

# A prognostic signature established based on genes related to tumor microenvironment for patients with hepatocellular carcinoma

Zhongfeng Cui<sup>1</sup>, Ge Li<sup>1</sup>, Yanbin Shi<sup>2</sup>, Xiaoli Zhao<sup>3</sup>, Juan Wang<sup>3</sup>, Shanlei Hu<sup>3</sup>, Chunguang Chen<sup>1</sup>, Guangming Li<sup>3</sup>

<sup>1</sup>Department of Clinical Laboratory, Henan Provincial Infectious Disease Hospital, Zhengzhou 450000, China

<sup>2</sup>Department of Radiology, Henan Provincial Infectious Disease Hospital, Zhengzhou 450000, China

<sup>3</sup>Department of Infectious Diseases and Hepatology, Henan Provincial Infectious Disease Hospital, Zhengzhou 450000, China

**Correspondence to:** Guangming Li; email: [lgm177@sina.com](mailto:lgm177@sina.com), <https://orcid.org/0000-0003-2286-7679>

**Keywords:** tumor microenvironment, hepatocellular carcinoma, molecular subtypes, risk score model, prognosis

**Received:** November 10, 2023

**Accepted:** March 13, 2024

**Published:** April 4, 2024

**Copyright:** © 2024 Cui et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Background:** Complex cellular signaling network in the tumor microenvironment (TME) could serve as an indicator for the prognostic classification of hepatocellular carcinoma (HCC) patients.

**Methods:** Univariate Cox regression analysis was performed to screen prognosis-related TME-related genes (TRGs), based on which HCC samples were clustered by running non-negative matrix factorization (NMF) algorithm. Furthermore, the correlation between different molecular HCC subtypes and immune cell infiltration level was analyzed. Finally, a risk score (RS) model was established by LASSO and Cox regression analyses (CRA) using these TRGs. Functional enrichment analysis was performed using gene set enrichment analysis (GSEA).

**Results:** HCC patients were divided into three molecular subtypes (C1, C2, and C3) based on 704 prognosis-related TRGs. HCC subtype C1 had significantly better OS than C2 and C3. We selected 13 TRGs to construct the RS model. Univariate and multivariate CRA showed that the RS could independently predict patients' prognosis. A nomogram integrating the RS and clinicopathologic features of the patients was further created. We also validated the reliability of the model according to the area under the receiver operating characteristic (ROC) curve value, concordance index (C-index), and decision curve analysis. The current findings demonstrated that the RS was significantly correlated with CD8+ T cells, monocytic lineage, and myeloid dendritic cells.

**Conclusion:** This study provided TRGs to help classify patients with HCC and predict their prognoses, contributing to personalized treatments for patients with HCC.

## INTRODUCTION

The liver is an immunological organ that harbors abundant immune cells such as dendritic cells (DCs), natural killer (NK) cells, and gamma delta ( $\gamma$   $\delta$ ) T cells, which will trigger immune response to protect the body against adverse pathogens and prevent tumorigenesis [1]. HCC as the most common subtype of liver cancer has the third highest death rate [2]. Unlike normal liver, the immunosuppressive TME of patients

is characterized by the presence of immune cells, for example, macrophages, myeloid-derived suppressor cells, tumor-correlated neutrophils, regulatory T cells, lymphocytes infiltrating tumors, and CD8+ cytotoxic T lymphocytes and tumor vasculature [3, 4]. Previous studies have explored the TME in the biological and prognostic classification of patients with HCC [5] and proposed an RME-based classification for HCC patients according to the mRNA and protein expression detected by multiplex-immunohistochemistry (IHC) and mass

cytometry (CyTOF) [6]. Patients with immunosuppressive TME, including high levels of regulatory T cells (Tregs) and myeloid-derived suppressor cells (MDSCs), will often develop poorer prognosis and benefit limitedly from conventional and immunotherapeutic treatments. Conversely, TME enriched with active immune components such as CD8+ T cells often indicates a favorable prognosis and better responses to certain therapies [7–10]. Therefore, the heterogeneity of TME and patients' immune status should be comprehensively explored in order to improve the clinical treatment of HCC.

TME is a complex network of signaling pathways as it contains a variety of cell types, including cancer, stromal, endothelial, and immune cells, and non-cellular components such as growth factors and cytokines [11]. Currently, targeting the TME as an option to treat patients with HCC is still difficult [12], which also points to the need to characterize the TME of patients with HCC and identify TME-related molecular signatures [5].

This study excavated TME-related genes (TRGs) in patients with HCC and developed a risk score (RS) model using these TRGs. Samples were clustered using TRGs and non-negative matrix factorization (NMF) algorithm. The RS model was established and validated by performing LASSO and Cox regression analysis (CRA). The present study developed a TME-based classification for HCC patients and filtered TME-related markers for prognostic prediction, contributing to the personalized treatment in HCC.

## MATERIALS AND METHODS

### Data acquisition and preprocessing

We collected the transcriptomic data and clinical features of 374 HCC patients and 50 control samples from The Cancer Genome Atlas (TCGA) database by visiting Genomic Data Commons (<https://portal.gdc.cancer.gov/>) [13]. Four patients without complete clinical follow-up and microdissection data were eliminated, while the remaining samples were included in the analyses. Moreover, for external validation, the GSE76427 dataset, which used the GPL10558 platform and included 115 primary HCC tumors and 52 adjacent nontumor tissues, were obtained from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>).

### Screening and analysis of differentially expressed TRGs (DE-TRGs)

We identified 4,061 TRGs from previously published studies [14, 15]. First, “limma” package was employed

to screen DE-TRGs between HCC and adjacent nontumor tissues [16]. The data in fpkm format were preprocessed by normalization and filtering under the criteria of the absolute value of  $FDR < 0.05$  and  $|\log_2\text{Fold Change}| > 1$ .

### NMF clustering

Prognosis-related DE-TRGs were screened using Univariate CRA (UCRA) under  $P < 0.01$  as the threshold. Next, NMF clustering analysis was performed on the expression matrix of prognosis-related DE-TRGs in the “NMF” R package [17]. The correlation coefficients were calculated, the inner feature structure of gene expression matrices was predicted, and 50 iterations were recycled during the clustering.

### Analysis of immune cell infiltration

Microenvironment cell populations-counter (MCP-counter) was applied to score the abundance of immune cells such as CD8+ T cells, NK and endothelial cells, myeloid DCs, neutrophils, cytotoxic and B lymphocytes, monocyte-derived cells, and fibroblasts based on the gene expression [18] using the “MCPcounter” package.

### Developing and validating the RS model

We divided 370 HCC patients from TCGA into the training ( $n = 262$ ) and validation ( $n = 108$ ) sets at the ratio of 7:3 ratio without control (Table 1). Notably, the GSE76427 dataset contained both tumor and adjacent non-tumor samples, only the tumor samples were used for developing a RS model. Next, prognosis-related genes were mined by performing UCRA on DE-TRGs, the model was refined by building a penalized feature using the “glmnet” package [19], and multivariate CRA (MCRA) was used to screen the characteristic genes for the RS model. Finally, we calculated the coefficient value of each gene and the riskscore as follows:

$$\text{Riskscore} = \sum_{i=1}^n (\text{coef}_i \times X_i)$$

*coef* represented the Cox regression coefficient for each gene, and *X* referred to gene expression.

The samples were clustered by their median RS value into the low-risk group (LRG) and high-risk group (HRG). The prognostic significance of the model was analyzed by Kaplan-Meier (KM) analysis with two-sided log-rank test using the “survminer” package. Area under the receiver operating characteristic (ROC) curve (AUC) in “timeROC” [20] package was used to reflect the prediction accuracy of the model.

**Table 1. 370 sample information about training set and verification set of TCGA-LIHC.**

Characteristics	TCGA-LIHC cohort (N = 370)	GSE76427 (N = 115)
<b>Age</b>		
Mean ± SD	59.45 ± 13.51	63.45 ± 12.68
Median (min-max)	61.00 (16.00, 90.00)	64.00 (14.00, 93.00)
<b>Gender</b>		
Female	121 (32.7%)	22 (19.13%)
Male	249 (67.3%)	93 (80.87%)
<b>Grade</b>		
G1	55 (14.86%)	NA
G2	177 (47.84%)	NA
G3	121 (32.7%)	NA
G4	12 (3.24%)	NA
Unknow	5 (1.35%)	NA
<b>Stage</b>		
Stage I	171 (46.22%)	55 (47.83%)
Stage II	85 (22.97%)	35 (30.43%)
Stage III	85 (22.97%)	21 (18.26%)
Stage IV	5 (1.35%)	4 (3.48%)
Unknow	24 (6.49%)	NA
<b>T</b>		
T1	181 (48.92%)	NA
T2	93 (25.14%)	NA
T3	80 (21.62%)	NA
T4	13 (3.51%)	NA
TX	1 (0.27%)	NA
Unknow	2 (0.54%)	NA
<b>M</b>		
M0	345 (93.24%)	NA
M1	4 (1.08%)	NA
MX	21 (5.68%)	NA
<b>N</b>		
N0	329 (88.92%)	NA
N1	4 (1.08%)	NA
NX	37(10%)	NA
<b>Survival status</b>		
Dead	240 (64.86%)	92 (80.00%)
Alive	130 (35.14%)	23 (20.00%)

Abbreviation: NA: Not Available.

We also created a nomogram by integrating clinicopathological features such as gender, age, TNM stages, grade, clinical stage, and the RS model using the “restricted mean survival (rms)” package. The reliability of the nomogram was tested according to the ROC curve, calibration curve (CC, bootstrap 1000), decision curve analysis (DCA), and C-index.

#### Analysis of biological functions

Annotated gene sets ‘c2.cp.kegg.v7.4’ and ‘c5.go.v7.4.symbols’ obtained from the Molecular Signatures Database (MSigDB, <http://www.gsea-msigdb.org/gsea/msigdb>) were used for GSEA (software version 4.1.0).

## Statistical analysis

Data were analyzed using the R (version 4.1.1) package. Volcano maps and violin plots were visualized by the “ggplot2” package [21] and “ggpubr” package, respectively. CRA was performed using the “survival” package. By employing the chi-squared or Fisher’s exact test, differences between both sets and the correlation between clinical characteristics and the RS were analyzed. KM survival curves with the log-rank tests were generated using the “survminer” package. The ROC and calibration curves were plotted using the “timeROC” and “rms” packages, respectively [22]. The restricted mean survival (RMS) package was used for computing the C-index of the model. A  $p < 0.05$  indicated a significant difference.

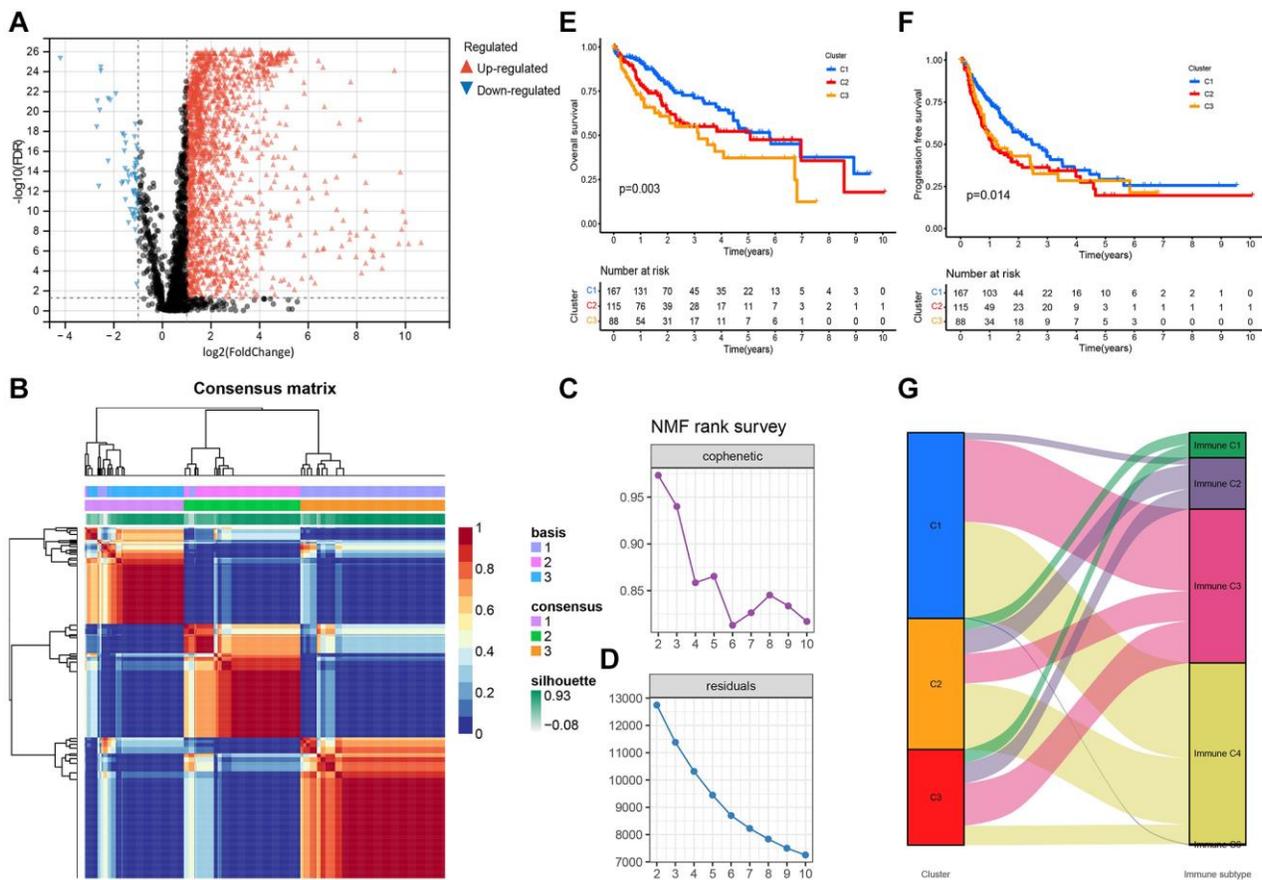
## Data availability statement

The dataset analyzed in this study is available in (GSE76427) at (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76427>).

## RESULTS

### NMF consensus clustering analysis classified three molecular HCC subtypes based on TRGs

A total of 1,717 TRGs were defined as DE-TRGs from 4061 TRGs between 370 (FDR  $< 0.05$  and  $|\log_2FC| > 1$ , Figure 1A), and UCRA identified 704 prognosis-related TRGs. NMF consensus clustering analysis was performed based on the expression profiles of 704 prognosis-related TRGs. The cophenetic correlation coefficient was significantly reduced at  $k = 3$ , which was therefore selected as the optimal number of clusters (Figure 1B–1D). Survival analysis showed that the OS and progression-free survival (PFS) of patients in C2 and C3 was significantly worse compared to C1 (Figure 1E, 1F). Moreover, immune scores of myeloid dendritic cells, B lineage, endothelial cells, cytotoxic lymphocytes, neutrophils, fibroblasts, monocytic lineage, NK cells, and CD8+ T cells were different across C1, C2, and C3 (Figure 2A–2I).



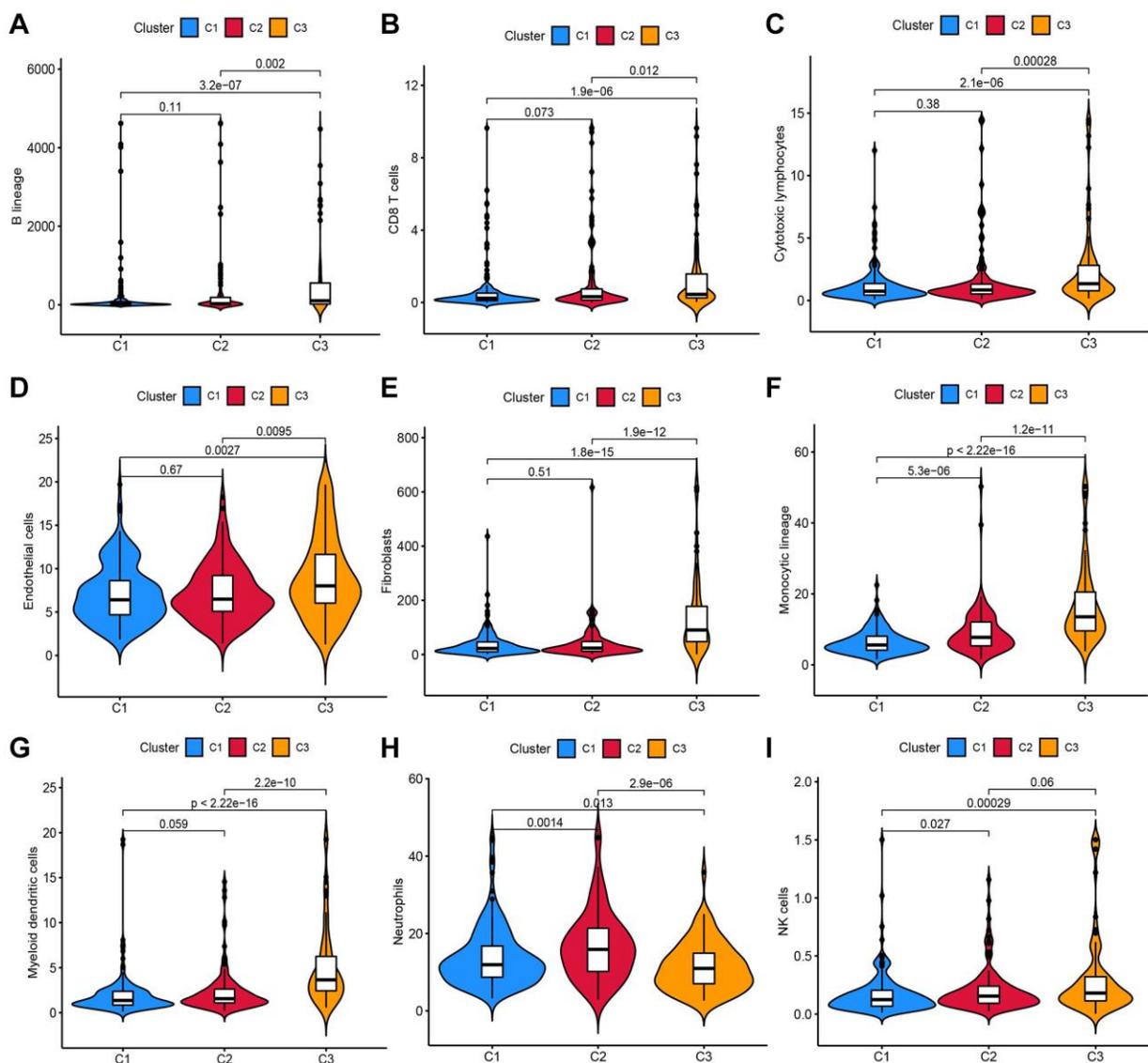
**Figure 1. Differential expression and NMF consensus clustering analysis of TRGs.** (A) Volcano plot shows DE-TRGs among 374 patients with HCC and 50 control samples. (B) NMF consensus clustering for the  $k = 3$ . (C) The cophenetic correlation coefficient indicates the stability of the cluster obtained using the NMF algorithm. (D) RSS reflects the performance of the model in clustering. (E) The KM survival curves show the OS and (F) the PFS of patients in the three subtypes. (G) The alluvial map shows the distribution of these subtypes in immune subtypes.

Wound healing (Immune C1), IFN-gamma dominant (Immune C2), inflammatory (Immune C3), lymphocyte depleted (Immune C4), C5 (Immunologically quiet), and TGF-beta dominant (Immune C6) are the six immune subtypes of solid tumors [23]. Hence, the distribution of these immune subtypes in three molecular clusters was analyzed. Specifically, Immune C3 and C4 were the primary immune subtypes in C1, the proportions of Immune C4 and C3 were the highest in C2 and C3, respectively, and Immune C1 was only detected in C2 and C3 (Figure 1G).

### Constructing and validating the RS model containing 13 DE-TRGs

UCRA and LASSO regression analysis was conducted using the DE-TRGs identified from the training set.

Independent variables gradually increased to zero with lower  $\lambda$  (Figure 3A). In addition, 32 genes were mined using the partial likelihood deviance method (Figure 3B). Finally, the RS model was constructed with the 13 TMGs identified by MCRA as follows:

$$\text{Risk score} = (-0.5974 \times \text{CAPN3 expression level}) + (0.3376 \times \text{HAVCR1 expression level}) + (0.4161 \times \text{GRIN2D expression level}) + (0.6536 \times \text{BMI1 expression level}) + (0.4733 \times \text{SLC30A3 expression level}) + (0.2293 \times \text{GLP1R expression level}) + (0.2211 \times \text{CSAG3 expression level}) + (-0.5401 \times \text{NR4A3 expression level}) + (-0.8607 \times \text{INPP4B expression level}) + (0.1339 \times \text{ETV4 expression level}) + (-0.2406 \times \text{MYH4 expression level}) + (-0.4394 \times \text{ICA1 expression level}) + (0.4485 \times \text{ST6GALNAC4 expression level}).$$


**Figure 2.** Analysis of the infiltration scores of (A) B lineage, (B) CD8+ T cells, (C) cytotoxic lymphocytes, (D) endothelial cells, (E) fibroblasts, (F) monocytic lineage, (G) myeloid DCs, (H) neutrophils and (I) NK cells in patients in these three molecular subtypes.

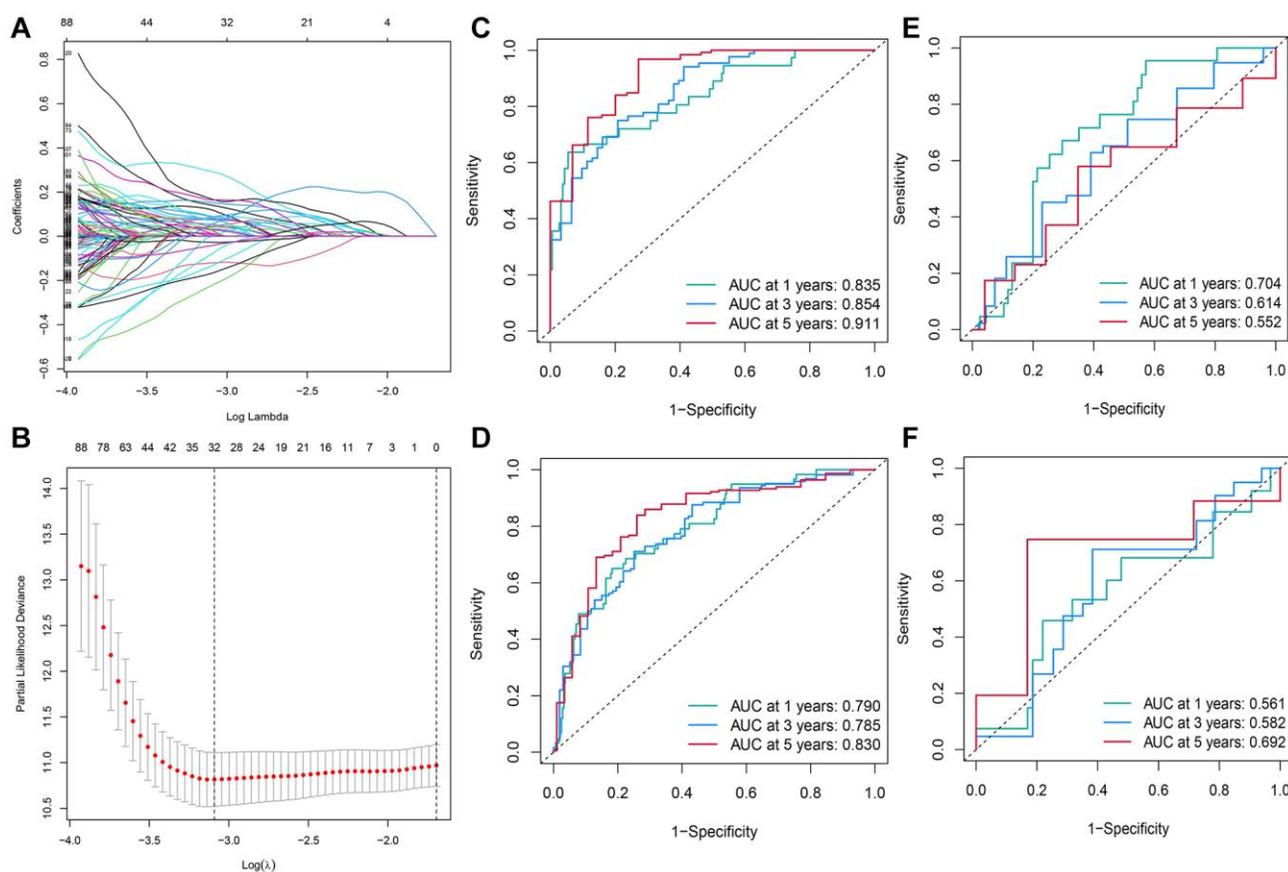
Next, the AUC value was calculated to assess the RS model in all cohorts. The AUC values for 1-, 3-, and 5-year OS of patients in the training set were 0.835, 0.854, and 0.911, respectively (Figure 3C). In the entire TCGA-LIHC cohort, the AUC values for 1, 3, and 5-year OS of patients were 0.790, 0.785, and 0.830, respectively (Figure 3D). In the TCGA validation cohort (the validation set), the AUC values for 1-, 3-, and 5-year OS of patients were 0.704, 0.614, and 0.552, respectively (Figure 3E). Though the AUC values of the 1- (0.561), 3- (0.614), and 5-year (0.692) OS of patients in the GSE76427 cohort were <0.7, the RS model was still accurate to some extent, especially in predicting the 5-year OS. Such a result might be related to the short median duration of follow-up of patients in this cohort (Figure 3F).

Patients from all HCC cohorts were assigned with a RS and divided into HRG and LRG based on the median RS. The KM survival curves showed that in the training set, the prognosis in the LRG was significantly better than in HRG (Figure 4A). The results from survival analyses on patients in the validation set, the entire TCGA, and the GSE76427 cohort were the same

results. In the HRG, the patient's OS was significantly poor than in LRG (Figure 4B–4D). Additionally, in the TCGA-LIHC cohort, the RS varied in patients with different clinicopathological characteristics and it was positively correlated with high pathological grades and advanced T, TNM stages (Figure 4I).

### Construction of a nomogram by integrating RS and clinicopathological features

We performed UCRA and MCRA to determine the correlation between the OS and clinicopathological features such as age, tumor grade, and stage, gender, T and M stage, as well as RS. The results demonstrated that only the RS (HR: 1.026, 95% CI: 1.014–1.037,  $P < 0.001$ ) could independently predict risk (Table 2). A nomogram integrating clinicopathological features and RS was developed (Figure 4E) to predict 1-, 3-, and 5-year survival. For example, the clinical features of a 50-year-old female with LIHC were T2NxM0 stage II, grade 1, and high RS. Then, final total score of the variables was 399. The AUC values of patients' 1-, 3-, and 5-year survival rates were 0.767, 0.561, and 0.322, respectively. The CC of the nomogram showed a high



**Figure 3. Construction of the RS model constituting 13 DE-TRGs.** (A) The LASSO coefficient profile of the DE-TRGs in patients from the training set. (B) The partial likelihood deviance method was used for screening genes. (C–F) Constructing and validating the RS model with the AUC value for 1-, 3- and 5-year OS of patients with HCC in (C) training set; (D) entire TCGA cohort; (E) validation set; (F) GEO cohort.

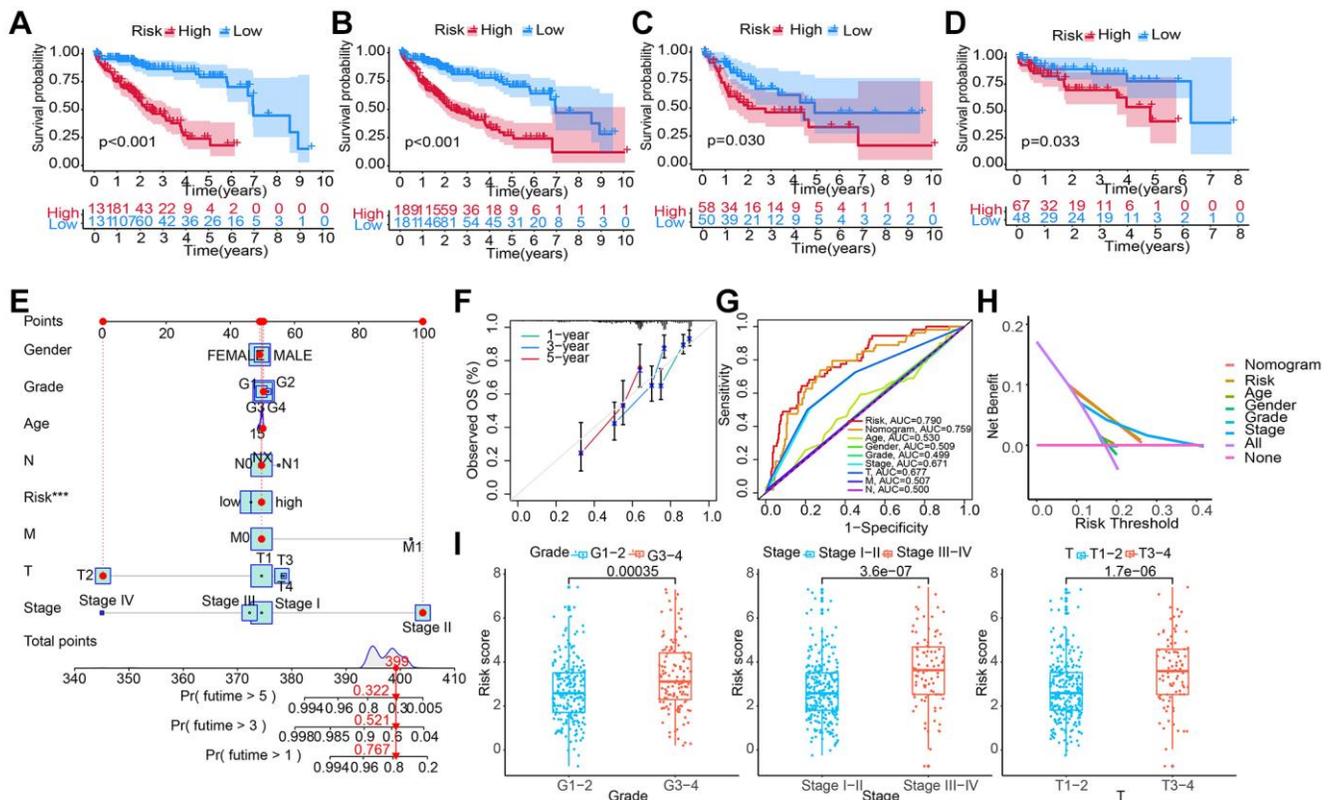
**Table 2. Univariate and multivariate Cox regression for risk score and clinicopathologic features.**

Variables	Univariable analysis			Multivariable analysis		
	HR	95% CI	p-value	HR	95% CI	p-value
Age	1.011	0.996–1.0258	0.147	–	–	–
Gender	0.765	0.524–1.119	0.167	–	–	–
Grade	1.123	0.872–1.445	0.369	–	–	–
Stage	1.675	1.364–2.0567	8.54E-07	0.958	0.430–2.132	0.915
T	1.655	1.361–2.012	4.48E-07	1.693	0.796–3.600	0.171
M	3.78	1.196–11.946	0.024	1.505	0.439–5.153	0.515
RiskScore	1.024	1.013–1.034	6.71E-06	1.026	1.014–1.0367	5.18E-06

consistency between the observed 1-, 3- and 5-year survival rates and the predicted values in the training sets (Figure 4F). In addition, compared to the other clinicopathological characteristics, the AUC value of the RS was higher (Figure 4G). DCA showed that the RS and nomogram could effectively predict the OS compared to all other clinicopathological features (Figure 4H).

**Comparing the TRG-RS model with other models**

We compared our 13-gene signature RS model to other models, including 8-gene signature related to inflammation response (a TME signature) [24], 3 immune-related prognostic genes [14], and 3 tumor doubling time-related immune genes [15]. Firstly, we



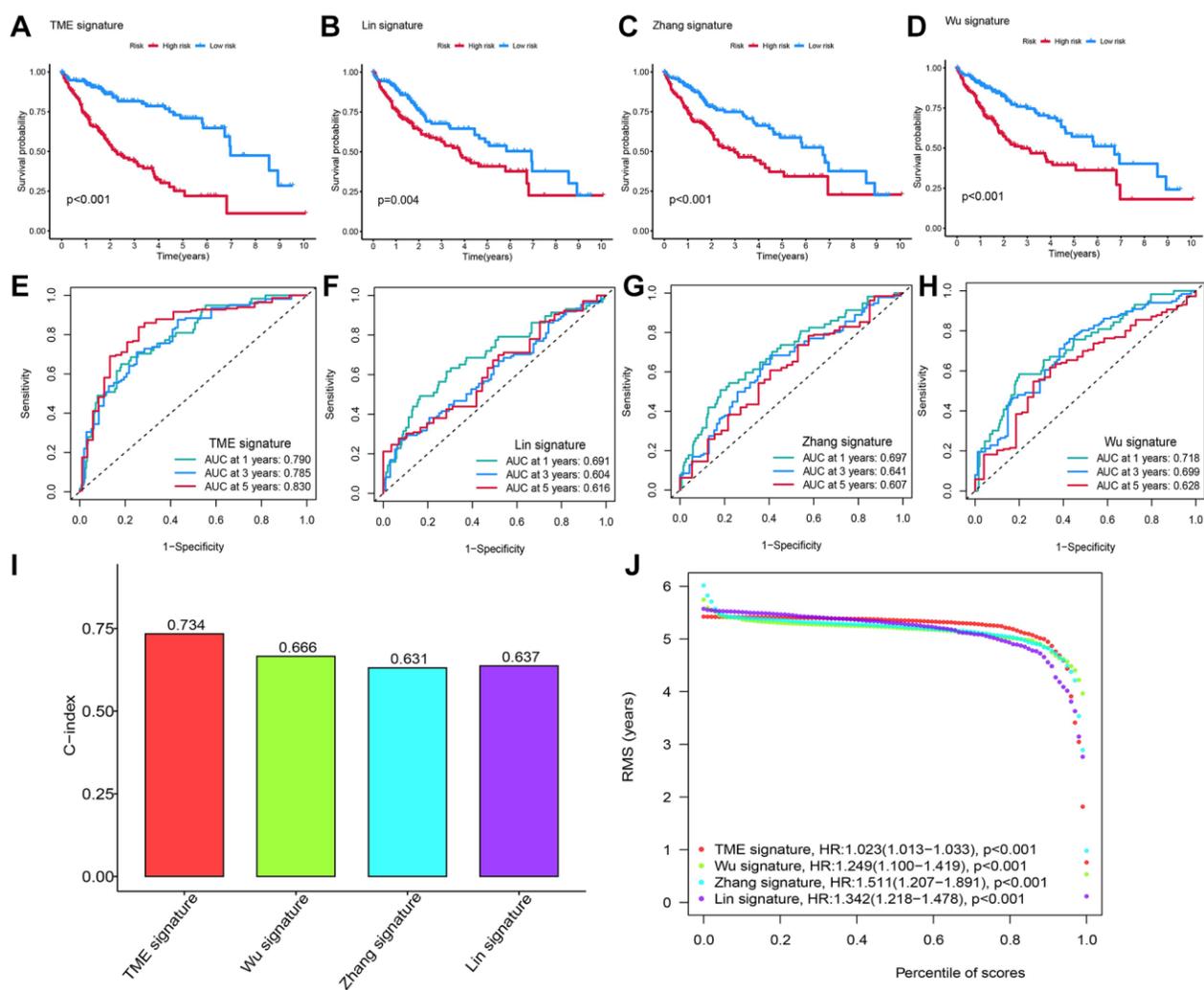
**Figure 4. Constructing and evaluating a nomogram by integrating the RS and clinicopathological features. (A–D)** Construction and validation of the TRG-RS in patients with HCC in (A) training set; (B) entire TCGA cohort; (C) TCGA testing cohort; (D) GEO cohort using the KM survival curves. (E) Nomogram predicting the 1-, 3- and 5-year patient’s OS. The points identified on the point scale of each variable are totaled. Finally, beneath the total points, the probability of 1-, 3- or 5-year survival is projected on the scales below. (F) CC shows the actually observed and nomogram-predicted 1-, 3- and 5-year OS of patients. (G) The ROC curve shows the AUC value of the nomogram, the RS, and clinicopathologic features for predicting the patient’s survival. (H) Evaluating the clinical benefit of the nomogram and the RS using DCA. None indicates that all samples were negative and untreated; therefore, the net benefit is zero. All indicates that all samples were positive and treated. The x-axis shows the threshold probability. (I) TRG-RS comparison of clinical information of patients with HCC in TCGA-LIHC.

calculated the RS for all patients using MCRA and plotted the ROC curve using their corresponding genes. Then, the patients were classified by the median RS value into HRG and LRG. A significant survival difference of patients in the two groups was observed. In addition, the prognosis of patients in LRG was better than in HRG (Figure 5A–5D). As shown by the ROC curve, the AUC value of our model in predicting average 1-year prognosis was the highest. The AUC values of our model for predicting 1-, 3-, and 5-year survival were higher compared to the other three models (Figure 5E–5H). The C-index of our model was 0.734, which was higher than other models (Figure 5I). Finally, we used the RMS time to assess the prediction accuracy of our model at different time points, and observed that our model outperformed the other three models when predicting survival time longer than 60 months. These results indicated that our

model was more effective and accurate in predicting 5-year OS and survival timer longer than 5 years (Figure 5J).

### Functions of our RS model

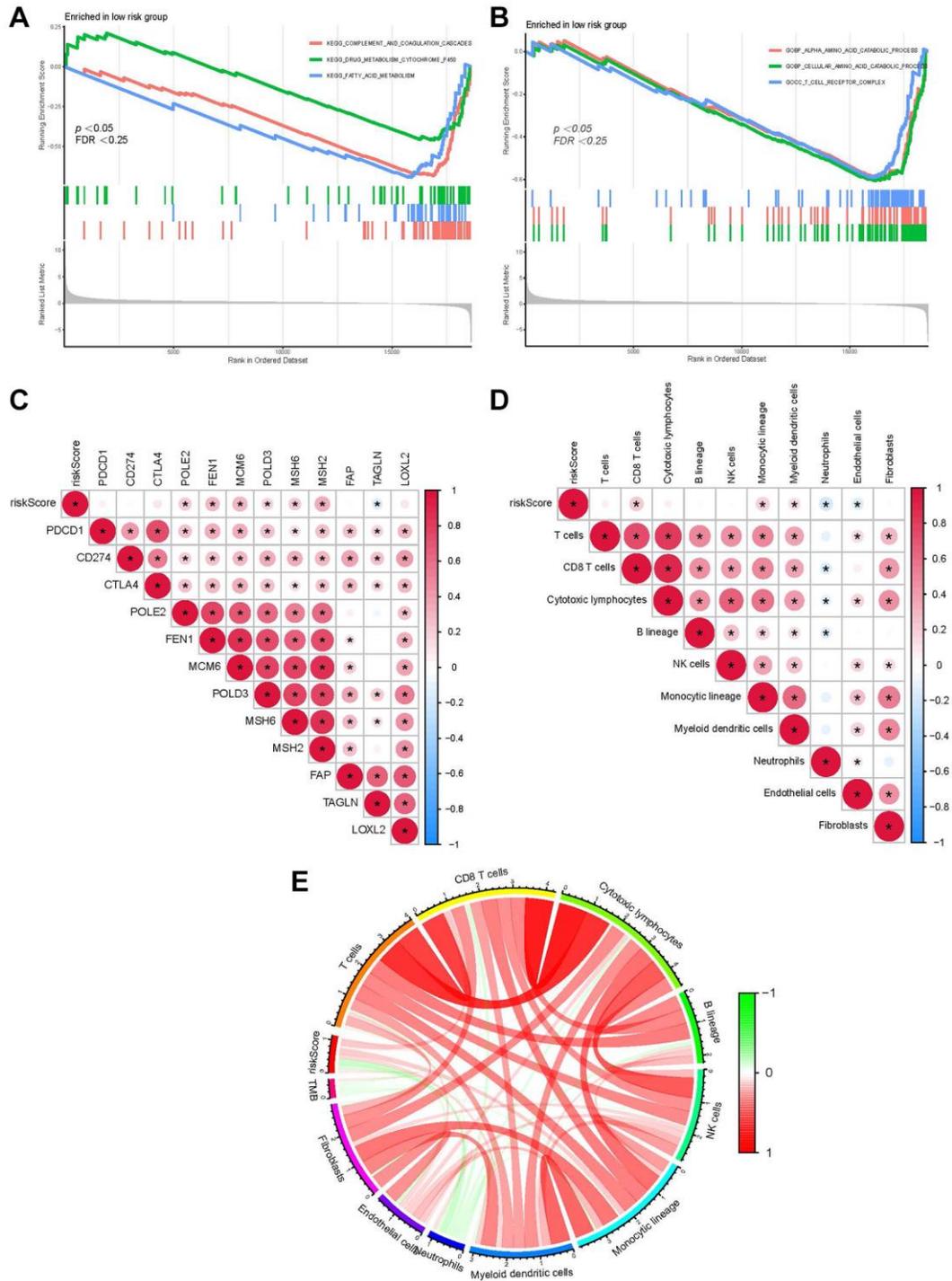
Based on the normalized enrichment score and adjusted *p*-value (*q*-value), GSEA was applied to identify enriched pathways in the two groups. The complement and coagulation pathways, fatty acid and cytochrome p450 drug metabolism was enriched by genes in the LRG (Figure 6A). Moreover, GO terms such as the process of cellular amino acid and alpha-amino acid catabolism and the T cell receptor complex were enriched genes in the LRG (Figure 6B). The correlation between the RS and target genes was analyzed according to the expression of genes related to classical cancer-related pathways including fatty acid



**Figure 5. Comparing the RS model and three previously published models. (A–D)** The KM survival curves of patients were predicted by our RS model and three previously published models. **(E–H)** The ROC curves and **(I)** the C-index of our RS model and three published models and the C-index of our RS model were the highest. **(J)** The RMS time curve of all four prognostic RS models revealed an overlap of 60 months.

metabolism, DNA replication, and cell cycle. DNA replication-related genes (*MCM6*, *MSH2*, and *MSH6*) were positively correlated with the RS (Figure 6C). Additionally, immune cell infiltration in all patients from TCGA-LIHC was analyzed. The RS showed a

positive relationship with CD8+ T cells, monocytic lineage, and myeloid DCs (Figure 6D). The correlation between the RS/tumor mutation burden (TMB) and infiltrating immune cells was analyzed using an MCP counter, and the results are presented in Figure 6E.



**Figure 6. Function of our RS model.** (A) GSEA shows the KEGG pathways enriched in the LRG. (B) GO terms enriched in the LRG. (C) The correlation between the RS of patients with LIHC and the expression of cancer-related pathway genes. (D) The correlation between immune cell score in patients from TCGA-LIHC and the RS. (E) Correlation between the RS/TMB and immune cell score. Among them, \*usually represents a  $p$ -value  $< 0.05$ . Red indicates a positive correlation, blue indicates a negative correlation, and the shade of the color indicates the strength of the correlation.

## Survival prediction for patients with different clinicopathological characteristics using the RS

We predicted the survival of patients with different clinical traits using the RS model. Compared to those in HRG, the KM survival showed better prognostic outcomes of patients with age >65 and ≤65, pathological grade (G1-2 and G3-4), gender, stage I–II, and III–IV in the LRG (Figure 7).

## DISCUSSION

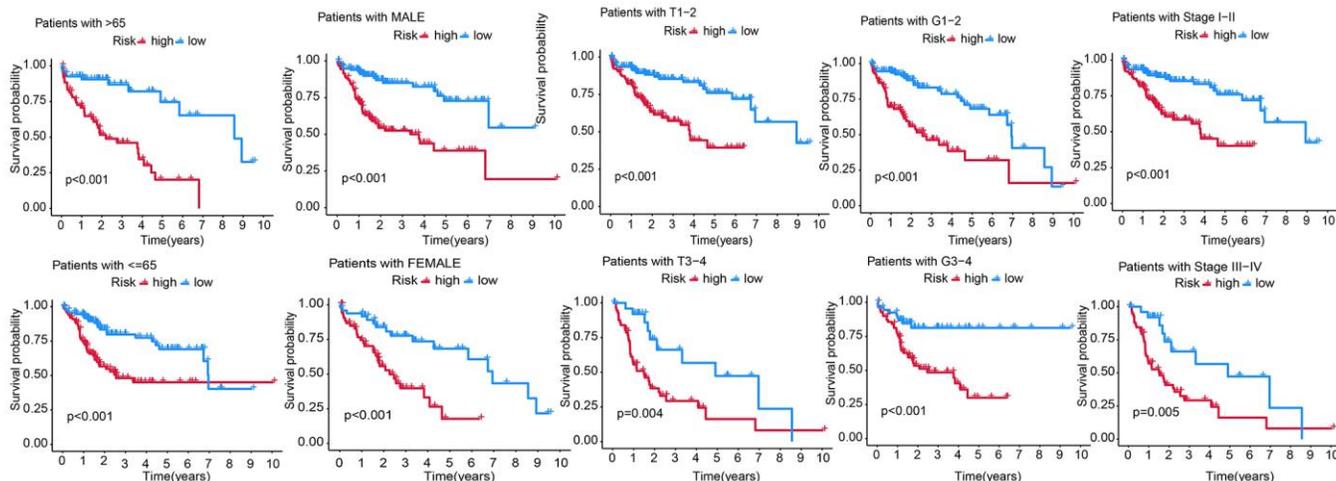
HCC is a typical inflammation-induced cancer. A complex interaction between TMEs is a prerequisite for HCC development [25]. Comprehensive and systematic profiling of these complex interactions in heterogeneous tumor samples could help effectively identify therapeutic agents and biomarkers [26]. This study analyzed TME-related genes to explore the molecular heterogeneity of HCC. Targeting a single molecule involved in the interactions between tumor and TME is unlikely to accurately predict the prognosis of HCC [27]. Hence, we developed an RS model based on multiple TRGs to predict patients' prognosis and determine the biological effects of HCC.

Classifying molecular subtypes using TME-associated genes represents a key advance in understanding the heterogeneity of HCC, and these isoforms could reflect TME characteristics, different stages of tumor progression, immune evasion capacity and response to therapy [28]. Zhang et al. revealed three unique HCC subtypes of immunoactive, immunodeficient, and immunosuppressive TME using CyTOF [29]. In this study, HCC patients were classified into different subtypes based on the transcriptomic characteristics

of TRGs. We identified 4,061 TRGs, of which 704 prognosis-related DE-TRGs were screened by differential expression analysis and UCRA. Based on the expression matrix of these DE-TRGs, we divided patients with HCC into C1, C2, and C3 molecular subtypes. Specifically, patients in subtype C1 exhibited a more favorable prognosis than those in subtypes C2 and C3. We then employed CRA and LASSO regression analysis to construct an RS model based on these 13 TRGs.

The effectiveness of the RS model in predicting the patient's prognosis was analyzed. UCRA and MCRA results revealed that only the RS was an independent risk (not clinicopathological features) for patients with HCC from TCGA-cohort. Furthermore, the reliability of our model was validated according to the CC, AUC value, C-index, and DCA. Furthermore, our RS models outperformed the other three previously published models in predicting HCC prognosis. These results confirmed the reliability of our prognostic model.

We also analyzed the potential pathways enriched by our RS model. Dysregulated pathways, such as fatty acid metabolism [30], DNA replication [31], and cell cycle [32], are involved during tumorigenesis. Correlation analysis between the RS and target genes also showed a positive relationship between DNA replication-related genes and the RS. CD8+ T cells were positively related to the RS. CD8+T cells primarily mediate anti-tumor immune responses [33]. A previous study showed better survival outcomes of cancer patients with high CD8+ T cell levels [34]. However, patients with a high RS associated with CD8+T cell score had poor prognosis, which might be related to tumor cell escape from immune surveillance mediated by CD8+T cells.



**Figure 7.** Predicting the survival of patients with different clinicopathological features, including age (age >65 and ≤65), gender, T stage (T1-2, T3-4), AJCC stage (stage I–II and stage III–IV), and pathological grade (G1-2 and G3-4) using the RS model.

In summary, we identified three molecular HCC subtypes based on TRGs using NMF. An RS model was established using these 13 TRGs and the prognostic value and application of the RS model was verified in clinical settings. However, additional experimental validation is needed to confirm the underlying mechanism of TRGs of the RS model in HCC.

## Abbreviations

TME: Tumor microenvironment; HCC: Hepatocellular carcinoma; NMF: Non-negative matrix factorization; ROC: Receiver Operating Characteristic; AUC: Area under curve; C-index: Concordance index; DCs: Dendritic cells; NK cell: Natural killer cell; MDSC: Myeloid-derived suppressor cell; TAM: Tumor-associated macrophage; TAN: Tumor-associated neutrophil; Treg: Regulatory T cell; TIL: Tumor-infiltrating lymphocyte; CTL: CD8+ cytotoxic T lymphocyte; IHC: Immunohistochemistry; CyTOF: Cytometry; GDC: Genomic Data Commons; TCGA: The Cancer Genome Atlas Consortium; GEO: Gene Expression Omnibus; MCP-counter: Microenvironment Cell Populations-counter; DCA: Decision curve analysis; GSEA: Gene set enrichment analysis; RMS: Restricted mean survival; LRG: Low-risk group; HRG: High-risk group.

## AUTHOR CONTRIBUTIONS

The study was designed by ZFC and GL. YBS participated in the literature review. XLZ and JW performed data analysis and interpretation. The original draft of the manuscript was done by SLH and CGC. GML and ZFC reviewed and edited the manuscript. The manuscript was reviewed and approved by all authors.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

## FUNDING

No funding was used for this paper.

## REFERENCES

1. Gao B, Jeong WI, Tian Z. Liver: An organ with predominant innate immunity. *Hepatology*. 2008; 47:729–36. <https://doi.org/10.1002/hep.22034> PMID:18167066
2. Fu Y, Liu S, Zeng S, Shen H. From bench to bed: the tumor immune microenvironment and current immunotherapeutic strategies for hepatocellular carcinoma. *J Exp Clin Cancer Res*. 2019; 38:396. <https://doi.org/10.1186/s13046-019-1396-4> PMID:31500650
3. Santhakumar C, Gane EJ, Liu K, McCaughan GW. Current perspectives on the tumor microenvironment in hepatocellular carcinoma. *Hepatol Int*. 2020; 14:947–57. <https://doi.org/10.1007/s12072-020-10104-3> PMID:33188512
4. Lu C, Rong D, Zhang B, Zheng W, Wang X, Chen Z, Tang W. Current perspectives on the immunosuppressive tumor microenvironment in hepatocellular carcinoma: challenges and opportunities. *Mol Cancer*. 2019; 18:130. <https://doi.org/10.1186/s12943-019-1047-6> PMID:31464625
5. Yang JD, Nakamura I, Roberts LR. The tumor microenvironment in hepatocellular carcinoma: current status and therapeutic targets. *Semin Cancer Biol*. 2011; 21:35–43. <https://doi.org/10.1016/j.semcancer.2010.10.007> PMID:20946957
6. Giraud J, Chalopin D, Blanc JF, Saleh M. Hepatocellular Carcinoma Immune Landscape and the Potential of Immunotherapies. *Front Immunol*. 2021; 12:655697. <https://doi.org/10.3389/fimmu.2021.655697> PMID:33815418
7. Yang M, Li J, Gu P, Fan X. The application of nanoparticles in cancer immunotherapy: Targeting tumor microenvironment. *Bioact Mater*. 2020; 6:1973–87. <https://doi.org/10.1016/j.bioactmat.2020.12.010> PMID:33426371
8. Xiao Y, Jiang J, Chen Y, Huang Y, Wei C. PD-1 Relevant Genes Predict the Prognosis of Breast Cancer and Their Prediction Effect in Tumor Response to Immunotherapy. *Oncologie*. 2022; 24:729–42. <https://doi.org/10.32604/oncologie.2022.026118>
9. Lin X, Tian C, Pan F, Wang R. A novel immune-associated prognostic signature based on the immune cell infiltration analysis for hepatocellular carcinoma. *Oncologie*. 2024; 26:91–103. <https://doi.org/10.1515/oncologie-2023-0360>
10. Gao L, Gui R, Zheng X, Wang Y, Gong Y, Wang TH, Wang J, Huang J, Liao X. Topical Application of *Houttuynia cordata* Thunb Ethanol Extracts Increases Tumor Infiltrating CD<sup>8+</sup> /Treg Cells Ratio and Inhibits Cutaneous Squamous Cell Carcinoma *in vivo*. *Oncologie*. 2022; 24:565–77.

- <https://doi.org/10.32604/oncologie.2022.022454>
11. Leonardi GC, Candido S, Cervello M, Nicolosi D, Raiti F, Travali S, Spandidos DA, Libra M. The tumor microenvironment in hepatocellular carcinoma (review). *Int J Oncol*. 2012; 40:1733–47. <https://doi.org/10.3892/ijo.2012.1408> PMID:22447316
  12. Sas Z, Cendrowicz E, Weinhäuser I, Rygiel TP. Tumor Microenvironment of Hepatocellular Carcinoma: Challenges and Opportunities for New Treatment Options. *Int J Mol Sci*. 2022; 23:3778. <https://doi.org/10.3390/ijms23073778> PMID:35409139
  13. Wilson S, Fitzsimons M, Ferguson M, Heath A, Jensen M, Miller J, Murphy MW, Porter J, Sahni H, Staudt L, Tang Y, Wang Z, Yu C, et al, and GDC Project. Developing Cancer Informatics Applications and Tools Using the NCI Genomic Data Commons API. *Cancer Res*. 2017; 77:e15–8. <https://doi.org/10.1158/0008-5472.CAN-17-0598> PMID:29092930
  14. Wu L, Quan W, Luo Q, Pan Y, Peng D, Zhang G. Identification of an Immune-Related Prognostic Predictor in Hepatocellular Carcinoma. *Front Mol Biosci*. 2020; 7:567950. <https://doi.org/10.3389/fmolb.2020.567950> PMID:33195412
  15. Zhang G, Su L, Lv X, Yang Q. A novel tumor doubling time-related immune gene signature for prognosis prediction in hepatocellular carcinoma. *Cancer Cell Int*. 2021; 21:522. <https://doi.org/10.1186/s12935-021-02227-w> PMID:34627241
  16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; 43:e47. <https://doi.org/10.1093/nar/gkv007> PMID:25605792
  17. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*. 2010; 11:367. <https://doi.org/10.1186/1471-2105-11-367> PMID:20598126
  18. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016; 17:218. <https://doi.org/10.1186/s13059-016-1070-5> PMID:27765066
  19. Blanco JL, Porto-Pazos AB, Pazos A, Fernandez-Lozano C. Prediction of high anti-angiogenic activity peptides in silico using a generalized linear model and feature selection. *Sci Rep*. 2018; 8:15688. <https://doi.org/10.1038/s41598-018-33911-z> PMID:30356060
  20. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013; 32:5381–97. <https://doi.org/10.1002/sim.5958> PMID:24027076
  21. Wickham H. Ggplot2: Elegant Graphics for Data Analysis. ggplot2. Springer New York, NY. 2009. <https://doi.org/10.1007/978-0-387-98141-3>
  22. Harrell FE Jr. With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Regression Modeling Strategies. Springer Cham. 2015. <https://doi.org/10.1007/978-3-319-19425-7>
  23. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, et al, and Cancer Genome Atlas Research Network. The Immune Landscape of Cancer. *Immunity*. 2018; 48:812–30.e14. <https://doi.org/10.1016/j.immuni.2018.03.023> PMID:29628290
  24. Lin Z, Xu Q, Miao D, Yu F. An Inflammatory Response-Related Gene Signature Can Impact the Immune Status and Predict the Prognosis of Hepatocellular Carcinoma. *Front Oncol*. 2021; 11:644416. <https://doi.org/10.3389/fonc.2021.644416> PMID:33828988
  25. Xia Y, Brown ZJ, Huang H, Tsung A. Metabolic reprogramming of immune cells: Shaping the tumor microenvironment in hepatocellular carcinoma. *Cancer Med*. 2021; 10:6374–83. <https://doi.org/10.1002/cam4.4177> PMID:34390203
  26. Clancy T, Dannenfelser R, Troyanskaya O, Malmberg KJ, Hovig E, Kristensen V. Bioinformatics Approaches to Profile the Tumor Microenvironment for Immunotherapeutic Discovery. *Curr Pharm Des*. 2017; 23:4716–25. <https://doi.org/10.2174/1381612823666170710154936> PMID:28699527
  27. Tsilimigras DI, Ntanasis-Stathopoulos I, Moris D, Pawlik TM. Liver Tumor Microenvironment. *Adv Exp Med Biol*. 2020; 1296:227–41. [https://doi.org/10.1007/978-3-030-59038-3\\_14](https://doi.org/10.1007/978-3-030-59038-3_14) PMID:34185296

28. Kurebayashi Y, Ojima H, Tsujikawa H, Kubota N, Maehara J, Abe Y, Kitago M, Shinoda M, Kitagawa Y, Sakamoto M. Landscape of immune microenvironment in hepatocellular carcinoma and its additional impact on histological and molecular classification. *Hepatology*. 2018; 68:1025–41.  
<https://doi.org/10.1002/hep.29904>  
PMID:[29603348](https://pubmed.ncbi.nlm.nih.gov/29603348/)
29. Zhang Q, Lou Y, Yang J, Wang J, Feng J, Zhao Y, Wang L, Huang X, Fu Q, Ye M, Zhang X, Chen Y, Ma C, et al. Integrated multiomic analysis reveals comprehensive tumour heterogeneity and novel immunophenotypic classification in hepatocellular carcinomas. *Gut*. 2019; 68:2019–31.  
<https://doi.org/10.1136/gutjnl-2019-318912>  
PMID:[31227589](https://pubmed.ncbi.nlm.nih.gov/31227589/)
30. Currie E, Schulze A, Zechner R, Walther TC, Farese RV Jr. Cellular fatty acid metabolism and cancer. *Cell Metab*. 2013; 18:153–61.  
<https://doi.org/10.1016/j.cmet.2013.05.017>  
PMID:[23791484](https://pubmed.ncbi.nlm.nih.gov/23791484/)
31. Macheret M, Halazonetis TD. DNA replication stress as a hallmark of cancer. *Annu Rev Pathol*. 2015; 10:425–48.  
<https://doi.org/10.1146/annurev-pathol-012414-040424>  
PMID:[25621662](https://pubmed.ncbi.nlm.nih.gov/25621662/)
32. Williams GH, Stoeber K. The cell cycle and cancer. *J Pathol*. 2012; 226:352–64.  
<https://doi.org/10.1002/path.3022>  
PMID:[21990031](https://pubmed.ncbi.nlm.nih.gov/21990031/)
33. van der Leun AM, Thommen DS, Schumacher TN. CD8<sup>+</sup> T cell states in human cancer: insights from single-cell analysis. *Nat Rev Cancer*. 2020; 20:218–32.  
<https://doi.org/10.1038/s41568-019-0235-4>  
PMID:[32024970](https://pubmed.ncbi.nlm.nih.gov/32024970/)
34. Oshi M, Asaoka M, Tokumaru Y, Yan L, Matsuyama R, Ishikawa T, Endo I, Takabe K. CD8 T Cell Score as a Prognostic Biomarker for Triple Negative Breast Cancer. *Int J Mol Sci*. 2020; 21:6968.  
<https://doi.org/10.3390/ijms21186968>  
PMID:[32971948](https://pubmed.ncbi.nlm.nih.gov/32971948/)