

# Development and validation of a robust immune-related risk signature for hepatocellular carcinoma

**Zaoqu Liu**

Zhengzhou University First Affiliated Hospital <https://orcid.org/0000-0002-0452-742X>

**Dechao Jiao**

Zhengzhou University First Affiliated Hospital

**Xueliang Zhou**

Zhengzhou University First Affiliated Hospital

**Yuan Yao**

Zhengzhou University First Affiliated Hospital

**Zhaonan Li**

Zhengzhou University First Affiliated Hospital

**Jing Li**

Zhengzhou University First Affiliated Hospital

**Jianjian Chen**

Zhengzhou University First Affiliated Hospital

**Qinyu Lei**

Zhengzhou University First Affiliated Hospital

**Xinwei Han** (✉ [fcchanxw@zzu.edu.cn](mailto:fcchanxw@zzu.edu.cn))

<https://orcid.org/0000-0003-4407-4864>

---

## Primary research

**Keywords:** hepatocellular carcinoma (HCC), immune-related genes, prognosis, tumor immunology, signature, nomogram

**Posted Date:** March 9th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-16346/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Medicine on March 12th, 2021. See the published version at <https://doi.org/10.1097/MD.0000000000024683>.

# Abstract

**Background:** A growing amount of evidence has suggested immune-related genes (IRGs) play a key role in the development of hepatocellular carcinoma (HCC). However, there have been no investigations proposing a reliable prognostic signature in terms of tumor immunology. This study aimed to develop a robust signature based on IRGs in HCC.

**Methods:** A total of 597 HCC patients were enrolled. The TCGA database was utilized for discovery, and the ICGC database was utilized for validation. Multiple algorithms (including univariate Cox, LASSO, and multivariate Cox regression) were performed to identify key prognostic IRGs and establish an immune-related risk signature. Bioinformatics analysis and R soft tools were utilized to annotate underlying biological functions.

**Results:** A total of 1416 differentially expressed mRNAs (DEMs) were screened in the TCGA cohort, of which 90 were differentially expressed IRGs (DEIRGs). Using univariate Cox regression analysis, we identified 33 prognostically relevant DEIRGs. Using LASSO regression and multivariate Cox regression analysis, we extracted 8 optimal DEIRGs (APLN, CDK4, CXCL2, ESR1, IL1RN, PSMD2, SEMA3F, and SPP1) to construct a risk signature with the ability to distinguish cases as having a high or low risk of unfavorable prognosis in the TCGA cohort, and the signature was verified in the ICGC cohort. The signature was prognostically significant in all stratified cohorts and was deemed an independent prognostic factor for HCC. We also built a nomogram with good performance by combining the signature with clinicopathological factors to increase the accuracy of predicting HCC prognosis. By investigating the relationship of the risk score and 8 risk genes from our signature with clinical traits, we found that the aberrant expression of the immune-related risk genes is correlated with the development of HCC. Moreover, the high-risk group was higher than the low-risk group in terms of tumor mutation burden (TMB), immune cell infiltration, and the expression of immune checkpoints (PD-1, PD-L1, and CTLA-4), and functional enrichment analysis indicated the signature enriched an intensive immune phenotype.

**Conclusions:** This study developed a robust immune-related risk signature and built a predictive nomogram that reliably predict overall survival in HCC, which may be helpful for clinical management and personalized immunotherapy decisions.

## Background

Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer and mostly results from viral infection and liver fibrosis [1]. It is the sixth leading cause of malignancy worldwide and the third most common cause of cancer-related death [2]. There are approximately 437000 people in the world who are diagnosed with HCC each year, of which approximately 50% of cases occur in China [3]. The main reason for the poor prognosis of HCC is the lack of early and effective biomarkers and therapeutic treatments. Thus, further exploring effective biomarkers and understanding the potential

molecular mechanisms of tumorigenesis and progression in HCC may improve the overall survival and therapeutic effect of HCC patients.

It has been confirmed that the derangement of the immune response in the tumor microenvironment (TME) actively participates in tumor aggressiveness and progression. The TME facilitates tumor cell survival, proliferation, and metastasis and enhances the ability to evade host immune responses [4]. Abundant efforts have been conducted to understand the relationship between tumors and the immune response, and tumor immunotherapy has achieved remarkable success in advancing tumor treatment [5]. The development of immune checkpoint inhibitors (including nivolumab and pembrolizumab) has boosted their application in HCC, but they are only effective in a subset of patients [6, 7]. Hence, it is necessary to identify a robust immune signature that predicts the sensitivity of patients to immunotherapy to realize individualized treatment and that may become a new target for immunotherapy in HCC.

Prior studies have shed light on immune-related genes (IRGs) that not only play an important role in the process of immunotherapy but are also associated with the prognosis of patients with HCC. However, these studies focused on investigating the relationships of a single IRG with the prognosis of HCC. Few studies have explored the correlations between multiple IRGs and the prognosis of HCC. A recent report illustrated that multigene immune signatures might be a more effective tool for selecting candidates for treatment[8].

In the present study, we utilized IRG expression data based on The Cancer Genome Atlas (TCGA) and Immunology Database and Analysis Portal (ImmPort) to establish an immune-related risk signature consisting of 8 IRGs for HCC and validated it in the International Cancer Genome Consortium (ICGC) database. The signature was combined with clinical factors to further develop a predictive nomogram, enabling improved prognostic assessment and clinical management of HCC patients. To investigate the ability of our signature to predict the progression of HCC, we analyzed the relationships between the risk factors from our signature (the risk score and risk genes) and the clinical parameters. Moreover, to further explore the underlying immune-related molecular mechanisms and functions of the genes in the signature, we performed research on tumor mutation burden (TMB), immune cell infiltration, and functional enrichment in the HCC.

## Methods

### Sample datasets and IRGs

The RNA-Seq gene expression data and corresponding clinical data of HCC samples were collected from the TCGA database (<https://portal.gdc.cancer.gov/>). The transcriptome profiling of RNA expression was measured in fragments per kilobase of exon model per million mapped reads or FPKM values. Log<sub>2</sub> transformation was performed for the normalization of RNA expression profiles. After eliminating genes with FPKM values equal to 0 in more than 50% of the samples, 18947 protein-coding genes were used for further analysis. A total of 1534 immune-related genes (IRGs) were obtained from the ImmPort database

(<https://import.niaid.nih.gov/>) (Additional file 1: Table S1). These IRGs play multiple roles in immune pathways, including antigen processing and presentation pathways, the B cell receptor signaling pathway; cytokine, cytokine receptor, interleukin, and NK cell cytotoxicity pathways; transforming growth factor- $\beta$  family member pathways; the T cell receptor signaling pathway; and TNF family member receptor pathways. The IRGs were prepared for the construction of an immune-related risk signature.

### **Differential expression analysis (DEA)**

Differentially expressed mRNAs (DEMs) were identified by the R package “limma”, and P values were corrected using the Benjamini & Hochberg false discovery rate (FDR) method. The following cut-off thresholds were set to screen DEMs:  $|\log_2$  fold change (FC)| $>1$  and FDR  $< 0.05$ . The intersection of these DEMs and 1534 IRGs was used to obtain differentially expressed immune-related genes (DEIRGs).

### **Construction of the immune-related risk signature**

A total of 365 HCC patients with survival data in the TCGA cohort were randomly divided into a training set (n=255; 70%) and a testing set (n=110; 30%). First, univariate Cox regression analysis of the DEIRGs was implemented to determine the correlation between the survival of HCC patients and the expression level of the DEIRGs in the training set. DEIRGs with P  $< 0.05$  were considered prognostic IRGs and extracted for least absolute shrinkage and selection operator (LASSO) Cox regression analysis to further select key prognostic IRGs for overall survival in patients. The hazard ratio (HR) from the univariate Cox regression analysis was used to identify the protective (HR  $< 1$ ) and risk genes (HR  $> 1$ ). Next, multivariate Cox regression analysis was performed on the DEIRGs obtained from the LASSO Cox regression analysis. Then, an immune-related risk signature was constructed based on a linear combination of the expression value of the prognostic DEIRGs weighted by the regression coefficient ( $\beta$ ) derived from the multivariate Cox regression analysis, as follows:

**See formula 1 in the supplementary files.**

where n is the number of key prognostic IRGs,  $exp_i$  is the expression value of the key prognostic IRGs i, and  $\beta_i$  is the regression coefficient of IRGs i. Using the median risk score in the training set as the cut-off value, HCC patients were classified into high- and low-risk groups in the training set, testing set and entire set. Subsequently, we analyzed the expression differences of the key prognostic IRGs in the high- and low-risk groups in the entire set. The area under the curve (AUC) of the time-dependent receiver operating characteristic (ROC) curve was calculated to evaluate the prognostic ability of the immune-related risk signature for overall survival using the R package “timeROC”. The Kaplan–Meier method was used to calculate survival curves, and the log-rank test was performed to compare the survival differences in the high- and low-risk groups using the R package “survival”. Then, the prognostic ability of the immune-related risk signature was further investigated in the testing set and entire set.

### **External validation of the immune-related risk signature**

To evaluate the predictive accuracy of the immune-related risk signature from the TCGA cohort, the ICGC database (<https://icgc.org/>) was further analyzed for validation. We processed the data from the ICGC cohort in the same way as the TCGA cohort. For each included patient, the risk score was calculated using the same formula. Subsequently, the ROC curve and the Kaplan-Meier survival curve were constructed to test the predictive value of the signature.

All data used were from the TCGA and ICGC databases, which are both publicly available and open access, so additional approval was not required from the ethics committee.

### **Independent prognostic role of the immune-related risk signature**

Next, we investigated whether the prognostic power of the immune-related risk signature was influenced by other clinicopathological factors, including sex, age, body mass index (BMI), alpha-fetoprotein (AFP), histological grade, BCLC (Barcelona Clinic Liver Cancer) stage, and vascular invasion. Univariate and multivariate Cox regression analyses were carried out to determine independent factors of prognosis in the entire set.  $P < 0.05$  was considered statistically significant. Factors with  $P < 0.05$  based on univariate and multivariate analyses were deemed independent prognostic factors.

### **Establishment of a new predictive nomogram**

All independent prognostic factors were included to establish a new predictive nomogram via a stepwise Cox regression model to investigate the probability of the 1-, 2-, and 3-year overall survival of HCC patients. Kaplan-Meier analysis, the AUC of the time-ROC curve, Harrell's concordance index (C-index) and a calibration curve were used to assess the performance of the nomogram. The C-index was calculated to evaluate the discrimination of the nomogram by a bootstrap method with 1000 resamples. The nomogram calibration curve was plotted to compare the predicted and observed overall survival.

### **Mutation and immune analysis of the signature**

The mutation data available for HCC patients were obtained from the TCGA database (<https://portal.gdc.cancer.gov/>). We calculated the total number of mutations for each sample and then compared the mutation counts of the high- and low-risk groups in the entire set. In addition, to determine whether our signature could influence the status of the tumor immune microenvironment in HCC patients, the tumor immune estimation resource (TIMER) algorithm (<https://cistrome.shinyapps.io/timer/>) was used to estimate the abundances of B cells, CD4+ T cells, CD8+ T cells, dendritic cells, neutrophils, and macrophages in the data from the TCGA database, and then the relationships between the risk score and immune cell infiltration in the entire set were investigated.

### **Functional enrichment analysis**

To explore the underlying biological processes and pathways of the immune-related risk signature, we utilized the R package "clusterProfiler" to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes

and Genomes (KEGG) pathway enrichment analysis. GO biological processes and KEGG pathways having both  $P < 0.05$  and  $FDR < 0.05$  were considered to be significant.

## Results

### DEMs and DEIRGs in patients with HCC

Fig. 1 displays the flow diagram of our study. The analysis of mRNA expression profiles between HCC tissues and adjacent normal liver tissues identified 1416 DEMs (Fig. 2a, b). A total of 1016 mRNAs were significantly upregulated and 400 mRNAs were significantly downregulated in HCC tissues compared to normal liver tissues. The IRGs among the 1416 DEMs were identified, and a total of 90 DEIRGs, including 49 upregulated mRNAs and 41 downregulated mRNAs, were filtered for further analysis (Fig. 2c, d).

### Construction of the immune-related risk signature

In the TCGA cohort, the training set was used to construct the immune-related risk signature. The testing set and entire set were used for validation. Using univariate Cox regression analysis, we identified 33 IRGs with prognostic prediction capacity from a total of 90 DEIRGs (Additional file 2: Figure S1). The 33 IRGs were enrolled in LASSO Cox regression analysis. After 100 rounds of 10-fold cross-validation, 8 optimal IRGs were identified when lambda took the minimum value of 0.059, and the 8 IRGs were used to establish an immune-related risk signature for HCC patients (Fig. 3a, b). The results of the univariate and multivariate Cox regression analyses of the 8 optimal IRGs are shown in Table 1. The formula of the risk score is as follows:

**Table 1 Univariate and multivariate analysis of the 8 optimal IRGs**

Gene ID	Univariate Analysis		Univariate Analysis	
	HR (95%CI)	P-value	HR (95%CI)	P-value
APLN	1.298 (1.066, 1.582)	0.010	1.178 (0.941, 1.476)	0.154
CDK4	1.721 (1.360, 2.176)	0.000	1.289 (0.967, 1.717)	0.083
CXCL2	0.826 (0.724, 0.943)	0.005	0.924 (0.790, 1.079)	0.318
ESR1	0.514 (0.321, 0.821)	0.005	0.790 (0.479, 1.302)	0.355
IL1RN	0.739 (0.624, 0.875)	0.000	0.819 (0.680, 0.987)	0.036
PSMD2	2.334 (1.610, 3.383)	0.000	1.180 (0.739, 1.886)	0.488
SEMA3F	1.591 (1.101, 2.299)	0.013	1.279 (0.877, 1.865)	0.201
SPP1	1.131 (1.060, 1.208)	0.000	1.078 (0.993, 1.170)	0.073

$$\text{Risk score} = (0.164 * \text{APLN}) + (0.254 * \text{CDK4}) + (-0.079 * \text{CXCL2}) + (-0.236 * \text{ESR1}) + (-0.199 * \text{IL1RN}) + (0.166 * \text{PSMD2}) + (0.246 * \text{SEMA3F}) + (0.075 * \text{SPP1})$$

The risk scores for patients were calculated, and patients were divided into high- and low-risk groups using the median risk score of the training set as the cut-off threshold in each set. As shown in Fig. 3c-j, among the 8 genes of the immune-related risk signature, significantly elevated APLN, CDK4, PSMD2,

SEMA3F, and SPP1 expression was found in high-risk group compared with low-risk group, while CXCL2, ESR1, and IL1RN expression was significantly decreased in high-risk group compared with low-risk group (all  $P < 0.001$ ). The ROC curves displayed that the AUCs of the immune-related risk signature at 1, 2 and 3 years were 0.813, 0.717, and 0.732 for the training set; 0.631, 0.690, and 0.755 for the testing set; and 0.764, 0.710, and 0.728 for the entire set, respectively (Fig. 4a-c left panel). The distribution of the risk scores, the survival status of the patients ranked according to the risk scores and the expression values of the 8 genes are shown in Fig. 4a-c middle panel. The Kaplan-Meier survival curve combined with the log-rank test showed that patients in the high-risk group had a significantly shorter survival time compared to patients in the low-risk group (all  $P < 0.05$ ) (Fig. 4a-c right panel). Our results indicated a good performance of the immune-related risk signature for the survival prediction of HCC patients.

### **External validation of the immune-related risk signature**

To avoid bias arising from only investigating the TCGA cohort, we enrolled a total of 232 HCC samples in the ICGC database to validate the predictive value of the signature. We utilized the same formula to calculate the risk score for each patient and then classified the patients into high- and low-risk groups using the same cut-off threshold. As shown in Fig. 4d, the AUCs for 1, 2, and 3 years were 0.704, 0.758, and 0.787, respectively. Consistent with the results in the TCGA cohort, patients in the high-risk group had significantly poorer overall survival than those in the low-risk group. Overall, our signature was able to accurately predict the overall survival of HCC patients.

### **Independent prognostic role of the immune-related risk signature**

A total of 238 HCC patients in the TCGA cohort with complete clinical information (including sex, age, BMI, AFP, histological grade, BCLC stage, and vascular invasion), were included for further analysis. Univariate and multivariate Cox regression analyses indicated that BCLC stage (HR=2.135,  $P=0.004$ ), vascular invasion (HR=2.142,  $P=0.003$ ) and risk score (HR=1.612,  $P < 0.001$ ) calculated from the immune-related risk signature were independent prognostic factors for the overall survival of HCC patients (Fig. 5a, b). In addition, the prediction accuracy of the signature was compared with all clinical factors. As shown in Fig. 5c-e, the signature had the largest AUC at 1, 2, and 3 year, this suggested the predictive performance of the signature was superior to any other clinical factors. Subsequently, all patients were stratified into subgroups based on their clinicopathological factors, such as sex (female/male), age (<60/ $\geq 60$ ), BMI (<25/ $\geq 25$ ), AFP (<400/ $\geq 400$ ), tumor grade (G1+2/G3+4), BCLC stage (I+II/III+IV), and vascular invasion (no/yes), and each subgroup was further divided into the low- and high-risk groups using the immune-related risk signature (Table 2). The results of stratification analysis showed that the high-risk patients in each stratum of all clinicopathological factors had significantly shorter survival times than the low-risk patients (all  $P < 0.05$ ) (Fig. 5f-s). Our findings suggest that the immune-related risk signature is a promising independent risk factor for the prognosis assessment of HCC.

### **Table 2** Clinicopathological factors of HCC patients from TCGA cohort in the low- and high-risk groups

Factor	Risk score		P value
	Low	High	
Sex			
Male	81 (72.3%)	80 (63.5%)	0.189
Female	31 (27.7%)	46 (36.5%)	
Age			
< 60	54 (48.2%)	61 (48.4%)	1.000
≥60	58 (51.8%)	65 (51.6%)	
BMI			
< 25	64 (57.1%)	60 (47.6%)	0.181
≥25	48 (42.9%)	66 (52.4%)	
AFP			
□400	95 (84.8%)	87 (69.0%)	0.007
≥400	17 (15.2%)	39 (31.0%)	
Grade			
G1+2	76 (67.9%)	57 (45.2%)	0.001
G3+4	36 (32.1%)	69 (54.8%)	
BCLC stage			
□+□	96 (85.7%)	99 (78.6%)	0.207
□+□	16 (14.3%)	27 (21.4%)	
Vascular invasion			
Yes	80 (71.4%)	72 (57.1%)	0.031
No	32 (28.6%)	54 (42.9%)	

### Establishment of a new predictive nomogram

We then established a new nomogram to predict 1-, 2-, and 3-year overall survival in 238 HCC patients from the TCGA cohort using three independent prognostic factors, including BCLC stage, vascular invasion and risk score (Fig. 6a). According to the total points of the nomogram, the patients were divided into high- and low-point groups using the median point value as the cut-off threshold. Survival curves for the high- and low-point groups were plotted by Kaplan-Meier analysis, and the results showed that those with higher points had significantly poorer overall survival ( $P = 0.0014$ ) (Fig. 6b). The AUCs of the 1-, 2-, and 3-year overall survival predictions for the nomogram were 0.750, 0.693, and 0.733, respectively (Fig. 6c). The C-index for the nomogram was 0.689 (95% confidence interval [CI] 0.629–0.749). The calibration curve showed that the nomogram performed well at predicting overall survival in HCC patients, but the nomogram might underestimate or overestimate mortality (Fig. 6d-f).

### Clinical correlation of the immune-related risk signature

To investigate the ability of our signature to predict the progression of HCC, we analyzed the relationships between the risk factors from our signature (the risk score and 8 risk genes) and clinical parameters (sex, age, BMI, AFP, histological grade, BCLC stage, and vascular invasion) in the entire set. The results indicated that the expression of CDK4 and SEMA3F was higher in female patients, and the expression of ESR1 was higher in male patients (all  $P < 0.05$ ) (Fig. 7a-c). In patients older than 60 years, the expression of CXCL2 and ESR1 was higher than in those younger than 60 years (both  $P < 0.05$ ) (Fig. 7d, e). The

expression of CXCL2 was higher in patients with BMI  $\geq 25$  than in those with BMI  $< 25$  ( $P < 0.05$ ) (Fig. 7f). Patients with AFP  $\geq 400$  had a higher expression of CDK4 and SEMA3F but a lower expression of CXCL2 and ESR1 than patients with AFP  $< 400$  (Fig. 7g-j). As the values of certain factors increased (CDK4, PSMD2, and SPD1 levels and the risk score), the histological grade of HCC patients increased, and histological grade decreased in patients with increased CXCL2 and ESR1 expression (all  $P < 0.05$ ) (Fig. 7k-p). And the higher the expression of SEMA3F, the higher the patient's BCLC stage ( $P < 0.05$ ) (Fig. 7q). Patients with vascular invasion had higher SPP1 levels and risk scores and lower ESR1 expression than patients without vascular invasion (all  $P < 0.05$ ) (Fig. 7r-t). These results indicated that the dysregulation of immune-related risk gene expression is correlated with the development of HCC.

### **Mutation and immune analysis of the signature**

As shown in Fig. 8a, the mutation count was greater in the high-risk group than in the low-risk group ( $P < 0.05$ ). Previous studies have suggested that the tumor mutation burden (TMB) is significantly related to the clinical effectiveness of immunotherapy[9]. The results demonstrated that our signature may stratify patients with different sensitivities to immunotherapy. To further identify more immune-related mechanisms of the signature, we analyzed the correlation between the risk score and immune checkpoints, including programmed cell death protein 1 (PD-1), programmed cell death ligand 1 (PD-L1), and cytotoxic T-lymphocyte-associated protein 4 (CTLA-4). We found that HCC patients in the high-risk group tended to have higher expression levels of PD-1 ( $P < 0.01$ ), PD-L1 ( $P < 0.05$ ), and CTLA-4 ( $P < 0.05$ ) (Fig. 8b-d). This finding suggests that high-risk HCC patients might better respond to immune checkpoint blockade (ICB) immunotherapy targeting PD-1 and CTLA-4. The analysis of immune cell infiltration showed that as the risk score increased, the abundances of immune cells (B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages and dendritic cells) in HCC tissues also increased (all  $P < 0.05$ ) (Fig. 8e-j).

### **Functional enrichment analysis**

To investigate the mechanisms underlying the immune-related risk signature, we examined the expression correlation between the risk genes from the signature and all IRGs using two-sided Pearson correlation coefficients in the entire set. As shown in Additional file 3: Table S2, a total of 37 IRGs were expressed as significantly correlated with at least one of the risk genes ( $|\text{Pearson correlation coefficient}| > 0.5$  and  $P < 0.01$ ). Then, we performed functional enrichment analysis with the clusterProfiler package to explore the underlying biological processes and pathways of the 37 IRGs, revealing 444 enriched biological processes (Additional file 4: Table S3) and 116 enriched KEGG pathways (Additional file 5: Table S4). The GO results of the top 10 immune-related biological processes are shown in Fig. 9a including Fc receptor signaling pathway, the immune response-regulating cell surface receptor signaling pathway, antigen receptor-mediated signaling pathway, regulation of innate immune response, response to tumor necrosis factor, cellular response to interleukin-1, neutrophil mediated immunity, regulation of leukocyte cell-cell adhesion, regulation of T cell activation, and regulation of leukocyte activation. The top 10 immune-related KEGG pathways are shown in Fig. 9b, including the T cell receptor signaling pathway, B cell receptor signaling pathway, viral carcinogenesis, chemokine signaling pathway, natural killer cell-

mediated cytotoxicity, PD-L1 expression and PD-1 checkpoint pathway in cancer, IL-17 signaling pathway, NF-kappa B signaling pathway, Th17 cell differentiation, and Th1 and Th2 cell differentiation. Therefore, our results indicated that the immune-related risk signature displays an intensive immune phenotype.

## Discussion

The immune response in the tumor microenvironment (TME) plays a critical role in tumor initiation and progression[10, 11]. The immune system can kill cancer cells and suppress cancer growth by upregulating or downregulating IRGs at certain immune checkpoints[12, 13]. However, because tumors have the characteristics of immune evasion, certain tumor cells can imitate the IRG expression patterns of normal cells to elude the attack from the immune system[14, 15]. Therefore, the expression of IRGs may be a vital predictor of HCC progression and prognosis. In the present study, we utilized the TCGA and ICGC cohorts to establish and validate a robust immune-related risk signature consisting of 8 IRGs for HCC that can predict patients' overall survival and was significantly related to clinicopathological factors. We also built a nomogram by combining the signature with clinicopathological factors to increase the accuracy in predicting HCC prognosis. By investigating the relationship of the risk score and 8 risk genes from our signature with clinical traits, we found that the aberrant expression of the immune-related risk genes was correlated with the development of HCC. More importantly, we found that the mutation status and the immune status were associated with this signature, and the signature enriched an intensive immune phenotype. These discoveries shed light on the value of our signature for HCC patient prognosis and on possible new immune targets for immunotherapy.

First, we screened 1416 DEMs, among which we found 90 DEIRGs, including 49 upregulated IRGs and 41 downregulated IRGs. We then sequentially performed univariate, LASSO, and multivariate Cox regression analyses to construct a immune-related risk signature. The signature consisted of 8 IRGs significantly related to HCC patient overall survival (including APLN, CDK4, CXCL2, ESR1, IL1RN, PSMD2, SEMA3F, and SPP1). For the genes in our signature, six of them were chemokines, cytokines or cytokine receptors that actively participate in the inflammatory process of tumor growth, progression and metastasis. Chemokines and cytokines produced by tumor cells and other cellular components of the TME are able to facilitate the occurrence of tumor-associated or tumor-elicited inflammation, and this type of inflammation is a key driver of tumor progression and an essential component of the TME[16, 17]. Chemokines and cytokines act in autocrine and paracrine manners through receptors on tumor cells to activate STAT3, AP1 and NFkB families of potential oncogenic transcription factors. They are also involved in the progression of tumor growth through the induction of survival factors and proliferation[18]. Thus, according to our signature, high-risk HCC patients may have increased inflammation in the tumor microenvironment, which promotes the progression of HCC and leads to the poor overall survival of patients. In addition, of the 8 IRGs that compose the signature, CDK4 was the most significant gene in the univariate Cox regression analysis. CDK4 is one of the core proteins of the cell cycle regulatory network, and changes in its expression activity directly affect cell cycle progression; moreover, its overexpression is closely related to tumor initiation, progression, and metastasis[19]. A recent study indicated that the expression of CDK4 can affect the prognosis of HCC patients, and the

higher its expression level is, the worse the prognosis[20]. CDK4 and CDK6 inhibitors have been considered the most promising cell cycle therapeutics, and intense efforts are now being conducted to expand the field of this paradigm[21]. PSMD2 is another gene in the signature that attracts our attention. The PSMD2 protein is a subunit of the 19S regulatory complex of the 26S proteasome and belongs to the ubiquitin–proteasome system (UPS)[22]. Prior studies demonstrated that knocking down PSMD2 can suppress tumor cell proliferation, and PSMD2 may serve as a potential therapeutic target for tumors[23, 24]. PSMD2 is able to modulate HCC cell proliferation by regulating cellular lipid metabolism. The knockdown of PSMD2 reduces the formation of cellular lipid droplets, and PSMD2 also regulates the expression level of genes involved in lipid synthesis through the p38-JNK and AKT signaling pathways[25]. These results further suggest the importance of our signature in the HCC microenvironment.

Next, we combined the signature with conventional clinical factors, including BCLC stage and vascular invasion, to develop a nomogram with good performance. This nomogram could be a valuable diagnostic and prognostic management tool for HCC. To explore the clinical utility of the signature, we evaluated the correlations of the risk factors in our signature with clinical factors. The risk score and 8 risk genes were significantly related to the progression of HCC. Our research identified the immune-related differences between different clinical subsets in HCC. Our findings also reflect that the expression differences of IRGs may lead to different states of the TME, which further affects the prognosis of different clinical subsets. A recent report suggested that the expression profile of IRGs in the TME of HCC can predict tumor progression and patient overall survival[26], which was consistent with our research. Thus, our signature displayed decent clinical applicability in predicting the development of HCC.

We also examined whether our signature is related to TMB. A previous study showed that tumors with high TMB have the potential to produce more neoantigens, which in turn activate the immune system to recognize and clear tumor cells, suggesting that high TMB can enhance tumor sensitivity to ICB therapy[27]. In our signature, the TMB was higher in the high-risk group than in the low-risk group, which not only indicated that our signature has good predictive power but can also stratify patients with different susceptibilities to immunotherapy. In addition, we found that as the risk score increased, the infiltration abundances of immune cells (B cells, CD4 + T cells, CD8 + T cells dendritic cells, neutrophils and macrophages) in HCC tissues also increased. Macrophages were the most statistically significant and most positively correlated with the risk score. A prior report demonstrated that tumor-associated macrophages (TAMs) serve as the main immune cells in the TME, and they could improve the cell proliferation, invasion, and metastatic ability of HCC by secreting multiple cytokines and inducing the occurrence of epithelial–mesenchymal transition (EMT) [28]. Culturing TAMs together with HCC cells can significantly enhance the stem cell properties of HCC as well as EMT progression, which promotes the progression of HCC[29]. We also found that patients in the high-risk group had more CD4 + and CD8 + T-cell infiltration than those in the low-risk group. Further research indicated that high-risk group patients tended to have higher expression levels of PD1, PD-L1 and CTLA-4. This finding suggests that despite the high infiltration abundance of T cells in the HCC microenvironment, they are unable to function normally due to inhibition by the PD-1- or CTLA-4-mediated suppression pathways, which leads to immune evasion

of the tumor. It was also reported that tumors with pre-existing intratumor T cells inhibited by the PD-1/PD-L1 signaling pathway are most likely to benefit from ICB[30]. Hence, our results revealed that the high-risk group of patients may be associated with a better response to ICB targeting PD-1 and CTLA-4, which enables the personalized treatment of HCC patients. Finally, we conducted functional enrichment analysis on the 37 IRGs that significantly co-expressed with the risk genes in the signature and obtained intensive immune-related phenotypes. This further proved that our signature was closely linked to the immune system.

Our study was the first to systematically establish and validate a multigene immune-related risk signature with independent prognostic power for HCC patients. We utilized multiple algorithms (including univariate Cox, LASSO, and multivariate Cox regression) to identify key prognostic IRGs and construct signatures based on large cohorts from the TCGA and ICGC databases. More importantly, we analyzed TMB, immune cell infiltration, immune checkpoints, and immune-related functions, which reflected the immune response intensity in the HCC microenvironment. Our signature may be helpful for clinical management and personalized immunotherapy decisions. Nevertheless, our study had several limitations. First, we did not perform a prospective clinical trial to validate the signature. Moreover, further investigation with in vivo and in vitro experiments should be conducted to explore the molecular mechanisms of the 8 IRGs in our signature during the progression of HCC.

## Conclusions

Our study developed a robust immune-related risk signature and built a predictive nomogram incorporating the signature and clinical factors to predict overall survival of HCC. The signature was not only closely related to the invasion, progression and prognosis of HCC, but also has a strong correlation with the local immune status. The nomogram reliably predicted overall survival of HCC, which may facilitate clinical management and making of medical decisions.

## Declarations

### Acknowledgements

Not applicable.

### Authors' contributions

Z. L. and X. H. designed the research. Z. L., D. J., X. Z. and Y. Y. performed data acquisition and data analysis. Y. Y., J. C. and J. L. assisted with data acquisition and data analysis. Z. L. and D. J. wrote the manuscript. All authors read and approved the

final version of the manuscript.

### Funding

This research was supported by the Key Scientific Research Projects of Higher Education Institutions in Henan Province (20A320024) and the Provincial and ministerial youth projects of Henan Medical Science and Technology Public Relations Program (SB201902015).

### **Availability of data and materials**

Publicly available datasets were analyzed in this study. These data can be found here: TCGA: <https://portal.gdc.cancer.gov/>; ICGC: <https://icgc.org/>; ImmPort: <https://immport.niaid.nih.gov/>.

### **Ethics approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable.

### **Competing interests**

The authors declare that they have no competing interests.

### **Supplementary information**

**Additional file 1: Table S1** | The 1534 IRGs and their immune category.

**Additional file 2: Figure S1** | The 33 IRGs associated with overall survival using univariate Cox regression model with  $p < 0.05$ .

**Additional file 3: Table S2** | The 37 IRGs significantly co-expressed with the risk genes in the signature ( $|\text{Pearson correlation coefficient}| > 0.5$  and  $P < 0.01$ ).

**Additional file 4: Table S3** | The 444 enriched biological processes.

**Additional file 5: Table S4** | The 116 enriched KEGG pathways.

## **Abbreviations**

IRGs: immune-related genes; HCC: hepatocellular carcinoma; DEMs: differentially expressed mRNAs; DEIRGs: differentially expressed IRGs; TMB: tumor mutation burden; TME: tumor microenvironment; TCGA: The Cancer Genome Atlas; ImmPort: Immunology Database and Analysis Portal; ICGC: International Cancer Genome Consortium; FPKM: fragments per kilobase of exon model per million mapped reads; DEA: Differential expression analysis; FDR: false discovery rate; FC: fold change; LASSO: least absolute shrinkage and selection operator; HR: hazard ratio; AUC: area under the curve; ROC: receiver operating characteristic; BMI: body mass index; AFP: alpha-fetoprotein; BCLC: Barcelona Clinic Liver Cancer; C-index: Harrell's concordance index; TIMER: tumor immune estimation resource; GO: Gene

Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; TAMs: tumor-associated macrophages; EMT: epithelial–mesenchymal transition.

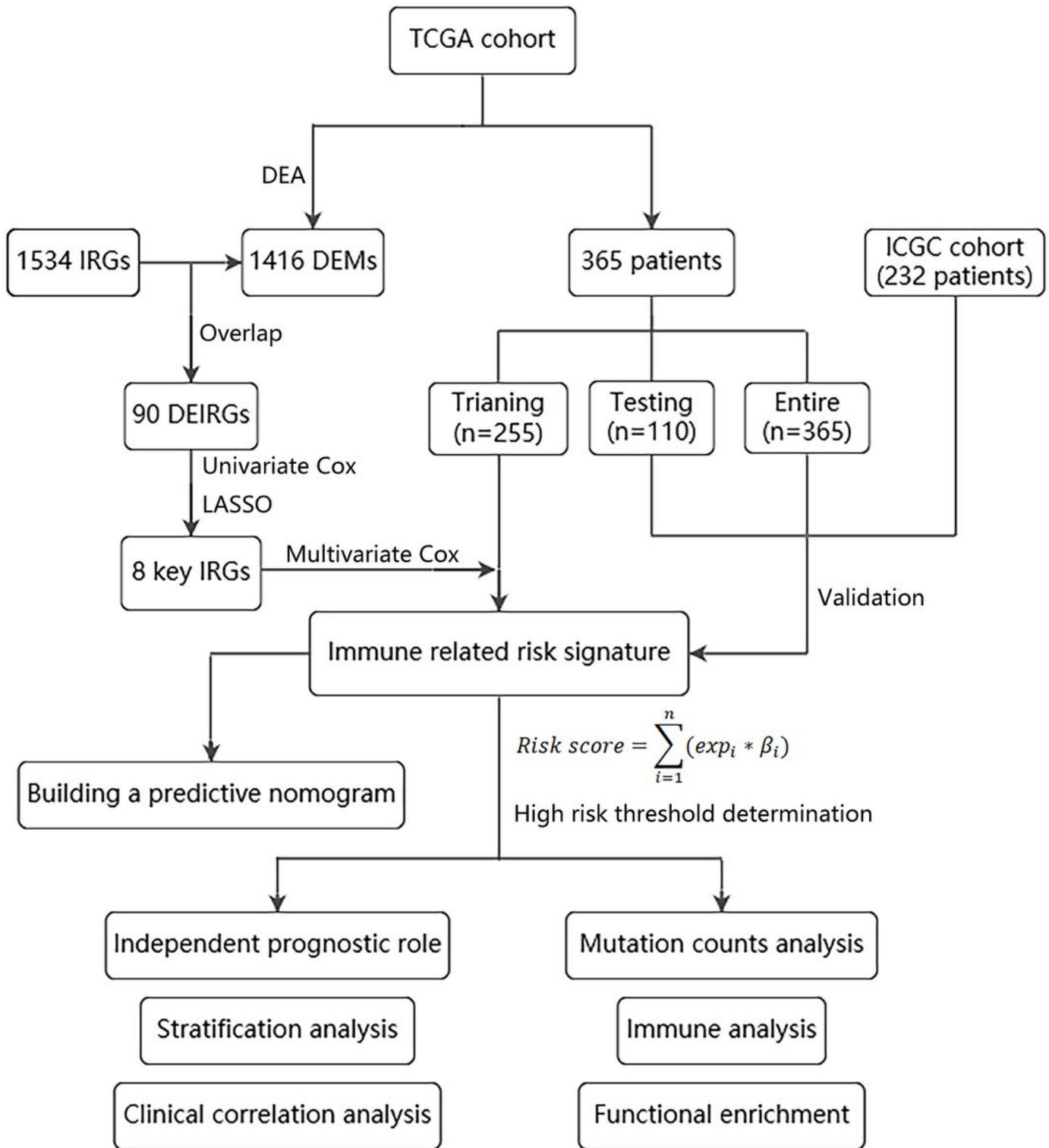
## References

1. Duran SR, Jaquiss RDB: Hepatocellular Carcinoma. *N Engl J Med* 2019, 381(1):e2.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018, 68(6):394-424.
3. Vineis P, Wild CP: Global cancer patterns: causes and prevention. *Lancet* 2014, 383(9916):549-557.
4. Zhang QM, He Y, Luo N, Patel SJ, Han YJ, Gao RR, Modak M, Carotta S, Haslinger C, Kind D *et al*: Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. *Cell* 2019, 179(4):829+.
5. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S *et al*: Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol* 2016, 17(1):174.
6. El-Khoueiry AB, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, Kim TY, Choo SP, Trojan J, Welling TH *et al*: Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet* 2017, 389(10088):2492-2502.
7. Zhu AX, Finn RS, Edeline J, Cattan S, Ogasawara S, Palmer D, Verslype C, Zagonel V, Fartoux L, Vogel A *et al*: Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial. *Lancet Oncol* 2018, 19(7):940-952.
8. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, Albright A, Cheng JD, Kang SP, Shankaran V *et al*: IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* 2017, 127(8):2930-2940.
9. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan JD, Wong P, Ho TS *et al*: Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 2015, 348(6230):124-128.
10. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD *et al*: The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* 2015, 21(8):938-945.
11. Angell H, Galon J: From the immune contexture to the Immunoscore: the role of prognostic and predictive immune markers in cancer. *Curr Opin Immunol* 2013, 25(2):261-267.
12. Sayour EJ, Mitchell DA: Manipulation of Innate and Adaptive Immunity through Cancer Vaccines. *Journal of Immunology Research* 2017.

13. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N: Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 2015, 160(1-2):48-61.
14. Friedrich M, Jasinski-Bergner S, Lazaridou MF, Subbarayan K, Massa C, Tretbar S, Mueller A, Handke D, Biehl K, Bukur J *et al*: Tumor-induced escape mechanisms and their association with resistance to checkpoint inhibitor therapy. *Cancer Immunol Immun* 2019, 68(10):1689-1700.
15. Schreiber RD, Old LJ, Smyth MJ: Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science* 2011, 331(6024):1565-1570.
16. Mantovani A, Savino B, Locati M, Zammataro L, Allavena P, Bonecchi R: The chemokine system in cancer biology and therapy. *Cytokine Growth Factor Rev* 2010, 21(1):27-39.
17. Lippitz BE: Cytokine patterns in patients with cancer: a systematic review. *Lancet Oncol* 2013, 14(6):e218-228.
18. Atretkhany KN, Drutskaya MS, Nedospasov SA, Grivennikov SI, Kuprash DV: Chemokines, cytokines and exosomes help tumors to shape inflammatory microenvironment. *Pharmacol Ther* 2016, 168:98-112.
19. Malumbres M, Barbacid M: Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer* 2009, 9(3):153-166.
20. Yuan P, Meng L, Wang N: SOX12 upregulation is associated with metastasis of hepatocellular carcinoma and increases CDK4 and IGF2BP1 expression. *Eur Rev Med Pharmacol* 2017, 21(17):3821-3826.
21. Ingham M, Schwartz GK: Cell-Cycle Therapeutics Come of Age. *J Clin Oncol* 2017, 35(25):2949-+.
22. Lander GC, Estrin E, Matyskiela ME, Bashore C, Nogales E, Martin A: Complete subunit architecture of the proteasome regulatory particle. *Nature* 2012, 482(7384):186-U175.
23. Li YH, Huang J, Zeng BL, Yang DJ, Sun JZ, Yin XD, Lu MQ, Qiu Z, Peng WY, Xiang TX *et al*: PSMD2 regulates breast cancer cell proliferation and cell cycle progression by modulating p21 and p27 proteasomal degradation. *Cancer Lett* 2018, 430:109-122.
24. Matsuyama Y, Suzuki M, Arima C, Huang QM, Tomida S, Takeuchi T, Sugiyama R, Itoh Y, Yatabe Y, Goto H *et al*: Proteasomal Non-catalytic Subunit PSMD2 as a Potential Therapeutic Target in Association With Various Clinicopathologic Features in Lung Adenocarcinomas. *Molecular Carcinogenesis* 2011, 50(4):301-309.
25. Tan YJ, Jin Y, Wu X, Ren ZQ: PSMD1 and PSMD2 regulate HepG2 cell proliferation and apoptosis via modulating cellular lipid droplet metabolism. *Bmc Mol Biol* 2019, 20(1).
26. Carone C, Olivani A, Valle RD, Manuguerra R, Silini EM, Trenti T, Missale G, Cariani E: Immune Gene Expression Profile in Hepatocellular Carcinoma and Surrounding Tissue Predicts Time to Tumor Recurrence. *Liver Cancer* 2018, 7(3):277-+.
27. Schumacher TN, Schreiber RD: Neoantigens in cancer immunotherapy. *Science* 2015, 348(6230):69-74.

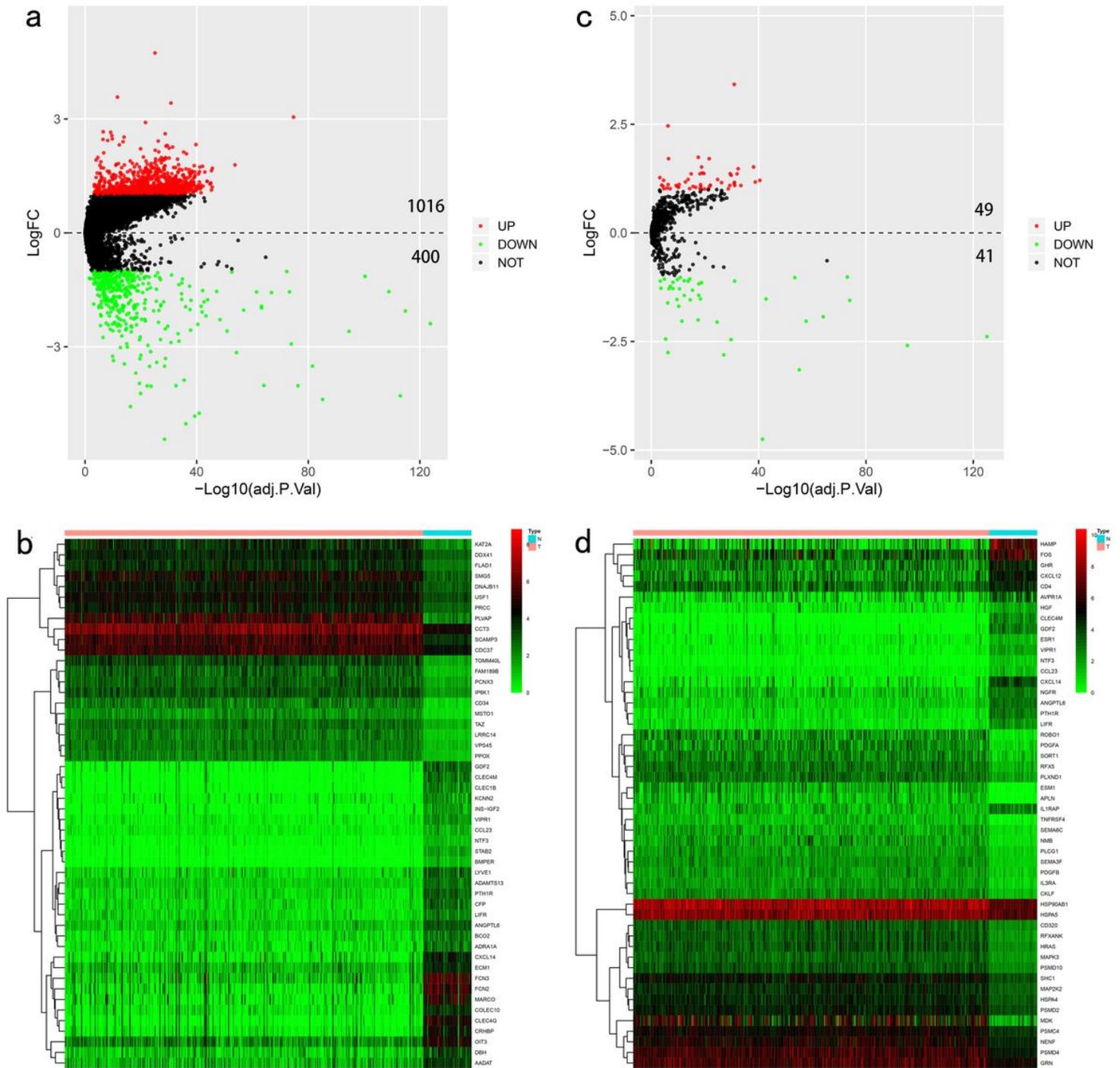
28. Xiao P, Long XX, Zhang LJ, Ye YN, Guo JC, Liu PP, Zhang R, Ning JY, Yu WW, Wei F *et al*: Neurotensin/IL-8 pathway orchestrates local inflammatory response and tumor invasion by inducing M2 polarization of Tumor-Associated macrophages and epithelial-mesenchymal transition of hepatocellular carcinoma cells. *Oncoimmunology* 2018, 7(7).
29. Zhu FY, Li XN, Chen SY, Zeng Q, Zhao Y, Luo F: Tumor-associated macrophage or chemokine ligand CCL17 positively regulates the tumorigenesis of hepatocellular carcinoma. *Med Oncol* 2016, 33(2).
30. Chen YP, Zhang Y, Lv JW, Li YQ, Wang YQ, He QM, Yang XJ, Sun Y, Mao YP, Yun JP *et al*: Genomic Analysis of Tumor Microenvironment Immune Types across 14 Solid Cancer Types: Immunotherapeutic Implications. *Theranostics* 2017, 7(14):3585-3594.

## Figures



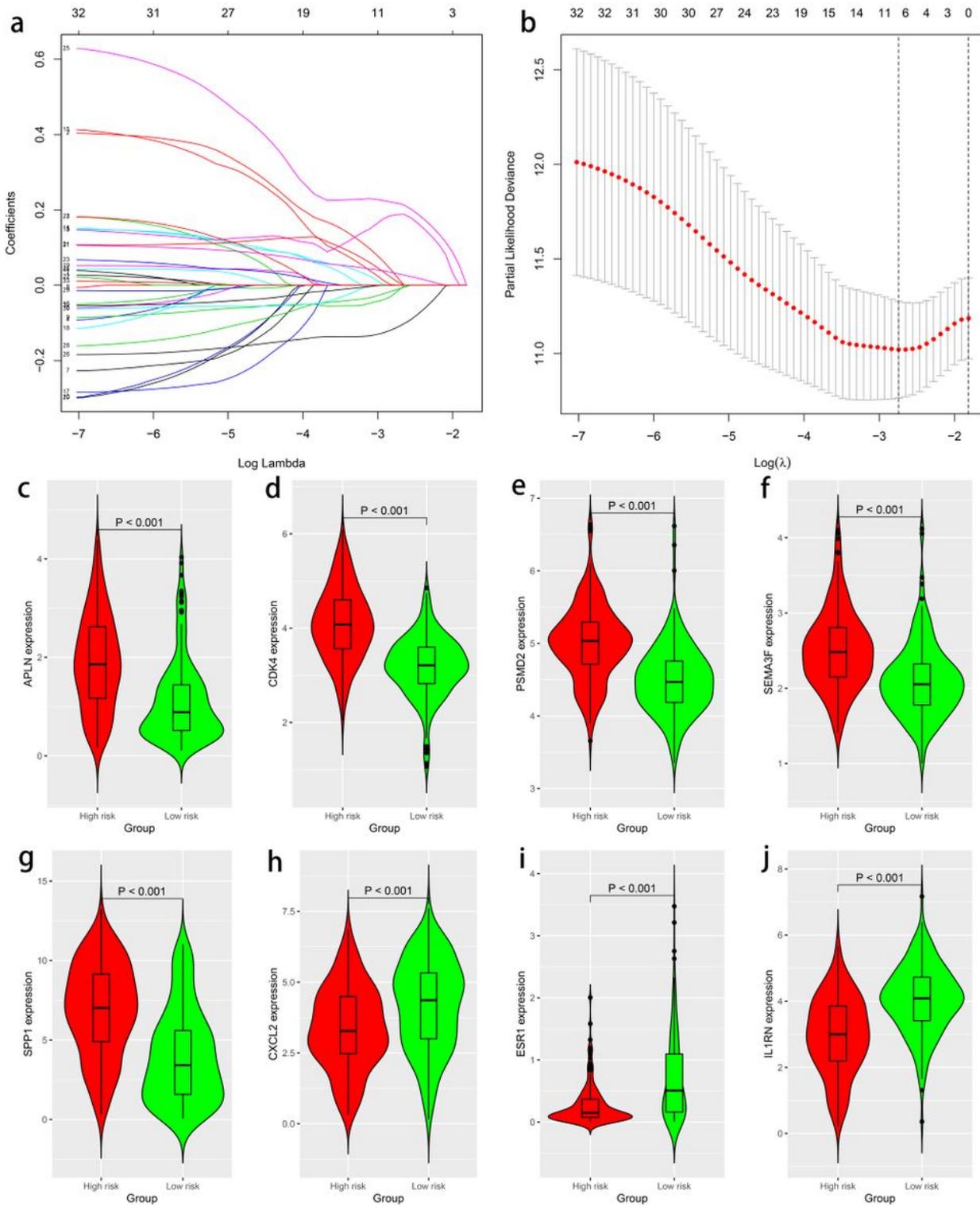
**Figure 1**

Study flow diagram. DEA, differential expression analysis. DEMs, differentially expressed mRNAs. IRGs, immune-related genes. DEIRGs, differentially expressed IRGs.



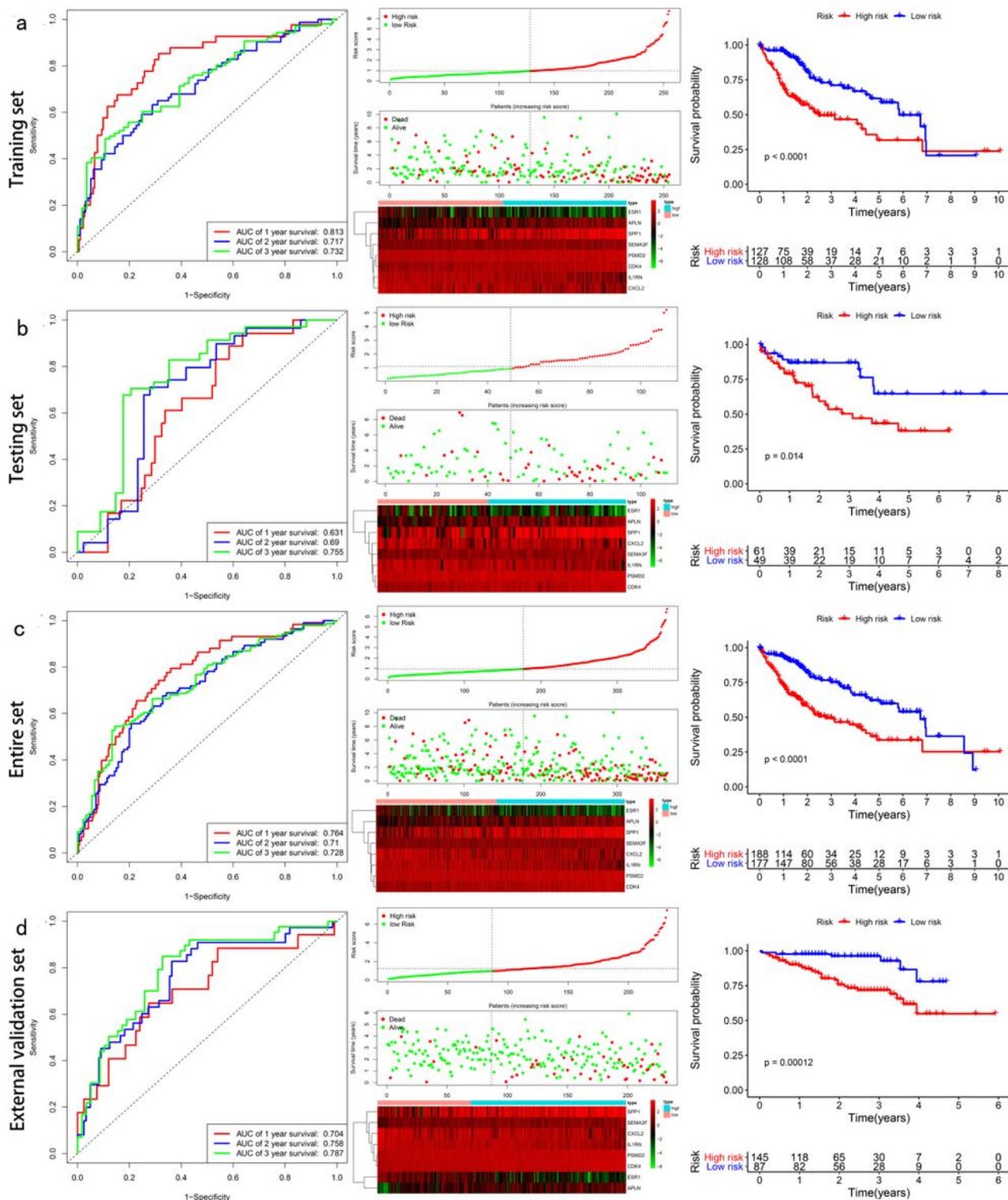
**Figure 2**

Volcano plot and hierarchical clustering analysis of DEMs and DEIRGs in the TCGA dataset. a Volcano plot of DEMs. b Cluster heat map of the top 50 DEMs. c Volcano plot of DEIRGs. d Cluster heat map of the top 50 DEIRGs.



**Figure 3**

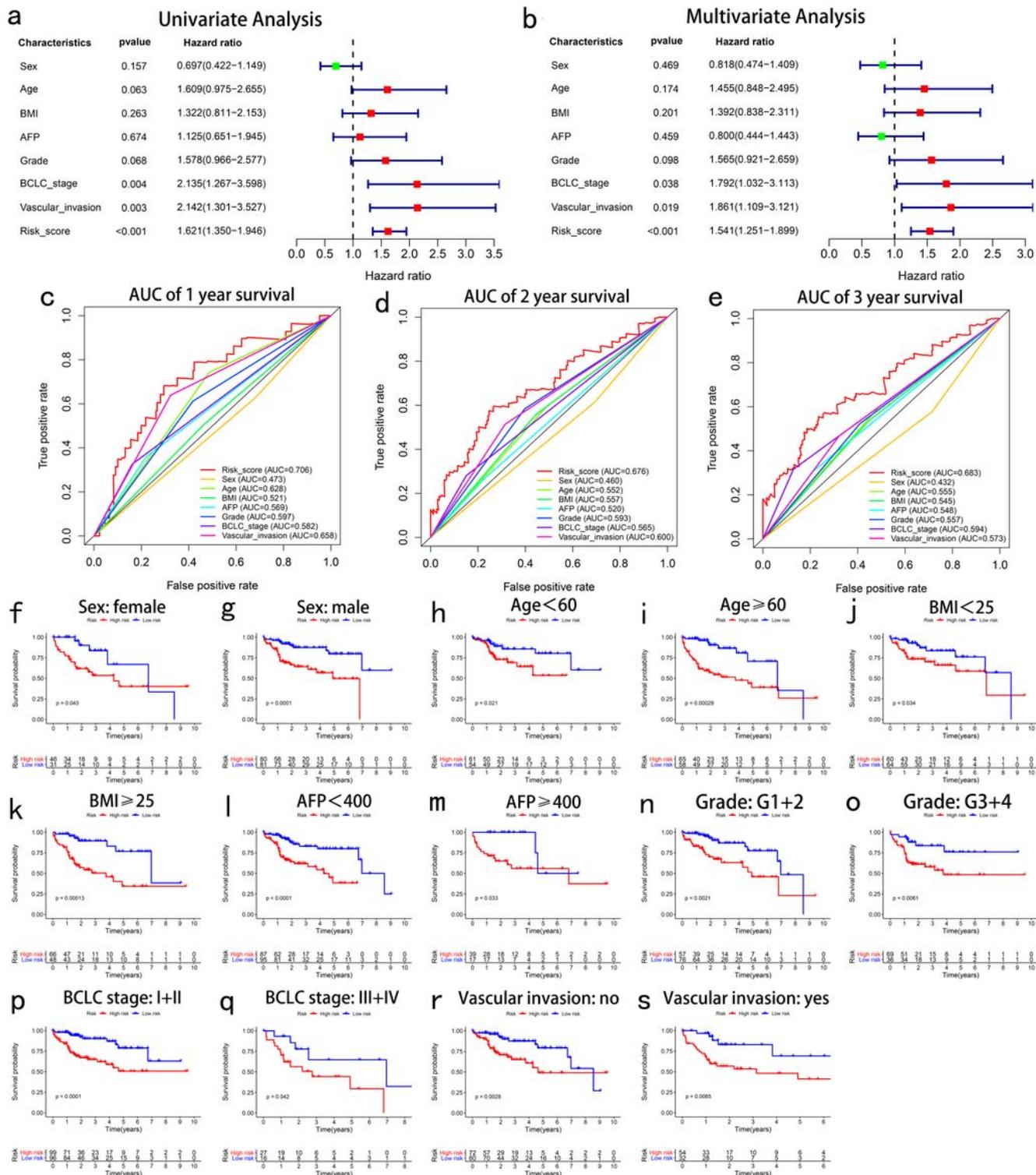
Identification and expression levels of the key prognostic IRGs genes in the risk group. a, b Identification of key prognostic IRGs by LASSO regression analysis. c-j The expression levels of the key prognostic IRGs in the high- and low-risk groups.



**Figure 4**

ROC analysis, risk score analysis, and Kaplan–Meier analysis for the immune-related signature in HCC. a Time-dependent ROC analysis, risk score, heatmap of mRNA expression, and Kaplan–Meier curve of the signature in the training set of TCGA cohort. b Time-dependent ROC analysis, risk score, heatmap of mRNA expression, and Kaplan–Meier curve of the signature in the testing set of TCGA cohort. c Time-dependent ROC analysis, risk score, heatmap of mRNA expression, and Kaplan–Meier curve of the

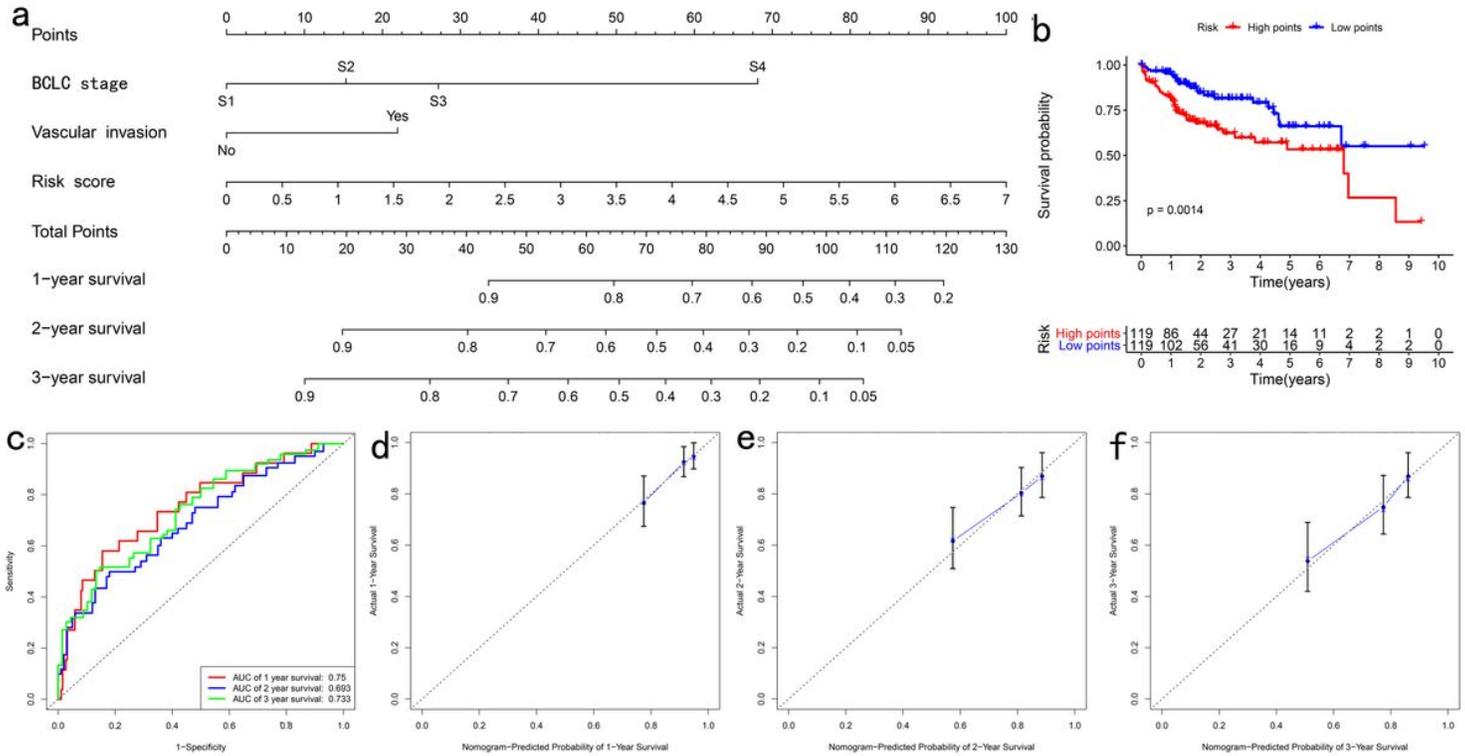
signature in the entire included set of TCGA cohort. d Time-dependent ROC analysis, risk score, heatmap of mRNA expression, and Kaplan–Meier curve of the signature in ICGC cohort.



**Figure 5**

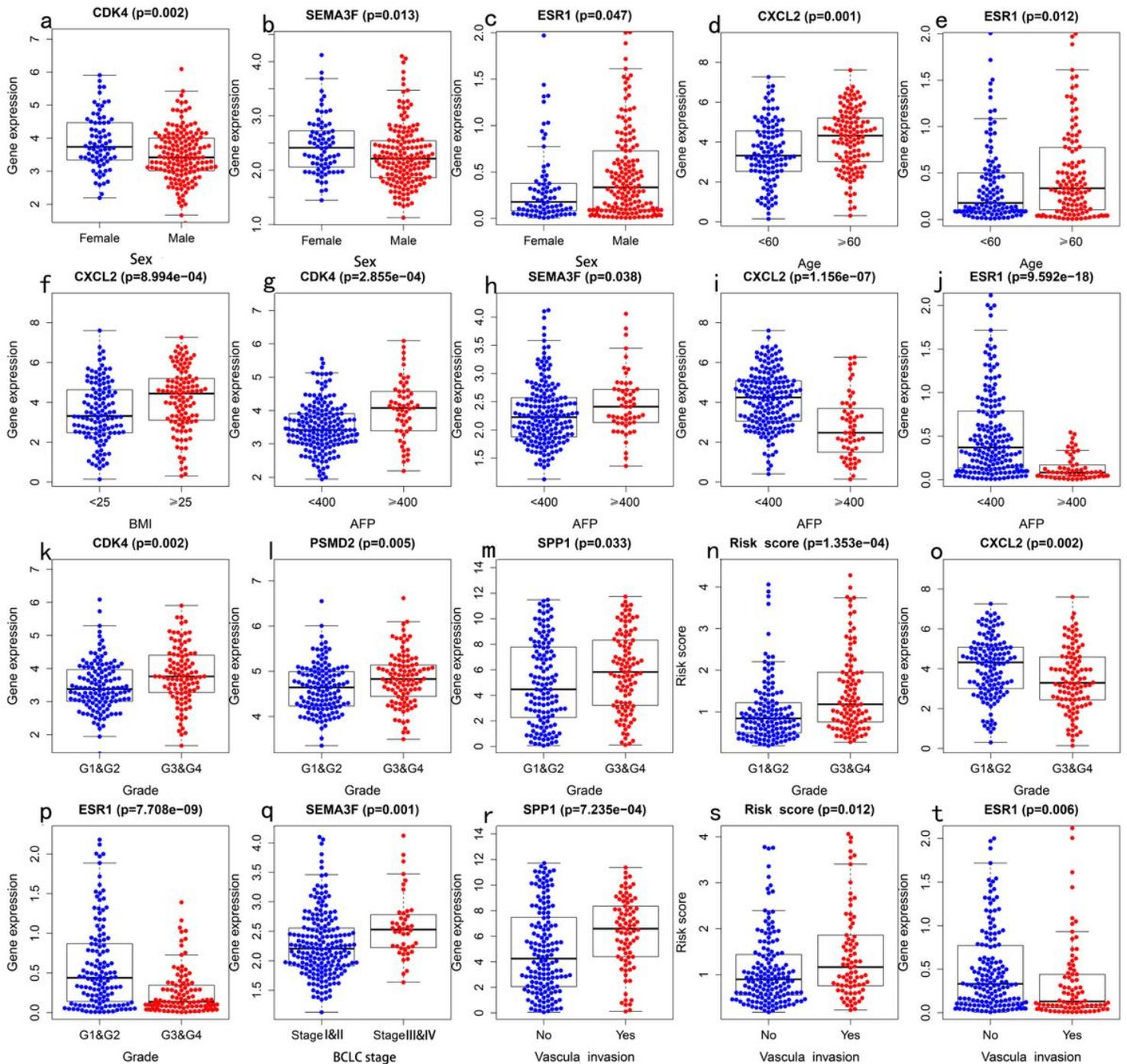
Univariate and multivariate Cox regression, the ROC contrastive and stratification analysis in the TCGA cohort. a, b Forrest plot of the univariate and multivariate Cox regression analysis in HCC. c-e The time-

dependent ROC contrastive analysis between the signature and all clinical factors at 1, 2, and 3 year. f- Kaplan-Meier analysis of overall survival for patients in the 14 subgroups.



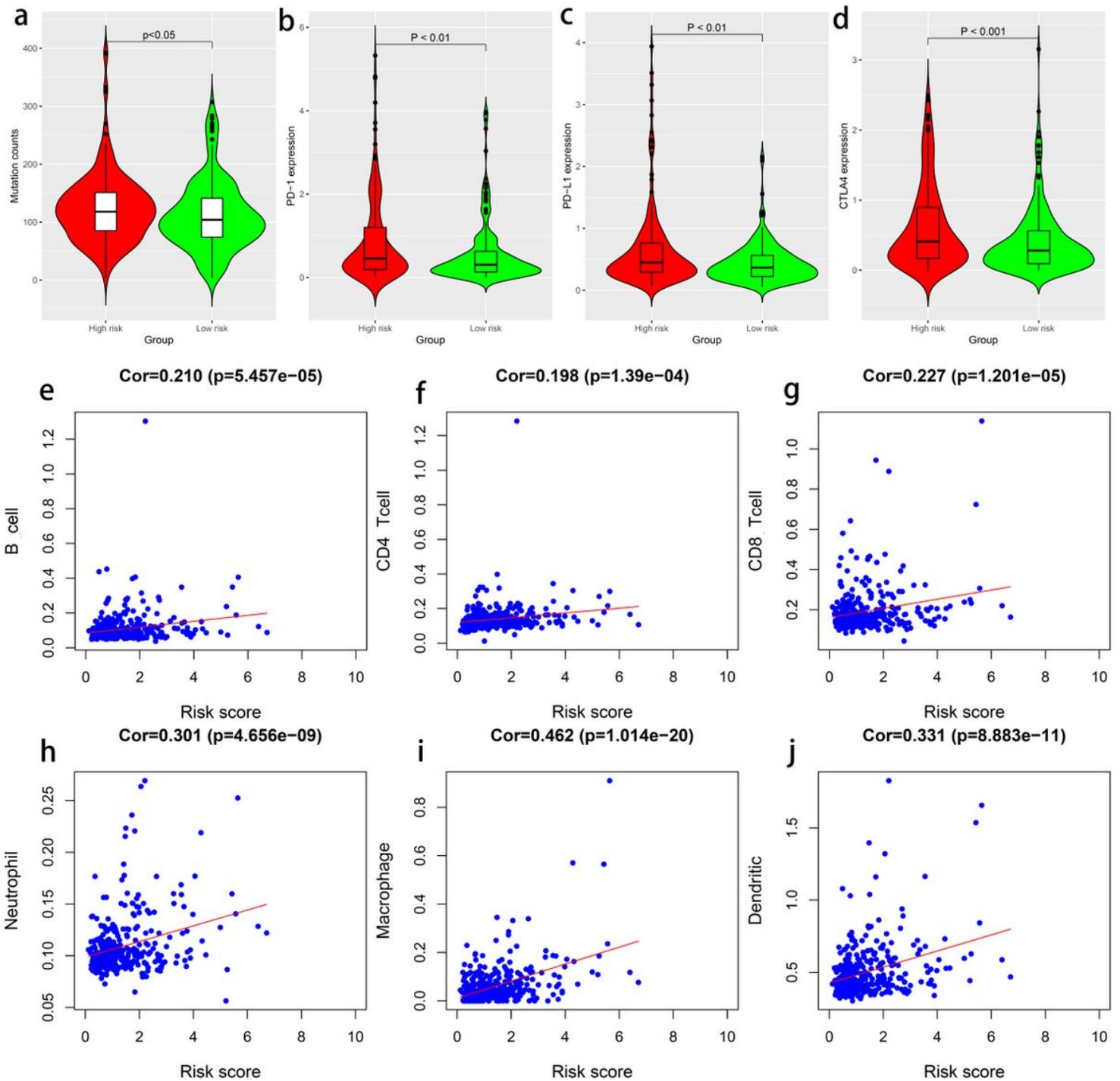
**Figure 6**

Nomogram predicting overall survival for HCC patients. a A prognostic nomogram predicting 1-, 2-, and 3-year overall survival of HCC patients. b Shows the Kaplan-Meier survival curves of the nomogram. c Shows the time-dependent ROC for 1-, 2-, and 3-year overall survival predictions for the nomogram. d-f The calibration plot for internal validation of the nomogram at 1, 2, and 3 year. The Y-axis represents actual survival, and the X-axis represents nomogram-predicted survival.



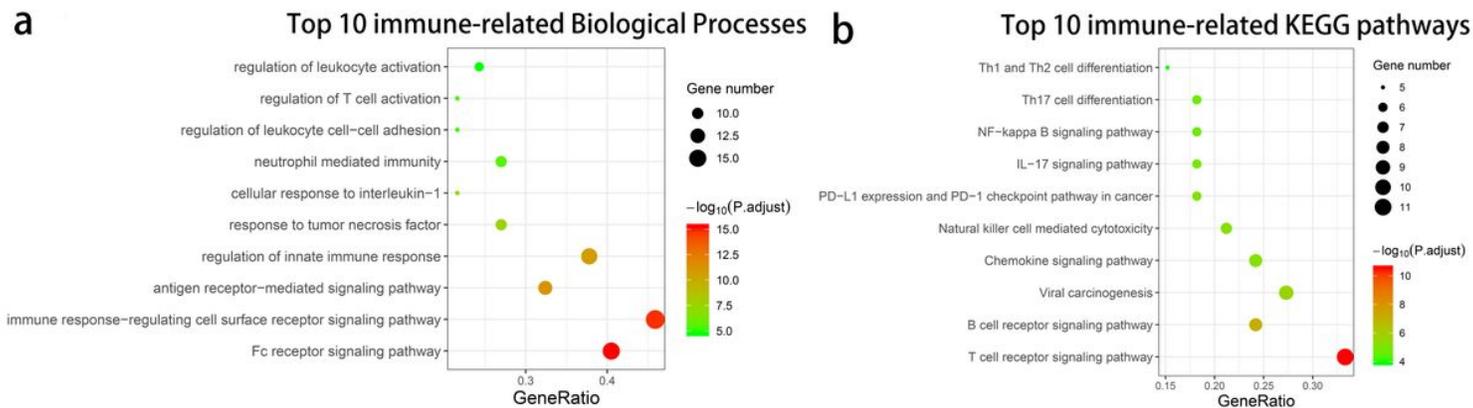
**Figure 7**

Relationships of the risk factors in the signature with the clinical parameters of HCC patients. a CDK4 expression and sex. b SEMA3F expression and sex. c ESR1 expression and sex. d CXCL2 expression and age. e ESR1 expression and age. f CXCL2 expression and BMI. g CDK4 expression and AFP. h SEMA3F expression and AFP. i CXCL2 expression and AFP. j ESR1 expression and AFP. k CDK4 expression and histological grade. l PSMD2 expression and histological grade. m SPP1 expression and histological grade. n Risk score and histological grade. o CXCL2 expression and histological grade. p ESR1 expression and histological grade. q SEMA3F expression and BCLC stage. r SPP1 expression and vascular invasion. s Risk score and vascular invasion. t ESR1 expression and vascular invasion.



**Figure 8**

Mutation and immune analysis of the signature in the TCGA cohort. a Comparison of mutation burden between the high- and low-risk groups. b-d Comparison of PD-1, PD-L1, and CTLA-4 expression between the high- and low-risk groups. e-g Analysis of the correlation between the risk score and immune cell infiltration.



**Figure 9**

Functional enrichment analysis of the 37 IRGs that significantly co-expressed with the risk genes. a Top 10 immune-related biological processes. b Top 10 immune-related KEGG pathways.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.pdf](#)
- [formulas.docx](#)
- [TableS4.xls](#)
- [TableS3.xls](#)
- [TableS2.xls](#)
- [TableS1.xls](#)