

## Research Article

# Machine Learning for Building Immune Genetic Model in Hepatocellular Carcinoma Patients

Jun Liu <sup>1,2</sup>, Zheng Chen <sup>3</sup>, and Wenli Li <sup>1</sup>

<sup>1</sup>Reproductive Medicine Center, Yue Bei People's Hospital, Shantou University Medical College, Shaoguan, Guangdong, China

<sup>2</sup>Medical Research Center, Yue Bei People's Hospital, Shantou University Medical College, Shaoguan 512025, China

<sup>3</sup>Liver Cancer Institute, Zhongshan Hospital, Fudan University, Key Laboratory of Carcinogenesis and Cancer Invasion, Ministry of Education, Shanghai, China

Correspondence should be addressed to Wenli Li; wangyiliwenli@163.com

Received 28 November 2020; Revised 18 February 2021; Accepted 1 March 2021; Published 18 March 2021

Academic Editor: Zhixiong Liu

Copyright © 2021 Jun Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Background.** Hepatocellular carcinoma (HCC) is the leading liver cancer with special immune microenvironment, which played vital roles in tumor relapse and poor drug responses. In this study, we aimed to explore the prognostic immune signatures in HCC and tried to construct an immune-risk model for patient evaluation. **Methods.** RNA sequencing profiles of HCC patients were collected from the cancer genome Atlas (TCGA), international cancer genome consortium (ICGC), and gene expression omnibus (GEO) databases (GSE14520). Differentially expressed immune genes, derived from ImmPort database and MSigDB signaling pathway lists, between tumor and normal tissues were analyzed with Limma package in R environment. Univariate Cox regression was performed to find survival-related immune genes in TCGA dataset, and in further random forest algorithm analysis, significantly changed immune genes were used to generate a multivariate Cox model to calculate the corresponding immune-risk score. The model was examined in the other two datasets with recipient operation curve (ROC) and survival analysis. Risk effects of immune-risk score and clinical characteristics of patients were individually evaluated, and significant factors were then used to generate a nomogram. **Results.** There were 52 downregulated and 259 upregulated immune genes between tumor and relatively normal tissues, and the final immune-risk model (based on SPP1, BRD8, NDRG1, KITLG, HSPA4, TRAF3, ITGAV and MAP4K2) can better differentiate patients into high and low immune-risk subpopulations, in which high score patients showed worse outcomes after resection ( $p < 0.05$ ). The differentially enriched pathways between the two groups were mainly about cell proliferation and cytokine production, and calculated immune-risk score was also highly correlated with immune infiltration levels. The nomogram, constructed with immune-risk score and tumor stages, showed high accuracy and clinical benefits in prediction of 1-, 3- and 5-year overall survival, which is useful in clinical practice. **Conclusion.** The immune-risk model, based on expression of SPP1, BRD8, NDRG1, KITLG, HSPA4, TRAF3, ITGAV, and MAP4K2, can better differentiate patients into high and low immune-risk groups. Combined nomogram, using immune-risk score and tumor stages, could make accurate prediction of 1-, 3- and 5-year survival in HCC patients.

## 1. Introduction

Hepatocellular carcinoma (HCC) is one of the most malignant tumors around the world, causing the second highest mortal rate with poor responses to therapies [1, 2]. The immune microenvironment of tumors has been testified to play vital roles in tumor progression and relapse, which incurred development of immune therapies, such as application of checkpoint inhibitors and Car-T

transfusion [3]. Though these drugs, approved by the federal government in several countries, demonstrated efficacy in tumor regression and prolonged overall survival, overall response was not satisfying in patients, which may be related to tumor mutation burden and immune infiltration levels [4]. Understanding immune microenvironment within HCC can better predict patients' survival, which can also be used to guide drug usage or treating strategies.

There have been some studies investigating prognostic signatures of HCC; however, high-precision prognostic models based on immune-related genes were few in HCC [5, 6]. In this study, we managed to construct an immune-risk model to differentiate HCC patients into high- and low-risk subgroups for survival prediction. High score patients had worse prognosis after resection, and the significantly changed immune genes between HCC tissues and normal tissues were highly enriched in pathways of cell growth and tyrosine kinase inhibitor resistance. Combining immune-risk score and tumor stage, we constructed a nomogram with high precision, which can help therapists make comprehensive clinical evaluation of patients in practice.

## 2. Materials and Methods

**2.1. Data Description and Derivation of Immune Gene List.** Expression data included in the analysis were from TCGA database (<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>), ICGA database (<https://icgc.org>), and GEO database (<https://www.ncbi.nlm.nih.gov/geo>) [7, 8]. Clinical information of all the related patients (TCGA: 370 cases, GSE14520: 209 cases, ICGC: 232 cases) was shown in Table 1. Immune gene list analyzed in this study was procured from ImmPort database (<https://www.immport.org>) and MSigDB pathway signature list (IMMUNE\_RESPONSE and IMMUNE\_SYSTEM\_PROCESS). The gene list of transcription factors was from CISTROME Project (<http://cistrome.org/>), which is an open web tool for tumor sequencing data analysis.

**2.2. Analysis of Differentially Expressed Immune Genes.** RNA expression levels of different genes from patients in TCGA dataset were analyzed between the hepatocellular carcinoma tissues and the para-tumor tissues, using Limma package in R environment [9]. The genes expressing more than 1-fold change with adjusted  $p$  value under 0.05 were considered significant after normalization and background correction. The package of Heatmap was used to create the heatmap of significantly up- and downregulated genes.

**2.3. Gene Ontology and Gene Enrichment.** To better understand the functional roles of the significantly changed genes, the package of clusterProfiler was used to demonstrate the gene ontology and enriched pathways, including the cellular compartment, biological process, molecular function, and KEGG pathways (Kyoto Encyclopedia of Genes and Genomes, <https://www.kegg.jp>) [10]. The further gene set enrichment analysis (GSEA) between high and low immune-score patients was performed with GSEA 4.0.1 software in order to locate the related gene sets [11, 12].

**2.4. Construction of Immune Prognostic Models and Nomogram.** The RNA sequencing data in TCGA LIHC dataset were used to find the differentially expressed immune genes, and through univariate Cox regression, survival-related immune genes (overall survival and

TABLE 1: Patients' information in the TCGA, ICGC, and GEO cohorts.

Clinical characteristics		Total (%)
TCGA		365
Survival status	Survival	239 (65.48)
	Death	126 (34.52)
Age	<61 years	173 (47.40)
	≥61 years	192 (52.60)
Gender	Male	246 (67.40)
	Female	119 (32.60)
Histological grade	G1	55 (15.07)
	G2	175 (47.95)
	G3	118 (32.33)
	G4	12 (3.29)
	Unknown	5 (1.36)
Stage	I	170 (46.58)
	II	84 (23.01)
	III	83 (22.74)
	IV	5 (1.10)
	Unknown	24 (6.57)
ICGC		232
Survival status	Survival	189 (81.47)
	Death	43 (18.53)
Age	<69 years	116 (50.00)
	≥69 years	116 (50.00)
Gender	Male	171 (73.71)
	Female	61 (26.29)
Stage	I	36 (15.52)
	II	106 (45.69)
	III	71 (30.60)
	IV	19 (8.19)
Prior malignancy	No	202 (87.07)
	Yes	30 (12.93)
GSE14520		209
Survival status	Survival	130 (62.20)
	Death	79 (37.80)
Age	<60 years	168 (80.38)
	≥60 years	41 (19.62)
Gender	Male	26 (12.44)
	Female	183 (87.56)
Stage (TNM)	I	90 (43.06)
	II	74 (35.41)
	III	43 (20.57)
	Unknown	2 (0.96)
		0
Stage (CLIP)	1	67 (32.57)
	2	32 (15.30)
	3	9 (4.30)
	4	3 (1.44)
	5	1 (0.48)
	Unknown	2 (0.96)

progression free survival) ( $p < 0.05$ ) were selected for risk evaluation. The overlapped gene set (68 signatures) was chosen to find the most relevant prognostic genes, and random forest algorithm was used to compress the gene list. The random forest algorithm was performed to determine the important of immune-related genes, and the relative importance  $> 0.2$  was identified as the final signature. The

finally yielded 8 immune genes (SPP1, BRD8, NDRG1, KITLG, HSPA4, TRAF3, ITGAV, and MAP4K2) were also evaluated for their variable relative importance and were further put into multivariate Cox analysis to generate the immune-risk model. These analyses were performed with *R* package of randomForestSRC and algorithm of randomSurvivalForest. In construction of the model, median risk score (0.961) was deployed to assign patients into high and low immune-risk groups, and, correspondingly, prognostic value of the model was examined in the training set of TCGA and the testing sets of ICGC and GSE14520 (immune-risk score =  $0.115 * \text{expSPP1} + 0.263 * \text{expBRD8} + 0.125 * \text{expNDRG1} + 0.233 * \text{KITLG} + 0.195 * \text{HSPA4} + 0.319 * \text{TRAF3} - 0.186 * \text{expITGAV} + 0.333 * \text{expMAP4K2}$ ).

Immune-risk score and clinical characteristics of patients, such as age, gender, tumor grades, and stages, were then put into Cox model to evaluate corresponding risk effect. Immune-risk score and tumor stages were significantly correlated with survival of patients before or after adjustment, which were then used to create a nomogram for prognosis prediction in 1-, 3-, and 5-year follow-up. C-index of the nomogram was calculated with bootstrap of 1000 resamples, ranging from 0.5 to 1.0. Precision of the model was examined by calibration graphs, in which the alignment of both lines was a fact of good performance, reflecting the actual probability. Recipient operation curves (ROCs) were used to demonstrate its specificity and sensitivity in comparison with other factors. Decision curve analysis (DCA) was also deployed to evaluate the potential clinical benefits of the model. ROC and DCA were performed with *R* packages of rms, survcomp, and survivalROC. Kaplan–Meier (K-M) method was used to estimate the accumulative incidence of survival between different subpopulations in each dataset. The survival curves were drawn through *R* package of survival and compared with log-rank test.

**2.5. Analysis of the Infiltrated Immune Cells.** The infiltration of six common immune cells was estimated through online database of TIMER (<https://cistrome.shinyapps.io/timer/>), which is an open tool with data access to 32 types of cancers in TCGA database [13, 14]. Infiltration levels of B cells, CD4+ T cells, CD8+ T cells, neutrophils, dendritic cells, and macrophages were calculated correspondingly, and the correlation between immune-risk score and infiltrated cell types was examined by Pearson's correlation test. Evaluation of immune status's difference between high and low immune-risk patients was performed with single sample gene set enrichment analysis (ssGSEA), using gene sets in MSigDB pathway lists. Expression of well-known immune checkpoints was also compared between groups. *p* value under 0.05 was considered significant.

### 3. Results

**3.1. Differentially Expressed Immune Genes between HCC and Para-Tumor Tissues.** Work flow of this study was shown in Figure 1. We found 311 differentially expressed immune genes between HCC and para-tumor tissues ( $\log_2$ -

transformed fold change > 1,  $p < 0.05$ ). In the list of significantly changed genes, 259 genes were upregulated and 52 genes were downregulated (Figures 2(a) and 2(b)). Gene ontology of those significantly changed genes was mainly involved in signal transduction and cytokine regulation (Figures 2(c) and 2(e)). The highly enriched pathways were MAPK signaling pathway, Rap1 signaling pathway, Ras signaling pathway, EGFR tyrosine kinase inhibitor resistance, ErbB signaling pathway, and so on, which were well-known signalings for tumor growths and metastasis, indicating the protumor immune microenvironment (Figures 2(d) and 2(f)). We also analyzed the differentially expressed transcription factors (TFs) between normal and tumor tissues in HCC patients (Supplemental Figures 1(a) and 1(b)). It turned out most TFs were upregulated in tumor tissues (108 upregulated, 9 downregulated).

Constructed immune-risk model with eight signatures can better stratify patients into subpopulations.

We performed univariate Cox analysis to locate survival-related genes in the training set from TCGA database. In association with overall survival (OS), we located 93 signatures, while in association with progression free survival (PFS), we located 116 signatures. The overlapped 68 signatures were used to build immune-risk model (Figure 3(a)). The regulating network between differentially expressed TFs and survival-related immune genes was also examined (Supplemental Figure 1(c)).

Random forest algorithm was used to additionally downsize the candidate signatures, and after iteration, signatures of SPP1, BRD8, NDRG1, KITLG, HSPA4, TRAF3, ITGAV, and MAP4K2 were used to construct immune-risk model (Figures 3(b) and 3(c)). Through multivariate Cox regression, we used the immune-risk model to calculate the immune-risk score of HCC patients, and using the mean immune-risk score, we dichotomized patients into high- and low-risk subpopulations (Figures 3(d), 3(e), and 3(f)). The difference of OS between high and low immune-risk score subgroups was consistent among all three datasets (TCGA, ICGC, and GSE14520), with high-risk patients suffering from poor outcomes in the follow-up ( $p < 0.05$ ) (Figures 4(a)–4(c)). The area under curves (AUCs) for the immune-risk model in prediction of 0.5-, 1-, 2-, 3-, and 5-year OS were 0.79, 0.80, 0.72, 0.67, and 0.68 in training set of TCGA dataset (Figure 4(d)). In testing dataset of ICGC, the corresponding AUC values were 0.77, 0.74, 0.76, 0.78, and 0.75, while, in GSE14520, AUCs for 0.5-, 2-, 3-, and 5-year OS were 0.68, 0.70, 0.68, and 0.66 (Figures 4(e) and 4(f)). In analysis of PFS, high immune-risk patients in TCGA dataset also suffered from worse outcomes, and the respective 0.5-, 1-, 2-, 3-, and 5-year AUCs were 0.64, 0.66, 0.63, 0.64, and 0.62 (Supplemental Figures 2(a) and 2(b)).

Independent risk effect for the immune-risk model and corresponding expression levels in patients' subgroups.

Then we tried to evaluate the risk effect of immune-risk score and other clinical characteristics in relation to OS in TCGA dataset. After assignment of patients according to age ( $\geq 60$  or  $< 60$ ), gender, and tumor grades (G1/2 or G3/4) and tumor stages (stage1/2 or stage3/4), we found high immune-risk patients shared worse outcomes consistently

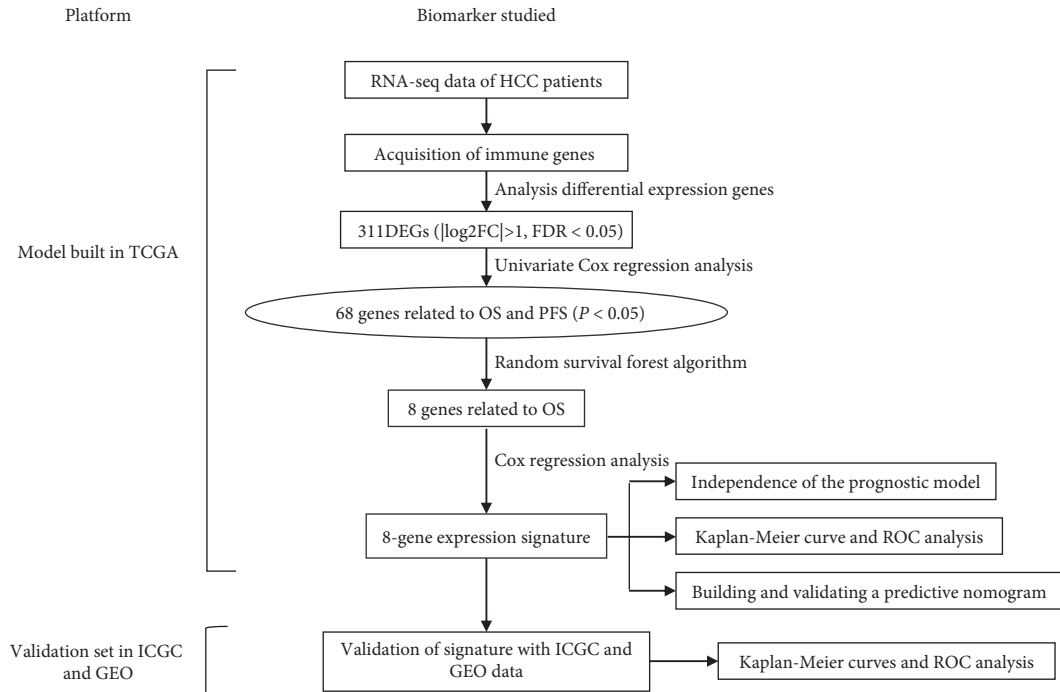


FIGURE 1: Workflow of the analysis.

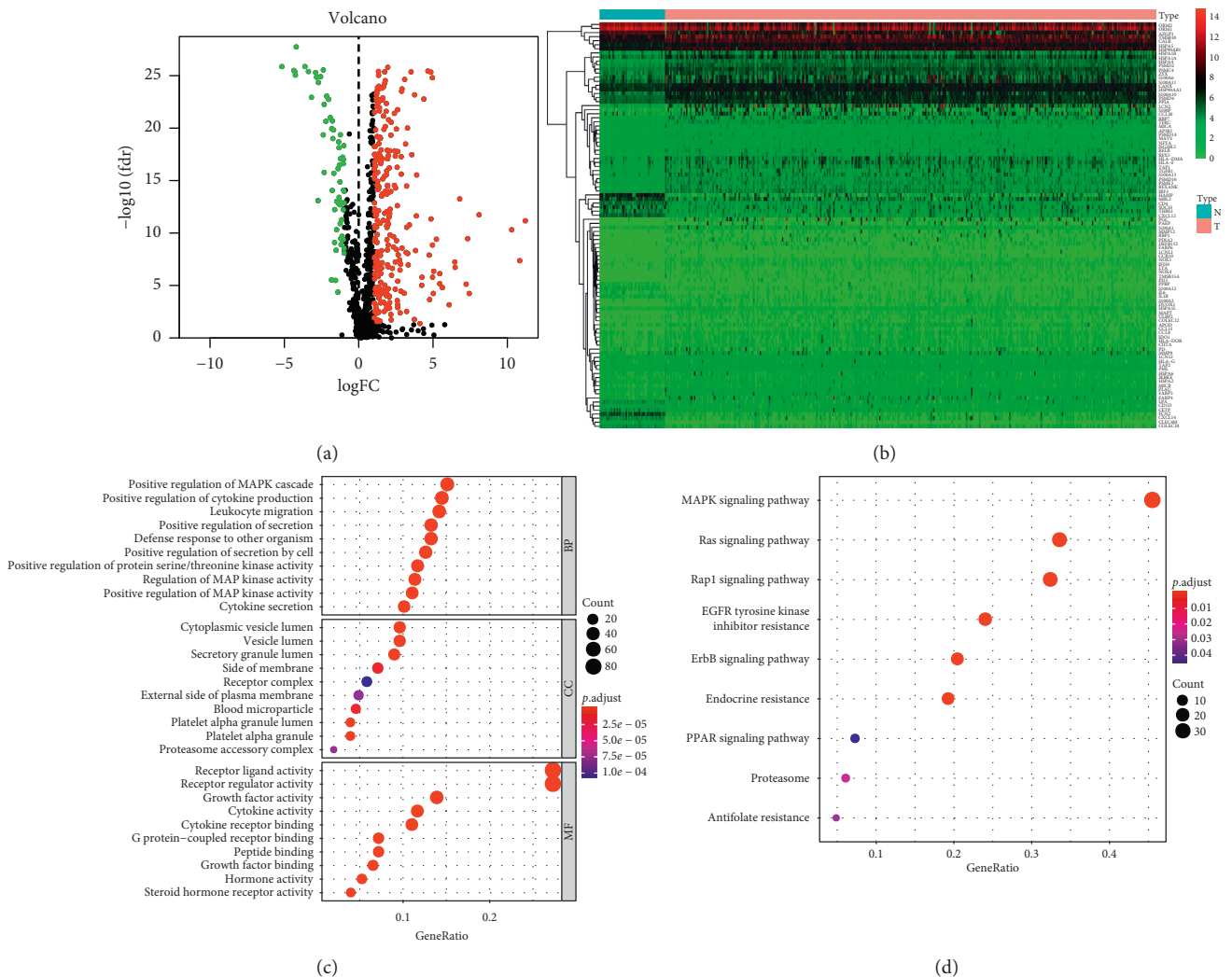


FIGURE 2: Continued.

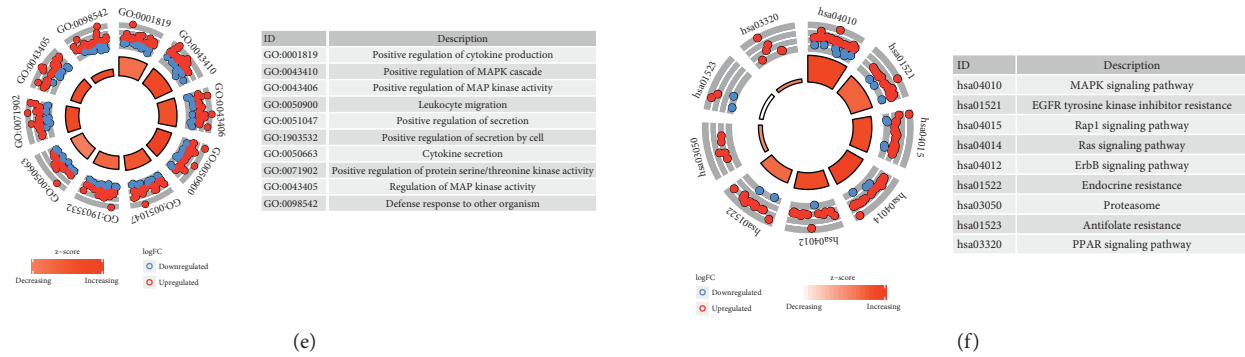


FIGURE 2: Differentially expressed immune genes (DEIGs) and gene ontology. (a) Volcano graph of DEIGs between HCC and para-tumor tissues. (b) Heatmap of DEIGs. (c) Gene ontology of DEIGs. (d) Enriched pathways for DEIGs through KEGG website. (e) Top ranked gene ontology for DEIGs. (f) Top ranked KEGG pathways for DEIGs.

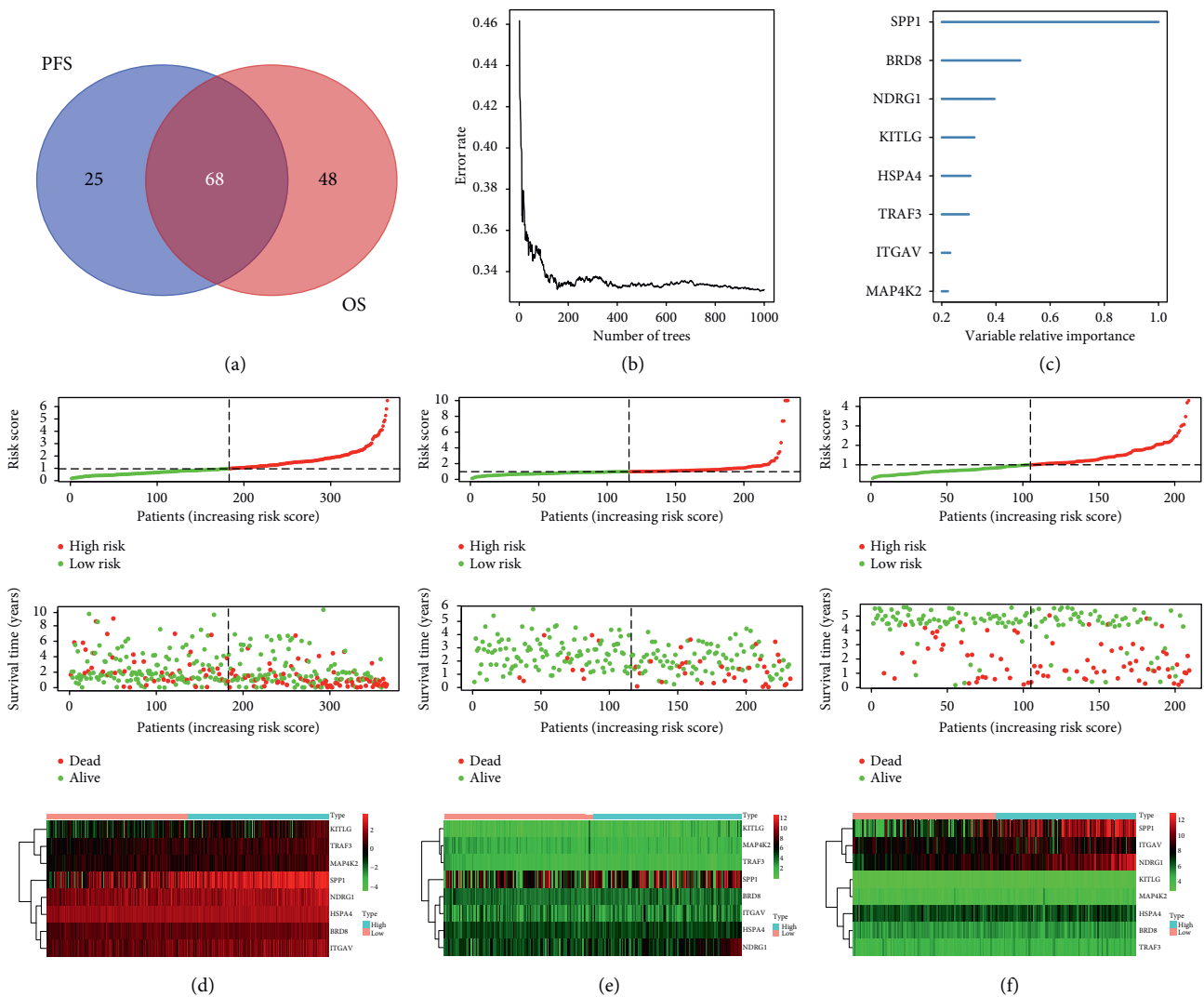


FIGURE 3: Construction of immune-risk model with DEIGs and further assignment of patients. (a) Venn diagram for survival-related (OS and PFS) DEIGs in TCGA dataset. (b) Random forest algorithm was used to downsize the survival-related DEIGs. (c) The relative importance for eight immune signatures generated by random forest algorithm. (d-f) The division of patients with median risk score calculated by 8-gene immune-risk model and the corresponding data spread in training dataset from TCGA and testing datasets from ICGC and GSE14520. The expression difference of the eight signatures between high and low immune-risk groups was shown by heatmap.



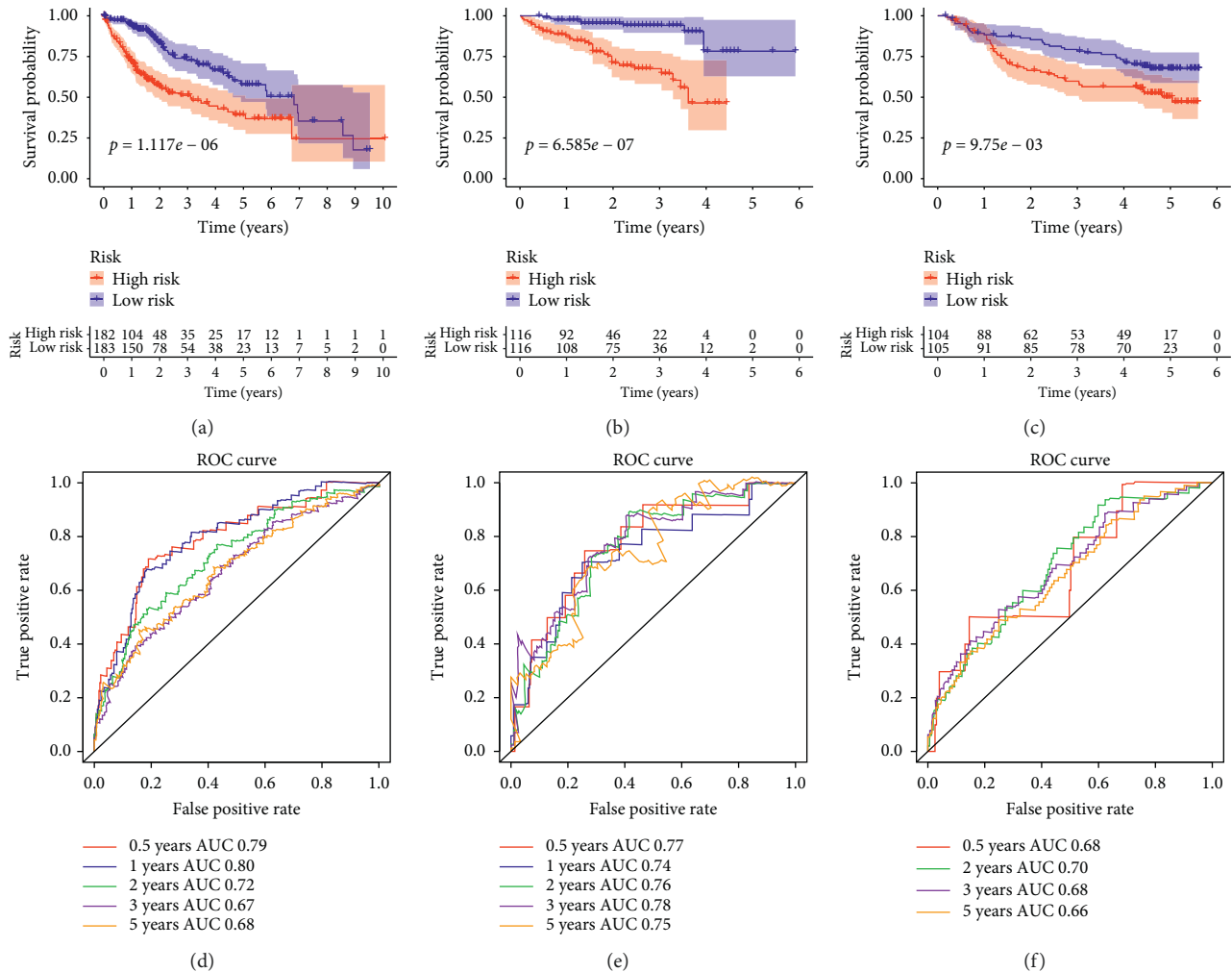


FIGURE 4: The overall survival difference between high and low immune-risk score patients in datasets of TCGA, ICGC, and GSE14520, and their respective predicting values for 0.5-, 1-, 2-, 3-, and 5-year survival. (a–c) Overall survival difference between high and low immune-risk patients in TCGA, ICGC, and GSE14520 datasets. (d–f) Recipient operation curve (ROC) for 0.5-, 1-, 2-, 3-, and 5-year survival prediction of immune-risk model in patients from TCGA, ICGC, and GSE14520 datasets.

(Figures 5(a)–5(h)). Also, the respective results in relation to PFS were similar for HCC patients in TCGA dataset, in which high immune-risk patients shared shorter PFS time (Supplemental Figure 2(c)). All those indicated immune-risk score was an independent risk factor for HCC patients.

We further examined clinical characteristics in single variate and multivariate Cox analysis in combination with immune-risk score. It showed in TCGA, ICGC, and GSE14520 datasets, tumor stages and immune-risk score were both significant risk factors to OS even after adjustment in multivariate model (Figures 6(a)–6(f)).

Enriched pathways in high immune-risk score patients demonstrated unique biological behaviors of HCC cells.

Through gene set enrichment analysis (GSEA), we tried to understand the differentially enriched signals between high and low immune-risk patients. The results showed that signatures, related to pathways of cell cycle, apoptosis, NOD like receptor signaling, Notch signaling, and VEGF signaling, were significantly enriched in high immune-risk

patients, indicating the proliferative status of HCC cells (Figure 7, Table 2).

HCC infiltrated immune cell populations were also analyzed to find whether the infiltration patterns were related to immune-risk score. It turned out infiltration of six types of immune cells (B cell, CD4+ T cell, CD8+ T cell, dendritic cell, macrophage, and neutrophil) were positively related to immune-risk score ( $p < 0.05$ ). High immune-risk score was correlated to high infiltration of immune cells, which could also be extrapolated as exhaustive immune status (Figure 8). Of the six cell types, macrophages and neutrophils showed the highest relevance and significance. Also, using ssGSEA analysis to evaluate immune status between groups, we found the overall immune activity score of different signals was relatively higher in high immune-risk patients. Expression levels of immune checkpoints were also higher in high immune-risk patients (Supplemental Figure 3).

We further analyzed the survival difference between immune cell infiltration groups, and there was no significant

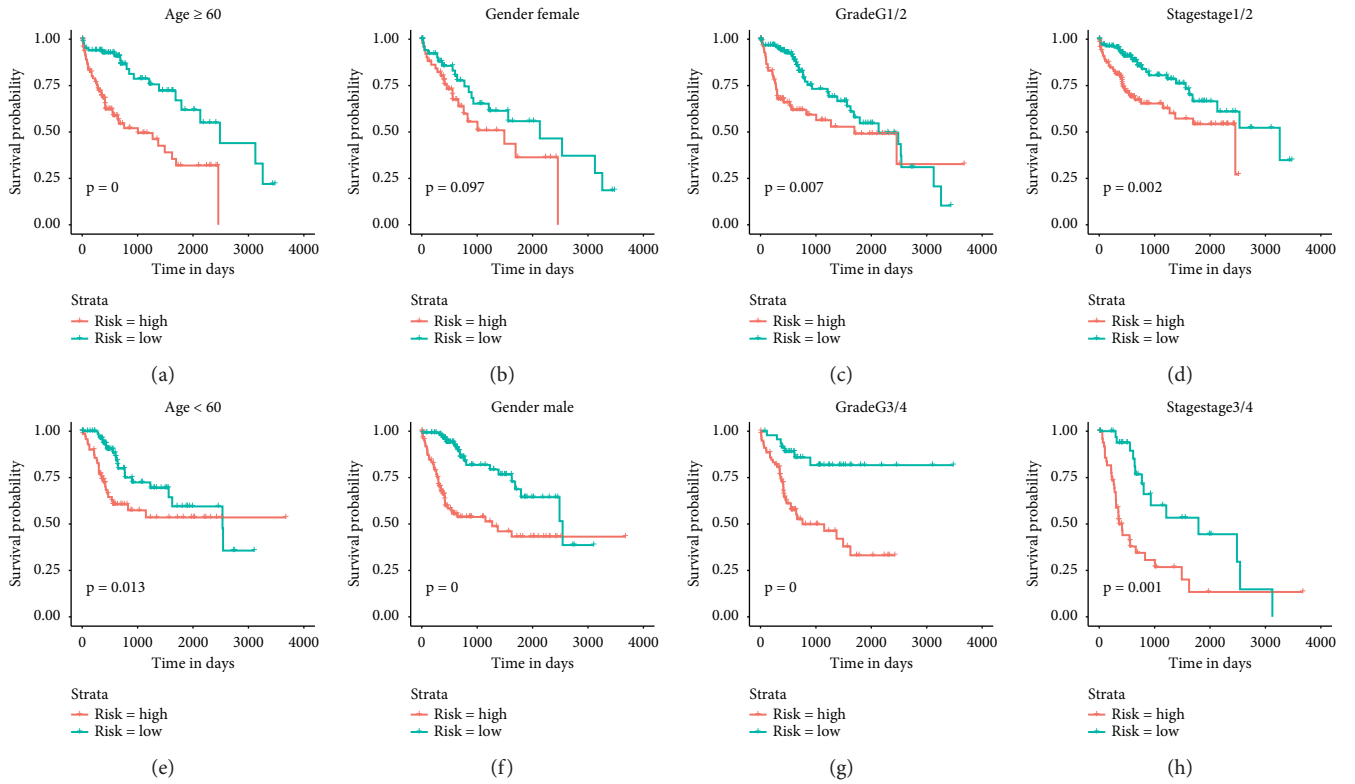


FIGURE 5: Overall survival (OS) difference between high and low immune-risk score patients in subpopulations with different clinical characteristics. (a-b) OS difference between high and low immune-risk patients in subgroups with age over or below 60 from TCGA dataset. (c-d) OS difference between high and low immune-risk patients in subgroups with different gender (female and male) from TCGA dataset. (e-f) OS difference between high and low immune-risk subgroups with tumor grades of grade 1/2 or grade 3/4 from TCGA dataset. (g-h) OS difference between high and low immune-risk patients in subgroups with tumor stage 1/2 or stage 3/4 from TCGA dataset.

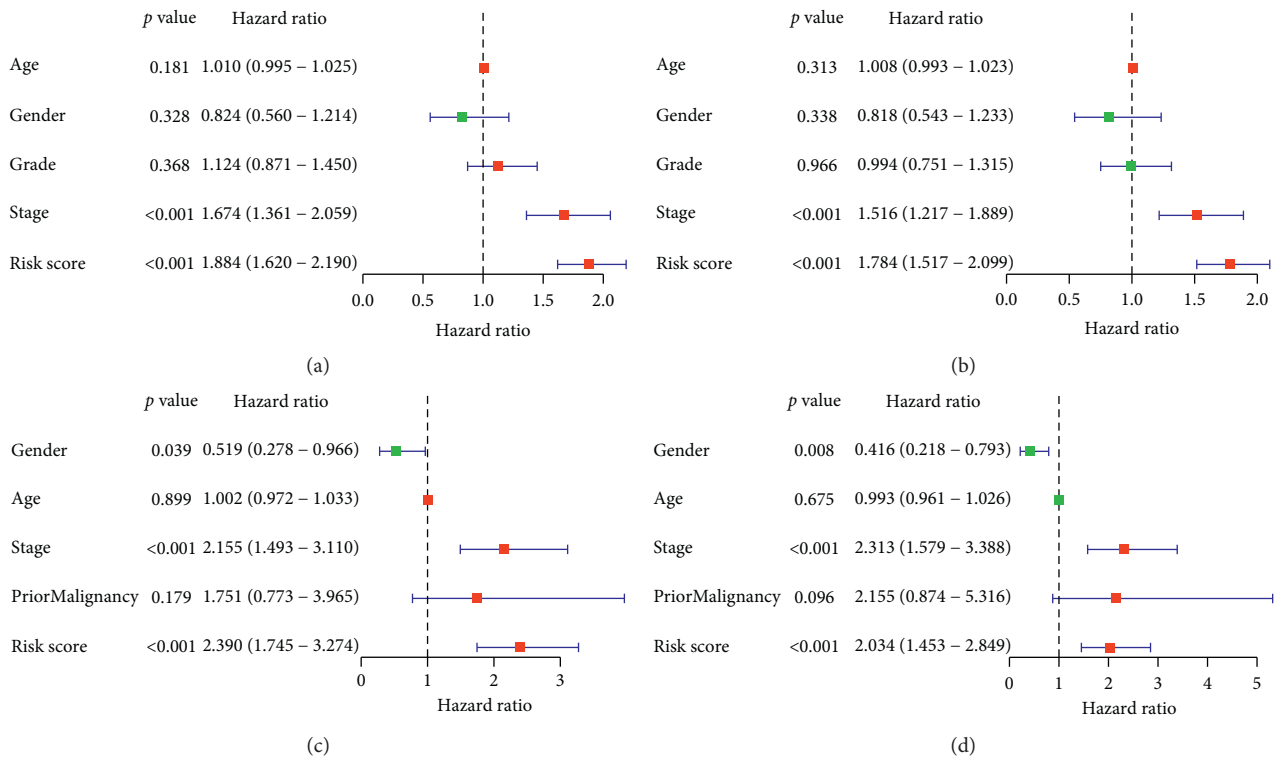


FIGURE 6: Continued.

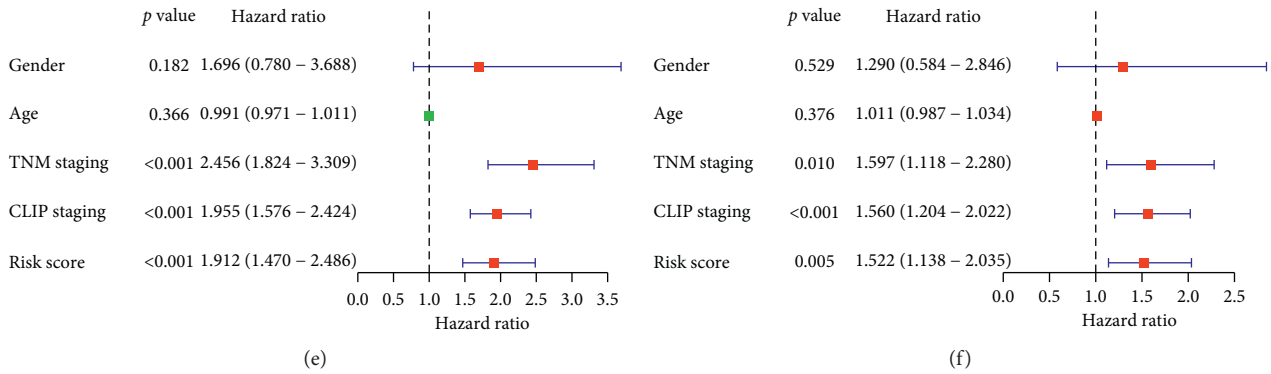


FIGURE 6: Univariate and multivariate analysis of overall survival-related risk factors in datasets of TCGA, ICGC, and GSE14520 (a, c, e). Univariate Cox analysis of the survival-related factors in TCGA, ICGC, and GSE14520 datasets (b, d, f). Multivariate Cox analysis of the survival-related factors in TCGA, ICGC and GSE14520 dataset.

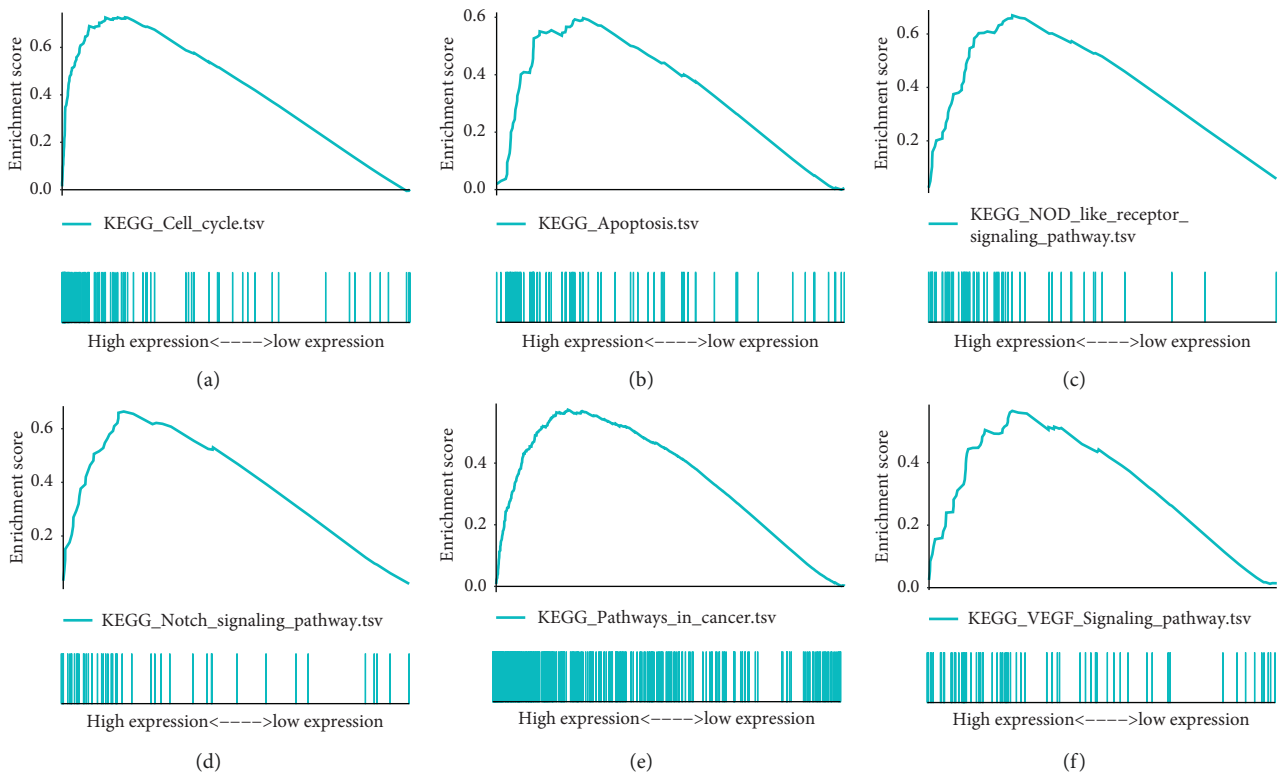


FIGURE 7: Highly enriched pathways in high immune-risk score patients through gene set enrichment analysis (GSEA). (a) Enriched score for signatures in the pathway of cell cycle. (b) Enriched score for signatures in the pathway of apoptosis. (c) Enriched score for signatures in the pathway of NOD like receptor signaling. (d) Enriched score for signatures in the pathway of Notch signaling. (e) Enriched score for signatures in pathways of cancer. (f) Enriched score for signatures in the pathway of VEGF signaling.

TABLE 2: Gene sets enriched in phenotype high.

NAME	ES	NES	FDR q-val
KEGG_PYRIMIDINE_METABOLISM	0.672451	2.173925	0
KEGG_CELL_CYCLE	0.737056	2.097065	7.27E-04
KEGG_P53_SIGNALING_PATHWAY	0.620247	2.06442	0.001371
KEGG_BASAL_TRANSCRIPTION_FACTORS	0.709419	2.012458	0.003543
KEGG_PATHWAYS_IN_CANCER	0.551037	1.914016	0.007377
KEGG_NOTCH_SIGNALING_PATHWAY	0.663903	1.91001	0.007401



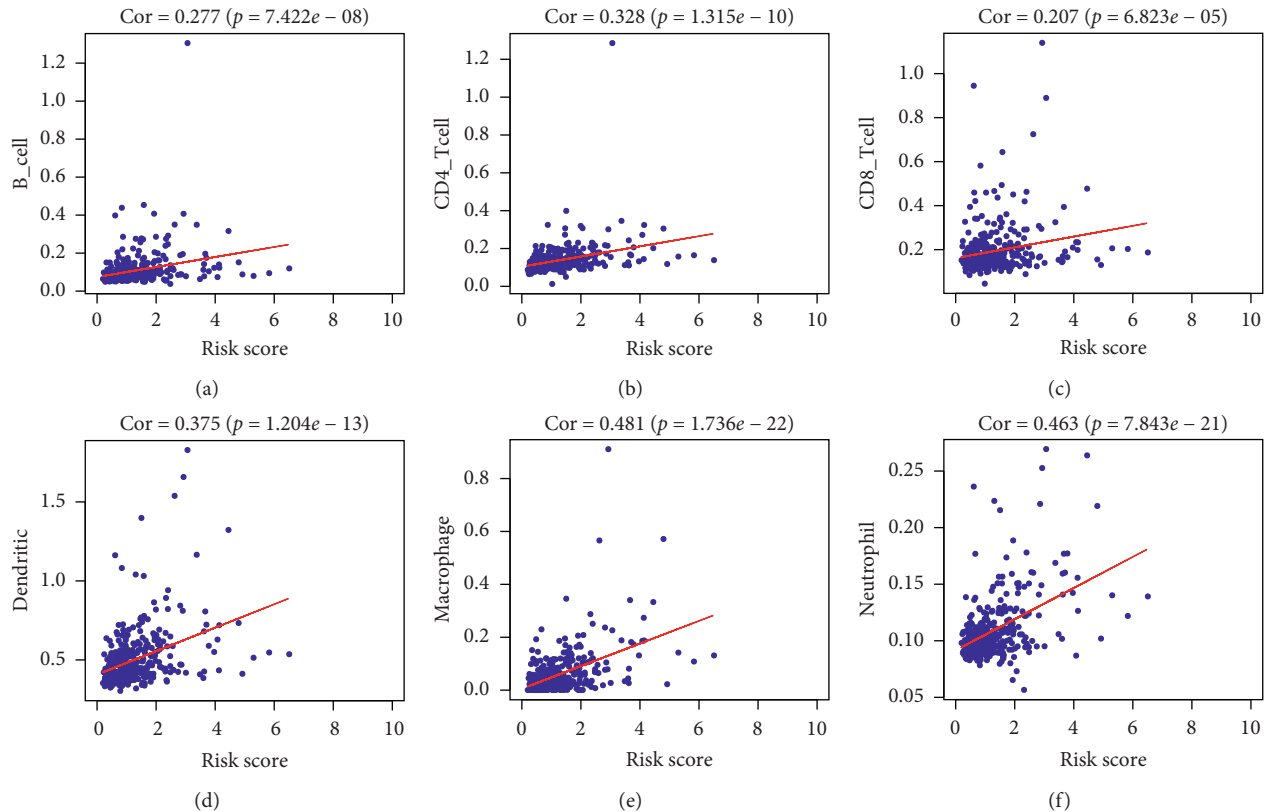


FIGURE 8: The correlation between immune-risk score and commonly infiltrated immune cells in HCC. (a) Relation between infiltrated B cell score and immune-risk score. (b) Relation between infiltrated CD4+ T cell score and immune-risk score. (c) Relations between infiltrated CD8+ T cell score and immune-risk score. (d) Relation between dendritic cell score and immune-risk score. (e) Relation between macrophage score and immune-risk score. (f) Relations between infiltrated neutrophil score and immune-risk score.

difference between high and low infiltration groups of six immune cell types (Figure 9(a)). However, after assigning patients according to immune-score and each cell type infiltration score, we found high immune-risk score was related to worse prognosis consistently, which was independent on immune cell infiltration levels (Figures 9(b)–9(g)). The patients with high-risk score and low B cell, CD4+T cell, CD8+T cell, or dendritic cell showed the worst OS, especially high-risk score combined with low B cell or CD 8+ T cell (Figures 9(b)–9(e)). However, HCC patients with high-risk score and high macrophage presented the worst OS, while the patients with low-risk score and low macrophage had the best OS (Figure 9(f)). This is perhaps because that tumor-associated macrophage can promote cancer-related malignancy. HCC patients with high-risk score had worse OS than those with low-risk score, regardless of neutrophil level (Figure 9(g)).

**3.2. Combined Nomogram Can Better Guide Treating Strategies in Clinical Practice.** In consideration of all the survival-related factors, we generated a nomogram with tumor stages and immune-risk score to evaluate 1-, 3- and 5-year OS of HCC patients (Figure 10(a)). Calibration curves for the 3 models were also generated to show the consistence between estimation and actual probability (Figure 10(b)). The AUCs for the combined model in prediction of 1-, 3-, and 5-year

OS were all over 0.7, which were better than the other two single-factor models (Figure 11(a)). DCAs of the three models also showed superiority of the combined model in prediction of 1-, 3-, and 5-year OS at various threshold probabilities (Figure 11(b)).

#### 4. Discussion

In this study, we focused on the immune status of HCC to find the changed immune genes between HCC and relatively normal liver tissues, and through regression analysis, we further located eight survival-related signatures to construct an immune-risk score model for prognostic prediction. High score patients shared a worse outcome in comparison to the low-risk ones significantly, which was independent on immune infiltration levels, though high immune-risk score positively correlated with immune cell infiltration. The further constructed nomogram, using immune-risk score and tumor stages, could better predict the overall survival of patients in 1, 3, and 5 years than single factor, which is very useful for patient monitoring and instructive for choice of therapy in clinical practice. Though, many previous studies tried to find the differentially expressed prognostic biomarkers in HCC, immune-related models were seldom used to evaluate patients. However, with the development of immune therapy in cancer treatment, such as immune

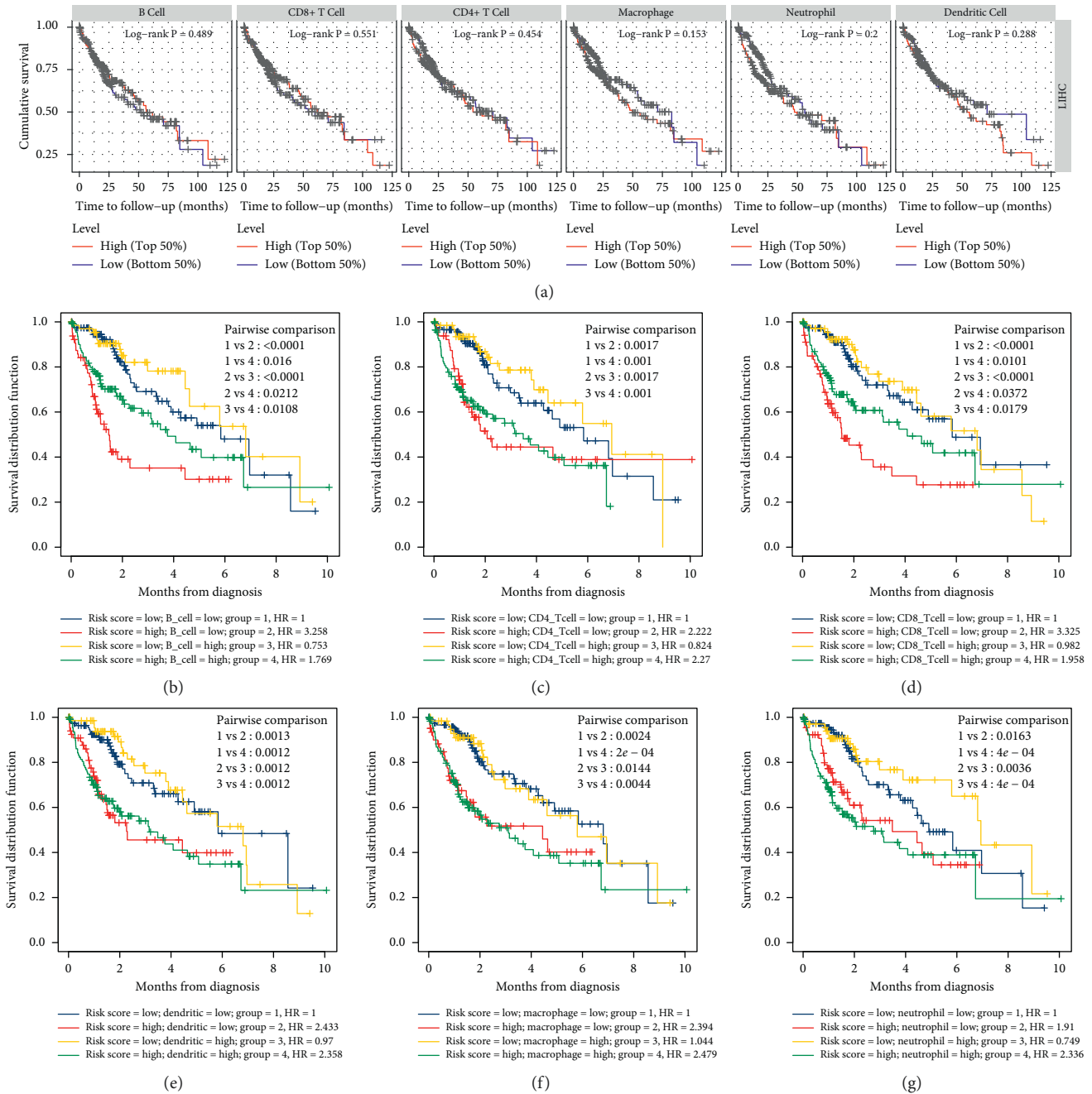


FIGURE 9: Prognostic value of immune-risk model was consistent between different immune-infiltration subgroups. (a) Survival difference was not obvious between subgroups of different immune infiltration groups of six immune cell types. (b) Survival difference between subgroups classified by B cell infiltration and immune-risk score levels. (c) Survival difference between subgroups classified by CD4 T cell infiltration and immune-risk score levels. (d) Survival difference between subgroups classified by CD8 T cell infiltration and immune-risk score levels. (e) Survival difference between subgroups classified by dendritic cell infiltration and immune-risk score levels. (f) Survival difference between subgroups classified by macrophage infiltration and immune-risk score levels. (g) Survival difference between subgroups classified by neutrophil infiltration and immune-risk score levels.

checkpoint blockade (ICB) and Car-T transfusion, immune-risk evaluation of cancer patients may better predict patients' survival and the following response to ICB or Car-T treatment.

In our study, MAPK signaling, Ras signaling, ErbB signaling, and EGFR tyrosine kinase inhibitor resistance pathways were enriched in high immune-risk patients,

specifying the highly proliferative status of cancer cells in high immune-risk patients. Immune microenvironment of high immune-risk patients was more in favor of tumor growth. Also, though the immune-risk score was positively related to tumor infiltrating immune cells, high immune-risk patients tended to share a worse outcome after HCC resection, which may be due to exhaustive immune status.

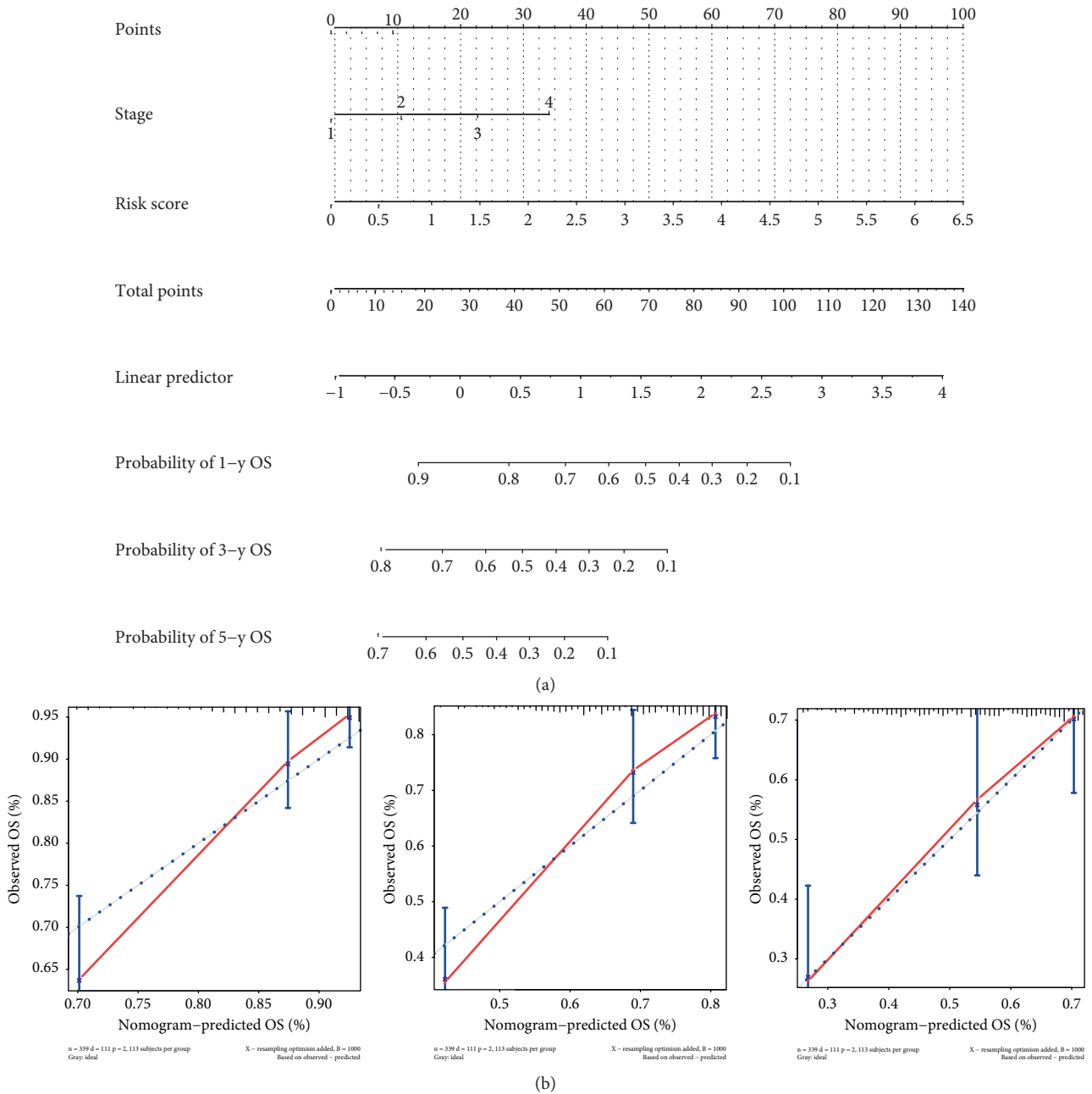


FIGURE 10: Construction of a nomogram with related factors for prognostic prediction. (a) Score model of nomogram in prediction of 1-, 3-, and 5-year overall survival. (b) Calibration graphs for 1-, 3-, and 5-year overall survival prediction within the nomogram (usage: the clinical stages and immune-risk score of patients were calculated and projected to the reference score bar, and total scores were compared to linear predictor to find the probability of 1-, 3-, and 5-year survival).

Former studies have shown that immune cell function depression was related to cancer progression, and we thought those high immune-risk patients could have better treating effect with ICB or Car-T therapy.

In the immune-risk model, secreted phosphoprotein 1 (SPP1 or OPN) is a widely studied signature in tumors, which is normally involved in the process of osteoclasts' attachment to mineralized bone matrix. It is elevated in tumors for progression and metastasis, and its alternatively spliced variants are related to many malignant traits in

cancers, such as epithelial-mesenchymal plasticity, cancer cell stemness, chemoresistance, and radioresistance [15–20]. However, the variants of SPP1 due to aberrantly processing may cause autoimmune reactions in tumors, which can lead to better outcomes in partial patients [21, 22]. In HCC, SPP1 is a validated prognostic biomarker, and it may induce chemoresistance through regulation of autophagy in HCC cells [23, 24]. The next step of exploration shall focus on the involved regulatory pathways in order to find the potential drugs and inhibitors.

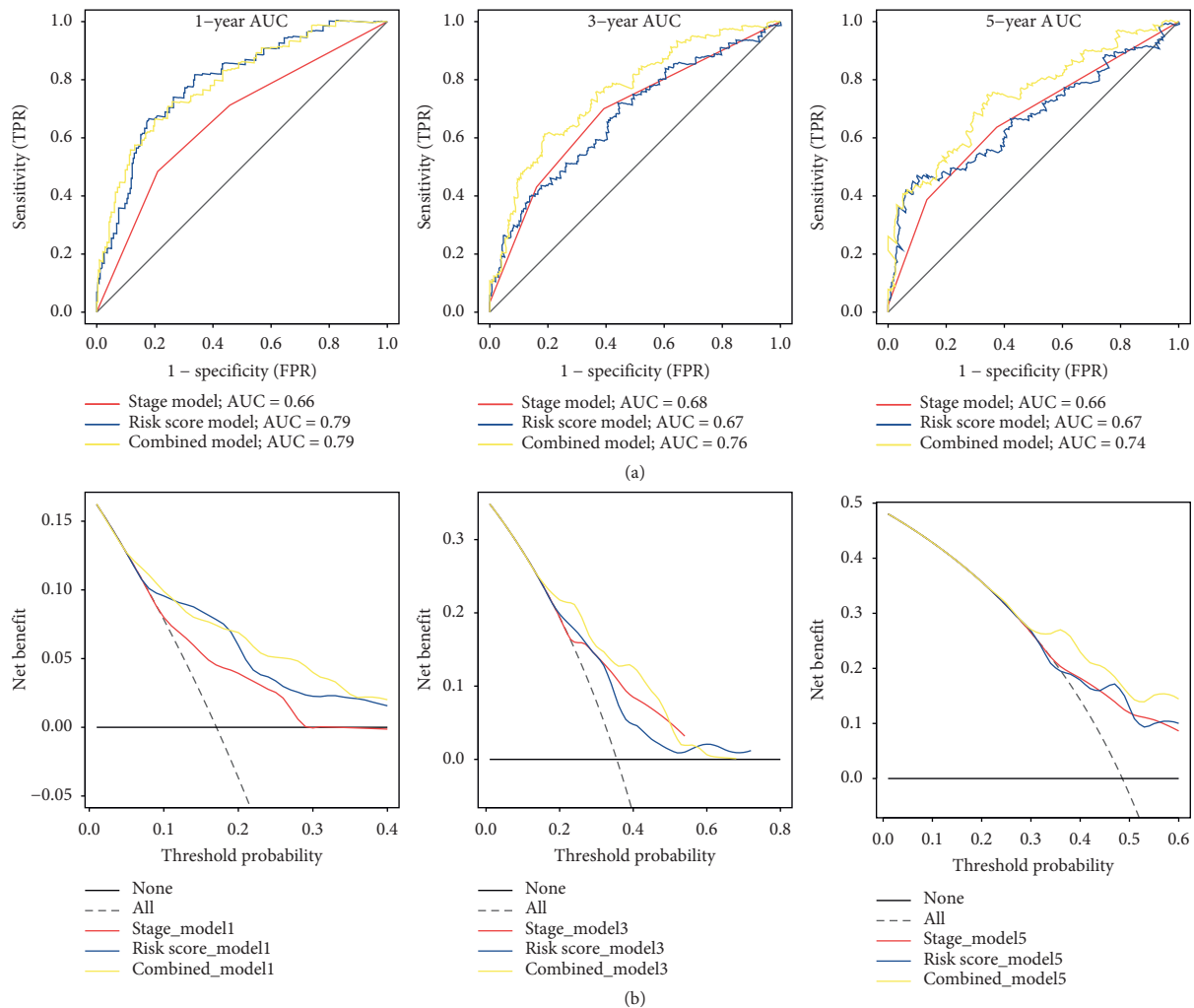


FIGURE 11: Comparison of prognostic values and clinical benefits between different models. (a) ROCs of prognostic models for 1-, 3-, and 5-year overall survival separately. (b) DCA curves of different models for 1-, 3-, and 5-year overall survival prediction. Dotted grey line represented all patients' benefits, and solid black line represented no benefits.

Two of the signatures in our immune-risk model were related to NF- $\kappa$ B and TGF- $\beta$  signal pathways. Heat shock protein family A (Hsp70) member 4 (HSPA4) was formerly reported to involve inflammation responses and could be used as a biomarker for early lung cancer diagnosis and glioma outcome evaluation [25, 26]. It has been found extracellular HSP70 could activate ERK1/2, NF- $\kappa$ B, and proinflammatory genes transcription in lung cancer cell, and in breast cancer, tumor-educated B cells could target HSPA4 to promote lymph node metastasis through Src/NF- $\kappa$ B pathway [27, 28]. Also, Th2 cells in allergic disease could increase Hsp70, which was involved in pathogenesis [29]. N-myc downregulated 1 (NDRG1) was a tumor suppressor, which has been reported in various studies; it could attenuate NF- $\kappa$ B and TGF- $\beta$  pathways in pancreatic cancer cells, and downregulated NDRG1 was related to increased proliferation, invasion, and migration of digestive cancers [30–33]. It additionally involves apoptosis, glycolytic, and lipid metabolism in cancer cells, and virus infection process

was also related to NDRG1 [34–40]. Another signature, mitogen-activated protein kinase 2 (MAP4K2), was involved in MAPK signaling pathway, which corroborated with the enrichment results. We thought these pathways regulated by the three signatures involved crucial proliferation and progression process of tumors, and they all influence HCC tumor immune microenvironment through unrevealed mechanisms.

In the immune-risk model, bromodomain containing 8 (BRD8) was an androgen receptor coactivator, while TNF receptor associated factor 3 (TRAF3) was also involved in functioning of a variety of receptors. The knowledge of BRD8 in immune regulation was rare; however, it was related to p53-dependent apoptosis and could be a chemosensitizing target in colorectal cancer [41–43]. TRAF3 has been found to interplay with toll like receptors and TNF receptors in lymphocytes, and previous studies have found TRAF3 could attenuate noncanonical NF- $\kappa$ B pathway, influencing B cell and T cell development and recruitment

through chemokine regulation [44–49]. Kit ligand (KITLG or stem cell factor, SCF) is a cytokine, which has been testified to be expressed by both cancer cells and immune cells and related to tumor growth, metastasis, and stemness [50, 51]. In HCC, KITLG has been found to be an independent prognostic factor, and it can bind to c-kit receptor expressed by various immune cell types, leading to pathological process of allergy [52–55]. High expressions of BRD8, TRAF3, and KITLG were related to hyperfunction of receptors in tumor or immune cell populations, which could in turn influence the survival of patients. We thought high expression of the three markers in HCC was related to exhaustive immune status, indicating highly communicative signals in tumor microenvironment, in which ICB treatment could yield more benefits.

Overall, the immune-risk model and the combined nomogram are of great value in clinical practice for prognostic prediction. The immune microenvironment difference between high and low immune-risk score patients may decide patients' responses to immune treatment, and understanding the immunogenetic changes or patterns of immune infiltration in HCC can optimize treating strategies and drug application [56].

There are several limitations about our investigation. Firstly, expression analysis through bioinformatic methods still needs tissue sample confirmation in consideration of other potential confounding factors, such as ethnic bias. Also, the immune status behind immunogenetic changes requires further exploration of infiltrated immune cells, which were simply estimated in our study.

## 5. Conclusion

The immune-risk model, based on expression of SPP1, BRD8, NDRG1, KITLG, HSPA4, TRAF3, ITGAV, and MAP4K2, can efficiently differentiate HCC patients into high and low immune-risk subpopulations, and in combination with tumor stages, the derived nomogram can precisely predict the 1-, 3-, and 5-year overall survival among HCC patients, providing a tool for prognostic prediction.

## Data Availability

The datasets generated and/or analyzed during the current study are available in the TCGA (<https://portal.gdc.cancer.gov/>), GEO (<https://www.ncbi.nlm.nih.gov/geo/>), and ICGC (<https://icgc.org/>) repository.

## Additional Points

**Summary.** Hepatocellular carcinoma has been one of the most malignant tumors, conflicting a large number of patients worldwide with poor responses to drugs. The high recurrent rate after resection makes outcomes of patients even worse. Though new immune targeting drugs approved by federal government showed some light in treatment, the overall responses among patients are still unsatisfactory. As a result, the prognostic prediction of patients is very important

in consideration of treating strategies and patient monitoring. Former studies tried to find the biomarkers in HCC for diagnosis and prognosis, the scope of immunogenetic changes have seldom been investigated. Our study managed to find the significantly changed immune signatures in HCC, and based on survival-related immune genes, we constructed an immune-risk model for prognostic prediction. The high-risk patients shared worse outcomes, and the enriched pathways demonstrated the protumor growth immune microenvironment. Also, the prognostic value of immune-risk score in HCC patients was independent on immune infiltration levels, while it did positively correlate with infiltrating immune cells, indicating high-risk patients may have better ICB therapy efficacy with immune exhaustive status. In combination with tumor stages, the nomogram showed high accuracy in prediction of 1-, 3-, and 5-year overall survival, which is valuable for clinical evaluation in HCC treatment.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Jun Liu and Zheng Chen were both involved in analysis and interpretation of the study and drafted the manuscript equally. Wenli Li took part in the design and data collection process of the study and is responsible for the content of the manuscript. All authors read and approved the final manuscript.

## Supplementary Materials

Supplemental Figure 1. Differentially expressed transcription factors in HCC tissues and the regulation network between transcription factors and immune signatures. Supplemental Figure 2. PFS difference between high and low immune-risk score patients in TCGA dataset and the respective prognostic values for 0.5-, 1-, 2-, 3-, and 5-year survival. Supplemental Figure 3. Comprehensive immune status evaluation of high and low immune-risk HCC patients. (*Supplementary Materials*)

## References

- [1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics," *CA: A Cancer Journal for Clinicians*, vol. 65, no. 2, pp. 87–108, 2012.
- [2] J. Hartke, M. Johnson, and M. Ghabril, "The diagnosis and treatment of hepatocellular carcinoma," *Seminars in Diagnostic Pathology*, vol. 34, no. 2, pp. 153–159, 2017.
- [3] A. Jindal, A. Thadi, and K. Shailubhai, "Hepatocellular carcinoma: etiology and current and future drugs," *Journal of Clinical and Experimental Hepatology*, vol. 9, no. 2, pp. 221–232, 2019.
- [4] L. Chen and X. Han, "Anti-PD-1/PD-L1 therapy of human cancer: past, present, and future," *Journal of Clinical Investigation*, vol. 125, no. 9, pp. 3384–3391, 2015.
- [5] J. Long, L. Zhang, X. Wan et al., "A four-gene-based prognostic model predicts overall survival in patients with



- hepatocellular carcinoma," *Journal of Cellular and Molecular Medicine*, vol. 22, no. 12, pp. 5928–5938, 2018.
- [6] Z. Wan, X. Zhang, Y. Luo, and B. Zhao, "Identification of hepatocellular carcinoma-related potential genes and pathways through bioinformatic-based analyses," *Genetic Testing and Molecular Biomarkers*, vol. 23, no. 11, pp. 766–777, 2019.
- [7] T. J. Hudson, T. J. Hudson, W. Anderson et al., "International network of cancer genome projects," *Nature*, vol. 464, no. 7291, pp. 993–998, 2010.
- [8] K. Tomczak, P. Czerwinska, and M. Wiznerowicz, "The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge," *Contemporary Oncology (Poznan, Poland)*, vol. 19, no. 1a, pp. A68–A77, 2015.
- [9] 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.
- [10] 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.
- [11] V. K. Mootha, C. M. Lindgren, K.-F. Eriksson et al., "PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes," *Nature Genetics*, vol. 34, no. 3, pp. 267–273, 2003.
- [12] A. Subramanian, P. Tamayo, V. K. Mootha et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences*, vol. 102, no. 43, pp. 15545–15550, 2005.
- [13] 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.
- [14] 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.
- [15] R. Jia, Y. Liang, R. Chen et al., "Osteopontin facilitates tumor metastasis by regulating epithelial-mesenchymal plasticity," *Cell Death & Disease*, vol. 7, no. 12, Article ID e2564, 2016.
- [16] C. Hao, Y. Cui, M. Hu et al., "OPN-a splicing variant expression in non-small cell lung cancer and its effects on the bone metastatic abilities of lung cancer cells in vitro," *Anti-cancer Research*, vol. 37, no. 5, pp. 2245–2254, 2017.
- [17] C. Xu, L. Sun, C. Jiang et al., "SPPI, analyzed by bioinformatics methods, promotes the metastasis in colorectal cancer by activating EMT pathway," *Biomedicine & Pharmacotherapy*, vol. 91, pp. 1167–1177, 2017.
- [18] G. D'Addazio, L. Artese, T. Traini, C. Rubini, S. Caputi, and B. Sinjari, "Immunohistochemical study of osteopontin in oral squamous cell carcinoma allied to fractal dimension," *Journal of Biological Regulators and Homeostatic Agents*, vol. 32, no. 4, pp. 1033–1038, 2018.
- [19] M. Assidi, W. Goma, M. Jafri et al., "Prognostic value of Osteopontin (SPP1) in colorectal carcinoma requires a personalized molecular approach," *Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine*, vol. 41, no. 9, Article ID 1010428319863627, 2019.
- [20] E. R. P. Gimba, M. C. M. Brum, and G. Nestal De Moraes, "Full-length osteopontin and its splice variants as modulators of chemoresistance and radioresistance (Review)," *International Journal of Oncology*, vol. 54, no. 2, pp. 420–430, 2019.
- [21] A. Kamal, R. K. Darwish, S. Saad et al., "Association of osteopontin gene polymorphisms with colorectal cancer," *Cancer Investigation*, vol. 35, no. 2, pp. 71–77, 2017.
- [22] L. K. Alsarkhi and G. F. Weber, "Antiosteopontin autoantibodies in various types of cancer," *Oncology Reports*, vol. 40, no. 6, pp. 3879–3889, 2018.
- [23] G. Liu, X. Fan, M. Tang et al., "Osteopontin induces autophagy to promote chemo-resistance in human hepatocellular carcinoma cells," *Cancer Letters*, vol. 383, no. 2, pp. 171–182, 2016.
- [24] M. Cabiati, M. Gaggini, M. M. Cesare et al., "Osteopontin in hepatocellular carcinoma: a possible biomarker for diagnosis and follow-up," *Cytokine*, vol. 99, pp. 59–65, 2017.
- [25] T. Tang, C. Yang, H. E. Brown, and J. Huang, "Circulating heat shock protein 70 is a novel biomarker for early diagnosis of lung cancer," *Disease Markers*, vol. 5, 2018.
- [26] F. Lämmer, C. Delbridge, S. Würstle et al., "Cytosolic Hsp70 as a biomarker to predict clinical outcome in patients with glioblastoma," *PLoS One*, vol. 14, Article ID e0221502, 2019.
- [27] N. Somensi, P. O. Brum, V. de Miranda Ramos et al., "Extracellular HSP70 activates ERK1/2, NF- $\kappa$ B and pro-inflammatory gene transcription through binding with RAGE in A549 human lung cancer cells," *Cellular Physiology and Biochemistry*, vol. 42, no. 6, pp. 2507–2522, 2017.
- [28] Y. Gu, Y. Liu, L. Fu et al., "Tumor-educated B cells selectively promote breast cancer lymph node metastasis by HSPA4-targeting IgG," *Nature Medicine*, vol. 25, no. 2, pp. 312–322, 2019.
- [29] H. J. Min, K. S. Kim, J.-H. Yoon, C.-H. Kim, and H.-J. Cho, "T-helper 2 cytokine-induced heat shock protein 70 secretion and its potential association with allergic rhinitis," *International Forum of Allergy & Rhinology*, vol. 7, no. 5, pp. 530–535, 2017.
- [30] S. V. Menezes, L. Fouani, M. L. H. Huang et al., "The metastasis suppressor, NDRG1, attenuates oncogenic TGF- $\beta$  and NF- $\kappa$ B signaling to enhance membrane E-cadherin expression in pancreatic cancer cells," *Carcinogenesis*, vol. 40, no. 6, pp. 805–818, 2019.
- [31] G. Cen, K. Zhang, J. Cao, and Z. Qiu, "Downregulation of the N-myc downstream regulated gene 1 is related to enhanced proliferation, invasion and migration of pancreatic cancer," *Oncology Reports*, vol. 37, no. 2, pp. 1189–1195, 2017.
- [32] K. Chen, X.-H. Liu, F.-R. Wang, H.-P. Liu, Z.-P. Huang, and X. Chen, "The prognostic value of decreased NDRG1 expression in patients with digestive system cancers," *Medicine*, vol. 97, no. 41, Article ID e12455, 2018.
- [33] T. Dai, Y. Dai, Y. Murata et al., "The prognostic significance of N-myc downregulated gene 1 in lung adenocarcinoma," *Pathology International*, vol. 68, no. 4, pp. 224–231, 2018.
- [34] D.-D. Guo, K.-F. Xie, and X.-J. Luo, "Hypoxia-induced elevated NDRG1 mediates apoptosis through reprogramming mitochondrial fission in HCC," *Gene*, vol. 741, Article ID 144552, 2020.
- [35] W. Liu, B. Zhang, Q. Hu et al., "A new facet of NDRG1 in pancreatic ductal adenocarcinoma: suppression of glycolytic metabolism," *International Journal of Oncology*, vol. 50, no. 5, pp. 1792–1800, 2017.
- [36] C. J. Sevinsky, F. Khan, L. Kokabee, A. Darehshouri, K. R. Maddipati, and D. S. Conklin, "NDRG1 regulates neutral lipid metabolism in breast cancer cells," *Breast Cancer Research*, vol. 20, p. 55, 2018.
- [37] X. Yang, F. Zhu, C. Yu et al., "N-myc downstream-regulated gene 1 promotes oxaliplatin-triggered apoptosis in colorectal cancer cells via enhancing the ubiquitination of Bcl-2," *Oncotarget*, vol. 8, no. 29, pp. 47709–47724, 2017.
- [38] L. Chen, C. Xing, G. Ma et al., "N-myc downstream-regulated gene 1 facilitates influenza A virus replication by suppressing

- canonical NF- $\kappa$ B signaling,” *Virus Research*, vol. 252, pp. 22–28, 2018.
- [39] P. Gupta, N. Shahzad, A. Harold et al., “Merkel cell polyomavirus downregulates N-myc downstream-regulated gene 1, leading to cellular proliferation and migration,” *Journal of Virology*, vol. 94, no. 3, 2020.
- [40] C. J. Schweitzer, F. Zhang, A. Boyer, K. Valdez, M. Cam, and T. J. Liang, “N-myc downstream-regulated gene 1 restricts hepatitis C virus propagation by regulating lipid droplet biogenesis and viral assembly,” *Journal of Virology*, vol. 92, no. 2, 2017.
- [41] C.-Y. Jiang, Y. Ruan, X.-H. Wang et al., “MiR-185 attenuates androgen receptor function in prostate cancer indirectly by targeting bromodomain containing 8 isoform 2, an androgen receptor co-activator,” *Molecular and Cellular Endocrinology*, vol. 427, pp. 13–20, 2016.
- [42] A. Lashgari, M. Fauteux, A. Maréchal, and L. Gaudreau, “Cellular depletion of BRD8 causes p53-dependent apoptosis and induces a DNA damage response in non-stressed cells,” *Scientific Reports*, vol. 8, p. 14089, 2018.
- [43] H. Y. Yamada and C. V. Rao, “BRD8 is a potential chemosensitizing target for spindle poisons in colorectal cancer therapy,” *International Journal of Oncology*, vol. 35, pp. 1101–1109, 2009.
- [44] G. A. Bishop, L. L. Stunz, and B. S. Hostager, “TRAF3 as a multifaceted regulator of B lymphocyte survival and activation,” *Frontiers in Immunology*, vol. 9, p. 2161, 2018.
- [45] H. Häcker, P.-H. Tseng, and M. Karin, “Expanding TRAF function: TRAF3 as a tri-faced immune regulator,” *Nature Reviews Immunology*, vol. 11, no. 7, pp. 457–468, 2011.
- [46] Z. Yi, A. M. Wallis, and G. A. Bishop, “Roles of TRAF3 in T cells: many surprises,” *Cell Cycle*, vol. 14, no. 8, pp. 1156–1163, 2015.
- [47] J. M. Hildebrand, Z. Yi, C. M. Buchta, J. Poovassery, L. L. Stunz, and G. A. Bishop, “Roles of tumor necrosis factor receptor associated factor 3 (TRAF3) and TRAF5 in immune cell functions,” *Immunological Reviews*, vol. 244, no. 1, pp. 55–74, 2011.
- [48] F. Liu, L. Cheng, J. Xu, F. Guo, and W. Chen, “miR-17-92 functions as an oncogene and modulates NF- $\kappa$ B signaling by targeting TRAF3 in MGC-803 human gastric cancer cells,” *International Journal of Oncology*, vol. 53, no. 5, pp. 2241–2257, 2018.
- [49] G. Perez-Chacon, M. Adrados, M. T. Vallejo-Cremades, S. Lefebvre, J. C. Reed, and J. M. Zapata, “Dysregulated TRAF3 and BCL2 expression promotes multiple classes of mature non-hodgkin B cell lymphoma in mice,” *Frontiers in Immunology*, vol. 9, p. 3114, 2018.
- [50] S. Q. Shen, R. Wang, and S. G. Huang, “Expression of the stem cell factor in fibroblasts, endothelial cells, and macrophages in periapical tissues in human chronic periapical diseases,” *Genetics and Molecular Research*, vol. 16, no. 1, 2017.
- [51] L. Wang, J. Wang, Z. Li et al., “Silencing stem cell factor attenuates stemness and inhibits migration of cancer stem cells derived from Lewis lung carcinoma cells,” *Tumor Biology*, vol. 37, no. 6, pp. 7213–7227, 2016.
- [52] X. Wang, H. Ren, T. Zhao et al., “Stem cell factor is a novel independent prognostic biomarker for hepatocellular carcinoma after curative resection,” *Carcinogenesis*, vol. 35, no. 10, pp. 2283–2290, 2014.
- [53] W. Fonseca, A. J. Rasky, C. Ptaschinski et al., “Group 2 innate lymphoid cells (ILC2) are regulated by stem cell factor during chronic asthmatic disease,” *Mucosal Immunology*, vol. 12, no. 2, pp. 445–456, 2019.
- [54] K.-A. Cho, M. Park, Y.-H. Kim, and S.-Y. Woo, “Th17 cell-mediated immune responses promote mast cell proliferation by triggering stem cell factor in keratinocytes,” *Biochemical and Biophysical Research Communications*, vol. 487, no. 4, pp. 856–861, 2017.
- [55] N. Al-Azzam, V. Kondeti, E. Duah, F. Gombedza, C. K. Thodeti, and S. Paruchuri, “Modulation of mast cell proliferative and inflammatory responses by leukotriene D4 and stem cell factor signaling interactions,” *Journal of Cellular Physiology*, vol. 230, no. 3, pp. 595–602, 2015.
- [56] Q.-F. Chen, W. Li, P.-H. Wu, L.-J. Shen, and Z.-L. Huang, “Significance of tumor-infiltrating immunocytes for predicting prognosis of hepatitis B virus-related hepatocellular carcinoma,” *World Journal of Gastroenterology*, vol. 25, no. 35, pp. 5266–5282, 2019.