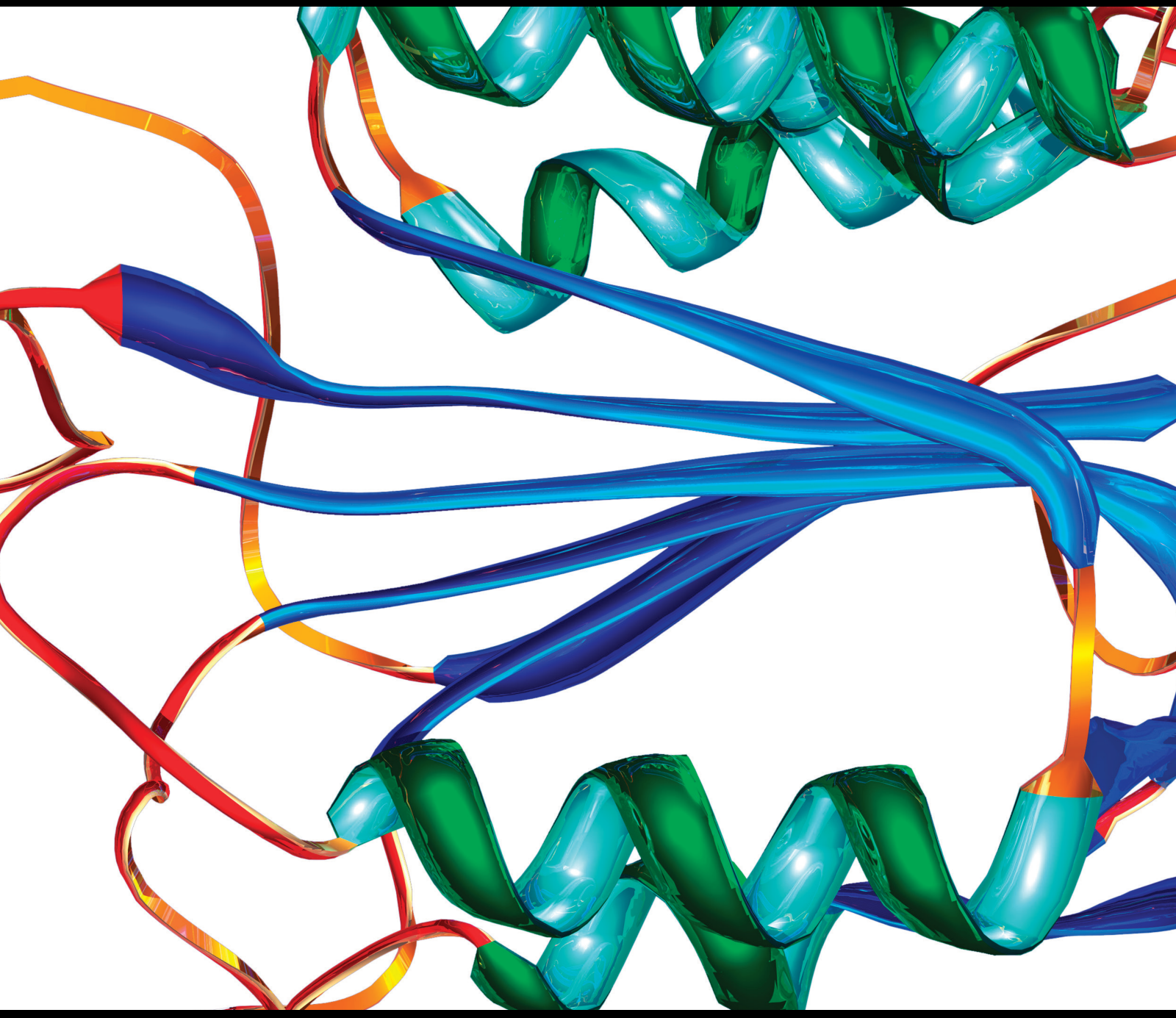
Disease Markers

High Resolution Omics for Biomarker Discovery

Lead Guest Editor: Xiang-Yu Meng
Guest Editors: Qiaoli Wang and Ming-Jun Shi



High Resolution Omics for Biomarker Discovery

Disease Markers

High Resolution Omics for Biomarker Discovery

Lead Guest Editor: Xiang-Yu Meng
Guest Editors: Qiaoli Wang and Ming-Jun Shi

Chief Editor

Paola Gazzaniga, Italy

Associate Editors

Donald H. Chace, USA
Mariann Harangi, Hungary
Hubertus Himmerich, United Kingdom
Yi-Chia Huang, Taiwan
Giuseppe Murdaca, Italy
Irene Rebelo, Portugal

Academic Editors

Muhammad Abdel Ghafar, Egypt
George Agrogiannis, Greece
Mojgan Alaeddini, Iran
Atif Ali Hashmi, Pakistan
Cornelia Amalinei, Romania
Pasquale Ambrosino, Italy
Paul Ashwood, USA
Faryal Mehwish Awan, Pakistan
Atif Baig, Malaysia
Valeria Barresi, Italy
Lalit Batra, USA
Francesca Belardinilli, Italy
Elisa Belluzzi, Italy
Laura Bergantini, Italy
Sourav Bhattacharya, USA
Anna Birková, Slovakia
Giulia Bivona, Italy
Luisella Bocchio-Chiavetto, Italy
Francesco Paolo Busardó, Italy
Andrea Cabrera-Pastor, Spain
Paolo Cameli, Italy
Chiara Caselli, Italy
Jin Chai, China
Qixing Chen, China
Shaoqiu Chen, USA
Xiangmei Chen, China
Carlo Chiarla, Italy
Marcello Ciaccio, Italy
Luciano Colangelo, Italy
Alexandru Corlateanu, Moldova
Miriana D'Alessandro, Saint Vincent and the Grenadines
Waaqo B. Daddacha, USA
Xi-jian Dai, China
Maria Dalamaga, Greece
Serena Del Turco, Italy
Jiang Du, USA
Xing Du, China
Benoit Dugue, France
Paulina Dumnicka, Poland
Nashwa El-Khazragy, Egypt
Zhe Fan, China
Rudy Foddis, Italy
Serena Fragiotta, Italy
Helge Frieling, Germany
Alain J. Gelibter, Italy
Matteo Giulietti, Italy
Damjan Glavač, Slovenia
Alvaro González, Spain
Rohit Gundamaraju, USA
Emilia Hadziyannis, Greece
Michael Hawkes, Canada
Shih-Ping Hsu, Taiwan
Menghao Huang, USA
Shu-Hong Huang, China
Xuan Huang, China
Ding-Sheng Jiang, China
Esteban Jorge Galarza, Mexico
Mohamed Gomaa Kamel, Japan
Michalis V. Karamouzis, Greece
Muhammad Babar Khawar, Pakistan
Young-Kug Kim, Republic of Korea
Mallikarjuna Korivi, China
Arun Kumar, India
Jinan Li, USA
Peng-fei Li, China
Yiping Li, China
Michael Lichtenauer, Austria
Daniela Ligi, Italy
Hui Liu, China
Jin-Hui Liu, China
Ying Liu, USA
Zhengwen Liu, China
César López-Camarillo, Mexico
Xin Luo, USA
Zhiwen Luo, China
Valentina Magri, Italy
Michele Malaguarnera, Italy
Erminia Manfrin, Italy
Upender Manne, USA

Contents

Hindawi
Disease Markers
Volume 2022, Article ID 4872001, 13 pages
https://doi.org/10.1155/2022/4872001

Research Article

lncRNAs AC156455.1 and AC104532.2 as Biomarkers for Diagnosis and Prognosis in Colorectal Cancer

Fei Kuang,[1] **De Luo**,[2] **Mengjia Zhou**,[3] **Juan Du**,[2] **and Jie Yang**[4]

[1]*Department of General Surgery, Changhai Hospital of the Second Military Medical University, Shanghai 200433, China*
[2]*Department of Hepatobiliary Surgery, Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan 646000, China*
[3]*Department of Ultrasound, Seventh People's Hospital of Shanghai University of Traditional Chinese Medicine, Shanghai 200137, China*
[4]*Department of Radiation Oncology, Peking University Third Hospital, Beijing 100191, China*

Correspondence should be addressed to Jie Yang; yang.jie@whu.edu.cn

Fei Kuang, De Luo, and Mengjia Zhou contributed equally to this work.

Received 13 August 2022; Revised 14 September 2022; Accepted 21 September 2022; Published 13 October 2022

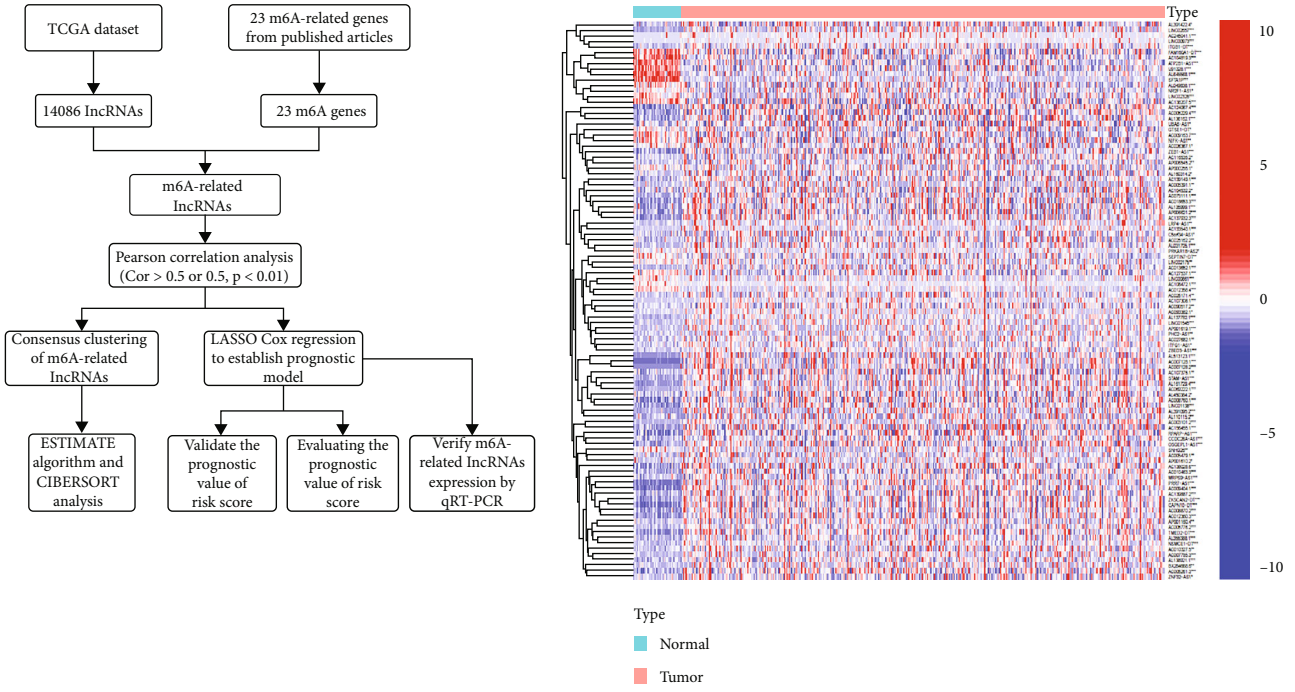
Academic Editor: Ming-Jun Shi

Background. There have been countless studies to date assessing specific oncogenic pathways in a range of tumor classes, but the role of N6-methyladenosine- (m6A-) related long noncoding RNAs (lncRNAs) in colorectal cancer (CRC) remains to be defined. *Methods*. We analyzed such m6A-related lncRNAs by conducting analyses of the Pearson correlation with information originating from the databank of The Cancer Genome Atlas (TCGA). The prognostic relevance of these lncRNAs in CRC was then assessed through a series of univariate Cox regression analyses, leading to the identification of two different m6A modification patterns; they are associated with clinical outcomes and have been used to estimate tumor immune microenvironment (TIME) by the CIBERSORT and ESTIMATE algorithms. We tested the expression of m6A-related lncRNAs in twelve pairs of colorectal cancer tissues and adjacent normal tissues from patients by qRT-PCR. *Results*. We discovered the prognostic risk signature composed of six m6A-related lncRNAs based upon TCGA data. When the overall survival of cases in the dataset of TCGA was investigated, the low-risk cases survived longer than the high-risk CRC cases in both the training and testing cohorts. ROC curves further indicated that m6A-related lncRNA prognostic signature (m6A-LPS) can effectively estimate the survival outcomes of patients in both of these cohorts. We found that lncRNAs AC156455.1 and AC104532.2 were upregulated in twelve colorectal cancer tissues compared with adjacent normal tissues using qRT-PCR. *Conclusions*. This data highlights that the lncRNAs AC156455.1 and AC104532.2 in CRC can be used as biomarkers for diagnostics and prognosis in CRC, demonstrating their potential as targets when designing novel immunotherapeutic regimens.

1. Introduction

Colorectal cancer (CRC) is well known as one of the most widespread and deadly cancers all around the globe, with roughly 400,000 and 212,000 new diagnoses and deaths annually, respectively [1]. Novel immunotherapy-based regimens developed in recent years can engage or enhance natural immunological pathways and cells within a treated host to aid in tumor cell clearance, and many of these immunotherapies have been effective when applied in combination with other treatments in individuals with a range of tumor classes [2–4]. However, immunotherapy is commonly beneficial to a poorly defined subset of cases, and therefore, novel strategies must be devised to determine which CRC patients are likely to respond to such treatment [2–5].

The most commonly found epigenetic modification on mRNAs and noncoding RNAs (ncRNAs) is N6-methyladenosine (m6A), which can alter the stability, translation, and splicing of modified RNAs in biologically important contexts [6, 7]. As a reversible and dynamic process, the modification of m6A is controlled through three primary classes of m6A-regulating proteins: "writers" (methyltransferases),



(a) (b)

Figure 1: m6A-related lncRNA identification in patients with CRC. (a) Research flow chart. (b) Gene expression heatmap for 101 prognostic m6A-related lncRNAs in pairs of tumor and paracancerous tissues. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

"readers" (signal transducers), and "erasers" (demethylases) [8]. According to the latest studies, m6A modification has been identified as a mechanism capable of modulating oncogenesis in a range of tumor types. For example, METTL14 can drive cancer advancement and maintenance in acute myeloid leukemia by enhancing the self-renewal of leukemia stem cells [9], while knocking down FTO can compromise lung squamous cell carcinoma cell proliferative and invasive activity and the self-renewal of glioblastoma stem cells [10, 11]. Furthermore, YTHDF2 can decrease the EGFR mRNA stability in cancer cells of the liver, thereby compromising their proliferation [12].

Recent research has underscored the relevance of ncRNAs in the context of CRC both through the therapeutic delivery of small interfering RNAs (siRNAs) as well as through in-depth analyses of the functional importance of long ncRNAs (lncRNAs) [13–18]. How m6A-related lncRNAs function in the regulation of CRC onset and progression, however, has yet to be defined, and there have been few studies exploring the impact of m6A modification on lncRNA-mediated CRC development. By studying the association between m6A modifications and lncRNAs in this tumorigenesis-related setting, it may be possible to define novel biomarkers that can guide therapeutic targeting efforts.

Herein, we leveraged the dataset of The Cancer Genome Atlas (TCGA) and conducted a series of bioinformatics analyses to clarify the prognostic relevance of m6A-related lncRNAs in CRC cases. The overall aim of the current investigation is to conduct a systematic assessment of the relationship among m6A-related lncRNAs and CRC patient prognosis, the composition of tumor immune microenvironment (TIME), and expression of programmed death ligand 1 (PD-L1). Through clustering analyses and risk modeling, we were able to initiate an m6A-related lncRNA prognostic signature (m6A-LPS). Associations among clustering subgroups, PD-L1 status, immune scores, and immune cell infiltration were analyzed in light of m6A-LPS as a means of further understanding the association between this risk signature. We analyzed and confirmed m6A-related lncRNAs AC156455.1 and AC104532.2 as biomarkers for diagnostics and prognosis in CRC, which will provide a new foundation for future efforts to develop effective immunotherapeutic treatments for CRC.

2. Materials and Methods

2.1. Data Selection. Transcriptomic and clinical outcomes from cases with CRC were acquired from the database of TCGA (https://portal.gdc.cancer.gov/). Initially, 23 regulators of m6A RNA methylation were selected based upon prior reports [19, 20], including 8 writers (RBM15, VIRMA, METTL3, WTAP, METTL14, RBM15B, METTL16, and ZC3H13), 2 erasers (FTO and ALKBH5), and 13 readers (IGFBP2, YTHDC1-2, LRPPRC, YTHDF1-3, FMR1, HNRNPA2B1, IGF2BP1, IGFBP3, HNRNPC, and RBMX).

2.2. lncRNA Annotation. The GRCh38 lncRNA annotation file was downloaded from GENCODE to aid in TCGA data annotation. In total, 14,086 lncRNAs were identified in this dataset based upon Ensemble IDs.

2.3. Bioinformatics Analysis. Initially, m6A-related lncRNAs were identified in each dataset via a series of Pearson's

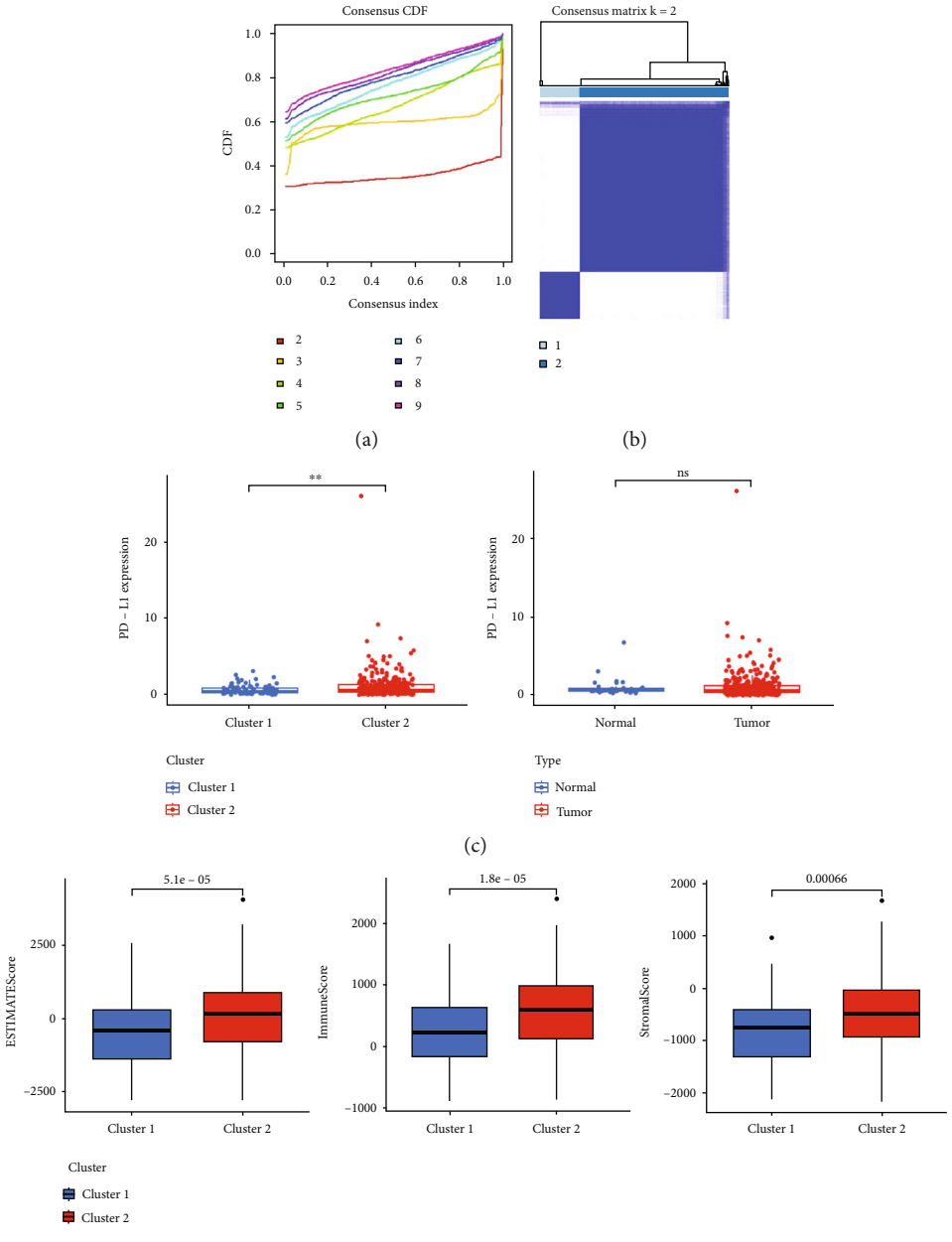
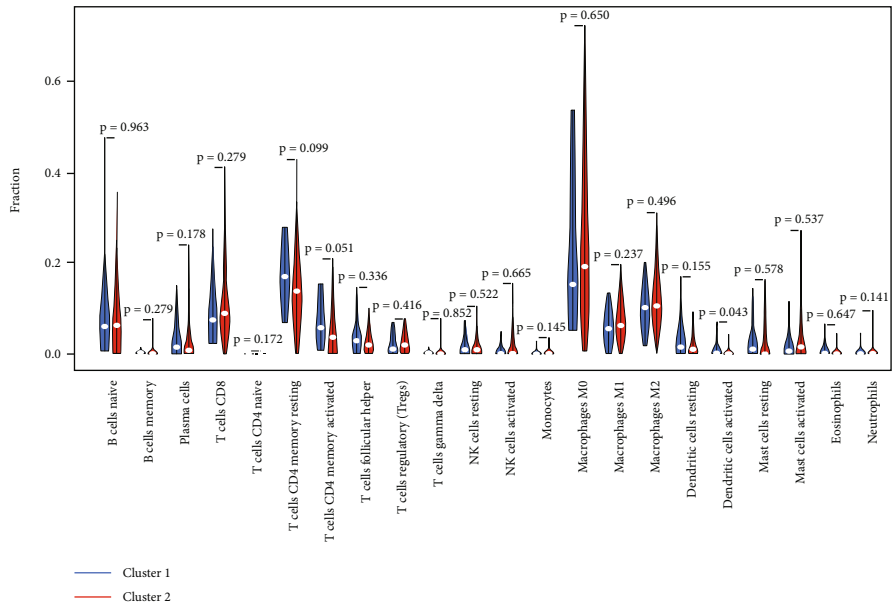


Figure 2: Continued.



(e)

Figure 2: m6A-related lncRNA identification and evaluation of the CRC-related intratumoral immune cell landscape. (a) Consensus clustering matrix at $k = 2$. (b) Consensus clustering cumulative distribution functions (CDFs) and relative area under CDF curve from k values of 2 through 9. (c) PD-L1 expression in normal/tumor samples and in the cluster1/2 subtypes. (d) ESTIMATE, immune, and stromal scores in the cluster 1/2 subtypes. (e) Levels of predicted the infiltration of immune cells for 22 different immune cell subtypes in the cluster 1/2 subtypes. $^{*}p < 0.05$ and $^{**}p < 0.01$.

correlation analyses ($|\text{Pearson's R}| > 0.5$ and $p < 0.001$). Prognostic m6A-related lncRNAs were subsequently identified through univariate Cox regression analyses, with 101 shared m6A-related prognostic lncRNAs being identified through comparison of overlap among TCGA datasets.

CRC patients were classified into two cohorts via a k-means clustering approach based upon the expression of 23 m6A-modulating genes using the R "kmeans function" using the ConsensusClusterPlus package, with 1,000 computational permutations being performed to guarantee stability and reliability [21].

The ESTIMATE algorithm was implemented to compute immune scores with the "estimate" R package [22]. The CIBERSORT algorithm was further employed to approximate the levels of intratumor infiltration via 22 various immune cell populations according to the expression data of RNA (https://cibersort.stanford.edu/), with 1,000 permutations of this analysis being performed and samples with a CIBERSORT $p < 0.05$ being retained for comparisons of differential immune cell infiltration among CRC patient subgroups defined according to risk scores and clustering subtypes.

The "h.all.V6.2.symbols.gmt" hallmark gene set from MSigDB was employed for a GSEA approach conducted using the JACA program to compare differences among CRC patient subtypes with respect to survival outcomes. For this analysis, 1,000 random sampling permutations were employed. Gene set enrichment was described according to the false discovery rate (FDR) < 0.05 and NES.

An analysis of LASSO regression was performed within the TCGA training cohort to define m6A-related lncRNA-based prognostic risk signatures [23], with the most appropriate signature being selected via choosing the optimal penalty criterion (l) associated with the minimum 10-fold cross-validation. LASSO regression algorithm-derived coefficients were employed to develop a risk score model with the following general equation: *risk score* = *sum* of *coef f icients* × *e xpression level* of *m6A regulator*. Risk scores were individually computed for each case in the training and testing cohorts, and cases were stratified into low- and high-risk groups, with median risk score values serving as the cutoff for patient stratification. Comparisons of patient outcomes were then made through Kaplan-Meier survival curves, while the sensitivity and specificity of this prognostic model were established through the use of receiver operating characteristic (ROC) curves.

Associations between immune cell infiltration and m6A-related lncRNAs were further assessed by utilizing the Tumor Immune Estimation Resource (TIMER) tool (https://cistrome.shinyapps.io/timer/), which assessed four types of immune cells (activated mast cells, memory B cells, T follicular helper cells, and resting memory CD4+ T cells). GISTIC 2.0 outcomes were implemented in the analyses of TIMER.

2.4. RNA Isolation and qRT-PCR. We collected twelve pairs of colorectal cancer tissues and adjacent normal tissues from patients who recently underwent surgical treatment in the Department of Gastrointestinal Surgery, Shanghai Changhai Hospital. Fresh tissues were frozen and stored at −80°C. This research was approved by the Medical Ethics Committee of

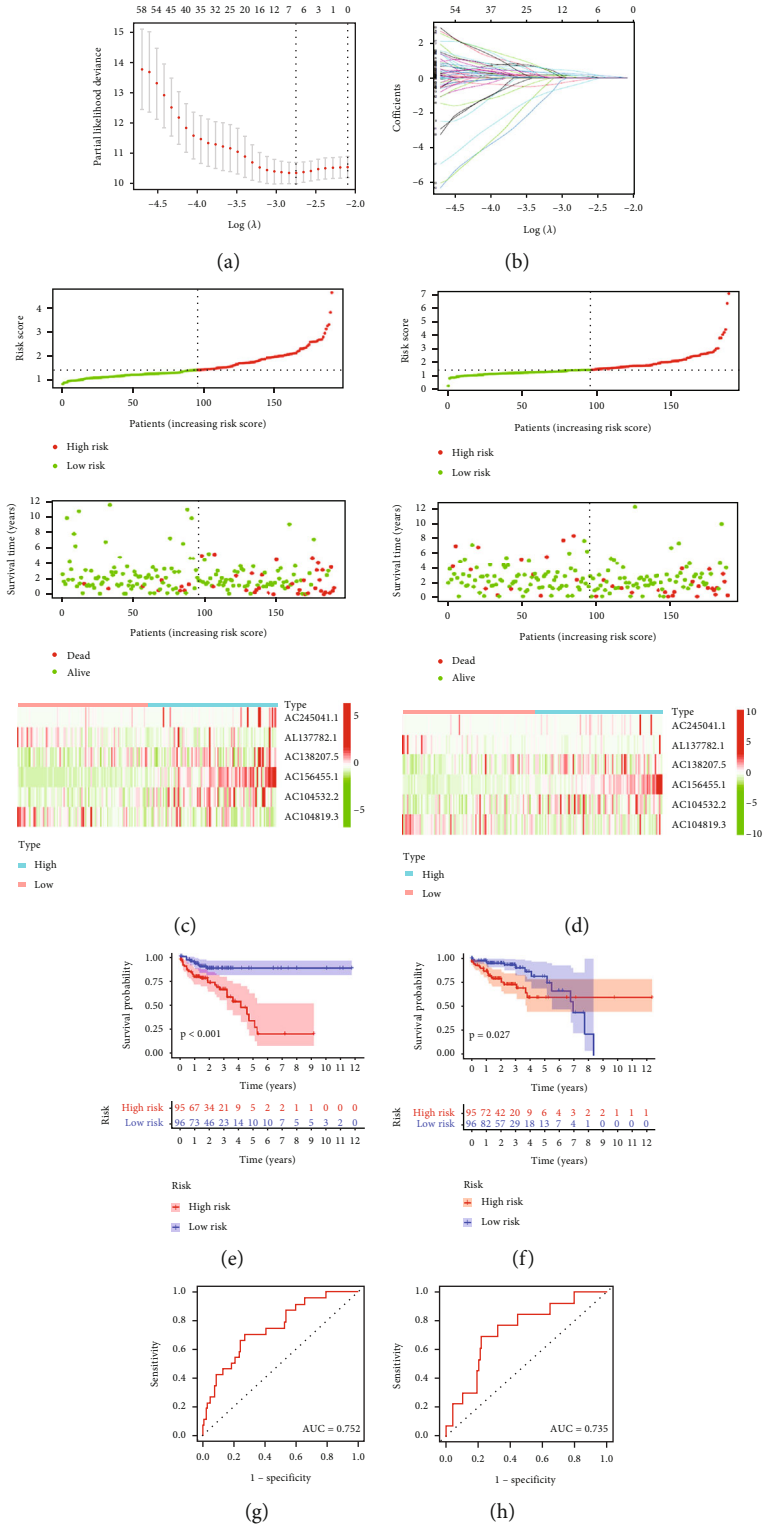
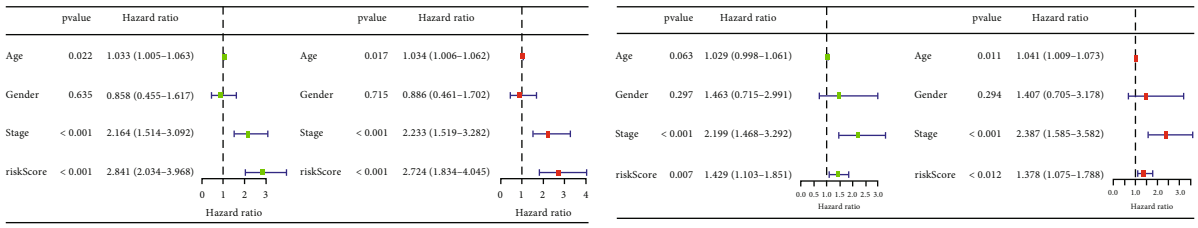


Figure 3: Continued.



(i) (j)

FIGURE 3: Validation and construction of a prognostic m6A-related lncRNA risk model. (a, b) Analysis of LASSO Cox regression for m6A-related lncRNAs. (c, d) Risk score, OS, and survival status distributions together with heatmaps demonstrating the expression of 6 prognostic m6A-Related lncRNAs in the training (c) and testing (d) TCGA cohorts. (e, f) CRC patient OS outcomes were compared based upon risk scores in the training ($p < 0.001$) and testing ($p = 0.027$) TCGA cohorts. (g, h) Risk score ROC curves for the TCGA training (g) and testing (h) cohorts. (i, j) Analyses of univariate and multivariate Cox regression for risk scores and other clinicopathological properties identifying independent prognostic indicators in the training (i) and testing (j) TCGA cohorts.

Changhai Hospital of the Second Military Medical University. Informed consent was acquired from each involved patient. Total RNA from tissues of CRC patients was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. For complementary DNA (cDNA) synthesis, 1 μg of total RNA and the PrimeScript RT reagent kit (Takara, Otsu, Shiga, Japan) were utilized. The SYBR Green assay (Takara) was used to perform qRT-PCR, and the progression was executed on a CFX-96 system (Bio-Rad Laboratories, Inc., Hercules, CA, USA). The GAPDH was used as an internal reference, and the relative lncRNA expression was calculated using the $2^{-\Delta\Delta Cq}$ method. Primer sequences for qRT-PCR used in this study are shown in Supplementary Table S1.

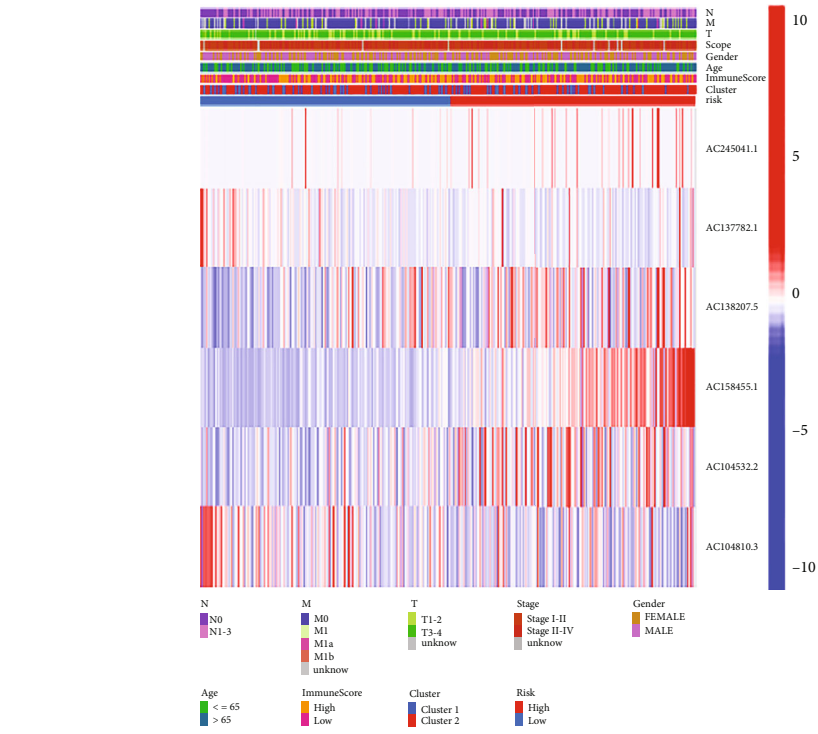
2.5. Statistical Studies. GraphPad Prism 8.0, R v 3.60, and SPSS 24.0 (IBM, NY, USA) were employed for statistical assessments. Mann-Whitney *U* tests were employed to compare lncRNA expression in normal and tumor tissue samples. Data in different subgroups or groups were thoroughly evaluated via Student's *t*-test and one-way ANOVAs. Chi-square experiments were employed to assess categorical variables. The curves of Kaplan-Meier and log-rank assessments were utilized to compare survival outcomes. Pearson correlation analyses were used to explore the relationships among risk scores, PD-L1 status, levels of infiltration of immune cells, clinicopathological factors, and subtypes. The independent prognostic relevance of the scores of risk and other clinical traits was analyzed through the analyses of the multivariate and univariate Cox regression. ROC curves were implemented to appraise the predictive efficiency of m6A-related lncRNA signatures when estimating CRC patient OS. $p < 0.05$ was the significance threshold for the current research.

3. Results

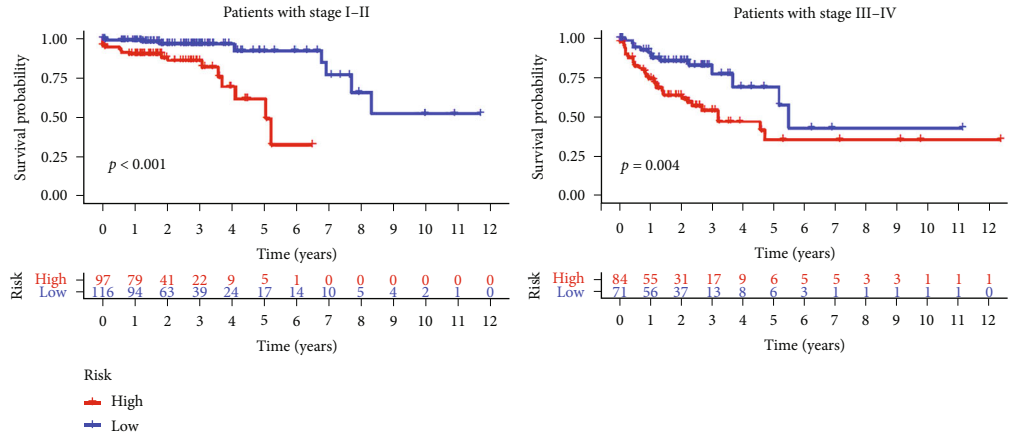
3.1. m6A-Related lncRNA Identification in Patients with CRC. We began by identifying and evaluating 14,086 lncRNAs present within the selected TCGA dataset. Initially, expression matrices for 23 m6A-associated genes were downloaded from the TCGA database, and those lncRNAs that correlated with the expression of one or more of these genes were defined as m6A-related lncRNAs ($|\text{Pearson R}| > 0.5$ and $p < 0.001$). Clustering analysis was conducted to separate cases in the TCGA-CRC cohort into different groups on the basis of their expression of m6A-related lncRNAs. The prognostic relevance of these m6A-related lncRNAs was then further evaluated through a series of univariate Cox regression analyses based upon a $p < 0.05$ cutoff within the analyzed TCGA datasets. A LASSO Cox analysis of the resultant 101 m6A-related prognostic lncRNAs was identified via this approach, with the overall workflow being detailed in Figure 1(a). Patterns of prognostic m6A-related lncRNA expression in CRC and normal tissues are shown in Figure 1(b).

3.2. m6A-Related lncRNA Identification and Evaluation of the CRC-Related Intratumoral Immune Cell Landscape. For this analysis, cumulative distribution functions (CDFs) were generated for consensus clusters for *k* values from 2-9 (Figure 2(a)), with the maximal area under the curve (AUC) value for this CDF function being evident at $k = 2$, at which time there was a clear difference in the expression of m6A-related lncRNAs between the two defined clusters (Figure 2(b)). The associations among m6A-related lncRNAs and PD-L1 expression were next assessed, revealing that the expression of PD-L1 was considerably greater in cluster 2 relative to cluster 1; there was also a trend toward increased PD-L1 expression in CRC tumor tissues relative to vicinal normal tissues (Figure 2(c)). Next, the algorithm of ESTIMATE was employed to measure stromal and immune scores for CRC case and tumor samples; these immune scores differed significantly between clusters, with higher immune ESTIMATE and stromal scores in cluster 2 patients relative to those in cluster 1 (Figure 2(d)). The CIBERSORT algorithm was utilized to evaluate the discrepancies in the levels of 22 populations of immune cells in these CRC tumors, revealing that there were high levels of M0 macrophages, CD8+ T cells, naïve B cells, and resting memory CD4+ T cells in cluster 2 patient tumors (Figure 2(e)).

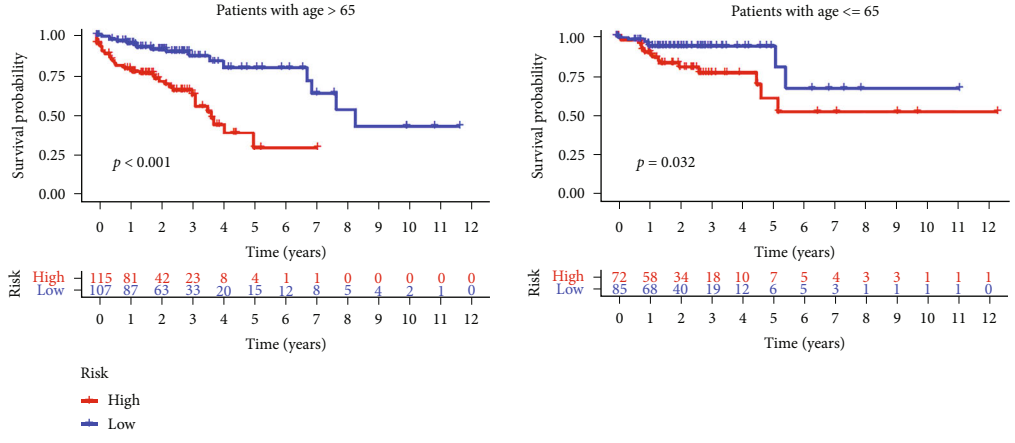
3.3. Validation and Construction of a Prognostic m6A-Related lncRNA Risk Model. Risk scores were then measured based the regression coefficients derived from the LASSO algorithm for cases in both the TCGA testing and training



(a)

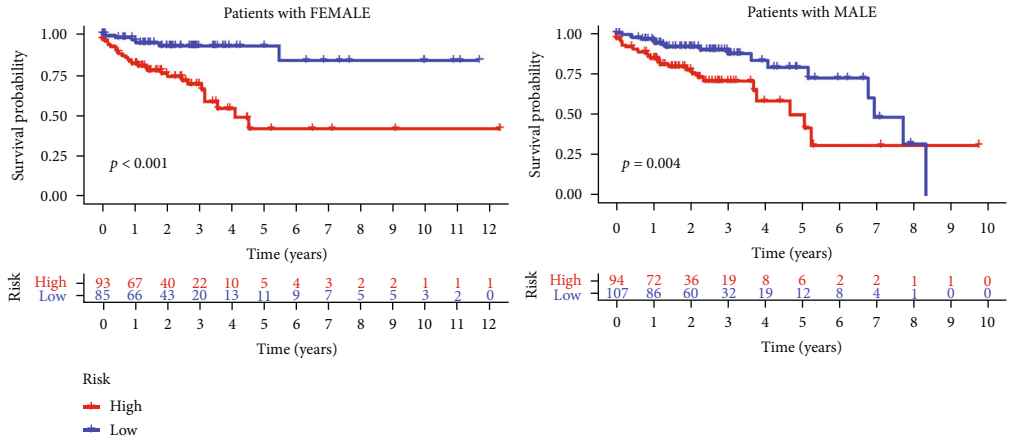


(b)

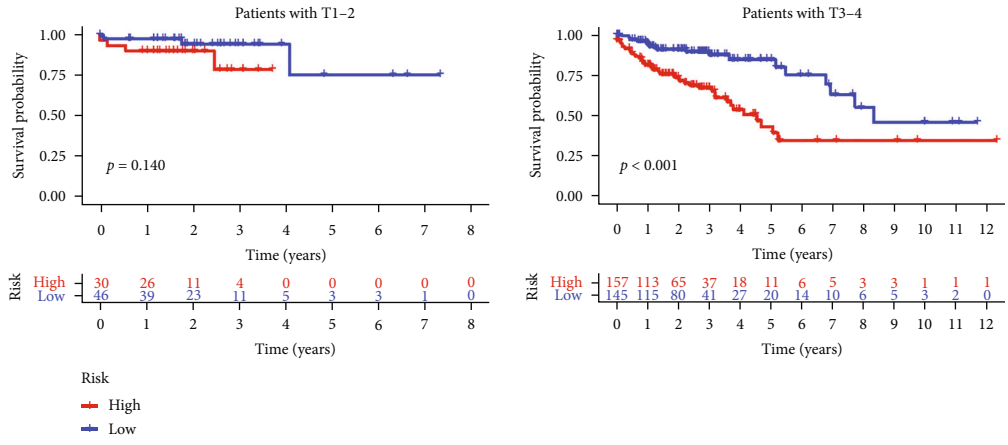


(c)

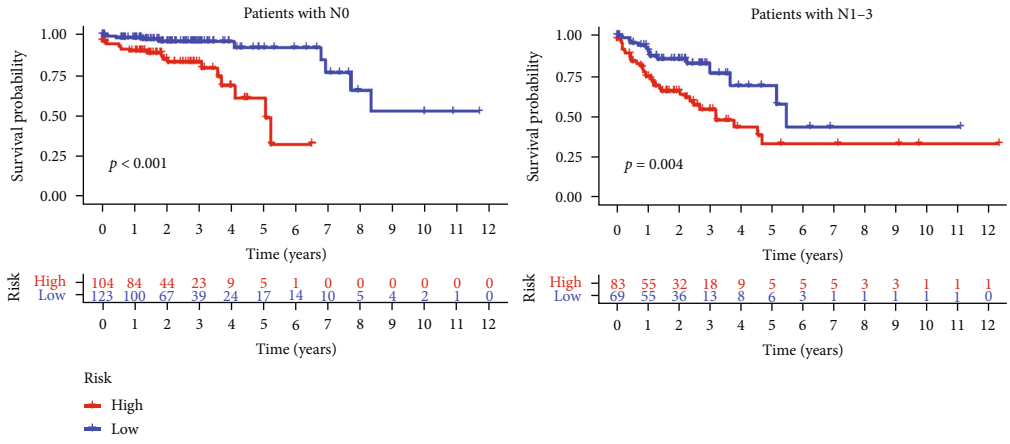
FIGURE 4: Continued.



(d)

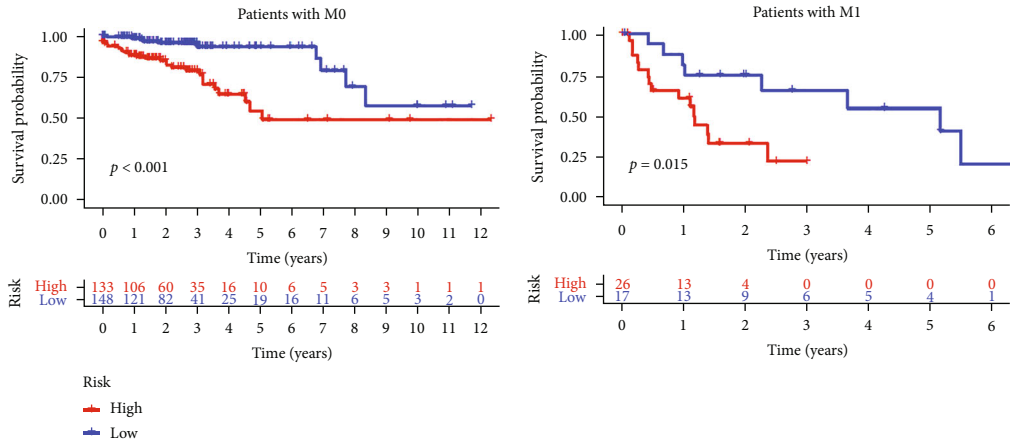


(e)



(f)

Figure 4: Continued.



(g)

Figure 4: Assessment of the prognostic utility of risk scores in different CRC patient subgroups. (a) Low- and high-risk patient group clinicopathological findings and heatmaps. (b) Comparison of the survival of low- and high-risk cases in the dataset of TCGA (*cases* > 65 years old: $p < 0.001$ and *cases* ≤ 65 years old: $p = 0.032$). (c) Comparison of the survival of low- and high-risk cases in the dataset of TCGA (males: $p = 0.004$ and females: $p < 0.001$). (d) Comparison of the survival of low- and high-risk cases in the dataset of TCGA (cases with stage I-II disease: $p < 0.001$ and cases with stage III-IV disease: $p = 0.004$). (e) Comparison of the survival of low- and high-risk cases in the dataset of TCGA (cases with T1-2 disease: $p = 0.140$ and cases with T3-4 disease: $p < 0.001$). (f) Comparison of the survival of low- and high-risk cases in the dataset of TCGA (cases with N0 disease: $p < 0.001$ and cases with N1-3 disease: $p = 0.004$). (g) Comparison of the survival of low- and high-risk cases in the dataset of TCGA (M0 cases: $p < 0.001$ and cases with M1: $p = 0.015$). All survival outcomes were compared through curves of Kaplan-Meier and log-rank experiments.

cohorts (Figures 3(a) and 3(b)). Median risk score values were subsequently used to separate cases into high- and low-risk groups, and the patterns of OS and the expression of the six m6A-related lncRNAs composing the risk score were next assessed (Figures 3(c) and 3(d)). We found that low-risk CRC cases exhibited a longer OS relative to high-risk cases in either of the training and testing cohorts (Figures 3(e) and 3(f)). ROC curves further indicated that the developed m6A-LPS was able to reliably predict the OS of cases in both cohorts (Figures 3(g) and 3(h)). Univariate and multivariate analyses were then conducted, which confirmed that stage, age, and risk score values were all independent predictors of patient outcomes within the TCGA testing and training cohorts (Figures 3(i) and 3(j)).

3.4. Assessment of the Prognostic Utility of Risk Scores in Different CRC Patient Subgroups. Next, we examined the relationship between risk scores and CRC patient clinical features. Heatmaps were used to evaluate the patterns of expression of the six m6A-related lncRNAs in low- and high-risk cases (Figure 4(a)). This revealed that AL137782.1 and AC104819.3 were expressed at lower levels within the high-risk group relative to the low-risk group, whereas AC245041.1, AC138207.5, AC156455.1, and AC104532.2 exhibited the opposite trend. To more fully understand the prognostic values of these risk scores, we stratified CRC cases based upon their disease status and found that compared to low-risk cases, those in the high-risk group exhibited worse OS for both individuals with stage I-II and stage III-IV disease (Figure 4(b)). Similarly, these prognostic m6A-related lncRNAs were also able to estimate the OS of cases irrespective of age (>65 vs. ≤65 years) (Figure 4(c)), gender (female vs. male) (Figure 4(d)), T status (T1-2 vs. T3-4) (Figure 4(e)), N status (N0 vs. N1-3) (Figure 4(f)), and M status (M0 vs. M1) (Figure 4(g)).

3.5. Association between Risk Scores and Immune Cell Infiltration. We explored the associations between risk scores and intratumoral infiltration by four different immune cell types to meticulously discover the influence of the 6 m6A-related lncRNAs composing our risk signature and the TIME in CRC. Risk scores are considerably negatively related to infiltration by resting memory CD4+ T cells ($p = 0.023$) and activated mast cells (Figures 5(a) and 5(b)), whereas they are positively correlated with infiltration by memory B cells ($p = 0.032$) and T follicular helper cells (Figures 5(c) and 5(d)).

3.6. Validation of the Expression Levels of lncRNAs AC156455.1 and AC104532.2 with Prognostic Signature. To further verify the accuracy of the m6A-related lncRNA signature, the expression levels of lncRNAs AC156455.1 and AC104532.2 were measured in twelve colorectal cancer tissues and twelve adjacent normal tissues using qRT-PCR. lncRNAs AC156455.1 and AC104532.2 were upregulated in colorectal cancer tissues compared with corresponding normal tissues (Figures 6(a) and 6(b)). Meanwhile, the same results were analyzed; lncRNAs AC156455.1 and AC104532.2 were also upregulated in colorectal cancer tissues compared with corresponding normal tissues from TCGA database (Figures 6(c) and 6(d)).

4. Discussion

Several prior reports have explored the link between m6A modifications and the regulation of cancer pathogenesis

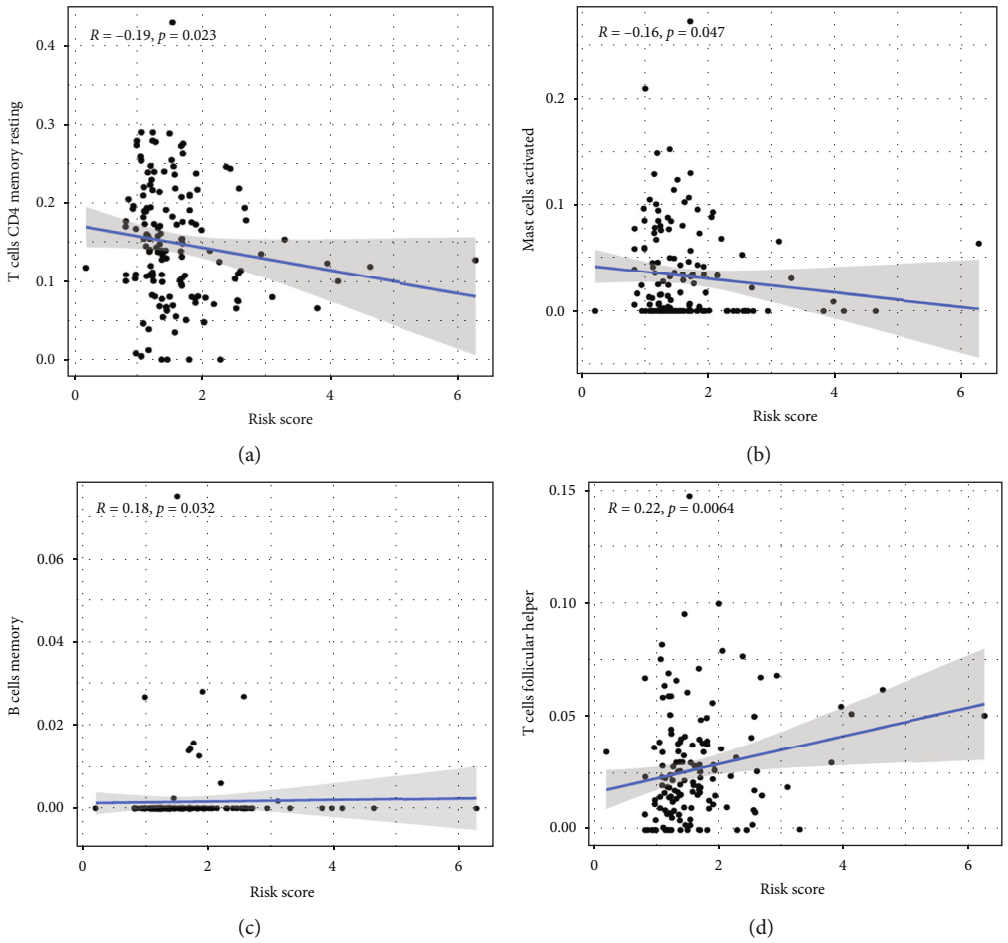


Figure 5: Associations between risk scores and immune cell infiltration. (a–d) Memory B cells (a), activated mast cells (b), resting CD4 memory T cells (c), and T follicular helper cells (d).

[24–28], but the mechanisms whereby lncRNAs may shape this relationship are yet to be defined. KIAA1429 has been indicated to drive the progression of liver cancer through the m6A modification of the lncRNA GATA3 [29]. In glioblastoma stem cells, the lncRNA FOXM1-AS has been illustrated to influence interactions among FOXM1 and ALKBH5 to shape cell maintenance [24]. In light of the above results, we speculate that m6A modifications of specific mRNAs may shape oncogenesis, and as such, further study of the impact of such m6A modifications on lncRNA function is warranted to better identify key therapeutic targets or prognostic biomarkers associated with particular cancers. LINC00265 has been shown to predict undesirable findings in cases with AML [30], while LINC00665 has been associated with enhanced activation of the pathway of PKR/NF-κB hepatocellular carcinoma and with concomitant increases in malignancy [31], while in gastric cancer, this same lncRNA can activate the Wnt pathway to promote tumor progression [32]. In our study, we explored the prognostic relevance of m6A-related lncRNAs by analyzing data from 437 CRC patients in the TCGA database. We ultimately defined 101 prognostically relevant m6A-related lncRNAs, of which 6 were employed to establish an m6A-related lncRNA prognostic signature (m6A-LPS) capable of estimating the OS of patients with CRC. When stratified into low- and high-risk groups following the median risk score values, high-risk CRC patients survived for significantly shorter periods relative to low-risk patients. Multivariate analyses further indicated that these m6A-LPS values were independent predictors of CRC patient OS. While several of the lncRNAs within our risk signature have been studied in oncogenic contexts, they have not been analyzed in the context of CRC, and there have been few reports regarding interactions between these lncRNAs and m6A-associated genes. As such, our findings offer novel insights regarding lncRNAs targeted by m6A regulators in the context of CRC, potentially shedding new light on their ability to promote CRC onset and progression.

The heterogeneous tumor microenvironment (TME) often harbors a diverse array of immunosuppressive signals, shaping tumor development, patient prognosis, and therapeutic responses [33–36]. The TME consists of an assorted immune cells, vascular structures, and stromal cells, all of which can impact the oncogenic progression associated with a given tumor type. Immune cell infiltration within the TME can predict patient outcomes and is often correlated with tumor grade, stage, and metastasis [37, 38]. For example, tumor-associated macrophages (TAMs) can generate

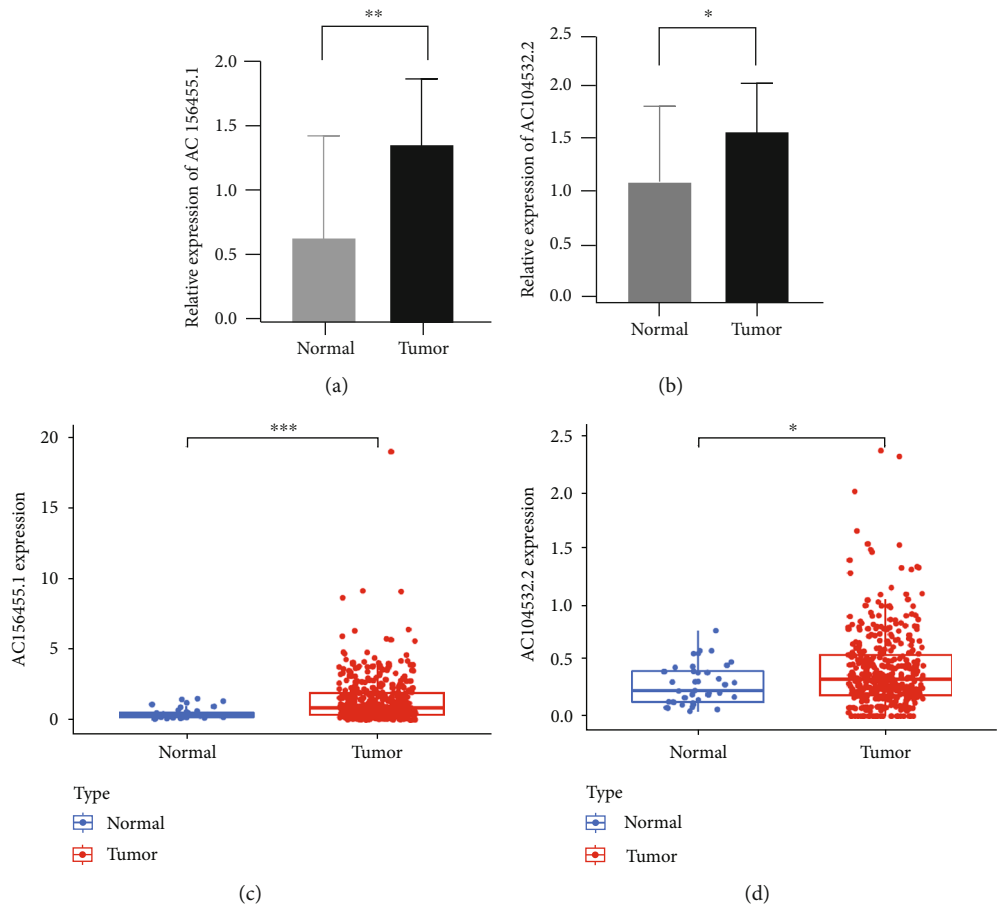


Figure 6: Validation of the expression Levels of lncRNAs AC156455.1 and AC104532.2 of prognostic signature. (a, b) AC156455.1 and AC104532.2 expressions in twelve colorectal cancer tissues and twelve adjacent normal tissues using RT-qPCR. (c, d) AC156455.1 and AC104532.2 expressions in colorectal cancer tissues compared with corresponding normal tissues from TCGA database. $^{*}p < 0.05$ and $^{**}p < 0.01$.

immunosuppressive cytokines including TGF-B and IL-10 for example, which can drive preferential tumor outgrowth and contribute to poor patient outcomes [39–41]. More potent tumor infiltration by CD4+ and CD8+ T cells, on the contrary, is often related to better patient survival and a higher response rate to immunotherapy [42]. We observed that PD-L1 expression levels in cluster 2 were considerably greater relative to cluster 1, and a trend towards increased PD-L1 expression in CRC tissues relative to normal tissues. It is critical that consensus criteria be established in order to determine the CRC cases that are most probable to respond to immunotherapeutic treatment. We found that the ESTIMATE, stromal, and immune scores of cluster 2 were greater than those in cluster 1. This strongly suggests a close relationship between patterns of m6A-related lncRNA expression and the ability of particular immune cells to enter or persist within the TIME, thereby altering patient responses to immunotherapeutic intervention. Risk scores were all negatively associated with activated mast cell and resting memory CD4+ T cell infiltration in this study, whereas they were positively associated with memory B cell and T follicular helper cell infiltration. We reported that lncRNAs AC156455.1 and AC104532.2 were upregulated in colorectal cancer tissues compared with corresponding normal tissues using qRT-PCR, compared with previous studies. Therefore, lncRNAs AC156455.1 and AC104532.2 can be used as biomarkers for diagnostic and prognosis in CRC, to provide new targets for future immunotherapy.

5. Conclusions

In summary, we herein conducted a systematic assessment of the prognostic relevance of m6A-associated lncRNAs in CRC and explored their associations with PD-L1 expression and the TIME. Risk scores derived from m6A-associated lncRNA-based expression signatures were also found to be independently related to CRC patient prognosis, and further predictive analyses suggested that these lncRNAs may be associated with the regulation of the TIME in CRC tumors. As such, lncRNAs AC156455.1 and AC104532.2 associated with tumor immune responses have the potential to guide the design of modern immunotherapeutic treatments for CRC.

Abbreviations

M6A: N6-methyladenosine

TIME: Tumor immune microenvironment
TCGA: The Cancer Genome Atlas
CRC: Colorectal cancer
LASSO: Least absolute shrinkage and selection operator
M6A-LPS: m6A-related lncRNA prognostic signature
siRNA: Small interfering RNAs
lncRNAs: Long noncoding RNAs
PD-L1: Programmed death-ligand 1
ROC: Receiver operating characteristics
FDR: False discovery rate
TIMER: Tumor Immune Estimation Resource
CDFs: Cumulative distribution functions
AUC: Area under the curve
GSEA: Gene set enrichment analysis
OS: Overall survival
TAMs: Tumor-associated macrophages.

Data Availability

All data generated or analyzed during this study are included in this published article.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Authors' Contributions

F.K and J.Y conceived and designed the whole project. D.L and M.J.Z analyzed the data and wrote the manuscript. J.D carried out data interpretation and data analysis. J.Y revised the manuscript. All authors read and approved the final manuscript. Fei Kuang, De Luo, and Mengjia Zhou contributed equally to this work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC, 82100611).

Supplementary Materials

Table S1: primer sequences for qRT-PCR used in this study. *(Supplementary Materials)*

References

[1] J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin, "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008," *International Journal of Cancer*, vol. 127, no. 12, pp. 2893–2917, 2010.

[2] M. A. Curran, W. Montalvo, H. Yagita, and J. P. Allison, "PD-1 and CTLA-4 combination blockade expands infiltrating T cells and reduces regulatory T and myeloid cells within B16 melanoma tumors," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 9, pp. 4275–4280, 2010.

[3] M. C. Garassino, S. Gadgeel, E. Esteban et al., "Patient-reported outcomes following pembrolizumab or placebo plus pemetrexed and platinum in patients with previously untreated, metastatic, non-squamous non-small-cell lung cancer (KEYNOTE-189): a multicentre, double-blind, randomised, placebo-controlled, phase 3 trial," *The Lancet Oncology*, vol. 21, no. 3, pp. 387–397, 2020.

[4] C. Robert, A. Ribas, J. D. Wolchok et al., "Anti-programmed-death-receptor-1 treatment with pembrolizumab in ipilimumab- refractory advanced melanoma: a randomised dose-comparison cohort of a phase 1 trial," *Lancet*, vol. 384, no. 9948, pp. 1109–1117, 2014.

[5] O. Hamid, C. Robert, A. Daud et al., "Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma," *The New England Journal of Medicine*, vol. 369, no. 2, pp. 134–144, 2013.

[6] B. S. Zhao, I. A. Roundtree, and C. He, "Post-transcriptional gene regulation by mRNA modifications," *Nature Reviews. Molecular Cell Biology*, vol. 18, no. 1, pp. 31–42, 2017.

[7] D. Dai, H. Wang, L. Zhu, H. Jin, and X. Wang, "N6-methyladenosine links RNA metabolism to cancer progression," *Cell Death & Disease*, vol. 9, no. 2, p. 124, 2018.

[8] S. Zaccara, R. J. Ries, and S. R. Jaffrey, "Reading, writing and erasing mRNA methylation," *Nature Reviews. Molecular Cell Biology*, vol. 20, no. 10, pp. 608–624, 2019.

[9] H. Weng, H. Huang, H. Wu et al., "METTL14 inhibits hematopoietic stem/progenitor differentiation and promotes leukemogenesis via mRNA m^6A modification," *Cell Stem Cell*, vol. 22, no. 2, pp. 191–205.e9, 2018.

[10] Q. Cui, H. Shi, P. Ye et al., "m^6A RNA methylation regulates the self-renewal and tumorigenesis of glioblastoma stem cells," *Cell Reports*, vol. 18, no. 11, pp. 2622–2634, 2017.

[11] J. Liu, D. Ren, Z. Du, H. Wang, H. Zhang, and Y. Jin, "m^6A demethylase FTO facilitates tumor progression in lung squamous cell carcinoma by regulating MZF1 expression," *Biochemical and Biophysical Research Communications*, vol. 502, no. 4, pp. 456–464, 2018.

[12] L. Zhong, D. Liao, M. Zhang et al., "YTHDF2 suppresses cell proliferation and growth via destabilizing the EGFR mRNA in hepatocellular carcinoma," *Cancer Letters*, vol. 442, pp. 252–261, 2019.

[13] Y. Dong, T. Yu, L. Ding et al., "A dual targeting dendrimer-mediated siRNA delivery system for effective gene silencing in cancer therapy," *Journal of the American Chemical Society*, vol. 140, no. 47, pp. 16264–16274, 2018.

[14] H. J. Kim, A. Kim, K. Miyata, and K. Kataoka, "Recent progress in development of siRNA delivery vehicles for cancer therapy," *Advanced Drug Delivery Reviews*, vol. 104, pp. 61–77, 2016.

[15] K. Katsushima, A. Natsume, F. Ohka et al., "Targeting the Notch-regulated non-coding RNA TUG1 for glioma treatment," *Nature Communications*, vol. 7, no. 1, p. 13616, 2016.

[16] X. Li, Z. Xie, C. Xie et al., "D-SP5 peptide-modified highly branched polyethylenimine for gene therapy of gastric adenocarcinoma," *Bioconjugate Chemistry*, vol. 26, no. 8, pp. 1494–1503, 2015.

[17] J. Wang, Y. Lei, C. Xie et al., "Retro-inverso CendR peptide-mediated polyethyleneimine for intracranial glioblastoma-targeting gene therapy," *Bioconjugate Chemistry*, vol. 25, no. 2, pp. 414–423, 2014.

[18] J. Wang, Y. Lei, C. Xie et al., "Targeted gene delivery to glioblastoma using a C-end rule RGERPPR peptide-functionalised polyethylenimine complex," *International Journal of Pharmaceutics*, vol. 2013, no. 458, pp. 48–56, 2013.

[19] H. Shi, J. Wei, and C. He, "Where, when, and how: context-dependent functions of RNA methylation writers, readers, and erasers," *Molecular Cell*, vol. 74, no. 4, pp. 640–650, 2019.

[20] Y. Yang, P. J. Hsu, Y. S. Chen, and Y. G. Yang, "Dynamic transcriptomic m^6A decoration: writers, erasers, readers and functions in RNA metabolism," *Cell Research*, vol. 28, no. 6, pp. 616–624, 2018.

[21] M. D. Wilkerson and D. N. Hayes, "ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking," *Bioinformatics*, vol. 26, no. 12, pp. 1572-1573, 2010.

[22] K. Yoshihara, M. Shahmoradgoli, E. Martínez et al., "Inferring tumour purity and stromal and immune cell admixture from expression data," *Nature Communications*, vol. 4, no. 1, p. 2612, 2013.

[23] H. M. Bovelstad, S. Nygard, H. L. Storvold et al., "Predicting survival from microarray data–a comparative study," *Bioinformatics*, vol. 23, no. 16, pp. 2080–2087, 2007.

[24] S. Zhang, B. S. Zhao, A. Zhou et al., "m6A demethylase ALKBH5 maintains tumorigenicity of glioblastoma stem-like cells by sustaining FOXM1 expression and cell proliferation program," *Cancer Cell*, vol. 31, no. 4, pp. 591–606.e6, 2017.

[25] J. Y. Roignant and M. Soller, "m6A in mRNA: an ancient mechanism for fine-tuning gene expression," *Trends in Genetics*, vol. 33, no. 6, pp. 380–390, 2017.

[26] J. Z. Ma, F. Yang, C. C. Zhou et al., "METTL14 suppresses the metastatic potential of hepatocellular carcinoma by modulating N6-methyladenosine-dependent primary MicroRNA processing," *Hepatology*, vol. 65, no. 2, pp. 529–543, 2017.

[27] Z. Li, H. Weng, R. Su et al., "FTO plays an oncogenic role in acute myeloid leukemia as a N^6-methyladenosine RNA demethylase," *Cancer Cell*, vol. 31, no. 1, pp. 127–141, 2017.

[28] S. Lin, J. Choe, P. Du, R. Triboulet, and R. I. Gregory, "The m^6A methyltransferase METTL3 promotes translation in human cancer cells," *Molecular Cell*, vol. 62, no. 3, pp. 335–345, 2016.

[29] T. Lan, H. Li, D. Zhang et al., "KIAA1429 contributes to liver cancer progression through N6-methyladenosine-dependent post-transcriptional modification of GATA3," *Molecular Cancer*, vol. 18, no. 1, p. 186, 2019.

[30] L. Ma, W. X. Kuai, X. Z. Sun, X. C. Lu, and Y. F. Yuan, "Long noncoding RNA LINC00265 predicts the prognosis of acute myeloid leukemia patients and functions as a promoter by activating PI3K-AKT pathway," *European Review for Medical and Pharmacological Sciences*, vol. 22, no. 22, pp. 7867–7876, 2018.

[31] J. Ding, J. Zhao, L. Huan et al., "Inflammation-induced long intergenic noncoding RNA (LINC00665) increases malignancy through activating the double-stranded RNA–activated protein kinase/nuclear factor kappa B pathway in hepatocellular carcinoma," *Hepatology*, vol. 72, no. 5, pp. 1666–1681, 2020, [Epub ahead of print].

[32] G. Yang, X. Lu, and L. Yuan, "LncRNA: a link between RNA and cancer," *Biochimica et Biophysica Acta*, vol. 1839, no. 11, pp. 1097–1109, 2014.

[33] W. H. Fridman, F. Pagès, C. Sautès-Fridman, and J. Galon, "The immune contexture in human tumours: impact on clinical outcome," *Nature Reviews. Cancer*, vol. 12, no. 4, pp. 298–306, 2012.

[34] J. Galon, B. Mlecnik, G. Bindea et al., "Towards the introduction of the 'Immunoscore' in the classification of malignant tumours," *The Journal of Pathology*, vol. 232, no. 2, pp. 199–209, 2014.

[35] L. Hui and Y. Chen, "Tumor microenvironment: sanctuary of the devil," *Cancer Letters*, vol. 368, no. 1, pp. 7–13, 2015.

[36] T. L. Whiteside, "The tumor microenvironment and its role in promoting tumor growth," *Oncogene*, vol. 27, no. 45, pp. 5904–5912, 2008.

[37] J. Galon, A. Costes, F. Sanchez-Cabo et al., "Type, density, and location of immune cells within human colorectal tumors predict clinical outcome," *Science*, vol. 313, no. 5795, pp. 1960–1964, 2006.

[38] J. Galon, W. H. Fridman, and F. Pages, "The adaptive immunologic microenvironment in colorectal cancer: a novel perspective," *Cancer Research*, vol. 67, no. 5, pp. 1883–1886, 2007.

[39] R. Noy and J. W. Pollard, "Tumor-associated macrophages: from mechanisms to therapy," *Immunity*, vol. 41, no. 1, pp. 49–61, 2014.

[40] M. De Palma and C. E. Lewis, "Macrophage regulation of tumor responses to anticancer therapies," *Cancer Cell*, vol. 23, no. 3, pp. 277–286, 2013.

[41] Y.-P. Chen, Y.-Q. Wang, J.-W. Lv, Y.-Q. Li, M. L. K. Chua, and Q.-T. Le, "Identification and validation of novel microenvironment-based immune molecular subgroups of head and neck squamous cell carcinoma: implications for immunotherapy," *Annals of Oncology*, vol. 30, no. 1, pp. 68–75, 2019.

[42] M. Vassilakopoulou, M. Avgeris, V. Velcheti et al., "Evaluation of PD-L1 expression and associated tumor-infiltrating lymphocytes in laryngeal squamous cell carcinoma," *Clinical Cancer Research*, vol. 22, no. 3, pp. 704–713, 2016.

Hindawi
Disease Markers
Volume 2022, Article ID 2148627, 19 pages
https://doi.org/10.1155/2022/2148627

Research Article

The Variation of Transcriptomic Perturbations is Associated with the Development and Progression of Various Diseases

Zehua Dong,[1,2] Qiyu Yan,[1,2] and Xiaosheng Wang[1,2]

[1]*Biomedical Informatics Research Laboratory, School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing 211198, China*
[2]*Big Data Research Institute, China Pharmaceutical University, Nanjing 211198, China*

Correspondence should be addressed to Xiaosheng Wang; xiaosheng.wang@cpu.edu.cn

Received 6 August 2022; Revised 31 August 2022; Accepted 7 September 2022; Published 26 September 2022

Academic Editor: Xiang-Yu Meng

Background. Although transcriptomic data have been widely applied to explore various diseases, few studies have investigated the association between transcriptomic perturbations and disease development in a wide variety of diseases. *Methods*. Based on a previously developed algorithm for quantifying intratumor heterogeneity at the transcriptomic level, we defined the variation of transcriptomic perturbations (VTP) of a disease relative to the health status. Based on publicly available transcriptome datasets, we compared VTP values between the disease and health status and analyzed correlations between VTP values and disease progression or severity in various diseases, including neurological disorders, infectious diseases, cardiovascular diseases, respiratory diseases, liver diseases, kidney diseases, digestive diseases, and endocrine diseases. We also identified the genes and pathways whose expression perturbations correlated positively with VTP across diverse diseases. *Results*. VTP values were upregulated in various diseases relative to their normal controls. VTP values were significantly greater in define than in possible or probable Alzheimer's disease. VTP values were significantly larger in intensive care unit (ICU) COVID-19 patients than in non-ICU patients, and in COVID-19 patients requiring mechanical ventilatory support (MVS) than in those not requiring MVS. VTP correlated positively with viral loads in acquired immune deficiency syndrome (AIDS) patients. Moreover, the AIDS patients treated with abacavir or zidovudine had lower VTP values than those without such therapies. In pulmonary tuberculosis (TB) patients, VTP values followed the pattern: active TB > latent TB > normal controls. VTP values were greater in clinically apparent than in presymptomatic malaria. VTP correlated negatively with the cardiac index of left ventricular ejection fraction (LVEF). In chronic obstructive pulmonary disease (COPD), VTP showed a negative correlation with forced expiratory volume in the first second (FEV1). VTP values increased with H. pylori infection and were upregulated in atrophic gastritis caused by H. pylori infection. The genes and pathways whose expression perturbations correlated positively with VTP scores across diseases were mainly involved in the regulation of immune, metabolic, and cellular activities. *Conclusions*. VTP is upregulated in the disease versus health status, and its upregulation is associated with disease progression and severity in various diseases. Thus, VTP has potential clinical implications for disease diagnosis and prognosis.

1. Introduction

With the recent development of next-generation sequencing (NGS) technologies, a substantial number of multiomics data associated with various diseases have been produced, including cancer, neurological disorders, cardiovascular disease, respiratory disease, digestive system disease, metabolic disease, endocrine disease, kidney and urinary system disorders, and infectious disease. In a previous study [1], we developed an algorithm, termed DEPTH, to quantify the variation of transcriptomic perturbations (VTP) in cancer, namely intratumor heterogeneity. We found that VTP value was significantly higher in cancer than in normal controls. Moreover, VTP values increased with cancer advancement, and its

increase were associated with worse clinical outcomes in cancer patients [1]. In this study, we generalized this algorithm to a wide variety of diseases and explored the association between VTP and prognosis-associated clinical features. The disease types we analyzed included neurological disorders, infectious diseases, cardiovascular diseases, respiratory diseases, liver diseases, kidney diseases, digestive diseases, and endocrine diseases. We compared VTP values between the disease state and normal controls and analyzed correlations between VTP and disease progression or severity.

2. Methods

2.1. Algorithm. The algorithm is described as follows: given a transcriptome dataset, which involves g genes and m disease samples and n normal control samples; the variation of transcriptomic perturbations (VTP) of a disease sample DS is defined as

$$\begin{aligned}
&\sqrt{\frac{1}{g-1}\left(\sum_{i=1}^{g}\left(\Delta\left(G_i,\mathrm{DS},\mathrm{NS}_j\right)^2-\frac{1}{g}\sum_{i=1}^{g}\Delta\left(G_i,\mathrm{DS},\mathrm{NS}_j\right)^2\right)^2\right)}\\
&\cdot = \sqrt{\frac{1}{g-1}\left(\sum_{i=1}^{g}\left(\left(\mathrm{ex}(G_i,\mathrm{DS})-\frac{1}{n}\sum_{j=1}^{n}\mathrm{ex}\left(G_i,\mathrm{NS}_j\right)\right)^2-\frac{1}{g}\sum_{i=1}^{g}\left(\mathrm{ex}(G_i,\mathrm{DS})-\frac{1}{n}\sum_{j=1}^{n}\mathrm{ex}\left(G_i,\mathrm{NS}_j\right)\right)^2\right)^2\right)},\\
&\cdot \Delta\left(G_i,\mathrm{DS},\mathrm{NS}_j\right)=\mathrm{ex}(G_i,\mathrm{DS})-\frac{1}{n}\sum_{j=1}^{n}\mathrm{ex}\left(G_i,\mathrm{NS}_j\right),
\end{aligned} \tag{1}$$

where $ex(G_i, \mathrm{DS})$ indicates the expression value of gene G_i in DS, and $\mathrm{ex}(G_i, \mathrm{NS}_j)$ indicates the expression value of G_i in the normal sample NS_j.

2.2. Datasets. We downloaded transcriptome datasets for various diseases from the NCBI Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) and analyzed these datasets with the algorithm. The datasets were associated with various types of diseases, including neurological disorders (Alzheimer's disease (AD) and schizophrenia (SCZ)), infectious diseases (COVID-19, acquired immune deficiency syndrome (AIDS), hepatitis B virus (HBV) infection, tuberculosis (TB), and malaria), cardiovascular diseases (acute myocardial infarction, dilated cardiomyopathy, idiopathic or ischemic cardiomyopathy, and heart failure), respiratory diseases (chronic obstructive pulmonary disease), liver diseases (chronic hepatitis B and liver cirrhosis), kidney diseases (nephrotic syndrome, uremia, focal segmental glomerulosclerosis and glomerular disease), digestive diseases (inflammatory bowel disease and helicobacter pylori infection), and endocrine diseases (diabetes). A description of these datasets is shown in Table 1.

2.3. Data Preprocessing. For RNA-Seq gene expression values, we normalized them by the TPM method. For microarray gene expression values, we used the normalization methods recommended by related platforms. A description of the normalization methods for the datasets analyzed was provided in Supplementary Table S1. All normalized expression values were transformed by $\log 2(x+1)$ before subsequent analyses.

2.4. Statistical Analysis and Visualization. We employed the Mann–Whitney U test (one-tailed) to compare VTP values between two classes of samples, and the Kruskal-Wallis test to compare VTP values among more than two classes of samples. We utilized the Spearman method to assess the correlation between VTP values and other variables and reported the correlation coefficients (ρ) and P values. To correct for P values in multiple tests, we utilized the Benjamini and Hochberg method to calculate the false discovery rate (FDR) [2]. All statistical analyses were performed in the R programming environment (version 4.1.2). The R packages "ggplot2", "ggpubr", and "ggstatsplot" were used for data visualization.

2.5. Identifying Genes and Pathways whose Expression Perturbations Have Significant Positive Correlations with VTP across Diverse Diseases. In each dataset, we identified the genes satisfying that $|\Delta(G_i, \mathrm{DS}, \mathrm{NS}_j)|$ significantly and positively correlated with VTP values in all disease samples using a threshold of the Spearman correlation test FDR < 0.1. For each disease with n datasets analyzed, we identified the genes which satisfied the prior condition at least $n-1$ datasets. These genes were defined as the genes having significant positive correlations of expression perturbations with VTP in specific diseases. Among them, the genes identified in common in at least 5 specific diseases were defined as the genes whose expression perturbations had significant positive correlations with VTP across diseases. By inputting the genes associated with VTP across diseases into the GSEA web tool [3], we obtained the KEGG pathways [4] having

Table 1: Summary of the datasets analyzed.

Disease	Dataset	Platform	Sample size	
			Patients	Controls
Alzheimer's disease	GSE63063	GPL6947	145	104
		GPL10558	139	134
	GSE84422	GPL96	737	214
			328 (definite AD)	
			180 (probable AD)	
			229 (possible AD)	
		GPL97	737	214
			328 (definite AD)	
			180 (probable AD)	
			229 (possible AD)	
		GPL570	74	28
			34 (definite AD)	
			23 (probable AD)	
			17 (possible AD)	
	GSE118553	GPL10558	167	100
	GSE140831	GPL15988	599	530
	GSE158233	GPL20828	20	10
	Danio rerio dataset			
Schizophrenia	GSE38484	GPL6947	106	96
	GSE53987	GPL570	48	55
	GSE87610	GPL13667	65	72
	GSE93577	GPL13667	70	71
	GSE93987	GPL13158	102	106
Corona virus Disease 2019	GSE152075	GPL18573	430	54
	GSE157103	GPL24676	102	
			50 (ICU)	26
			42 (MVS)	
	GSE161731	GPL24676	46	19
	GSE198449	GPL24676	149	
			85 (asymptomatic)	22
			64 (symptomatic)	
Acquired immune deficiency syndrome	GSE18233	GPL6884	153	3
			16 (EC)	
	GSE87620	GPL10558	83	
			51 (EC)	10
			32 (HAART-treated)	
	GSE104640	GPL10558	188	60
Hepatitis B virus infection	GSE83148	GPL570	122	6
	GSE114783	GPL15491	3	3
	GSE121248	GPL570	70	37
Tuberculosis	GSE28623	GPL4133	71	
			46 (TB)	37
			25 (LTB)	
	GSE153340	GPL21185	18	4
	GSE152532	GPL10558	50	
			42 (TB)	11

Table 1: Continued.

Disease	Dataset	Platform	Sample size	
			Patients	Controls
			8 (LTB)	
Malaria	GSE1124	GPL96	20 5 (asymptomatic) 5 (uncomplicated) 5 (severe)	5
		GPL97	20 5 (asymptomatic) 5 (uncomplicated) 5 (severe)	5
	GSE5418	GPL96	22	22
	GSE34404	GPL10558	94	61
Cardiovascular disease	GSE1145	GPL570	53	37
	GSE5406	GPL96	194	16
	GSE17800	GPL570	40	8
	GSE33463	GPL6947	49	41
	GSE48060	GPL570	31	21
	GSE62646	GPL6244	28	14
	GSE66360	GPL570	49	50
	GSE74144	GPL13497	28	8
	GSE109048	GPL17586	38	19
	GSE120895	GPL570	47	8
Respiratory disease	GSE5058	GPL570	15	24
	GSE42057	GPL570	94	42
	GSE55962	GPL13667	24	82
	GSE103174	GPL13667	37	16
	GSE112811	GPL570	20	18
	GSE151052	GPL17556	77	40
	GSE32147	GPL6101	56	4
	Rattus norvegicus dataset			
Liver disease	GSE14323	GPL96	58	19
	GSE77627	GPL14951	40 32 (cirrhosis) 18 (noncirrhosis)	14
	GSE135501	GPL13667	40 16 (white tongue) 24 (yellow tongue)	14
	GSE36533	GPL15354	11	10
	Marmota monax dataset			
Kidney disease	GSE37171	GPL570	63	20
	GSE104948	GPL22945	53	18
	GSE108113	GPL19983	269	5
	GSE133288	GPL19983	239	5
Digestive disease	GSE16879	GPL570	61	12
	GSE27411	GPL6255	12	6
Endocrine disease	GSE9006	GPL96	55	24

Table 1: Continued.

Disease	Dataset	Platform	Sample size	
			Patients	Controls
			T1D (43)	
			T2D (12)	
			55	
		GPL97	T1D (43)	24
			T2D (12)	
	GSE19420	GPL570	30	12
			57	
	GSE35725	GPL570	46 (recent onset)	44
			11 (longstanding)	

Note: EC: Elite controllers. HAART: Highly active antiretroviral therapy. LTB: Latently tuberculosis. T1D: Type 1 diabetes. T2D: Type 2 diabetes.

significant positive correlations of their expression perturbations with VTP across diseases using a threshold of FDR<0.1 .

3. Results

3.1. Neurological Disorder. AD is a progressive neurodegenerative disease [5]. In four transcriptome datasets for AD (GSE63063 [6], GSE118553 [7], GSE140831, and GSE84422 [8]), the VTP values were significantly larger in AD patients than in normal controls ($P < 0.001$) (Figure 1(a)). In GSE84422, VTP values were significantly larger in define than in possible or probable AD ($P = 0.02$) (Figure 1(a)). In addition, we analyzed correlations between VTP and several measures of the degree of AD progression in GSE84422, including clinical dementia rating, Braak neurofibrillary tangle score, average neuritic plaque density, sum of consortium to establish a registry for Alzheimer's diseases (CERAD) rating scores in multiple brain regions, and sum of neurofibrillary tangles density in multiple brain regions. Notably, VTP displayed significant positive correlations with these measures ($P < 0.01$) (Figure 1(a)). Mutations in PSEN2 may result in early-onset AD. In GSE158233 [9], Barthelson et al. generated transcriptomes of two-types of PSEN2-mutated ($psen2^{T141\_L142delinsMISLISV}$ and $psen2^{N140fs}$) lines of zebrafish brains and transcriptomes of their wild type siblings. We observed that VTP values were remarkedly greater in PSEN2-mutated zebrafish brains than in their wild type controls ($P < 0.03$) (Figure 1(a)).

Schizophrenia (SCZ) is a severe psychotic disorder characterized by relapsing incidences of psychosis [10]. In four transcriptome datasets (GSE38484 [11], GSE87610 [12], GSE93577 [13], and GSE93987 [14]) generated from SCZ patients and normal controls, VTP values were consistently greater in SCZ patients than in normal controls ($P < 0.02$) (Figure 1(b)).

Taken together, these results indicate that VTP is augmented in certain neurological disorders (such as AD and SCZ) and grows with disease progression.

3.2. Infectious Disease. COVID-19 is a highly contagious disease caused by SARS-CoV-2 infection and is currently widespread around the globe. This disease has caused more than 552 million cases and 6.3 million deaths as of July 1, 2022 [15]. In four transcriptome datasets (GSE152075 [16], GSE157103 [17], GSE161731 [18], and GSE198449 [19]) for COVID-19 patients, VTP values were significantly greater in COVID-19 patients than in normal controls ($P < 0.01$) (Figure 2(a)). Notably, in GSE157103, VTP values were significantly greater in intensive care unit (ICU) COVID-19 patients than in non-ICU patients ($P < 0.001$) (Figure 2(a)). Moreover, COVID-19 patients requiring mechanical ventilatory support (MVS) had greater VTP values than those not requiring MVS ($P < 0.001$) (Figure 2(a)). In addition, VTP displayed a significant positive correlation with the scores of the sequential organ failure assessment (SOFA) ($P = 0.006$; $\rho = 0.36$) (Figure 2(a)), which indicates the severity of ICU patients.

Acquired immune deficiency syndrome (AIDS) is a chronic condition resulting from infection with the human immunodeficiency virus (HIV) [20]. In three transcriptome datasets (GSE18233 [21], GSE87620 [22], and GSE104640 [23]) for AIDS patients, VTP values were significantly upregulated in AIDS patients versus normal controls ($P < 0.001$) (Figure 2(b)). In GSE18233, VTP correlated positively with viral loads ($P = 0.002$; $\rho = 0.28$) (Figure 2(b)). In GSE87620, the AIDS patients with highly active antiretroviral therapy had greater VTP values than elite controllers, who were the AIDS patients with undetectable levels of HIV replication not receiving antiretroviral therapy ($P = 0.01$) (Figure 2(b)). In addition, in GSE62117 [24], the AIDS patients treated with abacavir or zidovudine had lower VTP values than those without such therapies ($P < 0.05$) (Figure 2(b)).

Hepatitis B virus (HBV) infection is a major etiologic factor for hepatocellular carcinoma [25]. In three transcriptome datasets (GSE83148 [26], GSE114783 [27], and GSE121248 [28]), VTP values were greater in HBV-infected patients than in normal controls ($P < 0.01$) (Figure 2(c)).

Pulmonary tuberculosis (TB) is an infectious disease caused by Mycobacterium tuberculosis attacking lungs [29]. In three transcriptome datasets (GSE28623 [30], GSE153340 [31], and GSE152532 [32]), TB patients showed

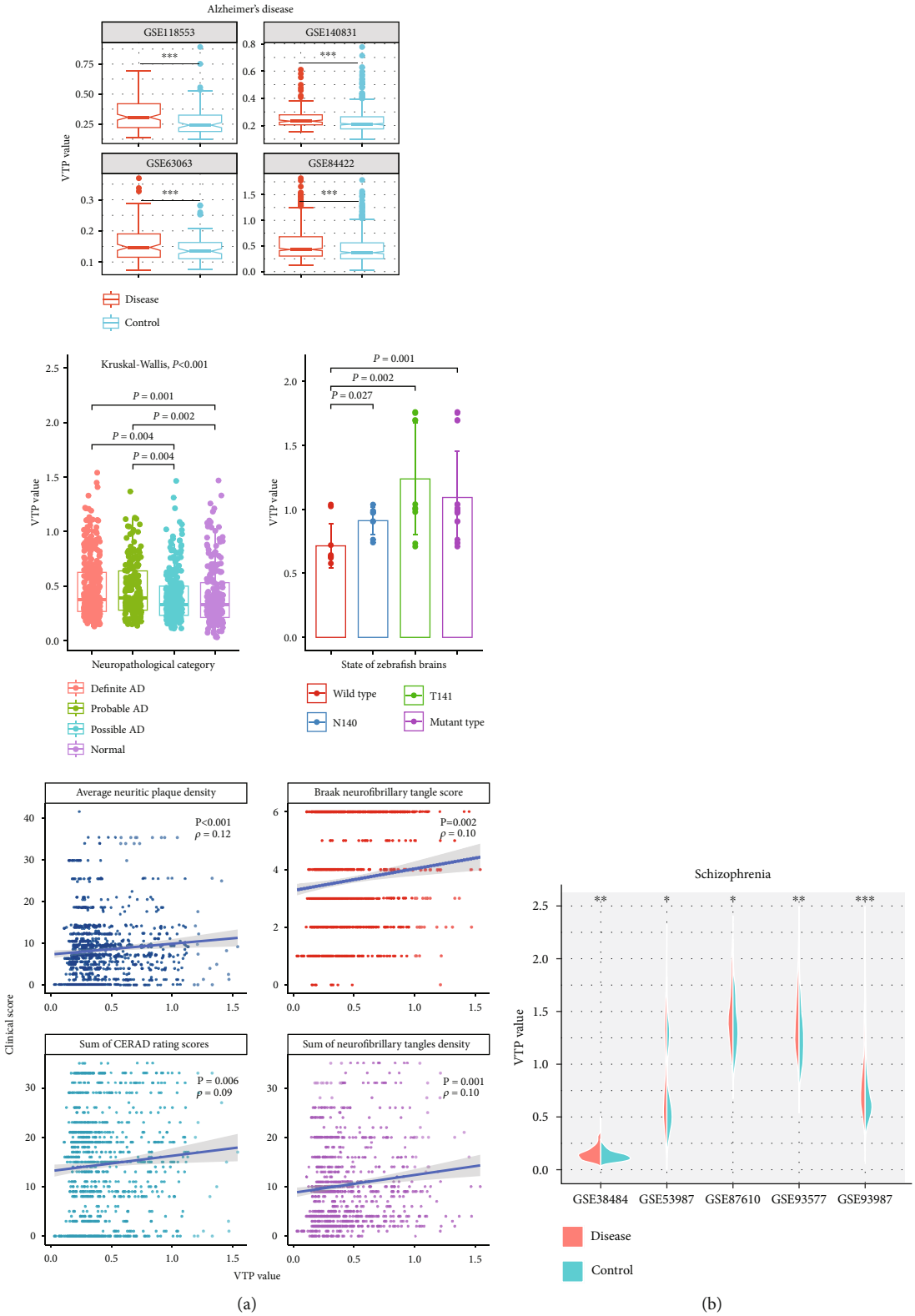
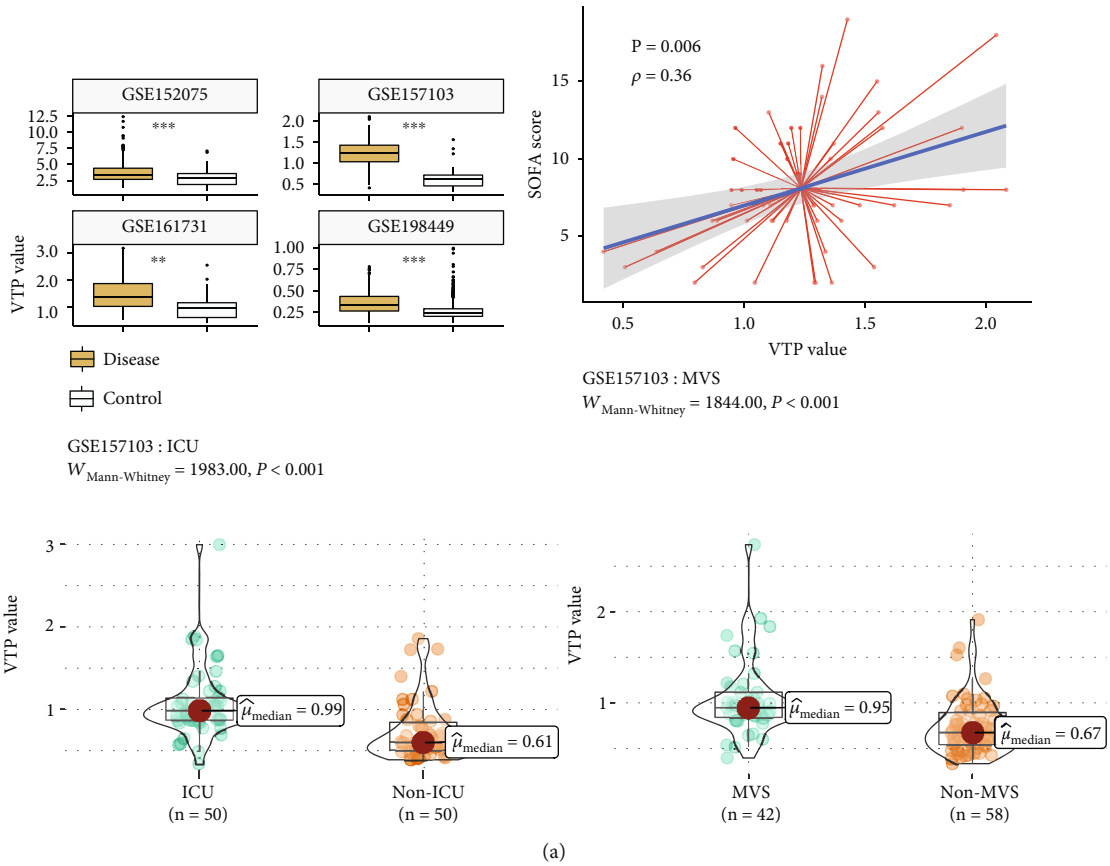
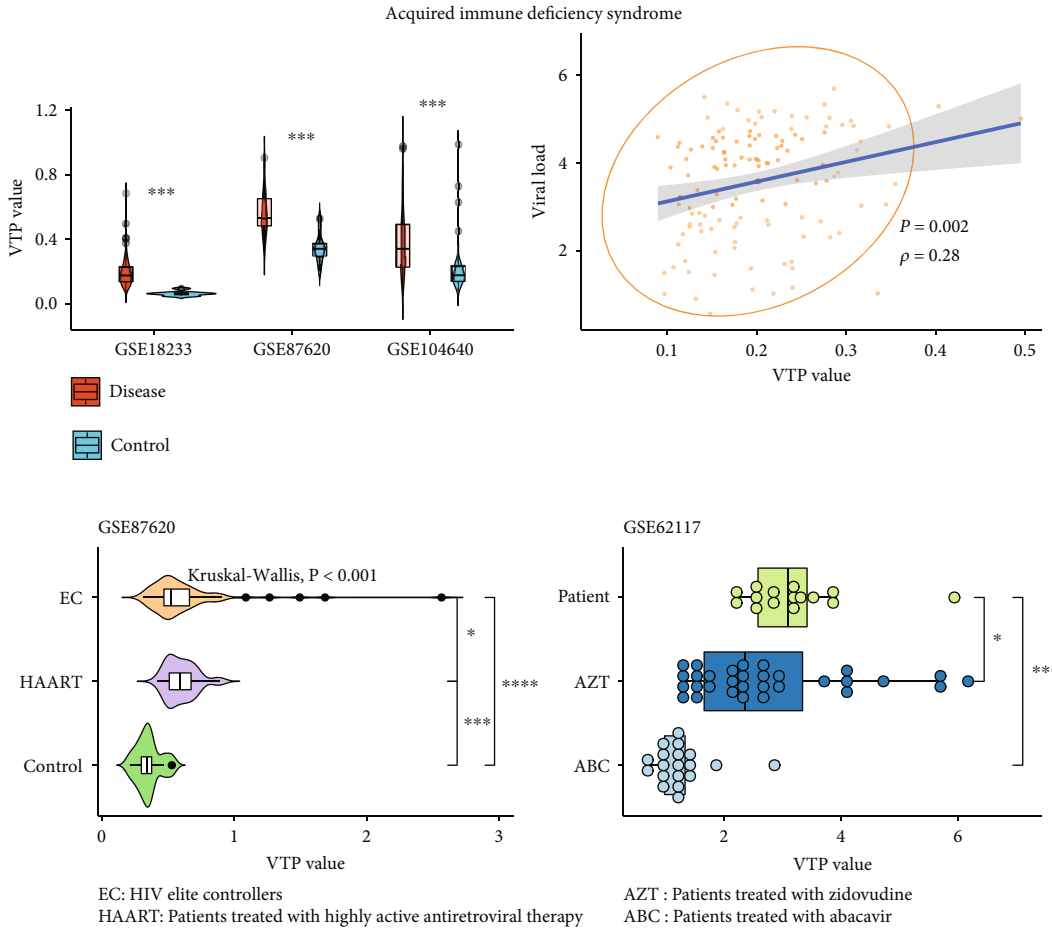


Figure 1: Associations between the VTP measure and disease development and progression in neurological disorder. (a) VTP values are significantly greater in AD patients than in normal controls, larger in define than in possible or probable AD, and increase with AD progression. The measures of Braak neurofibrillary tangle score, average neuritic plaque density, sum of CERAD rating scores in multiple brain regions, and sum of neurofibrillary tangles density in multiple brain regions represent the degree of AD progression. VTP values are remarkedly greater in PSEN2-mutated zebrafish brains than in their wild type controls. (b) VTP values are significantly greater in SCZ patients than in normal controls. AD: Alzheimer's disease. N140: psen2^{N140fs}. T141: psen2^{T141_L142delinsMISLISV}. SCZ: Schizophrenia. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. They also apply to the following figures.

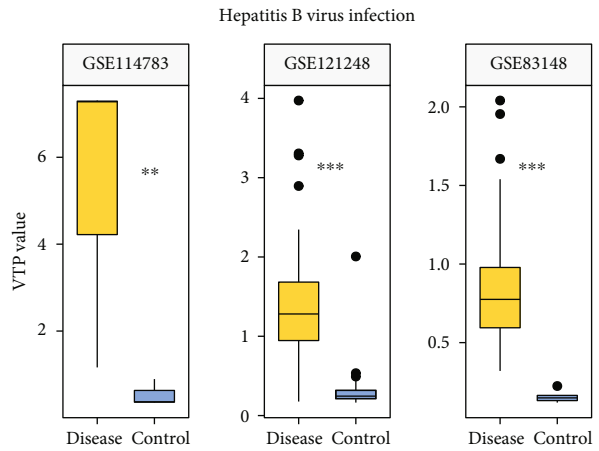


(a)

Figure 2: Continued.

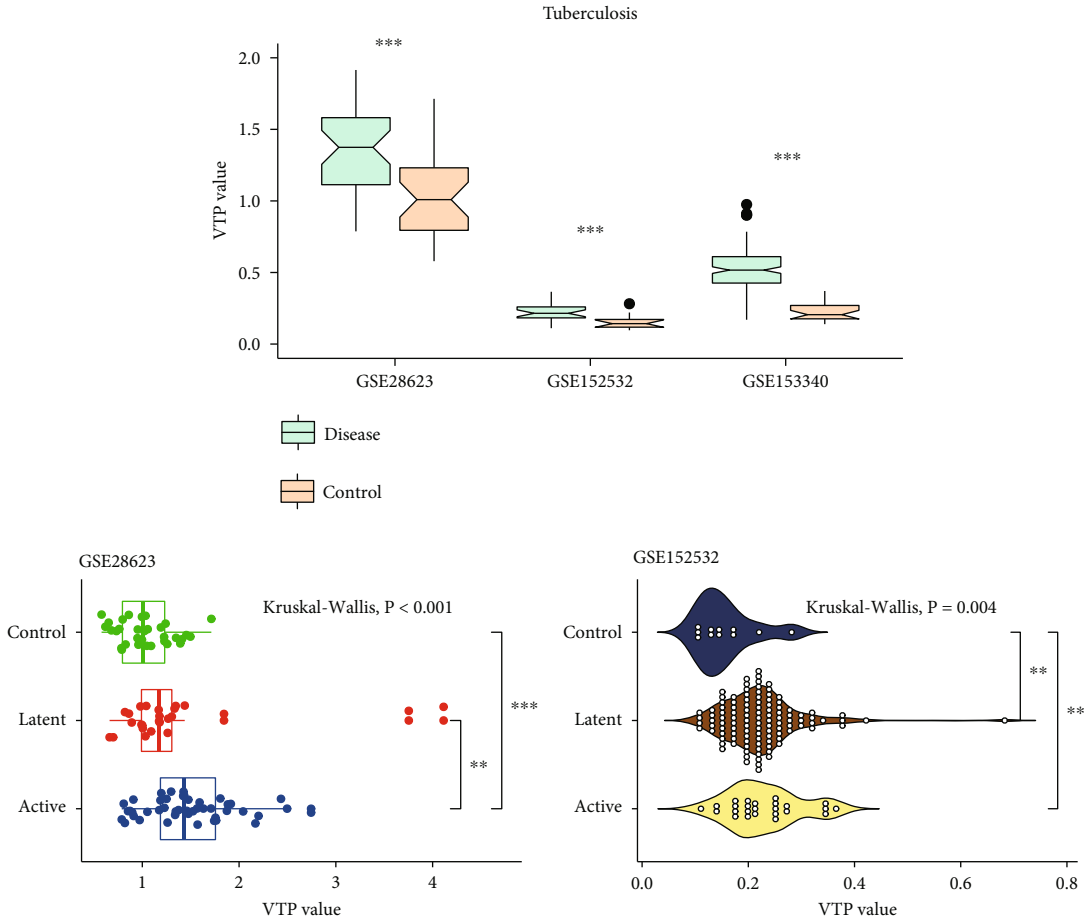


(b)



(c)

Figure 2: Continued.



(d)

Figure 2: Continued.

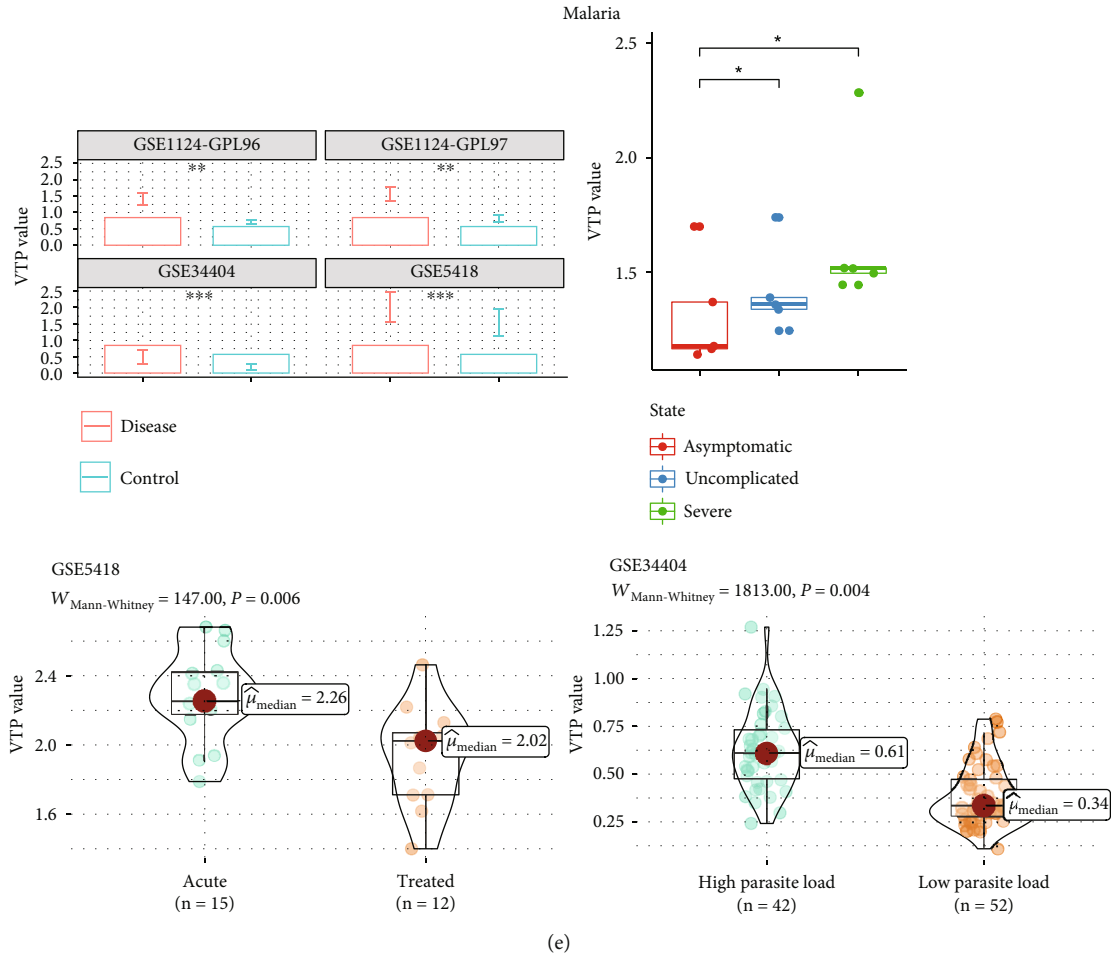


Figure 2: Associations between the VTP measure and disease development and progression in infectious disease. (a) VTP values are significantly greater in COVID-19 patients than in normal controls and increase with disease severity. (b) VTP values are significantly greater in AIDS patients than in normal controls, greater in AIDS patients without treatment than in AIDS patients with treatment, and increase with disease severity. (c). VTP values are significantly greater in HBV-infected patients than in normal controls. (d) VTP values are significantly greater in TB patients than in normal controls and increase with disease progression. (e) VTP values are significantly greater in malaria patients than in normal controls and increase with disease severity. ICU: intensive care unit. MVS: mechanical ventilatory support. SOFA: sequential organ failure assessment. AIDS: acquired immune deficiency syndrome. HBV: hepatitis B virus. TB: tuberculosis.

greater VTP values than normal controls ($P < 0.001$) (Figure 2(d)). Moreover, in GSE152532 and GSE28623 [30], VTP values likely followed the pattern: active TB > latent TB > normal controls (Figure 2(d)).

Malaria is a serious disease caused by a parasite and is a major cause of death globally [33]. In three transcriptome datasets (GSE1124 [34], GSE5418 [35], and GSE34404 [36]), malaria patients had greater VTP values than normal controls ($P < 0.01$) (Figure 2(e)). Moreover, in GSE34404, the high parasitemia group had significantly larger VTP values than the low parasitemia group of malaria patients ($P = 0.004$) (Figure 2(e)). In GSE1124, VTP values likely followed the pattern: malaria associated with severe anemia > uncomplicated malaria > asymptomatic infection (Figure 2(e)). In addition, in another transcriptome dataset (GSE5418), VTP values were greater in clinically apparent than in presymptomatic malaria ($P = 0.006$) (Figure 2(e)).

Collectively, these results support that VTP is upregulated in infectious diseases and increases with disease severity.

3.3. Cardiovascular Disease. Heart disease is the leading cause of death worldwide [37]. In numerous transcriptome datasets of heart disease, such as GSE1145, GSE5406 [38], GSE17800 [39], GSE48060 [40], GSE66360 [41], GSE109048 [42], and GSE120895 [43], VTP values were significantly greater in patients than in normal controls ($P < 0.05$) (Figure 3(a)). In GSE17800, VTP had a significant negative correlation with the cardiac index of left ventricular ejection fraction (LVEF) ($P = 0.035$; $\rho = -0.33$) (Figure 3(b)). In GSE62646 [44], the VTP values calculated based on gene expression patterns in leukocytes from acute myocardial infarction patients followed the pattern: the 1st day of myocardial infarction > after 4 – 6 days > after 6

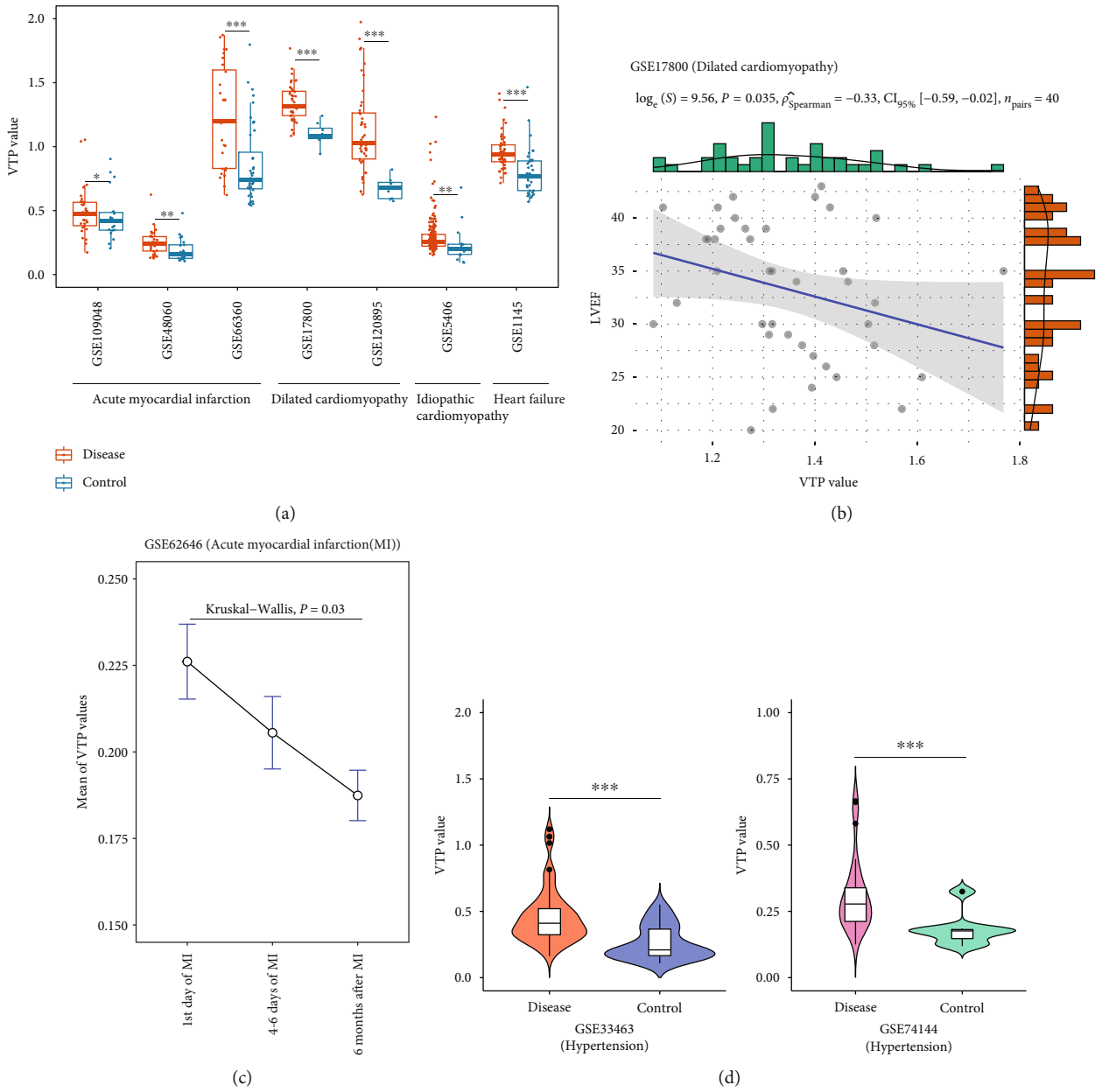


Figure 3: Associations between the VTP measure and disease development and progression in cardiovascular disease. (a) VTP values are significantly greater in heart disease patients than in normal controls and increase with disease severity. (b) VTP values correlate negatively with the cardiac index of LVEF. (c) VTP values decrease with the remission of acute myocardial infarction. (d) VTP values are significantly greater in hypertension patients than in normal controls. LVEF: left ventricular ejection fraction.

months (Figure 3(c)). In addition, in two transcriptome datasets (GSE33463 [45] and GSE74144) for hypertension, VTP values were significantly larger in patients than in normal controls ($P < 0.001$) (Figure 3(d)). Altogether, these results suggest that VTP is upregulated in cardiovascular diseases and decreases with disease remission.

3.4. Respiratory Disease. Respiratory diseases are the diseases affecting the organs and tissues involved in gas exchange in air-breathing animals [46]. Some of the most common respiratory diseases include obstructive lung disease, restrictive lung disease, and respiratory tract infections. In many transcriptome datasets of respiratory diseases, such as GSE112811, GSE42057 [47], GSE55962 [48], GSE103174, and GSE151052, VTP values were significantly larger in patients than in normal controls ($P < 0.05$) (Figure 4(a)). In chronic obstructive pulmonary disease (COPD), forced expiratory volume in the first second (FEV1) and ratio of FEV1 to forced vital capacity (FVC) are crucial in evaluating the severity of disease [49]. In GSE103174, which is a transcriptome dataset for COPD, VTP showed negative correlations with both FEV1 ($P = 0.018$; $\rho = -0.39$) and FEV1/FVC ($P = 0.067$; $\rho = -0.31$) (Figure 4(b)). The transcriptome dataset GSE32147 [50] is gene expression profiles in lung

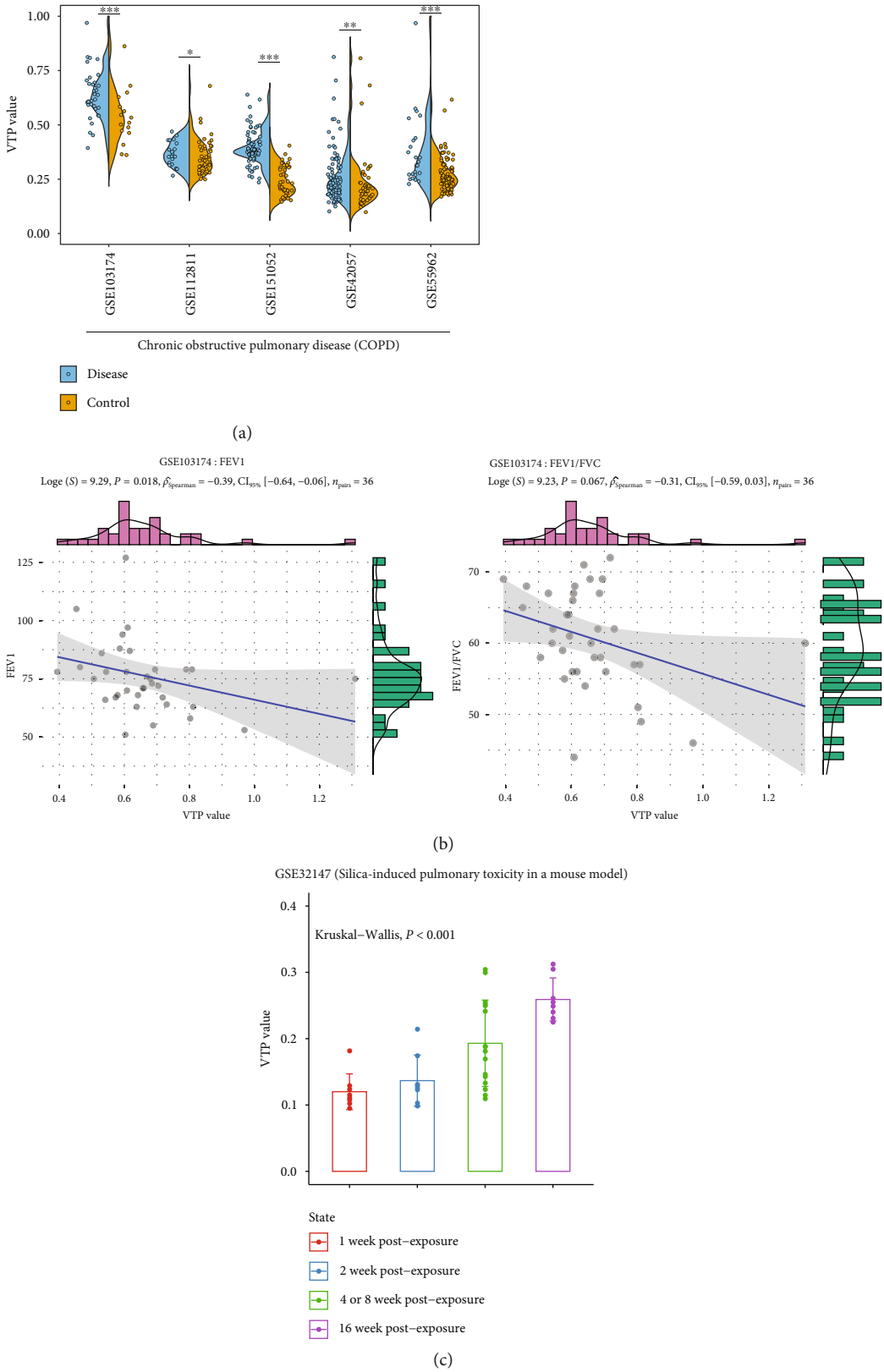
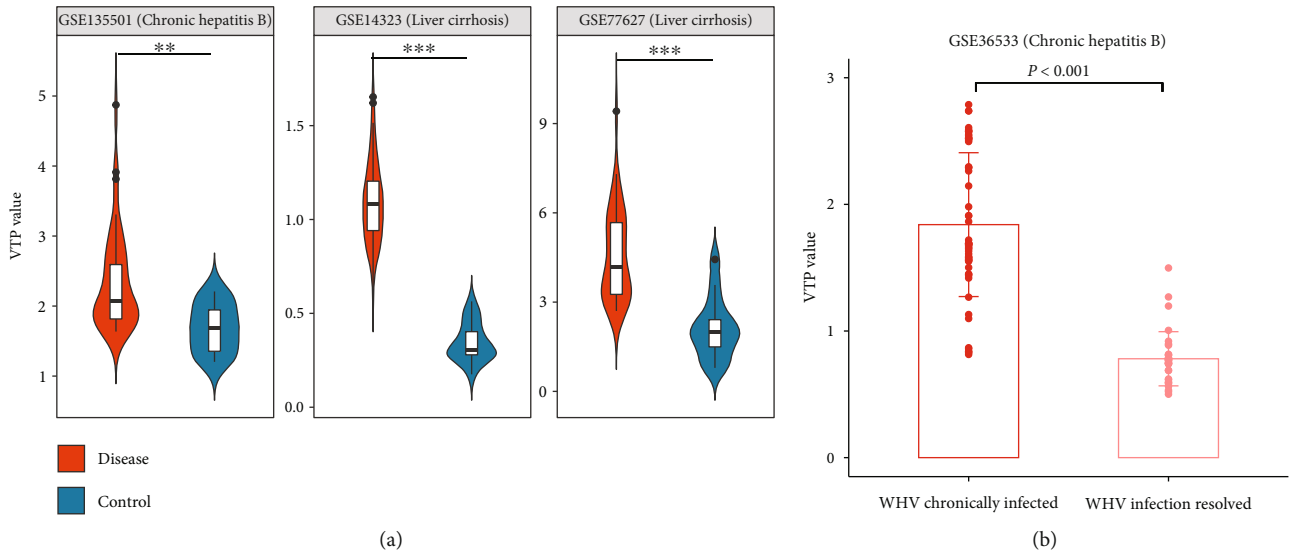
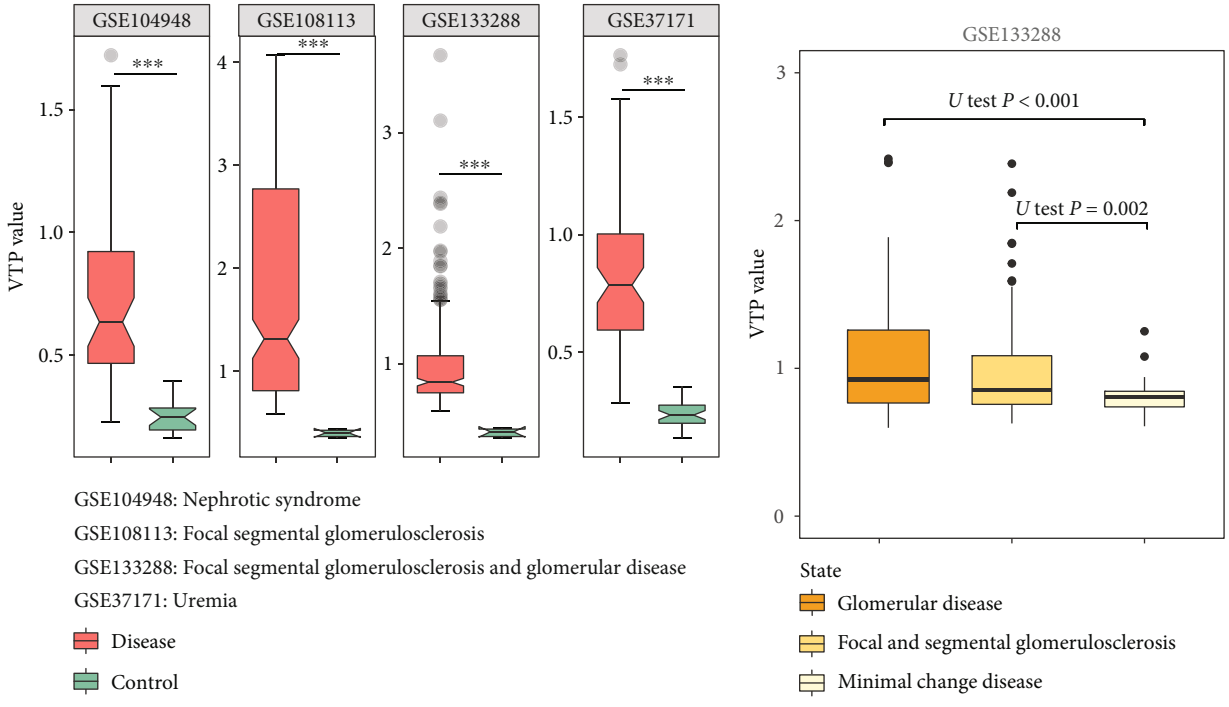


Figure 4: Associations between the VTP measure and disease development and progression in respiratory disease. (a) VTP values are significantly greater in respiratory disease patients than in normal controls. (b) VTP values correlate negatively with FEV1 and ratios of FEV1/FVC in COPD. (c) VTP values increase steadily with the progression of silica-induced pulmonary toxicity in a mouse model. COPD: chronic obstructive pulmonary disease. FEV1: forced expiratory volume in the first second. FVC: forced vital capacity.



(a) (b)

Figure 5: Associations between the VTP measure and disease development and progression in liver disease. (a) VTP values are significantly greater in liver disease patients than in normal controls. (b) VTP values are greater in WHV chronically infected than in infection resolved woodchuck in an animal model. WHV: woodchuck hepatitis virus.



(a) (b)

Figure 6: Associations between the VTP measure and disease development and progression in kidney disease. (a) VTP values are significantly greater in kidney disease patients than in normal controls. (b) VTP values correlate positively with disease severity in kidney disease.

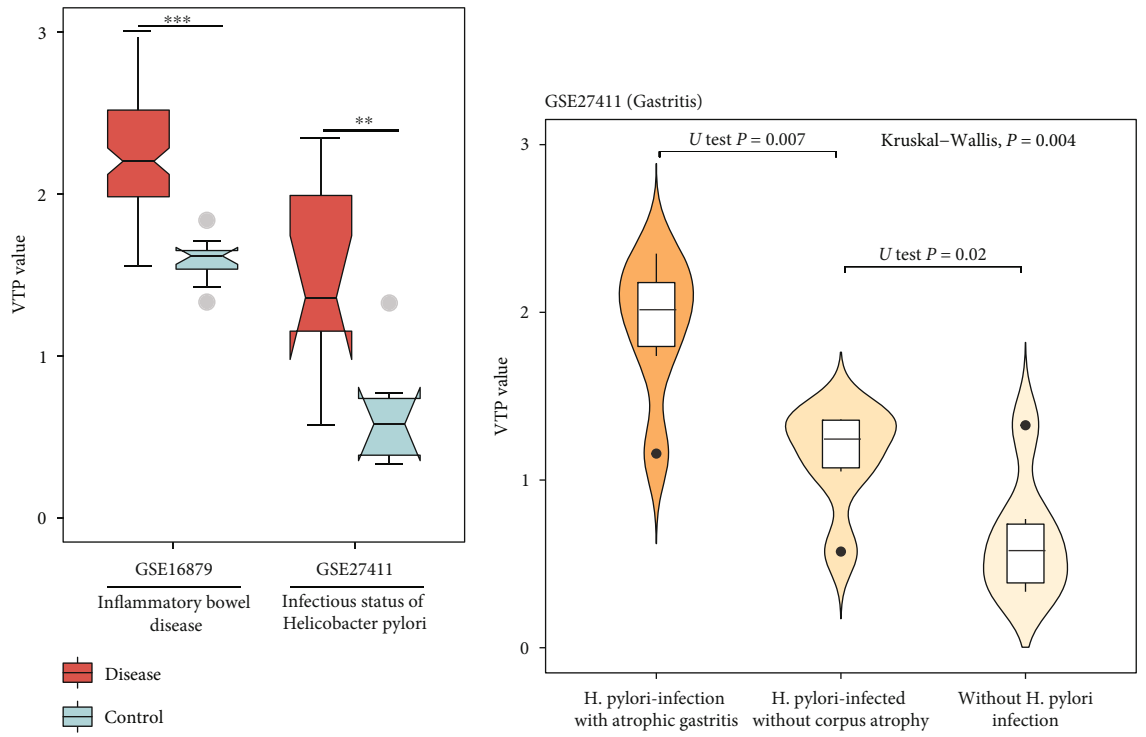


Figure 7: Associations between the VTP measure and disease development and progression in digestive disease. (a) VTP values are significantly greater in digestive disease patients than in normal controls. (b) VTP values correlate positively with disease severity in atrophic gastritis.

samples of rats exposed to crystalline silica. We observed that VTP values increased steadily with the progression of silica-induced pulmonary toxicity: 1 week of exposed to crystalline silica < 2 weeks < 4 or 8 weeks < 16 weeks (Figure 4(c)).

Collectively, these results support that VTP is upregulated in respiratory diseases and is negatively associated with their clinical outcomes.

3.5. Liver Disease. In three transcriptome datasets (GSE14323 [51], GSE77627, and GSE135501) for liver diseases, VTP values were significantly larger in patients than in normal controls ($P < 0.01$) (Figure 5(a)). The transcriptome dataset GSE36533 [52] is gene expression profiles in woodchuck infected with woodchuck hepatitis virus (WHV), an animal model for studying the human HBV. Notably, VTP values are greater in WHV chronically infected than in infection resolved woodchuck ($P < 0.001$) (Figure 5(b)).

3.6. Kidney Disease. In four transcriptome datasets (GSE37171 [53], GSE104948 [54], GSE108113 [54], and GSE133288 [55]) for kidney disease, VTP values were significantly greater in patients than in normal controls ($P < 0.001$) (Figure 6(a)). In addition, in GSE133228, VTP values were significantly larger in focal segmental glomerulosclerosis and glomerular disease than in minimal change disease ($P < 0.01$) (Figure 6(b)). It indicates that VTP values increase with disease progression in kidney disease.

3.7. Digestive Disease. In two transcriptome datasets (GSE16879 [56] and GSE27411 [57]) for digestive disease, VTP values were significantly larger in patients than in normal controls ($P < 0.01$) (Figure 7(a)). GSE27411 is a transcriptome dataset for patients with different stages of Helicobacter pylori (H. pylori) infection. Interestingly, we found that VTP values were significantly different among different stages of H. pylori infection and followed the pattern: without current H.pylori infection < H.pylori – infected without corpus atrophy < with current or past H.pylori – infection with corpus-predominant atrophic gastritis (Figure 7(b)). These results collectively support that VTP is upregulated in digestive diseases and increases with disease severity.

3.8. Endocrine Disease. Diabetes is a metabolic disease that causes high blood sugar to cause many chronic health problems, such as cardiovascular diseases, vision damage, and kidney disease [58]. In two transcriptome datasets (GSE9006 [59] and GSE19420 [60]) for diabetes, VTP values were significantly greater in patients than in normal controls ($P < 0.05$) (Figure 8(a)). Moreover, in the transcriptome dataset GSE35725 [61] for diabetes, VTP values were significantly greater in recent onset diabetes patients than in long-standing diabetes patients ($P < 0.001$) (Figure 8(b)).

3.9. Genes and Pathways whose Expression Perturbations Correlate Positively with VTP across Diseases. We identified

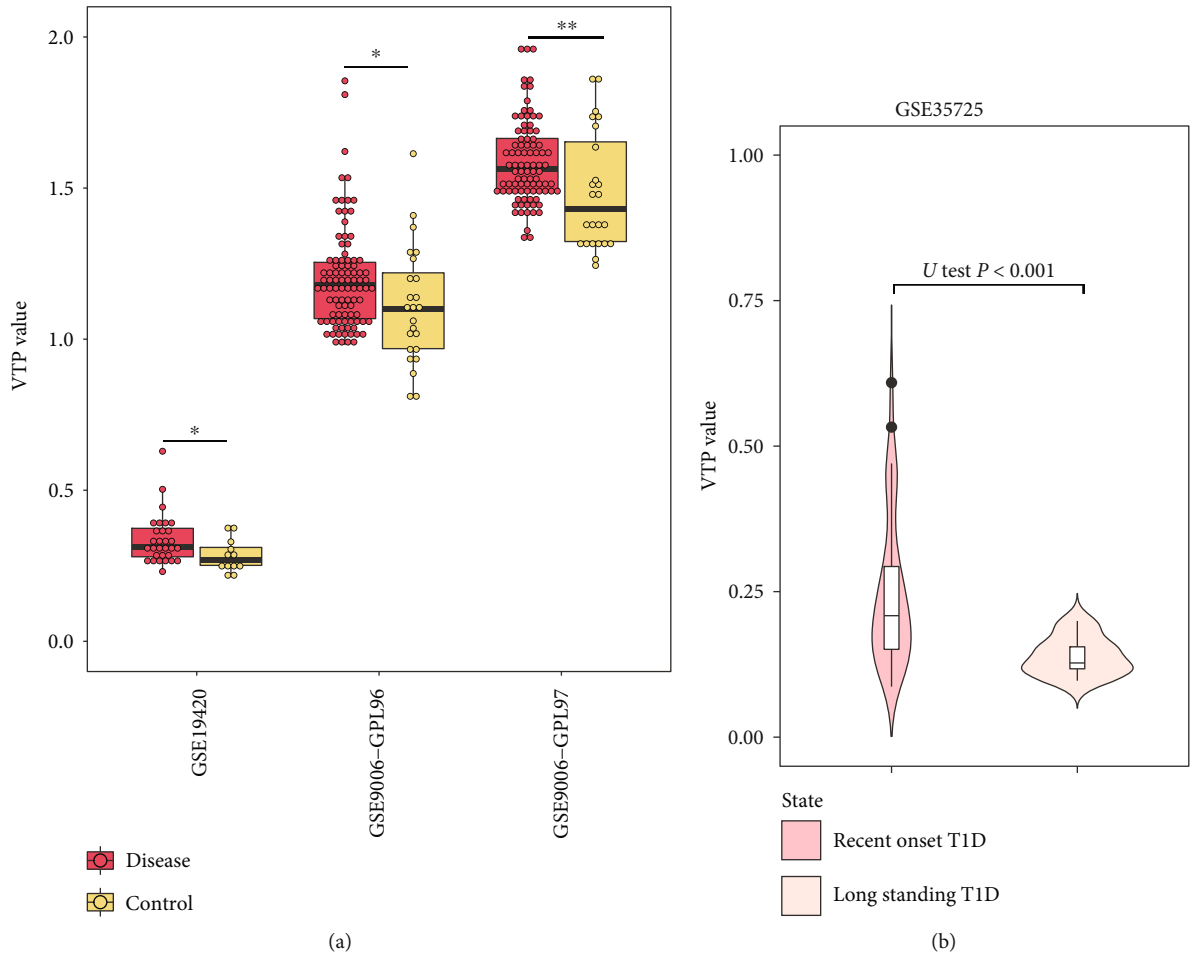


Figure 8: Associations between the VTP measure and disease development and progression in endocrine disease. (a) VTP values are significantly greater in diabetes patients than in normal controls. (b) VTP values are significantly greater in recent onset diabetes patients than in longstanding diabetes patients. T1D: Type 1 diabetes.

369 genes whose expression perturbations showed significant positive correlations with VTP values across diseases (Supplementary Table S2). Notably, many of these genes are involved in immune regulation (such as CD2, CD247, CD300A, CD2AP, CD28, CD47, CD53, CD7, and CXCR2), cell cycle (such as CCND2, CDK4, and SKP2), and metabolism (such as LDHA, LDHB, PDHA1, GLO1, and ME2). Furthermore, we identified 58 KEGG pathways showing significant positive correlations of expression perturbations with VTP across diseases. Notably, many of these pathways are immune pathways, including natural killer cell-mediated cytotoxicity, T cell receptor signaling, B cell receptor signaling, chemokine signaling, cell adhesion molecules, Fc gamma R-mediated phagocytosis, leukocyte transendothelial migration, Fc epsilon RI signaling, hematopoietic cell lineage, Toll-like receptor signaling, Jak-STAT signaling, cytokine-cytokine receptor interaction, intestinal immune network for IgA production, and NOD-like receptor signaling (Figure 9). The 58 pathways also included many metabolism-related pathways, such as pyruvate metabolism, inositol phosphate metabolism, propanoate metabolism, cysteine and methionine metabolism, fructose and mannose metabolism, riboflavin metabolism, β-alanine metabolism, and nicotinate and nicotinamide metabolism. Moreover, many pathways regulating cell growth and division were included in the list of the 58 pathways. Such pathways included MAPK signaling, Wnt signaling, calcium signaling, ErbB signaling, oocyte meiosis, and cell cycle. In addition, the 58 pathways also included many specific diseases-associated pathways, such as leishmania infection, AD, vibrio cholerae infection, epithelial cell signaling in Helicobacter pylori infection, amyotrophic lateral sclerosis, viral myocarditis, pathogenic Escherichia coli infection, arrhythmogenic right ventricular cardiomyopathy, pancreatic cancer, non-small-cell lung cancer, acute myeloid leukemia, colorectal cancer, glioma, and chronic myeloid leukemia.

4. Discussion

Although transcriptomic data have been widely applied to biomedical science, few studies have explored the association between transcriptomic perturbations and disease development and progression in a wide variety of diseases. For the

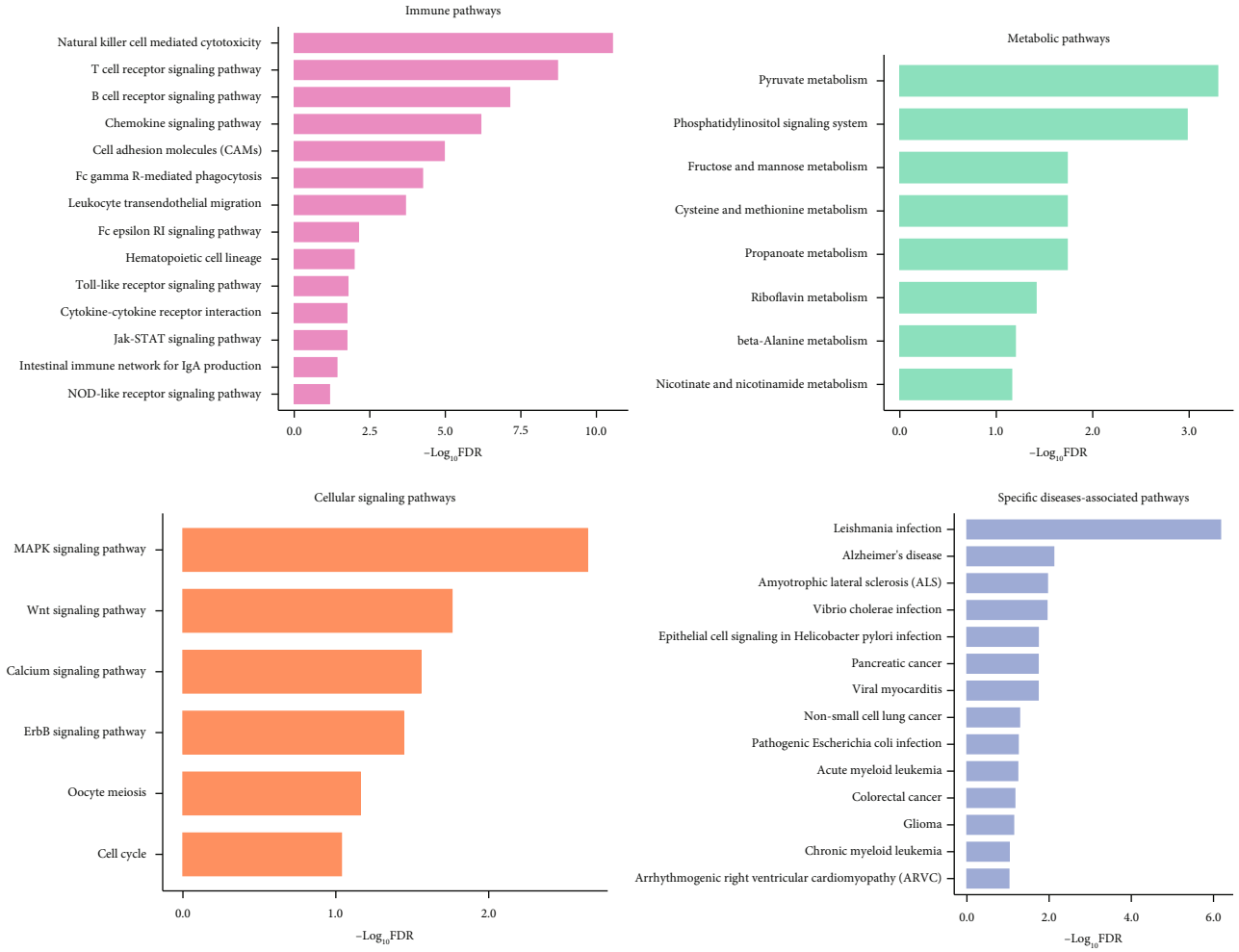


Figure 9: Pathways whose expression perturbations correlate positively with VTP scores across diseases.

first time, we investigated the association between the VTP and various diseases' onset and progression. Our analysis suggests that VTP values are upregulated in various diseases relative to their normal controls, and that VTP values increase with disease progression. Thus, this analysis uncovers a common characteristic of transcriptomic perturbations across various human diseases. In fact, the VTP measure reflects the asynchronous degree of transcriptomic perturbations in a disease status relative to the health status. Our results indicate that the asynchronous degree of transcriptomic perturbations is positively associated with disease progression or severity. That is, the higher asynchronous degree of transcriptomic perturbations suggests more unfavorable clinical outcomes in disease. This is consistent with the findings in cancer [1]. An intriguing question is whether the variation of perturbations in other molecules, such as genome, proteome, and metabolome, has similar associations with disease development and progression.

We identified numerous genes and pathways whose expression perturbations correlated positively with VTP scores across diseases. These genes and pathways are mainly involved in the regulation of immune, metabolic, and cellular activities. It is justified since deregulated immune, metabolic, and cellular activities have been associated with various diseases. Our data suggest that the disordered perturbations of the molecules modulating immune, metabolic, or cellular activities are associated with the development and progression of various diseases. Interestingly, by searching for the database of publicly available GWAS summary statistics (https://www.ebi.ac.uk/gwas/), we found that many of the 369 genes, which displayed significant expression perturbations' correlations with VTP values across diseases, had genetic variants that are statistically associated with the risk of the diseases we analyzed (Supplementary Table S3). For example, there were 16 genes, including RDX, PIP4K2A, PILRA, LPXN, LILRB2, ITGAX, IQGAP2, FOXN2, CR1, CELF2, CDC42SE2, CD2AP, PDK4, PARP8, HSPA6, and BNIP3, whose genetic variants are statistically associated with the risk of AD. Six genes (TKT, TCF4, SWAP70, DDHD2, ARHGAP31, and LTB) showed significant associations of genetic variants with the risk of cardiovascular disease. Notably, FOXN2 had genetic variants statistically associated with the risk of both AD and SCZ, and NOTCH2 displayed genetic variants that are

statistically associated with the risk of both endocrine disease and kidney disease. These data support the relevance of many of these genes with the diseases.

This study has several limitations. First, although we have analyzed numerous datasets for various diseases, more datasets are needed to be analyzed to bolster the validity of this analysis. Second, the mechanism underlying the association between VTP and disease development and progression needs to be explored. Finally, the prospect of translating the present findings into clinical practice remains unclear. Nevertheless, our further study is to implement further investigations to overcome these limitations.

5. Conclusions

VTP is upregulated in the disease relative to health status, and its upregulation is associated with disease progression and severity in various diseases. The molecules whose abundance perturbations correlate positively with VTP are mainly involved in the regulation of immune, metabolic, and cellular activities. Thus, VTP has potential clinical values in disease diagnosis and prognosis.

Data Availability

All data associated with this study are available within the paper and the database of NCBI Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/).

Ethical Approval

Ethical approval and consent to participate were waived since we used only publicly available data and materials in this study.

Consent

Not applicable.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

ZD performed data analyses, visualization, and manuscript editing. QY performed data analyses, visualization, and manuscript editing. XW conceived this study, designed analysis strategies, and wrote the manuscript. All the authors read and approved the final manuscript. Zehua Dong and Qiyu Yan contributed equally to this work.

Acknowledgments

This work was supported by the China Pharmaceutical University (grant number 3150120001 to XW).

Supplementary Materials

Supplementary 1. Table S1: a description of the normalization methods for the datasets analyzed.

Supplementary 2. Table S2: 369 genes having significant positive correlations of expression perturbations with VTP values across diseases.

Supplementary 3. Table S3: the genes displayed significant expression perturbations' correlations with VTP values across diseases and had genetic variants statistically associated with a risk of the diseases.

References

[1] M. Li, Z. Zhang, L. Li, and X. Wang, "An algorithm to quantify intratumor heterogeneity based on alterations of gene expression profiles," *Communications biology*, vol. 3, no. 1, p. 505, 2020.

[2] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society B*, vol. 57, no. 1, pp. 289–300, 1995.

[3] A. Subramanian, P. Tamayo, V. K. Mootha et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.

[4] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: new perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Research*, vol. 45, no. D1, pp. D353–D361, 2017.

[5] J. A. Soria Lopez, H. M. Gonzalez, and G. C. Leger, "Alzheimer's disease," *Handbook of Clinical Neurology*, vol. 167, pp. 231–255, 2019.

[6] S. Sood, I. J. Gallagher, K. Lunnon et al., "A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status," *Genome Biology*, vol. 16, no. 1, p. 185, 2015.

[7] H. Patel, A. K. Hodges, C. Curtis et al., "Transcriptomic analysis of probable asymptomatic and symptomatic alzheimer brains," *Brain, Behavior, and Immunity*, vol. 80, pp. 644–656, 2019.

[8] M. Wang, P. Roussos, A. McKenzie et al., "Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease," *Genome Medicine*, vol. 8, no. 1, p. 104, 2016.

[9] K. Barthelson, S. M. Pederson, M. Newman, H. Jiang, and M. Lardelli, "In-frame and frameshift mutations in zebrafish Presenilin 2 affect different cellular functions in young adult brains," *Journal of Alzheimer's Disease Reports*, vol. 5, no. 1, pp. 395–404, 2021.

[10] J. van de Leemput, J. L. Hess, S. J. Glatt, and M. T. Tsuang, "Genetics of schizophrenia: historical insights and prevailing evidence," *Advances in Genetics*, vol. 96, pp. 99–141, 2016.

[11] S. de Jong, M. P. M. Boks, T. F. Fuller et al., "A gene co-expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes," *PLoS One*, vol. 7, no. 6, article e39498, 2012.

[12] D. Arion, Z. Huo, J. F. Enwright, J. P. Corradi, G. Tseng, and D. A. Lewis, "Transcriptome alterations in prefrontal

pyramidal cells distinguish schizophrenia from bipolar and major depressive disorders," *Biological Psychiatry*, vol. 82, no. 8, pp. 594–600, 2017.

[13] D. Datta, J. F. Enwright, D. Arion et al., "Mapping phosphodiesterase 4D (PDE4D) in macaque dorsolateral prefrontal cortex: postsynaptic compartmentalization in layer III pyramidal cell circuits," *Frontiers in Neuroanatomy*, vol. 14, article 578483, 2020.

[14] D. Arion, J. P. Corradi, S. Tang et al., "Distinctive transcriptome alterations of prefrontal pyramidal neurons in schizophrenia and schizoaffective disorder," *Molecular Psychiatry*, vol. 20, no. 11, pp. 1397–1405, 2015.

[15] *COVID-19 Dashboard*, Johns Hopkins University, 2022, https://coronavirus.jhu.edu/map.html.

[16] N. A. P. Lieberman, V. Peddu, H. Xie et al., "In vivo antiviral host transcriptional response to SARS-CoV-2 by viral load, sex, and age," *PLoS Biology*, vol. 18, no. 9, article e3000849, 2020.

[17] K. A. Overmyer, E. Shishkova, I. J. Miller et al., "Large-scale multi-omic analysis of COVID-19 severity," *Cell Systems*, vol. 12, no. 1, pp. 23–40.e7, 2021.

[18] M. T. McClain, F. J. Constantine, R. Henao et al., "Dysregulated transcriptional responses to SARS-CoV-2 in the periphery," *Nature Communications*, vol. 12, no. 1, p. 1079, 2021.

[19] A. Soares-Schanoski, N. Sauerwald, C. W. Goforth et al., "Asymptomatic SARS-CoV-2 infection is associated with higher levels of serum IL-17C, matrix metalloproteinase 10 and fibroblast growth factors than mild symptomatic COVID-19," *Frontiers in Immunology*, vol. 13, article 821730, 2022.

[20] A. C. Collier and H. H. Handsfield, "Acquired immune deficiency syndrome," *The Urologic Clinics of North America*, vol. 11, no. 1, pp. 187–197, 1984.

[21] M. Rotger, K. K. Dang, J. Fellay et al., "Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals," *PLoS Pathogens*, vol. 6, no. 2, article e1000781, 2010.

[22] F. Z. Chowdhury, Z. Ouyang, M. Buzon, B. D. Walker, M. Lichterfeld, and X. G. Yu, "Metabolic pathway activation distinguishes transcriptional signatures of CD8+ T cells from HIV-1 elite controllers," *AIDS*, vol. 32, no. 18, pp. 2669–2677, 2018.

[23] A. Quach, S. Horvath, N. Nemanim et al., "No reliable gene expression biomarkers of current or impending neurocognitive impairment in peripheral blood monocytes of persons living with HIV," *Journal of Neurovirology*, vol. 24, no. 3, pp. 350–361, 2018.

[24] M. Shahmanesh, K. Phillips, M. Boothby, and J. W. Tomlinson, "Differential adipose tissue gene expression profiles in abacavir treated patients that may contribute to the understanding of cardiovascular risk: a microarray study," *PLoS One*, vol. 10, no. 1, article e0117164, 2015.

[25] C. Shih, C. C. Yang, G. Choijilsuren, C. H. Chang, and A. T. Liou, "Hepatitis B Virus," *Trends in Microbiology*, vol. 26, no. 4, pp. 386-387, 2018.

[26] W. Zhou, Y. Ma, J. Zhang et al., "Predictive model for inflammation grades of chronic hepatitis B: large-scale analysis of clinical parameters and gene expressions," *Liver International*, vol. 37, no. 11, pp. 1632–1641, 2017.

[27] Y. Lu, Z. Fang, M. Li et al., "Dynamic edge-based biomarker non-invasively predicts hepatocellular carcinoma with hepatitis B virus infection for individual patients based on blood testing," *Journal of Molecular Cell Biology*, vol. 11, no. 8, pp. 665–677, 2019.

[28] S. M. Wang, L. L. Ooi, and K. M. Hui, "Identification and validation of a novel gene signature associated with the recurrence of human hepatocellular carcinoma," *Clinical Cancer Research*, vol. 13, no. 21, pp. 6275–6283, 2007.

[29] D. Schlossberg, "Acute tuberculosis," *Infectious Disease Clinics of North America*, vol. 24, no. 1, pp. 139–146, 2010.

[30] J. Maertzdorf, M. Ota, D. Repsilber et al., "Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis," *PLoS One*, vol. 6, no. 10, article e26938, 2011.

[31] C. Broderick, J. M. Cliff, J. S. Lee, M. Kaforou, and D. A. J. Moore, "Host transcriptional response to TB preventive therapy differentiates two sub- groups of IGRA-positive individuals," *Tuberculosis (Edinburgh, Scotland)*, vol. 127, article 102033, 2021.

[32] J. G. Burel, A. Singhania, P. Dubelko et al., "Distinct blood transcriptomic signature of treatment in latent tuberculosis infected individuals at risk of developing active disease," *Tuberculosis (Edinburgh, Scotland)*, vol. 131, article 102127, 2021.

[33] L. S. Garcia, "Malaria," *Clinics in Laboratory Medicine*, vol. 30, no. 1, pp. 93–129, 2010.

[34] A. B. W. Boldt, H. van Tong, M. P. Grobusch et al., "The blood transcriptome of childhood malaria," *eBioMedicine*, vol. 40, pp. 614–625, 2019.

[35] C. F. Ockenhouse, W. C. Hu, K. E. Kester et al., "Common and divergent immune response signaling pathways discovered in peripheral blood mononuclear cell gene expression patterns in presymptomatic and clinically apparent malaria," *Infection and Immunity*, vol. 74, no. 10, pp. 5561–5573, 2006.

[36] Y. Idaghdour, J. Quinlan, J. P. Goulet et al., "Evidence for additive and interaction effects of host genotype and infection in malaria," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 42, pp. 16786–16793, 2012.

[37] Y. Kokubo and C. Matsumoto, "Hypertension is a risk factor for several types of heart disease: review of prospective studies," *Advances in Experimental Medicine and Biology*, vol. 956, pp. 419–426, 2017.

[38] S. Hannenhalli, M. E. Putt, J. M. Gilmore et al., "Transcriptional genomics associates FOX transcription factors with human heart failure," *Circulation*, vol. 114, no. 12, pp. 1269–1276, 2006.

[39] S. Ameling, L. R. Herda, E. Hammer et al., "Myocardial gene expression profiles and cardiodepressant autoantibodies predict response of patients with dilated cardiomyopathy to immunoadsorption therapy," *European Heart Journal*, vol. 34, no. 9, pp. 666–675, 2013.

[40] R. Suresh, X. Li, A. Chiriac et al., "Transcriptome from circulating cells suggests dysregulated pathways associated with long-term recurrent events following first-time myocardial infarction," *Journal of Molecular and Cellular Cardiology*, vol. 74, pp. 13–21, 2014.

[41] E. D. Muse, E. R. Kramer, H. Wang et al., "A whole blood molecular signature for acute myocardial infarction," *Scientific Reports*, vol. 7, no. 1, p. 12268, 2017.

[42] G. Gobbi, C. Carubbi, G. M. Tagliazucchi et al., "Sighting acute myocardial infarction through platelet gene expression," *Scientific Reports*, vol. 9, no. 1, p. 19574, 2019.

[43] E. Witt, E. Hammer, M. Dörr et al., "Correlation of gene expression and clinical parameters identifies a set of genes reflecting LV systolic dysfunction and morphological alterations," *Physiological Genomics*, vol. 51, no. 8, pp. 356–367, 2019.

[44] M. Kiliszek, B. Burzynska, M. Michalak et al., "Altered gene expression pattern in peripheral blood mononuclear cells in patients with acute myocardial infarction," *PLoS One*, vol. 7, no. 11, article e50054, 2012.

[45] C. Cheadle, A. E. Berger, S. C. Mathai et al., "Erythroid-specific transcriptional changes in PBMCs from pulmonary hypertension patients," *PLoS One*, vol. 7, no. 4, article e34951, 2012.

[46] F. García-Río, B. Alcázar-Navarrete, D. Castillo-Villegas et al., "Biological biomarkers in respiratory diseases," *Archivos de Bronconeumología*, vol. 58, no. 4, pp. 323–333, 2022.

[47] T. M. Bahr, G. J. Hughes, M. Armstrong et al., "Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease," *American Journal of Respiratory Cell and Molecular Biology*, vol. 49, no. 2, pp. 316–323, 2013.

[48] R. Faner, N. Gonzalez, T. Cruz, S. G. Kalko, and A. Agustí, "Systemic inflammatory response to smoking in chronic obstructive pulmonary disease: evidence of a gender effect," *PLoS One*, vol. 9, no. 5, article e97491, 2014.

[49] R. C. Langan and A. J. Goodbred, "Office spirometry: indications and interpretation," *American Family Physician*, vol. 101, no. 6, pp. 362–368, 2020.

[50] R. Sellamuthu, C. Umbright, J. R. Roberts et al., "Molecular insights into the progression of crystalline silica-induced pulmonary toxicity in rats," *Journal of Applied Toxicology*, vol. 33, no. 4, pp. 301–312, 2013.

[51] V. R. Mas, D. G. Maluf, K. J. Archer et al., "Genes involved in viral carcinogenesis and tumor initiation in hepatitis C virus-induced hepatocellular carcinoma," *Molecular Medicine*, vol. 15, no. 3-4, pp. 85–94, 2009.

[52] S. P. Fletcher, D. J. Chin, Y. Ji et al., "Transcriptomic analysis of the woodchuck model of chronic hepatitis B," *Hepatology*, vol. 56, no. 3, pp. 820–830, 2012.

[53] A. Scherer, O. P. Günther, R. F. Balshaw et al., "Alteration of human blood cell transcriptome in uremia," *BMC Medical Genomics*, vol. 6, no. 1, p. 23, 2013.

[54] P. C. Grayson, S. Eddy, J. N. Taroni et al., "Metabolic pathways and immunometabolism in rare kidney diseases," *Annals of the Rheumatic Diseases*, vol. 77, no. 8, pp. 1226–1233, 2018.

[55] L. H. Mariani, S. Martini, L. Barisoni et al., "Interstitial fibrosis scored on whole-slide digital imaging of kidney biopsies is a predictor of outcome in proteinuric glomerulopathies," *Nephrology, Dialysis, Transplantation*, vol. 33, no. 2, pp. 310–318, 2018.

[56] I. Arijs, G. de Hertogh, K. Lemaire et al., "Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment," *PLoS One*, vol. 4, no. 11, article e7984, 2009.

[57] I. Nookaew, K. Thorell, K. Worah et al., "Transcriptome signatures in helicobacter pylori-infected mucosa identifies acidic mammalian chitinase loss as a corpus atrophy marker," *BMC Medical Genomics*, vol. 6, no. 1, p. 41, 2013.

[58] J. Refardt, "Diagnosis and differential diagnosis of diabetes insipidus: update," *Best Practice & Research Clinical Endocrinology & Metabolism*, vol. 34, no. 5, article 101398, 2020.

[59] E. C. Kaizer, C. L. Glaser, D. Chaussabel, J. Banchereau, V. Pascual, and P. C. White, "Gene expression in peripheral blood mononuclear cells from children with diabetes," *The Journal of Clinical Endocrinology and Metabolism*, vol. 92, no. 9, pp. 3705–3711, 2007.

[60] F. H. van Tienen, S. F. E. Praet, H. M. de Feyter et al., "Physical activity is the key determinant of skeletal muscle mitochondrial function in type 2 diabetes," *The Journal of Clinical Endocrinology and Metabolism*, vol. 97, no. 9, pp. 3261–3269, 2012.

[61] S. Gao, N. Wolanyk, Y. Chen, S. Jia, M. J. Hessner, and X. Wang, "Investigation of coordination and order in transcription regulation of innate and adaptive immunity genes in type 1 diabetes," *BMC Medical Genomics*, vol. 10, no. 1, p. 7, 2017.

Hindawi
Disease Markers
Volume 2022, Article ID 8347125, 26 pages
https://doi.org/10.1155/2022/8347125

Research Article

Single-Cell RNA Sequencing Reveals the Role of Epithelial Cell Marker Genes in Predicting the Prognosis of Colorectal Cancer Patients

Kai-yu Shen[1], **Bin-yu Chen**[1] and **Wen-cang Gao**[2]

[1]*The Second Clinical Medical College of Zhejiang Chinese Medical University, Hangzhou, Zhejiang, China*
[2]*Department of Oncology, The Second Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, Zhejiang, China*

Correspondence should be addressed to Wen-cang Gao; 20164580@zcmu.edu.cn

Received 13 May 2022; Revised 9 July 2022; Accepted 21 July 2022; Published 1 August 2022

Academic Editor: Qiaoli Wang

Single-cell RNA sequencing (scRNA-seq) is increasingly used in studies on gastrointestinal cancers. This study investigated the prognostic value of epithelial cell-associated biomarkers in colorectal cancer (CRC) using scRNA-seq data. We downloaded and analysed scRNA-seq data from four CRC samples from the Gene Expression Omnibus (GEO), and we identified marker genes of malignant epithelial cells (MECs) using CRC transcriptome and clinical data downloaded from The Cancer Genome Atlas (TCGA) and GEO as training and validation cohorts, respectively. In the TCGA training cohort, weighted gene correlation network analysis, univariate Cox proportional hazard model (Cox) analysis, and least absolute shrinkage and selection operator regression analysis were performed on the marker genes of MEC subsets to identify a signature of nine prognostic MEC-related genes (MECRGs) and calculate a risk score based on the signature. CRC patients were divided into high- and low-risk groups according to the median risk score. We found that the MECRG risk score was significantly correlated with the clinical features and overall survival of CRC patients, and that CRC patients in the high-risk group showed a significantly shorter survival time. The univariate and multivariate Cox regression analyses showed that the MECRG risk score can serve as an independent prognostic factor for CRC patients. Gene set enrichment analysis revealed that the MECRG signature genes are involved in fatty acid metabolism, p53 signalling, and other pathways. To increase the clinical application value, we constructed a MECRG nomogram by combining the MECRG risk score with other independent prognostic factors. The validity of the nomogram is based on receiver operating characteristics and calibration curves. The MECRG signature and nomogram models were well validated in the GEO dataset. In conclusion, we established an epithelial cell marker gene-based risk assessment model based on scRNA-seq analysis of CRC samples for predicting the prognosis of CRC patients.

1. Introduction

Colorectal cancer (CRC) is among the most common malignant tumours of the digestive system worldwide, and its high incidence and mortality are second only to those of lung cancer and breast cancer [1, 2]. Unhealthy lifestyle habits, such as smoking, drinking, and consuming a high-fat diet, have led to an increase in CRC incidence [3], with 1.5% annual increase reported in people aged 30 to 39 years from 2007 to 2016 [4]. At present, tumour node metastasis (TNM) staging, histopathology, and completion of surgical resection are mainly evaluated to determine CRC prognosis, and molecular markers are widely used for CRC diagnosis and treatment. Various medical treatments, including surgery, postoperative adjuvant chemoradiotherapy, and molecular targeted therapy, have been used to treat CRC [5]. CRC symptoms are generally minor in early disease stages, and therefore, patients are often at an advanced stage of the disease by the time they seek medical attention. Furthermore, the optimal treatment timing can easily be missed

in cases of advanced, metastatic, and recurrent CRC, and the effects of conventional treatment are usually inconsistent, resulting in variable prognoses.

As a malignant tumour of epithelial origin, the development and progression of CRC are closely related to epithelial tissue [6, 7]. Normal epithelial cells have antitumor activities and are able to eliminate oncogenic transformed cells by regulating cytoskeletal proteins [8]. According to Lv et al., epithelial cells secrete periostin, which inhibits the growth of gastric cancer cells by stabilising P53 and E-cadherin proteins [9]. Epithelial cells exhibit apical-basal polarity and cell–cell adhesion. Correct regulation of polarity is essential to inhibit tumour growth [10]. Royer and Lu suggested that the malignant transformation of epithelial cells in the presence of oncogene activation is generally closely associated with the loss of cell polarity and disorganisation and that the disruption of epithelial cell polarity promotes epithelial-mesenchymal transition (EMT), which is a key step in the invasion of the surrounding stroma by epithelial tumour cells [11]. In addition, under hypoxic conditions, pulmonary epithelial cells can downregulate connexin (CX)26 and CX43 via the P53 signalling pathway to promote lung carcinogenesis [12]. Epithelial cells are both the structural basis of most organ tissues in the body and the source of most human tumours [13]. With the progressive research on epithelial cells, epithelial cell-related biomarkers have become a research hotspot in recent years. Keratins are widely used as epithelial cell biomarkers in the pathological diagnosis of tumours and in predicting survival prognosis. In CRC patients, decreased expression of keratin (K)8 and K20 is closely related to EMT and suggestive of poor survival and high tumour invasion [14]. Increased serum levels of a cleaved K18 fragment produced by apoptotic epithelial cells suggest a high risk of CRC recurrence within 3 years [15]. However, because of the uncontrollable pathological types, stages, and metastases of CRC [16], using conventional biomarkers to predict prognosis does not achieve adequate results. Therefore, to explore more epithelial cell-related biomarkers with clinical application potential, we used high-resolution omics tools to perform a more accurate analysis of CRC epithelial cells.

Single-cell RNA sequencing (scRNA-seq) allows the construction of a gene-regulatory network at the cellular level [17]. Analysis of the genome, transcriptome, or epigenome of single cells individually or simultaneously enables the detection of gene expression profiles and tracking of cell development at the single-cell level [18]. The RNA-seq technology has been widely used to evaluate the tumour immune microenvironment [19, 20]; however, few studies have applied this method to CRC epithelial cells to predict CRC prognosis.

In this study, we identified marker genes of CRC epithelial cell subsets using scRNA-seq analysis, determined the prognostic significance of nine malignant epithelial cell-(MEC-) related genes (MECRGs) using data from The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO), and integrated the MECRG-based risk score and clinical traits to construct a nomogram for the prediction of CRC patient prognosis.

2. Materials and Methods

2.1. Data Availability. ScRNA-seq data of four CRC samples were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) (GSE161277). Relevant clinical information related to the CRC samples is provided in Supplementary Table S1. The dataset comprises four CRC samples, four adenoma samples, three normal tissue samples, one paracancer sample, and one blood sample [21]. Tissue samples other than the CRC samples were not considered in this study. Transcriptome data of 473 CRC samples and 41 nontumour tissues were obtained from the TCGA database (https://portal.gdc.cancer.gov/), including 435 CRC samples with matched clinical data. The extracted clinical information included sex, age, and stage. TCGA-CRC samples with complete clinical information were used to construct the training set used to develop a prediction score. A total of 232 CRC samples with complete clinical information obtained from the GEO database (GSE17538) were used as the external validation set.

2.2. scRNA-seq Data Processing. The scRNA-seq data of the four CRC samples were processed using the R language (v4.1.0; https://www.r-project.org/) and the "Seurat" package [22] (https://cran.r-project.org/web/packages/Seurat/index.html). The "harmony" package was used to remove batch effects from the scRNA-seq data of the four CRC samples. First, we determined the percentage of mitochondrial genes in each cell using the "PercentageFeatureSet" function with the parameter set to pattern = "^MT-". Subsequently, using the "subset" function, genes expressed in <10 cells and cells expressing <200 genes were eliminated. Further, we excluded noncells and cell aggregates. Cell samples with a mitochondrial gene proportion < 15% were included in the subsequent analysis and logarithmically normalized. Principal component analysis (PCA) was used for unsupervised clustering, and the "JackStraw" function was used to determine and visualise the number of principal components. We used nonlinear dimension reduction (t-distributed stochastic neighbour embedding; t-SNE) clustering and the "FindAllMarkers" function (parameters: minimum% = 0.3, log function threshold = 0.25) to identify marker genes based on between-cluster differences. The cell clusters were then annotated based on reported cell-specific marker genes [23–29].

2.3. Analysis of Chromosome Copy Number Variation (CNV) and Screening of MEC Marker Genes. To distinguish epithelial cell malignancy, we identified chromosome CNV in each sample using the "infercnv" package [30] (https://github.com/broadinstitute/inferCNV; default parameters). Using immune cells as a reference [31], we calculated CNV scores for epithelial cell subsets and defined the epithelial cell subset showing a median CNV score > 1,000 as the MEC subset, which was used for subsequent analysis. Differences in CNV scores between epithelial cell clusters were compared using the Kruskal–Wallis test. We then identified marker genes in the MEC subset based on a $|\log 2(fold\ change)| > 0.5$ and a false discovery rate (FDR) < 0.01.

2.4. Analysis of Differential Gene Enrichment of Epithelial Cell Subsets. The potential biological mechanisms of the MEC marker genes were determined using the "clusterProfiler" (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) and "org.Hs.eg.db" (https://www.bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html) packages. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed, using a q value < 0.05 to determine statistically significant enrichment.

2.5. MEC Marker Gene Analysis Using Weighted Gene Correlation Network Analysis (WGCNA). WGCNA can be used to find modules of correlated genes and identify disease-related biomarkers. We used the WGCNA R package [32] (https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/) to identify MEC marker genes related to CRC. We generated a similarity matrix between MEC marker genes using the Pearson correlation analysis and then calculated an adjacency matrix and constructed a topological overlap matrix. Next, we plotted a tree diagram of modules and clustered closely related MEC marker genes within the modules using hierarchical clustering. The MEC marker genes in the final correlation module ($P < 0.05$) were used for subsequent analysis.

2.6. Construction and Validation of the MECRG Signature. The univariate Cox analysis was used to find MECRGs significantly related to overall survival (OS) in the TCGA-CRC cohort, and a forest map was plotted. MECRGs were identified using least absolute shrinkage and selection operator (LASSO) analysis, using 10-fold cross-validation and 1,000 iterations to determine the minimum value of the penalty parameter (λ) and construct a MECRG signature. The regression coefficients of the nine MECRGs thus identified were calculated using the "predict" function. The following formula was used to calculate the risk score:

$$\text{MECRG risk score} = \sum X, G * \text{coef } G, \tag{1}$$

where "coef G" is the regression coefficient and "X G" describes the expression levels of the MECRGs. Patients from the TCGA-CRC training cohort were divided into high- and low-risk groups according to the median value of the MECRG risk score. A Kaplan–Meier survival curve was used to explore differences in survival and prognosis between the two groups. We then used a receiver operating characteristic (ROC) curve to evaluate the predictive value of the MECRG signature. Finally, the signature was validated using the GSE17538 GEO dataset.

2.7. Analysis of the Prognostic Accuracy of the MECRG Risk Score. Combining the MECRG risk score with clinical features, we used univariate and multivariate Cox analyses to assess whether the risk score could serve as an independent prognostic factor. Using the training and validation sets, we then performed a survival analysis of the MECRG signature for patients of different clinical subgroups. Additionally, we identified the relationships between MECRG risk groups and clinical traits (including sex, age, and TNM stage) and generated a heat map.

2.8. Gene Set Enrichment Analysis (GSEA) of the MECRG Signature. Patients in the TCGA-CRC training cohort were grouped according to the median value of the risk score and we used GSEA to evaluate the functions of and signalling pathways associated with MECRGs in the high- and low-risk groups, using $P < 0.05$ as a threshold. The "AUCell" package (https://www.rdocumentation.org/packages/msigdbr/versions/7.4.1) was used to present the results of enrichment analysis according to cell subsets.

2.9. Construction of a MECRG Nomogram. Using the training set, we integrated the three independent prognostic factors of age, stage, and MECRG risk group to plot a MECRG nomogram capable of predicting the 1-, 3-, and 5-year OS of CRC patients for clinical application. ROC and calibration curves were used to evaluate the predictive value of the MECRG nomogram, and feasibility was confirmed by external validation using the GSE17538 GEO dataset.

2.10. Statistical Analysis. The "survival" (https://cran.r-project.org/web/packages/survival/index.html) and "survminer" (https://cran.r-project.org/web/packages/survminer/index.html) packages were used to construct Kaplan–Meier survival curves. The log-rank test was used to determine significant differences in survival between the high- and low-risk groups according to the training and validation datasets. LASSO regression was performed using the "glmnet" package (https://cran.r-project.org/web/packages/glmnet/index.html), and the "timeROC" package (https://cran.r-project.org/web/packages/timeROC/index.html) was used to generate the ROC curve to evaluate model accuracy. The univariate and multivariate Cox regression analyses were used to determine the independent predictors of OS outcomes in CRC patients. The "rms" package (https://cran.r-project.org/web/packages/rms/) was used to construct the nomogram model. Comparisons between two groups were made using the Wilcoxon rank sum test, and comparisons between multiple groups were made using the Kruskal–Wallis test. Correlation analysis was conducted using the Pearson method. $P < 0.05$ was considered significant.

3. Results

3.1. CRC scRNA-seq Data Analysis. We obtained 15,465 cells from the four CRC samples. After applying the screening criteria, 8,798 high-quality cell samples were obtained. Gene numbers (nFeature_RNA), sequencing depth (nCount_RNA), and mitochondrial gene percentage (percent.mt) are shown in Supplementary Figure S1. The Pearson correlation coefficient between gene count and sequencing depth was 0.92 (Supplementary Figure S2), suggesting a positive correlation. PCA used to classify the high-quality cells identified 40 principal components (Supplementary Figure S3), and t-SNE of the first 11 principal components ($P < 0.05$) (Supplementary Figure S4) allowed visualisation of the high-dimensional CRC scRNA-seq data and the distribution of the cell subsets, as well as classification of

the cells into 18 subclasses (Figures 1(a) and 1(b)). Among the 18 subclasses, clusters 0, 4, 7, 9, 10, and 11 were identified as epithelial cell subtypes 1, 2, 3, 4, 5, and 6, respectively, based on the presence of epithelial cell marker genes (*EPCAM*, *KRT19*, and *CDH1*), clusters 2, 12, 13, 15, and 17 were identified as B cell subsets (*CD79A*, *MS4A1*, and *CD19*), clusters 1, 3, 6, and 8 were identified as T cell subsets (*CD3D*, *CD8A*, and *CD3E*), clusters 5 and 16 were identified as macrophage subsets (*CD14*, CD68, and CD163), and cluster 14 was identified as an endothelial cell subset (*IL3RA*, *SERPINF1*, and *CCDC88A*). For the purpose of selecting epithelial cell subsets with high malignancy, we did not merge the epithelial cell subsets. Based on the above annotation effects, we summarised the final cell subpopulation annotations in Figure 2(a). In addition, we calculated the proportions of the five cell types in the four CRC samples, which revealed that the epithelial cell subpopulation accounted for a relatively large fraction (Figure 2(b)). The five most strongly expressed marker genes in each cell subset are shown in Figure 2(c).

3.2. Evaluation of Chromosome CNV in the Epithelial Cell Subsets. We next determined the chromosome CNV in each sample based on the transcriptome data to evaluate the degree of malignancy in the epithelial cell subsets. We observed low CNV in the immune cell subsets (macrophages, B cells, and T cells) in control samples, whereas high CNV was observed in epithelial cells. Chromosome amplification mainly occurred in chromosomes 7, 8, 9, 12, 13, 16, 19, 20, and 21, and deletions were most prevalent in chromosomes 4, 6, 8, 12, 14, 15, 17, 18, 19, and 22 (Figure 3(a)). The MEC subpopulation was screened using a median CNV score > 1,000 as the threshold. Notably, the CNV scores for epithelial cell subtypes 1 through 5 were more significant than epithelial cell subtype 6 (Figure 3(b) and Supplementary Table S2), suggesting a higher degree of malignancy of CRC lesions associated with these cell subsets. Therefore, we defined these as MEC subsets, and, using $|\log 2(fold\,change)| > 0.5$ and FDR < 0.01 as thresholds, we identified 1,259 marker genes (Supplementary Table S3), and we speculated that 1259 MEC marker genes function differently in CRC than in normal colonic epithelial cells and therefore require further analysis.

3.3. GO and KEGG Enrichment Analyses of MEC Subset Marker Genes. We performed GO function (Figure 4(a)) and KEGG pathway (Figure 4(b)) enrichment analyses of the 1,259 marker genes in the five MEC subsets to determine their possible biological functions. We found that the differentially expressed genes in the MEC subsets were mainly involved in processes associated with ATP metabolism, focal adhesion, formation of cell–substrate junctions, cadherin binding, formation of adherens junctions, reactive oxygen species (ROS), and thyroid cancer. Notably, cell–substrate junction formation, calmodulin binding, and adhesion junctions are closely related to the characteristics of epithelial cells themselves, ROS is strongly associated with tumor progression, and as thyroid cancer and CRC are both epithelial-derived malignancies, we hypothesised that common epithelial cell-related genes might be involved in the development of both. These results tentatively suggest that marker genes of the MEC subpopulation are involved in the occurrence and progression of CRC mainly through the above-mentioned biological mechanisms.

3.4. WGCNA of MEC Marker Genes. Using the TCGA-CRC cohort, we performed WGCNA of the expression profiles of the identified 1259 marker genes. We clustered the MEC genes into modules associated with clinical traits ("tumour" and "normal") based on a soft threshold of $\beta = 7$. As shown in Figure 5, to prevent meaningful MEC marker genes from being missed, with the exception of the turquoise module ($P > 0.05$), all genes in the remaining modules were significantly associated with clinical characteristics ("tumour" and "normal") ($P < 0.05$) and were used in subsequent analyses. To build a clinical prediction model by linking to the clinical characteristics of CRC patients, we initially identified 787 genes significantly associated with clinical characteristics ("tumour" and "normal") and further identified MEC marker genes associated with survival prognosis.

3.5. Establishment and Validation of a MECRG Signature for Predicting CRC Patient Survival. The univariate Cox analysis (Figure 6(a)) and LASSO regression analysis (Supplementary Figure S7) of the TCGA-CRC training cohort identified 47 MEC marker genes capable of predicting OS, the regression coefficients of the nine MECRGs were calculated using the "predict" function, and nine MECRGs were identified based on the minimum λ ($\lambda = 0.01970$): galectin 2 (*LGALS2*), glycerophosphodiester phosphodiesterase 1 (*GDE1*), monocyte chemoattractant protein 1 (*MPC1*), bone marrow stromal cell antigen 2 (*BST2*), tropomyosin 2 (*TPM2*), PRELI domain-containing 2 (*PRELID2*), G protein subunit gamma 5 (*GNG5*), calcyphosin (*CAPS*), and calcium voltage-gated channel subunit alpha 1D (*CACNA1D*). Validation using the GSE17538 data identified *BST2*, *TPM2*, and *CAPS* as risk genes (hazard ratio (HR) > 1) and *LGALS2*, *GDE1*, *MPC1*, *GNG5*, *PRELID2*, and *CACNA1D* as protective genes (HR < 1). The MECRG risk score was defined as follows: (−0.13614 ∗ *expression* of *LGALS2*) + (−0.60881 ∗ *expression* of *GDE1*) + (−0.49609 ∗ *expression* of *MPC1*) + (−0.20530 ∗ *expression* of *GNG5*) + (0.05613 ∗ *expression* of *BST2*) + (0.12894 ∗ *expression* of *TPM2*) + (−0.16624 ∗ *expression* of *PRELID2*) + (0.48635 ∗ *expression* of *CAPS*) + (−0.83941 ∗ *expression* of *CACNA1D*). The final risk scores for all CRC samples in the training and validation cohorts are shown in Supplementary Tables S4 and S5. The MECRGs were predominantly distributed in the epithelial cell subsets (Figure 6(b)).

Next, we divided the patients into high- and low-risk groups according to the median MECRG risk score. The Kaplan–Meier survival analysis demonstrated that in the TCGA-CRC training and GSE17538 validation cohorts, the survival time of the high-risk group was shorter than that of the low-risk group ($P < 0.001$) (Figures 6(c) and 6(d)). In both cohorts, an increase in the risk score was

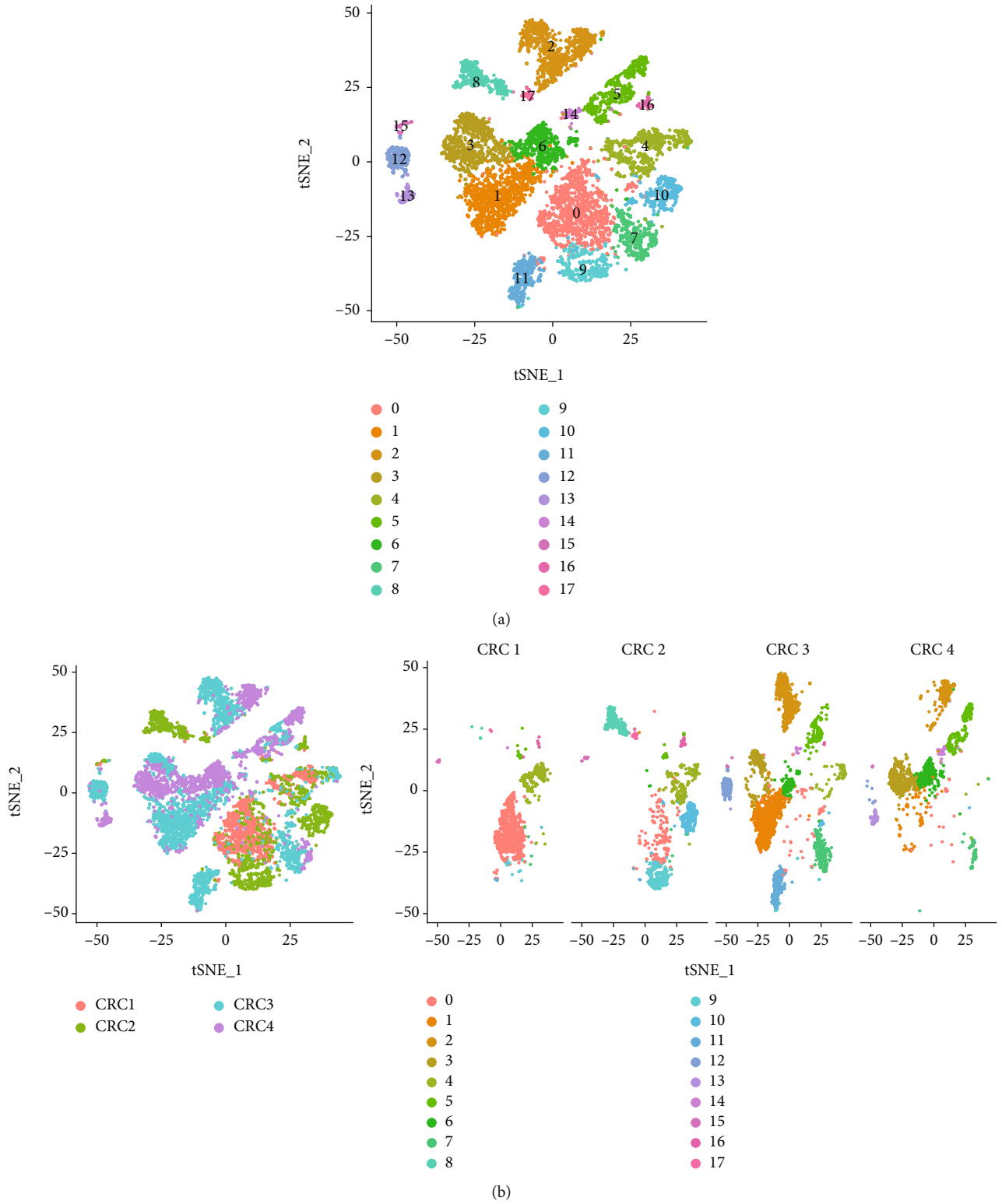


Figure 1: Characterisation of scRNA-seq from 15,465 cells. (a) The cells were classified into 18 subsets using the t-SNE algorithm. (b) Distribution ratios of the cell subsets in the four CRC samples.

accompanied by an increase in patient mortality (Figures 6(e) and 6(f)). The area under the ROC curve (AUC) at 1, 3, and 5 years was 0.746, 0.726, and 0.710, respectively, in the training cohort, and 0.618, 0.668, and 0.658, respectively, in the validation cohort (Figures 6(g) and 6(h)). These data indicated that the MECRG risk score shows good sensitivity and specificity for predicting the prognosis of CRC patients.

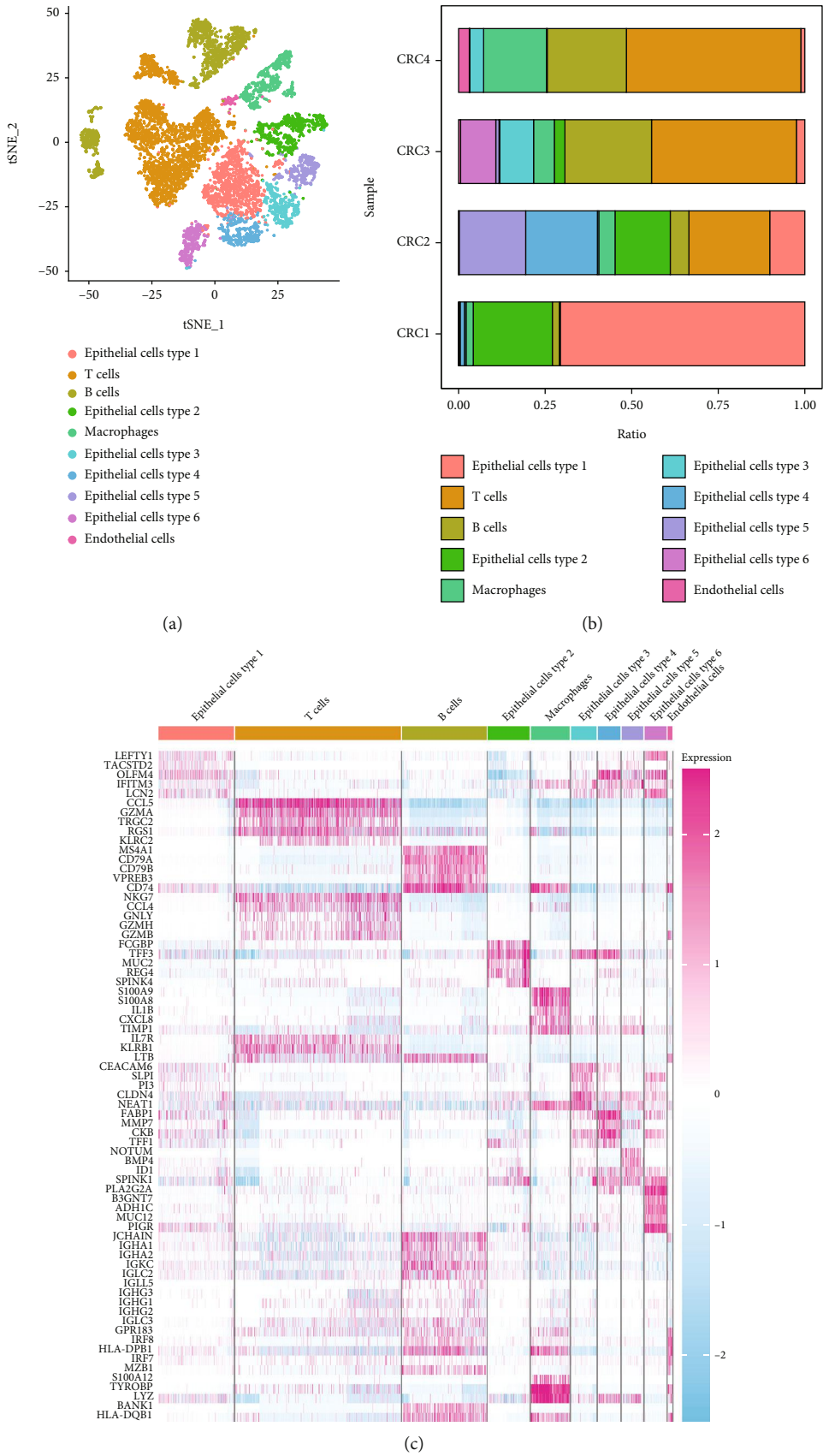
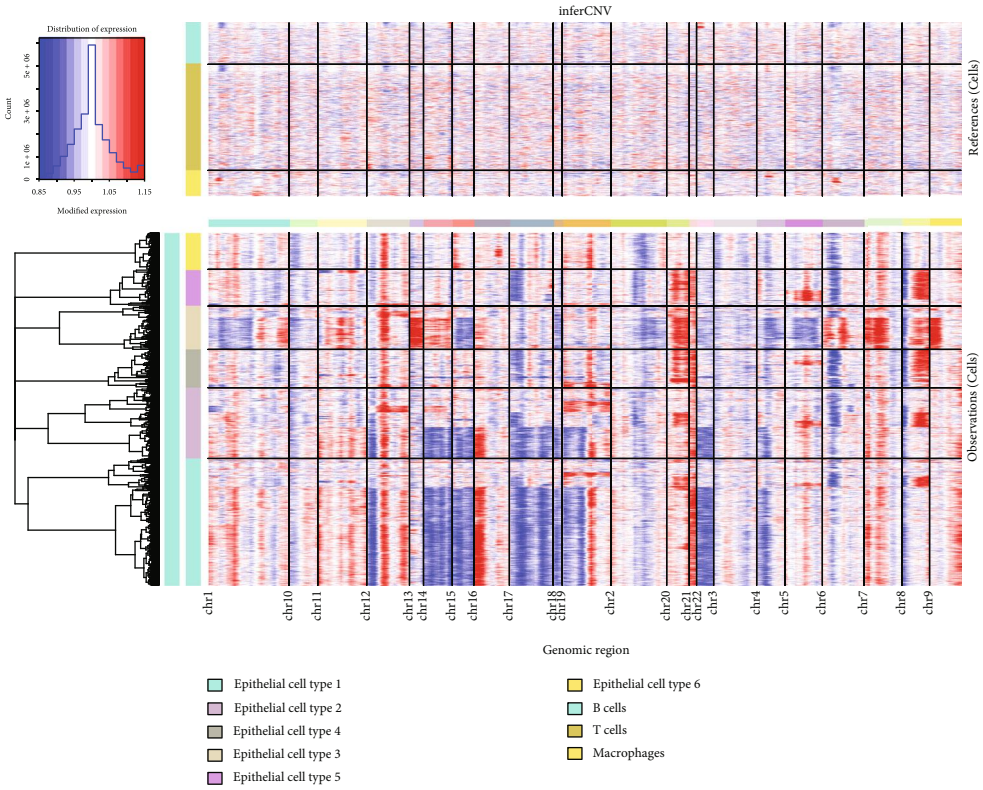
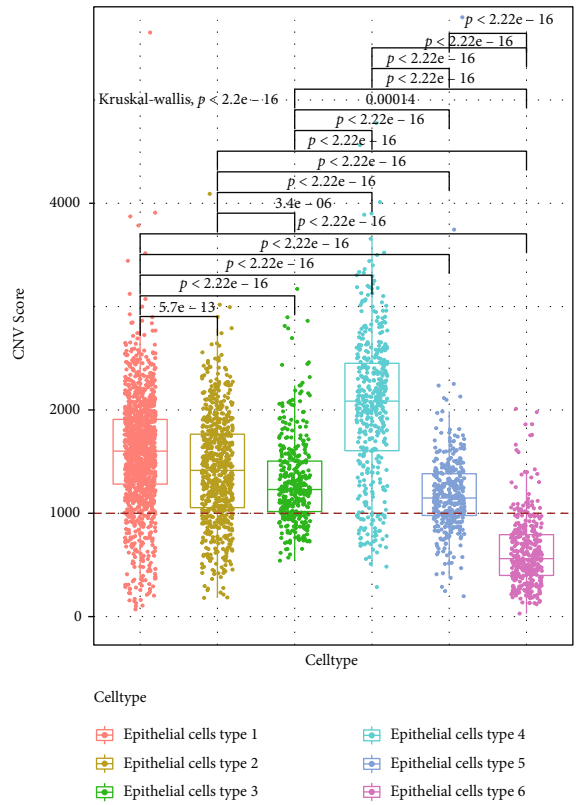


Figure 2: Characteristics of the cell subsets. (a) Annotated cell subsets. (b) Proportions of the various types of cells in the four CRC samples. (c) Heat map of the five most differentially expressed marker genes in each cell subset.

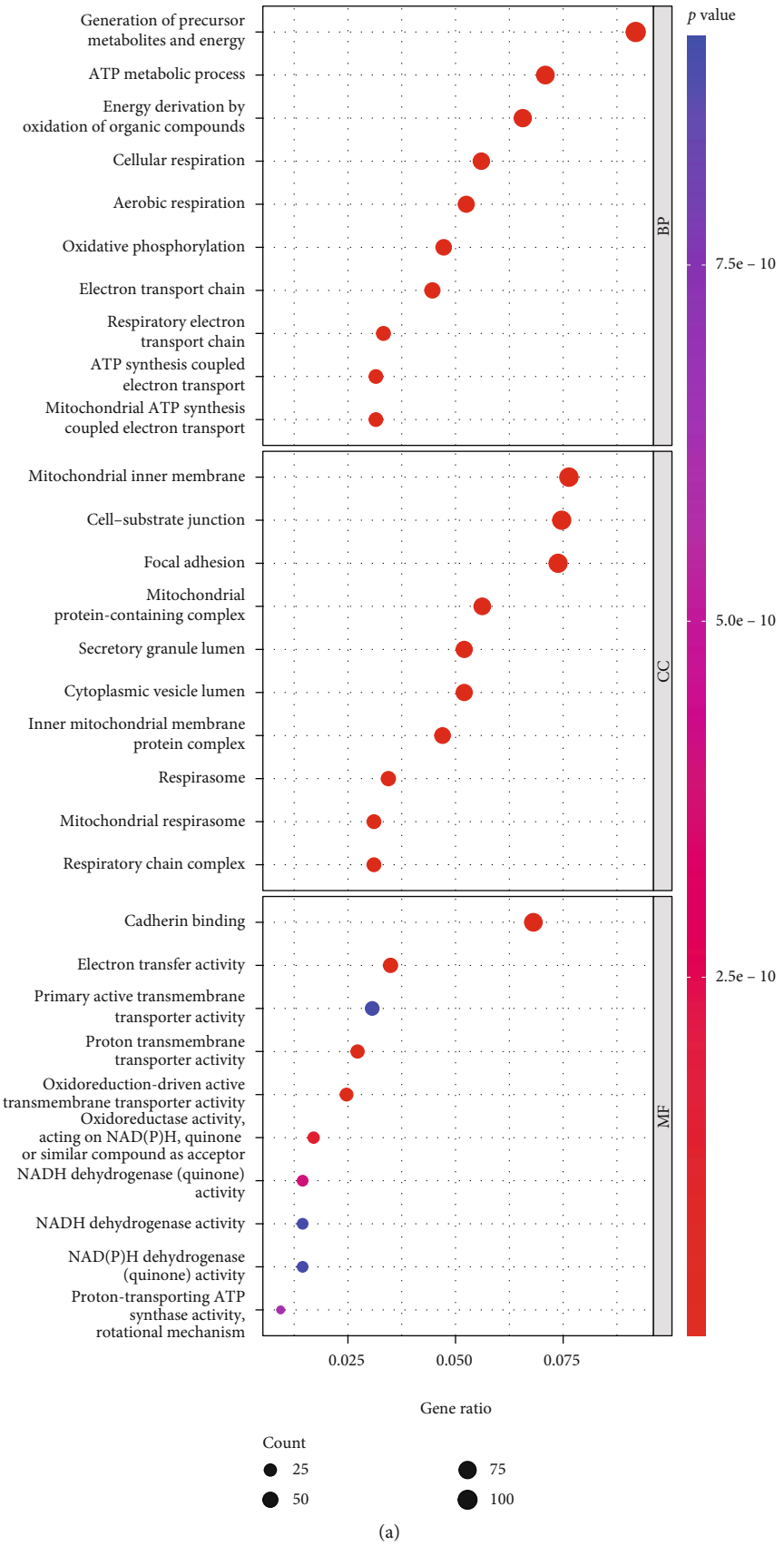


(a)



(b)

Figure 3: CNV analysis of epithelial cells from CRC patients. (a) Heat map of CNV in epithelial cells from the four CRC samples. (b) CNV score distribution among different epithelial cell subtypes.



(a)

Figure 4: Continued.

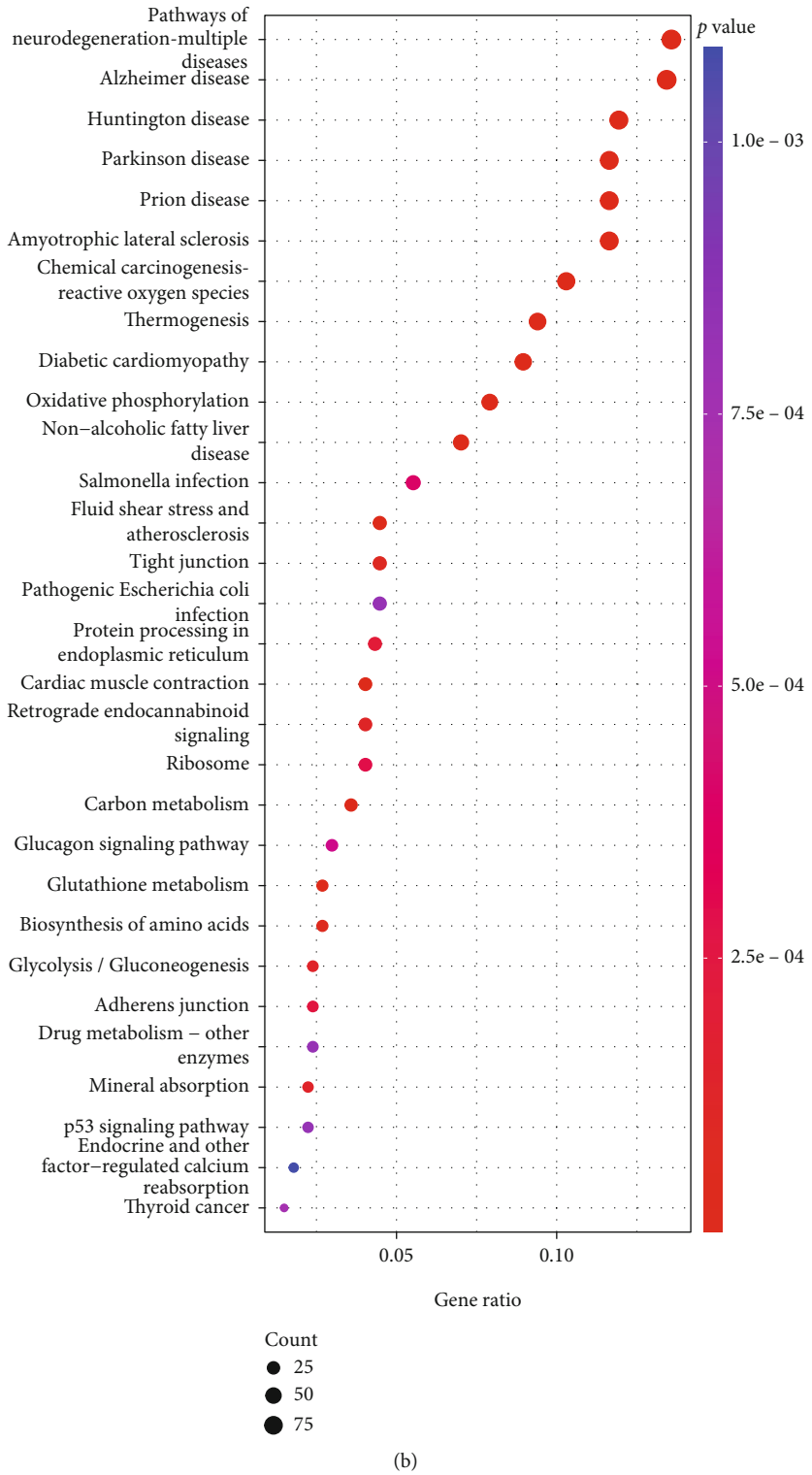


Figure 4: Functional enrichment analysis of marker genes in the MEC subsets. Results of GO function (a) and KEGG pathway (b) enrichment analyses.

3.6. Independent Prognostic Value of the MECRG Risk Score. To determine whether the MECRG risk score can predict prognosis independently of traditional clinical features, such as age, sex, and TNM stage, we performed univariate and multivariate Cox regression analyses. In the training cohort, age (HR = 1.029 and 1.040, 95% confidence interval (CI): 1.010–1.048 and 1.021–1.060, respectively; $P = 0.002$ and $P < 0.001$), TNM stage (HR = 2.068 and 2.106, 95% CI: 1.628–2.627 and 1.642–2.701, respectively; $P < 0.001$), and the MECRG risk score (HR = 1.294 and 1.234, 95% CI:

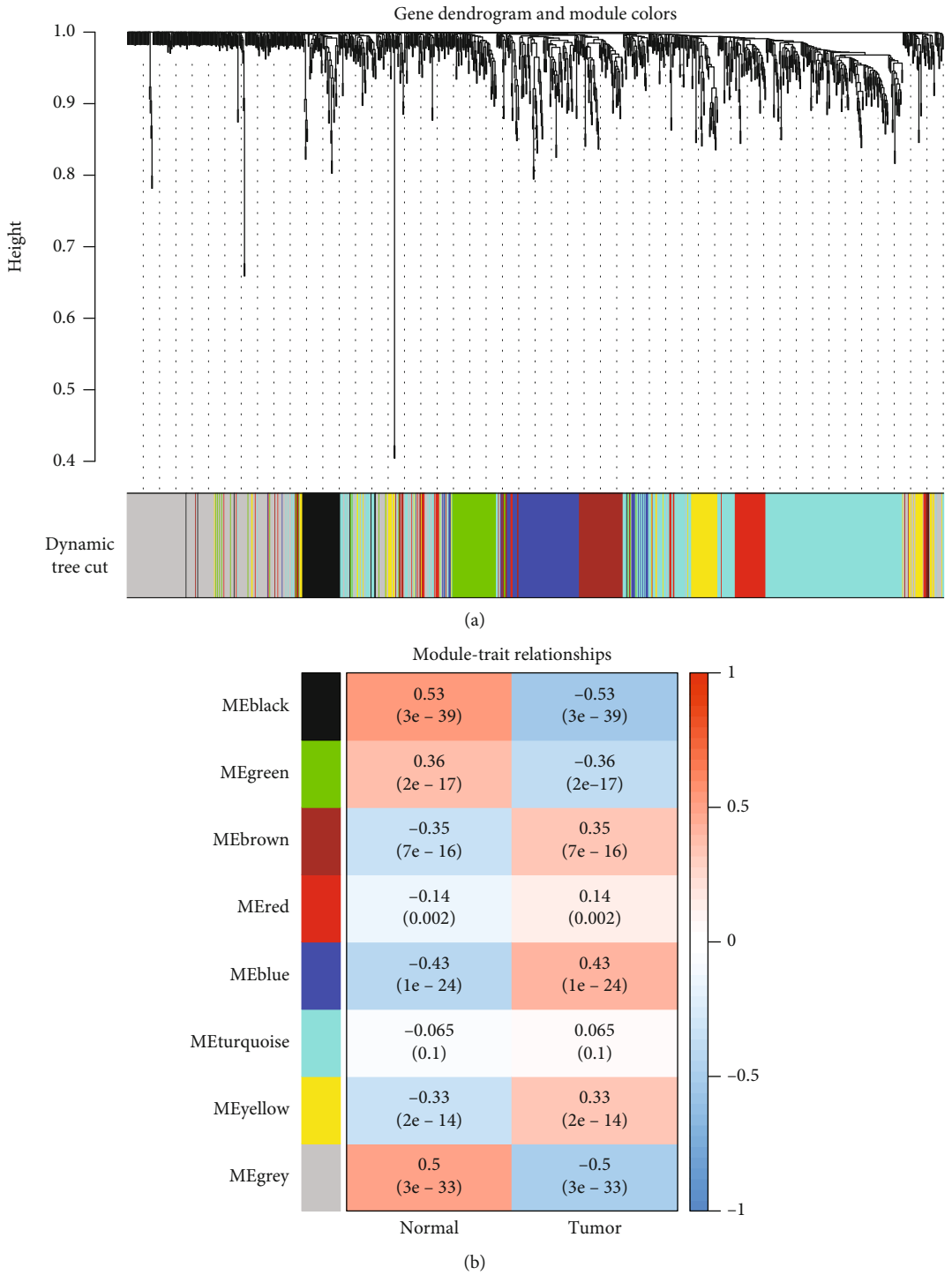
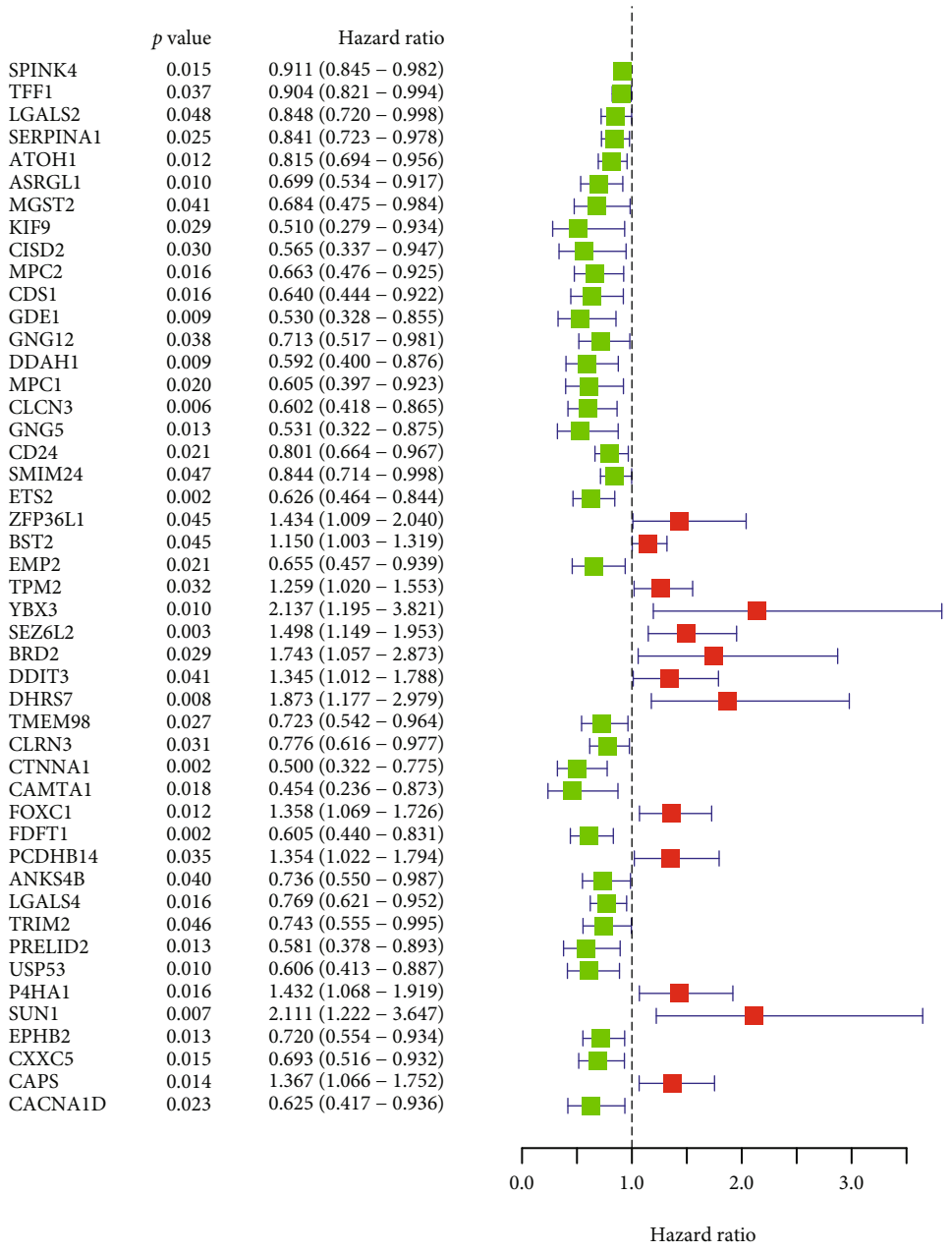


Figure 5: MEC marker genes associated with CRC. (a) Dendrogram of MEC subset marker genes obtained by WGCNA according to colour. (b) Correlations between characteristic genes of different modules and CRC.

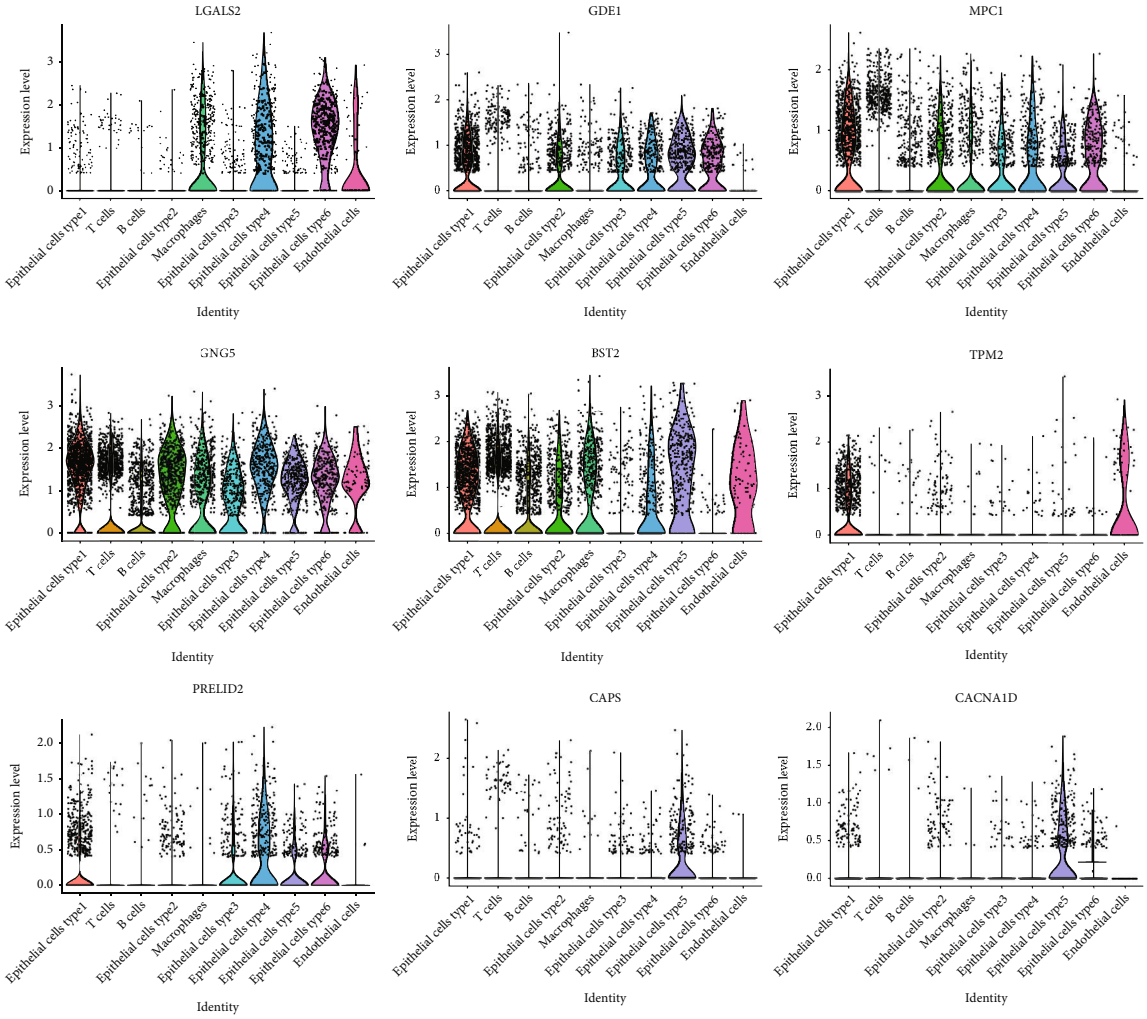
1.200–1.396 and 1.139–1.336, respectively; $P < 0.001$) were independent predictors of OS. Analysis of the GSE17538 validation cohort confirmed that TNM stage (HR = 2.712 and 3.027, 95% CI: 2.077–3.541 and 2.323–3.945, respectively; $P < 0.001$) and the MECRG risk score (HR = 1.138 and 1.298, 95% CI: 1.005–1.289 and 1.127–1.496, respectively; $P = 0.042$ and $P < 0.001$) were independent predictors of OS (Figure 7). The above results indicated that the MECRG risk score was an independent prognostic factor in both the training and validation sets.

3.7. Correlation between MECRG Risk Groups and Clinical Features. We evaluated the correlation between clinical traits and risk groups among the 435 patients in the TCGA-CRC training cohort and the 232 CRC patients in the GSE17538 validation cohort (with complete clinical information). The



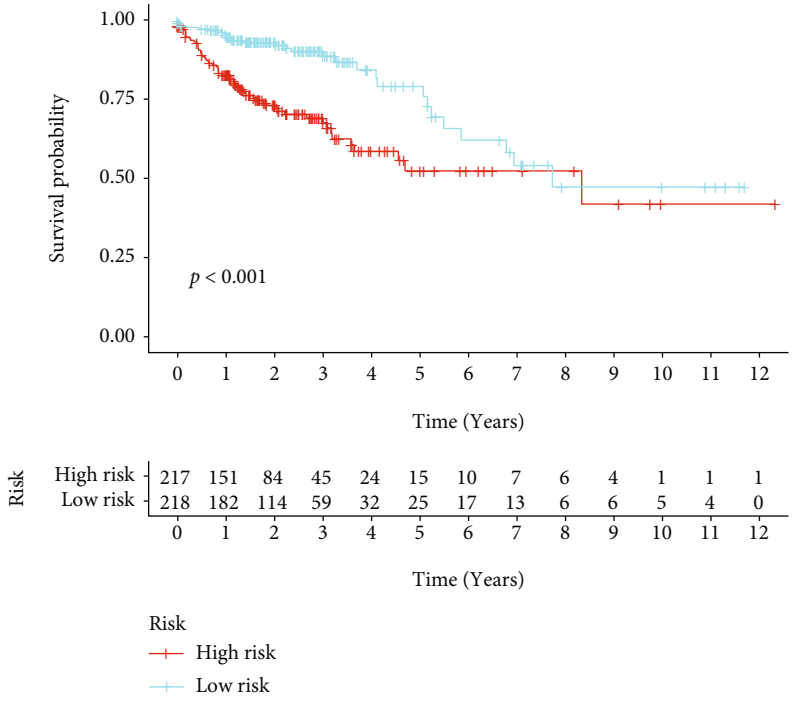
(a)

Figure 6: Continued.

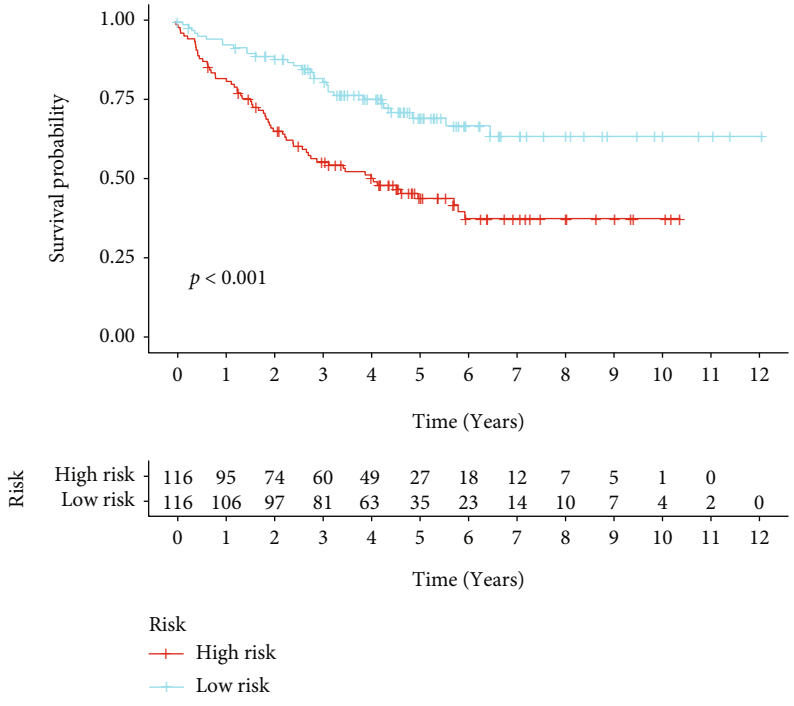


(b)

Figure 6: Continued.



(c)



(d)

Figure 6: Continued.

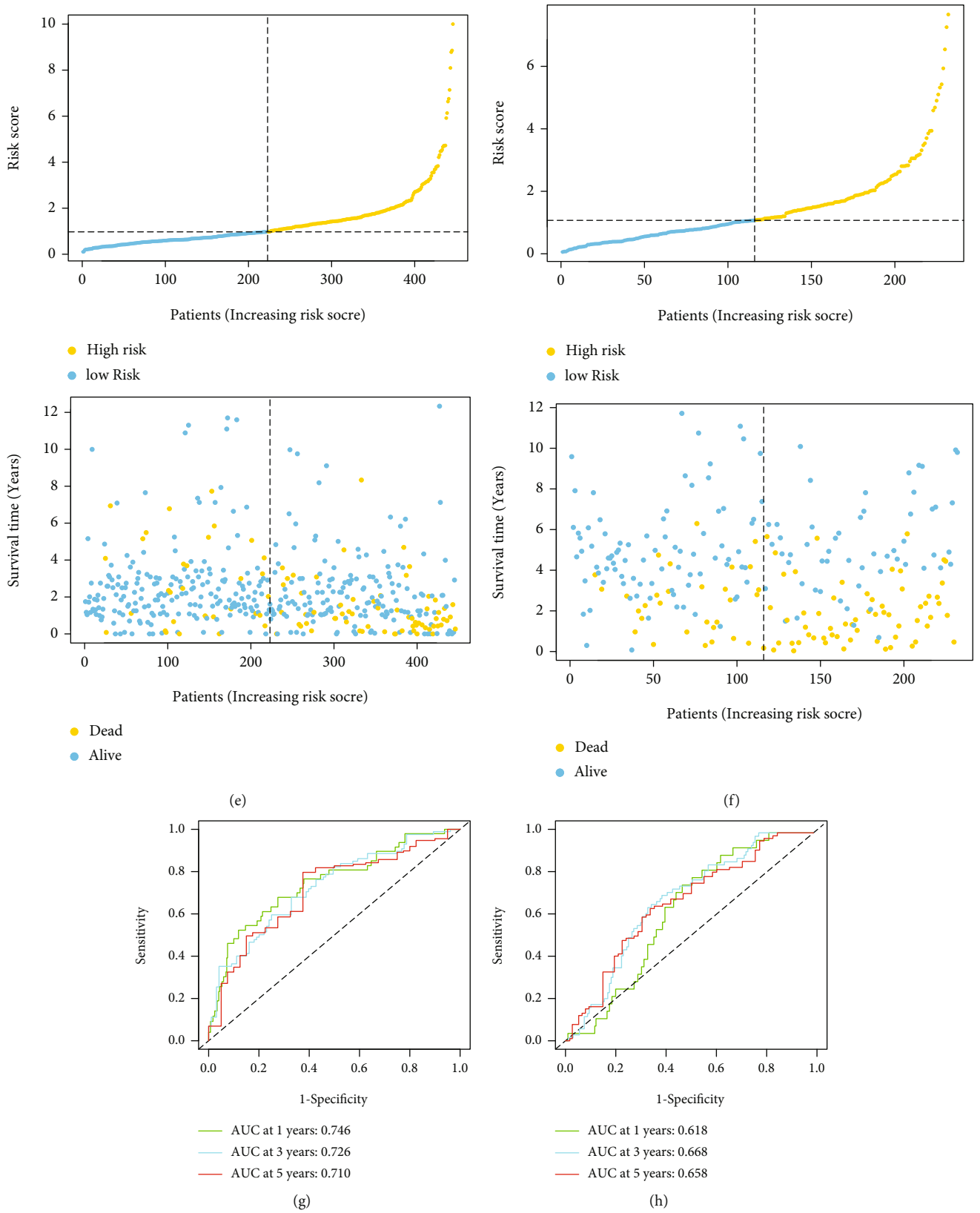
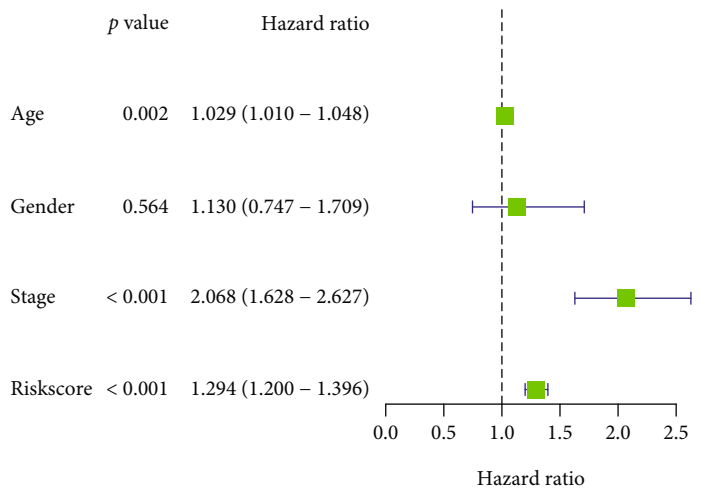
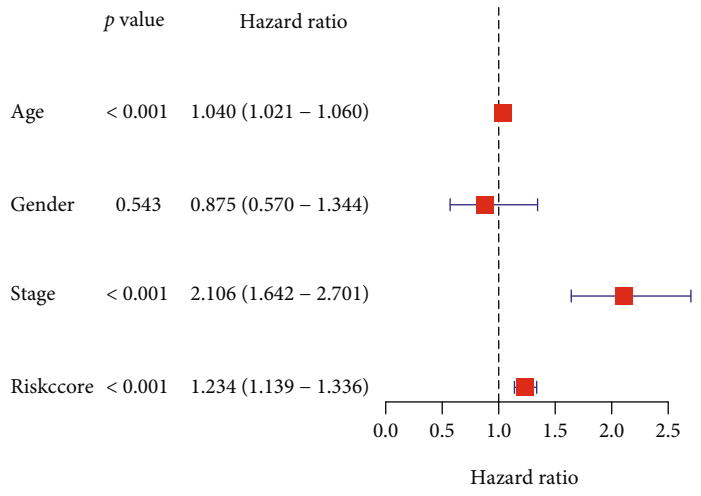


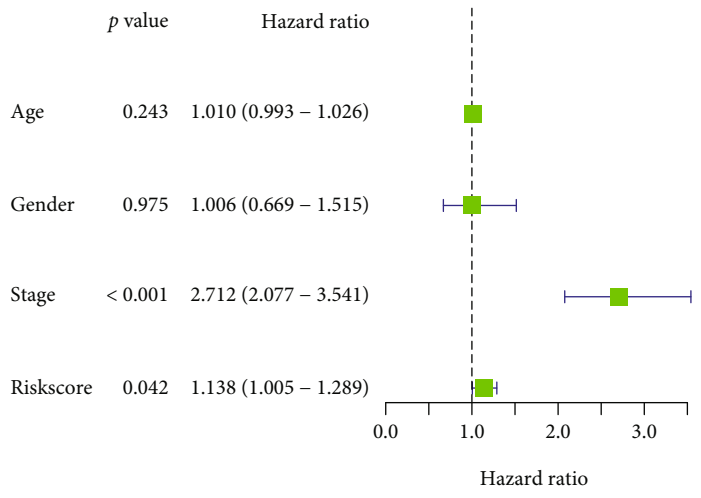
Figure 6: Identification and validation of MECRGs in the TCGA-CRC training and GSE17538 validation cohorts. (a) Univariate Cox analysis was used to screen MECRGs with prognostic significance. (b) Distribution of MECRGs in cell subsets. (c and d) Kaplan–Meier survival curve showing the prognostic value of the MECRG signature in the training cohort (c) and the validation cohort (d). (e and f) Distribution of the MECRG risk scores and survival status of CRC patients in the training cohort (e) and validation cohort (f). ROC curve representing the efficiency of the MECRG signature in predicting 1-, 3-, and 5-year OS in CRC patients in the training cohort (g) and validation cohort (h).



(a)



(b)



(c)

Figure 7: Continued.

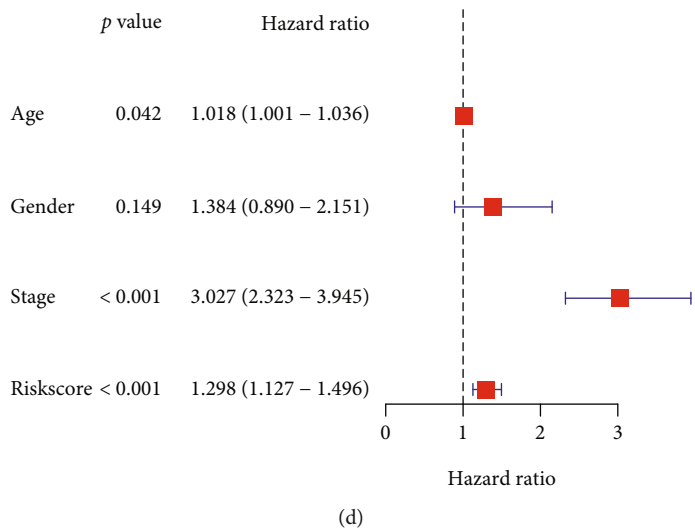


Figure 7: Independent prognostic value of the MECRG signature. Univariate and multivariate Cox analyses of the MECRG risk score in the TCGA-CRC training cohort (a and b) and GSE17538 validation cohort (c and d).

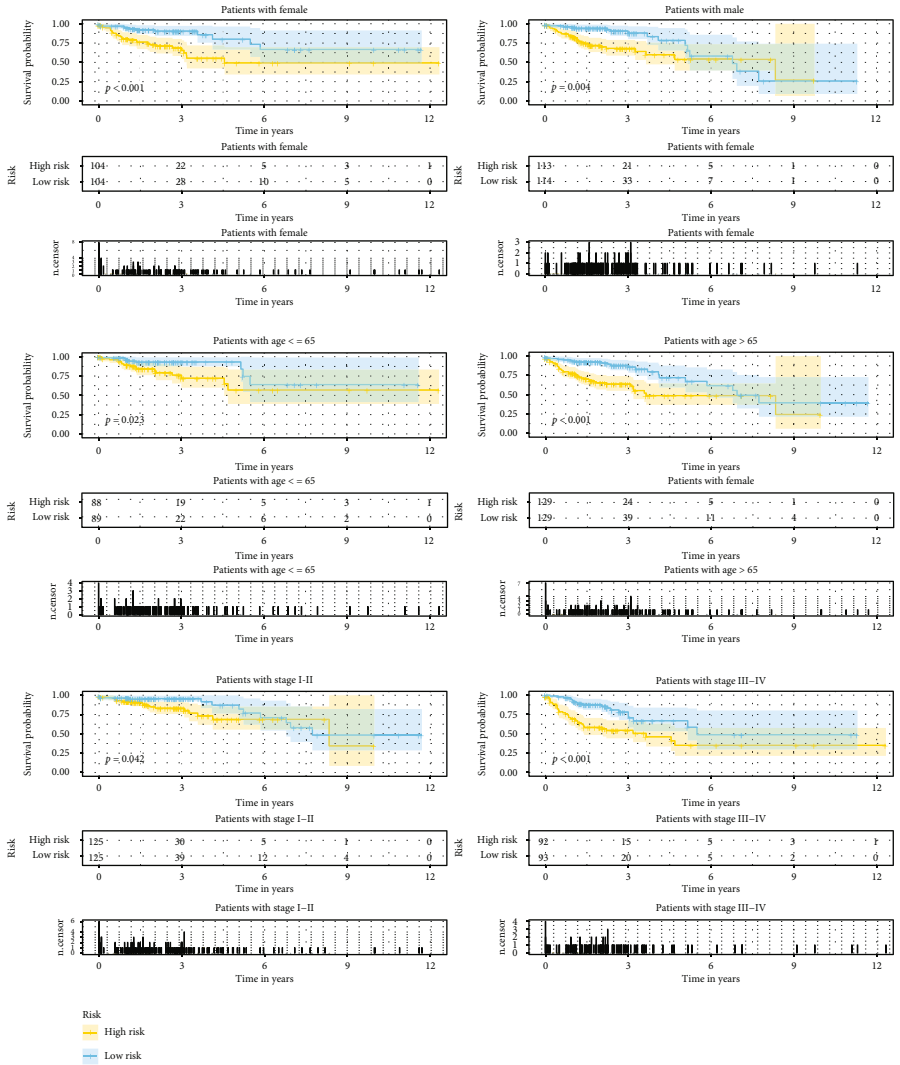
patients were grouped according to sex (male or female), age (≤65 or >65 years), and TNM stage (I–II or III–IV), and analyses were performed using the log-rank test and Kaplan–Meier analysis. We found that among female and male patients ($P < 0.05$), patients aged >65 years and ≤65 years ($P < 0.05$), and patients at TNM stage III–IV ($P \leq 0.001$), those in the high-risk group had a shorter survival time than those in the low-risk group (Figures 8(a) and 8(b)). A heat map revealed significant differences in tumour staging between the high- and low-risk groups (TCGA-CRC: $P = 0.024$; GSE17538: $P = 0.02$) (Figures 8(c) and 8(d)). These results suggested that the MECRG risk score can significantly affect the survival prognosis of CRC patients across gender, age, and stage.

3.8. Functional Enrichment Analysis of the MECRG Signature. We used GSEA to evaluate the functions of and signalling pathways associated with the genes in the high- and low-risk groups (Figure 9(a)). The results showed that the high-risk group was significantly enriched in functions related to the positive regulation of migration involved in sprouting angiogenesis, the phosphoinositide 3-kinase (PI3K)-AKT-mammalian target of rapamycin (mTOR) signalling pathway, basal cell carcinoma, extracellular matrix (ECM) receptor interaction, and angiogenesis, whereas the low-risk group was significantly enriched in apoptosis and the p53 signalling pathway. Additionally, enrichment analysis of the epithelial cell subsets indicated significant enrichment of activities involving collagen-containing ECM, fatty acid metabolism, basal cell carcinoma, PI3K-AKT signalling, and p53 signalling (Figure 9(b)). We found that the above signalling pathways are more likely to mediate the process of tumor progression in CRC epithelial cells than in other cell subpopulations.

3.9. Nomogram for Predicting the Prognosis of CRC Patients. To establish a practical method for predicting the probability of CRC patient survival, we constructed a MECRG nomogram using the TCGA-CRC training cohort to predict OS at 1, 3, and 5 years (Figure 10(a)). The predictors included three clinical features (age, sex, and TNM stage) and the risk score. ROC curve analysis of nomogram reliability revealed AUC values at 1, 3, and 5 years of 0.787, 0.806, and 0.777, respectively (Figure 10(b)). Analysis of the GSE17538 validation cohort yielded AUC values of 0.843, 0.812, and 0.827, respectively (Figure 10(c)). The calibration curve showed that the predicted survival rate agreed well with the actual survival rate (Figures 10(d) and 10(e)). Our results demonstrated that the MECRG nomogram constructed in combination with the clinical characteristics of CRC can provide good prediction.

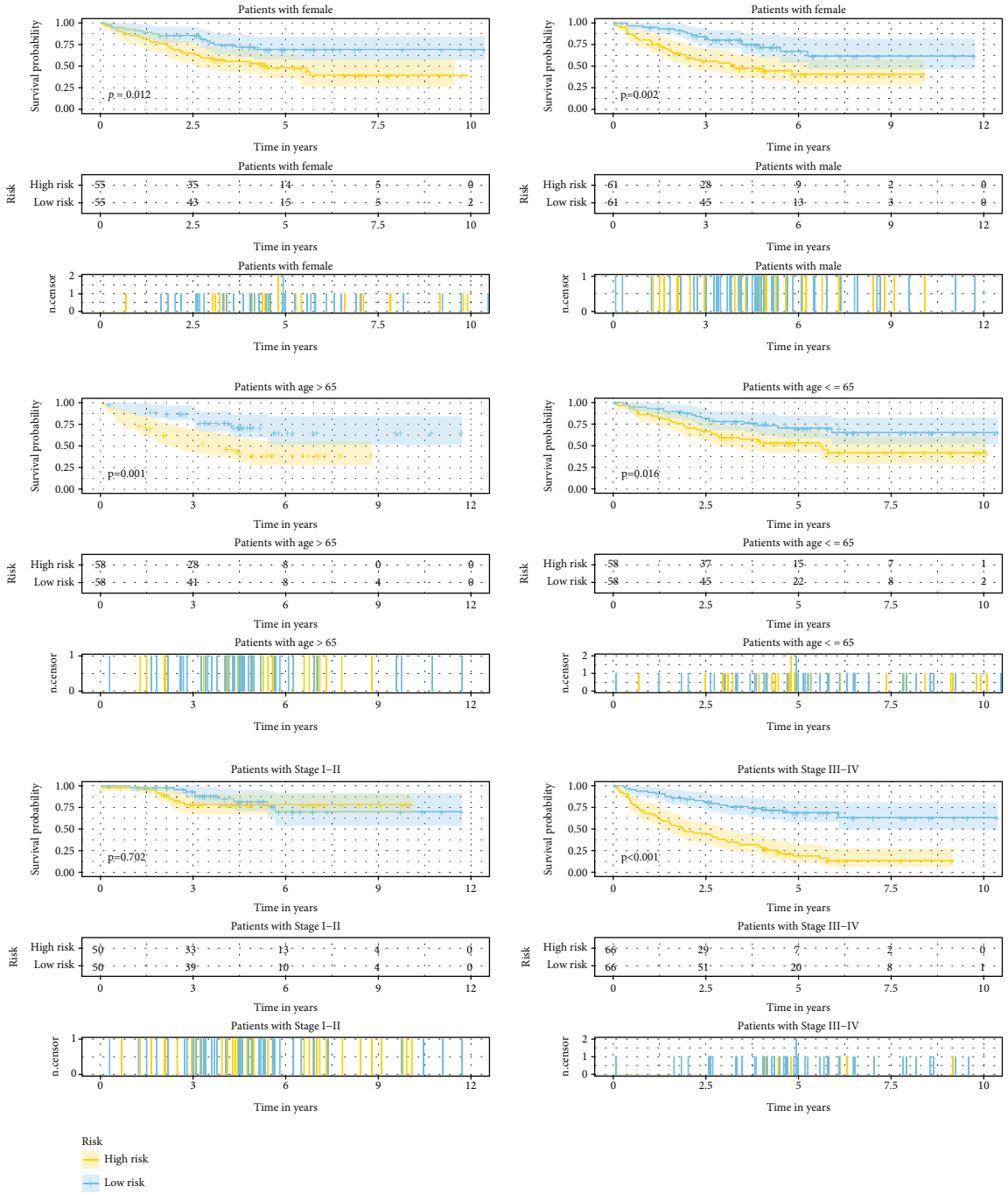
4. Discussion

Epithelial adenocarcinoma represents the most prevalent type of CRC and arises from benign colon adenoma [33]. Studies have shown that in epithelial malignant tumours, unstable adhesion between epithelial cells strongly correlates with an increased invasiveness of tumour cells. Recently, numerous mechanisms underlying cell–cell junctions regulated by E-cadherin expression in epithelial cells have been discovered [34–36]. Apical-basal polarity is the main characteristic of epithelial cells. Epithelial cell-associated polarity proteins are associated with the origin and poor prognosis of colorectal tumours, hepatocellular carcinoma, and endometrial cancer [37–39]. Therapeutic targets based on epithelial cell apical-basal polarity complexes have been reported. For example, partitioning-defective 6 (Par6) is expected to be a therapeutic target for breast cancer [37], and atypical protein kinase C (aPKC) has been suggested as a possible therapeutic target for gastric cancer [40]. Based on the above findings, we believed that the discovery of CRC epithelial cell-related biomarkers would facilitate the development of new therapeutic and predictive targets for CRC prognosis. ScRNA-seq technology allows the sequencing of RNA from individual cells to systematically track the dynamic changes



(a)

Figure 8: Continued.



(b)

Figure 8: Continued.

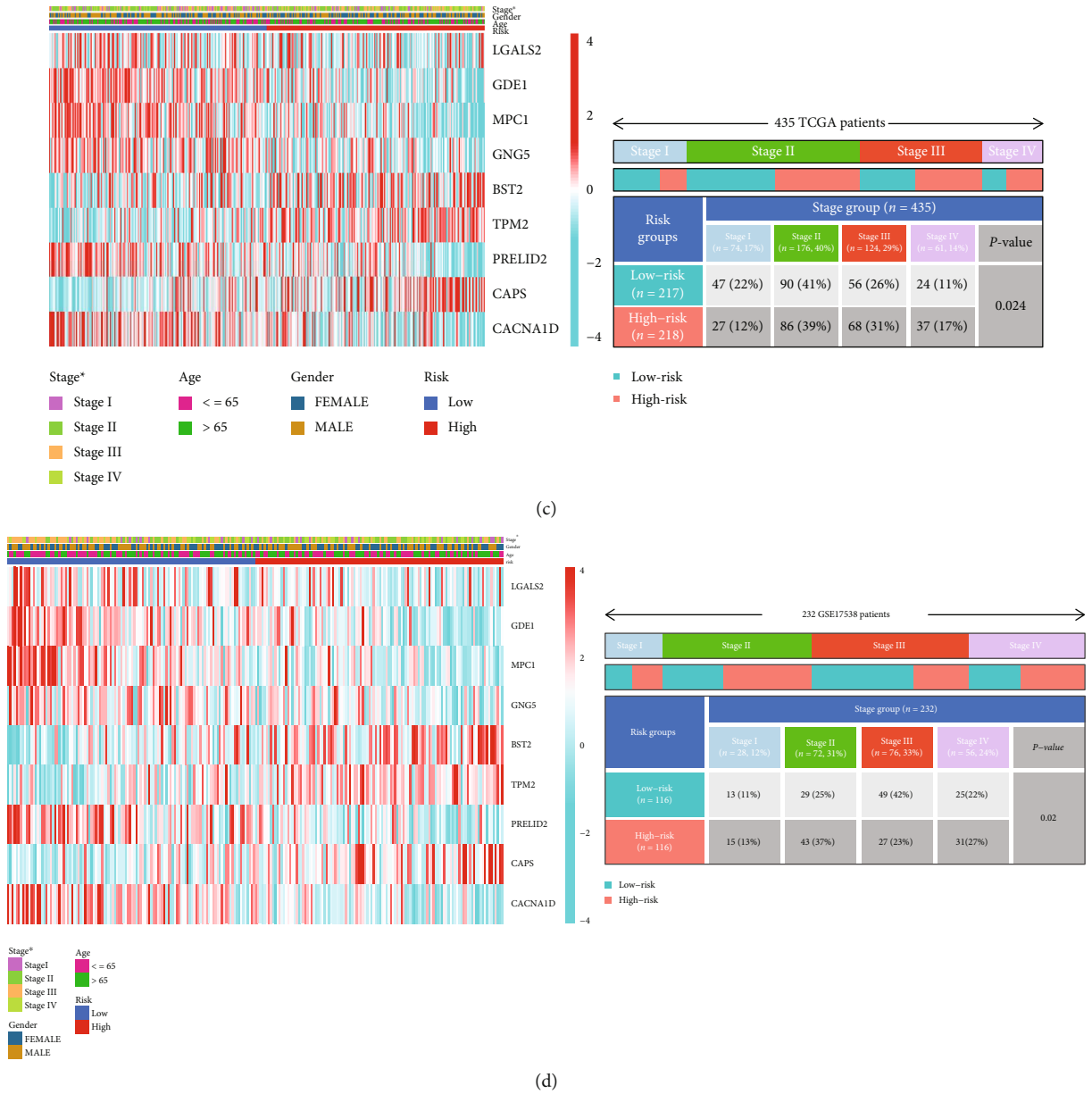
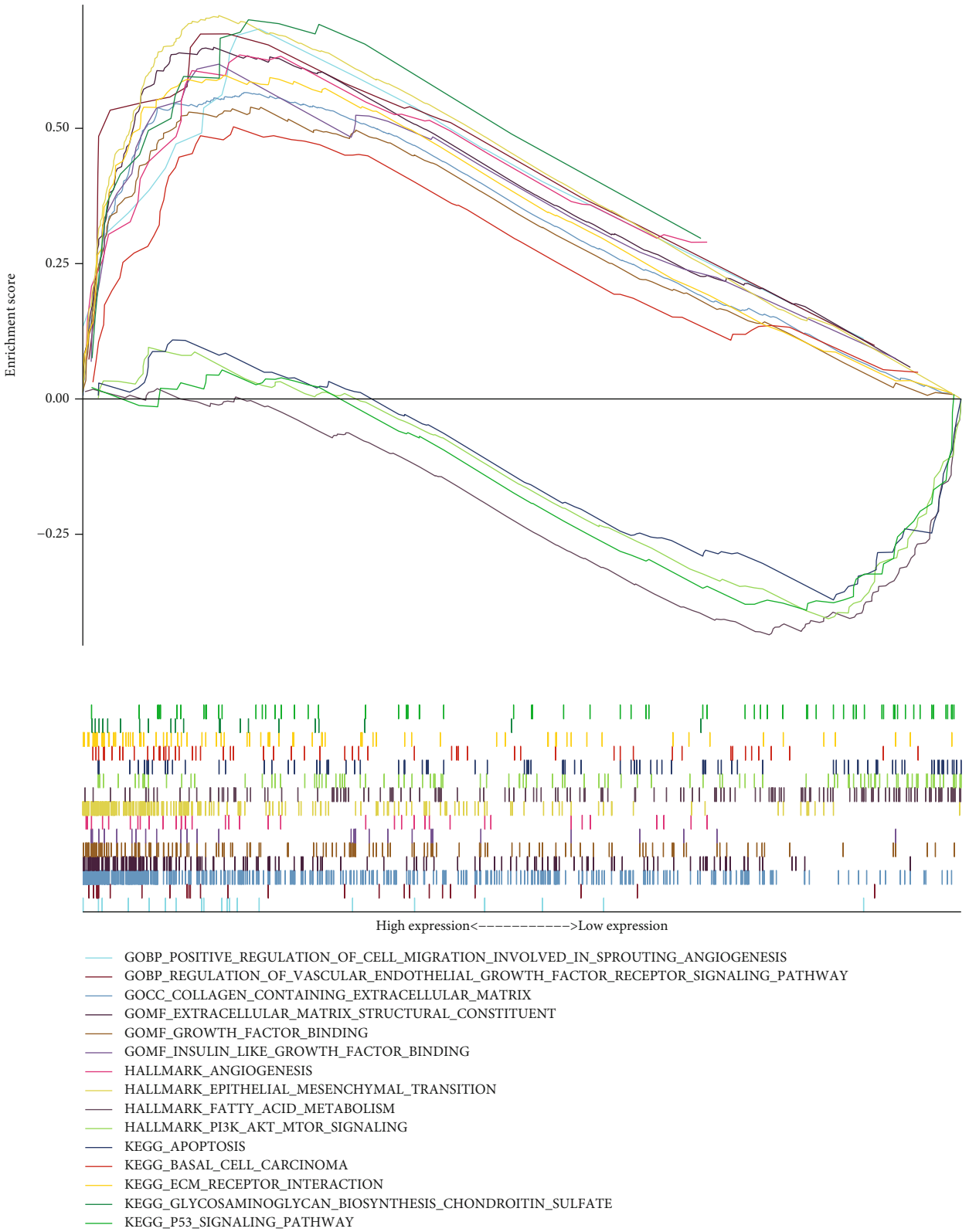


Figure 8: Analysis of the relationships between the MECRG signature and clinical features using the TCGA-CRC training and GSE17538 validation cohorts. (a) Survival analysis of the MECRG signature in clinical features based on the training cohort. (b) Survival analysis of the MECRG signature in clinical features based on the validation cohort. (c and d) Heat maps showing the correlation between MECRG risk grouping and TNM stage in the training cohort (c) and validation cohort (d). $^{***}P < 0.001$, $^{**}P < 0.01$, and $^{*}P < 0.05$.

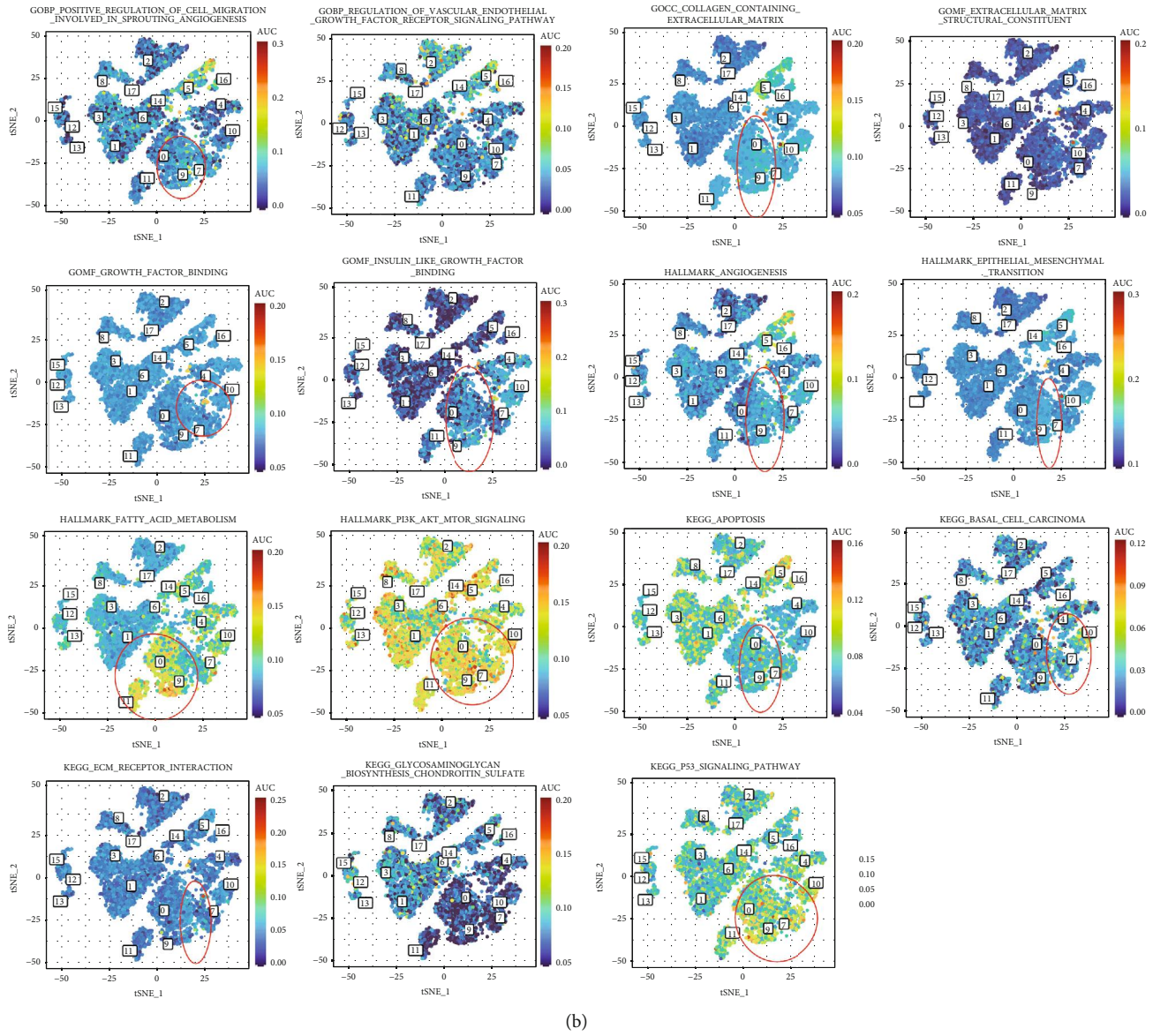
in individual cells and deepen the understanding of cellular states and gene expression regulation in pathological disease processes [41]. scRNA-seq analysis provides analytical detail on the cellular level [42]. The technology has been recently applied to study epithelial cells in serous epithelial ovarian cancer and endometrial cancer [43, 44]. In nontumour research, scRNA-seq has been used for gene expression profiling of breast epithelial cells during their development [45]. Xu et al. revealed the biological mechanisms implicated in the involvement of lung epithelial cells in idiopathic fibrosis using scRNA-seq [46]. Additionally, scRNA-seq technology has received increasing attention for predicting the prognosis of cancer patients. For example, Wang et al. constructed a model based on 10 biomarkers of pancreatic ductal epithelial cells to predict the prognosis of patients with pancreatic adenocarcinoma [47], Zheng et al. [48] screened nine fibroblast marker genes from scRNA-seq CRC data as potential prognostic markers, and Li et al. [49] identified seven macrophage marker genes from breast cancer scRNA-seq data as promising diagnostic and prognostic biomarkers.

In our study, we constructed a CRC prognosis model by analysing scRNA-seq data from CRC epithelial cells. Specifically, we analysed scRNA-seq data of 15,465 cells from four CRC patients, followed by PCA for unsupervised clustering and the identification of MEC subsets based on chromosome CNV analysis. We then used WGCNA, univariate Cox, and



(a)

Figure 9: Continued.



(b)

Figure 9: Functional enrichment analysis of the MECRG signature and distribution of enriched pathways in the cell subsets. (a) GSEA of the MECRG high- and low-risk groups. (b) Pathways enriched in the cell subsets of the MECRG risk groups.

LASSO regression analyses to construct a MECRG signature of nine genes related to CRC prognosis. We validated the MECRG signature using both training and validation cohorts. The Kaplan–Meier analysis confirmed a significantly shortened OS for CRC patients with a high MECRG risk score. Furthermore, the univariate and multivariate Cox regression analyses showed that the MECRG risk score may be an independent predictor of OS. Based on the MECRG risk score coupled with the clinical characteristics of sex, age, and TNM stage (III–IV), low-risk patients showed significantly longer survival times than high-risk patients. It is worth noting that there were significant differences between the MECRG signature-based risk groups at different CRC stages. Furthermore, the MECRG nomogram showed excellent prediction in both the training and validation sets, suggesting that it may efficiently predict CRC patient survival in the clinical setting.

Elevated expression of *LGALS2* reportedly inhibits the development of CRC and lymph node metastasis of gastric cancer [50, 51]. *GDE1* expression is significantly reduced in drug-resistant ovarian cancer samples [52]. Consistent with these findings, our results suggested that high expression of *LGALS2* and *GDE1* in CRC patients implies a good prognosis. However, the relationship between *GDE1* and CRC requires further study. Lysine demethylase 5A (KDM5A) regulates *MPC1* expression, and the KDM5A–MPC1 axis is involved in regulating the mesenchymal characteristics of cancer cells during EMT [53]. Schell et al. [54] found that loss of *MPC1* expression enhances the Warburg effect and promotes the proliferation of CRC cells. Deletion of *MPC1* is related to a poor prognosis in glioblastoma [55]. Furthermore, *BST2* activates the nuclear factor-*κ*B-Snail-Raf kinase inhibitor protein axis to promote tumour invasion and metastasis via EMT [56, 57]. *BST2* overexpression correlates with poor prognosis in CRC, stomach cancer, and

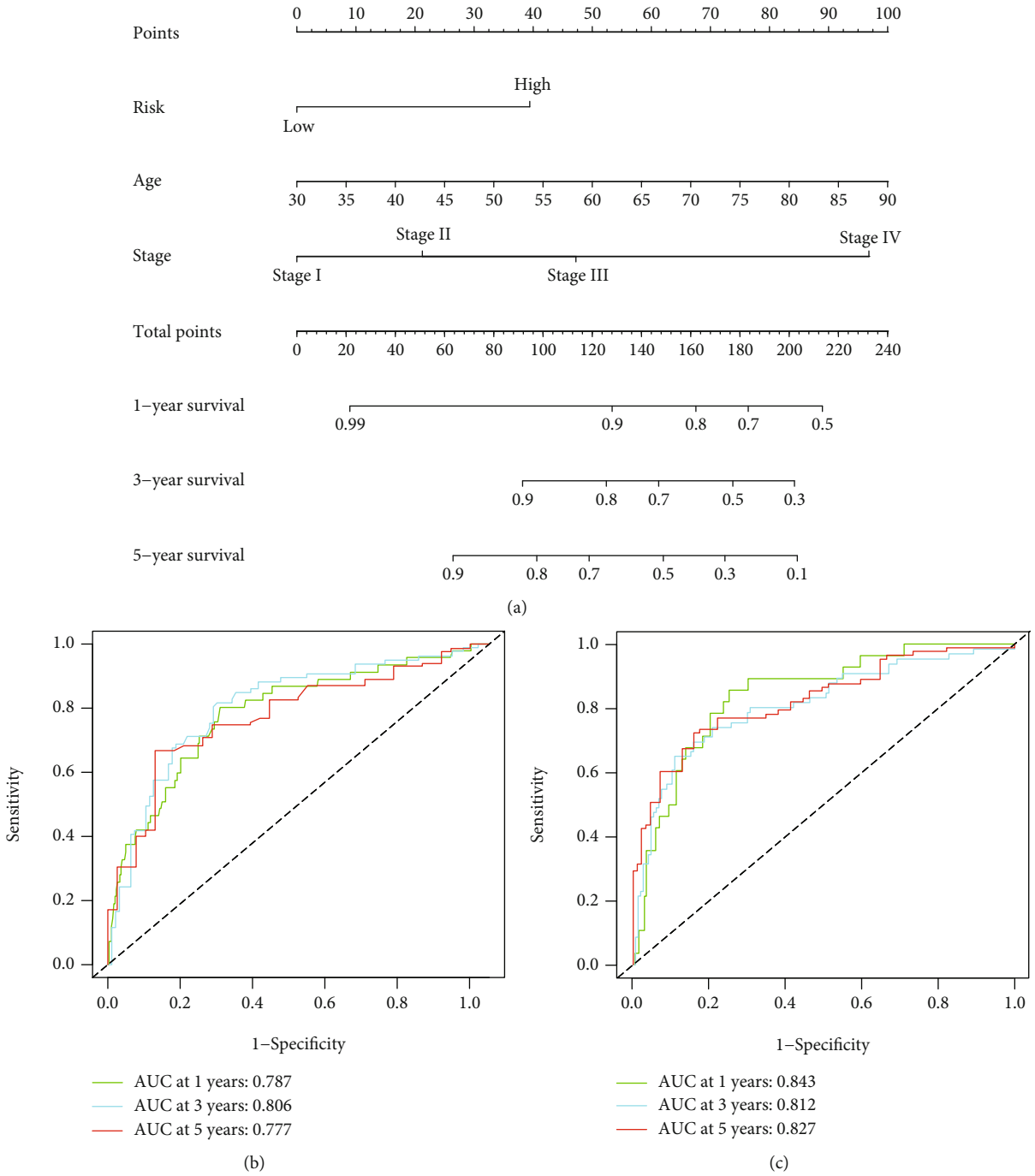


Figure 10: Continued.

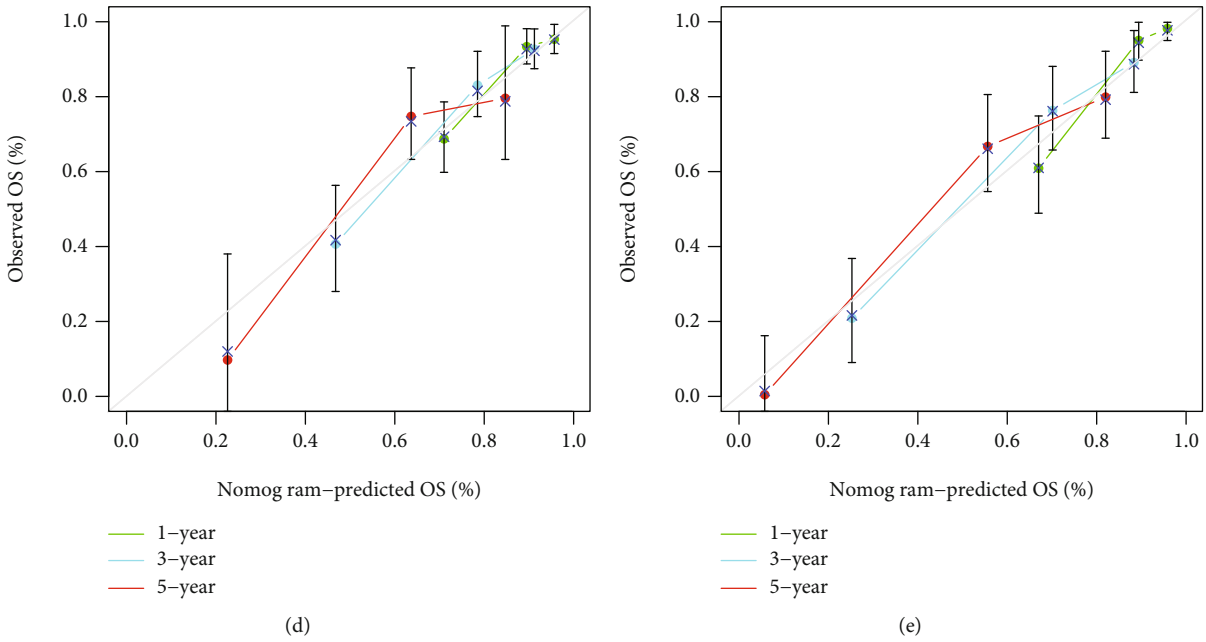


Figure 10: Construction of a nomogram based on the MECRG signature. ROC curve analysis of the nomogram (a) using the TCGA-CRC training cohort (b) and the GSE17538 validation cohort (c). (d and e) Calibration curves for the nomogram using the training cohort (d) and the validation cohort (e).

oesophageal cancer [58]. *BST2* expression is specifically upregulated in oral squamous cell carcinoma and is responsible for drug resistance [59]. *BST2* is more highly expressed in breast cancer cells derived from patients presenting bone metastasis than in human primary breast cancer cells [60]. Furthermore, Mukai et al. [58] showed that *BST2* knockout *in vitro* inhibited the proliferation of gastric cancer cells. Together, these findings suggest that *MPC1* is a protective gene and *BST2* a risk gene in CRC, which is consistent with the results of the present study. *CAPS* is a calcium-binding protein related to cell proliferation and differentiation signals [61] and associated with poor prognosis in CRC and gliomas and drug resistance in breast cancer [62–64], which is consistent with our results. Abnormal expression of *TPM2* is involved in actin cytoskeleton remodelling during EMT of lens epithelial cells [65]. Shibata et al. [66] showed that downregulation of *TPM2* expression decreased EMT in injured mouse lens epithelium, resulting in delayed lens wound healing. Additionally, *TPM2* is upregulated in ovarian cancer, liver cancer, and breast cancer [67–69]. Zhou et al. [70] found that elevated *TPM2* expression in CRC patients was predictive of poor prognosis, which is in line with the findings of the present study. However, Ma et al. [71] showed that *TPM2* expression was downregulated in CRC; therefore, the precise mechanism of *TPM2* in CRC requires clarification. Patients with glioma exhibiting elevated *GNG5* expression have a shorter survival time [72], and patients with head and neck squamous cell carcinoma and elevated *PRELID2* expression have a poor prognosis [73]. Tan et al. [74] reported that excessive aldosterone secretion from aldosteronoma is related to a *CACNA1D* mutation. We identified *GNG5*, *PRELID2*, and *CACNA1D* as protective genes in CRC patients; however, further basic experimental studies of these three genes in CRC are needed. Additionally, we showed that *MPC1*, *BST2*, and *TPM2* are closely related to EMT, suggesting that these molecules are potentially important EMT-related therapeutic targets.

GSEA suggested that genes related to collagen-containing ECM, fatty acid metabolism, PI3K-AKT signalling, p53 signalling, EMT, and other related pathways were enriched in epithelial cell subsets. Among these, EMT-related pathways were more significantly enriched in epithelial cell clusters 7, 9, and 10. Activation of the PI3K-AKT signalling pathway is a key feature of the EMT programme during tumour progression [75]. The expression of genes involved in fatty acid synthesis is upregulated in CRC epithelial cells, where the accumulation of polyunsaturated fatty acids contributes to the development of CRC [76]. Targeting fatty acid synthesis genes may become a new strategy for the treatment of CRC in the future. ECM remodelling can affect the signalling in the tumour microenvironment [77]. The P53 signalling pathway is a common tumour suppressor pathway [78], and we speculate that epithelial cells may promote CRC progression by interfering with this pathway. Thus, in the CRC microenvironment, epithelial cells may mediate the development of CRC through collagen-containing ECM, fatty acid metabolism, PI3K-AKT signalling, p53 signalling, EMT, and other pathological mechanisms.

This was a retrospective study that used scRNA-seq data and bulk data from public databases to construct a model for predicting survival in CRC patients. However, this study had some limitations. First, the mechanisms of *GDE1*, *PRELID2*, *GNG5*, and *CACNA1D* in CRC have not been clarified; therefore, the data suggesting their prognostic value need to be validated. Second, we did not evaluate tumour size, metastasis, surgery, postoperative chemoradiotherapy, and other prognostic factors in this study. This may have affected the predictive accuracy of the model. In future studies, we will include additional data to increase the accuracy of the model.

5. Conclusion

We evaluated epithelial cell marker genes with prognostic significance in CRC using scRNA-seq data and bulk data and generated a MECRG signature and risk score, which was confirmed to show independent prognostic efficacy for CRC patients. A nomogram based on the MECRG signature along with specific clinical features demonstrated accurate prediction of CRC patient survival, suggesting its potential utility for clinical application.

Data Availability

The datasets generated during and/or analysed in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This study was supported by the Zhejiang Chinese Medicine 13th Five-Year Key Discipline Construction Program (2017-XK-A12), the Zhejiang Provincial Traditional Chinese Medicine Science and Technology Programme Projects (2021ZB128 and 2018ZB064), and the Pang Dexiang Famous Veteran Chinese Medicine Specialist Inheritance Studio (2A12012014).

Supplementary Materials

Supplementary 1. Figure S1: quality control of the scRNA-seq data from the four CRC samples. Figure S2: gene numbers in the four CRC samples are positively correlated with sequencing depth. Figure S3: PCA classification of the cells based on the scRNA-seq data. Figure S4: eleven statistically significant principal components as identified by PCA. Figure S5: dot diagram showing the expression of specific marker genes in the cell subsets. Figure S6: distribution of specific marker genes in the four cell subsets. Figure S7: LASSO regression was used to screen MECRGs with prognostic significance.

Supplementary 2. Table S1: clinical characteristics of the four CRC patients. Table S2: CNV scores of epithelial cell subclusters. Table S3: MEC subclusters of marker genes. Table S4: MECRG signature using the TCGA-CRC training cohort. Table S5: MECRG signature using the GSE17538 validation cohort.

References

[1] Y. N. Peng, M. L. Huang, and C. H. Kao, "Prevalence of depression and anxiety in colorectal cancer patients: a literature review," *International Journal Environmental Research and Public Health*, vol. 16, no. 3, p. 411, 2019.

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

[3] F. Islami, A. Goding Sauer, K. D. Miller et al., "Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States," *CA: a Cancer Journal for Clinicians*, vol. 68, no. 1, pp. 31–54, 2018.

[4] K. D. Miller, M. Fidler-Benaoudia, T. H. Keegan, H. S. Hipp, A. Jemal, and R. L. Siegel, "Cancer statistics for adolescents and young adults, 2020," *CA: a Cancer Journal for Clinicians*, vol. 70, no. 6, pp. 443–459, 2020.

[5] S. B. Edge and C. C. Compton, "The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM," *Annals of Surgical Oncology*, vol. 17, no. 6, pp. 1471–1474, 2010.

[6] A. Jemal, R. Siegel, E. Ward et al., "Cancer statistics, 2008," *CA: a Cancer Journal for Clinicians*, vol. 58, no. 2, pp. 71–96, 2008.

[7] V. Karantza, "Keratins in health and cancer: more than mere epithelial cell markers," *Oncogene*, vol. 30, no. 2, pp. 127–138, 2011.

[8] N. Tanimura and Y. Fujita, "Epithelial defense against cancer (EDAC)," *Seminars in Cancer Biology*, vol. 63, pp. 44–48, 2020.

[9] H. Lv, R. Liu, J. Fu et al., "Epithelial cell-derived periostin functions as a tumor suppressor in gastric cancer through stabilizing p53 and E-cadherin proteins via the Rb/E2F1/p14ARF/Mdm2 signaling pathway," *Cell Cycle*, vol. 13, no. 18, pp. 2962–2974, 2014.

[10] L. M. McCaffrey and I. G. Macara, "Epithelial organization, cell polarity and tumorigenesis," *Trends in Cell Biology*, vol. 21, no. 12, pp. 727–735, 2011.

[11] C. Royer and X. Lu, "Epithelial cell polarity: a major gatekeeper against cancer?," *Cell Death & Differentiation*, vol. 18, no. 9, pp. 1470–1477, 2011.

[12] S. G. Zeng, X. Lin, J.-C. Liu, and J. Zhou, "Hypoxia-induced internalization of connexin 26 and connexin 43 in pulmonary epithelial cells is involved in the occurrence of non-small cell lung cancer via the P53/MDM2 signaling pathway," *International Journal of Oncology*, vol. 55, no. 4, pp. 845–859, 2019.

[13] J. Varga and F. R. Greten, "Cell plasticity in epithelial homeostasis and tumorigenesis," *Nature Cell Biology*, vol. 19, no. 10, pp. 1133–1141, 2017.

[14] T. Knösel, V. Emde, K. Schlüns, P. M. Schlag, M. Dietel, and I. Petersen, "Cytokeratin profiles identify diagnostic signatures in colorectal cancer using multiplex analysis of tissue microarrays," *Cell Oncology*, vol. 28, no. 4, pp. 167–175, 2006.

[15] C. Ausch, V. Buxhofer-Ausch, U. Olszewski et al., "Caspase-cleaved cytokeratin 18 fragment (M30) as marker of postoperative residual tumor load in colon cancer patients," *European Journal of Surgical Oncology*, vol. 35, no. 11, pp. 1164–1168, 2009.

[16] C. K. Cheong, K. R. Y. Nistala, C. H. Ng et al., "Neoadjuvant therapy in locally advanced colon cancer: a meta-analysis and systematic review," *Journal of Gastrointestinal Oncology*, vol. 11, no. 5, pp. 847–857, 2020.

[17] D. Lambrechts, E. Wauters, B. Boeckx et al., "Phenotype molding of stromal cells in the lung tumor microenvironment," *Nature Medicine*, vol. 24, no. 8, pp. 1277–1289, 2018.

[18] F. Tang, C. Barbacioru, Y. Wang et al., "mRNA-Seq whole-transcriptome analysis of a single cell," *Nature Methods*, vol. 6, no. 5, pp. 377–382, 2009.

[19] E. Azizi, A. J. Carr, G. Plitas et al., "Single-cell map of diverse immune phenotypes in the breast tumor microenvironment," *Cell*, vol. 174, no. 5, pp. 1293–1308.e36, 2018.

[20] Q. Zhang, Y. He, N. Luo et al., "Landscape and dynamics of single immune cells in hepatocellular carcinoma," *Cell*, vol. 179, no. 4, pp. 829–845.e20, 2019.

[21] X. Zheng, J. Song, C. Yu et al., "Single-cell transcriptomic profiling unravels the adenoma-initiation role of protein tyrosine kinases during colorectal tumorigenesis," *Signal Transduction and Targeted Therapy*, vol. 7, no. 1, 2022.

[22] S. Mangiola, M. A. Doyle, and A. T. Papenfuss, "Interfacing Seurat with the R tidy universe," *Bioinformatics*, vol. 37, no. 22, pp. 4100–4107, 2021.

[23] Z. Chen, H. Zhang, Y. Bai et al., "Single cell transcriptomic analysis identifies novel vascular smooth muscle subsets under high hydrostatic pressure," *Science China Life Sciences*, vol. 64, no. 10, pp. 1677–1690, 2021.

[24] J. H. Pan, H. Zhou, L. Cooper et al., "LAYN is a prognostic biomarker and correlated with immune infiltrates in gastric and colon cancers," *Frontiers in Immunology*, vol. 10, p. 6, 2019.

[25] G. Lombardo, M. Gili, C. Grange et al., "IL-3R-alpha blockade inhibits tumor endothelial cell-derived extracellular vesicle (EV)-mediated vessel formation by targeting the β-catenin pathway," *Oncogene*, vol. 37, no. 9, pp. 1175–1191, 2018.

[26] H. Ichimiya, K. Maeda, A. Enomoto, L. Weng, M. Takahashi, and T. Murohara, "Girdin/GIV regulates transendothelial permeability by controlling VE-cadherin trafficking through the small GTPase, R-Ras," *Biochemical and Biophysical Research Communications*, vol. 461, no. 2, pp. 260–267, 2015.

[27] O. Gires, M. Pan, H. Schinke, M. Canis, and P. A. Baeuerle, "Expression and function of epithelial cell adhesion molecule EpCAM: where are we after 40 years?," *Cancer and Metastasis Reviews*, vol. 39, no. 3, pp. 969–987, 2020.

[28] G. Corso, J. Figueiredo, S. P. De Angelis et al., "E-cadherin deregulation in breast cancer," *Journal of Cellular and Molecular Medicine*, vol. 24, no. 11, pp. 5930–5936, 2020.

[29] L. Fang, G. Yu, W. Yu, G. Chen, and B. Ye, "The correlation of WDR76 expression with survival outcomes and immune infiltrates in lung adenocarcinoma," *PeerJ*, vol. 9, article e12277, 2021.

[30] A. P. Patel, I. Tirosh, J. J. Trombetta et al., "Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma," *Science*, vol. 344, no. 6190, pp. 1396–1401, 2014.

[31] Q. Xu, S. Chen, Y. Hu, and W. Huang, "Single-cell RNA transcriptome reveals the intra-tumoral heterogeneity and regulators underlying tumor progression in metastatic pancreatic ductal adenocarcinoma," *Cell Death Discovery*, vol. 7, no. 1, p. 331, 2021.

[32] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, no. 1, p. 559, 2008.

[33] A. S. Takhar, O. Eremin, and S. A. Watson, "The role of gastrin in colorectal carcinogenesis," *The Surgeon*, vol. 2, no. 5, pp. 251–257, 2004.

[34] D. M. Bryant and J. L. Stow, "The ins and outs of E-cadherin trafficking," *Trends in Cell Biology*, vol. 14, no. 8, pp. 427–434, 2004.

[35] A. P. Kowalczyk and A. B. Reynolds, "Protecting your tail: regulation of cadherin degradation by p120-catenin," *Current Opinion in Cell Biology*, vol. 16, no. 5, pp. 522–527, 2004.

[36] G. Swaminathan and C. A. Cartwright, "Rack1 promotes epithelial cell-cell adhesion by regulating E-cadherin endocytosis," *Oncogene*, vol. 31, no. 3, pp. 376–389, 2012.

[37] Y. Kamei, K. Kito, T. Takeuchi et al., "Human scribble accumulates in colorectal neoplasia in association with an altered distribution of β-catenin," *Human Pathology*, vol. 38, no. 8, pp. 1273–1281, 2007.

[38] Z. Ouyang, W. Zhan, and L. Dan, "hScrib, a human homolog of Drosophila neoplastic tumor suppressor, is involved in the progress of endometrial cancer," *Oncology Research*, vol. 18, no. 11, pp. 593–599, 2009.

[39] S. Wan, A.-S. Meyer, S. M. E. Weiler et al., "Cytoplasmic localization of the cell polarity factor scribble supports liver tumor formation and tumor cell invasiveness," *Hepatology*, vol. 67, no. 5, pp. 1842–1856, 2018.

[40] R. Takagawa, K. Akimoto, Y. Ichikawa et al., "High expression of atypical protein kinase C lambda/iota in gastric cancer as a prognostic factor for recurrence," *Annals of Surgical Oncology*, vol. 17, no. 1, pp. 81–88, 2010.

[41] A. A. Kolodziejczyk, J. K. Kim, V. Svensson, J. C. Marioni, and S. A. Teichmann, "The technology and biology of single-cell RNA sequencing," *Molecular Cell*, vol. 58, no. 4, pp. 610–620, 2015.

[42] X. Ren, B. Kang, and Z. Zhang, "Understanding tumor ecosystems by single-cell sequencing: promises and limitations," *Genome Biology*, vol. 19, no. 1, p. 211, 2018.

[43] A. J. Shih, A. Menzin, J. Whyte et al., "Identification of grade and origin specific cell populations in serous epithelial ovarian cancer by single cell RNA-seq," *PLoS One*, vol. 13, no. 11, article e0206785, 2018.

[44] Y.-E. Guo, Y. Li, B. Cai et al., "Phenotyping of immune and endometrial epithelial cells in endometrial carcinomas revealed by single-cell RNA sequencing," *Aging*, vol. 13, no. 5, pp. 6565–6591, 2021.

[45] K. Bach, S. Pensa, M. Grzelak et al., "Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing," *Nature Communications*, vol. 8, no. 1, p. 2128, 2017.

[46] Y. Xu, T. Mizuno, A. Sridharan et al., "Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis," *JCI Insight*, vol. 1, no. 20, article e90558, 2016.

[47] X. Wang, X. Dou, X. Ren et al., "A ductal-cell-related risk model integrating single-cell and bulk sequencing data predicts the prognosis of patients with pancreatic adenocarcinoma," *Frontiers in Genetics*, vol. 12, article 763636, 2021.

[48] H. Zheng, H. Liu, Y. Ge, and X. Wang, "Integrated single-cell and bulk RNA sequencing analysis identifies a cancer associated fibroblast-related signature for predicting prognosis and therapeutic responses in colorectal cancer," *Cancer Cell International*, vol. 21, no. 1, 2021.

[49] Y. Li, X. Zhao, Q. Liu, and Y. Liu, "Bioinformatics reveal macrophages marker genes signature in breast cancer to predict prognosis," *Annals of Medicine*, vol. 53, no. 1, pp. 1019–1031, 2021.

[50] H. Li, L. Zhao, Y. S. Lau, C. Zhang, and R. Han, "Genome-wide CRISPR screen identifies LGALS2 as an oxidative stress-responsive gene with an inhibitory function on colon tumor growth," *Oncogene*, vol. 40, no. 1, pp. 177–188, 2021.

[51] J.-H. Jung, H.-J. Kim, J. Yeom et al., "Lowered expression of galectin-2 is associated with lymph node metastasis in gastric cancer," *Journal of Gastroenterology*, vol. 47, no. 1, pp. 37–48, 2012.

[52] J. Lu, Y. Li, Y. A. Li et al., "In vivo detection of dysregulated choline metabolism in paclitaxel-resistant ovarian cancers with proton magnetic resonance spectroscopy," *Journal of Translational Medicine*, vol. 20, no. 1, p. 92, 2022.

[53] J. H. You, J. Lee, and J. L. Roh, "Mitochondrial pyruvate carrier 1 regulates ferroptosis in drug-tolerant persister head and neck cancer cells via epithelial-mesenchymal transition," *Cancer Letters*, vol. 507, pp. 40–54, 2021.

[54] J. C. Schell, K. A. Olson, L. Jiang et al., "A role for the mitochondrial pyruvate carrier as a repressor of the Warburg effect and colon cancer cell growth," *Molecular Cell*, vol. 56, no. 3, pp. 400–413, 2014.

[55] Y. Chai, C. Wang, W. Liu, Y. Fan, and Y. Zhang, "*MPC1* deletion is associated with poor prognosis and temozolomide resistance in glioblastoma," *Journal of Neurooncology*, vol. 144, no. 2, pp. 293–301, 2019.

[56] A. Matsuda, Y. Suzuki, G. Honda et al., "Large-scale identification and characterization of human genes that activate NF-κB and MAPK signaling pathways," *Oncogene*, vol. 22, no. 21, pp. 3307–3318, 2003.

[57] K. Wu and B. Bonavida, "The activated NF-κB-Snail-RKIP circuitry in cancer regulates both the metastatic cascade and resistance to apoptosis by cytotoxic drugs," *Critical Reviews of Immunology*, vol. 29, no. 3, pp. 241–254, 2009.

[58] S. Mukai, N. Oue, T. Oshima et al., "Overexpression of transmembrane protein BST2 is associated with poor survival of patients with esophageal, gastric, or colorectal cancer," *Annals of Surgical Oncology*, vol. 24, no. 2, pp. 594–602, 2017.

[59] H. Jin, L. Zhang, S. Wang, and L. Qian, "BST2 promotes growth and induces gefitinib resistance in oral squamous cell carcinoma via regulating the EGFR pathway," *Archives of Medicinal Science*, vol. 17, no. 6, pp. 1772–1782, 2021.

[60] D. Cai, J. Cao, Z. Li et al., "Up-regulation of bone marrow stromal protein 2 (BST2) in breast cancer with bone metastasis," *BMC Cancer*, vol. 9, no. 1, p. 102, 2009.

[61] R. Lecocq, F. Lamy, C. Erneux, and J. E. Dumont, "Rapid purification and identification of calcyphosine, a Ca^{2+}-binding protein phosphorylated by protein kinase A," *Biochemistry Journal*, vol. 306, no. 1, pp. 147–151, 1995.

[62] W. Shao, Q. Wang, F. Wang, Y. Jiang, M. Xu, and J. Xu, "Abnormal expression of calcyphosine is associated with poor prognosis and cell biology function in colorectal cancer," *Onco Targets Therapy*, vol. 9, pp. 477–487, 2016.

[63] Z. Zhu, J. Wang, J. Tan et al., "Calcyphosine promotes the proliferation of glioma cells and serves as a potential therapeutic target," *Journal of Pathology*, vol. 255, no. 4, pp. 374–386, 2021.

[64] H. J. Johansson, B. C. Sanchez, J. Forshed et al., "Proteomics profiling identify CAPS as a potential predictive marker of tamoxifen resistance in estrogen receptor positive breast cancer," *Clinical Proteomics*, vol. 12, no. 1, p. 8, 2015.

[65] E. Kubo, N. Hasanova, N. Fatma, H. Sasaki, and D. P. Singh, "Elevated tropomyosin expression is associated with epithelial-mesenchymal transition of lens epithelial cells," *Journal of Cellular and Molecular Medicine*, vol. 17, no. 1, pp. 212–221, 2013.

[66] T. Shibata, S. Shibata, Y. Ishigaki et al., "Tropomyosin 2 heterozygous knockout in mice using CRISPR-cas9 system displays the inhibition of injury-induced epithelial-mesenchymal transition, and lens opacity," *Mechanisms of Ageing and Development*, vol. 171, pp. 24–30, 2018.

[67] D. Q. Li, L. Wang, F. Fei et al., "Identification of breast cancer metastasis-associated proteins in an isogenic tumor metastasis model using two-dimensional gel electrophoresis and liquid chromatography-ion trap-mass spectrometry," *Proteomics*, vol. 6, no. 11, pp. 3352–3368, 2006.

[68] Z. Tian, J. Zhao, and Y. Wang, "The prognostic value of TPM1–4 in hepatocellular carcinoma," *Cancer Medicine*, vol. 11, no. 2, pp. 433–446, 2022.

[69] H.-Y. Tang, L. A. Beer, J. L. Tanyi, R. Zhang, Q. Liu, and D. W. Speicher, "Protein isoform-specific validation defines multiple chloride intracellular channel and tropomyosin isoforms as serological biomarkers of ovarian cancer," *Journal of Proteomics*, vol. 89, pp. 165–178, 2013.

[70] Y. Zhou, S. Bian, X. Zhou et al., "Single-cell multiomics sequencing reveals prevalent genomic alterations in tumor stromal cells of human colorectal cancer," *Cancer Cell*, vol. 38, no. 6, pp. 818–828.e5, 2020.

[71] Y. Ma, T. Xiao, Q. Xu, X. Shao, and H. Wang, "iTRAQ-based quantitative analysis of cancer-derived secretory proteome reveals TPM2 as a potential diagnostic biomarker of colorectal cancer," *Frontiers in Medicine*, vol. 10, no. 3, pp. 278–285, 2016.

[72] W. Zhang, Z. Liu, B. Liu et al., "*GNG5* is a novel oncogene associated with cell migration, proliferation, and poor prognosis in glioma," *Cancer Cell International*, vol. 21, no. 1, 2021.

[73] G. Liu, X. Zeng, B. Wu, J. Zhao, and Y. Pan, "RNA-seq analysis of peripheral blood mononuclear cells reveals unique transcriptional signatures associated with radiotherapy response of nasopharyngeal carcinoma and prognosis of head and neck cancer," *Cancer Biology and Therapy*, vol. 21, no. 2, pp. 139–146, 2020.

[74] G. C. Tan, G. Negro, A. Pinggera et al., "Aldosterone-producing adenomas: histopathology-genotype correlation and identification of a novel *CACNA1D* mutation," *Hypertension*, vol. 70, no. 1, pp. 129–136, 2017.

[75] L. Larue and A. Bellacosa, "Epithelial-mesenchymal transition in development and cancer: role of phosphatidylinositol 3′ kinase/AKT pathways," *Oncogene*, vol. 24, no. 50, pp. 7443–7454, 2005.

[76] J. Hofmanová, J. Slavík, M. Ciganek et al., "Complex alterations of fatty acid metabolism and phospholipidome uncovered in isolated colon cancer epithelial cells," *International Journal of Molecular Science*, vol. 22, no. 13, p. 6650, 2021.

[77] V. Mohan, A. Das, and I. Sagi, "Emerging roles of ECM remodeling processes in cancer," *Seminars in Cancer Biology*, vol. 62, pp. 192–200, 2020.

[78] L. J. Hernández Borrero and W. S. El-Deiry, "Tumor suppressor p53: biology, signaling pathways, and therapeutic targeting," *Biochimica Biophysica Acta Reviews on Cancer*, vol. 1876, no. 1, article 188556, 2021.

Hindawi
Disease Markers
Volume 2022, Article ID 3653363, 11 pages
https://doi.org/10.1155/2022/3653363

Research Article

Characterization of Circular RNA Expression Profiles in Colon Specimens of Patients with Slow Transit Constipation

Changlei Xi,[1,2] **Yuntian Hong**,[1,3,4,5,6] **Baoxiang Chen**,[1,3,4,5,6] **Xiaoyu Xie**,[1,3,4,5,6] **Weicheng Liu**,[1,3,4,5,6] **Qun Qian**,[1,3,4,5,6] **Congqing Jiang**,[1,3,4,5,6] **and Xianghai Ren**[1,3,4,5,6]

[1]*Department of Colorectal and Anal Surgery, Zhongnan Hospital of Wuhan University, Wuhan 430071, China*
[2]*Department of Colorectal and Anal Surgery, The Second Clinical Medical College, Yangtze University, Jingzhou Central Hospital, Jingzhou 434020, China*
[3]*Clinical Center of Intestinal and Colorectal Diseases of Hubei Province, Wuhan 430071, China*
[4]*Key Laboratory of Intestinal and Colorectal Diseases of Hubei Province, Wuhan 430071, China*
[5]*Colorectal and Anal Disease Research Centre, Medical School of Wuhan University, Wuhan 430071, China*
[6]*Quality Control Centre of Colorectal and Anal Surgery of Health Commission of Hubei Province, Wuhan 430071, China*

Correspondence should be addressed to Congqing Jiang; wb002554@whu.edu.cn and Xianghai Ren; xhren@whu.edu.cn

Changlei Xi, Yuntian Hong, and Baoxiang Chen contributed equally to this work.

Received 13 April 2022; Accepted 26 May 2022; Published 10 June 2022

Academic Editor: Chiara Nicolazzo

Background. Slow transit constipation (STC) is a clinical syndrome characterized by a decreased urge to defecate and delayed colonic transit. Circular RNAs (circRNAs) are a recently discovered class of regulatory RNAs that have emerged as critical biomarkers and regulators of various diseases. However, the expression profiles and mechanisms underlying circRNA regulation in human STC tissues have not been explored. *Methods*. High-throughput RNA sequencing technology was used to compare the differences in circRNA expression profiles in colon samples taken from patients with STC or controls. Bioinformatics analyses were performed on the host genes of the differentially expressed circRNAs (DE-circRNAs), a competing endogenous RNA network was constructed, and the expression levels of some DE-circRNAs were verified using quantitative real-time polymerase chain reactions (qRT-PCR). *Results*. There were 190 DE-circRNAs identified in the STC group. Bioinformatics analysis predicted that the DE-circRNAs were enriched in the relaxation of smooth muscle, actin binding, actin cytoskeleton organization, dilated cardiomyopathy, and cardiac muscle contraction. These results suggest that muscle diseases may be related to the pathogenesis of STC. The expression levels of the 12 most differentially expressed circRNAs were verified using qRT-PCR. In addition, circRNA–microRNA–mRNA regulatory networks were constructed using the 8 most significant circRNAs. Some mRNAs predicted to be closely related to smooth muscle function were found in these networks. *Conclusions*. This study provides a helpful blueprint for researchers to select candidate circRNAs for further study of the pathogenesis of STC and screen potential biomarkers or targets for use in the diagnosis and treatment of STC.

1. Introduction

Functional constipation (FC) is a common digestive tract disorder [1]. A recent study showed that the incidence of FC was 8.73% in the Japanese population between the ages of 20 and 69 years [2]. Slow transit constipation (STC) is a typical type of functional constipation characterized by prolonged colonic transmission and a decreased frequency of defecation [3]. Patients with STC suffer from both physical and psychological burdens that seriously affect their quality of life.

Many hypotheses about the pathophysiology of STC have been proposed, including degenerative neuromuscular processes, interstitial cell dysfunction, dysbacteriosis, and autoimmune disorder [4–6]. However, the pathogenesis of STC remains unclear. Recently, the changes and functions

Table 1: Demographic information of STC and control.

Characteristics	STC for circRNA screening ($n = 6$)	Control for circRNA screening ($n = 6$)	STC for verification ($n = 15$)	Control for verification ($n = 15$)
Age (y)	62.3 ± 7.0	60.8 ± 10.1	57.3 ± 8.4	60.7 ± 10.6
Male/female	2/4	2/4	4/11	5/10

STC: slow transit constipation.

in noncoding RNAs (ncRNAs) in the colon tissue of STC patients have been explored, enriching the research directions on the pathogenesis of STC [7–10].

Circular RNAs (circRNAs) are a special class of ncRNAs that have a stable circular structure [11]. These RNAs play important roles in a variety of digestive diseases (e.g., Hirschsprung's disease and colitis) and physiological processes (e.g., the self-renewal of intestinal stem cells) [12–14]. However, the expression profiles of circRNAs in human STC tissues have yet to be explored.

In the present study, we performed whole-transcriptome sequencing of RNA samples from the colon tissues of STC patients and controls to identify differentially expressed circRNAs. Furthermore, the expression of potentially functional circRNAs was validated using reverse transcription quantitative real-time polymerase chain reactions (qRT-PCR). Bioinformatics analyses were also performed to explore the possible regulatory mechanisms of selected circRNAs. This study provides a basis for further research on the pathogenesis of STC, innovative diagnostic methods for STC, and new insights into STC gene therapy.

2. Materials and Methods

2.1. Tissue Samples and Cell Culture. There were 42 patients included in this study: 21 patients with STC undergoing subtotal colectomy and 21 controls undergoing radical surgery for colon cancer. The inclusion and exclusion criteria were based on those of our previous study [15]. Briefly, all included patients had a history of STC for more than five years, failed to respond to nonsurgical regimens, and had a strong desire for surgery. Patients with obstructed defecation syndrome, severe psychiatric disease, small bowel dysmotility, or megacolon/megarectum were excluded from this study. Tissue samples of these patients were collected during surgical treatment at the Department of Colorectal and Anal Surgery, Zhongnan Hospital of Wuhan University. All samples were obtained from the same region of the colon descendens in the STC and tumor-free control groups. Tissues were immediately frozen in liquid nitrogen for 15 min and stored at −80°C until use. This study was approved by the Ethics Committee of Zhongnan Hospital (ethical application ref: 2022061), and written informed consent was obtained from each participant.

Human HEK 293T cells were purchased from the China Center for Type Culture Collection and cultured in DMEM (Dulbecco's modified Eagle's medium) with 10% fetal bovine serum (Gibco, USA), 1% penicillin/streptomycin, in a humidified incubator with 5% CO_2 at 37°C.

2.2. circRNA Sequencing and Identification. We randomly selected 12 tissues (6 STC and 6 controls) from 42 samples for circRNA screening by Shanghai Majorbio Bio-Pharm Biotechnology Co., Ltd. (Shanghai, China). The remaining 15 pairs of samples were used to verify circRNA expression (Table 1). The sequencing library was prepared using the TruSeq total RNA kit (Illumina; San Diego, CA, USA). First, ribosomal RNA was depleted and fragmented. Next, cDNA was synthesized using random hexamer primers. The RNA template was then removed, and a replacement strand was synthesized incorporating dUTP instead of dTTP to generate double-stranded (ds) cDNA. AMPure XP beads were used to separate the ds-cDNA from the second-strand reaction mix. A single 'A' nucleotide was added to the $3'$ ends of these blunt fragments. Finally, multiple indexing adapters were ligated to the ends of the ds-cDNA. Libraries were size-selected for cDNA target fragments of 200–300 base pairs on 2% Low Range Ultra Agarose, followed by PCR amplification using Phusion DNA polymerase (NEB). After quantification using TBS380, the library was sequenced using an Illumina HiSeq Xten system (Illumina; San Diego, CA, USA). CircRNA Identifier (CIRI) tools were used to identify the circRNAs. The level of each circRNA was calculated using the *spliced reads per billion mapping* (SRPBM) method. Significantly differentially expressed circRNAs (DE-circRNAs) were defined as $|\log_2 FC| > 1$ and $p < 0.001$.

2.3. Quantitative Real-Time PCR (qRT-PCR). Total RNA was extracted from tissues using TRIzol reagent (Invitrogen; USA). RNA (1 μg) was reverse-transcribed into cDNA using HiScript II Reverse Transcriptase (Vazyme Biotech Co., Ltd., China). An Applied Biosystems 7500 Real-Time PCR System (ThermoFisher Scientific; USA) was used to perform qRT-PCR in a 10 μl reaction, including 5 μl SYBR Mix (Vazyme Biotech Co., Ltd., China), 1 μl cDNA, 3.6 μl ddH_2O, and 0.2 μl each of the forward and reverse primers. The qRT-PCR amplification conditions included an initial denaturation step (95°C for 2 min) followed by 40 cycles of denaturation at 95°C for 15 s and annealing at 60°C for 1 min. GAPDH was used as a normalization standard. The relative RNA levels were calculated using the $2^{-\Delta\Delta Ct}$ method. The experiments were repeated three times for each sample. The primers are listed in Supplementary Table S1.

2.4. Functional Enrichment Analysis of Host Genes of the circRNAs. The host genes of the significant DE-circRNAs were subjected to functional enrichment analysis using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. The analyses of GO terms

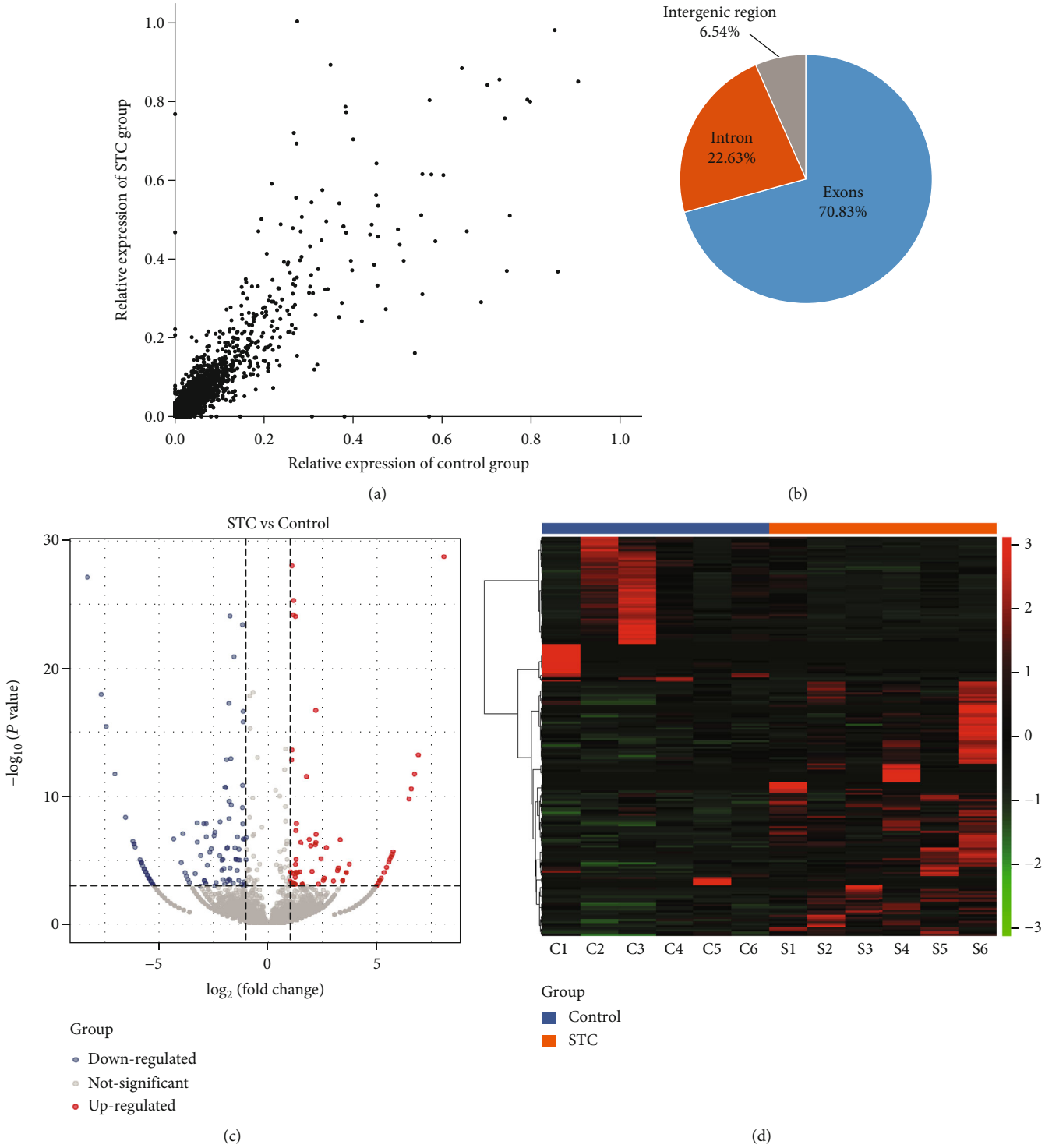
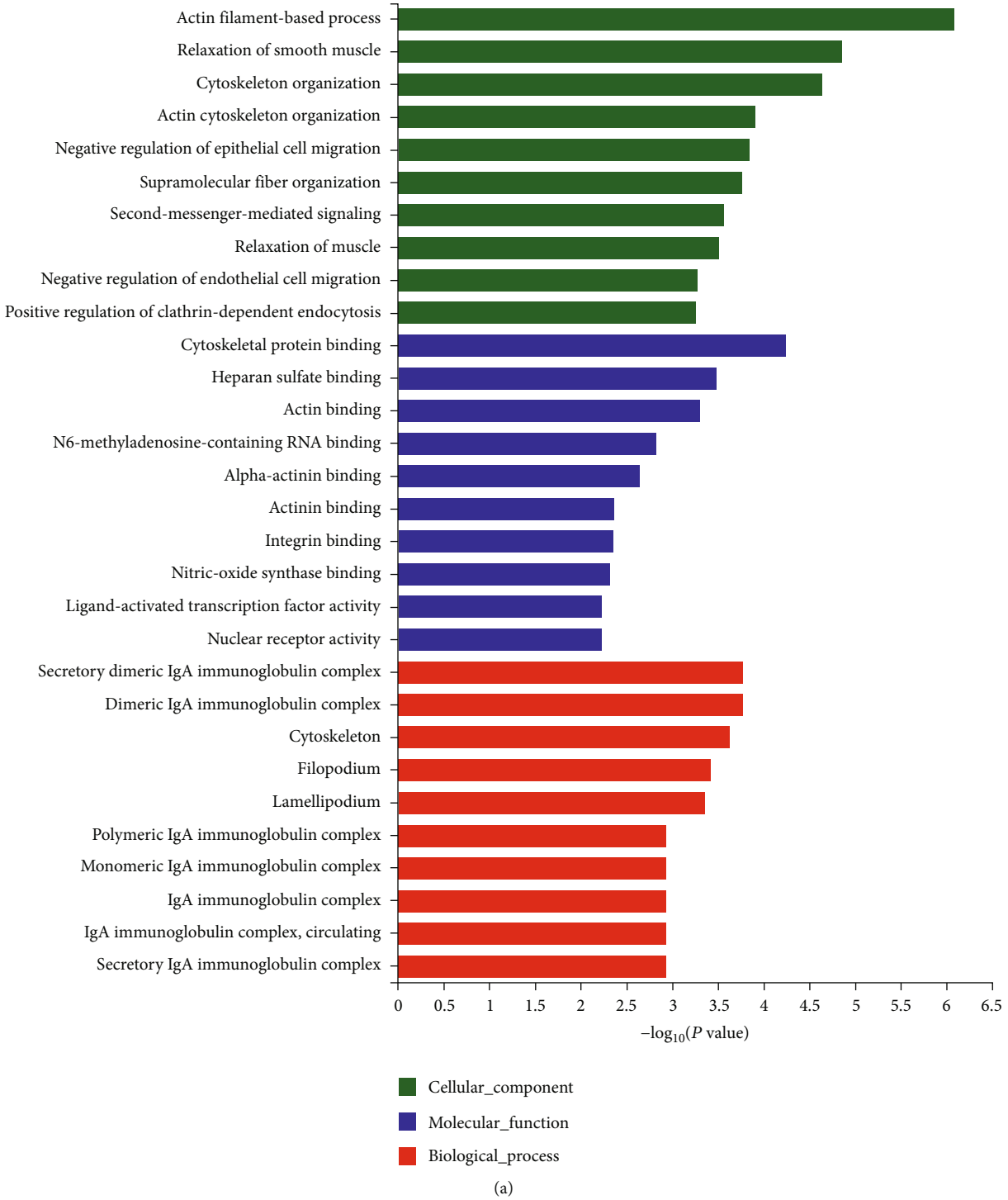


Figure 1: Analysis of the circRNA expression profile in STC. (a) Scatter plot of circRNA expression in the STC and control groups. (b) The circRNAs are classified by distribution. (c) A volcano plot represents the STC DE-circRNAs. Red dots and blue dots indicate the upregulated and downregulated DE-circRNAs, respectively. (d) Hierarchically clustered heatmap analysis of the circRNA expression profiles.

and KEGG pathways were performed using Goatools (https://github.com/tanghaibao/Goatools) and KOBAS (http://kobas.cbi.pku.edu.cn/home.do).

2.5. Construction of circRNA–microRNA–mRNA Network. The target microRNAs (miRNAs) of the circRNAs were predicted using starBase (https://starbase.sysu.edu.cn/) and circBank (http://www.circbank.cn/index.html). The target mRNAs of the miRNAs were predicted using TargetScan (https://www.targetscan.org/), miRDB (http://mirdb.org/), miRWalk (http://mirwalk.umm.uni-heidelberg.de/), and TarBase (http://microrna.gr/tarbase/). Next, circRNA–



(a)

Figure 2: Continued.

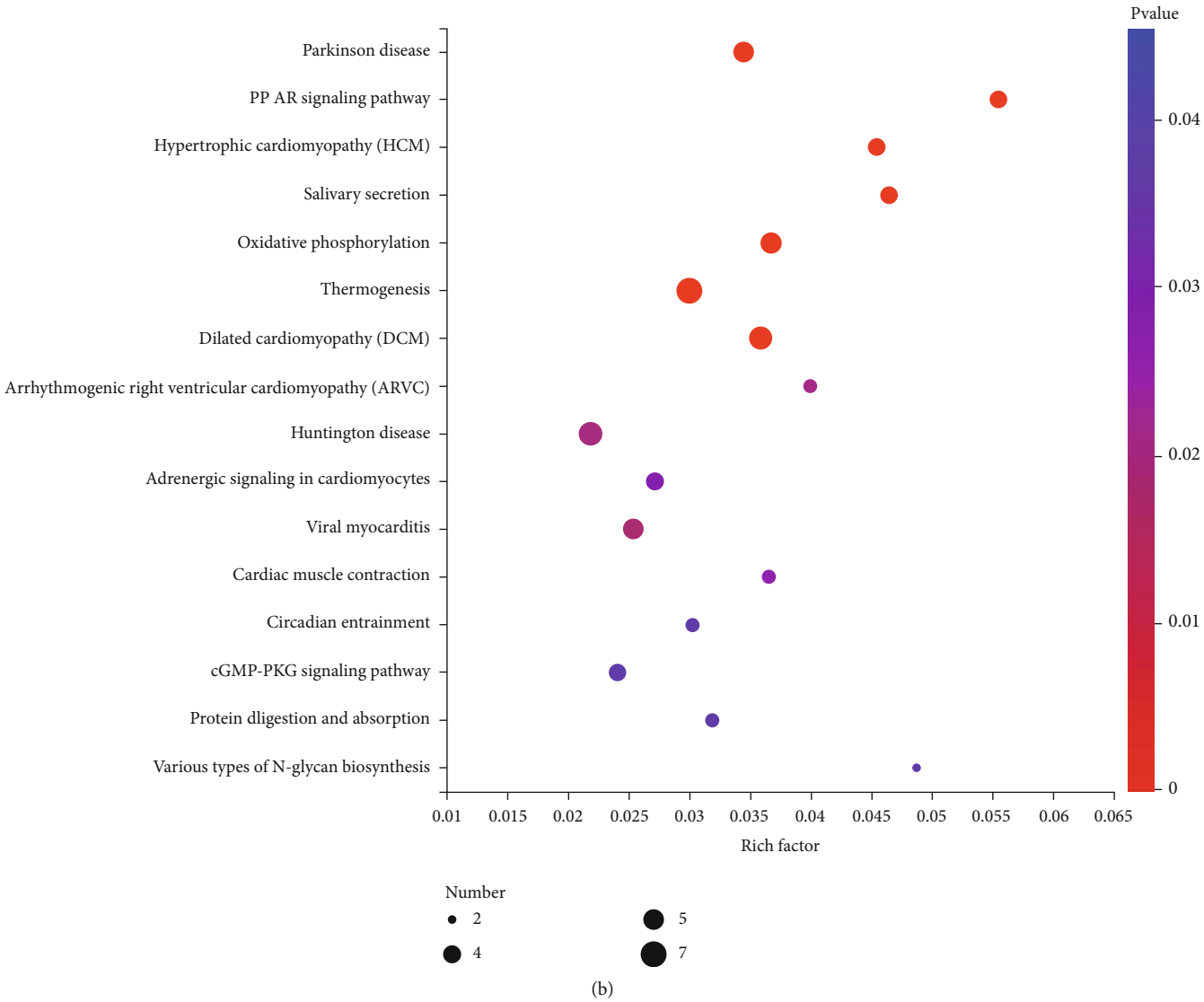


Figure 2: Functional enrichment analysis. (a) Top 10 GO enrichment terms in biological processes (BP), molecular functions (MF), and cellular component (CC) functions. (b) KEGG signaling pathways with $p < 0.05$.

miRNA–mRNA interaction networks were constructed and displayed using Cytoscape v3.8.2 [16].

2.6. Statistical Analysis. GraphPad Prism 8 was used for data analysis. Two-tailed Student's *t*-test was used to evaluate the differences between the two groups. Data are presented as the mean ± standard deviation. Statistical significance was defined as $p < 0.05$.

3. Results

3.1. circRNA Expression Profiling in STC. CIRI analysis of the high-throughput RNA sequencing results from the STC and control colon tissues identified 31082 circRNAs (Figure 1 (a)). Among the circRNAs, 70.83% was derived from exons, 22.63% were from introns, and 6.54% were intergenic (Figure 1(b)). There were 190 DE-circRNAs; 117 were upregulated and 73 were downregulated. Figure 1(c) shows a volcano plot of the DE-circRNAs. A hierarchically clustered heatmap revealed distinct differences between the circRNA expression profiles of the two groups (Figure 1(d)). The DE-circRNAs are listed in Supplementary Table S2.

3.2. Functional Enrichment Analysis of Host Genes. GO enrichment analysis of the host genes of the DE-circRNA transcripts revealed that they were enriched in 555 terms, comprising 413 biological processes (BP) terms, 64 cellular component (CC) terms, and 78 molecular function (MF) terms. The majority of BP host genes were enriched in actin filament–based processes (GO:0030029), the relaxation of smooth muscle (GO:0044557), cytoskeleton organization (GO:0007010), and actin cytoskeletal organization (GO:0030036). The most enriched MF terms were for cytoskeletal protein binding (GO:0008092), heparin sulfate binding (GO:1904399), and actin binding (GO:0003779). Among the CC terms, the host genes were associated primarily with

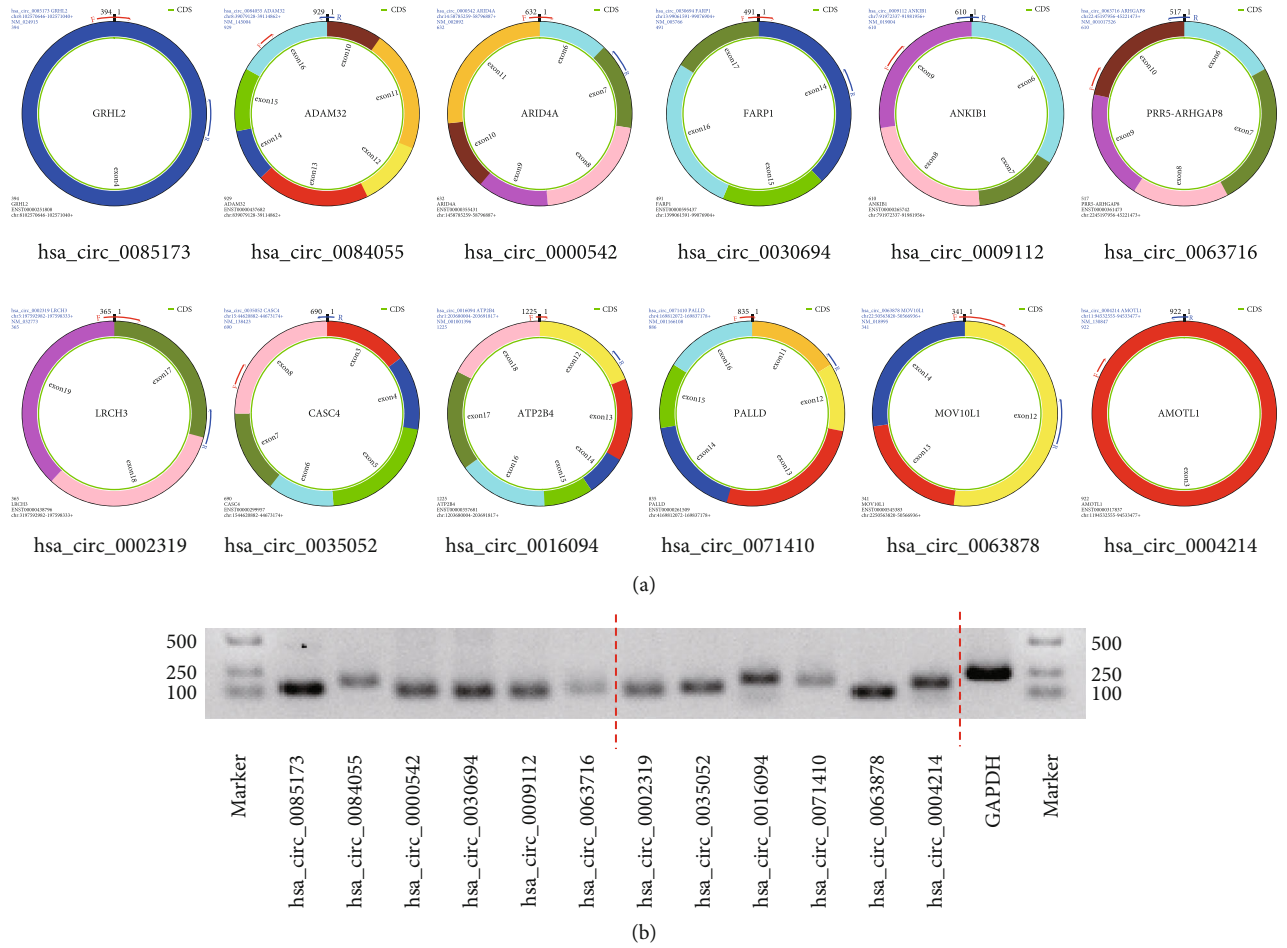


Figure 3: Presence of the selected circRNAs. (a) Detailed information on the circRNAs in circPrimer 2.0. (b) The circRNAs are amplified from cDNA of HEK 293T cells using divergent primers.

Table 2: Top six up-/downregulated circRNAs.

circRNA ID	circBase ID	Log2FC	*P* value	Regulate
8_101558419_101558812	hsa_circ_0085173	5.80	$1.44192E-05$	Up
8_39221610_39257343	hsa_circ_0084055	5.60	$7.49084E-05$	Up
14_58318542_58330169	hsa_circ_0000542	5.52	0.000132928	Up
13_98409338_98424650	hsa_circ_0030694	5.45	0.000237855	Up
7_92343024_92352642	hsa_circ_0009112	5.36	0.000400085	Up
22_44802077_44825593	hsa_circ_0063716	5.28	0.000688839	Up
3_197866112_197871462	hsa_circ_0002319	-5.42	$3.5989E-05$	Down
15_44328685_44380976	hsa_circ_0035052	-5.12	0.000370208	Down
1_203710877_203722689	hsa_circ_0016094	-5.12	0.000370208	Down
4_168890922_168916027	hsa_circ_0071410	-3.58	$8.67005E-05$	Down
22_50125392_50128507	hsa_circ_0063878	-3.58	$8.67005E-05$	Down
11_94799390_94800311	hsa_circ_0004214	-2.67	$8.83338E-07$	Down

secretory dimeric IgA immunoglobulin complexes (GO:0071752), dimeric IgA immunoglobulin complexes (GO:0071750), and the cytoskeleton (GO:0005856) (Figure 2(a)). In the KEGG signaling pathway analysis, the host genes were significantly enriched in 16 terms; dilated cardiomyopathy (map05414), thermogenesis (map04714), the PPAR signaling pathway (map03320), and oxidative phosphorylation (map00190) were the most significant

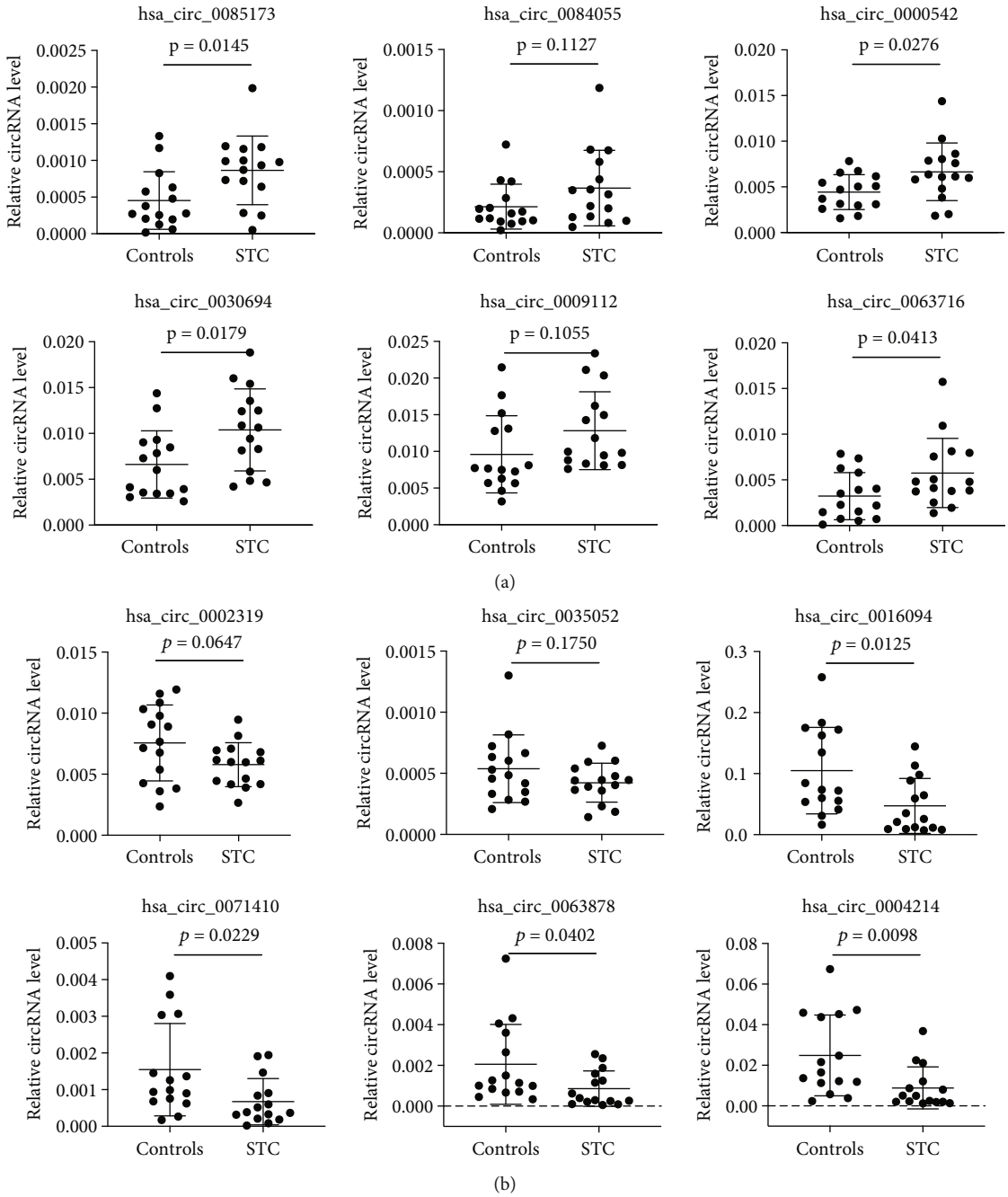
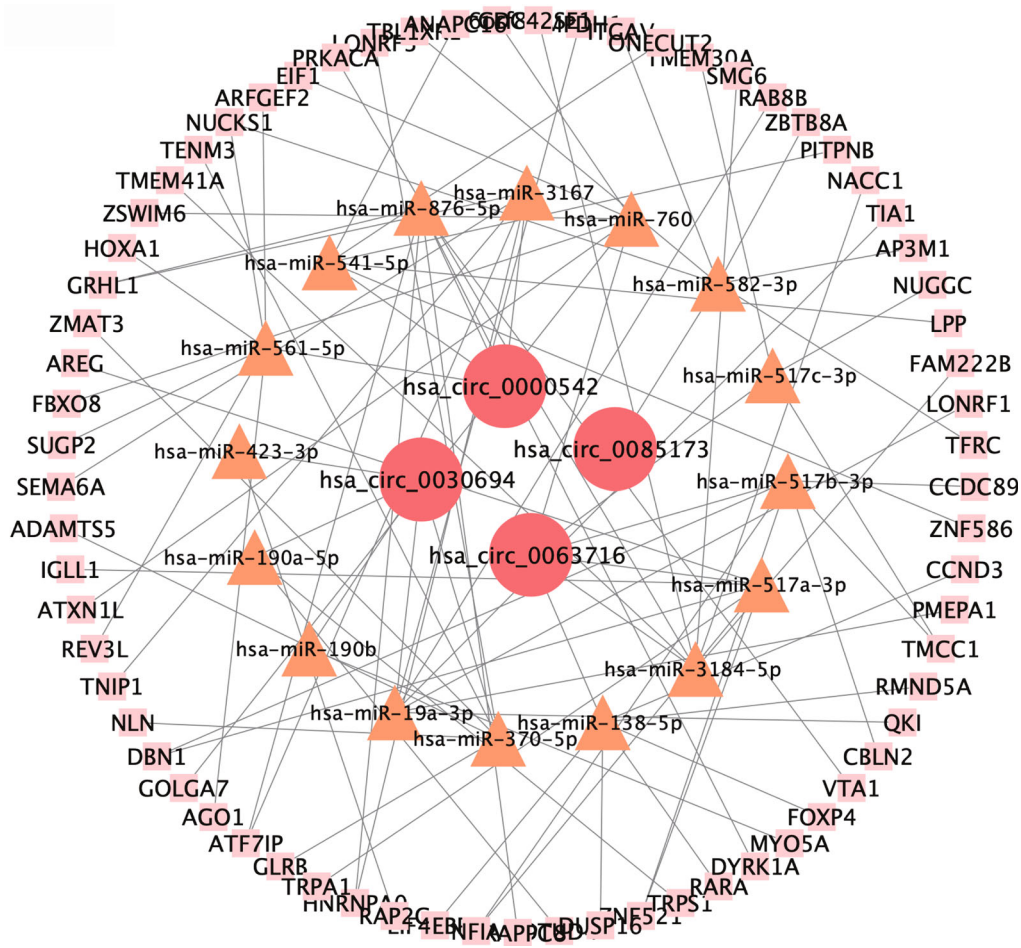


Figure 4: Validation of the expression of the selected circRNAs in STC tissues. (a, b) qRT-PCR analysis of the top 6 up-/downregulated circRNAs in the STC and control groups ($n = 15$ each). All data are means ± SD.

(Figure 2(b)). The detailed data are provided in Supplementary Tables S3 and S4.

3.3. Validation of DE-circRNAs. Further analysis of the DE-circRNAs using circBase identified 61 overlapping circRNAs and 129 novel circRNAs [17]. The top 6 DE-circRNAs that were derived from exons were selected for subsequent experiments (Figure 3(a), Table 2). First, divergent primers were designed to amplify the circRNAs from cDNA, and their presence was confirmed using agarose gel electrophoresis (Figure 3(b)). The expression of these DE-circRNAs was verified in 15 pairs of STC and non-STC tissues using qRT-PCR (15 vs. 15 samples). The results revealed that the candidate circRNAs were differentially expressed in the STC samples, which is consistent with the circRNA sequencing results. In the STC tissues, the levels of hsa_circ_0085173, hsa_circ_000542, hsa_circ_0030694, and hsa_circ_0063716 were significantly increased, whereas the levels of hsa_circ_0016094, hsa_circ_0071410, hsa_circ_0063878, and hsa_circ_0004214 were significantly decreased, compared with the control samples (Figures 4(a) and 4(b)).



(a)

Figure 5: Continued.

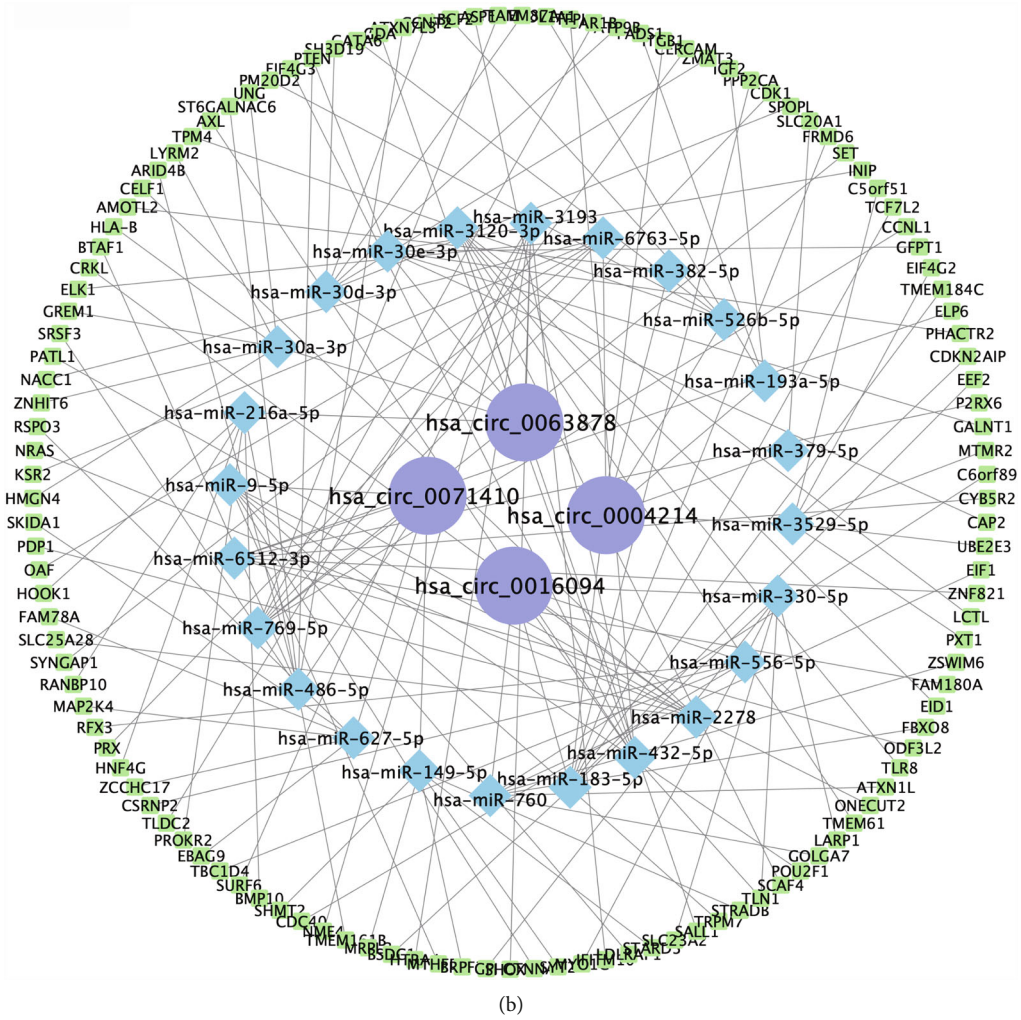


Figure 5: The ceRNA regulatory network. (a) The upregulated ceRNA network. (b) The downregulated ceRNA network.

3.4. Construction of the circRNA–miRNA–mRNA Interaction Network in STC. circRNAs can act as miRNA sponges to regulate miRNA activities and influence downstream mRNA expression. Thus, we constructed a circRNA–miRNA–mRNA interaction network to investigate the role of circRNAs in STC. The starBase and circBank databases were used to predict the target miRNAs of the selected circRNAs. TargetScan, miRDB, miRWalk, and TarBase were used to predict the downstream genes of the identified miRNAs. Because thousands of interaction pairs were predicted for the DE-circRNAs, we established and visualized competing endogenous RNA (ceRNA) networks for the validated circRNAs. Using the regulatory network map, we identified the top 6 miRNAs that potentially bind to the circRNAs and the 6 most likely target genes for each miRNA (Figures 5(a) and 5(b), Supplementary Table S5). This might provide a foundation for understanding the biological functions of circRNAs in STC.

4. Discussion

STC is a primary functional disease characterized by impaired colonic function and decreased motility. Surgery may be the definitive therapy for patients with refractory STC who fail to respond to medical treatment [18]. To date, the detailed mechanisms of STC have not been fully elucidated. In recent years, ncRNAs have been confirmed to participate in the occurrence and development of various diseases by directly or indirectly regulating gene expression [19]. Exploring the changes in ncRNAs in STC patients will help enrich our understanding of the pathogenesis and potential therapeutic targets of STC. Using human specimens, we compared circRNA expression patterns between STC and control colon tissues via whole-transcriptome sequencing.

This study identified 190 circRNAs as being significantly differentially expressed in patients with STC compared with controls. GO and KEGG enrichment analyses showed that the host genes of these circRNAs were enriched in the relaxation of smooth muscle, actin binding, dilated cardiomyopathy, and cardiac muscle contraction. These results suggest that STC pathogenesis may be related to myopathy. Smooth muscle cells are the final effectors of gastrointestinal motility [20]. Studies have shown that colonic smooth muscle cells are impaired in patients with STC [21]. Our previous study demonstrated thinning of the intestinal smooth muscle layer in chronically constipated mice [22]. However, whether

injury of smooth muscle cells is a primary pathological process of STC or a secondary result of fecal deposition of denervation remains to be further studied. The results of the present study suggest that abnormal expression of circRNAs may be involved in the regulation of smooth muscle injury in patients with STC.

The 12 circRNAs with the most significant abnormal expression were verified via qRT-PCR using 15 paired colon tissues from STC and controls. The results were consistent with the sequencing data. Among the identified circRNAs, four were significantly overexpressed in STC tissues, and four others were significantly downregulated compared with the controls. Since ceRNA mechanisms are important pathophysiological pathways of ncRNAs and ceRNA networks that are out of balance can disrupt life activities and cause disease [23], we constructed circRNA–miRNA–mRNA interaction networks for the top 8 circRNAs that were differentially expressed in both the whole-transcriptome sequencing and qRT-PCR tests.

Interestingly, some mRNAs predicted to be closely related to smooth muscle function were found in these networks. For example, it has been suggested that there may be a ceRNA relationship between hsa_circ_0004214, hsa-miR-526b-5p, and tropomyosin 4 (*TPM4*). TPM4 is a major F-actin–binding protein that plays important roles in modulating muscle contraction [24]. Common binding sites were predicted among hsa_circ_0071410, hsa-miR-149-5p, and synaptotagmin 2 (*SYT2*), thus constituting a potential ceRNA network. SYT2 is a key protein in the neuromuscular junction and is essential for fast synaptic vesicle exocytosis. Studies have shown that *SYT2* is one of the disease genes responsible for congenital myasthenic syndromes [25]. Neuromuscular junction disorders may also play crucial roles in STC; this provides ideas for further understanding of the pathogenesis of the disease. Moreover, hsa_circ_0004214 and hsa-miR-193a-5p shared binding sites with COL1A1, which is the major component of type 1 collagen, suggesting that they may participate in the pathogenesis of STC through ceRNA mechanisms. Although collagen is not a direct component of smooth muscle cell contraction elements, the extracellular matrix composed of collagen plays an important role in regulating muscle contraction [26]. Smooth muscle has been shown to mediate extracellular matrix remodeling, which indirectly regulates overall muscle tissue contractility [27].

Although the ability to bind to miRNAs is the best described mechanism of circRNAs, other functions of circRNAs should not be ignored, including participation as RNA-binding proteins [28], transcriptional regulators [29], and the ability to directly translate proteins [30]. Comprehensive exploration and innovative research on the differentially expressed circRNAs identified in the current study may contribute to further understanding the pathogenesis of STC and provide potential targets for its treatment.

5. Conclusions

This is the first study to summarize the differential expression of circRNAs in human STC colon tissues. Bioinformatics methods were used to identify the GO and KEGG pathways of these DE-circRNAs to understand their potential mechanisms of action. Moreover, 12 circRNAs with the most differentially expressed genes were verified using qRT-PCR, and their predicted ceRNA networks were constructed. Taken together, the findings of this study provide a helpful blueprint for researchers to select circRNAs for further study of their corresponding mechanisms. Moreover, it is hoped that our findings will highlight the role of circRNAs in STC and stimulate the exploration and development of new therapeutic targets.

Data Availability

The authors declare that the data supporting the findings of this study are available within the article and its Supplementary Information File, or from the corresponding authors upon request. Raw data is available at BioSample, accession number PRJNA826640.

Conflicts of Interest

The authors declares that they have no conflict of interests.

Authors' Contributions

Changlei Xi, Yuntian Hong, and Baoxiang Chen contributed equally to this work and share first authorship. Xianghai Ren and Congqing Jiang designed and conceived this project. Xianghai Ren and Yuntian Hong for paperwriting; Yuntian Hong and Baoxiang Chen for experimental verification; Changlei Xi, Yuntian Hong, and Baoxiang Chen analyzed the data; Qun Qian, Congqing Jiang, Weicheng Liu, Xianghai Ren, Changlei Xi, Xiaoyu Xie for surgery and sample collection; All authors read, revised, and approved the final manuscript.

Acknowledgments

This work was supported by grants from the Medical Science and Technology Innovation Platform of Health Commission of Hubei Province/Zhongnan Hospital of Wuhan University (grant number PTXM2019011 to C.J.), the Clinical Research Special Fund of Wu Jieping Medical Foundation (grant number 320.6750.2021-11-8 to C.J.), the Engineering construction project of improving diagnosis and treatment ability of difficult diseases (oncology) (grant number ZLYNXM202012 to Q.Q.), and Zhongnan Hospital of Wuhan University/Hubei Health Commission Joint Fund Project (grant number znpy2019086 to C.J.).

Supplementary Materials

Supplementary 1. Table S1: PCR primers.

Supplementary 2. Table S2: differentially expressed circRNAs in STC.

Supplementary 3. Table S3: the GO enrichment terms.

Supplementary 4. Table S4: the KEGG pathway enrichment terms.

Supplementary 5. Table S5: top 6 upregulated/downregulated circRNA–miRNA–mRNA regulatory network.

References

[1] H. Vriesman Mana and J. N. Koppen Ilan, "Management of functional constipation in children and adults," *Nature Reviews Gastroenterology & Hepatology*, vol. 17, no. 1, pp. 21–39, 2020.

[2] Y. Kawamura, S. Yamamoto, Y. Funaki et al., "Internet survey on the actual situation of constipation in the Japanese population under 70 years old: focus on functional constipation and constipation-predominant irritable bowel syndrome," *Journal of gastroenterology*, vol. 55, no. 1, pp. 27–38, 2020.

[3] S. Wan, W. Liu, C. Tian et al., "Differential proteomics analysis of colonic tissues in patients of slow transit constipation," *BioMed Research International*, vol. 2016, Article ID 4814702, 6 pages, 2016.

[4] C. H. Knowles and J. E. Martin, "Slow transit constipation: a model of human gut dysmotility. Review of possible aetiologies," *Neurogastroenterology & Motility*, vol. 12, pp. 181–196, 2000.

[5] G. L. Lyford, C.-L. He, E. Soffer et al., "Pan-colonic decrease in interstitial cells of Cajal in patients with slow transit constipation," *Gut*, vol. 51, no. 4, pp. 496–501, 2002.

[6] A. Attaluri, M. Jackson, J. Paulson, and S. S. C. Rao, "Methanogenic flora is associated with altered colonic transit but not stool characteristics in constipation without IBS," *The American journal of gastroenterology,*, vol. 105, no. 6, pp. 1407–1411, 2010.

[7] A. Mazzone, P. R. Strege, S. J. Gibbons et al., "microRNA overexpression in slow transit constipation leads to reduced NaV1.5 current and altered smooth muscle contractility," *Gut*, vol. 69, no. 5, pp. 868–876, 2020.

[8] S. Zhao, Q. Chen, X. Kang, B. Kong, and Z. Wang, "Aberrantly Expressed Genes and miRNAs in Slow Transit Constipation Based on RNA-Seq Analysis," *BioMed Research International*, vol. 2018, Article ID 2617432, 14 pages, 2018.

[9] H. Zheng, Y. J. Liu, Z. C. Chen, and G. Q. Fan, "miR-222 regulates cell growth, apoptosis, and autophagy of interstitial cells of Cajal isolated from slow transit constipation rats by targeting c-kit," *Indian Journal of Gastroenterology*, vol. 40, no. 2, pp. 198–208, 2021.

[10] Y. Y. Wang, R. Y. Lu, J. Shi, S. Zhao, X. Jiang, and X. Gu, "CircCORC2 is involved in the pathogenesis of slow transit constipation via modulating the signalling of miR-19a and neurotensin/motilin," *Journal of Cellular and Molecular Medicine*, vol. 25, no. 8, pp. 3754–3764, 2021.

[11] B. Li, J. Yang, J. He et al., "Spatiotemporal regulation and functional analysis of circular RNAs in skeletal muscle and subcutaneous fat during pig growth," *Biology (Basel)*, vol. 10, no. 9, p. 841, 2021.

[12] Z. Wen, Q. Shen, H. Zhang et al., "Circular RNA CCDC66 targets DCX to regulate cell proliferation and migration by sponging miR-488-3p in Hirschsprung's disease," *Journal of cellular physiology*, vol. 234, no. 7, pp. 10576–10587, 2019.

[13] B. Liu, B. Ye, X. Zhu et al., "An inducible circular RNA *circKcnt2* inhibits ILC3 activation to facilitate colitis resolution," *Nature communications*, vol. 11, no. 1, p. 4076, 2020.

[14] P. Zhu, X. Zhu, J. Wu et al., "IL-13 secreted by ILC2s promotes the self-renewal of intestinal stem cells through circular RNA circPan3," *Nature Immunology*, vol. 20, no. 2, pp. 183–194, 2019.

[15] W. Liu, Q. Zhang, S. Li et al., "The relationship between colonic macrophages and microRNA-128 in the pathogenesis of slow transit constipation," *Digestive diseases and sciences*, vol. 60, no. 8, pp. 2304–2315, 2015.

[16] 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.

[17] P. Glažar, P. Papavasileiou, and N. Rajewsky, "circBase: a database for circular RNAs," *RNA*, vol. 20, no. 11, pp. 1666–1670, 2014.

[18] A. E. Bharucha and A. Wald, "Chronic constipation," *Mayo Clinic Proceedings*, vol. 94, no. 11, pp. 2340–2357, 2019.

[19] G. J. Goodall and V. O. Wickramasinghe, "RNA in cancer," *Nature Reviews Cancer*, vol. 21, no. 1, pp. 22–36, 2021.

[20] J. Luo, A. Qian, L. K. Oetjen et al., "TRPV4 channel signaling in macrophages promotes gastrointestinal motility via direct effects on smooth muscle cells," *Immunity*, vol. 49, no. 1, pp. 107–119.e4, 2018.

[21] L. Leung, T. Riutta, J. Kotecha, and W. Rosser, "Chronic constipation: an evidence-based review," *The Journal of the American Board of Family Medicine*, vol. 24, no. 4, pp. 436–451, 2011.

[22] Y. Hong, X. Ren, W. Liu et al., "miR-128 participates in the pathogenesis of chronic constipation by regulating the p38α/M-CSF inflammatory signaling pathway," *American Journal of Physiology-Gastrointestinal and Liver Physiology*, vol. 321, no. 4, pp. G436–G447, 2021.

[23] P. Wang, Q. Guo, Y. Qi et al., "LncACTdb 3.0: an updated database of experimentally supported ceRNA interactions and personalized networks contributing to precision medicine," *Nucleic acids research*, vol. 50, no. D1, pp. D183–D189, 2022.

[24] Y. S. Borovikov, D. D. Andreeva, S. V. Avrova et al., "Molecular mechanisms of the deregulation of muscle contraction induced by the R90P mutation in Tpm3.12 and the weakening of this effect by BDM and W7," *International Journal of Molecular Sciences*, vol. 22, no. 12, p. 6318, 2021.

[25] D. N. Herrmann, R. Horvath, J. E. Sowden et al., "Synaptotagmin 2 mutations cause an autosomal-dominant form of Lambert-Eaton myasthenic syndrome and nonprogressive motor neuropathy," *The American Journal of Human Genetics*, vol. 95, no. 3, pp. 332–339, 2014.

[26] H. P. Ehrlich, G. M. Allison, M. J. Page, W. A. Kolton, and M. Graham, "Increased gelsolin expression and retarded collagen lattice contraction with smooth muscle cells from Crohn's diseased intestine," *Journal of cellular physiology*, vol. 182, no. 2, pp. 303–309, 2000.

[27] J. E. Bourke, X. Li, S. R. Foster et al., "Collagen remodelling by airway smooth muscle is resistant to steroids and β_2-agonists," *European Respiratory Journal*, vol. 37, no. 1, pp. 173–182, 2011.

[28] T. L. Okholm, S. Sathe, S. S. Park et al., "Transcriptome-wide profiles of circular RNA and RNA-binding protein interactions reveal effects on circular RNA biogenesis and cancer pathway expression," *Genome medicine*, vol. 12, no. 1, p. 112, 2020.

[29] W. Y. Zhou, Z. R. Cai, J. Liu, D. S. Wang, H. Q. Ju, and R. H. Xu, "Circular RNA: metabolism, functions and interactions with proteins," *Molecular cancer*, vol. 19, no. 1, p. 172, 2020.

[30] T. Sinha, C. Panigrahi, D. Das, and A. Chandra Panda, "CircularRNAtranslation, a path to hidden proteome," *Wiley Interdisciplinary Reviews: RNA*, vol. 13, no. 1, article e1685, 2022.