveness with more frequent first-pass isolation, decreased acute reconnection, and favorable 12-month outcomes.

4.1. High-Power Ablation. Myocardial lesion creation starts at 45°C , being partially reversible below 50°C (transient stunning), and definitive above 50°C (durable necrosis) [27]. RF energy delivery to tissue is a complex interaction [1]. Thermal injury induced by electrical current delivery with an irrigated-tip comprises resistive and conductive phases. Resistive heating, which probably occurs relatively early in the RF application, depends upon current delivered to the tissue and the resistance seen by the RF generator. Greater resistive heating can be achieved by the use of higher RF power or lower resistance. For instance, with standard power (25–30 W), temperature already rises above 50°C , but tissue necrosis is confined to the first 1 to 1.5 mm from the ablation catheter tip [28]. Conductive heating is secondary passive heating of deeper tissue which increases with longer duration RF applications. Tissue needs to be heated to 50°C or higher for several seconds to achieve irreversible coagulation necrosis which results in an electrically silent scar. Force sensing and stability monitoring have considerably facilitated the reproducibility of heat transmission to the tissue [29, 30]. Thus, the balance between power and duration parameters involved in resistive and conductive heating, respectively, has an increasing impact on

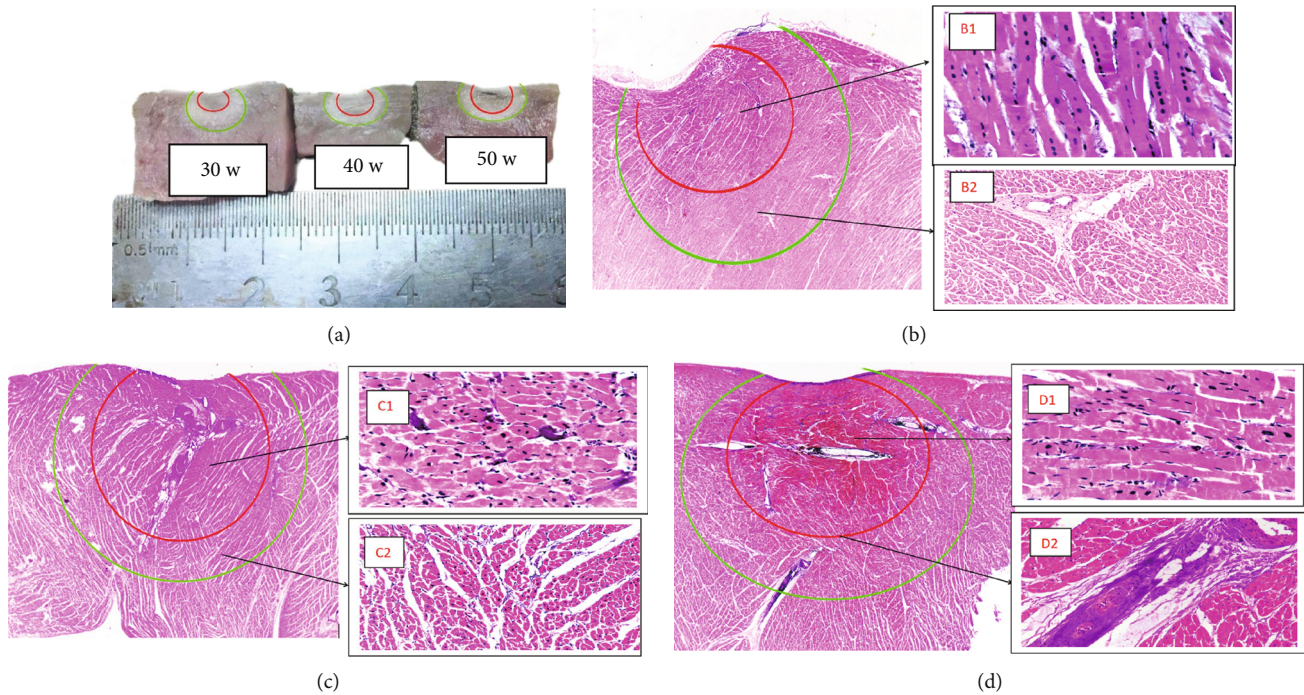


FIGURE 2: Ventricular lesions obtained with constant AI of 500 and with 50 W, 40 W, or 30 W of power output. (a) By visual inspection, increasing power was associated with larger lesion maximum diameter (green line) and deeper tissue necrosis (red line), with boundaries of tissue lesion and necrosis becoming clearer. (b–d) Lesions (green line) and necrosis (red line) generated with 30 W, 40 W, and 50 W were examined under 20 \times amplification; under 400 \times magnification at 1 mm from the ablation catheter tip in panels B1 (mainly basophilic changes of connective tissue with least myocardial cell changes and nuclear pyknosis), C1 (small number of myocardial cells with fuzzy sarcoplasm, no horizontal stripes and nuclear pyknosis), and D1 (largest number of affected myocardial cells); and under 200 \times magnification at 3 mm from the ablation catheter tip in panels B2 (no basophilic changes of fibrous connective tissue), C2 (basophilic changes only around cells), and D2 (basophilic changes around the blood vessels and cells).

TABLE 3: Clinical characteristics.

Variable	Study group High-power ($n = 50$)	Control group Standard power ($n = 50$)	p value
Age, y*	64.4 \pm 9.45	64.9 \pm 8.62	0.862
Male	34 (68)	32 (64)	0.673
LV ejection fraction, %	64.74 \pm 4.46	61.9 \pm 5.40	0.08
Left atrial size, mm	40.65 \pm 5.87	42.85 \pm 3.10	0.147
Hypertension	22 (44)	26 (52)	0.423
Diabetes mellitus	6 (12)	8 (16)	0.564
Body mass index	21.27 \pm 1.88	21.04 \pm 2.65	0.749
Prior stroke/transient ischemic attack	8 (16)	10 (20)	0.603
Coronary artery disease n	14 (28)	12 (24)	0.648
CHADS2 score	2.1 \pm 0.64	1.95 \pm 0.80	0.504
CHA2DS2-VASC score	3.4 \pm 1.14	3.15 \pm 1.04	0.474
Obstructive sleep apnea	6 (12)	7 (14)	0.766
Dilated cardiomyopathy	3 (6)	3 (6)	1

Data are presented as mean \pm SD or n (%).

lesion creation. By increasing resistive heating size, high-power (40–50 W) may theoretically be beneficial for the generation of durable lesions (temperature above 50 $^{\circ}$ C) whose dimensions may be particularly suitable for PVI as antral thickness is consistently below 4 mm [31].

4.2. *Animal Experiments.* Conventional LPLD (25–35 W) ablation is used to generate mature ablated lesions and reduce complications [12, 13]. But recent studies have shown that HPSD is as safe and effective as LPLD while shortening the operative time, and it has a higher first-pass

TABLE 4: Procedural characteristics.

Variable	Study group High power (40 W) <i>n</i> = 50	Control group Standard power (30 W); <i>n</i> = 50	<i>p</i> value
Procedure time, min	56.54 ± 1.81	76.55 ± 2.34	<0.001
Right encirclement procedure time, min	30.92 ± 1.31	38.33 ± 3.06	<0.001
Left encirclement procedure time, min	25.83 ± 1.12	37.92 ± 1.24	<0.001
Right encirclement points, <i>n</i>	51.25 ± 2.45	49.08 ± 3.58	0.1
Left encirclement points, <i>n</i>	44.42 ± 1.44	45.00 ± 1.91	0.407
Total ablation time, min	35.85 ± 14.87	51.01 ± 17.99	<0.001
Contact-force, g*	12.07 ± 5.34	11.85 ± 5.40	0.523
Ablation time per point, s	22.64 ± 9.39	32.56 ± 11.48	<0.001
Energy delivery per point, J*	909.02 ± 354.57	1045 ± 376.60	<0.001
Impedance drop per point, Ω *	10.13 ± 1.624	6.57 ± 1.012	<0.001
First-pass PVI	87 (87)	72 (72)	<0.01
Reconnection after 20 min	22 (11)	46 (23)	<0.01
Groin hematoma	1	1	1
Tamponade	0	0	1
Periprocedural stroke	0	0	1
Esophageal fistula	0	0	1
Sinus rhythm at 12 months	46 (92)	42 (84)	0.22

Data are presented as mean ± SD or *n* (%).

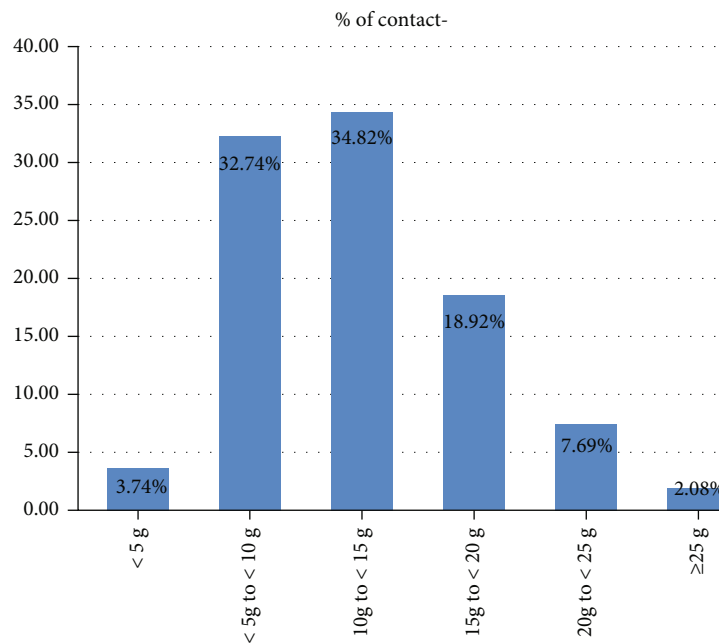


FIGURE 3: Percentage of total number of RF lesions by average CF ranges.

PVI rate and lower acute pulmonary vein reconnection (PVR) [17, 32, 33]. Additionally, Borne et al. [34] found using a porcine *in vivo* model that although a 50 W/5-second ablation is similar in volume to a 20 W/30-second ablation, it is shallower in depth while potentially causing less collateral damage. Hence, in this study, we divided the porcine left ventricle into control (30 W) and study groups (40 W and 50 W) and performed 30 RF applications followed by histo-

pathological evaluation. Results indicated that 50 W power ablations for short durations at 10 g of CF created larger and deeper necrosis than that observed with 40 W and 30 W of power; however, the use of 30 W resulted in deeper lesions but without basophilic changes which might render them more susceptible to tissue recovery. HPSD applications were more effective in creating larger and deeper necrosis yet shallower lesions may be due to reduced temperature rise in

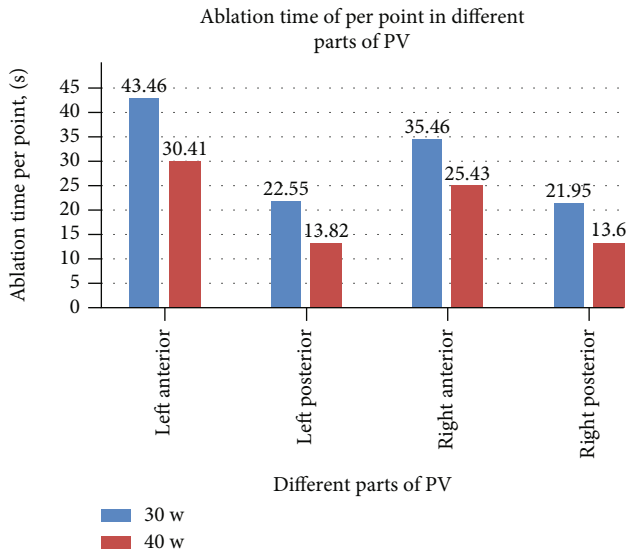


FIGURE 4: Ablation time per point throughout PV.

deeper tissue relative to standard lesions. However, the area of tissue necrosis near the surface of the ablation catheter tip is larger because it is more dependent on resistive heating, which is proportional to power. The larger diameter of necrosis in study group might contribute to complete encirclement of pulmonary veins, by ensuring better contiguity with adjacent necrosis. At the same time, it was found that higher power (50 W) ablation may lead to more complete cell necrosis. However, nuclear pyknosis was observed only more superficially with low-power ablation, and only cell edema was present in the deep part, which may lead to cell reactivation and consequent acute and late PV reconnection and AF recurrence. Since there was a small amount of scab on the catheter tip during 50 W ablation in animal experiments, we chose 40 W as high-power ablation and 30 W as control ablation in the clinical study.

4.3. Clinical Research

4.3.1. Procedural Time. In the present study, high-power ablation was associated with shorter procedural duration due to the shorter time required for lesion creation, more first-pass PVI, and fewer acute pulmonary vein reconnections. It is known that the first 10 seconds of RF application are the most important for lesion formation with diminishing effect beyond 20 seconds, and the ideal contact force is 10-20 g [32]. Nilsson et al. [15] reported that ablation with 45 W for 20 seconds vs. 30 W for 120 seconds was associated with shorter PV isolation time, mean fluoroscopy time, radiation dose, and total RF application time. In ventricles from freshly killed pigs, Goyal et al. [20] showed that for 20 g of CF, the time needed to create a 4 mm deep lesion decreased from just over 20 seconds for 20 W to 6-7 seconds for 50 W. In the present study, under the same pressure, the ablation time per point of the left atrial posterior wall in the high-power group was 8-9 seconds shorter than that in the control group and 10-13 seconds shorter for the left atrial anterior wall. Because catheter instability in a constantly beating

heart may also account for the difficulty to transmit heat to the tissue, RF application time shortening probably optimizes lesion creation by increasing the likelihood of catheter stability throughout the entire RF application, particularly for ablation of the left atrial anterior wall by the left atrial appendage ridge. The significantly smaller pressure of <10 g (60%) used for the left atrial anterior wall than for the other three surfaces helps avoid catheter slippage (Figure 6). However, low power use requires longer time which increases catheter slippage, and the additional time spent adjusting the catheter may cause discontinuous ablation tissue edema, lower first-pass PVI rate, and increased risk of acute PV reconnections. In the present study, ablation time of the left anterior wall was longer than that of other parts of the left atrium in both study and control groups (Figure 7); however, it was shorter in the study than the control group likely secondary to shorter time required to adjust the catheter or the ineffective point of ablation (procedure time minus total ablation time).

4.4. AI Value. AI is a novel marker incorporating contact force, time, and power in a weighted formula. The use of contact force targets and markers of ablation output such as FTI reduces recurrence and complication rates in cohorts of patients with AF undergoing PVI [3, 35, 36]. Single-center studies on AI-guided ablation reported very high rates of first-pass PVI (97-98%) and very low rates of acute PV reconnection (2-6%) [36, 37]. In the present study on AI-guided high-power ablation, rates of first-pass PVI were high at 87% albeit lower than in previous reports. Similarly, rates of acute PV reconnection were lower than in controls (11%). AI targets are arbitrary and published targets range from 380 to 400 for the posterior wall and 550 elsewhere. This study used a novel combination of high-power delivered by ST catheters. In this context, an AI target of 400 for the posterior wall and 500 elsewhere led to high rates of first-pass PVI and low rates of acute PV reconnection.

4.5. First-Pass Pulmonary Vein Isolation. An earlier study displayed increased efficacy of AI-guided catheter ablation while having comparable safety to non-AI catheter ablation [10]. Chen et al. [38] indicated that AI-guided high-power (50 W) ablation increases first-pass PVI rates. Leshem et al. [13] compared ablation using 90 W for 4 seconds to 25 W for 20 seconds. The 90 W/4-second ablation resulted in full-thickness lesions with no gaps in all cases, whereas the 25 W/20-second ablation resulted in some partial-thickness lesions and many gaps between lesions. At 25 W for 20 seconds, irrigation of the catheter tip seemed to cause “endocardial sparing,” thought to be a failure to create scar because of cooling of the endocardium by irrigation before resistive heating can destroy the tissue. In the present animal study experiment, use of higher power was associated with larger maximum diameter and more thorough tissue necrosis. Greater size and better consistency of the created tissue necrosis may explain why high-power increases procedure efficiency by ensuring more first-pass PVI and fewer reconnections at 20 minutes.

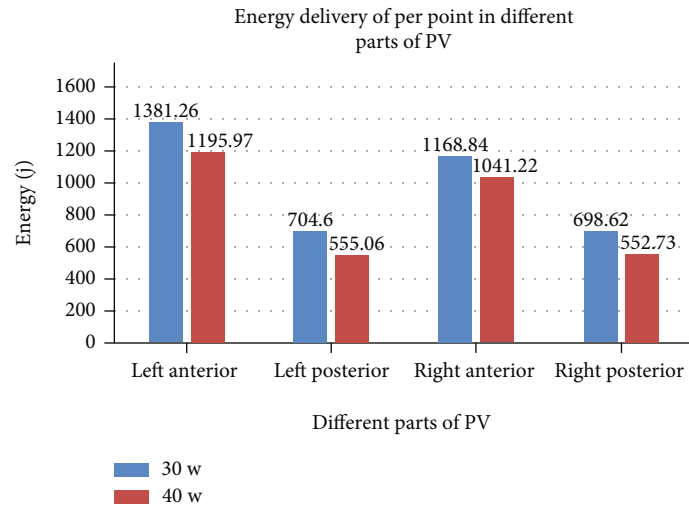


FIGURE 5: Energy delivery per point throughout PV.

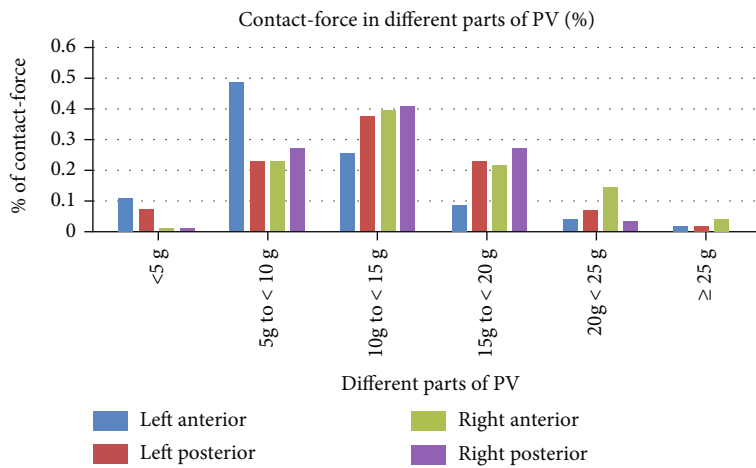


FIGURE 6: Distribution of contact-force throughout PV.

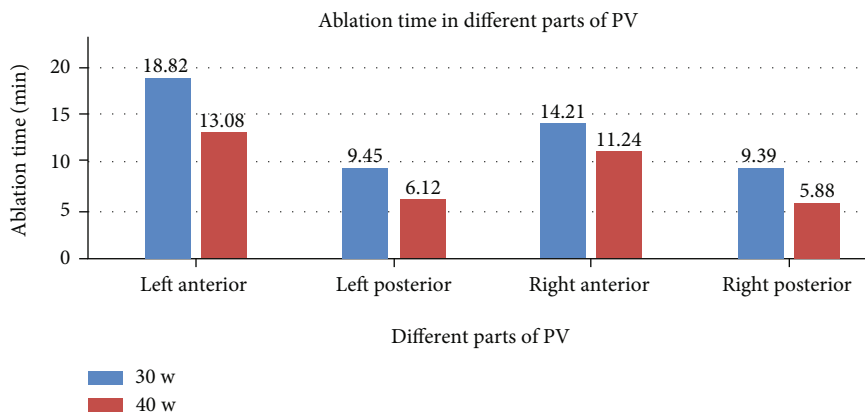


FIGURE 7: Ablation time throughout PV.

4.6. *Acute PV Reconnections.* Durability of PVI is important to procedural success; both acute and late PV reconnection have been associated with AF recurrence [5, 39]. In a study [14] using an open irrigated-tip catheter, patients undergoing ablation at 50 W vs. 35 W had greater freedom from

AF (82% vs. 66%). The idea of the “weakest link” predicting sites of reconnection was described recently by El Haddad and colleagues [40], who found that sites of reconnection were associated with poor lesion depth and lack of contiguous lesion sets. In addition, Dhillon et al. [41] found reduced

operative time and reduced acute PV reconnection in patients with paroxysmal atrial fibrillation who underwent AI-guided high-power ablation. As such, the present one documented fewer acute PV reconnections in the high- vs. standard-power group, and in an animal model, higher power caused more thorough tissue necrosis thereby precluding cell restoration and reducing acute pulmonary vein connection.

4.7. Complications. HPSD RF delivery is thought to destroy tissue, mostly through local resistive heating, which occurs early during an RF application, by avoiding the distant conductive heating tissue damage that predominates later during long RF applications. In freshly killed porcine ventricles, Goyal et al. [20] showed that the time to create a 4 mm deep lesion was 20 seconds for 20 W ablations and only 6–7 seconds for 50 W ablations, which suggested that high-power, short-duration RF applications might help reduce collateral injury. Bhaskaran et al. [17] showed that 50 W and 60 W ablations for 5 seconds achieved transmural lesions and were safer than 40 W ablations for 30 seconds. The incidence of steam pops was 8% in the 40 W/30-second ablations vs. none in the 50 W and 60 W ablations for 5 seconds. Winkle et al. [18] compared the use of open irrigated-tip catheters at 50 W for short durations of 3–10 seconds at each site to lower power applied for 25–40 seconds at each site. 50 W ablations had longer-term freedom from AF and shorter procedural, fluoroscopy times without an increase in complications. Winkle et al. [42] examined the complication rates of 4 experienced centers performing AF ablations at RF powers from 45–50 W for 2–15 seconds per lesion while power was reduced to 35 W for 20 seconds, AF ablations can be performed at 45–50 W for short durations with very low complication rates. Our clinical results cannot provide definitive information (for example using esophageal temperature monitoring) on whether high-power, short-duration lesions were safer than low-power energy delivered for a longer time. However, our histological results confirmed shallower tissue lesions with high-power than low-power ablation, with shorter ablation time reflecting shorter time for catheter to attach to the atrial wall tissue, thus leading to higher safety. In addition, the extremely low complication rate may be reassuring for using 40 W for short durations in left atrium, even posterior wall and may encourage considering use of short-duration, higher-power RF ablations to take advantage of the reduction in procedural, fluoroscopy, and total RF energy delivery times.

4.8. Limitations. Although the results of this study are encouraging, their robustness is limited by several factors. First, degree of esophageal injury during ablation is unclear because all patients were treated under local anesthesia and could not tolerate esophageal temperature monitoring, and no postoperative endoscopic examination was performed. Second, the relatively fixed AI values used may be insufficient or excessive for some patients, especially for thin women. Third, because ST was used instead of Smart Touch Surround Flow (STSF) catheter, there is a small amount of scab at the catheter tip during the 50 W ablation in the ani-

mal study, for which, 50 W was not selected for ablation in clinical research. Fourth, after a year, there was a similar sinus rhythm maintenance rate after ablation with 30 W and 40 W which may be because the numbers in each group were probably too small to look at a real difference in these two techniques. Fifth, the animal study is carried out in ventricles while the human study is in atrial. There may be a lack of consistency.

5. Conclusions

In the present study of high-power ablation, use of guidance by AI, which provides a rational local endpoint allowing for a tailored short-duration radiofrequency application associated with optimized lesion metrics, translated into shorter procedural time and improved acute efficacy, without compromising safety profile and long-term outcomes relative to standard-power ablation.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

Additional Points

Posted History. A preprint has previously been published [43]. The link is as follows <https://www.researchsquare.com/article/rs-21367/v2>.

Ethical Approval

The Ethics Committee of Yantai Yuhuangding Hospital (Yantai, China) approved the protocol of the present study (ID: 2017203).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Xuefeng Zhu and Chunxiao Wang contributed equally to this work.

References

- [1] S. Kumar, C. R. Barbaiya, S. Balindger et al., "Better lesion creation and assessment during catheter ablation," *Journal of Atrial Fibrillation*, vol. 8, no. 3, p. 1189, 2015.
- [2] W. Ullah, A. McLean, M. H. Tayebjee et al., "Randomized trial comparing pulmonary vein isolation using the SmartTouch catheter with or without real-time contact force data," *Heart Rhythm*, vol. 13, no. 9, pp. 1761–1767, 2016.
- [3] J. Kautzner, P. Neuzil, H. Lambert et al., "EFFICAS II: optimization of catheter contact force improves outcome of pulmonary vein isolation for paroxysmal atrial fibrillation," *Europace*, vol. 17, no. 8, pp. 1229–1235, 2015.
- [4] J. G. Andrade, G. Monir, S. J. Pollak et al., "Pulmonary vein isolation using "contact force" ablation: the effect on dormant

- conduction and long-term freedom from recurrent atrial fibrillation—a prospective study,” *Heart Rhythm*, vol. 11, pp. 1919–1924, 2014.
- [5] E. Anter, F. M. Contreras-Valdes, A. Shvilkin, C. M. Tschabrunn, and M. E. Josephson, “Acute pulmonary vein reconnection is a predictor of atrial fibrillation recurrence following pulmonary vein isolation,” *Journal of Interventional Cardiac Electrophysiology*, vol. 39, no. 3, pp. 225–232, 2014.
 - [6] E. Anter, C. M. Tschabrunn, F. M. Contreras-Valdes, A. E. Buxton, and M. E. Josephson, “Radiofrequency ablation annotation algorithm reduces the incidence of linear gaps and reconnection after pulmonary vein isolation,” *Heart Rhythm*, vol. 11, no. 5, pp. 783–790, 2014.
 - [7] Y. J. Wang, Y. Tian, L. Shi et al., “Pulmonary vein isolation guided by moderate ablation index targets combined with strict procedural endpoints for patients with paroxysmal atrial fibrillation,” *Journal of Cardiovascular Electrophysiology*, vol. 32, no. 7, pp. 1842–1848, 2021.
 - [8] F. Solimene, V. Schillaci, G. Shopova et al., “Safety and efficacy of atrial fibrillation ablation guided by ablation index module,” *Journal of Interventional Cardiac Electrophysiology*, vol. 54, no. 1, pp. 9–15, 2019.
 - [9] H. Okamoto, K. Okumura, S. Kaneko et al., “Ablation index-guided high-power radiofrequency application shortens the procedure time with similar outcomes to conventional power application in atrial fibrillation ablation,” *Circulation Reports*, vol. 3, no. 10, pp. 559–568, 2021.
 - [10] A. Ioannou, N. Papageorgiou, W. Y. Lim et al., “Efficacy and safety of ablation index-guided catheter ablation for atrial fibrillation: an updated meta-analysis,” *Europace*, vol. 22, no. 11, pp. 1659–1671, 2020.
 - [11] N. Reinsch, A. Fütting, J. Buchholz et al., “One-year outcome and durability of pulmonary vein isolation after prospective use of ablation index for catheter ablation in patients with persistent atrial fibrillation,” *Journal of Interventional Cardiac Electrophysiology*, vol. 62, no. 1, pp. 143–151, 2021.
 - [12] G. Liu, S. Zhang, H. Hu, T. Liu, and J. Huang, “The role of neutrophil-lymphocyte ratio and lymphocyte-monocyte ratio in the prognosis of type 2 diabetics with COVID-19,” *Scottish Medical Journal*, vol. 65, no. 4, pp. 154–160, 2020.
 - [13] E. Leshem, I. Zilberman, C. M. Tschabrunn et al., “High-power and short-duration ablation for pulmonary vein isolation: biophysical characterization,” *JACC Clin Electrophysiol*, vol. 4, no. 4, pp. 467–479, 2018.
 - [14] M. H. Kanj, O. Wazni, T. Fahmy et al., “Pulmonary vein antral isolation using an open irrigation ablation catheter for the treatment of atrial fibrillation: a randomized pilot study,” *Journal of the American College of Cardiology*, vol. 49, no. 15, pp. 1634–1641, 2007.
 - [15] B. Nilsson, X. Chen, S. Pehrson, and J. H. Svendsen, “The effectiveness of a high output/short duration radiofrequency current application technique in segmental pulmonary vein isolation for atrial fibrillation,” *Europace*, vol. 8, no. 11, pp. 962–965, 2006.
 - [16] R. A. Winkle, “High-power short-duration ablation: turn up the heat to cool down the esophagus,” *Journal of Cardiovascular Electrophysiology*, vol. 30, no. 10, pp. 1884–1885, 2019.
 - [17] A. Bhaskaran, W. Chik, J. Pouliopoulos et al., “Five seconds of 50-60 W radio frequency atrial ablations were transmural and safe: an in vitro mechanistic assessment and force-controlled in vivo validation,” *Europace*, vol. 19, no. 5, pp. 874–880, 2017.
 - [18] R. A. Winkle, R. H. Mead, G. Engel, and R. A. Patrawala, “Atrial fibrillation ablation: “perpetual motion” of open irrigated tip catheters at 50 W is safe and improves outcomes,” *Pacing and Clinical Electrophysiology*, vol. 34, no. 5, pp. 531–539, 2011.
 - [19] T. Pambrun, C. Durand, M. Constantin et al., “High-power (40-50 W) radiofrequency ablation guided by unipolar signal modification for pulmonary vein isolation: experimental findings and clinical results,” *Circulation. Arrhythmia and Electrophysiology*, vol. 12, no. 6, article e007304, 2019.
 - [20] F. Ali-Ahmed, V. Goyal, M. Patel, F. Orelaru, D. E. Haines, and W. S. Wong, “High-power, low-flow, short-ablation duration—the key to avoid collateral injury?,” *Journal of Interventional Cardiac Electrophysiology*, vol. 55, no. 1, pp. 9–16, 2019.
 - [21] H. Okamoto, J. Koyama, Y. Sakai et al., “High-power application is associated with shorter procedure time and higher rate of first-pass pulmonary vein isolation in ablation index-guided atrial fibrillation ablation,” *Journal of Cardiovascular Electrophysiology*, vol. 30, no. 12, pp. 2751–2758, 2019.
 - [22] S. Chen, B. Schmidt, S. Bordignon et al., “Ablation index-guided 50 W ablation for pulmonary vein isolation in patients with atrial fibrillation: procedural data, lesion analysis, and initial results from the FAFA AI high power study,” *Journal of Cardiovascular Electrophysiology*, vol. 30, no. 12, pp. 2724–2731, 2019.
 - [23] S. Chen, B. Schmidt, A. Seeger et al., “Catheter ablation of atrial fibrillation using ablation index-guided high power (50 W) for pulmonary vein isolation with or without esophageal temperature probe (the AI-HP ESO II),” *Heart Rhythm*, vol. 17, no. 11, pp. 1833–1840, 2020.
 - [24] R. A. Winkle, R. H. Mead, G. Engel, M. H. Kong, and R. A. Patrawala, “Peri-procedural interrupted oral anticoagulation for atrial fibrillation ablation: comparison of aspirin, warfarin, dabigatran, and rivaroxaban,” *Europace*, vol. 16, no. 10, pp. 1443–1449, 2014.
 - [25] H. Calkins, G. Hindricks, R. Cappato et al., “2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation,” *Europace*, vol. 20, no. 1, pp. e1–e160, 2018.
 - [26] T. Pambrun, S. Combes, P. Sousa et al., “Contact-force guided single-catheter approach for pulmonary vein isolation: feasibility, outcomes, and cost-effectiveness,” *Heart Rhythm*, vol. 14, no. 3, pp. 331–338, 2017.
 - [27] S. Nath, C. Lynch, J. G. Wayne, and D. E. Haines, “Cellular electrophysiological effects of hyperthermia on isolated Guinea pig papillary muscle. Implications for catheter ablation,” *Circulation*, vol. 88, pp. 1826–1831, 1993.
 - [28] F. H. Wittkampf and H. Nakagawa, “RF catheter ablation: lessons on lesions,” *Pacing and Clinical Electrophysiology*, vol. 29, no. 11, pp. 1285–1297, 2006.
 - [29] E. Marijon, S. Faza, K. Narayanan et al., “Real-time contact force sensing for pulmonary vein isolation in the setting of paroxysmal atrial fibrillation: procedural and 1-year results,” *Journal of Cardiovascular Electrophysiology*, vol. 25, no. 2, pp. 130–137, 2014.
 - [30] G. Zucchelli, G. Sirico, L. Rebellato et al., “Contiguity between ablation lesions and strict catheter stability settings assessed by VISITAGTM module improve clinical outcomes of paroxysmal atrial fibrillation ablation — results from the VISITALY study —,” *Circulation Journal*, vol. 82, no. 4, pp. 974–982, 2018.

- [31] S. Y. Ho, D. Sanchez-Quintana, J. A. Cabrera, and R. H. Anderson, "Anatomy of the left atrium: implications for radiofrequency ablation of atrial fibrillation," *Journal of Cardiovascular Electrophysiology*, vol. 10, no. 11, pp. 1525–1533, 1999.
- [32] R. A. Winkle, "HPSD ablation for AF high-power short-duration RF ablation for atrial fibrillation: a review," *Journal of Cardiovascular Electrophysiology*, vol. 32, no. 10, pp. 2813–2823, 2021.
- [33] D. Khanra, A. Hamid, S. Deshpande et al., "Meta-analysis comparing outcomes of high-power short-duration and low-power long-duration radiofrequency ablation for atrial fibrillation," *Anatolian Journal of Cardiology*, vol. 26, no. 1, pp. 2–14, 2022.
- [34] R. T. Borne, W. H. Sauer, M. M. Zipse, L. Zheng, W. Tzou, and D. T. Nguyen, "Longer duration versus increasing power during radiofrequency ablation yields different ablation lesion characteristics," *JACC: Clinical Electrophysiology*, vol. 4, no. 7, pp. 902–908, 2018.
- [35] M. Das, J. J. Loveday, G. J. Wynn et al., "Ablation index, a novel marker of ablation lesion quality: prediction of pulmonary vein reconnection at repeat electrophysiology study and regional differences in target values," *Europace*, vol. 19, pp. 775–783, 2016.
- [36] P. Taghji, M. el Haddad, T. Philips et al., "Evaluation of a strategy aiming to enclose the pulmonary veins with contiguous and optimized radiofrequency lesions in paroxysmal atrial fibrillation: a pilot study," *JACC Clin Electrophysiol*, vol. 4, no. 1, pp. 99–108, 2018.
- [37] A. Hussein, M. das, V. Chaturvedi et al., "Prospective use of ablation index targets improves clinical outcomes following ablation for atrial fibrillation," *Journal of Cardiovascular Electrophysiology*, vol. 28, no. 9, pp. 1037–1047, 2017.
- [38] S. Chen, B. Schmidt, S. Bordignon et al., "Catheter ablation of atrial fibrillation using ablation index-guided high-power technique: Frankfurt AI high-power 15-month follow-up," *Journal of Cardiovascular Electrophysiology*, vol. 32, no. 3, pp. 616–624, 2021.
- [39] M. Efremidis, K. Letsas, G. Giannopoulos et al., "Early pulmonary vein reconnection as a predictor of left atrial ablation outcomes for paroxysmal atrial fibrillation," *Europace*, vol. 17, no. 5, pp. 741–746, 2015.
- [40] M. El Haddad, P. Taghji, T. Philips et al., "Determinants of acute and late pulmonary vein reconnection in contact force-guided pulmonary vein isolation: identifying the weakest link in the ablation chain," *Circulation. Arrhythmia and Electrophysiology*, vol. 10, no. 4, 2017.
- [41] G. Dhillon, S. Ahsan, S. Honarbakhsh et al., "A multicentered evaluation of ablation at higher power guided by ablation index: establishing ablation targets for pulmonary vein isolation," *Journal of Cardiovascular Electrophysiology*, vol. 30, no. 3, pp. 357–365, 2019.
- [42] R. A. Winkle, S. Mohanty, R. A. Patrawala et al., "Low complication rates using high power (45-50 W) for short duration for atrial fibrillation ablations," *Heart Rhythm*, vol. 16, no. 2, pp. 165–169, 2019.
- [43] X. Zhu, C. Wang, J. Li, W. Li, H. Chu, and H. Zhou, *High-Power Radiofrequency Ablation Guided by Ablation Index for Pulmonary Vein Isolation*, Research Square, 2020.

Research Article

Comprehensive Multiomics Analysis Reveals Potential Diagnostic and Prognostic Biomarkers in Adrenal Cortical Carcinoma

Xiunan Li,¹ Jiayi Li ,² Leizuo Zhao,^{3,4} Zicheng Wang,⁵ Peizhi Zhang ,³ Yingkun Xu ,⁶ and Guangzhen Wu ¹

¹Department of Urology, The First Affiliated Hospital of Dalian Medical University, Dalian 116011, China

²School of Business, Hanyang University, Seoul 15588, Republic of Korea

³Department of Urology, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University, Jinan 250021, China

⁴Department of Urology, Dongying People's Hospital, Dongying 257000, China

⁵Department of Urology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan 250021, China

⁶Department of Breast and Thyroid Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400042, China

Correspondence should be addressed to Yingkun Xu; yingkunxu@hotmail.com and Guangzhen Wu; wuguang0613@hotmail.com

Received 22 April 2022; Revised 6 July 2022; Accepted 9 July 2022; Published 9 August 2022

Academic Editor: Lei Chen

Copyright © 2022 Xiunan Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Adrenal cortical carcinoma (ACC) is a severe malignant tumor with low early diagnosis rates and high mortality. In this study, we used a variety of bioinformatic analyses to find potential prognostic markers and therapeutic targets for ACC. Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) data sets were used to perform differential expressed analysis. WebGestalt was used to perform enrichment analysis, while String was used for protein-protein analysis. Our study first detected 28 up-regulation and 462 down-regulation differential expressed genes through the GEO and TCGA databases. Then, GO functional analysis, four pathway analyses (KEGG, REACTOME, PANTHER, and BIOCYC), and protein-protein interaction network were performed to identify these genes by WebGestalt tool and KOBAS website, as well as String database, respectively, and finalize 17 hub genes. After a series of analyses from GEPIA, including gene mutations, differential expression, and prognosis, we excluded one candidate unrelated to the prognosis of ACC and put the remaining genes into pathway analysis again. We screened out CCNB1 and NDC80 genes by three algorithms of Degree, MCC, and MNC. We subsequently performed genomic analysis using the TCGA and cBioPortal databases to better understand these two hub genes. Our data also showed that the CCNB1 and NDC80 genes might become ACC biomarkers for future clinical use.

1. Introduction

Adrenal cortical carcinoma (ACC) originates from the adrenal cortex and is a rare clinical malignant endocrine tumor [1], with a population incidence of 0.001‰ to 0.002‰ [2]. Still, it is also the most common primary malignant tumor of the adrenal gland [3] and is the second most common malignant tumor of the endocrine organ after thyroid cancer [4]. ACC can occur at any age, with two peaks in childhood and between 50 and 70, and is more common in women [5–7]. The clinical manifestations of ACC are diverse and prone to invasion and metastasis. Due to the low early diagnosis rate and high mortality, the survival period is generally

less than three years [8], and the 5-year survival rate is only 10% to 20% [9], which greatly threatens the life and health of patients. There is currently no effective early diagnosis and late treatment for ACC, and complete surgical resection is the only possible cure for ACC [10–13]. Therefore, finding novel biomarkers for efficient screening in the early stages of ACC may be valuable for long-term survival.

It is also worth noting that adrenocortical adenocarcinomas have distinct gene expression profiles from adrenocortical adenomas. The most widely recognized gene at present is the gene IGF2. The expression of IGF2 in adrenocortical adenocarcinoma is higher than that in adrenocortical adenoma. However, the differential diagnosis of adrenocortical

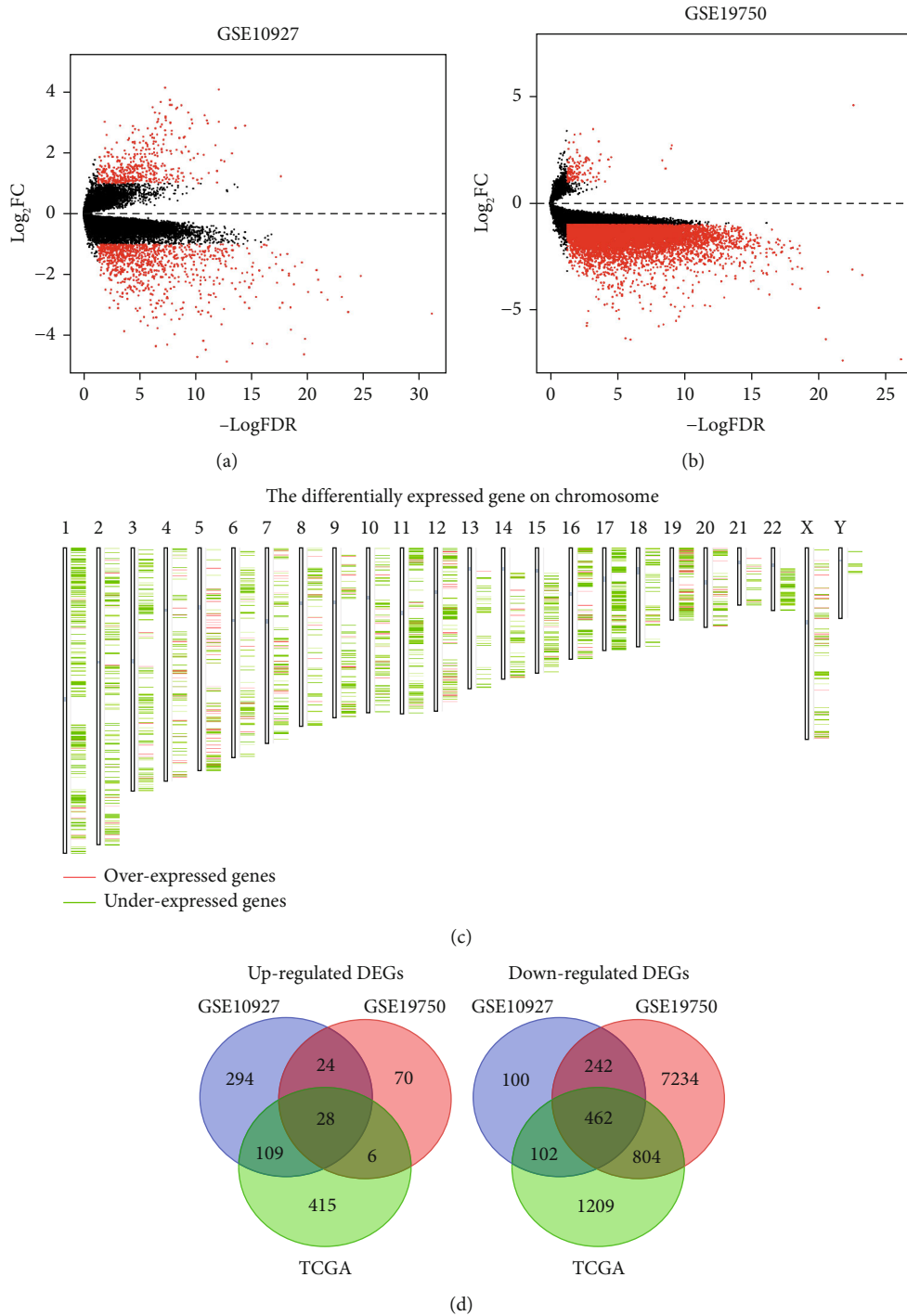


FIGURE 1: The process of identifying DEGs in ACC. (a-b) Volcano maps based on GSE10927 and GSE19750 data sets. (c) Schematic representation of differentially expressed genes on chromosomes. (d) Venn diagram based on DEGs in GSE10927, GSE19750, and TCGA data.

adenocarcinoma and adrenocortical adenoma cannot be accurately performed by only using IGF2 as an indicator [14–16]. In recent years, research on differential genetic screening of adrenal tumors has been on the ascendant. It has been reported that the combination of IGF2 and Ki-67 has high specificity and sensitivity in identifying benign and malignant adrenal cortical tumors [12, 14, 17]. Another

study reported that the most differentially significant genes were TOP2A, IGF2, CCNB2, CDC2, CDC25C, and CDKN1C [18]. The correlation between the differential gene expression fold and the survival time of patients with adrenocortical adenocarcinoma has also been confirmed [19], so it is possible to judge the prognosis of patients according to the gene expression level. In addition, steroidogenic factor

TABLE 1: 490 DEGs were identified from TCGA and GEO data sets, including 28 up-regulated and 285 down-regulated genes in ACC compared with normal tissues.

	Genes name
Up-regulated genes ($n=28$)	GGH TPIX2, CCNB1, PLA2G1B, ANLN, MND1, FOXM1, KIF11, RACGAP1, CENPH, RRM2, TOP2A, ZNF367, CENPU, APOBEC3B, GPX8, MAD2L1, GAS2L3, KIF4A, KIF20A, CENPK, PDE8B, CDC20, NDC80, PBK, NUF2, NCAPG, ESM1
Down-regulated genes ($n=462$)	CLMP, FSTL1, MMP2, RALYL, NR4A2, SERPINF1, SUGCT, SLC02B1, TEK, NEFH, GYPC, LINC00924, EMILIN1, ID1, CELA1, IGSF11, SLC9A3R1, FHL1, IRX3, IFITM10, BTK, SYTL5, USP9Y, AQP11, ZBED6CL, FAM49A, HOXA5, TAC1, HOTAIRM1, EPB41L3, TSTD1, ALASI, DAAM2, SMOC2, MAN1A1, NKAIN1, CSDC2, LRRC32, EMB, AXL, SHE, TCEAL2, IL10, ALOX5AP, FMO3, ABCA6, NBA, DDX3Y, MCOLN3, SLC16A4, MC2R, WISPI, BRE, SRPX, ZNF204P, ADAP2, EIF1AY, LRRN4CL, RARRES1, CLEC5A, MARCO, TIMP4, KCNMB4, C9orf3, AOX1, CYR61, TYMP, GGT5, APOC1, FLYCR2, DLGAP1-ASI, CHRDL1, LAMA2, C1QC, CD55, PLN, RERG, PLTP, MRPL33, PONI, DNASE1L3, RNASE2, ERMP1, SLC47A1, ABCB1, THBD, CHKB, TH, MAP3K8, SPON1, PLA2G4A, ABCC3, EDNRB, EGFLAM, DPYS, ADAMTSL2, C7, S100A8, NPY5R, ITGAM, FOSL2, SKAP1, CCR1, HTR2B, PYGL, HIBCH, COL4A4, SPOCK2, GPR34, CORO1A, EFEMP2, AEBP1, JAM2, RASD1, CYP11B1, GPRASP1, CDKN1C, TXLNGY, IL33, GPX3, NOV, GPM6B, AMT, HSD11B1, KCNJ5, ACOX2, ERN1, PTH1R, PHYHD1, NXPFI, DAPL1, NPC1, PARM1, MS4A14, FBLN5, FIBIN, MUMILL1, IGF1, CERK, SUSD2, CSF3R, SCUBE3, SERPINB9, GATA6, MRAP, SERPING1, PDZRN3, MFAP5, COLEC11, MGST1, STON1, PAX8-AS1, CEBPD, NGFR, NEDD4L, PDGFD, SGK1, KRT8, NFKBIZ, SLC25A34, PLCXD3, RAMP3, TINAGL1, S100A16, TNFSF13, EFEMP1, LUM, C1S, FCGRT, NGEF, PLAT, SRPX2, IGFBBP6, SLC37A2, AKAP12, HSD3B2, APOD, AKR1B1, MAPK13, TNFRSF14, ARFGAP3, CYP17A1, IL4R, OLFML3, FXYD1, FCER1G, C11orf96, RSP03, CCDC159, SREBF1, C2orf40, KCNJ8, CFD, CIQTNF1, AS3MT, PTPNMI, ACTR3C, ANKSI1, SYNPO2, ALPK3, NR2F1, EPHA2, FAM150B, RASGRP2, PTPRB, PNMAL2, ECM1, DNALI1, STEAP4, LILRA2, B4GALT6, TTC39C, STX11, ACO1, SFRP4, FAM166B, DNAJC12, RXFP1, RAPGEF4, EPHX2, CCDC68, DUOX1, ACSM5, PLIN1, SULT1E1, RARRES2, ADAMTSL1, TMEM173, GLUL, RASSF2, AVPR1A, TCIRG1, NPR2, ZEB2, PYY2, FMO2, MEST, SCNN1A, FAM65C, IGFBP4, NANOS1, RETSAT, OLFML1, NCF4, SIRPB2, KCNK3, FGR, HEPH, PHYHIP, C5ARI, APOE, GKN1, CRHBP, THRB, MCOLN2, LONRF2, SORBS2, MTMR6, PACSIN3, OMD, TCF21, SLA, KLF2, ACADVL, SLC16A2, SIGLECI, MGP, ECHDC3, CPA4, GIMAP6, MYC, GPM6A, STARD8, FNDC4, F13A1, GIPC2, OGN, SLC44A3, CXCL2, STAB1, THBS1, AMDHD1, TMEM200C, SYBU, FILIP1L, SCN7A, MCF2, PDGFRA, SLC27A2, CPE, DHRS1, DCN, CYB5A, C1orf162, CYP4B1, COL12A1, HOPX, EIF2D, ARHGEF10L, ZDHHC2, ST6GALNAC5, FCGR2B, MAP3K5, ACRC, CRYAB, PARVA, C8orf4, SLC40A1, CORO2B, ITGA8, IL1RL1, MS4A6A, IFITM2, BHMT2, FRMD6, GBP2, ATP1B2, LINC01314, USP53, G0S2, C10orf10, SHC3, CTGF, CD163, DPT, PTGDS, IGSF10, NKD2, CXCL12, ALDH1A1, CARTPT, PPAP2B, C1QB, CBLN4, BRINP2, IFI35, TLR4, ZNF185, C9orf24, TMOD1, LPARI, ADAMTSL3, CRISPLD2, SELENBP1, FOSL1, CNTN6, S100A9, KLHL2, LRFN5, GLT8D2, CNN1, SIGLEC9, ALDH3A2, PLEKH01, SLC02A1, MEIS2, PRPS2, TLE2, ACSF2, HCK, CSRP1, MAP7, DGATI, NPY1R, TPD52L1, SHISA8, GPR182, MRC1, IGFBBP5, PTGER4, KCNQ1, ANGP1L1, IGSF21, CD14, TRIP6, KLHDC8A, EMCN, SLC27A6, ISLR, DKK3, MS4A4A, MYLK, ACSBG1, PIDI, ADORA3, RAL2, GCKR, FBPI, ST3GAL4-AS1, VASN, ALDH1A3, DOK2, SELM, BOC, TMEM61, PRELP, WFDC1, CYBRD1, PLEKHA6, HGF, CYP11B1, VAMP8, C1R, SQRDL, TRIM22, CD33, NRIH3, AADAC, CACHD1, GSTA4, ABCA1, C3AR1, CFH, CHGA, SLC1A5, MTIM, TNNC1, DUSP26, FBLN1, SLC16A9, CD248, LMOD1, ZRANB1, LAT2, VSIG4, THRSF, FMO1, ARNTL, CCL2, FAM179A, RBKS, TAGLN, KCNK2, MOXD1, MFAP4, DLG2, ARHGAP9, PLEK2, RBP4, S100A4, PROK1, ACKR1, CREG1, FND5, PCDH10, NEXN, GATA5, B1CC1, INMT, ITM2A, MPDZ, TMEM220, ADH1B, CAB39L, FSTL3, FCN3, GATA6-AS1, GAREM, KDM5D, VIPR1, GRAMD3, HCLS1

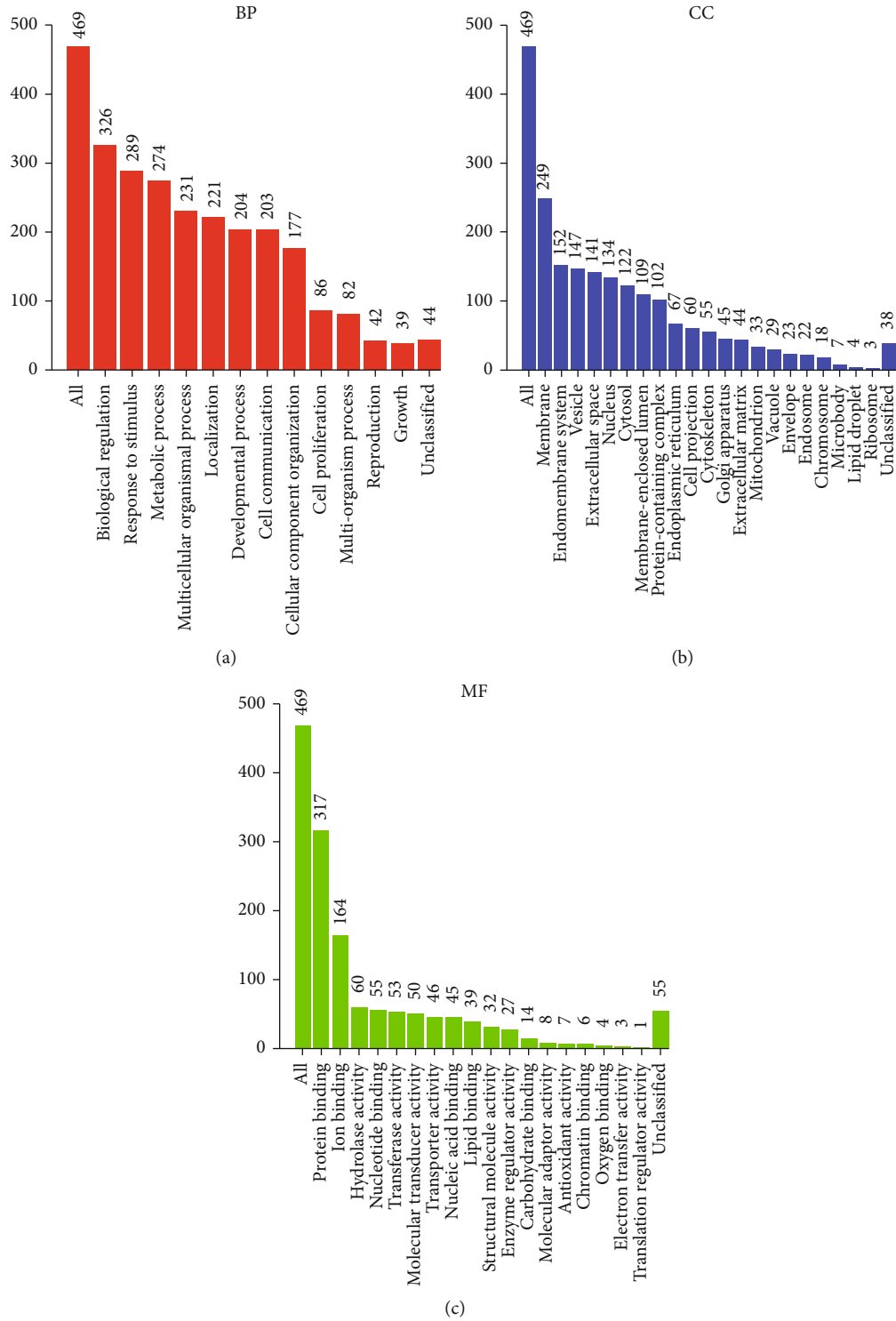


FIGURE 2: Continued.

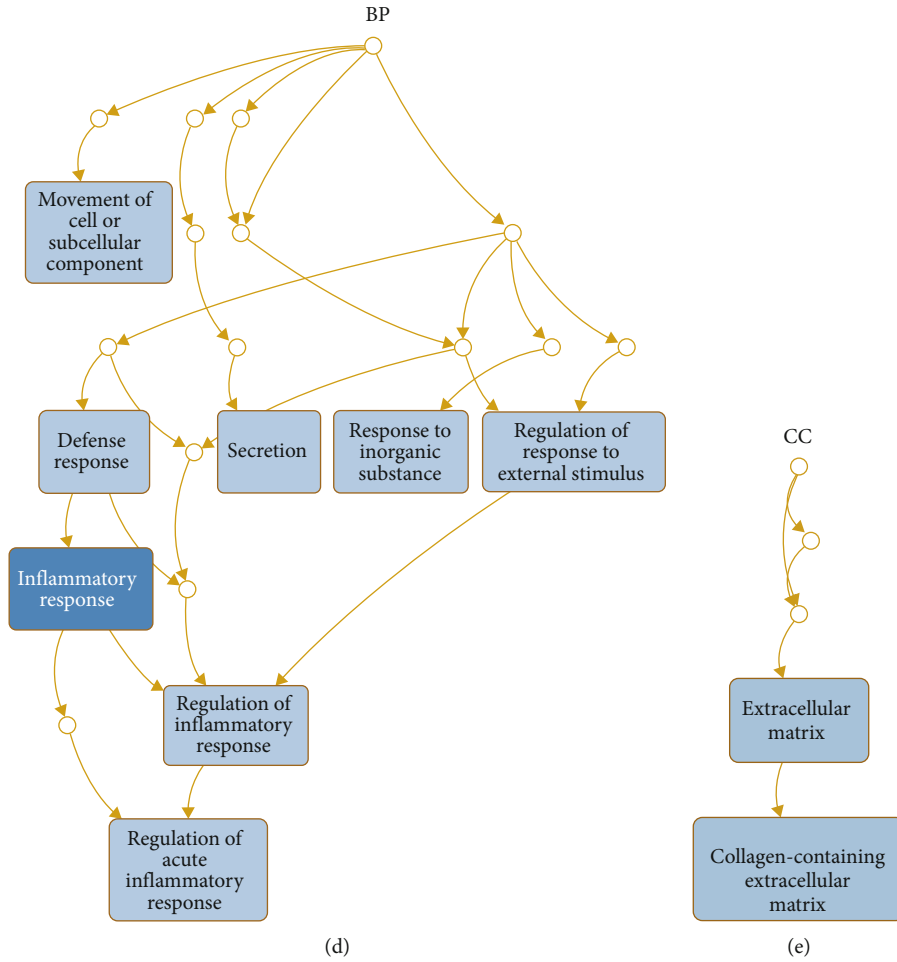


FIGURE 2: GO analysis was performed for DEGs in ACC. (a–c) Histograms show the results of GO analysis. (d–e) Hierarchical plots show the results of the GO analysis.

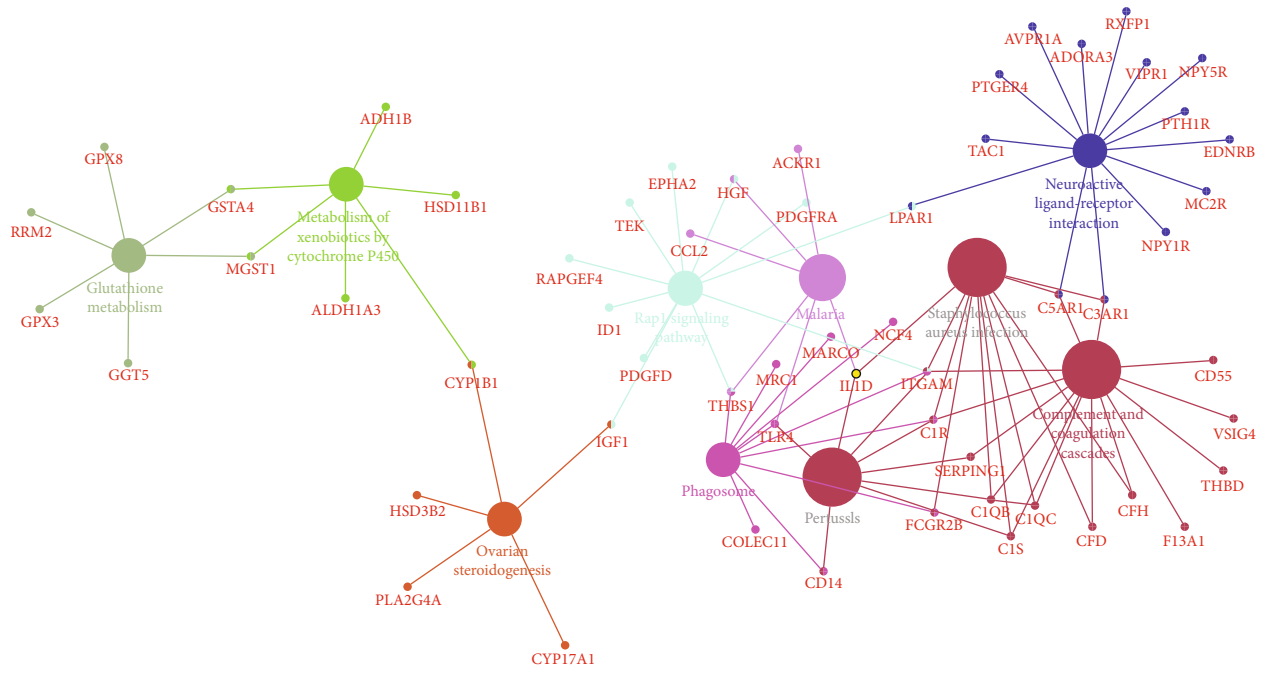
1 (SF-1), another gene that plays an essential role in promoting the occurrence and development of adrenal tumors, is of great significance to the growth and migration of adrenal tumor cells. In vivo experiments have proved that overexpression of SF-1 promotes the proliferation and migration of adrenocortical adenocarcinoma cells [20]. In addition, multiple studies have also confirmed that SF-1 has a high value in the diagnosis of adrenocortical carcinoma and the prognosis evaluation of patients [21–23], and it has been reported that SF-1 overexpression is associated with a low survival rate in patients with adrenocortical carcinoma. In addition, Snail is closely related to the metastasis and prognosis of adrenocortical carcinoma. The relevant research results show that more than 95% of the clinical stage III and IV adrenocortical carcinoma tumors have positive Snail expression [24]; ER-negative expression adrenal cortical carcinoma patients have a lower 5-year survival rate than those with ER-positive expression and have a greater chance of distant metastasis [25, 26]. In addition, the simultaneous high expression of BUB1B and PINK1 in tumor tissue may indicate a good prognosis in patients [27]. Therefore, the study of these differential gene expression profiles through

bioinformatics analysis plays a crucial role in understanding the pathogenesis of adrenocortical adenocarcinoma and the molecular signaling pathways involved [28].

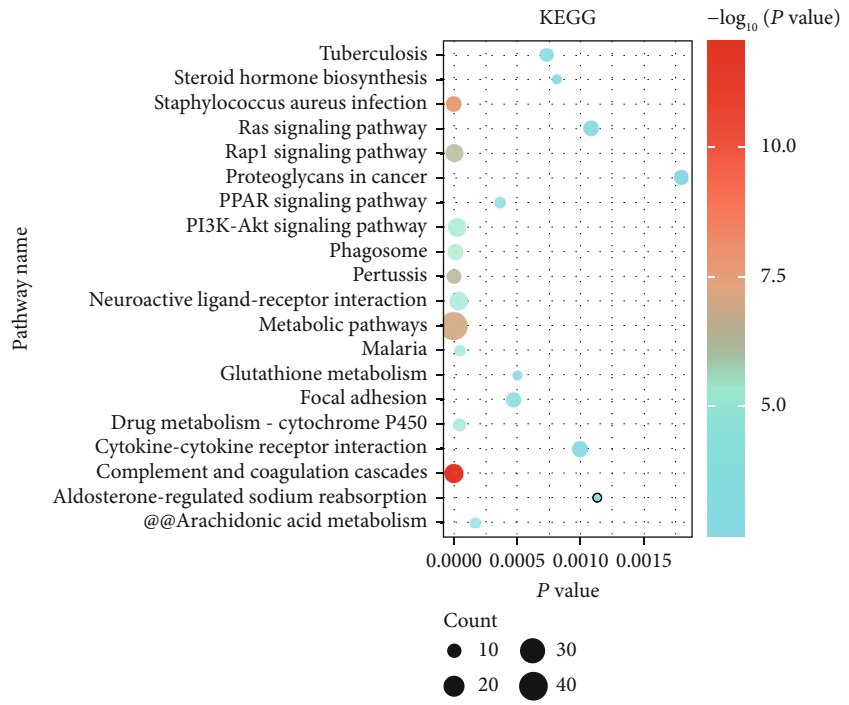
We first downloaded raw data from GEO and TCGA databases in this study to obtain differentially expressed genes (DEGs) in ACC. Then, we performed gene ontology, pathway enrichment analysis, and protein-protein interaction (PPI) network. GEPIA was adopted to observe these genes’ mutations, differential expression, and prognostic characteristics. Besides, TCGA and cBioPortal were used to determine the distribution in pan cancers, pathway enrichment, the features in pathological parameters, and the relationship with other genes. We attempted to seek specific hub genes that may serve as influential biomarkers for ACC.

2. Materials and Methods

2.1. GEO Database. GEO is a gene expression database created and maintained by the National Center for Biotechnology Information NCBI. The database was built in 2000 and contains high-throughput gene expression data from research institutions worldwide. In this study, GEO database

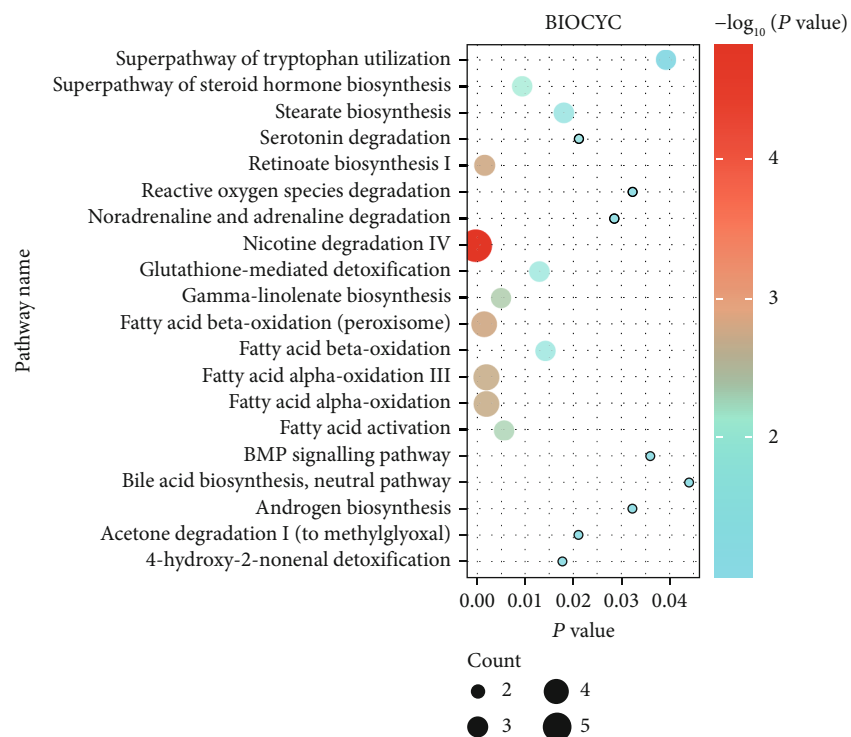


(a)

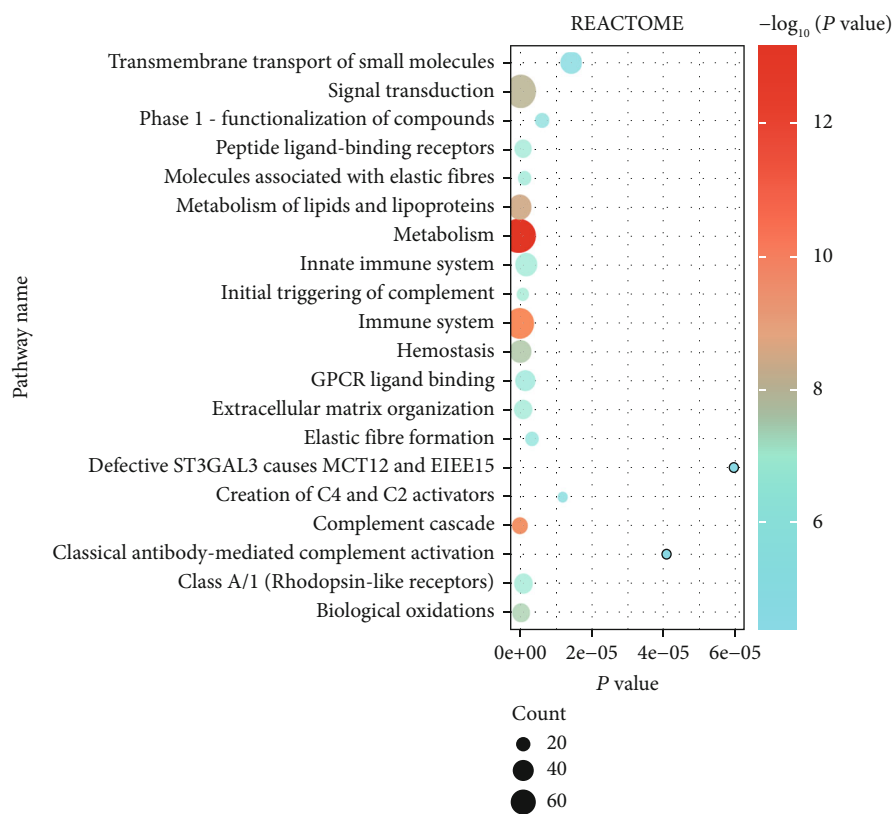


(b)

FIGURE 3: Continued.



(c)



(d)

FIGURE 3: Continued.

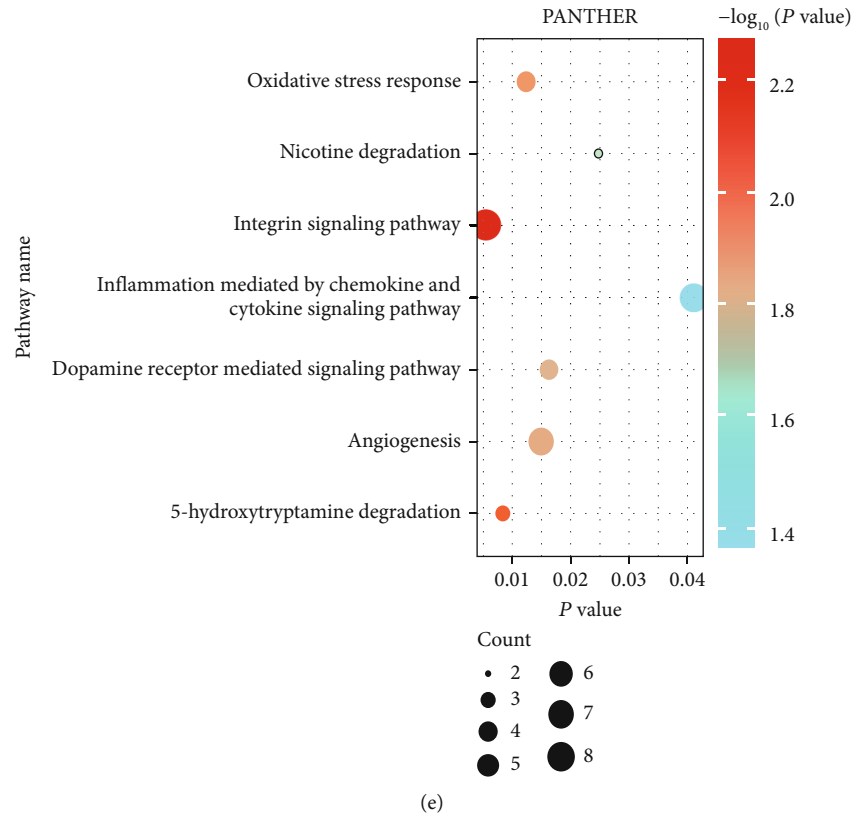


FIGURE 3: Pathway enrichment analysis was performed for DEGs in ACC. (a) Interaction plot showing the results of pathway enrichment analysis. (b–e) Bubble plots show KEGG, BIOCYC, REACTOME, and PANTHER pathway enrichment analysis results.

(<http://www.ncbi.nlm.nih.gov/geo/>) [29] was used for gene expression data sets between ACC tissues and normal tissues. Then, we further evaluated the complete information about the relevant data sets. Finally, in line with the Affymetrix Human Genome (GPL570) platform, two data sets (GSE19750 and GSE10927) were chosen for subsequent analysis. The GSE19507 data set contained 44 ACC and 4 normal samples [30, 31], and the GSE10927 data set included 33 ACC and 10 normal samples [32].

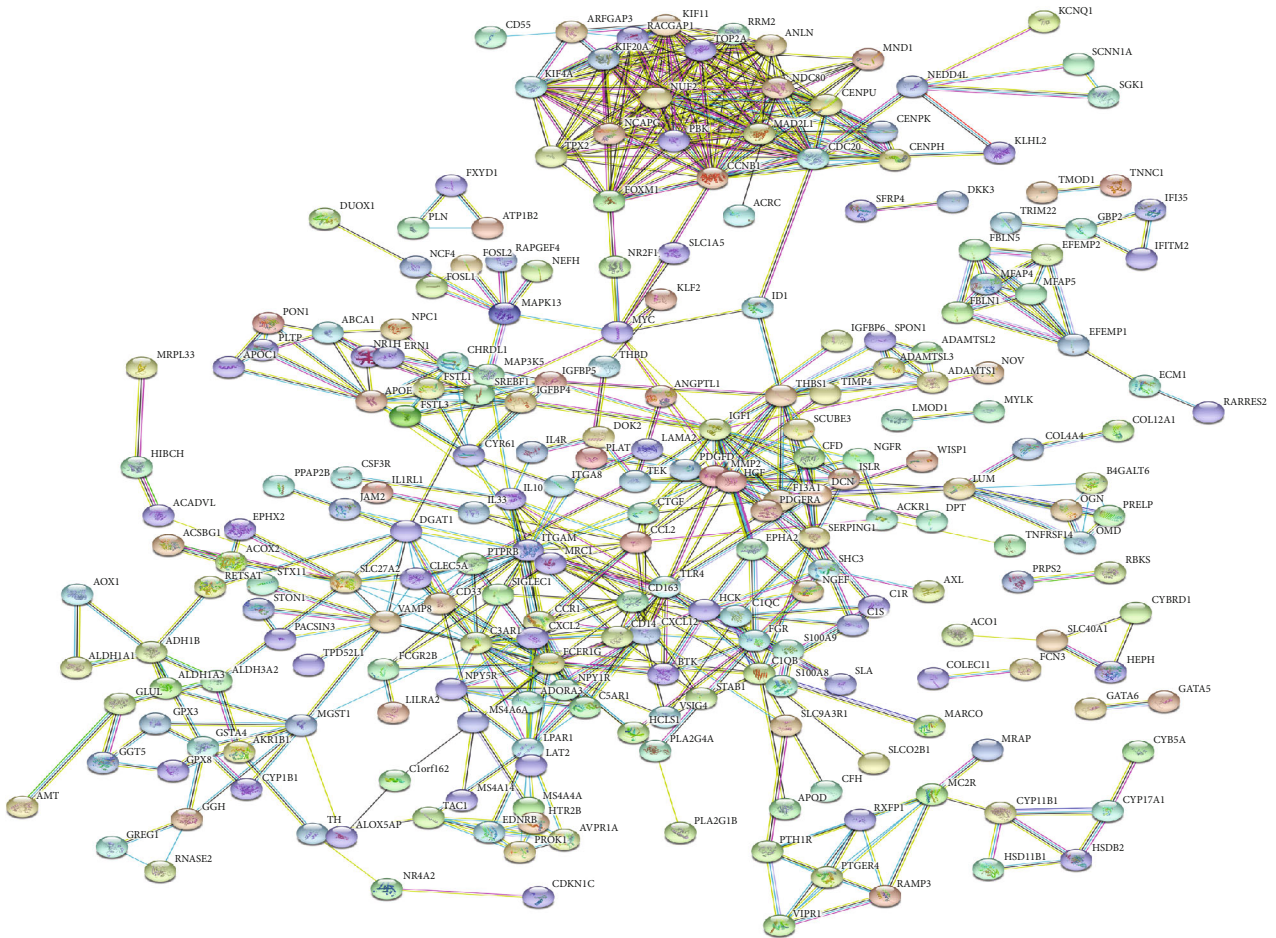
2.2. Differential Expression Analysis. R language was used to analyze GEO data and drew volcano maps and heat maps, and these two data sets were employed to get differential expressed genes (DEGs). $|\text{Log}_2\text{FC}| > 1$, P -value < 0.05 was considered the cutoff criterion. Besides, we put on these data to cross with TCGA data [33]. Then, an online tool, Bioinformatics & Evolutionary Genomics, was used to draw the Venn diagram for up-regulated and down-regulated DEGs (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) [34].

2.3. Gene Ontology and Pathway Enrichment Analysis. The up-regulated and down-regulated DEGs were integrated into the WEB-based Gene Set Analysis Toolkit (webgestalt) (<http://www.webgestalt.org/>) [35] for Gene Ontology (GO) functional annotation enrichment analysis. Furthermore, we performed KEGG pathway analysis for DEGs through the ClueGO plugin in Cytoscape software [36]. The KEGG [37], REACTOME [38], PANTHER [39], and BIOCYC

[40] pathways were downloaded from the KOBAS website [41]. A P -value of < 0.05 was considered statistically significant.

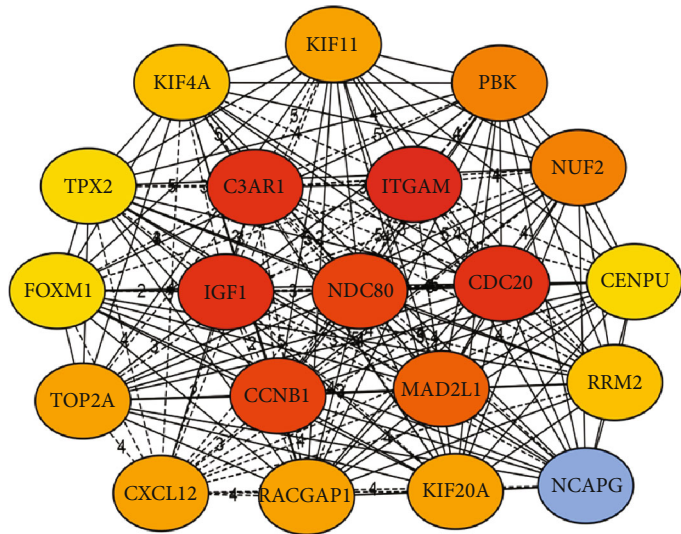
2.4. Protein-Protein Interaction (PPI) Network and Identification of Hub Genes. String database is a database that can be used to search for interactions between known and predicted proteins. In addition to generating beautiful protein-protein-interaction (PPI) maps of these proteins, an analysis of imported proteins is also provided. In this study, PPI network between DEGs was built by String database (<http://stringdb.org/>) [42]. First, entered the DEGs into the database and set the confidence score ≥ 0.7 . Then, removed unlinked DEGs and arranged the remaining DEGs protein interaction data and photos. The data acquired by String website was substituted into the Cytoscape software and the hub genes were captured through the cytoHubba plugin. Afterward, the top 20 genes were collected by three algorithms of Degree, MCC, and MNC [43]. The Venn diagram of these hub genes was gathered using the online tool Bioinformatics & Evolutionary Genomics.

2.5. Gene Expression Analysis and Survival Analysis. GEPIA (<http://gepia.cancerpku.cn/detail.php>) [44] is a newly developed interactive web server for analyzing RNA sequencing expression data of 9736 tumors and 8587 normal samples in TCGA and GTEX projects. Based on GEPIA database, we checked the differences in hub gene expression between



(a)

Top20 hub genes (degree)



(b)

FIGURE 4: Continued.

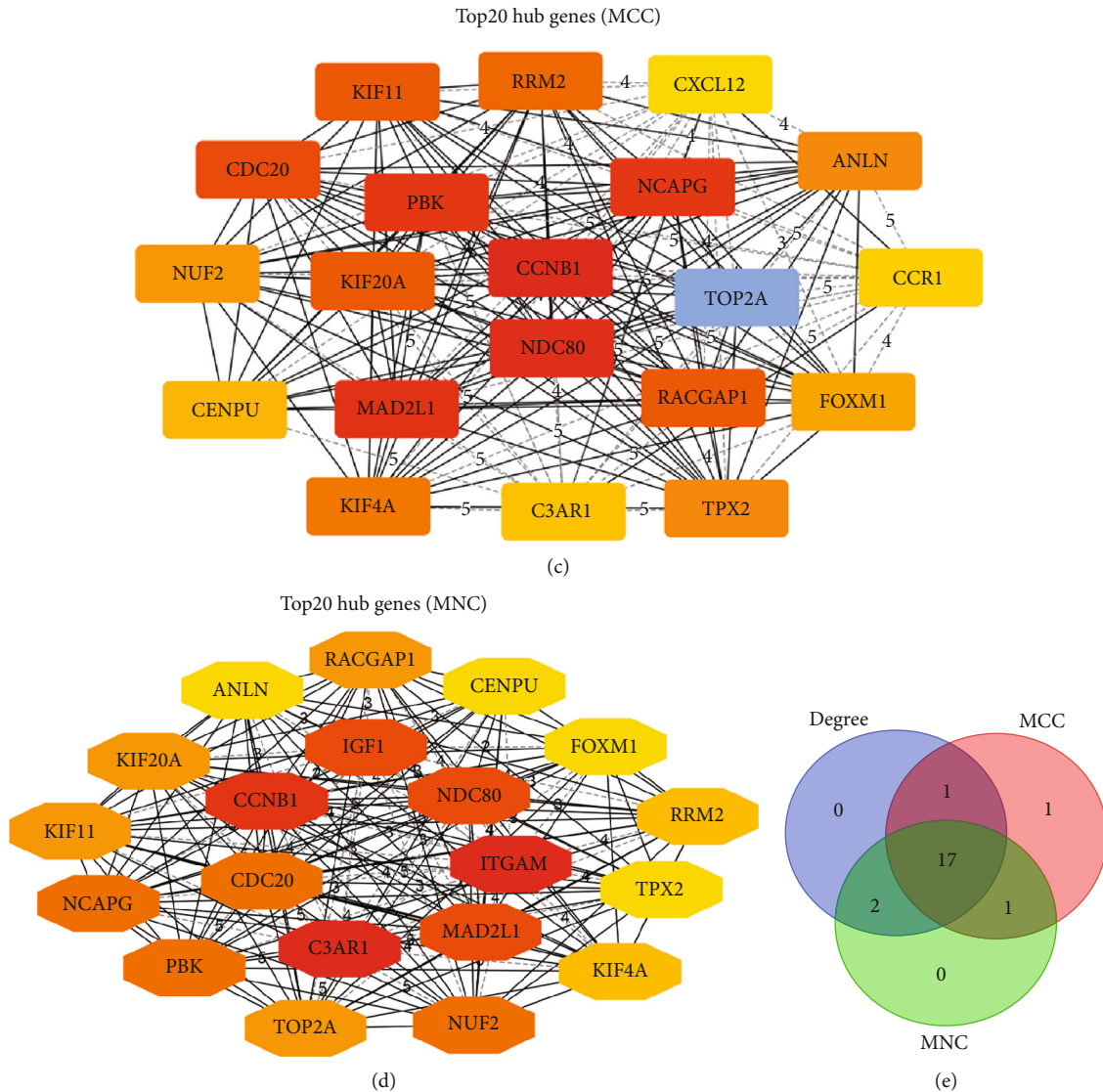


FIGURE 4: Protein-protein interaction analysis and screening of hub genes of DEGs. (a) The protein-protein interaction network of these DEGs molecules. (b) Top 20 hub genes screened by Degree algorithm. (c) Top 20 hub genes screened by MCC algorithm. (d) Top 20 hub genes screened by MNC algorithm. (e) A Venn diagram is drawn based on the hub genes obtained by Degree, MCC, and MNC algorithms.

ACC and normal tissues. The predictive value of these genes in ACC was analyzed using the GEPIA database, and the cutoff value was set to 50%. The website automatically calculated the hazard ratio (HR) of 95% confidence interval and log-rank P -value and displayed it directly on the web page. P -value < 0.05 was considered statistically significant.

2.6. TCGA and cBioPortal Data. The cancer genome map included sequencing and pathology data for 30 different cancers. The ACC (TCGA, Provisional) data set was selected, comprising data from 92 pathology reports. These DEGs were further conducted via cBioportal (<http://www.cbioportal.org/index.do>) [45]. The genomic analysis is covered with mutations and co-expression analysis. The co-expression and networking were calculated based on

cBioportal's online instructions. P -value < 0.05 was considered statistically significant.

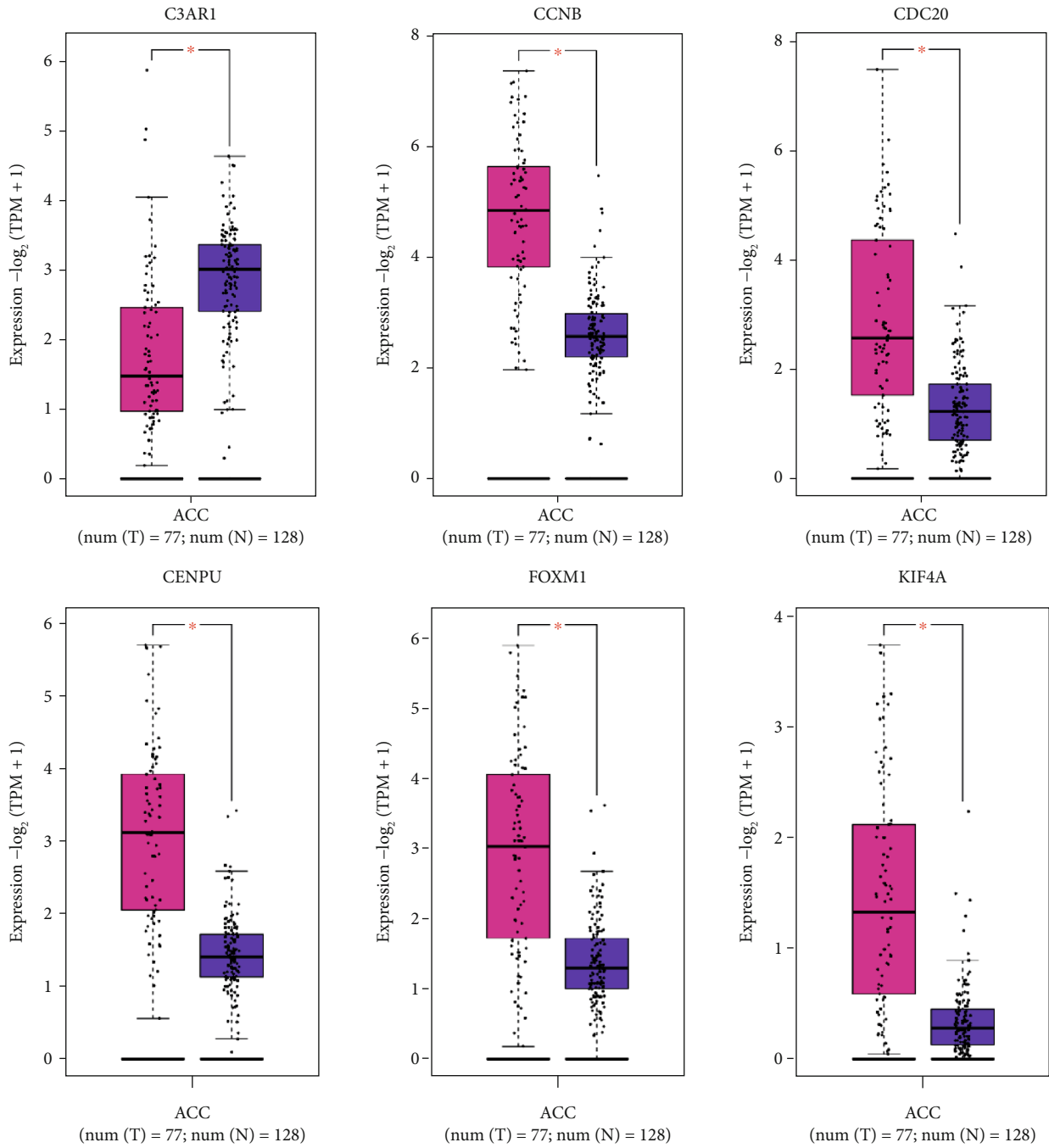
2.7. Statistical Analysis. Statistical analyses of all data were performed using statistical software from all online databases. Statistical significance of differences between and among groups was assessed using the t -test. Statistical significance was set at $*P < 0.05$; $**P < 0.01$; and $***P < 0.001$.

3. Results

3.1. DEGs in ACC. In recent decades, differentially expressed genes have been the focus of research in the field of cancer research. DEGs in ACC were identified by examining two GEO data sets and TCGA data (Figures 1(a) and 1(b)). 490 DEGs consisting of 28 up-regulated genes and 462 down-



FIGURE 5: Continued.



(c)

FIGURE 5: Continued.

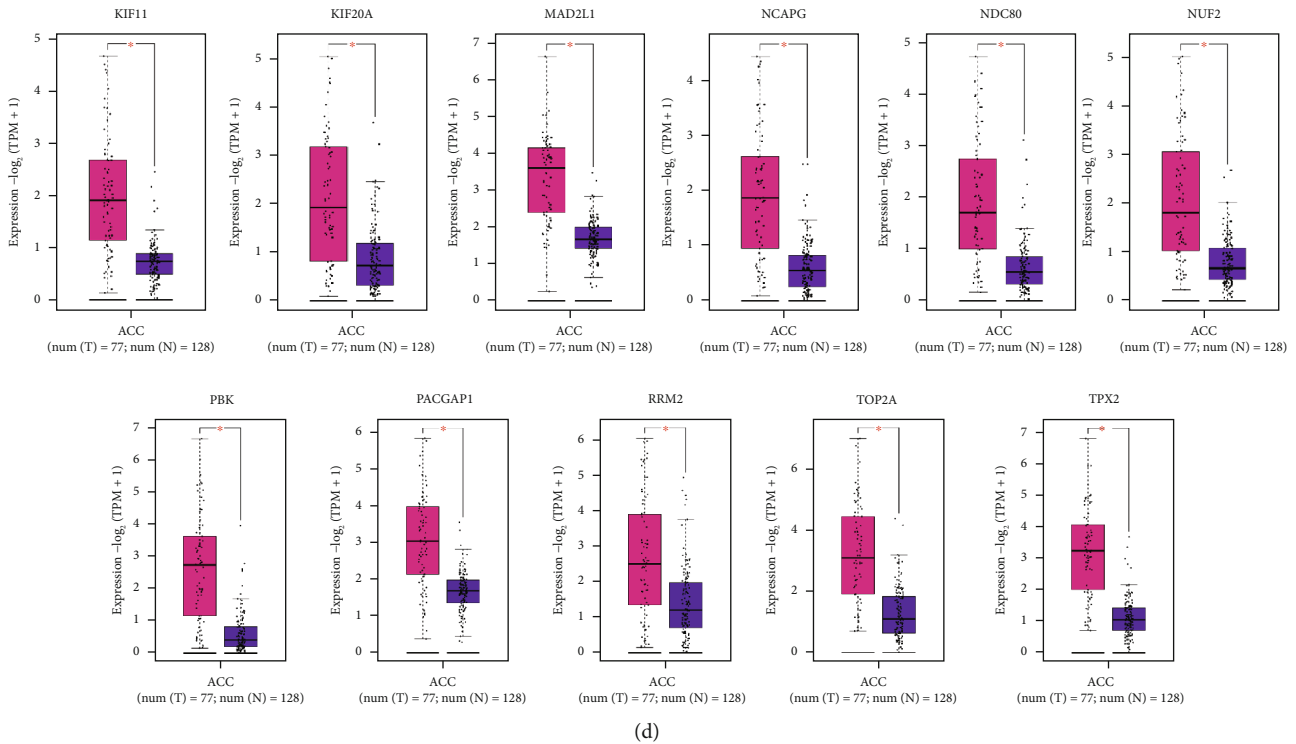


FIGURE 5: Overall variation and mRNA expression of 17 hub genes in urological tumors. (a) Overall variation of 17 hub genes in ACC. (b) Expression of 17 central genes in urologic tumors. (c-d) mRNA differential expression of 17 hub genes in ACC.

regulated genes were finally obtained in our work (Figure 1(d), Table 1). In addition, to show the distribution of these DEGs on human chromosomes more specifically, we draw the corresponding heatmaps. The results showed that over-expressed genes were mainly distributed on chromosomes 5, 7, and 12 (Figure 1(c)).

3.2. Functional Enrichment of DEGs. GO functional enrichment analysis was performed on these DEGs, demonstrating that biological regulation, membrane, and protein binding of most genes were enriched in terms of BP, CC, and MF, respectively (Figures 2(a)–2(e)). Four pathway databases with KEGG, BIOCYE, REACTOME, and PANTHER revealed that ACC-related DEGs mainly concentrated on complement and coagulation cascades, metabolic pathways, malaria, ovarian steroidogenesis, and so on (Figures 3(a)–3(e)).

3.3. Identification of ACC-Associated Hub Gene. String database was applied to analyze the protein interactions of DEGs and make a PPI network (Figure 4(a)). The top 20 ACC-related hub genes were screened through three algorithms involving Degree, MCC, and MNC. After taking the intersection of these three data sets, 17 hub genes containing C3AR1, CCNB1, CDC20, CENPU, FOXM1, KIF4A, KIF11, KIF20A, MAD2L1, NCAPG, NDC80, NUF2, PBK, RACGAP1, RRM2, TOP2A, and TPX2 were collected for further study (Figures 4(b)–4(e)).

3.4. Hub Gene Expression and Prognosis in ACC. To better make out the 17 hub genes, we analyzed the mutations of

17 hub genes. The results showed that CENPU, FOXM1, and PBK had higher mutation rates accounting for 13%, 12%, and 11%, respectively (Figure 5(a)). Subsequently, we detected the expression of these hub genes in six tumors, including ACC, KICH, KIRC, KIRP, PAAD, and BLCA. CCNB1, MAD2L1, ACGAP1, and CENPU were significantly higher expressed in all six tumors (Figure 5(b)). Another discovery is that the expression analysis of these genes in ACC manifested that except for C3AR1, which was down-regulated in ACC, the other 16 genes were up-regulated in ACC (Figures 5(c) and 5(d)). In addition, we also found no significant correlation between C3AR1 and the prognosis of patients with ACC. Still, the rest of the hub genes had a great connection with an unfavorable prognosis (Figures 6(a)–6(q) and 7(a)–7(q)).

3.5. Functional Enrichment of Hub Genes. In cancer research, gene function enrichment analysis has become a routine method for high-throughput omics data analysis, which is of great significance for revealing biomedical molecular mechanisms. To better understand these hub genes' function, pathway enrichment analysis was performed on these 16 hub genes again, which suggested that hub genes were mainly associated with classical tumor-associated pathways, such as the P53 signaling pathway, and cell cycle-related signaling pathways (Figures 8(a)–8(d)).

3.6. Identification of Two ACC Core Genes CCNB1 and NDC80. By duplicating protein interaction analysis on these

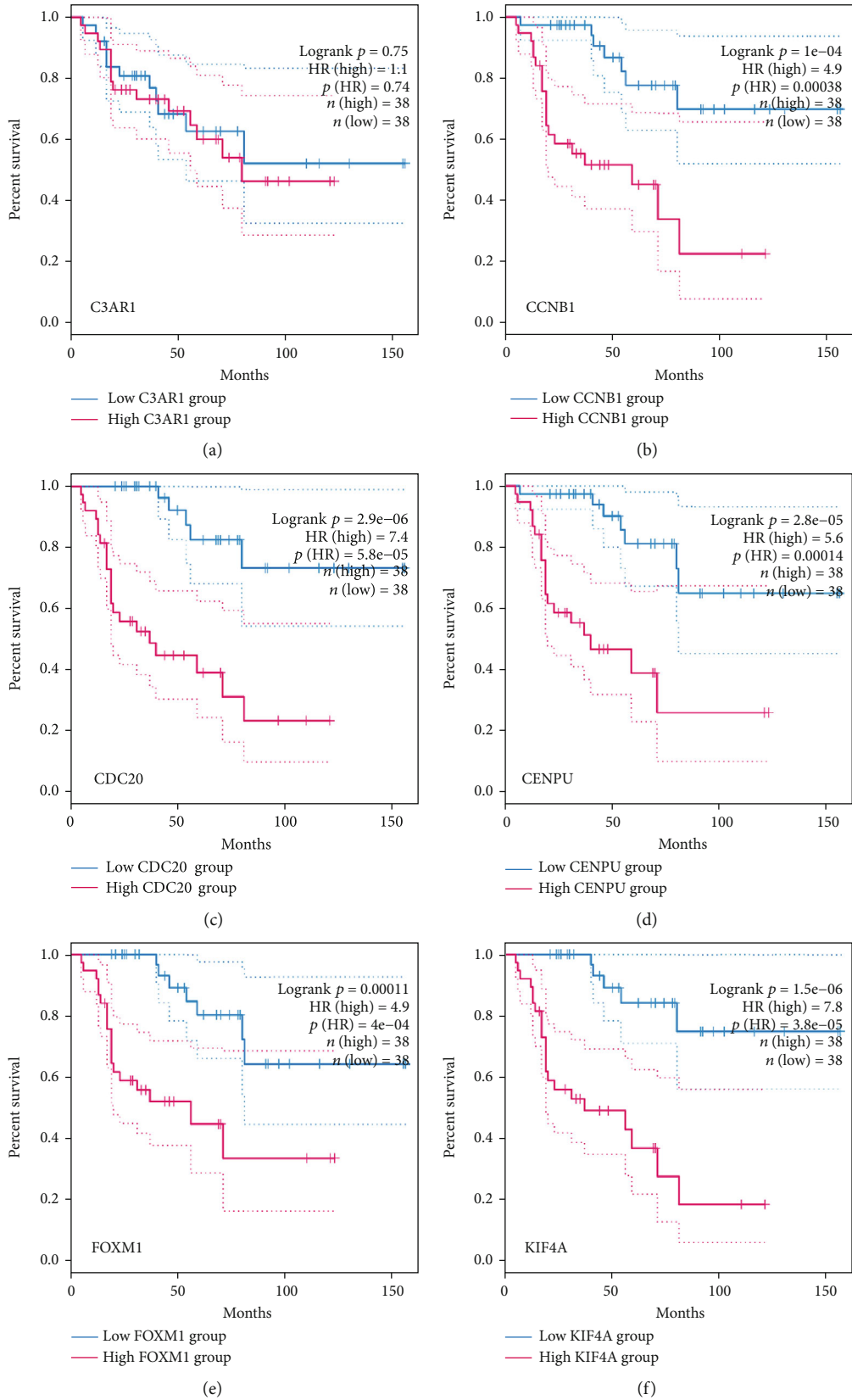


FIGURE 6: Continued.

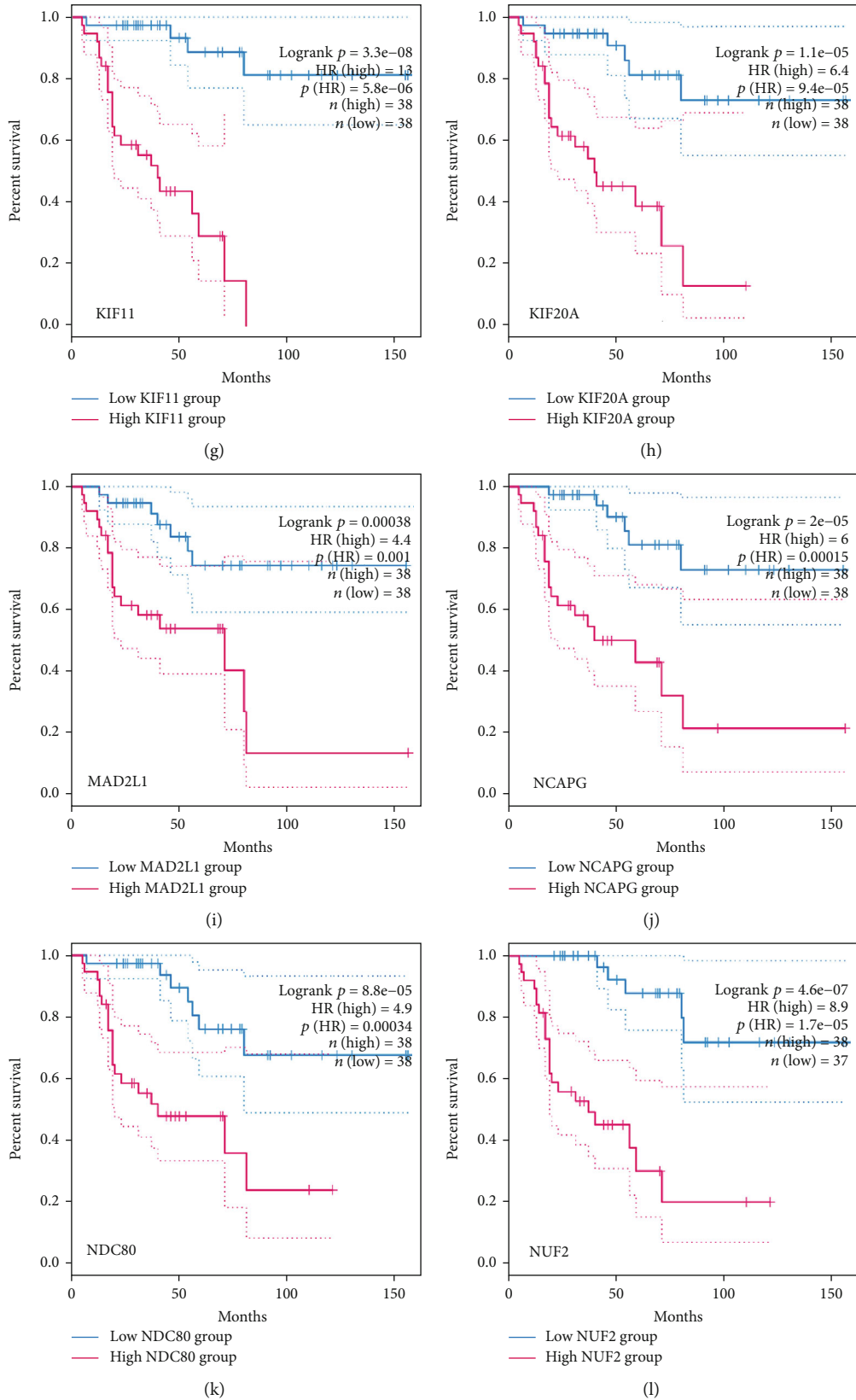


FIGURE 6: Continued.

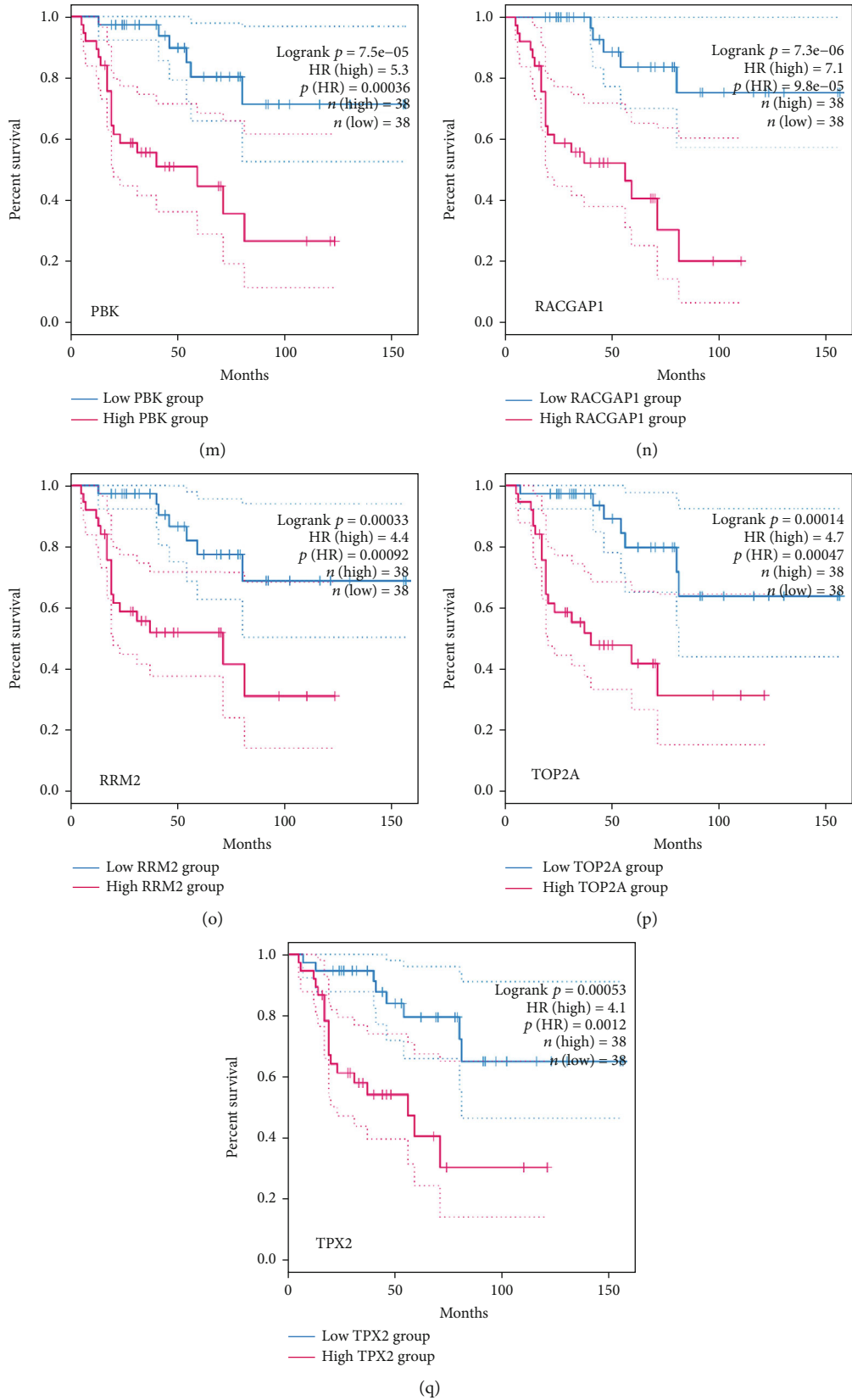


FIGURE 6: Overall survival analysis. (a–q) Survival graphs showing the overall survival of these 17 hub genes in ACC, in order of C3AR1, CCNB1, CDC20, CENPU, FOXM1, KIF4A, KIF11, KIF20A, MAD2L1, NCAPG, NDC80, NUF2, PBK, RACGAP1, RRM2, TOP2A, and TPX2.

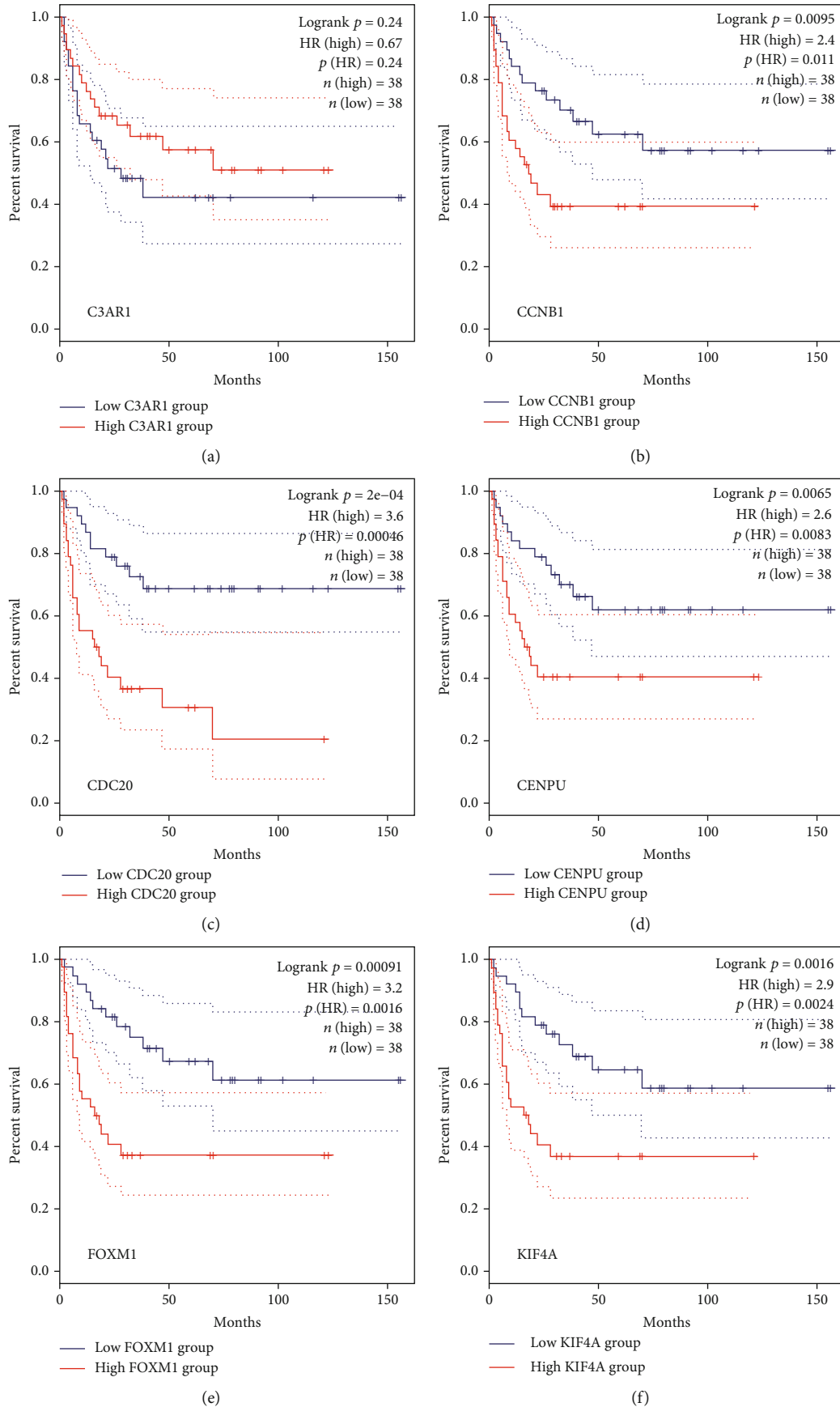


FIGURE 7: Continued.

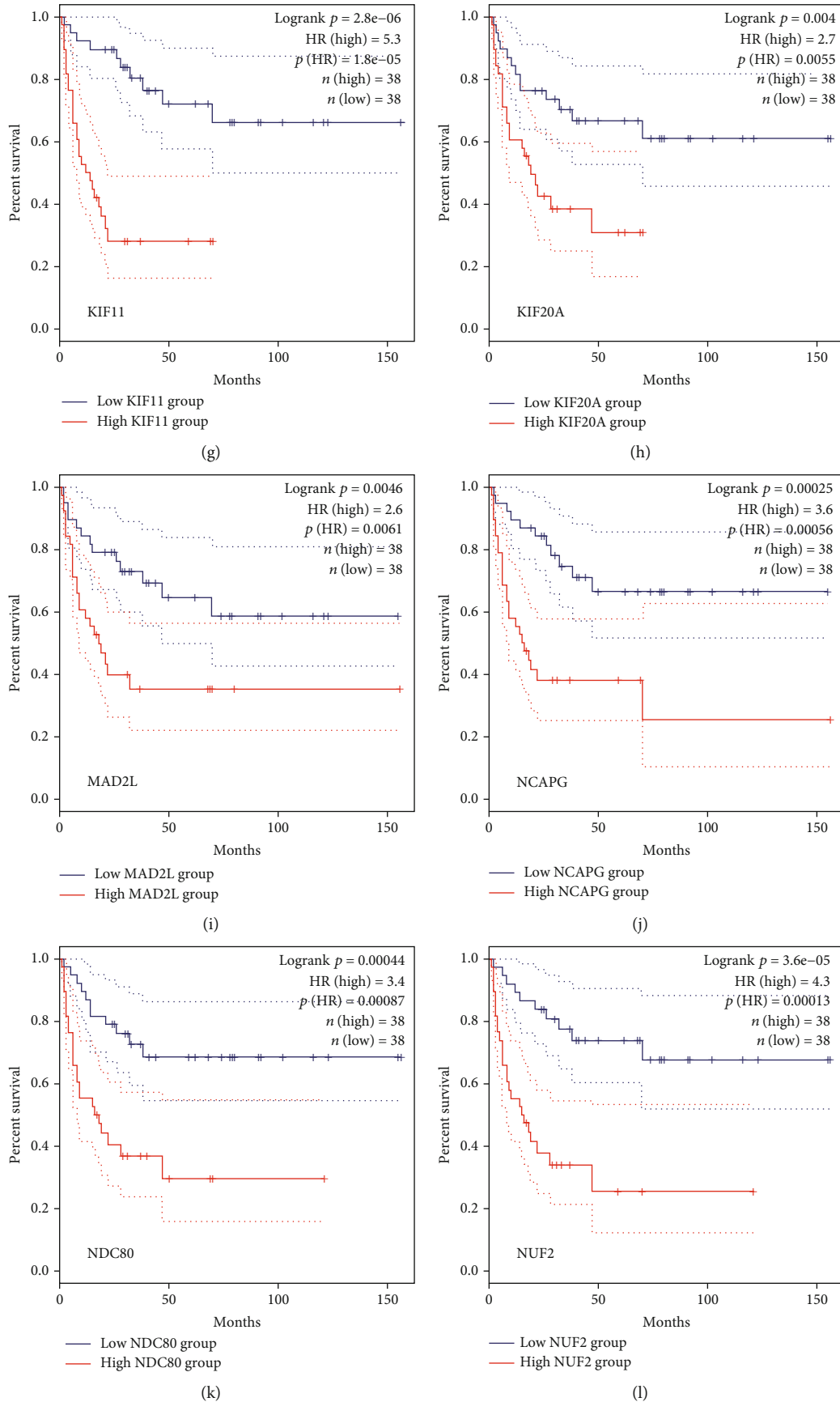


FIGURE 7: Continued.

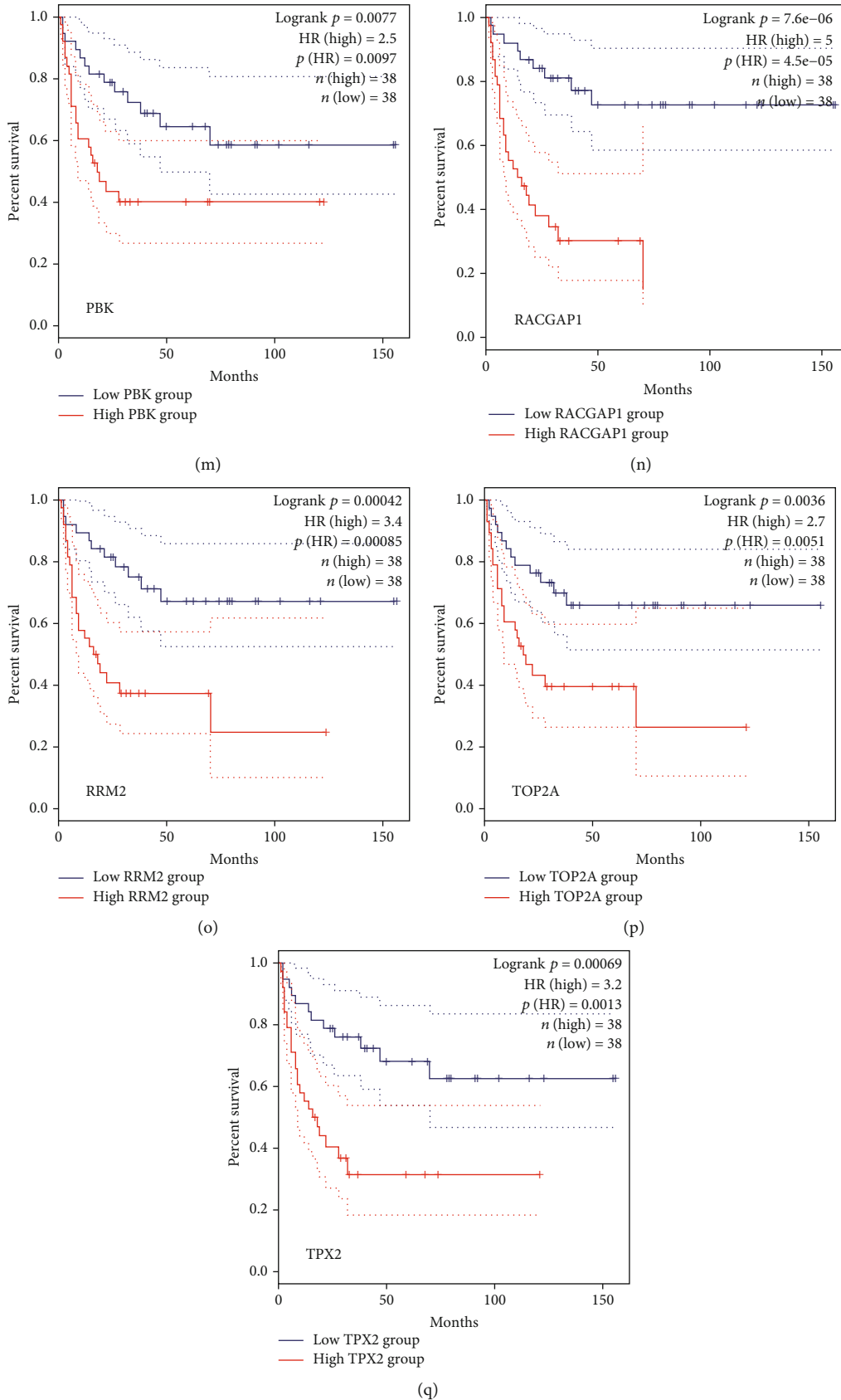


FIGURE 7: Disease-free survival analysis. (a–q) Survival graphs show the disease-free survival of these 17 hub genes in ACC, followed by C3AR1, CCNB1, CDC20, CENPU, FOXM1, KIF4A, KIF11, KIF20A, MAD2L1, NCAPG, NDC80, NUF2, PBK, RACGAP1, RRM2, TOP2A, and TPX2.

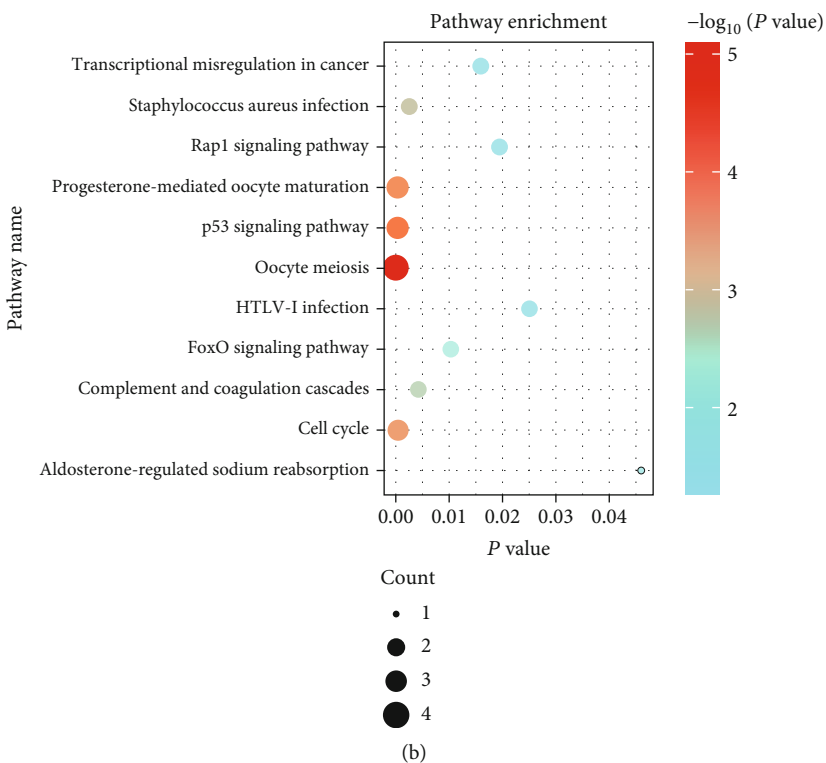
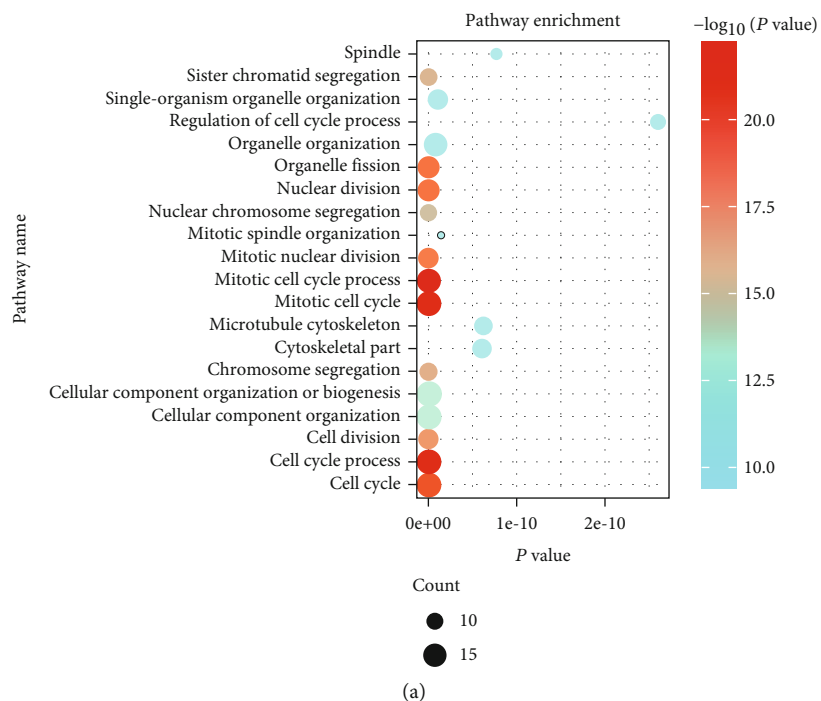


FIGURE 8: Continued.

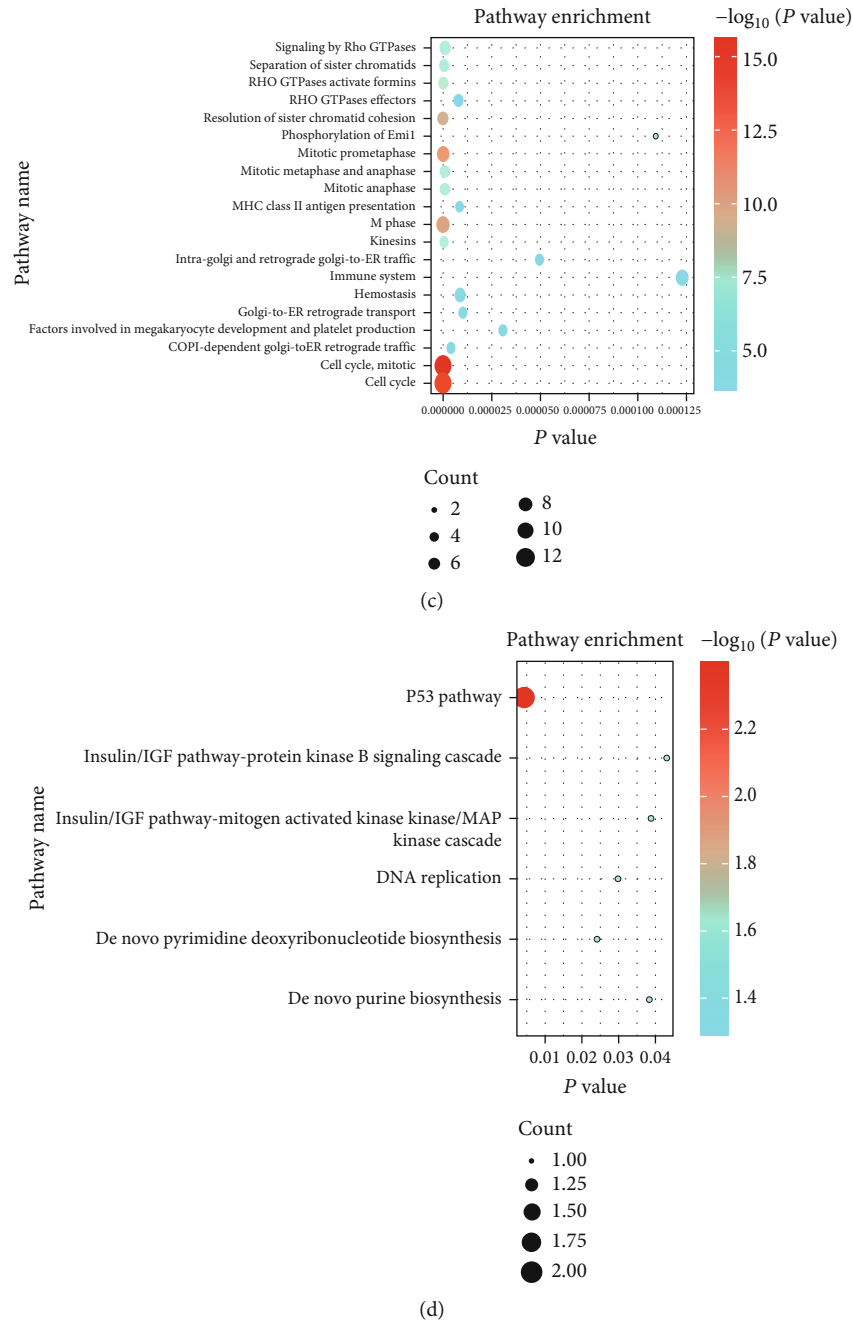
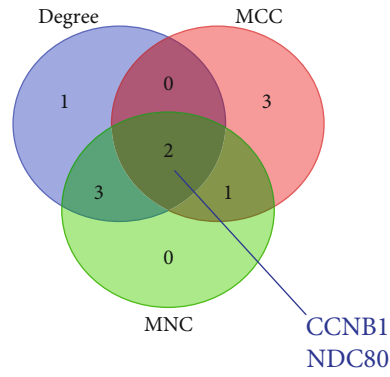


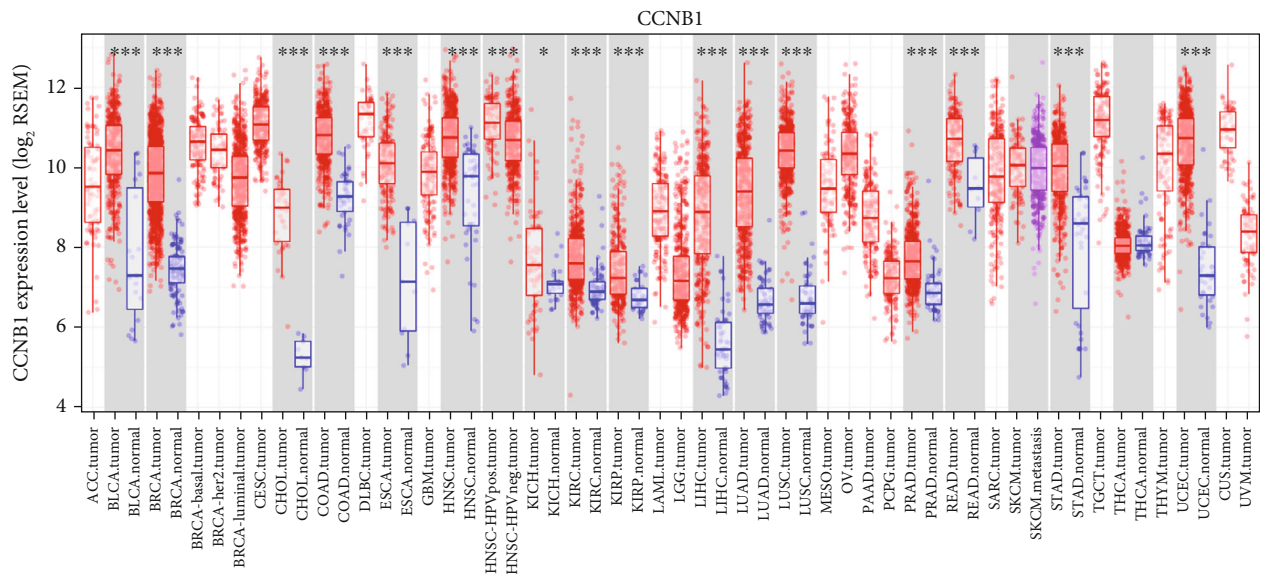
FIGURE 8: After removing the C3AR1 gene with no prognostic significance in ACC, pathway enrichment analysis was performed in ACC for the remaining 16 hub genes. (a) KEGG pathway. (b) BIOCYC pathway. (c) REACTOME pathway. (d) PANTHER pathway.

16 hub genes and narrowing the core gene range, we derived two core genes, CCNB1 and NDC80 (Figure 9(a)). Then, we evaluated the expression of these two genes in pan cancers, and the consequences proved that these two genes were highly expressed in various tumors (Figures 9(b) and 10(a)). Further analysis suggested that the expression of CCNB1 and NDC80 would increase with disease progression. The high expression could also predict adverse outcomes in ACC patients but has little to do with gender (Figures 9(c) and 9(d) and 10(b) and 10(c)). To improve our knowledge about the functions of the core genes CCNB1

and NDC80, ten related proteins were retrieved by the String database (Figures 9(e) and 10(d)). Later, we discovered that CCNB1 and NDC80 participate in the same pathway, incorporated with cell cycle, progesterone-mediated oocyte maturation, HTLV-1 infection, and oocyte meiosis (Figures 11(a) and 11(b)). CCNB1 co-expressed with its related proteins CDK1, CDK2, CCNB2, PLK1, CDC20, CDCA8, ESPL1, and FZR1 (Figures 11(c)–11(j)) in ACC patients. Pathway analysis for NDC80 showed that NDC80 was associated with Cell Cycle (Figure 12(a)). It was worth mentioning that CCNB1 and NDC80 were

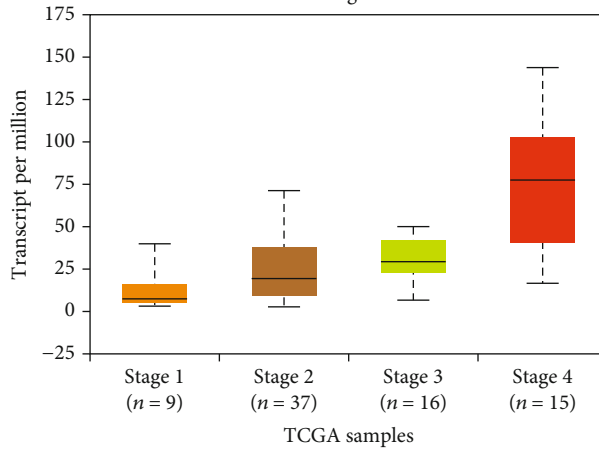


(a)



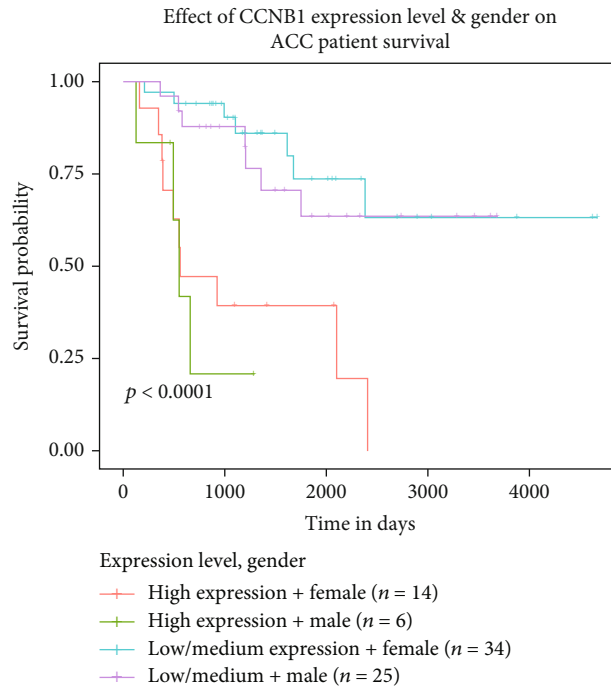
(b)

Expression of CCNB1 in ACC based on individual cancer stages



(c)

FIGURE 9: Continued.



(d)

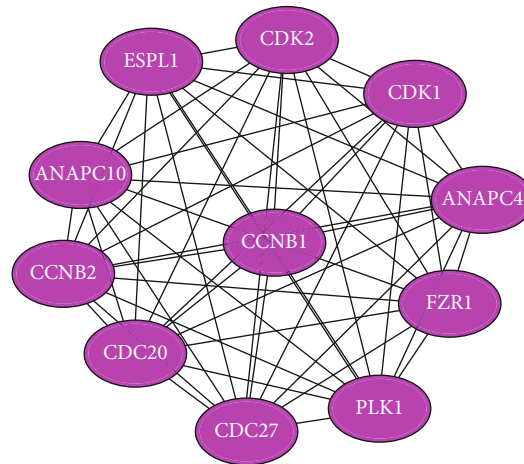


FIGURE 9: In-depth exploration of the biological value of the core gene CCNB1. (a) Venn diagram showing the identification of the core genes CCNB1 and NDC80. (b) mRNA expression of CCNB1 in pan-cancer. (c) mRNA expression of CCNB1 in different stages of ACC. The P -value between stage 1 and stage 4 is $2.2252E-04$. (d) The effect of CCNB1 mRNA expression level and patient gender on the overall survival of ACC patients. (e) PPI map between CCNB1 and the ten most closely related CCNB1 protein molecules.

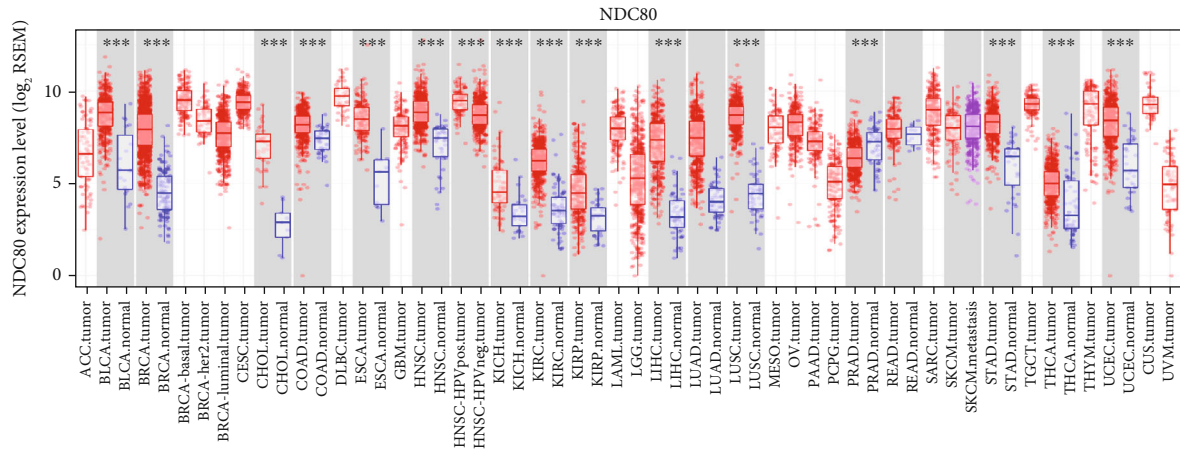
consistently expressed in ACC (Figure 12(b)). Simultaneously, the expression of NDC80 also has collinearity with several proteins, like AURKB, BUB1, SPC25, and CENPE (Figures 12(c)–12(f)).

4. Discussion

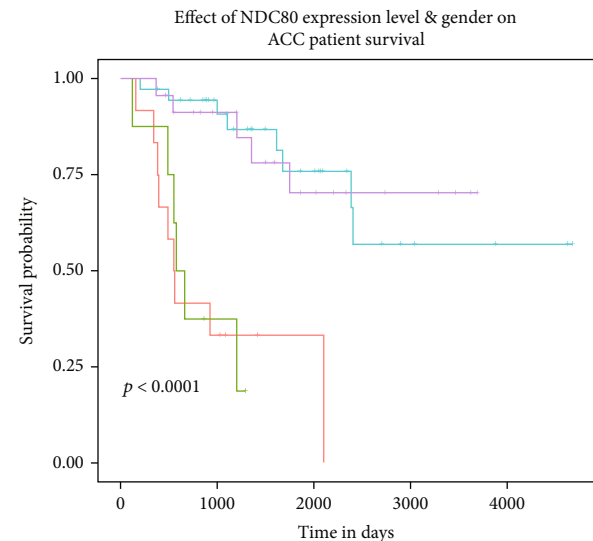
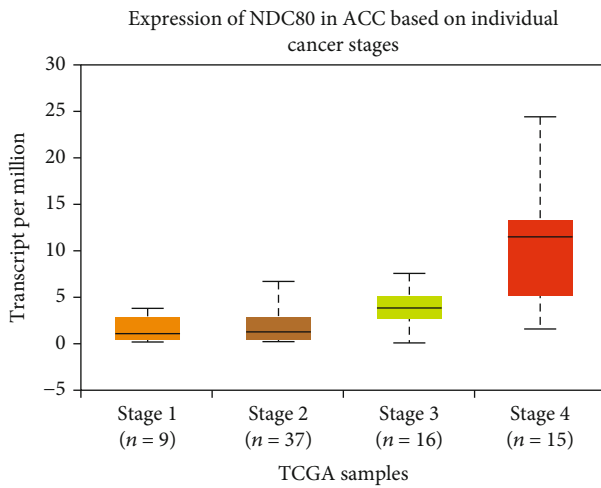
In the past 20 years, molecular biology studies on ACC have made significant progress [46, 47], but this cancer’s primary pathogenesis is still unclear. Moreover, recent epidemiological studies have shown that the incidence of ACC has increased yearly in the past 40 years, but the survival rate

of patients has not improved [3]. As a highly malignant tumor, there is an urgent need to find effective diagnostic and prognostic targets for identifying early-stage patients, developing proper treatments, and improving ACC’s poor prognosis. Therefore, using bioinformatics techniques to unravel the genomic properties of ACC at the molecular level is crucial for finding effective treatments and predicting patient survival and relapse risk, and there have been several successful cases of bioinformatics used in cancer research [48–51].

Our research selected GSE10927 (10 normal and 33 ACC tissues) and GSE19750 (4 normal and 44 ACC tissues)



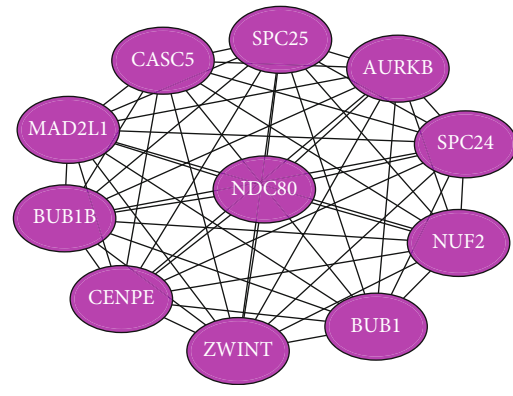
(a)



Expression level, gender
 — High expression + female (n = 12)
 — High expression + male (n = 8)
 — Low/medium expression + female (n = 36)
 — Low/medium + male (n = 23)

(b)

(c)



(d)

FIGURE 10: In-depth exploration of the biological value of the core gene NDC80. (a) mRNA expression of NDC80 in pan-cancer. (b) mRNA expression of NDC80 in different stages of ACC. The P -value between stage 1 and stage 4 is $5.7562E-03$. (c) The effect of NDC80 mRNA expression level and patient gender on the overall survival of ACC patients. (d) PPI map between NDC80 and the ten most closely related NDC80 protein molecules.

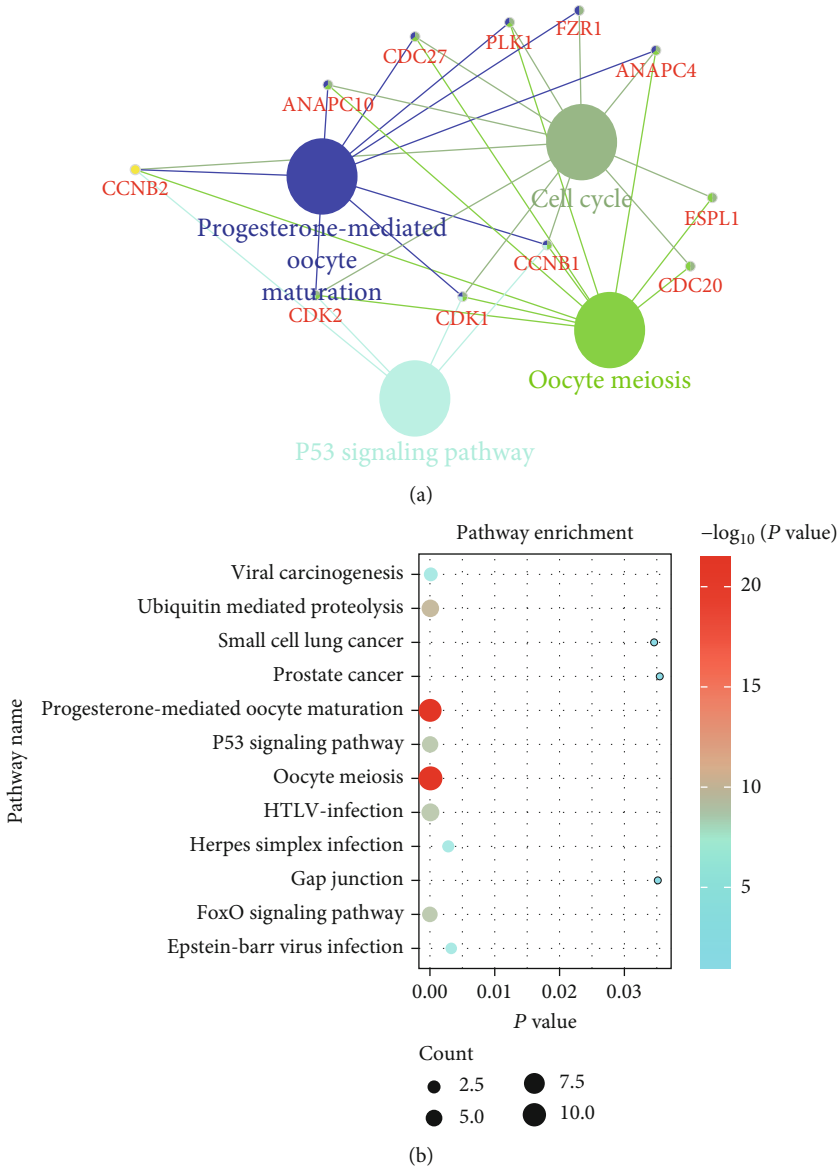
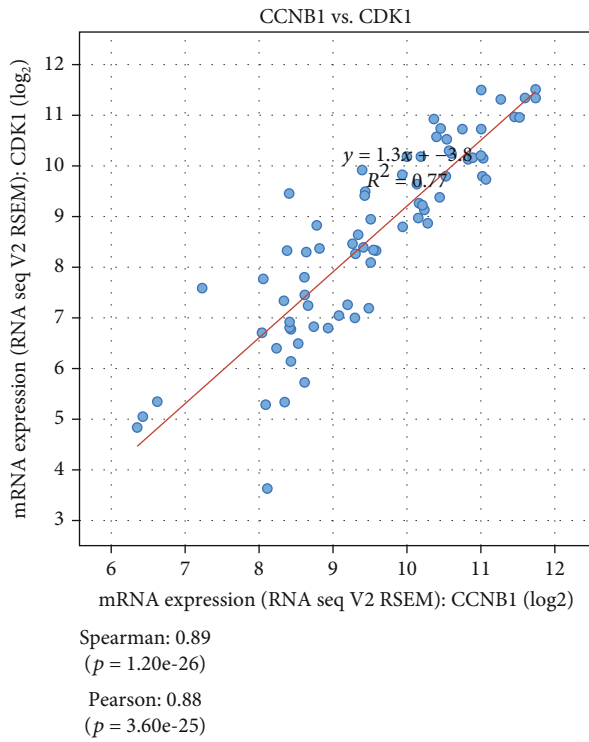
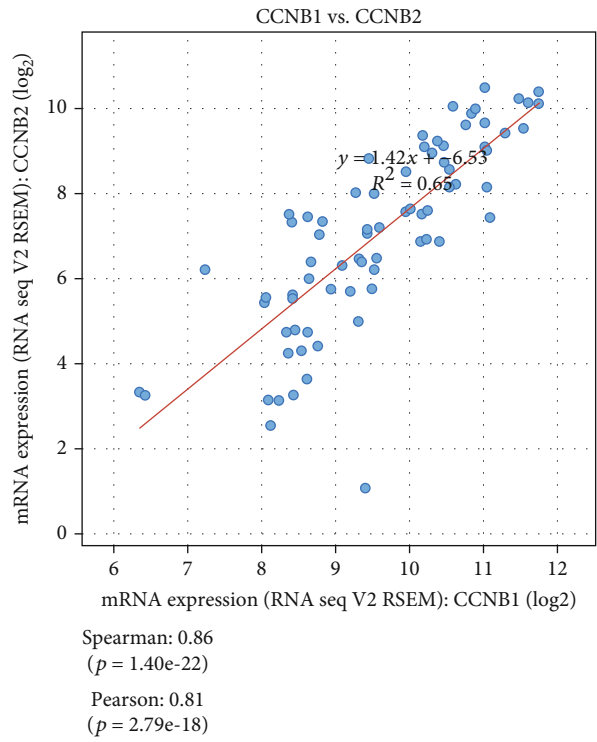


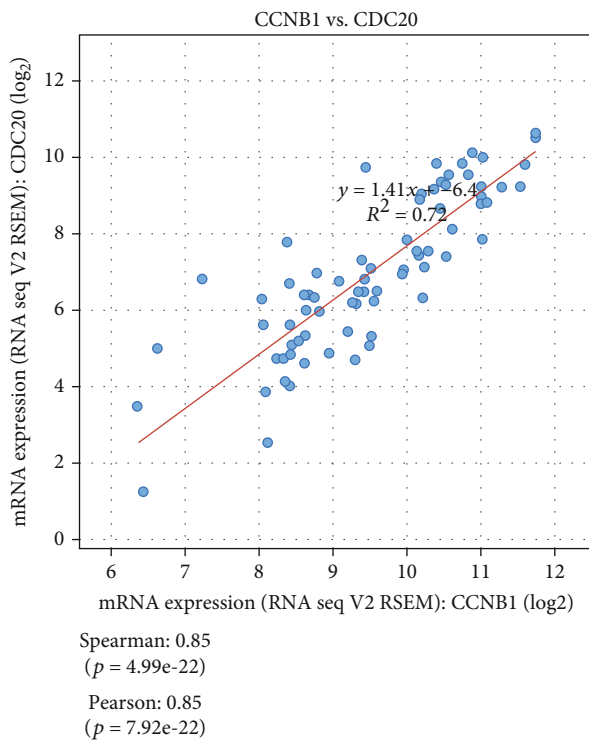
FIGURE 11: Continued.



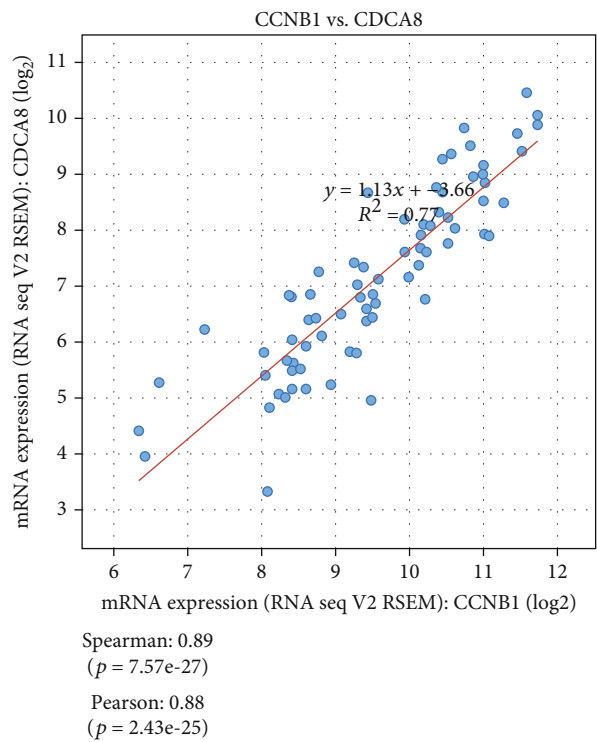
(c)



(d)



(e)



(f)

FIGURE 11: Continued.

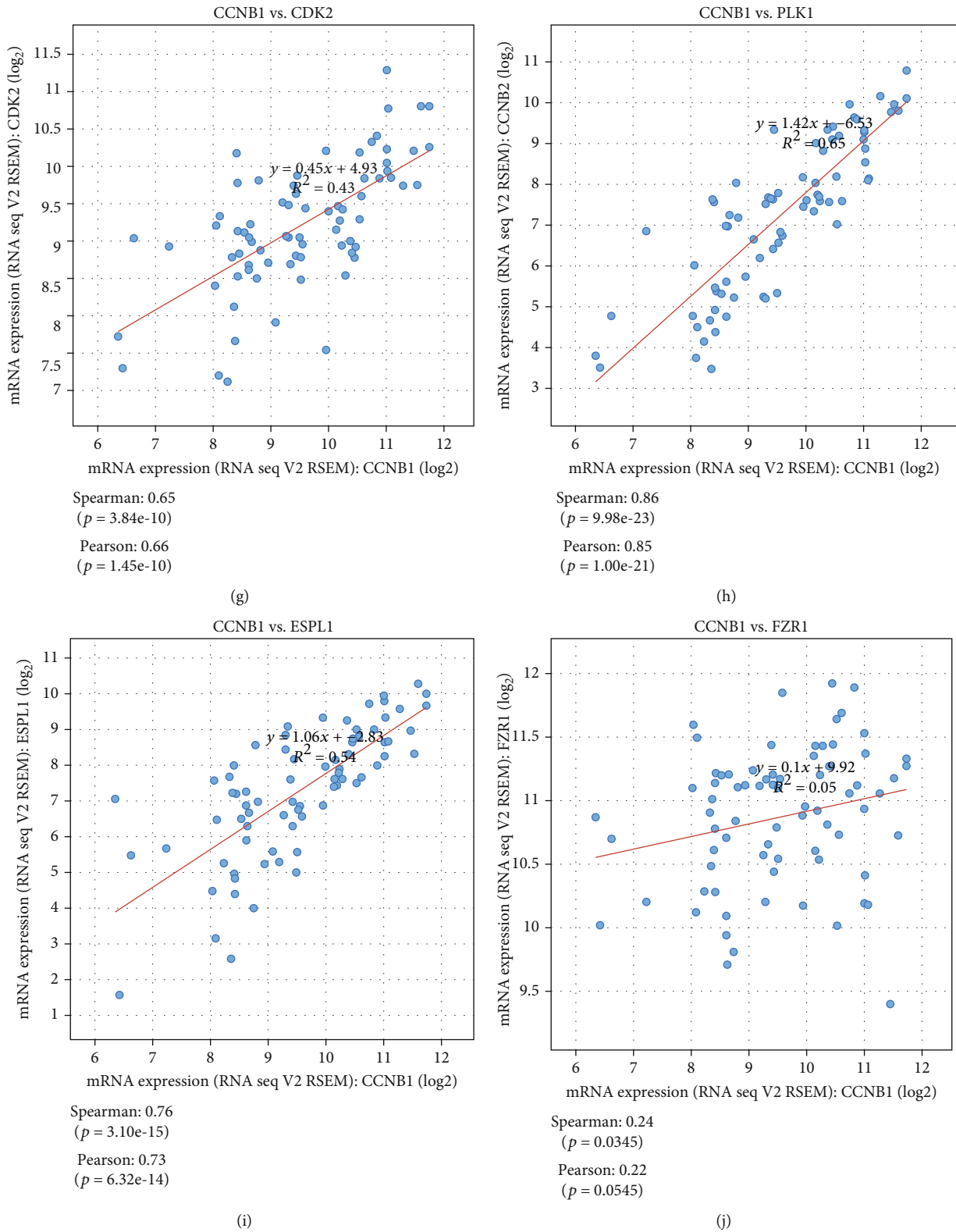


FIGURE 11: Functional and co-expression analysis of CCNB1. (a–b) Pathway enrichment analysis of CCNB1. (c–j) Co-expression analysis of CCNB1 and related genes.

from the GEO database. After analyzing R language, these results were cross-correlated with data from TCGA, and 28 up-regulated and 462 down-regulated DEGs were enrolled

for our study. Then, we carried out GO functional analysis and pathway analysis (KEGG, REACTOME, PANTHER, and BIOCYC) using WebGestalt and KOBAS websites to

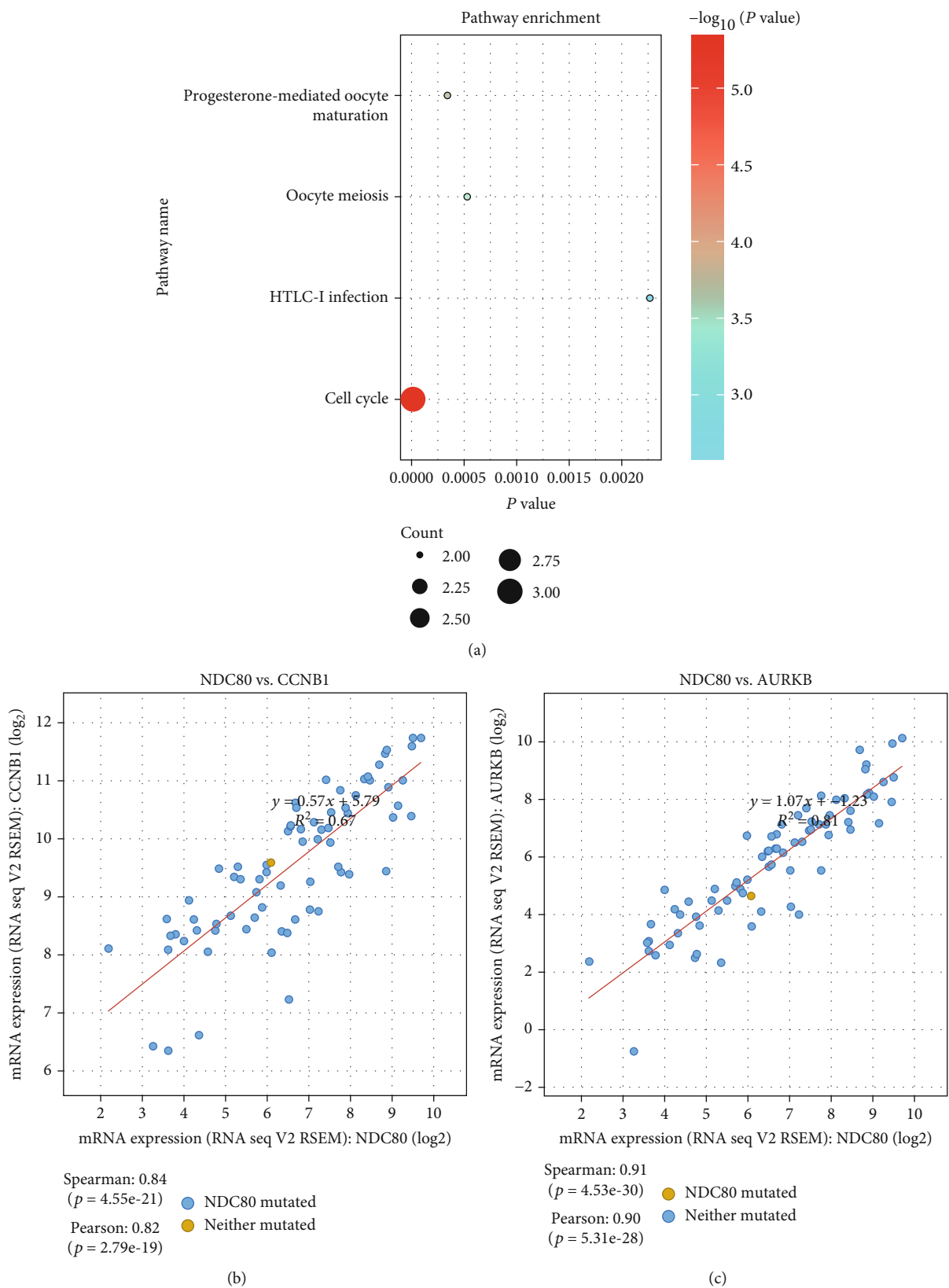


FIGURE 12: Continued.

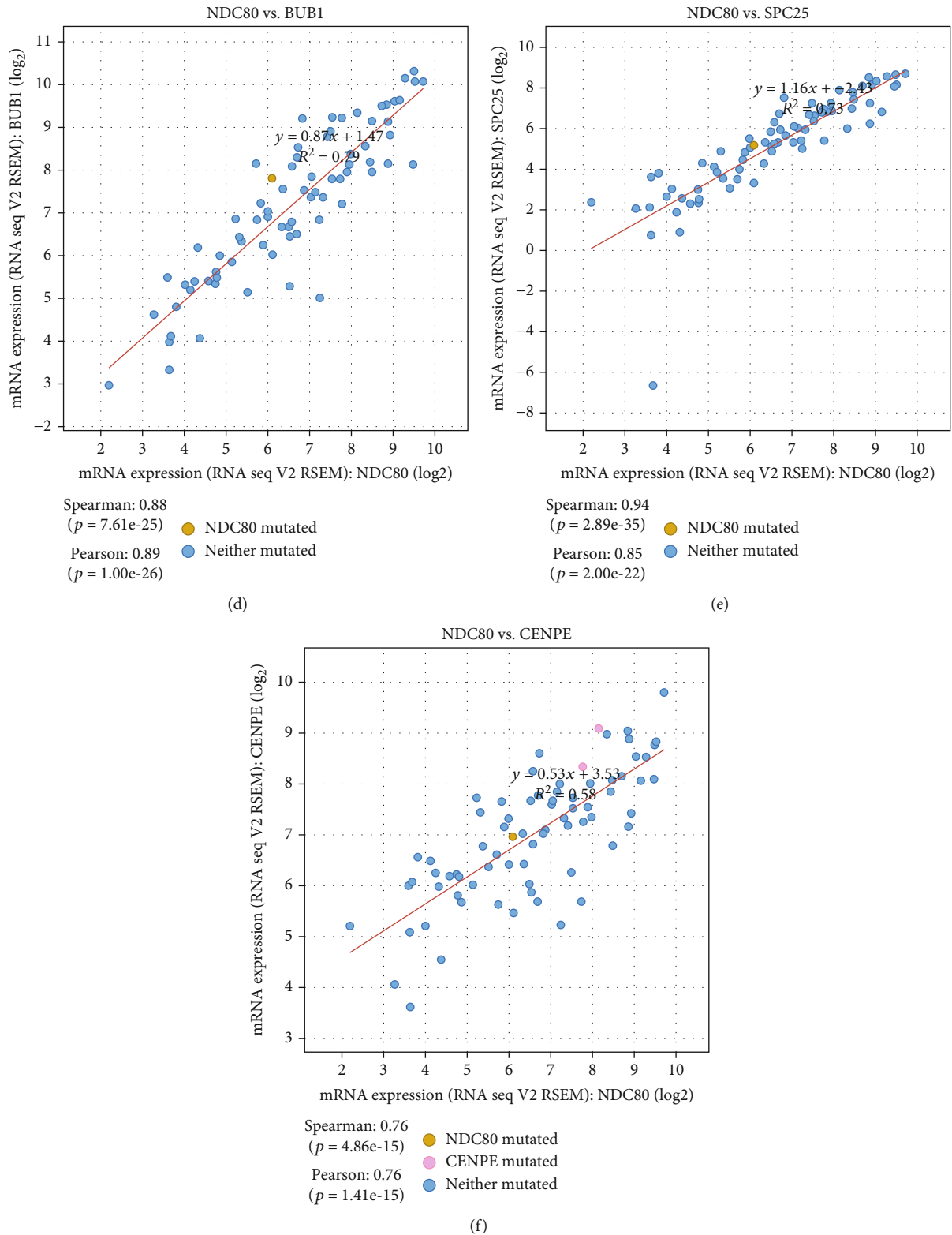


FIGURE 12: Functional and co-expression analysis of NDC80. (a) Pathway enrichment analysis of NDC80. (b–f) Co-expression analysis of NDC80 and related genes.

learn these candidates' gene function and regulatory process. Moreover, PPI network analysis was used to search for the hub genes through String database, and 17 dominant genes

were considered. In addition, the cBioPortal database helped investigate the mutations in these genes. The GEPIA website was applied to assess the extent of differential expression,

overall survival (OS), and disease-free survival (DFS). After excluding genes unrelated to ACC's prognosis, we repeated pathway analysis on the remaining genes and acquired two target genes by three different algorithms. Eventually, we demonstrated that CCNB1 and NDC80 were associated with ACC's diagnosis and prognosis and could be considered vital biomarkers for future clinical use.

CCNB1, also known as Cyclin B1, is essential for controlling cell cycle during the G2/M (mitosis) transition [52]. Our results showed that the expression of CCNB1 was elevated in many cancers compared to normal cases, such as esophageal cancer, gastric cancer, colorectal cancer, liver cancer, and breast cancer [53–56]. CCNB1 was positively correlated with the stage of ACC. As the degree of disease increased, the expression of this gene also increased. This denoted that CCNB1 can distinguish the severity of this cancer. Ten genes (ESPL1, CDK2, CDK1, ANAPC4, FZR1, PLK1, CDC27, CDC20, CCNB2, and ANAPC10) refer to 4 pathways (P53 signaling pathway, cell cycle, progesterone-mediated oocyte maturation, and oocyte meiosis) connected with CCNB1 were filtered out by our results. CCNB2 can compensate for CCNB1 in oocyte meiosis [57] and works consistently in ACC. CCNB1 and CDK1 were co-expressed in ACC, and this action was also acknowledged in breast cancer susceptibility, progression, and survival of Chinese women [58]. Lohberger et al. proposed that CCNB1 and CDK1/2 are involved in the G2/M cell cycle checkpoint, providing an inner relationship between CCNB1 and CDK family [59]. The combination of CCNB1 and CDC20 high expression could predict the poor prognosis of liver cancer [60], similar to what we got in ACC. In a word, CCNB1 was involved in the process of ACC disease progression and occupied the central position of several pathways, implying that it could become a potential gene for further study.

NDC80 is required for chromosome segregation and spindle checkpoint activity [61]. It could affect the growth of hepatocellular carcinoma [62] and promote proliferation and metastasis of colon cancer [62]. In our study, the expression of NDC80 was much higher in ACC stage 4 than in stage 1-3 but had nothing to do with gender. NDC80 was mainly centralized in cell cycle pathways and had protein interaction with CASC5, SPC25, AURKB, SPC24, NUF2, BUB1, ZWINT, CENPE, BUB1B, and MAD2L1. We should pay attention to whether NDC80 and CCNB1 had a co-expression in ACC, prompting that the combined detection of these two genes can improve the diagnostic rate of ACC. NDC80 could also be a promising marker to identify ACC and estimate the prognosis of this cancer.

5. Conclusions

Based on a series of bioinformatics analyses, our study concluded that CCNB1 and NDC80 are particularly relevant for the high risk and poor prognosis of ACC in theory, suggesting that these two genes can be beneficial for proper diagnosis and treatment of this disease. However, more efforts should be invested in clinical experiments to learn

these genes' biological functions and pathological evolution in ACC.

Abbreviations

ACC:	Adrenocortical carcinoma
TCGA:	The Cancer Genome Atlas
GEO:	Gene Expression Omnibus
GEPIA:	Gene expression profiling interactive analysis
C3AR1:	Complement C3a receptor 1
CCNB1:	Cyclin B1
CDC20:	Cell division cycle 20
CENPU:	Centromere protein U
FOXM1:	Forkhead box M1
KIF4A:	Kinesin family member 4A
KIF11:	Kinesin family member 11
KIF20A:	Kinesin family member 20A
MAD2L1:	Mitotic arrest deficient 2 like 1
NCAPG:	Non-SMC condensin I complex subunit G
NDC80:	NDC80 kinetochore complex component
NUF2:	NUF2 component of NDC80 kinetochore complex
PBK:	PDZ binding kinase
RACGAP1:	Rac GTPase activating protein 1
RRM2:	Ribonucleotide reductase regulatory subunit M2
TOP2A:	DNA topoisomerase II alpha
TPX2:	TPX2 microtubule nucleation factor
KICH:	Kidney chromophobe
KIRC:	Kidney renal clear cell carcinoma
KIRP:	Kidney renal papillary cell carcinoma
PAAD:	Pancreatic adenocarcinoma
BLCA:	Bladder urothelial carcinoma
ESPL1:	Extra spindle pole bodies like 1, separate
CDK2:	Cyclin-dependent kinase 2
CDK1:	Cyclin-dependent kinase 1
ANAPC4:	Anaphase promoting complex subunit 4
FZR1:	Fizzy and cell division cycle 20 related 1
PLK1:	Polo-like kinase 1
CDC27:	Cell division cycle 27
CCNB2:	Cyclin B2
ANAPC10:	Anaphase promoting complex subunit 10
SPC25:	SPC25 component of NDC80 kinetochore complex
AURKB:	Aurora kinase B
SPC24:	SPC24 component of NDC80 kinetochore complex
BUB1:	BUB1 mitotic checkpoint serine/threonine kinase
ZWINT:	ZW10 interacting kinetochore protein
CENPE:	Centromere protein E
BUB1B:	BUB1 mitotic checkpoint serine/threonine kinase B.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Guangzhen Wu and Yingkun Xu designed the research methods and analyzed the data. Xiunan Li and Jiayi Li participated in data collection. Leizuo Zhao, Zicheng Wang, and Peizhi Zhang drafted and revised the manuscript. All authors approved the version to be released and agreed to be responsible for all aspects of the work.

Acknowledgments

We thank Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) for providing publicly available data. This project is supported by the Scientific Research Fund of Liaoning Provincial Education Department (No. LZ2020071), the Doctoral Start-up Foundation of Liaoning Province (No. 2021-BS-209), the Dalian Youth Science and Technology Star (No. 2021RQ010), and the Doctoral Research Innovation Project of the First Affiliated Hospital of Chongqing Medical University (No. CYYY-BSYJSCXXM-202213).

References

- [1] M. D. Wooten and D. K. King, "Adrenal cortical carcinoma. Epidemiology and treatment with mitotane and a review of the literature," *Cancer*, vol. 72, no. 11, pp. 3145–3155, 1993.
- [2] S. E. Long and B. S. Miller, "Adrenocortical cancer treatment," *The Surgical Clinics of North America*, vol. 99, no. 4, pp. 759–771, 2019.
- [3] T. Chandrasekar, H. Goldberg, Z. Klaassen et al., "The who, when, and why of primary adrenal malignancies: insights into the epidemiology of a rare clinical entity," *Cancer*, vol. 125, no. 7, pp. 1050–1059, 2019.
- [4] I. Abe and A. K. Lam, "Anaplastic thyroid carcinoma: updates on WHO classification, clinicopathological features and staging," *Histology and Histopathology*, vol. 36, no. 3, pp. 239–248, 2021.
- [5] T. Else, A. C. Kim, A. Sabolch et al., "Adrenocortical carcinoma," *Endocrine Reviews*, vol. 35, no. 2, pp. 282–326, 2014.
- [6] R. Libé, "Adrenocortical carcinoma (ACC): diagnosis, prognosis, and treatment," *Frontiers in Cell and Development Biology*, vol. 3, p. 45, 2015.
- [7] A. Vaidya, M. Nehs, and K. Kilbridge, "Treatment of adrenocortical carcinoma," *Surg Pathol Clin*, vol. 12, no. 4, pp. 997–1006, 2019.
- [8] T. Else, A. R. Williams, A. Sabolch, S. Jolly, B. S. Miller, and G. D. Hammer, "Adjuvant therapies and patient and tumor characteristics associated with survival of adult patients with adrenocortical carcinoma," *The Journal of Clinical Endocrinology and Metabolism*, vol. 99, no. 2, pp. 455–461, 2014.
- [9] B. Alolio and M. Fassnacht, "Adrenocortical carcinoma: clinical update," *The Journal of Clinical Endocrinology and Metabolism*, vol. 91, no. 6, pp. 2027–2037, 2006.
- [10] I. Erdogan, T. Deutschbein, C. Jurowich et al., "The role of surgery in the management of recurrent adrenocortical carcinoma," *The Journal of Clinical Endocrinology and Metabolism*, vol. 98, no. 1, pp. 181–191, 2013.
- [11] E. Mirallie, C. Blanchard, C. Caillard et al., "Prise en charge d'une tumeur cortico-surrénalienne : impact du traitement chirurgical," *Annales d'Endocrinologie*, vol. 80, no. 5–6, pp. 308–313, 2019.
- [12] E. Baudin, "Adrenocortical carcinoma," *Endocrinology and Metabolism Clinics of North America*, vol. 44, no. 2, pp. 411–434, 2015.
- [13] R. Libé, "Clinical and molecular prognostic factors in adrenocortical carcinoma," *Minerva Endocrinologica*, vol. 44, no. 1, pp. 58–69, 2019.
- [14] P. S. Soon, A. J. Gill, D. E. Benn et al., "Microarray gene expression and immunohistochemistry analyses of adrenocortical tumors identify IGF2 and Ki-67 as useful in differentiating carcinomas from adenomas," *Endocrine-Related Cancer*, vol. 16, no. 2, pp. 573–583, 2009.
- [15] C. Wang, Y. Sun, H. Wu, D. Zhao, and J. Chen, "Distinguishing adrenal cortical carcinomas and adenomas: a study of clinicopathological features and biomarkers," *Histopathology*, vol. 64, no. 4, pp. 567–576, 2014.
- [16] S. Zheng, A. D. Cherniack, N. Dewal et al., "Comprehensive pan-genomic characterization of adrenocortical carcinoma," *Cancer Cell*, vol. 30, no. 2, p. 363, 2016.
- [17] S. Das, M. Sengupta, N. Islam et al., "Weineke criteria, Ki-67 index and p53 status to study pediatric adrenocortical tumors: is there a correlation?," *Journal of Pediatric Surgery*, vol. 51, no. 11, pp. 1795–1800, 2016.
- [18] Z. Tömböl, P. M. Szabó, V. Molnár et al., "Integrative molecular bioinformatics study of human adrenocortical tumors: microRNA, tissue-specific target prediction, and pathway analysis," *Endocrine-Related Cancer*, vol. 16, no. 3, pp. 895–906, 2009.
- [19] C. Laurell, D. Velázquez-Fernández, K. Lindsten et al., "Transcriptional profiling enables molecular classification of adrenocortical tumours," *European Journal of Endocrinology*, vol. 161, no. 1, pp. 141–152, 2009.
- [20] M. Doghman, J. Cazareth, and E. Lalli, "The T cell factor/ β -catenin antagonist PKF115-584 inhibits proliferation of adrenocortical carcinoma cells," *The Journal of Clinical Endocrinology and Metabolism*, vol. 93, no. 8, pp. 3222–3225, 2008.
- [21] S. Sbiera, S. Schmulz, G. Assie et al., "High diagnostic and prognostic value of steroidogenic factor-1 expression in adrenal tumors," *The Journal of Clinical Endocrinology and Metabolism*, vol. 95, no. 10, pp. E161–E171, 2010.
- [22] A. Sabolch, M. Feng, K. Griffith, G. Hammer, G. Doherty, and E. Ben-Josef, "Adjuvant and definitive radiotherapy for adrenocortical carcinoma," *International Journal of Radiation Oncology • Biology • Physics*, vol. 80, no. 5, pp. 1477–1484, 2011.
- [23] E. Duregon, M. Volante, J. Giorcelli, M. Terzolo, E. Lalli, and M. Papotti, "Diagnostic and prognostic role of steroidogenic factor 1 in adrenocortical carcinoma: a validation study focusing on clinical and pathologic correlates," *Human Pathology*, vol. 44, no. 5, pp. 822–828, 2013.
- [24] O. Barreau, G. Assié, H. Wilmot-Roussel et al., "Identification of a CpG island methylator phenotype in adrenocortical carcinomas," *The Journal of Clinical Endocrinology and Metabolism*, vol. 98, no. 1, pp. E174–E184, 2013.
- [25] X. C. Shen, C. X. Gu, Y. Q. Qiu, C. J. Du, Y. B. Fu, and J. J. Wu, "Estrogen receptor expression in adrenocortical carcinoma,"

- Journal of Zhejiang University. Science. B*, vol. 10, no. 1, pp. 1–6, 2009.
- [26] M. Jain, N. Rechache, and E. Kebebew, “Molecular markers of adrenocortical tumors,” *Journal of Surgical Oncology*, vol. 106, no. 5, pp. 549–556, 2012.
- [27] J. Long, Z. Zhang, Z. Liu, Y. Xu, and C. Ge, “Identification of genes and pathways associated with pancreatic ductal adenocarcinoma by bioinformatics analyses,” *Oncology Letters*, vol. 11, no. 2, pp. 1391–1397, 2016.
- [28] J. Rung and A. Brazma, “Reuse of public genome-wide gene expression data,” *Nature Reviews. Genetics*, vol. 14, no. 2, pp. 89–99, 2013.
- [29] T. Barrett, D. B. Troup, S. E. Wilhite et al., “NCBI GEO: mining tens of millions of expression profiles—database and tools update,” *Nucleic Acids Research*, vol. 35, no. Database, pp. D760–D765, 2007.
- [30] M. J. Demeure, K. E. Coan, C. S. Grant et al., “PTTG1 overexpression in adrenocortical cancer is associated with poor survival and represents a potential therapeutic target,” *Surgery*, vol. 154, no. 6, pp. 1405–1416, 2013, discussion 1416.
- [31] C. R. Legendre, M. J. Demeure, T. G. Whitsett et al., “Pathway implications of aberrant global methylation in adrenocortical cancer,” *PLoS One*, vol. 11, no. 3, article e0150629, 2016.
- [32] T. J. Giordano, R. Kuick, T. Else et al., “Molecular classification and prognostication of adrenocortical tumors by transcriptome profiling,” *Clinical Cancer Research*, vol. 15, no. 2, pp. 668–676, 2009.
- [33] K. Tomczak, P. Czerwinska, and M. Wiznerowicz, “The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge,” *Contemp Oncol (Pozn)*, vol. 19, no. 1A, pp. A68–A77, 2015.
- [34] M. S. Barker, K. M. Dlugosch, L. Dinh et al., “EvoPipes.net: bioinformatic tools for ecological and evolutionary genomics,” *Evolutionary Bioinformatics Online*, vol. 6, pp. 143–149, 2010.
- [35] J. Wang, D. Duncan, Z. Shi, and B. Zhang, “WEB-based GENE SeT AnaLYsis Toolkit (WebGestalt): update 2013,” *Nucleic Acids Research*, vol. 41, no. W1, pp. W77–W83, 2013.
- [36] P. Shannon, A. Markiel, O. Ozier et al., “Cytoscape: a software environment for integrated models of biomolecular interaction networks,” *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [37] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, “KEGG: new perspectives on genomes, pathways, diseases and drugs,” *Nucleic Acids Research*, vol. 45, no. D1, pp. D353–D361, 2017.
- [38] B. Jassal, L. Matthews, G. Viteri et al., “The reactome pathway knowledgebase,” *Nucleic Acids Research*, vol. 48, no. D1, pp. D498–D503, 2020.
- [39] H. Mi, S. Poudel, A. Muruganujan, J. T. Casagrande, and P. D. Thomas, “PANTHER version 10: expanded protein families and functions, and analysis tools,” *Nucleic Acids Research*, vol. 44, no. D1, pp. D336–D342, 2016.
- [40] R. Caspi, R. Billington, I. M. Keseler et al., “The MetaCyc database of metabolic pathways and enzymes - a 2019 update,” *Nucleic Acids Research*, vol. 48, no. D1, pp. D445–D453, 2020.
- [41] H. Ge, Y. Yan, D. Wu, Y. Huang, and F. Tian, “Potential role of LINC00996 in colorectal cancer: a study based on data mining and bioinformatics,” *Oncotargets and Therapy*, vol. Volume 11, pp. 4845–4855, 2018.
- [42] D. Szklarczyk, J. H. Morris, H. Cook et al., “The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible,” *Nucleic Acids Research*, vol. 45, no. D1, pp. D362–D368, 2017.
- [43] C. H. Chin, S. H. Chen, H. H. Wu, C. W. Ho, M. T. Ko, and C. Y. Lin, “CytoHubba: identifying hub objects and sub-networks from complex interactome,” *BMC Systems Biology*, vol. 8, Suppl 4, p. S11, 2014.
- [44] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, and Z. Zhang, “GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses,” *Nucleic Acids Research*, vol. 45, no. W1, pp. W98–W102, 2017.
- [45] J. Gao, B. A. Aksoy, U. Dogrusoz et al., “Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal,” *Science Signaling*, vol. 6, no. 269, p. p11, 2013.
- [46] C. A. Longui, “Adrenal cortical carcinoma in infancy,” *Rev Paul Pediatr*, vol. 37, no. 1, pp. 2-3, 2019.
- [47] L. Cerquetti, B. Bucci, G. Carpinelli et al., “Antineoplastic effect of a combined mitotane treatment/ionizing radiation in adrenocortical carcinoma: a preclinical study,” *Cancers (Basel)*, vol. 11, no. 11, p. 1768, 2019.
- [48] Y. Xu, G. Wu, J. Li et al., “Screening and identification of key biomarkers for bladder cancer: a study based on TCGA and GEO data,” *BioMed Research International*, vol. 2020, Article ID 8283401, 20 pages, 2020.
- [49] Z. Xu, B. Peng, Q. Liang et al., “Construction of a ferroptosis-related nine-lncRNA signature for predicting prognosis and immune response in hepatocellular carcinoma,” *Frontiers in Immunology*, vol. 12, article 719175, 2021.
- [50] Y. Cai, G. Wu, B. Peng et al., “Expression and molecular profiles of the AlkB family in ovarian serous carcinoma,” *Aging (Albany NY)*, vol. 13, no. 7, pp. 9679–9692, 2021.
- [51] X. Li, Y. Su, J. Zhang, Y. Zhu, Y. Xu, and G. Wu, “LAPTM5 plays a key role in the diagnosis and prognosis of testicular germ cell tumors,” *Int J Genomics*, vol. 2021, p. 8816456, 2021.
- [52] N. R. Brown, E. D. Lowe, E. Petri, V. Skamnaki, R. Antrobus, and L. N. Johnson, “Cyclin B and cyclin A confer different substrate recognition properties on CDK2,” *Cell Cycle*, vol. 6, no. 11, pp. 1350–1359, 2007.
- [53] T. Huang, S. Chen, H. Han et al., “Expression of Hsp90 α and cyclin B1 were related to prognosis of esophageal squamous cell carcinoma and keratin pearl formation,” *International Journal of Clinical and Experimental Pathology*, vol. 7, no. 4, pp. 1544–1552, 2014.
- [54] L. J. Yang, Y. Chen, Q. Ma et al., “Effect of betulinic acid on the regulation of Hiwi and cyclin B1 in human gastric adenocarcinoma AGS cells,” *Acta Pharmacologica Sinica*, vol. 31, no. 1, pp. 66–72, 2010.
- [55] Y. Fang, H. Yu, X. Liang, J. Xu, and X. Cai, “Chk1-induced CCNB1 overexpression promotes cell proliferation and tumor growth in human colorectal cancer,” *Cancer Biology & Therapy*, vol. 15, no. 9, pp. 1268–1279, 2014.
- [56] J. Gu, X. Liu, J. Li, and Y. He, “MicroRNA-144 inhibits cell proliferation, migration and invasion in human hepatocellular carcinoma by targeting CCNB1,” *Cancer Cell International*, vol. 19, no. 1, p. 15, 2019.
- [57] L. Zhuang, Z. Yang, and Z. Meng, “Upregulation of BUB1B, CCNB1, CDC7, CDC20, and MCM3 in tumor tissues predicted worse overall survival and disease-free survival in hepatocellular carcinoma patients,” *BioMed Research International*, vol. 2018, Article ID 7897346, 2018.

- [58] Y. Li, Y. L. Chen, Y. T. Xie et al., “Association study of germline variants in CCNB1 and CDK1 with breast cancer susceptibility, progression, and survival among Chinese Han women,” *PLoS One*, vol. 8, no. 12, article e84489, 2013.
- [59] B. Lohberger, A. Leithner, N. Stuendl, H. Kaltenegger, W. Kullich, and B. Steinecker-Frohnwieser, “Diacerein retards cell growth of chondrosarcoma cells at the G2/M cell cycle checkpoint via cyclin B1/CDK1 and CDK2 downregulation,” *BMC Cancer*, vol. 15, no. 1, p. 891, 2015.
- [60] Q. Shi, W. Wang, Z. Jia, P. Chen, K. Ma, and C. Zhou, “ISL1, a novel regulator of CCNB1, CCNB2 and c-MYC genes, promotes gastric cancer cell proliferation and tumor growth,” *Oncotarget*, vol. 7, no. 24, pp. 36489–36500, 2016.
- [61] Y. T. Lin, Y. Chen, G. Wu, and W. H. Lee, “Hec1 sequentially recruits Zwint-1 and ZW10 to kinetochores for faithful chromosome segregation and spindle checkpoint control,” *Oncogene*, vol. 25, no. 52, pp. 6901–6914, 2006.
- [62] X. K. Xing, H. Y. Wu, H. L. Chen, and H. G. Feng, “NDC80 promotes proliferation and metastasis of colon cancer cells,” *Genetics and Molecular Research*, vol. 15, no. 2, 2016.

Research Article

Identification of Human Retinal Organoid Cell Differentiation-Related Genes via Single-Cell Sequencing Data Analysis

He Dong ¹, Liang Yu ¹, Jian Song ¹, Lili Ji ¹, Xiaoxia Yu ² and Lijun Zhang ¹

¹Department of Ophthalmology, Dalian No. 3 People's Hospital, Dalian 116033, China

²Department of Blood Purification, Affiliated Zhongshan Hospital of Dalian University, Dalian 116001, China

Correspondence should be addressed to Xiaoxia Yu; yuxiaoxiayt@126.com and Lijun Zhang; zzzhanglj@163.com

Received 12 April 2022; Accepted 29 June 2022; Published 8 August 2022

Academic Editor: Chuan Lu

Copyright © 2022 He Dong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. To study the development process of the human retina, we analyzed the development track of main cell types and transitional cell populations, identifying the retinal organoid cell differentiation-related genes (RDRGs). **Methods.** Single-cell RNA sequencing data (scRNA-Seq) of human retinal organoids were downloaded from Gene Expression Omnibus (GEO) database in this study. Data were processed with quality analysis and analysis of variance. Principal component analysis and *t*-distributed stochastic neighbor embedding were used to conduct dimension reduction analysis and type annotation for the screened data. Marker genes and RDRGs were identified by differential analysis. Cell differentiation characteristics were determined by trajectory analysis. Enrichment pathways were analyzed by Gene Ontology(GO) and Kyoto Encyclopedia of Genes and Genomes(KEGG), and functional modules were obtained by protein-protein interaction (PPI) network analysis. **Results.** iPSCs were mainly located at the root of differentiation trajectory, while neurons and astrocytes were distributed in different branches, respectively. Meanwhile, 220 RDRGs were obtained. They were involved in the biological functions related to vision and visual development, as well as significantly enriched in signaling pathways associated with retinal vascular development and retinal neuroregulation. Protein-protein interaction network construction and functional subnetwork analysis were conducted on RDRGs, and two functional submodules were obtained. The enrichment analysis presented that the two submodules played a vital role in retinal development, visual perception, and cell respiration. **Conclusions.** This study identified RDRGs and revealed the biological functions involved in these genes, which are expected to provide evidence for researching retinal development and diseases.

1. Introduction

As the premise for vision, vertebrate retina is a thin layer of cells in the posterior lining of the eyeball. The retina is transformed from the front of the brain by a complex network of signals. At present, the human retina is mainly studied from the perspective of the organoid derived from stem cells [1]. Stem cells include human embryonic stem cells (hESCs) and induced pluripotent stem cells (hiPSCs). These two types of cells can differentiate into the precursors of retinal photoreceptor cells (cones and rods) [2]. Retinal differentiation methods have also evolved from 2D adherent cultures to 3D cultures that can more easily replicate development in humans [1]. For example, Zhong et al. [3] rebuilt a complete tissue structure of retina via hiPSCs. Wahlin et al. [4]

derived 3D microretinas from human pluripotent stem cells (hPSCs), which were more similar to the retina structure observed in vivo. The study of retinal organoid differentiation in vitro is of great significance to the cognition of human retinal diseases.

Abnormal gene expression could cause retinal degeneration. A previous study presented that in a mouse model of oxidative retinal degeneration, TMEM97 knockout exacerbated oxidative stress response, thereby aggravating oxidative retinal degeneration [5]. It was also elaborated that increasing HTRA1 expression played a role in AMD [6]. With the popularization of high-throughput sequencing technology in the last few years, accumulating research has applied single-cell RNA sequencing (scRNA-Seq) technology into the analysis of disease transcriptomics. Compared

with bulk RNA-Seq technology, scRNA-Seq shows superiority in revealing gene expression features of different cells and in studying cell differentiation trajectory [7]. Wang et al. [8] identified differentially expressed genes related to glioblastoma (GBM) cell differentiation via analyzing scRNA-Seq data of human GBM. Voigt et al. [9] analyzed scRNA-Seq data of choroid in human retinal pigment epithelium (RPE), revealing the gene expression characteristics of choroidal cell population, and pointed out that the upregulation of RGCC expression might be related to the occurrence of AMD [9].

scRNA-Seq data of human retinal organoids were downloaded from Gene Expression Omnibus (GEO) database and analyzed through a series of bioinformatics methods. Cell subpopulation was identified during differentiation of human retinal organoids through trajectory analysis, and retinal organoid cell differentiation-related genes (RDRGs) were also identified. Enrichment analysis exhibited that these genes were mainly significantly enriched in biological functions related to visual development and retinal neuroregulation-relative signaling pathways. This study is expected to provide a strong theoretic basis for retinal development and disease research.

2. Materials and Methods

2.1. Data Sources and Technical Lines. scRNA-Seq data (GSE119343) of human retinal organoids including cone cells were downloaded from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). Samples were retinal organs with 8 months of development. scRNA-Seq data included sequencing data of 1346 cells from 72 retinal organoids [10] (Data S1). The sequencing data were based on Illumina HiSeq 2500 (Homo sapiens) sequencing platform (GPL16791). The flowchart of this study was shown in Figure 1.

2.2. Processing of scRNA-Seq Data. Quality analysis was performed on scRNA-Seq data by using the R package “Seurat” (<https://github.com/satijalab/seurat>) to delete low-quality sequencing data. Data exclusion criteria were as follows: (1) genes that can be detected in <10 cells, (2) cells with total gene number <100, and (3) cells with mitochondrial expressed genes $\geq 10\%$. The preserved data were normalized by the LogNormalize method (<https://rdrr.io/github/LTLA/scuttle/man/logNormCounts.html>).

Seurat is an R package for quality control, analysis, and exploration of scRNA-Seq data. Seurat enables users to identify and reveal sources of heterogeneity from the single-cell transcript side and integrate various types of single-cell data to undertake variance heterogeneity analysis and recognize HVG parameter as scRNA, selection method = “VST” [11].

2.3. Dimension Reduction Analysis and Cell Type Annotation. Principal component analysis (PCA) was performed on the normalized data with the RunPCA function in “Seurat” package. PC with $p < 0.05$ was chosen through JackStraw analysis and ElbowPlot for the following analysis. Then, cell clustering analysis was conducted on data via the nonlinear dimension reduction method of t -distributed stochastic neighbor embed-

ding (tSNE) [12]. Differential analysis ($|\log_2FC| > 0.8$, p value < 0.05) was undertaken on different clusters with Wilcoxon test (<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/wilcox.test>) to identify marker genes of cells in each cluster. Cell annotation information HumanPrimaryCellAtlasData in R package “SingleR” was used to annotate cells (<https://www.bioconductor.org/packages/release/bioc/html/SingleR.html>).

2.4. Trajectory Analysis and RDRG Identification. Trajectory analysis was conducted on cells using R package “monocle” to identify differentiation characteristics of cells. Monocle2 algorithm constructs single-cell differentiation trajectories based on reversed graph embedding (RGE) unsupervised machine learning technique [13]. Firstly, high-dimensional scRNA-Seq data were projected into low-dimensional space by reducing dimension with RGE. Afterward, each cell was taken as a part of cell differentiation to learn gene expression changes. A tree structure with “root” and “branch” was obtained. Finally, cells were arranged in the trajectory of the tree structure in pseudotime order. After single-cell trajectory was obtained, gene differential expression analysis was performed on cells distributed in the root and in different branches ($|\log_2FC| > 0.585$, FDR < 0.01) with cells distributed in the root as the control. The obtained differential expression genes were defined as RDRGs.

To analyze biological functions and signaling pathways RDRGs involved, GO annotation and KEGG enrichment analyses (q value < 0.05) were conducted on RDRGs with R package “clusterProfiler” [14].

2.5. Construction of RDRG Protein-Protein Interaction (PPI) Network and Submodule Analysis. PPI network analysis was performed on RDRGs on STRING website (<https://string-db.org/>) (interaction score > 0.4) [15]. Clustering analysis was undertaken in PPI network by using MCODE [16] plug-in of Cytoscape software, and functional submodules were constructed. To analyze biological functions of functional submodules, GO and KEGG enrichment analyses were performed on functional modules obtained by clustering through ClueGO [17] plug-in (q value < 0.05).

3. Results

3.1. scRNA-Seq Data Preprocessing and PCA. Through quality analysis of scRNA-Seq data of human retinal organoids, 0 low-quality cells were eliminated, and 1,346 cell gene expression data from 72 retinal organoids were preserved (Figure 2(a)). Detected gene number was positively correlated with the total count number obtained from sequencing (Figure 2(b)). Variance analysis was undertaken on the 13,547 genes, and the top 1,500 high variation genes were acquired, among which top 9 genes were IFITM1, CCL2, LGALS1, CRYAB, IFIT2, GFAP, CXCL14, ISL1, and AQP4 (Figure 2(c)). PCA dimension reduction was performed on the top 1,500 high variation genes. As shown in Figure 2(d), PC_1 and PC_2 could not significantly separate cells derived from different retinal organoids. Lastly, 9 PCs were obtained through evaluation ($p < 0.05$) for the

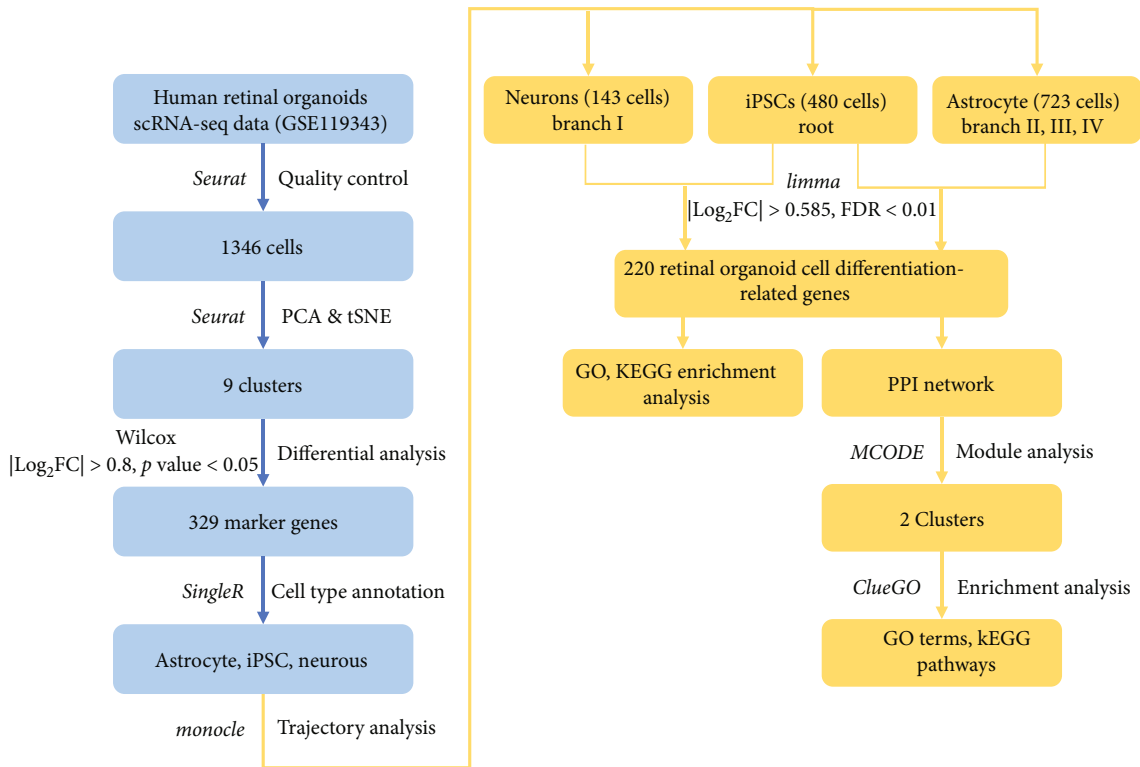


FIGURE 1: Technical route of scRNA-Seq data analysis of retinal organoids. Blue refers to dimension reduction clustering and cell annotation process of single cell sequencing data. Yellow refers to RDRG recognition and function analysis.

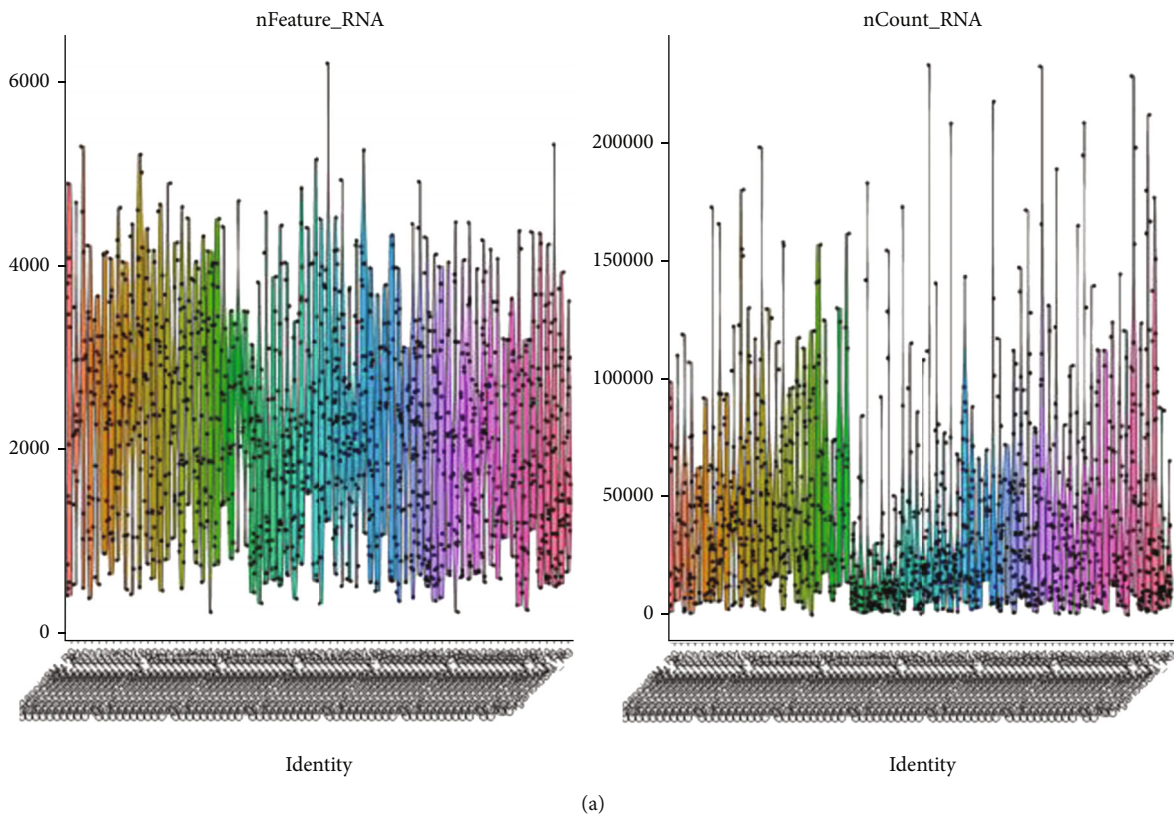
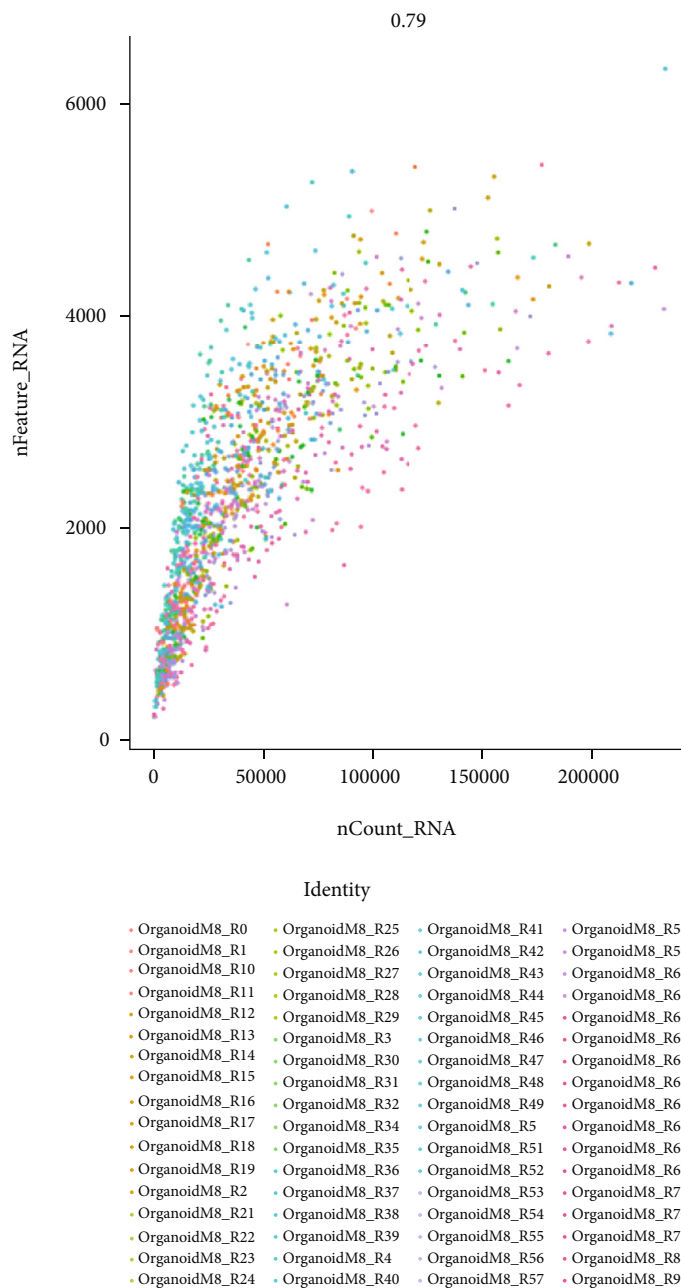
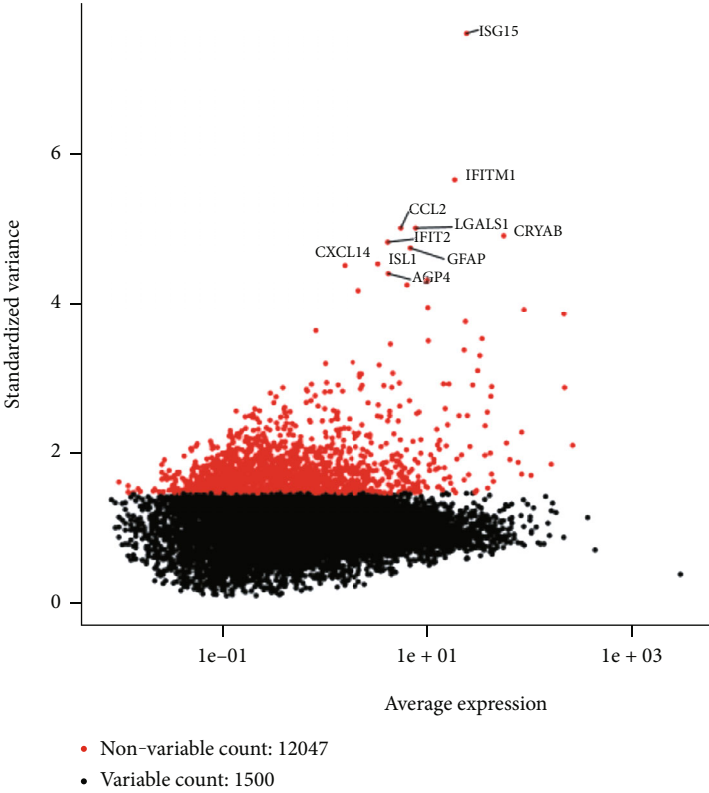


FIGURE 2: Continued.



(b)

FIGURE 2: Continued.



(c)

FIGURE 2: Continued.

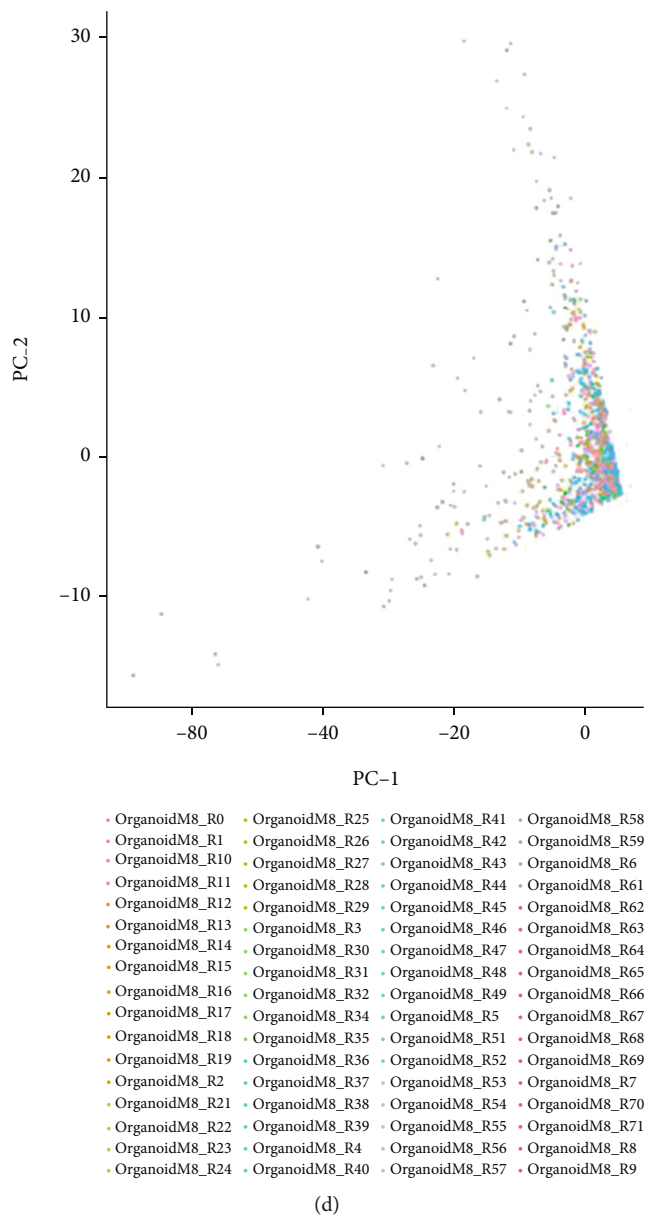
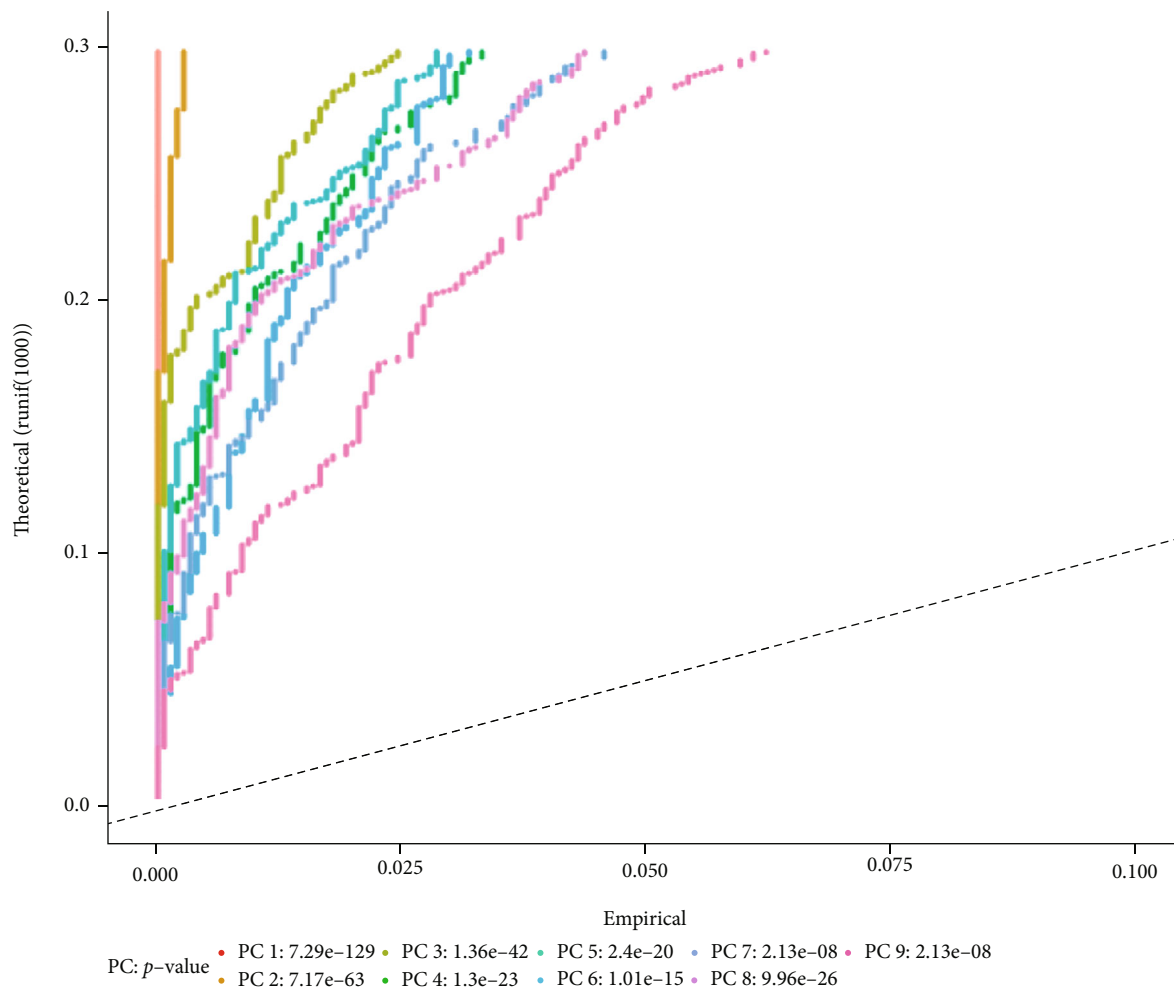


FIGURE 2: Continued.



(e)

FIGURE 2: Continued.

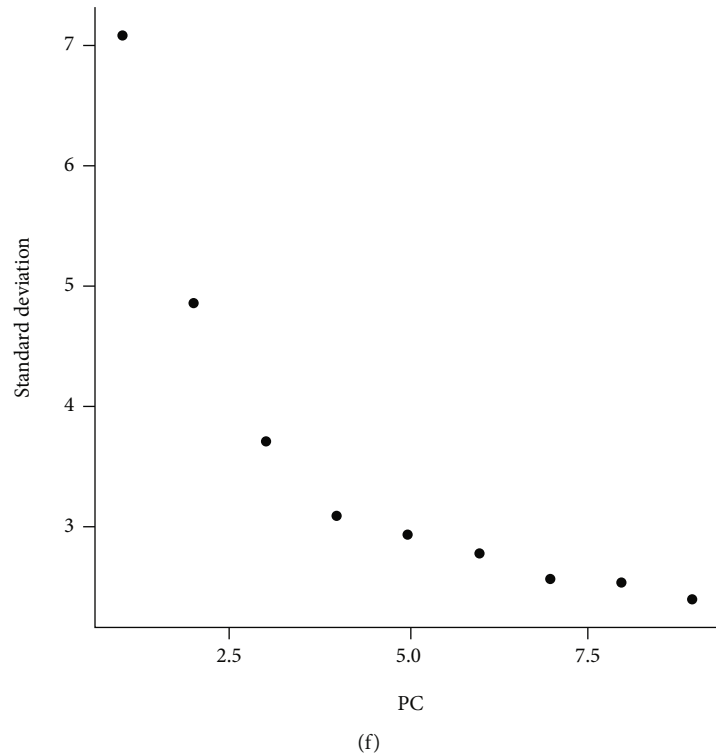


FIGURE 2: Quality analysis of scRNA-Seq data and PCA. (a) The left figure showed eliminated gene number of each cell data after quality analysis. The right figure showed total count number of each cell after quality analysis. (b) Correlation analysis between gene number detected by single cell and total count number obtained by sequencing. (c) Top 1,500 high variation genes obtained by variance analysis (red dots). (d) PCA plot. (e) PC with $p < 0.05$ through JackStraw analysis. (f) Standard deviation of PC shown in ElbowPlot.

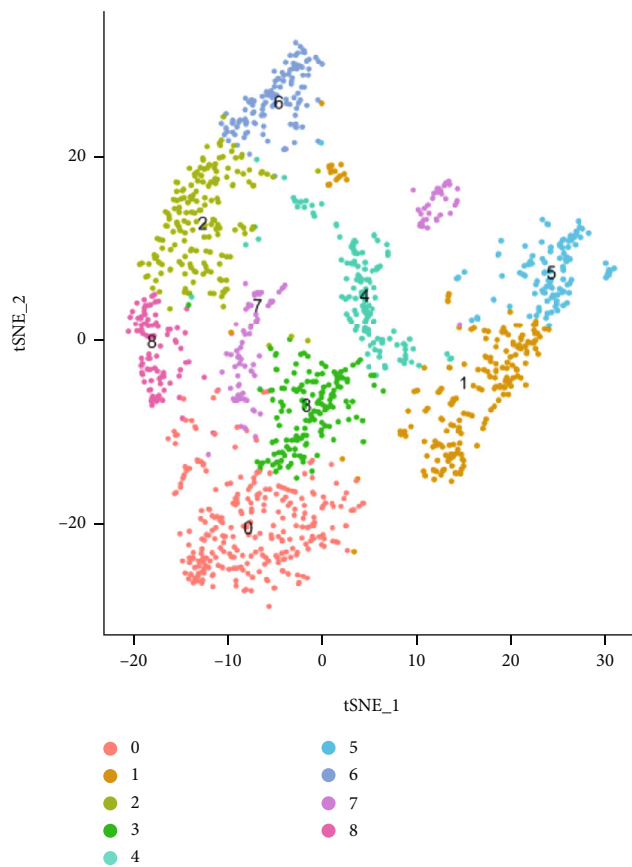
subsequent analyses (Figures 2(e) and 2(f)). Finally, we divided the cells of retinal organoids into 9 PCs.

3.2. Cell Type Annotation and Single-Cell Differentiation Trajectory Analysis. tSNE analysis was performed on 9 PCs obtained by linear dimension reduction, and cells were clustered into 9 clusters (Figure 3(a)). Differential analysis was conducted on different clusters to identify marker genes in each cell cluster (Figure 3(b)) (Data S1). A total of 3 types of cells were obtained by cell annotation with R package “SingleR.” Clusters 2, 6, 7, and 8 were annotated as iPSCs ($n = 480$). Cluster 3 including 143 cells was annotated as neurons. Clusters 0, 1, 4, and 5 including 723 cells were annotated as astrocytes (Figure 3(c)). Trajectory analysis was performed on annotated cells (Figures 3(d) and 3(e)). It was shown that iPSCs were mainly distributed in the root of trajectory. Neurons were mainly located in branch I. Astrocytes were mainly distributed in branch II, branch III, and branch IV. Together, through analysis of scRNA-Seq data of human retinal organoids, it was displayed that iPSCs could differentiate into astrocytes and neurons during the formation of human retinal organoids.

3.3. Identification of RDRGs and Enrichment Analyses. Cells in the root region were taken as the control group, and 58 RDRGs were obtained by differential expression analysis with branch I cells (Figure 4(a)). Differential expression analysis was performed with branch II, branch III, and

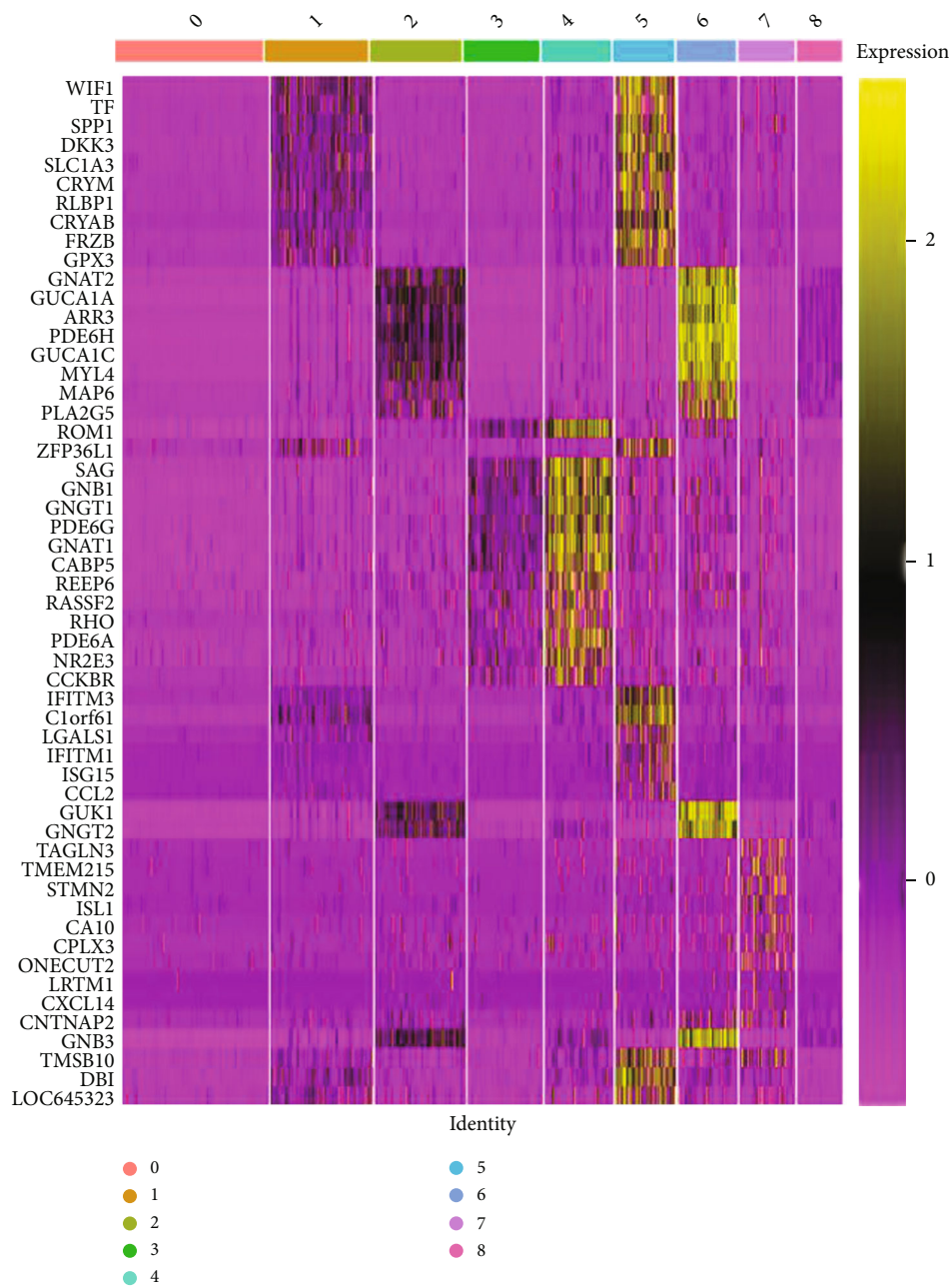
branch IV cells, and 191 RDRGs were obtained (Figure 4(b)). A total of 220 RDRGs were acquired by taking intersection with these RDRGs (Table S2) (Figure 4(c)). Further, GO enrichment analysis and KEGG pathway enrichment analysis were undertaken on these RDRGs. It was displayed that genes were significantly enriched in biological functions relevant to the vision and visual development, such as visual perception, sensory perception of light stimulus, response to radiation, response to light stimulus, eye development, visual system development, and sensory system development (Figure 4(d)). KEGG pathway enrichment analysis revealed that genes were mainly enriched in signaling pathways like phototransduction, apelin signaling pathway, retrograde endocannabinoid signaling, oxidative phosphorylation, and glutamatergic synapse (Figure 4(e)). On the whole, the 220 RDRGs were important when iPSCs differentiated into neurons and astrocytes and formed retina. Abnormal expression of these genes may also be relevant to degenerative retinal diseases.

3.4. Construction of PPI Network of RDRGs and Submodule Analysis. PPI network of RDRGs was analyzed using STRING website, including 178 nodes and 864 interactions (Figure 5(a)). Clustering analysis was performed on PPI network via MCODE. We obtained functional submodule Cluster 1 containing 23 genes and Cluster 2 containing 7 genes (Figures 5(b) and 5(c)). ClueGO was used to conduct GO



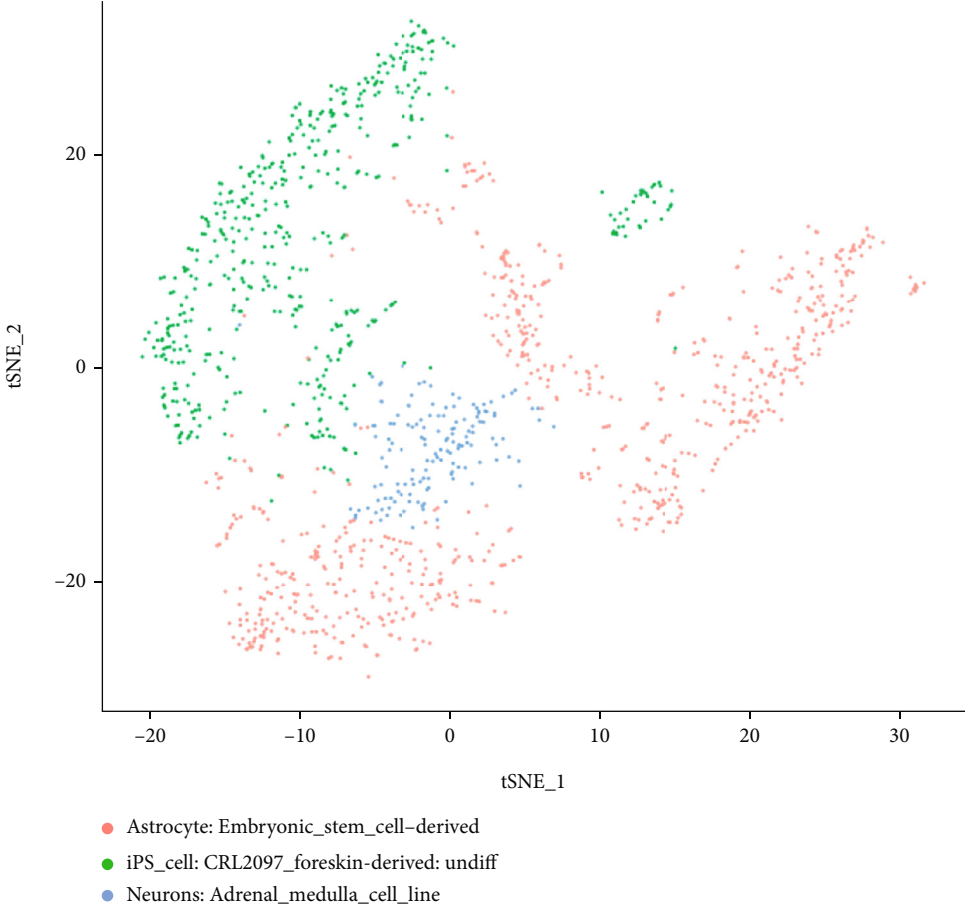
(a)

FIGURE 3: Continued.

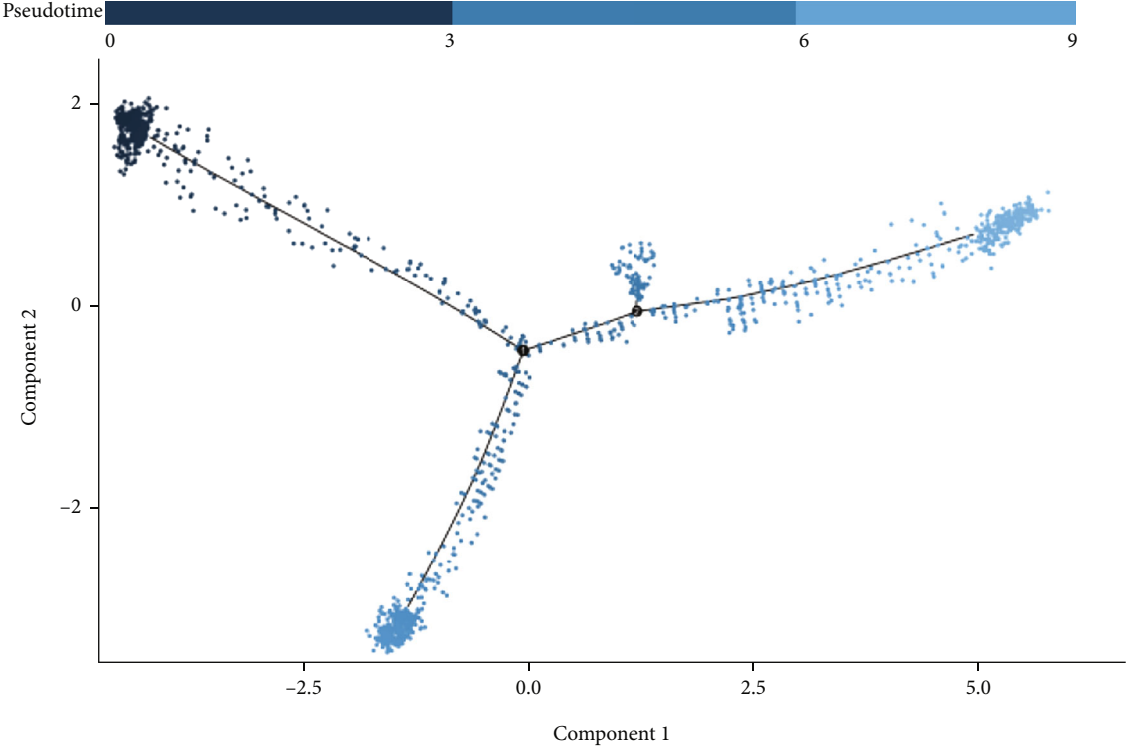


(b)

FIGURE 3: Continued.



(c)



(d)

FIGURE 3: Continued.

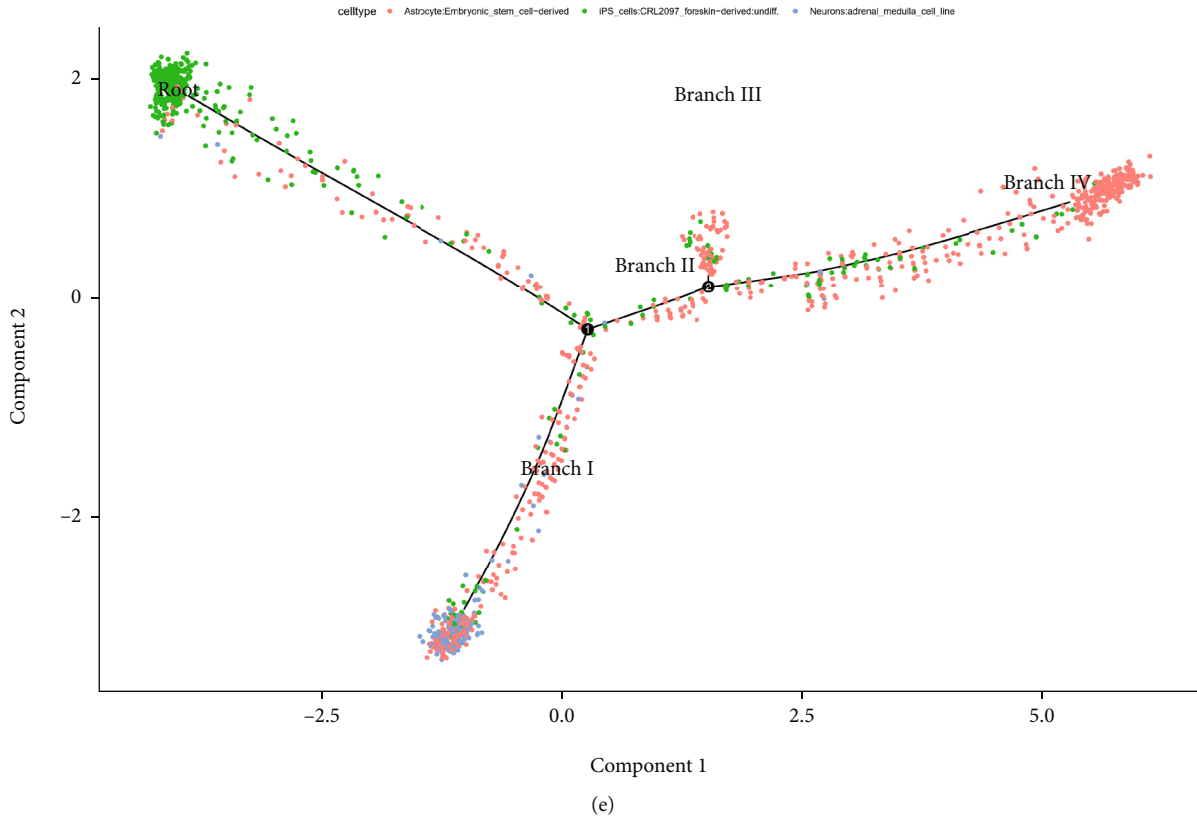


FIGURE 3: Cell annotation and trajectory analysis. (a) Nine clusters were obtained by cell clustering analysis through tSNE nonlinear dimension reduction. (b) Heat map of gene differential expression analysis of cells in each cluster (heat map only shows part of marker genes in each cluster of cells). (c) Cells in each cluster were annotated via marker genes to obtain iPSCs (green dots), neurons (blue dots), and astrocytes (pink dots). (d) Cells were sequenced along the trajectory according to pseudotime. Lighter color refers to higher differentiating level. (e) One root and four branches were obtained by trajectory analysis.

and KEGG enrichment analysis for the two submodules, and the results showed that Cluster 1 was mainly enriched in biological functions such as retinal development and visual perception and biological pathways of light transduction (Figure 5(d)). Cluster 2 was significantly enriched in respiratory-related biological functions, such as inner mitochondrial membrane complex, respirasome, and NADH dehydrogenase complex, and oxidative phosphorylation pathway (Figure 5(e)). There were two functional subsets in the PPI network of RDRGs. Genes in Cluster 1 were relevant to retinal development and visual signaling transduction. We speculated that aberrant expression of Cluster 1 genes may arise retinitis pigmentosa and degenerative retinal diseases.

4. Discussion

scRNA-Seq technology has gradually become the mainstream tool for cell heterogeneity investigation [18, 19]. Schafer et al. [20] performed scRNA-Seq on human cardiac fibroblasts and observed IL-11 as a critical factor for cardiovascular fibrosis. Der et al. [21] unraveled the molecular heterogeneity of lupus nephritis by scRNA-Seq. Guo et al. [22] applied scRNA-Seq to human embryonic stem cells and revealed bad effects of nicotine on the differentiation of these

cells. Cells are the basic units of life activities, and understanding cell heterogeneity and differentiation is of medical value. In this study, bioinformatics methods were used to identify three cell populations that are important in the process of organoid differentiation. Through trajectory analysis, RDRGs were obtained, providing reference for the study of organoid differentiation mechanism and retinal diseases.

In the present study, dimension reduction analysis and cell type annotation were performed on scRNA-Seq data of human retinal organoids. Then, three cell populations during human retinal organoids differentiation were identified: iPSCs, neurons, and astrocytes. iPSCs are pluripotent stem cells obtained by reprogramming terminally differentiated somatic cells by artificially introducing specific transcription factors, which have similar morphology and function to embryonic stem cells [23]. The application prospect of these cells is broad in regenerative medicine, new drug development, and disease modeling [24, 25]. In the research of retinal diseases, Mandai et al. used iPSC generated in skin fibroblasts of a patient with neovascular age-related macular degeneration and induced iPSC to differentiate into retinal pigment epithelial (RPE). After one year of transplantation for the RPE sheet, the sheet structure is complete, and the optimal corrected visual acuity does not change [26]. Astrocytes are one of the 3 types of glial cells contained in human

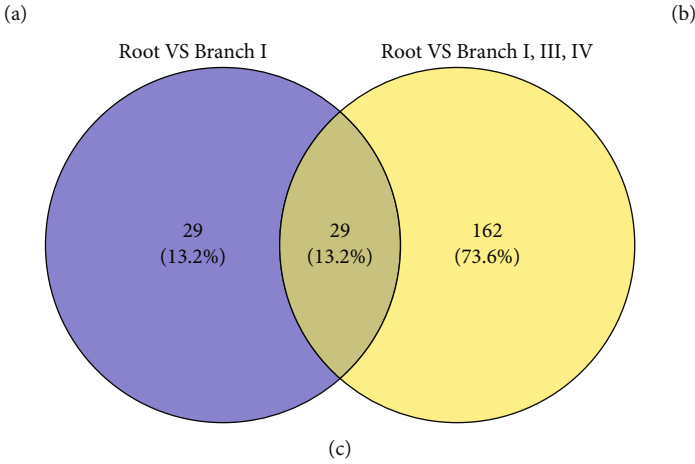
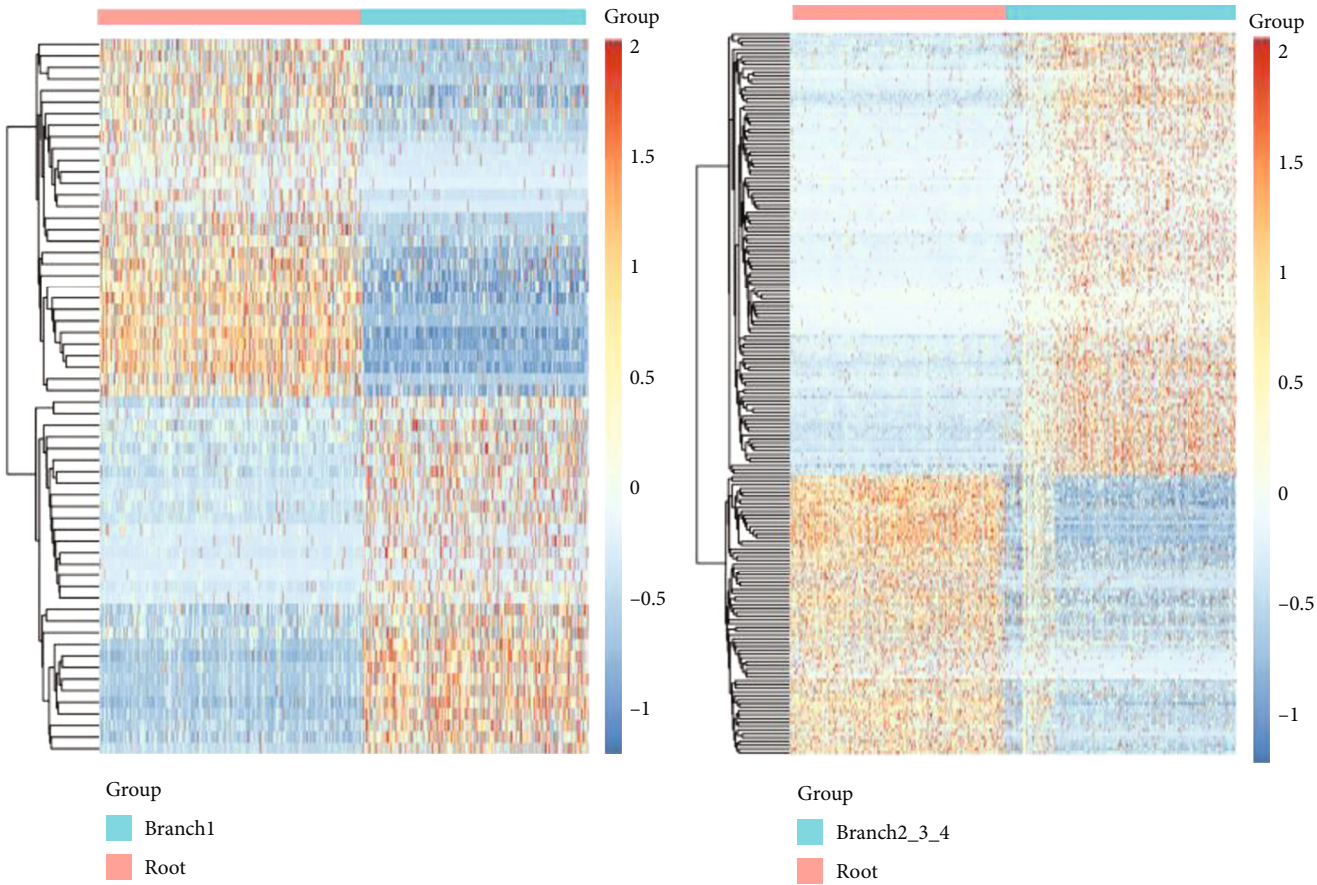
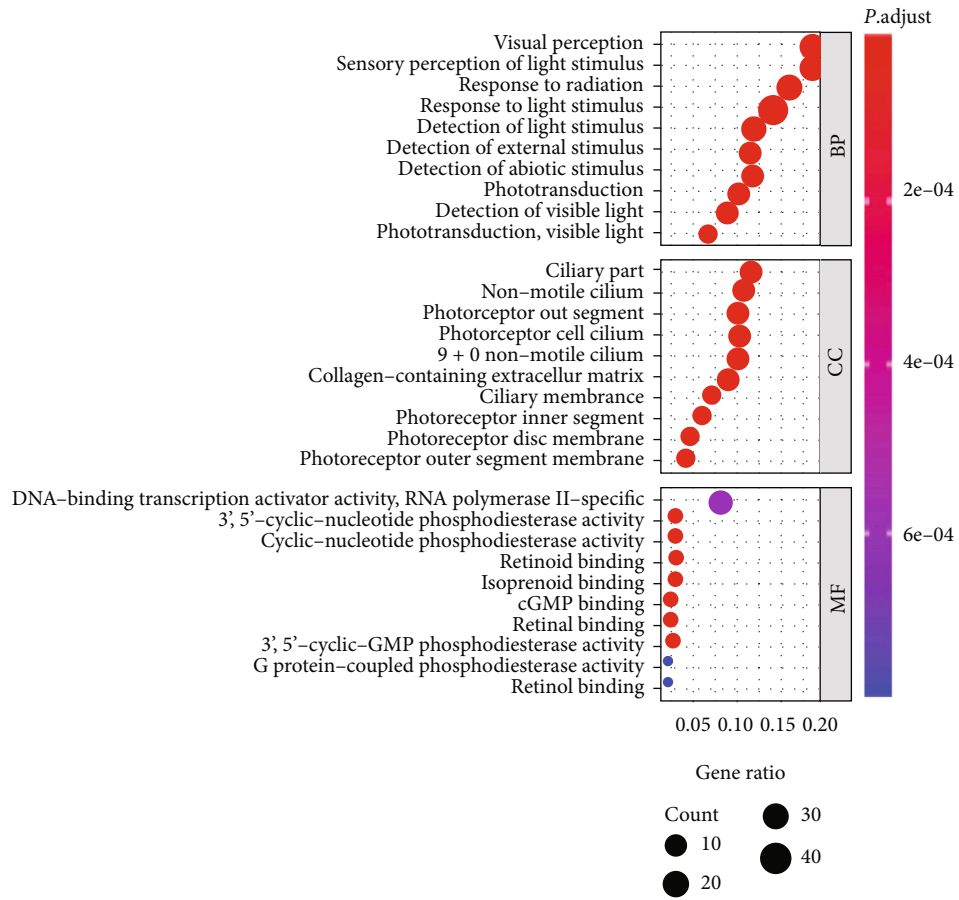


FIGURE 4: Continued.



(d)

FIGURE 4: Continued.

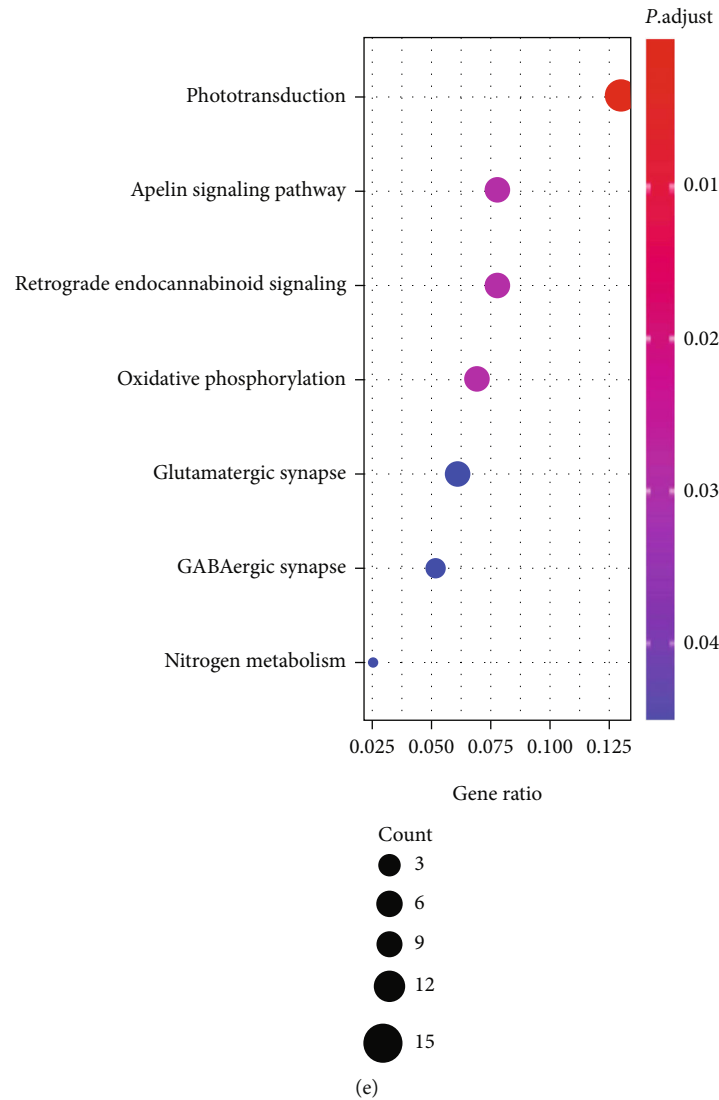
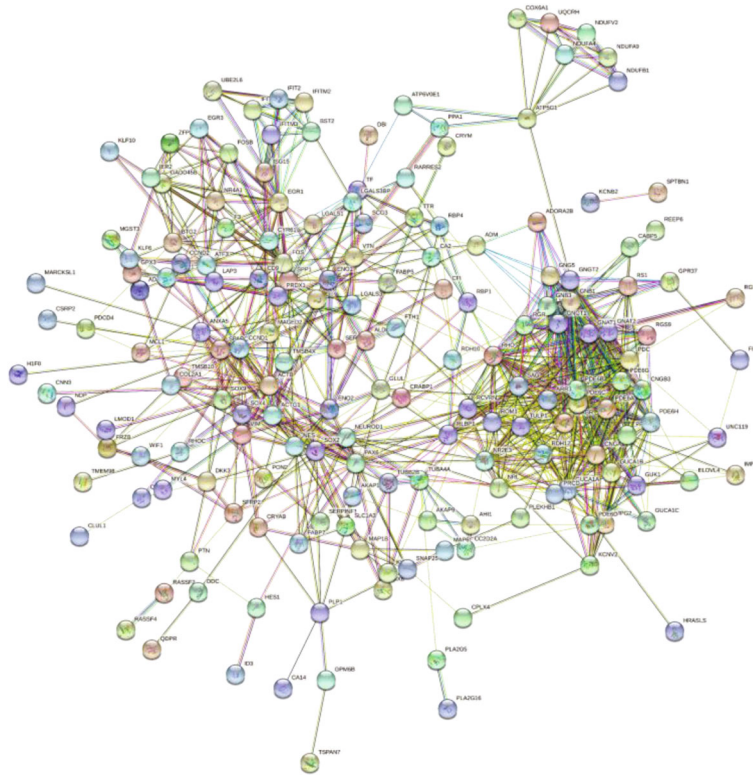


FIGURE 4: Identification of RDRGs and enrichment analyses. (a) Heat map of RDRGs obtained from differential analysis on root cells and branch I cells. (b) Heat map of RDRGs obtained from differential analysis on root cells and branch II, branch III, and branch IV cells. (c) Venn diagram of the two groups of RDRGs. 220 RDRGs are obtained by taking the union. (d) GO functional enrichment result of RDRGs. (e) KEGG pathway enrichment result of RDRGs.

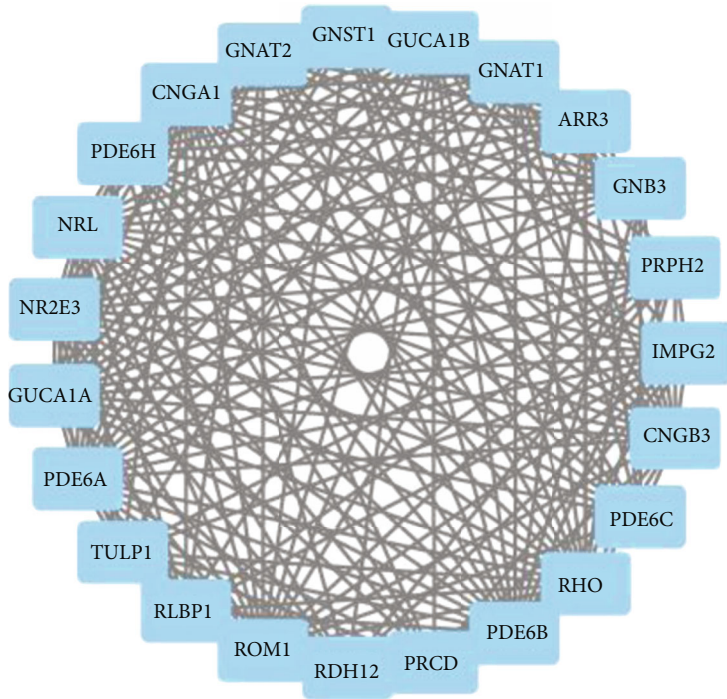
retina, and their presence and distribution are relevant to retinal blood vessels [27, 28]. Astrocytes show morphologic and functional changes in the early and late stages of almost every retinal vascular disease [29]. Guttenplan et al. [30] found that neurotoxic reactive astrocytes could produce neuronal death after retinal injury in a mouse glaucoma model. This suggests that astrocytes are the drivers of RGC death in chronic neurodegenerative eye diseases. Overall, this study used scRNA-Seq technology to identify three cell populations during differentiation of human retinoid organs. Our results are congruent with these reports, leading new directions for treatments of retinal diseases.

Trajectory analysis discovered two main differentiation features during retinal organoids formation: iPSCs differentiated into neurons and astrocytes. RDRGs were identified by gene differential expression analysis and then were subjected to enrichment analyses. GO functional annotation indicated

that RDRGs were significantly enriched in biological functions relevant to the vision and visual development. KEGG pathway enrichment analysis showed that RDRGs were markedly enriched in pathways including phototransduction, apelin signaling pathway, retrograde endocannabinoid signaling, oxidative phosphorylation, and glutamatergic synapse. Chu et al. [31] found that apelin inhibited apoptosis of astrocytes and promoted angiogenesis. Feng et al. [32] also elaborated that apelin signaling pathway was critical to modulate vascular formation both pathologically and physiologically. In addition, oxidative phosphorylation is also associated with the vascular formation in retina-related diseases [33]. Endocannabinoid (eCB) is key to signaling transduction and synaptic function, and anterograde eCB participates in RGC to regulate CB1R-mediated calcium signaling [34]. Armed with the previous studies, it was posited that RDRGs screened in this study may be closely relevant to retinal development and retinal



(a)



(b)

FIGURE 5: Continued.

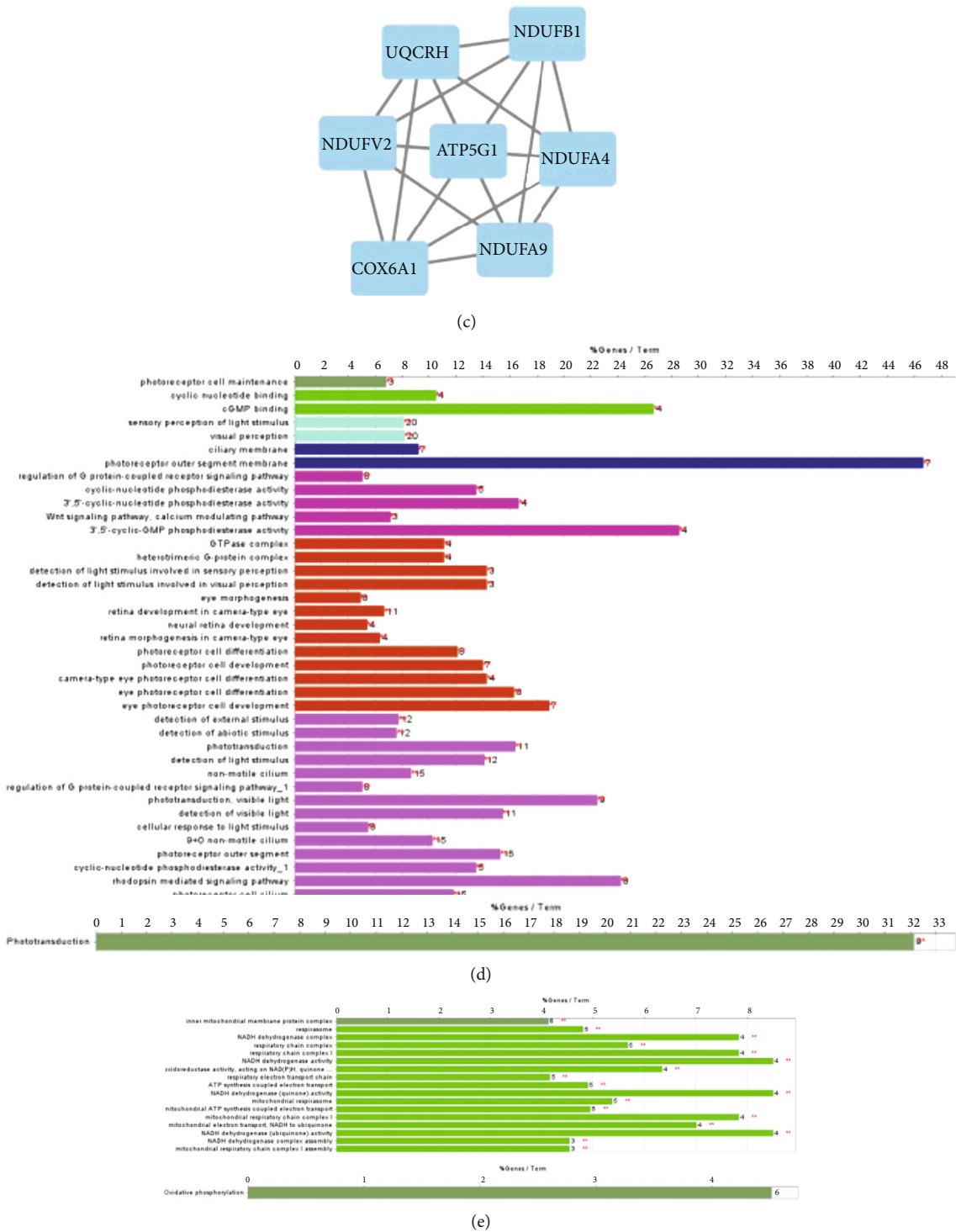


FIGURE 5: Construction of PPI network and functional submodule analysis. (a) PPI network constructed by RDRGs; nodes represent proteins, and connections represent protein-protein interactions predicted by STRING database. (b and c) Cluster 1 and Cluster 2 obtained by functional module analysis, respectively. (d and e) GO and KEGG enrichment analyses of Cluster 1 and Cluster 2, respectively.

neural neuroregulation. 220 RDRGs we identified here may affect iPSCs differentiating into neuron and astrocyte cells. Aberrant expression of these genes may trigger degenerative retina diseases.

PPI network analysis is a commonly used method in bioinformatics analysis, and it systemically analyzed the interac-

tion of massive proteins in the biologic system to reflect the role of genes in biological signaling transmission, substance metabolism, and cell cycle regulation under specific physiological status [35, 36]. Here, PPI network of RDRGs was analyzed through the STRING database, and then, two key submodules were obtained through subnetwork analysis. Enrichment

analysis illuminated that Cluster 1 genes were mainly relevant to biologic functions of visual sensation like photoreceptor cell maintenance, photoreceptor cell differentiation, and photoreceptor cell development. Besides, KEGG enrichment analysis presented that Cluster 1 genes were significantly enriched in phototransduction signaling pathway. Phototransduction is a biological process of photoreceptor cells to transform the optical signal into an electrical signal. Opsins in photoreceptor cells can form photoreceptors with chromophore [37]. NRL and NR2E3 in Cluster 1 are pertinent to degenerative retina diseases. Rod differentiation factor NRL induces NR2E3 level to inhibit cone photoreceptor development [38]. NR2E3 and RHO are the genes enriched in PPI network. We speculated that NRL inhibited the expression of rod-specific genes such as NR2E3, thus inhibiting cone and rod development. A previous study found that light receptor phosphatase (PDE6) was a core rod and cone photoreceptor visual pathway in effecting enzyme, involved in visual signal transmission and amplification [39]. Rod PDE6 is also an essential heterodimer of PDE6A and PDE6B [40]. PDE6A mutation results in progressive retinitis pigmentosa [41]. Mutation of PDE6 family genes arose blockage of visual signal transmission, contributing to ophthalmic diseases, which was supported by KEGG results in this study.

Cluster 2 genes mainly function in mitochondrial energy supplying including NADH dehydrogenase complex and oxidative phosphorylation. Eyes are one of the most active organs in energy metabolism. An increasing amount of evidence supported that mitochondrial dysfunction was a cause of visual dysfunction [42, 43].

On the whole, this study acquired RDRGs through exploring scRNA-Seq data of human retinal organoids, showing that these genes were associated with retinal development, visual sensation, and signaling transduction. PPI network and functional submodule analyses indicated that RDRG-based PPI network encompassed Cluster 1 related to light transduction and Cluster 2 relevant to mitochondrial energy metabolism. The results in this study are hopeful to provide reliable theoretic evidence for studying retinal organoids and retina-related diseases. However, the present study is only based on bioinformatics analysis without any experimental proof. Further biochemical and clinical data are needed to support the results of this study. In addition, we only used one database for analysis, which reduced the reliability of experimental results. In future studies, we will collect more relevant samples and establish our own database to further verify our experimental results and increase the credibility of the results.

Data Availability

The data used to support the findings of this study are included within the article. The data and materials in the current study are available from the corresponding author on reasonable request.

Consent

Consent is not necessary.

Conflicts of Interest

The authors declare that they have no potential conflicts of interest.

Authors' Contributions

All authors contributed to data analysis and drafting and revising the article, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work.

Acknowledgments

This study is supported by the startup funding for specific research project of Dalian Medical University Affiliated Dalian No. 3 People's Hospital (2021ky001).

Supplementary Materials

Data S1: sequencing data matrix of 1346 cells. Table S2: marker genes for each cell cluster. Table S3: retinal organoid cell differentiation-related genes. (*Supplementary Materials*)

References

- [1] M. O'Hara-Wright and A. Gonzalez-Cordero, "Retinal organoids: a window into human retinal development," *Development*, vol. 147, no. 24, p. dev189746, 2020.
- [2] D. Zerti, B. Dorgau, M. Felemban et al., "Developing a simple method to enhance the generation of cone and rod photoreceptors in pluripotent stem cell-derived retinal organoids," *Stem cells (Dayton, Ohio)*, vol. 38, no. 1, pp. 45–51, 2020.
- [3] X. Zhong, C. Gutierrez, T. Xue et al., "Generation of three-dimensional retinal tissue with functional photoreceptors from human iPSCs," *Nature Communications*, vol. 5, no. 1, p. 4047, 2014.
- [4] K. J. Wahlin, J. A. Maruotti, S. R. Sripathi et al., "Photoreceptor outer segment-like structures in long-term 3D retinas from human pluripotent stem cells," *Scientific Reports*, vol. 7, no. 1, p. 766, 2017.
- [5] H. Shen, J. Li, T. Heisler-Taylor et al., "TMEM97 ablation aggravates oxidant-induced retinal degeneration," *Cellular Signalling*, vol. 86, article 110078, 2021.
- [6] M. K. Lin, J. Yang, C. W. Hsu et al., "HTRA1, an age-related macular degeneration protease, processes extracellular matrix proteins EFEMP1 and TSP1," *Aging Cell*, vol. 17, no. 4, article e12710, 2018.
- [7] B. Hwang, J. H. Lee, and D. Bang, "Single-cell RNA sequencing technologies and bioinformatics pipelines," *Experimental & Molecular Medicine*, vol. 50, no. 8, pp. 1–14, 2018.
- [8] Z. Wang, X. Guo, L. Gao, Y. Wang, W. Ma, and B. Xing, "Glioblastoma cell differentiation trajectory predicts the immunotherapy response and overall survival of patients," *Aging (Albany NY)*, vol. 12, no. 18, pp. 18297–18321, 2020.
- [9] A. P. Voigt, K. Mulfaul, N. K. Mullin et al., "Single-cell transcriptomics of the human retinal pigment epithelium and choroid in health and macular degeneration," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 48, pp. 24100–24107, 2019.

- [10] S. Kim, A. Lowe, R. Dharmat et al., "Generation, transcriptome profiling, and functional validation of cone-rich human retinal organoids," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 22, pp. 10824–10833, 2019.
- [11] R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, and A. Regev, "Spatial reconstruction of single-cell gene expression data," *Nature Biotechnology*, vol. 33, no. 5, pp. 495–502, 2015.
- [12] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. 11, 2008.
- [13] X. Qiu, Q. Mao, Y. Tang et al., "Reversed graph embedding resolves complex single-cell trajectories," *Nature Methods*, vol. 14, no. 10, pp. 979–982, 2017.
- [14] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, "clusterProfiler: an R package for comparing biological themes among gene clusters," *Omics: a journal of integrative biology*, vol. 16, no. 5, pp. 284–287, 2012.
- [15] S. Wang, J. Wu, C. Guo et al., "Identification and validation of novel genes in anaplastic thyroid carcinoma via bioinformatics analysis," *Cancer Management and Research*, vol. Volume 12, pp. 9787–9799, 2020.
- [16] S. Chen, D. Yang, C. Lei et al., "Identification of crucial genes in abdominal aortic aneurysm by WGCNA," *PeerJ*, vol. 7, article e7873, 2019.
- [17] L. Zhang, L. Han, X. Wang et al., "Exploring the mechanisms underlying the therapeutic effect of *Salvia miltiorrhiza* in diabetic nephropathy using network pharmacology and molecular docking," *Bioscience Reports*, vol. 41, no. 6, 2021.
- [18] J. Eberwine, J. Y. Sul, T. Bartfai, and J. Kim, "The promise of single-cell sequencing," *Nature Methods*, vol. 11, no. 1, pp. 25–27, 2014.
- [19] T. K. Olsen and N. Baryawno, "Introduction to single-cell RNA sequencing," *Current Protocols in Molecular Biology*, vol. 122, no. 1, article e57, 2018.
- [20] S. Schafer, S. Viswanathan, A. A. Widjaja et al., "IL-11 is a crucial determinant of cardiovascular fibrosis," *Nature*, vol. 552, no. 7683, pp. 110–115, 2017.
- [21] E. Der, S. Ranabothu, H. Suryawanshi et al., "Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis," *JCI Insight*, vol. 2, no. 9, 2017.
- [22] H. Guo, L. Tian, J. Z. Zhang et al., "Single-cell RNA sequencing of human embryonic stem cell differentiation delineates adverse effects of nicotine on embryonic development," *Stem Cell Reports*, vol. 12, no. 4, pp. 772–786, 2019.
- [23] M. Bordoni, F. Rey, V. Fantini et al., "From neuronal differentiation of iPSCs to 3D neuro-organoids: modelling and therapy of neurodegenerative diseases," *International Journal of Molecular Sciences*, vol. 19, no. 12, p. 3972, 2018.
- [24] R. G. Rowe and G. Q. Daley, "Induced pluripotent stem cells in disease modelling and drug discovery," *Nature Reviews. Genetics*, vol. 20, no. 7, pp. 377–388, 2019.
- [25] P. Karagiannis, K. Takahashi, M. Saito et al., "Induced pluripotent stem cells and their use in human models of disease and development," *Physiological Reviews*, vol. 99, no. 1, pp. 79–114, 2019.
- [26] M. Mandai, A. Watanabe, Y. Kurimoto et al., "Autologous induced stem-cell-derived retinal cells for macular degeneration," *The New England Journal of Medicine*, vol. 376, no. 11, pp. 1038–1046, 2017.
- [27] E. Vecino, F. D. Rodriguez, N. Ruzafa, X. Pereiro, and S. C. Sharma, "Glia-neuron interactions in the mammalian retina," *Progress in Retinal and Eye Research*, vol. 51, pp. 1–40, 2016.
- [28] A. Reichenbach and A. Bringmann, "Glia of the human retina," *Glia*, vol. 68, no. 4, pp. 768–796, 2020.
- [29] C. E. Paisley and J. N. Kay, "Seeing stars: development and function of retinal astrocytes," *Developmental Biology*, vol. 478, pp. 144–154, 2021.
- [30] K. A. Guttenplan, B. K. Stafford, R. N. el-Danaf et al., "Neurotoxic reactive astrocytes drive neuronal death after retinal injury," *Cell Reports*, vol. 31, no. 12, article 107776, 2020.
- [31] H. Chu, X. Yang, C. Huang, Z. Gao, Y. Tang, and Q. Dong, "Apelin-13 protects against ischemic blood-brain barrier damage through the effects of aquaporin-4," *Cerebrovascular diseases (Basel, Switzerland)*, vol. 44, no. 1-2, pp. 10–25, 2017.
- [32] J. Feng, L. Chen, Y. Jiang, and Y. Tao, "The role of apelin/APJ in a mouse model of oxygen-induced retinopathy," *Investigative Ophthalmology & Visual Science*, vol. 61, no. 8, p. 47, 2020.
- [33] M. P. Bayona-Bafaluy, O. Esteban, J. Ascaso, J. Montoya, and E. Ruiz-Pesini, "Oxidative phosphorylation inducers fight pathological angiogenesis," *Drug Discovery Today*, vol. 24, no. 9, pp. 1731–1734, 2019.
- [34] A. O. Jo, J. M. Noel, M. Lakk et al., "Mouse retinal ganglion cell signalling is dynamically modulated through parallel anterograde activation of cannabinoid and vanilloid pathways," *The Journal of Physiology*, vol. 595, no. 20, pp. 6499–6516, 2017.
- [35] D. Szklarczyk, A. L. Gable, D. Lyon et al., "STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic Acids Research*, vol. 47, no. D1, pp. D607–D613, 2019.
- [36] J. Liu, S. L. Zhou, S. Y. Li et al., "Eleven genes associated with progression and prognosis of endometrial cancer (EC) identified by comprehensive bioinformatics analysis," *Cancer Cell International*, vol. 19, no. 1, p. 136, 2019.
- [37] J. Hanna, L. A. David, Y. Touahri, T. Fleming, R. A. Screaton, and C. Schuurmans, "Beyond genetics: the role of metabolism in photoreceptor survival, development and repair," *Frontiers in cell and developmental biology*, vol. 10, article 887764, 2022.
- [38] S. M. Moore, D. Skowronska-Krawczyk, and D. L. Chao, "Targeting of the NRL pathway as a therapeutic strategy to treat retinitis pigmentosa," *Journal of Clinical Medicine*, vol. 9, no. 7, p. 2224, 2020.
- [39] F. Chu, D. Hogan, R. Gupta, X. Z. Gao, H. T. Nguyen, and R. H. Cote, "Allosteric regulation of rod photoreceptor phosphodiesterase 6 (PDE6) elucidated by chemical cross-linking and quantitative mass spectrometry," *Journal of Molecular Biology*, vol. 431, no. 19, pp. 3677–3689, 2019.
- [40] H. Muradov, K. K. Boyd, and N. O. Artemyev, "Rod phosphodiesterase-6 PDE6A and PDE6B subunits are enzymatically equivalent," *The Journal of Biological Chemistry*, vol. 285, no. 51, pp. 39828–39834, 2010.
- [41] P. Nair, A. R. Hamzeh, E. M. Malik, D. Oberoi, M. T. al-Ali, and F. Bastaki, "Novel PDE6A mutation in an Emirati patient with retinitis pigmentosa," *Oman J Ophthalmol*, vol. 10, no. 3, pp. 228–231, 2017.
- [42] E. M. Rueda, Johnson JE Jr, A. Giddabasappa et al., "The cellular and compartmental profile of mouse retinal glycolysis, tricarboxylic acid cycle, oxidative phosphorylation, and ~P transferring kinases," *Molecular Vision*, vol. 22, pp. 847–885, 2016.
- [43] E. Jankauskaite, E. Bartnik, and A. Kodron, "Investigating Leber's hereditary optic neuropathy: cell models and future perspectives," *Mitochondrion*, vol. 32, pp. 19–26, 2017.

Research Article

Double-Balanced Loss for Imbalanced Colorectal Lesion Classification

Chang Yu ¹, Wei Sun ¹, Qilin Xiong ¹, Junbo Gao ¹ and Guoqiang Qu ²

¹Information Engineering College, Shanghai Maritime University, Shanghai 201306, China

²Department of Gastroenterology, Eastern Hospital, Shanghai Sixth People's Hospital, Shanghai 201306, China

Correspondence should be addressed to Wei Sun; weisun@shmtu.edu.cn

Received 23 April 2022; Revised 6 July 2022; Accepted 13 July 2022; Published 8 August 2022

Academic Editor: Lin Lu

Copyright © 2022 Chang Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Colorectal cancer has a high incidence rate in all countries around the world, and the survival rate of patients is improved by early detection. With the development of object detection technology based on deep learning, computer-aided diagnosis of colonoscopy medical images becomes a reality, which can effectively reduce the occurrence of missed diagnosis and misdiagnosis. In medical image recognition, the assumption that training samples follow independent identical distribution (IID) is the key to the high accuracy of deep learning. However, the classification of medical images is unbalanced in most cases. This paper proposes a new loss function named the double-balanced loss function for the deep learning model, to improve the impact of datasets on classification accuracy. It introduces the effects of sample size and sample difficulty to the loss calculation and deals with both sample size imbalance and sample difficulty imbalance. And it combines with deep learning to build the medical diagnosis model for colorectal cancer. Experimentally verified by three colorectal white-light endoscopic image datasets, the double-balanced loss function proposed in this paper has better performance on the imbalance classification problem of colorectal medical images.

1. Introduction

A survey shows that malignant tumors have become the first killer of Chinese residents' health by 2020. There is an evidence that the incidence rate and mortality of colorectal cancer are increasing at a particularly significant rate. In 2020, the number of new cases and deaths of colorectal cancer has increased to more than 500,000 and more than 300,000, respectively, which has seriously threatened the health of Chinese residents [1]. Colorectal white-light endoscopy is one of the most widely used of polyp detection and have been extensively used for early colorectal cancer screening. Previous research has shown that missed and misdiagnosed colorectal polyps will increase the possibility of colorectal cancer, and their survival rate is less than 10% [2]. Therefore, the proper classification of colorectal polyps by white-light endoscopic images can be used to assist physicians in the early screening of colorectal cancer.

With the application of artificial intelligence in the field of intelligent medicine, deep learning models are widely used

in the lesion detection and classification of medical images. This paper reviews and compares the main research work of large intestine white-light endoscope image recognition from two aspects: deep learning models and lesion classification methods; the results are shown in Table 1. The existing body of research on colorectal polyps' classification suggests that classification standards are not uniform. Komeda et al. [3, 4] used convolutional neural networks to classify lesions into neoplastic and nonneoplastic and adenomatous and nonadenomatous. Gao et al. [5] used ResNet50 [6] to distinguish whether colonoscopic images contain lesions, then detected specific lesions and divided them into adenoma, cancer and polyp, which detected AP50 up to 0.903. Taş and Yilmaz [7] proposed to perform super-resolution reconstruction of colonoscopic images to obtain high-resolution images, then use Faster R-CNN [8] for detection, which improved the accuracy of model detection by 8% through super-resolution processing. Shin et al. [9] used image enhancement such as rotation and scaling to increase the number of training samples and then used Faster R-CNN

TABLE 1: Deep learning in the diagnosis of colorectal white-light endoscopy.

Study	Date	Model	Classes
Komeda et al. [3]	2016	CNNs	Neoplastic, nonneoplastic
Eduardo et al. [4]	2017	CNNs	Adenoma, nonadenoma
Gao et al. [5]	2020	Mask R-CNN	Cancer, adenoma, polyp
Taş and Yilmaz [7]	2021	Faster R-CNN	Polyp, nonpolyp
Shin et al. [9]	2018	Faster R-CNN	Polyp, nonpolyp
Shin et al. [10]	2019	Faster R-CNN	Polyp, nonpolyp
Nogueira-Rodríguez et al. [11]	2020	YOLO	Polyp, nonpolyp

for detection. Shin et al. [10] proposed a framework for generating synthetic polyp images to augment the training data to detect polyps using Faster R-CNN with a 20% improvement in accuracy. Among them, the literature [7, 9–11] did not classify the lesions but only detected the presence of polyps.

In early colorectal cancer screening, the classification of colorectal polyps is important. Clinically, physicians need to determine whether to perform resection surgery based on the specific category of polyps. At present, there is little research on the classification of colorectal polyps under deep learning, mainly due to the lack of large public datasets for polyp classification [11]. Based on the experience of doctors in the Shanghai Sixth People’s Hospital (East Hospital) in the clinical diagnosis of colon white-light endoscopy, this paper adopts the classification criteria proposed in literature [4]; that is, colon lesions are divided into polyps (polyp), adenomatous polyps (adenoma), and cancerous structures (cancer). Furthermore, the collected colorectal images with lesions were labeled into three categories to carry out a classification study of colorectal white-light endoscopic images.

2. Problem Statement and Background

In recent years, deep learning has achieved good results in the detection and classification of medical image lesions, which can assist doctors in the diagnosis and treatment work to a certain extent. This mainly depends on the powerful learning ability of deep learning models in the image field and the recognition accuracy that can match or even exceed the human eyes. However, the deep learning model is based on the assumption that the samples are independently and equally distributed in the training set, but this premise does not hold in most medical image datasets. A survey showed that about 75% of colorectal polyps are adenomatous polyps [12], which leads to the fact that most of the colorectal white-light endoscopic images collected by doctors from clinical diagnosis are adenoma images. Unbalanced data makes it difficult for deep learning models to be applied in the field of medical-assisted diagnosis. There are two types of approaches designed to address imbalance classification. The first category is the data-level approach, which adjusts the adaptability of the model to the data by changing the data distribution of the training set, such as random oversampling and random undersampling. The second category

is the algorithmic-level approach, which does not change the training set but adjusts the training strategy or model structure, such as reconfiguring the classifier, two-stage training, and improving the loss function. Moreover, methods that combine the two types are available.

2.1. Data-Level Approach. Random undersampling and random oversampling are the classical methods to solve data imbalance; both of them change the original distribution of the dataset. In 2015, Bae and Yoon [13] proposed an upsampling enhancement framework based on data sampling to use rebalanced datasets to learn comprehensive classifiers and use them to detect different types of polyps. The experimental results show that the performance is improved compared with other most advanced detectors. However, random undersampling may cause the samples to lose important feature information during undersampling, while random oversampling increases the risk of overfitting. SMOTE [14] is a more advanced sampling method to overcome this problem, in which new samples are generated by adding data points from the nearest neighbors by interpolation; the limitation of this method is that it cannot overcome the problem of data distribution in unbalanced datasets, which increases the difficulty of classification algorithms to classify them. Another sampling method is class-aware sampling for stochastic gradient descent optimization neural networks [15], whose main idea is to ensure that the sample classes of each batch are evenly distributed in training.

2.2. Algorithm-Level Approach. One algorithm-level approach to address data imbalance is to reconfigure the classifier, which can be implemented in various ways, such as one-class classifier, which uses small classes as outliers and transforms the classification problem into abnormal detection [16]. Two-stage training [17] is first performed on a balanced dataset, and then, the final output layer is fine-tuned on the unbalanced original dataset. Although these methods can alleviate the data imbalance problem to some extent, the improvement of loss function has more attractive features, such as ease of implementation. Cross-entropy loss is widely used in classification problems, but it cannot handle data imbalance. A simple improvement is to use weighted cross-entropy (WCE) according to the number of categories, which is often ineffective in practice.

In 2019, Cui et al. [18] proposed the class-balanced loss framework, using the effective number of samples per class to inversely weight the loss, and this method can effectively solve the class number imbalance problem. Kim et al. [19] proposed complement cross-entropy (CCE) to solve the problem of imbalanced classification and proved that suppressing the probability of the error class helps the deep learning model to learn discriminative information. By neutralizing the confidence of the error sample, the minority class samples get more learning opportunities. In addition, the sample difficulty imbalance can be handled by improving the loss function. In 2016, Shrivastava et al. [20] proposed the OHEM algorithm to screen out difficult samples based on the loss value of the input samples, and then, the screened samples are applied to training in stochastic gradient descent. This method achieves online hard example mining, but it completely discards simple samples, which leads to the model's inability to improve their detection accuracy. In 2017, Lin et al. [21] proposed focal loss to increase the model's focus on difficult samples by dynamically adjusting the loss contribution of difficult samples by introducing a $(1-p)^{\gamma}$ modulation factor in the cross-entropy, but this approach assigns high weights to outliers as difficult samples. The GHM [22] loss function can improve this problem, and its basic idea is to start from the gradient parity of the samples and to dynamically weight the samples according to the proportion of samples accounted for by the gradient parity so that easy samples with small gradients are downweighted, difficult samples with medium gradients are upweighted, and outlier samples with large gradients are downweighted.

3. Materials and Methods

3.1. The Proposed Methods. The most common loss function used in deep learning multiclassification tasks is cross-entropy (CE) loss. The cross-entropy loss gives equal importance to each data instance, which will lead to the network monitoring the classes with fewer number of observations. Therefore, CE loss is inappropriate in classification tasks under class imbalance. This paper proposes a new loss function named double-balanced (DB) loss. We derive it from the perspective of sample size and sample difficulty.

3.1.1. Imbalance of Sample Size. Usually, the method to deal with the unbalanced sample size is to assign a weight to the sample that is inversely proportionate to the class frequency. Since the weights are chosen with a fixed value to the number of samples of each class in the total sample, it does not work well for deep learning when using the batch gradient descent optimization method SGD. This is due to the overlap of features between different samples, as shown in Figure 1. As the number of samples increases, the features carried by new samples already exist in the original samples, and the model does not learn new features from the new samples, so this increase of sample size is ineffective for model training.

To address the problem of sample size imbalance, this paper rebalances the loss by effective samples size [18]. First,

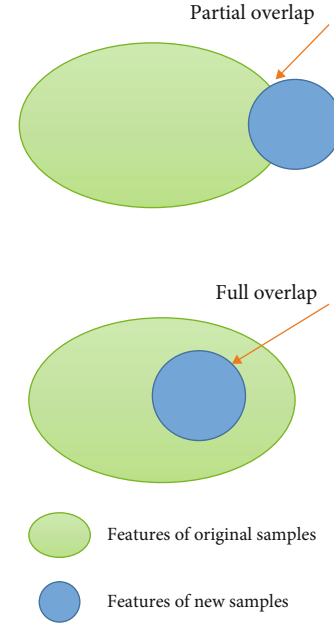


FIGURE 1: The features contained in the new samples may already be partially or fully contained in the original samples.

the effective sample size of each category is calculated using the following equation.

$$E_{n_i} = \frac{1 - \beta^{n_i}}{1 - \beta}, \quad (1)$$

where n_i denotes the true number of samples in each category, β is the hyperparameter that controls the growth rate of E_{n_i} with n_i , $\beta \in (0, 1)$, and here, it is set at $[0.9, 1)$. It can be seen from the above that a larger E_{n_i} indicates a larger effective sample size of category i in the training sample, and its loss proportion should be as small as possible. Therefore, the loss value is inversely related to the effective sample size as follows.

$$e_i = \frac{1}{E_{n_i}} = \frac{1 - \beta}{1 - \beta^{n_i}}. \quad (2)$$

Secondly, considering the large difference in the number of each category, the effective sample size is normalized to get the sample size balance factor α_i for each category:

$$\alpha_i = \frac{e_i}{\sum_{j=1}^k e_j} = \frac{(1 - \beta)/(1 - \beta^{n_i})}{\sum_{j=1}^k ((1 - \beta)/(1 - \beta^{n_j}))}. \quad (3)$$

where k denotes the number of categories. The weight α_i based on the effective sample size of each category is obtained from (3), and in turn, this weight is added to the loss calculation.

$$\text{Loss} = L(p) + \alpha_i L(p) = (1 + \alpha_i) L(p), \quad (4)$$

where $L(p)$ denotes the original loss and p denotes the predicted probability. Here, $\alpha_i L(p)$ is called the loss based on the effective sample size, and the total loss is the original loss plus the loss based on the effective sample size.

3.1.2. Imbalance of Sample Difficulty. The difficult samples refer to less clearly classified borders on the transition region between the foreground and background, while the easy samples refer to background samples that do not overlap with the real target or positive samples that have a high degree of overlap with the real target. We proposed a method to determine and calculate the difficulty of samples. The basic idea is to calculate the distance between the prior probability distribution and the predicted probability distribution of a category and then measure the difficulty of the sample based on the magnitude of the distance.

First, the empirical class frequencies of each category [17] were calculated as their prior probabilities:

$$p_p(c_i) = \frac{(1/n_i)^\rho}{\sum_{j=1}^k (1/n_j)^\rho}, \quad (5)$$

where n_i denotes the number of categories c_i and ρ , as a hyperparameter, denotes the degree of flexibility. After calculating the prior probability of each category, the prior distribution of the training samples is obtained. The cross-entropy function is then used to calculate the distance between the two distributions.

$$H(p_p, p_s) = \sum_{i=1}^k -p_p(c_i) \log p_s(c_i), \quad (6)$$

where p_p denotes the prior given matrix of the prior distribution and p_s denotes the matrix of each category of distribution obtained by the softmax function.

Considering the mutual exclusivity between categories in the classification problem, we only calculate the prior probabilities and predicted probabilities corresponding to the true categories in the two probability distribution matrices, where the true vector corresponding to a single category c_i is

$$\text{true}_{c_i} = [0, 0, \dots, 1, \dots, 0]^T. \quad (7)$$

true_{c_i} is a $k * 1$ matrix. In this matrix, the value of the category c_i is 1 at the corresponding position, and the rest are 0. Next, we do the following calculation.

$$p_p^T * \text{true}_{c_i} = [p_p(c_1), p_p(c_2), \dots, p_p(c_k)] * \begin{bmatrix} 0 \\ 0 \\ \dots \\ 1 \\ \dots \\ 0 \end{bmatrix} = p_p(c_i),$$

$$p_s^T * \text{true}_{c_i} = [p_s(c_1), p_s(c_2), \dots, p_s(c_k)] * \begin{bmatrix} 0 \\ 0 \\ \dots \\ 1 \\ \dots \\ 0 \end{bmatrix} = p_s(c_i). \quad (8)$$

In this way, we extract the individual category prior probability and the predicted probability, and this category must correspond to the true category, suppressing the interference of misinformation. In the following, a purer difficulty weight is obtained by calculating the distance between these two following the method in (6).

$$H(p_p^T * \text{true}_{c_i}, p_s^T * \text{true}_{c_i}) = -p_p(c_i) \log p_s(c_i), \quad (9)$$

where p_p denotes the prior probability for the category c_i and p_s denotes the predicted probability of the softmax output for the category c_i .

3.1.3. Double-Balanced Loss Function. The working principle of the double-balanced loss function is shown in Figure 2. In the classifier, the model obtains the prediction vector *Output* through the full connection layer, with the size of $[3 * 1]$, representing three kinds of lesions. Then, the model uses the softmax function to normalize this group of values, so that the probability vector *Prediction* of each model can be obtained. Before calculating the loss, the model calculates the quantity weight and the difficulty weight according to equations (4) and (9), respectively. Based on these two weights, the double-balanced loss calculates the distance between the predicted value (*Prediction* $[3 * 1]$) and the true value as the final loss value, and this loss value is used as a feedback signal passed to the optimizer, and then, the optimizer implements the neural network weight update through the back propagation algorithm.

In this paper, we use cross-entropy as the original loss, which is responsible for calculating the difference between the true value and the predicted value, and use the one-hot coding in the calculation to obtain a concise expression:

$$\text{CE}_{\text{loss}} = -\log p_s. \quad (10)$$

Bringing (10) into (4) yields the loss based on the effective sample.

$$\text{loss}' = -(1 + \alpha_i) \log p_s. \quad (11)$$

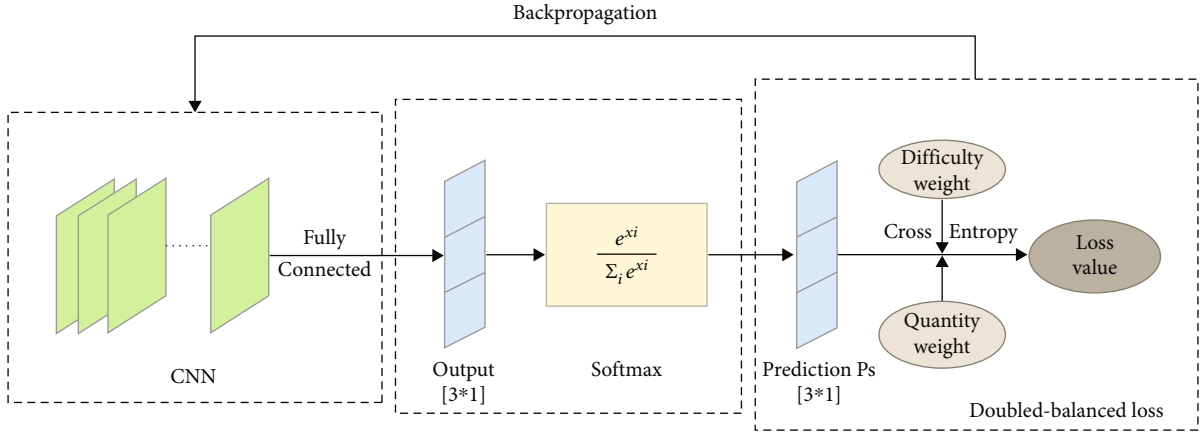


FIGURE 2: Double-balanced loss working principle, balancing loss in terms of both sample size and sample difficulty.

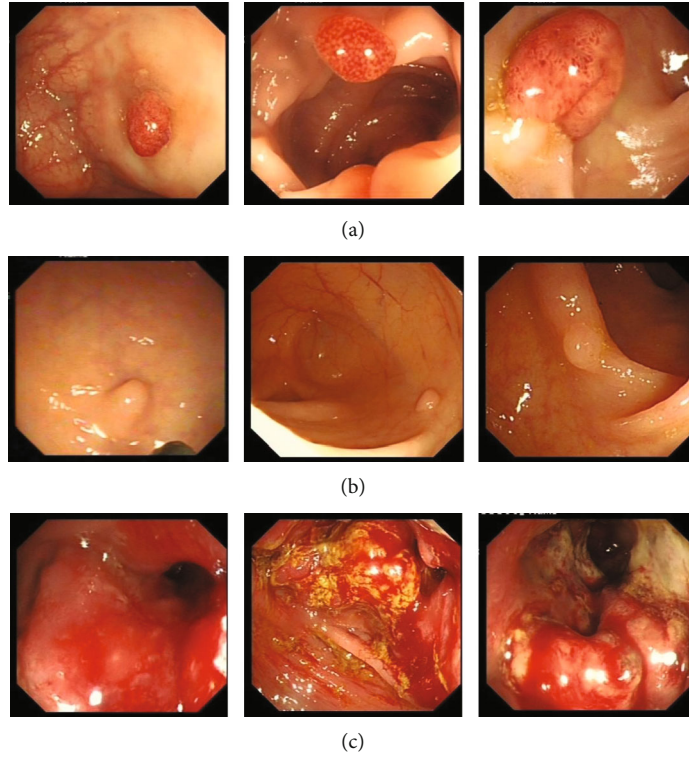


FIGURE 3: White-light endoscopy images of patients' colonoscopy: (a) adenoma, (b) polyp, and (c) cancer.

Then, the formula for calculating the difficulty weights in (9) is added to (11), and the double-balanced loss function can be obtained as follows.

$$DB_{\text{loss}} = \left(1 + \frac{(1-\beta)/(1-\beta^{n_i})}{\sum_{j=1}^k ((1-\beta)/(1-\beta^{n_j}))} \right) \left(\frac{(1/n_i)^\rho}{\sum_{j=1}^k (1/n_j)^\rho} \right) (\log p_s)^2, \quad (12)$$

where β and ρ are hyperparameters, k is the number of categories, and n_i is the number of training samples corresponding to category i .

3.2. Materials. The dataset used in this paper was obtained from white-light endoscopy images of patients' colonoscopy, provided by the Gastrointestinal Endoscopy Center of the East Hospital of Shanghai Sixth People's Hospital. We named this dataset SSPH_WL. The SSPH_WL was collected by doctors under ethical approval. The colorectal lesions were classified into three categories (polyp, adenoma, and cancer) in combination with clinical diagnosis. Sample images of the three categories are shown in Figure 3.

The dataset was collected from June 2015 to September 2019, and a total of 1709 white-light endoscopic images containing lesions were collected and labeled by physicians with more than 5 years' clinical diagnostic experience in

TABLE 2: The datasets used in this experiment: SSPH_WL-I is used for training, SSPH_WL-II, Kvasir, and CVC is used for test.

Dataset	Used for	Description	Resolution ($w \times h$)	Public
SSPH_WL-I	Train	1367 images	375 × 347	No
SSPH_WL-II	Test	342 images	375 × 347	No
Kvasir	Test	428 images	Various resolutions	Yes
CVC	Test	95 images	388 × 284 574 × 500	Yes

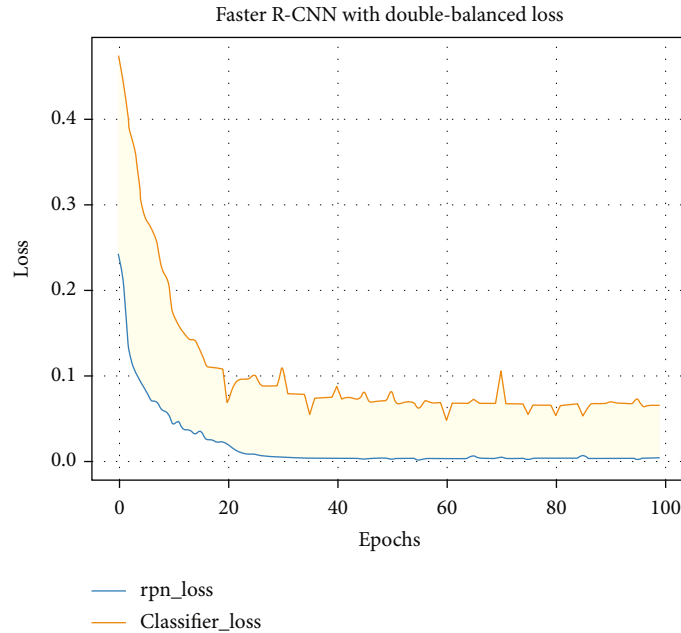


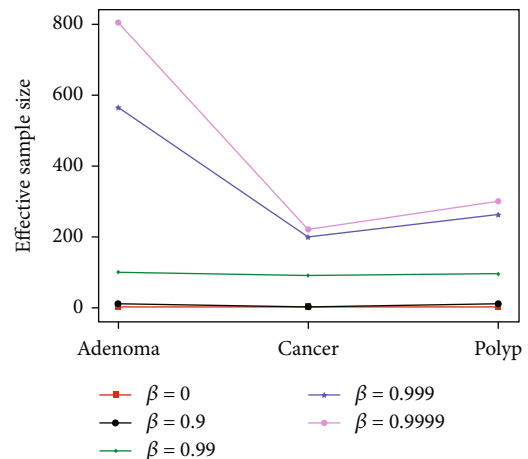
FIGURE 4: Loss variation curves for training: yellow curve represents RPN loss and blue curve represents classifier loss.

gastroenterology endoscopy. Among them, there were 1048 adenomatous cases, 381 polypoid cases, and 280 cancerous cases. The SSPH_WL was divided into the training set (SSPH_WL-I) and test set (SSPH_WL-II) according to 8:2.

We also use colorectal white-light endoscopic images from the public datasets CVC_ClinicDB [23], CVC_ColonDB [24], and Kvasir [25] as the test sets to further evaluate the generalization ability of the classification model. 428 colonoscopic images are selected from the Kvasir data, including 180 images of adenoma, 73 images of cancer, and 175 images of polyp. 95 images are selected from CVC_ClinicDB, CVC_ColonDB and our collection of videos, including 36 images of polyp and 40 images of adenoma, and 19 images of cancer from colorectal cancer videos, named CVC. All of the above datasets were annotated by experienced gastroenterology endoscopy clinicians. The detailed descriptions of these three datasets are shown in Table 2.

4. Experimental Results and Discussion

4.1. Training Strategy and Evaluation Metrics. Deep learning-based object detection models can be divided into two categories: two stage and one stage. Due to the complex

FIGURE 5: The effective sample size changes with the hyperparameter β .

background of the gut, the two-stage algorithm performs background filtering first to overcome this problem, so we choose the two-stage object detection model Faster R-CNN [8] as the representative and focus on the classification effect

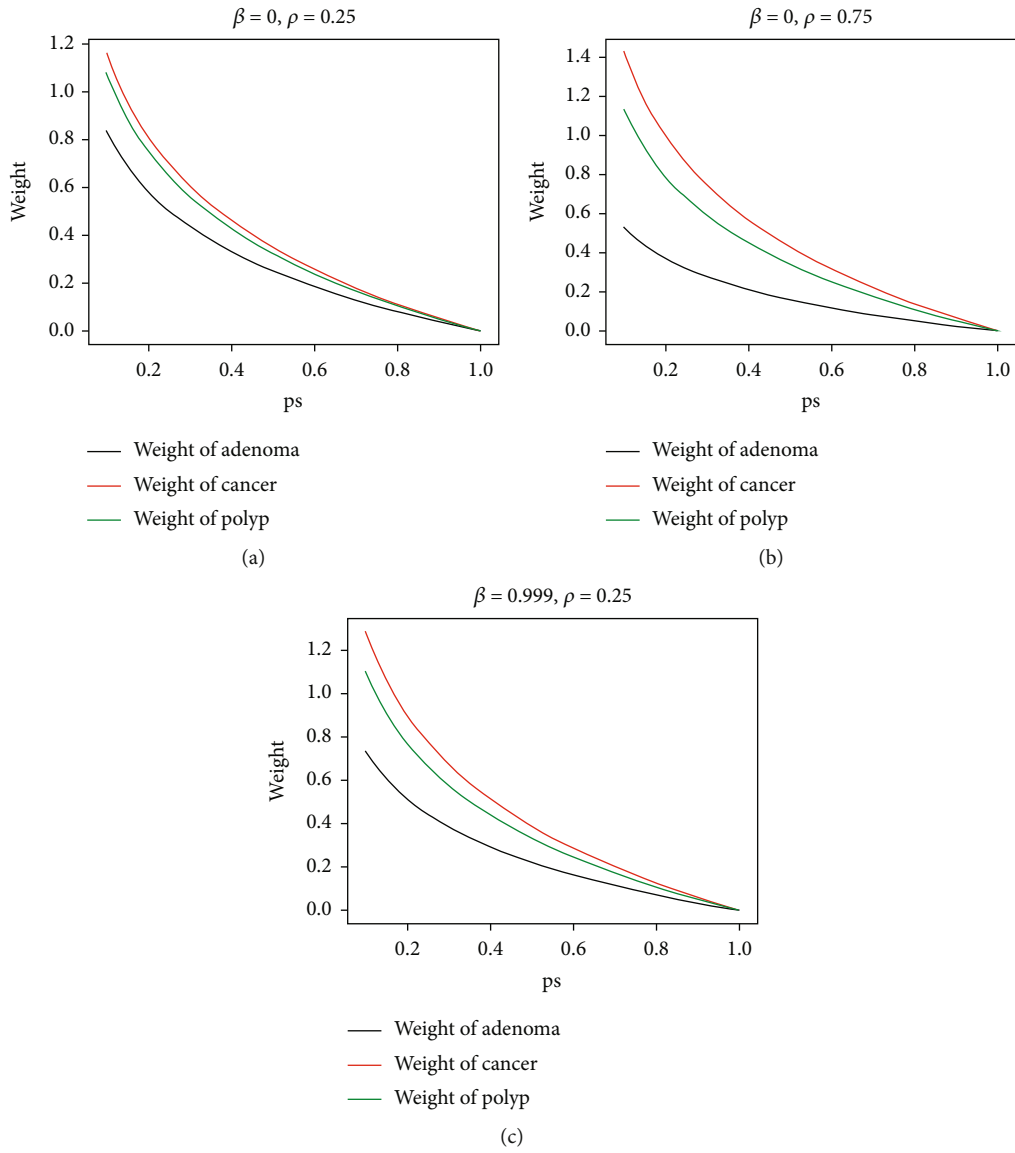


FIGURE 6: Analysis of β and ρ . Variation curve of weight.

TABLE 3: AP and AR of Faster R-CNN with different parameter values.

β	ρ	AP	AP ₅₀	AP ₇₅	AR	AR ₁₀	AR ₁₀₀
0.9	0.25	0.578	0.854	0.678	0.669	0.718	0.718
0.99	0.25	0.598	0.858	0.712	0.675	0.731	0.733
0.999	0.25	0.585	0.858	0.696	0.672	0.710	0.711
0.9999	0.25	0.57	0.833	0.694	0.656	0.707	0.707
0.9	0.5	0.587	0.853	0.695	0.662	0.717	0.717
0.99	0.5	0.582	0.838	0.7	0.671	0.726	0.726
0.999	0.5	0.582	0.843	0.675	0.653	0.715	0.716
0.9999	0.5	0.571	0.826	0.679	0.661	0.709	0.709
0.9	0.75	0.579	0.852	0.663	0.669	0.718	0.718
0.99	0.75	0.562	0.836	0.627	0.654	0.704	0.704
0.999	0.75	0.584	0.85	0.667	0.675	0.713	0.714
0.9999	0.75	0.597	0.848	0.717	0.674	0.719	0.719

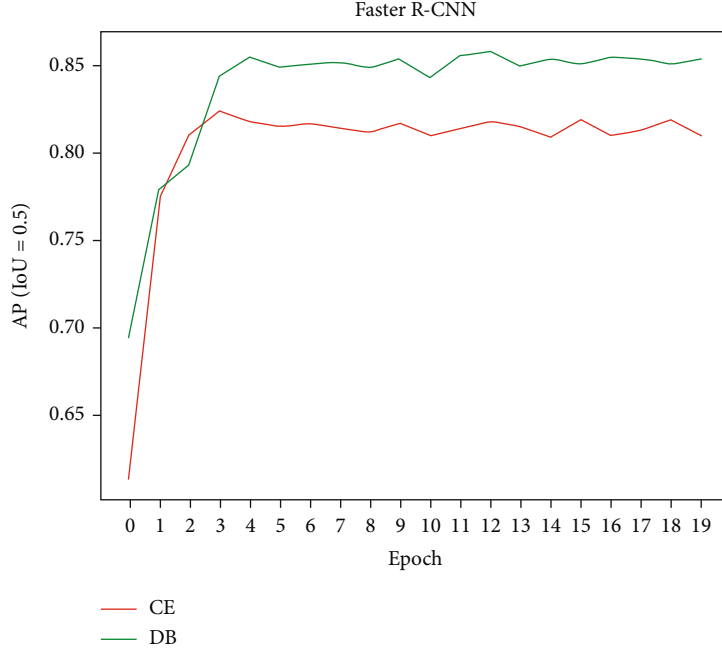


FIGURE 7: The AP50 (IoU = 0.5) changes with epoch on SSPH_WL-II by using CE and DB.

of the double-balanced loss in this model. The momentum is set to 0.9; the initial learning rate (lr) is 0.005 and automatically decreases by two-thirds every three epochs. For other parameters, we use the default values in the model. For the backbone feature extraction network, we choose ResNet50 based on natural image pretraining, which can make the initial performance of the model higher and converge faster by assisting us to train the model through transfer learning. As shown in Figure 4, Since the loss value no longer decreases, the model reaches the optimum at this time and stops training.

This study is a multilabel classification, and the accuracy is not suitable for evaluating the performance of a model for a single category due to the data imbalance. This paper uses AP and AR as model evaluation indicators. AP (Average Precision) is the area under the P-R curve and can be used to measure the performance of the model for a single category, and AR represents the average recall rate of the classification. These two indicators can be expanded as follows.

AP: AP at IoU = 0.50 : 0.05 : 0.95.

AP₅₀: AP at IoU = 0.50.

AP₇₅: AP at IoU = 0.75.

AR: AR given 1 detection per image.

AR₁₀: AR given 10 detections per image.

AR₁₀₀: AR given 100 detections per image.

IoU (Intersection-over-Union) represents the ratio of the intersection and the union of the predicted border and the ground truth bound. In addition, we add the missed detection rate (False Negative Rate, FNR) and the wrong detection rate (False Positive Rate, FPR), which are common evaluation indicators for computer-aided diagnosis. These two indicators are calculated as follows.

$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}}, \quad (13)$$

TABLE 4: Comparison of the improvement effects of Faster R-CNN and SSD on SSPH_WL-II.

Model	Loss	AP ₅₀	AR ₁₀₀
Faster R-CNN	DB	0.858	0.733
Faster R-CNN	CE	0.824	0.711
SSD	DB	0.835	0.701
SSD	CE	0.821	0.689

TABLE 5: Comparison of the improvement effects of Faster R-CNN on three test sets.

Dataset	Loss	AP	AP ₅₀	AP ₇₅	AR	AR ₁₀	AR ₁₀₀
SSPH_WL-II	DB	0.598	0.858	0.712	0.675	0.731	0.733
	CE	0.564	0.824	0.638	0.660	0.711	0.711
Kvasir	DB	0.528	0.833	0.552	0.601	0.661	0.661
	CE	0.475	0.750	0.505	0.594	0.652	0.652
CVC	DB	0.639	0.948	0.745	0.677	0.770	0.770
	CE	0.583	0.874	0.649	0.656	0.698	0.698

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \quad (14)$$

In equations (13) and (14), TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

4.2. *Experiment on Hyperparameter.* The double-balanced loss function has two parameters, where β is responsible for regulating the effective sample size and ρ is responsible for regulating the prior probability of the category. Usually, $\beta \in [0, 1)$; in this paper, the number of training samples of

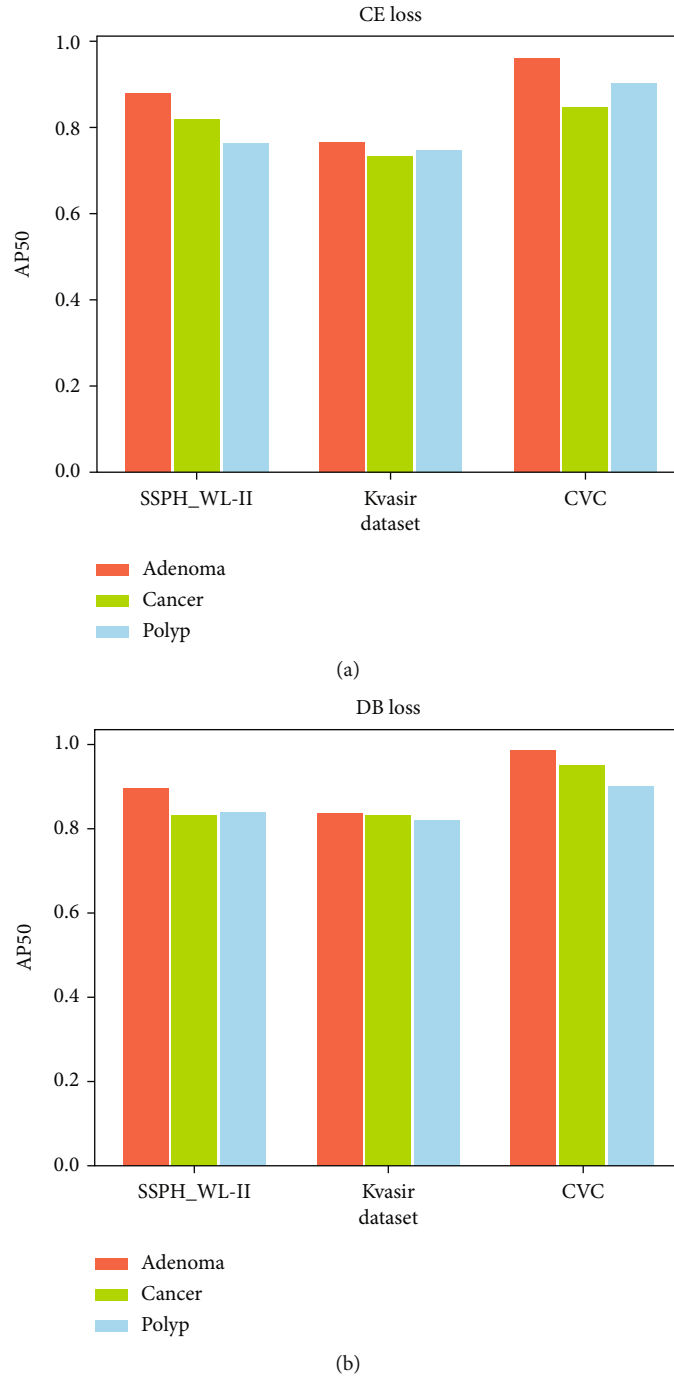


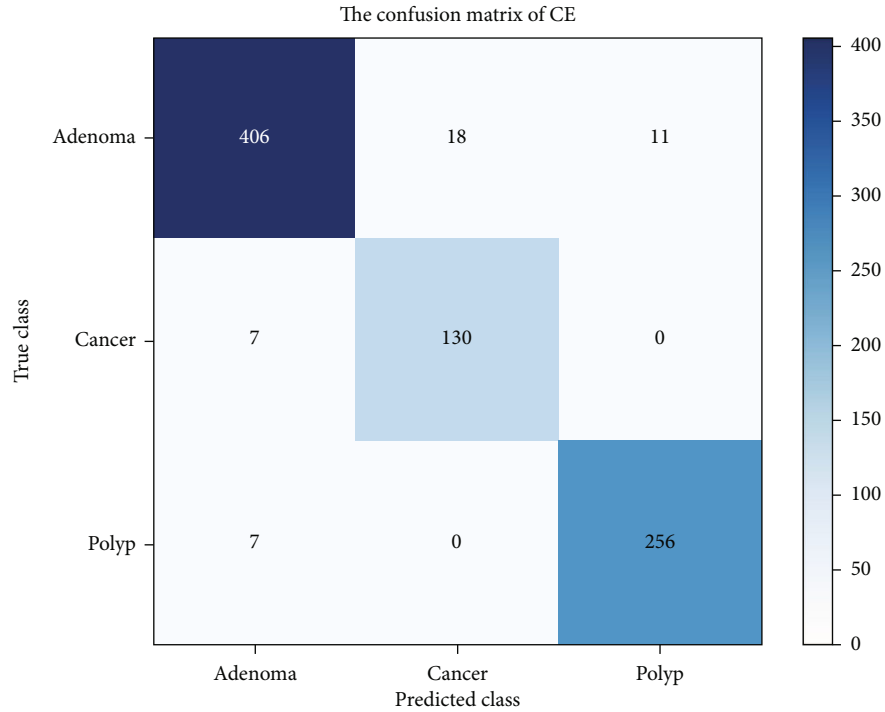
FIGURE 8: Comparison of classification effect by using CE and DB: (a) cross-entropy loss; (b) double-balanced loss.

each category is $\{\text{adenoma} = 838, \text{cancer} = 224, \text{polyp} = 305\}$, and when $\beta \in [0, 0.9)$, the three effective sample sizes calculated from the training samples are almost equal, and the difference of sample sizes cannot be reflected at this time, so $\beta \in [0.9, 1)$ is set. As shown in Figure 5, starting from $\beta = 0.99$, the difference of the effective sample size corresponding to the three categories starts to appear, and as β increases, the effective sample size is closer to the true number. For the parameter ρ , since the number of each category in the training data is fixed, when $\rho = 0$, it means

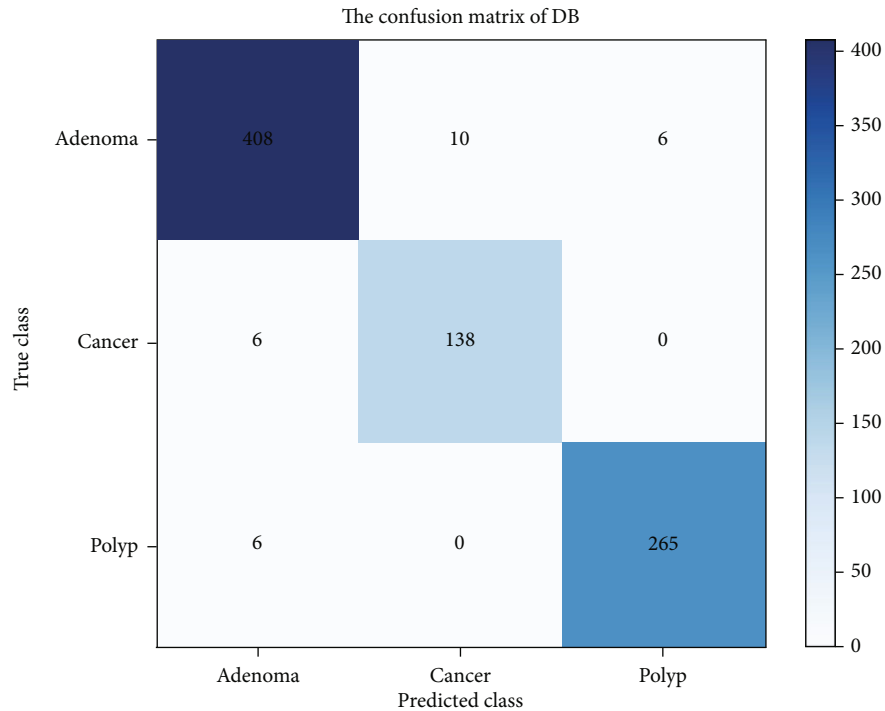
TABLE 6: Wrong detection rates and missed detection rates.

Loss	FPR	FNR
DB	3.24	3.01
CE	4.97	3.47

that each category has the same prior distribution, and when $\rho = 1$, it is equivalent to each category having a prior distribution based on the inverse class frequency.



(a)



(b)

FIGURE 9: Confusion matrix: (a) cross-entropy loss function; (b) double-balanced loss function.

As shown in Figures 6(a)–6(c), the change curve of the weighted value of the loss function with the prediction probability, looking at each line alone, there is a trend: the higher the predicted probability, the smaller the weighted value, which means that the loss of simple samples is suppressed; looking at the three category curves, the weighted value of

adenoma with the largest number of samples is always the smallest, which means that the loss of the multisample category is suppressed. In Figures 6(a) and 6(b), when β is constant, the larger ρ is and the larger the weighted gap between the three categories. In Figures 6(a) and 6(c), when ρ is constant, the larger β is, the larger the weighted gap between the

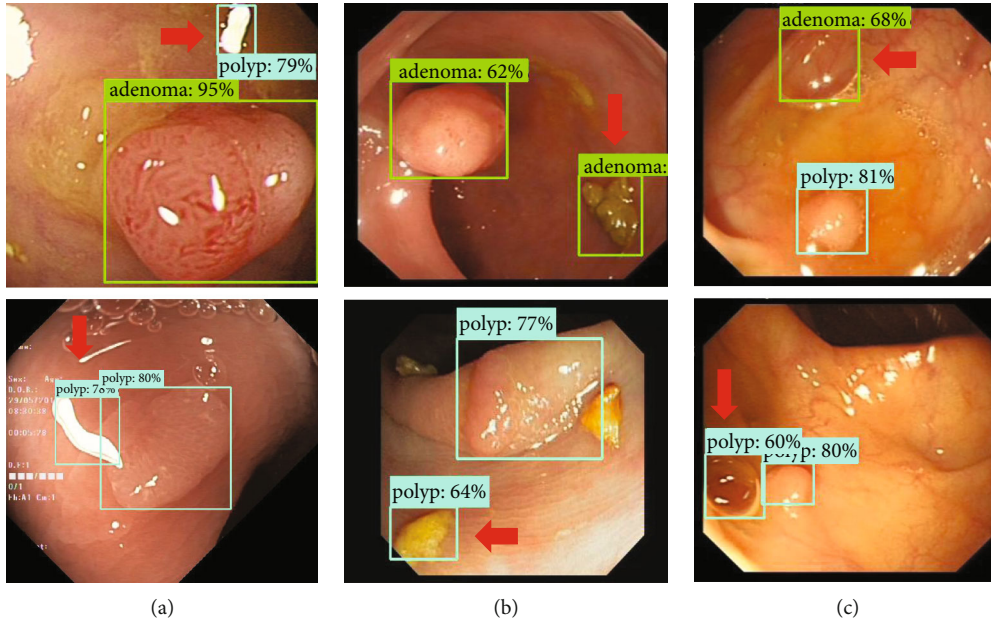


FIGURE 10: Three main categories of disturbances: (a) light spot, (b) foreign matter, and (c) bubble.

three categories will also be. By adjusting hyperparameter β and ρ , the model forms a dynamic weighting mechanism of loss function during training, which can make the model update parameters in a more reasonable way.

In this paper, the comprehensive experiments are conducted for the parameter values, and the corresponding results of different parameter values are shown in Table 3. From the overall performance, the average precision and average recall of detection are almost optimal for $\beta = 0.99$ and $\rho = 0.25$, so we mainly conduct experiments under this parameter value.

4.3. Improved Results on Faster R-CNN. The original loss function used for classification in Faster R-CNN is cross-entropy (CE) loss. Thus, comparing the CE loss, it is clear to see the improvement brought by the DB loss. The training set SSPH_WL-I and test set SSPH_WL-II are from the dataset SSPH_WL, so their data distribution are the same. Figure 7 shows the AP_{50} (IoU = 0.5) changes with epoch on SSPH_WL-II; DB loss can help the model steadily improve the AP_{50} value, where the AP_{50} using the DB loss in Faster R-CNN is stable at around 0.850 and up to 0.858, and the AP_{50} using the CE loss is stable at around 0.820 and up to 0.824. Table 4 compares the detection results of the one-stage model SSD [26]; for the colorectal polyp lesion detection, Faster R-CNN performs better than SSD, and the use of double-balanced loss function can significantly improve the detection performance of Faster R-CNN and SSD.

Table 5 shows a comparison of the results on SSPH_WL-II, Kvasir, and CVC. In these three test sets, DB loss achieves a great lead in all metrics. Kvasir and CVC are used as additional test sets to verify the generalization of DB loss, and the results show that Faster R-CNN trained based on DB loss has a good generalization ability.

In our study, we focus not only on the overall classification level of the model but also on the model's ability to distinguish the three categories. Figure 8 shows the comparison of the classification effect (AP_{50}) of the model on the three test sets. On SSPH_WL-II, the model has the worst classification effect on polyp, and after improvement, the classification precision of polyp is improved the most. On Kvasir and CVC, the model has the worst classification effect on cancer, and after improvement, the classification precision of cancer is improved the most. This indicates that the model can focus on the samples that are not good at classification after improvement, and the classification precision of the three categories is balanced in this way.

We use three test sets to further evaluate the classification ability of the model in terms of the wrong detection rate (FPR) and the missed detection rate (FNR). The wrong detection refers to the model that can locate the lesions but cannot classify them correctly. This is caused by the insufficient classification ability of the model. The missed detection refers to lesions that are not detected by the model, which are filtered out mainly because the model does not locate the lesion or the classification confidence is lower than the set threshold. As shown in Table 6, the FPR and FNR have decreased by using DB loss, indicating that the classification ability of the model has been improved.

The confusion matrices are shown in Figure 9. For the wrong detection, in the original model, 18 adenomas are wrongly classified as cancer, 7 cancers are wrongly classified as adenoma, 11 adenomas are wrongly classified as polyp, and 7 polyps are wrongly classified as adenoma; after using the double-balanced loss, 10 adenomas are wrongly classified as cancer, 6 cancers are wrongly classified as adenoma, 6 adenomas are wrongly classified as polyp, and 6 polyps are wrongly classified as adenoma. It is not difficult to find that the wrong detection never occurs between cancer and

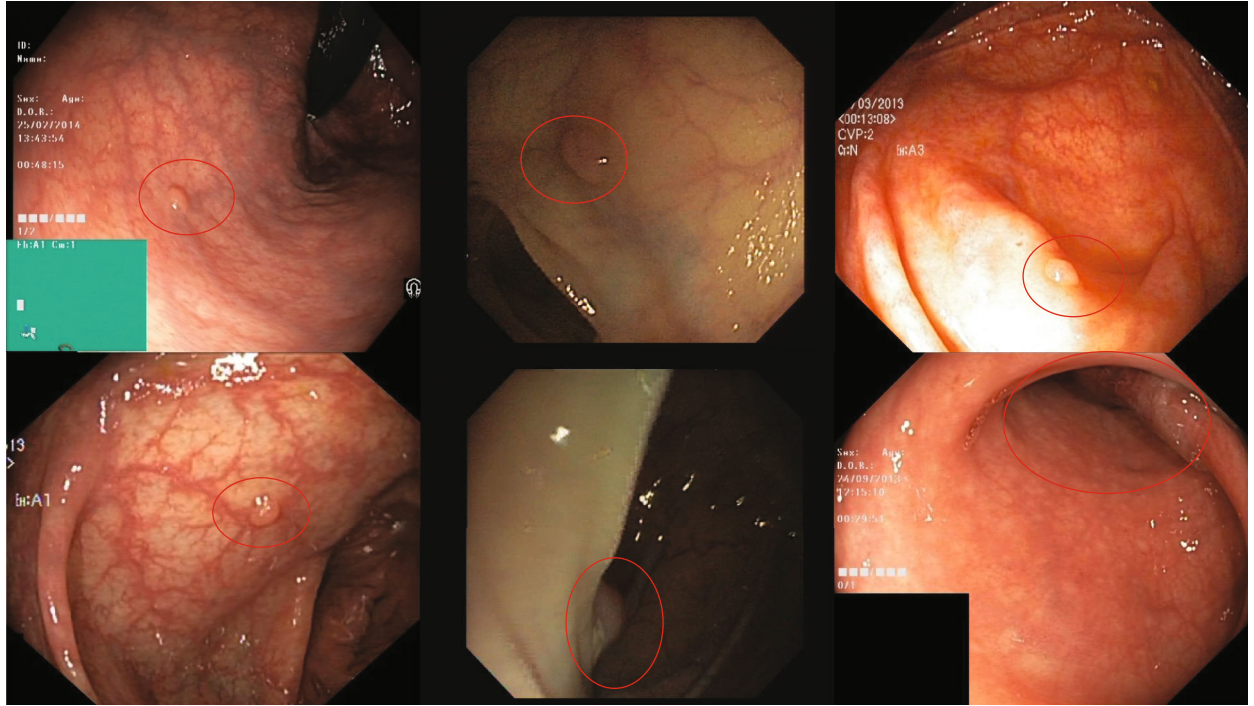


FIGURE 11: Missed detection images of three test datasets.

TABLE 7: Performance of different loss functions on the three test sets.

Dataset	Loss	AP	AP ₅₀	AP ₇₅	AR	AR ₁₀	AR ₁₀₀
SSPH_WL-II	DB	0.598	0.858	0.712	0.675	0.731	0.733
	WCE	0.577	0.842	0.669	0.658	0.724	0.724
	FL	0.571	0.813	0.662	0.667	0.719	0.719
	CB	0.574	0.834	0.681	0.661	0.728	0.726
Kvasir	DB	0.528	0.833	0.552	0.601	0.661	0.661
	WCE	0.458	0.764	0.459	0.581	0.637	0.637
	FL	0.453	0.739	0.459	0.581	0.639	0.639
	CB	0.500	0.788	0.570	0.605	0.661	0.661
CVC	DB	0.639	0.948	0.745	0.677	0.770	0.770
	WCE	0.593	0.885	0.668	0.662	0.705	0.705
	FL	0.597	0.864	0.668	0.663	0.711	0.712
	CB	0.619	0.905	0.676	0.672	0.735	0.735

polyp; this is due to the huge difference in characteristics between polyp and cancer.

In addition, some disturbances in the intestine may also lead to wrong detection of the model, as shown in Figure 10. There are three main categories of these disturbances: the first is the identification of the reflected light spots in the intestine as lesions, the second is the identification of foreign matter in the intestine as lesions, and the third is the identification of air bubbles in the intestine as lesions.

As for missed images, DB loss reduces the occurrence of missed detection, but there are still missed detections, as shown in Figure 11. Most of these images are characterized

by small target areas, dim light, and occlusion, which may be the main reason for missing model detection.

4.4. Comparison Experiments of Different Loss Functions. We compare the classification effects of weight cross-entropy loss (WCE), multiclassification focal loss (FL), and class-balanced cross-entropy (CB) loss in Faster R-CNN, all of which can be used to solve the imbalance problem.

As shown in Table 7, the classification effects of different loss functions on different test sets are compared. On SSPH_WL-II, the double-balanced loss achieves the advantage in all metrics, and on Kvasir, the double-balanced loss and

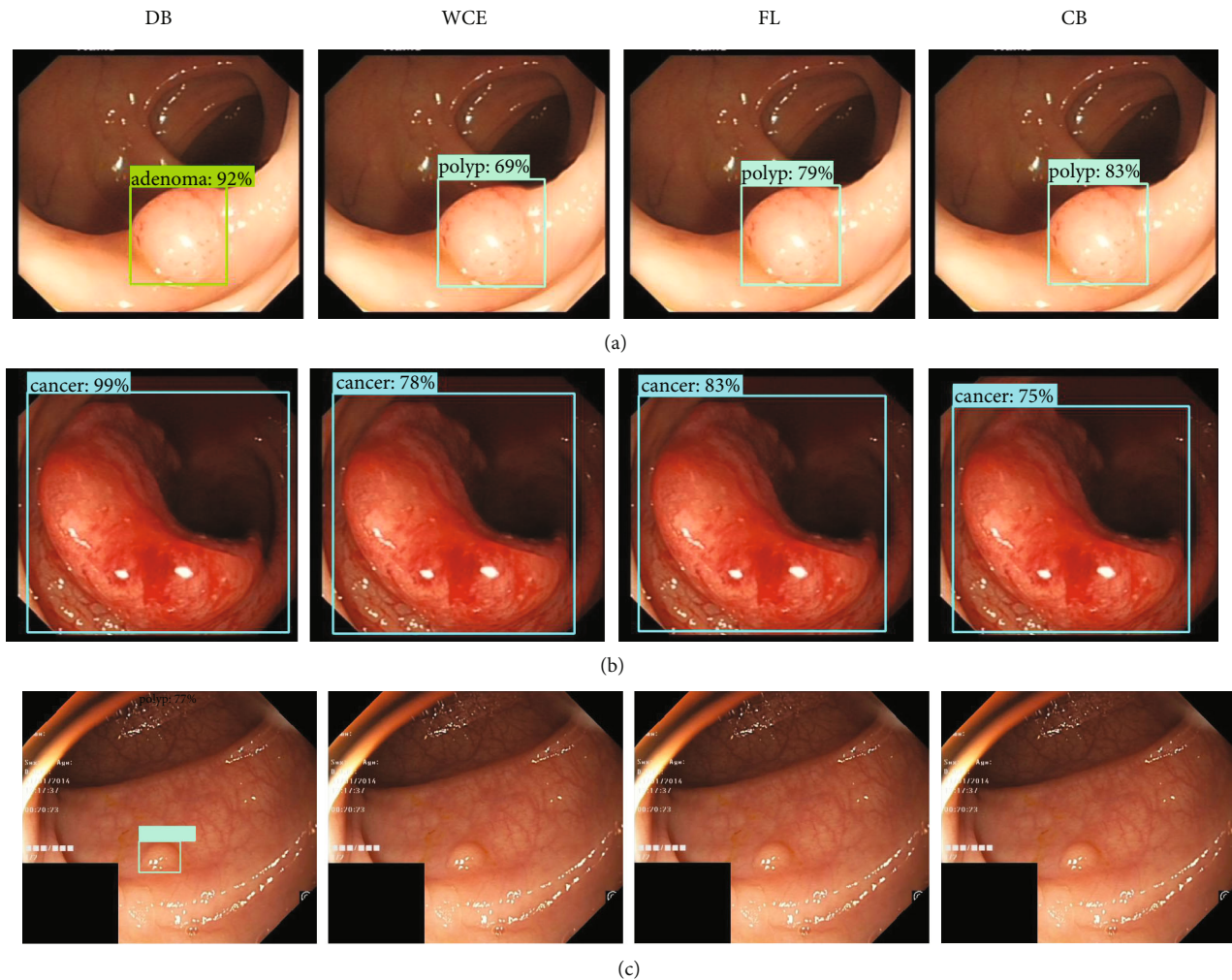


FIGURE 12: Comparison of different loss function detection results. True category is (a) adenoma, (b) cancer, and (c) polyp.

the class-balanced cross-entropy loss perform the best. On CVC, these four loss functions achieve better results, and the double-balanced loss function comes out ahead in all metrics. Experiments show that the double-balanced loss has good generalization ability on different test sets and is able to solve the data imbalance problem better than other loss functions.

Comparing the classification effects of weight cross-entropy (WCE), multiclassification focal loss (FL), and class-balanced cross-entropy (CB) loss functions in Faster R-CNN, as shown in Figure 12, we can see that the double-balanced loss function proposed in this paper has the following advantages: in example (a), the classification is more accurate; in example (b), the classification confidence is higher; and in example (c), the classification performance is better for small targets.

5. Conclusions

To address the imbalance problem in medical image classification, this paper proposes a new loss function, namely, the double-balanced loss. This loss function improves the classification ability of the model for this part of samples by

increasing the focus on fewer sample categories and difficult samples during model training. In this paper, we mainly achieve the double-balanced loss function in Faster R-CNN, and after three test set validations, the model achieves the best detection effect at $\text{IoU} = 0.5$, when the AP values are improved by 3.4%, 8.3%, and 2.9%, indicating that the double-balanced loss function achieves the expected effect on the classification of colorectal white-light endoscopic images. However, there are various types of medical images, and we will verify the effectiveness of the double-balanced loss function on other imbalanced datasets and further promote the double-balanced loss function in the next work.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no competing interests.

Acknowledgments

This research was funded by the Shanghai University of Medicine and Health Sciences Seed Foundation in China, grant number SFP-18-22-14-006.

References

- [1] Y. Li, "Colorectal cancer should not be feared proactive screening can prevent and treat it," *Chinese Medical Information Herald*, vol. 36, no. 8, p. 1, 2021.
- [2] D. H. Kim et al., "CT colonography versus colonoscopy for the detection of advanced neoplasia," *New England Journal of Medicine*, vol. 358, no. 1, pp. 88–90, 2008.
- [3] Y. Komeda, H. Handa, T. Watanabe et al., "Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: preliminary experience," *Oncology*, vol. 93, Suppl. 1, pp. 30–34, 2017.
- [4] E. Ribeiro, A. Uhl, G. Wimmer, and M. Häfner, "Exploring deep learning and transfer learning for colonic polyp classification," *Computational and Mathematical Methods in Medicine*, vol. 2016, Article ID 6584725, 16 pages, 2016.
- [5] J. Gao, Y. Guo, Y. Sun, and G. Qu, "Application of deep learning for early screening of colorectal precancerous lesions under white light endoscopy," *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 8374317, 8 pages, 2020.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, 2016.
- [7] M. Taş and B. Yılmaz, "Super resolution convolutional neural network based pre-processing for automatic polyp detection in colonoscopy images," *Computers and Electrical Engineering*, vol. 90, no. 12, p. 106959, 2021.
- [8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [9] Y. Shin, H. A. Qadir, L. Aabakken, J. Bergsland, and I. Balasingham, "Automatic colon polyp detection using region based deep CNN and post learning approaches," *IEEE Access*, vol. 6, pp. 40950–40962, 2018.
- [10] Y. Shin, H. A. Qadir, and I. Balasingham, "Abnormal colon polyp image synthesis using conditional adversarial networks for improved detection performance," *IEEE Access*, vol. 6, pp. 56007–56017, 2018.
- [11] A. Nogueira-Rodríguez, R. Domínguez-Carbajales, H. López-Fernández et al., "Deep neural networks approaches for detecting and classifying colorectal polyps," *Neurocomputing*, vol. 423, pp. 721–734, 2021.
- [12] S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, and M. Tzivras, "Computer-aided tumor detection in endoscopic video using color wavelet features," *IEEE Transactions on Information Technology in Biomedicine A Publication of the IEEE Engineering in Medicine & Biology Society*, vol. 7, no. 3, pp. 141–152, 2003.
- [13] S. H. Bae and K. J. Yoon, "Polyp detection via imbalanced learning and discriminative feature learning," *IEEE Transactions on Medical Imaging*, vol. 34, no. 11, pp. 2379–2393, 2015.
- [14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *The Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [15] L. Shen, Z. Lin, and Q. Huang, "Relay backpropagation for effective learning of deep convolutional neural networks," in *European Conference on Computer Vision*, Springer, Cham, 2016.
- [16] H. J. Lee and S. Cho, "The novelty detection approach for different degrees of class imbalance," in *Neural Information Processing*, pp. 21–30, Springer, 2006.
- [17] S. Zhang, Z. Li, S. Yan, X. He, and J. Sun, "Distribution alignment: a unified framework for long-tail visual recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2361–2370, 2021.
- [18] Y. Cui, M. Jia, T. Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9268–9277, Long Beach, CA, USA, 2019.
- [19] Y. Kim, Y. Lee, and M. Jeon, "Imbalanced image classification with complement cross entropy," *Pattern Recognition Letters*, vol. 151, article S016786552100266X, pp. 33–40, 2021.
- [20] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *IEEE Conference on Computer Vision & Pattern Recognition IEEE Computer Society*, pp. 761–769, Las Vegas, NV, USA, 2016.
- [21] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 99, pp. 2999–3007, 2017.
- [22] B. Li, Y. Liu, and X. Wang, "Gradient harmonized single-stage detector," *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 1, pp. 8577–8584, 2019.
- [23] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians," *Computerized Medical Imaging and Graphics*, vol. 43, pp. 99–111, 2015.
- [24] D. Vázquez, J. Bernal, F. J. Sánchez et al., "A benchmark for endoluminal scene segmentation of colonoscopy images," *Journal of Healthcare Engineering*, vol. 2017, Article ID 4037190, 9 pages, 2017.
- [25] D. Jha, P. H. Smedsrud, M. A. Riegler et al., "Kvasir-SEG: a segmented polyp dataset," in *International Conference on Multimedia Modeling*, pp. 451–462, Springer, Cham, 2020.
- [26] W. Liu, D. Anguelov, D. Erhan et al., *SSD: single shot multi box detector*, Springer, Cham, 2016.

Retraction

Retracted: Clinicopathological Features and Prognostic-Related Risk Factors of Gastric Signet Ring Cell Carcinoma: A Meta-Analysis

Computational and Mathematical Methods in Medicine

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

Copyright © 2023 Computational and Mathematical Methods in Medicine. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Y. Guo, Q. Wang, Q. Tian et al., "Clinicopathological Features and Prognostic-Related Risk Factors of Gastric Signet Ring Cell Carcinoma: A Meta-Analysis," *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 3473445, 11 pages, 2022.

Review Article

Clinicopathological Features and Prognostic-Related Risk Factors of Gastric Signet Ring Cell Carcinoma: A Meta-Analysis

Ying Guo ¹, Qian Wang ¹, Qing Tian ², Changwen Bo ¹, Na Li ¹, Sujing Zhang ¹, and Peishun Li ³

¹Department of Oncology, The First Hospital of Hebei Medical University, Shijiazhuang, Hebei 050000, China

²Thoracic Surgery Department, The First Hospital of Hebei Medical University, Shijiazhuang, Hebei 050000, China

³Department of Oncology, Tengzhou Central People's Hospital, Tengzhou, Shandong 277500, China

Correspondence should be addressed to Peishun Li; tzzxhlps@163.com

Received 25 May 2022; Revised 6 July 2022; Accepted 17 July 2022; Published 5 August 2022

Academic Editor: Tao Huang

Copyright © 2022 Ying Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Gastric signet ring cell carcinoma (SRCC) has shown a growth growing trend worldwide, but its clinicopathological features and prognostic-related risk factors have not been systematically studied. This systematic review was devoted to this. **Method.** PubMed, Embase, Cochrane Library, and Web of Science databases were retrieved, and retrospective cohort studies comparing clinicopathological features and related risk factors in SRCC patients were included. **Results.** In SRCC patient population, males were more than females (male, OR = 1.38, 95% CI: 1.20-1.60); N3 patients were more than N0-2 patients (N0-2, OR = 3.19, 95% CI: 1.98-5.15); M1 patients were more than M0 patients (M0, OR = 3.30, 95% CI: 1.88-5.80); patients with tumor > 5 cm were more than those with tumor (≤5 cm, OR = 7.36, 95% CI: 1.33-40.60). Patients with age < 60 years (age ≥ 60 years, OR = 1.03, 95% CI: 1.01-1.05), lymphatic vessel invasion (no, OR = 1.74, 95% CI: 1.03-2.45), T2 (T1, OR = 1.17, 95% CI: 1.07-1.28) and T4 (T1, OR = 2.55, 95% CI: 2.30-2.81) stages, and N1 (N0, OR = 1.73, 95% CI: 1.08-2.38), N2 (N0, OR = 2.24, 95% CI: 1.12-3.36), and N3 (N0, OR = 3.45, 95% CI: 1.58-5.32) stages had higher hazard ratio (HR). **Conclusion.** SRCC may occur frequently in male. Age, lymphatic vessel invasion, TN, and M stage may be risk factors for poor prognoses of SRCC patients.

1. Introduction

Gastric cancer is the second cause of cancer death worldwide, and patients in advanced stage have low survival and high recurrence rate [1–3]. More than 1 million people are newly diagnosed with gastric cancer worldwide each year; in 2018, about 783,000 people died of gastric cancer [4]. In the past few decades, a unique type of gastric cancer, gastric signet ring cell carcinoma (SRCC), has been increasingly developed in Asia, Europe, and the United States, representing 35–45% of new cases of adenocarcinoma [5, 6]. SRCC is a special histologic type among all gastric adenocarcinomas with myxoid changes [7]. SRCC is often diagnosed with lymph node metastasis, distant metastasis, or both [8]. Many patients relapse after radical surgical excision [9]. Currently, the treatment strategy of SRCC is based on the multidisciplinary collaborative concept, but some patients still develop

drug resistance or have relapse and metastasis after treatment [10–12]. Overall, current treatments have limited benefit for the overall survival of patients with SRCC, with a median survival of 12 to 20 months [11, 13–15].

SRCC faces a number of clinical challenges. Endoscopy and pathology are difficult to popularize in early screening, and most patients have progressed by the time they are diagnosed [6]. Preliminary trials of combination therapy for SRCC have produced conflicting results, and there is currently no treatment regimen for SRCC, with most chemotherapy regimens targeting common adenocarcinomas [16, 17]. In addition to these clinical challenges, there is growing recognition that the poor prognosis of SRCC may be closely related to particular biological behaviors [18, 19].

In recent years, important achievements have been made to help us understand the epidemiology, pathology, molecular mechanisms, treatment options, and strategies of SRCC

[20]. Previous works revealed the clinicopathological features of SRCC and their effect on prognostic risk factors, but the conclusions are conflicting [21]. In some reports, SRCC is associated with better outcomes [5, 22–24], while another study found no difference in 5-year survival between patients with SRCC and other types of gastric cancer [25].

Therefore, we conducted a meta-analysis to compare the clinicopathological features of patients with SRCC and to analyze prognostic risk factors, hoping to provide reference for clinical diagnosis and treatment of SRCC.

2. Method

2.1. Literature Retrieval. The methodology of this study strictly followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [26]. Through searching databases of PubMed, Cochrane Library, Embase, and Web of Science, we collected all relevant literature from the inception of each database to October 2021. The keywords searched included “gastric signet ring cell carcinoma”, “clinicopathological features”, and “prognostic risk factors”. Since more keywords would narrow the scope of articles, the specific literature retrieval strategies were as follows: (gastric signet ring cell carcinoma [MeSH Terms]) OR (gastric signet ring cell carcinoma [Title/Abstract]). In the retrieval process, “signet ring cell carcinoma of stomach” was entered as the main retrieval keyword. Two investigators independently completed the literature retrieval, and disputes were solved by consultation with the third investigator.

2.2. Literature Selection. Literature selection was completed by two investigators independently, and disputes were settled through consultation with a third investigator. The inclusion criteria were as follows: (1) SRCC was defined by the classification from World Health Organization (WHO) as the cancer with a more than 50% predominant signet ring cells (SRCs); (2) all patients were diagnosed with SRCC or had a history of SRCC; (3) at least one outcome of the following risk factors was reported, including sex, age, chemotherapy, and TNM stage; and (4) all included studies evaluated the clinicopathological features and prognostic significance of SRC histology in patients with gastric cancer and were designed as retrospective cohort studies. The exclusion criteria were listed as follows: (1) duplicate publications, case reports, reviews, conference abstracts, systematic reviews, case-control studies, meta-analyses, editorials, and letters; (2) studies without enough data to extract the results; (3) studies with only 1-3 references of randomized controlled trials, which was not enough to support the meta-analysis; and (4) studies failed to meet the classification from WHO.

2.3. Data Extraction and Quality Assessment. Data extracted from the literature included author information, year of publication, country, study design, sample size, and patient clinical features (sex, TNM stage, and tumor size), as well as data on comparisons in risk factors related to sex, age, chemotherapy, lymphatic vessel invasion, and TNM stage. Retrospective cohort studies were evaluated using the

Newcastle-Ottawa Scale (NOS), which consists of 3 main components, including participant selection (4 stars), inter-study comparability (2 stars), and outcome assessment (3 stars). The total score was 9, and the literature with the score ≥ 6 was considered as good quality. Disagreements between the two investigators were resolved by the majority opinion after a third investigator evaluated all items involved.

2.4. Statistical Analysis. Stata version 16.0 was used for meta-analysis. When evaluating patients’ clinical features and hazard ratio (HR), for dichotomous data, the odds ratio (OR) and 95% confidence interval (CI) were merged for analysis. I^2 statistics was used to assess the statistical heterogeneity of the included studies. The fixed effect model was first selected for detection. If $p > 0.1$ or $I^2 < 50\%$, it was indicated that there was no heterogeneity, and the fixed effect model was used. If $p < 0.1$ or $I^2 > 50\%$, it was indicated that there was significant heterogeneity, and the random effect model should be used for detection.

3. Results

3.1. Literature Retrieval and Study Selection. Based on the developed retrieval strategy, a total of 1,607 articles were retrieved, and 724 duplicated articles were excluded. By browsing titles and abstracts, 656 articles were discarded. The rest of 68 articles were read in full text, among which 48 articles reported irrelevant data, 8 articles did not have enough data to extract the results we wanted, and finally, 12 articles were included [18, 21, 27–36]. Figure 1 shows the literature screening process.

3.2. Study Features and Quality Assessment. A total of 12 articles involving 15,493 patients were included, all cohort studies. Table 1 shows the characteristics and quality assessment results of all the included literature. The literature quality score of each study was 6 points or above, indicating the high quality of the literature.

3.3. Meta-Analysis of Clinicopathological Features. By investigating the relationship between clinicopathological features and SRCC, we revealed that SRCC mainly occurred in males ($I^2 = 85.0\%$, $p < 0.001$, OR: 1.38, 95% CI: 1.20–1.60, Figure 2(a)). In addition, for TNM stage (Figures 2(b)–2(d)), SRCC patients presented high statistical heterogeneity and mainly occurred in T4 stage ($I^2 = 99.0\%$, $p < 0.001$, OR: 1.45, 95% CI: 0.85–2.47), N3 stage ($I^2 = 98.6\%$, $p < 0.001$, OR: 3.19, 95% CI: 1.98–5.15), and M1 stage ($I^2 = 99.2\%$, $p < 0.001$, OR: 3.30, 95% CI: 1.88–5.80). Meanwhile, more SRCC patients had tumors > 5 cm ($I^2 = 99.5\%$, $p < 0.001$, OR: 7.36, 95% CI: 1.33–40.60, Figure 2(e)). In sum, in the SRCC patient population, there were more male patients and patients at N3 stage, M1 stage, and those with tumors > 5 cm.

3.4. Meta-Analysis of Risk Factors. Thereafter, we analyzed the risk factors affecting overall survival in patients with SRCC. SRCC was revealed to have high statistical

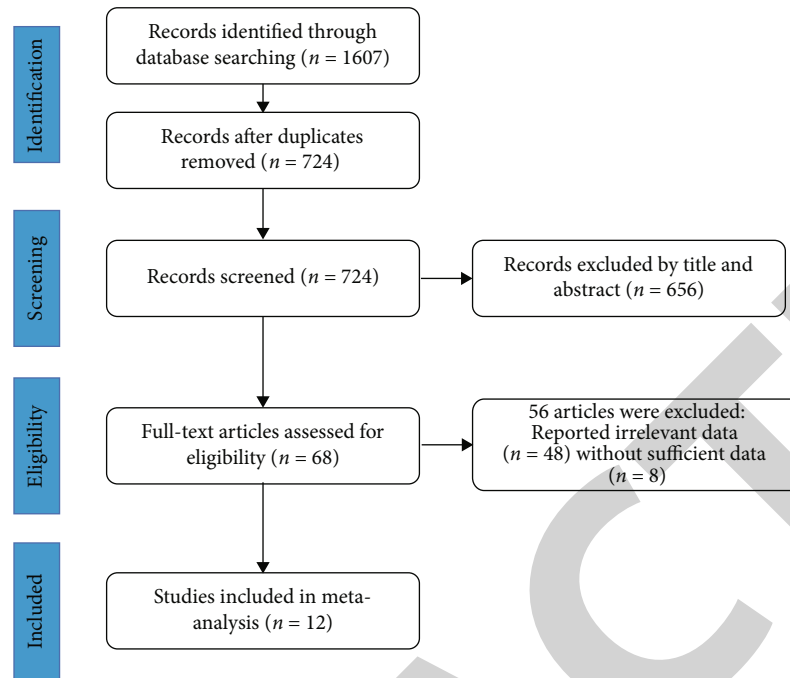


FIGURE 1: Flow chart of literature screening.

TABLE 1: Basic information of the included literature.

Author	Year	Country	Study design	Sample size (n)	Sex		T stage		N stage		M stage		Tumor size		NOS
					Male	Female	T1-3	T4	N0-2	N3	M0	M1	≤5 cm	>5 cm	
Yang	2018	China	Cohort study	375	209	166	116	259	206	169	337	38	NA	NA	7
Zhou	2020	China	Cohort study	403	259	144	176	227	292	111	360	43	227	176	8
Guo	2020	USA	Cohort study	7149	3758	3391	5845	1304	6621	528	4794	2355	NA	NA	9
Shi	2019	USA	Cohort study	4638	2446	2192	2424	1188	3606	626	2647	1991	NA	NA	9
Chen	2018	China	Cohort study	347	181	166	315	32	335	12	341	6	NA	NA	8
Kunisaki	2004	Japan	Cohort study	54	28	26	NA	NA	51	3	NA	NA	NA	NA	9
KYUNG	2011	Korea	Cohort study	41	19	22	NA	NA	NA	NA	NA	NA	9	32	9
Kyoung-Joo	2013	Korea	Cohort study	205	149	56	NA	NA	NA	NA	NA	NA	NA	NA	6
Kao	2018	Taiwan	Cohort study	185	96	89	NA	NA	181	4	NA	NA	NA	NA	6
Li	2020	China	Cohort study	144	96	48	61	83	NA	NA	NA	NA	NA	NA	6
Tang	2020	China	Cohort study	266	182	84	137	129	181	85	NA	NA	187	79	8
Wang	2021	China	Cohort study	1686	879	807	1021	665	1125	561	1436	250	935	751	9

heterogeneity in age ($I^2 = 67.5\%$, $p = 0.009$, Figure 3(a)), chemotherapy ($I^2 = 98.9\%$, $p < 0.001$, Figure 3(c)), lymphatic vessel invasion ($I^2 = 64.1\%$, $p = 0.016$, Figure 3(d)), and T3 ($I^2 = 77.9\%$, $p = 0.004$, Figure 3(f)), T4 ($I^2 = 79.8\%$, $p = 0.002$, Figure 3(g)), and N ($I^2 = 77.5\%$, $p = 0.001$, Figures 3(h)–3(j)) stages. Hence, we introduced a random effect model, and higher HR was found in age < 60 years (HR: 1.03, 95% CI: 1.01–1.05, Figure 3(a)), nonchemotherapy (HR: 0.93, 95% CI: 0.45–1.41, Figure 3(c)), and lymphatic vessel invasion (HR: 1.74, 95% CI: 1.03–2.45, Figure 3(d)). In addition, HR was higher in T3 stage (HR: 3.72, 95% CI: 0.03–7.40, Figure 3(f)) and T4 stage (HR: 2.55, 95% CI: 2.30–2.81, Figure 3(g)) in comparison with

T1 stage. N1 stage (HR: 1.73, 95% CI: 1.08–2.38, Figure 3(h)), N2 stage (HR: 2.24, 95% CI: 1.12–3.36, Figure 3(i)), and N3 stage (HR: 3.45, 95% CI: 1.58–5.32, Figure 3(j)) had higher HRs over N0 stage. However, SRCC showed low statistical heterogeneity in sex ($I^2 = 36.6\%$, $p = 0.137$, Figure 3(b)), T2 stage ($I^2 = 38.0\%$, $p = 0.184$, Figure 3(e)), and M stage ($I^2 = 40.9\%$, $p = 0.149$, Figure 3(k)). Subsequently, HRs were found to be higher in T2 stage (HR: 1.17, 95% CI: 1.07–1.28) and M1 stage (HR: 3.13, 95% CI: 2.94–3.32) with fixed effect model analysis. Accordingly, we suggested that age, lymphatic vessel invasion, and TNM stage might be risk factors affecting the overall survival of SRCC patients.

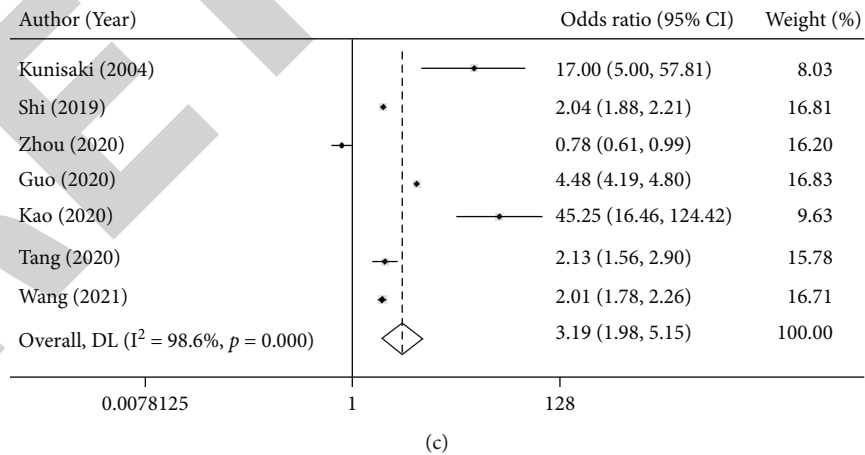
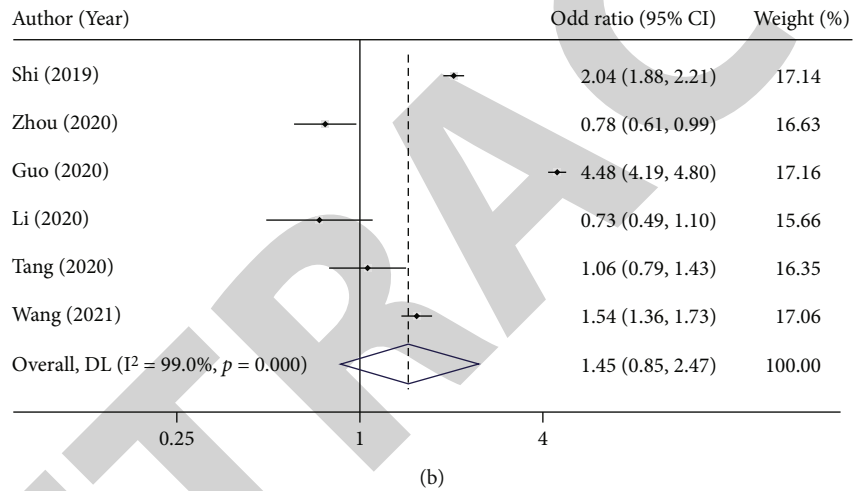
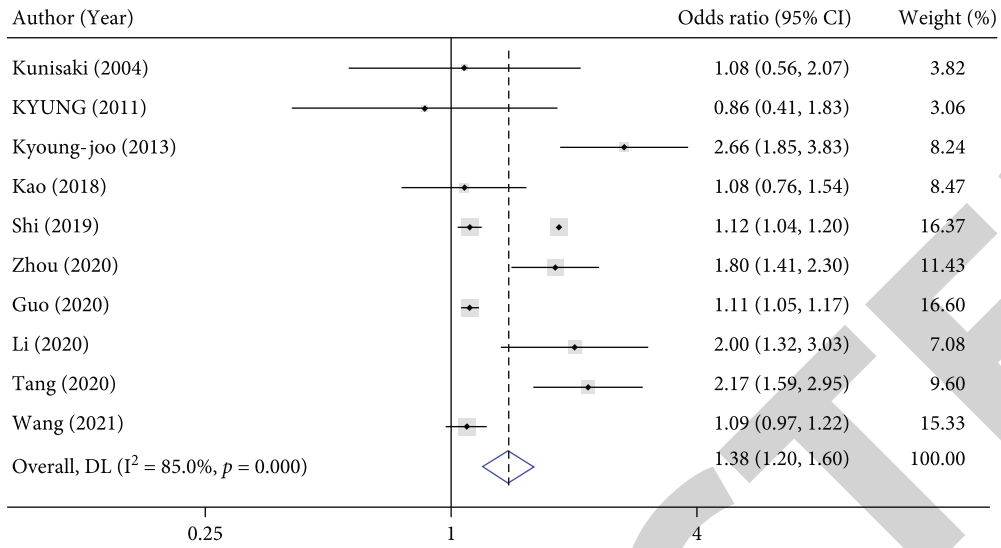


FIGURE 2: Continued.

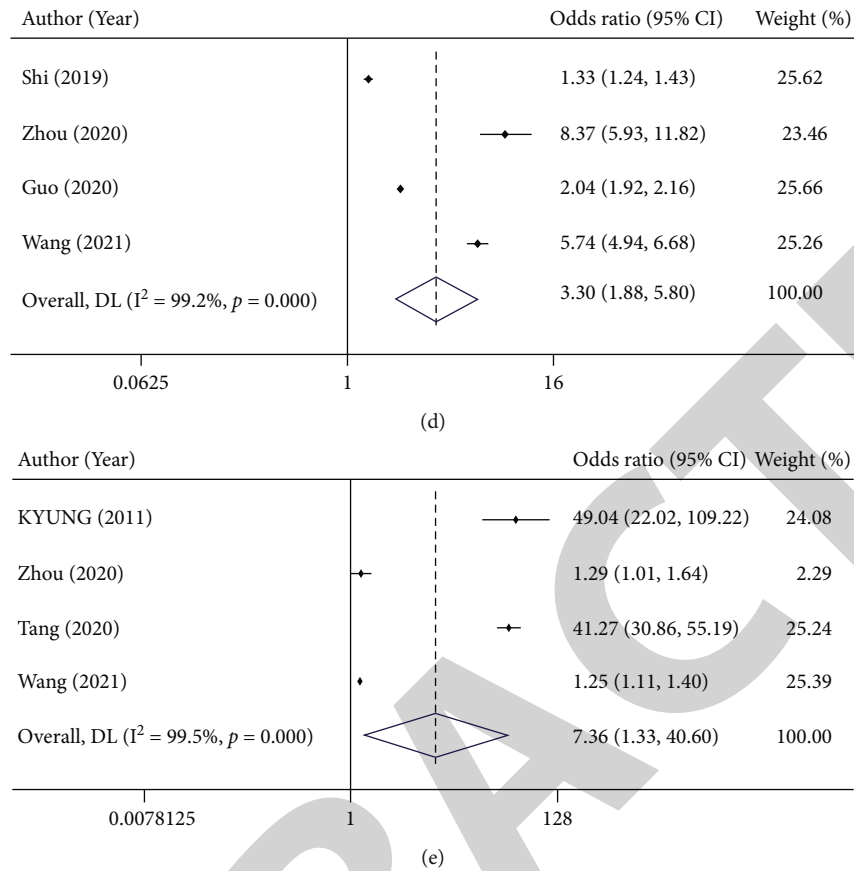


FIGURE 2: Forest plot was applied to assess different clinicopathological characteristics following SRCC and non-SRCC gastric cancer. (a) Sex, (b) T stage, (c) N stage, (d) M stage, and (e) tumor size.

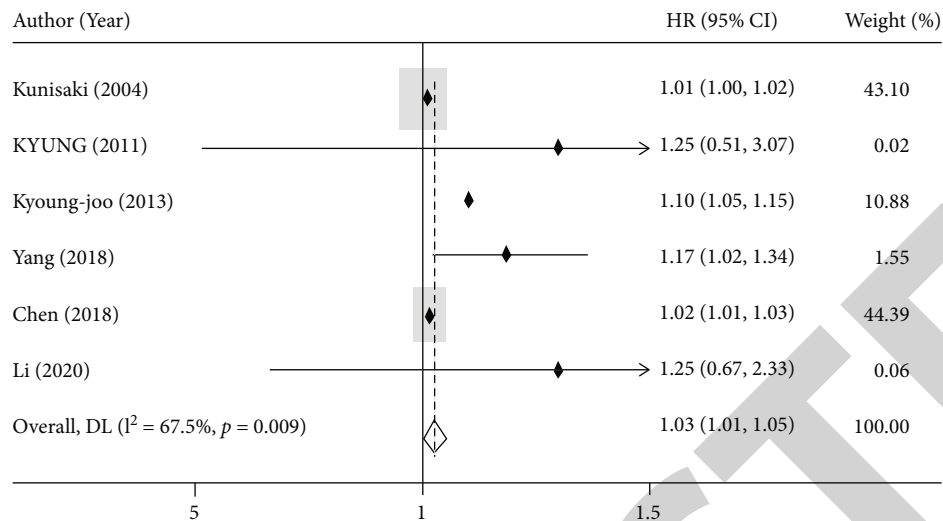
4. Discussion

The incidence of SRCC has been increasing obviously in recent years [5, 6]. There have been many studies on the clinicopathological features of SRCC and their influence on prognostic risk factors, but no consistent conclusions have been drawn [21]. Therefore, we conducted a meta-analysis on the clinicopathological features and prognostic risk factors of patients with SRCC.

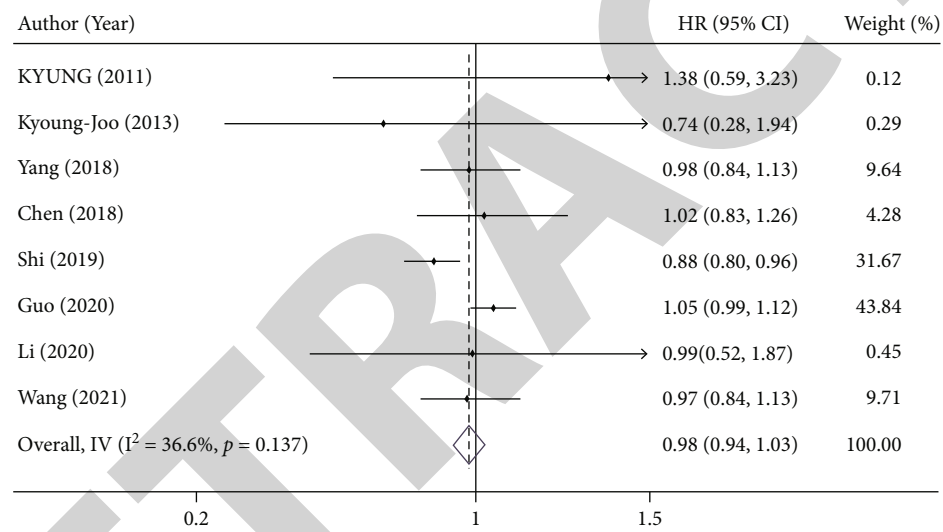
Studies have found that SRCC is a poorly differentiated adenocarcinoma with high malignancy, unique clinical features, and low survival rate [18, 37, 38]. We studied on the clinical features of SRCC patients, and the results showed that there were more males than females in SRCC patients. In respect of N stage, the population of patients in N3 stage was larger than that in N0 to N2 stage. In terms of M stage, the population of patients in M1 stage was larger than that in M0 stage. In a single-center database retrospective analysis, TNM stage significantly influences the poor prognosis of primary SRCC [39]. Besides, higher stage was an independent risk prognostic factor for patients with SRCC [40, 41]. Consistently, by further analysis of risk factors affecting the overall survival in SRCC patients, we found that patients had higher HR at T2 and T4 stages over T0 stage; N1, N2, and N3 stages over N0 stage; and M1 stage over M0 stage. However, in other studies, the correlation between TNM

stage and SRC prognosis remains controversial and may be related to gene mutation and the degree of invasion of SRC [42–44], which requires further analysis by combining multiple causes.

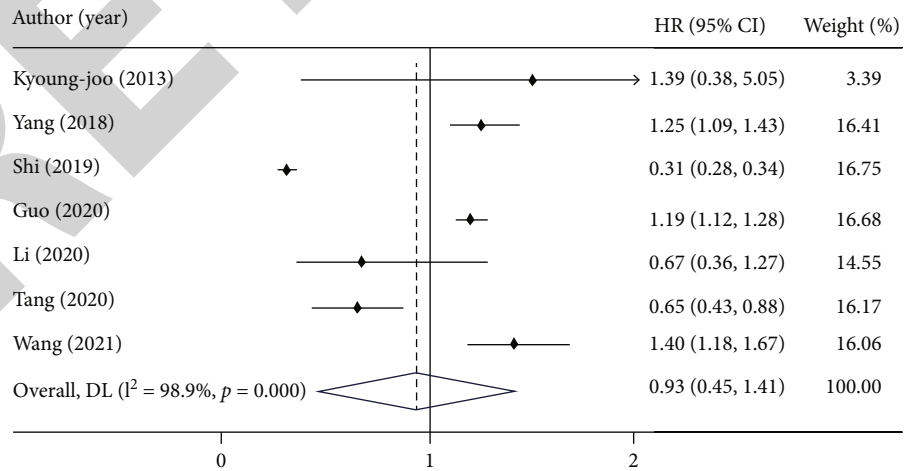
It is well known that SRCC patients with larger tumors have a poorer prognosis, which may be because larger tumors are coupled to greater depth of invasion, worse histological grade, higher risk of lymph node metastasis, and distant metastasis [21, 41, 45]. Our analysis showed that tumor size > 5 cm was more common in SRCC patients than in patients with tumors ≤ 5 cm. Larger tumors are strongly linked to lymph node metastasis and more common in patients with advanced gastric cancer and cause poorer prognoses [41, 46, 47]. Although tumor size was not evaluated as a risk factor for overall survival in patients with SRCC in this study, the results suggested that lymphatic vessel invasion caused a higher HR. The association of tumor size and lymph node metastasis is indicative of the association between larger tumors and poorer prognosis [41], but data collection remains necessary for further analysis. Therefore, less invasive gastric surgeries (such as endoscopic mucosal resection and endoscopic submucosal dissection) are more beneficial for patients with early SRC, and patients with advanced SRCC require more thorough surgical treatment [44]. Nevertheless, these conclusions still need to be treated with caution



(a)

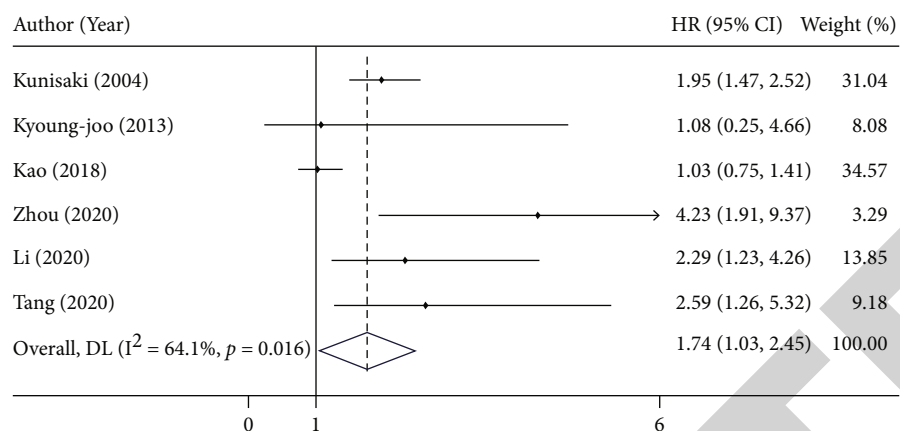


(b)

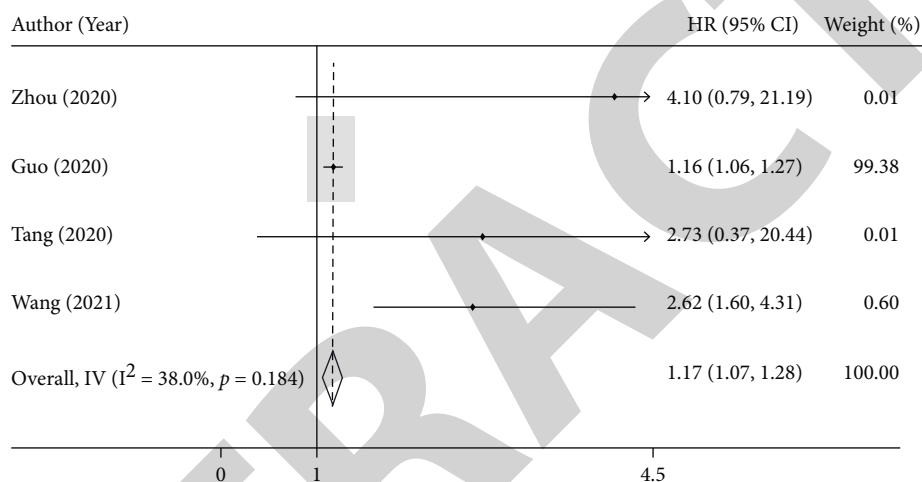


(c)

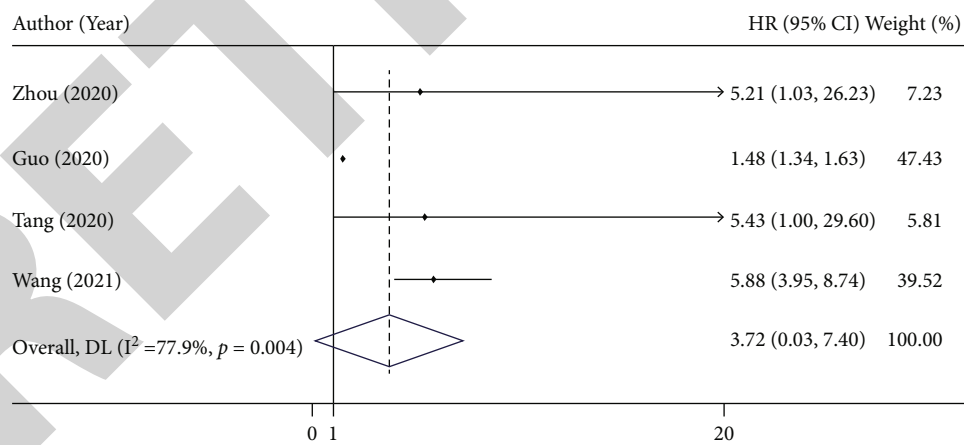
FIGURE 3: Continued.



(d)



(e)



(f)

FIGURE 3: Continued.

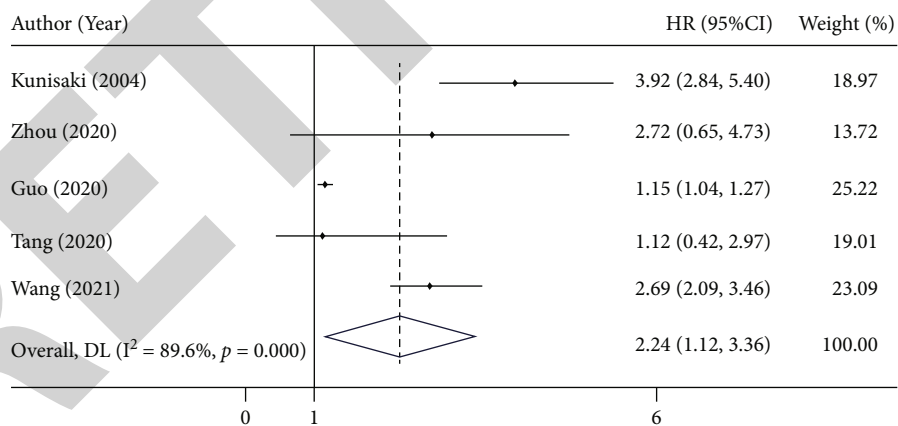
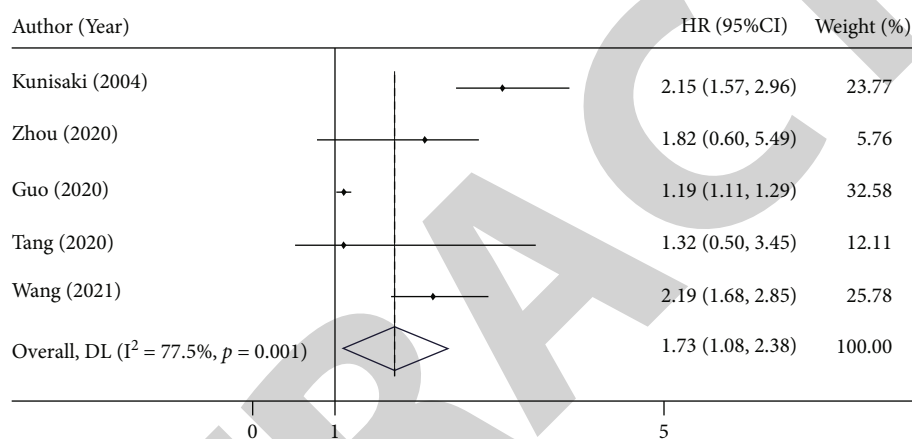
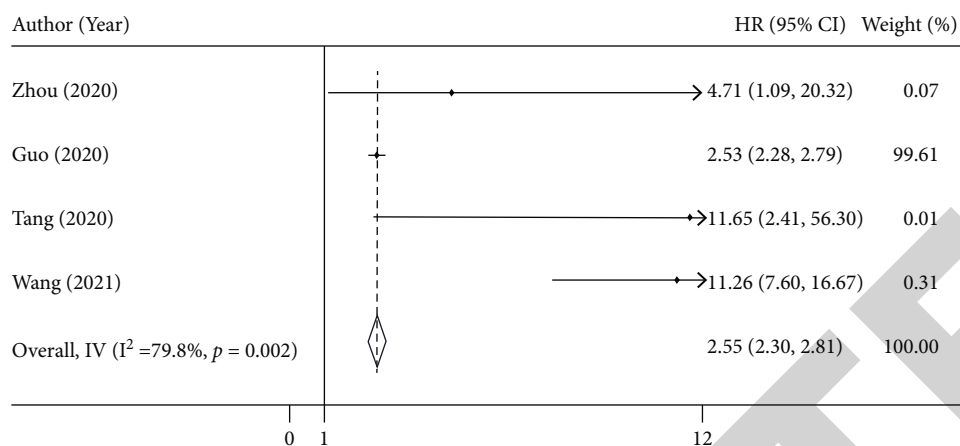


FIGURE 3: Continued.

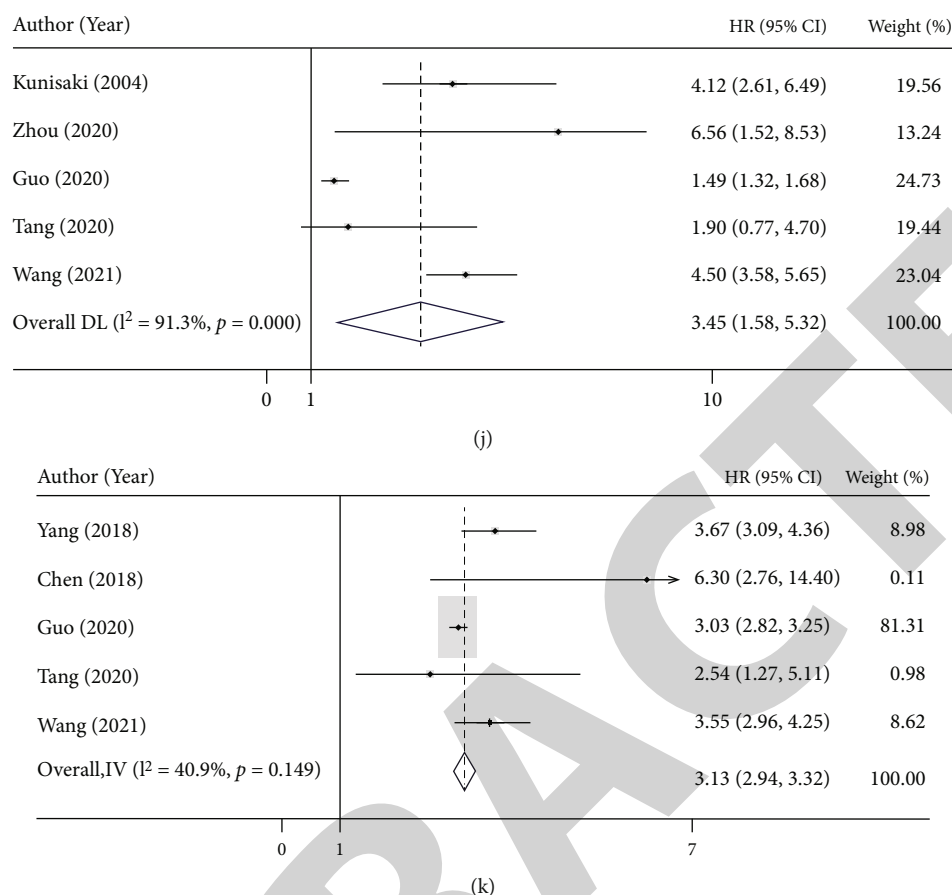


FIGURE 3: Forest plot of the influence of (a) age, (b) sex, (c) chemotherapy, (d) lymphatic vessel invasion, (e–g) T stage, (h–j) N stage, and (k) M stage on SRCC.

as the results of this analysis exhibited a large heterogeneity between studies.

SRCC is frequently found in the mid and distal stomach and is highly prevalent in younger populations [48]. Efares et al. [49] found that in 183 patients with gastric cancer, the mean age of SRCC patients was lower than that of non-SRCC patients, which was consistent with the preliminary analysis of our systematic review. In addition, an increasing number of early-onset gastric cancers [50] have been detected in people younger than 45 years old, and patients with early-onset SRCC are more prone to lymphatic vessel invasion [51]. Early-onset patients have fewer surgical complications than late-onset patients because the previous group receives chemotherapy in a larger proportion, which is conducive to a better prognosis [51, 52]. Consistently, the results of this study are also indicative of a lower HR in SRCC patients receiving chemotherapy. However, the prognosis of patients with early-onset and late-onset diseases remains controversial and may be related to more factors, which calls for further analyses.

Other limitations also call for future discussion. First of all, the current evidence is not strong enough, all of which was based on retrospective investigation that might cause selection bias. Additionally, 10 of the 12 studies we included were from East Asian countries. Given the possible epidemi-

ological and demographic differences between East and West, their data provided may not be a good representation for western populations. Hence, the clinical features and risk factors of SRCC patients from European and American countries may be different. Second, there was considerable heterogeneity among the included studies. More prospective randomized controlled trials are needed in the future to analyze and summarize the clinicopathological features and risk factors of SRCC accurately.

Taken together, our conclusions manifested that SRCC is associated more with males, larger tumor size, and higher TNM stage. In addition, younger age, lymphatic vessel invasion, and higher TNM stage may be linked to the poor prognosis of SRCC patients.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

All authors declare that they have no potential conflicts of interest.

Authors' Contributions

Y G and Q W contributed to the conceptualization and methodology. Q T contributed to the data curation. CW contributed to the formal analysis. N L contributed to the visualization. SJ contributed to the writing of the original draft. PS contributed to the writing, reviewing, and editing of the manuscript. All authors have reviewed and gave final approval of the version to be published. Ying Guo and Qian Wang contributed equally to this work.

Acknowledgments

This study was supported by the funds from the Medical Science Research Project of Hebei Provincial Health Commission (20180255).

References

- [1] A. Jemal, M. M. Center, C. DeSantis, and E. M. Ward, "Global patterns of cancer incidence and mortality rates and trends," *Cancer Epidemiology, Biomarkers & Prevention*, vol. 19, no. 8, pp. 1893–1907, 2010.
- [2] D. Cunningham, W. H. Allum, S. P. Stenning et al., "Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer," *The New England Journal of Medicine*, vol. 355, no. 1, pp. 11–20, 2006.
- [3] J. S. Macdonald, S. R. Smalley, J. Benedetti et al., "Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction," *The New England Journal of Medicine*, vol. 345, no. 10, pp. 725–730, 2001.
- [4] A. P. Thrift and H. B. El-Serag, "Burden of gastric cancer," *Clinical Gastroenterology and Hepatology*, vol. 18, no. 3, pp. 534–542, 2020.
- [5] S. Pernot, T. Voron, G. Perkins, C. Lagorce-Pages, A. Berger, and J. Taieb, "Signet-ring cell carcinoma of the stomach: impact on prognosis and specific therapeutic challenge," *World Journal of Gastroenterology*, vol. 21, no. 40, pp. 11428–11438, 2015.
- [6] C. Mariette, F. Carneiro, H. I. Grabsch, R. S. Van der Post, W. Allum, and G. de Manzoni, "Consensus on the pathological definition and classification of poorly cohesive gastric carcinoma," *Gastric Cancer*, vol. 22, no. 1, pp. 1–9, 2019.
- [7] J. R. Jass, L. H. Sobin, and H. Watanabe, "The World Health Organization's histologic classification of gastrointestinal tumors. A commentary on the second edition," *Cancer*, vol. 66, no. 10, pp. 2162–2167, 1990.
- [8] T. Voron, M. Messenger, A. Duhamel et al., "Is signet-ring cell carcinoma a specific entity among gastric cancers?," *Gastric Cancer*, vol. 19, no. 4, pp. 1027–1040, 2016.
- [9] E. C. Smyth, M. Nilsson, H. I. Grabsch, N. C. van Grieken, and F. Lordick, "Gastric cancer," *Lancet*, vol. 396, no. 10251, pp. 635–648, 2020.
- [10] N. Lemoine, A. Adenis, O. Bouche et al., "Signet ring cells and efficacy of first-line chemotherapy in advanced gastric or oesogastric junction adenocarcinoma," *Anticancer Research*, vol. 36, no. 10, pp. 5543–5550, 2016.
- [11] M. Messenger, J. H. Lefevre, V. Pichot-Delahaye et al., "The impact of perioperative chemotherapy on survival in patients with gastric signet ring cell adenocarcinoma: a multicenter comparative study," *Annals of Surgery*, vol. 254, pp. 684–693, 2011.
- [12] N. Charalampakis, G. M. Nogueras González, E. Elimova et al., "The proportion of signet ring cell component in patients with localized gastric adenocarcinoma correlates with the degree of response to pre-operative chemoradiation," *Oncology*, vol. 90, no. 5, pp. 239–247, 2016.
- [13] S. Bekkar, C. Gronnier, M. Messenger et al., "The impact of preoperative radiochemotherapy on survival in advanced esophagogastric junction signet ring cell adenocarcinoma," *The Annals of Thoracic Surgery*, vol. 97, no. 1, pp. 303–310, 2014.
- [14] G. Piessen, M. Messenger, J. H. Lefevre et al., "Signet ring cell adenocarcinomas: different clinical-pathological characteristics of oesophageal and gastric locations," *European Journal of Surgical Oncology*, vol. 40, no. 12, pp. 1746–1755, 2014.
- [15] Z. Wan, Z. Huang, and L. Chen, "Survival predictors associated with signet ring cell carcinoma of the esophagus (SRCCE): a population-based retrospective cohort study," *PLoS One*, vol. 12, no. 7, article e0181845, 2017.
- [16] U. Heger, L. Sisic, H. Nienhüser et al., "Neoadjuvant therapy improves outcomes in locally advanced signet-ring-cell containing esophagogastric adenocarcinomas," *Annals of Surgical Oncology*, vol. 25, no. 8, pp. 2418–2427, 2018.
- [17] A. M. Stessin, C. Sison, A. Schwartz, J. Ng, C. K. S. Chao, and B. Li, "Does adjuvant radiotherapy benefit patients with diffuse-type gastric cancer? Results from the surveillance, epidemiology, and end results database," *Cancer*, vol. 120, no. 22, pp. 3562–3568, 2014.
- [18] Y. Shu, W. Zhang, Q. Hou et al., "Prognostic significance of frequent CLDN18-ARHGAP26/6 fusion in gastric signet-ring cell cancer," *Nature Communications*, vol. 9, no. 1, p. 2447, 2018.
- [19] S. Ge, X. Xia, C. Ding et al., "A proteomic landscape of diffuse-type gastric cancer," *Nature Communications*, vol. 9, no. 1, p. 1012, 2018.
- [20] Y. Li, Z. Zhu, F. Ma, L. Xue, and Y. Tian, "Gastric Signet Ring Cell Carcinoma: Current Management and Future Challenges," *Cancer Management and Research*, vol. 12, pp. 7973–7981, 2020.
- [21] C. Kunisaki, H. Shimada, M. Nomura, G. Matsuda, Y. Otsuka, and H. Akiyama, "Therapeutic strategy for signet ring cell carcinoma of the stomach," *The British Journal of Surgery*, vol. 91, no. 10, pp. 1319–1324, 2004.
- [22] J. P. Kim, S. C. Kim, and H. K. Yang, "Prognostic significance of signet ring cell carcinoma of the stomach," *Surgical Oncology*, vol. 3, no. 4, pp. 221–227, 1994.
- [23] E. Otsuji, T. Yamaguchi, K. Sawai, and T. Takahashi, "Characterization of signet ring cell carcinoma of the stomach," *Journal of Surgical Oncology*, vol. 67, no. 4, pp. 216–220, 1998.
- [24] W. J. Hyung, S. H. Noh, J. H. Lee et al., "Early gastric carcinoma with signet ring cell histology," *Cancer*, vol. 94, no. 1, pp. 78–83, 2002.
- [25] C. P. Theuer, F. Nastanski, W. R. Brewster, J. A. Butler, and H. Anton-Culver, "Signet ring cell histology is associated with unique clinical features but does not affect gastric cancer survival," *The American Surgeon*, vol. 65, no. 10, pp. 915–921, 1999.
- [26] M. K. Swartz, "PRISMA 2020: An Update," *Journal of Pediatric Health Care*, vol. 35, no. 4, p. 351, 2021.
- [27] J. Liu, S. Nie, Z. Wu et al., "Exploration of a novel prognostic risk signatures and immune checkpoint molecules in

Research Article

Network Pharmacology and Molecular Docking Analysis on the Pharmacological Mechanisms of Modified Sanmiaosan in Treating Ulcerative Colitis

Yong Wang ¹, Ying Sun ², Ruoran Wang ³, Jisha Du ⁴, and Qingqing Wang ⁵

¹School of Traditional Chinese Medicine, Shandong University of Traditional Chinese Medicine, Jinan 250355, China

²Traditional Chinese Medicine Research Institute, Tai'an Hospital of Chinese Medicine, Tai'an 271000, China

³Department of Neurosurgery, West China Hospital, Sichuan University, Chengdu 610000, China

⁴Department of Nephrology, PLA Naval Medical Center, Shanghai 200052, China

⁵Department of Neurology, PLA Naval Medical Center, Shanghai 200052, China

Correspondence should be addressed to Ying Sun; 202072724@yangtzeu.edu.cn

Received 16 June 2022; Revised 17 July 2022; Accepted 19 July 2022; Published 4 August 2022

Academic Editor: Lei Chen

Copyright © 2022 Yong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Modified Sanmiaosan is an effective cure in the treatment of ulcerative colitis, but its mechanisms of action remain unclear. This study revealed the pharmacological mechanisms of Modified Sanmiaosan acting on ulcerative colitis through a pharmacology approach. **Materials and Methods.** The active compounds and the targets of Modified Sanmiaosan were selected from the Traditional Chinese Medicine Systems Pharmacology database according to the absorption and metabolism. The UC-related therapeutic targets were collected from the PharmGKB database, the GeneCards database, the GADA database, and the OMIM database. The networks of “drug-component-target-disease” and “herbal-component-target” were constructed by the Cytoscape software. Protein-protein interaction network was generated by the STRING database. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed by the R software. Molecular docking technology was used to identify the affinity and activity between active compounds and targets. **Results.** The 80 effective ingredients of MSM were collected. A total of 5180 UC-related genes and the 153 key targets of MSM and UC-related were obtained. JUN, Akt1, and MAPK1 were identified as the “hub targets” involved in the effects of Modified Sanmiaosan on ulcerative colitis. Hub targets were mainly involved in inflammatory response and oxidative stress. As the results of GO analysis, biological processes such as DNA-binding transcription and RNA polymerization may participate in the treatment process; KEGG pathway analysis showed that hub targets were mainly involved in IL-17 signal pathway and TNF signal pathway of ulcerative colitis. The high affinity and activity of the active compounds and targets were verified through molecular docking. **Conclusion.** These findings demonstrate the active ingredients in Modified Sanmiaosan reduce inflammatory response by TNF and IL-17 signaling pathways to treat ulcerative colitis. Anti-inflammation and immune regulation may be the main mechanism of Modified Sanmiaosan in the treatment of ulcerative colitis. This study not only provide new insights into the development of a natural therapy for the prevention and treatment of ulcerative colitis but also proves a feasible method for discovering potential activated compounds from Chinese herbs.

1. Introduction

Ulcerative colitis (UC) is an idiopathic, chronic inflammatory disorder of the colonic mucosa, which starts in the rectum and progresses proximally in a continuous manner through the entire colon. It is characterized by chronic mucosal inflammation that recurs for life. The most promi-

nent symptoms include hematochezia, diarrhea, and abdominal pain. Moreover, several patients with UC are affected by extraintestinal complications. The incidence and burden of UC are on the rise globally. The annual incidence of UC is between 8 and 14 cases per 100,000 persons in the Western population, and the prevalence is increasing in Asia [1, 2]. First-line therapy includes 5-aminosalicylic acid and

TABLE 1: Active ingredients of MSM.

Herbal	Mol ID	Molecule name	OB	DL
AE	MOL000006	Luteolin	36.16	0.25
AE	MOL000422	Kaempferol	41.88	0.24
AE	MOL001002	Ellagic acid	43.06	0.43
AE	MOL000098	Quercetin	46.43	0.28
AE	MOL000492	(+)-Catechin	54.83	0.24
RAB	MOL001006	Poriferasta-7,22E-dien-3beta-ol	42.98	0.76
RAB	MOL012461	28-Norolean-17-en-3-ol	35.93	0.78
RAB	MOL012505	Bidentatoside,ii_qt	31.76	0.59
RAB	MOL012537	Spinoside A	41.75	0.4
RAB	MOL012542	β -Ecdysterone	44.23	0.82
RAB	MOL001454	Berberine	36.86	0.78
RAB	MOL001458	Coptisine	30.67	0.86
RAB	MOL000173	Wogonin	30.68	0.23
RAB	MOL002643	Delta 7-stigmastenol	37.42	0.75
RAB	MOL002714	Baicalein	33.52	0.21
RAB	MOL002776	Baicalin	40.12	0.75
RAB	MOL002897	Epiberberine	43.09	0.78
RAB	MOL000358	Beta-sitosterol	36.91	0.75
RAB	MOL003847	Inophyllum E	38.81	0.85
RAB	MOL000422	Kaempferol	41.88	0.24
RAB	MOL004355	Spinasterol	42.98	0.76
RAB	MOL000449	Stigmasterol	43.83	0.76
RAB	MOL000785	Palmatine	64.6	0.65
RAB	MOL000085	Beta-daucosterol_qt	36.91	0.75
RAB	MOL000098	Quercetin	46.43	0.28
BS	MOL005755	1-(4-Hydroxybenzyl)-4-methoxy-9,10-dihydrophenanthrene-2,7-diol	54.18	0.55
BS	MOL005756	2,3,4,7-Tetramethoxyphenanthrene	39.09	0.29
BS	MOL005759	2,7-Dihydroxy-4-methoxyphenanthrene-2,7-O-diglucoside	30.22	0.74
BS	MOL005761	3-(p-Hydroxybenzyl)-4-methoxy-9,10-dihydrophenanthrene	37.98	0.55
BS	MOL005766	3,7-Dihydroxy-2,4-dimethoxyphenanthrene-3-O-glucoside	31.46	0.78
BS	MOL005768	4,7-Dihydroxy-1-p-hydroxybenzyl-2-methoxy-9,10-dihydrophenanthrene	30.54	0.55
BS	MOL005770	Bletlol A	54.43	0.55
BS	MOL005773	Blespirol	43.74	0.86
BS	MOL005776	1-(2,7-Dihydroxy-4-methoxy-1-phenanthryl)-4-methoxyphenanthrene-2,7-diol	35.22	0.67
AL	MOL000173	Wogonin	30.68	0.23
AL	MOL000179	2-Hydroxyisoxypopyl-3-hydroxy-7-isopentene-2,3-dihydrobenzofuran-5-carboxylic	45.2	0.2
AL	MOL000184	NSC63551	39.25	0.76
AL	MOL000186	Stigmasterol 3-O-beta-D-glucopyranoside_qt	43.83	0.76
AL	MOL000188	3 β -Acetoxyatractylone	40.57	0.22
AL	MOL000085	Beta-daucosterol_qt	36.91	0.75
AL	MOL000088	Beta-sitosterol 3-O-glucoside_qt	36.91	0.75
AL	MOL000092	Daucosterin_qt	36.91	0.76
AL	MOL000094	Daucosterol_qt	36.91	0.76
PC	MOL001454	Berberine	36.86	0.78
PC	MOL001458	Coptisine	30.67	0.86
PC	MOL002636	Kihadalactone A	34.21	0.82
PC	MOL013352	Obacunone	43.29	0.77
PC	MOL002641	Phellavin_qt	35.86	0.44

TABLE 1: Continued.

Herbal	Mol ID	Molecule name	OB	DL
PC	MOL002643	Delta 7-stigmastenol	37.42	0.75
PC	MOL002644	Phellopterin	40.19	0.28
PC	MOL002651	Dehydrotanshinone II A	43.76	0.4
PC	MOL002652	delta7-Dehydrosophoramine	54.45	0.25
PC	MOL002656	Dihydroniloticin	36.43	0.81
PC	MOL002659	Kihadanin A	31.6	0.7
PC	MOL002660	Niloticin	41.41	0.82
PC	MOL002662	Rutaecarpine	40.3	0.6
PC	MOL002663	Skimmianin	40.14	0.2
PC	MOL002666	Chelerythrine	34.18	0.78
PC	MOL000449	Stigmasterol	43.83	0.76
PC	MOL002668	Worenine	45.83	0.87
PC	MOL002670	Cavidine	35.64	0.81
PC	MOL002671	Candletoxin A	31.81	0.69
PC	MOL002672	Hericenone H	39	0.63
PC	MOL002673	Hispidone	36.18	0.83
PC	MOL000358	Beta-sitosterol	36.91	0.75
PC	MOL000622	Magnograndiolide	63.71	0.19
PC	MOL000762	Palmidin A	35.36	0.65
PC	MOL000785	Palmatine	64.6	0.65
PC	MOL000787	Fumarine	59.26	0.83
PC	MOL000790	Isocorypalmine	35.77	0.59
PC	MOL000098	Quercetin	46.43	0.28
PC	MOL001131	Phellamurin_qt	56.6	0.39
PC	MOL001455	(S)-Canadine	53.83	0.77
PC	MOL001771	Poriferast-5-en-3beta-ol	36.91	0.75
PC	MOL002894	Berberrubine	35.74	0.73
PC	MOL005438	Campesterol	37.58	0.71
PC	MOL006392	Dihydroniloticin	36.43	0.82
PC	MOL006401	Melianone	40.53	0.78
PC	MOL006413	Phellochin	35.41	0.82
PC	MOL006422	Thalifendine	44.41	0.73

AL: *Atractylodes lancea*; PC: *Phellodendri chinensis*; RAB: *Radix Achyranthis bidentatae*; BS: *Bletilla striata*; AE: *Agrimonia eupatoria*.

corticosteroids. Glucocorticoids and biological agents can be added when 5-aminosalicylic acid therapy is ineffective [2]. Because of the marked side effects of Western medicine, patients still have difficulty adhering to treatment, leading to the discontinuation of therapy. As an important component of complementary and alternative medicine, traditional Chinese medicine (TCM) has been widely used to treat and prevent UC and has shown significant positive effects and low recurrence rates [3].

In TCM, clearing heat and removing dampness is a key method for UC treatment [4]. Sanmiao powder is a classic TCM prescription for clearing heat and drying dampness, including *Atractylodes lancea*, *Phellodendri chinensis*, and *Radix Achyranthis bidentatae* [5].

Based on the formulation of Sanmiaosan, *Bletilla striata* and *Agrimonia eupatoria* were added to form Modified Sanmiaosan (MSM). MSM has been reported to have achieved a

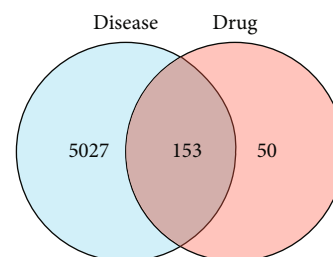


FIGURE 1: Venn diagram of the targets of the active ingredients of the drug and the disease-related targets.

good curative effect in the treatment of UC [6]. *Atractylodes lancea* has proven regulation of the nuclear factor κ B (NF- κ B) signaling pathway to reduce lipopolysaccharide-induced inflammatory injury of human colonic epithelial cells [7]. *Agrimonia eupatoria* ameliorates dextran sodium

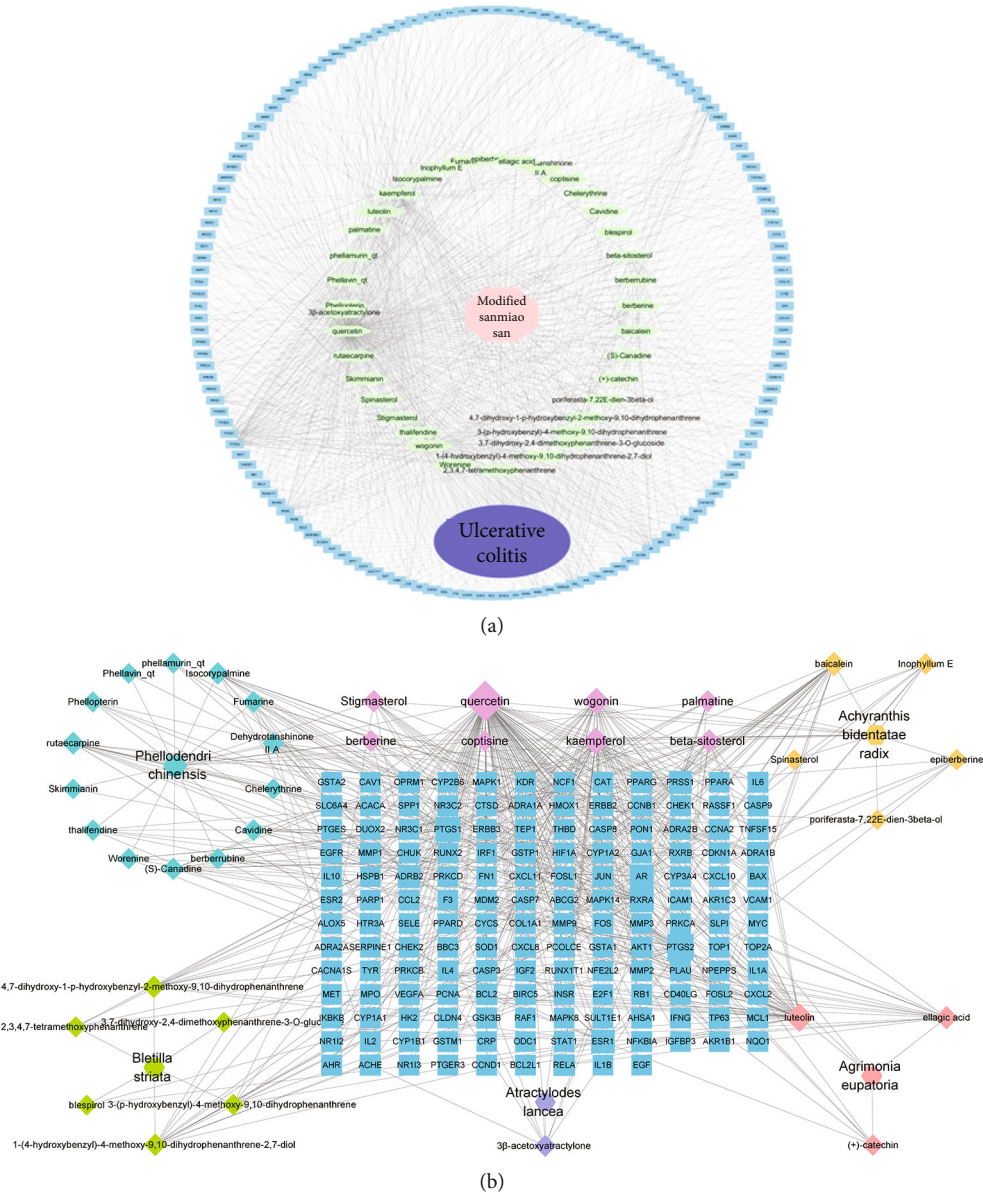


FIGURE 2: (a) “Drug-component-target-disease” network construction. The blue rectangle nodes represent the targets; the green diamond nodes represent ingredients of MSM. (b) “Herbal-component-target” network construction. The blue rectangle nodes represent the targets.

TABLE 2: The degree of active ingredients of MSM.

Active ingredients	Degree
Kaempferol	75
Luteolin	46
Quercetin	340

sulfate-induced colitis through NF- κ B signaling pathway [8]. *Bletilla striata* effectively inhibited the release of intestinal inflammatory cytokines such as IL-6, TNF- α , and IL-1 β and attenuated dysfunction of the intestinal barrier in UC mouse model induced by 3% dextran sodium sulfate [9]. MSM is effective in the treatment of UC, but its underlying

pharmacological properties and mechanisms of action remain unclear [6].

Network pharmacology is an innovative way to explore new drug targets, which can identify active drug ingredients by investigating the relationship between drug components and disease targets. Based on the intersection of drug targets with key genes, the drug-target network can be constructed, and the potential mechanism can be explored. This method can be helpful to discern the pharmacological action and mechanism of TCM [10]. In this study, we adopted a network pharmacology approach to investigate how MSM exerts therapeutic effects on UC by target prediction, biological function, and pathway analysis. This study may provide a useful reference for the identification of the therapeutic mechanisms of TCM.

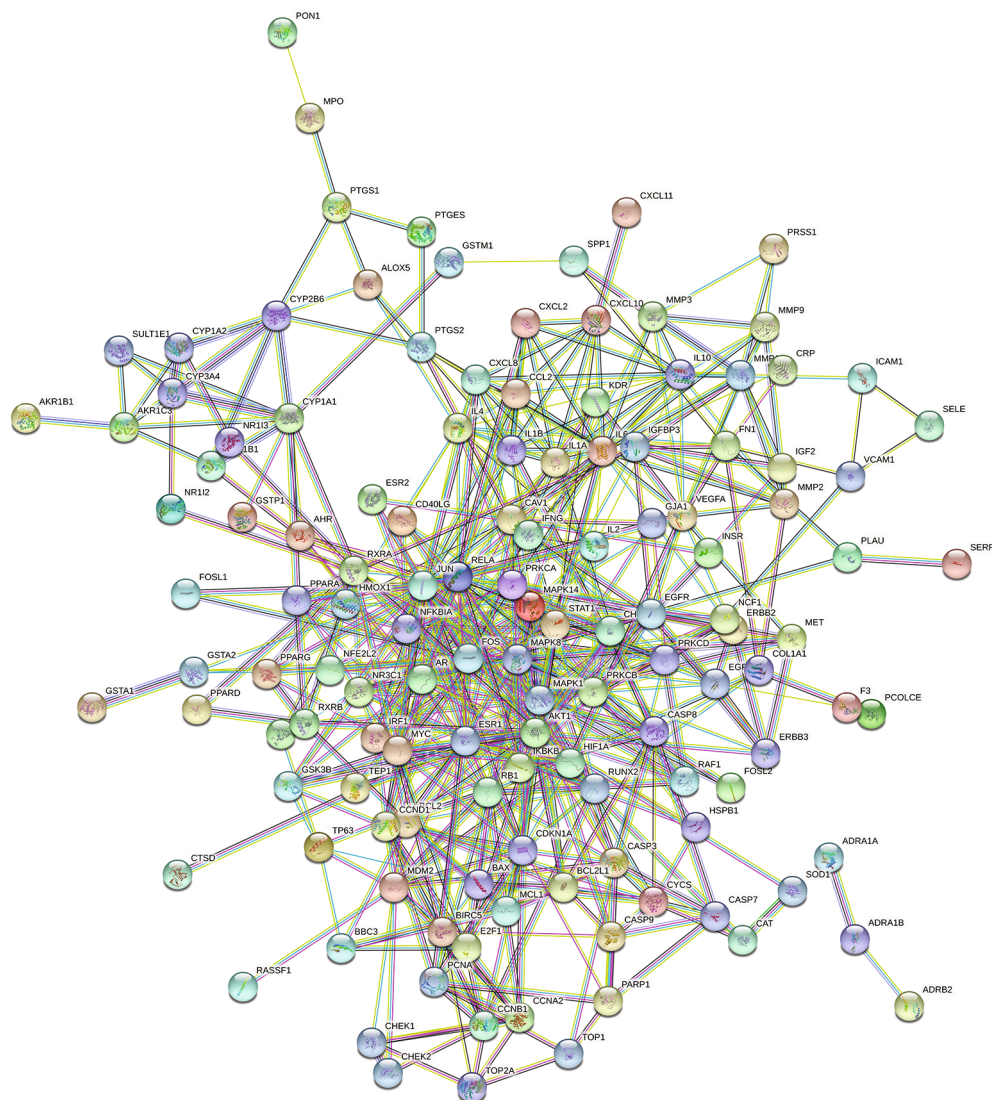


FIGURE 3: PPI network of the core targets.

2. Materials and Methods

2.1. MSM Ingredients. The components of MSM (*Atractylodes lancea*, *Phellodendri chinensis*, *Radix Achyranthis bidentatae*, *Bletilla striata*, *Agrimonia eupatoria*) were searched in the Traditional Chinese Medicine Systems Pharmacology (TCMSP) database (<https://old.tcmsp-e.com/tcmsp.php/>) by the ADME model. The ingredients with $OB \geq 30\%$ and $DL \geq 0.18$ were selected. OB represents the ability of a compound to circulate in the body after oral administration. DL is an indicator for determining the similarity of physicochemical properties of a compound with conventional drugs. The target names of all compounds were transformed into UniProt IDs using the UniProt database (<http://www.uniprot.org/>).

2.2. Databases to Identify and Predict the Targets of UC. By searching the keywords “Ulcerative colitis”, the targets related to UC were collected from the GeneCards database (<https://www.genecards.org>), the OMIM database (<https://omim.org>), the PharmGKB database (<https://www.pharmgkb.org>), and the Genetic Association Database (GADA, <https://www.Geneticassociationdb.nih.gov>) [11].

2.3. Key Targets. The Venn diagram generation software Venny 2.1 was used to identify potential key targets of MSM in treating UC.

2.4. Construction of Network Relationships. The “drug-component-target-disease” and “herbal-component-target” networks were constructed for visualization using Cytoscape version 3.6.1 (<http://www.cytoscape.org/>). The nodes in the network represent herbs, chemicals, and potential targets; the edges indicate interactions between them. In the network, the quantity of each edge was defined as “degree.”

2.5. Protein-Protein Interactions (PPI). Key targets were imported into the STRING (<https://string-db.org/>) database to build the PPI network model. The minimum required interaction score was set as “medium confidence = 0.9,”

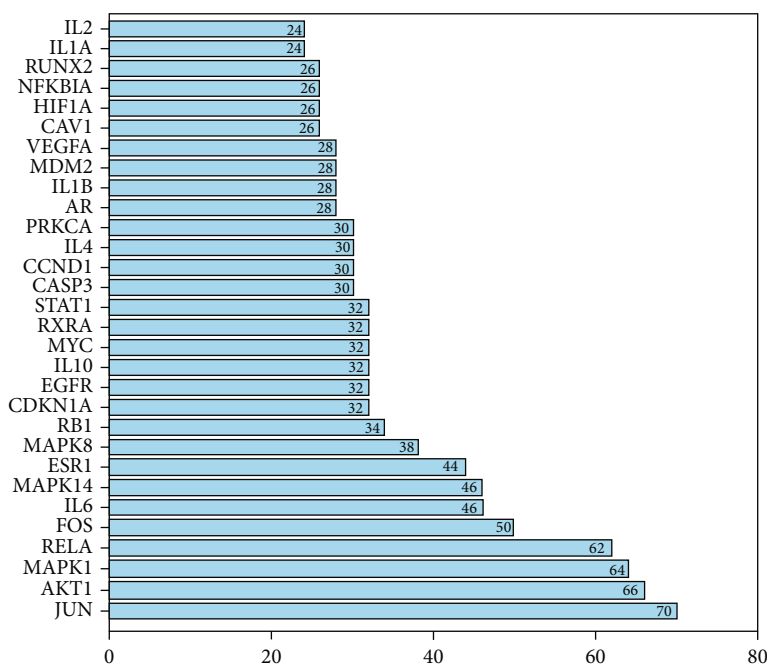


FIGURE 4: 30 core targets determined in the PPI network.

while the other parameters remain as default. In the network, the size of the nodes represents the degree size. The higher the degree, the better the correlation between the proteins. Targets with scores in the top 30 were selected as core targets [12].

2.6. Enrichment Analysis. Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of target genes were performed using the R software (version 3.6.0); the p value was defined as <0.05 and $Q < 0.05$. The top 20 enriched pathways of GO and KEGG were reported as bubble charts and bar graphs [12].

2.7. Molecular Docking Analysis. Download the SDF format of 3D structure of kaempferol, luteolin, and quercetin in PubChem, and then import it into ChemDraw 3D. Use the MM 2 module to minimize the energy, obtain the optimal conformation of the lowest energy, and save it as a mol 2 file. MAPK1, AKT1, and JUN were retrieved from the PDB database to obtain the protein structure; their PDB IDs are 6slg, 7nh5, and 1a02. Then, PyMOL is used for visualization. MGTools 1.5.6 is used for water removal, hydrogenation, charge calculation, and nonpolar hydrogen combination. Then, use AutoDock Vina 1.1.2 for docking the ligand with the receptor [13].

3. Results

3.1. Distribution of MSM Ingredients. 80 ingredients of MSM that satisfied the DL criteria were retrieved from the TCMSP database, including 9 *Atractylodes lancea* ingredients, 37 *Phellodendri chinensis* ingredients, 20 *Radix Achyranthis*

bidentatae ingredients, 9 *Bletilla striata* ingredients, and 5 *Agrimonia eupatoria* ingredients (Table 1).

3.2. Potential Targets of MSM in Treating UC. The targets of active ingredients were downloaded from the TCMSP database. A total of 1458 UC-related targets were obtained from the GADA database, 15 UC-related targets were obtained from the PharmGKB database, 4800 UC-related targets were obtained from the GeneCards database, and 7 were obtained from the OMIM database. A total of 5180 UC-related targets were retrieved after eliminating the duplicate entries. The MSM targets and UC targets were intersected to obtain 153 targets, which were noted as “key targets” to construct the PPI network and for GO\KEGG analysis (Figure 1).

3.3. Analysis of “Drug-Component-Target-Disease” Network and “Herbal-Component-Target” Network. The “drug-component-disease-target” network consisted of 80 active ingredients and 153 key targets. In Figure 2(a), the purple oval represents UC, the pink polygon represents MSM, the green diamond represents the active ingredient of MSM, and the blue rectangle represents the disease targets corresponding to the active ingredients of the drug. There are 192 nodes and 996 edges in the drug-component-disease-target network. The three active components with the highest “degree” values are quercetin, kaempferol, and luteolin (as shown in Table 2); it is suggested that it is the key active ingredient of the network. In the herbal-component-target network (Figure 2(b)), the blue rectangle nodes represent the targets.

3.4. PPI Network Construction and PPI-Based Hub Target Identification. The 153 key targets were uploaded to the STRING database (<https://string-db.org/cgi/input.pl>) to

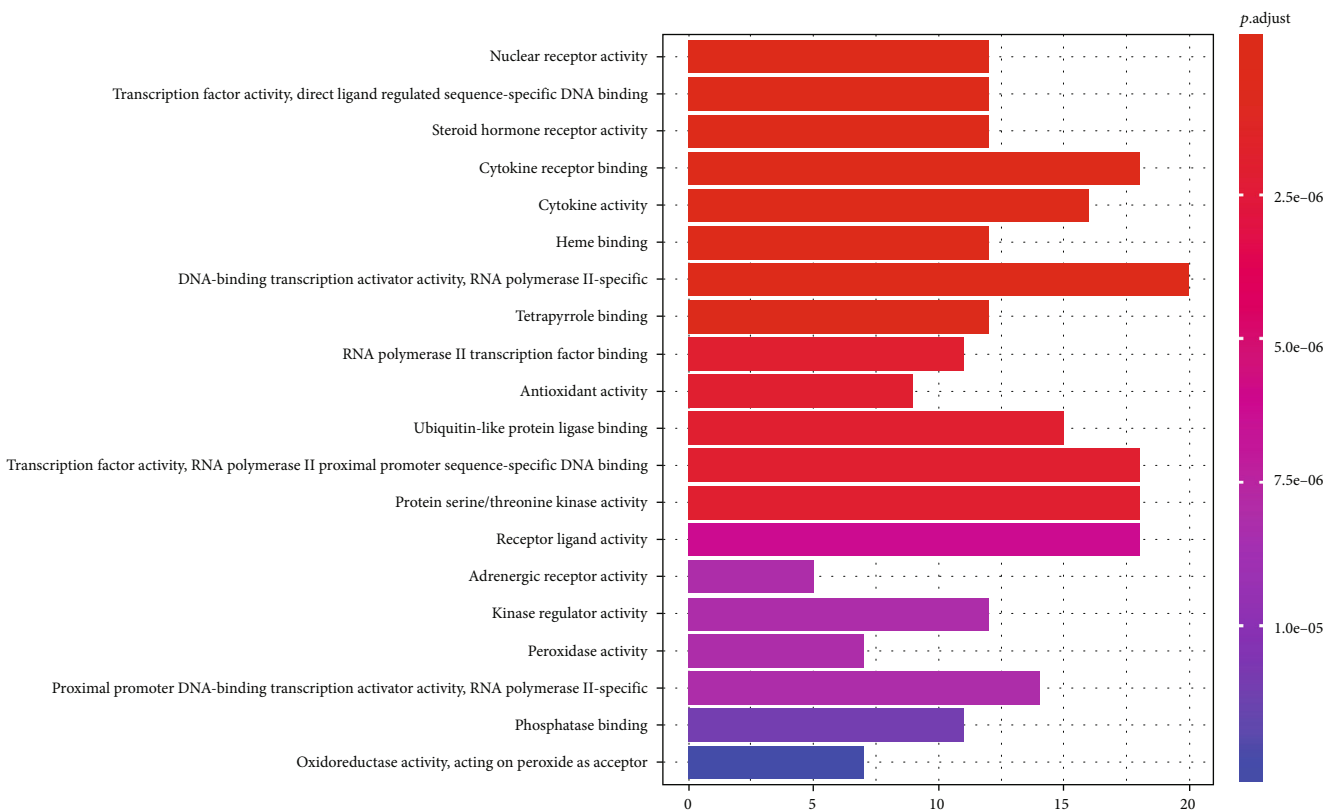
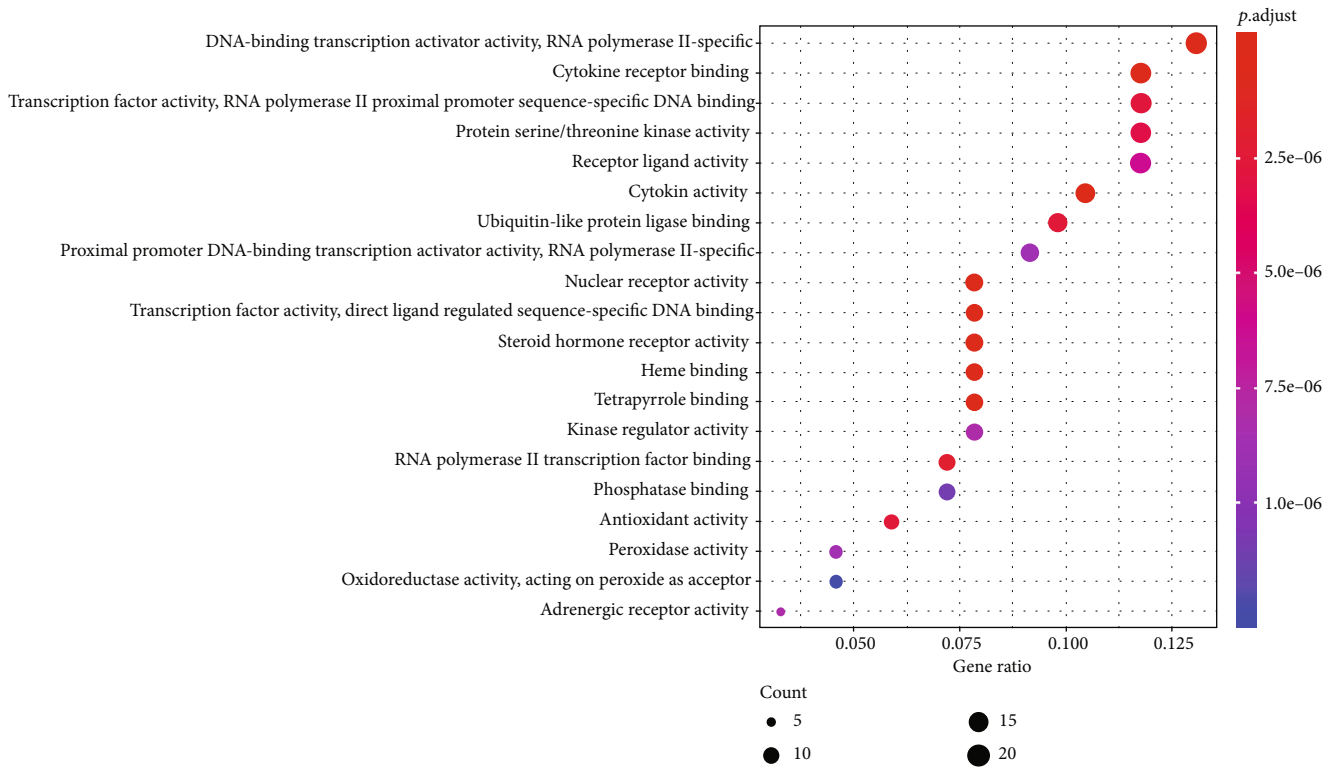


FIGURE 5: (a) Dot plot of GO enrichment analysis. The “x” coordinate indicates the number of enriched genes, and the color of the dot represents the p value of the corresponding term. The larger the dot, the more genes are enriched. (b) Bar plot of GO enrichment analysis.

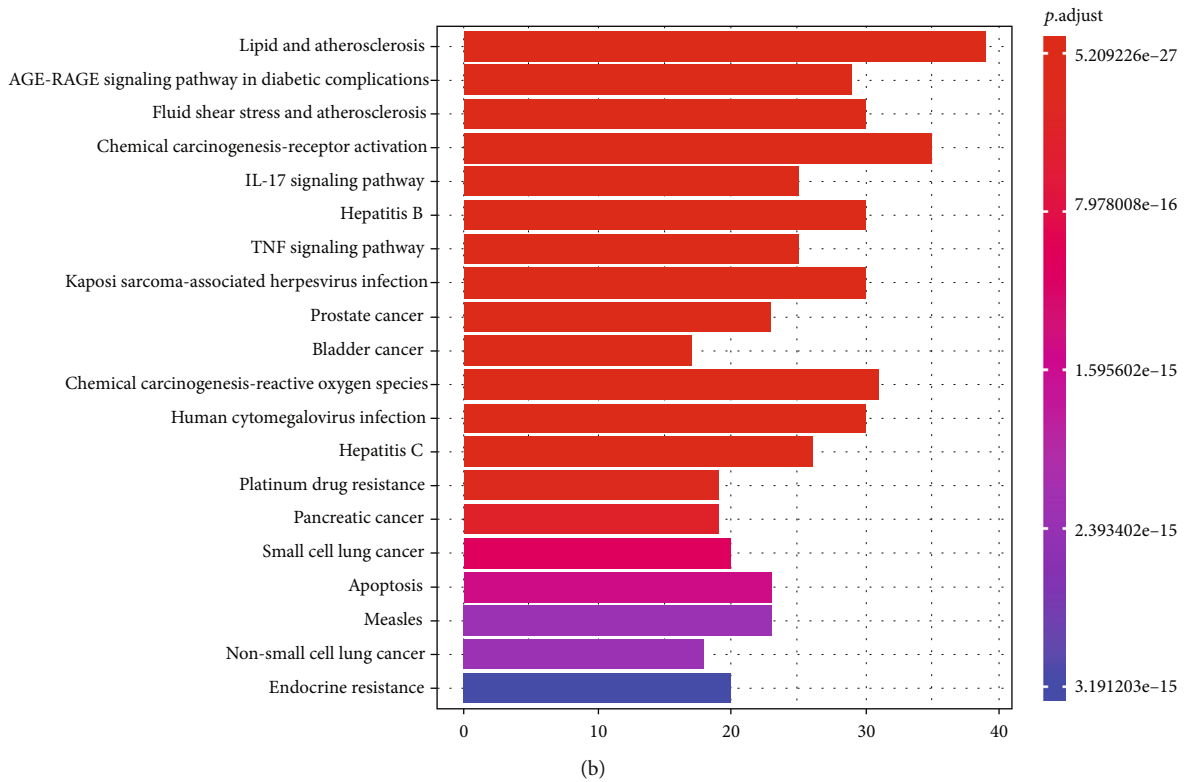
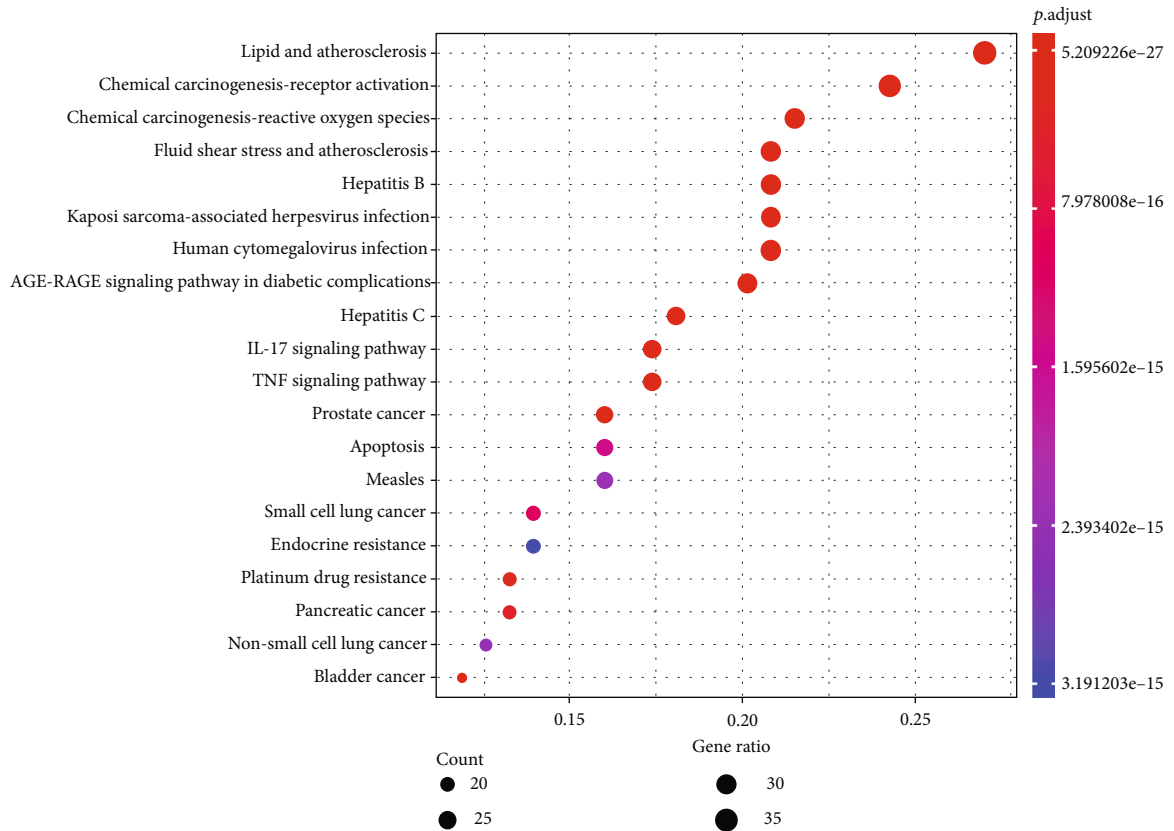
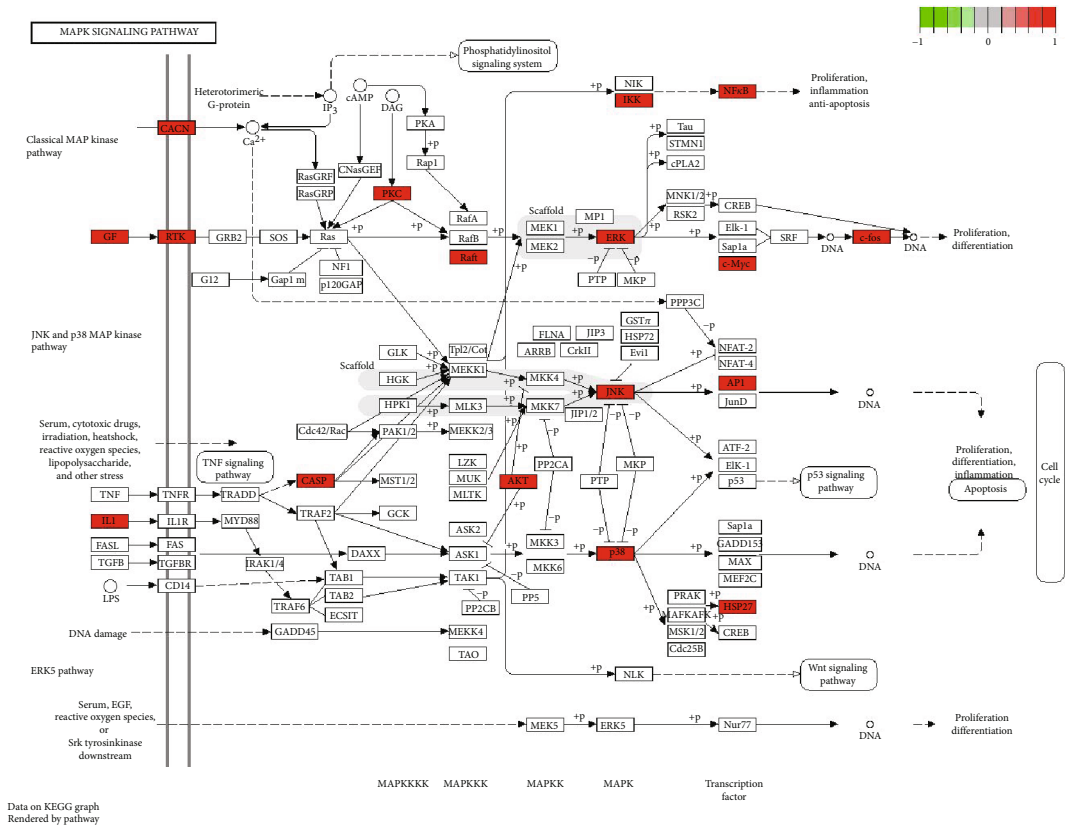
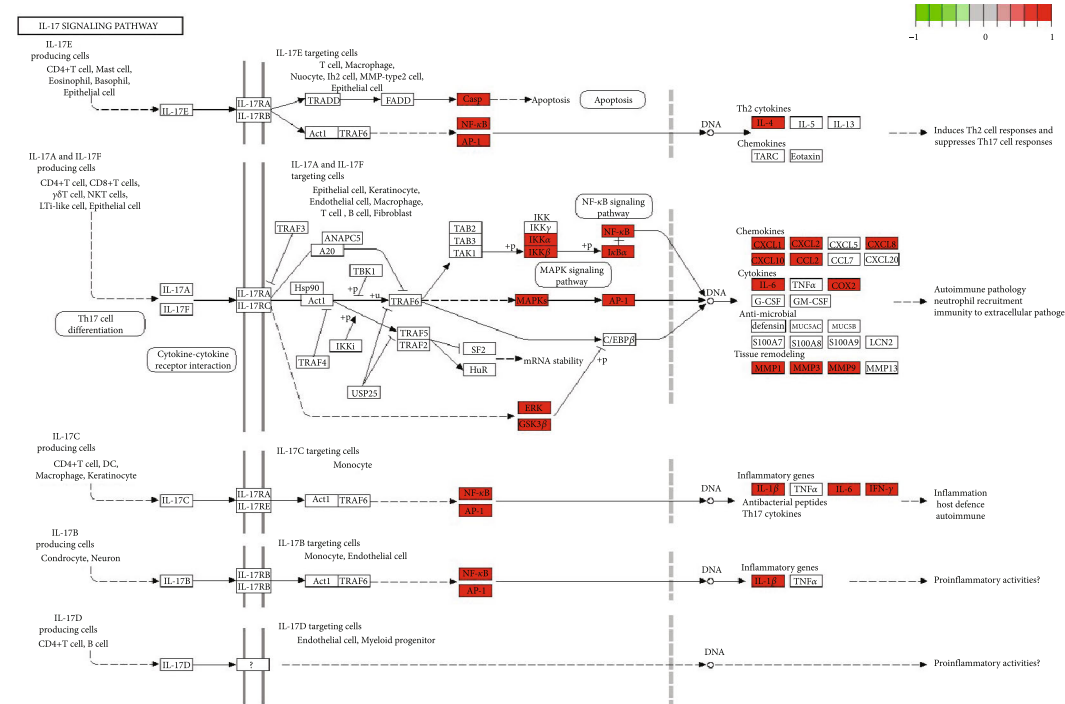


FIGURE 6: (a) Dot plot of KEGG enrichment analysis. The size of the bubble represents the number of targets enriched in the indicated pathway, and the color of the bubble represents the *p* value of enrichment. (b) Bar plot of KEGG enrichment analysis.



(a)



(b)

FIGURE 7: (a) TNF pathway map of MSM for their potential treatment of UC. Nodes in red represent MSM-UC-related target. (b) IL-17 pathway map of MSM for their potential treatment of UC. Nodes in red represent MSM-UC-related target.

TABLE 3: Energy docking scores (kcal mol⁻¹).

	Kaempferol	Luteolin	Quercetin
MAKP1	-8.1	-8	-8
AKT1	-9.2	-9.8	-10.1
JUN	-6.5	-6.6	-6.6

construct the PPI network. In Figure 3, there were 153 nodes and 547 edges in the PPI network (PPI enrichment p value: $< 1.0e-16$), which reflects the complexity of the disease mechanism and the characteristics of TCM for disease treatment. Gene targets with a high degree in the PPI network include JUN, Akt1, and mitogen-activated protein kinase-1 (MAPK1); these genes were considered “hub targets” involved in the effects of MSM on UC (Figures 3 and 4).

3.5. GO and KEGG Enrichment Analyses. 153 key targets were selected for GO and KEGG pathway enrichment analyses. The top 20 GO analysis results are shown in Figures 5(a) and 5(b). The GO enrichment analysis reported that the top results are chiefly involved in DNA-binding transcription activator activity, RNA polymerase II-specific, cytokine receptor binding, transcription factor activity, RNA polymerase II proximal, promoter sequence-specific DNA binding, protein serine/threonine kinase activity, receptor ligand activity, and cytokine activity.

The top 20 pathways are shown in Figures 6(a) and 6(b). Pathways unrelated to UC were excluded. The KEGG pathway analysis showed these targets to be mainly associated with IL-17 and TNF signaling pathways of UC (Figures 7(a) and 7(b)).

Active compounds from MSM could target multiple proteins and subsequently initiate complex signal transduction, regulate inflammation-related pathways, restrain the progression of UC, and provide a better prognosis in patients.

3.6. Molecular Docking Studies to Evaluate Interactions between MSM and UC Targets. The top three targets (JUN, Akt1, and MAPK1) in the PPI network and the top three active ingredients (quercetin, kaempferol, and luteolin) in the network were selected for molecular docking. Molecular docking showed that quercetin, kaempferol, and luteolin can closely bind three targets (JUN, Akt1, and MAPK1). Table 3 shows the energy docking scores. The lower the binding energy between small molecule ligands and protein receptors, the better the affinity between them and the more stable the concept. It is generally accepted that binding energy which is less than -4.25 kcal/mol, -5.0 kcal/mol, or -7.0 kcal/mol indicates a certain, good, or strong binding activity between the ligand and the receptor, respectively [13].

The PyMOL software was used for visual analysis of molecular docking. The 3D structures of JUN, Akt1, and MAPK1 are shown in Figures 8(a)–8(c). We selected the top three receptor proteins with the lowest energy value and the ligand that bound to these receptor proteins for visualization (see Figures 9(a)–9(c)). It shows the best docking combinations for Akt1 and the compound, including

kaempferol, luteolin, and quercetin, with binding energies of -9.2 , -9.8 , and -10.1 kcal/mol, respectively. It indicates that there is strong binding activity between receptor and ligand.

4. Discussion

In this study, network pharmacology methods were used to explore the potential targets and mechanisms of MSM in treating UC. A total of 80 active compounds were identified using the TCMSP database ($OB \geq 30\%$ and $DL \geq 0.18$) and 5180 UC-related targets were obtained from the PharmGKB database, the GeneCards database, the GADA database, and the OMIM database. The MSM targets and UC targets were intersected by using a Venn tool to obtain 153 targets, identified as “key targets” for PPI network construction and GO\KEGG analysis. The PPI network was constructed by uploading these 153 key targets to the STRING database. Targets with high degree value in the PPI network included JUN, Akt1, and MAPK1; these genes were considered “hub targets” involved in the effects of MSM on UC. GO and KEGG enrichment analyses of the 153 key targets revealed that IL-17 and TNF signaling pathways were the most enriched pathways. Cytoscape was used to generate the network showing the relationship between drug components and targets containing 80 active ingredients and 153 key targets. The treatment of diseases by TCM was mainly achieved by the synergistic action of many active compounds on multiple targets. Network pharmacology was used to reveal the underlying mechanisms with the above characteristics.

4.1. Active Ingredients of MSM. As previously reported, 153 active compounds of MSM in the treatment of UC were obtained. Through network analysis, the most promising compounds were identified as quercetin, kaempferol, luteolin, β -sitosterol, and baicalein. This finding suggests that these compounds may be important for the therapeutic role of MSM in UC and warrant further investigation. As reported, quercetin is a polyphenolic flavonoid compound whose antioxidant activity has been investigated in numerous studies. It is effective in treating and preventing various diseases due to its effect on glutathione, enzyme activity, signal transduction pathways, and reactive oxygen species (ROS) production [14]. Kaempferol is one of the most common polyphenols attached to sugar groups in fruits and vegetables [15]. Kaempferol is a proven regulator of ER stress and autophagy and exhibits protective effects on malfunctioning cells [16]. In a novel epithelial-endothelial cell coculture model, kaempferol is confirmed to reduce lipopolysaccharide- (LPS-) induced IL-8 secretion and barrier dysfunction of Caco-2 cells [17]. Luteolin is a flavonoid found in plants, with anticancer activity against various types of human malignancies [18]. Moreover, luteolin modulates many inflammatory pathways by suppressing proinflammatory mediators (interleukin- (IL-) 1β , IL-6, IL-17, IL-22, and TNF- α) and regulates various inflammation-related signaling pathways [19]. β -Sitosterol is a bioactive compound found in the plant cell membrane, with a chemical structure similar to cholesterol. β -Sitosterol is known to

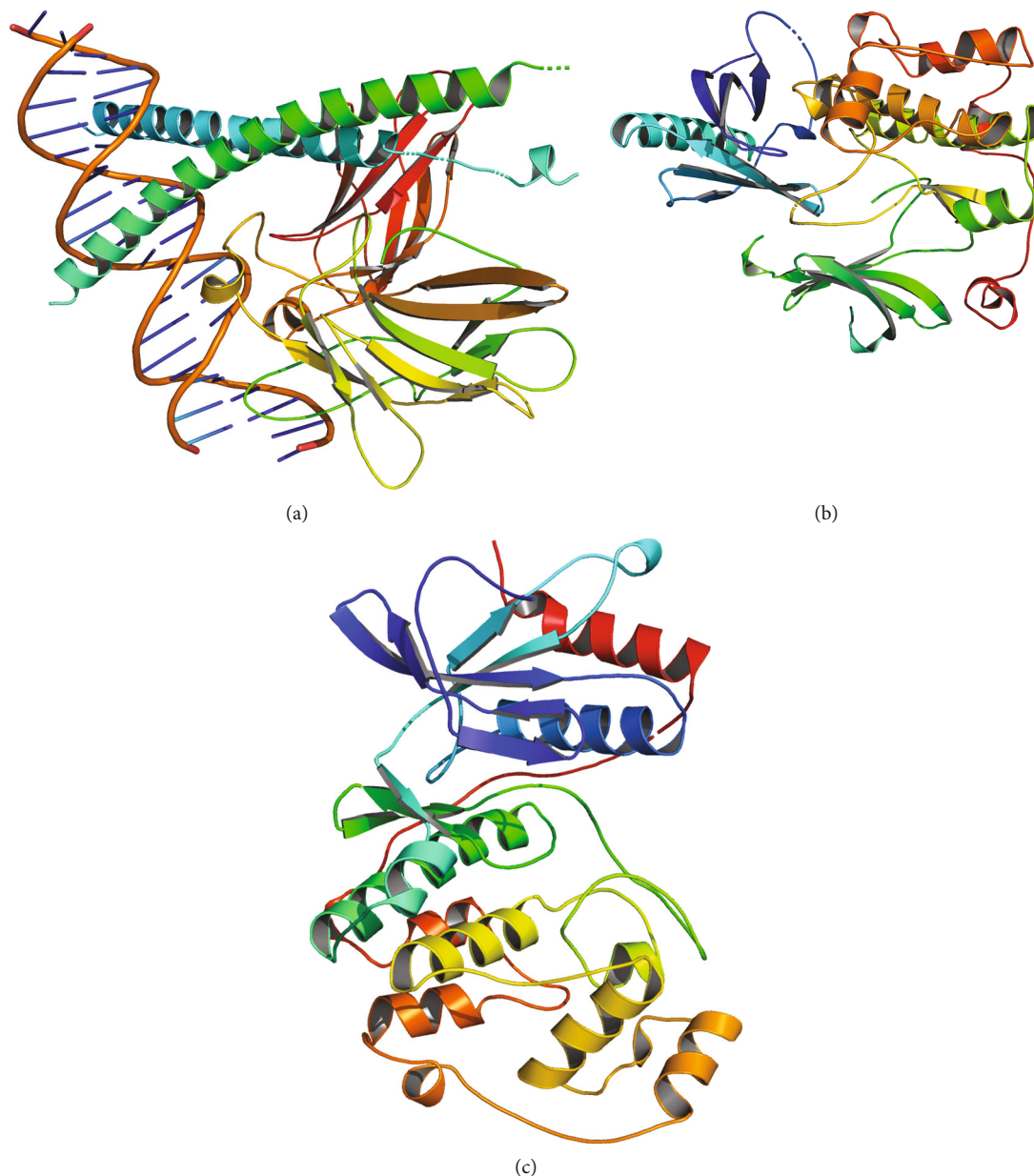


FIGURE 8: (a) The 3D structures of JUN. (b) The 3D structures of Akt1. (c) The 3D structures of MAPK1.

possess various benefit actions such as immunomodulatory effect, anticancer, and anti-inflammatory activity [20]. Various studies have demonstrated the antioxidative, anti-inflammatory, and antitumor activities of baicalein. Baicalein has also been reported to ameliorate dextran sulfate sodium-induced colitis by inhibiting NF- κ B and STAT3 signaling pathways and mitigating radiation-induced enteritis by improving endothelial dysfunction, thereby preventing the development of colitis-associated cancer [21, 22].

4.2. Core Targets of MSM in UC. PPI network identified JUN, Akt1, and MAPK1 as core targets. These targets involved in inflammation, apoptosis, and other processes may play an important role in the therapeutic effect of MSM in UC. JUN is a major component of the heterodi-

meric transcription factor AP-1 and a critical driver of cell proliferation and apoptosis [23]. Additionally, various growth factors, cytokines, and stress stimulate the activation of c-Jun and then regulate biological processes [24]. Akt1 is one of the serine/threonine protein kinases called Akt kinase, which is integral for insulin signaling, endothelial function, and metabolic regulation [25]. Akt plays an important role as a protumorigenic and metastatic regulator in cancer due to its specific effects on cancer cells, tumor endothelial cells, and stromal cells [26]. AKT1 has been demonstrated to be involved in cell survival, and proliferation, and as a downstream factor of PI3K. MAPK is an interconnected signaling cascade with involvement in cell growth, differentiation, inflammatory response, and other important cellular, physiological, and pathological pathways [27]. The

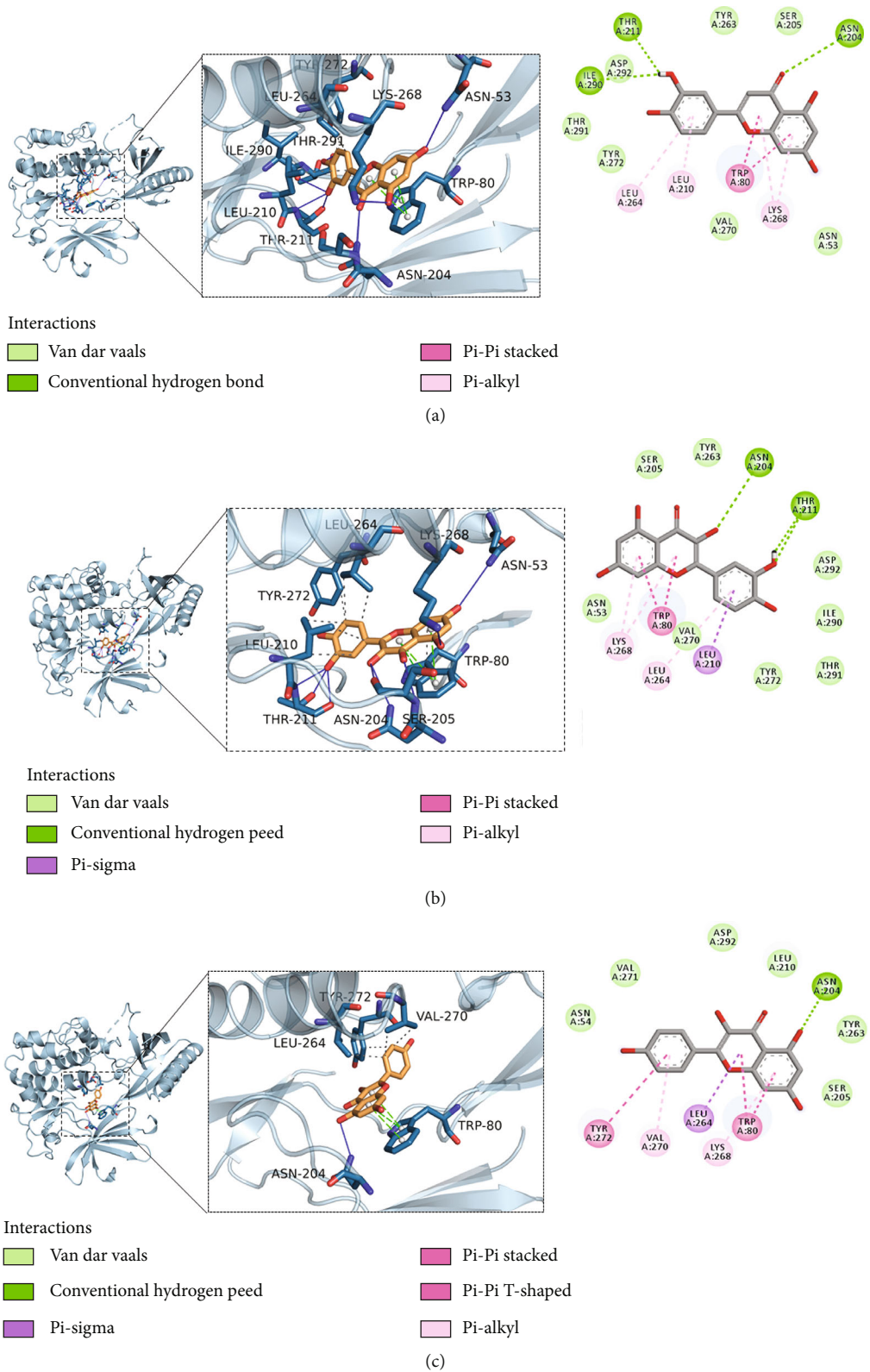


FIGURE 9: (a) Schematic diagram of the docking of AKT1 and kaempferol. (b) Schematic diagram of the docking of AKT1 and luteolin. (c) Schematic diagram of the docking of AKT1 and quercetin.

regulation of PI3K/Akt and MAPK/Akt pathways by MSM may play an important role in its therapeutic effect in UC. The results of molecular docking studies showed that quercetin, kaempferol, and luteolin can closely bind three targets (JUN, Akt1, and MAPK1).

4.3. MSM Pathways in Treating UC. We performed GO enrichment analysis and KEGG enrichment analysis of the 153 potential targets. GO analysis results showed that biological processes such as DNA-binding transcription activator activity, RNA polymerase II-specific cytokine receptor binding, transcription factor activity, RNA polymerase II proximal, promoter sequence-specific DNA binding, protein serine/threonine kinase activity, receptor ligand activity, and cytokine activity may play a role in the treatment of UC. KEGG pathway analysis showed that these targets were mainly involved in IL-17 and TNF signaling pathways of UC.

The pathogenesis of UC consists of immunoinflammatory pathways, including the epithelial barrier, commensal microflora, dysregulation of immunological responses, and leukocyte recruitment.

Proinflammatory cytokines (tumor necrosis factor- α (TNF- α), IL-1, IL-6, IL-9, IL-13, and IL-33) play a significant role in upregulation, while anti-inflammatory cytokines (transforming growth factor- β , IL-10, and IL-37) play a significant role in downregulation of disease progression. The imbalance of proinflammatory and anti-inflammatory cytokines is a key factor in the immune dysfunction of intestinal mucosal [28]. A clearer understanding of the inflammatory pathways of UC regulation by MSM may aid in the development of drug screening.

The proinflammatory cytokines IL-17 have been implicated in the pathogenesis of inflammatory bowel disease and have received a lot of attention for their activity. IL-17 signaling can induce a cascade of proinflammatory molecules such as TNF, IFN γ , IL-22, lymphotoxin, IL-1 β , and LPS [29]. TNF signaling pathways play a vital role in the typical immune response by regulating several downstream processes [30]. The IL-17 and TNF signaling pathways are considered classical signaling pathways in immune reaction. These two signaling pathways can activate the intracellular molecular signal, leading to the activation of cells in the ER, which results in inflammation of intestinal mucosal cells [28]. As biological response modifiers, antibodies neutralizing TNF- α have been used in studies and clinical practice [31].

In conclusion, chronic inflammation, the main pathological process of UC, may lead to the accumulation of high levels of proinflammatory cytokines within the colonic mucosa and could lead to cancer. The risk of colorectal cancer (CRC) in patients with UC is 2.4-fold higher compared to the general population [2]. Improving inflammatory response is a viable approach for the treatment of UC and prevention of cancer. This study showed that MSM improved the inflammatory response of UC via the IL-17 and TNF signaling pathways, which may prove beneficial to alleviate symptoms and prevent cancer. The results of molecular docking studies showed that the compound

(quercetin, kaempferol, and luteolin) can closely bind the core targets (JUN, Akt1, and MAPK1) [32].

However, this study has limitations. The mechanism of MSM in the treatment of UC through the IL-17 and TNF signaling pathways needs to be further explored. In addition, the mechanism of MSM determined in this study needs to be further analyzed and verified through experimental research. We will further use molecular biology methods to investigate the effects and molecular mechanism of MSM in treating UC and conduct random clinical trials to identify the effectiveness of MSM in patients with UC.

Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was sponsored in part by the Traditional Chinese Medicine Technology Program of Shandong Province of China (2021M068).

References

- [1] J. Burisch, T. Jess, M. Martinato, and P. L. Lakatos, "The burden of inflammatory bowel disease in Europe," *Journal of Crohn's and Colitis*, vol. 7, no. 4, pp. 322–337, 2013.
- [2] A. P. Rivera, G. V. F. Monar, H. Islam et al., "Ulcerative colitis-induced colorectal carcinoma: a deleterious concatenation," *Cureus*, vol. 14, no. 2, p. e22636, 2022.
- [3] C. A. O. Si-Yu, Y. E. Sheng-Jie, W. A. N. G. Wei-Wei, W. A. N. G. Bing, T. Zhang, and P. U. Yi-Qiong, "Progress in active compounds effective on ulcerative colitis from Chinese medicines," *Chinese Journal of Natural Medicines*, vol. 17, no. 2, pp. 81–102, 2019.
- [4] L. Ning, *The Clinical Efficacy and Data Minging of Professor Xu Jing-Fan Prescription for Treating Ulcerative Colitis*, Nanjing University of Chinese Medicine, 2018, <https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CDFDLAST2018&filename=1018124646.nh>.
- [5] H. Qian, Q. Jin, Y. Liu et al., "Study on the multitarget mechanism of Sanmiao pill on gouty arthritis based on network pharmacology," *Evidence-Based Complementary and Alternative Medicine*, vol. 2020, Article ID 9873739, 11 pages, 2020.
- [6] K. Wang, J. Guo, X. Chang, and S. Gui, "Painong-San extract alleviates dextran sulfate sodium-induced colitis in mice by modulating gut microbiota, restoring intestinal barrier function and attenuating TLR4/NF- κ B signaling cascades," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 209, p. 114529, 2022.
- [7] Y. Yu, Z. Wu, Y. Han et al., "Comparison of the effects of essential oil obtained from the crude and bran-processed *Atractylodes lancea* on lipopolysaccharide-induced inflammatory injury of human colonic epithelial cells by downregulating the IKK/NF- κ B signaling pathway," *Evidence-Based*

- Complementary and Alternative Medicine*, vol. 2021, Article ID 5219129, 16 pages, 2021.
- [8] C. Li, M. Wang, J. Sui, Y. Zhou, and W. Chen, "Protective mechanisms of *Agrimonia pilosa* Ledeb in dextran sodium sulfate-induced colitis as determined by a network pharmacology approach," *Acta Biochimica et Biophysica Sinica*, vol. 53, no. 10, pp. 1342–1353, 2021.
- [9] T. Zhu, B. Hu, C. Ye et al., "Bletilla striata oligosaccharides improve ulcerative colitis by regulating gut microbiota and intestinal metabolites in dextran sulfate sodium-induced mice," *Frontiers in Pharmacology*, vol. 13, p. 867525, 2022.
- [10] K. Wang, L. Lei, J. Cao et al., "Network pharmacology-based prediction of the active compounds and mechanism of Buyang Huanwu decoction for ischemic stroke," *Experimental and Therapeutic Medicine*, vol. 22, no. 4, p. 1050, 2021.
- [11] W. Hu, W. Fu, X. Wei, Y. Yang, C. Lu, and Z. Liu, "A network pharmacology study on the active ingredients and potential targets of *Tripterygium wilfordii* hook for treatment of rheumatoid arthritis," *Evidence-based Complementary and Alternative Medicine*, vol. 2019, 2019.
- [12] L. Qin, H. Chen, X. Ding et al., "Utilizing network pharmacology to explore potential mechanisms of YiSui Nong Jian formula in treating myelodysplastic syndrome," *Bioengineered*, vol. 12, no. 1, pp. 2238–2252, 2021.
- [13] J. Liu, J. Liu, X. Tong et al., "Network pharmacology prediction and molecular docking-based strategy to discover the potential pharmacological mechanism of Huai Hua San against ulcerative colitis," *Drug Design, Development and Therapy*, vol. 15, pp. 3255–3276, 2021.
- [14] D. Xu, M. J. Hu, Y. Q. Wang, and Y. L. Cui, "Antioxidant activities of quercetin and its complexes for medicinal application," *Molecules*, vol. 24, no. 6, p. 1123, 2019.
- [15] W. M. Dabeek and M. V. Marra, "Dietary quercetin and kaempferol: bioavailability and potential cardiovascular-related bioactivity in humans," *Nutrients*, vol. 11, no. 10, p. 2288, 2019.
- [16] M. Ashrafzadeh, S. Tavakol, Z. Ahmadi, S. Roomiani, R. Mohammadinejad, and S. Samarghandian, "Therapeutic effects of kaempferol affecting autophagy and endoplasmic reticulum stress," *Phytotherapy Research*, vol. 34, no. 5, pp. 911–923, 2020.
- [17] Y. Bian, Y. Dong, J. Sun et al., "Protective effect of kaempferol on LPS-induced inflammation and barrier dysfunction in a coculture model of intestinal epithelial cells and intestinal microvascular endothelial cells," *Journal of Agricultural and Food Chemistry*, vol. 68, no. 1, pp. 160–167, 2020.
- [18] M. Imran, A. Rauf, T. Abu-Izneid et al., "Luteolin, a flavonoid, as an anticancer agent: a review," *Biomedicine & Pharmacotherapy*, vol. 112, p. 108612, 2019.
- [19] F. Gendrisch, P. R. Esser, C. M. Schempp, and U. Wölfle, "Luteolin as a modulator of skin aging and inflammation," *Biofactors*, vol. 47, no. 2, pp. 170–180, 2021.
- [20] S. Babu and S. Jayaraman, "An update on β -sitosterol: a potential herbal nutraceutical for diabetic management," *Biomedicine & Pharmacotherapy*, vol. 131, p. 110702, 2020.
- [21] X. Zhong, Y. J. Surh, S. G. Do et al., "Baicalein inhibits dextran sulfate sodium-induced mouse colitis," *Journal of Cancer Prevention*, vol. 24, no. 2, pp. 129–138, 2019.
- [22] H. Jang, J. Lee, S. Park et al., "Baicalein mitigates radiation-induced enteritis by improving endothelial dysfunction," *Frontiers in Pharmacology*, vol. 10, p. 892, 2019.
- [23] M. F. Griffin, M. R. Borrelli, J. T. Garcia et al., "JUN promotes hypertrophic skin scarring via CD36 in preclinical in vitro and in vivo models," *Science Translational Medicine*, vol. 13, no. 609, p. eabb 3312, 2021.
- [24] W. Yu, B. Wang, L. Zhou, and G. Xu, "Endoplasmic reticulum stress-mediated p62 downregulation inhibits apoptosis via c-jun upregulation," *Biomolecules & Therapeutics*, vol. 2021, pp. 195–204, 2021.
- [25] A. Alwhaibi, A. Verma, M. S. Adil, and P. R. Somanath, "The unconventional role of Akt 1 in the advanced cancers and in diabetes-promoted carcinogenesis," *Pharmacological Research*, vol. 145, p. 104270, 2019.
- [26] A. Alwhaibi, A. Verma, M. S. Adil, and P. R. Somanath, "The unconventional role of Akt1 in the advanced cancers and in diabetes-promoted carcinogenesis," *Pharmacological Research*, vol. 145, p. 104270, 2019.
- [27] C. Braicu, M. Buse, C. Busuioc et al., "A comprehensive review on MAPK: a promising therapeutic target in cancer," *Cancers (Basel)*, vol. 11, no. 10, p. 1618, 2019.
- [28] N. Tatiya-Aphiradee, W. Chatuphonprasert, and K. Jarukamjorn, "Immune response and inflammatory pathway of ulcerative colitis," *Journal of Basic and Clinical Physiology and Pharmacology*, vol. 30, no. 1, pp. 1–10, 2018.
- [29] H. Schmitt, M. F. Neurath, and R. Atreya, "Role of the IL23/IL17 pathway in Crohn's disease," *Frontiers in Immunology*, vol. 12, p. 622934, 2021.
- [30] J. Holbrook, S. Lara-Reyna, H. Jarosz-Griffiths, and M. F. McDermott, "Tumour necrosis factor signalling in health and disease," *F1000Research*, vol. 8, p. 111, 2019.
- [31] Y. Li, J. Chen, A. A. Bolinger et al., "Target-based small molecule drug discovery towards novel therapeutics for inflammatory bowel diseases," *Inflammatory Bowel Diseases*, vol. 27, Supplement_2, pp. S38–S62, 2021.
- [32] L. Zhang, L. Han, X. Wang et al., "Exploring the mechanisms underlying the therapeutic effect of *Salvia miltiorrhiza* in diabetic nephropathy using network pharmacology and molecular docking," *Bioscience Reports*, vol. 41, no. 6, 2021.

Research Article

Research on Optimization Scheme for Blocking Artifacts after Patch-Based Medical Image Reconstruction

Yan Xu ^{1,2}, Shunbo Hu ², and Yuyue Du ¹

¹College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China

²College of Information Science and Engineering Linyi University, Shandong Province 276005, China

Correspondence should be addressed to Shunbo Hu; hushunbo@lyu.edu.cn and Yuyue Du; yydu001@163.com

Received 9 June 2022; Accepted 12 July 2022; Published 31 July 2022

Academic Editor: Lin Lu

Copyright © 2022 Yan Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to limitations of computer resources, when utilizing a neural network to process an image with a high resolution, the typical processing approach is to slice the original image. However, because of the influence of zero-padding in the edge component during the convolution process, the central part of the patch often has more accurate feature information than the edge part, resulting in image blocking artifacts after patch stitching. We studied this problem in this paper and proposed a fusion method that assigns a weight to each pixel in a patch using a truncated Gaussian function as the weighting function. In this method, we used the weighting function to transform the Euclidean-distance between a point in the overlapping part and the central point of the patch where the point was located into a weight coefficient. With increasing distance, the value of the weight coefficient decreased. Finally, the reconstructed image was obtained by weighting. We employed the bias correction model to evaluate our method on the simulated database BrainWeb and the real dataset HCP (Human Connectome Project). The results show that the proposed method is capable of effectively removing blocking artifacts and obtaining a smoother bias field. To verify the effectiveness of our algorithm, we employed a denoising model to test it on the IXI-Guys human dataset. Qualitative and quantitative evaluations of both models show that the fusion method proposed in this paper can effectively remove blocking artifacts and demonstrates superior performance compared to five commonly available and state-of-the-art fusion methods.

1. Introduction

In recent years, with the development of deep learning, medical imaging has gradually become one of the most promising fields of artificial intelligence. Especially in the aspects of localization and detection, recognition and classification, lesion segmentation, registration, and fusion in medical images, deep learning algorithms have played a crucial role in assisting doctors to diagnose accurately and efficiently [1]. Medical images have the characteristics of large image pixels [2], when using neural networks to process them, due to the limitations of computer video cards and video memory, directly using the original image as an input will often lead to insufficient graphics memory. The usual processing method involves slicing the original image, reconstructing the image patches, and then splicing the image patches. In deep learning, patch-based training methods

have been widely used because they can save GPU memory, are not affected by insufficient training data, and can obtain better local performance than other methods [3]. However, there is also a grid artifact phenomenon at the edge of patches when they are spliced for image reconstruction as shown in Figure 1 [3], these artifacts are called image blocking artifacts.

To solve this problem, Yang et al. [4] and Hu et al. [5] obtained the final predicted deformation field by averaging the overlapping regions of patches when reconstructing the deformation field in medical image registration. In brain image registration using dual-supervised fully convolutional networks, as proposed by Fan et al. [6], only the deformation field in the central region of the patch is estimated, the input patch size is $64 \times 64 \times 64$, and the output DDF (dense displacement field) is $24 \times 24 \times 24$. The MINScnn model proposed by Müller and Kramer [7] used the average fusion



FIGURE 1: Image blocking artifacts after patch-based image reconstruction.

method for the two overlapping patches in the segmentation field. In the patchify method proposed by Wu, the overlapping region is covered with the next patch [8]. The EDW (Exponential-Distance-Weighted) method proposed by Wu et al. [9] uses an exponential function to convert the distance between the point of the overlapping area and the center point of the patch into a weight coefficient and reconstruct the predicted deformation field by weighting. This method achieved good results.

During image reconstruction, most of the above methods reduce image blocking artifacts to a certain extent by reducing the step size or by performing estimates for the central area of the patch, but blocking artifacts are still obvious after image reconstruction. Considering the uncertainty of patch edge region prediction, we propose to use the truncated Gaussian function as a weighting function to convert the Euclidean distance from the points in the overlapping regions to the center point of the patch into a weight coefficient. The value of the weight coefficient will decrease with increasing distance, reducing the predicted value in the patch edge region. Finally, the reconstructed image will be obtained by weighted fusion. For distance calculation, we choose Euclidean distance as the distance measure [9]. This method can be applied to any patch-based deep learning network without modifying the network structure and loss function. Compared with the image reconstruction methods used in the above literature, the patch-based fusion method in this paper shows better performance in both qualitative and quantitative evaluations.

The paper is organized as follows. In Section 2, we briefly review the popularly available methods for extracting

patches in deep learning. Section 3 describes the proposed patch fusion method in detail. In Section 4, we evaluate the proposed patch fusion algorithm on synthetic and real datasets and present the experimental results. At the same time, a comparison with other state-of-the-art methods is also given. Conclusions and future works are discussed in Section 5.

2. Common Methods for Extracting Patches

Nonoverlapping slicing is performed directly on the image, as shown in Figure 2(a). Assuming that the image size is 512×512 , the image is sliced into four equal parts, and the size of each patch is 256×256 . The patches are input to the network to obtain the reconstructed image patches, and then, they are stitched together. Because the boundary information of each patch is inconsistent after reconstruction, the stitched image has obvious blocking artifacts.

In overlapping slicing, the image is sliced into patches, so that there are overlapping areas between the patches, as shown in Figure 2(b). The red region is the region where the two patches overlap. Assuming that the image size is still 512×512 and the red overlapping region is 8 pixels wide, then the size of the input patch is 260×260 . After reconstruction, in situ pixel stitching is performed, and the red overlapping part is taken as the average value after stitching. This scheme has a better effect in removing the blocking artifact. However, since the average weighting method is adopted for the overlapping region, the weight of the pixels in the overlapping region is the same no matter how far away from the center point, so grid artifacts appear.

When padding is used around the image block, a patch is obtained as shown in block 1 in Figure 2(c). First, the padding is made around the image, the image is sliced after padding, and the cyan line is the central axis of the image. For example, in block 1, the orange part is the size of the original image block, and the surrounding red area consists of the expanded pixels. If the original image block is 256×256 and the red area occupies a width of 8 pixels, the input patch size is 272×272 ($256 + 8 * 2$) for reconstruction. Then, the reconstructed image block is cut with a width of 8 pixels around to obtain a patch with a size of 256×256 , and finally, it is stitched in sequence. For image block 1, the red part is removed after reconstruction. For image block 2, the blue part on the left needs to be cut off. This scheme is better than the previous two schemes in removing blocking artifacts, but it will be affected by stride. The smaller the stride is, the better the effect of removing blocking artifacts. However, the smaller stride, the greater the amount of calculations. Since the weight of the entire estimated central region is the same, the image blocking artifacts will still occur.

3. Methods and Theories

3.1. *One-Dimensional Gaussian Distribution.* The one-dimensional Gaussian function has the following form [10]:

$$f(x) = ae^{-(x-b)^2/2c^2}, \quad (1)$$

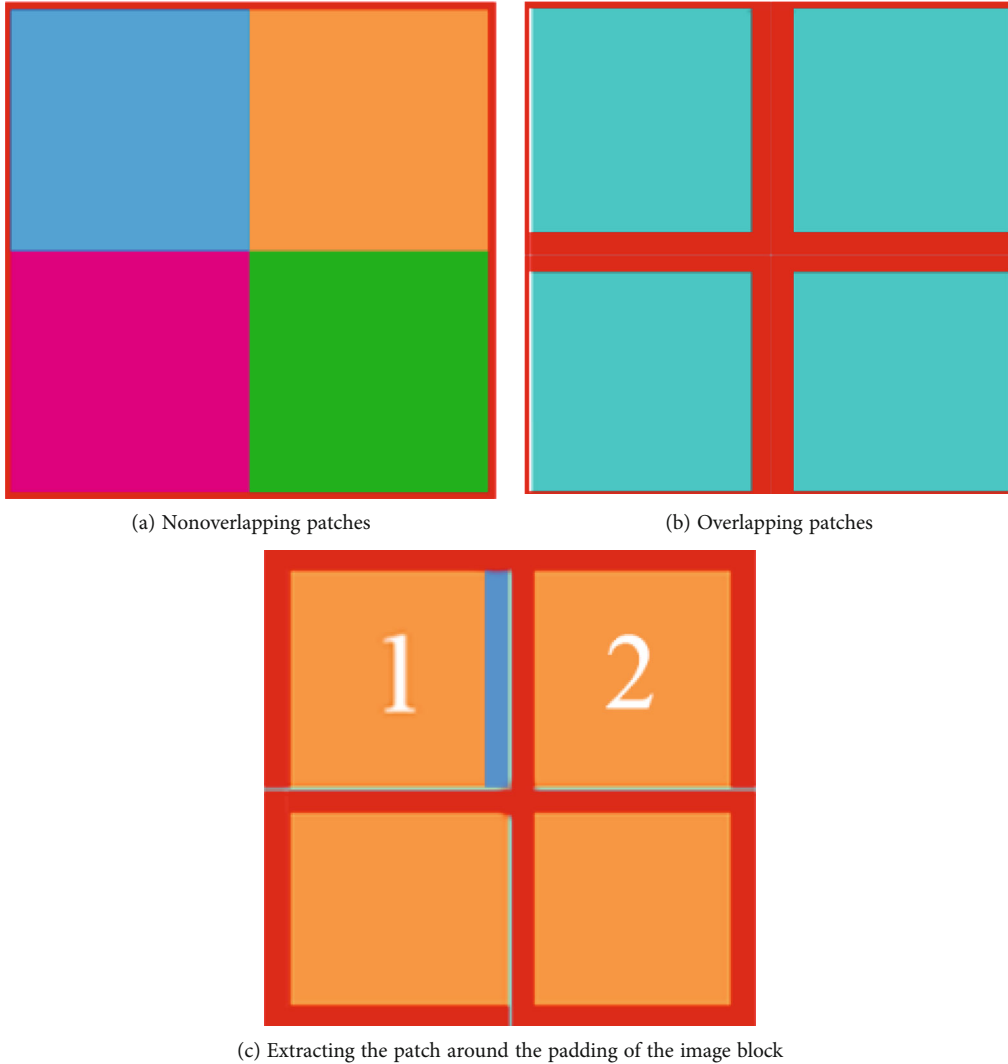


FIGURE 2: Three methods for extracting patches of images.

where a, b , and c are arbitrary real numbers. The Gaussian curve graph is a characteristically symmetrical “bell curve” shape, where a represents the height of the peak of the curve, b is the position of the center of the peak, and c is called the standard variance, which is a parameter used to control the width of the “bell.” Let $a = 1, b = \mu$ (mean value) and $c = \sigma$ (standard variance); the Gaussian distribution can be obtained as follows:

$$G(x) = e^{-(x-\mu)^2/2\sigma^2}. \tag{2}$$

Its function curve is shown in Figure 3.

The horizontal axis represents the possible values, the vertical axis represents the probability distribution density $G(x)$, the expected value μ of the normal distribution determines the center of the curve, and the standard variance σ determines the magnitude of the distribution. In Figure 3, the settings are $a = 1, b = \mu = 0$. Hence, the peak heights and center positions of all curves are the same. With increas-

ing standard variance, the graph becomes wider, the distribution becomes dispersed, and the area of the graph becomes increasingly larger. The closer the value of x is to the center, the greater the value of $G(x)$. Therefore, we can use this feature of the Gaussian function by assigning the distance from the pixel’s location to the center point of the patch as the input x to the Gaussian function and calculating the weight coefficient of each pixel point through Gaussian function [11].

3.2. The Proposed Patch Fusion Method. In Figure 4, the blue part is patch 1, B is the central point of patch 1, the orange part is patch 2, C is the central point of patch 2, and the red part is the overlapping part of these two patches. It can be seen that pixel A is closer to the central point B in patch 1 in the overlapping area, so its predicted value in patch 1 is more accurate. To reduce the weight of the predicted value of the pixel located at the boundary, we convert the Euclidean distance between the pixel and the center of the patch to the weight coefficient of the

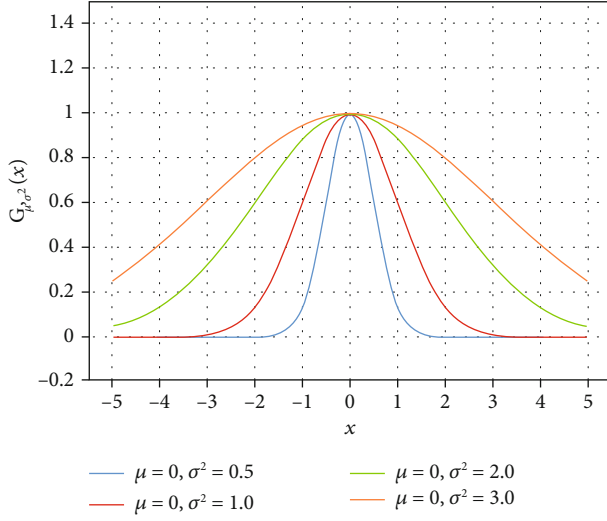


FIGURE 3: The Gaussian function curve ($a = 1, b = \mu, c = \sigma$).

pixel through the Gaussian truncation function. The formula is defined as follows [11, 12]:

$$W(y) = \begin{cases} \frac{1}{z} e^{-|y|^2/2\sigma^2}, & \text{for } |y| \leq \rho, \\ 0, & \text{else,} \end{cases} \quad (3)$$

where σ represents the standard variance, z represents the normalization factor of the normalized Gaussian kernel, and ρ represents the radius to measure the size of the local region. Therefore, in Figure 4, assuming that the predicted value of point A in patch 1 is P_{AB} and the predicted value in patch 2 is P_{AC} , the final predicted value of point A is $P_A = W(y1)P_{AB} + W(y2)P_{AC}$.

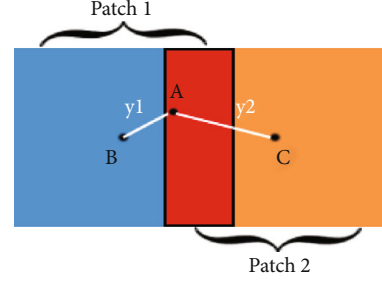


FIGURE 4: Schematic diagram of patch overlap. The red part is the overlapping area, and B and C are the center points of patch 1 and patch 2, respectively. A is a point in the overlapping area, and $y1$ and $y2$ are the distances from point A to the two center points.

We take a two-dimensional graph as an example. Assuming that the image size is $M \times N$, the patch size is $h \times w$, and the stride size is s , n patches can be generated, and the value of n can be calculated by

$$n = \left\lfloor \frac{M - (h - s)}{s} \right\rfloor \left\lfloor \frac{N - (w - s)}{s} \right\rfloor. \quad (4)$$

In these n patches, the Gaussian distribution of the distance between each pixel point and the center point of the patch is the same, which is calculated by Formula (2). The Gaussian function value corresponding to the pixel point (i, j) in the k th patch can be recorded as $G_k^{(i,j)}(y_k)$, where y_k is the Euclidean distance from this point to the central point of the patch. Therefore, Formula (3) can be expressed by $G_k^{(i,j)}(y_k)$.

$$W(y_k) = \begin{cases} \frac{1}{z} G_k^{(i,j)}(y_k), & \text{for } |y_k| \leq \rho, \\ 0, & \text{else,} \end{cases} \quad (5)$$

$$m_i = \begin{cases} \left\lfloor \frac{i}{s} \right\rfloor + 1, & i < h, \\ \frac{h}{s}, & h \leq i \leq \left\lfloor \frac{M - (h - s)}{s} \right\rfloor s - 1, \\ \left\lfloor \frac{(\lfloor (M + s)/s \rfloor - 1)s - i}{s} \right\rfloor, & \left\lfloor \frac{M - (h - s)}{s} \right\rfloor s - 1 < i < \left(\left\lfloor \frac{M + s}{s} \right\rfloor - 1 \right) s, \\ 0, & \left(\left\lfloor \frac{M + s}{s} \right\rfloor - 1 \right) s \leq i. \end{cases} \quad (6)$$

Assuming that the pixel point $O(i, j)$ overlaps m times, $m = m_i \times m_j$, $i \in [0, M - 1]$, $j \in [0, N - 1]$, then m_i can be calculated by Formula (6) [9], and the edge part that cannot

be obtained by the patch is discarded. The value of m_j can be obtained by replacing M in Formula (6) with N . Assuming that the m predicted values of the pixel point are $P^{(i,j)}$

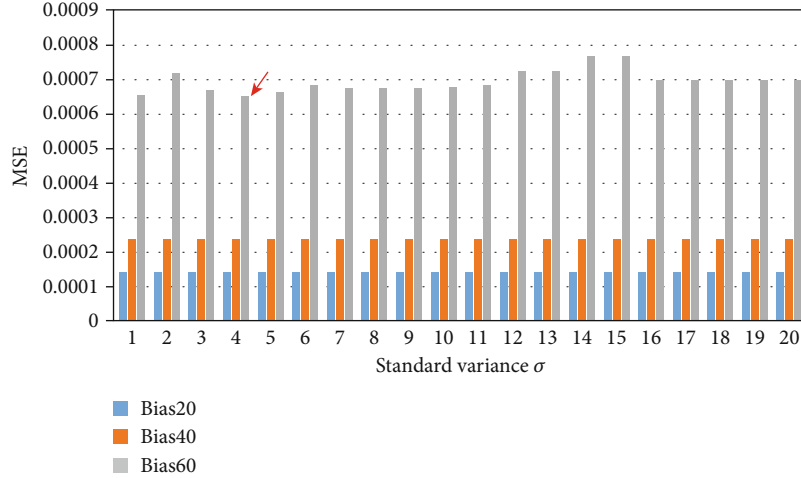


FIGURE 5: MSE values of different standard variance on different INU level datasets.

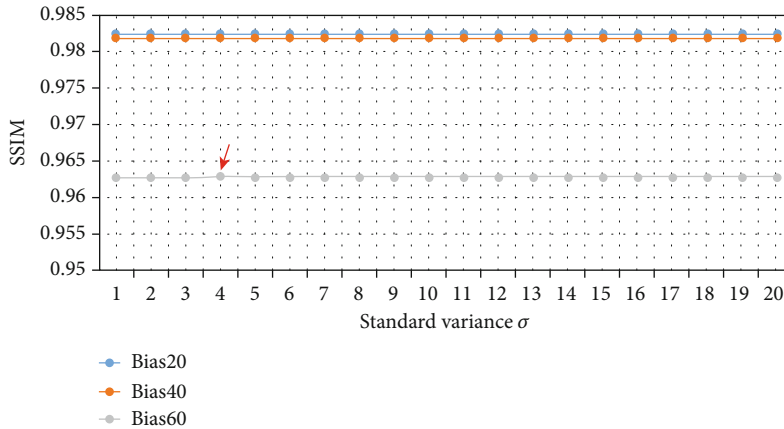


FIGURE 6: SSIM values of different standard variance on different INU level datasets.

$= \{p_1^{(i,j)}, p_2^{(i,j)}, \dots, p_m^{(i,j)}\}$ and the weight coefficients of the m predicted values calculated by Formula (5) are $W^{(i,j)} = \{W_{y_1}^{(i,j)}, W_{y_2}^{(i,j)}, \dots, W_{y_m}^{(i,j)}\}$, then the final predicted value of the pixel point $O(i, j)$ is

$$\hat{p}^{(i,j)} = p^{(i,j)} \cdot W^{(i,j)} = \sum_{k=1}^m p_k^{(i,j)} W_{y_k}^{(i,j)}, \quad (7)$$

where $\sum_{k=1}^m W_{y_k}^{(i,j)} = 1$.

4. Experimental Results and Analysis

To evaluate the effective fusion ability of the algorithm, we employed DN-RESnet (a deep convolutional neural network consisting of several residual blocks) [13] as the training model. The bias field removal experiment was carried out on the simulation database BrainWeb [14–15] and the real dataset HCP (Human Connectome Project) [16]. Meanwhile, we also used the DnCNN (De-noising Convolutional Neural Net-

work) denoising model [17] for validation on the IXI-Guys (<http://brain-development.org/ixidataset>) dataset [18].

4.1. Validation of the Proposed Fusion Method in the Bias Removal Model

4.1.1. Results Obtained on the Simulation Database BrainWeb. In the Simulated Brain Database (SBD) [15], the size of each MRI brain image volume is $181 \times 217 \times 181$ voxels with a resolution of $1 \times 1 \times 1$ mm³. 2D MRI slices were extracted from the volumes for training and testing, with different intensity nonuniformity (INU) levels (20%, 40%, and 60%) on the axial plane, respectively. We used rotation, mirror imaging, translation, and other methods to expand the samples. Finally, 604 slices were obtained for training and 30 slices for testing from the volume of each INU level. In order to obtain fewer redundant regions, we cropped the edges of each slice, so the slice size becomes 210×180 . In the training stage, we set the patch size to 60×60 , and obtained 99 patches from each image by using a sliding window with a stride of 15×15 . Finally, we obtained 59796 patches.

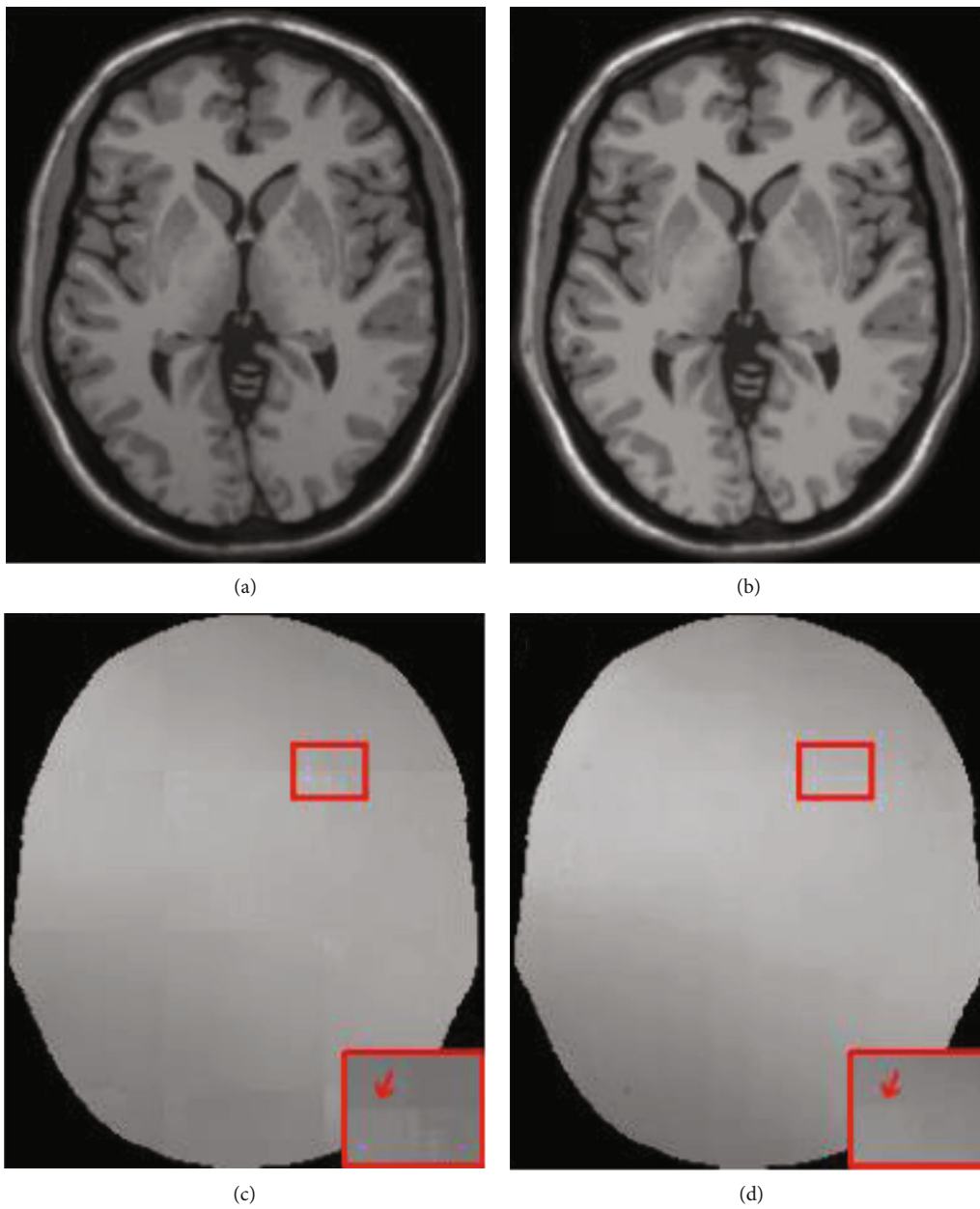


FIGURE 7: Continued.

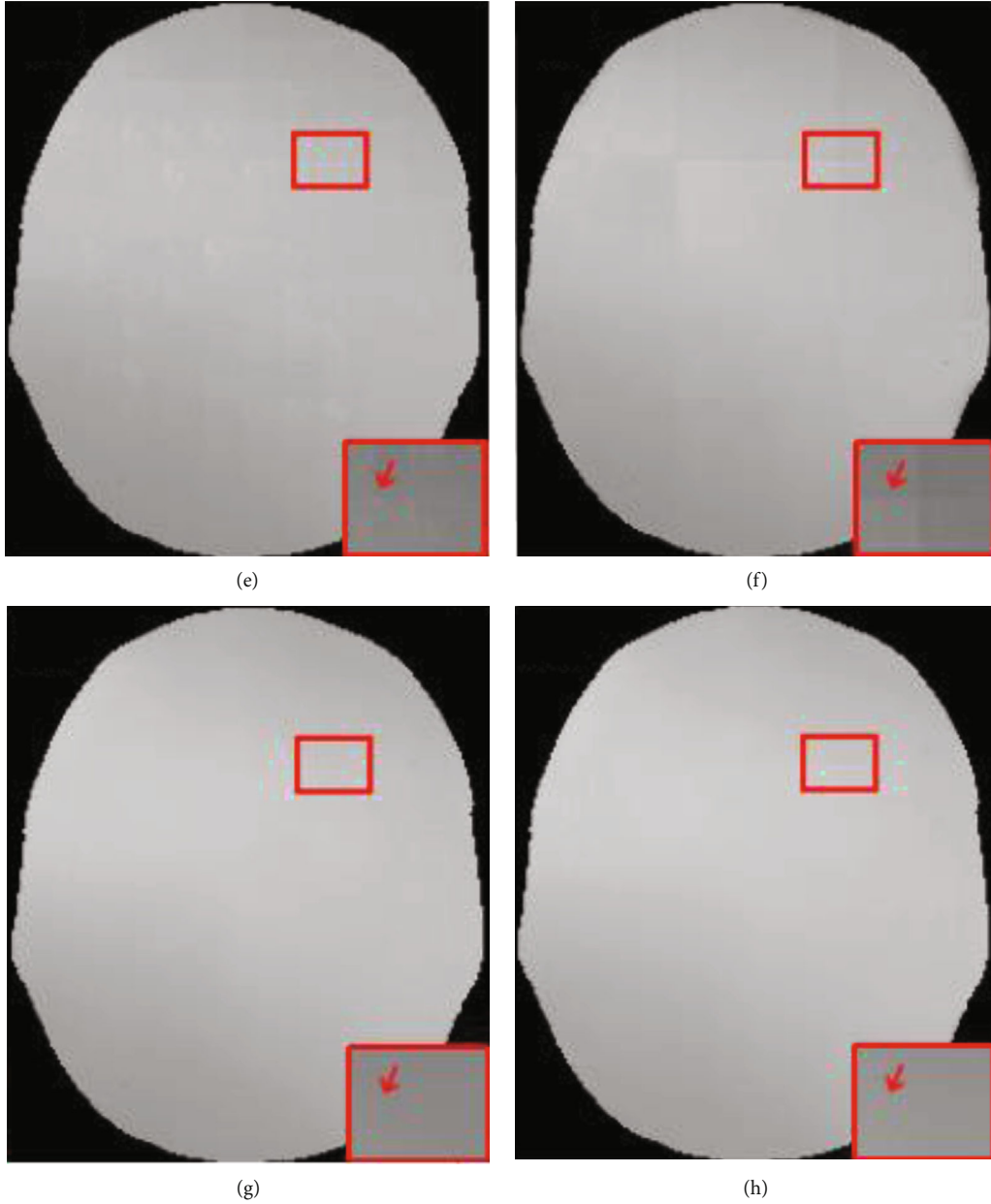


FIGURE 7: Contaminated image, real image, and bias field images obtained by different fusion methods from the BrainWeb database: (a) original image (INU = 40%, noise level = 0%); (b) real image; (c) NPS; (d) AAW; (e) MIScnn; (f) pathify; (g) EDW; (h) proposed method.

In our algorithm, we needed to set the value of the standard variance σ . We conducted experiments on datasets with different (INU) levels (20%, 40%, and 60%) setting the value of σ in the range of 1 to 20. The mean square error (MSE) and the structural similarity index (SSIM) were used as the quantitative evaluation criteria after image reconstruction. The results are shown in Figures 5 and 6.

It can be seen from Figures 5 and 6 that when the INU level is not high (20%, 40%), no matter what the value of the standard variance σ was, it had little effect on the MSE and SSIM after image reconstruction. However, when the INU level increased to 60%, MSE and SSIM values fluctuated for different values of σ . When $\sigma = 4$, MSE and SSIM showed better performance. At the same time, we also performed the

TABLE 1: MSE values between the real image and the corrected images obtained by different fusion methods.

Methods	n3-20	n0-40	n0-60
NPS	13.60×10^{-4}	2.139×10^{-4}	8.228×10^{-4}
AAW [5]	16.95×10^{-4}	2.054×10^{-4}	6.575×10^{-4}
MIScnn [7]	48.57×10^{-4}	2.108×10^{-4}	7.079×10^{-4}
Pathify [8]	13.53×10^{-4}	2.110×10^{-4}	7.985×10^{-4}
EDW [9]	6.593×10^{-4}	2.053×10^{-4}	7.887×10^{-4}
Proposed	4.193×10^{-4}	2.051×10^{-4}	6.212×10^{-4}

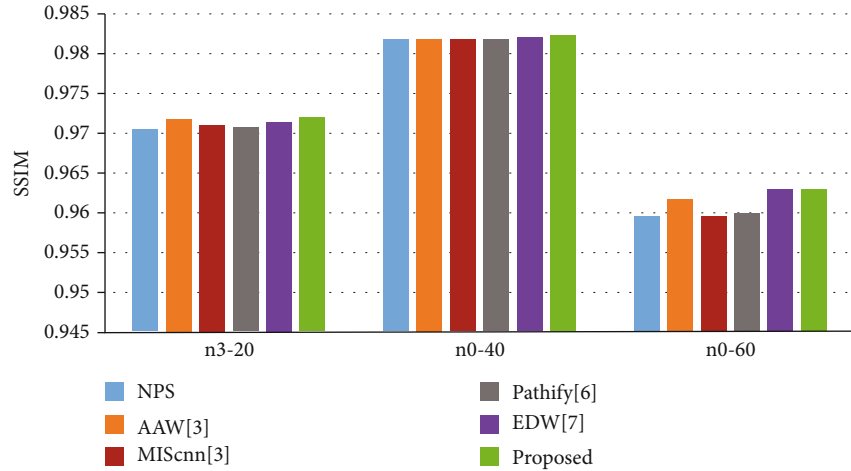


FIGURE 8: SSIM values between the real image and the corrected images obtained by different fusion methods.

same experiments on the real dataset HCP [16] and reached at consistent conclusions. Therefore, in the bias removal experiment of our algorithm, the standard variance σ was set to 4.

To evaluate the fusion ability of the proposed method after image reconstruction, we compared the proposed method with popularly available and state-of-the-art nonoverlapping patch splicing (NPS) method, arithmetic average weighted method (AAW) [5], MIScnn [7], Pathify [8], and Exponential-Distance-Weighted method (EDW) [9] on three datasets with different INU levels from BrainWeb. The dataset with INU = 20% and noise level = 3% was denoted as n3-20, the dataset with INU = 40% and noise level = 0% was denoted as n0-40, and the dataset with INU = 60% and noise level = 0% was denoted as n0-60. The experimental results are shown in Figure 7, Table 1, and Figure 8.

In Figure 7, we show the bias field images fused by different methods. The bias field images of NPS (Figure 7(c)) and Pathify (Figure 7(f)) demonstrate an obvious blocking artifact phenomenon, while the blocking artifact phenomena of AAW (Figure 7(d)) and MIScnn (Figure 7(e)) are greatly improved, but the seam line between patches are vaguely visible. The AAW method is greatly affected by the stride. The larger the stride is, the more obvious the blocking artifact phenomenon is. However, in Figure 7(e), the MIScnn method also shows the phenomenon of uneven bias field, indicating the inaccurate prediction of the edge region. EDW (Figure 7(g)) and our method (Figure 7(h)) consider the uncertainty of edges and assign different weights to the central region and the edge region, effectively eliminating the grid artifacts. From Table 1 and Figure 8, it is observed that on the n0-40 dataset, our method has little difference from the EDW method, but on the n3-20 and n3-60 datasets, our method shows better performance in the quantitative evaluation of MSE and SSIM than EDW, indicating that our method has better generalization ability.

4.1.2. The Results from the Real Database HCP. HCP [16] real human brain dataset contains T1-W and T2-W struc-

tural MR images of healthy adults from Siemens Skyra 3T scanner, along with the corresponding bias fields images. For detailed information, please refer to [19]. In the experiment, we randomly selected the T1-W structural images of 10 patients from the dataset. The size of each MRI brain image was $260 \times 311 \times 260$ voxels, and the resolution was $0.7 \times 0.7 \times 0.7$ mm³. We randomly selected 1000 slices from these 10 brain image volumes. To better train the model, we augmented the data with rotation and mirroring methods and finally obtained 3200 training samples and 200 testing samples. Similarly, to speed up the training process and obtain fewer redundant regions, we cropped the image edges of each sample with size 300×240 . In the training stage, we set the patch size to 60×60 and obtained 63 patches from each image by using a sliding window with a stride of 30×30 . Finally, we obtained 201600 patches.

In Figure 9, we show the bias field images from HCP fused by different methods and the corrected images. It can be seen that the bias field images fused by the NPS, AAW, MIScnn, and Pathify methods have obvious seam lines at the junction between patches, which indicates that the prediction at the boundary region is not accurate. EDW and our proposed method can effectively reduce grid artifacts by assigning different weights to pixels in different regions. In the enlarged red box of the corrected image, it can be seen that there are no grid artifacts in Figures 9(g) and 9(h). Meanwhile, it can be seen from Figures 9(g) and 9(h) that the bias field obtained by our method is smoother, and the intensity of each tissue area of the corrected image looks more consistent, which is obvious in the white matter region of Figure 9(h). This conclusion is also verified in Table 2. Our method can obtain a lower MSE value and a higher SSIM value.

In the AAW method, the stride size has a great influence on the patch fusion effect. Therefore, on the HCP dataset, we verified the effect of different stride sizes on the fusion effect of AAW [5], EDW [9], and the method proposed in this paper. Table 3 lists the number of patches obtained on each image with different stride sizes and the fusion time of these three methods. At the training stage, we used a sliding window

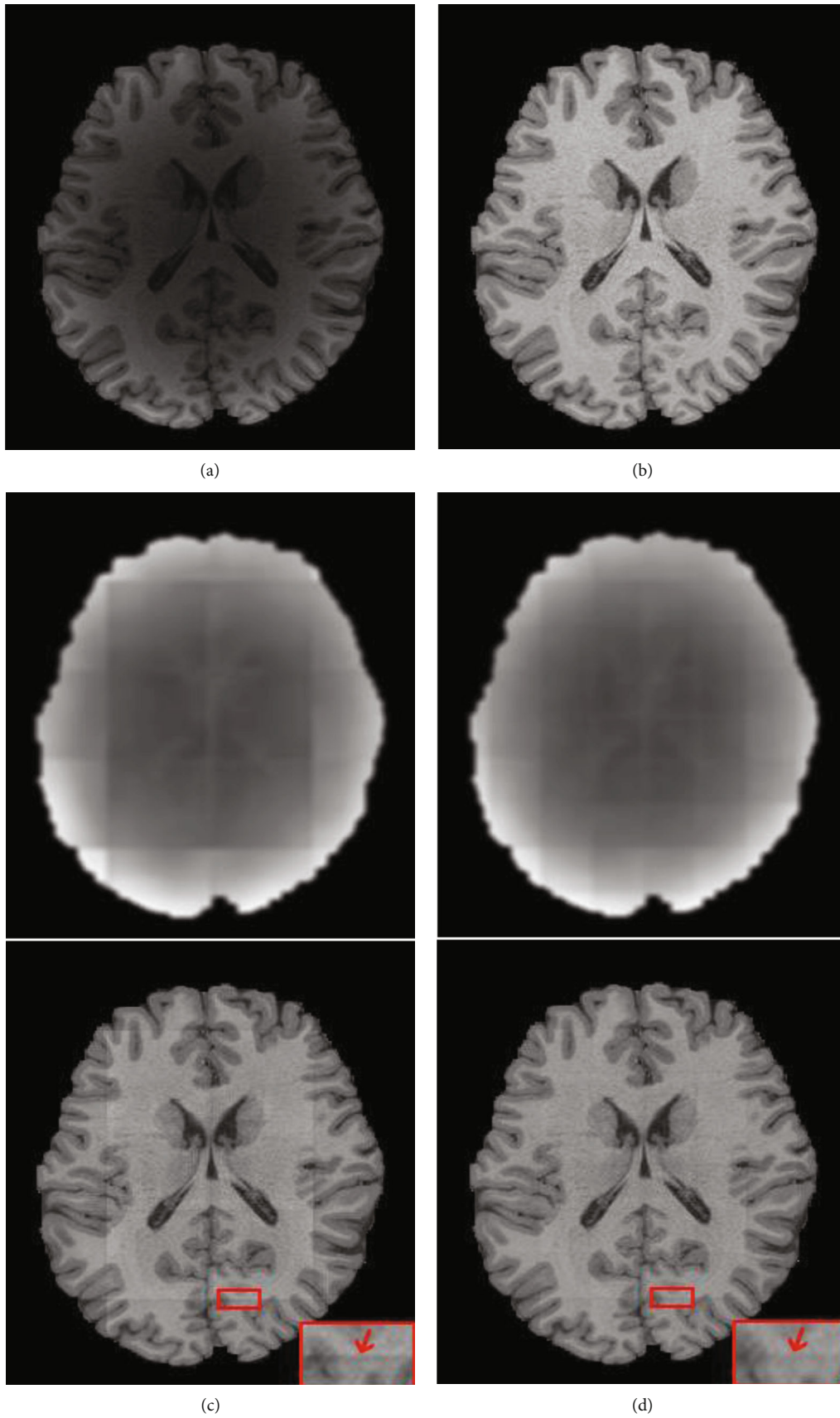


FIGURE 9: Continued.

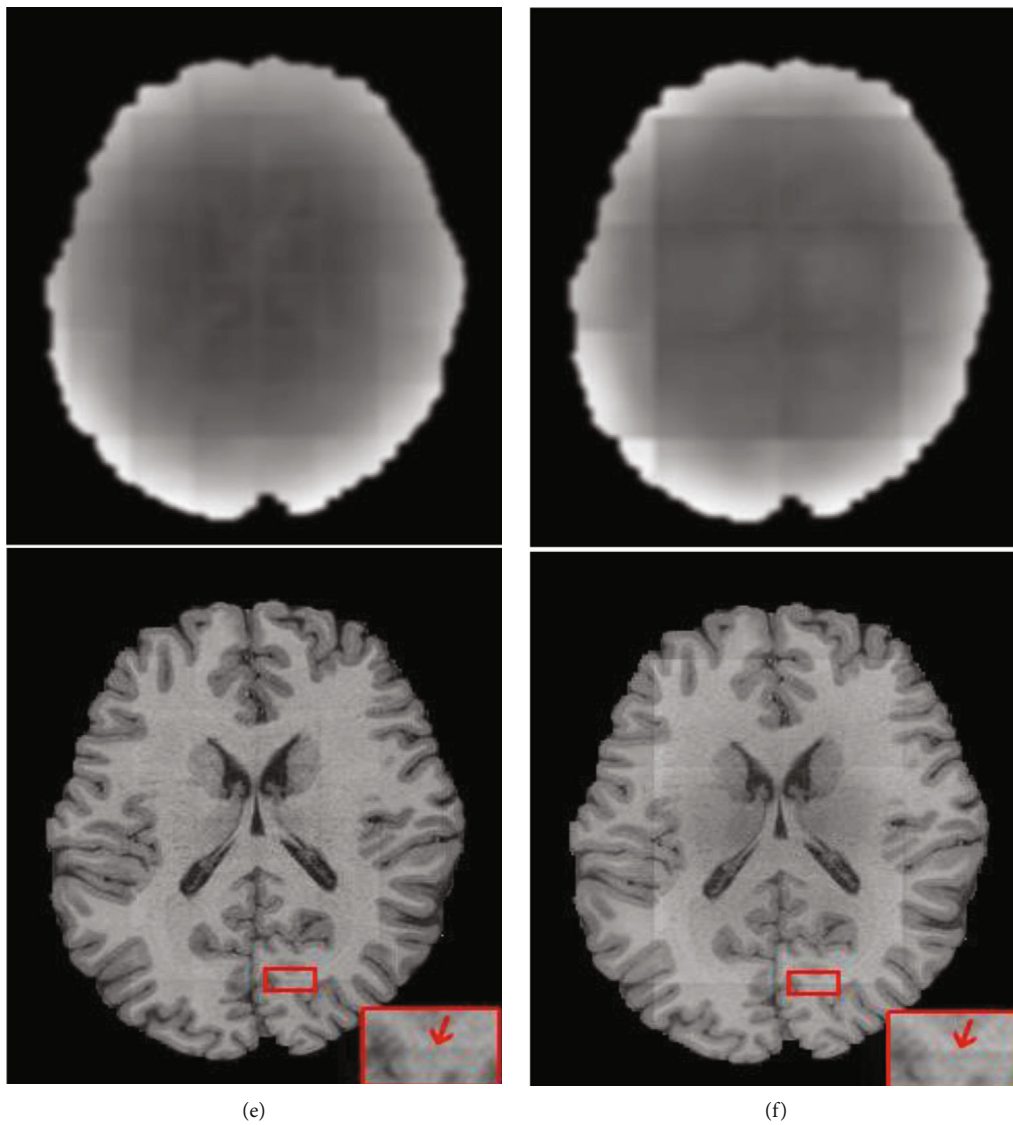


FIGURE 9: Continued.

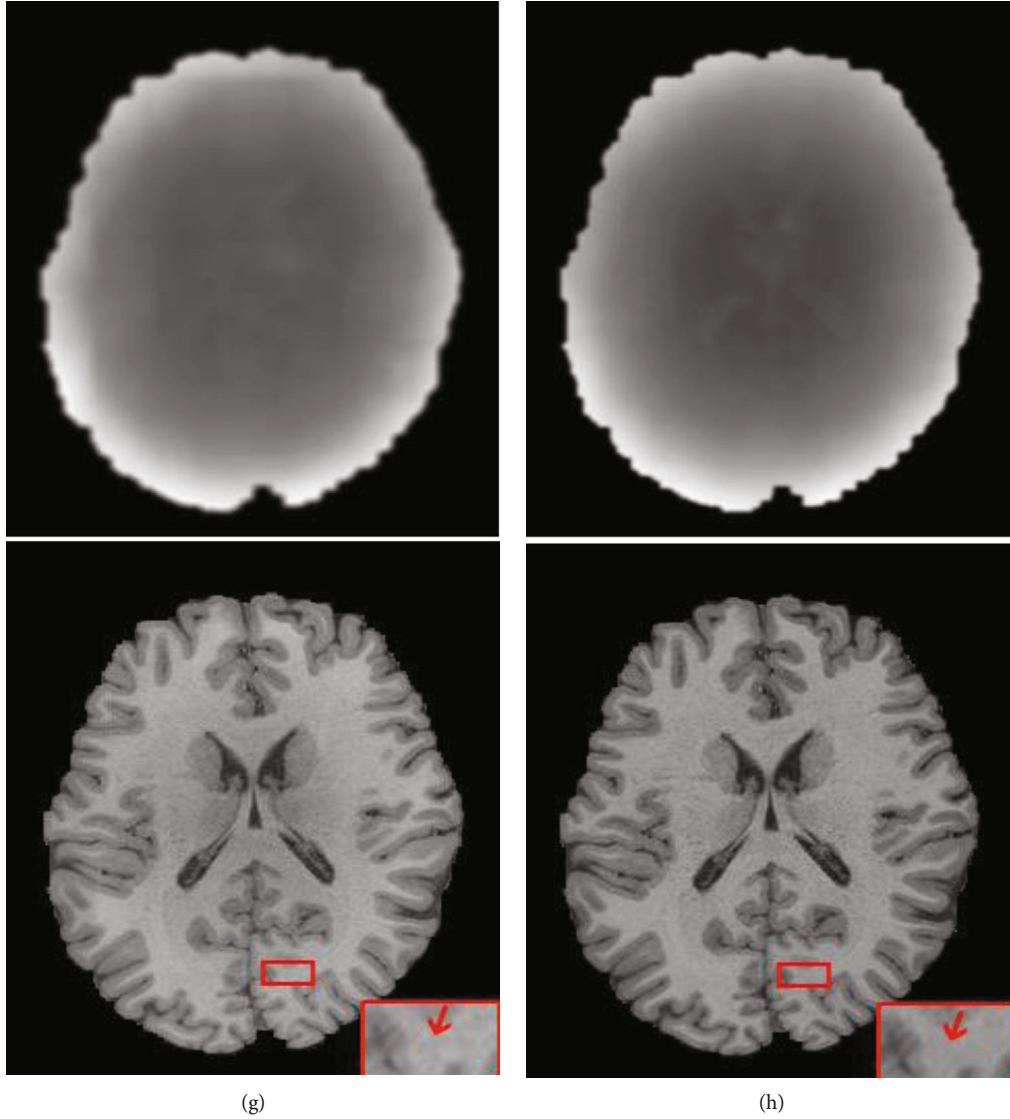


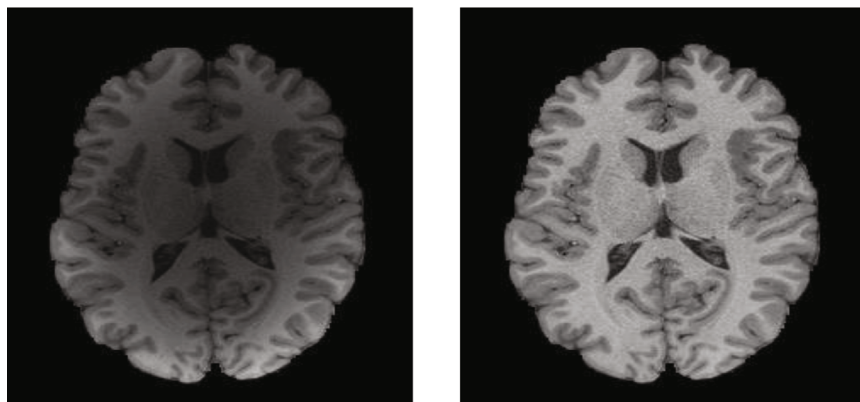
FIGURE 9: Bias field images obtained by different fusion methods on real HCP data and corresponding corrected images: (a) original image, (b) real image, (c) NPS, (d) AAW, (e) MIScnn, (f) pathify, (g) EDW, and (h) proposed method.

TABLE 2: The MSE values, SSIM values and synthesis time between the real image and the corrected image obtained by different fusion methods.

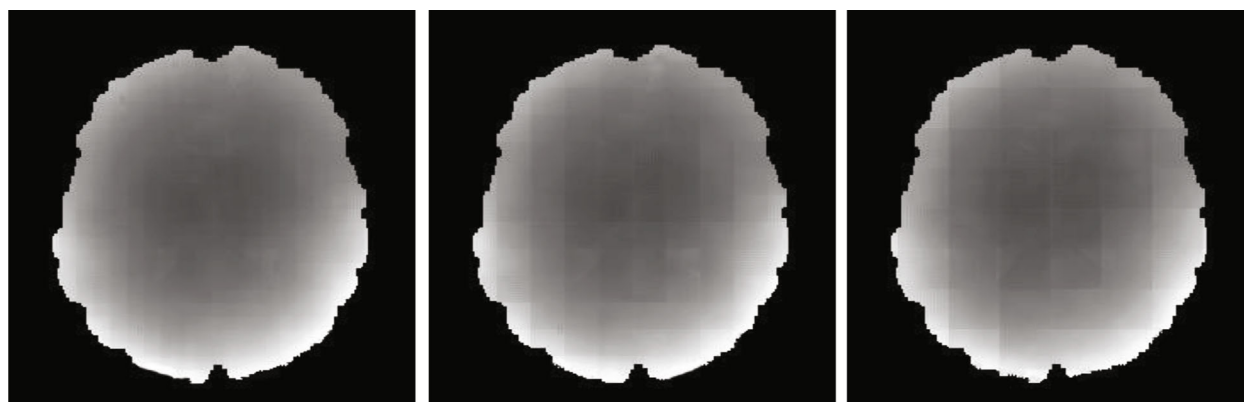
Method	NPS	AAW [5]	MIScnn [7]	Pathify [8]	EDW [9]	Proposed
MSE	$1.233 * 10^{-3}$	$1.130 * 10^{-3}$	$1.136 * 10^{-3}$	$1.468 * 10^{-3}$	$1.139 * 10^{-3}$	$1.104 * 10^{-3}$
SSIM	0.9882	0.9909	0.9913	0.9869	0.9936	0.9942
Time (s)	0.25	0.40	0.41	0.39	1.79	1.81

TABLE 3: The number of patches extracted from each image with different strides and the fusion time.

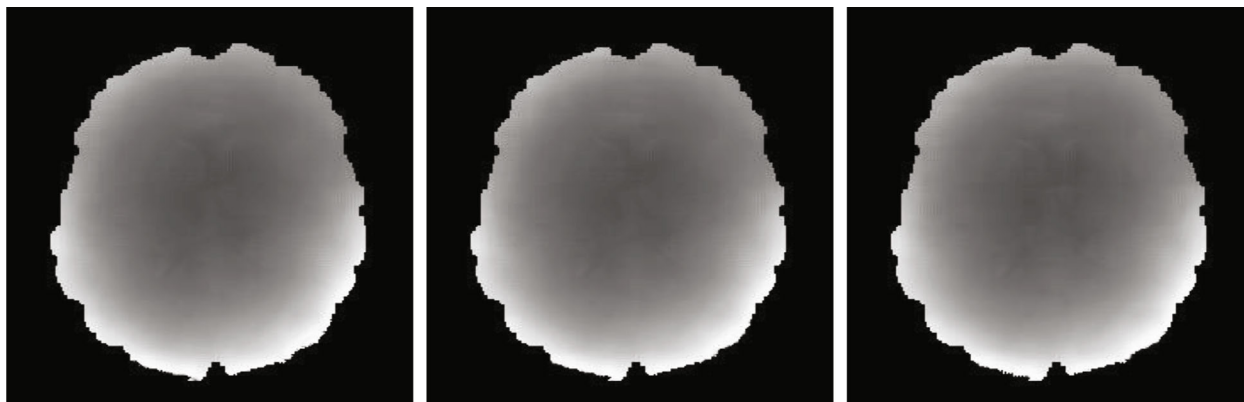
Stride	10×10			20×20			30×30		
Patch number	475			130			63		
Fusion time (s)	AAW	EDW	Proposed	AAW	EDW	Proposed	AAW	EDW	Proposed
	1.77	6.16	6.20	0.61	2.44	2.45	0.37	1.79	1.81



(a)



(b)



(c)

FIGURE 10: Continued.

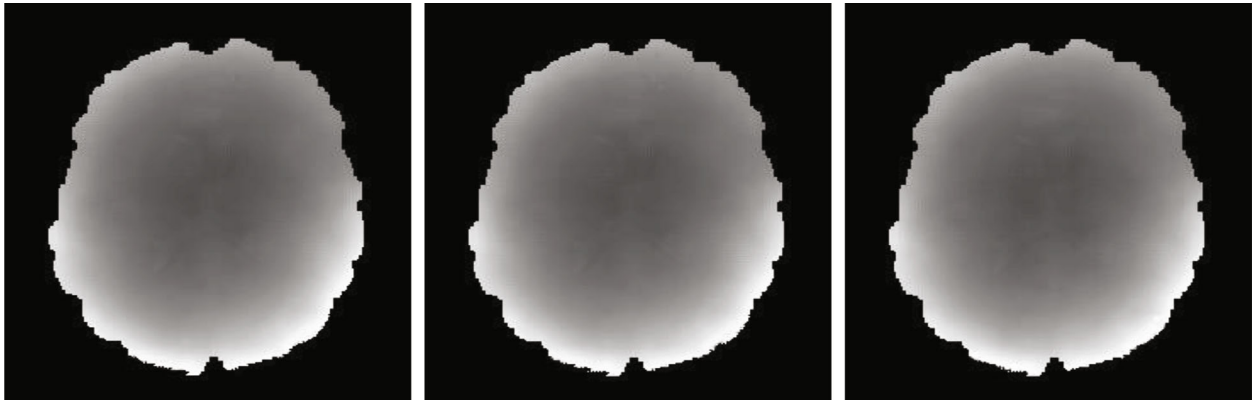


FIGURE 10: Bias field images obtained by fusing patches with AAW, EDW, and the proposed methods under different strides. (a) The original image and the real image. (b) The result of AAW. (c) The result of EDW. (d) The result of the proposed method. The step sizes from left to right are 10×10 , 20×20 , 30×30 .

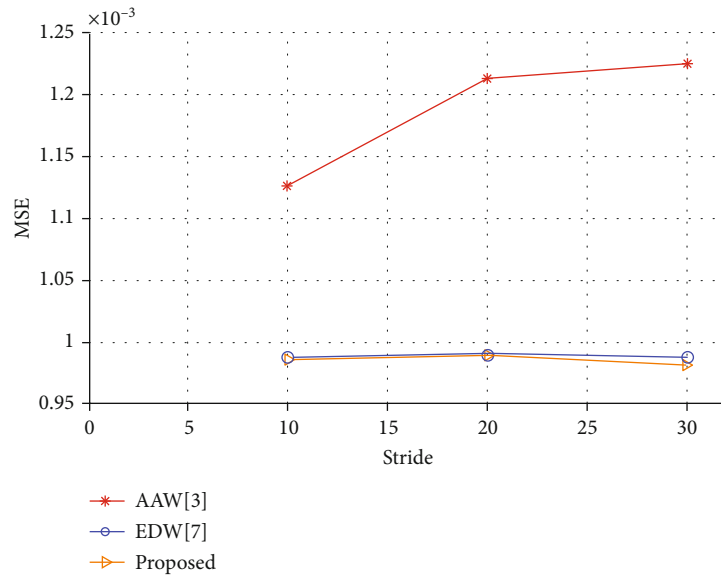


FIGURE 11: MSE values of AAW, EDW, and the proposed methods using different strides on the HCP dataset.

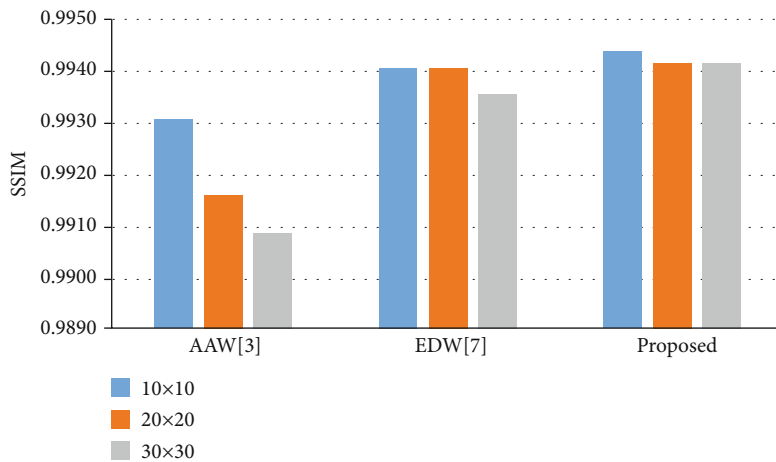


FIGURE 12: SSIM values of AAW, EDW, and the proposed methods using different strides on the HCP dataset.

TABLE 4: Average MSEs, PSNRs, and SSIMs of different fusion methods on the IXI-Guys dataset.

Method	NPS	AAW [5]	MIScnn [7]	Pathify [8]	EDW [9]	Proposed
MSE	$4.574 * 10^{-4}$	$3.034 * 10^{-4}$	$3.024 * 10^{-4}$	$3.064 * 10^{-4}$	$2.986 * 10^{-4}$	$2.971 * 10^{-4}$
PSNR	32.85	35.64	35.67	35.61	35.71	35.76
SSIM	0.9789	0.9833	0.9838	0.9835	0.9841	0.9843

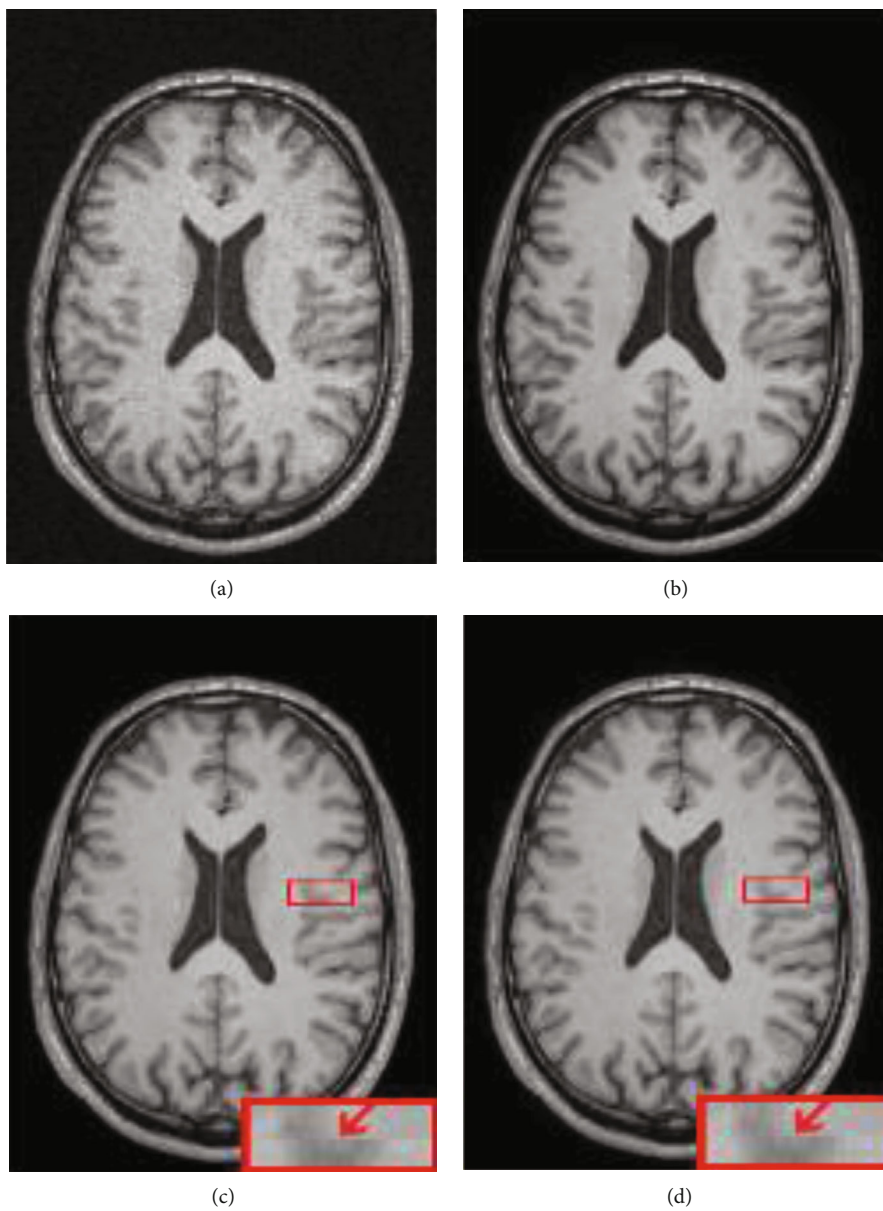


FIGURE 13: Continued.

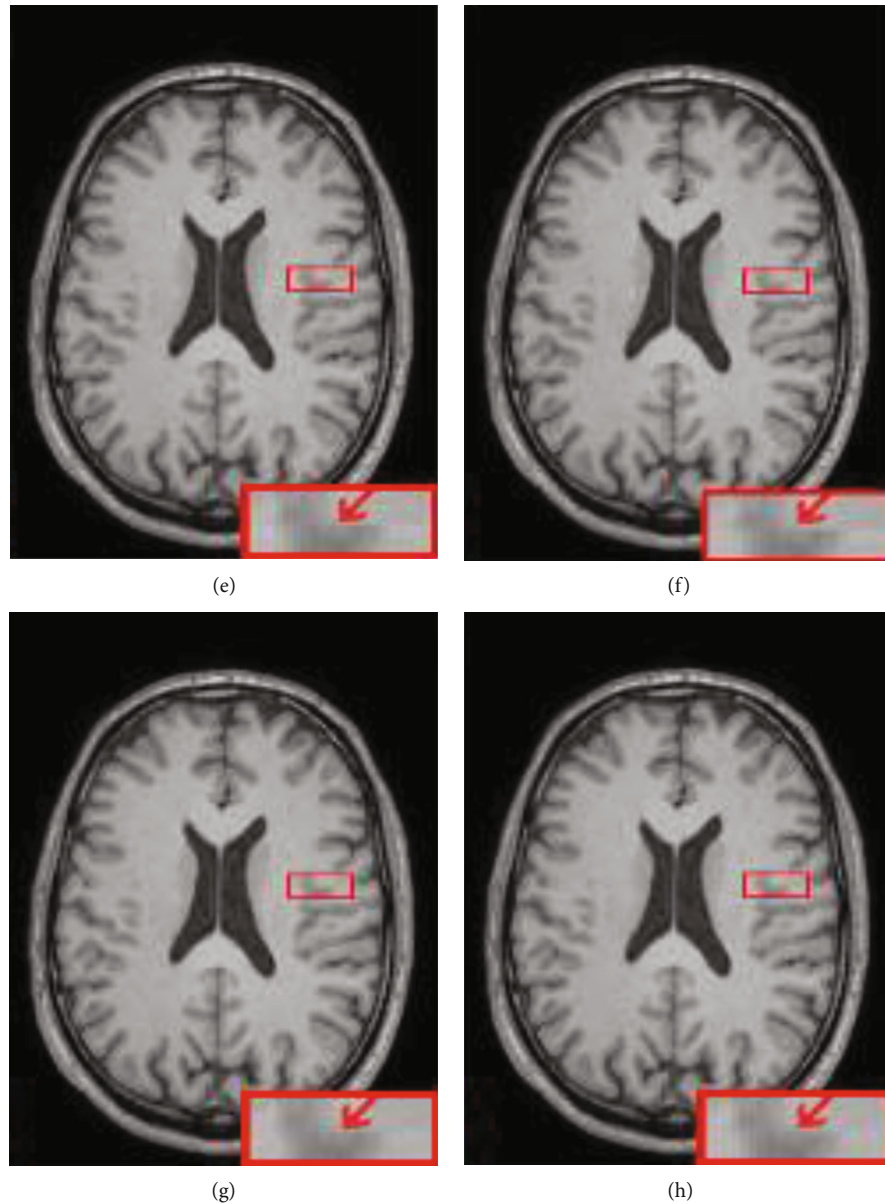


FIGURE 13: Images from the IXI-Guys dataset with 30% level Rician noise, real images, and denoised images obtained by different fusion methods: (a) noise image, (b) noise-free image, (c) NPS, (d) AAW, (e) MIScnn, (f) pathify, (g) EDW, and (h) proposed method.

with a stride of 30×30 to obtain 63 patches from each image and finally obtained 201,600 patches. At the test phase, according to Equation (6), we used sliding windows of different strides to extract patches as shown in Table 3 and explored the performance of the proposed fusion method.

In Figure 10, we can see that the AAW method (Figure 10(b)) is greatly affected by the stride size. When the stride size is 10×10 , the image blocking artifact is not obvious in the fused bias field image, but with increasing stride size, obvious grid artifacts appear in the fused image. The EDW method (Figure 10(c)) and our method (Figure 10(d)) are less affected by the stride size. From the line chart of the quantitative evaluation index MSE in Figure 11, we can see that the highest point of the AAW method fluctuates by approximately 9% relative to the lowest point, while the highest point of the EDW method and our method fluctuates by less than 1% rel-

ative to the lowest point %. In Figure 12, we show the change in SSIM value with the change in stride size. The SSIM value of the AAW method decreases obviously with increasing stride. Although the SSIM values of the EDW method and our method fluctuate with the change in stride, their amplitude is small. In addition, compared with the EDW method, our method resulted in a smoother bias field image. As seen from Figure 12, our method can obtain a higher SSIM value than the EDW method despite the INU level, regardless of stride size.

As seen in Table 3, the smaller the step size, the more the number of patches extracted from each image; hence, the fusion time increases accordingly. Although the AAW method can obtain a relatively smooth bias field image when the stride is 10×10 , the fusion time difference is very small compared with the other two methods when the stride is 30×30 . However, there are still blocking artifacts in the

images fused by the AAW method. Our method and the EDW method are less affected by the stride. Even when the stride is 30×30 , the bias field image obtained does not appear to contain image blocking artifacts, having a lower MES value and higher SSIM value than the AAW method. Therefore, in the experiment on HCP database, to shorten the fusion time, we uniformly used a sliding window with a stride of 30×30 to extract patches.

4.2. Validation of the Proposed Fusion Method on the Denoising Model for 3D Brain MR Images. To further evaluate the fusion performance of the proposed algorithm, we adopted the DnCNN denoising model [18] for validation on the IXI-Guys human dataset. We randomly selected 20 T1w brain images from the IXI-Guys dataset with an image size of $256 \times 256 \times 150$ and voxel resolution of $0.9375 \times 0.9375 \times 1.2 \text{ mm}^3$ [20, 21]. 12 images were randomly selected as the training set, 4 images were used for verification, and the other 4 images were used for testing. In this dataset, we manually added 30% Rician noise to simulate the noise image [22]. In order to better train the model, we cropped the image to $240 \times 240 \times 150$, used the mirroring method to augment the data, and finally obtained 36 training samples. For patch-based training, we set the patch size to $60 \times 60 \times 60$ and used a sliding window with stride $30 \times 30 \times 30$ to obtain 196 patches from each image. Finally, we obtained 7056 patches for training the 3D model. Our fusion algorithm is evaluated on the testing set based on three aspects: MSE, PSNR (peak signal-to-noise ratio), and SSIM. The results are shown in Table 4.

In Table 4, we quantitatively compared the image qualities of the proposed fusion method and the state-of-the-art fusion methods in terms of MSE, PSNR, and SSIM [23] by training the denoising model on the 3D IXI-Guys human dataset. As seen from Table 4, the fusion method proposed in this paper shows lower MSE and higher PSNR and SSIM than other methods, which further proves the effectiveness of our method. In addition, in the enlarged red box in Figure 13, it can be seen that the denoising images of NPS, AAW, MIScnn, and Pathify have obvious seam lines between patches, which further illustrates the difference in the prediction of boundary regions between adjacent patches. However, in Figures 13(g) and 13(h), this situation is much better, because both the EDW method and our method fully consider the uncertainty of the edge and assign different weights to the central and the edge regions to solve the problem. However, a small grid artifact appears faintly in Figure 13(g), while in Figure 13(h) obtained by our fusion method, there is no grid artifact and is smoother than EDW.

5. Discussion and Conclusions

In deep learning, when the resolution of the image to be processed is too large and the resources (such as video card and video memory) are limited, the image is divided into small patches for processing, and the image patches are reconstructed and then spliced. Since the zero-padding method is commonly used in deep learning networks to ensure the consistency of input and output sizes, this method can lead to uncertainty in edge prediction. After

image reconstruction, obvious blocking artifacts appear in the spliced image due to the inconsistency of the boundary information of each patch processed. To solve this problem, we have studied the most popularly available patch fusion methods and proposed a fusion method with a truncated Gaussian function as a weighting function to assign weights to each pixel in the patch. In this method, we used the weighting function to convert the Euclidean-distance between the overlapping point and the central point of the patch into a weight coefficient, considered the predicted pixel values in all patches, and reduced the weight of the predicted pixels at the boundary. Finally, we obtained the predicted pixel values through the weighted calculations.

We carried out experiments on the simulated database BrainWeb and the real dataset HCP using the bias removal model. After comparing the proposed method with the popularly available fusion methods, NPS, AAW, MIScnn, patchify, and EDW, our method can obtain a super seamless and smooth bias field image. As evidenced by quantitative analysis, our fusion method achieves lower MSE and greater SSIM on both simulated and real data, which is clearly superior to the other five methods. In addition, we also discussed the bias field images obtained by AAW, EDW, and our proposed method with different stride on the HCP dataset. To further demonstrate the robustness of our method, we conducted experiments by training a denoising model on the IXI-Guys human dataset. Experimental results show that our fusion method also performs better than other five methods in quantitative analysis of MSE, PSNR, and SSIM on 3D dataset. However, our method has two shortcomings: (1) the standard variance of Gaussian function σ needs to be determined. Different application backgrounds require different values of σ . For example, in the bias removal model, the best fusion performance is achieved when $\sigma = 4$, while on the 3D brain image denoising model, the best fusion performance is achieved when the value of σ is around 8; (2) while our approach for obtaining an artifact-free image under a wider sliding window takes about the same amount of time as the AAW method for obtaining the same quality image under a smaller stride, it still takes a lengthy period. In a future work, we hope to be able to adaptively select the hyperparameters of different scenes and further optimize the algorithm to shorten the running time.

In this paper, we introduced a truncated Gaussian function as a weighting function, which converted the Euclidean distance between each pixel in the patch and the center point into the weight coefficient of this pixel, to reduce the image blocking artifact in patch-based image reconstruction. We demonstrated that the method proposed in this paper has significant advantages over existing patch fusion methods. Additionally, our approach can be applied to any patch-based deep learning model, even when the model is already trained.

Data Availability

All datasets used in this study are discussed in Section 4. They are publicly available and cited in the list of references.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

YX and SH are assigned to the conceptualization. SH is assigned to the methodology. YX and SH did the validation. SH and YD did the formal analysis. YX and SH curated the data. YX did the writing—original draft preparation. SH and YD did the writing—review and editing. SH and YD did the supervision. SH and YD acquired funding. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61771230, Grant 61903229 and Grant 61973180 and in part by the Natural Science Foundation of Shandong Province of China under Grant ZR2019BF004, Grant ZR2019PF005, Grant ZR2020QF011, and Grant ZR2021MF115.

References

- [1] J. Lin, Y. Pang, L. Xu, and Z. Huang, "Research progress of medical image processing based on deep learning," *Life Science Instruments*, vol. 16, no. 4, pp. 45–54, 2018.
- [2] Q. Zhang, D. Zhao, and X. Chi, "Review for deep learning based on medical imaging diagnosis," *Computer Science*, vol. 44, no. 11A, pp. 1–7, 2017.
- [3] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep learning in medical image registration: a review," *Physics in Medicine and Biology*, vol. 65, no. 20, pp. 1–32, 2020.
- [4] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: fast predictive image registration - a deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, 2017.
- [5] S. Hu, L. Zhang, G. Li, M. Liu, D. Fu, and W. Zhang, "Brain deformable registration using global and local label-driven deep regression learning in the first year of life," *IEEE Access*, vol. 8, pp. 25691–25705, 2019.
- [6] J. Fan, X. Cao, P. T. Yap, and D. Shen, "BIRNet: brain image registration using dual-supervised fully convolutional networks," *Medical Image Analysis*, vol. 54, pp. 193–206, 2019.
- [7] D. Müller and F. Kramer, "MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning," *BMC Medical Imaging*, vol. 21, no. 1, pp. 1–11, 2021.
- [8] W. Wu, "Patchify," <https://github.com/dovahcrow/patchify.py>.
- [9] L. Wu, S. Hu, and C. Liu, "Exponential-distance weights for reducing grid-like artifacts in patch-based medical image registration," *Sensors*, vol. 21, no. 21, p. 7112, 2021.
- [10] L. S. Athanasiou, D. I. Fotiadis, and L. K. Michalis, "Plaque Characterization Methods Using Optical Coherence Tomography," in *Atherosclerotic Plaque Characterization Methods Based on Coronary Imaging*, pp. 95–113, Academic Press, 2017.
- [11] J. Song and Z. Zhang, "Brain tissue segmentation and bias field correction of MR image based on spatially coherent FCM with nonlocal constraints," *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 4762490, 13 pages, 2019.
- [12] C. Li, C. Xu, A. W. Anderson, and J. C. Gore, "MRI tissue classification and bias field estimation based on coherent local intensity clustering: a unified energy minimization framework," in *Proceedings of International Conference on Information Processing in Medical Imaging*, pp. 288–299, Springer-Verlag, Williamsburg, VA, USA, 2009.
- [13] H. Ren, M. El-Khamy, and J. Lee, "Dn-resnet: efficient deep residual network for image denoising," in *Asian Conference on Computer Vision*, pp. 215–230, Springer, Cham, 2018.
- [14] Y. Xu, S. Hu, and Y. Du, "Bias correction of multiple MRI images based on an improved nonparametric maximum likelihood method," *IEEE Access*, vol. 7, pp. 166762–166775, 2019.
- [15] C. A. Cocosco, V. Kollokian, R. K. Kwan, G. B. Pike, and A. C. Evans, "BrainWeb: online interface to a 3D MRI simulated brain database," *NeuroImage*, vol. 5, 1997.
- [16] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, and K. Ugurbil, "The WU-Minn human connectome project: an overview," *NeuroImage*, vol. 80, pp. 62–79, 2013.
- [17] A. Hammers, C.-H. Chen, L. Lemieux et al., "Statistical neuroanatomy of the human inferior frontal gyrus and probabilistic atlas in a standard stereotaxic space," *Human Brain Mapping*, vol. 28, no. 1, pp. 34–48, 2007.
- [18] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
- [19] J. Elam, "HCP Wu-minn, 900 Subjects Data Release: Reference Manual," 2015, <https://www.humanconnectome.org/study/hcp-young-adult/document/900-subjects-data-release>.
- [20] L. Wu, S. Hu, and C. Liu, "Denoising of 3D brain MR images with parallel residual learning of convolutional neural network using global and local feature extraction," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 5577956, 18 pages, 2021.
- [21] D. Jiang, W. Dou, L. Vosters, X. Xu, Y. Sun, and T. Tan, "Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network," *Japanese Journal of Radiology*, vol. 36, no. 9, pp. 566–574, 2018.
- [22] M. Ran, J. Hu, Y. Chen et al., "Denoising of 3D magnetic resonance images using a residual encoder-decoder Wasserstein generative adversarial network," *Medical Image Analysis*, vol. 55, pp. 165–180, 2019.
- [23] J. Chen, J. Chen, H. Chao, and M. Yang, "Image blind denoising with generative adversarial network based noise modeling," in *Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on*, Salt Lake, 2018.

Research Article

Effects of Roflumilast on Patients with Chronic Obstructive Pulmonary Disease Treated with Inhaled Corticosteroid/Long-Acting β_2 Agonist: A Meta-analysis

Shasha Zeng ¹, Haibing Bai ¹, and Mi Zou ²

¹Department of Internal Medicine, Chongqing Nanan Traditional Chinese Medicine Hospital, Chongqing 400060, China

²Respiratory Department, The First Branch of the First Affiliated Hospital of Chongqing Medical University, Chongqing 400015, China

Correspondence should be addressed to Mi Zou; zoumi222@163.com

Received 26 April 2022; Revised 23 June 2022; Accepted 27 June 2022; Published 23 July 2022

Academic Editor: Tao Huang

Copyright © 2022 Shasha Zeng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. Roflumilast is a novel therapeutic drug for chronic obstructive pulmonary disease (COPD). This study was designed to evaluate the efficacy and safety of roflumilast combining inhaled corticosteroid (ICS)/long-acting β_2 agonist (LABA) in treating COPD patients through the meta-analysis. **Methods.** Randomized controlled trials of roflumilast combining ICS/LABA in treating patients with severe and profound COPD were searched from PubMed, Cochrane Library, and Embase databases from their establishment to February 2022. The quality of included studies was assessed by Cochrane risk bias assessment tool. The main outcomes of these studies should include at least one of the following clinical outcome indicators: forced expiratory volume in one second (FEV_1), exacerbation rate, and adverse events (AEs) such as diarrhea, nasopharyngitis, and headache. **Results.** Six articles were included in the study, including 9,715 patients. Meta-analysis revealed that compared with placebo, roflumilast gained superiority for severe COPD patients treated with ICS/LABA combinations in FEV_1 before bronchodilator administration (MD = 46.62, 95% CI (30.69, 62.55), $P < 0.00001$), FEV_1 after bronchodilator administration (MD = 45.62, 95% CI (34.95, 56.28), $P < 0.00001$), and COPD exacerbation rate (RR = 0.90, 95% CI (0.87, 0.94), $P = 0.001$). In terms of safety, the incidence of diarrhea, headache, nausea, weight loss, back pain, loss of appetite, and insomnia was notably higher in the roflumilast group than in the placebo group. **Conclusion.** Roflumilast is suggested to be significantly effective for severe COPD patients with ICS/LABA combination therapy, which reduces the exacerbation rate but also leads to PDE4 inhibitor-related adverse reactions.

1. Introduction

Chronic obstructive pulmonary disease (COPD) is a chronic respiratory disease with a high risk of death, and its incidence is high among smokers, which gradually increases with age [1]. The obstruction often occurs in small airways with diameter < 2 mm in COPD patients, resulting in airflow limitation. After bronchodilator administration, the decrease of patient's forced expiratory volume in one second (FEV_1) can reflect airflow limitation and lung functions, which can be utilized to diagnose the disease severity of COPD patients [2, 3]. Increased airflow limitation in COPD patients may increase the exacerbation risk of symptoms, and this limita-

tion possesses incomplete reversibility, whereas bronchodilators relax airway smooth muscle and increase tension, making it relatively reversible [4]. Furthermore, inflammation in the lungs or the whole body would have occurred in patients with COPD exacerbation, while inhaled corticosteroid (ICS) can reduce the exacerbation rate of COPD by arresting inflammation [5]. A clinical trial revealed that ICS combining dual bronchodilators is effective for patients with severe COPD at high exacerbation risk [6]. However, the treatment of COPD has been gradually relying on the combination therapy of ICS and long-acting bronchodilators in recent years, which may have safety risks or limit the disease control and prevention management of COPD [7].

Therefore, drugs that are more suitable for treating severe COPD or that can be used in combination with ICS to increase the treatment efficiency are continuously required.

Roflumilast is a phosphodiesterase-4 inhibitors (PDE4), which has anti-inflammatory effects and can inhibit the release of inflammatory mediators, thus advantageously treating the respiratory diseases, such as pulmonary inflammation complicated by asthma and COPD [8]. In terms of treating severe to profound COPD, roflumilast could attenuate frequent exacerbation of symptoms in patients, thus enabling them enter a stable period and reducing the history of exacerbation and the times of hospitalization [9, 10]. Meanwhile, roflumilast is well tolerated and has a favorable affinity with phosphodiesterase 4, which could effectively ameliorate lung functions of patients with severe to profound COPD [8, 11]. Nowadays, roflumilast generally functions as an adjunct to combination therapy of ICS and long-acting bronchodilators to further improve the drug efficacy in patients with severe to profound COPD. Long-acting beta2-agonists (LABAs) and long-acting muscarinic antagonists (LAMAs) are the widely used bronchodilators in combination therapies for severe COPD [12]. Roflumilast combining LABA/LAMA was pointed out to ameliorate lung functions of COPD patients, which is more effective than those treated with LABA or LAMA alone. Besides, roflumilast combining ICS/LABA or ICS/LABA/LAMA can effectively reduce the exacerbation rate of COPD [10, 13].

Therefore, we conducted a meta-analysis to comprehensively evaluate the efficacy and safety of roflumilast combining ICS/LABA or ICS/LABA/LAMA in patients with severe COPD.

2. Methods

2.1. Literature Retrieval. Randomized controlled trials (RCTs) of roflumilast combining ICS/LABA in COPD patients were retrieved from PubMed, Cochrane library, and Embase databases. All related English literature was researched from the establishment of databases to February 2020, whose keywords consisted of “roflumilast”, “inhaled corticosteroid”, “long-acting β_2 agonist”, and “chronic obstructive pulmonary disease”.

2.2. Selection of Literature

2.2.1. Inclusion Criteria. (1) Subjects: patients diagnosed with severe COPD by histopathological examination. At the same time, their spirometry showed airflow obstruction (after passing bronchodilator, forced expiratory volume in one second and forced vital capacity (FEV_1/FVC) < 0.70); (2) study type: phase III/IV RCTs; (3) interventions: combination treatment of roflumilast and ICS/LABA for COPD; (4) control group: combination treatment of placebo and ICS/LABA for COPD; and (5) outcome indicators: the following descriptions were included: FEV_1 , exacerbation rate, and AEs (such as diarrhea, nasopharyngitis, and headache).

2.2.2. Exclusion Criteria. (1) The intervention contained only ICS or LABA, or the control drug was not a placebo; (2) animal experiments; (3) non-English literature, repeatedly pub-

lished literature, or guidelines, review, case analysis, expert experience, meeting records, technical reports, and editorials; (4) literature with inconsistent data or could not be extracted; and (5) therapeutic drugs mentioned in the literature were not approved by the drug administration.

2.3. Data Extraction and Quality Assessment. Two investigators extracted the data independently, and a third investigator mended their divergences. These data, including author, year of publication, trial stage, sample size, interventions, FEV_1 , exacerbation rate, and AEs, were extracted from the trials. Quality assessment of the included studies was performed employing Cochrane bias risk assessment tool. This scale mainly evaluated the bias risk with 7 items in 6 aspects, including random sequence generation, allocation concealment, blind evaluation of investigators and subjects, blind evaluation of outcomes, integration of outcome data, reporting bias, and other obvious biases. The results of “low-risk bias,” “high-risk bias,” and “unclear” were obtained.

2.4. Statistical Methods. The Review Manager 5.4 software was applied for meta-analysis. Risk ratio (RR) served as the effect index for the count (dichotomous) data, and mean difference (MD) was utilized as the effect index for measurement (continuous variable) data, with point estimate values and 95% confidence intervals (CI) given for each effect size. The heterogeneity among results was determined by chi-square test, and the size of heterogeneity was quantitatively determined by combining I^2 . $I^2 \leq 50\%$ and $P \geq 0.1$ were considered to indicate no statistical heterogeneity among studies, and a fixed-effect model was carried out. Otherwise, heterogeneity was considered, and a random-effect model was applied for meta-analysis.

3. Results

3.1. Literature Retrieval Results. 136 studies were discarded in the 146 preliminarily searched literature by browsing titles and abstracts (guidelines, review, case analysis, expert experience, meeting records, technical reports, editorials, and republications). Among the 10 remaining literature, 4 were excluded after a full-text review, and other 6 studies [10, 14–18] that met the criteria were included. Figure 1 displays the literature screening process.

3.2. Basic Features and Quality Assessments of the Included Literature. Finally, 9,715 COPD cases were enrolled in the 6 included literature. Among them, 5,045 patients were treated with roflumilast combining ICS/LABA, and 4,670 patients were treated with placebo combining ICS/LABA. In 3 involved literatures [15–17], some patients were dosed with short-acting β_2 receptor agonist (SABAs) during the trial according to the actual situation. In all included treatment methods, patients were treated with or without LAMAs in line with their actual situation. The basic features of the included studies are indicated in Table 1. Cochrane bias risk assessment revealed that, except the large bias caused by the withdrawal of more cases from the trial due to adverse reactions in the experimental group than in the control group, other kinds of bias risk were at a low level.

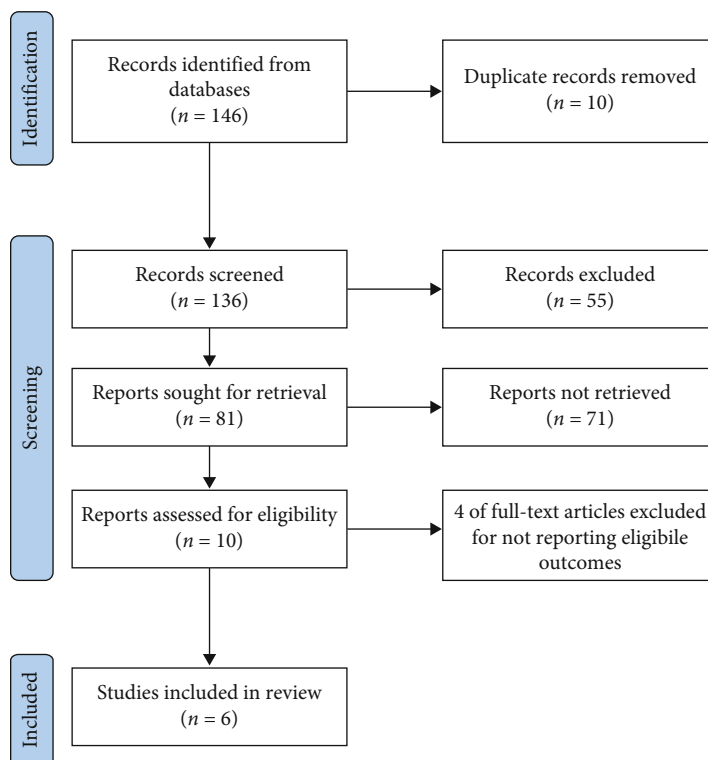


FIGURE 1: Flow chart of literature screening process.

TABLE 1: Primary characteristics of the eligible studies in more detail.

Author (year)	Phase	No. of patients		Male, <i>n</i> (%)		Age, mean (SD)		COPD severity, <i>n</i> (%)			
		Roflumilast	Placebo	Roflumilast	Placebo	Roflumilast	Placebo	Severe		Very severe	
								Roflumilast	Placebo	Roflumilast	Placebo
Calverley PM (2007)	IV	1,178	1,174	821 (70)	794 (68)	64 (8.8)	65 (8.4)	508 (67)	510 (68)	181 (24)	176 (23)
Calverley PM (2009)	IV	973	972	718 (74)	725 (75)	65 (8.4)	65 (8.4)	943 (61)	989 (64)	463 (30)	440 (28)
Rennard SI (2011)	IV	30	11	NR	NR	64 (7.4)	70 (6.8)	356 (63)	399 (66)	148 (26)	169 (28)
De Backer W (2014)	III	567	606	387 (68)	400 (66)	64 (8.7)	64 (8.8)	NR	NR	NR	NR
Martinez FJ (2015)	III	1,537	1,154	1,150 (75)	1,186 (76)	64 (9.0)	64 (9.0)	678 (70)	677 (70)	291 (30)	273 (28)
Martinez FJ (2016)	III	760	753	571 (75)	574 (76)	65 (9.6)	64 (9.1)	698 (59)	720 (61)	474 (40)	446 (38)

NR: no report.

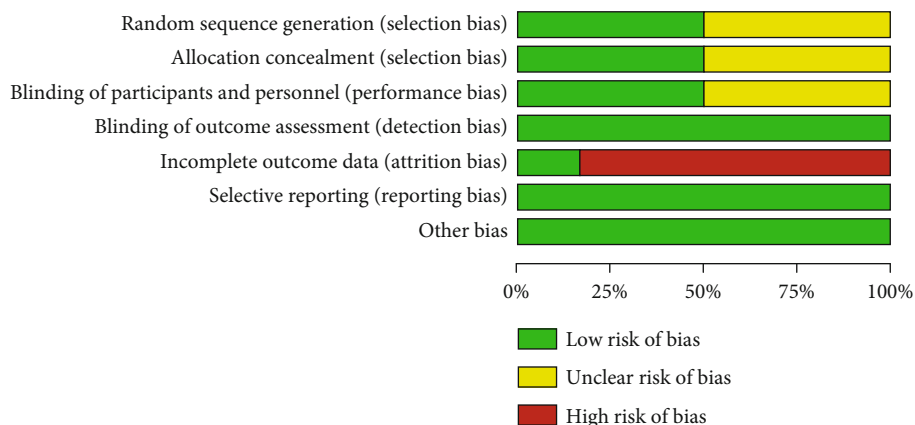
The overall quality of the included literature was relatively high, as represented in Figures 2(a) and 2(b).

3.3. Meta-analysis Results

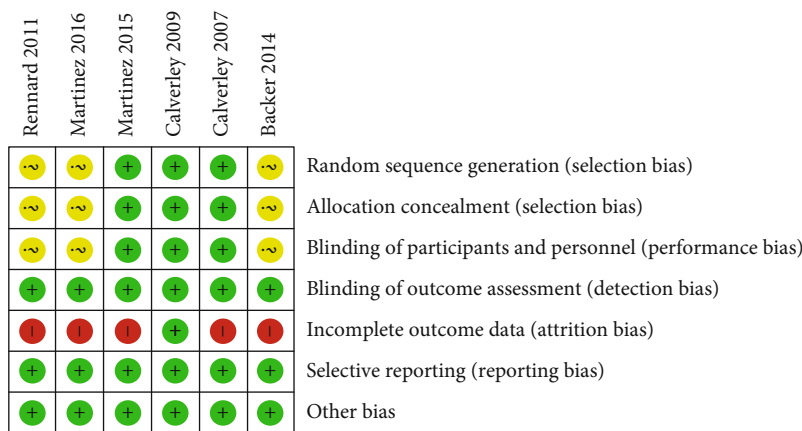
3.3.1. Changes in FEV₁ before Bronchodilator Administration Relative to Baseline. Four studies were included. FEV₁ was evidently increased in the roflumilast group before bronchodilator administration, whereas FEV₁ decrease was observed in the placebo group.

Meta-analysis of the random-effect model illustrated that the difference between experimental group and control group was significant when $\alpha = 0.05$ (MD = 46.62, 95% CI (30.69, 62.55), $P < 0.00001$), as shown in Figure 3.

3.3.2. Changes in FEV₁ after Bronchodilator Administration Relative to Baseline. Four studies were included. FEV₁ was notably increased in the roflumilast group after bronchodilator administration, whereas FEV₁ decrease was observed in the placebo group.



(a)



(b)

FIGURE 2: The quality assessment of the included literature: (a) overall bias risk; (b) bias risk for each RCT.

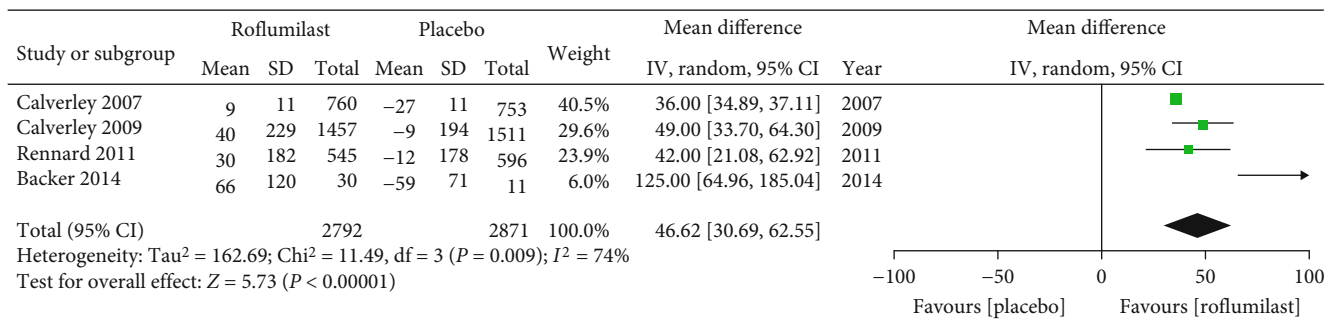


FIGURE 3: Comparison of FEV₁ changes before bronchodilator administration between the roflumilast group and placebo group.

the placebo group. Meta-analysis of the random-effect model illustrated that the difference between experimental group and control group was significant when $\alpha = 0.05$ (MD = 45.62, 95% CI (34.95, 56.28), $P < 0.00001$), as exhibited in Figure 4.

3.3.3. COPD Exacerbation. Three studies were eventually included for exploring COPD exacerbation. The results of meta-analysis demonstrated that COPD exacerbation rate

in the roflumilast group was remarkably lower than that in the placebo group, and the difference was statistically significant (RR = 0.90, 95% CI (0.87, 0.94), $P = 0.001$), as displayed in Figure 5.

3.3.4. Adverse Reactions. Among the adverse reactions, each of diarrhea, headache, nausea, and nasopharyngitis was included in 4 literatures. Each of weight loss, appetite loss, insomnia, back pain, influenza, pneumonia, hypertension,

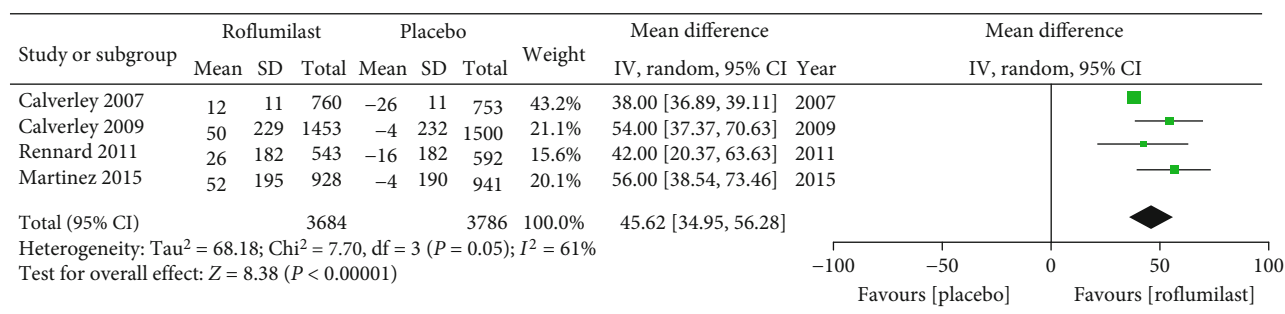


FIGURE 4: Comparison of FEV₁ changes after bronchodilator administration between the roflumilast group and placebo group.

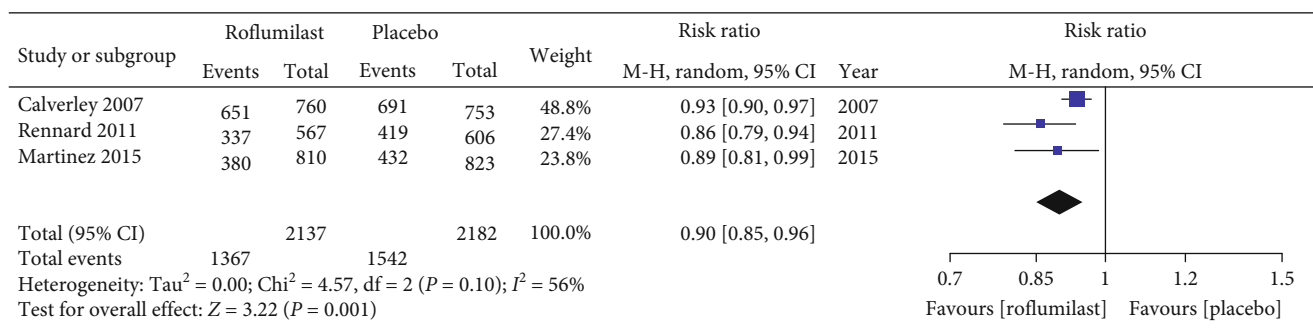


FIGURE 5: Comparison of COPD exacerbation rate between the roflumilast group and placebo group.

and death was included in 3 literature, and upper respiratory tract infection was included in 2 literatures. Meta-analysis of the fixed-effect model confirmed that the incidence of diarrhea (RR = 2.95, 95% CI (2.45, 3.56), P < 0.00001), headache (RR = 1.95, 95% CI (1.56, 2.43), P < 0.00001), nausea (RR = 2.58, 95% CI (2.01, 3.30), P < 0.00001), weight loss (RR = 3.41, 95% CI (2.74, 4.24), P < 0.00001), appetite loss (RR = 5.01, 95% CI (3.18, 7.90), P < 0.00001), insomnia (RR = 2.17, 95% CI (1.58, 2.96), P < 0.00001), and back pain (RR = 1.45, 95% CI (1.09, 1.94), P = 0.01) was markedly higher than that in the placebo group, and the differences were statistically significant, as presented in Figure 6.

As indicated in Figure 7, there was no statistically significant difference in incidence of nasopharyngitis (RR = 0.93, 95% CI (0.78, 1.10), P = 0.39), influenza (RR = 1.08, 95% CI (0.83, 1.41), P = 0.58), pneumonia (RR = 1.02, 95% CI (0.82, 1.28), P = 0.85), upper respiratory tract infection (RR = 0.88, 95% CI (0.68, 1.12), P = 0.30), hypertension (RR = 0.81, 95% CI (0.62, 1.05), P = 0.12), and death (RR = 1.05, 95% CI (0.78, 1.41), P = 0.74) between the roflumilast group and placebo group.

4. Discussion

COPD exacerbation is an acute event characterized by worsening of respiratory symptoms, which requires alterations of drug therapy and/or hospitalization [19]. Therefore, drug therapy for COPD is designed to relieve symptoms and reduce the risk of AEs such as exacerbation, disease progression, and death [20]. The combination of ICS/LABA has

been shown to reduce acute exacerbations of COPD and is often the preferred treatment for COPD [21, 22]. However, acute exacerbations of COPD are usually associated with high level of inflammation in the body [23]. Roflumilast can reduce the levels of inflammatory markers in the airway of COPD patients and is approved for long-term treatment in combination with ICS and long-acting bronchodilators of patients at high risk of acute exacerbation [24]. As roflumilast is recommended as an adjunct agent for patients with severe COPD, its combination with ICS/LABA effectively decreases the exacerbation risk in patients with severe COPD [25]. This study evaluated the efficacy and safety of roflumilast in patients with moderate to severe COPD using RCT data in combination with inhaled ICS/LABA.

The results suggested that compared with placebo, roflumilast remarkably ameliorated lung functions (as measured with FEV₁ before and after bronchodilator administration) and reduced the incidence of COPD exacerbation in patients with moderate to severe COPD. FEV₁ improvement plays a crucial role in determining the efficacy of new drugs for COPD [20]. In one included study, the end-point values of FEV₁ before and after bronchodilator administration are higher in the roflumilast group than the baseline level, while FEV₁ reduction was observed in the placebo group, which is in line with previous studies [26].

Heterogeneity is attributed to the differential definition of COPD exacerbation. In 3 literatures including this index, moderate to severe exacerbation is defined as the need for oral or parenteral glucocorticoid therapy (with or without antibiotics) and being hospitalized, or dead, or both. As a

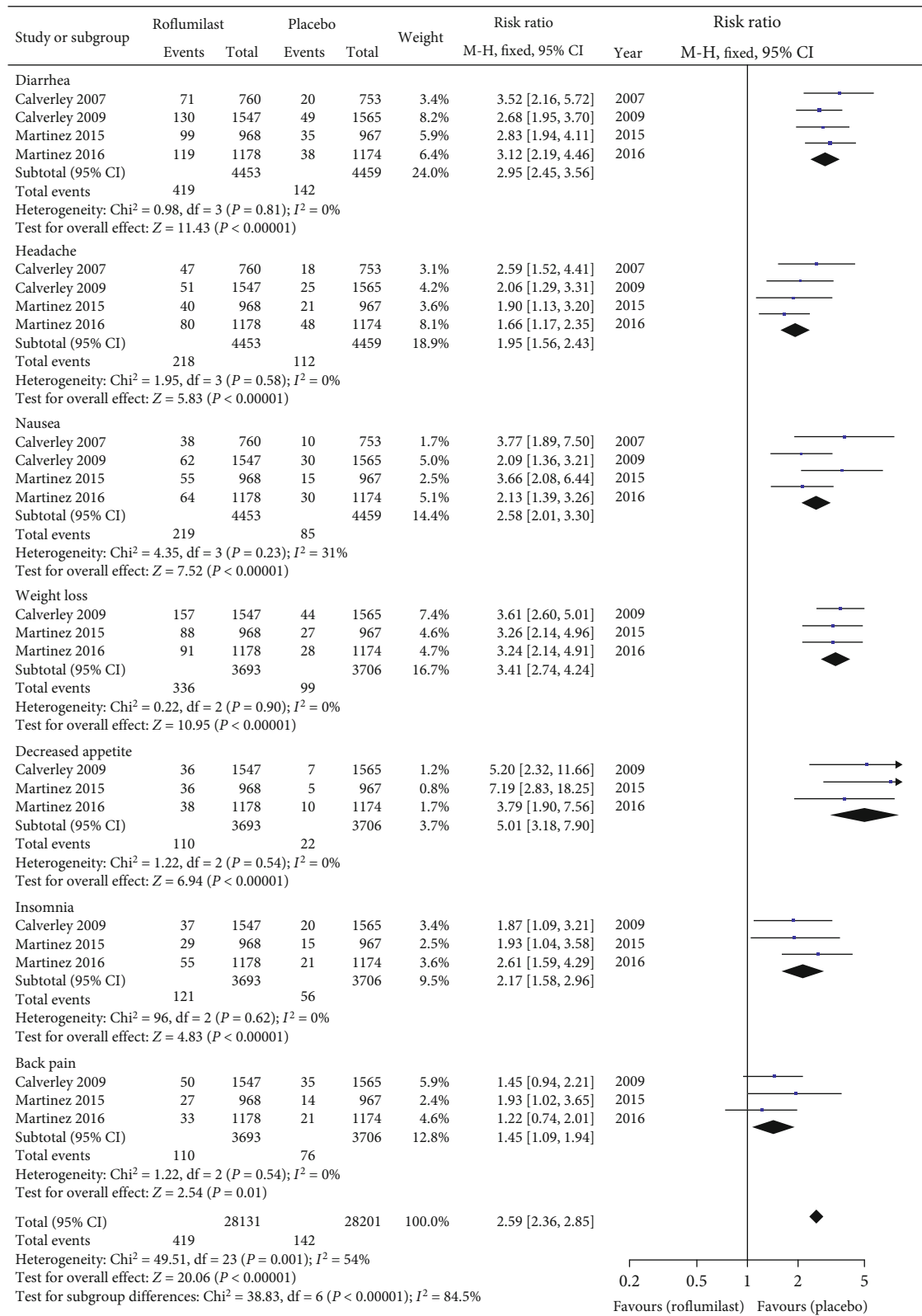


FIGURE 6: Comparison of incidence of adverse reactions between the roflumilast group and placebo group.

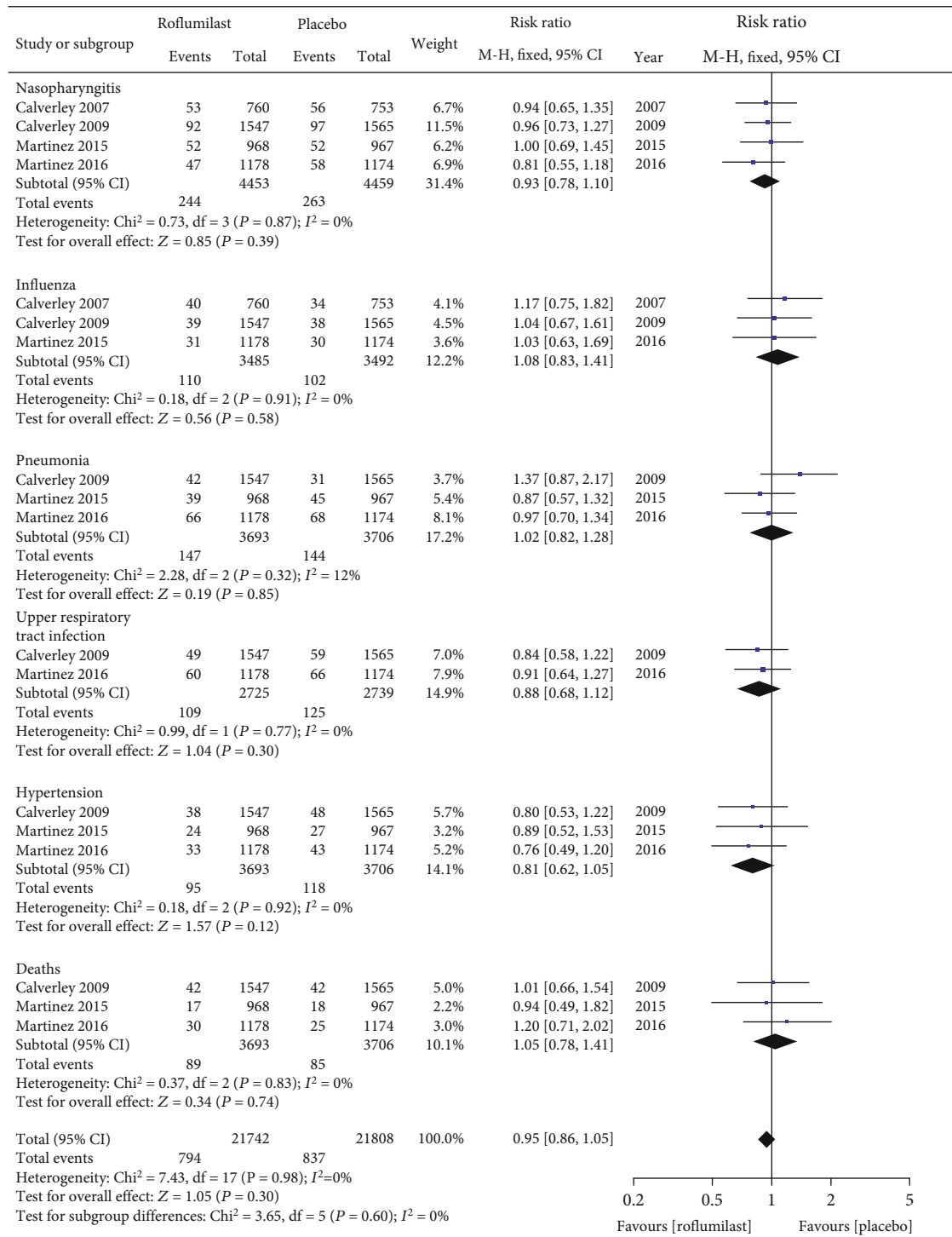


FIGURE 7: Comparison of incidence of adverse reactions between the roflumilast group and placebo group (no difference).

previous study described, roflumilast and ICS may reduce inflammation through different mechanisms. The combination of these agents may be cumulative or synergistic, and the utilization of their combination may provide clinical benefits beyond those achieved by single ICS, single PDE4 inhibitor, or ICS/LABA combination therapy [27]. Hajian et al. [25] suggested that roflumilast may enhance the effi-

cacy of ICS/LABA by reducing inflammation and edema via opening smaller airways or preventing airway collapse, thereby reducing regional overinflation. In a receiving appropriate combination therapy (REACT) trial, roflumilast in combination with ICS/LABA significantly reduced rates of exacerbation and hospitalization in moderate to severe COPD [10], which was consistent with our results.

More AEs and higher AE incidence were observed in the roflumilast group than the placebo group. In previously pooled analyses, this therapy mainly affects the gastrointestinal tract and nervous system, and the most common AEs comprise diarrhea, nausea, headache, and weight loss [9, 28], which are empirical PDE4 inhibitor-associated AEs [29]. And the incidence of this therapy is in accordance with that of a previous 6-month RCT of roflumilast [30]. In this study, compared with patients taking placebo, the most reported AEs were decreased appetite and weight loss in those taking roflumilast. And patients in the roflumilast group are elucidated to lost more weight [10, 17]. Among adverse reactions, pneumonia was proved to be correlated with an increased risk of death in COPD [31], and the use of ICS leads to an increased risk of pneumonia in COPD patients [32]. Roflumilast was not explored to increase the incidence of pneumonia in our study ($P=0.85$). During the study, approximately 2%-3% of patients died mainly due to COPD exacerbation and AEs [10], with no difference in mortality between the roflumilast and placebo groups ($P=0.74$).

However, several limitations existed in this study. First, little literature containing RCTs was included, whose results might not fully reflect the treatment situation. Second, various patients in the roflumilast group quit the trial due to AEs than those in the control group, which did not indicate the true incidence of AEs. Third, confined to the data provided by the original study authors, this study was unable to conduct a subgroup analysis, and expanded studies are needed in the future to characterize the population with the greatest benefit from this therapy. Finally, the evaluation of publication bias and funnel plot was not carried out in this study because less than 10 literatures were included.

Taken all together, according to the completed RCTs, roflumilast was dramatically effective to patients with severe and profound COPD treated with ICS/LABA and notably reduced the incidence of exacerbation but brought PDE4 inhibitor-associated AEs.

Data Availability

No additional data are available.

Conflicts of Interest

The authors declare that they have no conflicts of interest with the contents of this article.

Authors' Contributions

Conception and design were carried out by Shasha Zeng. Administrative support was carried out by Mi Zou. Shasha Zeng was responsible for the provision of study materials or patients. Collection and assembly of data were carried out by Haibing Bai. Data analysis and interpretation were carried out by Shasha Zeng. All authors contributed to the manuscript writing. All authors approved the final manuscript.

References

- [1] R. Monteleone Neto, "Current medical genetics," *Revista Paulista de Medicina*, vol. 108, no. 4, p. 147, 1990.
- [2] J. C. Hogg, P. D. Paré, and T. L. Hackett, "The contribution of small airway obstruction to the pathogenesis of chronic obstructive pulmonary disease," *Physiological Reviews*, vol. 97, no. 2, pp. 529–552, 2017.
- [3] P. Lange, D. M. Halpin, D. E. O'Donnell, and W. MacNee, "Diagnosis, assessment, and phenotyping of COPD: beyond FEV₁," *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 11 Spec Iss, pp. 3–12, 2016.
- [4] N. A. Hanania, B. R. Celli, J. F. Donohue, and U. J. Martin, "Bronchodilator reversibility in COPD," *Chest*, vol. 140, no. 4, pp. 1055–1063, 2011.
- [5] P. J. Barnes, "Inhaled corticosteroids," *Pharmaceuticals*, vol. 3, no. 3, pp. 514–540, 2010.
- [6] L. Vanfleteren, L. M. Fabbri, A. Papi, S. Petruzzelli, and B. Celli, "Triple therapy (ICS/LABA/LAMA) in COPD: time for a reappraisal," *International Journal of Chronic Obstructive Pulmonary Disease*, vol. Volume 13, pp. 3971–3981, 2018.
- [7] M. G. Matera, M. Cazzola, and C. Page, "Prospects for COPD treatment," *Current Opinion in Pharmacology*, vol. 56, pp. 74–84, 2021.
- [8] T. Kawamatawong, "Roles of roflumilast, a selective phosphodiesterase 4 inhibitor, in airway diseases," *Journal of Thoracic Disease*, vol. 9, no. 4, pp. 1144–1154, 2017.
- [9] J. A. Wedzicha, K. F. Rabe, F. J. Martinez et al., "Efficacy of roflumilast in the COPD frequent exacerbator phenotype," *Chest*, vol. 143, no. 5, pp. 1302–1311, 2013.
- [10] F. J. Martinez, P. M. A. Calverley, U. M. Goehring, M. Brose, L. M. Fabbri, and K. F. Rabe, "Effect of roflumilast on exacerbations in patients with severe chronic obstructive pulmonary disease uncontrolled by combination therapy (REACT): a multicentre randomised controlled trial," *Lancet*, vol. 385, no. 9971, pp. 857–866, 2015.
- [11] J. Zheng, J. Yang, X. Zhou et al., "Roflumilast for the treatment of COPD in an Asian population: a randomized, double-blind, parallel-group study," *Chest*, vol. 145, no. 1, pp. 44–52, 2014.
- [12] D. Singh, A. Agusti, A. Anzueto et al., "Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease: the GOLD science committee report 2019," *The European Respiratory Journal*, vol. 53, no. 5, article 1900164, 2019.
- [13] L. M. Fabbri, P. M. Calverley, J. L. Izquierdo-Alonso et al., "Roflumilast in moderate-to-severe chronic obstructive pulmonary disease treated with longacting bronchodilators: two randomised clinical trials," *Lancet*, vol. 374, no. 9691, pp. 695–703, 2009.
- [14] F. J. Martinez, K. F. Rabe, S. Sethi et al., "Effect of roflumilast and inhaled corticosteroid/long-acting β 2-agonist on chronic obstructive pulmonary disease exacerbations (RE²SPOND). A randomized clinical trial," *American Journal of Respiratory and Critical Care Medicine*, vol. 194, no. 5, pp. 559–567, 2016.
- [15] S. I. Rennard, P. M. Calverley, U. M. Goehring, D. Bredenbröker, and F. J. Martinez, "Reduction of exacerbations by the PDE4 inhibitor roflumilast—the importance of defining different subsets of patients with COPD," *Respiratory Research*, vol. 12, no. 1, p. 18, 2011.
- [16] P. M. Calverley, F. Sanchez-Toril, A. McIvor, P. Teichmann, D. Bredenbroeker, and L. M. Fabbri, "Effect of 1-year

- treatment with roflumilast in severe chronic obstructive pulmonary disease,” *American Journal of Respiratory and Critical Care Medicine*, vol. 176, no. 2, pp. 154–161, 2007.
- [17] P. M. Calverley, K. F. Rabe, U. M. Goehring et al., “Roflumilast in symptomatic chronic obstructive pulmonary disease: two randomised clinical trials,” *Lancet*, vol. 374, no. 9691, pp. 685–694, 2009.
- [18] W. De Backer, W. Vos, C. Van Holsbeke et al., “The effect of roflumilast in addition to LABA/LAMA/ICS treatment in COPD patients,” *The European Respiratory Journal*, vol. 44, no. 2, pp. 527–529, 2014.
- [19] J. Vestbo, S. S. Hurd, A. G. Agustí et al., “Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease,” *American Journal of Respiratory and Critical Care Medicine*, vol. 187, no. 4, pp. 347–365, 2013.
- [20] C. F. Vogelmeier, G. J. Criner, F. J. Martinez et al., “Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary,” *American Journal of Respiratory and Critical Care Medicine*, vol. 195, no. 5, pp. 557–582, 2017.
- [21] L. J. Nannini, P. Poole, S. J. Milan, and A. Kesterton, “Combined corticosteroid and long-acting beta(2)-agonist in one inhaler versus inhaled corticosteroids alone for chronic obstructive pulmonary disease,” *Cochrane Database of Systematic Reviews*, vol. 2013, no. 8, article Cd006826, 2013.
- [22] G. T. Ferguson, P. Darken, S. Ballal et al., “Efficacy of budesonide/glycopyrronium/formoterol fumarate metered dose inhaler (BGF MDI) versus other inhaled corticosteroid/long-acting muscarinic antagonist/long-acting β_2 -agonist (ICS/LAMA/LABA) triple combinations in COPD: a systematic literature review and network meta-analysis,” *Advances in Therapy*, vol. 37, no. 6, pp. 2956–2975, 2020.
- [23] W. R. Perera, J. R. Hurst, T. M. A. Wilkinson et al., “Inflammatory changes, recovery and recurrence at COPD exacerbation,” *The European Respiratory Journal*, vol. 29, no. 3, pp. 527–534, 2007.
- [24] D. C. Grootendorst, S. A. Gauw, R. M. Verhoosel et al., “Reduction in sputum neutrophil and eosinophil numbers by the PDE4 inhibitor roflumilast in patients with COPD,” *Thorax*, vol. 62, no. 12, pp. 1081–1087, 2007.
- [25] B. Hajian, W. Vos, J. De Backer et al., “Functional respiratory imaging to assess the interaction between systemic roflumilast and inhaled ICS/LABA/LAMA,” *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 11, pp. 263–271, 2016.
- [26] Y. Oba and N. A. Lone, “Efficacy and safety of roflumilast in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis,” *Therapeutic Advances in Respiratory Disease*, vol. 7, no. 1, pp. 13–24, 2013.
- [27] T. Moodley, S. M. Wilson, T. Joshi et al., “Phosphodiesterase 4 inhibitors augment the ability of formoterol to enhance glucocorticoid-dependent gene transcription in human airway epithelial cells: a novel mechanism for the clinical efficacy of roflumilast in severe chronic obstructive pulmonary disease,” *Molecular Pharmacology*, vol. 83, no. 4, pp. 894–906, 2013.
- [28] F. J. Martinez, K. F. Rabe, P. M. A. Calverley et al., “Determinants of response to roflumilast in severe chronic obstructive pulmonary disease. Pooled analysis of two randomized trials,” *American Journal of Respiratory and Critical Care Medicine*, vol. 198, no. 10, pp. 1268–1278, 2018.
- [29] G. F. Hebenstreit, K. Fellerer, K. Fichte et al., “Rolipram in major depressive disorder: results of a double-blind comparative study with imipramine,” *Pharmacopsychiatry*, vol. 22, no. 4, pp. 156–160, 1989.
- [30] K. F. Rabe, E. D. Bateman, D. O’Donnell, S. Witte, D. Bredenbröker, and T. D. Bethke, “Roflumilast—an oral anti-inflammatory treatment for chronic obstructive pulmonary disease: a randomised controlled trial,” *Lancet*, vol. 366, no. 9485, pp. 563–571, 2005.
- [31] M. I. Restrepo, E. M. Mortensen, J. A. Pugh, and A. Anzueto, “COPD is associated with increased mortality in patients with community-acquired pneumonia,” *The European Respiratory Journal*, vol. 28, no. 2, pp. 346–351, 2006.
- [32] S. Singh, A. V. Amin, and Y. K. Loke, “Long-term use of inhaled corticosteroids and the risk of pneumonia in chronic obstructive pulmonary disease,” *Archives of Internal Medicine*, vol. 169, no. 3, pp. 219–229, 2009.

Research Article

Mathematical Modeling and Computational Prediction of High-Risk Types of Human Papillomaviruses

Junchao Zhang  and Kechao Wang

Department of Mathematics, Harbin University, Harbin, Heilongjiang 150001, China

Correspondence should be addressed to Junchao Zhang; zhangjc@hrbu.edu.cn

Received 22 April 2022; Accepted 28 June 2022; Published 21 July 2022

Academic Editor: Lei Chen

Copyright © 2022 Junchao Zhang and Kechao Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cervical cancer is one of the main causes of cancer death all over the world. Most diseases such as cervical epithelial atypical hyperplasia and invasive cervical cancer are closely related to the continuous infection of high-risk types of human papillomavirus. Therefore, the high-risk types of human papillomavirus are the key to the prevention and treatment of cervical cancer. With the accumulation of high-throughput and clinical data, the use of systematic and quantitative methods for mathematical modeling and computational prediction has become more and more important. This paper summarizes the mathematical models and prediction methods of the risk types of human papillomavirus, especially around the key steps such as feature extraction, feature selection, and prediction algorithms. We summarized and discussed the advantages and disadvantages of existing algorithms, which provides a theoretical basis for follow-up research.

1. Introduction

Human papillomavirus (HPV) usually causes benign papillomas and epithelial malignancies [1]. About 30 years ago, researchers found a close relationship between HPV and cervical cancer. Since then, specific HPV type DNA has been found in almost all cervical cancer biopsies [2]. Epidemiological studies also continue to show that HPV is the main cause of cervical cancer. HPV is a papillomavacuolar virus of the family lactomaviridae. It is a spherical DNA virus with 72 shell particles on the surface, a 20 hedral three-dimensional symmetrical nucleocapsid structure, with a diameter of about 45-55 nm [3]. It contains 8000 base pairs, of which 88% are viral proteins. According to the function of HPV genome, it can be divided into three parts: noncoding region, early gene region, and late gene region. The noncoding regions include promoters, enhancers, and silencers, which play an important role in DNA replication and transcriptional regulation. The early gene region contains six genes E1, E2, E4, E5, E6, and E7, including proteins and oncogenes required in the process of replication [4]. The late

gene region contains L1 and L2 genes, which encode the structural proteins that constitute the virus capsid [5].

The early region encodes and produces six proteins. Their functions are as follows: E1 protein mainly controls virus replication and replication inhibition and is expressed in the early stage of virus infection; E2 protein is a specific DNA binding protein [6], which can not only regulate mRNA transcription and DNA replication but also inhibit the expression of E6 and E7 proteins; E4 protein is expressed during virus infection and plays an important role in virus replication and mutation; E5 protein is the smallest transforming protein [7]. It consists of two functional regions, one is the amino terminal hydrophobic region, which is related to the insertion position of E5 protein on the membrane, and the other is the hydrophilic region of carboxyl terminal. If the carboxyl terminal is injected into resting cells, it can induce cell DNA synthesis. In addition, it can also induce the expression of a variety of oncogenes; E6 and E7 proteins are the two most important proteins. They can not only regulate the cell cycle [8] but also play an important role in the proliferation of cancer cells.

More than 200 HPV types have been identified, and about 40 types can invade the female reproductive tract [9]. When the identified HPV has significant homology differences with the defined HPV types, some new types will be defined. Epidemiological studies have shown that there is a strong relationship between reproductive human papillomavirus and cervical cancer, which is not related to other risk factors. According to its relative malignancy, HPV can be divided into two or three types: low-risk type, medium-risk type, and high-risk type. The high-risk HPVs include HPV16, 18, 31, 33, 34, 35, 39, and 45, and the low-risk HPVs include HPV6, 11, 42, 43, and 44 [10–13].

The distribution of HPV types has obvious geographical characteristics. Research shows that among the high-risk types worldwide, HPV16 accounts for 51% [14], followed by HPV16, accounting for 16%, others, such as HPV45 accounts for 9%, HPV31 accounts for 6%, and HPV33 accounts for 3%. The sum of these four common HPV types exceeds 80%. However, in Latin America, HPV33 is the most common, followed by HPV39 and 59. In Asia, in addition to the most common HPV16 and 18 [15], HPV 52 and 58 account for far more cervical cancer than Western and African countries.

Different types of HPV have different clinical manifestations. According to the tissues invaded, HPV can be divided into high (low) risk type of skin and high (low) risk type of mucosa [16]. Low-risk skin types such as HPV1, 2, 3, and 4 are generally related to common warts, flat warts, and other diseases. High-risk skin types such as HPV5, 8, 14, and 17 are associated with anal cancer, prostate cancer, bladder cancer, and so on. For low-risk types of mucosa: HPV6, 11, 13, 32, 34, 40, 42, etc., they are generally related to diseases such as genital, anal, oropharyngeal, and esophageal mucosal infection. Mucosal high-risk types such as HPV16, 18, 30, 31, 33, 35, and 39 can lead to cervical cancer, rectal cancer, tonsillar cancer, and other diseases [17].

Now, there are many epidemiological and experimental methods to identify HPV types [18–21]. They mainly use polymerase chain reaction (PCR) technology to realize the rapid detection of clinical samples. With the rapid accumulation of human papillomavirus data, a reliable and effective calculation method is needed to predict high-risk HPV. In recent years, many computational models have been proposed to predict HPV high-risk types. Eom et al. designed genetic algorithm to predict HPV type through the sequence fragment distribution characteristics of HPV [22]. Joung et al. designed the prediction method of HPV type based on support vector machine prediction and hidden Markov model [23, 24]. Through systematic analysis, Park et al. suggested using decision tree to construct the typing model of human papillomavirus [25]. Kim and Zhang calculated the distance distribution of amino acid pairs and further predicted the risk type of HPV through E6 protein [26, 27]. Kim et al. extracted the differential features of protein secondary structure and designed a set of support vector machine (GSVM) to classify HPV types [28]. Esmaeili et al. calculated Chou’s pseudoamino acid composition of HPV sequence and classified HPV types using ROC [29]. Alemi et al. systematically compared the physical and chemical properties of high-risk and low-risk HPV and build a

prediction model of high-risk HPV based on the support vector machine [30]. Wang et al. used protein “sequence space” to explore this information to predict high-risk types of HPVs [31]. Xu et al. proposed a HPV high-risk prediction model based on reduced amino acids, and they divided 20 amino acids into several nonoverlapping groups and calculated the structure and chemical patterns of each group as the basis for distinguishing high-risk HPV [32].

In this paper, we introduce the development of HPV typing in cervical cancer, mainly focusing on sequence characteristics and predicted secondary structure characteristics. The feature fusion method and feature selection algorithm are discussed. Finally, we review the multiple classification and various machine learning algorithms in HPV typing of cervical cancer.

2. Benchmark Datasets

HPV database is from Los Alamos National Laboratory (LANL), with a total of 72 HPV types. HPV risk types were manually determined based on the HPV profile [33]. If HPV belongs to skin related or skin group, HPV is classified as a low-risk type. On the other hand, it is classified as a high-risk type if HPV is known to be a high-risk type of cervical cancer [28]. Some researchers may build their own HPV dataset from SWISS-PROT. For these sequences, some researchers will complete their own wonderful processing. Here, we focus on using mathematical models to detect the risk types of human papillomavirus. Therefore, the detailed processing of data sets is no longer our important discussion content.

3. Methods

3.1. Feature Extraction. The extraction of HPV sequence information is transforming from complex structure to extraction of mathematical features [34]. Directly extracting information from the complex structure of HPV protein will cause large errors, and the extracted information may not be comprehensive. However, it is very easy to directly extract the mathematical features from the protein sequence, there is no influence of error on the connection, and the extracted information contains all the protein information.

3.1.1. Sequence Features

(1) *K-Mer.* Protein sequences and peptides can be regarded as a set of signs [35], and we can analyze their characteristics through the frequency of k -mer to obtain the difference between the two HPV sequences. If k is 1, it means that it is the composition of amino acids, which is calculated as follows [36]:

$$F(S) = (f_1, f_2, \dots, f_{20})^T, \quad (1)$$

$$f_i = \frac{n_i}{\sum_{i=1}^{20} n_i}, \quad (2)$$

where n_i means the number of the amino acid i in the protein sequence.

(2) *Order-Based Features*. Order-based features reflect the physical and chemical interaction among the amino acids pairs, which can be described by sequence coupling score and quasisequence score [37].

Suppose S is a protein sequence with a length of L , τ_j is denoted as the sequence interaction factor describing the sequence influence:

$$\tau_j = \frac{1}{L-j} \sum_{i=1}^{L-j} J_{i,i+j} (j=1, 2, \dots, k, k < L), \quad (3)$$

where $J_{i,i+j}$ is the physical and chemical distance between the amino acid R_i and R_{i+j} , which can be calculated:

$$J_{i,i+j} = D^2(R_i, R_{i+j}). \quad (4)$$

The calculation method of D^2 can consider the hydrophobicity value, polarity value, and side chain volume of amino acids. Thus, S can be defined by the following features:

$$V_{QSOE}(S) = [v_1, v_2, \dots, v_{20}, v_{20+1}, \dots, v_{20+k}]^T. \quad (5)$$

The first 20 features are k -mer features, and the latter k -dimensional features represent the interaction information of amino acid sequences.

(3) *Amino Acid Distribution*. Amino acid distribution is composed of amino acid composition, transformation, and distribution. This component can be regarded as the amino acid component of HPV, which can be original sequence or reduced sequence to describe the sequence component [38]. The transformation can be described by calculating the number of conversions between a base and a subsequent base. It can be calculated by the following formula:

$$Trs = \frac{\text{Count}_{rs} + \text{Count}_{sr}}{N - 1}, \quad (6)$$

where N is the number of amino acids, and Count_{sr} is the number of conversions between the base s and a subsequent base r .

(4) *Pseudoamino Acid Composition (PseAAC)*. In order to avoid the loss of many important information hidden in HPV sequences, Chou proposed pseudoamino acid composition (PseAAC), which is not a simple AAC to represent protein samples [39]. Pseudoamino acid composition is a vector with size $R + \lambda$, where R is the composition size of the amino acids, and the latter λ dimension is the information of pseudoamino acid composition [29]. They can be calculated by the following formula:

$$v_u = \begin{cases} \frac{f_u}{\sum_{u=1}^R f_u + w \sum_{k=1}^{\lambda} \tau_k} (u \leq R), \\ \frac{w \tau_{u-20}}{\sum_{u=1}^R f_u + w \sum_{k=1}^{\lambda} \tau_k} (R \leq u \leq R + \lambda), \end{cases} \quad (7)$$

where f_u is the frequency of RedAA, w is the weight, and τ_k can be calculated by the following formula:

$$\tau_k = \frac{1}{L-k} \sum_{i=1}^{L-k} J_{i,i+k} (k < L), \quad (8)$$

$$J_{i,j+k} = \frac{1}{3} \{ [H_1(R_i) - H_1(R_{i+k})]^2 + [H_2(R_i) - H_2(R_{i+k})]^2 + [M(R_i) - M(R_{i+k})]^2 \}. \quad (9)$$

$H_i(R_i)$ represents molecular weight of amino acid residue R_i , and $M(R_i)$ is the molecular weight of amino acid residue R_i . Then, the protein sequence can be mapped to the following feature vectors by this way:

$$V_{\text{PseAAC}}(S) = [v_1, v_2, \dots, v_{20}, \dots, v_{20+\lambda}]^T. \quad (10)$$

3.1.2. Evolutionary Profile

(1) *Position Specific Scoring Matrix (PSSM)*. PSSM is developed on the basis of position frequency matrix [40], which can quantitatively describe the evolutionary characteristics of HPV sequences. At present, it is the most commonly used matrix feature model, especially for protein structural class prediction [41]. When a protein s with length L is given, its PSSM is defined as

$$\text{PSSM}_s = \begin{bmatrix} P_{1 \rightarrow 1} & P_{1 \rightarrow 2} & \dots & P_{1 \rightarrow j} & \dots & P_{1 \rightarrow 20} \\ P_{2 \rightarrow 1} & P_{2 \rightarrow 2} & \dots & P_{2 \rightarrow j} & \dots & P_{2 \rightarrow 20} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{i \rightarrow 1} & P_{i \rightarrow 2} & \dots & P_{i \rightarrow j} & \dots & P_{i \rightarrow 20} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{L \rightarrow 1} & P_{L \rightarrow 2} & \dots & P_{L \rightarrow j} & \dots & P_{L \rightarrow 20} \end{bmatrix}, \quad (11)$$

where $i \rightarrow j$ represents that the residue at the position of i - th mutated into the type of j during the course of biological evolution. $P_{i \rightarrow j}$ indicates the score of mutation, and the ordered 20 amino acids are marked as 1, 2, 3... 20. The close evolutionary relationship among these HPV sequences can be described by their PSSM [42]. We search homogeneous sequences from SWISS-PROT database with help of PSI-BLAST [43–45] and normalize PSSM using the following sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (12)$$

(2) *Structural Properties of PSSM*. When $[0, L - 1] \times \Sigma \rightarrow \mathfrak{R}$ is used to describe a function M , L is the length of a function M . The PSSM of the sequence Seq_0 will be transformed into

a RedPSSM of its reduced sequence Seq_s ,

$$[0, L] \times \sum_{Seq_0} \longrightarrow [0, L] \times \sum_{Seq_r}, \quad (13)$$

where \sum_{Seq_0} and \sum_{Seq_r} are finite alphabet sets of the original and reduced sequences. $[\text{RedPSSM}]_{ij}$ is defined as

$$[\text{RedPSSM}]_{ij} = \sum_{j=1}^{g_s(j)} \frac{P_{i \rightarrow j}}{g_s(j)}, \quad (14)$$

where $g_s(j)$ represents the size of the reduced group which is made up with the reduced amino acid j , and j belongs to $1 \leq j \leq |\sum_{Seq_r}|$.

Auto covariance (AC) is a correlation factor, which has been widely used in various fields of bioinformatics [46]. It can connect adjacent residues according to protein sequence and convert PSSM into equal length vector [47]. The AC variable can represent the average interaction between residues with a series of lags, which is defined as

$$AC_{j,g}(S) = \frac{1}{L-g} \sum_{i=1}^{L-g} (P_{i,j} - P_j)(P_{i,j+g} - P_j), \quad (15)$$

where $AC_{j,g}(S)$ can indicate the interaction between two spacer g residues. But it can just calculate the residue interaction in the same column [48]. In order to solve this problem, we calculated total correlations among all the columns by the following formula:

$$\begin{aligned} \text{RAC}_g(h, j) &= \frac{1}{L-g} \sum_{i=1}^{L-g} \left| [\text{RedPSSM}]_{i,h} - \frac{[\text{RedPSSM}]_{i,h} + [\text{RedPSSM}]_{i+g,j}}{2} \right| \\ &\quad \times \left| [\text{RedPSSM}]_{i+g,h} - \frac{[\text{RedPSSM}]_{i,h} + [\text{RedPSSM}]_{i+g,j}}{2} \right| \\ &= \frac{1}{4(L-g)} \sum_{i=1}^{L-g} \left([\text{RedPSSM}]_{i,h} - [\text{RedPSSM}]_{i+g,h} \right)^2. \end{aligned} \quad (16)$$

3.1.3. Secondary Structure Features. In order to improve the accuracy of HPV high-risk types, some researchers analyzed the protein secondary structures of HPV, extracted its characteristics, and constructed prediction models [28, 49]. Given a protein sequence, we predict its secondary structure sequence. Its secondary structure content ($\text{content}_{\text{SE}}$) can be calculated as

$$\text{content}_{\text{SE}} = \frac{\text{Count}_{\text{SE}}}{\sum_{x \in \{C,H,E\}} \text{Count}_x}, \quad (17)$$

where SE belong to $\{C, H, E\}$. H is α -helix, E is β -strand, and C is coil. Gk and Herand calculated the first and second order composition moment vector (CMV) as the structure features [50], which is defined as the following formula:

$$\text{CMV}_{\text{SE}}^k = \frac{\sum_{j=1}^{\text{Count}_{\text{SE}}} \text{PO}_{\text{SE}j}^k}{\prod_{d=1}^k (N-d)}, \quad (18)$$

where $\text{PO}_{\text{SE}j}$ represents the element at position j in the secondary structure sequence with length N , which k is the vector order.

In order to further study the arrangement of different structural elements, some important structural fragments or patterns have been proposed one after another [51]. The length of the longest segment ($\text{MaxSeg}_{\text{SE}}$) is defined as the following formula:

$$\text{MaxSeg}_{\text{SE}} = \text{MaxLen}(\text{SEG} : \text{SEG}_{\text{SE}}), \quad (19)$$

where MaxLen represents the function of the maximal segment length, and SEG_{SE} is the composed of each segment of the structure element SE [52]. In order to reduce the length effect, a normalized length of the longest segment ($\text{NMaxSeg}_{\text{SE}}$) is defined

$$\text{NMaxSeg}_{\text{SE}} = \frac{\text{MaxLen}(\text{SEG} : \text{SEG}_{\text{SE}})}{N}, \quad (20)$$

where N is the sequence length. In addition to the maximal segment length, the average length of the segment ($\text{AvgSeg}_{\text{SE}}$) is also an important feature of protein secondary structure, which is defined as

$$\text{AvgSeg}_{\text{SE}} = \frac{\sum \text{Len}(\text{SEG} : \text{SEG}_{\text{SE}})}{\text{Content}_{\text{SEG}_{\text{SE}}}}, \quad (21)$$

where $\text{Content}_{\text{SEG}_{\text{SE}}}$ represents the substance of SEG_{SE} , and Len is a function of segment length. The normalized average length of the segment ($\text{NAvgSeg}_{\text{SE}}$) is

$$\text{NAvgSeg}_{\text{SE}} = \frac{\sum \text{Len}(\text{SEG} : \text{SEG}_{\text{SE}})}{\text{Content}_{\text{SEG}_{\text{SE}}} \times N}, \quad (22)$$

where N represents the length of protein sequence.

Zhang et al. analyzed the secondary structure of protein, only considered the helical and folded structural units, ignoring the irregular curl [53]. According to this simplification rule, they got a simplified protein secondary structure sequence and defined the transition probability matrix (TPM) as follows:

$$\text{TPM} = \begin{pmatrix} P_{\alpha\alpha} & P_{\alpha\beta} \\ P_{\beta\alpha} & P_{\beta\beta} \end{pmatrix}, \quad (23)$$

where

$$P_{a_i, a_j} = \begin{cases} \frac{\text{Content}_{a_i a_j}}{\sum_{t=1}^2 \text{Content}_{a_i a_t}} & \sum_{t=1}^2 \text{Content}_{a_i a_t} \neq 0, \\ 0 & \sum_{t=1}^2 \text{Content}_{a_i a_t} = 0, \end{cases} \quad (24)$$

where a_i belongs to $\{\alpha, \beta\}$, a_i is followed by the a_j , and the size of the event is described by $\text{Content}_{a_i a_j}$.

Most researchers pay attention to the content of structural units, so sometimes they do not know the location information of these structural units in the protein structure. Dai et al. proposed some location-based structural features [54]. If the distance $\text{Dis}(\delta)$ of the given structural element δ is 1, they will be divided into a structural domain. If not, they will be divided into two different structural domains. If the $\text{Dis}(\delta)$ is given, its probability distribution can be achieved. The definitions of digital feature semimean $\text{Semi} - E_{(k)}(\delta)$ and semivariance $\text{Semi} - D_{(k)}(\delta)$ are as follows:

$$\text{Semi} - E_{(k)}(\delta) = \sum_{\text{Dis}(\delta)=1}^k \text{Dis}(\delta) \times P(\text{Dis}(\delta)), \quad (25)$$

$$\text{Semi} - D_{(k)}(\delta) = \sum_{\text{Dis}(\delta)=1}^k (\text{Dis}(\delta))^2 \times P(\text{Dis}(\delta)) - \left[\sum_{\text{Dis}(\delta)=1}^k \text{Dis}(\delta) \times P(\text{Dis}(\delta)) \right]^2. \quad (26)$$

The ratio of the standard $\text{Semi} - D_{(5)}$ to $\text{Semi} - E_{(5)}$ can be calculated by:

$$\text{PSSF}(\delta) = \frac{\text{Semi} - E_{(5)}(\delta)}{\sqrt{\text{Semi} - D_{(5)}(\delta)}}. \quad (27)$$

$\text{PSSF}(\delta)$ is used to describe the position distribution of predicted secondary structural elements.

3.2. Feature Selection. Feature selection (SF) is one of the important steps in data processing, which plays an important role in data mining, pattern recognition, and machine learning [32]. It solves the problem of how to select the input features corresponding to the optimal prediction results. Through feature selection, the complexity of the problem can be reduced, and the prediction accuracy, robustness, and interpretability of the learning algorithm can be improved [55]. Therefore, feature selection can be used to select the characteristics of HPV risk types [56]. More importantly, it helps to have a deeper understanding of HPV sequences when we analyze their characteristics. This section summarizes some feature selection methods.

3.2.1. Mutual Information. The feature selection method based on mutual information can capture the nonlinear relationship between variables, which is more suitable for dealing with complex classification problems [57]. Mutual information can be expressed as characteristic matrix X , and the relevant category is C . Mutual information can be defined as

$$\text{Rel}(X_i) = I(X_i; C) = \sum_{X_i, C} P(X_i, C) \log \frac{P(X_i, C)}{P(X_i)P(C)}. \quad (28)$$

When the set of S is selected, the redundant formula Eq. (29) is

$$\text{Red}(X_i|S) = \frac{1}{S} \sum_{X_j \in S} I(X_i; X_j). \quad (29)$$

As can be seen from the above definition, mutual information aims to select the features that are most relevant to the target category and have the least redundancy between the selected features [58]. Therefore, feature selection can be realized directly according to the value of mutual information.

3.2.2. Support Vector Machine Recursive Feature Elimination (SVM-RFE). Support vector machine has been widely used in the field of pattern recognition and is very suitable for small samples with high-dimensional data. SVM-RFE is a sequential reverse selection (SBS) algorithm based on the maximum interval principle of SVM [59]. SVM-RFE can be divided into linear SVM-RFE and nonlinear SVM-RFE.

(1) Linear SVM-RFE. There is a training sample set $\{x_i, y_i\}$, $y_i \in \{-1, 1\}$, $i = 1, 2, \dots, n$. The linear SVM can be calculated by the following formula:

$$f(x) = a \cdot x + b, \quad (30)$$

where a is the weight vector of the optimal hyperplane, and b is the threshold.

By introducing Lagrange's formula, the optimization problem of SVM can be transformed into the following dual programming problem:

$$L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j, \quad (31)$$

where α_i can be calculated by solving the maximum value of L_D under the range of $\alpha_i \geq 0$ and $\sum_{i=1}^n \alpha_i y_i = 0$. So a Eq. (32) can be defined as follows:

$$a = \sum_{i=1}^n \alpha_i y_i x_i. \quad (32)$$

The ranking criterion score of the k -th feature is defined as the following equation:

$$J(k) = w_k^2. \quad (33)$$

During the training process, the feature with the smallest score of the ranking criterion is removed, and the remaining features are used to train the SVM for the next iteration.

(2) Nonlinear SVM-RFE. When the sample size is larger than the number of features, nonlinear SVM-RFE can obtain better results [60]. Features can be mapped to new spaces in higher dimensions by nonlinear SVM-RFE:

$$x \in R^d \mapsto \varphi(x) \in R^h. \quad (34)$$

After being mapped to the new space, the samples can be

divided into linear features. It can be calculated by the formula of Lagrangian:

$$L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \varphi(x_i) \varphi(x_j). \quad (35)$$

Here, $\Phi(x_i)\Phi(x_j)$ can be changed into a Gaussian kernel formula $K(x_i, x_j)$:

$$K(x_i, x_j) = e^{-\lambda \|x_i - x_j\|^2}. \quad (36)$$

The ranking criteria of feature k can be described by the following formula:

$$J(k) = \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i^{(-k)}, x_j^{(-k)}), \quad (37)$$

where $x_i^{(-k)}$ represents that k has been dropped.

3.2.3. Genetic Algorithm. Genetic algorithm is an adaptive search strategy, and its principle is similar to the survival mechanism of the fittest in nature [61]. It has strong adaptability, strong independence of domain knowledge, and can carry out a large number of parallel computing. Therefore, it is more suitable for processing large-scale complex data, especially for solving multiobjective optimization problems [62]. Given an HPV data set, we construct its characteristic matrix X and use genetic algorithm to obtain an eigenvector, which is an optimal feature set [63].

3.2.4. Kurtosis and Skewness. Kurtosis and skewness are two characteristics describing distribution [64], which can distinguish the distribution of different characteristics. Therefore, many studies calculated the skewness and kurtosis of each feature and selected certain features according to the value of skewness and kurtosis. Given the distribution of a feature $\{x_1, x_2, \dots, x_n\}$, its kurtosis and skewness are defined as follows:

$$\text{Kurtosis} = \frac{\sum_{i=1}^n (x_i - x)^4}{(n-1)SD^4} - 3, \quad (38)$$

$$\text{Skewness} = \frac{\sum_{i=1}^n (x_i - x)^3}{(n-1)SD^3}. \quad (39)$$

3.2.5. ReliefF Algorithm. ReliefF is an independent evaluation method, which evaluates each feature separately and assigns weight to the feature [65, 66]. For each time, it randomly selects a sample R from the training set and then calculates the latest instance from the same class R and different sample sets, and then the weight update rules in each step are as follows:

$$W(A) = W(A) - \sum_{j=1}^k \frac{\text{diff}(A, R, H_j)}{mk} + \sum_{C \notin \text{class}(R)} \frac{p(C)/1 - p(\text{Class}(R)) \sum_{j=1}^k \text{diff}(A, R, M_j(C))}{mk}, \quad (40)$$

where $\text{diff}(A, R_1, R_2)$ represents the difference between the sample R_1 and the sample R_2 on the feature A , and $M_j(C)$ indicates the nearest neighbor of the position of j -th in class C Eq. (41). $\text{diff}(A, R_1, R_2)$ can be calculated by the following formula:

$$\text{diff}(A, R_1, R_2) = \begin{cases} \frac{|R_1[A] - R_2[A]|}{\max(A) - \min(A)}, \\ 0R_1[A] = R_2[A], \\ 1R_1[A] \neq R_2[A]. \end{cases} \quad (41)$$

3.3. Prediction Algorithms

3.3.1. Support Vector Machines (SVM). SVM is a discriminant classifier, which is defined by the classification hyperplane [27]. In other words, the labeled training samples are used to train the model, and then the test sample classification is realized by outputting the best hyperplane [67]. When the prediction problem is nonlinear, the objective function of SVM can be defined as

$$f(x) = \sum_{i=1}^N \alpha_i y_i K(x_i, x) + b, \quad (42)$$

where $K(x_i, x)$ is a kernel function, x_i is a support vector, and α_i belongs to $[0, C]$. The Gaussian radial basis kernel function with strong learning ability and small error is often selected as the kernel function, which is defined as

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right), \quad (43)$$

where σ is the coefficient of the kernel function and has high flexibility.

3.3.2. Principal Component Analysis (PCA). PCA is a statistical analysis method that converts multiple indicators into a few comprehensive indicators. The idea of PCA is to replace the original more with fewer comprehensive variables [68]. The first step of PCA is to normalize the matrix which can also be expressed as the following mathematical formula:

$$Z_{ij} = \frac{x_{ij} - x_j}{s_j} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, p), \quad (44)$$

$$x_j = \frac{\sum_{i=1}^n x_{ij}}{n}, s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - x_j)^2}{n-1}. \quad (45)$$

The second step of PCA is to obtain the correlation of

coefficient matrix for Z :

$$R = [r_{ij}]_p \cdot xp = \frac{Z^T Z}{n-1}, \quad (46)$$

$$r_{ij} = \frac{\sum z_{kj} \cdot z_{kj}}{n-1} (i, j = 1, 2, \dots, p). \quad (47)$$

3.3.3. *K-Nearest Neighbor Algorithm (KNN)*. KNN is an important nonparametric classification method, which classifies the samples according to the categories of most k -nearest neighbor samples in the feature space [69]. However, the basic k -nearest neighbor classification algorithm needs global search, which has high computational complexity and slow computing speed. In order to reduce the influence of the same feature function in the traditional KNN algorithm, different weights can be assigned to the features in the distance formula to measure the similarity. For example, in Euclidean distance formula, different weights are assigned to different features, as shown in the formula:

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (48)$$

3.3.4. *Partial Least Square Discriminant Analysis (PLS-DA)*. PLS-DA is a standard high-dimensional data analysis method, and it is especially suitable for situations where there are a large number of explanatory variables [70], multicollinearity samples, few observations, and large interference noises. PLS-DA first treats the sample category with dummy variables which can be calculated by the following formula:

$$\begin{cases} Yk = 1, Y = k \\ Yk = 0, Y \neq k \end{cases}, (k = 1, 2, \dots, q), \quad (49)$$

where Y represents categorical variables, and q is a dummy variable. The relationship model between explanatory variables, response variables is established by least square regression, and then, the category of each sample is determined by comparing the predicted values of model response variables. If the predicted value of the dummy variable component is the largest, it is determined that the sample belongs to the category corresponding to the dummy variable.

3.3.5. *Classification and Regression Tree (CART)*. CART is a nonparametric statistical process of data analysis. Its characteristic is to make full use of the binary tree structure in the calculation process, that is, the root node contains all samples, and the root node is divided into two children under certain partition rules [71]. The process of node is repeated on the subbook point and becomes a regression process until it can no longer be divided into leaf nodes [72]. When all nodes can be classified into class C ($k = 1, 2, \dots, C$), the Gini impurity of node A can be expressed by the following formula:

$$\text{Gini}(A) = 1 - \sum_{k=1}^C p_k^2, \quad (50)$$

where p^k is the proportion of the sample which belongs to class k . If class A can be divided into B and C . The probability of B in the sample A is p and C is q . And the size of impurities will be expressed by the following formula:

$$\text{Gini}(A) - p \cdot \text{Gini}(B) - q \cdot \text{Gini}(C). \quad (51)$$

3.3.6. *Linear Discriminant Analysis (LDA)*. LDA is a statistical analysis method used to determine the type of sample [73], which has been widely used in the prediction of protein structural classes. By finding the feature vector w , the k sets of m metadata are mapped into another lower-dimensional direction. Then classify with the sample in the new space.

$$w_{opt} = \arg \max_w \frac{|w^T S_B w|}{|w^T S_w w|}. \quad (52)$$

S_B is an interclass deviation matrix:

$$S_B = \sum (m_x - m_y)(m_x - m_y)^T. \quad (53)$$

S_w is an intraclass deviation matrix:

$$S_w = \sum_{i=1}^{N_x} (x_i - m_x)(x_i - m_x)^T + \sum_{i=1}^{N_y} (y_i - m_y)(y_i - m_y)^T, \quad (54)$$

where $\{x_i, i = 1, 2, \dots, N_x\}$ represents class p , $\{y_i, i = 1, 2, \dots, N_y\}$ represents class n . Where m_x and m_y , respectively, represent the average of class p and class n .

3.3.7. *Extreme Gradient Boosting (XGBoost)*. The gradient boosting algorithm framework supports a variety of different loss functions [74]. In addition to the exponential loss function, it also includes the mean square error loss function and logarithmic loss function. XGBoost algorithm is an enhanced version of gradient boosting algorithm. It is more efficient, flexible, and portable [75]. It used the training data x_i to predict a target variable y_i .

First of all, the objective function can be calculated as

$$obj = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^t \Omega(f_i), \quad (55)$$

where n represents the number of trees, l is the training loss function, and Ω is the regularization term.

Then, the XGBoost takes the Taylor expansion of the loss function up to the second order and removes all the constants, so the specific objective at step t can be describes as

$$L^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t), \quad (56)$$

where g_i and h_i can be described as follows:

$$\begin{cases} g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}), \\ h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}). \end{cases} \quad (57)$$

The value of the objective function only depends on g_i and h_i .

It can optimize every loss function, including logistic regression and pairwise ranking, and it is simple to parallel and can greatly enhance the program efficiency with a fast model exploration [76].

3.4. Evaluation Measure. After realizing HPV prediction, we need to use statistical test method to evaluate the efficiency of prediction model. Leave-one-out cross-validation (LOO-CV) is a more common method of Bayesian model and a special case of k -fold cross validation [77], because when k is equal to sample size n , and it can be regarded as n -fold cross validation.

HPV typing is a two or three classification problem. If the HPV model predicts a positive result (P) and the true result is also positive, it is called true positive (TP). If the predicted result is negative (N), the real result is also positive, which is called false positive (FP). The results can be calculated by the following formula:

$$\text{accuracy(ACC)} = \frac{\text{TP} + \text{TN}}{\text{TN} + \text{FP} + \text{FN} + \text{TP}}, \quad (58)$$

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (59)$$

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (60)$$

4. Conclusion

High-risk HPV accounts for a higher proportion of cervical cancer in the world. Therefore, the identification of high-risk HPV is of great significance for the treatment of some major cancers such as cervical cancer [78]. At present, there are common epidemiological detection techniques for HPV typing [79], such as nucleic acid imprinted in situ hybridization or dot blot hybridization, and the detection infection rate is about 10-20% [80]. In addition, there are test methods, such as polymerase chain reaction (PCR). All HPV types can be detected by PCR, and the detection infection rate is more than 40%. In addition to experimental methods, some computational methods have been proposed to predict HPV typing.

We summarize the mathematical models and prediction methods of the risk types of human papillomavirus, especially around the key steps such as feature extraction, feature selection, and prediction algorithms. From the above research papers, it can be found that the prediction accuracy of low-risk types is often higher than that of high-risk types. E7 performed better than other HPV proteins in low-risk experiments. However, E6 performs best among all HPV proteins for high-risk prediction and all types of prediction experiments, which is just consistent with experimental studies. E5, E6, and E7 proteins of high-risk HPV play an important role

in disease progression and cancer [81]. Therefore, E6 protein sequence is more suitable for HPV high-risk prediction, and E7 protein sequence is more suitable for HPV low-risk prediction. However, there are some exceptions. In the reduced amino acid modes, L1 protein performs better in predicting high-risk HPV, while L2 protein is more suitable for low-risk HPV. These conclusions can provide a theoretical basis for subsequent research and guide the establishment of HPV typing models.

Mutations usually lead to cell dysfunction, cell death, and even cancer in higher organisms. Therefore, mutations in HPV protein may make the virus more easily induced or carcinogenic and increase the chance of reinfecting the host or fleeing the host immune system. For example, the carcinogenicity of HPV is mainly controlled by E6 and E7 proteins, which often produce internal variation. The mutation frequency of E6 in cancer is 20%~90%, and that of E7 is 60%~90% [82]. A study in Japan showed that the mutation of aspartate to glutamate (d25e) at position 25 of HPV16 E6 protein was related to the DRB1 * 1502 allele of HLA II. This mutation is considered to be an important mutation in invasive cancer and cervical intraepithelial neoplasia. In the Netherlands, the frequency of HLV DRB1 * 07 allele increased in cancer patients with l83v mutation ($P = 0.08$). If the information of mutation information is considered in HPV classification model, it is also an effective way to improve HPV typing detection.

Data Availability

HPV database can be downloaded from Los Alamos National Laboratory (LANL, <https://lanl.gov>).

Conflicts of Interest

The authors declare no conflicts of interest, financial, or otherwise.

Acknowledgments

The authors thank all the anonymous referees for their valuable suggestions and support. This work is supported by the Natural Science Foundation of Heilongjiang Province of China (LH2019F046).

References

- [1] Y. R. Tzenov, P. G. Andrews, K. Voisey et al., "Human papilloma virus (HPV) E7-mediated attenuation of retinoblastoma (Rb) induces hPygopus2 expression via Elf-1 in cervical cancer," *Molecular Cancer Research*, vol. 11, no. 1, pp. 19-30, 2013.
- [2] Y. Yao, H. Xu, M. Li, Z. Qi, and B. Liao, "Recent advances on prediction of human papillomaviruses risk types," *Current Drug Metabolism*, vol. 20, no. 3, pp. 236-243, 2019.
- [3] J. Haedicke and T. Iftner, "Human papillomaviruses and cancer," *Radiotherapy and Oncology*, vol. 108, no. 3, pp. 397-402, 2013.
- [4] Y. J. Choi, E. Y. Ki, C. Zhang et al., "Analysis of sequence variation and risk association of human papillomavirus 52

- variants circulating in Korea,” *PLoS One*, vol. 11, no. 12, article e0168178, 2016.
- [5] E. M. Burd, “Human papillomavirus laboratory testing: the changing paradigm,” *Clinical Microbiology Reviews*, vol. 29, no. 2, pp. 291–319, 2016.
 - [6] M. M. Pan, C. Gao, X. L. Li et al., “The enhancement of DNA binding ability of a mutated E2 (A338V) protein of HPV-2,” *Chinese Journal of Virology*, vol. 26, no. 3, pp. 223–227, 2010.
 - [7] B. H. Horwitz, D. L. Weinstat, and D. Dimaiio, “Transforming activity of a 16-amino-acid segment of the bovine papillomavirus E5 protein linked to random sequences of hydrophobic amino acids,” *Journal of Virology*, vol. 63, no. 11, pp. 4515–4519, 1989.
 - [8] M. Tommasino and L. Crawford, “Human papillomavirus E6 and E7: proteins which deregulate the cell cycle,” *Bioessays News & Reviews in Molecular Cellular & Developmental Biology*, vol. 17, no. 6, pp. 509–518, 1995.
 - [9] I. Manini and E. Montomoli, “Epidemiology and prevention of human papillomavirus,” *Annali di Igiene : Medicina Preventiva e di Comunita*, vol. 30, 4 Supple 1, pp. 28–32, 2018.
 - [10] M. R. Pillai, S. Lakshmi, S. Sreekala et al., “High-risk human papillomavirus infection and E6 protein expression in lesions of the uterine cervix,” *Pathobiology*, vol. 66, no. 5, pp. 240–246, 1998.
 - [11] M. L. Tornesello, M. L. Duraturo, G. Botti et al., “Prevalence of alpha-papillomavirus genotypes in cervical squamous intraepithelial lesions and invasive cervical carcinoma in the Italian population,” *Journal of Medical Virology*, vol. 78, no. 12, pp. 1663–1672, 2006.
 - [12] M. Arbyn, M. Tommasino, C. Depuydt, and J. Dillner, “Are 20 human papillomavirus types causing cervical cancer?,” *Journal of Pathology*, vol. 234, no. 4, pp. 431–435, 2014.
 - [13] V. Cogliano, R. Baan, K. Straif, Y. Grosse, B. Secretan, and F. E. Ghissassi, “Carcinogenicity of human papillomaviruses,” *Lancet Oncology*, vol. 6, no. 4, pp. 204–204, 2005.
 - [14] D. A. Wick and J. R. Webb, “A novel, broad spectrum therapeutic HPV vaccine targeting the E7 proteins of HPV16, 18, 31, 45 and 52 that elicits potent E7-specific CD8T cell immunity and regression of large, established, E7-expressing TC-1 tumors,” *Vaccine*, vol. 29, no. 44, pp. 7857–7866, 2011.
 - [15] G. M. Clifford, T. Stephen, and F. Silvia, “Carcinogenicity of human papillomavirus (HPV) types in HIV-positive women: a meta-analysis from HPV infection to cervical cancer,” *Clinical Infectious Diseases*, vol. 64, no. 9, pp. 1228–1235, 2017.
 - [16] T. M. Castro, I. Bussoloti Filho, V. X. Nascimento, and S. D. Xavier, “HPV detection in the oral and genital mucosa of women with positive histopathological exam for genital HPV, by means of the PCR,” *Revista Brasileira de Oto-Rino-Laringologia*, vol. 75, pp. 167–171, 2009.
 - [17] M. J. Lace, J. R. Anson, A. J. Klingelutz et al., “Human papillomavirus (HPV) type 18 induces extended growth in primary human cervical, tonsillar, or foreskin keratinocytes more effectively than other high-risk mucosal HPVs,” *Journal of Virology*, vol. 83, no. 22, pp. 11784–11794, 2009.
 - [18] H. Furumoto and M. Irahara, “Human papilloma virus (HPV) and cervical cancer,” *Journal of Medical Investigation*, vol. 49, no. 3–4, pp. 124–133, 2002.
 - [19] F. X. Bosch, M. M. Manos, N. Muñoz et al., “Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) study group,” *Journal of the National Cancer Institute*, vol. 87, no. 11, pp. 796–802, 1995.
 - [20] R. D. Burk, G. Y. Ho, L. Beardsley, M. Lempa, M. Peters, and R. Bierman, “Sexual behavior and partner characteristics are the predominant risk factors for genital human papillomavirus infection in young women,” *The Journal of Infectious Diseases*, vol. 174, no. 4, pp. 679–689, 1996.
 - [21] N. Muñoz, F. X. Bosch, S. de Sanjosé et al., “Epidemiologic classification of human papillomavirus types associated with cervical cancer,” *The New England Journal of Medicine*, vol. 348, no. 6, pp. 518–527, 2003.
 - [22] J. H. Eom, S. B. Park, and B. T. Zhang, “Genetic mining of DNA sequence structures for effective classification of the risk types of human papillomavirus (HPV),” *Lecture Notes in Computer Science*, vol. 3316, pp. 1334–1343, 2004.
 - [23] J. G. Joung, S. June, and B. T. Zhang, “Prediction of the risk types of human papillomaviruses by support vector machines,” in *Pacific Rim International Conference on Artificial Intelligence*, pp. 723–731, Springer, Berlin, Heidelberg, 2004.
 - [24] J. G. Joung, S. June, and B. T. Zhang, “Protein sequence-based risk classification for human papillomaviruses,” *Computers in Biology and Medicine*, vol. 36, no. 6, pp. 656–667, 2006.
 - [25] S. B. Park, S. Hwang, and B. T. Zhang, “Mining the risk types of human papillomavirus (HPV) by AdaCost,” in *International Conference on Database and expert Systems Applications*, pp. 403–412, Prague, Czech Republic, September 2003.
 - [26] K. Sun and J. H. Eom, “Prediction of the human papillomavirus risk types using gap-spectrum kernels,” in *International Conference on Advances in Neural Networks*, pp. 710–715, Springer, Berlin, Heidelberg, 2006.
 - [27] S. Kim and B. T. Zhang, “Human papillomavirus risk type classification from protein sequences using support vector machines,” in *Workshops on Applications of Evolutionary Computation*, pp. 57–66, Springer, Berlin, Heidelberg, 2006.
 - [28] K. Sun, J. Kim, and B. T. Zhang, “Ensembled support vector machines for human papillomavirus risk type prediction from protein secondary structures,” *Computers in Biology and Medicine*, vol. 39, no. 2, pp. 187–193, 2009.
 - [29] M. Esmaeili, H. Mohabatkar, and S. Mohsenzadeh, “Using the concept of Chou’s pseudo amino acid composition for risk type prediction of human papillomaviruses,” *Journal of Theoretical Biology*, vol. 263, no. 2, pp. 203–209, 2010.
 - [30] M. Alemi, H. Mohabatkar, and M. Behbahani, “In silico comparison of low- and high-risk human papillomavirus proteins,” *Applied Biochemistry and Biotechnology*, vol. 172, no. 1, pp. 188–195, 2014.
 - [31] C. Wang, Y. Hai, X. Liu et al., “Prediction of high-risk types of human papillomaviruses using statistical model of protein sequence space,” *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 756345, 9 pages, 2015.
 - [32] X. Xu, R. Kong, X. Liu, P. He, and Q. Dai, “Prediction of high-risk types of human papillomaviruses using reduced amino acid modes,” *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 5325304, 10 pages, 2020.
 - [33] U. C. Megwalu, M. M. Chen, Y. Ma, and V. Divi, “Surrogate for oropharyngeal cancer HPV status in cancer database studies,” *Head & Neck*, vol. 39, no. 12, pp. 2494–2500, 2017.
 - [34] R. P. Bonidia, L. D. H. Sampaio, D. S. Domingues et al., “Feature extraction approaches for biological sequences: a

- comparative study of mathematical features,” *Briefings in Bioinformatics*, vol. 22, no. 5, 2021.
- [35] A. Fiannaca, M. L. Rosa, R. Rizzo, and U. Alfonso, “A k -mer-based barcode DNA classification methodology based on spectral representation and a neural gas network,” *Artificial Intelligence in Medicine*, vol. 64, no. 3, pp. 173–184, 2015.
- [36] D. Qi, W. Li, and L. Li, “Improving protein structural class prediction using novel combined sequence information and predicted secondary structural features,” *Journal of Computational Chemistry*, vol. 32, no. 16, pp. 3393–3398, 2011.
- [37] X. Nan, Q. Xu, and D. Cao, “Generating various numerical representation schemes of protein sequence,” 2015.
- [38] B. Yang, H. S. Yang, J. Li, Z. X. Li, and Y. M. Jiang, “Amino acid composition, molecular weight distribution and antioxidant stability of shrimp processing byproduct hydrolysate,” *American Journal of Food Technology*, vol. 6, no. 10, pp. 904–913, 2011.
- [39] K. C. Chou, “Prediction of protein cellular attributes using pseudo-amino acid composition,” *Proteins-structure Function & Bioinformatics*, vol. 43, no. 3, pp. 246–255, 2001.
- [40] T. K. Attwood, “Profile (Weight Matrix, Position Weight Matrix, Position-Specific Scoring Matrix, PSSM),” in *Dictionary of Bioinformatics and Computational Biology*, American Cancer Society, 2004.
- [41] R. Kong, X. Xu, X. Liu, P. He, M. Q. Zhang, and Q. Dai, “2SigFinder: the combined use of small-scale and large-scale statistical testing for genomic island detection from a single genome,” *BMC Bioinformatics*, vol. 21, no. 1, p. 159, 2020.
- [42] L. Mariño-Ramírez, K. Tharakaraman, O. Bodenreider, J. Spouge, and D. Landsman, “Identification of cis-regulatory elements in gene co-expression networks using A-GLAM,” *Computational Systems Biology*, vol. 541, pp. 3–22, 2009.
- [43] V. S. Gowri, K. G. Tina, O. Krishnadev, and N. Srinivasan, “Strategies for the effective identification of remotely related sequences in multiple PSSM search approach,” *Proteins-structure Function & Bioinformatics*, vol. 67, no. 4, pp. 789–794, 2010.
- [44] Z. H. Chen, Z. H. You, L. P. Li, Y. B. Wang, L. Wong, and H. C. Yi, “Prediction of self-interacting proteins from protein sequence information based on random projection model and fast Fourier transform,” *International Journal of Molecular Sciences*, vol. 20, no. 4, article 930, 2019.
- [45] D. T. Jones, “Protein secondary structure prediction based on position-specific scoring matrices¹,” *Journal of Molecular Biology*, vol. 292, no. 2, pp. 195–202, 1999.
- [46] J. Wu, Y. Z. Li, M. L. Li, and L. Z. Yu, “Two multi-classification strategies used on SVM to predict protein structural classes by using auto covariance,” *Interdisciplinary Sciences Computational Life Sciences*, vol. 1, no. 4, pp. 315–319, 2009.
- [47] T. Liu, X. Geng, X. Zheng, R. Li, and J. Wang, “Accurate prediction of protein structural class using auto covariance transformation of PSI-BLAST profiles,” *Amino Acids*, vol. 42, no. 6, pp. 2243–2249, 2012.
- [48] S. Yang, Y. Wang, Y. Chen, and Q. Dai, “MASQC: next generation sequencing assists third generation sequencing for quality control in N6-methyladenine DNA identification,” *Frontiers in Genetics*, vol. 11, article 269, 2020.
- [49] N. Choudhary, B. S. Biswal, and A. Mohapatra, “Prediction of HPV risk types from protein secondary structure,” in *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, Jabalpur, India, December 2015.
- [50] M. Gk and D. Herand, “Prediction of bacterial virulent proteins with composition moment vector feature encoding method,” *MATEC Web of Conferences*, vol. 49, article 07001, 2016.
- [51] M. Onesime, Z. Yang, and Q. Dai, “Genomic island prediction via chi-square test and random forest algorithm,” *Computational and Mathematical Methods in Medicine*, vol. 2021, 9 pages, 2021.
- [52] Y. Wang, W. Cai, L. Gu, X. Ji, and Q. Shen, “Comprehensive analysis of pertinent genes and pathways in atrial fibrillation,” *Computational and Mathematical Methods in Medicine*, vol. 2021, 20 pages, 2021.
- [53] S. Zhang, S. Ding, and T. Wang, “High-accuracy prediction of protein structural class for low-similarity sequences based on predicted secondary structure,” *Biochimie*, vol. 93, no. 4, pp. 710–714, 2011.
- [54] Q. Dai, Y. Li, X. Liu, Y. Yao, Y. Cao, and P. He, “Comparison study on statistical features of predicted secondary structures for protein structural class prediction: from content to position,” *BMC Bioinformatics*, vol. 14, no. 1, pp. 1–14, 2013.
- [55] J. Tao, X. Liu, S. Yang, C. Bao, P. He, and Q. Dai, “An efficient genomic signature ranking method for genomic island prediction from a single genome,” *Journal of Theoretical Biology*, vol. 467, pp. 142–149, 2019.
- [56] J. V. Kuzhali, G. Rajendran, V. Srinivasan, and G. S. Kumar, “Feature selection algorithm using fuzzy rough sets for predicting cervical cancer Risks,” *Modern Applied Science*, vol. 4, no. 8, pp. 134–143, 2010.
- [57] F. Fleuret, “Fast binary feature selection with conditional mutual information,” *Journal of Machine Learning Research*, vol. 5, no. 4941, pp. 1531–1555, 2004.
- [58] H. Lim and D. W. Kim, “Using mutual information for selecting features in multi-label classification,” *Journal of KISS: Software and Applications*, vol. 39, no. 10, pp. 806–811, 2012.
- [59] Y. Ying, “SVM-RFE algorithm for gene feature selection,” *Computer Engineering*, 2005.
- [60] Y. Mao, X. Zhou, Z. Yin, D. Pi, Y. Sun, and S. T. Wong, “Gene selection using Gaussian kernel support vector machine based recursive feature elimination with adaptive kernel width strategy,” in *International Conference on Rough Sets and Knowledge Technology*, pp. 799–806, Springer, Berlin, Heidelberg, 2006.
- [61] B. Xi, J. Tao, X. Liu, X. Xu, P. He, and Q. Dai, “RaaMLab: a MATLAB toolbox that generates amino acid groups and reduced amino acid modes,” *Biosystems*, vol. 180, pp. 38–45, 2019.
- [62] Z. Yang, W. Yi, J. Tao et al., “HPVMD-C: a disease-based mutation database of human papillomavirus in China,” *Database: The Journal of Biological Databases and Curation*, vol. 2022, 2022.
- [63] E. Topaka, “Application of genetic algorithms and other feature selection techniques in clinical decision support for cervical cancer diagnosis,” 2016.
- [64] F. J. Rubio, *Modelling of Kurtosis and Skewness: Bayesian Inference and Distribution Theory*, University of Warwick, England, 2013.
- [65] R. Durgabai, “Feature selection using ReliefF algorithm,” *IJARCCCE*, vol. 10, no. 30, pp. 8215–8218, 2014.

- [66] H. E. Tao, H. U. Jie, X. I. A. Peng, and G. U. Chaochen, "Feature selection of Emg signal based on ReliefF algorithm and genetic algorithm," *Journal of Shanghai Jiaotong University*, vol. 50, no. 2, article 204, 2016.
- [67] I. I. Suni, Y. Huang, and S. Schuckers, *Bioelectronic Tongue for Food Allergy Detection*, US, 2010.
- [68] I. M. A. Melo, M. R. P. Viana, B. Pupin, T. T. Bhattacharjee, and R. de Azevedo Canevari, "PCR-RFLP and FTIR-based detection of high-risk human papilloma virus for cervical cancer screening and prevention," *Biochemistry and Biophysics Reports*, vol. 26, article 100993, 2021.
- [69] E. Y. Boateng, J. Otoo, and D. A. Abaye, "Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review," *Journal of Data Analysis and Information Processing*, vol. 8, no. 4, pp. 341–357, 2020.
- [70] L. L. Chuen, L. Choong-Yeun, and J. A. Aziz, "Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps," *Analyst*, vol. 143, no. 15, pp. 3526–3539, 2018.
- [71] Q. Dai, C. Bao, Y. Hai et al., "MTGIpick allows robust identification of genomic islands from a single genome," *Briefings in Bioinformatics*, vol. 19, no. 3, pp. 361–373, 2016.
- [72] C. J. Yang, Y. C. Tsai, and J. J. Tien, "Patients with minor diseases who access high-tier medical care facilities: new evidence from classification and regression trees," *The International Journal of Health Planning and Management*, vol. 34, no. 2, 2019.
- [73] H. Jones, C. Gardner, E. Hull, K. Nixon, M. Robinson, and Rio Grande Medical Technologies Inc, "Within-sample variance classification of samples," 2003.
- [74] N. A. Bokulich, M. Dillon, E. Bolyen, B. Kaehler, G. Huttley, and J. Caporaso, "q2-sample-classifier: machine-learning tools for microbiome classification and regression," *The Journal of Open Source Software*, vol. 3, no. 30, 2018.
- [75] X. Wang, Y. Wang, Z. Xu, Y. Xiong, and D. Q. Wei, "Prediction of the classes of anatomical therapeutic chemicals using a network-based label space partition method," *Frontiers in Pharmacology*, vol. 10, article 971, 2019.
- [76] T. Chen and C. Guestrin, "A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, United States, August 2016.
- [77] T. Sivula, M. Magnusson, and A. Vehtari, "Uncertainty in Bayesian leave-one-out cross-validation based model comparison," 2020, <http://arxiv.org/abs/2008.10296>.
- [78] J. Wu, T. Zhou, J. Tao et al., "Similarity/dissimilarity analysis of protein structures based on Markov random fields," *Computational Biology & Chemistry*, vol. 75, pp. 45–53, 2018.
- [79] M. A. Cohenford and B. Lentricchia, "Detection and typing of human papillomavirus using PNA probes," 2007.
- [80] M. Kalantari, E. Blennow, B. Hagmar, and B. Johansson, "Physical state of HPV16 and chromosomal mapping of the integrated form in cervical carcinomas," *Diagnostic Molecular Pathology*, vol. 10, no. 1, pp. 46–54, 2001.
- [81] L. C. Barrow, "e7 proteins of high-risk (type 16) and low-risk (type 6) human papillomaviruses regulate p130 differently," 2010.
- [82] D. M. Da Silva, G. L. Eiben, S. C. Fausch et al., "Cervical cancer vaccines: emerging concepts and developments," *Journal of Cellular Physiology*, vol. 186, no. 2, pp. 169–182, 2001.

Research Article

TRIM66 Promotes Malignant Progression of Non-Small-Cell Lung Cancer Cells via Targeting MMP9

Yufen Xu ¹, Qi Yang ², Zhixian Fang ², Xiaoli Tan ², Ming Zhang ²,
and Wenyu Chen ²

¹Department of Oncology, Affiliated Hospital of Jiaxing University, No. 1882, Zhonghuan South Road, Nanhu District, Jiaxing, Zhejiang Province 314001, China

²Department of Respiratory Medicine, Affiliated Hospital of Jiaxing University, Jiaxing, Zhejiang Province 314001, China

Correspondence should be addressed to Wenyu Chen; 00135116@zjxu.edu.cn

Received 18 May 2022; Revised 1 July 2022; Accepted 7 July 2022; Published 21 July 2022

Academic Editor: Tao Huang

Copyright © 2022 Yufen Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lung cancer has a higher incidence and mortality rate than other cancers, and over 80% of lung cancer cases were classified as non-small-cell lung cancer (NSCLC). TRIM66 is one of the crucial members of TRIM, which has a deep connection with the behavior of various malignant tumors. But it remains uncertain regarding its exact function and underlying mechanism in NSCLC. In our study, qRT-PCR and Western blot were employed to validate that TRIM66 was overexpressed in NSCLC. The migration, invasion, and epithelial-mesenchymal transformation (EMT) progression of NSCLC cells were determined by Western blotting and Transwell experiments after knocking down TRIM66, and it was found that knockdown TRIM66 inhibited the migration, invasion, and EMT processes of NSCLC cells. Next, the binding relationship between TRIM66 and MMP9 was verified by Co-IP assay. After determining the interaction between them, rescue assays showed that overexpression of MMP9 was capable to promote the migration, invasion, and EMT of NSCLC cells. However, the transfection of si-TRIM66 could reverse this facilitating effectiveness. To sum up, we concluded that by targeting MMP9, TRIM66 could exert a cancer-promoting role in the progression of NSCLC cells.

1. Introduction

The global cancer statistics indicated that the number of lung cancer diagnoses was about 1.8 million each year, which accounts for 13% of all newly diagnosed cancer cases [1]. Non-small-cell lung cancer (NSCLC) and SCLC are two subtypes of lung cancer [2]. NSCLC accounts for more than 80% of lung cancer cases [3]. And most patients present with advanced NSCLC at diagnosis. Though recent years have witnessed lots of improvement in early diagnosis and treatment of lung cancer, most of the treatment for NSCLC is surgery, chemotherapy, and radiation therapy [4]; the 5-year survival rate of NSCLC patients does not exceed 20% [3]. However, the genes associated with NSCLC and potential therapeutic targets are unclear. Hence, the current

research focuses on identifying effective prognostic markers and possible therapeutic targets in NSCLC.

The tripartite motif-containing protein (TRIM) family participates in innate immunity to viruses, cell cycle, apoptosis, and so on [5, 6]. As one of the members, TRIM66 takes a part in the behavior of various cancers [7]. Existing studies indicate that the abnormal expression of TRIM66 in NSCLC can promote lymphatic and distant metastasis, which is a negative factor affecting prognosis [8]. Epithelial-mesenchymal transition (EMT) is an essential part of tumorigenesis as well as metastasis [8, 9]. The loss of E-cadherin and the increase in migratory and invasive behaviors and elevated levels of vimentin and N-cadherin are the features of EMT [10]. He et al. [11] found that the knockdown of TRIM66 can inhibit the EMT process. However, it remains

unclear about the role of TRIM66 in NSCLC progression. And the enzymes that regulate the role of TRIM66 in NSCLC are not clear to us. Hence, research is needed.

Matrix metalloproteinase 9 (MMP9) is a type IV collagenase that plays a crucial part in promoting cell migration and reepithelialization [12]. MMP9 is implicated with the malignant progression of cancer, including but not limited to invasion [13, 14], migration [15], metastasis [16], and angiogenesis [17]. On the basis of above findings, the speculation that MMP9 may be a potential biomarker and target has been investigated. However, how MMP9 works in NSCLC is rarely studied.

Our study focuses on TRIM66 by validating its expression level in NSCLC. Its targeted relationship in NSCLC with MMP9 was validated through CHIP experiment. Then, the migration, invasion, and epithelial-mesenchymal transformation (EMT) progression of NSCLC cells were determined by Western blotting and Transwell experiments after knocking down TRIM66. Our study can bring an improved comprehension of molecular mechanisms of NSCLC malignant progression and may provide potential targets for NSCLC.

2. Materials and Methods

2.1. Cell Culture. Human pulmonary alveolar epithelial cells (HPAEPiC) were procured from Shanghai Zhong Qiao Xin Zhou Biotechnology Co., Ltd, and NSCLC cell lines H460, H1299, and A549 were bought from Shanghai Cell Bank of the Chinese Academy of Sciences (Shanghai, China). Conditions for the cell culturing were as follows: alveolar epithelial cell culture medium for HPAEPiC, RPMI-1640 culture medium containing 10% high-quality fetal bovine serum (FBS) for H460 and H1299 (GIBCO, Art. No. 31800022, supplemented with 2.5 g/L glucose, 1.5 g/L NaHCO₃, and 0.11 g/L sodium pyruvate, USA), and F12K medium containing 10% high-quality FBS for A549 (Sigma, Art. No. N3520, supplemented with 2.5 g/L NaHCO₃, Germany). The culture conditions were 37°C and 5% CO₂.

2.2. Plasmid Construction and Cell Transfection. siRNA of TRIM66 (si-TRIM66) and corresponding negative control (si-NC), overexpressed plasmid of MMP9 (oe-MMP9), and corresponding negative control (oe-NC) was designed by Sangon Biotech (Shanghai, China). All transfections were performed with Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA). Cells were collected 48 h after transfection.

2.3. RNA Purification and qRT-PCR Analysis. RNA extraction was conducted complying strictly with the instructions of TRIzol (Life Technologies Corporation of Carlsbad, California, USA). Then, the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, USA) was employed for the reverse transcription of RNA into cDNA. The SYBR Green PCR kit (Takara Bio, Otsu, Japan) was utilized for performing PCR amplification on a StepOne Real-Time PCR System (Thermo Fisher Scientific, USA). And the relative gene expression normalized by β -actin was calculated

using the 2^{- $\Delta\Delta$ Ct} method. The sequences of the PCR primers were as follows:

TRIM66 forward primer 5'-GCCCTCTGTGCTACTTACTCTC-3', reverse primer 5'-GCTGGTTGTGGGGGTTACTCTC-3'

β -Actin forward primer 5'-CATGTACGTTGCTATCAGGC-3', reverse primer 5'-CTCCTTAATGTACAGCACGAT-3'

2.4. Western Blot. After lysing cells by lysis buffer, a bicinchoninic acid protein assay kit (Thermo Fisher Scientific, USA) was utilized for the measurement of protein concentration. Then, after electrophoresis in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), proteins were transferred onto polyvinylidene difluoride membranes (Millipore, USA). Next, membranes were blocked with 5% skimmed milk. 2 h later, membranes were incubated with primary antibodies overnight at 4°C, which were then incubated with secondary antibody at room temperature for 1 h the next day. The membranes were finally developed following the electrochemiluminescence (ECL) kit (Pierce Biotechnology, USA). Western blot images were acquired using a ChemiDoc imaging system (Bio-Rad, USA). Antibody information is detailed in Table 1, and the internal reference used here was β -actin. Antibodies were all purchased from Invitrogen (Thermo Fisher Scientific, USA).

2.5. Transwell Assays. Transwell inserts (8 μ M pore size, Costar, Cambridge, MA, USA) were employed for the assessment of cell migration as well as invasion. 2×10^4 cells were seeded in inserts without or precoated with Matrigel (BD, Franklin Lakes, USA) in the basolateral membrane using 200 μ l of FBS-free medium for migration or invasion analysis, respectively. 600 μ l of medium with 10% FBS was placed in the lower chamber. After incubation for 24 h at 37°C in 5% CO₂, cells from the upper chamber that failed to migrate as well as invade were swabbed. Then, methanol was used for the cell fixing, 0.5% crystal violet for cell staining, and phosphate buffered saline (PBS) (Gibco; Thermo Fisher Scientific, Inc., USA) for cell washing. They were then photographed and counted under a microscope (Zeiss, Germany).

2.6. Co-Immunoprecipitation (Co-IP) Assay. Cells were lysed for 30 min in Co-IP buffer supplemented with protease inhibitor mixture (Sigma-Aldrich, USA). The centrifugation of these lysates was conducted for 15 min at 12,000 rpm. The supernatant was then incubated with 20 μ l protein A/G beads (Santa Cruz, USA) for 30 min, followed by centrifugation at 1000g for 5 min at 4°C. Next, the immunoprecipitation of proteins lasted for more than 4 h by using TRIM66 antibody or control IgG antibody at 4°C. Protein A/G was added to capture antigen-antibody complexes. And through centrifugation, those agarose bead-antibody antigen complexes were collected and then washed three times using PBS. Next, they were eluted in boiling protein sample buffer under reducing conditions. In the end, proteins were separated by SDS-PAGE and analyzed by Western blot.

TABLE 1: Information on antibodies used in the experiment.

Type	Name	Art. no.	Dilution rate
Primary antibodies (rabbit antibody)	TRIM66 polyclonal antibody	PA5-69788	1.0 $\mu\text{g}/\text{mL}$
	MMP2 polyclonal antibody	PA5-85197	1 : 1000
	MMP9 polyclonal antibody	PA5-13199	1 : 2000
	E-cadherin polyclonal antibody	PA5-32178	1 : 1000
	N-cadherin polyclonal antibody	PA5-19486	1.0 $\mu\text{g}/\text{mL}$
	Vimentin polyclonal antibody	PA5-27231	1 : 2000
	SNAIL polyclonal antibody	PA5-11923	1 : 1000
	β -Actin polyclonal antibody	PA5-16914	1.0 $\mu\text{g}/\text{mL}$
Secondary antibody (goat anti-rabbit)	Goat anti-rabbit IgG H&L	A32731	0.1 $\mu\text{g}/\text{mL}$

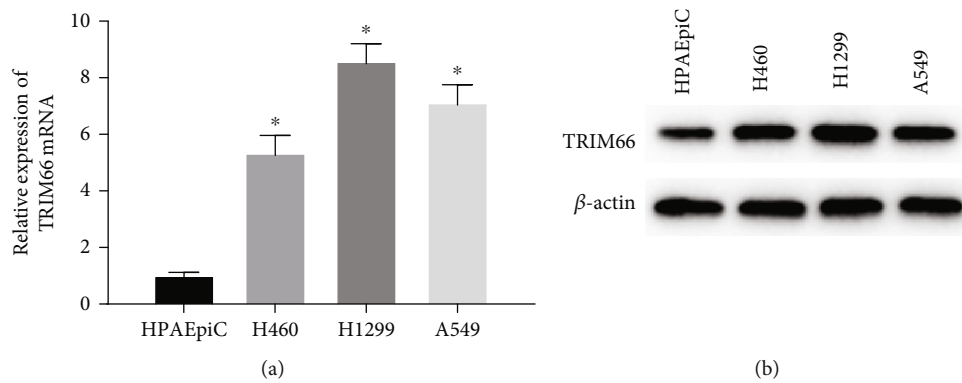


FIGURE 1: TRIM66 is highly expressed in NSCLC cells. (a, b) Relative expression of TRIM66 mRNA and protein in human pulmonary alveolar epithelial and NSCLC cells; $*p < 0.05$.

2.7. Statistical Analysis. SPSS 22.0 (IBM Corp., Armonk, NY, USA) and GraphPad Prism 6.0 software (GraphPad Inc., San Diego, CA, USA) were utilized for data analysis. All measured data were presented as mean \pm SD. The comparison between the two groups was testified by t -test, and $p < 0.05$ indicated a significant difference.

3. Results

3.1. TRIM66 Is Highly Expressed in NSCLC Cells. TRIM66 expression in HPAEpiC, H460, H1299, and A549 was analyzed by Q-PCR, the result of which revealed a higher level of TRIM66 in NSCLC cells than that in human pulmonary alveolar cells (Figure 1(a)). The result of Western blot assay is consistent with the finding of Q-PCR (Figure 1(b)). These all displayed that TRIM66 was abnormally increased in NSCLC cells. In addition, we selected H1299 and A549 which enjoyed the highest expression for subsequent experiments.

3.2. TRIM66 Downregulation Restrains Invasion, Migration, and EMT Process of NSCLC Cells. Transfection of si-TRIM66 plasmid into NCI-H1299 cells and A549 cells was performed to assess the expression of TRIM66 and biological behaviors of TRIM66 in NSCLC cells. Here, TRIM66 expression was knocked down at first. And the transfection efficiency was examined using Q-PCR as well as Western

blot (Figures 2(a) and 2(b)). Next, we conducted a Transwell assay for the detection of the effect that TRIM66 exerted on cell migration and invasion. Silencing TRIM66 did exert remarkable inhibiting effects on cell migration and invasion (Figures 2(c) and 2(d)). Then, we examined EMT-related proteins by Western blot, from the result of which we observed that the expression of N-cadherin, vimentin, and SNAIL was downregulated, and E-cadherin expression was upregulated in both cell lines. It was further demonstrated that the knockdown of siRNA-mediated TRIM66 restrained the EMT process (Figure 2(e)). Therefore, we suggested that knockdown of TRIM66 can restrain migration, invasion, and EMT process of NSCLC cells.

3.3. The Bindings between TRIM66 and MMP2 or MMP9 Are Verified. MMP2 and MMP9 are considered to exert an important role in tumor metastasis [18]. To dig deeper into the regulatory mechanism that TRIM66 works in NSCLC cells, we tried to figure out the interactions between TRIM66, MMP2, and MMP9 by Western blot. Knockdown of TRIM66 could downregulate both MMP2 and MMP9 expression when compared to the control group (Figure 3(a)). Further, Co-IP was performed to investigate whether there was a physical binding between TRIM66 and MMP2 or MMP9, and the results revealed that Co-IP occurred only between TRIM66 and MMP9 (Figure 3(b)). Next, MMP9 level in NSCLC and pulmonary alveolar

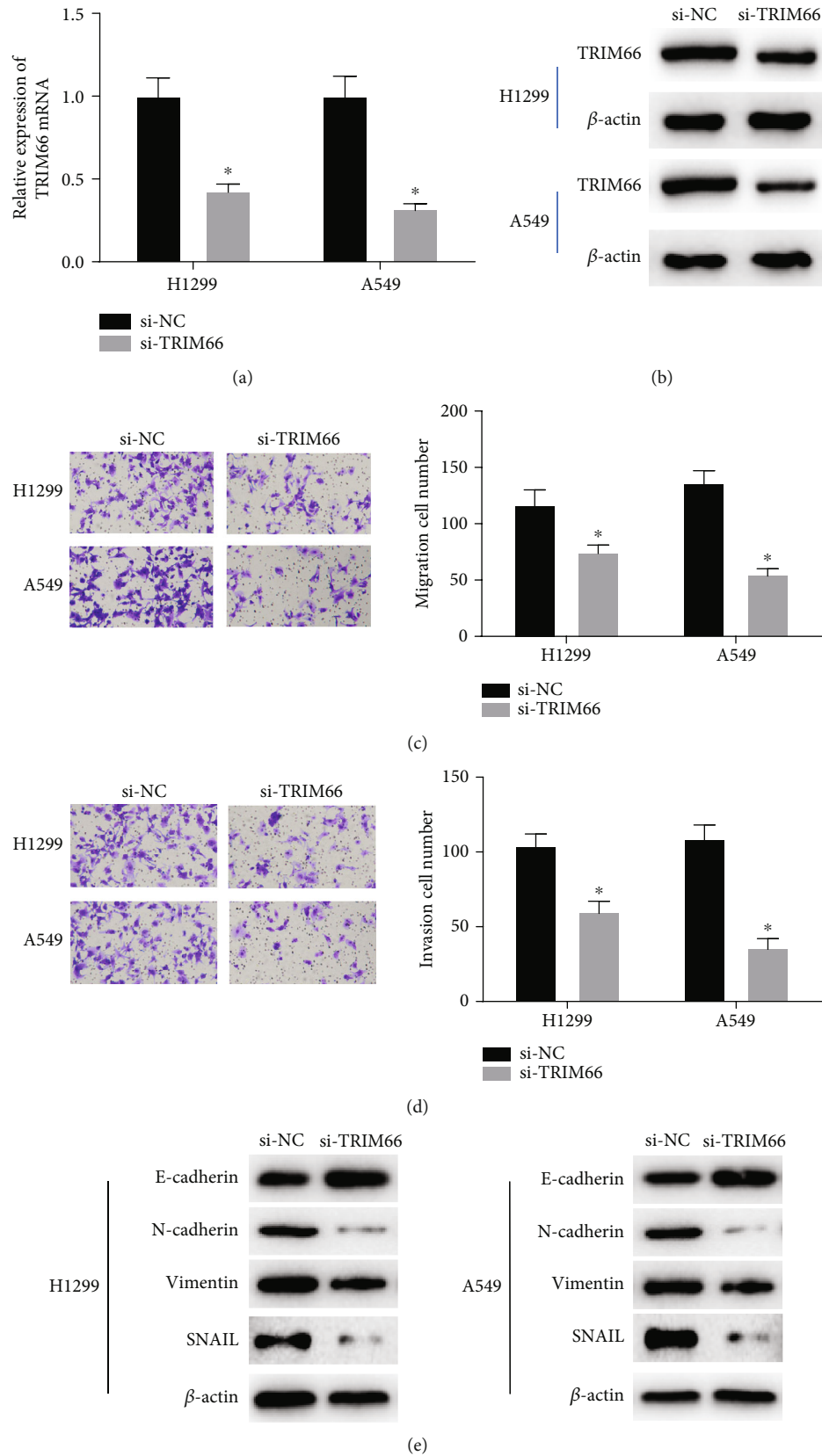


FIGURE 2: TRIM66 knockdown inhibits the migration and invasion of NSCLC cells and EMT procession. (a, b) Expression levels of TRIM66 mRNA and protein in NSCLC cells after transfection with si-TRIM66. (c, d) Influence of knockdown of TRIM66 on cell migration and invasion. (e) Levels of EMT-related proteins; * $p < 0.05$.

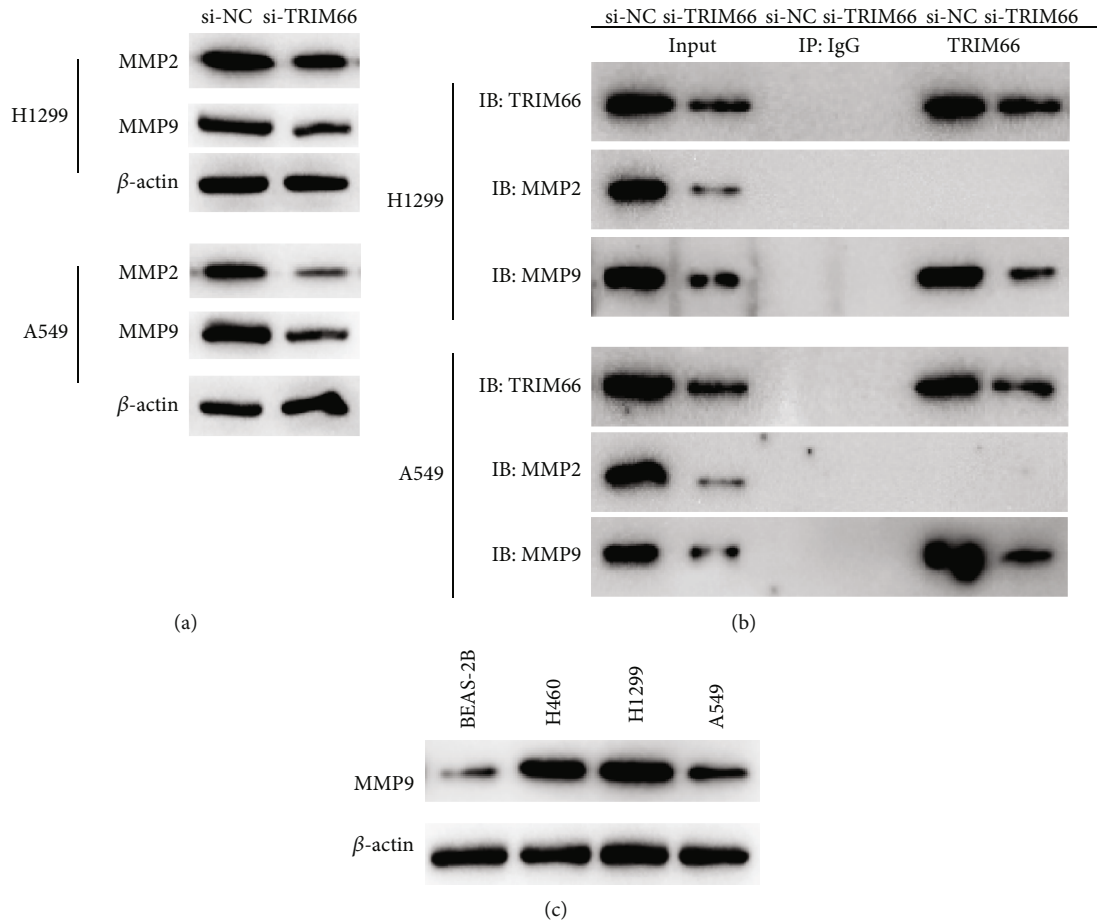


FIGURE 3: The bindings between TRIM66 and MMP2 or MMP9. (a) Protein levels of MMP2 and MMP9 in each transfection group. (b) Cells were lysed and immunoprecipitated with antibodies, and immunocomplexes were analyzed by Western blot. (c) MMP9 level in proteins extracted from NSCLC and pulmonary alveolar epithelial cells.

epithelial cells was also examined by Western blot, demonstrating that MMP9 was upregulated in NSCLC cell lines (Figure 3(c)). Based on the abovementioned findings, we could conclude that there was a direct interaction between TRIM66 and MMP9. And MMP9 displayed a high expression level in NSCLC cells.

3.4. TRIM66 Affects the Migration, Invasion, and EMT of NSCLC Cells by Mediating MMP9. To further verify the regulatory relationship between TRIM66 and MMP9 in NSCLC cells, we transfected si-NC + oe-NC, si-NC + oe-MMP9, and si-TRIM66 + oe-MMP9 into H1299 cells and A549 cells. Firstly, transfection efficiency testified by Western blot, illustrating that MMP9 displayed a remarkably upregulated expression in the si-NC + oe-MMP9 group relevant to the control group, but after simultaneous silencing TRIM66, the upregulation in MMP9 expression decreased (Figure 4(a)). Secondly, Transwell assay results demonstrated that enforced expression of MMP9 fostered migration and invasion of NCI-H1299 and A549 cells, while transfection of si-TRIM66 could reverse the promoting effect of oe-MMP9 (Figures 4(b) and 4(c)). Next, we found

that compared with controls, overexpression of MMP9 only upregulated N-cadherin, vimentin, and SNAIL expression and downregulated E-cadherin expression relevant to the control group, which was conducive to EMT. However, simultaneous silencing of TRIM66 expression could revert these results compared with overexpression of MMP9 alone (Figure 4(d)). Thus, influence of overexpressed MMP9 on NSCLC cell migration, invasion, and EMT was reversed by silencing TRIM66.

4. Discussion

In the past decade, a variety of oncogenes have been discovered, such as KRAS [19] and KAI1/CD82 [20]. And the previous studies have shown that these oncogenes can be taken as therapeutic targets to improve patient survival. The literature suggests that abnormal TRIM expression has a close association with the occurrence and progression of NSCLC. For example, highly expressed TRIM29 in NSCLC tissues indicates poor prognosis for NSCLC patients [21]. TRIM59 has a high expression level in a variety of lung cancer cell lines, while knockdown of which can affect the expression

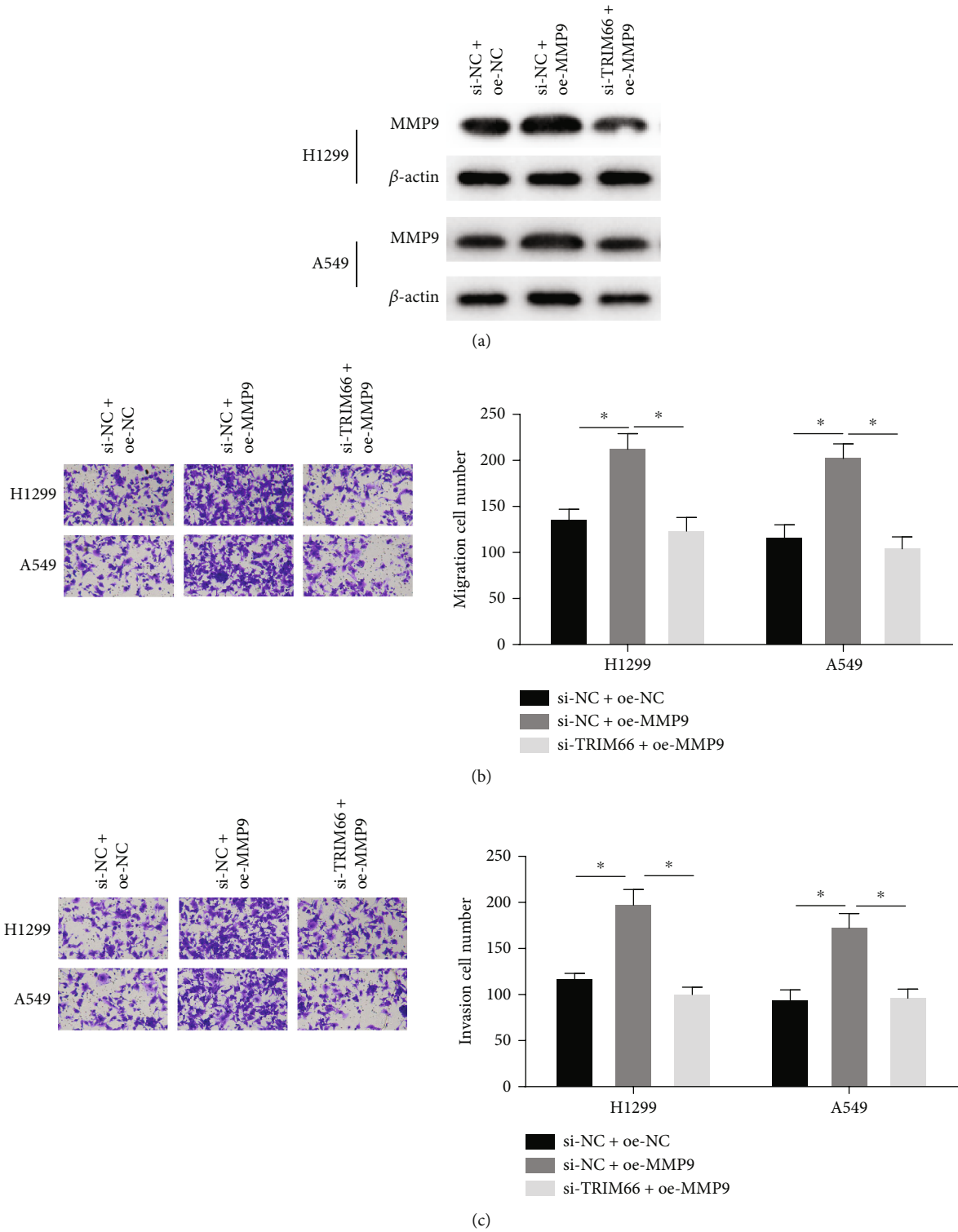


FIGURE 4: Continued.

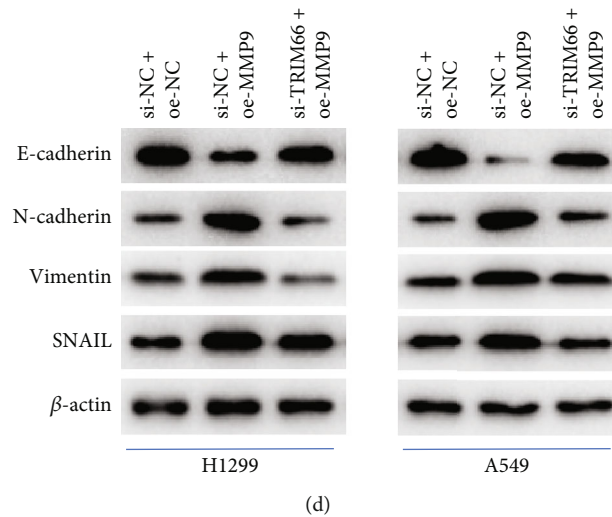


FIGURE 4: TRIM66 affects progression of NSCLC cells by regulating MMP9. (a) Protein level of MMP9 in each group. (b, c) Migration and invasion of cells in each group. (d) Expression levels of EMT-related proteins; * $p < 0.05$.

of cyclins (including CDC25C and CDK1), thereby enhancing the proliferative and migratory abilities of lung cancer cells [22]. TRIM44 displayed an upregulated expression in lung cancer cells, substantially facilitating invasion and migration of cancer cells [23]. However, the topic related to both TRIM66 and NSCLC is scarcely touched upon by researchers. In our study, q-PCR and Western blot pointed out that TRIM66 displayed a high expression level in NSCLC cells, which is consistent with the results of an existing study [24], which suggests that it may exert a promotive function in NSCLC.

Next, we knocked down TRIM66 expression to study its effect on the cellular phenotype of NSCLC and found that knocking down TRIM66 attenuated migration, invasion, and EMT process of cancer cells. Matrix metalloproteinases MMP2 and MMP9 are implicated in tumor angiogenesis and invasion [25], and these processes are regulated by some genes. For example, TGF- β mediates the regulation of MMP2 expression by exosome lnc-MMP2-2 to promote lung cancer cell invasion and migration [26]. EZH2 inhibits the transcription of TIMP2, which boosts activities of MMP-2 and MMP-9, which in turn increases the invasive activity of triple-negative breast cancer cells [27]. Therefore, influence of TRIM66 on migration and invasion of NSCLC cells may be realized by regulation of MMP2 or MMP9. Thus, we performed Co-IP to study the interaction between TRIM66 and MMP2 or MMP9 and found that there was a reciprocal relationship between TRIM66 and MMP9 only. Next, we found that silencing TRIM66 was capable of reversing promoting effect of MMP9 forced expression on migration, invasion, and EMT process of NSCLC cells by rescue experiment. Combined with the literature and our experimental results, TRIM66 promotes invasion, migration, and EMT in NSCLC by modulating MMP9.

Overall, we identified that both TRIM66 and MMP9 were upregulated in NSCLC cells. Silencing TRIM66 inhibits malignant progression of NSCLC cells, and this process is

achieved by regulating MMP9. The findings may provide clues for elucidating the tumorigenesis of NSCLC and may offer a theoretical basis for generating novel approaches for NSCLC diagnosis and therapy.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Authors' Contributions

Wenyu Chen participated in the design and interpretation of the data and drafted the manuscript. Yufen Xu and Qi Yang conceived the study, participated in its design and interpretation, and helped in drafting the manuscript. Zhixian Fang, Xiaoli Tan, and Ming Zhang participated in its design and interpretation and helped in revising the manuscript critically. Yufen Xu and Qi Yang contributed equally to this work.

Acknowledgments


This work was supported by the Natural Science Foundation of Zhejiang Province (No. LQ20H160057), the Key Discipline of Jiaxing Respiratory Medicine Construction Project (No. 2019-zc-04), the Scientific Technology Plan Program for Healthcare in Zhejiang Province (No. 2021RC031), the Science and Technology Project of Jiaxing (2019AY32030, 2020AY30012, 2021AY30024), and the Jiaxing Key Laboratory of Precision Treatment for Lung Cancer.

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: a Cancer Journal for Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.
- [3] D. S. Ettinger, D. E. Wood, D. L. Aisner et al., "Non-small cell lung cancer, version 5.2017, NCCN clinical practice guidelines in oncology," *Journal of the National Comprehensive Cancer Network*, vol. 15, no. 4, pp. 504–535, 2017.
- [4] M. Evison and U. K. L. AstraZeneca, "The current treatment landscape in the UK for stage III NSCLC," *British Journal of Cancer*, vol. 123, no. S1, pp. 3–9, 2020.
- [5] H. Y. Dai, Y. Ma, Z. Da, and X. M. Hou, "Knockdown of TRIM66 inhibits malignant behavior and epithelial-mesenchymal transition in non-small cell lung cancer," *Pathology, Research and Practice*, vol. 214, no. 8, pp. 1130–1135, 2018.
- [6] S. Hatakeyama, "TRIM family proteins: roles in autophagy, immunity, and carcinogenesis," *Trends in Biochemical Sciences*, vol. 42, no. 4, pp. 297–311, 2017.
- [7] K. Ozato, D. M. Shin, T. H. Chang, and H. C. Morse, "TRIM family proteins and their emerging roles in innate immunity," *Nature Reviews. Immunology*, vol. 8, no. 11, pp. 849–860, 2008.
- [8] Y. Ma, H. Y. Dai, F. Zhang, and D. Zhao, "TRIM66 expression in non-small cell lung cancer: a new predictor of prognosis," *Cancer Biomarkers*, vol. 20, no. 3, pp. 309–315, 2017.
- [9] H. Nakashima, N. Hashimoto, D. Aoyama et al., "Involvement of the transcription factor twist in phenotype alteration through epithelial-mesenchymal transition in lung cancer cells," *Molecular Carcinogenesis*, vol. 51, no. 5, pp. 400–410, 2012.
- [10] S. Thomson, F. Petti, I. Sujka-Kwok et al., "A systems view of epithelial-mesenchymal transition signaling states," *Clinical & Experimental Metastasis*, vol. 28, no. 2, pp. 137–155, 2011.
- [11] T. He, J. Cui, Y. Wu, X. Sun, and N. Chen, "Knockdown of TRIM66 inhibits cell proliferation, migration and invasion in colorectal cancer through JAK2/STAT3 pathway," *Life Sciences*, vol. 235, p. 116799, 2019.
- [12] H. Huang, "Matrix metalloproteinase-9 (MMP-9) as a cancer biomarker and MMP-9 biosensors: recent advances," *Sensors*, vol. 18, no. 10, p. 3249, 2018.
- [13] Q. Xue, L. Cao, X. Y. Chen et al., "High expression of MMP9 in glioma affects cell proliferation and is associated with patient survival rates," *Oncology Letters*, vol. 13, no. 3, pp. 1325–1330, 2017.
- [14] S. W. Chen, Q. Zhang, Z. F. Xu et al., "HOXC6 promotes gastric cancer cell invasion by upregulating the expression of MMP9," *Molecular Medicine Reports*, vol. 14, no. 4, pp. 3261–3268, 2016.
- [15] J. Tripathy, A. Tripathy, M. Thangaraju, M. Suar, and S. Elangovan, " α -Lipoic acid inhibits the migration and invasion of breast cancer cells through inhibition of TGF β signaling," *Life Sciences*, vol. 207, pp. 15–22, 2018.
- [16] H. Zhang, C. Hao, Y. Wang et al., "Sohlh2 inhibits human ovarian cancer cell invasion and metastasis by transcriptional inactivation of MMP9," *Molecular Carcinogenesis*, vol. 55, no. 7, pp. 1127–1137, 2016.
- [17] H. Dong, H. Diao, Y. Zhao et al., "Overexpression of matrix metalloproteinase-9 in breast cancer cell lines remarkably increases the cell malignancy largely via activation of transforming growth factor beta/SMAD signalling," *Cell Proliferation*, vol. 52, no. 5, article e12633, 2019.
- [18] G. Shay, C. C. Lynch, and B. Fingleton, "Moving targets: emerging roles for MMPs in cancer progression and metastasis," *Matrix Biology*, vol. 44–46, pp. 200–206, 2015.
- [19] M. Román, I. Baraibar, I. López et al., "KRAS oncogene in non-small cell lung cancer: clinical perspectives on the treatment of an old target," *Molecular Cancer*, vol. 17, no. 1, p. 33, 2018.
- [20] V. V. Prabhu and S. N. Devaraj, "KAI1/CD82, metastasis suppressor gene as a therapeutic target for non-small-cell lung carcinoma," *Journal of Environmental Pathology, Toxicology and Oncology*, vol. 36, no. 3, pp. 269–275, 2017.
- [21] X. Song, C. Fu, X. Yang, D. Sun, X. Zhang, and J. Zhang, "Tripartite motif-containing 29 as a novel biomarker in non-small cell lung cancer," *Oncology Letters*, vol. 10, no. 4, pp. 2283–2288, 2015.
- [22] W. Zhan, T. Han, C. Zhang et al., "TRIM59 promotes the proliferation and migration of non-small cell lung cancer cells by upregulating cell cycle related proteins," *PLoS One*, vol. 10, no. 11, article e0142596, 2015.
- [23] Q. Luo, H. Lin, X. Ye, J. Huang, S. Lu, and L. Xu, "Trim44 facilitates the migration and invasion of human lung cancer cells via the NF- κ B signaling pathway," *International Journal of Clinical Oncology*, vol. 20, no. 3, pp. 508–517, 2015.
- [24] Y. Zhang, L. Wu, C. Jiang, and B. Yan, "Reprogramming cellular signaling machinery using surface-modified carbon nanotubes," *Chemical Research in Toxicology*, vol. 28, no. 3, pp. 296–305, 2015.
- [25] P. Farina, E. Tabouret, P. Lehmann et al., "Relationship between magnetic resonance imaging characteristics and plasmatic levels of MMP2 and MMP9 in patients with recurrent high-grade gliomas treated by bevacizumab and irinotecan," *Journal of Neuro-Oncology*, vol. 132, no. 3, pp. 433–437, 2017.
- [26] D. M. Wu, S. H. Deng, T. Liu, R. Han, T. Zhang, and Y. Xu, "TGF- β -mediated exosomal lnc-MMP2-2 regulates migration and invasion of lung cancer cells to the vasculature by promoting MMP2 expression," *Cancer Medicine*, vol. 7, no. 10, pp. 5118–5129, 2018.
- [27] Y. C. Chien, L. C. Liu, H. Y. Ye, J. Y. Wu, and Y. L. Yu, "EZH2 promotes migration and invasion of triple-negative breast cancer cells via regulating TIMP2-MMP-2/-9 pathway," *American Journal of Cancer Research*, vol. 8, no. 3, pp. 422–434, 2018.

Research Article

Dysregulation of Circadian Clock Genes Associated with Tumor Immunity and Prognosis in Patients with Colon Cancer

Yongshan He, Yuanyuan Chen, Xuan Dai, and Shiyong Huang 

Department of Colorectal Surgery, School of Medicine, Xinhua Hospital Affiliated to Shanghai Jiao Tong University, No. 1665 Kongjiang Road, Shanghai 200092, China

Correspondence should be addressed to Shiyong Huang; huangshiyong@xinhuamed.com.cn

Yongshan He and Yuanyuan Chen contributed equally to this work.

Received 5 April 2022; Accepted 6 May 2022; Published 16 July 2022

Academic Editor: Tao Huang

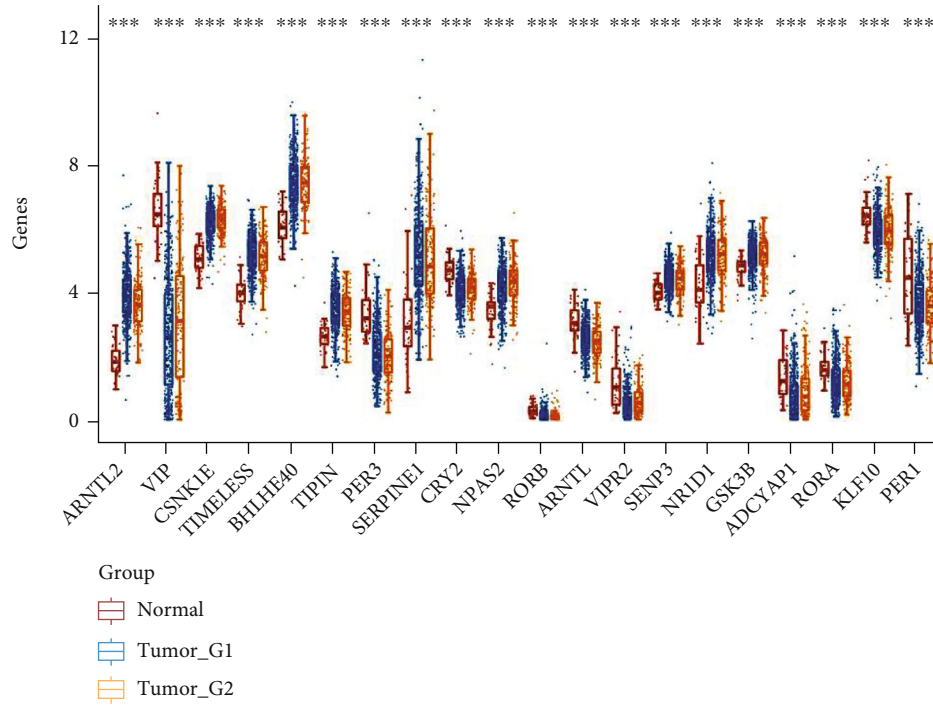
Copyright © 2022 Yongshan He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Early research shows that disrupting the circadian rhythm increases the risk of various cancers. However, the roles of circadian clock genes in colorectal cancer, which is becoming more common and lethal in China, remained to be unclear. In conclusion, the present study has demonstrated that multiple CCGs were dysregulated and frequently mutated in CRC samples by analyzing the TCGA database. The higher expression levels of REV1, ADCYAP1, CSNK1D, NR1D1, CSNK1E, and CRY2 had a strong link with shorter DFS time in CRC patients, demonstrating that CCGs had an important regulatory role in CRC development. Moreover, 513 CRC tumor samples were divided into 3 categories, namely, cluster1 ($n = 428$), cluster2 ($n = 83$), and cluster 3 ($n = 109$), based on the expression levels of the CCGs. Clinical significance analysis showed that the overall survival and disease-free survival of cluster 2 and cluster 3 were significantly shorter than those of cluster 1. The stemness scores in cluster 1 and cluster 2 were significantly higher than those of cluster 3 CRC samples. Clinically, we found that the C3 subtype had significantly higher percentage of T3/T4, N1/N2, and grades III and IV than groups C1 or C2. In addition, we reported that different CRC clusters had significantly different tumor-infiltrating immune cell signatures. Finally, pancancer analysis showed that higher expression of CSNK1D was correlated with shorter DFS time in multiple cancer types, such as COAD and LIHC, and was dysregulated in various cancers. In conclusion, we effectively developed a CCG-related predictive model and opened up new avenues for research into immune regulatory mechanisms and the development of immunotherapy for CRC.

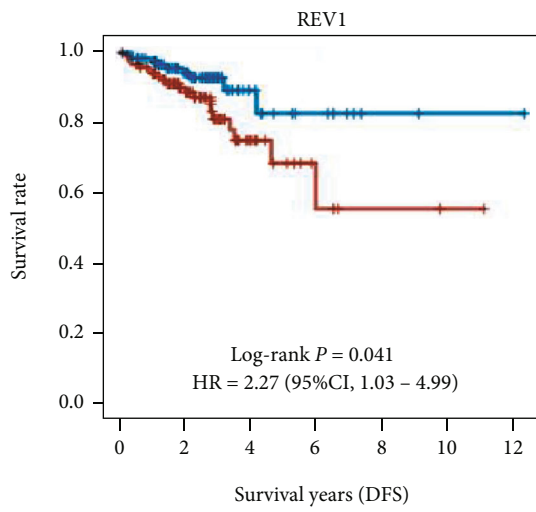
1. Introduction

Circadian rhythms are required for several biological activities, including metabolism, regeneration, immunology, and endocrinology [1]. Circadian rhythms regulate all human tissues through incredibly intricate mechanisms [2]. The molecular clock is comprised of a core clock gene loop. Circadian rhythm-controlled genes have key roles in tumor processes such as DNA damage and repair, apoptosis, cell proliferation, and metastasis [3]. A growing number of studies have sought to investigate the association between circadian rhythms and cancer in recent years [4, 5]. Early research shows that disrupting the circadian rhythm increases the risk of various cancers, including lung, pros-

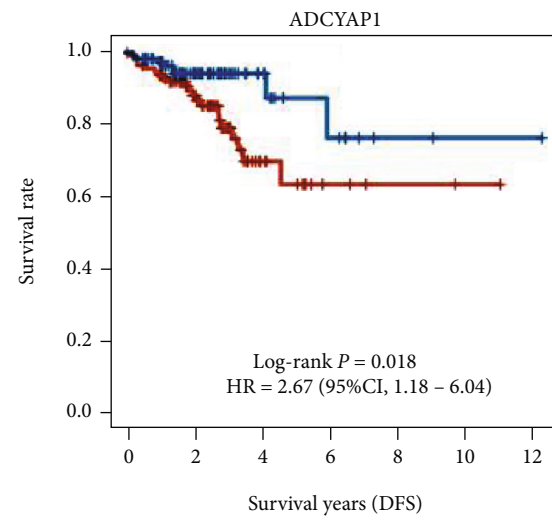
tate, breast, colon, endometrial, liver, pancreatic, and kidney cancers [6]. For instance, BMAL1 is vital in the prevention of breast cancer [7]. PER2 overexpression accelerates the growth of oral squamous cell carcinoma [8], and overexpression of NR1D1 may contribute to the development of kidney cell carcinoma [9]. In the pan renal cell carcinoma, circadian clock genes (CCGs) govern immunity, the cell cycle, and apoptosis [9]. Furthermore, persistent jet lag-induced gene deregulation and liver metabolic inefficiency can enhance hepatocarcinogenesis [10]. Nine circadian clock genes, including CCSNK1E, DBP, and NR1D2, were discovered as major prognostic markers in prostate cancer and were utilized to build a risk score model based on them [11]. Previous research has also demonstrated that disturbance of



(a)



(b)



(c)

FIGURE 1: Continued.

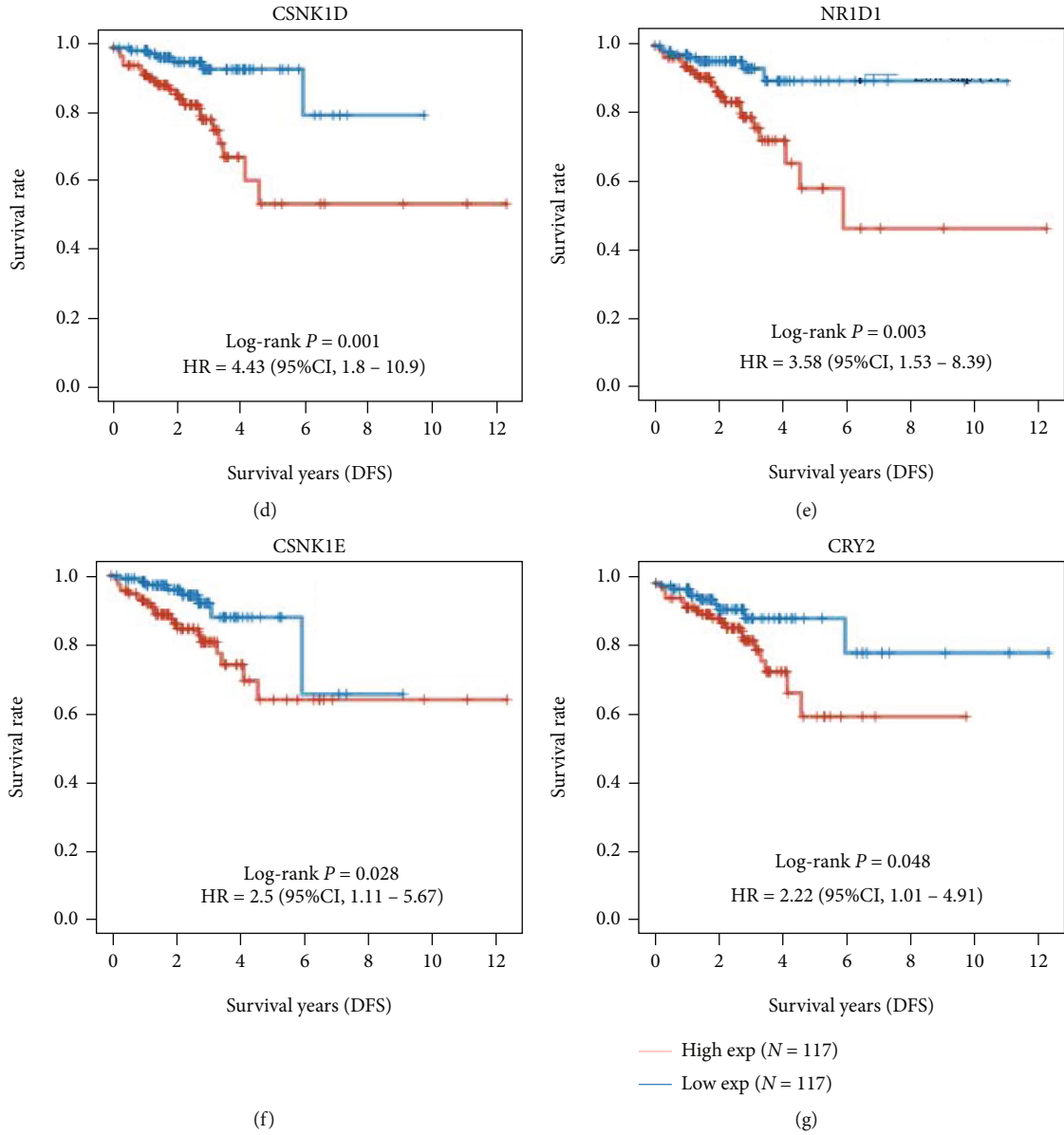
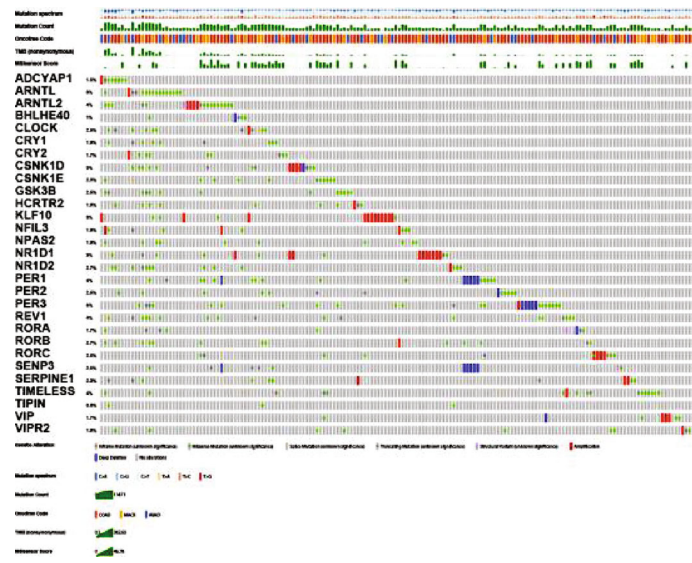
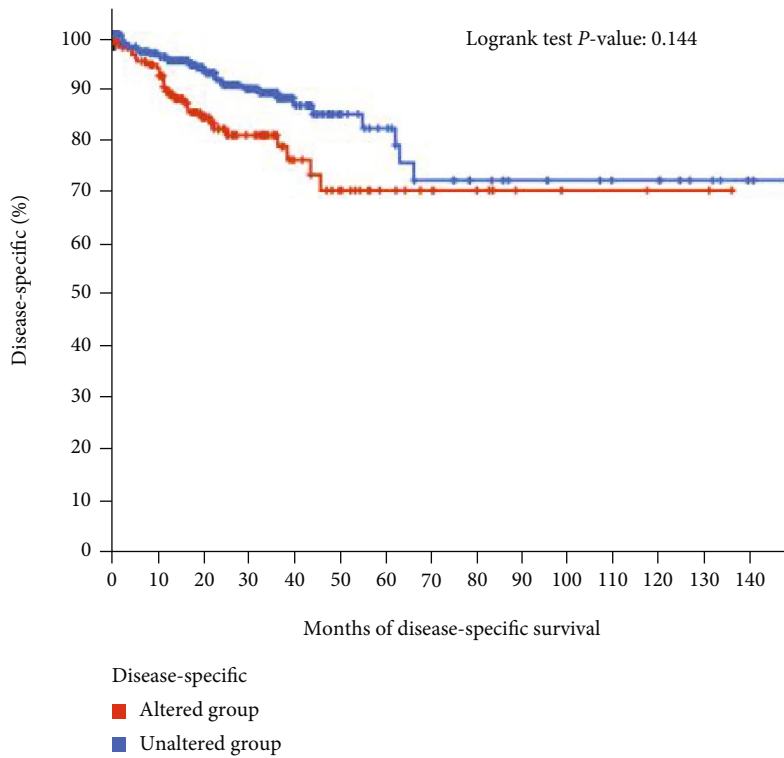


FIGURE 1: Continued.



(h)



(i)

FIGURE 1: The expression and genetic variation profile of CCGs in CRC. (a) The expression levels of CCGs in CRC and matched normal samples. (b) Correlation between disease-free survival time and expression of CCGs in CRC was present. (c) The horizontal histogram presented the genetic variation profile of CCGs in CRC. (d) CRC patients with mutations of CCGs had a shorter DFS time.

normal rhythms and malfunction of CCGs contribute to the formation and progression of various cancer types, as well as affect the tumor immune cycle function [12]. Previous research has indicated that cancer chronotherapy or scheduled chemotherapy delivery based on circadian rhythm may lessen drug toxicity [13]. Several animal model-based investigations have found that medicines targeting circadian

rhythm genes, such as ROR synthesis agonists, can increase anticancer immunity activation [14]. The significance of the circadian clock in prognostic evaluation and its clinical consequences in COAD, on the other hand, are rarely studied. According to one study, CLOCK, CRY1, and NR1D1 mRNA expression was raised in COAD tissue. A bioinformatic study revealed that circadian rhythm genes were mostly

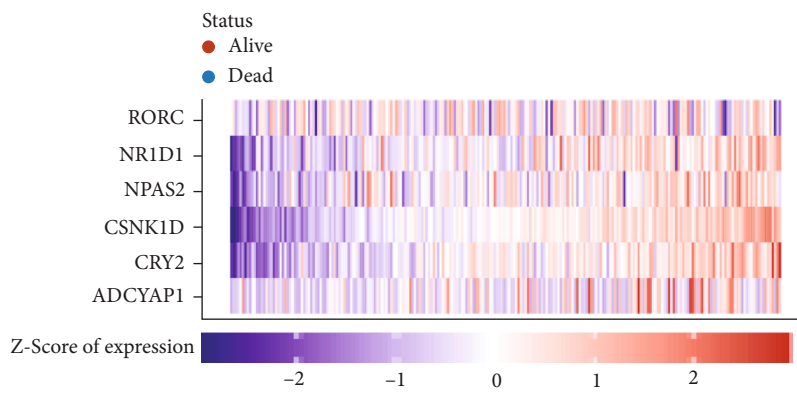
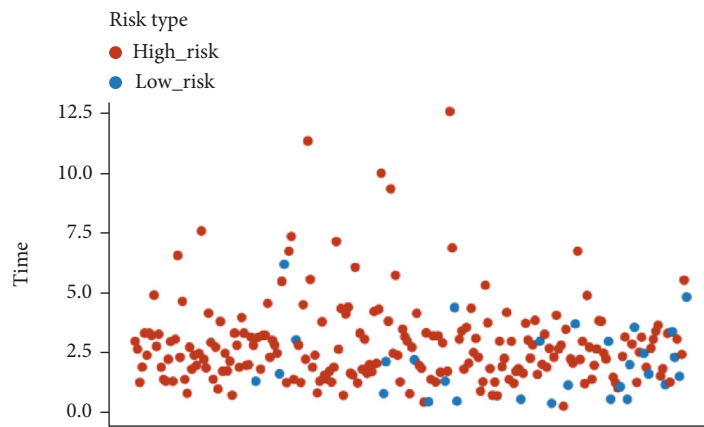
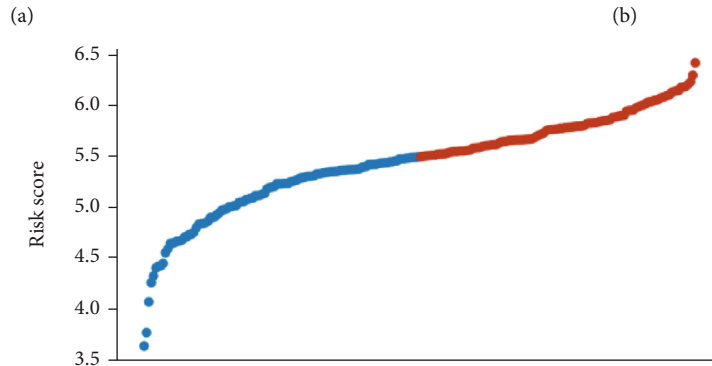
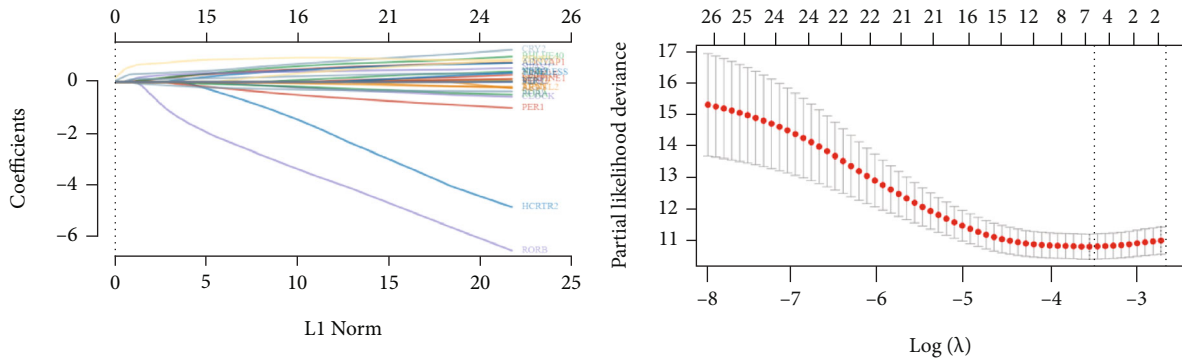


FIGURE 2: Continued.

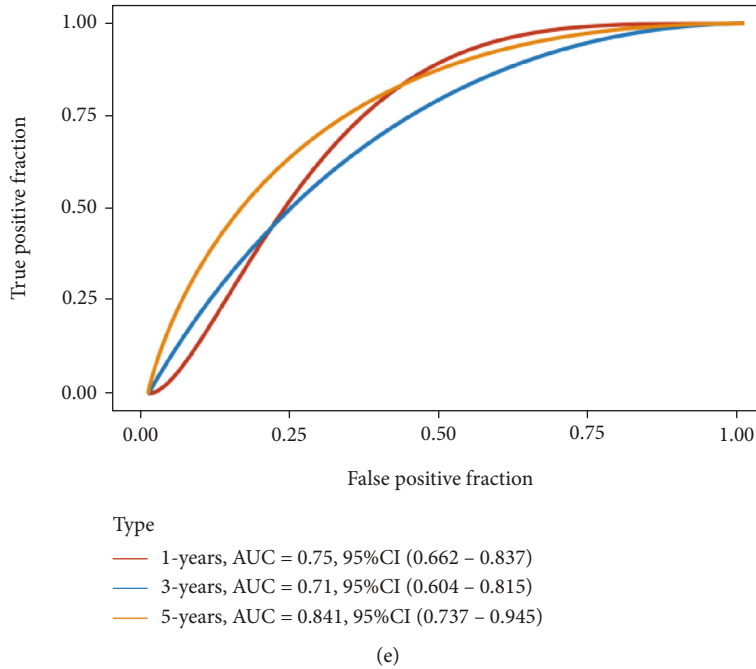
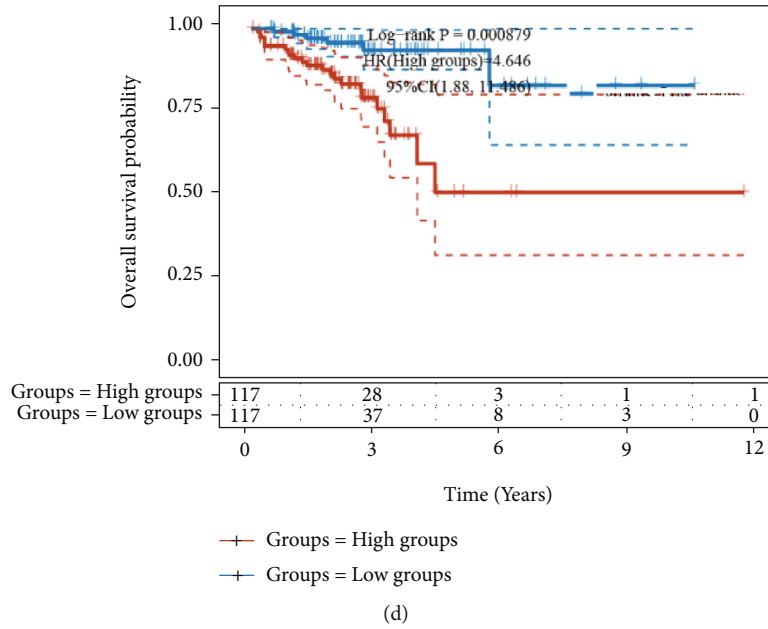


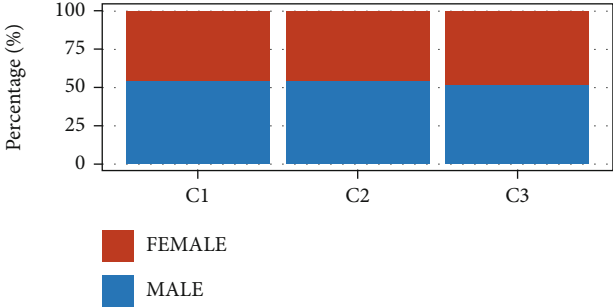
FIGURE 2: Construction and evaluation of the effectiveness of a prognostic signature. (a) LASSO coefficients for PRGs. Each curve represents a CCG. (b) 1000-fold crossvalidation of variable selection in LASSO regressions. (c) A significant survival benefit for low-risk CRC patients. (d, e) KM plotter analysis demonstrated that CRC patients with higher risk score had a shorter DFS time with an AUC value of 0.75, 0.71, and 0.841 for the 1-year, 3-year, and 5-year DFS, respectively.

associated with the glucocorticoid receptor pathway [15]. However, the functions of CCGs in the colorectal cancer remained largely to be unclear.

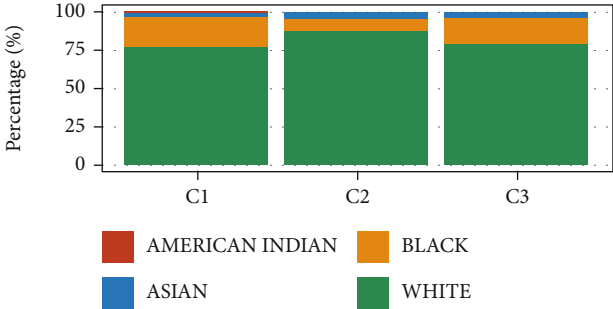
Colorectal cancer is becoming more common and lethal in China, with 0.37 million new cases and 0.19 million deaths, respectively [16, 17]. The incidence of colon cancer has considerably grown, and the majority of patients are detected in the middle or late stages [16, 17]. As a result, detecting CRC is a critical duty at the moment. Previous research has demonstrated that full loss of p53 expression

indicates a poor outcome in CRC [18]. TMED3 expression has been linked to a poor outcome in CRC [19]. Numerous research has been conducted to study the association between microsatellite instability (MSI) and the prognosis of CRC. The prognosis of CRC with MSI is much better than that of CRC with intact mismatch repair [20].

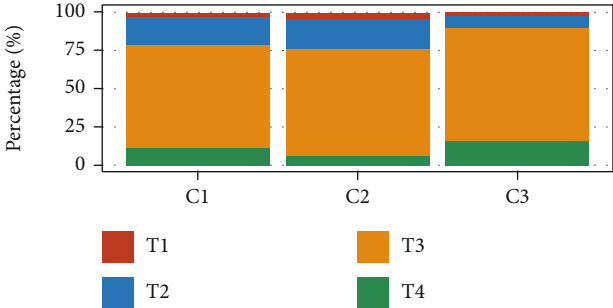
Immunotherapy has shown great progression in eliminating malignant cells by using the innate processes of the host immune system throughout the last decade, transforming the treatment landscape of many malignancies [21].



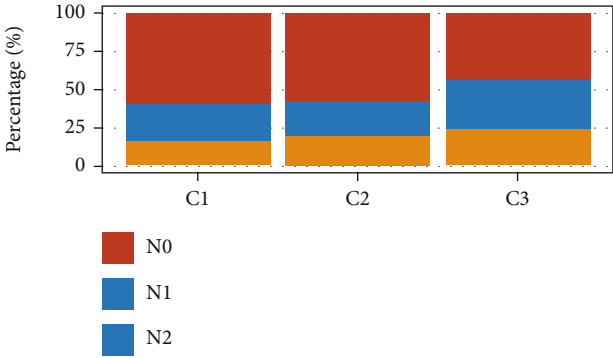
(a)



(b)

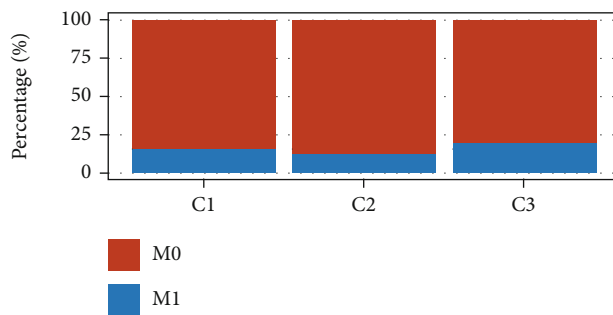


(c)

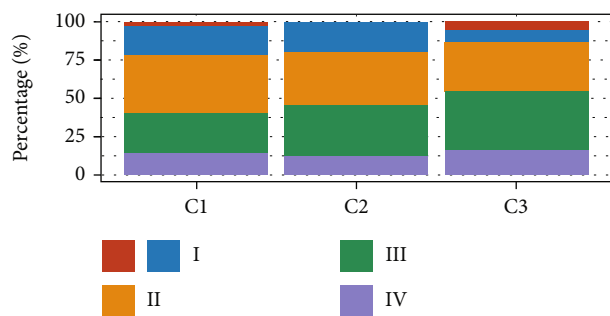


(d)

FIGURE 3: Continued.



(e)



(f)

FIGURE 3: Continued.

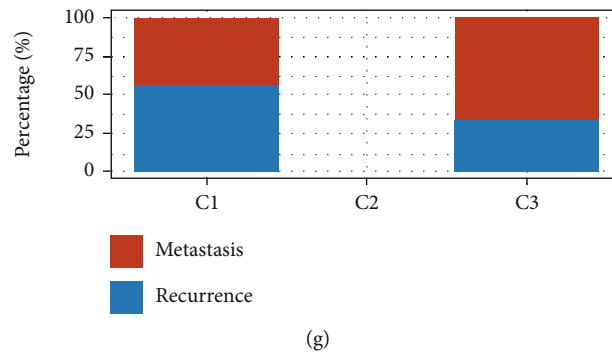


FIGURE 3: Clinical profile of the four subtypes. (a–g) The distribution of gender, race, T stage, N stage, M stage, grade, and metastasis was compared across patients with the three CCG subtypes to evaluate the clinical importance of CCG-related classification.

Immune checkpoint inhibition (ICI) has proven substantial effectiveness in cancer treatment techniques. ICI significantly increases overall survival (OS) time in patients with melanoma and lung cancer [21]. ICI stimulates the adaptive immune system to prevent immune escape caused by the activation of immune checkpoint cascades such as PD-1 or PD-L1 or CTLA-4 [22]. PD-1 is found on activated lymphocyte cells, which bind to PD-L1 expressed on tumor cells [22]. When activated, the PD-1/PD-L1 axis generates inhibitory signals, resulting in T cell depletion and inactivation [22]. The FDA-approved PD-1 inhibitors pembrolizumab for the treatment of dMMR/MSI-H CRC [23]. Recently, the FDA has also authorized ipilimumab, a CTLA-4 antibody, in conjunction with nivolumab in patients with CRC. Although PD-L1 is the most important prognostic biomarker for immunotherapy, PD-L1 expression was not related to better OS in the pembrolizumab phase II study in MSI-H CRC [24]. As a result, new biomarkers for predicting immunotherapy effectiveness are required.

2. Materials and Methods

2.1. Datasets and Preprocessing. The TCGA database was utilized to acquire CRC data. The TCGA database included 620 CRC tissues. RNA sequence data and clinicopathological characteristics were analyzed. R software was used for all subsequent statistical studies. $P < 0.05$ was used as the screening criterion for the differential expression of CCGs.

2.2. Characterization of Molecular Subtypes of CRC. The gene list of the core circadian clock genes was downloaded from The Molecular Signatures Database (MSigDB) [25, 26]. We aimed to see whether the expression profile of CCGs may help us differentiate between CRC subtypes. 29 CCGs were selected for further ConsensusClusterPlus analysis. The Euclidean distance metric was used to calculate the similarity distance between samples. The samples were clustered over 1000 iterations using the k-means clustering technique. The number of clusters ranged from 2 to 8, and the best partition was chosen by assessing the consensus cumulative distribution function (CDF). The PCA was performed using the R package for R v3.6.0.

2.3. Immune Signature Analysis in CRC Molecular Subtypes. The CIBERSORT method [27] was used to calculate the

expression scores of microenvironmental variables (tumor, immunological, and stromal purity). TIMER was used to examine the correlation between tumor samples and six tumor-infiltrating cells [28]. The immunological signature and checkpoint gene expression levels were also examined in all molecular subtypes. The analysis of variance (ANOVA) test was used to examine different CRC subtypes. Bonferroni correction was used for multiple testing.

2.4. Establishment of Prognostic Signature. CRC data from 620 patients from the TCGA database was retrieved for the identification of prognosis-related CCGs. We constructed a prediction signature for the CCGs stratified by the risk score using a univariate Cox proportional regression analysis and the least-absolute shrinkage and selection operator (Lasso) regression

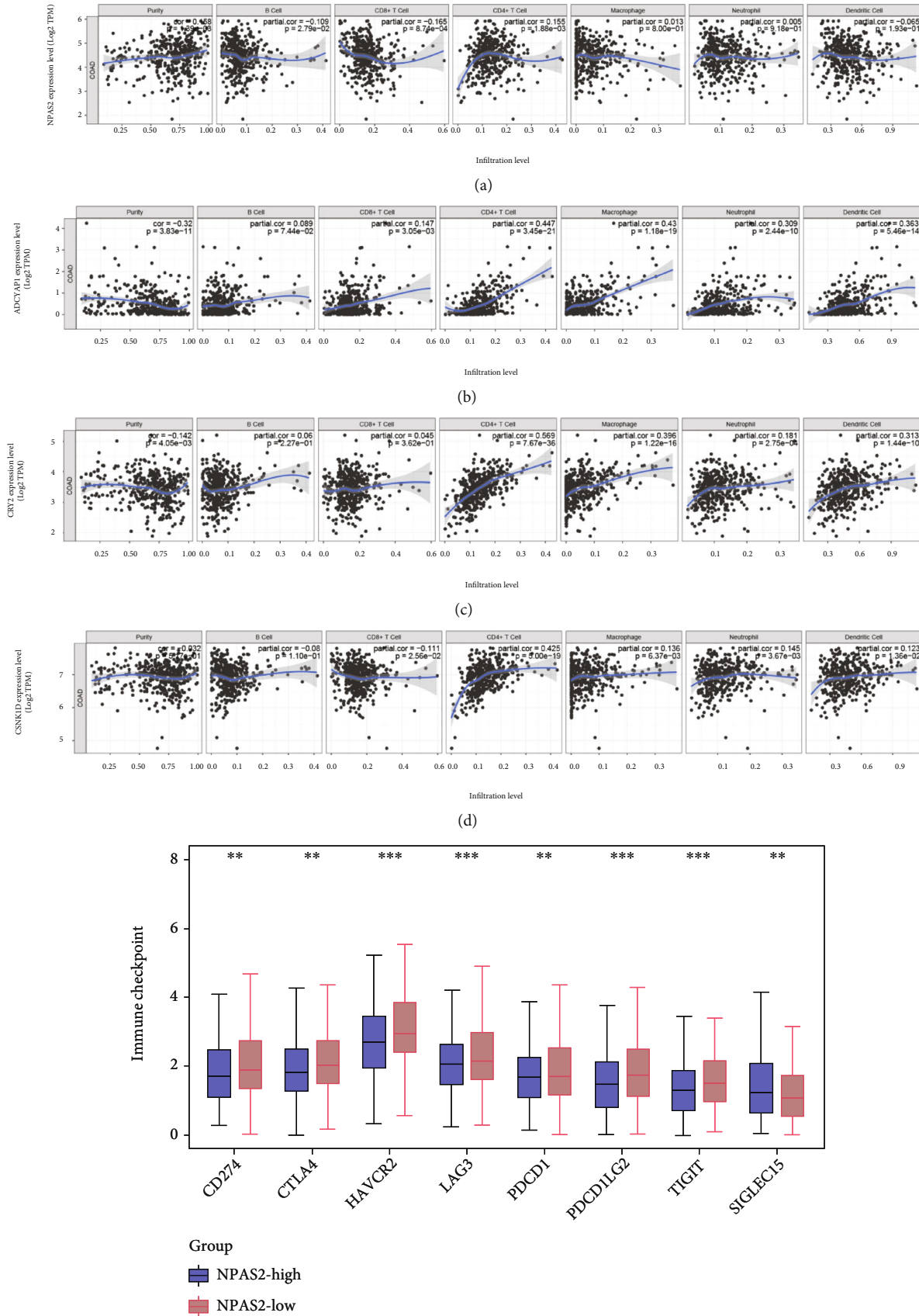
(- risk score = EXP CCGs n coefficient n + EXP CCGs 1 coefficient 1 + EXP CCGs 2 coefficient 2 + \dots + EXP CCGs n coefficient n). The risk score for each CRC patient was then computed. Based on the median risk score, the CRC patient was categorized as risk score-high or low group. The ROC curves were computed using the “survivalROC” program to evaluate the specificity and sensitivity of the prognosis model.

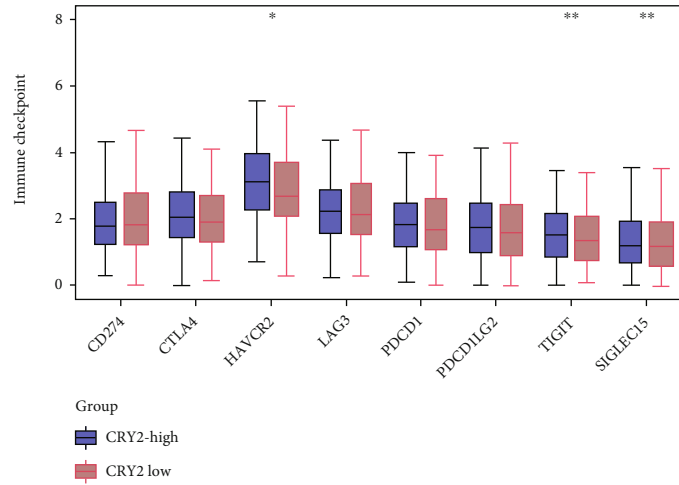
2.5. Analysis of mRNAsi Levels and MSI Levels. The mRNAsi was calculated using the OCLR machine-learning algorithm [29]. MSI analysis with MANTIS was performed as previously described [30].

2.6. Statistical Methods. To evaluate the connection between molecular subtypes and clinical factors, Fisher’s exact test or the chi-square test was utilized. These statistics were generated using the R software and were two sided.

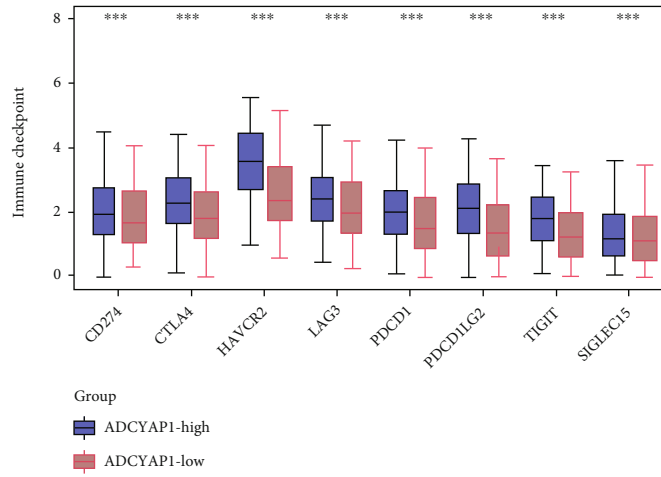
3. Results

3.1. The Expression and Genetic Variation Profile of CCGs in CRC. To begin with, we discovered that ARNTL2, CSNK1E, TIMELESS, BHLHE40, TIPIN, SERPINE1, NPAS2, SENP3, NR1D1, and GSK3B were highly upregulated, whereas VIP, PER3, CRY2, RORB, ARNTL, VIPR2, ADCYAP1, RORA, KLF10, and PER1 were significantly suppressed in both colon cancer and rectal cancer (Figure 1(a)). In order

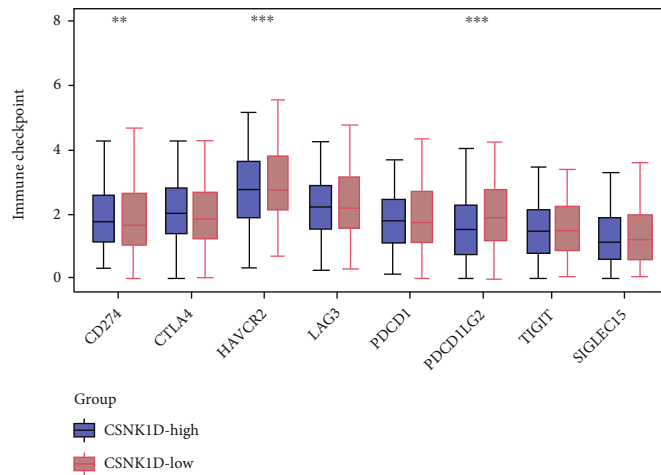




(f)



(g)



(h)

FIGURE 4: Four prognostic CCGs were correlated to immune cell infiltration in CRC. (a–d) The relationship between immune cell infiltration and expression of NPAS2, ADCYAP1, CRY2, and CSNK1D was analyzed by using TIMER database. (e–h) The expression levels SIGLEC15, CD274, HAVCR2, PDCD1LG2, LAG3, PDCD1, CTLA4, and TIGIT were analyzed between NPAS2, ADCYAP1, CRY2, and CSNK1D-high and low groups.

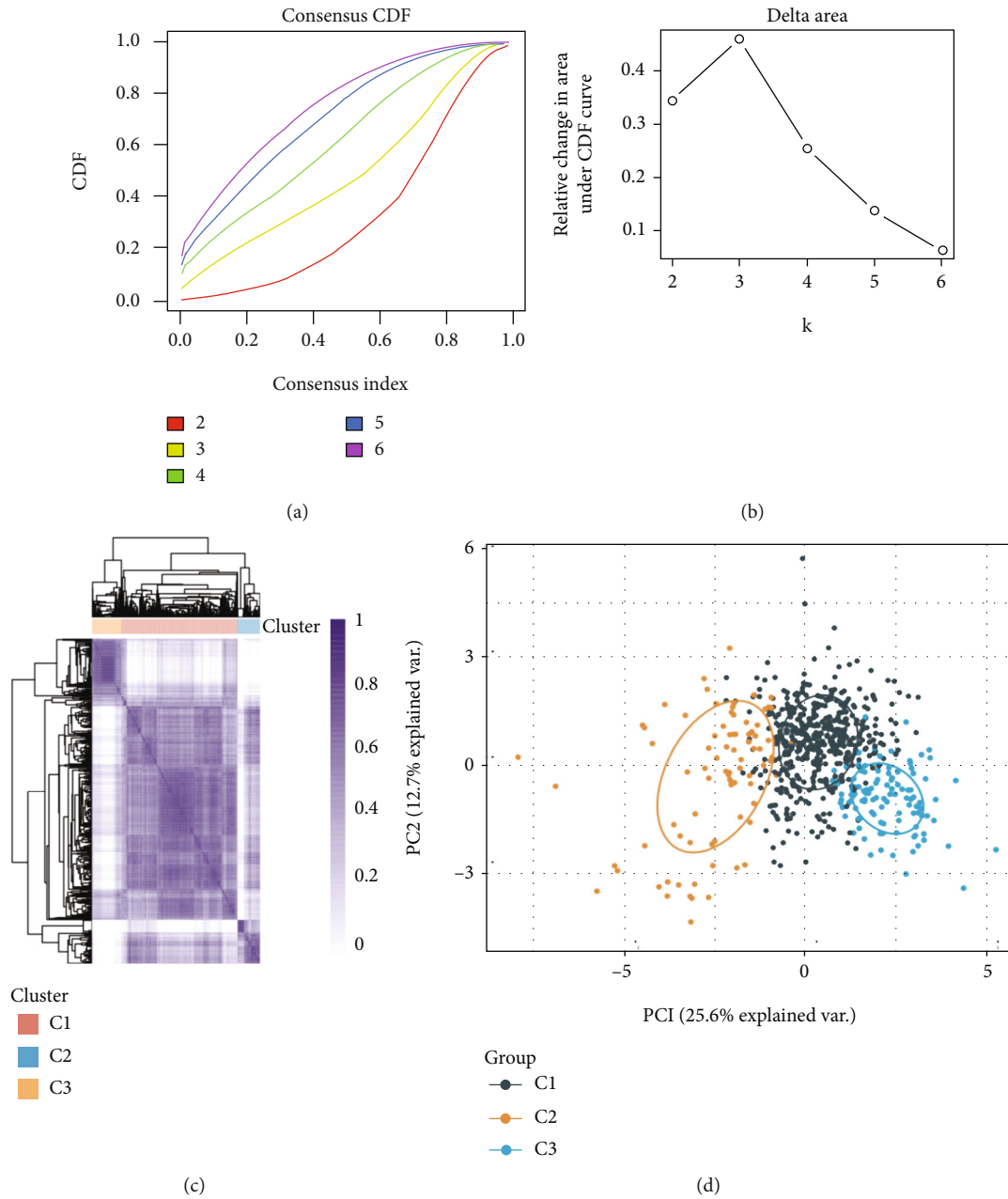


FIGURE 5: Continued.

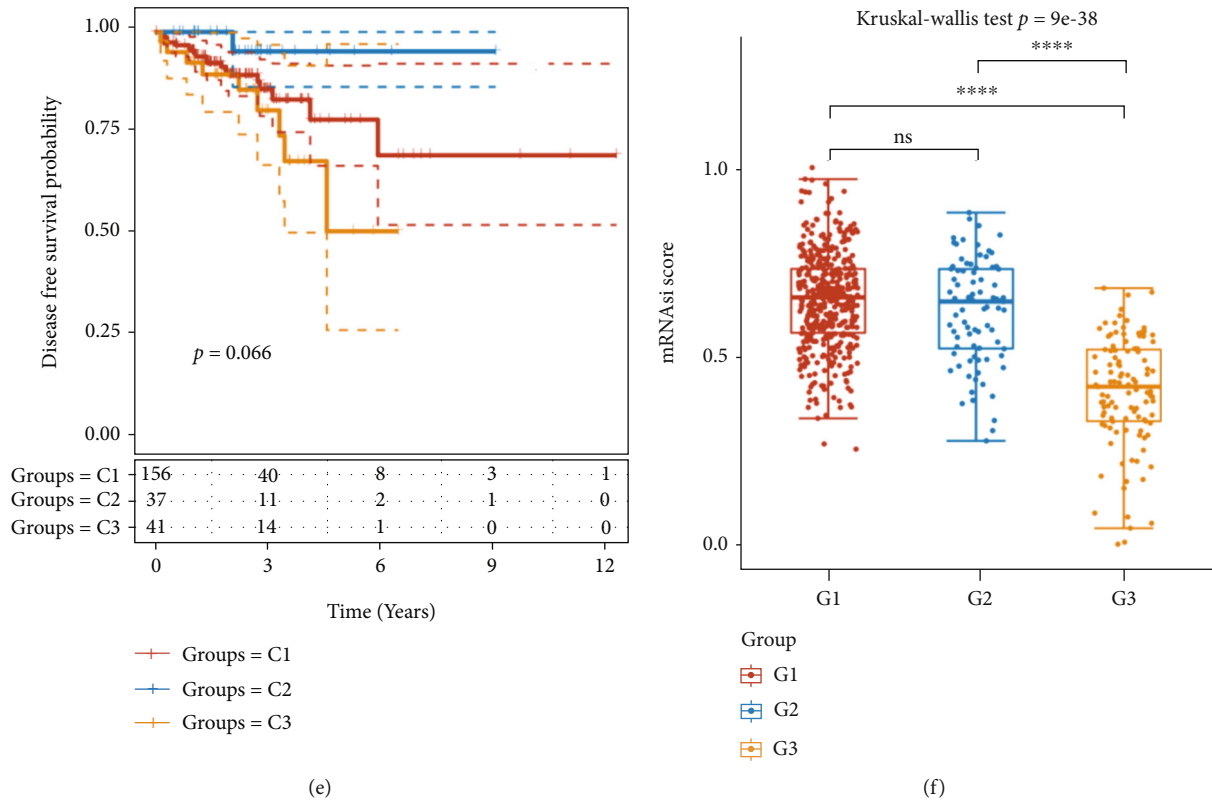


FIGURE 5: Four CRC subtypes were delineated based on the CCGs. (a, b) Consensus clustering cumulative distribution function (CDF) and relative change in the area under the CDF curve (CDF delta area) were analyzed. (c) Finally, 513 CRC tumor samples were divided into 3 categories, namely, cluster 1 ($n = 203$), cluster 2 ($n = 296$), and cluster 3 ($n = 296$), based on the expression levels of the CCGs. (d) PCA analysis of 3 cluster. (e) Clusters 2 and 3 had significantly shorter overall and disease-free survival time than cluster 1. (f) The results showed that the stemness scores in cluster 1 and cluster 2 were significantly higher than those in cluster3 CRC samples.

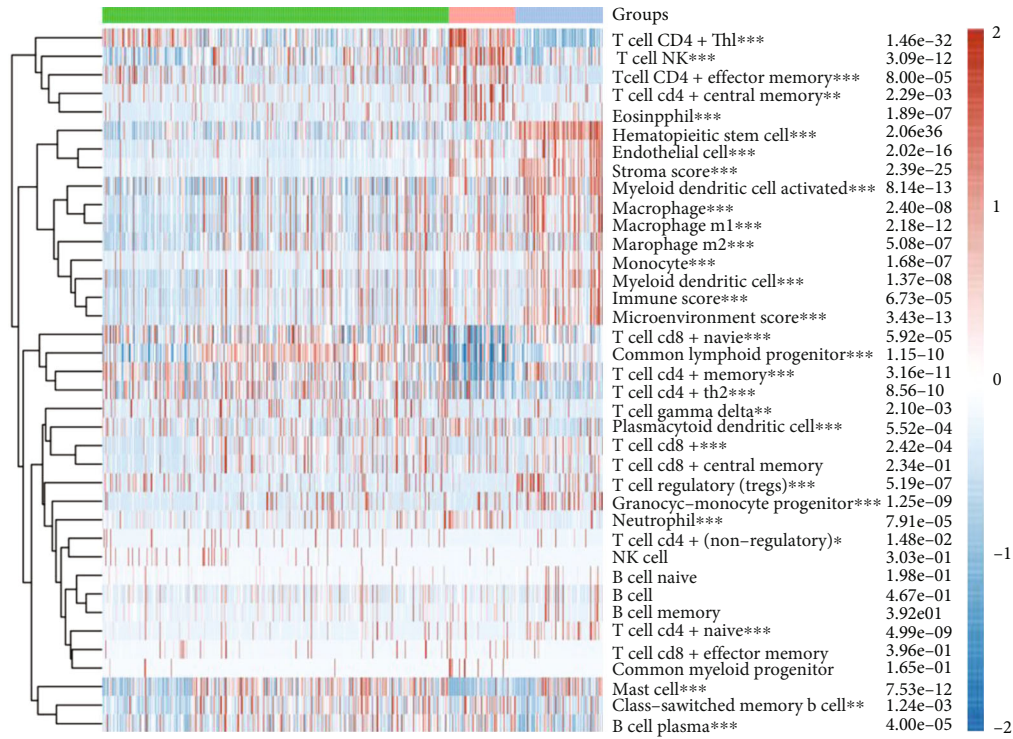
to further demonstrate the prognostic value of CCGs in CRC, we analyzed the correlation between disease-free survival time and the expression of CCGs in CRC. Higher levels of REV1, ADCYAP1, CSNK1D, NR1D1, CSNK1E, and CRY2 had a strong link with shorter DFS time in CRC patients (Figures 1(b)–1(g)). Mutation data was retrieved and shown using the “maftools” package in R software to find CCG somatic mutations in CRC. The mutation frequency of CCGs in CRC patients was shown in a horizontal histogram, such as PER3 (5%), ARNTL2 (4%), PER1 (4%), REV1 (4%), TIMELESS (4%), ARNTL (3%), CSNK1D (3%), KLF10 (3%), and NR1D1 (3%) (Figure 1(h)). Of note, we observed that CRC patients with mutations of CCGs had a shorter DFS time (Figure 1(i)), demonstrating that CCGs have an important regulatory role in CRC development.

3.2. Construction and Evaluation of the Effectiveness of Prognostic Signatures. The circadian clock gene-related signature was then created by further downscaling the CCGs using Lasso regression (Figures 2(a)–2(b)). The risk score was calculated for each CRC case, $\text{risk score} = (0.0184) * \text{ADCYAP1} + (0.2999) * \text{CRY2} + (0.5958) * \text{CSNK1D} + (0.0797) * \text{NPAS2} + (0.1034) * \text{NR1D1} + (-0.0141) * \text{RORC}$. Patients were divided into low- and high-risk groups based on the median risk score of all CRC cases, and the analysis revealed a significant survival benefit for low-risk CRC

patients (Figure 2(c)). KM plotter analysis demonstrated that CRC patients with a higher risk score had a shorter DFS time (Figure 2(d)) with an AUC value of 0.75, 0.71, and 0.841 for the 1-year, 3-year, and 5-year DFS, respectively (Figure 2(e)).

3.3. Four Prognostic CCGs Were Correlated to Immune Cell Infiltration in CRC. Figure 3(e) depicts the relationship between immune cell infiltration and expression of ADCYAP1, CRY2, NPAS2, and CSNK1D by using TIMER database. NPAS2 expression was considerably positively correlated with CD4+ T cell infiltration but dramatically negatively correlated to CD8+ T cell infiltration (Figure 4(a)). In CRC, ADCYAP1 expression was substantially associated with the levels of infiltration of CD4+ T cell, CD8+ T cell, macrophage, neutrophil, and dendritic cell (Figure 4(b)). CRY2 and CSNK1D expression was found to be substantially linked with levels of CD4+ T cell, macrophage, neutrophil, and dendritic cell infiltration in CRC (Figures 4(c) and 4(d)).

Moreover, we observed that SIGLEC15, CD274, HAVCR2, PDCD1LG2, LAG3, PDCD1, CTLA4, and TIGIT were higher in ADCYAP1 and NPAS2-high CRC samples than in ADCYAP1 and NPAS2-low CRC samples (Figures 4(e) and 4(g)). HAVCR2, TIGIT, and SIGLEC15 were higher in CRY2-high CRC samples than in CRY2-low



Groups
■ G1
■ G2
■ G3

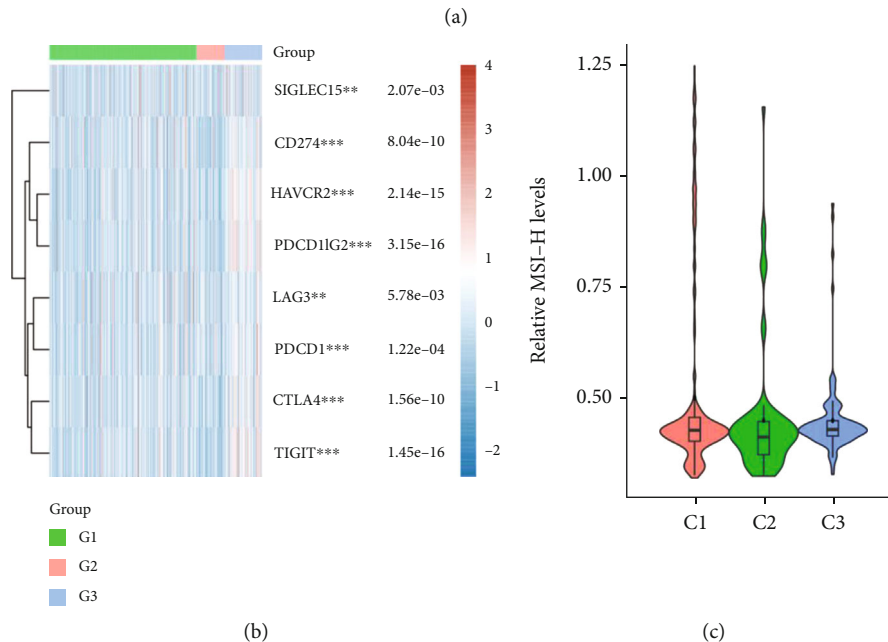
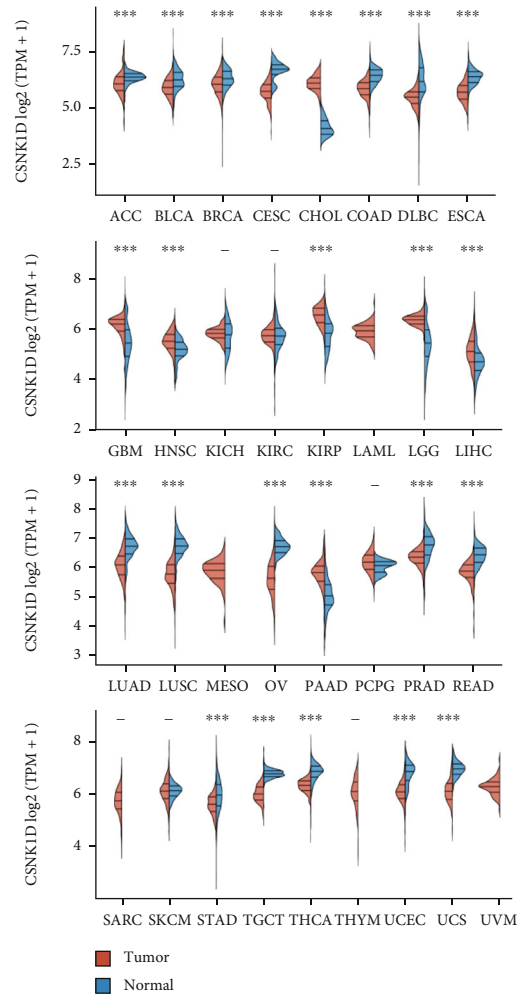
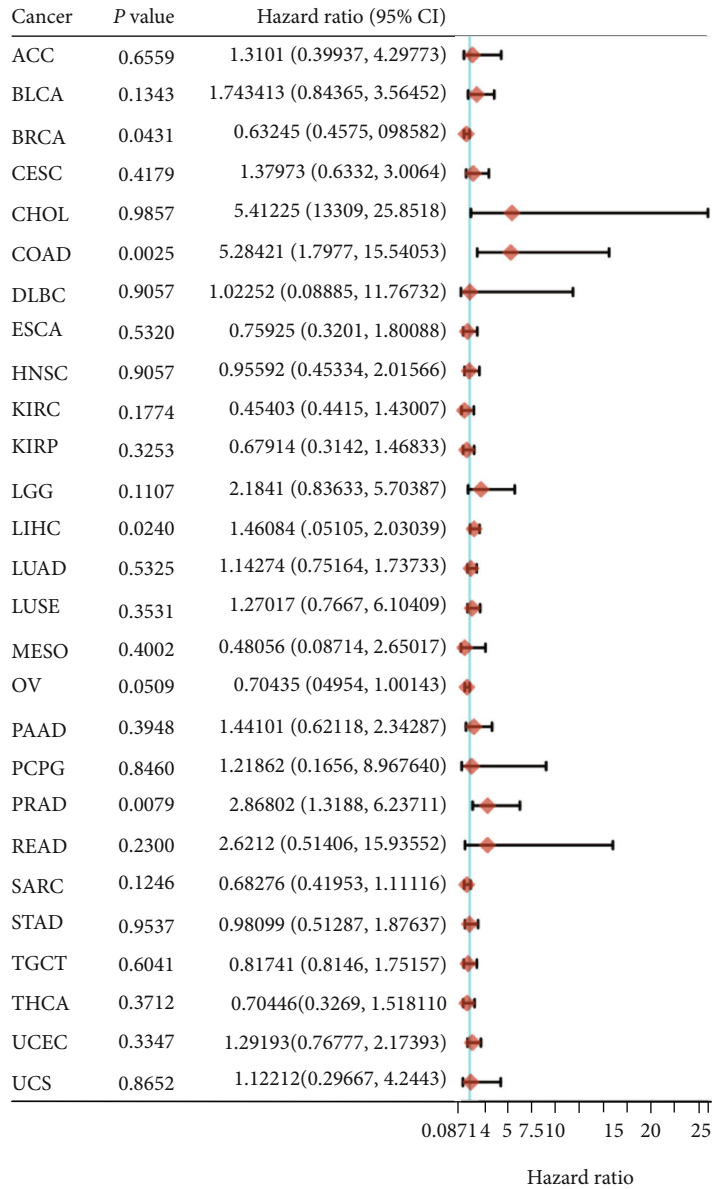


FIGURE 6: Distinct characteristics of immunogenicity of the CRC subtypes. (a) Tumor-infiltrating lymphocyte inflation among CRC subtypes was analyzed by using RNA expression data. (b) The expression levels of 8 immune checkpoint targets, which are critical for immunological control, were also evaluated, including SIGLEC15, CD274, HAVCR2, PDCD1LG2, LAG3, PDCD1, CTLA4, and TIGIT. (c) The MSI levels in CCG-related subtypes were analyzed.

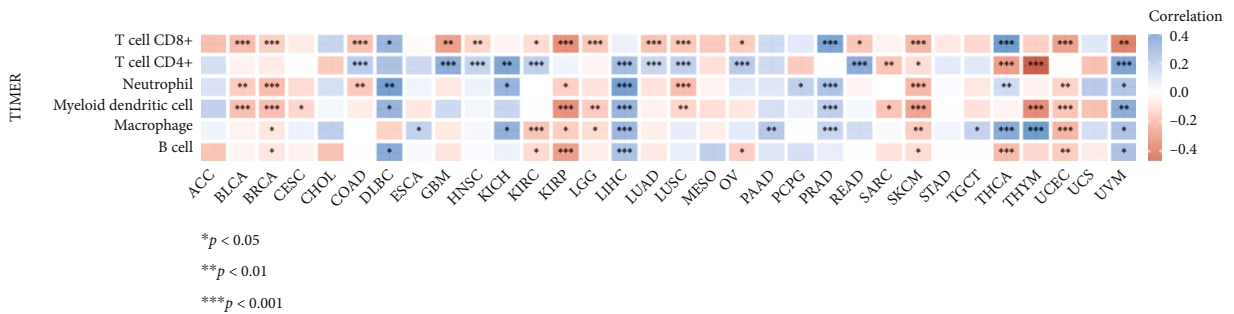
CRC samples (Figure 4(f)). However, CD274, HAVCR2, and PDCD1LG2 were lower in CSNK1D-high CRC samples than in CSNK1D-low CRC samples (Figure 4(h)).

3.4. Four CRC Subtypes Were Delineated Based on the CCGs. Next, 29 circadian clock gene expressions were used to categorize the TCGA CRC data into different CRC subtypes.



(a)

(b)



(c)

FIGURE 7: Pancancer analysis of CSNK1D. (a) Forest plot of CCGs in pan cancers. The graphic depicts the *P* value, hazard ratio, and associated 95 percent confidence interval for CCGs. (b) CSNK1D expression between tumor samples and normal samples were analyzed in pancancers. (c) Pancancer immune infiltration analysis showed that CSNK1D was correlated to multiple types of immune cell infiltration in various cancer types.

According to the CDF curves of the consensus score, a value of $k = 3$ was chosen to reflect stable clusters (Figures 5(a) and 5(b)). Finally, 620 CRC tumor samples were divided into 3 categories, namely, cluster1 ($n = 428$), cluster2 ($n = 83$), and cluster 3 ($n = 109$), based on the expression levels of the CCGs (Figures 5(c)–5(d)). Clusters 2 and 3 had significantly shorter overall and disease-free survival time than cluster 1 (Figures 5(e)). Cancer progression involves the progressive loss of a differentiated phenotype and the acquisition of progenitor/stem cell-like characteristics. Thus, we calculated the stemness of CRC samples in different subtypes using the logistic regression machine learning algorithm (OCLR) provided by Malta et al. based on the mRNA expression signature. The stemness score in cluster 1 and cluster 2 was considerably higher than that in the cluster 3 CRC samples (Figure 5(f)).

3.5. Clinical Profiles of the Four Subtypes. The distribution of gender, race, T stage, N stage, M stage, grade, and metastasis was compared across patients with the three CCG subtypes to evaluate the clinical importance of CCG-related classification. Clinically, we found that the C3 subtype had a considerably higher percentage of T3 and T4 than groups C1 or C2. In addition, we revealed that the C3 subtype had a considerably higher percentage of N1 and N2 samples than groups C1 or C2. However, the C1 subtype had a highest percentage of N0 samples than groups C2 or C3. Moreover, groups C3 had a higher percentage of grades III and IV and a lower proportion of grades I and II than cluster 1 and cluster 2. However, the distribution of gender, race, M stage, and metastasis does not have a significant difference among these clusters (Figures 3(a)–3(g)).

3.6. Distinct Characteristics of Immunogenicity of the CRC Subtypes. We next investigated microenvironmental variables and tumor-infiltrating lymphocyte inflation among CRC subtypes using RNA expression data. We observed that T cell CD4+ Th1, T cell NK, T cell CD4+ effector memory, T cell CD4+ central memory, and eosinophil cells were considerably enriched in cluster 2 CRC samples. Hematopoietic stem cell, endothelial cell, myeloid dendritic cell activated, M1 and M2 macrophages, monocyte, and myeloid dendritic cells were considerably enriched in cluster 3 CRC cases compared to those of clusters 1 and 2. Meanwhile, we observed that naive CD8+ T cell, common lymphoid progenitor cells, CD4+ memory T cells, CD4+ Th2 T cells, and T cell gamma delta were substantially more abundant in cluster 1 CRC samples than those of clusters 2 and 3 (Figure 6(a)).

Moreover, the expression levels of 8 immune checkpoint targets, which are critical for immunological control, were also evaluated, including SIGLEC15, CD274, HAVCR2, PDCD1LG2, LAG3, PDCD1, CTLA4, and TIGIT. As presented in Figure 7, we observed that these immune checkpoint markers in cluster 3 CRC samples were substantially higher than those in cluster 1 and 2 CRC samples (Figure 6(b)). Immune therapy was only approved for MSI-H/dMMR CRC. Thus, the destitution of MSI levels in 3 CCG subtypes was also evaluated. However, we do not observe a significant difference among 3 CCG subtypes.

3.7. Pancancer Analysis of CSNK1D. The present study has demonstrated the significant correlation between immune inflation and prognosis and CSNK1D in CRC. Next, we evaluated the clinical importance of CSNK1D across human cancers using the TCGA database. As presented by the forest plot, we observed that higher levels of CSNK1D were correlated to shorter DFS time in CHOL (cholangiocarcinoma), COAD (colon cancer), LIHC, and PRAD (prostate cancer); however, higher expression of CSNK1D was correlated to longer DFS time in BRCA and OV (ovarian cancer) (Figure 7(a)). By analyzing CSNK1D expression between tumor samples and normal samples, we revealed that CSNK1D was downregulated in ACC, BLCA, BRCA, CESC, COAD, DLBC, ESCA, LUAD, LUSC, PRAD, READ, STAD, TGCT, THCA, UCEC, and UCS but was increased in CHOL, GBM, HNSC, KIRP, LGG, LIHC, PAAD, and PCPG (Figure 7(b)).

Pancancer immune inflation analysis showed that CSNK1D was correlated to multiple types of immune cell inflation in various cancer types, such as LGG, COAD, KIRP, LIHC, PRAD, THCA, BLCA, BRCA, THYM, and UCEC (Figure 7(c)).

4. Discussion

The circadian system's significance in carcinogenesis is well recognized, and numerous studies have found differential clock gene expression in cancers compared to healthy tissues. By analyzing the TCGA database, the present study demonstrated that multiple CCGs were dysregulated and frequently mutated in CRC samples. The higher expression levels of REV1, ADCYAP1, CSNK1D, NR1D1, CSNK1E, and CRY2 had a strong link with shorter DFS time in CRC patients, demonstrating that CCGs had an important regulatory role in CRC development. ADCYAP1 is known to modulate the immune system and is involved in cell proliferation and apoptosis in normal cells [31]. ADCYAP1 upregulation or downregulation has been identified in a variety of malignancies [31]. A significant percentage of ADCYAP1 hypermethylation is frequently reported in ovarian cancer [32]. ADCYAP1 promoter hypermethylation levels have been linked to cervical cancer development [31]. The low rate of ADCYAP1 hypermethylation in early-stage lesions and its increase with the tumor stage suggest that ADCYAP1 hypermethylation may hinder ADCYAP1's apoptotic action. Cry1 expression was increased in right colon cancers but not in left colorectum tumors [33]. CRY2NK1 upregulation is associated with a worse outcome in patients with hepatocellular cancer. NPAS2 is a circadian gene that has attracted the interest of researchers due to its various effects on cells and various roles in disease development, especially cancer. Differential NPAS2 expression has been associated with patient outcomes in tumors, lung tumors, non-Hodgkin's lymphoma, and other diseases [34], and nucleotide variants in the NPAS2 gene were related to cancer patients' outcomes. In endometrial cancer of the uterus, the circadian gene NPAS2 operates as a potential tumor stimulator. NPAS2 increases liver fibrosis in hepatocellular carcinoma by direct transcriptional activation of Hes1 in hepatic stellate cells and

induces glucose metabolism reprogramming and cell survival by transactivating CDC25A in liver cancer cells. Here, we constructed a prognostic signature based on CCG expression in CRC. Risk score = $(0.0184) * ADCYAP1 + (0.2999) * CRY2 + (0.5958) * CSNK1D + (0.0797) * NPAS2 + (0.1034) * NR1D1 + (-0.0141) * RORC$. KM plotter analysis demonstrated that CRC patients with a higher risk score had a shorter DFS time.

CRC is the leading cause of cancer mortality worldwide. CRC, similar with many other cancers, is a heterogeneous disease, making it a clinical challenge to enhance treatment efficacy to reduce morbidity and mortality. CRC can be caused by a number of pathogenic mechanisms, such as DNA mismatch repair failure (MMR), and epigenetic changes. A greater understanding of biology, particularly the clinical features that distinguish CRC patients, is required for more robust targeted therapy development and the application of personalized therapy. A rising corpus of research has addressed the topic of molecular categorization of CRC in recent decades in order to supplement the current staging systems, enhance therapy options, and better predict survival following treatment. The application of high-throughput sequencing has heralded a new age in subtyping research; for example, the CRC Subtyping Consortium reported the finding of four consensus molecular subtypes (CMS) in 2015, giving the most rigorous CRC classification system to date [35]. The colorectal cancer molecular classification consists mostly of mutation-centered CRC classification and transcriptome-centered CRC classification [35]. According to recent studies, intratumoral heterogeneity is best described at the transcriptome level, as it provides more comprehensive genomic information about the disorder process. The classification of CRCs based on transcriptomes has switched from supervised to unsupervised. Transcriptome-based techniques, as opposed to histopathological subtypes and traditional mutation-oriented stratification, use genome-wide expression profiling for unsupervised data processing [35]. This approach has been utilized successfully in a wide range of malignancies, including lung cancer, prostate cancer, and gastric cancers. Recent efforts have been made, like with other malignancies, to subtype CRC. Wang and colleagues, for example, used fresh-frozen tissue samples from a large multicenter cohort (CIT cohort) of 566 CRC patients to perform gene expression analysis, identifying six CRC subgroups with distinct molecular signatures and clinical correlations [35]. In this study, 513 CRC tumor samples were divided into 3 categories, namely, cluster 1 ($n = 203$), cluster 2 ($n = 296$), and cluster 3 ($n = 296$), based on the expression levels of the CCGs (Figure 2(a)). Clinical significance analysis showed that clusters 2 and 3 had significantly shorter overall and disease-free survival time than cluster 1. The stemness scores in cluster 1 and cluster 2 were significantly higher than those of cluster 3 CRC samples. Clinically, we found that the C3 subtype had a higher percentage of T3/T4, N1/N2, and grades III and IV than groups C1 or C2.

A growing body of evidence supports the importance of immune infiltration in cancer, which comprises lymphocytes, dendritic cells, and macrophages, demonstrating a

wide range of patient-patient variability [36]. TILs reflect the host immune response to tumor cells, which is related to CRC patients' prognosis [37]. CD8+ cytotoxic T cells can directly destroy tumor cells. CTL is activated by type 1 helper T lymphocytes (Th1s), whereas humoral immunity is stimulated by type 2 helper T lymphocytes (Th2s). Immune responses and TILs have been studied in CRC as a strategy of classifying tumors and as prognostic indicators. TILs situated near the tumor boundary are activated and assault the tumor when PD-1 is suppressed. Most tumor-infiltrating lymphocyte types have recently been investigated, and it appears that CD8+ T cells have the greatest impact on patient outcome [38]. More than a decade ago, researchers looked into the prognostic relevance of CD8+ CTLs in a large CRC cohort. According to several studies, increased CTL levels in the tumor microenvironment have been linked to antitumor effects and improved prognosis in a variety of malignancies, including CRC. The dMMR group had more CD56+ cells, CD4+ cells, and higher CD8 protein levels than the pMMR group, according to Bai and colleagues [39]. In the present study, we observed that T cell CD4+ Th1, T cell NK, T cell CD4+ effector memory, T cell CD4+ central memory, and eosinophil cells were enriched in cluster 2 CRC samples. Hematopoietic stem cell, endothelial cell, myeloid dendritic cell activated, M1 and M2 macrophages, monocyte, and myeloid dendritic cells were considerably enriched in cluster 3 CRC cases compared to clusters 1 and 2. Meanwhile, we observed that CD8+ naive T cell, common lymphoid progenitor, and CD4+ memory/Th2 T cells were substantially more abundant in cluster 1 CRC samples than in clusters 2 and 3. Immune therapy was only approved for MSI-H/dMMR CRC. Thus, the distribution of MSI levels in 3 CCG subtypes were also evaluated. However, we do not observe a significant difference among 3 CCG subtypes. These results indicate that microsatellite-stable (mss) CRC patients may also benefit from immune therapy by using CCG-related biomarkers.

Here, we for the first time demonstrated the significant correlation between immune inflation and prognosis and CSNK1D in CRC. CK1 δ is a member of the CK1 family. The role of CK1 has been increasingly described during the last few decades, both physiologically and pathologically. Indeed, dysregulated CK1 expression and activity have been identified in several cancers as well as neurological illnesses such as Alzheimer's disease. CSNK1D has been identified as a possible gene driver in cutaneous squamous cell cancer [40]. In patients with superficial and invasive bladder cancer, the CSNK1D gene is elevated. In bladder cancer cells, CK1 knockdown reduced catenin expression and hindered cell proliferation. Next, we evaluated the clinical importance of CSNK1D across human cancers using the TCGA database. Pancancer analysis indicates that the expression of CSNK1D was correlated to shorter DFS time in multiple cancer types, such as COAD and LIHC, and was dysregulated in various cancers, which was consistent previous reports. For example, Liu et al. reported that CSNK1D levels are strongly upregulated in HCC samples [41]. Upregulation of CSNK1D is associated with a poor prognosis for HCC patients. CSNK1D expression was higher in hepatocellular carcinoma

(HCC) with distant metastasis than in HCC without metastasis [42]. According to the GSEA enrichment analysis, CSNK1D influences the HCC prognosis mostly through the cell cycle and the WNT pathway. In prostate cancer, 9 CCGs, including CSNK1D, were identified as key prognostic genes [11]. These results showed that CSNK1D may serve as an important cancer regulator.

Several limitations should also be noted in this study. First, only the TCGA database was used for the analysis of CCGs. More datasets should be analyzed and used to confirm the TCGA results. Secondly, the clinical samples would also be collected for further validation of the correlation between tumor-infiltrating lymphocytes and CCGs in the future study. Thirdly, the molecular functions of CCGs in CRC remained to be confirmed with experimental assays.

In conclusion, the present study has demonstrated that multiple CCGs were dysregulated and frequently mutated in CRC samples by analyzing the TCGA database. Higher expression levels of REV1, ADCYAP1, CSNK1D, NR1D1, CSNK1E, and CRY2 had a strong link with shorter DFS time in CRC patients, demonstrating that CCGs had an important regulatory role in CRC development. Moreover, 513 CRC tumor samples were divided into 3 categories, based on the expression levels of the CCGs. Clinical significance analysis showed that clusters 2 and 3 had significantly shorter overall and disease-free survival time than cluster 1. Stemness scores in cluster 1 and cluster 2 were significantly higher than that of cluster 3 CRC samples. Clinically, we found that the C3 subtype had a significantly higher percentage of T3/T4, N1/N2, and grades III and IV than groups C1 or C2. In addition, we reported that different CRC clusters had a significantly different tumor-infiltrating immune cell signature. Finally, pancancer analysis showed that higher expression of CSNK1D was correlated with shorter DFS time in multiple cancer types, such as COAD and LIHC, and was dysregulated in various cancers. In conclusion, we effectively developed a CCG-related predictive model and opened new avenues for research into immune regulatory mechanisms and the development of immunotherapy for CRC.

Data Availability

All the data can be acquired by reasonable request.

Conflicts of Interest

There is no conflict of interest.

Authors' Contributions

Yongshan He and Yuanyuan Chen contributed equally to this work.

References

- [1] P. Dierickx, B. du Pré, D. A. M. Feyen et al., "Circadian rhythms in stem cell biology and function," in *Stem Cells and Cardiac Regeneration*, Springer, 2016.
- [2] J. Richards and M. L. Gumz, "Mechanism of the circadian clock in physiology," *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, vol. 304, no. 12, pp. R1053–R1064, 2013.
- [3] H. X. Li, "The role of circadian clock genes in tumors," *Oncotargets and Therapy*, vol. Volume 12, pp. 3645–3660, 2019.
- [4] E. Cash, S. Sephton, C. Woolley et al., "The role of the circadian clock in cancer hallmark acquisition and immune-based cancer therapeutics," *Journal of Experimental & Clinical Cancer Research*, vol. 40, no. 1, p. 119, 2021.
- [5] F. Battaglin, P. Chan, Y. Pan et al., "Clocking cancer: the circadian clock as a target in cancer therapy," *Oncogene*, vol. 40, no. 18, pp. 3187–3200, 2021.
- [6] S. Lee, L. A. Donehower, A. J. Herron, D. D. Moore, and L. Fu, "Disrupting circadian homeostasis of sympathetic signaling promotes tumor development in mice," *PLoS One*, vol. 5, no. 6, article e10995, 2010.
- [7] W. E. Samsa, A. Vasani, R. J. Midura, and R. V. Kondratov, "Deficiency of circadian clock protein BMAL1 in mice results in a low bone mass phenotype," *Bone*, vol. 84, pp. 194–203, 2016.
- [8] H. Liu, X. Gong, and K. Yang, "Overexpression of the clock gene Per2 suppresses oral squamous cell carcinoma progression by activating autophagy via the PI3K/AKT/mTOR pathway," *Journal of Cancer*, vol. 11, no. 12, pp. 3655–3666, 2020.
- [9] S. Liu, Y. Cheng, S. Wang, and H. Liu, "Circadian clock genes modulate immune, cell cycle and apoptosis in the diagnosis and prognosis of pan-renal cell carcinoma," *Frontiers in Molecular Biosciences*, vol. 8, article 747629, 2021.
- [10] N. Kettner, H. Voicu, M. J. Finegold et al., "Circadian homeostasis of liver metabolism suppresses hepatocarcinogenesis," *Cancer Cell*, vol. 30, no. 6, pp. 909–924, 2016.
- [11] W. Yue, X. du, X. Wang et al., "Prognostic values of the core components of the mammalian circadian clock in prostate cancer," *PeerJ*, vol. 9, article e12539, 2021.
- [12] C. Savvidis and M. Koutsilieris, "Circadian rhythm disruption in cancer biology," *Molecular Medicine*, vol. 18, no. 9, pp. 1249–1260, 2012.
- [13] T. T. Kuo and A. G. Ladurner, "Exploiting the circadian clock for improved cancer therapy: perspective from a cell Biologist," *Biologist*, vol. 10, p. 1210, 2019.
- [14] G. Sulli, M. T. Y. Lam, and S. Panda, "Interplay between circadian clock and cancer: new frontiers for cancer treatment," *Trends in Cancer*, vol. 5, no. 8, pp. 475–494, 2019.
- [15] A. He, Z. Huang, R. Zhang et al., "Circadian clock genes are correlated with prognosis and immune cell infiltration in colon adenocarcinoma," *Computational and Mathematical Methods in Medicine*, vol. 2022, 1709924 pages, 2022.
- [16] P. Rawla, T. Sunkara, and A. Barsouk, "Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors," *Gastroenterology Review*, vol. 14, no. 2, pp. 89–103, 2019.
- [17] F. A. Hagggar and R. P. Boushey, "Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors," *Clinics in Colon and Rectal Surgery*, vol. 22, no. 4, pp. 191–197, 2009.
- [18] K. Nagao, A. Koshino, A. Sugimura-Nagata et al., "The complete loss of p53 expression uniquely predicts worse prognosis in colorectal cancer," *International Journal of Molecular Sciences*, vol. 23, no. 6, p. 3252, 2022.
- [19] R. F. Wang, Y. G. Hong, L. Q. Hao, and H. T. Yu, "Expression of TMED3 is independently associated with colorectal cancer

- prognosis,” *Experimental and Therapeutic Medicine*, vol. 23, no. 4, p. 286, 2022.
- [20] N. R. Parine, R. S. Varsha, and M. S. Alanazi, “Microsatellite instability in colorectal cancer,” in *Microsatellite Markers*, 2016.
- [21] K. Esfahani, L. Roudaia, N. Buhlaiga, S. V. del Rincon, N. Papneja, and W. H. Miller, “A review of cancer immunotherapy: from the past, to the present, to the future,” *Current Oncology*, vol. 27, no. 12, pp. 87–97, 2020.
- [22] P. Zheng and Z. Zhou, “Human cancer immunotherapy with PD-1/PD-L1 blockade,” *Biomarkers in Cancer*, vol. 7s2, Suppl 2, pp. BIC.S29325–BIC.S29318, 2015.
- [23] A. Ooki, E. Shinozaki, and K. Yamaguchi, “Immunotherapy in colorectal cancer: current and future strategies,” *Journal of the Anus, Rectum and Colon*, vol. 5, no. 1, pp. 11–24, 2021.
- [24] A. F. Oliveira, L. Bretes, and I. Furtado, “Review of PD-1/PD-L1 inhibitors in metastatic dMMR/MSI-H colorectal cancer,” *Frontiers in Oncology*, vol. 9, 2019.
- [25] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, and P. Tamayo, “The molecular signatures database hallmark gene set collection,” *Cell Systems*, vol. 1, no. 6, pp. 417–425, 2015.
- [26] A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, and J. P. Mesirov, “Molecular signatures database (MSigDB) 3.0,” *Bioinformatics*, vol. 27, no. 12, pp. 1739–1740, 2011.
- [27] B. Chen, M. S. Khodadoust, C. L. Liu, A. M. Newman, and A. A. Alizadeh, “Profiling tumor infiltrating immune cells with CIBERSORT,” *Methods in Molecular Biology*, vol. 1711, pp. 243–259, 2018.
- [28] T. Li, J. Fan, B. Wang et al., “TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells,” *Cell*, vol. 77, no. 21, pp. e108–e110, 2017.
- [29] T. M. Malta, A. Sokolov, A. J. Gentles et al., “Machine learning identifies stemness features associated with oncogenic dedifferentiation,” *Cell*, vol. 173, no. 2, pp. 338–354.e15, 2018.
- [30] E. A. Kautto, R. Bonneville, J. Miya et al., “Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS,” *Oncotarget*, vol. 8, no. 5, pp. 7452–7463, 2017.
- [31] S. Jung, L. Yi, D. Jeong et al., “The role of ADCYAP1, adenylate cyclase activating polypeptide 1, as a methylation biomarker for the early detection of cervical cancer,” *Oncology Reports*, vol. 25, no. 1, pp. 245–252, 2011.
- [32] Z. K. Hassan, M. M. Hafez, M. M. Kamel, and A. R. Zekri, “Human papillomavirus genotypes and methylation of CADM1, PAX1, MAL and ADCYAP1 genes in epithelial ovarian cancer patients,” *Asian Pacific Journal of Cancer Prevention*, vol. 18, no. 1, pp. 169–176, 2017.
- [33] K. Hasakova, M. Vician, R. Reis, M. Zeman, and I. Herichova, “The expression of clock genes cry1 and cry2 in human colorectal cancer and tumor adjacent tissues correlates differently dependent on tumor location,” *Neoplasma*, vol. 65, no. 6, pp. 986–992, 2018.
- [34] L. U. Peng, G. Bai, and Y. Pang, “Roles of NPAS2 in circadian rhythm and disease,” *Acta Biochimica et Biophysica Sinica*, vol. 53, no. 10, pp. 1257–1265, 2021.
- [35] W. Wang, R. Kandimalla, H. Huang et al., “Molecular subtyping of colorectal cancer: recent progress, new challenges and emerging opportunities,” *Seminars in Cancer Biology*, vol. 55, pp. 37–52, 2019.
- [36] S. I. Grivennikov, F. R. Greten, and M. J. C. Karin, “Immunity, inflammation, and cancer,” *Cell*, vol. 140, no. 6, pp. 883–899, 2010.
- [37] J. L. Markman and S. L. Shiao, “Impact of the immune system and immunotherapy in colorectal cancer,” *Journal of Gastrointestinal Oncology*, vol. 6, no. 2, pp. 208–223, 2015.
- [38] A. Durgeau, Y. Virk, S. Corgnac, and F. Mami-Chouaib, “Recent advances in targeting CD8 T-cell immunity for more effective cancer immunotherapy,” *Frontiers in Immunology*, vol. 9, p. 14, 2018.
- [39] Z. Bai, Y. Zhou, Z. Ye, J. Xiong, H. Lan, and F. Wang, “Tumor-infiltrating lymphocytes in colorectal cancer: the fundamental indication and application on immunotherapy,” *Frontiers in Immunology*, vol. 12, article 808964, 2021.
- [40] Y. Zheng, S. Chi, and C. Li, “Identification of potential gene drivers of cutaneous squamous cell carcinoma: analysis of microarray data,” *Medicine*, vol. 99, no. 39, article e22257, 2020.
- [41] H. Liu, Y. Gao, S. Hu, Z. Fan, X. Wang, and S. Li, “Bioinformatics analysis of differentially expressed rhythm genes in liver hepatocellular carcinoma,” *Frontiers in Genetics*, vol. 12, article 680528, 2021.
- [42] H. Zhang, C. Qiu, H. Zeng, W. Zhu, W. Lyu, and X. Lao, “Upregulation of stress-induced protein kinase CK1 delta is associated with a poor prognosis for patients with hepatocellular carcinoma,” *Genetic Testing and Molecular Biomarkers*, vol. 25, no. 7, pp. 504–514, 2021.

Research Article

Screening of Prognostic Markers for Hepatocellular Carcinoma Patients Based on Multichip Combined Analysis

Yong Dong ¹, Qian Miao ², and Da Li ¹

¹Department of Medical Oncology, Sir Run Run Shaw Hospital, College of Medicine, Zhejiang University, No. 3 Qingchun East Road, Jianggan District, Hangzhou, Zhejiang 310016, China

²Department of Medical Oncology, Quzhou People's Hospital, No. 2, Zhongludi, Kecheng District, Quzhou, Zhejiang 324000, China

Correspondence should be addressed to Da Li; lidaonconew@zju.edu.cn

Received 13 April 2022; Accepted 28 June 2022; Published 14 July 2022

Academic Editor: Lei Chen

Copyright © 2022 Yong Dong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Object. Due to challenges in HCC prognosis evaluation, novel prognosis-related markers are urgently needed. This study attempted to screen reliable HCC prognostic markers by bioinformatics methods. **Methods.** GSE (14520, 36376, 57957, 76427) datasets were accessed from GEO database. 55 differential mRNAs (DEGs) were obtained by differential analysis based on the datasets. GO and KEGG analysis results indicated that the DEGs were enriched in xenobiotic metabolic process and other pathways. Expression profiles and clinical data of TCGA-LIHC mRNAs were from TCGA database. We established a prognostic model of HCC through univariate and multivariate Cox risk regression analyses. ROC curve analysis was used to examine the prognostic model performance. GSEA analysis was performed between the high- and low-risk score sample groups. **Results.** A 4-gene HCC prognostic model was constructed, in which the gene expressions correlated to HCC patients' survival. The AUC value presented 0.734 in the ROC analysis for the prognostic model. **Conclusion.** The four-gene model could be introduced as an independent prognostic factors to assess HCC patients' survival status.

1. Introduction

Liver cancer is the sixth most commonly diagnosed malignancy, with increasing incidence and mortality year by year [1]. Primary liver cancer consists of 75%-85% hepatocellular carcinoma (HCC) and 10%-15% intrahepatic cholangiocarcinoma along with other rare types [2]. Clinical staging, age, and vascular invasion as important clinical factors may contribute to improvement of survival prediction [3]. Due to the complex molecular mechanism of tumor regulation, traditional clinical information prediction ability is limited. Though a lot of strategies such as hepatectomy, radiofrequency ablation, transcatheter arterial chemoembolization, liver transplantation, and chemotherapy are used in HCC treatment, HCC patient's prognosis is still poor [4]. About 70% of HCC patients have recurrence or metastasis within 5 years after surgery [5]. With the continuous improvement of tumor heterogeneity and molecular mechanism research, new tools like molecular markers are in urgent need for accurate prediction of patient's prognosis.

Genome sequencing technology and data emerged in this genomic era [6]. Microarray and bioinformatics analysis are applied to screen differentially expressed genes (DEGs) at the genomic level, which may aid in the identification of DEGs and pathways correlated with occurrence and development of HCC. Gene chips can rapidly detect DEGs, generate slicing data, and store them in a public database, which is a reliable technology [7]. Great contributions have been made to tumor diagnosis and prognosis prediction. On the basis of these data, valuable evidence can be found for new research. For example, through public databases, more and more potential biomarkers are being mined [8–10]. Feng et al. used lung adenocarcinoma (LUAD) dataset from TCGA and found that LUAD patients expressing high HMGB1 level had dismal overall survival (OS) [11]. Wang et al. take advantage of RNA sequencing (RNA-seq) data from TCGA-KIRC to confirm DEGs [12]. For HCC, Kong et al. selected the top 25% DEGs from the GSE62232 dataset and screened out prognostic-related modules to build protein-protein interaction (PPI) networks, who obtained 5 candidate genes (PCNA, RFC4, PTTG1, H2AFZ,

and RRM1) finally [13]. It is indicated that these fresh next-generation sequencing methods and data can screen biomarkers.

Despite the prognostic ability of signature genes being explored by bioinformatics analysis, these prognostic prediction effects are limited due to the heterogeneity of HCC [14]. These prognostic prediction effects are subject to certain challenges, and thus, it is necessary to develop reliable novel signature genes to design a more personalized diagnostic and therapeutic plan. In this study, we downloaded 4 sets of HCC mRNA chip data from GEO database and analyzed them to obtain the differentially expressed mRNAs (DEmRNAs) between HCC tissues and noncancer tissues. Subsequently, biological significance and key pathways of the DEGs were explored through GO and KEGG enrichment analyses. Cox regression analysis, risk assessment, and clinical information analysis helped us find markers that could predict HCC patients' prognoses. Here, a total of 55 DEmRNAs and 4 independent prognostic factors for HCC were identified, which may be candidate biomarkers for HCC, being utilized for prediction of HCC progression and prognosis. A new method for adjuvant therapy of clinical HCC patients was put forward in this study.

2. Materials and Methods

2.1. Data Downloading and Processing. HCC-related matrix was downloaded from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). Criteria for obtaining matrix were (1) human HCC tissue samples, (2) cancer and noncancer samples, (3) the total number of samples ≥ 30 , and (4) based on the same sequencing platform. Four groups of HCC mRNA chip data including GSE14520 (normal: 220, tumor: 225), GSE36376 (normal: 193, tumor: 240), GSE57957 (normal: 193, tumor: 240), and GSE76427 (normal: 52, tumor: 115) were selected for this research. The sequencing platform for GSE36376, GSE57957, and GSE76427 was GPL10558 Illumina HumanHT-12 V4.0 chip, and the sequencing platform for GSE14520 was GPL3921 HT_HG-U133A. Then, the combined differential analysis ($\log_{FC} > |1.0|$, $\text{padj} < 0.05$) was done with "limma" (PMID: 25605792) and "RobustRankAggreg" (PMID: 31638362) in R package to identify DEmRNAs in HCC tissues.

The mRNA expression data and clinical data of TCGA-LIHC were downloaded from TCGA database (<https://portal.gdc.cancer.gov/>). DEmRNA expression and clinical information screened by differential expression analyses were extracted for subsequent validation.

2.2. Functional Enrichment Analysis. GO and KEGG enrichment analyses of DEGs were conducted using R package "clusterProfiler" (PMID: 22455463). GO enrichment analysis investigates the biological significance of DEmRNAs, including long-term enrichment analysis of biological processes, molecular functions, and cellular components. KEGG pathway enrichment analysis seeks key pathways closely related to DEmRNAs. P value < 0.05 and false discovery rate (FDR) < 0.05 were set as thresholds. P value was calculated based on hypergeometric separation.

2.3. Cox Regression Analysis and Model Assessment. Univariate and multivariate Cox risk regression analyses were done on DEmRNAs using the R package "survival" [15]. At first, the clinical information provided by TCGA was used to conduct univariate Cox regression analysis on DEmRNAs, and the DEmRNAs with P value < 0.01 were selected (omnibus test). Then, these DEmRNAs were further analyzed by multivariate Cox regression analysis to establish a multigene risk model. Samples were divided into high- and low-risk groups by median risk score, and OS of two groups was compared by R package "survival". To assess the accuracy and predictive value of risk assessment model, we used R package "survivalROC" (PMID: 33790572) to draw receiver operating characteristic (ROC) curves and conduct independent ROC tests for genes in risk assessment model, respectively, to obtain area under curve (AUC) value.

2.4. GSEA Analysis for Hub Genes. To investigate pathway changes in the high- and low-risk groups, the GSEA software (V 4.1.0) was utilized to perform KEGG enrichment pathway analysis to explore potential mechanism of action (PMID: 16199517). Normalized enrichment score (NES) and FDR were used to quantify enrichment and statistical significance, respectively [16].

2.5. Correlation Analysis of Risk Scores and Clinical Characteristics. The expression levels of feature genes in the high- and low-risk groups were analyzed by using heat map, and their differences in different clinical information and risk score coefficients were analyzed. The performance of risk scores in different tumor stages was analyzed by box plots, and the reliability of the prognostic model in liver cancer was further validated. Kaplan-Meier survival curve was used to assess the effect of the combination of high- or low-risk score and tumor stage on survival rate (log-rank test was introduced for statistical test).

2.6. Expressions of Hub Genes and Survival Analysis. GEPIA database (<http://gepia.cancer-pku.cn/index.html>) is an integration of RNA expression profiles of tumor and normal samples from TCGA and GTEx projects [17], which can be used for individual analysis. Kaplan-Meier plotter is a popular web tool for assessing the impact of a large number of genes on survival based on EGA, TCGA, and GEO databases [18]. In this study, GEPIA was utilized for expression confirmation and survival analysis of hub genes, and log-rank tests were conducted to measure statistical significance.

3. Results

3.1. Identification of DEmRNAs in HCC. Four datasets GSE14520, GSE36376, GSE57957, and GSE76427 were obtained from GEO database. Subsequently, a differential gene expression analysis was performed among the 4 datasets and then integrated to remove batch effects; finally, 6 significantly upregulated mRNAs and 49 prominently downregulated mRNAs were screened (Figure 1). These DEGs were selected for further analysis.

3.2. GO and KEGG Enrichment Analyses for the DEmRNAs. Then, functional enrichment analysis was conducted on

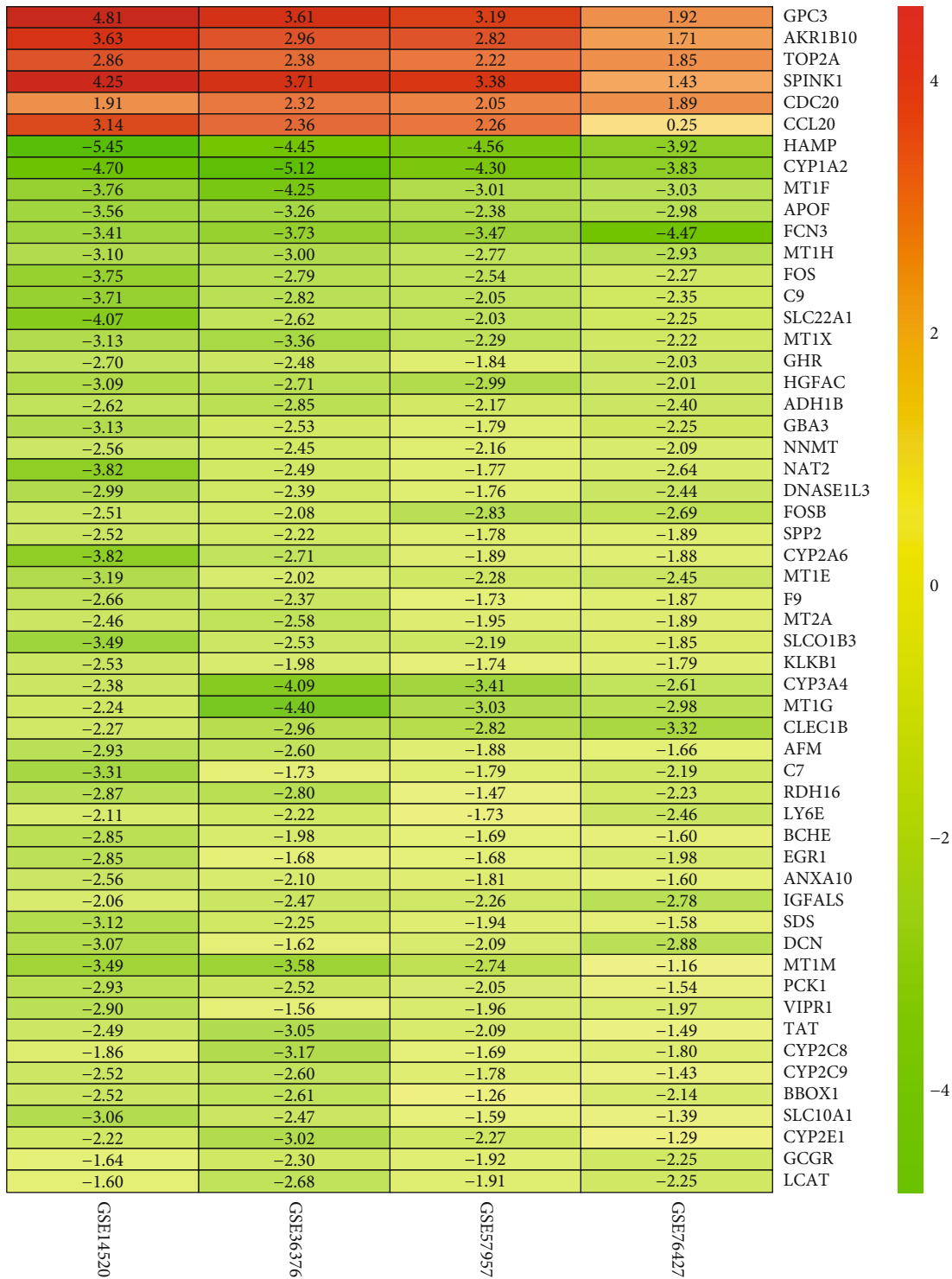
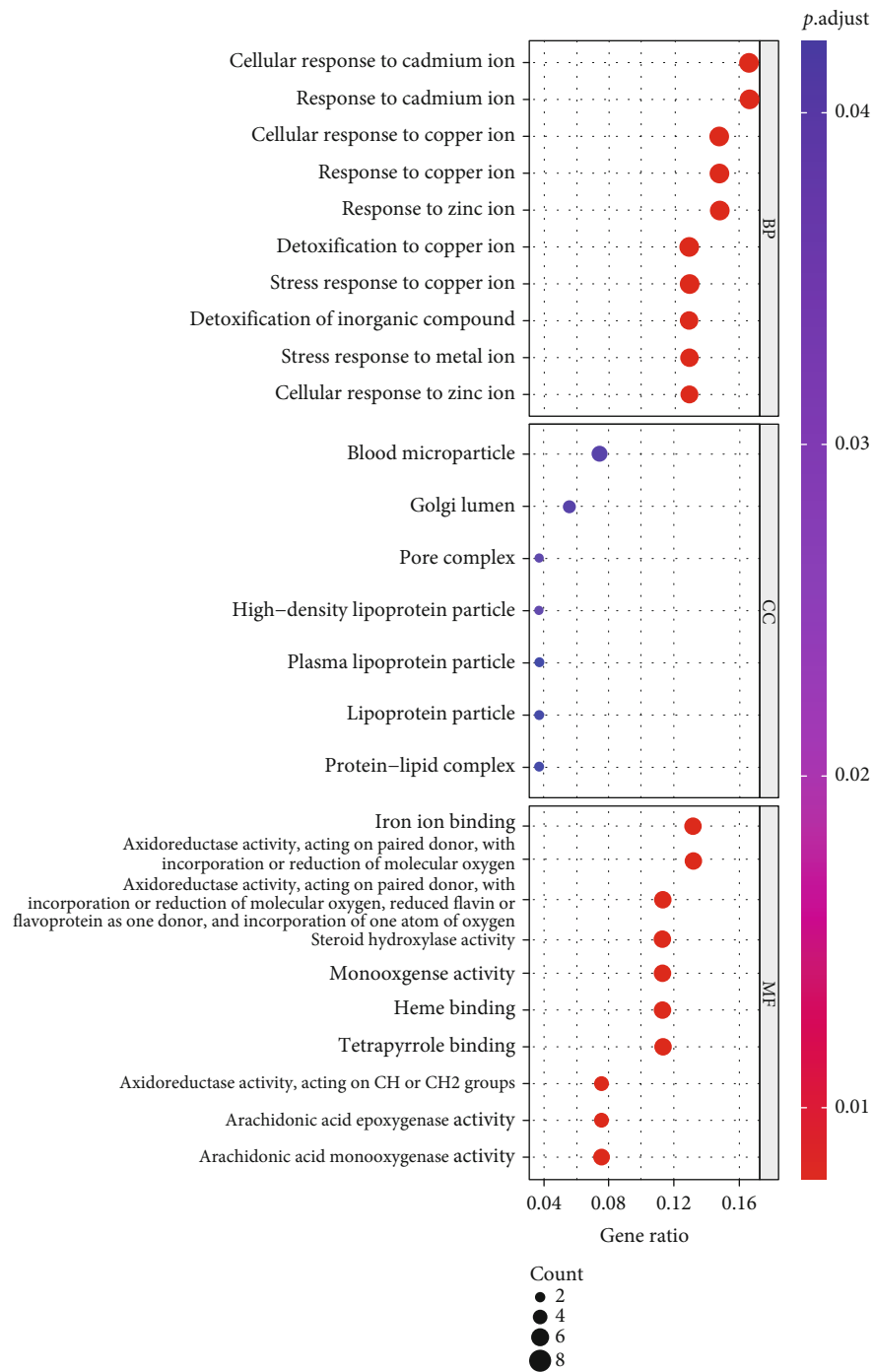


FIGURE 1: Heat map of DEGs screened by combined analysis of 4 datasets from GEP database. The numbers in the grids represent the fold changes of genes in different datasets between the tumor group and the normal group.

above 55 DEMRNAs to explore their potential biological functions. GO enrichment analysis revealed that these genes were mainly concentrated in cellular response to cadmium ion, detoxification of inorganic compound, and cellular response to zinc ion (Figure 2(a)). KEGG unveiled those genes were

enriched mainly in mineral absorption, chemical carcinogenesis DNA adducts, and drug metabolism-cytochrome P450 (Figure 2(b)). Therefore, these DEMRNAs may influence the progression of HCC by influencing these biological functions and pathways.



(a)

FIGURE 2: Continued.

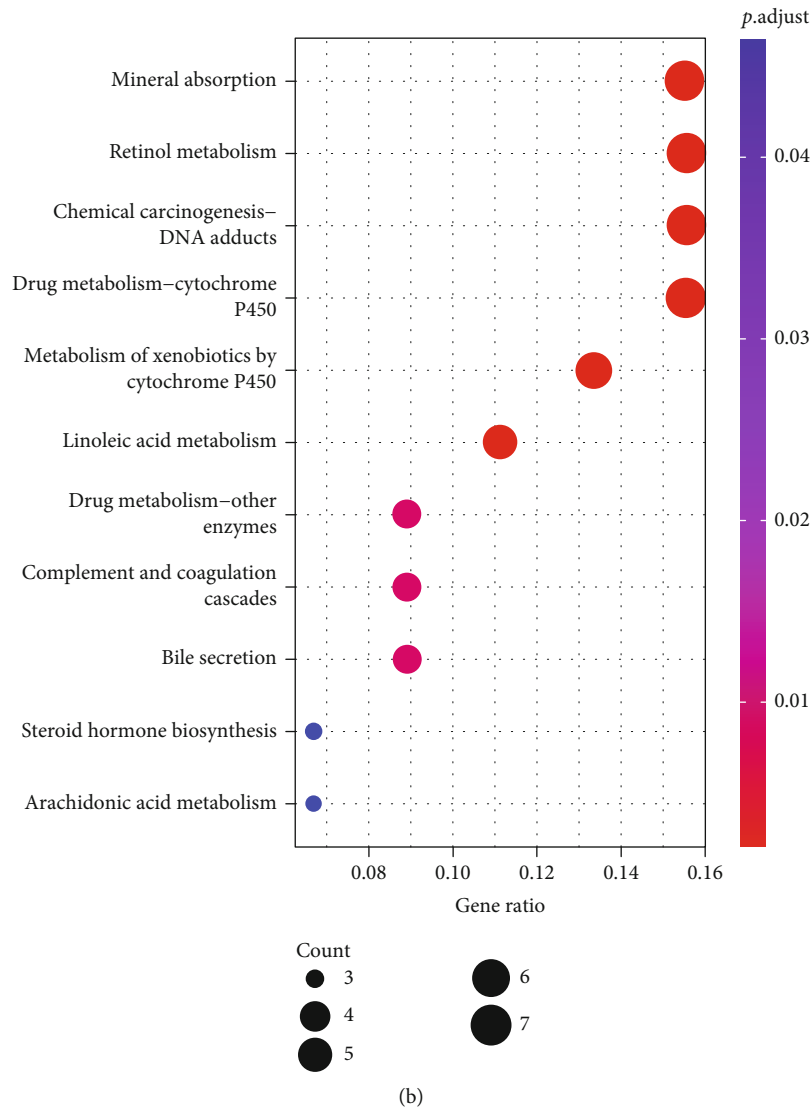


FIGURE 2: GO and KEGG pathway analyses of DE mRNAs in HCC: (a, b) GO and KEGG enrichment analyses of DEGs.

3.3. *Screening and Verifying the Optimal Prognostic Hub Genes.* To screen out hub genes with the best prognosis, we carried out univariate Cox regression analysis on 55 DE mRNAs and selected 20 genes noticeably associated with HCC patients' prognoses (Table 1). Multivariate Cox regression analysis was done on these 20 related genes to establish a risk model of four genes, among which CDC20 was a high-risk factor while CYP2C9, CLEC1B, and LCAT were low-risk factors (Figure 3(a)). Next, patients were assorted into high- and low-risk groups by median risk score. R package "survival" was utilized to compare survival time of two groups, and survival curves were plotted. A significant difference in OS was seen between the two groups, which demonstrated that survival time of patients in the high-risk group was dramatically shorter than the other group, indicating that the risk score could be used for risk grading and prognosis assessment of HCC patients (Figure 3(b)). The ROC curves manifested that AUC value of the model was 0.734 (Figure 3(c)), indicating

that the model had a certain accuracy in diagnosing patients' prognoses. Thus, risk scores of these four hub genes were dramatically associated with HCC patients' prognoses.

3.4. *Functional Analysis of 4 Selected Genes via GSEA.* In order to analyze function changes of the high- and low-risk groups, we used the GSEA software to carry out KEGG pathway analysis on the two groups, and the results displayed that compared with the low-risk group, complement and blood coagulation cascade, fatty acid metabolism, peroxidase, and primary bile acid biosynthesis pathways are significantly activated in the high-risk group (Figure 4). These pathways are closely related to the liver and are the major metabolic processes in the liver. Hence, changes in these pathways can affect the metabolic function of the liver. Therefore, we speculated that hub genes may affect the biological function of cancer cells by regulating the metabolic process of the liver, so as to play a role in prognosis.

TABLE 1: Univariate Cox regression analysis screened genes that were significantly associated with the prognosis of HCC patients.

Gene	HR	z	P value
CDC20	1.303474527	4.659540595	3.17E-06
LCAT	0.800676874	-4.422698682	9.75E-06
DNASE1L3	0.834137148	-4.228589383	2.35E-05
GHR	0.813122352	-4.187942095	2.81E-05
CYP2C9	0.887041102	-4.126833672	3.68E-05
ANXA10	0.877021937	-3.969016876	7.22E-05
SPP2	0.907350691	-3.964707036	7.35E-05
SLC10A1	0.914033265	-3.67849838	0.000234611
SLC22A1	0.909870575	-3.598475946	0.000320087
AFM	0.902676346	-3.557458905	0.00037446
TAT	0.912717953	-3.472545816	0.000515547
ADH1B	0.901724063	-3.429051233	0.000605695
KLKB1	0.842899464	-3.320604451	0.000898227
VIPR1	0.851232022	-3.312650071	0.000924165
TOP2A	1.210739633	3.308652423	0.000937461
RDH16	0.903679968	-3.278095563	0.0010451
CLEC1B	0.860469147	-3.090652798	0.00199717
FCN3	0.8707386	-2.958448406	0.00309192
C7	0.919120279	-2.896960425	0.003767973
IGFALS	0.912348703	-2.779238461	0.005448651

3.5. Relationship between the Hub Genes and Clinical Features.

To validate the reliability of the prognostic model, we explored the performance of the risk score on clinical features. Heat map results showed that CDC20 level increased gradually, while CYP2C9, CLEC1B, and LCAT levels decreased gradually, with increased risk score (Figure 5(a)). Significant differences were seen in different grades, pathological stages, and T stages between high- and low-risk group (Figure 5(a)). Survival time of patients with risk score combined with different clinical stages, pathological grades, and T stages was significantly different (Figure 5(b)), which may be reference factors predicting patient's prognosis. With progression of tumor, the risk value increased (Figure 5(c)). This further indicated that the four-gene model could predict the risk of tumors, and the risk scores and some clinical features established by us (stage, grade, T) could be utilized to comprehensively evaluate patient's prognosis.

3.6. Analysis of Hub Genes by the Kaplan-Meier Plotter and GEPIA.

GEPIA dataset was used to conduct survival and ROC analysis on 4 hub genes in cancer and normal tissues to test significance of the 4 hub genes to patients' prognosis. We searched the Kaplan-Meier OS curves and progression-free survival (PFS) curves of the four genes in the risk model in the GEPIA database and manifested that these four-gene levels had a dramatic impact on OS of patients (Figure 6(a)), among which the HCC patients with high expressions of CLEC1B, CYP2C9, and LCAT presented poor prognosis. Moreover, CDC20, CLEC1B, and LCAT expression levels also

had significant effects on PFS (Figure 6(b)). Independent ROC tests were conducted for the four genes in the risk model, and it was noted that the AUC value of each gene reached more than 0.6 (Figure 6(c)). Therefore, the above results suggested that the expressions of these four genes in HCC were related to the progression of cancer and correlated with prognosis, and these four genes could be used as independent prognostic factors for HCC.

4. Discussion

Liver cancer is a malignant tumor with different histological characteristics, which is mainly caused by chronic hepatitis virus infection, gene mutation, cell damage, alcoholic liver disease, and aflatoxin poisoning. Despite advances in cancer treatment, prognosis is unfavorable with increasing morbidity and mortality [19]. Most HCC patients without early detection are not suitable for radical treatment, which may be a reason for dismal prognosis. Many studies have demonstrated that biomarkers can determine the progression of cancer. For example, CCND1, c-myc, and RAS mutations, and hypermethylation of CCND2 promoter have been associated with HCC [20, 21]. Splicing changes of NT5E, Sulf1, and SLC39A14 were also associated with HCC [22–24]. All of these factors can be used as markers to predict the progression of HCC. However, the molecular mechanism of HCC is very complex, including many genetic and epigenetic changes, chromosomal aberrations, gene mutations, and alternative molecular pathways [25]. Therefore, potential and efficient diagnostic and therapeutic markers are in great need. Microarray technology enables the exploration of genetic changes in HCC and has been proved to be an effective strategy for mining novel biomarkers in other diseases.

We analyzed four microarray datasets by multichip combined analysis to obtain DEMRNAs between HCC and non-cancer tissues. 55 DEMRNAs were identified, containing 6 upregulated genes and 49 downregulated mRNAs. It is well known that main metabolic pathways in the liver include bile acid synthesis, fatty acid metabolism, complement and coagulation cascade, etc., which play a vital role in disease and human homeostasis [26–29]. Therefore, we used GO and KEGG enrichment analyses to explore biological functions and pathway enrichment of DEMRNAs. The results exhibited that DEMRNAs were mainly concentrated in biological functions such as response to exotic biological stimulation, drug catabolism, complement activation, cAMP response, and signaling pathways such as cytochrome P450, drug metabolism, complement and coagulation cascade, fatty acid metabolism, peroxisome, and primary bile acid biosynthesis. Another study has suggested that the activation of complement can promote the tumor [30]. Oxidoreductase activity may lead to antioxidant defense and can encode tumor repressors that are often altered in tumors [31, 32]. The accumulation of bile acids in the liver inhibits cell growth and improves survival [33]. These conclusions are all consistent with our predictions.

Subsequently, Cox regression analysis and the construction of risk model obtained the 4-gene risk model consisting of CDC20, CLEC1B, CYP2C9, and LCAT. The risk values of these four-gene sets were evaluated and verified to be

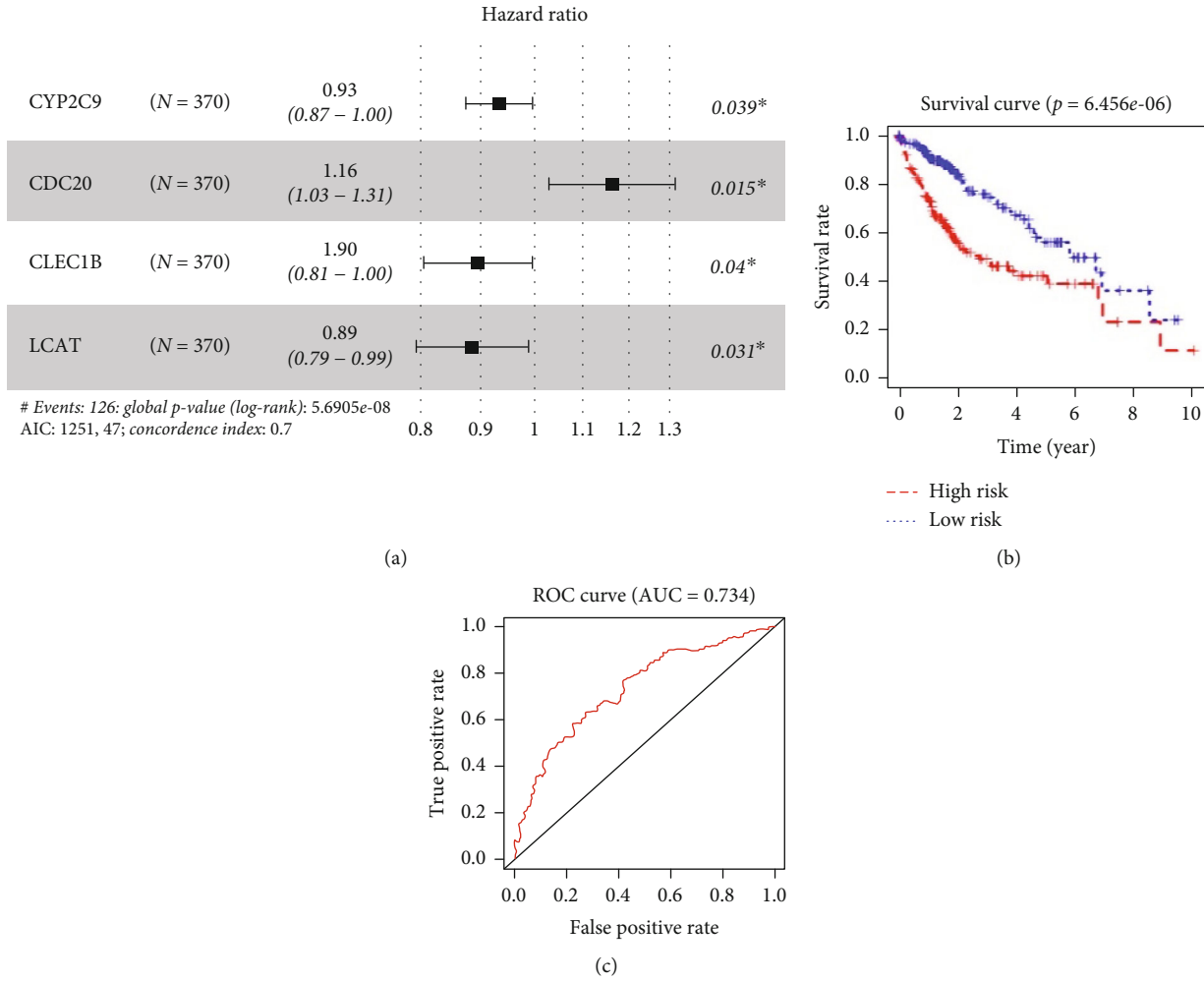


FIGURE 3: Constructing and verification the prognostic model. (a) Forest map of the 4-gene prognostic model established by multivariate Cox risk regression analysis; (b) Kaplan-Meier OS curves for the high- and low-risk groups are plotted based on risk score. Horizontal axis: survival time; vertical axis: overall rate; red line: high-risk samples; blue line: low-risk samples. (c) ROC curves based on the risk model.

prognostic factors for HCC. We also found from the extensive literature that these hub genes played an essential role in many cancers including HCC. Based on the gene expression and clinical information of liver cancer patients obtained from the TCGA database, Long et al. used univariate, LASSO, and multivariate Cox regression analyses to establish the prognostic model of liver cancer and used GEO database to verify the feasibility of the model. Finally, a prognostic model based on four marker genes is obtained, and the model can predict the overall survival of liver cancer patients [34]. Similarly, in our study, based on 4 GSE datasets in the GEO database, the liver cancer prognostic model was established by univariate and multivariate Cox regression models and finally verified by the GEPIA database. Since we built the model based on multiple gene sets, the results of this study are more reliable to a certain extent as compared with the earlier studies. Dysregulation of cell cycle processes is vital in the development of tumors [20, 35, 36]. CDC20 is an important regulatory factor in the cell cycle process, which forms ubiquitin-proteolytic enzyme complex

by binding with APC and participates in the degradation of various proteins to modulate cell cycle process [37]. It has been reported that upregulation of CDC20 may predict the decrease of OS and DFS in HCC patients [38]. Increased CDC20 in HCC was related to gender, differentiation, and TNM stage [39], which is similar to our results. Through risk model, we manifested that CDC20 was a high-risk factor and was also associated with clinical features. CYP2C9 is a drug metabolism enzyme and is a decreased low-risk factor in HCC [40]. Shuaichen and Guangyi have discovered that CYP2C9 may promote the development of HCC, especially can be a diagnostic biomarker in drug metabolism [41]. CLEC2 is expressed in platelets and some hematopoietic cells. Wang et al. have put forward that CLEC2 inhibits gastric cancer metastasis, prevents the activation of AKT and glycogen synthase signals, and affects the invasion and expression of EMT markers, which can be a potential biomarker in gastric cancer [42]. However, fewer studies on CLEC2 in HCC have been reported. So, our study provided a data source for the role of this gene in HCC. Finally, LCAT

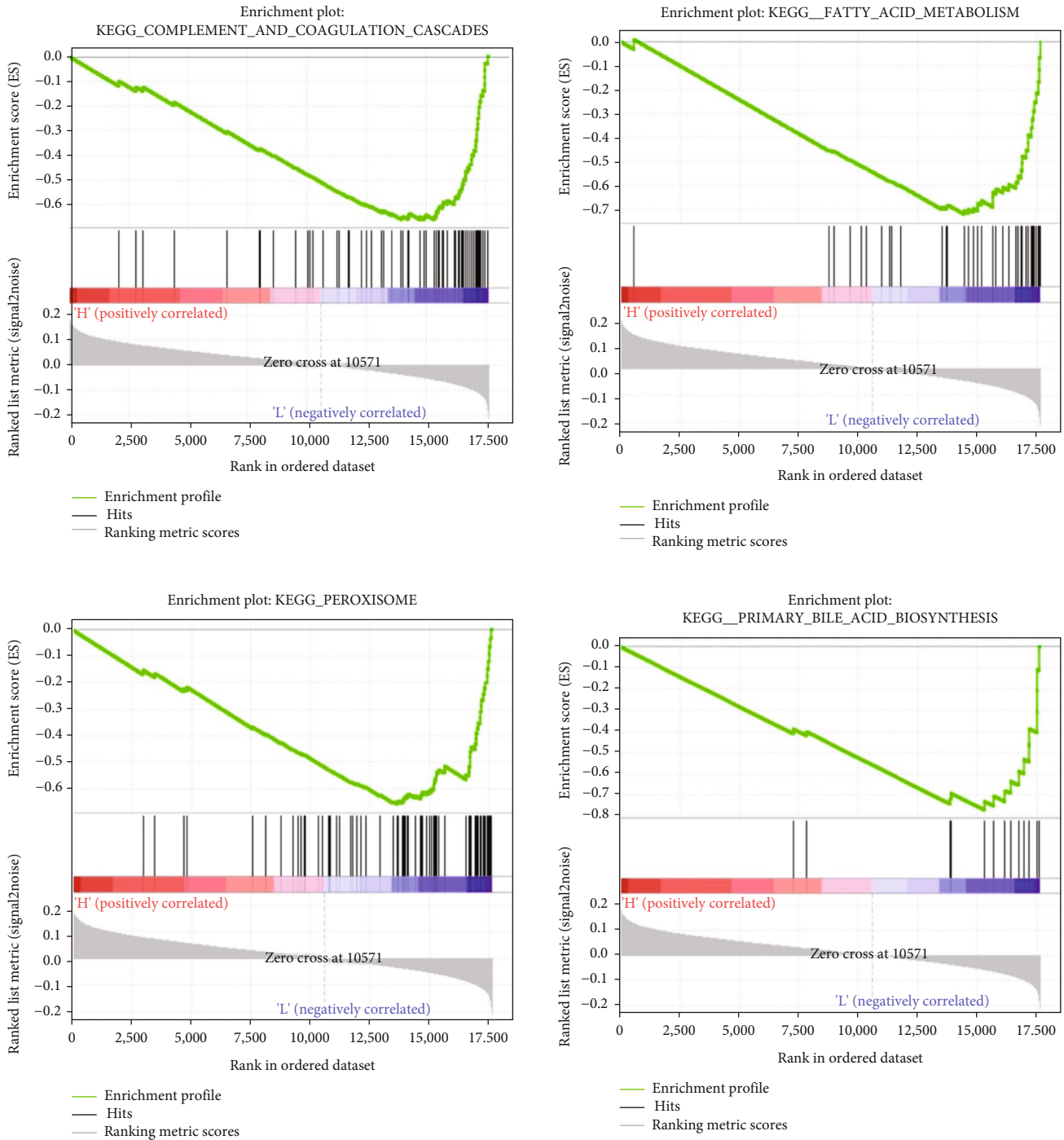


FIGURE 4: GSEA pathway enrichment results based on median risk score grouping.

is the only enzyme that can esterify cholesterol in plasma, which determines maturation of high-density lipoprotein as a key enzyme for reverse cholesterol transport with reports of its role in atherosclerosis, cholesterol deposition, and kidney [43–45]. Besides, Russell et al. have pointed out that LCAT plays a role as a diagnostic marker in epithelial ovarian cancer [46]. It can also predict OS of HCC [47]. In conclusion, the four hub genes found in this study played a crucial role in HCC. These genes also function in pathways

such as complement and blood coagulation cascade, fatty acid metabolism, peroxisome, and primary bile acid biosynthesis. Their expressions in HCC will indirectly affect the changes of these pathways, thus affecting the incidence of HCC, which fully illustrates the importance of these hub genes of HCC progression.

Taken together, we attempted to screen DEmRNAs that may be related to HCC occurrence and development. 55 DEmRNAs and 4 hub genes had been identified as diagnostic

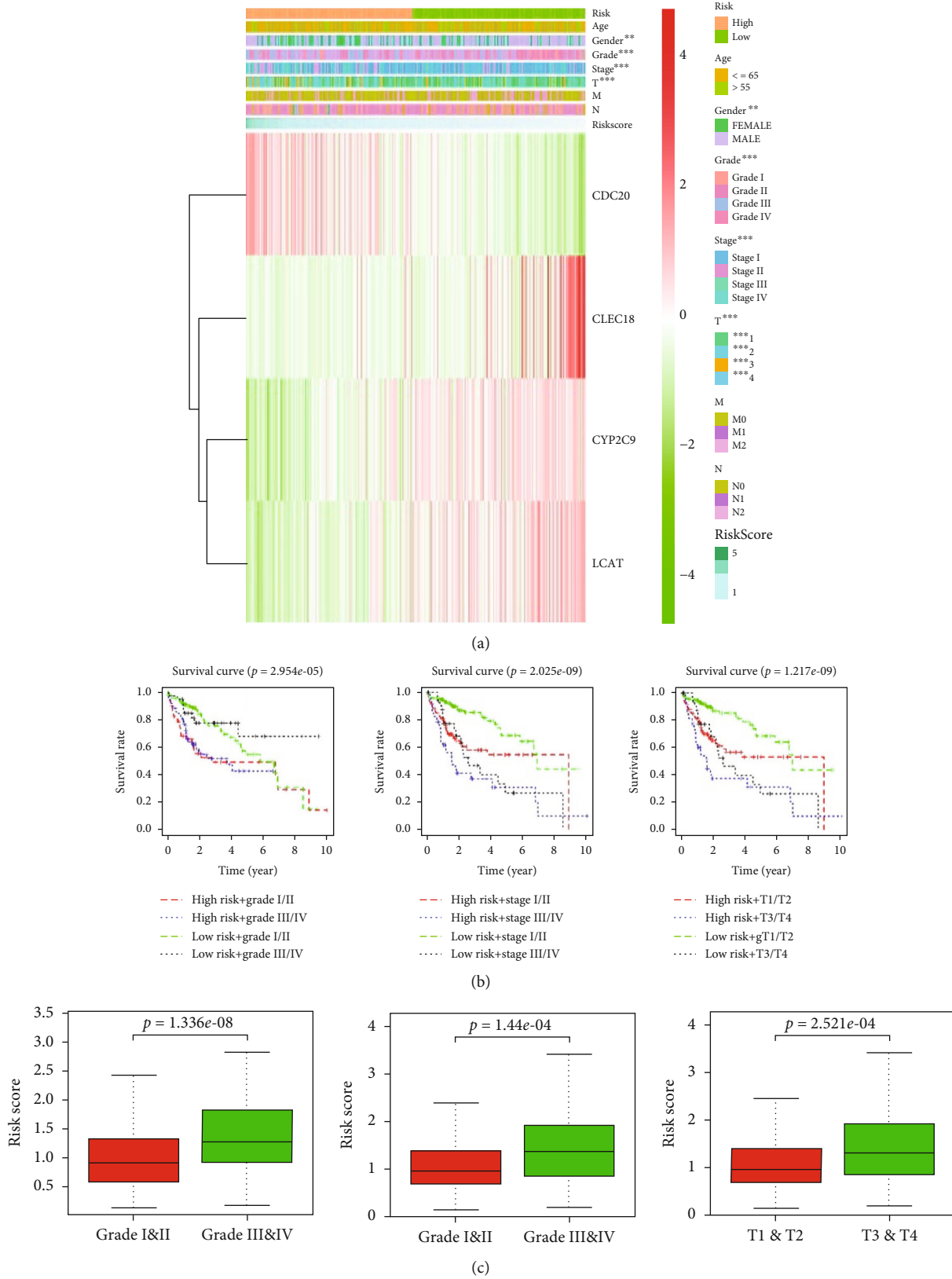


FIGURE 5: Analysis of DEMRNAs and clinical relationship and patients' prognosis. (a) Heat map of 4 gene expressions in the high- and low-risk groups of risk model and clinicopathological differences between the two groups; (b) OS curves combined risk score and the patients' different tumor stages, with the horizontal axis representing survival time and the vertical axis representing OS rate. Different color curves represent different combinations of risk and tumor stage. (c) Boxplots of differences in risk scores of different clinical stages, grades, and T stages.

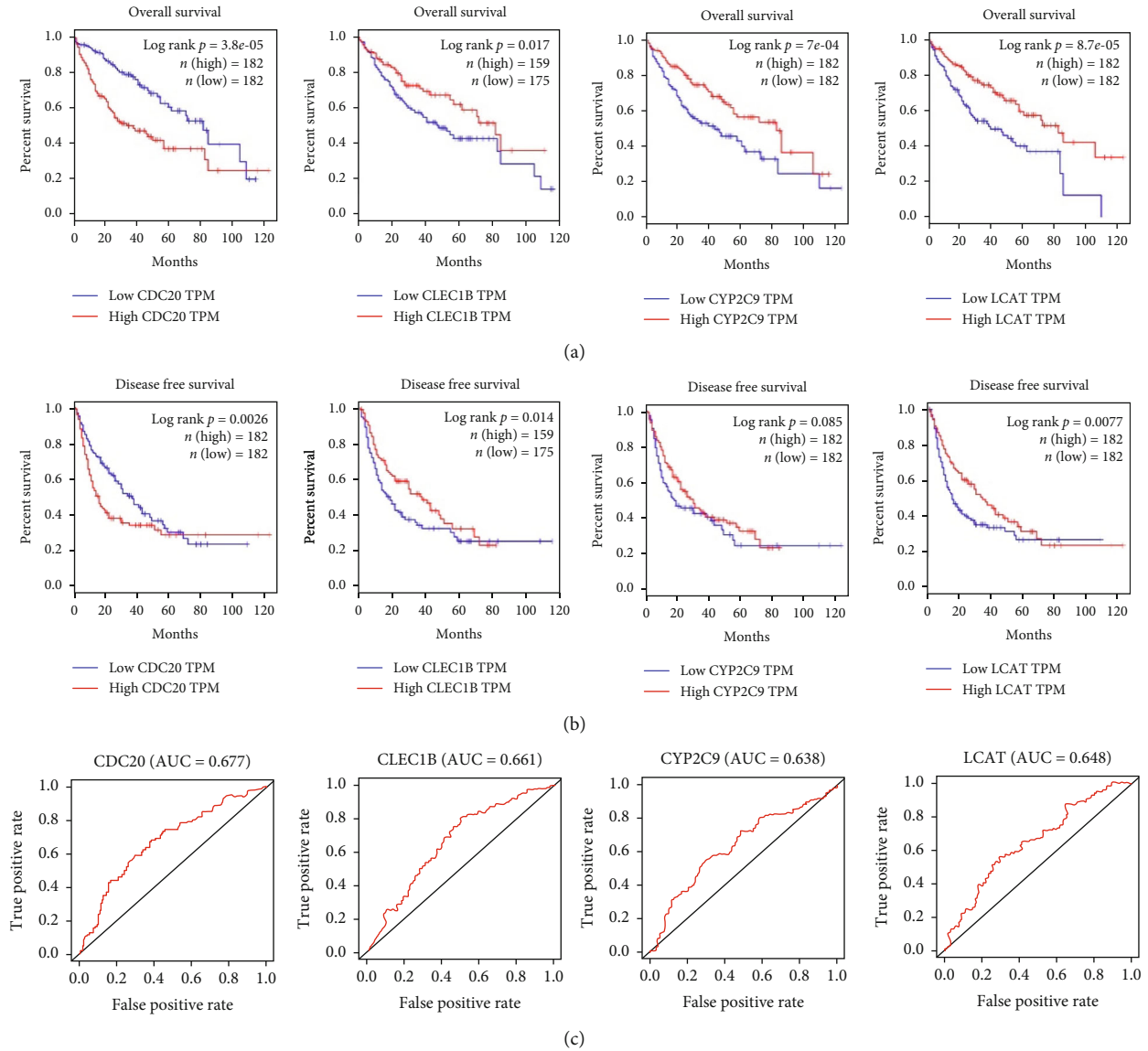


FIGURE 6: Analysis of hub genes by the Kaplan-Meier plotter and GEPIA. (a, b) Kaplan-Meier OS curves and PFS curves of four genes in the risk model are retrieved in the GEPIA database. Horizontal axis: survival time; vertical axis: survival rate; red line: high-expression group; blue line: low-expression group. (c) Independent ROC curves of four genes in the risk model for HCC patient's prognosis.

biomarkers for HCC, which offered an effective basis for the treatment of HCC. But this experiment still has certain limitations. This study only conducted a pure bioinformatics analysis through the GEO database, further verification in multicenter clinical trials, and prospective studies is required. Nevertheless, biological functions of these genes in HCC need further study.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Consent

All authors consent to submit the manuscript for publication.

Conflicts of Interest

The authors declare that they have no potential conflicts of interest.

Authors' Contributions

DY and LD both contribute to the work, including conception and design and article drafting and revising. MQ is the guarantor for the article who accept full responsibility for the work.

References

- [1] H. B. El-Serag and K. L. Rudolph, "Hepatocellular carcinoma: epidemiology and molecular carcinogenesis," *Gastroenterology*, vol. 132, no. 7, pp. 2557–2576, 2007.

- [2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [3] K. Ikeda, S. Saitoh, A. Tsubota et al., "Risk factors for tumor recurrence and prognosis after curative resection of hepatocellular carcinoma," *Cancer*, vol. 71, no. 1, pp. 19–25, 1993.
- [4] A. Villanueva, Y. Hoshida, C. Battiston et al., "Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma," *Gastroenterology*, vol. 140, no. 5, pp. 1501–1512.e2, 2011.
- [5] F. Ji, Y. Liang, S. Fu et al., "Prognostic value of combined preoperative prognostic nutritional index and body mass index in HCC after hepatectomy," *HPB: the official journal of the International Hepato Pancreato Biliary Association*, vol. 19, no. 8, pp. 695–705, 2017.
- [6] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nature Reviews Genetics*, vol. 10, no. 1, pp. 57–63, 2009.
- [7] B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler, "Cancer genome landscapes," *Science*, vol. 339, no. 6127, pp. 1546–1558, 2013.
- [8] B. Li, K. Pu, and X. Wu, "Identifying novel biomarkers in hepatocellular carcinoma by weighted gene co-expression network analysis," *Journal of cellular biochemistry*, vol. 120, no. 7, pp. 11418–11431, 2019.
- [9] E. Martínez-Ledesma, R. G. Verhaak, and V. Treviño, "Identification of a multi-cancer gene expression biomarker for cancer clinical outcomes using a network-based algorithm," *Scientific Reports*, vol. 5, no. 1, p. 11966, 2015.
- [10] D. Huang, W. Sun, Y. Zhou et al., "Mutations of key driver genes in colorectal cancer progression and metastasis," *Cancer Metastasis Reviews*, vol. 37, no. 1, pp. 173–187, 2018.
- [11] A. Feng, Z. Tu, and B. Yin, "The effect of HMGB1 on the clinicopathological and prognostic features of non-small cell lung cancer," *Oncotarget*, vol. 7, pp. 20507–20519, 2016.
- [12] Y. Wang, L. Chen, G. Wang et al., "Fifteen hub genes associated with progression and prognosis of clear cell renal cell carcinoma identified by coexpression analysis," *Journal of Cellular Physiology*, vol. 234, no. 7, pp. 10225–10237, 2019.
- [13] J. Kong, T. Wang, Z. Zhang, X. Yang, S. Shen, and W. Wang, "Five core genes related to the progression and prognosis of hepatocellular carcinoma identified by analysis of a coexpression network," *DNA and Cell Biology*, vol. 38, no. 12, pp. 1564–1576, 2019.
- [14] Z. Wang, L. Pan, D. Guo et al., "A novel five-gene signature predicts overall survival of patients with hepatocellular carcinoma," *Cancer Medicine*, vol. 10, no. 11, pp. 3808–3821, 2021.
- [15] T. M. Therneau and P. M. Grambsch, *Modeling Survival Data: Extending the Cox Model*, Springer, New York, 2000.
- [16] A. Eraso-Pichot, M. Brasó-Vives, A. Golbano et al., "GSEA of mouse and human mitochondriomes reveals fatty acid oxidation in astrocytes," *Glia*, vol. 66, no. 8, pp. 1724–1735, 2018.
- [17] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, and Z. Zhang, "GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses," *Nucleic Acids Research*, vol. 45, no. W1, pp. W98–w102, 2017.
- [18] A. M. Szász, A. Lánckzy, Á. Nagy et al., "Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients," *Oncotarget*, vol. 7, no. 31, pp. 49322–49333, 2016.
- [19] D. Sia, A. Villanueva, S. L. Friedman, and J. M. Llovet, "Liver cancer cell of origin, molecular class, and effects on patient prognosis," *Gastroenterology*, vol. 152, no. 4, pp. 745–761, 2017.
- [20] Y. Wang, J. Cheng, C. Xu et al., "Quantitative methylation analysis reveals gender and age differences in p16INK4a hypermethylation in hepatitis B virus-related hepatocellular carcinoma," *Liver international: official journal of the International Association for the Study of the Liver*, vol. 32, no. 3, pp. 420–428, 2011.
- [21] M. Gorath, T. Stahnke, T. Mronga, O. Goldbaum, and C. Richter-Landsberg, "Developmental changes of tau protein and mRNA in cultured rat brain oligodendrocytes," *Glia*, vol. 36, no. 1, pp. 89–101, 2001.
- [22] N. T. Snider, P. J. Altshuler, S. Wan, T. H. Welling, J. Cavalcoli, and M. B. Omary, "Alternative splicing of human NT5E in cirrhosis and hepatocellular carcinoma produces a negative regulator of ecto-5'-nucleotidase (CD73)," *Molecular Biology of the Cell*, vol. 25, no. 25, pp. 4024–4033, 2014.
- [23] R. B. Gill, A. Day, A. Barstow, G. Zaman, C. Chenu, and G. K. Dhoot, "Mammalian Sulfl RNA alternative splicing and its significance to tumour growth regulation," *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine*, vol. 33, no. 5, pp. 1669–1680, 2012.
- [24] R. B. Franklin, B. A. Levy, J. Zou et al., "ZIP14 zinc transporter downregulation and zinc depletion in the development and progression of hepatocellular cancer," *Journal of Gastrointestinal Cancer*, vol. 43, no. 2, pp. 249–257, 2012.
- [25] S. T. Awwad and J. P. McCulley, "Wavefront-guided LASIK: recent developments and results," *International Ophthalmology Clinics*, vol. 46, no. 3, pp. 27–38, 2006.
- [26] C. M. Pusec, A. de Jesus, M. W. Khan et al., "Hepatic HKDC1 expression contributes to liver metabolism," *Endocrinology*, vol. 160, no. 2, pp. 313–330, 2019.
- [27] J. Muciño-Bermejo, R. Carrillo-Esper, M. Uribe, and N. Méndez-Sánchez, "Coagulation abnormalities in the cirrhotic patient," *Annals of Hepatology*, vol. 12, no. 5, pp. 713–724, 2013.
- [28] W. Li, W. Zhang, M. Chang et al., "Metabonomics reveals that triclocarban affects liver metabolism by affecting glucose metabolism, β -oxidation of fatty acids, and the TCA cycle in male mice," *Toxicology Letters*, vol. 299, pp. 76–85, 2018.
- [29] M. Li, S. Y. Cai, and J. L. Boyer, "Mechanisms of bile acid mediated inflammation in the liver," *Molecular Aspects of Medicine*, vol. 56, pp. 45–53, 2017.
- [30] M. M. Markiewski and J. D. Lambris, "Unwelcome complement," *Cancer Research*, vol. 69, no. 16, pp. 6367–6370, 2009.
- [31] M. Abu-Remaileh and R. I. Aqeilan, "The tumor suppressor WW domain-containing oxidoreductase modulates cell metabolism," *Experimental biology and medicine*, vol. 240, no. 3, pp. 345–350, 2015.
- [32] P. Roszczenko, K. A. Radomska, E. Wywiał, J. F. Collet, and E. K. Jagusztyn-Krynicka, "A novel insight into the oxidoreductase activity of Helicobacter pylori HP0231 protein," *PLoS One*, vol. 7, no. 10, article e46563, 2012.
- [33] W. Tang, V. Putluri, C. R. Ambati, T. H. Dorsey, N. Putluri, and S. Ambs, "Liver- and microbiome-derived bile acids accumulate in human breast tumors and inhibit growth and improve patient survival," *Clinical cancer research: an official*

- journal of the American Association for Cancer Research*, vol. 25, no. 19, pp. 5972–5983, 2019.
- [34] J. Long, L. Zhang, X. Wan et al., “A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma,” *Journal of Cellular and Molecular Medicine*, vol. 22, no. 12, pp. 5928–5938, 2018.
- [35] Y. L. Choi, S. H. Park, J. J. Jang, and C. K. Park, “Expression of the G1-S modulators in hepatitis B virus-related hepatocellular carcinoma and dysplastic nodule: association of cyclin D1 and p53 proteins with the progression of hepatocellular carcinoma,” *Journal of Korean Medical Science*, vol. 16, no. 4, pp. 424–432, 2001.
- [36] V. Tripathi, Z. Shen, A. Chakraborty et al., “Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB,” *PLoS Genetics*, vol. 9, no. 3, article e1003368, 2013.
- [37] L. Wang, J. Zhang, L. Wan, X. Zhou, Z. Wang, and W. Wei, “Targeting Cdc20 as a novel cancer therapeutic strategy,” *Pharmacology & Therapeutics*, vol. 151, pp. 141–151, 2015.
- [38] L. Zhuang, Z. Yang, and Z. Meng, “Upregulation of BUB1B, CCNB1, CDC7, CDC20, and MCM3 in tumor tissues predicted worse overall survival and disease-free survival in hepatocellular carcinoma patients,” *BioMed Research International*, vol. 2018, Article ID 7897346, 8 pages, 2018.
- [39] N. Rahman, “Mainstreaming genetic testing of cancer predisposition genes,” *Clinical medicine*, vol. 14, pp. 436–439, 2014.
- [40] D. Yu, B. Green, A. Marrone et al., “Suppression of CYP2C9 by microRNA hsa-miR-128-3p in human liver cells and association with hepatocellular carcinoma,” *Scientific Reports*, vol. 5, no. 1, p. 8534, 2015.
- [41] L. Shuaichen and W. Guangyi, “Bioinformatic analysis reveals CYP2C9 as a potential prognostic marker for HCC and liver cancer cell lines suitable for its mechanism study,” *Cellular and molecular biology*, vol. 64, no. 7, pp. 70–74, 2018.
- [42] L. Wang, J. Yin, X. Wang et al., “C-type lectin-like receptor 2 suppresses AKT signaling and invasive activities of gastric cancer cells by blocking expression of phosphoinositide 3-kinase subunits,” *Gastroenterology*, vol. 150, no. 5, pp. 1183–1195.e16, 2016.
- [43] A. Ossoli, S. Simonelli, C. Vitali, G. Franceschini, and L. Calabresi, “Role of LCAT in atherosclerosis,” *Journal of Atherosclerosis and Thrombosis*, vol. 23, no. 2, pp. 119–127, 2016.
- [44] R. Flores, X. Jin, J. Chang et al., “LCAT, ApoD, and ApoA1 expression and review of cholesterol deposition in the cornea,” *Biomolecules*, vol. 9, no. 12, p. 785, 2019.
- [45] A. Ossoli, E. B. Neufeld, S. G. Thacker et al., “Lipoprotein X causes renal disease in LCAT deficiency,” *PLoS One*, vol. 11, no. 2, article e0150083, 2016.
- [46] M. R. Russell, C. Graham, A. D’Amato et al., “Diagnosis of epithelial ovarian cancer using a combined protein biomarker panel,” *British Journal of Cancer*, vol. 121, no. 6, pp. 483–489, 2019.
- [47] Y. Zheng, Y. Liu, S. Zhao et al., “Large-scale analysis reveals a novel risk score to predict overall survival in hepatocellular carcinoma,” *Cancer Management and Research*, vol. Volume 10, pp. 6079–6096, 2018.

Research Article

Effect of Enhanced Recovery after Surgery with Integrated Traditional Chinese and Western Medicine on Postoperative Stress Response of Patients with Gastrointestinal Tumors

Haiping Zhao  and Wenhui Sun 

Department of General Surgery, The First People's Hospital of Fuyang, Hangzhou, Zhejiang, China

Correspondence should be addressed to Wenhui Sun; sunwenhui_shw@163.com

Received 11 April 2022; Revised 16 June 2022; Accepted 23 June 2022; Published 8 July 2022

Academic Editor: Tao Huang

Copyright © 2022 Haiping Zhao and Wenhui Sun. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. To investigate the effect of enhanced recovery after surgery (ERAS) with integrated traditional Chinese and Western medicine on postoperative stress response of patients suffering from gastrointestinal tumors. **Methods.** A total of 74 patients with gastrointestinal tumors who underwent surgical treatment in our hospital from April 2019 to March 2021 were recruited and randomized into the control group and the observation group (1:1). The control group received routine treatment and care, while the observation group received ERAS plus integrated traditional Chinese and Western medicine. Clinical observation was performed regarding changes in preoperative mood and postoperative pain level in each group. Changes in expression levels of plasma cortisol, C-reactive protein (CRP), interleukin-6 (IL-6), interleukin-8 (IL-8), and tumor necrosis factor- β (TNF- β) before and after surgery were detected in each group. Time of patients' first fart and defecation and complications after surgery in each group were recorded. **Results.** The visual analogue scale (VAS) of patients in the observation group after 12 and 24 h of surgery was significantly lower than that in the control group (12 h: observation group 2.0 (1.00, 3.00) vs. control group 4.00 (3.00, 5.00), $p = 0.001$; 24 h: observation group 2.00 (1.00, 3.00) vs. control group 3.00 (2.00, 5.00), $p = 0.005$). The preoperative anxiety degree of patients in the two groups was not statistically significant ($p = 0.489$). The plasma cortisol level of patients in the observation group after 24 and 48 h of surgery was significantly lower than that in the control group (24 h: observation group 426.54 ± 52.15 nmol/L vs. control group 508.32 ± 41.08 nmol/L, $p = 0.001$; 48 h: observation group 287.19 ± 44.24 nmol/L vs. control group 362.57 ± 43.46 nmol/L, $p = 0.001$). Patients' postoperative CRP, IL-6, IL-8, and TNF- β expression levels in the observation group were remarkably lower than those in the control group at all time points. The first postoperative defecation came earlier in the observation group than that in the control group (observation group 76.00 h (64.50, 87.50) vs. control group 89.00 h (73.50, 116.00), $p = 0.007$). There was 1 postoperative urinary tract infection in the observation group and 1 postoperative intestinal obstruction and 1 incisional wound infection in the control group. **Conclusion.** ERAS with integrated traditional Chinese and Western medicine could effectively reduce the postoperative stress response and inflammatory reaction in patients with gastrointestinal tumors, contributing to the safe and quick recovery of gastrointestinal functions of patients.

1. Introduction

Gastrointestinal tumors are the most common type of malignant tumors in the stomach or intestines and include gastric cancer and colorectal cancer. Gastric cancer is the fifth most prevalent malignancy worldwide and can be divided into Epstein-Barr virus- (EBV-) positive subtype, Microsatellite Instability (MSI) subtype, Genomically Stable (GS) subtype,

and Chromosomal Instability (CIN) subtype, according to the latest research of The Cancer Genome Atlas (TCGA) [1]. Chronic inflammation of the stomach is considered one of the risk factors for gastric cancer and may progress through steps encompassing atrophic gastritis, intestinal metaplasia, and dysplasia [2]. Colorectal cancer is the third major cancer in the digestive system and the fourth leading cause of cancer death in the world [3]. Its heritability is approximately 35%

(95% confidence interval: 10%-48%) [4, 5]. Surgical resection is thought of as the only treatment regimen that can cure gastric cancer and colorectal cancer completely [6, 7]. Tumor resection and clearance of surrounding lymph nodes are the basis of surgical regimens for both gastric cancer and colorectal cancer [8]. Nevertheless, surgical injury triggers new problems like stress response.

Gastrointestinal tumor surgery features large surgical trauma, evident stress response, serious postoperative pain, and great risks for various postoperative complications [9]. All these postoperative traumas and stress responses prolong hospital days and increase treatment cost [9]. Systematic stress responses arising from surgical traumas mainly include neuroendocrine disorder and abnormal metabolisms. A relevant study elaborated that abdominal surgery will cause the abnormal activation of the Hypothalamic-Pituitary-Adrenal (HPA) axis and disrupt the normal circadian rhythm of cortisol secretion [10]. After a few hours of surgery, total plasma cortisol concentration had 4-fold elevation in relation to the normal level and reached peak [10]. Besides, high circulating levels of cortisol may directly result in organ damages [11]. In addition, a study illustrated that some serum cytokines such as interleukin-1 (IL-1), interleukin-6 (IL-6), and tumor necrosis factor- β (TNF- β) are related to abnormal secretion of plasma cortisol, and the expression level of these cytokines elevates after surgery [12]. Reducing stress response which is caused by surgical trauma and severity of neuroendocrine disorder or abnormal metabolisms can shorten patients' recovery after surgery, improve patients' life quality, and decrease the incidence of complications during the perioperative period.

Enhanced recovery after surgery (ERAS) refers to a series of optimized perioperative measures based on evidence-based medicine taken by a team consisting of surgeons, anesthetists, and nurses [13, 14]. This measure is aimed at reducing physical and psychological traumatic stress of patients with surgery and help patients recover rapidly [13, 14]. ERAS can quicken postoperative recovery without increasing complication incidence [15]. A meta-analysis of gastric cancer showed that the C-reactive protein (CRP) level is significantly declined in patients applying ERAS after 3/4 and 7 days of surgery [16]. Furthermore, the IL-6 level is declined after 1 and 3/4 days of surgery, and the TNF- α level is also declined after 3/4 day of surgery [16]. Gustafsson et al. [17] researched 911 patients who received surgery for colon carcinoma. They discovered that risk of 5-year cancer-specific death of patients whose compliance to ERAS intervention $\geq 70\%$ is decreased by 42% compared with other patients. In view of these findings, application of ERAS may help patients shorten the perioperative period and quickly recover.

Traditional Chinese Medicine (TCM) is a repository of experience in improving postoperative recovery effect. Auricular point sticking and TCM plaster are characteristic techniques of TCM and widely used in clinics for convenience and low cost. Owing to the connection of the pain center and auditory center in the cerebral cortex, the auditory center can be stimulated by auricular acupressure, then effectively inhibit the adjacent pain center to relieve pain

[18]. Meanwhile, the mechanism regulates the excitement and inhibition of the cerebral cortex and autonomic nerve center under the cerebral cortex to diminish inflammation, relieve pain, and soothe nerves [19]. Auricular point sticking can improve postoperative gastrointestinal function recovery and is relevant to somatostatin downregulation and motilin elevation [20]. Despite the increasing number of relevant studies, little is reported about the effect of ERAS with integrated TCM and Western medicine on postoperative stress response of patients with gastrointestinal tumors.

This study is aimed at exploring the effect of routine care with ERAS plus integrated TCM and Western medicine on mood change, postoperative pain, gastrointestinal function recovery, and inflammatory reaction indicators of patients with gastrointestinal tumors.

2. Materials and Methods

2.1. Data Sources and Patients' Inclusion Criteria

2.1.1. Data Sources. Patients who underwent surgery for gastrointestinal tumors from April 2019 to March 2021 in First People's Hospital of Fuyang were selected as the study objects. All patients were diagnosed with gastrointestinal tumors by medical iconography and pathology. Patients who met the following criteria were excluded: (i) patients who had received chemotherapy or radiation before surgery; (ii) patients with cardiovascular, liver, kidney, brain, lung, or other serious diseases or patients with basic diseases like hypertension, diabetes, and mental illness; (iii) patients who had severe complications during the perioperative period and transferred to the intensive care unit; and (iv) patients during pregnancy or lactation. A total of 74 patients were finally included in this study and equally divided into the control group and observation group (37 patients for each group) according to different treatment and care regimens during the perioperative period. Patients' detailed baseline characteristics are shown in Table 1. Informed consent of patients was not included because it was a retrospective study. The presented study had been approved by the ethics committee of our hospital.

2.2. Research Methods

2.2.1. Intervention

- (i) *Basic treatment.* Patients in two groups both were given basic treatment. Patients' cardiomotility was monitored by electrocardiogram. Water and electrolyte were given to maintain acid-base balance. Intravenous nutrition was given during the period of fasting. Antibodies were also applied.
- (ii) *Treatment or Care Plan for the Control Group.* Treatment or care plan for the control group includes routine treatment or care only.

Western medicine intervention: (a) surgical plan and risk were explained to patients before surgery. The fasting started 10h before surgery. Water deprivation began 6h before

TABLE 1: Patients' clinical baseline characteristics in each group.

Baseline characteristics	Group		<i>p</i> value
	Control group (37)	Observation group (37)	
Age (years)	46.49 ± 9.57	47.65 ± 10.04	0.612
<i>Gender</i>			0.353
Men (%)	16 (43)	21 (57)	
Women (%)	21 (57)	16 (43)	
Weight index	21.74 ± 1.66	22.08 ± 1.51	0.355
<i>Smoking history</i>			1.000
No (%)	27 (73)	27 (73)	
Yes (%)	10 (27)	10 (27)	
<i>Tumor type</i>			0.245
Gastric cancer (%)	22 (59)	16 (43)	
Colorectal cancer (%)	15 (41)	21 (57)	
<i>Tumor stage</i>			0.350
Stage I (%)	18 (49)	23 (62)	
Stage II (%)	19 (51)	14 (38)	

surgery. Routine bowel preparation was performed. (b) Two drainage tubes were placed during surgery and were removed after 7-9 d of surgery. Intraoperative thermal insulation was not emphasized. (c) Patient-Controlled Intravenous Analgesia (PCIA) and opioid drugs were used after surgery to relieve pain. Patients were allowed to drink water and eat after intestinal ventilation. (d) The catheter was removed after 2 d of surgery. (e) Patients were not guided to get early ambulation.

- (iii) *Treatment or Care Plan in the Observation Group.* The treatment plan in the observation group contained Western medicine intervention and TCM medicine intervention.

Western medicine intervention: (a) ERAS-related concepts, measures, surgical plan, and risk were explained to patients before surgery. (b) 1000 mL 10% glucose solution was given to patients 10 h before surgery. 500 mL 10% glucose solution was given to patients 2 h before surgery. Routine bowel preparation was not performed. (c) A drainage tube was placed during surgery and removed after 3-5 d of surgery. Gastric tubes were not routinely placed in principle (or removed after 24 h of surgery if placed). Thermal insulation was emphasized during surgery. (d) PCIA and nonsteroidal anti-inflammatory drugs (by oral) were used to relieve pain. Opioid drug use was reduced. (e) Water drinking and liquid diet were allowed on the day of surgery. (f) The catheter was removed after 24 h of surgery. (vii) Early ambulation was guided, including performing a semireclining position and turn-over after postoperative awakening from anesthesia, standing near the bed or walking under assistance after 18-24 h of surgery, indoor activities under assistance after 24-36 h of surgery, and walking in the corridor out of the ward and washing by themselves

under assistance after 48 h of surgery. Activity amount and time increased gradually without making patients feel tired.

TCM intervention: (a) TCM plaster was given to patients: 300 g Sargentodoxa, 300 g rheum, 300 g salvia, 200 g fructus, 150 g officinal magnolia bark, and 75 g fructus Evodiae were quickly washed and dried in a drying oven at 50°C low temperature. After mixing, medicines were ground to superfine powder in 100,000 mesh size for 1 h. Then, the medicines were packed as powder bags (8.0 g/bag) by using a powder packing machine in the clean zone. Next, the powder was placed into plasters, fixed, and sealed to make TCM plasters. The TCM plaster was used immediately after surgery by sticking on the center of the abdomen taken navel (Shenque acupoint). The plaster was changed at 4:00 pm every day from the first day after surgery, and this procedure was repeated for 7 d. (b) Auricular acupoint sticking: this was conducted within 0.5 h after patients returned to the ward. That was to say, after being routinely disinfected with 75% alcohol wipes, patients' auricular acupoints (Shenmen, subcortical auricular point, endocrine auricular point) were accurately fixed with wangbuliuxing seed auricular acupoint sticking. The sticking was pressed with the index finger and thumb. Each acupoint was pressed for 1 min a time and repeated every 15 min. The pressing was undertaken on two ears in the meanwhile from gentle to heavy until the patient felt distension, acid, and radiation. This process was performed three times a day for 2 successive days.

2.2.2. Observation Indexes, Time, and Methods

(1) Main Outcome Indexes.

- (a) The visual analogue scale (VAS) is a validated subjective measure of acute and chronic pain to document patients' pain progression or to compare pain severity in patients with similar conditions [21]. Basically, it uses a vernier scale about 10 cm in length. On the one side, there are "0" and "10" at both ends and 10 marks in between. 0 score indicates no pain, and 10 scores represent unbearable severe pain.

VAS was conducted on patients after 3, 6, 12, 24, and 48 h of surgery to evaluate the pain when patients rested and positively coughed.

- (b) Stress response index: plasma cortisol changes were measured before surgery and after 3, 24, and 48 h of surgery. At each time point, venous blood was collected from the antecubital vein to the heparin anticoagulant tube, centrifuged at 3,000 rpm for 5 min, and then preserved at 4°C until use
- (c) Inflammatory factor index: changes in CRP, IL-6, IL-8, and TNF- β levels were measured before surgery and after 3, 24, and 48 h of surgery. At each time, venous blood was collected from the antecubital vein to the heparin anticoagulant tube, centrifuged at 3,000 rpm for 5 min, and then preserved at 4°C until use

(2) Secondary Outcome Indexes.

- (a) Self-rating anxiety scale (SAS): patients' anxiety was determined by SAS designed by Zung [22] at enrollment and when entering the operating room. There were 20 items in SAS (including 5 reverse coded items) with each item having 4 choices scoring from 1 to 4. Scores of the 20 items were added and then multiplied by 1.25 to select the integer part as the standard scores. According to results of the Chinese national norm, a standard score < 50 refers to no anxiety, $50 \leq$ standard score < 60 refers to mild anxiety, $60 \leq$ standard score < 70 refers to moderate anxiety, and standard score ≥ 70 refers to severe anxiety [23]
- (b) Gastrointestinal function recovery: first time to fart and defecate after surgery
- (c) Postoperative complication incidence rate and hospitalization expense

2.2.3. Discharge Criteria and Follow-Up Visit

- (i) Discharge criteria: unified discharge criteria were applied to two groups. Details were shown as follows: (a) complete recovery of oral feeding and no need for intravenous infusion, (b) no drainage or decompression tubes required, (c) simple off-bed activities were available, and (d) patients who were evaluated to be able to be discharged to leave the hospital according to their willingness
- (ii) Follow-up: all patients were followed up for 6 m by outpatient follow-up and telephone follow-up. Complications were recorded

2.3. *Statistical Analysis.* SPSS 23.0 and Prism 8.0 were applied for statistical analysis. Measurement data were shown as the mean \pm standard deviation (SD) or M (P25, P75). Enumeration data were shown as the percentage (%). T test or U test was applied for comparison of measurement data between groups. The Chi-squared test or Fisher's exact test was applied for comparison of enumeration data between groups. $p < 0.05$ indicated that the difference between groups was statistically significant.

3. Results

3.1. *Patients' Clinical Baseline Level.* For the control group, average age was 46.49 ± 9.57 , there were 16 men and 21 women, the weight index was 21.74 ± 1.66 , 10 patients had a smoking history, 22 patients were diagnosed with gastric cancer, 15 patients were diagnosed with colorectal cancer, 18 patients were in I stage, and 19 patients were in II stage.

For the observation group, average age was 47.65 ± 10.04 , there were 21 men and 16 women, the weight index was 22.08 ± 1.51 , 10 patients had a smoking history, 16 patients were diagnosed with gastric cancer, 21 patients were diagnosed with colorectal cancer, 23 patients were in I stage, and 14 patients were in II stage.

Differences in each baseline characteristic (Table 1) between groups had no statistical significance ($p > 0.05$), and the factors were comparable.

3.2. *Main Outcome Indexes.* Patients' postoperative pain degree in two groups was assessed by VAS. The pain degree in the two groups both elevated first and then declined. The pain degree of patients in the observation group was lower than that of the control group after 12 h (2.00 ($1.00, 3.00$) vs. 4.00 ($3.00, 5.00$), $p = 0.001$) and 24 h (2.00 ($1.00, 3.00$) vs. 3.00 ($2.00, 5.00$), $p = 0.005$) of surgery (Figure 1(a)).

Changes in patients' stress response in two groups were evaluated by plasma cortisol concentration. The plasma cortisol concentration of patients in the two groups both reached the peak after 24 h of surgery (control group: 508.32 ± 41.08 nmol/L vs. observation group: 426.54 ± 52.15 nmol/L, $p = 0.001$) and then declined after 24-48 h of surgery (control group: 362.57 ± 43.46 nmol/L vs. observation group: 287.19 ± 44.24 nmol/L, $p = 0.001$) (Figure 1(b)).

These results suggested that the treatment and care of integrating traditional Chinese and Western medicine with ERAS reduced patients' pain levels and plasma cortisol levels in comparison with conventional treatment and care.

Afterwards, the expression level of CRP, IL-6, IL-8, and TNF- β of patients in each group before and after surgery was detected. CRP of patients in two groups continuously elevated after surgery (Figure 2(a)), while the elevation trend in the observation group weakened after 48 h of surgery (control group: 187.35 ± 15.97 mg/L vs. observation group: 171.65 ± 19.67 mg/L, $p = 0.001$). The IL-6 expression level of patients in two groups presented a statistically significant difference after 3 h (control group: 95.16 ± 23.63 pg/L vs. observation group: 78.57 ± 20.97 pg/L, $p = 0.005$), 24 h (control group: 112.03 ± 24.66 pg/L vs. observation group: 94.19 ± 32.57 pg/L, $p = 0.018$), and 48 h (control group: 123.03 ± 28.59 pg/L vs. observation group: 105.78 ± 33.42 pg/L, $p = 0.038$) of surgery (Figure 2(b)). The IL-8 expression level of patients in two groups continuously elevated within 24 h after surgery while declined during 24-48 h after surgery and showed a statistically significant difference after 48 h (control group: 48.76 ± 3.63 pg/L vs. observation group: 41.92 ± 5.83 pg/L, $p = 0.001$) (Figure 2(c)). The TNF- β expression level of patients in two groups started to decline after 3 h of surgery (Figure 2(d)). The declined trend of TNF- β expression level after 24 h (control group: 1994.38 ± 220.75 pg/L vs. observation group: 1634.38 ± 185.40 pg/L, $p = 0.001$) and 48 h (control group: 1452.22 ± 200.29 pg/L vs. observation group: 1230.54 ± 100.13 pg/L, $p = 0.001$) of surgery was more significant in the observation group than the control group. In conclusion, the postoperative CRP, IL-6, IL-8, and TNF- β levels of patients in the observation group were remarkably lower than those in the control group at different time points.

3.3. *Secondary Outcome Indexes.* The patients' psychological anxiety degree at enrollment and when entering the operating room in each group was evaluated by Zung's SAS. Interestingly, it disclosed that patients in the observation group had more significantly severe anxiety at enrollment (control

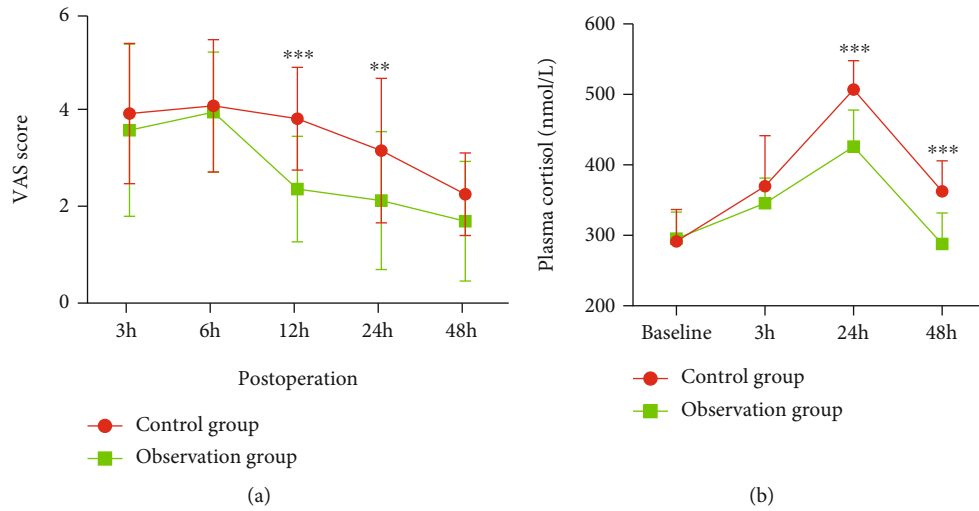


FIGURE 1: Changes in patients' preoperative and postoperative painful degree and plasma cortisol level in each group: (a) VAS scores compared changes in patients' postoperative pain degree in each group; (b) changes in patients' postoperative plasma cortisol level in each group. VAS: visual analogue scale; ** $p < 0.01$, *** $p < 0.001$.

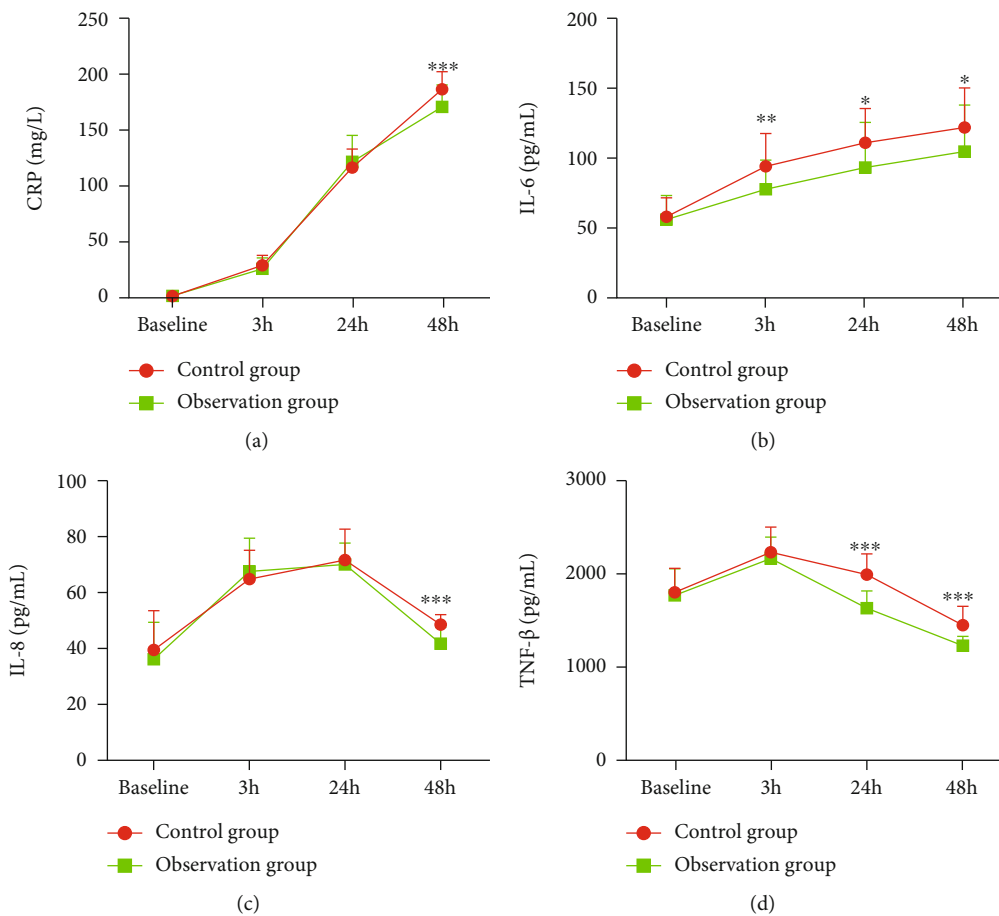


FIGURE 2: Patients' preoperative and postoperative inflammatory reaction-related indexes in each group: (a) patients' preoperative and postoperative CRP expression level in each group; (b) patients' preoperative and postoperative IL-6 expression level in each group; (c) patients' preoperative and postoperative IL-8 expression level in each group; (d) patients' preoperative and postoperative TNF-β expression level in each group. Abbreviations: CRP: C-reactive protein; IL-6: interleukin-6; IL-8: interleukin-8; TNF-β: tumor necrosis factor-β. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

TABLE 2: Secondary outcome indexes.

Secondary indexes	Control group ($N = 37$)	Observation group ($N = 37$)	p value
<i>SAS score</i>			
Enrollment	32.00 (25.00, 44.00)	39.00 (30.00, 48.50)	0.015
Before operation	49.00 (39.50, 53.50)	48.00 (42.00, 55.00)	0.489
Postoperative first fart time (day)	39.00 (28.50, 46.50)	34.00 (30.50, 42.50)	0.372
Postoperative first defecation time (day)	89.00 (73.50, 116.00)	76.00 (64.50, 87.50)	0.007
Hospitalization expense (RMB)	61365 ± 5840	57607 ± 6801	0.013

Abbreviation: SAS: self-rating anxiety scale.

group: 32.00 (25.00, 44.00) vs. observation group: 39.00 (30.00, 48.50), $p = 0.015$). Patients' first postoperative defecation time in the observation group was earlier than that in the control group (control group: 89.00 h (73.50, 116.00) vs. observation group: 76.00 h (64.50, 87.50), $p = 0.007$). Patients' hospitalization expense in the observation group was significantly lower than that in the control group (control group: 61365 ± 5840 RMB vs. observation group: 57607 ± 6801 RMB, $p = 0.013$). In conclusion, the observation of postoperative defecation time and hospitalization costs was superior to the control group. A specific comparison of secondary outcome indexes is shown in Table 2.

3.4. Postoperative Complications. We performed a 6-month follow-up visit on patients in each group and recorded their complications during the period, as shown in Table 3. There was 1 case of postoperative intestinal obstruction and 1 case of postoperative incision infection in the control group. In the observation group, there was 1 case of postoperative urinary tract infection. Postoperative complications were less developed in both groups.

4. Discussion

ERAS has been widely applied in the postoperative recovery of tumor surgery, such as liver resection for hepatocellular carcinoma [24], esophagectomy [25], pulmonary lobectomy for lung cancer [26], gastrectomy for gastric cancer [27], and colorectal cancer surgery [28]. Besides, a study on Japanese patients indicated that ERAS can help colorectal cancer patients rapidly recover, and the postoperative hospitalization time of patients who applied ERAS was evidently shorter than that of patients in the control group [29]. However, the study did not compare the indexes including patients' postoperative stress response and inflammatory factors in two groups; therefore, the results were relatively simple [29]. Another study about the recovery of patients who received gastrectomy and applied ERAS found that patients in the ERAS group recovered normal diet more quickly than patients in the control group [30]. This study combined TCM methods on the basis of ERAS and explored the effect of ERAS with integrated TCM and Western medicine on postoperative stress response of patients with gastrointestinal tumors.

This study assessed patients' postoperative pain degree in two groups through VAS. The pain degree within 6 h after

TABLE 3: Patients' postoperative complications in each group.

Postoperative complications	Control group ($N = 37$)	Observation group ($N = 37$)
Intestinal obstruction	1	0
Abdominal infection	0	0
Incision infection	1	0
Postoperative urinary tract infection	0	1
Total	2	1

surgery both elevated. As time passed, the pain degree of patients in the observation group was significantly weaker than that in the control group. Moreover, the effect of ERAS plus integrated TCM and Western medicine on postoperative stress response was evaluated by the plasma cortisol level of patients in each group. A relevant study revealed that following colorectal carcinoma resection, the patients' plasma cortisol level elevates on d 1 and 5 after surgery, while the level of patients who applied ERAS only elevates on day 5 after surgery, suggesting that ERAS can help patients reduce stress response [31]. Our study demonstrated that the postoperative plasma cortisol level of patients in the observation group who received ERAS with integrated TCM and Western medicine recovered to the normal level more quickly than that in the control group. The inflammatory index of patients in each group was also compared in this study. A relevant study illustrated that ERAS could make inflammatory-related factors recover to a normal level more quickly, and the CRP level was lower than that in the control group [32]. Our study manifested that CRP of patients in two groups within 48 h after surgery continuously elevated, and the elevation slowed then in the observation group. After 48 h of surgery, the CRP level of patients in the observation group was significantly lower than that in the control group, and the difference exhibited a statistical significance. A study detected the IL-6 level of patients in the ERAS group and the control group after colorectal cancer surgery, and it was disclosed that the IL-6 level in the ERAS group after 1, 3, and 5 d of surgery was lower than that in the control group [33]. Similarly, it was discovered here that the postoperative IL-6 level of patients in the observation group and control group continuously elevated,

and overall, the IL-6 expression level of patients in the observation group was lower than that in the control group. Besides, the expression level of IL-8 and TNF- β was also detected. It was indicated that ERAS with integrated TCM and Western medicine could effectively reduce patients' postoperative inflammatory response.

A study elaborated that the first defecation time of patients who underwent duodenectomy and received ERAS was 3 ± 2 days and was shorter than 4 ± 2 days of patients in the control group, exhibiting a statistically significant difference [34]. According to our study, it was discovered that ERAS with integrated TCM and Western medicine could effectively shorten postoperative first defecation time of gastrointestinal cancer patients. The time of patients in the observation group was significantly shorter than the time in the control group. In terms of safety, a relevant study showed that colorectal cancer patients in the control group and the ERAS group were both more likely to have intestinal obstruction [35]. We tracked postoperative complications of patients in each group and found that there was 1 postoperative intestinal obstruction and 1 postoperative incision infection in the control group and 1 postoperative urinary tract infection in the observation group. The overall postoperative complication incidence of patients in the two groups was relatively low, suggesting that ERAS with integrated TCM and Western medicine was safe to improve patients' postoperative recovery.

Taken together, this study investigated the effect of ERAS with integrated TCM and Western medicine on the stress response of patients with gastrointestinal tumors. According to the results of this study, it was believed that ERAS with integrated TCM and Western medicine could effectively and safely decline the postoperative stress response and inflammatory reaction of gastrointestinal tumor patients. Furthermore, patients' gastrointestinal functions could quickly recover.

Data Availability

The data and materials in the current study are available from the corresponding author on reasonable request.

Ethical Approval

The materials and ethics of the article have been approved by the First People's Hospital of Fuyang (2020.021).

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

All authors contributed to data analysis, drafting, and revising the article, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work. All authors consent to submit the manuscript for publication.

Acknowledgments

The study was supported by the Zhejiang TCM Science and Technology Plan (2021ZA24).

References

- [1] T. Seidlitz, Y. T. Chen, H. Uhlemann et al., "Mouse models of human gastric cancer subtypes with stomach-specific CreERT2-mediated pathway alterations," *Gastroenterology*, vol. 157, no. 6, pp. 1599–1614.e2, 2019.
- [2] G. Yu, J. Torres, N. Hu et al., "Molecular characterization of the human stomach microbiota in gastric cancer patients," *Frontiers in Cellular and Infection Microbiology*, vol. 7, p. 302, 2017.
- [3] L. Zhu, Y. X. Tong, X. S. Xu, A. T. Xiao, Y. J. Zhang, and S. Zhang, "High level of unmet needs and anxiety are associated with delayed initiation of adjuvant chemotherapy for colorectal cancer patients," *Support Care Cancer*, vol. 28, no. 11, pp. 5299–5306, 2020.
- [4] N. Song, J. Lee, S. Cho, J. Kim, J. H. Oh, and A. Shin, "Evaluation of gene-environment interactions for colorectal cancer susceptibility loci using case-only and case-control designs," *BMC Cancer*, vol. 19, no. 1, p. 1231, 2019.
- [5] K. J. Monahan, N. Bradshaw, S. Dolwani et al., "Guidelines for the management of hereditary colorectal cancer from the British Society of Gastroenterology (BSG)/Association of Coloproctology of Great Britain and Ireland (ACPGBI)/United Kingdom Cancer Genetics Group (UKCGG)," *Gut*, vol. 69, no. 3, pp. 411–444, 2020.
- [6] C. A. Liu, K. H. Huang, M. H. Chen et al., "Comparison of the surgical outcomes of minimally invasive and open surgery for octogenarian and older compared to younger gastric cancer patients: a retrospective cohort study," *BMC Surgery*, vol. 17, no. 1, p. 68, 2017.
- [7] Z. F. Yang, D. Q. Wu, J. J. Wang, Z. J. Lv, and Y. Li, "Short- and long-term outcomes following laparoscopic vs open surgery for pathological T4 colorectal cancer: 10 years of experience in a single center," *World Journal of Gastroenterology*, vol. 24, no. 1, pp. 76–86, 2018.
- [8] M. Salibasic, S. Pusina, E. Bicakcic et al., "Colorectal cancer surgical treatment, our experience," *Medical Archives*, vol. 73, no. 6, pp. 412–414, 2019.
- [9] V. Manou-Stathopoulou, M. Korbonits, and G. L. Ackland, "Redefining the perioperative stress response: a narrative review," *British Journal of Anaesthesia*, vol. 123, no. 5, pp. 570–583, 2019.
- [10] J. Cerejeira, P. Batista, V. Nogueira, A. Vaz-Serra, and E. B. Mukaetova-Ladinska, "The stress response to surgery and postoperative delirium," *Journal of Geriatric Psychiatry and Neurology*, vol. 26, no. 3, pp. 185–194, 2013.
- [11] E. J. Skovira, E. N. Behrend, L. G. Martin, L. E. Palmer, R. J. Kempainen, and H. P. Lee, "Effect of laparotomy on the pituitary-adrenal axis in dogs," *American Journal of Veterinary Research*, vol. 78, no. 8, pp. 919–925, 2017.
- [12] W. Siekmann, C. Eintrei, A. Magnuson et al., "Surgical and not analgesic technique affects postoperative inflammation following colorectal cancer surgery: a prospective, randomized study," *Colorectal Disease*, vol. 19, no. 6, pp. O186–O195, 2017.
- [13] O. Ljungqvist, M. Scott, and K. C. Fearon, "Enhanced recovery after surgery," *JAMA Surgery*, vol. 152, no. 3, pp. 292–298, 2017.

- [14] L. Bardram, P. Funch-Jensen, P. Jensen, M. E. Crawford, and H. Kehlet, "Recovery after laparoscopic colonic surgery with epidural analgesia, and early oral nutrition and mobilisation," *Lancet (London, England)*, vol. 345, no. 8952, pp. 763-764, 1995.
- [15] J. C. Senturk, G. Kristo, J. Gold, R. Bleday, and E. Whang, "The development of enhanced recovery after surgery across surgical specialties," *Journal of Laparoendoscopic & Advanced Surgical Techniques. Part A*, vol. 27, no. 9, pp. 863-870, 2017.
- [16] I. J. Y. Wee, N. L. Syn, A. Shabbir, G. Kim, and J. B. Y. So, "Enhanced recovery versus conventional care in gastric cancer surgery: a meta-analysis of randomized and non-randomized controlled trials," *Gastric Cancer*, vol. 22, no. 3, pp. 423-434, 2019.
- [17] U. O. Gustafsson, H. Oettelstrup, A. Thorell, J. Nygren, and O. Ljungqvist, "Adherence to the ERAS protocol is associated with 5-year survival after colorectal cancer surgery: a retrospective cohort study," *World Journal of Surgery*, vol. 40, no. 7, pp. 1741-1747, 2016.
- [18] S. Gao, Z. Zhu, and S. Han, "Effect of auricular plaster therapy combined with soothing touch on perioperative stress response of patients undergoing abdominal surgery," *Journal of emergency in traditional chinese medicine*, vol. 3, 2010.
- [19] Z. He, G. Liu, X. Li, and J. Yao, "Clinical observation on the postoperative pain after cesarean section with the intervention of auricular point pressing combined with perioperative psychological nursing," *Shanxi Journal of Traditional Chinese Medicine*, vol. 33, p. 3, 2017.
- [20] X. Wei, H. Qiu, Q. Zhang et al., "Shenhuang powder paste for 110 patients with gastrointestinal dysfunction after abdominal surgery: a prospective multi-center randomized controlled clinical study," *Journal of Traditional Chinese Medicine*, vol. 55, p. 5, 2014.
- [21] D. A. Delgado, B. S. Lambert, N. Boutris et al., "Validation of digital visual analog scale pain scoring with a traditional paper-based visual analog scale in adults," *Journal of the American Academy of Orthopaedic Surgeons. Global research & reviews*, vol. 2, no. 3, article e088, 2018.
- [22] D. A. Dunstan and N. Scott, "Norms for Zung's self-rating anxiety scale," *BMC Psychiatry*, vol. 20, no. 1, p. 90, 2020.
- [23] Y. Du, Y. Cui, X. Cai, Y. Li, and D. Yang, "Analysis of influencing factors of preoperative anxiety or depression in patients with lung cancer surgery," *Zhongguo Fei Ai Za Zhi*, vol. 23, no. 7, pp. 568-572, 2020.
- [24] Q. P. Ren, Y. L. Luo, F. M. Xiao et al., "Effect of enhanced recovery after surgery program on patient-reported outcomes and function recovery in patients undergoing liver resection for hepatocellular carcinoma," *Medicine (Baltimore)*, vol. 99, no. 20, article e20062, 2020.
- [25] K. Kano, T. Aoyama, Y. Maezawa et al., "Postoperative level of C-reactive protein is a prognosticator after esophageal cancer surgery with perioperative steroid therapy and enhanced recovery after surgery care," *In Vivo*, vol. 33, no. 2, pp. 587-594, 2019.
- [26] F. Chen and G. Wang, "Enhanced recovery after surgery for lung cancer patients," *Open Med (Wars)*, vol. 15, no. 1, pp. 198-203, 2020.
- [27] S. Cao, T. Zheng, H. Wang et al., "Enhanced recovery after surgery in elderly gastric cancer patients undergoing laparoscopic total gastrectomy," *The Journal of Surgical Research*, vol. 257, pp. 579-586, 2021.
- [28] U. O. Gustafsson, J. Hausel, A. Thorell et al., "Adherence to the enhanced recovery after surgery protocol and outcomes after colorectal cancer surgery," *Archives of Surgery*, vol. 146, no. 5, pp. 571-577, 2011.
- [29] D. Shida, K. Tagawa, K. Inada et al., "Enhanced recovery after surgery (ERAS) protocols for colorectal cancer in Japan," *BMC Surgery*, vol. 15, no. 1, p. 90, 2015.
- [30] J. Desiderio, C. L. Stewart, V. Sun et al., "Enhanced recovery after surgery for gastric cancer patients improves clinical outcomes at a US cancer center," *Journal of gastric cancer*, vol. 18, no. 3, pp. 230-241, 2018.
- [31] L. Ren, D. Zhu, Y. Wei et al., "Enhanced Recovery After Surgery (ERAS) program attenuates stress and accelerates recovery in patients after radical resection for colorectal cancer: a prospective randomized controlled trial," *World Journal of Surgery*, vol. 36, no. 2, pp. 407-414, 2012.
- [32] H. E. Jaloun, I. K. Lee, M. K. Kim et al., "Influence of the enhanced recovery after surgery protocol on postoperative inflammation and short-term postoperative surgical outcomes after colorectal cancer surgery," *Annals of coloproctology*, vol. 36, no. 4, pp. 264-272, 2020.
- [33] G. Mari, J. Crippa, A. Costanzi, M. Mazzola, M. Rossi, and D. Maggioni, "ERAS protocol reduces IL-6 secretion in colorectal laparoscopic surgery: results from a randomized clinical trial," *Surgical Laparoscopy, Endoscopy & Percutaneous Techniques*, vol. 26, no. 6, pp. 444-448, 2016.
- [34] X. Deng, X. Cheng, Z. Huo et al., "Modified protocol for enhanced recovery after surgery is beneficial for Chinese cancer patients undergoing pancreaticoduodenectomy," *Oncotarget*, vol. 8, no. 29, pp. 47841-47848, 2017.
- [35] D. Shida, K. Tagawa, K. Inada et al., "Modified enhanced recovery after surgery (ERAS) protocols for patients with obstructive colorectal cancer," *BMC Surgery*, vol. 17, no. 1, p. 18, 2017.

Research Article

Influences of Oral Administration of Probiotics on Posthepatectomy Recovery in Patients in Child-Pugh Grade

Hao Huang , Fang Fang , Zheng Jia , Wen Peng , and Yi Wu 

Department of Nursing, Shanghai General Hospital, Shanghai 200080, China

Correspondence should be addressed to Yi Wu; wu85yi@163.com

Received 22 April 2022; Revised 10 June 2022; Accepted 18 June 2022; Published 8 July 2022

Academic Editor: Tao Huang

Copyright © 2022 Hao Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. This study is aimed at investigating the influences of oral administration of probiotics on posthepatectomy recovery in patients in Child-Pugh grade. **Methods.** 100 patients (50 cases in Child-Pugh A grade and 50 cases in Child-Pugh B grade) underwent hepatectomy in our hospital from January 2018 to January 2020 were involved in this study. Subsequently, Child-Pugh A grade and Child-Pugh B grade patients were set as probiotics group (taking *Clostridium butyricum*, $n = 25$) and control group (no probiotics, $n = 25$). The general information, infectious indexes, and liver function indexes on days 1, 3, and 5 after operation were collected. **Results.** In Child-Pugh B grade subgroup patients, the procalcitonin, alanine aminotransferase, and prothrombin time of the probiotics group were statistically significantly lower than that of the control group on days 3 ($P < 0.05$) and 5 ($P < 0.05$) after surgery. In Child-Pugh A grade subgroup patients, there were no significant differences between probiotics group and control group after operation. **Conclusion.** Child-Pugh A grade subgroup patients with hepatectomy could not benefit from oral probiotics. However, Child-Pugh B grade subgroup patients taking probiotics after hepatectomy could reduce postoperative infection and accelerate recovery of liver function.

1. Introduction

Partial hepatectomy is a pivotal method of treating benign and malignant liver tumors that can effectively reduce the mortality of patients and improve their quality of life [1]. But patients are prone to suffering from the abdominal, incisional wound, and urinary system infections after surgery, due to liver dysfunction and body immunity reduction, which increases the pain and affects patients' recovery. Meanwhile, patients are at risk for adverse outcomes such as prolonged hospitalization stays, extra medical costs, and increased mortality [2–4].

Recent attention has focused on the relationship between gut-microbiota-liver axis and the intestinal flora and its pathological significance. Small amount of endotoxin produced by the intestinal flora enters the liver through the portal system and is phagocytosed and cleared by hepatocytes under physiological conditions. And liver dysfunction can cause reduced body immunity, alteration of intestinal flora, and increased endotoxin level, leading to aggravation of hepatocyte damage and formation of a

vicious circle [5, 6]. Clinical research and animal experiments both presented potential clinical application of probiotics in hepatic injury caused by alcoholic hepatitis or fatty liver [7–9]. Probiotic preparations are ecological preparations made from normal microorganisms that are beneficial to the host or their microbial growth-promoting substances, which can effectively adjust intestinal flora alteration, restore the intestinal microecological environment, repress the excessive reproduction of Gram-negative (G-) bacilli, and decrease serum endotoxin level [10, 11]. Furthermore, probiotics have been used prophylactically to prevent bacterial infections after abdominal surgery [12]. Clinical trials also manifest that the bacterial infections after pancreatic surgery, hepatectomy, and liver transplantation can be reduced by prebiotics and probiotics (synbiotics) [13–15].

Preoperative Child grading is a quantitative grading standard for evaluating liver reserve function in patients with liver disease and can indicate the severity of liver damage [16]. Patients in Child-Pugh C grade who underwent partial hepatectomy are implicated in high death rate and

complication rate [17]. Thus, patients with Child-Pugh C grade were not included in this study, although Child-Pugh C grade is now no longer a contraindication to surgical treatment [18]. In this study, we investigated the impact of preoperative oral *Clostridium butyricum* on postoperative infection and recovery of liver function in patients in Child-Pugh A grade and Child-Pugh B grade after partial hepatectomy.

2. Objects and Methods

2.1. Research Objects. 100 patients underwent partial hepatectomy in Shanghai General Hospital from January 2018 to January 2020 were collected, including 80 males and 20 females with a mean age of 50.3 years (25-75 years). 92 patients were diagnosed with primary liver cancer, and 8 patients were diagnosed with hepatic hemangioma preoperatively. According to the preoperative Child classification (see Table 1), 50 cases were in grade A and 50 cases were in grade B.

Patients who met the following inclusion criteria were deemed eligible for inclusion: (1) patient underwent partial hepatectomy; (2) patient did not have infection before surgery; (3) patient is older than 18 years old; (4) patient did not take probiotics in the past 1 week; (5) patient signed the informed consent. Patients who were not conforming to the inclusion criteria were excluded, detailed as follows: (1) patient had chronic constipation or diarrhea; (2) patient had liver function failure due to excessive liver resection, without postoperative complications such as bile leakage and massive bleeding; (3) patient underwent hepatic enterostomy or choledochojejunostomy; (4) patient had poor compliance; (5) patient had mental disorders.

2.2. Grouping and Treatment. Patients in grade A/B were randomly grouped into probiotics group ($n = 25$) and control group ($n = 25$). Patients in the probiotics group received *Clostridium butyricum* (MIYAIRI, Japan) 3×2 tablets daily for three consecutive days before surgery. Prior to surgery, the patients were put on a 12 h fasting and 6 h water fasting. Then, the *Clostridium butyricum* was taken for 4 days after surgery. In the contrast, patients in the control group did not receive probiotics. This research protocol received approval from the Ethics Committee of Shanghai General Hospital, and all patients signed the written informed consent.

2.3. Data Collection. Baseline parameters of patients were as follows: sex, age, body mass index (BMI), presence or absence of liver cirrhosis, serum total bilirubin (TBIL), serum albumin (ALB), coagulation index (prothrombin time (PT)), and surgical method.

Detection indexes on days 1, 3, and 5 after surgery were as follows: fasting blood samples from each patient and healthy subjects were collected. Their infection indexes (procalcitonin, limulus test for endotoxins, and interleukin-6 (IL-6)) and liver function (alanine aminotransferase (ALT) and TBIL) were detected. Procalcitonin and IL-6 were measured by procalcitonin and IL-6 ELISA kit, respectively. Endotoxin levels in patients were detected using a chromogenic LAL assay kit (Lonza, USA) according to the

manufacturer's instructions. Liver function indexes and coagulation indexes were assessed using Roche Cobas 8000 automatic biochemical analyzer (Roche, Germany).

2.4. Statistical Methods. Statistical analysis was processed by GraphPad Prism 5.0 (GraphPad Software Inc., San Diego, CA, USA). Measurement data were presented as mean \pm SD, and a statistical comparison between the two groups was done by means of the *t*-test. Enumeration data were shown as numbers and percentages and were analyzed by Fisher's exact test or Pearson's χ^2 test. Statistical significance was set at $\alpha = 0.05$.

3. Results

3.1. Comparison of Baseline Information of Patients. There were 50 cases both in Child-Pugh A grade and Child-Pugh B grade. In the two subgroups, no statistically significant differences in sex, age, and BMI were observed between the probiotics group and the control group (Table 2). With regard to liver conditions, we found no statistically significant differences in presence or absence of liver cirrhosis, serum TBIL, and serum albumin index between the results. Moreover, there was also an absence of statistically significant differences in PT (a coagulation index indicating liver synthesis function), creatinine (a renal function index), and surgical method.

3.2. Comparison of Infection Indexes of Patients. The following infection indexes were selected: procalcitonin is a specific index of severe bacterial inflammation and fungal infection and is associated with sepsis; limulus test for endotoxins indicates is used to reflect G-bacterial infection; IL-6 is an index that reflects early infection. The specific information of infection indexes is listed in Table 3.

In Child-Pugh A grade subgroup, there were no statistically significant differences in procalcitonin, limulus test for endotoxins, and IL-6 between the probiotics group and the control group on days 1, 3, and 5 after surgery.

In Child-Pugh B grade subgroup, the procalcitonin in the probiotics group and the control group on day 3 after surgery was 0.80 ± 0.96 and 1.98 ± 1.88 , respectively, with a statistically significant difference ($P = 0.007$). The procalcitonin levels in the probiotics group and the control group on day 5 after surgery were 0.29 ± 0.47 and 1.14 ± 1.59 , respectively, with a statistically significant difference ($P = 0.014$). These observations suggested that the postoperatively early bacterial infection rate in the probiotics group was lower than that in the control group.

3.3. Comparison of Liver Function Recovery Indexes in Patients. In the Child-Pugh A grade subgroup, there were no statistically significant differences in ALT, TBIL, and PT between the probiotics group and the control group on days 1, 3, and 5 after surgery.

In the Child-Pugh A grade subgroup, the ALT and the PT of the probiotics group were statistically significantly lower than that of the control group on days 3 and 5 after surgery ($P < 0.05$). Table 4 displays the comparison of liver

TABLE 1: Child-Pugh score.

Parameter	Score		
	1 point	2 points	3 points
Hepatic encephalopathy	None	1-2	3-4
Ascites	None	Mild	Moderate or severe
Serum total bilirubin (TBIL) ($\mu\text{mol/L}$)	<34	34-51	>51
Serum albumin (ALB) (g/L)	≥ 35	28-35	<28
Prothrombin time (PT) (s)	≤ 14	15-17	≥ 17

Grade A = 5-6 points; grade B = 7-9 points; grade C = 10-15 points.

TABLE 2: Baseline information of patients.

	Child-Pugh A grade		<i>P</i> value	Child-Pugh B grade		<i>P</i> value
	Probiotics group (<i>n</i> = 25)	Control group (<i>n</i> = 25)		Probiotics group (<i>n</i> = 25)	Control group (<i>n</i> = 25)	
Sex						
Male	19	21	0.480	20	23	0.221
Female	6	4		5	2	
Age						
≥ 60 years	13	12	0.777	10	12	0.569
<60 years	12	13		15	13	
Body mass index (BMI)	23.34 \pm 0.36	23.53 \pm 0.85	0.309	24.32 \pm 0.74	24.39 \pm 0.96	0.774
Liver cirrhosis						
Yes	5	4	0.713	20	21	0.713
No	20	21		5	4	
Serum TBIL ($\mu\text{mol/L}$)	17.23 \pm 2.34	16.64 \pm 0.56	0.226	23.23 \pm 4.87	24.09 \pm 3.56	0.479
Serum ALB (g/L)	38.34 \pm 0.40	38.50 \pm 0.55	0.245	33.10 \pm 0.75	33.50 \pm 0.83	0.080
PT (s)	11.36 \pm 0.39	11.54 \pm 0.32	0.081	12.20 \pm 0.63	12.54 \pm 0.92	0.134
Serum creatinine ($\mu\text{mol/L}$)	66.21 \pm 2.34	65.21 \pm 4.21	0.053	70.21 \pm 5.09	71.43 \pm 3.02	0.244
Resecting more than 30% of the liver						
Yes	14	11	0.396	10	16	0.089
No	11	14		15	9	

function recovery. The results presented that probiotics helped to promote the recovery of liver function.

4. Discussion

Traditionally, bowel preparation was considered as a pivotal part of preoperative nursing work for partial hepatectomy, which was beneficial to reduce postoperative infection caused by intestinal flora translocation [19]. Although mechanical bowel preparation (MBP) is generally used for partial hepatectomy [20], it has been questioned in recent years. Hokuto et al. [20] compared the perioperative outcomes of patients who received or did not receive MBP before liver resection, and they manifested that the use of MBP has no significant effect on short-term outcomes after liver resection in patients with hepatocellular carcinoma (HCC); therefore, liver surgery can omit MBP. To sum up, MBP could not provide effective benefits to patients before liver surgery, and drinking plenty of water and diarrhea significantly increased the discomfort at preoperative preparation. Therefore, our patients were not

treated with MBP according to current guidelines and research results.

It is well known that HCC is generally the last stage of the development of chronic liver disease, and liver resection is the preferred treatment for most patients [21]. But surgical stress induces dysbiosis, promotes the release of inflammatory cytokines, and increases the permeability of the intestinal barrier, leading to bacterial translocation [21, 22]. The liver is exposed to microbiota and microbial metabolites through portal flow, which ultimately leads to increased infection and poor prognosis due to the limited detoxification function of the liver [21, 23]. The bidirectional relationship between the gut with its microbiota and the liver is known as the “gut-liver axis” [24]. The gut microbiota, as a key factor in this axis, contributes to the progression of liver disease at different stages, thereby promoting the development of HCC and can be used as a noninvasive biomarker for the disease [25–27]. Hence, maintaining intestinal homeostasis is pivotal.

Consideration attention has been given to physiopathological significance of gut microbiota and gut

TABLE 3: Comparison of infection indexes of patients.

	Child-Pugh A grade		<i>P</i> value	Child-Pugh B grade		<i>P</i> value
	Probiotics group (<i>n</i> = 25)	Control group (<i>n</i> = 25)		Probiotics group (<i>n</i> = 25)	Control group (<i>n</i> = 25)	
Procalcitonin (ng/mL)						
Day 1 after surgery	1.94 ± 7.29	1.28 ± 2.08	0.665	4.33 ± 11.77	4.09 ± 11.78	0.943
Day 3 after surgery	1.49 ± 3.72	1.41 ± 2.91	0.933	0.80 ± 0.96	1.98 ± 1.88	0.007
Day 5 after surgery	0.75 ± 1.25	0.51 ± 0.86	0.433	0.29 ± 0.47	1.14 ± 1.59	0.014
Limulus test for endotoxins (Eu/mL)						
Day 1 after surgery	0.11 ± 0.19	0.13 ± 0.21	0.726	0.27 ± 0.45	0.16 ± 0.24	0.286
Day 3 after surgery	0.21 ± 0.43	0.15 ± 0.27	0.534	0.17 ± 0.39	0.22 ± 0.39	0.652
Day 5 after surgery	0.16 ± 0.16	0.15 ± 0.18	0.836	0.22 ± 0.53	0.41 ± 0.78	0.319
Interleukin-6 (pg/mL)						
Day 1 after surgery	153.20 ± 437.00	144.20 ± 165.80	0.924	135.40 ± 116.20	155.30 ± 103.20	0.525
Day 3 after surgery	106.80 ± 173.50	81.24 ± 88.14	0.515	97.52 ± 119.60	94.94 ± 105.30	0.936
Day 5 after surgery	48.25 ± 53.45	46.41 ± 45.38	0.896	34.65 ± 30.77	28.99 ± 16.43	0.421

TABLE 4: Comparison of liver function indexes of patients.

	Child-Pugh A grade		<i>P</i> value	Child-Pugh B grade		<i>P</i> value
	Probiotics group (<i>n</i> = 25)	Control group (<i>n</i> = 25)		Probiotics group (<i>n</i> = 25)	Control group (<i>n</i> = 25)	
Serum ALT (U/L)						
Day 1 after surgery	288.30 ± 196.20	271.30 ± 183.70	0.753	258.80 ± 253.50	254.90 ± 181.80	0.950
Day 3 after surgery	108.50 ± 77.33	113.70 ± 91.00	0.829	115.30 ± 79.56	216.20 ± 184.1	0.015
Day 5 after surgery	69.86 ± 37.49	75.58 ± 48.91	0.645	90.28 ± 61.36	138.80 ± 78.28	0.019
Serum TBIL (μmol/L)						
Day 1 after surgery	23.99 ± 18.33	23.70 ± 13.75	0.950	25.34 ± 29.23	24.80 ± 9.65	0.931
Day 3 after surgery	21.23 ± 19.79	21.78 ± 9.19	0.900	23.75 ± 15.97	23.57 ± 8.89	0.961
Day 5 after surgery	18.92 ± 15.79	19.84 ± 8.36	0.798	20.00 ± 12.50	23.37 ± 11.21	0.321
PT (s)						
Day 1 after surgery	12.82 ± 1.609	13.06 ± 1.416	0.579	12.63 ± 1.28	12.92 ± 1.08	0.391
Day 3 after surgery	12.33 ± 1.369	12.96 ± 1.311	0.103	12.50 ± 1.40	13.63 ± 1.56	0.010
Day 5 after surgery	11.99 ± 2.292	12.82 ± 1.358	0.126	11.38 ± 2.736	12.79 ± 1.116	0.021

microecological balance, which lays a theoretical basis for rational bowel preparation methods [24]. Probiotics, which can prevent bacterial translocation by stabilizing the intestinal barrier and stimulating the proliferation, mucus secretion, and motility of the intestinal epithelium, have been suggested as a treatment for different types of chronic liver injury [28, 29]. In a randomized controlled trial, daily intake of a probiotic preparation (VSL#3) is found to

significantly reduce the risk of hospitalization in patients with hepatic encephalopathy, as well as the Child-Turcotte-Pugh and end-stage liver disease (MELD) scoring model [30]. Alisi et al. [31] also demonstrated that taking a VSL#3 supplement notably improves NAFLD in children. A meta-analysis illustrated that a combination of probiotics and prebiotics given to patients before or on the day of liver transplantation noticeably reduces postoperative infection

rates and antibiotic use compared with patients in the prebiotic-only group [29]. Additionally, using probiotics in animal HCC models can alleviate gut dysbiosis and produce anti-inflammatory mediators to repress tumor growth [32]. Clinically, Rifatbegovic et al. [33] emphasized that cirrhotic HCC patients who took probiotics before and after surgery had faster recovery of liver function and fewer complications. Meanwhile, taking probiotics aids in reducing morbidity and mortality. Furthermore, preoperative application of probiotics to modulate intestinal flora is deemed as a functional bowel preparation strategy, and it has achieved better postoperative recovery results than MBP in patients with colon cancer surgery, pancreaticoduodenectomy, and electroprostatectomy [34–36]. But no study has reported the impact of preoperative probiotic administration on infection and liver function recovery after partial hepatectomy in patients with different Child grades.

Clostridium butyricum is a type of probiotics, whose pharmacological effects are as follows: *Clostridium butyricum* can facilitate the growth of beneficial bacteria in the intestine and repress the reproduction of harmful bacteria. Its metabolites can facilitate the repair and regeneration of intestinal epithelium and restore and maintain the intestinal microecological balance. In addition, its metabolites can also correct intestinal flora alternation and repress the overgrowth of G-bacteria and bacterial translocation, thereby reducing intestinal permeability and restoring intestinal mucosal barrier [37]. One study unveiled that *Clostridium butyricum* could reduce generation and release of endotoxin in acute hepatic failure and relieve IETM, thereby reducing the secondary lesion of liver tissue. Moreover, butyric acid produced by *Clostridium butyricum* can effectively constrain the transfer of nuclear factor kappa B (NF- κ B) into the nucleus while repressing the binding of NF- κ B and DNA, thus restraining gene expression of a series of proinflammatory factors such as TNF- α and exerting effective effects on the treatment of liver disease [38]. Clinical study indicated that intestinal barrier and liver function in patients with liver cirrhosis could be effectively improved by using *Clostridium butyricum* [39]. Hence, *Clostridium butyricum* has been proved to be an effective agent against infection and hepatic injury in patients with chronic liver disease. In this study, we manifested that probiotics markedly reduced calcitonin content in the Child-Pugh B grade subgroup on postoperative day 3 and day 5, indicating that probiotics could reduce early postoperative bacterial infections. In addition, probiotics also noticeably decreased the content of ALT in the Child-Pugh B grade subgroup, while reducing PT, indicating that probiotics can accelerate the recovery of liver function. But in the Child-Pugh A grade subgroup, probiotics had no notable effect on infection markers, liver function markers, or coagulation markers. Patients in the Child-Pugh A grade subgroup may not present obvious intestinal flora disturbance due to mild liver disease, while most of the patients in Child-Pugh B grade subgroup have a long history of liver disease with significant intestinal flora disturbance. Thus, prominent differences exist in patients' response to treatment with probiotic preparations.

This study may be somewhat limited as a single-center study with the small sample size. Further verification by a multicenter study with large sample size is warranted. Additionally, we will perform liver cancer classification based on a multicenter large sample to observe whether probiotics in different types of liver cancer also work best in Child-Pugh B grade patients. Besides, mechanisms and detection of intestinal flora are still underexplored.

To sum up, the value of probiotics in the treatment of liver disease was verified previously. In this research, a subgroup analysis of infection and liver function after partial hepatectomy and oral probiotics in patients with different Child-Pugh grades was carried out. The results illustrated that patients in the Child-Pugh A grade subgroup did not benefit from oral probiotics. But patients in Child-Pugh B grade subgroup had fewer postoperative infections and accelerated recovery of liver function. This study provides a basis for the improvement and individualized application of bowel preparation measures in patients who underwent hepatectomy and unravels the clinical effect of oral probiotics in improving postoperative infection and liver function recovery in those patients.

Data Availability

The data and materials in the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the funds from the 2020 Shanghai “Rising Stars of Medical Talent” Program. This project was supported by the Science and Technology Fund of Shanghai Jiao Tong University School of Medicine (Jyh1903) and Nursing Project of Shanghai General Hospital (02.06.01.19.46).

References

- [1] W. Meilong, Y. Shizhong, and D. Jiahong, “Relationship between neutrophil -lymphocyte ratio and overall survival after liver resection for hepatocellular carcinoma,” *Chinese Journal of Current Advances in General Surgery*, vol. 21, pp. 855–859, 2018.
- [2] J. Yan and L. Li, “Risk factors of postoperative infection in patients underwent laparoscopic resection of hepatic hemangioma,” *Journal of Practical Hepatology*, vol. 20, no. 3, pp. 341–344, 2017.
- [3] G. Longchamp, I. Labgaa, N. Demartines, and G. R. Joliat, “Predictors of complications after liver surgery: a systematic review of the literature,” *HPB*, vol. 23, no. 5, pp. 645–655, 2021.
- [4] L. Hui, Y. Miao, L. Chengke, S. Hexian, and K. Jie, “Analysis of influencing factors for infections after partial hepatectomy in patients with liver cancer,” *Chinese Journal of Nosocomiology*, vol. 27, p. 4, 2017.

- [5] H. Fukui, "Leaky gut and gut-liver axis in liver cirrhosis: clinical studies update," *Gut Liver*, vol. 15, no. 5, pp. 666–676, 2021.
- [6] J. Mu, F. Tan, X. Zhou, and X. Zhao, "Lactobacillus fermentum CQPC06 in naturally fermented pickles prevents non-alcoholic fatty liver disease by stabilizing the gut-liver axis in mice," *Food & Function*, vol. 11, no. 10, pp. 8707–8723, 2020.
- [7] H. Huang, Z. Lin, Y. Zeng, X. Lin, and Y. Zhang, "Probiotic and glutamine treatments attenuate alcoholic liver disease in a rat model," *Experimental and Therapeutic Medicine*, vol. 18, pp. 4733–4739, 2019.
- [8] M. Meroni, M. Longo, and P. Dongiovanni, "The role of probiotics in nonalcoholic fatty liver disease: a new insight into therapeutic strategies," *Nutrients*, vol. 11, no. 11, p. 2642, 2019.
- [9] S. Leclercq, P. de Timary, and P. Starkel, "Targeting the gut microbiota to treat alcoholic liver diseases: evidence and promises," *Acta Gastroenterologica Belgica*, vol. 83, pp. 616–621, 2020.
- [10] N. Haghghat, M. Mohammadshahi, S. Shayanpour, and M. H. Haghghizadeh, "Effects of synbiotics and probiotics supplementation on serum levels of endotoxin, heat shock protein 70 antibodies and inflammatory markers in hemodialysis patients: a randomized double-blinded controlled trial," *Probiotics Antimicrob Proteins*, vol. 12, no. 1, pp. 144–151, 2020.
- [11] J. Ye, L. Lv, W. Wu et al., "Butyrate protects mice against methionine-choline-deficient diet-induced non-alcoholic steatohepatitis by improving gut barrier function, attenuating inflammation and reducing endotoxin levels," *Frontiers in Microbiology*, vol. 9, p. 1967, 2018.
- [12] N. Rayes, T. Pilarski, M. Stockmann, S. Bengmark, P. Neuhaus, and D. Seehofer, "Effect of pre- and probiotics on liver regeneration after resection: a randomised, double-blind pilot study," *Beneficial Microbes*, vol. 3, no. 3, pp. 237–244, 2012.
- [13] N. Rayes, D. Seehofer, S. Hansen et al., "Early enteral supply of lactobacillus and fiber versus selective bowel decontamination: a controlled trial in liver transplant recipients," *Transplantation*, vol. 74, no. 1, pp. 123–128, 2002.
- [14] N. Rayes, D. Seehofer, T. Theruvath et al., "Effect of enteral nutrition and synbiotics on bacterial infection rates after pylorus-preserving pancreaticoduodenectomy: a randomized, double-blind trial," *Annals of Surgery*, vol. 246, no. 1, pp. 36–41, 2007.
- [15] G. Sugawara, M. Nagino, H. Nishio et al., "Perioperative synbiotic treatment to prevent postoperative infectious complications in biliary cancer surgery: a randomized controlled trial," *Annals of Surgery*, vol. 244, no. 5, pp. 706–714, 2006.
- [16] J. Jianqiang, "Correlation of preoperative Child grade, microvascular invasion and prognosis of patients with primary liver cancer after allogeneic liver transplantation," *Chinese Journal of Public Health Engineering*, vol. 19, p. 2, 2020.
- [17] Liver Surgery Group, C. M. A. S. B., "Expert consensus on perioperative management of hepatectomy," *Chinese Journal of Practical Surgery*, vol. 37, pp. 525–530, 2017.
- [18] H. Yuelei, S. Dawei, L. Feixiang, L. Huan, and L. Guoyue, "Influencing factors for microvascular invasion in patients with single hepatocellular carcinoma and their prognosis," *Journal of Clinical Hepatology*, vol. 34, p. 6, 2018.
- [19] T. Orimo, T. Kamiyama, H. Yokoo et al., "Effect of preoperative bowel preparation on surgical site infection in liver surgery," *Gan to Kagaku Ryoho*, vol. 42, no. 12, pp. 1887–1889, 2015.
- [20] D. Hokuto, T. Nomi, I. Yamato, S. Yasuda, S. Obara, and Y. Nakajima, "Impact of mechanical bowel preparation on post-operative outcomes after liver resection for patients with hepatocellular carcinoma: a single-center retrospective cohort study," *Digestive Surgery*, vol. 33, no. 1, pp. 51–57, 2016.
- [21] J. Kahn, G. Pregartner, and P. Schemmer, "Effects of both pro- and synbiotics in liver surgery and transplantation with special focus on the gut-liver axis—a systematic review and meta-analysis," *Nutrients*, vol. 12, no. 8, p. 2461, 2020.
- [22] I. Spadoni, E. Zagato, A. Bertocchi et al., "A gut-vascular barrier controls the systemic dissemination of bacteria," *Science*, vol. 350, no. 6262, pp. 830–834, 2015.
- [23] I. Bartolini, M. Risaliti, R. Tucci et al., "Gut microbiota and immune system in liver cancer: promising therapeutic implication from development to treatment," *World Journal of Gastrointestinal Oncology*, vol. 13, no. 11, pp. 1616–1631, 2021.
- [24] A. Albillos, A. de Gottardi, and M. Rescigno, "The gut-liver axis in liver disease: pathophysiological basis for therapy," *Journal of Hepatology*, vol. 72, no. 3, pp. 558–577, 2020.
- [25] L. X. Yu and R. F. Schwabe, "The gut microbiome and liver cancer: mechanisms and clinical translation," *Nature Reviews. Gastroenterology & Hepatology*, vol. 14, no. 9, pp. 527–539, 2017.
- [26] Z. D. Fu and J. Y. Cui, "Remote sensing between liver and intestine: importance of microbial metabolites," *Current Pharmacology Reports*, vol. 3, no. 3, pp. 101–113, 2017.
- [27] C. Zhang, M. Yang, and A. C. Ericsson, "The potential gut microbiota-mediated treatment options for liver cancer," *Frontiers in Oncology*, vol. 10, article 524205, 2020.
- [28] C. Cesaro, A. Tiso, A. del Prete et al., "Gut microbiota and probiotics in chronic liver diseases," *Digestive and Liver Disease*, vol. 43, no. 6, pp. 431–438, 2011.
- [29] T. Sawas, S. Al Halabi, R. Hernaez, W. D. Carey, and W. K. Cho, "Patients receiving prebiotics and probiotics before liver transplantation develop fewer infections than controls: a systematic review and meta-analysis," *Clinical Gastroenterology and Hepatology*, vol. 13, no. 9, pp. 1567–1574.e3, 2015, e1563; quiz e1143–1564.
- [30] R. K. Dhiman, B. Rana, S. Agrawal et al., "Probiotic VSL3 reduces liver disease severity and hospitalization in patients with cirrhosis: a randomized, controlled trial," *Gastroenterology*, vol. 147, no. 6, pp. 1327–1337.e3, 2014.
- [31] A. Alisi, G. Bedogni, G. Baviera et al., "Randomised clinical trial: the beneficial effects of VSL#3 in obese children with non-alcoholic steatohepatitis," *Alimentary Pharmacology & Therapeutics*, vol. 39, no. 11, pp. 1276–1285, 2014.
- [32] J. Li, C. Y. J. Sung, N. Lee et al., "Probiotics modulated gut microbiota suppresses hepatocellular carcinoma growth in mice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 9, pp. E1306–E1315, 2016.
- [33] Z. Rifatbegovic, D. Mesic, F. Ljuca et al., "Effect of probiotics on liver function after surgery resection for malignancy in the liver cirrhotic," *Medicinski Arhiv*, vol. 64, no. 4, pp. 208–211, 2010.
- [34] Y. Pengfei, Y. Yanan, Y. Lin, and Z. Cuiping, "Effect of probiotics to reduce postoperative infection in patients with colorectal cancer: a meta analysis," *Chinese Journal of Microecology*, vol. 29, pp. 639–642, 2017.

- [35] Y. Guowei, *Efficacy Analysis of Probiotic Preparations in the Perioperative Period of Pancreaticoduodenectomy*, Zhengzhou University, 2018.
- [36] C. Ling and G. Junping, "Evaluation of the application effect of probiotics in intestinal preparation for patients with benign prostatic hyperplasia," *Chinese Nursing Research*, vol. 33, pp. 1997–1999, 2019.
- [37] A. Yasueda, T. Mizushima, R. Nezu et al., "The effect of *Clostridium butyricum* MIYAIRI on the prevention of pouchitis and alteration of the microbiota profile in patients with ulcerative colitis," *Surgery Today*, vol. 46, no. 8, pp. 939–949, 2016.
- [38] J. Liu, Y. Fu, H. Zhang et al., "The hepatoprotective effect of the probiotic *Clostridium butyricum* against carbon tetrachloride-induced acute liver damage in mice," *Food & Function*, vol. 8, no. 11, pp. 4042–4052, 2017.
- [39] G. Yongchun, "Combination of glutamin and *clostridium butyricus* improved intestinal barrier and liver function in patients with liver cirrhosis," *Chinese Journal of Clinical Gastroenterology*, vol. 27, pp. 16–19, 2015.

Research Article

Construction of Lymph Node Metastasis-Related Prognostic Model and Analysis of Immune Infiltration Mode in Lung Adenocarcinoma

Wujin Li , Debin Ou , Jiguang Zhang , and Mingfan Ye 

Thoracic Surgery Department, Fujian Provincial Hospital, Fuzhou 350001, China

Correspondence should be addressed to Mingfan Ye; yemingfan282@163.com

Received 28 April 2022; Accepted 9 June 2022; Published 29 June 2022

Academic Editor: Tao Huang

Copyright © 2022 Wujin Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Lung adenocarcinoma (LUAD) is a major cause for global cancer-related deaths. Research reports demonstrate that lymph node metastasis (LNM) is pertinent to the survival rate of LUAD patients, and crux lies in the lack of biomarkers that could distinguish patients with LNM. We aimed to verify the LNM-related prognostic biomarkers in LUAD. **Methods.** We firstly accessed the expression data of mRNA from The Cancer Genome Atlas (TCGA) database and then obtained samples with LNM (N+) and without LNM (N-). Differential expression analysis was conducted to acquire differentially expressed genes (DEGs). Univariate-LASSO-multivariate Cox regression analyses were performed on DEGs to build a risk model and obtain optimal genes. Afterwards, effectiveness and independence of risk model were assessed based on TCGA-LUAD and GSE31210 datasets. Moreover, a nomogram was established combining clinical factors and riskscores. Nomogram performance was measured by calibration curves. The infiltration abundance of immune cells was scored with CIBERSORT to explore the differences between high- and low-risk groups. Lastly, gene set enrichment analysis (GSEA) was used to investigate differences in immune features between the two risk groups. **Results.** Nine optimal feature genes closely related to LNM in LUAD were identified to construct a risk model. Prognostic ability of the risk model was verified in independent databases. Patients were classified into high- and low-risk groups in accordance with their median riskscores. CIBERSORT score displayed differences in immune cell infiltration like T cells CD4 memory resting between high/low-risk groups. LNM-related genes may also be closely relevant to immune features. Additionally, GSEA indicated that differential genes in the two risk groups were enriched in genes related to immune cells. **Conclusion.** This research built a risk model including nine optimal feature genes, which may be potential biomarkers for LUAD.

1. Introduction

Lung cancer is the leading cause of cancer death globally, with adenocarcinoma as the most prevalent histologic type [1]. Recent decades have witnessed great progress in cancer treatment; however, the prognosis of patients with lung adenocarcinoma (LUAD) still fails to be satisfying [2]. Studies demonstrated that lymph node metastasis (LNM) is pertinent to the survival rate of LUAD patients [3, 4]. As reported, the 5-year survival of LUAD patients with LNM is only 26-53%, while the survival rate of LUAD patients without LNM is more than 95% [5, 6]. Therefore, it is urgent to predict prognostic biomarkers that are related to LNM

occurrence in LUAD and to identify patients with high mortality risk in advance.

The establishment and improvement of many databases further promote our understanding of disease genomic alterations, including identifying biomarkers related to tumor diagnosis and prognosis. For instance, Fu et al. [7] analyzed CEP55 expression in LUAD and lung squamous cell carcinoma (LUSC) by using The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases and validated by receiver operating characteristic (ROC) curves and univariate and multivariate Cox regression analyses. Ma et al. [8] screened survival related key genes by random survival forest based on TCGA-LUAD database, KM

survival curve, and C-index method, and we validated the performance of these genes in GEO database, in which 13 genes as the prognostic biomarkers of LUAD were first reported. Zhang et al. [9] obtained sample cases based on TCGA database, and Cox regression analysis screened the signature genes and constructed the model, which finally yielded the correlation between a 9-gene signaling and glycolysis, and provided a new biomarker for LUAD patient's poor prognosis and metastasis. The above studies presented that abnormally expressed genes can act as prognostic or diagnostic biomarkers of LUAD. Nevertheless, rare was researched on LNM-related differentially expressed genes (DEGs) as prognostic biomarkers of LUAD.

Here, LUAD samples with or without LNM were obtained through TCGA database, which were subjected to differential expression analysis. Enrichment analysis was conducted on the acquired differential genes, and a risk model was built through regression analyses. Besides, immune infiltration abundance of LUAD samples was scored to unravel the correlation between immune infiltration and riskscores. The results may provide an effective prognostic tool for LUAD patients and assist doctors in identifying patients with high risk of mortality to increase their survival rate.

2. Materials and Methods

2.1. Data Source and Preprocessing. Firstly, mRNA data were accessed from TCGA (<https://portal.gdc.cancer.gov/>) database along with clinical data (Supplementary Table 1). Samples followed for more than 30 days were screened as the training set, including 330 non-LNM samples (N-) and 171 LNM samples (N+). LUAD gene expression data (tumor: 226) of GSE31210 (GPL570) were downloaded from GEO along with clinical information (Supplementary Table 2) as the validation set.

2.2. Differential Expression Analysis. This step was aimed at recognizing differentially expressed genes (DEGs) between LNM samples (N+) and non-LNM samples (N-). To that end, "edgeR" package [10] was applied to undertake differential expression analysis on N+ samples relative to N- samples. The screening threshold value was $|\logFC| > 1$ and $FDR < 0.0510$ [10].

2.3. Differential Gene Enrichment Analyses. For better understanding of biological processes that DEGs may participate, "clusterProfiler" package [11] was employed to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses. $padjust < 0.05$ and q value < 0.05 were considered statistically significant [11].

2.4. Construction of the Risk Model. Univariate Cox regression analysis was performed on DEGs with "survival" package ($p < 0.001$) [12]. To avoid overfitting of the statistical model, LASSO regression analysis was conducted on genes screened by univariate analysis by using "glmnet" package [13]. Penalty parameter "lambda" was selected by cross validation method. Genes with strong correlation were removed to reduce model complexity ($maxit = 5000$) [14]. Multivariate Cox regression analysis was undertaken on

LASSO regression analysis-screened genes by using "survival" package [15] to build a risk model and obtain optimal genes. Riskscore was calculated as follows:

$$\text{Riskscore} = \sum_{i=1}^n (\text{Coef}_i \times x_i). \quad (1)$$

In this formula, Coef_i is the cooperativity coefficient and x_i is the relative gene expression standardized by Z-score.

2.5. Model Assessment. We assessed the validity of the model in training set and validation set. Riskscores of samples were calculated according to the formula. Samples were divided into high- and low-risk groups by median score. Survival analysis was undertaken with survminer package (<https://cran.r-project.org/web/packages/survminer/index.html>). High- and low-risk groups were further divided by LNM occurrence to undertake survival analysis. Receiver operator characteristic (ROC) curve was drawn by using "timeROC" package [16] to calculate the area under the curve (AUC) values of 1-, 3-, and 5-year overall survival (OS) [16].

Riskscore was taken as a separate feature in the validation of the independence of the model. Univariate and multivariate Cox regression analyses were further conducted combining clinical data to evaluate the ability of the prognostic risk model to predict patient's survival status.

2.6. Construction and Evaluation of the Nomogram. A nomogram was generated by using "rms" package [17] combining clinical information and the risk model so as to predict the possibility of patient's 1-, 3-, and 5-year OS [17]. Calibration curves corresponded to 1-, 3-, and 5-year were plotted to validate the prediction efficacy of the nomogram.

2.7. Immune Analysis of High- and Low-Risk Groups. CIBERSORT was used to score the abundance of each immune cells in the training set. To increase the accuracy of deconvolution results, we only reserved data with CIBERSORT p value < 0.05 and analyzed differences in the infiltration abundance of immune cells between high- and low-risk groups [18].

Immune function enrichment analysis was conducted in high- and low-risk groups by using the GSEA software. Gene sets used for analysis were $c7$: immunologic signature gene sets. Significant screening criterion was $FDR < 0.25$.

3. Results

3.1. Differential Expression Analysis. Differential analysis was conducted on N+ samples relative to N- samples, and 637 DEGs were obtained including 196 upregulated genes and 441 downregulated ones (Figure 1(a)). Enrichment analyses were undertaken on these 637 DEGs. GO analysis suggested that these genes were mainly enriched in biological functions like chromatin assembly, DNA packaging, protein-DNA complex assembly, and ligand (Figure 1(b)). The result of KEGG manifested that these genes were mainly enriched in biological pathways like retinol metabolism, metabolism of xenobiotics by cytochrome P450, and viral carcinogenesis

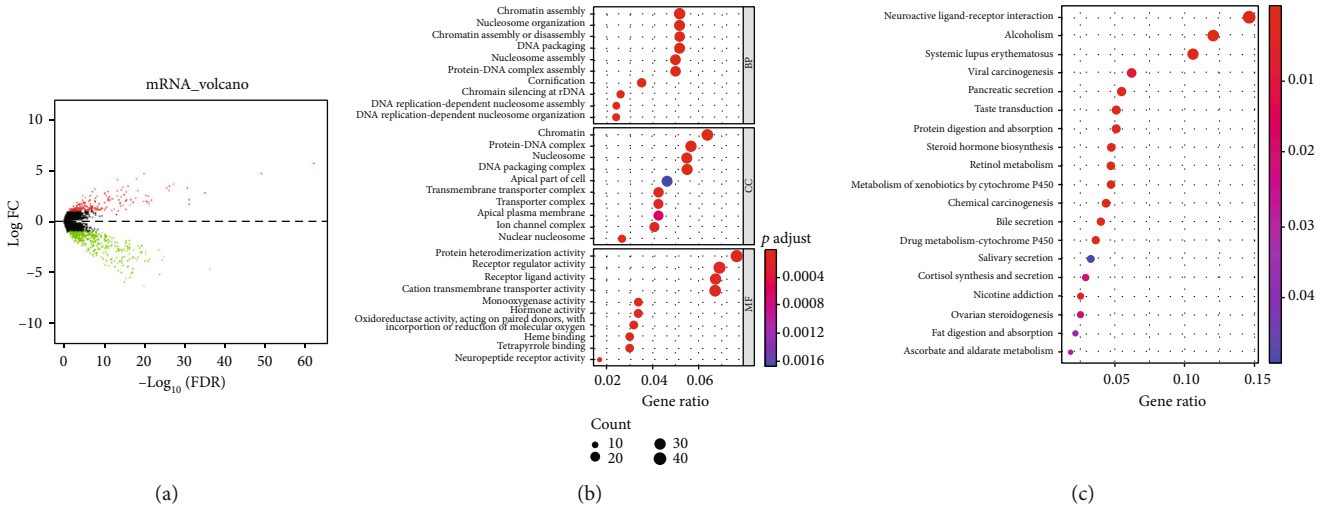


FIGURE 1: Acquisition and analysis of DEGs. (a) Volcano plot of differential analysis on N+ relative to N- group. Red refers to upregulated genes, and green refers to downregulated genes. (b) GO enrichment analysis. (c) KEGG enrichment analysis.

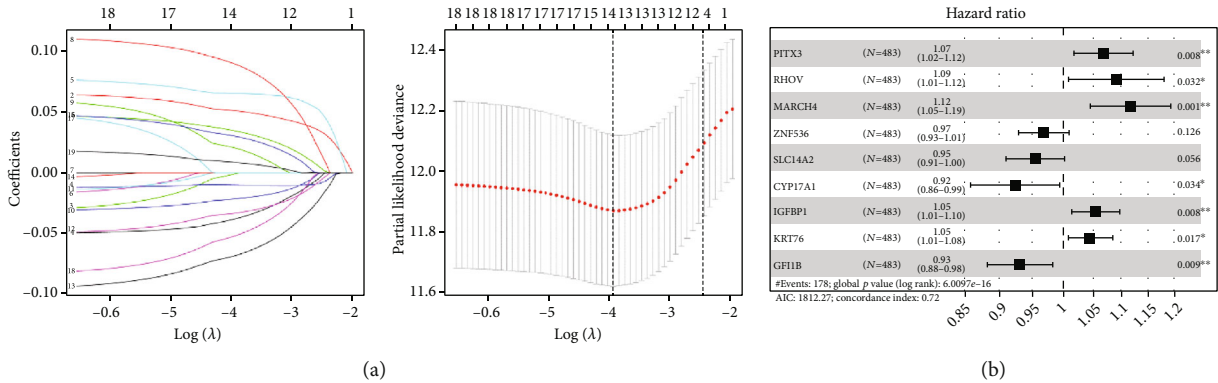


FIGURE 2: Establishment of the risk model. (a) LASSO regression analysis was undertaken on 19 genes obtained. Fourteen feature genes were finally screened. (b) Forest plot of 9 optimal feature genes which obtained from multivariate Cox regression analysis. * $p < 0.05$. ** $p < 0.01$.

(Figure 1(c)). The above results indicated that these DEGs may participate in chromatin assembly, DNA packaging, retinol metabolism, and so on.

3.2. Construction of the Risk Model. Univariate Cox regression analysis was performed on the above 637 DEGs and obtained 19 prognosis-related genes (Supplementary Table 3). To avoid overfitting of the model, LASSO regression analysis of these 19 genes screened 14 feature genes (Figure 2(a)), which were then subjected to multivariate Cox regression analysis, and 9 optimal feature genes were finally screened (Figure 2(b)). Riskscore = $0.066 * \text{PITX3} + 0.087 * \text{RHOV} + 0.111 * \text{MARCH4} - 0.033 * \text{ZNF536} - 0.047 * \text{SLC14A2} - 0.079 * \text{CYP17A1} + 0.053 * \text{IGFBP1} + 0.044 * \text{KRT76} - 0.072 * \text{GFI1B}$. Hence, we obtained 9 optimal feature genes through regression analyses to assess LUAD patient's prognostic risk.

3.3. Assessment of the Risk Model. Afterwards, we assessed validity and independence of the model. According to heat map of the levels of these 9 optimal feature genes in high- and low-risk groups (Figure 3(a)) combined with clinical

features, significant differences were exhibited in stages M and N and survival status between the two risk groups. Patient's survival rate in high-risk group was prominently lower than that in the low-risk group through distribution of the riskscore of 9 optimal feature genes (Figure 3(b)), patient's survival status (Figure 3(c)), and survival curves (Figure 3(d)) in two risk groups. Moreover, the survival rate of N+ patients was lower relative to N- patients in two risk groups (Figures 3(e) and 3(f)). Besides, the performance of the risk model on determining patient's prognosis was evaluated by ROC curve. The 1-, 3-, and 5-year AUC values in training set were, respectively, 0.75, 0.75, and 0.74 (Figure 3(g)), while those in validation set were 0.74, 0.76, and 0.75, respectively (Figure 3(h)). This testified that the risk model based on the training set possessed good determining function on LUAD patient's prognosis. In addition, expression of these 9 optimal feature genes was closely associated with LNM occurrence. Univariate and multivariate Cox regression analyses were conducted on riskscores and clinical data to assess the independence of the model. It was found that riskscores were statistically significant to patient's prognosis (Figures 3(i) and 3(j)). To conclude, the

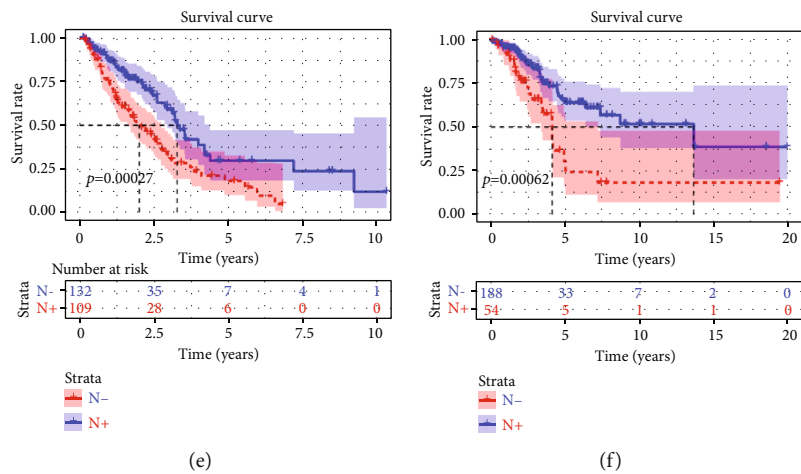
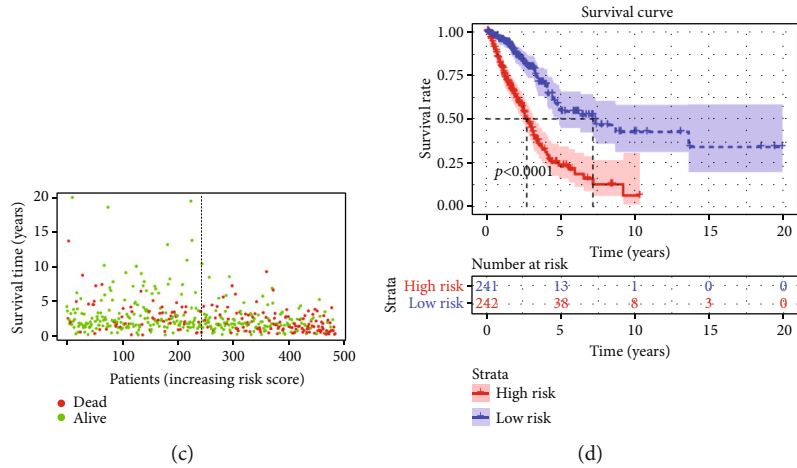
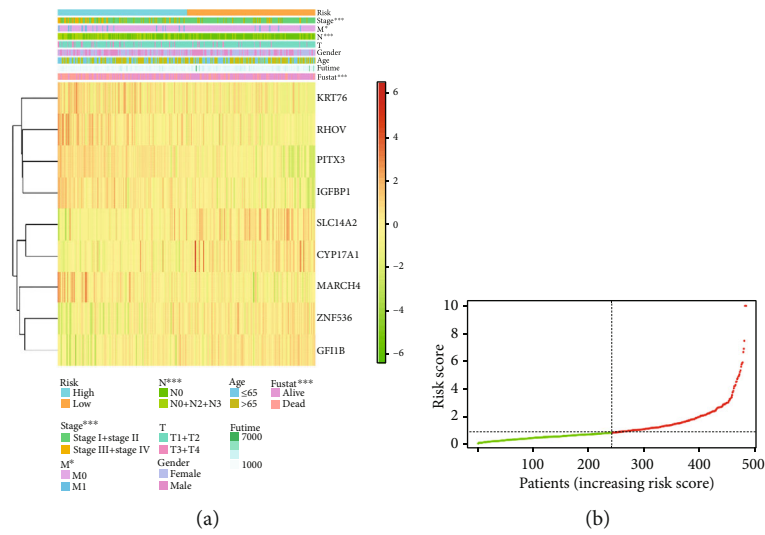


FIGURE 3: Continued.

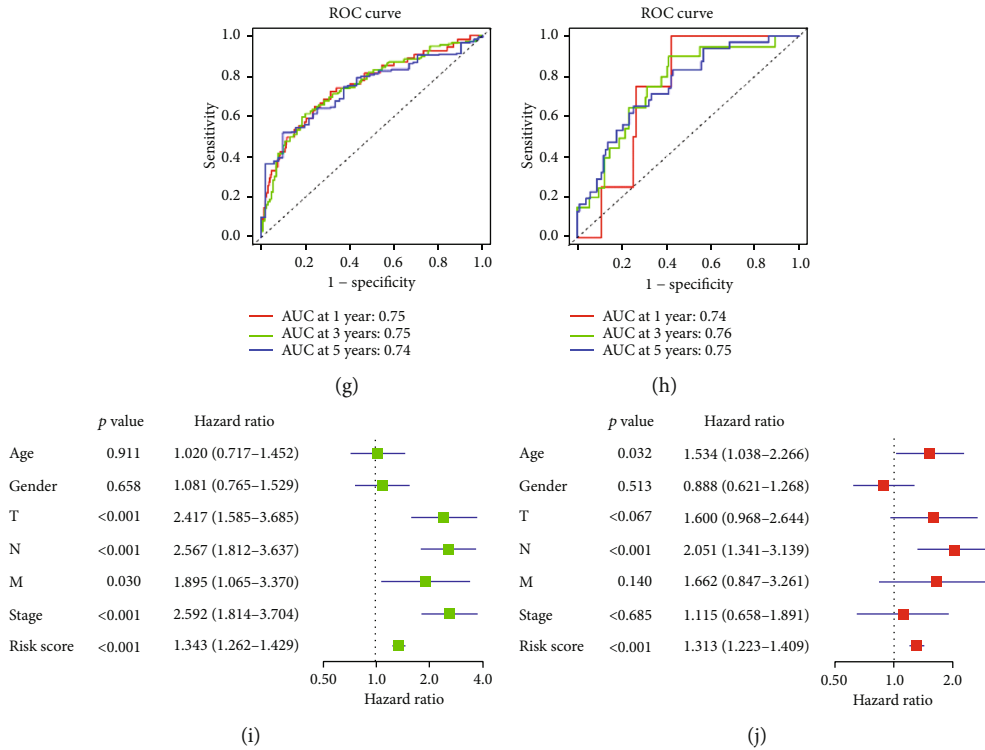


FIGURE 3: Evaluation of the risk model. (a) Heat map of the expression of 9 genes. (b) Distribution plot of riskscore in each group. (c) Distribution plot of survival status in each group. (d) Survival curve graph of patients (red: high-risk group and blue: low-risk group). (e) Survival curve graph of in N+ and N- patients in the high-risk group (red: N+ patients and blue: N- patients). (f) Survival curve graph of patients in N+ group and N- group in the low-risk group (red: N+ patients and blue: N- patients). (g) ROC curve graph of the risk model in TCGA dataset. (h) ROC curve of the risk model in GSE31210 dataset. (i) Forest plot of univariate and (j) multivariate Cox regression analyses.

9-gene risk model suggested a good prognostic effect, and riskscore could be used as an independent factor for the prognosis of LUAD patients.

3.4. Construction and Evaluation of the Nomogram. To better apply our model to clinical practice, we established a nomogram which could better predict LUAD patient’s prognosis. The nomogram was drawn with risk type (low/high) along with patient’s clinical data (age, sex, TNM stage, and clinical stages) to predict the possibility of patient’s 1-, 3-, and 5-year survival (Figure 4(a)). Its performance was visualized by calibration curve (45° line referred to the optimal performance). The high fitting of 1-, 3-, and 5-year calibration curves demonstrated a good performance (Figures 4(b)–4(d)). The above results suggested that the nomogram may assist doctors to decide the plan for following treatment of LUAD patients.

3.5. Riskscore Was Related to Tumor Immune Infiltration. The infiltration abundance of each immune cells in samples was scored by using CIBERSORT algorithm. A total of 184 low-risk samples and 170 high-risk samples were obtained after screening the samples with *p* value<0.05 (Supplementary Table 4). Histogram and heat map based on CIBERSORT exhibited the degree of immunity in the high- and low-risk groups (Figures 5(a) and 5(b)). Infiltration abundance of T cells CD4 memory resting, NK cells activated, dendritic cells

resting, and mast cells resting in the high-risk group was prominently downregulated relative to the low-risk group, while infiltration abundance of T cells CD4 memory activated, macrophages M0, and macrophages M1 was significantly upregulated (Figure 5(c)). The above results showed remarkable differences in immune cell infiltration between two risk groups.

3.6. GSEA Enrichment Analysis in High- and Low-Risk Groups. Lastly, to explore the differences in immune features between the groups, GSEA was performed in these two groups. As the results suggested, differential genes in the high- and low-risk groups were enriched in FETAL_VS_AUDULT_TREG_DN, NAIVE_TCELL_VS_MONOCYTE_UP, and CD16_POS_MONOCYTE_VS_DC_DN (Figures 6(a)–6(c)). It was showed that the two groups manifested statistical significances in immune features of immune cells including regulatory T cells and dendritic cells. These pathways may trigger significant differences in prognosis.

4. Discussion

Accumulating evidence demonstrated that LUAD patients with LNM often have a poor response to standard treatment and shorter survival time [19]. It is urgent to classify these patients in advance and predict their prognosis to help clinicians to better make targeted treatment plans. Nevertheless,

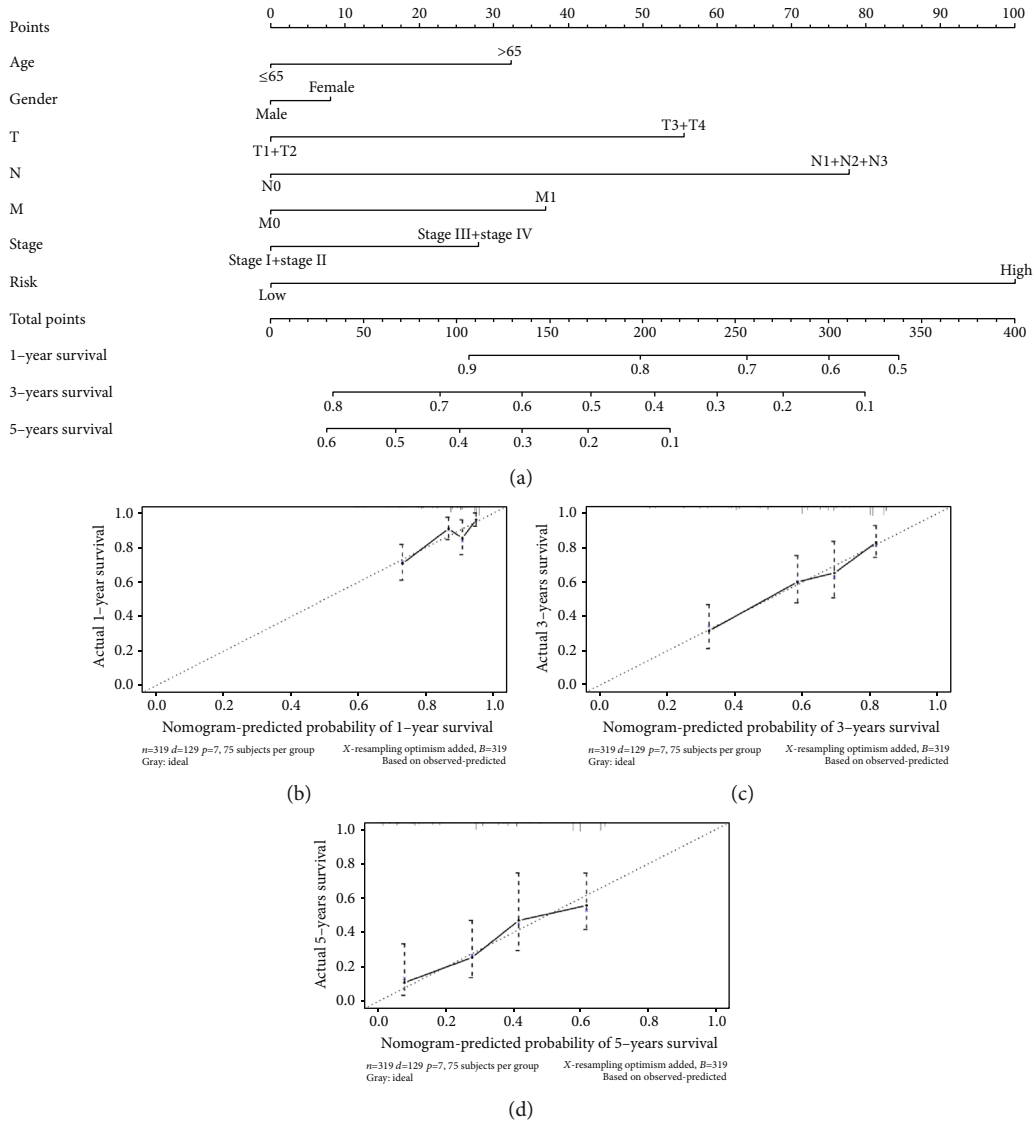


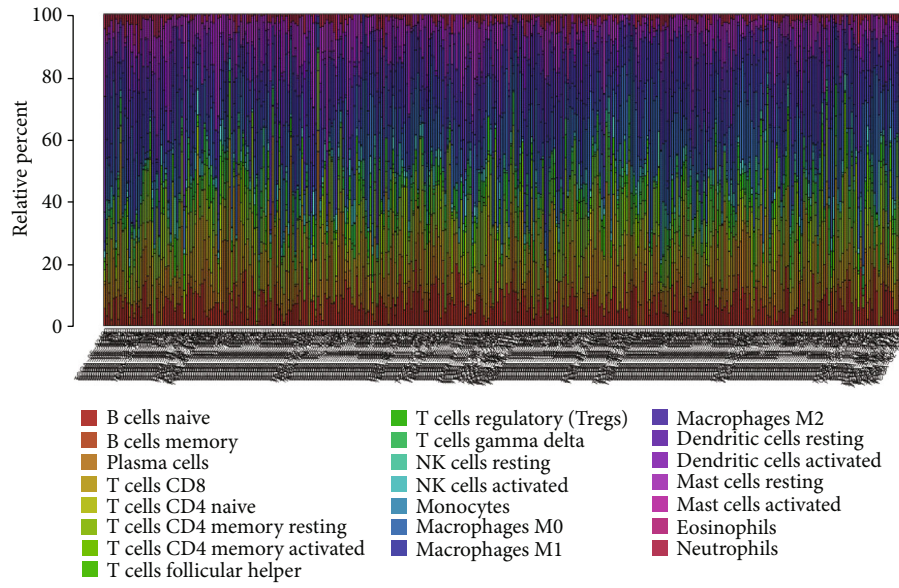
FIGURE 4: Nomogram establishment and evaluation. (a) Nomogram based on the prognostic riskscore of the 9 feature genes combining clinical data. (b)–(d) Calibration curves of 1-, 3-, and 5-year.

a single biomarker cannot accurately or independently evaluate patient’s prognosis and is often affected by other clinical factors [19]. Furthermore, clinical stages established by patient- and tumor-related factors are limited in accuracy and specificity, such as AJCC-TNM stage [20]. A study found that multiple-gene signature is a better choice for predicting patient’s prognosis and survival [21]. Hence, our study was aimed at identifying molecular biomarkers related to LNM in LUAD for better prediction of patient’s prognosis.

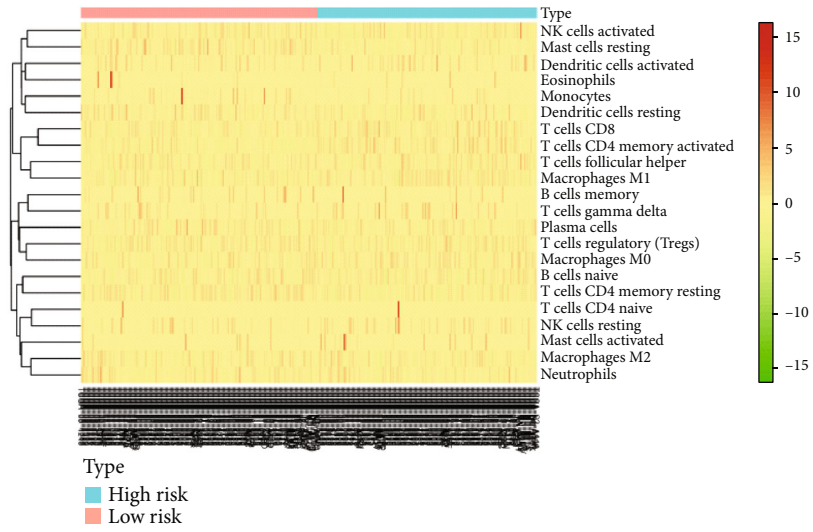
In the present study, DEGs were obtained via analyzing gene expression profiles of N+ and N- samples in TCGA-LUAD dataset. Enrichment analyses suggested that the expression of the DEGs was associated with tumor development. To further screen genes relevant to patient’s prognosis, regression analyses were undertaken. Finally, 9 optimal feature genes were acquired, and a risk model was constructed. $Riskscore = 0.066 * PITX3 + 0.087 * RHOV + 0.111 * MARCH4 - 0.033 * ZNF536 - 0.047 * SLC14A2 - 0.079 * CYP17A1 + 0.053$

$* IGFBP1 + 0.044 * KRT76 - 0.072 * GFI1B$. Previous studies demonstrated that the above genes are relevant to patient’s survival and prognosis. For instance, elevated RHOV expression level correlates with NSCLC patient’s poor survival [22]. High KRT76 expression is associated with increased tumor susceptibility [23]. Subsequently, we evaluated the constructed model and found that the high-risk group showed short survival. N+ patients had shorter survival relative to N- patients. Moreover, ROC curve showed the favorable performance of the model. To help clinicians to predict patients with high mortality risk, we built a nomogram with riskscores and clinical factors. The calibration curve suggested the good performance of the nomogram.

A clinical study illustrated that immune activation in tumor cells is closely relevant to LNM [24]. To this end, we also scored infiltration abundance. It was exhibited that relative to the low-risk group, the infiltration abundance of NK cells activated, T cells CD4 memory resting, dendritic

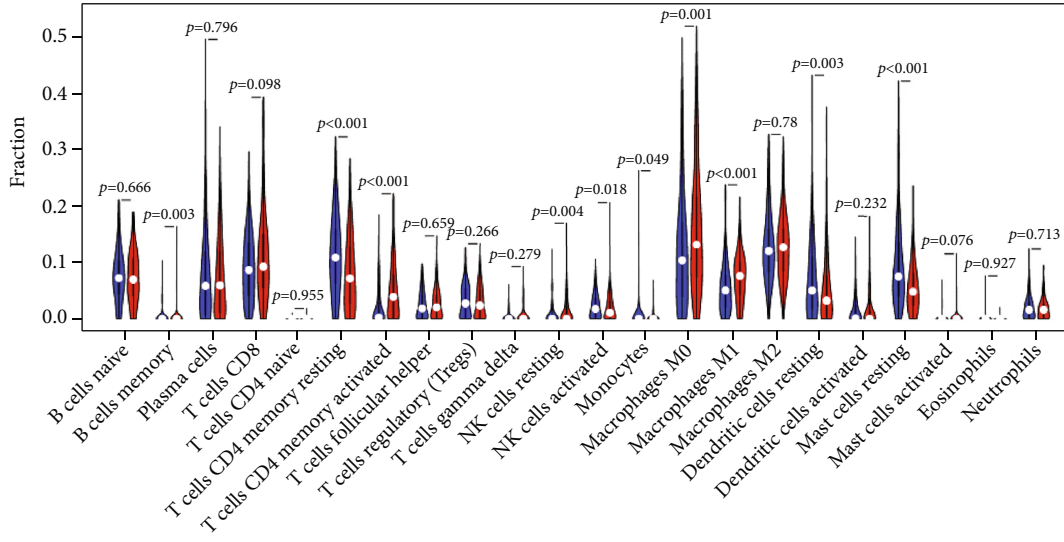


(a)



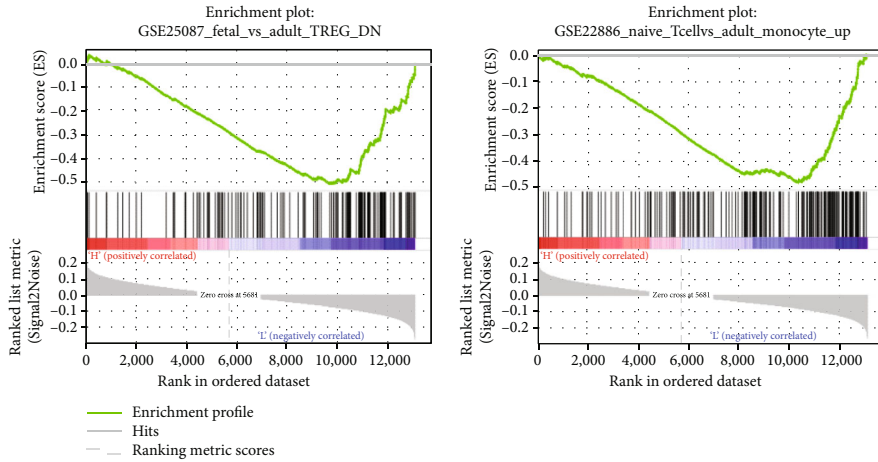
(b)

FIGURE 5: Continued.



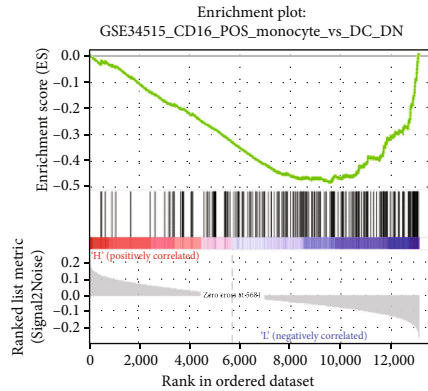
(c)

FIGURE 5: Riskscore was related to tumor immune infiltration. (a) Histogram of infiltration abundance ratio of each immune cells. (b) Heat map of infiltration abundance. (c) Violin plot of differences in each immune cell infiltration abundance (red: high-risk group and blue: low-risk group).



(a)

(b)



(c)

FIGURE 6: GSEA analysis of feature genes. Differential gene enrichment in high- and low-risk groups in (a) FETAL_VS_AUDULT_TREG_DN, (b) NAIVE_TCELL_VS_MONOCYTE_UP, and (c) CD16_POS_MONOCYTE_VS_DC_DN.

cells resting, and mast cells resting was remarkably downregulated, while the infiltration abundance of macrophages M0, T cells CD4 memory activated, and macrophages M1 was prominently upregulated. Melaiu *et al.* [25] presented that high density of tumor infiltrating NK cells is associated with the excellent prognosis of various solid tumors. Mast cells, known as mastocytes, are key regulators of immune effector cells [26–28]. The infiltration of mast cells positively pertains to the prognosis of gastric cancer [26–28]. Padoan *et al.* [29] indicated that memory CD4+ T cells generate interleukin to accelerate tumorigenesis, which explains potential factors for upregulation of activated CD4+ T cells in the high-risk group. Macrophages were considered as a major cell type to connect inflammation and cancer, among which M1 can activate inflammation to stimulate cancer progression [30]. In addition, Xiao *et al.* [31] confirmed that M1 macrophages initiated by exosome-delivered THBS1 exacerbate malignant progression of oral squamous cell carcinoma. Results of the above references were similar to the results here, further demonstrating that optimal feature genes related to LNM were closely associated with immune infiltration degree. GSEA enrichment analysis was conducted to better understand the difference between the two groups. They displayed statistical differences in immune degree of immune cells like regulatory T cells and dendritic cells. A reference elaborated that Treg cells abundantly infiltrate tumor tissue, relating cancer patient's poor prognosis [32]. Haak *et al.* [33] found that the OS of tumor patients with high infiltration of CD16+ cells is evidently longer. Dendritic cells are the main modulators of adaptive immune response and indispensable for T cell-manipulated cancer immunity [34]. This explains that differences in immune features may be one of the reasons for poor prognosis.

In sum, according to TCGA-LUAD data, we constructed an effective 9-gene risk prognostic model that could predict patient's prognosis independent of other clinical factors. Our model was able to divide LUAD patients into two groups and effectively distinguish patients with poor prognosis. Moreover, the identified feature genes may play a predictive role to a certain extent in immune treatment. However, limitations still exist here. In the future, we plan to analyze the expression of optimal feature genes and immune checkpoints to further validate underlying value of the optimal feature genes in predicting the efficacy of treatment with immune checkpoint inhibitors.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Ethical Approval

Not applicable.

Consent

Not applicable.

Conflicts of Interest

The authors declare that they have no potential conflicts of interest.

Authors' Contributions

All authors contributed to data analysis, drafting and revising the article, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work.

Supplementary Materials

Supplementary 1. Supplementary Table 1. Clinical data of TCGA-LUAD samples.

Supplementary 2. Supplementary Table 2. Clinical data of LUAD samples in GSE31210 dataset.

Supplementary 3. Supplementary Table 3. The 19 prognosis-related genes obtained.

Supplementary 4. Supplementary Table 4. Screened samples after CIBERSORT score.

References

- [1] W. Chen, R. Zheng, P. D. Baade *et al.*, "Cancer statistics in China, 2015," *CA: a Cancer Journal for Clinicians*, vol. 66, no. 2, pp. 115–132, 2016.
- [2] D. S. Ettinger, D. E. Wood, W. Akerley *et al.*, "Non-small cell lung cancer, version 6.2015," *Journal of the National Comprehensive Cancer Network*, vol. 13, no. 5, pp. 515–524, 2015.
- [3] C. E. Nwogu, A. Groman, D. Fahey *et al.*, "Number of lymph nodes and metastatic lymph node ratio are associated with survival in lung cancer," *The Annals Of Thoracic Surgery*, vol. 93, pp. 1614–1619, 2012.
- [4] S. H. Ou, J. A. Zell, A. Ziogas, and H. Anton-Culver, "Prognostic factors for survival of stage I nonsmall cell lung cancer patients," *Cancer*, vol. 110, no. 7, pp. 1532–1541, 2007.
- [5] P. Goldstraw, K. Chansky, J. Crowley *et al.*, "The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer," *Journal of Thoracic Oncology*, vol. 11, no. 1, pp. 39–51, 2016.
- [6] A. Matsuda and K. Katanoda, "Five-year relative survival rate of lung cancer in the USA, Europe and Japan," *Japanese Journal of Clinical Oncology*, vol. 43, no. 12, pp. 1287–1288, 2013.
- [7] L. Fu, H. Wang, D. Wei *et al.*, "The value of CEP55 gene as a diagnostic biomarker and independent prognostic factor in LUAD and LUSC," *PLoS One*, vol. 15, no. 5, article e0233283, 2020.
- [8] B. Ma, Y. Geng, F. Meng, G. Yan, and F. Song, "Identification of a sixteen-gene prognostic biomarker for lung adenocarcinoma using a machine learning method," *Journal of Cancer*, vol. 11, no. 5, pp. 1288–1298, 2020.
- [9] L. Zhang, Z. Zhang, and Z. Yu, "Identification of a novel glycolysis-related gene signature for predicting metastasis and survival in patients with lung adenocarcinoma," *Journal of Translational Medicine*, vol. 17, no. 1, p. 423, 2019.
- [10] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: a bioconductor package for differential expression analysis of

- digital gene expression data,” *Bioinformatics*, vol. 26, no. 1, pp. 139–140, 2010.
- [11] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, “clusterProfiler: an R package for comparing biological themes among gene clusters,” *OmicS: A Journal Of Integrative Biology*, vol. 16, no. 5, pp. 284–287, 2012.
- [12] T. M. Therneau and P. M. Grambsch, “Modeling Survival Data: Extending the Cox Model,” *Technometrics*, vol. 44, no. 1, 2013.
- [13] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for Cox’s proportional hazards model via coordinate descent,” *Journal of Statistical Software*, vol. 39, no. 5, pp. 1–13, 2011.
- [14] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
- [15] Z. Wang, J. Zhu, Y. Liu et al., “Development and validation of a novel immune-related prognostic model in hepatocellular carcinoma,” *Journal of Translational Medicine*, vol. 18, no. 1, p. 67, 2020.
- [16] P. Blanche, J. F. Dartigues, and H. Jacqmin-Gadda, “Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks,” *Statistics in Medicine*, vol. 32, no. 30, pp. 5381–5397, 2013.
- [17] C. Huang, Z. Liu, L. Xiao et al., “Clinical significance of serum CA125, CA19-9, CA72-4, and fibrinogen-to-lymphocyte ratio in gastric cancer with peritoneal dissemination,” *Frontiers in Oncology*, vol. 9, p. 1159, 2019.
- [18] A. M. Newman, C. L. Liu, M. R. Green et al., “Robust enumeration of cell subsets from tissue expression profiles,” *Nature Methods*, vol. 12, no. 5, pp. 453–457, 2015.
- [19] H. Kuroda, Y. Sakao, M. Mun et al., “Lymph node metastases and prognosis in left upper division non-small cell lung cancers: the impact of Interlobar lymph node metastasis,” *PLoS One*, vol. 10, no. 8, article e0134674, 2015.
- [20] Y. Wang, Q. Zhang, Z. Gao et al., “A novel 4-gene signature for overall survival prediction in lung adenocarcinoma patients with lymph node metastasis,” *Cancer Cell International*, vol. 19, no. 1, p. 100, 2019.
- [21] Y. L. Chen, G. J. Ge, C. Qi et al., “A five-gene signature may predict sunitinib sensitivity and serve as prognostic biomarkers for renal cell carcinoma,” *Journal of Cellular Physiology*, vol. 233, no. 10, pp. 6649–6660, 2018.
- [22] M. V. Shepelev and I. V. Korobko, “The *_RHOV_* gene is over-expressed in human non -small cell lung cancer,” *Cancer Genetics*, vol. 206, no. 11, pp. 393–397, 2013.
- [23] I. Sequeira and F. M. Watt, “The role of keratins in modulating carcinogenesis via communication with cells of the immune system,” *Cell Stress*, vol. 3, no. 4, pp. 136–138, 2019.
- [24] Z. Xiang, C. Zhong, A. Chang et al., “Immune-related key gene *CLDN10* correlates with lymph node metastasis but predicts favorable prognosis in papillary thyroid carcinoma,” *Aging (Albany NY)*, vol. 12, no. 3, pp. 2825–2839, 2020.
- [25] O. Melaiu, V. Lucarini, L. Cifaldi, and D. Fruci, “Influence of the tumor microenvironment on NK cell function in solid tumors,” *Frontiers in Immunology*, vol. 10, p. 3038, 2019.
- [26] S. A. Oldford and J. S. Marshall, “Mast cells as targets for immunotherapy of solid tumors,” *Molecular Immunology*, vol. 63, no. 1, pp. 113–124, 2015.
- [27] T. Nakano, C. Y. Lai, S. Goto et al., “Immunological and regenerative aspects of hepatic mast cells in liver allograft rejection and tolerance,” *PLoS One*, vol. 7, no. 5, article e37202, 2012.
- [28] M. Wang, Z. Li, Y. Peng et al., “Identification of immune cells and mRNA associated with prognosis of gastric cancer,” *BMC Cancer*, vol. 20, no. 1, p. 206, 2020.
- [29] A. Padoan, M. Plebani, and D. Basso, “Inflammation and pancreatic cancer: focus on metabolism, cytokines, and immunity,” *International Journal of Molecular Sciences*, vol. 20, no. 3, p. 676, 2019.
- [30] E. M. Conway, L. A. Pikor, S. H. Y. Kung et al., “Macrophages, inflammation, and lung cancer,” *American Journal of Respiratory and Critical Care Medicine*, vol. 193, no. 2, pp. 116–130, 2016.
- [31] M. Xiao, J. Zhang, W. Chen, and W. Chen, “M1-like tumor-associated macrophages activated by exosome-transferred THBS1 promote malignant migration in oral squamous cell carcinoma,” *Journal of Experimental & Clinical Cancer Research*, vol. 37, no. 1, p. 143, 2018.
- [32] A. Tanaka and S. Sakaguchi, “Targeting Treg cells in cancer immunotherapy,” *European Journal of Immunology*, vol. 49, no. 8, pp. 1140–1146, 2019.
- [33] F. Haak, I. Obrecht, N. Tosti et al., “Tumor infiltration by OX40+ cells enhances the prognostic significance of CD16+ cell infiltration in colorectal cancer,” *Cancer Control*, vol. 27, no. 1, p. 1073274820903383, 2020.
- [34] A. Gardner and B. Ruffell, “Dendritic cells and cancer immunity,” *Trends in Immunology*, vol. 37, no. 12, pp. 855–865, 2016.

Research Article

Combination of the NRF2 Inhibitor and Autophagy Inhibitor Significantly Inhibited Tumorigenicity of Castration-Resistant Prostate Cancer

Yong Zhang, Zhixiang Xin, Baijun Dong , and Wei Xue 

Department of Urology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China

Correspondence should be addressed to Baijun Dong; dongbaijun@renji.com and Wei Xue; xuewei@renji.com

Received 10 April 2022; Accepted 23 May 2022; Published 20 June 2022

Academic Editor: Tao Huang

Copyright © 2022 Yong Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Prostate cancer (PCa) is the most frequent cancer in men. Developing new treatment methods for CRPC will be a significant challenge in the clinical treatment of PCa. In conclusion, the results of this study show that NRF2 is downregulated in untreated PCa samples compared to normal PCa samples; however, it was upregulated in mCRPC samples compared to HSPC samples. These results demonstrated that NRF2 may serve as a tumor suppressor in tumorigenesis but promote PCa androgen-independent transferring after ADT treatment. Bioinformatics analysis showed that NRF2 was related to multiple signaling, such as the AGE-RAGE pathway, MAPK pathway, NF-kappa B signaling, PI3K-Akt signaling pathway, and VEGF signaling pathway. Moreover, we revealed that the NRF2 inhibitor significantly inhibited tumorigenicity of CRPC cells in vitro. Of note, combination of the NRF2 inhibitor and autophagy inhibitor had a more significantly suppressive role than either ML385 or CQ, indicating that combination of CQ (autophagy inhibitor) and ML385 (NRF2 inhibitor) is a potential treatment of CRPC. Finally, we confirmed that high levels of autophagy regulators LC3B, ULK1, and beclin1 significantly correlated to longer PSA recurrence-free survival time. We think that this study could provide more evidence to confirm that NRF2 is a crucial regulator and targeting NRF2 and autophagy is a potential therapy option for CRPC.

1. Introduction

Prostate cancer (PCa) is the most frequent cancer in men in Western countries [1]. With the increasing popularity of PCa screening in East Asia, particularly China, the incidence of PCa is rising year by year [2]. The number of patients diagnosed with PCa is rising, and it will soon pose a major threat to global public health. For men with PCa, androgen deprivation therapy (ADT) is the mainstay of treatment. However, most PCas become resistant to ADT after prolonged treatment and progress to castration-resistant PCa (CRPC) [3, 4]. The progression of CRPC is also dependent on androgen receptor- (AR-) related signaling pathways. As a result, a second generation of ADT therapy for AR was created. Unfortunately, after second-generation ADT therapy, almost all CRPC patients progress to AR-independent PCa (AIPC), including small cell carcinoma of the prostate (SCC) and double-negative PCa (DNPC)

[5–7]. As a result, developing new treatment methods for CRPC will be a significant challenge in the clinical treatment of PCa.

As a bZIP transcription factor, NRF2, along with small Maf proteins, regulates gene expression via antioxidant response elements (AREs) [8]. ARE is a promoter and enhancer element that responds to ROS, carcinogens, antibiotics, and xenobiotics and regulates basal transcription and antioxidant enzyme induction [9]. Keap1 is a NRF2 binding protein found in the cytoplasm that acts as a negative regulator of NRF2 in vivo. Keap1's ubiquitination, phosphorylation, and nuclear shuttling mechanisms can all influence NRF2 activity [10, 11]. Some studies have found that NRF2 and AR can regulate each other in the prostate, and it is speculated that they, along with AR, may play a critical regulatory role in the formation, development, and treatment of prostate tumors [12, 13]. The crosstalk between AR and NRF2 signaling had been demonstrated in multiple

studies [11, 12]. NRF2, for example, has been demonstrated to reduce AR transactivation by increasing nuclear accumulation of p120-NRF1 [11]. Overexpression of NRF2 was demonstrated to dramatically decrease AR activity generated by dihydrotestosterone (DHT) [11]. As a result, targeted NRF2 inhibition is expected to become a new therapeutic modality for the treatment of PCa, particularly in CRPC patients. A large number of clinical trials on autophagy and tumor therapy, on the other hand, are currently underway, but the results are unsatisfactory. It is hypothesized that when autophagy is inhibited, other pathways may take its place. When autophagy is inhibited, cancer cells can rely on NRF2 signaling to maintain protein levels, according to Andrew Thorburn's research team. Therefore, we propose that combination of CQ (autophagy inhibitor) and ML385 (NRF2 inhibitor) is a potential treatment of CRPC.

The current study explored the correlation between NRF2 expression and clinicopathological information of PCa. In addition, the current research investigated the effect of the NRF2 inhibitor in the growth of CRPC cells. We think that this study could provide more evidence to confirm that NRF2 is a crucial regulator and a potential therapy target of CRPC.

2. Materials and Methods

2.1. Clinical PCa Samples. This study included 472 male patients with PCa who underwent resection or biopsy at Renji Hospital between June 2010 and August 2015. All patients enrolled in this study gave written informed consent to the use of their tissue samples for clinical research purposes. Patients receiving NCHT prior to radical prostatectomy were treated with docetaxel together with goserelin plus bicalutamide. In this study, biochemical recurrence (BCR) was defined as a PSA value of 0.2 ng/ml after RP, confirmed by at least two consecutive measurements.

2.2. Enrichment Analysis. The DEGs between NRF2-high PCa and NRF2-low PCa samples were identified by using the "edgeR" R package [14] with a P value < 0.01 and $|\log_{2}FC| > 1$. Next, KEGG enrichment analysis was carried out with the DAVID system (<https://david.ncicrf.gov/summary.jsp>) [15].

2.3. Cell Culture. Human PCa cell lines, including 22Rv1 and DU145, were obtained from the Chinese Academy of Sciences' typical culture collection center (Shanghai, China) and cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum in a 5% CO₂ atmosphere at 37° C. All reagents were purchased from GIBCO (GIBCO; Thermo Fisher Scientific Inc.)

2.4. Cell Viability Analysis. CellTiter-Glo 3D cell viability assay was conducted to determine cell viability. Cells were plated in a low-attachment 96-well plate at a density of 200 per well with prostate organoid medium with 2% Matrigel. At the third day of plating, treatments were started by addition of medium containing 10 μ M chloroquine (CQ) (Sigma), together with 5 μ M ML385 in five technical replicates per treatment group. Following four days of treatment,

cell viability of the treated tumor cells was assessed using the CellTiter Glo 3D Reagent (Promega) according to the manufacturer's instructions.

2.5. Immunohistochemistry. Immunostaining was performed as previously reported [16]. The proportion of positive cells was calculated in the total number of cells in each field. Areas with strong dyeing intensity are marked as "3," those with medium intensity are marked as "2," and those with low and negative intensity are marked as "1 and 0."

2.6. Statistical Analysis. For statistical analysis, GraphPad prism 9 (GraphPad software, USA) was used. We used the t -test and Tukey's multiple comparison test to assess the differences between experimental variables. The difference is considered as significant when the P value is equal to or less than 0.05.

3. Results

3.1. TCGA Analysis Showed That NRF2 Was Downregulated in Untreated PCa. In order to evaluate the expression levels of NRF2 in PCa, we firstly analyzed TCGA database. The results showed that NRF2 was downregulated in N0 stage and N1 stage PCa samples compared to the normal sample (Figure 1(a)). However, no significant difference of NRF2 expression between N0 and N1 PCa was observed (Figure 1(a)). Moreover, we found that NRF2 was suppressed in stage 1, stage 2, and stage 3 PCa samples compared to normal samples, however, not significantly differently expressed among different stages of PCa (Figure 1(b)).

3.2. NRF2 Protein Levels Was Downregulated in Untreated PCa. Next, we evaluate the protein levels of NRF2 in PCa using human protein atlas database. As present in Figure 1, NRF2 protein levels were significantly lower in high-grade PCa samples compared to low-grade PCa samples (Figure 1(c)).

3.3. TCGA Analysis Showed that NRF2 Low Expression Was Correlated to Poor Outcome in Untreated PCa. Then, the PCa samples were divided into NRF2-high and low groups. As present in Figure 2, we observed that the number of death cases was significantly lower in NRF2-high groups (Figure 2(a)). High NRF2 expression levels were found to be substantially linked with a favorable survival rate; however, reduced NRF2 expression was linked to a worse survival rate in PCa (Figure 2(b)).

3.4. NRF2 Is Suppressed in Untreated PCa Tissues and Is Associated with Disease Progression. To validate the TCGA dataset, we performed immunohistochemical staining assay by using a tissue chip containing 281 hormone-sensitive PCa specimens established by the department of urology, Renji Hospital. By analyzing the correlation between NRF2 protein levels and clinical data of PCa patients, we revealed higher total, nuclear, and cytoplasm protein levels of NRF2 correlated to longer biochemical recurrence-free (BCR-free) survival time and lower protein

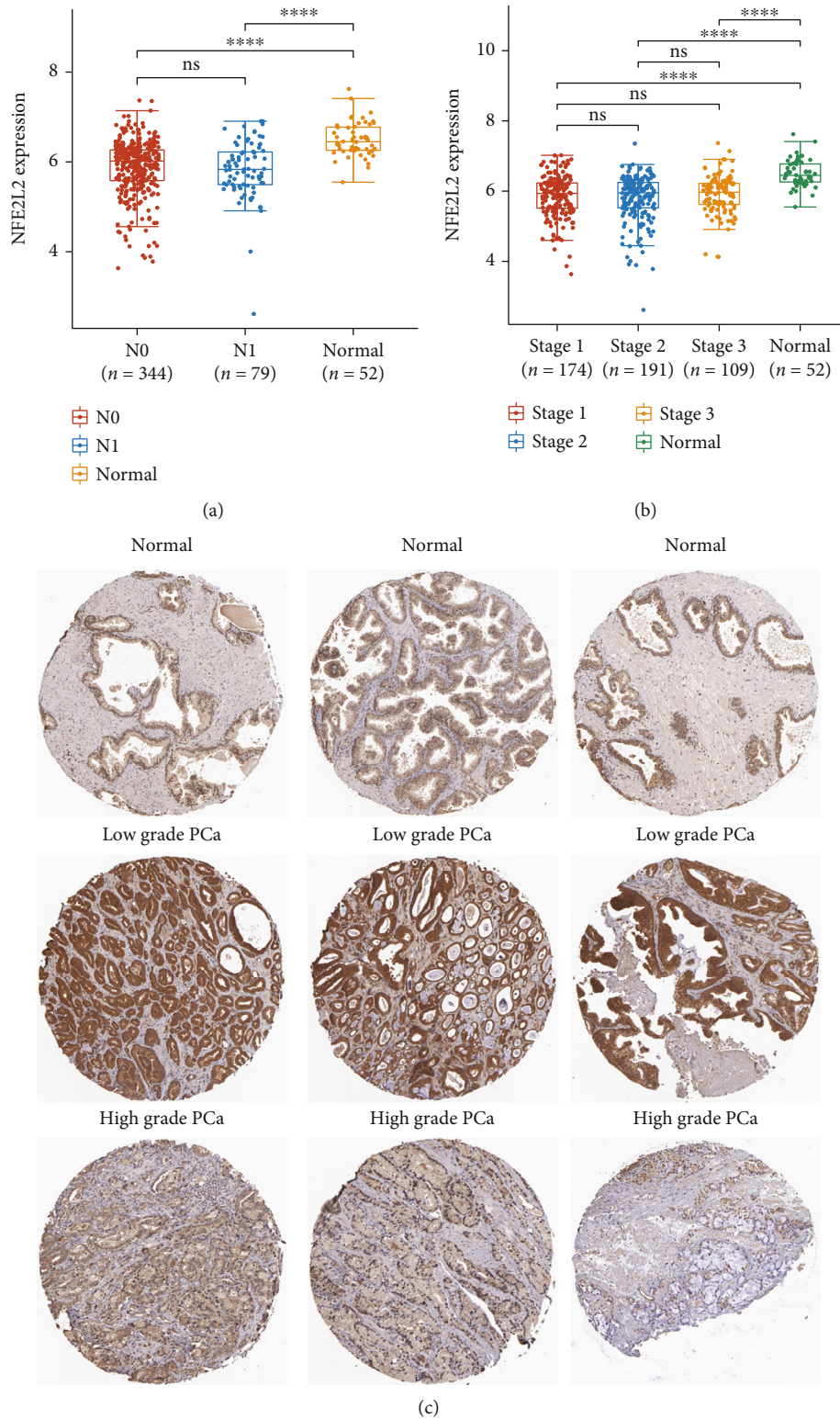


FIGURE 1: NRF2 was downregulated in untreated PCa. (a) The expression levels of NRF2 in normal prostate, N0 stage, and N1 stage PCa samples were analyzed using TCGA database. (b) The expression levels of NRF2 in normal prostate, stage 1, stage 2, and stage 3 PCa samples were analyzed using TCGA database. (c) The protein levels of NRF2 in PCa were analyzed using Human Protein Atlas database.

levels of NRF2 correlated to shorter BCR-free survival time, which was consistent with TCGA data analysis (Figures 3(a)–3(d)). Finally, we analyzed the Tayler dataset

and observed a similar result (Figure 3(e)). Collectively, these results indicated that NRF2 low expression was correlated to poor outcome in PCa.

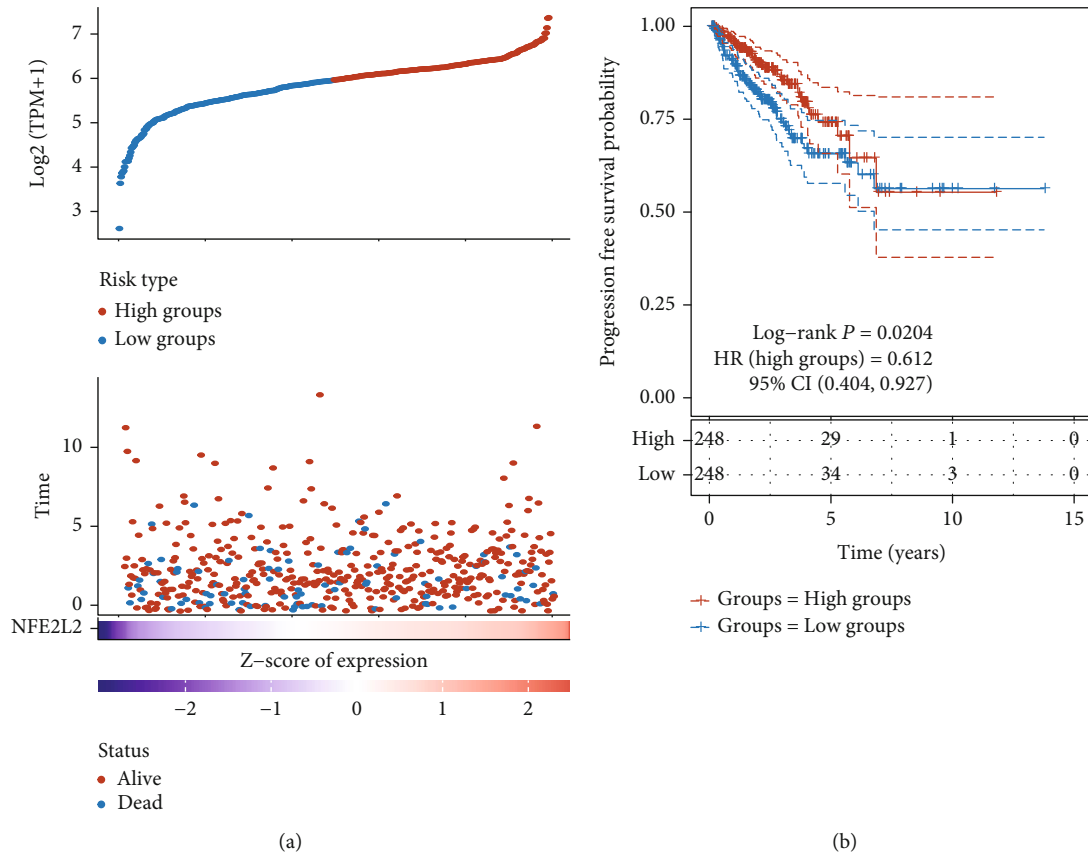


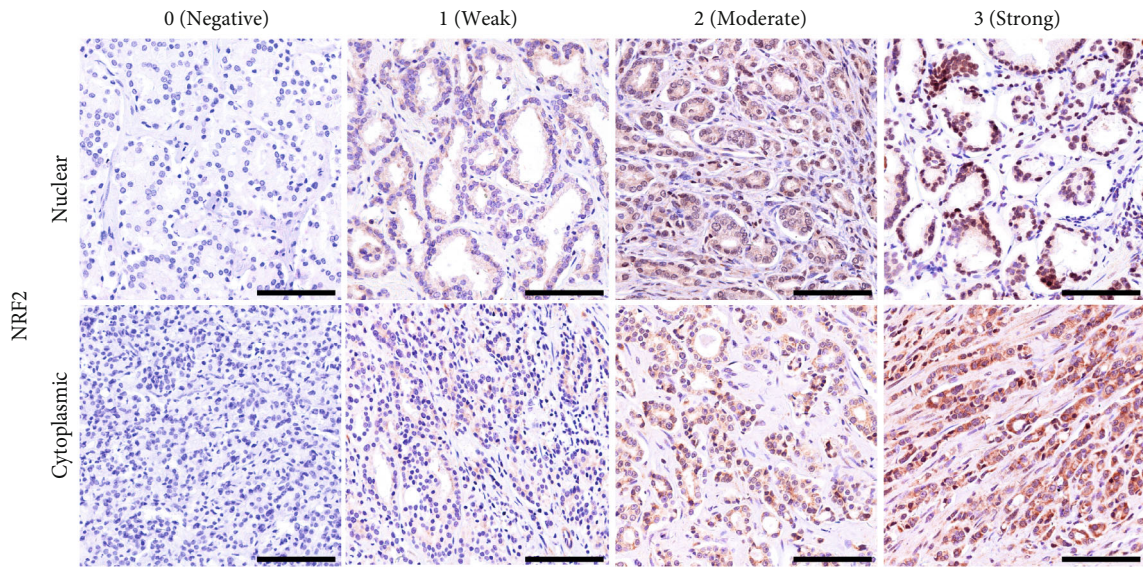
FIGURE 2: NRF2 low expression was correlated to poor outcome in untreated PCa. (a) The number of death cases was significant lower in NRF2-high groups. (b) Kaplan-Meier analysis of the correlation between NRF2 expression levels and progression-free survival time in patients with PCa.

3.5. *NRF2 Is Upregulated in mCRPC Samples Compared to HSPC Samples.* Next, we detected the protein levels of NRF2 between HSPC and mCRPC samples collecting from the same patient using immunohistochemical staining. Very interestingly, we found that NRF2 was highly expressed in mCRPC samples compared to HSPC samples, indicating that NRF2 was activated during PCa androgen-independent transformation (Figure 4(a)). To further confirm these findings, we detected NRF2 protein levels in HSPC and mCRPC samples using a TMA containing 84 neoadjuvant endocrine therapy (NHT) PCa samples, 54 neoadjuvant endocrine + neoadjuvant chemotherapy (NCHT) PCa samples, 281 untreated PCa samples, and 43 CRPC and 10 SCC samples. The results showed that NRF2 levels were higher in 43 CRPC and 10 SCC samples compared to untreated PCa samples. Interestingly, NRF2 levels were significantly suppressed after neoadjuvant endocrine therapy and lower in 84 neoadjuvant endocrine therapy (NHT) PCa samples and 54 neoadjuvant endocrine + neoadjuvant chemotherapy (NCHT) PCa samples compared to 281 untreated PCa samples (Figure 4(b)). These results indicated that NRF2 protein may have a crucial regulatory role in PCa androgen-independent transformation.

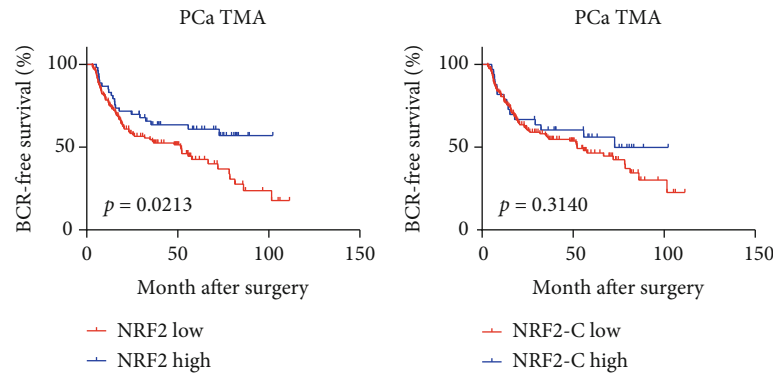
3.6. *Bioinformatics Analysis of NRF2 in Prostate Cancer.* In order to evaluate the mechanism of NRF2 in PCa, we

divided all PCa samples into NRF2-high and NRF2-low groups based on the median levels of NRF2. Then, the differently expressed gene between NRF2-high and low groups was identified. As present in Figure 5, a total of 4317 overexpressed genes and 477 suppressed genes were identified in NRF2-high groups (Figure 6(a)). Among them, NFE2L2, LONRF3, OSMR, RAP1B, and PLEKHA1 were the most significantly upregulated genes and SUMF1, FBXO25, INPP5K, IGHMBP2, and BLOC1S3 were the most significantly downregulated genes. Bioinformatics analysis showed that these upregulated DEGs were significantly related to the AGE-RAGE pathway, chemokine signaling, ECM-receptor interaction, focal adhesion, Hippo signaling, leishmaniasis, MAPK pathway, NF-kappa B signaling, PI3K-Akt signaling pathway, Rap1 pathway, TNF pathway, and Th17 cell differentiation (Figure 6(b)). The downregulated DEGs were significantly related to the neurotrophin signaling pathway, neutrophil extracellular trap formation, pyrimidine metabolism, retrograde endocannabinoid signaling, ribosome, systemic lupus erythematosus, thermogenesis, and VEGF signaling pathway (Figure 6(c)).

3.7. *NRF2 Inhibitor Significantly Inhibited Tumorigenicity of CRPC Cells In Vitro.* We used a nonadherent sphere assay to detect whether NRF2 affected the tumorigenic ability of CRPC cells. In 3D cultures, the effects of NRF2 inhibition

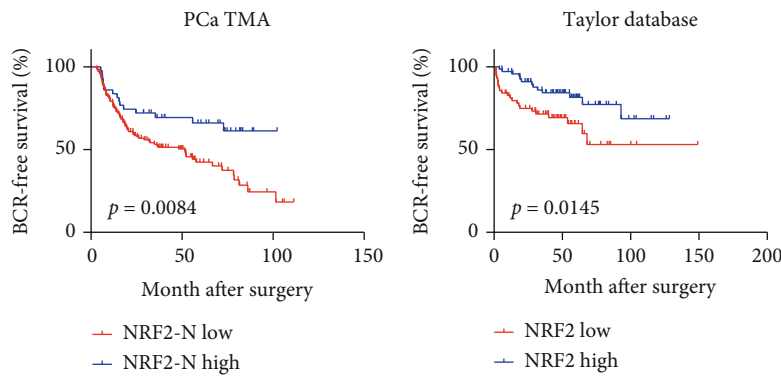


(a)



(b)

(c)



(d)

(e)

FIGURE 3: NRF2 is suppressed in untreated PCa tissues and is associated with disease progression. (a) The representative images of IHC 0, 1, 2, and 3 of nuclear and cytoplasmic levels of NRF2 in PCa. (b) Higher total, (c) nuclear, and (d) cytoplasm protein levels of NRF2 correlated to longer BCR-free survival time. (e) Taylor dataset analysis showed that NRF2 low expression was correlated to poor outcome in PCa.

(ML385) on 22Rv1 and DU145 cell growth were further investigated (Figures 5(a) and 5(c)).

Compared to the control, ML385 significantly suppressed the tumorigenicity of CRPC cells in vitro. The cell viability of DU145 and 22Rv1 cells was suppressed by 79.3% and 44.5%, respectively. The number of

DU145 and 22Rv1 cell spheres was decreased by 36.3% and 38.6%, respectively. The Feret diameter of DU145 and 22Rv1 cell spheres was decreased by 41.3% and 23.5%, respectively. Together, these results indicate that ML385 reduced the tumorigenicity of CRPC cells (Figures 5(a)–5(d)).

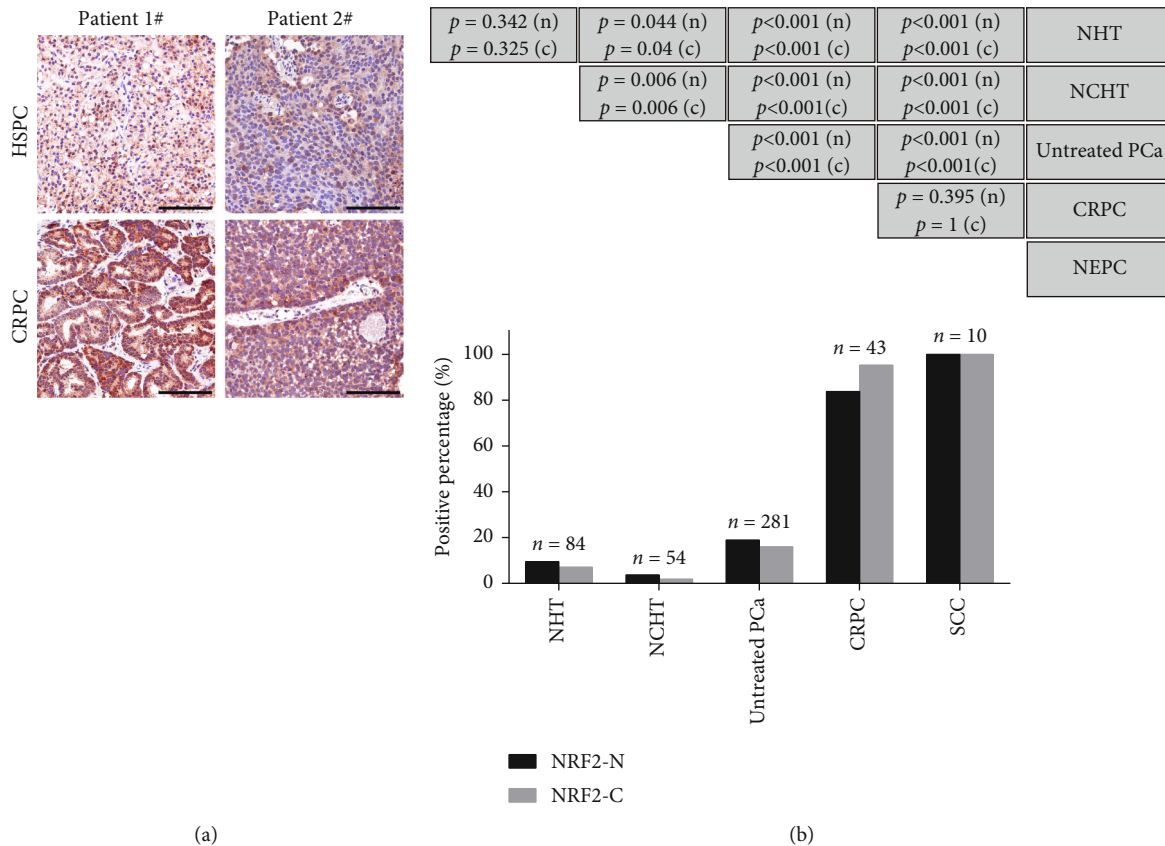


FIGURE 4: NRF2 is upregulated in mCRPC samples compared to HSPC samples. (a) NRF2 was highly expressed in mCRPC samples compared to HSPC samples. (b) The protein levels of NRF2 in a TMA containing 84 neoadjuvant endocrine therapy (NHT) PCa samples, 54 neoadjuvant endocrine + neoadjuvant chemotherapy (NCHT) PCa samples, 281 untreated PCa samples, and 43 CRPC and 10 SCC samples were detected using IHC methods.

3.8. Combination of the NRF2 Inhibitor and Autophagy Inhibitor Significantly Inhibited Tumorigenicity of CRPC Cells. To evaluate whether the combination of the NRF2 inhibitor and autophagy inhibitor could be a potential therapy strategy for CRPC, we treated DU145 and 22Rv1 cells with the NRF2 inhibitor (ML385) and autophagy inhibitor (CQ). As present in Figure 5, the cell viability, sphere number, and organized size of DU145 and 22Rv1 cells were not affected by CQ treatment (Figures 5(a)–5(d)). Meanwhile, the cell viability, sphere number, and organized size of 22Rv1 cells were slightly suppressed after CQ treatment, indicating that treatment with the autophagy inhibitor (CQ) alone did not affect the tumorigenicity of CRPC cells in vitro (Figures 5(a)–5(d)).

As expected, we found that a combination of the NRF2 inhibitor and autophagy inhibitor had a more significantly suppressive role than either ML385 or CQ. The cell viability of DU145 and 22Rv1 cells was suppressed by 93.1% and 76.5%, respectively, after treatment with combination. The number of DU145 and 22Rv1 cell spheres was decreased by 84.5% and 64.5%, respectively, after treatment with combination. The Feret diameter of DU145 and 22Rv1 cell spheres was decreased by 51.3% and 57.5%, respectively, after treatment with combination (Figures 5(a)–5(d)).

3.9. The Dysregulation of Autophagy Markers Was Correlated to the Prognosis of PCa. The abovementioned analysis demonstrated that the combination of the NRF2 inhibitor and autophagy inhibitor significantly suppressed PCa growth. Thus, we detected whether autophagy markers, including LC3B, ULK1, and beclin1, were correlated to PCa prognosis. LC3B, ULK1 and Beclin1 are three important regulators in the regulation of autophagy. Here, we detected whether these autophagy markers were correlated to PCa prognosis. Phosphatidylethanolamine (PE) interconjugates with LC3I in the cytoplasm to generate lipidated forms of LC3II that are embedded in the bilayer membrane. This lipid-bound form of LC3 is often used as a marker for autophagosomes. Beclin1 is phosphorylated by ULK1 and can act as an integral scaffold for the PI3K complex, facilitating the localization of autophagic proteins to autophagic vacuoles. Therefore, we selected these three proteins as markers of autophagy. IHC assay was conducted to analyze LC3B, ULK1, and beclin1 protein levels in prostate tissue specimens. Figure 7 showed the representative staining of LC3B, ULK1, and beclin1 protein in TMAs. As present in Figure 7(b), we observed that high levels of LC3B, ULK1, and beclin1 significantly correlated to longer BCR-free survival time.

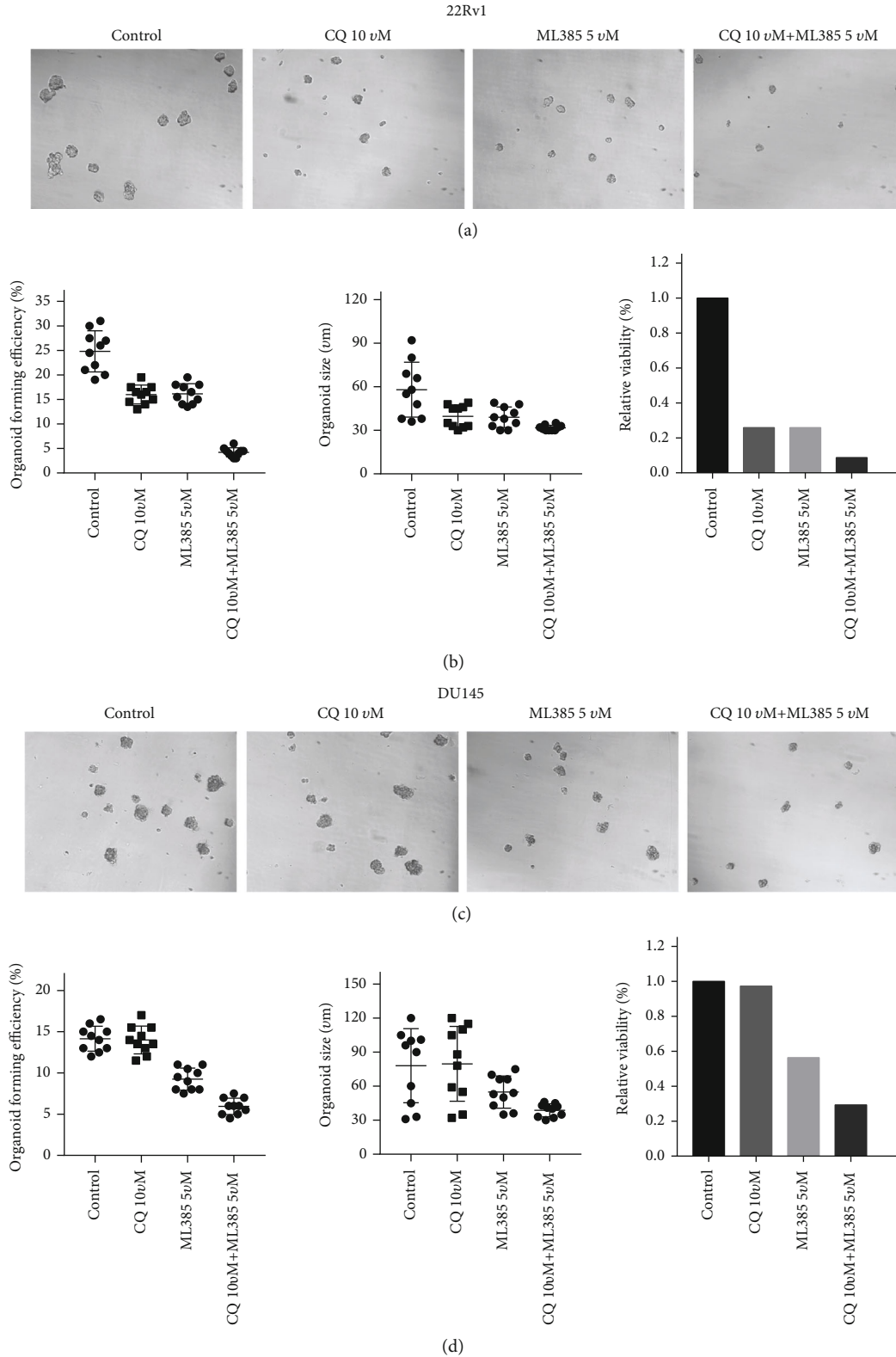


FIGURE 5: Combination of the NRF2 inhibitor and autophagy inhibitor significantly inhibited tumorigenicity of PCa cells. (a) The representative images of PCa 22Rv1 spheres after treatment with the NRF2 inhibitor, autophagy inhibitor, and combination. (b) The cell viability, sphere number, and organized size of PCa 22Rv1 spheres after treatment with the NRF2 inhibitor, autophagy inhibitor, and combination. (c) The representative images of PCa DU145 spheres after treatment with the NRF2 inhibitor, autophagy inhibitor, and combination. (d) The cell viability, sphere number, and organized size of PCa DU145 spheres after treatment with the NRF2 inhibitor, autophagy inhibitor, and combination.

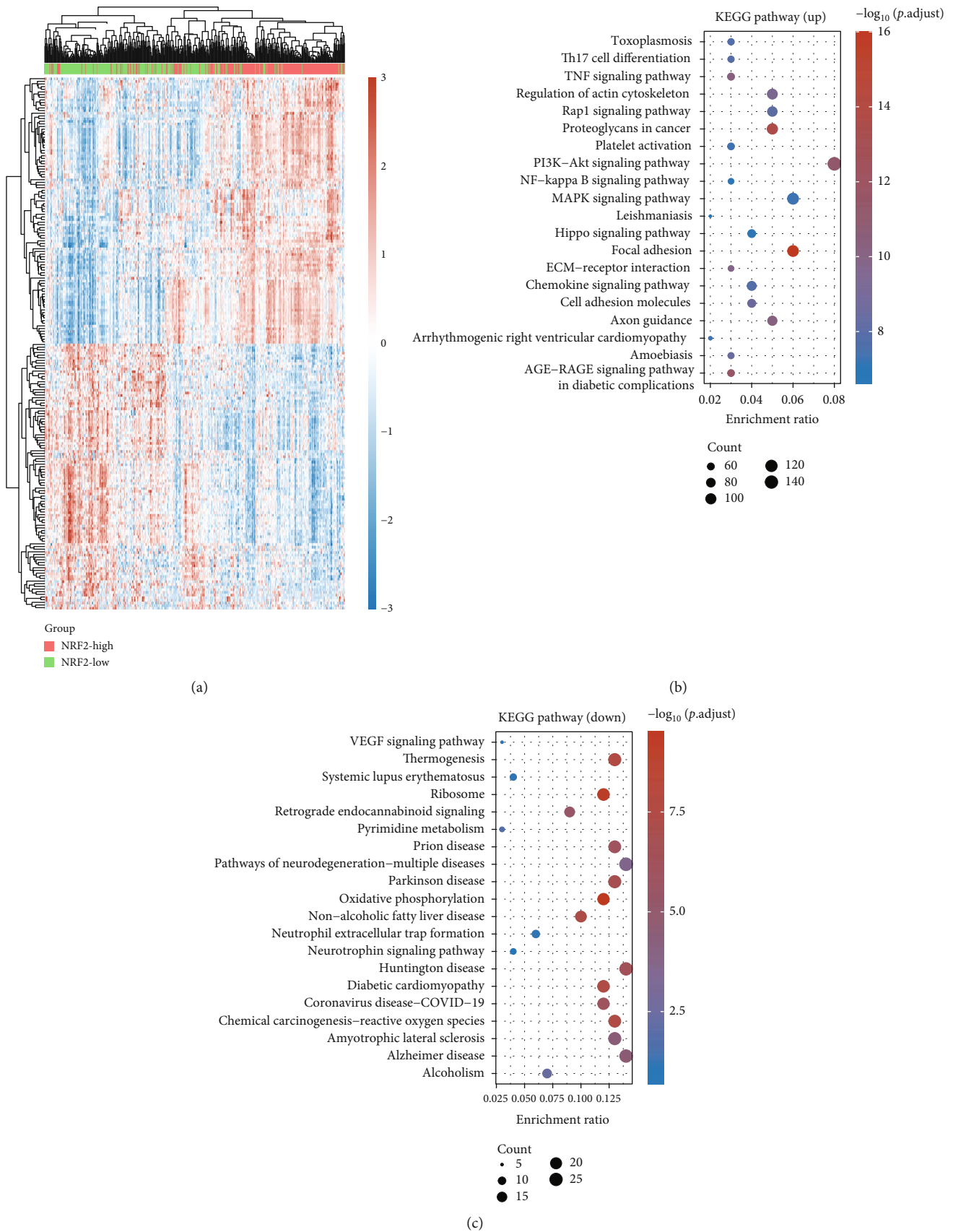


FIGURE 6: Bioinformatics analysis of NRF2 in prostate cancer. (a) The differentially expressed genes between NRF2-high and NRF2-low groups were identified. (b, c) KEGG analysis of upregulated and downregulated DEGs in PCA.

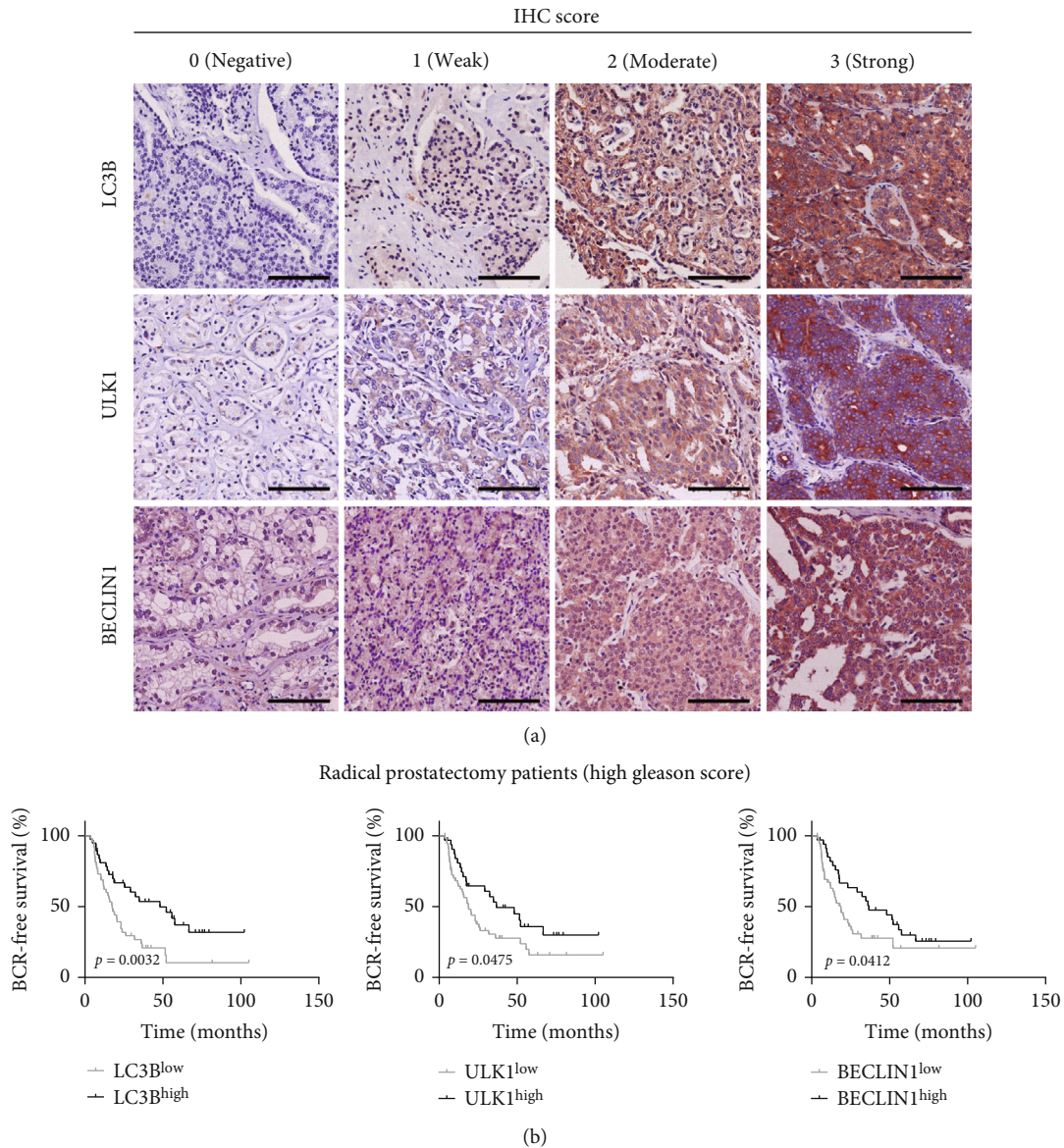


FIGURE 7: The dysregulation of autophagy markers was correlated to the prognosis of PCa. (a) The representative images of IHC 0, 1, 2, and 3 levels of LC3B, ULK1, and beclin1 in PCa. (b) high levels of LC3B, ULK1, and beclin1 significantly correlated to longer BCR-free survival time.

4. Discussion

NRF2 is the master regulator of a group of cytoprotective genes that protect cells from environmental stressors such as reactive oxygen species (ROS) and electrophiles [17]. In addition to its importance in cell physiology and stress response, new research has revealed that NRF2-regulated pathways are frequently activated and play a significant role in initiating malignant transformation, providing a growth advantage, and modulating cancer metabolism [18]. About 30% of human lung cancers have Keap1 or nfe2l2 mutations, resulting in the stability of nfe2l2 gene product NRF2, which controls the oxidation balance. NRF2 accumulation in lung cancer stabilizes Bach1 by inducing HO1, an enzyme that breaks down heme [19]. C-myc-directed NRF2 drives the malignant progression of head and neck cancer through

the activation of glucose-6-phosphate dehydrogenase and transketolase [20]. NRF2 overexpression is linked to enhanced tumor-infiltrating lymphocytes and tumor immunity in ER-positive/HER2-negative breast cancer [20]. NRF2 activation promotes invasive lung cancer and is associated with adverse clinical outcomes. Decoupling of NRF2 expression in nonsmall cell lung cancer promotes the maintenance of the mesenchymal state. However, the roles of NRF2 in PCa remained to be further explored. In this study, we for the first time revealed that NRF2 was downregulated in untreated PCa samples. Moreover, high NRF2 levels were found to be substantially related with a favorable survival rate in untreated PCa patients. In our study, the positive percentage of NRF2 was less than 20% in untreated patients (data not shown), while that was 100% in SCC patients, which indicated that NRF2 could serve as a

marker in the diagnosis of SCC. Thus, we believe that NRF2 plays an important role in castration resistance and even neuroendocrine differentiation. In order to validate the TCGA dataset, we performed immunohistochemical staining assay by using collecting samples and revealed higher total, nuclear, and cytoplasm protein levels of NRF2 correlated to longer BCR-free survival time. These results indicated that NRF2 low expression was correlated to poor outcome in untreated PCa.

ADT is the mainstay of treatment for PCa. However, most PCas become resistant to ADT after prolonged treatment and progress to castration-resistant PCa (CRPC). Understanding the mechanisms underlying the androgen-independent transferring is important for CRPC treatment. Previous studies had demonstrated the interaction between NRF2 signaling and AR signaling. Very interestingly, we found that NRF2 was highly expressed in mCRPC samples compared to HSPC samples, which were higher in 43 CRPC and 10 SCC samples compared to untreated PCa samples and were significantly suppressed after neoadjuvant endocrine therapy and lower in 84 neoadjuvant endocrine therapy (NHT) PCa samples. We used a nonadherent sphere experiment to see if NRF2 altered the tumorigenic potential of CRPC cells and found that ML385 decreased the tumorigenicity of CRPC cells.

Programmed cell death (autophagy, apoptosis, and necrotic ptosis) plays a key role in tumor metastasis and drug resistance [21]. Autophagy plays dual roles in PCa progression via promoting and suppressing PCa growth based on different cellular characteristics. For example, beclin1 is deleted in many PCa patients, which is an important autophagy regulator, suggesting that autophagy may be a tumor-suppressive mechanism in the PCa [22]. Autophagy may also serve as a survival mechanism for cells under stress, making it a tumor-promoting mechanism in PCa. A great number of literatures indicates that the inhibition of AR activity, microtubule, and cell signaling stimulates the adaptive response of autophagy in PCa models. Regulating autophagy can improve the therapeutic effect of the above-mentioned treatment on PCa. Autophagy induction has been shown to make cells more susceptible to apoptosis stimulation and radiation in androgen-independent PCa cells [23, 24]. Very interestingly, as a key regulator of autophagy, we found that NRF2 is downregulated in untreated PCa samples compared to normal PCa samples; however, it was upregulated in mCRPC samples compared to HSPC samples. These results further indicated that NRF2 may serve different roles in the progression of PCa and was activated during PCa androgen-independent transformation. Despite that no autophagy-targeting drugs were approved for PCa therapy, several preclinical studies indicated the promising progression about the using of autophagy-related drugs in PCa [25–27]. For example, diosgenin has high antitumor activity in prostate cancer via inducing autophagy [26]. Moreover, Eriocalyxin B (EriB), a promising candidate in cancer therapy, induced apoptosis and autophagy in PCa via AKT-MTOR signaling [27]. Here, we also evaluate the effect of the autophagy inhibitor on PCa growth. We found that treatment with the autophagy inhib-

itor (CQ) alone did not affect the tumorigenicity of CRPC cells in vitro.

In tumors, a functional link between autophagy dysregulation and NRF2 signaling pathways has been discovered. Recent research has shown that the autophagy and NRF2 signaling pathways are inextricably linked via p62, which binds and transports ubiquitinated protein aggregates to autophagosomes [28]. The interaction of p62 and Keap1 increases NRF2 stability and transcriptional activity. In this study, we found that combination of the NRF2 inhibitor and autophagy inhibitor had a more significantly suppressive role than either ML385 or CQ, indicating that combination of CQ (autophagy inhibitor) and ML385 (NRF2 inhibitor) is a potential treatment of CRPC.

In conclusion, the results of this study show that NRF2 is downregulated in untreated PCa samples compared to normal PCa samples; however, it was upregulated in mCRPC samples compared to HSPC samples. These results demonstrated that NRF2 may serve as a tumor suppressor in tumorigenesis but promote PCa androgen-independent transferring after ADT treatment. Moreover, we revealed that the NRF2 inhibitor significantly inhibited tumorigenicity of CRPC cells in vitro. Of note, combination of the NRF2 inhibitor and autophagy inhibitor had a more significantly suppressive role than either ML385 or CQ, indicating that combination of CQ (autophagy inhibitor) and ML385 (NRF2 inhibitor) is a potential treatment of CRPC.

Data Availability

All data generated or analyzed during this study are included in this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Baijun Dong and Wei Xue conceived the study. Yong Zhang and Zhixiang Xin collected the samples and performed the experiments. All authors read, edited, and approved the final paper. Yong Zhang and Zhixiang Xin contributed equally to this work.

References

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.
- [2] W. Chen, R. Zheng, P. D. Baade et al., "Cancer statistics in China, 2015," *CA: A Cancer Journal for Clinicians*, vol. 66, no. 2, pp. 115–132, 2016.
- [3] C. D. Chen, D. S. Welsbie, C. Tran et al., "Molecular determinants of resistance to antiandrogen therapy," *Nature Medicine*, vol. 10, no. 1, pp. 33–39, 2004.
- [4] K. E. Knudsen and H. I. Scher, "Starving the addiction: new opportunities for durable suppression of AR signaling in prostate cancer," *Clinical Cancer Research*, vol. 15, no. 15, pp. 4792–4798, 2009.

- [5] H. Beltran, D. S. Rickman, K. Park et al., "Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets," *Cancer Discovery*, vol. 1, no. 6, pp. 487–495, 2011.
- [6] C. J. Logothetis, G. E. Gallick, S. N. Maity et al., "Molecular classification of prostate cancer progression: foundation for marker-driven treatment of prostate cancer," *Cancer Discovery*, vol. 3, no. 8, pp. 849–861, 2013.
- [7] J. L. Yao, R. Madeb, P. Bourne et al., "Small cell carcinoma of the prostate: an immunohistochemical study," *The American Journal of Surgical Pathology*, vol. 30, no. 6, pp. 705–712, 2006.
- [8] M. McMahon, K. Itoh, M. Yamamoto et al., "The Cap'n'Collar basic leucine zipper transcription factor Nrf2 (NF-E2 p45-related factor 2) controls both constitutive and inducible expression of intestinal detoxification and glutathione biosynthetic enzymes," *Cancer Research*, vol. 61, no. 8, pp. 3299–3307, 2001.
- [9] R. K. Thimmulappa, K. H. Mai, S. Srisuma, T. W. Kensler, M. Yamamoto, and S. Biswal, "Identification of Nrf2-regulated genes induced by the chemopreventive agent sulforaphane by oligonucleotide microarray," *Cancer Research*, vol. 62, no. 18, pp. 5196–5203, 2002.
- [10] N. M. Reddy, S. R. Kleeberger, M. Yamamoto et al., "Genetic dissection of the Nrf2-dependent redox signaling-regulated transcriptional programs of cell proliferation and cytoprotection," *Physiological Genomics*, vol. 32, no. 1, pp. 74–81, 2007.
- [11] M. A. Schultz, S. S. Hagan, A. Datta et al., "Nrf1 and Nrf2 transcription factors regulate androgen receptor transactivation in prostate cancer cells," *PLoS One*, vol. 9, no. 1, article e87204, 2014.
- [12] N. Khurana and S. C. Sikka, "Targeting crosstalk between Nrf2, NF-kappaB and androgen receptor signaling in prostate cancer," *Cancers*, vol. 10, no. 10, p. 352, 2018.
- [13] M. Shiota, A. Yokomizo, and S. Naito, "Oxidative stress and androgen receptor signaling in the development and progression of castration-resistant prostate cancer," *Free Radical Biology & Medicine*, vol. 51, no. 7, pp. 1320–1328, 2011.
- [14] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: a bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, vol. 26, no. 1, pp. 139–140, 2010.
- [15] W. Huang da, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.
- [16] X. Wan, S. Yang, W. Huang et al., "UHRF1 overexpression is involved in cell proliferation and biochemical recurrence in prostate cancer after radical prostatectomy," *Journal of Experimental & Clinical Cancer Research*, vol. 35, no. 1, p. 34, 2016.
- [17] S. Vomund, A. Schafer, M. J. Parnham, B. Brune, and A. von Knethen, "Nrf2, the master regulator of anti-oxidative responses," *International Journal of Molecular Sciences*, vol. 18, no. 12, p. 2772, 2017.
- [18] E. Panieri and L. Saso, "Potential applications of NRF2 inhibitors in cancer therapy," *Oxidative Medicine and Cellular Longevity*, vol. 2019, Article ID 8592348, 34 pages, 2019.
- [19] L. Lignitto, S. E. LeBoeuf, H. Homer et al., "Nrf2 activation promotes lung cancer metastasis by inhibiting the degradation of Bach 1," *Cell*, vol. 178, no. 316–329, article e18, 2019.
- [20] Y. C. Tang, J. R. Hsiao, S. S. Jiang et al., "c-MYC-directed NRF2 drives malignant progression of head and neck cancer via glucose-6-phosphate dehydrogenase and transketolase activation," *Theranostics*, vol. 11, no. 11, pp. 5232–5247, 2021.
- [21] Z. Su, Z. Yang, Y. Xu, Y. Chen, and Q. Yu, "Apoptosis, autophagy, necroptosis, and cancer metastasis," *Molecular Cancer*, vol. 14, no. 1, p. 48, 2015.
- [22] C. Liu, P. Xu, D. Chen et al., "Roles of autophagy-related genes Beclin-1 and LC3 in the development and progression of prostate cancer and benign prostatic hyperplasia," *Biomedical Reports*, vol. 1, no. 6, pp. 855–860, 2013.
- [23] Q. Jiang, S. Yeh, X. Wang et al., "Targeting androgen receptor leads to suppression of prostate cancer via induction of autophagy," *The Journal of Urology*, vol. 188, no. 4, pp. 1361–1368, 2012.
- [24] E. Ziparo, S. Petrungero, E. S. Marini et al., "Autophagy in prostate cancer and androgen suppression therapy," *International Journal of Molecular Sciences*, vol. 14, no. 6, pp. 12090–12106, 2013.
- [25] M. Ashrafizadeh, M. D. A. Paskeh, S. Mirzaei et al., "Targeting autophagy in prostate cancer: preclinical and clinical evidence for therapeutic response," *Journal of Experimental & Clinical Cancer Research*, vol. 41, no. 1, p. 105, 2022.
- [26] C. Nie, J. Zhou, X. Qin et al., "Diosgenin-induced autophagy and apoptosis in a human prostate cancer cell line," *Molecular Medicine Reports*, vol. 14, no. 5, pp. 4349–4359, 2016.
- [27] Z. Yu, Y. Chen, and C. Liang, "Eriocalyxin B induces apoptosis and autophagy involving Akt/mammalian target of rapamycin (mTOR) pathway in prostate cancer cells," *Medical Science Monitor*, vol. 25, pp. 8534–8543, 2019.
- [28] T. Jiang, B. Harder, M. Rojo de la Vega, P. K. Wong, E. Chapman, and D. D. Zhang, "p62 links autophagy and Nrf2 signaling," *Free Radical Biology & Medicine*, vol. 88, Part B, pp. 199–204, 2015.

Research Article

Effect of Probiotics Therapy on Nonalcoholic Fatty Liver Disease

Yuanshe Huang,^{1,2} Xiaodong Wang,³ Lai Zhang,¹ Ke Zheng,⁴ Jie Xiong,⁵ Jing Li,¹
Chunlei Cong ,¹ Zhaomiao Gong ,³ and Jingxin Mao ^{2,3}

¹Anshun University, Guizhou, Anshun 561000, China

²College of Pharmaceutical Sciences, Southwest University, Chongqing 400715, China

³Chongqing Medical and Pharmaceutical College, Chongqing 400030, China

⁴Department of Endocrine and Breast Surgery, The First Affiliated Hospital of Chongqing Medical University, No. 1 Youyi Road, Yuzhong District, Chongqing 400016, China

⁵Department of Pharmacy, Ministry of Education Key Laboratory of Child Development and Disorders, National Clinical Research Center for Child Health and Disorders/Chongqing Key Laboratory of Pediatrics/Children's Hospital of Chongqing Medical University, Chongqing 400014, China

Correspondence should be addressed to Zhaomiao Gong; 657224511@qq.com and Jingxin Mao; maomao1985@email.swu.edu.cn

Received 4 April 2022; Revised 30 April 2022; Accepted 13 May 2022; Published 30 May 2022

Academic Editor: Lei Chen

Copyright © 2022 Yuanshe Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. Nonalcoholic fatty liver disease (NAFLD) is the most prevalent chronic liver disease in the world. The pathogenesis of NAFLD is complex and multifactorial. Clinical studies have shown that alterations in the gut microbiota play a key role in NAFLD. The purpose of this study was to analyze the effect of probiotic supplementation on the treatment of NAFLD patients based on various indicators. **Methods.** We conducted a meta-analysis investigating the relationship between NAFLD and probiotic supplementation. Embase, PubMed, and Web of Science databases were searched by computer, and then, eligible studies were identified. Finally, a total of high-quality randomized controlled trials were selected involving 1403 participants. Meta-analysis was performed using the RevMan 5.3 software which was systematically searched for works published through Dec. 1, 2021, in the present study. **Results.** The meta-analysis results showed that the probiotics supplementation improved hepatocyte injury and significantly reduced the level of ALT ($P = 0.00001$), AST ($P = 0.0009$), GGT ($P = 0.04$), TG ($P = 0.01$), LDL-C ($P = 0.0005$), HDL-C ($P = 0.0002$), insulin ($P = 0.003$), IR ($P = 0.03$), BMI ($P = 0.03$), TNF- α ($P = 0.03$), and CRP ($P = 0.02$), respectively, in NAFLD patients. **Conclusion.** The present study suggests that probiotics therapy may improve liver enzyme levels, regulated lipid metabolism, reduced insulin resistance, and improved inflammation in NAFLD patients. It supports the potential role of probiotics supplementation in the treatment of NAFLD.

1. Introduction

Nonalcoholic fatty liver disease (NAFLD) was always considered to be a disorder caused by excessive deposition of fat in liver cells in addition to alcohol and other definite factors [1]. The incidence of NAFLD is increasing year by year, and it has become the most common cause of chronic liver disease in both developed and developing countries [2]. The NAFLD disease mainly includes nonalcoholic fatty liver (NAFL) and nonalcoholic steatohepatitis (NASH) which commonly associated cirrhosis and liver cancer [3]. It was

reported that 2%-3% of NAFLD patients and 15%-20% of NASH patients may finally develop into cirrhosis or even liver cancer [4]. Although the complex pathogenesis of NAFLD has not been fully elucidated, its pathogenesis is mainly related to metabolic abnormalities, such as insulin resistance (IR), type 2 diabetes, visceral obesity, and abnormal metabolism of blood lipids [5]. An international expert consensus statement was issued by a multinational expert group to change the name of NAFLD to a new definition of metabolic fatty liver disease (MAFLD) in 2020 [6]. The "two-hit theory" was the first hypothesis proposed for the

pathogenesis of NAFLD. The “first hit” is characterized by lipid accumulation in the liver due to IR [7] while the “second hit” is characterized by lipid peroxidation, secretion of proinflammatory cytokines, and mitochondrial dysfunction that determines disease progression [8, 9]. It is currently known that these mechanisms are not sufficient to explain all NAFLD pathogenesis, so the “multiple parallel hit theory” has received increasing attention. According to this concept, several processes including adipose tissue-derived signaling, gut barrier dysfunction, genetic factors, endoplasmic reticulum stress, and related signaling networks may work together to contribute to the progression of steatosis to NASH development [10].

It was reported that the human gut microbiota has emerged as a major participant in human health and disease. The gut and liver “communicate extensively” through the biliary tract, portal vein, and systemic circulation. This bidirectional connection is called the “gut-liver axis.” The liver becomes a key first-line immune organ [11]. Previous studies have shown that the intestinal flora may play a regulatory role in NAFLD and other metabolic diseases through the “gut-liver axis” [12], and it has also been found in clinical practice that patients with NAFLD may easily suffered from imbalance of gut microbiota and microbial metabolic dysfunction. The metabolic function of microorganisms in the tract is disordered, and the number of pathogenic bacteria significantly increases [13]. The liver receives a variety of gut-derived signals that including bacterial products, environmental toxins, and food antigens and normally strikes a balance between immunity and tolerance that is critical to its function. Even in the absence of pathogens, changes in the gut microbiota brought about disruption of intestinal homeostasis which leading to disturbances in immune status and various liver diseases. Excessive immune responses may also cause sterile liver inflammation, chronic inflammation, and liver disease-related cancer [14]. Nowadays, there is a lack of effective drugs for the treatment of NAFLD both domestically and abroad. The main treatment is to improve lifestyle and metabolic disorder. Among the commonly used clinical drugs, the therapeutic effect of vitamin E on IR is not clear, and it will increase the risk of stroke. Pioglitazone may significantly increase the patient’s weight, and obexolate will lead to abnormal lipid metabolism [15].

Probiotics are known as nonpathogenic living microorganisms which may affect the health of the host by regulating the intestinal microbiota, producing antibacterial substances, improving epithelial barrier function, and reducing intestinal inflammation [16]. At birth, the gastrointestinal tract is sterile. Within a few months after birth, a relatively stable microbial population was established. This kind of rich, diverse, and dynamic intestinal flora usually has a complex symbiotic relationship with mucosal eukaryotic cells. Intestinal microbial imbalance is closely related to the occurrence and progress of NAFLD. It was reported that probiotics may improve the risk factors of NAFLD by affecting the intestinal microbiota [17]. With the in-depth study of probiotics, many clinical and basic trials have been proved that probiotics are a potential treatment for NAFLD. In the present study, we conducted a meta-analysis of the

clinical literature of probiotics in the treatment of NAFLD in recent 10 years, so as to provide evidence-based medical basis for the clinical treatment of NAFLD.

2. Methods

2.1. Search Strategy. Systematic review and meta-analysis were conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines to ensure the reliability and integrity of data and conclusions [18]. A comprehensive literature search was conducted on the database (Embase, PubMed, and Web of Science) using the keyword words as follows: “non alcoholic fatty liver disease” OR “NAFLD” OR “nonalcoholic fatty liver disease” OR “nonalcoholic fatty liver” OR “non-alcoholic steatohepatitis” OR “NASH” OR “NAFL” AND “Lactobacillus” OR “probiotics” OR “synbiotic”. The publication time range of the literature is from January 1, 2011, to December 1, 2021.

2.2. Selection Criteria. We followed the methods of Mao et al. in 2020 and 2022, respectively [19, 20]. The following selection criteria had been used for the decision-making process.

Inclusion criteria are as follows: (1) the clinical trials in the literature were randomized controlled trials (RCTs); (2) the participants were patients with NAFLD who were not restricted by age, sex, or race; (3) the intervention was probiotics; (4) studies that directly assess the effects of probiotics based on any outcome measure; and (5) complete outcome data were available to assess treatment effects.

Exclusion criteria are as follows: (1) the trial was not a RCT trial; (2) patients with hepatic steatosis or fibrosis due to other causes, such as viral hepatitis, autoimmune hepatitis, drug-induced hepatitis, and liver disease due to genetic causes, were all excluded; (3) the experimental measures were interfered with by other therapeutic measures; (4) literature was conference case reports, letters to the editor, editorials, or abstract; (5) studying period beyond 10 years; and (6) insufficient data or only from big data.

2.3. Data Extraction. The data of the included literature were extracted using the preset table by two researchers (Table 1), including the first author, the year of publication, case number, intervention time, diagnostic method, probiotic ingredients, control group, and Newcastle Ottawa Scale (NOS).

2.4. Literature Quality Evaluation. Three researchers independently evaluated the quality of the included literature according to Cochrane Collaboration tool (version 5.1.0), including whether the research object was generated by random method, whether the distribution scheme was hidden, whether blind method was used, data integrity, whether there was selective report, and other bias.

2.5. Heterogeneity, Sensitivity Test, and Publication Bias Analysis. The Q test was used to analyze the heterogeneity of the pooled effect sizes, and the results were expressed by I^2 using the RevMan 5.3 software. If there is homogeneity among studies ($P > 0.05$ or $I^2 \leq 50\%$), the fixed-effect model is used for analysis; otherwise, if there is heterogeneity

TABLE 1: Basic traits and characteristics of the included studies.

First author	Country	Publication years	Case number	Intervention time	Diagnostic method	Probiotic ingredients	Control group	NOS
Abdel [45]	Egypt	2017	30 (15/15)	4 weeks	Liver biopsy	Lactobacillus	Nonplacebo	5
Abhari [46]	Egypt	2020	45 (23/22)	12 weeks	N/A	Bacillus	Placebo	6
Ahn [47]	South Korea	2019	65 (30/35)	12 weeks	Liver biopsy	6 probiotic mixtures	Placebo	7
Alisi [48]	Italy	2014	24 (10/14)	4 months	Liver biopsy	VSL #3	Placebo	8
Aller [49]	Spain	2011	28 (14/14)	3 months	Liver biopsy	Lactobacillus bulgaricus, Streptococcus thermophilus	Placebo	6
Asgharian [50]	Iran	2016	74 (36/38)	12 weeks	Ultrasonic examination	Lactobacillus, Bifidobacterium, Streptococcus	Placebo	8
Bakhshimoghaddam [51]	Iran	2018	68 (34/34)	24 weeks	Ultrasonic examination	Streptococcus thermophilus, Lactobacillus, Bifidobacterium	Nonplacebo	9
Behrouz [52]	Iran	2017	60 (30/30)	12 weeks	Ultrasonic examination	Lactobacillus, Bifidobacterium	Placebo	7
Cai [53]	China	2020	140 (70/70)	12 weeks	Liver biopsy	Lactobacillus, Bifidobacterium, Enterococcus	Nonplacebo	6
Dagan [54]	Israel	2017	80 (40/40)	24 weeks	N/A	Lactobacillus, Bifidobacterium, Streptococcus, Lactococcus	Placebo	6
Duseja [55]	India	2019	30 (17/13)	48 weeks	Liver biopsy	Lactobacillus, Bifidobacterium	Placebo	5
Ekhlesi [56]	Iran	2016	30 (15/15)	8 weeks	Ultrasonic examination	Lactobacillus, Bifidobacterium, Streptococcus	Placebo	6
Eslamparast [57]	Iran	2014	52 (26/26)	28 weeks	Ultrasonic examination	Lactobacillus, Bifidobacterium, Streptococcus	Placebo	7
Ferolla [58]	Brazil	2016	50 (27/23)	12 weeks	Liver biopsy	Lactobacillus	Nonplacebo	7
Famouri [59]	Iran	2017	64 (32/32)	12 weeks	Ultrasonic examination	Lactobacillus, Bifidobacterium	Placebo	7
Javadi [60]	Iran	2017	38 (19/19)	18 weeks	Liver biopsy	Bifidobacterium longum, Lactobacillus acidophilus	Placebo	6
Kobyliak [61]	Ukraine	2018	58 (30/28)	8 weeks	Liver biopsy	Bifidobacterium, Lactobacillus, Lactococcus, Propionibacterium	Placebo	7
Malaguarnera [62]	Italy	2012	66 (34/32)	24 weeks	N/A	Bifidobacterium longum	Placebo	6
Manzhalii [63]	Germany	2017	75 (38/37)	12 weeks	Ultrasonic examination	Lactobacillus casei, L. rhamnosus, L. bulgaricus, Bifidobacterium longum, and Streptococcus thermophilus	Nonplacebo	7
Mofidi [64]	Iran	2017	42 (21/21)	28 weeks	Ultrasonic examination	Lactobacillus, Bifidobacterium, Streptococcus	Placebo	6
Nabavi [65]	Iran	2014	72 (36/36)	18 weeks	Ultrasonic examination	Lactobacillus bulgaricus and Streptococcus thermophilus	Placebo	6
Sepideh [66]	Iran	2015	42 (21/21)	8 weeks	Ultrasonic examination	Lactobacillus, Bifidobacterium, Streptococcus	Placebo	5
Shavakhi [67]	Iran	2013	63 (32/31)	6 months	Liver biopsy	Protexin tablet	Placebo	5
Vajro [68]	Italy	2011	20 (10/10)	8 weeks	Ultrasonic examination	Lactobacillus	Placebo	6

The NOS presents Newcastle-Ottawa Scale.

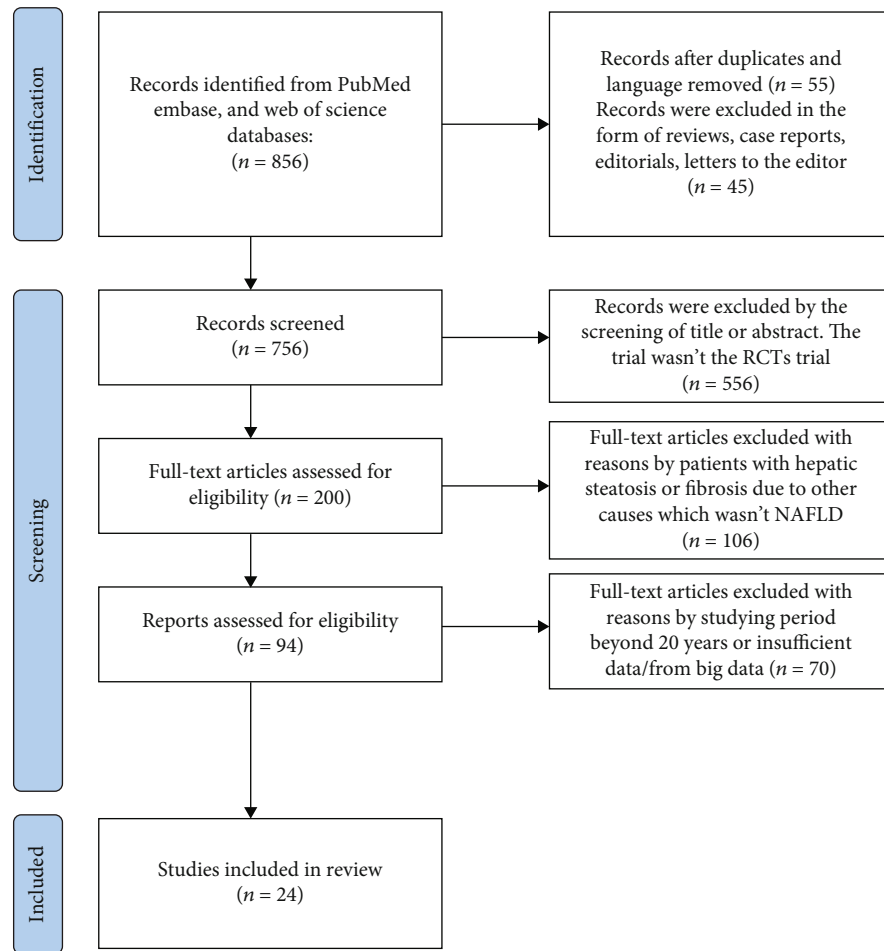


FIGURE 1: Flow chart of the study selection process.

($P \leq 0.05$ or $I^2 > 50\%$), subgroup analysis should be performed to reduce heterogeneity as much as possible. If the cause of heterogeneity is not found, the random effect model is used for meta-analysis. Sensitivity analysis was carried out by comparing the difference between point estimation and interval estimation of combined effect values in different effect models. Begg's test was used for publication bias analysis; it indicates that there is no obvious publication bias if $P > 0.1$.

2.6. Statistical Evaluation. Meta-analysis was performed using the RevMan 5.3 software. When the measurement methods or units of the effect of the same intervention were exactly the same, the mean difference (MD) pooled statistic was selected. When different measurement methods or units were used for the effect of the same intervention, the standard mean difference (SMD) was selected as the pooled statistic. All data were analyzed with 95% confidence interval (CI). Data analyses were performed using the Q test and I^2 statistic to detect heterogeneity. Using a forest plot to describe the results, with P values less than 0.05 considered statistically significant. The NOS was used to assess the quality of the research. In addition, Begg funnel plots have been used to search for viable publication bias.

3. Results

Through the search strategy, 24 studies that met the criteria in the past 10 years were finally included, including a total of 1403 patients. The flow chart of the study selection process is presented in Figure 1. The study population includes children and adults, with the diagnosis method of ultrasonic examination and liver biopsy combined with clinical biochemical indicators. The probiotics preparations included in the literature mainly include *Lactobacillus*, *Bifidobacterium*, *Enterococcus*, *Streptococcus*, *Bacillus*, and *Lactococcus*. The basic characteristics of the included studies and the quality assessment of the literature are listed in Table 1. The quality of the literature was assessed using the Cochrane Collaboration tool. The evaluation criteria of NOS are based on whether the definition and diagnosis of the disease are appropriate, whether the pathology is representative, whether the selection of control is reasonable, whether there are reliable data sources, whether double-blind analysis is used, etc. A total score of 1 to 4 was considered low quality, and a score of 5 to 9 was considered high quality. A total of 24 articles were all high-quality, and none of the articles had a high risk of bias. The related factors for probiotics therapy on NAFLD patients are presented in Table 2.

TABLE 2: Related factors for probiotics therapy on NAFLD patients.

Related factors	MD	95% CI	P value
ALT	-7.25	-10.11--4.39	0.00001
AST	-3.53	-5.62--1.44	0.0009
GGT	-2.27	-4.49--0.05	0.04
TG	-0.42	-0.53--0.05	0.01
LDL-C	-1.38	-2.15--0.60	0.0005
HDL-C	-19.92	-30.56--9.29	0.0002
IR	-0.61	-1.02--0.21	0.003
Insulin	-1.27	-2.39--0.15	0.03
GLU	-0.03	-0.32--0.25	0.82
BMI	-0.80	-1.51--0.08	0.03
TNF- α	-2.43	-6.56--1.71	0.03
CRP	-1.06	-1.94--0.18	0.02

3.1. Effects of Probiotics on Liver Enzymes in NAFLD Patients. The effect of supplementing probiotics on alanine aminotransferase (ALT) was explored in 19 included literatures in the study. The heterogeneity test results showed that heterogeneity exists in the literature ($I^2 = 71% > 87%$, Q test $P < 0.00001$). Therefore, random effects were used to meta-analysis. The MD value of the 19 literature summary was -7.25; the 95% CI was among -10.11 to -4.39, $Z = -4.97$, and $P < 0.00001$ (Figure 2(a)). The effect of supplementing probiotics on aspartate aminotransferase (AST) was investigated in 18 included literatures. The heterogeneity test results investigated that heterogeneity exists in the present study ($I^2 = 71% > 50%$ and the Q test $P < 0.00001$). Qualitative, random effects were selected for meta-analysis. The MD value of the 18 literature summary was -3.53; the 95% CI was among -5.62 to -1.44, $Z = 2.69$, and $P = 0.0009 < 0.05$ (Figure 2(b)). The effect of probiotics on γ -glutamyltransferase (GGT) was studied in a total of 9 literatures. The heterogeneity assessment results found that heterogeneity exists in the literature ($I^2 = 92% > 50%$ and the Q test $P = 0.00001$). Therefore, random effects were selected for meta-analysis. The MD value of the 9 literatures was -2.27; the 95% CI was among -4.49 to -0.05, $Z = 2$, and $P = 0.04 < 0.05$ (Figure 2(c)). It was suggested that probiotics supplementation in the treatment of NAFLD patients may improve hepatocyte injury and intrahepatic biliary obstruction and significantly reduce the level of ALT, AST and GGT.

3.2. Effect of Probiotics on Lipid Metabolism in Patients with NAFLD. Effect of probiotics supplement on triglyceride (TG) was studied in 15 literatures. The heterogeneity test results revealed that heterogeneity exists in the inclusive literatures ($I^2 = 53% > 50%$, $P = 0.008 < 0.05$ in Q test). Random effects were selected for meta-analysis. The MD value of 15 literatures was -0.42; the 95% CI is among -0.53 to 0.05, $Z = 1.59$, and $P = 0.01$ (Figure 3(a)). Effect of probiotics supplementation on low density lipoprotein (LDL-C) was investigated in 15 studies. The heterogeneity test results explored that heterogeneity exists in the present study ($I^2 = 90% > 50%$ and $P < 0.00001$ of Q test). Random effects were selected for meta-analysis. The MD value of 15 literatures

is -1.38; the 95% CI is among -2.15 to -0.60, $Z = 3.47$, and $P = 0.0005 < 0.05$ (Figure 3(b)). Effect of probiotics on high density lipoprotein (HDL-C) was investigated in 18 literatures. The test results of statistical heterogeneity for the meta-analysis of literatures are heterogeneity that exists ($I^2 = 53% > 50%$ and $P < 0.00001$ of Q test). Random effects were selected for meta-analysis. The MD value of 18 literatures is -19.92; the 95% CI is among -30.56 to -9.29, $Z = 3.67$, and $P = 0.0002 < 0.05$ (Figure 3(c)). It was indicated that probiotics supplementation in the treatment of NAFLD patients may significantly reduce the level of TG, LDL-C, and HDL-C.

3.3. Effect of Probiotics on Blood Glucose-Related Indexes in Patients with NAFLD. Effect of probiotics on insulin resistance (IR) was studied in the present study. Test results of statistical heterogeneity for the meta-analysis of literatures ($I^2 = 89% > 50%$ and $P < 0.00001$ of Q test). Random effects were selected for meta-analysis. The MD value of 13 literatures is -0.61; the 95% CI is -1.02 to -0.21, $Z = 2.98$, and $P = 0.003 < 0.05$ (Figure 4(a)). Effect of probiotics on insulin was explored in 11 literatures. The heterogeneity test results revealed that heterogeneity exists in the present study ($I^2 = 86% > 50%$ and $P < 0.00001$ of Q test). Therefore, the random effects were selected for meta-analysis. The MD value of 11 literatures is -1.27; the 95% CI is among -2.39 to -0.15, $Z = 2.23$, and $P = 0.03 < 0.05$ (Figure 4(b)). Effects of probiotics on blood glucose (GLU) were studied in total of 10 articles. The heterogeneity test results showed that no heterogeneity found in the literature ($I^2 = 0% < 50%$ and $P = 0.45 > 0.1$ of Q test), and the fixed effects were selected for meta-analysis. The MD value of 10 literatures is -0.03; the 95% CI is among -0.32 to 0.25, $Z = 0.23$, and $P = 0.82$ (Figure 4(c)). The above suggests that probiotics supplementation in the treatment of NAFLD may significantly reduce the levels of insulin and IR.

3.4. Effects of Probiotics on BMI and Inflammation in NAFLD Patients. The effect of probiotics supplementation on body mass index (BMI) was investigated in 24 literatures. The heterogeneity test results revealed that heterogeneity exists in the present study ($I^2 = 89% > 50%$ and the Q test $P < 0.00001$). A random effect was selected for meta-analysis. The MD value of the 24 literature pooled was -0.80; the 95% CI was among -1.51 to -0.08, $Z = 2.18$, and $P = 0.03 < 0.05$, suggesting that probiotics treatment of NAFLD may significantly reduce the BMI of patients (Figure 5(a)). Effect of probiotics supplementation on tumor necrosis factor- α (TNF- α) was explored in 10 included literatures. The test results of statistical heterogeneity for the meta-analysis were $I^2 = 99% > 50%$ and Q test $P < 0.00001$, respectively (heterogeneity exists), and random effects were selected for meta-analysis. The MD value of the 10 articles was -2.43; the 95% CI was among -6.56 to 1.71, $Z = 1.15$, and $P = 0.03$ (Figure 5(b)). Effect of probiotics supplementation on C-reactive protein (CRP) was investigated in 6 included literatures. The heterogeneity test results revealed that no heterogeneity exists in the studied researches ($I^2 = 0% < 50%$ and Q test $P = 0.15 > 0.1$). Therefore, fixed

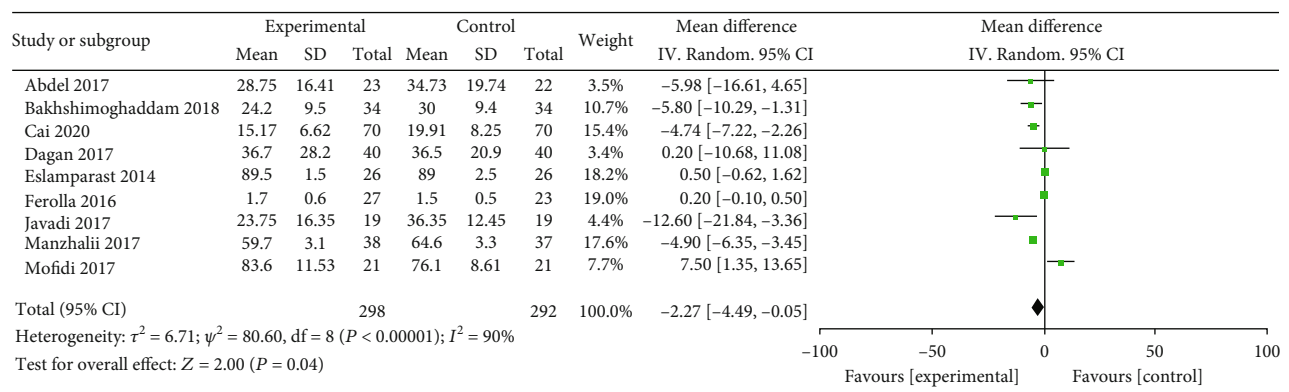
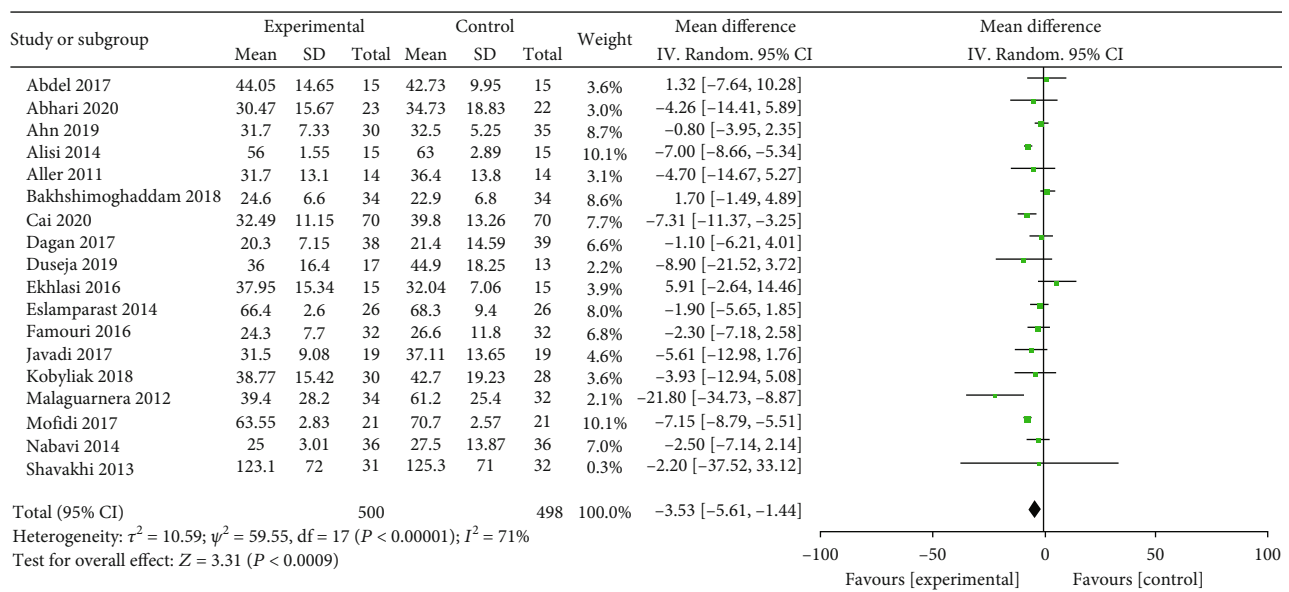
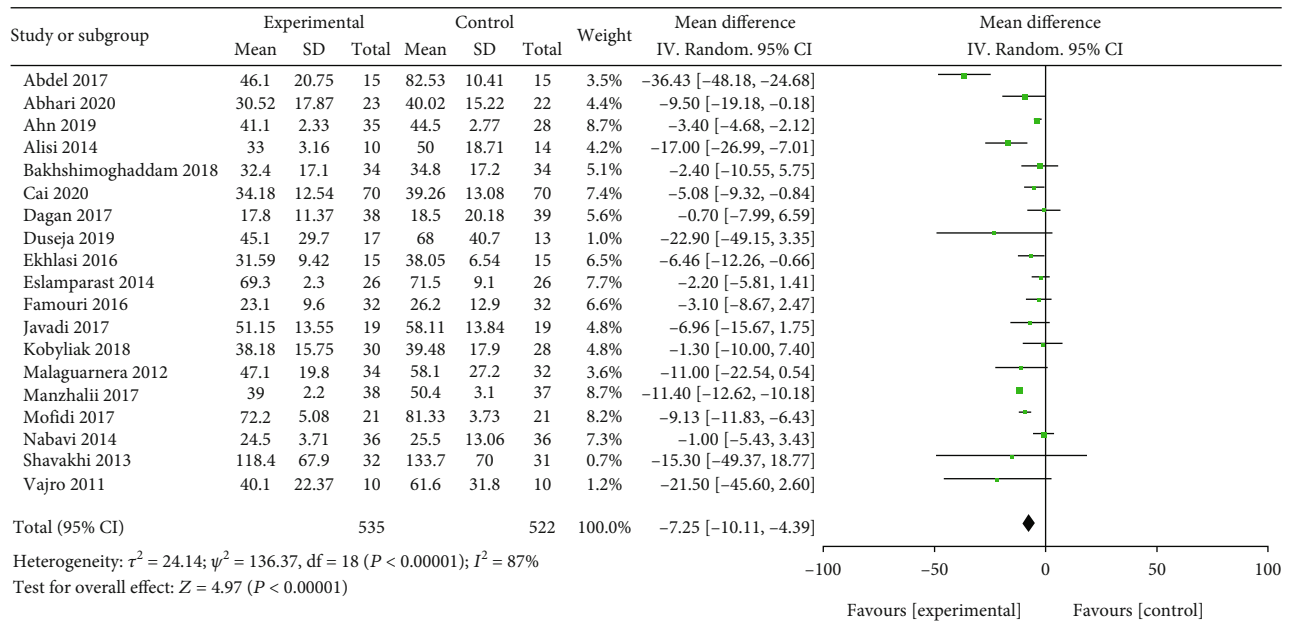


FIGURE 2: Forest plots of the effect of probiotics therapy on (a) ALT, (b) AST, and (c) GGT levels.

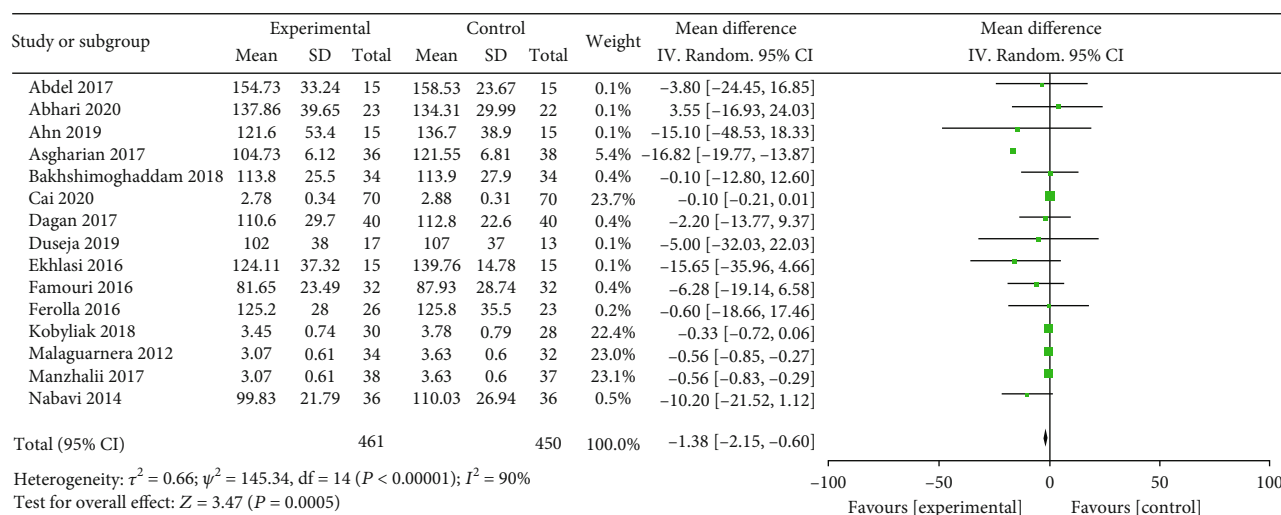
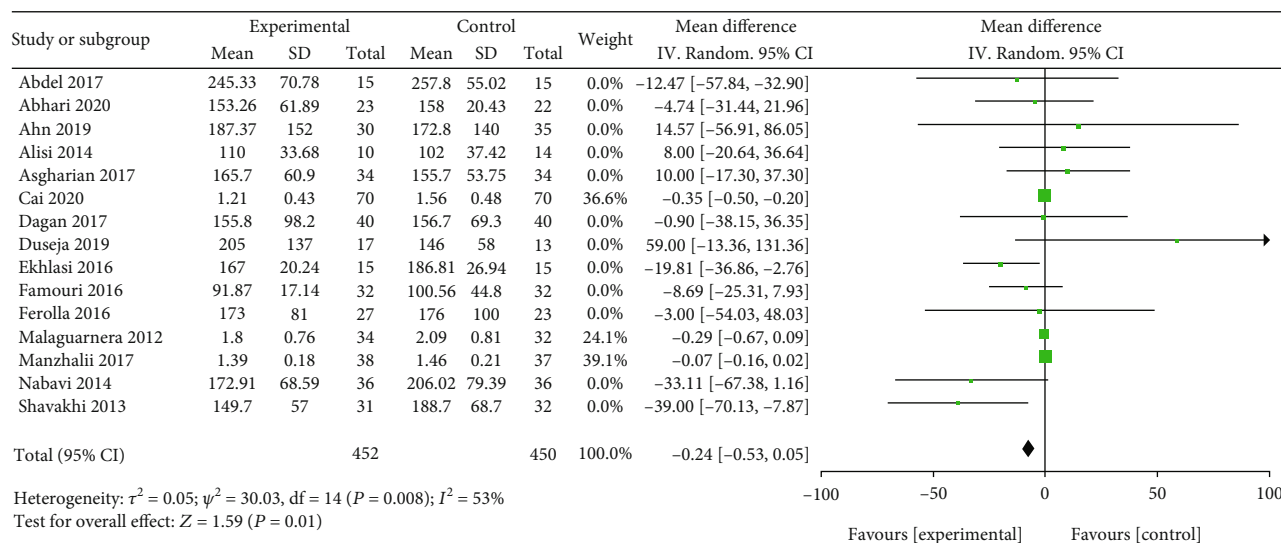


FIGURE 3: Continued.

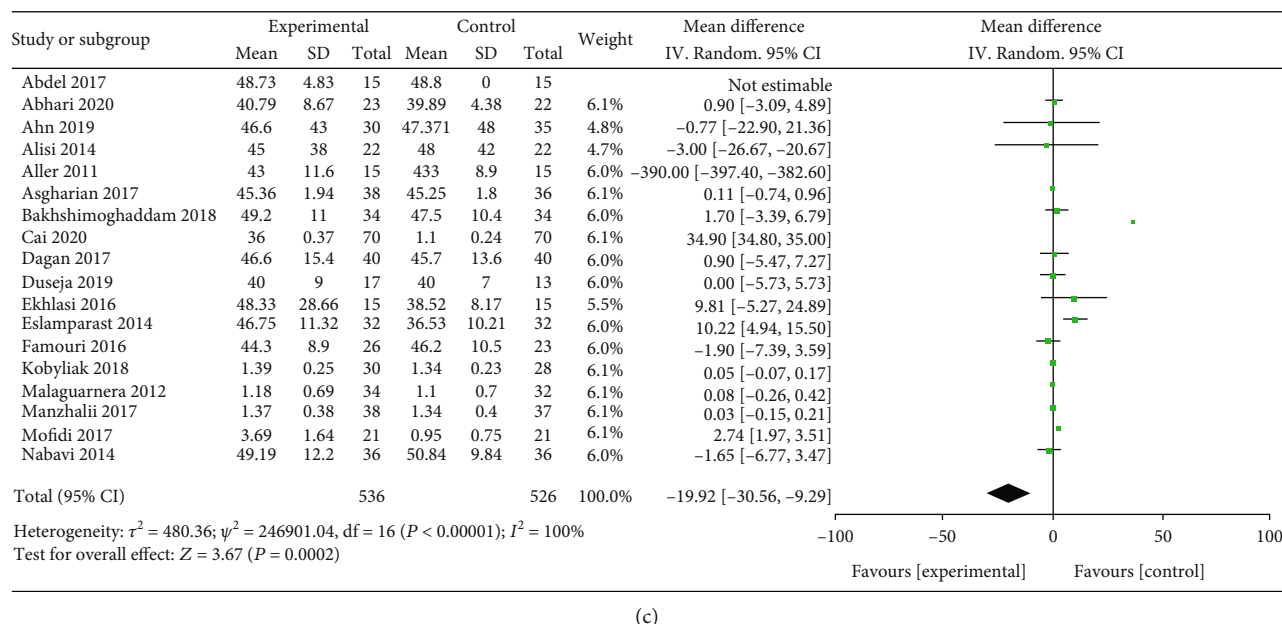


FIGURE 3: Forest plots of the effect of probiotics therapy on (a) TG, (b) LDL-C, and (c) HDL-C levels.

effects can be selected for meta-analysis. The MD value of the six literature pools was -1.06; the 95% CI was among -1.94 to -0.18, $Z = 2.36$, and $P = 0.02 < 0.05$ (Figure 5(c)). It was suggested that supplementing probiotics in the treatment of NAFLD may significantly decrease the levels of BMI, TNF- α , and CRP.

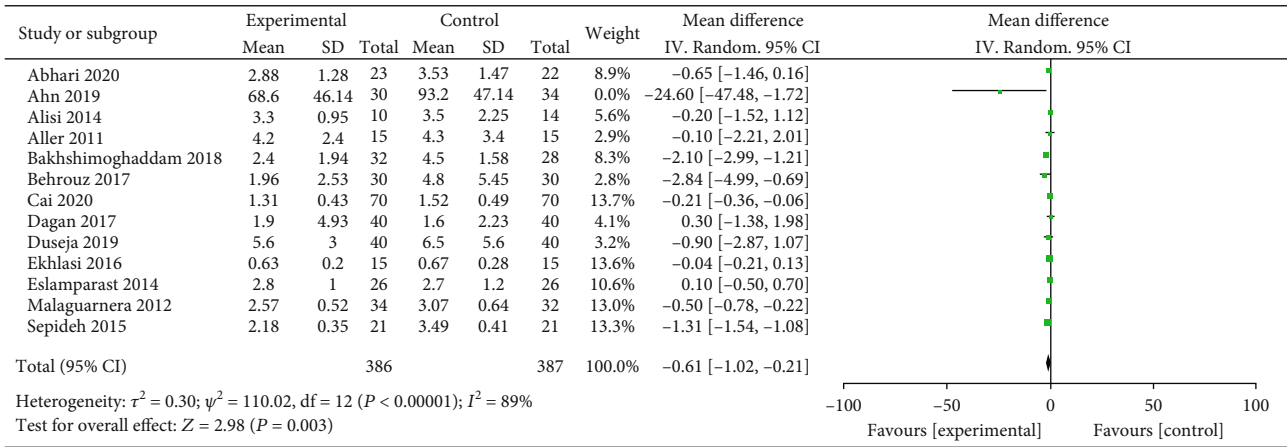
3.5. Results of Heterogeneity, Sensitivity Test, and Bias Publication Analysis. Due to the obvious heterogeneity of some indicators, further subgroup analysis showed that there was no correlation between the country of publication and the heterogeneity, while the intervention time of the study group was related to the heterogeneity of the literature. Sensitivity analyses were performed by comparing the differential changes in point estimates and interval estimates of pooled effect sizes when different effect models were compared. The results showed that there was no change in the conclusions of each indicator, indicating that the results of the meta-analysis of each indicator in this study were stable. Cochrane funnel plot was used to explore the publication bias. There is no obvious asymmetric distribution exhibits in Figure 6 which indicating that there was no publication bias.

4. Discussion

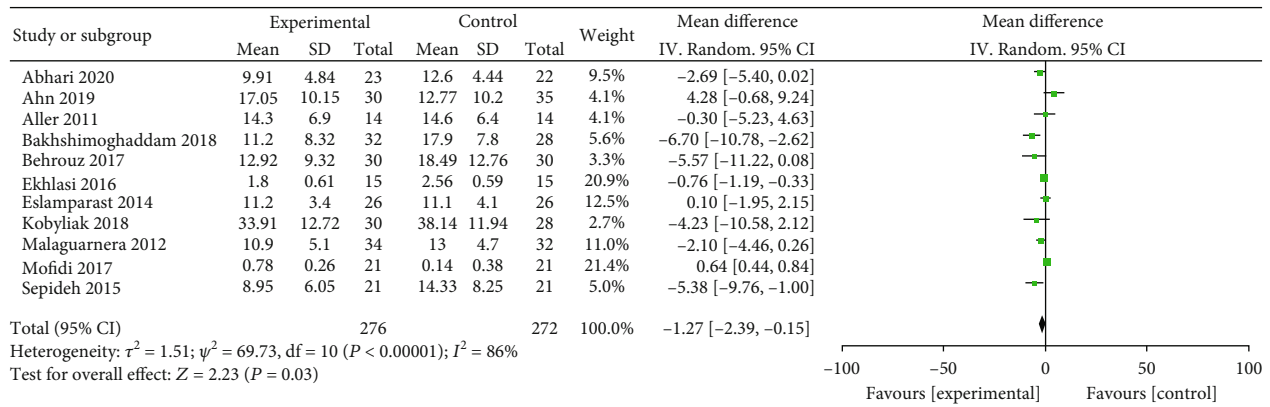
NAFLD usually refers to the excessive accumulation of triglycerides in hepatocytes $\geq 5\%$ without heavy drinking (less than 20 or 30 g/day for women and men, respectively) or other liver diseases such as viral, autoimmune, metabolic, or drug-induced [21]. NAFLD pathology is mainly characterized by the deposition of fat particles in the liver, ballooning of hepatocytes, increase of inflammatory cells, and infiltration of liver tissue. It is the manifestation of systemic metabolic syndrome in liver tissue. Although NAFLD is a

benign disease, it may lead to liver fibrosis, liver cirrhosis, and even liver cancer [22]. NAFLD is becoming an important public health problem because it is now the main cause of chronic liver disease worldwide. About 80 million to 100 million adults suffer from NAFLD in the United States alone. Approximately 20% of NAFLD patients in the United States have nonalcoholic steatohepatitis (NASH) [23]. In addition, the prevalence of NAFLD is about 15% in adults while 1.3% of children and adolescents in China. The incidence rate of NAFLD in 2 nonalcoholic health volunteers in Shanghai was 6.1% for 2 years [24].

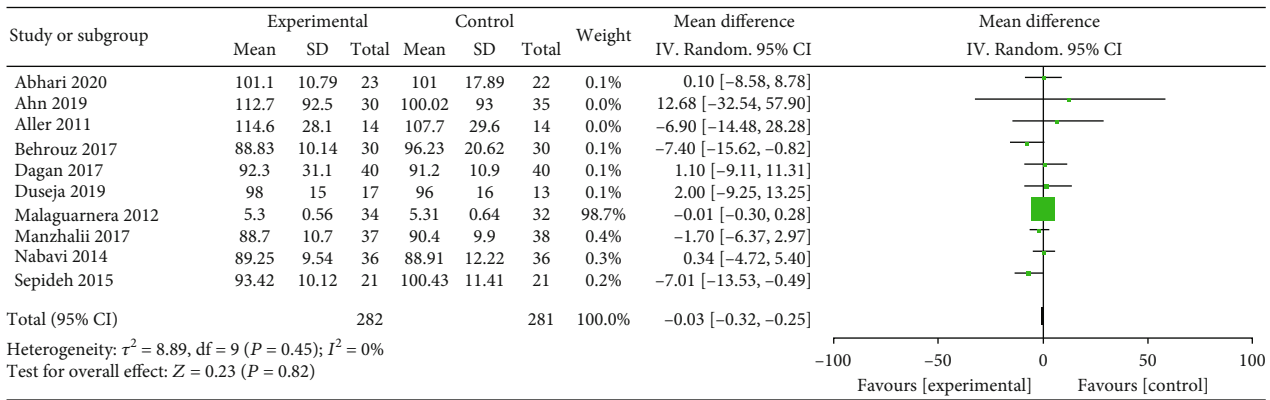
There are 10 to 100 trillion microorganisms in the human gut, called the gut microbiota, which combined with their genetic material, making up the gut microbiome [25]. The bacteria in the gut microbiota are dominated by *Gram-positive Firmicutes* and *Gram-negative Bacteroidetes* [26]. The gut microbiome maintains gut homeostasis by assisting in nutrient digestion, metabolism, immune function, and barrier protection and is therefore considered a functional organ. The exact function of the gut microbiota remains largely unknown. However, it plays an important role in the processing of complex indigestible polysaccharides from short-chain fatty acids, providing energy for the host, and also participating in the synthesis of vitamins, bile acids, and amino acids; the metabolism of drugs and toxins; and the integrity of the intestinal barrier. Although the gut microbiome is already formed at age 3 years, a variety of factors that may alter its diversity over the course of a person's life, including medications, geography, stress, and diet. Using 16S rRNA sequence analysis, differences in bacterial community structure were found in humans eating different diets [27]. Specifically, the term "dysbiosis" refers to any imbalance between beneficial and pathogenic bacteria or alterations in the taxonomic composition and function of the gut microbiota. In both mice and humans, the



(a)



(b)

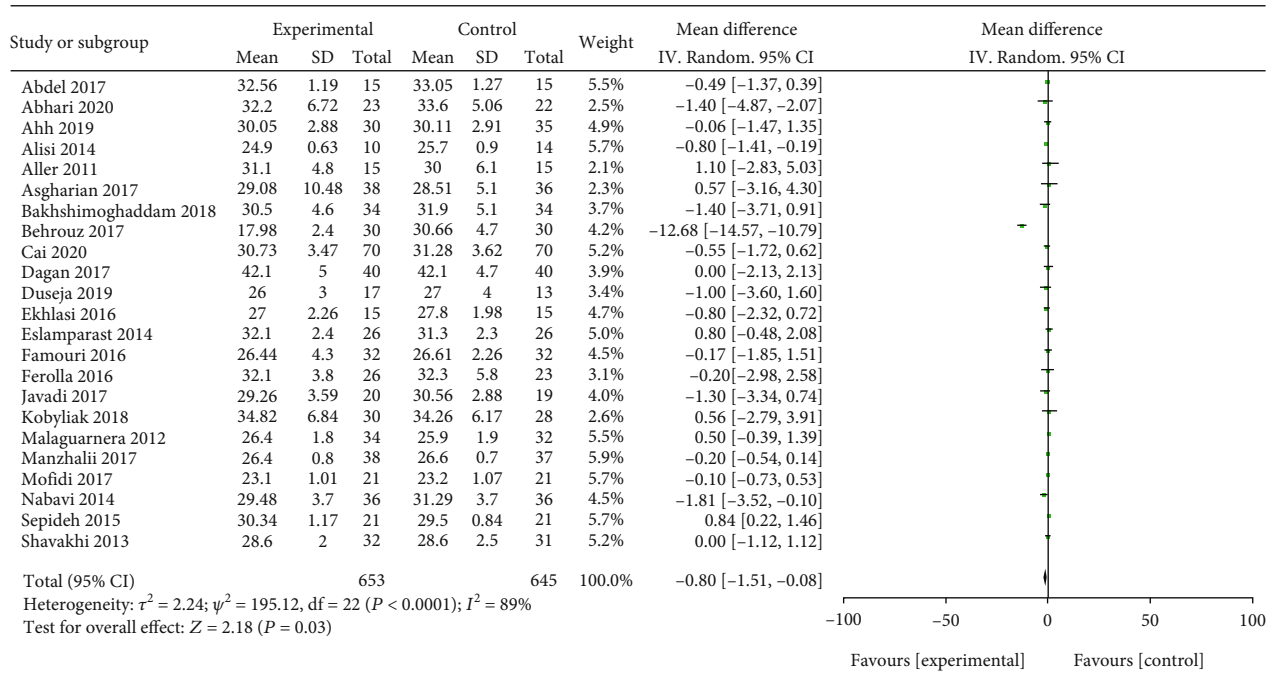


(c)

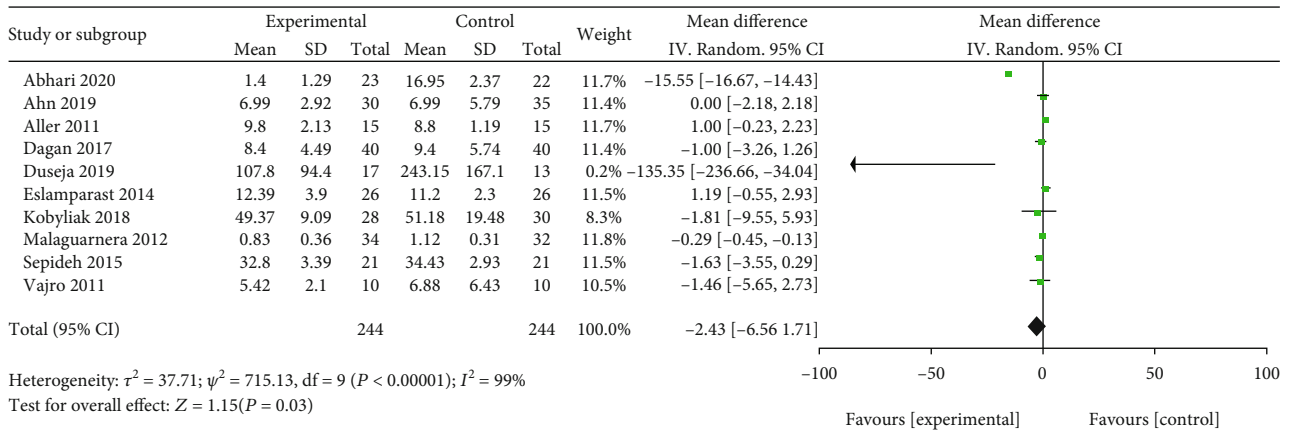
FIGURE 4: Forest plots of the effect of probiotics therapy on (a) IR, (b) insulin, and (c) GLU levels.

abundance of *Firmicutes* and *Bacteroidetes* varies with body fat composition, with an increase in *Firmicutes* and a decrease in *Bacteroidetes* when body fat content increases [28]. The study showed that germ-free mice, while consuming more food than conventional mice, had leaner physical characteristics and increased fat due to increased dietary energy intake after transferring fecal material from conventional mice. When humans are obese, dietary modification to restrict fat or carbohydrates shifts the gut microbiota from an “obese” to a “thin” phenotype [29]. The quantitative

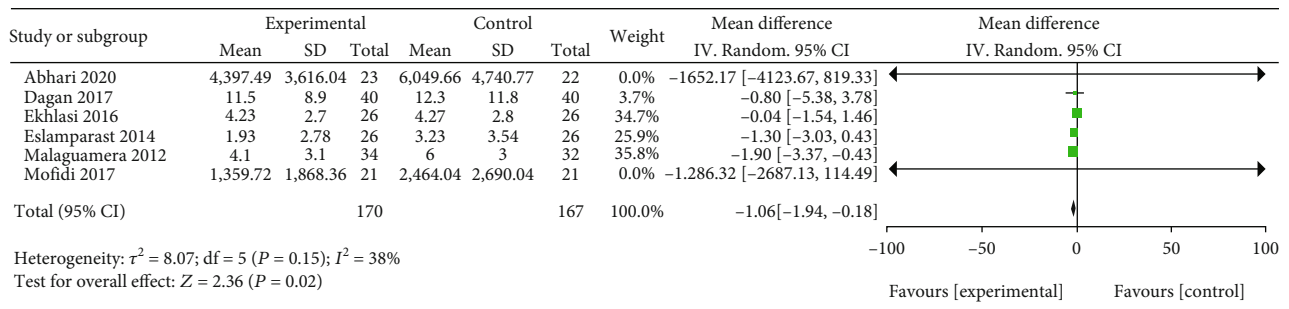
changes in gut microbial composition may independently associated with the development of NAFLD that caused progression to NASH and HCC. Therefore, species-specific biomes may reflect the staging of NAFLD. Although liver biopsy is still the gold standard for disease diagnosis, gut microbiota sequencing has become a noninvasive auxiliary test for NAFLD diagnosis. As a biomarker of disease phenotype or to provide prognostic value for the possible development of cirrhotic liver cancer, supplementation of probiotics to modulate the structure of the



(a)



(b)



(c)

FIGURE 5: Forest plots of the effect of probiotics therapy on (a) BMI, (b) TNF, and (c) CRP levels.

microbiota has become a potential therapeutic option for the treatment of NAFLD [30].

Abnormal ALT, AST, and GGT indicators are usually used as markers of liver damage, but only as general markers of liver damage, not markers of specific liver function. In

NAFLD patients, intestinal inflammation is thought to promote the translocation of bacteria and their products, which in turn stimulate Kupffer cells and stellate cells of the liver, promoting hepatic inflammation [31], leading to hepatocyte death and concomitant hepatic enzyme release, so elevated

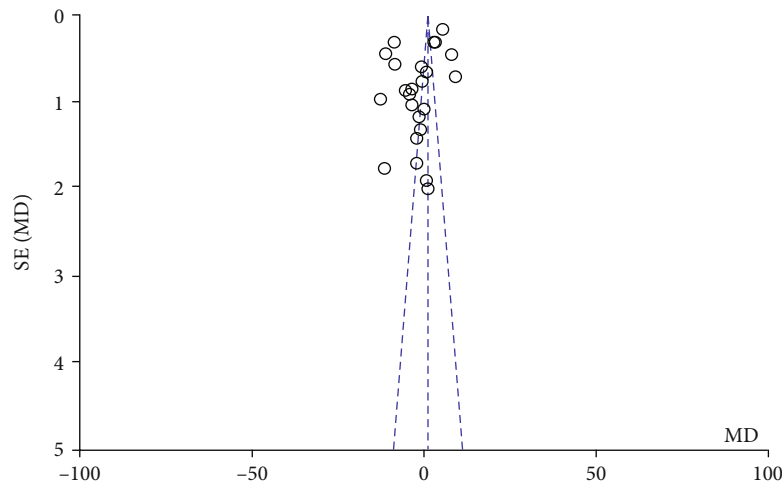


FIGURE 6: Funnel plots for publication bias analysis of the included articles.

liver enzymes are considered a reliable indicator of liver damage. In the present study, the improvement of liver function in NAFLD patients after probiotics treatment was a basic marker for evaluating the therapeutic effect by quantifying clinical liver function. In addition, the level of ALT, AST, and GGT in the probiotics group was significantly lower than that in the control group, with statistical significance.

Intrahepatic lipid accumulation exhibits the effect of increased lipolysis, hepatic free fatty acid and relatively low-density lipoprotein synthesis, decreased free fat acid oxidation, and triglyceride transport, leading to lipid accumulation in the liver [32]. Intrahepatic lipid accumulation interferes with hepatic sinusoidal microcirculation and the ability of hepatocytes to clear microbial and gut-derived danger signals, enhancing the reactivity of Kupffer cells, which are critical for the progression of NAFLD [33]. In the present study, the levels of LDL-C, HDL-C, and TG in the probiotics group were significantly lower than that in the control group. Therefore, using probiotics could improve NAFLD by improving lipid status and correcting lipid metabolism.

IR is also part of a preemptive mechanism in the pathogenesis of “two-hit” nonalcoholic fatty liver disease. The effectiveness of IR in the pathogenesis of NAFLD occurs through the following mechanisms: increased activation of sterol regulatory element binding and carbohydrate response element binding protein [34]. It was revealed that probiotics supplementation is helpful to improve IR and reduce insulin levels in patients with NAFLD in the present study. In addition, the level of GLU was significantly increased in the probiotics group which indicating that probiotics may help to increase the level of GLU.

The high prevalence of NAFLD in obese people shows that obesity is one of the most important factors related to the disease, and obesity plays an important role in the development of the disease to NASH. It has shown that specific microbial species are related to the reduction or increase of body weight. Previous studies demonstrated that surgical

resection of intra-abdominal fat can reverse liver IR and steatosis while [35]. Therefore, the clinical evaluation of BMI improvement after probiotics treatment is a useful index of treatment effect among NAFLD patients. During the progression of NAFLD, inflammation through the activation of nuclear factor kappa-B (NF- κ B) and mitogen-activated protein kinase (MAPK) pathways increases the activity of Kupffer cells and aggravates the progression of liver and systemic inflammation. Studies have shown that the severity of NAFLD is directly related to the levels of inflammatory markers such as TNF- α and CRP [36]. Elevated levels of free fatty acids in the liver provide a source of oxidative stress, which is an important cause of the development of steatosis to steatohepatitis. TNF- α is a pleiotropic cytokine that activates multiple signaling mechanisms leading to hepatocyte apoptosis, hepatic stellate cell activation, and hepatic cell aggregation [37]. In the present study, it was showed that the levels of BMI, TNF- α , and CRP are significantly induced in the probiotics group. The possible reason is that intrahepatic free fatty acids have direct cytotoxic effects, increasing lysosomal permeability and hepatocyte synthesis of TNF- α [38]. However, according to the probiotics therapy, the levels of free fatty acids were inhibited which leads to positive effect on NAFLD.

Although the molecular mechanism of probiotics has not been fully elucidated, current studies suggest that probiotics in the treatment of NAFLD may modulating the composition of intestinal flora, improving the permeability of intestinal mucosa, inhibiting the inflammatory response, and regulating the immune system by probiotics. It was found that several strains of *Bifidobacterium* may produce bacteriocin-like compounds that are toxic to both Gram-positive and Gram-negative bacteria. Furthermore, some secreted probiotic factors are able to inhibit the binding of pathogenic bacteria to specific receptors expressed on the epithelial surface except the antibacterial effects [39]. Several strains of *Lactobacillus* and *Bifidobacterium* are able to compete with and displace pathogenic bacteria including

Staphylococcus aureus, *Clostridium difficile*, *Enterobacter aerogenes*, *Listeria Special bacteria*, and *Salmonella enterica* [40]. Therefore, probiotics may improve the intestinal ecology and microbial composition to compete and replace with pathogenic bacteria and prevent the overgrowth of small intestinal bacteria. As the incidence of NAFLD continues to rise, research into therapeutic approaches to mitigate the occurrence and progression of NAFLD remains essential [41]. More and more researches are expanding our understanding of the mechanisms by which NAFLD is affected by gut microbes especially beneficial bacteria [42]. However, further well-designed prospective clinical studies combined with preclinical models are needed to establish pathogenic microbe-host interactions in the pathogenesis of NAFLD. Therefore, we hope that the intestinal microbiome will become an indispensable part of personalized medicine, especially in multifactorial chronic metabolic diseases such as NAFLD in the next few years. Previous research shows that dietary and lifestyle changes have been shown to improve NAFLD disease markers, but patient compliance has been problematic, making treatment limited [43]. With increasing research on the gut microbiota and its role in obesity and liver disease, fecal microbial signatures can serve as noninvasive diagnostic tools or provide prognostic value [44], and future research directions are to use the gut microbiota as a treatment options for NAFLD.

5. Conclusion

Gut dysbiosis is a risk factor that may influence and contribute to the pathogenesis of NAFLD. The ability of probiotics to reverse gut dysbiosis has generated increasing interest in investigating probiotics as an alternative treatment option for patients with NAFLD. In the present study, it was found that probiotics supplementation has a significant positive effect on the treatment of NAFLD. In addition, probiotics supplementation plays a positive role in improving patients' lipid metabolism, reducing IR, regulating body immunity, improving liver function, and delaying disease progression in the treatment of NAFLD. In this growing area of NAFLD-related therapeutic research, further research is needed to more clearly elucidate the role of probiotics.

6. Limitations of the Study

(1) All the included literatures and studies were randomized controlled trials, but some trials did not accurately explain the random sampling method, distribution scheme, and blind method. (2) The gold standard for the diagnosis of NAFLD is normal liver biopsy, but few studies are based on ultrasound, MRI, and biochemical indexes. (3) A few of included articles have obvious heterogeneity, which may be related to the information bias in the process of data collection, the area of the included population, the type, the dose, the treatment time, and the duration of the disease. Therefore, the therapeutic effect of probiotics in the treatment of NAFLD still needs to be confirmed by further randomized double-blind controlled trials. (4) The actual action mechanism of probiotics in NAFLD and the efficacy of NAFLD

in children and adults are different. (5) The comparison of the efficacy of effective probiotics, the specific action targets of each probiotic, and the long-term efficacy have not been elucidated well.

Abbreviations

ALT:	Alanine aminotransferase
AST:	Aspartate aminotransferase
BMI:	Body mass index
GLU:	Blood glucose
CRP:	C-reactive protein
IR:	Insulin resistance
LDL-C:	Low density lipoprotein
MAFLD:	Metabolic fatty liver disease
GGT:	γ -Glutamyltransferase
NASH:	Nonalcoholic steatohepatitis
NAFL:	Nonalcoholic fatty liver
NAFLD:	Nonalcoholic fatty liver disease
TNF- α :	Tumor necrosis factor- α
TG:	Triglyceride.

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest

All authors declare that there are no any commercial or associative interests that represent competing interests in connection with the work submitted.

Authors' Contributions

Jingxin Mao conceived and designed the research. Yuanshe Huang, Xiaodong Wang, and Jingxin Mao conducted the statistical analysis and wrote the paper. Lai Zhang, Ke Zheng, and Jie Xiong abstracted the total data from the included articles in the meta-analysis. Jing Li, Chunlei Cong, and Zhaomiao Gong finished the data analysis process in the included articles. Yuanshe Huang and Xiaodong Wang contributed equally to this work. Jingxin Mao and Zhaomiao Gong contributed equally to the study. All authors contributed to the manuscript revision and read and approved the submitted version.

Acknowledgments

This work was supported by the 2020 Ministerial Project of China (No. 2020YYCXQCSJ050), the Guizhou Province Department of Education Project (QJH KY [2020] 063 [2013]130), the Science and Technology Bureau of Anshun (grant no. ASKN (2020) 07), and the Anshun University/Innovation Center for Efficient Agricultural of Guizhou Mountain Characteristics/Branch of learning in Agricultural Resources and Environment, respectively.

References

- [1] F. Idalsoaga, A. V. Kulkarni, O. Y. Mousa, M. Arrese, and J. P. Arab, "Non-alcoholic fatty liver disease and alcohol-related liver disease: two intertwined entities," *Frontiers in Medicine*, vol. 7, p. 448, 2020.
- [2] K. Wójcik-Cichy, E. Koślińska-Berkan, and A. Piekarska, "The influence of NAFLD on the risk of atherosclerosis and cardiovascular diseases," *Clinical and Experimental Hepatology*, vol. 4, no. 1, pp. 1–6, 2018.
- [3] E. Hashimoto, K. Tokushige, and J. Ludwig, "Diagnosis and classification of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis: current concepts and remaining challenges," *Hepatology Research*, vol. 45, no. 1, pp. 20–28, 2015.
- [4] J. Ertle, A. Dechêne, J. P. Sowa et al., "Non-alcoholic fatty liver disease progresses to hepatocellular carcinoma in the absence of apparent cirrhosis," *International Journal of Cancer*, vol. 128, no. 10, pp. 2436–2443, 2011.
- [5] M. Doulberis, G. Kotronis, D. Gialamprinou, J. Kountouras, and P. Katsinelos, "Non-alcoholic fatty liver disease: an update with special focus on the role of gut microbiota," *Metabolism*, vol. 71, pp. 182–197, 2017.
- [6] K. I. Zheng, J. G. Fan, J. P. Shi et al., "From NAFLD to MAFLD: a "redefining" moment for fatty liver disease," *Chinese Medical Journal*, vol. 133, no. 19, pp. 2271–2273, 2020.
- [7] A. Takaki, D. Kawai, and K. Yamamoto, "Multiple hits, including oxidative stress, as pathogenesis and treatment target in non-alcoholic steatohepatitis (NASH)," *International Journal of Molecular Sciences*, vol. 14, no. 10, pp. 20704–20728, 2013.
- [8] P. Puri and A. J. Sanyal, "The intestinal microbiome in nonalcoholic fatty liver disease," *Clinics in Liver Disease*, vol. 22, no. 1, pp. 121–132, 2018.
- [9] V. T. Samuel and G. I. Shulman, "Nonalcoholic fatty liver disease, insulin resistance, and ceramides," *New England Journal of Medicine*, vol. 381, no. 19, pp. 1866–1869, 2019.
- [10] Y. L. Fang, H. Chen, C. L. Wang, and L. Liang, "Pathogenesis of non-alcoholic fatty liver disease in children and adolescence: from "two hit theory" to "multiple hit model"," *World Journal of Gastroenterology*, vol. 24, no. 27, pp. 2974–2983, 2018.
- [11] A. Albillos, A. De Gottardi, and M. Rescigno, "The gut-liver axis in liver disease: pathophysiological basis for therapy," *Journal of Hepatology*, vol. 72, no. 3, pp. 558–577, 2020.
- [12] S. X. Fan, J. Wang, Q. Li, Y. S. Li, W. X. Guan, and J. S. Li, "Mechanism of gut-microbiota-liver axis in the pathogenesis of intestinal failure-associated liver disease," *Zhonghua wei Chang Wwai ke za zhi= Chinese Journal of Gastrointestinal Surgery*, vol. 24, no. 1, pp. 94–100, 2021.
- [13] S. Carding, K. Verbeke, D. T. Vipond, B. M. Corfe, and L. J. Owen, "Dysbiosis of the gut microbiota in disease," *Microbial Ecology in Health and Disease*, vol. 26, p. 26191, 2015.
- [14] P. Bist and S. Choudhary, "Impact of heavy metal toxicity on the gut microbiota and its relationship with metabolites and future probiotics strategy: A review," *Biological Trace Element Research*, 2022.
- [15] K. G. Parhofer, C. Otto, H. C. Geiss, E. Laubach, and B. Göke, "Effect of pioglitazone on lipids in well controlled patients with diabetes mellitus type 2 - results of a pilot study," *Experimental and Clinical Endocrinology & Diabetes*, vol. 113, no. 1, pp. 49–52, 2005.
- [16] C. P. Champagne, A. G. da Cruz, and M. Daga, "Strategies to improve the functionality of probiotics in supplements and foods," *Current Opinion in Food Science*, vol. 22, pp. 160–166, 2018.
- [17] M. Peng and D. Biswas, "Short chain and polyunsaturated fatty acids in host gut health and foodborne bacterial pathogen inhibition," *Critical Reviews in Food Science and Nutrition*, vol. 57, no. 18, pp. 3987–4002, 2017.
- [18] M. J. Page and D. Moher, "Evaluations of the uptake and impact of the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and extensions: a scoping review," *Systematic Reviews*, vol. 6, no. 1, pp. 1–14, 2017.
- [19] G. Jiang, C. Sun, X. Wang et al., "Hepatoprotective mechanism of Silybum marianum on nonalcoholic fatty liver disease based on network pharmacology and experimental verification," *Bioengineered*, vol. 13, no. 3, pp. 5216–5235, 2022.
- [20] J. Mao, Q. Zhang, H. Zhang, K. Zheng, R. Wang, and G. Wang, "Risk factors for lymph node metastasis in papillary thyroid carcinoma: a systematic review and meta-analysis," *Frontiers in Endocrinology*, vol. 11, p. 265, 2020.
- [21] N. N. Than and P. N. Newsome, "A concise review of non-alcoholic fatty liver disease," *Atherosclerosis*, vol. 239, no. 1, pp. 192–202, 2015.
- [22] S. Milić and D. Štimac, "Nonalcoholic fatty liver disease/steatohepatitis: epidemiology, pathogenesis, clinical presentation and treatment," *Digestive Diseases*, vol. 30, no. 2, pp. 158–162, 2012.
- [23] Z. M. Younossi, R. Loomba, Q. M. Anstee et al., "Diagnostic modalities for nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, and associated fibrosis," *Hepatology*, vol. 68, no. 1, pp. 349–360, 2018.
- [24] J. G. Fan, Q. Zhou, and Q. H. Wo, "Effect of body weight mass and its change on the incidence of nonalcoholic fatty liver disease," *Zhonghua Ggan Zang Bing za zhi= Zhonghua Gan-zangbing Zazhi= Chinese Journal of Hepatology*, vol. 18, no. 9, pp. 676–679, 2010.
- [25] J. A. Gilbert, M. J. Blaser, J. G. Caporaso, J. K. Jansson, S. V. Lynch, and R. Knight, "Current understanding of the human microbiome," *Nature Medicine*, vol. 24, no. 4, pp. 392–400, 2018.
- [26] P. B. Eckburg, E. M. Bik, C. N. Bernstein et al., "Diversity of the human intestinal microbial flora," *Science*, vol. 308, no. 5728, pp. 1635–1638, 2005.
- [27] E. R. Davenport, J. G. Sanders, S. J. Song, K. R. Amato, A. G. Clark, and R. Knight, "The human microbiome in evolution," *BMC Biology*, vol. 15, no. 1, pp. 1–12, 2017.
- [28] F. Bäckhed, H. Ding, T. Wang et al., "The gut microbiota as an environmental factor that regulates fat storage," *Proceedings of the National Academy of Sciences*, vol. 101, no. 44, pp. 15718–15723, 2004.
- [29] R. E. Ley, P. J. Turnbaugh, S. Klein, and J. I. Gordon, "Human gut microbes associated with obesity," *Nature*, vol. 444, no. 7122, pp. 1022–1023, 2006.
- [30] A. M. Thomas, P. Manghi, F. Asnicar et al., "Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation," *Nature Medicine*, vol. 25, no. 4, pp. 667–678, 2019.
- [31] E. Org, Y. Blum, S. Kasela et al., "Relationships between gut microbiota, plasma metabolites, and metabolic syndrome traits in the METSIM cohort," *Genome Biology*, vol. 18, no. 1, p. 70, 2017.

- [32] C. Ress and S. Kaser, "Mechanisms of intrahepatic triglyceride accumulation," *World Journal of Gastroenterology*, vol. 22, no. 4, pp. 1664–1673, 2016.
- [33] S. A. Polyzos, K. N. Aronis, J. Kountouras, D. D. Raptis, M. F. Vasiloglou, and C. S. Mantzoros, "Circulating leptin in non-alcoholic fatty liver disease: a systematic review and meta-analysis," *Diabetologia*, vol. 59, no. 1, pp. 30–43, 2016.
- [34] E. F. Khaleel, G. A. Abdel-Aleem, and D. G. Mostafa, "Resveratrol improves high-fat diet induced fatty liver and insulin resistance by concomitantly inhibiting proteolytic cleavage of sterol regulatory element-binding proteins, free fatty acid oxidation, and intestinal triglyceride absorption," *Canadian Journal of Physiology and Pharmacology*, vol. 96, no. 2, pp. 145–157, 2018.
- [35] V. Menon, X. Zhi, T. Hossain et al., "The contribution of visceral fat to improved insulin signaling in Ames dwarf mice," *Aging Cell*, vol. 13, no. 3, pp. 497–506, 2014.
- [36] S. Abiru, K. Migita, Y. Maeda et al., "Serum cytokine and soluble cytokine receptor levels in patients with non-alcoholic steatohepatitis," *Liver International*, vol. 26, no. 1, pp. 39–45, 2006.
- [37] M. Manco, M. Marcellini, G. Giannone, and V. Nobili, "Correlation of serum TNF- α levels and histologic liver injury scores in pediatric nonalcoholic fatty liver disease," *American Journal of Clinical Pathology*, vol. 127, no. 6, pp. 954–960, 2007.
- [38] G. Svegliati-Baroni, C. Candelaresi, S. Saccomanno et al., "A model of insulin resistance and nonalcoholic steatohepatitis in rats: role of peroxisome proliferator-activated receptor- α and n-3 polyunsaturated fatty acid treatment on liver injury," *The American Journal of Pathology*, vol. 169, no. 3, pp. 846–860, 2006.
- [39] M. C. Collado, M. Hernandez, and Y. Sanz, "Production of bacteriocin-like inhibitory compounds by human fecal Bifidobacterium strains," *Journal of Food Protection*, vol. 68, no. 5, pp. 1034–1040, 2005.
- [40] R. Tojo, A. Suárez, M. G. Clemente et al., "Intestinal microbiota in health and disease: role of bifidobacteria in gut homeostasis," *World Journal of Gastroenterology: WJG*, vol. 20, no. 41, pp. 15163–15176, 2014.
- [41] C. Estes, H. Razavi, R. Loomba, Z. Younossi, and A. J. Sanyal, "Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease," *Hepatology*, vol. 67, no. 1, pp. 123–133, 2018.
- [42] S. R. Sharpton, B. Schnabl, R. Knight, and R. Loomba, "Current concepts, opportunities, and challenges of gut microbiome-based personalized medicine in nonalcoholic fatty liver disease," *Cell Metabolism*, vol. 33, no. 1, pp. 21–32, 2021.
- [43] A. Mazzotti, M. T. Caletti, L. Brodosi et al., "An internet-based approach for lifestyle changes in patients with NAFLD: two-year effects on weight loss and surrogate markers," *Journal of Hepatology*, vol. 69, no. 5, pp. 1155–1163, 2018.
- [44] T. Hrnčir, L. Hrnčirova, M. Kverka et al., "Gut microbiota and NAFLD: pathogenetic mechanisms, microbiota signatures, and therapeutic interventions," *Microorganisms*, vol. 9, no. 5, p. 957, 2021.
- [45] S. M. A. Monem, "Probiotic therapy in patients with nonalcoholic steatohepatitis in Zagazig University hospitals," *Euroasian journal of hepato-gastroenterology*, vol. 7, no. 1, pp. 101–106, 2017.
- [46] K. Abhari, S. Saadati, Z. Yari et al., "The effects of Bacillus coagulans supplementation in patients with non-alcoholic fatty liver disease: a randomized, placebo-controlled, clinical trial," *Clinical nutrition ESPEN*, vol. 39, pp. 53–60, 2020.
- [47] S. B. Ahn, D. W. Jun, B. K. Kang, J. H. Lim, S. Lim, and M. J. Chung, "Randomized, double-blind, placebo-controlled study of a multispecies probiotic mixture in nonalcoholic fatty liver disease," *Scientific Reports*, vol. 9, no. 1, pp. 1–9, 2019.
- [48] A. Alisi, G. Bedogni, G. Baviera et al., "Randomised clinical trial: the beneficial effects of VSL#3 in obese children with non-alcoholic steatohepatitis," *Alimentary Pharmacology & Therapeutics*, vol. 39, no. 11, pp. 1276–1285, 2014.
- [49] R. Aller, D. A. De Luis, O. Izaola et al., "Effect of a probiotic on liver aminotransferases in nonalcoholic fatty liver disease patients: a double blind randomized clinical trial," *European Review for Medical and Pharmacological Sciences*, vol. 15, no. 9, pp. 1090–1095, 2011.
- [50] A. Asgharian, V. Mohammadi, Z. Gholi, A. Esmailzadeh, A. Feizi, and G. Askari, "The effect of synbiotic supplementation on body composition and lipid profile in patients with NAFLD: a randomized, double blind, placebo-controlled clinical trial study," *IRANIAN RED CRESCENT MEDICAL JOURNAL (IRCMJ)*, vol. 19, no. 4, 2017.
- [51] F. Bakhshimoghaddam, K. Shateri, M. Sina, M. Hashemian, and M. Alizadeh, "Daily consumption of synbiotic yogurt decreases liver steatosis in patients with nonalcoholic fatty liver disease: a randomized controlled clinical trial," *The Journal of Nutrition*, vol. 148, no. 8, pp. 1276–1284, 2018.
- [52] V. Behrouz, S. Jazayeri, N. Aryaeian, M. J. Zahedi, and F. Hosseini, "Effects of probiotic and prebiotic supplementation on leptin, adiponectin, and glycemic parameters in non-alcoholic fatty liver disease: a randomized clinical trial," *Middle East Journal of Digestive Diseases*, vol. 9, no. 3, pp. 150–157, 2017.
- [53] G. S. Cai, H. Su, and J. Zhang, "Protective effect of probiotics in patients with non-alcoholic fatty liver disease," *Medicine*, vol. 99, no. 32, p. e21464, 2020.
- [54] S. Sherf-Dagan, S. Zelber-Sagi, G. Zilberman-Schapira et al., "Probiotics administration following sleeve gastrectomy surgery: a randomized double-blind trial," *International Journal of Obesity*, vol. 42, no. 2, pp. 147–155, 2018.
- [55] A. Duseja, S. K. Acharya, M. Mehta et al., "High potency multi-strain probiotic improves liver histology in non-alcoholic fatty liver disease (NAFLD): a randomised, double-blind, proof of concept study," *BMJ Open Gastroenterology*, vol. 6, no. 1, p. e000315, 2019.
- [56] G. Ekhlasi, R. K. Mohammadi, S. Agah et al., "Do symbiotic and vitamin E supplementation have favorable effects in nonalcoholic fatty liver disease? A randomized, double-blind, placebo-controlled trial," *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences*, vol. 21, 2016.
- [57] T. Eslamparast, H. Poustchi, F. Zamani, M. Sharafkhan, R. Malekzadeh, and A. Hekmatdoost, "Synbiotic supplementation in nonalcoholic fatty liver disease: a randomized, double-blind, placebo-controlled pilot study," *The American Journal of Clinical Nutrition*, vol. 99, no. 3, pp. 535–542, 2014.
- [58] F. Famouri, Z. Shariat, M. Hashemipour, M. Keikha, and R. Kelishadi, "Effects of probiotics on nonalcoholic fatty liver disease in obese children and adolescents," *Journal of Pediatric Gastroenterology and Nutrition*, vol. 64, no. 3, pp. 413–417, 2017.
- [59] S. M. Ferolla, C. A. Couto, L. Costa-Silva et al., "Beneficial effect of synbiotic supplementation on hepatic steatosis and

- anthropometric parameters, but not on gut permeability in a population with nonalcoholic steatohepatitis,” *Nutrients*, vol. 8, no. 7, p. 397, 2016.
- [60] L. Javadi, M. Ghavami, M. Khoshbaten, A. Safaiyan, A. Barzegari, and B. P. Gargari, “The effect of probiotic and/or prebiotic on liver function tests in patients with nonalcoholic fatty liver disease: a double blind randomized clinical trial,” *Iranian Red Crescent Medical Journal*, vol. 19, no. 4, p. e46017, 2017.
- [61] N. Kobyliak, L. Abenavoli, G. Mykhalchyshyn et al., “A multi-strain probiotic reduces the fatty liver index, cytokines and aminotransferase levels in NAFLD patients: Evidence from a Randomized Clinical Trial,” *Journal of Gastrointestinal and Liver Diseases*, vol. 27, no. 1, pp. 41–49, 2018.
- [62] M. Malaguarnera, M. Vacante, T. Antic et al., “Bifidobacterium longum with fructo-oligosaccharides in patients with non alcoholic steatohepatitis,” *Digestive Diseases and Sciences*, vol. 57, no. 2, pp. 545–553, 2012.
- [63] E. Manzhali, O. Virchenko, T. Falalyeyeva, T. Beregova, and W. Stremmel, “Treatment efficacy of a probiotic preparation for non-alcoholic steatohepatitis: a pilot trial,” *Journal of Digestive Diseases*, vol. 18, no. 12, pp. 698–703, 2017.
- [64] F. Mofidi, H. Poustchi, Z. Yari et al., “Synbiotic supplementation in lean patients with non-alcoholic fatty liver disease: a pilot, randomised, double-blind, placebo-controlled, clinical trial,” *British Journal of Nutrition*, vol. 117, no. 5, pp. 662–668, 2017.
- [65] S. Nabavi, M. Rafraf, M. H. Somi, A. Homayouni-Rad, and M. Asghari-Jafarabadi, “Effects of probiotic yogurt consumption on metabolic factors in individuals with nonalcoholic fatty liver disease,” *Journal of Dairy Science*, vol. 97, no. 12, pp. 7386–7393, 2014.
- [66] A. Sepideh, P. Karim, A. Hossein et al., “Effects of multistrain probiotic supplementation on glycemic and inflammatory indices in patients with nonalcoholic fatty liver disease: a double-blind randomized clinical trial,” *Journal of the American College of Nutrition*, vol. 35, no. 6, pp. 500–505, 2016.
- [67] A. Shavakhi, M. Minakari, H. Firouzian, R. Assali, A. Hekmatdoost, and G. Ferns, “Effect of a probiotic and metformin on liver aminotransferases in non-alcoholic steatohepatitis: a double blind randomized clinical trial,” *International Journal of Preventive Medicine*, vol. 4, no. 5, pp. 531–537, 2013.
- [68] P. Vajro, C. Mandato, M. R. Licenziati et al., “Effects of Lactobacillus rhamnosus strain GG in pediatric obesity-related liver disease,” *Journal of Pediatric Gastroenterology and Nutrition*, vol. 52, no. 6, pp. 740–743, 2011.