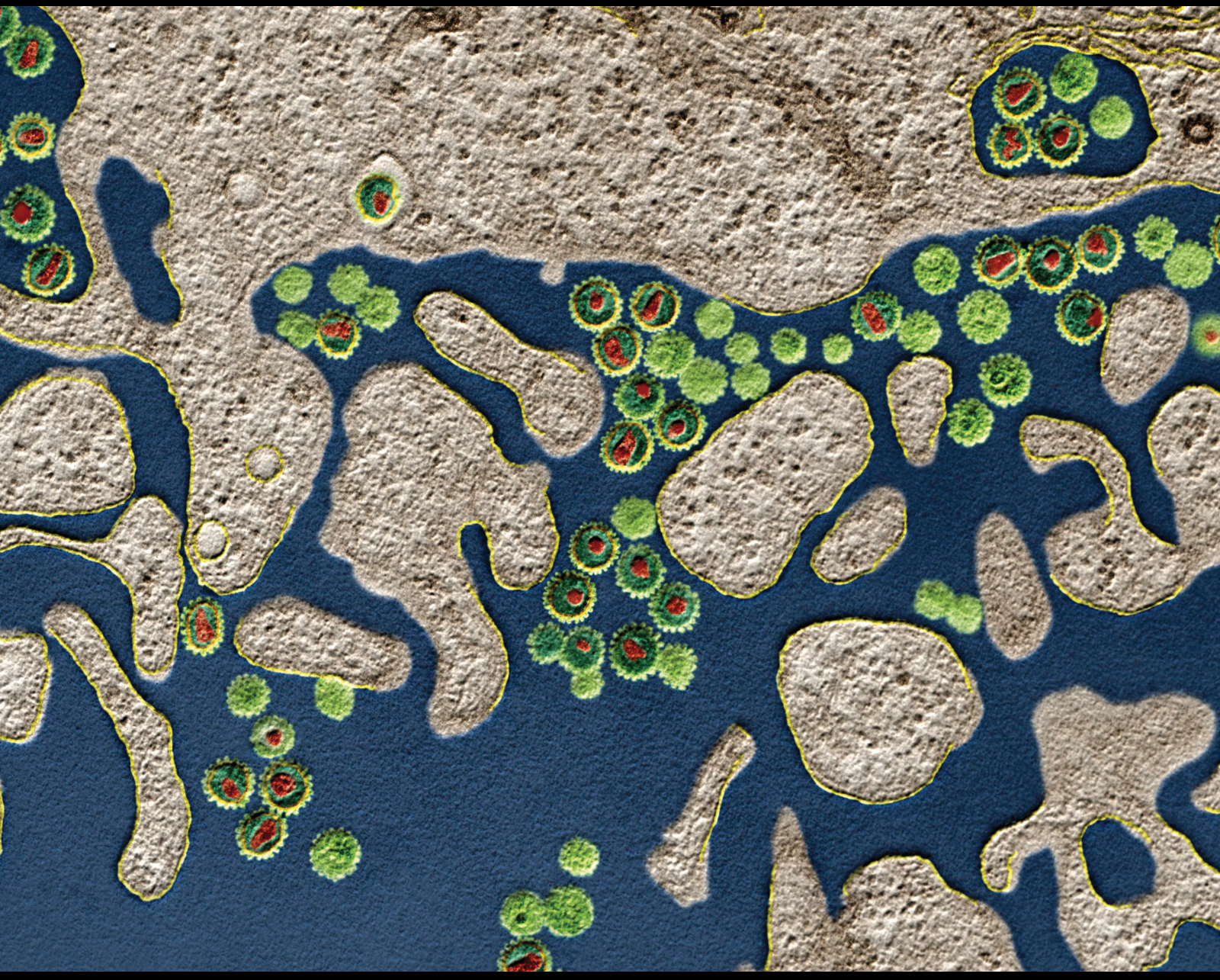
Journal of Immunology Research

The Crosstalk Between Circadian Clocks, Immunity, and Tumour Genesis

Lead Guest Editor: Mingyi Zhao
Guest Editors: Ying-Jie Zhang, Pan Chen, and Jialiang Liang



The Crosstalk Between Circadian Clocks, Immunity, and Tumour Genesis

Journal of Immunology Research

The Crosstalk Between Circadian Clocks, Immunity, and Tumour Genesis

Lead Guest Editor: Mingyi Zhao
Guest Editors: Ying-Jie Zhang, Pan Chen, and Jialiang Liang

Copyright © 2021 Hindawi Limited. All rights reserved.

This is a special issue published in "Journal of Immunology Research." 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

Douglas C. Hooper, USA
Senthamil R. Selvan, USA
Jacek Tabarkiewicz, Poland
Baohui Xu, USA

Academic Editors

Nitin Amdare, USA
Lalit Batra, USA
Kurt Blaser, Switzerland
Dimitrios P. Bogdanos, Greece
Srinivasa Reddy Bonam, USA
Carlo Cavaliere, Italy
Cinzia Ciccacci, Italy
Robert B. Clark, USA
Marco De Vincentiis, Italy
M. Victoria Delpino, Argentina
Roberta Antonia Diotti, Italy
Lihua Duan, China
Nejat K. Egilmez, USA
Theodoros Eleftheriadis, Greece
Eyad Elkord, United Kingdom
Weirong Fang, China
Elizabeth Soares Fernandes, Brazil
Steven E. Finkelstein, USA
JING GUO, USA
Luca Gattinoni, USA
Alvaro González, Spain
Manish Goyal, USA
Qingdong Guan, Canada
Theresa Hautz, Austria
Weicheng Hu, China
Giannicola Iannella, Italy
Juraj Ivanyi, United Kingdom
Ravirajsinh Jadeja, USA
Peirong Jiao, China
Youmin Kang, China
Sung Hwan Ki, Republic of Korea
Bogdan Kolarz, Poland
Vijay Kumar, USA
Esther Maria Lafuente, Spain
Natalie Lister, Australia
Daniele Maria-Ferreira, Saint Vincent and the Grenadines
Eiji Matsuura, Japan
Juliana Melgaço, Brazil
Cinzia Milito, Italy
Prasenjit Mitra, India
Chikao Morimoto, Japan
Paulina Niedźwiedzka-Rystwej, Poland
Enrique Ortega, Mexico
Felipe Passero, Brazil
Anup Singh Pathania, USA
Keshav Raj Paudel, Australia
Patrice Xavier Petit, France
Luis Alberto Ponce-Soto, Peru
Massimo Ralli, Italy
Pedro A. Reche, Spain
Eirini Rigopoulou, Greece
Ilaria Roato, Italy
Suyasha Roy, India
Francesca Santilli, Italy
Takami Sato, USA
Rahul Shivahare, USA
Arif Siddiqui, Saudi Arabia
Amar Singh, USA
Benoit Stijlemans, Belgium
Hiroshi Tanaka, Japan
Bufu Tang, China
Samanta Taurone, Italy
Mizue Terai, USA
Ban-Hock Toh, Australia
Shariq M. Usmani, USA
Ran Wang, China
Shengjun Wang, China
Paulina Wlasiuk, Poland
Zhipeng Xu, China
Xiao-Feng Yang, USA
Dunfang Zhang, China
Qiang Zhang, USA
Qianxia Zhang, USA
Bin Zhao, China
Jixin Zhong, USA
Lele Zhu, China

Contents

Hindawi
Journal of Immunology Research
Volume 2021, Article ID 1893882, 7 pages
https://doi.org/10.1155/2021/1893882

Research Article

Association between Sleep Traits and Lung Cancer: A Mendelian Randomization Study

Jie Wang,[1,2] **Haibo Tang**,[3] **Yumei Duan**,[4] **Siyu Yang**,[5] **and Jian An**[6]

[1]*Health Management Center, The Third Xiangya Hospital, Central South University, Changsha 410013, China*
[2]*Department of Cardiology, The Third Xiangya Hospital, Central South University, Changsha 410013, China*
[3]*Department of Metabolic and Bariatric Surgery, The Third Xiangya Hospital, Central South University, Changsha 410013, China*
[4]*Department of Pathology, Xiangya Hospital, Central South University, Changsha 410008, China*
[5]*Suzhou Science and Technology Town Foreign Language School, China*
[6]*Department of Respiratory Medicine, Xiangya Hospital, Central South University, Changsha 410008, China*

Correspondence should be addressed to Jian An; 175162349@qq.com

Received 5 May 2021; Revised 15 May 2021; Accepted 4 June 2021; Published 21 June 2021

Academic Editor: Ilaria Roato

Copyright © 2021 Jie 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.

Multidimensional sleep trait, which is related to circadian rhythms closely, affects some cancers predominantly, while the relationship between sleep and lung cancer is rarely illustrated. We aimed to investigate whether sleep is causally associated with risk of lung cancer, through a two-sample Mendelian randomization study. The main analysis used publicly available GWAS summary data from two large consortia (UK Biobank and International Lung Cancer Consortium). Two-sample Mendelian randomization (MR) analysis was used to examine whether chronotype, getting up in the morning, sleep duration, nap during the day, or sleeplessness was causally associated with the risk of lung cancer. Additionally, multivariate MR analysis was also conducted to estimate the direct effects between sleep traits and lung cancer risks independent of smoking status including pack years of smoking or current tobacco smoking. There was no evidence of causal association between chronotype, getting up in the morning, or nap during the day and lung cancer. Sleeplessness was associated with higher risk of lung adenocarcinoma (odds ratio 5.75, 95% confidence intervals 2.12-15.65), while sleep duration played a protective role in lung cancer (0.46, 0.26-0.83). In multivariate MR analysis, sleeplessness and sleep duration remained to have similar results. In conclusion, we found robust evidence for effect of sleeplessness on lung adenocarcinoma risk and inconsistent evidence for a protective effect of sleep duration on lung cancer risk.

1. Introduction

Lung cancer, which accounts for 11.6% of all newly diagnosed cancer cases and 18.4% of cancer-related deaths [1], brings a growing global burden of disease. Smoking has been identified as the most common risk factor for lung cancer, and a large number of epidemiological researches support this connection [2–4]. Smoking cessation before middle age can effectively decrease lung cancer risk. However, more and more nonsmokers were diagnosed with lung cancer over the past decades [5–7]. Based on this fact, attention has been focused on modified lifestyle risk factors other than smoking, such as sleep.

Many studies have shown that sleep plays an important role in cancer by affecting circadian rhythms, especially in breast cancer [8–11]. Nevertheless, only limited observational studies illustrated associations between sleep duration and lung cancer with inconsistent results [12–16]. These inconsistent results from epidemiological studies tend to be biased by small sample size, insufficient follow-up, and many unmeasured confounding, making inaccurate causation. Meanwhile, fewer studies have examined the relationship between sleep and lung cancer at the genetic level.

Mendelian randomization (MR) can use genetic variants that are associated robustly with exposure as instrumental variables to evaluate causal effects between the modifiable

risk factors and the diseases [17, 18]. The selected instrumental variables used in MR must meet three important assumptions [19] including the following: (1) SNP should be associated with sleep traits, (2) SNP should not be associated with confounding, and (3) SNP must influence lung cancer through exposure without direct association. Thus, this approach may avoid measurement error, confounding, and reverse causation that always exist in conventional clinical studies.

Furthermore, sleep is a multidimensional concept, including chronotype, getting up in the morning, sleep duration, nap during the day, and sleeplessness. Therefore, the exploration of association between sleep and lung cancer should not be finite to sleep duration. Based on the limited evidence for effects of sleep traits on lung cancer and the significant association between unfavorable sleep duration and lung function [20], we aimed to conduct a two-sample MR study to estimate the causal inferences between sleep traits and lung cancer risks.

2. Materials and Methods

2.1. GWAS Data on Exposure. Our exposure data were extracted from the UK Biobank, a large cohort study with deep genetic and phenotypic data collected on more than 500,000 individuals from across the United Kingdom [21]. Genome-wide association study (GWAS) of chronotype, getting up in the morning, sleep duration, nap during the day, sleeplessness/insomnia, pack years of smoking, and current tobacco smoking was performed among individuals of European ancestry ($n = 413,343$-$462,434$). With statistically significant threshold [$P < 5 \times 10^{-8}$; linkage disequilibrium (LD) $r^2 < 0.001$, LD distance > 10,000 kb], we identified single nucleotide polymorphisms (SNPs) robustly associated with sleep traits to generate genetic instruments. F statistic represents the strength of relationship between SNPs and sleep traits. It is related to the explained variance for exposure (R^2), sample size (n), and number of SNPs (k) by the formula $F = [(n - k - 1)/k]/[R^2/(1 - R^2)]$. Generally, $F > 10$ indicating that selected SNPs may strongly predict sleep traits [22].

2.2. GWAS Data on Outcome. GWAS summary data of lung cancer were extracted from the International Lung Cancer Consortium (ILCCO) [23] with 27,209 participants (11,348 cases and 15,861 controls). ILCCO also provided information of histological subtypes including squamous cell cancer and adenocarcinoma. For each of the SNP associated with sleep traits, we retrieved its effect on lung cancer from ILCCO and proxy SNP (LD $r^2 > 0.8$) from the 1000 Genomes Project, which were absent in outcome dataset.

2.3. Statistical Analysis

2.3.1. Univariate Two-Sample MR Analysis. The associations between exposure (sleeping traits) and outcome (lung cancer) were calculated with two-sample MR analysis [24]. We used inverse variance weighted (IVW) to clarify the causal associations. We also performed the same procedure for its subtypes (squamous cell cancer and adenocarcinoma). The results were shown as odds ratios (OR) and 95% confidence intervals (CI). To account for sensitivity of results, we used MR Egger regression, weighted median [25], and weighted mode to evaluate causal association. Moreover, we performed heterogeneity test which can suggest reliability of MR estimates. We also used Egger regression intercept to estimate the magnitude of horizontal pleiotropy, which can further illustrate whether SNPs influence the lung cancer risks through the sleep traits.

To further detect causal estimates for potential violation of the MR assumptions, we also performed RadialMR [26] to ascertain outliers in MR analysis and conducted reanalysis after excluding these outliers. RadialMR analysis was conducted using modified second-order weights and an α level of 0.05.

2.3.2. Multivariate Two-Sample MR Analysis. Considering that smoking is recognized as the common risk factor for lung cancer, we conducted IVW multivariable MR to estimate the effect of each sleeping traits after adjusting for pack years of smoking or current tobacco smoking status. To further eliminate the interaction effect between different exposures and avoid the multicollinearity, we also performed IVW multivariable MR after applying LASSO feature selection to identify effects of sleep duration, nap during the day, and sleeplessness for lung cancer. All analyses were replicated on squamous cell cancer and adenocarcinoma.

MR analyses were performed using the R package "TwoSampleMR" (version 0.5.5) in R (version 4.0.3).

3. Results

3.1. Character of SNP for Analysis. Table 1 shows the source of GWAS data. Each SNP extracted from different sleep traits and its F statistic and R^2 are shown in Supplementary Table 3. There were 156 SNPs for chronotype, 75 for getting up in the morning, 91 for nap during the day, 70 for sleep duration, and 42 for sleeplessness. After harmonization of the SNP effects, the SNPs available in univariate two-sample MR analysis are presented in Supplementary Table 4. Finally, 147 SNPs were used to instrument chronotype, 72 for getting up in the morning, 87 for nap during the day, 65 for sleep duration, and 42 for sleeplessness. F statistics range from 45 to 59, representing strong instruments in the MR analysis.

3.2. Causal Effect from Sleeping Traits to Lung Cancer

3.2.1. Lung Cancer. We found adverse effects of sleeplessness (OR 2.53, 95% CI 1.25-5.12) and protective effects of sleep duration (0.46, 0.26-0.83) on lung cancer risk. However, the effects of chronotype, getting up in the morning, and nap during the day were not statistically significant (0.98, 0.70-1.16 for chronotype; 0.99, 0.62-1.60 for getting up in the morning; 1.37, 0.77-2.24 for sleep duration).

3.2.2. Squamous Cell Lung Cancer. All MR results were not statistically significant (0.87, 0.99-3.79 for chronotype; 1.08,

Table 1: Characteristics of sleep traits in UK Biobank and lung cancer consortium.

(a)

Exposure	Consortium	Sample size	Population
Chronotype	MRC-IEU	413343	European
Getting up in the morning	MRC-IEU	461658	European
Sleep duration	MRC-IEU	460099	European
Nap during the day	MRC-IEU	462400	European
Sleeplessness/insomnia	MRC-IEU	462341	European
Current tobacco smoking	MRC-IEU	462434	European
Pack years of smoking	MRC-IEU	142387	European

(b)

Outcomes	Consortium	Cases/control	Sample size	Population
Lung cancer	ILCCO	11348/15861	27209	European
Squamous cell cancer	ILCCO	3275/15038	18313	European
Adenocarcinoma	ILCCO	3442/14894	18336	European

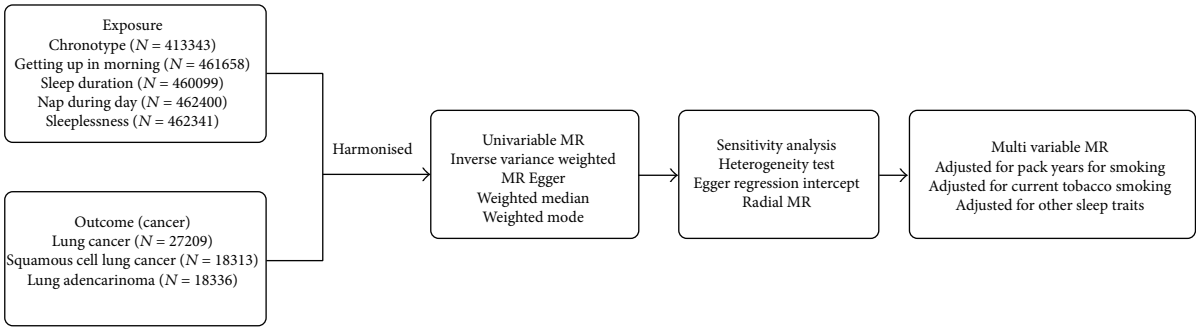


Figure 1: Flow diagram of Mendelian randomization.

0.55-2.12 for getting up in the morning; 0.46, 0.18-1.18 for sleep duration; 1.07, 0.48-2.35 for nap during the day; and 2.46, 0.83-7.34 for sleeplessness).

3.2.3. Lung Adenocarcinoma. We observed a strongly hazardous effect of sleeplessness (5.75, 2.12-15.65) on the risk of lung adenocarcinoma, while little evidence of causal effects of other sleeping traits was obtained (0.85, 0.61-1.2 for chronotype; 2.21, 0.81-5.99 for getting up in the morning; 0.62, 0.29-1.31 for sleep duration; and 2.04, 0.66-6.35 for nap during the day).

In multivariate MR analysis, sleeplessness still showed an adverse effect on lung adenocarcinoma adjusted for pack years of smoking (4.55, 1.23-16.87) or current tobacco smoking (4.99, 1.79-13.90), while sleep duration showed a protective influence on lung cancer adjusted for these two smoking statuses (0.47, 0.25-0.90, and 0.53, 0.31-0.90, respectively). Figure 1 showed the study design. All MR results are shown in Table 2 and Figure 2.

Through the LASSO feature selection function, only relevant features and instruments were retained. The results of MVMR performed on remaining SNP data were also similar with univariate analysis (in Supplementary Table 6).

3.3. Sensitivity Analyses. Other results estimated by MR Egger, weighted median, and weighted mode are available in Supplementary Table 1. There was no evidence supporting the presence of horizontal pleiotropy in the MR Egger regression analysis (Supplementary Table 2). Heterogeneity was observed in the chronotype and nap during the day analysis. Sleep duration showed heterogeneity only in lung cancer analysis. We did not observe heterogeneity in other MR results. A detailed heterogeneity test and pleiotropy are available in Supplementary Table 2. After excluding outliers of these results with heterogeneity, MR results were consistent with the results before excluding (in Supplementary Table 5).

4. Discussion

In this study, we explored the causal effects of five sleep traits including chronotype, getting up in the morning, sleep duration, nap during the day, and sleeplessness on lung cancer, squamous cell lung cancer, and lung adenocarcinoma. Insomnia was causally associated with a higher risk of lung adenocarcinoma, while sleep duration showed a protective effect on lung cancer risk.

Table 2: Two-sample Mendelian randomization estimations showing the effect of sleep traits on cancer using the IVW method.

Exposure	Method	OR (95% CI)	*P* value
	Inverse variance weighted	0.90 (0.70-1.16)	0.42
Chronotype	MVMR adjusted for pack years for smoking	0.90 (0.68-1.19)	0.46
	MVMR adjusted for current tobacco smoking	0.92 (0.71-1.18)	0.50
	Inverse variance weighted	0.99 (0.62-1.60)	0.98
Getting up in the morning	MVMR adjusted for pack years for smoking	1.18 (0.66-2.12)	0.57
	MVMR adjusted for current tobacco smoking	1.30 (0.80-2.10)	0.29
	Inverse variance weighted	0.46 (0.26-0.83)	0.01
Sleep duration	MVMR adjusted for pack years for smoking	0.47 (0.25-0.90)	0.02
	MVMR adjusted for current tobacco smoking	0.53 (0.31-0.90)	0.02
	Inverse variance weighted	1.37 (0.77-2.44)	0.29
Nap during the day	MVMR adjusted for pack years for smoking	1.04 (0.53-2.01)	0.92
	MVMR adjusted for current tobacco smoking	1.17 (0.66-2.07)	0.60
	Inverse variance weighted	2.53 (1.25-5.12)	0.01
Sleeplessness	MVMR adjusted for pack years for smoking	1.64 (0.59-4.59)	0.34
	MVMR adjusted for current tobacco smoking	1.94 (0.99-3.79)	0.05
	Inverse variance weighted	0.87 (0.59-1.27)	0.46
Chronotype	MVMR adjusted for pack years for smoking	0.83 (0.57-1.21)	0.33
	MVMR adjusted for current tobacco smoking	0.86 (0.59-1.24)	0.42
	Inverse variance weighted	1.08 (0.55-2.12)	0.82
Getting up in the morning	MVMR adjusted for pack years for smoking	1.20 (0.59-2.43)	0.61
	MVMR adjusted for current tobacco smoking	1.20 (0.60-2.38)	0.61
	Inverse variance weighted	0.46 (0.18-1.18)	0.11
Sleep duration	MVMR adjusted for pack years for smoking	0.50 (0.21-1.19)	0.12
	MVMR adjusted for current tobacco smoking	0.57 (0.25-1.30)	0.18
	Inverse variance weighted	1.07 (0.48-2.35)	0.87
Nap during the day	MVMR adjusted for pack years for smoking	0.81 (0.36-1.80)	0.60
	MVMR adjusted for current tobacco smoking	0.86 (0.39-1.88)	0.70
	Inverse variance weighted	2.46 (0.83-7.34)	0.11
Sleeplessness	MVMR adjusted for pack years for smoking	1.36 (0.43-4.24)	0.60
	MVMR adjusted for current tobacco smoking	1.76 (0.70-4.41)	0.23
	Inverse variance weighted	0.85 (0.61-1.20)	0.37
Chronotype	MVMR adjusted for pack years for smoking	0.83 (0.57-1.20)	0.32
	MVMR adjusted for current tobacco smoking	0.86 (0.60-1.22)	0.39
	Inverse variance weighted	1.60 (0.73-3.49)	0.24
Getting up in the morning	MVMR adjusted for pack years for smoking	2.01 (0.89-4.55)	0.10
	MVMR adjusted for current tobacco smoking	2.03 (0.95-4.33)	0.07
	Inverse variance weighted	0.62 (0.29-1.31)	0.21
Sleep duration	MVMR adjusted for pack years for smoking	0.63 (0.28-1.43)	0.27
	MVMR adjusted for current tobacco smoking	0.70 (0.34-1.43)	0.32
	Inverse variance weighted	1.67 (0.72-3.87)	0.23
Nap during the day	MVMR adjusted for pack years for smoking	1.20 (0.48-2.99)	0.70
	MVMR adjusted for current tobacco smoking	1.56 (0.67-3.65)	0.30
	Inverse variance weighted	5.75 (2.12-15.65)	<0.01
Sleeplessness	MVMR adjusted for pack years for smoking	4.55 (1.23-16.87)	0.02
	MVMR adjusted for current tobacco smoking	4.99 (1.79-13.90)	<0.01

OR: odds ratios; 95% CI: 95% confidence interval; IVW: inverse variants weighted; MVMR: multivariable variant Mendelian randomization.

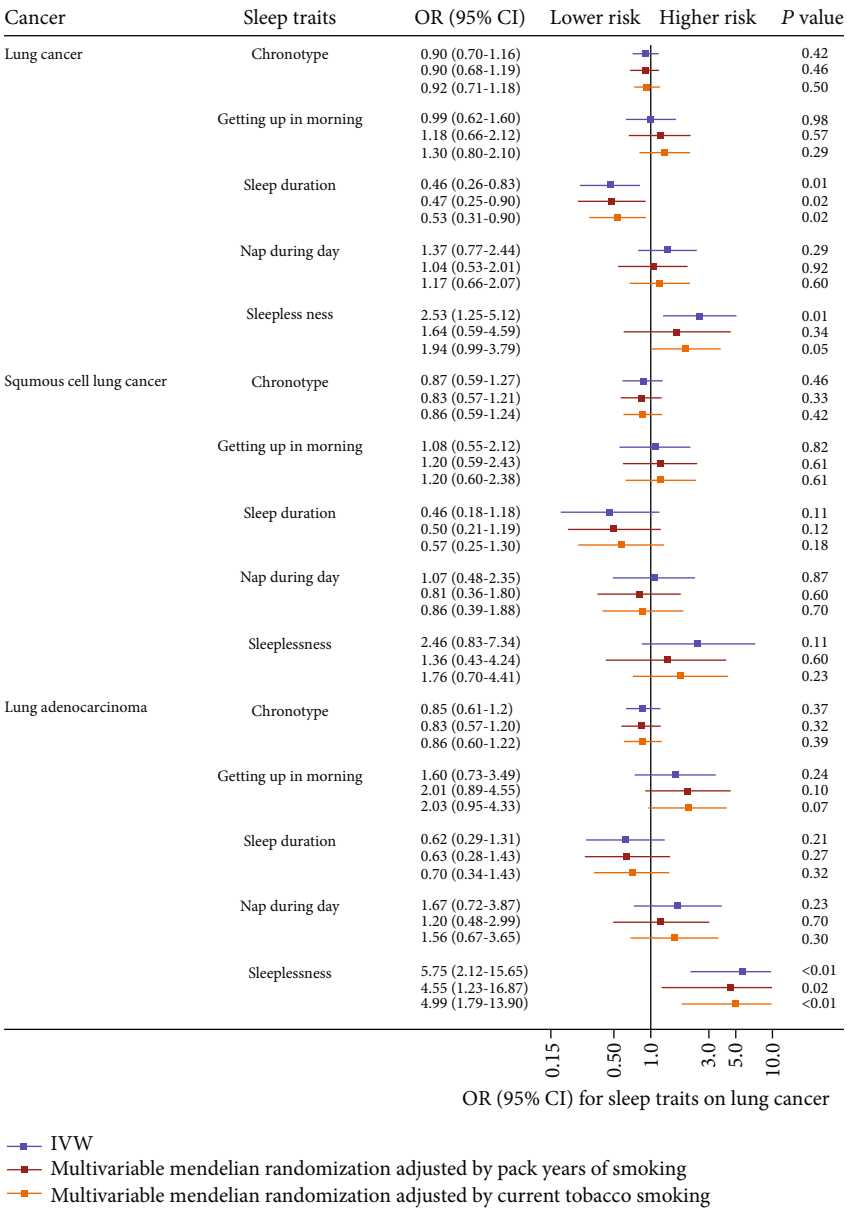


Figure 2: Forest plot of Mendelian randomization (MR) estimates for association between sleep traits and cancer risk. OR: odds ratios; 95% CI: 95% confidence interval; IVW: inverse variants weighted; MVMR: multivariable variant Mendelian randomization.

Previous epidemiological studies have just focused on the relationship between sleep duration and lung cancer. Some studies have reported the U-shaped association [13, 14], indicating that longer sleep duration and short sleep duration are both associated with unhealthy outcomes. Furthermore, a meta-analysis including 32 studies also suggested that long sleep duration increases cancer-specific mortality, especially for lung cancer [12]. However, a US male physician cohort study with a mean follow-up of 7.5 years had a different conclusion that altered sleep duration (≤6 h/day or ≥8 h/day) failed to increase lung cancer incidence. Another prospective cohort study including 21,804 participants in Canada also identified no significant effects of unfavorable sleep duration (<7 h/day or >9 h/day) while night shift work may contribute to lung cancer incidence. Unlike observational studies, our study showed that sleep duration was a protective factor for lung cancer, suggesting that longer sleep duration could decrease the risk for lung cancer.

In addition to sleep duration, other sleep traits also reflect sleep conditions; a comprehensive evaluation should contain the impacts of chronotype, getting up in the morning, and sleeplessness on lung cancer. Only Xie and his colleagues [13] explored associations of other sleep traits and lung cancer, indicating that evening chronotype also increases lung cancer risk except for unfavorable sleep duration while sleeplessness has no effects. Chronotype and getting up in the morning, related to circadian rhythms closely, were reported as risk factors for cancer such as breast cancer [8] and epithelial ovarian cancer [27]. However, compared with Xie et al.'s study, our study showed opposed findings that chronotype

did not contribute to lung cancer incidence and sleep duration showed a protective effect. Given the heterogeneity of different subtypes, we replicated all analyses on other subtypes such as squamous cell lung cancer and lung adenocarcinoma. Although sleeplessness may not be harmful to lung cancer, there surprisingly appeared a strong association with lung adenocarcinoma. For the cancer patient, sleeplessness is often a common and enduring symptom [28, 29], especially for patients in the terminal stage of lung cancer [30].

The mechanisms underlying these associations are poorly understood. One possible pathway is that sleep disturbances may lead to chronic lung disease through circadian rhythm disruption [31]. Sleep deprivation leads to a more severe lung inflammation [32], which is essential for the risk of lung cancer [33]. These findings may support the adverse effect of short sleep duration sleeplessness and are consistent with our results partially. However, there is lack of evidence on the histology-specific impact of sleeplessness.

To our knowledge, this study is the first to explore connections between sleep traits and lung cancer risks at the level of genes. Although random control trial (RCT) can provide the most compelling evidence, it involves many ethical issues and costs much money. For observation studies, despite these results from observed studies that were adjusted by other relative variables, undetected biases could not be ignored. Therefore, the results provided by MR are the most convincing. Bias due to confounding and reverse sources could be decreased by MR. To minimize the potential violation of the MR assumption, we also conducted serials of sensitivity analysis and detected any outliers by RadialMR analysis. We also conducted multivariable MR to adjust for smoking, the most common and important risk factor of lung cancer.

Several limitations should be considered in our study. First, our study was based on the European population. Thus, whether our study could be generalizable to other populations requires further investigations. Second, the summary data used in our MR analyses were not stratified by gender or smoking. Finally, all sleep traits were self-reported. Thus, it is possible to lead to misclassification of exposure.

In conclusion, MR analysis provides stronger evidence for the causal effect of sleeplessness on lung adenocarcinoma and highlights the importance of sleep duration in lung cancer incidence. Although other sleep traits did not show protective or adverse effects on lung cancer, these findings imply that we still need to pay attention to sleep health to mitigate the risk of incident lung cancer. Our results may further emphasize the importance of enough sleep for health. Further studies are needed to illustrate the association between sleep traits and lung cancer in females and nonsmokers.

Data Availability

Our data was from the UK Biobank and the International Lung Cancer Consortium, the two open-access datasets (https://www.mrbase.org/;https://www.ukbiobank.ac.uk/ ;https://ilcco.iarc.fr/).

Conflicts of Interest

The authors declare no financial or commercial conflict of interest.

Authors' Contributions

Jie Wang and Haibo Tang contributed equally to this work.

Acknowledgments

This work was funded by the following grants and associations: National Natural Science Foundation of China (81974465 and 81900199), Hunan province natural science funds for Excellent Young Scholars (2019JJ30043), and the recruitment program for Huxiang talents (2019RS1009).

Supplementary Materials

Supplementary Table 1: two-sample Mendelian randomization estimations showing the effect of sleep traits on cancer using the MR Egger, weighted median, and weighted mode method. Supplementary Table 2: sensitivity analysis performed by Egger regression intercept and heterogeneity test. Supplementary Table 3: SNPs of sleep traits extracted from UK Biobank with statistically significant threshold [$P < 5 \times 10^{-8}$; linkage disequilibrium (LD) $r^2 < 0.001$, LD distance > 10000 kb]. Supplementary Table 4: SNPs used in two-sample Mendelian randomization analysis. Supplementary Table 5: outliers selected by RadialMR and the reanalysis results after excluding outliers. Supplementary Table 6: multivariable two-sample Mendelian randomization estimation showing the effects of different sleep traits on lung cancer. *(Supplementary Materials)*

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] S. C. Larsson, P. Carter, S. Kar et al., "Smoking, alcohol consumption, and cancer: a Mendelian randomisation study in UK Biobank and international genetic consortia participants," *PLoS Medicine*, vol. 17, no. 7, article e1003178, 2020.

[3] J. V. Aredo, S. J. Luo, R. M. Gardner et al., "Tobacco smoking and risk of second primary lung cancer," *Journal of Thoracic Oncology*, vol. 16, no. 6, pp. 968–979, 2021.

[4] R. Peto, S. Darby, H. Deo, P. Silcocks, E. Whitley, and R. Doll, "Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies," *BMJ*, vol. 321, no. 7257, pp. 323–329, 2000.

[5] D. A. Siegel, S. A. Fedewa, S. J. Henley, L. A. Pollack, and A. Jemal, "Proportion of never smokers among men and women with lung cancer in 7 US states," *JAMA Oncology*, vol. 7, no. 2, pp. 302–304, 2021.

[6] A. G. Pallis and K. N. Syrigos, "Lung cancer in never smokers: disease characteristics and risk factors," *Critical Reviews in Oncology/Hematology*, vol. 88, no. 3, pp. 494–503, 2013.

[7] J. M. Samet, E. Avila-Tang, P. Boffetta et al., "Lung cancer in never smokers: clinical epidemiology and environmental risk

factors," *Clinical Cancer Research*, vol. 15, no. 18, pp. 5626–5645, 2009.

[8] R. C. Richmond, E. L. Anderson, H. S. Dashti et al., "Investigating causal relations between sleep traits and risk of breast cancer in women: Mendelian randomisation study," *BMJ*, vol. 365, p. l2327, 2019.

[9] L. Jiao, Z. Duan, H. Sangi-Haghpeykar, L. Hale, D. L. White, and H. B. el-Serag, "Sleep duration and incidence of colorectal cancer in postmenopausal women," *British Journal of Cancer*, vol. 108, no. 1, pp. 213–221, 2013.

[10] C. L. Thompson, E. K. Larkin, S. Patel, N. A. Berger, S. Redline, and L. Li, "Short duration of sleep increases risk of colorectal adenoma," *Cancer*, vol. 117, no. 4, pp. 841–847, 2011.

[11] J. McNeil, A. M. Barberio, C. M. Friedenreich, and D. R. Brenner, "Sleep and cancer incidence in Alberta's Tomorrow Project cohort," *Sleep*, vol. 42, no. 3, 2019.

[12] C. R. Stone, T. R. Haig, K. M. Fiest, J. McNeil, D. R. Brenner, and C. M. Friedenreich, "The association between sleep duration and cancer-specific mortality: a systematic review and meta-analysis," *Cancer Causes & Control*, vol. 30, no. 5, pp. 501–525, 2019.

[13] J. Xie, M. Zhu, M. Ji et al., "Relationships between sleep traits and lung cancer risk: a prospective cohort study in UK Biobank," *Sleep*, 2021.

[14] M. K. Luojus, S. M. Lehto, T. Tolmunen, A. T. Erkkilä, and J. Kauhanen, "Sleep duration and incidence of lung cancer in ageing men," *BMC Public Health*, vol. 14, no. 1, p. 295, 2014.

[15] J. McNeil, E. Heer, R. F. Willemsen, C. M. Friedenreich, and D. R. Brenner, "The effects of shift work and sleep duration on cancer incidence in Alberta`s Tomorrow Project cohort," *Cancer Epidemiology*, vol. 67, article 101729, 2020.

[16] O. Khawaja, A. B. Petrone, S. Aleem, K. Manzoor, J. M. Gaziano, and L. Djousse, "Sleep duration and risk of lung cancer in the physicians' health study," *Zhongguo Fei Ai Za Zhi*, vol. 17, no. 9, pp. 649–655, 2014.

[17] G. Davey Smith and S. Ebrahim, "'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?," *International Journal of Epidemiology*, vol. 32, no. 1, pp. 1–22, 2003.

[18] G. Davey Smith and G. Hemani, "Mendelian randomization: genetic anchors for causal inference in epidemiological studies," *Human Molecular Genetics*, vol. 23, no. R1, pp. R89–R98, 2014.

[19] M. M. Glymour, E. J. Tchetgen Tchetgen, and J. M. Robins, "Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions," *American Journal of Epidemiology*, vol. 175, no. 4, pp. 332–339, 2012.

[20] G. Yang, Y. Y. Han, T. Sun et al., "Sleep duration, current asthma, and lung function in a nationwide study of U.S. adults," *American Journal of Respiratory and Critical Care Medicine*, vol. 200, no. 7, pp. 926–929, 2019.

[21] C. Bycroft, C. Freeman, D. Petkova et al., "The UK Biobank resource with deep phenotyping and genomic data," *Nature*, vol. 562, no. 7726, pp. 203–209, 2018.

[22] S. Burgess, S. G. Thompson, and CRP CHD Genetics Collaboration, "Avoiding bias from weak instruments in Mendelian randomization studies," *International Journal of Epidemiology*, vol. 40, no. 3, pp. 755–764, 2011.

[23] Y. Wang, J. D. McKay, T. Rafnar et al., "Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer," *Nature Genetics*, vol. 46, no. 7, pp. 736–741, 2014.

[24] EPIC- InterAct Consortium, S. Burgess, R. A. Scott, N. J. Timpson, G. Davey Smith, and S. G. Thompson, "Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors," *European Journal of Epidemiology*, vol. 30, no. 7, pp. 543–552, 2015.

[25] J. Bowden, G. Davey Smith, P. C. Haycock, and S. Burgess, "Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator," *Genetic Epidemiology*, vol. 40, no. 4, pp. 304–314, 2016.

[26] J. Bowden, W. Spiller, F. del Greco M et al., "Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the radial plot and radial regression," *International Journal of Epidemiology*, vol. 47, no. 4, pp. 1264–1278, 2018.

[27] L. Leung, A. Grundy, J. Siemiatycki et al., "Shift work patterns, chronotype, and epithelial ovarian cancer risk," *Cancer Epidemiology, Biomarkers & Prevention*, vol. 28, no. 5, pp. 987–995, 2019.

[28] J. Savard and C. M. Morin, "Insomnia in the context of cancer: a review of a neglected problem," *Journal of Clinical Oncology*, vol. 19, no. 3, pp. 895–908, 2001.

[29] J. Savard, H. Ivers, J. Villa, A. Caplette-Gingras, and C. M. Morin, "Natural course of insomnia comorbid with cancer: an 18-month longitudinal study," *Journal of Clinical Oncology*, vol. 29, no. 26, pp. 3580–3586, 2011.

[30] K. Skaug, G. E. Eide, and A. Gulsvik, "Prevalence and predictors of symptoms in the terminal stage of lung cancer: a community study," *Chest*, vol. 131, no. 2, pp. 389–394, 2007.

[31] H. Hadden, S. J. Soldin, and D. Massaro, "Circadian disruption alters mouse lung clock gene expression and lung mechanics," *Journal of Applied Physiology*, vol. 113, no. 3, pp. 385–392, 2012.

[32] J. O. F. Nunes, J. . S. Apostolico, D. A. G. Andrade et al., "Sleep deprivation predisposes allergic mice to neutrophilic lung inflammation," *Journal of Allergy and Clinical Immunology*, vol. 141, no. 3, pp. 1018–1027.e4, 2018.

[33] E. A. Engels, "Inflammation in the development of lung cancer: epidemiological evidence," *Expert Review of Anticancer Therapy*, vol. 8, no. 4, pp. 605–615, 2008.

Hindawi
Journal of Immunology Research
Volume 2021, Article ID 9995869, 13 pages
https://doi.org/10.1155/2021/9995869

Research Article

Suppression of DLBCL Progression by the E3 Ligase Trim35 Is Mediated by CLOCK Degradation and NK Cell Infiltration

Xiyan Tan,[1,2] Fuyang Cao,[1,2] Feiyu Tang,[1,2] Can Lu,[3] Qiaoyan Yu,[1,2] Songshan Feng,[4] Zhanghuan Yang,[1,2] Songming Chen,[1,2] Xiang He,[1,2] Jiang He,[1,2] Liang Weng,[1,2,5] and Lunquan Sun[1,2,5,6,7]

[1]*Xiangya Cancer Center, Xiangya Hospital, Central South University, Changsha 410008, China*
[2]*Key Laboratory of Molecular Radiation Oncology Hunan Province, Changsha 410008, China*
[3]*Department of Pathology, Xiangya Hospital, Central South University, Changsha 410078, China*
[4]*Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha 410078, China*
[5]*Hunan International Science and Technology Collaboration Base of Precision Medicine for Cancer, Changsha 410008, China*
[6]*Institute of Gerontological Cancer Research, National Clinical Research Center for Gerontology, Changsha 410008, China*
[7]*Center for Molecular Imaging of Central South University, Xiangya Hospital, Changsha 410008, China*

Correspondence should be addressed to Liang Weng; wengliang@csu.edu.cn and Lunquan Sun; lunquansun@csu.edu.cn

Received 25 March 2021; Accepted 10 May 2021; Published 24 May 2021

Academic Editor: Jialiang Liang

Copyright © 2021 Xiyan Tan 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.

The majority of diffuse large B-cell lymphoma (DLBCL) patients develop relapsed or refractory disease after standard ruxolitinib, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) chemotherapy, which is partly related to a dysregulated tumor immune microenvironment. However, how the infiltration of immune cells is appropriately regulated is poorly understood. Herein, we show that the E3 ubiquitin ligase Trim35 is expressed at low levels in human DLBCL tissues. We also show that overexpression of Trim35 suppresses DLBCL cell proliferation and correlates with inferior survival in DLBCL patients. Our mechanistic study shows that Trim35 functions as an E3 ligase to mediate the ubiquitination and degradation of CLOCK, a key regulator of circadian rhythmicity. High expression of Trim35 correlates with NK cell infiltration in DLBCL, partly due to the degradation of CLOCK. Consistently, patients with high expression of CLOCK show poor overall survival. Overall, these findings suggest that Trim35 suppresses the progression of DLBCL by modulating the tumor immune microenvironment, indicating that it may be a promising diagnostic and prognostic biomarker in DLBCL.

1. Introduction

Diffuse large B-cell lymphoma (DLBCL), the most common subtype of non-Hodgkin lymphoma (NHL) in adults, has one of the highest mortality rates among cancers in the most developed areas of the world. Three gene expression subgroups, germinal center B cell-like (GCB-like), activated B cell-like (ABC-like), and primary mediastinal B cell lymphoma, have been identified, and the last two groups are also called non-GCB-like lymphoma [1, 2]. Patients with GCB-like DLBCL have significantly better overall survival than those with non-GCB-like DLBCL [3, 4]. Likewise, a clinical indicator of prognosis, the International Prognostic Indicator (IPI), has been successfully used to define prognostic subgroups of DLBCL. This indicator takes into account the patients' age and performance status, the extent and location of disease, and the serum LDH level [2]. Although current therapeutic strategies, including standard ruxolitinib, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) chemotherapy; unlabeled or radiolabeled monoclonal antibodies; and high-dose chemotherapy following autologous peripheral blood stem cell transplantation, have significantly improved the outcome of DLBCL, the majority of patients relapse or become resistant to prior therapies [5], which is partly due to the dysregulated tumor immune microenvironment. For example, the ABC-like DLBCL

subtype was found to have inactivation of the CD58 gene, which indicates the loss of recognition of tumor cells by CTLs and NK cells and may be a cause of relapse or refractory disease in a patient with the ABC-like DLBCL subtype. Therefore, it is necessary to further investigate the underlying mechanism regulating the infiltration of immune cells into the tumor microenvironment.

Circadian rhythmicity is an approximately 24-hour cell-autonomous period driven by transcription-translation feedback loops of "circadian clock genes," which are expressed in most cell types, including cells of the immune system [6–9]. The core circadian clock is an autoregulatory transcriptional feedback loop involving the activators CLOCK and BMAL1 and the repressor complex, including Per1, Per2, Cry1, and Cry2 [9–12]. The circadian system regulates virtually all physiological processes and influences the states of health and disease in humans. Recent advances have shown that circadian rhythmicity disruption contributes to disturbed immune responses, obesity, type 2 diabetes mellitus, and cancer. To date, how the expression of circadian clock genes is regulated in tumors is still poorly understood.

Ubiquitin is a small modifier molecule that labels proteins in a highly specific manner. Initially, ubiquitination was described as the process by which proteins are labeled for degradation by the proteasome [13], which is responsible for degrading 80-90% of intracellular proteins that are aberrantly folded or typically short-lived. Protein ubiquitination also regulates protein trafficking and protein-protein interactions via different linkages of polyubiquitin chains [14, 15]. The Ub-activation enzyme (E1), Ub-conjugation enzymes (E2), and Ub ligases (E3) are required to attach ubiquitin to a substrate [16, 17]. E3 ligases are considered to be the most important components of the ubiquitin conjugation machinery, as they bind directly to their target proteins and have substrate specificity. Protein ubiquitination affects many cellular processes, from gene transcription and DNA repair to the cell cycle and apoptosis [18]. Polyubiquitination often tags proteins for proteasomal degradation [19]. Recent advances indicate that the linear ubiquitin chain assembly complex (LUBAC) is crucially involved in B-cell lymphomagenesis through protection against DNA damage-induced cell death and is a suitable therapeutic target for B-cell lymphomas [20]. Moreover, recent research indicates that the CRL3-SPOP ubiquitin ligase complex suppresses the growth of diffuse large B-cells by negatively regulating MYD88/NF-κB signaling [21]. Accumulating evidence has shown that ubiquitination is critical for DLBCL development. It has also been reported that the stability of the clock gene PER and CRY proteins are regulated by SCF E3 ubiquitin ligase complexes involving β-TrCP and FBXL3, respectively [22]. However, it remains largely unknown how the ubiquitination of circadian clock proteins is involved in DLBCL progression.

Tripartite motif 35 (Trim35), also called hemopoietic lineage switch (Hls5), is a member of the RING finger, B box, coiled coil (RBCC), or TRIM family E3 ligases expressed in a wide variety of hemopoietic cell types, including fetal liver progenitors [23]. Trim35 is reported to be a tumor suppressor gene [24]. Enforced expression of Trim35 in HeLa cells inhibits cell growth, clonogenicity, and tumorigenicity. Trim35 downregulation is a frequent event in hepatocellular carcinoma, and the expression level of Trim35 is negatively correlated with tumor size, histological grade, and serum alpha-fetoprotein concentration [24]. Moreover, TRIM family proteins are important effectors of innate immunity against viral infections. Recent findings revealed novel roles of Trim35; it catalyzes the Lys63- or Lys48-linked polyubiquitination of TRAF3, regulates RIG-I antiviral immunity, and is involved in the mechanism of defense against influenza A virus (IAV) infection [25]. However, the roles of Trim35 in the immune microenvironment and DLBCL have not been previously reported.

In this study, we identified that Trim35 is expressed at low levels in DLBCL tissues. Trim35 overexpression significantly inhibited DLBCL proliferation and is correlated with a better prognosis than Trim35 deficiency in DLBCL patients. In addition, we found that Trim35 interacts with CLOCK and promotes its ubiquitination and degradation. Furthermore, Trim35 and CLOCK promote and inhibit NK cell infiltration in DLBCL. Our study revealed how the ubiquitination of circadian clock proteins contributes to tumor immune microenvironment remodeling and DLBCL progression.

2. Methods

2.1. Patients. In total, 69 patients who were diagnosed with DLBCL at Xiangya Hospital of Central South University between 2013 and 2019 were included in this study. Clinical data were obtained from their medical records. The follow-up periods ranged from 4 to 88 months.

As an independent validation cohort, a total of 61 patients were diagnosed at DLBCL between 2013 and 2019 and, homogeneously treated with R-CHOP, were collected. The age of patients ranged from 10 to 86 years, and the follow-up duration was from 5 to 88 months.

2.2. Immunohistochemistry. Whole sections of representative formalin-fixed, paraffin-embedded (FFPE) tumor tissue blocks were submitted for immunohistochemistry (IHC). The immunohistochemical subgroup of DLBCL was determined to be non-GCB or GCB type according to Hans' criteria [26]. IHC was performed following standard protocols. In detail, immunohistochemical cytokeratin staining was performed on FFPE tissue using an in direct immunoperoxidase technique. Sections mounted on a slide were dewaxed in xylene, dehydrated in ethanol, boiled in 0.01 M citrate buffer (pH 6.0) for 30 minutes in a microwave oven, and then incubated with 3% hydrogen peroxide for 15 minutes. After washing with PBS, the slides were incubated in 5% normal BSA for 1 hour, followed by incubation overnight with rabbit polyclonal antibodies recognizing Trim35 (HPA019647, 1 : 200, Sigma-Aldrich, Shanghai, China), CLOCK (18094-1-AP,1 : 50, Protein-Tech, Wuhan, China), CD3 (Kit-0003, MaiXin Biotechnologies, Fuzhou, China), CD19 (MAB-0705, MaiXin Biotechnologies, Fuzhou, China), CD56 (MAB-0743, MaiXin Biotechnologies, Fuzhou, China), CD16 (16559-AP,1 : 100, Protein-Tech, Wuhan, China), and CD68 (Kit-0026, MaiXin Biotechnologies, Fuzhou, China).

After washing, the sections were incubated with 3,3′-diaminobenzidine (DAB) (PV-6000D, ZSGB-BIO, Beijing, China). The sections were then counterstained with hematoxylin, dehydrated, cleared, and mounted. CD3 was used as a marker of T cells, CD19 was used as a marker of B-cells, CD56 and CD16 were used as markers of NK cells, and CD68 was used as a marker of tumor-associated macrophages (TAMs). Immunostaining for CD3, CD16, CD19, CD56, CD68, Trim35, and CLOCK was performed.

2.3. Cell Culture and Transfection. Pfeiffer cells (a human DLBCL cell line) were purchased from the American Type Culture Collection (ATCC) (Manassas, VA, USA). This cell line was cultured in RPMI 1640 medium (Gibco-Invitrogen), which was supplemented with penicillin (100 U/ml), streptomycin (100 μg ml^{-1}; Gibco-Invitrogen), and 10% fetal bovine serum (FBS); the cells were cultures in an incubator with 5% CO_2 at 37°C. Human embryonic kidney 293FT cells were cultured in DMEM (Gibco-Invitrogen). All transfection experiments were performed using Lipofectamine 8000 (Beyotime Biotechnology, Shanghai, China) according to the manufacturer's recommendations. In detail, the cells were seeded in culture plates. When the cell density reached 70-80%, the cells were transfected with the plasmid. The cells were collected at 36-48 hours of posttransfection for western blot analyses.

For the generation of Pfeiffer cells stably expressing Flag tagged Trim35 or Trim35 KO cell lines, a lentiviral vector encoding Flag-Trim35 was transfected in a lentivirus packaging cell line, and the produced recombinant lentiviruses were then infected with the recombinant lentiviruses followed by a selection of infected cells using puromycin (A1113803, Thermo Fisher Scientific, Shanghai, China).

2.4. Plasmids and Small Guide RNAs (sgRNAs). To overexpress Trim35, the coding regions were amplified from the cDNA of 293FT cells, and Trim35 was cloned into the PQCXIP-Flag vector between the SalI and BamHI sites. And the Trim35 sgRNA was constructed by annealing double complementary oligomers to encode sgRNA into the AgeI and EcoRI sites of the lentiCRISPR V2 vector, which is referred to as pSilencer throughout this report. sgRNAs were designed using the CRISPR tool (http://crispr.mit.edu) to minimize potential off-target effects. sgRNA sequences and genomic primers were shown as follows: Trim35 sgRNA (CACGTCGGGACTCCGCTCCA).

2.5. Cell Viability Assay. Cell viability was measured by using CCK-8 assays. Briefly, cell suspensions with a density of 4,000 cells per well were seeded in 96-well plates and incubated at 37°C. After overnight incubation, the cells were treated under the indicated conditions. At the end of the treatment, CCK-8 assays were used to calculate the number of viable cells by measuring the absorbance at 450 nm and normalized to the absorbance of a blank (fill with CCK-8 reagent only).

2.6. RNA Preparation and Quantitative Real-Time Polymerase Chain Reaction (PCR). Total RNA was extracted from DLBCL cell lines by using TRIzol (Invitrogen) according to the manufacturer's instructions and was inversely transcribed to cDNA using PrimeScript RT-polymerase (RR047A-6, Takara, Dalian, China). The mRNA level of CLOCK was detected by using the iTaq™ Universal SYBR® Green (Bio-Rad), and GAPDH was used as an internal control.

The following primers were used for real-time quantitative reverse transcription PCR (qRT-PCR):

(1) GAPDH primer:

(i) Forward: 5′- GGAGCGAGATCCCTCCAAAAT

(ii) Reverse: 5′- GGCTGTTGTCATACTTCTCATGG

(2) CLOCK primer:

(i) Forward: 5′- TGCGAGGAACAATAGACCCAA

(ii) Reverse: 5′- ATGGCCTATGTGTGCGTTGTA

2.7. Immunoblotting and Immunoprecipitation (IP). For IP, extraction of proteins with a modified buffer from cultured cells was followed by IP and immunoblotting with the corresponding antibodies. Rabbit polyclonal antibodies recognizing Trim35 (HPA019647, 1 : 1000, Sigma-Aldrich, Shanghai, China) were obtained from Abcam; those for CLOCK (18094-1-AP,1 : 1000, Protein-Tech, Wuhan, China), Cryptochrome 2 (13997-1-AP,1 : 1000, Protein-Tech, Wuhan, China), and ARNTL (14268-AP,1 : 1000, Protein-Tech, Wuhan, China) were obtained from Protein Tech. Mouse polyclonal antibodies recognizing PER2 (67513-1-Ig,1 : 5000, Protein-Tech, Wuhan, China), GAPDH (60004-1-Ig,1 : 3000, Protein-Tech, Wuhan, China), and Flag (20543-1-AP, 1 : 3000, Protein-Tech, Wuhan, China) were purchased from Protein Tech.

2.8. In Vitro sgRNA Testing. In vitro testing of sgRNA was performed by lentivirus infection into the human DLBCL cell line Pfeiffer. Genomic DNA was harvested by DNA Extraction Kit (B518215, Sangon Biotech, Shanghai, China) and used as a PCR template, followed by T7 Endonuclease I (T7EI) assay.

2.9. T7 Endonuclease I (T7EI) Assay. PCR amplicons of target sites were purified and used as an input to the T7 Endonuclease I (T7EI) (D0508S, Genome-Editing Mutation Detection Kit, Beyotime Biotechnology, Shanghai, China). The manufacturer's recommended protocol was followed. T7 EI-treated products were run on a 2% agarose gel (Invitrogen), stained with nuclear acid dye, and imaged on a gel imager (Bio-Rad).

2.10. TIMER Analysis. The relationship between Trim35 expression and tumor-infiltrating immune cells (TIICs) in

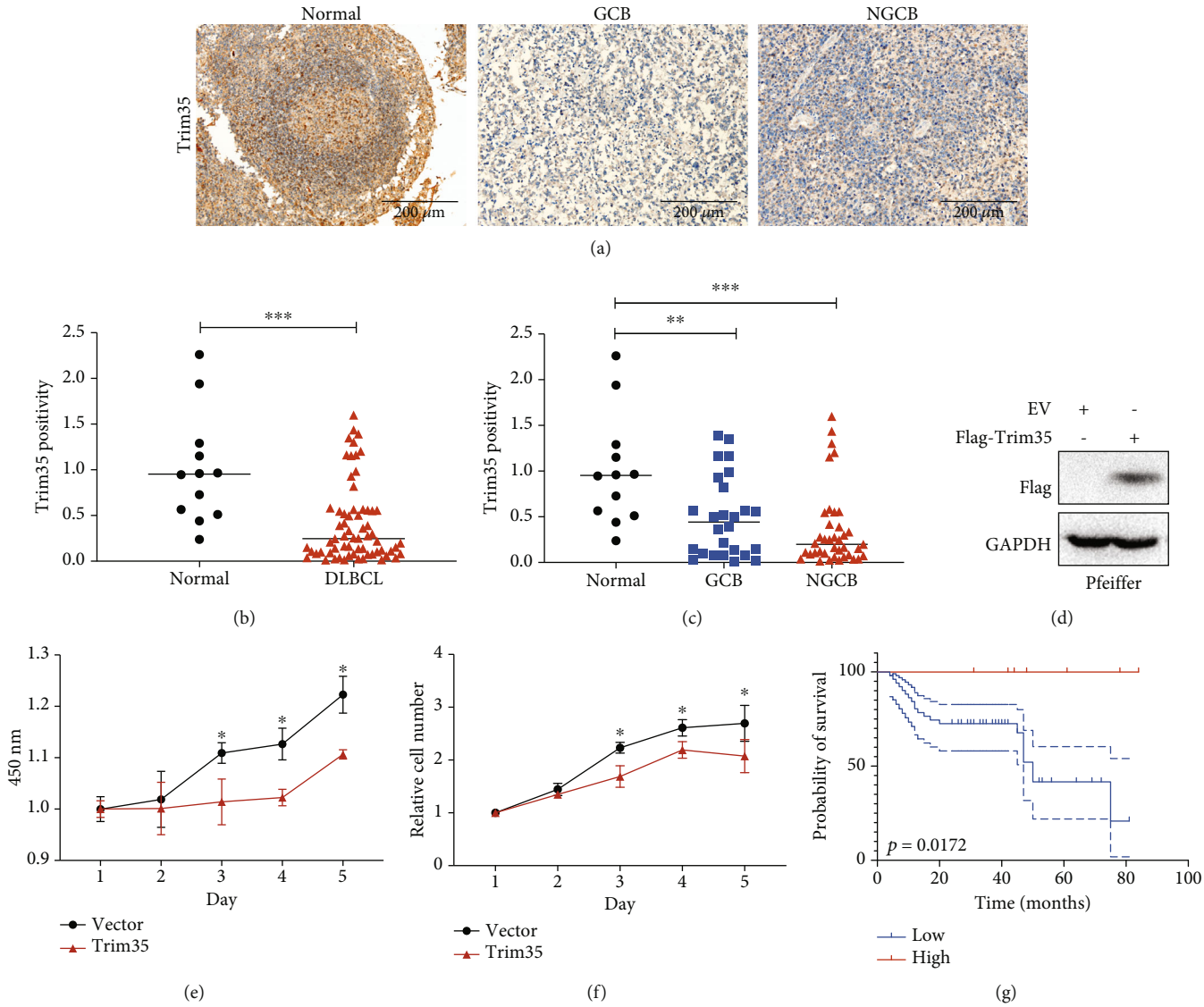


Figure 1: Low expression of Trim35 is correlated with poor prognosis in DLBCL. (a) Immunohistochemical staining for Trim35 in human DLBCL tissues and normal lymph nodes. The specific Trim35 signal is shown in brown (DAB staining), and counterstaining was performed using hematoxylin (blue nuclei). Scale bar, 200 μm. (b, c) Summary for the entire cohort of DLBCL patients, as assessed by IHC. The p value was calculated by the chi-square test. (d) Pfeiffer cells were infected with virus expressing Flag-Trim35 or empty vector. After being selected, the protein extracts were immunoblotted for the indicated proteins. (e) CCK-8 assays of stable Trim35-overexpressing cells and empty vector-expressing cells. (f) Growth curves of stable Trim35 overexpression group and empty vector group were counted. (g) Overall survival analysis of patients with high (red) versus low (blue) Trim35 protein expression based on IHC staining. Samples with Trim35 staining in <5% of cells were considered to have low expression, and those with staining >5% were considered to have high expression. All experiments were repeated at least 3 times. $^{NS}p < 0.05$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

32 cancer types was determined using the TIMER (https://cistrome.shinyapps.io/timer/) [27]. TIMER infers the abundance of TIICs applying the statistical analysis of gene expression profiles [28]. We analyzed the association between the level of Trim35 gene expression and the abundance of infiltrating immune cells, including tumor-associated macrophages (TAMs), monocytes, Tregs, myeloid dendritic cells (DCs), NK cells, B-cells, and $CD4^{+}$ T cells based on the expression of specific marker genes in different cancers including DLBCL. The marker genes used for the analysis of the tumor-infiltrating immune cells including T cells, B-cells, monocytes, TAMs, M1 macrophages, M2 macrophages, natural killer (NK) cells, dendritic cells (DCs), Tregs, and myeloid-derived suppressor cells (MDSCs) were based on data from previous studies [29, 30].

2.11. Statistical Analysis. SPSS 24.0 was used to perform all statistical analyses. Numerical data are expressed as the mean ± standard error and were calculated using a two-tailed Student's ***t***-test, chi-square test, and Spearman correlation analysis. ***p*** values are indicated by asterisks in the figures: $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. All experiments were repeated at least 3 times.

Table 1: Monofactor analysis of overall (OS) in patients with DLBCL.

Characteristics	Variables	Total ($n = 61$)	Hazard ratio (95% CI)	p value
Age (years)	<60	46	1.429 (0.5667-3.604)	0.4046
	>60	15		
Sex	Male	39	0.9404 (0.3972-2.226)	0.8881
	Female	22		
Cell of origin	GCB	24	0.5581 (0.2379-1.309)	0.1587
	Non-GCB	37		
Cancer stage at diagnosis	I/II	23	3.530 (1.521-8.193)	0.0132
	III/IV	38		
Serum LDH	Normal	37	2.116 (0.8643-5.178)	0.036
	Elevated	24		
IPI score	0/1/2	42	2.906 (1.065-7.931)	0.0078
	3/4/5	19		

3. Results

3.1. Low Expression of Trim35 Is Correlated with a Poor Prognosis in DLBCL. To assess the effect of Trim35 on DLBCL cell growth, we established Pfeiffer cells stably expressing Trim35 (Trim35-OE) by lentiviral transduction (Figure 1(d)). Then, CCK-8 assays and cell counting analysis were performed to assess the effect of Trim35 on DLBCL cell proliferation. The results showed that overexpression of Trim35 greatly inhibited the proliferation of DLBCL cells (Figures 1(e) and 1(f)), which indicated that Trim35 may be involved in the suppression of DLBCL. To confirm the relevance of Trim35 to DLBCL, we subjected clinical tissue samples from 61 patients with primary DLBCL to IHC analysis. Compared with that in adjacent normal lymph nodes, nuclear Trim35 immunoreactivity in tumor tissues was markedly lower (Figures 1(a) and 1(b)). Notably, nongerminal center B-cell type (NGCB) DLBCL tissues expressed lower Trim35 levels than GCB DLBCL tissues (Figure 1(c)). Furthermore, Kaplan-Meier survival analysis demonstrated that high expression of Trim35 was significantly associated with a longer OS in DLBCL patients ($p = 0.0172$) (Figure 1(g)). Next, in a monofactor analysis, sex and cell of origin had no prognostic significance. However, cancer stage at diagnosis, serum LDH, and IPI score were found to be independent prognostic indicators for OS (cancer stage of diagnosis, $p = 0.0132$, HR = 3.530 [95% CI, 1.521-8.193]; serum LDH, $p = 0.036$, HR = 2.116 [95% CI, 0.8643-5.178]; and IPI score, $p = 0.0078$, HR = 2.906 [95% CI, 1.065-7.931]) (Table 1). Overall, these data showed Trim35 functions as a suppressor of DLBCL.

3.2. Trim35 Is Positively Associated with NK Cell Infiltration and Inhibits DLBCL Progression. To address how Trim35 suppresses DLBCL progression, we assessed whether Trim35 upregulation leads to antitumor effects by remodeling the immune microenvironment. To verify this hypothesis, we conducted Gene-Immune Analysis using the Tumor IMmune Estimation Resource (TIMER) (https://cistrome.shinyapps.io/timer/). As expected, the results showed that the expression of Trim35 was highly correlated with the infiltration of memory B-cells, $CD4^+$ (nonregulatory) T cells, $CD4^+$ memory T cells, resting $CD4^+$ memory T cells, $CD4^+$ Th2 cells, natural killer (NK) cells, resting NK cells, M1 macrophages, and monocytes (Figure 2(a)). To verify the correlation of Trim35 expression and immune cell infiltration, IHC was conducted to determine the numbers of tumor-infiltrating T cells ($CD3^+$), B-cells ($CD19^+$), macrophages ($CD68^+$), and NK cells ($CD16^+$ and $CD56^+$) [31] in DLBCL. Representative IHC images are shown in Figure 2(b), and the cell densities (cell counts/mm) in 61 patients with DLBCL were quantified and plotted; the results showed a positive correlation between tumor-infiltrating NK cells ($CD16^+$ and $CD56^+$) and Trim35 (Figure 2(c)). Furthermore, better overall survival was observed in the patients with high infiltration of NK cells (CD56 positivity > 15%) than in those with low infiltration (CD56 positivity < 15%) ($p = 0.0205$). And the better overall survival was correlated with high expression of CD16 (samples with CD16 positivity < 0.8% of cells were considered to have low expression, and those with staining > 0.8% were considered to have high expression) (Figures 2(d) and 2(e)). Taken together, these results showed that Trim35 may suppress DLBCL progression by regulating NK cell infiltration.

3.3. Trim35 Is an E3 Ubiquitin Ligase for CLOCK and Promotes Its Proteasomal Degradation. The clock genes are reported to be involved in the regulation of the tumor microenvironment and cancer progression. To explore molecular mediators of the tumor suppressive function of Trim35, we overexpressed Flag-Trim35 in 293FT cells and then performed IP analysis to check the interaction between Trim35 and several clock proteins (Figure 3(a)). Among these proteins, CLOCK was successfully coimmunoprecipitated by Flag-Trim35, suggesting an interaction between the two proteins. Trim35 is reported as an E3 ligase, so we explored whether Trim35 could promote the ubiquitination and degradation of CLOCK. Trim35 overexpression induced CLOCK ubiquitination (Figure 3(b)) in 293FT cells. Besides, Trim35 knockout cell line was constructed, to verify the effective sgRNA for Trim35; we quantified the editing efficiency using the T7 Endonuclease I assay. The Trim35

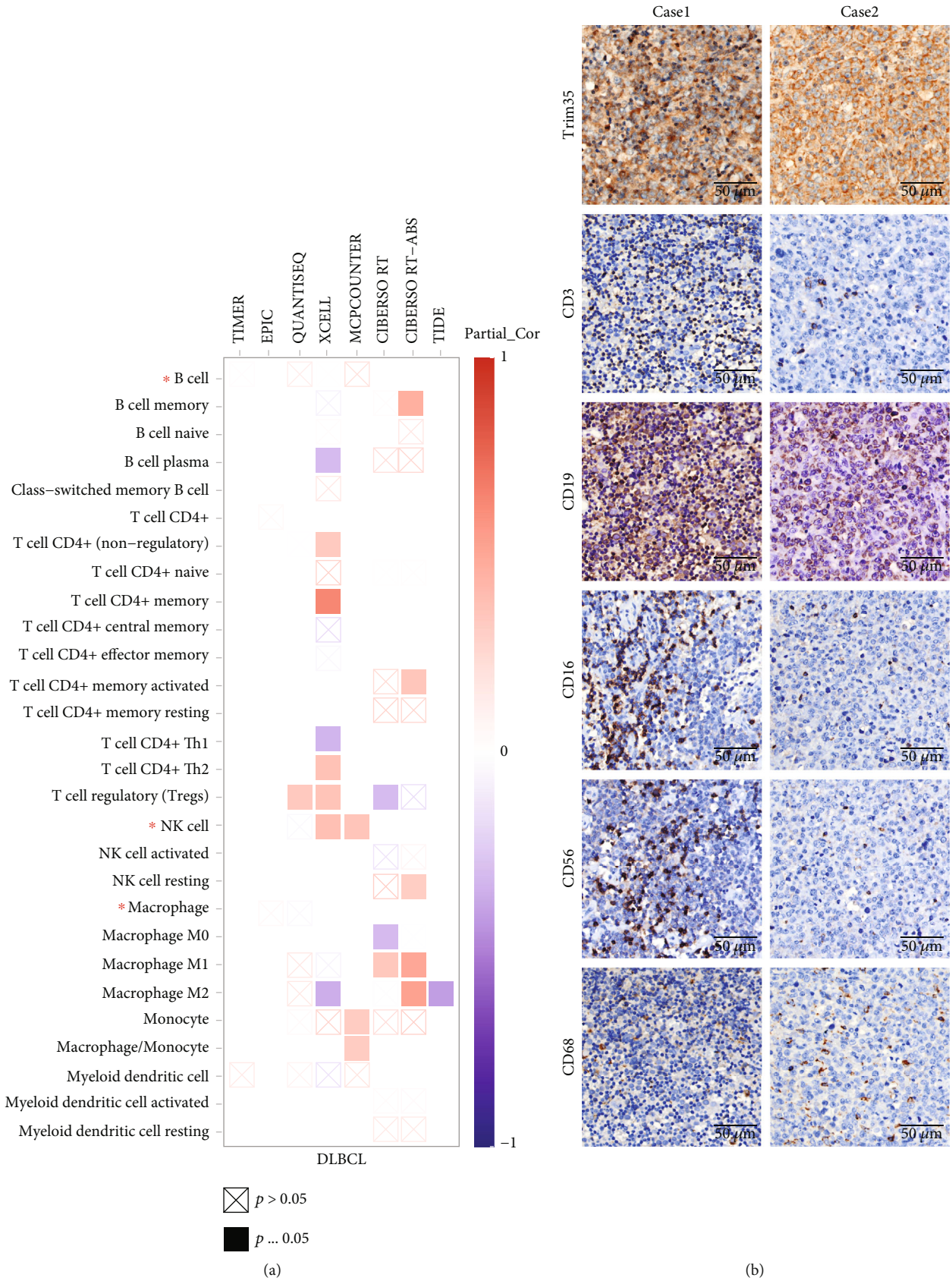


Figure 2: Continued.

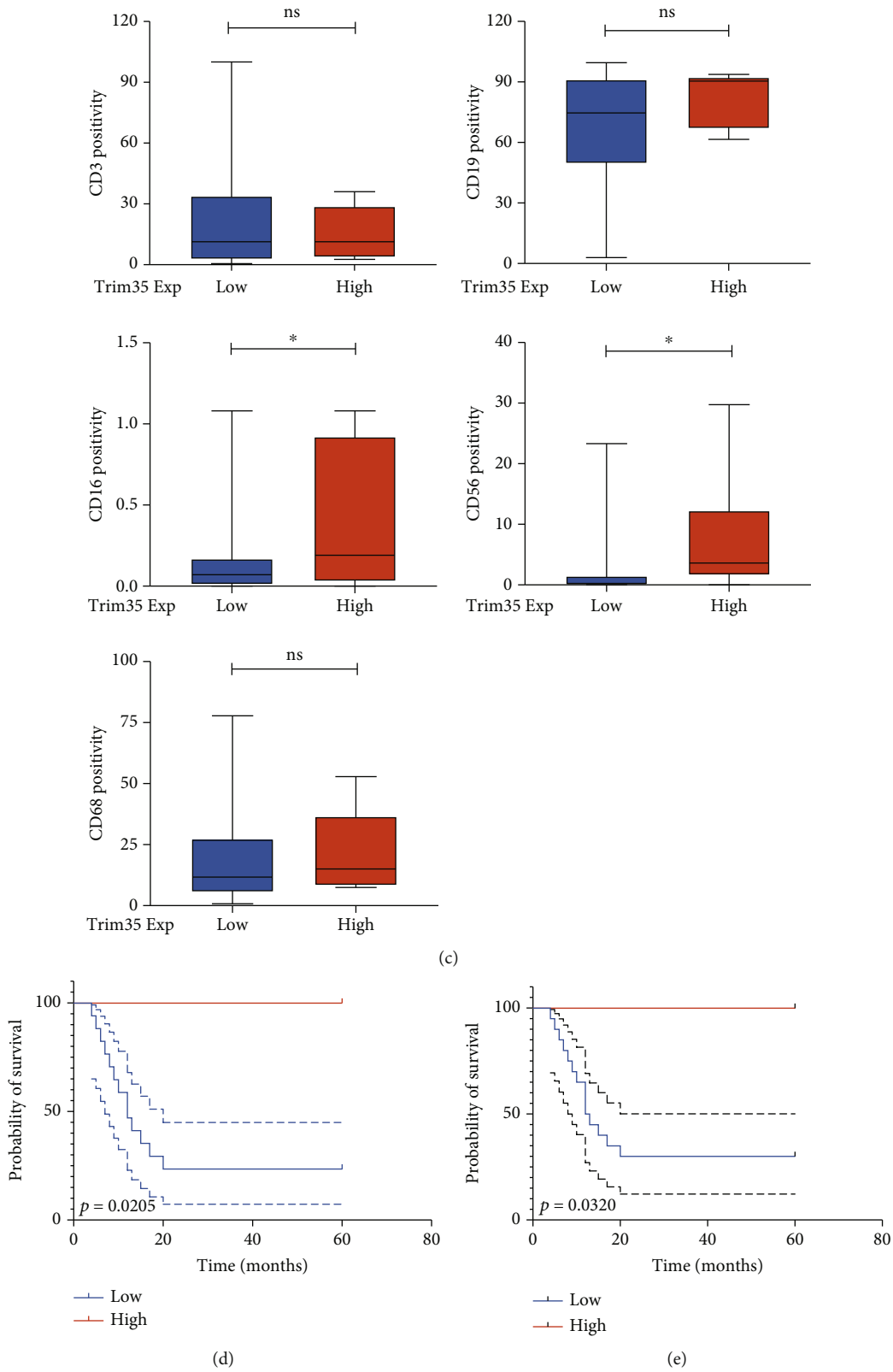


Figure 2: Trim35 is positively associated with NK cell infiltration and inhibits DLBCL progression. (a) Correlation of Trim35 expression with immune cells in DLBCL. (b) Representative IHC images of tumor-infiltrating $CD3^+$, $CD19^+$, $CD16^+$, $CD56^+$, and $CD68^+$ cells from two patients with high and low expression of Trim35. Scale bar, 50 μm. (c) The difference in CD3/CD19/CD16/CD56/CD68 positivity between the Trim35 subgroups was assessed by unpaired t test analysis. (d, e) Kaplan-Meier OS analysis in DLBCL with respect to phenotypic characteristics. Analysis of an NK cell marker ($CD56^+$ and $CD16^+$) was performed with evaluable immunostaining. $^{NS}p < 0.05$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

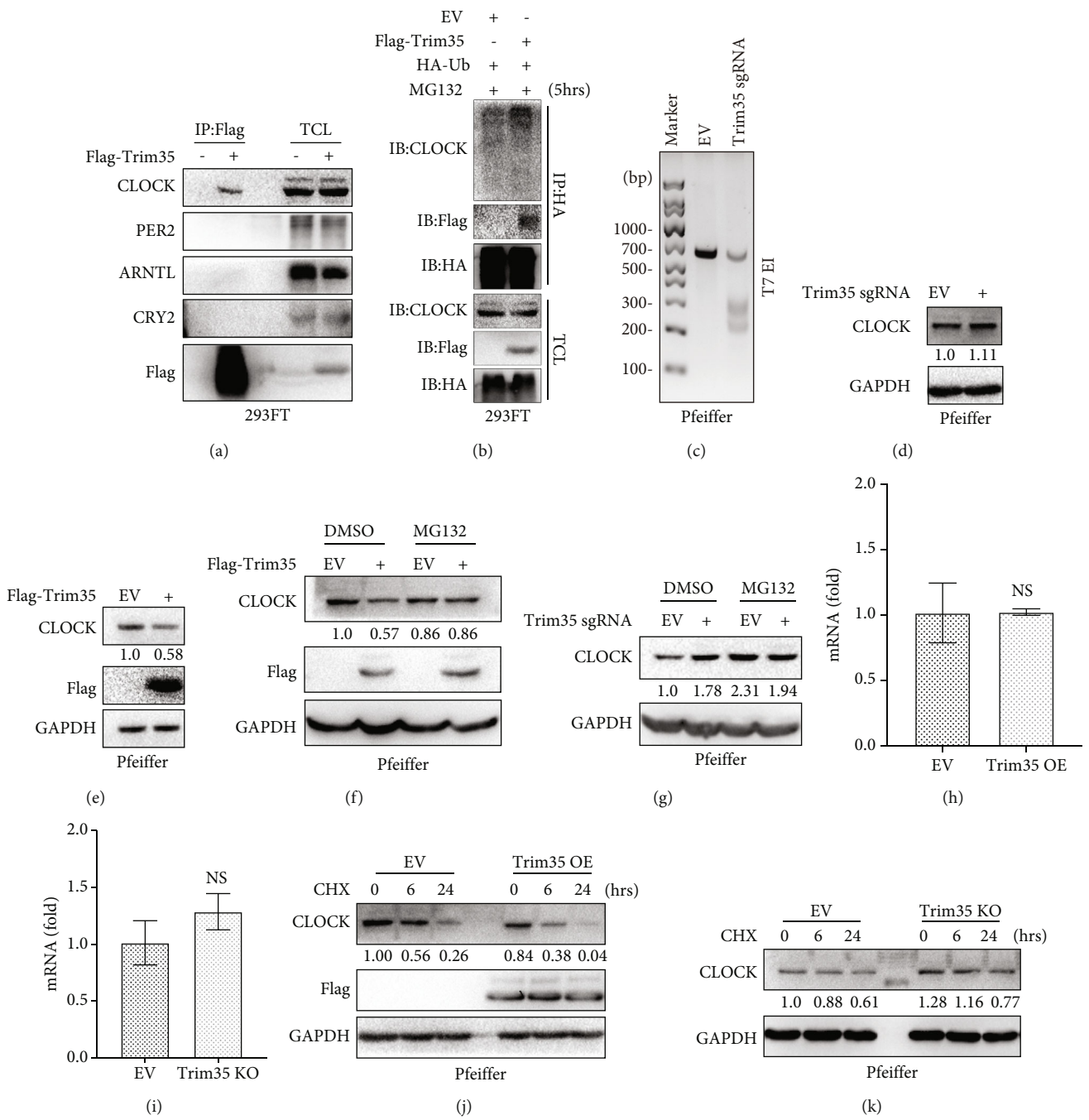


Figure 3: Trim35 controls the ubiquitination and degradation of CLOCK. (a) 293FT cells were transfected with Flag-tagged Trim35 or an empty vector. Forty-eight hours after transfection, cell lysates were prepared and subjected to IP with anti-Flag beads and immunoblotting as indicated. (b) Immunoblotting of lysed 293FT cells transfected with Trim35 or empty vector, along with HA-Ub. Cell lysates were prepared and subjected to IP with anti-HA beads and immunoblotting as indicated. (c) Trim35 sgRNAs are designed to target the human Trim35 loci. Agarose gel shows modification at both loci in transfected cells. (d) Pfeiffer cells were infected with viruses expressing Trim35 sgRNA or a control sgRNA and selected; cell lysates were immunoblotted for the indicated proteins. Densitometry quantified protein bands, presented as relative expression. (e) Pfeiffer cells were infected with Flag-Trim35 expression virus. Protein extracts were immunoblotted for the indicated proteins. CLOCK bands were quantified by densitometry and presented as relative expression. Pfeiffer cells with stable Trim35 overexpression (f) and Trim35-depleted Pfeiffer cells (g) were treated with a proteasome inhibitor MG132 (10 μg/mL) during the last 5 hours before lysis. CLOCK protein levels were detected and analyzed. Protein bands were quantified by densitometry and presented as relative expression. (h, i) The same cells as described in (e, d) were used for qRT-PCR analysis of CLOCK mRNA expression. Results are normalized to GAPDH mRNA level and expressed fold changes in mRNA expression compared with control. (j, k) The same cells as described in (e, d) were cultured for 6 h and 24 h before being further incubated with cycloheximide (CHX) for the indicated time points. The level of CLOCK at different time points was detected by western blot. Densitometry quantified CLOCK bands, presented as relative expression. Each experiment was successfully carried out three times. $^{NS}p > 0.05$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

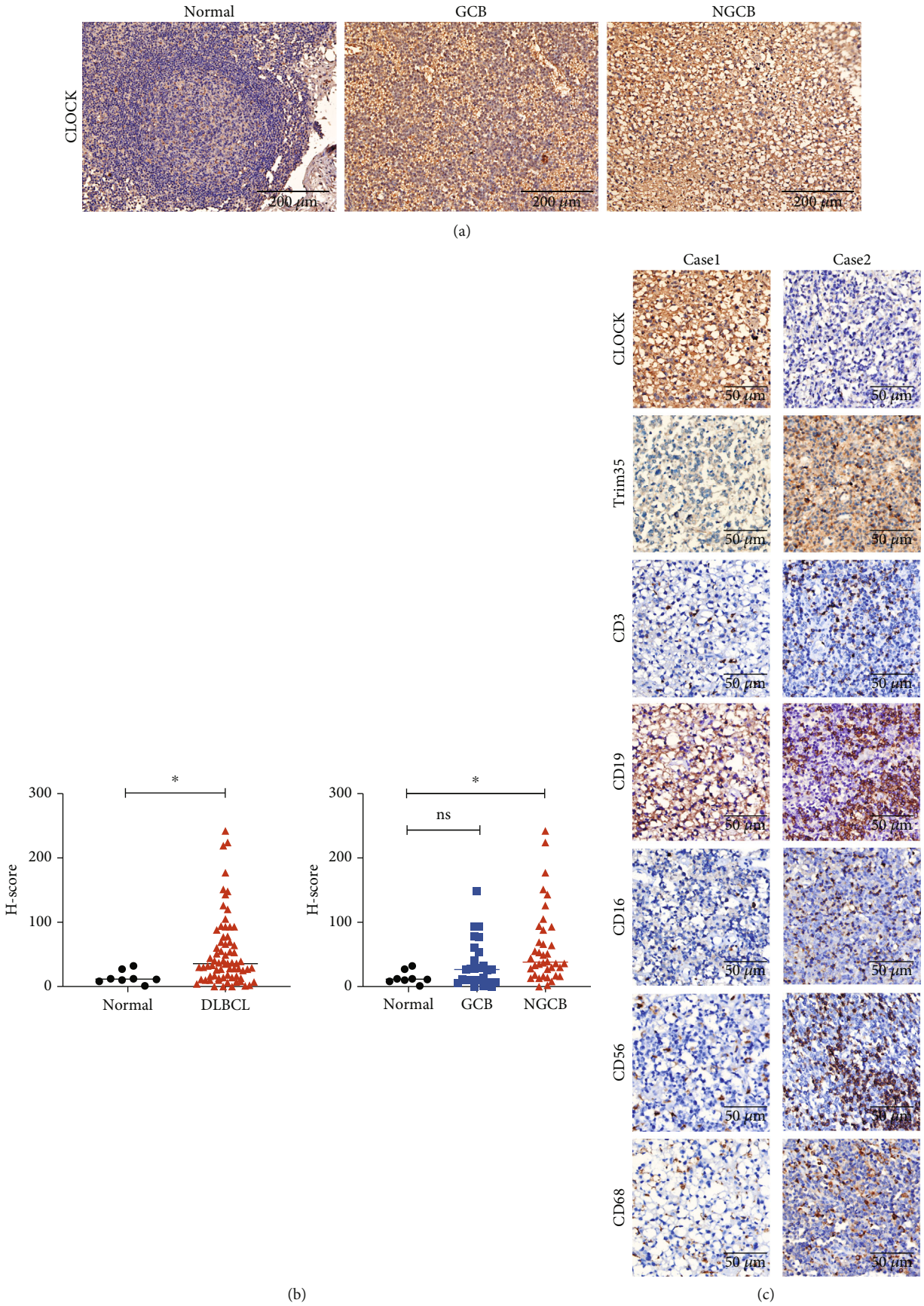


Figure 4: Continued.

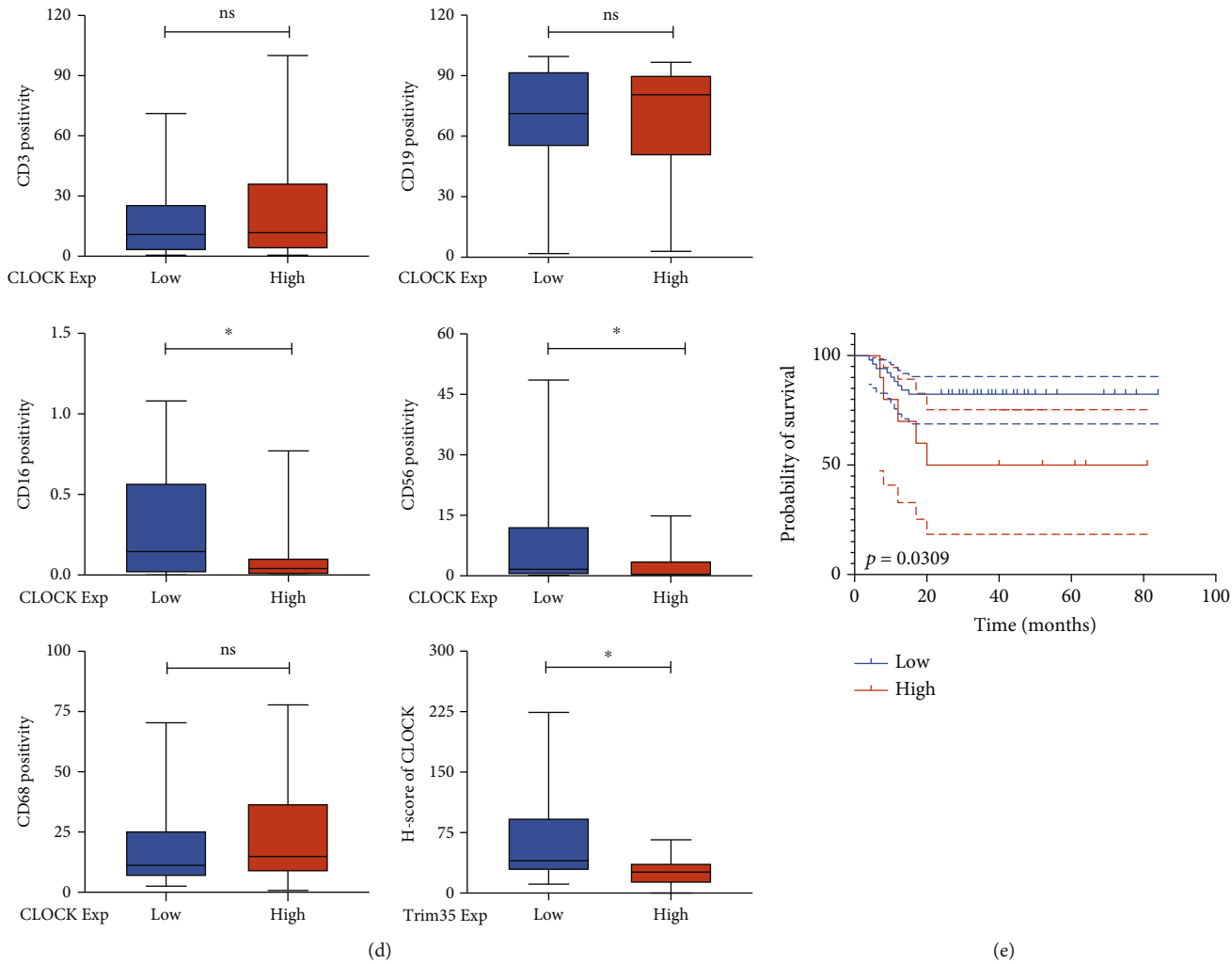


Figure 4: CLOCK is highly expressed in DLBCL and negatively correlated with NK cell infiltration. (a) Immunohistochemical staining for CLOCK in human DLBCL tissues and normal lymph node tissues. The specific CLOCK signal is seen in brown (DAB staining), and counterstaining was performed using hematoxylin (blue nuclei). Scale bar, 200 μm. (b) Summary for the entire cohort of DLBCL patients, as assessed by IHC. The p value was calculated by the chi-square test. (c) Representative IHC images of Trim35 and tumor-infiltrating $CD3^+$, $CD19^+$, $CD16^+$, $CD56^+$, and $CD68^+$ cells from two patients with high and low expression of CLOCK. Scale bar, 50 μm. (d) Differences in CD3/CD19/CD16/CD56/CD68 positivity between the two CLOCK subgroups were assessed by an unpaired t test. (e) Overall survival of patients with high (red) versus low (blue) CLOCK expression based on IHC staining. $^{NS}p > 0.05$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

sgRNAs were capable of inducing indels in specific targeted loci (Figure 3(c)). Consistent with this finding, depletion of Trim35 increased the CLOCK protein level in DLBCL cell lines (Figure 3(d)). Interestingly, stable overexpression of Trim35 in DLBCL cell lines reduced the endogenous protein level of CLOCK (Figure 3(e)). These observations suggested that Trim35 is involved in regulating CLOCK expression. Next, we analyzed endogenous CLOCK ubiquitination in the DLBCL cell line in the presence or absence of the proteasome inhibitor MG132. As expected, we found that Trim35-induced suppression of CLOCK could be efficiently abolished by MG132 (Figures 3(f) and 3(g)), which suggested that CLOCK was degraded through proteasome-mediated proteolysis degradation pathway. Further, we observed Trim35 has no effect on the mRNA level of CLOCK (Figures 3(h) and 3(i)). To further define whether Trim35 modulated the stability of endogenous CLOCK, we inhibited *de novo* protein levels of CLOCK in the Trim35 overexpressed and Trim35 knock out cells. Our data showed that Trim35 significantly induced degradation of CLOCK proteins in these cycloheximide-treated DLBCL cells (Figures 3(j) and 3(k)). These results suggest that Trim35 functions as an E3 ligase to promote the degradation of CLOCK through the proteasome pathway.

3.4. CLOCK Is Highly Expressed in DLBCL and Is Negatively Correlated with NK Cell Infiltration. To investigate whether Trim35 suppresses DLBCL progression via CLOCK, we first assessed the clinical relevance of trim35 in human DLBCL. CLOCK expression was analyzed in human DLBCL samples, and representative IHC images are shown in Figure 4(a). The expression of CLOCK was significantly increased in DLBCL

tissues compared with normal lymph node tissue (Figure 4(b)) and was inversely correlated with the expression of Trim35. Next, to further clarify the relevance of CLOCK in the immune microenvironment of human DLBCL, we determined the correlation between CLOCK expression and markers of several immune cells. Consistently, we found a negative correlation between CLOCK and CD56 and CD16, two biomarkers of NK cells (Figure 4(d)), and the representative IHC image is shown in (Figure 4(c)). Otherwise, the overall survival of patients with high CLOCK expression was significantly worse than that of patients with low CLOCK expression ($p = 0.0309$) (Figure 4(e)). These results indicate that Trim35 may remodel the tumor immune microenvironment via CLOCK.

4. Discussion

Our present study found a significant decrease in Trim35 expression in human DLBCL tissues compared with normal lymph node tissues, indicating the clinical relevance of Trim35 in human lymphomagenesis. We further showed that overexpression of Trim35 in human DLBCL cells can suppress their proliferation and correlates with poor survival in DLBCL. A mechanistic study showed that Trim35 functions as an E3 ligase to mediate the ubiquitination and degradation of CLOCK. Consistently, bioinformatics methods and IHC analysis of human DLBCL samples revealed that Trim35 and CLOCK expression were positively and negatively correlated with NK cell infiltration, respectively. In addition, our study confirmed that Trim35 and CLOCK expression and NK cell infiltration are independent prognostic variables of DLBCL. Future studies are required to uncover the detailed mechanism by which CLOCK regulates NK cell infiltration in DLBCL.

Human tumors can escape immune surveillance to enhance their survival [32, 33]. Anticancer immunotherapy has been shown to be a powerful way to cure cancer. NK cells reside in the peripheral blood and in some lymphoid and nonlymphoid organs and are promptly recruited to tumor sites as an important component of tumor immunosurveillance [31, 34]. Activated NK cells also rapidly secrete a variety of cytokines and chemokines, such as interferon γ (IFNγ), that promote the recruitment and activation of other participants in the antitumor response [35]. Their effector functions also represent a crucial factor in determining the response to anticancer therapy; for example, NK cells play an important role in rituximab-dependent killing of lymphoma cells via an antibody-dependent cellular cytotoxicity (ADCC) mechanism [36, 37]. A previous study also confirmed that an absolute decrease in NK cell count was predictive of no response and of shorter event-free survival and progression-free survival (PFS) in DLBCL, which suggested that the peripheral blood natural killer cell count (NKCC) could be a valuable biomarker in clinical practice and could pave the way for the development of novel combination treatment approaches for B-cell non-Hodgkin lymphoma (B-NHL) that are aimed at enhancing the activity of anti-CD20 antibodies by providing them with a sufficient number of functional effector cells [38–40]. Furthermore, the prognostic and therapeutic significance of the host response represented by NK cell infiltration has been described in Hodgkin lymphoma and DLBCL [41]. All of these findings highlight the crucial role of NK cells in DLBCL.

The target protein CLOCK is a key regulator of circadian rhythm, which is a biological mechanism that dictates an array of rhythmic physiological processes. The feature of this circadian control relies on cellular metabolism, both within the tumor microenvironment and organism systemically [42]. Previous studies have characterized the impact of the circadian clock on tumorigenesis and specific immune cells. For example, a recent study revealed that $CD4^+$ and $CD8^+$ T cells are correlated with core clock molecules, especially in lung adenocarcinomas and lung squamous cell carcinomas [43]. Additionally, it was reported that the tumor growth rate increased and latency decreased under circadian disruption conditions compared to normal light-dark (LD) schedules in a murine melanoma model [44]. However, little is known about the role of the clock gene in tumor microenvironment regulation and DLBCL progression. Our results reveal that the circadian gene CLOCK has a strong inverse correlation with NK cell infiltration in DLBCL tissues. In addition, we found that CLOCK is degraded by Trim35 via the ubiquitination-proteasome pathway. Consistently, CLOCK is overexpressed in DLBCL patients and is inversely correlated with the expression of Trim35.

To our knowledge, Trim35 is the first E3 ligase of CLOCK identified so far, which also extends our understanding of the suppressive role of Trim35 in tumor progression.

5. Conclusions

Overall, the present study revealed that Trim35 functions as an E3 ubiquitin ligase of CLOCK to promote its ubiquitination and proteasomal degradation; this suppresses DLBCL development by enhancing the infiltration of cytotoxic NK cells into tumors. Trim35 is a promising prognostic biomarker in DLBCL, and future studies should be conducted to further uncover the detailed mechanism by which the ubiquitination of clock proteins is related to remodeling of the tumor microenvironment and treatment response.

Data Availability

The data used to support the finding 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

We thank all individuals who participated in this work. This work was supported by the National Natural Science Foundation of China (81900199 and 81974465 to L.W.), the Hunan Natural Science Fund for Excellent Young Scholars (2019JJ30043 to L.W.), and the Talents Recruitment Project of Huxiang (2019RS1009 to L.W.)

References

[1] A. Rosenwald, G. Wright, W. C. Chan et al., "The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma," *The New England Journal of Medicine*, vol. 346, no. 25, pp. 1937–1947, 2002.

[2] A. A. Alizadeh, M. B. Eisen, R. E. Davis et al., "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, no. 6769, pp. 503–511, 2000.

[3] N. Khan and R. I. Fisher, "Subtype-specific therapy for DLBCL: are we there yet?," *Blood*, vol. 126, no. 16, pp. 1869-1870, 2015.

[4] G. Wright, B. Tan, A. Rosenwald, E. H. Hurt, A. Wiestner, and L. M. Staudt, "A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 17, pp. 9991–9996, 2003.

[5] H. Tilly, M. Gomes da Silva, U. Vitolo et al., "Diffuse large B-cell lymphoma (DLBCL): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†," *Annals of Oncology*, vol. 26, Suppl 5, pp. v116–v125, 2015.

[6] J. A. Mohawk, C. B. Green, and J. S. Takahashi, "Central and peripheral circadian clocks in mammals," *Annual Review of Neuroscience*, vol. 35, no. 1, pp. 445–462, 2012.

[7] P. L. Lowrey and J. S. Takahashi, "Mammalian circadian biology: elucidating genome-wide levels of temporal organization," *Annual Review of Genomics and Human Genetics*, vol. 5, no. 1, pp. 407–441, 2004.

[8] S. M. Reppert and D. R. Weaver, "Coordination of circadian timing in mammals," *Nature*, vol. 418, no. 6901, pp. 935–941, 2002.

[9] M. K. Bunger, L. D. Wilsbacher, S. M. Moran et al., "_Mop3_ is an essential component of the master circadian pacemaker in mammals," *Cell*, vol. 103, no. 7, pp. 1009–1017, 2000.

[10] N. Gekakis, D. Staknis, H. B. Nguyen et al., "Role of the CLOCK protein in the mammalian circadian mechanism," *Science*, vol. 280, no. 5369, pp. 1564–1569, 1998.

[11] D. P. King, Y. Zhao, A. M. Sangoram et al., "Positional cloning of the mouse circadian clock gene," *Cell*, vol. 89, no. 4, pp. 641–653, 1997.

[12] K. Kume, M. J. Zylka, S. Sriram et al., "mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop," *Cell*, vol. 98, no. 2, pp. 193–205, 1999.

[13] D. Nandi, P. Tahiliani, A. Kumar, and D. Chandu, "The ubiquitin-proteasome system," *Journal of Biosciences*, vol. 31, no. 1, pp. 137–155, 2006.

[14] D. Komander and M. Rape, "The ubiquitin code," *Annual Review of Biochemistry*, vol. 81, no. 1, pp. 203–229, 2012.

[15] F. Ikeda, N. Crosetto, and I. Dikic, "What determines the specificity and outcomes of ubiquitin signaling?," *Cell*, vol. 143, no. 5, pp. 677–681, 2010.

[16] D. Hoeller and I. Dikic, "Targeting the ubiquitin system in cancer therapy," *Nature*, vol. 458, no. 7237, pp. 438–444, 2009.

[17] E. J. Bennett and J. W. Harper, "DNA damage: ubiquitin marks the spot," *Nature Structural & Molecular Biology*, vol. 15, no. 1, pp. 20–22, 2008.

[18] Y. Yang, J. Kitagaki, H. Wang, D. X. Hou, and A. O. Perantoni, "Targeting the ubiquitin-proteasome system for cancer therapy," *Cancer Science*, vol. 100, no. 1, pp. 24–28, 2009.

[19] L. Hicke, "PtdIns(3,5)P_2 finds a partner," *Developmental Cell*, vol. 5, no. 3, pp. 363-364, 2003.

[20] T. Jo, M. Nishikori, Y. Kogure et al., "LUBAC accelerates B-cell lymphomagenesis by conferring resistance to genotoxic stress on B cells," *Blood*, vol. 136, no. 6, pp. 684–697, 2020.

[21] X. Jin, Q. Shi, Q. Li et al., "CRL3-SPOP ubiquitin ligase complex suppresses the growth of diffuse large B-cell lymphoma by negatively regulating the MyD88/NF-κB signaling," *Leukemia*, vol. 34, no. 5, pp. 1305–1314, 2020.

[22] S. M. Siepka, S. H. Yoo, J. Park et al., "Circadian mutant _overtime_ reveals F-box protein FBXL3 regulation of _cryptochrome_ and _period_ gene expression," *Cell*, vol. 129, no. 5, pp. 1011–1023, 2007.

[23] A. Reymond, G. Meroni, A. Fantozzi et al., "The tripartite motif family identifies cell compartments," *The EMBO Journal*, vol. 20, no. 9, pp. 2140–2151, 2001.

[24] D. Jia, L. Wei, W. Guo et al., "Genome-wide copy number analyses identified novel cancer genes in hepatocellular carcinoma," *Hepatology*, vol. 54, no. 4, pp. 1227–1236, 2011.

[25] N. Sun, L. Jiang, M. Ye et al., "TRIM35 mediates protection against influenza infection by activating TRAF3 and degrading viral PB2," *Protein & Cell*, vol. 11, no. 12, pp. 894–914, 2020.

[26] C. P. Hans, D. D. Weisenburger, T. C. Greiner et al., "Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray," *Blood*, vol. 103, no. 1, pp. 275–282, 2004.

[27] T. Li, J. Fu, Z. Zeng et al., "TIMER2.0 for analysis of tumor-infiltrating immune cells," *Nucleic Acids Research*, vol. 48, no. W1, pp. W509–w514, 2020.

[28] B. Li, E. Severson, J. C. Pignon et al., "Comprehensive analyses of tumor immunity: implications for cancer immunotherapy," *Genome Biology*, vol. 17, no. 1, p. 174, 2016.

[29] P. Danaher, S. Warren, L. Dennis et al., "Gene expression markers of tumor infiltrating leukocytes," *Journal for Immunotherapy of Cancer*, vol. 5, no. 1, p. 18, 2017.

[30] S. Sousa and J. Määttä, "The role of tumour-associated macrophages in bone metastasis," *J Bone Oncol*, vol. 5, no. 3, pp. 135–138, 2016.

[31] M. Cooper, T. Fehniger, and M. Caligiuri, "The biology of human natural killer-cell subsets," *Trends in Immunology*, vol. 22, no. 11, pp. 633–640, 2001.

[32] G. P. Dunn, A. T. Bruce, H. Ikeda, L. J. Old, and R. D. Schreiber, "Cancer immunoediting: from immunosurveillance to tumor escape," *Nature Immunology*, vol. 3, no. 11, pp. 991–998, 2002.

[33] J. B. Swann and M. J. Smyth, "Immune surveillance of tumors," *The Journal of Clinical Investigation*, vol. 117, no. 5, pp. 1137–1146, 2007.

[34] M. A. Caligiuri, "Human natural killer cells," *Blood*, vol. 112, no. 3, pp. 461–469, 2008.

[35] G. P. Dunn, C. M. Koebel, and R. D. Schreiber, "Interferons, immunity and cancer immunoediting," *Nature Reviews. Immunology*, vol. 6, no. 11, pp. 836–848, 2006.

[36] M. J. Robertson and J. Ritz, "Biology and clinical relevance of human natural killer cells," *Blood*, vol. 76, no. 12, pp. 2421–2438, 1990.

[37] A. S. Chretien, A. le Roy, N. Vey et al., "Cancer-induced alterations of NK-mediated target recognition: current and investigational pharmacological strategies aiming at restoring NK-mediated anti-tumor activity," *Frontiers in Immunology*, vol. 5, p. 122, 2014.

[38] N. R. Lee, E. K. Song, K. Y. Jang et al., "Prognostic impact of tumor infiltrating FOXP3 positive regulatory T cells in diffuse

large B-cell lymphoma at diagnosis," *Leukemia & Lymphoma*, vol. 49, no. 2, pp. 247–256, 2008.

[39] R. Shen, P. P. Xu, N. Wang et al., "Influence of oncogenic mutations and tumor microenvironment alterations on extranodal invasion in diffuse large B-cell lymphoma," *Clinical and Translational Medicine*, vol. 10, no. 7, article e221, 2020.

[40] A. Plonquet, C. Haioun, J. P. Jais et al., "Peripheral blood natural killer cell count is associated with clinical outcome in patients with aaIPI 2-3 diffuse large B-cell lymphoma," *Annals of Oncology*, vol. 18, no. 7, pp. 1209–1215, 2007.

[41] F. Vari, D. Arpon, C. Keane et al., "Immune evasion via PD-1/PD-L1 on NK cells and monocyte/macrophages is more prominent in Hodgkin lymphoma than DLBCL," *Blood*, vol. 131, no. 16, pp. 1809–1819, 2018.

[42] A. Verlande and S. Masri, "Circadian clocks and cancer: timekeeping governs cellular metabolism," *Trends in Endocrinology and Metabolism: TEM*, vol. 30, no. 7, pp. 445–458, 2019.

[43] Y. Yang, G. Yuan, H. Xie et al., "Circadian clock associates with tumor microenvironment in thoracic cancers," *Aging (Albany NY)*, vol. 11, no. 24, pp. 11814–11828, 2019.

[44] I. Aiello, M. L. M. Fedele, F. Román et al., "Circadian disruption promotes tumor-immune microenvironment remodeling favoring tumor cell proliferation," *Science Advances*, vol. 6, no. 42, p. eaaz4530, 2020.

Hindawi
Journal of Immunology Research
Volume 2021, Article ID 9989321, 17 pages
https://doi.org/10.1155/2021/9989321

Research Article

The Value of Immune-Related Genes Signature in Osteosarcoma Based on Weighted Gene Co-expression Network Analysis

Xin Wang,[1] Li Gan,[2] Ju Ye,[3] Mengjie Tang,[4] and Wei Liu[5]

[1]*Department of Bone and Soft Tissue, Hunan Cancer Hospital, The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha 410013, China*
[2]*Department of Anesthesiology, Hunan Cancer Hospital, The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha 410013, China*
[3]*Department of Pharmacy, Zunyi Medical University, Zunyi 563000, China*
[4]*Department of Pathology, Hunan Cancer Hospital, The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha 410013, China*
[5]*Department of Pharmacy, The Third Xiangya Hospital, Central South University, Changsha 410013, China*

Correspondence should be addressed to Mengjie Tang; tangmengjie@hnca.org.cn and Wei Liu; liuweixy3@csu.edu.cn

Received 16 March 2021; Accepted 25 April 2021; Published 15 May 2021

Academic Editor: Jialiang Liang

Copyright © 2021 Xin 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. Osteosarcoma (OS) is a serious malignant tumor that is more common in adolescents or children under 20 years of age. This study is aimed at obtaining immune-related genes (IRGs) associated with the progression and prognosis of OS. *Method*. Expression profiling data and clinical data for OS were downloaded from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) database. ESTIMATE calculates immune scores and stromal scores of samples and performs the prognostic analysis. Weighted gene coexpression network analysis (WGCNA) was used to find modules correlated with immune and stromal scores. Cox regression analysis and least absolute shrinkage and selection operator (LASSO) analysis were used to explore IRGs associated with OS prognosis and construct and validate a hazard score model. Finally, we verified the expression and function of EVI2B in OS. *Results*. WGCNA selected twenty-eight IRGs, 10 of which were associated with OS prognosis, and LASSO further obtained three key prognostic genes. A prognostic model of EVI2B was constructed, and according to the risk score model, patients in the high-risk group had a worse prognosis than those in the low-risk group, and the prognosis was statistically significant in the high- and low-risk groups. Receiver operating characteristic (ROC) curves were used to assess the prognostic model's accuracy and externally validate the independent GSE21257 cohort. The results of immunohistochemical staining and qPCR showed that EVI2B was a tumor suppressor gene. The differential genes in the high- and low-risk groups were analyzed by enrichment analysis of GO and KEGG, indicating that the EVI2B model is associated with immune response. *Conclusion*. In this study, IRG EVI2B is closely related to OS's prognosis and can be used as a potential biomarker for prognosis and treatment of OS.

1. Introduction

Osteosarcoma is a malignant tumor that occurs more commonly in adolescents or children under 20 years of age [1]. Conventional therapies for osteosarcoma include surgery, adjuvant chemotherapy, and neoadjuvant chemotherapy [2–4]. For patients with osteosarcoma, the use of standard multiagent chemotherapy in combination with surgical resection produces a long-term survival of approximately 70% for the localized disease at diagnosis [5, 6] and 20–30% for metastatic disease at diagnosis or recurrence [7]. Although advances have been made in surgical techniques, targeted therapy and tumor immunity, and complications such as infection and poor survival due to limb salvage surgery, there are still many issues to be addressed in the treatment of OS; therefore, there is an urgent need to develop new prediction methods to improve OS patients' survival [5, 8].

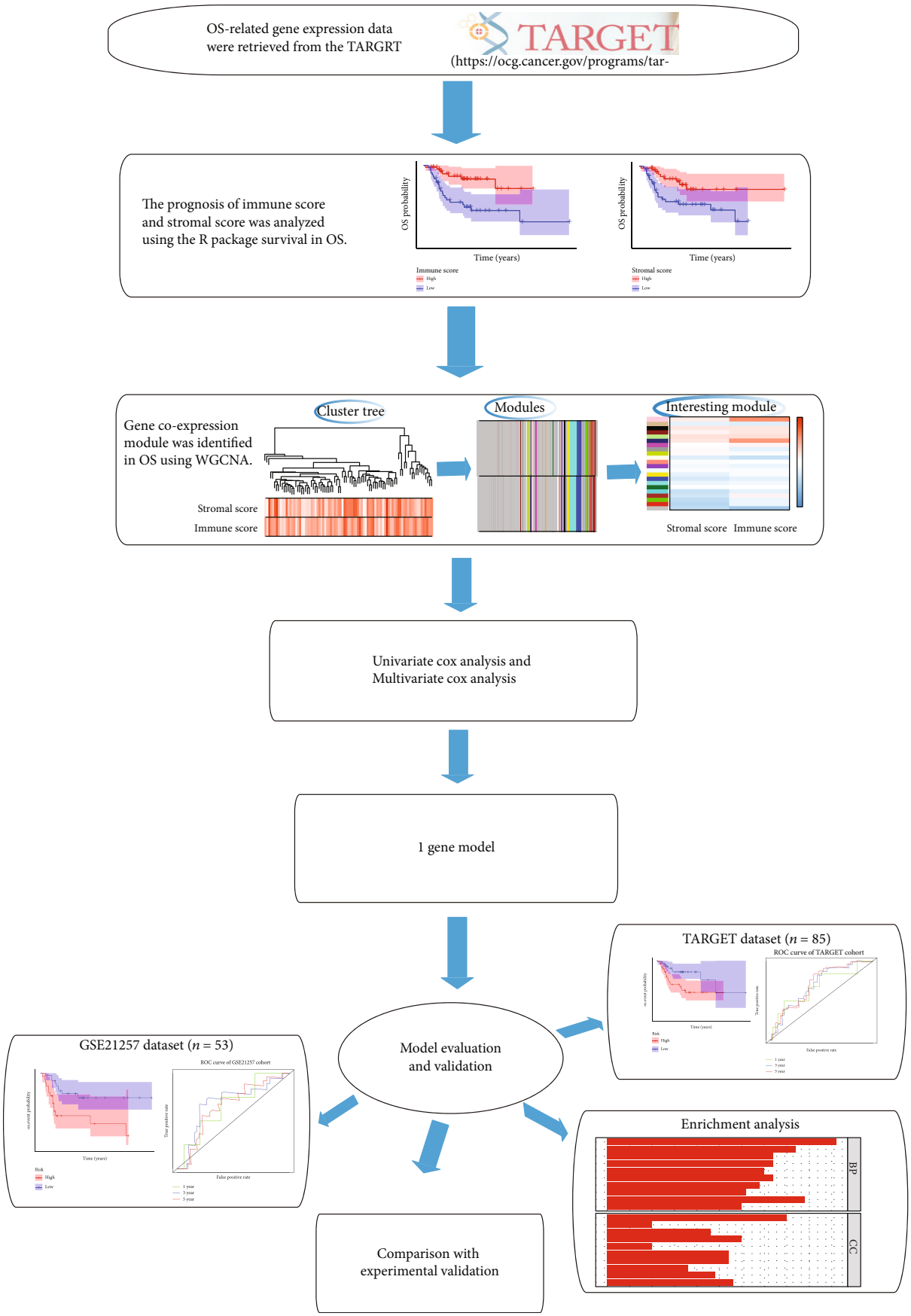


Figure 1: Flowchart of the present study.

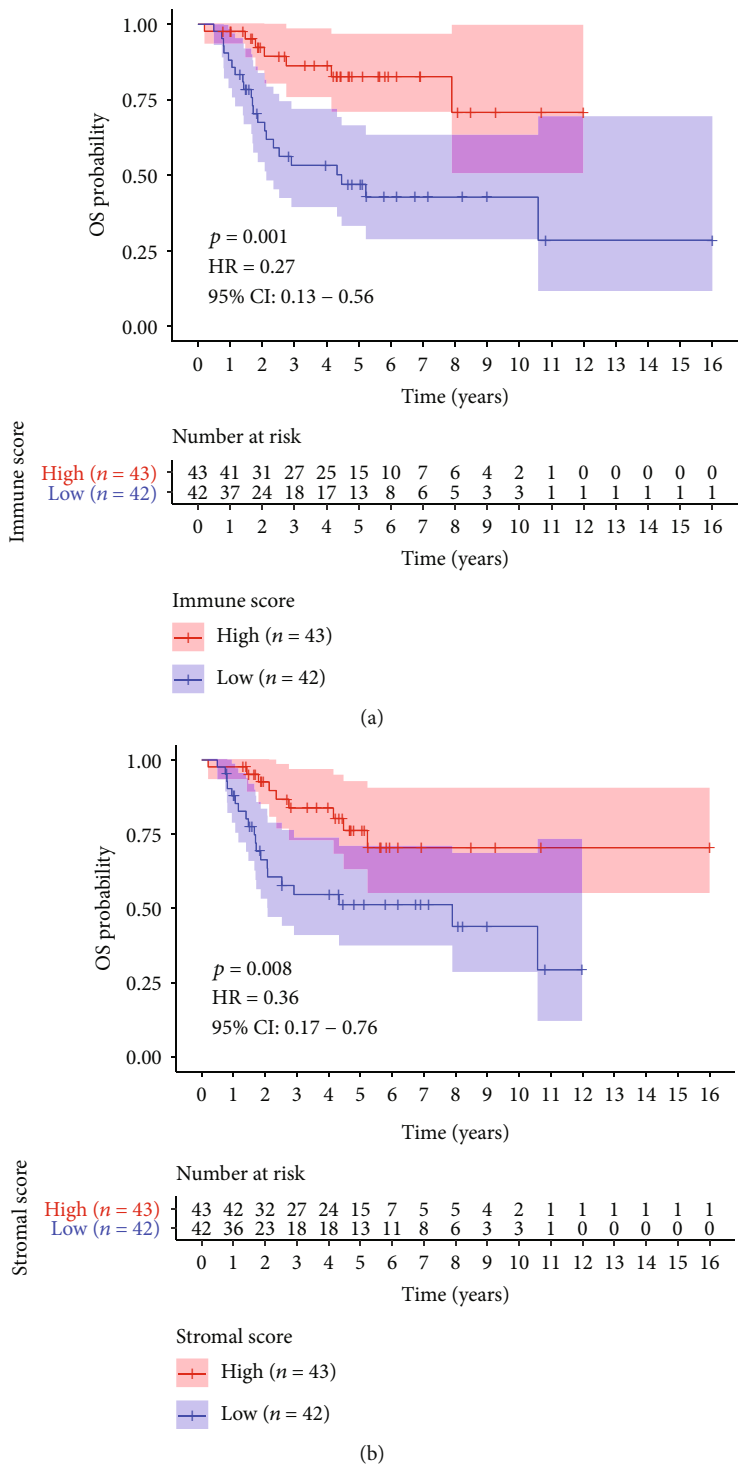
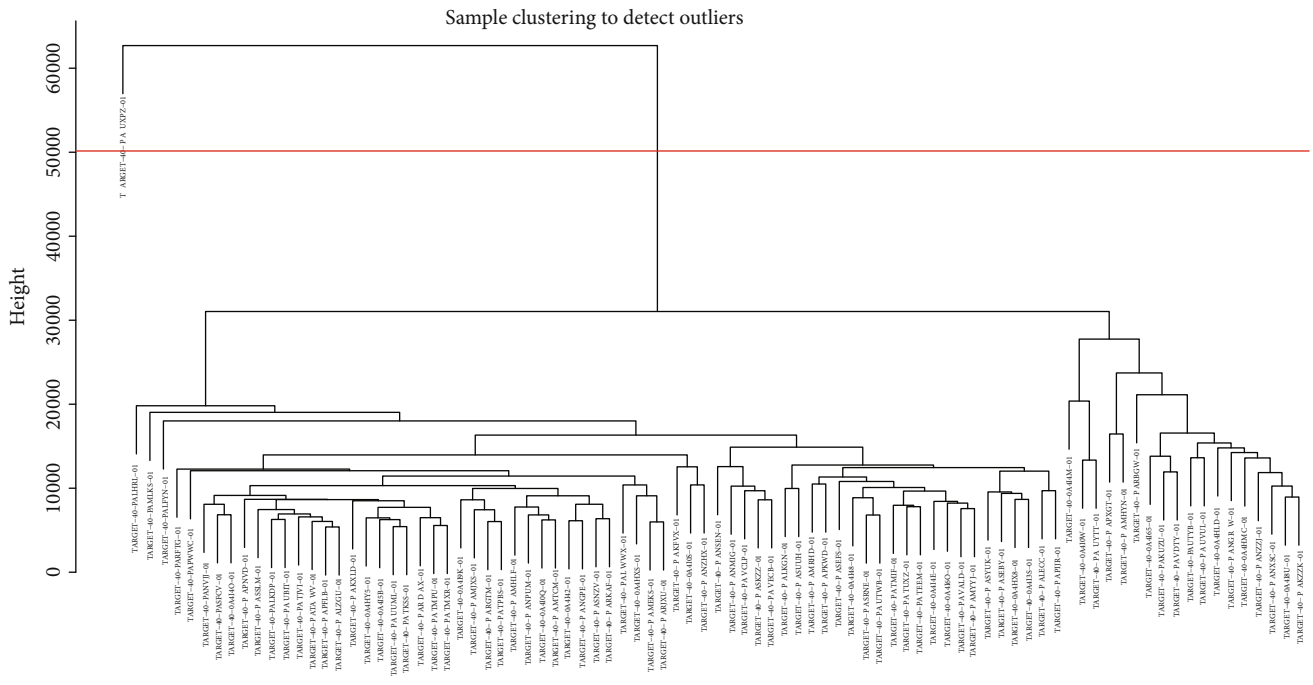
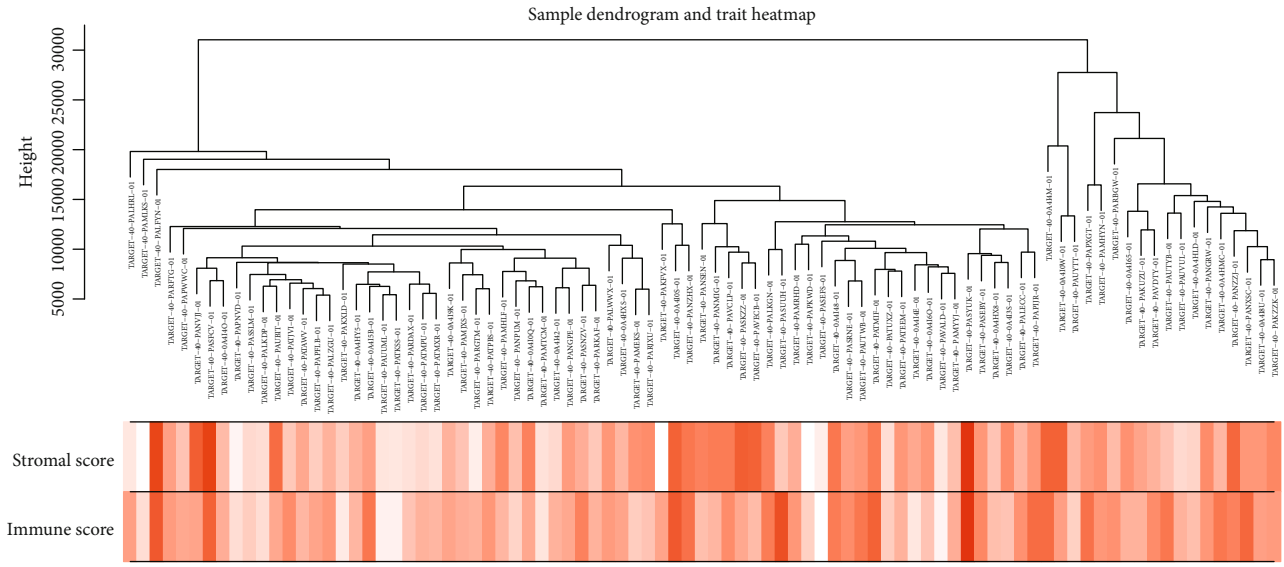


Figure 2: The relationship between immune score and stromal score for OS prognosis. (a) Survival analysis of the correlation between immune score and overall survival time in OS. (b) Survival analysis of the correlation between stromal score and overall survival time in OS. Immune/stromal scores: the significance is to predict tumor purity by immune and stromal scores. Log-rank p values were less than 0.05 with statistically significant. OS: osteosarcoma.

A large body of evidence suggests that OS has an immune system with multiple therapeutic targets, including receptor T cell therapy [9], HER2-specific immunity [10–12], and adjuvant immune therapy with mifamurtide [13]. Immune checkpoint inhibitors may develop immune tolerance to tumor immunosuppressants, thereby increasing endogenous antitumor activity, and could improve therapy-induced tumor immunogenic chemotherapy, radiotherapy, and targeted therapy [13–17]. To date, the impact of IRGs on OS prognosis is not well clear, so prognostic

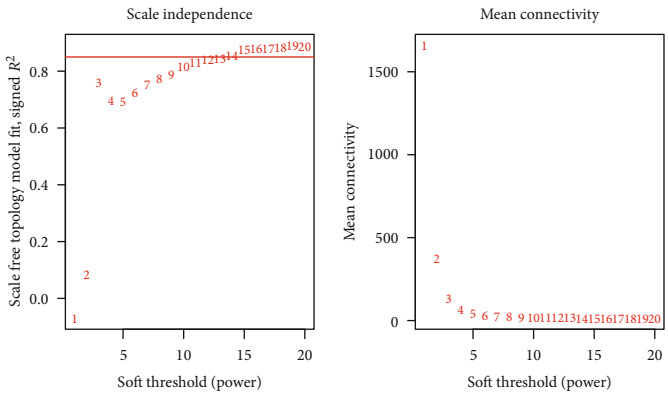


(a)

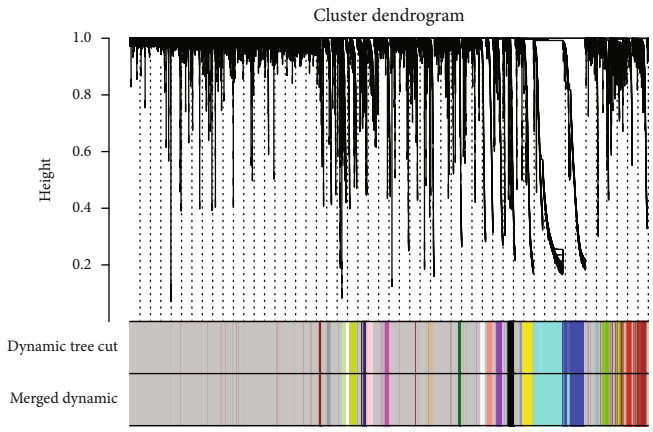


(b)

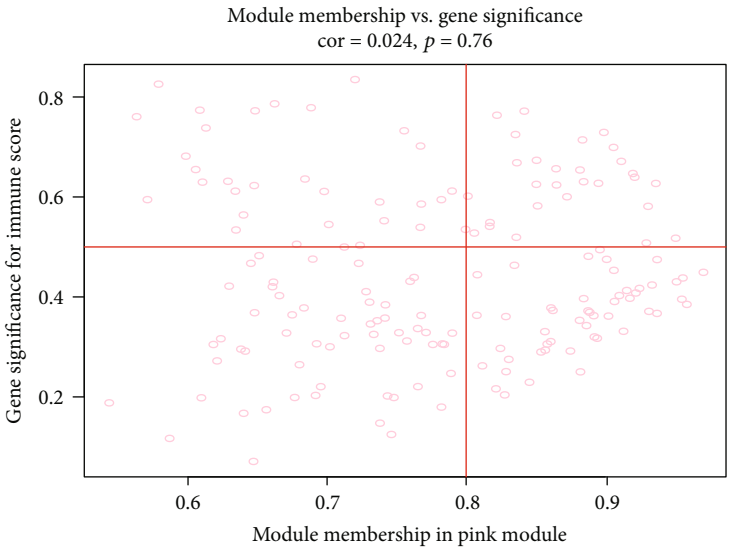
Figure 3: Continued.



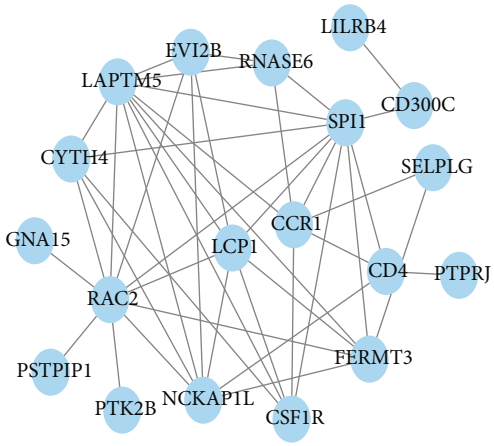
(c) (d)



(e)

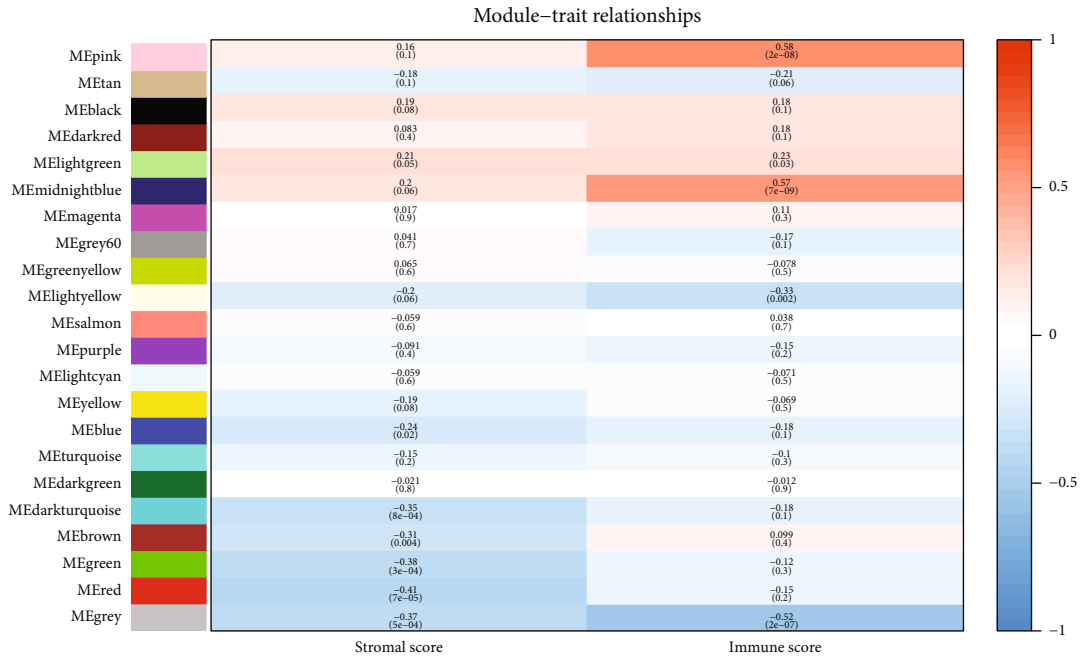


(f)



(g)

Figure 3: Continued.



(h)

Figure 3: The samples clustering dendrogram, immune score, stromal score, and the determination of soft-thresholding power analyzed by WGCNA. (a) Sample clustering was performed to detect outlying samples based on immune score and stromal score. The red line indicates the cut-off point for data filtering in the data preprocessing step. (b) Clustering based on incorporating immune scores and stromal scores of expression data from osteosarcomas with clinical data and color intensity is proportional to immune and stromal scores. (c) Analysis of the scale-free fit index for various soft-thresholding powers (β). (d) Analysis of the mean connectivity for various soft-thresholding powers. (e) Clustering dendrogram of genes based on a dissimilarity measure (1-TOM). (f) Pink coexpression modules based on immune scores yielded scatter plots of gene significance (GS) versus module membership (MM). The screening conditions were the correlation between genes, and the pink module was more significant than 0.8, and the correlation coefficient between genes and the immune score was greater than 0.5. (g) The DEG PPI network module analysis via Cytoscape software. (h) Heatmap of the association with MES and immune score and stromal score of OS. OS: osteosarcoma; TOM: topological overlap stromal; MES: module eigengenes; PPI: protein-protein interaction; DEGs: differentially expressed genes.

biomarkers with IRG are clinically essential for OS prognosis identification.

A standard analysis method relates the measurement of these genomic covariates to patients' survival time, which often censored survival data [18]. A popular strategy is to use these covariates to fit a Cox regression model to the censored survival data and then predict new cancer prognosis based on this fitted model [18–20]. Compared with Cox risk regression analysis, data overfitting can be solved entirely by the LASSO analysis [21]. LASSO variants are a popular strategy to provide variable selection in regression analysis and have extended to the Cox regression model [18, 22, 23]. Regression with LASSO penalty is a commonly used method for selection in a high-dimensional variable. Still, the results depend heavily on the value of the shrinkage setting λ [23]. Systems biology algorithms for WGCNA have been used to assess the association between gene sets and clinical features by constructing scale-free gene coexpression networks [24–27].

This study focused on revealing IRGs involved in OS prognosis. The WGCNA analysis and PPI network analysis were first used by constructing a Cox proportional hazards (PH) prognostic model for LASSO [18, 19, 22]. Univariate and multivariate Cox regression analyses were performed for IRGs in PPI, and the hub genes identified could be used for prognostic risk assessment in OS patients. Finally, we validated the prediction model's performance and accuracy by ROC analysis. Furthermore, the prognostic significance of this model was validated on an independent GEO dataset. To reveal the role of differential genes in these training groups, functional and pathway enrichment analysis was performed using differential genes. The results showed that these differential genes were mainly associated with immune response and inflammatory response. We also performed immunohistochemical staining and qPCR experimental validation, and the results were consistent with the bioinformatics results. In this study, IRG is involved in the immune process to affect the prognosis of OS patients, thereby intervening in patients' immune response to improve the poor prognosis and enhance the quality of life of cancer patients.

2. Materials and Methods

2.1. Data Collection and Processing. Gene expression data (FPKM, fragments per kilobase million) and clinical data for OS obtained from the TARGET database (https://ocg.cancer.gov/programs/target). Log_2 (FPKM+1) conversion for the expression value of the TARGET database. Clinicopathological data of the corresponding patients retrieved

Table 1: 28 key genes selected in the pink module.

Probes	Module color	GS. immune score	p.GS. immune score	MM pink	p.MM pink
PTPRJ	Pink	0.601266559	7.37*E*-10	0.800403944	1.38*E*-20
SCARF1	Pink	0.527434914	1.53*E*-07	0.805025206	5.65*E*-21
CD300C	Pink	0.540904961	6.35*E*-08	0.816077901	6.05*E*-22
ANPEP	Pink	0.548497065	3.80*E*-08	0.816287387	5.79*E*-22
TBXAS1	Pink	0.763212372	8.47*E*-18	0.821317927	1.99*E*-22
RNASE6	Pink	0.724597637	2.13*E*-15	0.83463483	9.93*E*-24
HSD3B7	Pink	0.5189629	2.61*E*-07	0.835214408	8.67*E*-24
EVI2B	Pink	0.668107494	1.56*E*-12	0.835636508	7.84*E*-24
NCKAP1L	Pink	0.771339655	2.31*E*-18	0.840965442	2.17*E*-24
PTK2B	Pink	0.624883531	9.88*E*-11	0.849623447	2.43*E*-25
SPI1	Pink	0.673046116	9.26*E*-13	0.849796638	2.32*E*-25
GRN	Pink	0.581838777	3.41*E*-09	0.850541409	1.91*E*-25
SQOR	Pink	0.656087588	5.28*E*-12	0.863819471	4.90*E*-27
GLRX	Pink	0.62368047	1.10*E*-10	0.864140646	4.46*E*-27
APOBR	Pink	0.600016784	8.15*E*-10	0.871667349	4.66*E*-28
GNA15	Pink	0.653638843	6.73*E*-12	0.881020282	2.28*E*-29
LILRB4	Pink	0.713838334	8.46*E*-15	0.882807578	1.25*E*-29
LAPTM5	Pink	0.630129195	6.18*E*-11	0.88363821	9.39*E*-30
CCR1	Pink	0.626964574	8.21*E*-11	0.894293907	1.99*E*-31
FERMT3	Pink	0.728808967	1.22*E*-15	0.898203844	4.36*E*-32
CYTH4	Pink	0.698812331	5.23*E*-14	0.905139124	2.52*E*-33
LCP1	Pink	0.670851401	1.17*E*-12	0.910716239	2.16*E*-34
CD4	Pink	0.646606036	1.33*E*-11	0.91904082	4.02*E*-36
CSF1R	Pink	0.639120086	2.71*E*-11	0.9203441	2.07*E*-36
PSTPIP1	Pink	0.507580891	5.23*E*-07	0.928727774	2.19*E*-38
LAT2	Pink	0.580893451	3.67*E*-09	0.930001029	1.05*E*-38
SELPLG	Pink	0.626714866	8.39*E*-11	0.935560532	3.49*E*-40
RAC2	Pink	0.517122434	2.92*E*-07	0.949814884	1.14*E*-44

Table 2: 10 prognostic genes for overall survival of OS patients assessed by univariate regression analysis.

Gene	HR	*z*	*p* value	Lower	Upper
EVI2B	0.627242	-2.6835	0.007286	0.446155	0.881827
GRN	0.588554	-2.57038	0.010159	0.392866	0.881716
NCKAP1L	0.570277	-2.51025	0.012065	0.367825	0.884159
SELPLG	0.653439	-2.50146	0.012368	0.46818	0.912005
LILRB4	0.625401	-2.22715	0.025937	0.413782	0.945248
FERMT3	0.698225	-2.22338	0.02619	0.508714	0.958334
SQOR	0.601923	-2.18736	0.028716	0.381945	0.948596
CYTH4	0.676584	-2.13248	0.032968	0.472465	0.968888
GNA15	0.686282	-2.06624	0.038806	0.480192	0.980824
RNASE6	0.752659	-1.99037	0.04655	0.568959	0.995669

from the database, including gender, ethnicity, age, tumor location, time to recurrence, and survival information. In this study, data from the GEO database were used (http://www.ncbi.nlm.nih.gov/geo/) of an independent dataset for external validation, downloaded, and collected a high-throughput gene expression dataset, numbered GSE21257 (series matrix file). The Ensemble IDs of the mRNAs in the TARGET database were extracted from the GENCODE project. The ID conversion for the GSE21257 data set is provided by GPL10295 (Illumina human-6 v2.0 expression beadchip).

2.2. Correlation between Prognosis and Stromal/Immune Score. The ESTIMATE algorithm was applied to the normalized expression stromal [28] to estimate each osteosarcoma sample's stromal and immune score. The overall survival deeds the primary prognostic endpoint, the immune score, and the stromal score of each sample calculated with the R package ESTIMATE; it analyzed overall prognosis survival using the R package survival.

2.3. Construction of Coexpression Network. TARGET's data set used as the training set, the WGCNA program package [29, 30] in R software used to remove outlier samples and find the modules related to immune score and stromal score toolkit of heat map drawn to analyze the strength of interacting. Module-trait correlations were estimated using correlations between module signature genes and traits (immune

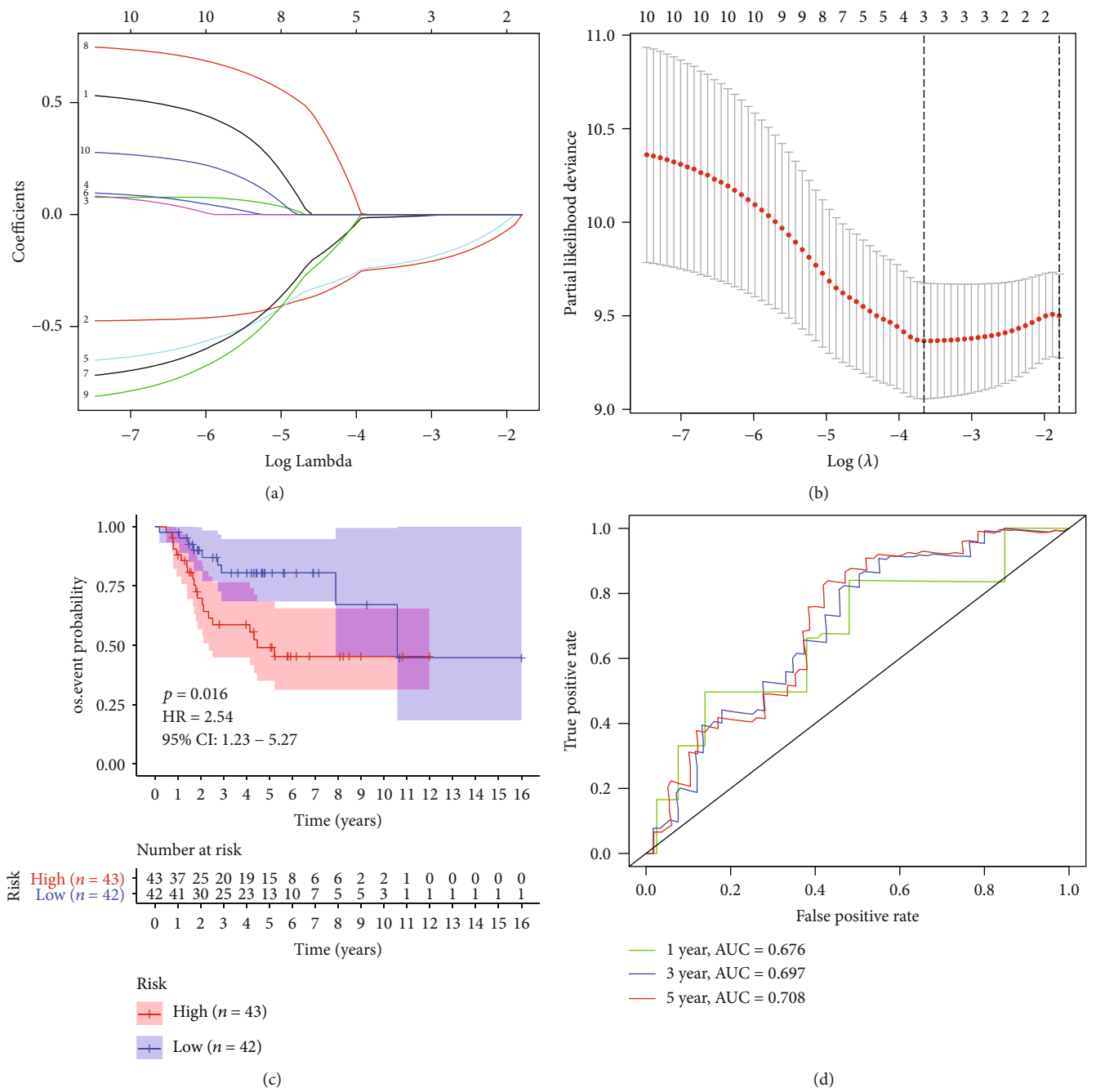


Figure 4: The results of the Cox prognostic model and verification. (a) Determine parameter Lambda in LASSO. (b) Crossvalidation determination Lambda. (c) Survival analysis between high- and low-risk groups in TARGET. (d) The result for the ROC curve of the TARGET cohort.

score and stromal score). The soft-threshold power of β calculates using a scale-free topology criterion. A weighted adjacency stromal generates, and the soft-threshold power β can emphasize weak and robust correlations between penalized genes [24, 31]. Modules of RNA were obtained using the dynamic tree cutting method. It is classifying genes with similar expression profiles by modules; the average linkage was performed to perform hierarchical clustering by TOM-based dissimilarity [24]. Finally, the difference of module eigengenes (MES) of the module dendrogram was calculated, and some modules were merged. The pink module with the highest correlation coefficient was selected as the next research object. The conditions for the screening of IRGs in the pink module were that the correlation between genes and the pink module was more significant than 0.8, the correlation coefficient between genes and the immune score was greater than 0.5, and a total of 28 IRGs were selected.

2.4. PPI Network Analysis. Using search tools for reciprocal gene/protein retrieval (https://string-db.org/) database

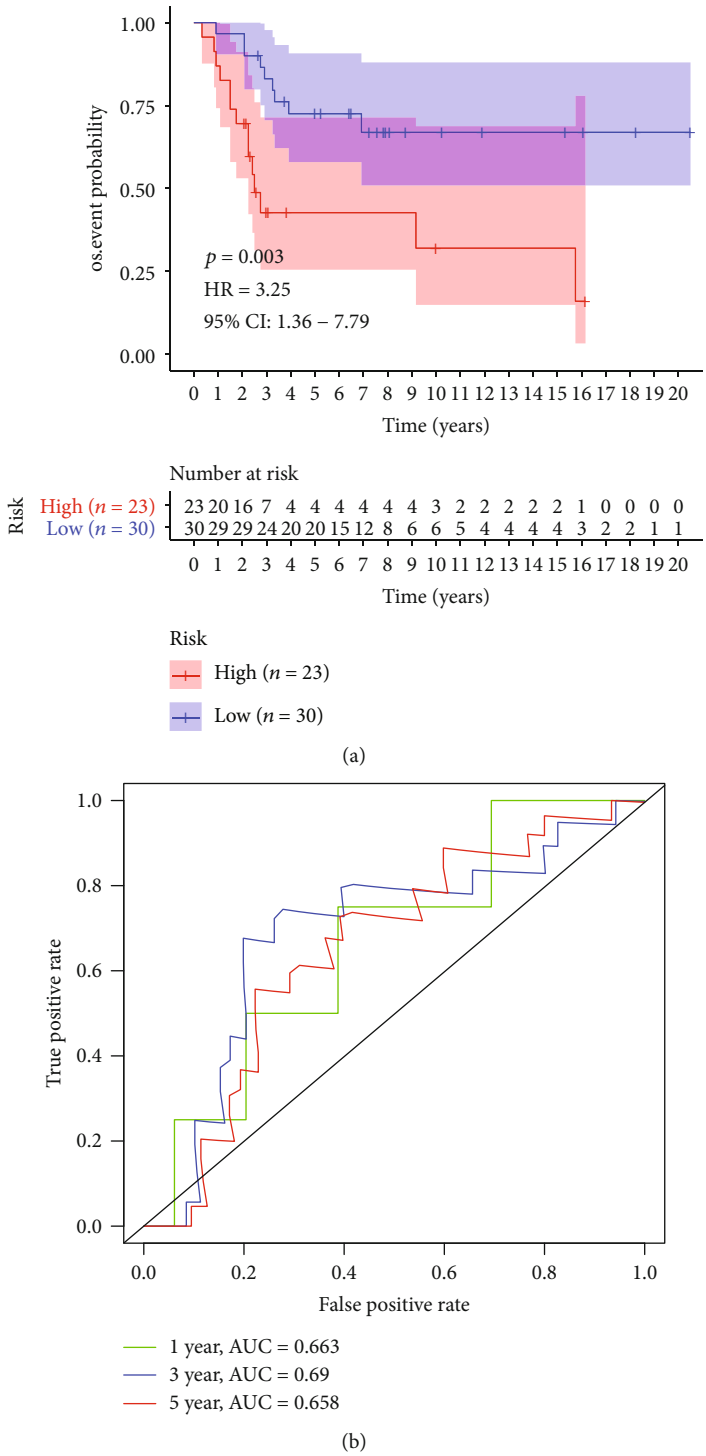
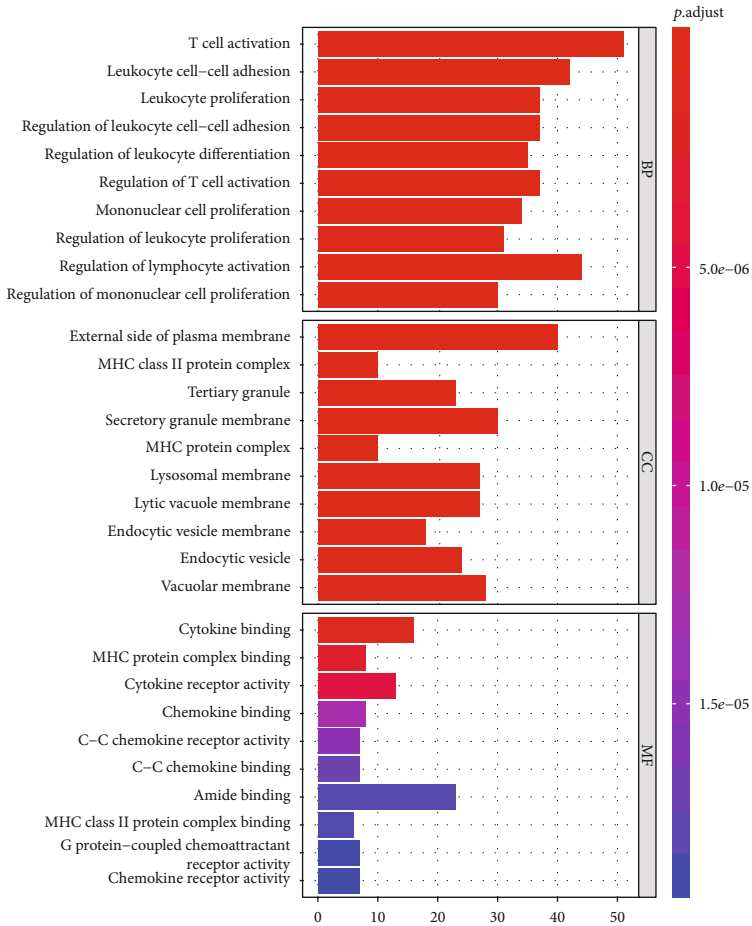


Figure 5: The results of evaluating the accuracy of the EVI2B model. (a) Survival analysis between high- and low-risk groups in GSE21257. (b) The result for the ROC curve of the GSE21257 cohort.

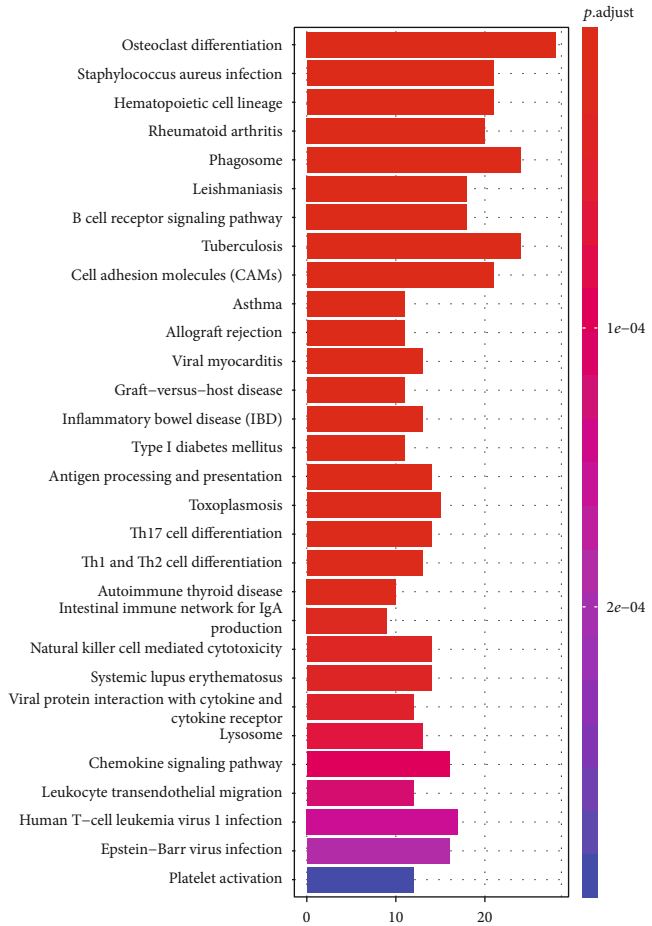
(Version 10.0) [32], interactions with a composite score > 0.4 were considered statistically significant. PPI network results were visualized using Cytoscape (3.4.0); the most important modules were identified by Cytoscape's [33] Plugin Molecular Complex Detection (MCODE) (version 1.4.2).

2.5. Construction of Risk Assessment Model. Then, we performed univariate Cox regression analysis of 28 IRGs followed by screening for IRGs with significant prognostic differences ($p < 0.05$); IRGs were further analyzed by LASSO to reduce genes. Finally, a multivariate Cox prognostic model was built by selecting the IRGs closely related to survival. The sample was divided into two groups based on the median risk score. Besides, we validated the model in the training set. External validation was performed in an independent GEO dataset, GSE21257.



(a)

Figure 6: Continued.



(b)

Figure 6: The results of GO and KEGG enrichment analysis in OS. (a) Gene ontology analysis: biological process, cellular component and molecular function of differentially expressed genes, respectively. (b) KEGG analysis: the top 30 significant pathways of differentially expressed genes.

2.6. Establishment and Validation of the Prognostic Risk Scoring System. To generate a risk scoring system for genes, we performed multivariate Cox proportional hazards regression. First, we used the survival package of R to obtain the regression coefficients for each gene. The coefficient of each selected gene (parameter coefficient R) represents the estimated logarithm of the hazard ratio (HR, parameter exp (coefficient R)) in R. Then, a risk score formula was established for all patients. ROC curve (AUC) predictions over time calculated for 1-, 3-, and 5-year survival using the survival ROC package in R. The optimal cut-off point was selected as the maximum sensitivity and specificity. According to the optimal cut-off point, patients were divided into high- and low-risk groups; the survival difference between the two groups were assessed with the R package survival analysis.

2.7. GO and KEGG Enrichment Analysis. Differential genes between TARGET high- and low-risk groups (R for differential analysis wrapped as limma package, FDR < 0.05, |Log FC| >1) were subjected to enrichment analysis. GO analysis is a standard method to define genes and their RNA or protein products to identify high-throughput transcriptome or genomic data [34]. KEGG is a collection of databases dealing with genomes, diseases, biological pathways, drugs, and chemicals [35]. We used the R package ClusterProfiler [36] to enrich differential genes in the training set's high- and low-risk groups.

2.8. Sample Collection. Osteosarcoma samples were collected from Hunan Cancer Hospital in 2020. OS tissues and adjacent normal tissues from 20 patients were collected, immediately placed in liquid nitrogen, and stored at −80°C. None of the OS patients received antineoplastic therapy. Both the patients and their families in this study were fully informed, and informed consent was obtained from the participants. The Hunan Cancer Hospital ethics committee approved the study.

2.9. Immunohistochemical Staining and Analysis. Immunohistochemistry was performed in 20 OS tissues and adjacent normal tissues. Paraffin-embedded sections were stained to determine the expression level of EVI2B protein. Sections were incubated overnight at 4°C with anti-EVI2B 1 antibody

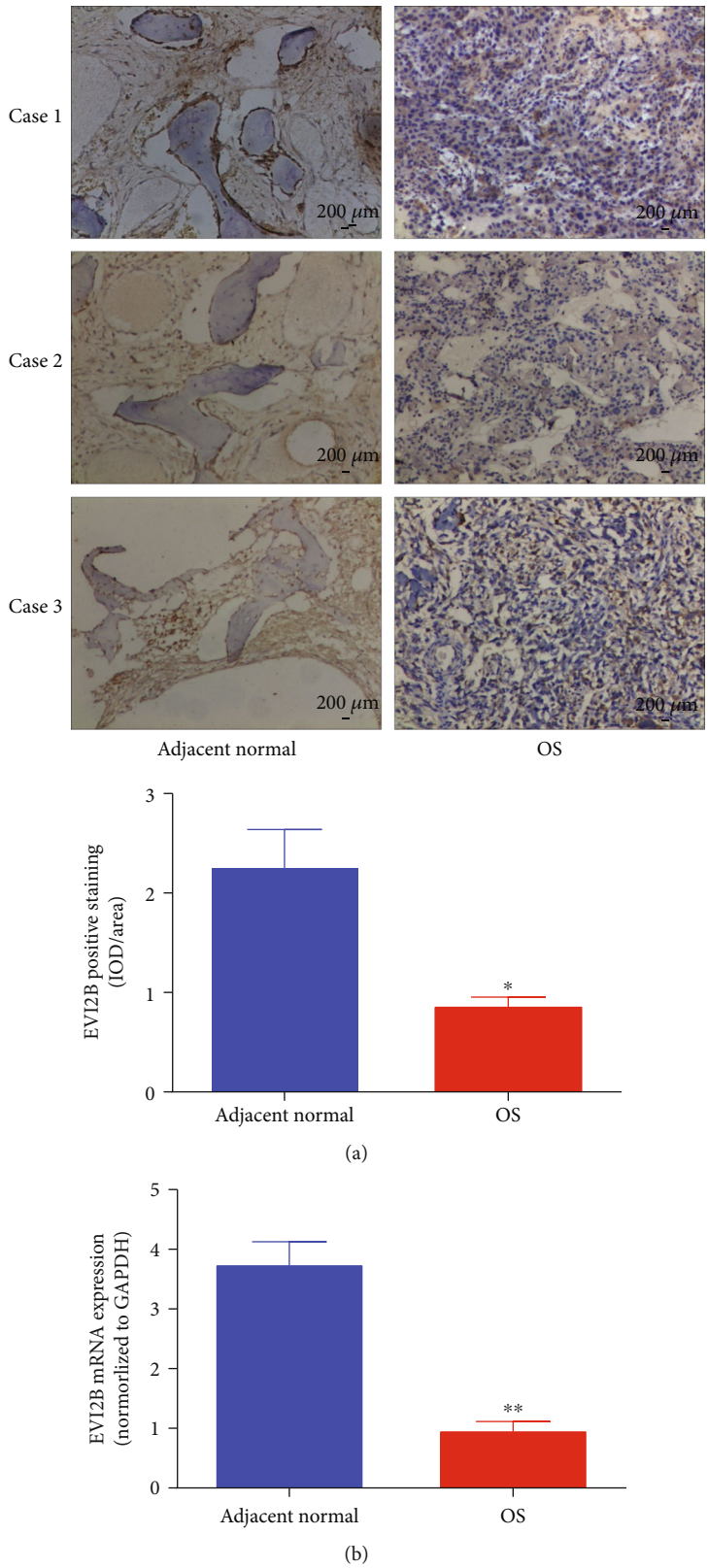


Figure 7: The relative EVI2B expression in OS tissues and adjacent normal tissues. (a) The immunohistochemical detection results showed the expression level of EVI2B in osteosarcoma tissues and adjacent normal tissues. (b) RT-qPCR analysis of relative EVI2B expression in OS tissues and adjacent normal tissues.

(Invitrogen) at a 1 : 100 dilution. After washing with phosphate-buffered saline (PBS), the slides were incubated with a goat anti-mouse IgG secondary antibody conjugated to fluorescein isothiocyanate (ZSDB-BIO, Beijing, China) for 30 mins. They were washed with PBS and then incubated with an antifade reagent (Invitrogen, Carlsbad, USA). Finally, staining was observed using an Olympus CX41 fluorescence microscope (Olympus, Tokyo, Japan). Positive staining intensity in OS and adjacent normal tissues were analyzed by integrated optical density (IOD) using Image-Pro Plus software (version 6.0; Media Cybernetics, United States). All images were taken using the same microscope and camera set. Image-Pro Plus software was used to calculate the mean IOD (μm^2) per stained area (IOD/area) for positive staining. T-test was used to analyze the results. A p value < 0.05 was considered statistically significant.

2.10. Real-Time Quantitative PCR. Total RNA was isolated from tissues from the above patients using Trizol reagent (TaKaRa, Japan) according to the manufacturer's instructions. One microgram of RNA was reverse transcribed into cDNA using the Revert Aid First Strand cDNA Synthesis Kit (Thermo, USA). Quantitative RT-PCR was then performed with Pro Taq HS Premix Probe qPCR Kit (Accurate, Hunan, China). The amplification program consisted of one cycle of predenaturation at 95°C for 5 mins, 37 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30 s, and extension at 72°C for 10 mins. The GAPDH gene was used as an endogenous control gene for normalizing the expression of target genes. Each sample was analyzed in triplicate. The sequences of primers were used for RT-qPCR and annealing temperature (Table S1).

2.11. Statistical Analyses. All statistical analyses were performed using the R software 3.5.0. Statistical significance set at a probability value of $p < 0.05$. Univariate, LASSO and multivariate Cox regression analyses were applied to predict the overall survival of OS patients. The effect of ROC and calibration curve on the prediction accuracy of the prognostic model was compared. Enrichment analysis of differential genes in the high- and low-risk groups in the training set was performed with the R package cluster profile.

3. Results

3.1. Relationship between Immune Score and Stromal Score for Prognosis in OS. The flowchart of this study design is shown in Figure 1. Overall survival analysis did with R package survival analysis of the immune score's prognosis and stromal score. The immune score and stromal score were calculated based on the ESTIMATE algorithm can promote the quantification of immune and stromal components in tumors; in this algorithm, the immune and stromal score is calculated by analyzing the specific gene expression characteristics of the immune and stromal cells to predict the infiltration of nontumor cells. Patients with high immune scores had better overall survival than those with low immune scores (Figure 2(a)), log-rank p value = 0.001. Patients with a high stromal score had a more favorable prognosis than those with a low stromal score (Figure 2(b)), log-rank p value = 0.008. The above results indicated that the immune score and stromal score significantly correlated with OS patients' prognosis. Patients with high immune scores and the stromal score had a good prognosis.

3.2. IRGs of OS Screened with WGCNA and PPI Network Diagram. The sample cluster tree had an outlying sample, and the red line was the cut-off value to filter data (Figure 3(a)). All samples were in clusters after removing one outlying sample based on immune score and stromal score (Figure 3(b)). Sample dendrogram and trait heatmap were drawn according to the immune score and stromal score (Figure 3(b)). The above results were obtained by further analysis of the modules associated with immune score and stromal score by WGCNA. In this study, R^2 and connectivity were highest when the power β set at 15 (Figures 3(c) and 3(d)).

Thus, β identifies distinct gene coexpression modules in OSs. The cluster dendrogram of the selected genes was clustered with an adjacency stromal, and we obtained a module visualization of the genes (Figure 3(e)). We also received the correlation between modules and traits (Figure 3(h)). The pink module has the highest correlation coefficient with an immune score, with a correlation coefficient of 0.58, $p < 0.05$, which was statistically significant, indicating that the immune score significantly correlated with clinical traits—the above results were based on the WGCNA analysis performed by the immune system stromal scoring. Therefore, we chose this module as the next research object. We screened the pink module's IRGs with a correlation between genes and the pink module greater than 0.8 and a correlation coefficient between genes and immune scores greater than 0.5 (Figure 3(f)). With conditions, we screened a total of 28 IRGs PTPRJ, SCARF1, CD300C, ANPEP, TBXAS1, RNASE6, HSD3B7, EVI2B, NCKAP1L, PTK2B, SPI1, GRN, SQOR, GLRX, APOBR, GNA15, LILRB4, LAPTM5, CCR1, FERMT3, CYTH4, LCP1, CD4, CSF1R, PSTPIP1, LAT2, SELPLG, and RAC2 shown in Table 1.

A PPI regulatory network of 28 IRGs was obtained based on the STRING online database (https://string-db.org/). We found that the direct interaction between hub genes was through the PPI network analysis (Figure 3(g)). Nineteen genes (PTPRJ, CD300C, RNASE6, EVI2B, NCKAP1L, PTK2B, SPI1, GNA15, LILRB4, LAPTM5, CCR1, FERMT3, CYTH4, LCP1, CD4, CSF1R, PSTPIP1, SELPLG, and RAC2) were correlated well with immune score.

3.3. Recognize IRG Prognostic Gene and Survival Analysis and ROC Analysis. The 28 IRGs were subjected to univariate Cox analysis, resulting in 10 IRGs with prognostic significance (Table 2) (p value < 0.05). These ten genes, EVI2B, GRN, NCKAP1L, SELPLG, LILRB4, FERMT3, SQOR, CYTH4, GNA15, and RNASE6, were subjected to Lasso regression analysis further to screen prognostic genes (Figures 4(a) and 4(b)) after 1,000 stimuli run through the crossvalidation possibility, the optimal lambda determined, and three genes, EVI2B, GRN, and NCKAP1L, were screened. Next, multivariate Cox analysis was performed to construct a prognostic

model based on EVI2B, GRN, and NCKAP1L, resulting in a model containing one IRG EVI2B. EVI2B is highly expressed in normal bone tissue controls and lowly expressed in osteosarcoma (Table 2), which can be regarded as a tumor suppressor gene consistent with the results of a colorectal cancer study by Yuan et al. [37].

By dividing the sample into high- and low-risk groups according to the model's median risk score, we found a significant difference in survival between the high- and low-risk groups (log-rank p value = 0.016) (Figure 4(c)). ROC analysis showed AUC values of 0.676, 0.697, and 0.708 for ROC curves at 1, 3, and 5 years, respectively (Figure 4(d)). All results showed that the prediction effect of the IRG EVI2B model was moderately accurate.

3.4. Evaluating the Accuracy of the IRG EVI2B Model. The time-dependent ROC curve analysis evaluated the accuracy of the OS prediction model constructed by EVI2B. The IRG EVI2B model evaluates in an independent GEO dataset GSE21257. By dividing GSE21257 samples into high- and low-risk groups according to the TARGET dataset samples' median risk score in the model, we found a significant difference in survival between the high- and low-risk groups (log-rank p value = 0.003) (Figure 5(a)). ROC analysis showed that the AUC values for 1, 3, and 5 years of the ROC curve were 0.663, 0.690, and 0.658, respectively (Figure 5(b)). The survival rate of the low-risk group of the GSE21257 sample in the validation set based on the IRG EVI2B model was higher than that of the high-risk group, and the prolonged survival time was consistent with the results in the training set. The AUC values for the validation set GSE21257 samples were like the AUC results in the training set. The result fully validates the IRG EVI2B prognostic model's accuracy and illustrates that the IRG EVI2B is feasible as a prognostic indicator in OS patients.

3.5. GO and KEGG Enrichment Analysis. Next, to further elucidate the molecular functions and signaling pathways in which differential genes between the TARGET high- and low-risk groups are involved, we performed functional enrichment analysis of all DEGs. GO results: these genes are involved not only in T cell activation, regulation of leukocyte differentiation, regulation of lymphocyte activation, and several other biological processes but also in cytokine binding, cytokine receptor complex activity, MHC protein binding, and molecular functions in other tumors. Some of the encoded proteins are important components of the plasma membrane's external side, secretory granule membrane, and vacuolar membrane (Figure 6(a)). We further analyzed these genes' signaling pathways, and KEGG results showed that these genes play a regulatory role in important immune-related pathways (such as the B cell receptor signaling pathway and chemokine signaling pathway) (Figure 6(b)). These results suggest that the OS prognostic model is significantly associated with immune response and inflammatory response, affecting cancer patients' prognosis by participating in immune processes, indicating the correctness of our analysis results.

3.6. IRG EVI2B Expression Is Downregulated in OS Tissues. Finally, immunostaining analysis of IRG EVI2B showed that EVI2B expression was higher in adjacent normal tissues (Figure 7(a)). EVI2B-positive staining was significantly different between the control and OS, and positive staining was low and statistically significant in OS (Figure 7(a)). We examined IRG EVI2B expression in clinical samples from OS patients by qPCR. The qPCR results showed that IRG EVI2B was highly expressed at adjacent normal bone tissue levels (Figure 7(b)). The above results showed that IRG EVI2B was highly expressed in adjacent normal bone tissue (control) and lowly expressed in osteosarcoma, and Yuan et al. included colorectal cancer study [37].

4. Discussion

In the past decades, traditional cancer treatments, including surgical treatment and chemotherapy/radiotherapy, have incredibly prolonged osteosarcoma patients' survival time [38, 39]. Although the 5-year overall survival rate is high [9, 40, 41], the survival rate of patients with metastatic disease (most commonly in the lung parenchyma and distal bone) is meager, 19-30% [39, 42]. Also, chemotherapy/radiotherapy acquired resistance to new antitumor drugs and severe side effects of OS treatment's drug resistance characteristics after extensive surgical resection of OS treatment reach the platform [39, 43].

The use of targeted therapy [44–46] also benefits OS patients; for example, therapies targeting the unfolded protein response (UPR) pathway may be a feasible approach for treating osteosarcoma. The components targeting the UPR may prove beneficial for patients refractory to conventional chemotherapy [47]. In addition to being a marker of metastatic disease, therapeutic targeting of cathepsin D (CTSD) may be a promising novel approach for osteosarcoma treatment; it may produce a good response in metastatic disease [47]. However, targeted therapy also has some limitations and cannot be suitable for all types of osteosarcoma. In recent years, immunotherapy [48] has dramatically developed in cancer. Immune checkpoint-based therapy has been shown to encouraging the antitumor effect by restoring immune response in the tumor microenvironment [49, 50]. At the basement of immune checkpoint inhibitors, ipilimumab (monoclonal antibody) anticytotoxic T lymphocyte antigen four antibodies (mAb) (CTLA4) and monoclonal antibody death protein 1 (PD1) or PD1 ligand (PDL1) against programmed cells show that immune checkpoints may be immune-tolerant to the tumor [13–17, 51]. A study has shown that for osteosarcoma, the immune checkpoint inhibitor PD-L1 is negatively correlated with prognosis, and PD-1 is negatively correlated with overall survival [13–17, 51]. To date, although immunotherapy [10, 11, 13, 51–53] has used in OS, the impact on OS prognosis is not well understood, so there is an urgent need for a new biomarker that can use to predict the prognosis of OS patients, and our study is to affect the prognosis of cancer patients by participating in the immune process.

At present, the use of EVI2B [54] is still relatively rare. The EVI2B is located in the intron of the neurofibromatosis type 1 (NF1) gene and transcribed in the opposite direction

to the NF1 gene [55–58]. Still, their expression is not related, indicating that they are independently regulated [59]. EVI2B is a transmembrane protein [60], while NF1 is a tumor suppressor gene [61]. EVI2B is expressed in many different cell types, including myeloid cells [62]. The EVI2B is also located within EVI2, a common viral integration site found in retroviral induced myeloid tumors [54]. It has postulated that viral integration on EVI2 alters EVI2B expression and that this altered expression predisposes mice to the development of bone marrow tumors [54].

The successful construction of the IRG EVI2B prognostic model and the evaluation and validation of the accuracy of the IRG EVI2B prognostic model with the ROC curve method and in the GEO dataset GSE21257 give us hope to use the IRG EVI2B gene as a new biomarker and therapeutic target for predicting the prognosis of osteosarcoma patients. In the same study, EVI2B was found to correlate with the prognosis of patients significantly negatively with colon cancer ($p < 0.05$), and EVI2B may be involved in the prognosis of colon cancer patients [37, 63]. EVI2B can similarly be used as a prognosis for osteosarcoma compared to the study by Yang et al. [64].

EVI2B is expressed in various tumors [62] and breast cancer mutations [65]. EVI2B is involved in the differentiation of melanocytes and keratinocytes [66, 67]. In fibroblast-like cells derived from neurofibromas, increased EVI2B mRNA levels were found [66]. EVI2B gene, their related studies indicate that the EVI2B gene is a direct target of CCAAT/enhancer-binding protein (C/EBP). The product of this gene, the transmembrane glycoprotein EVI2B (CD361) [68], shows to be abundantly expressed on the surface of primary hematopoietic cells, reaching the highest expression level in mature granulocytes [69]. They use shRNA-mediated downregulation of EVI2B in human and murine cell lines and primary hematopoietic stem and progenitor cells (HSPC), demonstrating impaired myeloid lineage development and altered progenitor function in EVI2B-depleted cells. Its study suggests that EVI2B is an essential modulator of HSPC and bone marrow differentiation and that low levels of EVI2B may contribute to the differentiation inhibitory profile of acute myelocytic leukemia (AML) [69]. It has shown that EVI2B appears highly expressed in patients with 11q deletion in chronic lymphocytic leukemia (CLL). It speculated that the EVI2B gene might be associated with poor prognosis in patients with CLL 11q deletion [70]. It also studied that EVI2B was transiently expressed on cellular adjuvants and used to directly immunize BALB/c mice to obtain a higher specific anti-EVI2B Ab response in immunized mice than stimulated by other cellular adjuvants, which used to accelerate the development of Ab drugs [71]. Our study also showed that EVI2B is associated with the prognosis of osteosarcoma. These results suggest that it is feasible for us to use EVI2B as a marker and potential target for the diagnosis and treatment of osteosarcoma, but further studies should be on its possible action mechanism.

5. Conclusions

In conclusion, we selected IRG EVI2B by WGCNA analysis and constructed an IRG-based prognostic model, which was validated and experimentally analyzed to obtain that IRG EVI2B can significantly predict the prognosis of osteosarcoma patients.

Data Availability

The publicly available datasets were analyzed in this study; these can be found in the TARGET and the NCBI Gene Expression Omnibus (GSE21257).

Ethical Approval

The Hunan Cancer Hospital ethics committee approved 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

Li Gan and Xin Wang contributed equally and are co-first authors to this work.

Acknowledgments

This study was supported by the Scientific Research Project of Hunan Health Commission (Nos. 202102041763 and 20200985), Changsha Municipal Natural Science Foundation (No. kq2014267), and Hunan Cancer Hospital Climb Plan (No. 2020QH001).

Supplementary Materials

Table S1: the sequences of primers used for RT-qPCR and annealing temperature. *(Supplementary Materials)*

References

[1] S. J. Cotterill, C. M. Wright, M. S. Pearce, and A. W. Craft, "Stature of young people with malignant bone tumors," *Pediatric Blood & Cancer*, vol. 42, no. 1, pp. 59–63, 2004.

[2] M. S. Isakoff, S. S. Bielack, P. Meltzer, and R. Gorlick, "Osteosarcoma: current treatment and a collaborative pathway to success," *Journal of Clinical Oncology*, vol. 33, no. 27, pp. 3029–3035, 2015.

[3] A. Misaghi, A. Goldin, M. Awad, and A. A. Kulidjian, "Osteosarcoma: a comprehensive review," *SICOT J*, vol. 4, p. 12, 2018.

[4] S. S. Bielack, S. Hecker-Nolting, C. Blattmann, and L. Kager, "Advances in the management of osteosarcoma," *F1000Research*, vol. 5, 2016.

[5] F. Liu, L. Xing, X. Zhang, and X. Zhang, "A four-pseudogene classifier identified by machine learning serves as a novel prognostic marker for survival of osteosarcoma," *Genes (Basel)*, vol. 10, no. 6, p. 414, 2019.

[6] N. M. Bernthal, N. Federman, F. R. Eilber et al., "Long-term results (>25 years) of a randomized, prospective clinical trial evaluating chemotherapy in patients with high-grade, operable osteosarcoma," *Cancer*, vol. 118, no. 23, pp. 5888–5893, 2012.

[7] C. C. Wu, H. C. Beird, J. Andrew Livingston et al., "Immunogenomic landscape of osteosarcoma," *Nature Communications*, vol. 11, no. 1, 2020.

[8] A. Luetke, P. A. Meyers, I. Lewis, and H. Juergens, "Osteosarcoma treatment - where do we stand? A state of the art review," *Cancer Treatment Reviews*, vol. 40, no. 4, pp. 523–532, 2014.

[9] Y. Wang, W. Yu, J. Zhu et al., "Anti-CD166/4-1BB chimeric antigen receptor T cell therapy for the treatment of osteosarcoma," *Journal of Experimental & Clinical Cancer Research*, vol. 38, no. 1, 2019.

[10] N. Ahmed, V. S. Brawley, M. Hegde et al., "Human epidermal growth factor receptor 2 (HER2) -specific chimeric antigen receptor-modified T cells for the immunotherapy of HER2-positive sarcoma," *Journal of Clinical Oncology*, vol. 33, no. 15, pp. 1688–1696, 2015.

[11] N. J. Mason, J. S. Gnanandarajah, J. B. Engiles et al., "Immunotherapy with a HER2-targeting listeria induces HER2-specific immunity and demonstrates potential therapeutic effects in a phase I trial in canine osteosarcoma," *Clinical Cancer Research*, vol. 22, no. 17, pp. 4380–4390, 2016.

[12] N. Rainusso, V. S. Brawley, A. Ghazi et al., "Immunotherapy targeting HER2 with genetically modified T cells eliminates tumor-initiating cells in osteosarcoma," *Cancer Gene Therapy*, vol. 19, no. 3, pp. 212–217, 2012.

[13] M. Kansara, M. W. Teng, M. J. Smyth, and D. M. Thomas, "Translational biology of osteosarcoma," *Nature Reviews. Cancer*, vol. 14, no. 11, pp. 722–735, 2014.

[14] P. W. Kantoff, C. S. Higano, N. D. Shore et al., "Sipuleucel-T immunotherapy for castration-resistant prostate cancer," *The New England Journal of Medicine*, vol. 363, no. 5, pp. 411–422, 2010.

[15] J. R. Brahmer, S. S. Tykodi, L. Q. Chow et al., "Safety and activity of anti-PD-L1 antibody in patients with advanced cancer," *The New England Journal of Medicine*, vol. 366, no. 26, pp. 2455–2465, 2012.

[16] A. D. Garg and P. Agostinis, "Cell death and immunity in cancer: from danger signals to mimicry of pathogen defense responses," *Immunological Reviews*, vol. 280, no. 1, pp. 126–148, 2017.

[17] G. Kroemer, L. Galluzzi, O. Kepp, and L. Zitvogel, "Immunogenic cell death in cancer therapy," *Annual Review of Immunology*, vol. 31, no. 1, pp. 51–72, 2013.

[18] C. Liu, X. Wang, G. Z. Genchev, and H. Lu, "Multi-omics facilitated variable selection in Cox-regression model for cancer prognosis prediction," *Methods*, vol. 124, pp. 100–107, 2017.

[19] Y. Zhang, H. Li, W. Zhang, Y. Che, W. Bai, and G. Huang, "LASSO-based Cox-PH model identifies an 11-lncRNA signature for prognosis prediction in gastric cancer," *Molecular Medicine Reports*, vol. 18, no. 6, pp. 5579–5593, 2018.

[20] M. Bagnoli, S. Canevari, D. Califano et al., "Development and validation of a microRNA-based signature (MiROvaR) to predict early relapse or progression of epithelial ovarian cancer: a cohort study," *Lancet Oncology*, vol. 17, no. 8, pp. 1137–1146, 2016.

[21] Y. Xiong, L. Yuan, J. Xiong et al., "An outcome model for human bladder cancer: a comprehensive study based on weighted gene co-expression network analysis," *Journal of Cellular and Molecular Medicine*, vol. 24, no. 3, pp. 2342–2355, 2020.

[22] R. Tibshirani, "The lasso method for variable selection in the Cox model," *Statistics in Medicine*, vol. 16, no. 4, pp. 385–395, 1997.

[23] N. Ternes, F. Rotolo, and S. Michiels, "Empirical extensions of the lasso penalty to reduce the false discovery rate in high-dimensional Cox regression models," *Statistics in Medicine*, vol. 35, no. 15, pp. 2561–2573, 2016.

[24] P. F. Chen, F. Wang, J. Y. Nie et al., "Co-expression network analysis identified CDH11 in association with progression and prognosis in gastric cancer," *OncoTargets and Therapy*, vol. Volume 11, pp. 6425–6436, 2018.

[25] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, no. 1, 2008.

[26] F. Wang, Y. Chang, J. Li et al., "Strong correlation between _ASPM_ gene expression and HCV cirrhosis progression identified by co-expression analysis," *Digestive and Liver Disease*, vol. 49, no. 1, pp. 70–76, 2017.

[27] Z. Zhou, S. Liu, M. Zhang et al., "Overexpression of topoisomerase 2-alpha confers a poor prognosis in pancreatic adenocarcinoma identified by co-expression analysis," *Digestive Diseases and Sciences*, vol. 62, no. 10, pp. 2790–2800, 2017.

[28] 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, 2019.

[29] M. J. Mason, G. Fan, K. Plath, Q. Zhou, and S. Horvath, "Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells," *BMC Genomics*, vol. 10, no. 1, p. 327, 2009.

[30] X. Zhai, Q. Xue, Q. Liu, Y. Guo, and Z. Chen, "Colon cancer recurrence associated genes revealed by WGCNA coexpression network analysis," *Molecular Medicine Reports*, vol. 16, no. 5, pp. 6499–6505, 2017.

[31] Z. Liu, M. Li, Q. Hua, Y. Li, and G. Wang, "Identification of an eight-lncRNA prognostic model for breast cancer using WGCNA network analysis and a Cox-proportional hazards model based on L1-penalized estimation," *International Journal of Molecular Medicine*, vol. 44, no. 4, pp. 1333–1343, 2019.

[32] 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.

[33] 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.

[34] M. Ashburner, C. A. Ball, J. A. Blake et al., "Gene Ontology: tool for the unification of biology," *Nature Genetics*, vol. 25, no. 1, pp. 25–29, 2000.

[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] 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.

[37] Y. Yuan, J. Chen, J. Wang et al., "Identification hub genes in colorectal cancer by integrating weighted gene co-expression network analysis and clinical validation in vivo and vitro," *Frontiers in Oncology*, vol. 10, 2020.

[38] P. A. Meyers, C. L. Schwartz, M. Krailo et al., "Osteosarcoma: a randomized, prospective trial of the addition of ifosfamide and/or muramyl tripeptide to cisplatin, doxorubicin, and high-dose methotrexate," *Journal of Clinical Oncology*, vol. 23, no. 9, pp. 2004–2011, 2005.

[39] M. W. Bishop, K. A. Janeway, and R. Gorlick, "Future directions in the treatment of osteosarcoma," *Current Opinion in Pediatrics*, vol. 28, no. 1, pp. 26–33, 2016.

[40] N. Ahmed, V. S. Salsman, E. Yvon et al., "Immunotherapy for osteosarcoma: genetic modification of T cells overcomes low levels of tumor antigen expression," *Molecular Therapy*, vol. 17, no. 10, pp. 1779–1787, 2009.

[41] D. C. Allison, S. C. Carney, E. R. Ahlmann et al., "A meta-analysis of osteosarcoma outcomes in the modern medical era," *Sarcoma*, vol. 2012, Article ID 704872, 10 pages, 2012.

[42] F. Fagioli, E. Biasin, O. M. Mereuta et al., "Poor prognosis osteosarcoma: new therapeutic approach," *Bone Marrow Transplantation*, vol. 41, Supplement 2, pp. S131–S134, 2008.

[43] J. Luo, Y. Xia, Y. Yin et al., "ATF4 destabilizes RET through nonclassical GRP78 inhibition to enhance chemosensitivity to bortezomib in human osteosarcoma," *Theranostics.*, vol. 9, no. 21, pp. 6334–6353, 2019.

[44] P. Angulo, G. Kaushik, D. Subramaniam et al., "Natural compounds targeting major cell signaling pathways: a novel paradigm for osteosarcoma therapy," *Journal of Hematology & Oncology*, vol. 10, no. 1, p. 10, 2017.

[45] L. C. Sayles, M. R. Breese, A. L. Koehne et al., "Genome-informed targeted therapy for osteosarcoma," *Cancer Discovery*, vol. 9, no. 1, pp. 46–63, 2019.

[46] A. B. Shaikh, F. Li, M. Li et al., "Present advances and future perspectives of molecular targeted therapy for osteosarcoma," *International Journal of Molecular Sciences*, vol. 17, no. 4, p. 506, 2016.

[47] J. Burns, C. P. Wilding, R. L. Jones, and P. H. Huang, "Proteomic research in sarcomas - current status and future opportunities," *Seminars in Cancer Biology*, vol. 61, pp. 56–70, 2020.

[48] M. F. Wedekind, L. M. Wagner, and T. P. Cripe, "Immunotherapy for osteosarcoma: where do we go from here?," *Pediatric Blood & Cancer*, vol. 65, no. 9, p. e27227, 2018.

[49] K. M. Mahoney, P. D. Rennert, and G. J. Freeman, "Combination cancer immunotherapy and new immunomodulatory targets," *Nature Reviews. Drug Discovery*, vol. 14, no. 8, pp. 561–584, 2015.

[50] W. Wu, D. Jing, Z. Meng et al., "FGD1 promotes tumor progression and regulates tumor immune response in osteosarcoma via inhibiting PTEN activity," *Theranostics.*, vol. 10, no. 6, pp. 2859–2871, 2020.

[51] B. Zheng, T. Ren, Y. Huang et al., "PD-1 axis expression in musculoskeletal tumors and antitumor effect of nivolumab in osteosarcoma model of humanized mouse," *Journal of Hematology & Oncology*, vol. 11, no. 1, 2018.

[52] M. F. Heymann, K. Schiavone, and D. Heymann, "Bone sarcomas in the immunotherapy era," *British Journal of Pharmacology*, vol. 178, no. 9, 2021.

[53] M. Burgess and H. Tawbi, "Immunotherapeutic approaches to sarcoma," *Current Treatment Options in Oncology*, vol. 16, no. 6, 2015.

[54] A. M. Buchberg, H. G. Bedigian, N. A. Jenkins, and N. G. Copeland, "Evi-2, a common integration site involved in murine myeloid leukemogenesis," *Molecular and Cellular Biology*, vol. 10, no. 9, pp. 4658–4666, 1990.

[55] D. E. Jenne, S. Tinschert, H. Reimann et al., "Molecular characterization and gene content of breakpoint boundaries in patients with neurofibromatosis type 1 with 17q11.2 microdeletions," *The American Journal of Human Genetics*, vol. 69, no. 3, pp. 516–527, 2001.

[56] L. M. Kayes, W. Burke, V. M. Riccardi et al., "Deletions spanning the neurofibromatosis 1 gene: identification and phenotype of five patients," *American Journal of Human Genetics*, vol. 54, no. 3, pp. 424–436, 1994.

[57] H. Kehrer-Sawatzki, C. Maier, E. Moschgath, G. Elgar, and W. Krone, "Genomic characterization of the neurofibromatosis type 1 gene of _Fugu rubripes_," *Gene*, vol. 222, no. 1, pp. 145–153, 1998.

[58] D. Viskochil, R. Cawthon, P. O'Connell et al., "The gene encoding the oligodendrocyte-myelin glycoprotein is embedded within the neurofibromatosis type 1 gene," *Molecular and Cellular Biology*, vol. 11, no. 2, pp. 906–912, 1991.

[59] G. Serra, V. Antona, G. Corsello, F. Zara, E. Piro, and R. Falsaperla, "NF1 microdeletion syndrome: case report of two new patients," *Italian Journal of Pediatrics*, vol. 45, no. 1, p. 138, 2019.

[60] R. M. Cawthon, L. B. Andersen, A. M. Buchberg et al., "cDNA sequence and genomic structure of _EVI2B_ , a gene lying within an intron of the neurofibromatosis type 1 gene," *Genomics*, vol. 9, no. 3, pp. 446–460, 1991.

[61] A. Pemov, NISC Comparative Sequencing Program, H. Li et al., "The primacy of _NF1_ loss as the driver of tumorigenesis in neurofibromatosis type 1-associated plexiform neurofibromas," *Oncogene*, vol. 36, no. 22, pp. 3168–3177, 2017.

[62] B. Andreopoulos and D. Anastassiou, "Integrated analysis reveals hsa-miR-142 as a representative of a lymphocyte-specific gene expression and methylation signature," *Cancer Information*, vol. 11, pp. 61–75, 2012.

[63] M. Y. Huang, H. M. Wang, T. S. Tok et al., "EVI2B, ATP2A2, S100B, TM4SF3, and OLFM4 as potential prognostic markers for postoperative Taiwanese colorectal cancer patients," *DNA and Cell Biology*, vol. 31, no. 4, pp. 625–635, 2012.

[64] B. Yang, Z. Su, G. Chen et al., "Identification of prognostic biomarkers associated with metastasis and immune infiltration in osteosarcoma," *Oncology Letters*, vol. 21, no. 3, p. 180, 2021.

[65] E. Y. Leung, M. E. Askarian-Amiri, D. C. Singleton et al., "Derivation of breast cancer cell lines under physiological (5%) oxygen concentrations," *Frontiers in Oncology*, vol. 8, 2018.

[66] D. Kaufmann, S. Gruener, F. Braun et al., "EVI2B, a gene lying in an intron of the neurofibromatosis type 1 (NF1) gene, is as the NF1 gene involved in differentiation of melanocytes and keratinocytes and is overexpressed in cells derived from NF1 neurofibromas," *DNA and Cell Biology*, vol. 18, no. 5, pp. 345–356, 1999.

[67] J. Voisey, G. Kelly, and A. Van Daal, "Agouti signal protein regulation in human melanoma cells," *Pigment Cell Research*, vol. 16, no. 1, pp. 65–71, 2003.

[68] J. Matesanz-Isabel, J. Sintes, L. Llinas, J. de Salort, A. Lazaro, and P. Engel, "New B-cell CD molecules," *Immunology Letters*, vol. 134, no. 2, pp. 104–112, 2011.

[69] P. Zjablovskaja, M. Kardosova, P. Danek et al., "Correction to: EVI2B is a C/EBP _α_ target gene required for granulocytic differentiation and functionality of hematopoietic progenitors," *Cell Death & Differentiation*, vol. 26, no. 1, p. 198, 2019.

[70] Y. Aalto, W. El-Rifa, L. Vilpo et al., "Distinct gene expression profiling in chronic lymphocytic leukemia with 11q23 deletion," *Leukemia*, vol. 15, no. 11, pp. 1721–1728, 2001.

[71] C. C. Huang, K. W. Cheng, Y. C. Hsieh et al., "Use of syngeneic cells expressing membrane-bound GM-CSF as an adjuvant to induce antibodies against native multi-pass transmembrane protein," *Scientific Reports*, vol. 9, no. 1, p. 9931, 2019.

Hindawi
Journal of Immunology Research
Volume 2021, Article ID 6661625, 23 pages
https://doi.org/10.1155/2021/6661625

Research Article

Identification of Immune-Related Prognostic Biomarkers Associated with HPV-Positive Head and Neck Squamous Cell Carcinoma

Yifei Chen, Jin Nie, Xiangsheng Li, Tao Fan, Xiaowen Deng, Dan Liang, and Guilin Song

Department of Otolaryngology Head and Neck Surgery, Affiliated Changsha Hospital of Hunan Normal University, 70 Lushan Road, Changsha, Hunan 410013, China

Correspondence should be addressed to Guilin Song; songguilin@hunnu.edu.cn

Received 29 November 2020; Revised 20 December 2020; Accepted 24 December 2020; Published 8 January 2021

Academic Editor: Jialiang Liang

Copyright © 2021 Yifei Chen 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. As a type of malignant tumor, head and neck squamous cell carcinoma (HNSCC) seriously threatens human health. This study is aimed at constructing a new, reliable prognostic model. *Method*. The gene expression profile data of HNSCC patients were downloaded from the Gene Expression Omnibus and The Cancer Genome Atlas databases. The immune-related differentially expressed genes (IRDEGs) related to HNSCC were identified. We then used Cox regression analysis and least absolute shrinkage and selection operator (LASSO) analysis to explore IRDEGs related to the HNSCC prognosis and to construct and validate a risk scoring model and used ESTIMATE to evaluate tumor immune infiltration in HNSCC patients. Finally, we validated IGSF5 expression and function in HNSCC cells. *Results*. A total of 1,195 IRDEGs were found from the GSE65858 dataset. Thirty-one of the 1,195 IRDEGs were associated with the prognosis of HNSCC. Nine key IRDEGs were further selected using the LASSO method, and a risk scoring model was established for predicting the survival of HNSCC patients. According to the risk scoring model, the prognosis of patients in the high-risk group was worse than that of the low-risk group; the high-risk group had significantly higher immune scores than the low-risk group; and between the high- and low-risk samples, there were significant differences in the proportion of 10 types of cells, including naive cells, plasma cells, and resting $CD4^+$ memory T cells. IGSF5 has low expression in HNSCC, and overexpression of IGSF5 significantly impaired HNSCC cell proliferation. *Conclusion*. This prognostic risk assessment model can help systematically evaluate the survival prognosis of HNSCC patients and provides a new research direction for the improvement of the survival prognosis of HNSCC patients in clinical practice.

1. Introduction

Head and neck squamous cell carcinoma (HNSCC) is an epithelial tumor that develops in the mucous membranes of the mouth, oropharynx, larynx, or hypopharynx [1]. According to the "Global Cancer Statistics 2018," it is estimated that there are more than 800,000 new HNSCC cases each year [2]. HNSCC is associated with several risk factors, including tobacco use, alcohol consumption, and betel nut chewing (*Areca catechu*) [3]. Despite the reduction in HNSCC cases caused by these risk factors, the number of HNSCC cases caused by human papillomavirus (HPV) is increasing, and the overall prevalence is gradually rising [4]. Recently, convincing results have clearly demonstrated that HNSCC patients with HPV positivity and HPV negativity are completely different in terms of molecular markers, clinical manifestations, and response to treatment. As the most common subtype of infection, HPV16 infection accounted for approximately 80% of the HPV infections in HNSCC patients [5]. Moreover, HNSCC patients who did not smoke or drink alcohol but had HPV infection were already in the middle and late stages when they were diagnosed. Approximately 40-60% of these patients experience relapse after comprehensive treatment, including surgery, radiotherapy,

and chemotherapy, and they do not respond to subsequent treatment measures [6]. Therefore, it is crucial to evaluate the prognosis of HNSCC patients with HPV infection. Currently, the clinical evaluation of the prognosis of HNSCC patients is mainly done through traditional tumor, node, metastases (TNM) staging. The accuracy of this prediction method is limited, and the prognostic stratification of patients is not accurate. Therefore, to better guide the clinical improvement of the prognosis of HNSCC patients with HPV infection, it is imperative to explore a stable and efficient prognostic model.

Immune escape is an important factor affecting the occurrence and development of malignant tumors, including HNSCC [7]. Immune-related biomarkers can define not only the immune status of patients but also the biological behavior of HNSCC. Many studies have provided evidence for the role of immune cells in the development of HPV-induced HNSCC. For example, $CD4^+$ T cell infiltration in HPV^+ HNSCC patients was higher than that of HPV^- HNSCC patients [8]. The levels of cytokines and chemokines in HPV^+ HNSCC patients were higher than those of HPV^- patients [9]. Some studies have already explored a model for assessing the survival prognosis of HNSCC using immune-related genes [10]. These studies have had some limitations, such as the failure to perform their analyses in terms of the immune and stromal cell components in the tumor microenvironment, and their descriptions of the molecular characteristics of the tumor-immune interaction have been unclear, especially regarding the prognostic evaluation of HNSCC. Many components of the immune system are closely related to the occurrence and development of HNSCC [8]. Immune scores and stromal scores are emerging methods that can be used to characterize the tumor microenvironment to determine the invasion and infiltration capabilities of this tumor and can play a role in prognostic assessment [11].

Based on the gene expression profile data of HNSCC patients with HPV infection in the Gene Expression Omnibus (GEO) database and incorporating the known immune-related genes (IRGs) in the InnateDB database and the ImmPort database, we selected novel immune biomarkers related to HPV infection using bioinformatic methods and established a prognostic evaluation model. This model can be used to assess survival prognosis and immune infiltration in HNSCC patients. This study will help improve the clinical diagnostic and prognostic evaluation of HNSCC patients and provide new targets for the treatment of HNSCC patients.

2. Materials and Methods

2.1. Data Possession. The gene expression profile data and clinical information of HNSCC patients in the GSE65858 dataset were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). There were 60 cases of HPV16 infection and 196 cases of HPV negativity. On September 12, 2019, the gene expression profile data and survival information of patients (494 cases) with HNSCC in The Cancer Genome Atlas (TCGA) database were obtained from the University of California, Santa Cruz, Xena website (https://xenabrowser.net/datapages/).

2.2. Selection of Immune-Related Differentially Expressed Genes (DEGs). We used the package limma of R software (Version 3.44.3) to analyze the gene expression profile data of HNSCC patients in the GSE65858 dataset and screened the results according to adjusted p value ($p_{\text{adj.}}$) < 0.05 to select DEGs in HPV16-infected samples vs. HPV-uninfected samples. On September 12, 2019, IRGs were downloaded from the InnateDB database (https://www.innatedb.ca/) and the ImmPort database (https://www.immport.org). The intersection of DEGs and IRGs made up the list of immune-related differentially expressed genes (IRDEGs) for HNSCC patients.

2.3. IRDEG Enrichment Analysis. The functional enrichment of IRDEGs was analyzed using the ClusterProfiler package of R (Version 3.16.1). The significantly enriched Gene Ontology (GO) terms were identified based on the GO functional enrichment analysis results ($p_{\text{adj.}} < 0.05$). Significantly enriched biological pathways were identified by the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis ($p_{\text{adj.}} < 0.05$).

2.4. Weighted Gene Coexpression Network Analysis (WGCNA), Survival Analysis, and Gene Set Enrichment Analysis (GSEA). WGCNA was performed on IRDEGs, in which the clinical information of patients and the expression levels of IRDEGs were all calculated by WGCNA. Base

d on the topological overlap measure model, genes were clustered using the average linkage hierarchical clustering method to obtain different gene modules. Module eigengenes were defined as the first principal component of each gene module. The Pearson correlation coefficient between the model eigengene of each module and the clinical information of the samples was calculated to obtain the gene significance (GS). The higher the GS was, the more important the module was, indicating how more correlated this module was to HPV16-infected samples. Then, the IRDEGs in the key modules derived from WGCNA were used for survival analysis using the Kaplan-Meier method, and the survival-related genes among IRDEGs were identified.

2.5. Establishment and Validation of a Prognostic Model. The survival-associated core IRDEGs were further selected from among the survival-related IRDEGs using the least absolute shrinkage and selection operator (LASSO) method. A risk scoring model that could predict the prognosis of the samples was constructed using these core IRDEGs. The risk score of each sample was then calculated based on the model. Based on the median risk score of all HNSCC patients, HNSCC patients were divided into high-risk and low-risk groups. Kaplan-Meier survival analysis was used to analyze the survival prognosis of patients in the high-risk and low-risk groups, and the diagnostic value of the model was analyzed using the receiver operating characteristic (ROC) curve and was validated on the data of HNSCC patients in the TCGA database. On this basis, we grouped all patients with HNSCC according to different clinical characteristics and used the

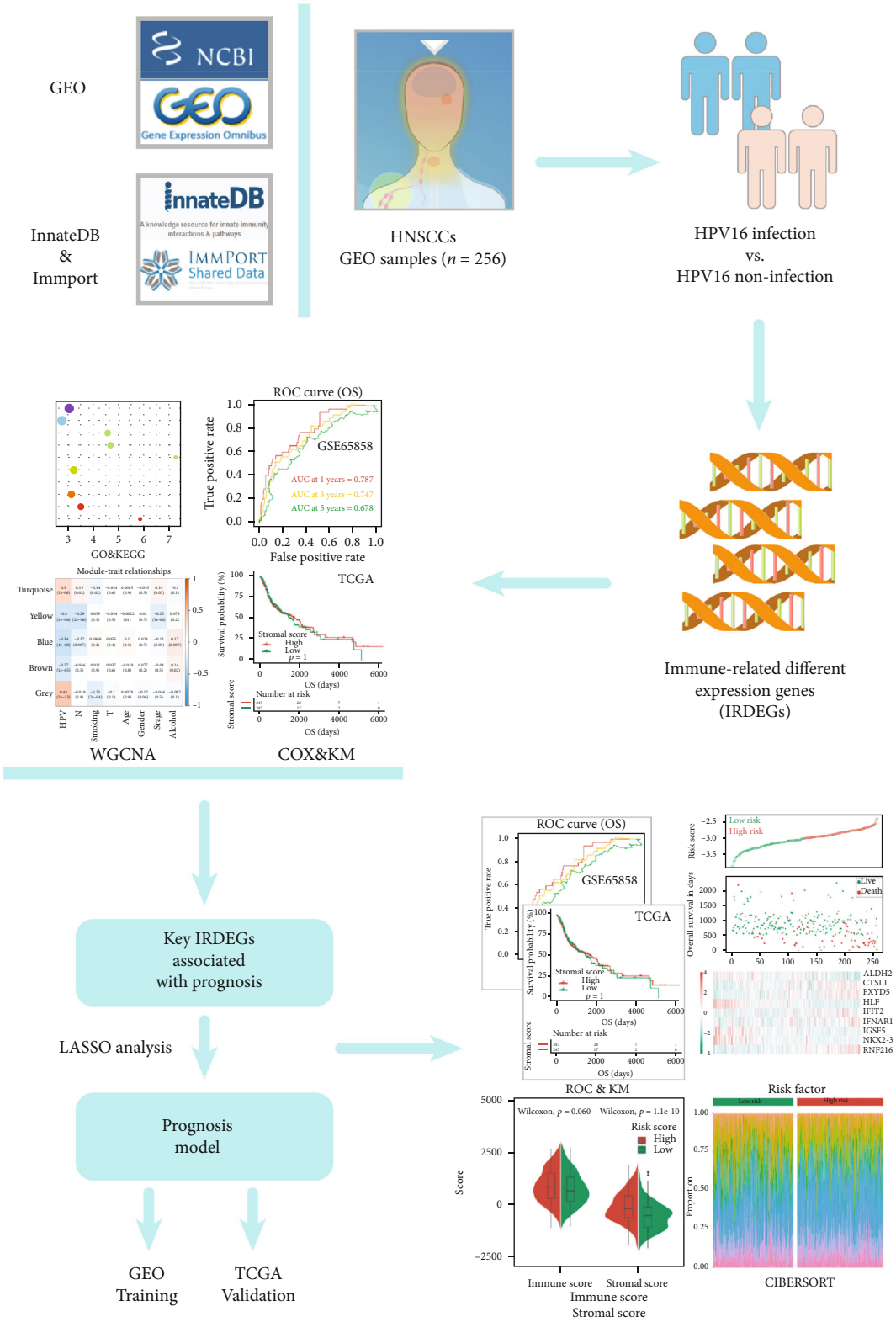


Figure 1: Flowchart of the present study.

Kaplan-Meier method to analyze the stability of the predictive ability of the prognostic model after grouping. Finally, we used univariate and multivariate Cox regression analyses to investigate the correlation between the prognostic model and other clinical characteristics (N stage, T stage, age, stage, and sex) to determine the independence of the model.

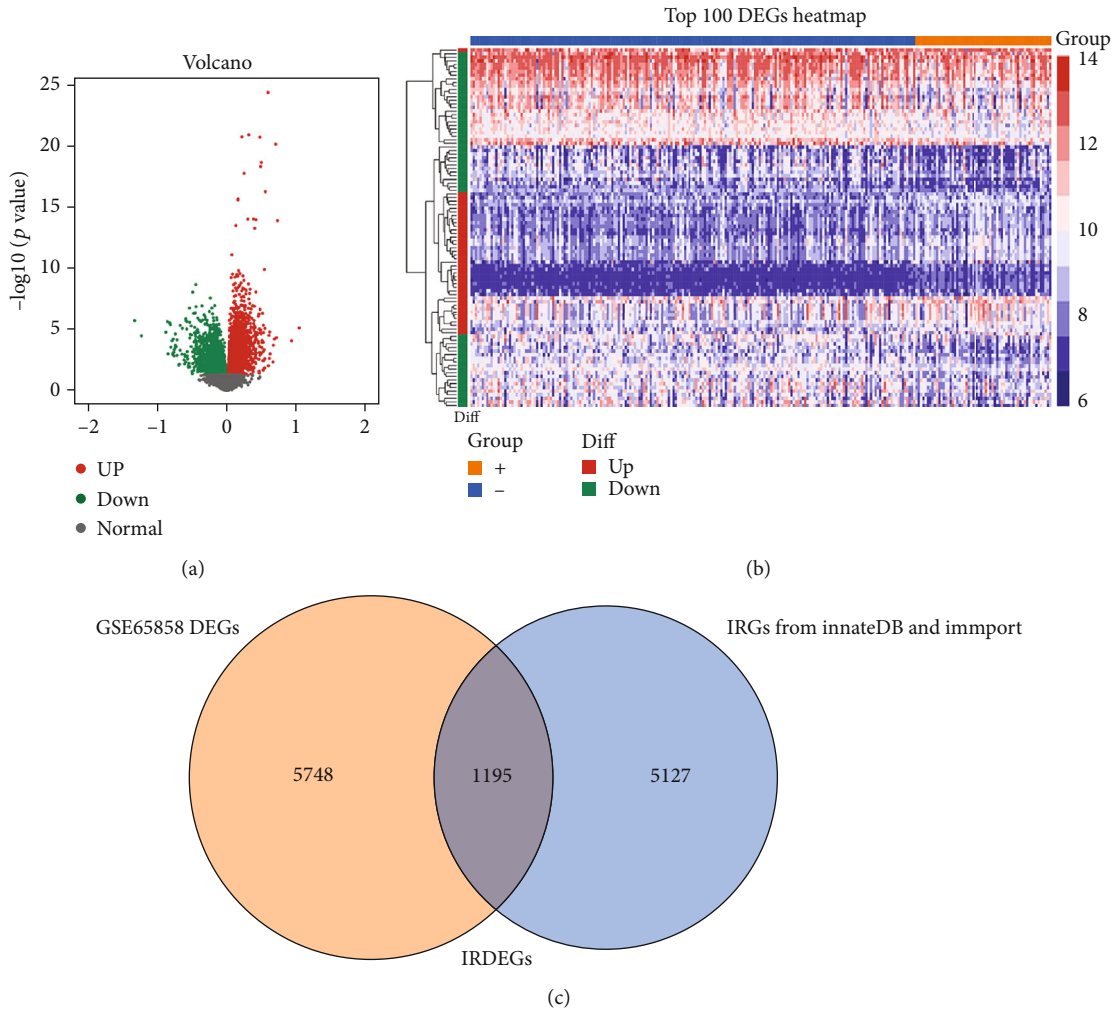


Figure 2: Overview of IRDEG profiling in HNSCC. (a) The DEGs between HNSCC tissues and normal tissues of dataset GSE65858. The DEGs identified in HNSCC were visualized in a volcano plot. The red and green points in the plot represent DEGs with statistical significance ($p_{adj.} < 0.05$). (b) Heat map of the top 100 DEGs. (c) The two datasets had an overlap of 1,195 genes.

2.6. Immune Infiltration and Calculation of Immune Cell Components. We used the estimate package of R to calculate the immune score and stromal score of each HNSCC patient by inputting the gene expression profile of each HNSCC patient, combined with the specific gene expression characteristics of immunization and stromal cells. Based on the median risk score of all HNSCC patients, HNSCC patients were divided into high-risk and low-risk groups, and the differences in the immune score and stromal score between the two groups were analyzed. We further used CIBERSORT and the LM22 matrix (a signature matrix containing 22 functionally defined human immune subsets) to calculate the proportions of 22 phenotypes of human hematopoietic cells in each HNSCC patient. The sum of the proportions of all estimated immune cell types in each sample was equal to 1. Based on the median risk score of all HNSCC patients, HNSCC patients were divided into high-risk and low-risk groups, and the differences in the proportions of immune cells between the two groups were analyzed.

2.7. Cell Culture and Transfection. The human HNSCC cell lines Hep-2 and TU212 were gifts from the Cancer Research Institute of Central South University. These two cell lines were cultured in the RPMI-1640 medium supplemented with 10% FBS, in a humidified incubator at 37°C with 5% carbon dioxide. Lentiviruses including IGSF5 overexpression plasmid and vector for IGSF5 were obtained from GeneChem (Shanghai, China) for in vivo experiments. Cells were seeded at 60-70% confluence and infected with lentivirus culture (MOI is 10). The noninfected cells were eliminated by puromycin, and stable cell lines were selected with 4 μg/ml puromycin treatment after 72 h of transfection. The efficiency of transfection was determined by Western blot.

2.8. Western Blot and Antibody. Antibodies against IGSF5 were from Thermo Fisher Scientific (catalog # PA5-80713, 1 : 1000, Thermo Fisher Scientific, MA, USA); GAPDH was from Proteintech (1 : 5000, Proteintech Group, Wuhan, China). The human HNSCC cell lines were lysed by the RIPA lysis buffer, which contains a protease inhibitor cocktail and phosphatase protease inhibitor cocktail. The BCA Protein Assay Kit (Beyotime Institute of Biotechnology,

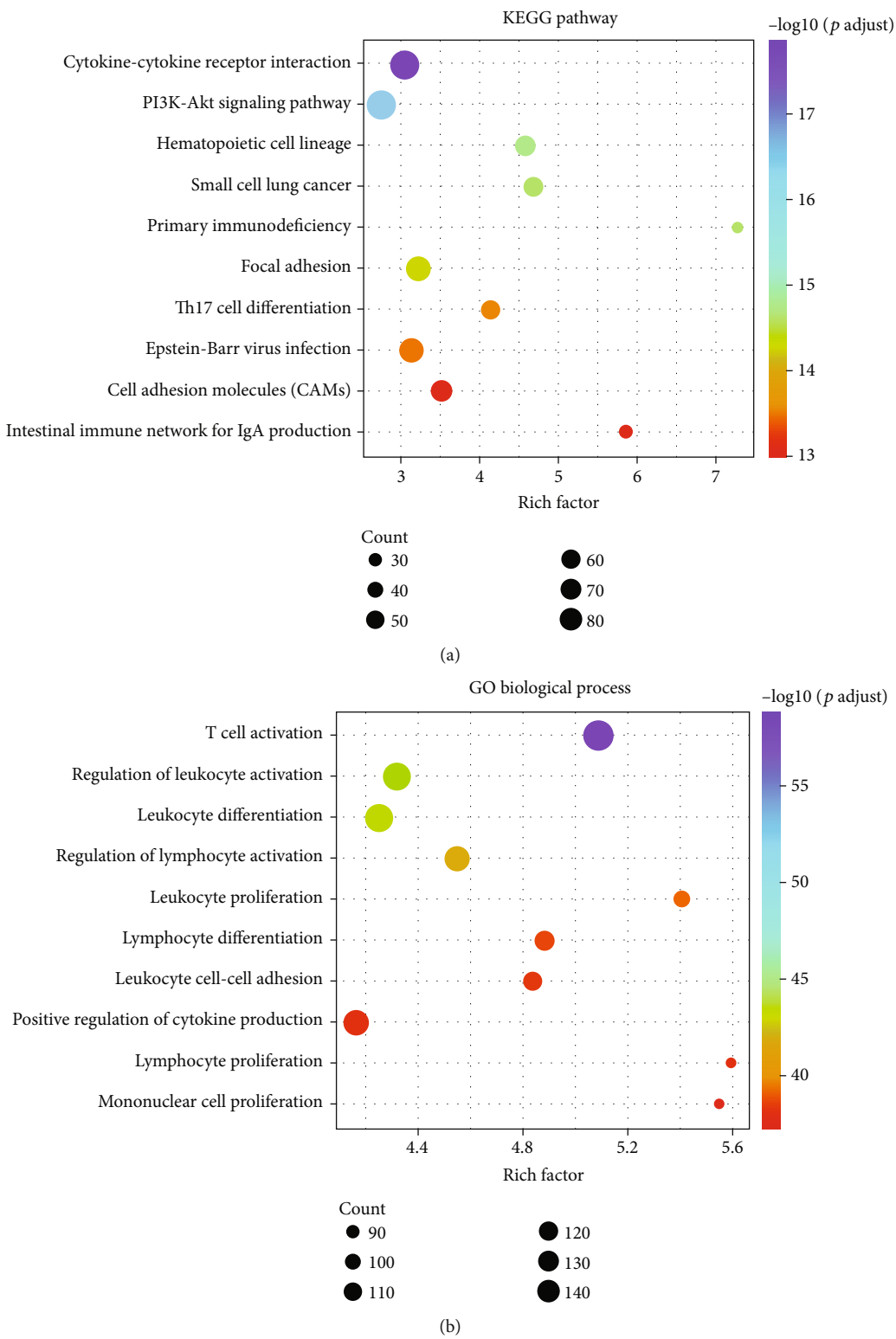
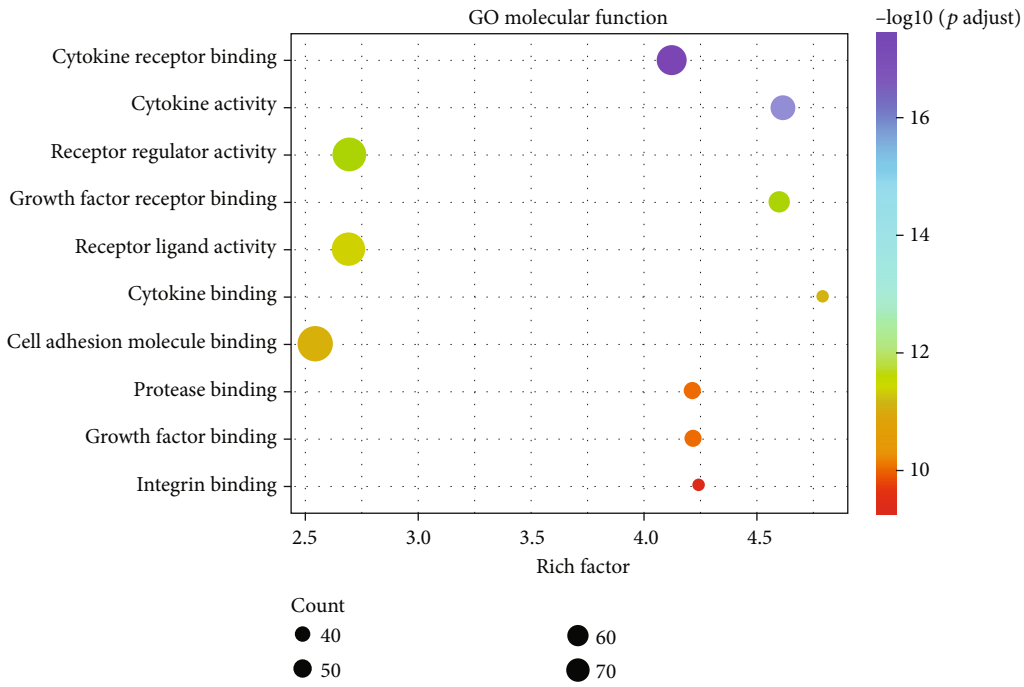
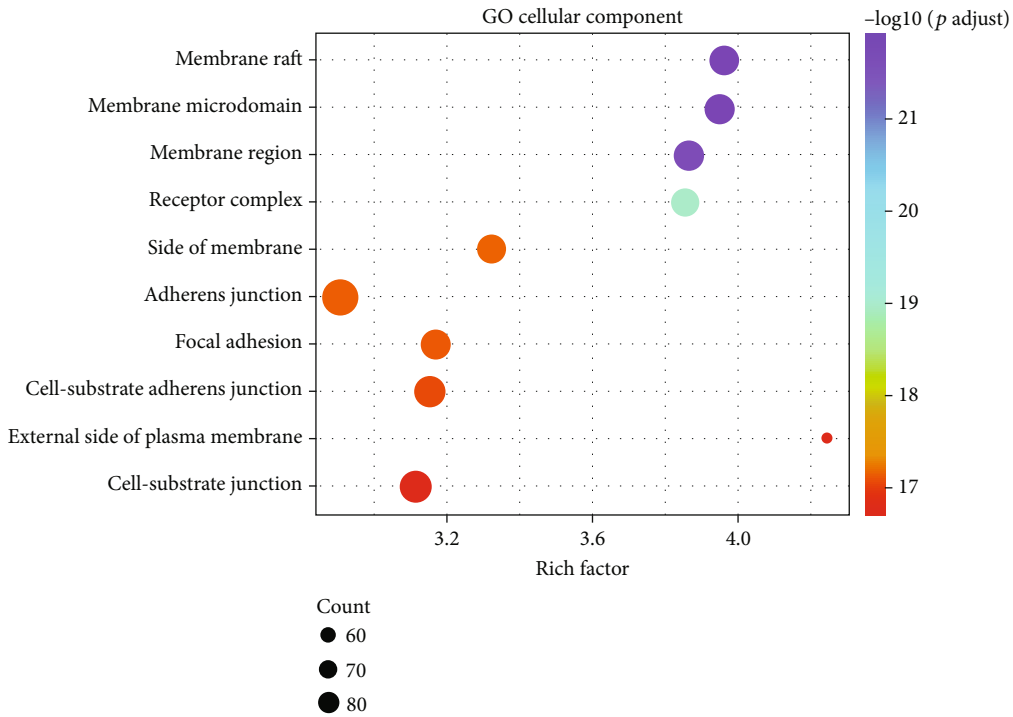


Figure 3: Continued.



(c)



(d)

Figure 3: KEGG and GO analysis of the overlapping IRDEGs in HNSCC: (a) KEGG pathways; (b) biological processes; (c) molecular functions; (d) cellular components. Functional and pathway enrichment of IRDEGs is displayed in bubble charts.

Shanghai, China) was used to detect the concentration of total protein. 10% SDS-PAGE was used to separate the total protein, and then, the protein was transferred to the PVDF membrane (Merck Millipore, USA). We block the membrane with 5% BSA (Sigma-Aldrich, Shanghai, China) for 1 hour. Then, the membrane was incubated with the antibody of IGSF5 at 4°C overnight; the membrane was washed with TBST six times. After incubation with secondary antibodies for 1 h at room temperature, the membrane was washed with TBST six times again. Finally, the bands were detected by the Odyssey CLx Infrared Imaging System (LI-COR Biosciences, NE, USA).

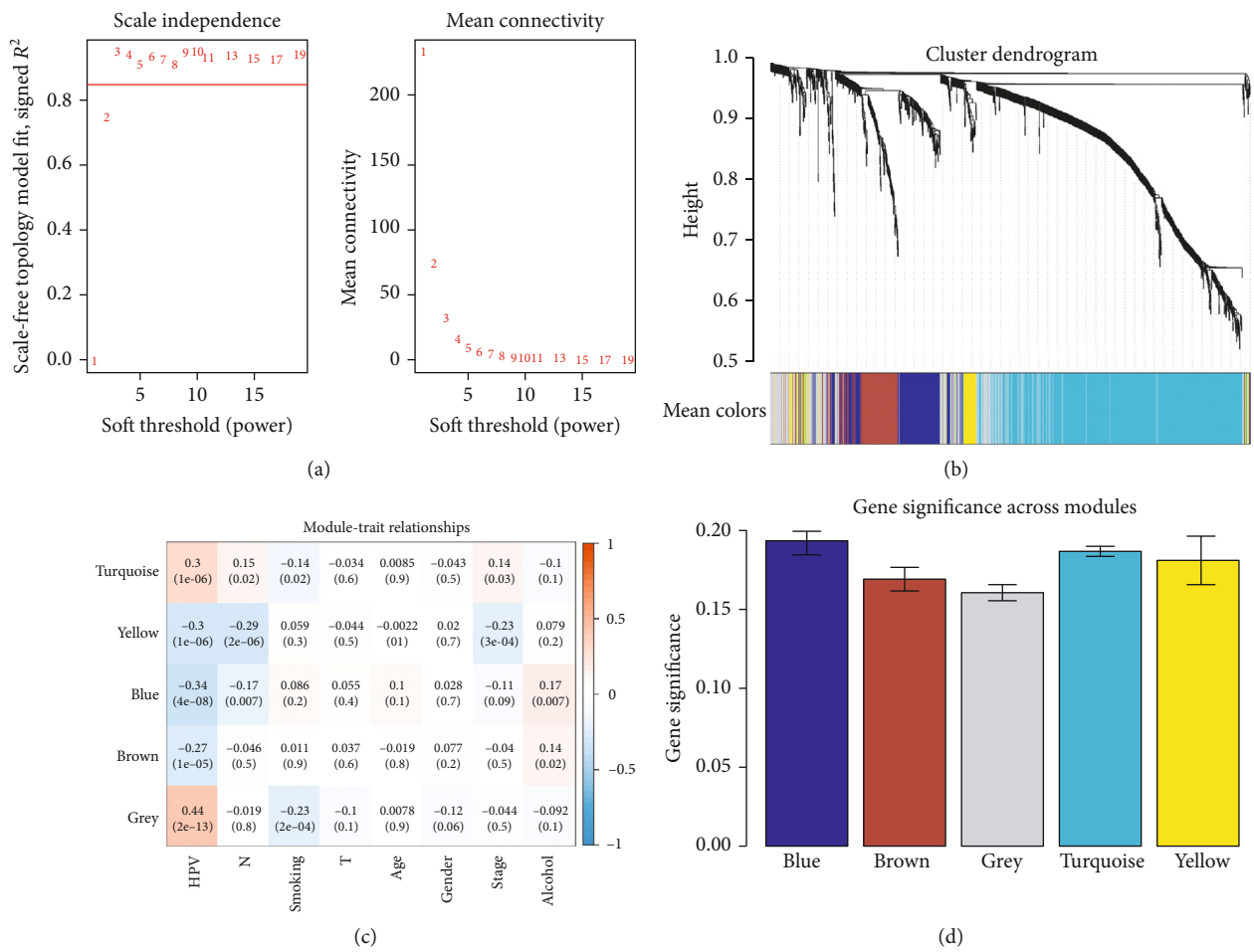


Figure 4: Weighted IRDEGs of the HNSCC expression network. (a) The scale-free fit index for soft-thresholding powers. The soft-thresholding power in the WGCNA was determined based on the scale-free R^2 ($R^2 = 0.85$). The left panel presents the relationship between the soft threshold and the scale-free R^2. The right panel presents the relationship between the soft threshold and mean connectivity. (b) A dendrogram of the differentially expressed genes clustered based on different metrics. Each branch in the figure represents one gene, and the color below it represents one expression module. (c) Heatmap showing the correlation between the gene module and clinical traits. The turquoise module contained 596 IRGs, while the blue module contained 186 IRGs. The correlation coefficient in each cell represents the correlation between gene module and clinical traits, which decrease in magnitude from red to blue. The turquoise module showed the highest positive correlation with survival, while the blue module showed the highest negative correlation with survival. (d) Distribution of average GS and errors in the modules associated with the overall survival of HNSCC patients. Based on the average linkage hierarchical clustering and the soft-thresholding power, six modules were identified. To determine the significance of each module, GS was calculated to measure the correlations between genes and sample traits. GS was defined as the $\log_{10}$ conversion of the p value in the linear regression between gene expression and clinical data ($GS = \log_{10} p$). The turquoise and blue modules showed high correlations with the survival of HNSCC patients.

2.9. Cell Proliferation and Colony Formation Assays. We used the cell counting kit 8 (CCK8) assay (Dojindo, Kumamoto, Japan) to measure cell proliferation in 96-well plates. Cells were seeded at 4×10^3 per well, with five replicates for each condition. CCK8 was added at 24, 48, 72, and 96 h and incubated at 37°C for 2 h. Cell numbers were determined by measuring the absorbance at 450 nm. We seeded 500 cells in each well of a 6-well plate in triplicate for each condition and incubated for 10 days. Then, the colonies were fixed by methanol and stained with crystal violet. The average colony counts were calculated. Each experiment was repeated three times.

2.10. Cell Apoptosis Analysis. Cells were harvested, washed twice with ice-cold phosphate-buffered solution (PBS), and fixed in 70% ice-cold ethanol, then incubated at 4°C overnight. Then, the cells were suspended following the manufacturer's protocol. Finally, a FACSCalibur flow cytometer was used for detection and FlowJo/ModFit LT software for data analysis (Becton Dickinson, NJ, USA). All experiments were performed three times.

3. Results

3.1. Selecting IRDEGs. The design of this study design is illustrated, as shown in Figure 1. To clarify the effects of HPV16 infection on the immune system of HNSCC patients, we first downloaded the GSE65858 data from the GEO database (containing 60 HNSCC patients with HPV16 and 196 HNSCC

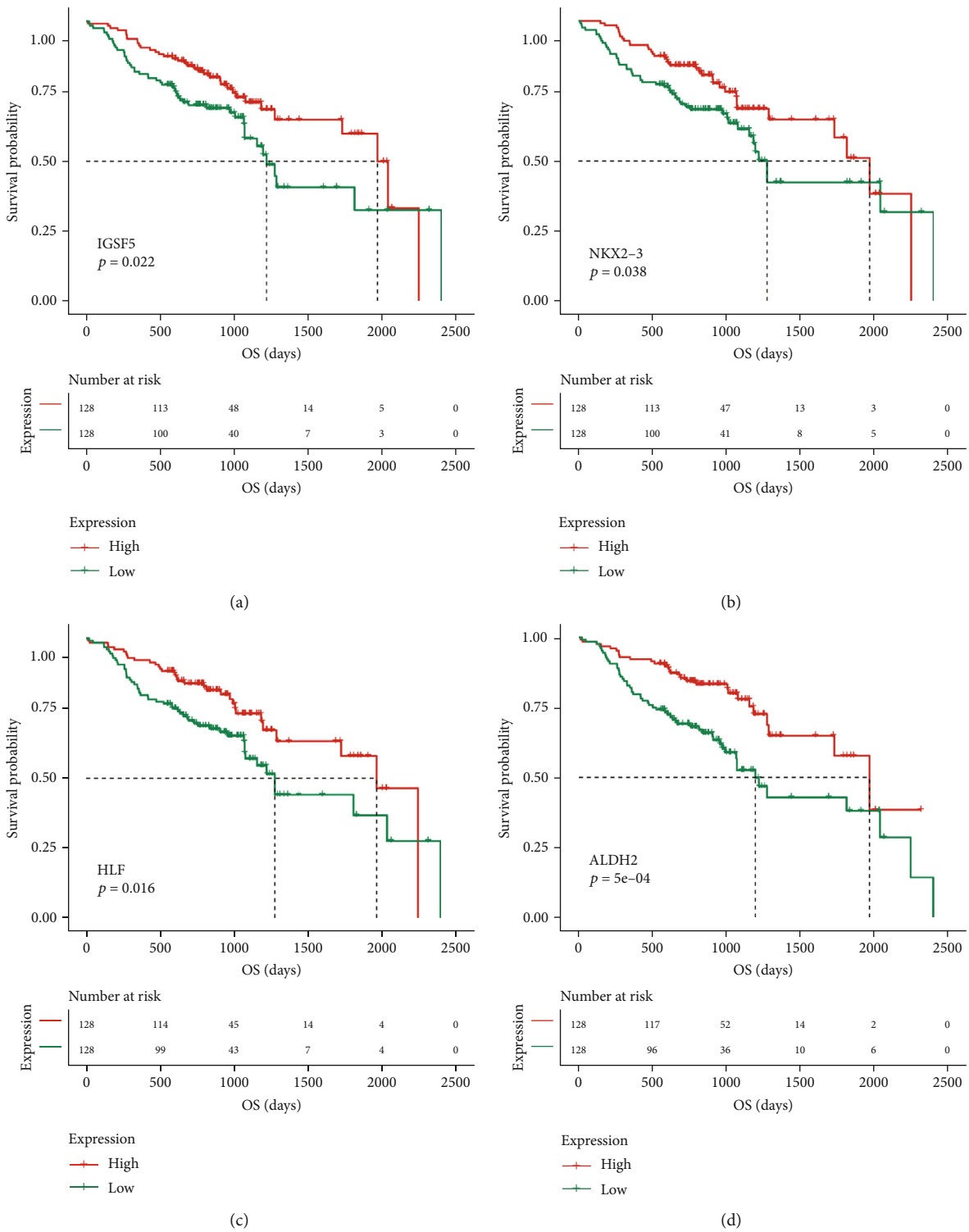


Figure 5: Continued.

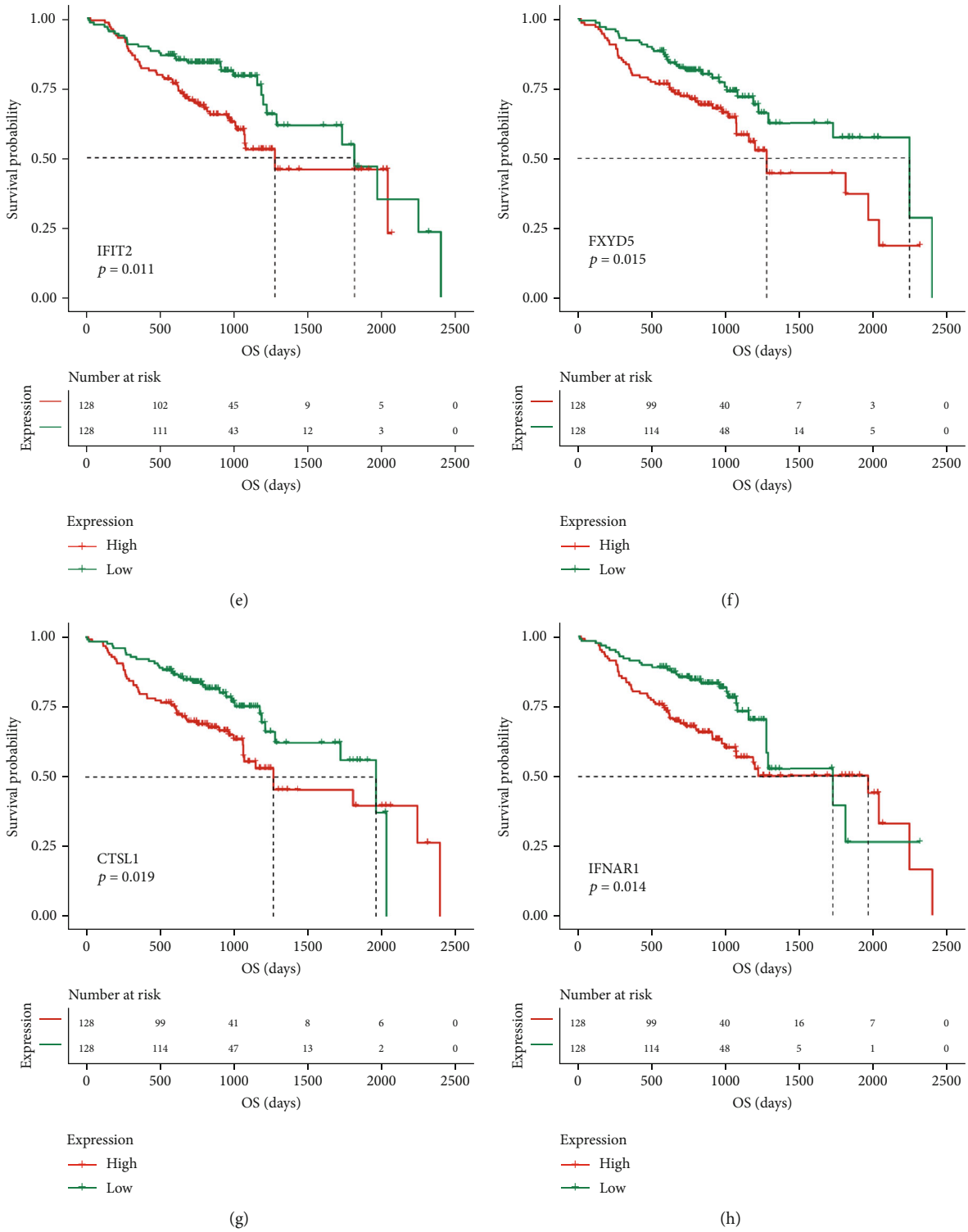
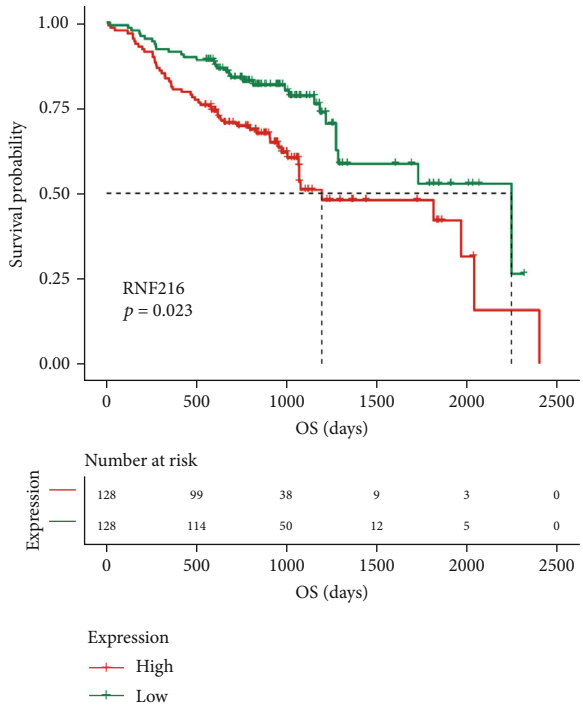


Figure 5: Continued.



(i)

Figure 5: The prognostic value of the nine key IRDEGs ((a) *IGSF5*, (b) *NKX2-3*, (c) *HLF*, (d) *ALDH2*, (e) *IFIT2*, (f) *FXYD5*, (g) *CTSL1*, (h) *IFNAR1*, and (i) *RNF216*) in HNSCC patients in the overall survival curve.

patients without HPV infection) to screen for DEGs. A total of 5,748 DEGs were identified ($p < 0.05$), of which 3218 were upregulated and 2,530 were downregulated (Figures 2(a) and 2(b)). Then, a total of 5,127 IRGs were downloaded from the InnateDB database and the ImmPort database. The intersection of the IRGs and DEGs was a total of 1,195 genes, called IRDEGs associated with HPV16-related HNSCC (Figure 2(c)). These findings suggest that HPV infection could cause an immune response in HNSCC patients.

3.2. IRDEG Enrichment Analysis. Next, to further clarify the signaling pathways and molecular functions in which these IRDEGs were involved, functional enrichment analysis was performed on all 1,195 IRDEGs. The KEGG results showed that these genes played regulatory roles in important immune-related pathways, such as primary immunodeficiency, and in tumor-related signaling pathways, such as the PI3K-Akt signaling pathway (Figure 3(a)). Our group further used GO analysis to categorize the functions of these genes, and similar results were obtained: these genes not only participate in T cell activation, regulation of leukocyte activation, leukocyte differentiation, and other immune-related biological processes (Figure 3(b)) but also regulate cytokine receptor binding, cytokine activity, receptor regulator activity, and other tumor-related molecular functions (Figure 3(c)). Some encoded proteins that are important components of membrane rafts, membrane microdomains, and membrane regions (Figure 3(d)). These results further indicate that HPV infection indeed had a certain impact on the immune system of HNSCC patients, thus affecting tumor occurrence and development.

3.3. WGCNA and Survival Analysis. After preliminary validation of the presence of IRDEGs and their signaling pathways, we further constructed a weighted gene coexpression network based on the 1,195 IRDEGs to identify key IRGs to the survival prognosis of HNSCC patients. To ensure that the network was a scale-free network, we chose the optimal R^2 of 0.85 (Figure 4(a)). We clustered these genes using the average linkage hierarchical clustering method. A total of five modules were obtained (Figure 4(b)). The number of genes in each module was statistically analyzed as shown in Supplement Table 1. As shown in Figure 4(c), the color changing from red to blue represents the correlation coefficient decreasing from high to low. We also calculated the GS value of each gene module (Figure 4(d)). A higher GS meant that the module was more correlated with samples with HPV infection. Based on the results of the above two methods of analyzing the correlation between module and phenotype, the modules most correlated with HPV infection are colored turquoise (positive) and blue (negative).

We next performed univariate Cox survival analysis on the 782 key module genes ($N = 596 + 186$) included in the positive and negative correlation modules. A total of 31 IRDEGs that were correlated with survival prognosis were selected (the survival curves of some prognosis-related factors are shown in Figure 5, $p < 0.05$). In a word, high expression of IGSF5, NKX2-3, HLF, and ALDH2 showed a better

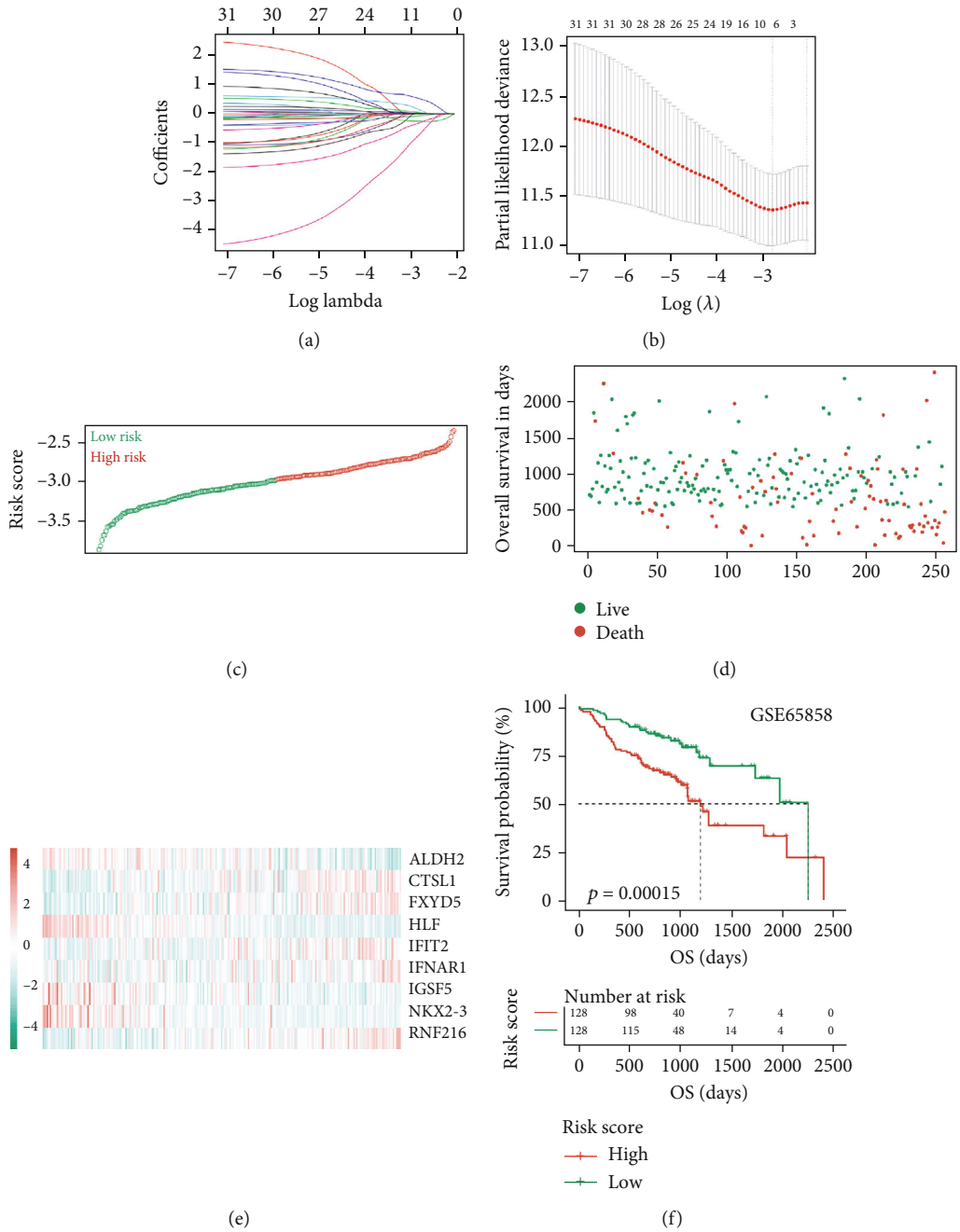


Figure 6: Continued.

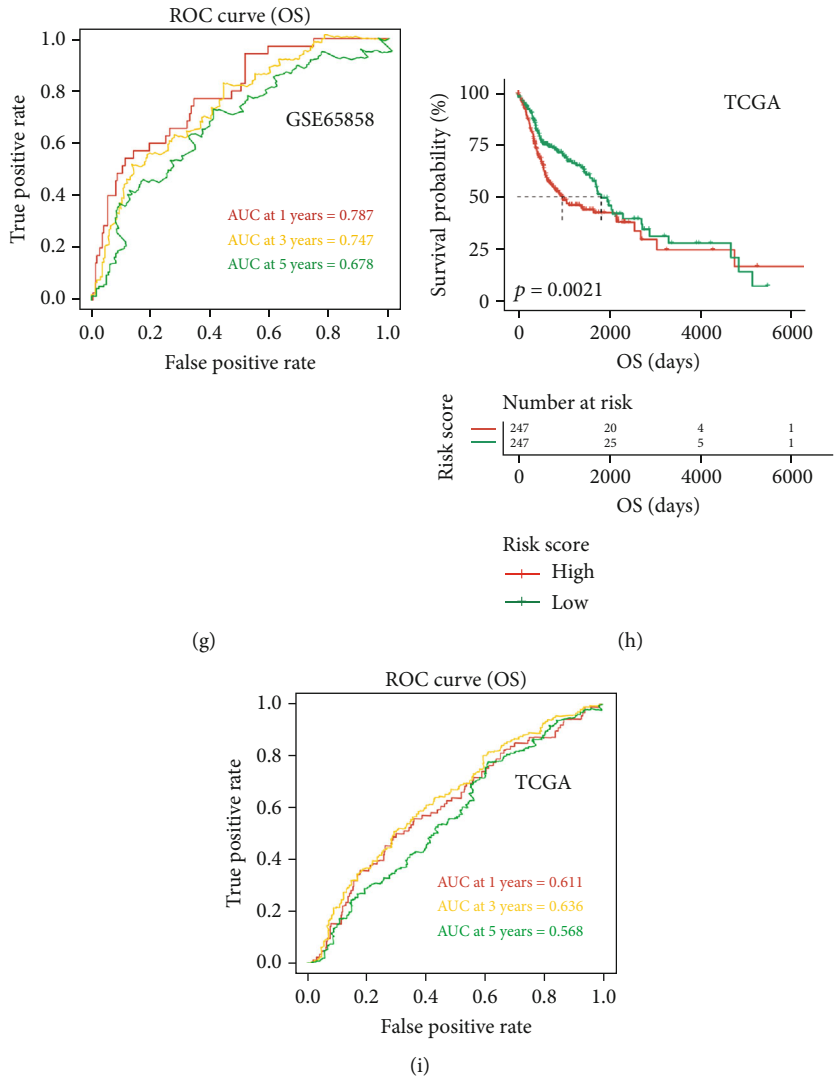


Figure 6: Construction of the IRDEG prognostic classifier, the distribution of time-dependent ROC curves, and Kaplan-Meier survival: (a, b) determination of the number of factors by LASSO analysis; (c) distribution of risk scores; (d) survival duration and status of patients; (e) heatmap of IRDEGs in the classifier; (f) Kaplan-Meier curve of the GSE65858 cohort; (g) ROC curve for the GSE65858 cohort; (h) Kaplan-Meier curve of the TCGA HNSCC cohort; (i) ROC curve for the TCGA HNSCC cohort.

Table 1: 9-gene immune signature.

Gene symbol	Univariate Cox regression analysis				LASSO coefficients
	HR	Low 95% CI	High 95% CI	*p* value	
IGSF5	1.65	1.07	2.54	0.023	-0.68
NKX2-3	1.58	1.02	2.43	0.039	-0.3
HLF	1.71	1.1	2.64	0.017	-0.26
ALDH2	2.18	1.39	3.43	0.001	-0.07
IFIT2	0.56	0.36	0.88	0.012	0.02
FXYD5	0.59	0.38	0.91	0.016	0.02
CTSL1	0.6	0.38	0.92	0.021	0.04
IFNAR1	0.58	0.37	0.9	0.015	0.13
RNF216	0.51	0.33	0.79	0.003	0.52

overall survival rate compared with the low-expression group based on the Kaplan-Meier analysis. Low expression of IFIT2, FXYD5, CTSL1, IFNAR1, and RNF216 showed a poor overall survival rate compared with the high-expression group based on the Kaplan-Meier analysis.

3.4. Construction and Validation of Prognostic Models. To better evaluate the clinical prognosis of HNSCC patients through these IRDEGs, we selected nine key prognostic genes, including immunoglobulin superfamily member 5 (*IGSF5*), NK2 homeobox 3 (*NKX2-3*), hevein-like protein (*HLPF*), aldehyde dehydrogenase 2 family member (*ALDH2*), interferon-induced protein with tetratricopeptide repeats 2 (*IFIT2*), FXYD domain-containing ion transport regulator 5 (*FXYD5*), cathepsin L1 (*CTSL1*), interferon alpha and beta receptor subunit 1 (*IFNAR1*), and ring finger protein 216 (*RNF216*), using the LASSO method based on 31 survival-related IRDEGs (Figures 6(a)–6(i)). The expression levels of

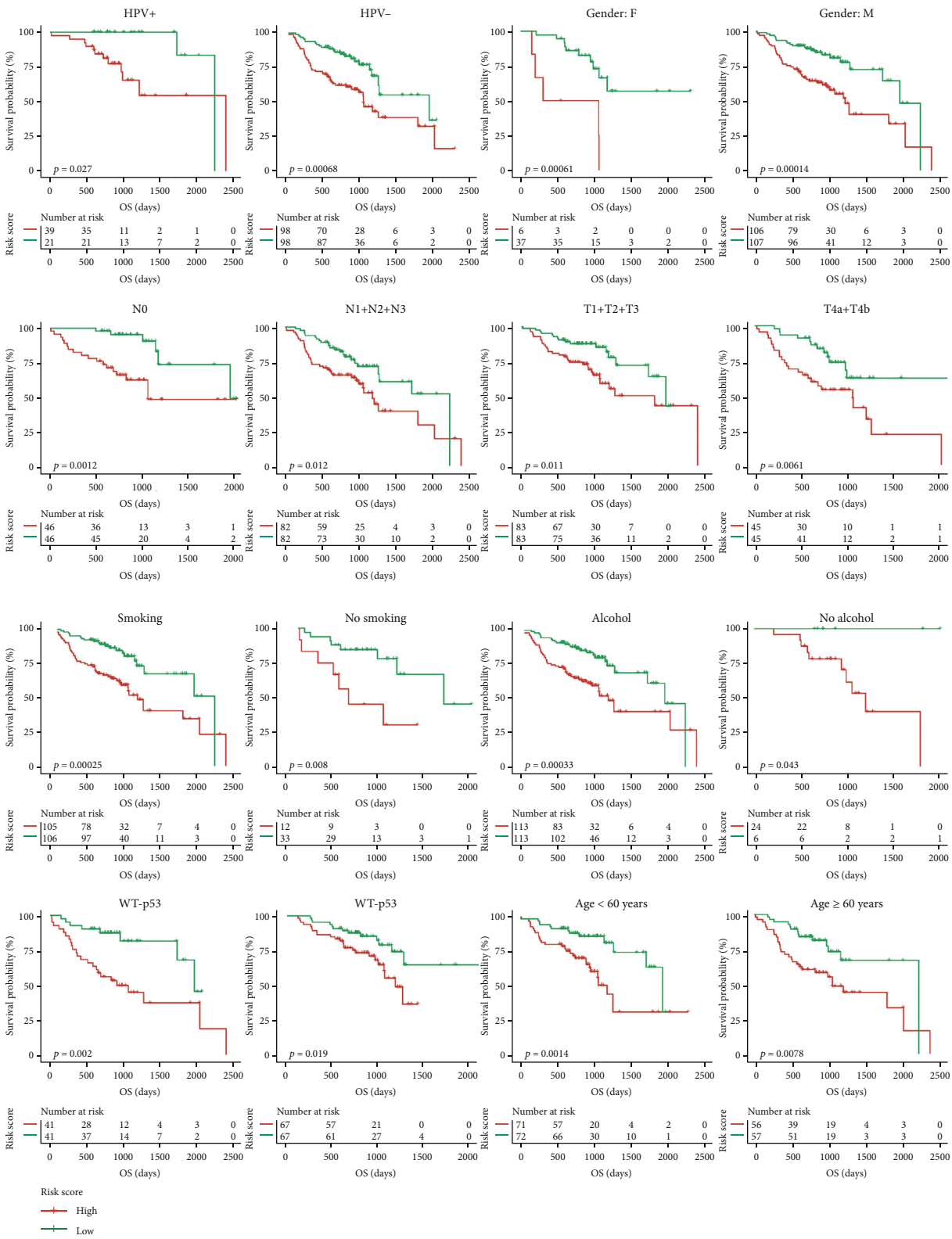


Figure 7: Effects of different clinical characteristics on the prognosis of HNSCC patients. Based on the risk score of 256 HNSCC patients in the GSE65858 dataset calculated by the prognostic evaluation model, patients were divided into high- and low-risk groups after first being stratified by clinical features such as HPV infection, sex, lymph node metastasis, degree of tumor infiltration, smoking, alcohol consumption, *TP53* gene mutation, and age. All these features had an impact on the prognosis of HNSCC patients.

Table 2: Univariate and multivariate analyses of prognostic factors and overall survival of HNSCC patients in the GSE65858 cohort.

	Univariate analysis				Multivariate analysis			
	HR	Low 95% CI	High 95% CI	*p* value	HR	Low 95% CI	High 95% CI	*p* value
N	1.41	1.12	1.78	0.004	1.53	1.06	2.21	0.024
T	1.47	1.18	1.83	0.001	1.29	0.97	1.7	0.076
Age	1.03	1.01	1.05	0.006	1.03	1.01	1.05	0.015
Stage	1.56	1.16	2.09	0.003	0.88	0.53	1.45	0.614
Gender	1	0.57	1.75	0.999	—	—	—	—
Risk score	11.82	4.72	29.56	<0.001	8.84	3.67	21.26	<0.001

these nine key factors were weighted by the LASSO regression coefficients (Table 1), and the risk scoring model for predicting the survival prognosis of HNSCC patients was successfully established. The risk score of each sample was calculated based on the model. According to the median risk score of all HNSCC patients in GSE65858, HNSCC patients were divided into the high-risk and low-risk groups. There was a significant difference in prognosis between the two groups (Figure 6(f)). The prediction results of the model were evaluated using the ROC curve. The area under the ROC curve (AUC) for predicting the prognosis of HNSCC patients in the GNS65858 dataset at 1 year, 3 years, and 5 years reached 0.787, 0.747, and 0.678, respectively (Figure 6(g)), evidencing the effectiveness of this prognostic model. That is, this model could be used to evaluate the clinical prognosis of HNSCC patients. The above results were verified in HNSCC samples from the TCGA database to further confirm the effectiveness of the model in determining the prognosis of HNSCC patients (Figures 6(h) and 6(i)).

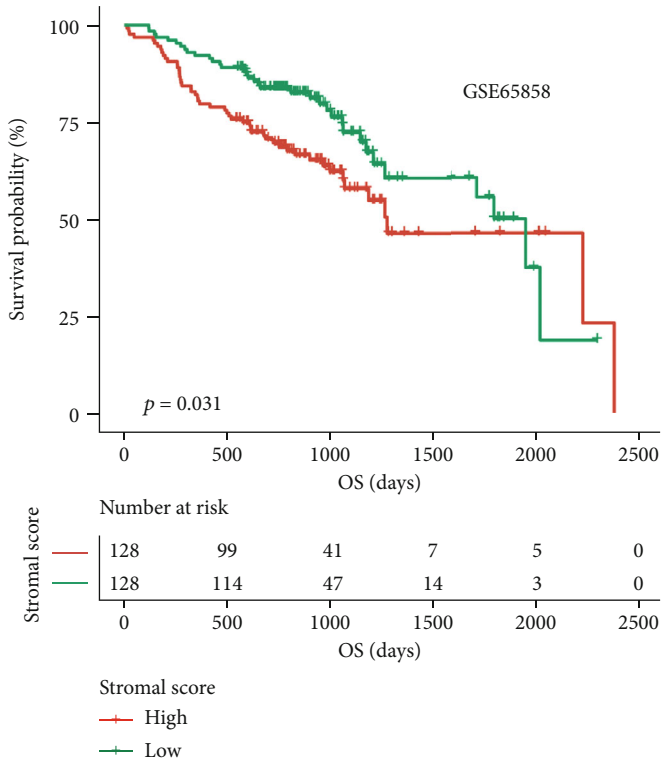
In addition, to validate the stability of the model, 256 HNSCC samples in GSE65858 were grouped according to clinical characteristics. These HNSCC samples were divided into a high-risk group and a low-risk group based on the median model score. The results showed that whether we grouped according to HPV^+, HPV^-, female sex, male sex, N0, N1+N2 +N3, T1+T2+T3, T4a+T4b, smoking, nonsmoking, alcohol, no alcohol, MT53, WT53, age ≥ 60, or age < 60, there was a significant difference in the prognosis between the high-risk group and the low-risk group (Figure 7). These findings further indicate that HPV infection is an important factor affecting the survival prognosis of HNSCC patients and that the predictive power of the predictive model is stable. Similarly, the stability of this prognostic model was also confirmed in HNSCC samples from TCGA (Supplement Figure 1).

Finally, to verify the independence of this prognostic evaluation model, we used univariate and multivariate Cox regression analyses to analyze the relationship between the prognostic evaluation model and other clinical factors in the experimental group. Univariate Cox regression analysis showed that the N stage, T stage, stage, age, and our prognostic model risk score were all significantly correlated with survival. Multivariate Cox regression analysis showed that the N stage, age, and prognostic model score were independent prognostic factors (Table 2). Applied to the TCGA validation set, univariate Cox regression analysis showed that age, sex, and prognostic model score were all significantly associated with survival; multivariate Cox regression analysis showed that age and the constructed prognostic model score were independent prognostic factors (Supplement Table 2). The above results indicate that the prognostic evaluation model based on IRDEGs is stable and reliable and can be used to evaluate the survival prognosis of HNSCC patients.

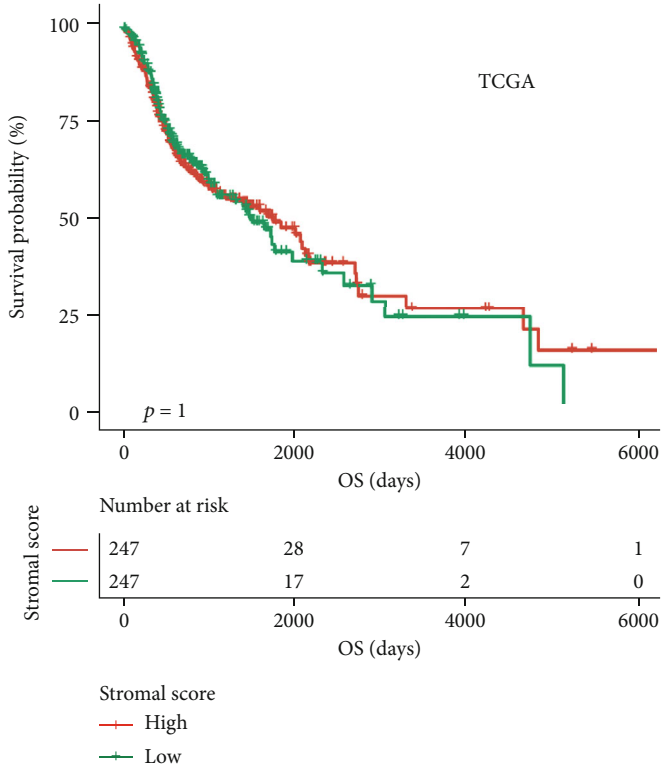
3.5. Immune Score and Stromal Score. To further assess the relationship between immune infiltration and survival prognosis in HNSCC patients, we evaluated the immune score and stromal score. The HNSCC patients in the experimental group and the validation group were divided into two subgroups according to the median value of the immune score or the stromal score. There were significant differences in the prognosis between patients in the two subgroups of the experimental group (Figures 8(a) and 8(c), $p < 0.05$). There was no significant difference in the prognosis between the two subgroups divided according to the stromal score in the validation group (Figure 8(b), $p > 0.05$), but significant differences were identified between two subgroups divided according to the immune score (Figure 8(d), $p < 0.05$).

HNSCC patients in the experimental group and the validation group were divided into high-risk and low-risk groups according to the prognostic evaluation model. There was no significant difference in the immune score between patients in the high- and low-risk subgroups of the experimental group or the validation group (Figures 8(e) and 8(f), left part, $p > 0.05$), but there was a difference in the stromal score between the high- and low-risk patients of each group (Figures 8(e) and 8(f), right part, $p < 0.05$). These results suggest that the degree of immune infiltration affects the survival prognosis of HNSCC patients, and the prognostic evaluation model constructed by our research group could evaluate the degree of immune infiltration in HNSCC patients.

3.6. Immune Landscape of HNSCC Samples with Low/High Risk. After confirming the correlation between immune infiltration and the survival prognosis of HNSCC patients, we further explored the proportion of immune cells in the tumor tissues of HNSCC patients. The results showed that the proportion of immune cells in HNSCC samples differed between low- and high-risk HNSCC samples (Figure 9). Specifically, the analysis of the experimental group showed that the proportions of naive B cells, plasma cells, resting $CD4^+$ memory T cells, follicular helper T cells, gamma delta T cells, M0 macrophages, resting dendritic cells, activated dendritic cells, activated mast cells, and neutrophils were significantly different between the two

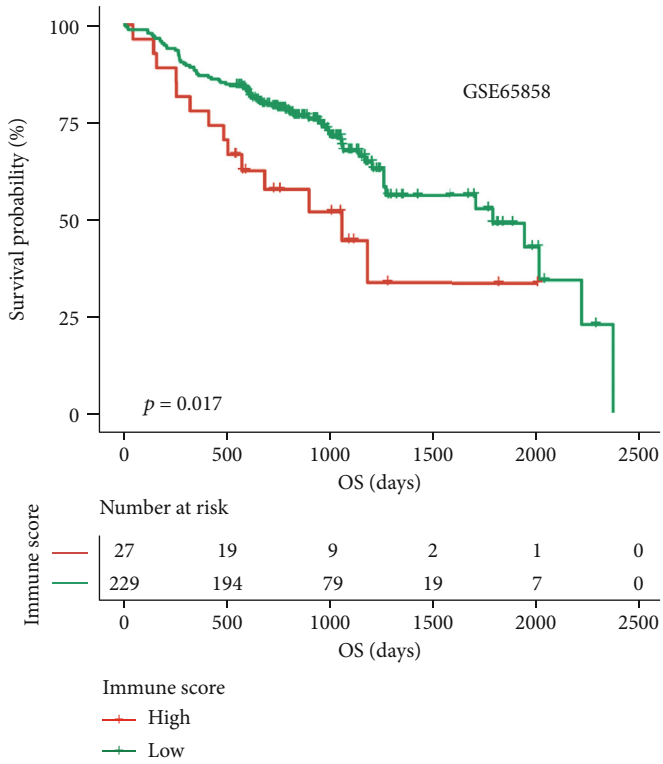


(a)

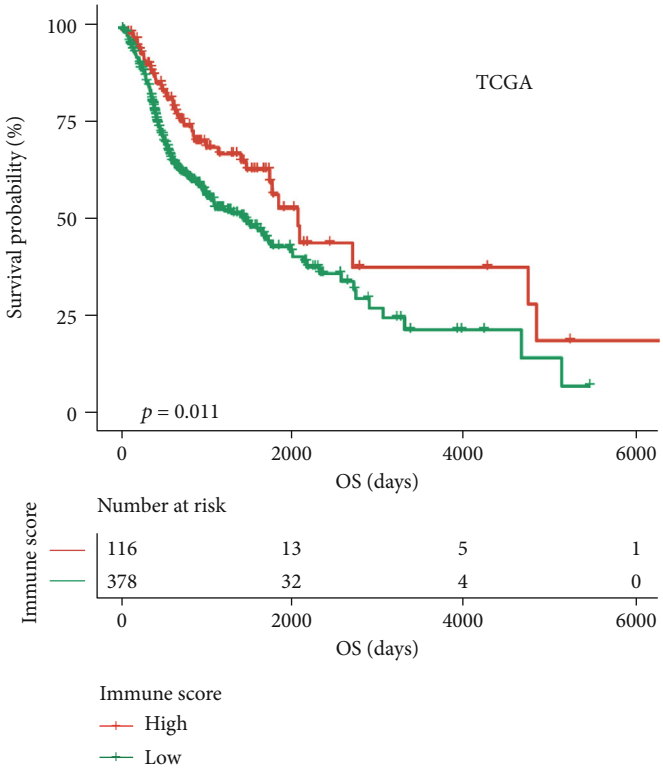


(b)

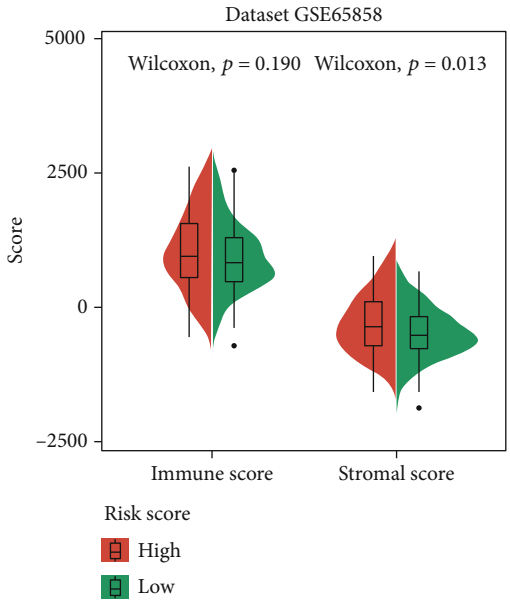
Figure 8: Continued.



(c)



(d)



(e)

Figure 8: Continued.

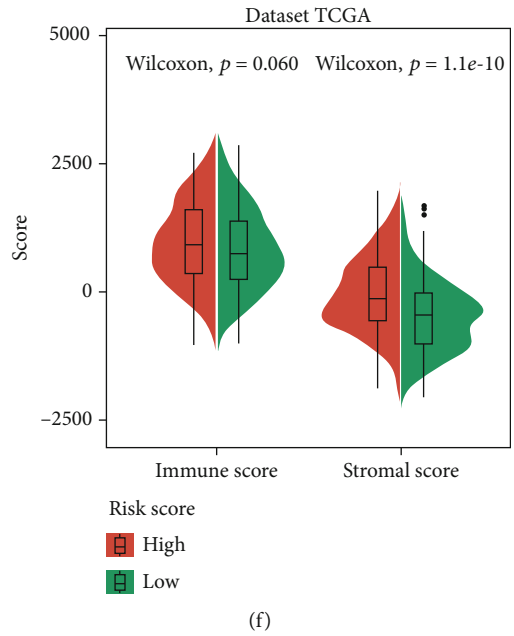


Figure 8: Effect of immune infiltration on the prognosis of HNSCC patients: (a, c) impact of the stromal score and immune score on overall survival in the GSE65858 cohort based on Kaplan-Meier analysis; (b, d) impact of the stromal score and immune score on overall survival in the TCGA HNSCC cohort based on Kaplan-Meier analysis; (e) association between immune score, stromal score, and risk score in the GSE65858 cohort; (f) association between immune score, stromal score, and risk score in the TCGA HNSCC cohort.

groups of samples ($p < 0.05$, Table 3). In the validation group, the proportion of 11 cell types, including naïve B cells, memory B cell, plasma cells, $CD8^+$ T cells, resting $CD4^+$ memory T cells, follicular helper T cells, regulatory T cells, gamma delta T cells, resting NK cells, M0 macrophages, and M2 macrophages, were significantly different between the two groups (Table 4).

4. Validation of IGSF5 Expression in HNSCC Cells from the Prognostic Model

Since the IGSF5 gene is the most relevant factor in the model (LASSO coefficients = −0.68, Table 1), we selected it for preliminary verification. Firstly, we found that IGSF5 has low expression in HNSCC according to the data from starBase v3.0 (Figure 10(a)). And IGSF5 expression was significantly elevated in HNSCC patients infected with HPV (Figure 10(b)), which was consistent with the prognostic model. To validate the function of IGSF5, we upregulated the expression of IGSF5 in HNSCC cells (Figure 10(c)). Then, we found that overexpression of IGSF5 significantly impaired HNSCC cell proliferation and in Hep-2 and TU212 cells (Figures 10(d)–10(f)) and dramatically promotes cell apoptosis (Figures 10(g) and 10(h)).

5. Discussion

In recent years, tumor immunity and treatments related to it have been hot spots and challenges in HNSCC research. In this study, we first selected 31 key IRDEGs that affected the prognosis of HNSCC patients in terms of HNSCC patients with or without HPV16 infection and further used nine key IRDEGs to construct a stable and highly efficient prognostic evaluation model to evaluate the survival prognosis and immune infiltration of HNSCC patients in clinical practice. In our study, 5,127 IRGs were downloaded from the InnateDB database and ImmPort database, and 1,195 IRDEGs were obtained. In addition, KEGG and GO analysis confirmed that these genes not only are involved in many immunoregulatory pathways but also play important roles in tumor occurrence and development. She et al. previously studied the differences in IRGs among HNSCC patients with or without HPV infection. However, the IRGs in their study were limited to the 1,073 genes included in the ImmPort website. Moreover, their study did not examine the differential expression of IRGs between cancerous and paracancerous tissues of HNSCC patients [12]. Therefore, compared with our study, theirs has more limitations that greatly reduced the reliability of their results.

To evaluate the clinical prognosis of HNSCC patients, we built a prognostic model using the LASSO method. The nine core genes of the model, *IGSF5* [13], *NKX2-3* [14], *HLF* [15], *ALDH2* [16], *IFIT2* [17], *FXYD5* [18], *CTSL1* [19], *IFNAR1* [20], and *RNF216* [21], have been associated with tumor occurrence and development. *ALDH2* plays an important regulatory role in HNSCC [22] and the tumor immune response [23]. *HLF* is closely related to prognosis and drug resistance in HNSCC patients [24] and participates in the immune response of $CD4^+$ and other immune cells [25]. *IFNAR1* is highly expressed in HNSCC patients, where it promotes the expression of programmed death-ligand 1 (PDL1) and programmed cell death protein 1 (PD1) in tumor cells [26]. These findings indirectly show that our risk prognosis model constructed based on the key IRDEGs related to prognosis was reliable. More importantly, the AUC values for predicting the survival of HNSCC patients in the GNS65858 dataset at 1 year, 3 years, and 5 years

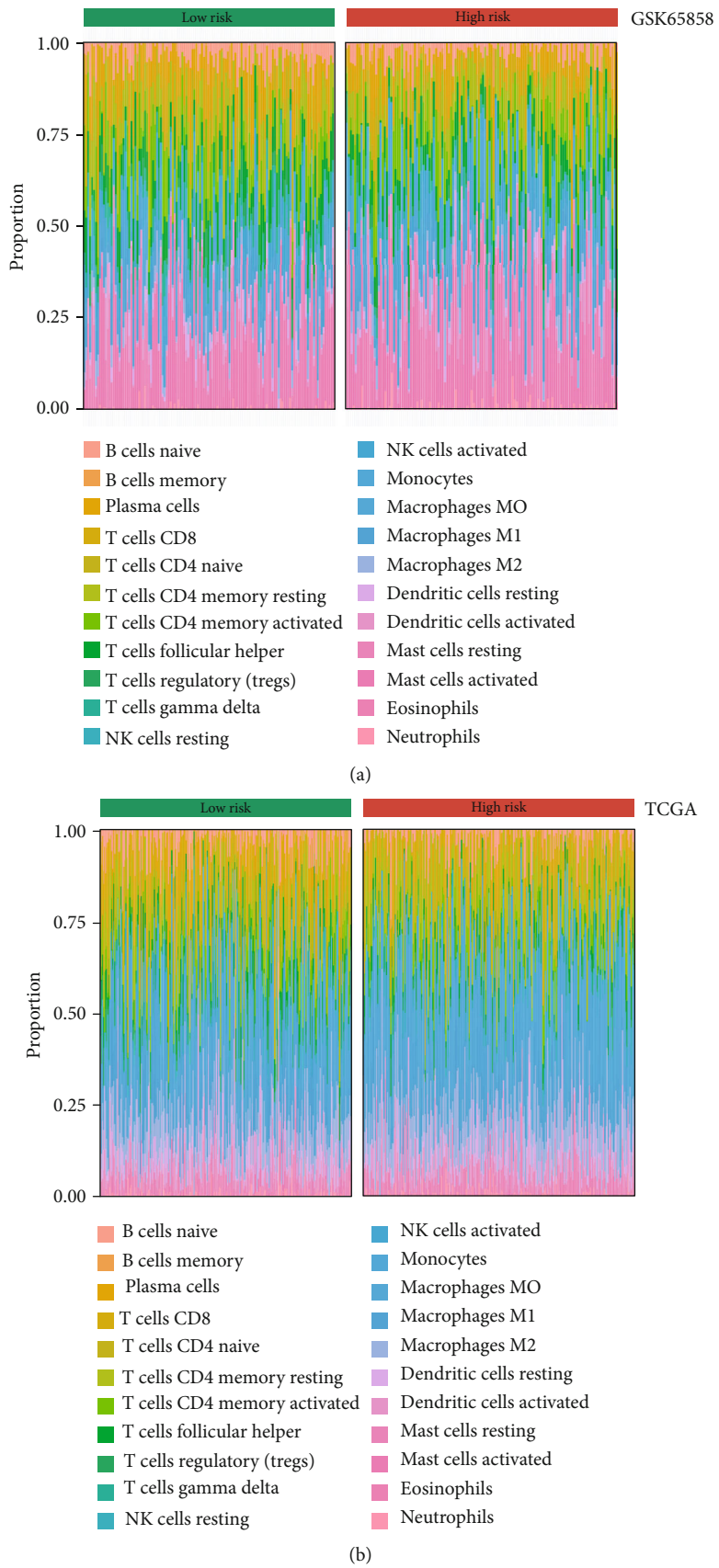


Figure 9: The proportions of immune cell types in HNSCC patients: (a) the mean proportions of 22 immune cell types in the GSE65858 cohort; (b) the mean proportions of 22 immune cell types in the TCGA HNSCC cohort.

Table 3: Differential immune cell type expression was observed between the high and low RS groups in the GSE65858 cohort.

Cell type	Low risk	High risk	*p* value
Naive B cells	0.039804215	0.028293595	0.002849395
Memory B cells	0.003513616	0.004345782	0.691435185
Plasma cells	0.084088896	0.057558826	0.002982023
CD8 T cells	0.042427714	0.043767342	0.840225177
Naive CD4 T cells	0.012275351	0.012559372	0.938789785
Resting CD4 memory T cells	0.083830748	0.064374727	0.006572652
Activated CD4 memory T cells	0.080999535	0.081630012	0.948165785
Follicular helper T cells	0.087385368	0.074676429	0.019497779
Regulatory T cells (Tregs)	0.013609875	0.009772032	0.214696852
Gamma delta T cells	0.042151413	0.028734392	0.002338151
Resting NK cells	0.000999625	0.00106397	0.933127189
Activated NK cells	0.053364353	0.053284618	0.983751584
Monocytes	0.024330133	0.026670409	0.505003128
M0 macrophages	0.063840447	0.078844898	0.049219689
M1 macrophages	0.055566688	0.061100091	0.312529981
M2 macrophages	0.026315633	0.030067857	0.323696097
Resting dendritic cells	0.019759167	0.013892642	0.040243924
Activated dendritic cells	0.035258645	0.046148972	0.042403999
Resting mast cells	0.005683298	0.002321107	0.213976349
Activated mast cells	0.220380575	0.271912742	0.002425909
Eosinophils	0.000213091	0.000188235	0.914308033
Neutrophils	0.004201615	0.00879195	0.011608708

Table 4: Differential immune cell type expression was observed between the high- and low-RS groups in the TCGA cohort.

Cell type	Low risk	High risk	*p* value
Naive B cells	0.040267	0.02912	0.005183
Memory B cells	0.004514	0.001859	0.026328
Plasma cells	0.058277	0.03223	8.52*E*-06
CD8 T cells	0.108232	0.087539	0.006781
Naive CD4 T cells	0.002498	0.003162	0.638828
Resting CD4 memory T cells	0.072697	0.096681	0.000199
Activated CD4 memory T cells	0.051323	0.044211	0.100323
Follicular helper T cells	0.045584	0.029642	7.57*E*-08
Regulatory T cells (Tregs)	0.035	0.021083	2.45*E*-07
Gamma delta T cells	0.002973	0.001212	0.019656
Resting NK cells	0.024675	0.034967	0.000234
Activated NK cells	0.014214	0.014061	0.939139
Monocytes	0.002964	0.002761	0.78523
M0 macrophages	0.214181	0.256082	0.002617
M1 macrophages	0.087477	0.095573	0.146756
M2 macrophages	0.091557	0.109234	0.000378
Resting dendritic cells	0.046756	0.040619	0.204292
Activated dendritic cells	0.035284	0.034777	0.902059
Resting mast cells	0.027813	0.023826	0.160864
Activated mast cells	0.023828	0.03231	0.057819
Eosinophils	0.000594	0.000908	0.30239
Neutrophils	0.00929	0.008142	0.502245

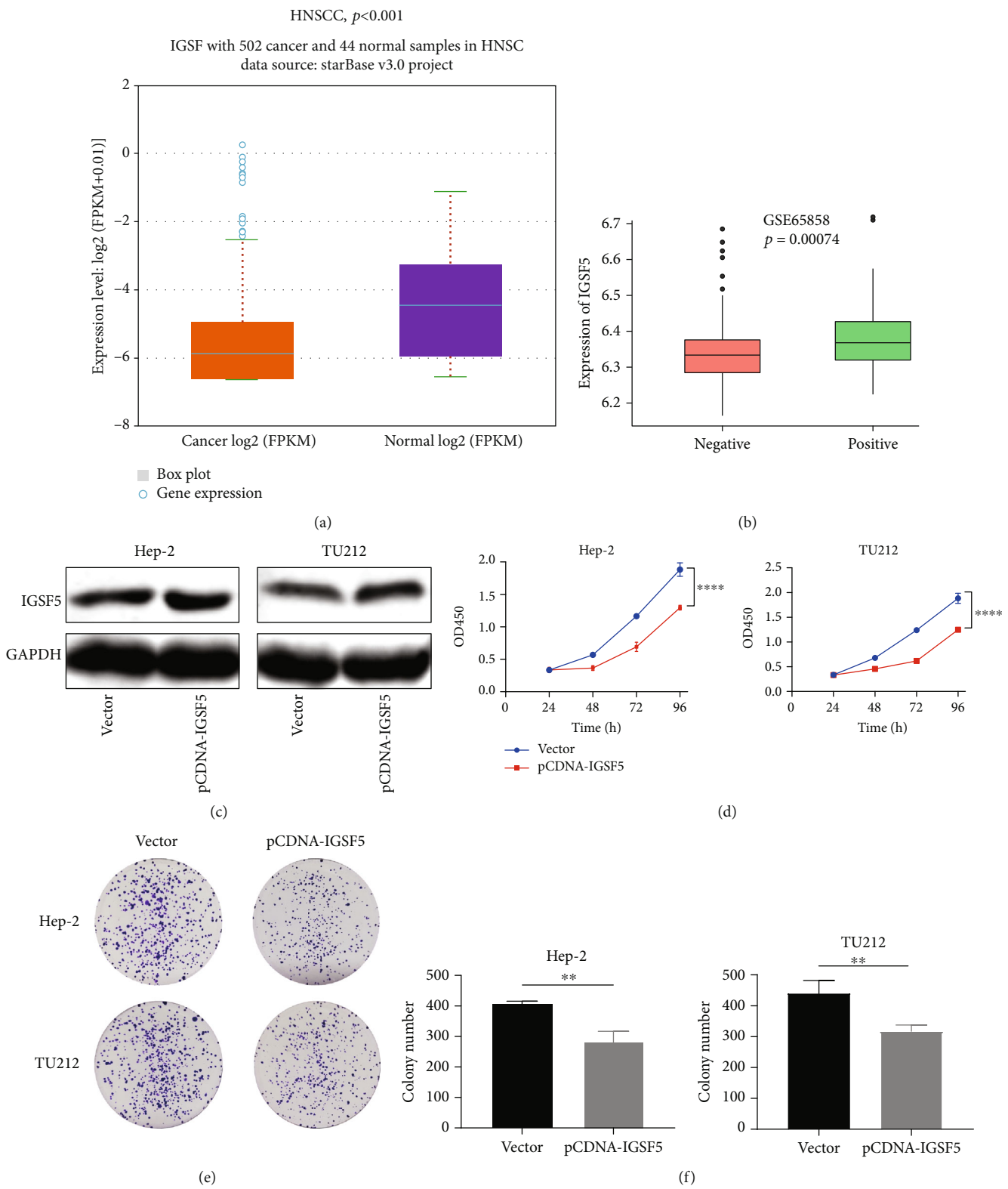


Figure 10: Continued.

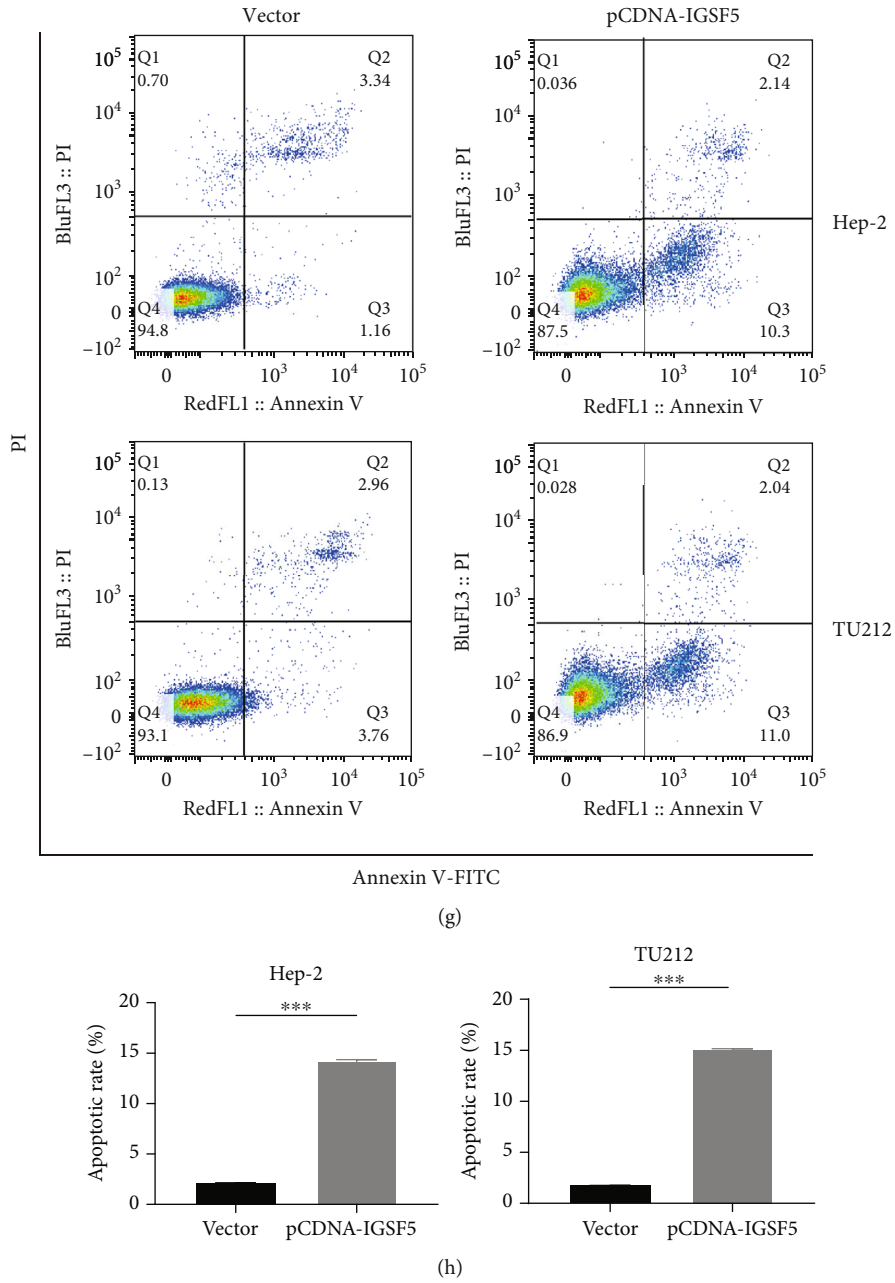


Figure 10: Overexpression of IGSF5 inhibits the proliferation of HNSCC cells. (a) IGSF5 had lower expression in HNSCC samples than in matched normal tissues from the TCGA database (normal = 44, tumor = 519). (b) IGSF5 in HPV-positive HNSCC tissues from the GSE65858 cohort was upregulated compared with that in HPV-negative tissues. (c) The expression of IGSF5 was detected by Western blot, and cells transfected with the IGSF5 overexpression plasmid differed significantly between the *IGSF5* overexpression (pCDNA-IGSF5) and control (vector) groups in Hep-2 and TU212 cells. GAPDH was used as an internal control. (d) Reduction in the proliferation ability of pCDNA-IGSF5 Hep-2 and TU212 cells compared with the control (vector) cells by the CCK8 assay. (e, f) Reduction in colony formation ability of pCDNA-IGSF5 Hep-2 and TU212 cells compared with the control (vector) cells by a colony formation assay. The bar graph indicates the number of colonies. (g, h) Cell apoptosis of pCDNA-IGSF5 Hep-2 and TU212 cells compared with the control (vector) was analyzed by flow cytometry. The rate of cell apoptosis is shown in the graphs. The results are presented as the mean ± s.d. and are representative of at least three independent experiments. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$. ns: $p > 0.05$.

reached 0.787, 0.747, and 0.678, respectively, indicating that the prognostic evaluation results of the model were highly efficient. Our analysis results also demonstrated that there was a significant difference in survival prognosis in HNSCC patients with vs. without HPV infection. Sex, age, lymph node metastasis, tumor infiltration depth, smoking status, alcohol consumption, and *TP53* gene mutation were also closely correlated with the survival prognosis of HNSCC patients. Finally, we proved that IGSF5 was downregulated in HNSCC patients; overexpression of IGSF5 could inhibit

HNSCC cell proliferation and induce cell apoptosis, which means that our predicted model is reliable.

In the tumor microenvironment, immune and stromal cells are the two main nontumor components. Information on immune cells and stromal cells in the tumor microenvironment can be used as an important indicator in evaluating the survival prognosis of cancer patients [27]. Therefore, based on the constructed risk prognostic model, we divided HNSCC patients into the high-risk group and the low-risk group and found that the two groups had significant differences not only in tumor infiltration depth and lymph node metastasis but also in the degree of immune infiltration in tumor tissue. The survival prognosis was even poorer in the group with higher immune and stromal scores. On this basis, we found that the immune cell components of patients in these two groups were also different, mainly in naïve B cells, memory B cells, plasma cells, $CD8^+$ T cells, naive $CD4^+$ T cells, resting $CD4^+$ memory T cells, activated $CD4^+$ memory T cells, and regulatory T cells, in line with previous findings [28]. For example, $CD8^+$ T cells play an important role in the tumor microenvironment, which inhibits the proliferation and metastasis of tumor cells [29]. Our results indicate that $CD8^+$ T cells were highly expressed in low immune risk groups. These findings further confirm the importance of including more IRGs in survival prognosis models.

In summary, this study constructed a more comprehensive immune-associated survival prognostic evaluation system for HNSCC patients. The core genes of this system, *IGSF5*, *NKX2-3*, *HLF*, *ALDH2*, *IFIT2*, *FXYD5*, *CTSL1*, *IFNAR1*, and *RNF216*, can be used as markers of HNSCC. This system can help predict survival, immune infiltration, and tumor metastasis in HNSCC patients. It provides an important reference for understanding HNSCC and finding new targets for diagnosis and treatment in clinical practice. In future studies, the expression and function of the nine key genes need further validation.

6. Conclusion

We constructed a prognostic risk assessment model to help systematically evaluate the survival prognosis of HNSCC patients and provide a new research direction for the improvement of the survival prognosis of HNSCC patients in clinical practice.

Data Availability

The gene expression profile data and clinical information of HNSCC patients in the GSE65858 dataset were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The gene expression profile data and survival information of patients (494 cases) with HNSCC in The Cancer Genome Atlas (TCGA) database were obtained from the University of California, Santa Cruz, Xena website (https://xenabrowser.net/datapages/).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We thank all individuals who participated in this work. This work was supported by the Hunan Provincial Innovation Foundation for Postgraduate (No. CX20190411).

Supplementary Materials

Supplement Figure 1: the stability of this prognostic model was confirmed in HNSCC samples from TCGA. Supplement Table 1: number of genes corresponding to each module. Supplement Table 2: univariate and multivariate analyses of prognostic factors and overall survival of HNSCC patients in the TCGA cohort. *(Supplementary Materials)*

References

[1] A. Jou and J. Hess, "Epidemiology and molecular biology of head and neck cancer," *Oncology Research and Treatment*, vol. 40, no. 6, pp. 328–332, 2017.

[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. D. McDermott and D. W. Bowles, "Epidemiology of head and neck squamous cell carcinomas: impact on staging and prevention strategies," *Current Treatment Options in Oncology*, vol. 20, no. 5, p. ???, 2019.

[4] T. Sheedy and C. Heaton, "HPV-associated oropharyngeal cancer," *JAAPA.*, vol. 32, no. 9, pp. 26–31, 2019.

[5] C. Fakhry, A. L. Blackford, G. Neuner et al., "Association of oral human papillomavirus DNA persistence with cancer progression after primary treatment for oral cavity and oropharyngeal squamous cell carcinoma," *JAMA Oncology*, vol. 5, no. 7, pp. 985–992, 2019.

[6] G. Tolstonog and C. Simon, "Trends in surgical research in head and neck cancer," *Current Treatment Options in Oncology*, vol. 18, no. 6, p. 38, 2017.

[7] D. S. Chen and I. Mellman, "Elements of cancer immunity and the cancer-immune set point," *Nature*, vol. 541, no. 7637, pp. 321–330, 2017.

[8] A. Lechner, H. A. Schlößer, M. Thelen et al., "Tumor-associated B cells and humoral immune response in head and neck squamous cell carcinoma," *Oncoimmunology*, vol. 8, no. 3, article 1535293, 2018.

[9] W. Ma, F. Concha-Benavente, S. Santegoets et al., "EGFR signaling suppresses type 1 cytokine-induced T-cell attracting chemokine secretion in head and neck cancer," *PLoS One*, vol. 13, no. 9, article e0203402, 2018.

[10] L. A. Koneva, Y. Zhang, S. Virani et al., "HPV integration in HNSCC correlates with survival outcomes, immune response signatures, and candidate drivers," *Molecular Cancer Research*, vol. 16, no. 1, pp. 90–102, 2018.

[11] L. Wei, Z. Delin, Y. Kefei, W. Hong, H. Jiwei, and Z. Yange, "A classification based on tumor budding and immune score for patients with hepatocellular carcinoma," *Oncoimmunology.*, vol. 9, no. 1, article 1672495, 2020.

[12] Y. She, X. Kong, Y. Ge et al., "Immune-related gene signature for predicting the prognosis of head and neck squamous cell carcinoma," *Cancer Cell International*, vol. 20, no. 1, p. 22, 2020.

[13] J. Wang, L. G. Carvajal-Carmona, J. H. Chu et al., "Germline variants and advanced colorectal adenomas: adenoma prevention with celecoxib trial genome-wide association study," *Clinical Cancer Research*, vol. 19, no. 23, pp. 6430–6437, 2013.

[14] E. F. Robles, M. Mena-Varas, L. Barrio et al., "Homeobox NKX2-3 promotes marginal-zone lymphomagenesis by activating B-cell receptor signalling and shaping lymphocyte dynamics," *Nature Communications*, vol. 7, no. 1, p. 11889, 2016.

[15] D. M. Xiang, W. Sun, T. Zhou et al., "Oncofetal HLF transactivates c-Jun to promote hepatocellular carcinoma development and sorafenib resistance," *Gut*, vol. 68, no. 10, pp. 1858–1871, 2019.

[16] W. Seo, Y. Gao, Y. He et al., "ALDH2 deficiency promotes alcohol-associated liver cancer by activating oncogenic pathways via oxidized DNA-enriched extracellular vesicles," *Journal of Hepatology*, vol. 71, no. 5, pp. 1000–1011, 2019.

[17] H. Shen, M. Zhan, Y. Zhang et al., "PLZF inhibits proliferation and metastasis of gallbladder cancer by regulating IFIT2," *Cell Death & Disease*, vol. 9, no. 2, p. 71, 2018.

[18] R. A. Tassi, A. Gambino, L. Ardighieri et al., "FXYD5 (dysadherin) upregulation predicts shorter survival and reveals platinum resistance in high-grade serous ovarian cancer patients," *British Journal of Cancer*, vol. 121, no. 7, pp. 584–592, 2019.

[19] N. F. Ajeawung, R. Maltais, C. Jones, D. Poirier, and D. Kamnasaran, "Viability screen on pediatric low grade glioma cell lines unveils a novel anti-cancer drug of the steroid biosynthesis inhibitor family," *Cancer Letters*, vol. 330, no. 1, pp. 96–105, 2013.

[20] Y. Kawano, O. Zavidij, J. Park et al., "Blocking IFNAR1 inhibits multiple myeloma-driven Treg expansion and immunosuppression," *The Journal of Clinical Investigation*, vol. 128, no. 6, pp. 2487–2499, 2018.

[21] R. Guan, S. Cai, M. Sun, and M. Xu, "Upregulation of miR-520b promotes ovarian cancer growth," *Oncology Letters*, vol. 14, no. 3, pp. 3155–3161, 2017.

[22] S. Abiko, Y. Shimizu, S. Miyamoto et al., "Risk assessment of metachronous squamous cell carcinoma after endoscopic resection for esophageal carcinoma based on the genetic polymorphisms of alcoholdehydrogense-1B aldehyde dehydrogenase-2: temperance reduces the risk," *Journal of Gastroenterology*, vol. 53, no. 10, pp. 1120–1130, 2018.

[23] S. Ramakrishnan, V. Granger, M. Rak et al., "Inhibition of EZH2 induces NK cell-mediated differentiation and death in muscle-invasive bladder cancer," *Cell Death and Differentiation*, vol. 26, no. 10, pp. 2100–2114, 2019.

[24] R. M. L. Lapa, M. C. Barros-Filho, F. A. Marchi et al., "Integrated miRNA and mRNA expression analysis uncovers drug targets in laryngeal squamous cell carcinoma patients," *Oral Oncology*, vol. 93, pp. 76–84, 2019.

[25] M. Iigo, D. B. Alexander, J. Xu et al., "Inhibition of intestinal polyp growth by oral ingestion of bovine lactoferrin and immune cells in the large intestine," *Biometals*, vol. 27, no. 5, pp. 1017–1029, 2014.

[26] H. Ma, W. Yang, L. Zhang et al., "Interferon-alpha promotes immunosuppression through IFNAR1/STAT1 signalling in head and neck squamous cell carcinoma," *British Journal of Cancer*, vol. 120, no. 3, pp. 317–330, 2019.

[27] D. B. Rivadeneira, K. DePeaux, Y. Wang et al., "Oncolytic viruses engineered to enforce leptin expression reprogram tumor- infiltrating T cell metabolism and promote tumor clearance," *Immunity*, vol. 51, no. 3, pp. 548–560.e4, 2019, e4.

[28] X. Chen, B. Yan, H. Lou et al., "Immunological network analysis in HPV associated head and neck squamous cancer and implications for disease prognosis," *Molecular Immunology*, vol. 96, pp. 28–36, 2018.

[29] B. Huang, W. Han, Z. F. Sheng, and G. L. Shen, "Identification of immune-related biomarkers associated with tumorigenesis and prognosis in cutaneous melanoma patients," *Cancer Cell International*, vol. 20, no. 1, p. 195, 2020.