Prognostic Value of a Novel Signature With Nine Hepatitis C Virus-Induced Genes in Hepatic Cancer by Mining GEO and TCGA Databases

Jianming Wei[1†], Bo Wang[2†], Xibo Gao[3†] and Daqing Sun[2*]

[1] Department of General Surgery, Tianjin Medical University General Hospital, Tianjin, China, [2] Department of Paediatric Surgery, Tianjin Medical University General Hospital, Tianjin, China, [3] Department of Dermatology, Tianjin Children's Hospital, Tianjin, China

Background: Hepatitis C virus-induced genes (HCVIGs) play a critical role in regulating tumor development in hepatic cancer. The role of HCVIGs in hepatic cancer remains unknown. This study aimed to construct a prognostic signature and assess the value of the risk model for predicting the prognosis of hepatic cancer.

Methods: Differentially expressed HCVIGs were identified in hepatic cancer data from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases using the library ("limma") package of R software. The protein–protein interaction (PPI) network was constructed using the Cytoscape software. Functional enrichment analysis was performed using the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Univariate and multivariate Cox proportional hazard regression analyses were applied to screen for prognostic HCVIGs. The signature of HCVIGs was constructed. Gene Set Enrichment Analysis (GSEA) compared the low-risk and high-risk groups. Finally, the International Cancer Genome Consortium (ICGC) database was used to validate this prognostic signature. Polymerase chain reaction (PCR) was performed to validate the expression of nine HCVIGs in the hepatic cancer cell lines.

Results: A total of 143 differentially expressed HCVIGs were identified in TCGA hepatic cancer dataset. Functional enrichment analysis showed that DNA replication was associated with the development of hepatic cancer. The risk score signature was constructed based on the expression of *ZIC2*, *SLC7A11*, *PSRC1*, *TMEM106C*, *TRAIP*, *DTYMK*, *FAM72D*, *TRIP13*, and *CENPM*. In this study, the risk score was an independent prognostic factor in the multivariate Cox regression analysis [hazard ratio (HR) = 1.433, 95% CI = 1.280–1.605, $P < 0.001$]. The overall survival curve revealed that the high-risk group had a poor prognosis. The Kaplan–Meier Plotter online database showed that the survival time of hepatic cancer patients with overexpression of HCVIGs in this signature was significantly shorter. The prognostic signature-associated GO and

KEGG pathways were significantly enriched in the risk group. This prognostic signature was validated using external data from the ICGC databases. The expression of nine prognostic genes was validated in HepG2 and LO-2.

Conclusion: This study evaluates a potential prognostic signature and provides a way to explore the mechanism of HCVIGs in hepatic cancer.

INTRODUCTION

Hepatocellular carcinoma (HCC) is the sixth most common cancer and was the third leading cause of cancer-related deaths worldwide in 2020 (Sung et al., 2021). The incidence of HCC has increased sharply over the recent decades (Zhang et al., 2012; McGlynn et al., 2021). The 5—year overall survival (OS) rate of HCC is still unsatisfactory, and its pathogenesis remains controversial.

The development of liver cancer is a complex process, including chronic inflammation and epigenetic modifications (Plissonnier et al., 2018). Hepatitis C virus (HCV) infection is a prominent risk factor for the development of HCC (Ali et al., 2011). Each year, 4–5% of patients with chronic hepatitis C develop HCC (Bartosch et al., 2009). The combination of host, environmental, and viral factors results in HCV-related carcinogenesis (Vescovo et al., 2016). HCV enhances the invasiveness of HCC *via* epidermal growth factor receptor (EGFR)-mediated invadopodia formation and activation (Ninio et al., 2019). Many non-coding RNAs (ncRNAs) play important roles in biological processes, such as differentiation, proliferation, and cell death, in HCC (Yang et al., 2014; Refai et al., 2019). However, the molecular mechanism of HCV-induced genes (HCVIGs) in hepatocarcinogenesis is not well-understood.

To explore HCVIGs in hepatic cancer, we analyzed HCV-induced expression profiling of hepatic and non-hepatic cancer cell lines from GSE70781 and identified a total of 1,582 differentially expressed genes (DEGs). Then, a total of 1,286 DEGs were common genes in the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. We extracted 1,284 DEGs from TCGA hepatic cancer gene expression, including 407 hepatic cancer tissues and 58 adjacent normal tissues. Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed to analyze the molecular processes. We used univariate Cox proportional risk regression analysis to identify 21 prognostic HCVIGs from TCGA hepatic cancer dataset. To predict the prognosis of hepatic cancer, we constructed nine prognostic HCVIGs signatures using *ZIC2*, *SLC7A11*, *PSRC1*, *TMEM106C*, *TRAIP*, *DTYMK*, *FAM72D*, *TRIP13*, and *CENPM*. The risk score was identified as an independent factor. OS analysis showed that patients in the low-risk group had a better prognosis than those in the high-risk group according to the median risk score. Finally, the International Cancer Genome Consortium (ICGC) databases were analyzed to validate the prognostic signature. Polymerase chain reaction (PCR) revealed that *ZIC2*, *SLC7A11*, *PSRC1*,

TMEM106C, *TRAIP*, *DTYMK*, *FAM72D*, *TRIP13*, and *CENPM* were upregulated in the hepatic cancer cell line HepG2.

MATERIALS AND METHODS

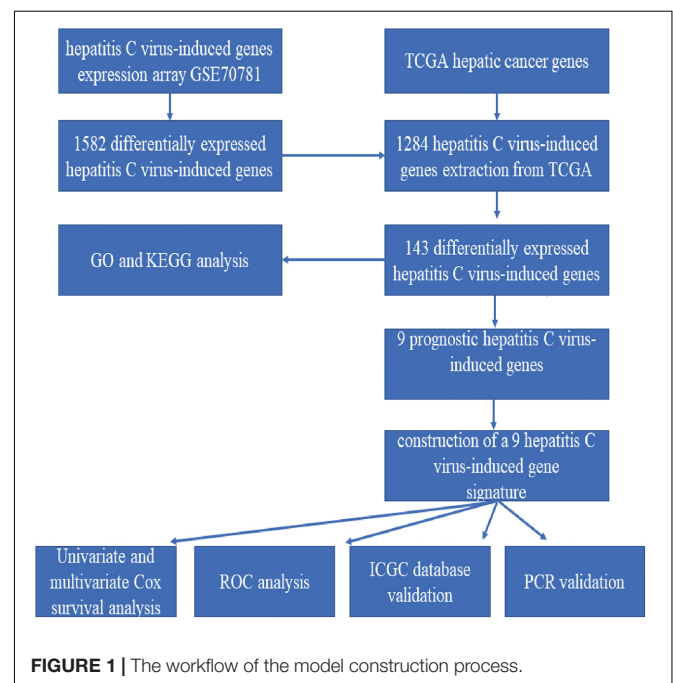### The Workflow of the Model Construction Process

The detailed workflow is shown in **Figure 1**.

Data Acquisition and Processing

Hepatitis C virus-induced genes (HCVIGs) were downloaded from the GEO[1] including hepatic and non-hepatic cancer cell lines in GSE70781. The original gene expression profiles and clinical data of hepatic cancer were obtained from TCGA data portal[2]. Differentially expressed HCVIGs from the GEO database were extracted using TCGA data. The JP Project from the International Cancer Genome Consortium (ICGC-LIRI-JP) was downloaded from the Database of Hepatocellular Carcinoma Expression Atlas (HCCDB) for external validation.

---

[1]https://www.ncbi.nlm.nih.gov/geo/
[2]https://tcga-data.nci.nih.gov/tcga/

FIGURE 1 | The workflow of the model construction process.

FIGURE 2 | Identification of differentially expressed hepatitis C virus-induced genes. **(A)** Heat map of differentially expressed HCVIGs expression levels in GSE70781. **(B)** Volcano plot for differentially expressed HCVIGs expression levels in GSE70781. **(C)** Heat map of differentially expressed HCVIGs expression levels in TCGA dataset. **(D)** Volcano plot for differentially expressed HCVIGs expression levels in TCGA dataset.

FIGURE 3 | Gene Ontology and KEGG pathway enrichment analyses. **(A)** GO analysis of 143 differentially expressed hepatitis C virus-induced genes. "BP" represents "biological process," "CC" represents "cellular component," and "MF" represents "molecular function." **(B)** Kyoto Encyclopedia of Genes and Genomes analysis of differentially expressed hepatitis C virus-induced genes.

Function Annotation and Gene Set Enrichment Analysis

Differentially expressed HCVIGs were identified using the limma package. Fold change > 2.0 and adjusted <0.05 were selected as the significantly differentially expressed HCVIGs. GO is a community-based bioinformatics resource that includes biological processes (BP), cell components (CC), and molecular

functions (MF) (Peng et al., 2017). The KEGG is a knowledge base for systematic analysis of gene functions, linking genomic information with higher-order functional information (Kanehisa and Goto, 2000). GO and KEGG enrichment results were generated by R package "ggplot2," "enrichplot," "clusterProfiler," and "GOplot" for the purpose of analysis. Gene Set Enrichment Analysis (GSEA) comparing the low-risk and high-risk
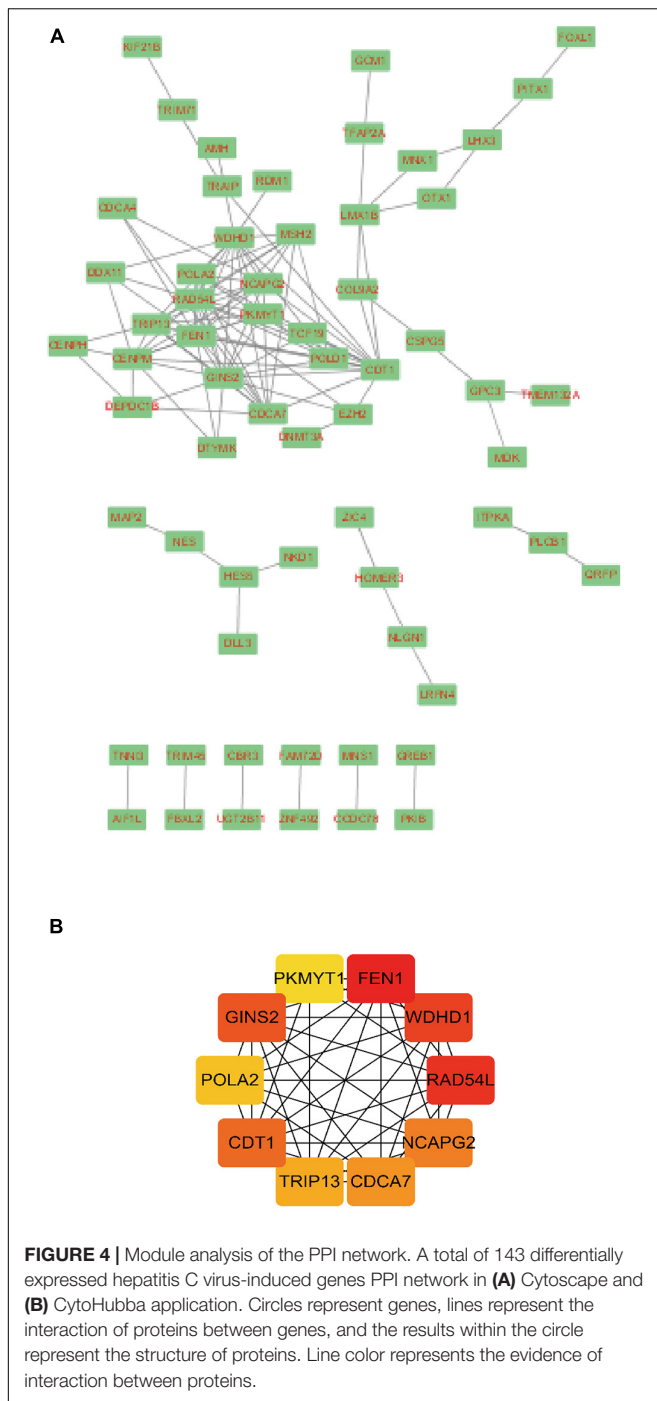
FIGURE 4 | Module analysis of the PPI network. A total of 143 differentially expressed hepatitis C virus-induced genes PPI network in **(A)** Cytoscape and **(B)** CytoHubba application. Circles represent genes, lines represent the interaction of proteins between genes, and the results within the circle represent the structure of proteins. Line color represents the evidence of interaction between proteins.

groups was performed by GSEA software (version 4.0.3). Functional annotation with a *P*-value < 0.05 was considered statistically significant.

PPI Network and Hub Genes Analysis

STRING database (Szklarczyk et al., 2017)[3] was used to analyze the functional interactions between expressed proteins.

[3]https://string-db.org/

The STRING network was then input to Cytoscape software (version 3.6.0) (Yang et al., 2019) to construct a protein–protein interaction (PPI) network of 143 differentially expressed HCVIGs. The top 10 genes were selected using the CytoHubba application (Zhou et al., 2019).

Construction of the Nine Prognostic HCVIGs Signature

Differentially expressed HCVIGs were selected to construct a multivariate Cox regression model. OS curves were plotted using the Kaplan–Meier (K–M) method. The relationships were tested using the log-rank test. The nine prognostic HCVIGs-associated signature was constructed using coefficients in multivariate Cox regression analysis (Qiu et al., 2020). Nine prognostic values of HCVIGs were validated using the Kaplan–Meier Plotter (Menyhárt et al., 2018).

Real-Time RT-qPCR Assay

Hepatic cancer cell line HepG2 and normal hepatic cell line LO-2 were used to test the expression of nine prognostic values of HCVIGs. Real-time RT-qPCR experiments were performed as previously described. The primer sequences used are shown in **Supplementary Table 1**. Relative mRNA levels after correction for GAPDH control mRNA were expressed.
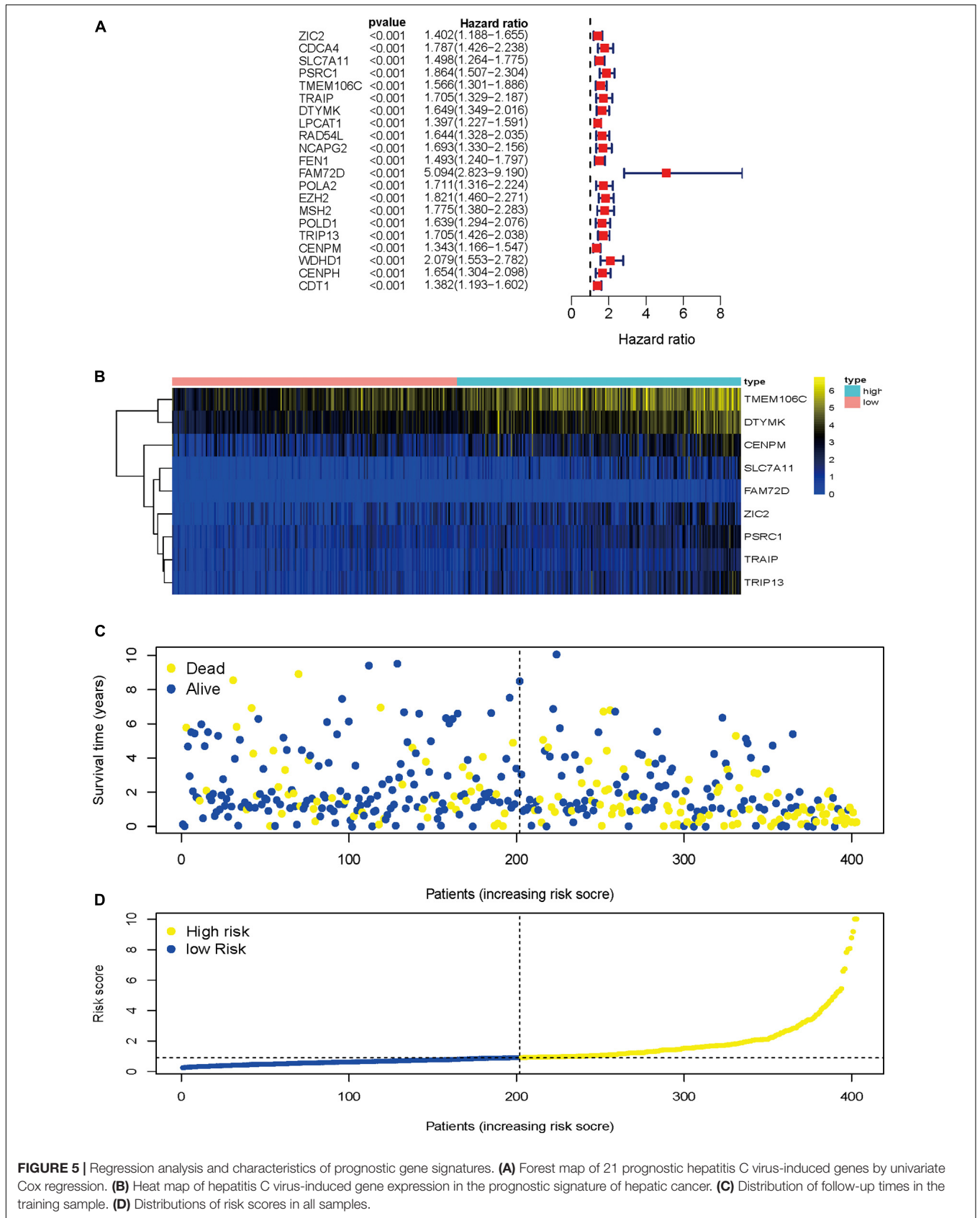
Statistical Analysis

R software (version 3.5.2) was used to perform all statistical analyses in this study. Statistical significance was set at $P < 0.05$. Univariate Cox regression analyses were performed to assess the relationship between expression profiles and prognosis. Multivariate Cox regression was used to construct a risk model according to the clinical factors correlated with survival. Receiver operating characteristic (ROC) curve and the corresponding area under the ROC curve (AUC) were analyzed by the package of "survivalROC" in R.

RESULTS

Identification of Differentially Expressed HCVIGs

A total of 1,582 differentially expressed HCVIGs were identified using the R language software in GSE70781. Among the differentially expressed HCVIGs, 744 were downregulated, whereas a total of 838 were upregulated. These results are presented as a heat map (**Figure 2A**) and a volcano plot (**Figure 2C**). We obtained the gene expression and clinical data for all 465 patients, including 407 hepatic cancer tissues and 58 adjacent normal tissues from TCGA database. A total of 1,286 differentially expressed HCVIGs were common genes between GSE70781 and TCGA. A total of 1,284 differentially expressed HCVIGs were extracted from TCGA database. Finally, a total of 143 differentially expressed HCVIGs were selected from TCGA, including 2 low-expression genes and 141 over-expression genes, as shown in the heat map (**Figure 2B**) and the volcano plot (**Figure 2D**).

FIGURE 5 | Regression analysis and characteristics of prognostic gene signatures. (A) Forest map of 21 prognostic hepatitis C virus-induced genes by univariate Cox regression. (B) Heat map of hepatitis C virus-induced gene expression in the prognostic signature of hepatic cancer. (C) Distribution of follow-up times in the training sample. (D) Distributions of risk scores in all samples.

GO Enrichment and KEGG Pathway Analysis

A total of 143 differentially expressed HCVIGs were subjected to GO and KEGG functional annotation. A total of three GO terms of BP and one GO term of CC were statistically significant ($P < 0.05$). The four GO terms included "regionalization," "negative regulation of cell morphogenesis involved in differentiation," "DNA-dependent DNA replication," and "nuclear replication fork." The results are shown in **Figure 3A**. In addition, in **Figure 3B**, a total of 148 KEGG pathways were analyzed; the top three KEGG pathways were DNA replication, endocrine and other factor-regulated calcium reabsorption, and mismatch repair. These results are closely associated with HCV infection.

PPI and Hub Genes Analysis

A total of 141 nodes and 120 edges were downloaded from the STRING database, and then Cytoscape software was used to visualize the regulatory network of the genes (**Figure 4A**). The CytoHubba application identified the top 10 hub genes (**Figure 4B**). These genes included *FEN1*, *WDHD1*, *RAD54L*, *GINS2*, *CDT1*, *NCAPG2*, *CDCA7*, *TRIP13*, *POLA2*, and *PKMYT1*.

Nine Prognostic HCVIGs Risk Score Model Construction

To identify HCVIGs associated with hepatic cancer, we screened out 21 prognostic HCVIGs using univariate analysis (**Figure 5A**). Finally, nine prognostic genes (*ZIC2*, *SLC7A11*, *PSRC1*, *TMEM106C*, *TRAIP*, *DTYMK*, *FAM72D*, *TRIP13*, and *CENPM*) of hepatic cancer were identified. The coefficients of the nine prognostic genes are shown in **Table 1**. Nine prognostic genes were determined to construct the risk model. The study formula for the risk score was as follows (Wang et al., 2019): Risk score = $(0.169963967 \times$ expression value of *ZIC2*) + $(0.335235013 \times$ expression value of *SLC7A11*) + $(0.35212068 \times$ expression value of *PSRC1*) + $(0.3061227 \times$ expression value of *TMEM106C*) + $(-0.537475693 \times$ expression value of *TRAIP*) + $(0.461239921 \times$ expression value of *DTYMK*) + $(1.079744869 \times$ expression value of *FAM72D*) +

TABLE 1 | Genes included in prognostic hepatitis C virus-induced genes signature.

Id	Coef	HR	HR.95L	HR.95H	p-value
ZIC2	0.169964	1.185262	0.982667	1.429627	0.075546
SLC7A11	0.335235	1.398269	1.148343	1.702589	0.000848
PSRC1	0.352121	1.42208	0.986723	2.049524	0.058988
TMEM106C	0.306123	1.358149	1.064037	1.733556	0.013954
TRAIP	−0.53748	0.584221	0.335822	1.016356	0.057101
DTYMK	0.46124	1.586039	1.053333	2.388154	0.02719
FAM72D	1.079745	2.943928	1.194113	7.257866	0.019012
TRIP13	0.343541	1.409931	1.038792	1.913671	0.027514
CENPM	−0.48932	0.613044	0.429589	0.874844	0.006998

$(0.3435407 \times$ expression value of *TRIP13*) + $(-0.489318594 \times$ expression value of *CENPM*).

Heat map was drawn to investigate genes expression profiles in the high-risk and low-risk HCC groups. Patients were divided into the high-risk and low-risk groups based on the median risk score. The risk score distribution, follow-up time, and gene heat map are shown in **Figures 5B–D**.

Nine Prognostic HCVIGs Predict a Prognosis in Hepatic Cancer

To validate the potential values of *ZIC2*, *SLC7A11*, *PSRC1*, *TMEM106C*, *TRAIP*, *DTYMK*, *FAM72D*, *TRIP13*, and *CENPM* predicting prognosis, we generated Kaplan–Meier survival curves from the Kaplan–Meier Plotter online database (Nagy et al., 2018) using data for 364 liver cancer patients. Among the nine HCVIGs, high expression of *ZIC2*, *SLC7A11*, *PSRC1*, *TMEM106C*, *TRAIP*, *DTYMK*, *FAM72D*, *TRIP13*, and *CENPM* correlated with a poor prognosis. This result is shown in **Figures 6A–I**.

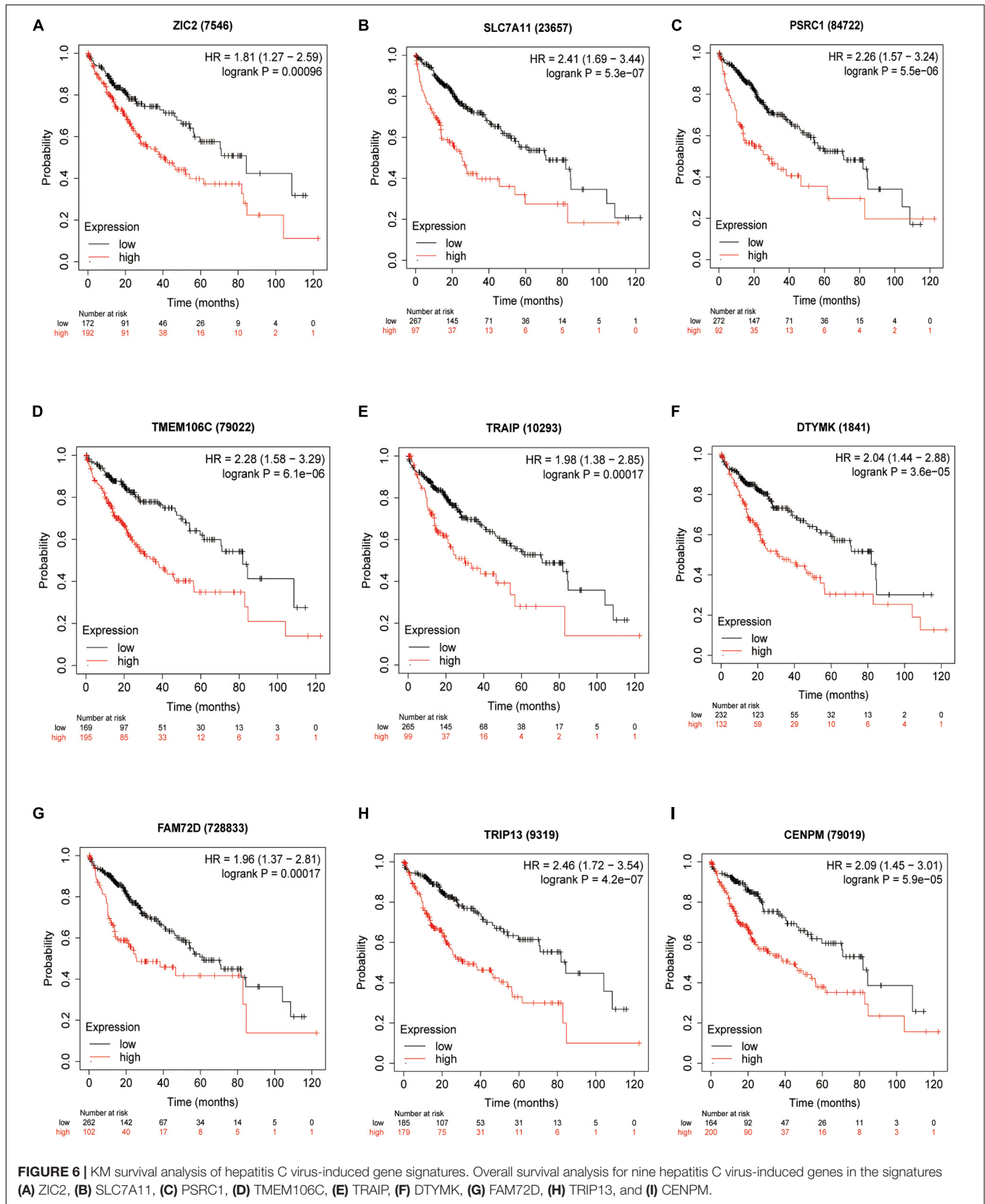## Cox Proportional Hazard Regression Analysis

In the univariate Cox regression analysis, we found that stage [hazard ratio (HR) = 1.865, 95% CI = 1.456–2.388, $P < 0.001$], tumor (T; HR = 1.804, 95% CI = 1.434–2.270, $P < 0.001$), metastasis (M; HR = 3.850, 95% CI = 1.207–12.281, $P = 0.023$), and risk score (HR = 1.482, 95% CI = 1.334–1.645, $P < 0.001$) were significantly correlated with OS. Furthermore, in the multivariate Cox regression analysis, the risk score (HR = 1.433, 95% CI = 1.280–1.605, $P < 0.001$) was found to be an independent risk factor for prognosis in patients with hepatic cancer (**Figures 7A,B**). Using packages "survival" and "survminer" of the R language, the survival curves revealed that the low-risk group had a better survival outcome (**Figure 7C**). In this study, the AUC of the ROC curve included the following: risk score (AUC = 0.806), age (AUC = 0.546), sex (AUC = 0.504), grade (AUC = 0.478), stage (AUC = 0.703), T (AUC = 0.708), M (AUC = 0.508), and node (N) (AUC = 0.508). The AUC value of the risk score was higher than that of other clinical parameters (**Figure 7D**), revealing that this Cox regression model predicted prognosis well in hepatic cancer patients.
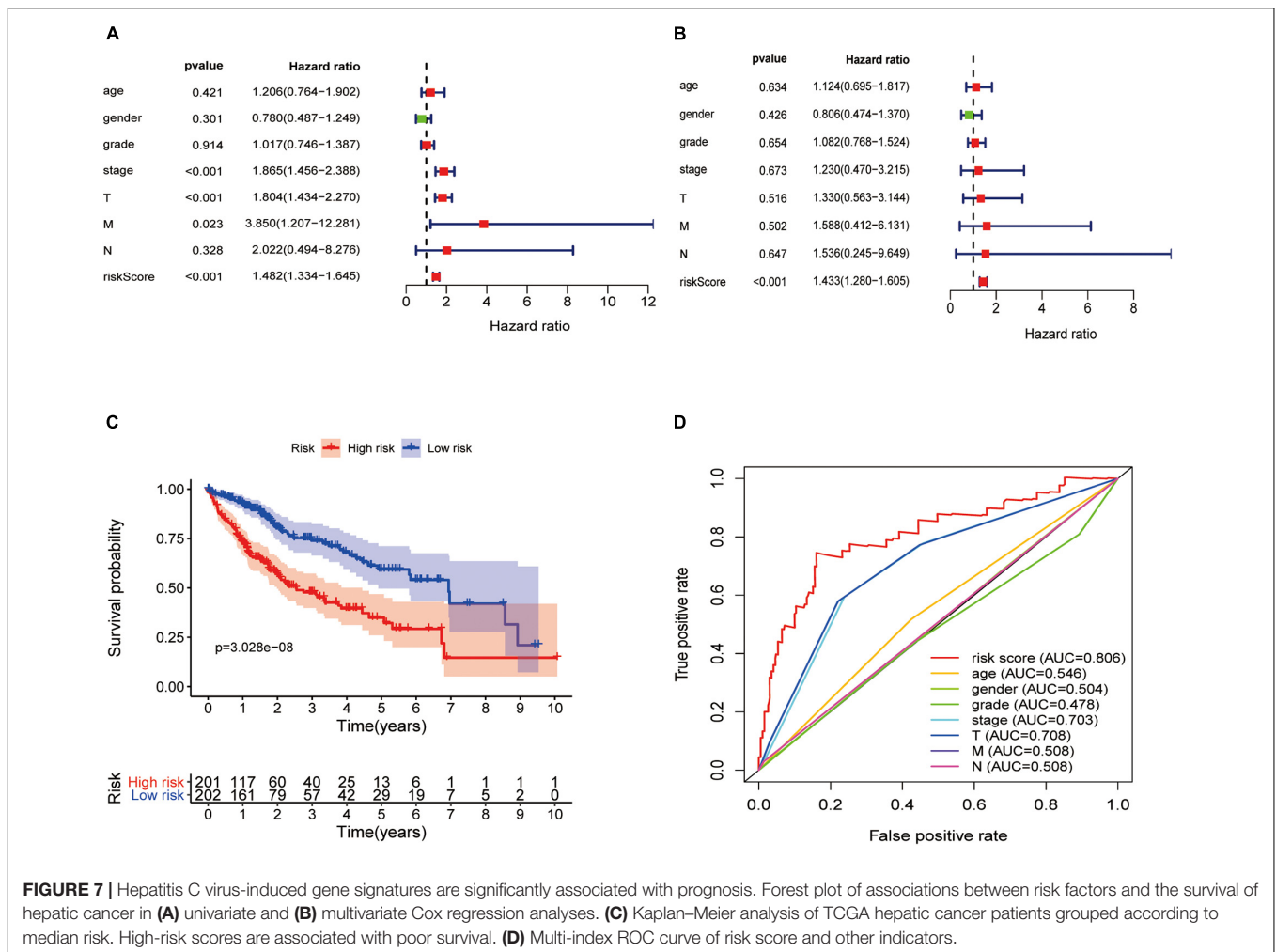
GSEA of the Prognostic Signature in the Low-Risk and High-Risk Groups

GSEAs were performed to compare the low-risk group with the high-risk group of the HCVIG signature. The results showed that these prognostic signature-associated GO and KEGG pathways are significantly enriched in the high-risk (**Figures 8A,B**) and low-risk groups (**Figures 8C,D**). The top five KEGG pathways in the high-risk group were RNA degradation, pyrimidine metabolism, ubiquitin mediated proteolysis, cell cycle, and oocyte meiosis. However, the top five KEGG pathways in the low-risk group were complement and coagulation cascades, drug metabolism cytochrome P450, tryptophan metabolism, linoleic acid metabolism, and retinol metabolism.

FIGURE 6 | KM survival analysis of hepatitis C virus-induced gene signatures. Overall survival analysis for nine hepatitis C virus-induced genes in the signatures **(A)** ZIC2, **(B)** SLC7A11, **(C)** PSRC1, **(D)** TMEM106C, **(E)** TRAIP, **(F)** DTYMK, **(G)** FAM72D, **(H)** TRIP13, and **(I)** CENPM.

FIGURE 7 | Hepatitis C virus-induced gene signatures are significantly associated with prognosis. Forest plot of associations between risk factors and the survival of hepatic cancer in **(A)** univariate and **(B)** multivariate Cox regression analyses. **(C)** Kaplan–Meier analysis of TCGA hepatic cancer patients grouped according to median risk. High-risk scores are associated with poor survival. **(D)** Multi-index ROC curve of risk score and other indicators.

External Verification of the Prognostic HCVIG Signature

The HCVIG signature was tested in a hepatic cancer cohort from the ICGC database for validation using the same risk score formula and the same cut-off value.

The gene expression pattern, risk distribution, and survival status are shown in **Figures 9A–C**. The Kaplan–Meier plot showed that patients in the high-risk group had a significantly poorer OS than those in the low-risk group in the validation cohort (**Figure 9D**). The HCVIG signature had a good accuracy to predict OS in the validation cohort, with an AUC of 0.763 (**Figure 9E**). Multivariate Cox regression analyses showed that gender, stage, prior malignancy, and risk score were found to be independent risk factors for prognosis in patients with hepatic cancer (**Table 1**). The relations between risk score and clinicopathological factors are shown in **Figure 10**.

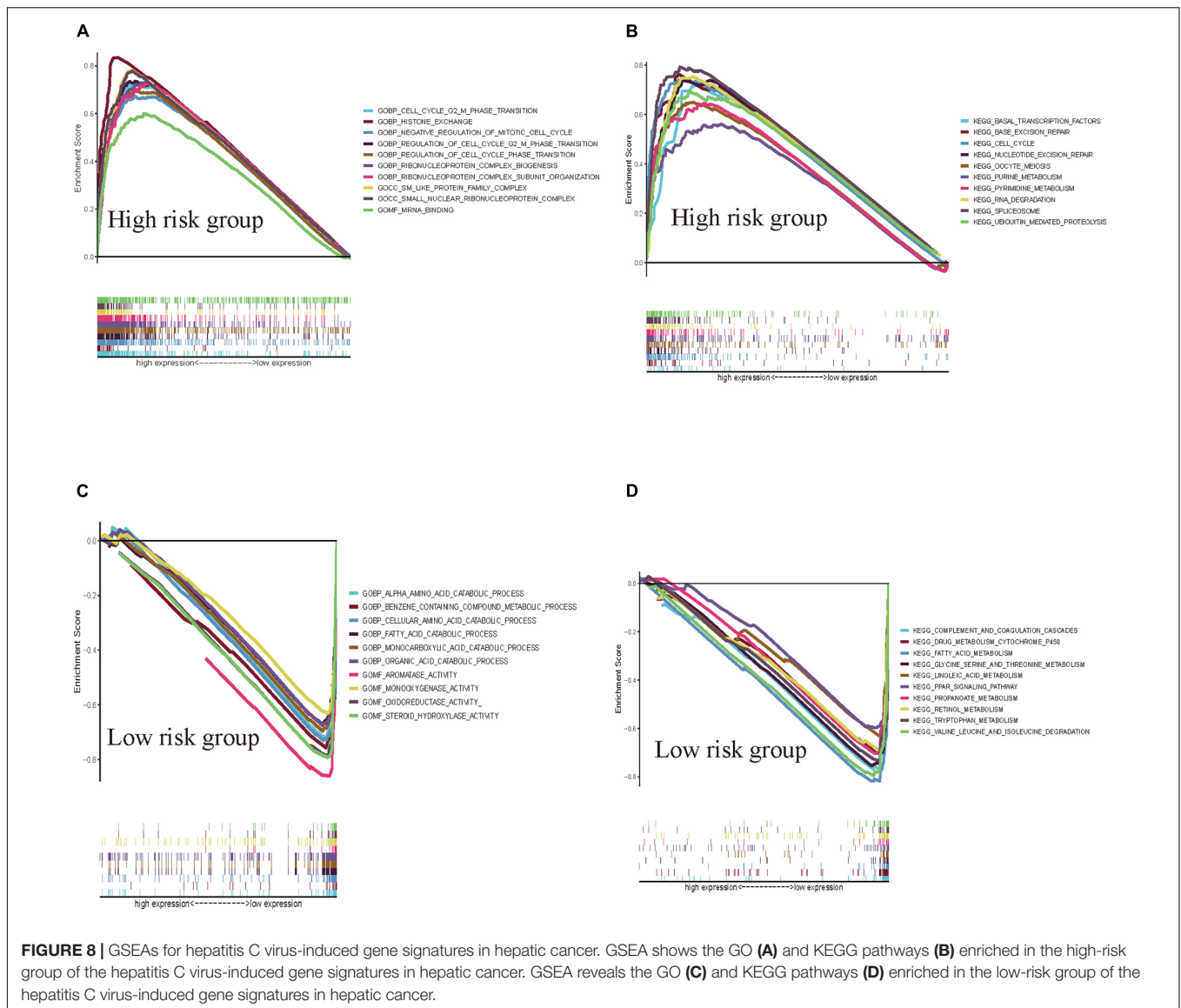## Nine Prognostic HCVIGs Are Overexpressed in Hepatic Cancer

To explore the expressions of *ZIC2*, *SLC7A11*, *PSRC1*, *TMEM106C*, *TRAIP*, *DTYMK*, *FAM72D*, *TRIP13*, and *CENPM*,

these genes were tested in hepatic cancer cell line (HepG2) and normal hepatic cell line (LO-2). The expressions of *ZIC2*, *SLC7A11*, *PSRC1*, *TMEM106C*, *TRAIP*, *DTYMK*, *FAM72D*, *TRIP13*, and *CENPM* are shown in **Figure 11**.

DISCUSSION

HCV is an RNA virus that integrates its genetic material into the host genome (Hoshida et al., 2014). Persistent HCV infection is a critical factor in fibrosis/cirrhosis and HCC. HCV can infect and replicate in hepatocytes, impair normal functions of other liver cells, and promote fibrosis/cirrhosis (Sasaki et al., 2017). However, the mechanism of HCV infection in the development of HCC remains unclear.
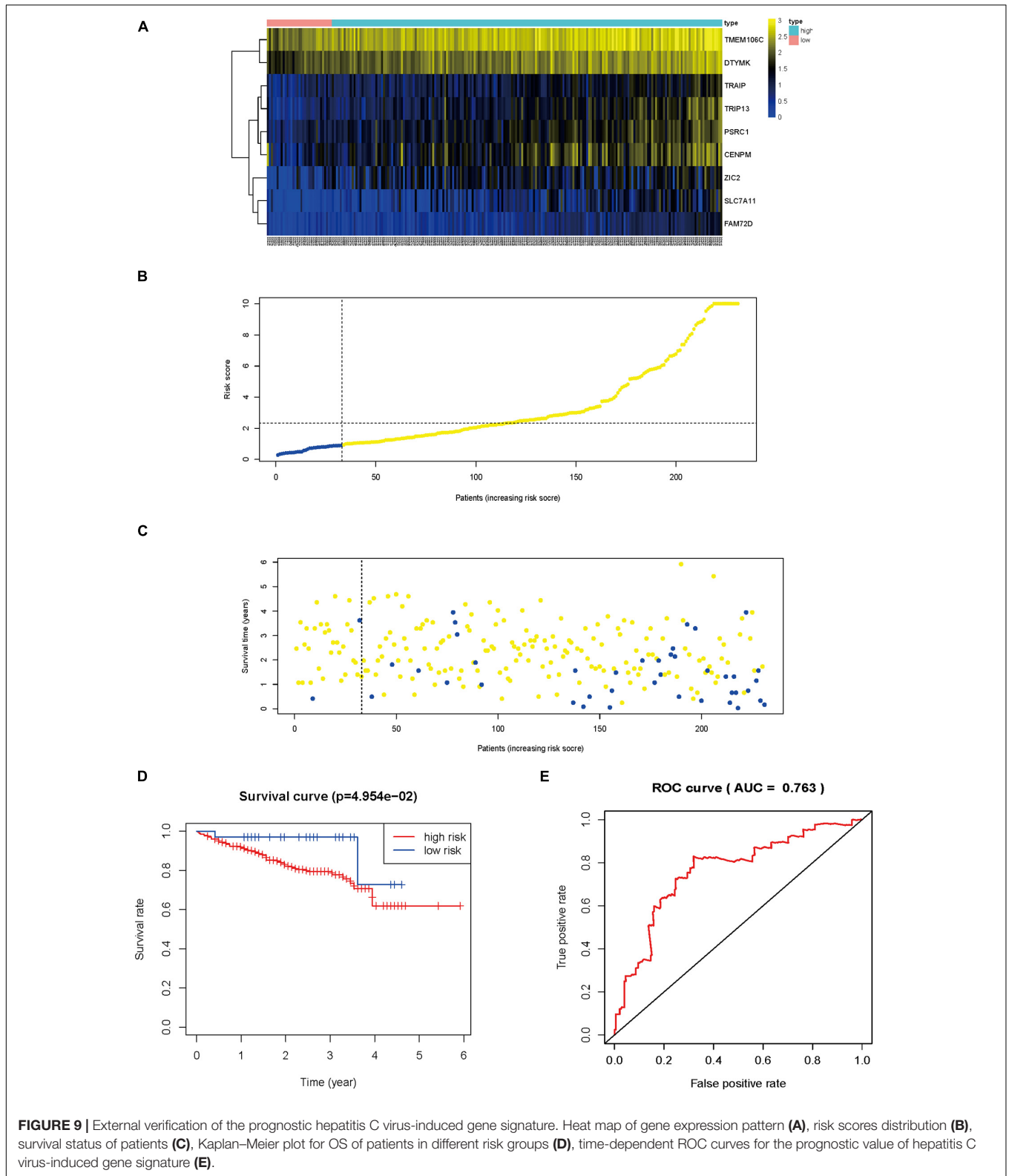
Many reports have shown that ncRNAs play important roles in many cancers (Hansen et al., 2013; Riquelme et al., 2016; Chen et al., 2020). Aberrant expression of ncRNAs is a key player in HCV-related HCC metastasis, invasion, dissemination, and recurrence (Bandiera et al., 2016; Li et al., 2019; Shehab-Eldeen et al., 2019). However, the functional involvement of HCV-induced genes in liver pathogenesis remains to be explored.

FIGURE 8 | GSEAs for hepatitis C virus-induced gene signatures in hepatic cancer. GSEA shows the GO **(A)** and KEGG pathways **(B)** enriched in the high-risk group of the hepatitis C virus-induced gene signatures in hepatic cancer. GSEA reveals the GO **(C)** and KEGG pathways **(D)** enriched in the low-risk group of the hepatitis C virus-induced gene signatures in hepatic cancer.

In the present study, we analyzed a total of 143 differentially expressed HCVIGs in TCGA hepatic cancer dataset using the mRNA expression of 1,582 differentially expressed HCVIGs in GSE70781. With univariate and multivariate Cox regressions, a total of 21 prognostic genes were identified. Finally, the signature of nine prognostic HCVIGs was constructed. The Kaplan–Meier Plotter was used in this study to validate the prognostic value of the nine HCVIGs. To the best of our knowledge, this study was the first to establish a nine prognostic HCVIG signature predicting prognosis in hepatic cancer.

The zinc finger of the cerebellum (ZIC) gene family is composed of five members, Zic1–Zic5 (Zhao et al., 2019). The transcription factor zinc family member 2 (*ZIC2*) has been reported to promote proliferation, invasion, and progression in HCC (Zhu et al., 2015; Lu et al., 2017); we found that the overexpression of *ZIC2* had a poor prognosis. The cystine/glutamate antiporter solute carrier family 7 member

11 (*SLC7A11*) is a novel prognostic biomarker for hepatic carcinoma (Zhang et al., 2018). Proline and serine-rich coiled-coil 1 (*PSRC1*), which is encoded by *PSRC1* and is also known as DDA3, has been shown to be involved in HCC (Yang et al., 2011). Transmembrane protein 106C (*TMEM106C*) is differentially expressed and promotes the development of HCC (Luo et al., 2020). TRAF-interacting protein (*TRAIP*) is a master regulator of DNA interstrand crosslink repair (Wu et al., 2019). Guo et al. (2020) showed that *TRAIP* promotes malignant behaviors and is correlated with poor prognosis in liver cancer. Deoxythymidylate kinase (*DTYMK*) is a novel gene associated with mitochondrial DNA depletion syndrome (MDDS) (Lam et al., 2019). Wang et al. (2020) reported that *DTYMK* is a crucial gene associated with immune cell infiltration in HCC. Family with sequence similarity 72-member D (*FAM72D*) is a prognostic gene signature for kidney renal cell carcinoma (Hu et al., 2019); however, the mechanism of *FAM72D* in HCC remains unclear. Thyroid

FIGURE 9 | External verification of the prognostic hepatitis C virus-induced gene signature. Heat map of gene expression pattern **(A)**, risk scores distribution **(B)**, survival status of patients **(C)**, Kaplan–Meier plot for OS of patients in different risk groups **(D)**, time-dependent ROC curves for the prognostic value of hepatitis C virus-induced gene signature **(E)**.

hormone receptor interactor 13 (*TRIP13*) is an AAA+ ATPase that plays an important role in mitotic checkpoint (Zhang et al., 2019). Zhu et al. (2019) showed that elevated TRIP13 drives

the AKT/mTOR pathway to induce the progression of HCC by interacting with ACTN4. Zheng et al. (2020) reported that the upregulation of Centromere protein M (*CENPM*) facilitates
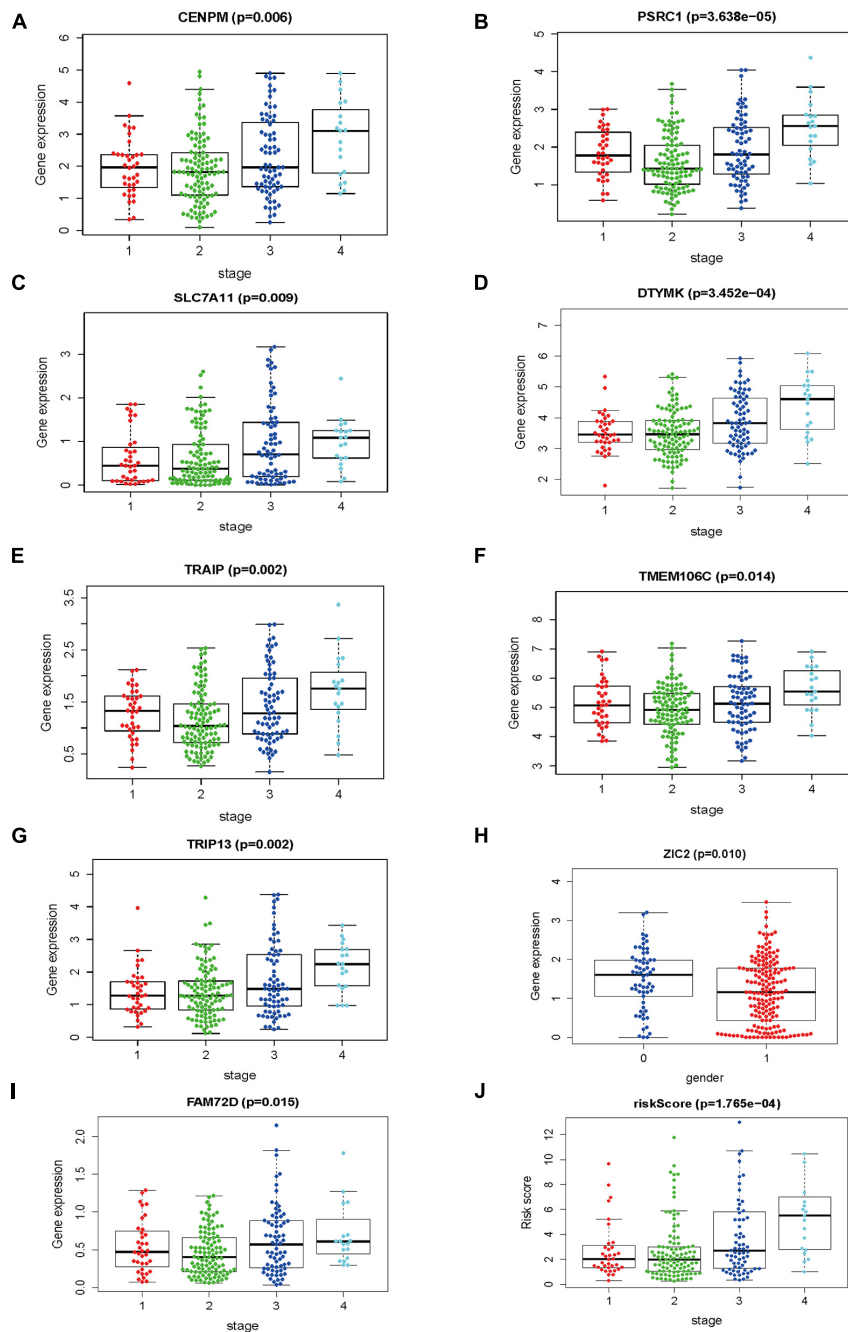
FIGURE 10 | The correlations between risk score and clinicopathological factors. **(A)** CENPM and stage, **(B)** *PSRC1* and stage, **(C)** *SLC7A11* and stage, **(D)** *DTYMK* and stage, **(E)** *TRAIP* and stage, **(F)** *TMEM106C* and stage, **(G)** *TRIP13* and stage, **(H)** *ZIC2* and gender **(I)** *FAM72D* and stage, and **(J)** risk score and stage.

tumor metastasis *via* the mTOR/p70S6K signaling pathway in pancreatic cancer. *CENPM* plays an important role in HCC (Xiao et al., 2019; Wu and Yang, 2020).

Bioinformatics enrichment analysis showed that 143 differentially expressed HCVIGs were mainly related to DNA replication. DNA replication is a crucial biological process that is tightly regulated to ensure accurate and complete transmission of genetic information to daughter cells (Boyer et al., 2016).

DNA replication is fundamental for cellular proliferation and genome stability. Interestingly, many studies have found that DNA replication plays a vital role in liver cancer development (Cervello et al., 2012; Kitao et al., 2018). According to this result, HCVIGs may be associated with host cell DNA replication.

In the present study, we developed and validated the prognostic signature of nine HCVIGs using TCGA, GEO, and ICGC databases. Then, *ZIC2, SLC7A11, PSRC1, TMEM106C,*

FIGURE 11 | The expression of nine prognostic HCVIGs in hepatic cancer. **(A)** *ZIC2*, **(B)** *SLC7A11*, **(C)** *PSRC1*, **(D)** *TMEM106C*, **(E)** *TRAIP*, **(F)** *DTYMK*, **(G)** *FAM72D*, **(H)** *TRIP13*, and **(I)** *CENPM*.
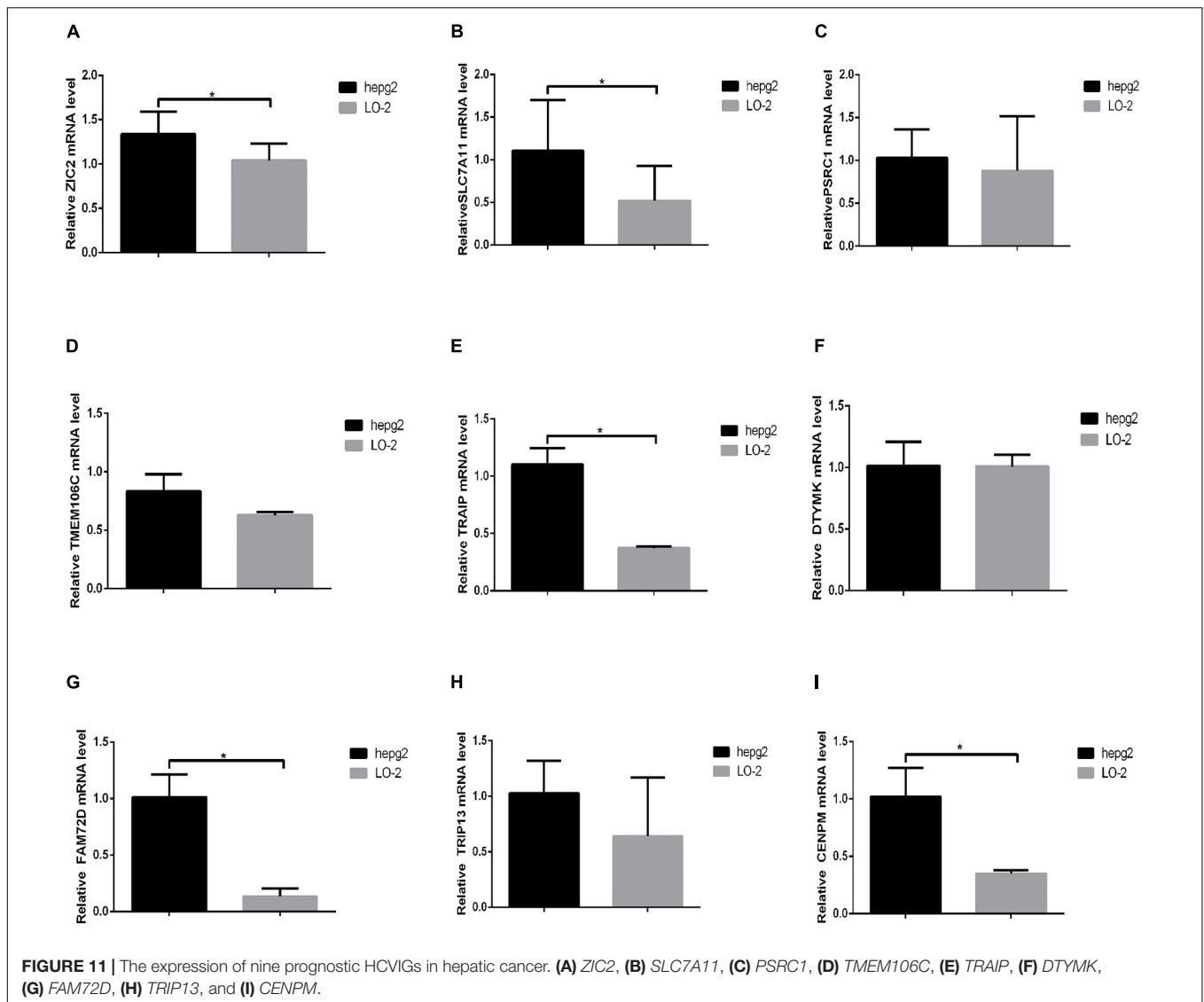
TRAIP, *DTYMK*, *FAM72D*, *TRIP13*, and *CENPM* were identified. Finally, *ZIC2*, *SLC7A11*, *PSRC1*, *TMEM106C*, *TRAIP*, *DTYMK*, *FAM72D*, *TRIP13*, and *CENPM* were upregulated in the hepatic cancer cell line HepG2. In conclusion, this prognostic nine HCVIGs signature could predict prognosis in HCC patients.

This study has two limitations. First, we were unable to collect sufficient hepatic cancer cases to validate the predictive power of this signature. Second, the mechanism of HCVIGs contained in gene signatures requires further study.

DATA AVAILABILITY STATEMENT

Hepatic cancer data was downloaded from the TCGA database (https://tcga-data.nci.nih.gov/tcga/) under the accession number(s) gdc_download_20201008_115541.090406. Hepatitis C virus-induced genes chip data sets was downloaded from Gene Expression Omnibus database GSE78092

(https://www.ncbi.nlm.nih.gov/gds/?term=GSE70781) under the accession number(s) GSE70781_series_matrix.txt.gz.

AUTHOR CONTRIBUTIONS

JW analyzed hepatitis C virus-induced gene expression data of hepatic cancer from TCGA database. BW and XG contributed to the writing and editing of the manuscript. All authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Ali, N., Allam, H., May, R., Sureban, S. M., Bronze, M. S., Bader, T., et al. (2011). Hepatitis C virus-induced cancer stem cell-like signatures in cell culture and murine tumor xenografts. *J. Virol.* 85, 12292–12303. doi: 10.1128/jvi. 05920-11

Bandiera, S., Pernot, S., El Saghire, H., Durand, S. C., Thumann, C., Crouchet, E., et al. (2016). Virus-induced upregulation of MicroRNA miR-146a-5p in hepatocytes promotes viral infection and deregulates metabolic pathways associated with liver disease pathogenesis. *J. Virol.* 90, 6387–6400. doi: 10.1128/jvi.00619-16

Bartosch, B., Thimme, R., Blum, H. E., and Zoulim, F. (2009). Hepatitis C virus-induced hepatocarcinogenesis. *J. Hepatol.* 51, 810–820. doi: 10.1016/j.jhep. 2009.05.008

Boyer, A. S., Walter, D., and Sørensen, C. S. (2016). DNA replication and cancer: from dysfunctional replication origin activities to therapeutic opportunities. *Semin. Cancer Biol.* 37-38, 16–25. doi: 10.1016/j.semcancer.2016.01.001

Cervello, M., Bachvarov, D., Lampiasi, N., Cusimano, A., Azzolina, A., McCubrey, J. A., et al. (2012). Molecular mechanisms of sorafenib action in liver cancer cells. *Cell Cycle* 11, 2843–2855. doi: 10.4161/cc.21193

Chen, J., Liu, X., Ke, K., Zou, J., Gao, Z., Habuchi, T., et al. (2020). LINC00992 contributes to the oncogenic phenotypes in prostate cancer via targeting miR-3935 and augmenting GOLM1 expression. *BMC Cancer* 20:749. doi: 10.1186/s12885-020-07141-4

Guo, Z., Zeng, Y., Chen, Y., Liu, M., Chen, S., Yao, M., et al. (2020). TRAIP promotes malignant behaviors and correlates with poor prognosis in liver cancer. *Biomed. Pharmacother.* 124:109857. doi: 10.1016/j.biopha.2020. 109857

Hansen, T. B., Kjems, J., and Damgaard, C. K. (2013). Circular RNA and miR-7 in cancer. *Cancer Res.* 73, 5609–5612. doi: 10.1158/0008-5472.can-13-1568

Hoshida, Y., Fuchs, B. C., Bardeesy, N., Baumert, T. F., and Chung, R. T. (2014). Pathogenesis and prevention of hepatitis C virus-induced hepatocellular carcinoma. *J. Hepatol.* 61(1 Suppl), S79–S90.

Hu, F., Zeng, W., and Liu, X. (2019). A gene signature of survival prediction for kidney renal cell carcinoma by multi-omic data analysis. *Int. J. Mol. Sci.* 20:5720. doi: 10.3390/ijms20225720

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.

Kitao, H., Iimori, M., Kataoka, Y., Wakasa, T., Tokunaga, E., Saeki, H., et al. (2018). DNA replication stress and cancer chemotherapy. *Cancer Sci.* 109, 264–271. doi: 10.1111/cas.13455

Lam, C. W., Yeung, W. L., Ling, T. K., Wong, K. C., and Law, C. Y. (2019). Deoxythymidylate kinase, DTYMK, is a novel gene for mitochondrial DNA depletion syndrome. *Clin. Chim. Acta* 496, 93–99. doi: 10.1016/j.cca.2019. 06.028

Li, J., Jin, B., Wang, T., Li, W., Wang, Z., Zhang, H., et al. (2019). Serum microRNA expression profiling identifies serum biomarkers for HCV-related hepatocellular carcinoma. *Cancer Biomark.* 26, 501–512. doi: 10.3233/cbm-181970

Lu, S. X., Zhang, C. Z., Luo, R. Z., Wang, C. H., Liu, L. L., Fu, J., et al. (2017). Zic2 promotes tumor growth and metastasis via PAK4 in hepatocellular carcinoma. *Cancer Lett.* 402, 71–80. doi: 10.1016/j.canlet.2017.05.018

Luo, X., Han, G., Lu, R., Guan, S., Wang, Y., Ju, L., et al. (2020). Transmembrane protein 106C promotes the development of hepatocellular carcinoma. *J. Cell. Biochem.* 121, 4484–4495. doi: 10.1002/jcb.29678

McGlynn, K. A., Petrick, J. L., and El-Serag, H. B. (2021). Epidemiology of hepatocellular carcinoma. *Hepatology* 73(Suppl. 1), 4–13.

Menyhárt, O., Nagy, Á, and Győrffy, B. (2018). Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma. *R. Soc. Open Sci.* 5:181006. doi: 10.1098/rsos.181006

Nagy, A., Lanczky, A., Menyhart, O., and Gyorffy, B. (2018). Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci. Rep.* 8:9227.

Ninio, L., Nissani, A., Meirson, T., Domovitz, T., Genna, A., Twafra, S., et al. (2019). Virus enhances the invasiveness of hepatocellular carcinoma via EGFR-mediated invadopodia formation and activation. *Cells* 8:1395. doi: 10.3390/cells8111395

Peng, J., Wang, H., Lu, J., Hui, W., Wang, Y., and Shang, X. (2017). Identifying term relations cross different gene ontology categories. *BMC Bioinformatics* 18(Suppl. 16):573. doi: 10.1186/s12859-017-1959-3

Plissonnier, M. L., Herzog, K., Levrero, M., and Zeisel, M. B. (2018). Non-coding RNAs and hepatitis C virus-induced hepatocellular carcinoma. *Viruses* 10:591. doi: 10.3390/v10110591

Qiu, J., Sun, M., Wang, Y., and Chen, B. (2020). Identification and validation of an individualized autophagy-clinical prognostic index in gastric cancer patients. *Cancer Cell Int.* 20:178.

Refai, N. S., Louka, M. L., Halim, H. Y., and Montasser, I. (2019). Long non-coding RNAs (CASC2 and TUG1) in hepatocellular carcinoma: clinical significance. *J. Gene Med.* 21:e3112.

Riquelme, I., Letelier, P., Riffo-Campos, A. L., Brebi, P., and Roa, J. C. (2016). Emerging role of miRNAs in the drug resistance of gastric cancer. *Int. J. Mol. Sci.* 17:424. doi: 10.3390/ijms17030424

Sasaki, R., Devhare, P., Ray, R. B., and Ray, R. (2017). Hepatitis C virus-induced tumor-initiating cancer stem-like cells activate stromal fibroblasts in a xenograft tumor model. *Hepatology* 66, 1766–1778. doi: 10.1002/hep.29346

Shehab-Eldeen, S., Nada, A., Abou-Elela, D., El-Naidany, S., Arafat, E., and Omar, T. (2019). Diagnostic performance of microRNA-122 and microRNA-224 in hepatitis C virus-induced hepatocellular carcinoma (HCC). *Asian Pac. J. Cancer Prev.* 20, 2515–2522. doi: 10.31557/apjcp.2019.20.8.2515

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi: 10.3322/caac.21660

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368.

Vescovo, T., Refolo, G., Vitagliano, G., Fimia, G. M., and Piacentini, M. (2016). Molecular mechanisms of hepatitis C virus-induced hepatocellular carcinoma. *Clin. Microbiol. Infect.* 22, 853–861. doi: 10.1016/j.cmi.2016.07.019

Wang, D., Liu, J., Liu, S., and Li, W. (2020). Identification of crucial genes associated with immune cell infiltration in hepatocellular carcinoma by weighted gene co-expression network analysis. *Front. Genet.* 11:342. doi: 10.3389/fgene.2020.00342

Wang, Z., Wang, Z., Niu, X., Liu, J., Wang, Z., Chen, L., et al. (2019). Identification of seven-gene signature for prediction of lung squamous cell carcinoma. *Onco Targets Ther.* 12, 5979–5988. doi: 10.2147/ott.s198998

Wu, R. A., Semlow, D. R., Kamimae-Lanning, A. N., Kochenova, O. V., Chistol, G., Hodskinson, M. R., et al. (2019). TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature* 567, 267–272. doi: 10.1038/s41586-019-1002-0

Wu, Z. H., and Yang, D. L. (2020). High CENPM mRNA expression and its prognostic significance in hepatocellular carcinoma: a study based on data mining. *Cancer Cell Int.* 20:406.

Xiao, Y., Najeeb, R. M., Ma, D., Yang, K., Zhong, Q., and Liu, Q. (2019). Upregulation of CENPM promotes hepatocarcinogenesis through mutiple mechanisms. *J. Exp. Clin. Cancer Res.* 38:458.

Yang, D., Meng, X., Xue, B., Liu, N., Wang, X., and Zhu, H. (2014). MiR-942 mediates hepatitis C virus-induced apoptosis via regulation of ISG12a. *PLoS One* 9:e94501. doi: 10.1371/journal.pone.0094501

SUPPLEMENTARY MATERIAL

Yang, G., Jian, L., Lin, X., Zhu, A., and Wen, G. (2019). Bioinformatics analysis of potential key genes in trastuzumab-resistant gastric cancer. *Dis. Markers* 2019:1372571.

Yang, J. D., Seol, S. Y., Leem, S. H., Kim, Y. H., Sun, Z., Lee, J. S., et al. (2011). Genes associated with recurrence of hepatocellular carcinoma: integrated analysis by gene expression and methylation profiling. *J. Korean Med. Sci.* 26, 1428–1438. doi: 10.3346/jkms.2011.26.11.1428

Zhang, G., Zhu, Q., Fu, G., Hou, J., Hu, X., Cao, J., et al. (2019). TRIP13 promotes the cell proliferation, migration and invasion of glioblastoma through the FBXW7/c-MYC axis. *Br. J. Cancer* 121, 1069–1078. doi: 10.1038/s41416-019-0633-0

Zhang, L., Huang, Y., Ling, J., Zhuo, W., Yu, Z., Luo, Y., et al. (2018). Overexpression of SLC7A11: a novel oncogene and an indicator of unfavorable prognosis for liver carcinoma. *Future Oncol.* 14, 927–936. doi: 10.2217/fon-2017-0540

Zhang, Y., Wei, W., Cheng, N., Wang, K., Li, B., Jiang, X., et al. (2012). Hepatitis C virus-induced up-regulation of microRNA-155 promotes hepatocarcinogenesis by activating Wnt signaling. *Hepatology* 56, 1631–1640. doi: 10.1002/hep.25849

Zhao, Z., Wang, L., Bartom, E., Marshall, S., Rendleman, E., Ryan, C., et al. (2019). β-Catenin/Tcf7l2-dependent transcriptional regulation of GLUT1 gene expression by Zic family proteins in colon cancer. *Sci. Adv.* 5:eaax0698. doi: 10.1126/sciadv.aax0698

Zheng, C., Zhang, T., Li, D., Huang, C., Tang, H., Ni, X. F., et al. (2020). Upregulation of CENPM facilitates tumor metastasis via the mTOR/p70S6K signaling pathway in pancreatic cancer. *Oncol. Rep.* 44, 1003–1012. doi: 10.3892/or.2020.7673

Zhou, Z., Li, Y., Hao, H., Wang, Y., Zhou, Z., Wang, Z., et al. (2019). Screening hub genes as prognostic biomarkers of hepatocellular carcinoma by bioinformatics analysis. *Cell Transplant.* 28(1_suppl), 76S–86S. doi: 10.1177/0963689719893950

Zhu, M. X., Wei, C. Y., Zhang, P. F., Gao, D. M., Chen, J., Zhao, Y., et al. (2019). Elevated TRIP13 drives the AKT/mTOR pathway to induce the progression of hepatocellular carcinoma via interacting with ACTN4. *J. Exp. Clin. Cancer Res.* 38:409.

Zhu, P., Wang, Y., He, L., Huang, G., Du, Y., Zhang, G., et al. (2015). ZIC2-dependent OCT4 activation drives self-renewal of human liver cancer stem cells. *J. Clin. Invest.* 125, 3795–3808. doi: 10.1172/jci81979