



# A Four Amino Acid Metabolism-Associated Genes (AMGs) Signature for Predicting Overall Survival Outcomes and Immunotherapeutic Efficacy in Hepatocellular Carcinoma

Lu-sheng Liao<sup>1,2,3</sup> · Zi-jun Xiao<sup>3</sup> · Jun-li Wang<sup>4</sup> · Ting-jun Liu<sup>3</sup> · Feng-die Huang<sup>3</sup> · Yan-ping Zhong<sup>3</sup> · Xin Zhang<sup>3</sup> · Ke-heng Chen<sup>3</sup> · Run-lei Du<sup>5</sup> · Ming-you Dong<sup>1,2,3</sup>

Received: 30 January 2023 / Accepted: 10 August 2023 / Published online: 1 September 2023  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

Metabolites are important indicators of cancer and mutations in genes involved in amino acid metabolism may influence tumorigenesis. Immunotherapy is an effective cancer treatment option; however, its relationship with amino acid metabolism has not been reported. In this study, RNA-seq data for 371 liver cancer patients were acquired from TCGA and used as the training set. Data for 231 liver cancer patients were obtained from ICGC and used as the validation set to establish a gene signature for predicting liver cancer overall survival outcomes and immunotherapeutic responses. Four reliable groups based on 132 amino acid metabolism-related DEGs were obtained by consistent clustering of 371 HCC patients and a four-gene signature for prediction of liver cancer survival outcomes was developed. Our data show that in different clinical groups, the overall survival outcomes in the high-risk group were markedly low relative to the low-risk group. Univariate and multivariate analyses revealed that the characteristics of the 4-gene signature were independent prognostic factors for liver cancer. The ROC curve revealed that the risk characteristic is an efficient predictor for 1-, 2-, and 3-year HCC survival outcomes. The GSVA and KEGG pathway analyses revealed that high-risk score tumors were associated with all aspects of the degree of malignancy in liver cancer. There were more mutant genes and greater immune infiltrations in the high-risk groups. Assessment of the three immunotherapeutic cohorts established that low-risk score patients significantly benefited from immunotherapy. Then, we established a prognostic nomogram based on the TCGA cohort. In conclusion, the 4-gene signature is a reliable diagnostic marker and predictor for immunotherapeutic efficacy.

**Keywords** Hepatocellular carcinoma · Amino acid metabolism · Gene signature · Overall survival · Immunotherapy · Prognosis

Extended author information available on the last page of the article

## Abbreviations

aDC	Activated dendritic cell
AFP	Alpha fetoprotein
AMGs	Amino acid metabolism-related genes
APC	Antigen-presenting cell
AUC	Area under the curve
CCR	Cytokine–cytokine receptor
CI	Confidence interval
DEGs	Differentially expressed genes
FDR	False discovery rate
HCC	Hepatocellular carcinoma
HLA	Human leukocyte antigen
HR	Hazard ratio
ICGC	International Cancer Genome Consortium
iDC	Immature dendritic cell
KEGG	Kyoto Encyclopedia of Genes and Genomes
LASSO	Least absolute shrinkage and selection operator
OS	Overall survival
GO	Gene ontology
PCA	Principal component analysis
pDC	Plasmacytoid dendritic cell
ROC	Receiver operating characteristic
ssGSEA	Single-sample gene set enrichment analysis
TCGA	The Cancer Genome Atlas
Tfh	T follicular helper cell
TIL	Tumor-infiltrating lymphocyte
t-SNE	T-distributed stochastic neighbor embedding

## Introduction

Mutations in amino acid metabolism-related genes are reported to promote the tumorigenesis and metastasis of various cancers (Pavlova and Thompson 2016; Jain et al. 2012). Amino acid metabolism is a crucial mediator involved in cell growth, proliferation, and in the maintenance of cell redox, genetic as well as epigenetic states (Fu et al. 2023; Li and Zhang 2023). Amino acids are associated with metabolisms of lipids, glucose, and nucleotides, which are very important for cancer cell proliferation and metastasis (Li and Zhang 2016; Vettore et al. 2020; Zhu et al. 2022). Various genes encoding metabolic enzymes have been implicated in tumorigenesis. For instance, glutaminase 2 (GLS2) is reported to enhance tumor drug resistance via a p53-mediated signaling pathway (Mates et al. 2020).

Immunotherapy can enhance the capacity of the immune system to detect and clear cancer cells. Recently, immunotherapy has emerged as a highly effective treatment option for various cancers, such as melanoma, lung cancer, head and neck cancers, and renal cell carcinoma (RCC) (Reda et al. 2022; Fasano et al. 2022; Yao

et al. 2022; Huang and Zappasodi 2022). However, it is not clear how immunotherapy affects amino acid metabolism in cancer.

The overall survival outcomes of hepatocellular carcinoma (HCC) patients vary significantly across the world, with a 5-year survival rate of 19% in the USA (Dasgupta et al. 2020). The risk factors for HCC include chronic HBV/HCV infections, alcohol consumption, exposure to aflatoxins, and non-alcoholic fatty liver disease (Yang et al. 2019). Chemotherapy and radiotherapy have markedly improved HCC patients' survival outcomes. Pathologically, HCC is a highly heterogeneous disease, whose treatment has been documented at the interpatient, intertumoral and intratumoral levels (Torrecilla et al. 2017; Alawyia and Constantinou 2023). Currently, there is no effective immunotherapy developed for liver cancer. Considering the limited HCC treatment strategies, effective markers are urgently needed for the identification of HCC patients who are likely to benefit from immunotherapy.

In this study, we systematically and comprehensively investigated the characteristics of gene sets related to amino acid metabolism in liver cancer. First, we demonstrated that gene sets associated with amino acid metabolism could stratify liver cancer based on clinical and molecular features. Next, we developed a 4-gene signature that is based on amino acid metabolism-related genes using liver cancer RNA-seq data from TCGA and validated it with a liver cancer dataset from ICGC. Our results show that the 4-gene signature exhibited an ability to accurately predict the prognosis and responses to immunotherapy, which offer further insight into individualized treatment of liver cancer.

## Materials and Methods

### Data Collection and Processing

The level 3 RNA expressions as well as the clinical data associated with 50 normal liver tissues and 374 HCC tissues were retrieved from the TCGA database (<https://portal.gdc.cancer.gov/repository>). The RNA-seq data and associated clinical data for 231 liver cancer cases were downloaded from the ICGC database (<https://dcc.icgc.org/projects/LIRI-JP>). Then, the “limma” R package was used to normalize read count values using the scale method. Since the ICGC and TCGA datasets are publicly available, the ethical approval requirement was waived. The gene sets associated with amino acid metabolism (REACTOME\_METABOLISM\_OF\_AMINO\_ACIDS\_AND\_DERIVATIVES, Supplementary Table S1) were acquired from the Molecular Signatures Database v7.1 (MSigDB).

### Identification of Differentially Expressed AMGs

The “Limma” R package was used for identification of differentially expressed AMGs, with  $\log_2$ fold change (FC)  $> 1$  and FDR  $< 0.05$  as cut-offs. The differentially expressed AMGs were then subjected to KEGG pathway and GO analyses via

“clusterprofiler” R package to identify the enriched terms and pathways (Wu et al. 2021a, b).

### **Molecular Subgroup Classification by Consistent Clustering**

The prognostic-related AMGs were screened by univariate regression analysis and used for unsupervised clustering analysis. “ConsensusClusterPlus” R package (Wilkerson and Hayes 2010) was used to determine the cluster number of HCC samples from TCGA using the continuous clustering algorithm. The Euclidean square distribution metric, K-Merge cluster algorithm, and KM Clustering algorithm were used for sample classification into  $k$  clusters ( $k=2:9$ ), with 100 iterations and around 80% of samples selected in every iteration. Pheatmap in R was used to visualize the results on a consistency matrix heatmap. Correlation cumulative distributor (CDF) and Delta region maps were used to establish the optimal cluster number (Wilkerson and Hayes 2010). The maximum number of clusters was established based on the criteria: high consistency in the cluster, low variability coefficient, and no significant increase under the CDF curve.

### **Establishment and Verification of a Prognostic AMGs Signature**

Univariate Cox regression analyses were conducted to identify prognostic genes related to amino acid metabolism based on these DEGs ( $p < 0.05$ ). Then, the LASSO regression analysis was performed to further screen hub genes using the “glmnet” R package (Tibshirani 1997). We conducted 1000 replacement samples of the dataset and chose markers whose repetition frequency  $> 900$ . Finally, the multivariate Cox regression analysis was performed to develop a certain gene signature and to define the importance of gene expression. The regression coefficient was derived from multivariate Cox regression analysis and the prognostic index (PI) determined via the formula: (mRNA1 coefficient \* expressions of mRNA1) + (mRNA2 coefficient \* expressions of mRNA2) + ... + (mRNAn coefficient \* expression of mRNAn). Using the median risk score, patients were classified into the high- and low-risk groups. With regards to gene expressions in the signature, PCA analysis were done using “prcomp” R package. In addition, t-sne analysis was performed using “rtsne” package in R to assess the distributions of various groups. “Survivalroc” R package was used for time-dependent ROC curve analysis to assess the gene signature’s predictive ability. At last, we analyzed the correlations of hub genes in HCC with Xiantao love online tool (<https://www.xiantao.love/products>).

### **Independence of the AMG-Based Model from Other LIHC Patients’ Clinical Features**

Based on other clinical features (grade, age, TNM stage, and stage) of LIHC patients, univariate and multivariate Cox regression analyses were conducted to check if the prognostic model is an independent variable. To confirm the prognostic significance of the predictive model, LIHC samples were randomized into two



groups depending on various clinical characteristics. Patients were separately classified into the following subgroups: grade I/II, grade III/IV, stage I/II, stage III/IV, age < 65, age ≥ 65, T1–T2, and T3–T4 subgroups. Then, survival outcome analysis was performed to verify the independent prognostic significance of the gene signature in particular subtypes. The ideal cut-off value of the risk score was established using the `surv_cutpoint` function of “survminer” R package.

### Gene Set Variation Analysis (GSVA) and KEGG Pathway Analysis

To establish the biological pathways or processes that were enriched in different risk groups, we used GSVA to assess differences in bioprocess activities and signaling pathways in various risk groups (Hanzelmann et al. 2013). To this end, we acquired the “c2.cp.kegg.v7.2.symbols” file from MSigDB (<https://www.gsea-msigdb.org/gsea/msigdb/>) and performed GSVA analysis using “GSVA” R package to identify significantly enriched pathways (adjusted  $p < 0.05$ ). Next, DEGs in low-risk vs high-risk groups ( $|\log_2FC| \geq 1$ ,  $FDR < 0.05$ ) were identified and KEGG pathway analyses done using “clusterProfiler” R package (Wu et al. 2021a, b) to predict their functions.  $p < 0.05$  indicated the biological functions in which the DEGs were significantly enriched.

### Analysis of Tumor Mutation Burden

The liver cancer patients’ mutation burden data were retrieved from TCGA database (<https://www.cancer.gov/tcga/>) and the total non-synonymous mutations counted to determine the mutational burden. The “Maftools” R package was used for driver gene identification (Mayakonda et al. 2018), with  $p < 0.05$  indicating genes that were significantly differentially mutated in low- vs high-risk groups. The top 20 driver genes with the highest mutation frequencies were evaluated further. The copy number increase and loss burden was calculated at the arm and focal levels between high- and low-risk patient groups as previously described (Shen et al. 2019).

### Assessment of Immune Cell Type Fractions

The activities of 13 immune-associated pathways and infiltration levels of 16 types of immune cells were evaluated by ssGSEA (Rooney et al. 2015) using the “GSVA” R package. The annotated gene set file is as shown in Supplementary Table S2.

### Prediction of Patients’ Responses to ICI and Chemotherapy

The PD-1, CTLA4, PD-L1, and LAG3 immune checkpoints were used to evaluate the associations between risk scores and immunotherapeutic efficacies (Charoentong et al. 2017). The independent datasets (IMvigor210, GSE135222, and GSE91061) were analyzed to assess the ability of the gene signature to predict immunotherapeutic responses. Clinical information and expression data in the IMvigor210 dataset were retrieved from <http://research-pub.gene.com/IMvigor210CoreBiologies>. The expression data as well

as clinical information for GSE91061 and GSE135222 datasets were acquired from GEO (<https://www.ncbi.nlm.nih.gov/geo/>).

### Estimation of Drug Responses

The sensitivity of low- and high-risk groups to chemotherapy was assessed as previously described (Villanueva 2019). Briefly, the drug sensitivity datasets (CTRP and PRISM) were obtained from CTRP (<https://portals.broadinstitute.org/ctrp>) and PRISM (<https://depmap.org/portal/prism/>). The CCLE expression data for drug sensitivity analyses were extracted from CTRP and PRISM datasets. Significantly different drugs between the low- and high-risk groups were indicated by  $\log_2FC > 0.10$  in the CTRP and PRISM datasets, and the Pearson correlation coefficient was  $< -0.30$  in the CTRP dataset. Pearson correlation coefficient was  $< -0.35$  in the PRISM dataset.

### Establishment and Evaluation of a Predictive Nomogram

A nomogram was build based on gender, stage, grade, age, and risk score as described before (Iasonos et al. 2008). The receiver operating characteristic (ROC) curve (AUC), 1-, 3-, and 5-year calibration curves, as well as decision curve analyses (DCA) were used to assess the nomogram's predictive accuracy and discriminatory capacity (Vickers and Elkin 2006).

### Drug Susceptibility Analysis

To screen for FDA-approved drugs and clinical trial data, the drug susceptibility based on hub genes was analyzed with the CellMiner database. We mainly analyzed the expressions of four hub genes and drug sensitivity. The correlations between hub genes expression levels and drug susceptibility were conducted by spearman correlation analysis, and the top 16 drugs with the most significant correlations were selected.

### Transcription Factor (TF) Regulatory Network Analysis

The regulatory networks between transcription factors (TFs) and hub genes were conducted with the online analysis tool, NetworkAnalyst (<https://www.networkanalyst.ca/NetworkAnalyst/>). The hub genes (TXNRD1, PSMD14, SMOX, and EEF1E1) were input and the human transcription factor targets were derived from the JASPAR TF binding site profile database to establish the TF–gene interactions regulatory network.

### Statistical Analysis

Gene expressions in normal vs tumor tissues were compared using the student's *t* test, while the chi-square test was used to compare proportional differences. The ssGSEA scores and differences in immune cells or immune pathways in the low- vs high-risk group were compared using the Mann–Whitney U test.

Comparisons of survival differences in the low- vs high-risk groups were conducted by Kaplan–Meier analysis. Independent predictors were identified by univariate and multivariate Cox regression analyses. The R (version 4.0.1) or SPSS (version 22.0) software were used for analyses. Unless otherwise stated,  $p < 0.05$  was set as the threshold for statistical significance.

## Results

### Prognostic and Amino Acid Metabolism-Associated DEGs in the TCGA Dataset

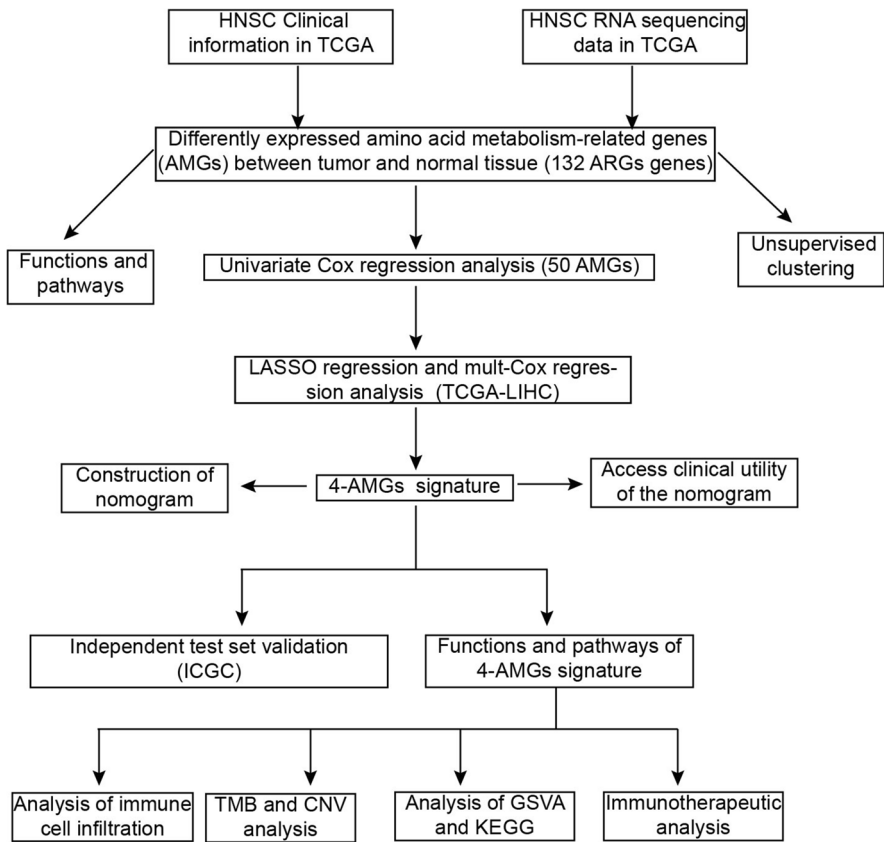
This study involved 374 liver cancer patients in the TCGA dataset and 231 liver cancer patients in the ICGC (LIRI-JP) dataset, and the flow chart is shown in Fig. 1. Analysis of these datasets using the LIMMA package in R identified 374 AMGs, of which, 132 AMGs were differentially expressed (23 suppressed and 109 elevated) between liver cancer tissues and normal tissues (Fig. S1A, B). GO analysis revealed that these AMGs were enriched in responses to mRNA catabolic processes, protein localization of the endoplasmic reticulum, and protein targeting to the membrane (Fig. S1C). KEGG pathway analysis revealed that the modular genes were enriched in the metabolism of arginine, proline, alanine, aspartate, glutamate, cysteine, methionine, tryptophan, amino acid biosynthesis, and the selenocompound metabolism signaling pathways (Fig. S1D).

### Identification of Distinct Molecular Clusters Based on Prognostic AMGs

Consistent clustering analysis based on the 132 DEGs was done using Consensus-ClusterPlus package in R. First, a consensus matrix graph was constructed and the 132 DEGs assigned to 2, 3, and 4 clusters to evaluate the clustering quantity (Fig. S2A–C). As shown in Fig. S2D,  $k=4$  was associated with good clustering. CDF delta area curve analysis revealed that the area was stable with a cluster number of 4 (Fig. S2E). PCA showed that the 4 clusters could be better distinguished (Fig. S2F). The cluster subgroup was significantly correlated with gender (Fig. S2G). Kaplan–Meier analysis revealed that cluster 1 had markedly better survival probability relative to the other 3 clusters, while cluster 2 had the worst survival probability (Fig. S2H).

### The TCGA Dataset-Based Prognostic Model

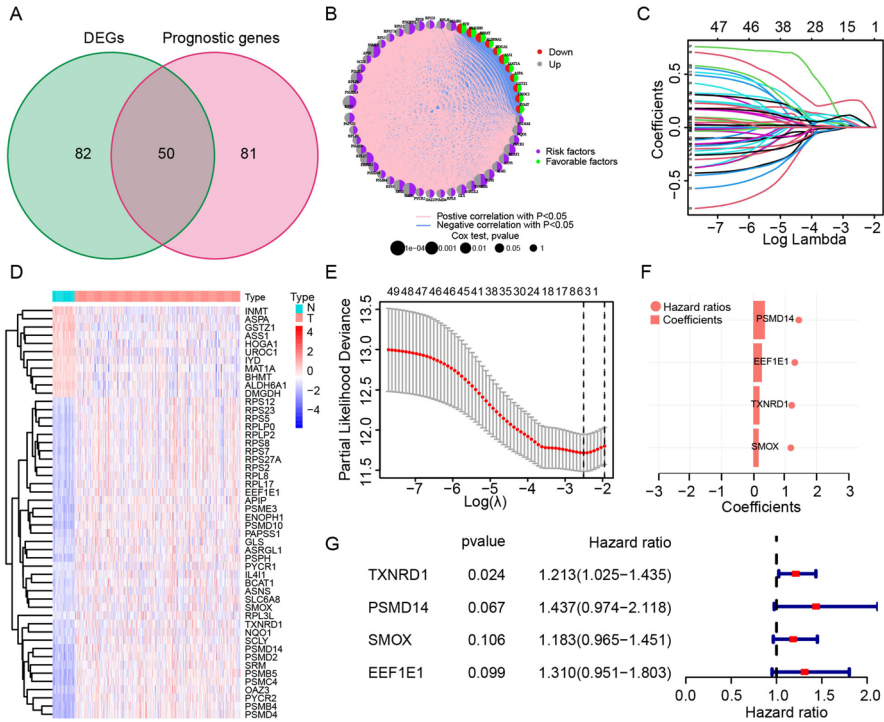
To establish an amino acid metabolism-associated gene signature, we obtained 50 OS-associated amino acid metabolism-related DEGs based on univariate Cox regression analysis of the TCGA dataset, 11 of which are favorable factors of liver cancer (Fig. 2A, B). A total of 50 AMGs were differentially expressed between liver cancer tissues and normal tissues (Fig. 2D). Then, LASSO regression analysis was performed and 30 genes were reserved (Fig. 2C, E). Finally, 4 AMGs were selected using multiple stepwise Cox regression analysis and used to build a



**Fig. 1** Flowchart showing the process of analyzing AMG in liver cancer

predictive model (Table 1, Fig. 2F, G). The risk score for every HCC patient was determined using the formula:  $e (0.193 * \text{expression levels of TXNRD1} + 0.362 * \text{expression levels of PSMD14} + 0.168 * \text{expression levels of SMOX} + 0.270 * \text{expression level of EEF1E1})$ .

The patients were assigned into low-risk ( $n = 183$ ) and high-risk ( $n = 182$ ) groups based on the median risk score (Fig. 3A). Next, the prognostic gene expression patterns in the high- vs low-risk groups of the TCGA training set were visualized with a heatmap (Fig. 3B). Time-dependent ROC curve analysis of the risk scores capacity for OS prediction revealed respective AUC values of 0.757, 0.678, and 0.669 at 1, 2, and 3 years (Fig. 3C). High-risk score patients had a high chance of dying earlier, relative to those with low-risk scores (Fig. 3D). Then, PCA and t-SNE analysis showed that patients in various risk groups were scattered in two directions (Fig. 3E, F). Kaplan–Meier analysis revealed that the high-risk patient group had markedly worse OS and PFI outcomes, relative to low-risk patients (Fig. 3G, I). However, DFI did not differ

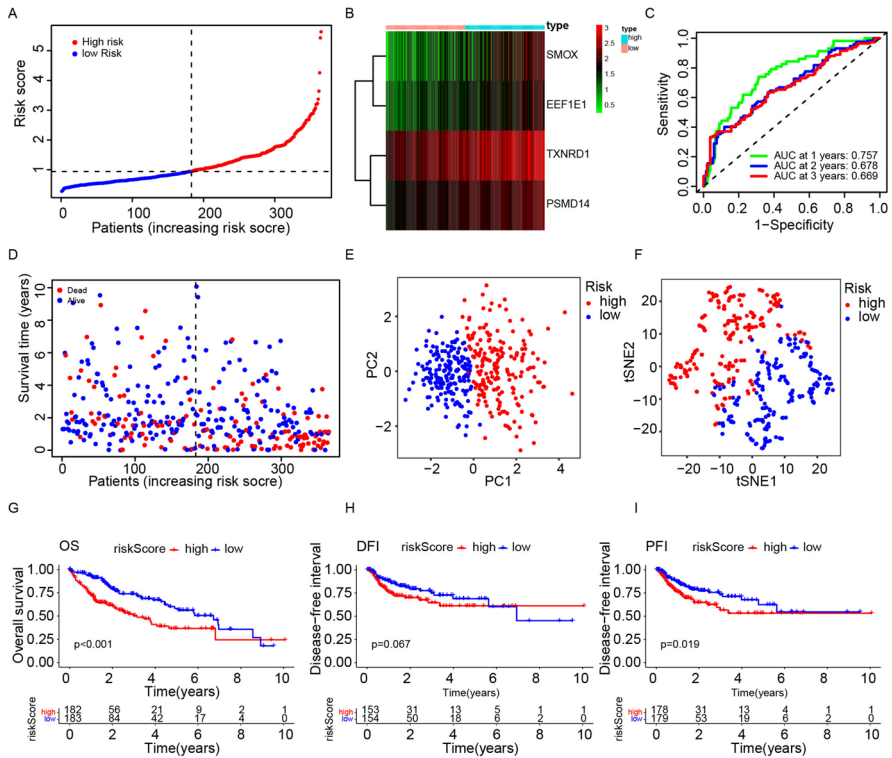


**Fig. 2** Identification of amino acid metabolism-associated genes in the TCGA dataset. **A** Venn diagram of DEGs and survival-related genes. **B** Correlation network map of candidate genes. **C** Distribution map of LASSO coefficients for 50 amino acid metabolism genes. **D** Heatmap of 26 differentially expressed genes significantly associated with amino acid metabolism. **E** Six candidate genes were identified using LASSO regression analyses with ten-fold cross-validation. **F** Multivariate Cox regression coefficient distribution of 4 amino acid metabolism-associated genes. **G** Multivariate Cox regression forest plot of the prognostic significance of the 4 amino acid metabolism-related genes

**Table 1** Identification of prognosis-related hub AMGs using multivariate Cox regression analysis

Gene	Coef	HR	HR.95L	HR.95H	<i>p</i> value
TXNRD1	0.193153	1.213069	1.025215	1.435344	0.024446
PSMD14	0.362236	1.436539	0.974382	2.1179	0.06741
SMOX	0.168397	1.183406	0.965026	1.451206	0.10568
EEF1E1	0.269667	1.309529	0.95098	1.803261	0.098525

markedly in the high- vs low-risk groups (Fig. 3H). The risk score was markedly associated with grade, stage, T stage, as well as the survival status of LIHC patients (Table 2; Fig. S3).



**Fig. 3** Prognostic assessment of the 10-gene signature model in the TCGA dataset (**A**). Distribution characteristics and median risk score values in TCGA dataset. **B** Heatmap analysis indicating differences in levels of the 4-gene signature in high- vs low-risk groups. **C** AUC of the time-dependent ROC curve was used to evaluate the risk scores' prognostic value in the TCGA dataset. **D** Survival rate of patients in low and high-risk groups. **E**, **F** PCA and t-SNE analyses of the risk model's accuracy in TCGA dataset. **G–I** Kaplan–Meier analyses of DFI, OS, and PFI in the low- and high-risk groups in the TCGA dataset

### Validity of the 4-Gene Signature in the ICGC Dataset

To assess the robustness of the developed model using the TCGA dataset, the ICGC dataset patients were classified into the high- and low-risk groups based on median risk scores and an expression profile heatmap of prognostic risk genes in the high- vs low-risk group generated (Fig. 4A–C). As with findings from the ICGC cohort, t-SNE and PCA analyses verified that patients in the 2 groups were scattered in distinct directions (Fig. 4D, E). Respective AUC value of the 4-gene signature was 0.697, 0.693, and 0.678 at 1, 2, and 3 years (Fig. 4F). Relative to gender, age, and stage, the 4-gene signature had higher AUC value at 1 year (Fig. 4G). The high-risk group had low survival outcomes, relative to the low-risk group (Fig. 4H,  $p = 0.001$ ).

**Table 2** Relationship between patients' clinical features and the risk score in the training set

Variables	Group	High risk	Low risk	chi	<i>p</i> value
Age	< 65	111 (60.99%)	105 (57.38%)	0.3546	0.5515
	≥ 65	71 (39.01%)	78 (42.62%)		
Gender	Female	51 (28.02%)	68 (37.16%)	3.0632	0.0801
	Male	131 (71.98%)	115 (62.84%)		
Grade	G1	15 (8.24%)	40 (21.86%)	20.9648	<b>1.00E–04</b>
	G2	85 (46.7%)	94 (51.37%)		
	G3	76 (41.76%)	43 (23.5%)		
	G4	6 (3.3%)	6 (3.28%)		
Stage	I	67 (36.81%)	108 (59.02%)	18.7608	<b>3.00E–04</b>
	II	56 (30.77%)	32 (17.49%)		
	III	57 (31.32%)	41 (22.4%)		
	IV	2 (1.1%)	2 (1.09%)		
T	T1	69 (37.91%)	111 (60.66%)	19.9241	<b>2.00E–04</b>
	T2	58 (31.87%)	35 (19.13%)		
	T3	46 (25.27%)	32 (17.49%)		
	T4	9 (4.95%)	4 (2.19%)		
	Unknown	0 (0%)	1 (0.55%)		
Radiation	No	176 (96.7%)	179 (97.81%)	0.1085	0.7418
	Yes	6 (3.3%)	4 (2.19%)		
Survival_status	Alive	105 (57.69%)	130 (71.04%)	6.5176	<b>0.0107</b>
	Dead	77 (42.31%)	53 (28.96%)		

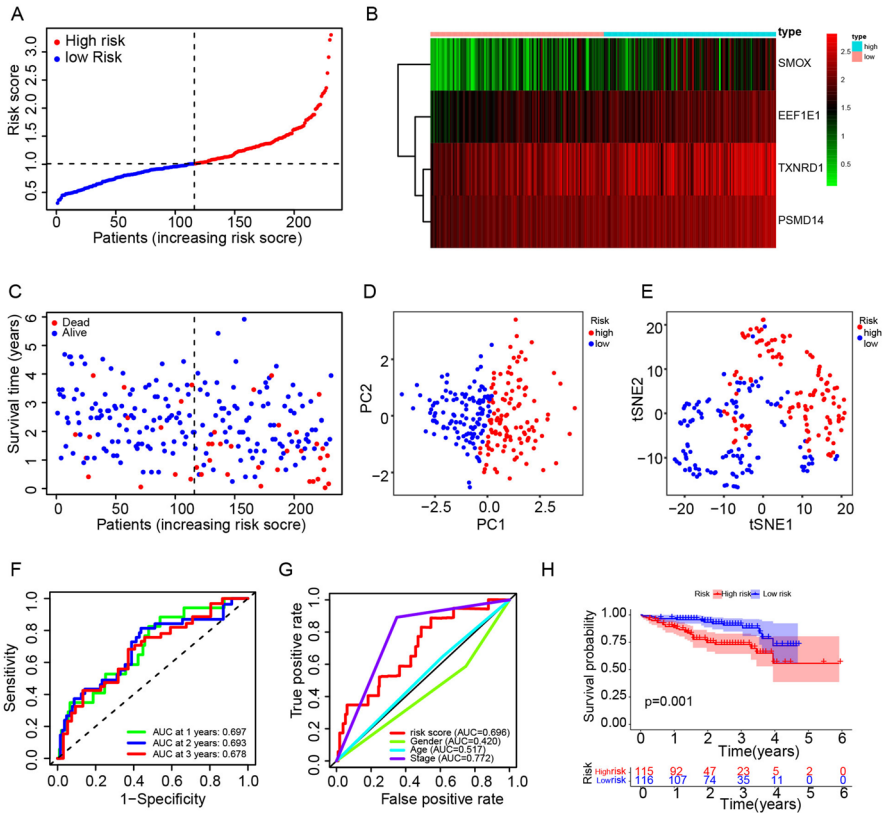
The bold values indicate  $p < 0.05$

### Prognostic Significance of the 4-Gene Signature

To assess the prognostic value of the model in various clinicopathological settings, the clinical variables and samples were randomized into 2 subgroups with regards to TNM stage, age, grade, and gender. Patients in the subgroups were then assigned into high- and low-risk groups and the best cut-off value of the prognostic model is used as the cut-off. KM survival analysis of the 8 subgroups (age < 65 years old, grade I–II, age ≥ 65 years old, grade III–IV, sex, stage I–II, III, stage IV, T0–T2 and T3–T4) revealed that except for the female subgroup, the prognostic models markedly correlated with LIHC patient survivals outcomes (Fig. 5).

### Immune Microenvironments in High- vs Low-Risk Groups

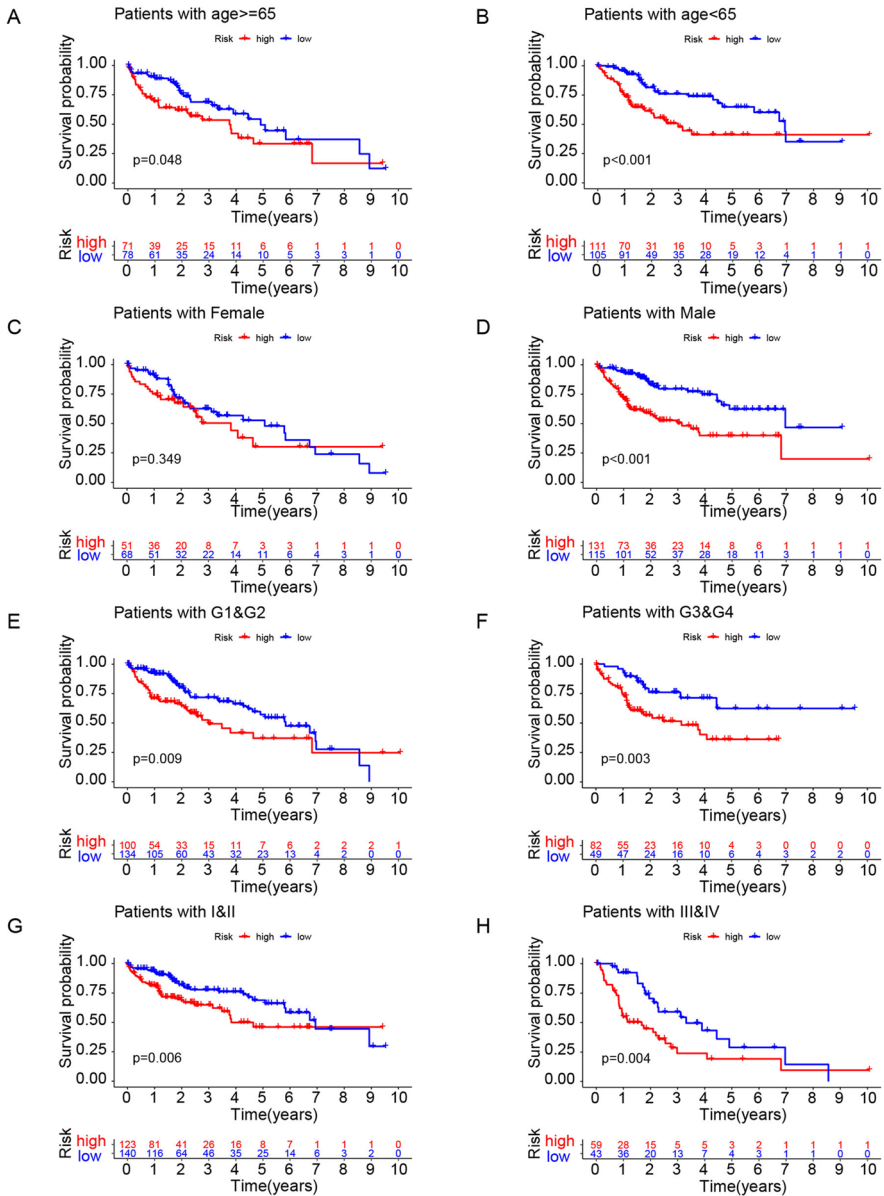
To analyze the function of the risk model, enrichment scores of 16 immune cell types and activities of 13 immune-associated pathways in TCGA and ICGC datasets were compared. In the TCGA cohort (Fig. 6A), the high-risk patient subgroup



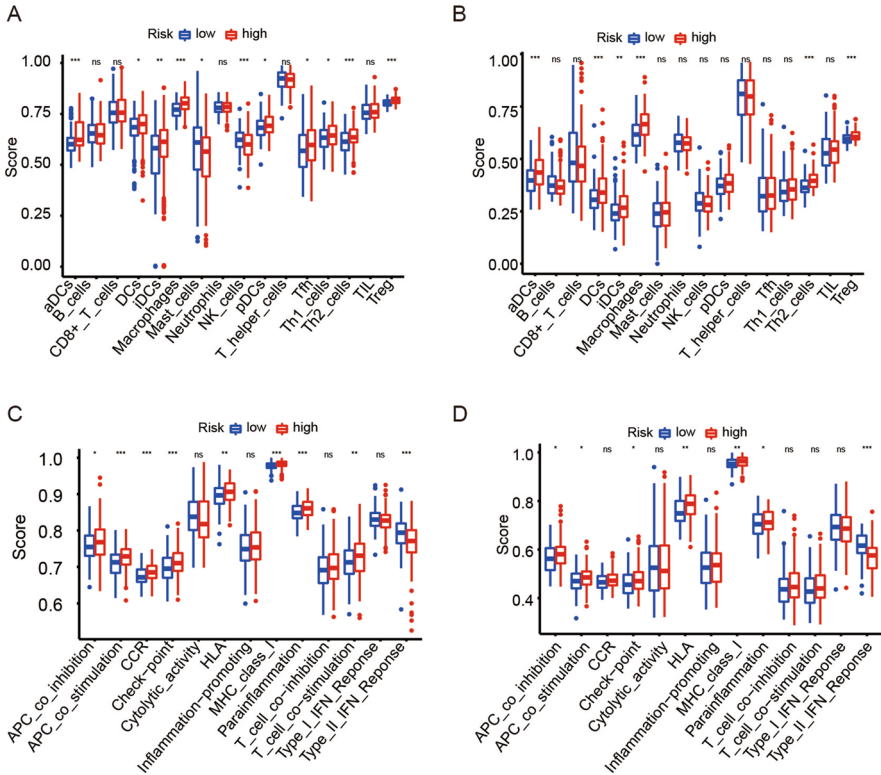
**Fig. 4** Verification of the 4-gene signature in the ICGC dataset. **A** Median risk scores and distribution characteristics in the ICGC dataset. **B** Heatmap analysis of differences in the expression of the 4-gene signature in low- and high-risk groups. **C** Survival rates of high- and low-risk patient groups. **D**, **E** t-SNE and PCA analyses of the model's accuracy on the TCGA dataset. **F** AUC of time-dependent ROC curve was used to evaluate the risk score's prognostic value in the ICGC dataset. **G** AUC of the time-dependent ROC curve was used to assess the prognostic significance of the risk score, gender, age, as well as stage in the ICGC dataset. **H** Kaplan–Meier analysis of OS in the high- and low-risk groups in the ICGC dataset

usually exhibited high levels of immune cell infiltrations, especially for activated dendritic cells (aDC), dendritic cells (DC), induced dendritic cells (iDC), natural killer (NK) cells, macrophages, T helper (Th) cells (Tfh, Th1, and Th2 cells), and regulatory T (Treg) cells. Except for cytolytic activities, inflammation promotion, T-cell co-suppression, type I interferon response pathway, and type II interferon response pathway, the rest of the 7 immune pathways were more active in the high-risk patient group than that in low-risk in TCGA cohort (Fig. 6C). Similar observations were made in the ICGC dataset (Fig. 6B, D). Assessment of associations between expressions of the gene signature, immune cells, and





**Fig. 5** KM survival subgroup analyses for LIHC patients in TCGA dataset based on the 4-gene signature stratified by clinical characteristics. **A** Aged ≥ 65 years. **B** Aged < 65 years. **C** Female. **D** Male. **E** Grade I–II. **F** Grade III–IV. **G** Stage I–II. **H** Stage III–IV

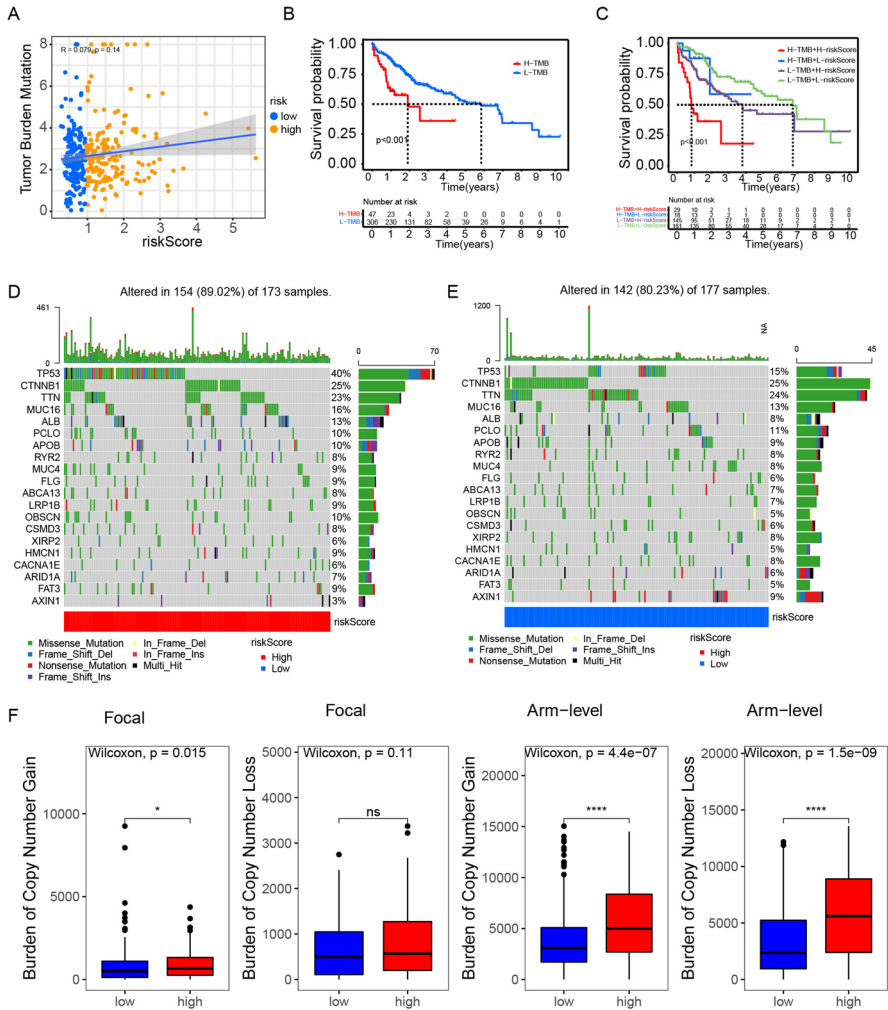


**Fig. 6** Comparison of the immune cells and immune-associated roles between various risk groups in the TCGA dataset (**A, B**) and ICGC dataset (**C, D**). Comparison of difference in immune cell scores (**A, C**) and immune-associated function scores (**B, D**) in TCGA and ICGC datasets. ns = not significant. \*, \*\*, and \*\*\*, respectively, denote  $p < 0.05$ ,  $p < 0.01$ , and  $p < 0.001$

immune-associated pathways revealed that most of them were significantly positively or negatively associated with immune cell infiltrations in cancer tissues (Fig. S4).

### The 4-Gene Signature and TMB

The correlations between TMB and the risk scores were not significant (Fig. 7A). Grouping patients into low- and high-TMB group revealed that low TMB correlated with better OS relative to high TMB (log-rank test,  $p = 0.001$ , Fig. 7B). Given the contraindicated prognostic significance of TMB and the 4-gene signature, we assessed the synergy of these scores for prognostic stratification of TCGA-HCC. Stratified survival analysis showed that TMB status did not interfere with predictions based on the 4-gene signature ( $p < 0.001$ , Fig. 7C). These findings imply that the risk score is a potential predictor that is independent of TMB and an effective evaluator of immunotherapeutic responses.



**Fig. 7** Associations between the 4-gene signature and somatic variations. **A** TMB differences in low- and high-risk groups. Wilcoxon test,  $p < 0.0001$ . **B** Kaplan–Meier analysis of low and high-TMB groups in LIHC dataset.  $p = 0.0067$ , by Log rank test. **C** Kaplan–Meier analysis of LIHC dataset patients stratified using TMB and the 2-gene signature. **D, E** OncoPrint was build using high- and low-risk scores. **F** Focal and broad copy number alterations in low- and high-risk patient groups

Then, we assessed the distribution of somatic variations in TCGA-HCC driver genes in low- and high-risk groups of the TCGA-HCC dataset using the maftools package in R and selected the top 20 driver genes with highest frequencies (Fig. 7D, E). Twenty genes, including TP53 ( $p = 2.63e-07$ ), SPEG ( $p = 0.007$ ), NLRP12 ( $p = 0.007$ ), and DYNC2H1 ( $p = 0.009$ ) differed markedly with regards to somatic variations in low- vs high-risk group (Table 3). To assess

**Table 3** Relationships between risk scores and somatic variations

Genes	H-wild	H-mutation	L-wild	L-mutation	<i>p</i> value
TP53	103 (59.54%)	70 (40.46%)	150 (84.75%)	27 (15.25%)	2.63E-07
SPEG	160 (92.49%)	13 (7.51%)	175 (98.87%)	2 (1.13%)	0.007262
NLRP12	173 (100%)	0 (0%)	168 (94.92%)	9 (5.08%)	0.007652
DYNC2H1	172 (99.42%)	1 (0.58%)	166 (93.79%)	11 (6.21%)	0.009223
KEAP1	160 (92.49%)	13 (7.51%)	174 (98.31%)	3 (1.69%)	0.018763
DNAH7	157 (90.75%)	16 (9.25%)	172 (97.18%)	5 (2.82%)	0.021171
MAGEL2	173 (100%)	0 (0%)	170 (96.05%)	7 (3.95%)	0.023796
TENM1	173 (100%)	0 (0%)	170 (96.05%)	7 (3.95%)	0.023796
DNAH10	163 (94.22%)	10 (5.78%)	175 (98.87%)	2 (1.13%)	0.036019
MYO18B	163 (94.22%)	10 (5.78%)	175 (98.87%)	2 (1.13%)	0.036019
KALRN	167 (96.53%)	6 (3.47%)	177 (100%)	0 (0%)	0.036857
SETD2	165 (95.38%)	8 (4.62%)	176 (99.44%)	1 (0.56%)	0.039294
BSN	165 (95.38%)	8 (4.62%)	176 (99.44%)	1 (0.56%)	0.039294
TRERF1	173 (100%)	0 (0%)	171 (96.61%)	6 (3.39%)	0.042269
CEMIP	173 (100%)	0 (0%)	171 (96.61%)	6 (3.39%)	0.042269
GRM7	173 (100%)	0 (0%)	171 (96.61%)	6 (3.39%)	0.042269
FMN2	171 (98.84%)	2 (1.16%)	167 (94.35%)	10 (5.65%)	0.043786
PIK3CA	172 (99.42%)	1 (0.58%)	169 (95.48%)	8 (4.52%)	0.046413
HECW2	172 (99.42%)	1 (0.58%)	169 (95.48%)	8 (4.52%)	0.046413
EPB41L3	172 (99.42%)	1 (0.58%)	169 (95.48%)	8 (4.52%)	0.046413

the differences in genetic changes in high- and low-risk group, we next assessed their copy number changes and found that the high-risk patient group had a high burden of copy number gain at focal and arm levels and a higher burden of copy number loss at the arm level, relative to low-risk group patients (Fig. 7F).

### KEGG and GSVA Analyses of Low- and High-Risk Groups

To elucidate risk score associated of biological functions, GSVA was used to determine biological differences in high- vs low-risk groups. The analysis results revealed the cell cycle, ubiquitin-mediated proteolysis, spliceosome, and RNA degradation pathways were highly enriched in high-risk groups of ICGC and TCGA cohorts (Fig. S5A), while PRAR, primary bile acid biosynthesis, and linoleic acid metabolism pathways were enriched in the low-risk group (Fig. S5B). Comparisons of the high- and low-risk groups identified 933 and 1772 DEGs ( $\log_2\text{FC} > 1$ ,  $\text{FDR} < 0.05$ ) in ICGC and TCGA cohorts, respectively. KEGG pathway analysis of the DEGs from both cohorts revealed that they were significantly enriched in glycolysis/gluconeogenesis, carbon metabolism, bile secretion, metabolism of xenobiotics by cytochrome P450, fructose as well as mannose metabolism, and drug metabolism pathways ( $p < 0.05$ , Fig. S5C, D). Moreover, DEGs from the TCGA cohort were markedly enriched in retinol metabolism, ECM–receptor interaction, and leishmaniasis pathways ( $p < 0.05$ , Fig. S5C).

## Responses of Patients to ICI

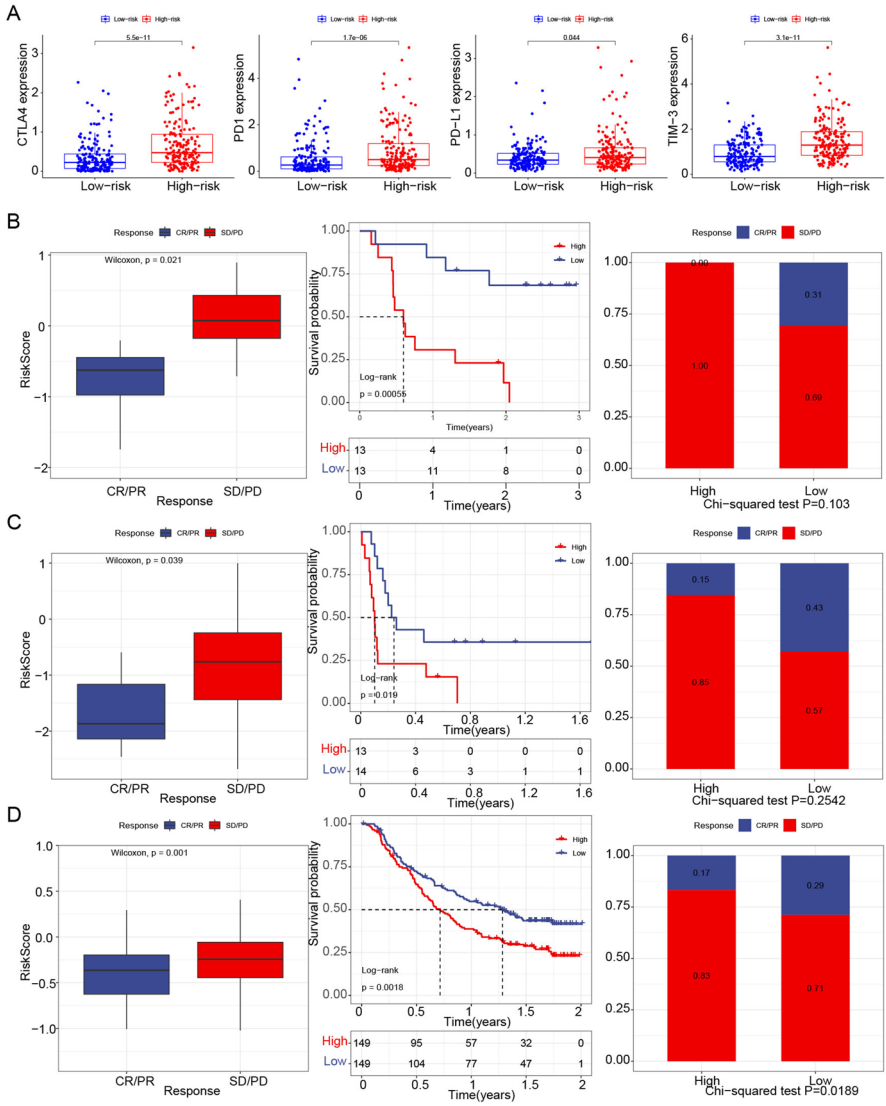
We examined the expressions of immune checkpoints (PD-1, PD-L1, LAG3, and CTLA4) in the risk groups and found that they were markedly upregulated in high-risk patient group, indicating a hot immune microenvironment (Fig. 8A). To assess the value of the risk score in predicting the therapeutic advantage to patients, we analyzed the GSE91061 dataset and classified patients who received immunotherapy by high levels or low levels of the 4-gene signature. Notably, exploration of the IMvigor210 cohort revealed that the high-risk patient group markedly outlives those in low-risk group (log-rank test,  $p=0.0012$ , Fig. 8B). The high-risk IMvigor210 patient cohort had a high probability of dying earlier, relative to those in the low-risk group (log-rank test,  $p=0.00056$ , Fig. 8B). The clinical response rate (including CR/PR and SD/PD) was also higher in the high-risk group (Fig. 8B). Comparable findings were obtained from the validation dataset (GSE135222) and the IMvigor210 cohort (Fig. 8C, D). These data indicate that the 4-gene signature can predict responses to immunotherapy.

## Drug Responses

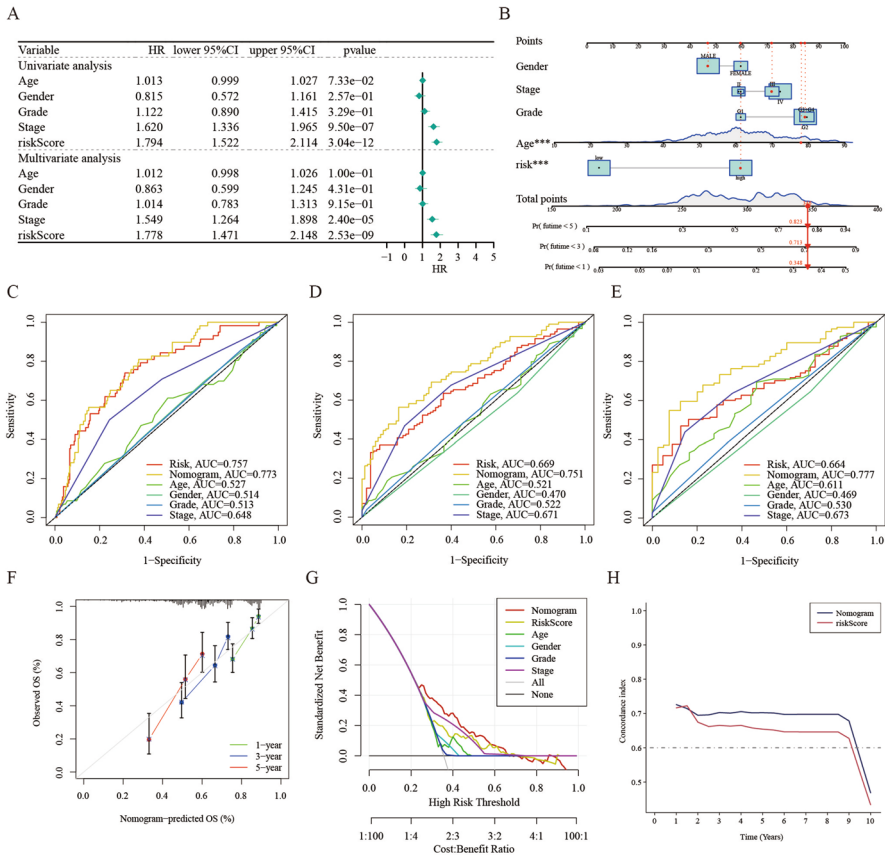
Analyses of PRISM and CTRP drug response datasets were aimed at identifying drug candidates with high sensitivities among high-risk patients. The high-risk patients were highly sensitive to compounds MLN2238, SB-743921, SGK461364, clofarabine, paclitaxel, and BI-2536 from the CTRP dataset and the compounds volasertib, epothilone-b, and ispinesib from the PRISM dataset. These compounds had low AUC values in the high-risk group and correlated negatively with the risk score (Fig. S6A, B), indicating that they may exhibit therapeutic efficacies in high-risk liver cancer patients.

## Establishment and Validation of the Predictive Nomogram

To assess the risk model's clinical prognostic value, we used univariate and multivariate Cox regression analyses to ascertain risk factors with independent prognostic value in LIHC. The risk score and stage were found to be important independent factors (Fig. 9A). To test the risk model in clinical settings, we constructed a nomogram and tested its capacity to predict the OS outcomes on LIHC datasets at 1, 3, and 5 years based on gender, grade, age, stage, and risk group (Fig. 9B). This analyze revealed that relative to the 4-gene signature and the other 4 clinical indicators, the nomogram had a better prognostic ROC value and it could predict OS outcomes for 1-, 3-, and 5-year (Fig. 9C–E). Moreover, the 1-, 3-, and 5-year OS calibration curves for the LIHC data revealed that the nomogram had a good predictive discrimination capacity as well as accuracy (Fig. 9F). Comparisons of net benefits of various models, such as none, risk score, all, nomogram, and clinical indicators, revealed that the nomogram had a higher net income and a wider



**Fig. 8** Role of the 4-gene signature in predicting immunotherapeutic benefits. **A** Levels of immune checkpoints, CD274, PDCD1, CTLA4, and TIM-3 in low- and high-risk groups. **B** Risk scores in various immunotherapy response groups. Kaplan–Meier analysis and rate of clinical response in indicated risk score groups in dataset GSE91061. **C** Risk scores in different immunotherapy response groups. Kaplan–Meier curves and rate of clinical response in different risk score groups in dataset GSE135222. **D** Risk scores in different immunotherapy response groups. Kaplan–Meier curves and rate of clinical response in different risk score groups in dataset IMvigor210



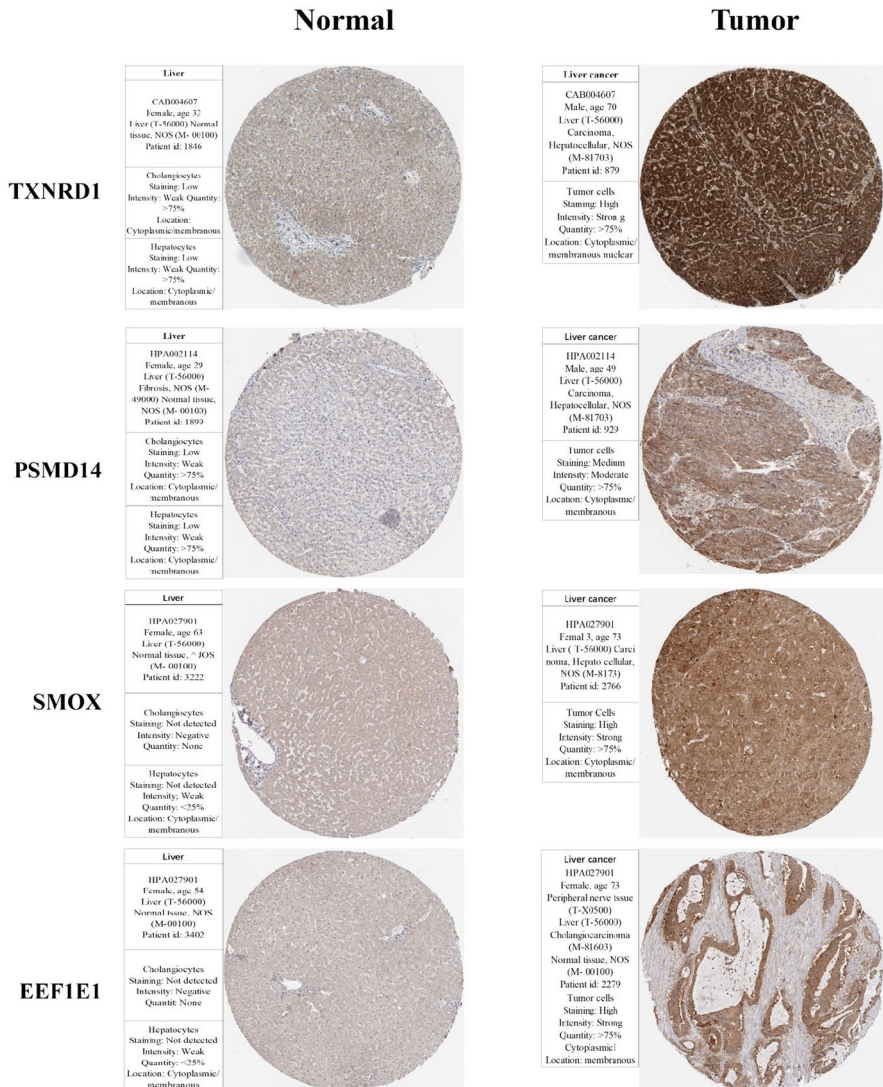
**Fig. 9** Forest plot from univariate and multivariate analyses of the 4-gene signature and nomogram for predicting 1-, 3-, and 5-year OS in the LIHC training set. **A** Univariable and multivariable analyses of the 4-gene signature in LIHC patients. **B** Nomogram for OS prediction at 1-, 3-, and 5-year. **C–E** ROC analysis of 1-, 3-, and 5-year OS prediction. **F** Calibration curves for prediction of 1-, 3-, and 5-year OS. **G** DCA curve. **H** Concordance index revealing measure of concordance of the predictor with patient survival

threshold probability (Fig. 9G). Relative to other clinical markers and risk score, the nomogram also had a higher consistency index (C index, Fig. 9H). Thus, ROC, DCA, calibration curve, and C index analyses indicate that the nomogram has better clinical benefits than the risk score based on the 4 AMGs signature alone.

**Protein Expressions of Hub Genes in HPA Database**

To further verify the functions of hub genes, the 4 AMGs were selected for immunohistochemical analysis with the Human Protein Atlas (HPA) database. Which show that, the protein levels of the four hub genes (TXNRD1, PSMD14, EEF1E1, and SMOX) were significantly higher in LIHC tumor tissues, compared to normal tissues (Fig. 10). Moreover, it was discovered that the aforementioned





**Fig. 10** Immunohistochemical detection of the hub genes in liver cancer and normal tissues from the HPA database

hub genes were expressed not only in the liver tissue but also in 26 other human organ types (Fig. S8).

### Drug Susceptibility Based on Hub Genes

Correlations between drug Z-scores and hub genes were analyzed, with the first 16 significant drug–gene pairs shown in Fig. S9. A total of 246 drugs showed statistical



differences, of which, Irofulven, Staurosporine, Amonafide, and 3-Bromopyruvate were highly positively correlated with hub gene expressions. In contrast, the other 12 drugs were negatively correlated with hub gene expressions (Fig. S9).

### Transcription Factor (TF) Regulatory Network Based on Hub Genes

To further understand the mechanisms of the hub genes, a regulatory network of transcription factors (TFs) centered around hub genes was constructed. As exhibited in Fig. S10, hub genes (TXNRD1, PSMD14, SMOX, and EEF1E1) play an important role in the regulatory network, they interact more or less with other TFs. In addition, there was also a mutual regulatory relationship between hub genes. Furthermore, it shows that all of the hub genes in this study were significantly correlated with each other (Fig. S7).

## Discussion

Recent studies have associated metabolism with cancer epigenetics (Lee and Kim 2022; Thakur and Chen 2019). Abnormal metabolism enhances tumor proliferation and metastasis. Numerous metabolic genes are effective prognostic biomarkers and amino acid metabolism is a vital metabolic variation in HCC. Bioinformatics approaches have been used to investigate how metabolism affects the risk of HCC (Liu et al. 2020; Tang et al. 2020). Immunotherapy is effective against cancers; however, its relationship with amino acid metabolism genes in HCC have not been investigated. Here, we established a 4 amino acid metabolism-associated genes signature and found that it is an effective prognostic biomarker and predictor of immunotherapeutic efficacy in HCC. In this study, we identified a 4-gene signature also displaying a relative high prognostic value of HCC, and some studies require more genes to achieve similar results (Zhao et al. 2021).

Based on differentially expressed genes that associated with amino acid metabolism, we divided the TCGA dataset of 375 liver cancer cases into 4 subtypes ( $k=4$ ) and found that their survival outcomes and clinical characteristics differed significantly (Fig. S2), indicating that amino acid metabolism is associated with liver cancer occurrence and development.

Then, we established a 4-gene signature to assess the amino acid metabolic status of liver cancer patients. The patients were assigned into low- and high-risk groups based on the risk score. K–M analysis revealed that low-risk patients had long survival times relative to high-risk patients. ROC curve analysis showed that the risk characteristics can efficiently predict 1-, 2-, and 3-year survival outcomes of HCC patients. The independent prognostic significance of the risk signature was verified by univariate and multivariate Cox analyses. Recent studies have reported that metabolic gene signatures can efficiently predict OS outcomes of HCC patients (Hu et al. 2020; Wu et al. 2021a). Prognostic markers based on lipid metabolism have been developed and shown to be closely associated with clinical features, immune cells, and various biological roles in HCC (Zhu et al.

2021). Wu et al. developed a six-gene metabolism risk signature for HCC that was highly based on lipid and nucleotide metabolism (Wu et al. 2021b). Liu et al. developed a prognostic marker for glioma that was based on amino acid metabolism and showed that the risk score was closely associated with various aspects of glioma malignancy (Wu et al. 2021a, b). In a recent study, a comprehensive approach incorporating risk model construction, analysis of immune cell infiltration, and gene expression analysis was employed to develop a 9-gene signature associated with amino acid metabolism, and the signature was further utilized to develop a prognostic nomogram for predicting OS in HCC (Zhao et al. 2021). However, in this study, the risk model comprising 4 amino acid metabolism genes revealed that the risk score significantly correlated with immunotherapy.

Notably, after adjusting molecular and clinical features, we found that the amino acid metabolism-associated risk signature was an independent prognostic factor. Next, we constructed a nomogram to predict 1-, 3-, and 5-year OS outcomes in the LIHC dataset. The ROC, DCA, calibration curves, and concordance index analyses showed that the nomogram exhibited superior clinical value relative to the risk score obtained by the 4-gene signature alone. The amino acid metabolism status was used to refine clinicopathological characteristics and which shows great promise in accurately predicting prognosis in liver cancer. Consequently, by integrating the identified risk signatures with other pertinent features, a more comprehensive and precise prognosis prediction for liver cancer can be achieved. This integrated approach improves the design of prognostic models and hence the clinical management of patients with liver cancer.

Risk score-based biomarkers offer valuable insights for prognostication and guiding targeted therapy in precision oncology. An extensive analysis of 6125 compounds revealed promising candidates, including MLN2238, SB-743921, SGK461364, clofarabine, paclitaxel, and BI-2536, identified from CTRP-derived drug response data. High-risk patients with liver cancer showed significant correlations with compounds from both PRISM and CTRP data, such as volasertib, epothilone-b, and ispinesib. These findings highlight the potential of risk score-based biomarkers in identifying targeted therapeutic options for liver cancer patients.

Pathway enrichment analysis using GSEA and KEGG analyses uncovered key insights in high-risk patients. Including cell cycle regulation and biological macromolecule synthesis pathways were enriched, highlighting the importance of amino acid metabolism-associated pathways in high toxicity metabolism, which indicated that amino acid metabolism-associated genes can impact amino acids levels, indicating that many high-risk patients were influenced by toxicity and drug metabolism. Ferroptosis-related genes were closely associated with immune microenvironment in HCC (Zhu et al. 2023). In this study, immune response-associated pathways are enriched in these patients. We also found that increased infiltration of memory B cells, follicular helper T cells, activated memory CD4<sup>+</sup> T cells and naive CD4<sup>+</sup> T cells, and reduced infiltration of naïve B cells in high-risk patients. B cells infiltration has been reported to be higher in liver cancer patients than that in liver cirrhosis patients or healthy subjects (Zhang et al. 2020). Elevated plasma cell levels and low levels of immature B cells are associated with poor prognosis (Zhang et al. 2019).

A recent study found that reduced levels of CD8<sup>+</sup> T cells result in immune dysregulation in HCC patients, which may promote HCC progression. Our results also indicated low infiltration of NK cells in high-risk patients. Elevated of Trp and Arg catabolism was reported to trigger NK cell apoptosis and to enhance tumor immune escape (Grohmann and Bronte 2010).

Further analysis indicated that there was no significant correlation (correlation=0.14) between risk score and TMB, which is more sensitive to immunotherapies. Stratified analysis showed that the prognostic significance of the risk score in LIHC was independent of TMB. The absence of association, along with individual predictive values as well as GSEA outcomes, suggests that the TMB and risk score are distinct tumor immunobiology aspects. Furthermore, the risk score demonstrates its independence to predict immunotherapeutic responses. Analysis of data from patients undergoing immunotherapy (datasets GSE91061, GSE135222, and IMvigor210) revealed significantly higher risk score in patients who responded to immunotherapy, highlighting the predictive value of risk scores. Together, this study indicates that a single dose of immunotherapy may benefit high-risk patients. Previous studies have been reported that immunodiagnostic markers exhibited important value in early prediction of HCC (Xing et al. 2021). The 4 amino acid metabolism-related genes identified in this study may also act as immunotherapeutic markers for hepatocellular carcinoma.

This study has several limitations that should be acknowledged. Firstly, the risk signature is based on 4 genes involved in amino acids metabolism was validated solely using an ICGC dataset. Thus, the clinical utility of this risk signature needs to be validated using real-world prospective data. Secondly, our analysis based on a single hallmark to establish the prognostic model, which may exclude other important factors associated with HCC prognosis. Additionally, the relationship between risk score and immunity warrants further investigation.

In conclusion, we have successfully developed a prognostic model based on four genes associated with amino acid metabolism. Our analysis shows that this model is independently correlated with overall survival in both validation and derivation cohorts, providing valuable insights into prognostic prediction of HCC. Moreover, our study highlights the effectiveness of the 4-gene signature in predicting HCC prognosis and response to immunotherapy.

## Conclusion

In summary, we developed a 4-gene amino acid metabolism-associated genes signature. Univariate and multivariate analyses revealed that the characteristics of 4-gene signature were independent prognostic factors in liver cancer. GSVA and KEGG analyses demonstrated a significant association between high-risk score tumors and various malignant characteristics of liver cancer. Moreover, the high-risk groups exhibited a higher number of mutant genes and elevated levels of immune infiltration. This observation was further validated in three immunotherapeutic cohorts, where patients with a low-risk score exhibited notable therapeutic and clinical

advantages. Finally, a prognostic nomogram was established according to the TCGA cohort. On a general perspective, this study demonstrates that the 4-gene signature serves as a reliable diagnostic marker and predictive marker for immunotherapy.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10528-023-10502-w>.

**Acknowledgements** The authors are grateful for the invaluable support and useful discussions with other members of the Department of Youjiang Medical College for Nationalities. We thank the TCGA and GEO databases as well as IMvigor210 package for the availability of the data. And we also want to thank the support of the Key Laboratory of Molecular Pathology (For Hepatobiliary Diseases) of Guangxi, Affiliated Hospital of Youjiang Medical University for Nationalities.

**Author Contributions** Designed study and wrote the paper: Ming-you Dong, Lu-sheng Liao, and Run-lei Du. Analyzed data: Lu-sheng Liao, Ming-you Dong, and Jun-li Wang. Performed research: Zi-jun Xiao, Ting-jun Liu, Feng-die Huang, Yan-ping Zhong, Xin Zhang, and Ke-heng Chen. Contributed to methodology: Lu-sheng Liao, Ming-you Dong, and Run-lei Du. All the authors contributed to the article and approved the submitted version.

**Funding** This study was supported by the Basic Ability Improvement Project for Young and Middle-aged Teachers in Colleges and Universities of Guangxi (Grant No. 2022KY0542), and the Basic Ability Improvement Project for Young and Middle-aged Teachers in Colleges and Universities of Guangxi (Grant No. 2022KY0532). The work here was also supported by the Project of Basic Scientific Research and Technology Development Plan in 2021 (Grant No. 20212348) and the Project of Baise Scientific Research and Technology Development Plan in 2021 (Grant No. 20212347).

**Data Availability** The data underlying this study are freely available from the TCGA, ICGC, and the GEO databases. The R code and raw data of this article can be obtained from the following link <https://www.jianguoyun.com/p/DRReQzUQufXRCxi9xYYFIAA>. The data from TCGA, ICGC, and GEO are all publicly available. Therefore, this study was exempted from the approval of the local ethics committee. The current study follows TCGA, ICGC, and GEO data access policies and publication guidelines.

## Declarations

**Conflict of interest** The authors have not disclosed any conflict of interest.

## References

- Alawya B, Constantinou C (2023) Hepatocellular Carcinoma: a narrative review on current knowledge and future prospects. *Curr Treat Options Oncol*. <https://doi.org/10.1007/s11864-023-01098-9>
- Charoentong P, Finotello F, Angelova M et al (2017) Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* 18:248–262. <https://doi.org/10.1016/j.celrep.2016.12.019>
- Dasgupta P, Henshaw C, Youlden DR et al (2020) Global trends in incidence rates of primary adult liver cancers: a systematic review and meta-analysis. *Front Oncol* 10:171. <https://doi.org/10.3389/fonc.2020.00171>
- Fasano M, Corte C, Liello RD et al (2022) Immunotherapy for head and neck cancer: present and future. *Crit Rev Oncol Hematol* 174:103679. <https://doi.org/10.1016/j.critrevonc.2022.103679>
- Fu S, Xu S, Zhang S (2023) The role of amino acid metabolism alterations in pancreatic cancer: from mechanism to application. *Biochim Biophys Acta Rev Cancer* 1878:188893. <https://doi.org/10.1016/j.bbcan.2023.188893>
- Grohmann U, Bronte V (2010) Control of immune response by amino acid metabolism. *Immunol Rev* 236:243–264. <https://doi.org/10.1111/j.1600-065X.2010.00915.x>
- Hanzelmann S, Castelo R, Guinney J (2013) GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14:7. <https://doi.org/10.1186/1471-2105-14-7>

- Hu B, Yang XB, Sang XT (2020) Construction of a lipid metabolism-related and immune-associated prognostic signature for hepatocellular carcinoma. *Cancer Med* 9:7646–7662. <https://doi.org/10.1002/cam4.3353>
- Huang AC, Zappasodi R (2022) A decade of checkpoint blockade immunotherapy in melanoma: understanding the molecular basis for immune sensitivity and resistance. *Nat Immunol* 23:660–670. <https://doi.org/10.1038/s41590-022-01141-1>
- Iasonos A, Schrag D, Raj GV et al (2008) How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* 26:1364–1370. <https://doi.org/10.1200/JCO.2007.12.9791>
- Jain M, Nilsson R, Sharma S et al (2012) Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science* 336:1040–1044. <https://doi.org/10.1126/science.1218595>
- Lee JE, Kim MY (2022) Cancer epigenetics: past, present and future. *Semin Cancer Biol* 83:4–14. <https://doi.org/10.1016/j.semcancer.2021.03.025>
- Li Z, Zhang H (2016) Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. *Cell Mol Life Sci* 73:377–392. <https://doi.org/10.1007/s00018-015-2070-4>
- Li X, Zhang HS (2023) Amino acid metabolism, redox balance and epigenetic regulation in cancer. *Febs J*. <https://doi.org/10.1111/febs.16803>
- Liu GM, Xie WX, Zhang CY et al (2020) Identification of a four-gene metabolic signature predicting overall survival for hepatocellular carcinoma. *J Cell Physiol* 235:1624–1636. <https://doi.org/10.1002/jcp.29081>
- Mates JM, Campos-Sandoval JA, de Los SJ et al (2020) Glutaminases regulate glutathione and oxidative stress in cancer. *Arch Toxicol* 94:2603–2623. <https://doi.org/10.1007/s00204-020-02838-8>
- Mayakonda A, Lin DC, Assenov Y et al (2018) Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* 28:1747–1756. <https://doi.org/10.1101/gr.239244.118>
- Pavlova NN, Thompson CB (2016) The emerging hallmarks of cancer metabolism. *Cell Metab* 23:27–47. <https://doi.org/10.1016/j.cmet.2015.12.006>
- Reda M, Ngamcherdtrakul W, Nelson MA et al (2022) Development of a nanoparticle-based immunotherapy targeting PD-L1 and PLK1 for lung cancer treatment. *Nat Commun* 13:4261. <https://doi.org/10.1038/s41467-022-31926-9>
- Rooney MS, Shukla SA, Wu CJ et al (2015) Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160:48–61. <https://doi.org/10.1016/j.cell.2014.12.033>
- Shen R, Li P, Li B et al (2019) Identification of distinct immune subtypes in colorectal cancer based on the stromal compartment. *Front Oncol* 9:1497. <https://doi.org/10.3389/fonc.2019.01497>
- Tang C, Ma J, Liu X et al (2020) Identification of a prognostic signature of nine metabolism-related genes for hepatocellular carcinoma. *PeerJ* 8:e9774. <https://doi.org/10.7717/peerj.9774>
- Thakur C, Chen F (2019) Connections between metabolism and epigenetics in cancers. *Semin Cancer Biol* 57:52–58. <https://doi.org/10.1016/j.semcancer.2019.06.006>
- Tibshirani R (1997) the lasso method for variable selection in the cox model. *Stat Med* 16:385–395. [https://doi.org/10.1002/\(SICI\)1097-0258\(19970228\)16:4<385::AID-SIM380>3e3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3e3.0.CO;2-3)
- Torreccilla S, Sia D, Harrington AN et al (2017) Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma. *J Hepatol* 67:1222–1231. <https://doi.org/10.1016/j.jhep.2017.08.013>
- Vettore L, Westbrook RL, Tennant DA (2020) New aspects of amino acid metabolism in cancer. *Br J Cancer* 122:150–156. <https://doi.org/10.1038/s41416-019-0620-5>
- Vickers AJ, Elkin EB (2006) Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 26:565–574. <https://doi.org/10.1177/0272989X06295361>
- Villanueva A (2019) Hepatocellular carcinoma. *N Engl J Med* 380:1450–1462. <https://doi.org/10.1056/NEJMr1713263>
- Wilkerson MD, Hayes DN (2010) ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26:1572–1573. <https://doi.org/10.1093/bioinformatics/btq170>
- Wu T, Hu E, Xu S et al (2021a) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (NY)* 2:100141. <https://doi.org/10.1016/j.xinn.2021.100141>
- Wu X, Lan T, Li M et al (2021) Six metabolism related mRNAs predict the prognosis of patients with hepatocellular carcinoma. *Front Mol Biosci* 8:621232. <https://doi.org/10.3389/fmolb.2021.621232>
- Xing M, Wang X, Kiken RA et al (2021) Immunodiagnostic biomarkers for hepatocellular carcinoma (HCC): the first step in detection and treatment. *Int J Mol Sci* 22:6139. <https://doi.org/10.3390/ijms22116139>
- Yang JD, Hainaut P, Gores GJ et al (2019) A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastro Hepat* 16:589–604. <https://doi.org/10.1038/s41575-019-0186-y>

- Yao C, Zhang T, Wu T et al (2022) Facts and hopes for immunotherapy in renal cell carcinoma. *Clin Cancer Res* 28:5013–5020. <https://doi.org/10.1158/1078-0432.CCR-21-2372>
- Zhang Z, Ma L, Goswami S et al (2019) Landscape of infiltrating B cells and their clinical significance in human hepatocellular carcinoma. *Oncoimmunology* 8:e1571388. <https://doi.org/10.1080/2162402X.2019.1571388>
- Zhang S, Liu Z, Wu D et al (2020) Single-cell RNA-Seq analysis reveals microenvironmental infiltration of plasma cells and hepatocytic prognostic markers in HCC with cirrhosis. *Front Oncol* 10:596318. <https://doi.org/10.3389/fonc.2020.596318>
- Zhao Y, Zhang J, Wang S et al (2021) Identification and validation of a nine-gene amino acid metabolism-related risk signature in HCC. *Front Cell Dev Biol* 9:731790. <https://doi.org/10.3389/fcell.2021.731790>
- Zhu P, Li FF, Zeng J et al (2021) Integrative analysis of the characteristics of lipid metabolism-related genes as prognostic prediction markers for hepatocellular carcinoma. *Eur Rev Med Pharmacol Sci* 25:116–126. [https://doi.org/10.26355/eurrev\\_202101\\_24355](https://doi.org/10.26355/eurrev_202101_24355)
- Zhu L, Zhu X, Wu Y (2022) Effects of glucose metabolism, lipid metabolism, and glutamine metabolism on tumor microenvironment and clinical implications. *Biomolecules* 12:580. <https://doi.org/10.3390/biom12040580>
- Zhu J, Xu X, Jiang M et al (2023) Comprehensive characterization of ferroptosis in hepatocellular carcinoma revealing the association with prognosis and tumor immune microenvironment. *Front Oncol* 13:1145380. <https://doi.org/10.3389/fonc.2023.1145380>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

Lu-sheng Liao<sup>1,2,3</sup> · Zi-jun Xiao<sup>3</sup> · Jun-li Wang<sup>4</sup> · Ting-jun Liu<sup>3</sup> · Feng-die Huang<sup>3</sup> · Yan-ping Zhong<sup>3</sup> · Xin Zhang<sup>3</sup> · Ke-heng Chen<sup>3</sup> · Run-lei Du<sup>5</sup> · Ming-you Dong<sup>1,2,3</sup>

✉ Run-lei Du  
runleidu@whu.edu.cn

✉ Ming-you Dong  
mydong@ymun.edu.cn

<sup>1</sup> The Key Laboratory of Molecular Pathology (For Hepatobiliary Diseases) of Guangxi, Affiliated Hospital of Youjiang Medical University for Nationalities, No. 98, Chengxiang Road, Youjiang District, Baise 533000, Guangxi, China

<sup>2</sup> School of Medical Laboratory Medicine, Youjiang Medical University for Nationalities, Baise 533000, Guangxi, China

<sup>3</sup> Modern Industrial College of Biomedicine and Great Health, Youjiang Medical University for Nationalities, Baise, China

<sup>4</sup> Department of Reproductive Medicine, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise 533000, Guangxi, China

<sup>5</sup> Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan 430072, Hubei, China