Immune Score-based Molecular Subtypes and Signature Associated with Clinical Outcome in Hepatoblastoma

Yongping Huang[1], Jinlong Yan[2], Ruiqi Liu[3], Guang Tang[1], Qi Dong[4], Dong Liu[1], Xiaolan Yang[1], Shouhua Zhang[1, *] and Dejun Tang[5, **]

[1]Department of General Surgery, Longgang District People's Hospital of Shenzhen, Shenzhen, Guangdong, China
[2]Department of General Surgery, Second Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China
[3]WenZhou Medical University, WenZhou, China
[4]Department of General Surgery, Jiangxi Provincial Children's Hospital, Nanchang, Jiangxi, China
[5]Department of General Surgery, Longgang District People's Hospital of Shenzhen, Shenzhen, Guangdong, China

*Corresponding author: Department of General Surgery, Jiangxi Provincial Children's Hospital, Nanchang, 330006, Jiangxi, China. Email: zshouhua416@163.com
**Corresponding author: Department of General Surgery, Longgang District People's Hospital of Shenzhen, Shenzhen, 518000, Guangdong, China. Email: tangdejun97@163.com

Abstract

Background: This study aimed to identify genes related to the immune score of hepatoblastoma, examine the characteristics of the immune microenvironment of hepatoblastoma, and construct a risk scoring system for predicting the prognosis of hepatoblastoma.

Methods: Through using the gene chip data of patients with hepatoblastoma with survival data in the ArrayExpress and GEO databases, the immune score of hepatoblastoma was calculated by the ESITIMATE algorithm, and the prognostic value of immune score in patients with hepatoblastoma was studied by the survival analysis. Genes related to the immune score were identified by the WGCNA algorithm. According to these genes, patients with hepatoblastoma were clustered unsupervised. Finally, the risk scoring system was constructed according to the immune score-related genes.

Results: The immune score calculated by the ESTIMATE algorithm had a good prognostic value in patients with hepatoblastoma. Patients with high immune scores had better OS than those with low immune scores (P < 0.001). A total of 146 immune score-related genes were identified by WGCNA analysis, and univariate COX regression analysis indicated that 59 of the genes had prognostic value. According to the unsupervised clustering results of the 146 immune score-related genes, patients with hepatoblastoma could be divided into two subtypes with different prognoses, namely molecular subtype 1 and subtype 2, with molecular subtype 1 having a better prognosis. The immunocyte infiltration analysis results showed that the difference between the two subtypes was mainly in activated CD4 T cells, activated dendritic cells, CD56 bright natural killer cells, the macrophage, and regulatory T cells. According to the immune score-related genes, a risk scoring system was constructed based on a five-gene signature. After the cut-off value was determined, patients with hepatoblastoma were divided into a high-risk group and a low-risk group. The prognosis of the two groups was different.

Conclusions: The immune score has a good prognostic value in patients with hepatoblastoma. Based on the different expression patterns of immune score-related genes, hepatoblastoma can be divided into two different prognostic molecular subtypes, showing different immunocyte infiltration patterns. The established risk scoring system based on a five-gene signature has a good predictive value in patients with hepatoblastoma.

Keywords: Hepatoblastoma, Prognosis, Immune Microenvironment, Model, Immune Cell

1. Background

Hepatoblastoma, which originates from hepatic primordial embryonic cells (1-3), is the most common primary epithelial malignant tumor in children. The incidence rate of hepatoblastoma has been increasing in the past 30 years (4), and the treatment of hepatoblastoma includes chemotherapy, surgical resection, and liver transplantation (5-7). At present, surgical resection is still the first-line treatment for the initial diagnosis of hepatoblastoma. With the development of these treatments, the curative effect of hepatoblastoma has made great progress, and the survival outcome has also been greatly improved (8, 9). In the past ten years, immunotherapy has made a great breakthrough in cancer treatment, showing good

therapeutic effect in various malignant tumors, especially in melanoma, lung cancer, urothelial carcinoma, head and neck squamous cell carcinoma, renal cell carcinoma, and Hodgkin's lymphoma ([10], [11]). However, there is still a great need to investigate hepatoblastoma, and a clear understanding of the tumor immune microenvironment will contribute to the application of this treatment in hepatoblastoma.

Tumor microenvironment (TME) is mainly composed of stromal cells and immune cells. An increasing number of studies have shown that TME plays a crucial role in the progression and treatment of many cancers ([12-14]). The immune cells include cytotoxic T cells, helper T cells, dendritic cells (DCS), tumor-associated macrophages, and mesenchymal stem cells, and the composition of these cells determines the prognosis of a variety of tumors. Several studies based on the gene expression profile data of the immune-related score have predicted the prognosis of the tumor and have been applied in a variety of tumors ([15-18]). Therefore, TME and immune-related genes may play an essential role in tumor therapy. Recently, algorithms based on gene expression profiles have made great progress, which can evaluate the composition of various cellular components in TME ([19-21]).

In this study, we first determined the immune score of hepatoblastoma by the ESTIMATE algorithm ([21]) and identified immune-related genes by WGCNA ([22]). We performed consensus cluster analysis on hepatoblastoma based on the expression characteristics of these immune genes and analyzed the immunocyte infiltration by the ssGSEA algorithm ([23]). Finally, we constructed a prognostic risk scoring system based on these immune score-related genes, which could be used to evaluate the prognosis of patients with hepatoblastoma.

2. Methods

2.1. Hepatoblastoma Data Sets and Preprocessing

We searched the GEO and ArrayExpress databases to find the hepatoblastoma data set with survival information published in the public database. Finally, a total of two data sets, ie, GSE75271 and E-MEXP-1851, were retrieved ([24], [25]). The analysis platform used for both groups was the expression chip of Affymetrix. The former was the hgu133plus2 chip, containing five normal tissue chip expression profiles and 50 tumor tissue chip expression profiles, and the latter was the hgu133a chip, containing four normal tissue chip expression profiles and 24 tumor tissue chip expression profiles. After downloading the original

chip data, the RMA algorithm in the R package AFFY was used for background correction, data standardization, and other processing to obtain the expression matrix. We converted the probe ID into a gene symbol and then combined the data according to the gene symbol shared by the two expression matrices to obtain an expression matrix containing 83 samples and 12402 genes. Further, the combat function of the sva package was used to correct batch effect and remove normal control samples, and a total of 68 samples with survival information were used for subsequent analysis.

2.2. Inference of Sample Immune Score

The ESTIMATE algorithm is a tool for predicting tumor purity. It uses gene expression data to predict the infiltration degree of stromal and immune cells in tumor tissue. It obtains three kinds of scores, namely stromal score, immune score, and estimate score, representing the proportion of stromal in tumor tissue, immune cell infiltration, and tumor purity, respectively. The calculated immune score was used to search for an optimal cut-off by the survminer package, and the Kaplan-Meier method was used to evaluate the prognostic value of the immune score in patients with hepatoblastoma.

2.3. Identification of Co-expression Network Modules and Immune-related Modules

The WGCNA software package was used to build co-expression networks. The algorithm first calculated Pearson's correlation coefficient for each gene and used its absolute values to construct the gene expression similarity matrix. The optimal β-value was selected to construct the proximity matrix so that our gene distribution fitted the connection-based scale-free network. The adjacent matrix and the topological matrix were obtained based on the β-value, and the distance between the genes was represented by the dissimilarity between the genes calculated by the topological overlap matrix (TOM). Then, the gene cluster tree was divided into different modules (the minimum number of genes in each module was 30). A power of $\beta = 7$ and a scale-free $R^2 = 0.91$ were set as soft-threshold parameters to ensure a signed scale-free co-expression gene network. Eigengene connectivity was the correlation between a gene's expression profile and the module eigengene. We kept genes whose connectivity to their module gene was greater than 0.8 to increase the stability of the module, and a total of 2489 genes were filtered and then analyzed by WGCNA. MEDissThres was set to 0.7 to merge similar modules.

The hierarchical clustering module closely related to the immune score was selected as the module for subsequent analysis, and gene significance (GS), module significance (MS), and module eigengene (ME) were calculated. GS was defined as the correlation between gene expression and clinical information, calculated by the $\log_{10}$ conversion of p-value in linear regression. MS was the average importance of all genes in the module. ME was the first principal component obtained through the principal component analysis of the gene expression matrix of each module, which represented the value of the gene expression profile in the module. Univariate COX analysis was performed on all genes in the module to further evaluate the prognostic value of each gene in the immune score-related module.

2.4. Identification of Molecular Subtypes

The unsupervised clustering algorithm was used to identify the expression patterns of black module gene expression, and finally, the number and stability of the categories were determined. This step was completed by the "CancerSubtyeps" package (26). Then, the ConsensusClusterPlus algorithm (27) was applied, and 1000 operations were repeated to ensure the stability of the results. The Kaplan-Meier method was used to evaluate the prognosis of the two subtypes.

2.5. Analysis of Differentially Expressed Genes

Differentially expressed genes (DEGs) were analyzed using the "Limma" package (28). The software package used the empirical Bayes method and improved *t*-test to analyze gene expression changes of the molecular subtypes. The Benjamini-Hochberg method was used to correct the corrected P-values for multiple tests. Genes with a corrected P-value < 0.05 and |log FoldChange| > 1 were identified as differential genes between the molecular subtypes.

2.6. Immune Cell Infiltration in Tumor Microenvironment

The ssGSEA was introduced to calculate the relative infiltration of immune cells in TME. Markers of immune cells were obtained from recently published literature (29). The relative abundance of each type of immune cell was expressed by the enrichment fraction in ssGSEA analysis. Then, the enrichment fraction was standardized, with 0 being the lowest abundance and 1 the highest abundance.

2.7. Functional Annotation and PPI Network Construction

For inferring the potential biological functions of genes in the immune-related modules, the clusterProfiler package was used for gene ontology (GO) enrichment analysis and the Kyoto encyclopedia of genes and genomes enrichment analysis. The threshold was set to P < 0.05 after adjustment. At the same time, the identified genes in the immune-related modules were used to construct PPI gene network interaction analysis by Metascape (30). The Metascape database searches for known and predicted protein interactions and studies interaction networks between proteins to help identify core regulatory genes with the highest MCODE score performed by screening with MCODE (31).

2.8. Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) (32) was used to identify up-regulated or down-regulated gene pathways between different molecular subtypes. All the genes were ranked from large to small according to log-fc (log fold change) after difference analysis as the input gene set, and weighted enrichment statistics were used to calculate the enrichment fraction of each gene set. One thousand phenotype permutations were used to evaluate significance. The reference gene set was the Hallmark gene set (H.A.V7.0 data set downloaded from MsigDB database), and the gene sets were defined as significant at the 5% level with a false discovery rate under 25%.

2.9. Prognostic Gene Signature-based Risk Score

The whole cohort was divided into training set and verification set according to 7:3. Then, the L1 regularization (lasso) of the glmnet package of R language was used to fit the Cox-PH model of immune score-related genes to determine the gene signature for prognosis. L1 regularization (lasso) is a useful method to determine the interpretable prediction rules in high-dimensional data (33). The optimal lambda value was selected through 1000 cross-validations, and a set of prognostic genes were identified. According to the expression levels of these prognostic genes and their regression coefficients from the COX-PH model, an equation for calculating risk score was generated as follows:

Risk score $= \beta$gene 1 $\times$ exprgene 1 $+ \beta$gene 2 $\times$ exprgene 2 $+\cdots+ \beta$gene n $\times$ exprgene n

The risk score was calculated and assigned to each patient in the training group. The survminer R package was used to determine the best cut-off value, and all patients in the training set were divided into a high-risk group and

a low-risk group. The overall survival (OS) time of the two risk groups was compared by Kaplan-Meier survival analysis and the log-rank test. The robustness of the risk scoring system was verified in the validation set and the entire cohort. Kaplan-Meier survival analysis and the log-rank test were used to analyze OS time between the risk groups.

2.10. Statistical Analysis

The Shapiro Wilk test was used to assess the normality of the variables. Statistical significance differences among the normally distributed variables were estimated using the unpaired Student *t*-test, and the non-normally distributed variables were analyzed using the Mann-Whitney U test. The Kruskal-Wallis and one-way ANOVA tests were used as nonparametric and parametric methods for three or more data sets, respectively. The correlation was calculated by Pearson's correlation coefficient. The Kaplan-Meier method was used to calculate the survival rate, and the log-rank test was used to determine the significance of differences between the survival curves. Regarding the heterogeneity among different types of cancer, the optimal cut-off value of each continuous prognostic marker was recalculated using the survminer R package for different tumor types. Univariate and multivariate analyses were performed using the Cox proportional risk model. The survival prediction accuracy of the prediction model was evaluated according to time-dependent receiver operating characteristic curve (ROC) analysis. All statistical analyses were performed using the R package (version 3.6.3), with P-values of two tails, and the statistical significance was set as å = 0.05.

3. Results

3.1. High Immune Score in Hepatoblastoma Benefiting Survival of Patients

We first combined the data and removed the effect of batch processing, as shown in Appendix 1 in Supplementary File. We first analyzed four kinds of scores for hepatoblastoma using the ESTIMATES package to clarify the relationship between the immune score and the survival of patients with hepatoblastoma. The basic information and immune score values of the patients are shown in Appendices 2 and 3 in Supplementary File. We realized that patients with high immune scores had a better prognosis than those with a low immune score. These results suggested that high immune scores were associated with longer survival in hepatoblastoma (Figure 1A).

3.2. Identification of a Gene Signature Associated with Immune Score

The immune score-related genes were identified and obtained by WGCNA analysis. The genes were clustered into 13 modules (Figure 1B). Pearson's correlation coefficient (Figure 1C) was used to express the correlation between modules and immune scores. The black module showed the highest correlation with the immune score (cor: 0.99, P < 0.001).

The diagram of module membership and gene importance illustrated the significant correlation for each gene in the black module (cor: 0.99, Figure 1D). Then, univariate Cox regression analysis was used to analyze each gene in the black module, and 59 genes significantly associated with survival in patients with hepatoblastoma were identified (Figure 1F). The heat map showed the relative expression of 146 genes in the black module (Figure 1E). We defined these 146 genes as immune score-related genes, and their expressions are shown in Appendix 4 in Supplementary File.

3.3. GO Analysis and Protein-Protein Interaction Analysis for Immune Score Related Genes

GO analysis revealed that T cell activation, antigen processing and presentation, T cell differentiation, and other immune-associated pathways were associated with the immune score-related genes (Figure 2A). Protein-protein interaction (PPI) enrichment was done among the list of immune score-related genes. The PPI network represented interactions between proteins. The PPI diagram of the input genes (Figure 2B) showed that CD247, LCK, HLA-DPA1, and HLA-DPB1 were the core of the network, and these genes were mainly involved in the immune response cell surface receptor signaling pathway and cell-cell adhesion signaling pathway regulation. The molecular complex detection (MCODE) method was applied to identify closely related proteins from the PPI network. The MCODE algorithm subclustered the PPI network into five subclusters, and five MCODE components were made (Figure 2C).

3.4. Identification of Molecular Subtypes Based on Immune Score Related Genes in Hepatoblastoma
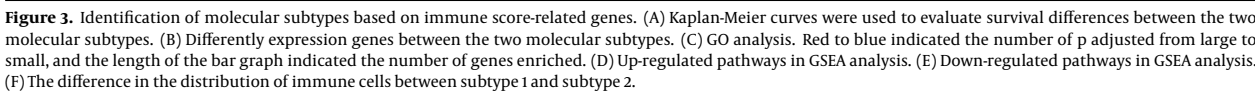
Two different expression patterns of these genes could be observed in the heat map of the expression profile of the immune score-related genes in hepatoblastoma. Therefore, we wondered whether these immune score-related genes could be used to distinguish the molecular subtypes of hepatoblastoma. We applied the consistent clustering method to unsupervised clustering based on the immune
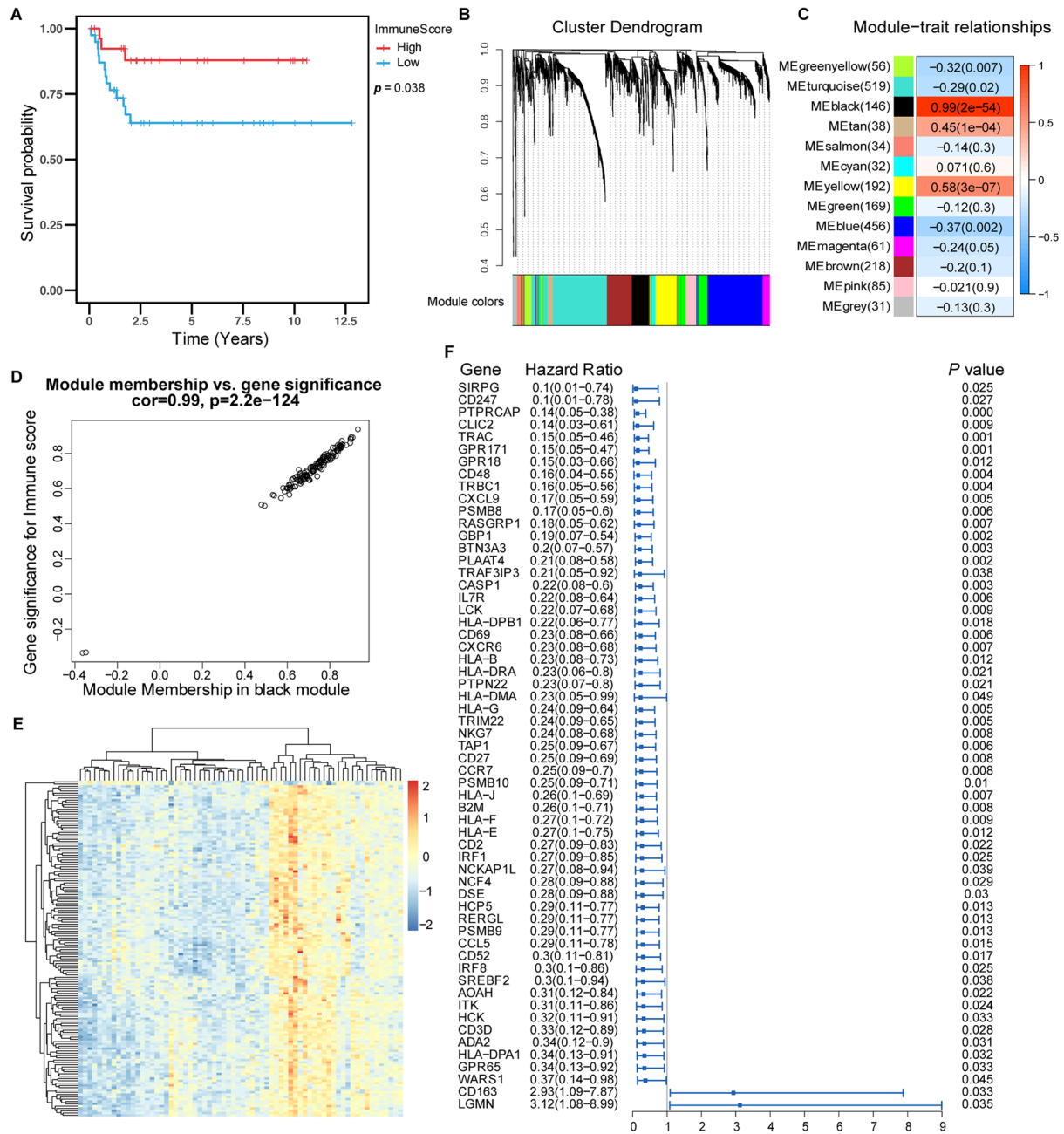
Figure 1. Prognostic value of immune score and identification of immune-related genes. (A) Kaplan-Meier curves showed that in the GSE75271, E-MAXP-1851 cohort, patients with higher immune scores had longer OS than patients with lower immune scores. (B) WGCNA identified 13 modules by unsupervised clustering. (C) The black module had the highest correlation with the immune score (r = 0.99, P = 2e-54), and the black module gene was called the immune score-related gene. (D) The gene significance and module membership of the genes in the black module exhibited a high correlation. (F) The forest plot with a hazard ratio for the genes of the univariable model in the black module. The hazard ratio below one indicated that a gene was negatively associated with the event probability and thus positively associated with survival time. The box size was based on precision, and the x-axis had a logarithmic scale (a bigger box size represented a more precise confidence interval (95% CI)).

Figure 2. GO annotation and protein-protein interaction of immune score-related genes. (A) GO analysis was performed based on the 146 immune score-related genes. (B) The PPI network of immune score-related genes. (C) The MCODE algorithm was applied to this network to identify neighborhoods where proteins were densely connected. Each of the five colors represents five different MCODES.

score-related genes in hepatoblastoma and identified two molecular subtypes, subtype 1 and subtype 2. The two molecular subtypes had different prognostic characteristics (P = 0.0036) (Figure 3A), and the average silhouette width was used to evaluate the clustering effect of the samples (Appendix 1 in Supplementary File). These results suggested that these immune score-related genes could classify hepatoblastoma into two molecular subtypes (subtype 1 and subtype 2) with different prognostic and molecular characteristics, and patients with molecular subtype 1 had better survival than those with molecular subtype 2.

Differential genes of the two molecular subtypes were identified by differential gene analysis, and the heat map showed the expression of differential genes in the two molecular subtypes (Figure 3B). Then, DEGs were subjected to GO analysis (Figure 3C). The results revealed enrichments in the small molecule catabolic process, the carboxylic acid biosynthetic process, the organic acid biosynthetic process, and the organic acid catabolic process. GSEA was performed on molecular subtype 1 and subtype 2 hepatoblastoma. Up-regulated pathways included pathways related to cell cycle, DNA replication, and P53 signaling pathway in subtype 2 (Figure 3D). Down-regulated

pathways included pathways related to drug metabolism cytochrome p450, fatty acid metabolism, and peroxisome in subtype 2 (Figure 3E). The immune cell population distribution in subtype 1 and subtype 2 further illustrated different tumor immune microenvironments in the two molecular subtypes of hepatoblastoma (Figure 3F). Among all immune cell populations, the macrophage and the activated CD4 T cell showed the most significant difference between subtype 1 and subtype 2, but most immune cells did not show much difference (Figure 3F).

3.5. Development and Validation of a Risk Scoring System Based on Immune Score Related Genes

Based on the expression of these immune score-related genes in the training set and the Cox-PH model of LASSO, five gene signatures, including CXCL9, PSMB8, MYO1F, GZMK, and FAM49A, significantly related to the survival rate were identified. The risk score for each patient was calculated using the following formula:

Risk score = -0.2187 × expression of CXCL9 + 0.1361 × expression of GZMK + -0.7372 × expression of PSMB8 + 0.8170 × expression of FAM49A + -0.1239 × expression of MYO1F.

The risk score and survival of all patients in the whole cohort of hepatoblastoma and the expression of five genes

Figure 3. Identification of molecular subtypes based on immune score-related genes. (A) Kaplan-Meier curves were used to evaluate survival differences between the two molecular subtypes. (B) Differently expression genes between the two molecular subtypes. (C) GO analysis. Red to blue indicated the number of p adjusted from large to small, and the length of the bar graph indicated the number of genes enriched. (D) Up-regulated pathways in GSEA analysis. (E) Down-regulated pathways in GSEA analysis. (F) The difference in the distribution of immune cells between subtype 1 and subtype 2.

are shown in Figure 4D. All patients in the training set were divided into a high-risk group and a low-risk group according to the optimal cut-off value. Compared to the low-risk group, the high-risk group had a significantly shorter OS time (Figure 4A, log-rank, P < 0.0001). In the internal validation set, we tested the prognostic performance of the risk scoring system based on the above five gene signature. Patients in the internal validation set were divided into a high-risk group and a low-risk group according to the above cut-off value. Similarly, OS time was significantly longer in the low-risk group than in the high-risk group (Figure 4B, log-rank, P = 0.003). Finally, we validated the risk scoring system in the whole hepatoblastoma cohort. Similarly, the OS time of the high-risk group was shorter than that of the low-risk group (Figure 4C, log-rank, P < 0.0001). At the same time, we evaluated the risk scoring system to judge the prognosis of OS in the whole cohort. The results showed that the receiver operating characteristic (ROC) curve illustrated a high accuracy rate with the area under the curve of 1-year OS, 3-year OS, and 5-year OS reaching 0.84, 0.83, and 0.81, respectively (Figure 4E). Furthermore, the risk score was negatively correlated with the immune score (r = -0.4, P < 0.001), (Figure 4F). We conducted correlation analysis with quantified immune cells to further conclude the role of the risk score. The results showed that the immune score was negatively correlated with the activated CD8 T cell and the type 1 T helper cell, and positively correlated with neutrophil, suggesting that the risk score was correlated with the components of immune cells in the immune microenvironment (Figure 4G).

4. Discussion

In the past ten years, tumor-related microenvironment and immunotherapy have made great progress. Hepatoblastoma, as the most common liver malignant tumor in children and newborns, has not been reported on a large scale in this field, which is related to the low incidence of hepatoblastoma and its relatively few cases. Based on the data mining of previously published microarray data, we conducted a series of data analyses on the microenvironment and immune-related genes of hepatoblastoma. The composition of TME and molecular subtypes was preliminarily clarified based on the immune score-related genes of hepatoblastoma. Two molecular subtypes with different prognoses were identified, and the composition of different immune cells in the TME of the two molecular subtypes was discussed. Finally, a prognostic risk score was fitted. According to the prognostic risk score, patients with hepa-

toblastoma were divided into a high-risk group and a low-risk group. As a result, the prognosis score was negatively correlated with the immune score, the activated CD8 T cell, and the type 1 T helper cell but positively correlated with neutrophil.

Our results showed that the immune score calculated by the ESTIMATE algorithm had a prognostic value in hepatoblastoma. Patients with high immune scores had relatively good prognoses. Previous studies also confirmed that the immune score obtained by the ESTIMATE algorithm had a prognostic value in a variety of malignant tumors, such as colorectal cancer (34, 35), lung adenocarcinoma (36), breast cancer (37), prostate cancer, and liver cancer (38, 39), which is consistent with our analysis results. Our results further confirmed that the immune score had a good prognostic value in a variety of tumors. However, the calculation of the immune score involves a large number of genes, which usually requires the microarray analysis or second-generation sequencing of tumor tissue samples, which limits the clinical application of this index.

Therefore, WCGNA analysis was used to identify the most relevant modules of the immune score. The functional enrichment analysis of genes in the black module indicated that most of these genes were involved in immune-related cellular pathways. Protein interaction analysis showed obvious interaction among these genes. The MCODE results further indicated that some genes were involved in the process of antigen presentation, such as HLA-A, HLA-B, and HLA-C. Some genes were mainly concentrated in the chemotactic process of cells, and studies have shown that these cytokines are involved in the recruitment of immune cells from TME. Univariate Cox regression analysis showed that most of these genes had prognostic values, indicating that the immune score-related genes identified by WGCNA played an essential role in hepatocytes.

We obtained two molecular subtypes of hepatoblastoma with different prognoses by congruent cluster analysis to further investigate the role of these immune score-related genes. The different prognoses of the two subtypes further suggested that the different expression patterns of these immune score-related genes impacted the survival of patients with hepatoblastoma. The results of immune infiltration analysis showed that the different expression patterns of the immune score-related genes were associated with different expressions in infiltrating immune cells. These results supported the importance of the gene sets identified in hepatoblastoma. Two hepatoblastoma molecular subtypes with different prognoses showed no significant differences in the analysis of immune cell infil-

tration, which mainly focused on the activated CD4 T cell, the activated dendritic cell, the CD56 bright natural killer cell, the macrophage, and the regulatory T cell. Our results also confirmed, to a certain extent, that different patterns of immune cell infiltration impacted the prognosis of patients with hepatoblastoma.

The variables were screened by LASSO, and CXCL9, PSMB8, MYO1F, GZMK, and FAM49A were selected to construct a risk scoring system based on immune scoring genes. Patients with hepatoblastoma could be divided into two groups with different prognoses and risk scores according to the calculated best cut-off value of the risk score. The stability of the scoring system was also confirmed in the internal validation set. Finally, the prognostic value of the risk scoring system was verified across the entire cohort, and the robustness was further verified. Among the screened genes, PSMB8 was confirmed to promote the occurrence and metastasis of gastric cancer and was a potential biomarker for predicting the poor prognosis (40). PSMB8 is closely related to the migration, proliferation, and apoptosis of glioma cells and can be used as a new prognostic indicator of glioma (41). CXCL9 is produced by macrophages, endothelial cells, hepatocytes, and tumors. As a CXCR3 ligand, CXCL9 mainly acts as a chemokine that activates immune cells, including T cells and natural killer (NK) cells (42). CXCL9 is expressed in a variety of tumors, and its biological functions are diverse. Recently, studies by Fukuda et al. have shown that CXCL9 can be used as a prognostic indicator of intrahepatic cholangiocarcinoma (43). The upregulation of CXCL9 might provide a therapeutic strategy for intrahepatic cholangiocarcinoma expressing CXCL9 by enhancing anti-tumor immune monitoring. GZMK gene products are members of a group of related serine proteases in cytotoxic lymphocyte cytoplasmic granules. They are involved in the biological processes of cytolytic T lymphocytes and natural killer (NK) cells to recognize, bind and lyse specific target cells. At present, little is known about the role of FAM49A. It has been reported that CYRI-A, the post-translational product of FAM49A, is a dynamic regulator of large-scale pinocytosis, and it adjusts integrin together with CYPTI-B (the post-translational product of FAM49B, a family gene of FAM49A) (44). It has been reported that the mutation of MYO1F can increase the tumorigenicity of cells in vitro, which is characterized by accelerated growth and enhanced invasion. In thyroid cancer, the mutation of MYOF can lead to tumor proliferation (45). Our risk scoring system based on these genes can well predict the prognosis of patients with hepatoblastoma and verify that these cells might play an essential role

in hepatoblastoma. However, further rigorous biological experiments are needed.

This study analyzed the prognostic value of the immune score in hepatoblastoma and identified genes related to the immune score according to the previously published hepatoblastoma chip data. According to the different expression patterns of these genes, two different molecular subtypes were identified and showed different patterns of immune cell infiltration. Finally, we constructed a risk scoring system based on five genes, which could be used to predict the prognosis of patients with hepatoblastoma. However, because hepatoblastoma was a rare disease, we could not further validate the findings in a larger case cohort. Our study was based on the data of previous studies. Thus, it was necessary to collect more specimens from patients with hepatoblastoma and complete clinical data for verification.

Supplementary Material

Supplementary material(s) is available here [To read supplementary materials, please refer to the journal website and open PDF/HTML].

Footnotes

Figure 4. Immune risk scoring system predicted OS in patients with hepatoblastoma. (A-C) Kaplan-Meier curves were used to evaluate the impact of the immune risk score on OS in the training set, the internal validation set, and the entire cohort. The red curves represented the high-risk score, and the blue curves represented the low-risk score. (D) The five-gene signature-based immune risk score in the prognosis of overall survival in the whole data. The black dot plots represented the distribution of immune risk scores, the blue and red dot plots represented the survival status of patients with hepatoblastoma, and the heat maps represented the expression of five genes (E). An ROC was used to evaluate the predictive ability of the immune risk scoring system in patients with hepatoblastoma in 1, 3, and 5 years. AUC, the area under the ROC curve. (F) The correlation between the risk score and the immune score. (E) The correlation between the immune risk score and the ssGSEA score of immune cells. The X-axis was the $-\log_{10}$P-value of the correlation coefficient. The lower right quadrant represented P < 0.05 and positive correlation, while the upper right quadrant represented P < 0.05 and negative correlation. The red dots represented immune cells with anti-tumor effects, the blue dots represented immune cells with protective effects on tumor cells, and the green dots represented cells with unclear effects on tumor cells.

References

1. Bosch FX, Ribes J, Diaz M, Cleries R. Primary liver cancer: Worldwide incidence and trends. *Gastroenterology.* 2004;**127**(5 Suppl 1):5–16. doi: 10.1053/j.gastro.2004.09.011. [PubMed: 15508102].

2. Ross JA, Gurney JG. Hepatoblastoma incidence in the United States from 1973 to 1992. *Med Pediatr Oncol.* 1998;**30**(3):141–2. doi: 10.1002/(sici)1096-911x(199803)30:3<141::aid-mpo1>3.0.co;2-h.

3. Fuchs J, Rydzynski J, Von Schweinitz D, Bode U, Hecker H, Weinel P, et al. Pretreatment prognostic factors and treatment results in children with hepatoblastoma: a report from the German Cooperative Pediatric Liver Tumor Study HB 94. *Cancer.* 2002;**95**(1):172–82. doi: 10.1002/cncr.10632. [PubMed: 12115331].

4. Spector LG, Birch J. The epidemiology of hepatoblastoma. *Pediatr Blood Cancer.* 2012;**59**(5):776–9. doi: 10.1002/pbc.24215. [PubMed: 22692949].

5. Perilongo G, Malogolowkin M, Feusner J. Hepatoblastoma clinical research: lessons learned and future challenges. *Pediatr Blood Cancer.* 2012;**59**(5):818–21. doi: 10.1002/pbc.24217. [PubMed: 22678761].

6. Hong JC, Kim J, Browning M, Wagner A, Lerret S, Segura AD, et al. Modified Associating Liver Partition and Portal Vein Ligation for Staged Hepatectomy for Hepatoblastoma in a Small Infant: How Far Can We Push the Envelope? *Ann Surg.* 2017;**266**(2):16–17s. doi: 10.1097/SLA.0000000000002217. [PubMed: 28288067].

7. Pham TA, Gallo AM, Concepcion W, Esquivel CO, Bonham CA. Effect of Liver Transplant on Long-term Disease-Free Survival in Children With Hepatoblastoma and Hepatocellular Cancer. *JAMA Surg.* 2015;**150**(12):1150. doi: 10.1001/jamasurg.2015.1847. [PubMed: 26308249].

8. Kremer N, Walther AE, Tiao GM. Management of hepatoblastoma: an update. *Curr Opin Pediatr.* 2014;**26**(3):362–9. doi: 10.1097/MOP.0000000000000081. [PubMed: 24759227].

9. Ismail H, Broniszczak D, Kalicinski P, Dembowska-Baginska B, Perek D, Teisseyre J, et al. Changing treatment and outcome of children with hepatoblastoma: analysis of a single center experience over the last 20 years. *J Pediatr Surg.* 2012;**47**(7):1331–9. doi: 10.1016/j.jpedsurg.2011.11.073. [PubMed: 22813792].

10. Gong J, Chehrazi-Raffle A, Reddi S, Salgia R. Development of PD-1 and PD-L1 inhibitors as a form of cancer immunotherapy: a comprehensive review of registration trials and future considerations. *J Immunother Cancer.* 2018;**6**(1):8. doi: 10.1186/s40425-018-0316-z. [PubMed: 29357948]. [PubMed Central: PMC5778665].

11. Tang J, Shalabi A, Hubbard-Lucey VM. Comprehensive analysis of the clinical immuno-oncology landscape. *Ann Oncol.* 2018;**29**(1):84–91. doi: 10.1093/annonc/mdx755. [PubMed: 29228097].

12. Guo S, Deng CX. Effect of Stromal Cells in Tumor Microenvironment on Metastasis Initiation. *Int J Biol Sci.* 2018;**14**(14):2083–93. doi: 10.7150/ijbs.25720. [PubMed: 30585271]. [PubMed Central: PMC6299363].

13. Cheng HS, Lee JXT, Wahli W, Tan NS. Exploiting vulnerabilities of cancer by targeting nuclear receptors of stromal cells in tumor microenvironment. *Mol Cancer.* 2019;**18**(1):51. doi: 10.1186/s12943-019-0971-9. [PubMed: 30925918]. [PubMed Central: PMC6441226].

14. Nilendu P, Sarode SC, Jahagirdar D, Tandon I, Patil S, Sarode GS, et al. Mutual concessions and compromises between stromal cells and cancer cells: driving tumor development and drug resistance. *Cell Oncol (Dordr).* 2018;**41**(4):353–67. doi: 10.1007/s13402-018-0388-2. [PubMed: 30027403].

15. Hua X, Chen J, Su Y, Liang C. Identification of an immune-related risk signature for predicting prognosis in clear cell renal cell carcinoma. *Aging (Albany NY).* 2020;**12**(3):2302–32. doi: 10.18632/aging.102746. [PubMed: 32028264]. [PubMed Central: PMC7041771].

16. Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro de Moura M, et al. Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. *Gastroenterology.* 2017;**153**(3):812–26. doi: 10.1053/j.gastro.2017.06.007. [PubMed: 28624577].

17. Chen G, Wang L, Diao T, Chen Y, Cao C, Zhang X. Analysis of immune-related signatures of colorectal cancer identifying two different immune phenotypes: Evidence for immune checkpoint inhibitor therapy. *Oncol Lett.* 2020;**20**(1):517–24. doi: 10.3892/ol.2020.11605. [PubMed: 32565977]. [PubMed Central: PMC7285802].

18. Wang Q, Li M, Yang M, Yang Y, Song F, Zhang W, et al. Analysis of immune-related signatures of lung adenocarcinoma identified two distinct subtypes: implications for immune checkpoint blockade therapy. *Aging (Albany NY).* 2020;**12**(4):3312–39. doi: 10.18632/aging.102814. [PubMed: 32091408]. [PubMed Central: PMC7066911].

19. Li B, Li X, Zai J, Qian X. Facile Synthesis of Porous Zn-Sn-O Nanocubes and Their Electrochemical Performances. *Nanomicro Lett.* 2016;**8**(2):174–81. doi: 10.1007/s40820-015-0075-z. [PubMed: 30460278]. [PubMed Central: PMC6223671].

20. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;**12**(5):453–7. doi: 10.1038/nmeth.3337. [PubMed: 25822800]. [PubMed Central: PMC4739640].

21. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;**4**:2612. doi: 10.1038/ncomms3612. [PubMed: 24113773]. [PubMed Central: PMC3826632].

22. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;**9**:559. doi: 10.1186/1471-2105-9-559. [PubMed: 19114008]. [PubMed Central: PMC2631488].

23. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;**14**:7. doi: 10.1186/1471-2105-14-7. [PubMed: 23323831]. [PubMed Central: PMC3618321].

24. Sumazin P, Chen Y, Trevino LR, Sarabia SF, Hampton OA, Patel K, et al. Genomic analysis of hepatoblastoma identifies distinct molecular and prognostic subgroups. *Hepatology.* 2017;**65**(1):104–21. doi: 10.1002/hep.28888. [PubMed: 27775819].

25. Cairo S, Armengol C, De Reyniès A, Wei Y, Thomas E, Renard C, et al. Hepatic stem-like phenotype and interplay of Wnt/β-catenin and Myc signaling in aggressive childhood liver cancer. *Cancer Cell.* 2008;**14**(6):471–84.

26. Xu T, Le TD, Liu L, Su N, Wang R, Sun B, et al. CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization. *Bioinformatics.* 2017;**33**(19):3131–3. doi: 10.1093/bioinformatics/btx378. [PubMed: 28605519].

27. Seiler M, Huang CC, Szalma S, Bhanot G. ConsensusCluster: a software tool for unsupervised cluster discovery in numerical data. *OMICS J Integr Biol.* 2010;**14**(1):109–13. doi: 10.1089/omi.2009.0083. [PubMed: 20141333].

28. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;**43**(7). e47. doi: 10.1093/nar/gkv007. [PubMed: 25605792]. [PubMed Central: PMC4402510].

29. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep.* 2017;**18**(1):248–62. doi: 10.1016/j.celrep.2016.12.019. [PubMed: 28052254].

30. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the

analysis of systems-level datasets. *Nat Commun.* 2019;**10**(1):1523. doi: 10.1038/s41467-019-09234-6. [PubMed: 30944313]. [PubMed Central: PMC6447622].

31. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003;**4**:2. doi: 10.1186/1471-2105-4-2. [PubMed: 12525261]. [PubMed Central: PMC149346].

32. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;**102**(43):15545–50. doi: 10.1073/pnas.0506580102. [PubMed: 16199517]. [PubMed Central: PMC1239896].

33. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. *Biom J.* 2010;**52**(1):70–84. doi: 10.1002/bimj.200900028. [PubMed: 19937997].

34. Anitei MG, Zeitoun G, Mlecnik B, Marliot F, Haicheur N, Todosi AM, et al. Prognostic and predictive values of the immunoscore in patients with rectal cancer. *Clin Cancer Res.* 2014;**20**(7):1891–9. doi: 10.1158/1078-0432.CCR-13-2830. [PubMed: 24691640].

35. Mlecnik B, Bindea G, Angell HK, Maby P, Angelova M, Tougeron D, et al. Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity.* 2016;**44**(3):698–711. doi: 10.1016/j.immuni.2016.02.025. [PubMed: 26982367].

36. Karlsson A, Jonsson M, Lauss M, Brunnstrom H, Jonsson P, Borg A, et al. Genome-wide DNA methylation analysis of lung carcinoma reveals one neuroendocrine and four adenocarcinoma epitypes associated with patient outcome. *Clin Cancer Res.* 2014;**20**(23):6127–40. doi: 10.1158/1078-0432.CCR-14-1087. [PubMed: 25278450].

37. Dalin MG, Desrichard A, Katabi N, Makarov V, Walsh LA, Lee KW, et al. Comprehensive Molecular Characterization of Salivary Duct Carcinoma Reveals Actionable Targets and Similarity to Apocrine Breast Cancer. *Clin Cancer Res.* 2016;**22**(18):4623–33. doi: 10.1158/1078-0432.CCR-16-0637. [PubMed: 27103403]. [PubMed Central: PMC5026550].

38. Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature.* 2017;**541**(7637):359–64.

39. Shen Q, Hu G, Wu J, Lv L. A new clinical prognostic nomogram for liver cancer based on immune score. *PLoS One.* 2020;**15**(7). e0236622. doi: 10.1371/journal.pone.0236622. [PubMed: 32730361]. [PubMed Central: PMC7392298].

40. Kwon CH, Park HJ, Choi YR, Kim A, Kim HW, Choi JH, et al. PSMB8 and PBK as potential gastric cancer subtype-specific biomarkers associated with prognosis. *Oncotarget.* 2016;**7**(16):21454–68. doi: 10.18632/oncotarget.7411. [PubMed: 26894977]. [PubMed Central: PMC5008298].

41. Yang BY, Song JW, Sun HZ, Xing JC, Yang ZH, Wei CY, et al. PSMB8 regulates glioma cell migration, proliferation, and apoptosis through modulating ERK1/2 and PI3K/AKT signaling pathways. *Biomed Pharmacother.* 2018;**100**:205–12. doi: 10.1016/j.biopha.2018.01.170. [PubMed: 29428669].

42. Ding Q, Lu P, Xia Y, Ding S, Fan Y, Li X, et al. CXCL9: evidence and contradictions for its role in tumor progression. *Cancer Med.* 2016;**5**(11):3246–59. doi: 10.1002/cam4.934. [PubMed: 27726306]. [PubMed Central: PMC5119981].

43. Fukuda Y, Asaoka T, Eguchi H, Yokota Y, Kubo M, Kinoshita M, et al. Endogenous CXCL9 affects prognosis by regulating tumor-infiltrating natural killer cells in intrahepatic cholangiocarcinoma. *Cancer Sci.* 2020;**111**(2):323–33. doi: 10.1111/cas.14267. [PubMed: 31799781]. [PubMed Central: PMC7004525].

44. Le AH, Yelland T, Paul NR, Fort L, Nikolaou S, Ismail S, et al. CYRI-A limits invasive migration through macropinosome formation and integrin uptake regulation. *J Cell Biol.* 2021;**220**(9). doi: 10.1083/jcb.202012114. [PubMed: 34165494]. [PubMed Central: PMC8236918].

45. Diquigiovanni C, Bergamini C, Evangelisti C, Isidori F, Vettori A, Tiso N, et al. Mutant MYO1F alters the mitochondrial network and induces tumor proliferation in thyroid cancer. *Int J Cancer.* 2018;**143**(7):1706–19. doi: 10.1002/ijc.31548. [PubMed: 29672841].