Original Article

Co-amplification of genes in chromosome 8q24: a robust prognostic marker in hepatocellular carcinoma

Yongjian Zheng[1#], Yuan Cheng[1#], Cheng Zhang[1#], Shunjun Fu[1], Guolin He[1], Lei Cai[1], Ling Qiu[2], Kunhua Huang[1], Qunhui Chen[1], Wenzhuan Xie[3], Tingting Chen[3], Mengli Huang[3], Yuezong Bai[3], Mingxin Pan[1]

[1]Second Department of Hepatobiliary Surgery, Zhujiang Hospital of Southern Medical University, Guangzhou, China; [2]Second Department of Surgery, Dongfeng People's Hospital, Guangzhou, China; [3]The Research and Development Center of Precision Medicine, 3D Medicines Inc., Shanghai, China

Contributions: (I) Conception and design: M Pan; (II) Administrative support: Y Cheng; (III) Provision of study materials or patients: G He, L Cai; (IV) Collection and assembly of data: M Huang, Y Bai, S Fu; (V) Data analysis and interpretation: Y Zheng, C Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Mingxin Pan. Second Department of Hepatobiliary Surgery, Zhujiang Hospital of Southern Medical University, Guangzhou, China. Email: pmxwxy@sohu.com.

Background: Hepatocellular carcinoma (HCC) is a leading cause of tumor-associated death worldwide, owing to its high 5-year postoperative recurrence rate and inter-individual heterogeneity. Thus, a prognostic model is urgently needed for patients with HCC. Several researches have reported that copy number amplification of the 8q24 chromosomal region is associated with low survival in many cancers. In the present work, we set out to construct a multi-gene model for prognostic prediction in HCC.

Methods: RNA sequencing and copy number variant data of tumor tissue samples of HCC from The Cancer Genome Atlas (n=328) were used to identify differentially expressed messenger RNAs of genes located on the chromosomal 8q24 region by the Wilcox test. Univariate Cox and Lasso-Cox regression analyses were carried out for the screening and construction of a prognostic multi-gene signature in The Cancer Genome Atlas cohort (n=119). The multi-gene signature was validated in a cohort from the International Cancer Genome Consortium (n=240). A nomogram for prognostic prediction was built, and the underpinning molecular mechanisms were studied by Gene Set Enrichment Analysis.

Results: We successfully established a 7-gene prognostic signature model to predict the prognosis of patients with HCC. Using the model, we divided individuals into high-risk and low-risk sets, which showed a significant difference in overall survival in the training dataset (HR =0.17, 95% CI: 0.1–0.28; P<0.001) and in the testing dataset (HR = 0.42, 95% CI: 0.23–0.74; P=0.002). Multivariate Cox regression analysis showed the signature to be an independent prognostic factor of HCC survival. A nomogram including the prognostic signature was constructed and showed a better predictive performance in short-term (1 and 3 years) than in long-term (5 years) survival. Furthermore, Gene Set Enrichment Analysis identified several pathways of significance, which may aid in explaining the underlying molecular mechanism.

Conclusions: Our 7-gene signature is a reliable prognostic marker for HCC, which may provide meaningful information for therapeutic customization and treatment-related decision making.

Keywords: Hepatocellular carcinoma (HCC); 8q24; amplification; The Cancer Genome Atlas (TCGA); International Cancer Genome Consortium (ICGC); prognostic model

Submitted Mar 16, 2021. Accepted for publication May 06, 2021.
doi: 10.21037/jgo-21-205
View this article at: http://dx.doi.org/10.21037/jgo-21-205

Introduction

Hepatocellular carcinoma (HCC) ranks among the most prevalent liver malignancies globally, and its morbidity is rising in most areas of the world (1). HCC has an exceedingly poor prognosis, with only 12% of patients with the disease surviving more than 5 years (2-4). Thus, prognostic models are urgently needed to aid in predicting the outcomes of HCC and to inform individualized management decisions for patients. The majority of studies that have attempted to construct a prognostic model for HCC to date have used clinical characteristics and serum tumor markers (5-7). However, with the rapid development of the genome sequencing technologies, evidence has accumulated showing gene signatures and traditional parameters to have immense promise for prognostic prediction in HCC. Li *et al.*, for instance, constructed a 3-gene prediction signature which proved effective in identifying HCC patients with a high mortality risk (8). Also, Bai *et al.* identified a 6-microRNA signature as an independent factor to predict HCC recurrence (9). Furthermore, a number of other studies constructed messenger RNA (mRNA) expression signatures including different numbers of genes by adopting similar methods (10-13).

Accumulating research suggests that tumors are caused by multiple gene mutations occurring sequentially in cell lines, at chromosomal and nucleotide levels. Chromosomal DNA amplification is among the molecular mechanisms capable of changing genes. The overexpression of mRNA contributes to the occurrence and progression of a variety of tumors (14,15). Amplification of the chromosome 8q24 is frequently detected in human cancer, and several studies have emphasized the significance of this subchromosomal region in the occurrence and progression of cancers, including prostate (16,17), gastrointestinal (18,19), and breast (20) cancer. For example, Cussenot *et al.* revealed that 8q24 amplification at the Myc locus correlated with Myc protein expression and was associated with disease progression and biochemical recurrence after radical prostatectomy in prostate cancer (16). Besides, the 8q24 chromosome locus is one of the most common amplified genomic regions in HCC (18,21-23).

Although the functional annotation genes in this region have yet to be fully expounded, several studies have reported the amplification of some genes separately located on the 8q24 chromosome region, such as MYC (24), MAL2 (22), and BOP1 (4), which have been linked to HCC occurrence and progression in recent years. Through informatics analysis of genomic data from 150 HCC tumors, we further confirmed that an increase of 8q24 is the major contributing event in the occurrence and development of HCC (25). Zeng *et al.* has been genotyped the rs9642880 G >T polymorphism using DNA isolated from blood samples of 271 hepatocellular carcinoma (HCC) patients who received radiotherapy treatment (21). The results showed that patients who carried the GT or TT genotypes had significantly shorter median survival times compared to patients with the GG genotype. Recently Zhang *et al.* has been reported that Copy number gains and amplification of chromosome 8q24.3 in HCC were determined to be positively correlated with the PRL-3 overexpression, which provided an example of how co-amplified genes work together in HCC (26). In our previous study, we also retrospectively analyzed the genetic testing and clinical follow-up information of 80 patients with HCC. Our results indicated that a gain of 8q24 was significantly associated with poor survival of HCC ($P<0.05$), with an alteration frequency of 22–30%.

In recent years, the construction of an effective and powerful predictive model which can provide valuable information to inform clinical medical decisions in a convenient way has drawn much attention. In the present study, we focused on copy number variation of genes located on the chromosomal 8q24 region to build a prognostic prediction model for HCC. The mRNA expression data of HCC from The Cancer Genome Atlas (TGCA) were analyzed to identify differentially expressed genes (DEGs). Lasso-Cox regression analysis was performed to build a 7-gene signature, which we validated in an International Cancer Genome Consortium (ICGC) cohort. A nomogram combining the 7-gene prognostic signature and clinical prognostic factors to predict survival was finally constructed. Gene Set Enrichment Analysis (GSEA) was applied for further analysis of the mechanisms potentially underlying the predictive effect of our model. We present the following article in accordance with the REMARK reporting checklist (available at http://dx.doi.org/10.21037/jgo-21-205).

Methods

Data collection

We collected mRNA expression and clinical data from the TCGA-Liver Hepatocellular Carcinoma dataset (TCGA-LIHC, https://portal.gdc.cancer.gov/), cBioportal for Cancer Genomics, and International Cancer Genome Consortium (ICGC, http://cbioportal.org/) (27). The study was conducted in accordance with the Declaration

of Helsinki (as revised in 2013). All data were obtained from publicly available, open-access databases; All data can be found here: https://portal.gdc.cancer.gov/repository, https://dcc.icgc.org/projects/LIRI-JP. thus, additional ethical approval was not required.

Identifying and intersecting differentially expressed mRNAs (DEM) in the TCGA-LIHC dataset

The transcripts per million (TPM) method was applied to standardize raw counts of HCC mRNA expression data, which were subjected to log2 conversion. A total of 167 protein-coding genes located in the chromosome 8q24 region were accurately annotated. The R package limma (version 3.36.2) was used to identify the DEMs. The DEMs which had an absolute log2 fold change (FC) >1 and an adjusted P value <0.05 were included in subsequent analyses (https://cancerci.biomedcenral.com/).

Construction of the prognostic gene signature

The prognostic gene signature model was established in a discovery cohort comprising 331 patients from TCGA. For the analysis, we only selected patients with initial treatment whose overall survival (OS) exceeded 1 month. Patients with amplification of ≥2 genes located in the chromosome 8q24 region were assigned to the amplification group (mutation group), while those with amplification of 0–1 gene were included in the non-amplification group (wild-type group). DEGs were identified by univariate Cox regression analysis. Genes for predicting OS of HCC were further screened by Lasso-penalized Cox regression analysis designed using the R package glmnet (28). Finally, a 7-gene signature prognosis model was constructed in the TCGA discovery cohort on the basis of a linear combination of the regression coefficient of the Lasso-Cox regression model coefficients (β) and the mRNA expression level. The prognostic gene signature was presented as risk score = (β_{mRNA1}*expression level of mRNA1) + (β_{mRNA2}*expression level of mRNA2) +…+ (β_{mRNAn}*expression level of mRNAn). For verification of the prediction efficiency of the predictive gene signature, patients form TCGA cohort were randomly separated into training and testing datasets at a ratio of 6:4 using the createDataPartition function in the R package caret. The optimal cutoff values were investigated using the survminer R package. Subsequently, we divided patients into groups with high and low risk based on the correlation of the signature gene expression levels with OS in the training dataset. With the survival R package, we analyzed the survival difference between patients with high and low risk scores by combining Kaplan–Meier (KM) survival curve analysis with the log-rank test. We used time-dependent receiver operating characteristic (ROC) curve analysis and the concordance index (C-index) to assess the accuracy of the prognostic signature in predicting 1-, 3-, and 5-year survival. The prognostic gene signature's estimated value was further verified in the testing dataset and TCGA discovery cohort. Using the Wilcoxon signed-rank test, we also performed an analysis in the TCGA discovery cohort to compare the mRNA expression levels of the 7 prognosis-related genes included in the signature in the high- and low-risk groups.

External validation of the prognostic gene signature

To validate the prognostic gene signature, we used another cohort comprising 243 patients with HCC from the ICGC. Patients' risk scores were calculated according to the prognostic gene characteristics. ROC curve and KM curves were drawn to validate the value of the prognostic gene signature for predicting HCC prognosis.

Independence of the prognostic gene signature from other clinical parameters in the TCGA

To determine whether the prediction efficiency of the prognostic model was able to serve independently of other clinical parameters [such as age, sex, alpha-fetoprotein (AFP), body mass index (BMI), tumor grade, vascular tumor invasion (MVI) inflammation, residual tumor, tumor mutation burden (TMB), and tumor stage], we performed univariate and multivariate analyses using a Cox regression model. P value <0.05 indicated statistical significance. We also calculated the hazard ratios (HRs) and 95% confidence intervals (CIs).

Building and validation of the predictive nomogram

Nomograms are commonly applied to predict the prognosis of cancer. All independent prognostic factors were selected for the evaluation of the probability of OS at 1, 3, and 5 years using the multivariable Cox regression method. We validated the nomogram by assessing its discrimination and calibration. To test the nomogram's predicted probability of OS against the observed rates, we plotted its calibration

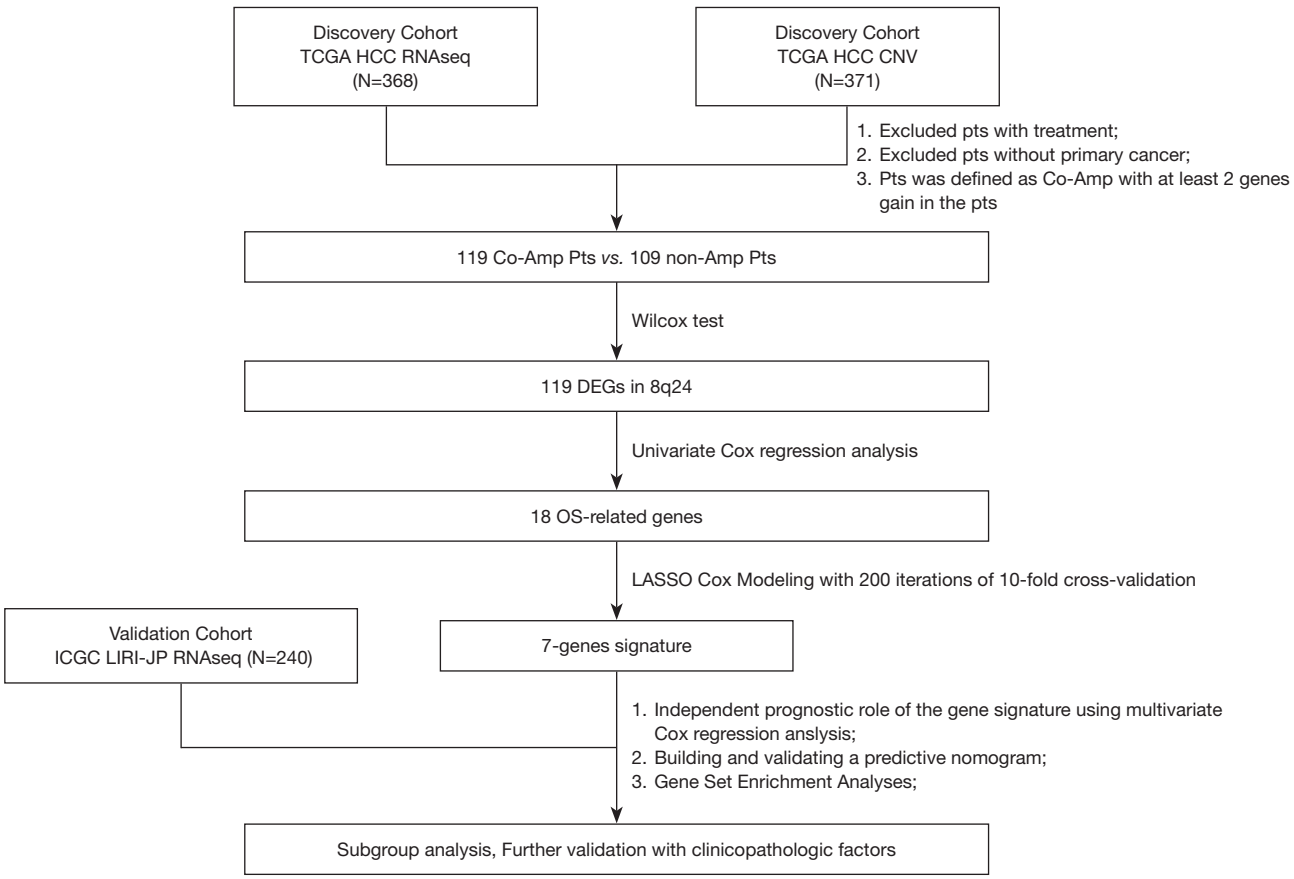

Figure 1 Flow chart of multi-gene signature identification and validation.

curve. To examine whether the nomogram was reliable in comparison with tumor stage or risk group, decision curve analysis (DCA) was carried out.

GSEA

To reveal the potential molecular mechanism of the prognostic gene signature, GSEA (v.3.0.) was used to investigate patients in high- and low-risk subgroups in the TCGA discovery cohort.

Pathways related to OS in the 2 risk groups were separately retrieved by searching Molecular Signatures Database v.7.0. We identified gene sets with a P value <0.05 and a false discovery rate (FDR) <25% as having significant enrichment.

Statistical analysis

Data were statistically analyzed with R software v3.6.0 (R Foundation for Statistical Computing, Vienna, Austria). P value <0.05 indicated statistical significance.

Results

Identified DEMs

Figure 1 shows a flow chart of the analytical process. In total, 109 genes were screened out as DEMs between the 8q24 amplification group (n=184) and the non-amplification group (n=109; Table S1).

Construction and validation of the prognostic gene signature

At random, we assigned 184 patients with primary HCC in the TCGA discovery cohort at a ratio of 6:4 to a training dataset (n=111) and a testing dataset (n=73). *Table 1* summarizes the baseline characteristics of patients in the 2 datasets. None of the clinical characteristics were significantly different between the testing dataset and training dataset.

We further narrowed candidate mRNAs using Lasso-penalized Cox analysis, finally identifying 7 target genes,

Table 1 The relationship between clinical characteristics and overall survival in the TCGA cohort

	Testing set	Training set	P value
N	73	111	
Age, years, mean (SD)	58.55 (13.53)	59.79 (10.45)	0.483
Sex (male), n (%)	56 (76.7)	82 (73.9)	0.794
Race, n (%)			0.434
Asian	41 (56.2)	61 (55.0)	
Black/African American	4 (5.5)	4 (3.6)	
Not reported	4 (5.5)	2 (1.8)	
White	24 (32.9)	44 (39.6)	
AFP, ng/mL, mean (SD)	6,296.75 (41,597.65)	32,503.79 (220,471.63)	0.385
BMI, kg/m^2, mean (SD)	25.46 (6.76)	25.36 (5.92)	0.924
Inflammation, n (%)			0.098
Mild	28 (38.4)	25 (22.5)	
None	19 (26.0)	30 (27.0)	
Severe	3 (4.1)	9 (8.1)	
N/A	23 (31.5)	47 (42.3)	
Tumor grade, n (%)			0.19
Grade 1	10 (13.7)	10 (9.0)	
Grade 2	29 (39.7)	56 (50.5)	
Grade 3	31 (42.5)	37 (33.3)	
Grade 4	2 (2.7)	8 (7.2)	
N/A	1 (1.4)	0 (0.0)	
Tumor stage, n (%)			0.302
Not reported	4 (5.5)	4 (3.6)	
Stage I	30 (41.1)	61 (55.0)	
Stage II	23 (31.5)	22 (19.8)	
Stage IIIA	15 (20.5)	20 (18.0)	
Stage IIIB	0 (0.0)	2 (1.8)	
Stage IIIC	1 (1.4)	2 (1.8)	
Residual tumor, n (%)			0.928
R0	67 (91.8)	99 (89.2)	
R1	3 (4.1)	5 (4.5)	
RX	2 (2.7)	5 (4.5)	
N/A	1 (1.4)	2 (1.8)	
Vascular tumor invasion, n (%)			0.684
Macro	3 (4.1)	3 (2.7)	
Micro	22 (30.1)	26 (23.4)	
None	36 (49.3)	63 (56.8)	
N/A	12 (16.4)	19 (17.1)	
TMB, Mut/Mb, mean (SD)	6.46 (4.41)	6.51 (6.31)	0.953

TGCA, The Cancer Genome Atlas; AFP, alpha-fetoprotein; BMI, body mass index; TMB, tumor mutation burden.

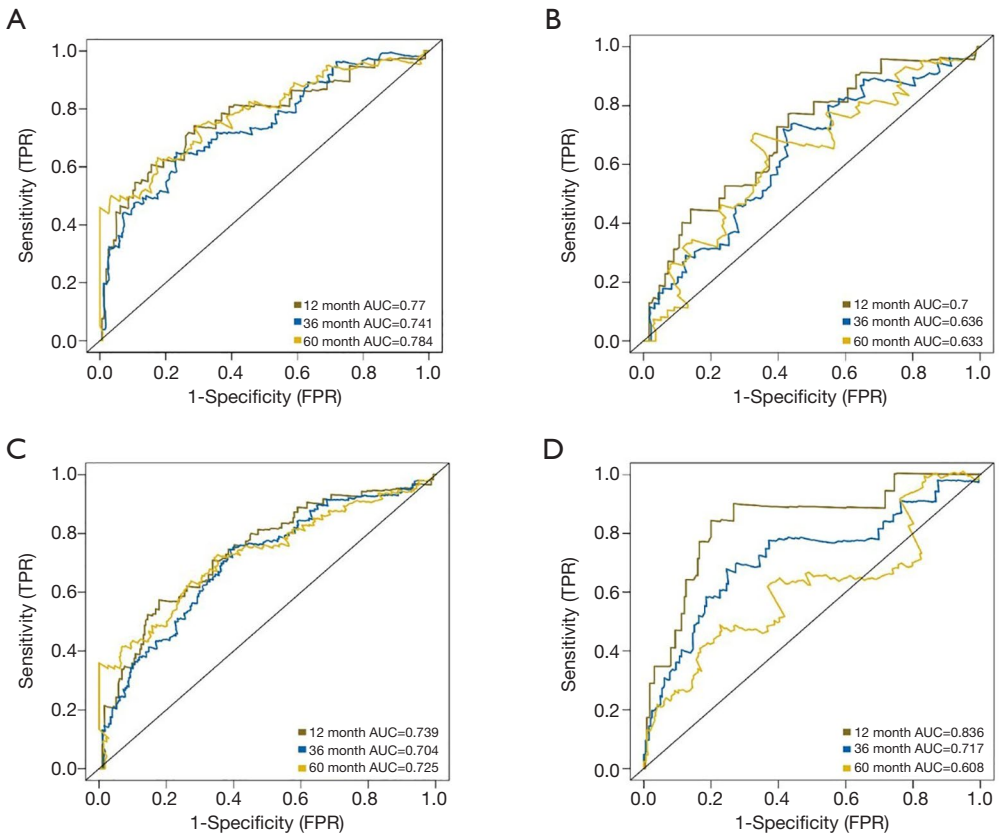

Figure 2 Time dependent receiver operating characteristic (ROC) analysis of the 7-gene signature for predicting overall survival at 1, 3, and 5 years in the training set (A), the testing set (B), the TCGA discovery cohort (C), and the International Cancer Genome Consortium (ICGC) validation cohort (D).

which were selected to build the prognostic gene signature. These genes included zinc finger protein 7 (ZNF7), trafficking protein particle complex 9 (TRAPPC9), transmembrane protein 65 (TMEM65), POU domain class 5 transcription factor 1B (POU5F1B), DEP domain-containing MTOR interacting protein (DEPTOR), BOP1 ribosomal biogenesis factor (BOP1), and Rho GTPase-activating protein 39 (ARHGAP39), which were found to have a prognostic connection HCC using the regsubsets function of the leaps R package. The risk score= $(-1.465*Expr_{ZNF7)} + (-1.317*Expr_{TRAPPC9}) + (0.432*Expr_{TMEM65}) + (-0.728*Expr_{POU5F1B}) + (-0.253*Expr_{DEPTOR}) + (0.357*Expr_{BOP1}) + (0.553*Expr_{ARHGAP39})$.

The optimal risk-score cutoff value was finally determined to be 2.074. According to this cutoff, patients in the training dataset were dissected into groups with high (n=44) and low (n=169) risk scores. Subsequently, all stratification was based on this optimal cutoff value. Time-dependent ROC and KM survival curves were calculated to evaluate the ability of our 7-gene signature to predict HCC prognosis. The same procedure was employed for both the testing and the entire TCGA datasets. The area under the ROC curve (AUC) values of the signature for OS at 1, 3, and 5 years were: 0.77, 0.741, and 0.784, respectively, in the training dataset; 0.7, 0.636, and 0.633, respectively, in the testing dataset; and 0.739, 0.704, and 0.725, respectively, in the whole TCGA cohort (*Figure 2*). In all 3 datasets, the OS was significantly poorer in the high-risk group than in the low-risk group (P<0.001, P=0.002, and P<0.001 in the training, testing, and whole TCGA datasets; *Figure 3*). This result indicated a moderately sufficient specificity of the 7-gene signature for prognostic prediction in HCC. Among the 368 HCC patients in the TCGA cohort, 64 (18%) had at least 1 genetic variation of the 7 selected genes, with amplification being the most common type of genetic variation. In TCGA cohort, the mRNA expression levels of BOP1, TMEM65, and ARHGAP39 were significantly

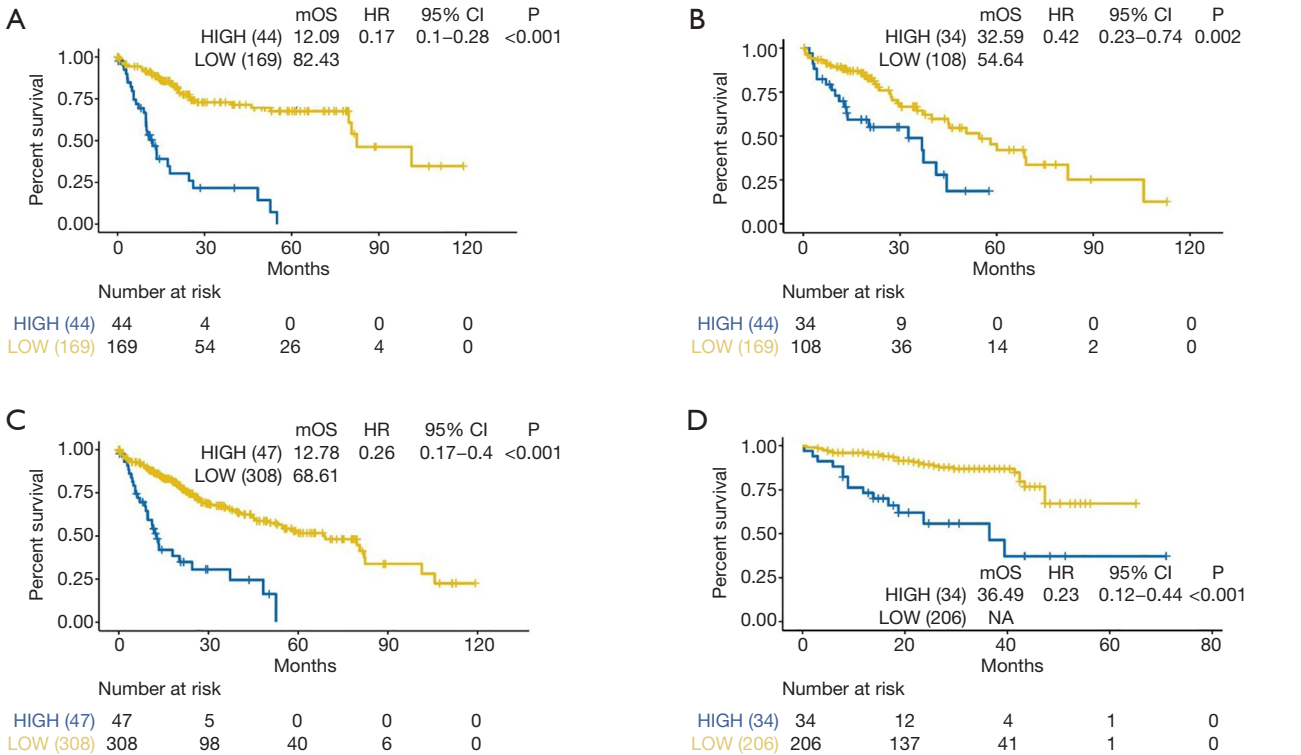

Figure 3 Kaplan–Meier analysis to compare patients between high-risk and low-risk groups based on the 7-gene signature in the training set (A), the testing set (B), The Cancer Genome Atlas (TCGA) discovery cohort (C), and (D) the International Cancer Genome Consortium (ICGC) validation cohort.

upregulated compared with those in the high-risk group; in contrast, those of TRAPPC9, POU5F1B and DEPTOR were significantly downregulated (*Figure 4*).

Externally validating the prognostic gene signature in the ICGC cohort

For further assessment of its prediction ability, we tested the 7-gene signature in the ICGC cohort. A total of 240 patients with HCC were classified as high (n=34) or low (n=206) risk using the optimal risk cutoff value described above. In accordance with our findings from the ICGC cohort, patients with a high-risk score had significantly worse OS than patients with a low-risk score (P<0.001, *Figure 2D*). The AUC for OS at 1, 3, and 5 years was 0.836, 0.717, and 0.608, respectively (*Figure 3D*).

Independent prognostic role of the gene signature

Next, we conducted Cox univariate and multivariate regression analyses in the TCGA cohort to assess risk group (classified by the 7-gene signature) and clinicopathological variables [including age, sex, AFP, BMI, inflammation, vascular tumor invasion, residual tumor, tumor grade, tumor stage, and tumor mutational burden (TMB)] as independent prognostic predictors. Risk group and tumor stage were independent prognostic factors in the TCGA cohort (risk group, HR: 0.32; 95% CI: 0.2–0.52; P<0.001; tumor stage, HR: 2.13; 95% CI: 1.4–3.19; P<0.001). Details are provided in *Table 2*.

Building and validating a predictive nomogram

For the purpose of constructing a clinically applicable survival prediction model for patients with HCC, a nomogram was built based on tumor stage and risk score (*Figure 5A*) (29). Calibration plots showed that the nomogram's predictive ability was better for short-term (1 and 3 years) than long-term (5 years) survival, which was consistent with the predicted and actual survival rates (*Figure 5B*). DCA demonstrated that the combined predictive model showed the best net benefit for 1-year

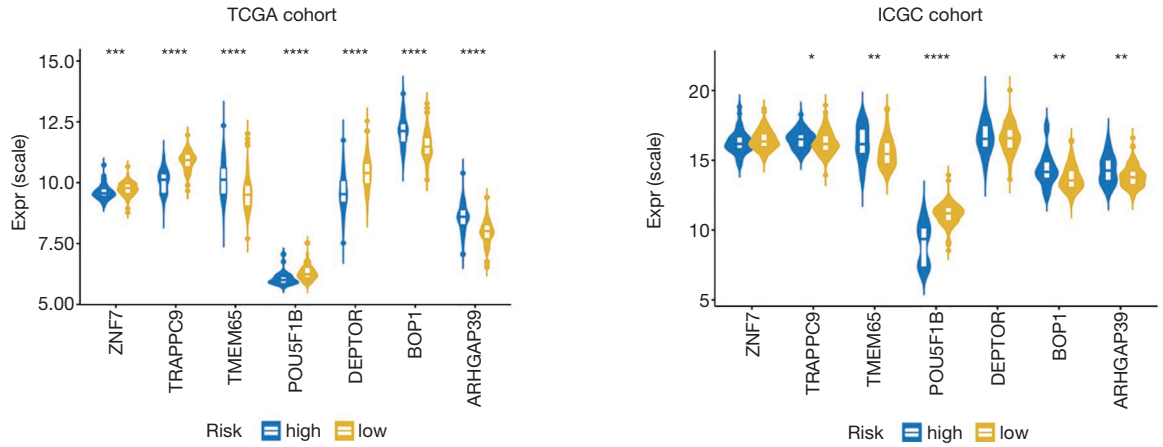

Figure 4 The mRNA expression level of 7 genes between the high-risk and low-risk groups of hepatocellular carcinoma (HCC) patients in The Cancer Genome Atlas (TCGA) cohort and International Cancer Genome Consortium (ICGC) cohort; P<0.05 (*), P<0.01 (**), P<0.001 (***), P<0.0001 (****).

Table 2 Univariate and multivariate Cox regression analyses of gene signature and overall survival of HCC in the TCGA cohort

Characteristics	Hazard ratio	95% CI	P value	Hazard ratio	95% CI	P value
AFP (≥400 *vs.* <400 ng/mL)	1.1062	0.67–1.84	0.697			
Age (≥60 *vs.* <60 years)	1.1606	0.81–1.67	0.419			
BMI (≥25 *vs.* <25 kg/m^2)	0.8099	0.55–1.19	0.285			
Sex (male *vs.* female)	0.7974	0.55–1.15	0.226			
Residual tumor (R1/2 *vs.* R0)	1.8877	0.95–3.74	0.068			
Risk (low *vs.* high)	0.2415	0.16–0.37	1.82E-10	0.3225	0.20–0.52	3.78E-06
TMB, Mut/Mb (≥10 *vs.* <10)	1.5874	0.97–2.59	0.0643			
Tumor grade (G3/4 *vs.* G1/2)	1.0238	0.70–1.49	0.903			
Tumor stage (III/IV *vs.* I/II)	2.6608	1.81–3.91	5.71E-07	2.1295	1.42–3.19	0.000245
Vascular tumor invasion (macro/micro *vs.* none)	1.5169	0.99–2.33	0.057			
Inflammation (mild/severe *vs.* none)	1.2809	0.76–2.12	0.334			

HCC, hepatocellular carcinoma; TGCA, The Cancer Genome Atlas.

and 3 year but not for 5 year as well (*Figure 5C*).

GSEA

With the aim of uncovering the potential molecular-level mechanism of our 7-gene signature, we performed GSEA analysis in the TCGA cohort. As demonstrated in *Table 3* and *Figure 6*, the tumors of patients classified as high risk using the 7-gene signature were significantly enriched in the G2M checkpoint (P<0.001, FDR =0.005), DNA repair (P=0.010, FDR =0.048), MYC targets (P=0.002, FDR =0.006), and mTORC1 signaling (P=0.006, FDR =0.028) pathways or terms. Furthermore, pathways with significant enrichment in patients with low risk were associated with several metabolic processes (including xenobiotic metabolism, bile acid, and fatty acid), adipogenesis, coagulation, and peroxisome function.

Discussion

Because of the complex molecular mechanisms and cellular heterogeneity of HCC, morbidity and death

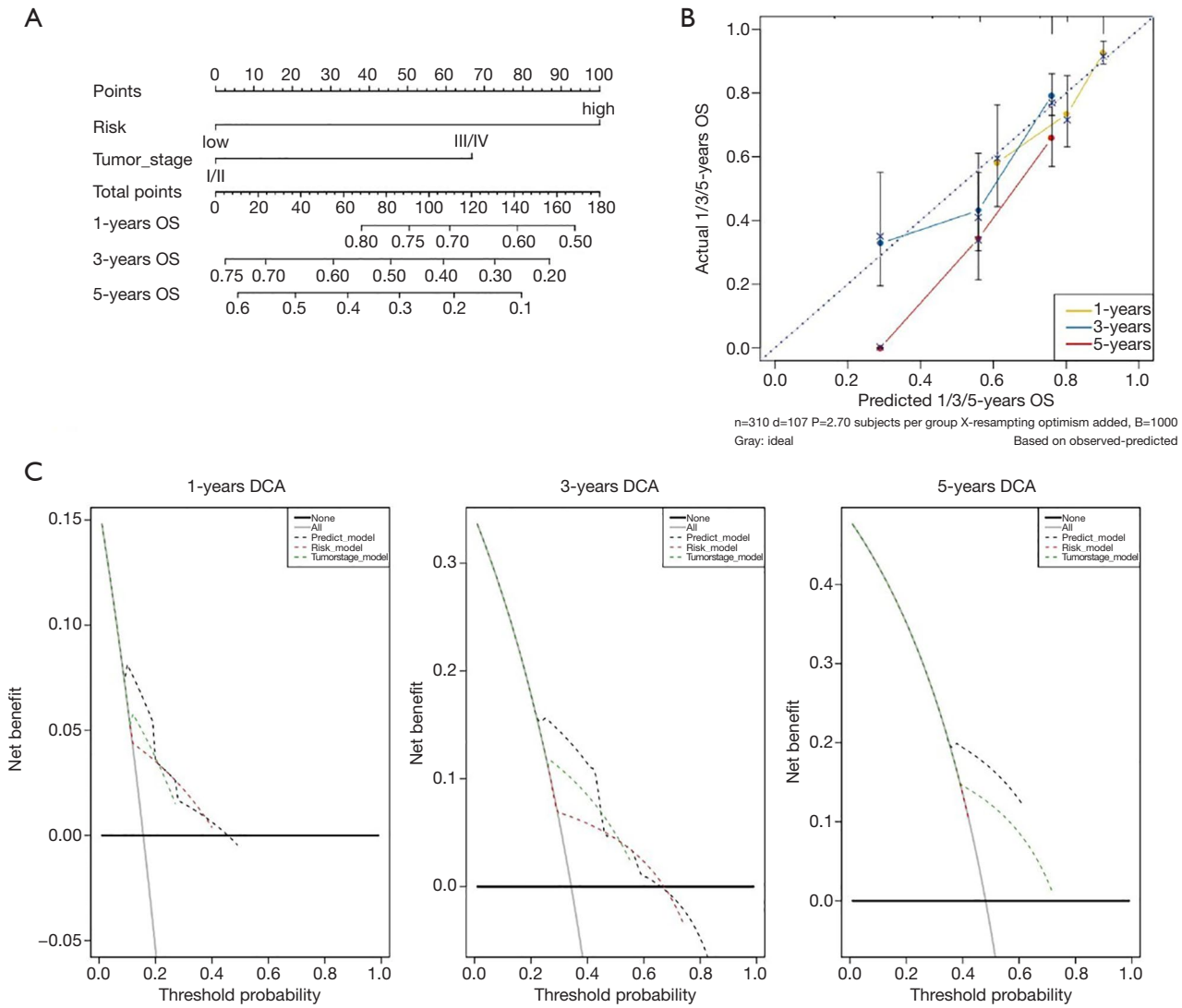

Figure 5 Nomogram to predict the probability of overall survival (OS) of hepatocellular carcinoma (HCC) in The Cancer Genome Atlas (TCGA) cohort. (A) Nomogram predicting the proportion of patients with OS; (B) calibration plot for the nomogram to predict the probability of OS at 1, 3, and 5 years; (C) DCA curves of the nomograms compared for 1-, 3-, and 5-year OS, respectively.

associated with this disease are continuing to rise in multiple countries (11,25). The discovery of new prognostic biomarkers and the development of prognostic models with higher accuracy are urgently required to improve the clinical dilemma posed by HCC. Compared with a single biomarker, a genetic prognostic prediction model constructed with traditional clinical parameters may offer superior prediction efficiency. In this study, we focused on the amplification of genes located in the 8q24 chromosomal region. Using a TCGA dataset, we successfully established a novel 7-gene signature, which was subjected to validation in ICGC dataset. We found that the risk score based on the signature was an independent factor of HCC prognosis, with patients with a high-risk score having a lower survival rate than patients with a low-risk score. ROC curve analysis and DCA results revealed that the nomogram combining the 7-gene signature with tumor stage showed the best predictive efficacy for short-term survival (1 and 3 years) comparing with the nomograms build with a single prognostic factor, but not for long-term survival (5 years). Altogether, the results demonstrated that our 7-gene signature could be

Table 3 Significantly enriched hallmarks in the TCGA cohort by GSEA

Name	Size	NES	NOM p-val	FDR q-val	Signal score
HALLMARK_BILE_ACID_METABOLISM	112	2.35	0.00E+00	7.00E-04	Low
HALLMARK_XENOBIOTIC_METABOLISM	200	2.24	0.00E+00	7.00E-04	Low
HALLMARK_COAGULATION	138	2.05	0.00E+00	7.19E-03	Low
HALLMARK_FATTY_ACID_METABOLISM	158	2.03	1.84E-03	6.52E-03	Low
HALLMARK_PEROXISOME	103	1.94	1.89E-03	1.45E-02	Low
HALLMARK_ADIPOGENESIS	200	1.93	3.74E-03	1.24E-02	Low
HALLMARK_HEME_METABOLISM	197	1.62	8.13E-03	1.06E-01	Low
HALLMARK_G2M_CHECKPOINT	197	–2.27	0.00E+00	4.52E-03	High
HALLMARK_E2F_TARGETS	200	–2.14	0.00E+00	5.00E-03	High
HALLMARK_MYC_TARGETS_V1	199	–2.09	2.04E-03	6.13E-03	High
HALLMARK_MTORC1_SIGNALING	200	–1.94	6.21E-03	2.81E-02	High
HALLMARK_MYC_TARGETS_V2	58	–1.87	2.08E-02	4.07E-02	High
HALLMARK_DNA_REPAIR	149	–1.83	9.94E-03	4.81E-02	High
HALLMARK_MITOTIC_SPINDLE	198	–1.69	3.41E-02	1.09E-01	High
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	112	–1.68	2.29E-02	1.05E-01	High
HALLMARK_UV_RESPONSE_UP	158	–1.62	2.52E-03	1.32E-01	High
HALLMARK_SPERMATOGENESIS	133	–1.54	2.91E-02	1.79E-01	High

TGCA, The Cancer Genome Atlas; GESA, Gene Set Enrichment Analysis; NES, Normalized Enrichment Score; NOM p-val, Nominal p value.

a useful and promising indicator to predict the survival of patients with HCC. Among the 7 genes in our signature, 3 of them have been linked to HCC. Block of proliferation 1 (BOP1) is a WD40 protein, the isolation of which was initially performed through complementary DNA library screening for growth-related sequences in fibroblasts from mouse embryos. BOP1 promotes HCC cell migration and invasion by inducing epithelial–mesenchymal transition (EMT), which is a vital regulatory process, the activation of which occurs during tumor invasion and metastasis (4). BOP1 also stimulates actin stress fiber assembly and RhoA activation, and makes a vital contribution to the enhancement of cell contractility via actin rearrangement and stress fiber formation (30-32). Recently, DEP-domain containing mammalian target of rapamicin (mTOR)-interacting protein (DEPTOR), also known as DEP-domain containing protein 6 (DEPDC6), has emerged as a research hotspot, owing to the discovery that it negatively regulates mTOR. The main components of the protein are mTORC1 and mTORC2 complexes, which inhibit its kinase activity by interacting with mTOR (33,34). The deregulation of mTOR has commonly been linked to tumor pathogenesis. Consistent with inhibitory mTOR activity, the expression of DEPTOR is low in the majority of tumors, although there are some exceptions, including multiple myeloma (MM), thyroid carcinoma, and lung cancer (35,36). DEPTOR promotes EMT and HCC cell metastasis through activation of the TGF-β1-SMAD3/SMAD4-Snail pathway by inhibiting mTOR (37). In embryonic stem cells, OCT4 (also called POU5F1) is an essential participant in the processes of self-renewal, development, and somatic cell reprogramming to pluripotent stem cells. OCT4 has 6 pseudogenes, which are transcribed in many tumors (38,39). Among these OCT4 pseudogenes, POU domain class5 transport factor 1B (POU5F1B) encodes 95% homology to OCT4A. The survival of HCC patients with a high expression POU5F1B is shorter than that of patients with a low expression, with the former group also having a higher level of AKT phosphorylation. POU5F1B was found to

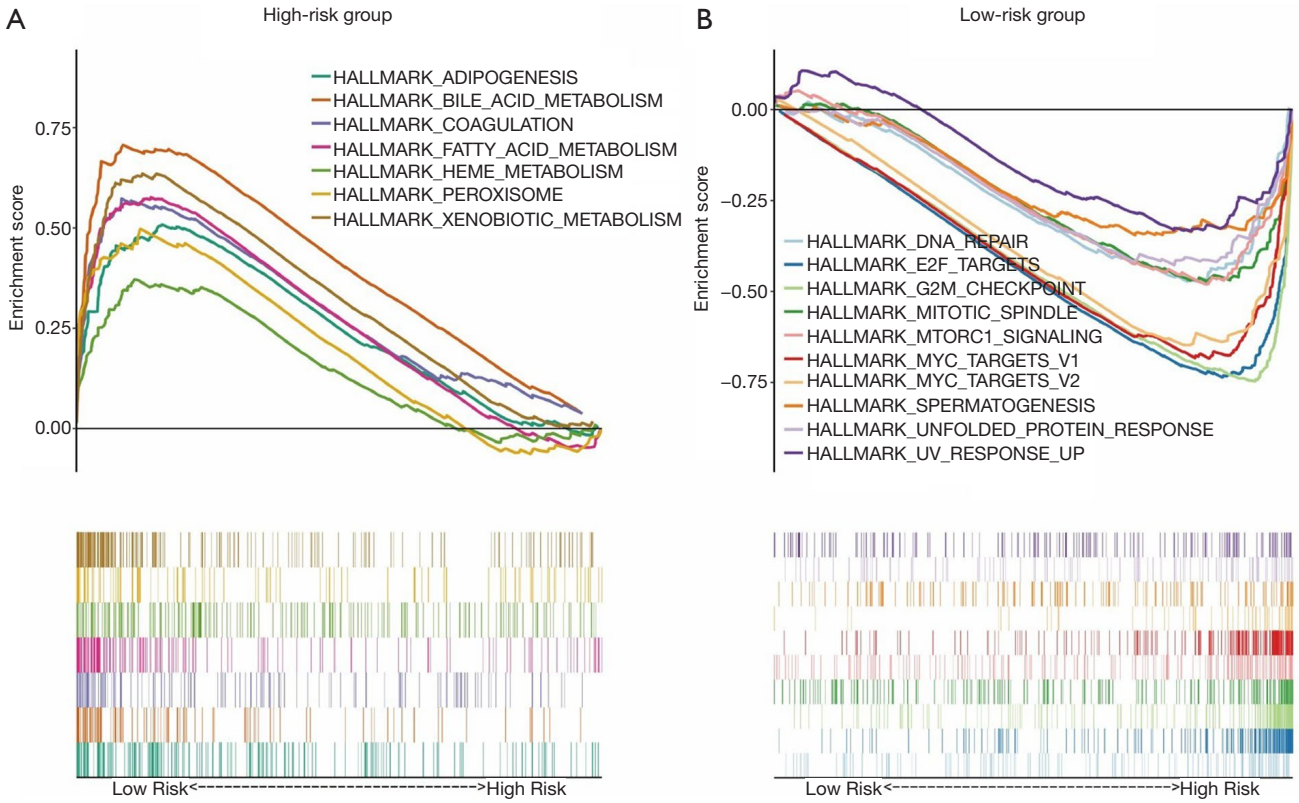

Figure 6 Significantly enriched KEGG pathways in the TCGA cohort by GSEA. (A) Ten representative KEGG pathways in high-risk patients; (B) seven representative KEGG pathways in low-risk patients. TGCA, The Cancer Genome Atlas; GSEA, Gene Set Enrichment Analysis.

promote hepatoma cell proliferation via its activation of AKT; thus, inhibiting AKT in POU5F1B-overexpressing hepatoma cells can in turn inhibit cell proliferation. AKT phosphorylation by TSC2 can also take place at Ser473 sites, which activates Rheb's stimulation of mTORC1 activity, increasing protein synthesis via phosphorylation of eukaryotic translation initiation factor 4 E (eIF4E) and ribosomal proteins S6 (40).

The roles of Arhgap39, TRAPPC9, ZNF7, and TMEM65 in HCC have not been reported. Rho GTPases are widely known to be participants in cell growth, cell dynamics, apoptosis, and intracellular membrane trafficking. Arhgap39 [also referred to as preoptic regulatory factor-2 (Porf-2) or Vilse] is a multidomain protein containing WW, myosin tail homolog 4 (MyTH4), and RhoGAP domains, which can inactivate Rho GTPases (41). Studies have shown that Arhgap39 regulates Rac/cdc42-dependent cytoskeleton and neurogenesis in hippocampal neurons of Drosophila and mice through WW domains (42). Transport protein particle complex 9 (TRAPPC9), a protein subunit of transport protein particle II (TRAPPII), has been illustrated to play a vital role in the transport of cargo from the endoplasmic reticulum to the Golgi body and endosome-to-Golgi trafficking in yeast cells (43). Besides, studies have indicated that TRAPPC9 mutations are associated with an autosomal recessive disorder characterized by mental retardation, autism in mothers, neonatal microcephaly, hearing loss, and stroke (44,45). A20 (Tumor necrosis factor alpha-induced protein 3, TNFAIP3) is a deubiquitinating enzyme with significant anti-inflammatory functions. Genome-wide studies have illustrated that A20 is a susceptibility gene in a variety of human inflammatory and autoimmune diseases, such as systemic lupus erythematosus, psoriasis, rheumatoid arthritis, and Crohn's disease (46,47). The binding ability of A20's ZnF domains 4 and 7 affects its ability to bind to ubiquitin chains. Notably, ZNF7 in A20 is involved in the regulation of the TNF receptor 1 (TNFR1) signaling pathway, and specifically binds to

the linear ubiquitin chain (48,49) and inhibits TNFR1-mediated apoptosis and death (50). ZnF7 can inhibit NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells), the constitutive activation of which is associated with tumorigenesis. Therefore, it can be speculated that the dysregulation of NF-κB activation leads to A20 mutation or deletion, which is the main cause of B-cell lymphoma (51). Transmembrane protein 65 (TMEM65) is an endometrial protein that plays an important role in the function of the mitochondrial respiratory chain (52,53). The biological function of TMEM65 protein has yet to be completely illuminated. One case of a mutation in the TMEM65 gene has been reported in a patient with mitochondrial encephalomyopathy, suggesting that TMEM65 functional changes in mitochondrial function can be caused by TMEM65 gene mutations, resulting in severe cellular and clinical consequences. Besides, the low expression of TMEM65 in skin fibroblasts has been reported to affect the content and respiration rate of mitochondria (54).

In recent years, many studies have constructed effective multi-gene prognostic signature models for HCC. However, a gene signature prognostic model of HCC based on copy number amplification of genes has not been reported previously. Gains of 8q, as the most common changing, have been reported in 27.7% of cancers arising in 27 different tissues and in 31% to 66% of HCCs (23). Amplification of DNA in the 8q24 chromosome region is thought to be associated with progression of multiple cancers including HCC, which indicates that the 8q region harbors 1 or more targeted genes whose amplification promotes their oncogenicity. Our study focused on the type of copy number variant of genes located in the 8q24 chromosome region, and we constructed a predictive signature based on the expression levels of 7 genes. The nomogram combined the 7-gene signature and clinical parameters to determine patient prognosis. The 7 genes may be potential molecular targets to fight HCC and include 4 genes for which a correlation with HCC has not previously been documented, which may be potential research interests or molecular therapeutic targets.

However, some limitations of our study need to be pointed out. Firstly, the mRNA expression data and clinical data were obtained from the TCGA and ICGC databases, in which Europeans and Americans account for the majority of patients. Secondly, our gene signature prognostic model gave a better performance in predicting short-term (1 and 3 years) than long-term (5 years) survival. Thirdly, further investigation and experimentation are needed to explore the underlying functional mechanisms of the 7 predictive signature genes in HCC development and progression.

Conclusions

In summary, our study selected the mRNA expression data of 8q24 genes of HCC patients and focused on the copy number variations of mutation types. A 7-gene signature was successfully constructed to predict the prognosis of HCC. Our prognostic nomogram, which combines the 7-gene signature with clinical factors to predict HCC prognosis, could serve as a reliable tool to identify high-risk HCC patients with poor survival and aid in the development of precise individualized therapy.

Acknowledgments

We thank the patients and investigators who participated in the TCGA, GEO and ICGC for providing data.
Funding: None.

Footnote

Reporting Checklist: The authors have completed the REMARK reporting checklist. Available at http://dx.doi.org/10.21037/jgo-21-205

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jgo-21-205). Author Tingting Chen, Wenzhuan Xie, Mengli Huang, and Yuezong Bai were employed by the company 3D Medicines Inc. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). All data were obtained from publicly available, open-access databases; thus, additional ethical approval was not required.

References

1. Katoh H, Shibata T, Kokubu A, et al. Genetic profile of hepatocellular carcinoma revealed by array-based comparative genomic hybridization: identification of genetic indicators to predict patient outcome. J Hepatol 2005;43:863-74.
2. Llovet JM, Montal R, Sia D, et al. Molecular therapies and precision medicine for hepatocellular carcinoma. Nat Rev Clin Oncol 2018;15:599-616.
3. Villanueva A. Hepatocellular Carcinoma. N Engl J Med 2019;380:1450-62.
4. Chung KY, Cheng IK, Ching AK, et al. Block of proliferation 1 (BOP1) plays an oncogenic role in hepatocellular carcinoma by promoting epithelial-to-mesenchymal transition. Hepatology 2011;54:307-18.
5. Budhu A, Forgues M, Ye QH, et al. Prediction of venous metastases, recurrence, and prognosis in hepatocellular carcinoma based on a unique immune response signature of the liver microenvironment. Cancer Cell 2006;10:99-111.
6. He YZ, He K, Huang RQ, et al. Preoperative evaluation and prediction of clinical scores for hepatocellular carcinoma microvascular invasion: a single-center retrospective analysis. Ann Hepatol 2020;19:654-61.
7. Yoo J, Lee MW, Lee DH, et al. Evaluation of a serum tumour marker-based recurrence prediction model after radiofrequency ablation for hepatocellular carcinoma. Liver Int 2020;40:1189-200.
8. Li B, Feng W, Luo O, et al. Development and Validation of a Three-gene Prognostic Signature for Patients with Hepatocellular Carcinoma. Sci Rep 2017;7:5517.
9. Bai F, Zhou H, Ma M, et al. A novel RNA sequencing-based miRNA signature predicts with recurrence and outcome of hepatocellular carcinoma. Mol Oncol 2018;12:1125-37.
10. Chen PF, Li QH, Zeng LR, et al. A 4-gene prognostic signature predicting survival in hepatocellular carcinoma. J Cell Biochem 2019;120:9117-24.
11. Ke K, Chen G, Cai Z, et al. Evaluation and prediction of hepatocellular carcinoma prognosis based on molecular classification. Cancer Manag Res 2018;10:5291-302.
12. Qiao GJ, Chen L, Wu JC, et al. Identification of an eight-gene signature for survival prediction for patients with hepatocellular carcinoma based on integrated bioinformatics analysis. PeerJ 2019;7:e6548.
13. Wang Z, Teng D, Li Y, et al. A six-gene-based prognostic signature for hepatocellular carcinoma overall survival prediction. Life Sci 2018;203:83-91.
14. Brisbin AG, Asmann YW, Song H, et al. Meta-analysis of 8q24 for seven cancers reveals a locus between NOV and ENPP2 associated with cancer development. BMC Med Genet 2011;12:156.
15. Wang XC, Wu YP, Ye B, et al. Suppression of anoikis by SKP2 amplification and overexpression promotes metastasis of esophageal squamous cell carcinoma. Mol Cancer Res 2009;7:12-22.
16. Fromont G, Godet J, Peyret A, et al. 8q24 amplification is associated with Myc expression and prostate cancer progression and is an independent predictor of recurrence after radical prostatectomy. Hum Pathol 2013;44:1617-23.
17. Ciampa J, Yeager M, Amundadottir L, et al. Large-scale exploration of gene-gene interactions in prostate cancer using a multistage genome-wide association study. Cancer Res 2011;71:3287-95.
18. White NM, Maher CA The potential use of lncRNAs found in the 8q24 region as biomarkers for colon cancer. Ann Oncol 2017;28:1688-9.
19. Wang X, Liu Y, Shao D, et al. Recurrent amplification of MYC and TNFRSF11B in 8q24 is associated with poor survival in patients with gastric cancer. Gastric Cancer 2016;19:116-27.
20. Meng F, Liu B, Xie G, et al. Amplification and overexpression of PSCA at 8q24 in invasive micropapillary carcinoma of breast. Breast Cancer Res Treat 2017;166:383-92.
21. Zhao XM, Xiang ZL, Chen YX, et al. A sequence polymorphism on 8q24 is associated with survival in hepatocellular carcinoma patients who received radiation therapy. Sci Rep 2018;8:2264.
22. Lopez-Coral A, Del Vecchio GJ, Chahine JJ, et al. MAL2-Induced Actin-Based Protrusion Formation is Anti-Oncogenic in Hepatocellular Carcinoma. Cancers (Basel) 2020;12:422.
23. Okamoto H, Yasui K, Zhao C, et al. PTK2 and EIF3S3 genes may be amplification targets at 8q23-q24 and are associated with large hepatocellular carcinomas. Hepatology 2003;38:1242-9.
24. Wu M, Zhou Y, Fei C, et al. ID1 overexpression promotes HCC progression by amplifying the AURKA/Myc signaling pathway. Int J Oncol 2020;57:845-57.

25. Poon TC, Wong N, Lai PB, et al. A tumor progression model for hepatocellular carcinoma: bioinformatic analysis of genomic data. Gastroenterology 2006;131:1262-70.
26. Zhou Q, Zhou Q, Liu Q, et al. PRL-3 facilitates Hepatocellular Carcinoma progression by co-amplifying with and activating FAK. Theranostics 2020;10:10345-59.
27. Liu GM, Xie WX, Zhang CY, et al. Identification of a four-gene metabolic signature predicting overall survival for hepatocellular carcinoma. J Cell Physiol 2020;235:1624-36.
28. Friedman J, Hastie T, Tibshirani R Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33:1-22.
29. Long J, Zhang L, Wan X, et al. A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. J Cell Mol Med 2018;22(12):5928-5938.
30. Narumiya S, Tanji M, Ishizaki T. Rho signaling, ROCK and mDia1, in transformation, metastasis and invasion. Cancer Metastasis Rev 2009;28:65-76.
31. Hall A. The cytoskeleton and cancer. Cancer Metastasis Rev 2009;28:5-14.
32. Etienne-Manneville S, Hall A Rho GTPases in cell biology. Nature 2002;420:629-35.
33. Peterson TR, Laplante M, Thoreen CC, et al. DEPTOR is an mTOR inhibitor frequently overexpressed in multiple myeloma cells and required for their survival. Cell 2009;137:873-86.
34. Caron A, Briscoe DM, Richard D, et al. DEPTOR at the Nexus of Cancer, Metabolism, and Immunity. Physiol Rev 2018;98:1765-803.
35. Catena V, Fanciulli M. Deptor: not only a mTOR inhibitor. J Exp Clin Cancer Res 2017;36:12.
36. Yao H, Tang H, Zhang Y, et al. DEPTOR inhibits cell proliferation and confers sensitivity to dopamine agonist in pituitary adenoma. Cancer Lett 2019;459:135-44.
37. Chen J, Zhu H, Liu Q, et al. DEPTOR induces a partial epithelial-to-mesenchymal transition and metastasis via autocrine TGFbeta1 signaling and is associated with poor prognosis in hepatocellular carcinoma. J Exp Clin Cancer Res 2019;38:273.
38. Hayashi H, Arao T, Togashi Y, et al. The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. Oncogene 2015;34:199-208.
39. Yi J, Zhou LY, Yi YY, et al. Low Expression of Pseudogene POU5F1B Affects Diagnosis and Prognosis in Acute Myeloid Leukemia (AML). Med Sci Monit 2019;25:4952-9.
40. Pan Y, Zhan L, Chen L, et al. POU5F1B promotes hepatocellular carcinoma proliferation by activating AKT. Biomed Pharmacother 2018;100:374-80.
41. Nowak FV. Porf-2 = Arhgap39 = Vilse: A Pivotal Role in Neurodevelopment, Learning and Memory. eNeuro 2018;5:ENEURO.0082-18.2018.
42. Lee JY, Lee LJ, Fan CC, et al. Important roles of Vilse in dendritic architecture and synaptic plasticity. Sci Rep 2017;7:45646.
43. Mbimba T, Hussein NJ, Najeed A, et al. TRAPPC9: Novel insights into its trafficking and signaling pathways in health and disease (Review). Int J Mol Med 2018;42:2991-7.
44. Dookeran KA, Zhang W, Stayner L, et al. Associations of two-pore domain potassium channels and triple negative breast cancer subtype in The Cancer Genome Atlas: systematic evaluation of gene expression and methylation. BMC Res Notes 2017;10:475.
45. Mohamoud HS, Ahmed S, Jelani M, et al. A missense mutation in TRAPPC6A leads to build-up of the protein, in patients with a neurodevelopmental syndrome and dysmorphic features. Sci Rep 2018;8:2053.
46. Skaug B, Chen J, Du F, et al. Direct, noncatalytic mechanism of IKK inhibition by A20. Mol Cell 2011;44:559-71.
47. Martens A, Priem D, Hoste E, et al. Two distinct ubiquitin-binding motifs in A20 mediate its anti-inflammatory and cell-protective activities. Nat Immunol 2020;21:381-7.
48. Wertz IE, Newton K, Seshasayee D, et al. Phosphorylation and linear ubiquitin direct A20 inhibition of inflammation. Nature 2015;528:370-5.
49. Draber P, Kupka S, Reichert M, et al. LUBAC-Recruited CYLD and A20 Regulate Gene Activation and Cell Death by Exerting Opposing Effects on Linear Ubiquitin in Signaling Complexes. Cell Rep 2015;13:2258-72.
50. Yamaguchi N, Yamaguchi N. The seventh zinc finger motif of A20 is required for the suppression of TNF-alpha-induced apoptosis. FEBS Lett 2015;589:1369-75.
51. Karin M. Nuclear factor-kappaB in cancer development and progression. Nature 2006;441:431-6.
52. Nishimura N, Gotoh T, Oike Y, et al. TMEM65 is a mitochondrial inner-membrane protein. PeerJ 2014;2:e349.
53. Sharma P, Abbasi C, Lazic S, et al. Evolutionarily

conserved intercalated disc protein Tmem65 regulates cardiac conduction and connexin 43 function. Nat Commun 2015;6:8391.
54. Nazli A, Safdar A, Saleem A, et al. A mutation in the TMEM65 gene results in mitochondrial myopathy with severe neurological manifestations. Eur J Hum Genet 2017;25:744-51.

(English Language Editor: J. Reynolds)

Cite this article as: Zheng Y, Cheng Y, Zhang C, Fu S, He G, Cai L, Qiu L, Huang K, Chen Q, Xie W, Chen T, Huang M, Bai Y, Pan M. Co-amplification of genes in chromosome 8q24: a robust prognostic marker in hepatocellular carcinoma. J Gastrointest Oncol 2021;12(3):1086-1100. doi: 10.21037/jgo-21-205

Supplementary

Table S1 The relationship between clinical characteristic and overall survival of TCGA cohort

	Amplification	Non-amplification	P value
N	199	109	
Age, years, mean (SD)	59.74 (11.88)	60.78 (14.22)	0.496
Gender (male), n (%)	150 (75.4)	63 (57.8)	0.002
Race, n (%)			<0.001
Asian	109 (54.8)	34 (31.2)	
Black or African American	9 (4.5)	5 (4.6)	
Not reported	6 (3.0)	2 (1.8)	
White	75 (37.7)	68 (62.4)	
AFP, ng/ml, mean (SD)	20,600.93 (166,204.61)	3,702.72 (26,575.26)	0.374
BMI, kg/m^2, mean (SD)	26.03 (9.95)	26.84 (6.43)	0.465
Inflammation, n (%)			0.005
Mild	58 (29.1)	24 (22.0)	
None	51 (25.6)	47 (43.1)	
Severe	15 (7.5)	2 (1.8)	
N/A	75 (37.7)	36 (33.0)	
Tumor grade, n (%)			<0.001
G1	20 (10.1)	30 (27.5)	
G2	91 (45.7)	52 (47.7)	
G3	76 (38.2)	25 (22.9)	
G4	11 (5.5)	0 (0.0)	
N/A	1 (0.5)	2 (1.8)	
Tumor stage, n (%)			0.122
Not reported	9 (4.5)	10 (9.2)	
Stage I	98 (49.2)	51 (46.8)	
Stage II	51 (25.6)	20 (18.3)	
Stage III	0 (0.0)	1 (0.9)	
Stage IIIA	35 (17.6)	18 (16.5)	
Stage IIIB	3 (1.5)	4 (3.7)	
Stage IIIC	3 (1.5)	5 (4.6)	
Residual tumor, n (%)			0.164
R0	180 (90.5)	95 (87.2)	
R1	9 (4.5)	2 (1.8)	
RX	7 (3.5)	9 (8.3)	
N/A	3 (1.5)	3 (2.8)	
Vascular tumor invasion, n (%)			0.854
Macro	7 (3.5)	5 (4.6)	
Micro	53 (26.6)	25 (22.9)	
None	108 (54.3)	63 (57.8)	
N/A	31 (15.6)	16 (14.7)	
TMB, Mut/Mb, mean (SD)	6.72 (6.31)	6.04 (5.82)	0.38

TGCA, The Cancer Genome Atlas; AFP, alpha-fetoprotein; BMI, body mass index; TMB, tumor mutation burden.

http://dx.doi.org/10.21037/jgo-21-205