ORIGINAL ARTICLE

WILEY

Development and validation of a CTNNB1-associated metabolic prognostic model for hepatocellular carcinoma

Junyu Huo [ID] | Liqun Wu | Yunjin Zang

Liver Disease Center, The Affiliated Hospital of Qingdao University, Qingdao, China

Correspondence
Liqun Wu, Liver Disease Center, The Affiliated Hospital of Qingdao University, No. 59 Haier Road, Qingdao, 266003, China.
Email: wulq5810@126.com

Abstract

Hepatocellular carcinoma (HCC) is a heterogeneous malignancy closely related to metabolic reprogramming. We investigated how CTNNB1 mutation regulates the HCC metabolic phenotype and thus affects the prognosis of HCC. We obtained the mRNA expression profiles and clinicopathological data from The Cancer Genome Atlas (TCGA), the International Cancer Genomics Consortium (ICGC) and the Gene Expression Omnibus database (GSE14520 and GSE116174). We conducted gene set enrichment analysis on HCC patients with and without mutant CTNNB1 through TCGA dataset. The Kaplan-Meier analysis and univariate Cox regression analysis assisted in screening metabolic genes related to prognosis, and the prognosis model was constructed using the Lasso and multivariate Cox regression analysis. The prognostic model showed good prediction performance in both the training cohort (TCGA) and the validation cohorts (ICGC, GSE14520, GSE116174), and the high-risk group presented obviously poorer overall survival compared with low-risk group. Cox regression analysis indicated that the risk score can be used as an independent predictor for the overall survival of HCC. The immune infiltration in different risk groups was also evaluated in this study to explore underlying mechanisms. This study is also the first to describe an metabolic prognostic model associated with CTNNB1 mutations and could be implemented for determining the prognoses of individual patients in clinical practice.

KEYWORDS
CTNNB1 mutation, hepatocellular carcinoma, metabolic, prognostic

1 | INTRODUCTION

Hepatocellular carcinoma (HCC) acts as a heterogeneous disease of which the prognosis is dismal. However, recently, with the rapid development of gene sequencing technology, our understanding of the molecular pathogenesis of HCC has improved significantly.[1,2] Accumulated data from high-throughput analysis of a large number of samples suggest that it can be used to identify key biomarkers associated with HCC progression. However, the number of biomarkers known to be associated with HCC prognosis is limited.

CTNNB1 mutations have been reported in approximately 18% to 40% of HCC patients.[3-7] The oncogenic Wnt/β-catenin pathway, activated by the mutated CTNNB1, plays a key role in the metabolic regulation in the liver.[1,8] CTNNB1-mutated HCC has a distinctive

metabolic morphotype and is often cholestatic and infrequently steatotic.[3,9,10] A recent report showed that proteins involved in different metabolic activities such as drug metabolism, amino acid metabolism, glycolysis and gluconeogenesis are enriched in CTNNB1 mutant tumours,[11] indicating that it is closely related to metabolic reprogramming. However, the potential mechanism is not fully understood.

Metabolic reprogramming refers to the significant changes in metabolic patterns that occur during the process of cell carcinogenesis, which involves many processes such as glycolysis, oxidative phosphorylation, tricarboxylic acid cycle, amino acid metabolism, nucleic acid metabolism and fatty acid metabolism.[12-14] Tumorigenesis is a multistep process involving modifications to pathways that promote uncontrolled proliferation and eliminate cell death, which requires metabolic reprogramming to provide large molecules for cell survival, growth and migration. In recent years, targeting metabolic reprogramming has become a promising novel treating strategy, but the metabolic reprogramming of HCC has not yet been deciphered.

The study conducted gene set enrichment analysis on the CTNNB1 mutant HCC and CTNNB1 wild-type HCC through the Cancer Genome Atlas (TCGA) database, finding that gene sets significantly up-regulated in CTNNB1 mutant HCC were all related to metabolism, further confirming the close relationship between CTNNB1 mutation and metabolic reprogramming. We further explored the prognostic value of these genes in HCC and constructed a prognostic risk model consisting of five metabolic genes. The results of multiple datasets (including TCGA, ICGC, GSE14520 and GSE11 6174; total of 876 HCC patients) indicate that this risk score is highly accurate in evaluating the prognosis of HCC. In addition, we found that the prognostic model may reflect the immune microenvironment of the tumour and thus has high clinical application potential.

2 | MATERIALS AND METHODS

2.1 | Data collection from TCGA

We obtained the sequence data for 374 samples with HCC from the TCGA website (https://portal.gdc.cancer.gov/repository). The corresponding clinical data including overall survival time, survival status, sex, age, race, alpha-fetoprotein (AFP) level, body mass index
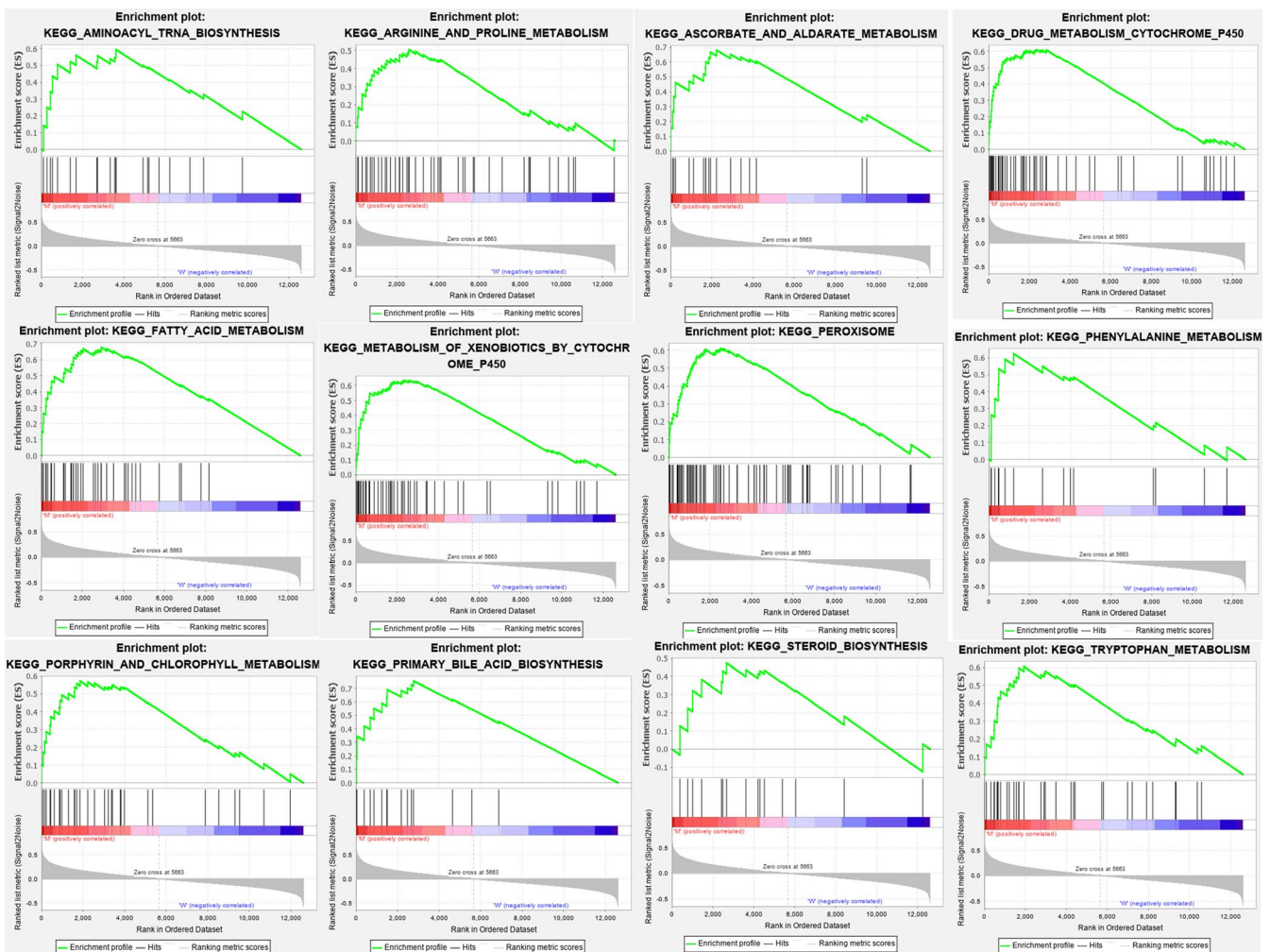

FIGURE 1 Gene set enrichment analysis between with and without CTNNB1 mutation HCC in TCGA

(BMI), vascular invasion, histological grading, AJCC TNM stage, family history of cancer, tumour history, presence of new tumours after initial treatment and individual tumour status of the TCGA cohort were obtained from the UCSC Xena website (https://xenab rowser.net/). The CTNNB1 mutation sample list was obtained from the cBioPortal website (https://www.cbioportal.org/). We retained genes of which the average expression value is over 1 and mean-while removed RNA-sequencing data with low abundance.[15,16] This study meets the publication requirement of TCGA (http://cance rgenome.nih.gov/publications/publicationguidelines).

Illumina HiSeq RNA-seq platform, GSE14520 gene expression files from the GPL571 platform and GSE116174 gene expression files from the GPL13158 platform) and corresponding clinical data from the Gene Expression Omnibus (GEO) database (https:// www. ncbi.nlm.nih.gov/geo/) and the International Cancer Genomics Consortium (ICGC, https://icgc.org/). The batch effects of RNA-seq data sets were eliminated via the combat function contained in the SVA R package (R Core Team, R Foundation for Statistical Computing, Vienna, Austria). The downloaded profiles all meet the GEO and ICGC data access rules.

2.2 | Data collection from ICGC and GEO

Three independent cohorts were used for external validation (ICGC-LIRI-JP, GSE14520 and GSE116174). We obtained the gene expression files (ICGC-LIRI-JP gene expression files from the

2.3 | Gene set enrichment analysis (GSEA)

Gene set enrichment analysis was utilized in this study to un-covering the differences in the metabolic-related genes be-tween the CTNNB1 mutant group (n = 98) and the non-mutant

TABLE 1 Gene set enrichment analysis between with and without CTNNB1 mutant HCC

NAME	NES	NOM P-value
KEGG_METABOLISM_OF_XENOBIOTICS_BY_ CYTOCHROME_P450	1.914358	0.002132
KEGG_DRUG_METABOLISM_CYTOCHROME_P450	1.81373	0.006383
KEGG_PEROXISOME	1.767487	0.02439
KEGG_TYROSINE_METABOLISM	1.758547	0.012876
KEGG_PORPHYRIN_AND_CHLOROPHYLL_ METABOLISM	1.745985	0.026477
KEGG_STEROID_HORMONE_BIOSYNTHESIS	1.725979	0.015317
KEGG_PRIMARY_BILE_ACID_BIOSYNTHESIS	1.710733	0.010331
KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	1.707652	0.025052
KEGG_FATTY_ACID_METABOLISM	1.707306	0.046939
KEGG_ASCORBATE_AND_ALDARATE_METABOLISM	1.700901	0.014737
KEGG_PHENYLALANINE_METABOLISM	1.627591	0.028509

TABLE 2 The prognostic gene list related to overall survival in TCGA cohort

Gene	K-M	HR	HR.95L	HR.95H	Cox P-value
CYP27A1	0.004001	0.995377	0.992811	0.99795	0.000435
AGPS	0.002754	1.14859	1.083584	1.217495	3.15E-06
CYP3A5	0.000394	0.994496	0.991012	0.997993	0.002057
HMBS	0.005014	1.085504	1.021233	1.15382	0.008422
ADH1A	0.001722	0.997199	0.995436	0.998966	0.001897
SRM	0.00237	1.015787	1.00748	1.024162	0.000185
ACSL3	0.002632	1.047003	1.024575	1.069922	3.22E-05
ALAS1	0.009778	0.992547	0.988309	0.996803	0.000611
PRDX1	0.002291	1.004368	1.002789	1.00595	5.63E-08
GOT2	0.003586	0.983121	0.974783	0.99153	8.95E-05
HCCS	0.000666	1.113809	1.047345	1.184489	0.000596
AMD1	4.61E-06	1.110299	1.068779	1.153433	7.42E-08
CYP2C9	0.00024	0.99541	0.992975	0.997852	0.000233
SMS	3.01E-06	1.035489	1.022567	1.048574	5.23E-08

group (n = 276) in the TCGA cohort. An annotated gene set file (c2.cp.kegg.v7.0.symbols.gmt) was selected as the reference. The threshold was confirmed as NOM *P*-value < 0.05.

2.4 | Construction and validation of a metabolic-related prognostic model

We first used univariate Cox regression and the Kaplan-Meier (K-M) for survival analysis of each gene and selected genes with

P < 0.05 in both algorithms as candidate genes to construct the model.[17] Next, the Lasso regression analysis assisted in further narrowing down the candidate gene number. Finally, we constructed a risk score system, multiplying the normalized expression level exhibited by each metabolic gene by virtue of the regression coefficients that were obtained from the multivariate Cox regression analysis. The median risk score of the TCGA cohort (n = 343) was used to classify the group with a high risk and group with a low risk. The Kaplan-Meier (K-M) survival analysis (log-rank test), together with the time-related receiver operating
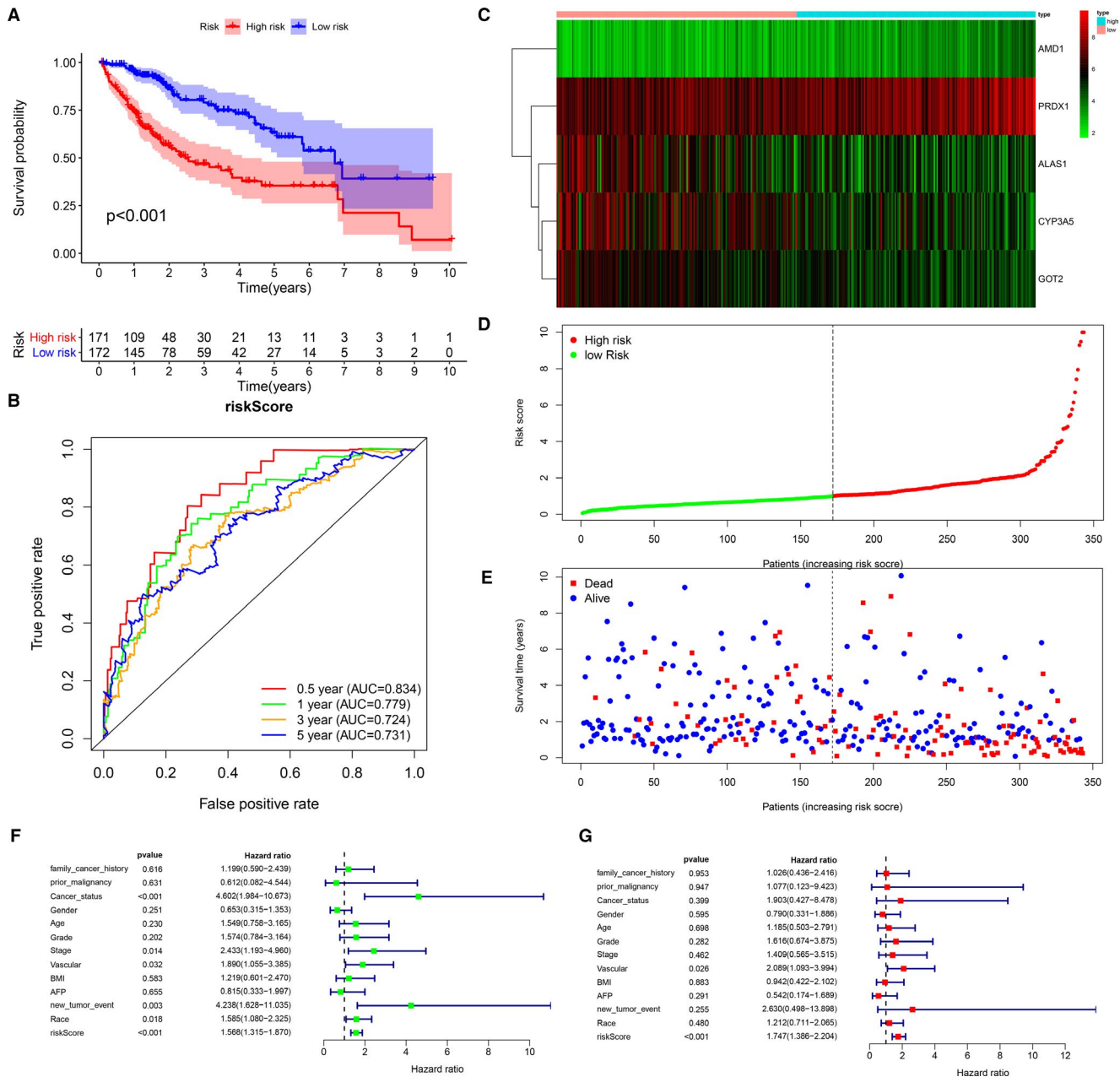

FIGURE 2 Construction of the prognostic model in TCGA cohort (A) The Kaplan-Meier survival analysis for overall survival (OS) of patients in TCGA cohort. (B) The time-dependent ROC analysis for risk score in the TCGA cohort. (C-E) The heat map of the five genes and the distribution of risk score and the survival status of patients in the TCGA cohort. (F-G) Forest plot of the univariate and multivariate Cox regression analysis in HCC regarding OS (green represents univariate analysis, and red represents multivariate analysis)

feature curve (ROC) analysis applied assessing the predictive ability exhibited by the prognostic model in above-mentioned four independent cohorts (TCGA, ICGC, GSE14520 and GSE11 6174). Univariate and multivariate Cox regression analyses were performed in this study to confirm whether the risk score could independently predict the prognosis. $P < 0.05$ was considered with statistical significance. K-M method in the R package 'survminer' was used to generate the survival curves, and R package 'survivalROC' was used to generate ROC curves. What needs to be pointed out is that in order to avoid the impact of other factors

on the prognosis of patients, we excluded patients with a survival time of less than one month.

2.5 | Correlation analysis between clinicopathological parameters and risk score

We used the Wilcoxon signed-rank test (2 groups) or the Kruskal-Wallis test (>=2 groups) for analysing the correlation between clinicopathology (including AJCC TNM stage, Barcelona stage, CLIP
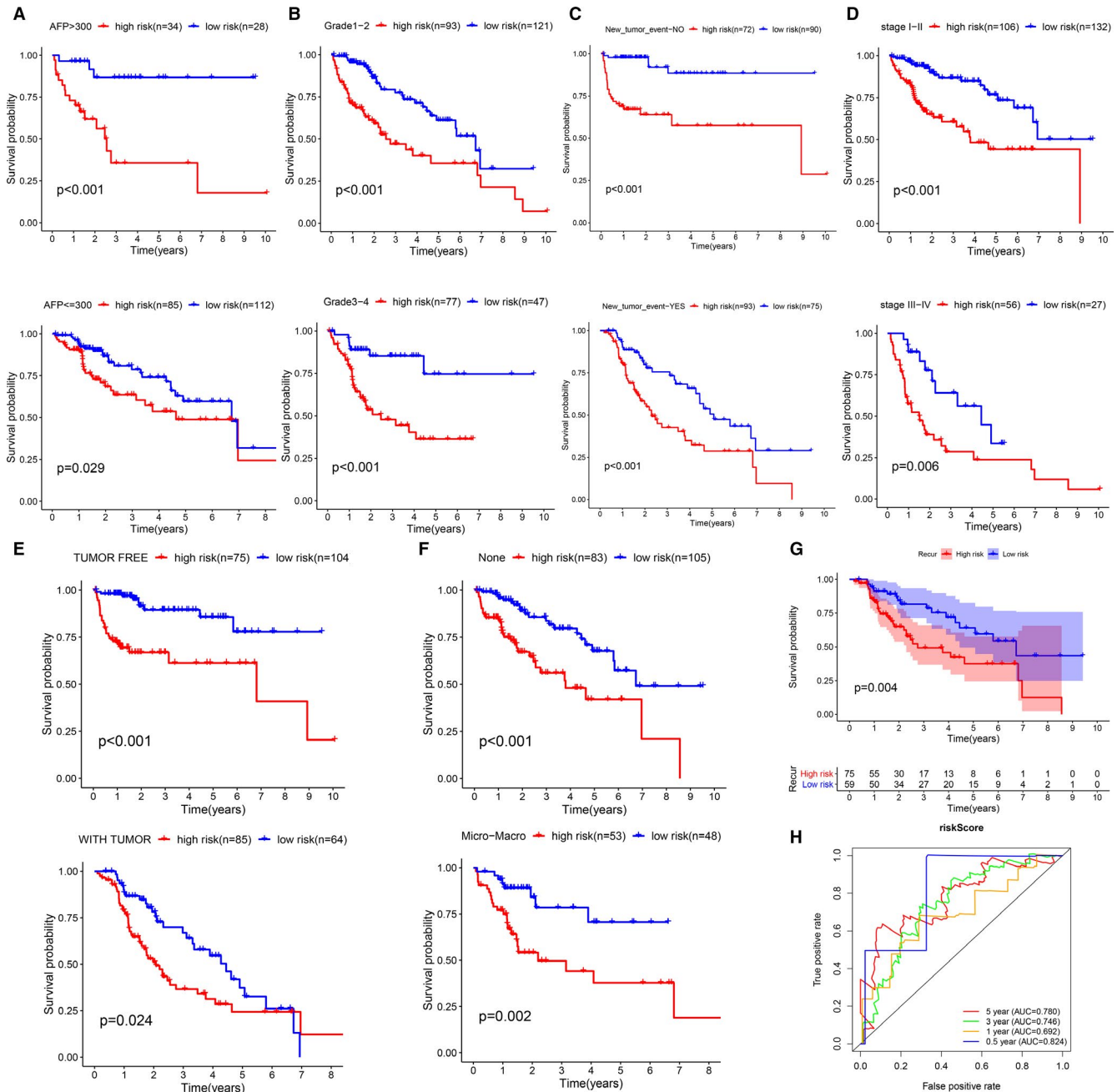

FIGURE 3 Internal validation of the prognostic model in TCGA cohort based on clinical features (A) AFP. (B) Histopathological grade. (C) New tumour event after initiate treatment. (D) AJCC TNM stage. (E) Cancer status. (F) Vascular tumour cell type. (G-H) Recurrence

stage, main tumour size, tumour differentiation, AFP and vascular tumour cell type) and the risk score. *P* < 0.05 was considered with statistical significance, and box plots were generated via the beeswarm packages implemented in R software.

2.6 | Identification of differentially expressed genes (DEGs) between different risk groups

DEGs in the tissues of the two HCC groups were examined using the Wilcoxon test method in R package 'limma'. The thresholds were confirmed to be |log2-fold change (FC)| > 1.0 and FDR < 0.05. Gene

ontology (GO) enrichment analysis was further used for biological function annotation of the DEGs using the R package 'clusterProfiler'.

2.7 | Evaluation of immune cell infiltration in different risk groups

Relevant data about the infiltration level exhibited by immune cells of HCC patients were downloaded from the tumour immune assessment resource (TIMER) website to compare the immune cell infiltration level between different risk groups in the TCGA dataset. The single sample gene set enrichment analysis (ssGSEA, the
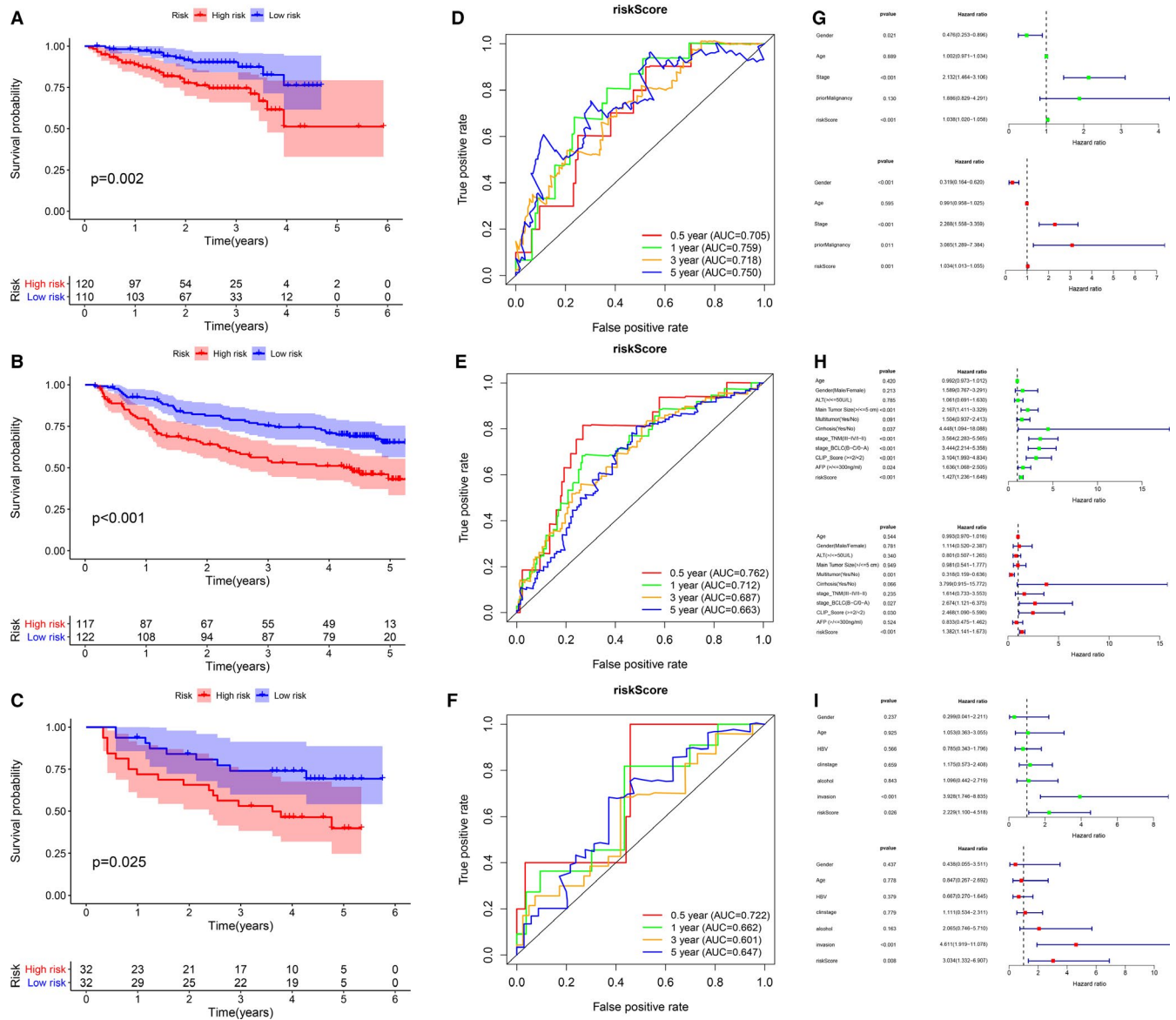

FIGURE 4 External validation of the prognostic model in three independent cohort (A-C) The Kaplan-Meier curve of overall survival (OS) in ICGC cohort, GSE14520 cohort and GSE116174 cohort, respectively. (D-F) The time-dependent ROC analysis for risk score in ICGC cohort, GSE14520 cohort and GSE116174 cohort, respectively. (G-I) Forest plot of the univariate and multivariate Cox regression analysis regarding OS in ICGC cohort, GSE1452 cohort and GSE116174 cohort, respectively (green represents univariate analysis, and red represents multivariate analysis)

immune cell type, function and pathway represented by 29 immune-related gene sets) was employed to quantify the activity or enrichment levels of immune cells, functions or pathways in the high- and low-risk samples from the four independent cohorts. The normalized enrichment score (NES) that calculated from the ssGSEA was used in the 'GSVA' and 'GSEABase' R package. The independent-samples t tests were used for comparing the differences in immune infiltration levels and immune function between high- and low-risk groups, and P-values < 0.05 were suggested to exhibit statistical significance.

2.8 | Statistical analyses

R software v3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) together with GraphPad Prism v7.00 (GraphPad Software Inc, USA) assisted in performing statistical analyses. Fisher's exact test or Pearson's chi-square test helped to analyse qualitative variables. Analysis on quantitative variables relied on a non-parametric Wilcoxon rank-sum test (for unpaired samples). The Kruskal-Wallis test was applied for normalizing multiple groups. If not specified above, P < 0.05 was considered with statistical significance.

3 | RESULTS

3.1 | Identification of differential metabolic gene sets between HCC samples with and without CTNNB1 mutations

Although CTNNB1 mutant HCC has been reported to have unique metabolic characteristics, the associated metabolic genes are

still unknown.[3] GSEA showed that the gene sets significantly up-regulated in CTNNB1 mutant HCCs were all associated with metabolic pathways (NOM P < 0.05) (Figure 1). These included 'metabolism of xenobiotics by cytochrome P450' (NES = 1.914, NOM P = 0.002), 'drug metabolism—cytochrome P450' (NES = 1.813, NOM P = 0.006), 'peroxisome' (NES = 1.767, NOM P = 0.02), 'tyrosine metabolism' (NES = 1.758, NOM P = 0.01), 'porphyrin and chlorophyll metabolism' (NES = 1.745, NOM P = 0.002), 'steroid hormone biosynthesis' (NES = 1.725, NOM P = 0.015), 'primary bile acid biosynthesis' (NES = 1.710, NOM P = 0.01), 'aminoacyl-tRNA biosynthesis' (NES = 1.707, NOM P = 0.02), 'fatty acid metabolism' (NES = 1.707, NOM P = 0.04), 'ascorbate and aldarate metabolism' (NES = 1.700, NOM P = 0.01), 'phenylalanine metabolism' (NES = 1.62, NOM P = 0.02) and 'arginine and proline metabolism' (NES = 162, NOM P = 0.04) (Table 1). In contrast, no gene sets related to metabolic pathways were significantly enriched in wild-type CTNNB1 HCC.

3.2 | Construction of prognostic model in the TCGA cohort

Although CTNNB1 has been confirmed to be one of the genes with the highest mutation frequency in HCC,[1-4,7,18] whether the significantly enriched metabolic gene set in CTNNB1 mutant HCC affects the prognosis of HCC is still unclear. Through univariate Cox regression analysis and K-M survival analysis, we identified 14 genes most associated with overall survival of HCC (Table 2). We used the Lasso regression and multi–Cox regression to further narrow down the range of model genes to optimize the model. Finally, a risk score consisting of five metabolic genes was constructed: risk score (RS) = normalized expression level of CYP3A5 * −0.00248 + normalized expression level of ALAS1 *
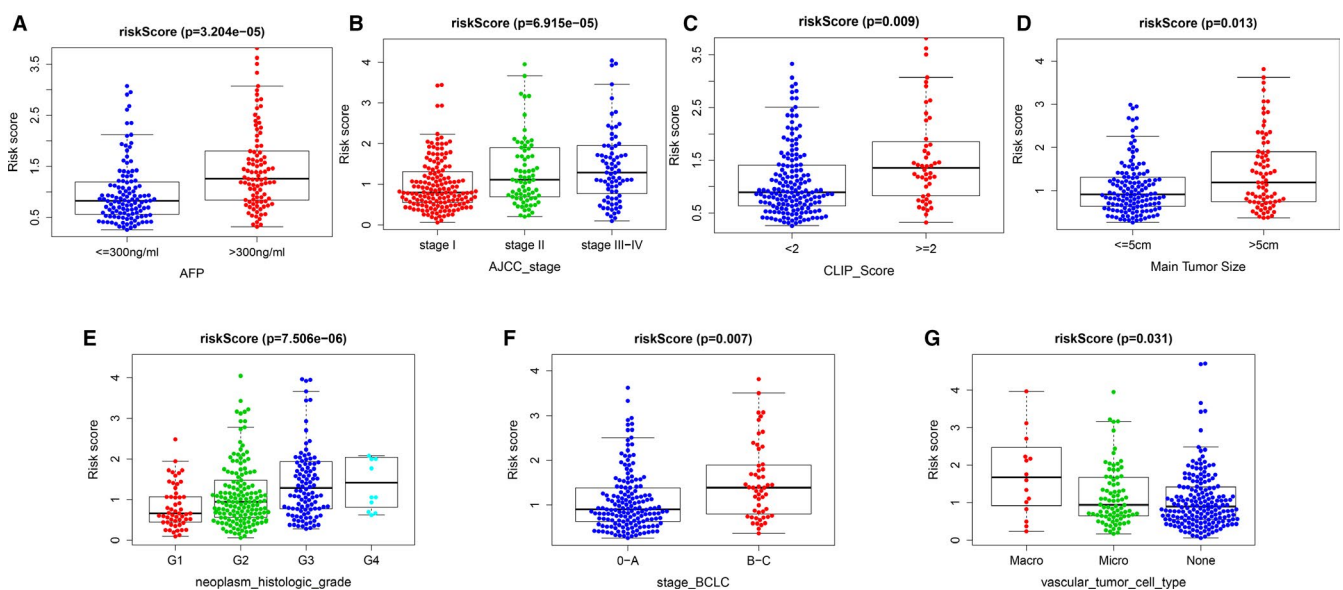

FIGURE 5 Correlation analysis between risk score and clinicopathological characteristics (A) AFP. B, AJCC stage. C, CLIP stage. D, Main tumour size. E, Histopathological grade. F, BCLC stage. G, Vascular tumour cell type

	Risk score			
Clinical features	Low risk n (%)	High risk n (%)	χ^2	P-value
AFP				
>300 ng/mL	28 (20.00%)	34 (28.57%)	2.596	0.1072
≤300 ng/mL	112 (80.00%)	85 (71.43%)		
Age				
>65	64 (37.21%)	63 (36.84%)	0.004959	0.9439
≤65	108 (62.79%)	108 (63.16%)		
Gender				
Female	51 (29.65%)	59 (34.50%)	0.9266	0.3358
Male	121 (70.35%)	112 (65.50%)		
BMI				
>25	82 (51.25%)	71 (45.51%)	1.041	0.3076
≤25	78 (48.75%)	85 (54.49%)		
Histological grade				
G1-2	121 (72.02%)	93 (54.71%)	10.91	0.001
G3-4	47 (27.98%)	77 (45.29%)		
New tumour event after initiate treatment				
Yes	75 (45.45%)	93 (56.36%)	3.929	0.0475
No	90 (54.55%)	72 (43.64%)		
Prior malignancy				
Yes	18 (10.47%)	13 (7.60%)	0.8548	0.3552
No	154 (89.53%)	158 (92.40%)		
Race				
White	80 (50.63%)	89 (55.97%)	0.9086	0.3405
Asian	78 (49.37%)	70 (44.03%)		
Family cancer history				
Yes	44 (38.60%)	29 (35.37%)	0.213	0.6444
No	70 (61.40%)	53 (64.63%)		
AJCC TNM stage				
I-II	132 (83.02%)	106 (65.43%)	12.95	0.0003
III-IV	27 (16.98%)	56 (34.57%)		
Cancer status				
With tumour	64 (38.10%)	85 (53.13%)	7.467	0.0063
Tumour free	104 (61.90%)	75 (46.88%)		
Vascular invasion				
None	105 (68.63%)	83 (61.03%)	1.828	0.1763
Yes	48 (31.37%)	53 (38.97%)		
Survival status				
Alive	90 (70.31%)	134 (62.33%)	2.259	0.1328
Dead	38 (29.69%)	81 (37.67%)		

−0.00445 + normalized expression level of PRDX1 * 0.003267 + normalized expression level of GOT2 * −0.00887 + normalized expression level of AMD1 * 0.078805. The median risk score of the TCGA cohort was used as the unified cut-off for dividing the group with a high risk (RS > 1) and the group with a low risk (RS < 1). The group with a high risk presented an obviously lower OS relative to the group with a low risk (Figure 2A), and the area under the ROC curve of the prognosis model at 0.5, 1, 3 and 5 years was 0.834, 0.779, 0.724 and 0.731, respectively (Figure 2B). The expression levels of ALAS1, CYP3A5 and GOT2 were higher in the low-risk group, whereas the expression levels of AMD1 and PRDX1 were higher in the high-risk group (Figure 2C). The risk of death in HCC patients increased with the increase in risk

score (Figure 2D-E). As found by the univariate and multivariate Cox regression analyses, the risk score could independently predict the prognosis (Figure 2F-G).

3.3 | Internal validation of the prognostic model in TCGA cohort

Patients fell into 12 subgroups for survival analysis according to their clinical characteristics such as AFP levels, vascular invasion, histological grade, AJCC TNM stage, new tumour after initial treatment and individual tumour status. In each subgroup, the OS rate of the group with a high risk appeared lower relative to the group with a low risk (Figure 3A-F). Also, the prognostic model accurately evaluated the prognosis of 134 patients with recurrence (including 65 cases of intrahepatic recurrence, 46 cases of local recurrence and 23 cases of extrahepatic recurrence). The area under the ROC curve at 0.5, 1, 3 and 5 years was 0.824, 0.692, 0.746 and 0.780, respectively (Figure 3H).

3.4 | External validation of the prognostic model in three different independent cohort

For verifying whether the prognostic model is robust, we used three independent cohorts (ICGC (n = 230), GSE14520 (n = 239) and GSE116174 (n = 64)) for external validation. Using the same formula and cut-off obtained from the TCGA-HCC cohort, patients were assigned to the group with a high risk and group with a low risk. Consistent with the results of TCGA, the group with a high risk showed significantly poorer OS relative to the group with a low risk (Figure 4A-C). In the ICGC cohort, the areas under the ROC curve at 0.5, 1, 3 and 5 years were 0.705, 0.759, 0.718 and 0.750, respectively (Figure 4D). In the GSE14520 cohort, the area under the ROC curve at 0.5, 1, 3 and 5 years was 0.762, 0.712, 0.687 and 0.663, respectively (Figure 4E). In the GSE116174 cohort, the areas under the ROC curve at 0.5, 1, 3 and 5 years were 0.722, 0.662, 0.601 and 0.647, respectively (Figure 4F). As demonstrated in the univariate and multivariate Cox regression analysis, the risk score could independently predict the prognosis in each cohort (Figure 4G-I). These results indicate that the prognostic model we constructed was capable of general application.

3.5 | Relationship between risk score and clinicopathological characteristics

Previous studies have shown that AFP levels, histological grade, clinical stage (TNM, BCLC and CLIP), tumour diameter and vascular invasion are correlated with the prognosis of HCC.[19-22] The present study analysed the correlation between risk score and the prognostic factors by making use of the clinical data of the four independent cohorts. Results showed that higher-risk scores were significantly

TABLE 4 The chi-square test of the relation between risk score and clinical features in ICGC cohort

| Clinical feature | Risk score | | χ^2 | P-value |
	High risk n (%)	Low risk n (%)		
Survival status				
Alive	91 (48.15%)	98 (51.85%)	6.886	0.009
Dead	29 (70.73%)	12 (29.27%)		
Gender				
Male	85 (50.30%)	84 (49.70%)	0.901	0.343
Female	35 (57.38%)	26 (42.62%)		
Age				
>65	75 (53.19%)	66 (46.81%)	0.151	0.698
≤65	45 (50.56%)	44 (49.43%)		
Prior malignancy				
Yes	18 (60%)	12 (40%)	0.847	0.357
No	102 (51%)	98 (49%)		
Stage				
I-II	66 (46.48%)	76 (53.52%)	4.824	0.028
III-IV	54 (61.36%)	34 (38.64%)		

associated with higher AFP levels, higher histological grade, larger tumour diameter, vascular invasion and advanced clinical staging (TNM, BCLC and CLIP) (Figure 5). We also performed the chi-square tests on different risk groups for correlation analysis of clinical features. The results showed that the two groups were remarkably different regarding the clinical stage, histological grade, number of tumours, AFP levels and survival status (Tables 3-6).

3.6 | Gene Ontology functional enrichment analysis

We used Gene Ontology (GO) functional enrichment analysis to annotate the function of DEGs (Figure 6A) in the group with a high risk and the group with a low risk. As found, the genes significantly upregulated in the group with a high risk were associated with a variety of immune regulatory processes including regulation of lymphocyte activation, regulation of immune effector processes and humoral immune response (Figure 6B). The estimation of immune cell infiltration used by TIMER method showed that the infiltration level of six types of immune cells in high-risk group was all higher than that in low-risk group (Figure 6C), indicating that the prognostic signature may affect the prognosis of HCC patients through regulating immune microenvironment of tumour.

Clinical feature	Risk score		χ^2	P-value
	High risk n (%)	Low risk n (%)		
Survival status				
Alive	54 (46.24%)	78 (53.76%)	8.5	0.004
Dead	52 (49.59%)	33 (50.41%)		
Gender				
Male	91 (50.30%)	98 (49.70%)	0.287	0.592
Female	15 (57.38%)	13 (42.62%)		
Age				
>65	4 (53.19%)	15 (46.81%)	7.715	0.005
≤65	102 (50.56%)	96 (49.43%)		
ALT				
>50 U/L	50 (56.18%)	39 (53.82%)	3.471	0.062
≤50 U/L	55 (43.31%)	72 (56.69%)		
AFP				
>300 ng/mL	65 (67.01%)	32 (32.99%)	23.157	<0.001
≤300 ng/mL	41 (34.17%)	79 (65.83%)		
Stage_TNM				
I-II	72 (46.48%)	96 (53.52%)	10.686	0.001
III-IV	34 (61.36%)	15 (38.64%)		
Main tumour size				
>5 cm	44 (57.14%)	33 (42.86%)	3.487	0.062
≤5 cm	61 (43.88%)	78 (56.12%)		
Multitumour				
Solitary	75 (44.38%)	94 (55.62%)	5.569	0.018
Multiple	30 (63.83%)	17 (36.17%)		
Cirrhosis				
Yes	100 (50.25%)	99 (49.75%)	2.723	0.099
No	5 (29.41%)	12 (70.59%)		
Stage_BCLC				
0-A	70 (42.68%)	94 (57.32%)	9.584	0.002
B-C	35 (67.31%)	17 (32.69%)		
CLIP_Score				
≥2	35 (72.92%)	13 (27.08%)	14.595	<0.01
<2	70 (41.67%)	98 (58.33%)		

3.7 | Relationship between risk score and immune cell infiltration

We further analysed the correlation between the prognostic model and immune cells' infiltration, and found that the risk score exhibited a positive relation to the infiltration of six types of immune cells: CD4 T cells ($r = 0.175$, $P = 0.001$), CD8 T cells ($r = 0.295$, $P < 0.001$), B cells ($r = 0.256$, $P < 0.001$), macrophages ($r = 0.472$, $P < 0.001$), dendritic cells ($r = 0.412$, $P < 0.001$) and neutrophils ($r = 0.448$, $P < 0.001$) (Figure 7A). CTNNB1-mutated HCC is characterized by immune rejection,[23] and a recent clinical study showed that this subgroup is not sensitive to the treatment

of immune-checkpoint inhibitors.[8] Our study found that the risk score positively corrected with the expression levels of six major immune checkpoints (PDL1, LAG3, CTLA4, CD276, PDCD1 and TIGIT) (Figure 7B), which were consistent with the risk score of the CTNNB1-mutated subgroup, which was lower than that of the non-mutated subgroup (Figure 7C).

3.8 | Relationship between risk score and tumour immune microenvironment

We used ssGSEA to further explore the internal relationship between prognostic model and tumour immune microenvironment (TIME). In

TABLE 6 The chi-square test of the relation between risk score and clinical features in GSE116174 cohort

Clinical feature	Risk score		χ²	P-value
	High risk n (%)	Low risk n (%)		
Survival status				
Alive	14 (37.84%)	23 (62.16%)	5.189	0.023
Dead	18 (66.67%)	9 (33.33%)		
Gender				
Male	3 (50%)	3 (50%)	0.184	0.668
Female	29 (50%)	29 (50%)		
Age				
>65	3 (33.33%)	6 (66.67%)	2.283	0.131
≤65	29 (52.73%)	26 (47.27%)		
HBV				
+	21 (44.68%)	26 (55.32%)	2.003	0.157
−	11 (64.71%)	6 (35.29%)		
Clinical stage				
I-II	24 (45.28%)	29 (54.72%)	3.952	0.047
III	8 (72.73%)	3 (27.27%)		
Alcohol				
Yes	4 (30.77%)	9 (69.23%)	3.475	0.062
No	28 (54.90%)	23 (45.10%)		
Invasion				
Yes	14 (48.28%)	15 (51.72%)	0.063	0.802
No	18 (51.43%)	17 (48.57%)		

terms of immune cell infiltration, the abundance of Tregs, aDCs, Th2 cells and macrophages of the high-risk group was significantly higher than those of the low-risk group in the four independent cohorts (Figure S1A-D). In terms of immune function, the results revealed that the type II IFN response in the low-risk group was significantly stronger than that in the low-risk group (Figure S1A-D). In addition, the expression of immune checkpoint-related gene set was up-regulated in the high-risk group, which was consistent with the previous results (Figure S1A-B).

4 | DISCUSSION

In the past few decades, considerable progress has been made in understanding the risk factors and molecular characteristics of HCC.[24-26] However, morbidity and cancer-specific mortality continue to increase in many countries.[26,27] The existing prognostic staging system still has many limitations in guiding accurate treatment and predicting more accurate clinical outcome.[28,29] We need to further improve the existing prognostic staging system using gene sequencing technology, so that patients can benefit from the accurate treatment.[8] There is increasing evidence that metabolic reprogramming can remarkably affect HCC development and may be related to advanced diseases and adverse clinical outcome[30,31]; hence, it seems

that targeting tumour metabolism can promisingly assist in effectively treating HCC.

A recent report found that proteins involved in different metabolic activities such as drug metabolism, gluconeogenesis, glycolysis and amino acid metabolism are enriched in CTNNB1 mutant HCC.[11] NRF2 mutation is an early event in the carcinogenesis of HCC and is considered to be an important factor in promoting the progression of precancerous hepatocytes to HCC.[32] Existing studies have found that the abnormal transduction of NRF2 signal pathway is associated with simvastatin.[33,34] In contrast, CTNNB1 mutation usually occurs in the later stage of HCC progression, and the relevant research on its specific mechanism is still lacking. Considering this, we conducted GSEA of mRNA expression profiles in the TCGA database. According to whether CTNNB1 was mutated, 374 tumour tissues were divided into two groups, among which 98 were mutated and 276 were not mutated. Interestingly, the significantly enriched genes in the mutant group were associated with 12 metabolic pathways. We used univariate Cox regression, K-M survival analysis, Lasso regression and multi–Cox regression to analyse metabolic genes step by step, and ultimately constructed a five-metabolic gene (including AMD1, PRDX1, ALAS1, CYP3A5 and GOT2) risk score model. The unified risk score formula and the threshold value of risk classification were taken into account for dividing all included patients in group with a high risk and in group with a low risk. We first conducted internal validation of the prognostic model based on the clinical characteristics of the TCGA cohort such as the presence of a new tumour after initial treatment, individual tumour status, AJCC stage, histological grade, vascular invasion and AFP levels. The patients were then divided into 12 subgroups. Based on the K-M survival curve, in each subgroup, the group with a high risk exhibited an obviously lower OS relative to the group with a low risk. Subsequently, we conducted external verification of the prognostic model by using three independent queues, namely ICGC (n = 230), GSE14520 (n = 239) and GSE116174 (n = 64). Consistent with the TCGA results, the group with a high risk showed an obviously worse prognosis relative to the group with a low risk. As demonstrated by the univariate and multivariate Cox regression analysis, this prognostic model could independently predict the prognosis of HCC. These results support that our prognostic model has a strong general applicability. GO enrichment analysis showed that genes with up-regulated expression in the group with a high risk showed a close relation to the immune regulation. The correlation between tumours with infiltrating macrophages and risk score was the most prominent. It is well known that macrophages are the most abundant in tumour tissues and significantly regulate tumours, which is be capable of promoting tumour cells in terms of the proliferation, invasion and metastasis as well as inducing immune tolerance in these cells.[35-37] As expected, the risk score presented a positive relation to the expression levels of six commonly used immune checkpoints (PDL1, LAG3, CTLA4, CD276, PDCD1 and TIGIT). We conclude that the immunosuppressive tumour microenvironment may be
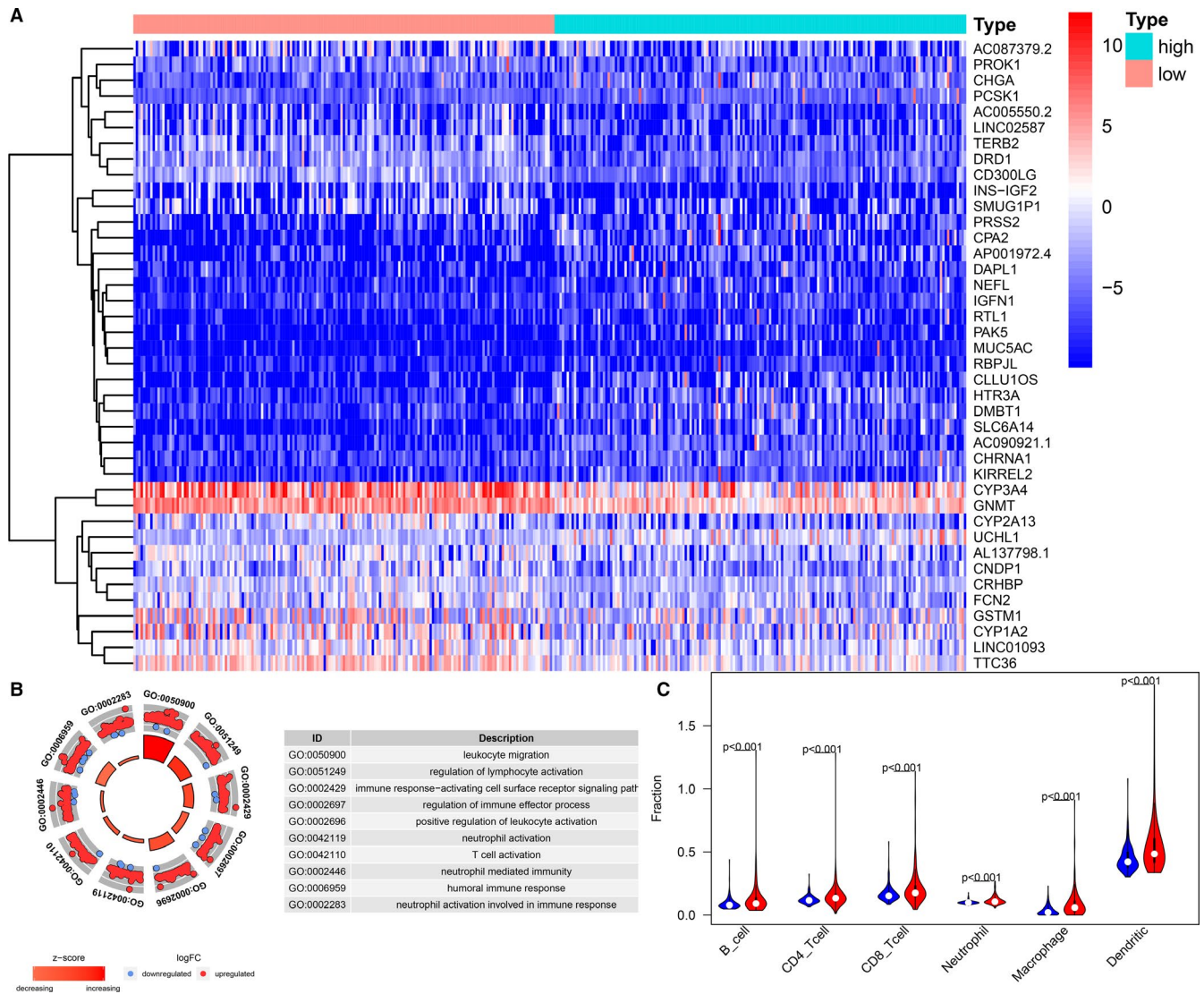
FIGURE 6 Identification of differentially expressed genes in high-risk and low-risk HCC patients. A, Heat map of differentially expressed genes samples between high-risk and low-risk groups. B, GO circle plot of immune pathways differentially enriched in patients with different risk scores. C, Violin plot of relationships between the risk score and infiltration abundances of six types of immune cells (red represents high-risk group, and blue represents low-risk group)

a major factor contributing to poor prognosis in the group with a high risk. It is noteworthy that the risk score is most strongly associated with the emerging immune checkpoint CD276 (B7-H3), which has recently been reported to be highly overexpressed in a variety of human solid cancers and is often associated with adverse clinical outcome in patients, making it a potential target for tumour immunotherapy.[38,39] Besides, high-risk score could reflect adverse survival outcome-related clinical characteristics, such as AFP levels (>300 ng/mL), tumour vascular invasion, low tumour differentiation, advanced AJCC stage, advanced BCLC stage, advanced CLIP stage, new tumour after initial treatment, main tumour size > 5 cm and multiple tumours, which may also help to explain the reason for the poor prognosis of high-risk patients.

Adenosylmethionine decarboxylase 1 (AMD1) acts as an essential enzyme that affects the biosynthesis of polyamines including spermine. Zabala-Letona found that the up-regulation of AMD1 could activate the PTEN-PI3K-mTORC1 pathway to maintain the growth and proliferation of prostate cancer cells.[40] Furthermore, Xu reported that AMD1 has tumorigenic effects on the prognosis of human gastric cancer, but the potential function of AMD1 in HCC is still unclear.[41] Peroxiredoxin 1 (PRDX1) acts as a peroxidase family member of the antioxidant enzymes. Fang Y and Sun et al found that high expression of PRDX1 in HCC tissues corresponds to adverse clinical outcome, and the mechanism may be related to promoting tumour angiogenesis and regulating cell migration and invasion.[42,43] 5'-Aminolevulinate synthase 1 (ALAS1) plays the role of an rate-limiting enzyme during haem biosynthesis. As demonstrated by studies performed recently, ALAS1 can affect many cellular functions and has important effects on non–small cell lung cancer, colorectal cancer and oral cancer.[44,45] However, the role of ALAS1 in HCC remains
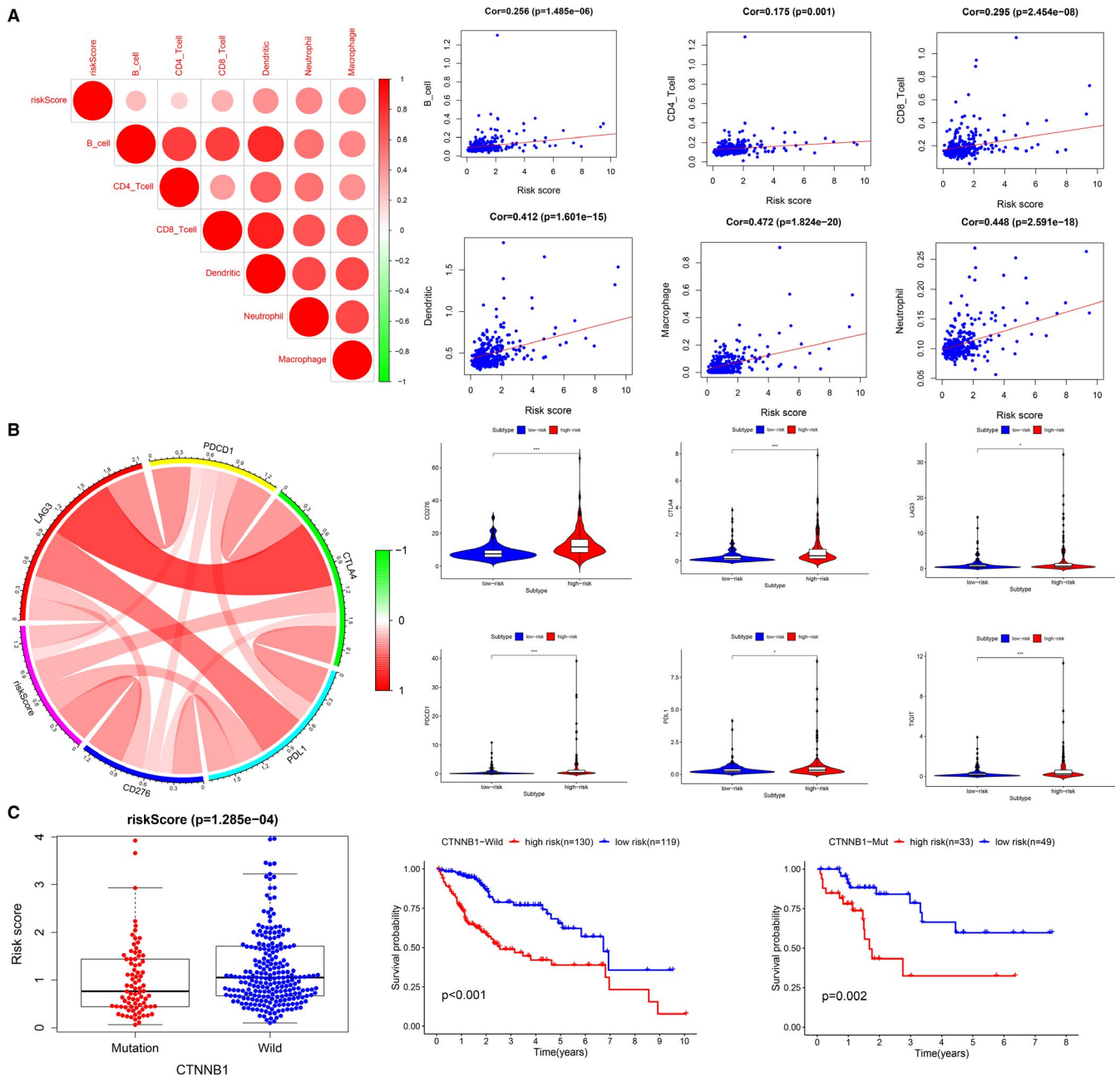
FIGURE 7 The landscape of immune infiltration in high- and low-risk HCC patients. A, Correlation analysis of risk score and immune cell infiltration. B, The relationship between risk score and expression level of immune checkpoint. C, The relationship between risk score and CTNNB1 status (*P*-value significant codes: 0 ≤ *** < 0.001 ≤ ** < 0.01 ≤ * < 0.05 ≤ . < 0.1)

not well known. CYP3A5 is a cytochrome P450 protein that plays a role in the metabolism of many carcinogens and anticancer drugs in the liver. Jiang et al found that CYP3A5 plays an antitumour role in HCC by regulating the mTORC2/Akt signalling pathway.[46] Glutamate oxaloacetate transaminase 2 (GOT2) has been repeatedly reported in recent years to be associated with the progression of pancreatic cancer,[47,48] but its role in HCC remains unclear.

Although the above studies support that our prognostic model has a high potential for clinical application, the study faces some limitations. Although the accumulated data of high-throughput analysis

from a large number of samples have been optimally applied, further verification through prospective studies is necessary. Furthermore, the specific biological functions of the five genes in HCC need to be explored experimentally.

5 | CONCLUSIONS

We constructed a novel prognostic model that is useful for further improving the prognostic evaluation system of patients with HCC,

and the five metabolic genes in this model are expected to become potential targets for the treatment of HCC.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

AUTHOR CONTRIBUTION

Junyu Huo: Conceptualization (equal); Data curation (equal); Formal analysis (lead); Investigation (lead); Methodology (lead); Resources (lead); Software (lead); Supervision (equal); Validation (equal); Visualization (lead); Writing-original draft (lead); Writing-review & editing (equal). **Liqun Wu:** Conceptualization (equal); Project administration (equal); Supervision (equal); Validation (equal); Writing-review & editing (equal). **Yunjin Zang:** Conceptualization (supporting); Data curation (supporting); Project administration (supporting); Writing-review & editing (supporting).

DATA AVAILABILITY STATEMENT

The data sets analysed for this study were obtained from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/), the International Cancer Genome Consortium (ICGC) (https://icgc.org/), the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) and UCSC Xena website (https://xenabrowser.net/).

ORCID

Junyu Huo https://orcid.org/0000-0002-1345-1893

REFERENCES

1. Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. Genetic landscape and biomarkers of hepatocellular carcinoma. *Gastroenterology*. 2015;149(5):1226-1239.e4.
2. Nakagawa H, Fujita M, Fujimoto A. Genome sequencing analysis of liver cancer for precision medicine. *Semin Cancer Biol*. 2019;55:120-127.
3. Senni N, Savall M, Cabrerizo Granados D, et al. beta-catenin-activated hepatocellular carcinomas are addicted to fatty acids. *Gut*. 2019;68(2):322-334.
4. Cleary SP, Jeck WR, Zhao X, et al. Identification of driver genes in hepatocellular carcinoma by exome sequencing. *Hepatology (Baltimore, MD)*. 2013;58(5):1693-1702.
5. Hlady RA, Sathyanarayan A, Thompson JJ, et al. Integrating the epigenome to identify drivers of hepatocellular carcinoma. *Hepatology (Baltimore, MD)*. 2019;69(2):639-652.
6. Bayo J, Fiore EJ, Dominguez LM, et al. A comprehensive study of epigenetic alterations in hepatocellular carcinoma identifies potential therapeutic targets. *J Hepatol*. 2019;71(1):78-90.
7. Nakagawa H, Shibata T. Comprehensive genome sequencing of the liver cancer genome. *Cancer Lett*. 2013;340(2):234-240.
8. Rebouissou S, Nault JC. Advances in molecular classification and precision oncology in hepatocellular carcinoma. *J Hepatol*. 2020;72(2):215-229.
9. Rebouissou S, Franconi A, Calderaro J, et al. Genotype-phenotype correlation of CTNNB1 mutations reveals different ss-catenin activity associated with liver tumor progression. *Hepatology (Baltimore, MD)*. 2016;64(6):2047-2061.
10. Calderaro J, Couchy G, Imbeaud S, et al. Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. *J Hepatol*. 2017;67(4):727-738.
11. Gao Q, Zhu H, Dong L, et al. Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell*. 2019;179(2):561-577.e22.
12. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-674.
13. Ward PS, Thompson CB. Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. *Cancer Cell*. 2012;21(3):297-308.
14. Icard P, Shulman S, Farhat D, Steyaert JM, Alifano M, Lincet H. How the Warburg effect supports aggressiveness and drug resistance of cancer cells? *Drug Resist Updat*. 2018;38:1-11.
15. Long J, Wang A, Bai Y, et al. Development and validation of a TP53-associated immune prognostic model for hepatocellular carcinoma. *EBioMedicine*. 2019;42:363-374.
16. Li G, Xu W, Zhang L, et al. Development and validation of a CIMP-associated prognostic model for hepatocellular carcinoma. *EBioMedicine*. 2019;47:128-141.
17. Liu F, Liao Z, Song J, et al. Genome-wide screening diagnostic biomarkers and the construction of prognostic model of hepatocellular carcinoma. *J Cell Biochem*. 2020;121(3):2582-2594.
18. Harding JJ, Nandakumar S, Armenia J, et al. Prospective genotyping of hepatocellular carcinoma: clinical implications of next-generation sequencing for matching patients to targeted and immune therapies. *Clin Cancer Res*. 2019;25(7):2116-2126.
19. Hanazaki K, Kajikawa S, Koide N, Adachi W, Amano J. Prognostic factors after hepatic resection for hepatocellular carcinoma with hepatitis C viral infection: univariate and multivariate analysis. *Am J Gastroenterol*. 2001;96(4):1243-1250.
20. Cillo U, Vitale A, Grigoletto F, et al. Prospective validation of the Barcelona Clinic Liver Cancer staging system. *J Hepatol*. 2006;44(4):723-731.
21. Llovet JM, Bruix JJH. Prospective validation of the Cancer of the Liver Italian Program (CLIP) score: a new prognostic system for patients with cirrhosis and hepatocellular carcinoma. *Hepatology*. 2000;32(3):679-680.
22. Leung TW, Tang AM, Zee B, et al. Construction of the Chinese University Prognostic Index for hepatocellular carcinoma and comparison with the TNM staging system, the Okuda staging system, and the Cancer of the Liver Italian Program staging system: a study based on 926 patients. *Cancer*. 2002;94(6):1760-1769.
23. Kurebayashi Y, Ojima H, Tsujikawa H, et al. Landscape of immune microenvironment in hepatocellular carcinoma and its additional impact on histological and molecular classification. *Hepatology (Baltimore, MD)*. 2018;68(3):1025-1041.
24. Zheng R, Qu C, Zhang S, et al. Liver cancer incidence and mortality in China: temporal trends and projections to 2030. *Chin J Cancer Res*. 2018;30(6):571-579.
25. Bruix J, da Fonseca LG, Reig M. Insights into the success and failure of systemic therapy for hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol*. 2019;16(10):617-630.
26. Park JW, Chen M, Colombo M, et al. Global patterns of hepatocellular carcinoma management from diagnosis to death: the BRIDGE Study. *Liver Int*. 2015;35(9):2155-2166.
27. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol*. 2019;16(10):589-604.
28. Hoshida Y, Moeini A, Alsinet C, Kojima K, Villanueva A. Gene signatures in the management of hepatocellular carcinoma. *Semin Oncol*. 2012;39(4):473-485.
29. Huitzil-Melendez FD, Capanu M, O'Reilly EM, et al. Advanced hepatocellular carcinoma: which staging systems best predict prognosis? *J Clin Oncol*. 2010;28(17):2889-2895.
30. Lee NCW, Carella MA, Papa S, Bubici C. High expression of glycolytic genes in cirrhosis correlates with the risk of developing liver cancer. *Front Cell Dev Biol*. 2018;6:138.

31. Satriano L, Lewinska M, Rodrigues PM, Banales JM, Andersen JB. Metabolic rearrangements in primary liver cancers: cause and consequences. *Nat Rev Gastroenterol Hepatol*. 2019;16(12):748-766.

32. Zavattari P, Perra A, Menegon S, et al. Nrf2, but not β-catenin, mutation represents an early event in rat hepatocarcinogenesis. *Hepatology*. 2015;62(3):851-862.

33. Jang HJ, Hong EM, Kim M, et al. Simvastatin induces heme oxygenase-1 via NF-E2-related factor 2 (Nrf2) activation through ERK and PI3K/Akt pathway in colon cancer. *Oncotarget*. 2016;7(29):46219.

34. Habeos IG, Ziros PG, Chartoumpekis D, Psyrogiannis A, Kyriazopoulou V, Papavassiliou AG. Simvastatin activates Keap1/Nrf2 signaling in rat liver. *J Mol Med*. 2008;86(11):1279.

35. Mantovani A, Marchesi F, Malesci A, Laghi L, Allavena P. Tumour-associated macrophages as treatment targets in oncology. *Nat Rev Clin Oncol*. 2017;14(7):399-416.

36. Degroote H, Van Dierendonck A, Geerts A, Van Vlierberghe H, Devisscher L. Preclinical and clinical therapeutic strategies affecting tumor-associated macrophages in hepatocellular carcinoma. *J Immunol Res*. 2018;2018:7819520.

37. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol*. 2013;14(10):1014-1022.

38. Picarda E, Ohaegbulam KC, Zang X. Molecular pathways: targeting B7–H3 (CD276) for human cancer immunotherapy. *Clin Cancer Res*. 2016;22(14):3425-3431.

39. Brown ZJ, Greten TF, Heinrich B. Adjuvant treatment of hepatocellular carcinoma: prospect of immunotherapy. *Hepatology (Baltimore, MD)*. 2019;70(4):1437-1442.

40. Zabala-Letona A, Arruabarrena-Aristorena A, Martin-Martin N, et al. mTORC1-dependent AMD1 regulation sustains polyamine metabolism in prostate cancer. *Nature*. 2017;547(7661):109-113.

41. Xu L, You X, Cao Q, et al. Polyamine synthesis enzyme AMD1 is closely associated with tumorigenesis and prognosis of human gastric cancers. *Carcinogenesis*. 2019;41(2):214-222.

42. Fang Y, He J, Janssen H, Wu J, Dong L, Shen XZ. Peroxiredoxin 1, restraining cell migration and invasion, is involved in hepatocellular carcinoma recurrence. *J Dig Dis*. 2018;19(3):155.

43. Sun Q-K, Zhu J-Y, Wang W, et al. Diagnostic and prognostic significance of peroxiredoxin 1 expression in human hepatocellular carcinoma. *Med Oncol*. 2014;31(1):786.

44. Zhao Y, Zhang X, Liu Y, et al. Inhibition of ALAS1 activity exerts anti-tumour effects on colorectal cancer in vitro. *Saudi J Gastroenterol*. 2020;26(3):144-152.

45. Sarkar R, Kishida S, Kishida M, et al. Effect of cigarette smoke extract on mitochondrial heme-metabolism: an in vitro model of oral cancer progression. *Toxicol In Vitro*. 2019;60:336-346.

46. Jiang F, Chen L, Yang Y-C, et al. CYP3A5 functions as a tumor suppressor in hepatocellular carcinoma by regulating mTORC2/Akt signaling. *Cancer Res*. 2015;75(7):1470-1481.

47. Yang H, Zhou L, Shi Q, et al. SIRT3-dependent GOT2 acetylation status affects the malate-aspartate NADH shuttle activity and pancreatic tumor growth. *EMBO J*. 2015;34(8):1110-1125.

48. Son J, Lyssiotis CA, Ying H, et al. Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. *Nature*. 2013;496(7443):101-105.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.