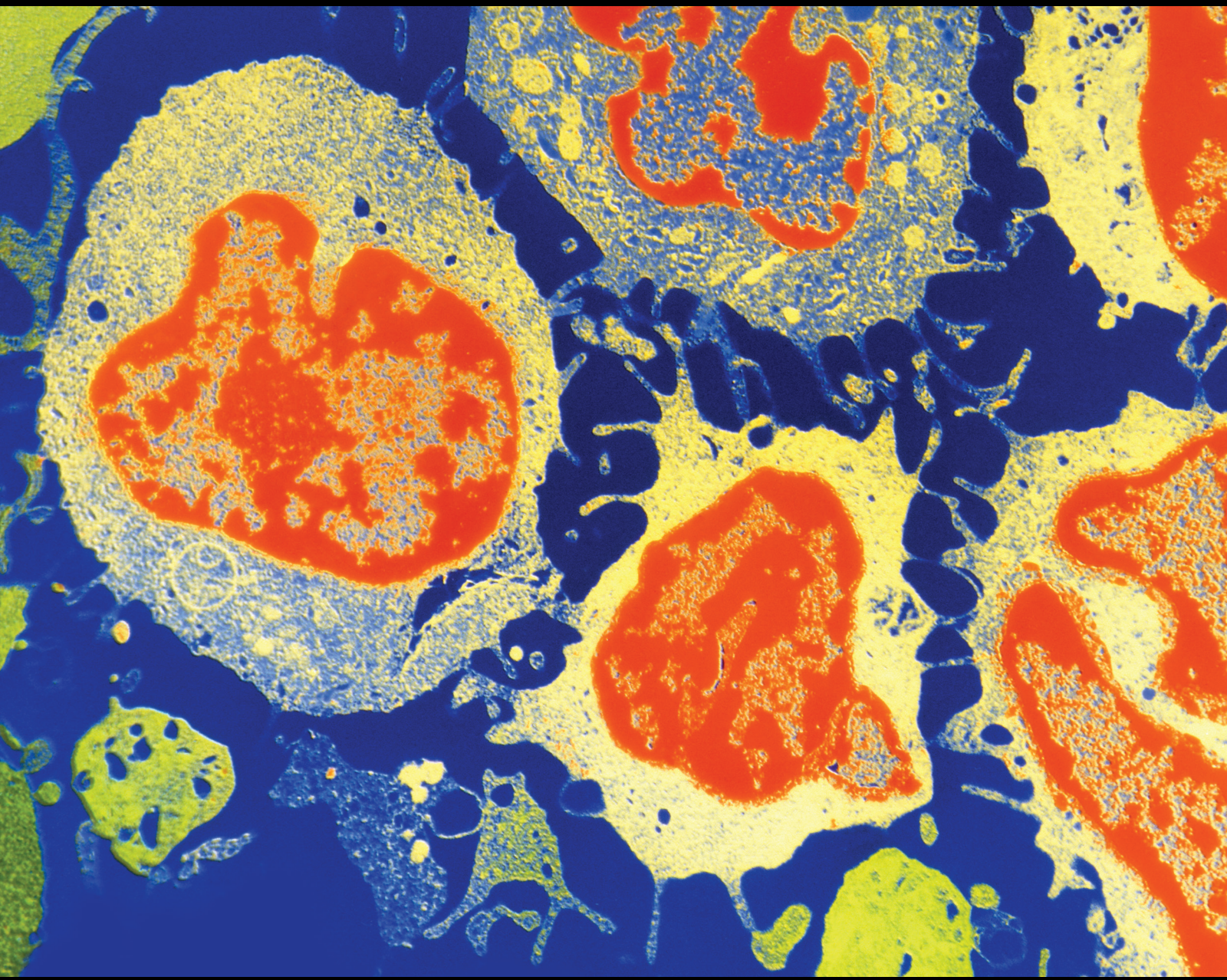
Journal of Oncology

# Machine Learning-Based Methods for Multi-Omics Data Analysis in Cancer

Lead Guest Editor: Jincheng Guo
Guest Editors: Dechao Bu, Hongqing Cai, Jian Zhang, and Xinyi Liu



# Machine Learning-Based Methods for Multi-Omics Data Analysis in Cancer

Journal of Oncology

# Machine Learning-Based Methods for Multi-Omics Data Analysis in Cancer

Lead Guest Editor: Jincheng Guo
Guest Editors: Dechao Bu, Hongqing Cai, Jian Zhang, and Xinyi Liu

# Contents

Hindawi
Journal of Oncology
Volume 2022, Article ID 9039110, 18 pages
https://doi.org/10.1155/2022/9039110

*Research Article*

# Methylation-Mediated Silencing of *RBP7* Promotes Breast Cancer Progression through PPAR and PI3K/AKT Pathway

**Hong Lin**,[1,2] **Qizheng Han**,[2] **Junhao Wang**,[2] **Zhaoqian Zhong**,[2] **Haihua Luo**,[2] **Yibin Hao**,[1] **and Yong Jiang**[2]

[1]*The fifth Clinical Medical College of Henan University of Chinese Medicine, Henan University of Chinese Medicine, No. 33 Huanghe Road, Zhengzhou, 410105 Henan, China*
[2]*Guangdong Provincial Key Laboratory of Proteomics, State Key Laboratory of Organ Failure Research, Department of Pathophysiology, School of Basic Medical Sciences, Southern Medical University, No. 1023, South Shatai Road, Baiyun District, Guangzhou, 510515 Guangdong, China*

Correspondence should be addressed to Yibin Hao; haoyibin450@126.com and Yong Jiang; jiang48231@163.com

Received 25 March 2022; Revised 6 July 2022; Accepted 26 September 2022; Published 13 October 2022

Academic Editor: Dechao Bu



Retinoid-binding protein7 (*RBP7*) is a member of the cellular retinol-binding protein (CRBP) family, which is involved in the pathogenesis of breast cancer. The study aims to illustrate the prognostic value and the potential regulatory mechanisms of *RBP7* expression in breast cancer. Bioinformatics analysis with the TCGA and CPTAC databases revealed that the mRNA and protein expression levels of *RBP7* in normal were higher compared to breast cancer tissues. Survival analysis displayed that the lower expression of *RBP7,* the worse the prognosis in ER-positive ($ER^+$) breast cancer patients. Genomic analysis showed that low expression of *RBP7* correlates with its promoter hypermethylation in breast cancer. Functional enrichment analysis demonstrated that downregulation of *RBP7* expression may exert its biological influence on breast cancer through the PPAR pathway and the PI3K/AKT pathway. In summary, we identified *RBP7* as a novel biomarker that is helpful for the prognosis of $ER^+$ breast cancer patients. Promoter methylation of *RBP7* is involved in its gene silencing in breast cancer, thus regulating the occurrence and development of $ER^+$ breast cancer through the PPAR and PI3K/AKT pathways.

## 1. Introduction

Breast cancer is the most common cancer worldwide, accounting for 30% of female cancers [1]. Estrogen receptor-positive ($ER^+$) breast cancer is driven by ER-mediated transcriptional activity, composing the major subtype (approximately 75%) of breast cancer [2]. Although endocrine therapy, including estrogen suppression and direct ER targeting, is widely applied in the treatment of $ER^+$ breast cancer, acquired resistance often occurs and remains a major challenge for the treatment of $ER^+$ patients [3]; thus, novel targets and effective therapeutic strategies for breast cancer patients are urgently needed.

Previous studies have confirmed that cellular retinol-binding protein (CRBP) family members play an important role in the pathological progression of breast cancer. CRBPs belong to the family of fatty acid-binding proteins and are required for vitamin A stability and metabolism [4]. Epigenetic silencing of CRBPs is a common event in cancers [5]. For example, Kuppumbatti et al. reported that CRBPs were underexpressed in 24% of human breast cancer [6]. Previous studies demonstrated that inhibition of the PI3K/AKT pathway by CRBPs was involved in the growth inhibition of mammary epithelial cells [7]. In breast epithelial cells, CRBP1 inhibits PI3K/AKT signalling through a retinoic acid receptor-dependent mechanism that regulates p85-p110 heterodimerization [8]. In addition, physiological retinoic acid receptor (RAR) activation is dependent on CRBP1-mediated retinol storage, and CRBP1 downregulation chronically compromises RAR activity, leading to loss of cell

differentiation and tumor progression [9]. Thus, CRBPs may be used as a potential biomarker for the diagnosis and treatment of breast cancer.

Retinoid-binding protein 7 (*RBP7*), also named CRBP4, belongs to a clearly distinct CRBP subfamily, representing a relatively different mode of retinol binding for this protein. The *RBP7* gene is located on human chromosome 1p36.22, encoding a protein of 134 amino acids in length. The size of the translated exon sequences and intron position of *RBP7* is highly conserved. With a structure similar to other CRBPs, the *RBP7*-encoded protein binds all-trans-retinol with a lower binding affinity than other CRBPs [10]. It was reported that *RBP7* regulates the occurrence and development of many diseases. For example, the PPAR$\gamma$-*RBP7*-adiponectin pathway plays a protective role in hypertensive diseases by regulating transcriptional activity [11]. *RBP7* also plays an important role in adipose tissue during adipogenesis, cold exposure, and nutritional treatment [12]. Recent studies have demonstrated that *RBP7* is a strong prognostic biomarker contributing to the malignant phenotype in colon cancer [13]. However, as a novel member of the CRBP family, the clinical and prognostic significance of *RBP7* in breast cancer is still unknown, and its functional role in breast cancer has never been documented.

In this study, we first analyzed the differential expression of *RBP7* in breast cancer and normal tissues and evaluated the prognostic value of *RBP7* by using data from the TCGA and GEO databases. Next, we determined the localization of *RBP7* expression in breast cancer, the functional enrichment of its coexpressed genes, and the association between the mRNA expression and DNA methylation of *RBP7*. Then, we explored the association between *RBP7* and multiple molecular subtypes of breast cancer and the significant KEGG pathways involved in *RBP7* in $ER^+$ breast cancer. Finally, we screened *RBP7*-targeting drugs from computational analysis of resistance (CARE) databases, which may provide new ideas for the treatment of breast cancer.

## 2. Materials and Methods

*2.1. Gene Expression and Survival Analysis.* The Human Protein Atlas [14, 15] (HPA; http://www.proteinatlas.org) database was used to illustrate *RBP7* mRNA distribution, protein expression, and immunohistochemical maps of *RBP7* in normal and breast cancer tissues. *RBP7* gene expression levels in pan cancers were identified in Tumor Immune Estimation Resource [16] (TIMER; https://cistrome.shinyapps.io/timer/) and ONCOMINE [17] (http://www.oncomine.org). UALCAN [18] (http://ualcan.path.uab.edu/index.html) and the Gene Expression Omnibus [19] (GEO; https://www.ncbi.nlm.nih.gov/geo/) were used to investigate the different expression of *RBP7* in normal and breast cancer tissues. To explore the prognostic role of *RBP7* expression in breast cancer, Gene Expression Profiling Interactive Analysis [20] (GEPIA; gepia2.cancer-pku.cn/#index), Kaplan–Meier plotter [21] (http://kmplot.com/), and PROGgeneV2 [22] (http://www.compbio.iupui.edu/proggene) were used to determine the prognostic significance. In this study, we analyzed the prognosis of *RBP7* in breast cancer with detailed hazard ratios (HRs) by setting the expression threshold at the medium or best cutoff and *a* log-rank *p value* less than 0.05.

*2.2. Tumor Immune Single-Cell Hub (TISCH).* TISCH [23] (http://tisch.comp-genomics.org) integrates single-cell transcriptome profiles of nearly 2 million cells for 27 cancer types. In this study, we utilized the "multiple-dataset comparison" model to visualize the averaged gene expression distributed in single cells and the "Gene" module to display the heat map of the cell-type averaged expression of *RBP7*.

*2.3. LinkedOmics.* The coexpressed genes of *RBP7* were screened from the TCGA BRCA (breast invasive carcinoma) cohort through the "LinkFinder" module in LinkedOmics [24] (http://www.linkedomics.org/login.php) databases, and the correlative significance was tested by the spearman correlation coefficient. The top 50 positively and negatively correlated genes are presented as heat maps. Gene Ontology biological process (GO_BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed with gene set enrichment analysis (GSEA) in the "LinkInterpreter" module.

*2.4. RBP7 DNA Methylation Analysis.* Heat maps of *RBP7* in the cohort of breast cancer patients were constructed through data mining in TCGA BRCA by using the University of California Santa Cruz (UCSC) Xena [25, 26] (http://xena.ucsc.edu/). MethSurv (http://biit.cs.ut.ee/methsurv/) is used to survival analysis ground on CpG methylation patterns [27], we verified the methylation levels of probes of *RBP7,* and four of them with high methylation in the promoter were chosen to display the distribution of methylation under different clinical stages. SMART [28, 29] (http://www.bioinfo-zs.com/smartapp/) is to further identify the association of the mRNA expression and methylation of *RBP7*.

*2.5. Bc-GenExMiner Online Tool and OSbrca.* Based on common clinical parameters, we utilized Bc-GenExMiner (v4.7) [30] (http://bcgenex.ico.unicancer.fr) to analyze the expression data and survival curves of *RBP7* in different molecular subtypes of breast cancer, including ER, PR, and HER-2 (IHC). The OSbrca [31] (http://bioinfo.henu.edu.cn/BRCA/BRCAList.jsp) was utilized to validate prognostic value of *RBP7* in breast cancer.

*2.6. Protein–Protein Interaction (PPI) Network Analysis.* PPI network analysis of *RBP7* was conducted in the STRING (https://string-db.org/) database [32]. The regulatory relationships between genes were visualized *via* Cytoscape (ver. 3.4.0). Then, we use the starBase v3.0 [33] (http://starbase.sysu.edu.cn/index.php) to analyze the correlation between *RBP7* and PIK3R3 in breast cancer.

*2.7. Differentially Expressed Genes (DEGs) Analysis.* We used the Limma package to screen DEGs by filtering the *p.adj* value of Student's *t*-test and the fold change (FC) and dividing the DEGs into two groups with high or low *RBP7* expression. A volcano plot was generated by using the ggplot2 R software package to display the DEGs with statistical significance, i.e., *p*.adjust value <0.05 and absolute FC value >1.

KEGG pathway analysis was performed on those DEGs by using the cluster profiler package, and the pathways with statistical significance (adjusted $p < 0.05$) were visualized by hierarchical clustering of a heat map [34].

*2.8. Computational Analysis of Resistance (CARE).* A positive CARE score represented a high expression value, which was related to drug response in CARE [35] (http://care.dfci.harvard.edu/), and vice versa. In this study, we utilized 3 databases, i.e., Cancer Cell Line Encyclopedia (CCLE), Cancer Therapeutics Response Portal (CGP), and Genomics of Drug Sensitivity in Cancer (CTRP), to analyze the drugs targeting *RBP7*.

*2.9. SwissDock.* The PDF file of the *RBP7* protein was downloaded from the RCSB Protein Data Bank (PDB) database, and the ligand and water molecules were then removed by using PyMOL software. The mol2 file of the nilotinib small molecule was downloaded from the PubChem database and converted using Open Babel software. Finally, we uploaded the two files to the SwissDock [36] (http://www.swissdock.ch) page for docking.

## 3. Results

*3.1. Gene Expression Profiles of RBP7 in Normal and Cancer Tissues.* We utilized the HPA database to analyze the mRNA and protein expression profiles of *RBP7* in human normal tissues. We found that the *RBP7* mRNA expression was mainly in breast and adipose tissues in normal human tissues by using the GTEx (genotype-tissue expression) database (Figure 1(a)). Consistently, the protein expression of *RBP7* was highly expressed in the breast and adipose tissues (Figure 1(b)). The mRNA expression levels of *RBP7* were explored by TIMER in many cancer types. Additionally, the results revealed that *RBP7* mRNA expression levels were significantly lower in most cancer samples than their corresponding normal samples, including breast invasive carcinoma (BRCA), uterine corpus endometrial carcinoma (UCEC), and lung adenocarcinoma (LUAD). Besides, the data also showed that *RBP7* expression was aberrantly higher in liver hepatocellular carcinoma (LIHC) and kidney renal clear cell carcinoma (KIRC) (Figure 1(c)).

To further verify the significance of *RBP7* expression in cancers, the differential expression of *RBP7* in tumor and normal tissues was analyzed by using ONCOMINE. We found that *RBP7* was overexpressed in liver cancer and lymphoma, but decreased expression of *RBP7* was found in brain and CNS cancer, breast cancer, esophageal cancer, head and neck cancer, leukemia, and ovarian cancer (Supplementary Figure 1A).

*3.2. Prognostic Value of RBP7 Expression in Breast Cancer.* To investigate the role of *RBP7* expression in breast cancer, the UALCAN database was utilized to analyze the expression of *RBP7* in 114 normal tissues and 1097 primary breast cancer tissues. The TCGA database results revealed that the mRNA expression level of *RBP7* was lower in breast cancer than in normal tissues (Figure 2(a)), which was further validated by the GSE37751 dataset from the GEO database (Supplementary Figure 1B). Then, we utilized the CPTAC database to analyze the protein expression of *RBP7* in breast cancer and found that the protein expression of *RBP7* in breast cancer was lower than that in normal tissues (Figure 2(b)), which was consistent with the result of *RBP7* expression in the TCGA database.

We subsequently utilized the transcriptomic sequencing data in the GEPIA database to assess the prognostic value of *RBP7* in breast cancer and found that a high level of expression of *RBP7* was favorable to the prognosis of breast cancer (Figure 2(c)). Kaplan–Meier survival analysis was performed with the GSE20685 (Figure 2(d)) and GSE42568 (Figure 2(e)) datasets from the GEO database to evaluate the prognostic value of *RBP7* in breast cancer, which showed that lower expression of *RBP7* was associated with a poorer prognosis in breast cancer. Furthermore, we used FROGgeneV2 to confirm the effect of *RBP7* on the OS of breast cancer patients. The results also showed that low expression of *RBP7* was significantly associated with a poor prognosis (Figure 2(f)). These results reveal that there is a significant association between *RBP7* expression and breast cancer prognosis and that *RBP7* serves as a protective factor in the prognosis of breast cancer.

*3.3. RBP7 Expression in Different Cells from Breast Cancer Tissues.* We used the TISCH database to analyze the expression of *RBP7* in different cells from breast cancer tissues. It demonstrated that *RBP7* is mainly expressed in endothelial and epithelial cells in TISCH database (Figures 3(a) and 3(b)). A heat map of the gene module analysis depicts the expression of *RBP7* in different cell types of breast cancer datasets, in which endothelial and epithelial cells are characterized by high expression of *RBP7* (Figure 3(c)). Then, we utilized the immunohistochemistry (IHC) data detected by the HPA-034749 antibody from the HPA database to determine the protein expression of *RBP7* in breast cancer and normal tissues. The results showed that adipocytes were highly stained in normal breast tissues, while glandular and myoepithelial cells were mildly stained, mainly in the nucleus (Figure 3(d)). In the breast cancer samples, the expression level of *RBP7* in tumor cells was ranked as weak, moderate, and strong, which was scored by the staining intensity in the pathological IHC (Figures 3(e)–3(g)).Interestingly, as a nuclear receptor, *RBP7* was found to be mainly localized in the nucleus, indicating the important role of *RBP7* in the regulation of gene expression in epithelial cells of breast cancer tissues.

*3.4. RBP7 Coexpression Networks in Breast Cancer.* To gain insight into the biological meaning of *RBP7* in breast cancer, the functional module of LinkedOmics was used to examine *RBP7* coexpression genes in the breast cancer cohort. The top 50 significant genes that were positively and negatively correlated with *RBP7* were selected as heat maps (Supplementary Figures 2A and 2B) in which *RBP7* displayed a strong positive association with the expression of FAM107A ($R = 0.4767$, $p$ *value* =4.252e-01), GPIHBP1 ($R = 0.4757$, $p$ *value* =8.560e-63), and FXYD1 ($R = 0.4756$, $p$ *value* =9.258e-63).Remarkably, the top 50 negatively coexpressed genes had

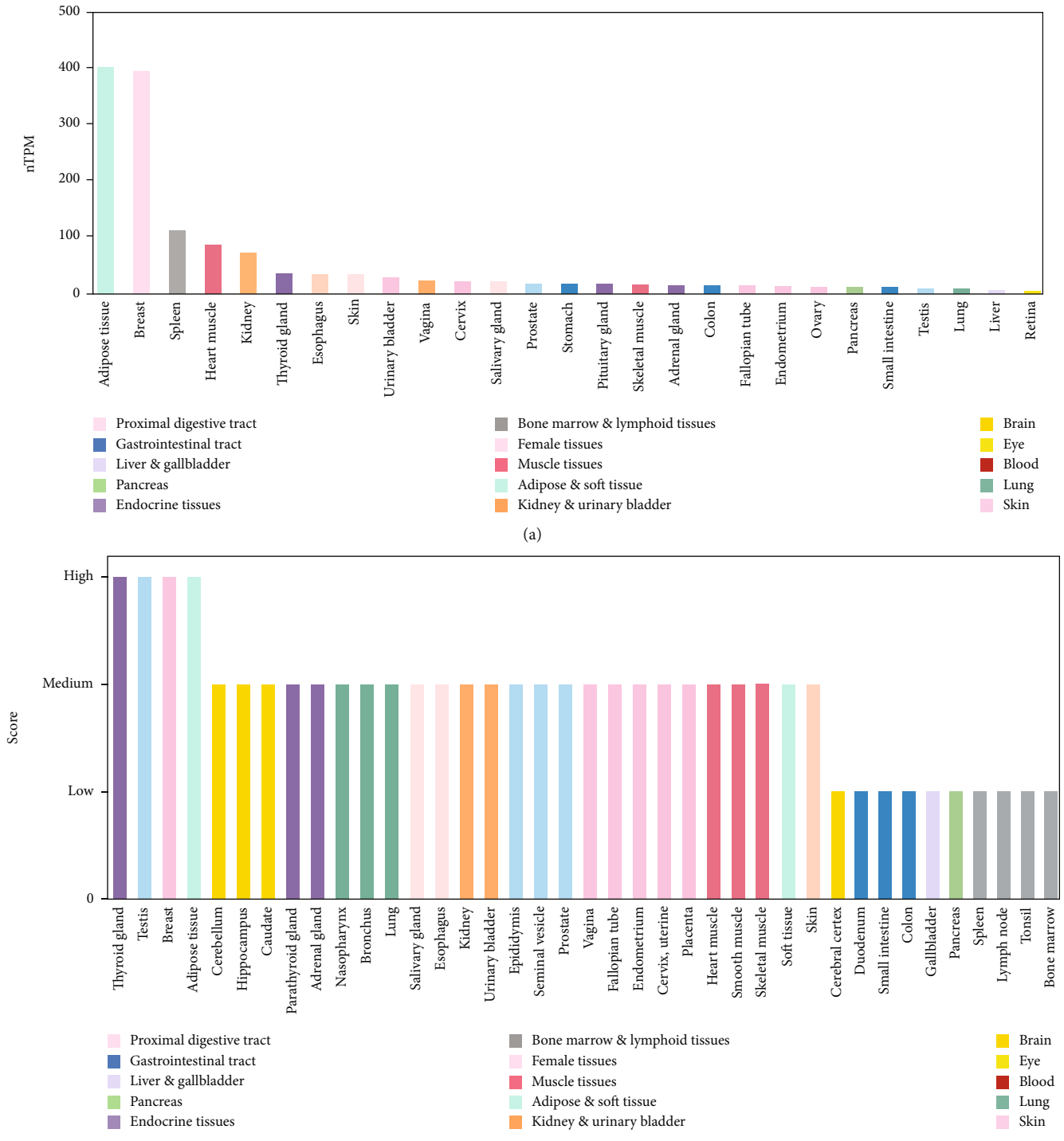
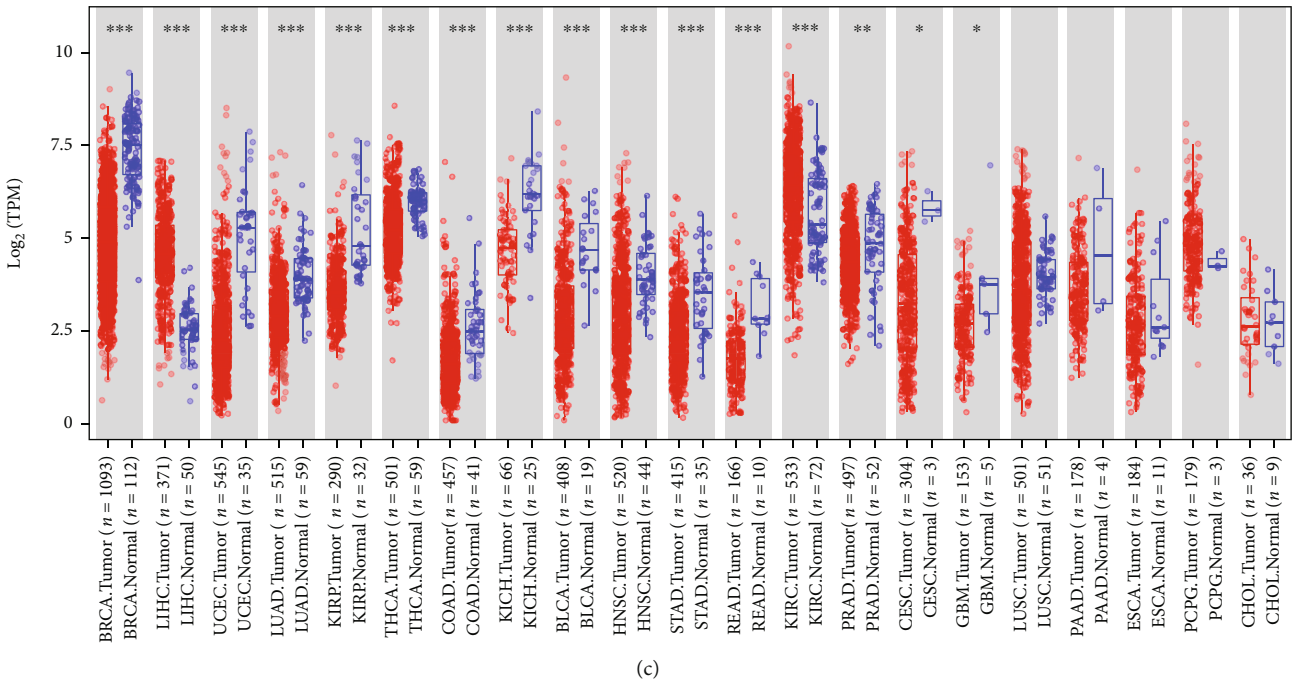



Figure 1: Continued.




(c)

Figure 1: The expression of *RBP7* in different human tissues and pan cancers. (a) *RBP7* expression profiles in normal human tissues from the GTEx project. nTPM, normalized protein-coding transcripts per million. (b) The protein expression of *RBP7* in normal human tissues. (c) The mRNA expression levels of *RBP7* in different cancer types were explored by TIMER. It is presented by ranking the *p* value. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

a high probability of being high-risk markers in breast cancer, of which 15 genes had significantly high HRs (*p value*<0.05). In contrast, there were no genes with high HRs ($p < 0.05$) in the top 50 positively coexpressed genes. These results further confirmed that *RBP7* performs a protective role in the progression of breast cancer (Figure 4(a)).

GO term annotation of biological processes showed that *RBP7* coexpressed genes mainly participate in the adrenergic signalling pathway, excretion, endothelium development, regulation of transporter activity, cell communication by electrical coupling and G protein-coupled receptor signalling pathway, with inhibition of the biological processes including double-strand break repair, cargo loading into vesicle, DNA conformation change, ncRNA transcription, and protein localization to chromosome (Figure 4(b)). KEGG pathway analysis showed that there was an enrichment in the regulation of lipolysis in adipocytes, PPAR signalling pathway, ovarian steroidogenesis, and drug metabolism (Figure 4(c)), indicating a widespread impact of *RBP7* on the global transcriptome.

*3.5. RBP7 Methylation in Breast Cancer.* To further explore the mechanism of the differential expression of *RBP7* in breast cancer and normal tissues, we performed hierarchical clustering analysis of *RBP7* mRNA expression related to DNA methylation by using the UCSC Cancer Genomics Browser. The results showed that *RBP7* methylation mainly occurred in primary breast cancer, and the mRNA expression of RBP7 corresponding to hypermethylated samples is low (Figure 5(a)), indicating a potential correlation between mRNA expression and DNA methylation of *RBP7*.

Next, we performed methylation analysis with the Methsurv database and found that 4 methylation probes, namely, cg20413202, cg03406535, cg10796749, and cg14202757, in the promoter of *RBP7* were highly methylated in breast cancer (Figure 5(b)). As routinely done, the level of methylation was represented by a beta value: a beta value ≥0.6 was considered completely methylated, a beta value ≤0.2 was considered completely unmethylated, and a beta value between 0.2 and 0.6 was considered partially methylated. According to this standard, we found that the beta values of 3 of the 4 detected methylation probes, i.e., cg20413202 (beta value = 0.947), cg10796749 (beta value = 0.798), and cg14202757 (beta value = 0.694), were higher than 0.6, indicating that almost complete methylation occurred on the promoter of *RBP7* (Supplementary Figure 3A-D). Consistently, we found that most of the median beta values of different clinical stages were also above 0.6 (Supplementary Figure 3E-H), suggesting that promoter methylation leads to the inhibition of *RBP7* gene transcription in breast cancer.

Subsequently, we used the SMART App to verify the Spearman correlation between gene expression and DNA methylation of the probes of *RBP7* in breast cancer. The expression of *RBP7* was significantly negatively correlated with the methylation probes cg03406535, cg27083689, cg18086187, cg03994053, cg27561954, cg15090005, and cg07224455. The aggregation plot for all methylation

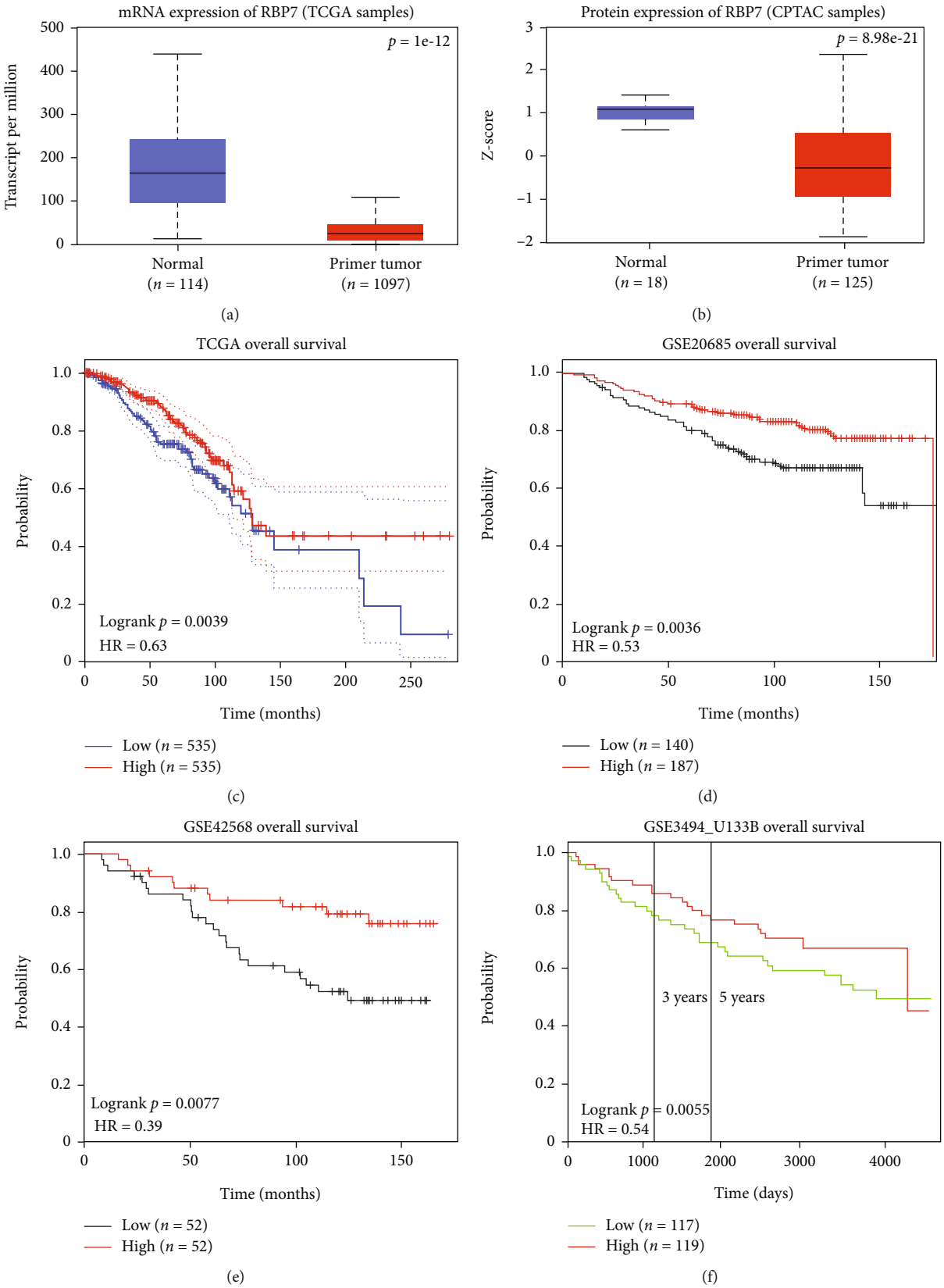



Figure 2: The mRNA and protein expression of *RBP7* in breast cancer and the overall survival analysis with *RBP7* mRNA expression. (a) Boxplot of the mRNA expression of *RBP7* in normal and breast cancer tissues. (b) Boxplot of the protein expression of *RBP7* in normal and breast cancer tissues. (c–f) Kaplan–Meier curve of OS based on the high and low expression of *RBP7* in breast cancer patients from different databases. A log-rank $p$ value <0.05 was considered statistically significant.

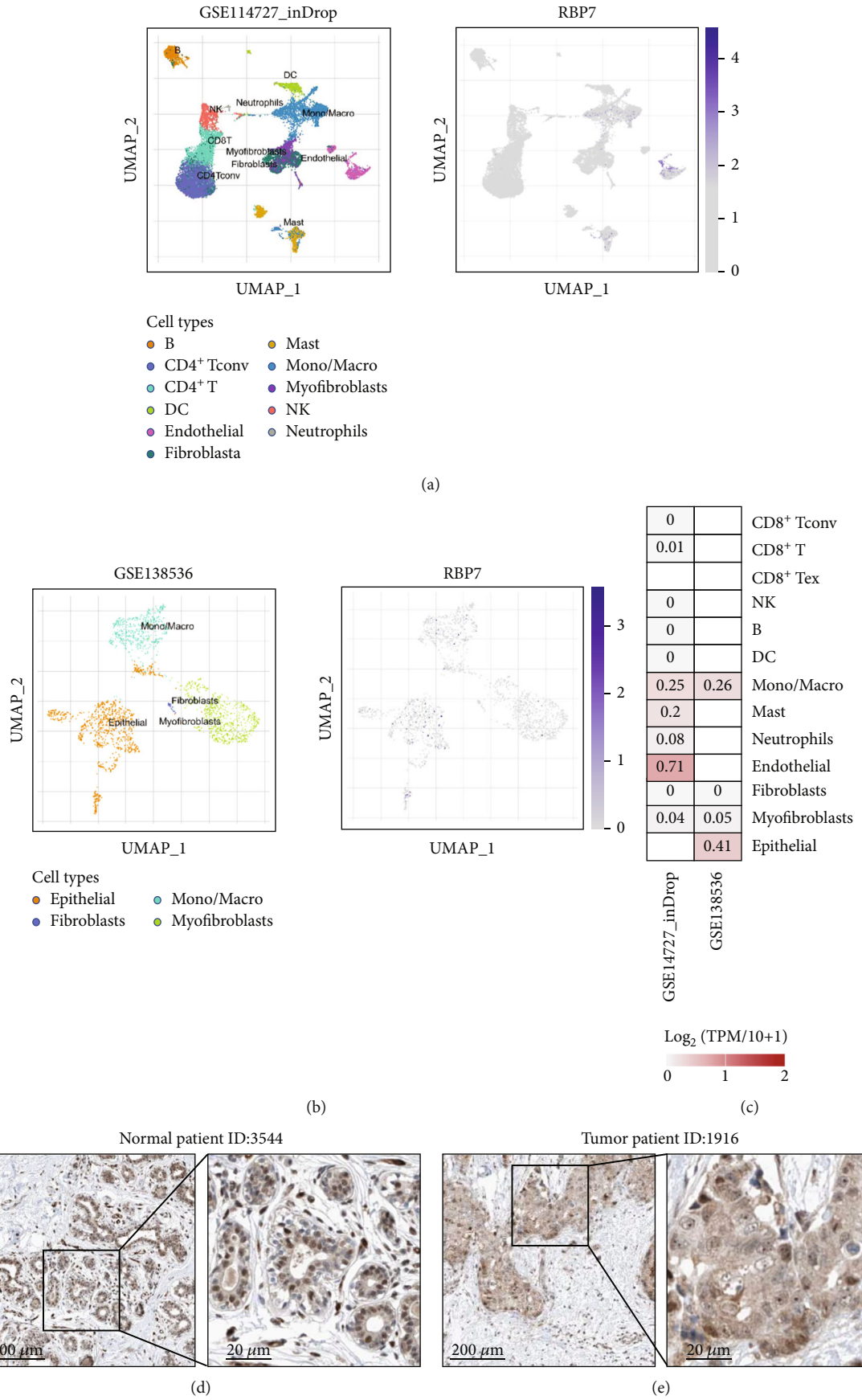



Figure 3: Continued.

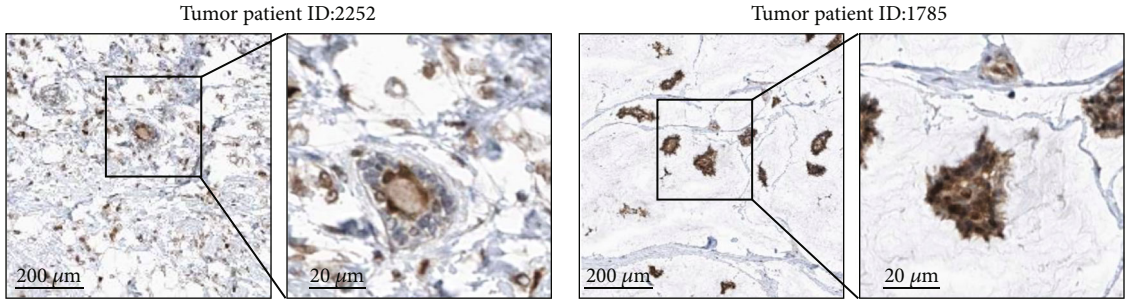



Figure 3: The cellular and subcellular localization of *RBP7* in breast cancer. (a–b) The cellular localization of *RBP7* mRNA in single breast cancer cells. The expression level is colored by marker intensity. UMAP: uniform manifold approximation and projection. (c) Heatmap showing the mRNA expression of *RBP7* in different cell types. (d–g) Representative immunohistochemical staining of normal tissues (https://www.proteinatlas.org/ENSG00000162444-RBP7/tissue/breast) (d) and breast cancer with weak (e), moderate (f), or strong (g) *RBP7* expression (https://www.proteinatlas.org/ENSG00000162444-RBP7/pathology/breast+cancer). Scale bars with lengths of 200 $\mu$m or 20 $\mu$m are displayed in the left and right panels, respectively.

probes showed that the expression level of *RBP7* was negatively correlated with the DNA methylation of *RBP7* (Figure 5(c)).

*3.6. Clinicopathological Association of RBP7 and Its Prognostic Value.* Breast cancer is a complex disease characterized by many morphological, clinical, and molecular features. Molecular subtypes and optimal treatments for breast cancer are usually based on immunohistochemical markers such as estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). We checked the relevance of *RBP7* expression and different clinicopathological features by using the web-based tool bc-GenExMiner. We found that the expression of *RBP7* was highly expressed in ER-positive breast cancer patients compared with ER-negative patients (Figure 6(a)). *RBP7* expression was higher in PR-positive breast cancer patients than in PR-negative patients (Figure 6(b)), whereas *RBP7* expression was higher in HER2-negative breast cancer patients than in HER2-positive patients (Figure 6(c)). To further determine the correlation of *RBP7* expression and hormone receptors (HRs), we utilized DNA microarray data to perform correlation analysis of *RBP7* expression with different combinations of ER and PR expression statuses, i.e., $ER^+/PR^+$, $ER^+/PR^-$, $ER^-/PR^+$, and $ER^-/PR^-$. The results showed that there were remarkable differences in *RBP7* mRNA expression in some hormone receptor combinations, i.e., $ER^+/PR^+$ vs. $ER^-/PR^-$, $p < 0.01$, and $ER^+/PR^+$ vs. $ER^+/PR^-$, $p < 0.01$ (Figure 6(d)).

Furthermore, we performed survival analysis of breast cancer patients with different ER/PR combinations. Down-regulated *RBP7* expression was only significantly associated with poor prognosis in $ER^+/PR^+$ and $ER^+/PR^-$ patients but not in $ER^-/PR^+$ and $ER^-/PR^-$ breast cancer patients (Figures 6(e)–6(l)). OSbrca was used to verify the prognostic value of *RBP7* in these 4 groups with the TCGA database. Consistently, we found that lower *RBP7* expression was significantly correlated with poorer OS in $ER^+/PR^+$ and $ER^+/PR^-$ patients (Supplementary Figure 4).

*3.7. Potential Regulatory Mechanisms and Target Drugs of RBP7 in Breast Cancer.* To further explore the PPI network of *RBP7* in breast cancer, we used STRING to identify genes that may interact with *RBP7*. The interactions between *RBP7* and proteins encoded by the functional genes, including SLC45A1, SLC25A33, UBE4B, TMEM201, NMNAT1, SOCS7, SOCS4, GPR157, PIK3R3, and POLR2G, are shown in Figure 7(a). Interestingly, in this interacting network, PIK3R3, a regulatory subunit of phosphatidylinositol 3-kinase (PI3K), was reported to play an important role in breast cancer. Therefore, we further analyzed the correlation between *RBP7* and PIK3R3 in breast cancer and found that the expression levels of *RBP7* and PIK3R3 were negatively correlated ($R = -0.231$, $p = 6.79e-15$) (Figure 7(b)).

Then, we acquired $ER^+$ breast cancer data from the TCGA database, which were quarterly ranked according to the expression level of *RBP7*. The high and low *RBP7* expression groups were defined as the first and fourth quarters of $ER^+$ breast cancer data, respectively, and the differential genes ($|\log 2(FC)| > 1$) between these two groups were analyzed by using the Limma software package for display as a volcano plot (Figure 7(c)).

KEGG pathway analysis was conducted to investigate the functional implications of the DEGs, and it was found that several tumor-related pathways, such as the PPAR signalling pathway, regulation of lipolysis in adipocytes, and tyrosine metabolism, were significantly enriched (Figure 7(d)). In addition, the PI3K-AKT signalling pathway, which is important in $ER^+$ breast cancers, was also enriched (Figure 7(d)).

Furthermore, we used the CARE database to analyze the association of the molecular alteration of *RBP7* with drug efficacy and found that *RBP7* was negatively correlated with drug efficacy in the CCLE, CGP, and CTRP databases (Figure 7(e)). Interestingly, we found that only one drug (nilotinib) was common among the resistant drugs from these 3 databases (Figure 7(f)). Finally, we visualized the binding site of *RBP7* with nilotinib by SwissDock (Figure 7(g)).

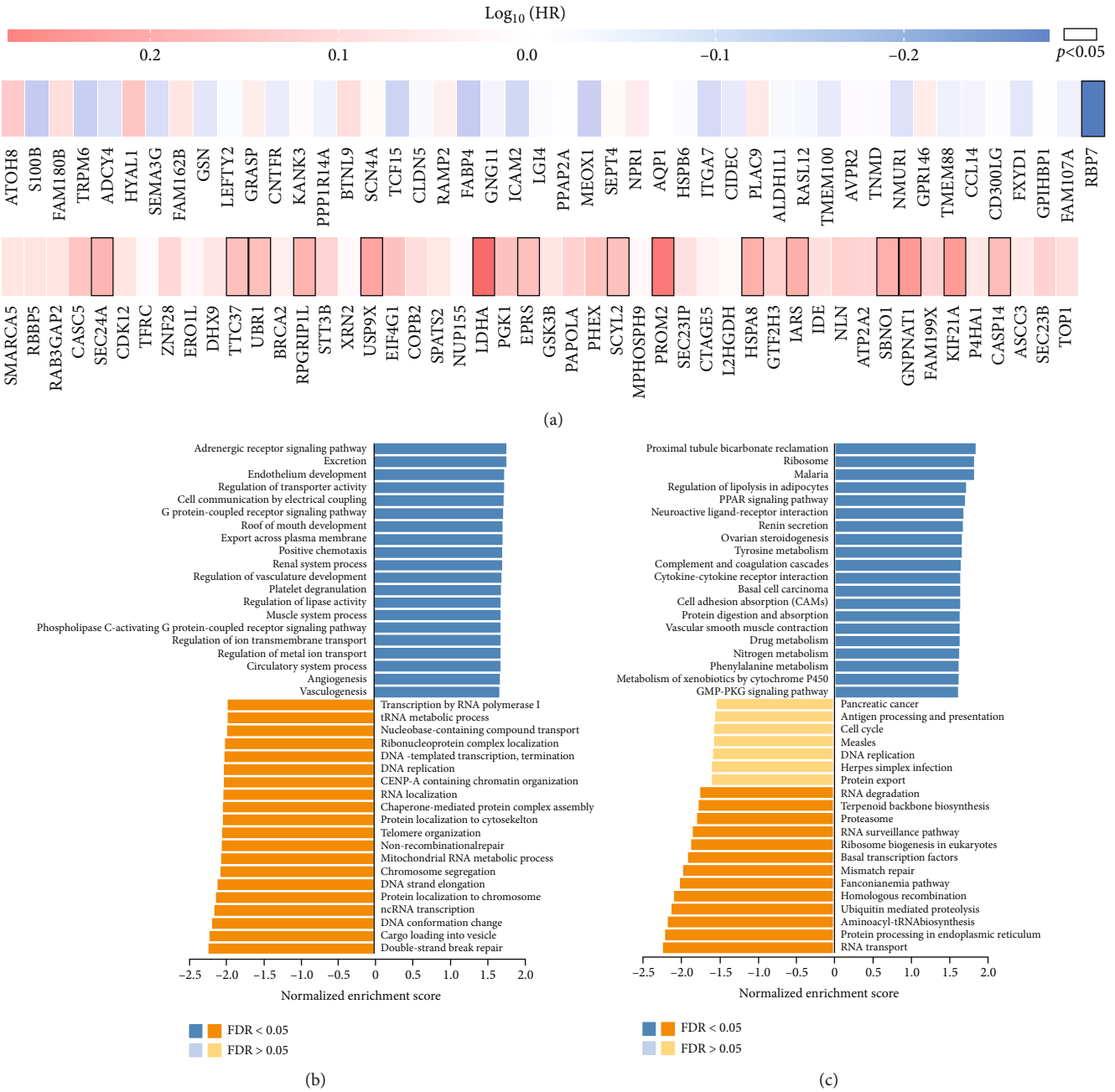



Figure 4: GO and KEGG pathway enrichment analysis of the coexpressed genes of *RBP7* in breast cancer. (a) Survival maps of the top 50 positive or negative coexpressed genes of *RBP7* in breast cancer. (b) GO_BP enrichment analysis of the coexpressed genes of *RBP7* in breast cancer. (c) KEGG pathway analysis of the coexpressed genes of *RBP7* in breast cancer.

## 4. Discussion

*RBP7* is a member of the CRBP family and is involved in retinoic acid-mediated cellular responses [37]. Previous studies demonstrated that the retinol signalling pathway might be relevant to breast cancer progression. However, the prognostic value of *RBP7*, a new member of the CRBP family, in breast cancer is still unclear. In this study, we utilized various databases to explore the expression, prognosis, cellular localization, coexpression network, DNA methylation, and function of *RBP7* in breast cancer.

Gene expression analysis showed that *RBP7* is widely expressed in various normal tissues, including thyroid, testis, breast, and adipose tissues. Notably, in breast cancer tissues, *RBP7* is mainly expressed in epithelial cells with nuclear localization. Importantly, we found that both the mRNA and protein expression levels of *RBP7* in breast cancer tissues were significantly lower than those in normal tissues. Survival analysis with 4 different databases showed that low expression of *RBP7* was significantly associated with poor OS in breast cancer patients. To further illuminate the role of *RBP7* in the progression of breast cancer, we

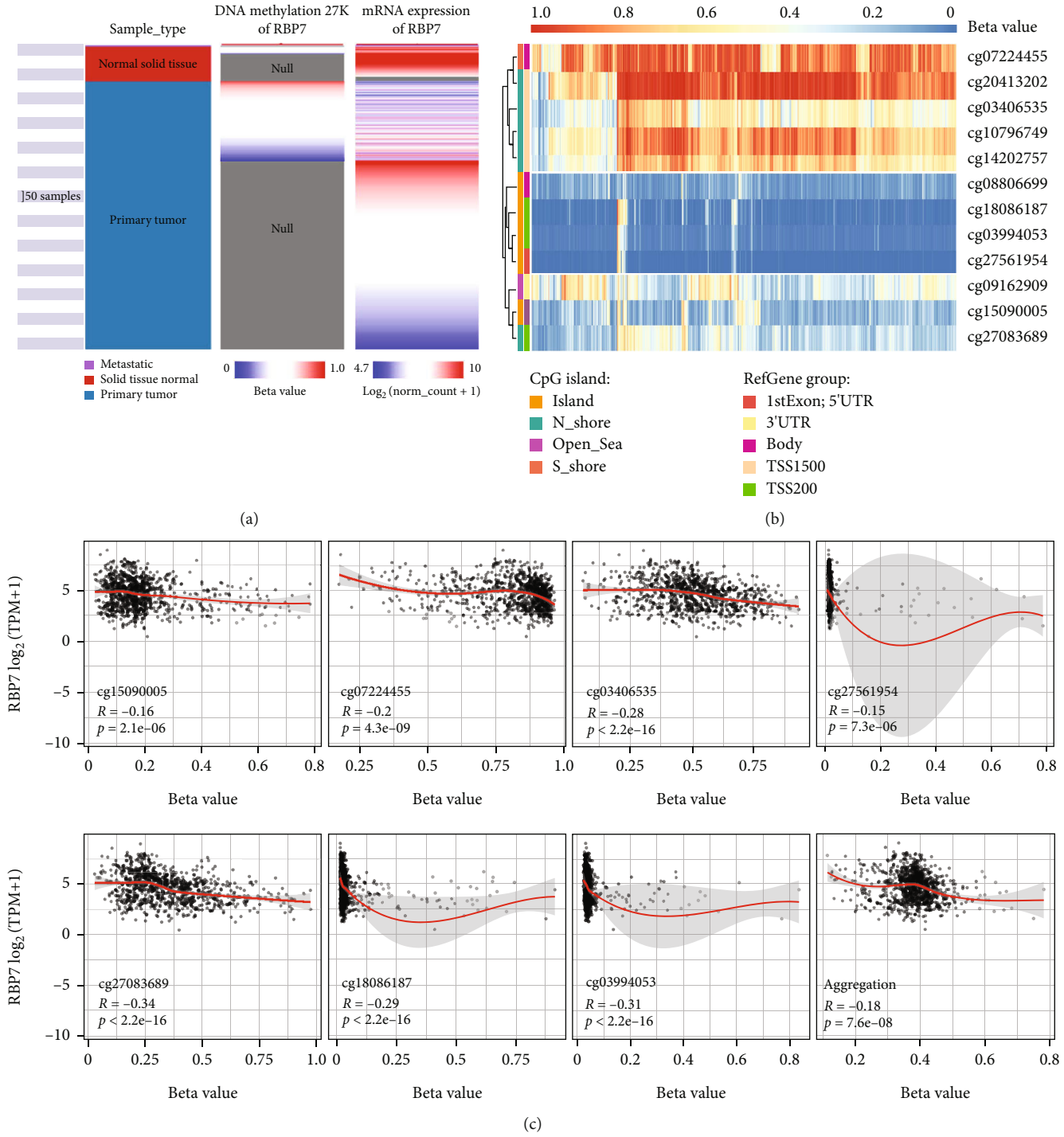



Figure 5: The methylation analysis of *RBP7* in breast cancer. (a) Heat maps for the mRNA expression and DNA methylation of *RBP7* in the TCGA database. (b) Visualization of methylation level and *RBP7* expression in different regions of the *RBP7* gene. (c) Analysis of the correlation between methylation on different methylation probes and mRNA expression of *RBP7* in breast cancer.

analyzed the expression and prognosis of *RBP7* in different molecular subtypes of breast cancer. The results showed that *RBP7* mRNA expression in ER-positive patients was higher than that in ER-negative patients, and higher expression of *RBP7* was associated with better OS and DFS in $ER^+$ breast cancer patients, indicating that *RBP7* was mainly related to the prognosis of patients with $ER^+$ breast cancer.

Our findings were consistent with other studies on the role of *RBP7* in breast cancer. For example, Kinyamu et al. used genome-wide transcriptional profiling to demonstrate that *RBP7* is positively regulated by estradiol (E2) in breast cancer cells [38]; Calvo et al. reported that blockers of estrogen receptors inhibited the expression of estradiol-modulated genes, including *RBP7*, in the mouse mammary

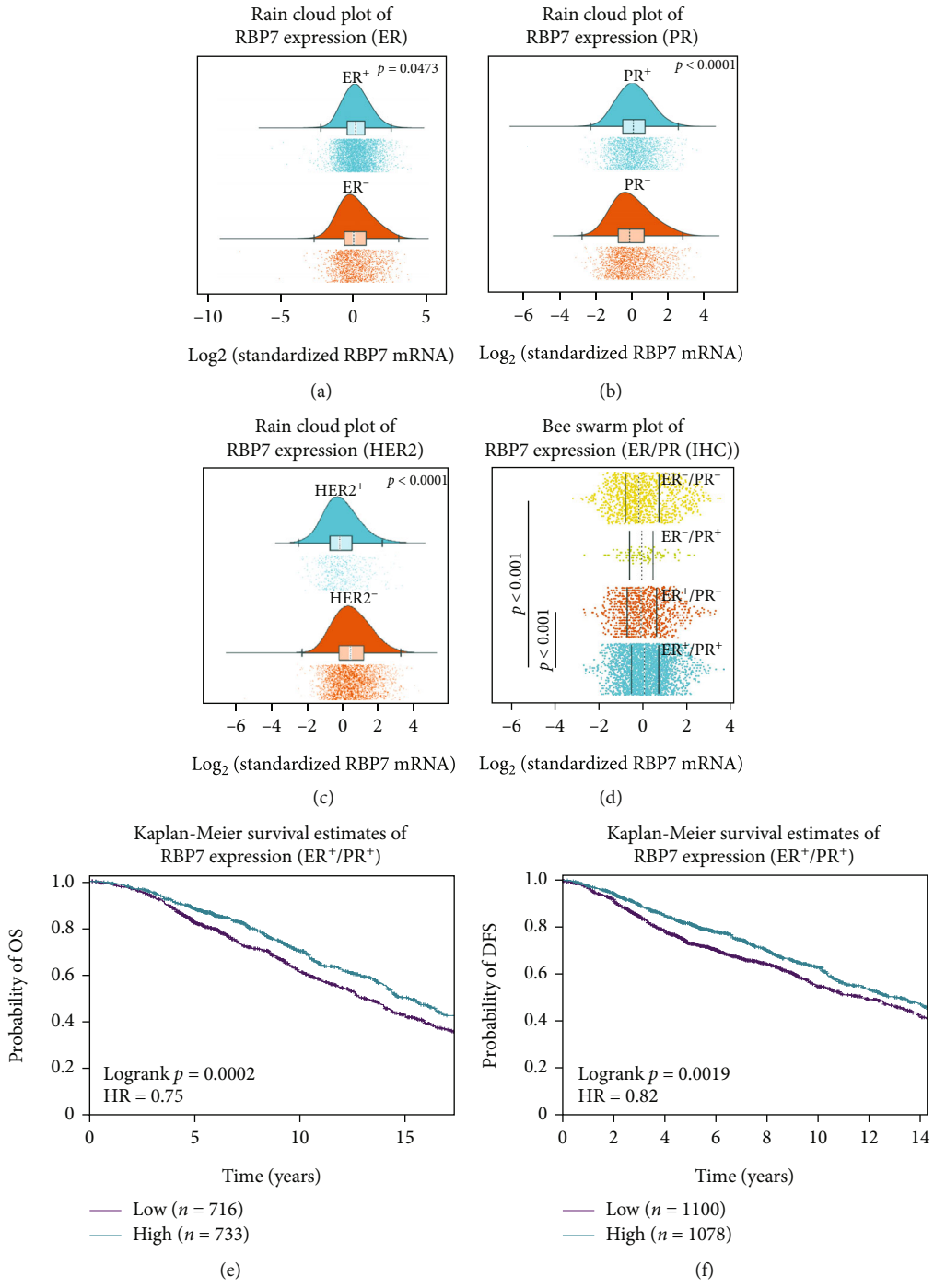



Figure 6: Continued.

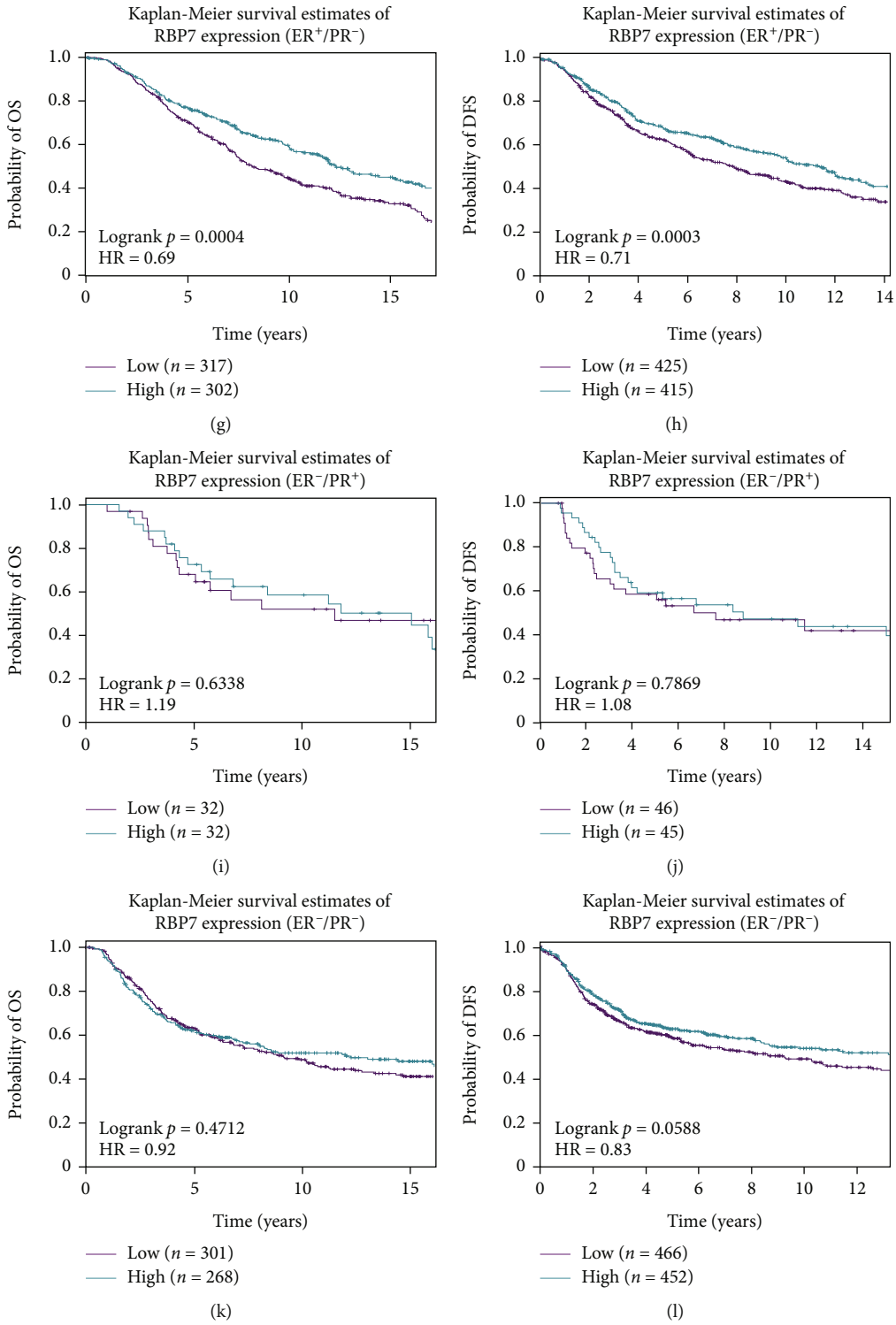



Figure 6: The mRNA expression of *RBP7* and the Kaplan–Meier survival curve in different molecular subtypes of breast cancer. (a–c) The mRNA expression of *RBP7* in different molecular subtypes, including ER (a), PR (b), and HER2 (c), of breast cancer. (d) Bee swarm plot of *RBP7* expression in breast cancer with positive or negative ER or PR expression. (e–l) Kaplan–Meier curves of breast cancer with the DNA microarray results of $ER^+/PR^+$ (e, f), $ER^+/PR^-$ (g, h), $ER^-/PR^+$ (i, j), or $ER^-/PR^-$ (k, l).

gland [39]. It is reasonable to propose that upregulation of *RBP7* by E2 leads to a good prognosis in $ER^+$ breast cancer. Previous studies demonstrated that the growth of human breast tumor cells is regulated by signalling pathways involving nuclear steroid thyroxine receptors [40], especially RARs, which show growth inhibitory activity against breast cancer cells both *in vitro* and *in vivo* [41].Interestingly, RAR and ER share a common coactivator, estradiol. Pemrick

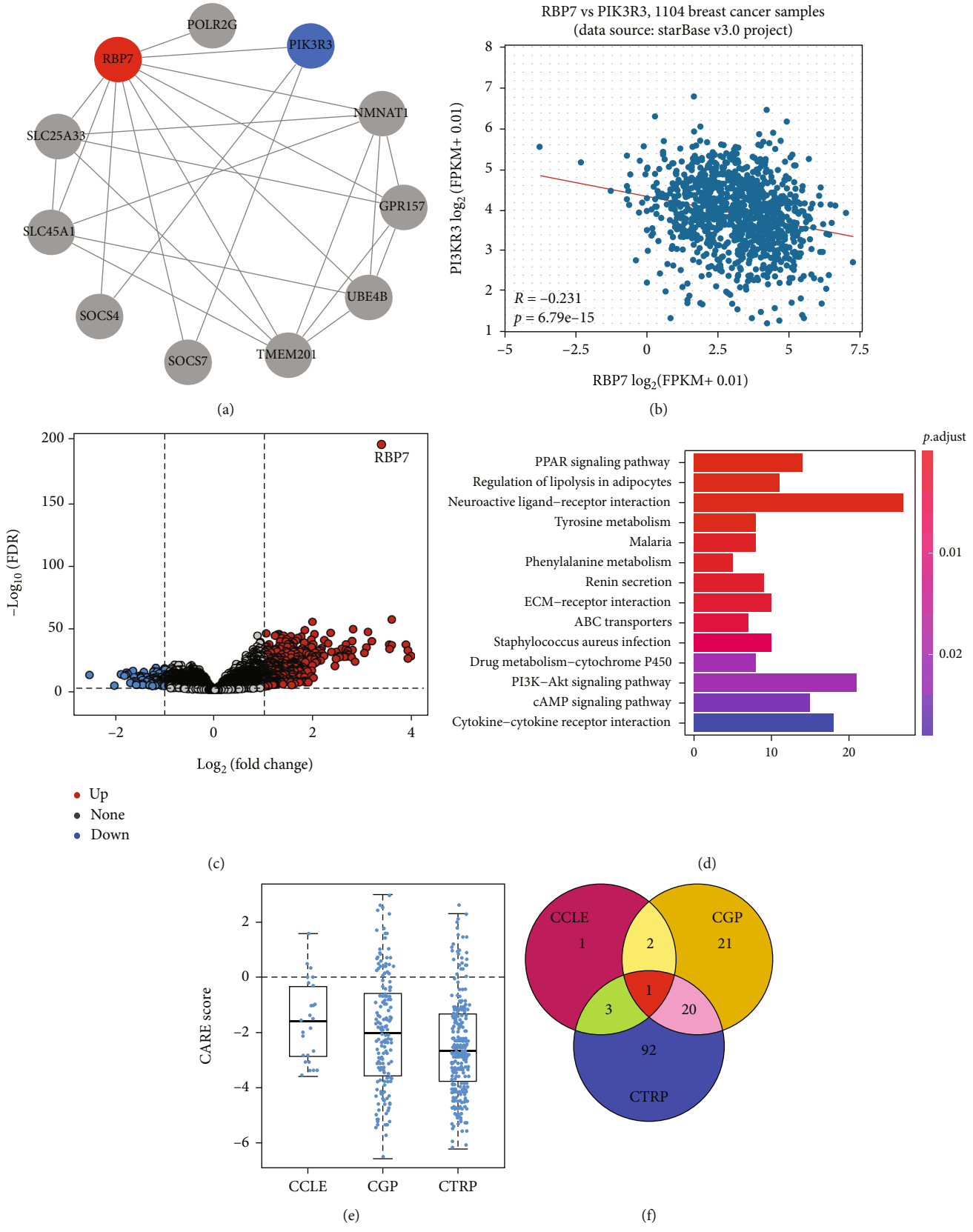
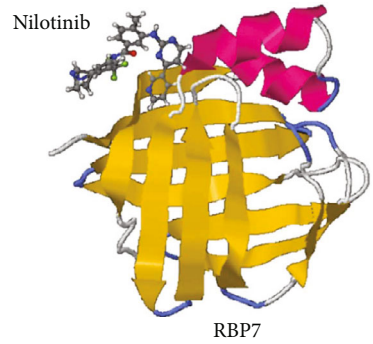



Figure 7: Continued.




(g)

Figure 7: Analysis of the potential regulatory functions and *RBP7*-targeting drugs for breast cancer. (a) The PPI network for *RBP7* and its coexpressed genes by STRING. (b) Correlation analysis of *RBP7* and PIK3R3. (c) Volcano plot showing the differentially expressed genes in $ER^+$ breast cancer with high or low expression of *RBP7*. (d) KEGG pathway analysis of DEGs in $ER^+$ breast cancer with high or low expression of *RBP7*. (e) CARE analysis of the resistance module gene *RBP7* in the CCLE, CTRP and CGP databases. (f) Venn diagram showing the overlap of the *RBP7*-targeting drugs in the 3 databases. (g) Prediction of the binding sites between nilotinib and *RBP7*.

et al. proved that both RAR and ER have high affinity for $\beta$-estradiol by constructing chimeric RARs containing the ligand-binding domain of ER [42]. Furthermore, Fonja et al. found that anti-estrogen and Herceptin induced the expression of PDCD4, revealing that the intracellular crosstalk of RAR, ER, and Her-2 may play a role in growth inhibitory signalling in breast cancer cells [43]. In the retinoid pathway, RAR activation is associated with CRBP1-mediated retinol storage [9] [44]. *RBP7*, another member of the CRBP family, may also engage in crosstalk with RAR; thus, it may play an important role in mechanism regulation via RAR/ER in breast cancer. Based on the results above, we conclude that *RBP7* may be a tumor suppressor gene and that high expression of *RBP7* is associated with a good prognosis for $ER^+$ breast cancer patients.

Our results demonstrated that the gene expression of *RBP7* in breast cancer was significantly lower than that in normal controls. Thus, we further explored the mechanism of *RBP7* downregulation in breast cancer. Bioinformatic analysis showed that there was a significant negative correlation between *RBP7* mRNA expression and promoter methylation. Breast cancer is a highly complex heterogeneous disease that forms different tumor subpopulations with distinct phenotypic characteristics. The different DNA methylation patterns between cell subpopulations drive the phenotypic changes in breast cancer, which is valuable for providing novel insights into intratumor epigenetic heterogeneity [45]. Recent studies have shown that methylation of the promoter as well as intragenic and intergenic regions is involved in the modulation of tumor development and invasion [46]. The methylation of CRBPs was reported to be associated with tumor development. For example, the methylation of CRBPs is common in the esophageal mucosa of patients with esophageal squamous cell carcinoma (ESCC) in the high-risk population and tends to increase in prevalence in foci with worse pathological changes [47]. The methylation profile of CRBPs in bladder cancers is also correlated with the clinicopathological features of poor prognosis [48]. DNA hypermethylation brings about epigenetic silencing of CRBPs in human and mouse breast cancer [49]. For example, CRBP1 gene silencing was found in 60% of G2 and 66.7% of G3 carcinoma cells due to CRBP1 promoter methylation [50]. DNA methylation can occur in the whole genome, including the promoter, gene body, 3′-untranslated region (UTR) and intergenic regions, while the promoter can be further divided into TSS200, TSS1500, 5′-UTR, and the 1st exon. DNA methylation in gene promoters generally has a negative regulatory effect on gene expression, whereas methylation in intragenic regions is not always associated with gene repression [51]. We utilized the MethSurv database to identify the methylation sites of *RBP7* in breast cancer and found 4 probes, namely, cg20413202, cg03406535, cg10796749, and cg14202757, located in TSS1500-N_Shore with high methylation. The analysis using SMART APP web tools showed that promoter methylation was negatively correlated with *RBP7* mRNA expression. In summary, our results demonstrate that the promoter methylation of *RBP7* be related to its transcriptional silencing, which may be a reasonable explanation for the gene downregulation of *RBP7* in breast cancer.

With the concept that low expression of *RBP7* leads to poor prognosis of $ER^+$ breast cancer patients, we further performed KEGG pathway analysis in $ER^+$ breast cancer and found that *RBP7* exerts its biological function through crosstalk with the PPAR and PI3K/AKT signalling pathways. PPARs represent a nuclear receptor superfamily that includes PPAR$\alpha$, PPAR$\beta/\delta$, and PPAR$\gamma$. It was reported that the activation of PPAR$\gamma$ inhibits the cell growth of different cancers, such as colon cancer, gastric cancer, and liposarcoma [52]. For example, after cleavage by caspase-1 at Asp64, PPAR$\gamma$ translocates to mitochondria, leading to attenuation of medium-chain acyl-CoA dehydrogenase (MCAD) activity and inhibition of fatty acid oxidation, which brings about the accumulation of lipid droplets and differentiation of tumor-associated macrophages (TAMs), thus resulting in an ultimate suppression of tumor growth [53, 54]. Another study by Mueller et al. reported that PPAR$\gamma$ is highly expressed in human primary and metastatic

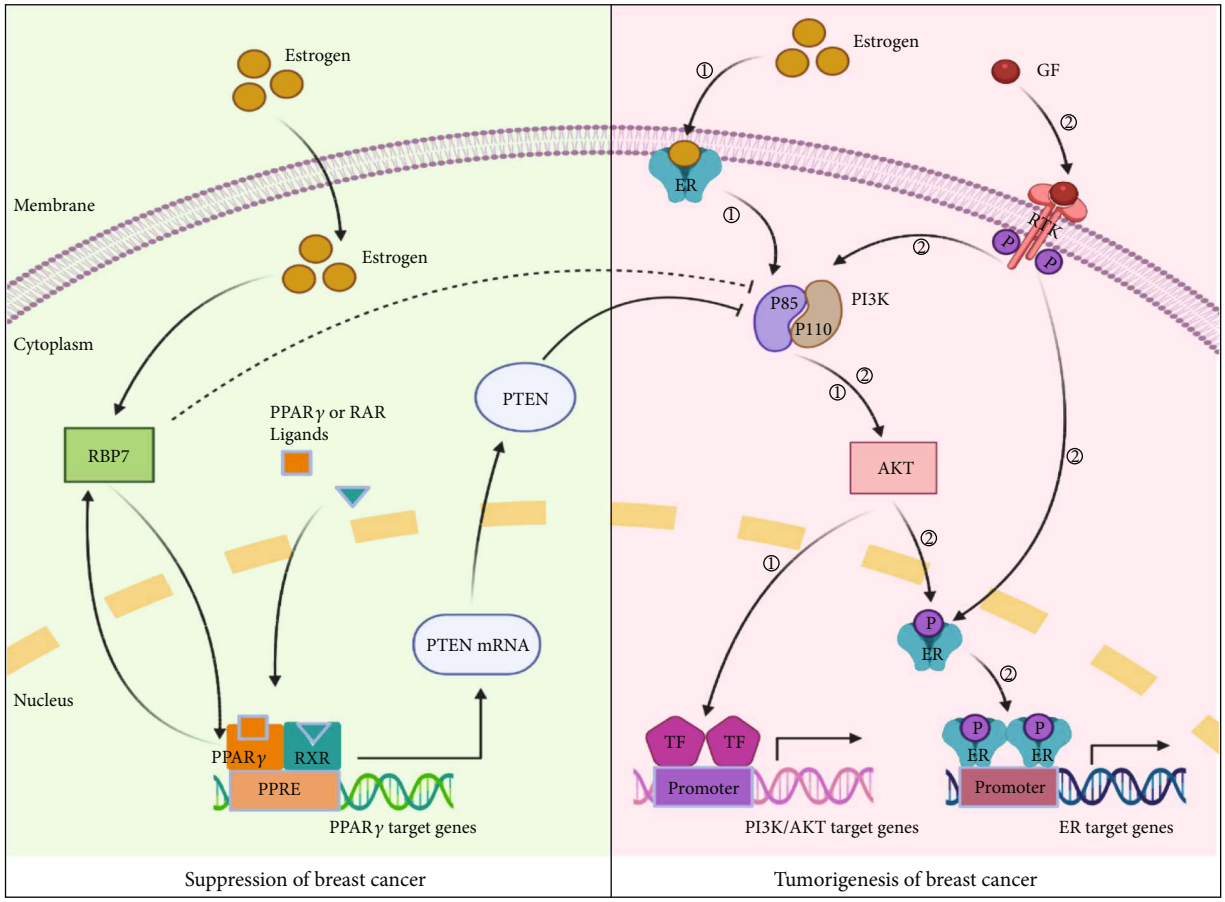



FIGURE 8: The hypothetical function of *RBP7* in the regulation of the PPAR and PI3K/AKT pathways in $ER^+$ breast cancer.

breast cancer, and ligand activation of this receptor in breast cancer cells causes extensive lipid accumulation, which results in a reduction in the growth rate and clonogenic capacity of tumor cells [55] [56]. Intriguingly, application of the PPARγ agonist rosiglitazone in combination with the MEK inhibitor trametinib can terminally differentiate breast cancer cells that have undergone epithelial-mesenchymal transition (EMT) into adipocytes [57]. As a PPARγ target gene, *RBP7* is also an upstream regulator of some other PPARγ target genes in the endothelium. PPARγ and *RBP7* control the oxidative state of blood vessels by forming a transcriptional regulatory circuit (or hub) in the endothelium. Loss of *RBP7* impairs this regulatory circuit, resulting in oxidative stress and dysfunction in endothelial cells [58].Cancer cells have to endure oxidative stress throughout tumorigenesis, including during initiation, matrix detachment, transmission in the circulation, and relapse after therapy [59]. In addition, endothelial injury is closely related to tumorigenesis and accompanies malignant cancer cells in almost every stage of the metastatic process [60]. Thus, impairment of the regulatory circuit between *RBP7* and PPARγ increases the opportunity to promote the occurrence of cancer.

PI3K/AKT is the most frequently activated signalling pathway that promotes tumor growth [61] and progression of breast cancer [62]. Bonofiglio *et al.* revealed that the ERα and PPARγ pathways have an opposite effect on the regulation of the PI3K/AKT signal transduction cascade [63]. The nuclear receptor ERα has been shown to be involved in the pathophysiological process of breast cancer. Membrane-anchored ERα can activate various cytoplasmic kinases, including components of the PI3K/AKT pathway, through rapid nongenomic actions [64]. There are two signalling pathways involved in the activation of the PI3K/AKT pathway in $ER^+$ breast cancer cells. The estrogen-dependent pathway activates the PI3K/AKT signalling process by directly binding to the p85α regulatory subunit of PI3K, thus enhancing the transcriptional activity of targeted genes in breast cancer cells [65]. For the estrogen-independent pathway, the interaction of EGFR with growth factor can directly induce ERα transcriptional activity through the PI3K/AKT signalling pathway [66]. Consequently, the downstream signalling of both pathways is activated, leading to the proliferation and survival of tumor cells. In contrast, PPARγ can inhibit the PI3K/AKT pathway by upregulating PTEN transcription in breast cancer cells [63]. As an antagonist of the PI3K/AKT pathway, PTEN plays a key role in preventing tumorigenesis [67]. Based on our results and previous studies, we hypothesize that *RBP7* may regulate PTEN by targeting PPARγ, thereby suppressing the activation of the PI3K/AKT pathway. The PPI network from

bioinformatics analysis revealed that *RBP7* may directly interact with PIK3R3, resulting in activation of the PI3K/AKT pathway (Figure 8).

## 5. Conclusions

In conclusion, this study provides the first evidence that *RBP7* downregulation in breast cancer is associated with promoter methylation. Furthermore, we found that *RBP7* has prognostic value for $ER^+$ breast cancer. Deep bioinformatics analysis of *RBP7*-related pathways reveals some vital information for the regulatory mechanism in $ER^+$ breast cancer. However, these results need to be further validated by both in vitro and in vivo experiments. This study is helpful in providing novel approaches for clinical diagnosis and treatment.

## Data Availability

The original contributions presented in the study are included in the article/Supplementary material, and further inquiries can be directed to the corresponding authors.

## Ethical Approval

All data of this study were public and required no ethical approval by an institutional review board or ethics committee.

## Conflicts of Interest

No conflicts of interest, financial or otherwise, are declared by the authors.

## Authors' Contributions

HL, QH, YH, and YJ conceived and designed research. HL and QH analyzed data. HL, JW, and ZZ interpreted results. HL, QH, and JW prepared figures. HL drafted manuscript. HHL, YH, and YJ edited and revised manuscript. HL, JW, QH, ZZ, HHL, YH, and YJ approved the final version of the manuscript.

## Acknowledgments

This work has benefited from TCGA and GEO. We thank the TCGA and GEO network for its generous sharing large amounts of data. The manuscript has been presented as a preprint according to the following link "https://www.researchsquare.com/article/rs-1135383/v1" [68]. This study was supported by grants from the National Natural Science Foundation of China (No. 81971895), the Special Support Plan for Outstanding Talents of Guangdong Province (No. 2019JC05Y340), and the Guangdong Provincial Science and Technology Projects (No. 2016A020216015).

## Supplementary Materials

Supplementary Figure 1. Analysis of the mRNA expression of RBP7 in breast cancer. Supplementary Figure 2. The genes correlated with the expression of RBP7. Supplementary Figure 3. Density maps of CpG island methylation and the methylation distribution in different clinical stages of breast cancer. Supplementary Figure 4. Effect of RBP7 expression on the overall survival curves of breast cancer patients of different molecular subtypes. *(Supplementary Materials)*

## References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA: a Cancer Journal for Clinicians*, vol. 70, no. 1, pp. 7–30, 2020.

[2] R. Siersbaek, S. Kumar, and J. S. Carroll, "Signaling pathways and steroid receptors modulating estrogen receptor $\alpha$ function in breast cancer," *Genes & Development*, vol. 32, no. 17-18, pp. 1141–1154, 2018.

[3] A. B. Hanker, D. R. Sudhan, and C. L. Arteaga, "Overcoming endocrine resistance in breast cancer," *Cancer Cell*, vol. 37, no. 4, pp. 496–513, 2020.

[4] J. L. Napoli, "Cellular retinoid binding-proteins, CRBP, CRABP, FABP5: effects on retinoid metabolism, function and related diseases," *Pharmacology & Therapeutics*, vol. 173, pp. 19–33, 2017.

[5] M. Esteller, M. Guo, V. Moreno et al., "Hypermethylation-associated inactivation of the cellular retinol-binding-protein 1 gene in human cancer," *Cancer Research*, vol. 62, no. 20, pp. 5902–5905, 2002.

[6] Y. S. Kuppumbatti, I. J. Bleiweiss, J. P. Mandeli, S. Waxman, and R. Mira-y-Lopez, "Cellular retinol-binding protein expression and breast cancer," *Journal of the National Cancer Institute*, vol. 92, no. 6, pp. 475–480, 2000.

[7] Y. S. Kuppumbatti, B. Rexer, S. Nakajo, K. Nakaya, and R. Mira-y-Lopez, "CRBP suppresses breast cancer cell survival and anchorage-independent growth," *Oncogene*, vol. 20, no. 50, pp. 7413–7419, 2001.

[8] E. F. Farias, C. Marzan, and R. Mira-y-Lopez, "Cellular retinol-binding protein-I inhibits PI3K/Akt signaling through a retinoic acid receptor-dependent mechanism that regulates p85-p110 heterodimerization," *Oncogene*, vol. 24, no. 9, pp. 1598–1606, 2005.

[9] E. F. Farias, D. E. Ong, N. B. Ghyselinck, S. Nakajo, Y. S. Kuppumbatti, and R. Mira y Lopez, "Cellular retinol-binding protein I, a regulator of breast epithelial retinoic acid receptor activity, cell differentiation, and tumorigenicity," *Journal of the National Cancer Institute*, vol. 97, no. 1, pp. 21–29, 2005.

[10] C. Folli, V. Calderone, I. Ramazzina, G. Zanotti, and R. Berni, "Ligand binding and structural analysis of a human putative cellular retinol-binding protein," *The Journal of Biological Chemistry*, vol. 277, no. 44, pp. 41970–41977, 2002.

[11] S. Fang and C. D. Sigmund, "PPAR$\gamma$ and RhoBTB1 in hypertension," *Current Opinion in Nephrology and Hypertension*, vol. 29, no. 2, pp. 161–170, 2020.

[12] J. Ahn, D. H. Kim, Y. Suh, J. W. Lee, and K. Lee, "Adipose-specific expression of mouse _Rbp7_ gene and its developmental and metabolic changes," *Gene*, vol. 670, pp. 38–45, 2018.

[13] M. Elmasry, L. Brandl, J. Engel, A. Jung, T. Kirchner, and D. Horst, "RBP7 is a clinically prognostic biomarker and

linked to tumor invasion and EMT in colon cancer," *Journal of Cancer*, vol. 10, no. 20, pp. 4883–4891, 2019.

[14] M. Uhlen, L. Fagerberg, B. M. Hallstrom et al., "Tissue-based map of the human proteome," *Science*, vol. 347, no. 6220, 2015.

[15] M. Uhlen, C. Zhang, S. Lee et al., "A pathology atlas of the human cancer transcriptome," *Science*, vol. 357, no. 6352, 2017.

[16] T. Li, J. Fan, B. Wang et al., "TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells," *Cancer Research*, vol. 77, no. 21, pp. e108–e110, 2017.

[17] D. R. Rhodes, J. Yu, K. Shanker et al., "ONCOMINE: a cancer microarray database and integrated data-mining platform," *Neoplasia*, vol. 6, no. 1, pp. 1–6, 2004.

[18] D. S. Chandrashekar, B. Bashel, S. A. H. Balasubramanya et al., "UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses," *Neoplasia*, vol. 19, no. 8, pp. 649–658, 2017.

[19] R. Edgar, M. Domrachev, and A. E. Lash, "Gene expression omnibus: NCBI gene expression and hybridization array data repository," *Nucleic Acids Research*, vol. 30, no. 1, pp. 207–210, 2002.

[20] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, and Z. Zhang, "GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses," *Nucleic Acids Research*, vol. 45, no. W1, pp. W98–W102, 2017.

[21] B. Gyorffy, A. Lanczky, A. C. Eklund et al., "An online survival analysis tool to rapidly assess the effect of 22, 277 genes on breast cancer prognosis using microarray data of 1, 809 patients," *Breast Cancer Research and Treatment*, vol. 123, no. 3, pp. 725–731, 2010.

[22] C. P. Goswami and H. Nakshatri, "PROGgeneV2: enhancements on the existing database," *BMC Cancer*, vol. 14, no. 1, p. 970, 2014.

[23] D. Sun, J. Wang, Y. Han et al., "TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment," *Nucleic Acids Research*, vol. 49, no. D1, pp. D1420–D1430, 2021.

[24] S. V. Vasaikar, P. Straub, J. Wang, and B. Zhang, "LinkedOmics: analyzing multi-omics data within and across 32 cancer types," *Nucleic Acids Research*, vol. 46, no. D1, pp. D956–D963, 2018.

[25] A. S. Zweig, D. Karolchik, R. M. Kuhn, D. Haussler, and W. J. Kent, "UCSC genome browser tutorial," *Genomics*, vol. 92, no. 2, pp. 75–84, 2008.

[26] J. Casper, A. S. Zweig, C. Villarreal et al., "The UCSC genome browser database: 2018 update," *Nucleic Acids Research*, vol. 46, no. D1, pp. D762–D769, 2018.

[27] V. Modhukur, T. Iljasenko, T. Metsalu, K. Lokk, T. Laisk-Podar, and J. Vilo, "MethSurv: a web tool to perform multivariable survival analysis using DNA methylation data," *Epigenomics*, vol. 10, no. 3, pp. 277–288, 2018.

[28] Y. Li, D. Ge, and C. Lu, "The SMART app: an interactive web application for comprehensive DNA methylation analysis and visualization," *Epigenetics & Chromatin*, vol. 12, no. 1, p. 71, 2019.

[29] W. Xu, M. Xu, L. Wang et al., "Integrative analysis of DNA methylation and gene expression identified cervical cancer-specific diagnostic biomarkers," *Signal Transduction and Targeted Therapy*, vol. 4, no. 1, p. 55, 2019.

[30] P. Jezequel, W. Gouraud, F. Ben Azzouz et al., "bc-GenExMiner 4.5: new mining module computes breast cancer differential gene expression analyses," *Database: The Journal of Biological Databases and Curation*, vol. 2021, article baab007, 2021.

[31] Z. Yan, Q. Wang, X. Sun et al., "OSbrca: a web server for breast cancer prognostic biomarker investigation with massive data from tens of cohorts," *Frontiers in Oncology*, vol. 9, p. 1349, 2019.

[32] D. Szklarczyk, A. Franceschini, M. Kuhn et al., "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored," *Nucleic Acids Research*, vol. 39, no. Database, pp. D561–D568, 2011.

[33] J. H. Li, S. Liu, H. Zhou, L. H. Qu, and J. H. Yang, "starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data," *Nucleic Acids Research*, vol. 42, no. D1, pp. D92–D97, 2014.

[34] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe, "KEGG for integration and interpretation of large-scale molecular data sets," *Nucleic Acids Research*, vol. 40, no. D1, pp. D109–D114, 2012.

[35] P. Jiang, W. Lee, X. Li et al., "Genome-scale signatures of gene interaction from compound screens predict clinical efficacy of targeted cancer therapies," *Cell Systems*, vol. 6, no. 3, pp. 343–354.e5, 2018.

[36] A. Grosdidier, V. Zoete, and O. Michielin, "Swiss Dock, a protein-small molecule docking web service based on EADock DSS," *Nucleic Acids Research*, vol. 39, pp. W270–W277, 2011.

[37] M. Choder, "Rpb4 and Rpb7: subunits of RNA polymerase II and beyond," *Trends in Biochemical Sciences*, vol. 29, no. 12, pp. 674–681, 2004.

[38] H. K. Kinyamu, J. B. Collins, S. F. Grissom, P. B. Hebbar, and T. K. Archer, "Genome wide transcriptional profiling in breast cancer cells reveals distinct changes in hormone receptor target genes and chromatin modifying enzymes after proteasome inhibition," *Molecular Carcinogenesis*, vol. 47, no. 11, pp. 845–885, 2008.

[39] E. Calvo, V. Luu-The, P. Belleau, C. Martel, and F. Labrie, "Specific transcriptional response of four blockers of estrogen receptors on estradiol-modulated genes in the mouse mammary gland," *Breast Cancer Research and Treatment*, vol. 134, no. 2, pp. 625–647, 2012.

[40] S. M. Schneider, M. Offterdinger, H. Huber, and T. W. Grunt, "Involvement of nuclear steroid/thyroid/retinoid receptors and of protein kinases in the regulation of growth and of c-erbB and retinoic acid receptor expression in MCF-7 breast cancer cells," *Breast Cancer Research and Treatment*, vol. 58, no. 2, pp. 171–181, 1999.

[41] X. Peng, T. Maruo, Y. Cao et al., "A novel RARbeta isoform directed by a distinct promoter P 3 and mediated by retinoic acid in breast cancer cells," *Cancer Research*, vol. 64, no. 24, pp. 8911–8918, 2004.

[42] S. M. Pemrick, P. Abarzua, C. Kratzeisen et al., "Characterization of the chimeric retinoic acid receptor RARα/VDR," *Leukemia*, vol. 12, no. 4, pp. 554–562, 1998.

[43] O. Afonja, D. Juste, S. Das, S. Matsuhashi, and H. H. Samuels, "Induction of PDCD4 tumor suppressor gene expression by RAR agonists, antiestrogen and HER-2/neu antagonist in breast cancer cells. Evidence for a role in apoptosis," *Oncogene*, vol. 23, no. 49, pp. 8135–8145, 2004.

[44] N. Bushue and Y. J. Wan, "Retinoid pathway and cancer therapeutics," *Advanced Drug Delivery Reviews*, vol. 62, no. 13, pp. 1285–1298, 2010.

[45] V. Almendro and G. Fuster, "Heterogeneity of breast cancer: etiology and clinical relevance," *Clinical & Translational Oncology*, vol. 13, no. 11, pp. 767–773, 2011.

[46] G. H. Rauscher, J. K. Kresovich, M. Poulin et al., "Exploring DNA methylation changes in promoter, intragenic, and intergenic regions as early and late events in breast cancer formation," *BMC Cancer*, vol. 15, no. 1, p. 816, 2015.

[47] M. J. Roth, C. C. Abnet, N. Hu et al., "P16, MGMT, RARbeta2, CLDN3, CRBP and MT1G gene methylation in esophageal squamous cell carcinoma and its precursor lesions," *Oncology Reports*, vol. 15, no. 6, pp. 1591–1597, 2006.

[48] M. Brait, S. Begum, A. L. Carvalho et al., "Aberrant promoter methylation of multiple genes during pathogenesis of bladder cancer," *Cancer Epidemiology, Biomarkers & Prevention*, vol. 17, no. 10, pp. 2786–2794, 2008.

[49] A. Arapshian, S. Bertran, Y. S. Kuppumbatti, S. Nakajo, and R. Mira-y-Lopez, "Epigenetic CRBP downregulation appears to be an evolutionarily conserved (human and mouse) and oncogene-specific phenomenon in breast cancer," *Molecular Cancer*, vol. 3, no. 1, p. 13, 2004.

[50] E. Doldo, G. Costanza, A. Ferlosio et al., "CRBP-1 expression in ovarian cancer: a potential therapeutic target," *Anticancer Research*, vol. 34, no. 7, pp. 3303–3312, 2014.

[51] N. Shenker and J. M. Flanagan, "Intragenic DNA methylation: implications of this epigenetic mechanism for cancer research," *British Journal of Cancer*, vol. 106, no. 2, pp. 248–253, 2012.

[52] A. Z. Mirza, I. I. Althagafi, and H. Shamshad, "Role of PPAR receptor in different diseases and their ligands: physiological importance and clinical implications," *European Journal of Medicinal Chemistry*, vol. 166, pp. 502–513, 2019.

[53] Z. Niu, Q. Shi, W. Zhang et al., "Caspase-1 cleaves PPAR$\gamma$ for potentiating the pro-tumor action of TAMs," *Nature Communications*, vol. 8, no. 1, p. 766, 2017.

[54] Y. Fu, T. Zou, X. Shen et al., "Lipid metabolism in cancer progression and therapeutic strategies," *MedComm*, vol. 2, no. 1, pp. 27–59, 2021.

[55] E. Mueller, P. Sarraf, P. Tontonoz et al., "Terminal differentiation of human breast cancer through PPAR$\gamma$," *Molecular Cell*, vol. 1, no. 3, pp. 465–470, 1998.

[56] A. J. Apostoli, J. M. Roche, M. M. Schneider et al., "Opposing roles for mammary epithelial-specific PPAR$\gamma$ signaling and activation during breast tumour progression," *Molecular Cancer*, vol. 14, no. 1, p. 85, 2015.

[57] D. Ishay-Ronen, M. Diepenbruck, R. K. R. Kalathur et al., "Gain fat–lose metastasis: converting invasive breast cancer cells into adipocytes inhibits cancer metastasis," *Cancer Cell*, vol. 35, no. 1, pp. 17–32.e6, 2019.

[58] C. Hu, H. L. Keen, K. T. Lu et al., "Retinol-binding protein 7 is an endothelium-specific PPAR$\gamma$ cofactor mediating an antioxidant response through adiponectin," *JCI Insight*, vol. 2, no. 6, article e91738, 2017.

[59] J. D. Hayes, A. T. Dinkova-Kostova, and K. D. Tew, "Oxidative stress in cancer," *Cancer Cell*, vol. 38, no. 2, pp. 167–197, 2020.

[60] A. Blazejczyk, D. Papiernik, K. Porshneva, J. Sadowska, and J. Wietrzyk, "Endothelium and cancer metastasis: perspectives for antimetastatic therapy," *Pharmacological Reports*, vol. 67, no. 4, pp. 711–718, 2015.

[61] T. W. Miller, B. N. Rexer, J. T. Garrett, and C. L. Arteaga, "Mutations in the phosphatidylinositol 3-kinase pathway: role in tumor progression and therapeutic implications in breast cancer," *Breast Cancer Research*, vol. 13, no. 6, p. 224, 2011.

[62] L. H. Saal, P. Johansson, K. Holm et al., "Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 18, pp. 7564–7569, 2007.

[63] D. Bonofiglio, S. Gabriele, S. Aquila et al., "Estrogen receptor alpha binds to peroxisome proliferator-activated receptor response element and negatively interferes with peroxisome proliferator-activated receptor gamma signaling in breast cancer cells," *Clinical Cancer Research*, vol. 11, no. 17, pp. 6139–6147, 2005.

[64] A. S. Khatpe, A. K. Adebayo, C. A. Herodotou, B. Kumar, and H. Nakshatri, "Nexus between PI3K/AKT and estrogen receptor signaling in breast cancer," *Cancers (Basel)*, vol. 13, no. 3, p. 369, 2021.

[65] T. Simoncini, A. Hafezi-Moghadam, D. P. Brazil, K. Ley, W. W. Chin, and J. K. Liao, "Interaction of oestrogen receptor with the regulatory subunit of phosphatidylinositol-3-OH kinase," *Nature*, vol. 407, no. 6803, pp. 538–541, 2000.

[66] T. W. Miller, J. M. Balko, and C. L. Arteaga, "Phosphatidylinositol 3-kinase and antiestrogen resistance in breast cancer," *Journal of Clinical Oncology*, vol. 29, no. 33, pp. 4452–4461, 2011.

[67] J. Huang, J. Yan, J. Zhang et al., "SUMO1 modification of PTEN regulates tumorigenesis by controlling its association with the plasma membrane," *Nature Communications*, vol. 3, no. 1, p. 911, 2012.

[68] H. Lin, Q. Han, J. Wang et al., *Methylation-mediated silencing of RBP7 promotes breast cancer progression through PPAR and PI3K/AKT pathway*, Research Square, 2021.

Hindawi
Journal of Oncology
Volume 2022, Article ID 2466006, 17 pages
https://doi.org/10.1155/2022/2466006

*Research Article*

# The Systematic Analyses of RING Finger Gene Signature for Predicting the Prognosis of Patients with Hepatocellular Carcinoma

**Chunfeng Zhang,[1] Yang Yang,[2] Kun Wang,[2] Muhua Chen,[2] Min Lu,[3] Chenyu Hu,[2] Xiaojuan Du,[4] Baocai Xing,[2] and Xiaofeng Liu [2]**

[1]*Department of Medical Genetics, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China*
[2]*Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Hepatopancreatobiliary Surgery Department I, Peking University Cancer Hospital & Institute, Beijing 100142, China*
[3]*Department of Pathology, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China*
[4]*Department of Cell Biology, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China*

Correspondence should be addressed to Xiaofeng Liu; liuxiaofeng100@bjmu.edu.cn

Received 28 October 2021; Accepted 8 September 2022; Published 26 September 2022

Academic Editor: Junmin Zhang



RING finger (RNF) proteins are frequently dysregulated in human malignancies and are tightly associated with tumorigenesis. However, the expression profiles of RNF genes in hepatocellular carcinoma (HCC) and their relations with prognosis remain undetermined. Here, we aimed at constructing a prognostic model according to RNF genes for forecasting the outcomes of HCC patients using the data from The Cancer Genome Atlas (TCGA) program. We collected HCC datasets to validate the values of our model in predicting prognosis of HCC patients from International Cancer Genome Consortium (ICGC) platform. Then, functional experiments were carried out to explore the roles of the representative RNF in HCC progression. A total of 107 differentially expressed RNFs were obtained between TCGA-HCC tumor and normal tissues. After comprehensive evaluation, a prognostic signature composed of 11 RNFs (RNF220, RNF25, TRIM25, BMI1, RNF216P1, RNF115, RNF2, TRAIP, RNF157, RNF145, and RNF19B) was constructed based on TCGA cohort. Then, the Kaplan-Meier (KM) curves and the receiver operating characteristic curve (ROC) were employed to evaluate predictive power of the prognostic model in testing cohort (TCGA) and validation cohort (ICGC). The KM and ROC curves illustrated the good predictive power in testing and validation cohort. The areas under the ROC curve are 0.77 and 0.76 in these two cohorts, respectively. Among the prognostic signature genes, BMI1 was selected as a representative for functional study. We found that BMI1 protein level was significantly upregulated in HCC tissues. Moreover, the inhibitor of BMI1, PTC-209, displayed an excellent anti-HCC effect *in vitro*. Enrichment analysis of BMI1 downstream targets showed that BMI1 might be involved in tumor immunotherapy. Together, our overall analyses revealed that the 11-RNFs prognostic signature might provide us latent chances for evaluating HCC prognosis and developing novel HCC therapy.

## 1. Introduction

Hepatocellular carcinoma (HCC) is one of the most common malignancies worldwide. HCC resulted in approximately 781,000 deaths worldwide in 2018, ranking as the fourth leading cause of cancer-related death according to the assessment by GLOBOCAN [1, 2]. Despite a series of treatment strategies for HCC have been developed, the overall outcome of HCC patients is poor. In the current, the optimal therapy is curative resection for HCC at early stage, while lots of cases are diagnosed in advanced stage with missing surgical timing [3]. Upon tumor progression, the accumulated somatic DNA

alterations constantly help cancer cells gain malignant behaviors [4, 5]. Due to the wide application of high-throughput sequencing method, researchers gain the chance to globally understand the molecular changes in hepatic cancer cells and establish molecular model for evaluating the status of HCC [6]. It has been found that HCC patients with the same clinical stage own specific molecular subtypes and gene signatures [7]. This further supports the possibility to predict HCC patients' outcomes at molecular level. The establishment of prognosis-related molecular model and the discovery of new therapeutic targets for HCC will be helpful for improving the survival rates of HCC patients.

RING finger (RNF) proteins comprise a large family of proteins which play pivotal roles in protein ubiquitination. Ubiquitination is mainly involved in mediating protein degradation, which in turn regulates cellular activities [8, 9]. It has been reported that ubiquitination participates in lots of intracellular biological processes, such as affecting DNA damage repair, modulating cell metabolism, regulating cell death, and altering therapeutic effect [10]. Ubiquitination is defined as a multistep biochemical reaction, which transfers ubiquitin molecules to the substrates. The indispensable enzymes in this reaction include ubiquitin-activating enzyme (E1), ubiquitin-binding enzyme (E2), and ubiquitin ligase (E3) [11]. Among the enzymes, E3 ligases are responsible for specifically recognizing the substrates and transferring ubiquitin to substrates. In eukaryotes, hundreds of E3 ligases have been identified. Generally, E3s mainly fall into three classes based on the conserved domains for ligase activity, namely HECT, RING finger, and U-box [12]. RNF proteins belong to RING finger E3 harboring RINF finger domain [13]. Dysfunction of RNFs leads to intricate alterations of the transcriptome and proteome in tumor cells, further causing changes in cellular activities, including cell growth, proliferation, apoptosis, migration, and invasion [14, 15]. Increasing number of studies have reported that some RNFs are exceptionally expressed in human cancers and are associated with poor prognosis of patients [16, 17], indicating that the certain RNFs might be latent targets for cancer diagnosis and therapy.

Recent studies have found that some RNF proteins play momentous roles in development and progression of HCC. RNF147, also named TRIM25, enhances the HCC cell survival upon cellular stress by targeting Keap1-Nrf2 pathway [18]. RNF2 promotes ubiquitination of SIK1 in HCC cells and promotes cell growth [19]. The overexpression of RNF40, as an E3 ligase of H2B ubiquitination, indicates poor prognosis of HCC patients [20, 21]. These studies indicate that some RNFs are tightly associated with the progression of HCC. The aberrant expression and function of these RNFs offer us new chances for developing inhibitors of HCC. However, the prognostic roles of RNFs in HCC remain undetermined and this urges us to explore the comprehensive roles of RNF-related genes in HCC.

In the present study, we collected RNA-sequencing data of HCC samples from TCGA and ICGC platforms. After evaluating transcriptomic alterations of RNF genes between HCC and nontumorous tissues, we constructed a risk score model with 11 prognostic RNFs. Moreover, BMI1 was selected as the representative to explore its roles in HCC through functional experiments. Ultimately, we uncovered an RNF-related signature related to the pathogenesis of HCC, which might be applied as latent prognosis-related biomarkers and drug targets for HCC.

## 2. Materials and Methods

*2.1. Patient Samples and Immunoblot.* In total, 18 paired HCC tissues and noncancerous tissues were obtained from the Peking University Cancer Hospital. The study was approved by the ethics committee of the Peking University Cancer Hospital. Western blot was carried out according to the previous reports [22, 23]. Anti-BMI1 antibody (A0211) and Anti-$\beta$-actin antibody (AC026-100) were purchased from Abclonal Technology (China). PTC-209 (S7372-PTC-209) was purchased from Selleck Chemicals (USA).

*2.2. Cell Culture.* HepG2, SMMC-7721, or Huh7 cell lines were purchased from the National Infrastructure of Cell Line Resource (NSTI, China). We cultured cells using DMEM or RPMI 1640 medium added with 10% fetal bovine serum. After passage or indicated treatment, cells were cultured in a humidified chamber in 5% $CO_2$ at 37°C.

*2.3. Cell Proliferation Evaluation.* Cell proliferation was analyzed using MTS kit (Promega, USA). Briefly, cells were digested and seeded into the 96-well plate. Then, the drugs were added to the cells as indicated. Cell number was determined for each day by MTS assay according to the manufacturer' protocol.

*2.4. Colony Formation Experiment.* Colony formation assay was carried out according to previously published protocol [24]. Briefly, cells were treated with the indicated concentrations of PTC-209 and subsequently seeded into 6-well plate. Fourteen days later, the colonies were fixed with paraformaldehyde and stained with 0.1% crystal violet (Beyotime, China). The visible colonies were counted using ImageJ software.

*2.5. Data Collection and Analysis.* The RNA sequencing data and the related clinical data of HCC were obtained from TCGA website (https://portal.gdc.cancer.gov/). The RNF gene set including 227 genes was obtained from GEPIA website [25]. We preprocessed the raw data using Limma package in R software. Then, we collected the RNFs expression profile from LIHC (Liver Hepatocellular Carcinoma) dataset including 374 tumorous and 50 normal samples for the following analyses. The differentially expressed RNFs were identified using the criteria: $|\log_2 FC \geq 1|$ and FDR < 0.05. The "pheatmap" package was employed for unsupervised clustering analysis in R software. The validation dataset of gene expression and clinical trait data (the Liver Cancer-RIKEN JP) was collected from the ICGC database (https://dcc.icgc.org/). GSE97172 dataset was downloaded from GEO database (https://www.ncbi.nlm.nih.gov/gds/). Similarly, the raw data was preprocessed using R software.

*2.6. KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathway and GO (Gene Ontology) Enrichment Analyses.* Briefly, GO and KEGG pathway enrichment analyses were carried out

using the DAVID platform (https://david.ncifcrf.gov/) [26]. The GO analysis terms contain molecular function (MF), cellular component (CC), and biological process (BP). The GO enrichment results and KEGG pathways were visualized through the "GOplot package" in R software. We used $p < 0.05$ as the threshold for statistical significance.

*2.7. PPI (Protein-Protein Interaction) Network Construction and Key Modules Identification.* The PPI network of the differentially expressed genes was established through the STRING database (http://www.string-db.org/) [27]. The results gained by STRING database were further analyzed and visualized using Cytoscape 3.7.1 software. The key modules were discovered using Molecular Complex Detection (MCODE) plug-in based on MCODE score and node counts.

*2.8. RNFs-Based Prognostic Model Construction.* We performed univariate Cox regression analysis to identify the prognosis-related RNF genes using the "survival" package in R software. $p < 0.05$ was used as the statistically significant. Lasso Cox regression analyses were further performed for uncovering prognostic signatures using the "glmnet" package in R software. The risk score for each sample was calculated according to RNFs expression (Expi) and coefficient value ($\beta$i): Risk score = exp (0.20*RNF220 + 0.33*RNF25 + 0.04*TRIM25 + 0.11*BMI1 + 0.057*RNF216P1 + 0.007*RNF115 + 0.019*TRAIP + 0.12*RNF2 + 0.039*RNF157 + 0.14*RNF145 + 0.21*RNF19B).

According to the median of risk score values, HCC patients were divided into high-risk and low-risk groups for the subsequent analyses. The "survival" package was utilized to calculate the differences of overall survival (OS) time between the two groups. Besides, the ROC curve analysis was also performed to evaluate the prognostic capability of this model using the "survivalROC" package in R software.

*2.9. Genomic Analysis and Drug Prediction for Prognostic RNFs.* The mutation of RNF genes was analyzed through cBioPortal (https://www.cbioportal.org/), which is a public platform used for analyzing and visualizing the cancer genomics datasets. All data for RNFs-related drugs were analyzed through PharmacoDB (https://pharmacodb.pmgenomics.ca/) [28].

*2.10. Detection of the Risk Genes in Protein Level.* The protein expression level of these risk genes was evaluated through the Human Protein Atlas (HPA) database for further verifying the transcriptional level of the related genes (https://www.proteinatlas.org/).

*2.11. Statistical Analyses.* Statistical analyses were conducted using the SPSS software (Inc., Chicago, IL) or R software. Univariate and multivariate Cox analyses were utilized to identify independent prognostic factors. The overall survival (OS) and disease-free survival (DFS) curves were calculated using Kaplan-Meier analysis, and the statistical significance between different groups was calculated by log-rank test. A Student's $t$-test was used to calculate the significance between two groups of the indicated samples. $p < 0.05$ was used as statistically significant.

## 3. Results

*3.1. The Differently Expressed RNFs Were Identified between Hepatic Normal and Tumorous Tissues.* To study the roles of RNFs in HCC, we downloaded the transcriptomic file from the TCGA-LIHC dataset consisting of 50 normal tissue samples and 374 HCC samples. Then, we screened the differently expressed RNFs using the cut-off values of FDR < 0.05 and $|\log_2 FC| \geq 1$. Totally, 105 upregulated and 2 downregulated RNF genes were uncovered (Figure 1(a)). The expression profile of these identified RNFs is described in Figure 1(b).

*3.2. GO and KEGG Pathway Analyses Were Performed with the Differently Expressed RNFs.* To analyze the function of these identified RNFs, we conducted GO enrichment and KEGG pathway analyses. GO analyses categorized RNFs into three groups including GOBP, GOCC, and GOMF. The top 10 enriched GOBP, GOCC, and GOMF are presented in Figures 2(a)–2(c), respectively. Based on the GO analyses, differently expressed RNFs were mainly involved in protein polyubiquitination, ubiquitin ligase complex, and zinc-ion binding pathways. Furthermore, KEGG analysis suggest that the top 5 enriched pathways were "ubiquitin mediated proteolysis", "Notch signaling pathway", "signaling pathways regulating pluripotency of stem cells", "pathways in cancer", and "protein processing in endoplasmic reticulum" (Figure 2(d)). These pathways are tightly associated with protein ubiquitination and tumorigenesis. Thus, these enrichment analyses of these differently expressed RNFs in HCC revealed that the alterations of ubiquitination network caused by these RNFs contribute to HCC development and progression.

*3.3. The PPI Network Was Constructed Using the Differently Expressed RNFs.* Next, we explored the critical protein association networks of these 107 RNFs. STRING platform were utilized to establish PPI network for exploring the interactions among these RNFs. We obtained 103 nodes and 548 edges using a $p$ value of PPI concentration<1.0e-16 as selection criteria. The top two enriched clusters in the PPI network were identified using the Cytoscape with MCODE plug-in (Figures 2(e) and 2(f)). The function of each cluster was next analyzed by pathway enrichment analysis. The results revealed that Module 1 was mainly involved in ubiquitin-mediated proteolysis. Module 2 mainly participated in regulating pluripotency of stem cells and ER (endoplasmic reticulum)-related protein processing. Thus, the PPI analysis confirmed that RNFs play an important role in protein ubiquitination, which is tightly associated with HCC progression.

*3.4. Recognition of Prognosis-Related RNFs and Establishment of Prognostic Model in HCC.* To explore the prognostic value of the differently expressed RNFs, the univariate Cox regression analyses were conducted using corresponding TCGA clinical data. As shown in Supplementary Figure S1, 29 candidate RNFs were found to be associated with the overall survival (OS) of HCC patients. Moreover, we performed Lasso Cox regression analysis to select the prognostic RNFs for constructing prognostic model (Figures 3(a) and 3(b)). According to the integrated prognostic relevance, 11 RNF genes (RNF220, RNF25, TRIM25, BMI1, RNF216P1,

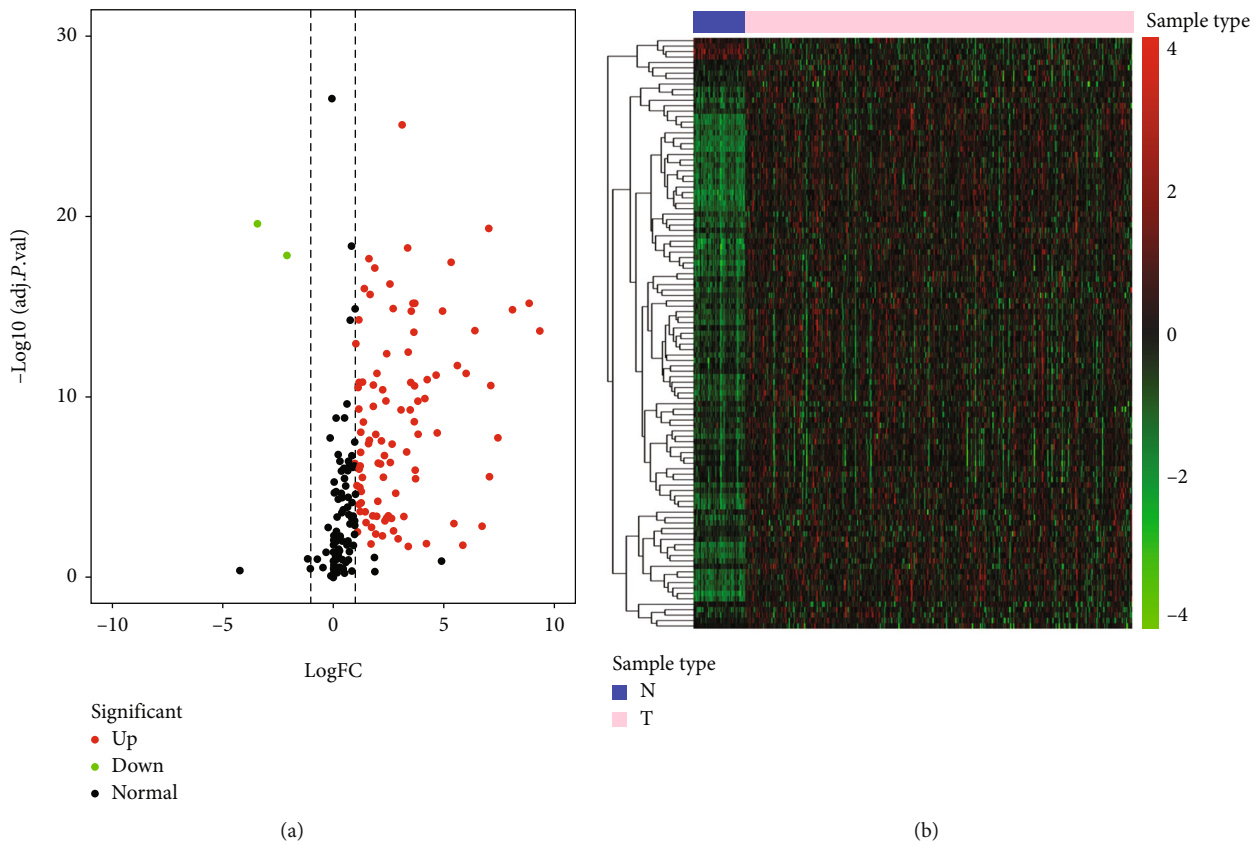


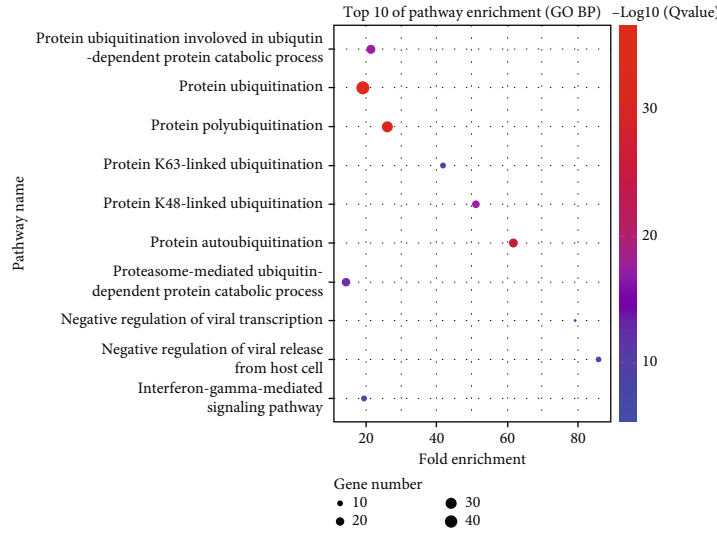

Figure 1: The differentially expressed RNFs were identified between hepatic normal and tumor tissues. (a) Volcano plot displaying RNFs. The upregulated RNF is shown in red and downregulated RNF is shown in green. Black means no difference. (b) Hierarchical clustering of differentially expressed RNFs is presented. The columns mean samples and the rows indicate RNFs. The green means downregulation while the red means upregulation.

RNF115, RNF2, TRAIP, RNF157, RNF145, and RNF19B) were selected to construct a risk score model. We calculated the risk score for each patient according to expression values of 11 RNFs as follows: Risk score = exp (0.20*RNF220 + 0.33*RNF25 + 0.04*TRIM25 + 0.11*BMI1 + 0.057*RNF216 P1 + 0.007*RNF115 + 0.019*TRAIP + 0.12*RNF2 + 0.039* RNF157 + 0.14*RNF145 + 0.21*RNF19B). According to results from Lasso-penalized Cox regression, all these RNFs had positive coefficients and acted as independent prognostic factors for OS of the patients with HCC.

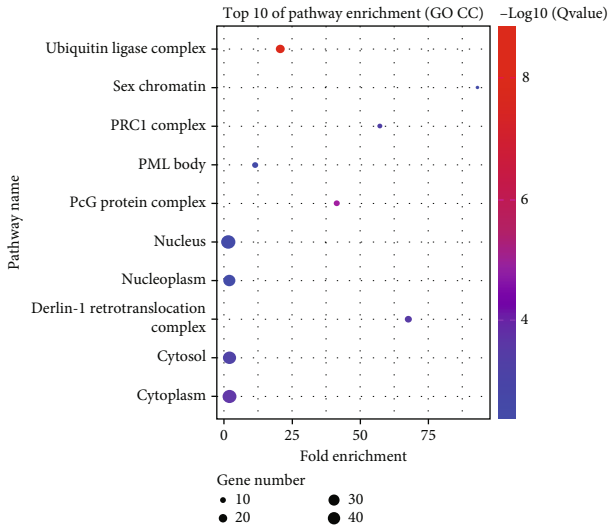
Based on the median of risk score values, the patients with HCC were divided into low-risk and high-risk groups. Compared with the patients in the low-risk, HCC patients in high-risk group had a worse outcome by OS analysis in TCGA cohort (Figure 3(c)). In addition, we employed the ROC analysis to evaluate the prognostic ability of the risk model. As shown in Figure 3(d), the areas under the ROC curve (AUC) of our model were 0.778, 0.675 and 0.698 at 1 year, 3 years, and 5 years, respectively, using the data of TCGA cohort. This result suggests that the risk score model is more accurate in the short-term follow-up. Moreover, the risk scores of these patients were ranked and exhibited according to risk score (Figure 3(e)). The survival status of each HCC patient in TCGA-LIHC cohort was shown in Figure 3(f). Consistently, there were shorter OS time and higher mortality rates in the patients with high-risk than those in the low-risk (Figure 3(f)). These data confirmed the utility of our risk score model in evaluating the prognosis of HCC patients. Additionally, the expression profiles of these prognosis-related RNFs between the two groups were presented in Figure 3(g). The result revealed that all the prognostic RNFs were upregulated in the HCC patients with high-risk.

*3.5. Independent Prognostic Value of the Risk Score Model Was Analyzed in HCC.* Subsequently, we performed univariate Cox regression analysis to explore the relations of clinicopathological characteristics and the risk score with prognosis in the TCGA-LIHC patients. As shown in Figure 4(a), tumor stage and risk score were tightly related to the overall survival. The following multivariate Cox regression analysis further confirmed that risk score and TNM stage were two independent prognostic factors of survival for patients with HCC (Figure 4(b)). Next, we compared the AUC values of these clinical factors at 1-year, 3-year, or 5-year. The results indicated that our model more precisely forecasted 1-year OS rate compared to TNM stage (Figures 4(c)–4(e)).

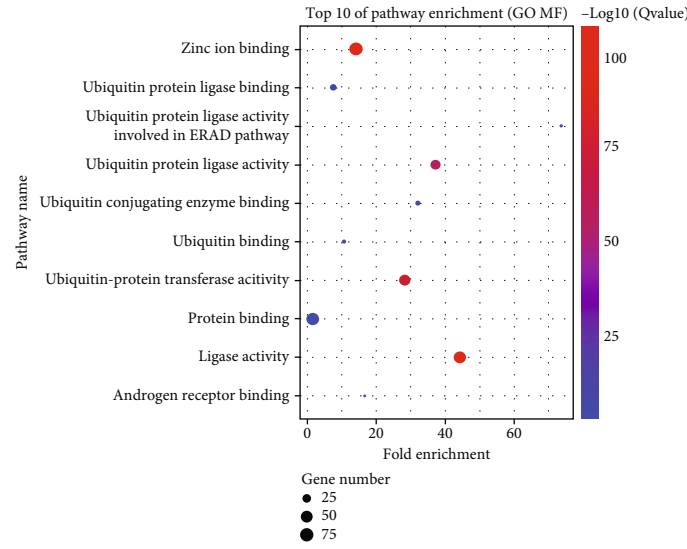
*3.6. The Prognostic Signature Was Validated for OS Prediction in ICGC Cohort.* For confirming the predictive power of our model, HCC patients with clinical information from the ICGC were enrolled as a validation cohort. Based on the expression level of the 11 RNFs, the risk score of each patient were calculated using the risk score formula. Then, we used the median



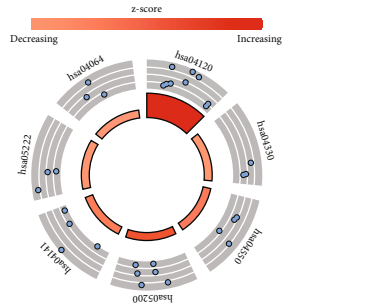

(a)




(b)



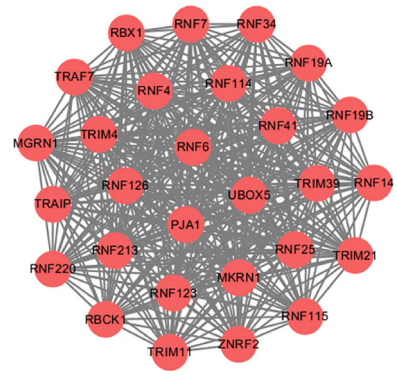

(c)




| ID | Description |
|---|---|
| hsa04120 | Ubiquitin mediated proteolysis |
| hsa04330 | Notch signaling pathway |
| hsa04550 | Signaling pathways regulating pluripotency of stem cells |
| hsa05200 | Pathways in cancer |
| hsa04141 | Protein processing in endoplasmic reticulum |
| hsa05222 | Small cell lung cancer |
| hsa04064 | NF-kappa B signaling pathway |

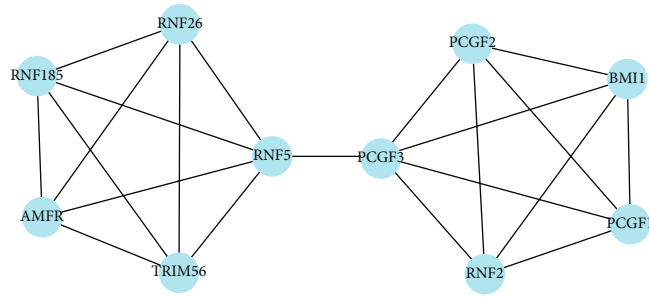
LogFC
Upregulated

(d)




(e)

Figure 2: Continued.




(f)

Figure 2: GO enrichment, KEGG pathway, and PPI network analyses of differentially expressed RNFs. (a–c) The top 10 items enriched GOBP (a), GOCC (b), and GOMF (c) in GO analyses. (d) The top seven enriched pathways are shown in KEGG analysis. (e) Module 1. MCODE score = 28.00, Nodes = 28, Edges = 378. (f) Module 2.MCODE score = 4.67, Nodes = 10, Edges = 21.

of risk score values as cut-off and subdivided the patients into high-risk and low-risk groups. We analyzed the difference of survival of these two groups by survival analysis and found that patients with high-risk had a shorter OS compared with the patients with low risk (Figure 4(f)). Then, the risk scores of these HCC patients were ranked and exhibited in Figure 4(g). The survival status of each ICGC-HCC patient was estimated. As shown in Figure 4(h), there were shorter OS time and higher mortality rates in the patients with high-risk than those in the low-risk group. ROC curve was further calculated and the AUC for the OS model at 1-year and 3-year were 0.766 and 0.662, respectively, in ICGC cohort (Figure 4(i)). Together, these data indicated that our risk score model is useful for estimating the outcome of ICGC HCC patients. Together, our risk score model based on RNF gene expression precisely predicts the prognosis of HCC patients.

*3.7. The Relationships between the Prognosis-Related RNFs and Clinicopathological Features Were Analyzed.* Next, we explored the associations of these prognostic RNFs and clinicopathological features, including TNM stage and tumor grade. As shown in Figure 5(a), nine RNFs (RNF220, RNF25, TRIM25, RNF216P1, RNF115, TRAIP, RNF2, RNF157, and RNF145) were found to be upregulated in HCC patients with advanced grade ($p < 0.05$). Moreover, six RNFs were increased in patients with advanced stage, including RNF220, BMI1, RNF216P1, TRAIP, RNF145, and RNF19B (Figure 5(b)). Importantly, the risk scores of patients with advanced grade or stage are much higher, indicating that our prognostic model was associated with both tumor grade and TNM stage.

The genetic changes of these RNFs were further determined using the cBioPortal website. Mainly, the genetic alterations of these RNF genes include truncating mutation, missense mutation, deep deletion, structural variant, and amplification. The top 5 most significantly altered genes are RNF115, RNF2, RNF157, TRIM25, and TRAIP in HCC samples (Figure 5(c)).

*3.8. The Protein Levels of Prognostic RNFs Expression Were Analyzed through the HPA Database.* We next assessed the protein levels of the 11 prognostic RNFs through the HPA database. We found that eight RNFs including RNF220, RNF25, TRIM25, BMI1, RNF115, TRAIP, RNF157, and RNF19B were overexpressed in HCC cells compared to normal cells by immunohistochemistry (IHC) staining (Figures 6(a)–6(h)). Finally, the staining of RNF2 and RNF145 proteins was missing and needs further analysis.

*3.9. Functional Study of the Roles of BMI1 in HCC Cells.* Due to the prognostic values of the 11 RNFs for HCC patients, we next explored the potential drugs for these RNFs through the PharmacoDB database. Only BMI1 was identified as small molecule targets among these prognostic RNFs. As a core element of polycomb repressive complex 1(PRC1), BMI1 has been found to be associated with various human cancers and become an attractive therapeutic target. One of the specific BMI1 inhibitors, PTC-209, displays high potency in repressing the growth of some types of cancer cells [29, 30]. However, it is unknown whether PTC-209 shows the potential capability in anti-HCC therapy.

Survival analysis using TCGA-LIHC cohort revealed that the high BMI1 expression was associated with shorter OS and DFS in the HCC patients (Figure 7(a)). Moreover, the expression of BMI1 on 18 paired cancerous and matched noncancerous sections of HCC tissues from our center was evaluated by immunoblotting. As shown in Figure 7(b), BMI1 was upregulated in cancerous tissues compared to peritumoral tissues. Importantly, the MTS assay showed that the proliferation of HCC cells was significantly repressed by PTC-209 treatment (Figures 7(c)–7(e)). To further confirm our results, we randomly selected SMMC-7721 and HepG2 cells to perform colony formation assay. Consistently, colony formation of HCC cells was significantly decreased upon PTC-209 treatment (Figure 7(f)). Together, these data indicated that PTC-209 inhibits HCC cell proliferation and growth *in vitro*.

To explore the mechanisms underlying BMI1-induced HCC progression, we analyzed the differentially expressed genes in liver tissues from BMI1-knockout mice (GSE97172). As shown in Figure 7(g), 624 genes were found to be differentially expressed. KEGG analysis suggested that immune system-related pathways were significantly altered upon BMI1 loss (Figure 7(h)). These results indicated that BMI1 might be involved in regulating tumor immune microenvironment. Tumor-infiltrating lymphocytes can be used as an independent indicator of the survival and sentinel lymph node status in human cancers [31]. Therefore, we analyzed

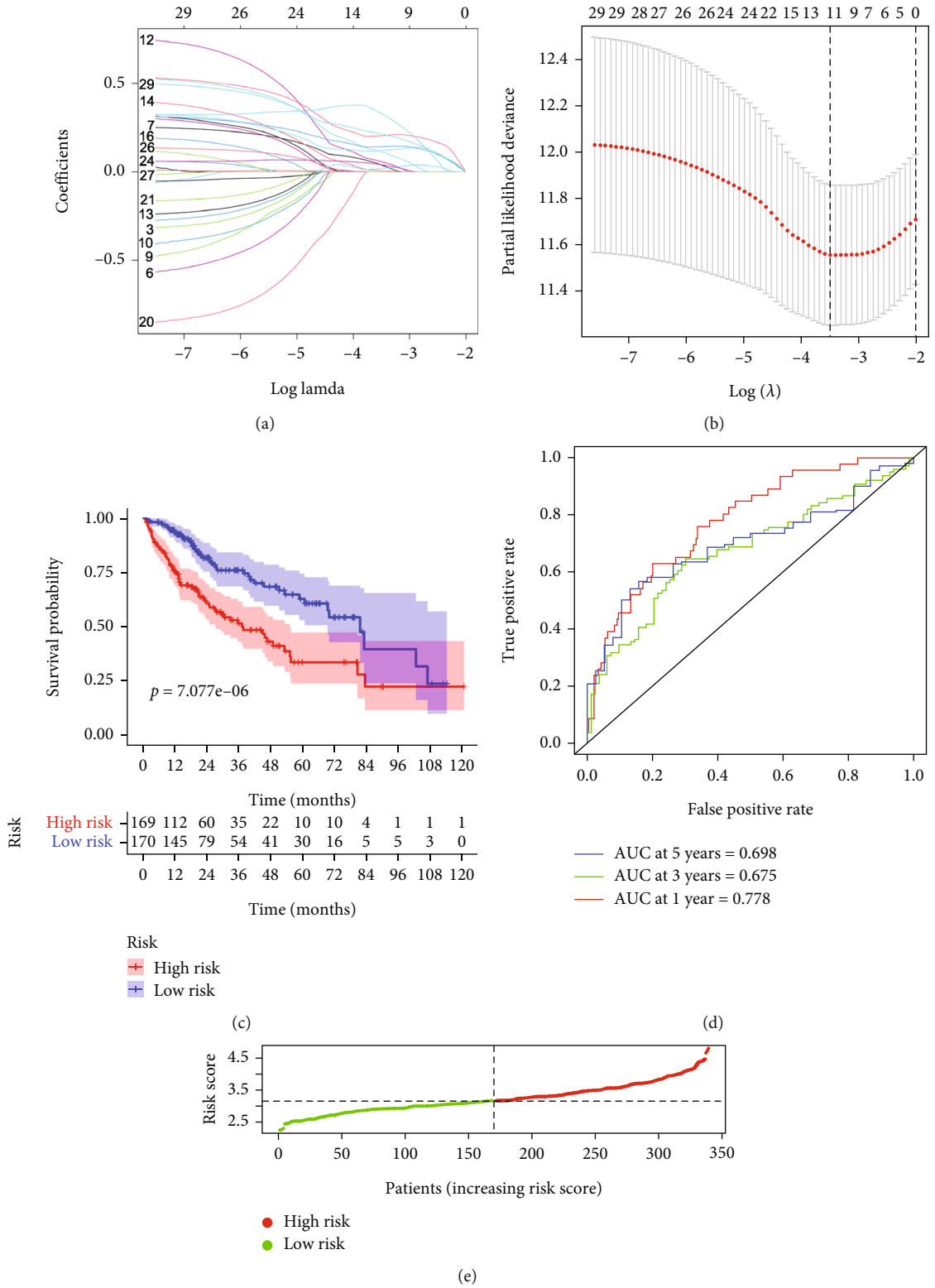



Figure 3: Continued.

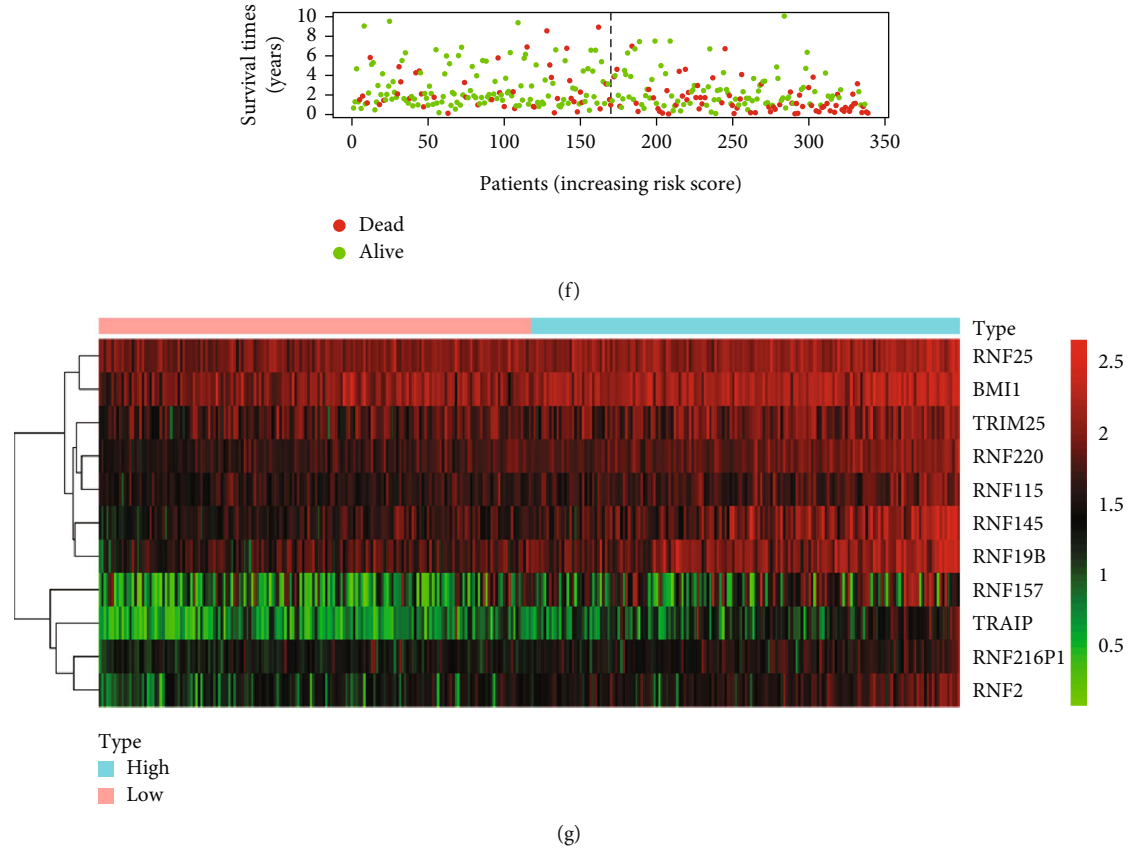



Figure 3: The risk score model was established based on prognostic RNFs using the TCGA HCC cohort. (a) Each curve indicates an *RNF* gene and the best lambda was computed to minimize mean cross-validated error. (b) The proportional hazards model was employed for cross-validation to select tuning parameter. (c) Kaplan-Meier analysis of TCGA-LIHC patients based on risk score. (d) Time-dependent ROC curves were established to estimate OS according to the risk score. (e) HCC patients were subdivided into high-risk and low-risk groups using the median of risk scores as cut-off. (f) Scatter plots displayed the associations of risk score status with survival outcome in TCGA HCC patients. (g) The heat map displayed the expression levels of prognostic RNFs in each HCC patient from TCGA cohort. The red represents increased expression and the green indicates decreased expression.

whether BMI1 expression was correlated with the infiltration of immune cells in HCC using TIMER website. The results show that BMI1 expression has positive correlations with $CD8^+$ T cell, $CD4^+$ T cell, neutrophil, B cell, and dendritic cell infiltration levels while BMI1 expression is negatively associated with natural killer cell infiltration level (Figure 7(i)). We also assessed the correlation between BMI1 expression and immune checkpoint-related molecules, including PD-L1, Galectin 9, HVEM, and IDO1, all of which are important marker genes for cancer immune therapy [32]. As shown in Figure 7(j), there were positive correlations between BMI1 expression level and the expression levels of these four marker genes. Together, these data suggest that BMI1 might be a novel target in HCC immunotherapy.

## 4. Discussion

Metastasis and recurrence after resection are common for HCC, which causes treatment failure and cancer-related death [33–35]. Development of prognostic assessment system will be favorable for follow-up after treatment, in order to attenuate tumor progression caused by metastasis or relapse, especially in patients with high-risk. Clinically, prognostic evaluation includes tumor status, cancer-related symptoms, and liver function of the patient [36]. Recently, with the improvement of high-throughput technologies, it is possible to develop molecular typing for cancer diagnosis and treatment. In the present study, we identified RNF-based molecular biomarkers and constructed a risk score model to forecast outcomes of HCC patients using TCGA-LIHC cohort. We further validated the risk model using ICGC-LIHC dataset. Moreover, we also explored the roles of a typical prognostic RNFs and BMI1, in HCC progression. Recently, some latent biomarkers and therapeutic targets have been identified for HCC by bioinformatics strategies. The eleven RNA-binding proteins (RBPs) were screened to construct a prognostic model for indicating overall outcomes of HCC patients [37]. A four-gene metabolic signature predicting OS for HCC was built [38]. The immune-related gene prognostic signature for HCC was also constructed [39]. Currently, the prognosis model based on the RNFs of HCC has not been reported. Thus, our study determines the prognostic values of RNF genes in HCC and provides a new idea for HCC diagnosis and treatment.

Dysregulation of some RNFs leads to abnormal ubiquitination of the important proteins in tumor cells and drives tumorigenesis, including HCC [40]. Here, we aimed at studying the

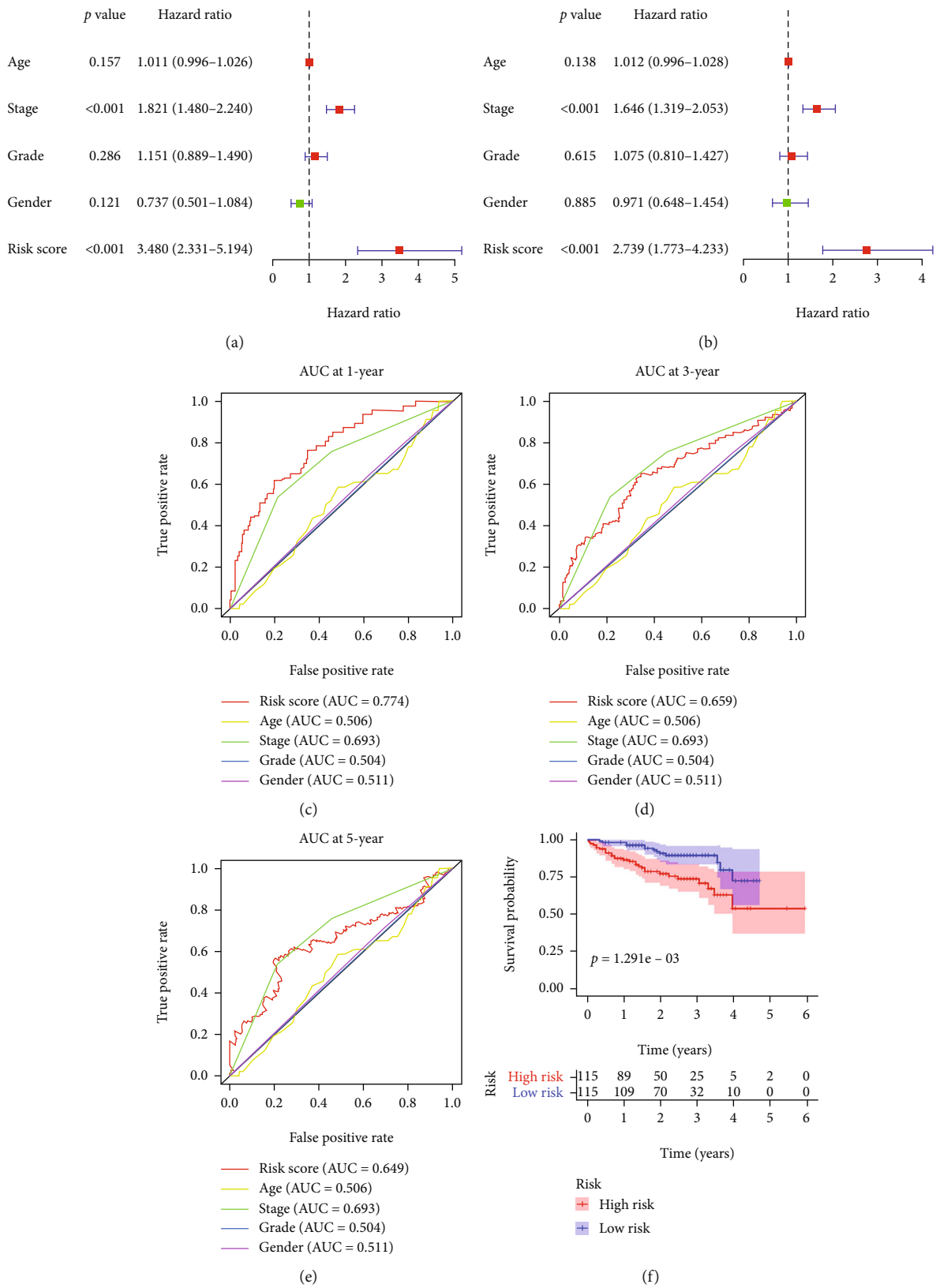



Figure 4: Continued.

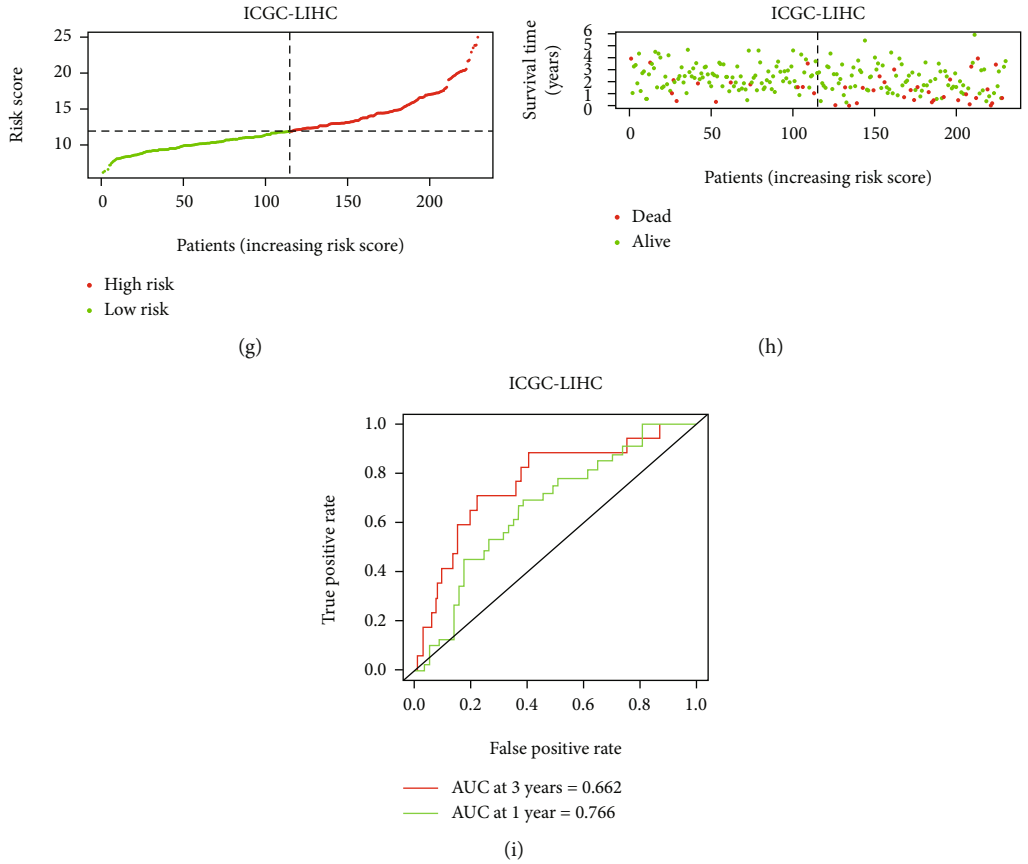
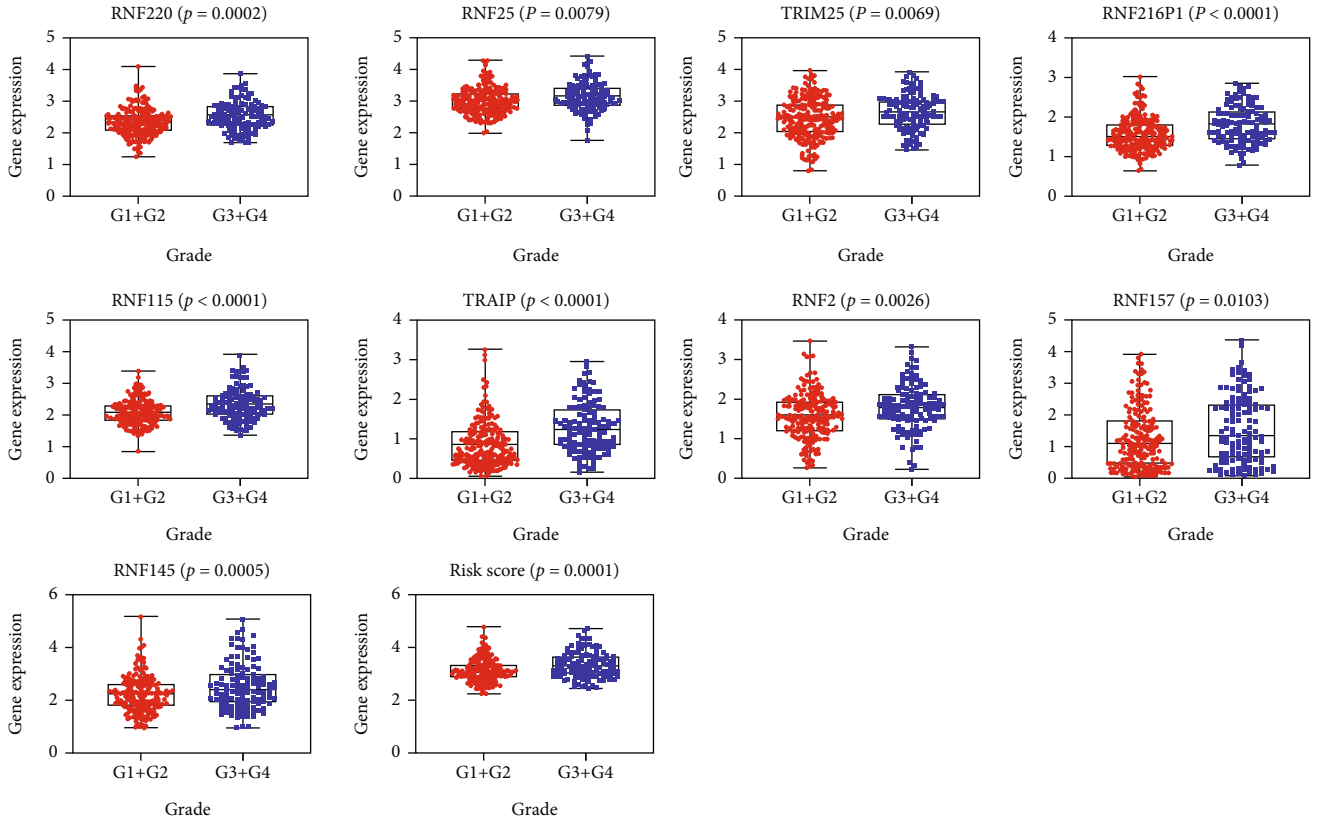



Figure 4: Validation of the reliability of the risk score model. (a) Univariate Cox regression analysis of the indicators. (b) Multivariate Cox regression analysis of the indicators. (c–e) Time-dependent ROC curves were used to analyze the predicting ability of the indicated factors in the TCGA HCC cohort. (f) Kaplan-Meier curve of ICGC patients was analyzed according to the risk score. (g) ICGC-LIHC patients were divided into high-risk and low-risk groups based on the median of risk score. (h) Scatter plots displayed the relationships of risk score status with the survival outcome in the ICGC HCC patients. (i) Time-dependent ROC curves were calculated according to risk scores in the ICGC HCC cohort.

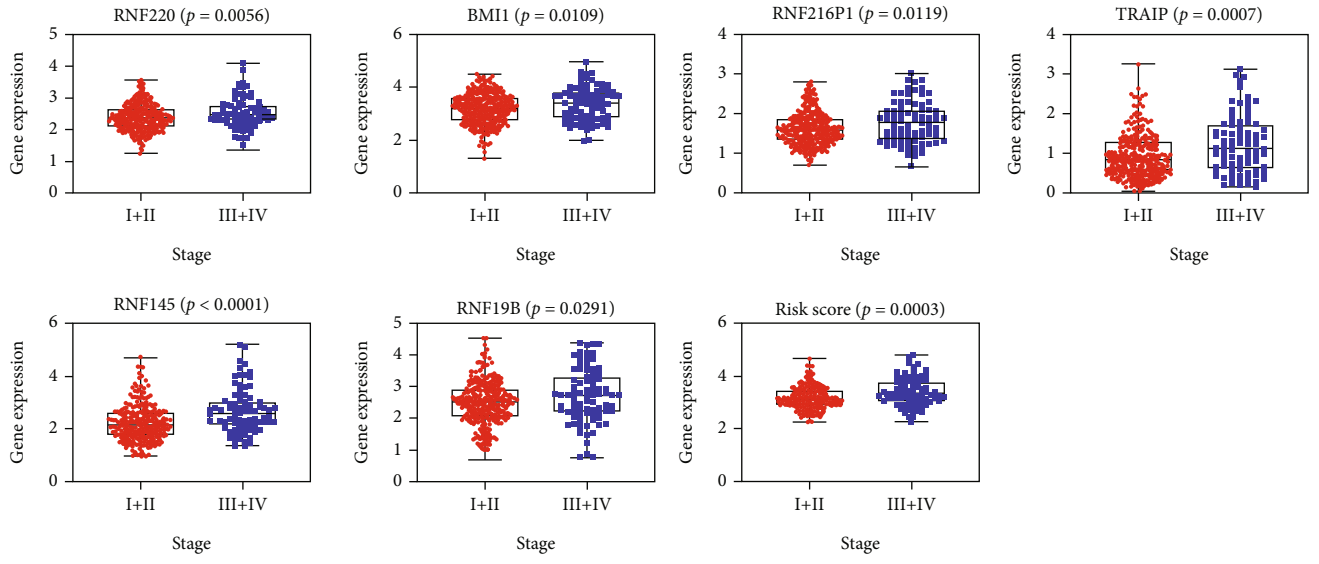
universal roles of RNF genes in forecasting OS of HCC patients. Through screening differentially expressed genes between HCC and normal tissues, total 107 RNFs were identified as differentially expressed RNFs using TCGA cohort. Then, GO enrichment and KEGG pathway analyses revealed that the differentially expressed RNFs were greatly involved in ubiquitin-mediated proteolysis. We uncovered 29 prognosis-associated candidate RNFs by univariate Cox regression analysis. With Lasso Cox regression, eleven RNFs genes (RNF220, RNF25, TRIM25, BMI1, RNF216P1, RNF115, RNF2, TRAIP, RNF157, RNF145, and RNF19B) were identified to construct a risk score model. We confirmed the stability and reliability of this model using ICGC data as the validation set. The results suggested that the model is accurate for distinguishing HCC patients with different survival outcomes. Univariate and multivariate analyses further confirmed that this prognosis model could independently indicate overall prognosis of patients with HCC. ROC curve also manifested that our model based on the 11 RNFs had a good predictive ability. These results suggest that our risk model might be applied to screen high-risk patients for personalized detection or follow-up.

Among the eleven RNF genes, the majority (RNF220, RNF25, TRIM25, RNF115, BMI1, TRAIP, RNF2, RNF157, RNF145, and RNF19B) have been reported to function in ubiquitination and play roles in tumorigenesis. RNF220 is associated with progression of leukemia or medulloblastoma [41, 42]. RNF25 upregulates gefitinib resistance via promoting ERK reactivation in EGFR-mutant NSCLC cells [43]. TRIM25 takes part in tumor growth, metastasis, and chemoresistance with its ubiquitin ligase activities [18, 44]. RNF115 is correlated with the prognosis of patients with lung adenocarcinoma or invasive breast cancer [45, 46]. As a master regulator of DNA repair, dysfunction of TRAIP is associated with tumor development and progression [47, 48]. Overexpression of RNF2 is positively correlated with progression of many cancers, including HCC, melanoma, pancreatic cancer, and gastric cancer [49].

BMI1 is a core element of the PRC1 complex which mediates gene silencing via monoubiquitination of histone H2A. The polycomb group (PcG) proteins encoding transcriptional repressors are indispensable for maintenance of stem cell pluripotency [50, 51]. The PcG proteins form multimeric protein complexes to regulate transcription of development-related genes, which are called as polycomb repressive complexes (PRCs) [52]. Currently, two major PRCs have been identified, PRC1 and PRC2, both of which modify chromatin to stably




(a)




(b)

Figure 5: Continued.


Figure 5: The relevance between prognostic RNFs and clinicopathologic features was analyzed. (a) The association of these prognostic RNFs with tumor grade was evaluated. (b) The relationship between prognostic RNFs and tumor stage was analyzed. (c) Oncoplot for each prognostic *RNF* gene.

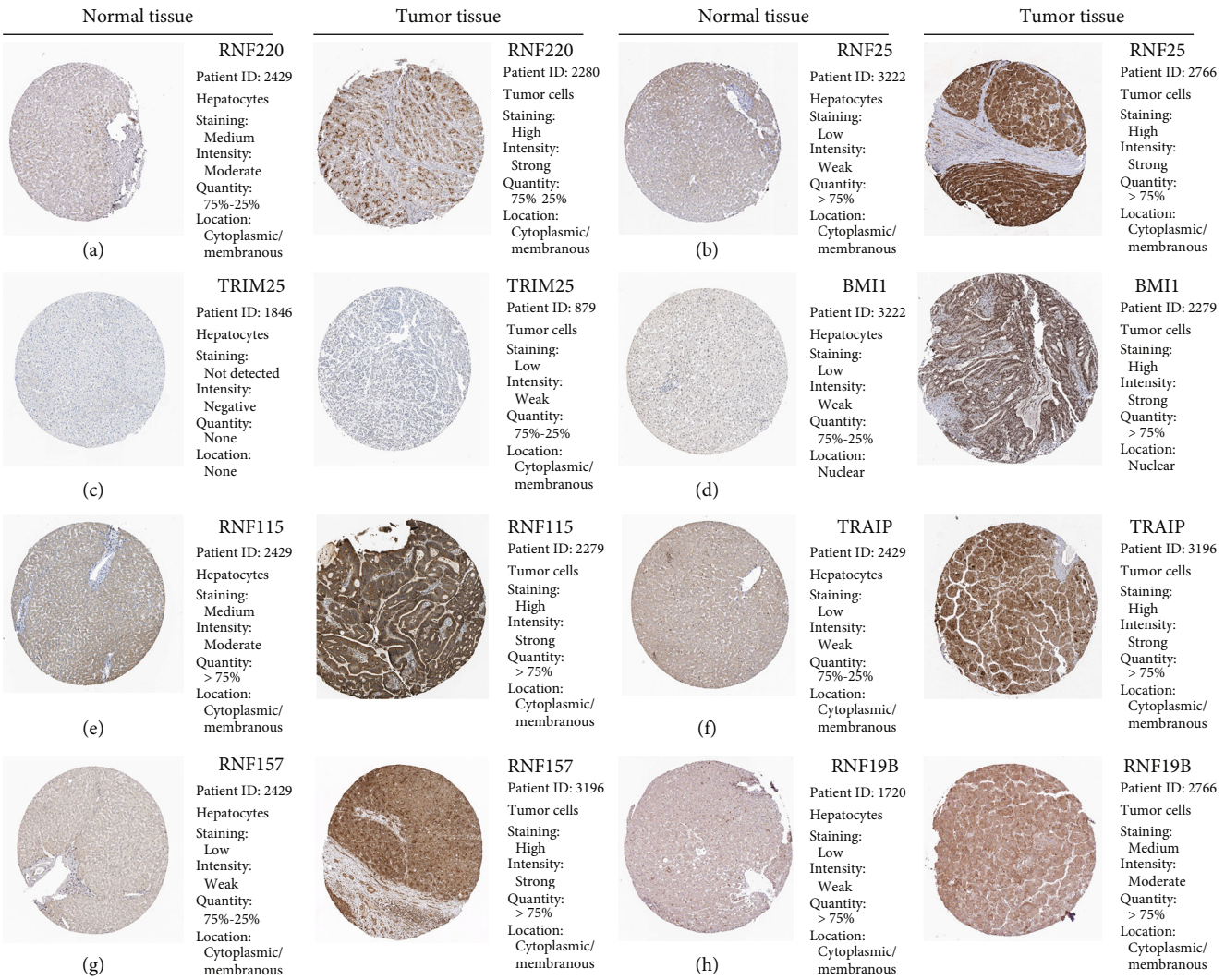



Figure 6: The protein levels of the prognostic RNFs were evaluated in the HPA database. (a–h) IHC staining of prognostic RNF in HCC tumor tissues and normal liver tissues.

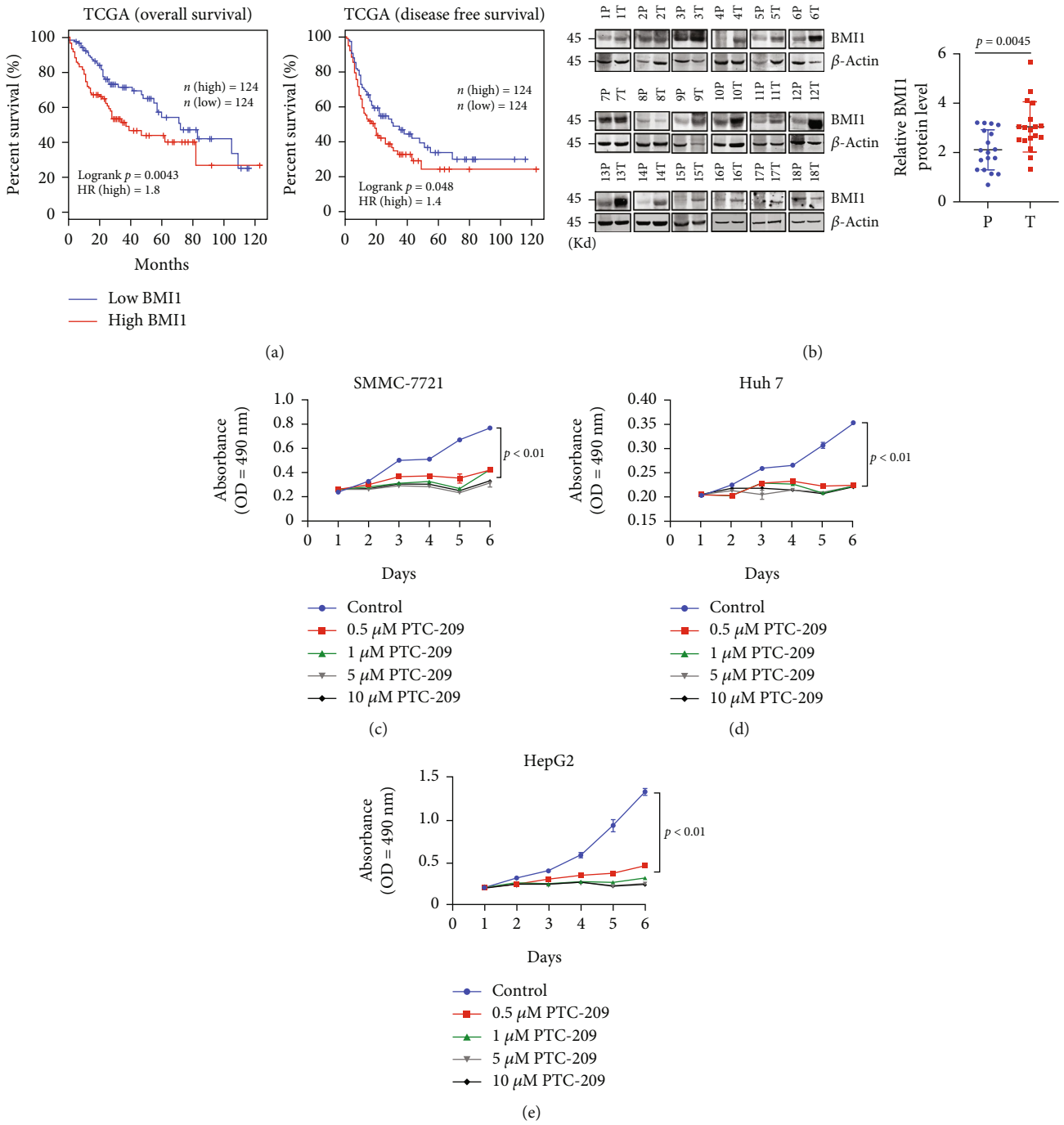



Figure 7: Continued.

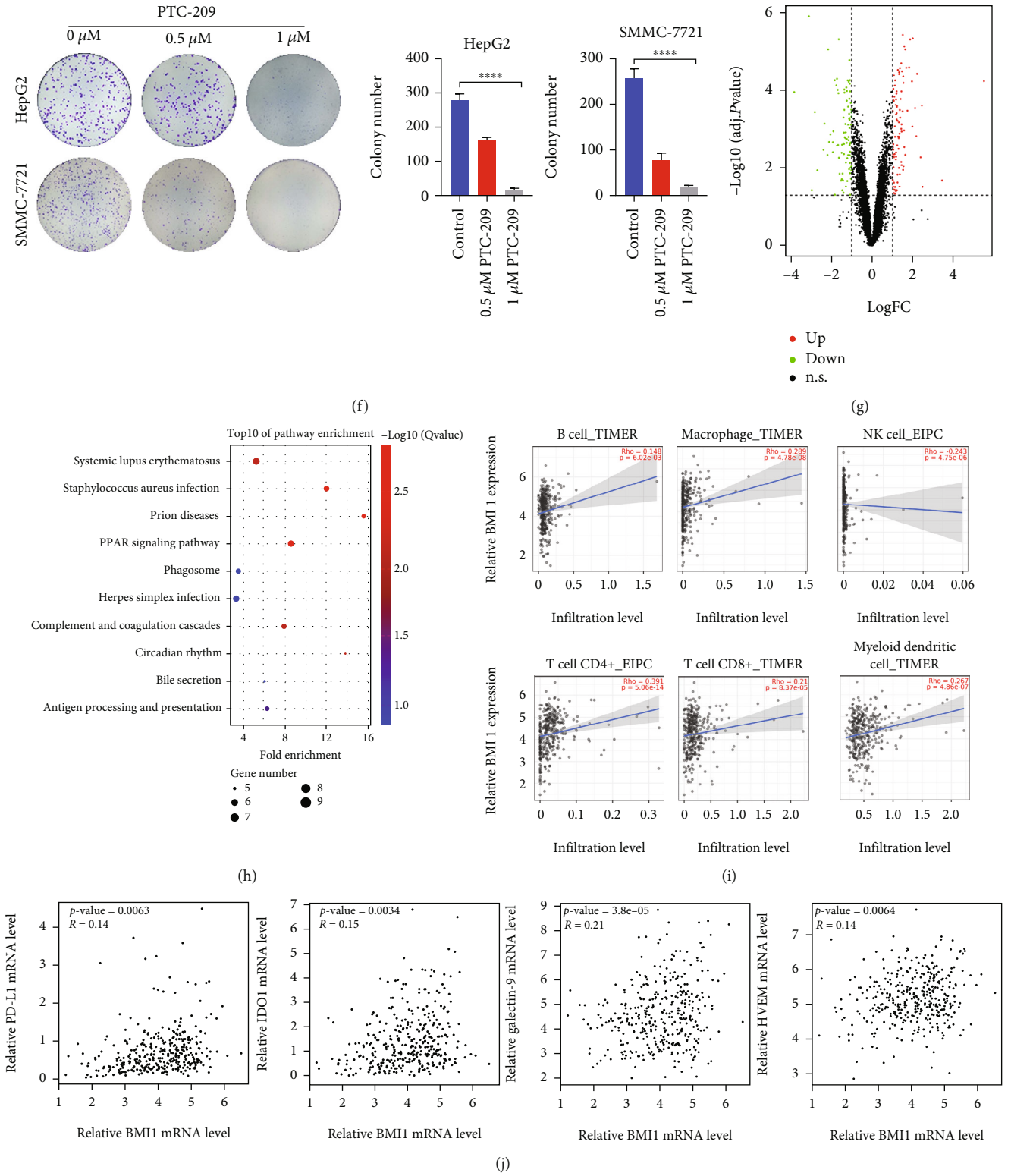



Figure 7: Functional studies of BMI1 in HCC progression. (a) Kaplan-Meier survival analysis for overall survival and disease-free survival of the TCGA-LIHC patients based on BMI1 mRNA level. (b) Western blot analysis of BMI1 expression in 18 individual paired HCC tissues. (c–e) Cell proliferation assay was performed in HCC cells after PTC-209 treatment. (f) Colony formation assay was conducted in HCC cells treated with PTC-209. (****$p < 0.0001$) (g) Volcano plot for DEGs between BMI1 wild-type and BMI1 knockout mice liver tissues using GSE97172 dataset. (h) KEGG pathway analysis for DEGs identified in (g). (i) The correlation of BMI1 expression with immune infiltration level was analyzed in HCC tissues. (j) Coexpression analysis between the expression level of PD-L1, Galectin 9, HVEM, IDO, and BMI1 using TCGA-LIHC data.

silence transcription at targeted genes [53]. As an E3 ubiquitin ligase, BMI1 works with its partners to catalyze the PcG-dependent ubiquitination of histone H2A in order to modulate transcription [54]. In tumorigenesis, BMI1 plays important roles in promoting cancer stemness, leading to tumor metastasis, recurrence, and drug resistance [55]. Thus, the development of small molecule inhibitors against BMI1 will offer potential opportunities for cancer treatment. PTC-209, as an important inhibitor of BMI1, downregulated BMI1 by reducing mRNA level. PTC-209 exerts inhibitory effects for several cancers, such as breast cancer, non-small cell lung cancer, and acute myeloid leukemia [30, 56, 57]. However, the effect of PTC-209 in anti-HCC is unclear. Here, we found that BMI1 is critical in constructing the risk score model. Our further studies show that BMI1 is upregulated in HCC tissues and the upregulation of BMI1 is associated with poor outcomes of HCC patients, confirming that BMI1 plays important roles in hepatocarcinogenesis. Importantly, we found that PTC-209 significantly inhibits HCC cell growth and proliferation. Thus, our results identify the inhibition of BMI1 as a potential strategy for HCC treatment. We also analyzed the downstream targets of BMI1 by comparing expression profiles of BMI1 wild-type and BMI1-knockout tissues. Strikingly, we found that these targets are enriched in immune-related events. Moreover, the expression of BMI1 correlates with immune cells' infiltration level in HCC, suggesting that BMI1 might be a novel target for improving HCC immune therapy.

Recently, the proteolysis targeting chimeras (PROTACs) technology attracts growing attention of scientific institutes and pharmaceutical companies [58]. PROTACs are designed based on the ubiquitin-proteasome system to induce the degradation of targeted protein. Briefly, the ligands in PROTACs combine with E3 ligase and the targeted protein, respectively, and the linker connects the two ligands and pulls them closer together. PROTACs show positive results for degrading the "undruggable" oncoproteins which lack of binding pockets by small molecule inhibitors [59]. The PROTACs-related new drugs are tested in clinical trials for cancer therapy [59]. Due to the importance of RNF protein-related ubiquitination in tumorigenesis, it is possible that the PROTACs based on RNFs could be useful for cancer treatment. This will be further studied in the future.

In addition, some limitations should be addressed in the future to increase the possibility of our risk model in HCC diagnosis. First, our study was a retrospective study based on the public datasets. It will be better to validate our model using data from prospective clinical trials. Second, the detailed molecular mechanisms of the RNF genes in hepatocellular carcinogenesis are not fully understood. Moreover, discoveries of effective drugs for targeting prognostic RNFs will be more helpful for HCC treatment by *in vivo* experiments and clinical trials. In the future study, we will try to address this issue in subsequent studies.

In summary, using a systematic and comprehensive biomarker discovery and validation approach, we uncovered that an RNF-related gene signature could act as a prognostic indicator for evaluating prognosis of HCC patients and guide HCC treatment. We also identified that BMI1 is tightly associated with HCC progression, which might be a new therapeutic target for HCC.

## Data Availability

All data generated or analyzed during this study are included in this article and its supplementary files. The analyzed data during the current study are available from the corresponding authors on reasonable request.

## Ethical Approval

The study protocol was approved by the Ethics Committee of Peking University Cancer Hospital.

## Consent

All the authors agree to the content of the paper and their being listed as a coauthor of the paper. Written informed consent was acquired from all the patients.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Authors' Contributions

Xiaofeng Liu worked on conception and designs. Chunfeng Zhang and Xiaofeng Liu worked on bioinformatic analysis. Chunfeng Zhang worked on experiments. Chunfeng Zhang, Min Lu, Yang Yang, Muhua Chen, Chenyu Hu, and Kun Wang worked on data collection. Xiaojuan Du, Xiaofeng Liu, and Baocai Xing worked on manuscript preparation and supervision.

## Acknowledgments

This work was supported by the grants from the National Natural Science Foundation of China (Grant No. 81802305, 31971192, and 81874143), the Beijing Natural Science Foundation (Grant No. 7192035), and the Science Foundation of Peking University Cancer Hospital (2020-1).

## Supplementary Materials

Supplementary Figure S1: univariate Cox regression analysis of OS for each prognosis-associated candidate RNF gene. *(Supplementary Materials)*

## References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] M. Chen, C. Zhang, W. Liu, X. Du, X. Liu, and B. Xing, "Long noncoding RNA LINC01234 promotes hepatocellular carcinoma progression through orchestrating aspartate metabolic reprogramming," *Molecular Therapy*, vol. 30, no. 6, pp. 2354–2369, 2022.

[3] A. Koulouris, C. Tsagkaris, V. Spyrou, E. Pappa, A. Troullinou, and M. Nikolaou, "Hepatocellular carcinoma: an overview of the changing landscape of treatment options," *Journal of Hepatocellular Carcinoma*, vol. 8, pp. 387–401, 2021.

[4] S. Li and M. Mao, "Next generation sequencing reveals genetic landscape of hepatocellular carcinomas," *Cancer Letters*, vol. 340, no. 2, pp. 247–253, 2013.

[5] Q. Li, X. Liu, K. Jin et al., "NAT10 is upregulated in hepatocellular carcinoma and enhances mutant p53 activity," *BMC Cancer*, vol. 17, no. 1, p. 605, 2017.

[6] J. M. Llovet, R. Montal, D. Sia, and R. S. Finn, "Molecular therapies and precision medicine for hepatocellular carcinoma," *Nature Reviews Clinical Oncology*, vol. 15, no. 10, pp. 599–616, 2018.

[7] P. H. D. Nguyen, S. Ma, C. Z. J. Phua et al., "Intratumoural immune heterogeneity as a hallmark of tumour evolution and progression in hepatocellular carcinoma," *Nature Communications*, vol. 12, no. 1, p. 227, 2021.

[8] X. Liu, Y. Tan, C. Zhang et al., "NAT10 regulates p53 activation through acetylating p53 at K120 and ubiquitinating Mdm2," *EMBO Reports*, vol. 17, no. 3, pp. 349–366, 2016.

[9] Y. Wang, J. Dai, Y. Zeng, J. Guo, and J. Lan, "E3 ubiquitin ligases in breast cancer metastasis: a systematic review of pathogenic functions and clinical implications," *Frontiers in Oncology*, vol. 11, article 752604, 2021.

[10] C. W. Fhu and A. Ali, "Dysregulation of the ubiquitin proteasome system in human malignancies: a window for therapeutic intervention," *Cancers (Basel)*, vol. 13, no. 7, p. 1513, 2021.

[11] I. Dikic, S. Wakatsuki, and K. J. Walters, "Ubiquitin-binding domains - from structures to functions," *Nature Reviews Molecular Cell Biology*, vol. 10, no. 10, pp. 659–671, 2009.

[12] C. E. Berndsen and C. Wolberger, "New insights into ubiquitin E3 ligase mechanism," *Nature Structural & Molecular Biology*, vol. 21, no. 4, pp. 301–307, 2014.

[13] M. B. Metzger, V. A. Hristova, and A. M. Weissman, "HECT and RING finger families of E3 ubiquitin ligases at a glance," *Journal of Cell Science*, vol. 125, no. 3, pp. 531–537, 2012.

[14] T. Okamoto, K. Imaizumi, and M. Kaneko, "The Role of Tissue-Specific Ubiquitin Ligases, RNF183, RNF186, RNF182 and RNF152, in Disease and Biological Function," *International Journal of Molecular Sciences*, vol. 21, no. 11, 2020.

[15] A. M. Jaworska, N. A. Wlodarczyk, A. Mackiewicz, and P. Czerwinska, "The role of TRIM family proteins in the regulation of cancer stem cell self-renewal," *Stem Cells*, vol. 38, no. 2, pp. 165–173, 2020.

[16] J. Fu, L. Liao, K. S. Balaji, C. Wei, J. Kim, and J. Peng, "Epigenetic modification and a role for the E3 ligase RNF40 in cancer development and metastasis," *Oncogene*, vol. 40, no. 3, pp. 465–474, 2021.

[17] A. Mattioni, L. Castagnoli, and E. Santonico, "RNF11 at the Crossroads of Protein Ubiquitination," *Biomolecules*, vol. 10, no. 11, 2020.

[18] Y. Liu, S. Tao, L. Liao et al., "TRIM25 promotes the cell survival and growth of hepatocellular carcinoma through targeting Keap1-Nrf2 pathway," *Nature Communications*, vol. 11, no. 1, p. 348, 2020.

[19] C. Qu and Y. Q. Qu, "Down-regulation of salt-inducible kinase 1 (SIK1) is mediated by RNF2 in hepatocarcinogenesis," *Oncotarget*, vol. 8, no. 2, pp. 3144–3155, 2017.

[20] C. C. So, S. Ramachandran, and A. Martin, "E3 Ubiquitin ligases RNF20 and RNF40 are required for double-stranded break (DSB) repair: evidence for monoubiquitination of histone H2B lysine 120 as a novel axis of DSB signaling and repair," *Molecular and Cellular Biology*, vol. 39, no. 8, p. 39, 2019.

[21] X. Y. Zheng, K. Chen, X. L. Liu, Y. Pan, and H. Liu, "High RNF40 expression indicates poor prognosis of hepatocellular carcinoma," *International Journal of Clinical and Experimental Pathology*, vol. 11, no. 5, pp. 2901–2906, 2018.

[22] C. Zhang, C. Hu, K. Su et al., "The integrative analysis of thrombospondin family genes in pan-cancer reveals that THBS2 facilitates gastrointestinal cancer metastasis," *Journal of Oncology*, vol. 2021, Article ID 4405491, pp. 1–19, 2021.

[23] X. Liu, D. Xu, Z. Liu et al., "THBS1 facilitates colorectal liver metastasis through enhancing epithelial-mesenchymal transition," *Clinical & Translational Oncology*, vol. 22, no. 10, pp. 1730–1740, 2020.

[24] X. F. Liu, K. Q. Su, X. Y. Sun et al., "Sec62 promotes stemness and chemoresistance of human colorectal cancer through activating Wnt/beta-catenin pathway," *Journal of Experimental & Clinical Cancer Research*, vol. 40, no. 1, 2021.

[25] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, and Z. Zhang, "GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses," *Nucleic Acids Research*, vol. 45, no. W1, pp. W98–W102, 2017.

[26] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists," *Nucleic Acids Research*, vol. 37, no. 1, pp. 1–13, 2009.

[27] D. Szklarczyk, A. L. Gable, D. Lyon et al., "STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic Acids Research*, vol. 47, no. D1, pp. D607–D613, 2019.

[28] P. Smirnov, V. Kofia, A. Maru et al., "PharmacoDB: an integrative database for mining in vitro anticancer drug screening studies," *Nucleic Acids Research*, vol. 46, no. D1, pp. D994–D1002, 2018.

[29] J. H. Wang, Y. J. Xing, Y. Y. Wang et al., "A novel BMI-1 inhibitor QW24 for the treatment of stem-like colorectal cancer," *Journal of Experimental & Clinical Cancer Research*, vol. 38, no. 1, 2019.

[30] H. T. Shen, P. J. Chien, S. H. Chen et al., "BMI1-mediated pemetrexed resistance in non-small cell lung cancer cells is associated with increased SP1 activation and cancer stemness," *Cancers (Basel)*, vol. 12, no. 8, p. 2069, 2020.

[31] F. Azimi, R. A. Scolyer, P. Rumcheva et al., "Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma," *Journal of Clinical Oncology*, vol. 30, no. 21, pp. 2678–2683, 2012.

[32] K. Sideras, K. Biermann, K. Yap et al., "Tumor cell expression of immune inhibitory molecules and tumor-infiltrating lymphocyte count predict cancer-specific survival in pancreatic and ampullary cancer," *International Journal of Cancer*, vol. 141, no. 3, pp. 572–582, 2017.

[33] D. Xu, X. Liu, L. Wang, and B. Xing, "Hepatectomy plus adjuvant transcatheter arterial chemoembolization improves the survival rate of patients with multicentric occurrence of hepatocellular carcinoma," *Oncology Letters*, vol. 16, no. 5, pp. 5882–5890, 2018.

[34] Y. Tan, J. Zheng, X. Liu et al., "Loss of nucleolar localization of NAT10 promotes cell migration and invasion in hepatocellular carcinoma," *Biochemical and Biophysical Research Communications*, vol. 499, no. 4, pp. 1032–1038, 2018.

[35] Y. Yang, P. Ren, X. Liu et al., "PPP1R26 drives hepatocellular carcinoma progression by controlling glycolysis and epithelial-mesenchymal transition," *Journal of Experimental & Clinical Cancer Research*, vol. 41, no. 1, p. 101, 2022.

[36] J. Bruix, M. Reig, and M. Sherman, "Evidence-based diagnosis, staging, and treatment of patients with hepatocellular carcinoma," *Gastroenterology*, vol. 150, no. 4, pp. 835–853, 2016.

[37] L. Wang, Z. Zhang, Y. Li, Y. Wan, and B. Xing, "Integrated bioinformatic analysis of RNA binding proteins in hepatocellular carcinoma," *Aging (Albany NY)*, vol. 13, no. 2, pp. 2480–2505, 2020.

[38] G. M. Liu, W. X. Xie, C. Y. Zhang, and J. W. Xu, "Identification of a four-gene metabolic signature predicting overall survival for hepatocellular carcinoma," *Journal of Cellular Physiology*, vol. 235, no. 2, pp. 1624–1636, 2020.

[39] W. B. Chen, M. L. Ou, O. G. Tang, Y. Dai, and W. Du, "Identification and validation of immune-related gene prognostic signature for hepatocellular carcinoma," *Journal of Immunology Research*, vol. 2020, Article ID 5494858, 14 pages, 2020.

[40] J. Yin, J. M. Zhu, and X. Z. Shen, "The role and therapeutic implications of RING-finger E3 ubiquitin ligases in hepatocellular carcinoma," *International Journal of Cancer*, vol. 136, no. 2, pp. 249–257, 2015.

[41] P. C. Ma, T. An, L. Zhu et al., "RNF220 is required for cerebellum development and regulates medulloblastoma progression through epigenetic modulation of Shh signaling," *Development*, vol. 147, no. 21, 2020.

[42] Y. M. Pan, N. An, X. P. Deng, Q. X. Zhang, and X. Du, "RNF220 promotes the proliferation of leukaemic cells and reduces the degradation of the cyclin D1 protein through USP22," *Blood Cells, Molecules, and Diseases*, vol. 86, article 102490, 2021.

[43] J. H. Cho, Y. M. You, Y. I. Yeom et al., "RNF25 promotes gefitinib resistance in EGFR-mutant NSCLC cells by inducing NF-κB-mediated ERK reactivation," *Cell Death & Disease*, vol. 9, no. 6, p. 587, 2018.

[44] Y. M. He, X. M. Zhou, S. Y. Jiang et al., "TRIM25 activates AKT/mTOR by inhibiting PTEN via K63-linked polyubiquitination in non-small cell lung cancer," *Acta Pharmacologica Sinica*, vol. 43, no. 3, pp. 681–691, 2022.

[45] A. M. Burger, Y. G. Gao, Y. Amemiya et al., "A novel RING-type ubiquitin ligase breast cancer-associated gene 2 correlates with outcome in invasive breast cancer," *Cancer Research*, vol. 65, no. 22, pp. 10401–10412, 2005.

[46] X. T. Wu, Y. H. Wang, X. Y. Cai et al., "RNF115 promotes lung adenocarcinoma through Wnt/beta-catenin pathway activation by mediating APC ubiquitination," *Cancer & Metabolism*, vol. 9, no. 1, 2021.

[47] R. A. Wu, D. R. Semlow, A. N. Kamimae-Lanning et al., "TRAIP is a master regulator of DNA interstrand crosslink repair," *Nature*, vol. 567, no. 7747, pp. 267–272, 2019.

[48] M. E. Harley, O. Murina, A. Leitch et al., "TRAIP promotes DNA damage response during genome replication and is mutated in primordial dwarfism," *Nature Genetics*, vol. 48, no. 1, pp. 36–43, 2016.

[49] Q. Yan, B. J. Chen, S. Hu et al., "Emerging role of RNF2 in cancer: from bench to bedside," *Journal of Cellular Physiology*, vol. 236, no. 8, pp. 5453–5465, 2021.

[50] S. Aranda, G. Mas, and L. Di Croce, "Regulation of gene transcription by polycomb proteins," *Science Advances*, vol. 1, no. 11, 2015.

[51] U. Grossniklaus and R. Paro, "Transcriptional silencing by polycomb-group proteins," *Cold Spring Harbor Perspectives in Biology*, vol. 6, no. 11, 2014.

[52] J. A. Simon and R. E. Kingston, "Mechanisms of polycomb gene silencing: knowns and unknowns," *Nature Reviews Molecular Cell Biology*, vol. 10, no. 10, pp. 697–708, 2009.

[53] Z. Wang, M. D. Gearhart, Y. W. Lee et al., "A non-canonical BCOR-PRC1.1 complex represses differentiation programs in human ESCs," *Cell Stem Cell*, vol. 22, no. 2, pp. 235–251.e9, 2018, e239.

[54] A. Fitieh, A. J. Locke, M. Motamedi, and I. H. Ismail, "The role of polycomb group protein BMI1 in DNA repair and genomic stability," *International Journal of Molecular Sciences*, vol. 22, no. 6, p. 2976, 2021.

[55] H. R. Siddique and M. Saleem, "Role of BMI1, a stem cell factor, in cancer recurrence and chemoresistance: preclinical and clinical evidences," *Stem Cells*, vol. 30, no. 3, pp. 372–378, 2012.

[56] K. Barbosa, A. Deshpande, B. R. Chen et al., "Acute myeloid leukemia driven by the CALM-AF10 fusion gene is dependent on BMI1," *Experimental Hematology*, vol. 74, pp. 42–51.e3, 2019.

[57] Q. Wang, Z. Li, Y. Wu et al., "Pharmacological inhibition of BMI1 by PTC-209 impaired tumor growth in head neck squamous cell carcinoma," *Cancer Cell International*, vol. 17, no. 1, p. 107, 2017.

[58] Z. Hu and C. M. Crews, "Recent developments in PROTAC-mediated protein degradation: from bench to clinic," *Chembiochem*, vol. 23, no. 2, article e202100270, 2022.

[59] M. Lv, W. Hu, S. Zhang, L. He, C. Hu, and S. Yang, "Proteolysis-targeting chimeras: a promising technique in cancer therapy for gaining insights into tumor development," *Cancer Letters*, vol. 539, article 215716, 2022.

Hindawi
Journal of Oncology
Volume 2022, Article ID 5300523, 17 pages
https://doi.org/10.1155/2022/5300523

*Research Article*

# On the Core Prescriptions and Their Mechanisms of Traditional Chinese Medicine in Hepatitis B, Liver Cirrhosis, and Liver Cancer Treatment

**Zhendong Wang**,[1] **Yong Zhang**,[2] **Qiuyun Zhang**,[3] **Qiang Ao**,[4] **Changyong Luo**,[5] **Bochuan Wang**,[1] **Chen Bai**,[1] **Xueyi Ge**,[6] **Yuhan Wang**,[1] **Jing Wang**,[5] **Ying Qian**,[3] **He Yu**,[1] **and Xiaohong Gu**[1]

[1]*School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China*
[2]*Liver Disease Branch, Shandong Hospital of Traditional Chinese Medicine, Jinan, China*
[3]*Capital Medical University, Beijing, China*
[4]*Beijing Tcmages Pharmaceutical Co., Ltd., Beijing, China*
[5]*Dongfang Hospital, Beijing University of Chinese Medicine, Beijing, China*
[6]*Shandong University of Traditional Chinese Medicine, Jinan, China*

Correspondence should be addressed to He Yu; yuhe@bucm.edu.cn and Xiaohong Gu; guxiaohong1962@163.com

Received 12 January 2022; Accepted 23 July 2022; Published 23 September 2022

Academic Editor: Jayaprakash Kolla



*Background*. As a frequent cause of death in cancer patients, liver cancer usually occurs in hepatitis B and cirrhosis. In China, Chinese people have been using traditional Chinese medicine (TCM) in treating various chronic liver diseases, which could effectively improve the symptoms and slow down the progression of liver diseases. However, due to the complexity rules of TCM prescription, their action mechanisms are still not clearly understood, which may affect the popularization of effective prescriptions. This study aims to identify the core TCM herbs in the treatment of hepatitis B, liver cirrhosis, and liver cancer so as to clarify the mechanism of action of the core herb networks. *Methods*. There were 1,673 prescriptions for chronic liver diseases collected in this study, of which 854 were hepatic B prescriptions, 530 were for liver cirrhosis, and 289 were for liver cancer. The basic characteristics of herbal medicine were firstly explained via descriptive analysis, then the core prescriptions of herbal medicine were analyzed through association rule, and finally, the mechanism of core prescriptions was explored with the help of systematic network pharmacology and by applying such databases as TCMIP, HERB, OMIM, GeneCards, KEGG, and software like RStudio and Cytoscape. *Results*. The rule of the core prescriptions in these cases was characterized by the application of herbs with both cold and warm properties, in which bitter herbs with cold property took priority. Tonifying deficiency, clearing heat, and activating blood circulations to remove stasis were common treatment principles for the three liver diseases. Turmeric Root Tuber (YuJin), White Peony Root (BaiShao), Bupleurum (ChaiHu), Salvia miltiorrhiza (DanShen), and Astragali Radix (HuangQi) were prescribed the most in hepatitis B treatment to invigorate the spleen and soothe the liver. Astragali Radix (HuangQi), Tuckahoe (FuLing), Atractylodis Macrocephalae Rhizoma (BaiZhu), Fructus Polygoni Orientalis (ShuiHongHuaZi), and Curcumae Rhizome (EZhu) were most frequently applied in liver cirrhosis treatment to replenish qi and activate blood. *Oldenlandia* (BaiHuaSheSheCao), Bearded Scutellaria (BanZhiLian), Curcumae Rhizome (EZhu), and Cardamom (DouKou) were most frequently prescribed to eliminate cancer toxin, invigorate the spleen, and activate blood. These core herbs mainly act through signal transduction and immune system pathways, in which the PI3K-Akt pathway plays a key role. The core prescription for liver cirrhosis regulated more endocrine system pathways than the hepatitis B prescription, and liver cancer prescription regulated more nervous system-related pathways. *Conclusion*. Three core prescriptions for hepatitis B, liver cirrhosis, and liver cancer treatment were identified, which acted mainly through signal transduction and immune system pathways to regulate immunity and cell growth and participate in inflammation inhibition, in which liver cancer prescription regulated more pathways, especially more nervous system-related pathways than the other two.

## 1. Introduction

The global morbidity and mortality rate of liver cancer has been increasing in the past years, which has become the third most common cause of death from tumors [1]. Asia and Africa have the highest incidence rate of it in the world [2]. It was reported by Surveillance, Epidemiology, and End Results (SEER) that liver cancer has become the fastest growing cause of cancer-related death in the United States ever since the beginning of the 21st century [3]. It was estimated that more than 1 million people worldwide would be affected by liver cancer each year by 2025. Hepatitis B virus infection is the main risk factor for the progress of liver cancer, accounting for 60% of primary hepatocellular carcinoma cases in Asia and Africa [4–6]. The long-term liver inflammation caused by hepatitis B virus infection triggers compensatory liver repair and regeneration and ultimately leads to liver fibrosis or cirrhosis [7]. This pathological state plays an important role in the precancerous environment of the liver; more than 80% of hepatocellular carcinomas occur in the fibrotic or cirrhotic liver [8]. It is known that the incidence and progression of hepatitis B, cirrhosis, and liver cancer are correlated to a certain degree, and the medical and family burden they brought about is gradually increasing. Therefore, it is of great significance to block the progression of liver diseases and reduce the social burdens they incur.

With the development of complementary and alternative medicine around the world, the role of traditional Chinese medicine (TCM) in tumor prevention and treatment has drawn more and more attention. Over 60% of the antitumor medicines were developed from natural herbal products these years [9]. In China, TCM has been widely used in liver cancer treatment combined with chemotherapy, radiotherapy, surgical resection, and liver transplantation and even used alone in the middle and late stages of liver cancer [10, 11]. TCM participation could reduce the adverse reactions of conventional treatment, relieve symptoms, protect liver function, and improve overall survival [12–14]. Studies have shown that, in hepatitis B cases, patients who received TCM treatment had a significantly lower risk of liver cirrhosis than those who did not [15]. TCM intervention after hepatitis B and liver cancer surgery was clinically proved to be able to prevent the occurrence of liver cancer [16, 17]. Compared with applying antiviral herbs alone in the treatment of liver fibrosis, a combining application of TCM herbs could increase the reversal rate of liver fibrosis, reduce the risk of liver cancer [18], and effectively alleviate the symptoms caused by liver dysfunction [19].

By analyzing the application characteristics, the prescription rules of the herbs, and constructing diagrams, data mining could help to discover the potential correlation between herbs, which might provide support for doctors in developing treatment protocols [20]. Analysis of frequency statistics and association rules is a common technical method of it [21, 22]. Studies have identified the core prescriptions through retrospective analysis of patients with chronic kidney diseases, as well as prescription frequency and association rules analysis, and have explored the TCM pathogenesis of chronic kidney diseases [23].

The "multicomponent and multitarget network" in systematic pharmacology happened to coincide with the holism concept of TCM prescription, which might effectively bridge the gap between modern medicine and TCM and provide new research methods for TCM theories and would greatly promote the synergistic development between TCM and modern medicine [24]. By analyzing the compound targets and the enrichment pathways of TCM prescriptions, the mechanism of core prescriptions would be studied further. Some researchers have applied data mining combined with network pharmacology in exploring new prescriptions for TCM treatment of recurrent respiratory infections and demonstrated the mechanism of action of the core prescriptions [20].

This study proposed a comprehensive method based on data mining and systematic pharmacology, analyzed the rules of TCM prescriptions, and revealed their mechanisms. Clinical data of patients with liver diseases from the two centers were collected, and the prescriptions of these cases were analyzed by descriptive statistics and association rules to explore the prescription rules and core prescriptions in the cases. The important targets were retrieved from the database, and Metascape and KEGG were used to further analyze the biomarker pathways. Workflow is shown in Figure 1. The findings of this study would help us to identify the core herbs and prescriptions and analyze the mechanism of action of the core prescriptions, thus providing a deeper understanding of TCM prescriptions and promoting the development of medicine for chronic liver disease treatment as well as TCM studies in the future.

## 2. Materials and Methods

*2.1. Data Collection and Processing.* In this study, 1,673 prescriptions from the outpatient departments of the Affiliated Hospital of Shandong University of Traditional Chinese Medicine by Professor Changjian Yin and the Affiliated Hospital of Capital Medical University by Professor Ying Qian were analyzed retrospectively from January 2020 to June 2021, among which, 854 cases were diagnosed as chronic hepatitis B, 530 were cirrhosis, and 289 were liver cancer.

Hepatitis B diagnostic criteria referred to the guidelines for the prevention and treatment of chronic hepatitis B [25]. Diagnostic criteria included HBsAg and HBV-DNA (+) in serum samples. Inclusion criteria included those who met the diagnostic criteria. Exclusion criteria included ① simultaneous infection with HIV, HCV, and HDV; ② those who suffered from liver cirrhosis, liver cancer, and other liver diseases; ③ patients with severe mental and other internal organ diseases.

The diagnostic criteria of decompensated liver cirrhosis referred to the Chinese guidelines on the management of liver cirrhosis [26]. Diagnostic criteria were as follows: ① Histology results should meet the diagnosis of liver cirrhosis.

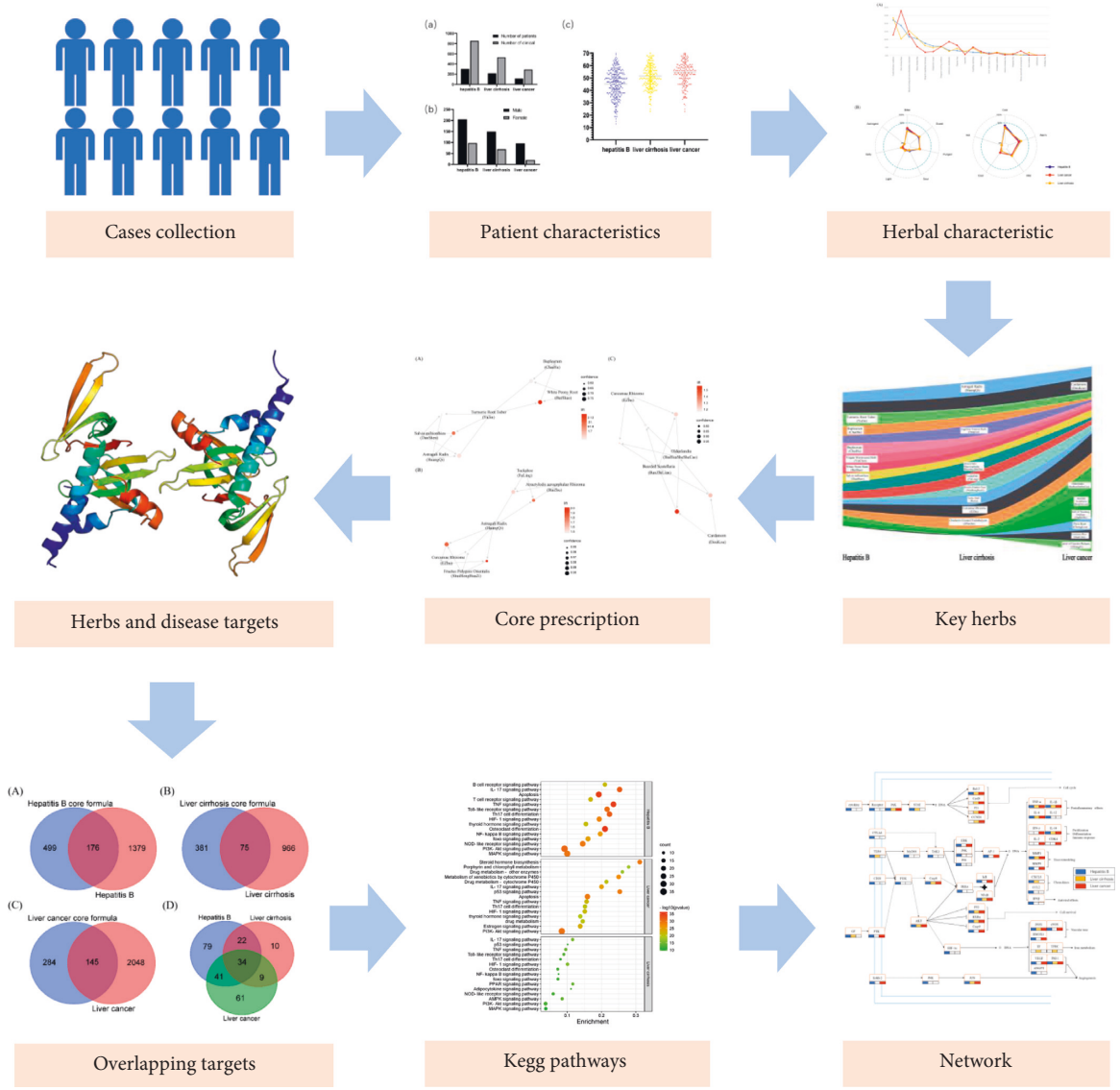



Figure 1: Workflow.

② There should be complications related to portal hypertension, such as ascites and gastric varices. Inclusion criteria included patients who met the diagnostic criteria. Exclusion criteria included ① liver cancer; ② serious cardiovascular, renal, or mental diseases or coagulation dysfunction.

The diagnostic criteria of primary liver cancer referred to the diagnostic criteria of the guidelines for diagnosis and treatment of primary liver cancer [27]. Diagnostic criteria were as follows: it should be confirmed by imaging examination, pathological examination, and clinical manifestations. Inclusion criteria included patients who met the diagnostic criteria. Exclusion criteria were as follows: ① patients with metastatic tumor; ② patients with a history of surgery; ③ patients with serious cardiovascular, renal, or mental diseases, with coagulation dysfunction, and so on.

The general clinical information and the information on Chinese herbal medicine prescriptions of these cases were collected, including the patients' age and gender and the formulae prescribed to them. The data were entered into Microsoft Office Excel, and the database was subsequently established. The herbs in the database were further standardized, and their properties, flavors, and efficacies of the herbs were completed according to the National Pharmacopoeia Commission and Chinese Materia Medica. Two researchers completed the data entry separately.

This study was reviewed by the Ethics Committee of Beijing University of Chinese Medicine, with ethical batch number 2019BZHYLL0204, which was strictly implemented in accordance with the ethical system.

*2.2. Analysis of Basic Information and the Characteristics of Prescriptions.* Microsoft Office Excel software was used to statistically analyze the distribution of patients' visits, age, and gender, and GraphPad Prism was used for visual display. Firstly, a general descriptive analysis was performed on the properties, flavors, and efficacies of the herbs in the prescriptions. Then the properties, flavors, and other aspects of the high-frequency herbs were analyzed and visually displayed so as to comprehensively show the therapeutic ideas of the prescriptions. Finally, association rules were used to screen the core herbs with strong association.

*2.2.1. Descriptive Analysis of the Properties, Flavors, and Frequency of the Herbs.* Descriptive analysis was performed on the characteristics (the properties, flavors, and categories) of the herbs. The siqi (four properties) included five items: cold, heat, warm, cool and mild, and the wuwei (five flavors) included seven items: sour, bitter, sweet, pungent, salty, astringent, and light. The herbs were divided into 21 categories according to their efficacies, which are deficiency-tonifying herbs, heat-clearing herbs, blood-activating and stasis-removing herbs, damp-draining herbs, qi-regulating herbs, exterior-relieving herbs, aromatic damp-resolving herbs, phlegm-resolving, cough-suppressing and panting-calming herbs, digestant herbs, blood-stanching herbs, wind-damp-dispelling herbs, tranquilizing herbs, liver-wind-calming herbs, astringent herbs, interior-warming herbs, purgative herbs, detoxification, parasiticide and itching-relieving herbs, worm-expelling herbs, resuscitative herbs, emetic herbs, and suppuration and granulation-promoting herbs. The top 10 core herbs in terms of frequency used in the three diseases were chosen for visual analysis of their properties, flavors, and proportion of use in all cases. RStudio 3.6.1 software and Mul-chart software were applied for analysis and visual display.

*2.2.2. Association Rules on the Screening of Core Prescriptions.* Apriori algorithm was a frequent itemset algorithm that formed the association rules, which was used to analyze the clear rules in TCM treatment and obtain the core herbs in this study. Each herb was treated as an itemset, each prescription was treated as a transaction, and they were used to find out the frequent itemset, to mine the association rules between the herbs in the prescriptions, and to discover the core prescriptions. RStudio 4.0.2 software was applied for analysis and visual display.

*2.3. Analysis of Therapeutic Mechanisms of Prescriptions*

*2.3.1. Acquisition of Core Herb Targets.* In this study, the integrative medicine pharmacology research platform TCMIP v2.0 (https://www.TCMIP.cn/TCMIP/index.php/Home/Login/login.html) [28] and the HERB database (https://herb.ac.cn/) [29] were used to gain the molecular targets of the core prescriptions and merge the obtained targets to get the relevant herb targets.

*2.3.2. Acquisition of Disease-Related Targets.* Relevant targets for the three diseases were collected via the integration of the GeneCards database (https://www.genecards.org/) and Online Mendelian Inheritance In Man (OMIM) database (https://www.omim.org/). A score ≥10 was set in the GeneCards database to select the targets of high correlation with the disease, and the obtained targets in the two databases were merged with each other to screen out the disease-related targets.

*2.4. Construction of PPI Network.* The overlapped targets were selected as the core herbs of the relevant ones for the treatment of hepatitis B, cirrhosis, and liver cancer. The target was then processed with String (https://string-db.org/), and the species was selected as "*Homo sapiens*" to derive the data of protein-protein interaction (PPI). The resulting PPI data were imported into Cytoscape 3.8 (https://www.Cytoscape.org/) for processing the PPI network diagram.

*2.4.1. Functional Annotation and Enrichment Analysis.* Common targets were acquired after overlapping the obtained herb targets, and Metascape (https://metascape.org/gp/index.html#/main/step1) was applied for KEGG pathway analysis. The data were uploaded as a multiple gene list, the species were selected as "*Homo sapiens*," and the mode was set as custom analysis, $P$ value <0.01, min overlap >3. After removing the disease-related pathways, each of the top 15 important pathways was selected, and the Microbiology Information website (https://www.bioinformatics.com.cn) was applied to plot the KEGG pathway bubble map.

*2.4.2. Construction of Core Herb Pathway Network.* Enrichment analysis was applied to get the pathways, and the pathways that were related to signal transduction or immune system were retained and summarized into a Subnetwork using the Kyoto Encyclopedia of Genes and Genomes (KEGG) online tool Mapper-Search and Color Pathway. For better presentation, the intermediate genes were hidden.

## 3. Results

*3.1. Patient Characteristics.* As shown in Figure 2, the cases with more than four herbs in a prescription were qualified and were collected in this study, involving 302 cases of hepatitis B with 205 males and 97 females, 218 cases of cirrhosis with 149 males and 69 females, and 116 cases of liver cancer with 96 males and 20 females, with an average visit of 2.8, 2.4, and 2.5, respectively. The average age of hepatitis B patients was 45.3 years, that of cirrhosis patients was 51.6 years, and that of liver cancer patients was 57.7 years. With the progression of disease, the number of patients decreased while their age increased. In terms of gender distribution, the number of male patients was higher than that of females, and the male-to-female ratios of hepatitis B, cirrhosis, and liver cancer were 2.1 : 1, 2.2 : 1, and 4.8 : 1, respectively.

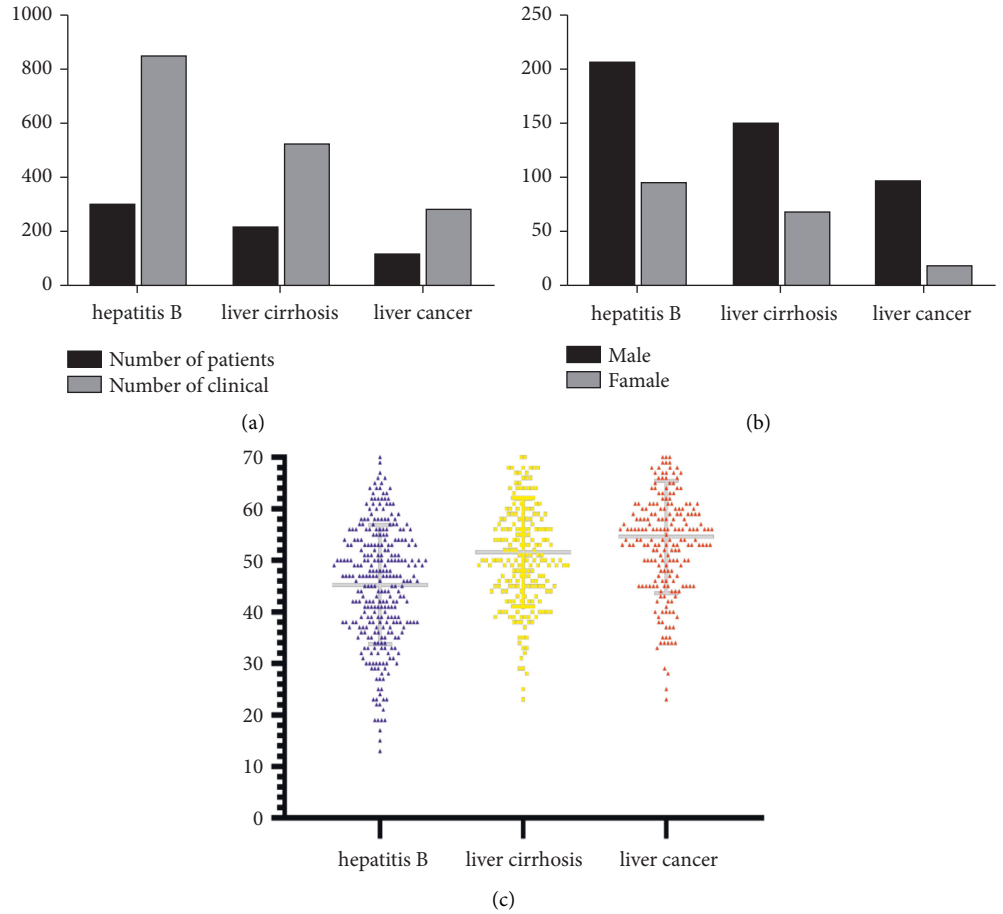
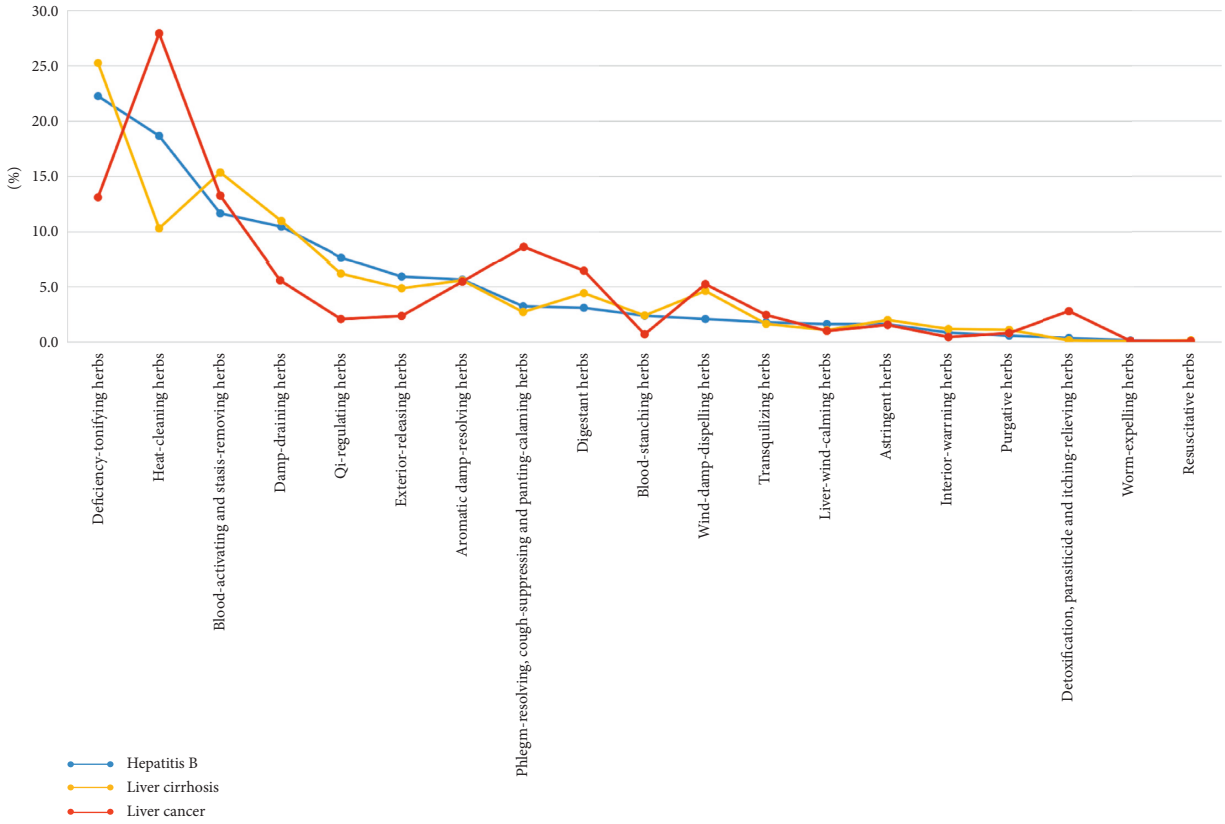



Figure 2: Basic information of patients. (a) Number of patients and visits. (b) Gender distribution of patients. (c) Age distribution of patients.

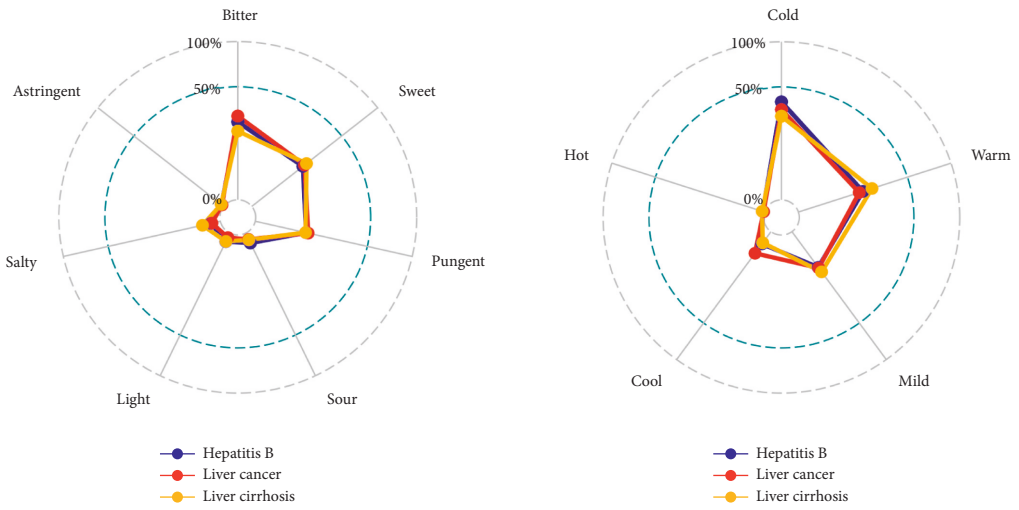
*3.2. Basic Characteristics of Herbs.* There were a total of 1,673 prescriptions, including 854 prescriptions for hepatitis B, 530 prescriptions for cirrhosis, and 289 prescriptions for liver cancer. The classification of herbs is shown in Figure 3(a), in which deficiency-tonifying herbs, heat-clearing herbs, and blood-activating and stasis-removing herbs were commonly used in the prescriptions for hepatitis B, liver cirrhosis, and liver cancer, accounting for more than 10%. In addition, it showed that, compared with the prescriptions for the other two diseases, qi-regulating herbs and exterior-resolving herbs were more commonly used in the prescriptions for hepatitis B. Deficiency-tonifying herbs and blood-activating and stasis-removing herbs were more commonly used in the prescriptions for liver cirrhosis, and heat-clearing herbs and phlegm-resolving herbs were more commonly used in the prescriptions for liver cancer. Figure 3(b) showed that the cold and warm herbs were most frequently used, followed by the mild ones. As for the flavors, bitter herbs were the most used, pungent and sweet ones were of similar total frequency, and the proportion of the characteristics and the flavors of herbs used for the three diseases were basically similar.

In this study, the top 10 herbs in frequency used in three liver diseases were further extracted, with a total of 21 herbs, and their properties, flavors, and frequency of use are shown in Figure 3(c). Generally, the cold herbs were mainly bitter cold, warm herbs were mostly sweet and warm, and mild herbs were mainly sweet and mild in this study. The proportion of medication for the three diseases was basically similar. Among them, the total herb frequency for hepatitis B was 12,491 times; the highest frequency herb in it was Turmeric Root Tuber (Yujin), with a frequency of 395 times. The total herb frequency of liver cirrhosis was 8,714 times, and its highest frequency herb was Astragali Radix (HuangQi), 301 times. And for liver cancer, its total herb frequency was 4,401 times, and the highest frequency herb was *Oldenlandia* (BaiHuaSheSheCao), 227 times.

Figure 4 shows the proportion of the top 10 frequently used herbs in each of the three diseases, in which Astragali Radix (HuangQi), Cardamom (DouKou), Turmeric Root Tuber (YuJin), Bupleurum (ChaiHu), Liquorice Root (GanCao), Virgate Wormwood Herb (YinChen), White Peony Root (BaiShao), Salvia miltiorrhiza (DanShen), and so on were frequently used in treating hepatitis B. Astragali Radix (HuangQi), Angelicae Sinensis Radix (DangGui), Atractylodis Macrocephalae Rhizoma (BaiZhu), Tuckahoe (FuLing), Curcumae Rhizome (EZhu), Fructus Polygoni Orientalis (ShuiHongHuaZi), Chicken Gizzard Linning

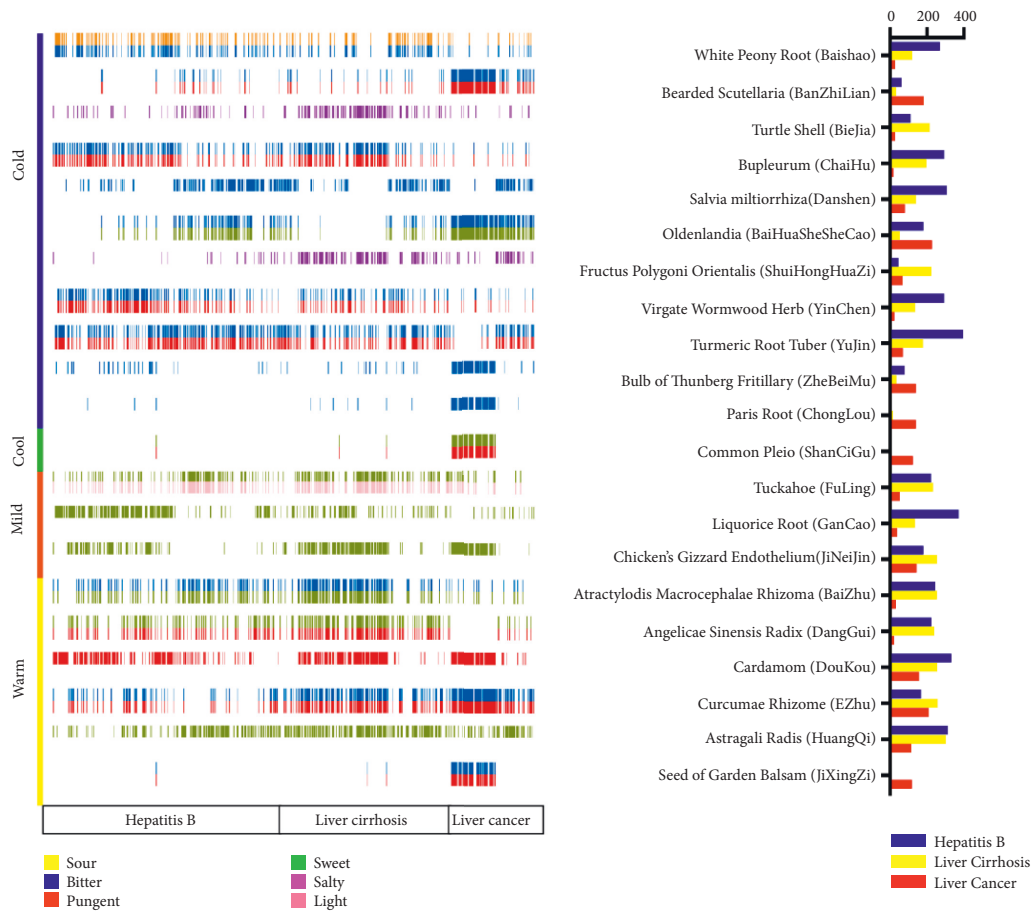



(a)




(b)

Figure 3: Continued.




(c)

Figure 3: (a) The *x*-axis represents the proportion of herbs, and the *y*-axis represents the category of herb efficacies. (b) Radar chart showed the total frequency of each property in all prescriptions. (c) The *x*-axis was the name code of patients, which was divided into hepatitis B, liver cirrhosis, and liver cancer, and the *y*-axis was the herb, which was divided into cold, cool, mild, and warm according to herbal properties. The color in the figure represented the flavors of the herbs, and the histogram corresponded to herb frequency.

(JiNeiJin), and so on were frequently used in treating liver cirrhosis. And Cardamom (DouKou), Bearded Scutellaria (BanZhiLian), *Oldenlandia* (BaiHuaSheSheCao), Curcumae Rhizome (EZhu), Paris Rhizome (ChongLou), Common Pleione Pseudobulb (ShanCiGu), and so on were frequently used in treating liver cancer.

*3.3. Association Rules Analysis.* Apriori algorithm was used in analyzing the association rules of the herbs in each of the three diseases. The data were firstly transformed by transactions, and the top 10 frequent items for each disease are listed in Figure 5. The top 5 rules of "support" associated with the most high-frequency herbs are shown in Figure 6 (the association rules are shown in attached file S1). It could be seen that, among the hepatitis B treatments, Turmeric Root Tuber (YuJin), White Peony Root (BaiShao), Bupleurum (ChaiHu), Salvia miltiorrhiza (DanShen), and Astragali Radix (HuangQi) have the strongest association. Astragali Radix (HuangQi), Tuckahoe (FuLing), Atractylodis Macrocephalae Rhizoma (BaiZhu), Princes-Feather Fruit (ShuiHongHuaZi), and Curcumae Rhizome (EZhu) have the strongest association among cirrhotic diseases. *Oldenlandia* (BaiHuaSheSheCao), Bearded Scutellaria (BanZhiLian), Curcumae Rhizome (EZhu), and Cardamom (DouKou) have the strongest association with liver cancer and are all at the core, also called the core prescriptions of the disease.

*3.4. Collection of Common Targets.* The targets of the three core prescriptions were collected on TCMIP and HERB, respectively, and 675 hepatitis B targets, 456 cirrhosis targets, and 429 liver cancer targets were obtained after deduplication. Hepatitis B, cirrhosis, and liver cancer were searched on the GeneCards and OMIM databases, respectively. Score >10 was set in GeneCards screening condition, and 1,555, 1,041, and 2,193 disease targets were acquired after intersection with the obtained targets on the OMIM database, respectively. After overlapping and deduplication of the herb and disease targets with Bioinformatics and Evolutionary Genomics (https://bioinformatics.psb.ugent.be/cgi-bin/liste/Venn) (shown in

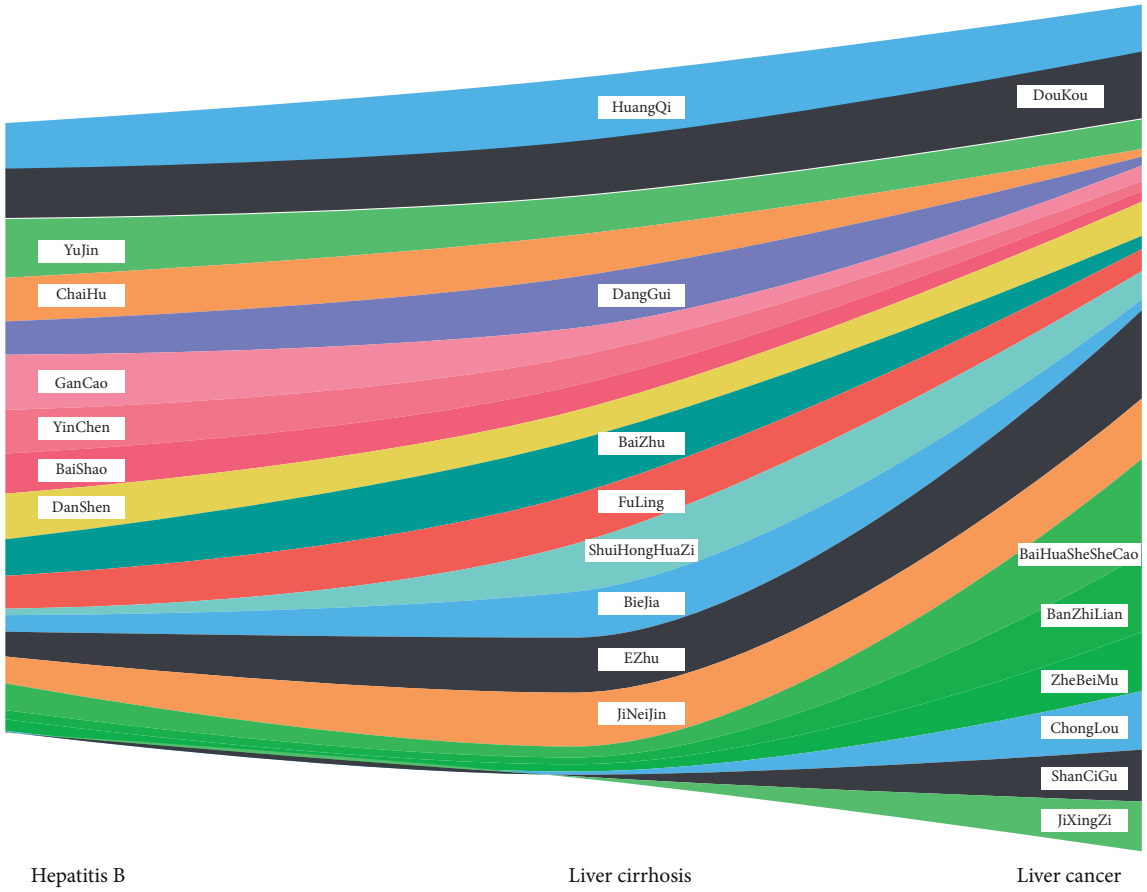



Figure 4: The $x$-axis is the disease, and the $y$-axis is the herb. The same color band represents a certain herb, and the width of the band represents the proportion of the frequency of the herb in a disease accounting for the total herb frequency of the disease.

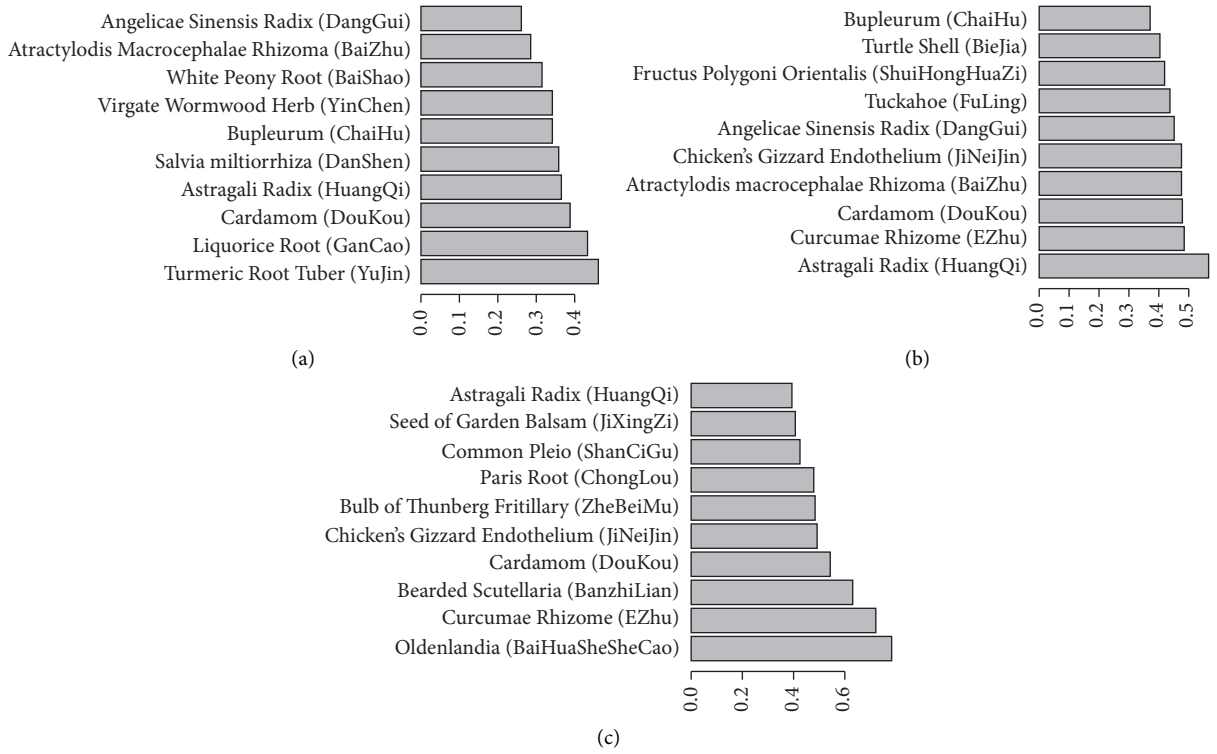



Figure 5: The $x$-axis is the support, and larger support represents a higher percentage and importance of the herbs. The first 10 frequency items of hepatitis B (a), cirrhosis (b), and liver cancer (c) are displayed, respectively.

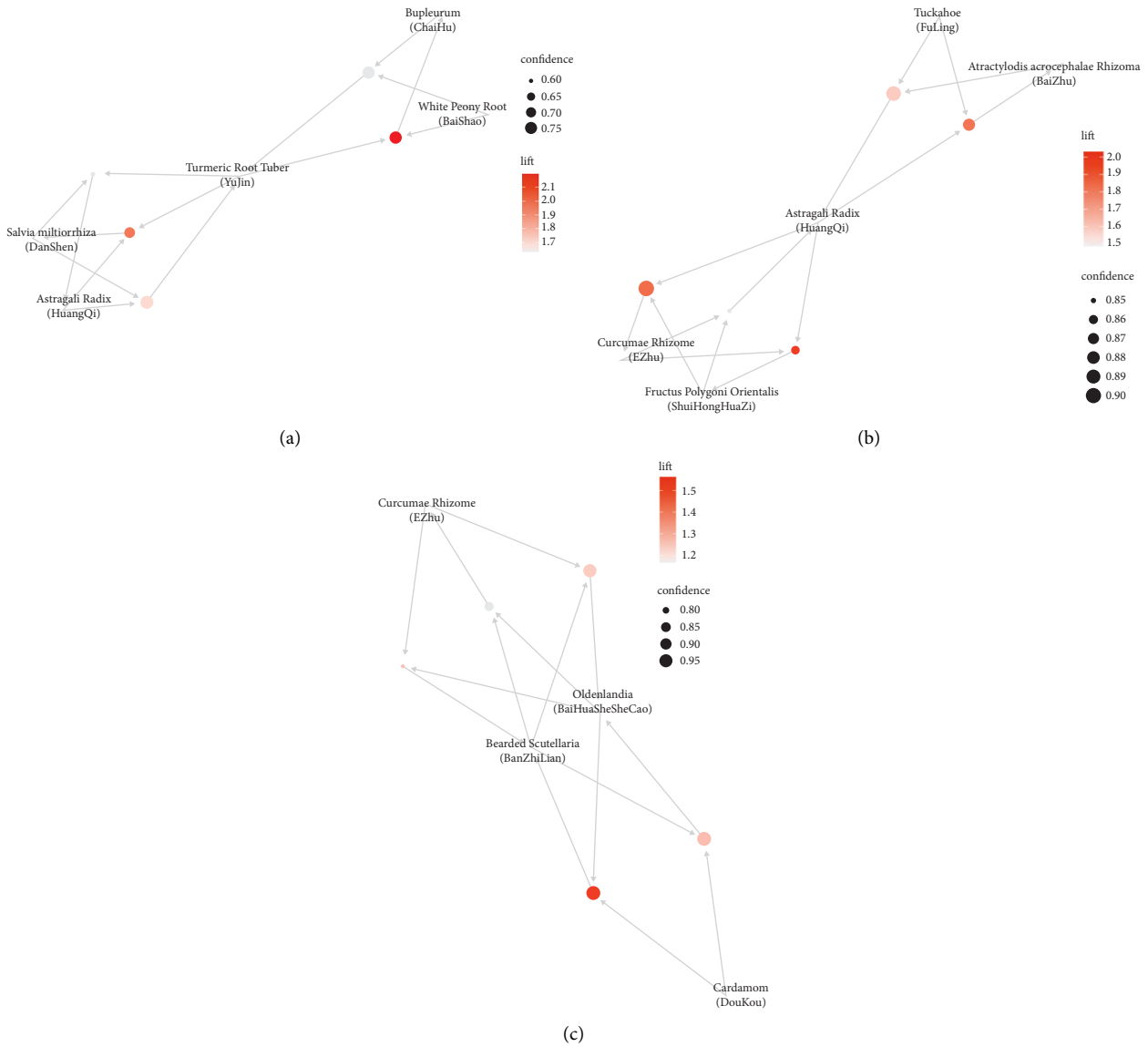



Figure 6: The network diagram of association rules in the first 5 support levels of hepatitis B (a), cirrhosis (b), and liver cancer (c), respectively.

Figure 7), 176, 75, and 145 common targets were obtained, respectively (see the attached file S2).

*3.4.1. PPI Network Diagram.* The minimum required interaction score in the String database was set to the highest confidence (0.9); after the free targets were filtered, there were 155 targets for hepatitis B, 58 targets for cirrhosis, and 126 targets for liver cancer. The data of the PPI network were then obtained, the file with "tsv" format was downloaded, and the PPI network diagram (shown in Figure 8) was obtained after processing with Cytoscape 3.8. The core genes of hepatitis B were AKT1, RELA, TP53, JUN, and STAT3, for cirrhosis were AKT1, JUN, and RXRA and for liver cancer were TP53, JUN, AKT1, and HSP90AA1.

*3.5. Functional Enrichment Analysis of Core Herbs.* KEGG enrichment of three interacting genes was carried out by Metascape (min overlap = 3, $P$ value cut-off <0.01). The diseases-related pathways in the KEGG pathway were eliminated in order to present the pathways of the core herbs on the three chronic liver diseases directly, and a total of 116 pathways were obtained with 64 pathways for hepatitis B, 68 pathways for cirrhosis, and 96 pathways for liver cancer (see the attached file S3). Important pathway categories modulated by the prescription are listed in Table 1, and signal transduction, immune system, endocrine transduction system, cell growth, and death all played a role in the three prescriptions, in which the signal transduction and immune system pathways were the most extensive and critical ones.

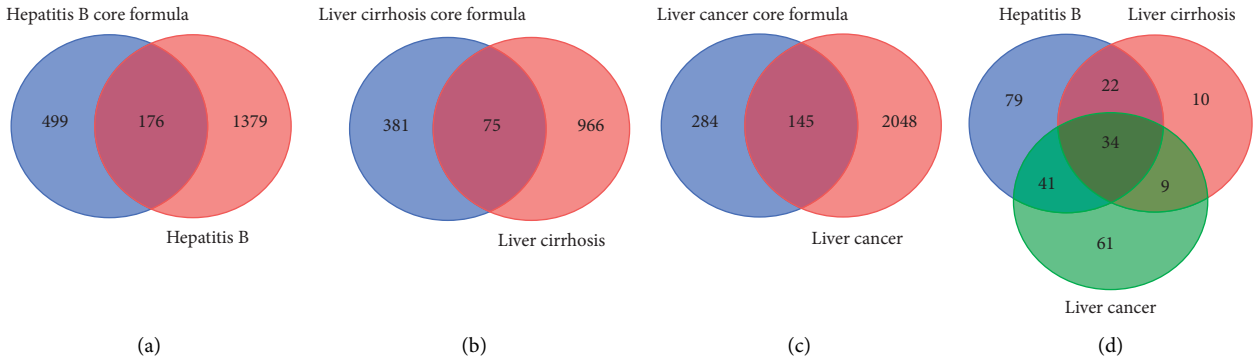



Figure 7: (a) Hepatitis B. (b) Cirrhosis. (c) Liver cancer. (d) Wayne diagram of the herb targets of three diseases and their core prescriptions.

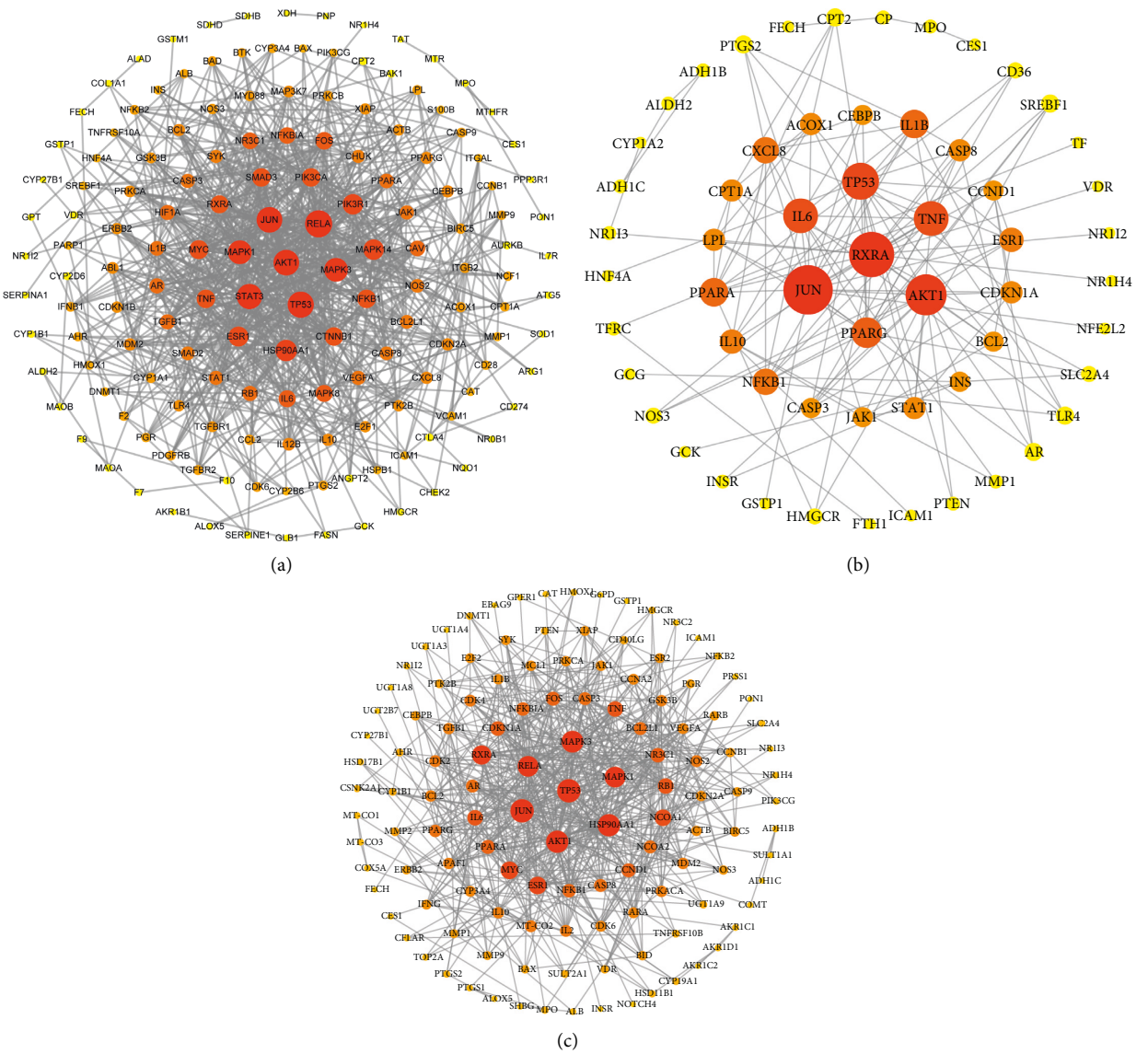



Figure 8: PPI network diagram. The graph was arranged in a circle under the condition of degree; those with a larger degree were placed in the center of the circle. (a) Hepatitis B. (b) Cirrhosis. (c) Liver cancer.

Table 1

| | Hepatitis B | Liver cirrhosis | Liver cancer |
|---|---|---|---|
| Signal transduction | 20 | 20 | 22 |
| Immune system | 15 | 12 | 16 |
| Endocrine system | 7 | 11 | 13 |
| Cell growth and death | 4 | 5 | 5 |
| Cellular community—eukaryotes | 4 | 1 | 5 |
| Nervous system | 1 | 1 | 8 |
| Other systems | 13 | 18 | 27 |

In the treatment of liver cancer, the nervous system-related pathways were more regulated than others. As for the cellular community system-related pathways, the core prescriptions of hepatitis B and liver cancer regulated more than what the core prescription of cirrhosis did.

The important pathways are shown in Figure 9. Among the top 15 important pathways, the important pathways cointervened by the core prescriptions of the three diseases were signal transduction-related pathways, including PI3K-Akt, TNF, and HIF-1 signaling pathways. The immune system-related pathways included Th17 cell differentiation and IL-17 signaling pathways.

The pathways related to signal transduction and immune system were summarized in order to explore the key pathways in detail, and some target genes that lack the upstream and downstream parts were omitted, as shown in Figure 10. TCM intervenes through a variety of ways, including the regulation of biological processes, such as cell cycle, cell survival, inflammatory response, immune response, and iron metabolism. It was found in this study that the prescription for hepatitis B played an extensive role in the related pathways, and cirrhosis prescription played a role in the middle and downstream of the pathways, mainly in inflammatory response, cell survival, cell cycle, and biological functions of iron metabolism. The liver cancer prescription mainly played a role in the middle and downstream of the pathway network, as well as in the biological functions of inflammatory response, cellular processes, angiogenesis, and vascular tension.

## 4. Discussion

Integrated analysis was used in this study to analyze the properties and flavors of herbs. In general, herbs of cold nature accounted for the greatest proportion, followed by the herbs of warm nature. The cold herbs in the prescriptions were mainly bitter in taste, and the warm herbs were mainly sweet. Salty cold herbs were slightly more prescribed in liver cirrhosis treatment. A similar proportion of herbs with other properties was prescribed for all three liver diseases. Cold herbs were related to inflammation/immunity modulation, and warm herbs affected cell growth and proliferation, reflecting that TCM herbs are widely involved in immune and cellular processes [30, 31].

In the framework of TCM theory, bitter cold herbs cleared away heat and removed pathogenic qi, and sweet warm and sweet mild herbs had a tonic effect, which suggested that removing excess substances and supplementing what lacked to restore balance were the main principle for TCM in treating the three liver diseases [32]. It can be seen from the classification of the efficacy of herbs that tonifying deficiency, clearing away heat and toxin, and activating blood circulation to remove blood stasis were the three main therapeutic approaches to liver diseases, regulating qi was emphasized in the treatment of hepatitis B, tonifying deficiency and activating blood circulation were often used for liver cirrhosis, and liver cancer was mainly treated by clearing away heat and toxin and resolving phlegm based on TCM theory. Then, the core prescriptions for the three diseases were obtained through association rules analysis. The prescriptions of the three diseases were different in that hepatitis B prescription was mainly to invigorate the spleen and soothe the liver, the liver cirrhosis prescription was mainly to invigorate qi and activate blood, and the liver cancer prescription was to remove cancer toxins while invigorating the spleen and activate blood. The three therapeutics above were the core treatments for these diseases, which were different from those in the current guidelines. The current guidelines, therapeutics, and prescriptions for the three diseases are listed in Table 2. And it is the approach of traditional Chinese medicine to treat diseases on the basis of one therapeutics while applying multiple therapeutics comprehensively [33]. There is a close relationship between hepatitis B, liver cirrhosis, and hepatocellular carcinoma. In a certain sense, they are recognized as one disease in certain stages of progression. TCM believes that patients are with more severe lesions in the blood phase than in the qi phase. The results of the analysis of patient characteristics show that, with the progression of the disease and increase of the patients' age, the disease becomes worse, which goes from the qi phase deep to the blood phase, showing that the disease is getting worse. TCM believes that [34] damp heat epidemic pathogenic factor and deficiency of vital qi are the key factors in the pathogenesis of hepatitis B, and the location of the disease is in the qi phase (shallow level). Blood stasis and phlegm are the key factors in the progression of hepatitis B into liver cirrhosis and liver cancer, and the location of the disease is in the blood phase (deep level). Compared with liver cirrhosis, phlegm and blood stasis transforming into toxin accumulation is the key to the formation of liver cancer [35], which is consistent with the core prescription treatment principles in this study.

DNA of hepatitis B virus participates directly in the onset of liver cancer through DNA integration of viral oncoproteins. At the same time, hepatocellular death and inflammatory infiltration caused by long-term infection could

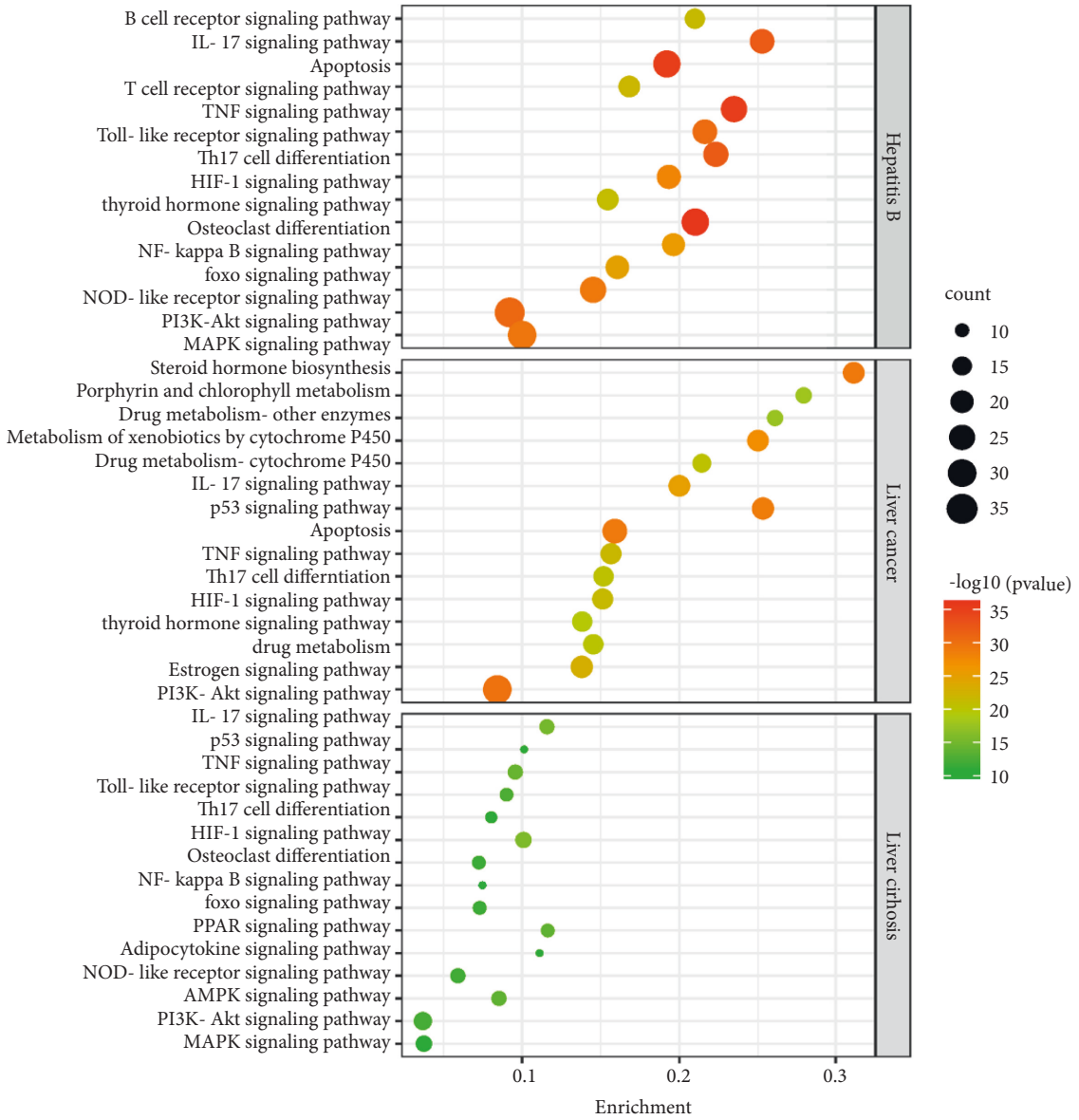



Figure 9: KEGG enrichment bubble plot presentation. Each bubble represents a KEGG process. Enrichment indicates the proportion of the target in the list of targets. The size of bubble is positively correlated with the number of targets in each pathway; the larger the bubble is, the richer the gene is. −log10 indicates *P* value, the greener the color is, the larger *P* value is, and a redder color indicates a smaller *P* value. The top 15 KEGG for *P* values are shown above.

accelerate liver cell renewal, promote mutation accumulation, and cause liver tumors indirectly [38]. When chronic liver diseases worsened into cirrhosis, changes such as inflammation, stellate cell activation accompanied by fibrosis, angiogenesis accompanied by liver microvascular changes, and others might happen, which would increase the risk of liver cancer [7, 39, 40].

As an important factor in liver disease progression, chronic inflammation and cell growth and apoptosis are involved in the whole process of hepatitis B, liver cirrhosis, and liver cancer. The pathways related to signal transduction and the immune system were mostly intervened in the core prescriptions of the three diseases, which intervened jointly in the PI3K-Akt signaling pathway, TNF signaling pathway, HIF-1 signaling pathway, Th17 cell differentiation, and IL-17 signaling pathways, and were involved in immune regulation, inflammation, cell cycle and other important roles [41–43]. It could be seen from the PPI network interaction diagram that AKT1 was the common key target of the three prescriptions, was widely expressed in liver, which affected cell survival, proliferation, and migration when activated, and was closely related to angiogenesis [44, 45]. The PI3K-Akt signaling pathway was jointly regulated by the core prescriptions of the three diseases, which was an important inflammatory regulatory pathway, involved in cell survival and cell growth, and had a key role in tumorigenesis and development [46–48]. The hepatitis B virus core protein promoted the occurrence of liver cancer after activating the

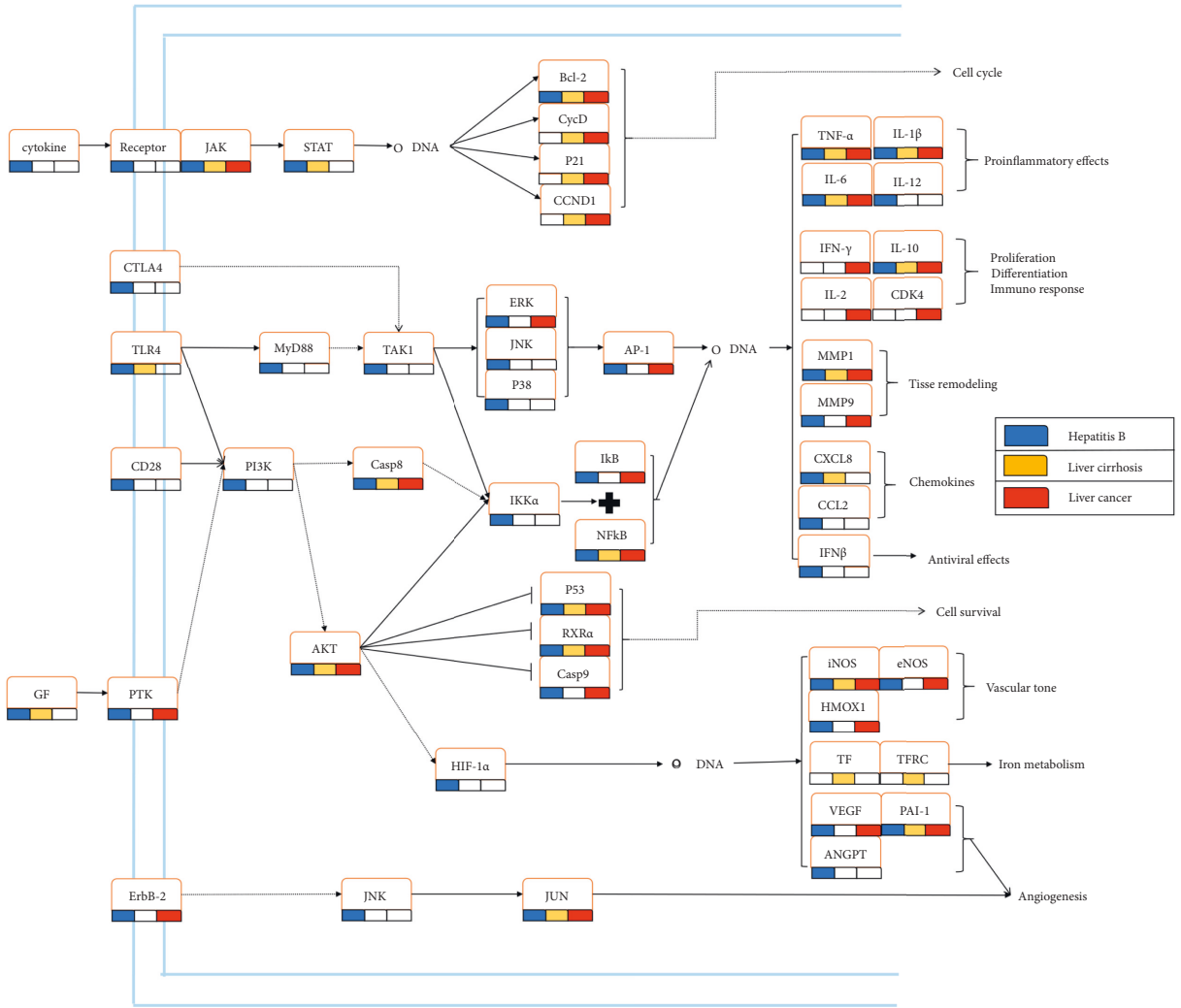



Figure 10: The gene regulatory effects of the core prescriptions of the three liver diseases. Light blue lines indicate cell membrane, and gene names are shown in rectangles. The markers below the genes show the aiming pattern of the three core prescriptions. Genes that lacked the upstream and downstream parts were omitted and indicated by dashed arrows.

Table 2

| Disease | Source | Treatment principle | Prescription |
|---|---|---|---|
| Hepatitis B | The clinical guidelines of diagnosis and treatment of chronic hepatitis B with traditional Chinese medicine [36] | Replenish the spleen and soothe the liver, nourish the liver and kidney, promote blood circulation and free the collaterals, and clear heat and drain dampness | Yinchenhao Decoction, Xiaoyao Powder, Yiguan Decoction, and Gexia Zhuyu Decoction |
| Liver cirrhosis | Chinese guidelines on the management of liver cirrhosis [26] | Strengthen qi and promote blood circulation, free the network vessels and resolve the stasis, and clear heat and drain dampness | Anluo Huaxian Pill, Fuzheng Huayu Capsule, Compound Biejiaruangan Troche, and so on |
| Liver cancer | Standardization for diagnosis and treatment of hepatocellular carcinoma [37] | Clear the cancer toxin, replenish the spleen and promote blood circulation, soothe the liver, and nourish the blood | Xiaoyao Powder, Guipi Decoction, Huaier Granule, Cinobufagin, Jiedu Granule, and so on |

Src/PI3K/Akt pathway [49]. The largest surface antigen encoded by HBV is the large hepatitis B virus (LHBs) glycoprotein. Studies have shown that LHBs activate the Src/PI3K/Akt signaling pathway through proximal stimulation of the PKC$\alpha$/Raf1 signaling pathway, which could promote the tumorigenesis of hepatoma cells [50]. The hepatitis B

virus *X* protein-binding protein (HBXIP) may promote the proliferation and migration of liver cancer cells through the PI3K/AKT signaling pathway [51]. When there was continuous inflammation, the production of a large number of cytokines would activate the FAK-PI3K-Akt-P70 signaling cascades, result in the proliferation of HSCs, and accelerate the fibrotic response [52]. Turmeric Root Tuber (YuJin) was the most important herb in hepatitis B prescription, and its main ingredient, the essential oil of Rhizoma curcumae (EORC), significantly affected the PI3K/AKT pathway and effectively inhibited the progression of liver fibrosis [53]. Astragali Radix (HuangQi) was important in the prescriptions for hepatitis B and liver cirrhosis. Reports showed that Astragaloside IV, the main component contained in Astragali Radix (HuangQi), could improve liver damage through the PI3K/Akt pathway [54]. *Oldenlandia* (BaiHuaSheSheCao) and Bearded Scutellaria (BanZhiLian) were important herbs in the core prescription of liver cancer. They significantly induced apoptosis of cancer cells by regulating PI3K/Akt and effectively inhibited tumor growth in vitro and in vivo [55].

Apart from regulating immunity and inhibiting inflammation, the prescription for hepatitis B had a regulatory effect on the endocrine system as well, which might be an important mechanism for preventing the progression and degeneration of hepatitis B. Compared with hepatitis B prescription, liver cirrhosis prescription regulated more endocrine system pathways. Studies have found that growth hormone resistance and low insulin-like growth factor-1, thyroid dysfunction, and other problems are common in patients with liver cirrhosis, which will affect negatively its prognosis [56]. It has been found that obesity, diabetes, and fatty liver are independently associated with the increased risk of hepatocellular carcinoma [57]. The prescriptions for liver cirrhosis had a regulatory effect on the key molecules RXR and PPAR in the PPAR signaling pathway, which was involved in lipid regulation and glucose metabolism. Actively regulating hormones and maintaining stable glucose and lipid metabolism could effectively protect liver tissue and prevent liver cancer. Compared with the other two prescriptions, the liver cancer prescription regulated the most extensive pathways and regulated more cell processes and endocrine system-related pathways as well. In addition, it had a prominent regulatory effect on the nervous system. It has been found clinically that neural-related factors could significantly affect the prognosis of liver cancer, which have received increasing attention in these years [58], while the methods for liver tumors and nervous system treatment are still limited [59], for which TCM could be an important supplement.

It was found from the subnetworks (Figure 10) compiled in this study that the three prescriptions for the three liver diseases interfered extensively with the inflammatory process, and the proinflammatory cytokines IL-6, IL-1$\beta$, and TNF$\alpha$, as well as the anti-inflammatory cytokine IL-6 and its upstream factor NF-$\kappa$B, were important targets that were regulated by the three prescriptions [60]. NF-$\kappa$B could be activated in almost all chronic liver diseases [61–63], which has a wide range of functions in various cell compartments, including liver cell survival, inflammation in Kupffer cells, and HSC activation, and is at the key position in chronic liver diseases and cellular wound repair response [64]. Compared with the other two prescriptions, the prescription for hepatitis B regulated more proinflammatory factors and chemokines. Another interesting finding was that hepatitis B regulated angiogenesis and VEGF pathways, while to the best of our knowledge, there has been no report on treating hepatitis B through angiogenesis intervention. Hepatocellular carcinoma is a highly vascularized tumor characterized by active neovascularization [65]. Intervention in angiogenesis before hepatitis B worsening into liver cancer might be effective in preventing the occurrence of liver tumors. The prescription for liver cirrhosis could additionally act in the biological function of iron metabolism; while there have been few studies on the application of iron metabolism in treating liver diseases, neither has this topic received enough attention. As liver is the main organ for iron storing, it is particularly vulnerable to damage from iron overload. Excessive iron produces free radicals, aggravates cell damage, leads to fibrosis, and accelerates liver cirrhosis worsening into liver cancer [66, 67]. As iron overload has often been observed in chronic liver diseases, iron metabolism would become a potential therapeutic target [68].

## 5. Innovation and Limitations

This research applied a combined method of data mining and integrated network pharmacology and explored the prescription rules and mechanism of action of TCM in the treatment of various chronic liver diseases. To the best of our knowledge, this study is the first of its kind in TCM to analyze various related diseases jointly, which has created a new paradigm in methodology for the analysis of related diseases.

With the application of this method, the core prescriptions and the herbs involved for the three liver diseases were identified, which mainly intervened in the process of chronic hepatitis B, cirrhosis, and liver cancer through the intervention of signal transduction and immune system-related pathways and regulated the immune and cell growth processes, in which the PI3K-Akt was an important signaling pathway of it. Hepatitis B prescription was helpful for angiogenesis regulation, which might be the reason why it prevented the disease progression. The prescriptions for liver cirrhosis and liver cancer interfered more with the endocrine system pathways, and the liver cancer prescription had a broader regulating effect on the nervous system. The findings may cast some light on the treatment of chronic liver diseases as well as prevention of liver cancer using TCM. The prescriptions in this study came from two professors in outpatient clinics; even though there was a lack of evaluation on the efficacy of prescriptions and the results of the analysis may be biased, the two professors are prestigious Chinese medicine experts who boast leading positions in the field of liver disease in China. Their prescriptions have been an important source of learning for doctors in the hepatology department of traditional Chinese medicine, and thus, it is still of great significance to analyze their

prescriptions. In addition, the source of the cases would be expanded in future studies, and further experimental studies are required for the evaluation and validation of the action mechanisms of the core prescriptions.

## 6. Conclusion

Three core prescriptions for chronic liver disease treatment were identified in this study, and it was found that they mainly functioned through regulating signal transduction and immune system-related pathways. Compared with the other two liver diseases, the core prescription for liver cancer regulated more neural system-related pathways.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Xiaohong Gu designed, supervised, and sponsored the research. He Yu designed and supervised the research and drafted the manuscript. Zhendong Wang cleaned and analyzed data and drafted the manuscript. Yong Zhang and Qiuyun Zhang collected the raw data and drafted the manuscript. Qiang Ao collected the raw data. Xueyi Ge and Ying Qian collected the clinical data. Changyong Luo, Bochuan Wang, and Chen Bai contributed to data processing. Yuhan Wang and Jing Wang embellished the pictures. All the authors approved the final edited version of the manuscript. Zhendong Wang, Yong Zhang, and Qiuyun Zhang have contributed equally to this work and share the first authorship.

## Acknowledgments

This study was financially supported by the National Key R&D Program of China (2018YFC1704100 and 2018YFC1704101). The authors gratefully acknowledge the data assistance of Changjian Yin and also present heartfelt thanks to Tingyao Hu and Shuiyuan Dai for language modification and to all the patients for their participation.

## Supplementary Materials

S1: The important association rules of the prescriptions of the three diseases were obtained via the Apriori algorithm. S2: The targets of the core herbs of hepatitis B, cirrhosis, and liver cancer. S3: KEGG enrichment analysis for the three core herbs of the three diseases. (*Supplementary Materials*)

## References

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

[2] J. L. Petrick, A. A. Florio, A. Znaor et al., "International trends in hepatocellular carcinoma incidence, 1978-2012," *International Journal of Cancer*, vol. 147, no. 2, pp. 317–330, 2020.

[3] L. Rahib, B. D. Smith, R. Aizenberg, A. B. Rosenzweig, J. M. Fleshman, and L. M. Matrisian, "Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States," *Cancer Research*, vol. 74, no. 11, pp. 2913–2921, 2014.

[4] T. Akinyemiju, S. Abera, M. Ahmed et al., "The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level: results from the global burden of disease study 2015," *JAMA Oncology*, vol. 3, no. 12, pp. 1683–1691, 2017.

[5] S. Affo, L. X. Yu, and R. F. Schwabe, "The role of cancer-associated fibroblasts and fibrosis in liver cancer," *Annual Review of Pathology: Mechanisms of Disease*, vol. 12, no. 1, pp. 153–186, 2017.

[6] Y. Xie, "Hepatitis B virus-associated hepatocellular carcinoma," *Advances in Experimental Medicine and Biology*, vol. 1018, pp. 11–21, 2017.

[7] E. A. Tsochatzis, J. Bosch, and A. K. Burroughs, "Liver cirrhosis," *The Lancet*, vol. 383, no. 9930, pp. 1749–1761, 2014.

[8] A. Villanueva, "Hepatocellular carcinoma," *New England Journal of Medicine*, vol. 380, no. 15, pp. 1450–1462, 2019.

[9] D. J. Newman and G. M. Cragg, "Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019," *Journal of Natural Products*, vol. 83, no. 3, pp. 770–803, 2020.

[10] B. Hu, S. S. Wang, and Q. Du, "Traditional Chinese medicine for prevention and treatment of hepatocarcinoma: from bench to bedside," *World Journal of Hepatology*, vol. 7, no. 9, pp. 1209–1232, 2015.

[11] C. Q. Ling, J. Fan, H. S. Lin et al., "Clinical practice guidelines for the treatment of primary liver cancer with integrative traditional Chinese and Western medicine," *Journal of Integrative Medicine*, vol. 16, no. 4, pp. 236–248, 2018.

[12] M. C. Wu, "Traditional Chinese medicine in prevention and treatment of liver cancer: function, status and existed problems," *Journal of Chinese Integrative Medicine*, vol. 1, no. 3, pp. 163-164, 2003.

[13] X. Liu, M. Li, X. Wang et al., "Effects of adjuvant traditional Chinese medicine therapy on long-term survival in patients with hepatocellular carcinoma," *Phytomedicine*, vol. 62, Article ID 152930, 2019.

[14] L. Ma, B. Wang, Y. Long, and H. Li, "Effect of traditional Chinese medicine combined with Western therapy on primary hepatic carcinoma: a systematic review with meta-analysis," *Frontiers of Medicine*, vol. 11, no. 2, pp. 191–202, 2017.

[15] Z. J. Hou, J. H. Zhang, X. Zhang et al., "Long-term traditional Chinese medicine combined with na antiviral therapy on cirrhosis incidence in chronic hepatitis B patients in the real-world setting: a retrospective study," *Evidence-Based Complementary and Alternative Medicine*, vol. 2020, Article ID 3826857, 8 pages, 2020.

[16] T. Y. Tsai, H. Livneh, T. H. Hung, I. H. Lin, M. C. Lu, and C. C. Yeh, "Associations between prescribed Chinese herbal medicine and risk of hepatocellular carcinoma in patients with chronic hepatitis B: a nationwide population-based cohort study," *BMJ Open*, vol. 7, no. 1, Article ID e014571, 2017.

[17] X. F. Zhai, X. L. Liu, F. Shen, J. Fan, and C. Q. Ling, "Traditional herbal medicine prevents postoperative recurrence of small hepatocellular carcinoma: a randomized controlled study," *Cancer*, vol. 124, no. 10, pp. 2161–2168, 2018.

[18] H. Li, "Advances in anti hepatic fibrotic therapy with Traditional Chinese Medicine herbal formula," *Journal of Ethnopharmacology*, vol. 251, Article ID 112442, 2020.

[19] S. Dong, Q. L. Chen, and S. B. Su, "Curative effects of fuzheng huayu on liver fibrosis and cirrhosis: a meta-analysis," *Evidence-Based Complementary and Alternative Medicine*, vol. 2015, Article ID 125659, 11 pages, 2015.

[20] C. Luo, H. Yu, T. Yang et al., "Data mining and systematic pharmacology to reveal the mechanisms of traditional Chinese medicine in recurrent respiratory tract infections' treatment," *Evidence-Based Complementary and Alternative Medicine*, vol. 2020, Article ID 8979713, 13 pages, 2020.

[21] Y. L. Chen, C. Y. Lee, K. H. Huang, Y. H. Kuan, and M. Chen, "Prescription patterns of Chinese herbal products for patients with sleep disorder and major depressive disorder in Taiwan," *Journal of Ethnopharmacology*, vol. 171, pp. 307–316, 2015.

[22] J. Cao, "The common prescription patterns based on the hierarchical clustering of herb-pairs efficacies," *Evidence-Based Complementary and Alternative Medicine*, vol. 2016, Article ID 6373270, 7 pages, 2016.

[23] P. Xia, K. Gao, J. Xie et al., "Data mining-based analysis of Chinese medicinal herb formulae in chronic kidney disease treatment," *Evidence-Based Complementary and Alternative Medicine*, vol. 2020, Article ID 9719872, 14 pages, 2020.

[24] H. Yuan, Q. Ma, H. Cui et al., "How can synergism of traditional medicines benefit from network pharmacology?" *Molecules*, vol. 22, no. 7, p. 1135, 2017.

[25] G. Q. Wang, F. S. Wang, H. Zhuang, T. S. Li, S. J. Zheng, and H. Zhao, "Guidelines for the prevention and treatment of chronic hepatitis B," *Chinese Journal of Viral Diseases*, vol. 10, no. 1, pp. 1–25, 2020.

[26] X. Y. Xu, H. G. Ding, W. G. Li et al., "Chinese guidelines on the management of liver cirrhosis," *Journal of Practical Hepatology*, vol. 22, no. 6, pp. 770–786, 2019.

[27] Bureau of Medical Administration, "Guidelines for diagnosis and treatment of primary liver cancer," *Infectious Disease Information*, vol. 33, no. 6, pp. 481–500, 2020.

[28] H. Y. Xu, Y. Q. Zhang, Z. M. Liu et al., "ETCM: an encyclopaedia of traditional Chinese medicine," *Nucleic Acids Research*, vol. 47, pp. D976–d982, 2019.

[29] S. Fang, L. Dong, L. Liu et al., "HERB: a high-throughput experiment- and reference-guided database of traditional Chinese medicine," *Nucleic Acids Research*, vol. 49, pp. D1197–d1206, 2021.

[30] F. Liang, L. Li, M. Wang et al., "Molecular network and chemical fragment-based characteristics of medicinal herbs with cold and hot properties from Chinese medicine," *Journal of Ethnopharmacology*, vol. 148, no. 3, pp. 770–779, 2013.

[31] X. Liao, Y. Bu, and Q. Jia, "Traditional Chinese medicine as supportive care for the management of liver cancer: past, present, and future," *Genes & Diseases*, vol. 7, no. 3, pp. 370–379, 2020.

[32] W. Wang and W. Du, "On the theory of four properties and five flavors and their corelations," *Chinese Journal of Ethnopharmacy*, vol. 26, no. 24, pp. 61–63, 2017.

[33] Y. R. Xing, "Research on the relationship between TCM diseases and syndromes," *China Journal of Traditonal Chinese Medicine and Pharmacy*, vol. 33, no. 12, pp. 5290–5294, 2018.

[34] J. Dong, J. Zhang, and H. Zhang, "The application of the idea of "preventive treatment of disease" in the prevention and treatment of hepatitis B," *Journal of New Chinese Medicine*, vol. 44, no. 6, pp. 179-180, 2012.

[35] G. D. Tong, Y. X. Wang, Y. F. Xing, C. S. Wei, H. H. Tang, and J. Z. Huang, "On the treatment of chronic hepatitis B and related liver cirrhosis and liver cancer from the aspect of "deficiency, toxin, blood stasis and accumulation"," *Journal of Traditional Chinese Medicine*, vol. 62, no. 16, pp. 1404–1407, 2021.

[36] Medicine HSCoCAoC and Minorities LDSCoCMAo, "The clinical guidelines of diagnosis and treatment of chronic hepatitis B with traditonal Chinese medicine," *Journal of Clinical Hepatol*, vol. 34, no. 12, pp. 2520–2525, 2018.

[37] Bureau of Medical Administration, "Standardization for diagnosis and treatment of hepatocellular carcinoma," *Chinese Journal of Practical Surgery*, vol. 42, no. 3, pp. 241–273, 2022.

[38] S. L. Chan, V. W. Wong, S. Qin, and H. L. Chan, "Infection and cancer: the case of hepatitis B," *Journal of Clinical Oncology*, vol. 34, no. 1, pp. 83–90, 2016.

[39] M. Ringehan, J. A. McKeating, and U. Protzer, "Viral hepatitis and liver cancer," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, Article ID 20160274, 2017.

[40] Y. Nakamoto and S. Kaneko, "Mechanisms of viral hepatitis induced liver injury," *Current Molecular Medicine*, vol. 3, no. 6, pp. 537–544, 2003.

[41] C. Corrado and S. Fontana, "Hypoxia and HIF signaling: one Axis with divergent effects," *International Journal of Molecular Sciences*, vol. 21, no. 16, p. 5611, 2020.

[42] G. Chen and D. V. Goeddel, "TNF-R1 signaling: a beautiful pathway," *Science*, vol. 296, no. 5573, pp. 1634-1635, 2002.

[43] N. Amatya, A. V. Garg, and S. L. Gaffen, "IL-17 signaling: the Yin and the yang," *Trends in Immunology*, vol. 38, no. 5, pp. 310–322, 2017.

[44] S. Revathidevi and A. K. Munirajan, "Akt in cancer: mediator and more," *Seminars in Cancer Biology*, vol. 59, pp. 80–91, 2019.

[45] A. Santel, J. Ribera, W. Jiménez, and M. Morales-Ruiz, "The role of akt in chronic liver disease and liver regeneration," *Seminars in Liver Disease*, vol. 37, no. 1, pp. 011–016, 2017.

[46] E. C. Lien, C. C. Dibble, and A. Toker, "PI3K signaling in cancer: beyond AKT," *Current Opinion in Cell Biology*, vol. 45, pp. 62–71, 2017.

[47] M. Aoki and T. Fujishita, "Oncogenic roles of the PI3K/AKT/mTOR Axis," *Current Topics in Microbiology and Immunology*, vol. 407, pp. 153–189, 2017.

[48] I. Pal and M. Mandal, "PI3K and Akt as molecular targets for cancer therapy: current clinical outcomes," *Acta Pharmacologica Sinica*, vol. 33, no. 12, pp. 1441–1458, 2012.

[49] W. Liu, T. F. Guo, Z. T. Jing et al., "Hepatitis B virus core protein promotes hepatocarcinogenesis by enhancing Src expression and activating the Src/PI3K/Akt pathway," *The FASEB Journal*, vol. 32, no. 6, pp. 3033–3046, 2018.

[50] H. Liu, J. Xu, L. Zhou et al., "Hepatitis B virus large surface antigen promotes liver carcinogenesis by activating the Src/PI3K/Akt pathway," *Cancer Research*, vol. 71, no. 24, pp. 7547–7557, 2011.

[51] Y. Liu, J. Wei, B. Xiao, and Q. Zeng, "Effect of hepatitis B X-interacting protein on proliferation and migration of hepatoma cells by regulating PI3 K/Akt signaling pathway," *Chinese Journal of Immunology*, vol. 36, no. 7, pp. 837–841, 2020.

[52] C. J. Parsons, M. Takashima, and R. A. Rippe, "Molecular mechanisms of hepatic fibrogenesis," *Journal of Gastroenterology and Hepatology*, vol. 22, pp. S79–S84, 2007.

[53] Y. Chen, W. Liao, Z. Zhu et al., "Essential oil from the raw and vinegar-processed Rhizoma Curcumae ameliorate CCl4-induced liver fibrosis: integrating network pharmacology and molecular mechanism evaluation," *Food & Function*, vol. 12, no. 9, pp. 4199–4220, 2021.

[54] R. Wei, H. Liu, R. Chen, Y. Sheng, and T. Liu, "Astragaloside IV combating liver cirrhosis through the PI3K/Akt/mTOR signaling pathway," *Experimental and Therapeutic Medicine*, vol. 17, no. 1, pp. 393–397, 2019.

[55] L. Lu, S. Zhan, X. Liu, X. Zhao, X. Lin, and H. Xu, "Antitumor effects and the compatibility mechanisms of herb pair sclerotrion diffusum (willd.) R. J. Wang-sculellaria barbata D. Don," *Frontiers in Pharmacology*, vol. 11, p. 292, 2020.

[56] A. Eshraghian and S. A. Taghavi, "Systematic review: endocrine abnormalities in patients with liver cirrhosis," *Archives of Iranian Medicine*, vol. 17, no. 10, pp. 713–721, 2014.

[57] G. A. Michelotti, M. V. Machado, and A. M. Diehl, "NAFLD, NASH and liver cancer," *Nature Reviews Gastroenterology & Hepatology*, vol. 10, no. 11, pp. 656–665, 2013.

[58] S. Jeong, B. Zheng, H. Wang, Q. Xia, and L. Chen, "Nervous system and primary liver cancer," *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1869, no. 2, pp. 286–292, 2018.

[59] R. Kazi and Y. L. Bunimovich, "The nervous system: a new target in the fight against cancer," *Anti-Cancer Drugs*, vol. 29, no. 10, pp. 929–934, 2018.

[60] B. B. Aggarwal, S. Shishodia, S. K. Sandur, M. K. Pandey, and G. Sethi, "Inflammation and cancer: how hot is the link?" *Biochemical Pharmacology*, vol. 72, no. 11, pp. 1605–1621, 2006.

[61] J. K. Reddy and M. Sambasiva Rao, "Lipid metabolism and liver inflammation. II. Fatty liver disease and fatty acid oxidation," *American Journal of Physiology—Gastrointestinal and Liver Physiology*, vol. 290, no. 5, pp. G852–G858, 2006.

[62] D. I. Tai, S. L. Tsai, Y. M. Chen et al., "Activation of nuclear factor kappaB in hepatitis C virus infection: implications for pathogenesis and hepatocarcinogenesis," *Hepatology*, vol. 31, no. 3, pp. 656–664, 2000.

[63] A. Kosters and S. J. Karpen, "The role of inflammation in cholestasis: clinical and basic aspects," *Seminars in Liver Disease*, vol. 30, no. 2, pp. 186–194, 2010.

[64] T. Luedde and R. F. Schwabe, "NF-$\kappa$B in the liver--linking injury, fibrosis and hepatocellular carcinoma," *Nature Reviews Gastroenterology & Hepatology*, vol. 8, no. 2, pp. 108–118, 2011.

[65] K. Vrancken, J. Paeshuyse, and S. Liekens, "Angiogenic activity of hepatitis B and C viruses," *Antiviral Chemistry & Chemotherapy*, vol. 22, no. 4, pp. 159–170, 2012.

[66] A. Pietrangelo, "Iron and the liver," *Liver International*, vol. 36, pp. 116–123, 2016.

[67] P. Sorrentino, S. D'Angelo, U. Ferbo, P. Micheli, A. Bracigliano, and R. Vecchione, "Liver iron excess in patients with hepatocellular carcinoma developed on non-alcoholic steato-hepatitis," *Journal of Hepatology*, vol. 50, no. 2, pp. 351–357, 2009.

[68] K. J. Mehta, S. J. Farnaud, and P. A. Sharp, "Iron and liver fibrosis: mechanistic and clinical aspects," *World Journal of Gastroenterology*, vol. 25, no. 5, pp. 521–538, 2019.

Hindawi
Journal of Oncology
Volume 2022, Article ID 3528142, 18 pages
https://doi.org/10.1155/2022/3528142

*Research Article*

# Characteristic of Molecular Subtypes in Lung Squamous Cell Carcinoma Based on Autophagy-Related Genes and Tumor Microenvironment Infiltration

**Jinjie Wang,[1,2] Jiaqi Zhu,[1,2] Yijie Tang,[1,2] Anping Zhang,[1,2] Tingting Zhou,[1,2] Youlang Zhou,[3] and Jiahai Shi[1,2,4]**

[1]*Nantong Key Laboratory of Translational Medicine in Cardiothoracic Diseases, and Research Institution of Translational Medicine in Cardiothoracic Diseases, Affiliated Hospital of Nantong University, Medical School of Nantong University, Nantong 226001, Jiangsu, China*
[2]*Department of Thoracic Surgery, Affiliated Hospital of Nantong University, Medical School of Nantong University, Nantong 226001, Jiangsu, China*
[3]*Research Center of Clinical Medicine, Affiliated Hospital of Nantong University, Medical School of Nantong University, Nantong 226001, Jiangsu, China*
[4]*School of Public Health, Nantong University, Nantong 226019, Jiangsu, China*

Correspondence should be addressed to Youlang Zhou; zhouyoulang@ntu.edu.cn and Jiahai Shi; sjh@ntu.edu.cn

Received 2 April 2022; Revised 27 July 2022; Accepted 23 August 2022; Published 13 September 2022

Academic Editor: Xinyi Liu



*Background*. Recently, a large number of studies have sought personalized treatment for lung squamous cell carcinoma (LUSC) by dividing patients into different molecular subtypes. Autophagy plays an important role in maintaining the tumor microenvironment and immune-related biological processes. However, the molecular subtypes mediated by autophagy in LUSC are not clear. *Methods*. Based on 490 LUSC samples, we systematically analyzed the molecular subtype modification patterns mediated by autophagy-related genes. The ssGSEA and CIBERSORT algorithm were utilized to quantify the relative abundance of TME cell infiltration. Principal component analysis was used to construct autophagy prognostic score (APS) model. *Results*. We identified three autophagy subtypes in LUSC, and their clinical outcomes and TME cell infiltration had significant heterogeneity. Cluster A was rich in immune cell infiltration. The enrichment of EMT stromal pathways and immune checkpoint molecules were significantly enhanced, which may lead to its immunosuppression. Cluster B was characterized by relative immunosuppression and relative stromal activation. Cluster C was activated in biological processes related to repair. Patients with high APS were significantly positively correlated with TME stromal activity and poor survival. Meanwhile, high APS showed an advantage in response to anti-PD1 and anti-CTLA4 immunotherapy. *Conclusion*. This study explored the autophagy molecular subtypes in LUSC. We also discovered the heterogeneity of TME cell infiltration driven by autophagy-related genes. The established APS model is of great significance for evaluating the prognosis of LUSC patients, the infiltration of TME cells, and the effect of immunotherapy.

## 1. Introduction

Lung cancer ranks second among common cancers in the world, after breast cancer. The mortality rate of lung cancer is still impressive, accounting for about one-tenth of diagnosed cancers and one-fifth of deaths in 2020 worldwide [1, 2]. Lung squamous cell carcinoma (LUSC), as a major subtype of non-small cell lung cancer (NSCLC), is generally considered to be closely related to smoking [3]. Although, with the help of imaging development and physical examination, great progress has been made in the early diagnosis and treatment of lung cancer, the survival rate of lung cancer is still very low [4, 5]. In addition, due to the heterogeneity of individuals and the insensitivity to molecular targeted

therapy, the treatment of LUSC still has great limitations [6]. Therefore, it is urgent to seek new treatment directions and individualized treatment plans for LUSC.

Autophagy is a strictly regulated multistep lysosomal degradation pathway in eukaryotes. It degrades damaged organelles, unfolded proteins, and harmful toxins and transports them to the lysosome for digestion to promote the metabolism and renewal of the cell itself [7]. The imbalance of this regulation process often leads to the disorder of the microenvironment, which may lead to the occurrence of tumor. Autophagy plays the role of double-edged sword in the field of tumors [8]. On the one hand, in the initial stage of cancer, autophagy achieves quality control by removing damaged organelles and toxic components of certain proteins, thereby limiting the transmission of carcinogenic signal pathways and playing a role in suppressing cancer [9]. On the other hand, once the tumor progresses to an advanced stage, cancer cells maintain their metabolism and proliferate migration through autophagy-mediated circulation, thereby promoting the development of the tumor [10]. This indicates that regulating autophagy may become one of the effective intervention strategies for cancer treatment.

Tumor microenvironment (TME) refers to the complex and constantly changing ecosystem in which tumor cells are located. It includes stromal cells, fibroblasts, and endothelial cells around the tumor, as well as natural and adaptive immune cells [11, 12].

The components of TME play a key role in tumor growth and immune response, and their heterogeneity may mediate different immune activation states [13]. Therefore, assessing the status of TME may help us predict the potential response of immunotherapy, so as to carry out separate treatment options.

In this study, we successfully divided the LUSC patients into three autophagy clusters based on autophagy-related genes (ARGs). We then explored the different autophagy modification patterns and the characteristics of TME cell infiltration in the autophagy clusters. In addition, LUSC patients were separated into three gene clusters based on differentially expressed genes (DEGs) identified from the three clusters. At the same time, an autophagy prognostic score (APS) model was established to quantify the autophagy characteristics of individuals and predict the clinical response of patients to immune checkpoint inhibitor (ICI) treatment. These findings indicated that APS can be used as an indicator for prognostic evaluation and selection of immunotherapy for LUSC patients.

## 2. Materials

*2.1. Data Collection and Processing of the LUSC Cohort.* We searched available public databases TCGA (https://portal.gdc.cancer.gov/) and GEO (https://www.ncbi.nlm.nih.gov/geo/) for gene expression, mutation information, and related clinical data of LUSC. Patients with incomplete survival information were excluded. For TCGA data, we obtained RNA-sequencing data (FPKM value) of 490 LUSC patients. The FPKM value was then transformed into Transcripts Per Kilobase Million (TPM) value, which is more comparable to microarray datasets from GEO. Moreover, we downloaded the LUSC microarray datasets with complete clinical information including GSE73403 ($n = 69$), GSE74777 ($n = 107$), and GSE157011 ($n = 484$) from GEO for verification. R (version 4.0.2) was used for data analysis.

*2.2. Establishment of Autophagy Clusters in LUSC Patients.* Initially, 232 ARGs were downloaded from the Human Autophagy Database (HADb: https://www.autophagy.lu/) [14]. The expression of ARGs in LUSC patients was extracted from TCGA. Univariate Cox analysis was then used to screen out ARGs associated with prognosis of LUSC ($P < 0.05$). Then, based on the expression of these genes, unsupervised cluster analysis was performed on LUSC patients to identify different autophagy clusters. The potential number of clusters was determined by the cumulative distribution function (CDF) curve of the consensus score [15]. When using the ConsensusClusterPlus R package to perform the above steps, 1000 repetitions were undertaken to ensure classification stability [16].

*2.3. Functional Annotation in Autophagy Clusters.* Gene set variation analysis (GSVA) was performed by GSVA R package to analyze the biological differences among different autophagy clusters [17]. The "c2.cp.kegg.v6.2.-symbols" gene sets were download from MSigDB database to run the GSVA analysis. A $P$ value less than 0.05 was considered statistically significant.

*2.4. Estimation of TME Cell Infiltration in LUSC Patients.* The single-sample gene set enrichment analysis (ssGSEA) algorithm was utilized to quantify the relative abundance of TME cell infiltration in each LUSC patient, which was then normalized to unity distribution from 0 to 1 [18]. To further quantify the specific composition ratio of immune cells in LUSC patients, the CIBERSORT R package was used for specific quantification [19]. The percentages of 22 distinct immune subsets in each LUSC patient and each cluster were then obtained. Then, through the ESTIMATE R package, we calculated the immune score, stromal score, and estimate score of each LUSC patient to predict the level of immune cell and stromal cell infiltration in the tumor, which form the basis for inferring tumor purity [20].

*2.5. Establishment of Autophagy Gene Clusters.* The limma R package was utilized to determine DEGs between autophagy clusters. The identification standard was adjusted: $P$ value <0.01 and $|\log FC| > 1$. Then the DEGs associated with the prognosis of LUSC were further determined through univariate Cox analysis ($P < 0.01$). Then, according to the expression of these prognostic-related DEGs, unsupervised cluster analysis was performed again on LUSC patients to identify autophagy gene clusters.

*2.6. Construction and Evaluation of the APS Model.* To quantify the autophagy characteristics of every individual patient, we constructed a comprehensive scoring system

called APS. The expression of these prognostic-related DEGs was subjected to the principal component analysis (PCA), in which principal component 1 and principal component 2 were both extracted to construct APS. The advantage of this method is to keep the main components as much as possible while reducing the dimensionality and eliminate the mutual influence factors between the components. After obtaining the expression of each gene, we defined the APS using a method similar to the study of Zhang et al. [21]:

$$\mathrm{APS} = \sum (\mathrm{PC1}i + \mathrm{PC2}i), \quad (1)$$

where $i$ is the expression of prognostic-related DEGs.

According to the best cut-off value determined by survminer R package, LUSC patients were divided into high APS group and low APS group. At the same time, the univariate and multivariate Cox regression models including age, gender, stage, and APS were utilized to assess the prognostic factors of LUSC patients. The hazard ratio (HR) value distinguished the prognostic predictors of risk factors and protective factors (HR > 1 was a risk factor and HR < 1 was a protective factor, $P < 0.05$). To verify whether the APS model has wide applicability, three GEO datasets GSE73403, GSE74777, and GSE157011 were downloaded as validation sets. Each LUSC patient had an APS value calculated by this model, and the Kaplan–Meier curve was utilized to reflect its survival value.

*2.7. Correlation of APS Characteristics with Tumor Mutation Burden (TMB) and Immunotherapy.* In order to determine the relationship between APS and TMB, the total numbers of mutations in high APS group and low APS group were calculated, respectively. The maftools R package was used to draw the mutation waterfall plot of the top 20 genes [22]. Meanwhile, the overall mutation rates of the two groups were calculated to obtain the TMB score. After that, the correlation between APS and TMB was analyzed, and their survival curves were drawn, respectively. In addition, to further study the significance of APS in immunotherapy, Wilcoxon test was utilized to explore the differential expression in immune checkpoints such as PD-L1 and CTLA4 among different APS groups. Finally, data of immunophenoscore (IPS) of the immune checkpoint inhibitor (ICI) were download from The Cancer Immunity Database (TCIA, https://tcia.at/home), which can predict the intergroup differences in response to immunotherapy [23, 24].

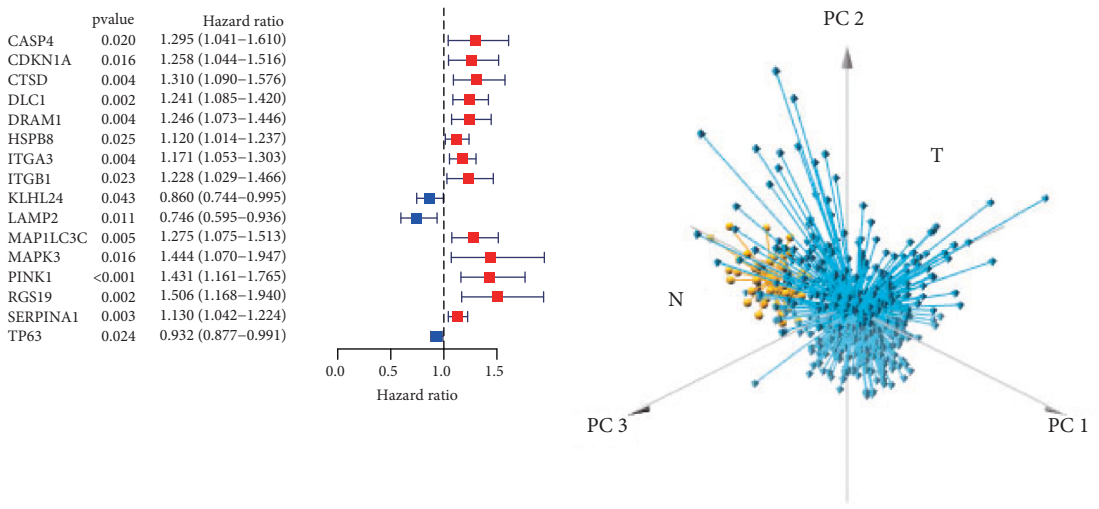
*2.8. Statistics.* In this study, Wilcoxon test was utilized to analyze the differences between two groups. Kruskal-Wallis and one-way ANOVA tests were utilized to analyze the differences between three or more groups. Spearman analysis was utilized for correlation analysis. The surv_cutpoint function module in the survival R package was utilized for the best grouping. Besides, the survminer R package was utilized to test all potential cut-off points to get the maximum rank statistic. Log-rank test and Kaplan–Meier test were utilized to draw the survival curve.

## 3. Results

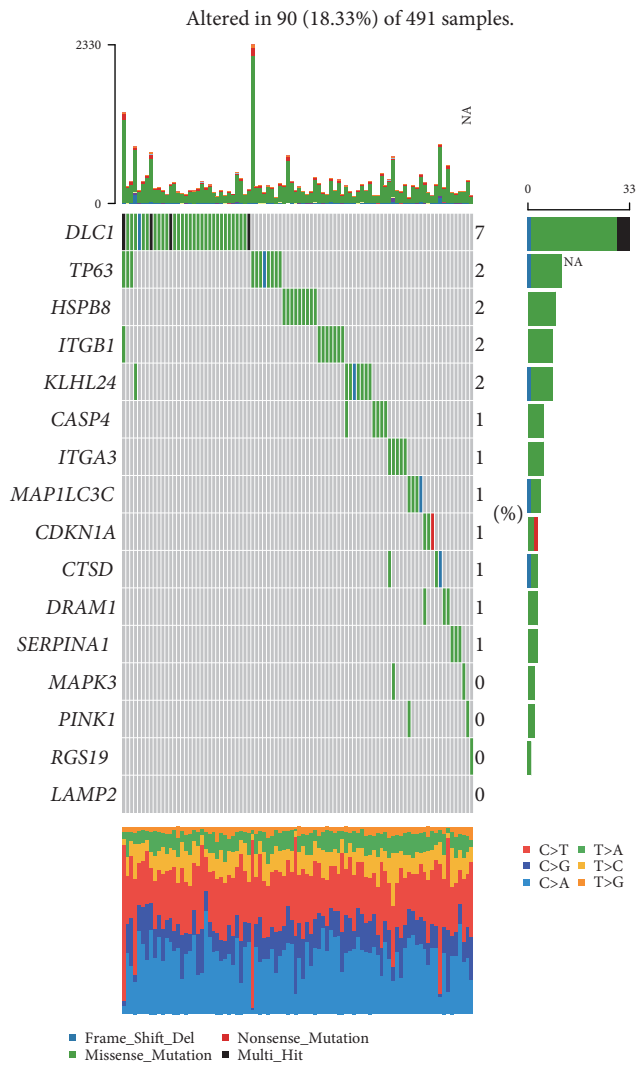
*3.1. Landscape of Prognostic-Related ARGs in LUSC.* A total of 210 ARGs were extracted from the TCGA database, among which 16 ARGs were associated with the prognosis of LUSC (Figure 1(a)). By reducing the dimensions of these prognostic-related ARGs, tumor and normal samples showed separate populations (Figure 1(b)). Then, the total mutation of these prognostic-related ARGs was observed in 90 out of 491 LUAD samples (18.33%). It was also pointed out that DLC1 had the highest mutation frequency, followed by TP63 and HSPB8 (Figure 1(c)). At the same time, there were significant differences in the expression of these genes in tumor samples and normal samples, with the exception of TP63, KLHL24, and LAMP2, which were all downregulated in tumors (Figure 1(d)). The above analysis indicated that the expression of these prognostic-related ARGs in normal and LUSC tissues was highly heterogeneous, which may be one of the mechanisms that mediate tumorigenesis.

*3.2. Identification of Autophagy Clusters in LUSC.* Depending on the expression of above prognostic-related ARGs, unsupervised cluster analysis was performed on LUSC patients. Based on the CDF curve, we selected three clusters as our classification (Figures 2(a) and 2(b), Figures S1A–S1D), and all LUSC patients were then divided into three different clusters (Figure 2(c)). There were 129 cases in cluster A, 210 cases in cluster B, and 151 cases in cluster C. Besides, we found that the ARGs expression in cluster A significantly exceeded those in the other two clusters (Figure S1E). Principal component analysis (PCA) further confirmed the distinct characteristics of these three clusters (Figure 2(d)). At the same time, survival analysis showed that cluster A had a significant survival disadvantage compared with the other two clusters, while cluster C had a prominent survival advantage (Figure 2(e)).

*3.3. TME Cell Infiltration and Immune Cell Characteristics among Three Autophagy Clusters.* To further explore the biological differences between the three autophagy clusters, we performed GSVA enrichment analysis in pairs (Figures 3(a) and 3(b)). Cluster A was strongly enriched in pathways associated with immune activation such as T cell receptor, B cell receptor signaling pathway, and natural killer cell mediated cytotoxicity. Clusters B and C were both enriched in base excision repair, mismatch repair, and cell cycle. Moreover, cluster C showed stronger immunosuppression. Next, ssGSEA was used to detect the infiltration of TME cell in the autophagy clusters. The results showed that the overall TME cell infiltration of cluster A was significantly abundant and the overall TME cell infiltration of cluster C was the lowest. Cluster B was somewhere in between (Figure 3(c)). However, patients in cluster A did not show the survival advantage brought by abundant TME cells. The composition of different immune cell may constitute different immune microenvironment, thereby mediating

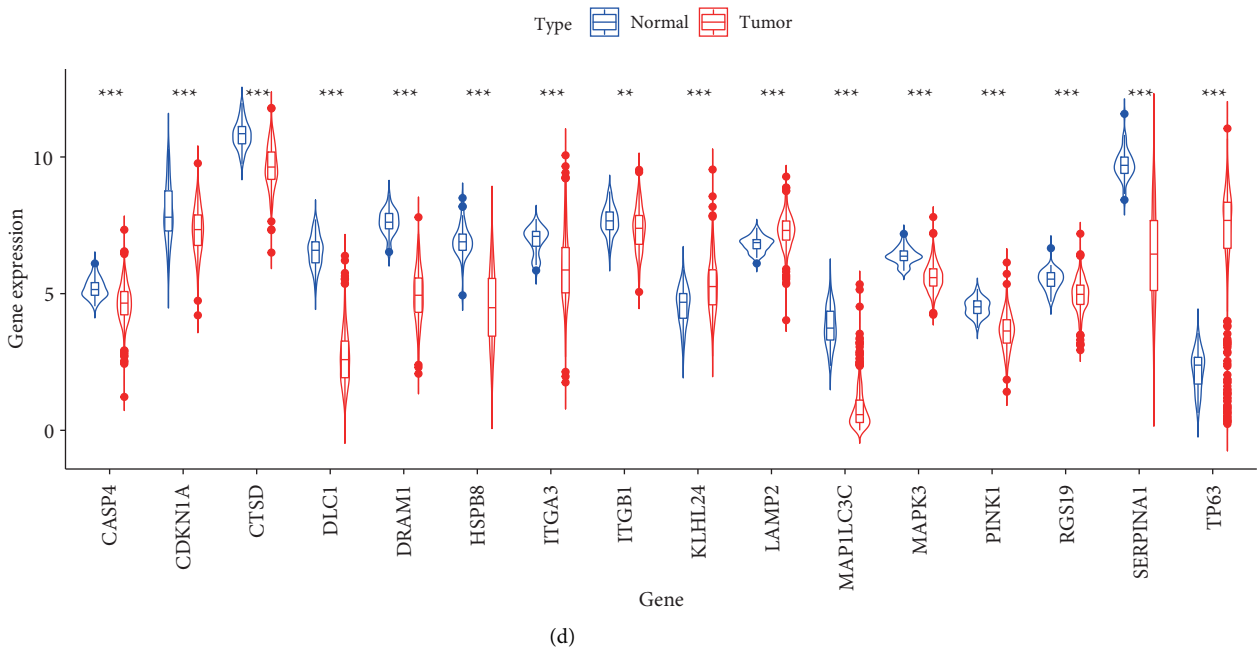



(a) (b)



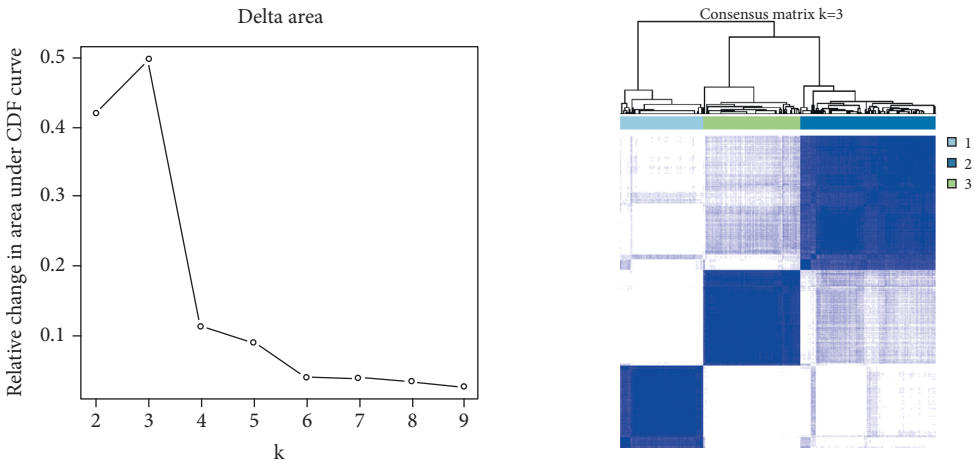

(c)

Figure 1: Continued.




(d)

Figure 1: Landscape of autophagy-related genes (ARGs) associated with prognosis of LUSC. (a) Forest plot showing the prognostic values of 16 prognostic-related ARGs in LUSC ($P < 0.05$). (b) Principal component analysis (PCA) for the prognostic-related ARGs between tumor and normal showed separate populations. (c) The waterfall plot showing the mutation landscape of prognostic-related ARGs. (d) The relative expression of prognostic-related ARGs in LUSC.




(a) (b)

Figure 2: Continued.

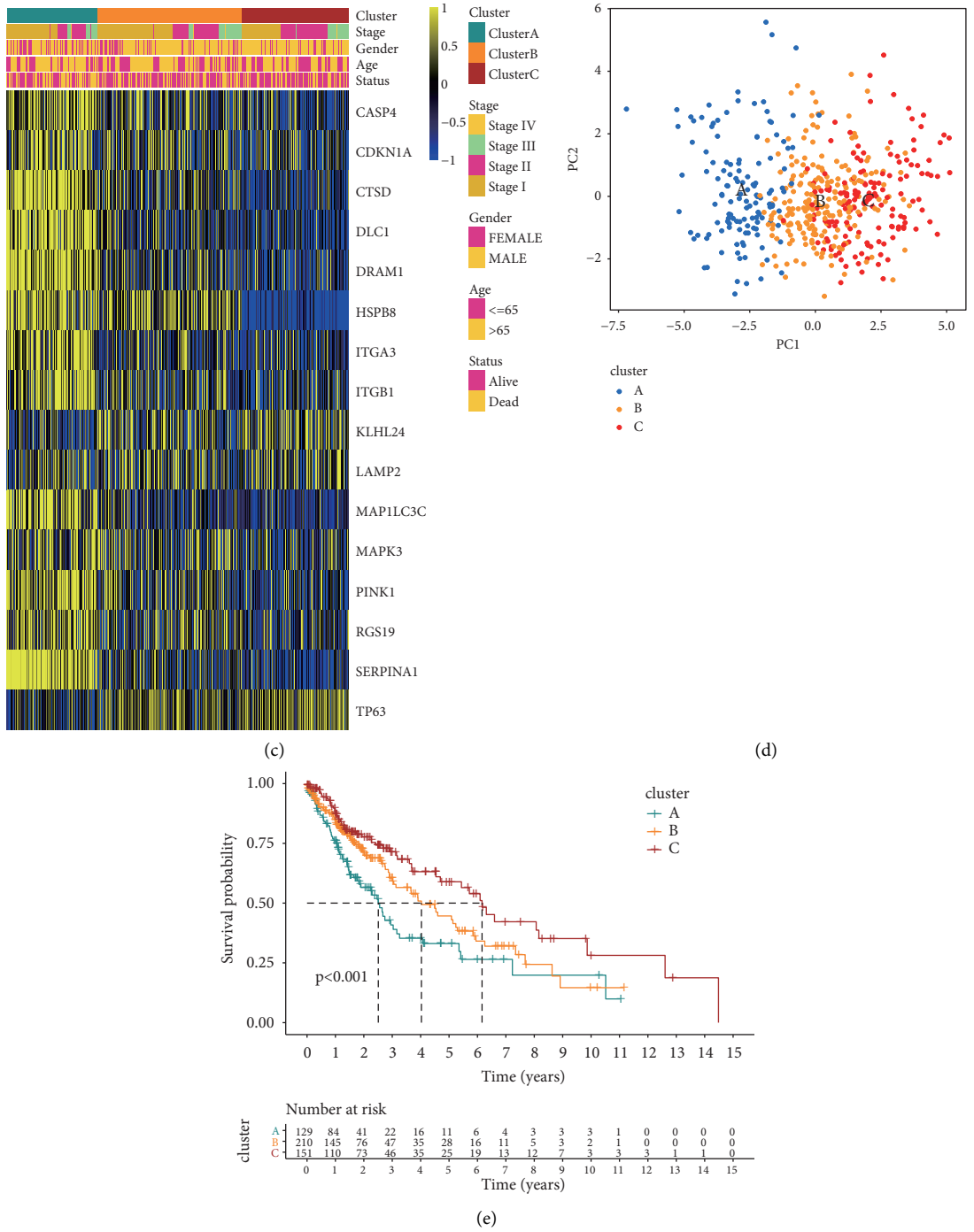
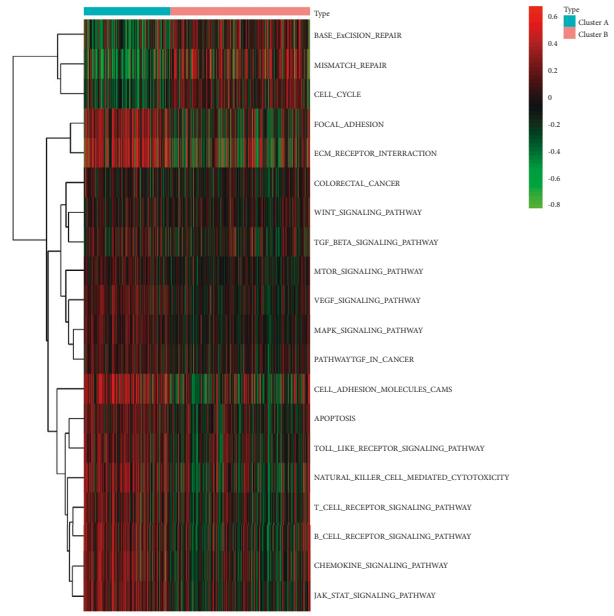
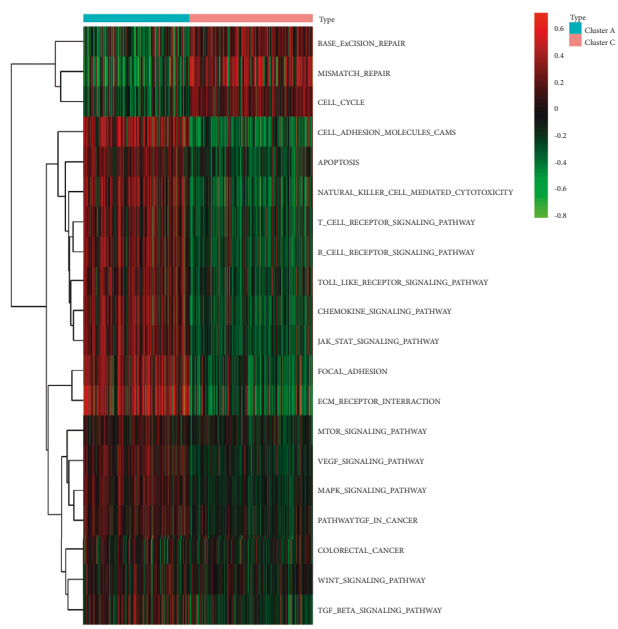



Figure 2: Identification of autophagy clusters in LUSC. (a) Delta area curve for clustering, representing the relative change in area under the cumulative distribution function (CDF) curve. (b) Heat map of sample clustering under $k = 3$ in LUSC. (c) Unsupervised clustering of 16 prognostic-related ARGs in LUSC. Blue, orange, and red represent autophagy clusters A, B, and C, respectively. Cluster, stage, gender, age, and survival status were used as patient annotations. Yellow and blue represent high and low expressions of ARGs, respectively. (d) PCA for the ARGs of three clusters among LUSC patients. (e) Survival analysis of the three autophagy clusters among all LUSC patients in TCGA.

different immune responses. So we further analyzed the specific components of immune cells in the three clusters. The relative percentages of 22 immune cells in each LUSC patient were obtained through CIBERSORT R package (Figure S2A). The composition of immune infiltrating cells among the three autophagy clusters also showed heterogeneity (Figure 3(d)). Cluster A had higher percentages of CD4 memory resting T cells and M2 macrophages compared to the other two clusters. Meanwhile, cluster C was rich in M1 macrophages and activated mast cells. Cluster B had more M0 macrophages and activated dendritic cells compared to the other two clusters. Then we performed

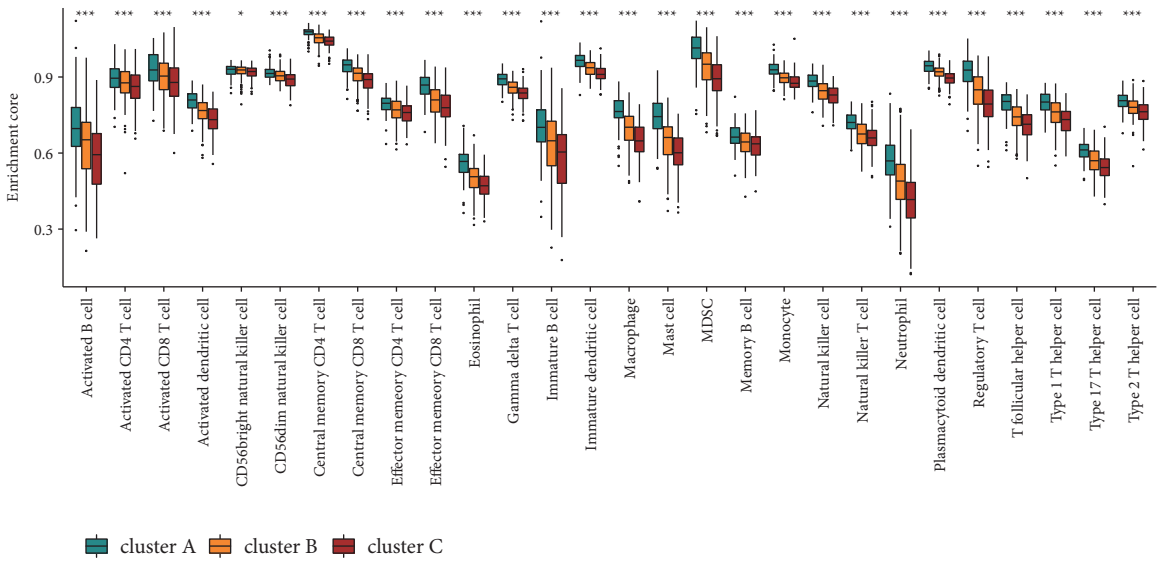



(a)

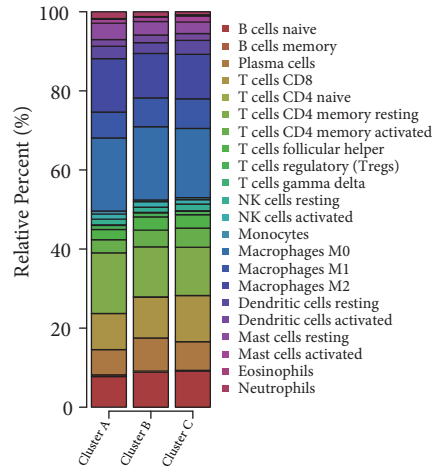



(b)

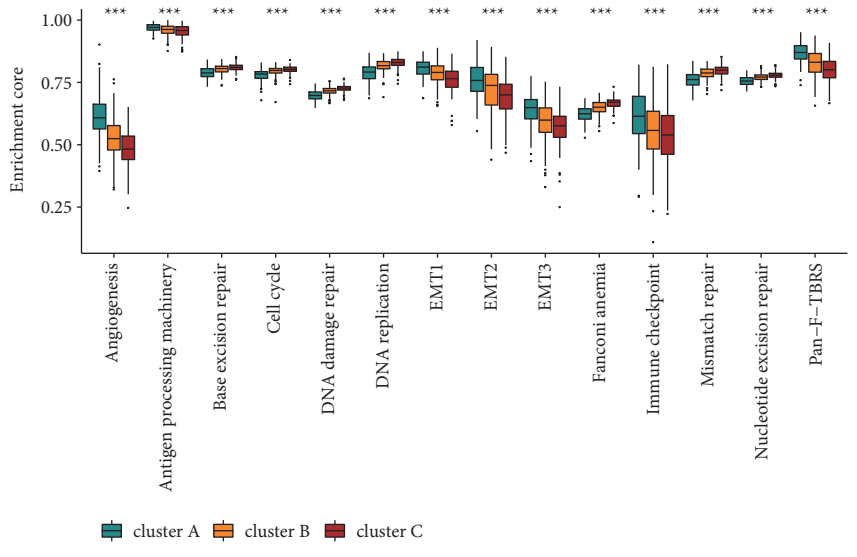
Figure 3: Continued.




(c)




(d)




(e)

Figure 3: TME cell infiltration and immune cell characteristics in the autophagy clusters. ((A) and (B)) GSVA analysis showing the biological pathways in the three autophagy clusters. Yellow represents activated pathways and blue represents inhibited pathways. (a) Cluster A versus cluster B; (b) cluster A versus cluster C. (c) TME cell infiltration of the three autophagy clusters. (d) The relative percentage of 22 immune infiltrating cells among the three clusters. (e) Differences in stroma-activated pathways among the three clusters. The asterisks represent the $P$ value ($^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$).

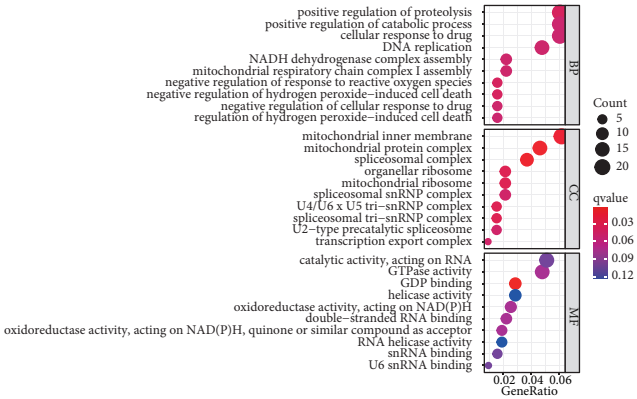
a comprehensive score based on the content of the immune cells and stromal cells and found that cluster A was at a relatively high level (Figures S2B–S2D). In addition, we were surprised to find that cluster A not only had a high level of immune cell infiltration but also had strong TME mesenchymal activities such as Epithelial-Mesenchymal Transition (EMT), pan-fibroblast transforming growth factor $\beta$ response (Pan-F-TBRs), and angiogenesis pathways. It was worth noting that cluster A was significantly higher than the other two clusters in terms of immune checkpoint. Meanwhile, cluster C showed higher activity in biological pathways related to repair such as base excision repair, DNA damage repair, and mismatch repair (Figure 3(e)).

*3.4. Reclassification of Autophagy Subtypes.* To further explore autophagy modification pattern in LUSC, 373 DEGs were then identified among the three autophagy clusters (Figure S3A). Gene ontology enrichment analysis showed that these DEGs were involved in the positive regulation of proteolysis, mitochondrial inner membrane, and catalytic activity, acting on RNA (Figure 4(a)). Then, the DEGs associated with the prognosis of LUSC were screened out, and unsupervised cluster analysis was performed again on 490 LUSC patients. The reclassification of autophagy subtypes was defined as autophagy gene clusters, of which 204 cases belong to gene cluster A, 74 cases belong to gene cluster B, and 212 cases belong to gene cluster C (Figures 4(b) and 4(c), Figures S3B–S3F). Similarly, these three gene clusters can clearly separate LUSC (Figure S3G). It was worth noting that these gene clusters also have heterogeneity in the expression of prognostic-related ARGs (Figure 4(d)). Among them, gene cluster C showed a low level of expression in most prognostic-related ARGs. The survival curve also suggested that gene cluster C had a clear survival advantage (Figure 4(e)).

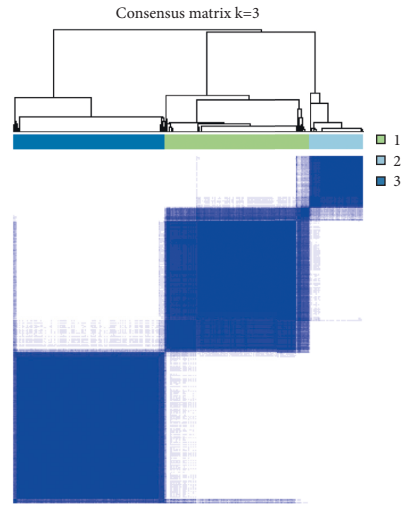
*3.5. Construction and Evaluation of the APS Model.* Considering that the autophagy subtype is only an assessment of different groups of LUSC, it is impossible to accurately assess the autophagy characteristics of every single patient. Therefore, we constructed the APS model to quantify the autophagy characteristics of each LUSC patient. We classified age, gender, TNM status, and APS of every LUSC patient into univariate and multivariate Cox regression analysis (Figures 5(a) and 5(b)). We can conclude that APS can be served as an independent prognostic factor of LUSC. Next, we tried to further verify the value of APS in predicting the prognosis of LUSC patients. According to the cut-off value of 0.935 determined by the survminer R package, LUSC patients were divided into low APS group ($n = 334$) and high APS group ($n = 156$). The survival benefit of patients in the low APS group was significant (Figure 5(c)). The GSE73403 and GSE157011 cohorts further verified that the APS model had a good predictive ability (Figures S4A and S4B). However, in the GSE74777 cohort, although the low APS group had a better prognostic advantage, it was not statistically significant ($P > 0.05$, Figure S4C). Next, we associated autophagy clusters, gene clusters, and APS through alluvial diagram (Figure 5(d)). Most patients in cluster C belonged to gene cluster C and both had low APS. Meanwhile, most patients in gene cluster B belonged to cluster A and had high APS. After that, APS was calculated on different autophagy clusters and gene clusters, which showed a great difference (Figures 5(e) and 5(f)). The scores of cluster C and gene cluster C were both at the lowest level, while the scores of cluster A and gene cluster B were at a high level. The analysis of matrix-related pathway activity indicated that low APS may be closely related to repair activation, while high APS was enriched in EMT and immune checkpoint (Figure S4D). In addition, LUSC patients with low APS had higher survival rates (Figure S4E), while surviving patients had lower APS (Figure S4F). Combining APS with the patient's TNM staging, we can see that, with the increase of T and N, the APS value showed a decreasing trend ($P < 0.05$, Figures S4B–S4C). Stage and M also had the same trend, but it was not statistically significant ($P > 0.05$, Figures S4A and S4D).

*3.6. Correlation between APS and TMB.* Previous studies have shown that TMB has guiding significance in helping patients choose immune therapy [25, 26]. In view of the differential enrichment of immune checkpoint pathway in different APS groups, we tried to associate APS with TMB. We separately analyzed the tumor somatic mutations of LUSC patients in both high and low APS groups and found that TP53 and TIN had high mutation rates in both groups. In general, the overall mutation rate in the low APS group was higher than that in the high APS group (Figures 6(a) and 6(b)), and the TMB was also at a higher level (Figure 6(c)). Correlation analysis showed a negative correlation between APS and TBM (Spearman coefficient: $R = -0.16$, $P = 7e - 04$; Figure 6(d)). Next, we jointly studied the impact of TBM and APS on the survival of LUSC. As shown in Figure 6(e), the overall survival rate of patients with high TMB was higher than that of patients with low TMB. At the same time, patients with low APS combined with high TMB had the best survival advantage (Figure 6(f)).

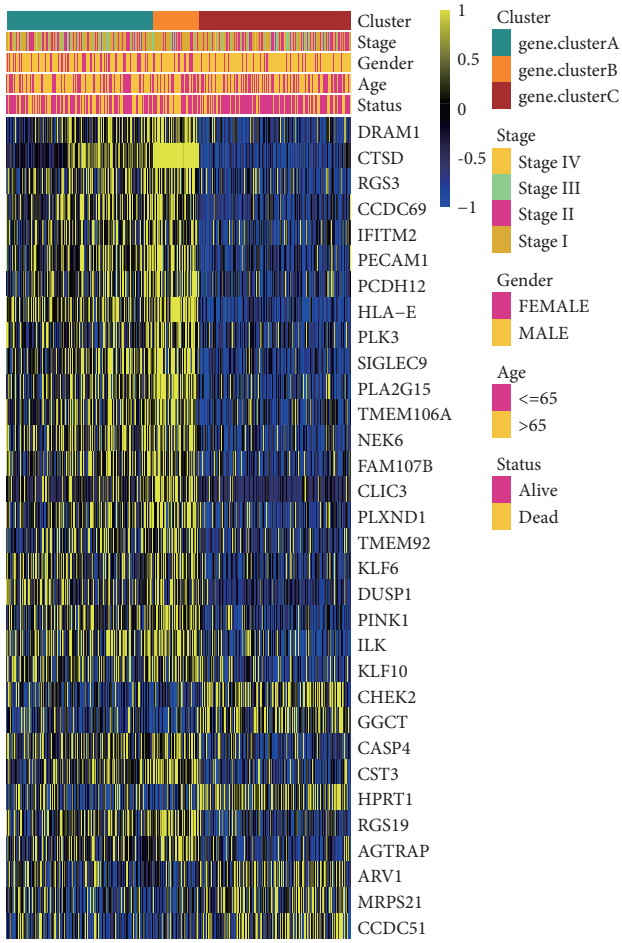
*3.7. The Effect of APS in Predicting Immunotherapy.* Immunotherapy represented by anti-PD-1 and anti-CTLA4 has become one of the promising options for cancer treatment [27, 28]. Taking into account the differential enrichment of APS in the immune checkpoint pathway, we assessed whether APS can predict the response of LUSC patients to ICI treatment. Patients with high APS showed upregulation of PD-L1 and CTLA4, which indicated that APS can be used to predict the differential expression of related ICI (Figures 7(a) and 7(b)). Also consistent with expectations, whether anti-PD-1 or anti-CTLA4 alone or combination, patients in the high APS group had higher IPS (Figures 7(c)–7(f)), which means better clinical treatment response. The results suggested that APS can be used to predict the clinical response of LUSC patients to immunotherapy.



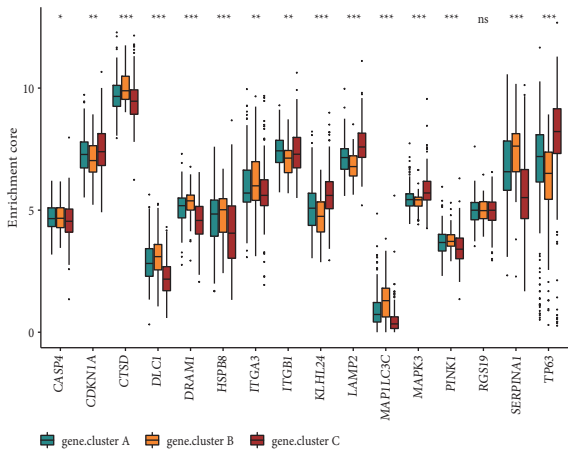

(a)




(b)

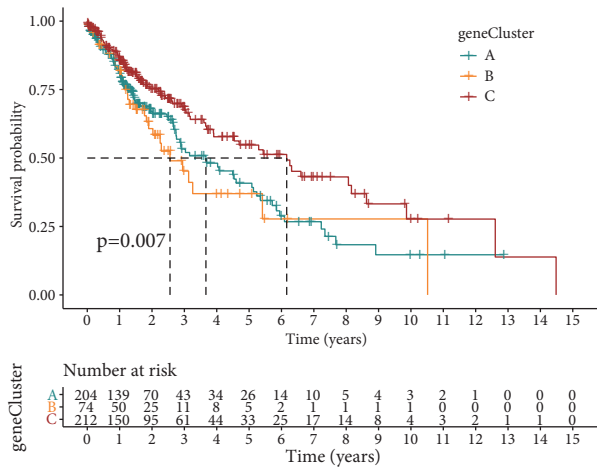



(c)



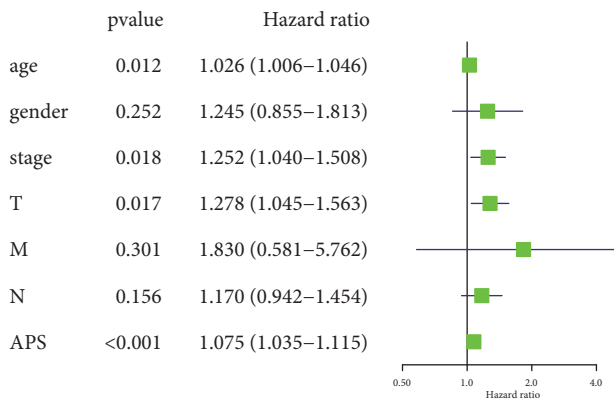

(d)

Figure 4: Continued.

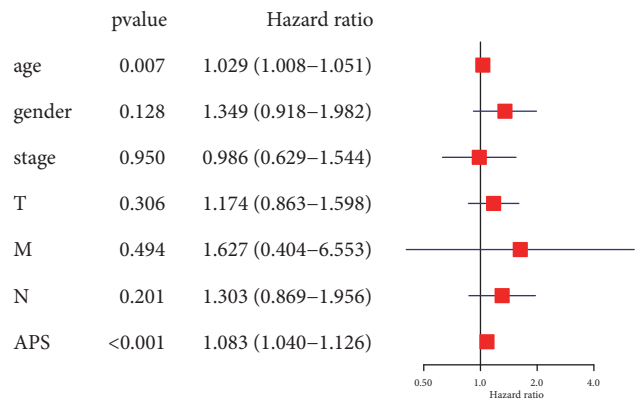



(e)

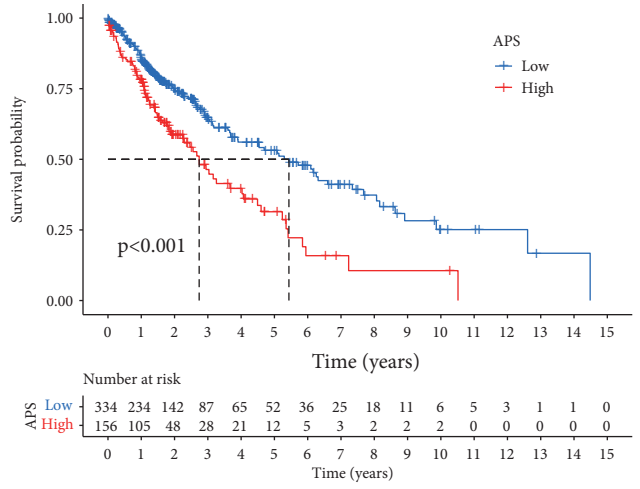
Figure 4: Reclassification of autophagy subtypes in LUSC. (a) Gene ontology analysis of the differentially expressed genes (DEGs). The size of bubbles represents the count of gene enrichment, and the color represents the q value. (b) Heat map of sample clustering under $k = 3$ in LUSC. (c) Unsupervised clustering of prognostic-related DEGs in LUSC. Blue, orange, and red represent gene clusters A, B, and C, respectively. Gene cluster, stage, gender, age, and survival status were used as patient annotations. Yellow and blue represent high and low expressions of prognostic-related DEGs, respectively. (d) The expression of prognostic-related ARGs among the three gene clusters. (e) Survival analysis of the three gene clusters among all patients.



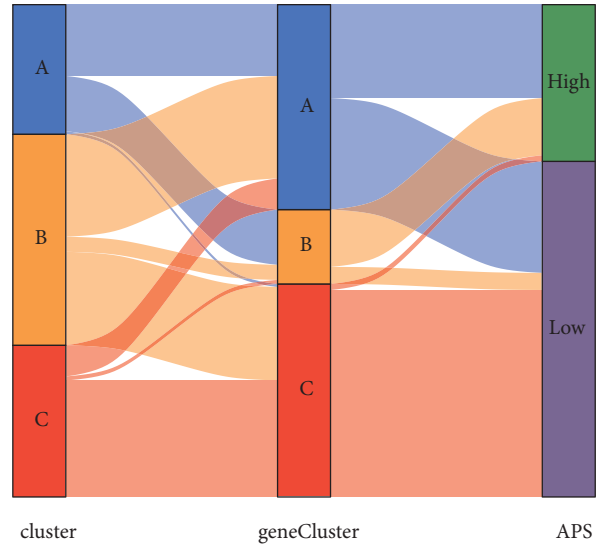

(a)




(b)




(c)




(d)

Figure 5: Continued.

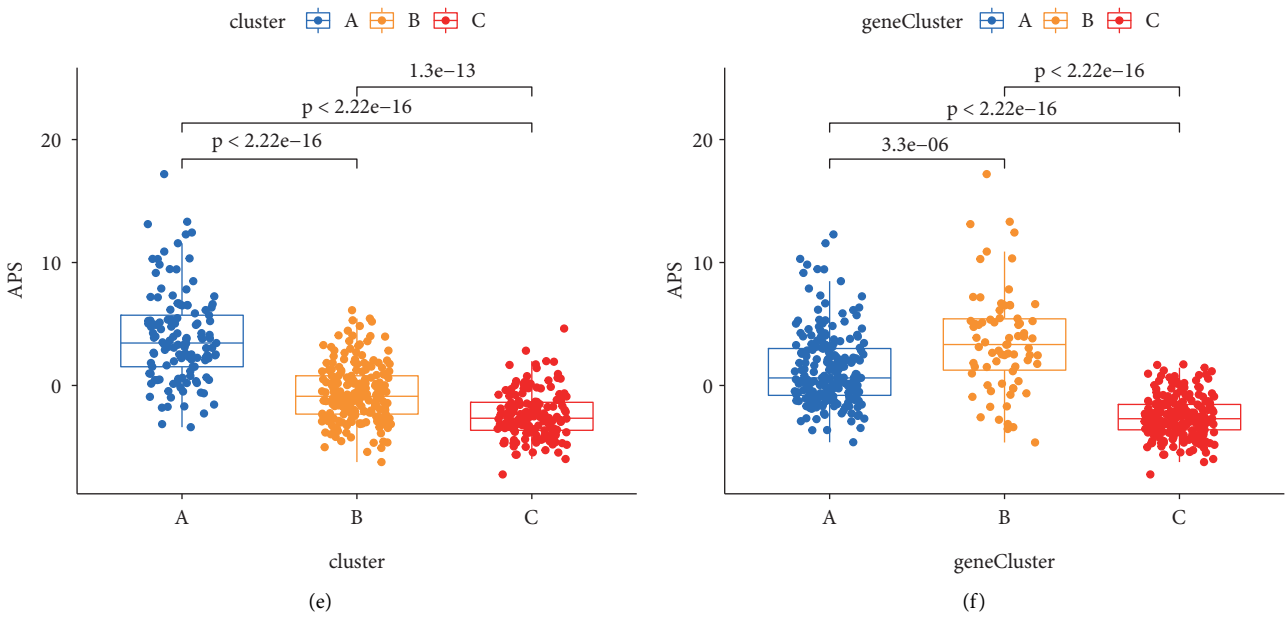



Figure 5: Construction and evaluation of APS model. (a) Univariate Cox regression analysis for the age, gender, TNM status, and APS signature in LUSC patients. (b) Multivariate Cox regression analysis for the age, gender, TNM status, and APS signature in LUSC patients. (c) Kaplan–Meier curves were used to analyze the overall survival of LUSC patients with high (156 cases) and low (334 cases) APS. (d) Autophagy cluster, gene cluster, and APS were associated through alluvial diagram. (e) Difference of APS among autophagy clusters in LUSC. (f) Difference of APS among gene clusters in LUSC.

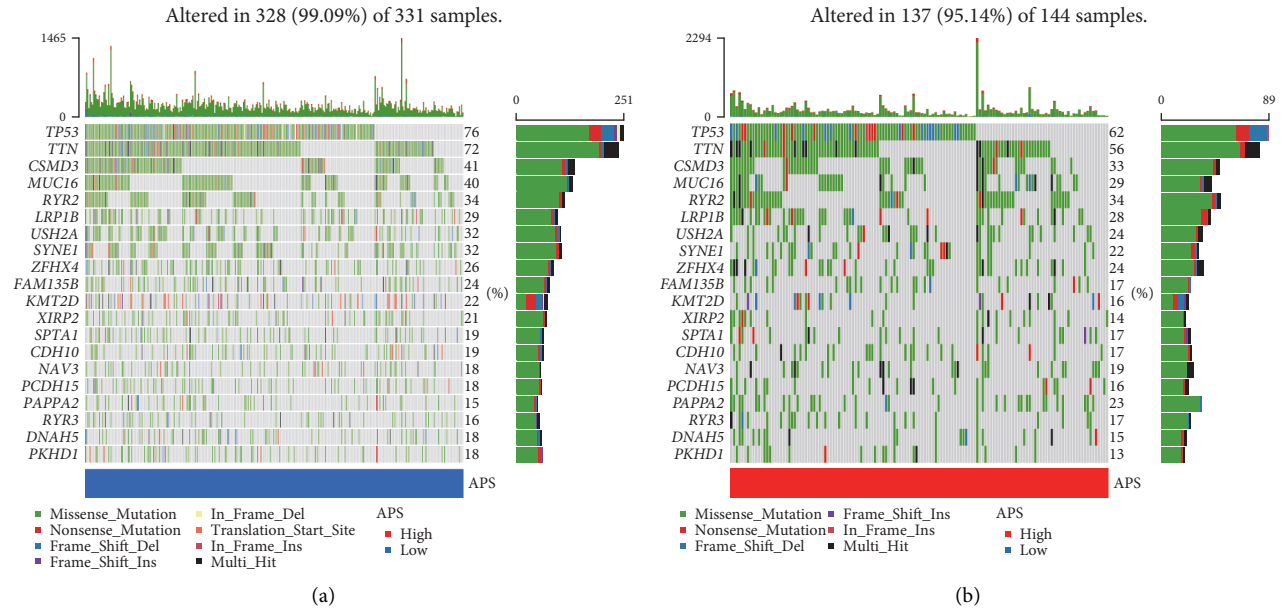



Figure 6: Continued.

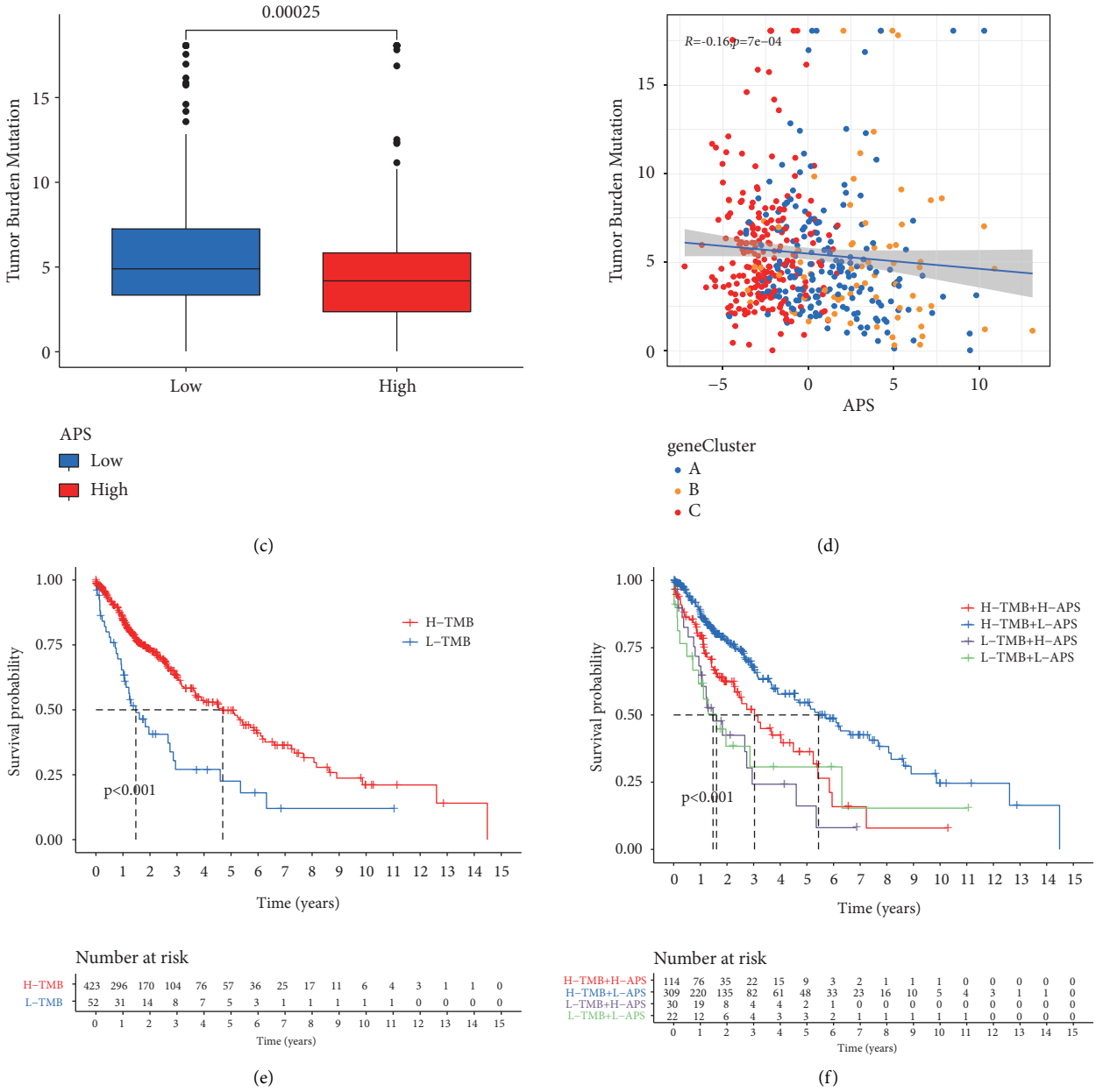



Figure 6: Correlation between APS and tumor mutation burden (TMB). (a) The waterfall plot shows the frequency of tumor somatic mutations landscape in low APS group. (b) The waterfall plot shows the frequency of tumor somatic mutations landscape in high APS group. (c) The relative score of TMB in low and high APS groups. (d) Correlation between APS and TMB. (e) Kaplan–Meier curves of overall survival rates in low and high TMB score groups. (f) Kaplan–Meier curves of overall survival rates by both APS and TMB.

## 4. Discussion

Although in recent years, with the help of the imaging development and physical examination, more lung cancers have been treated at an early stage, there are still a considerable number of lung cancers, especially LUSC, which are already in the advanced stage when they are discovered [29]. At the same time, with recurrence and distant metastasis, surgical treatment still has great limitations for LUSC. As LUSC is a strongly heterogeneous cancer, radiotherapy, chemotherapy, and molecular targeted therapy cannot benefit all LUSC patients [6, 30]. Therefore, it is urgent to seek new molecular biomarkers to predict and guide the personalized treatment of LUSC. Previous studies have tried to explore different molecular subtypes in LUSC. For example, Wilkerson et al. divided 382 LUSC patients into four molecular subtypes based on mRNA levels [31]; Li et al. divided LUSC patients into seven subtypes based on 965 DNA methylation sites [32]; Xu et al. defined four molecular subtypes of LUSC based on DNA copy number or methylation-related gene expression [33]. Although these molecular subtypes tried to classify LUSC to find their prognostic markers, these classifications did not explore the differences in TME among subtypes and the targeted differences in prognostic treatment were not

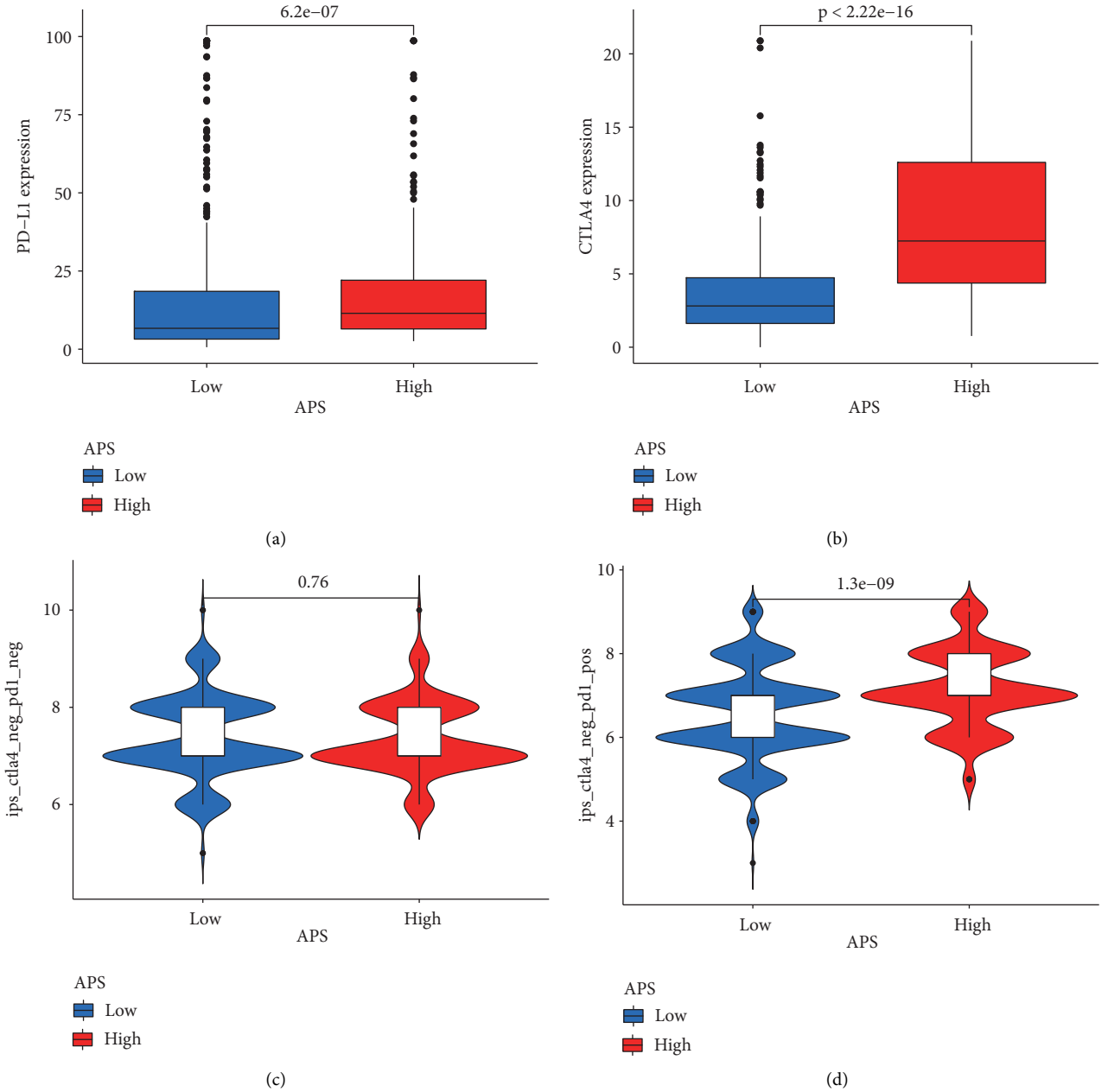



Figure 7: Continued.

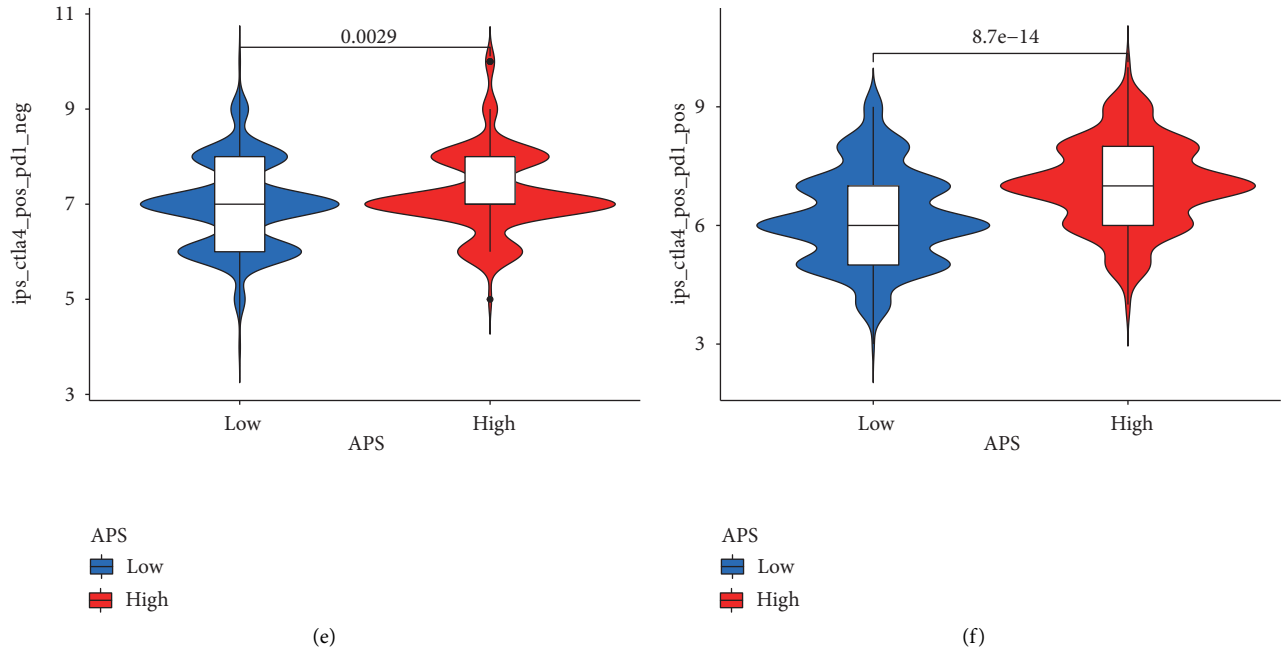



Figure 7: The effect of APS in predicting immunotherapy. (a, b) The expression of PD-L1 and CTLA4 between low and high APS groups. (c–f) The immunophenoscore (IPS) of the low and high APS groups with different immune checkpoint inhibitor (ICI) treatment. (c) CTLA4 (−)/PD1 (−); (d) CTLA4 (−)/PD1 (+); (e) CTLA4 (+)/PD1 (−); (f) CTLA4 (+)/PD1 (+).

clear. Therefore, it is of great significance to carry out more accurate and comprehensive classification for LUSC clinical personalized treatment.

Autophagy plays different roles in different TME, and tumors can often change the growth and metastasis of tumors by regulating autophagy and thereby regulating the immune response [34, 35]. Therefore, further analysis of the differences in autophagy and immune pathways in different TME may help understand the specific mechanisms of tumor autophagy. The role of autophagy and autophagy-related genes has been explained in a variety of cancers. For example, a prognostic model based on autophagy-related genes was established in gastric cancer [36]; a prognostic model based on autophagy-related lncRNA was established in pancreatic cancer [37]; three colon cancer molecular subtypes were defined through the expression of autophagy-related genes [38]; similar research works have also been carried out to explore the relationship between autophagy and breast cancer, bladder cancer, and endometrial cancer [39–41]. However, there is a lack of research on autophagy in LUSC. Therefore, this study investigated the comprehensive role of autophagy in different LUSC phenotype and TME.

In this study, we comprehensively constructed the autophagy-related characteristics of LUSC through TCGA database and then divided LUSC patients into three different autophagy clusters through sixteen prognostic-related ARGs. Because tumors can regulate metabolism in TME through autophagy, regulate oxidative stress and hypoxia, and evade host immune surveillance to support cancer growth [42], we studied the relationship between three autophagy clusters and TME cell infiltration. Through ssGSEA, we can see that the TME cell infiltration in these three clusters showed different characteristics. Cluster A was rich in a large number of innate and adaptive immune cells. However, to our surprise, patients in cluster A did not show the expected survival advantage compared with the other two clusters. On the contrary, patients in cluster A had the worst prognosis. Previous studies have shown that there is a type of inflammatory tumor rich in infiltrating lymphocytes. In order to evade immune surveillance, such tumor recruits a large number of myeloid-derived immune cells or secrete factors including TGF-$\beta$ to create an immunosuppressive tumor microenvironment. The tumor keeps immune cells around the tumor cell nest by producing highly expressed reactive mesenchyme and dense extracellular matrix, which makes the tumor exhibit strong immunosuppression despite the abundant immune cell infiltrating gallbladder [43–45]. Therefore, we focused on the TME stromal activation in cluster A, and the R package of ESTIMATE confirmed the significant increase in the stromal activity of the cluster. In the patients of cluster A, the expression of angiogenesis and EMT stromal signaling pathways and immune checkpoint molecules also increased significantly, which may limit the immune attack ability of the T cells. Meanwhile, it can be seen from the content of immune cells that cluster A contains more M2 macrophages, which was consistent with previous studies [46, 47]. For cluster C, although TME immune cells were poor, the patient's survival time was significantly longer than those of the other two clusters. From the GSVA enrichment analysis, patients in cluster C mainly activated the biological processes related to repair, including matrix repair and mismatch repair. The mesenchymal activation pathway also indicated that cluster C was also enriched in DNA damage repair. In addition, the expression of immune checkpoint molecule in this cluster was

relatively low. Immune cell content analysis also showed that cluster C had a high proportion of quiescent mast cells and CD4 helper T cells. The better prognosis was also consistent with previous studies [46]. Therefore, we emphasized the role of repairing activated and functioning T cells in cluster C. Cluster B was characterized by relative immune suppression and relative stromal activation. These findings may promote our understanding of the relationship between tumor autophagy, TME cell infiltration, and matrix repair. Next, we reclassified LUSC by DEGs between the three clusters to obtain three gene clusters. The three gene clusters also had obvious heterogeneity, and their prognosis had obvious differences. From the mulberry diagram, it can be seen that the autophagy clusters had a rough correspondence with the autophagy gene clusters. Taking into account the heterogeneity of LUSC, we constructed APS model based on these prognostic-related DEGs to further evaluate the autophagy characteristics of each LUSC patient. Cluster C and gene cluster C had low APS, corresponding to a better prognosis, and their autophagy gene expression was at a low level. Cluster A and most gene cluster A had high APS, corresponding to a poor prognosis, and their autophagy gene expression was at a high level. The independent prognostic value of APS in LUSC patients has been further confirmed. Then we observed a negative correlation between APS and TMB, which also means that lower autophagy gene expression has higher TMB, corresponding to a better prognosis. At the same time, the expression of immune checkpoint molecules like PD-L1 and CTLA4 was positively correlated with APS, indicating that APS may have a guiding role in predicting immunotherapy. In the TCIA database, significant clinical benefits were observed in patients with high APS who received ICI therapy, which may provide new guidance for future immunotherapy of LUSC patients. In conclusion, this study revealed that the three autophagy clusters have different autophagy and TME infiltration characteristics in LUSC, and the APS model constructed at the same time can be used as a biomarker to predict patient prognosis and guide immunotherapy.

## Data Availability

The datasets generated and/or analyzed during this study are available from GEO (https://www.ncbi.nlm.nih.gov/geo/) and TCGA official website (https://portal.gdc.cancer.gov/repository).

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Jinjie Wang and Jiaqi Zhu contributed equally to this work.

## Acknowledgments

This study was funded by the National Natural Science Foundation of China (81770266), Clinical Medical Research Center of Cardiothoracic Diseases in Nantong (HS2019001), Innovation Team of Cardiothoracic Disease in Affiliated Hospital of Nantong University (TECT-A04), and Nantong Key Laboratory of Translational Medicine of Cardiothoracic Diseases.

## Supplementary Materials

Figure S1: identification of autophagy clusters in LUSC. (A) Consensus clustering cumulative distribution function (CDF) with the number of subtypes $k = 2$ to 9. ((B)–(D)) Unsupervised clustering of prognostic-related ARGs in LUSC and consensus matrices for $k = 2,4,5$. (E) The relative expression of 16 prognostic-related ARGs in three autophagy clusters. Figure S2: immune cell characteristics in the three autophagy clusters. (A) The relative percentages of 22 different immune subsets in each LUSC patient. (B) Difference in immune score among three clusters in LUSC. (C) Difference in stromal score among three clusters in LUSC. (D) Difference in estimate score among three clusters in LUSC. (E) The relationship between 22 immune cells. Figure S3: identification of autophagy gene clusters in LUSC. (A) 373 differentially expressed genes (DEGs) between three autophagy clusters were shown in the Venn diagram. (B) Consensus clustering CDF with the number of subtypes $k = 2$ to 9. (C) Delta area curve for clustering, representing the relative change in area under the CDF curve. ((D)–(F)) Unsupervised clustering of prognostic-related DEGs in LUSC and consensus matrices for $k = 2, 4, 5$. (G) PCA for the three gene clusters among LUSC patients. Figure S4: evaluation of the prognosis and immune value of the APS model. (A) Survival analysis of APS in GSE73403 cohort. (B) Survival analysis of APS in GSE157011 cohort. (C) Survival analysis of APS in GSE74777 cohort. (D) Differences in stroma-activated pathways between low and high APS groups. The asterisks represent the $P$ value ($^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$). (E) The difference in the percentage of survival status (Fustat) between the low and high APS groups. (F) The APS in different survival status (Fustat) of LUSC patients. Figure S5: the relationship between APS and clinicopathological characteristics. The clinicopathological features of LUSC include (A) TNM stage, (B) T stage, (C) N stage, and (D) M stage. Figure S6: covariance of principal components. (*Supplementary Materials*)

## References

[1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2021," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 1, pp. 7–33, 2021.

[2] 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] P. Perez-Moreno, E. Brambilla, R. Thomas, and J. C. Soria, "Squamous cell carcinoma of the lung: molecular subtypes and therapeutic opportunities," *Clinical Cancer Research*, vol. 18, no. 9, pp. 2443–2451, 2012.

[4] K. Bogos, Z. Kiss, G. Galffy et al., "Revising incidence and mortality of lung cancer in central europe: an epidemiology review from Hungary," *Frontiers in Oncology*, vol. 9, p. 1051, 2019.

[5] H. Asamura, T. Goya, Y. Koshiishi et al., "A Japanese Lung Cancer Registry study: prognosis of 13, 010 resected lung cancers," *Journal of Thoracic Oncology*, vol. 3, no. 1, pp. 46–52, 2008.

[6] "Targeted drugs fall short in squamous lung cancer," *Cancer Discovery*, vol. 11, p. OF3, 2021.

[7] H. Morishita and N. Mizushima, "Diverse cellular roles of autophagy," *Annual Review of Cell and Developmental Biology*, vol. 35, no. 1, pp. 453–475, 2019.

[8] E. White, "Deconvoluting the context-dependent role for autophagy in cancer," *Nature Reviews Cancer*, vol. 12, no. 6, pp. 401–410, 2012.

[9] R. A. Barnard, D. P. Regan, R. J. Hansen, P. Maycotte, A. Thorburn, and D. L. Gustafson, "Autophagy inhibition delays early but not late-stage metastatic Disease," *Journal of Pharmacology and Experimental Therapeutics*, vol. 358, no. 2, pp. 282–293, 2016.

[10] W. K. K. Wu, S. B. Coffelt, C. H. Cho et al., "The autophagic paradox in cancer therapy," *Oncogene*, vol. 31, no. 8, pp. 939–953, 2012.

[11] D. C. Hinshaw and L. A. Shevde, "The tumor microenvironment innately modulates cancer progression," *Cancer Research*, vol. 79, no. 18, pp. 4557–4566, 2019.

[12] M. O. Li, N. Wolf, D. H. Raulet et al., "Innate immune cells in the tumor microenvironment," *Cancer Cell*, vol. 39, no. 6, pp. 725–729, 2021.

[13] R. Rosenthal, E. L. Cadieux, R. Salgado et al., "Neoantigen-directed immune escape in lung cancer evolution," *Nature*, vol. 567, no. 7749, pp. 479–485, 2019.

[14] Y. Deng, L. Zhu, H. Cai, G. Wang, and B. Liu, "Autophagic compound database: a resource connecting autophagy-modulating compounds, their potential targets and relevant diseases," *Cell Proliferation*, vol. 51, no. 3, Article ID e12403, 2018.

[15] S. Dudoit and J. Fridlyand, "A prediction-based resampling method for estimating the number of clusters in a dataset," *Genome Biology*, vol. 3, no. 7, Article ID RESEARCH0036, 2002.

[16] 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.

[17] M. R. Ferreira, G. A. Santos, C. A. Biagi, W. A. Silva Junior, and W. F. Zambuzzi, "GSVA score reveals molecular signatures from transcriptomes for biomaterials comparison," *Journal of Biomedical Materials Research Part A*, vol. 109, no. 6, pp. 1004–1014, 2021.

[18] K. K. Elam, S. Clifford, D. S. Shaw, M. N. Wilson, and K. Lemery-Chalfant, "Gene set enrichment analysis to create polygenic scores: a developmental examination of aggression," *Translational Psychiatry*, vol. 9, no. 1, p. 212, 2019.

[19] Y. Kim, J. W. Kang, J. Kang et al., "Novel deep learning-based survival prediction for oral cancer by analyzing tumor-infiltrating lymphocyte profiles through CIBERSORT," *OncoImmunology*, vol. 10, no. 1, Article ID 1904573, 2021.

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

[21] B. Zhang, Q. Wu, B. Li, D. Wang, L. Wang, and Y. L. Zhou, "m (6 A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer," *Molecular Cancer*, vol. 19, no. 1, p. 53, 2020.

[22] A. Mayakonda, D. C. Lin, Y. Assenov, C. Plass, and H. P. Koeffler, "Maftools: efficient and comprehensive analysis of somatic variants in cancer," *Genome Research*, vol. 28, no. 11, pp. 1747–1756, 2018.

[23] S. H. Patel, L. M. Poisson, D. J. Brat et al., "T2-FLAIR mismatch, an imaging biomarker for idh and 1p/19q status in lower-grade gliomas: a TCGA/TCIA project," *Clinical Cancer Research*, vol. 23, no. 20, pp. 6078–6085, 2017.

[24] P. Charoentong, F. Finotello, M. Angelova et al., "Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade," *Cell Reports*, vol. 18, no. 1, pp. 248–262, 2017.

[25] A. Galvano, V. Gristina, U. Malapelle et al., "The prognostic impact of tumor mutational burden (TMB) in the first-line management of advanced non-oncogene addicted non-small-cell lung cancer (NSCLC): a systematic review and meta-analysis of randomized controlled trials," *ESMO Open*, vol. 6, no. 3, Article ID 100124, 2021.

[26] A. Friedlaender, T. Nouspikel, Y. Christinat, L. Ho, T. McKee, and A. Addeo, "Tissue-plasma TMB comparison and plasma TMB monitoring in patients with metastatic non-small cell lung cancer receiving immune checkpoint inhibitors," *Frontiers in Oncology*, vol. 10, p. 142, 2020.

[27] J. R. Brahmer, "Harnessing the immune system for the treatment of non-small-cell lung cancer," *Journal of Clinical Oncology*, vol. 31, no. 8, pp. 1021–1028, 2013.

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

[29] J. Ferlay, M. Colombet, I. Soerjomataram et al., "Cancer incidence and mortality patterns in Europe: estimates for 40 countries and 25 major cancers in 2018," *European Journal of Cancer*, vol. 103, pp. 356–387, 2018.

[30] F. Wu, J. Fan, Y. He et al., "Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer," *Nature Communications*, vol. 12, no. 1, p. 2540, 2021.

[31] M. D. Wilkerson, X. Yin, K. A. Hoadley et al., "Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types," *Clinical Cancer Research*, vol. 16, no. 19, pp. 4864–4875, 2010.

[32] X. S. Li, K. C. Nie, Z. H. Zheng et al., "Molecular subtypes based on DNA methylation predict prognosis in lung squamous cell carcinoma," *BMC Cancer*, vol. 21, no. 1, p. 96, 2021.

[33] Y. Xu, Y. She, Y. Li et al., "Multi-omics analysis at epigenomics and transcriptomics levels reveals prognostic subtypes of lung squamous cell carcinoma," *Biomedicine & Pharmacotherapy*, vol. 125, Article ID 109859, 2020.

[34] K. Degenhardt, R. Mathew, B. Beaudoin et al., "Autophagy promotes tumor cell survival and restricts necrosis, inflammation, and tumorigenesis," *Cancer Cell*, vol. 10, no. 1, pp. 51–64, 2006.

[35] M. S. Sosa, P. Bragado, and J. A. Aguirre-Ghiso, "Mechanisms of disseminated cancer cell dormancy: an awakening field," *Nature Reviews Cancer*, vol. 14, no. 9, pp. 611–622, 2014.

[36] L. Chen, G. Ma, P. Wang et al., "Establishment and verification of prognostic model for gastric cancer based on autophagy-related genes," *American Journal of Cancer Research*, vol. 11, no. 4, pp. 1335–1346, 2021.

[37] Z. Deng, X. Li, Y. Shi, Y. Lu, W. Yao, and J. Wang, "A novel autophagy-related IncRNAs signature for prognostic prediction and clinical value in patients with pancreatic cancer," *Frontiers in Cell and Developmental Biology*, vol. 8, Article ID 606817, 2020.

[38] S. Zhu, Q. Wu, B. Zhang et al., "Autophagy-related gene expression classification defines three molecular subtypes with distinct clinical and microenvironment cell infiltration characteristics in colon cancer," *International Immunopharmacology*, vol. 87, Article ID 106757, 2020.

[39] R. Zhang, Q. Zhu, D. Yin et al., "Identification and validation of an autophagy-related lncRNA signature for patients with breast cancer," *Frontiers in Oncology*, vol. 10, Article ID 597569, 2020.

[40] X. Yan, H. H. Wu, Z. Chen et al., "Construction and validation of an autophagy-related prognostic signature and a nomogram for bladder cancer," *Frontiers in Oncology*, vol. 11, Article ID 632387, 2021.

[41] X. Wang, C. Dai, M. Ye, J. Wang, W. Lin, and R. Li, "Prognostic value of an autophagy-related long-non-coding-RNA signature for endometrial cancer," *Aging (Albany NY)*, vol. 13, no. 4, pp. 5104–5119, 2021.

[42] S. Mukhopadhyay, K. K. Mahapatra, P. P. Praharaj, S. Patil, and S. K. Bhutia, "Recent progress of autophagy signaling in tumor microenvironment and its targeting for possible cancer therapeutics," *Seminars in Cancer Biology*, 2021.

[43] P. S. Hegde, V. Karanikas, and S. Evers, "The where, the when, and the how of immune monitoring for cancer immunotherapies in the era of checkpoint inhibition," *Clinical Cancer Research*, vol. 22, no. 8, pp. 1865–1874, 2016.

[44] M. Pickup, S. Novitskiy, and H. L. Moses, "The roles of TGF$\beta$ in the tumour microenvironment," *Nature Reviews Cancer*, vol. 13, no. 11, pp. 788–799, 2013.

[45] O. Tredan, C. M. Galmarini, K. Patel, and I. F. Tannock, "Drug resistance and the solid tumor microenvironment," *JNCI Journal of the National Cancer Institute*, vol. 99, no. 19, pp. 1441–1454, 2007.

[46] M. Tamminga, T. J. N. Hiltermann, E. Schuuring, W. Timens, R. S. Fehrmann, and H. J. Groen, "Immune microenvironment composition in non-small cell lung cancer and its association with survival," *Clinical & Translational Immunology*, vol. 9, no. 6, Article ID e1142, 2020.

[47] R. Li, Y. Lin, Y. Wang et al., "Characterization of the tumor immune microenvironment in lung squamous cell carcinoma using imaging mass cytometry," *Frontiers in Oncology*, vol. 11, Article ID 620989, 2021.

Hindawi
Journal of Oncology
Volume 2022, Article ID 7831001, 11 pages
https://doi.org/10.1155/2022/7831001

*Research Article*

# A Novel Prognostic Nomogram and Risk Classification System for Predicting Cancer-Specific Survival of Postoperative Fibrosarcoma Patients: A Large Cohort Retrospective Study

**Chao Huang, Zhangheng Huang, and Zongke Zhou**

*Department of Orthopedics, West China Hospital of Sichuan University, No. 37 Guoxue Alley, Wuhou District, Chengdu 610041, Sichuan, China*

Correspondence should be addressed to Zongke Zhou; zhouzongke1968@126.com

Received 7 April 2022; Accepted 29 July 2022; Published 27 August 2022

Academic Editor: Xinyi Liu



*Background*. Fibrosarcoma (FS) is a typically invasive sarcoma formed by fibroblasts and collagen fibers. Currently, the standard treatment for FS is the surgical resection, but the high recurrence rate and poor prognosis limit the benefits of postoperative patients. Exploring what factors affect the benefit of postoperative patients is significant for guiding the implementation of surgical resection. Therefore, this study aims to construct a novel nomogram to predict the cancer-specific survival (CSS) of postoperative fibrosarcoma (POFS) patients. *Methods*. The included patients were randomly assigned to the training and validation sets at a ratio of 7 : 3. CSS was indexed as the research endpoint. Firstly, univariate and multivariate Cox regression analyses were used on the training set to determine independent prognostic predictors and build a nomogram for predicting the 1-, 3-, and 5-year CSS of POFS patients. Secondly, the nomogram's discriminative power and prediction accuracy were evaluated by receiver operating characteristic (ROC) and the calibration curve, and a risk classification system for POFS patients was constructed. Finally, the nomogram's clinical utility was evaluated using decision curve analysis (DCA). *Results*. Our study included 346 POFS patients, divided into the training (244) and validation sets (102). Multivariate Cox regression analysis demonstrated that tumor size, SEER stage, and tumor grade were independent prognostic predictors of CSS for POFS patients. They were used to create a nomogram. In the training and validation sets, the ROC curve showed that the 1-, 3-, and 5-year area under the curve (AUC) were higher than 0.700, indicating that the nomogram had good reliability and accuracy. DCA also showed that the nomogram has high application value in clinical practice. *Conclusion*. The larger tumor size, higher tumor grade, and distant metastasis were independently related to the poor prognosis of POFS patients. The nomogram constructed based on the above variables could accurately predict the 1-, 3-, and 5-year CSS of POFS patients. So, the nomogram and risk classification system we built might help make accurate judgments in clinical practice, optimize patient treatment decisions, maximize postoperative benefits, and ultimately improve the prognosis of POFS patients.

## 1. Introduction

Fibrosarcoma (FS) is a malignant mesenchymal tumor, composed of fibroblasts with variable collagen production [1]. All body parts containing fibrous tissue may be the birthplace of FS, but it is more common in head, neck, trunk, and limbs, accounting for approximately 5% of all soft tissue sarcomas [2]. FS could be divided into adult FS and infantile FS according to age. Infantile FS usually occurs in children under 5 years old. Congenital infantile FS is rare, accounting for less than 1% of all childhood cancers. Infantile FS was defined as a moderately malignant and rarely metastatic tumor by the World Health Organization (WHO), while adult FS was classified as a highly malignant tumor [3–5]. Adult FS accounts for approximately 3.6% of all adult sarcomas and mainly occurs in people between 25 and 79 years old, and the incidence of men and women is roughly the same [5, 6]. Other more common subtypes of FS include dermatofibrosarcoma protuberans (DFSP). DFSP mainly occurs in middle-aged men. Generally, DFSP's growth rate is

relatively slow, with a low metastasis rate (<5%), but the local recurrence rate is high (20–50%), especially when the resection is not enough [7, 8].

Although FS is mainly treated by surgical resection, it was reported that 10–20% of patients whose tumors were fully resected will experience recurrence within 5 years, and the prognosis of those patients was worse [9]. Thus, conducting a separate analysis to find the most relevant prognostic factors related to the survival rate of POFS patients and carry out individualized management for postoperative patients to improve the effectiveness of the surgical treatment is necessary. Compared with overall survival (OS), cancer-specific survival (CSS) can provide a closer relationship with tumor-mediated patient prognosis and provide more precise guidance for treating those patients. However, as far as we know, no research has focused on the development of predictive models for CSS in patients with POFS. Therefore, our study aims to find the prognostic predictors related to CSS of POFS patients by analyzing relevant data from the surveillance, epidemiology, and end results (SEER) database and to develop a new nomogram model and a risk classification system to predict the 1-, 3-, and 5-year CSS of POFS patients.

## 2. Methods

*2.1. Database.* Sponsored by the National Cancer Institute, the SEER data set (https://seer.cancer.gov/seerstat/) is a cancer database based on the US population, providing systematic evidence support and valuable direct information for clinicians' practice and medical research. It collected data from 18 registries, nearly 30% of the US population [10]. SEER Stat 8.3.9.2 was used to identify the data of all POFS patients from 1975 to 2016 in SEER database [Incidence-SEER 18 Regs Custom Data (with additional treatment fields), Nov 2018 Sub (1975–2016 varying)]. We obtained access to the SEER database after obtaining permission to access research data files with the reference number 16336-Nov2020. Since the acquired data has no specific personal information disclosed, it does not require the ethics committee's approval and the patient's informed consent. This study was conducted and reported in line with STROCSS 2019 criteria [11].

*2.2. Patient Selection.* The inclusion criteria of our study were as follows: (i) FS is the patient's ICD-O-3 histological type; (ii) complete follow-up data; (iii) primary tumor; (iv) surgery performed, while the exclusion criteria were as follows: (i) it is not the primary tumor; (ii) the information about age, sex, race, marriage, tumor stage, tumor grade, tumor size, surgery, radiotherapy, and chemotherapy was unknown; (iii) survival time is less than one month. Finally, 346 POFS patients were found to be suitable for inclusion in our study.

All included patients were randomly divided into a training set (70%) and a validation set (30%) according to a ratio of 7 : 3. We used the training set to determine independent prognostic predictors and establish the prognostic nomogram for postoperative patients, and the validation set was used to verify the nomogram.

*2.3. Variable Definitions.* Variables included in our study were POFS patients' demographic characteristics (age, race, sex, and marital status), disease characteristics (tumor size, tumor grade, and SEER histological stage), and information on treatment (radiotherapy and chemotherapy). The X-tile software (version 3.6.1) determined the best cut-off values for age and tumor size, and the results showed that the best cut-off values for age were 43 and 71 years, and those for tumor size were 6.4 and 11 cm, respectively (Supplementary File 1) [12]. Sex was divided into male and female, and the race was divided into white, black, and others. Marital status was divided into married and unmarried. Radiotherapy and chemotherapy were divided into yes and no. Tumor grades were divided into grades I, II, III, and IV, and SEER histological stages were classified as local, regional, and distant. Our study's primary endpoint was CSS, which was defined as the time interval between the day of diagnosis and death caused by this tumor alone.

*2.4. Statistical Analysis.* SPSS (version 22.0) and R software (version 4.0.3) was used to do statistical analyses in this study, and a $p$-value of <0.05 was considered statistically significant. First, values were assigned to CSS-related variables (Supplementary File 2). The Kaplan-Meier method was generated to show the statistical difference between the included variables, and univariate Cox regression analysis was performed. Then, the variables with $p$-value <0.05 obtained in the univariate Cox regression analysis were selected and included in the multivariate Cox regression analysis to determine the independent prognostic predictors of CSS for the POFS patients. A nomogram for predicting 1-, 3-, and 5-year CSS was established based on these independent prognostic predictors, and the corresponding point assignments for independent prognostic predictors were also obtained (Supplementary File 3). After that, 1-, 3-, and 5-year calibration curves were established to exhibit the nomogram's correction ability, and a decision curve analysis (DCA) was performed to demonstrate the nomogram's clinical benefits. Meanwhile, receiver operating characteristic (ROC) curves for 1-, 3-, and 5- years CSS were established, and the corresponding area under the curve (AUC) value was used to evaluate the nomogram's discriminative ability. In addition, points assigned to independent prognostic predictors were used to calculate the patients' total points, and the best cut-off value for the total points was obtained using the X-tile software. Then, the enrolled patients were divided into low-, middle-, and high-risk subgroups to create a risk classification system, stratifying the death risk of all POFS patients. Finally, the difference in CSS between the three subgroups was obtained using the Kaplan-Meier method.

## 3. Results

*3.1. Baseline Characteristics.* 346 patients with POFS were enrolled in this study and randomly divided into a training

set (244, 70%) and a validation set (102, 30%), of which 172 (49.7%) were aged 43–71 years; sex and marital status disparity were not apparent; the majority of patients were white (237, 68.5%); the diameter of the tumor was within 6.4 cm in 62.42% of all patients; besides, patients with low-grade (I-II) tumor accounted for 67.63% of the study population and 77.74% of them were diagnosed as localized metastasis; in addition to surgical treatment for those POFS patients, 11.56% of patients underwent chemotherapy, and 33.24% received radiotherapy (Table 1).

*3.2. Identification of Prognostic Predictors for CSS.* Univariate and multivariate Cox regression analyses were performed to explore independent prognostic predictors of CSS for POFS patients. Age, sex, race, radiotherapy, chemotherapy, tumor size, SEER stage, tumor grade, and marital status were included in univariate Cox regression analysis and the Kaplan-Meier method was performed for CSS in POFS patients (Figure 1). The results of univariate Cox regression analysis revealed that age, race, radiotherapy, chemotherapy, tumor size, tumor grade, and tumor stage were identified as CSS-related variables ($p < 0.05$), while sex and marital status had no significant difference ($p > 0.05$) (Table 2). Then, the multivariate Cox regression analysis was performed to eliminate confounding effects between the above CSS-related variables ($p < 0.05$), and the result showed that tumor size, tumor grade, and SEER stage were identified as independent prognostic predictors of CSS for POFS patients (Table 2). POFS patients with larger tumor size, higher tumor grade, and distant metastasis would be associated with poor CSS.

*3.3. Establishment and Verification of the Prognostic Nomogram for the CSS.* To predict the CSS of POFS patients, we developed a nomogram based on all the above independent CSS-related predictors from multivariate Cox regression analysis (Figure 2). The nomogram also endowed each independent prognostic predictor with a point. Adding these points could predict the 1-, 3-, and 5-year CSS of POFS patients. As shown in Figure 2, smaller tumor size, lower tumor grade, and localized metastasis are protective predictors for POFS patients. The poor prognosis of POFS patients included larger tumor size, higher tumor grade, and distant metastasis. The excellent agreement between the predicted results and the actual survival rate of POFS patients was reflected by the 1-, 3-, and 5-year calibration curves (Figure 3). The AUCs for CSS at 1-, 3-, and 5-year in the ROC curve of the training set were 0.879, 0.876, and 0.843, and those in validation set were 0.855, 0.786, and 0.822, respectively (Figure 4), indicating that the nomogram had good reliability and accuracy.

Meanwhile, we also compared the prediction accuracy of a single independent prognostic predictor with the nomogram (Figure 5). As shown in Figure 5, the AUCs of each independent prognostic predictor of CSS in training set at 1-, 3-, and 5-year calibration curves were lower than the nomogram, indicating that the nomogram had better prediction accuracy for CSS, while, in the validation set, nomogram's AUCs in the 1- and 3-yearROC curves were not as large as those of some individual independent predictors, which might be related to the small number of postoperative death patients in the 1- and 3-yearROC curves in the validation set (2/6, 33.3% and 8/36, 25%, respectively), but as the follow-up time increased, the AUC of the nomogram gradually increased and showed better prediction accuracy, as shown in Figure 5(f). Besides, the DCA showed that the nomogram had a high clinical application value and could be used as an effective auxiliary tool in clinical practice to maximize the benefit of postoperative patients (Figure 6).

Table 1: The baseline characteristics of the CSS-related variables of postoperative fibrosarcoma patients.

| Variables | Training set | | Validation set | | Total | |
|---|---|---|---|---|---|---|
| | 244 | 70.00 (%) | 102 | 30.00 (%) | 346 | 100.00 (%) |
| Age (years old) | | | | | | |
| <43 | 98 | 40.16 | 39 | 38.23 | 137 | 39.60 |
| 43–71 | 123 | 50.41 | 49 | 48.04 | 172 | 49.71 |
| >71 | 23 | 9.43 | 14 | 13.73 | 37 | 10.69 |
| Sex | | | | | | |
| Male | 126 | 51.64 | 60 | 58.82 | 186 | 53.76 |
| Female | 118 | 48.36 | 42 | 41.18 | 160 | 46.24 |
| Marital status | | | | | | |
| Unmarried | 127 | 52.05 | 47 | 46.08 | 174 | 50.29 |
| Married | 117 | 47.95 | 55 | 53.92 | 172 | 49.71 |
| Race | | | | | | |
| Black | 49 | 20.08 | 26 | 25.49 | 75 | 21.67 |
| White | 176 | 72.13 | 61 | 59.80 | 237 | 68.50 |
| Other | 19 | 7.79 | 15 | 14.71 | 34 | 9.83 |
| Tumor size (mm) | | | | | | |
| <64 | 150 | 61.48 | 66 | 64.71 | 216 | 62.42 |
| 64–110 | 52 | 21.31 | 24 | 23.53 | 76 | 21.97 |
| >110 | 42 | 17.21 | 12 | 11.76 | 54 | 15.61 |
| SEER stage | | | | | | |
| Localized | 185 | 75.82 | 84 | 82.35 | 269 | 77.74 |
| Regional | 49 | 20.08 | 15 | 14.71 | 64 | 18.50 |
| Distant | 10 | 4.10 | 3 | 2.94 | 13 | 3.76 |
| Tumor grade | | | | | | |
| Grade I | 100 | 40.98 | 30 | 29.41 | 130 | 37.57 |
| Grade II | 71 | 29.10 | 33 | 32.35 | 104 | 30.06 |
| Grade III | 40 | 16.39 | 21 | 20.59 | 61 | 17.63 |
| Grade IV | 33 | 13.53 | 18 | 17.65 | 51 | 14.74 |
| Radiotherapy | | | | | | |
| No | 162 | 66.40 | 69 | 67.65 | 231 | 66.76 |
| Yes | 82 | 33.60 | 33 | 32.35 | 115 | 33.24 |
| Chemotherapy | | | | | | |
| No | 212 | 86.90 | 94 | 92.16 | 306 | 88.44 |
| Yes | 32 | 13.10 | 8 | 7.84 | 40 | 11.56 |

*3.4. Risk Classification System.* We constructed a disease risk classification system for POFS patients based on the tumor size, tumor grade, and SEER historical stage to further verify the nomogram from different dimensions. And we calculated the patient's total point based on the assignment of the nomogram for the included independent predictors. The best cut-off values for the total point were 87 and 156 using the X-tile software (Supplementary File 1), and the patients

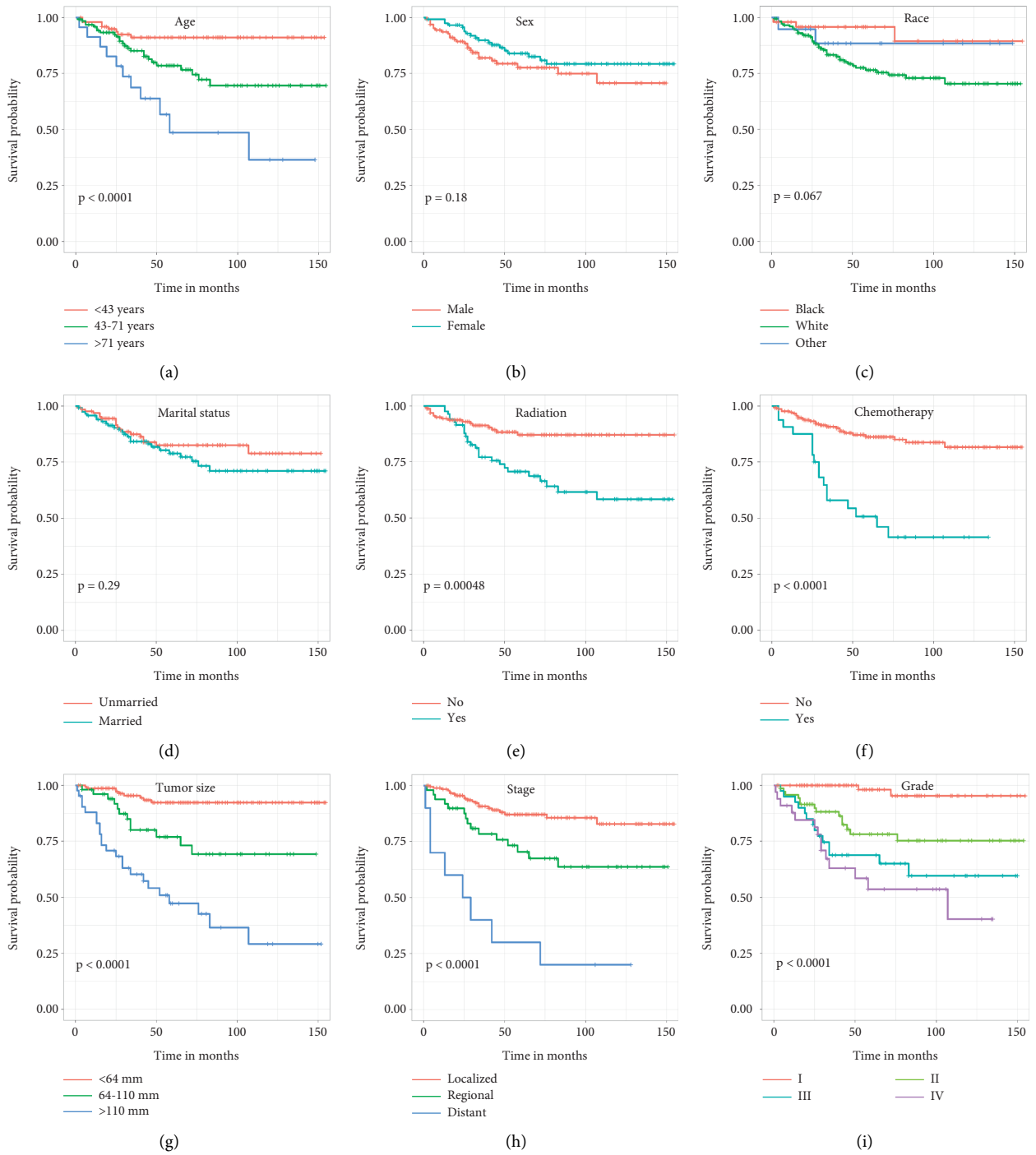



Figure 1: Kaplan–Meier curves of variables were performed for cancer-specific. Survival (CSS) in postoperative fibrosarcoma (POFS) patients. (a) Age, (b) sex, (c) race, (d) marital status, (e) radiation, (f) chemotherapy, (g) tumor size, (h) tumor stage, and (i) tumor grade.

were further divided into three different death risk classification subgroups based on the total point: low- (<87), middle- (87–156), and high- (>156) risk subgroups, and a Kaplan-Meier curve was drawn (Figure 7). As shown in Figure 7, whether it is a training set or a validation set, risk classification system can efficiently divide POFS patients into three subgroups with significant differences ($p < 0.05$), indicating that the nomogram has a significant predictive value in the prognosis of the subgroups POFS patients.

## 4. Discussion

FS is defined as fibroblast/myofibroblastic sarcoma and a rare high-grade malignant tumor derived from mesenchymal cells according to the WHO classification of soft tissue sarcoma [9]. The overall 5-year survival rate of FS is about 40–60%, and the recurrence rate is between 12 and 79% [13, 14]. Numerous studies have shown that the unfavorable prognostic factors of FS included the following

Table 2: The univariate and multivariate Cox regression analyses of the CSS-related variables of postoperative fibrosarcoma patients.

| Variables | Univariate analysis | | Multivariate analysis | |
|---|---|---|---|---|
| | Hr (95% CI) | $p$ value | Hr (95% CI) | $p$ value |
| Age (years) | | | | |
| <43 | References | | | |
| 43–71 | 2.873 (1.294–6.379) | 0.01 | | |
| >71 | 6.866 (2.755–17.113) | ≤0.001 | | |
| Sex | | | | |
| Male | References | | | |
| Female | 0.667 (0.367–1.214) | 0.185 | | |
| Marital status | | | | |
| Unmarried | References | | | |
| Married | 1.376 (0.758–2.499) | 0.295 | | |
| Race | | | | |
| Black | References | | | |
| White | 3.341 (1.032–10.814) | 0.044 | | |
| Other | 1.597 (0.267–9.563) | 0.608 | | |
| Tumor size (mm) | | | | |
| <64 | References | | References | |
| 64–110 | 4.016 (1.692–9.533) | 0.002 | 3.997 (1.640–9.745) | 0.002 |
| >110 | 11.567 (5.347–25.025) | ≤0.001 | 8.140 (3.624–18.279) | ≤0.001 |
| SEER stage | | | | |
| Localized | References | | References | |
| Regional | 2.544 (1.130–4.939) | 0.006 | 1.549 (0.751–3.195) | 0.236 |
| Distant | 10.234 (4.528–23.133) | ≤0.001 | 4.598 (1.929–10.962) | ≤0.001 |
| Tumor grade | | | | |
| Grade I | References | | References | |
| Grade II | 9.815 (2.230–43.193) | 0.003 | 9.844 (2.229–43.480) | 0.003 |
| Grade III | 17.895 (4.066–78.755) | ≤0.001 | 11.704 (2.631–52.056) | ≤0.001 |
| Grade IV | 24.892 (5.655–109.574) | ≤0.001 | 16.751 (3.731–75.206) | ≤0.001 |
| Radiotherapy | | | | |
| No | References | | | |
| Yes | 2.811 (1.532–5.159) | ≤0.001 | | |
| Chemotherapy | | | | |
| No | References | | | |
| Yes | 4.505 (2.453–8.274) | ≤0.001 | | |

aspects: (i) large tumor size (>5 cm in diameter); (ii) high histological grade; (iii) deep tumor; (iv) a decrease of collagen fibers; (v) massive mitotic phase (>20/10 high power field); and (vi) massive tissue necrosis (>50%) [6, 9, 13]. It was not until the last two years that articles about the prediction of the survival rate of FS were reported. In 2020, Xiang et al. established a nomogram based on 663 patients with FS through the SEER database, which showed that surgery, sex, tumor size, SEER stage, and pathological grade were the independent prognostic predictors of OS and CSS of FS patients [15]. Subsequently, Yang et al. constructed a novel nomogram based on 357 elderly FS patients, which showed that those patients' OS was related to age, tumor grade, surgery, chemotherapy, and SEER stage [16]. Surgical resection is essential for the treatment of FS. However, recurrence would occur in patients with fully resected tumors, causing poor progress. Thus, exploring prognostic factors that affect the survival benefits of surgical patients is necessary and will help evaluate and determine whether the patients need surgery and ensure the maximum benefits of surgery to FS patients.

Nomogram is a model for multi-index joint diagnosis or prediction of disease incidence or progression and is widely used in tumor diseases. The advantage of the nomogram is to simplify the complex statistical prediction model involving a large number of variables into a single short numerical estimation model to predict the probability of an event [17]. Each independent risk factor included in the model is assigned a value to evaluate the impact of the factor on the occurrence of the event. Therefore, when predicting the survival rate of cancer patients, nomograms can help clinicians optimize treatment plans for specific individual variables. Meanwhile, compared with the traditional American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) staging system, the nomogram has better performance and has the unique advantage of being a tailor-made survival prediction model.

Our study selected 346 POFS patients from the SEER database, and tumor size, SEER stage, and tumor grade were identified as independent prognostic predictors for CSS of POFS patients according to the multivariate Cox regression analysis. Therefore, we established a nomogram model for

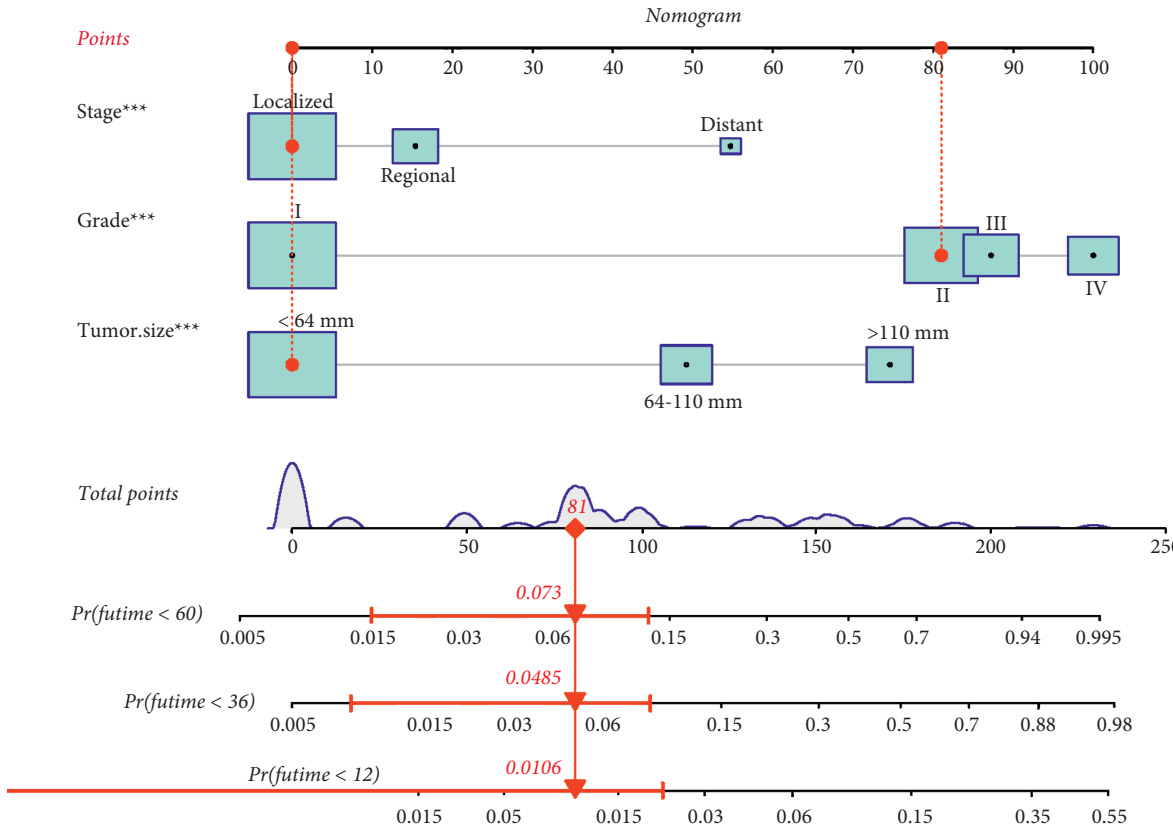



Figure 2: The nomogram model predicts the 1-, 3-, and 5-year CSS of POFS patients.

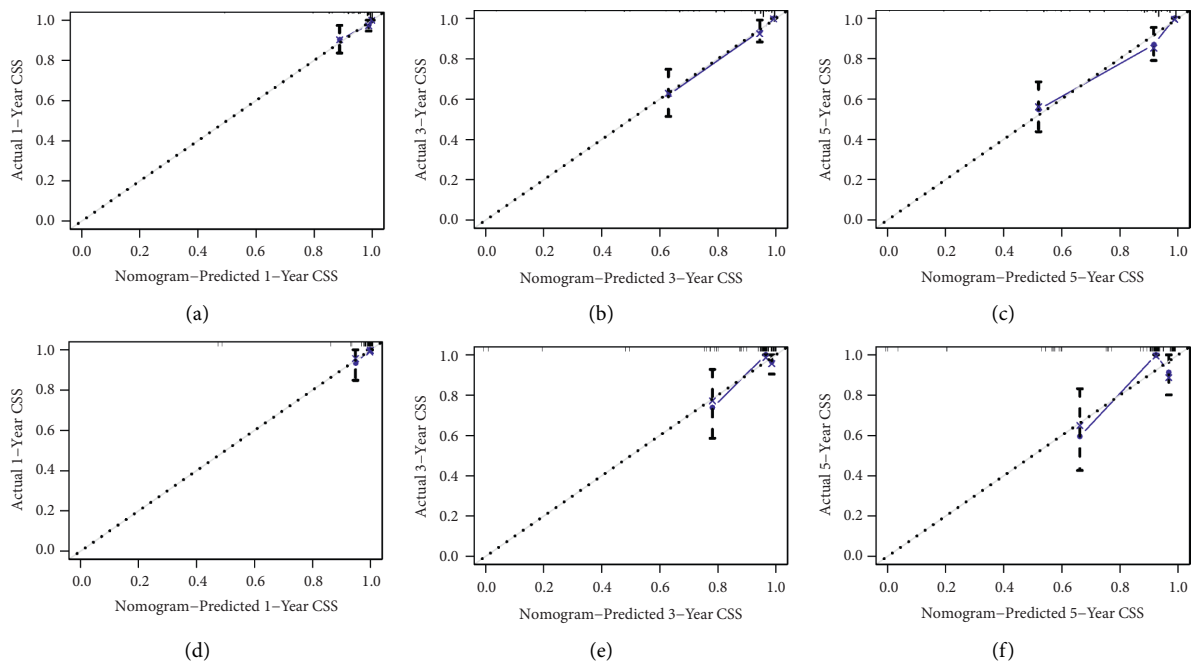



Figure 3: The training and the validation sets' calibration curves in our study. The nomogram's calibration curves for the 1-, 3-, and 5-year CSS prediction of POFS patients in the training set (a–c) and the validation set (d–f).

predicting the 1-, 3-, and 5-year CSS of POFS patients. Both the training and the validation sets showed that the nomogram had reasonable discrimination. There is no significant deviation between the actual survival rate and the predicted survival rate. The nomogram can be used as a practical clinical prediction tool and applied in the clinic.

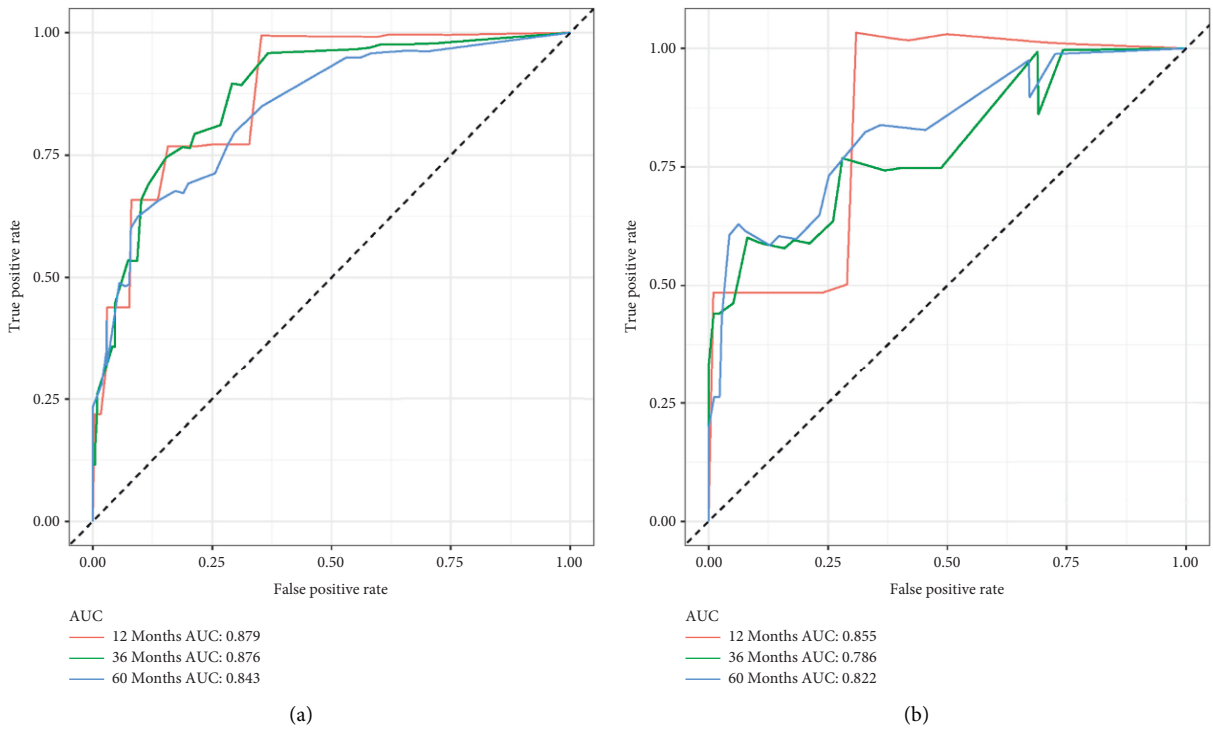



Figure 4: 1-, 3-, and 5-year receiver operating characteristic (ROC) curves in the training (a) and validation (b) sets of POFS patients.

Besides, the constructed risk classification system can effectively divide the patients in the training and validation sets into low-, middle-, and high-risk subgroups with significant differences ($p < 0.05$), which can achieve better patient risk differentiation and effective intervention treatment. Currently, the specific etiology of FS has not yet been definitively concluded, and FS is mainly treated by surgical resection [7]. The principle of surgical resection is radical resection or barrier resection, with negative margins in all directions. The surgical procedure varies according to the location, size, and degree of the tumor malignancy. If the tumor was located in the extremities, some bone might need to be removed and replaced with a prosthesis or bone graft. When the tumor involves the nerves and blood vessels of the limb, the limb must be amputated. However, surgical treatment has major drawbacks for the removal of giant tumors. When the surgical margin is not complete, the residual lesions will lead to local tumor recurrence, further leading to tumor metastasis and a poor prognosis for patients after FS. In a single-center study carried out by Bahrami and Andrew, the 2-year OS rate of FS patients was <70%, while that for 5-year was <55% [14].

Since FS is a malignant tumor, does a giant tumor predict a poor prognosis? According to the researches of Zhang et al. and Sulkowski et al., tumor size had no differential influence on the survival prognosis of FS patients with ovarian and infantile, respectively [18, 19]. However, Ma et al. found that a tumor size more than 5 cm and a Ki-67 index over 30% were associated with poor OS in patients with primary intracranial FS [20]. Furthermore, according to a study by Xiang et al., tumor size was the most important factor that determines CSS in FS patients [15]. The possible explanation is that there are certain differences in the conclusions drawn for a specific population, the location of the disease, and the number of enrolled patients, but they can also assist in disease management. Our research also concluded that tumor size was associated with the CSS of POFS patients based on a larger number of patients, so it is meant to include tumor size in the survival prediction of POFS patients. In addition, tumor grade is closely associated with the prognosis of soft tissue sarcoma. According to reports, the 10-year mortality rate for low-grade tumors is 40%, while that for high-grade tumors is as high as 70% [2]. We have also reached a similar conclusion that patients with high-grade tumors have a worse prognosis than patients with low-grade tumors. At the same time, the SEER stage is also closely associated with the prognosis of soft tissue sarcoma. The prognosis of patients with distant metastasis is significantly worse than that of local or regional metastasis, illustrating the importance of early diagnosis and surgical intervention for FS patients.

Adjuvant radiotherapy and chemotherapy for soft tissue sarcoma are controversial and are not standard treatments for these tumors [21]. Andrew L. Folpe found that almost 80% of adult FSs were high-grade (FNCLCC grade 2–3) malignant tumors, with one out of every four low-grade lesions progressing to high-grade sarcoma in local recurrence [14]. At the same time, 9–63% of adult FS patients who had hematological spread and metastasis, including patients whose tumors were fully resected, would experience recurrence, resulting in a poor prognosis [4, 9]. Although radiotherapy and chemotherapy have a low response rate for FS, they are still utilized as neoadjuvant/adjuvant tumor therapy, and patients with high-grade FS who are at risk of

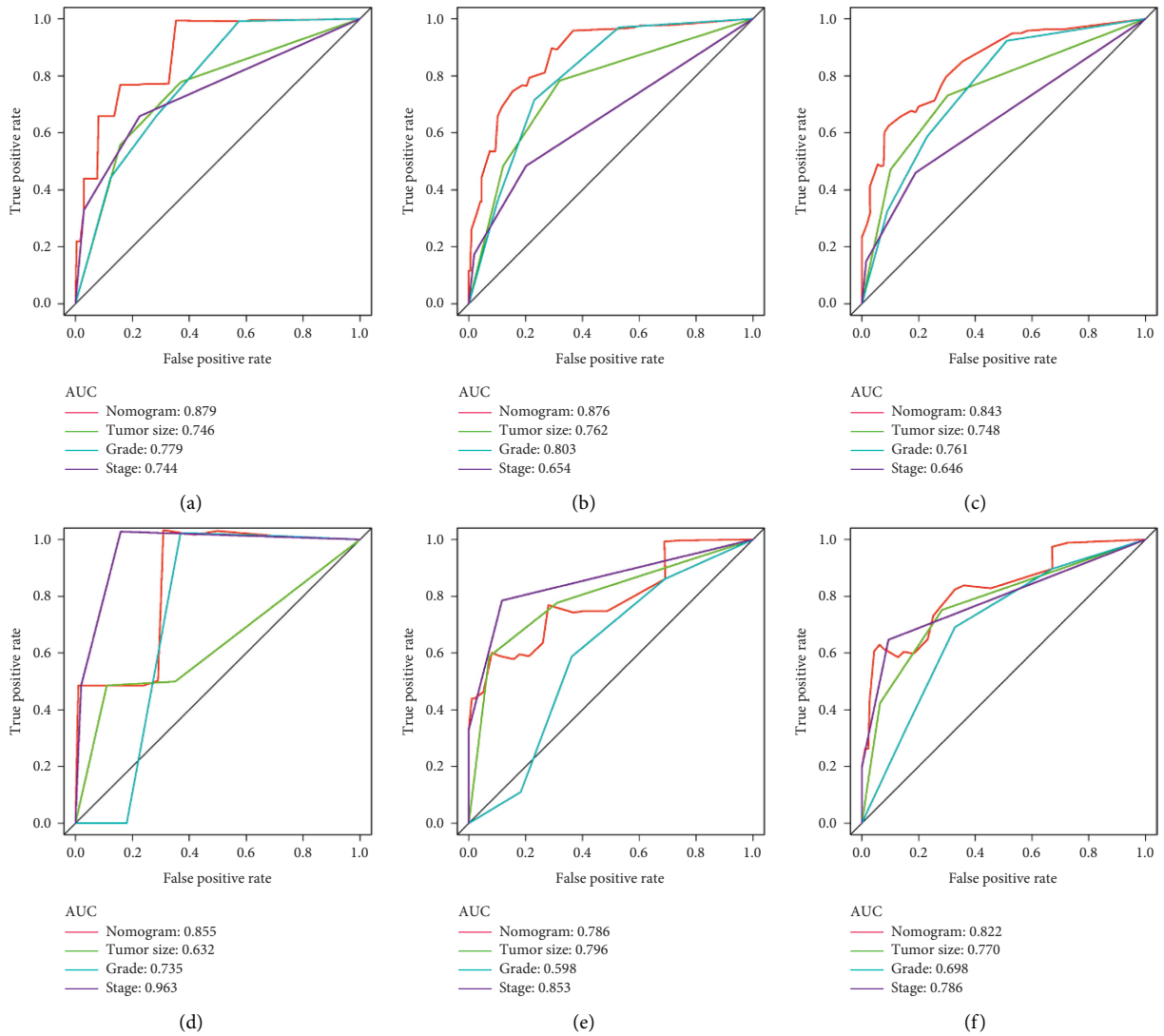



Figure 5: Comparison of the prediction accuracy between the nomogram model and independent predictors in our study. The nomogram and all independent predictors' ROC curves at 1- (a), 3- (b), and 5-year (c) in the training set and at 1- (d), 3- (e), and 5-year (f) in the validation set.

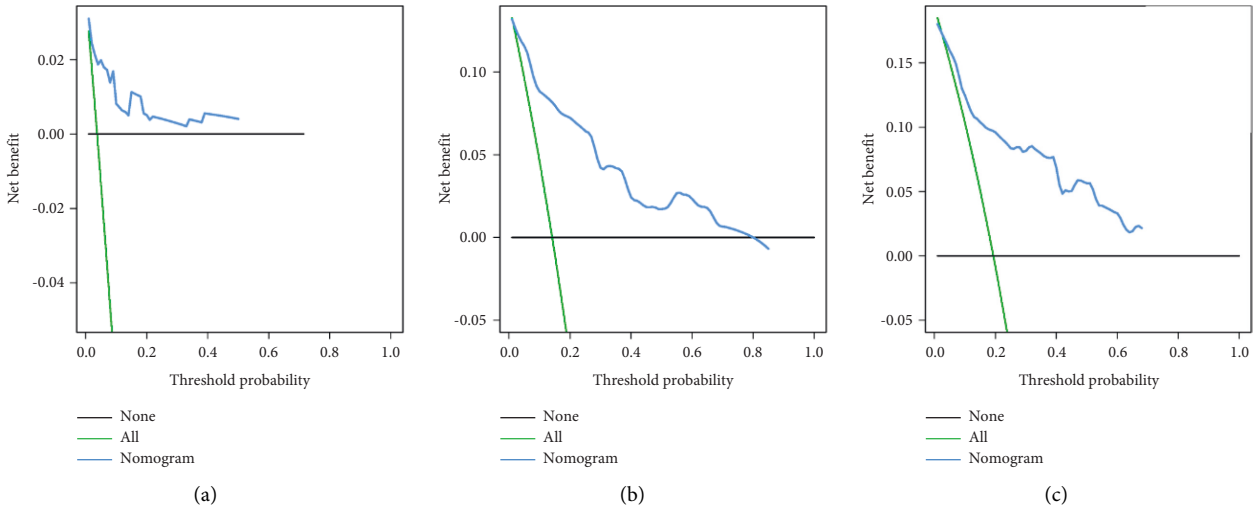



Figure 6: Continued.

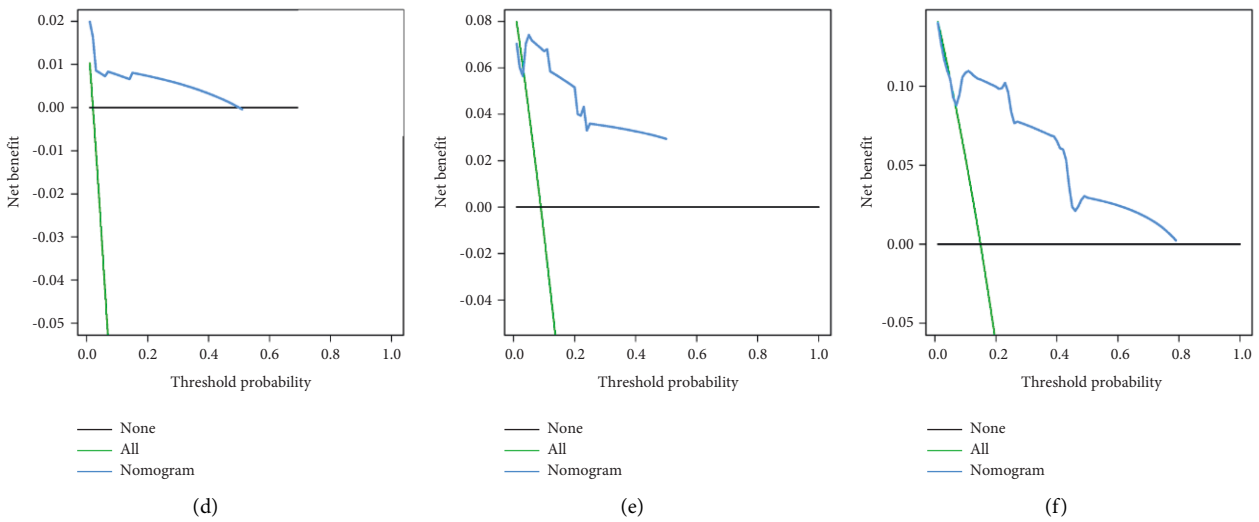



FIGURE 6: The training and the validation sets' decision curve analysis (DCA) in our study. DCA of the nomogram for predicting the 1- (a), 3- (b), and 5- year (c) CSS in the training set and the 1- (d), 3- (e), and 5- year (f) CSS in the validation set of POFS patients.

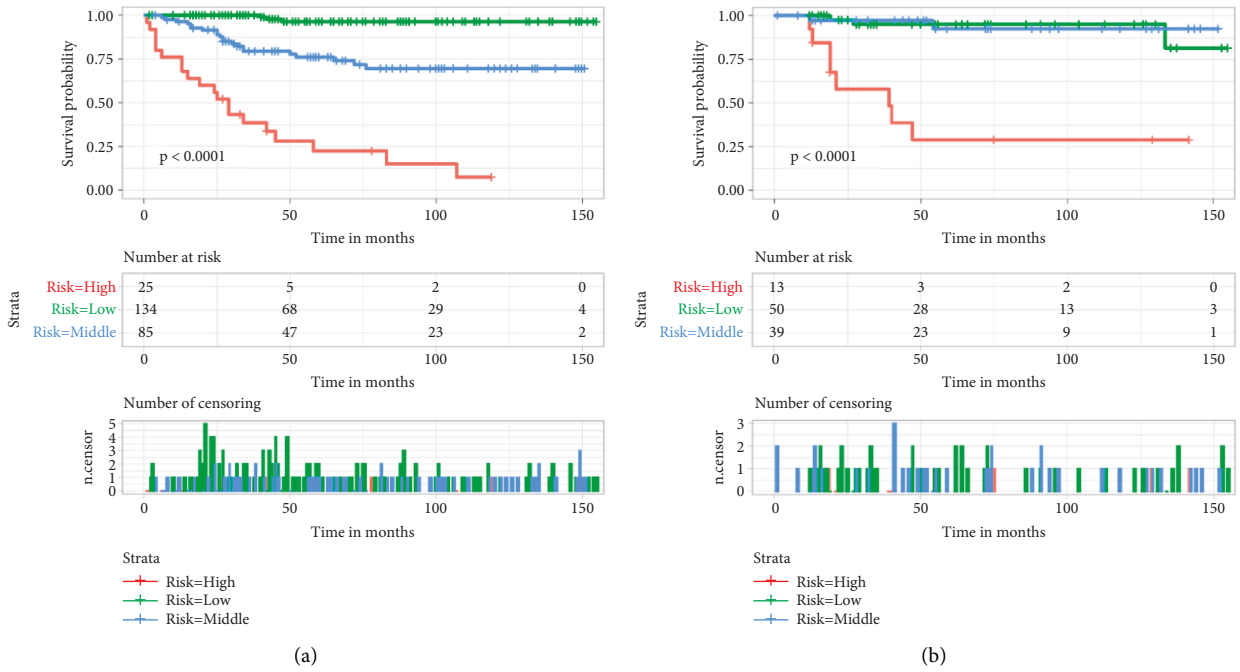



FIGURE 7: The training and the validation sets' Kaplan–Meier survival analysis in our study. Patients with lower risk scores showed a better prognosis than patients with high-risk scores in the training set (a) and the validation set (b) for the CSS of POFS patients in our cohort retrospective study.

metastasis are the most likely to benefit from adjuvant therapy [13, 22]. Muehlhofer et al. published a retrospective single-center investigation spanning more than 15 years, which showed that soft tissue sarcoma patients had a 5-year OS rate of 68.9%. Radiation therapy type (adjuvant or neoadjuvant radiation therapy) had no influence on survival [23]. However, data from 6,960 soft tissue sarcoma patients in the SEER database suggested that radiotherapy was related to the improved survival rate of high-grade tumors patients [24]. Augsburger et al. believed that for FS with deep location, high-grade, and tumor size beyond 5 cm, performing radiotherapy after R0 resection was strongly recommended. However, the necessity of adjuvant radiotherapy has not yet been determined for other types of tumor classification, size, and location [6, 25]. As for chemotherapy, it targets and kills rapidly dividing and proliferating cells, such as malignant tumor cells. Doxorubicin and other chemotherapy drugs are the primary drugs used for those patients. Ma et al. found that giving Apatinib to recurrent FS patients with high VEGF-2 expression could

provide short-term benefits and reduce the incidence of adverse reactions. However, the number of FS patients who do not respond well to chemotherapy was large, limiting the therapeutic effectiveness of chemotherapy drugs [26]. Interestingly, in our study, chemotherapy and radiotherapy were not independent prognostic indicators of CSS in POFS patients.

The SEER database has sufficient sample data of cancer patients to ensure the reliability of the research conclusions. No nomogram, however, can precisely quantify the impact of predictors on prognosis 100% of the time. There are some drawbacks in our nomogram: (i) selection bias is unavoidable in a clinical retrospective study; (ii) the tumor size is a crucial variable included in our study, but the tumor size has only been recorded in the SEER database since 2004, shortening the data period and the number of patients. Therefore, a longer time frame and larger population may help further to improve the credibility and persuasiveness of model predictions. At the same time, the lack of some essential variables in the SEER database also limits the use of the model, such as surgical margin status; (iii) more data from other research centers for external verification will increase the applicability and accuracy of the new nomogram.

## 5. Conclusion

In conclusion, soft tissue sarcomas involving FS are a rare and aggressive variety. In the management of FS patients, surgical resection is crucial. By analyzing 346 POFS patients with complete data in the SEER database, it is concluded that tumor size, SEER stage, and tumor stage are independent prognostic predictors for CSS in POFS patients. Based on the three variables, a nomogram and a risk classification system were built to predict the 1-, 3-, and 5-year CSS of POFS patients. They were helpful for the treatment decision-making, surveillance, and counseling, thus maximizing the benefit of POFS patients.

## Abbreviations

SEER: Surveillance, Epidemiology, and end results
POFS: Postoperative fibrosarcoma
WHO: World Health Organization
DFSP: Dermatofibrosarcoma protuberans
OS: Overall survival
CSS: Cancer-specific survival
ROC: Receiver operating characteristic
DCA: Decision curve analysis
AUCs: Area under the curves
TNM: Tumor-node-metastasis.

## Data Availability

Publicly available datasets were analyzed in this study. These data can be found in the SEER dataset repository (https://seer.cancer.gov/).

## Disclosure

The paper was not commissioned; it was externally peer-reviewed.

## Conflicts of Interest

The authors declare no conflicts of interests.

## Authors' Contributions

C H and ZH H designed the study, performed the literature review, extracted the data, and analyzed the pooled data. C H drew the figures and organized the tables. ZK *Z* provided critical comments and revised the manuscript. All authors read and approved the final manuscript. C H and ZH H contributed equally to this work.

## Acknowledgments

The authors are thankful for the contribution of the SEER database and the 18 registries supplying cancer research information and thank all colleagues involved in the study for their contributions.

## Supplementary Materials

Supplementary Figure 1: according to the X-tile software, the best cut-off values for the age were determined to be 43 and 71 (years). Supplementary Figure 2: according to the X-tile software, the best cut-off values for the tumor size were determined to be 64 and 110 (mm). Supplementary Figure 3: according to the X-tile software, the best cut-off values for the overall survival score were determined to be 87 and 156. Table S1: the values assigned to CSS-related variables in our study. Table S2: the detailed scores of independent prognostic factors in the CSS nomogram. (*Supplementary Materials*)

## References

[1] H. Wang, P. Nie, C. Dong et al., "CT and MRI findings of soft tissue adult fibrosarcoma in extremities," *BioMed Research International*, vol. 2018, Article ID 6075705, 7 pages, 2018.

[2] C. Wibmer, A. Leithner, N. Zielonke, M. Sperl, and R. Windhager, "Increasing incidence rates of soft tissue sarcomas? A population-based epidemiologic study and literature review," *Annals of Oncology*, vol. 21, no. 5, pp. 1106–1111, 2010.

[3] V. Y. Jo and C. D. Fletcher, "WHO classification of soft tissue tumours: an update based on the 2013 (4th) edition," *Pathology*, vol. 46, no. 2, pp. 95–104, 2014.

[4] C. Fisher, E. V. D. Berg, and W. M. Molenaar, *Adult Fibrosarcoma. World Health Organization Classification of Tumours*, 2002.

[5] J. R. Toro, L. B. Travis, H. J. Wu, K. Zhu, C. D. Fletcher, and S. S. Devesa, "Incidence patterns of soft tissue sarcomas, regardless of primary site, in the surveillance, epidemiology and end results program, 1978–2001: an analysis of 26, 758

cases," *International Journal of Cancer*, vol. 119, no. 12, pp. 2922–2930, 2006.

[6] D. Augsburger, P. J. Nelson, T. Kalinski et al., "Current diagnostics and treatment of fibrosarcoma -perspectives for future therapeutic targets and strategies," *Oncotarget*, vol. 8, no. 61, pp. 104638–104653, 2017.

[7] T. Iwasaki, H. Yamamoto, and Y. Oda, "Current update on the molecular biology of cutaneous sarcoma: dermatofibrosarcoma protuberans," *Current Treatment Options in Oncology*, vol. 20, no. 4, 2019.

[8] C. A. Liang, A. Jambusaria-Pahlajani, P. S. Karia, R. Elenitsas, P. D. Zhang, and C. D. Schmults, "A systematic review of outcome data for dermatofibrosarcoma protuberans with and without fibrosarcomatous change," *Journal of the American Academy of Dermatology*, vol. 71, no. 4, pp. 781–786, 2014.

[9] L. Guillou and A. L. Folpe, "Fibroblastic and fibrohistiocytic tumors," *Bone Soft Tissue Pathology*, 2010.

[10] Z. Lin, S. Yan, J. Zhang, and Q. Pan, "A nomogram for distinction and potential prediction of liver metastasis in breast cancer patients," *Journal of Cancer*, vol. 9, no. 12, pp. 2098–2106, 2018.

[11] R. Agha, A. Abdall-Razak, E. Crossley et al., "STROCSS 2019 Guideline: strengthening the reporting of cohort studies in surgery," *International Journal of Surgery*, vol. 72, pp. 156–165, 2019.

[12] R. L. Camp, M. Dolled-Filhart, and D. L. Rimm, "X-tile: a new bio-informatics tool for biomarker assessment and outcome-basedcut-point optimization," *Clinical Cancer Research*, vol. 10, no. 21, pp. 7252–7259, 2004.

[13] K. Thway, "Pathology of soft tissue sarcomas," *Clinical Oncology*, vol. 21, no. 9, pp. 695–705, 2009.

[14] A. Bahrami and A. L. Folpe, "Adult-type fibrosarcoma: a reevaluation of 163 putative cases diagnosed at a single institution over a 48-year period," *The American Journal of Surgical Pathology*, vol. 34, no. 10, pp. 1504–1513, 2010.

[15] G. H. Xiang, J. J. Zhu, C. R. Ke et al., "Nomograms predict overall survival and cancer-specific survival in patients with fibrosarcoma: a SEER-based study," *Journal of Oncology*, vol. 2020, Article ID 8284931, 9 pages, 2020.

[16] F. Yang, H. Xie, and Y. Wang, "Prognostic nomogram and a risk classification system for predicting overall survival of elderly patients with fibrosarcoma: a population-based study," *JAMA Oncology*, vol. 2021, Article ID 9984217, 9 pages, 2021.

[17] Q. Guo, Y. Wang, J. An, S. Wang, X. Dong, and H. Zhao, "A prognostic model for patients with gastric signet ring cell carcinoma," *Technology in Cancer Research and Treatment*, vol. 20, Article ID 15330338211027912, 2021.

[18] Z. Zhang, A. Yao, Q. Xin, Y. Shi, and A. Zhang, "Nuclear atypia is a necessary factor for diagnosis of primary ovarian fibrosarcoma: a case report and literature review," *Gynecologic and Obstetric Investigation*, vol. 85, no. 5, pp. 437–445, 2020.

[19] J. P. Sulkowski, M. V. Raval, and M. Browne, "Margin status and multimodal therapy in infantile fibrosarcoma," *Pediatric Surgery International*, vol. 29, no. 8, pp. 771–776, 2013.

[20] X. J. Ma, D. Li, H. Li et al., "The clinicoradiological features and surgical outcomes of primary intracranial fibrosarcoma: a single-institute experience with a systematic review," *Neurosurgical Review*, vol. 44, no. 1, pp. 543–553, 2021.

[21] L. R. Randall, *Approach to the Diagnosis of Bone and Soft Tissue Tumors—Clinical, Radiologic, and Classification Aspects*, 2010.

[22] J. E. Tepper and H. D. Suit, "Radiation therapy of soft tissue sarcomas," *Cancer*, vol. 55, pp. 2273–2277, 1985.

[23] H. M. L. Muehlhofer, B. Schlossmacher, U. Lenze et al., "Oncological outcome and prognostic factors of surgery for soft tissue sarcoma after neoadjuvant or adjuvant radiation therapy: a retrospective analysis over 15 years," *Anticancer Research*, vol. 41, no. 1, pp. 359–368, 2021.

[24] M. Sumi, "Radiation therapy for management of soft tissue sarcomas," *Cancer and Science*, vol. 43, no. 1, pp. 39–43, 2016.

[25] C. V. Sole, F. A. Calvo, A. Polo et al., "Anticipated intraoperative electron beam boost, external beam radiation therapy, and limb-sparing surgical resection for patients with pediatric soft-tissue sarcomas of the extremity: a multicentric pooled analysis of long-term outcomes," *International Journal of Radiation Oncology, Biology, Physics*, vol. 90, no. 1, pp. 172–180, 2014.

[26] S. Frustaci, A. De Paoli, E. Bidoli et al., "Ifosfamide in the adjuvant therapy of soft tissue sarcomas," *Oncology*, vol. 65, pp. 80–84, 2003.

Hindawi
Journal of Oncology
Volume 2022, Article ID 2513813, 25 pages
https://doi.org/10.1155/2022/2513813

*Research Article*

# M$^6$A Modifier-Mediated Methylation Characterized by Diverse Prognosis, Tumor Microenvironment, and Immunotherapy Response in Hepatocellular Carcinoma

**Fei Liu,[1] Xinyue Zhang,[2] Ziyu Liu,[3] Weiye Cai,[1] Chao Song,[1] Yan Jiang,[4] Ji Yin,[4] Zongchao Liu ,[1] and Chenyi Huang [1]**

[1]*Department of Orthopedics, The Affiliated Traditional Chinese Medicine Hospital of Southwest Medical University, Luzhou 646000, Sichuan, China*
[2]*College of Integrated Chinese and Western Medicine, The Affiliated Traditional Chinese Medicine Hospital of Southwest Medical University, Luzhou 646000, Sichuan, China*
[3]*Center for Phenomics of Traditional Chinese Medicine, The Affiliated Traditional Chinese Medicine Hospital of Southwest Medical University, Luzhou 646000, Sichuan, China*
[4]*Department of Otorhinolaryngology, The Affiliated Traditional Chinese Medicine Hospital of Southwest Medical University, Luzhou 646000, Sichuan, China*

Correspondence should be addressed to Zongchao Liu; 565409672@qq.com and Chenyi Huang; xnykdfszyyhcy@swmu.edu.cn

Received 24 April 2022; Revised 31 May 2022; Accepted 5 July 2022; Published 16 August 2022

Academic Editor: Dechao Bu



*Objective*. Emerging evidence highlights the clinical implications of N$^6$-methyladenosine (m$^6$A) modification in HCC. Yet, the roles of m$^6$A modification in modulating cancer immunity and shaping tumor microenvironment (TME) are undefined in hepatocellular carcinoma (HCC). *Methods*. Here, m$^6$A modification classification was determined for HCC through 23 m$^6$A modifier levels by employing consensus clustering approach. Prognosis analysis was presented for comparing the differences in survival outcomes. The ssGSEA and ESTIMATE approaches were adopted for evaluating the abundances of tumor-infiltrating immune cell populations. The m$^6$A scoring system was computed for reflecting m$^6$A modification classification via PCA algorithm. *Results*. Three m$^6$A modifier-mediated modification patterns were established among HCC specimens, which were characterized by different prognosis, signaling pathways, and TME features. After extracting m$^6$A phenotype-associated DEGs, we determined m$^6$A scores in individual HCC and stratified patients into high- and low-score groups. Patients with low m6A score displayed the survival advantage and higher sensitivity to gemcitabine. Moreover, those with low m$^6$A score possessed the better anti-PD-1/PD-L1 therapeutic response in the IMvigor210 immunotherapy cohort. *Conclusion*. Our findings highlighted that m$^6$A modification exerted a nonnegligible role in remodeling diverse and complex TME. Quantification of the m$^6$A modification patterns of individual HCC may enhance the comprehension of TME features and facilitate immunotherapeutic plans.

## 1. Introduction

Hepatocellular carcinoma (HCC) represents a complex neoplasm with multiple etiologies, comprising 75% to 85% of liver cancer cases [1]. Over 1 million HCC patients will die from HCC in 2030, as estimated by the World Health Organization [2]. Early-stage HCC patients suitably receive curative therapy including resection, ablation, and transplantation, with expected 5-year survival rate up to 60% to 80% [3]. Nevertheless, less than 20% patients are eligible for curative therapy [4]. Intermediate-stage patients usually experience locoregional therapy [5]. Meanwhile, systemic therapy is reserved for advanced patients. For instance, sorafenib is the first systemic agent with efficacy for advanced HCC [6]. In recent years, immunotherapy like antiprogrammed cell death-1 (anti-PD-1), anti-PD-1 ligand (anti-PD-L1), and anti-cytotoxic T-lymphocyte antigen-4 (anti-CTLA-4) that may activate the host's natural

defense system, and identify and eliminate the tumor cells [7], have emerged as a prospective alternative treatment strategy against advanced HCC with durable responses [8]. Despite this, only a minority of patients benefit from the immunotherapy [9]. Hence, it urgently demands novel therapeutic predictors for identifying the ideal HCC subgroups for immunotherapy.

$N^6$-methyladenosine ($m^6A$) is the most abundant messenger RNA (mRNA) modification that occurs in humans, occupying 0.1% to 0.4% total adenosine residues [10]. The $m^6A$ modification represents a dynamic reversible process in humans [11]. Hence, exploring the regulatory genes may assist to uncover the roles and mechanisms of $m^6A$ modification at posttranscriptional levels. Emerging evidences have confirmed that deregulation and genetic alterations of $m^6A$ modifiers contribute to HCC initiation and progression [12]. For instance, $m^6A$ reader YTHDF1 accelerates HCC progression via inducing FZD5 mRNA translation with an $m^6A$-dependent manner [13]. $M^6A$ eraser ALKBH5 inhibits malignancy of HCC through $m^6A$-dependent epigenetic suppression of LYPD1 [14]. $M^6A$ writer KIAA1429 facilitates migration and invasion of HCC through elevating $m^6A$-mediated ID2 level [15]. HCC progression represents a multistep event, comprising the genetic and epigenetic alterations within tumor cells and the surrounding tumor microenvironment (TME) [16]. Cancer cells elicit various biological behavior alterations via the direct or indirect interplay with TME [17]. The in-depth comprehending of the diverse and complex TME may reveal its key roles in tumor development, immune escape, and immunotherapeutic responsiveness [18]. Emerging evidence suggests that TME is specially correlated to $m^6A$ modification [19]. For instance, $m^6A$ eraser ALKBH5 may enhance the effects of anti-PD-1 agent through modulating lactate accumulation along with immunosuppressive cell populations in the TME [20]. Inhibiting $m^6A$ writers METTL3/14 may increase the responsiveness to anti-PD-1 therapy [21]. Nevertheless, above findings are limited to one or two $m^6A$ modifiers due to limited technology. Hence, comprehensively discerning the TME traits modulated by different $m^6A$ modifiers may enhance the cognition of antitumor immunity.

Herein, we presented an overall evaluation concerning the interactions of $m^6A$ modification with TME traits through integration of the HCC transcriptomic and genomic profiles from public data sets. We established three $m^6A$ modification patterns with diverse outcomes and TME. Also, an $m^6A$ scoring system was proposed for quantifying the $m^6A$ modification patterns, which could predict survival outcomes and immunotherapy responses. Thus, $m^6A$ machinery exerts a nonnegligible function in shaping diverse TME and regulating cancer immunity in HCC.

## 2. Materials and Methods

*2.1. Acquisition of HCC Cohorts and Preprocessing.* Transcriptome profiling and clinicopathological annotation of HCC were gathered from The Cancer Genome Atlas (TCGA) together with Gene-Expression Omnibus (GEO) repositories. Specimens with incomplete follow-up data were removed. For TCGA data set, RNA-seq profiling (FPKM value) of 373 HCC samples and 50 normal samples was gained from the Genomic Data Commons (GDC, https://portal.gdc.cancer.gov/) with TCGAbiolinks package [22]. Afterwards, FPKM was converted to TPM form. Somatic mutations and copy number variations (CNVs) were also retrieved from TCGA. For microarray data from the GSE14520 data set, the raw "CEL" file of 242 HCC samples was retrieved, which was corrected by background and normalized by quantile with robust multiarray averaging algorithm [23]. Through Rcircos package [24], the locations of 23 $m^6A$ modifiers in chromosome were drawn. The clinical information of TCGA and GSE14520 data sets was shown in Supplementary Tables 1 and 2.

*2.2. Unsupervised Clustering for 23 $m^6A$ Regulators.* Totally, this study extracted 23 $m^6A$ modifiers comprising 8 writers (CBLL1, KIAA1429, METTL3/14, RBM15/15B, WTAP, and ZC3H13), 2 erasers (ALKBH5 and FTO), and 13 readers (ELAVL1, FMR1, HNRNPA2B1, HNRNPC, IGF2BP/2/3, LRPPRC, YTHDC1/2, and YTHDF1/2/3) from the obtained data sets. Distinct $m^6A$ regulator-mediated modification patterns were classified for HCC through unsupervised clustering analyses in the light of the level of aforementioned modifiers. By employing consensus clustering approach, the number and consistency of clustering were determined via ConsensusClusterPlus package with 1000 times repetitions [25].

*2.3. Clinical Specimens.* Three fresh HCC and matched adjacent normal liver tissues were harvested in the Affiliated Traditional Chinese Medicine Hospital of Southwest Medical University from February 2021 to May 2021. No patients experienced preoperative chemo- or radiotherapy before operation. Each subject provided written informed consent following the guideline of the Declaration of Helsinki. The study was approved by the Ethics Committee of The Affiliated Traditional Chinese Medicine Hospital of Southwest Medical University (approval id: 2021017).

*2.4. Western Blot.* Tissue and cell specimens were lysed by RIPA lysis reagent (Beyotime, China) on the ice for half hour and sonicated in an ice bath for 3 min. Following centrifugation at 4°C at 12,000 r/min for 10 min, the supernatant was harvested. The protein concentrations were calculated with BCA kit (Beyotime, China). Then, lysate was boiled with 5 × SDS loading buffer at 100°C for 5 min. Then, protein was separated by SDS-PAGE electrophoresis as well as transferred onto PVDF membranes (Millipore, Germany). Following being washed with TBST, the membranes were blocked by 5% milk/TBST lasting 1 h. Then, the membranes were probed with primary antibodies against METTL3 (1 : 1000; 15073-1-AP; Proteintech, Wuhan, China), ZC3H13 (1 : 1000; DF4623; AFFINITY, USA), YTHDF2 (1 : 1000; 24744-1-AP; Proteintech, Wuhan, China), or GAPDH (1 : 5000; ATA00013Rb; AtaGenix, Wuhan, China) at 4°C overnight. The membranes were washed by PBST for three

times. Afterwards, the membranes were exposed to HRP-labeled goat antirabbit secondary antibody (SA00001-2; Proteintech, Wuhan, China) at room temperature for 1 h. The membranes were developed with luminescent buffer and investigated by ChemiDoc™ XRS + gel imaging system (Bio-Rad, Shanghai, China).

*2.5. Immunofluorescence.* Immunofluorescence was performed for detecting METTL3, ZC3H13, and YTHDF2 expression in HCC and normal tissues. In brief, paraformaldehyde-fixed and paraffin-embedded tissue slices were cut into 5 $\mu$m thickness. The slices were incubated by anti-METTL3 (1 : 50; 15073-1-AP; Proteintech, Wuhan, China), anti-ZC3H13 (1 : 50; DF4623; AFFINITY, USA), and anti-YTHDF2 (1 : 50; 24744-1-AP; Proteintech, Wuhan, China) antibodies lasting 2 h. Following being washed, the slices were probed with ALexa Fluor 488-conjugated Affinipure goat antirabbit IgG ($H + L$) (SA00006-2; Proteintech, Wuhan, China) and DAPI (D9542; Sigma, USA). Then, the sections were mounted with glycerol and investigated under a fluorescence microscope.

*2.6. Functional Annotation Analyses.* The activity of biological processes and pathways between different clusters was compared through gene set variation analysis (GSVA) [26] that represents a nonparametric and unsupervised gene set enrichment method. The Hallmark gene set was retrieved from the Molecular Signatures Database as a reference. Functional annotation analyses of $m^6A$ regulators or $m^6A$-related genes were carried out through clusterProfiler package [27].

*2.7. Cell Culture and Transfection.* HCC cell lines (Hep3B, HUH-7; Chinese Academy of Sciences; Shanghai, China) were grown in DMEM (Thermo Fisher Scientific, USA) plus 10% FBS (Thermo Fisher Scientific, USA), 100 units/mL ampicillin together with 100 $\mu$g/mL streptomycin at 37°C with 5% $CO_2$. ZC3H13 plasmid was gained from GenePharma company (USA), which was transfected into Hep3B and HUH-7 cells via Lipofectamine™ 2000 transfection reagent (Thermo Fisher Scientific, USA).

*2.8. 5-Ethynyl-2′-deoxyuridine (EdU) Assay.* Cellular proliferation was determined utilizing BeyoClick™ EdU-594 cell proliferation detection kit (C0078S; Beyotime, Shanghai, China). The transfected cells were inoculated onto 24-well plates ($8 \times 10^5$ cells/well). All operations were carried out following the instructions. Under a fluorescence microscopy, the images were acquired and analyzed.

*2.9. Transwell Assay.* Migration and invasion were examined utilizing Transwell chambers (Corning, Shanghai, China). For invasion test, the chambers were coated by Matrigel (BD, USA), without Matrigel for migration test. HCC cells were inoculated onto the upper chambers plus serum-free media ($3 \times 10^4$ cells/well). DMEM media with 10% FBS were added to the lower chambers. Following 24 h, migrated or invasive cells were fixed by 4% PFA (Beyotime, China), and dyed utilizing crystal violet. The number of migrated or invaded cells was counted at ×100 magnification utilizing an inverted light microscope.

*2.10. Assessment of the TME.* Single-sample gene set enrichment analyses (ssGSEA) were adopted to infer the relative infiltrations of 28 immune cell populations in the TME. The enrichment scores ranging from 0 to 1 were used to denote the relative infiltrations of each immune population based on the markers of each immune population [28, 29].

*2.11. Quantifying Immune Response Predictive Factors.* The Estimation of Stromal and Immune Cells in Malignant Tumors using Expression Data (ESTIMATE) method was adopted for computing immune/stromal score that could be predictive immune/stromal cell abundance [30]. The Tumor Immune Dysfunction and Exclusion (TIDE) score may infer cancer immunotherapy response [31]. This score was based on two major tumor immune evasion mechanisms: dysfunctional tumor-infiltrating cytotoxic T lymphocytes (CTLs) as well as CTL exclusion via immunosuppressors. Immunophenoscore (IPS) that was developed by four types of immune-related genes: MHC, checkpoint or immunomodulator, effector/suppressor cell population was used to estimate anti-CTLA-4/PD-1 therapeutic response [29].

*2.12. Dimension Reduction and Ferroptosis Score.* Differentially expressed genes (DEGs) with adjusted $p < 0.05$ and $|\text{fold-change}| > 1.5$ were screened between $m^6A$ methylation patterns with limma package, called as $m^6A$ [32]. The overlapped DEGs between distinct $m^6A$ methylation patterns were chosen, called as $m^6A$ phenotype-associated DEGs. Univariate Cox regression analysis was conducted for screening prognostic $m^6A$ phenotype-associated DEGs with $p < 0.05$. According to the expression profiling of prognostic $m^6A$ phenotype-relevant DEGs, HCC subjects were clustered into distinct $m^6A$ genomic phenotypes. The expression profiles of the prognostic $m^6A$ phenotype-associated DEGs were utilized for performing PCA, followed by extraction of principal components 1 and 2 as $m^6A$ score. The approach mostly depended upon the scores on the gene sets with the most favorable association (or inverse association) genes blocks, and down-weighted contribution of genes that cannot be tracked with other set members. The formula [33, 34] was adopted for defining the $m^6A$ score: $m^6A\ \text{score} = \sum \text{PC1i} + \sum \text{PC1i}$, in which $i$ denoted $m^6A$ phenotype-associated DEG level.

*2.13. Prediction of Chemotherapy and Immunotherapy Response.* The response to two commonly chemotherapeutic agents (gemcitabine and cisplatin) was inferred through the Genomics of Drug Sensitivity in Cancer (GDSC; https://www.cancerrxgene.org/) [35]. The half maximal inhibitory concentration (IC50) values were determined through pRRophetic package [36]. The available data of

immunotherapy were obtained from IMvigor210 data set in the light of Creative Commons 3.0 License [37]. The immunotherapeutic efficacy was estimated based on subclass mapping (SubMap) analyses [38].

*2.14. Statistical Analysis.* All the computational and statistical analyses were carried out with R programming and GraphPad Prism. Pearson test was applied for assessing the interactions between variables. T-distributed stochastic neighbor embedding (t-SNE) was implemented for the differences between HCC and controls in the light of the mRNA levels of $m^6A$ modifiers. Kaplan–Meier curves of overall survival (OS), disease-free survival (DFS), disease-specific survival (DSS) together with progression-free interval (PFI) were constructed and the survival differences were computed utilizing log-rank tests. Univariate and multivariate analyses were utilized for assessing the independency of variables in predicting prognosis. Comparisons between two groups were presented with student's $t$ or Wilcoxon tests, with one-way analysis of variance or Kruskal–Wallis test among more than two groups. $p$ values $< 0.05$ were regarded as statistical significance.

## 3. Results

*3.1. Gene Mutations and Expression of $m^6A$ Regulators in HCC.* Here, the present research observed the roles of 23 $m^6A$ modifiers across HCC comprising 8 writers, 2 erasers, and 13 readers (Figure 1(a)). The prevalence of somatic variations of above regulators in HCC was summarized in Figure 1(b). Totally, 39 of the 364 (10.71%) HCC specimens displayed somatic variations of $m^6A$ regulators, mainly containing missense mutation, nonsense mutation, splice site, in frame deletion, and frame shift deletion. HNRNPC, LRPPRC, ZC3H13, ELAVL1, FMR1, RNPA2B1, YTHDC1, YTHDC2, WTAP, and IGF2BP3 occurred somatic variations in HCC. Further analysis revealed the prevalent CNVs in $m^6A$ regulators and gain was the most frequent CNV type (Figure 1(c)). We further ascertained whether the aforementioned gene mutations affected the mRNA level of $m^6A$ modifiers across HCC. Compared to control specimens, most exhibited higher expression in HCC specimens (Figure 1(d)). The $m^6A$ regulators with gain CNVs significantly had increased expression in HCC. This indicated that CNVs might prominently contribute to the perturbation on the $m^6A$ modifier level. In accordance with 23 $m^6A$ modifiers, HCC specimens were distinctly distinguished from control specimens (Figure 1(e)). Pearson correlation analyses indicated the significantly mutual regulation between regulators, as shown in Figure 1(f). Among 23 $m^6A$ regulators, we selected three regulators METTL3, ZC3H13, and YTHDF2 to verify their expression in three paired HCC and normal tissues. Our Western blot (Figures 1(g) and 1(h)) and immunofluorescence assays (Figures 1(i) and 1(j)) confirmed that METTL3 and YTHDF2 were significantly up-regulated, while ZC3H13 possessed distinct down-regulation in HCC *versus* normal specimens. These data unveiled the heterogeneity in mutations and levels of $m^6A$ modifiers between HCC and controls together with their important implications in liver tumorigenesis.

*3.2. Prognostic Implications and Biological Functions of $m^6A$ Regulators in HCC.* Functional annotation analyses confirmed the influence of 23 $m^6A$ regulators on mRNA methylation, RNA modification, RNA stability, and the like (Figure 2(a)). As depicted in Figure 2(b), there were close correlations between 23 writer, reader, and eraser modifiers. This indicated the functions of distinct $m^6A$ modifiers on HCC pathogenesis. Univariate and multivariate analyses showed the survival implication of above regulators (Figures 2(c) and 2(d)). Among them, METTL3, ZC3H13, and YTHDF2 could act as independent prognostic indicators. Previous research has reported that METTL3 and YTHDF2 may facilitate HCC progression [39]. Here, we observed the biological functions of ZC3H13 in HCC. In HepG2 and HUH-7 cells, ZC3H13 was successfully overexpressed following transfection with ZC3H13 plasmid (Figures 2(e) and 2(f)). EdU staining showed that ZC3H13 overexpression markedly weakened proliferation of HepG2 and HUH-7 cells (Figures 2(g) and 2(h)). Also, the migrated (Figures 2(i) and 2(j)) and invasive capacities (Figures 2(k) and 2(l)) were distinctly suppressed in HCC cells after overexpressing ZC3H13. Figure 2(m) showed the top-30 genes associated with ZC3H13 in HCC. Collectively, $m^6A$ regulators possessed the important prognostic implications and biological functions in HCC.

*3.3. Characterization of $m^6A$ Regulator Expression Patterns with Distinct Prognosis and TME Landscape.* Our aforementioned data indicated that the interplay between $m^6A$ regulators exerted a key function in taking shape diverse $m^6A$ modification patterns of HCC. Through ConsensusClusterPlus package, we classified HCC subjects in TCGA cohort as diverse $m^6A$ modification patterns in the light of the mRNA levels of 23 $m^6A$ regulatory genes. As a result, three distinct modification patterns were characterized with unsupervised clustering analyses (Supplementary Figures 1(a)–1(d)). Our data showed the distinct discrepancy in $m^6A$ regulator levels among three $m^6A$ modification patterns (Figure 3(a)). For most $m^6A$ regulators, cluster B displayed the highest expression, with the moderate expression in cluster A and the lowest expression in cluster C. For exploring the biological molecular alterations underlying three clusters, GSVA was carried out based on the Hallmark gene set (Figure 3(b)). We found that cluster A was markedly enriched in metabolism-related processes (fatty acid/bile acid/xenobiotic metabolisms, etc.). Meanwhile, cluster B displayed prominently enriched pathways that were in relation to carcinogenic activation, stromal and immune pathways like PI3K-Akt-mTOR, P53, Hedgehog, Notch, Wnt-$\beta$-catenin, TGF-$\beta$, complement, IL2-STAT5, and IL6-JAK-STAT3 pathways. Nevertheless, most pathways were down-regulated in cluster C. Prognosis analysis showed that cluster A presented a distinct survival superiority, while cluster B possessed the poorest outcomes in

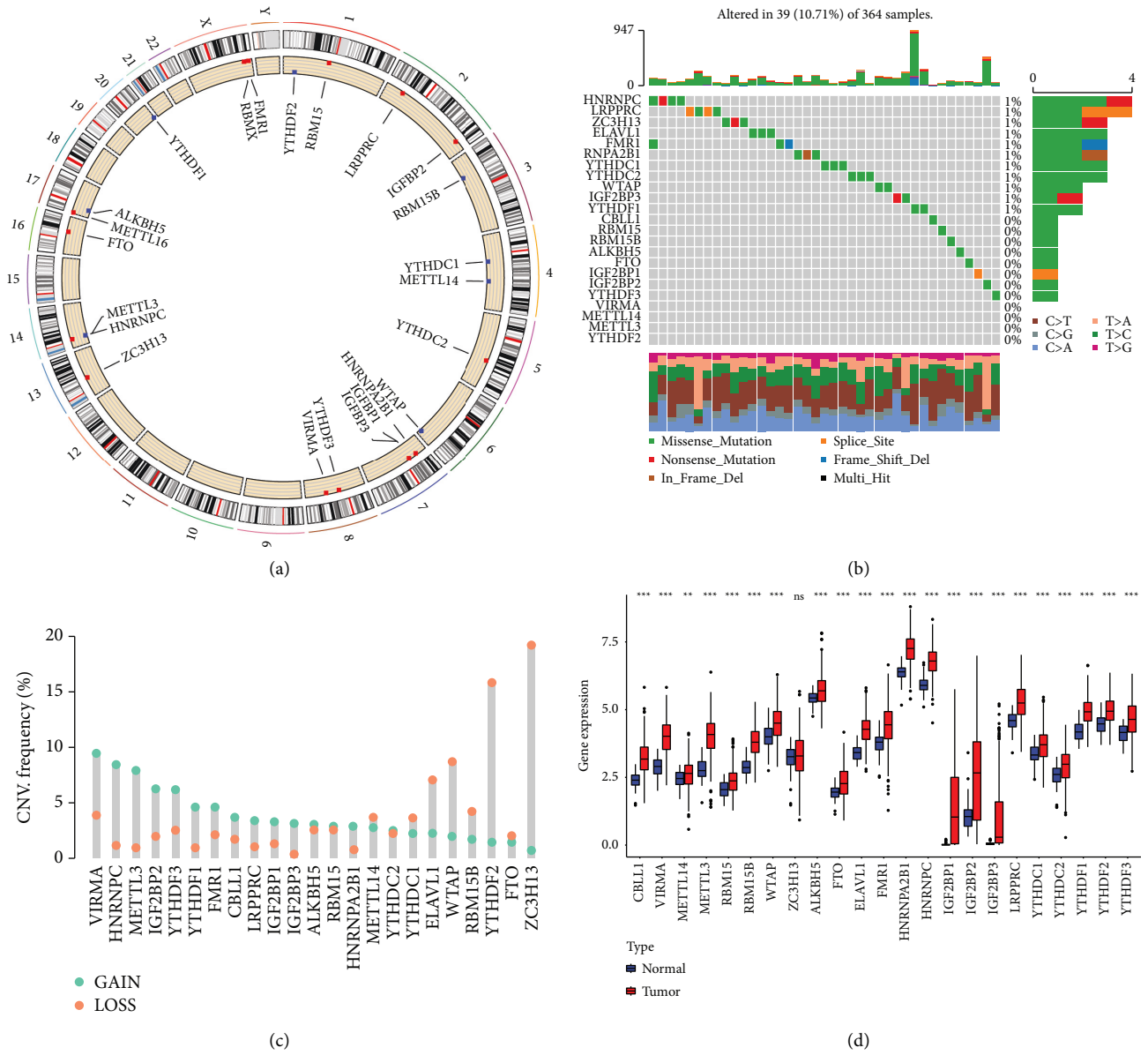
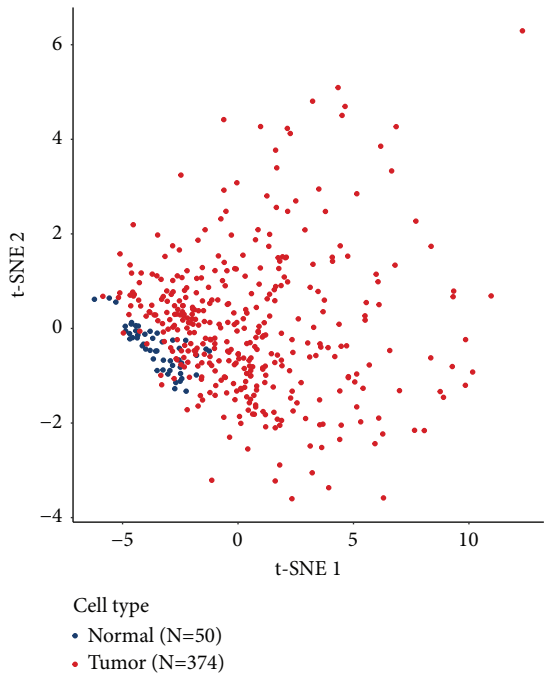
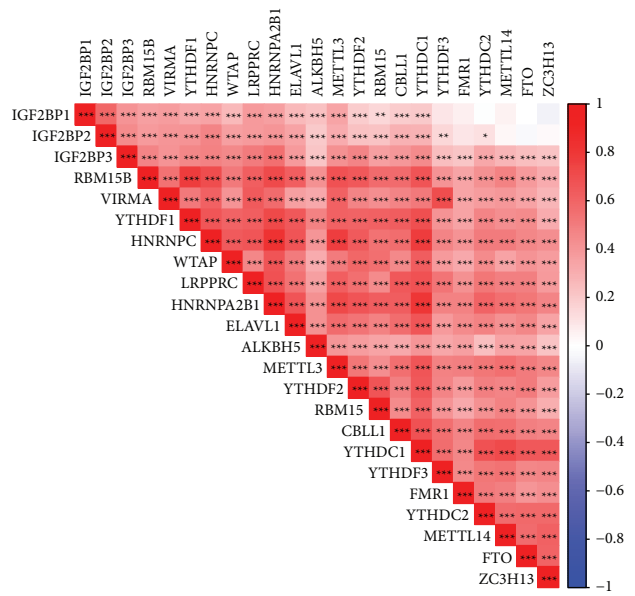


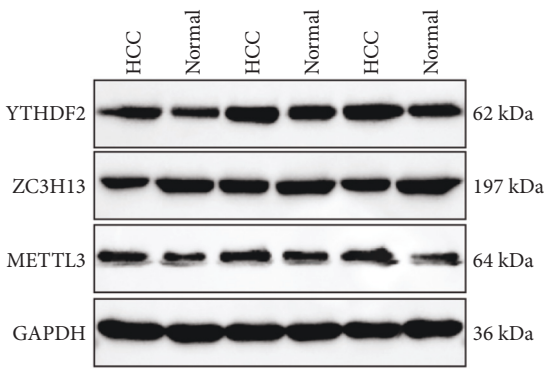

Figure 1: Continued.



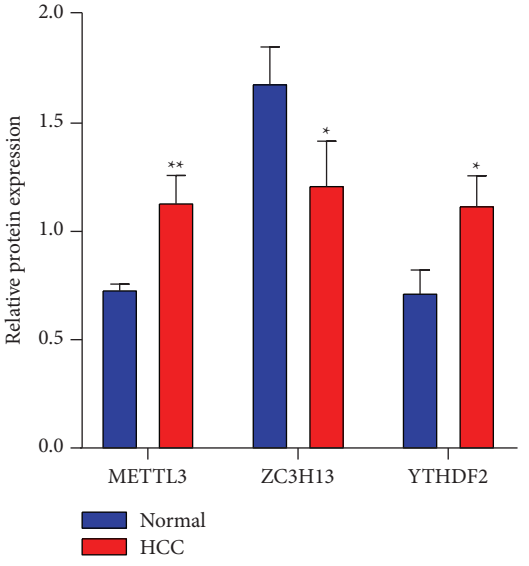

(e)




(f)

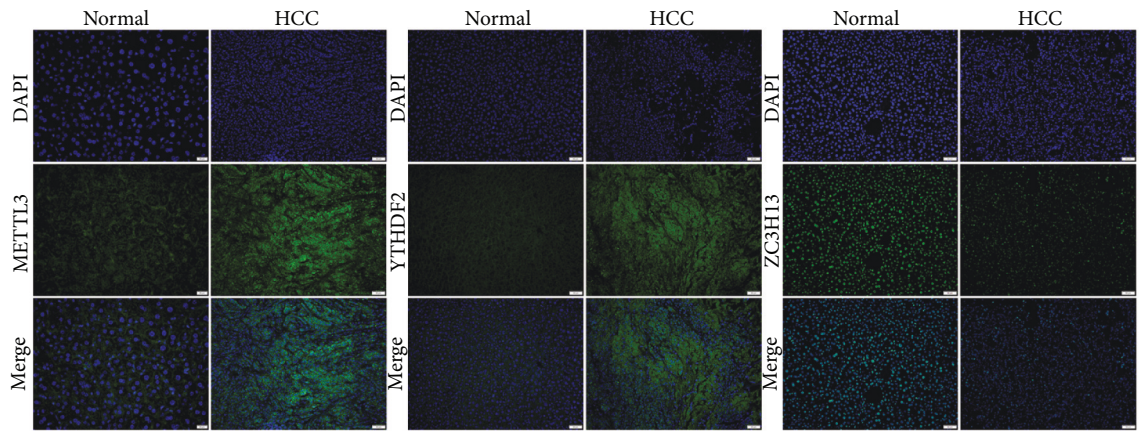



(g)

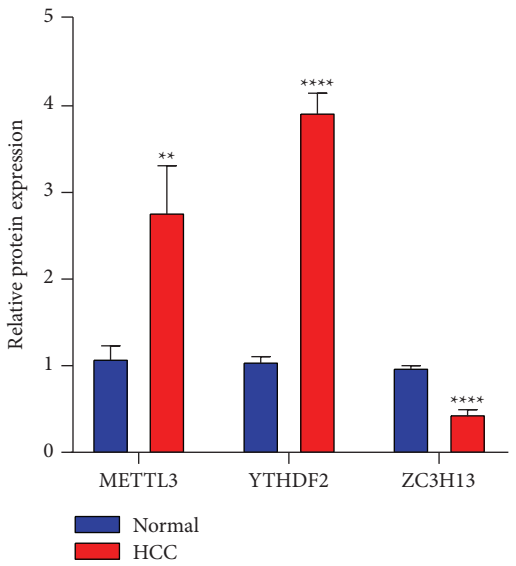



(h)

Figure 1: Continued.




(i)




(j)

Figure 1: Overall mutations and expression of $m^6A$ regulatory genes across HCC: (a) the locations of CNV alterations of 23 $m^6A$ modifiers on human chromosome. (b) The frequency of somatic mutations of 23 $m^6A$ modifiers across HCC specimens from TCGA cohort. The number on the right denoted the variation frequencies of modifies and the right bars displayed the proportions of mutation types. Each column represented individual samples. The stacked bars below indicated the fractions of transformation across specimens. (c) The CNV frequencies of $m^6A$ modifiers across HCC specimens from TCGA cohort. The column height indicated the mutation frequencies. Blue represented the frequency of gain and orange represented the frequency of loss. (d) The differences in mRNA levels of 23 $m^6A$ modifiers between HCC and normal tissues in TCGA cohort. (e) The t-SNE for distinguishing HCC (red dot) from control (blue dot) specimens based on the regulators. (f) Pearson analyses for the mutual regulation among regulators at the mRNA expression. The darker the red, the stronger the correlation. (g, h) Western blot of the expression of three regulators METTL3, ZC3H13, YTHDF2 in 3 paired HCC and normal tissues. (i, j) Immunofluorescence for the levels of above regulators. Bar = 50 $\mu$m. Ns: no significance; $^{*}p < 005$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$.

TCGA data set (Figure 3(c)). Further analyses confirmed that stromal activation was distinctly strengthened in cluster B, as displayed by EMT and WNT-target processes (Figure 3(d)). Genetic mutation pathways were also markedly activated in cluster B including DNA damage repair, DNA replication, nucleotide excision repair, homologous recombination, and mismatch repair, with cell cycle-related processes in cluster B. The ssGSEA approaches were adopted for estimating the immune population abundance across HCC. Pearson correlation analysis indicated the significantly positive or negative interactions between 23 $m^6A$ modifiers and infiltrating immune populations (Figure 3(e)). Heatmap was depicted for visualizing the differences in immune population abundance among patterns. Antitumor lymphocyte populations like effector memory CD4+ T, activated CD4+ T, and natural killer cells exhibited primary activation in cluster B (Figure 3(f)). By ESTIMATE approach, this study quantified the overall infiltrations of immune together with stromal cell populations. Cluster A was characterized by enhanced stromal score, cluster C displayed relatively high immune score and cluster B had the highest CD274 (PD-L1) expression (Figure 3(g)).

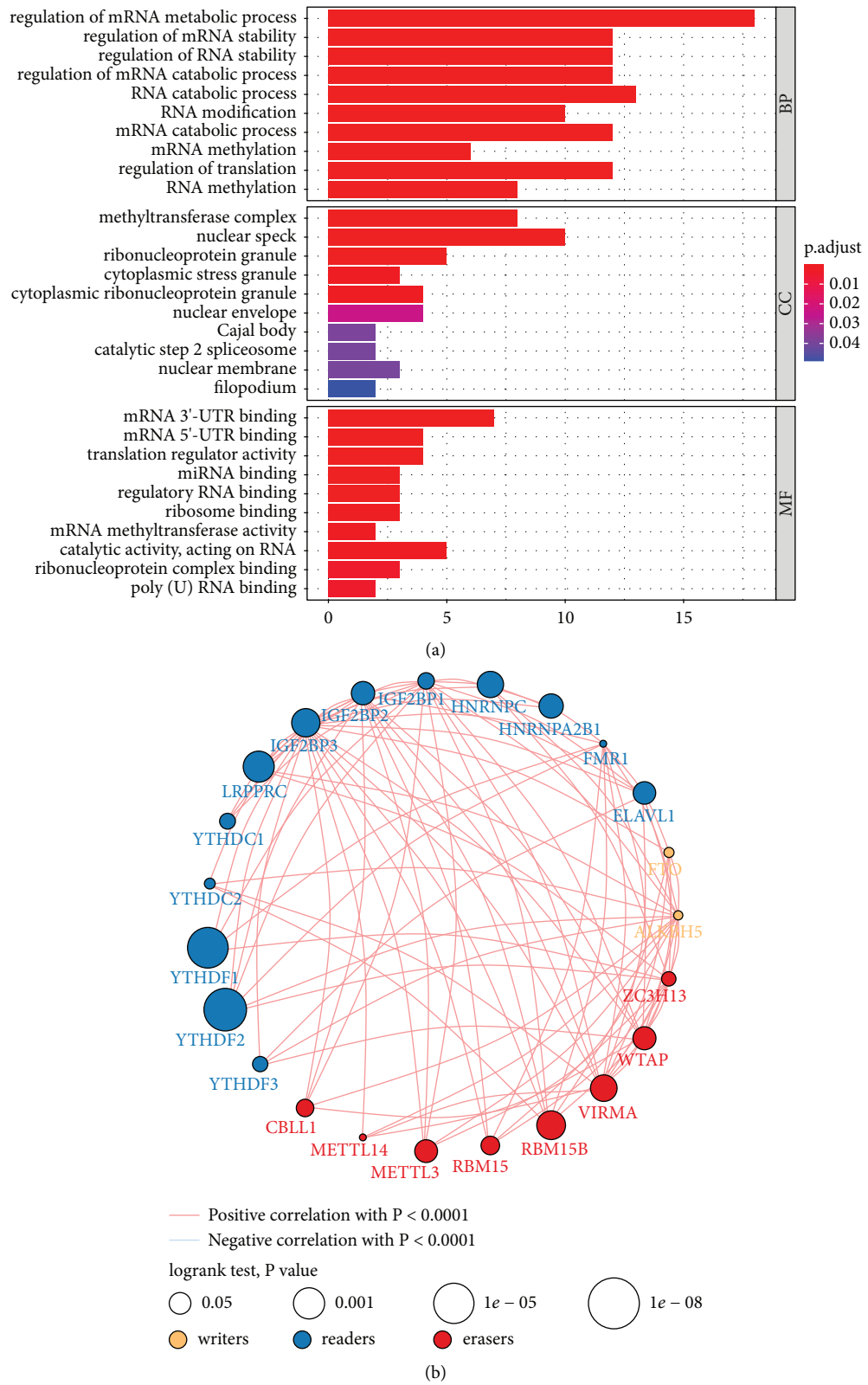
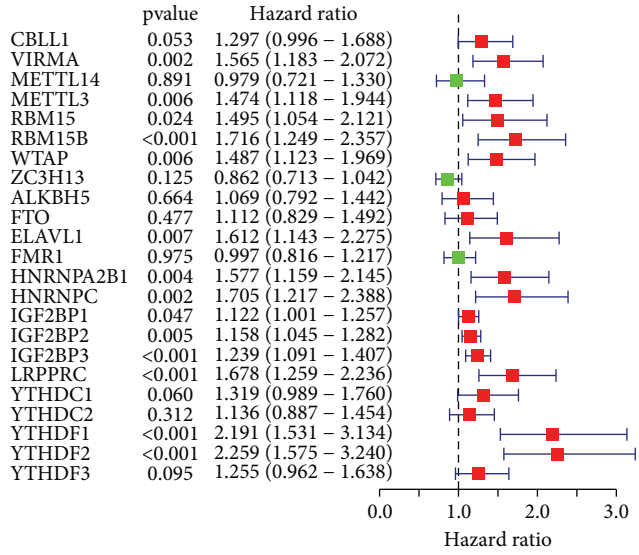
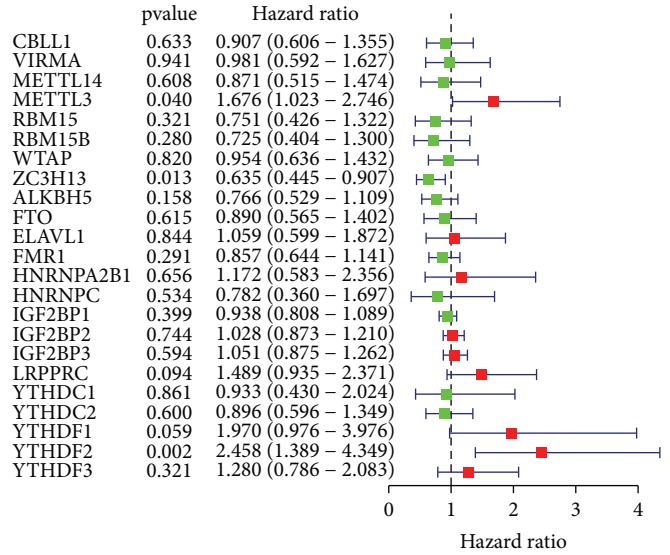


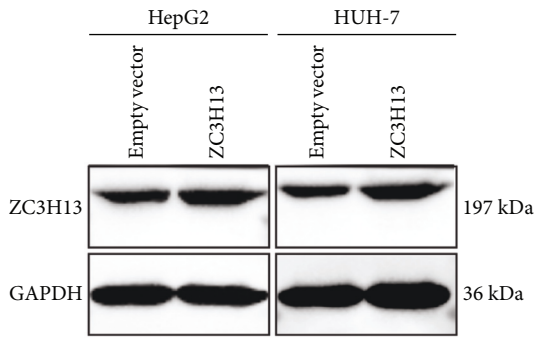

Figure 2: Continued.



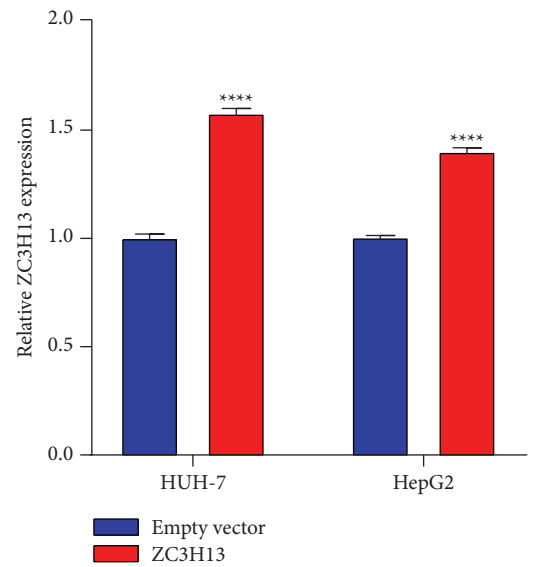

(c)



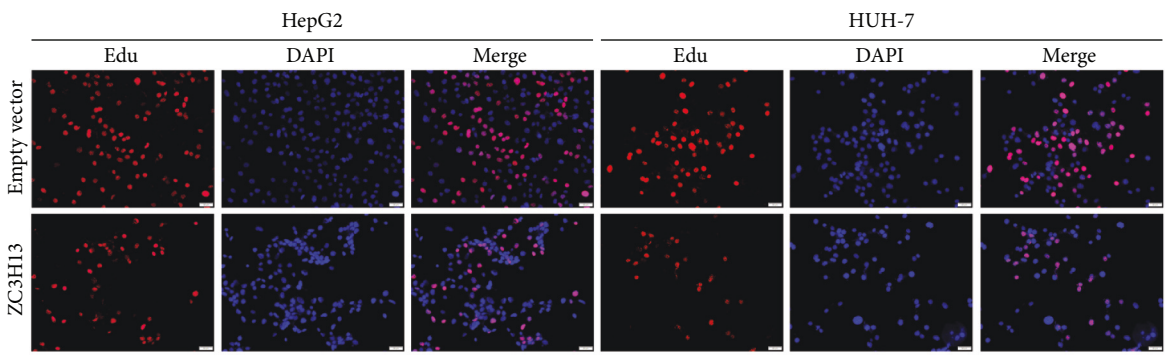

(d)




(e)

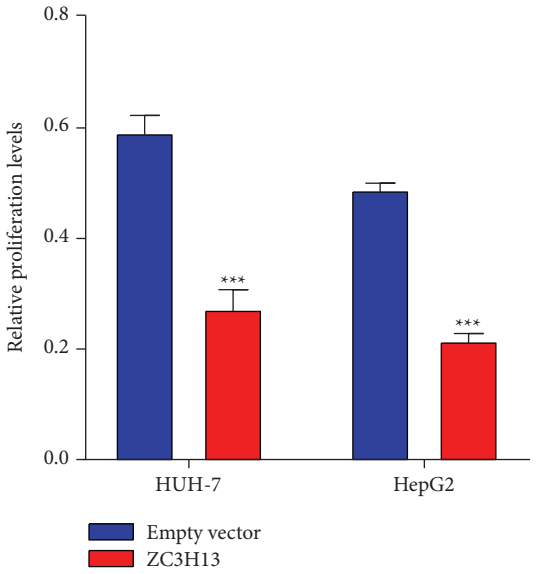



(f)

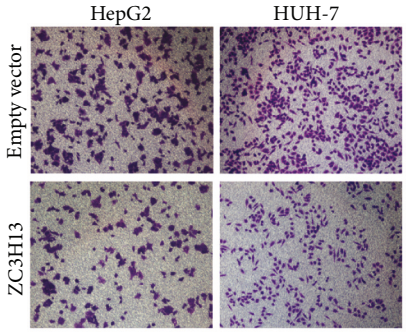



(g)

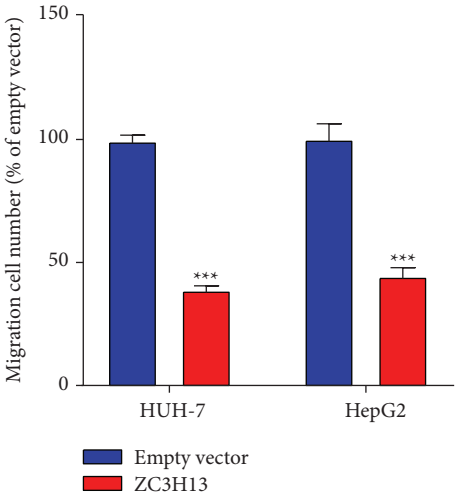
Figure 2: Continued.



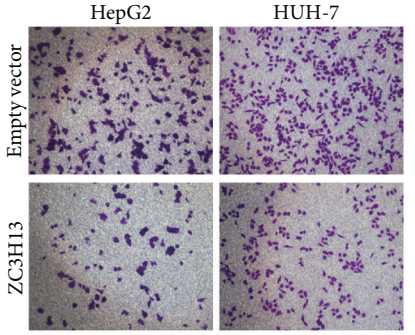

(h)




(i)




(j)




(k)

Figure 2: Continued.

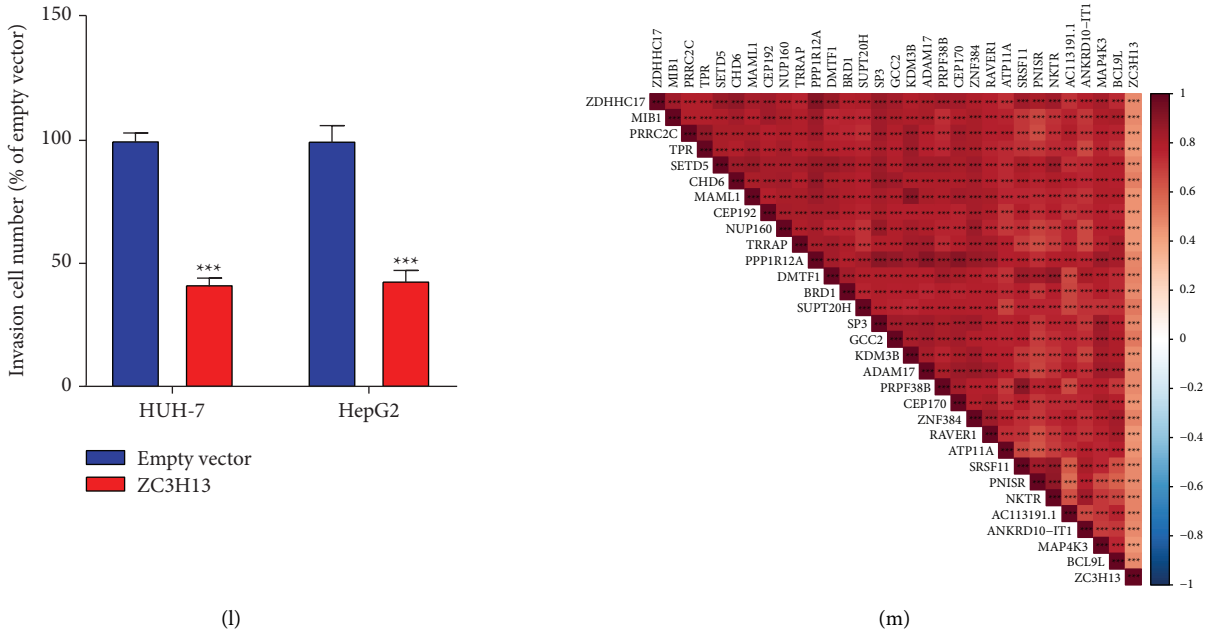



FIGURE 2: Characterization of prognostic implications and biological functions of 23 m$^6$A regulators across HCC: (a) functional annotations of the regulators with GO enrichment analyses. The bar length represented the number of enriched regulators. (b) The interaction of expression on 23 m$^6$A regulators (writers, yellow; readers, blue; erasers, red) in HCC. The lines that connected modifiers indicated their interactions, and the circle size indicated the survival value of regulators and scaled by $p$ values. (c) Uni- and (d) multivariate cox regression analyses of 23 m$^6$A regulators in HCC prognosis. (e, f) Western blot of the expression of ZC3H13 in HeG2 and HUH-7 cells transfected with empty vector and ZC3H13 up-regulation. (g, h) EdU staining for cell proliferation of HeG2 and HUH-7 cells with empty vector and ZC3H13 up-regulation. (i–l) Transwell assays for the migration and invasion of HeG2 and HUH-7 cell lines following transfection with empty vector and ZC3H13 overexpression. (m) Heatmap showing the top 30 genes associated with ZC3H13 in HCC. *** $p < 0.001$; **** $p < 0.0001$.

*3.4. Construction of m$^6$A Genomic Phenotypes in HCC.* For more deeply investigating the underlying biological behaviors in each m$^6$A cluster, 331 overlapping genes with deregulation were selected among distinct m$^6$A modification patterns with limma approach, called as m$^6$A phenotype-associated DEGs (Figure 4(a); Supplementary Table 3). Functional annotation analyses showed the distinct enrichment of the m$^6$A phenotype-associated DEGs in metabolic processes (Figure 4(b)) and pathways (Figure 4(c)), confirming the critical roles of m$^6$A methylation in HCC progression. For validating the regulation mechanisms, unsupervised clustering analysis was carried out according to above 331 m$^6$A phenotype-associated DEGs. Consistent with the m$^6$A modification patterns, patients were classified as three genomic phenotypes (Supplementary Figures 2(a)–2(d)). Prognosis analysis showed that gene cluster B displayed the worst survival outcomes among three genomic phenotypes (Figure 4(d)). Heatmap visualized the expression of the m$^6$A phenotype-associated DEGs among three gene clusters (Figure 4(e)). Consistent with m$^6$A modification patterns, gene cluster B displayed the largest stromal score, with the highest immune score in cluster C and the activated CD274 expression in cluster B (Figure 4(f)).

*3.5. Establishment of an m$^6$A Scoring System and Characterization of Its Clinical Implications.* For accurately predicting m$^6$A machinery classification across each HCC sample, this study generated the m$^6$A scoring system according to the prognostic m$^6$A phenotype-associated DEGs (Figure 5(a)). We compared m$^6$A score among distinct m$^6$A genomic phenotypes. Consequently, genomic cluster B exhibited the largest m$^6$A score, whereas cluster A possessed the lowest m$^6$A score (Figure 5(b)). Prognosis analysis showed that high m$^6$A score indicated poorer OS than low m$^6$A score (Figure 5(c)). Following univariate and multivariate analyses, m$^6$A score acted as an independent survival indicator (Figures 5(d) and 5(e)). After stratifying HCC cases into diverse subgroups in the light of clinicopathological traits (age, sex, grade, and stage), patients with high m$^6$A score also displayed unfavorable outcomes than those with low m$^6$A score (Supplementary Figure 3(a)–3(h)). No notable differences in stromal score were measured between high and low m$^6$A scores (Figure 5(f)). Nevertheless, low m$^6$A score displayed the increased immune score compared to high m$^6$A score (Figure 5(g)). CD274 expression was also compared between groups. In Figure 5(h), high m$^6$A score was characterized by elevated CD274 expression. We also found that m$^6$A score was positively associated with stromal pathways (including EMT and WNT-target) and genetic mutation-related pathways (Figure 5(i)). Furthermore, high m$^6$A score indicated the poorer DFI (Figure 5(j)), DFS (Figure 5(k)), DSS (Figure 5(l)), and PFI (Figure 5(m)) compared to low m$^6$A

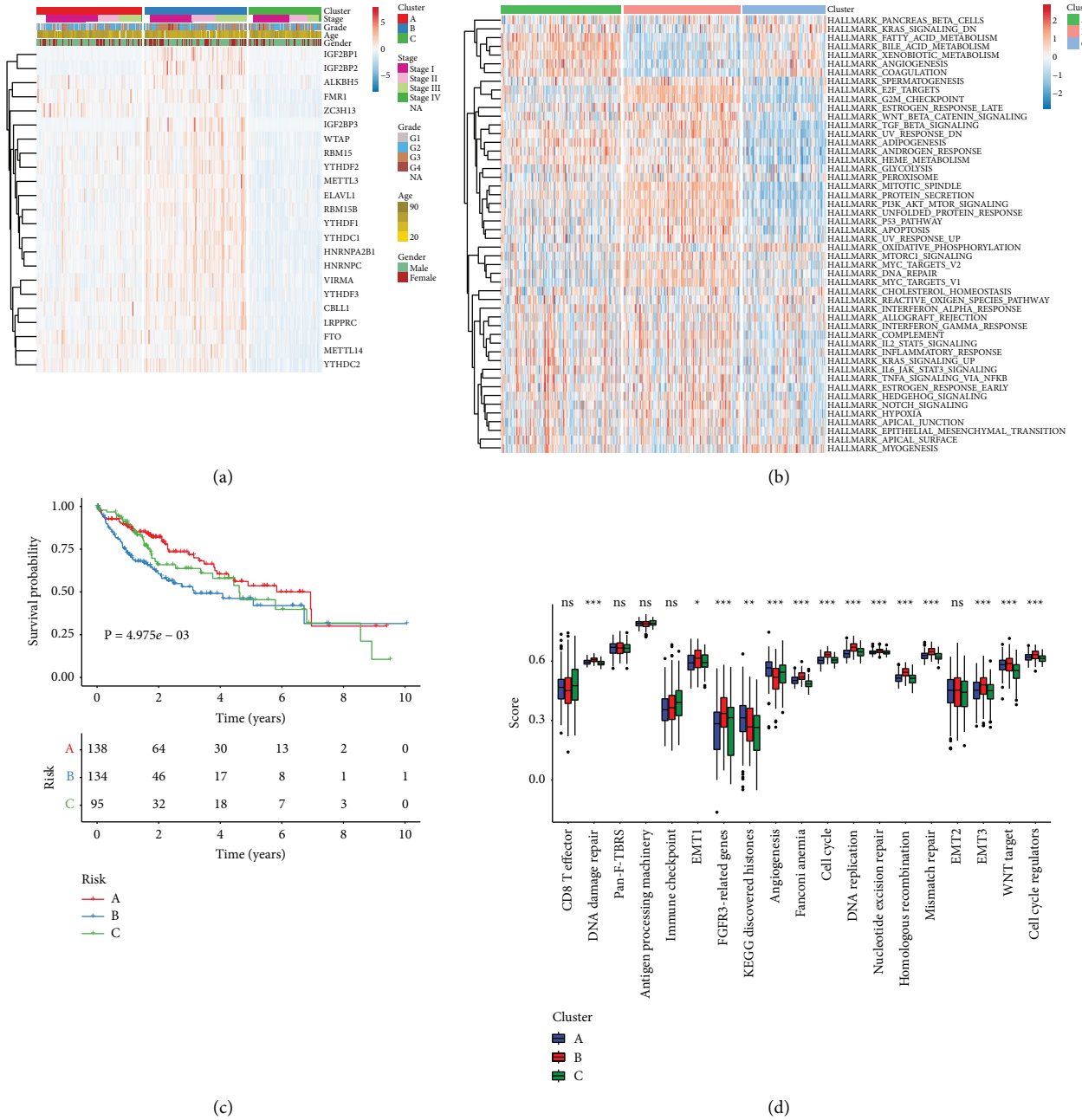
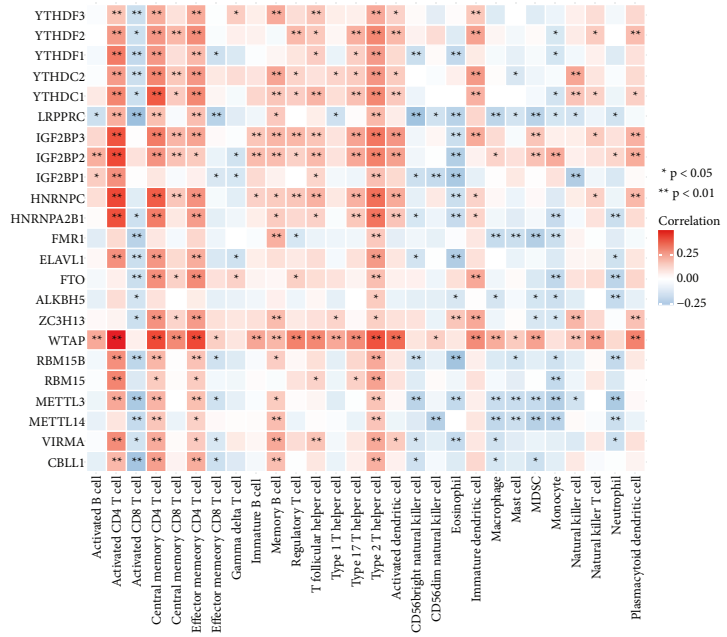
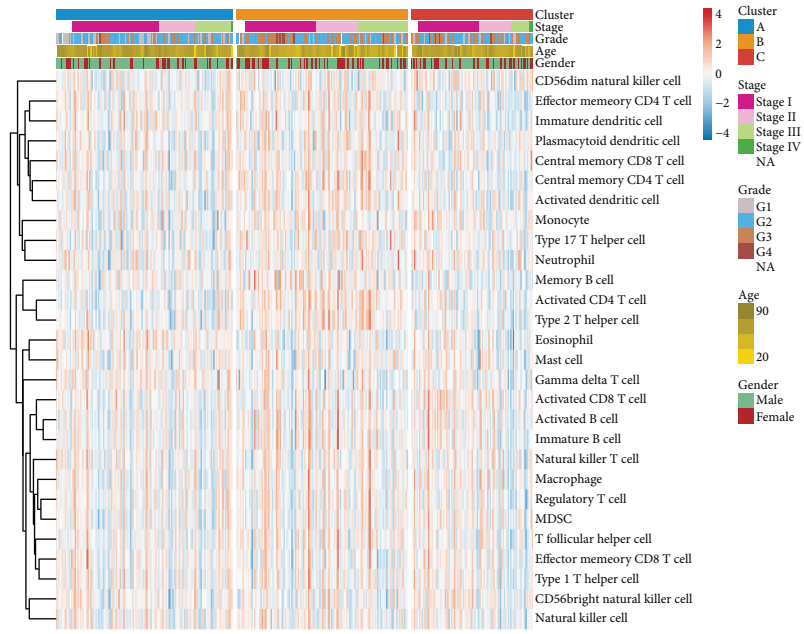



Figure 3: Continued.

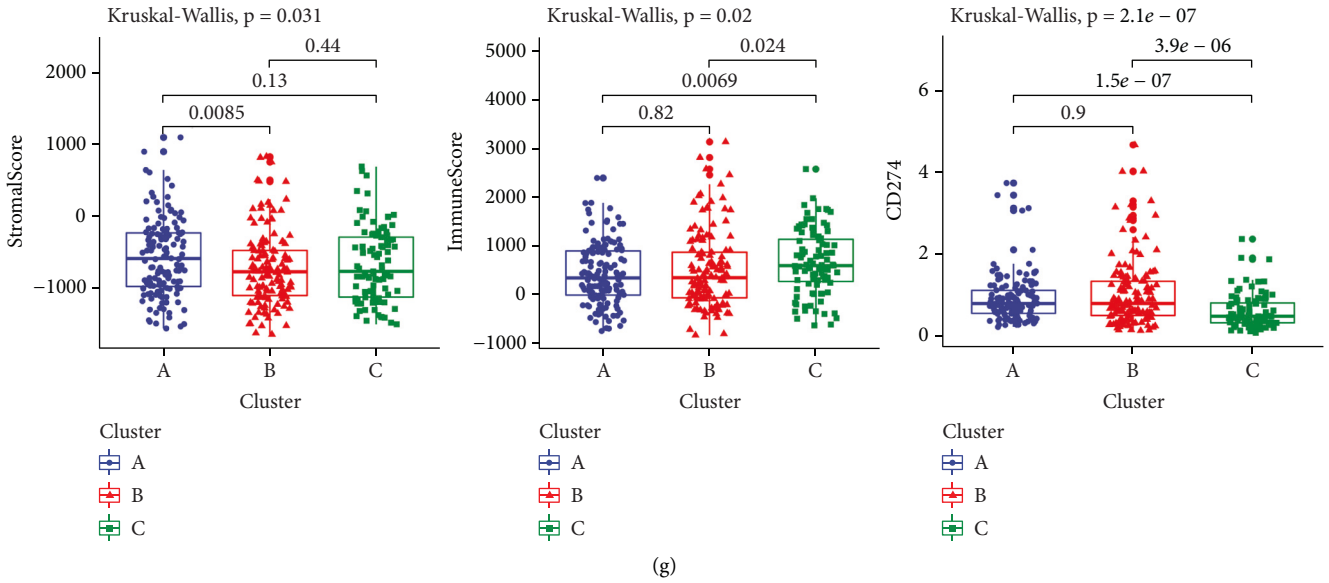



(e)




(f)

Figure 3: Continued.




(g)

Figure 3: Characterization of m$^6$A methylation patterns with diverse prognosis and TME landscape in TCGA cohort: (a) Heatmaps for the mRNA levels of 23 m$^6$A modifiers among three m$^6$A machinery phenotypes. (b) Heatmap of the GSVA scores of Hallmark pathways in distinct m$^6$A modification patterns. (c) Kaplan–Meier OS curves for HCC patients from TCGA data set with diverse patterns (log-rank tests). (d) The differences in the activation of several pathways among three patterns. Kruskal-Wallis H test was applied for comparing three clusters. Ns: no significance; $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. (e) Heatmap of the correlations of 23 m$^6$A modifiers with infiltrating immune cells based on ssGSEA method. Positive correlation, red; negative correlation, blue. (f) Heatmap for the differences in 28 immune population abundance among three patterns. (g) The discrepancy in stromal/immune score and CD274 expression among three patterns.

score, indicating that m$^6$A score could be utilized for predicting HCC recurrence and progression.

*3.6. The m$^6$A Score Can Predict Chemotherapeutic and Immunotherapeutic Benefits.* Tumor somatic mutations were compared between m$^6$A score subgroups in TCGA cohort. As depicted in Figures 6(a) and 6(b), more prevalent somatic mutations occurred in high m$^6$A group compared to low m$^6$A group. The responses to two commonly chemotherapeutic drugs were assessed by GDSC database. After comparison, the estimated IC50 value of gemcitabine was markedly lower in high m$^6$A score population (Figure 6(c)). This demonstrated that patients possessing high m$^6$A score were more likely to benefit from gemcitabine. Nevertheless, no significant difference in cisplatin was found between groups (Figure 6(d)). TIDE score, an emerging predictor of immunotherapy, was calculated in each HCC specimen. Compared to high m$^6$A score group, decreased TIDE score was found in low m$^6$A score population (Figure 6(e)), indicating the more benefits from immunotherapy. Figure 6(f) visualized the proportions of patients that responded to anti-PD-L1 immunotherapy in the IMvigor210 data set. The responser/nonresponser was 15%/85% in high m$^6$A score population and 32%/68% in another population. This confirmed that cases with low m$^6$A score exhibited the distinct therapeutic advantage in anti-PD-L1 therapy. In addition, we compared the survival differences between populations in this cohort. In Figure 6(g), cases with low m$^6$A score possessed the significant prognostic advantage. SubMap analyses also indicated that low m$^6$A score cases possessed the greater possibility of responding to anti-PD-1 therapy (Figure 6(h)). The IPS score was determined for evaluating the immune response, which was markedly increased in low m$^6$A score population (Figure 6(i)). Thus, m$^6$A modification patterns could be involved in mediating the tumor immune response.

*3.7. External Validation of m$^6$A Methylation Patterns and m$^6$A Score.* The m$^6$A regulator-based m$^6$A methylation patterns were validated among 242 HCC patients in the GSE14520 cohort. Consistent with the TCGA cohort, HCC cases were classified as three m$^6$A methylation patterns (Figure 7(a)). Heatmap visualized the notable discrepancy in 23 m$^6$A modifier levels among three m$^6$A methylation patterns (Figure 7(b)). Similarly, cluster B exhibited the poorest OS outcomes (Figure 7(c)). Based on the m$^6$A-related DEGs, we also calculated m$^6$A score of each HCC case in the GSE14520 data set. Data demonstrated that cases with low m$^6$A score exhibited the significant survival advantage (Figure 7(d)). These data confirmed the accurate classifications of HCC based on 23 m$^6$A regulators.

## 4. Discussion

Mounting evidences suggest that m$^6$A machinery is closely in relation to innate immunity, inflammatory response along with anticancer effects by interaction with different m$^6$A regulatory genes [40]. Here, we established three m$^6$A regulator-based m$^6$A modification phenotypes with diverse prognostic outcomes, biological processes along with TME traits across HCC according to the expression profiling of 8 writers, 2 erasers, and 13 readers. Based on m$^6$A-associated

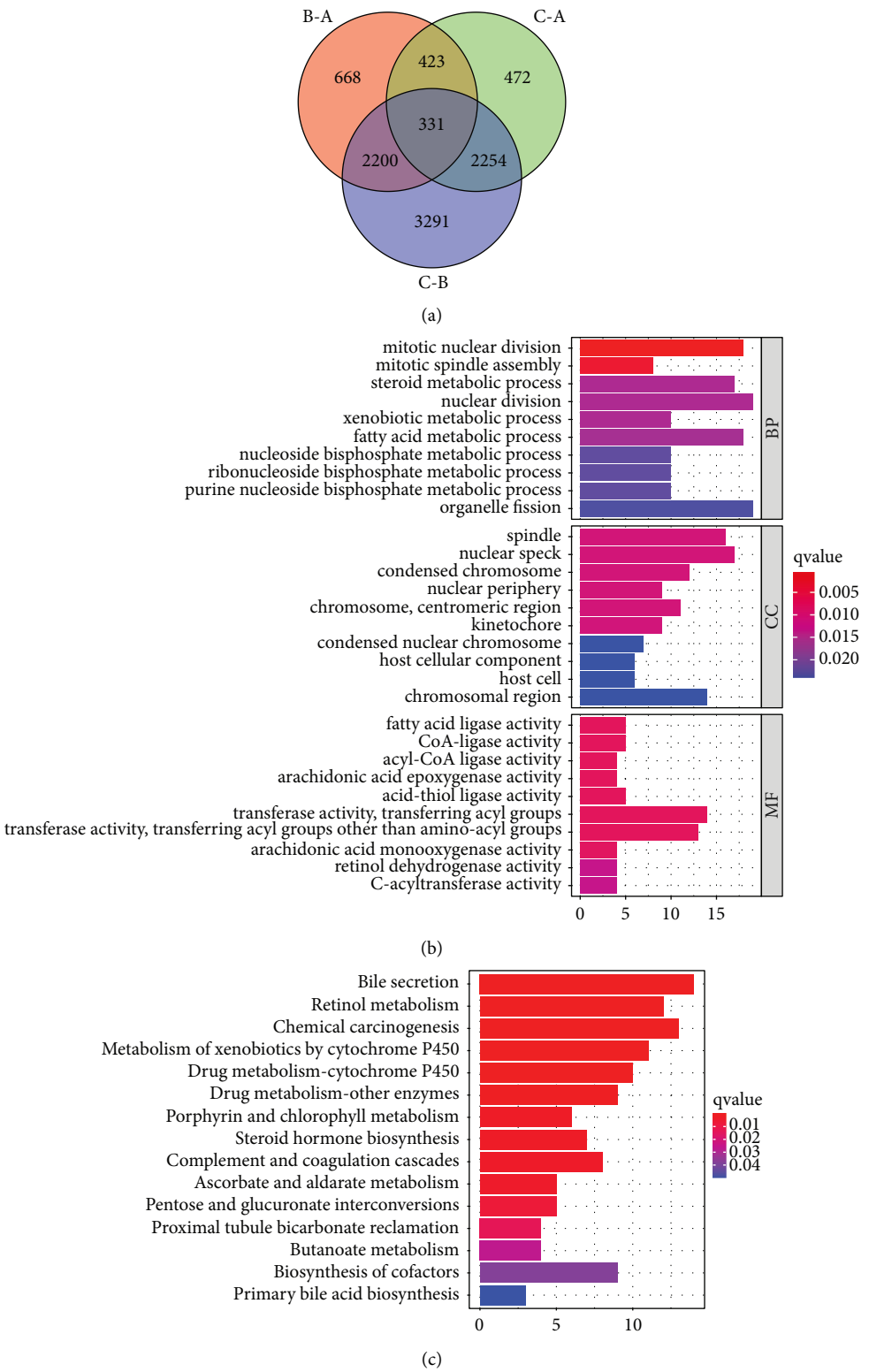



Figure 4: Continued.

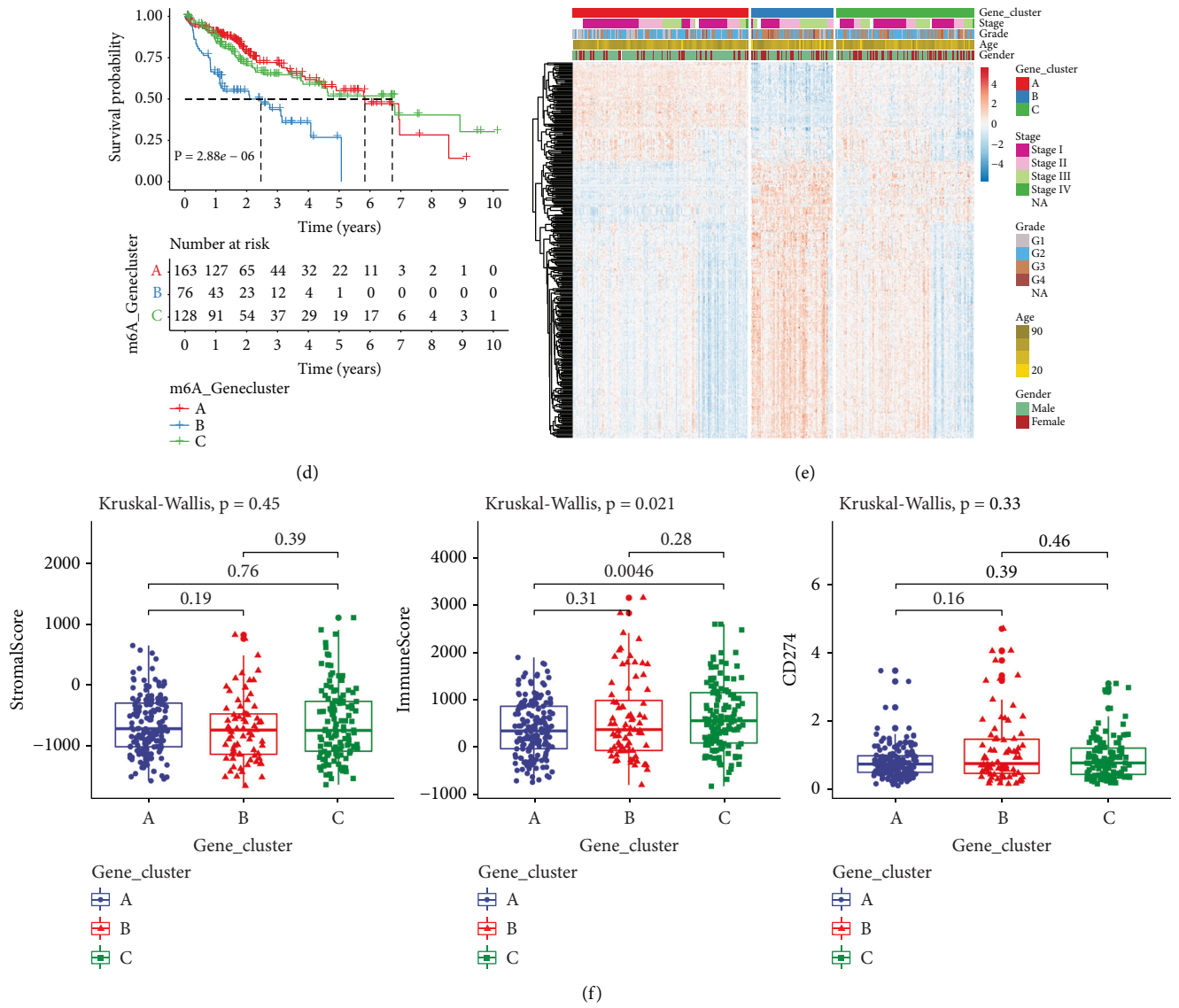
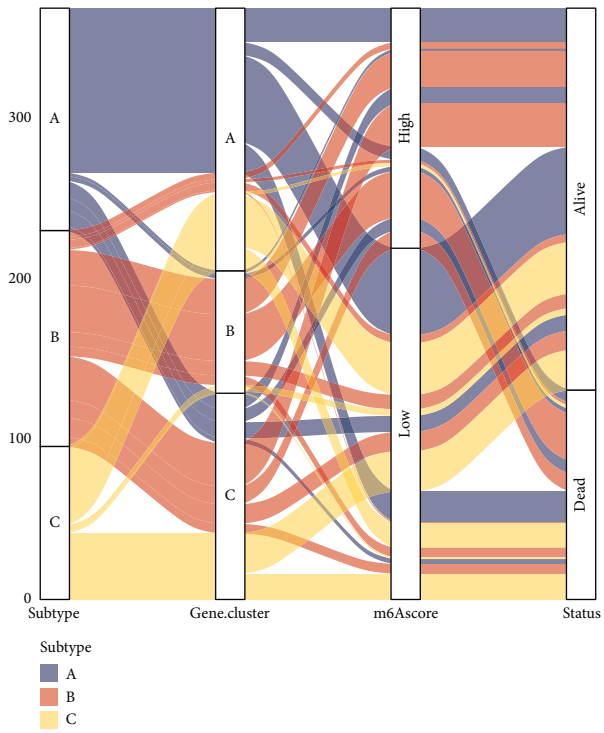



FIGURE 4: Construction of m$^6$A genomic phenotypes in HCC from TCGA cohort: (a) Venn diagram for the 331 m$^6$A phenotype-associated DEGs between three genomic phenotypes. (b) GO and (c) KEGG enrichment results of the m$^6$A phenotype-associated DEGs. (d) Kaplan–Meier curves of OS for three gene clusters (log-rank tests). (e) Heatmap for the levels of the m$^6$A phenotype-associated DEGs among clusters. (f) The discrepancy in stromal score, immune score and CD274 expression among clusters.

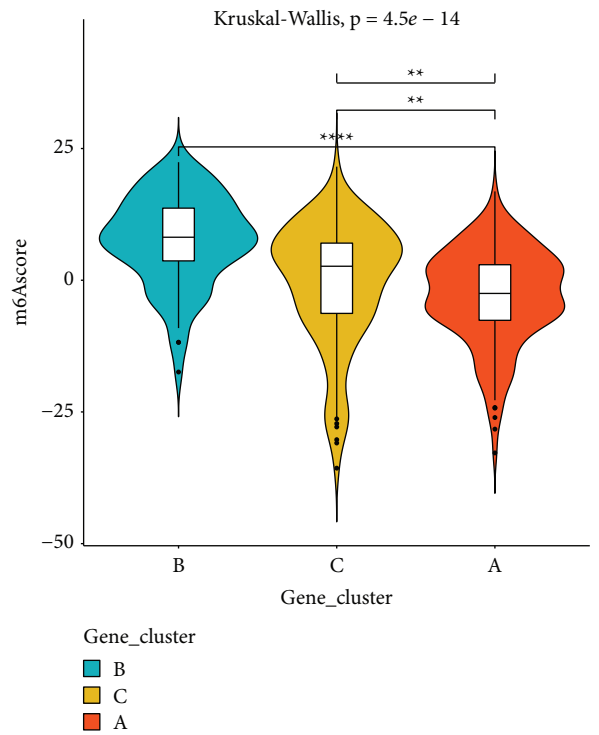
DEGs, we developed an m$^6$A scoring system for each HCC specimen. The m$^6$A score enabled to infer immunotherapeutic response. Hence, the current research highlighted the roles of m$^6$A machinery in shaping TME along with immunity modulation, which might promote precision immunotherapeutic strategies.

Herein, we comprehensively uncovered the somatic mutations, CNVs and expression patterns of 23 m$^6$A regulators. Most displayed the gain CNVs and high expression in HCC, and their expression profiling could distinguish HCC from normal liver tissues. Also, most were distinctly correlated to HCC prognosis. These data suggested that m$^6$A regulators participated in HCC initiation and progress. We verified METTL3, ZC3H13, and YTHDF2 levels across HCC and normal specimens via Western blot and immunofluorescence. Our data confirmed the up-regulation of METTL3 and YTHDF2 as well as the down-regulation of ZC3H13 in HCC. So far, there is no study about the implications of ZC3H13 across HCC. However, previous research has found that ZC3H13 acts as a tumor suppressor gene in breast carcinoma [41]. It mitigates growth along with invasive capacity of colorectal cancer via inactivation of Ras-ERK signaling [42]. Herein, our experiments showed that ZC3H13 up-regulation exerted an inhibitory effect on proliferation and invasion of HCC cells, confirming the anti-HCC effects of ZC3H13.

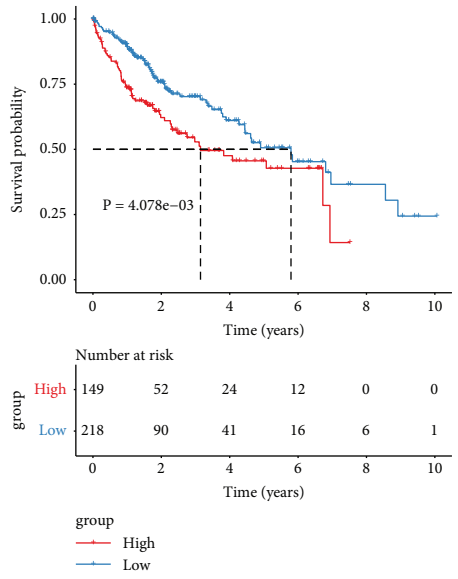
At the mRNA levels, 23 m$^6$A regulators exhibited significant synergistic effects in HCC. Hence, the current research established three m$^6$A machinery patterns in the light of their expression. In cluster A, metabolism-related processes were distinctly activated while cluster B displayed the activation of carcinogenic, stromal, and immune pathways.



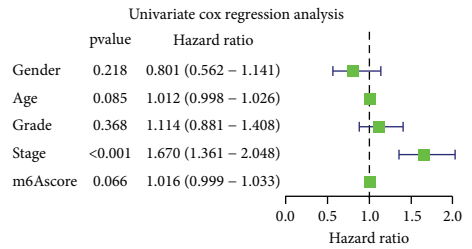

(a)



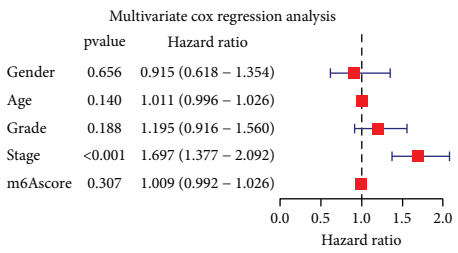

(b)



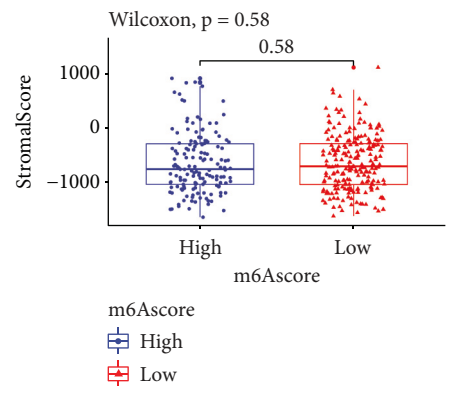

(c)




(d)

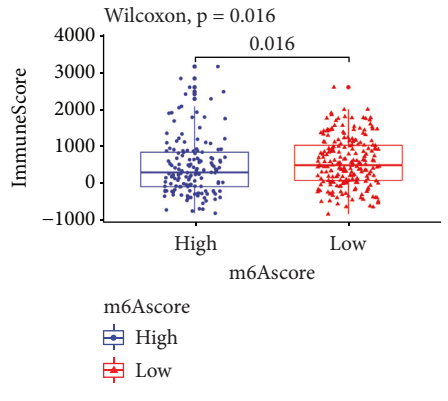



(e)

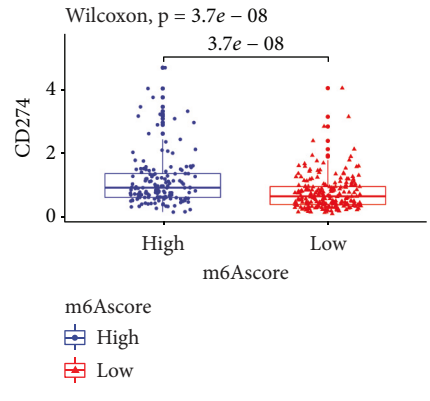



(f)

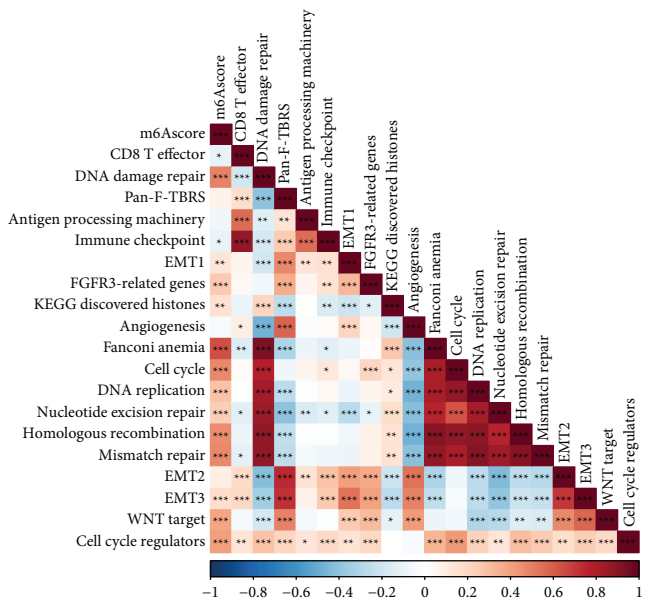
Figure 5: Continued.



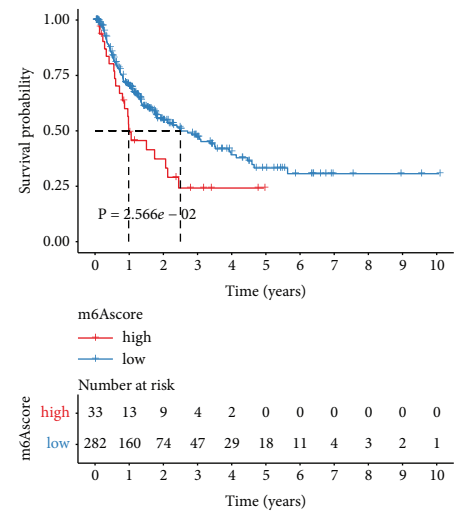

(g)



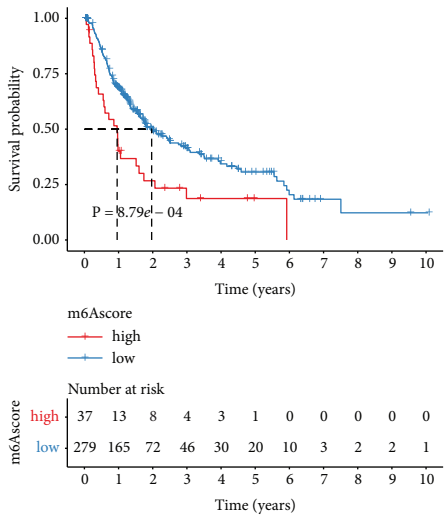

(h)



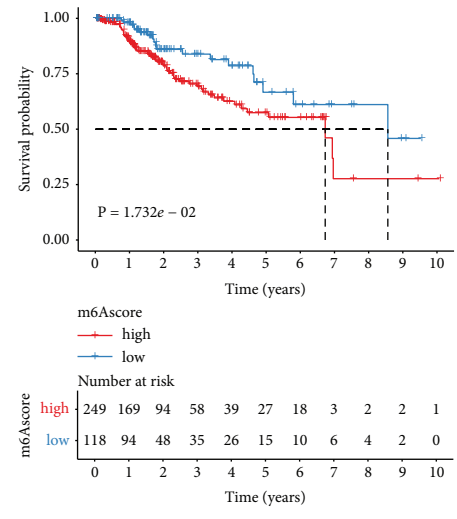

(i)




(j)

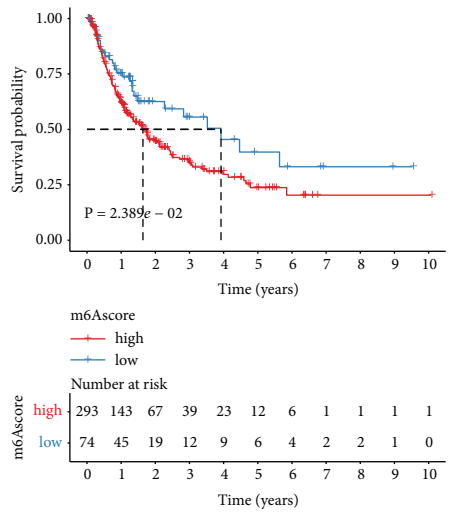



(k)



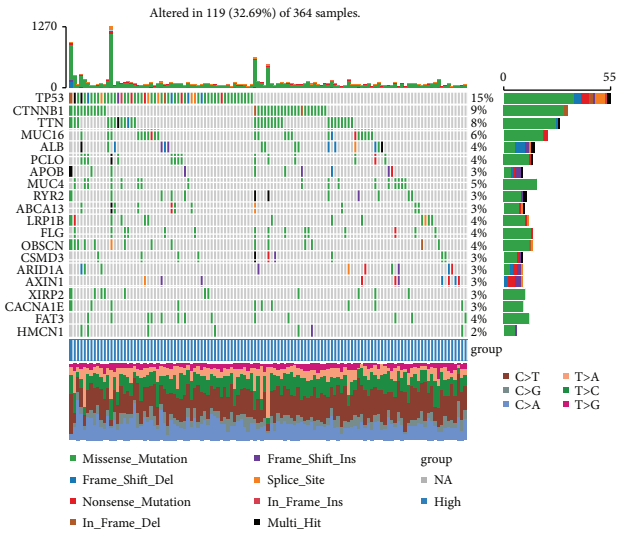

(l)

Figure 5: Continued.

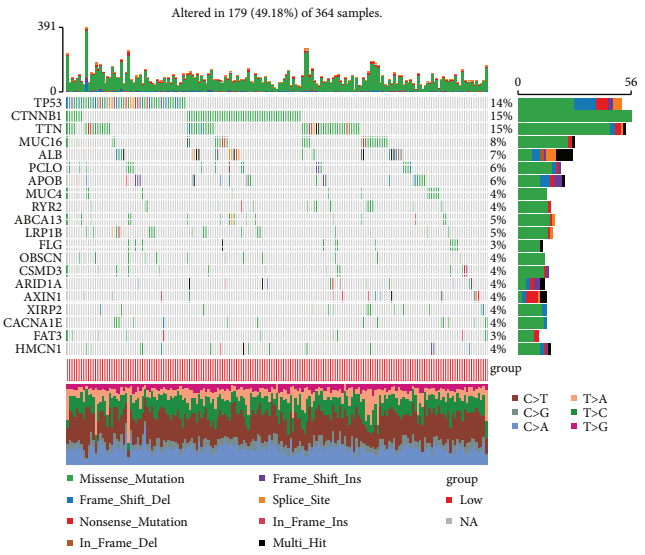



(m)

Figure 5: Establishment of an m$^6$A scoring system along with characterization of its clinical implications in TCGA cohort: (a) alluvial chart of m$^6$A modification, gene cluster, m$^6$A score along with survival status. (b) The discrepancy in m$^6$A score between three m$^6$A-modified gene clusters. *P* values were calculated with Kruskal-Wallis tests. (c) Kaplan–Meier OS curves for patients high or low m$^6$A score (log-rank test). (d) Uni- and (e) multivariate analyses of the independency of the m$^6$A score in predicting HCC outcomes. (f–h) The discrepancy in (f) stromal score, (g) immune score and (h) CD274 expression between groups. *P* values were determined using Wilcoxon tests. (i) Associations between m$^6$A score and several pathways in HCC specimens. (j–m) Kaplan–Meier curves of (j) DFI, (k) DFS, (l) DSS and (m) PFI for cases with high or low m$^6$A scores (log-rank tests). $^*p < 005$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$.




(a)




(b)

Figure 6: Continued.

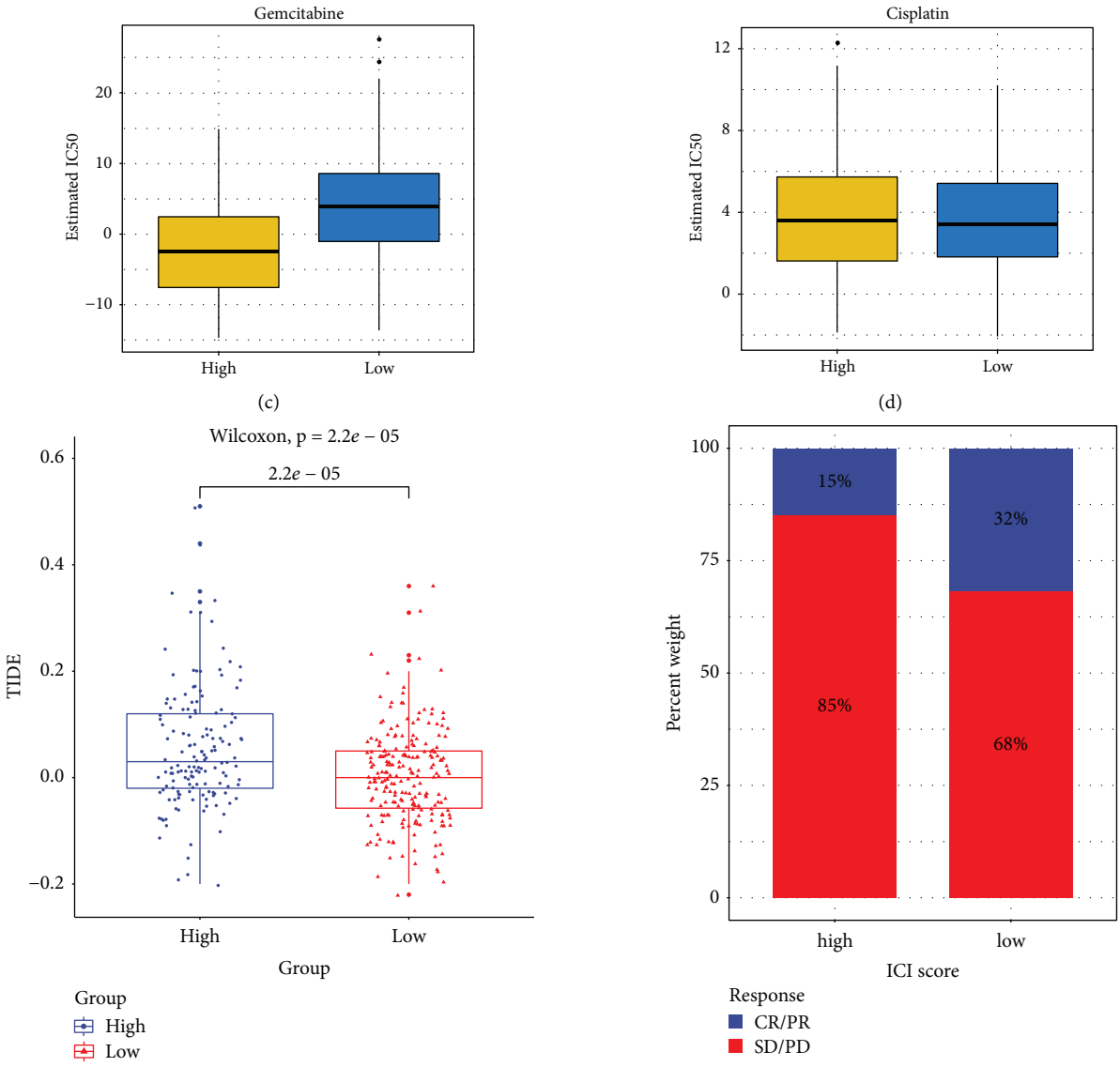



Figure 6: Continued.

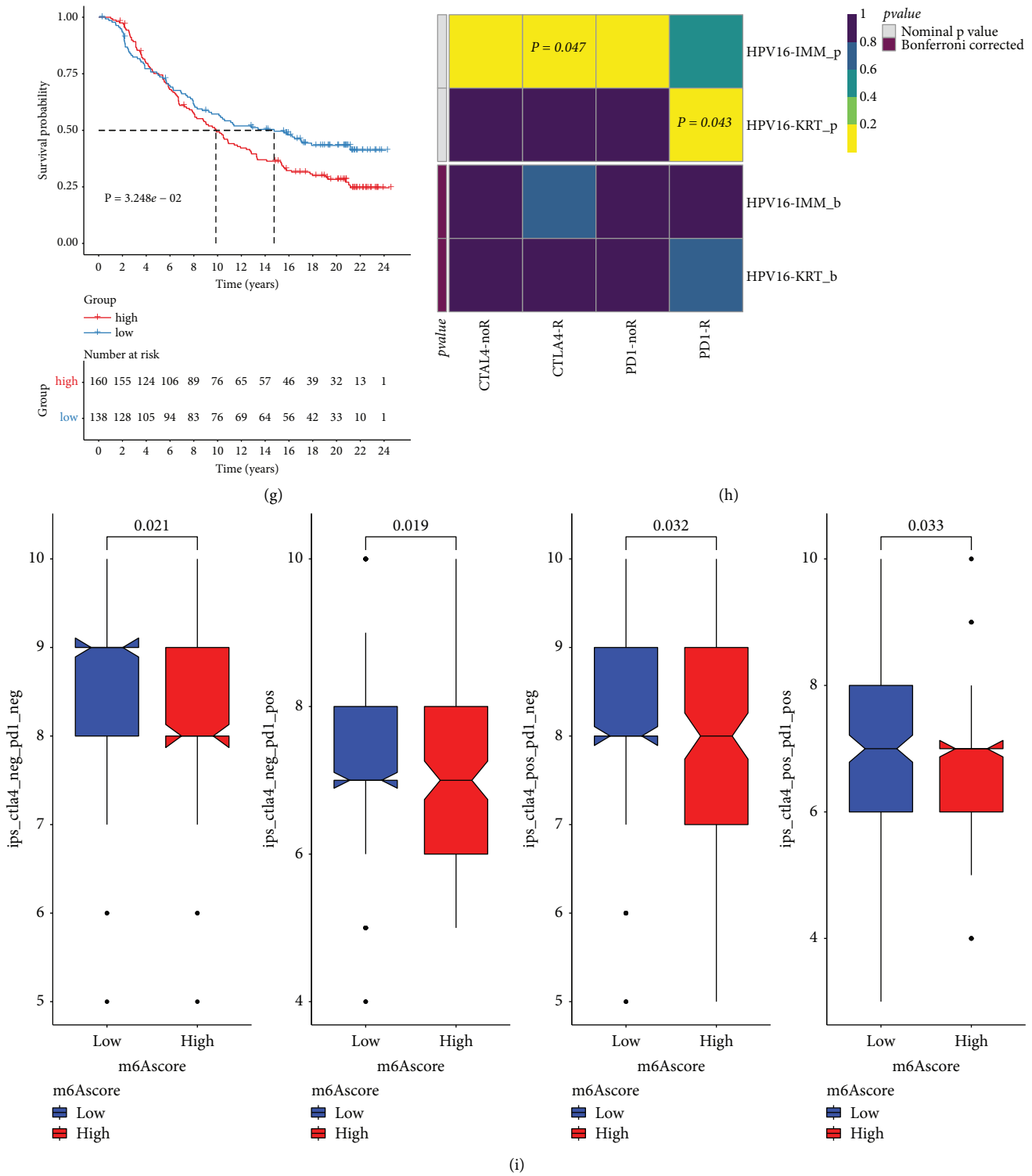



FIGURE 6: The roles of m$^6$A score in prediction of chemo- and immunotherapeutic benefits of HCC patients in TCGA cohort: (a, b) Somatic mutation landscape of significantly mutated genes across (a) high or (b) low m$^6$A score populations. (c, d) The differences in estimated IC50 values between populations. (e) The comparisons in TIDE score between populations. *P* values were determined with Wilcoxon tests. (f) The proportions of subjects responding to anti-PD-L1therapy in two populations in the IMvigor210 cohort. (g) Kaplan–Meier OS curves for cases with high and low m$^6$A scores (log-rank test). (h) The response of cases with high and low m$^6$A scores to anti-CTLA-4/PD-1 therapies through SubMap analyses. (i) The comparisons of IPS scores in two populations.

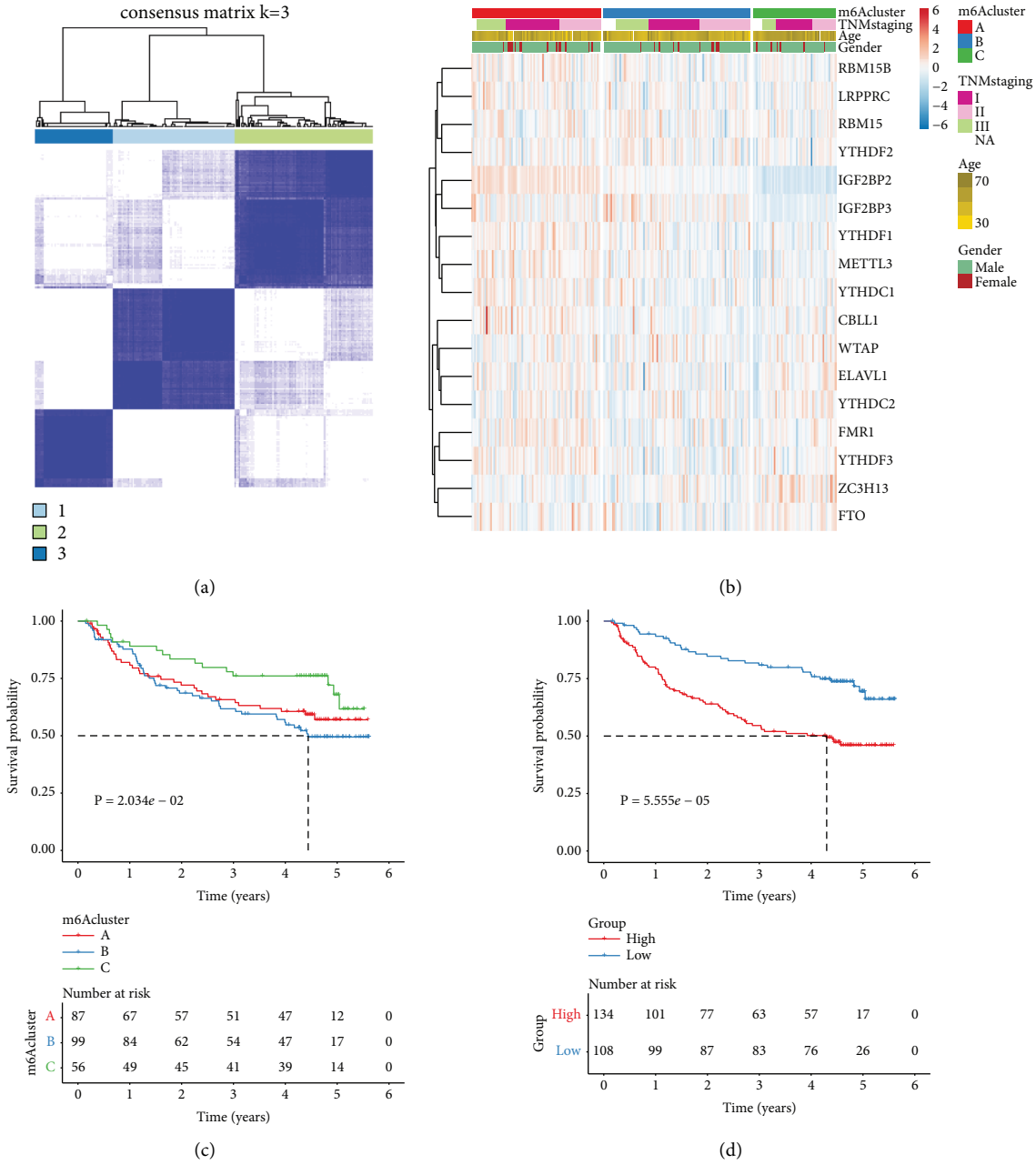



FIGURE 7: Verification of the m$^6$A methylation patterns and m$^6$A score for HCC in the GSE14520 data set: (a) Consensus clustering analysis for classifying 242 HCC patients into three m$^6$A machinery phenotypes in the light of the expression profiles of 23 m$^6$A modifiers. (b) Heatmap for the expression of 23 m$^6$A regulators among three patterns. (c) Kaplan–Meier OS curves among three phenotypes. (d) Kaplan-Meier OS curves for cases with high or low m$^6$A score. *P* values were calculated with log-rank tests.

This explained why cluster B had the worst survival outcomes. The three m$^6$A methylation patterns were characterized by distinct TME. For instance, anti-tumor lymphocyte cells were mainly activated in cluster B. Totally, 331 DEGs were identified among patterns, which were mainly in relation to metabolic processes. The evidence suggests the biological implications of metabolic process in HCC outcomes [2]. On the basis of m$^6$A-associated DEGs, three m$^6$A gene phenotypes were clustered. Consistent with the m$^6$A methylation patterns, three gene phenotypes were characterized by distinct prognosis and TME. For defining m$^6$A modification patterns, we developed an m$^6$A scoring system, which might guide treatment plans for each subject. Our prognosis analyses unveiled that the m$^6$A score acted as a credible independent survival predictor for HCC. Elevated m$^6$A score was markedly correlated to undesirable OS, recurrence, and progression of HCC patients. Moreover, m$^6$A score exhibited the strong correlations to immune predictors such as TIDE and IPS. Low m$^6$A score was more likely to be responsive to anti-PD-1/PD-L1 therapy in the IMvigor210

cohort. Our data confirmed that the m$^6$A machinery patterns can be adopted in clinical application as well as immunotherapy plans for HCC.

There are advantages in our study. First, several experimental studies have verified the roles of single m$^6$A regulators in HCC [43, 44]. For instance, RBM15 is highly expressed in HCC, and its up-regulation is indicative of undesirable survival outcomes as well as triggers HCC progression by modulating m$^6$A-modified YES1 depending upon IGF2BP1 [45]. However, the function of ZC3H13 in HCC progression remains unclear. Herein, this research for the first time confirmed the low level of ZC3H13 in HCC tissue and ZC3H13 up-regulation inhibited the proliferation, migration along with invasion of HCC cells through experiments. Second, although a recent study constructed distinct m$^6$A regulator-based m$^6$A modification patterns for HCC in TCGA cohort, this classification was not externally verified in independent cohorts. Herein, we determined three m$^6$A regulator-mediated modification patterns for HCC, and confirmed the classification accuracy in the GSE14520 cohort. Third, we developed an m$^6$A score that could predict survival outcomes and therapeutic responses for HCC patients. Despite this, several limitations should be pointed out. First, although the current research reviewed the literature and collected 23 known m$^6$A modifiers, new identified regulators should be included for optimizing the accuracy of the m$^6$A machinery phenotypes. Second, the m$^6$A machinery phenotypes and m$^6$A score were proposed based on retrospective cohorts. Therefore, more prospective cohorts of HCC cases are needed to verify our conclusion.

## 5. Conclusion

Collectively, the current research synthetically evaluated the m$^6$A regulator-based m$^6$A machinery phenotypes for HCC that were characterized by different prognosis, TME features, and activation of biological processes. Moreover, quantification of the m$^6$A modification patterns by m$^6$A score may enhance the cognition of TME features as well as offer key insights into immunotherapy responses.

## Abbreviations

HCC: Hepatocellular carcinoma
m$^6$A: N$^6$-methyladenosine
mRNA: Messenger RNA
TME: Tumor microenvironment
TCGA: The cancer genome atlas
GEO: Gene-expression omnibus
CNV: Copy number variation
GSVA: Gene set variation analysis
EdU: 5-Ethynyl-2′-deoxyuridine
ssGSEA: Single sample gene set enrichment analyses
TIDE: Tumor immune dysfunction and exclusion
DEGs: Differentially expressed genes
GDSC: Genomics of drug sensitivity in cancer
IC50: Half maximal inhibitory concentration
t-SNE: T-distributed stochastic neighbor embedding
OS: Overall survival
DFS: Disease-free survival
DSS: Disease-specific survival
PFI: Progression-free interval.

## Data Availability

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

## Conflicts of Interest

The authors declare no conflicts of interest.

## Authors' Contributions

Fei Liu and Xinyue Zhang contributed equally.

## Acknowledgments

This research was supported by the Sichuan Provincial Administration of Traditional Chinese Medicine Special Project of Traditional Chinese Medicine (grant nos. 2020LC0227 and 2020LC0228).

## Supplementary Materials

Supplementary Figure 1. Consensus clustering analyses of stratifying HCC cases in TCGA cohort into three m$^6$A methylation patterns according to 23 m$^6$A regulators. (A) Heatmap for the consensus matrix $k = 3$. (B) Cumulative distribution function (CDF) under diverse $k$ values. (C) Delta area diagram for relative alterations in area under CDF curves. (D) The tracking plot for HCC samples under different $k$ values. Supplementary Figure 2. Consensus clustering analyses for clustering three m$^6$A genomic phenotypes in the light of the expression profiling of m$^6$A-associated genes in TCGA cohort. (A) Heatmap for the consensus matrix $k = 3$. (B) CDF under diverse $k$ values. (C) Delta area diagram for relative alterations in area under CDF curves. (D) The tracking plot for HCC samples under different $k$ values. Supplementary Figure 3. Subgroup analysis of the prognosis value of m$^6$A score among HCC patients in TCGA data set. Kaplan-Meir curves of cases with high or low m$^6$A score in each subgroup: (A) age ≥ 65; (B) age < 65; (C) female; (D) male; (E) G1-2; (F) G3-4; (G) stage I-II; (H) stage III-IV. $P$ values were determined through log-rank tests. Supplementary Table 1. The clinical information of HCC samples in the TCGA data set. Supplementary Table 2. The clinical information of HCC samples in the GSE14520 data set. Supplementary Table 3. The list of 331 m$^6$A phenotype-associated DEGs. (*Supplementary Materials*)

## References

[1] V. Calvaruso and J. Bruix, "Towards personalized screening for hepatocellular carcinoma: still not there," *Journal of Hepatology*, vol. 73, pp. 1319–1321, 2020.

[2] C. Yang, X. Huang, Z. Liu, W. Qin, and C. Wang, "Metabolism-associated molecular classification of

hepatocellular carcinoma," *Molecular Oncology*, vol. 14, pp. 896–913, 2020.

[3] B. H. Zhang, J. Yang, L. Jiang et al., "Development and validation of a 14-gene signature for prognosis prediction in hepatocellular carcinoma," *Genomics*, vol. 112, no. 4, pp. 2763–2771, 2020.

[4] Y. H. Lee, D. Tai, C. Yip, S. P. Choo, and V. Chew, "Combinational immunotherapy for hepatocellular carcinoma: radiotherapy, immune checkpoint blockade and beyond," *Frontiers in Immunology*, vol. 11, Article ID 568759, 2020.

[5] J. D. Yang and J. K. Heimbach, "New advances in the diagnosis and management of hepatocellular carcinoma," *BMJ*, vol. 371, Article ID m3544, 2020.

[6] A. Vogel and A. Saborowski, "Current strategies for the treatment of intermediate and advanced hepatocellular carcinoma," *Cancer Treatment Reviews*, vol. 82, Article ID 101946, 2020.

[7] X. Kong, M. Fu, X. Niu, and H. Jiang, "Comprehensive analysis of the expression, relationship to immune infiltration and prognosis of TIM-1 in cancer," *Frontiers in Oncology*, vol. 10, p. 1086, 2020.

[8] L. Zhang, J. Ding, H. Y. Li, Z. H. Wang, and J. Wu, "Immunotherapy for advanced hepatocellular carcinoma, where are we?" *Biochimica et Biophysica Acta (BBA)—Reviews on Cancer*, vol. 1874, no. 2, Article ID 188441, 2020.

[9] Z. J. Brown, T. F. Greten, and B. Heinrich, "Adjuvant treatment of hepatocellular carcinoma: prospect of immunotherapy," *Hepatology*, vol. 70, no. 4, pp. 1437–1442, 2019.

[10] L. He, H. Li, A. Wu, Y. Peng, G. Shu, and G. Yin, "Functions of N6-methyladenosine and its role in cancer," *Molecular Cancer*, vol. 18, no. 1, p. 176, 2019.

[11] X. Niu, J. Xu, J. Liu, L. Chen, X. Qiao, and M. Zhong, "Landscape of N (6)-methyladenosine modification patterns in human ameloblastoma," *Frontiers in Oncology*, vol. 10, Article ID 556497, 2020.

[12] T. Wang, S. Kong, M. Tao, and S. Ju, "The potential role of RNA N6-methyladenosine in Cancer progression," *Molecular Cancer*, vol. 19, no. 1, p. 88, 2020.

[13] X. Liu, J. Qin, T. Gao et al., "RETRACTED: YTHDF1 facilitates the progression of hepatocellular carcinoma by promoting FZD5 mRNA translation in an m6A-dependent manner," *Molecular Therapy—Nucleic Acids*, vol. 22, pp. 750–765, 2020.

[14] Y. Chen, Y. Zhao, J. Chen et al., "ALKBH5 suppresses malignancy of hepatocellular carcinoma via m (6)A-guided epigenetic inhibition of LYPD1," *Molecular Cancer*, vol. 19, no. 1, p. 123, 2020.

[15] X. Cheng, M. Li, X. Rao et al., "KIAA1429 regulates the migration and invasion of hepatocellular carcinoma by altering m6A modification of ID2 mRNA," *OncoTargets and Therapy*, vol. 12, pp. 3421–3428, 2019.

[16] 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, 2019.

[17] L. Chen, X. Niu, X. Qiao et al., "Characterization of interplay between autophagy and ferroptosis and their synergistical roles on manipulating immunological tumor microenvironment in squamous cell carcinomas," *Frontiers in Immunology*, vol. 12, Article ID 739039, 2021.

[18] Q. Wu, L. Zhou, D. Lv, X. Zhu, and H. Tang, "Exosome-mediated communication in the tumor microenvironment contributes to hepatocellular carcinoma development and progression," *Journal of Hematology & Oncology*, vol. 12, no. 1, p. 53, 2019.

[19] Y. Li, J. Gu, F. Xu et al., "Molecular characterization, biological function, tumor microenvironment association and clinical significance of m6a regulators in lung adenocarcinoma," *Briefings in Bioinformatics*, vol. 22, 2020.

[20] N. Li, Y. Kang, L. Wang et al., "ALKBH5 regulates anti-PD-1 therapy response by modulating lactate and suppressive immune cell accumulation in tumor microenvironment," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 33, pp. 20159–20170, 2020.

[21] L. Wang, H. Hui, K. Agrawal et al., "m (6 A RNA methyltransferases METTL3/14 regulate immune responses to anti-PD-1 therapy," *EMBO Journal*, vol. 39, no. 20, Article ID e104514, 2020.

[22] A. Colaprico, T. C. Silva, C. Olsen et al., "TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data," *Nucleic Acids Research*, vol. 44, no. 8, p. e71, 2016.

[23] R. A. Irizarry, B. Hobbs, F. Collin et al., "Exploration, normalization, and summaries of high density oligonucleotide array probe level data," *Biostatistics*, vol. 4, no. 2, pp. 249–264, 2003.

[24] H. Zhang, P. Meltzer, and S. Davis, "RCircos: an R package for Circos 2D track plots," *BMC Bioinformatics*, vol. 14, no. 1, p. 244, 2013.

[25] 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.

[26] S. Hänzelmann, R. Castelo, and J. Guinney, "GSVA: gene set variation analysis for microarray and RNA-seq data," *BMC Bioinformatics*, vol. 14, no. 1, p. 7, 2013.

[27] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, "clusterProfiler: an R package for comparing biological themes among gene clusters," *OMICS: A Journal of Integrative Biology*, vol. 16, no. 5, pp. 284–287, 2012.

[28] Q. Jia, W. Wu, Y. Wang et al., "Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer," *Nature Communications*, vol. 9, no. 1, p. 5361, 2018.

[29] P. Charoentong, F. Finotello, M. Angelova et al., "Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade," *Cell Reports*, vol. 18, no. 1, pp. 248–262, 2017.

[30] 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.

[31] P. Jiang, S. Gu, D. Pan et al., "Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response," *Nature Medicine*, vol. 24, no. 10, pp. 1550–1558, 2018.

[32] M. E. Ritchie, B. Phipson, D. Wu et al., "Limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, p. e47, 2015.

[33] W. Chong, L. Shang, J. Liu et al., "m(6 A regulator-based methylation modification patterns characterized by distinct tumor microenvironment immune profiles in colon cancer," *Theranostics*, vol. 11, no. 5, pp. 2201–2217, 2021.

[34] B. Zhang, Q. Wu, B. Li, D. Wang, L. Wang, and Y. L. Zhou, "m(6 A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer," *Molecular Cancer*, vol. 19, no. 1, p. 53, 2020.

[35] W. Yang, J. Soares, P. Greninger et al., "Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells," *Nucleic Acids Research*, vol. 41, no. D1, pp. D955–D961, 2012.

[36] P. Geeleher, N. Cox, and R. S. Huang, "pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels," *PLoS One*, vol. 9, Article ID e107468, 2014.

[37] S. Mariathasan, S. J. Turley, D. Nickles et al., "TGF$\beta$ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells," *Nature*, vol. 554, no. 7693, pp. 544–548, 2018.

[38] Y. Hoshida, J. P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov, "Subclass mapping: identifying common subtypes in independent disease data sets," *PLoS One*, vol. 2, no. 11, Article ID e1195, 2007.

[39] M. Chen, L. Wei, C. T. Law et al., "RNA N6-methyladenosinemethyltransferase-like 3 promotes liver cancer progression through YTHDF2-dependent post-transcriptional silencing of SOCS2," *Hepatology*, vol. 67, no. 6, pp. 2254–2270, 2018.

[40] R. Su, L. Dong, Y. Li et al., "Targeting FTO suppresses cancer stem cell maintenance and immune evasion," *Cancer Cell*, vol. 38, no. 1, pp. 79–96, 2020.

[41] P. J. Gong, Y. C. Shao, Y. Yang et al., "Analysis of N6-methyladenosine methyltransferase reveals METTL14 and ZC3H13 as tumor suppressor genes in breast cancer," *Frontiers in Oncology*, vol. 10, Article ID 578963, 2020.

[42] D. Zhu, J. Zhou, J. Zhao et al., "ZC3H13 suppresses colorectal cancer proliferation and invasion via inactivating Ras-ERK signaling," *Journal of Cellular Physiology*, vol. 234, no. 6, pp. 8899–8907, 2019.

[43] J. Li, L. Zhu, Y. Shi, J. Liu, L. Lin, and X. Chen, "m6A demethylase FTO promotes hepatocellular carcinoma tumorigenesis via mediating PKM2 demethylation," *American Journal of Translational Research*, vol. 11, no. 9, pp. 6084–6092, 2019.

[44] C. Zhang, S. Huang, H. Zhuang et al., "YTHDF2 promotes the liver cancer stem cell phenotype and cancer metastasis by regulating OCT4 expression via m6A RNA methylation," *Oncogene*, vol. 39, no. 23, pp. 4507–4518, 2020.

[45] X. Cai, Y. Chen, D. Man et al., "RBM15 promotes hepatocellular carcinoma progression by regulating N6-methyladenosine modification of YES1 mRNA in an IGF2BP1-dependent manner," *Cell Death & Disease*, vol. 7, no. 1, p. 315, 2021.

Hindawi
Journal of Oncology
Volume 2022, Article ID 5436988, 19 pages
https://doi.org/10.1155/2022/5436988

*Research Article*

# EMT-Related Gene Signature Predicts the Prognosis in Uveal Melanoma Patients

**Yufei Lv**,[1,2,3] **Lixian He**,[2] **Mengyi Jin**,[2] **Wenxin Sun**,[2] **Gang Tan**,[2] **and Zuguo Liu**[1,2,3,4]

[1]*Postdoctoral Station for Basic Medicine, Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China*
[2]*The First Affiliated Hospital, Department of Ophthalmology, Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China*
[3]*Department of Ophthalmology, Xiang'an Hospital of Xiamen University, Fujian Provincial Key Laboratory of Ophthalmology and Visual Science, Fujian Engineering and Research Center of Eye Regenerative Medicine, Eye Institute of Xiamen University, School of Medicine, Xiamen University, Xiamen, Fujian 361100, China*
[4]*Xiamen University Affiliated Xiamen Eye Center, Xiamen, Fujian 361005, China*

Correspondence should be addressed to Gang Tan; tangang99@hotmail.com and Zuguo Liu; zuguoliu@xmu.edu.cn

Received 21 April 2022; Revised 23 June 2022; Accepted 1 July 2022; Published 12 August 2022

Academic Editor: Jincheng Guo



*Background*. Uveal melanoma (UVM) is the most common primary intraocular malignancy in adults. Epithelial-mesenchymal transition (EMT) is an essential regulator of the UVM's immune microenvironment. However, the precise role of EMT in UVM remains to be explored and the development of a related treatment strategy is urgently needed. *Methods*. Multiomics data and clinical information for TCGA-UVM were used to identify the EMT subtypes and analyze their regulatory role in the immune microenvironment in UVM. A machine-learning method based on the identified subtypes was utilized to construct the EMT feature-based prognostic model. External validation cohorts GSE84976 and GSE22138 were employed to validate the model's robustness. Immunotherapy cohort IMvigor210 was used to explore the model's potential to predict immunotherapy responsiveness. *Results*. Two EMT subtypes were identified in UVM. The role of EMT in shaping the immune microenvironment and regulating cancer-immunity circle of UVM was analyzed. A robust prognostic model was presented and validated to predict patient prognosis. The model also predicted patient's immune features and immunotherapy responsiveness. *Conclusion*. The EMT-mediated immune features in UVM were illustrated, providing a reliable model to facilitate precise UVM treatment. This research may assist in decision-making during clinical UVM therapy.

## 1. Introduction

As the most common primary intraocular malignancy in adults, uveal melanoma (UVM) represents 85% of ocular melanomas [1, 2]. UVM arises from melanocytes of the uveal tract. About 85%–90% of UVM cases originate from the choroid, while the remaining occur in the iris or ciliary body [3]. Traditional first-line therapy strategy, including radiotherapy, surgical treatment, and enucleation, has reached a satisfactory rate of local disease control and long-term survival [4]. However, due to the high metastatic rate and high mortality rate secondary to metastasis, traditional therapy strategies for systemic UVM treatment remain abundant [5, 6]. Targeting therapy and immunotherapy in recent years have become the emerging components of systemic UVM treatment and have resulted in an impressive therapeutic effect in clinical practice [7–10]. However, a robust method to help identify the UVM patients and potentially obtain satisfactory clinical benefits is still lacking. The underlying mechanism of the therapy's nonresponse in UVM needs to be further explored.

Epithelial-mesenchymal transition (EMT) is a process in which epithelial cells lose their junctions and polarity and acquire the characteristics of migratory mesenchymal cells

[11]. This phenomenon of cellular plasticity usually takes place during the embryo development. However, it can also be observed during cancer progression [11, 12]. A tumor can obtain a greater migration capability and is more likely to result in distant metastasis via EMT [13]. EMT can also contribute to the formation of immune-suppressive microenvironment and raise the activity of the immune checkpoints [14]. Furthermore, EMT is correlated with the activation of tumor drug efflux pumps and antiapoptotic effects [15]. In these mechanisms, an EMT-active tumor will have more aggressive clinicopathological features and a lower responsiveness to traditional antitumor drugs and can be a potential candidate for immunotherapy [16–19].

Therefore, a robust method to help estimate UVM EMT activity in order to develop a more appropriate multidisciplinary therapy strategy for patients is urgently needed. In this research study, genomic information for UVM samples derived from The Cancer Genome Atlas (TCGA) was employed to comprehensively assess the EMT activity and illustrate the tumoral microenvironment characteristics. Two UVM subtypes were identified: EMT inactive and EMT active. These two subtypes demonstrated a distinct pattern in clinicopathological features, somatic mutation features, immune microenvironment features, and prognosis. To further facilitate the clinical application, an EMT feature-based prognostic model to predict UVM prognosis and responsiveness for targeting therapy and immunotherapy was developed using the least absolute shrinkage and selection operator (LASSO) Cox regression. Transcriptome and somatic mutation data acquired from multiple external validation cohorts were employed to comprehensively evaluate the prognostic model's efficiency. The results demonstrated that the model has great potential to be utilized as a decision-making tool to assist doctors in evaluating tumor aggressiveness and choosing optimal therapy strategy during a precise UVM treatment.

## 2. Materials and Methods

*2.1. Data Acquisition and Processing.* TCGA-UVM patient clinical information, RNA-seq data, and gene mutation data were acquired from TCGA database (https://portal.gdc.cancer.gov) and processed using *R* packages "TCGAbiolinks" and "maftools" [20–22]. Transcriptome data and clinical follow-up information of GSE84976 and GSE22138 cohorts were acquired from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/gds). Immunotherapy cohort data were acquired from published literature and obtained by the *R* package "IMvigor210CoreBiologies" [23]. All clinical data utilized in this research are publicly available. Thus, local ethical approval was not required.

*2.2. EMT Signature Analysis and Identification of the UVM Subtype.* To identify the EMT inactive subtype and EMT active subtype in UVM and to conduct further analysis, 18 EMT signatures were acquired from three published sources [24–26]. EMT signature interaction was analyzed based on the Pathway Commons (https://www.pathwaycommons.org/) database and visualized using Cytoscape software. The EMT signature's coexpression status was analyzed using the Pearson correlation and visualized with *R* package "ggcorrplot". The EMT signature's protein-protein interaction (PPI) network was constructed using Cytoscape software to visualize the landscape of the EMT signature-related protein interaction. *K*-means clustering based on these EMT signatures was used to identify the EMT inactive and active subtypes with *R* package "pheatmap." To further validate the robustness of the subtype identification, EMT subtype activity was estimated using Gene Set Variation Analysis (GSVA) according to the EMT signatures [27].

*2.3. Analysis of the Subtypes' Immune Characteristics.* To quantify the samples' immune cell infiltration levels, CIBERSORT based on the CIBERSORT tool (https://cibersort.stanford.edu/) was utilized. Samples' immune score, stromal score, and ESTIMATE score were calculated in *R* package "ESTIMATE" to evaluate the samples' immune activity [28]. The process consisted of seven steps during anticancer immune response. The steps include the release of cancer cell antigens (Step 1), cancer antigen presentation (Step 2), priming and activation (Step 3), trafficking of immune cells to tumors (Step 4), infiltration of immune cells into tumors (Step 5), recognition of cancer cells by T cells (Step 6), and killing of cancer cells (Step 7). The cancer-immunitycircle-related pathway was acquired from the published literature. Detailed pathway information is presented in Table S1. The pathway activity was evaluated using GSVA in *R* package "GSVA" [27].

*2.4. Acquisition and Analysis of DEGs between the EMT Inactive Subtype and the EMT Active Subtype.* Subtypes of differentially expressed genes (DEGs) were acquired using the threshold of $|\log FC| > 1$ and $\text{adj} - P < 0.01$ with *R* package "limma." To investigate the DEG-mediated biological function, enrichment analysis was conducted based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) database [29, 30]. The enriched terms were visualized, and clustering was analyzed using Metascape (https://metascape.org/).

*2.5. Construction of EMT Feature-Based Prognostic Model.* To construct the EMT feature-based prognostic model, patient outcome-related genes were first identified among all DEGs. Univariate Cox regression was employed to analyze the gene impact on patient outcomes. The result with $P < 0.001$ was selected. Next, the selected genes were analyzed using LASSO Cox regression to further screen the candidate genes for model construction and to calculate the coefficient of the selected genes. Finally, the EMT feature-based prognostic model (EMT score) was developed based on the selected genes and their coefficients. All analyses were conducted by the *R* package "glmnet."

*2.6. Statistical Analysis.* The *K–M* survival analysis combined with a log-rank test was used to analyze the differences in patient prognosis between the two groups. TimeROC

analysis was performed to evaluate the prediction accuracy of the EMT score. The Wilcoxon rank sum test was used to compare the continuous variables between the two groups. The differences in sample distribution in the two groups were analyzed with a chi-square test. If not specifically mentioned, $P < 0.05$ was considered statistically significant.

## 3. Results

*3.1. Identification of the EMT Inactive Subtype and the EMT Active Subtype in UVM.* A total of 18 EMT regulators acquired form published literature were analyzed in this study. Detailed interaction patterns and coexpression states of the EMT regulators are represented in Figure S1. Given the critical role of EMT in UVM, EMT inactive subtype and EMT active subtype were identified according to the EMT regulators based on the *K*-means clustering analysis. The expression level of most EMT regulators was significantly higher in the EMT active subtype (Figure 1(a)). Somatic mutation landscape analysis of the two UVM subtypes was also performed. GNA Q had the highest mutation frequency in the EMT inactive subtype (Figure 1(b)). GNA11 was the most frequent mutation gene in the EMT active subtype (Figure 1(c)). The samples' EMT activity was estimated to further validate the accuracy of subtype identification. Sample EMT scores were significantly higher in the EMT active subtype, which confirmed the EMT active subtype's high-EMT activity (Figure 1(d), $P = 1.5e - 08$). The Kaplan–Meier survival analysis demonstrated that the EMT active subtype had a significantly worse prognosis (Figure 1(e), $P = 8.4e - 05$).

*3.2. Characterization of Immune Microenvironment in the Two Subtypes.* To further investigate the two subtypes' microenvironment patterns, the immune cells' infiltration levels of the two subtypes was calculated (Figure 2(a)). In the EMT inactive subtypes, "T cells CD4 memory resting," "B cells naive," and "monocytes" had significantly higher infiltration levels (Figure 2(b)). The infiltration levels of "macrophages M1," "T cells CD4 memory activated," and "T cells CD8" were upregulated in the active EMT subtype (Figure 2(c)). The results implied that the EMT active subtype may have a higher proinflammatory immune response activity.

To elucidate the two subtypes' immune heterogeneity in detail, the two subtypes' immunity-circle-related pathway activities were compared. The EMT active subtype had a significantly higher activity of Step 1 (release of cancer cell antigens), Step 4 (trafficking of immune cells to tumors), Step 5 (infiltration of immune cells into tumors), Step 6 (recognition of cancer cells by T cells), and Step 7 (killing of cancer cells) (Figure 2(d)). It is worth noting that in the "trafficking of immune cells to tumors" process, the active EMT subtype's activity of CD4 and CD8 T cell recruiting was significantly higher, which is consistent with the results of the immune infiltration analysis (Figure 2(d)).

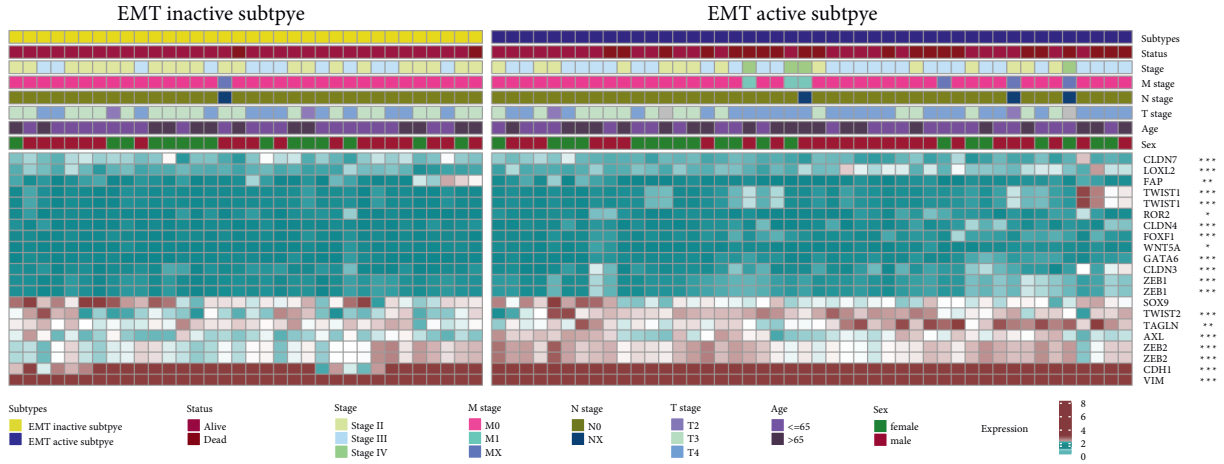
*3.3. Construction and Validation of the EMT Feature-Based Prognostic Model.* An EMT feature-based prognostic model (EMT score) was constructed using a machine-learning-based method. First, the DEGs between the EMT inactive and active subtypes were acquired. A total of 317 genes were identified (Figure S2A). Enrichment analysis of the DEGs also indicated that the two subtypes had a distinct immune microenvironment pattern (Figures S2B and S2C). Next, univariate Cox regression was employed to identify the patient outcome-related genes. A total of 117 genes were identified and submitted for the subsequent analysis (Table S2). Finally, LASSO Cox regression was used to identify the most robust prognostic genes among them and to calculate the coefficients of the selected genes (Figures 3(a) and 3(b)). The EMT feature-based gene prognostic model was described as follows:

$$\text{EMT score} = \sum_{a} \text{Coefficient}\left(\text{Gene}_a\right) * \text{Expression level}\left(\text{Gene}_a\right). \quad (1)$$

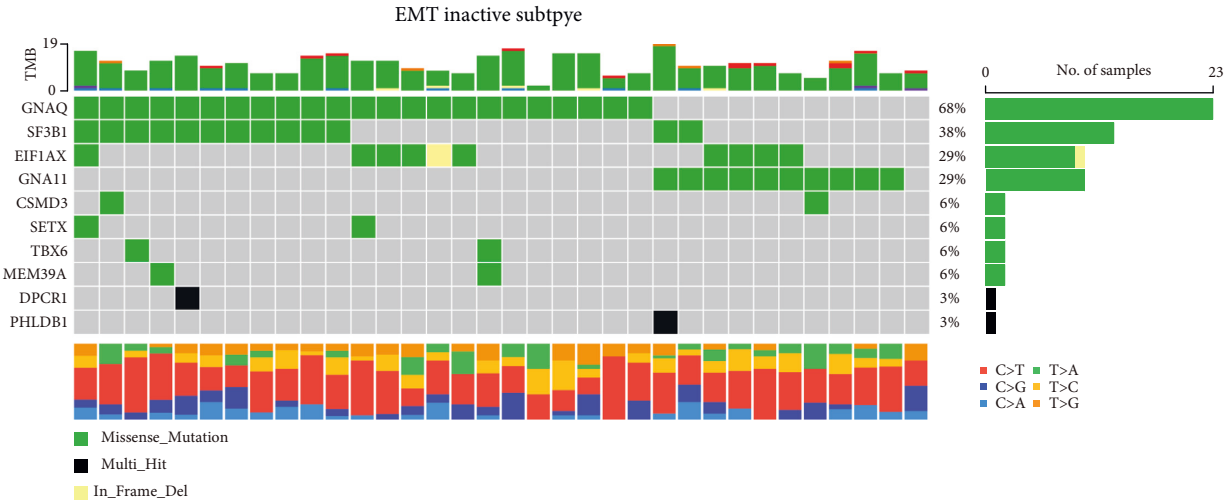
Selected genes and their coefficients are presented in Figure 3(c).

To validate the robustness of the EMT score, patient outcomes in the high- and low-EMT score groups were compared. Results indicated that patients in the high-EMT score group had a significantly poorer prognosis (Figures 3(d), 3(e), $P = 6.6e - 09$, and Figure 3(f), AUC = 0.958). Then, the correlation between the EMT score and clinical pathology characteristics was analyzed. The high-EMT score sample was more likely to be in the advanced-stage cancer (Figures S3A–S3C). The external validation cohorts GSE84976 and GSE22138 were also included to further test the EMT score's accuracy. According to the results, the model had a great prognostic value in the two validation cohorts. A high-EMT score predicted poor prognosis in UVM patients (Figures 4(a), 4(b), $P = 3.9e - 04$, Figure 4(c), AUC = 0.847, Figure 4(d), Figure 4(e), $P < 0.0001$, and Figure 4(f), AUC = 0.75). Then, in order to explore the EMT score's potential in reflecting the tumor's immune features, the immune score of the high- and low-EMT score groups was calculated in the two external validation cohorts. The high-EMT score group had a significantly higher immune score, stromal score, and ESTIMATE score in the two cohorts, which indicated that the high-EMT score group had a significantly higher immune activity (Figures 4(g) and 4(h)).

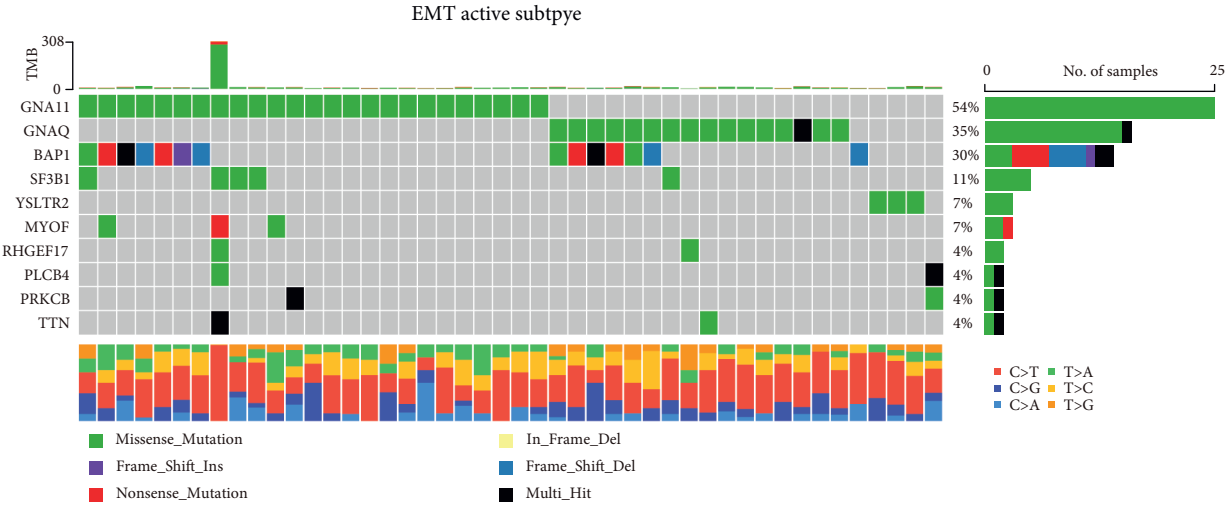
*3.4. Exploration of a Potential Therapy Strategy Targeting the High-EMT Score Tumor.* To further explore the potential therapy strategy targeting EMT active UVM, the UVM sample EMT scores were calculated according to the model and divided into high- and low-EMT score groups. According to the immune infiltration analysis, the high-EMT score group had a significantly higher infiltration level of "macrophages M1," "T cells CD4 memory activated," "T cells CD8," "T cells follicular helper," and "T cells gamma delta," which indicated the high-EMT score's potential in reflecting a proinflammatory immune response (Figure 5(a)). On the contrary, a low-EMT score predicted the resting-like immune response pattern (Figure 5(b)). To further explore the EMT score's potential




(a)




(b)




(c)

Figure 1: Continued.

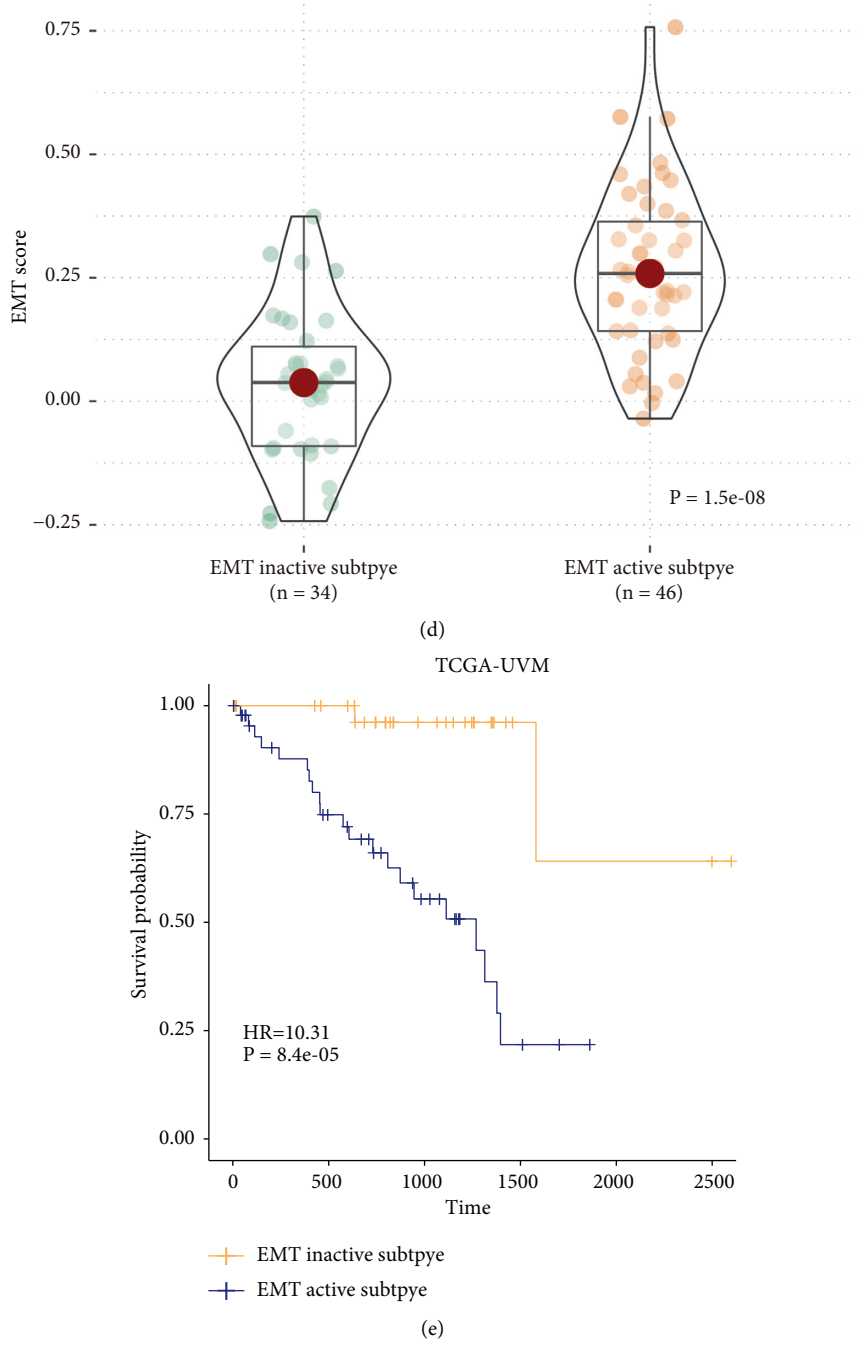



Figure 1: Identification of UVM's EMT inactive subtype and EMT active subtype. (a) Landscape of UVM EMT subtypes and clinicopathological features. (b) EMT inactive subtype's somatic mutation features. (c) EMT active subtype's somatic mutation features. (d) Comparison of two subtypes' EMT activity. (e) Comparison of two subtypes' patient prognosis. $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$.

in elucidating the immune characteristics of the UVM's immunity circle, the correlation between the EMT score and the activity of the cancer-immunitycircle-related pathway was analyzed. The EMT score was able to reflect the activity of the immune cell-recruiting and cancer cell-killing process rather than the cancer-immunity circle's initial process (Figure 5(c)).

To validate the model's efficiency in predicting immunotherapy responsiveness, the immunotherapy cohort Imvigor210 was employed for subsequent analysis. In the high-EMT score group, samples had a significantly higher tumor mutation burden, which inferred the high-EMT score group's potentially high immunotherapy response rate (Figure 5(d)). To validate our hypothesis, the immunotherapy response rate of the high-EMT score group was compared to that of the low-EMT score group. The high-EMT score group had a relatively higher frequency of PR and CR (Figure 5(e)). These results demonstrated that the EMT score had great potential in predicting cancer immunotherapy responsiveness.

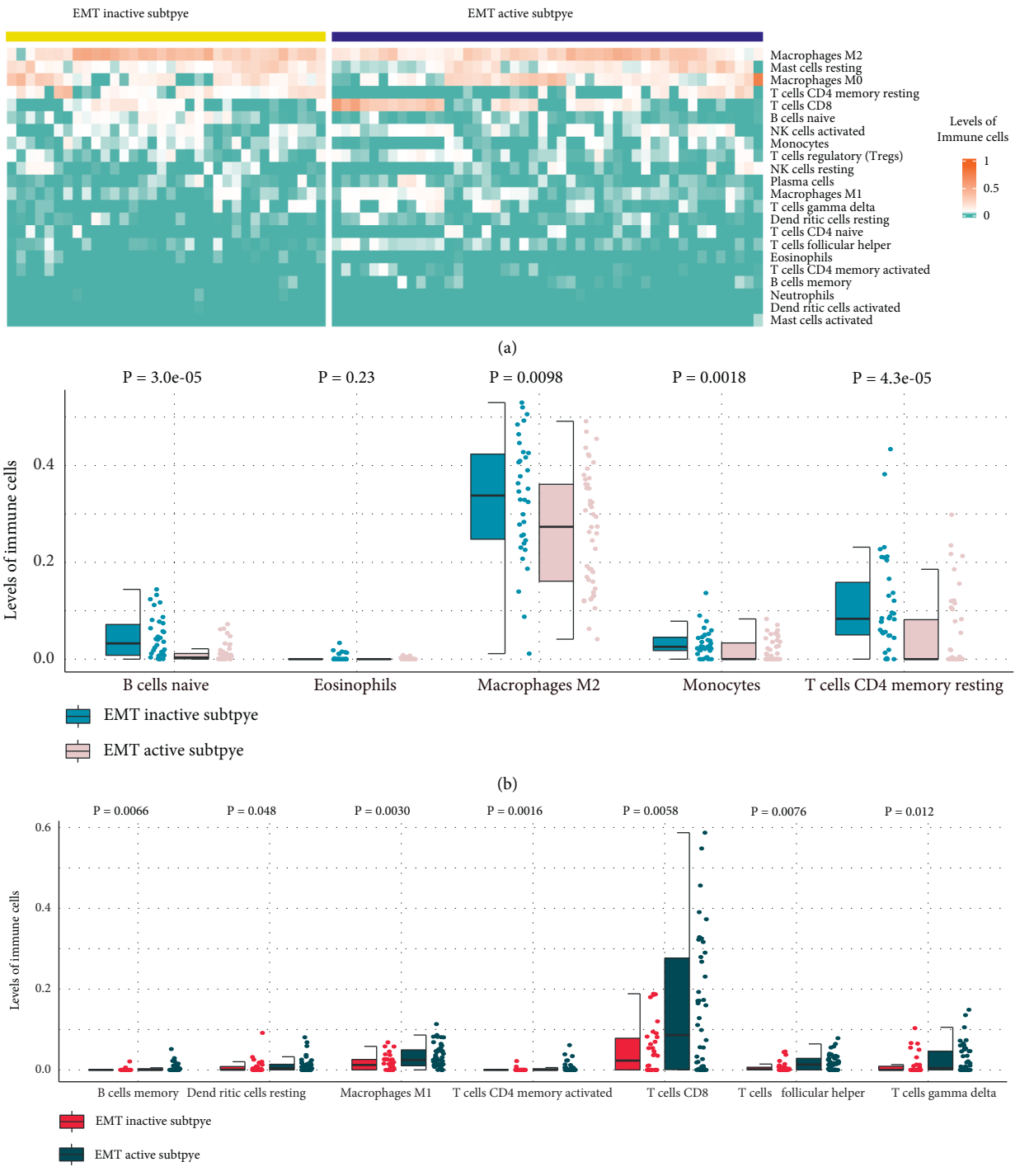
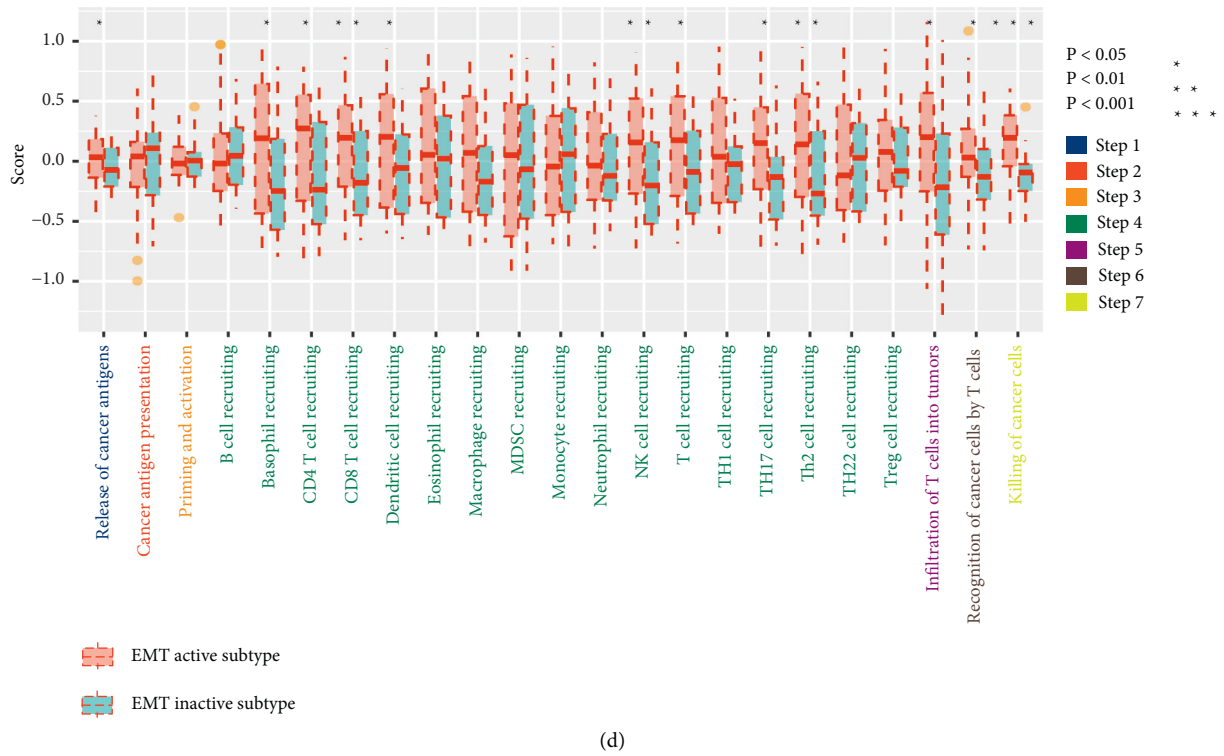



Figure 2: Continued.



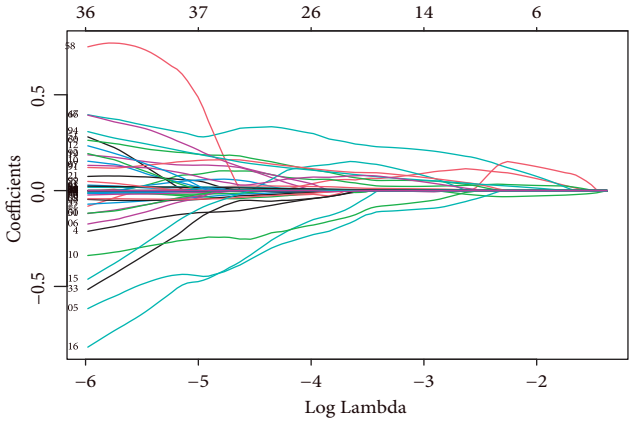

(d)

Figure 2: Comparison of immune characteristics of the UVM's EMT inactive subtype and EMT active subtype. (a) Subtypes' immune infiltration level was calculated using CIBERSHOT. (b, c) Comparison of two subtypes' immune infiltration levels. (d) Comparison of two subtypes' cancer immunity-related pathway activities.

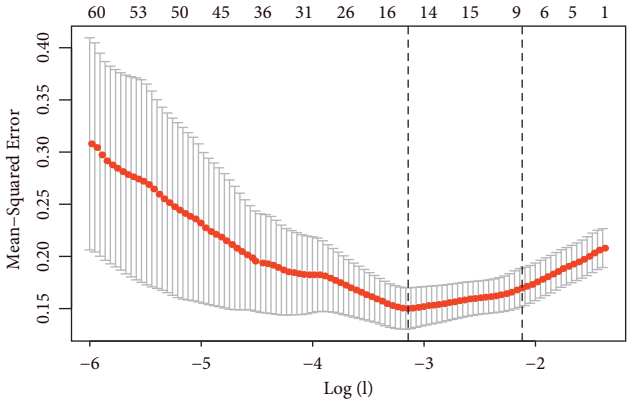
## 4. Discussion

UVM is a malignancy with a relatively low incidence rate but poor prognosis [4]. Although surgery and radiotherapy are effective treatment strategies for primary tumors, therapy options are limited once UVM becomes metastatic [31, 32]. Recent research has demonstrated that EMT plays an essential role in promoting UVM metastasis and contributing to the disease's poor prognosis [33–35]. However, systematic analysis to illustrate EMT-mediated tumor heterogeneity in UVM is still lacking. Robust biomarkers based on EMT features to reflect UVM's aggressiveness are also limited. Therefore, in this study, we integrated multiomics data to develop an EMT feature-based prognostic model and systematically analyzed the EMT-mediated immune microenvironment in UVM. First, EMT inactive and active subtypes were identified in UVM according to the EMT signatures acquired from multiple published sources. Somatic mutation in a GNA family gene, such as GNA11 and GNAQ, which encodes guanine nucleotide-binding protein G$\alpha$ subunits of the G$\alpha$q family, is the driver of UVM initiation [36, 37]. GNAQ had the highest mutation frequency in the EMT inactive subtype, while GNA11 had the highest mutation frequency in the EMT active subtype. These results implied the potentially different therapy targets for EMT inactive and active UVM. Then, DEGs between the two subtypes were acquired and patient outcome-related genes were obtained using univariate Cox regression. Next, the EMT feature-based gene prognostic model was trained using patient outcome-related genes. Finally, the model's efficiency in predicting patient outcomes and therapy responsiveness was verified. This research may assist doctors in evaluating patient prognosis and choosing suitable therapy strategies in clinical practice.

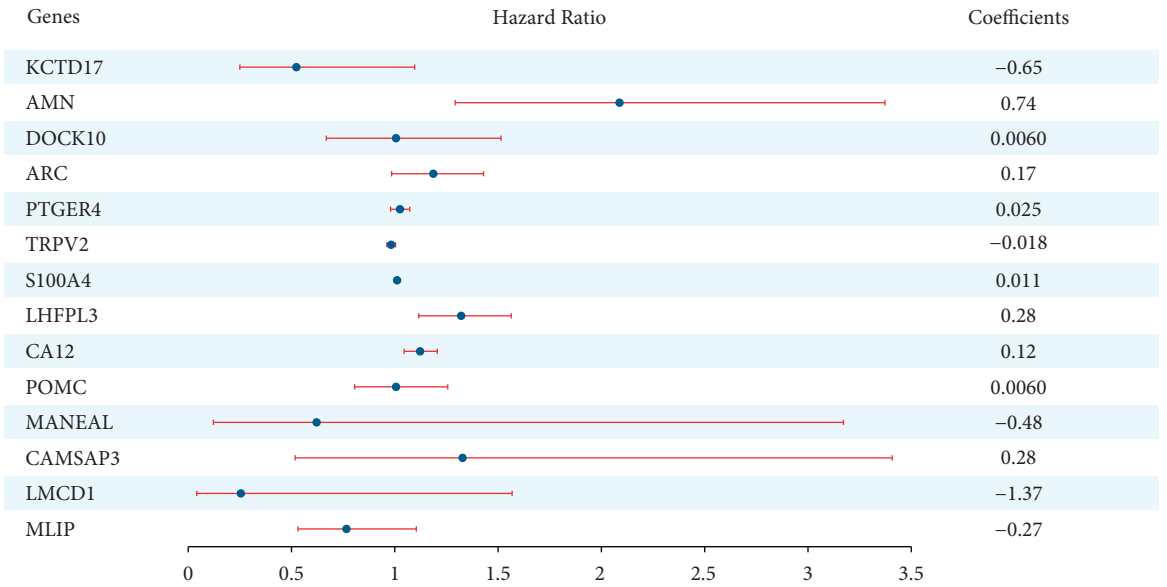
The eye is considered an immune-privileged organ with partial or even completely suppressed immune responses [38]. While the concentration of antitumor immune cells in the microenvironment is correlated with a better outcome in most cancer types, the immune infiltration in UVM can direct to poor prognosis [39]. Emerging studies have emphasized the strong interaction between EMT and tumor immune microenvironment [40]. Here, we analyzed the immune microenvironment characteristics of EMT inactive and active subtypes. The results demonstrated that EMT may result in the high infiltration and high activity of proinflammatory immune cells. EMT plays a crucial role in tumor microenvironment progression. For example, EMT transcriptional factors, including Snail, Zeb1, and Twist1, can attract cancer-related immune cells and shape tumor microenvironment into a protumor subtype [41, 42]. In turn, the modulated microenvironment can promote cancer EMT [43, 44]. Thus, therapy strategies that can interfere with this positive feedback system may introduce clinical benefits to the EMT active UVM subtype. When comparing the two groups' antitumor immune process activities, we found that the EMT active subtype's active anticancer-immune-related pathways were concentrated on the immune cells recruiting the related process and "infiltration of T cells into tumors,"




(a)




(b)




(c)

Figure 3: Continued.

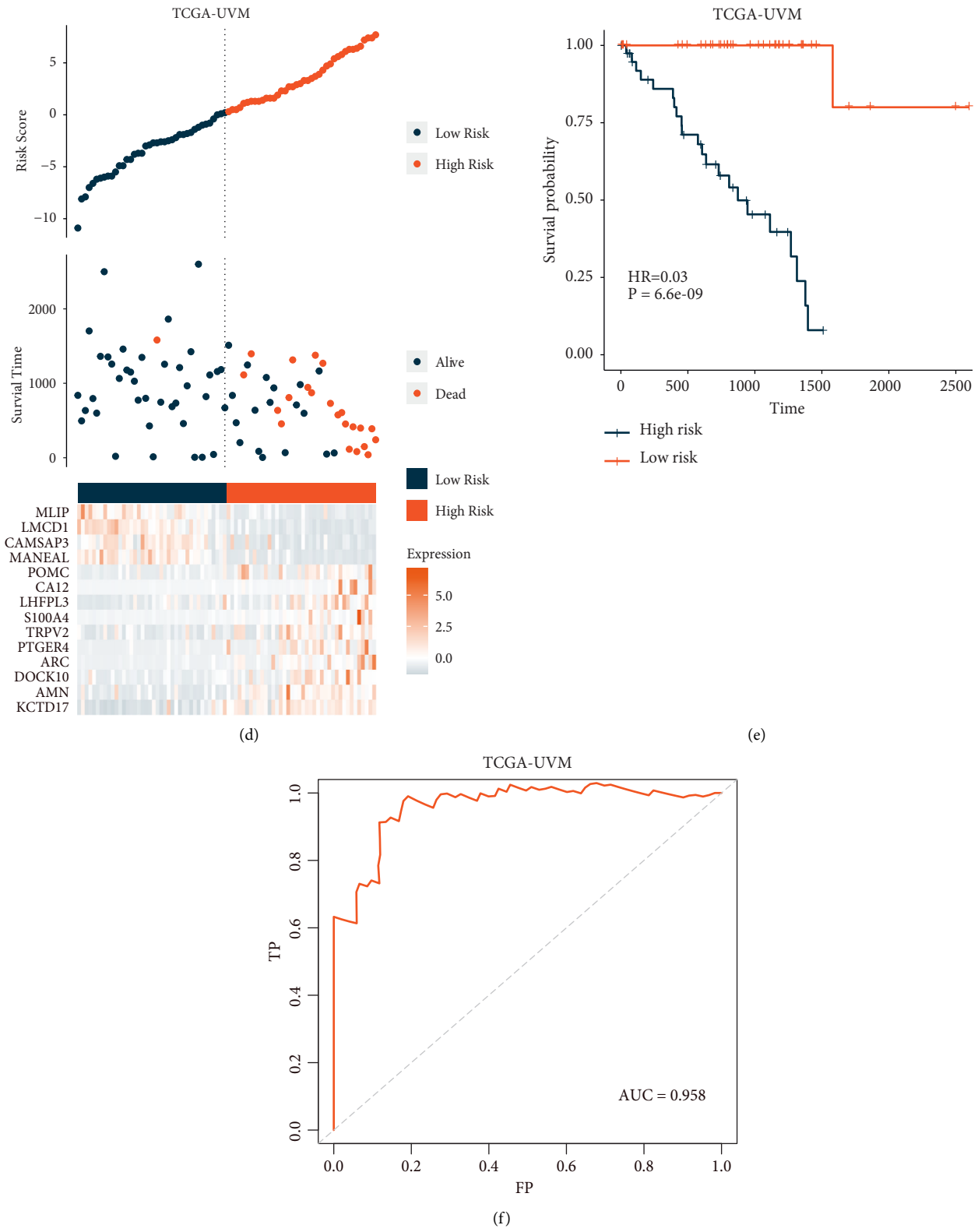
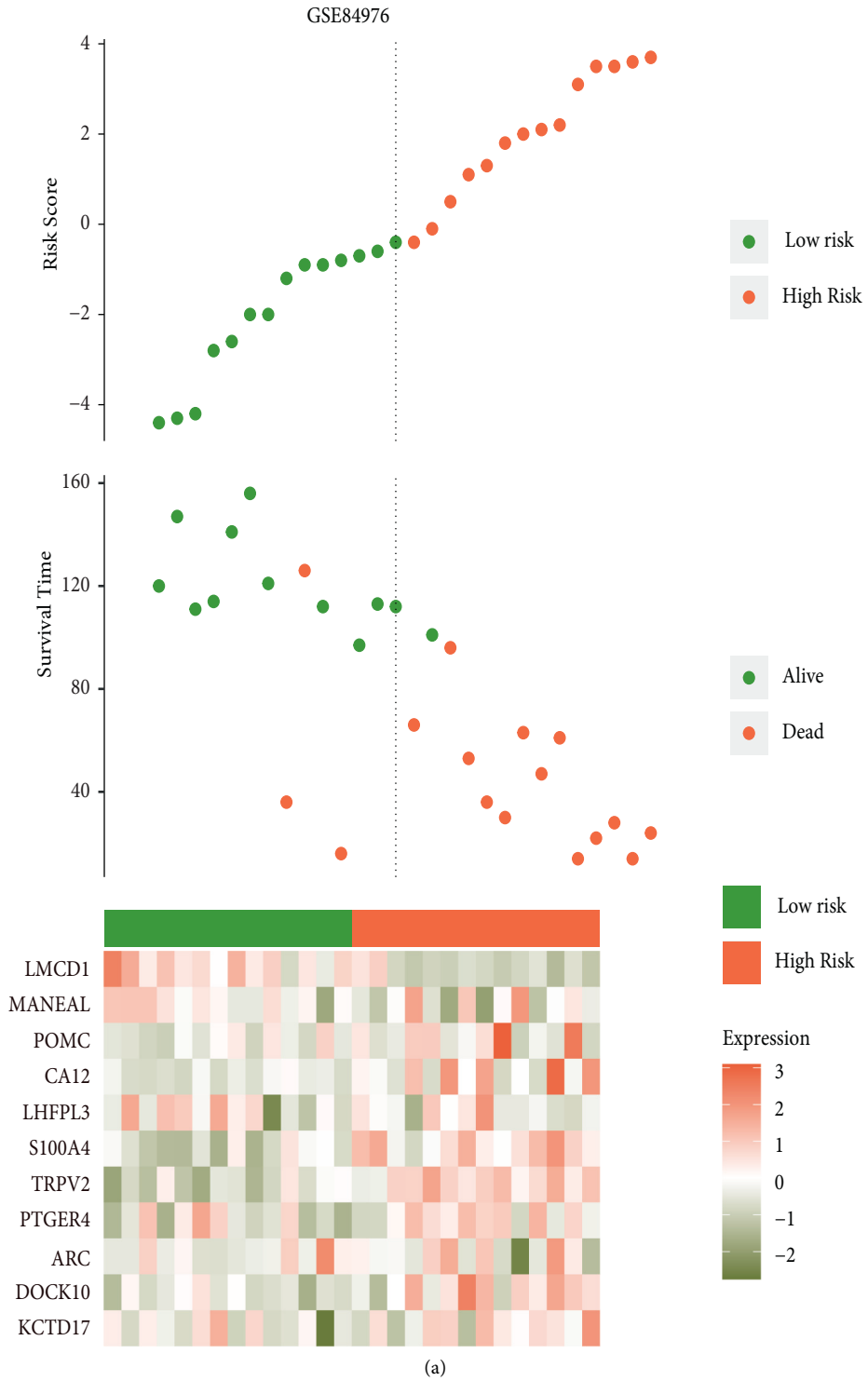
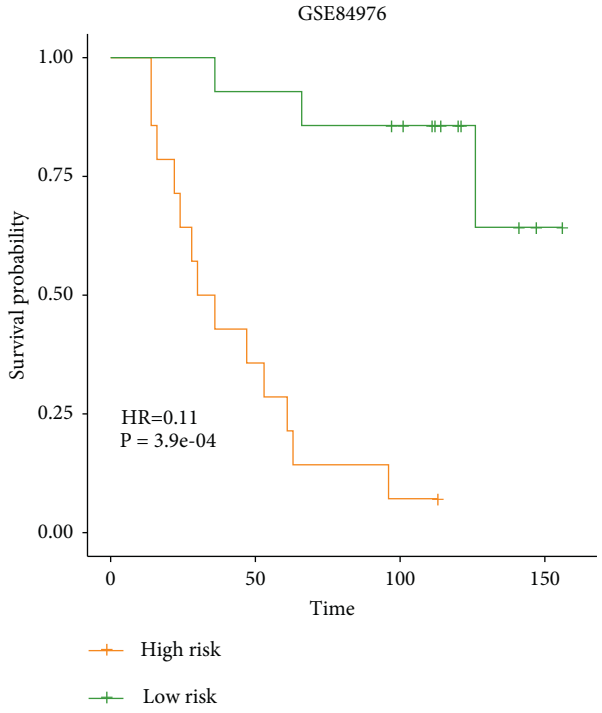



Figure 3: Construction of the EMT score using a machine-learning-based method. (a, b) Gene prognostic model was constructed using LASSO Cox regression. (c) Regression coefficients of model component genes. (d) EMT score component genes' expression status in TCGA-UVM cohort patients. (e) Comparison of patient prognosis in high- and low-EMT score groups from the TCGA-UVM cohort. (f) EMT score ROC curve in the TCGA-UVM cohort.

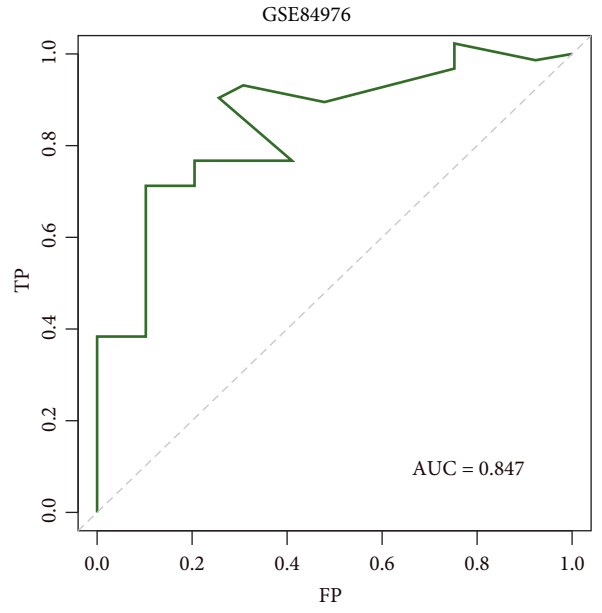



(a)

Figure 4: Continued.

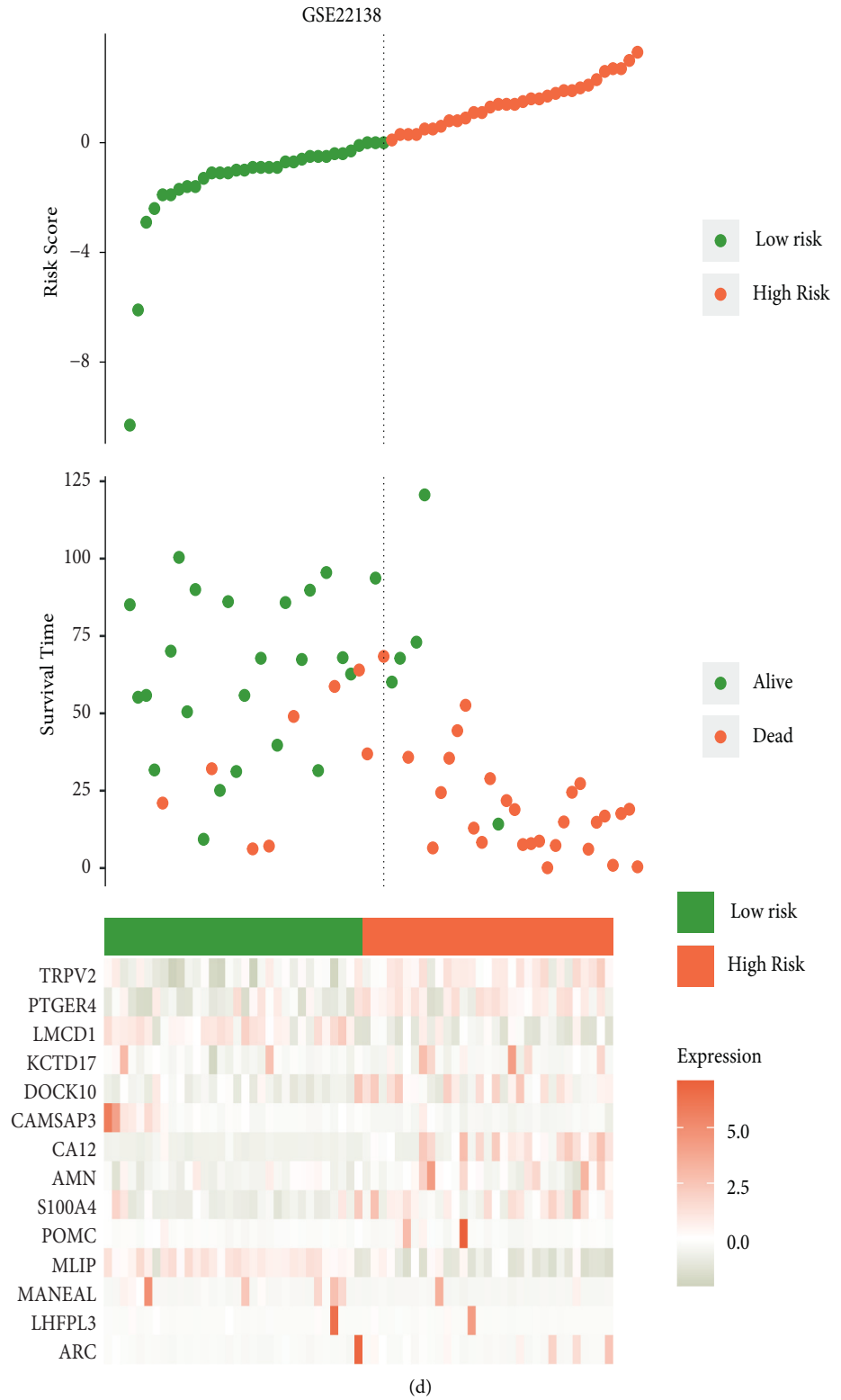



(b)



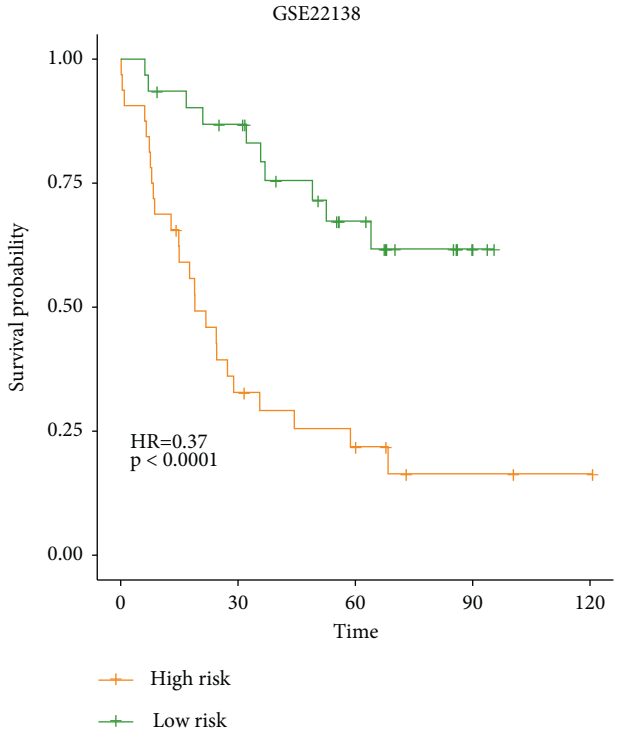

(c)

Figure 4: Continued.

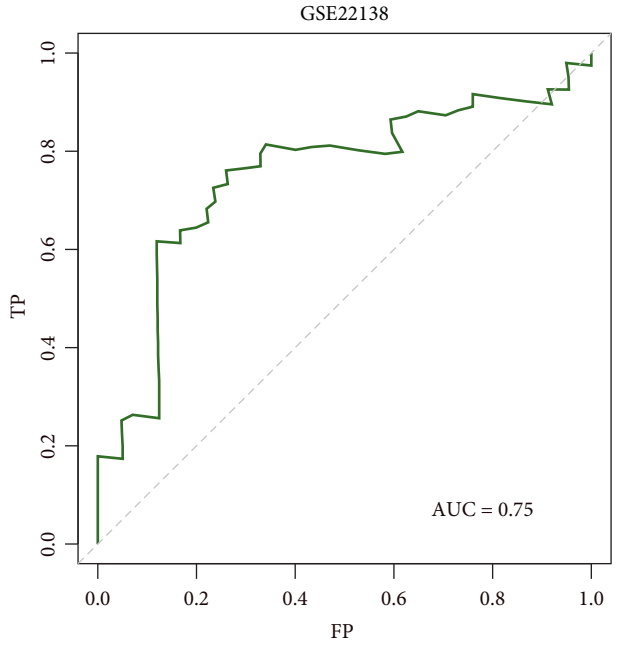



(d)

Figure 4: Continued.

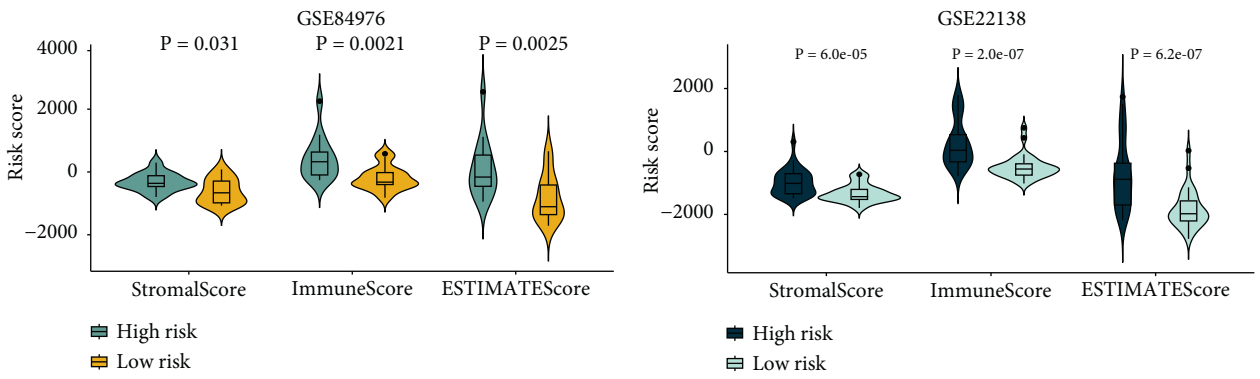



(e)



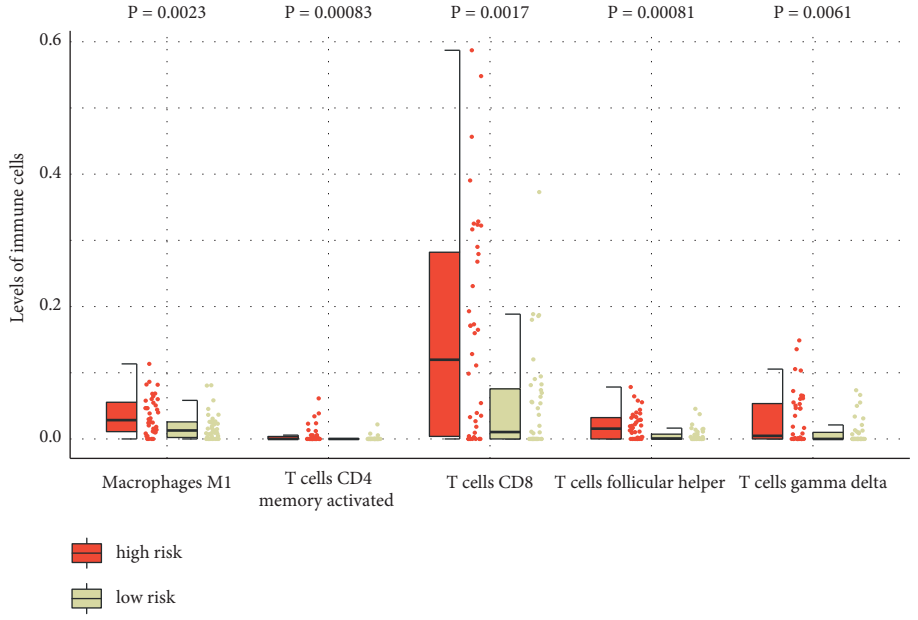

(f)

Figure 4: Continued.



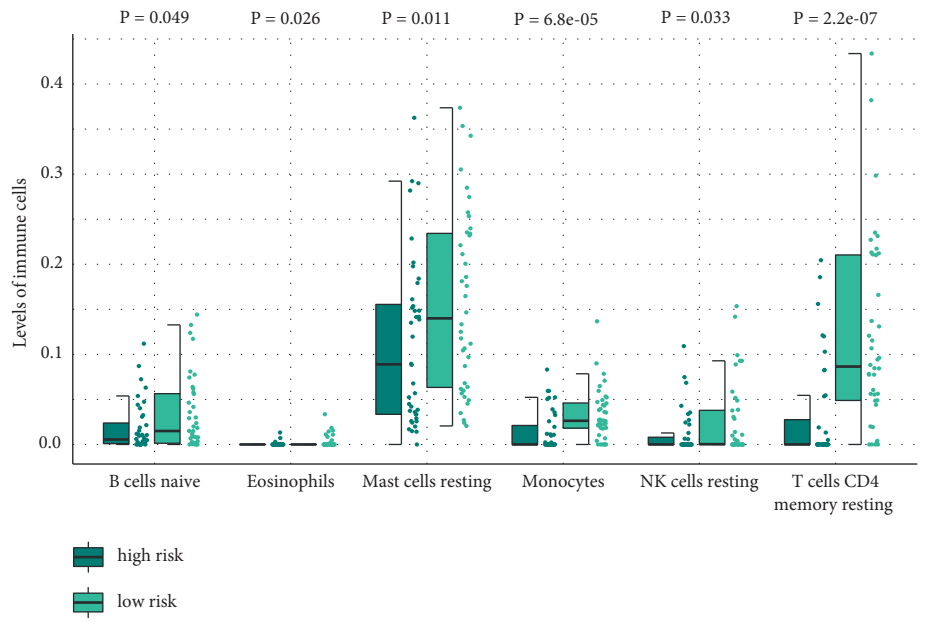

(g) (h)

Figure 4: Validation of EMT score robustness in external validation sets. (a) EMT score component genes' expression status in patients from the GSE84976 cohort. (b) Comparison of patient prognosis in high- and low-EMT score groups from the GSE84976 cohort. (c) EMT score's ROC curve in the GSE84976 cohort. (d) EMT score component genes' expression status in patients from the GSE22138 cohort. (e) Comparison of patient prognosis in high- and low-EMT score groups in the GSE22138 cohort. (f) EMT score's ROC curve in the GSE22138 cohort. (g) Comparison of patient's ESTIMATE scores in high- and low-EMT score groups from the GSE84976 cohort. (h) Comparison of patient's ESTIMATE scores in high- and low-EMT score groups from the GSE22138 cohort.

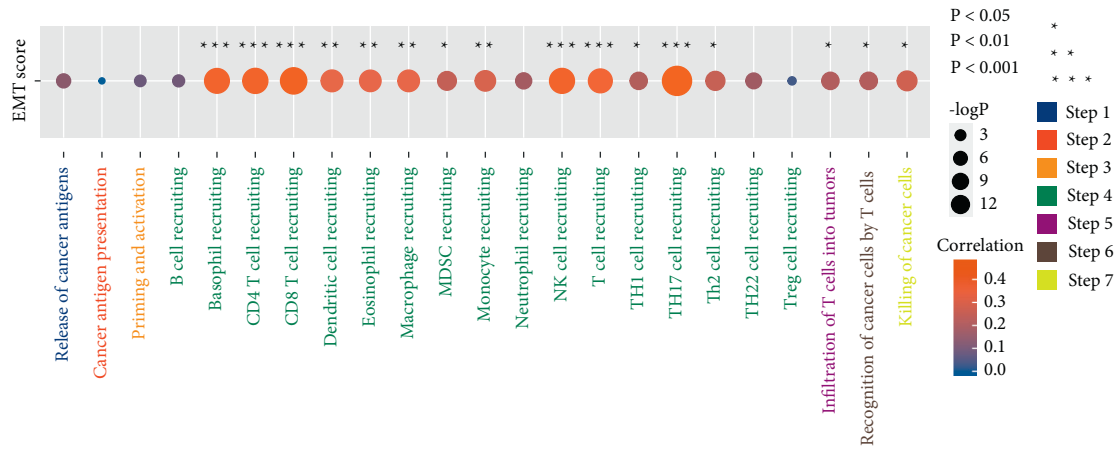



(a)

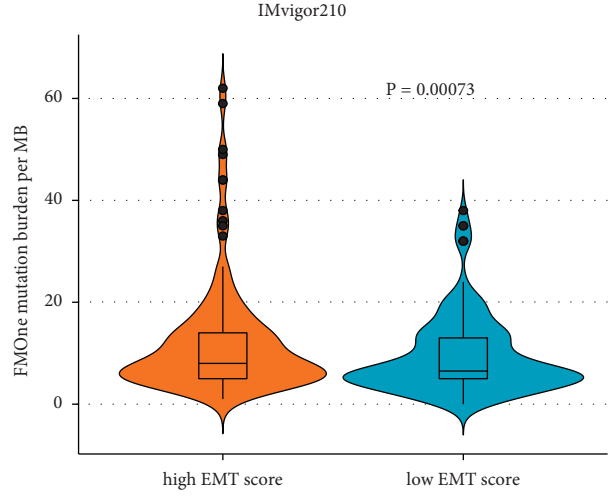
Figure 5: Continued.




(b)




(c)




(d)

Figure 5: Continued.

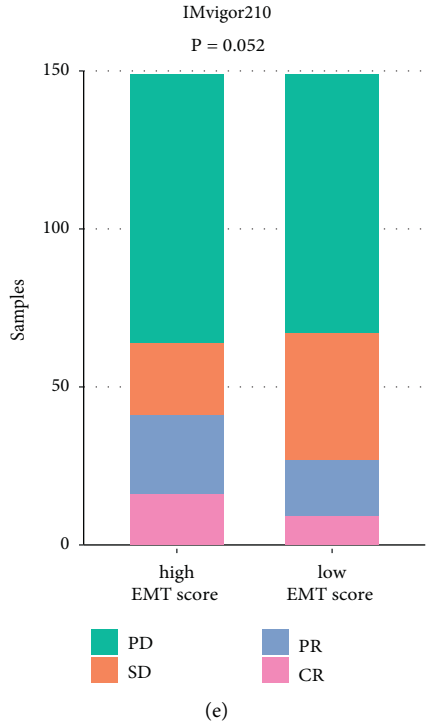



Figure 5: Exploration of EMT score's potential in predicting immunotherapy responsiveness. (a, b) Comparison of immune infiltration level between high- and low-EMT score groups. (c) Correlation analysis between EMT score and activity of the cancer immune-related pathway. (d) Comparison of tumor mutation burden in high- and low-EMT score groups. (e) Comparison of immunotherapy responsiveness rate in high- and low-EMT score groups. PD, progressive disease; SD, stable disease; PR, partial response; CR, complete response.

"recognition of cancer cells by T cells," and "killing of cancer cells" pathway. These EMT-mediated immune features may be the potential underlying mechanism for EMT-mediated UVM's poor prognosis. This EMT-mediated immune characteristic also suggested that a therapy targeting these immune processes may be suitable for the EMT active UVM.

UVM has a high metastatic rate, and clinical outcomes for metastatic UVM are unsatisfactory [45]. While the nonmetastatic UVM has a relatively good prognosis, once distant UVM metastases have occurred, the clinical treatment strategy will be limited [2]. Thus, apart from the traditional pathological detection methods, a supplementary method to assess the UVM's clinical pathology features is important. It was found that the tumor EMT score was significantly higher in advanced UVM (T4, M1, and stage IV). The EMT score also had great efficiency in evaluating UVM prognosis and immune activity. Therefore, the EMT score may be developed as a novel biomarker to predict UVM prognosis, metastasis status, and immune features.

The cancer-immunity cycle reflects the immune response of the inherent and adaptive immune systems to UVM. The goal of cancer immunotherapy is to initiate or reinitiate a self-sustaining cycle of cancer immunity to help the immune system conduct the cytotoxic tumor-killing process [46]. Every step of the cancer-immunity circle plays an essential role in immunotherapy response. Cancer immunotherapy strategies need to be developed based on the features of the cancer-immunity circle. For example, a cancer vaccine would be suitable for Step 1 (release of cancer antigens) dysfunction cancer subtype [47]. The activities of the cancer-immunitycycle-related pathways reflect the comprehensive immunomodulatory interactions in UVM tumor microenvironment. The present study explored the EMT score's potential in predicting the activity of the pathways. According to the results, the EMT score was able to reflect the activity of the immune cell recruiting process and tumor cytotoxic killing-related processes, while the correlation between the EMT score and activity of tumor antigen-based immune cell activation-related processes was not observed. A novel therapy strategy targeting these immunity-related pathways might introduce clinical benefits to UVM patients with high-EMT scores.

Cancer immunotherapy, which targets the tumor immune escape mechanisms and activates the body's immune system to recognize and attack cancer cells, has become the emerging strategy for comprehensive cancer treatments in clinical practice [48–51]. Immunotherapy has brought revolutionary progress to clinical tumor treatment. The application of immune checkpoint blockade in malignancies, including melanoma, urothelial bladder cancer, head and neck squamous cell carcinoma, and classical Hodgkin's lymphoma, has brought significant clinical benefits to patients [52–55]. However, the nonignorable nonresponsiveness rate and the related side effects in the clinical practice have been the major obstacles for its implementation [56]. In the present research, the EMT score

provides a method to identify the potential candidates for UVM immunotherapy. The results indicated that patients with high-EMT scores may have the potentially high responsiveness rate for immunotherapy.

The present study had some limitations. First, the enrolled UVM and immunotherapy cohorts were limited. More clinical information and transcriptome data should be utilized to validate the EMT score efficiency. Second, although the EMT-mediated immune microenvironment alteration was analyzed, the detailed underlying immune regulatory network remains to be further elucidated. Third, the EMT score's robustness should be further tested based on clinical trials. Related research would be important for the EMT score's clinical application. These shortcomings will be alleviated with the development of large data pools and further research.

## 5. Conclusions

In summary, we developed an EMT score to predict UVM patient prognosis, immune microenvironment characteristics, and immunotherapy responsiveness in a machine-learning-based method. The EMT score robustness was validated by two external validation cohorts. The EMT score predicted the UVM patient outcomes and immune activity in the training and validation cohorts. The immunotherapy cohort-based analysis revealed the EMT score's potential in the preliminary identification of immunotherapy candidates. This research may facilitate precise treatment in a further clinic-integrated oncology therapy of UVM.

## Data Availability

The microarray data supporting this study are from previously reported databases, which have been cited. The processed data are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported by grants from the National Key R&D Program of China (no. 2018YFA0107304, ZL) and National Natural Science Foundation of China (no. 81870627, ZL). The authors thank International Science Editing (http://www.internationalscienceediting.com) for editing this manuscript.

## Supplementary Materials

Figure S1: EMT signature interaction status, coexpression status, and PPI network. Figure S2: differential expression gene acquisition and enrichment analysis. Figure S3: analysis of correlation between clinicopathological features and EMT score. Table S1: cancer-immunity circle pathway. Table S2: Cox regression analysis for genes. (*Supplementary Materials*)

## References

[1] A. E. Chang, L. H. Karnell, and H. R. Menck, "The national cancer data base report on cutaneous and noncutaneous melanoma: a summary of 84,836 cases from the past decade. The American college of surgeons commission on cancer and the American cancer society," *Cancer*, vol. 83, pp. 1664–1678, 1998.

[2] A. D. Singh, M. E. Turell, and A. K. Topham, "Uveal melanoma: trends in incidence, treatment, and survival," *Ophthalmology*, vol. 118, pp. 1881–1885, 2011.

[3] C. C. McLaughlin, X. C. Wu, A. Jemal, H. J. Martin, L. M. Roche, and V. W. Chen, "Incidence of noncutaneous melanomas in the USA," *Cancer*, vol. 103, no. 5, pp. 1000–1007, 2005.

[4] S. Kaliki and C. L. Shields, "Uveal melanoma: relatively rare but deadly cancer," *Eye*, vol. 31, no. 2, pp. 241–257, 2017.

[5] C. L. Shields, S. Kaliki, S. U. Shah, W. Luo, M. Furuta, and J. A. Shields, "Iris melanoma: features and prognosis in 317 children and adults," *Journal of American Association for Pediatric Ophthalmology and Strabismus*, vol. 16, no. 1, pp. 10–16, 2012.

[6] S. Kaliki, C. L. Shields, and J. A. Shields, "Uveal melanoma: estimating prognosis," *Indian Journal of Ophthalmology*, vol. 63, no. 2, pp. 93–102, 2015.

[7] T. J. de Vries, D. Trancikova, D. J. Ruiter, and G. N. van Muijen, "High expression of immunotherapy candidate proteins gp100, MART-1, tyrosinase and TRP-1 in uveal melanoma," *British Journal of Cancer*, vol. 78, no. 9, pp. 1156–1161, 1998.

[8] R. D. Carvajal, G. K. Schwartz, T. Tezel, B. Marr, J. H. Francis, and P. D. Nathan, "Metastatic disease from uveal melanoma: treatment options and future prospects," *British Journal of Ophthalmology*, vol. 101, no. 1, pp. 38–44, 2017.

[9] X. Chen, Q. Wu, P. Depeille et al., "RasGRP3 mediates MAPK pathway activation in GNAQ mutant uveal melanoma," *Cancer Cell*, vol. 31, no. 5, pp. 685–696.e6, 2017.

[10] R. D. Carvajal, S. Piperno-Neumann, E. Kapiteijn et al., "Selumetinib in combination with dacarbazine in patients with metastatic uveal melanoma: a phase III, multicenter, randomized trial (SUMIT)," *Journal of Clinical Oncology*, vol. 36, no. 12, pp. 1232–1239, 2018.

[11] G. Babaei, S. G. G. Aziz, and N. Z. Z. Jaghi, "EMT, cancer stem cells and autophagy; the three main axes of metastasis," *Biomedicine & Pharmacotherapy*, vol. 133, Article ID 110909, 2021.

[12] M. Saitoh, "Involvement of partial EMT in cancer progression," *The Journal of Biochemistry*, vol. 164, no. 4, pp. 257–264, 2018.

[13] N. M. Aiello and Y. Kang, "Context-dependent EMT programs in cancer metastasis," *Journal of Experimental Medicine*, vol. 216, no. 5, pp. 1016–1026, 2019.

[14] Y. Jiang and H. Zhan, "Communication between EMT and PD-L1 signaling: new insights into tumor immune evasion," *Cancer Letters*, vol. 468, pp. 72–81, 2020.

[15] B. Du and J. S. Shim, "Targeting epithelial-mesenchymal transition (EMT) to overcome drug resistance in cancer," *Molecules*, vol. 21, no. 7, p. 965, 2016.

[16] D. C. Voon, R. Y. Huang, R. A. Jackson, and J. P. Thiery, "The EMT spectrum and therapeutic opportunities," *Molecular Oncology*, vol. 11, no. 7, pp. 878–891, 2017.

[17] M. Singh, N. Yelle, C. Venugopal, and S. K. Singh, "EMT: mechanisms and therapeutic implications," *Pharmacology & Therapeutics*, vol. 182, pp. 80–94, 2018.

[18] H. T. Chen, H. Liu, M. J. Mao et al., "Crosstalk between autophagy and epithelial-mesenchymal transition and its application in cancer therapy," *Molecular Cancer*, vol. 18, no. 1, p. 101, 2019.

[19] I. Georgakopoulos-Soares, D. V. Chartoumpekis, V. Kyriazopoulou, and A. Zaravinos, "EMT factors and metabolic pathways in cancer," *Frontiers in Oncology*, vol. 10, p. 499, 2020.

[20] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "Review the cancer genome atlas (TCGA): an immeasurable source of knowledge," *Współczesna Onkologia*, vol. 1A, pp. A68–A77, 2015.

[21] A. Colaprico, T. C. Silva, C. Olsen et al., "TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data," *Nucleic Acids Research*, vol. 44, no. 8, p. e71, 2016.

[22] A. Mayakonda, D. C. Lin, Y. Assenov, C. Plass, and H. P. Koeffler, "Maftools: efficient and comprehensive analysis of somatic variants in cancer," *Genome Research*, vol. 28, no. 11, pp. 1747–1756, 2018.

[23] S. Mariathasan, S. J. Turley, D. Nickles et al., "TGF$\beta$ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells," *Nature*, vol. 554, no. 7693, pp. 544–548, 2018.

[24] J. S. Damrauer, K. A. Hoadley, D. D. Chism et al., "Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology," *Proceedings of the National Academy of Sciences*, vol. 111, no. 8, pp. 3110–3115, 2014.

[25] J. Hedegaard, P. Lamy, I. Nordentoft et al., "Comprehensive transcriptional analysis of early-stage urothelial carcinoma," *Cancer Cell*, vol. 30, no. 1, pp. 27–42, 2016.

[26] W. Hugo, J. M. Zaretsky, L. Sun et al., "Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma," *Cell*, vol. 165, no. 1, pp. 35–44, 2016.

[27] S. Hänzelmann, R. Castelo, and J. Guinney, "GSVA: gene set variation analysis for microarray and RNA-seq data," *BMC Bioinformatics*, vol. 14, no. 1, p. 7, 2013.

[28] 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.

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

[30] 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.

[31] C. Chattopadhyay, D. W. Kim, D. S. Gombos et al., "Uveal melanoma: from diagnosis to treatment and the science in between," *Cancer*, vol. 122, no. 15, pp. 2299–2312, 2016.

[32] M. J. Jager, C. L. Shields, C. M. Cebulla et al., "Uveal melanoma," *Nature Reviews Disease Primers*, vol. 6, no. 1, p. 24, 2020.

[33] L. Asnaghi, G. Gezgin, A. Tripathy et al., "EMT-associated factors promote invasive properties of uveal melanoma cells," *Molecular Vision*, vol. 21, pp. 919–929, 2015.

[34] Q. Hou, S. Han, L. Yang et al., "The interplay of MicroRNA-34a, LGR4, EMT-associated factors, and MMP2 in regulating uveal melanoma cells," *Investigative Opthalmology & Visual Science*, vol. 60, no. 13, pp. 4503–4510, 2019.

[35] S. Liu, J. Zhang, S. Fang et al., "Antitumor efficacy of oncolytic HSV-1 expressing cytosine deaminase is synergistically enhanced by DPD down-regulation and EMT inhibition in uveal melanoma xenograft," *Cancer Letters*, vol. 495, pp. 123–134, 2020.

[36] M. J. C. Vader, M. C. Madigan, M. Versluis et al., "GNAQ and GNA11 mutations and downstream YAP activation in choroidal nevi," *British Journal of Cancer*, vol. 117, no. 6, pp. 884–887, 2017.

[37] C. D. Van Raamsdonk, V. Bezrookove, G. Green et al., "Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi," *Nature*, vol. 457, no. 7229, pp. 599–602, 2009.

[38] J. V. Forrester and H. Xu, "Good news-bad news: the Yin and Yang of immune privilege in the eye," *Frontiers in Immunology*, vol. 3, p. 338, 2012.

[39] I. H. G. Bronkhorst, T. H. K. Vu, E. S. Jordanova, G. P. M. Luyten, S. H. v. d. Burg, and M. J. Jager, "Different subsets of tumor-infiltrating lymphocytes correlate with macrophage influx and monosomy 3 in uveal melanoma," *Investigative Opthalmology & Visual Science*, vol. 53, no. 9, pp. 5370–5378, 2012.

[40] S. Singh and R. Chakrabarti, "Consequences of EMT-driven changes in the immune microenvironment of breast cancer and therapeutic response of cancer cells," *Journal of Clinical Medicine*, vol. 8, no. 5, p. 642, 2019.

[41] T. Brabletz, R. Kalluri, M. A. Nieto, and R. A. Weinberg, "EMT in cancer," *Nature Reviews Cancer*, vol. 18, no. 2, pp. 128–134, 2018.

[42] U. D. Kahlert, J. V. Joseph, and F. A. E. Kruyt, "EMT- and MET-related processes in nonepithelial tumors: importance for disease progression, prognosis, and therapeutic opportunities," *Molecular Oncology*, vol. 11, no. 7, pp. 860–877, 2017.

[43] E. Romeo, C. A. Caserta, C. Rumio, and F. Marcucci, "The vicious cross-talk between tumor cells with an EMT phenotype and cells of the immune system," *Cells*, vol. 8, no. 5, p. 460, 2019.

[44] L. A. Horn, K. Fousek, and C. Palena, "Tumor plasticity and resistance to immunotherapy," *Trends in Cancer*, vol. 6, no. 5, pp. 432–441, 2020.

[45] K. Mahendraraj, C. S. Lau, I. Lee, and R. S. Chamberlain, "Trends in incidence, survival, and management of uveal melanoma: a population-based study of 7, 516 patients from the surveillance, epidemiology, and end results database (1973– 2012) (1973-2012)," *Clinical Ophthalmology*, vol. 10, pp. 2113–2119, 2016.

[46] D. S. Chen and I. Mellman, "Oncology meets immunology: the cancer-immunity cycle," *Immunity*, vol. 39, pp. 1–10, 2013.

[47] T. A. Ferguson, J. Choi, and D. R. Green, "Armed response: how dying cells influence T-cell functions," *Immunological Reviews*, vol. 241, no. 1, pp. 77–88, 2011.

[48] A. Steven, S. A. Fisher, and B. W. Robinson, "Immunotherapy for lung cancer," *Respirology*, vol. 21, no. 5, pp. 821–833, 2016.

[49] J. S. O'Donnell, M. W. L. Teng, and M. J. Smyth, "Cancer immunoediting and resistance to T cell-based immunotherapy," *Nature Reviews Clinical Oncology*, vol. 16, no. 3, pp. 151–167, 2019.

[50] C. Y. Soekojo, M. Ooi, S. de Mel, and W. J. Chng, "Immunotherapy in multiple myeloma," *Cells*, vol. 9, no. 3, p. 601, 2020.

[51] Y. Jadoon and M. A. Siddiqui, "Immunotherapy in multiple myeloma," *Cancer Treatment and Research Communications*, vol. 29, Article ID 100468, 2021.

[52] F. S. Hodi, S. J. O'Day, D. F. McDermott et al., "Improved survival with ipilimumab in patients with metastatic melanoma," *New England Journal of Medicine*, vol. 363, no. 8, pp. 711–723, 2010.

[53] J. Larkin, V. Chiarion-Sileni, R. Gonzalez et al., "Combined nivolumab and ipilimumab or monotherapy in untreated melanoma," *New England Journal of Medicine*, vol. 373, no. 1, pp. 23–34, 2015.

[54] L. Q. Chow, R. Haddad, S. Gupta et al., "Antitumor activity of pembrolizumab in biomarker-unselected patients with recurrent and/or metastatic head and neck squamous cell carcinoma: results from the phase ib KEYNOTE-012 expansion cohort," *Journal of Clinical Oncology*, vol. 34, no. 32, pp. 3838–3845, 2016.

[55] S. M. Ansell, A. M. Lesokhin, I. Borrello et al., "PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma," *New England Journal of Medicine*, vol. 372, no. 4, pp. 311–319, 2015.

[56] L. B. Kennedy and A. K. S. Salama, "A review of cancer immunotherapy toxicity," *CA: A Cancer Journal for Clinicians*, vol. 70, no. 2, pp. 86–104, 2020.

Hindawi
Journal of Oncology
Volume 2022, Article ID 2151396, 15 pages
https://doi.org/10.1155/2022/2151396

*Research Article*

# Assessing the Prognostic Capability of Immune-Related Gene Scoring Systems in Lung Adenocarcinoma

**Wenhao Liu,[1] Ruihong Dong,[2] Shuai Gao,[3] Xiaodi Shan,[1] Mian Li,[4] Zhaoyan Yu,[5] and Liang Sun[1]**

[1]*College of Artificial Intelligence and Big Data For Medical Sciences, Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan, Shandong, China*
[2]*Beijing Mentougou Hospital of Traditional Chinese Medicine, Beijing, China*
[3]*Department of Rehabilitation, Beijing Rehabilitation Hospital of Capital Medical University, Beijing, China*
[4]*First Affiliated Hospital of Shandong First Medical University, Biomedical Sciences College & Shandong Medicinal Biotechnology Centre, Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan, Shandong 250000, China*
[5]*Department of Otorhinolaryngology, Shandong Public Health Clinical Center, Jinan, Shandong, China*

Correspondence should be addressed to Zhaoyan Yu; yuzhaoyan@163.com and Liang Sun; sunliang@sdfmu.edu.cn

Received 7 April 2022; Revised 26 May 2022; Accepted 10 June 2022; Published 31 July 2022

Academic Editor: Hongqing Cai



*Background*. Lung adenocarcinoma (LUAD) is the commonest of the subtypes of lung cancer histologically. For this study, we intended to analyze the expression profiling of the immune-related genes (IRGs) from an independently available public database and developed a potent signature predictive of patients' prognosis. *Methods*. Gene expression profiles and the clinical data of lung adenocarcinoma were gathered from the Gene Expression Omnibus database (GEO) and The Cancer Genome Atlas (TCGA), and the obtained data were split into a training set ($n = 226$), test set ($n = 83$), and validation set ($n = 400$). IRGs were then gathered from the ImmPort database. A prognostic model was constructed by analyzing the training set. Then the GO and KEGG analysis was performed, and a gene correlation prognostic nomogram was constructed. Finally, external validation, such as immune infiltration and immunohistochemistry, was performed. *Results*. The 110 genes were significant by univariate Cox regression analysis and randomized survival forest algorithm for the training set and showed a good distinction between the low-risk-score and high-risk-score groups in the training set ($P < 0.0001$) by screening for four prognosis-related genes (HMOX1, ARRB1, ADM, PDIA3) and validated by the test set GSE30219 ($P = 0.0025$) and TCGA dataset ($P = 0.00059$). Multivariate Cox showed that the four gene signatures were an individual risk factor for LUAD. In addition, the genes in the signatures were externally verified using an online database. In particular, PDIA3 and HMOX1 are essential genes in the prognostic nomogram and play an important role in the model of immune-related genes. *Conclusion*. Four immune-related genetic signatures are reliable prognostic indicators for patients with LUAD, providing a relevant theoretical basis and therapeutic rationale for immunotherapy.

## 1. Introduction

Lung cancer is major cancer in the world and adenocarcinoma patients lack tumor-specific clinical symptoms in the early stage, and local infiltration and even distant metastasis occur in the middle and late stages of lung cancer, with poorer efficacy and overall survival rate. Currently, 60% of NSCLC is LUAD, of which NSCLS is the major component of lung cancer [1, 2]. In the United States, in 2021, the number of new incidences in a year was 235,760, and the number of deaths reached 131,880 [3]. For treating patients with early LUAD, surgical lobectomy is the most frequently used method. However, 10–44% of them have a less favorable prognosis five years after surgery [4]. Therefore, it is necessary to develop more effective biomarkers to obtain more effective prognostic models of LUAD patients to help early diagnosis of LUAD and treat different patients with reasonable treatment plans so that patients can receive more

appropriate treatment for their own conditions and get the best treatment results.

It is generally accepted that cancer is an incredibly complex disease that involves the interplay between the tumor and the immune system [5]. The human immune-related systems have been confirmed to play a critical part in the development and progression of aggressive cancer [6, 7]. More recently, immune checkpoint blockade therapy has shown remarkable efficacy in treating solid tumors (melanoma, lung cancer, and so on). This treatment is based on a dynamic process between tumor and immunity, and most current studies predict efficacy based on marker levels before or at a point in time during treatment. Programmed death-ligand (PD-1) and tumor mutation burden (TMB) are widely used as prognostic biomarkers for immunotherapy [8]. However, immune-related therapies are only available for some patients. There are remarkable personal differences in therapeutic effectiveness, illustrating the complexity of cancer-causing mechanisms and the existence of tumor heterogeneity [9, 10]. A few studies have reported that immune-related genes can predict prognosis in tumor survivors and provide potential targets for immunotherapeutic treatment [11–13]. However, the prognostic model of immune-related genes in early-stage LUAD patients is relatively rare.

This study explores early LUAD-related prognostic models based on immune-related genes based on gene expression datasets from GEO and TCGA. After we matched the immune-related genes list to the three cohorts, the resulting genes were screened by a series of methods. An immune signature based on four genes for prognosis with outstanding predictive capability was established. Then enriched for the function of related genes. The infiltration of tumors into immune cells was also analyzed, and finally, the differential expression of the four genes in different cancers was analyzed. It helps physicians in the prognosis of early LUAD patients. Furthermore, it provides some theoretical basis for personalized treatment.

## 2. Materials and Methods

Figure 1 shows the flow chart of this work.

*2.1. Expression Data.* The Gene Expression Omnibus (GEO, GSE31210, GSE30219) and The Cancer Genome Atlas (TCGA) databases, which were widely used and generally recognized, were selected as the data sources for the study, from which clinical information and gene expression profiles of lung adenocarcinoma patients. GSE31210 were selected as training and GSE30219 and TCGA were selected as external validation sets, respectively. The total number of patients of the three cohorts was 709. The distribution of cases in the three data was training set GSE31210 ($N = 226$) and the validation sets are GSE30219 ($N = 83$) and TCGA ($N = 400$), respectively. Clinical information and gene expression profiles related to the dataset were collected according to the following methods. In addition, the GSE50081 ($N = 128$) dataset was also collected for the validation of the Kaplan–Meier analysis. In addition, the GPL570 microarray platform was annotated by probes to obtain the final expression profiles of the GEO data [13].

*2.2. Selection of IRGs set.* Immunology Database and Analysis Portal (ImmPort) is an open repository of discipline-level human immunology databases for translation and clinical research. The IRGs set was obtained from ImmPort, and the related *R* package "clusterProfiler" was used to match the corresponding ensemble and other related information. Finally, 1455 IRGs were obtained.

*2.3. Development of the Prognostic Gene Signature.* Prognostic immunity-related genes were obtained by univariate Cox regression screening on survival status for each gene in the training set. Genes that showed significant differences ($P < 0.01$) in the Cox regression were subsequently analyzed using the stochastic survival forest algorithm (RSFA) for dimensionality reduction [14]. In addition, the equation is shown below:

$$\textbf{Risk Score} = \sum_{\mathbf{i}=1}^{\mathbf{N}} (\textbf{Exp}\,\mathbf{r_i} + \textbf{coeff}_\mathbf{i}), \tag{1}$$

The meaning of each parameter is as follows: in which N is the amount of IRGs, Expr is the IRGs expression value, and coeff is the coefficient value of IRGs. In the following section, patients are classified into low and high-score-risk groups according to the median value.

We combined the genes generated in RSFA with sufficient scale to prevent overfitting during modeling. The obtained combinations of IRGs were tested for their performance using time-correlated receiver operating characteristic (time-ROC) analysis. The gene combinations with the largest area under the curve (AUC) were used for subsequent predictions, and the model is statistically significant. The accuracy of the model for prognosis was subsequently validated in internal and external validation.

Clinical information about cancer patients in the dataset may also be an influential factor in prognosis. So we analyzed the relationship between the two in the following way. Evaluation of patient prognostic outcomes and their relationship to clinical characteristics Chi-square analysis was used to determine the association between the IRGs prediction model and clinical information. KM survival curves and log-rank test were used to characterize the interaction between the model and survival time. In addition, the clinical information in the training and validation sets was related to OS using multivariate Cox regression analysis. The final univariate Cox analysis was utilized to determine whether clinical features could assist immune-related gene models used together for prognosis.

*2.4. Construction and Evaluation of an IRG Nomogram.* A predictive nomogram for gene expression correlation was developed. Subsequently, we used the calibration curve to detect its accuracy in the GEO and TCGA cohorts. In

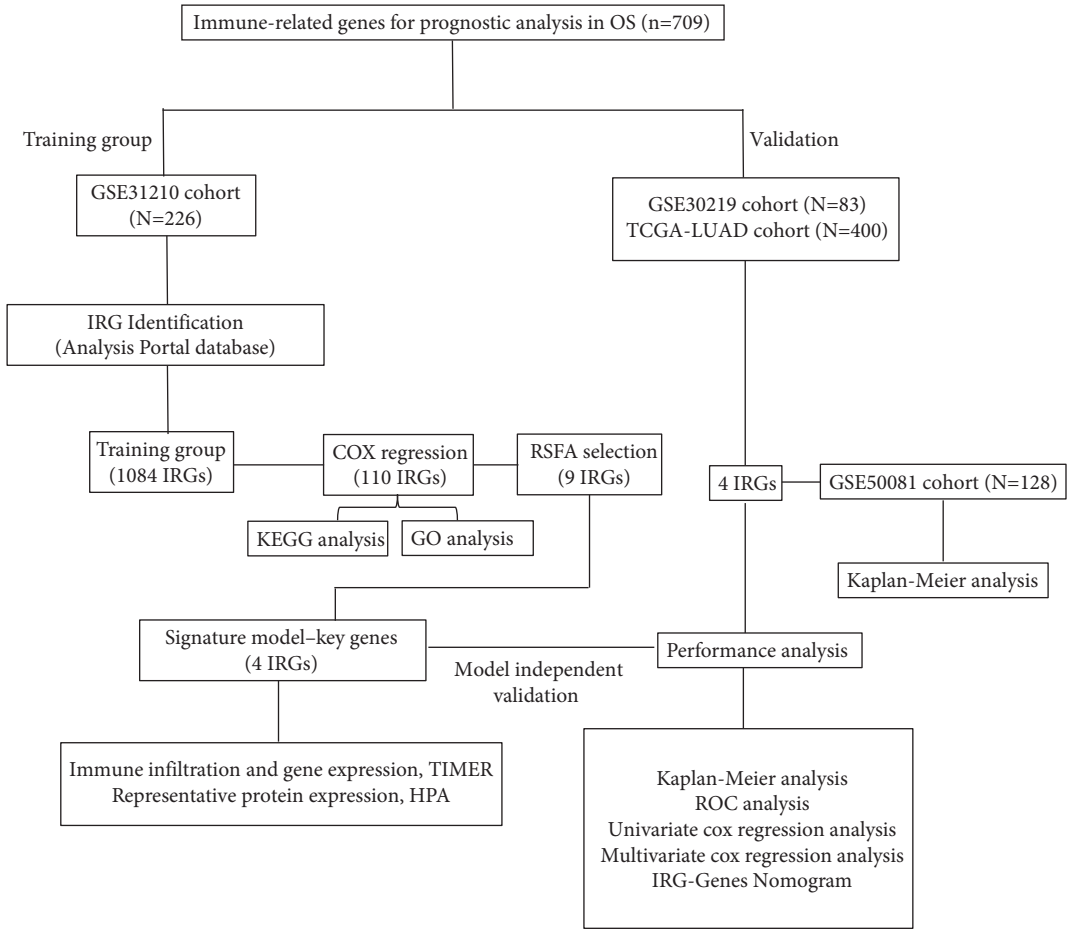



Figure 1: Flow chart of this study.

addition, the prediction bias of the nomogram was assessed for evaluation.

*2.5. Detection of the Infiltration of Immune Cells with Prognosis-Related Irgs in Tumors.* Tumor Immune Estimation Resource (TIMER) is a website that uses gene expression profiling data to detect the infiltration of immune cells in tumor tissue. We applied it to explore the infiltration of immune cells with the four genes (HMOX1, ARRB1, ADM, and PDIA3) in LUAD and six infiltrating immune cells (B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells) were available for analysis [9, 15].

*2.6. External Validation of the IRGs Signature.* Furthermore, TIMER database was utilized to validate the differential expression of the four prognostic IRGs in different cancers. The Human Protein Atlas (HPA) is based on proteomic, transcriptomic, and systems biology data and covers protein expression not only in tumors but also in normal tissues. It was used to verify the protein levels in the model.

*2.7. Functional Enrichment Analysis.* Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were applied using the clusterProfiler package to analyze the underlying biological processes of the prognostic IRGs [16].

*2.8. Statistical Analysis.* Data analysis was performed by *R* (version 4.1.2). KM and cox analyses were applied in three separate datasets using the *R* package "survival." Cox analysis was utilized to identify prognostic IRGs and detect the models. Chi-square or rank-sum tests were applied to stratified variables. "ROC" and "TimeROC" can be used to validate the model's viability. Functional annotation is performed using the "Clusterprofiler" package. The *P*-values involved in the analysis were $< 0.05$, which was a significant statistical value.

## 3. Results

*3.1. Patient Population Information.* The number of LUAD patients was obtained from GSE31210 and GSE30219, which were 226 and 83, respectively, and 400 LUAD patients were gathered from the TCGA database. A sum of 1084 IRGs was

Table 1: Clinical information of the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) datasets.

| Characteristic | GSE31210 | GSE30219 | TCGA |
|---|---|---|---|
| Age (years) | | | |
| >61 | 122 | 37 | 251 |
| ≤61 | 104 | 46 | 149 |
| Sex | | | |
| Female | 121 | 65 | 217 |
| Male | 105 | 18 | 183 |
| Vital status | | | |
| Alive | 191 | 39 | 278 |
| Dead | 35 | 42 | 122 |
| Pathological stage | | | |
| Stage I | 168 | — | 280 |
| Stage II | 58 | — | 120 |
| T stage | | | |
| T1 | — | 69 | — |
| T2 | — | 12 | — |

identified for expression in the GSE31210 dataset. In Table 1, we can easily obtain the median age of the patients as 61 years. The number of male and female cases in the training set was 105 and 121, respectively, with 35 deaths and 191 living cases, and the median OS was 5.33 years. Data per sample from GSE31210 and TCGA data were distributed in pathological stages I-II of LUAD, and data per sample from GSE30219 data were distributed in *T* stages T1-T2 of LUAD.

*3.2. Identifying four Prognostic IRGs in the Training Set.* After univariate cox analysis and ROC curve analysis, 110 prognosis-related genes were obtained by screening based on $P < 0.01$ and AUC > 0.6, and the results are shown in Supplementary Table S1. Nine IRGs were next obtained by importance scoring in a randomized survival forest, and subsequent permutations of these nine genes yielded $2^9 - 1 = 511$ prediction models (Figure 2(a) and 2)(b). Then, these models were evaluated by AUC, and the combination of four genes, HMOX1, ARRB1, ADM, and PDIA3, was found to have the most considerable AUC value of 0.779 (Figure 2)(c), which had the optimal predictive power. The risk score was shown in the following formula:

$$\begin{aligned}\textbf{Risk Score} = {} & (1.32004 \times \textbf{HMOX}1) + (-1.36259 \times \textbf{ARRB}1) \\ & + (1.17434 \times \textbf{ADM}) + (1.44023 \times \textbf{PDIA3})\end{aligned} \tag{2}$$

where the gene name represents the expression level of this gene in a particular sample. From the equation, we can get the coefficients of 1.32004, −1.36259, 1.17434, and 1.44023 for HMOX1, ARRB1, ADM, and PDIA3, respectively.

We computed the risk scores for each patient using the RSF formula and plotted heat maps for the four genes (Figures 3(a)–3)(c). The number of deaths increased with the increase of risk score in the three datasets. In the high-risk-score group, two genes, HMOX1, ADM, were highly expressed and ARRB1 was lowly expressed, with the same pattern in all three data sets. However, there was no significant pattern in the expression of the PDIA3 gene. Among the relevant clinical features, relapses become more frequent as risk scores increase, but no significant trends were found for age and gender following the score during the three cohort. By cox analysis, ARRB1 was highly expressed in the low-risk-score group, which was a protective factor (Table 2).

*3.3. Evaluation of a Prognostic IRGs Signature.* A risk score was computed for each individual with the prognostic models of immune-related genes. In the training cohort, the KM analysis was used to confirm the difference in survival between the high-risk-score ($N = 113$) and low-risk-score ($N = 113$) populations classified by a median score [17, 18]. In Figure 4(a), the 5-year survival incidence was 58.07% ($N = 66$) in the low -risk-score group vs. 32.74% ($N = 37$) in the high-risk-score group. In the training cohort, significant OS was observed in the low-risk-score group. To explore this in the validation set, the same methodology was then adopted for the GSE30219 (Figure 4)(b), and the model showed significant differentiation capability (the 5-year survival rate was 69.05% in the low-risk-score group ($N = 42$) vs. 36.59% in the high-score-risk group ($N = 41$), log rank $P = 0.0025$). In Figure 4(c), TCGA data ($N = 400$) with a large sample size was used for survival prediction. LUAD in this cohort was also classified into high- or low-risk-score groups (the 5-year survival incidence was 13.50% in the low-risk-score group ($N = 200$) vs. 9.00% in the high-risk-score group ($N = 200$), log rank $P = 0.0025$). In Figure 4(d), additional selected GSE50081 data were used for the validation of the survival model (the 5-year survival incidence was 50.00% in the low-score-risk group ($N = 64$) vs. 35.94% in the high-score-risk group ($N = 64$), log rank $P = 0.16$). Although not significant, it still has some predictive power. From the results of the four cohorts, it can be obtained that the four prognostic IRGs model had excellent prognostic ability.

*3.4. Correlation Analysis Between Signature and Clinical Features.* In Table 3, the association between the IRGs signature and clinical features in the three datasets was validated by the chi-square test. There was a significant correlation between the pathological stage and the prognostic IRGs signature in the GSE31210 and TCGA cohort ($P < 0.05$). In the GSE31210 cohort, gender and age both had no obvious correlation with the prognostic IRGs signature. In the TCGA dataset, gender and the prognostic IRGs signature had a significant relationship ($P < 0.05$), but the age had not with this model. The GSE30219 cohort has a different clinical characteristic T-stage than the GSE31210 and TCGA cohort, and no statistically significant between the three clinical features and the prognostic IRGs signature in the GSE30219.

In Table 4, the IRGs signature was shown to be statistically significant by multivariate COX regression, affirming it as an independent adverse predictor. In all three data sets, signatures proved to be powerful predictors of clinical outcomes in LUAD patients (high- vs. low-risk, GSE31210, HR = 32.11, 95% CI 4.32–238.91, $P < 0.001$, $n = 226$;

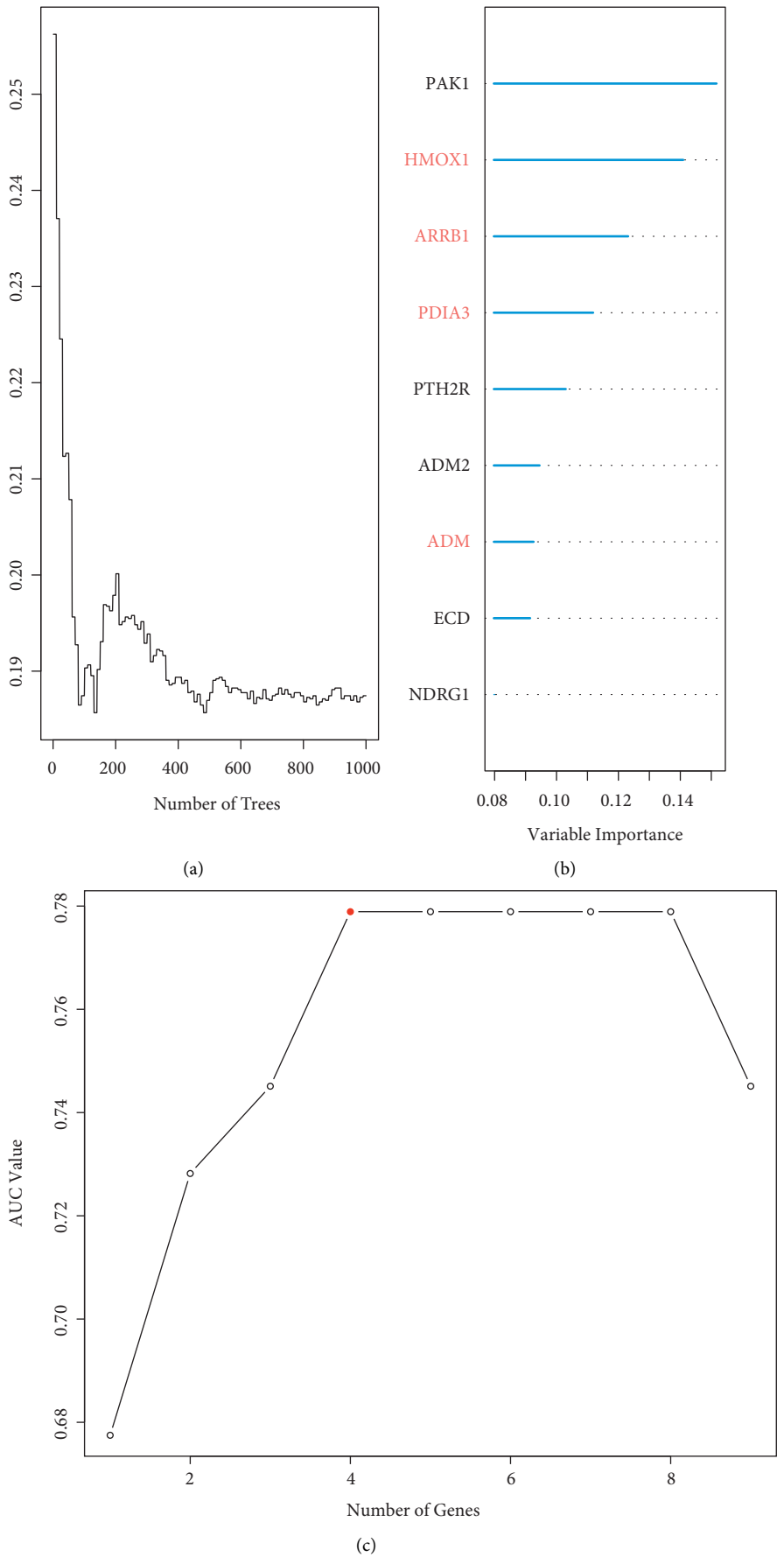



Figure 2: Prognosis prediction. (a, b) Selection of prognosis-related IRGs using RSFA. (c) AUC values of different IRGs prognostic models.

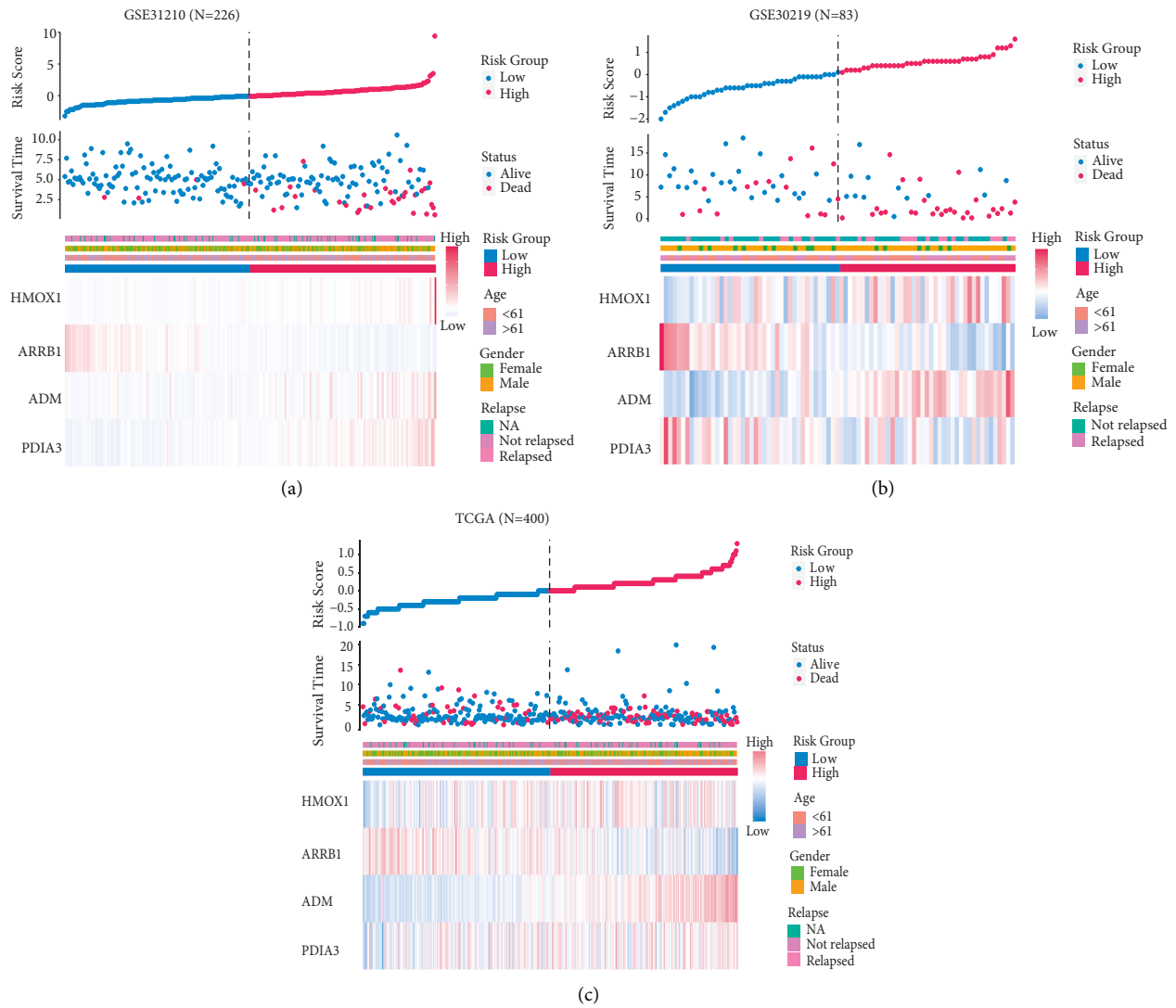



Figure 3: Evaluation of the risk predictive model. (a–c) Clustering Heatmap, the scatter plot of survival time, and the risk score curve of GSE31210, GSE30219, and TCGA sets. Gene expression levels, survival information, and clinical information are given.

Table 2: Prognosis of the four genes in the signature.

| ENSEMBL ID | Symbol ID | Gene name | Coef | *P*-value | Prognostic indicator |
|---|---|---|---|---|---|
| ENSG00000100292 | HMOX1 | Heme Oxygenase 1 | 1.32 | <0.01 | high |
| ENSG00000137486 | ARRB1 | Arrestin Beta 1 | −1.36 | <0.01 | low |
| ENSG00000148926 | ADM | Adrenomedullin | 1.17 | <0.01 | high |
| ENSG00000167004 | PDIA3 | Protein Disulfide Isomerase Family A Member 3 | 1.44 | <0.01 | high |

GSE30219, HR = 2.49, 95% CI 1.29–4.78, $P < 0.001$, $n = 83$; TCGA, HR = 1.73, 95% CI 1.18–2.54, $P < 0.001$, $n = 400$). Univariate Cox also suggests that prognostic IRGs signature is a risk factor. Interestingly, the *P*-values for the pathological stage in GSE31210 and TCGA were less than 0.5. Unfortunately, this clinical feature was not available in GSE30219, so this information could not be further judged by the three datasets. Gender and age cannot be considered as risk factor for prognosis in the three cohorts.

*3.5. Exploring the Functions of the IRG Panel.* To begin with, we obtained 110 IRGs by survival analysis and AUC analysis. Then we analyzed the pathways and functions of IRGs using GO and KEGG. Top 10 functional annotations of biological processes (BPs), cellular components (CCs), and molecular functions (MFs) were selected, among which the main outcomes of BPs were regulation of chemotaxis, leukocyte migration, and cytokine production. The preliminary results of CCs were linked to the membrane. MFs enrichment

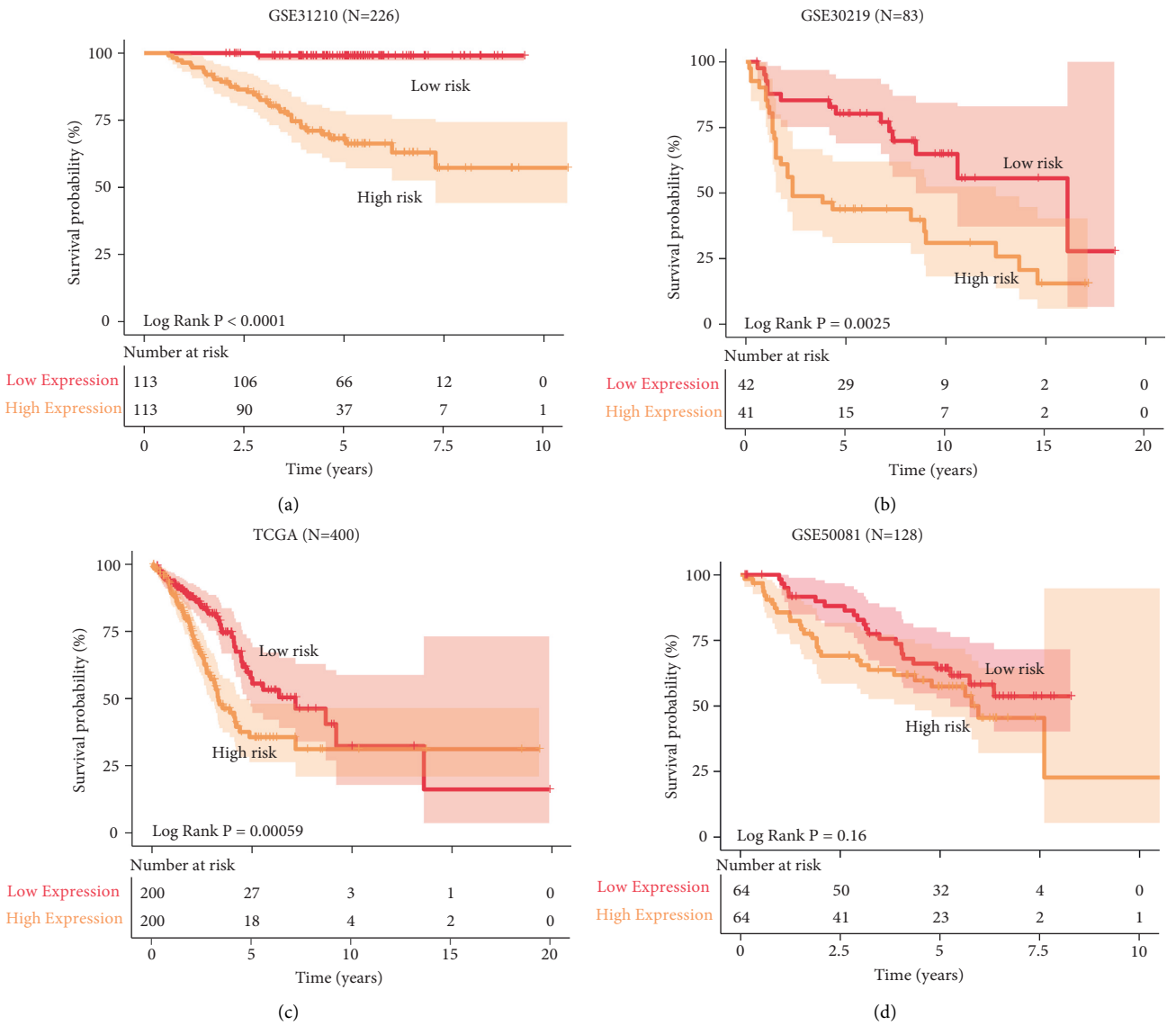



Figure 4: Immune-related gene modeling predicts overall survival in patients with LUAD. (a–d) Kaplan–Meier survival curves in the GSE31210, GSE30219, GSE50081, and TCGA. *P* values were calculated by log-rank test.

analysis can be obtained concerning molecular activity and molecular binding (Figures 5(a)–5(c).

In Figure 5(d), the KEGG-Gene-Concept Network provides a clear view of the distribution of IRGs in different pathways. Most genes are associated with neuroactive ligand-receptor interactions in the KEGG-Gene-Concept Network.

*3.6. IRGs Survival Prediction Nomogram.* In Figure 6(a), to enable comprehensive prediction, predictive models for the four genes were translated into a nomogram to provide a visual projection of OS at 1, 3, and 5 years. For instance, different scores were obtained from the expression of the four genes, and the total scores were summed to calculate the survival rates at different years. As can be seen from the nomogram, the two genes HMOX1 and PDIA3 in the prognostic IRGs model require focused attention in the model for prediction.

To evaluate how well this nomogram simulates the real situation, calibration curves using 1000 bootstrap tests were plotted. As shown in Figure 6(b), the actual case and the predicted situation show a good agreement in the training set. Furthermore, the calibration curve still showed a good deal in the validation set (Figures 6(c) and 6)(d). These results indicate that our nomogram is a good predictor of reality, and both independent validation databases show that the nomogram has excellent utility.

In the Figure 6(e), the time-AUC values in the GSE31210 at years of 1,3 and 5 were 0.755 (95% CI: 0.722–0.787), 0.752 (95% CI: 0.690–0.814), and 0.803 (95% CI: 0.748–0.859), respectively. All three AUC values of GSE31210 were greater than 0.7, which indicates a high degree of confidence in its prediction. Other cohorts (GSE30219 and TCGA) were used

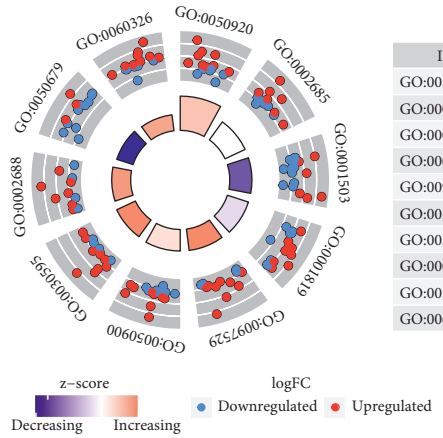
Table 3: The IRG signature and clinical characteristics Chi-square table in LUAD patients.

| Variables | Status | low | high | *P* |
|---|---|---|---|---|
| *GSE31210 dataset (N = 226)* | | | | |
| | | | | 0.69 |
| Age | ≤61 | 63 | 59 | |
| | >61 | 50 | 54 | |
| | | | | 0.11 |
| Gender | Female | 67 | 54 | |
| | Male | 46 | 59 | |
| | | | | <0.01 |
| Pathological stage | I | 103 | 65 | |
| | II | 10 | 48 | |
| *GSE30219 dataset (N = 83)* | | | | |
| | | | | 1.00 |
| Age | ≤61 | 23 | 22 | |
| | >61 | 18 | 18 | |
| | | | | 0.30 |
| Gender | Female | 11 | 6 | |
| | Male | 30 | 34 | |
| | | | | 0.32 |
| T stage | T1 | 37 | 32 | |
| | T2 | 4 | 8 | |
| *TCGA dataset (N = 400)* | | | | |
| | | | | 0.62 |
| Age | ≤61 | 67 | 73 | |
| | >61 | 128 | 123 | |
| | | | | 0.04 |
| Gender | Female | 119 | 98 | |
| | Male | 81 | 102 | |
| | | | | 0.02 |
| Pathological stage | I | 151 | 129 | |
| | II | 49 | 71 | |

Table 4: Cox regression analysis of the IRG signature and clinical information with lung adenocarcinoma (LUAD) survival.

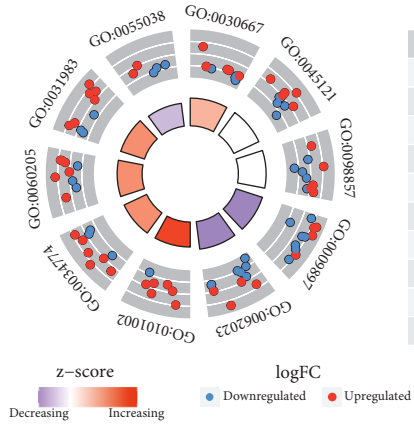
| Variables | | Univariable cox | | | | Multivariable cox | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | HR | 95% CI of HR | | *P* | HR | 95% CI of HR | | *P* |
| | | | right | left | | | right | left | |
| *GSE31210 (N = 226)* | | | | | | | | | |
| Age | >61 vs. ≤61 | 1.4 | 0.73 | 2.78 | 0.29 | 1.38 | 0.7 | 2.71 | 0.35 |
| Sex | Male vs. Female | 1.52 | 0.78 | 2.96 | 0.22 | 1.22 | 0.63 | 2.39 | 0.56 |
| Pathological stage | II vs. I | 4.23 | 2.17 | 8.23 | <0.01 | 2.07 | 1.04 | 4.1 | 0.03 |
| Signature | High risk vs. low risk | 42.31 | 5.79 | 309.3 | <0.01 | 32.11 | 4.32 | 238.91 | <0.01 |
| *GSE30219 (N = 83)* | | | | | | | | | |
| Age | >61 vs. ≤61 | 1.63 | 0.88 | 3.01 | 0.12 | 1.55 | 0.83 | 2.88 | 0.17 |
| Sex | Male vs. Female | 1.37 | 0.61 | 3.09 | 0.44 | 1.11 | 0.49 | 2.56 | 0.8 |
| T stage | T2 vs. T1 | 2.14 | 1.06 | 4.32 | 0.03 | 1.68 | 0.82 | 3.42 | 0.15 |
| Signature | High risk vs. Low risk | 2.65 | 1.39 | 5.06 | <0.01 | 2.49 | 1.29 | 4.78 | <0.01 |
| *TCGA (N = 400)* | | | | | | | | | |
| Age | >61 vs. ≤61 | 1.07 | 0.73 | 1.56 | 0.73 | 1.18 | 0.8 | 1.73 | 0.4 |
| Sex | Male vs. Female | 1.03 | 0.72 | 1.47 | 0.87 | 0.92 | 0.63 | 1.33 | 0.64 |
| Pathological stage | II vs. I | 2.48 | 1.73 | 3.57 | <0.01 | 2.2 | 1.52 | 3.18 | <0.01 |
| Signature | High risk vs. Low risk | 1.88 | 1.3 | 2.7 | <0.01 | 1.73 | 1.18 | 2.54 | <0.01 |

to demonstrate the accuracy of OS prediction.: The time-AUCs of GSE30219 at years of 1,3 and 5 were 0.593 (95% CI: 0.394–0.791), 0.707 (95% CI: 0.605–0.808), and 0.700 (95% CI: 0.595–0.805) (Figure 6)(f). The AUCs of TCGA set at years of 1,3 and 5 were 0.563 (95% CI: 0.475–0.650), 0.643 (95% CI: 0.576–0.710), and 0.604 (95% CI: 0.516–0.692)



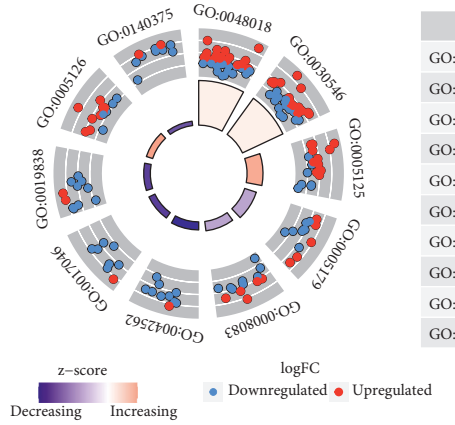

| ID | Description |
|---|---|
| GO:0050920 | Regulation of chemotaxis |
| GO:0002685 | Regulation of leukocyte migration |
| GO:0001503 | Ossification |
| GO:0001819 | Positive regulation of cytokine production |
| GO:0097529 | Myeloid leukocyte migration |
| GO:0050900 | Leukocyte migration |
| GO:0030595 | Leukocyte chemotaxis |
| GO:0002688 | Regulation of leukocyte chemotaxis |
| GO:0050679 | Positive regulation of epithelial cell proliferation |
| GO:0060326 | Cell chemotaxis |

(a)




| ID | Description |
|---|---|
| GO:0030667 | Secretory granule membrane |
| GO:0045121 | Membrane raft |
| GO:0098857 | Membrane microdomain |
| GO:0009897 | External side of plasma membrane |
| GO:0062023 | Collagen−containing extracellular matrix |
| GO:0101002 | Ficolin−1−rich granule |
| GO:0034774 | Secretory granule lumen |
| GO:0060205 | Cytoplasmic vesicle lumen |
| GO:0031983 | Vesicle lumen |
| GO:0055038 | Recycling endosome membrane |

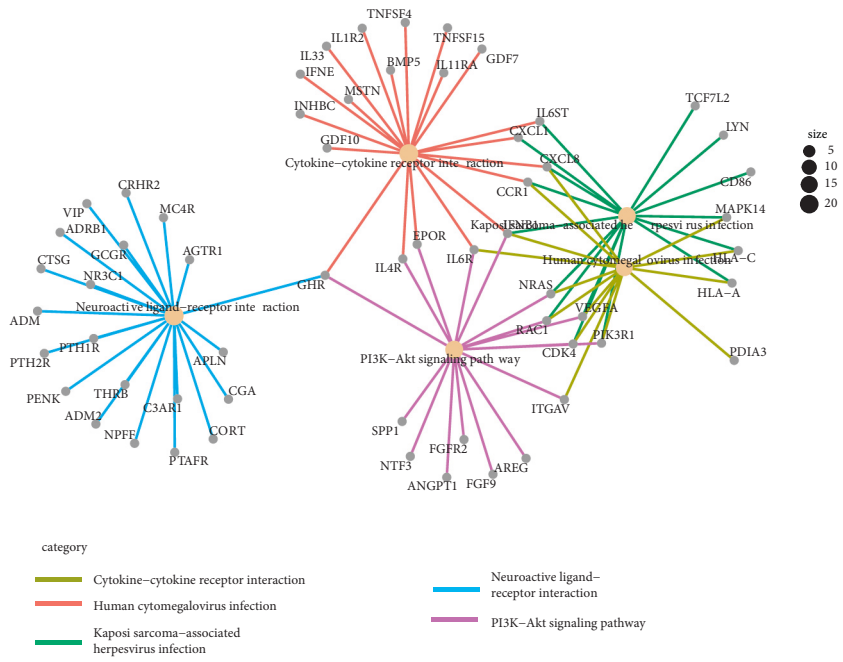
(b)




| ID | Description |
|---|---|
| GO:0048018 | Receptor ligand activity |
| GO:0030546 | Signaling receptor activator activity |
| GO:0005125 | Cytokine activity |
| GO:0005179 | Hormone activity |
| GO:0008083 | Growth factor activity |
| GO:0042562 | Hormone binding |
| GO:0017046 | Peptide hormone binding |
| GO:0019838 | Growth factor binding |
| GO:0005126 | Cytokine receptor binding |
| GO:0140375 | Immune receptor activity |

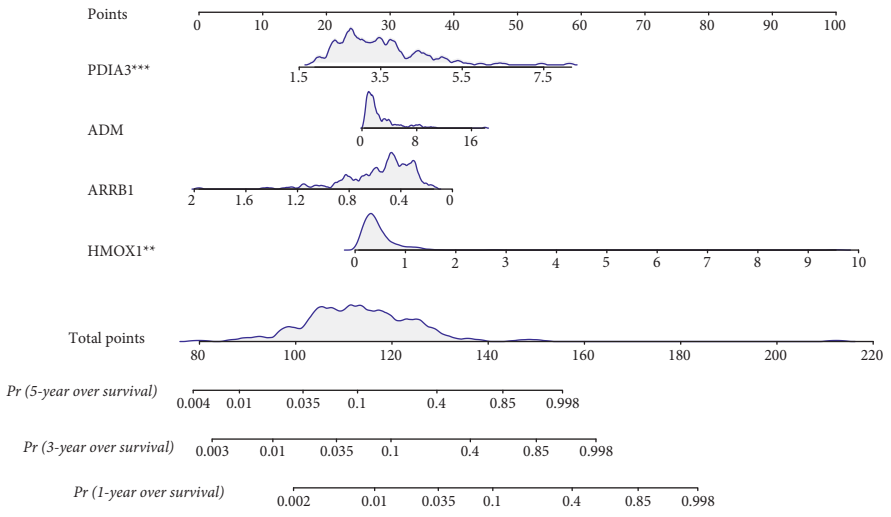
(c)

Figure 5: Continued.

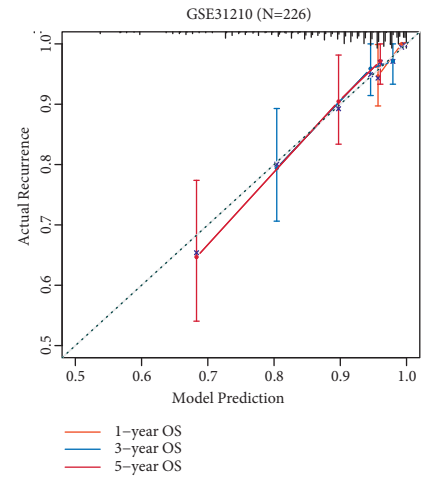



(d)

Figure 5: GO and KEGG analysis for the prognosis-related IRGs. (a–c) Function enrichment of biological process (BP), cellular component (CC), and molecular functions (MFs). (d) KEGG-Gene-Concept Network.




(a)




(b)

Figure 6: Continued.

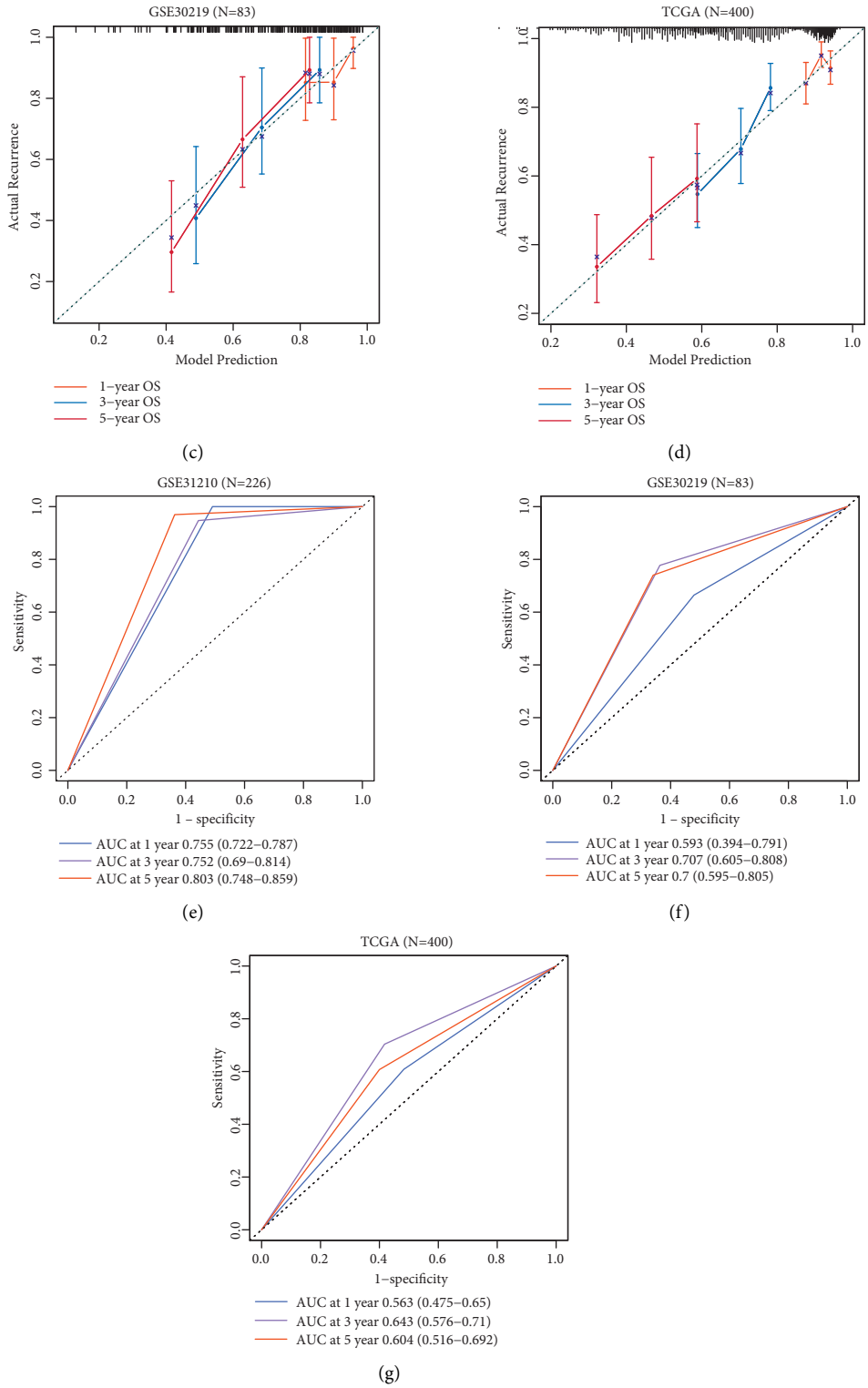



Figure 6: The nomogram including IRGs factor. (a) Nomogram for survival time. (b–d) Calibration curves of GSE31210, GSE30219, and TCGA. (E-G) Time-ROC curves of GSE31210, GSE30219, and TCGA.

(Figure 6)(g). The results from these two datasets show that our prognostic IRGs model has excellent prognostic accuracy.

*3.7. Association of Infiltration of Immune Cells with Prognosis-Related IRGs in Tumors.* The tumor microenvironment consists of stromal cells, tumor cells, and infiltrating

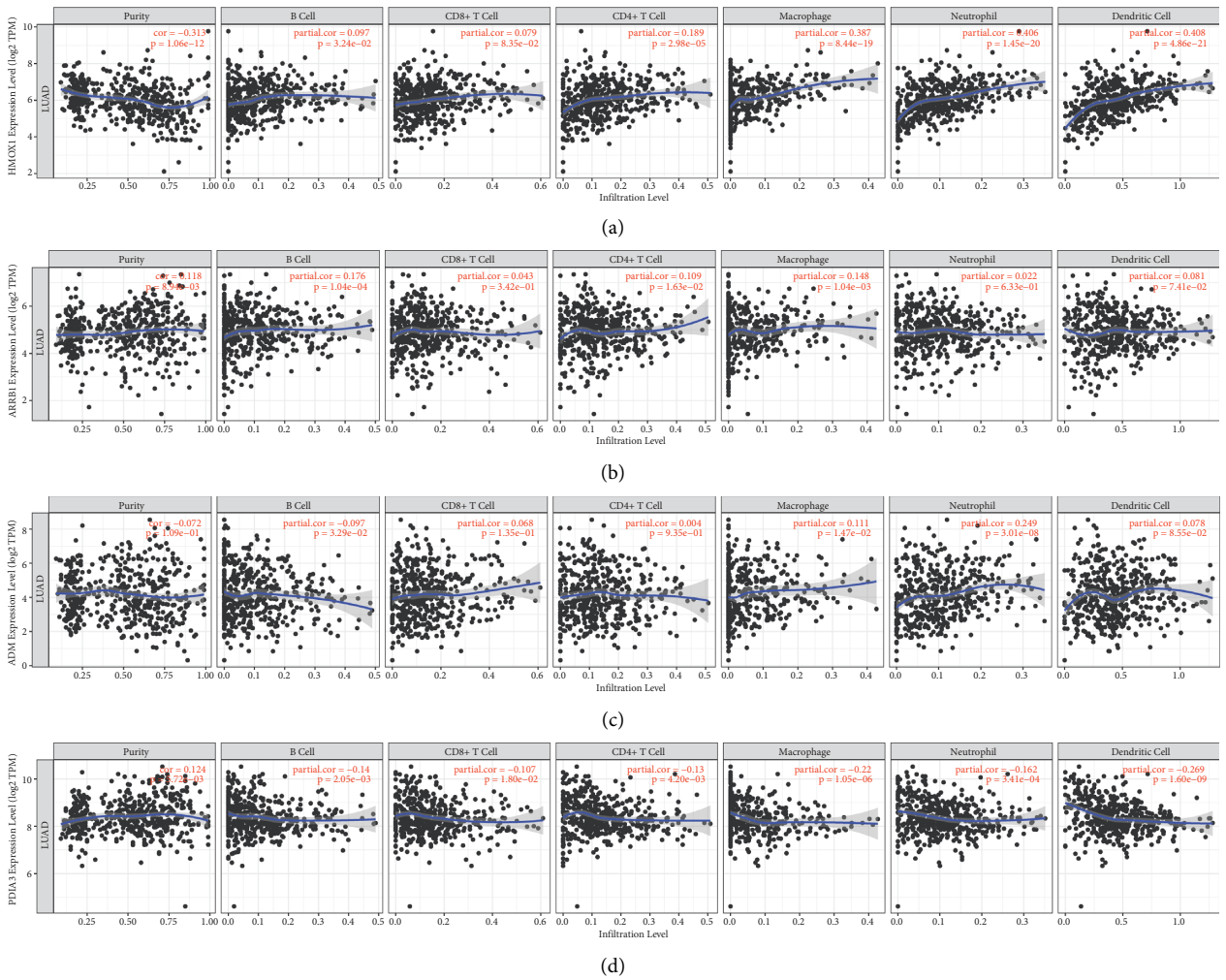



Figure 7: Association of infiltration of immune cells with prognosis-related IRGs in LUAD. (a) HMOX1. (b) ARRB1. (c) ADM. (d) PDIA3. $P < 0.05$ is significant. Each point means a case in the GSE31210 dataset.

immune cells. We utilized TIMER to explore the potential association of infiltration of immune cells with prognosis-related IRGs in tumors. Interestingly, HMOX1, ARRB1, ADM, and PDIA3 were positively associated with B cells and macrophages ($P < 0.05$). These results validate that the four IRGs are closely associated with the level of B cells and macrophages infiltration. Conversely, there is no significant association between these four genes and infiltration of purity, CD8+ T cells, CD4+ T cells, neutrophils, and dendritic cells (Figure 7(a)–7(d)).

*3.8. IRGs Expression and Immunohistochemistry in Online Databases.* In Figure 8(a), PDIA3 was highly expressed, and the results were consistent with the training set and TCGA set. ARRB1 was lowly expressed, and the results were consistent with the three independent data sets, interestingly, in LUAD, statistically significant differences in the expression of HMOX1, ARRB1, and PDIA3. However, ADM were not differentially expressed between LUAD-normal and LUAD-cancer. The differential expression levels of HMOX1, ARRB1, and PDIA3 in the prognostic IRGs model in LUAD also provide positive evidence that the model and LUAD have a strong correlation and can be used effectively for prognosis.

According to the immunohistochemical analyses in the HPA database, representative protein expressions of HMOX1, ARRB1, ADM, and PDIA3 in normal and cancerous tissues were compared (Figures 8(b) and 8)(c). The staining intensity of these two genes (HMOX1, ADM) in tumor cells is higher than in normal pneumocytes. The staining intensity of ARRB1 in tumor cells is lower than in normal pneumocytes, while PDIA3 did not show a striking difference.

## 4. Discussion

When immunotherapy was not well established, early treatment of LUAD was importantly based on surgical lobectomy, with suboptimal outcomes because of the malignant nature of LUAD and the limited results of surgery. However, with the increasing maturity and clinical use of immune checkpoint inhibitors and targeted therapies, the prognosis of patients has improved significantly [19, 20]. In

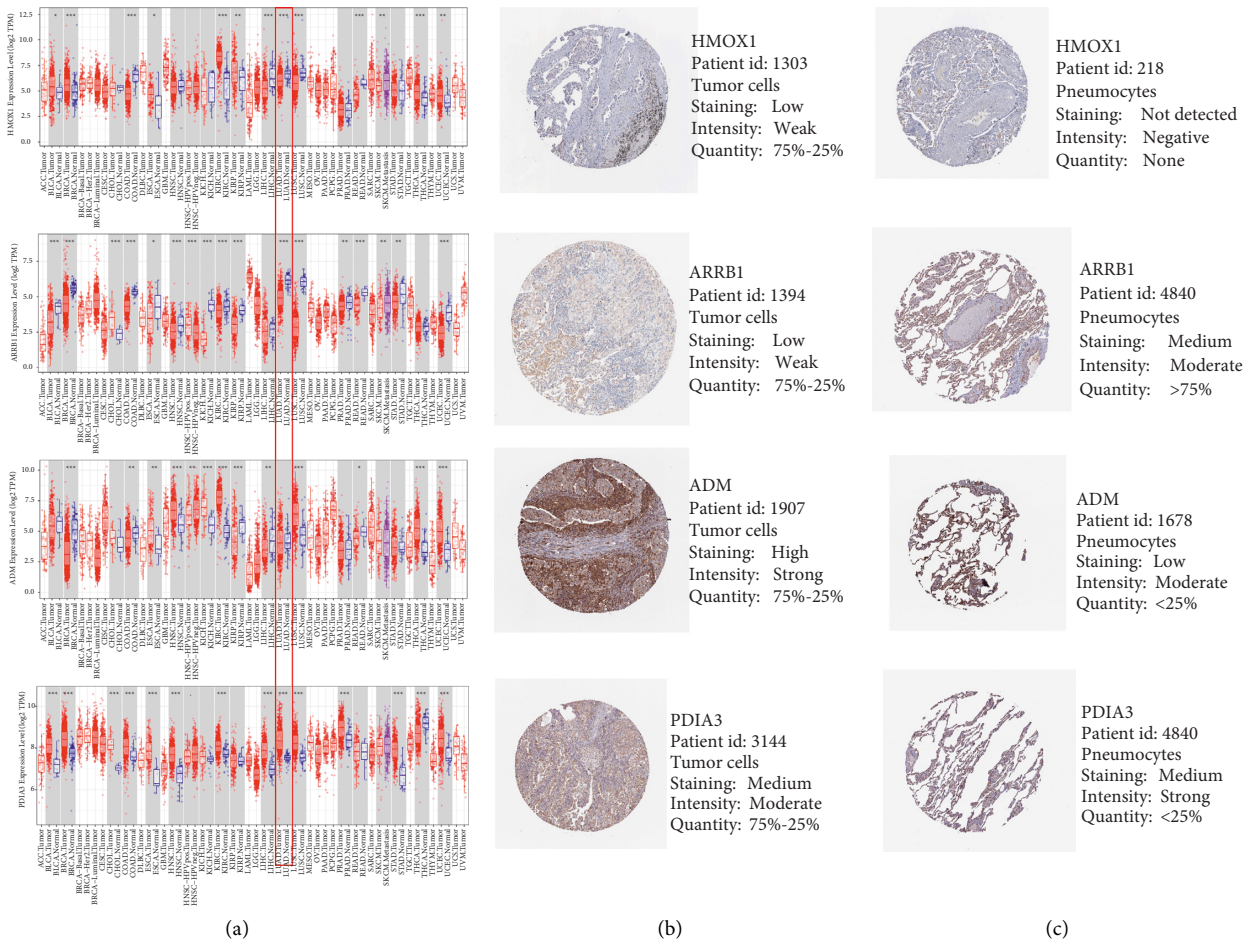



Figure 8: Expression of the four predictive genes. (a) HMOX1, ARRB1, ADM, and PDIA3 expression level in LUAD-tumor vs. LUAD-normal in TIMER (https://cistrome.shinyapps.io/timer/). (b) Expression of four IRGs-encoded proteins in tumor cells and normal pneumocytes in the HPA database.

this study, we choose three separate datasets, GSE31210, GSE30219, and TCGA, for the establishment and validation of an IRGs prognostic model. First, we carried out a dimensionality reduction on the obtained IRGs and subsequently avoided overfitting the model. Finally, we obtained 511 combinations by recombination and performed screening to obtain four gene prediction signatures.

Several recent studies have used IRG characteristics to predict the prognosis of LUAD patients. However, not all models showed an excellent prognosis. Wu et al. developed a 21-IRGs prognostic model. In this study, multivariate regression was used to screen for genes of relevance, but there is no practical method to prevent overfitting. Unfortunately, the AUC was only 0.61, 0.65, and 0.62 for 1, 3, and 5-year survival in the training set [21]. While the present study avoided overfitting and predicted 1, 3, and 5-year survival by four immune-related genes models, the AUC obtained was 0.755, 0.752, and 0.803 in the training set.

The included IRGs for signature were HMOX1, ARRB1, ADM, and PDIA3. HMOX1 (Heme Oxygenase 1) is a protein-coding gene whose related functions include protein homodimerization activity and oxidoreductase activity [22]. Several published pieces of evidence support the overexpression of HMOX-1 in several human malignant tumors, including kidney, gastrointestinal, lung, and breast cancers [23]. Tsai et al. research discovered that HMOX1 is a negative prognostic NSCLC gene. Its high expression may increase the metastasis ability of cancer cells of NSCLC patients, and HMOX1 has the potential as a therapeutic target for NSCLC in the future [24].

ARRB1 (Arrestin Beta 1) plays an important role in transcription factor binding and ubiquitin protein ligase binding functions, and it is also a protein-coding gene [25]. Li et al. suggest that ARRB1 may be associated with the prognosis of LUAD, and it is considered as a new molecular biomarker for the diagnosis and prognosis of LUAD [26].

ADM (Adrenomedullin) gene encodes a protein that is a prohormone with several functions, including vasodilation, hormone secretion regulation, angiogenesis promotion, and antimicrobial activity [27]. Some studies suggest that NSCLC cells are essential targets for ADM, and it may regulate the activity of these malignant lung cells through differential induction of different early response genes [28].

PDIA3 (Protein disulfide isomerase A3) gene encodes a protein that is a member of the PDI family. PDIA3 has high expression levels in response to cellular stress and blocks

apoptotic cell death associated with endoplasmic reticulum (ER) stress and protein misfolding. PDIA3 is highly expressed in most cancers, and its expression is associated with overall low cell survival, metastasis, and invasiveness [29]. Wang et al. demonstrated the potential prognostic value of PDIA3 through a proteomic biomarker discovery approach [30].

Using nomogram to show the prediction of survival by IRGs signatures, it visually exhibits excellent prognostic capability. The calibration curves also show good agreement between the predictions value and the actual situation in training. We also present the cellular functions and pathway enrichment of the selected IRGs based on GO analysis and KEGG analysis. According to GO analysis, regulation of chemotaxis, leukocyte migration, and cytokine production were significantly enriched functions and most genes associated with neuroactive ligand-receptor interactions based on KEGG-Gene-Concept Network.

By exploring the potential association of infiltration of immune cells with prognosis-related IRGs in tumors, we found that the four IRGs were closely associated with the level of B-cell and macrophage infiltration. HMOX1, ARRB1, and PDIA3 were significantly differentially expressed in LUAD-normal and LUAD-cancer. The proteins encoded by the three genes HMOX1, ADM, and ARRB1 also differ significantly in tumor cells and normal pneumocytes. These results suggest that the four IRGs have prognostic potential.

This study also has some restrictions. Firstly, the cohorts of GEO and TCGA databases we used for the construction of immune-related gene prognostic models are retrospective experiments and lack large clinical samples to prospectively validate the prognostic value of LUAD patients. Additionally, due to the lack of experimental exploration in these immune genes' potential functions and mechanisms, the potential functions and mechanisms of these immune genes need to be further validated in clinical trials.

## 5. Conclusion

In summary, our study systematically analyzed the expression, prognostic value, and potential functions of IRGs. Ultimately, a four-immune-related genes signature was constructed that can be independently used as a biomarker for OS prediction in LUAD. Eventually, our study provides an essential theoretical basis for further research on the role of IRGs in LUAD.

## Data Availability

https://www.ncbi.nlm.nih.gov/geo/, https://portal.gdc.cancer.gov/

## Conflicts of Interest

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

## Authors' Contributions

Wenhao Liu, Ruihong Dong, and Shuai Gao have contributed equally to this work and share first authorship.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (31970636).

## Supplementary Materials

After univariate cox analysis and ROC curve analysis, 110 prognosis-related genes were obtained by screening based on $P < 0.01$ and AUC > 0.6, and the results are shown in Supplementary Table S1. Supplementary Table S1 |m6A gene sets of Univariate Cox regression analysis in the GSE31210 set ($P < 0.01$, AUC > 0.6, $n = 226$). (*Supplementary Materials*)

## References

[1] M. Behera, T. K. Owonikoko, A. A. Gal et al., "Lung adenocarcinoma staging using the 2011 iaslc/ats/ers classification: a pooled analysis of adenocarcinoma in situ and minimally invasive adenocarcinoma," *Clinical Lung Cancer*, vol. 17, no. 5, pp. e57–e64, 2016.

[2] The Cancer Genome Atlas Research Network, "Comprehensive molecular profiling of lung adenocarcinoma," *Nature*, vol. 511, no. 7511, pp. 543–550, 2014.

[3] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2021," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 1, pp. 7–33, 2021.

[4] R. D. Neal, F. Sun, J. D. Emery, and M. E. Callister, "Lung cancer," *BMJ*, vol. 365, p. l1725, 2019.

[5] A. Abdelatty, Q. Sun, J. Hu et al., "Pan-cancer study on protein kinase C family as a potential biomarker for the tumors immune landscape and the response to immunotherapy," *Frontiers in Cell and Developmental Biology*, vol. 9, p. 798319, 2021.

[6] H. Angell and J. Galon, "From the immune contexture to the immunoscore: the role of prognostic and predictive immune markers in cancer," *Current Opinion in Immunology*, vol. 25, no. 2, pp. 261–267, 2013.

[7] A. J. Gentles, A. M. Newman, C. L. Liu et al., "The prognostic landscape of genes and infiltrating immune cells across human cancers," *Nature Medicine*, vol. 21, no. 8, pp. 938–945, 2015.

[8] G. T. Gibney, L. M. Weiner, and M. B. Atkins, "Predictive biomarkers for checkpoint inhibitor-based immunotherapy," *The Lancet Oncology*, vol. 17, no. 12, pp. e542–e551, 2016.

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

[10] M. S. Rooney, S. A. Shukla, C. J. Wu, G. Getz, and N. Hacohen, "Molecular and genetic properties of tumors associated with local immune cytolytic activity," *Cell*, vol. 160, no. 1-2, pp. 48–61, 2015.

[11] C. Luo, M. Lei, Y. Zhang et al., "Systematic construction and validation of an immune prognostic model for lung adenocarcinoma," *Journal of Cellular and Molecular Medicine*, vol. 24, no. 2, pp. 1233–1244, 2020.

[12] Q. Song, J. Shang, Z. Yang et al., "Identification of an immune signature predicting prognosis risk of patients in lung adenocarcinoma," *Journal of Translational Medicine*, vol. 17, no. 1, p. 70, 2019.

[13] M. Zhang, K. Zhu, H. Pu et al., "An immune-related signature predicts survival in patients with lung adenocarcinoma," *Frontiers Oncology*, vol. 9, p. 1314, 2019.

[14] Q. F. Chen, L. Dai, Y. Wu, Z. Huang, M. Chen, and M. Zhao, "Surveillance strategy for barcelona clinic liver cancer B hepatocellular carcinoma achieving complete response: an individualized risk-based machine learning study," *Frontiers in Bioengineering and Biotechnology*, vol. 9, Article ID 667641, 2021.

[15] T. Li, J. Fan, B. Wang et al., "Timer: a web server for comprehensive analysis of tumor-infiltrating immune cells," *Cancer Research*, vol. 77, no. 21, pp. e108–e110, 2017.

[16] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, "Clusterprofiler: an R package for comparing biological themes among gene clusters," *OMICS: A Journal of Integrative Biology*, vol. 16, no. 5, pp. 284–287, 2012.

[17] N. Sun, J. Chu, W. Hu, X. Chen, N. Yi, and Y. Shen, "A novel 14-gene signature for overall survival in lung adenocarcinoma based on the bayesian hierarchical cox proportional hazards model," *Scientific Reports*, vol. 12, no. 1, p. 27, 2022.

[18] L. Wang, M. Shang, Q. Dai, and P. A. He, "Prediction of lncrna-disease association based on a laplace normalized random walk with restart algorithm on heterogeneous networks," *BMC Bioinformatics*, vol. 23, no. 1, p. 5, 2022.

[19] Z. Liu, D. Sun, Q. Zhu, and X. Liu, "The screening of immune-related biomarkers for prognosis of lung adenocarcinoma," *Bioengineered*, vol. 12, no. 1, pp. 1273–1285, 2021.

[20] Y. Zheng, H. Tian, Z. Zhou et al., "A novel immune-related prognostic model for response to immunotherapy and survival in patients with lung adenocarcinoma," *Frontiers in Cell and Developmental Biology*, vol. 9, Article ID 651406, 2021.

[21] P. Wu, Y. Zheng, Y. Wang, Y. Wang, and N. Liang, "Development and validation of a robust immune-related prognostic signature in early-stage lung adenocarcinoma," *Journal of Translational Medicine*, vol. 18, no. 1, p. 380, 2020.

[22] Z. Meng, H. Liang, J. Zhao et al., "Hmox1 upregulation promotes ferroptosis in diabetic atherosclerosis," *Life Sciences*, vol. 284, Article ID 119935, 2021.

[23] L. Y. Chau, "Heme oxygenase-1: emerging target of cancer therapy," *Journal of Biomedical Science*, vol. 22, no. 1, p. 22, 2015.

[24] J. R. Tsai, H. M. Wang, P. L. Liu et al., "High expression of heme oxygenase-1 is associated with tumor invasiveness and poor clinical outcome in non-small cell lung cancer patients," *Cellular Oncology*, vol. 35, no. 6, pp. 461–471, 2012.

[25] Y. Lei, X. Xu, H. Liu et al., "Hbx induces hepatocellular carcinogenesis through arrb1-mediated autophagy to drive the G1/S cycle," *Autophagy*, vol. 17, no. 12, pp. 4423–4441, 2021.

[26] X. Li, K. Che, L. Wang et al., "Subcellular localization of beta-arrestin1 and its prognostic value in lung adenocarcinoma," *Medicine (Baltimore)*, vol. 96, no. 45, Article ID e8450, 2017.

[27] A. A. Voors, D. Kremer, C. Geven et al., "Adrenomedullin in heart failure: pathophysiology and therapeutic application," *European Journal of Heart Failure*, vol. 21, no. 2, pp. 163–171, 2019.

[28] L. You and S. B. Jakowlew, "Identification of early growth response gene-1 (Egr-1) as a phorbol myristate acetate-induced gene in lung cancer cells by differential mRNA display," *American Journal of Respiratory Cell and Molecular Biology*, vol. 17, no. 5, pp. 617–624, 1997.

[29] F. Mahmood, R. Xu, M. U. N. Awan et al., "Pdia3: structure, functions and its potential role in viral infections," *Biomedicine & Pharmacotherapy*, vol. 143, Article ID 112110, 2021.

[30] K. Wang, H. Li, R. Chen et al., "Combination of CALR and PDIA3 is a potential prognostic biomarker for non-small cell lung cancer," *Oncotarget*, vol. 8, no. 57, pp. 96945–96957, 2017.

Hindawi
Journal of Oncology
Volume 2022, Article ID 7117014, 21 pages
https://doi.org/10.1155/2022/7117014

*Research Article*

# Identification of Prognosis-Related Molecular Subgroups and Construction of a Prognostic Prediction Model Using Immune-Related Genes in Pancreatic Cancer

**Xiang Fei,[1] Lingming Kong,[2] Chao Shi,[3] Gang Wang,[2] Chenhai Liu,[2] Cheng Wang,[2] Peng Liu,[4] and Xiaodong Tan[4]**

[1]*The Third Department of General Surgery, People's Hospital of China Medical University (Liaoning Provincial People's Hospital), Shenyang, China*
[2]*Department of General Surgery, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China*
[3]*Department of Day Surgery Ward, The First Affiliated Hospital of Harbin Medical University, Harbin, China*
[4]*Department of General Surgery, Shengjing Hospital of China Medical University, Shenyang, China*

Correspondence should be addressed to Lingming Kong; konglingmingklm@163.com and Xiaodong Tan; tanxdcmu@163.com

Received 9 November 2021; Revised 30 April 2022; Accepted 23 May 2022; Published 7 June 2022

Academic Editor: Dechao Bu



*Background*. Pancreatic cancer patients with similar clinicopathological status exhibit substantially different therapeutic responses, which might be caused by the vast molecular heterogeneity of tumors. In this study, we attempted to identify specific molecular subgroups and construct a prognostic prediction model based on the expression level of immune-related genes in pancreatic cancer. The transcriptome profiling, single nucleotide variation, copy number variation, clinicopathological information, and follow-up data of pancreatic cancer patients were obtained from The Cancer Genome Atlas database. Thereafter, the immune-related genes with prognostic significance were identified for further consensus cluster analysis. The molecular characteristics and clinicopathological information were compared between the identified subgroups, and a weighted correlation network analysis was performed to identify the hub genes associated with the subgroups. Finally, the prognostic prediction model based on immune-related genes was established using the least absolute shrinkage and selection operator (LASSO) analysis. *Results*. A total of 67 immune-relevant genes with prognostic significance were selected and used for the consensus cluster analysis. The total samples were divided into two groups, C1 and C2. The subgroup C1 had a significantly worse prognosis than C2, as well as lower levels of immune cell infiltration, which indicate an immunosuppressed state. The mutational rate of the cancer-related genes including *KRAS*, *TP*53, and *RNF*43 was higher in the C1 subgroup. The C1 subgroup was associated with more advanced tumor grade and T stage and with higher mortality. Using LASSO regression, we developed a prognostic prediction model based on the expression levels of 19 immune-related genes, which we validated in three external data sets. In addition, we identified four potential therapeutic and prognostic biomarkers (*TNNT1*, *KCNN4*, *SH2D3A*, and *PHLDA2*). *Conclusion*. We identified two novel molecular subgroups of pancreatic cancer and developed a prognostic prediction model based on the expression levels of immune-related genes, which could be used in a clinical setting and could aid in unraveling the molecular processes leading to the development of pancreatic cancer.

## 1. Introduction

Pancreatic cancer is one of the most lethal malignancies affecting the digestive system. It is characterized by an insidious onset with nonspecific early symptoms and early metastasis [1, 2]. As a result, pancreatic cancer patients often are at an advanced stage or present distant metastases when they are diagnosed. This greatly hinders surgical management, which is the most effective therapeutic approach for pancreatic cancer nowadays [3]. In addition, pancreatic

cancer is associated with a high risk of relapse. Altogether, these factors contribute to the high mortality associated with this disease, which is the fourth leading cause of cancer-related deaths in America, and that has a five-year survival rate of less than 9% [4]. The traditional tumor staging methods that merely rely on basic clinicopathological information such as the tumor-node-metastasis classification or the age group in the same category individuals with substantial phenotypic differences have a considerable impact on prognosis. Therefore, traditional staging methods are not sensitive enough to provide individualized diagnosis and treatment for pancreatic cancer patients [5, 6]. In addition, classical serum tumor biomarkers such as carbohydrate antigen 19-9, carbohydrate antigen 125, or carcinoembryonic antigen perform poorly in the early diagnosis and outcome prediction of pancreatic cancer [7, 8]. Improving early diagnosis and developing new risk stratification methods will contribute to more effective management and a better prognosis for pancreatic cancer patients. With the rapid development and progress of sequencing methods, huge amounts of high-throughput data from different omics technologies are now available in public databases such as The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). Wise use of this information may help us further investigate the molecular mechanisms underlying the development of pancreatic cancer.

Previous studies have already investigated the molecular classification and the internal heterogeneity of pancreatic cancer using omics data [9, 10]. Integrating genomics data, Bailey et al. established four distinct molecular subtypes of pancreatic cancer (squamous, pancreatic progenitor, immunogenic, and aberrantly differentiated endocrine-exocrine), a new classification system that might contribute to understanding the molecular evolution of pancreatic cancer, as well as to developing new therapeutic methods [11]. Using transcriptional data, Collisson et al. distinguished three pancreatic ductal adenocarcinoma subtypes (classical, quasimesenchymal, and exocrine-like) that are associated with different survival and progress rates, as well as distinct therapeutic responses [12]. Finally, a classifier constructed using the expression levels of 19 miRNAs was reported to accurately predict the prognosis of pancreatic cancer patients with high sensitivity [13]. These studies highlight the potential ways in which the use of omics data can help molecularly characterize tumors, contributing to the discovery of new diagnostic and prognostic biomarkers.

The malignant behaviors of pancreatic cancer largely depend on the complex cross-talk between tumor cells and the tumor immune microenvironment [14]. A deeper understanding of the immune landscape in pancreatic cancer could help elucidate these interactions and develop new immunotherapy approaches [15]. A previous study identified four molecular subtypes of prostate cancer that markedly differed in their prognosis by analyzing the expression levels of immune-related genes. Six of these immune-related genes were used to develop a prognostic prediction model using least absolute shrinkage and selection operator (LASSO) regression [16]. To our knowledge, no molecular subgroups of pancreatic cancer have been identified using immune-related genes to date.

In this study, we attempted to identify specific molecular subtypes of pancreatic cancer that are closely associated with the immune system signature. Analyzing the expression levels of immune-related genes with prognostic value using consensus cluster analysis, we identified two specific molecular subgroups, C1 and C2. Significant differences in the overall survival (OS) and clinical features including tumor grade, T stage, and survival status were observed between the C1 and C2 subgroups. To molecularly characterize these two subgroups, we annotated the differentially expressed genes (DEGs) according to gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) and performed gene set enrichment analysis (GSEA). In addition, we also performed immune infiltration, mutation spectrum, and copy number variation analyses. Weighted correlation network analysis (WGCNA) identified four hub genes associated with the molecular subgroups. In the future, it is necessary to further study the function of the four genes identified, *TNNT1*, *KCNN4*, *SH2D3A*, and *PHLDA2*, and evaluate their potential use as biomarkers. Finally, we constructed a prognostic prediction model using the expression levels of 19 immune-related genes and validated it using three external data sets. In addition, uni- and multivariate Cox analyses revealed that the constructed model was an independent prognostic factor in pancreatic cancer. These results indicate that the model could be used in a clinical setting in the near future to aid clinicians in making management-related decisions.

## 2. Materials and Methods

*2.1. Data Download and Preprocess.* Transcriptome data (RNA-Seq, HTSeq-Counts type), single nucleotide variation data (MuTect2, Annotation type), copy number variation (Copy Number Segment, Masked type), clinicopathological information, and the latest follow-up data of pancreatic cancer patients from TCGA database were downloaded from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov) on December 8, 2020 [17]. Genes with RNA-sequencing missing values in more than half of the total samples and data from patients with overall survival (OS) of less than 30 days were excluded from the subsequent analyses. A total of 171 pancreatic cancer samples met the selection criteria and were used for consensus cluster analysis. In addition, three GEO data sets, GSE62452 [18], GSE71729 [19], and GSE78229 [20], were used for external validation [21, 22]. The detailed information on pancreatic cancer samples from TCGA and the three validation data sets are provided in Supplementary Tables 1 and 2, respectively.

*2.2. Identification of Prognosis-Related Molecular Subgroups Based on the Immune-Related Genes.* To study the existence of distinct molecular subgroups in pancreatic cancer, we selected immune-related genes with prognostic significance for further analyses. First, 1,811 immune-relevant genes were obtained from the Immunology Database and Analysis

Portal (ImmPort) website (https://www.immport.org/) [23]. Thereafter, log-rank and Cox survival analyses were performed to identify survival-relevant genes based on the gene expression data of pancreatic cancer samples from the TCGA data set. A total of 1,350 genes had associated $p$-values < 0.01 in both the log-rank and Cox analyses and were selected for further study. Immune-related genes with prognostic significance were defined as the intersection of potential prognosis-related markers and immune-related genes, and a total of 67 genes were selected. Detailed information about the selected genes is provided in Supplementary Table 3. Subsequently, consensus cluster analysis based on the above 67 genes was performed using the R package "ConsensusClusterPlus" [24]. The number of consensus clusters was set to 2, based on the package guidelines. Finally, the pancreatic cancer samples were separated into groups C1 ($n = 103$) and C2 ($n = 68$). OS and principal component analysis (PCA) between the two subgroups were performed with the R packages "survival," "survminer," and "DESeq2" [25–28].

*2.3. Molecular Characteristics and Clinicopathological Information Comparisons between the Two Subgroups.* To further elucidate the mechanisms underlying the two subgroups, the DEGs between the C1 and C2 subgroups were identified and selected using the criteria |Log2 Fold Change| > 1 and adjusted $p$-value < 0.05. A total of 2,698 DEGs were identified using the R package "DESeq2" [28]. Furthermore, GO, KEGG, and GSEA were performed using the R packages "clusterProfiler" and "enrichplot" to identify the related molecular signaling pathways of the specific subgroups [29, 30]. Detailed information on the DEGs is provided in Supplementary Table 4. In addition, the TIMER2.0 database was used to estimate and compare the immune infiltrate levels between the two subgroups [31–33]. Next, mutation spectrum analysis of the two subgroups was performed using the R package "maftools" [34]. The function "mafComapre" of the R package "maftools" was used to perform Fisher's test on all genes to detect differentially mutated genes between C1 and C2 subgroups. Correlation analysis between gene expression levels and copy numbers was analyzed. Genes with different frequencies of copy number variations were compared between C1 and C2 subgroups. Finally, the clinicopathological features were compared between the C1 and C2 subgroups.

*2.4. Identification of Genes Associated with the Molecular Subgroups Using WGCNA.* We performed WGCNA using the R package "WGCNA" to identify the genes associated with each molecular subgroup [35, 36]. The hub genes were selected and visualized using the software Cytoscape (version 3.8.2) [37, 38] and the website Metascape [39]. The web-based tool gene expression profiling interaction analysis (GEPIA) was used to visualize the expression levels of the hub genes in 179 pancreatic tumor samples and 171 normal tissue samples [40]. Overall survival (OS) and relapse-free survival (RFS) analyses of the selected hub genes in pancreatic cancer patients were performed based on the Kaplan-Meier Plotter website [41].

*2.5. Construction of a Prognostic Prediction Model Based on 19 Immune-Related Genes.* To aid clinicians in stratifying pancreatic cancer patients based on the risk level, we developed a prognostic predictor model based on the expression data of 67 immune-relevant genes via LASSO analysis using the R package "glmnet" according to the official recommendations [42, 43]. First, the total samples from the TCGA data set were randomly divided into the training (70% of the total sample, $n = 120$) and validation sets (30% of the total sample, $n = 51$). We developed a prognostic prediction model based on 19 immune-related genes using the training set. Detailed information on the 19 immune-related genes is provided in Table 1. A risk score was assigned to each patient in the training set, according to the following formula:
Risk score = $\sum_{n=1}^{19}$ (Coefficient$_n$ ∗ Expression of gene$_n$).
Samples were categorized as high or low risk, considering the threshold of the median value of the overall risk scores. OS analysis was performed between the high- and low-risk groups. The predictive efficiency of the model was assessed by determining the receiver operating characteristic (ROC) curve and the area under the curve (AUC) using the R package "timeROC" [44]. The validation set and three external data sets (GSE62452, GSE71729, and GSE78229) were used to assess the validity of the model. The expression data of all data sets used were standardized using z-scores.

*2.6. Uni- and Multivariate Cox Analyses of the Prognostic Factors.* We performed univariate analysis using the constructed model and common clinicopathological factors such as age, gender, tumor grade, clinical stage, T stage, M stage, and N stage, to identify the prognosis-related factors ($p < 0.05$). Next, these significantly prognostic factors were extracted for further multivariate Cox analysis to identify the independent prognostic factors ($p < 0.05$) in pancreatic cancer.

*2.7. Statistical Analysis.* In this study, the R (version 3.6.3) and RStudio software were utilized to carry out the statistical analysis and figure preparation. $p$-values less than 0.05 were defined as statistically significant.

## 3. Results

*3.1. Identification of Specific Molecular Subgroups Based on Immune-Related Genes Using Consensus Cluster Analysis.* To identify the molecular subgroups associated with immune-relevant genes in pancreatic cancer, we extracted 1,811 immune-relevant genes from the ImmPort website, which provides user-friendly bioinformatic analysis tools for basic and clinical immunology. Thereafter, log-rank and Cox survival analyses were performed to identify survival-relevant genes based on the gene expression data of pancreatic cancer samples from the TCGA data set. The genes with

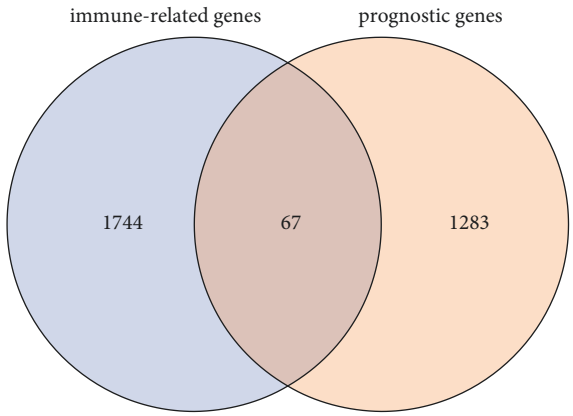
Table 1: The 19 immune-related genes and their corresponding coefficients that were selected to construct the prognostic prediction model based on the LASSO regression model.

| Number | Gene name | LASSO coefficient |
|---|---|---|
| 1 | MET | 0.0951 |
| 2 | SLURP1 | 0.8368 |
| 3 | IL1RAP | 0.0332 |
| 4 | IL18 | 0.3994 |
| 5 | IL31RA | 0.6839 |
| 6 | TNFSF10 | 0.1371 |
| 7 | CXCL11 | 0.4979 |
| 8 | IL20RB | 0.2564 |
| 9 | PLAU | 0.1469 |
| 10 | ERAP2 | 0.2966 |
| 11 | PIK3CB | −0.9853 |
| 12 | EIF2AK2 | −1.1102 |
| 13 | IL6R | −0.2665 |
| 14 | NPR1 | −0.2134 |
| 15 | GDF11 | −1.2324 |
| 16 | INSL5 | −0.0406 |
| 17 | LRSAM1 | −0.6822 |
| 18 | TYK2 | −0.0659 |
| 19 | SEMA6C | −0.3146 |

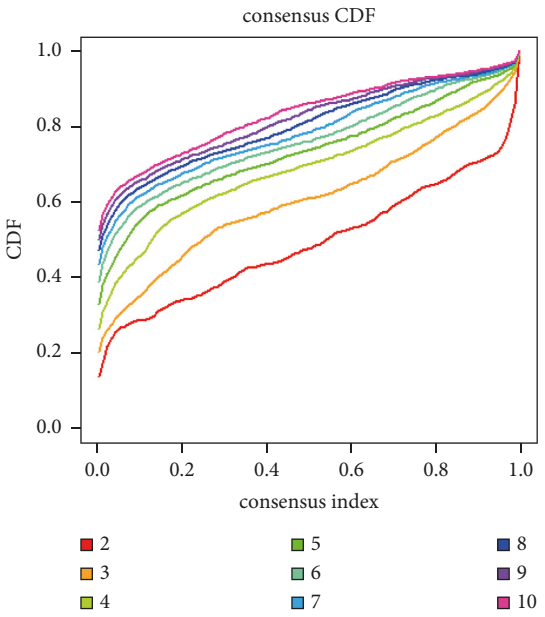
$p < 0.01$ in both log-rank and Cox analyses were selected and defined as genes with a prognostic value. Immune-related genes with prognostic significance were defined as the intersection of potential prognosis-related markers and immune-related genes, and a total of 67 genes were selected (Figure 1(a)). Expression data of the immune-related genes with prognostic significance were selected for further consensus cluster analysis. Samples of patients with survival time inferior to 30 days were excluded for further analyses to avoid possible disturbances. The consensus cumulative distribution function (CDF) plot showed that the slope of the CDF curve changes the most when the consensus index is set at 0 or 1 and the number of clusters is set as 2, as per official recommendations (Figure 1(b)). The relative change in the area under the CDF curve also indicated no substantial changes occur when the $k$ value is higher than 2 (Figure 1(c)). A tracking plot was constructed to reflect the distribution of the samples when $k$ ranges from 2 to 10 (Figure 1(d)). According to the cluster-consensus and consensus matrix plots, the total sample is divided into two distinct subgroups when $k = 2$ (Figures 1(e) and 1(f)). The pancreatic cancer samples were divided into groups C1 ($n = 103$) and C2 ($n = 68$). The expression level of the 67 immune-related genes among C1 and C2 subgroups was represented in a heat map (Figure 2(a)). The subgroup C1 has a significantly worse prognosis in comparison to the subgroup C2 (Figure 2(b); $p < 0.0001$). The existence of two distinct subgroups was further confirmed via PCA (Figure 2(c)). These results indicate the existence of two distinct molecular subgroups in pancreatic cancer according to the expression levels of prognosis-associated immune genes.

*3.2. Identification of Signaling Pathways Related to the Molecular Subgroups.* Since the molecular subgroups of pancreatic cancer samples were significantly correlated to the OS, the DEGs between the two subgroups might be related with the initiation and development of pancreatic cancer. Therefore, we first determined the DEGs between the C1 and C2 subgroups, using the standard criteria $|\text{Log}_2 \text{ Fold Change}| > 1$ and adjusted $p$-value $< 0.05$. A total of 2,698 DEGs were identified, of which 827 were upregulated and 1,871 were downregulated (Figures 3(a) and 3(b)). To determine the function of the DEGs, GO, KEGG, and GSEA were used. The top 10 enriched terms of the GO analysis including the biological process, cellular component, and molecular function are shown in Figure 3(c) (Supplementary Table 5). Several important molecular mechanisms were enriched in the GO analysis, including regulation of trans-synaptic signaling, multicellular organismal signaling, signal release, synaptic membrane, postsynaptic membrane, ion channel activity, ion gated channel activity, and potassium ion transmembrane transporter activity. These results indicate that there is a close relationship between cellular signal transduction pathways and the molecular subgroups. KEGG enrichment analysis revealed several molecular pathways that may play a vital role in the development of pancreatic cancer, such as cytokine-cytokine receptor interaction, cAMP signaling pathway, cell adhesion molecules, pancreatic secretion, and primary immunodeficiency (Figure 3(d) and Supplementary Table 6). GSEA showed that apical junction, glycolysis, mitotic spindle, mTORC1 signaling, and p53 pathway were significantly activated, whereas bile acid metabolism and pancreas $\beta$ cell pathways were significantly inhibited (Figures 3(e) and 3(f)). Further study of these alterations might contribute to a better understanding of the distinct molecular mechanisms underlying the two molecular subgroups identified.

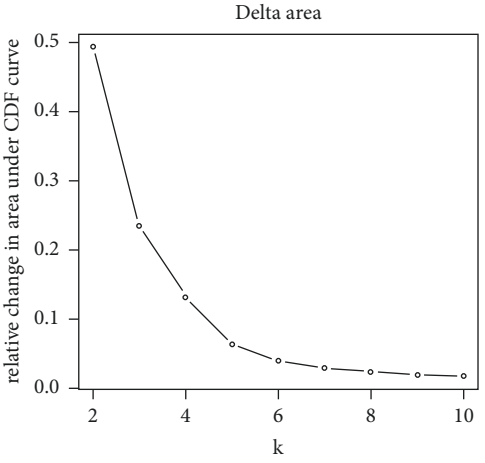
*3.3. Comparisons of Immune Infiltration, Mutation Spectrum, Clinical Features, and Copy Number Variation between the C1 and C2 Subgroups.* As the subgroups identified based on the immune-related genes were significantly associated with prognosis, we compared the molecular characteristics and some clinical features between the C1 and C2 subgroups. The infiltration level of six immune-cell populations was higher in the C2 subgroup compared with the C1 subgroup (Figure 4(a)). The difference was significant for the infiltration level of macrophages, myeloid dendritic cells, T cells CD4+, and T cells CD8+ (Figure 4(b)). These results indicate an immunosuppression state in the C1 subgroup, which might contribute to the poor prognosis associated with this group. Mutational spectrum analysis of the two subgroups was performed. The top 20 mutated genes in the C1 and C2 subgroups are shown in Figures 4(c) and 4(d), respectively. The mutational frequency of *KRAS*, *T753*, *RNF43*, *FLG*, *PCDH15*, and *ADAMTS16* was significantly higher in the C1 subgroup compared with the C2 subgroup (Supplementary Table 7; Figure 4(e)). A comparison of common clinicopathological characteristics indicated that the two subgroups significantly differed in tumor grade, T stage, and survival (Table 2). The C1 subgroup, with a worse prognosis, was associated with a more advanced tumor grade and T stage



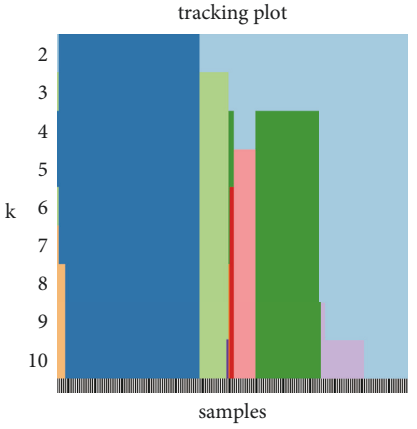

(a)




(b)

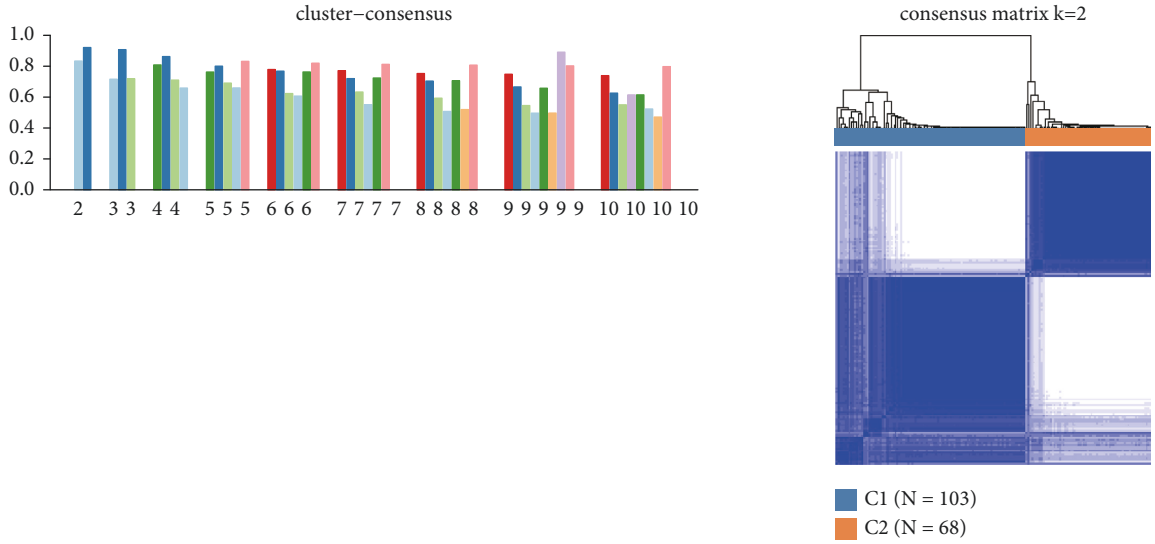



(c)




(d)

Figure 1: Continued.




(e) (f)

Figure 1: Identification of molecular subgroups using consensus cluster analysis. (a) Venn diagram that shows the intersection of the immune-related genes and the prognostic genes in pancreatic cancer. (b) Consensus cumulative distribution function (CDF) plot when the $k$ value ranges from 2 to 10. (c) Relationship between the relative change in area under the CDF curve and a different number of clusters ($k$ value). (d) Tracking plot of the total samples when the $k$ value ranges from 2 to 10. Each color represents different clusters. (e) Relationship between the average value of the consensus matrix and each cluster when $k$ value ranges from 2 to 10. The $Y$-axis stands for the average value of the consensus matrix and $X$-axis means the different number of clusters. (f) Heat map of consensus matrix when the total samples are divided into two groups, that is, C1 and C2 subgroups.

and higher mortality than the C2 subgroup (Figure 4(f)). Further analyses of genes with the difference in copy number variation were performed between the C1 and C2 subgroups. Interestingly, our results demonstrated that PTK2 and PLEC expression levels were significantly correlated with their copy number (Figures 5(a) and 5(d)). The PTK2, also known as FAK, upregulation of its expression could accelerate progression and contribute to an immunosuppressive environment of pancreatic cancer [45, 46]. PTK2 gene expression was higher in the C1 than the C2 subgroup (Figure 5(b)). And the frequencies of amplification and single gain were also higher in the C1 subgroup (Figure 5(c)). In addition, previous research proved that the PLEC gene could serve as a novel biomarker to identify preinvasive, primary, and metastatic pancreatic ductal adenocarcinoma, and its expression was continuously increasing along with tumor progression [47]. Similarly, PLEC expression was also significantly higher in the C1 subgroup compared with the C2 subgroup, and frequencies of amplification and single gain were lower in the C2 subgroup than the C1 subgroup (Figures 5(e) and 5(f)). The above results might contribute to elucidating the underlying mechanisms behind the heterogeneity between different molecular subgroups.

*3.4. Hub Genes Associated with the Molecular Subgroups Were Identified via WGCNA.* We performed WGCNA of 2,698 DEGs to identify specific gene coexpression modules, and subsequently, we identified the gene coexpression module most correlated with the clinical traits. The complete clinical information is shown in the clustering dendrogram with the trait heat map (Figure 6(a)). The soft threshold of WGCNA was defined as 6 to maintain the balance between scale independence and mean connectivity (Figure 6(b)). Genes were separated into eight modules depicted in different colors in Figure 6(c). The correlation analysis between the gene modules and clinical traits identified that the module labeled as brown was the one most significantly correlated with the C1 subgroup (Figure 6(d)). In addition, this module was also significantly correlated with the T stage and the clinical stage. Gene significance, defined as the correlation between gene expression and clinical traits, was put in relation to module membership, defined as the correlation between the module and the gene expression profile. The correlation between module membership and gene significance for the T stage, the clinical stage, or the subgroup in the brown module is shown in Figures 6(e)–6(g). Detailed information on the genes in the brown module is provided in Supplementary Table 8. To further identify the hub genes in the brown module, the hub genes were selected using the software Cytoscape (Figure 7(a)) and identified using the database Metascape (Figure 7(b)). Analysis of the expression level of the hub genes indicated that the genes *TNNT1*, *KCNN4*, *SH2D3A*, and *PHLDA2* were differentially expressed between 179 pancreatic tumors and 171 normal tissue samples (Figure 7(c)). The expression level of the four genes was significantly correlated with the OS, with higher expression levels indicating a worse prognosis ($p < 0.05$; Figure 7(d)), and with RFS ($p < 0.05$; Figure 7(e)). These results suggest that these four genes might play an important

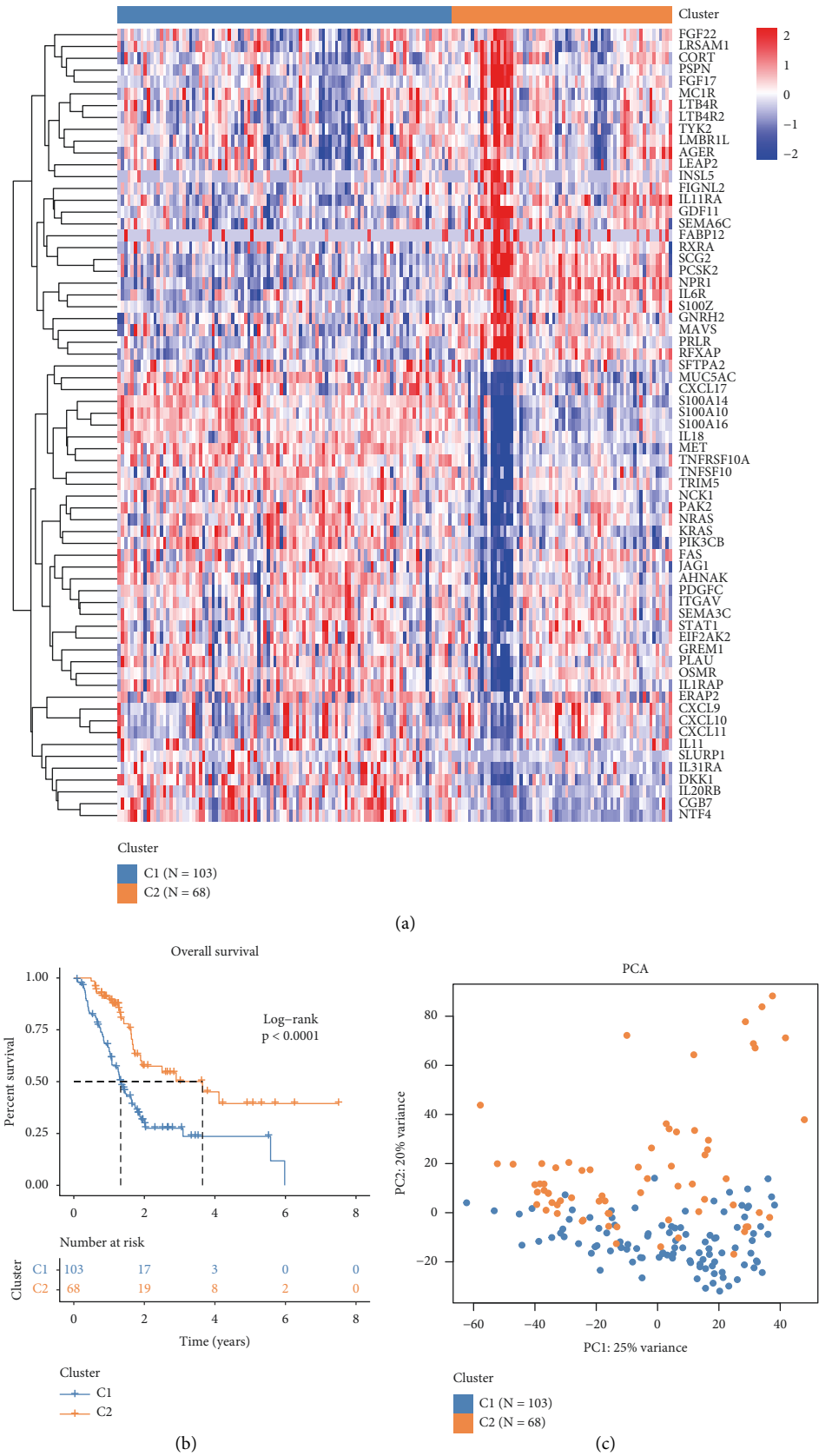



FIGURE 2: Overall survival and principal component analysis (PCA) of the C1 and C2 subgroups: (a) heat map of the expression level of the 67 immune-related genes used to identify molecular subgroups, (b) overall survival analysis between the C1 and C2 subgroups, and (c) PCA plot of the C1 and C2 subgroups.

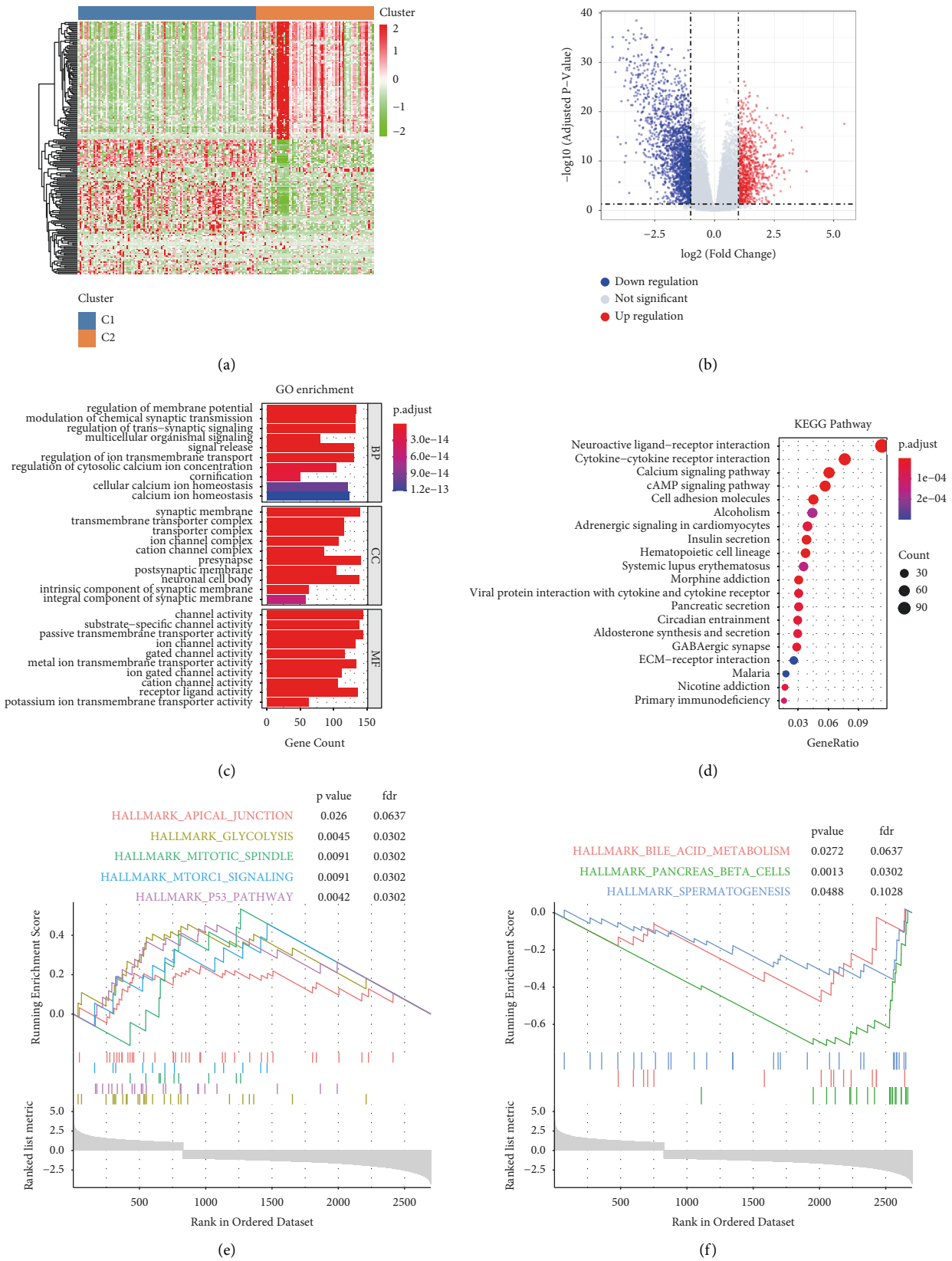



FIGURE 3: GO, KEGG, and GSEA based on the differentially expressed genes (DEGs) between the C1 and C2 subgroups: (a) heat map of the DEGs between the C1 and C2 subgroups; (b) volcano plot of the DEGs and genes with |Log2 Fold Change| > 1 and adjusted $p$-value < 0.05 were identified as significantly differentially expressed; (c) top 10 enriched items of the GO analysis including the biological process, cellular component, and molecular function; (d) top 20 signaling pathways of the KEGG enrichment analysis; (e) significantly activated signal pathways identified via GSEA; and (f) significantly inhibited signal pathways identified via GSEA.

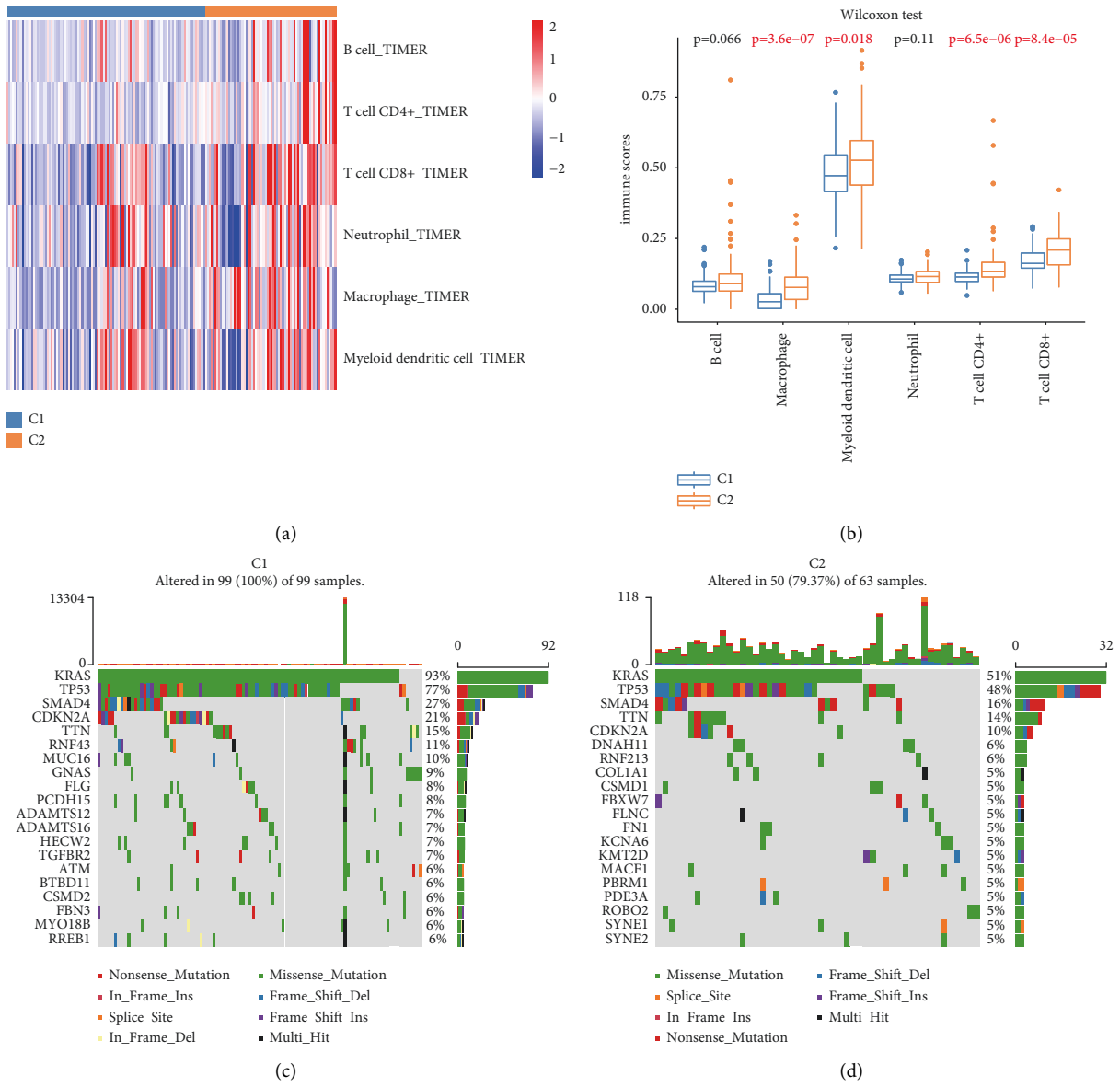



Figure 4: Continued.

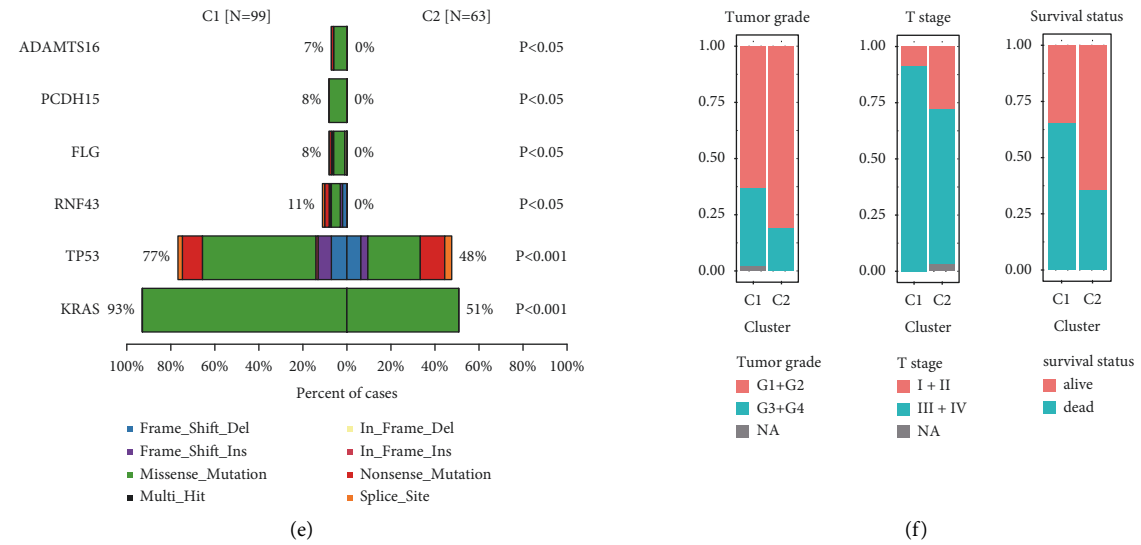



Figure 4: Comparisons of the immune infiltration, mutation spectrum, and clinical features between the C1 and C2 subgroups: (a) heat map of the immune infiltration levels of six types of immune cells between the C1 and C2 subgroups, (b) boxplot of the immune infiltration levels of the six types of immune cells between the C1 and C2 subgroups, (c) mutation spectrum of the top 20 genes in the C1 subgroup, (d) mutation spectrum of the top 20 genes in the C2 subgroup, (e) top six genes with different mutational frequencies between the C1 and C2 subgroups, and (f) distribution plots of different clinical features including tumor grade, T stage, and survival status between the C1 and C2 subgroups.

Table 2: Comparisons of clinicopathological information between the C1 and C2 subgroups in pancreatic cancer.

| Clinicopathological factors | Consensus cluster | | $p$-value |
|---|---|---|---|
| | C1 ($n = 103$) | C2 ($n = 68$) | |
| Age (year), median (IQR) | 65.0 (56.5–74.0) | 65.0 (57.0–71.0) | 0.4298[a] |
| Gender | | | 0.3495[b] |
| Female | 44 | 34 | |
| Male | 59 | 34 | |
| Tumor grade | | | 0.0202[b] |
| G1 + G2 | 65 | 55 | |
| G3 + G4 | 36 | 13 | |
| NA | 2 | 0 | |
| Clinical stage | | | 0.7055[c] |
| I + II | 97 | 64 | |
| III + IV | 5 | 2 | |
| NA | 1 | 2 | |
| T stage | | | 0.0006[b] |
| T1 + T2 | 9 | 19 | |
| T3 + T4 | 94 | 47 | |
| NA | 0 | 2 | |
| M stage | | | 0.6355[c] |
| M0 | 44 | 33 | |
| M1 | 3 | 1 | |
| NA | 56 | 34 | |
| N stage | | | 0.8867[b] |
| N0 | 29 | 18 | |
| N1 | 72 | 47 | |
| NA | 2 | 3 | |
| Survival status | | | 0.0001[b] |
| Alive | 36 | 44 | |
| Dead | 67 | 24 | |

[a]The Mann–Whitney U test, [b]chi-square test, and [c]Fisher's exact test were used to calculate the $p$-values. IQR: Interquartile range. NA: not available.

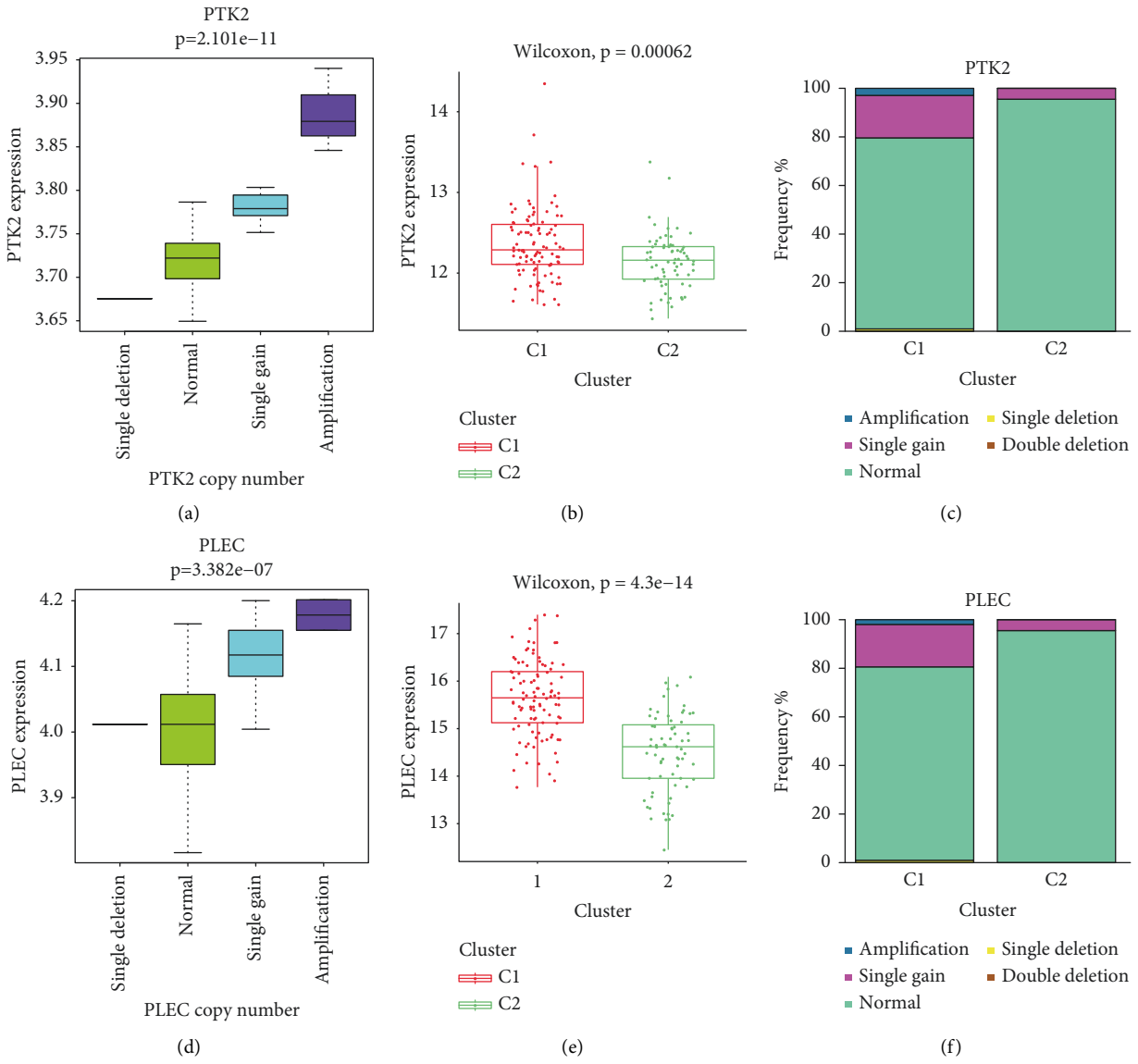



Figure 5: Comparisons of genes with copy number variation between the C1 and C2 subgroups: (a) correlation analysis between PTK2 gene expression level and its copy number variation, (b) PTK2 gene expression level between C1 and C2 subgroups, (c) frequency of copy number variation of PTK2 gene between C1 and C2 subgroups, (d) correlation analysis between PLEC gene expression level and its copy number variation, (e) PLEC gene expression level between C1 and C2 subgroups, and (f) frequency of copy number variation of PLEC gene between C1 and C2 subgroups.

role in the progression of pancreatic cancer. Further *in vitro* and *in vivo* experiments should be performed to verify its biological function in pancreatic cancer.

*3.5. Construction of the Prognostic Prediction Model Based on the 19 Immune-Related Genes.* In this study, a prognostic prediction model was developed to accurately stratify pancreatic cancer patients according to the risk level, which could be a tool of great importance in a clinical setting. To select the genes to construct the model, LASSO regression was performed using the 67 immune-relevant genes with prognostic significance. Samples in the TCGA data set were randomly divided into the training (70% of the total samples, $n = 120$) and validation sets (30% of the total samples, $n = 51$). The prognostic prediction model was constructed using the training data set using the LASSO regression model. The c-index was highest when $\log(\lambda) = -3.08$ (Figure 8(a)). With these parameter values, a total of 19 immune-related genes were selected to construct the model. The coefficients of different genes corresponding to various combination models are shown in Figure 8(b).

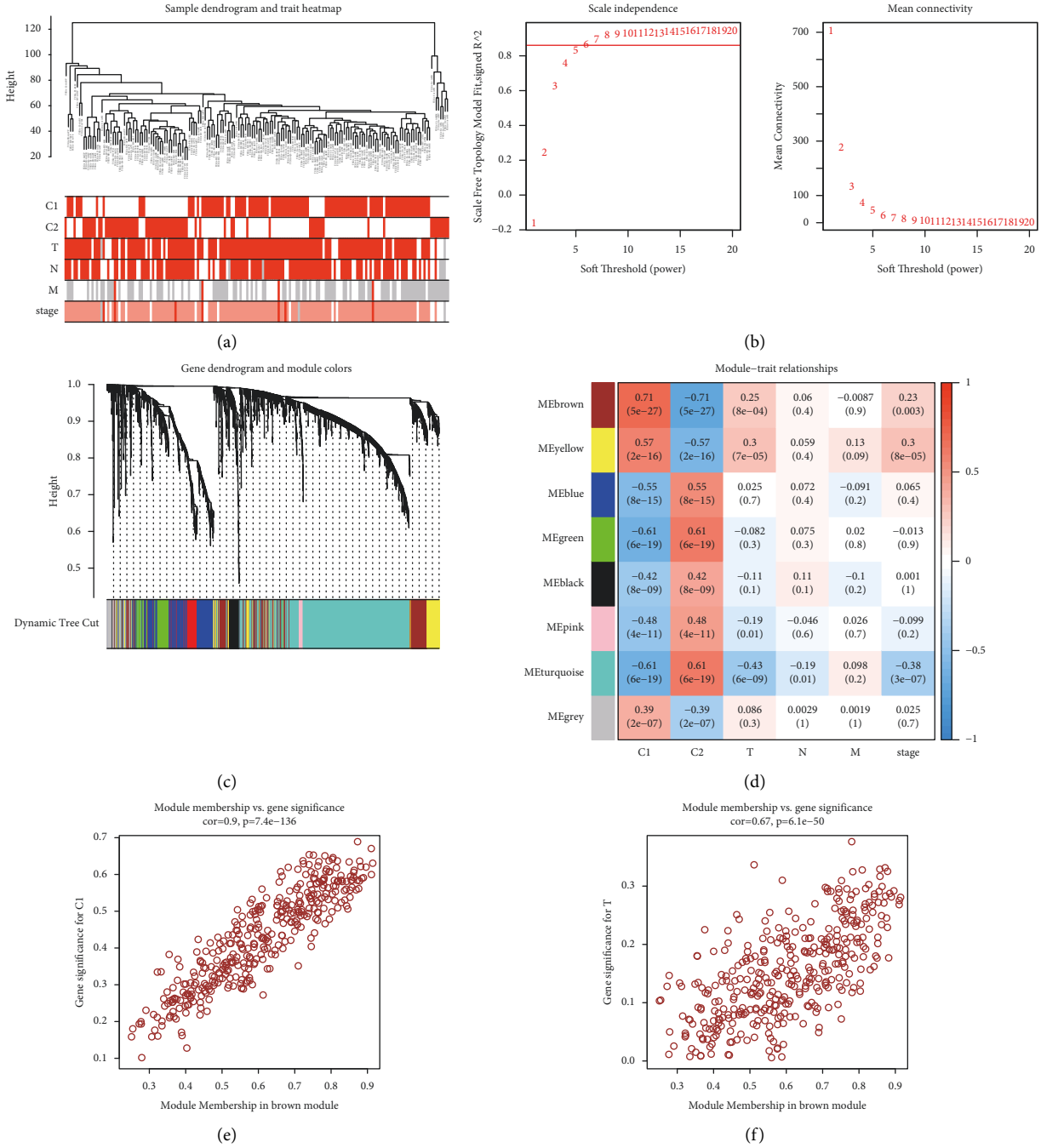
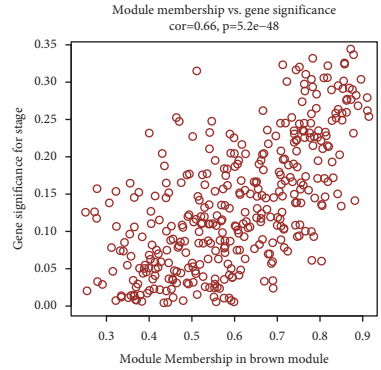
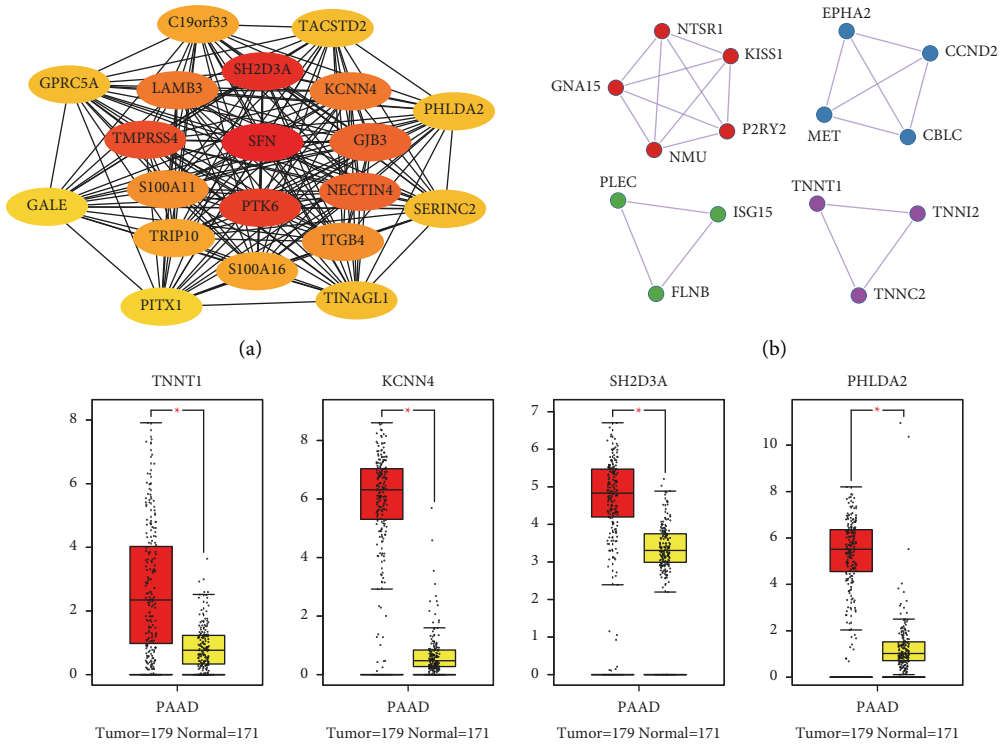



Figure 6: Continued.




(g)

Figure 6: Identification of genes associated with the molecular subgroups using WGCNA: (a) dendrogram of total samples and heat map of clinical traits, (b) correlation between the soft threshold and scale-free topology model fit (left plot) and correlation between the soft threshold and mean connectivity (right plot), (c) gene dendrogram and modules in different colors, (d) correlation analysis between the gene modules and clinical traits, (e) correlation analysis between module membership in brown module and gene significance for C1 subgroup, (f) correlation analysis between module membership in brown module and gene significance for T stage, and (g) correlation analysis between module membership in brown module and gene significance for the clinical stage.




(a) (b) (c)

Figure 7: Continued.

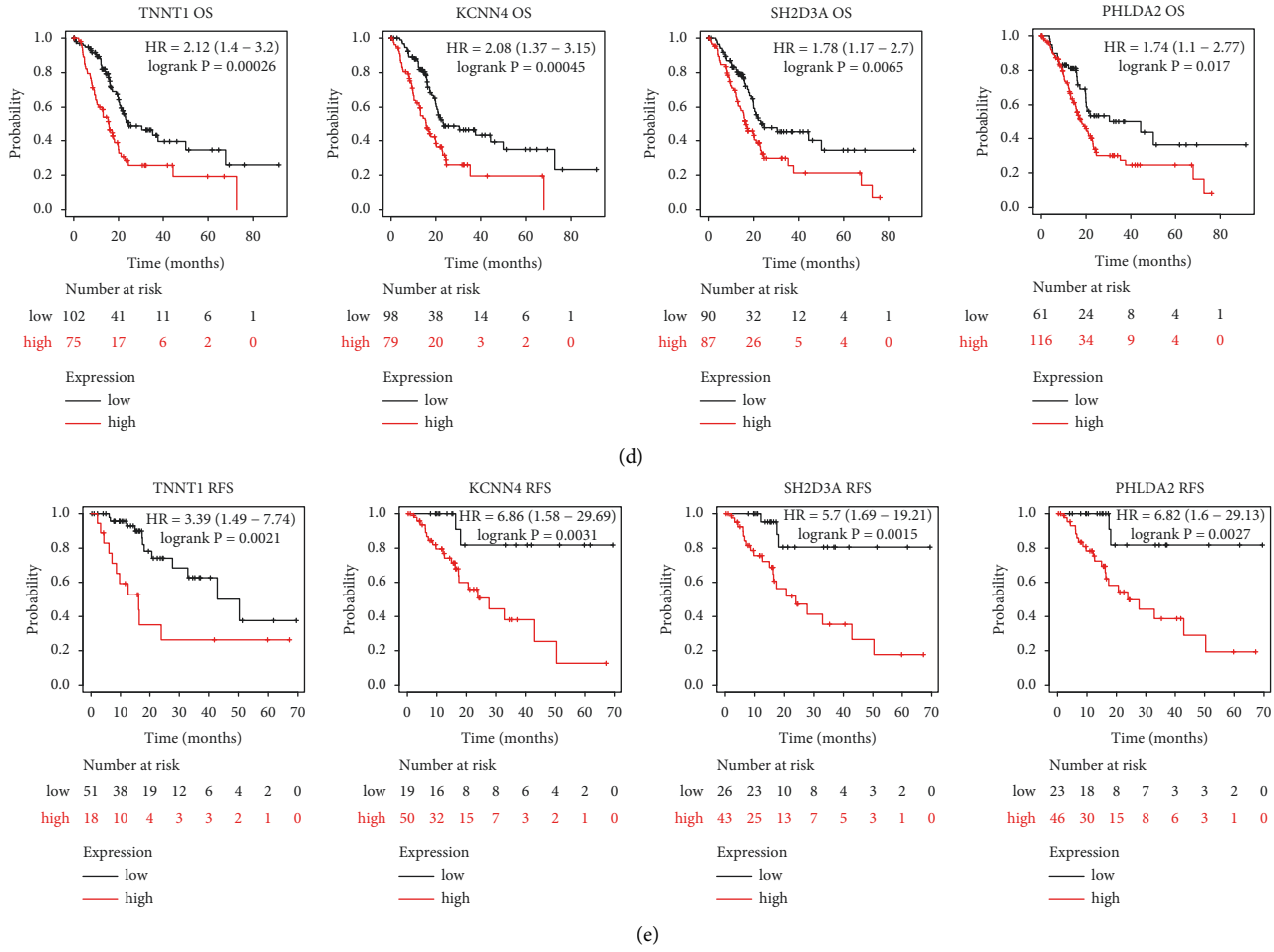



Figure 7: Identification of four potential target genes in pancreatic cancer: (a) top 20 hub genes of the brown module are selected by using the Cytoscape software; (b) hub genes of the brown module are selected by using the metascape database; (c) expression levels of *TNNT1*, *KCNN4*, *SH2D3A*, and *PHLDA2* between the 179 pancreatic tumors and 171 normal tissue samples based on the GEPIA database; (d) overall survival analysis of *TNNT1*, *KCNN4*, *SH2D3A*, and *PHLDA2* in pancreatic cancer using the Kaplan-Meier Plotter website; and (e) relapse-free survival of TNNT1, *KCNN4*, *SH2D3A*, and *PHLDA2* in pancreatic cancer using the Kaplan-Meier Plotter website.

Detailed information on 19 genes and their coefficients is listed in Table 1. A risk score was calculated for each patient according to the following formula:

$$\text{Risk score} = \sum_{n=1}^{19} (\text{Coefficient}_n * \text{Expression of gene}_n). \quad (1)$$

The distribution plots of risk score and OS between the high- and low-risk groups in the training set are shown in Figure 8(c). A heat map of the expression levels of the selected genes in the training set is depicted in Figure 8(e). The OS analysis between the high-risk ($n = 60$) and low-risk ($n = 60$) groups showed that the high-risk group had a significantly worse prognosis (log-rank $p < 0.001$, Cox $p < 0.001$, HR = 4.7, 95% CI: 3 – 7.3; Figure 8(f)). The efficiency of the model constructed based on the expression levels of the 19 genes was assessed using a ROC curve. The AUC of 1-, 3-, and 5-year survival time were 0.80, 0.89, and 0.96, respectively (Figure 8(g), c-index: 0.76, 95% CI: 0.66–0.87). Thereafter, the samples in the validation set were classified as high or low risk based on the median value of the overall risk scores. The risk score and OS distribution plots and the heat map of the genes for the internal validation set are provided in Figures 8(d) and 8(h), respectively. A significant difference between the high-risk ($n = 25$) and low-risk ($n = 26$) groups was also identified in the internal validation set (log-rank $p < 0.014$, Cox $p < 0.001$, HR = 3, 95% CI: 1.7 – 5.2; Figure 8(i)). The AUC of 1, 3, and 5 years in the ROC curve were 0.84, 0.91, and 1.00 in the internal validation set, respectively (Figure 8(j), c-index: 0.79, 95% CI: 0.65–0.94). These results indicated that the model based on the expression levels of the selected immune-related genes could serve as an accurate prognostic prediction tool in pancreatic cancer.

The universality of the model was investigated via external validation using multiple data sets obtained from the GEO database (GSE62452, GSE71729, and GSE78229). According to the risk score formula, the samples of the three external data sets were separated into high- and low-risk groups. The risk score and survival time distribution plots in

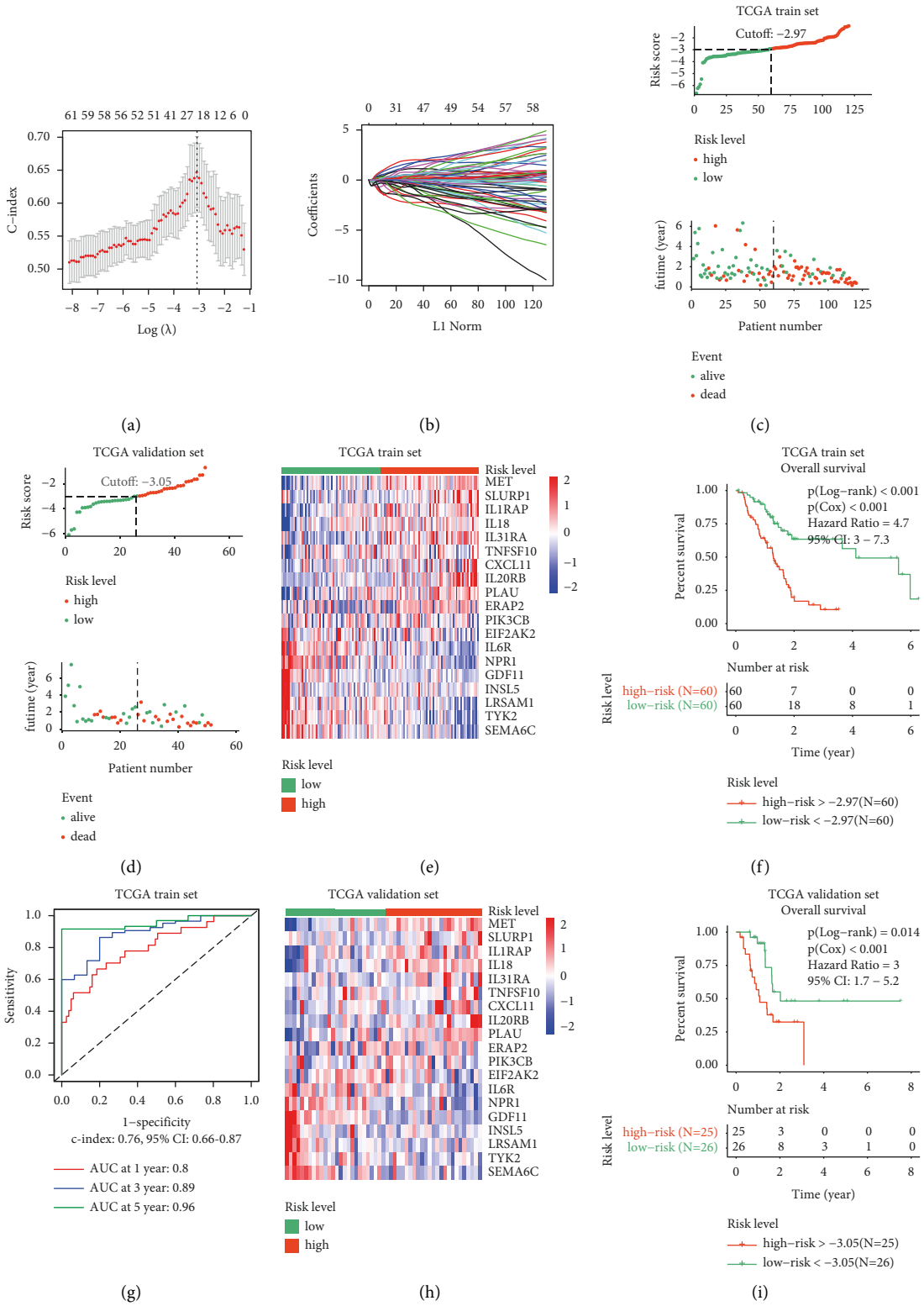



Figure 8: Continued.

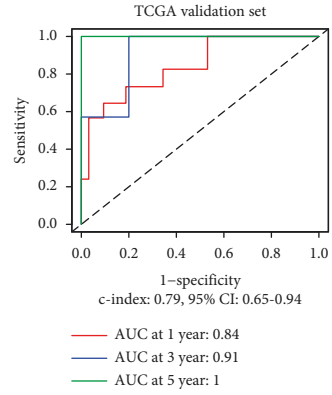



Figure 8: Construction and validation of the prognostic prediction model based on the 19 immune-related genes: (a) relationship between c-index and log ($\lambda$) value in the LASSO regression model, (b) coefficients of different genes corresponding to different combination models, (c) distribution plots of risk score and survival time between the high- and low-risk groups in the TCGA training set, (d) distribution plots of risk score and survival time between the high- and low-risk groups in the TCGA validation set, (e) heat map of 19 genes in the TCGA training set, (f) overall survival (OS) analysis between the high- and low-risk groups in the TCGA training set, (g) the ROC curve was used to evaluate the efficiency of the 19 gene model in the TCGA training set, (h) heat map of 19 genes in the TCGA validation set, (i) OS analysis between the high- and low-risk groups in the TCGA validation set, and (j) the ROC curve was used to evaluate the efficiency of the 19 gene model in the TCGA validation set.

the high- and low-risk groups from the GSE62452 data set are shown in Figure 9(a), those from GSE71729 in Figure 9(e), and those from GSE78229 in Figure 9(i). The heat map of the expression levels of the 19 immune-related genes in the GSE62452, GSE71729, and GSE78229 data sets are shown in Figures 9(b), 9(f), and 9(j), respectively. OS analysis showed that the prognosis of the low-risk group was better than that of the high-risk group in the three data sets (for GSE62452, log-rank $p = 0.035$, Cox $p = 0.009$, HR = 1.9, 95% CI: 1.2–3.1, Figure 9(c); for GSE71729, Log-rank $p = 0.018$, Cox $p = 0.016$, HR = 1.3, 95% CI: 1.1–1.7, Figure 9(g); and for GSE78229, Log-rank $p = 0.028$, Cox $p = 0.012$, HR = 2.1, 95% CI: 1.2–3.8, Figure 9(k)). The AUC for 1-year OS in the ROC curve in GSE62452 was 0.9 (Figure 9(d), c-index: 0.62, 95% CI: 0.47–0.76). The AUC for 1- and 3-year OS in GSE71729 were 0.66 and 0.56, respectively (Figure 9(h), c-index: 0.67, 95% CI: 0.56–0.78). The AUC for 1-year OS in the GSE78229 data set was 0.87 (Figure 9(l), c-index: 0.65, 95% CI: 0.49–0.81). These results demonstrated that the prognostic prediction model based on the 19 immune-related genes also had a satisfactory function in the external validation data sets.

*3.6. Uni- and Multivariate Cox Analyses of the Prognostic Factors.* To identify the prognosis-associated factors for pancreatic cancer patients, univariate Cox regression analysis was performed on the risk score model and common clinicopathological information, including age, gender, tumor grade, clinical stage, T stage, M stage, and N stage based on the data from TCGA data set. Risk score, age, tumor grade, T stage, and N stage could serve as prognosis-related factors ($p < 0.05$). Multivariate Cox regression analysis indicated that from these five prognosis-related factors, risk score and N stage were independent prognostic predictor factors ($p < 0.05$). Detailed results of the uni- and multivariate Cox analysis are provided in Table 3. These results demonstrate that the prognostic prediction model based on the 19 immune-related genes could serve as an independent prognostic factor.

## 4. Discussion

Pancreatic cancer patients with similar tumor morphology and clinicopathological status often show considerable differences in responses to the same therapeutic method, which may be caused by the vast molecular heterogeneity of the tumor tissue [10, 48, 49]. Accurately stratifying patients in molecular subgroups with specific OS, clinical outcome, and therapeutic responses would aid clinicians in making accurate decisions and administering individualized treatment, leading to a better prognosis in pancreatic cancer [50–52]. As the immune landscape is closely associated with the development of pancreatic cancer and immunotherapy stands as a promising therapeutic option in the near future [53, 54], we attempted to identify distinct molecular subgroups in pancreatic cancer by analyzing the expression level of immune-related genes.

For this purpose, we selected the expression data of the immune-related genes with $p < 0.01$ in both log-rank and Cox survival analyses for further analysis. To identify the distinct molecular subgroups, we performed consensus cluster analysis, an unsupervised technique that allows the grouping of similar objects and division of the data and has been widely used to study the existence of distinct subgroups in various cancer types [55, 56]. Using this approach, the samples were divided into two distinct molecular subgroups, C1 ($n = 103$) and C2 ($n = 68$), which presented significant differences in the prognosis.

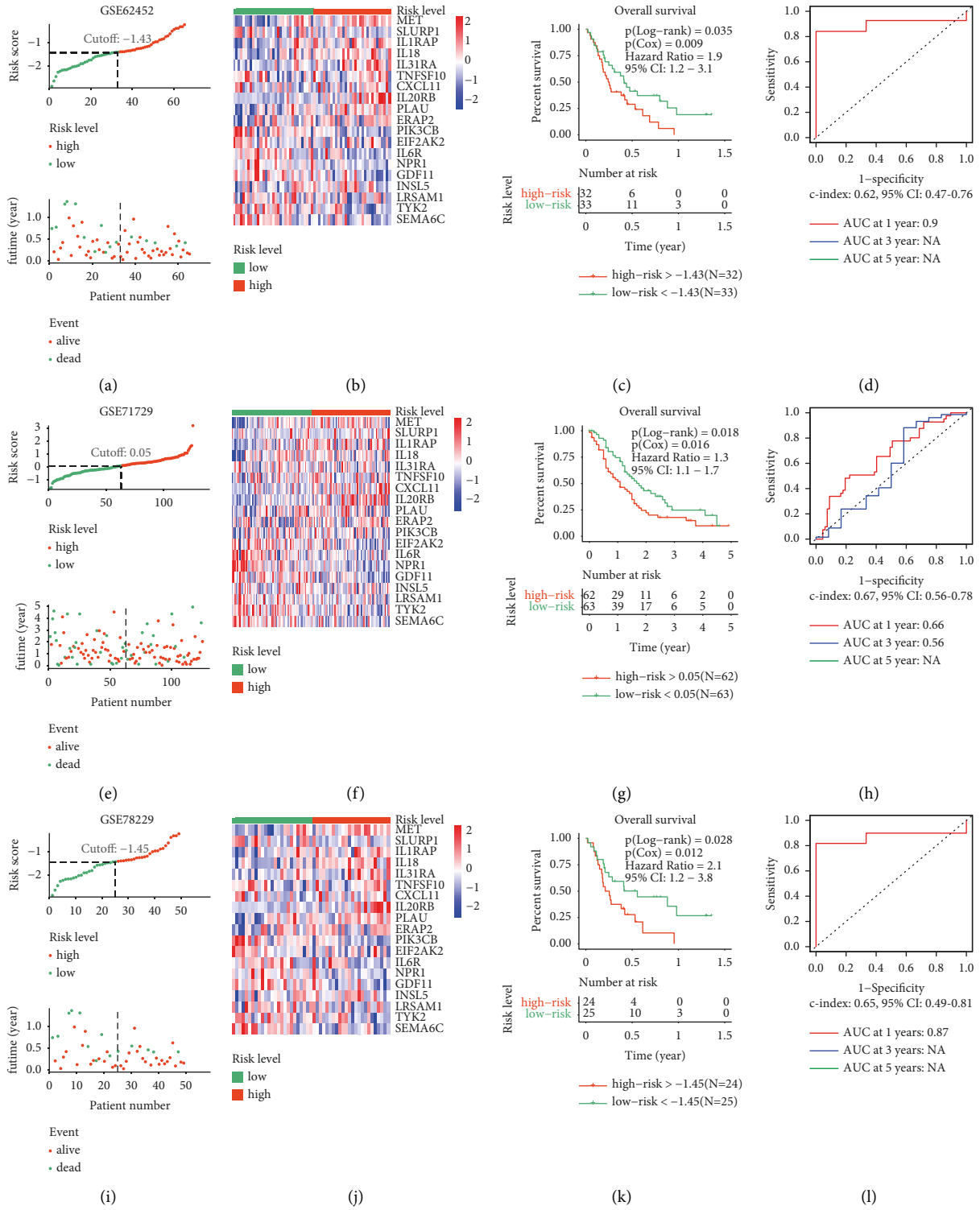



Figure 9: Validation of the prognostic prediction model in the three external validation sets from the GEO database: (a) distribution plots of risk score and survival time between the high- and low-risk groups in the GSE62452 data set, (b) heat map of 19 genes in the GSE62452 data set, (c) overall survival (OS) analysis between the high- and low-risk groups in the GSE62452 data set, (d) the ROC curve was used to evaluate the efficiency of the 19 gene model in the GSE62452 data set, (e) distribution plots of risk score and survival time between the high- and low-risk groups in the GSE71729 data set, (f) heat map of 19 genes in the GSE71729 data set, (g) OS analysis between the high- and low-risk groups in the GSE71729 data set, (h) the ROC curve was used to evaluate the efficiency of the 19 gene model in the GSE71729 data set, (i) distribution plots of risk score and survival time between the high- and low-risk groups in the GSE78229 data set, (j) heat map of 19 genes in the GSE78229 data set, (k) OS analysis between the high- and low-risk groups in the GSE78229 data set, and (l) the ROC curve was used to evaluate the efficiency of the 19 gene model in the GSE78229 data set.

Table 3: Uni- and multivariate Cox analyses of the prognostic factors in pancreatic cancer.

| Prognostic factors | Univariate Cox analysis | | | | Multivariate Cox analysis | | | |
|---|---|---|---|---|---|---|---|---|
| | HR | Lower 95% CI | Upper 95% CI | *p*-value | HR | Lower 95% CI | Upper 95% CI | *p*-value |
| Risk score | 3.949 | 2.801 | 5.566 | <0.001 | 3.838 | 2.626 | 5.610 | <0.001 |
| Age | 1.030 | 1.008 | 1.052 | 0.006 | 1.016 | 0.995 | 1.036 | 0.135 |
| Gender | 0.802 | 0.531 | 1.211 | 0.294 | | | | |
| Grade | 1.457 | 1.094 | 1.941 | 0.010 | 0.964 | 0.698 | 1.331 | 0.823 |
| Stage | 1.314 | 0.897 | 1.926 | 0.161 | | | | |
| T | 1.555 | 1.002 | 2.413 | 0.049 | 0.968 | 0.575 | 1.629 | 0.904 |
| M | 1.028 | 0.246 | 4.297 | 0.970 | | | | |
| N | 2.082 | 1.238 | 3.501 | 0.006 | 1.891 | 1.097 | 3.260 | 0.022 |

To further elucidate the underlying molecular mechanisms behind the established molecular subgroups, we utilized GO, KEGG, and GSEA by employing the DEGs to identify the signaling pathways specifically associated with each subtype. Immune infiltration analysis revealed that the immune scores of six immune cell types were higher in the C2 subgroup than in the C1 subgroup. These differences were significant for macrophages, myeloid dendritic cells, and the T cell populations CD4+ and CD8+. These results suggest a state of immune suppression in the C1 subgroup, which might underlie the poor prognosis associated with this subgroup. Mutation analysis indicated that classical cancer-related genes such as KRAS, TP53, and RNF43 were more frequently mutated in the C1 subgroup [57–59]. Analyses of genes with the difference in copy number variation were performed between the C1 and C2 subgroups. The expressions of the cancer-promoting PTK2 and PLEC genes were significantly higher in the C1 subgroups than the C2 subgroups. Besides, the two genes' frequencies of amplification and single gain were lower in the C2 subgroup than the C1 subgroup. Finally, analysis of the clinical features in the two subgroups revealed that a more advanced tumor grade and T stage and higher mortality were associated with the C1 subgroup. Our analysis provides a comprehensive perspective of the differential characteristics of the two identified groups.

To identify the DEGs specifically associated with each molecular subgroup, we performed WGCNA. The coexpression gene module labeled in brown in Figure 6(d) was the one most significantly correlated with the molecular subgroups. The hub genes in this module were identified using the Cytoscape software and the Metascape database. The expression level of genes *TNNT1*, *KCNN4*, *SH2D3A*, and *PHLDA2* was significantly different between the 179 pancreatic tumors and 171 normal tissue samples. In addition, the expression level of these genes was significantly correlated with the OS and RFS of pancreatic cancer patients. A previous study reported that *TNNT1* is significantly upregulated in breast tumor samples, and it facilitated their uncontrolled proliferation of tumor cells [60]. In addition, another study reported *TNNT1* overexpression in colorectal cancer cells, where it enhances their proliferation, migration, and invasion capacities [61]. *KCNN4* has been shown to modulate epithelial-mesenchymal transition and cell apoptosis, increasing the malignant behavior of papillary thyroid cancer cells [62]. Finally, high expression of *SH2D3A* has been reported to enhance the progression of ovarian cancer [63], and downregulation of *PHLDA2* has been reported to significantly inhibit the development of colorectal cancer through the PI3K/AKT signaling pathway [64]. In the future, we aim to study in detail the function of these four genes using *in vitro* and *in vivo* approaches.

We constructed a risk model using the expression levels of 19 immune-related genes using LASSO regression. The resulting model performed well in the internal training and validation TCGA subsets, as well as in three external validation data sets, although the AUC for 3- and 5-year OS were not all available in the three external data sets due to time limitations. Finally, uni- and multivariate Cox analyses demonstrated that the prognostic-prediction model developed using these genes could serve as an independent prognostic factor, indicating its potential use in a clinical setting.

## 5. Conclusions

In this study, we established two specific molecular subgroups based on the immune-related genes with prognostic significance using consensus cluster analysis. The two subgroups demonstrated significant differences in the OS and clinical features, including tumor grade, T stage, and survival status. The signaling pathways identified using GO, KEGG, and GSEA could contribute to understanding the underlying mechanisms behind the molecular classifications. The immune infiltration, copy number variation, and mutation spectrum analysis provided novel insights into the molecular subgroups. Four potential biomarkers, TNNT1, KCNN4, SH2D3A, and PHLDA2, were identified using WGCNA. The prognostic prediction model based on 19 immune-related genes could serve as an effective tool to predict the overall survival of pancreatic cancer patients.

## Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

## Authors' Contributions

XF, LK, and XT contributed to the conception and design of the study. CS, GW, and CL contributed to data acquisition and preprocessing. XF and LK performed the statistical analysis. XF and LK wrote the first draft of the manuscript. CW, PL, and XT reviewed and edited the manuscript. All authors contributed to manuscript revision and read and approved the submitted version. Xiang Fei and Lingming Kong contributed equally to this article.

## Acknowledgments

The results published here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. The authors would like to thank Editage (https://www.editage.cn) for English language editing. This study was supported by the National Natural Science Foundation of China (grant number 81902953), Natural Science Foundation of Liaoning Province (grant number 180530068), and 345 Talent Project of Shengjing Hospital of China Medical University.

## Supplementary Materials

Supplementary Tables 1–8 were uploaded to the submission system. Supplementary Table 1: Detailed information on pancreatic cancer samples from the TCGA data set. Supplementary Table 2: Detailed information on pancreatic cancer samples from the three validation data sets including GSE62452, GSE71729, and GSE78229. Supplementary Table 3: A total of 67 immune-related genes with prognostic significance ($p$-values < 0.01 in both the log-rank and Cox analyses) were selected for further analysis. Supplementary Table 4: Detailed information on the differentially expressed genes between the C1 and C2 subgroups. Supplementary Table 5: Detailed results of the gene ontology (GO) analysis including the biological process (BP), cellular component (CC), and molecular function (MF). Supplementary Table 6: Detailed results of the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. Supplementary Table 7: Genes with the difference in the mutational frequency between the C1 and C2 subgroups. Supplementary Table 8: Detailed information on the genes in the brown module of the weighted correlation network analysis (WGCNA). (*Supplementary Materials*)

## References

[1] J. D. Mizrahi, R. Surana, J. W. Valle, and R. T. Shroff, "Pancreatic cancer," *The Lancet*, vol. 395, no. 10242, p. 2008, 2020.

[2] J. Kleeff, M. Korc, M. Apte et al., "Pancreatic cancer," *Nature Reviews Disease Primers*, vol. 2, no. 1, p. 16022, 2016.

[3] O. Strobel, J. Neoptolemos, D. Jäger, and M. W. Büchler, "Optimizing the outcomes of pancreatic cancer surgery," *Nature Reviews Clinical Oncology*, vol. 16, no. 1, pp. 11–26, 2019.

[4] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA: A Cancer Journal for Clinicians*, vol. 70, no. 1, pp. 7–30, 2020.

[5] I. Regel, J. Mayerle, and M. Ujjwal Mukund, "Current strategies and future perspectives for precision medicine in pancreatic cancer," *Cancers*, vol. 12, no. 4, p. 1024, 2020.

[6] J. Xiang, L. Liu, W. Wang et al., "Metabolic tumor burden: a new promising way to reach precise personalized therapy in PDAC," *Cancer Letters*, vol. 359, no. 2, pp. 165–168, 2015.

[7] Y. Yu, Y. Tong, A. Zhong, Y. Wang, R. Lu, and L. Guo, "Identification of Serum microRNA-25 as a novel biomarker for pancreatic cancer," *Medicine*, vol. 99, no. 52, 2020.

[8] Y. Liu, W. Feng, W. Liu et al., "Circulating lncRNA ABHD11-AS1 serves as a biomarker for early pancreatic cancer diagnosis," *Journal of Cancer*, vol. 10, no. 16, pp. 3746–3756, 2019.

[9] C. Torres and P. J. Grippo, "Pancreatic cancer subtypes: a roadmap for precision medicine," *Annals of Medicine*, vol. 50, no. 4, pp. 277–287, 2018.

[10] E. A. Collisson, P. Bailey, D. K. Chang, and A. V. Biankin, "Molecular subtypes of pancreatic cancer," *Nature Reviews Gastroenterology & Hepatology*, vol. 16, no. 4, pp. 207–220, 2019.

[11] P. Bailey, D. K Chang, D. K. Chang et al., "Genomic analyses identify molecular subtypes of pancreatic cancer," *Nature*, vol. 531, no. 7592, pp. 47–52, 2016.

[12] E. A. Collisson, A. Sadanandam, P. Olson et al., "Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy," *Nature Medicine*, vol. 17, no. 4, pp. 500–503, 2011.

[13] W. Kwon, J. Namkung, S. Han et al., "Molecular subtypes of pancreatic cancer based on miRNA expression profiles have independent prognostic value," *Journal of Gastroenterology and Hepatology*, vol. 18, no. 6, 2016.

[14] M. Huber, C. U. Brehm, T. M. Gress et al., "The immune microenvironment in pancreatic cancer," *International Journal of Molecular Sciences*, vol. 21, no. 19, p. 7307, 2020.

[15] R. J. Torphy, R. D. Schulick, and Y. Zhu, "Understanding the immune landscape and tumor microenvironment of pancreatic cancer to improve immunotherapy," *Molecular Carcinogenesis*, vol. 59, no. 7, pp. 775–782, 2020.

[16] E. Zhang, J. He, H. Zhang et al., "Immune-related gene-based novel subtypes to establish a model predicting the risk of prostate cancer," *Frontiers in Genetics*, vol. 11, Article ID 595657, 2020.

[17] Cancer Genome Atlas Research Network, "Electronic address aadhe, cancer genome atlas research N. Integrated genomic characterization of pancreatic ductal adenocarcinoma," *Cancer Cell*, vol. 32, no. 2, pp. 185–203, 2017.

[18] S. Yang, P. He, J. Wang et al., "A novel MIF signaling pathway drives the malignant character of pancreatic cancer by targeting NR3C2," *Cancer Research*, vol. 76, no. 13, pp. 3838–3850, 2016.

[19] R. A. Moffitt, R. Marayati, E. L. Flate et al., "Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma," *Nature Genetics*, vol. 47, no. 10, pp. 1168–1178, 2015.

[20] J. Wang, S. Yang, P. He et al., "Endothelial nitric oxide synthase traffic inducer (NOSTRIN) is a negative regulator of disease aggressiveness in pancreatic cancer," *Clinical Cancer Research*, vol. 22, no. 24, pp. 5992–6001, 2016.

[21] R. Edgar, M. Domrachev, and A. E. Lash, "Gene expression omnibus: NCBI gene expression and hybridization array data repository," *Nucleic Acids Research*, vol. 30, no. 1, pp. 207–210, 2002.

[22] T. Barrett, S. E. Wilhite, P. Ledoux et al., "NCBI GEO: archive for functional genomics data sets-update," *Nucleic Acids Research*, vol. 41, no. D1, pp. D991–D995, 2012.

[23] S. Bhattacharya, P. Dunn, C. G. Thomas et al., "ImmPort, toward repurposing of open access immunological assay data for translational and clinical research," *Scientific Data*, vol. 5, no. 1, Article ID 180015, 2018.

[24] 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.

[25] T. M. Therneau and P. M. Grambsch, *Modeling Survival Data: Extending the Cox Model*, Springer Science & Business Media, Berlin, Germany, 2000.

[26] T. M. Therneau, "A package for survival analysis in R," 2021, https://cran.r-project.org/web/packages/survival/vignettes/survival.pdf.

[27] A. Kassambara, M. Kosinski, and P. Biecek, "Survminer: drawing survival curves using 'ggplot2'," 2020, https://cran.r-project.org/web/packages/survminer/index.html.

[28] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biology*, vol. 15, no. 12, p. 550, 2014.

[29] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, "ClusterProfiler: an R package for comparing biological themes among gene clusters," *OMICS: A Journal of Integrative Biology*, vol. 16, no. 5, pp. 284–287, 2012.

[30] G. Yu, "Enrichplot: visualization of functional enrichment result," 2019, https://www.bioconductor.org/packages/devel/bioc/manuals/enrichplot/man/enrichplot.pdf.

[31] T. Li, J. Fu, Z. Zeng et al., "TIMER2.0 for analysis of tumor-infiltrating immune cells," *Nucleic Acids Research*, vol. 48, no. W1, pp. W509–W514, 2020.

[32] T. Li, J. Fan, B. Wang et al., "TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells," *Cancer Research*, vol. 77, no. 21, pp. e108–e110, 2017.

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

[34] A. Mayakonda, D.-C. Lin, Y. Assenov, C. Plass, and H. P. Koeffler, "Maftools: efficient and comprehensive analysis of somatic variants in cancer," *Genome Research*, vol. 28, no. 11, pp. 1747–1756, 2018.

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

[36] P. Langfelder and S. Horvath, "Fast R functions for robust correlations and hierarchical clustering," *Journal of Statistical Software*, vol. 46, no. 11, 2012.

[37] 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.

[38] D. Otasek, J. H. Morris, J. Bouças, A. R. Pico, and B. Demchak, "Cytoscape automation: empowering workflow-based network analysis," *Genome Biology*, vol. 20, no. 1, p. 185, 2019.

[39] Y. Zhou, B. Zhou, L. Pache et al., "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets," *Nature Communications*, vol. 10, no. 1, p. 1523, 2019.

[40] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, and Z. Zhang, "GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses," *Nucleic Acids Research*, vol. 45, no. W1, pp. W98–W102, 2017.

[41] Á. Nagy, G. Munkácsy, and B. Győrffy, "Pancancer survival analysis of cancer hallmark genes," *Scientific Reports*, vol. 11, no. 1, p. 6047, 2021.

[42] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.

[43] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for cox's proportional hazards model via coordinate descent," *Journal of Statistical Software*, vol. 39, no. 5, pp. 1–13, 2011.

[44] P. Blanche, J.-F. Dartigues, and H. Jacqmin-Gadda, "Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks," *Statistics in Medicine*, vol. 32, no. 30, pp. 5381–5397, 2013.

[45] J. Xu and W. Zhang, "EZR promotes pancreatic cancer proliferation and metastasis by activating FAK/AKT signaling pathway," *Cancer Cell International*, vol. 21, no. 1, p. 521, 2021.

[46] S. N. Symeonides, S. M. Anderton, and A. Serrels, "FAK-inhibition opens the door to checkpoint immunotherapy in Pancreatic Cancer," *Journal for Immuno Therapy of Cancer*, vol. 5, no. 1, p. 17, 2017.

[47] D. Bausch, S. Thomas, M. Mino-Kenudson et al., "Plectin-1 as a novel biomarker for pancreatic cancer," *Clinical Cancer Research*, vol. 17, no. 2, pp. 302–309, 2011.

[48] T. J. Grant, K. Hua, and A. Singh, "Molecular pathogenesis of pancreatic cancer," *Progress in Molecular Biology and Translational Science*, vol. 144, pp. 241–275, 2016.

[49] W. Yao, A. Maitra, and H. Ying, "Recent insights into the biology of pancreatic cancer," *eBioMedicine*, vol. 53, Article ID 102655, 2020.

[50] T. Y. S. Le Large, M. Mato Prado, J. Krell et al., "Bioinformatic analysis reveals pancreatic cancer molecular subtypes specific to the tumor and the microenvironment," *Expert Review of Molecular Diagnostics*, vol. 16, no. 7, pp. 733–736, 2016.

[51] N. A. Juiz, J. Iovanna, and N. Dusetti, "Pancreatic cancer heterogeneity can be explained beyond the Genome," *Frontiers in Oncology*, vol. 9, p. 246, 2019.

[52] K. L. Aung, S. E. Fischer, R. E. Denroche et al., "Genomics-driven precision medicine for advanced pancreatic cancer: early results from the compass trial," *Clinical Cancer Research*, vol. 24, no. 6, pp. 1344–1354, 2018.

[53] A. H. Morrison, K. T. Byrne, and R. H. Vonderheide, "Immunotherapy and prevention of pancreatic cancer," *Trends in Cancer*, vol. 4, no. 6, pp. 418–428, 2018.

[54] D. Schizas, N. Charalampakis, C. Kole et al., "Immunotherapy for pancreatic cancer: a 2020 update," *Cancer Treatment Reviews*, vol. 86, Article ID 102016, 2020.

[55] T. Liu, S. Yang, Y.-P. Cheng et al., "The N6-methyladenosine (m6A) methylation gene YTHDF1 reveals a potential diagnostic role for gastric cancer," *Cancer Management and Research*, vol. 12, pp. 11953–11964, 2020.

[56] Z. Li, C. Jiang, and Y. Yuan, "TCGA based integrated genomic analyses of ceRNA network and novel subtypes revealing potential biomarkers for the prognosis and target therapy of tongue squamous cell carcinoma," *PLoS One*, vol. 14, no. 5, Article ID e0216834, 2019.

[57] A. M. Waters and C. J. Der, "KRAS: the critical driver and therapeutic target for pancreatic cancer," *Cold Spring Harbor Perspectives in Medicine*, vol. 8, no. 9, Article ID a031435, 2018.

[58] M. Aslan, R. Shahbazi, K. Ulubayram, and B. Ozpolat, "Targeted therapies for pancreatic cancer and hurdles ahead," *Anticancer Research*, vol. 38, no. 12, pp. 6591–6606, 2018.

[59] X. Jiang, H.-X. Hao, J. D. Growney et al., "Inactivating mutations of RNF43 confer Wnt dependency in pancreatic ductal adenocarcinoma," *Proceedings of the National Academy of Sciences*, vol. 110, no. 31, pp. 12649–12654, 2013.

[60] Y. Shi, Y. Zhao, Y. Zhang et al., "TNNT1 facilitates proliferation of breast cancer cells by promoting G1/S phase transition," *Life Sciences*, vol. 208, pp. 161–166, 2018.

[61] Y. Chen, J. Wang, D. Wang et al., "TNNT1, negatively regulated by miR-873, promotes the progression of colorectal cancer," *The Journal of Gene Medicine*, vol. 22, no. 2, Article ID e3152, 2020.

[62] J. Wen, B. Lin, L. Lin, Y. Chen, and O. Wang, "KCNN4 is a diagnostic and prognostic biomarker that promotes papillary thyroid cancer progression," *Aging*, vol. 12, no. 16, pp. 16437–16456, 2020.

[63] G. Wang, H. Zhang, and P. Li, "Upregulation of hsa_circRNA_102958 indicates poor prognosis and promotes ovarian cancer progression through miR-1205/SH2D3A Axis," *Cancer Management and Research*, vol. 12, pp. 4045–4053, 2020.

[64] Z. Ma, S. Lou, and Z. Jiang, "PHLDA2 regulates EMT and autophagy in colorectal cancer via the PI3K/AKT signaling pathway," *Aging*, vol. 12, no. 9, pp. 7985–8000, 2020.

Hindawi
Journal of Oncology
Volume 2022, Article ID 5286251, 16 pages
https://doi.org/10.1155/2022/5286251

*Research Article*

# Identification and Validation of a Hypoxia-Immune-Based Prognostic mRNA Signature for Oral Squamous Cell Carcinoma

**Shaohua Lv,[1,2] Zhipeng Qian,[3] Jianhao Li,[1] Songlin Piao,[1] and Jichen Li [1,2]**

[1]*Department of Oral and Maxillofacial Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin 150081, China*
[2]*Stomatology School, Harbin Medical University, 143 Yiman Street, Nangang District, Harbin, Heilongjiang, China*
[3]*College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China*

Correspondence should be addressed to Jichen Li; lijichen@163.com

Received 20 September 2021; Revised 22 November 2021; Accepted 20 December 2021; Published 7 February 2022

Academic Editor: Jincheng Guo



*Background*. Oral squamous cell carcinoma (OSCC) is a commonly encountered head and neck malignancy. Increasing evidence shows that there are abnormal immune response and chronic cell hypoxia in the development of OSCC. However, there is a lack of a reliable hypoxia-immune-based gene signature that may serve to accurately prognosticate OSCC. *Methods*. The mRNA expression data of OSCC patients were extracted from the TCGA and GEO databases. Hypoxia status was identified using the t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm. Both ESTIMATE and single-sample gene-set enrichment analysis (ssGSEA) were used for further evaluation of immune status. The DEGs in different hypoxia and immune status were determined, and univariate Cox regression was used to identify significantly prognostic genes. A machine learning method, least absolute shrinkage and selection operator (LASSO) Cox regression analysis, allowed us to construct prognostic gene signature to predict the overall survival (OS) of OSCC patients. *Results*. A total of 773 DEGs were identified between hypoxia high and low groups. According to immune cell infiltration, patients were divided into immune high, medium, and low groups and immune-associated DEGs were identified. A total of 193 overlapped DEGs in both immune and hypoxia status were identified. With the univariate and LASSO Cox regression model, eight signature mRNAs (FAM122C, RNF157, RANBP17, SOWAHA, KIAA1211, RIPPLY2, INSL3, and DNAH1) were selected for further calculation of their respective risk scores. The risk score showed a significant association with age and perineural and lymphovascular invasion. In the GEO validation cohort, a better OS was observed in patients from the low-risk group in comparison with those in the high-risk group. High-risk patients also demonstrated different immune infiltration characteristics from the low-risk group and the low-risk group showed potentially better immunotherapy efficacy in contrast to high-risk ones. *Conclusion*. The hypoxia-immune-based gene signature has prognostic potential in OSCC.

## 1. Introduction

The incidence of lip and oral cavity cancers has been on the rise, with an estimated 246,420 cases and 119,693 deaths occurring globally in 2018 [1]. Oral squamous cell carcinoma (OSCC) is the dominating subtype of oral cancer, which unfortunately harbors a dismal 5-year survival rate [2]. There is still an urgent need to discover accurate prognostic biomarkers and effective drug targets for OSCC.

Solid tumors such as OSCC are usually characterized by cellular hypoxia, which is generally considered to be a manifestation of poor prognosis [3, 4]. Cell hypoxia has evolved to represent a significant feature in cancer [5]. Hypoxic conditions could trigger the migration and invasion of OSCC cells [6–8]. Reports also suggest that hypoxia can induce epithelial to mesenchymal transition (EMT) in several types of cancer, including OSCC [9]. It can also stimulate angiogenesis through the activation of proangiogenic factors [10]. Zhang et al. reported that the expression of HIF-1$\alpha$, Glut-1, and CA9 can predict malignant transformation to OSCC [11]. It is undeniable that cellular hypoxia contributes significantly to the malignant properties of OSCC.

The immune system is crucial in OSCC tumorigenesis, progression, and prognosis [12–14]. Enhanced PD-1 and PD-L1 expressions were observed in OSCC patient serum and tissue samples [15]. Premalignant lesions were also found to harbor raised PD-L1 expressions [16]. Inhibiting the PD-1/PD-L1 axis is one of the most promising means to activate antitumor immunity [17]. A number of immune-related prognostic biomarkers have been identified in OSCC. Meehan et al. had previously constructed an immune-related signature that correlated with OSCC patient prognosis [18]. Some immune checkpoints have been documented to be closely related to poorer patient outcomes in OSCC [19]. One of them is the CTSG gene, which is an immune-related gene and has been identified as an independent biomarker and therapeutic target of OSCC [20]. Although these immune-related prognostic biomarkers have been discovered, few studies have explored the potential benefit of a combined assessment of both immunological characteristics and the hypoxic microenvironment in OSCC [21].

Several studies have alluded to the clinical significance of potential interactions between immune function and cellular hypoxia across various malignancies [22, 23]. Computational framework for prognosis assessment for cancer patients has been widely used in various tumors [24–27]. Although some hypoxia-related and immune-related models have been constructed to predict the prognosis of OSCC [28, 29], a reliable hypoxia-immune-integrated prognostic gene signature has not yet been established for OSCC. Here, we first accessed the mRNA expression profiles of OSCC patients from the TCGA database. Cell hypoxia was identified using the t-SNE algorithm, while the immune status and immune-related DEGs were identified using the ESTIMATE and ssGSEA algorithms. DEGs that were significant in both immune and cellular hypoxia phenomena were identified. This pool of genes was then subjected to a machine learning assessment to discern those which were potentially prognostic. We successfully constructed a hypoxia-immune-based gene signature that was useful in predicting the overall survival (OS) of OSCC patients.

## 2. Materials and Methods

*2.1. Data Collection and Preprocessing.* The discovery cohort included 303 OSCC individual data from the "TCGA-HNSC" project in The Cancer Genome Atlas (TCGA). The Genomic Data Commons (available at https://portal.gdc.cancer.gov) database was then accessed to extract the corresponding gene expression data. In this study, samples without data on the survival state and survival time were eliminated. Data were externally validated using an independent cohort. The expression profile of GSE41613 was downloaded from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) based on the GPL570 Affymetrix Human Genome U133 Plus 2.0 Array platform. A total of 97 OSCC patient data were included in this validated study. The IMvigor210 ($n = 348$) cohort with immunotherapy data and corresponding clinicopathological information was obtained from the IMvigor210CoreBiologies R package.

RNA-sequencing data (FPKM values) extracted from the TCGA database were translated into transcripts per kilobase million (TPM) values and normalized using log2 (TPM + 1). Raw GEO database-derived data were obtained using the RMA algorithm. Background adjustment, quantile normalization, and final summation of oligonucleotides per transcript were carried out using the median Polish algorithm of the Affy software package. Finally, the different probe IDs were converted into their respective gene symbols to annotate them, and the repeated gene expression values were averaged. All patients with included clinical information and survival data were also incorporated in this study.

*2.2. Identification of Hypoxia Status and Hypoxia-Related Differentially Expressed Genes (DEGs).* The presence of cell hypoxia was determined using the t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm. This nonparametric, unsupervised method was able to distinguish between patient clusters based on provided hallmarks or signatures. Based on the seven hypoxic signatures of Buffa, Elvidge, Eustace, Hu, Ragnum, Sorensen, and Winter [30], a nonlinear dimensionality reduction algorithm t-SNE was used to measure the Euclidean distance of any two patients in the TCGA cohort, which was then condensed into two-dimensional points. Three clusters were identified. Moreover, we obtained genes using the KEGG HIF-1 signaling pathway (https://www.kegg.jp/; ID:04066) to assess the hypoxic status. Of these genes, 14 were identified as "increased oxygen delivery" genes and 12 were identified as "reduced oxygen consumption" genes. Finally, the "DESeq2" software package was used to discern between DEGs significant to low- or high-hypoxic groups ($P$-value <0.05 and | log2 (fold change) |>1).

*2.3. Identification of Immune Status and Immune-Related DEGs.* We evaluated enrichment degrees of a total of 28 immune cells within each TCGA OSCC sample using the ssGSEA method. The OSCC samples were arbitrarily divided into high-, medium-, and low-immune groups using hierarchical clusters (namely, "Immune_Low," "Immune_Medium," and "Immune_High") based on the above immune matrix. Both degrees of immunocyte infiltration degree (immune score) and stromal level (stromal score) were assessed using the Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) algorithm in TCGA OSCC samples to validate the above immune status grouping. The immune-related DEGs between the Immune_Low group and the Immune_High group were identified by the "DESeq2" software package using similar parameters as in the cell hypoxia analysis. The Metascape (https://metascape.org/) database was used to construct GO and KEGG enrichment pathways of the selected DEGs.

*2.4. Construction and Verification of Prognostic Signatures Related to Hypoxia and Immune Status.* We took the intersection between hypoxia- and immune-related

differentially expressed genes (DEGs) and selected those overlapped genes for subsequent analysis. First, using the "survival" package in R, we employed univariate Cox regression on these overlapped genes and overall survival (OS) of OSCC in the TCGA to identify survival-related hypoxia-immune-related DEGs. Those with a $P < 0.05$ were considered as significant. Then, using the "glmnet" package in R, the Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression model was applied to select the optimal variables from all identified hypoxia-immune-related prognostic DEGs in the discovery cohort. LASSO is a kind of linear regression that uses shrinkage and can be applied to high-dimensional data. In this study, fivefold cross-validation was employed to select the minimal penalty term ($\lambda$). All patients were then subjected to a risk score calculation as follows:

$$\text{risk score} = \sum (\text{coefficient} * \text{expression of the signature gene}).$$

The risk score was defined as the sum of the product of the prognostic signature gene expression levels and the corresponding LASSO model-derived coefficients. The optimal cut point was determined by a method of maximally selected rank statistics based on an individual level risk score to further refine a means to stratify patients with OSCC according to their prognosis.

*2.5. Relationship between Risk Scores and the Immune Microenvironment of OSCC.* Profiles of 22 immune cell infiltration were determined in both the low- and high-risk groups using the CIBERSORT algorithm. Using gene expression data, the CIBERSORT is a deconvolution algorithm specifically designed to evaluate cellular composition in tissues. In addition, two immune checkpoints (CTLA-4 and PD-1), APM, CYT, TILs, and TIS allowed for further analysis of the relationship between the low- and high-risk groups.

The major histocompatibility complex (MHC) molecules were subjected to relative antigen presentation mechanism (APM) calculation. The degree of immune cytotoxic activity (CYT) based on granzyme A (GZMA) and perforin-1 (PRF1), both of which are related to CD8 + T cell activation, was calculated on previously reported formulas. Tumor prognosis is known to be associated with the density of T cells tumor infiltration. The proportion of tumor-infiltrating lymphocytes (TILs) was calculated. Lastly, we used the tumor inflammation signature (TIS), which represents a commonly used 18-gene signature in research, to quantify background adaptive immune responses, which are normally inhibited in tumors. The TIS score was derived in terms of the mean of log2-FPKM gene expression of the selected marker genes.

*2.6. Statistical Analysis.* The R version 3.6.1 and its related packages were used to carry out all data analyses. The t-SNE algorithm was performed using the "Rtsne" of the R package on the basis of nonlinear dimensionality reduction. The "estimate" package was used to determine immune scores. The "glmnet" package was used to carry out LASSO Cox regression modeling. To identify independent risk factors for survival, univariate and multivariate Cox regression analyses were employed after the adjustment of covariates. The "survival" package was used to perform the Cox regression model and Kaplan–Meier analyses. A $P$ value of <0.05 suggested that the results were statistically significant.

## 3. Results

*3.1. Hypoxia Status and Hypoxia-Related DEGs in OSCC.* In the TCGA cohort, cell hypoxia status was evaluated using ssGSEA analysis based on eight hypoxia signatures (Buffa, Elvidge, Eustace, Hu, Ragnum, Seigneuric, Sorensen, and Winter) [30] (Figure 1(a)). Based on the quantitative score of each hypoxia signature, we found that the "Seigneuric" signature had the lowest correlation with the others and was therefore not included in this study (Figure 1(b)). The remaining seven gene signatures were used in the nonlinear dimensionality reduction algorithm calculation of two-dimensional points between any two patients (see Methods for details). As shown in Figure 1(c), three clusters of patients were identified, and each patient was assigned to their nearest cluster. A total of 64, 142, and 97 cases were classified into Cluster 1, Cluster 2, and Cluster 3. There was no significant difference in terms of patient survival among the three clusters (log-rank test, $P = 0.088$) (Figure 1(d)). However, there was a significant difference when comparing Cluster 1 with Cluster 2 and Cluster 3 (log-rank test, $P = 0.029$). Clusters 2 and 3 were then combined for further comparison against Cluster 1, which appeared to possess the most favorable overall survival outcome. Clusters 2 and 3 may differ from Cluster 1 in terms of the degree of cellular hypoxia. We further explored differences in expressions of KEGG HIF-1 pathway molecules between the two new clusters (Cluster 1 versus Clusters 2 and 3). The resultant identified genes were classified as being involved in "reduced oxygen consumption" (12 genes) or "increased oxygen delivery" (14 genes). Of those 14 genes involved in increased oxygen delivery, 11 (78.57%) were found to be overexpressed in Clusters 2 and 3 compared with those in Cluster 1 (Figure 1(e)). Seven of the 12 genes related to reduced oxygen consumption (58.33%) were noted to be overexpressed in Clusters 2 and 3 (Figure 1(f)). We therefore proved that both these clusters were of different cell hypoxia status. Patients in Clusters 2 and 3 and Cluster 1 were then renamed to Hypoxia_High or Hypoxia_Low groups, respectively. A total of 773 DEGs were then identified using the DESeq2 package between the Hypoxia_High and Hypoxia_Low groups. Of these genes, 220 were upregulated, whereas 553 were downregulated (Figures 1(g) and 1(h)).

*3.2. Immune Status and Immune-Related DEGs in OSCC.* Based on the proportions of 28 immune cells quantified by ssGSEA, OSCC samples in the TCGA database were classified into high-, medium-, and low-immune groups (named Immune_ High, Immune_ Medium and Immune_ Low, resp. (Figure 2(a)). In accordance with calculations using the ESTIMATE approach, both immune and stromal scores of

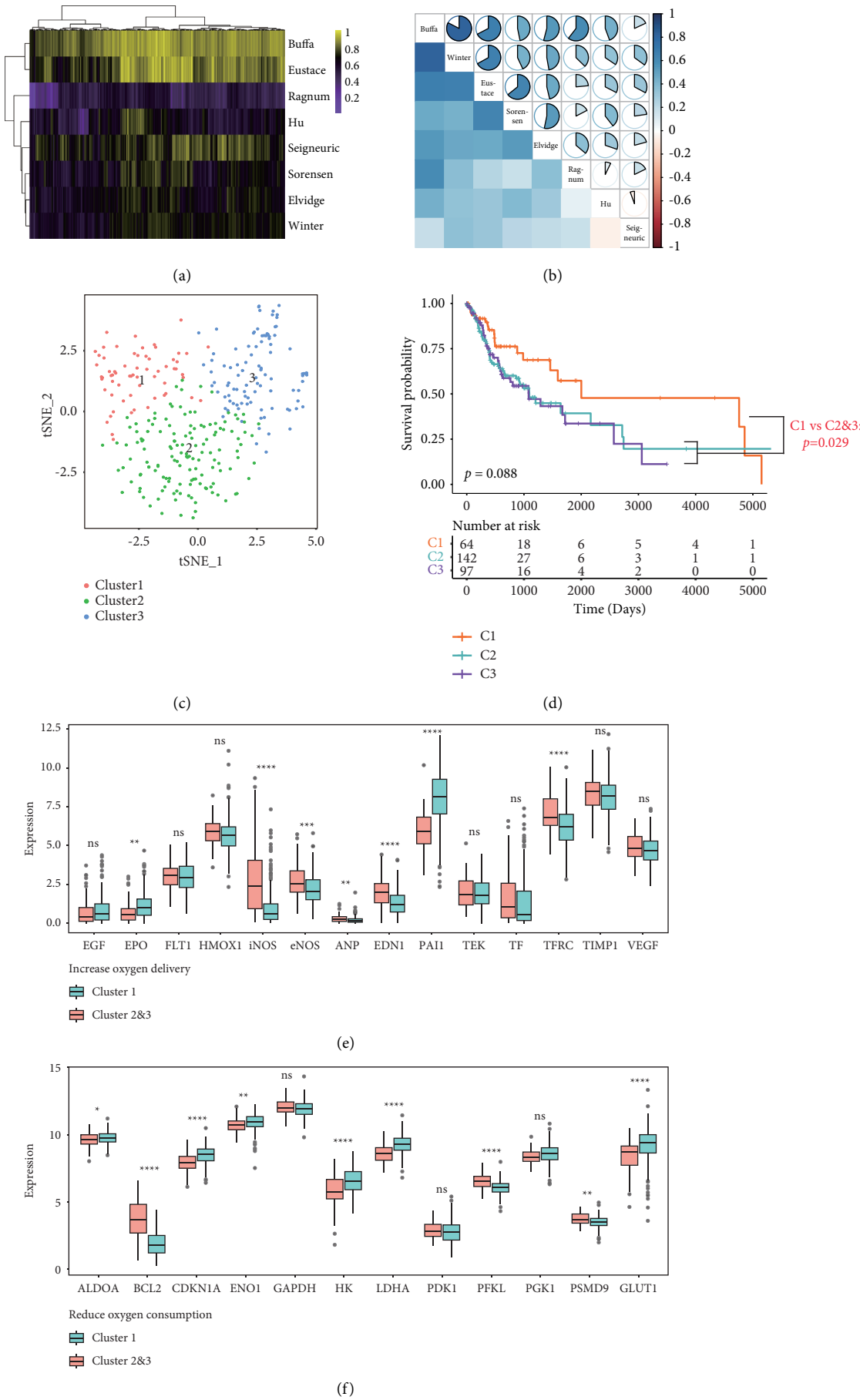
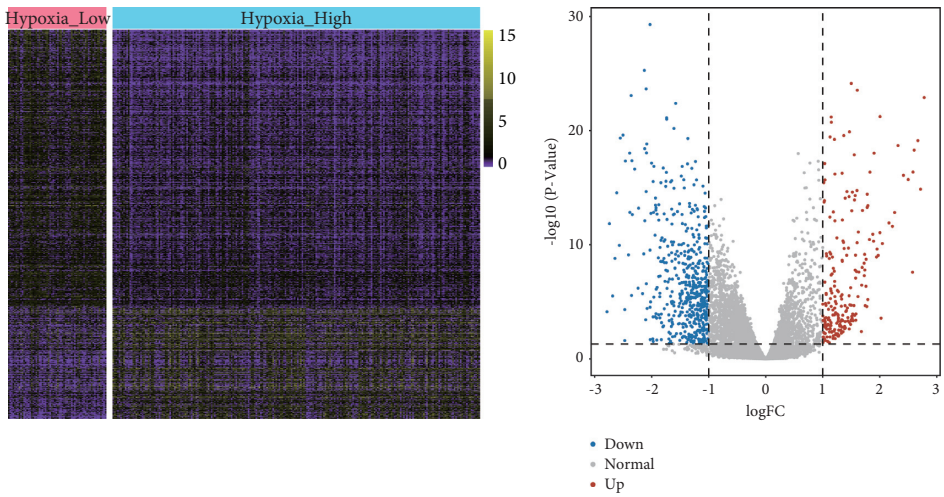
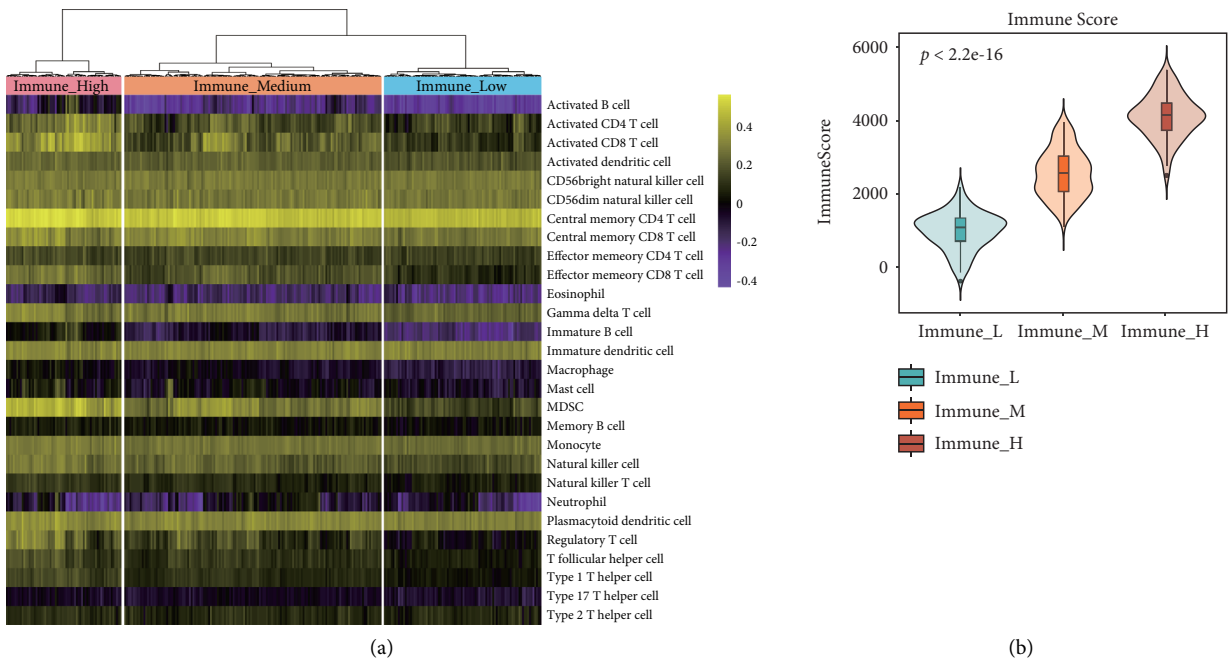



Figure 1: Continued.



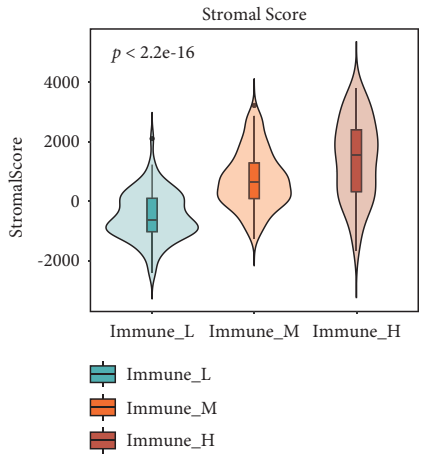

(g) (h)

Figure 1: Identification of hypoxia status and hypoxia-associated DEGs. Quantified heatmap (a) and correlation heatmap (b) calculated by ssGSEA with respect to the eight hypoxia signature sets. (c) Dot plot of three distinct clusters determined using the t-SNE algorithm to analyze 7 hypoxia signature gene sets. (d) Overall survival as shown by the Kaplan–Meier plot for patients in three clusters. (e), (f) HIF-1 KEGG pathway gene expression changes based on hypoxia status (Hypoxia_High versus Hypoxia_Low). Heatmap (g) and volcano plot (h) show the differentially expressed hypoxia-related genes in oral squamous cell carcinoma (OSCC). Red dots represent upregulated DEGs, blue dots represent downregulated DEGs, and gray dots represent genes with no differential expression.

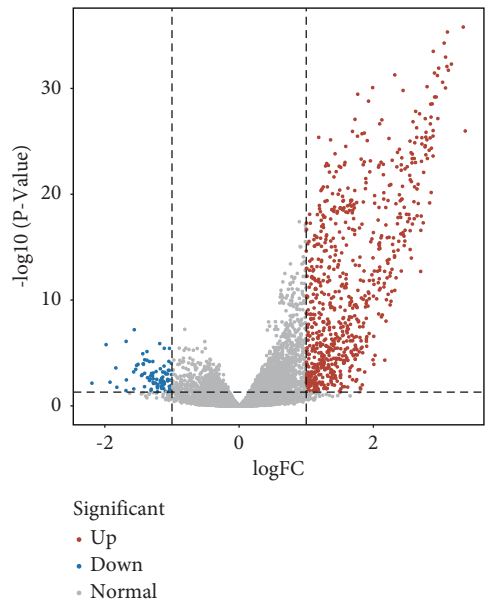



(a) (b)

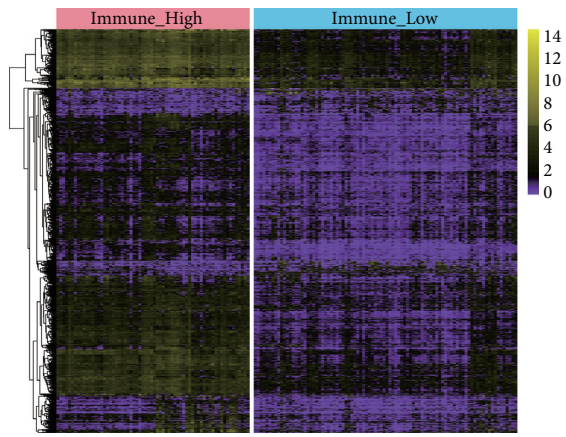
Figure 2: Continued.



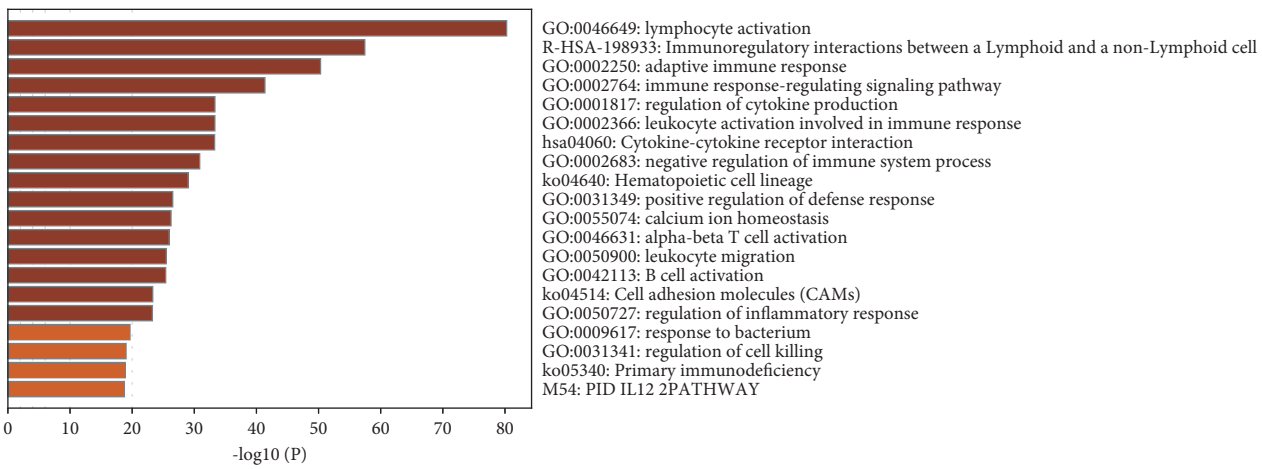

(c)




(d)




(e)




(f)

Figure 2: Identification of immune status and immune-associated DEGs. (a) Data from the TCGA database were classified into three subtypes (Immune_High group, Immune_Medium group, and Immune_Low group) based on immune status. Comparison of Immune_High and Immune_Low groups in terms of (b) immune score and (c) stromal score. (d) Volcano plot and (e) heatmap depicting DEGs between Immune_High and Immune_Low groups. Red dots represent upregulated DEGs, blue dots represent downregulated DEGs, and gray dots represent genes with no differential expression. Bar plot (f) of functional enrichment analysis of immune-related DEGs.

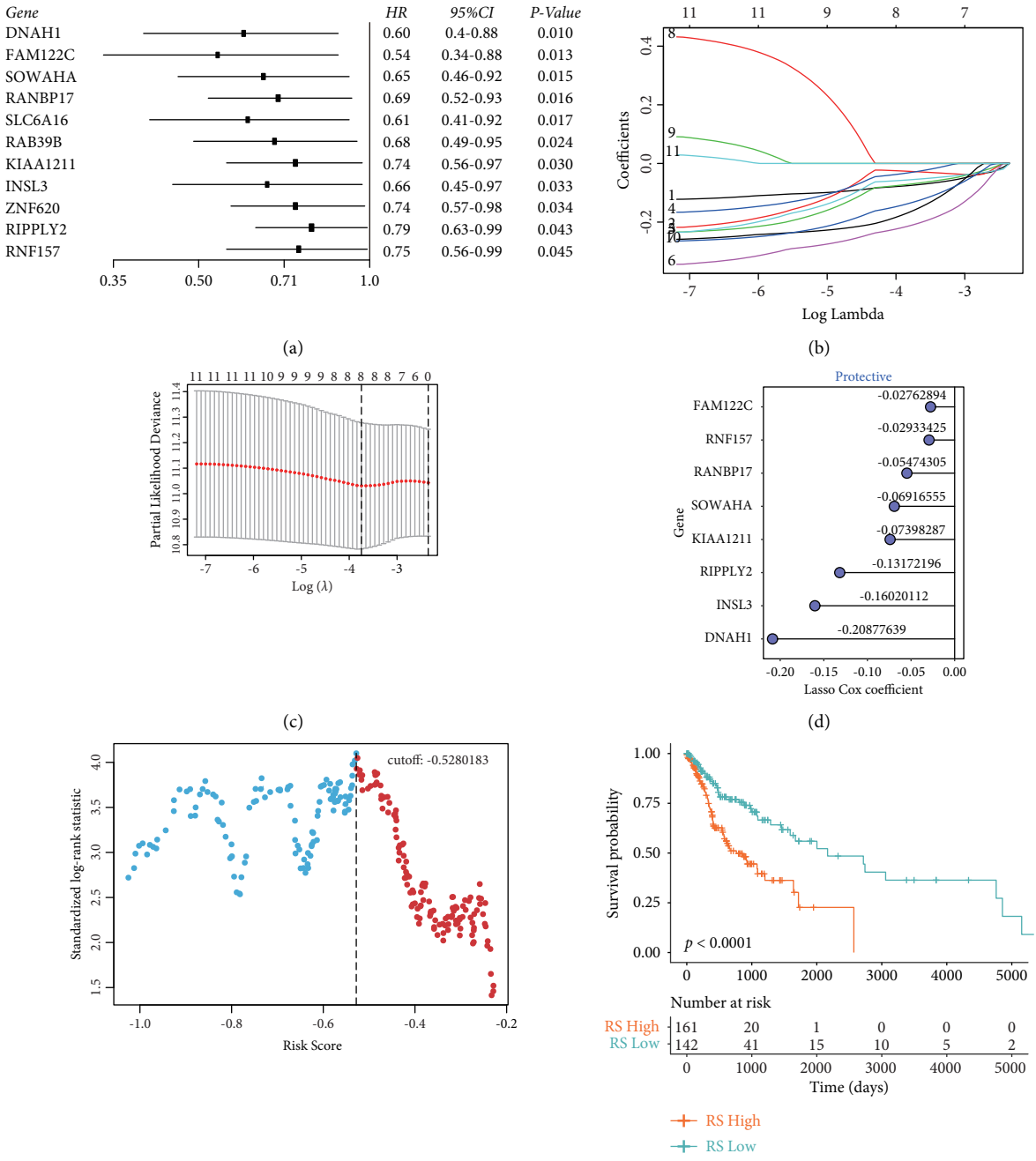
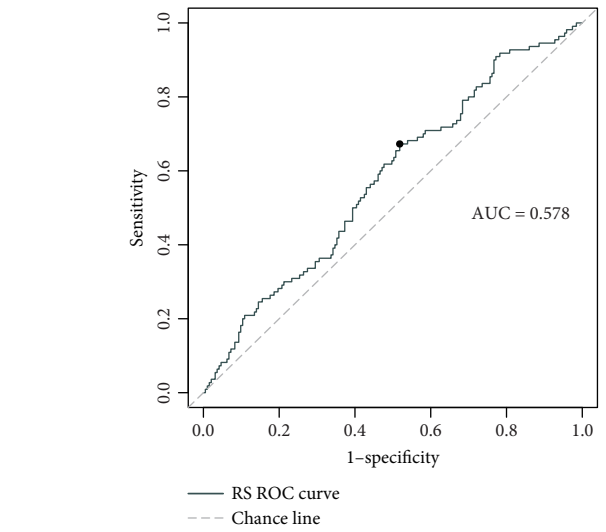
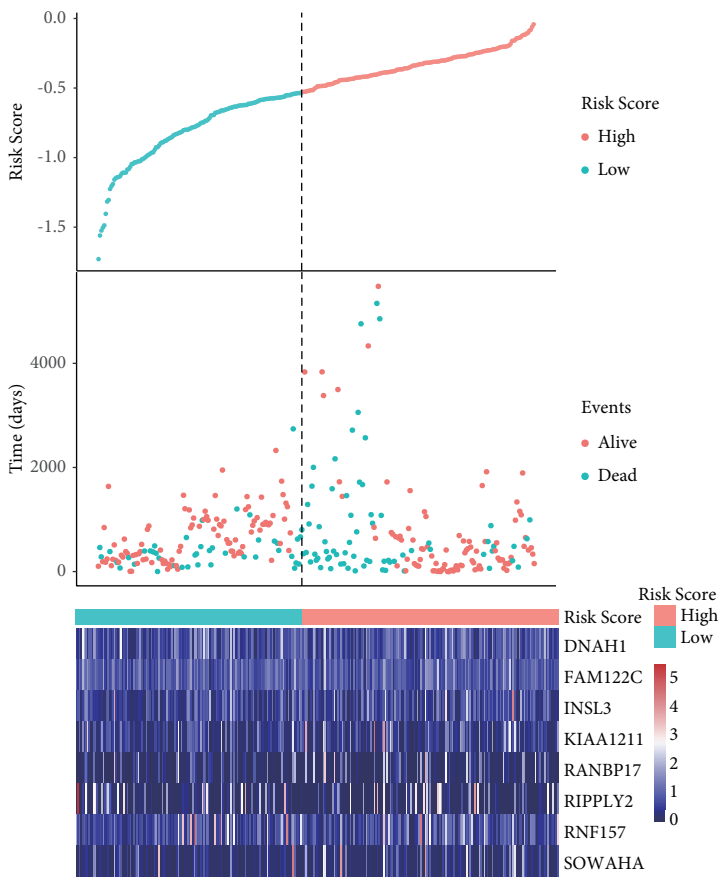



Figure 3: Continued.




(g)




(h)

Figure 3: The construction of hypoxia-immune-based mRNA prognosis model. (a) Hazard ratios for 11 hypoxia-immune-associated prognostic mRNAs. (b) LASSO coefficient profiles. (c) The LASSO model is used to choose the tuning parameter (lambda) using fivefold cross-validation based on minimum criteria for OS; average OS genes are represented by the upper $x$-axis, and log (lambda) is represented by the lower $x$-axis. Partial likelihood deviance error is represented by the $y$-axis. (d) An ensemble of 8 genes with their respective individual coefficients. (e) Scatter plot illustrates standardized log-rank statistic value for each corresponding cutoff of hypoxia-immune-based risk score. The optimal cutoff with the maximum standard log-rank statistic is marked with a vertical dashed line. (f) Kaplan–Meier plot of overall survival for patients in low-risk and high-risk groups by hypoxia-immune-based prognosis classifier in the discovery cohort. (g) The effectiveness of the risk scores is predicted using the ROC curve. (h) Expression profiles, survival status, and risk score distributions of the gene signature.

the Immune_ High group were elevated in contrast to those of the Immune_ Low group (Figures 2(b) and 2(c)). Comparisons between high- and low-immune groups revealed 854 immune-related DEGs (Figures 2(d) and 2(e)), which were subjected to Metascape tool functional enrichment analysis. The identified DEGs were predominately involved in immune response processes such as "lymphocyte activation," "adaptive immune response," "leukocyte activation involved immune response," and "immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell" (Figure 2(f)).

*3.3. Development and Validation of the Risk Score.* A total of 193 mRNAs were significant in both hypoxia- and immune-related genes. Eleven of these mRNAs were associated with OS ($P < 0.05$) according to univariate Cox regression analysis (Figure 3(a)). Of these, a total of 8 mRNA signatures were selected by the LASSO Cox regression model (Figures 3(b) and 3(c)). All the 8 mRNAs were protective against OSCC (Figure 3(d)). Subsequent risk scores were derived from the following equation: risk score = [DNAH1 expression * (−0.20877639)] + [FAM122C expression * (−0.02762894)] + [INSL3 expression * (−0.16020112)] + [KIAA1211 expression * (−0.07398287)] + [RANBP17 exp ression * (−0.05474305)] + [RIPPLY2 expression * (−0.13172196)] + [RNF157 expression * (−0.02933425)] + [SOWAHA expression * (−0.06916555)]. Using this formula, patients were stratified into either low- or high-risk groups. The method of maximally selected rank statistics was used to determine the optimal risk score cutoff (Figure 3(e)). Significantly improved survival was observed in those in the low-risk group compared to their high-risk counterparts (log-rank test, $p < 0.001$) (Figure 3(f)). The ROC curve showed that the classifier had relatively strong predictive power in the TCGA cohort, with a value of 0.578 for the area under the curve (AUC) (Figure 3(g)). Furthermore, the TCGA cohort was also analyzed for distribution of selected gene expressions and risk scores as well as survival status (Figure 3(h)). Table 1 depicts all relevant clinical parameters.

To further explore the efficacy of the constructed risk score, we performed validation tests using the GEO cohort. Using the method of maximally selected rank statistics, all OSCC patients in the GEO cohort were divided (Figure 4(a)). Comparison of survival showed that improved survival rates were experienced by low-risk patients in contrast to high-risk patients (log-rank test, $p = 0.00084$) (Figure 4(b)). In addition, we plotted the ROC curves for predicting one- and three-year survival according to the OS of patients, and the values of the area under the curve (AUC) were 0.552 and 0.605, respectively (Figure 4(c)). The GEO OSCC cohort was then analyzed for distribution of selected gene expressions and risk scores as well as survival status (Figures 4(d)–4(f)). Furthermore, we investigate the expression difference of eight signature genes between high and low groups. The result showed that the expression of the eight signature genes was significantly higher in the low-risk group than in the high-risk group in both TCGA and GEO cohorts (TCGA: Supplementary Figure 1A; GEO: Supplementary Figure 1B).

*3.4. Relationship between Prognostic mRNA Signature and Clinical Parameters.* We next investigate the relationship between the risk score and clinical parameters in the TCGA database. Only OSCC samples with complete clinical information (including patient tumor stage and grade (including N and T stages), genders, age, and the presence of perineural and lymphovascular invasion) were used. We found that, in addition to age (Figure 5(a)), perineural and lymphovascular invasion were found to correlate significantly with the hypoxia-immune mRNA signature, but not with patient tumor stage and grade (including N and T stages) and genders (Figures 5(b)–5(h)).

Both univariate and multivariate analyses were done on the TCGA dataset to further prove the significance of the constructed hypoxia- and immune-related gene signatures. The risk score, age, and pathological stage may be able to be combined to predict the prognosis of the TCGA OSCC cohort (Table 2).

Finally, we investigated whether our model could be used to determine survival outcomes in subgroups with different clinicopathological features. Supplementary Figures 2A–2I show that the risk score could be used to predict the prognosis of patients with different clinicopathological features.

*3.5. Variability of Degree of Immune Cell Infiltration between the Low- and High-Risk TCGA OSCC Cohorts.* The degree of immune cell infiltration between low- and high-risk groups was evaluated using the CIBERSORT algorithm. Those in the low-risk group were more likely to have higher infiltration degrees of CD8+ T cells, plasma cells, follicular helper T cells, regulatory T cells (Tregs), and naive B cells. On the other hand, high-risk patients demonstrated higher degrees of infiltration of activated DCs, activated mast cells, and neutrophils. Furthermore, the immune indicator scores and the immune checkpoint expressions in the low-risk group were all remarkably raised in comparison to those in the high-risk group patients (Figures 6(b)–6(g)).

To further assess the ability of our model to predict immunotherapy efficacy, the IMvigor210 cohort of MIBC patients treated with PD-L1 inhibitors was used. The Kaplan–Meier analysis showed that patients with high-risk scores had a poorer survival rate than those with low-risk scores (Supplementary Figure 3A, $p = 0.018$). ROC curve analyses showed that the risk score combined with tumor mutational burden (TMB) and tumor neoantigen burden (TNB) output a higher area under the curve (AUC) value (AUC = 0.699) than TMB (AUC = 0.659), TNB (AUC = 0.690), or risk score (AUC = 0.560), respectively (Supplementary Figure 3B). Furthermore, we further discussed the differences in immunotherapy response between the high-risk group and the low-risk group based on the immune signature, and the results found that patients in the low-risk group had a higher complete response (CR)/partial response (PR) rate than those in the high-risk group (Supplementary Figure 3C).

## 4. Discussion

OSCC is the most common oral cancer characterized by a higher recurrence rate and lower overall survival rate of patients [11]. Hypoxia has been found to play an important role in the prognosis and treatment of OSCC [28]. Hypoxia also influences the activity of immune cells in the tumor microenvironment [31]. For example, hypoxia can impair the maturation and activity of dendritic cells (DCs) and natural killer (NK) cells [31, 32]. Both hypoxia and immune status play important roles in OSCC [33–35]. However, there is a lack of a reliable prognostic model of integration of hypoxia- and immune-related signatures for OSCC. In this study, we have a comprehensive analysis and constructed a hypoxia-immune-integrated indicator in OSCC.

Based on hypoxia-related signatures, the samples in our study were grouped into 3 clusters (Cluster 1, Cluster 2, and Cluster 3) (Figures 1(a)–1(c)). The presence of cell hypoxia induces the expression of HIF-1$\alpha$. Enhanced HIF-1$\alpha$ expression has previously been associated with poor prognosis and lymph nodes metastasis in OSCC patients [36]. Further evaluation of overall patient survival resulted in the combination of Clusters 2 and 3 to form the Hypoxia_High group, while those in Cluster 1 were designated as the Hypoxia_Low group (Figure 1(d)). The expression of genes from two hypoxia-related gene sets, "increased oxygen delivery" and "reduced oxygen consumption," was compared between Cluster 1 and Clusters 2 and 3. Eleven of the 14 genes belonging to the "increased oxygen delivery" were expressed at higher levels in Clusters 2 and 3 in contrast to Cluster 1 (Figure 1(e)). On the other hand, 7 of the 12 genes associated with "reduced oxygen consumption" were highly expressed in Cluster 1 compared to Clusters 2 and 3 (Figure 1(f)). Therefore, Cluster 1 and Clusters 2 and 3 were considered as the Hypoxia_Low and Hypoxia_High groups, respectively. A total of 773 DEGs were identified across both these groups (Figures 1(e) and 1(f)).

We then divided these OSCC patients into three groups (Immune_High, Immune_Medium, and Immune_Low) according to the abundance of immune cells (Figure 2(a)). The immune score of Immune_High group was higher than that of the Immune_Medium and Immune_Low groups. The Immune_Low group has the lowest immune score among the three groups (Figures 2(b) and 2(c)). A total of 854 immune-related genes in OSCC were determined between the Immune_High and Immune_Low groups (Figures 2(e) and 2(f)). All genes were found to be enriched in immune-related functions, including "leukocyte activation involved immune response," "adaptive immune response," "lymphocyte activation," and "immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell" (Figure 2(f)). These immune processes are also related to the malignant properties of OSCC. Previous reports found that a higher neutrophil-to-lymphocyte ratio was a negative predictor for overall survival for patients with OSCC [37]. Dendritic cell immune response activation was able to be induced by IFN-$\gamma$-inhibited OSCC growth in tumor-bearing mice [38].

Table 1: Clinical characteristics of the TCGA OSCC patients.

| Characteristic | High ($n = 161$) | Low ($n = 142$) | TCGA ($n = 303$) |
|---|---|---|---|
| *Age* | | | |
| ≤62 | 80 | 81 | 161 |
| >62 | 81 | 61 | 142 |
| *Gender* | | | |
| Female | 51 | 48 | 99 |
| Male | 110 | 94 | 204 |
| *Survival status* | | | |
| Living | 93 | 100 | 193 |
| Dead | 68 | 42 | 110 |
| *Pathologic M* | | | |
| Unknown | 100 | 88 | 188 |
| M0 | 61 | 54 | 115 |
| *Pathologic N* | | | |
| Unknown | 27 | 17 | 44 |
| N0 | 59 | 54 | 113 |
| N1 | 27 | 19 | 46 |
| N2 | 47 | 51 | 98 |
| N3 | 1 | 1 | 2 |
| *Pathologic T* | | | |
| Unknown | 11 | 7 | 18 |
| T1 | 12 | 15 | 27 |
| T2 | 50 | 43 | 93 |
| T3 | 31 | 26 | 57 |
| T4 | 57 | 51 | 108 |
| *Tumor stage* | | | |
| Unknown | 13 | 10 | 23 |
| Stage I | 9 | 8 | 17 |
| Stage II | 27 | 24 | 51 |
| Stage III | 33 | 21 | 54 |
| Stage IV | 79 | 79 | 158 |

To further integrate the hypoxia- and immune-related genes, the overlapping genes between the two gene sets were screened. Eight prominent mRNA signatures associated with OS from the 193 overlapping genes were selected, which were FAM122C, RNF157, RANBP17, SOWAHA, KIAA1211, RIPPLY2, INSL3, and DNAH1 (Figures 3(a)–3(d)). Among these mRNAs, KIAA1211 is known to be an oncogenic gene. Non-small-cell lung cancer tissues were found to have raised KIAA1211 expressions in contrast to adjacent normal tissues. Knockdown of KIAA1211 inhibited the proliferative abilities of NSCLC cells while promoting apoptosis both *in vitro* and *in vivo* [39]. Small cell lung cancer patients with a KIAA1211 mutation possess a longer survival period than those with wild-type KIAA1211 mutations [40]. RIPPLY2 represented one of the mRNAs in a five-gene signature verified to be able to predict the survival of endometrial cancer patients [41]. The tumor-promoting effect of INSL3 in cancer has also been widely addressed. The plasma level of INSL3 was found to be raised in an individual with metastatic ovarian cancer [42]. Other reports highlight the potential role of INSL3 as a marker of human testicular Leydig cell tumors [43]. INSL3 could promote tumor growth and angiogenesis in nude mice model of thyroid cancer in a manner that appeared to be related to the action of RXFP2 and the secretion of S100A4 and (pro-) cathepsin-L [44]. In pancreatic cancer patients, a higher serum level of INSL3 was

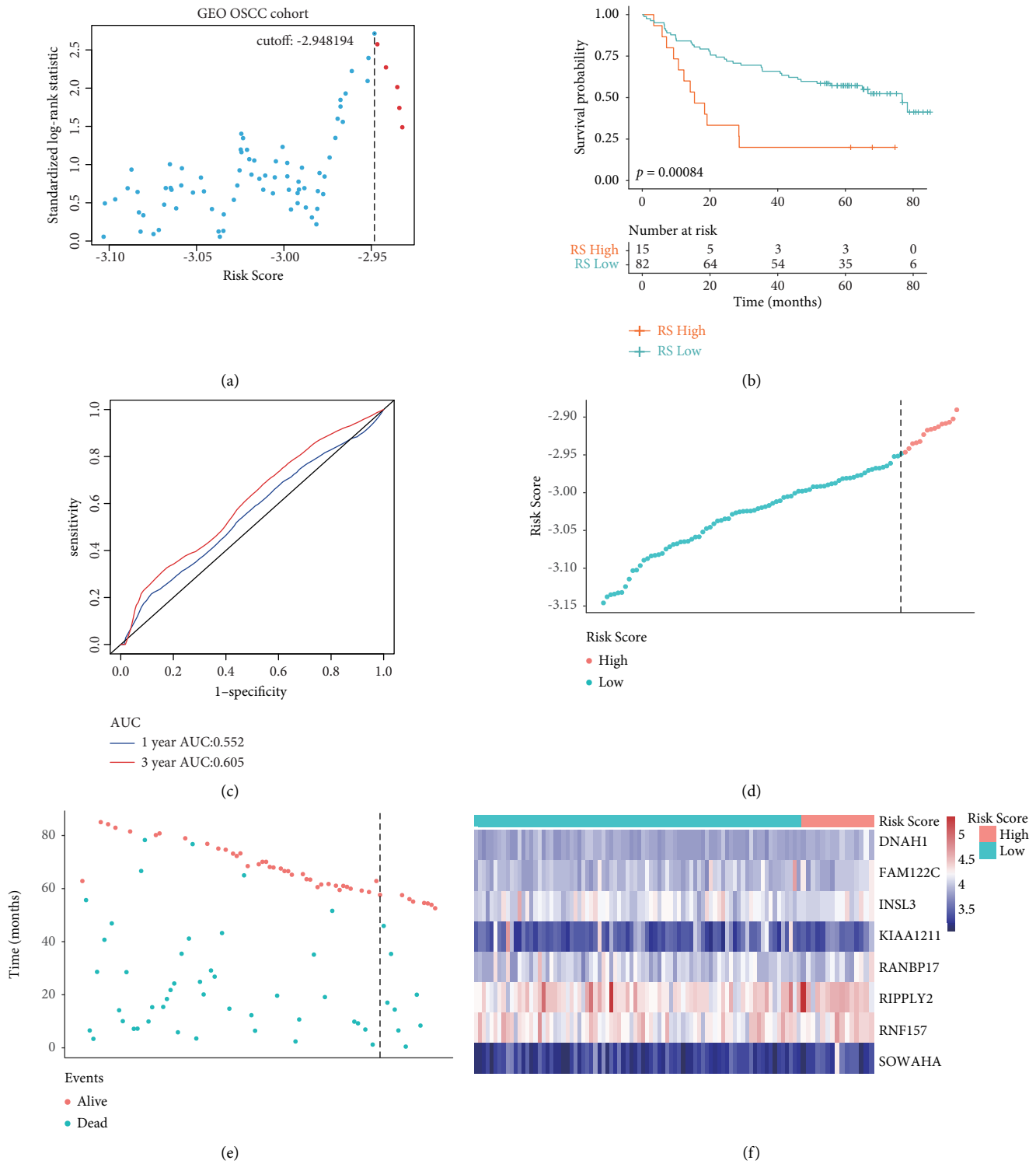



Figure 4: The GEO cohort was used to verify the hypoxia-immune-based prognosis classifier. (a) The scatter plot demonstrates the standardized log-rank statistic value for each corresponding cutoff of hypoxia-immune-based risk score. Vertical dashed lines represent the optimal cutoff with the maximum standard log-rank statistic. (b) The hypoxia-immune-based prognosis classifier is used to label patients from the GEO cohort as either low or high risk, followed by the construction of an overall survival Kaplan–Meier plot. (c) One-year and 3-year survival rates of OSCC patients in the GEO cohort as shown using ROC curves. (d) Distributions of risk score, (e) survival status, and (f) expression profile of signature genes.

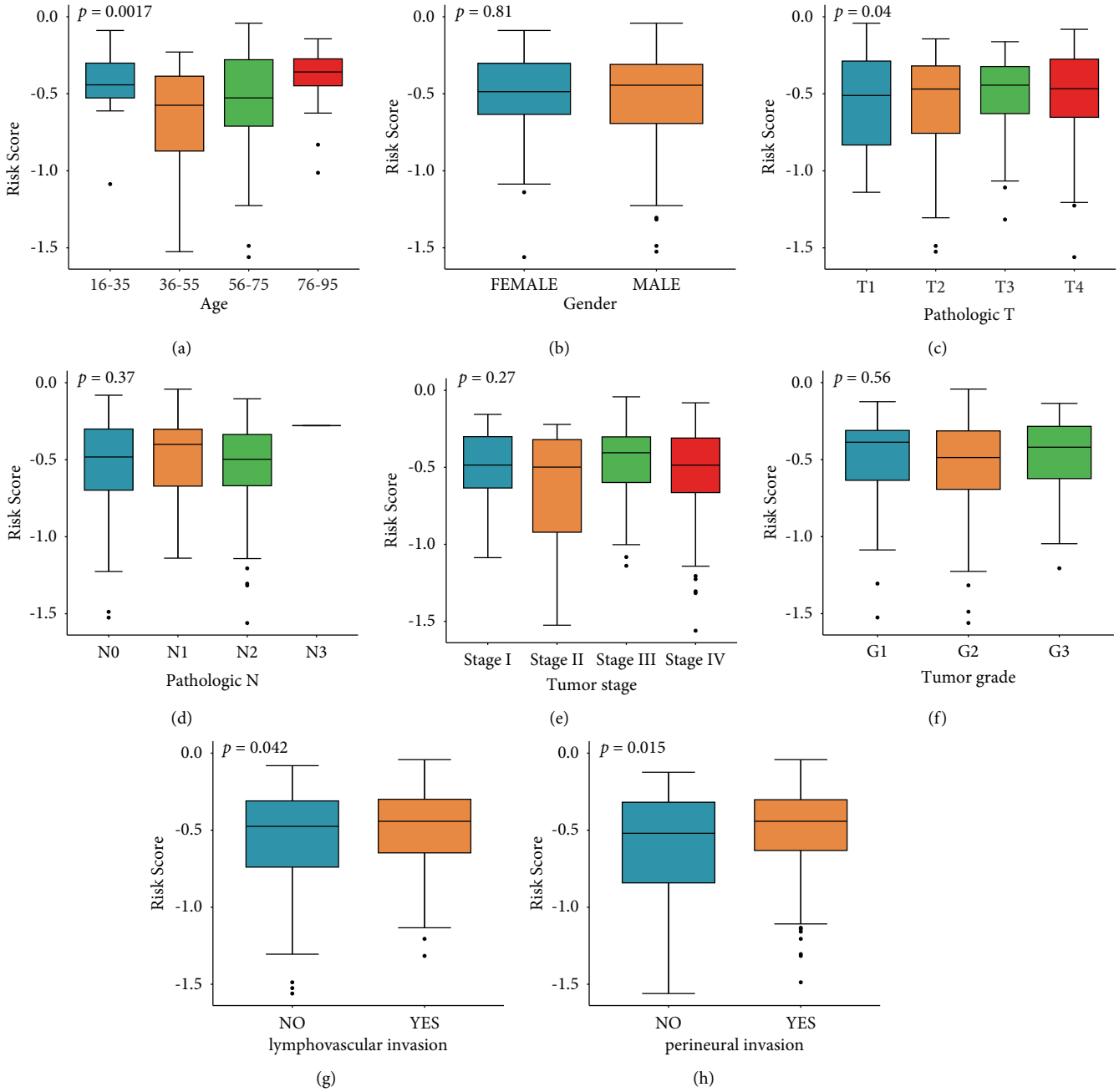
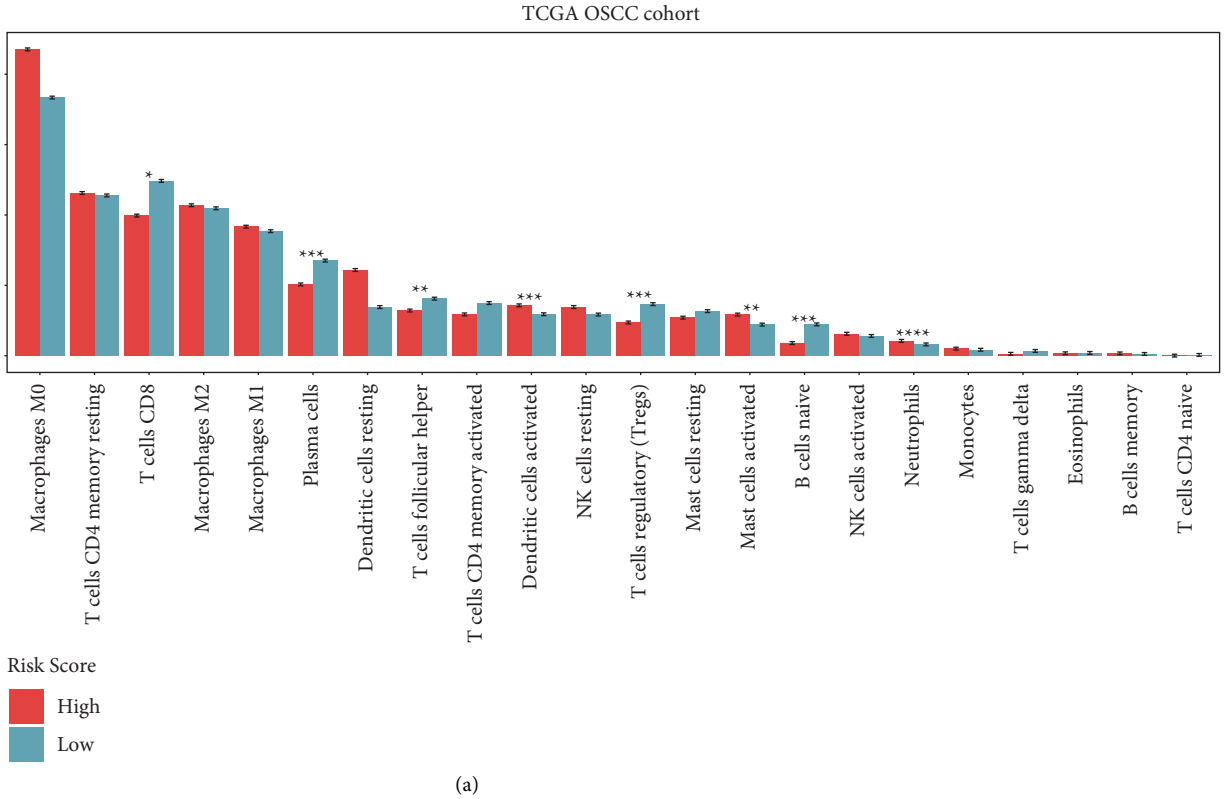



Figure 5: The TCGA cohort was used to determine the association between clinical parameters and the gene signature. Significant association was only found with (a) age, but not with (b) gender, (c) T stage, (d) N stage, (e) tumor stage, (f) tumor grade, (g) lymphovascular invasion, and (h) perineural invasion.

associated with increased anorexia [45]. These mRNAs, which have previously been found to be of significant value in other cancers, should be further investigated for their role in OSCC.

These 8 mRNAs were used to construct an mRNA signature, which may have prognostic potential in OSCC (Figures 3(e)–3(h)). This constructed risk score was then validated in a cohort from the GEO dataset (Figures 4(a)–4(f)). The risk score showed a significant association only with age and perineural and lymphovascular invasion, but not with other features, including gender, T stage, N stage, and tumor stage (Figures 5(a)–5(h)). Therefore, this established gene signature may be an independent prognostic indicator.

Cancer immunotherapy has obtained much attention in recent years and is considered as a direction of tumor therapy, such as checkpoint blocking therapy. The specific difference in the tumor immune environment can reflect the heterogeneity of clinical samples in response to current immunotherapy. Finally, the immune profile variability




(a)

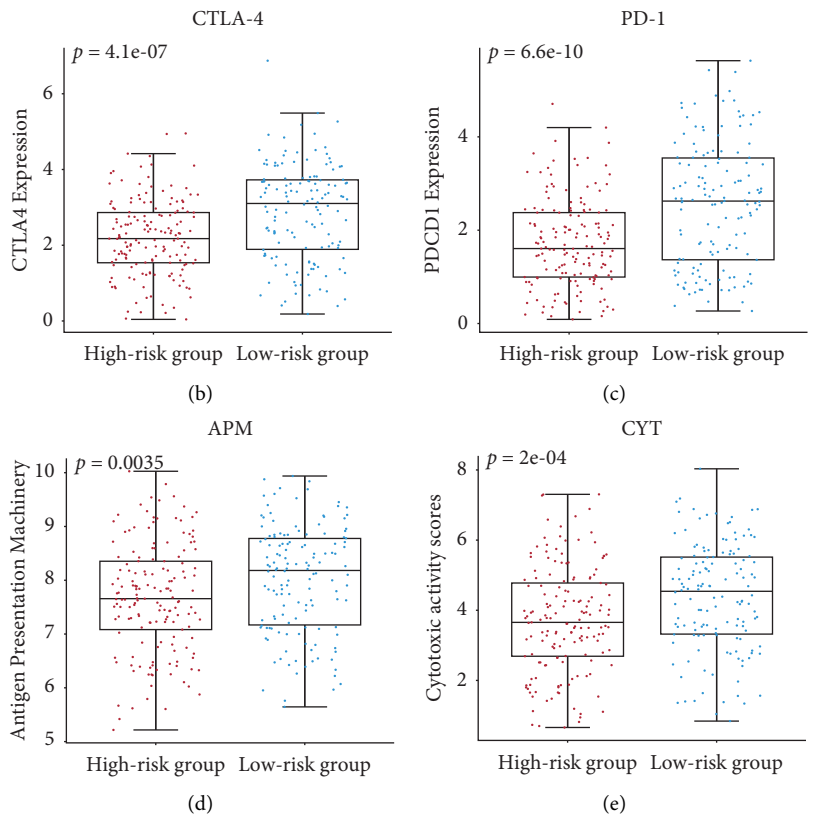



Figure 6: Continued.

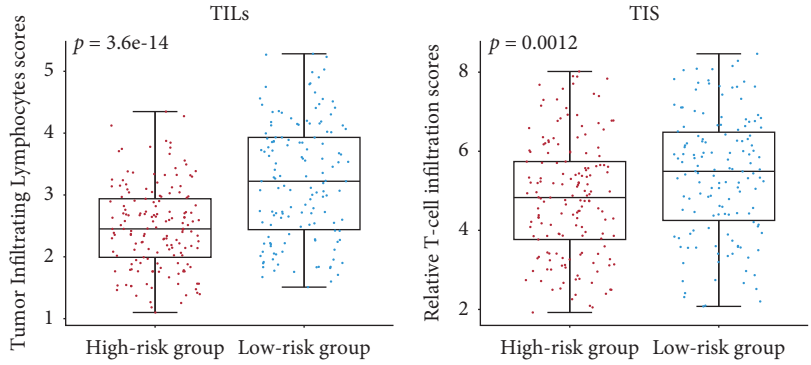



Figure 6: Analysis of immune status variability of the high- and low-risk groups in the TCGA OSCC cohort. (a) The bar plot shows various immune cell infiltration between the high- and low-risk patients. (b-g) Box plots illustrating markedly different immune checkpoints between the high- and low-risk groups.

Table 2: Univariate and multivariate Cox regression analysis of clinicopathological features associated with overall survival in TCGA data.

| Variables | Univariate analysis | | Multivariate analysis | |
|---|---|---|---|---|
| | HR (95% CI) | *P*-value | HR (95% CI) | *P*-value |
| Age ( > 62/≤62) | 1.021 (1.005–1.037) | 0.00942* | 1.0329 (1.0103–1.056) | 0.00413* |
| Gender (male/female) | 1.021 (0.6894–1.511) | 0.919 | 0.7058 (0.4183–1.191) | 0.19194 |
| Tumor grade (G4/G3/G2/G1) | 1.35 (1.01–1.803) | 0.0424* | 1.4226 (0.9819–2.061) | 0.06242 |
| Stage (IV/III/II/I) | 1.482 (1.172–1.874) | 0.00102* | 1.7513 (1.0060–3.049) | 0.04758* |
| Pathologic T (T4/T3/T2/T1) | 1.447 (1.181–1.773) | 0.000365* | 1.0449 (0.7235–1.509) | 0.81483 |
| Pathologic N (N3/N2/N1/N0) | 1.411 (1.116–1.785) | 0.00403* | 1.1601 (0.8620–1.561) | 0.32721 |
| Risk score | 4.322 (2.096–8.912) | 0.0000736* | 3.5246 (1.4818–8.383) | 0.00438* |

*Note.* *Statistically significant. HR: hazard ratio; CI: confidence interval.

between the high- and low-risk groups was compared in the TCGA OSCC cohort. High-risk group samples showed different degree of immune cell infiltration (Figure 6(a)). The low-risk group showed higher immune status than the low-risk group, as evidenced by a higher amount of immune checkpoint expression, including CTLA-4, PD-1, APM, CYT, and TILs (Figures 6(b)–6(g)). These checkpoints are indictors of OSCC risk and may function as therapeutic targets in OSCC. The genetic variants of CTLA-4 were associated with tobacco-related OSCC risk in the North Indian population [46]. A high number of CTLA-4 cells is associated with poor 5-year metastasis-free survival of OSCC patients [47]. An anti-PD-1 antibody is an agent which may prevent the initiation and progression of OSCC while prolonging patient survival time [48–50]. Analysis of the IMvigor210 cohort revealed that patients with low-risk scores had better survival and tended to have a higher complete response (CR)/partial response (PR) rate.

In conclusion, hypoxia and immune status play a key role in the prognosis of OSCC. Combining hypoxia- and immune-related genes, we established an OSCC prognostic model based on hypoxia and immunity, which provides a reliable reference for clinical decision-making.

A preprint has previously been published on the preprint website (https://www.researchsquare.com/article/rs-596220/v1) [51].

## Data Availability

The data used are obtained from TCGA Data Portal: https://portal.gdc.cancer.gov/, GEO Datasets: https://www.ncbi.nlm.nih.gov/gds/, and Metascape: https://metascape.org/.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Authors' Contributions

Jichen Li and Shaohua Lv designed the study. Shaohua Lv and Zhipeng Qian collected data. Zhipeng Qian and Songlin Piao developed the computational model and analyzed it. Shaohua Lv and Jianhao Li wrote the paper. All authors reviewed and approved this paper.

## Acknowledgments

This research was supported by the foundation of the National Conference on Undergraduate Innovation and Entrepreneurship (no. S202110226059).

## Supplementary Materials

*Supplementary Figure 1.* The expression of the eight signature genes. (A) Differential expression profiles of the 8

signature genes in low and high-risk groups. (B) Differential expression profiles of the 8 signature genes in low and high-risk groups. *Supplementary Figure 2.* Kaplan–Meier analyses of the eight-gene-based model in subgroups with different clinicopathological features. Kaplan–Meier survival curves for (A) age >62, (B) age ≤62, (C) male, (D) female, (E) T1 + T2, (F) T3 + T4, (G) N0 + N1, (H) N2 + N3, (I) stage I + stage II, (J) stage III + stage IV, (K) G1 + G2, and (L) G3 for the low- and high-risk groups. T: stage-T, N: stage-N, M: stage-M, and G: tumor grade. *Supplementary Figure 3*. The risk model predicts immunotherapy efficacy. (A) Kaplan–Meier overall survival curves for patients assigned to low- and high-risk groups in the IMvigor210 cohort. (B) ROC curves of tumor mutational burden (TMB), tumor neoantigen burden (TNB), risk score, and the combination (TMB, TNB, and risk score). (C) Rate of clinical response (CR/PR and SD/PD) to immunotherapy in the low- or high-risk groups (CR, complete response; PR, partial response; PD, progressive disease; SD, stable disease)." (*Supplementary Materials*)

## References


[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA: A Cancer Journal for Clinicians*, vol. 70, no. 1, pp. 7–30, 2020.

[3] A. W. Eckert, M. H. W. Lautner, A. Schütze, H. Taubert, J. Schubert, and U. Bilkenroth, "Coexpression of hypoxia-inducible factor-1$\alpha$ and glucose transporter-1 is associated with poor prognosis in oral squamous cell carcinoma patients," *Histopathology*, vol. 58, no. 7, pp. 1136–1147, 2011.

[4] A. W. Eckert, M. Kappler, J. Schubert, and H. Taubert, "Correlation of expression of hypoxia-related proteins with prognosis in oral squamous cell carcinoma patients," *Oral and Maxillofacial Surgery*, vol. 16, no. 2, pp. 189–196, 2012.

[5] M. E. Hubbi and G. L. Semenza, "Regulation of cell proliferation by hypoxia-inducible factors," *American Journal of Physiology-Cell Physiology*, vol. 309, no. 12, pp. C775–C782, 2015.

[6] P. L. B. Domingos, M. G. Souza, T. A. Guimarães et al., "Hypoxia reduces the E-cadherin expression and increases OSCC cell migration regardless of the E-cadherin methylation profile," *Pathology, Research & Practice*, vol. 213, no. 5, pp. 496–501, 2017.

[7] G. Zhu, S. Wang, J. Chen et al., "Long noncoding RNA HAS2-AS1 mediates hypoxia-induced invasiveness of oral squamous cell carcinoma," *Molecular Carcinogenesis*, vol. 56, no. 10, pp. 2210–2222, 2017.

[8] T. Ishikawa, K. Nakashiro, S. K. Klosek et al., "Hypoxia enhances CXCR4 expression by activating HIF-1 in oral squamous cell carcinoma," *Oncology Reports*, vol. 21, pp. 707–712, 2009.

[9] J. P. Joseph, M. K. Harishankar, A. A. Pillai, and A. Devi, "Hypoxia induced EMT: a review on the mechanism of tumor progression and metastasis in OSCC," *Oral Oncology*, vol. 80, pp. 23–32, 2018.

[10] X. Lv, J. Li, C. Zhang et al., "The role of hypoxia-inducible factors in tumor angiogenesis and cell metabolism," *Genes & Diseases*, vol. 4, no. 1, pp. 19–24, 2017.

[11] X. Zhang, S. Han, H.-Y. Han et al., "Risk prediction for malignant conversion of oral epithelial dysplasia by hypoxia related protein expression," *Pathology*, vol. 45, no. 5, pp. 478–483, 2013.

[12] S.-L. Jian, H.-Y. Hsieh, C.-T. Liao et al., "G$\alpha$12 drives invasion of oral squamous cell carcinoma through up-regulation of proinflammatory cytokines," *PLoS One*, vol. 8, no. 6, Article ID e66133, 2013.

[13] M.-x. Cao, W.-l. Zhang, X.-h. Yu et al., "Retracted article: interplay between cancer cells and M2 macrophages is necessary for miR-550a-3-5p down-regulation-mediated HPV-positive OSCC progression," *Journal of Experimental & Clinical Cancer Research*, vol. 39, no. 1, p. 102, 2020.

[14] P. Diao, Y. Wu, J. Li et al., "Preoperative systemic immune-inflammation index predicts prognosis of patients with oral squamous cell carcinoma after curative resection," *Journal of Translational Medicine*, vol. 16, no. 1, p. 365, 2018.

[15] T. S. de Souza Malaspina, T. H. Gasparoto, M. R. S. N. Costa et al., "Enhanced programmed death 1 (PD-1) and PD-1 ligand (PD-L1) expression in patients with actinic cheilitis and oral squamous cell carcinoma," *Cancer Immunology, Immunotherapy*, vol. 60, no. 7, pp. 965–974, 2011.

[16] K. Dave, A. Ali, and M. Magalhaes, "Increased expression of PD-1 and PD-L1 in oral lesions progressing to oral squamous cell carcinoma: a pilot study," *Scientific Reports*, vol. 10, no. 1, p. 9705, 2020.

[17] N. Kondoh, M. Mizuno-Kamiya, N. Umemura et al., "Immunomodulatory aspects in the progression and treatment of oral malignancy," *Japanese Dental Science Review*, vol. 55, no. 1, pp. 113–120, 2019.

[18] K. Meehan, C. Leslie, M. Lucas et al., "Characterization of the immune profile of oral tongue squamous cell carcinomas with advancing disease," *Cancer Medicine*, vol. 9, no. 13, pp. 4791–4807, 2020.

[19] M. Sieviläinen, R. Almahmoudi, A. Al-Samadi, T. Salo, M. Pirinen, and A. Almangush, "The prognostic value of immune checkpoints in oral squamous cell carcinoma," *Oral Diseases*, vol. 25, no. 6, pp. 1435–1445, 2019.

[20] G.-z. Huang, Q.-q. Wu, Z.-n. Zheng et al., "Bioinformatics analyses indicate that cathepsin G (CTSG) is a potential immune-related biomarker in oral squamous cell carcinoma (OSCC)," *OncoTargets and Therapy*, vol. 14, pp. 1275–1289, 2021.

[21] L. Li, B. Cao, X. Liang et al., "Microenvironmental oxygen pressure orchestrates an anti- and pro-tumoral $\gamma\delta$ T cell equilibrium via tumor-derived exosomes," *Oncogene*, vol. 38, no. 15, pp. 2830–2843, 2019.

[22] S. Zheng, Y. Zou, J. y. Liang et al., "Identification and validation of a combined hypoxia and immune index for triple-negative breast cancer," *Molecular Oncology*, vol. 14, no. 11, pp. 2814–2833, 2020.

[23] Y. Liu, J. Wu, W. Huang et al., "Development and validation of a hypoxia-immune-based microenvironment gene signature for risk stratification in gastric cancer," *Journal of Translational Medicine*, vol. 18, no. 1, p. 201, 2020.

[24] Z. Zhang, S. Bao, C. Yan, P. Hou, M. Zhou, and J. Sun, "Computational principles and practice for decoding immune contexture in the tumor microenvironment," *Briefings in Bioinformatics*, vol. 22, no. 3, 2021.

[25] Z. Zhang, C. Yan, K. Li et al., "Pan-cancer characterization of lncRNA modifiers of immune microenvironment reveals

clinically distinct de novo tumor subtypes," *Npj Genomic Medicine*, vol. 6, no. 1, p. 52, 2021.

[26] H. Zhao, S. Gu, S. Bao et al., "Mechanistically derived patient-level framework for precision medicine identifies a personalized immune prognostic signature in high-grade serous ovarian cancer," *Briefings in Bioinformatics*, vol. 22, 2021.

[27] J.-C. Guo, Y. Wu, Y. Chen et al., "Protein-coding genes combined with long noncoding RNA as a novel transcriptome molecular staging model to predict the survival of patients with esophageal squamous cell carcinoma," *Cancer Communications*, vol. 38, no. 1, p. 4, 2018.

[28] Y. Han, X. Wang, K. Xia, and T. Su, "A novel defined hypoxia-related gene signature to predict the prognosis of oral squamous cell carcinoma," *Annals of Translational Medicine*, vol. 9, no. 20, p. 1565, 2021.

[29] Z. Qian, D. Shang, L. Fan et al., "Heterogeneity analysis of the immune microenvironment in laryngeal carcinoma revealed potential prognostic biomarkers," *Human Molecular Genetics*, vol. 332, 2021.

[30] V. Bhandari, C. Hoey, L. Y. Liu et al., "Molecular landmarks of tumor hypoxia across cancer types," *Nature Genetics*, vol. 51, no. 2, pp. 308–318, 2019.

[31] K. Mortezaee and J. Majidpoor, "The impact of hypoxia on immune state in cancer," *Life Sciences*, vol. 286, Article ID 120057, 2021.

[32] M. H. Bao and C. C. Wong, "Hypoxia, metabolic reprogramming, and drug resistance in liver cancer," *Cells*, vol. 10, no. 7, 2021.

[33] L. Gao, Z.-C. Dou, W.-H. Ren, S.-M. Li, X. Liang, and K.-Q. Zhi, "CircCDR1as upregulates autophagy under hypoxia to promote tumor cell survival via AKT/ERK½/mTOR signaling pathways in oral squamous cell carcinomas," *Cell Death & Disease*, vol. 10, no. 10, p. 745, 2019.

[34] W. Dong, Y. Chen, N. Qian et al., "SATB2 knockdown decreases hypoxia-induced autophagy and stemness in oral squamous cell carcinoma," *Oncology Letters*, vol. 20, no. 1, pp. 794–802, 2020.

[35] H. Quan, Z. Shan, Z. Liu et al., "The repertoire of tumor-infiltrating lymphocytes within the microenvironment of oral squamous cell carcinoma reveals immune dysfunction," *Cancer Immunology, Immunotherapy*, vol. 69, no. 3, pp. 465–476, 2020.

[36] M. Uehara, K. Sano, H. Ikeda, M. Nonaka, and I. Asahina, "Hypoxia-inducible factor 1 alpha in oral squamous cell carcinoma and its relation to prognosis," *Oral Oncology*, vol. 45, no. 3, pp. 241–246, 2009.

[37] Y. Wang, P. Wang, O. Andrukhov et al., "Meta-analysis of the prognostic value of the neutrophil-to-lymphocyte ratio in oral squamous cell carcinoma," *Journal of Oral Pathology & Medicine*, vol. 47, no. 4, pp. 353–358, 2018.

[38] H. Wei, P. Hongya, J. Linlin et al., "IFN-$\gamma$ enhances the anti-tumour immune response of dendritic cells against oral squamous cell carcinoma," *Archives of Oral Biology*, vol. 56, no. 9, pp. 891–898, 2011.

[39] Z. Liu, H. Cao, Y. Shi, and R. Yang, "KIAA1211 plays an oncogenic role in human non-small cell lung cancer," *Journal of Cancer*, vol. 10, no. 26, pp. 6747–6753, 2019.

[40] L. Wang, T. Luan, S. Zhou et al., "LncRNA HCP5 promotes triple negative breast cancer progression as a ceRNA to regulate BIRC3 by sponging miR-219a-5p," *Cancer Med*, vol. 8, no. 9, 2019.

[41] X. Li, F. Yin, Y. Fan et al., "Establishment and validation of a prognostic nomogram based on a novel five-DNA methylation signature for survival in endometrial cancer patients," *Cancer Medicine*, vol. 10, no. 2, pp. 693–708, 2021.

[42] M. Rossato, M. Barban, and R. Vettor, "Elevated plasma levels of the novel hormone INSL3 in a woman with metastatic ovarian cancer," *The International Journal of Biological Markers*, vol. 22, no. 2, pp. 159-160, 2007.

[43] M. Rossato, I. M. Tavolini, A. Calcagno, M. Gardiman, F. Dal Moro, and W. Artibani, "The novel hormone INSL3 is expressed in human testicular Leydig cell tumors: a clinical and immunohistochemical study," *Urologic Oncology: Seminars and Original Investigations*, vol. 29, no. 1, pp. 33–37, 2011.

[44] S. Hombach-Klonisch, J. Bialek, Y. Radestock et al., "INSL3 has tumor-promoting activity in thyroid cancer," *International Journal of Cancer*, vol. 127, no. 3, pp. 521–531, 2010.

[45] E. Yeom, H. Shin, W. Yoo et al., "Tumour-derived Dilp8/INSL3 induces cancer anorexia by regulating feeding neuropeptides via Lgr3/8 in the brain," *Nature Cell Biology*, vol. 23, no. 2, pp. 172–183, 2021.

[46] V. Bharti, B. K. Mohanti, and S. N. Das, "Functional genetic variants of CTLA-4 and risk of tobacco-related oral carcinoma in high-risk North Indian population," *Human Immunology*, vol. 74, no. 3, pp. 348–352, 2013.

[47] K. Koike, H. Dehari, K. Ogi et al., "Prognostic value of FoxP3 and CTLA-4 expression in patients with oral squamous cell carcinoma," *PLoS One*, vol. 15, no. 8, Article ID e0237465, 2020.

[48] R. L. Ferris, G. Blumenschein Jr., J. Fayette et al., "Nivolumab for recurrent squamous-cell carcinoma of the head and neck," *New England Journal of Medicine*, vol. 375, no. 19, pp. 1856–1867, 2016.

[49] J. Wang, T. Xie, B. Wang et al., "PD-1 blockade prevents the development and progression of carcinogen-induced oral premalignant lesions," *Cancer Prevention Research*, vol. 10, no. 12, pp. 684–693, 2017.

[50] J. A. Monteiro de Oliveira Novaes, T. Hirz, I. Guijarro et al., "Targeting of CD40 and PD-L1 pathways inhibits progression of oral premalignant lesions in a carcinogen-induced model of oral squamous cell carcinoma," *Cancer Prevention Research*, vol. 14, no. 3, 2020.

[51] J. L. Shaohua Lv, S. Piao, and J. Li, "Identification and validation of a hypoxia-immune-based prognostic mRNA signature for oral squamous cell carcinoma," inpress, 2021.

Hindawi
Journal of Oncology
Volume 2022, Article ID 4656661, 12 pages
https://doi.org/10.1155/2022/4656661

*Research Article*

# Identification of LTF as a Prognostic Biomarker for Osteosarcoma

**Xiaoqi Liu**,[1] **Zengqiang Wang**,[2] **Meijiao Liu**,[3] **Fengnan Zhi**,[4] **Pengpeng Wang**,[5] **Xingyu Liu**,[6] **Shanxiao Yu**,[7] **Bing Liu**,[3,8] **and Yanan Jiang**[4,8]

[1]*Department of Orthopedic Surgery, The Second Affiliated Hospital, Harbin Medical University, Harbin, China*
[2]*Department of Pharmacy, Anqiu People's Hospital, Anqiu, China*
[3]*Department of Oral and Maxillofacial Surgery, The First Affiliated Hospital, Harbin Medical University, Harbin, China*
[4]*Department of Pharmacology (State-Province Key Laboratories of Biomedicine-Pharmaceutics of China, Key Laboratory of Cardiovascular Research, Ministry of Education), College of Pharmacy, Harbin Medical University, Harbin, China*
[5]*Continuing Education Office, The Second Affiliated Hospital, Harbin Medical University, Harbin, China*
[6]*Academic Affairs Office, The Second Affiliated Hospital, Harbin Medical University, Harbin, China*
[7]*College of Humanities and Social Sciences, Harbin Medical University, Harbin, China*
[8]*Translational Medicine Research and Cooperation Center of Northern China, Heilongjiang Academy of Medical Sciences, Harbin, China*

Correspondence should be addressed to Bing Liu; liubingdoctor123@126.com and Yanan Jiang; jiangyanan@hrbmu.edu.cn

Received 30 September 2021; Revised 17 December 2021; Accepted 23 December 2021; Published 21 January 2022

Academic Editor: Dechao Bu



Osteosarcoma remains a major health problem in teenagers. However, its pathogenesis mechanism remains not fully elucidated. This study aims to identify the prognostic biomarkers for osteosarcoma. In this study, we selected genes with a median absolute deviation (MAD) value of the top 5000 in the GSE32981 dataset for subsequent analysis. Weighted correlation network analysis (WGCNA) was used to construct a coexpression network. WGCNA showed that the tan module and midnight blue module were highly correlated with origin and metastases of osteosarcoma, respectively. Enrichment analysis was conducted using genes in the tan module and midnight blue module. A gene coexpression network was constructed by calculating the Spearman correlation coefficients. Four key genes (LTF, C10orf107, HIST1H2AK, and NEXN) were identified to be correlated with the prognosis of osteosarcoma patients. LTF has the highest AUC value, and its effect on osteosarcoma cells was then evaluated. The effect of LTF overexpression on proliferation, migration, and invasion of MG63 and 143B cells was detected by the CCK-8 assay, transwell cell migration assay, and transwell invasion assay, respectively. The overexpression of LTF promoted the proliferation, migration, and invasion of MG63 and 143B cells. In conclusion, LTF may serve as a prognostic biomarker for osteosarcoma.

## 1. Introduction

Osteosarcoma is a kind of primary malignant bone tumors that prefer to occur in teenagers. It is estimated to occur in 2% and 3% cancer cases in age at birth to 14 and 15 to 19, respectively [1]. Despite the advances in therapy, the 5-year survival rate of osteosarcoma was only about 67%~69% [1]. Therefore, osteosarcoma remains a major health problem. The identification of novel biomarkers for osteosarcoma is still needed. Finding more accurate biomarkers would promote the outcome prediction and individualized therapy of osteosarcoma patients.

With the development of high-throughput detection techniques, a large number of genes have been found to be differentially expressed in osteosarcoma [2]. Namløs et al. performed microarray analysis and identified gene expression profiles in osteosarcoma samples of primary and metastatic origin [3]. Buddingh et al. identified gene signatures related to the metastases of osteosarcoma patients [4]. Some differentially expressed genes have been identified

as biomarkers for osteosarcoma. For example, CBX3 and ABCA5 are identified as putative biomarkers for tumor stem cells in osteosarcoma [5]. High MMP9 expression was associated with poor overall survival of osteosarcoma patients [6]. Myc was highly expressed in human osteosarcoma cell lines and tissues. Higher Myc expression was correlated with metastasis and poor prognosis of patients with osteosarcoma [7]. However, the effect of altered genes in osteosarcoma has still not been fully addressed. Also, there is still a lack of understanding in the relationships among these genes.

Weighted gene coexpression network analysis (WGCNA) could find clusters (modules) of highly correlated genes [8]. WGCNA has been applied to screening biomarkers and drug targets in various cancers, including lung cancer [9], colon cancer [10], and bladder cancer [11]. This method could be also used in screening biomarkers for osteosarcoma.

In this study, osteosarcoma-related high-throughput data were retrieved from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). The data were analyzed using bioinformatic methods. We screened 326 candidate genes related to the origin and metastasis of osteosarcoma by WGCNA. Subsequently, we identified 4 key genes correlated with the survival of osteosarcoma patients, including lactoferrin/lactotransferrin (LTF), C10orf107, histone cluster 1 H2ak (HIST1H2AK), and nexilin F-actin binding protein (NEXN). Furthermore, the effect of LTF on proliferation, migration, and invasion of osteosarcoma cells was confirmed.

## 2. Material and Methods

*2.1. Gene Expression Data Collection.* We searched the osteosarcoma-related datasets in the GEO database (https://www.ncbi.nlm.nih.gov/gds/). Two datasets were involved in this work. The dataset GSE32981 from GPL3307 was used as the training dataset. The dataset GSE21257 from the GPL10295 platform was used as the testing dataset.

*2.2. WGCNA.* The median absolute deviation (MAD) value was used to screen the top 5000 genes from the GSE32981 dataset. A scale-free coexpression network was established using the WGCNA package in R software [8]. The gene modules were identified by calculating the topological overlap matrix (TOM). Metascape (https://metascape.org/) was used to perform enrichment analysis [12].

*2.3. Coexpression Network Establishment.* Spearman correlation coefficient (SCC) values of gene pairs were calculated. Gene pairs with $|SCC| > 0.6$ were involved in the coexpression network. Cytoscape software was used to screen the top ten nodes with the highest degree in tan and midnight blue modules, respectively.

*2.4. The Prognostic Value of Genes.* The prognostic value of hub nodes in the constructed coexpression network was calculated. Kaplan–Meier analysis of these hub nodes was conducted using data in GSE21257. Receiver operating characteristic (ROC) curves are used to describe the prediction accuracy of the hub nodes. The area under the ROC curve (AUC) was analyzed.

*2.5. Cell Culture and Transfection.* MG63 and 143B cells were cultured in MEM medium (Gibco, Cat. No. C11095500BT) with FBS (Hyclone, Cat. No. SH30087.01) and penicillin-streptomycin (Hyclone, Cat. No. SH30010). These two cell lines were incubated under 5% $CO_2$ at 37°C.

The overexpression plasmids containing the whole coding sequence of LTF with the Flag-tag and pcDNA3.1 vector served as the negative control. Real-time PCR product sequencing was conducted to confirm that the LTF sequence was successfully cloned into pcDNA3.1 vector. MG63 and 143B cells were cultured with a confluence of 70%–80% and transfected with Lipofectamine 2000 reagent (Invitrogen, Cat. No. 11668019) with the plasmid.

*2.6. CCK-8 Assay.* The proliferation of MG63 and 143B cells was detected before and 1, 2, 3, and 4 days after transfection using a CCK-8 assay kit (Jiangsu KeyGEN BioTECH Co. Ltd, Cat. No. KGA317). MG63 and 143B cells were cultured with the CCK-8 reagent for 4 h, and then the optical density was detected at 450 nm by using a microplate reader (Thermo Fisher Scientific, Multiscan MK3).

*2.7. Cell Migration and Invasion Assay.* The migration and invasion ability of MG63 and 143B cells was detected by the transwell cell migration assay and transwell cell invasion assay, respectively. The difference between the two experiments is that, for the transwell cell invasion assay, the upper chamber was coated with Matrigel (BD Biosciences). In detail, the upper chambers of the transwell plate (BD Biosciences) contain a culture medium without FBS, and the lower chambers contain a complete culture medium. Forty-eight hours after incubation, invasion cells were fixed with 4% paraformaldehyde and stained with crystal violet. After washed with PBS, the images of cells were captured. Data were analyzed using ImageJ software.

*2.8. Statistical Analysis.* Statistical analysis was conducted using GraphPad Prism 8 and R version 3.6.1. Data were presented as mean ± standard deviation. Comparison between two groups was performed using Student's *t*-tests. $P < 0.05$ was considered statistically significant.

*2.9. Ethics Statement.* This work did not involve human participants or animals.

## 3. Results

*3.1. Clustering of Coexpression Module Eigengenes in Osteosarcoma.* Genes in the GSE32981 dataset with the top 5000 MAD values were used for WGCNA analysis. There were no discrete samples as revealed by clustering of the samples (Figure 1(a)). Next, to make the constructed

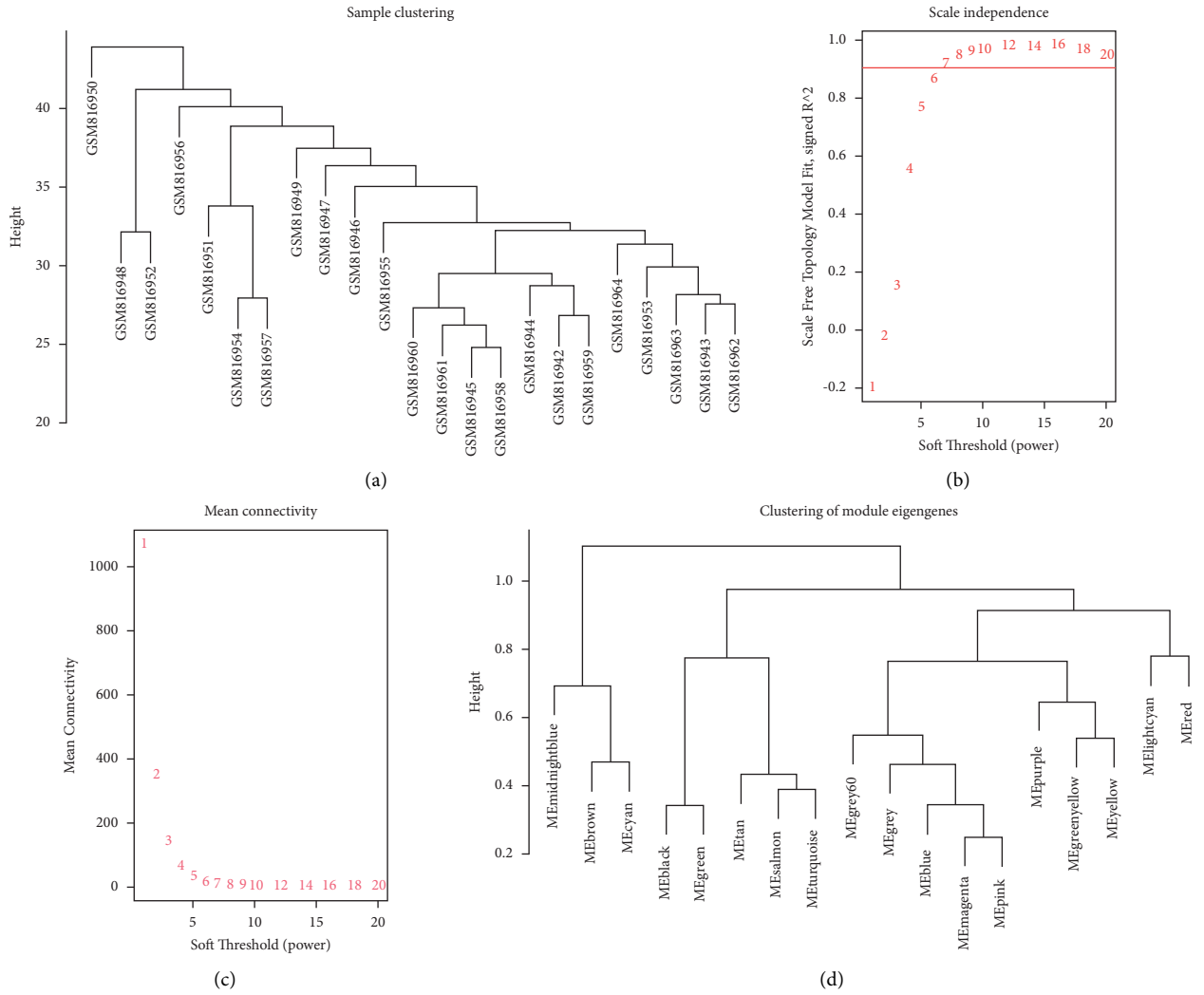



FIGURE 1: The WGCNA of potential biomarkers for osteosarcoma. (a) Sample clustering of GSE32981 to detect outliers. (b) The correlation coefficients between log ($K$) and log ($P(k)$) corresponding to different soft thresholds. (c) A genetic network corresponding to different soft thresholds. (d) Sample cluster of module eigengenes.

network conform to the characteristics of the scale-free network, we performed the screening with a soft threshold. In this study, the scale-independent value reached 0.9 when the power value $\beta$ was set to 7, and the average connectivity ratio was low (Figures 1(b) and 1(c)). Therefore, the coexpression matrix was calculated under the condition of determined $\beta = 7$. The cluster of module eigengenes is shown in Figure 1(d).

*3.2. Identification of Key Modules in Osteosarcoma.* Similar clusters were combined into a new module using two settings: height = 0.2 and min module size = 50. Eighteen modules with similar patterns of connected genes were obtained (Figure 2(a)). Through the correlation study of the network heatmap plot (Figure 2(b)), there was little correlation among the 18 modules. The adjacency relationship between gene and gene in the module is shown in Figure 2(c). The relevance between the 18 modules and characteristics (origin, age, gender, and metastases) of osteosarcoma samples was evaluated based on module-trait relationships (MTRs). The tan module included 195 genes, which showed the highest positive correlation with origin. The midnight blue module included 131 genes, which showed the highest negative correlation with metastases (Figure 2(d)). The tan module was correlated with cancer origin (primary or metastasis). The midnight blue module was correlated with cancer metastasis. Therefore, the tan module and midnight blue module were selected for further analyses.

*3.3. Enrichment Analysis of Genes Involving in Tan Module and Midnight Blue Module.* The enrichment analysis was performed using genes involving in the tan module and midnight blue module, respectively. The results showed that genes involving in the tan module were mainly enriched in biological processes including "extracellular structure organization," "regulation of peptidase activity," and "actin cytoskeleton organization" (Figure 3(a)). Also, genes involving in

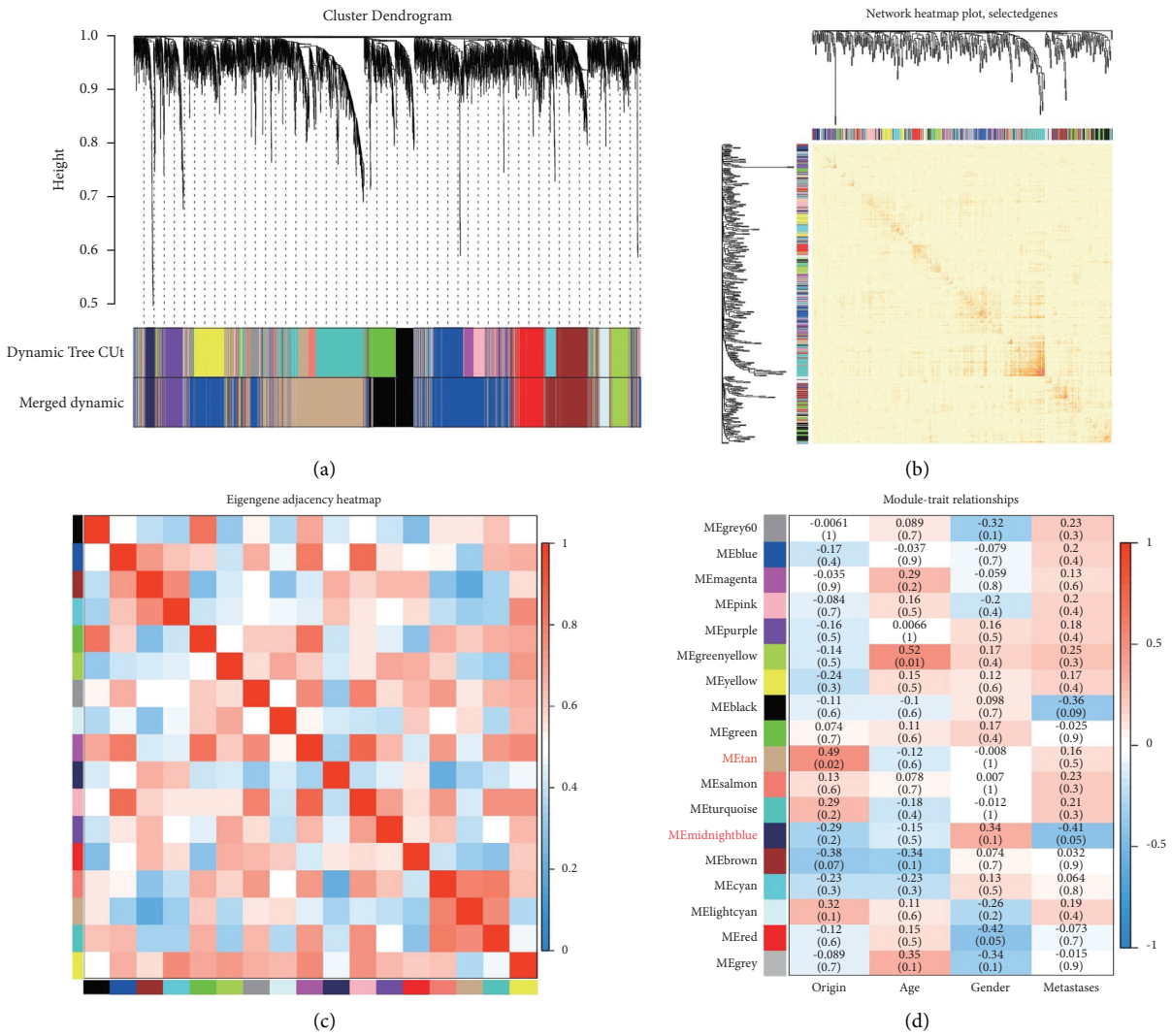



FIGURE 2: Identification of modules and key genes related to osteosarcoma. (a) Gene cluster tree classification diagram. (b) Heatmap plot of topological overlap in the gene network. (c) The eigengene of each colored module was calculated and an adjacency matrix was established. (d) The relevance between eigengenes of 18 modules and status of osteosarcoma samples.

the midnight blue module were mainly enriched in biological processes including "DNA damage/telomere stress-induced senescence," "metalloprotease DUBs," and "negative regulation of chromosome organization" (Figure 3(b)).

*3.4. Coexpression Network of Genes Involving in the Modules.* A coexpression network was constructed using genes involving in the tan module and midnight blue module based on the gene coexpression relationship. Spearman correlation coefficient was calculated using genes in the tan module and midnight blue module. 177 gene pairs with $|SCC| > 0.6$ were screened to construct the coexpression network, including 160 nodes and 177 edges (Figure 4).

*3.5. Validation of Key Genes in the Coexpression Network.* Subsequently, we analyzed the topological properties of the established coexpression network. The top 10 hub nodes with high degree involved in the tan module and midnight blue module are shown in Tables 1 and 2, respectively. Kaplan–Meier analysis of these key genes was performed using the GSE21257 dataset. LTF in the tan module was correlated with the prognosis of osteosarcoma patients (Figure 5(a)). C10orf107, HIST1H2AK, and NEXN in the midnightblue module were correlated with the prognosis of osteosarcoma patients (Figures 5(b)–5(d)). The ROC of specificity and sensitivity was analyzed, and the area under the curve (AUC) was then computed. The AUC of LTF, C10orf107, HIST1H2AK, and NEXN was 0.646, 0.583, 0.565, and 0.584, respectively (Figures 5(e)–5(h)). Among these genes, the AUC of LTF was the highest, which indicates its effectiveness as a prognostic biomarker in osteosarcoma. In addition, two independent datasets (GSE36001 and GSE99671) from the GEO database were used to compare the LTF expression in osteosarcoma samples and normal controls. The results showed that the expression of LTF was lower in osteosarcoma samples compared with that in normal controls (Figure S1).

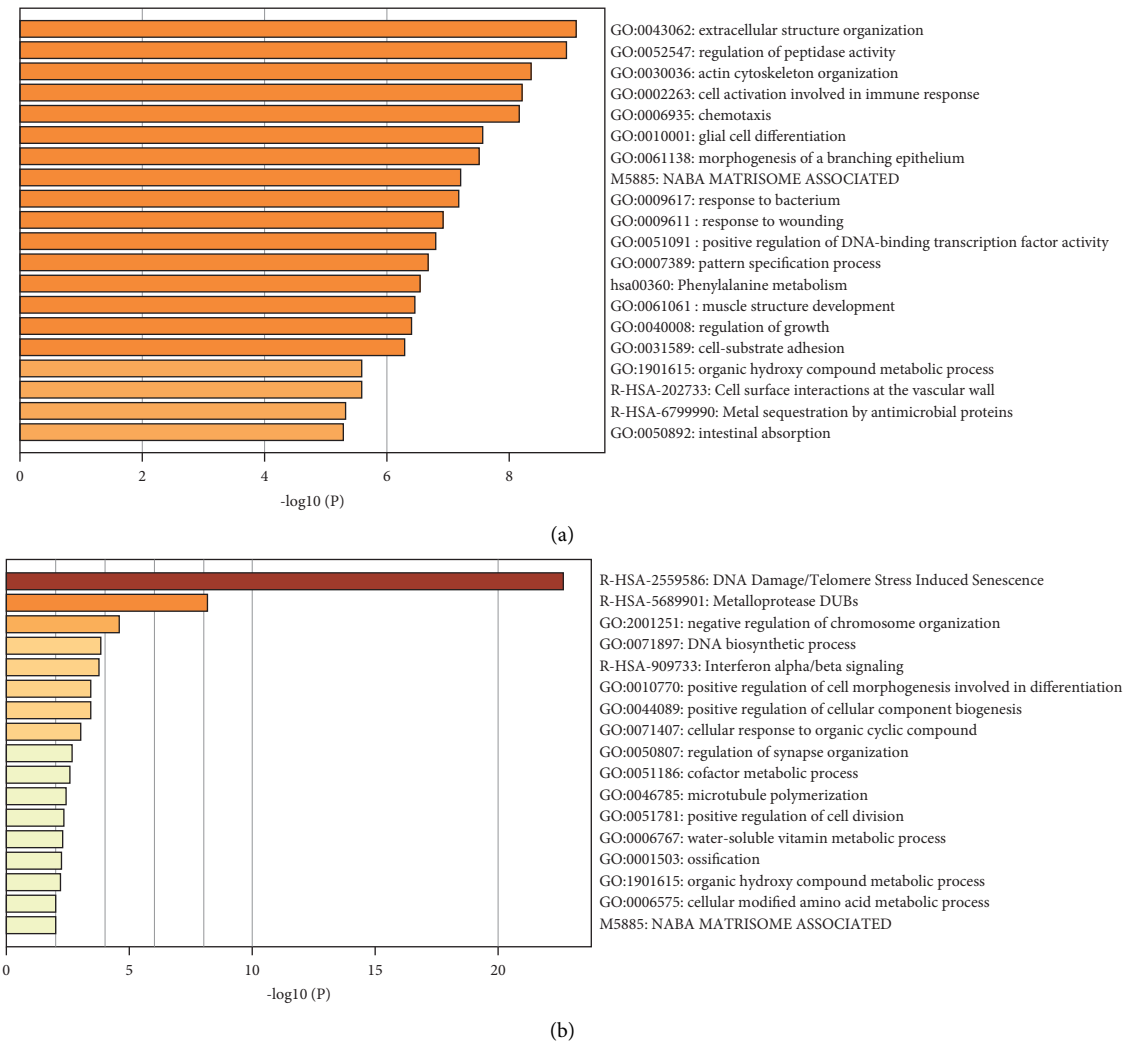



Figure 3: Functional enrichment of genes involving in the tan module and midnight blue module. (a) Functional enrichment of genes involving in the tan module. (b) Functional enrichment of genes involving in the midnightblue module.

*3.6. The Effect of LTF on Osteosarcoma Cells.* The CCK-8 assay showed that the OD value in the LTF overexpression group was lower than that in the pcDNA3.1-transfected group (Figures 6(a) and 6(c)). Therefore, the overexpression of LTF inhibited the proliferation of MG63 and 143B cells (Figures 6(b) and 6(d)). Transwell migration and invasion assays showed that the overexpression of LTF inhibited the migration and invasion ability of MG63 and 143B cells (Figure 7).

## 4. Discussion

Osteosarcoma is the commonest primary bone malignant tumor with a high rate of metastasis [13]. There is still a need for specific prognostic biomarkers and drug targets for osteosarcoma. Identifying specific prognostic biomarkers may contribute to the clinical management of osteosarcoma.

In the present study, we collected osteosarcoma-associated gene profiles in the GEO database and screened the top 5000 genes from GSE32981 using MAD. A weighted gene coexpression network was constructed, and the modules of this network were calculated by TOM. The relevance between eigengenes of 18 modules and origin, age, gender, and metastases status of osteosarcoma samples was calculated. The results showed that the tan module and midnight blue module are correlated with origin and metastases of osteosarcoma, respectively. The origin and metastases status of osteosarcoma were related to the response to therapies and may influence clinical therapeutic decision making [14, 15]. Therefore, these two modules were selected for further investigation.

We then performed enrichment analysis using Metascape. Genes involving in the tan module were mainly enriched in osteosarcoma-related biological processes including "regulation of peptidase activity" [16–18], "phenylalanine metabolism" [19, 20], and "regulation of growth" [21, 22]. Also, genes involving in the midnight blue module were mainly enriched in osteosarcoma-related biological processes, including "DNA damage/telomere stress-induced senescence" [23–25], "metalloprotease DUBs" [26–28], and "interferon alpha/beta signaling" [29–31].

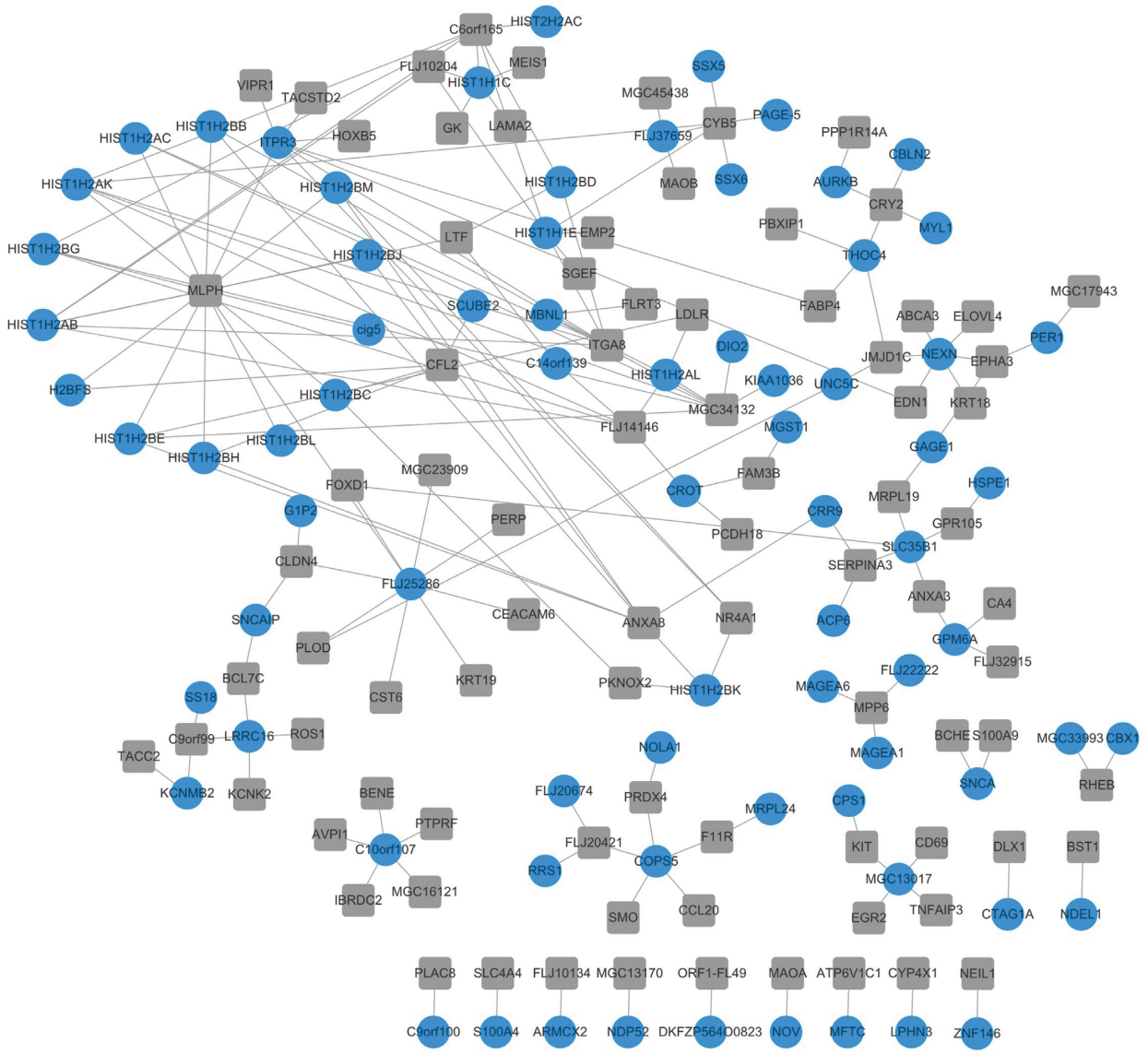



Figure 4: Coexpression network of genes involving in the tan module and midnightblue module. Square gray and circular blue nodes represent genes involving in the tan module and midnightblue module, respectively.

In addition, the correlation among these genes was calculated by Spearman correlation analysis. Gene pairs with $|SCC| > 0.6$ were identified to be screened to construct the coexpression network with 160 nodes and 177 edges. The hub nodes with high degree from the network were identified in the tan module and midnight blue module. Kaplan–Meier analysis of these key genes was performed. Among the 29 genes, four of them was correlated with the prognosis of osteosarcoma patients. C10orf107, HIST1H2AK, and NEXN were from the midnight blue module, and LTF was from the tan module. The expression of C10orf107 and NEXN was positively correlated with poor survival of osteosarcoma patients, whereas the expression of HIST1H2AK and LTF was negatively correlated with poor survival of osteosarcoma patients.

The effect of C10orf107 in cancer was poorly understood, and only a rare translocation t(3; 10) (q26; q21) was observed in an acute myeloid leukemia patient, presented as a fusion of MECOM (chromosome 3q26.2) and C10orf107 (chromosome 10q21.2) [32]. NEXN (encode Nexilin) is related to cardiovascular diseases, including hypertrophic cardiomyopathy, coronary artery disease, and septal defects [33–35]. However, its role in cancer has not been reported.

LTF has been considered as a tumor suppressor in multiple cancers. LTF is deficient or lowly expressed in prostate cancer, nasopharyngeal carcinoma, oral squamous cell carcinoma, etc. [36–38]. Osteosarcoma patients with lower LTF have a poor survival rate compared with those with higher LTF (Figure 5). However, its effect on osteosarcoma is still not known. We then detected its effect on

Table 1: The top 10 hub genes in the tan module.

| Gene name | Class | Degree |
|---|---|---|
| MLPH | tan | 15 |
| ITGA8 | tan | 9 |
| MGC34132 | tan | 8 |
| C6orf165 | tan | 7 |
| ANXA8 | tan | 7 |
| CYB5 | tan | 6 |
| FLJ14146 | tan | 6 |
| CRY2 | tan | 4 |
| CFL2 | tan | 4 |
| MPP6 | tan | 3 |
| EPHA3 | tan | 3 |
| LTF | tan | 3 |
| FLJ10204 | tan | 3 |
| CLDN4 | tan | 3 |
| SERPINA3 | tan | 3 |
| JMJD1C | tan | 3 |
| C9orf99 | tan | 3 |
| FLJ20421 | tan | 3 |
| NR4A1 | tan | 3 |

Table 2: The top 10 hub genes in the midnight blue module.

| Gene name | Class | Degree |
|---|---|---|
| FLJ25286 | Midnight blue | 9 |
| ITPR3 | Midnight blue | 8 |
| HIST1H2AK | Midnight blue | 6 |
| HIST1H1E | Midnight blue | 6 |
| NEXN | Midnight blue | 6 |
| HIST1H2AB | Midnight blue | 6 |
| COPS5 | Midnight blue | 5 |
| SLC35B1 | Midnight blue | 5 |
| HIST1H1C | Midnight blue | 5 |
| C10orf107 | Midnight blue | 5 |

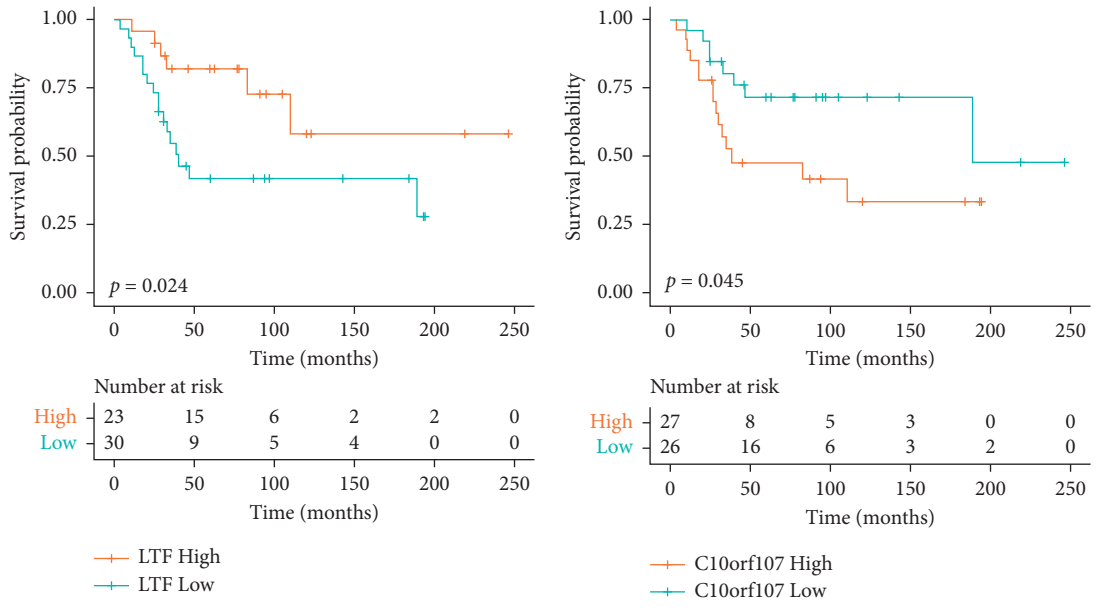



Figure 5: Continued.

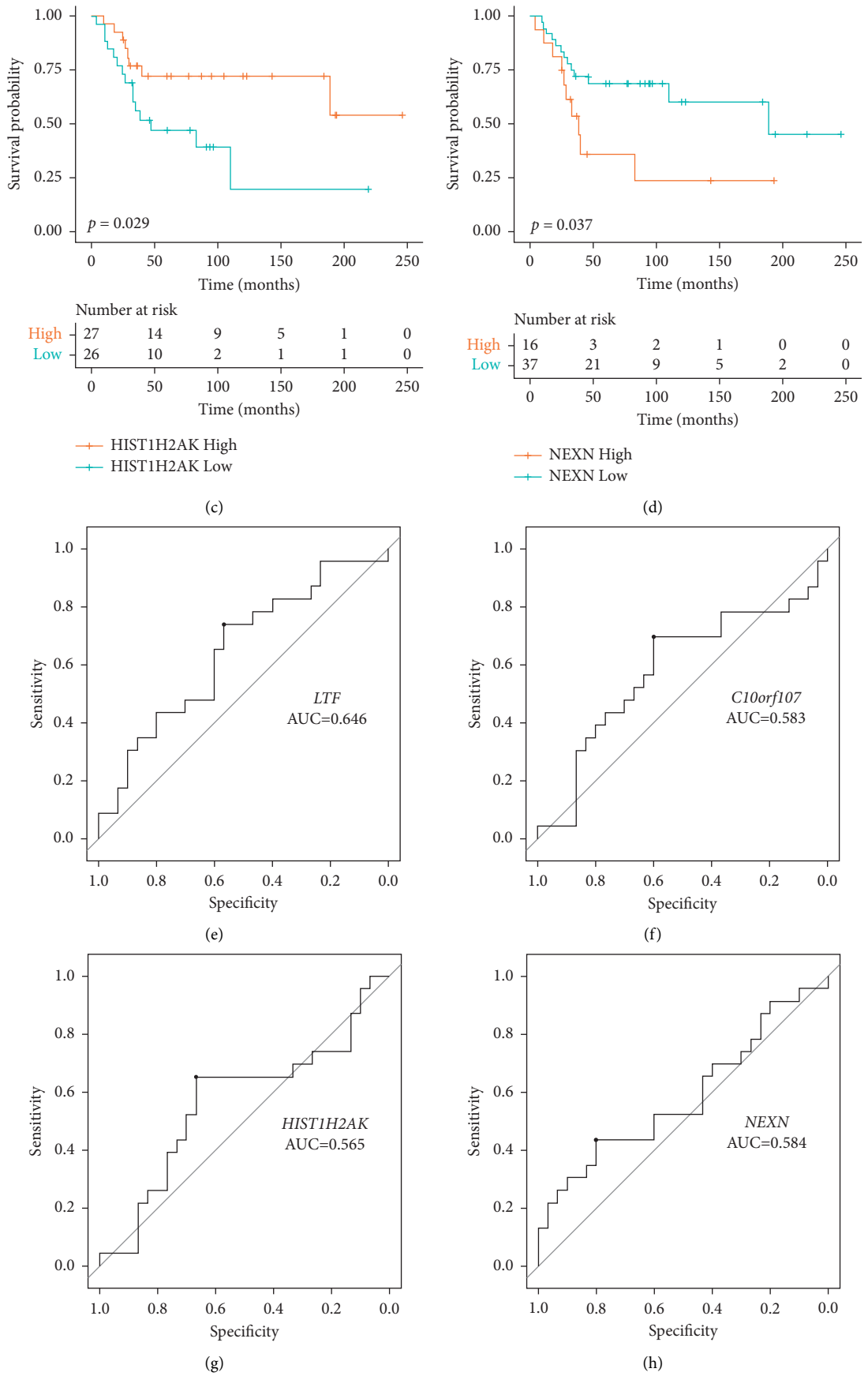



Figure 5: Kaplan–Meier survival curve and ROC curve of key genes. Kaplan–Meier survival cures for LTF (a), C10orf107 (b), HIST1H2AK (c), and NEXN (d), ROC curve of LTF (e), C10orf107 (f), HIST1H2AK (g), and NEXN (h).

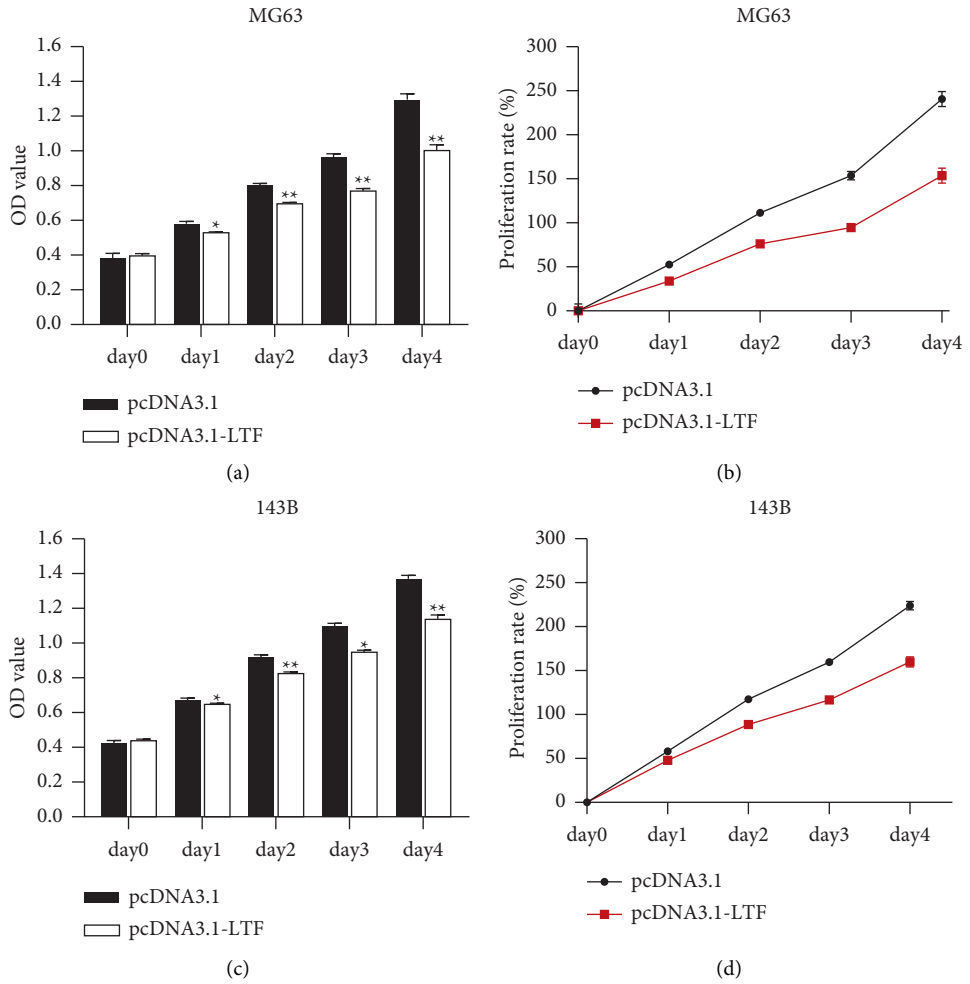



Figure 6: The effect of LTF on the proliferation of osteosarcoma cells. (a) The OD value of MG63 cells. (b) The proliferation rate of MG63 cells. (c) The OD value of 143B cells. (d) The proliferation rate of 143B cells. $^{*}P < 0.05$, $^{**}P < 0.01$ vs. pcDNA3.1, $n = 3$.

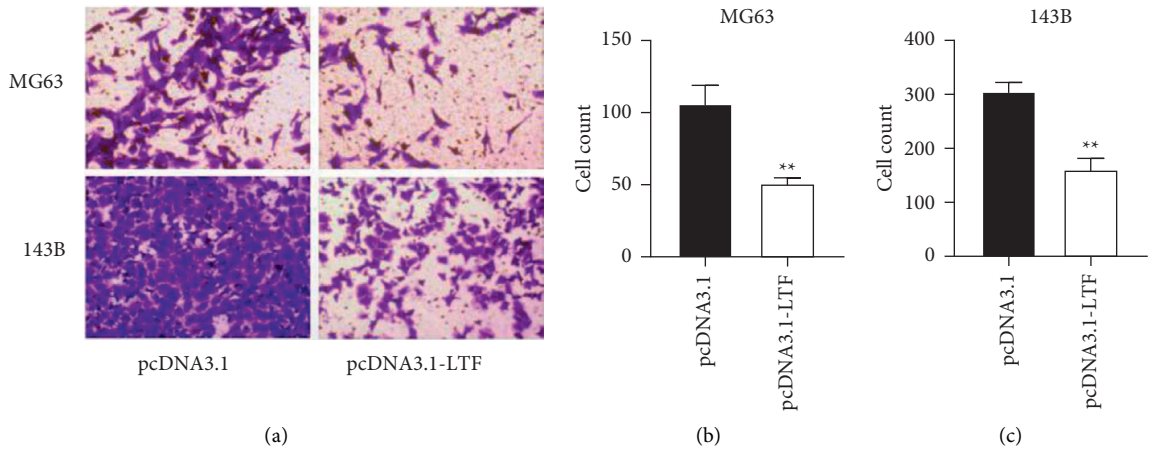



Figure 7: Continued.

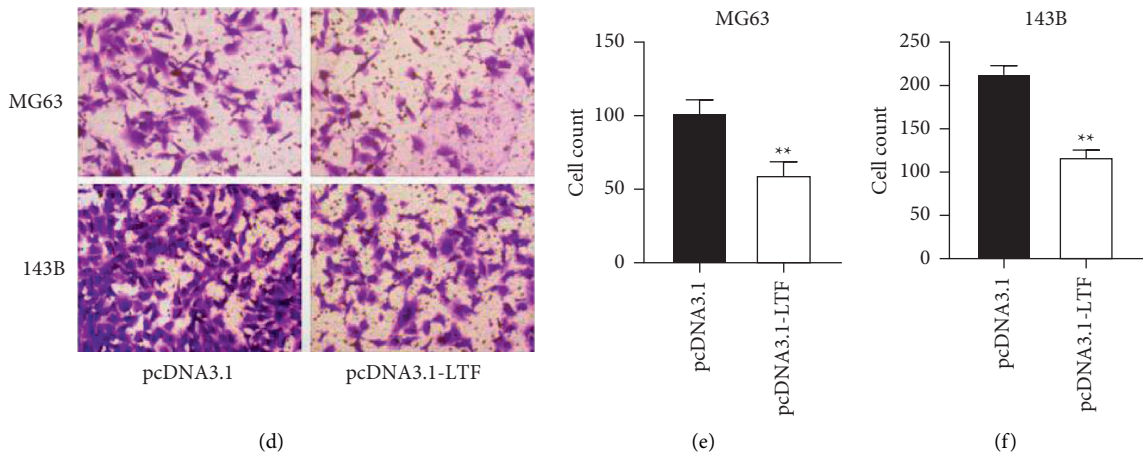



Figure 7: The effect of LTF on migration and invasion of osteosarcoma cells. (a) Representative images of cell migration assay. (b) The effect of LTF on the migration of MG63 cells. (c) The effect of LTF on the migration of 143B cells. (d) Representative images of cell invasion assay. (e) The effect of LTF on the invasion of MG63 cells. (f) The effect of LTF on the invasion of 143B cells. $^{**}P < 0.01$ vs. pcDNA3.1, $n = 3$.

osteosarcoma cells. The results showed that the overexpression of LTF inhibited the proliferation of osteosarcoma cells (Figure 6). Moreover, the upregulation of LTF also promoted the migration and invasion of osteosarcoma cells (Figure 7). These results were in accordance with the view that LTF is a tumor suppressor. LTF has inhibitive effects on cancer progression. Xiao et al. revealed that the growth inhibitory effects of LTF were through a p27/cyclin E-dependent pathway in head and neck cancer cells [39]. Similarly, Zhang et al. found that lactoferrin exerts an inhibitive effect on breast cancer cell growth through inducing cell cycle arrest, with little effect on normal breast cancer cells [40]. LTF also showed migration and/or invasion inhibitory effect in different types of cancer cells.

LTF could inhibit the migration of colorectal adenocarcinoma cells (Caco-2) and gastric adenocarcinoma cells (AGS) [41]. LTF suppressed and even reversed epithelial-to-mesenchymal transition process in oral squamous cell carcinoma [42] and glioblastoma [43]. Moreover, the administration of LTF inhibited the liver and spleen metastasis of L5178Y-ML25 cells and lung metastasis of L5178Y-ML25 cells [44]. A recent study found that LTF deficiency enhanced lung metastasis of melanoma in an LTF KO mouse model, which was related to the enhancing of the TLR9 pathway [30].

In summary, osteosarcoma-related datasets from the GEO database had been systematically analyzed. WGCNA showed that the tan module and midnight blue module are highly correlated with origin and metastases of osteosarcoma, respectively. Enrichment analyses showed that genes in these two modules were associated with cancer-related pathways. A gene coexpression network was constructed using these genes, and the key genes were identified with a high degree. Four key genes (C10orf107, HIST1H2AK, NEXN, and LTF) were found to be correlated with the prognosis of osteosarcoma patients. LTF has the highest AUC value, and its inhibitive effect on osteosarcoma cells was validated. Still, further studies are needed to reveal the precise effect and mechanism of these genes.

## 5. Conclusions

The key gene we identified, such as LTF, could be a prognosis biomarker and therapeutic drug target for osteosarcoma. Our study brought new insights into the investigation of osteosarcoma.

## Data Availability

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

## Additional Points

*Key Messages*. (1) The tan module and midnightblue module are with the highest correlation with origin and metastases of osteosarcoma, respectively. (2) Overexpression of LTF promoted the proliferation, migration, and invasion of osteosarcoma cells. (3) Key genes including LTF, C10orf107, HIST1H2AK, and NEXN may serve as prognostic biomarkers for osteosarcoma.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 81803012), the Natural Science Foundation of Heilongjiang Province, China (Grant no. YQ2019H008), the Higher Education Reform Research Project of Heilongjiang Province (Grant nos. SJGY20190429 and SJGY20190430), the Key Research Project of Heilongjiang Province (Grant no. GZ20210011), Transverse Research Project from the Second Affiliated Hospital of Harbin Medical University (Grant nos. 070500020146 and 070500020278), and Standardized Training Project for

Resident Doctors, the Second Affiliated Hospital of Harbin Medical University (Grant nos. 2020020208 and 2020020223).

## Supplementary Materials

Figure S1: the expression of LTF in osteosarcoma samples and normal controls. (A) The expression of LTF in osteosarcoma samples and normal controls calculated using the GSE36001 dataset. (B) The expression of LTF in osteosarcoma samples and normal controls calculated using the GSE99671 dataset. (*Supplementary Materials*)

## References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA: A Cancer Journal for Clinicians*, vol. 70, no. 1, pp. 7–30, 2020.

[2] J.-C. Guo, S.-S. Fang, Y. Wu et al., "CNIT: a fast and accurate web tool for identifying protein-coding and long non-coding transcripts based on intrinsic sequence composition," *Nucleic Acids Research*, vol. 47, no. W1, pp. W516–W522, 2019.

[3] H. M. Namløs, S. H. Kresse, C. R. Müller et al., "Global gene expression profiling of human osteosarcomas reveals metastasis-associated chemokine pattern," *Sarcoma*, vol. 2012, Article ID 639038, 12 pages, 2012.

[4] E. P. Buddingh, M. L. Kuijjer, R. A. J. Duim et al., "Tumor-infiltrating macrophages are associated with metastasis suppression in high-grade osteosarcoma: a rationale for treatment with macrophage activating agents," *Clinical Cancer Research*, vol. 17, no. 8, pp. 2110–2119, 2011.

[5] V. Saini, C. D. Hose, A. Monks et al., "Identification of CBX3 and ABCA5 as putative biomarkers for tumor stem cells in osteosarcoma," *PLoS One*, vol. 7, no. 8, Article ID e41401, 2012.

[6] H. Li, K. Zhang, L.-H. Liu et al., "A systematic review of matrix metalloproteinase 9 as a biomarker of survival in patients with osteosarcoma," *Tumor Biology*, vol. 35, no. 6, pp. 5487–5491, 2014.

[7] W. Feng, D. C. Dean, F. J. Hornicek et al., "Myc is a prognostic biomarker and potential therapeutic target in osteosarcoma," *Therapeutic Advances in Medical Oncology*, vol. 12, Article ID 1758835920922055, 2020.

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

[9] J. Hu, L. Xu, T. Shou, and Q. Chen, "Systematic analysis identifies three-lncRNA signature as a potentially prognostic biomarker for lung squamous cell carcinoma using bioinformatics strategy," *Translational Lung Cancer Research*, vol. 8, no. 5, pp. 614–635, 2019.

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

[11] Y. Di, D. Chen, W. Yu, and L. Yan, "Bladder cancer stage-associated hub genes revealed by WGCNA co-expression network analysis," *Hereditas*, vol. 156, no. 1, p. 7, 2019.

[12] Y. Zhou, B. Zhou, L. Pache et al., "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets," *Nature Communications*, vol. 10, no. 1, p. 1523, 2019.

[13] X. Huang, J. Zhao, J. Bai et al., "Risk and clinicopathological features of osteosarcoma metastasis to the lung: a population-based study," *Journal of Bone Oncology*, vol. 16, Article ID 100230, 2019.

[14] D. Mercatelli, M. Bortolotti, A. Bazzocchi, A. Bolognesi, and L. Polito, "Immunoconjugates for osteosarcoma therapy: preclinical experiences and future perspectives," *Biomedicines*, vol. 6, no. 1, p. 19, 2018.

[15] S. D'Oronzo, R. Coleman, J. Brown, and F. Silvestris, "Metastatic bone disease: pathogenesis and therapeutic options: up-date on bone metastasis management," *Journal of Bone Oncology*, vol. 15, no. 004, 2019.

[16] D. J. Luo, L. J. Li, H. F. Huo, X. Q. Liu, H. W. Cui, and D. M. Jiang, "MicroRNA-29b sensitizes osteosarcoma cells to doxorubicin by targeting matrix metalloproteinase 9 (MMP-9) in osteosarcoma," *European Review for Medical and Pharmacological Sciences*, vol. 23, no. 4, pp. 1434–1442, 2019.

[17] J. F. Liu, P. C. Chen, T. M. Chang, and C. H. Hou, "Thrombospondin-2 stimulates MMP-9 production and promotes osteosarcoma metastasis via the PLC, PKC, c-Src and NF-κB activation," *Journal of Cellular and Molecular Medicine*, vol. 24, no. 21, pp. 12826–12839, 2020.

[18] W. Xie, J. Xiao, T. Wang, D. Zhang, and Z. Li, "MicroRNA-876-5p inhibits cell proliferation, migration and invasion by targeting c-Met in osteosarcoma," *Journal of Cellular and Molecular Medicine*, vol. 23, no. 5, pp. 3293–3301, 2019.

[19] C. Ferrari, C. Zonta, L. Cansolino et al., "Selective uptake of p-boronophenylalanine by osteosarcoma cells for boron neutron capture therapy," *Applied Radiation and Isotopes*, vol. 67, no. 7-8, pp. S341–S344, 2009.

[20] M. A. Hilton, S. Bertolone, and C. C. Patel, "Daily profiles of plasma phenylalanine and tyrosine in patients with osteogenic sarcoma during treatment with high-dose methotrexate-citrovorum rescue," *Medical and Pediatric Oncology*, vol. 17, no. 4, pp. 265–270, 1989.

[21] X. Ji, L. Shan, P. Shen, and M. He, "Circular RNA circ_001621 promotes osteosarcoma cells proliferation and migration by sponging miR-578 and regulating VEGF expression," *Cell Death & Disease*, vol. 11, no. 1, p. 18, 2020.

[22] L. Zhang, C. Yang, Y. Huang et al., "Cardamonin inhibits the growth of human osteosarcoma cells through activating P38 and JNK signaling pathway," *Biomedicine & Pharmacotherapy*, vol. 134, Article ID 111155, 2021.

[23] Z. Zhang, S. H. Ha, Y. J. Moon et al., "Inhibition of SIRT6 potentiates the anti-tumor effect of doxorubicin through suppression of the DNA damage repair pathway in osteosarcoma," *Journal of Experimental & Clinical Cancer Research*, vol. 39, no. 1, p. 247, 2020.

[24] X. Li, D. C. Dean, G. M. Cote et al., "Inhibition of ATR-Chk1 signaling blocks DNA double-strand-break repair and induces cytoplasmic vacuolization in metastatic osteosarcoma," *Therapeutic Advances in Medical Oncology*, vol. 12, Article ID 1758835920956900, 2020.

[25] A. Kiss, A. P. Raduly, Z. Regdon et al., "Targeting nuclear NAD(+) synthesis inhibits DNA repair, impairs metabolic adaptation and increases chemosensitivity of U-2OS osteosarcoma cells," *Cancers (Basel)*, vol. 12, no. 5, 2020.

[26] M. Zhang, J. Zhang, and Q. Zhou, "Elevated expression of microRNA-328-3p suppresses aggressive malignant behaviors via targeting matrix metalloprotease 16 in osteosarcoma," *OncoTargets and Therapy*, vol. 12, pp. 2063–2070, 2019.

[27] S. Z. Ingvarsen, H. Gårdsvoll, S. van Putten et al., "Tumor cell MT1-MMP is dispensable for osteosarcoma tumor growth, bone degradation and lung metastasis," *Scientific Reports*, vol. 10, no. 1, Article ID 19138, 2020.

[28] A. Notaro, S. Emanuele, F. Geraci et al., "WIN55,212-2-Induced expression of Mir-29b1 favours the suppression of osteosarcoma cell migration in a SPARC-independent manner," *International Journal of Molecular Sciences*, vol. 20, no. 20, 2019.

[29] J. P. Cheng, B. Huang, J. H. Duan, K. J. Yi, and Z. L. Zhuang, "miR-4295 promotes cell proliferation, migration and invasion of osteosarcoma through targeting interferon regulatory factor 1," *Oncology Letters*, vol. 20, no. 5, p. 260, 2020.

[30] L. Wei, X. Zhang, J. Wang et al., "Lactoferrin deficiency induces a pro-metastatic tumor microenvironment through recruiting myeloid-derived suppressor cells in mice," *Oncogene*, vol. 39, no. 1, pp. 122–135, 2020.

[31] J.-Y. Wang, J.-Q. Wang, and S.-B. Lu, "miR-628-5p promotes growth and migration of osteosarcoma by targeting IFI44L," *Biochemistry and Cell Biology*, vol. 98, no. 2, pp. 99–105, 2020.

[32] T. Jancuskova, R. Plachy, L. Zemankova et al., "Molecular characterization of the rare translocation $t(3; 10)(q26; q21)$ in an acute myeloid leukemia patient," *Molecular Cytogenetics*, vol. 7, no. 1, p. 47, 2014.

[33] H. Wang, Z. Li, J. Wang et al., "Mutations in NEXN, a Z-disc gene, are associated with hypertrophic cardiomyopathy," *The American Journal of Human Genetics*, vol. 87, no. 5, pp. 687–693, 2010.

[34] C. Wu, H. Yan, J. Sun et al., "NEXN is a novel susceptibility gene for coronary artery disease in Han Chinese," *PLoS One*, vol. 8, no. 12, Article ID e82135, 2013.

[35] F. Yang, L. Zhou, Q. Wang et al., "NEXN inhibits GATA4 and leads to atrial septal defects in mice and humans," *Cardiovascular Research*, vol. 103, no. 2, pp. 228–237, 2014.

[36] S. Shaheduzzaman, A. Vishwanath, B. Furusato et al., "Silencing of Lactotransferrin expression by methylation in prostate cancer progression," *Cancer Biology & Therapy*, vol. 6, no. 7, pp. 1088–1095, 2007.

[37] H. M. Yi, Y. C. Li, and R. H. Zhong, "Expression, genetic and epigenetic alterations of LTF gene in nasopharyngeal carcinoma cell lines," *Zhonghua Zhong Liu Za Zhi*, vol. 32, no. 10, pp. 729–733, 2010.

[38] J. Zhang, T. Ling, H. Wu, and K. Wang, "Re-expression of lactotransferrin , a candidate tumor suppressor inactivated by promoter hypermethylation, impairs the malignance of oral squamous cell carcinoma cells," *Journal of Oral Pathology & Medicine*, vol. 44, no. 8, pp. 578–584, 2015.

[39] Y. Xiao, C. L. Monitto, K. M. Minhas, and D. Sidransky, "Lactoferrin down-regulates G1 cyclin-dependent kinases during growth arrest of head and neck cancer cells," *Clinical Cancer Research*, vol. 10, no. 24, pp. 8683–8686, 2004.

[40] Y. Zhang, A. Nicolau, C. F. Lima, and L. R. Rodrigues, "Bovine lactoferrin induces cell cycle arrest and inhibits mTOR signaling in breast cancer cells," *Nutrition and Cancer*, vol. 66, no. 8, pp. 1371–1385, 2014.

[41] M. Nakajima, I. Shinoda, Y. Samejima, H. Miyauchi, Y. Fukuwatari, and H. Hayasawa, "Lactoferrin as a suppressor of cell migration of gastrointestinal cell lines," *Journal of Cellular Physiology*, vol. 170, no. 2, pp. 101–105, 1997.

[42] C. Chea, M. Miyauchi, T. Inubushi et al., "Bovine lactoferrin reverses programming of epithelial-to-mesenchymal transition to mesenchymal-to-epithelial transition in oral squamous cell carcinoma," *Biochemical and Biophysical Research Communications*, vol. 507, no. 1-4, pp. 142–147, 2018.

[43] A. Cutone, B. Colella, A. Pagliaro et al., "Native and iron-saturated bovine lactoferrin differently hinder migration in a model of human glioblastoma by reverting epithelial-to-mesenchymal transition-like process and inhibiting interleukin-6/STAT3 axis," *Cellular Signalling*, vol. 65, Article ID 109461, 2020.

[44] Y.-C. Yoo, S. Watanabe, R. Watanabe, K. Hata, K.-I. Shimazaki, and I. Azuma, "Bovine lactoferrin and lactoferricin inhibit tumor metastasis in mice," *Advances in Lactoferrin Research*, vol. 443, pp. 285–291, 1998.

Hindawi
Journal of Oncology
Volume 2022, Article ID 9273628, 10 pages
https://doi.org/10.1155/2022/9273628

*Research Article*

# Evaluation of the Prognostic Value of Long Noncoding RNAs in Lung Squamous Cell Carcinoma

**Xiaoting Zhang**,[1] **Yue Su**,[2] **Xian Fu**,[1] **Jing Xiao**,[1] **Guicheng Qin**,[1] **Mengli Yu**,[1] **Xiaofeng Li**,[3] **and Guihong Chen**[1,2]

[1]*Shenzhen Bao'an District Songgang People's Hospital, Shenzhen, China*
[2]*School of Pharmaceutical Sciences, Guangzhou Medical University, Guangzhou, China*
[3]*Department of Laboratory Medicine, Peking University Shenzhen Hospital, Shenzhen, China*

Correspondence should be addressed to Xiaofeng Li; 13530597138@163.com

Received 8 October 2021; Accepted 16 December 2021; Published 13 January 2022

Academic Editor: Jincheng Guo



Lung squamous cell carcinoma (LUSC) is the most common type of lung cancer accounting for 40% to 51%. Long noncoding RNAs (lncRNAs) have been reported to play a significant role in the invasion, migration, and proliferation of lung cancer tissue cells. However, systematic identification of lncRNA signatures and evaluation of the prognostic value for LUSC are still an urgent problem. In this work, LUSC RNA-seq data were collected from TCGA database, and the limma R package was used to screen differentially expressed lncRNAs (DElncRNAs). In total, 216 DElncRNAs were identified between the LUSC and normal samples. lncRNAs associated with prognosis were calculated using univariate Cox regression analysis. The overall survival (OS) prognostic model containing 10 lncRNAs and the disease-free survival (DFS) prognostic model consisting of 11 lncRNAs were constructed using a machine learning-based algorithm, systematic LASSO-Cox regression analysis. We found that the survival rate of samples in the high-risk group was lower than that in the low-risk group. Results of ROC curves showed that both the OS and DFS risk score had better prognostic effects than the clinical characteristics, including age, stage, gender, and TNM. Two lncRNAs (LINC00519 and FAM83A-AS1) that were commonly identified as prognostic factors in both models could be further investigated for their clinical significance and therapeutic value. In conclusion, we constructed lncRNA prognostic models with considerable prognostic effect for both OS and DFS of LUSC.

## 1. Introduction

Lung cancer is one of the most common types of cancer. In 2018, lung cancer accounted for 11.6% of global cancer [1], and more than 1,600,000 new cases are diagnosed yearly [2]. Due to its indistinct early symptoms, it is often diagnosed in the middle or late stages, which usually leads to a very poor prognosis [3]. Non-small-cell lung cancer (NSCLC) accounts for more than 80% of total lung cancer, including lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), and large cell carcinoma (LCLC), among which LUAD and LUSC are the most prevalent ones [4, 5]. Despite advances in the treatment methods of LUSC, the mortality is still high, and the 5-year overall survival (OS) rate of LUSC patients with clinical I and II stages is about 40%. Notably, the 5-year OS rate for the stage III–IV LUSC patients is less than 5% [6, 7]. However, the basic methods for assessing the diagnosis and prognosis of LUSC are based on disease stage and histological grade.

Long noncoding RNAs (lncRNAs) are a type of noncoding RNAs of over 200 base pairs with limited protein-coding potential [8]. Recently, an increasing number of lncRNAs have been identified in humans, and the number continues to rise [9–11]. However, only a small fraction of human lncRNAs have been comprehensively investigated and functionally annotated, resulting in the majority of the rest being still annotated as unknown functions [12]. In recent years, numerous studies have indicated that the

dysregulation of certain lncRNAs plays an important role in a variety of tumors [13–16]. Researchers have paid increasing attention to the potential of lncRNAs in cancer diagnosis and prognosis because the aberrant expression of lncRNAs is associated with the cancer onset and progression [17–19]. Therefore, finding effective prognostic lncRNA biomarkers to prompt therapy and improve the patient survival rate has great significance in cancers.

Recent developments of sequencing and omics technologies provide the opportunity to perform large-scale measurements of diseases at the expression level. High-dimension data problems including prognostic analysis can be addressed using machine learning algorithms. In this study, LUSC data with a large sample size were downloaded from TCGA database [20] and were systematically integrated and analyzed based on bioinformatics methods including differentially expressed gene analysis (DEGA), Gene Ontology (GO) enrichment analysis, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. Then, we constructed OS and DFS prognostic models of LUSC using least absolute shrinkage and selection operator (LASSO) and Cox regression analysis and explored the key lncRNAs as potentially valuable prognosticators associated with LUSC.

## 2. Materials and Methods

*2.1. Data Collection.* The RNA-seq data of 543 samples and their corresponding clinical information, including 494 LUSC patients and 49 normal controls, were collected from TCGA database via UCSC Xena (https://xenabrowser.net/hub/). lncRNA-RNA interaction relationships were obtained from starBase 3.0 (http://starbase.sysu.edu.cn/index.php), and these RNAs were used as potential target genes for lncRNAs. The clinical characteristics of LUSC patients are listed in Table 1, and the research procedure is indicated in Figure 1.

*2.2. Differential Analysis of lncRNA.* The limma R package was used to identify the differentially expressed lncRNAs (DElncRNAs) between the LUSC and normal samples. Absolute value of fold change (FC) > 2 and FDR-adjusted $P$ value <0.05 were used as thresholds. The ggplot2 package was used to draw the volcano plot of lncRNA expression, and the pheatmap package was used to plot the heatmap of the identified DElncRNAs.

*2.3. Screening of the Prognostic lncRNA.* Univariate Cox regression analysis was applied for the clinical data and the lncRNA expression data using the survival R package. lncRNAs related to OS and DFS were separately screened with the $P$ value of 0.05 as the threshold. Then, a machine learning-based algorithm, LASSO-Cox regression analysis, was used to screen a panel of lncRNAs that were significantly related to OS and DFS. Next, the LUSC samples were randomly divided into a training set and a test set at a ratio of 1 : 1, and 10-fold cross-validation was performed to tune lncRNAs related to OS and DFS in the training set.

Table 1: Clinical information of samples in the training set and test set.

| Parameters | OS ($n = 494$) | | DFS ($n = 373$) | |
|---|---|---|---|---|
| | Training (247) | Test (247) | Training (186) | Test (187) |
| *Age* | | | | |
| >60 | 187 | 195 | 149 | 130 |
| ≤60 | 57 | 50 | 36 | 53 |
| *Gender* | | | | |
| Male | 188 | 178 | 137 | 136 |
| Female | 59 | 69 | 49 | 51 |
| *Stage* | | | | |
| 1 and 2 | 196 | 204 | 158 | 144 |
| 3 and 4 | 48 | 42 | 26 | 41 |
| *PN* | | | | |
| N0 | 161 | 156 | 124 | 109 |
| N1–3 | 86 | 122 | 62 | 78 |
| *PM* | | | | |
| M0 | 206 | 200 | 154 | 150 |
| M1–3 | 38 | 46 | 31 | 36 |
| *PT* | | | | |
| T0–2 | 196 | 205 | 158 | 143 |
| T3-4 | 51 | 42 | 28 | 44 |

*2.4. Functional Enrichment Analysis.* Target genes of lncRNAs were obtained from the starBase 3.0 database [21, 22]. These target genes were further analyzed using Gene Ontology (GO) [23] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway [24] functional enrichment analyses via the clusterProfiler R package [25]. An FDR-adjusted $P$ value <0.05 was considered to be statistically significant. GO enrichment analysis was performed for ontologies of biological process (BP), cellular component (CC), and molecular function (MF). In order to compare the difference of immune and stromal scores in the high- and low-risk groups, the expression signature of LUSC samples was calculated by the estimate package in R.

*2.5. Construction of the lncRNA Prognostic Model and Survival Analysis.* The univariate Cox regression method was separately conducted to select lncRNAs related to OS and DFS using the survival package in R. Furthermore, a multivariate Cox regression analysis was performed to confirm their independence, and log-rank $P$ value <0.05 was considered as statistically significant. Subsequently, the prognostic risk score model of OS and DFS was established, respectively, with the use of survival-related lncRNAs (formula (1)) by using the survival R package. The prognostic risk score model of OS and DFS was as follows: $\text{Risk\_score} = \sum coef_i * lncRNA_i$, where $coef_i$ is the coefficient of the $i$ lncRNA in multivariate Cox regression analysis and $lncRNA_i$ is the expression level.

LUSC samples were further divided into the high-risk group and low-risk group according to their median risk score [26, 27]. The survival R package was then separately used to map survival curves of the high-risk group and low-

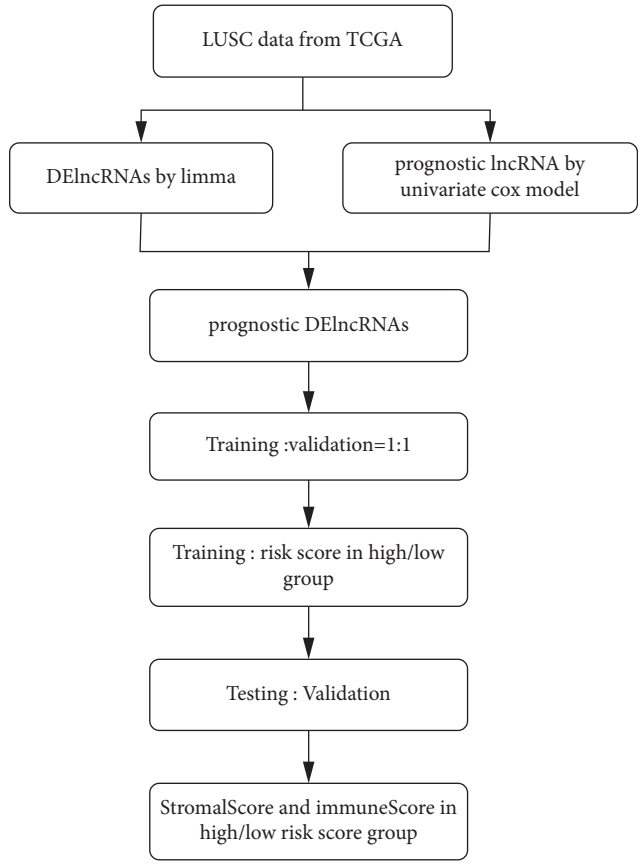



Figure 1: The workflow of this study.

risk group. In addition, a log-rank test was used to estimate the significance between survival curves and further analyze the difference of survival between the two groups.

*2.6. Assessment of the Prognostic Predictive Risk Models.* To validate the prediction accuracy of the prognostic risk model, the receiver operating characteristic (ROC) curves were used to compare the high-risk and low-risk LUSC patients. Furthermore, to verify whether the lncRNA prognostic model was an independent prognostic factor, the univariate and multivariate Cox regression tests were conducted for OS and DFS separately, using the risk score and clinical features (such as stage, gender, age, and TNM) in the first, third, and fifth year by calculating the area under the ROC curves (AUCs).

## 3. Results

*3.1. Screening of Differentially Expressed lncRNAs.* To obtain differentially expressed lncRNAs (DElncRNAs) between LUSC and normal samples, the expression signature of lncRNAs in 494 LUSC patients and 49 normal samples was obtained from TCGA database and was screened by the limma R package. A two-fold change and an FDR-adjusted $P$ value of 0.05 were set as the thresholds for DElncRNA identification. A total of 216 DElncRNAs were screened, including 95 downregulated and 75 upregulated DElncRNAs (Figure 2(a)). The expression abundance of these DElncRNAs was illustrated in a heatmap (Figure 2(b)).

*3.2. Screening of lncRNAs Related to Prognosis.* In order to get lncRNAs related to prognosis, a univariate Cox regression analysis was applied to compare the clinical features (including the OS, DFS, and corresponding survival status) of the LUSC samples and normal samples from TCGA database. Using a $P$ value of 0.05 as the threshold, we obtained 489 lncRNAs significantly related to OS and 920 lncRNAs positively related to DFS, among which there were 36 DElncRNAs related to OS (Figure 2(c)) and 40 DElncRNAs related to DFS, respectively (Figure 2(d)). In other words, these overlapped lncRNAs were both differentially expressed and survival related.

The starBase 3.0 database was used to identify the target genes of the OS- and DFS-related lncRNAs. Functional enrichment analysis showed that the target genes were enriched in several biological processes, including protein localization to the endoplasmic reticulum, SRP-dependent cotranslational protein targeting the membrane, cotranslational protein targeting the membrane, and mRNA catabolic process (Figure 3(a)). Structural constituent of the ribosome, cell adhesion molecule binding, and cadherin binding are the main molecular functions these lncRNAs are involved in (Figure 3(b)). The products of these targeting genes were

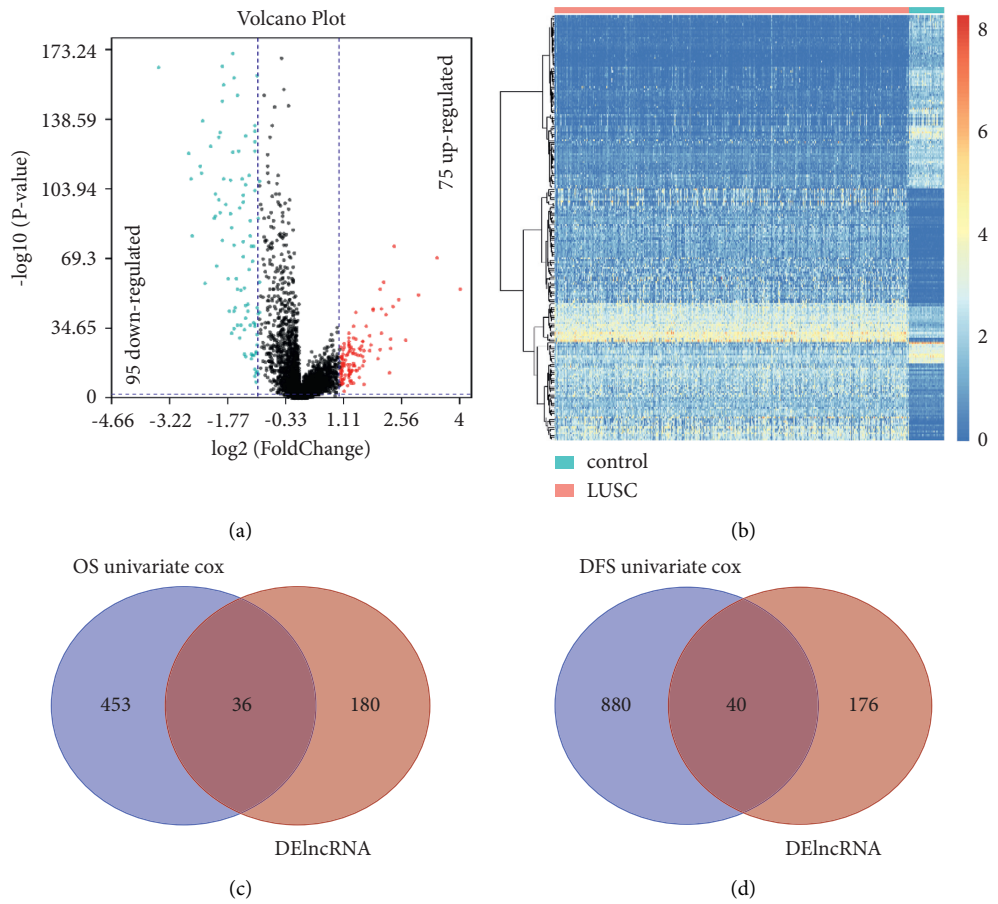



Figure 2: (a) Volcano plot of the screened DElncRNA. (b) Heatmap showing the expression abundance of the DElncRNAs. (c) Venn diagram of the DElncRNAs and OS-related lncRNAs in LUSC. (d) Venn diagram of the DElncRNAs and DFS-related lncRNAs in LUSC.

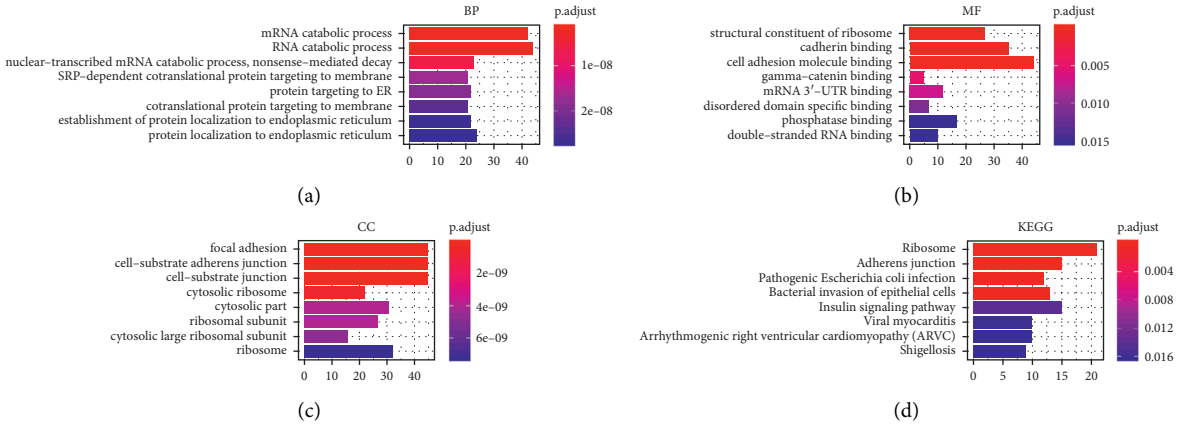



Figure 3: Functional enrichment analysis of GO and KEGG. Functional categories including biological process (BP), molecular function (MF), and cellular component (CC) were analyzed, respectively. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

located in the cell-substrate junction and ribosome-related compartment (Figure 3(c)). Also, the results of KEGG enrichment analysis showed that the target genes are mainly involved in pathways of the ribosome, pathogenic *Escherichia coli* infection, and the insulin signaling pathway (Figure 3(d)).

Many previous works have studied the interactions between ribosomes and lncRNAs using ribosome profiling techniques, with a primary focus on probing lncRNAs interacting with ribosomes related to protein synthesis as well as other unclear biological functions. Several cytoplasmic lncRNAs have recently been reported to interact with ribosomes. In footprinting experiments to map ribosome-bound transcripts genome-wide, a considerable number of lncRNAs were identified directly involved in the translation machinery.

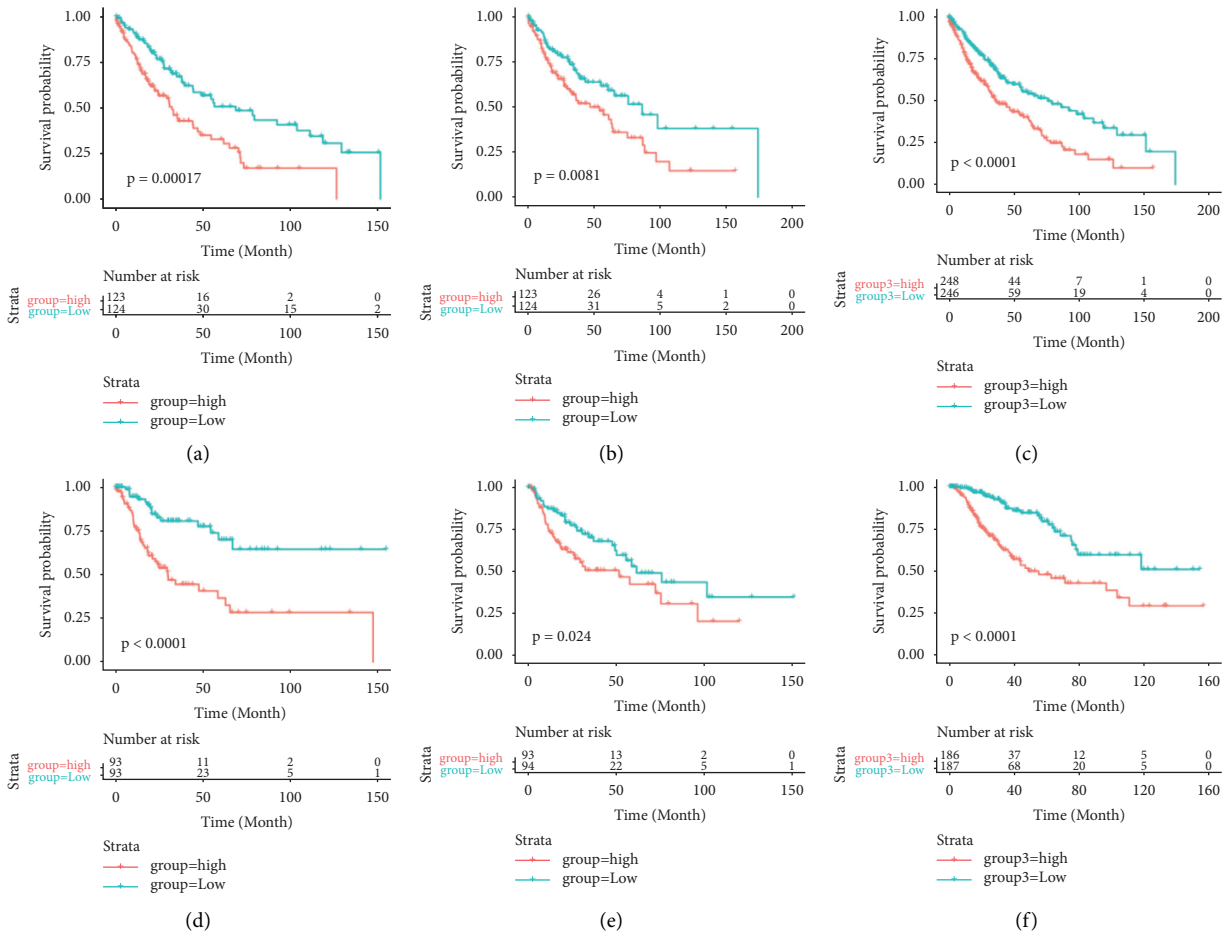



FIGURE 4: OS survival curve of high- and low-risk groups in the training set (a), test set (b), and entire set (c). DFS survival curve of high- and low-risk groups in the training set (d), test set (e), and entire set (f).

*3.3. Construction of Prognostic Models.* To build a prognostic model, samples with OS information were randomly divided into the training set and the test set at a ratio of 1 : 1. In the training set, LASSO-Cox regression analysis was used to calculate the 36 DElncRNAs related to OS, and 10 of them were considered as independent markers with significant prognostic value for LUSC (Figure S1). The coefficients and DElncRNAs are described as follows:

$$\begin{aligned} OS\_risk_{score} = {} & AC013457.1 * 0.156001 + AC124067.2 * (-0.068019) + AP001189.1 \\ & * 0.697599 + AP002360.1 * (-0.168550) + BANCR * (-0.569796) \\ & + LINC00519 * (-0.042434) + LINC01807 * (0.288462) \\ & + MIR3945HG * 0.278341 + (FAM83A - AS1) * (-0.007179) \\ & + (POU6F2 - AS2) * (-0.106658). \end{aligned} \tag{1}$$

Similarly, for samples with DFS information, they were randomly split into a training set and a test set of fifty-fifty. 40 DElncRNAs related to DFS were filtered in the training set, and 11 DElncRNAs significantly related to DFS were obtained (Figure S2). The coefficients and DElncRNAs of the DFS prognostic model are described as follows:

$$DFS\_risk_{score} = (FAM83A - AS1) * 0.23834 + AC010275.1 * (-0.08770) + AC015922.3 * 0.27021 + AL132712.1 * (-0.15284) + LINC00261 * 0.20817 + LINC00511 * (-0.02819) + LINC00519 * (-0.11217) + LINC01980 * (-0.07532) + TEM99 * (-0.49311) + MYOSLID * 0.34051 + (NUP50 - DT) * (-0.34856). \quad (2)$$

*3.4. Survival Analysis of the Prognostic Model.* To conduct the survival characteristics of the prognostic model related to OS, the risk scores were calculated for each sample according to the abovementioned formulas. Samples were categorized into high-risk and low-risk groups by the median of the prognostic risk scores, and the survival curve was mapped. The survival differences were calculated using the log-rank test. It revealed that, in the training set, the survival rate of samples in the high-risk group was significantly lower than that in the low-risk group ($P$ value <0.05, log-rank test) (Figures 4(a)–4(c)). Similar results were also observed in the test set and the entire set. We can draw the same conclusion for the DFS prognostic model when the same analysis procedure was performed. The survival rate of samples in the high-risk group was significantly shorter than that in the low-risk group in the training set, test set, and entire dataset ($P$ value <0.05, log-rank test) (Figures 4(d) and 4(e)).

*3.5. Correlation Analysis of the Prognostic Model and Clinical Characteristics.* We then evaluated the correlations between the risk scores and the clinical characteristics, including age, stage, gender, and TNM. The differences of risk score between groups of age, stage, gender, and TNM were calculated by the log-rank test. We observed that neither the OS risk score nor the DFS risk score has a significant correlation with the clinical characteristics (Tables S1 and S2), and no significant difference was found between groups of the stage (high and low), T (T0–2 and T3–4), N (N0 and N1–3), and M (M0 and M1–3) (Figure 5), where T refers to the size and extent of the primary tumor, N refers to the number of nearby lymph nodes that have cancer, and M refers to whether the cancer has metastasized. These results revealed that OS risk score and DFS risk score were independent predictors of survival risk for the clinical factors.

*3.6. Evaluation of the Efficiency of Prognostic Models.* The prognostic model of OS was used to compare with stage, gender, age, and TNM in the third year. The AUCs of the OS prognostic model in the training set, test set, and entire set were consistently higher than those of the clinical characteristics (Figures 6(a)–6(c)). We can draw the same conclusion for the first- and fifth-year samples (Figure S3). Similarly, the prognostic model of DFS was used to compare with stage, gender, age, and TNM in the first, third, and fifth years. The AUCs of the prognostic model of DFS in the training set, the test set, and the entire dataset were overall higher than those of other clinical characteristics (Figures 6(d)–6(f) and Figure S4). These results showed that the prognostic model of OS and DFS possessed better prognostic ability than the clinical characteristics.

Stromal cells are crucial components of TME, and the proportion of stromal cells in TME represents the stromal score. The tumor immune microenvironment plays a key role in the development of numerous cancers. The prognostic models of OS and DFS were built based on the RNA-seq data, for which the immune score and the stromal score for each sample can also be generated, using the estimate R package. For prognostic models of OS and DFS, immune scores and stromal scores of the high-risk group were significantly higher than those of the low-risk group in the training set, test set, and entire dataset ($P$ value <0.05, Wilcoxon test) (Figure 7).

## 4. Discussion

lncRNAs have been found to play an important role in many biological processes, including the onset and development of cancer [28–32], which intuitively could serve as prognostic markers for cancers. In this study, we leveraged the TCGA RNA-seq data to build prognostic lncRNA models to evaluate the clinical outcomes of LUSC patients [33]. We first screened the differentially expressed lncRNAs, and then, we picked up those with a significant prognostic value. Finally, we constructed two prognostic models using LASSO-Cox regression analysis for OS and DFS, respectively. Ten lncRNAs were determined with significant contribution to the OS prognosis of LUSC, including *AC013457.1, AC124067.2, AP001189.1, AP002360.1, BANCR, LINC00519, LINC01807, MIR3945HG, FAM83A-AS1,* and *POU6F2-AS2*. For the DFS prognostic signature, 11 lncRNAs were identified, including *FAM83A-AS1, AC010275.1, AC015922.3, AL132712.1, LINC00261, LINC00511, LINC00519, LINC01980, TMEM99, MYOSLID*, and *NUP50-DT*.

Importantly, two lncRNAs, *FAM83A-AS1 and LINC00519*, were commonly identified as prognosticators for both OS and DFS analysis. *FAM83A-AS1* has been reported to be a key role in NSCLC. For instance, a study found that the overexpression of *FAM83A-AS1* increased *FAM38A* protein levels and induced the downstream *ERK1/2* phosphorylation in cells. Moreover, the overexpression of *FAM83A-AS1* promoted LUAD cell proliferation and invasion, which was consistent with our results [34]. Recent studies demonstrated that *LINC00519* was upregulated in LUSC, and silenced *LINC00519* prohibited proliferation, migration, invasion, and stimulated apoptosis in the LUSC cells [35]. Additionally, lncRNA *BANCR* in the OS model has been reported to function as an oncogene or tumor

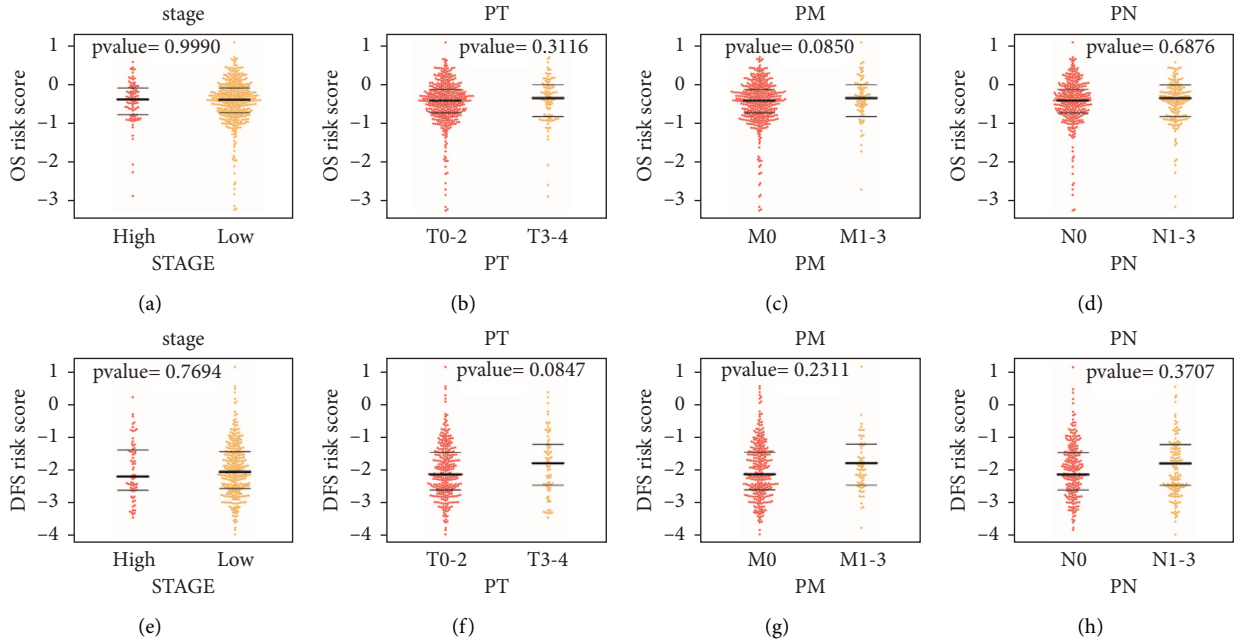



Figure 5: The distributions of the risk scores for OS in stage (a), PN (b), PM (c), and PT (d). The distributions of the risk score for DFS in stage (e), PN (f), PM (g), and PT (h). The bold line in the middle indicates the median, while the other two lines represent the first and third quantile of the distribution, respectively.

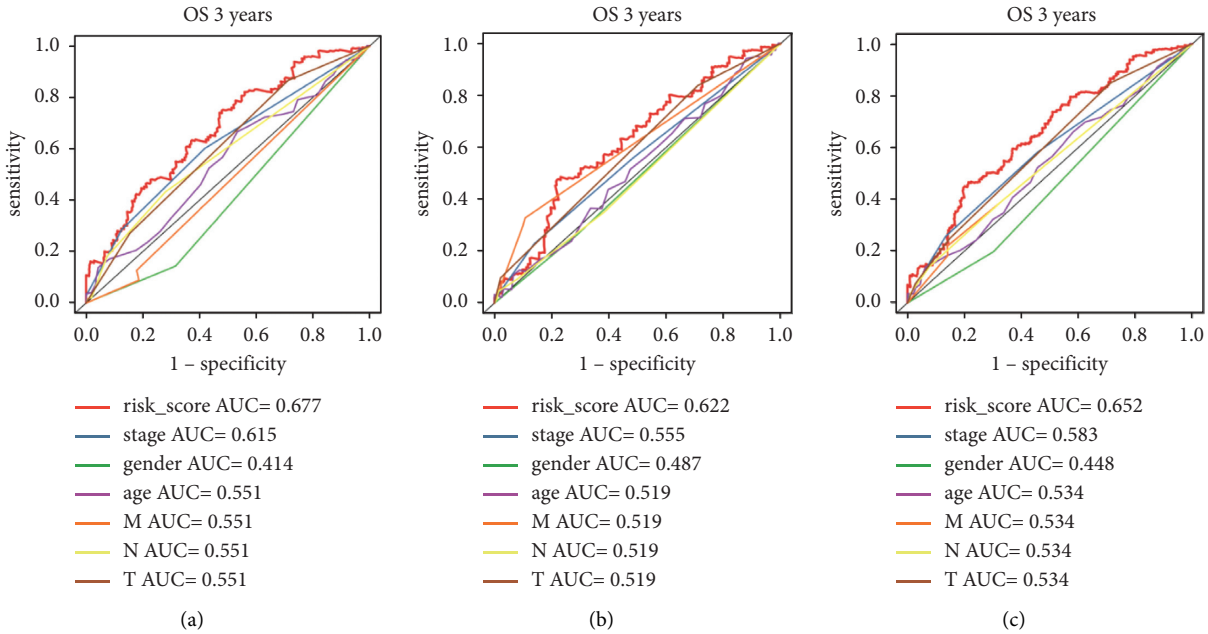



Figure 6: Continued.

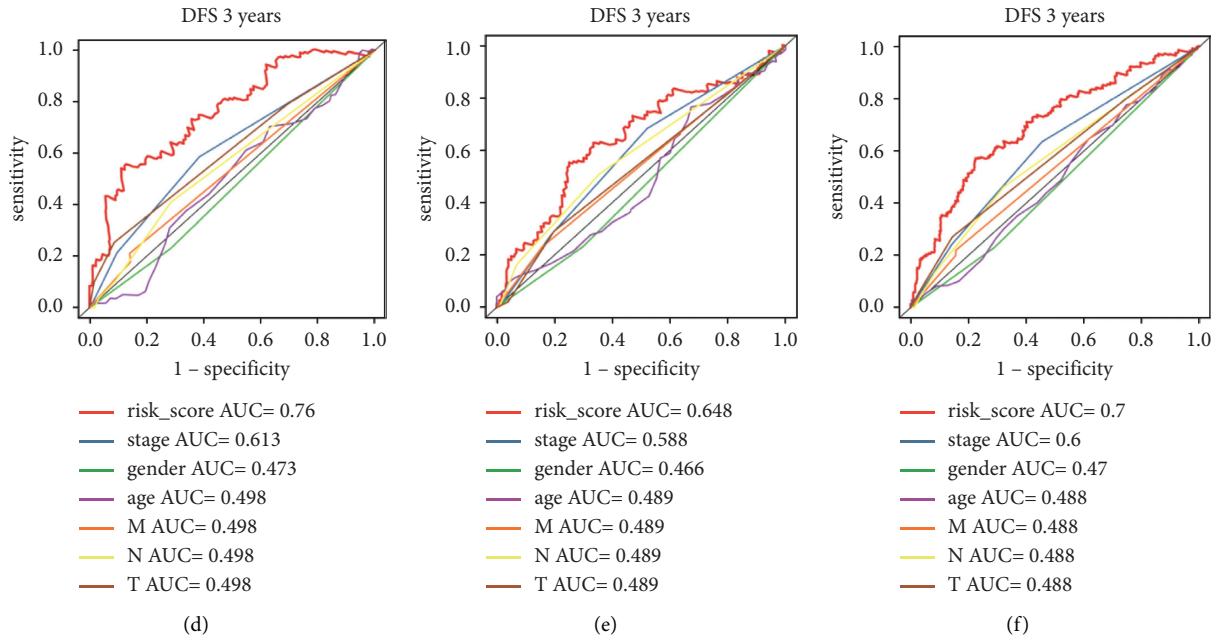



Figure 6: Performance evaluation of the models. ROC curves of the OS prognostic model in the training set (a), test set (b), and entire set (c) in the third year. ROC curves of the DFS prognostic model in the three sets (d–f).

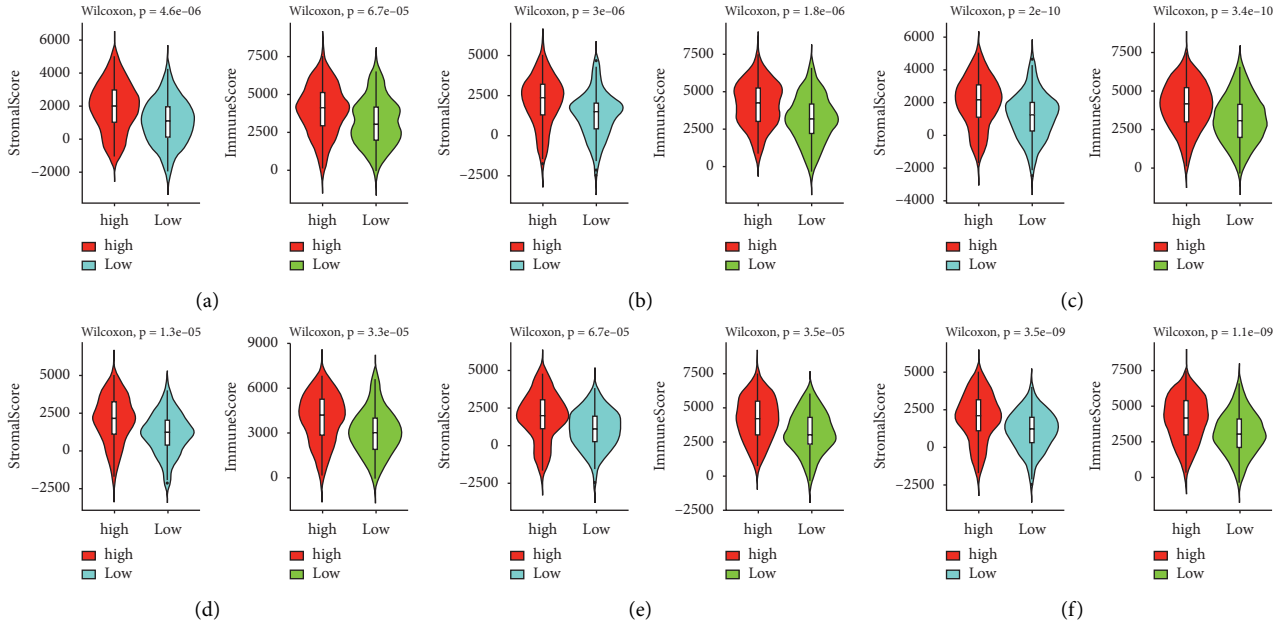



Figure 7: Difference of the stromal score and the immune score between the high- and low-risk group in OS in the training set (a), test set (b), and entire set (c). The difference of the stromal score and the immune score between the high- and low-risk groups in DFS (d–f).

suppressor gene, which was often dysregulated in human cancers, including lung cancer [36, 37]. Meanwhile, lncRNA *MYOSLID* in the DFS model was closely related to important biological processes and pathways that regulate cancer metastasis [38–40]. Notably, the coefficients of the two lncRNAs are the second biggest among the panels.

We used prognostic and differentially expressed lncRNAs to construct predictive OS and DFS models that were superior to other clinical indicators. The method can be applied to other diseases or symptoms, once sufficient gene expression datasets are provided. Although these results had certain clinical significance, some limitations must be noted. The sample numbers of the patient and normal control are imbalanced in this study, which influence the machine learning results and lead to low AUC scores. A widely adopted method for imbalanced classification is resampling, which consists of picking up a subset of samples from the large group (undersampling) or bootstrapping samples from

the small group (oversampling). For undersampling, the randomly chosen samples may be a biased sample set, resulting in an inaccurate representation of the population. Additionally, it can discard potentially useful information that could be important for model training. Unlike undersampling, oversampling leads to no information loss, but it increases the likelihood of overfitting because it replicates the samples in the small group.

Our models were built based on the public datasets and were not verified in external datasets. For clinical application and assessing the prognostic value of the proposed models, multicohort analysis integrating all the available LUSC expression data will be executed in our future studies. LINC00519 and FAM83A-AS1 were commonly identified as prognosticators in both models. The two lncRNAs have been reported to be associated with cell proliferation and invasion of lung cancer, suggesting the prognostic value of them, and further in vivo validation is warranted.

## Data Availability

The datasets used in this study are available in TCGA database.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Authors' Contributions

XZ and YS conceived of the idea and wrote the manuscript. XF, JX, GQ, and MY prepared the data and analyzed the results. XL and GC supervised this work. Xiaoting Zhang and Yue Su contributed equally to this study.

## Acknowledgments

The study was supported by the Shenzhen Project of Science and Technology (Grant nos. JCYJ20180302145109198 or JCYJ20210324123012034 to Z.X.T and JCYJ20190809094407602 to L.X.F), National Natural Science Foundation of China (81801517 to L.X.F), the fund of "San-ming" Project of Medicine in Shenzhen (No. SZSM201812088), and Scientific Research Foundation of Peking University Shenzhen Hospital (KYQD2021104).

## Supplementary Materials

Table S1. Correlation between the OS risk score and clinical characteristics. Table S2. Correlation between the DFS risk score and clinical characteristics. Figure S1. The optimal feature number of the LASSO-Cox regression model in the prognostic model of OS. Figure S2. The optimal feature number of the LASSO-Cox regression model in the prognostic model of DFS. Figure S3. Performance evaluation of the models and clinical characteristics. ROC curves of the OS prognostic model in year one, three, and five on different sets. Columns represent the observation year, while the rows represent the training set, test set, and entire set, respectively. Figure S4. Performance evaluation of the models and clinical characteristics. ROC curves of the DFS prognostic model in year one, three, and five on different sets. Columns represent the observation year, while the rows represent the training set, test set, and entire set, respectively. (*Supplementary Materials*)

## References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] A. J. Alberg, M. V. Brock, J. G. Ford, J. M. Samet, and S. D. Spivack, "Epidemiology of lung cancer," *Chest*, vol. 143, no. 5, pp. e1S–e29S, 2013.

[3] P. M. de Groot, C. C. Wu, B. W. Carter, and R. F. Munden, "The epidemiology of lung cancer," *Translational Lung Cancer Research*, vol. 7, no. 3, pp. 220–233, 2018.

[4] C. Gridelli, A. Rossi, D. P. Carbone et al., "Non-small-cell lung cancer," *Nature Reviews. Disease Primers*, vol. 1, no. 1, pp. 15009–15016, 2015.

[5] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.

[6] C. G. A. R. Network, "Comprehensive genomic characterization of squamous cell lung cancers," *Nature*, vol. 489, no. 7417, p. 519, 2012.

[7] Y.-H. Lai, W.-N. Chen, T.-C. Hsu, C. Lin, Y. Tsao, and S. Wu, "Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning," *Scientific Reports*, vol. 10, no. 1, p. 4679, 2020.

[8] J. T. Y. Kung, D. Colognori, and J. T. Lee, "Long noncoding RNAs: past, present, and future," *Genetics*, vol. 193, no. 3, pp. 651–669, 2013.

[9] M. K. Iyer, Y. S. Niknafs, R. Malik et al., "The landscape of long noncoding RNAs in the human transcriptome," *Nature Genetics*, vol. 47, no. 3, pp. 199–208, 2015.

[10] X. Chen and G.-Y. Yan, "Novel human lncRNA-disease association inference based on lncRNA expression profiles," *Bioinformatics*, vol. 29, no. 20, pp. 2617–2624, 2013.

[11] T. Zhou and Y. Gao, "Increased expression of LncRNA BANCR and its prognostic significance in human hepatocellular carcinoma," *World Journal of Surgical Oncology*, vol. 14, no. 1, p. 8, 2016.

[12] J.-C. Guo, S.-S. Fang, Y. Wu et al., "CNIT: a fast and accurate web tool for identifying protein-coding and long non-coding transcripts based on intrinsic sequence composition," *Nucleic Acids Research*, vol. 47, no. W1, pp. W516–W522, 2019.

[13] T. R. Mercer, M. E. Dinger, and J. S. Mattick, "Long noncoding RNAs: insights into functions," *Nature Reviews Genetics*, vol. 10, no. 3, pp. 155–159, 2009.

[14] R. Zhang, L. Q. Xia, W. W. Lu, J. Zhang, and J.-S. Zhu, "LncRNAs and cancer," *Oncology Letters*, vol. 12, no. 2, pp. 1233–1239, 2016.

[15] A. Nandwani, S. Rathore, and M. Datta, "LncRNAs in cancer: regulatory and therapeutic implications," *Cancer Letters*, vol. 501, pp. 162–171, 2021.

[16] J.-C. Guo, Y. Wu, Y. Chen et al., "Protein-coding genes combined with long noncoding RNA as a novel transcriptome molecular staging model to predict the survival of patients with esophageal squamous cell carcinoma," *Cancer Communications*, vol. 38, no. 1, p. 4, 2018.

[17] T. Liu, Z. Han, H. Li, Y. Zhu, Z. Sun, and A. Zhu, "LncRNA DLEU1 contributes to colorectal cancer progression via activation of KPNA3," *Molecular Cancer*, vol. 17, no. 1, pp. 118–213, 2018.

[18] Y. Liang, X. Song, Y. Li et al., "LncRNA BCRT1 promotes breast cancer progression by targeting miR-1303/PTBP3 axis," *Molecular Cancer*, vol. 19, pp. 85–20, 2020.

[19] X. Kong, Y. Duan, Y. Sang et al., "LncRNA-CDC6 promotes breast cancer progression and function as ceRNA to target CDC6 by sponging microRNA-215," *Journal of Cellular Physiology*, vol. 234, no. 6, pp. 9105–9117, 2019.

[20] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge," *Contemporary Oncology*, vol. 19, no. 1A, pp. A68–77, 2015.

[21] J.-H. Li, S. Liu, H. Zhou, L.-H. Qu, and J.-H. Yang, "starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data," *Nucleic Acids Research*, vol. 42, no. D1, pp. D92–D97, 2014.

[22] W. J. Shangguan, H. T. Liu, Z. J. Que, F. F. Qian, L. S. Liu, and J. H. Tian, "TOB1-AS1 suppresses non-small cell lung cancer cell migration and invasion through a ceRNA network," *Experimental and Therapeutic Medicine*, vol. 18, no. 6, pp. 4249–4258, 2019.

[23] C. Gene Ontology, "The Gene Ontology resource: enriching a GOld mine," *Nucleic Acids Research*, vol. 49, no. D1, pp. D325–D334, 2021.

[24] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.

[25] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, "clusterProfiler: an R package for comparing biological themes among gene clusters," *OMICS: A Journal of Integrative Biology*, vol. 16, no. 5, pp. 284–287, 2012.

[26] J. Li, Z. Chen, L. Tian et al., "LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma," *Gut*, vol. 63, no. 11, pp. 1700–1710, 2014.

[27] S. Shen, G. Wang, R. Zhang et al., "Development and validation of an immune gene-set based Prognostic signature in ovarian cancer," *EBioMedicine*, vol. 40, pp. 318–326, 2019.

[28] B. Rathinasamy and B. K. Velmurugan, "Role of lncRNAs in the cancer development and progression and their regulation by various phytochemicals," *Biomedicine & Pharmacotherapy*, vol. 102, pp. 242–248, 2018.

[29] J. R. Prensner and A. M. Chinnaiyan, "The emergence of lncRNAs in cancer biology," *Cancer Discovery*, vol. 1, no. 5, pp. 391–407, 2011.

[30] M. C. Jiang, J. J. Ni, W. Y. Cui, B. Y. Wang, and W. Zhuo, "Emerging roles of lncRNA in cancer and therapeutic opportunities," *American journal of cancer research*, vol. 9, no. 7, pp. 1354–1366, 2019.

[31] K. Grillone, C. Riillo, F. Scionti et al., "Non-coding RNAs in cancer: platforms and strategies for investigating the genomic "dark matter"," *Journal of Experimental & Clinical Cancer Research*, vol. 39, no. 1, pp. 1–19, 2020.

[32] G. Hu, F. Niu, B. A. Humburg et al., "Molecular mechanisms of long noncoding RNAs and their role in disease pathogenesis," *Oncotarget*, vol. 9, no. 26, pp. 18648–18663, 2018.

[33] J. Man, X. Zhang, H. Dong et al., "Screening and identification of key biomarkers in lung squamous cell carcinoma by bioinformatics analysis," *Oncology Letters*, vol. 18, no. 5, pp. 5185–5196, 2019.

[34] R. Shi, Z. Jiao, A. Yu, and T. Wang, "Long noncoding antisense RNA FAM83A-AS1 promotes lung cancer cell progression by increasing FAM83A," *Journal of Cellular Biochemistry*, vol. 120, no. 6, pp. 10505–10512, 2019.

[35] P. Ye, X. Lv, R. Aizemaiti, J. Cheng, P. Xia, and M. Di, "H3K27ac-activated LINC00519 promotes lung squamous cell carcinoma progression by targeting miR-450b-5p/miR-515-5p/YAP1 axis," *Cell Proliferation*, vol. 53, no. 5, Article ID e12797, 2020.

[36] Y. Zou, J. Li, Y. Chen et al., "BANCR: a novel oncogenic long non-coding RNA in human cancers," *Oncotarget*, vol. 8, no. 55, pp. 94997–95004, 2017.

[37] X. Yu, H. Zheng, M. T. Chan, and W. K. K. Wu, "BANCR: a cancer-related long non-coding RNA," *American journal of cancer research*, vol. 7, no. 9, pp. 1779–1787, 2017.

[38] H. G. Xiong, H. Li, Y. Xiao et al., "Long noncoding RNA MYOSLID promotes invasion and metastasis by modulating the partial epithelial-mesenchymal transition program in head and neck squamous cell carcinoma," *Journal of Experimental & Clinical Cancer Research : Climate Research*, vol. 38, no. 1, pp. 278–314, 2019.

[39] Y. Luo, J. Ye, J. Wei, J. Zhang, and Y. Li, "Long non-coding RNA-based risk scoring system predicts prognosis of alcohol-related hepatocellular carcinoma," *Molecular Medicine Reports*, vol. 22, no. 2, pp. 997–1007, 2020.

[40] C.-S. Lei, H.-J. Kung, and J.-W. Shih, "Long non-coding rnas as functional codes for oral cancer: translational potential, progress and promises," *International Journal of Molecular Sciences*, vol. 22, no. 9, p. 4903, 2021.