# Integrative analysis of the characteristic of lipid metabolism-related genes for the prognostic prediction of hepatocellular carcinoma

**Wenbiao Chen**
Zhejiang University
**Jun Zeng**
University of the Chinese Academy of Sciences
**Minglin Ou**
Guilin Medical University
**Donge Tang** ( donge66@126.com )
Shenzhen People's Hospital https://orcid.org/0000-0001-7011-196X

**Primary research**

**Keywords:** lipid metabolism, gene, prognosis, hepatocellular carcinoma, bioinformatics analysis

**Posted Date:** May 15th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-27431/v1

## Abstract

# Background

Dysregulation of lipid metabolism has been implicated in the progression of hepatocellular carcinoma (HCC). We therefore investigated the molecular characteristics of lipid-metabolism-related genes for the prognostic prediction of HCC.

# Methods

Multi-dimensional bioinformatics analysis was conducted to comprehensively analyzed the lipid metabolism-related genes (IMRG) and construct the prognostic prediction signature.

# Results

A total of 770 HCC patients and their corresponding 776 IMRGs were downloaded from three databases. The HCC patients were classified into 2 molecular clusters, which were associated with overall survival, clinical characteristics, and immune cells. The biological function of the differentially expressed IMRGs in the 2 clusters showed that the genes were associated with tumor-related metabolism pathways. A 6 IMRGs signature (6-IS), including *FMO3, SLC11A1, RNF10, KCNH2, ME1*, and *ZIC2* were established for HCC prognostic prediction, which was found to be an independent prognostic factor. Performance of the 6-IS prognostic signature was verified in a validation set and compared with an external data set. Results revealed that the 6-IS signature could effectively predict the prognosis of patients with HCC.

# Conclusion

This study provides new insights into the role of IMRG in the pathogenesis of HCC and presents a novel signature 6-IS to predict the prognosis of HCC.

## Introduction

HCC is one of the most common malignant tumors characterized by high malignancy, high metastatic potential, and poor prognosis [1]. Most patients with HCC are diagnosed at the advanced stage, at this stage they cannot benefit from surgery or chemoradiotherapy [2]. In recent years, therapies based on biological targets have been proposed for patients with HCC [3]. However, the clinical benefit of available biomarker for early diagnosis and prognostic assessment was still limited. Thus, it is important to study the pathogenesis of HCC and find specific targets that can be used to improve early diagnosis and prognostic prediction.

In human, many metabolic functions take place in liver cells [4]. Being an important site for lipid metabolism, HCC results in many lipid metabolic abnormalities [5]. Previous studies have shown that HCC is accompanied with abnormal changes such as increased de novo synthesis of fatty acids, suppressed oxidation levels, high secretion of insulin and insulin-like growth factors, and abnormal metabolism of phosphatidylcholine [6]. These metabolic processes provide intermediate energy substrates that enable HCC cells to grow, proliferate and metastasize [7]. In addition, several enzymes and signaling molecules, such as 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGGR) and AKT/mTORC1 pathway regulate lipid metabolism of HCC cells. Thus, the metabolic enzymes and pathways associated with these processes can be used as biomarker for diagnosis and treatment of HCC[8].

Numerous studies have been carried out to uncover the biological phenotypes and molecular classification of HCC on the basis of lipid metabolic patterns [9]. Fox example, the de novo fatty acid synthesis phenotype in tumor cells has been associated with up-regulated lipid-related genes at multiple levels, such as transcription, translation and post-translation modification, and enzyme activity, as well as the influence of these genes on oncogenes [10]. Moreover, molecular classification of HCC based on lipid metabolism-related genes reveals distinct tumor subtypes. Using bioinformatic methods, Gholamreza *et al* subdivided HCC patients in three clusters with distinct metabolic and signaling pathways at the genome, transcriptome, and proteome levels. These clusters were associated with clinical features and survival rate [11]. However, lipid metabolic programme and molecules have not been fully exploited in the prognostic prediction of HCC patients.

In this study, data of 770 HCC patients were divided into two molecular clusters based on 776 lipid IMRGs. The two molecular clusters were associated with the clinical features, immune infiltration, and tumor metabolism-related biological processes. We also established a prognostic signature for the HCC patients. A flow chart showing the protocol of this study is shown in Supplementary Fig. 1. This study extends our understanding of the molecular basis of lipid metabolism involved in the pathogenesis of HCC and suggested the lipid metabolism-related genes for the hallmark of prognostic prediction in HCC patient.

# Methods

# Patients information and genome expression dataset

Multiple datasets were downloaded from several databases including the cancer genome atlas (TCGA), Gene Expression Omnibus (GEO), and Database of Hepatocellular Carcinoma Expression Atlas (HCCDB). 371 samples obtained from TCGA were subjected to quality control and filtration procedure and we collected 342 samples that met the conditions, which were randomly divided into training and validation sets. To avoid random allocation bias which may affect the stability of subsequent modeling, all samples were sampled 100 times in advance to ensure that training and validation sets were consistent in clinical features. The GSE15654 data set obtained from GEO database were preprocessed for quality control and

filtration, and we finally got 216 samples. The HCCDB18 dataset containing 212 samples and corresponding clinical information derived from HCCDB database were downloaded directly. Detailed information of the three datasets are presented in supplementary Table 1.

# Molecular Classification Of HCC Based On Lipid Metabolism-related Genes

The IMRGs were obtained from 6 lipid metabolism-related pathways (supplementary Table 2) in Molecular Signature Database v7.0 (MSigDB) (www.gsea-msigdb.org/gsea/msigdb). A total of 776 IMRGs were retained after exclusion of overlapping genes. We extracted 776 IMRGs from the TCGA expression profile data, and genes with expression value above 0 in more than half of samples were retained. Finally, 739 IMRGs were enrolled for subsequent analysis. Univariate cox analysis was performed on 739 IMRGs with coxph R package to mine out HCC-related IMRGs. Next, the HCC-related IMRGs were processed through non-negative matrix factorization (NMF) clustering algorithm using the NMF R package. The NMF analysis and 50 iterations were carried out with the standard "brunet" pattern [12]. *k* values which indicates the optimal number of clusters ranged from 2 to 10. The average contour width of the common member matrix was determined with the NMF R package, and the minimum member of each subclass was set to 10. The optimal *k* value was determined from indicators of cophenetic, residual sum of squares (RSS), and silhouette. Differences in clinical features between the clusters based on HCC-related IMRGs were compared using Chi-square test. The Tumor Immune Estimation Resource (TIMER) (https://cistrome.shinyapps.io/timer/) algorithm was employed to investigate the association between clusters and immune score.

# Construction Of A Prognostic Signature Based On IMRGs

Differential expression of HCC-related IMRGs between clusters was analyzed by the DESeq2 algorithm using limma R package. Significant IMRGs were those with false discovery rate (FDR) < 0.05 and absolute value of log2 fold change > 1. Next, significant differentially expressed HCC-related IMRGs were subjected to univariate cox analysis to determine their association with survival of HCC using survival coxph function R package. The log rank $P < 0.01$ was set as the threshold. To narrow the gene range and build a prognostic model with high accuracy, we used the LASSO method to reduce the dimensionality and select the most significant differentially expressed HCC-related IMRGs. The 10-fold cross validations methods were employed to select optimal values of the penalty parameter lambda [13]. Next, multivariate cox analysis was performed on the genes obtained in the above steps, and the least value of Akaike information criterion (AIC) within cox proportional regression model was calculated to retain the most significant genes to construct an IMRGs signature. A risk score based on IMRGs signature set was calculated as follows: risk score = $\text{expression}_{\text{gene 1}} \times \beta_{\text{gene 1}} + \text{expression}_{\text{gene 2}} \times \beta_{\text{gene 2}} + \cdots + \text{expression}_{\text{gene x}} \times \beta_{\text{gene x}}$, where x was the number of IMRGs, $\beta$ was the coefficient value for each IMRGs.

We normalized the risk score to z-score using binormalization process algorithm, and the z-score value > 0 and < 0 for the samples were classified into high and low risk groups, respectively.

# Statistical analysis

The pheatmap R package was used to display unsupervised hierarchical clustering heatmap of HCC-related IMRGs and a volcano Plot of the differentially expressed IMRGs between clusters was plotted using a ggplot2 R package. The OS was estimated using the Kaplan–Meier (KM) method, and the sensitivity and specificity of the survival curve were assessed through receiver operating characteristic (ROC) curve by calculating the area under curve (AUC) of ROC using pROC R package. The GO and KEGG analyses were performed for the differentially expressed IMRGs using clusterprofiler R package. $p < 0.05$ was set as the threshold of statistical significance. Independent *t*-test and Mann Whitney *U* test were conducted to compare variables between groups, for variables following normal and abnormal distribution, respectively. The association between IMRGs signature and clinical features was analyzed by univariate and multivariate survival analyses. The association between the IMRGs signature and immune/stromal score was determined by calculating the immune and stromal scores of each sample using estimate R package comparing high and low risk groups. The potential mechanisms of IMRGs signature were analyzed by Gene Set Enrichment Analysis (GSEA) analysis using GSVA R package. Pearson correlation coefficients were used to analyze the association between IMRGs signature and biological functions. The prognostic value of the IMRGs signature and other signatures was assessed by Harrell's concordance index (c-index) using rms R package. Restricted mean survival time (RMST) is an index of the area under the KM curve at a specific timepoint. It was used to evaluate the predictive value of the IMRGs signature at different timepoints. All statistical analyses were performed using the SPSS Version 25.0 software and R software version 3.4.0, and a $p < 0.05$ was considered statistically significant.

## Results

# Identification of molecular subtypes based on IMRGs

Univariate cox analysis was performed on preprocessed 739 IMRGs obtained from TCGA dataset. In total, 324 HCC-related IMRGs were identified and used for HCC classification. The cophenetic coefficients, which indicate the stability of classified cluster was used to calculate the optimal *k* value. We performed comprehensive analysis on the cophenetic, RSS, and silhouette index, from which we selected the $k = 2$ as the optimal value. Consequently, 2 molecular subtypes (cluster 1 and cluster 2) were identified based on IMRGs (Supplementary Fig. 2). Notably, the matrix heat map exhibited clear boundaries based on *k* value of 2, suggesting that the molecular subtypes classification was stable (Fig. 1A). The gene cluster heatmap of 324 HCC-related IMRGs revealed marked differences between cluster 1 (C1) and cluster 2 (C2). Specifically, the expression level of HCC-related IMRGs in C2 was significantly higher than C1. In addition, distribution of clinical features between C1 and C2 exhibited significant differences (Fig. 1B). KM analysis revealed that C2 had significant shorter OS than C1 ($P = 0.0099$) (Fig. 1C). The accuracy of

molecular subtypes classification based on IMRGs was determined by comparing the association between the 2 clusters and clinical features using Chi-square test. Results showed that the pathological classification of tumor (T) ($P$ = 0.0002), stage ($P$ = 0.0478), and grade ($P$ = 0.0391) were significantly different between the 2 clusters (Supplementary Table 3). Further comparison of the immune scores between the 2 clusters was performed using TIMER algorithm. Except CD8 cells, the immune scores of B cell ($P$ = 0.045), CD4 T cell ($P$ = 0.015), neutrophil ($P$ = 0.015), macrophage ($p$ = 0.001), and dendritic ($p$ = 0.001) in clusters 2 were higher than that of cluster 1 (Fig. 1C). Collectively, these results revealed that the IMRGs signature could classify HCC into distinct molecular subtypes and was associated with clinical characteristic.

# Construction And Validation Of An IMRGs Signature

First, we screened for differentially expressed IMRGs between cluster 1 and clusters 2. A total of 400 IMRGs were found to be significantly differentially expressed. A volcano and clustering map revealed a distinct distribution of up-regulated and down-regulated IMRGs between the 2 clusters (Supplementary Fig. 3A, B). Results of GO analysis showed that the differentially expressed IMRGs were primarily enriched in metabolic process, such as glutamate and lactate metabolism, as well as in tumorigenesis-related process, including cell-cell adhesion and cell migration (Supplementary Fig. 3C). In the KEGG analysis, IMRGs were predominantly enriched in metabolic pathways, such as glucagon signaling, metabolism of xenobiotics by cytochrome P450, and retinol metabolism (Supplementary Fig. 3D). Functionally, the differentially expressed IMRGs were involved in tumorigenesis and metabolism related pathways.

Univariate Cox regression and LASSO (Fig. 2A, B) analyses were conducted to select suitable genes from the 400 differentially expressed IMRGs. 20 significant genes were revealed through above these two analyses were further subjected to a multivariate Cox regression analysis with the mimic AIC value = 466.72. Finally, 6 IMRGs (Supplementary Table 4) were used to construct an IMRGs signature (termed as 6-IS) using the risk score formula. Next, we explored the association between 6 IMRGs and HCC survival. Unlike *FMO3* which correlated with good prognosis in high-risk group, *SLC11A1*, *RNF10*, *KCNH2*, *ME1*, and *ZIC2* correlated with shorter survival time in the high-risk group than in the low-risk group (Fig. 2C). Moreover, the 6 IMRGs signature predicted significant differences in survival outcomes between C1 and C2. We then analyzed the expression profile of the 6 IMRGs in the two clusters. Except *FMO3*, the expression of *SLC11A1*, *RNF10*, *KCNH2*, *ME1*, and *ZIC2* in C2 were significantly higher than in C1 (Fig. 2D). Thus, *SLC11A1*, *RNF10*, *KCNH2*, *ME1*, and *ZIC2* were considered as the hazard indexes and *FMO3* as the protective index for construction of an independent prognostic IMRGs signature.

According to the calculation of the risk scores of 6-IS in each sample, we depicted the risk score plot, survival status, and expression profiles of the 6 IMRGs in patients from training set. We found that HCC patients with high-risk scores had higher mortality rates than those with low risk scores, and the changes in expression of 6 IMRGs with increased risk score revealed that *SLC11A1*, *RNF10*, *KCNH2*, *ME1*, and *ZIC2*

as hazard index and *FMO3* as protective index (Fig. 3A). In the ROC analysis, area under ROC curve (AUC) for the 6-IS signature was 0.80 for 1 year, 0.82 for 3 years, and 0.84 for 5 years, indicating a high prognostic prediction accuracy of the 6-IS signature (Fig. 3B). In the training cohort, patients were divided into high and low risk groups. KM analysis based on 6-IS showed that the OS of low risk groups was significantly better than that of high risk group (Fig. 3C). These results showed that the 6-IS could serve as an independent signature predicting the survival outcomes of HCC patients in the validation set (Supplementary Fig. 4A), GSE15654 (Supplementary Fig. 4B), and HCCDB18 sets (Supplementary Fig. 4C). Taken together, these findings show that the 6-IS could effectively predict the prognosis of HCC patients.

### Association of 6-IS signature with clinical features and molecular characteristics of HCC

KM analysis showed that the clinical features, including alpha-fetoprotein (AFP) ($p$ = 0.02871), stage ($p$ = 3e − 05), T ($p$ = 2e − 05), N (Lymph node) ($p$ = 0.02519), and M (Metastasis) ($p$ = 0.00223) could divide the HCC patients of training set based on OS analysis (Supplementary Fig. 5). Next, we predicted the OS of HCC patients using the 6-IS signature according to the above clinical features (AFP > 20, AFP < = 20, T, N, M, Stage I + II, and Stage III + IV). Consistently, we found that 6-IS signature could distinguish the low-risk group from high-risk group. This analysis also revealed that HCC patients in the high-risk groups had significantly shorter survival time than those in low-risk group (Fig. 4). Univariate Cox and multivariable Cox analyses were performed to verify the prognostic value of 6-IS in HCC patients from TCGA and HCCDB18 databases. Thus, we concluded that 6-IS is an independent prognostic marker associated with survival when treated as continuous variable both in TCGA ($P$ = 2.97E-09 and $P$ = 0.0049) and HCCDB18 ($P$ = 0.032 and $P$ = 0.0453) sets (Table 1). In subsequent analyses, we calculated immune, stromal and estimate scores for each sample from TCGA. Except stromal score, immune and estimate scores of high-risk group were significantly higher than those of the low-risk group (Supplementary Fig. 6A). Similar results were obtained in the HCCDB18 dataset (Supplementary Fig. 6B). We then compared expression of IMRGs in samples from TCGA, HCCDB18, and GSE15654 sets using GSEA analysis. An ssGSEA value was obtained which was used to infer the association between 6-IS risk score and biological function. Most of the biological functions were negatively associated with 6-IS risk score, such as glyoxylate and dicarboxylate metabolism, drug metabolism cytochrome p450, and beta alanine metabolism. In contrast, biological functions related to tumorigenesis, including glycerophospholipid metabolism, fatty acid metabolism, cell cycle, and RNA degradation were positively linked to the 6-IS risk score (Supplementary Fig. 7A). The clustering heatmap based on ssGSEA values significantly revealed the biological pathways positively or negatively correlated with 6-IS risk scores (Supplementary Fig. 7B).

Table 1
Univariate and multivariable Cox analyses to identify prognostic-related clinical factors.

| Variables | Univariate analysis | | | Multivariable analysis | | |
|---|---|---|---|---|---|---|
| | HR | 95%CI of HR | P value | HR | 95%CI of HR | P value |
| **Entire TCGA cohort** | | | | | | |
| Risk score (High/Low) | 3.025 | 2.099–4.361 | 2.97E-09 | 1.997 | 1.234–3.233 | 0.0049 |
| Age | 1.008 | 0.994–1.022 | 0.231 | 1.021 | 1.001–1.041 | 0.0434 |
| Gender (Male/Female) | 0.8 | 0.556–1.150 | 0.229 | 0.881 | 0.542–1.43 | 0.6072 |
| AFP | 1.748 | 1.105–2.765 | 0.017 | 2.455 | 1.392–4.333 | 0.0019 |
| T3/T4 vs T1/T2 | 2.838 | 1.981–4.067 | 1.32E-08 | 1.285 | 0.775–2.13 | 0.3318 |
| N1/N2 vs N0 | 1.617 | 1.112–2.349 | 0.012 | 1.030 | 0.554–1.916 | 0.9248 |
| M1/MX vs M0 | 1.795 | 1.235–2.608 | 0.002 | 2.972 | 1.611–5.482 | 0.0005 |
| Stage □/□ vs Stage □/□ | 2.767 | 1.893–4.046 | 1.51E-07 | 0.851 | 0.585–1.239 | 0.4010 |
| G3/G4 vs G1/G2 | 1.069 | 0.736–1.553 | 0.724 | 1.737 | 1.054–2.864 | 0.0304 |
| **ICGA cohort** | | | | | | |
| Risk score (High/Low) | 2.088 | 1.066–4.089 | 0.032 | 1.955 | 0.989–3.859 | 0.0453 |
| Age | 1.015 | 0.979–1.052 | 0.406 | 1.004 | 0.968–1.041 | 0.8422 |
| Gender (Male/Female) | 0.516 | 0.256–1.039 | 0.064 | 0.360 | 0.166–0.782 | 0.0098 |
| Stage □/□ vs Stage □/□ | 2.737 | 1.415–5.295 | 0.0028 | 3.462 | 1.711–7.003 | 0.0006 |

# Comparison Of 6-is Signature With External Models

Six genes signature [14], eight genes signature [15], six genes-based prognostic signature [16], and four genes signature [17] were used as the external data set for validation tests. These signatures were established to calculate risk score and assess the OS of patients in TCGA using similar methods as in our study. In line with 6-IS, KM analysis showed that all the four models could divide HCC patients into high and low-risk groups, and the high-risk groups had significantly shorter survival time than low-risk groups (Fig. 5A-D). However, except that eight-genes signature (0.813) which showed similar results with our study, ROC analysis revealed that the average AUC value of 1, 3, 5-year from six genes signature (0.613), six genes-based prognostic signature (0.770), and four genes signature (0.686) were lower than that of 6-IS (0.82) (Fig. 5A-D). In the c-index analysis, the 6-IS showed better prognostic ability than other four models (Fig. 5E). Besides, RMST analysis revealed that 6-IS performed better than other four models in

the prognostic prediction of HCC patients (Fig. 5F). These results confirmed that the 6-IS is a robust prognostic prediction signature.

# Discussion

Accumulating evidence show that tumorigenesis is accompanied with metabolic reprogramming of various nutrients that not only sustain cancer cell survival, but also regulate gene expression, emergence of mutations, and immune tumor microenvironment [18]. The most classic example of tumor metabolic reprogramming is the "Warburg effect", in which cancer cells tend to use glycolysis to replace normal cells that thrive on aerobic metabolism for survival [19]. Abnormalities in glucose and lipid metabolism in tumors have been the focus of recent studies [20]. Currently, abnormal lipid metabolism in HCC, especially genes related to lipid metabolism have not fully explored to determine their role in the pathogenesis, diagnosis, and treatment of HCC. In this study, we used multi-dimensional bioinformatics methods to screen out abnormally regulated genes related to lipid metabolism in HCC, and used these genes to construct a prognostic prediction signature.

Deregulation of genes related to lipid metabolism have been implicated in the tumorigenesis of HCC. For example, *ADH1A* triggered oncogenic transformation of hepatocytes leading to poor survival [21], whereas extracellular *PEDF* inhibited angiogenesis in HCC by inducing lipid metabolic disorders [22]. These insights into the molecular mechanisms and genes markers involved in the pathogenesis of HCC have extended our understanding of the metabolic profile of HCC. However, the clinical utility of single genes targets in HCC has been challenging. Here, we used data-mining bioinformatic approaches to explore the relationships between lipid metabolism related genes and clinical features of HCC. A prognostic signature 6-IS was constructed which showed good prediction results. The prognostic signature 6-IS was also associated with the overall survival, clinical features and metabolic signaling pathways of patients of HCC. The 6-IS comprised six genes obtained using multidimensional algorithms, including *FMO3*, *SLC11A1*, *RNF10*, *KCNH2*, *ME1*, and *ZIC2.* This signature is superior as it overcomes the shortcomings of single genes such as interference from other factors. Consistent with a previous study, we found that *FMO3* suppresses tumor progression by decreasing cell viability, hence it is a protective index [23]. In addition, our analysis revealed that *ME1* or *ZIC2*, which exhibit stem-cell features, correlated with poor prognosis, hence are risk indexes [24, 25]. However, these previous studies reported the presence of potential errors and bias in their analyses. For this reason, we constructed a statistical signature with multiple genes comprising clinical information to improve the efficiency of prognostic prediction. Notably, our 6-IS signature performed better than other models from external datasets in terms of prognostic prediction. Nevertheless, we acknowledge that further studies are needed to validate the performance of 6-IS in a prospective cohort.

Immune estimation tools, such as TIMER and ESTIMATE were used to calculate immune scores to assess the immune infiltration cells across groups. Similar to other studies, we found that HCC is characterized by heavy infiltration by immune cells, and that the inflammatory response in the liver is the main mechanism contributing to hepatitis, cirrhosis, and HCC [26]. Unlike other studies that used global

transcriptome of HCC to analyze the immune cells composition and perform molecular classification, we only used IMRGs to investigate the immune score and carry out metabolism stratification. A strong link between metabolism and immunity has been demonstrated during tumorigenesis[27]. Results in this study suggested that IMRGs regulate immune cells during the development of HCC. Accumulating evidence show that metabolic and epigenetic reprogramming are important mechanisms that regulate tumor immunity, and most epigenetic reprogramming genes are those related to fatty acid, cholesterol esters and phosphatidylcholine metabolism [28]. In KEGG and GO analyses, we found that the differentially expressed IMRGs between clusters were enriched in diverse metabolic pathways. Moreover, GSEA analysis revealed that upregulated or downregulated IMRGs used to construct the 6-IS were associated with metabolic pathways. We hypothesized that the IMRGs participated in several metabolic pathways that modulate functions and phenotypes of immune cells during the pathogenesis of HCC.

Clinical features, such as AFP, TNM, tumor stage and grade, and other pathological classifications are widely used in the clinical management of HCC [29]. However, these data are biased and lack specificity. This calls for identification of more accurate indicators and phenotypes to improve existing diagnostic and therapeutic guidelines [30]. In this study, we predicted the OS based on AFP, stage, and TNM indicators and then compared the performance of 6-IS with the above clinical features. Results showed that 6-IS was superior to other clinical features in predicting the prognosis of HCC. In addition, 6-IS was found to be an independent factor when compared with other clinical features. Given that 6-IS is yet to be verified in prospective patients, we suggest that it is combined with the traditional clinical features to improve the clinical management of HCC.

# Conclusion

In conclusion, we present an IMRG-based signature to classify HCC patients into molecular clusters based on metabolic profiles. The 6-IS was found to be a robust prognostic prediction marker for HCC patients. This signature was associated with clinical features, immune cells and various functions. This study provides novel insights to the prognostic value of lipid metabolism in HCC.

# Abbreviations

HCC: hepatocellular carcinoma; IMRG: lipid metabolism-related genes; LASSO: least absolute shrinkage and selection operator; 6-IS: 6 IMRGs signature; HMGGR: 3-hydroxy-3-methylglutaryl-coenzyme A reductase; CPT: carnitine palmitoyltransferase; TCGA: the cancer genome atlas; GEO: Gene Expression Omnibus; FPKM: Fragments per Kilobase Million; HCCDB: Hepatocellular Carcinoma Expression Atlas; NMF: non-negative matrix factorization; TIMER: Tumor Immune Estimation Resource; FDR: false discovery rate; AIC: Akaike information criterion; KM: Kaplan–Meier; ROC: receiver operating characteristic; AUC: area under curve; GSEA: Gene Set Enrichment Analysis; RMST: Restricted mean survival time

# Declarations

### Acknowledgements

Not applicable.

### Authors' contributions

Conceptualization, DT; Interpretation of data, WC, JZ, MO, and DT; Formal analysis, WC, MO, and JZ; Funding acquisition, DT; Investigation, WC, JZ, MO, and DT; Project administration, DT; Visualization, WC and JZ; Writing—original draft, WC; revision and edition, DT. All authors read and approved the final manuscript.

### Funding

This work was supported by Key Research & Development Plan of Zhejiang Province (2019C04005), the major national S&T projects for infectious diseases (2018ZX10301401-005), the National Natural Science Foundation of China (No. 81571953, 81671596). Basic research discipline layout of Shenzhen science and technology innovation committee (JCYJ20180507183428877).

### Availability of data and materials

The datasets generated and analyzed during the current study are available in the public database.

### Ethics approval and consent to participate

The study was approved by the clinical ethical committee of Shenzhen People's Hospital.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

## References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. Cancer J Clin. 2018;68(1):7–30.
2. Kulik L, El-Serag HB. Epidemiology and Management of Hepatocellular Carcinoma. Gastroenterology. 2019;156(2):477–91.e471.
3. Liu Z, Lin Y, Zhang J, Zhang Y, Li Y, Liu Z, Li Q, Luo M, Liang R, Ye J. Molecular targeted and immune checkpoint therapy for advanced hepatocellular carcinoma. Journal of experimental clinical cancer research: CR. 2019;38(1):447.
4. Rui L. Energy metabolism in the liver. Comprehensive Physiology. 2014;4(1):177–97.

5. Hu B, Lin JZ, Yang XB, Sang XT. **Aberrant lipid metabolism in hepatocellular carcinoma cells as well as immune microenvironment: A review**. *Cell proliferation* 2020:e12772.
6. Nakagawa H, Hayata Y, Kawamura S, Yamada T, Fujiwara N, Koike K. **Lipid Metabolic Reprogramming in Hepatocellular Carcinoma**. *Cancers* 2018, 10(11).
7. Wu JM, Skill NJ, Maluccio MA. Evidence of aberrant lipid metabolism in hepatitis C and hepatocellular carcinoma. HPB: the official journal of the International Hepato Pancreato Biliary Association. 2010;12(9):625–36.
8. Currie E, Schulze A, Zechner R, Walther TC, Farese RV Jr. Cellular fatty acid metabolism and cancer. Cell Metabol. 2013;18(2):153–61.
9. Beyoglu D, Imbeaud S, Maurhofer O, Bioulac-Sage P, Zucman-Rossi J, Dufour JF, Idle JR. Tissue metabolomics of hepatocellular carcinoma: tumor energy metabolism and the role of transcriptomic classification. Hepatology. 2013;58(1):229–38.
10. Che L, Pilo MG, Cigliano A, Latte G, Simile MM, Ribback S, Dombrowski F, Evert M, Chen X, Calvisi DF. Oncogene dependent requirement of fatty acid synthase in hepatocellular carcinoma. Cell cycle (Georgetown Tex). 2017;16(6):499–507.
11. Bidkhori G, Benfeitas R, Klevstig M, Zhang C, Nielsen J, Uhlen M, Boren J, Mardinoglu A. Metabolic network-based stratification of hepatocellular carcinoma reveals three distinct tumor subtypes. Proc Natl Acad Sci USA. 2018;115(50):E11874–83.
12. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci USA. 2004;101(12):4164–9.
13. Gui J, Li H. Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. Bioinformatics. 2005;21(13):3001–8.
14. Liu GM, Zeng HD, Zhang CY, Xu JW. Identification of a six-gene signature predicting overall survival for hepatocellular carcinoma. Cancer cell international. 2019;19:138.
15. Qiao GJ, Chen L, Wu JC, Li ZR. Identification of an eight-gene signature for survival prediction for patients with hepatocellular carcinoma based on integrated bioinformatics analysis. PeerJ. 2019;7:e6548.
16. Wang Z, Teng D, Li Y, Hu Z, Liu L, Zheng H. A six-gene-based prognostic signature for hepatocellular carcinoma overall survival prediction. Life sciences. 2018;203:83–91.
17. Zheng Y, Liu Y, Zhao S, Zheng Z, Shen C, An L, Yuan Y. Large-scale analysis reveals a novel risk score to predict overall survival in hepatocellular carcinoma. Cancer management research. 2018;10:6079–96.
18. Pavlova NN, Thompson CB. The Emerging Hallmarks of Cancer Metabolism. Cell Metabol. 2016;23(1):27–47.
19. Yoshida GJ. Metabolic reprogramming: the emerging concept and associated therapeutic strategies. Journal of experimental clinical cancer research: CR. 2015;34:111.

20. Buechler C, Aslanidis C. Role of lipids in pathophysiology, diagnosis and therapy of hepatocellular carcinoma. Biochimica et biophysica acta Molecular cell biology of lipids. 2020;1865(5):158658.
21. Zahid KR, Yao S, Khan ARR, Raza U, Gou D. mTOR/HDAC1 Crosstalk Mediated Suppression of ADH1A and ALDH2 Links Alcohol Metabolism to Hepatocellular Carcinoma Onset and Progression in silico. Frontiers in oncology. 2019;9:1000.
22. Li C, Huang Z, Zhu L, Yu X, Gao T, Feng J, Hong H, Yin H, Zhou T, Qi W, et al. The contrary intracellular and extracellular functions of PEDF in HCC development. Cell death disease. 2019;10(10):742.
23. Hlady RA, Sathyanarayan A, Thompson JJ, Zhou D, Wu Q, Pham K, Lee JH, Liu C, Robertson KD. Integrating the Epigenome to Identify Drivers of Hepatocellular Carcinoma. Hepatology. 2019;69(2):639–52.
24. Mihara Y, Akiba J, Ogasawara S, Kondo R, Fukushima H, Itadani H, Obara H, Kakuma T, Kusano H, Naito Y, et al. Malic enzyme 1 is a potential marker of combined hepatocellular cholangiocarcinoma, subtype with stem-cell features, intermediate-cell type. Hepatology research: the official journal of the Japan Society of Hepatology. 2019;49(9):1066–75.
25. Zhu P, Wang Y, He L, Huang G, Du Y, Zhang G, Yan X, Xia P, Ye B, Wang S, et al. ZIC2-dependent OCT4 activation drives self-renewal of human liver cancer stem cells. J Clin Investig. 2015;125(10):3795–808.
26. 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(3):1025–41.
27. Zhang Q, Lou Y, Bai XL, Liang TB. Immunometabolism: A novel perspective of liver cancer microenvironment and its influence on tumor progression. World journal of gastroenterology. 2018;24(31):3500–12.
28. O'Neill LA, Kishton RJ, Rathmell J. A guide to immunometabolism for immunologists. Nature reviews Immunology. 2016;16(9):553–65.
29. Juarez-Hernandez E, Motola-Kuba D, Chavez-Tapia NC, Uribe M, Barbero Becerra V. Biomarkers in hepatocellular carcinoma: an overview. Expert Rev Gastroenterol Hepatol. 2017;11(6):549–58.
30. De Stefano F, Chacon E, Turcios L, Marti F, Gedaly R. Novel biomarkers in hepatocellular carcinoma. Digestive liver disease: official journal of the Italian Society of Gastroenterology the Italian Association for the Study of the Liver. 2018;50(11):1115–23.

# Supplementary Figure Legends

Supplementary Fig. 1. A flow chart of the study.

Supplementary Fig. 2. The classification of HCC patients using NMF analysis. The optimal *k* value was determined using indicators of cophenetic, residual sum of squares (RSS), and silhouette.

Supplementary Fig. 3. Analysis of differentially expressed IMRGs and their biological function. The (A) volcano and (B) clustering maps of differentially expressed IMRGs between the 2 clusters. (C) KEGG and (D) GO analyses for the differentially expressed IMRGs.

Supplementary Fig. 4. The performance of 6-IS in the prognostic prediction in the (A) validation set, (B) GSE15654, and (C) HCCDB18 sets.

Supplementary Fig. 5. KM analysis of overall survival of HCC patients based on clinical features, including age, AFP, race, gender, family history, grade, stage, and TNM.

Supplementary Fig. 6. Association between immune and stromal scores and 6-IS determined using ESTIMATE algorithm. Distribution of stromal score, immune score, and the estimate score between high- and low-risk groups in the (A) TCGA and (B) HCCDB18 sets.

Supplementary Fig. 7. The biological function of 6-IS determined using GSEA method. (A) The association between 6-IS and biological functions, and the clustering map of the biological functions. (B) The clustering heatmap of the 6-IS related biological functions. HCC samples are shown on horizontal axis while risk score is shown increasing from left to right.

# Figures

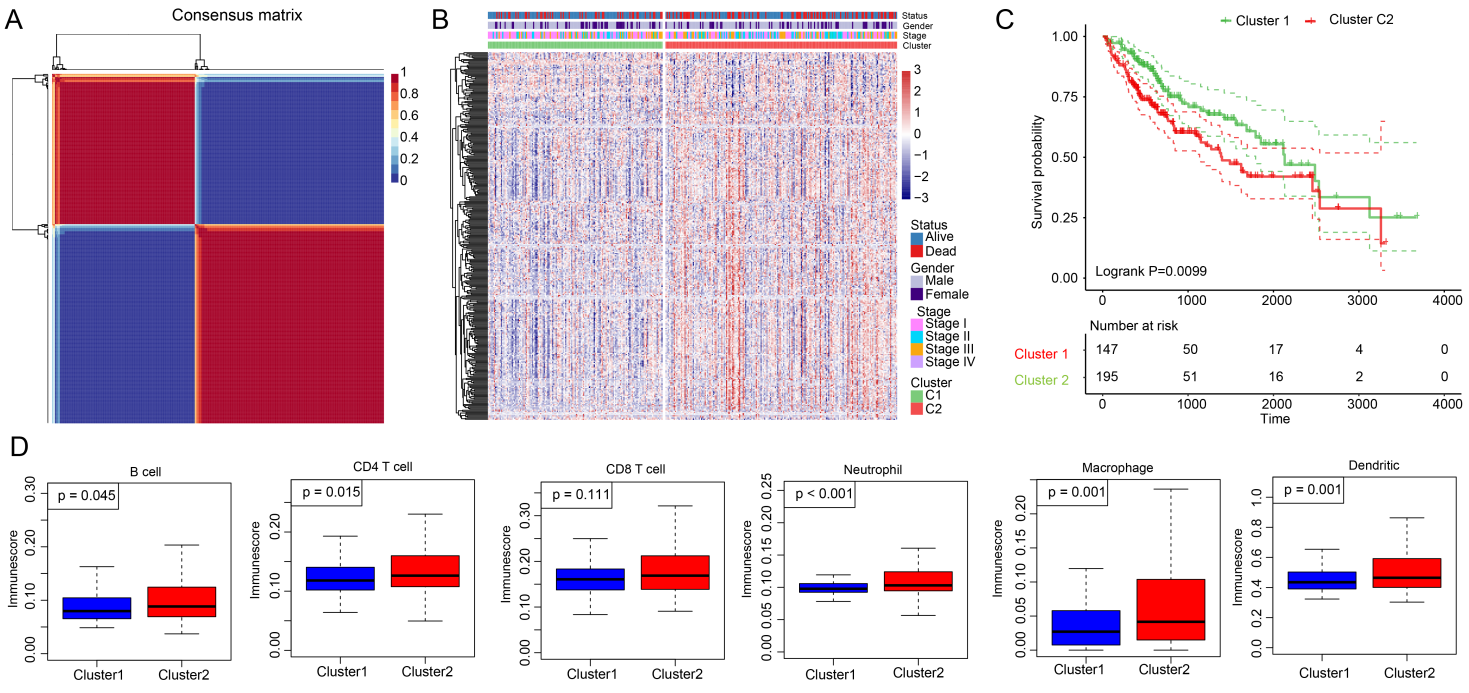
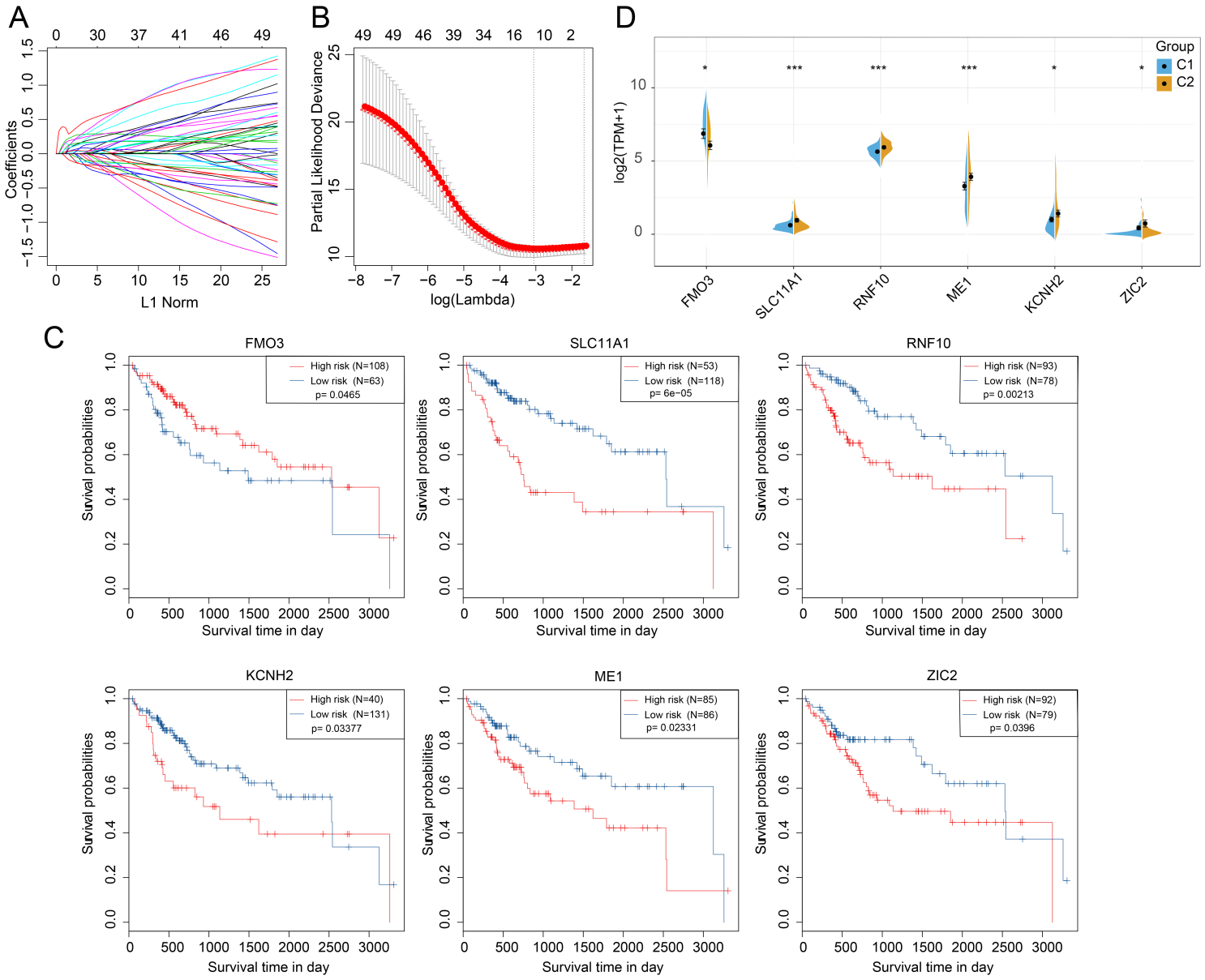


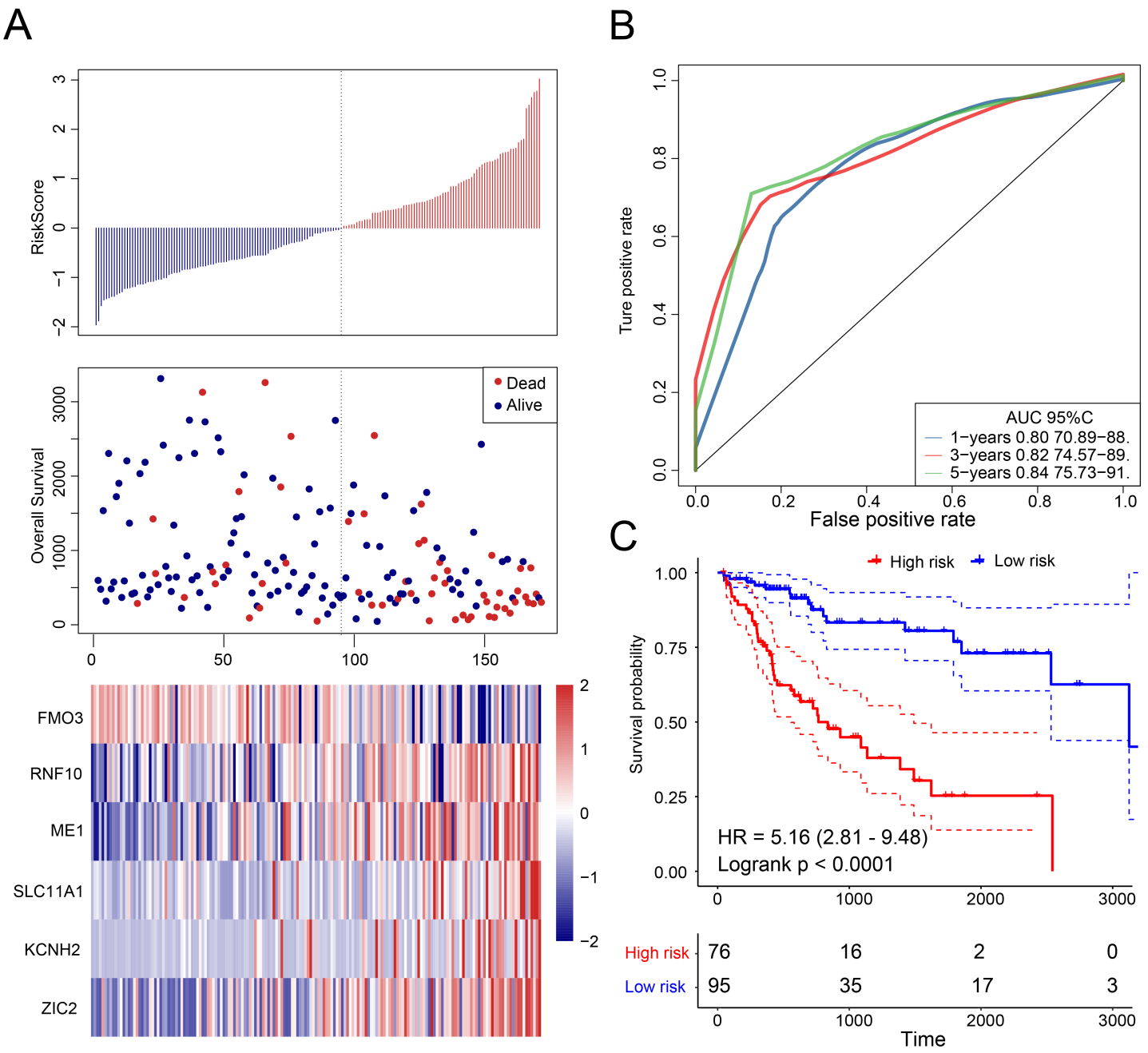

**Figure 1**

Classification of HCC based on IMRGs. (A) NMF algorithm of consensus map of HCC patients for k=2. (B) The cluster heatmap of 324 prognostic related IMRGs in the 2 HCC clusters. (C) KM analysis of overall survival in the 2 HCC clusters. (D) TIMER analysis of immune scores in the 2 clusters.



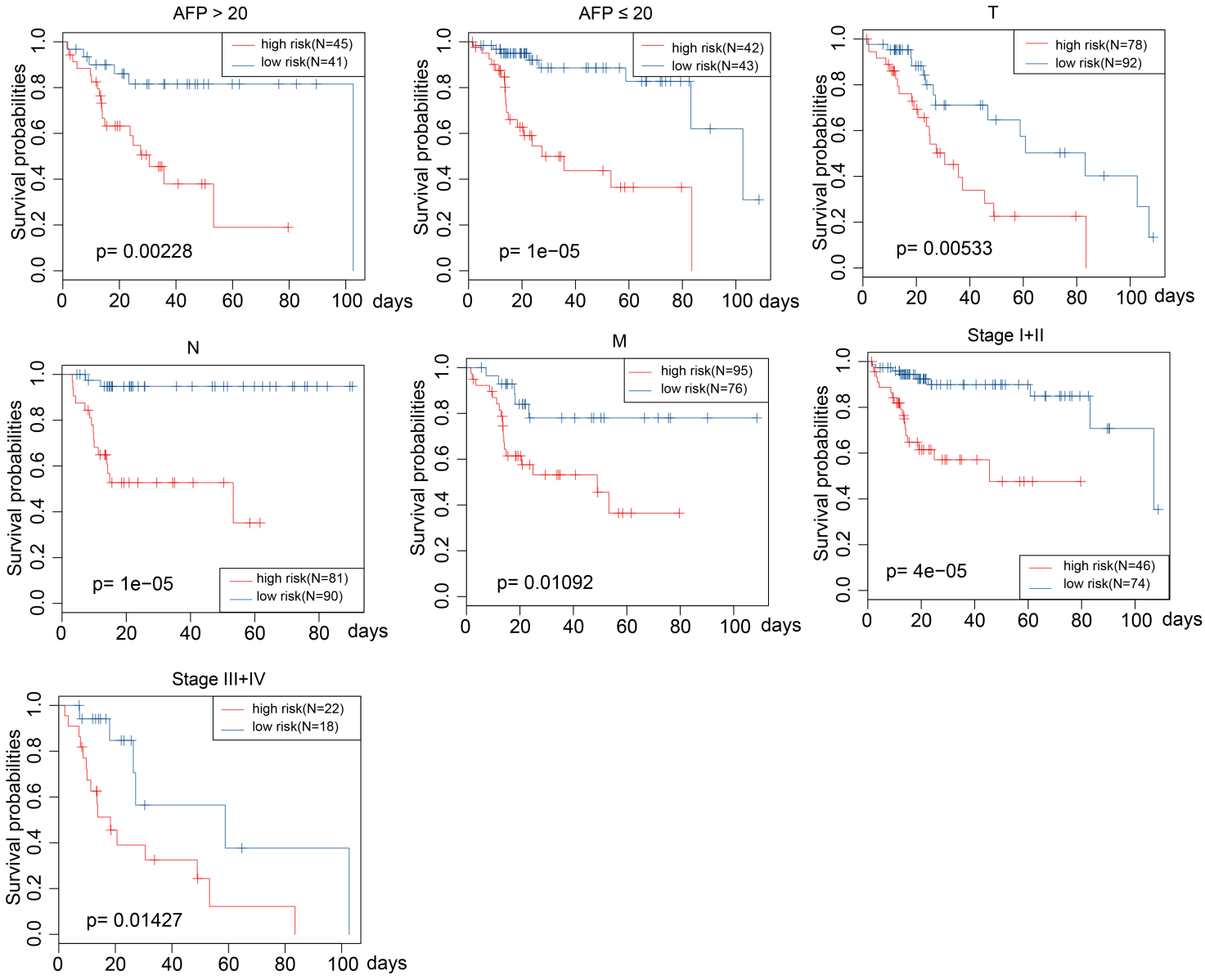

## Figure 2

The selection of IMRGs for construction of a prognostic signature. (A) LASSO analysis of coefficient profiles of the IMRGs and the distribution of the trajectory of each independent IMRG. (B) The confidence intervals under each lambda using 10-fold cross validation. (C) KM analysis of overall survival of HCC patients based on each 6 IMRG. (D) The expression of 6 IMRGs between the 2 clusters.




## Figure 3

Prognostic prediction by 6-IS in the training set. (A) The distributions of the risk score, survival status, and expression of the 6 IMRGs in patients. (B) ROC curve analysis of the 6-IS for 1, 3, 5 years. (C) KM analysis of overall survival of HCC patients based on 6-IS.




## Figure 4

KM analysis of overall survival of HCC patients based on 6-IS when these patients was classified by clinical features, including AFP, TNM, and stage.

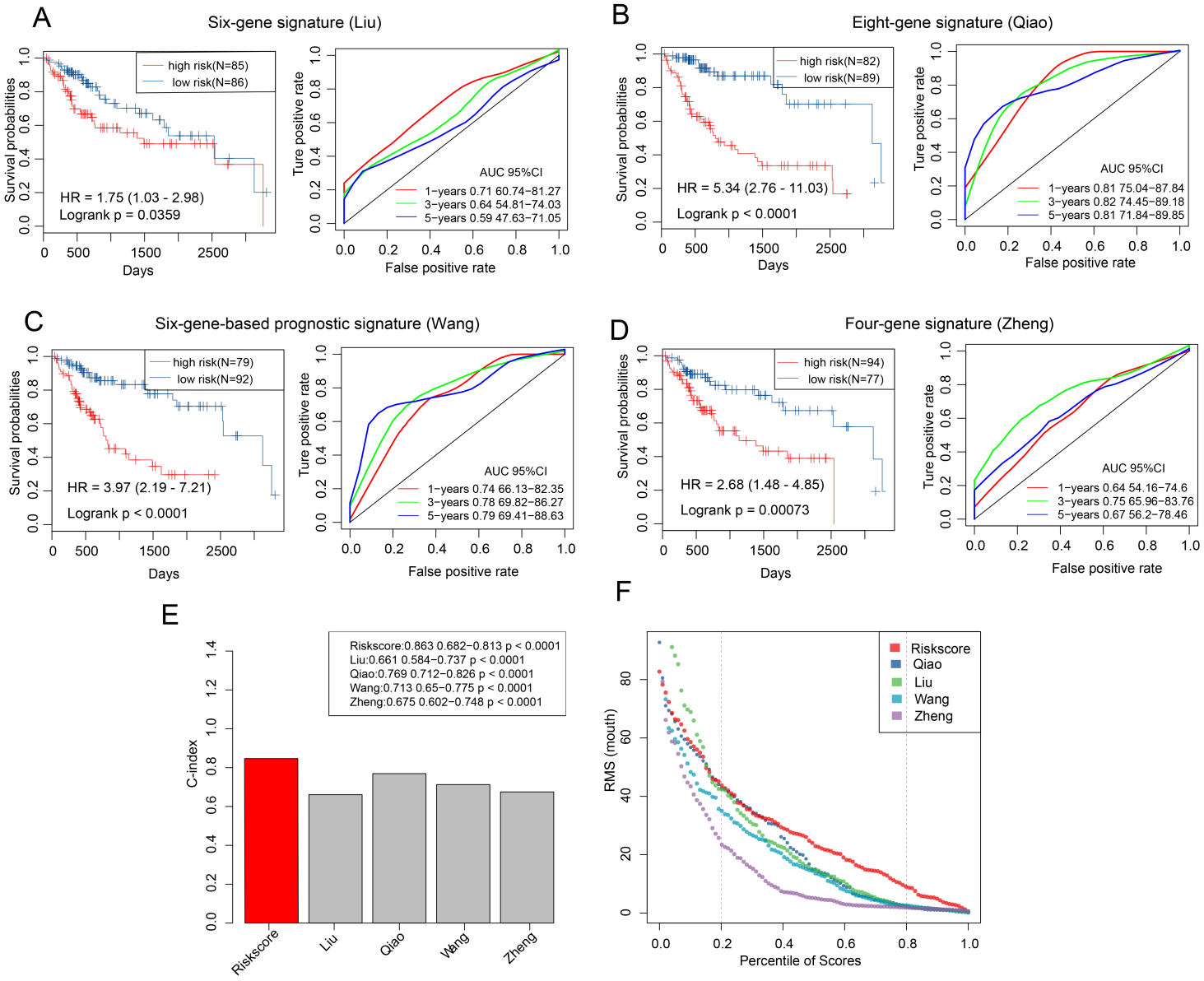



**Figure 5**

Comparison between 6-IS and external models. ROC curve analysis of the 6-IS for 1, 3, 5 years and KM analysis of overall survival in HCC patients according to the (A) six genes signature, (B) eight genes signature, (C) six genes-based prognostic signature, and (D) four genes signature. (E) C-index analysis of 6-IS and other four external models. (F) The RMST analysis of 6-IS and other four external models.

# Supplementary Files

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

- FigureS6.tif
- FigureS7.tif
- FigureS3.tif

- FigureS4.tif
- FigureS5.tif
- FigureS2.tif
- FigureS1.tif
- Tablesupplementary.docx