World Journal of
Gastrointestinal Oncology

Contents

ABOUT COVER

Editorial Board Member of *World Journal of Gastrointestinal Oncology*, Joseph Chao, MD, Assistant Professor, Medical Oncology and Therapeutics Research, City of Hope Comprehensive Cancer Center, Duarte, CA 91010, United States

AIMS AND SCOPE

The primary aim of *World Journal of Gastrointestinal Oncology* (*WJGO*, *World J Gastrointest Oncol*) is to provide scholars and readers from various fields of gastrointestinal oncology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.
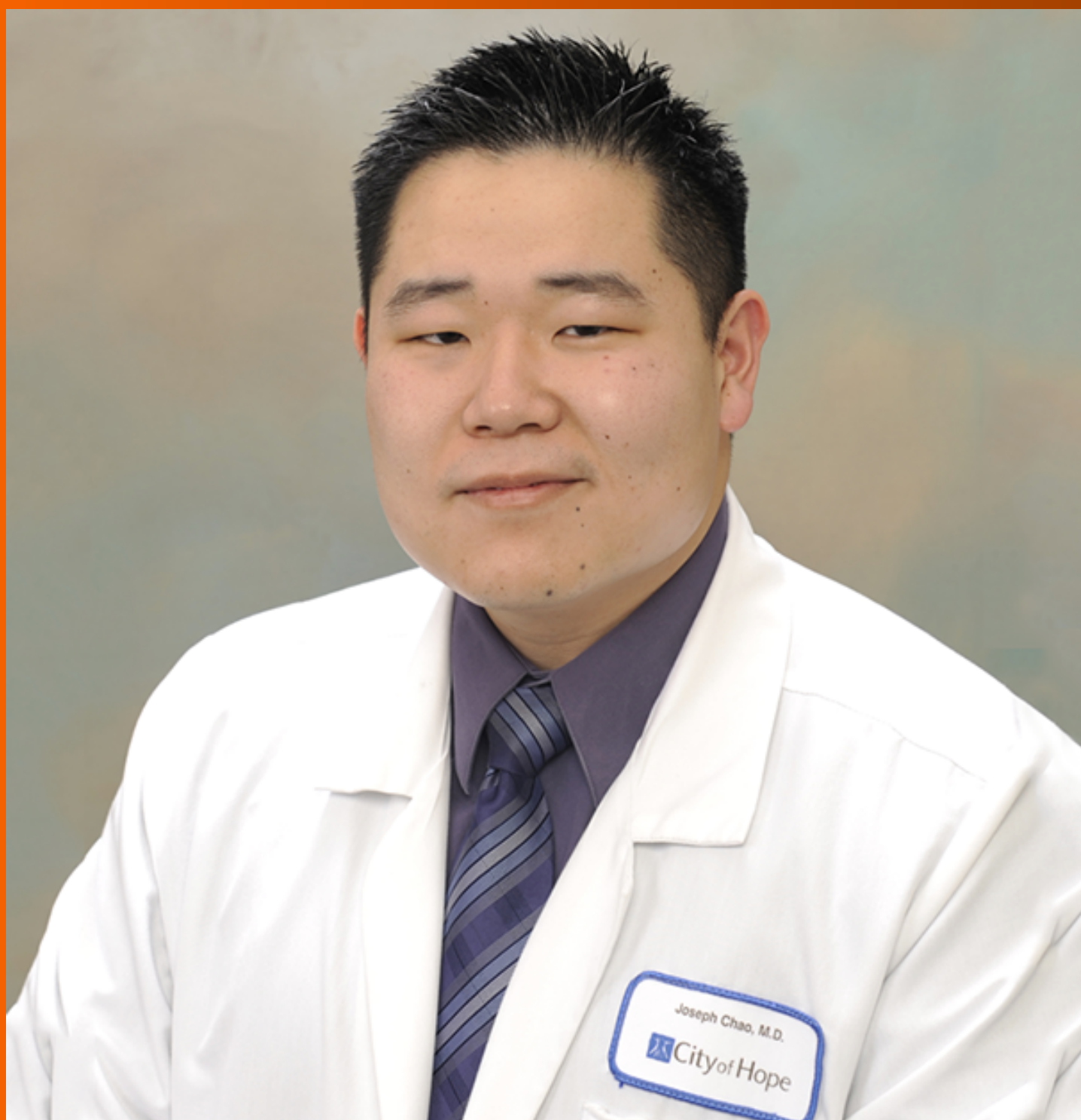
   WJGO mainly publishes articles reporting research results and findings obtained in the field of gastrointestinal oncology and covering a wide range of topics including islet cell adenoma, liver cell adenoma, adenomatous polyposis coli, appendiceal neoplasms, bile duct neoplasms, biliary tract neoplasms, hepatocellular carcinoma, islet cell carcinoma, pancreatic ductal carcinoma, cecal neoplasms, colonic neoplasms, colorectal neoplasms, hereditary nonpolyposis colorectal neoplasms, common bile duct neoplasms, duodenal neoplasms, esophageal neoplasms, gallbladder neoplasms, *etc*.

INDEXING/ABSTRACTING

The *WJGO* is now indexed in Science Citation Index Expanded (also known as SciSearch®), PubMed, and PubMed Central. The 2019 edition of Journal Citation Reports® cites the 2018 impact factor for *WJGO* as 2.758 (5-year impact factor: 3.220), ranking *WJGO* as 52 among 84 journals in gastroenterology and hepatology (quartile in category Q3), and 131 among 229 journals in oncology (quartile in category Q3).

RESPONSIBLE EDITORS FOR THIS ISSUE

Basic Study

Eight key long non-coding RNAs predict hepatitis virus positive hepatocellular carcinoma as prognostic targets

Zi-Lin Huang, Wang Li, Qi-Feng Chen, Pei-Hong Wu, Lu-Jun Shen

ORCID number: Zi-Lin Huang (0000-0002-1643-1994); Wang Li (0000-0002-4293-8182); Qi-Feng Chen (0000-0001-8998-3450); Pei-Hong Wu (0000-0001-8008-1872); Lu-Jun Shen (0000-0002-7936-0206).

Author contributions: Huang ZL, Li W, and Chen QF contributed equally to this work; Huang ZL, Li W, and Chen QF contributed to study conceptualization; Chen QF contributed to the methodology; Wu PH contributed to software; Chen QF, Huang ZL, and Li W contributed to data validation; Chen QF, Wu PH, Huang ZL, and Shen LJ analyzed the data and contributed to manuscript writing and editing; Chen QF and Li W contributed to manuscript drafting; Chen QF contributed to visualization and supervised the final paper.

Institutional review board statement: Not applicable, because the data were publicly available.

Institutional animal care and use committee statement: Not applicable, because no animal was used in the present study.

Conflict-of-interest statement: The authors deny any conflict of interest.

Data sharing statement: The data used in this manuscript are accessible through: https://portal.gdc.cancer.gov/

Zi-Lin Huang, Wang Li, Qi-Feng Chen, Pei-Hong Wu, Lu-Jun Shen, Department of Medical Imaging and Interventional Radiology, Sun Yat-sen University Cancer Center, Guangzhou 510060, Guangdong Province, China

Zi-Lin Huang, Wang Li, Qi-Feng Chen, Pei-Hong Wu, Lu-Jun Shen, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, Guangdong Province, China

Corresponding author: Qi-Feng Chen, MD, Doctor, Department of Medical Imaging and Interventional Radiology, Sun Yat-sen University Cancer Center, 651 East Dongfeng Road, Yuexiu District, Guangzhou 510060, Guangdong Province, China. ch_peaks@126.com
Telephone: +86-15626062848
Fax: +86-20-87343392

Abstract

BACKGROUND
Hepatitis B virus, together with hepatitis C virus, has been recognized as the leading causes of hepatocellular carcinoma (HCC). Long non-coding RNAs (lncRNAs) have been suggested in increasing studies to be the potential prognostic factors for HCC. However, the role of combined application of lncRNAs in estimating overall survival (OS) for hepatitis virus positive HCC (VHCC) is uncertain.

AIM
To construct an lncRNA signature related to the OS of VHCC patients to enhance the accuracy of prognosis prediction.

METHODS
The expression patterns of lncRNAs, as well as related clinical data were collected from 149 VHCC patients from The Cancer Genome Atlas database. The R package was adopted to obtain the differentially expressed lncRNAs (DElncRNAs). LncRNAs significantly associated with OS were screened by means of univariate Cox regression analysis, so as to construct a least absolute shrinkage and selection operator (LASSO) model. Subsequently, the constructed lncRNA signature was developed and validated. Afterwards, the prognostic nomogram was established, which combined the as-established lncRNA signature as well as the clinical features. Meanwhile, subgroup analysis stratified by the virus type was also performed. Finally, the above-mentioned lncRNAs were enriched to corresponding pathways according to the markedly co-

expressed genes.

RESULTS

A total of 1420 DElncRNAs were identified, among which 406 were significant in univariate Cox regression analysis. LASSO regression confirmed 8 out of the 406 lncRNAs, including AC005722.2, AC107959.3, AL353803.1, AL589182.1, AP000844.2, AP002478.1, FLJ36000, and NPSR1-AS1. Then, the prognostic risk score was calculated. Our results displayed a significant association between the risk model and the OS of VHCC [hazard ratio = 1.94, 95% confidence interval (CI): 1.61-2.34, log-rank P = 2e-10]. The inference tree suggested that the established lncRNA signature was useful in the risk stratification of VHCC. Furthermore, a nomogram was plotted, and the concordance index of internal validation was 0.763 (95%CI: 0.700-0.826). Moreover, the subgroup analysis regarding etiology confirmed this risk model. In addition, the Wnt signaling pathway, angiogenesis, the p53 pathway, and the PI3 kinase pathway were the remarkably enriched pathways.

CONCLUSION

An eight-lncRNA signature has been established to predict the prognosis for VHCC, which contributes to providing a novel foundation for the targeted therapy of VHCC.

Key words: Long non-coding RNAs; Hepatitis virus; Hepatocellular carcinoma; Prognostic signature; Least absolute shrinkage and selection operator

Core tip: The existing long non-coding RNA (lncRNA) signatures for prognosis prediction are still controversial and need to be optimized for hepatocellular carcinoma (HCC). Typically, there is less knowledge concerning the lncRNA signature for hepatitis virus positive HCC (VHCC). As a result, this study was carried out to construct a robust lncRNA signature to predict the prognosis for VHCC. Typically, the eight-lncRNA signature constructed in this study displayed a potent potential to predict the prognosis of VHCC, which might contribute to risk stratification, and provide the individualized clinical suggestion for an individual case.

INTRODUCTION

Hepatocellular carcinoma (HCC), one of the most frequent liver cancer types, is a severe worldwide health problem[1]. Chronic hepatitis B virus (HBV) infection and/or hepatitis C virus (HCV) infection have been identified as the leading causes of HCC[2]. Specifically, the prevalence of HCV infection is predominant in European countries, which accounts for about 60% of all HCC cases in Italy and Spain; by contrast, HBV infection is more commonly seen in eastern Asian countries such as China, with an incidence rate of over 2/3[3]. Specifically, hepatitis viruses have been epidemiologically associated with the development of HCC[4,5].

It has been extensively determined that hepatitis viruses are associated with HCC etiologically; nonetheless, it is still not completely clear about the molecular mechanism of hepatitis viruses in inducing liver cancer. An increasing number of mutation genes, like TP53, CTNNB1, and mTOR[6], are found to participate in HCC genesis as well as progression. Nonetheless, HCC is a highly heterogeneous disease, which can determine the altered biological progression of HCC and add to the difficulty in predicting prognosis. As a result, it is urgently needed to search for the efficient prognostic biomarkers for HCC.

Long non-coding RNAs (LncRNAs) have been identified to be the superior

prognostic biomarkers over other cancer hallmarks, which can be ascribed to their unique advantages, as shown below. First, lncRNA expression is greatly distinct among different tissues, in various diseases, as well as at different disease progression stages; as a result, they can better represent the disease characteristics[7]. Second, lncRNAs have been found to regulate the expression of genes at the transcriptional, post-transcriptional, and epigenetic levels[8,9]; on this account, their levels as well as functions show a close association with cancer progression. As a result, increasing studies have been carried out to illustrate the clinical value of lncRNAs among cancers, such as HCC[10].

In this study, hepatitis virus positive HCC (VHCC) patients were appropriately selected to explore the difference in the lncRNA expression profiles, so as to identify the candidate lncRNA biomarkers based on The Cancer Genome Atlas (TCGA) database. Moreover, the key lncRNAs were determined using the least absolute shrinkage and selection operator (LASSO) algorithm; thereafter, the risk score system for VHCC was established. This study aimed to establish a potent lncRNA signature based on the expression profiles through thorough genomic data analysis, thus improving the accuracy in predicting the prognosis of VHCC.

MATERIALS AND METHODS

Patient datasets as well as processing

The data were downloaded from TCGA database. Gene expression, together with the clinical information of HCC patients (https://tcga-data.nci.nih.gov/), was downloaded using the Data Transfer Tool (provided by GDC Apps). Among them, 151 patients were virus-related. Patients with unknown lncRNA expression were eliminated from this study (n = 2); as a result, a total of 149 HCC cases were finally enrolled for analysis. The flowchart of analysis is shown in Figure 1. A total of 149 virus-induced HCC cases were enrolled for comprehensive integrated analysis. The data were publicly available, so approval was exempted from the Ethics Committee. Data were processed in accordance with the data access policies as well as the NIH TCGA human subject protection policies (http://cancergenome.nih.gov/publications/publicationguidelines).

LncRNA expression in HCC patients was derived from Illumina HiSeq RNASeq platform, which was then standardized based on TCGA. Then, the R software "edgeR" package was employed to detect the differentially expressed RNAs, so as to identify the differentially expressed lncRNAs (DElncRNAs), and the thresholds were set at $\log_2$ fold change > 2.0 as well as adjusted P value < 0.05.

Construction of lncRNA signature

The relationship of each lncRNA expression with the overall survival (OS) of patients was calculated using the univariate Cox model. Typically, lncRNAs would be deemed as statistically significant at P < 0.05 upon univariate Cox analysis. Subsequently, the chosen lncRNAs were screened and verified through LASSO regression with the R software "glmnet" package. Finally, the lncRNA-based prognostic risk score was constructed by linearly combining the products of expression level with the regression model (β) according to the formula below: Risk_score = βlncRNA1 × lncRNA1 expression + βlncRNA2 × lncRNA2 expression + · ···· + βlncRNAn × lncRNAn expression.

Confirmation of lncRNA signature

Patients together with their survival information were distributed according to the risk_score. Furthermore, patients were classified into a high-risk group and low-risk group according to the median risk_score. Then, the Kaplan-Meier survival curves were plotted, which could predict the high or low risk of patients. Thereafter, the univariate Cox proportional hazards regression analysis was performed. Then, the sensitivity and specificity regrading survival prediction were compared by the risk_score, and the receiver operating characteristic (ROC) curves with time dependence was then employed to evaluate the accuracy in predicting the 5-year prognosis. Moreover, multivariate Cox regression was also performed to examine whether lncRNA risk_score was predicted independently from other clinical factors. The R-package "party" "ctree" function was used to construct a conditional inference tree for better illustration. Besides, a nomogram was also established with the R project "rms" package. In addition, the predictive performance of the nomogram was verified for discrimination and calibration. Subsequently, subgroup analysis based on virus type was carried out. All the two-sided P < 0.05 would be deemed to be statistically significant. The R software (version 3.4.1, R Foundation) was employed
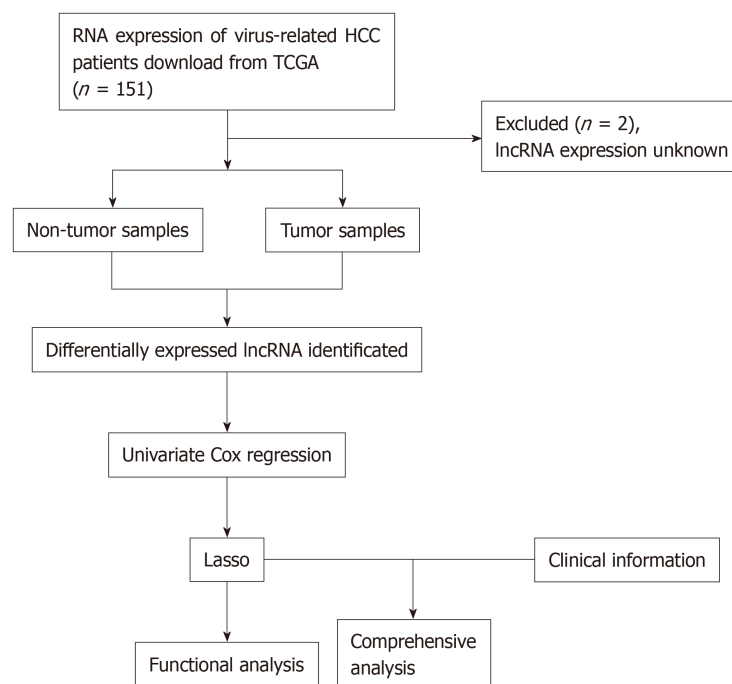
Figure 1 Flowchart of study design. HCC: Hepatocellular carcinoma; TCGA: The Cancer Genome Atlas; lncRNA: Long non-coding RNAs.

for all analyses.

Functional analysis of lncRNAs for prognosis prediction

Pearson correlation analysis was carried out for the screened lncRNAs as well as the protein coding genes in the TCGA database according to their different expression quantities. Typically, a coefficient of correlation > 0.4 and $P < 0.001$ would be deemed to show a significant correlation. Enrichment analysis was performed with the DAVID Bioinformatics Tool (version 6.8, https://david-d.ncifcrf.gov/).

RESULTS

DElncRNA identification

The significant DElncRNAs were identified by comparing the tumor samples with non-tumor samples. Altogether, 1420 DElncRNAs (including 1358 up-regulated and 62 down-regulated ones) were recognized using the R software "edgeR" package. Afterwards, the heat map (Figure 2A) and volcano of top 20 DElncRNAs (Figure 2B) were built.

Construction of an lncRNA signature

Univariate Cox regression analysis was performed between DElncRNAs and patients' survival; as suggested by the results, altogether 406 lncRNAs were significantly associated with OS ($P < 0.05$). Next, LASSO regression was employed for verifying further variables (Figure 3). Eight lncRNAs were generated in this process, including AC005722.2, AC107959.3, AL353803.1, AL589182.1, AP000844.2, AP002478.1, FLJ36000, and NPSR1-AS1. Then, the prognostic risk_score was imputed as follows: (0.15103*AC005722.2 expression)+(0.07857*AC107959.3 expression)+(0.09283* AL353803.1 expression)+(0.17744*AL589182.1 expression) + (0.07857*AP000844.2 expression)+(0.09888*AP002478.1 expression)+(0.13460*FLJ36000 expression)+ (0.06309*NPSR1-AS1 expression).

Confirmation of the lncRNA signature

The risk score of every case was calculated; then, patients were classified into a high-risk group or low-risk group based on the median threshold (Figure 4A and B). The Kaplan-Meier curves for the high-risk as well as low-risk groups are presented in Figure 4C. Patients in the high-risk group were associated with a shorter OS compared with those in the low-risk group ($P = 2e-10$). Meanwhile, the risk score hazard ratio (HR) upon univariate Cox proportional hazards regression was 5.360

A

B

Figure 2 Heat map and volcano plot. A: Heat map demonstrating the differential expression of long non-coding RNAs (lncRNAs). Th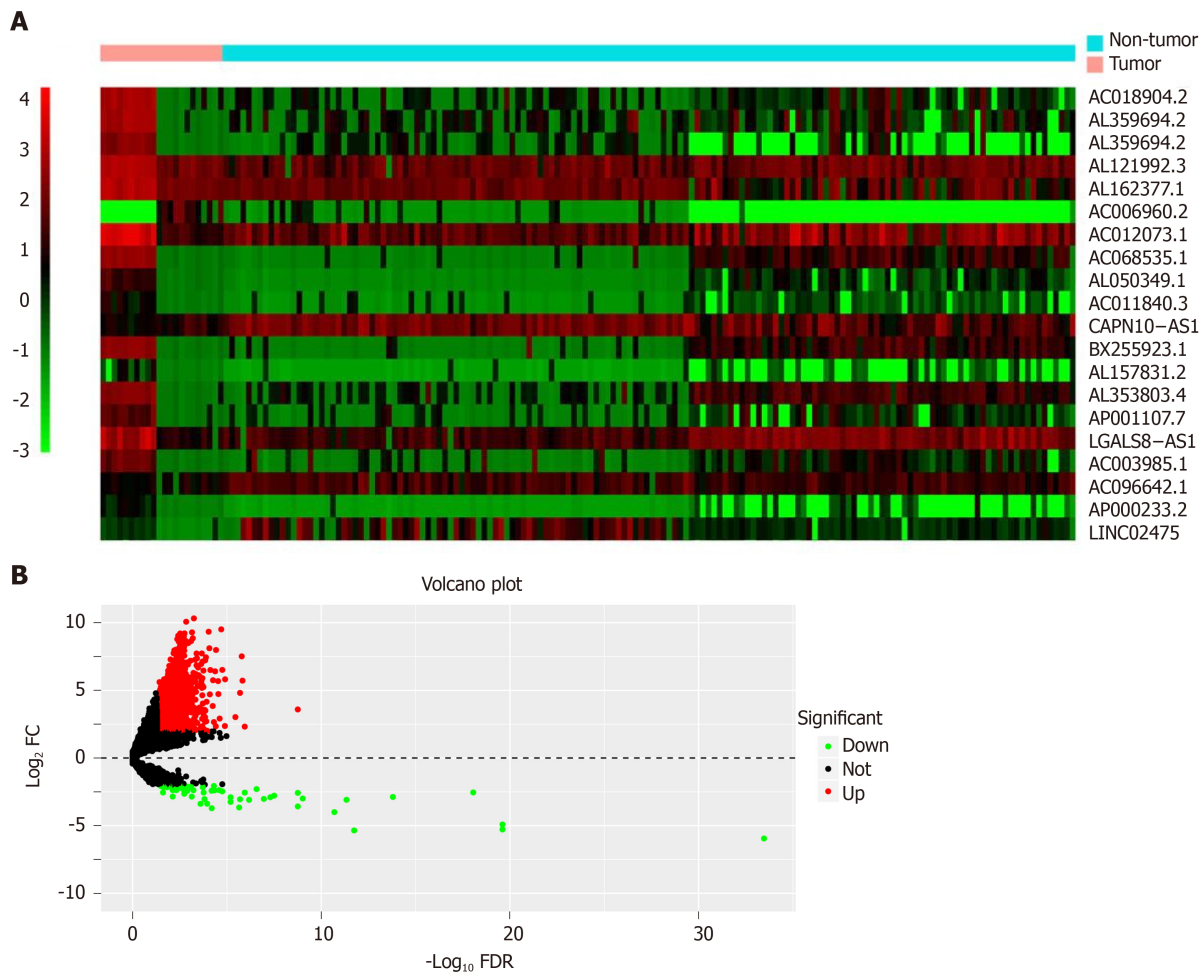e X-axis represents the samples, while the Y-axis denotes the differentially expressed lncRNAs. Green color is indicative of the down-regulated genes, while the red color stands for the up-regulated genes; B: Volcano plot showing the differential expression of lncRNAs. The X-axis indicates the false discovery rate values after log transformation, whereas the Y-axis suggests the log$_2$ (fold change) values for lncRNA expression. FDR: False discovery rate; FC: Fold change.

[95% confidence interval (CI): 3.012-9.538] (Table 1). Besides, multivariate Cox proportional hazards regression with adjusted clinical covariate also came to the same results (HR=1.94, 95%CI: 1.61-2.34) (Table 2). Subsequently, the ROC curves with time dependency were adopted to assess the prognostic value of the eight-lncRNA signature. In addition, the areas under curves (AUCs) of the prognostic model constructed based on the eight-lncRNA signature were 0.73, 0.758, and 0.788 for 1-, 3-, and 5-year OS, respectively (Figure 4D).

The first node of the conditional inference tree divided the data according to the risk_score, and the lower risk_score (≤0.855) was related to a better OS ($P < 0.001$) (Figure 5). For the lower risk_score (≤0.855), the branch was further divided depending on the stage ($P = 0.006$), and it was obvious that the risk_score ≤ 0.855 in stage I or II patients was associated with the best survival, while patients with the risk_score > 0.855 had the worst survival.

A nomogram for prognosis was constructed based on the above-mentioned independent factors for prognosis for the sake of visualization as well as prediction, so as to construct an integrated predicting factor for HCC. According to the multivariate Cox regression analysis, the eight-lncRNA risk_score, TNM stage, and race were recognized as independent risk factors ($P < 0.05$) (Table 2). Specifically, the nomogram was constituted by three factors (Figure 6A), with a C-index of 0.763 (95%CI: 0.700-0.826). Moreover, the calibrated curves for the risk of 1-, 3-, and 5-year recurrence were highly consistent between the nomogram predicted results and the observed results (Figure 6B-D).

Subgroup analysis by etiology

Subsequently, the patients were further classified into three subgroups, namely, HBV + HCV ($n = 82$), HBV alone ($n = 58$), and HCV alone ($n = 9$) groups. Kaplan-Meier
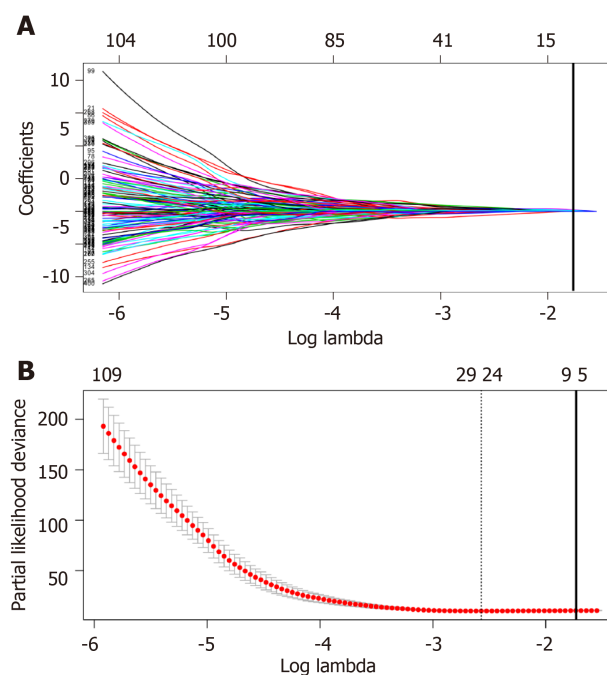
Figure 3 Plots of regression coefficient based on least absolute shrinkage and selection operator regression. A: Least absolute shrinkage and selection operator (LASSO) coefficient profiles of the fractions of the 406 significant long non-coding RNAs (lncRNAs) in univariate Cox regression analysis. The solid vertical line implies the optimal lambda value determined by cross-validation; B: Cross-validation for tuning parameter selection in the LASSO model. The vertical lines are drawn at the optimal values by minimum criteria and 1-standard error criteria. And the solid vertical line represents that eight lncRNAs were finally determined.

curves for the high-risk and low-risk groups are displayed in Figure 7. The OS of patients benefited from the low-risk score in the HBV + HCV group (P = 6e-5, Figure 7A) and HBV alone group (P = 5e-6, Figure 7B). At the same time, the AUCs regarding the 1-, 3-, and 5-year OS in the HBV + HCV model were 0.681, 0.759, and 0.724, respectively (Figure 7D), while those in the HBV model were 0.787, 0.720, and 0.826, respectively (Figure 7E). Curves and corresponding P values for HCV patients are also shown in Figure 7C and F.

Functional analysis of prognostic lncRNAs

LncRNA-related mRNAs were first filtered out by Pearson correlation analysis and then applied into online tool DAVID for pathway enrichment, so as to examine the possible biological functions of the above-mentioned identified lncRNAs. The top 10 pathways with the highest enrichment are presented in Figure 8. Among those, the Wnt signaling pathway, angiogenesis, the p53 pathway, and the PI3 kinase pathway were recognized as the canonical pathways in the initiation as well as progression of HCC.

DISCUSSION

Few HCC prognostic models have been constructed on the basis of lncRNAs up to now. To our knowledge, little is known about lncRNA signature for hepatitis virus infection associated HCC patients. Meanwhile, the majority of studies analyzed virus combined with non-virus related HCC at the same time[11,12]. Different from previous studies that carried out analysis regardless of etiology, we conducted the current research in virus positive patients only. The DElncRNAs with marked correlation with OS were comprehensively screened by applying biostatistical methods and univariate Cox analysis. Then, the LASSO regression was applied to analyze the lncRNA data from the carefully chosen patients from TCGA, and eight lncRNAs were identified, which were recognized to be the independent factors for the prognosis of VHCC. Afterwards, the identified significant lncRNAs were employed for constructing a prognostic model, and its prognostic value was verified through Kaplan-Meier analysis, time-dependent ROC curve, and Cox regression analysis. A nomogram was developed to present the constructed risk_score system, and it displayed a favorable accuracy in OS prediction for VHCC. Further functional

Table 1 Stratified overall survival analysis

Characteristic	Overall survival		
	High-risk/low-risk	HR (95%CI)	*P* value
Overall	75/74	5.360 (3.012-9.538)	<0.001
Age (yr)			
≤65	45/50	4.962 (2.361-10.428)	<0.001
>65	29/25	5.512 (2.174-13.980)	<0.001
Gender			
Male	51/49	5.324 (2.628-10.789)	<0.001
Female	23/26	5.509 (2.023-15.003)	0.001
Race			
White	38/47	7.473 (3.259-17.137)	<0.001
Asian	30/28	4.032 (1.774-9.164)	0.001
Grade			
I/II	45/57	4.959 (2.492-9.869)	<0.001
III/IV	29/18	5.033 (1.721-14.717)	0.003
Stage			
I/II	45/47	11.154 (3.882-32.049)	<0.001
III/IV	29/28	3.398 (1.652-6.988)	0.001

HR: Hazard ratio; CI: Confidence interval.

analysis uncovered that several pathways play crucial roles in the effects of the eight lncRNAs on the progression of HCC.

Our research displayed obvious advantages when compared with other research on the prognostic value of lncRNAs in HCC. First, patient homogeneity was markedly enhanced through careful patient selection before analysis, since the ignorance of patient features might result in false relationships, thus invaliding the conclusions. Existing treatments can treat chronic viral hepatitis, and some research has been performed previously to domenstrate the influence of antiviral treatment on disease progression as well as HCC development[13]. Moreover, many studies suggest that the survival of virus-related HCC patients can be prolonged through curative hepatectomy and/or local tumor ablation combined with adjuvant antiviral therapy[14-16]. On the other hand, virus and non-virus HCCs have been recognized to have apparently different outcomes, and different therapeutic strategies were recommended for them. Therefore, the prognostic models should also be developed in a distinct way. All the 149 patients were virus positive in our study. Second, with regard to the methodology, the use of LASSO penalized regression could boost the accuracy of bioinformatic analysis. Different from the conventional stepwise regression applied in prior research, the LASSO algorithm contributed to analyzing all independent factors at the same time, which tended to screen the most significant variables[17]. Then, the coefficients of the less significant variables would become zero following introduction into the penalty after regularization[18]. Consequently, such formulation approach would display a higher accuracy than stepwise regression using the multivariate Cox model, especially in the presence of huge datasets, such as genomics[19]. Third, an easy-to-use nomogram model was constructed on the basis of Cox analysis among the carefully screened patients with clinicopathological features, which would facilitate the prediction for individual patients.

As hepatitis viruses vary in different countries[3], our cohort was also divided into three subgroups based on the virus type (namely, HBV alone, HCV alone, and combined). Subgroup survival analysis showed that the as-constructed eight-lncRNA signature performed well in patients with combined virus infection (5-year AUC = 0.724), especially for those with HBV infection (5-year AUC = 0.826). No obvious OS heterogeneity was detected among the HCV patients (*P* = 0.51), which could be resulted from the small sample size (*n* = 9).

No study had investigated the biological functions for the eight lncRNAs identified in HCC; however, they could affect the genesis as well as the progression of HCC by the Wnt signaling pathway[20-24], angiogenesis[25,26], p53 pathway[27,28], and PI3 kinase pathways[29-31], as suggested by the pathway enrichment results. More importantly, this research could provide new evidence that lncRNAs, whose biological functions

Table 2 Cox regression analysis for the overall survival-related clinicopathological features

Variable	Multivariate analysis		
	HR	95%CI	*P* value
Risk_score	1.944	1.613-2.343	2.94e-12
Age	1.004	0.986-1.023	0.662
Gender	0.832	0.480-1.442	0.512
Race	1.531	1.005-2.330	0.047
Grade	1.018	0.702-1.477	0.925
Stage	1.584	1.167-2.151	0.003

HR: Hazard ratio; CI: Confidence interval.

had not been reported before, might potentially serve as the potent markers to predict the prognosis of HCC. However, such findings should be further validated in future studies, and the molecular characteristics should also be investigated.

To our knowledge, this is the first study that constructs an lncRNA nomogram to predict the prognosis of VHCC. Nonetheless, some limitations should be noted in this study. First, no experimental study *in vitro* or *in vivo* was performed to verify the effect of the eight-lncRNA signature, which was constructed based on online datasets by the bioinformatic approach, in predicting the prognosis of HCCs. Second, data regarding more valuable clinical features are not available, including viral hepatitis duration, state of virus, Milan criteria, HCC treatment, and anti-viral treatment. Moreover, our results should be further validated in future studies.

To sum up, a novel lncRNA signature has been uncovered in this study based on comprehensive bioinformatic analysis combined with the expression patterns as well as the clinical information from the carefully selected cohort. Specifically, this risk model can be used as a biomarker to independently predict the prognosis of VHCC. Nonetheless, our results should be validated in studies that investigate the precise mechanisms regarding VHCC progression, as well as those that explore the functions of the eight identified lncRNAs.
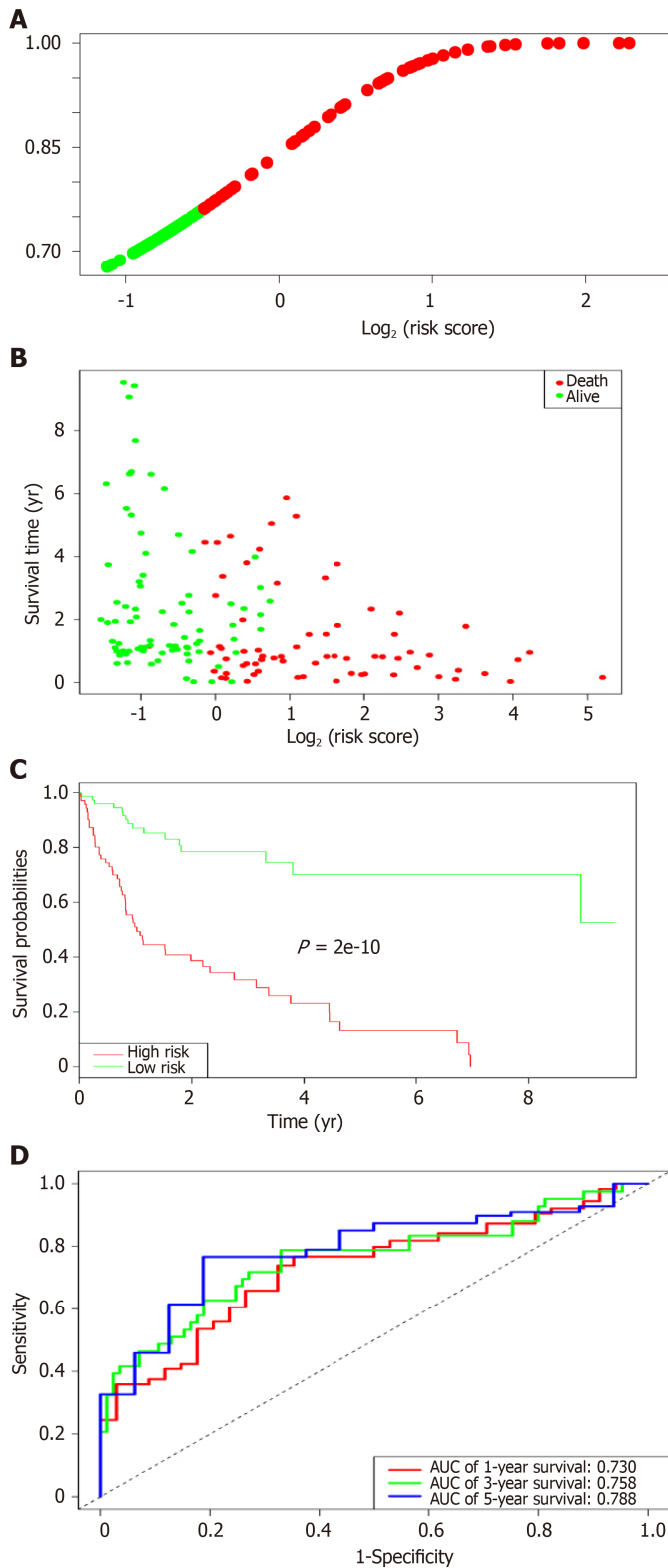
Figure 4 Establishment and verification of a long non-coding RNA signature for predicting the prognosis of hepatitis virus positive hepatocellular carcinoma using the cut-off value 0.798. A: Distribution of the long non-coding RNA (lncRNA) risk scores; B: Survival status as well as overall survival (OS); C: Kaplan-Meier curve regarding the OS of the high-risk as well as low-risk group classified based on median risk score; D: Receiver operating characteristic curves regarding the 1-, 3-, and 5-year survival discriminated by the lncRNA signature. AUC: Area under the curve.
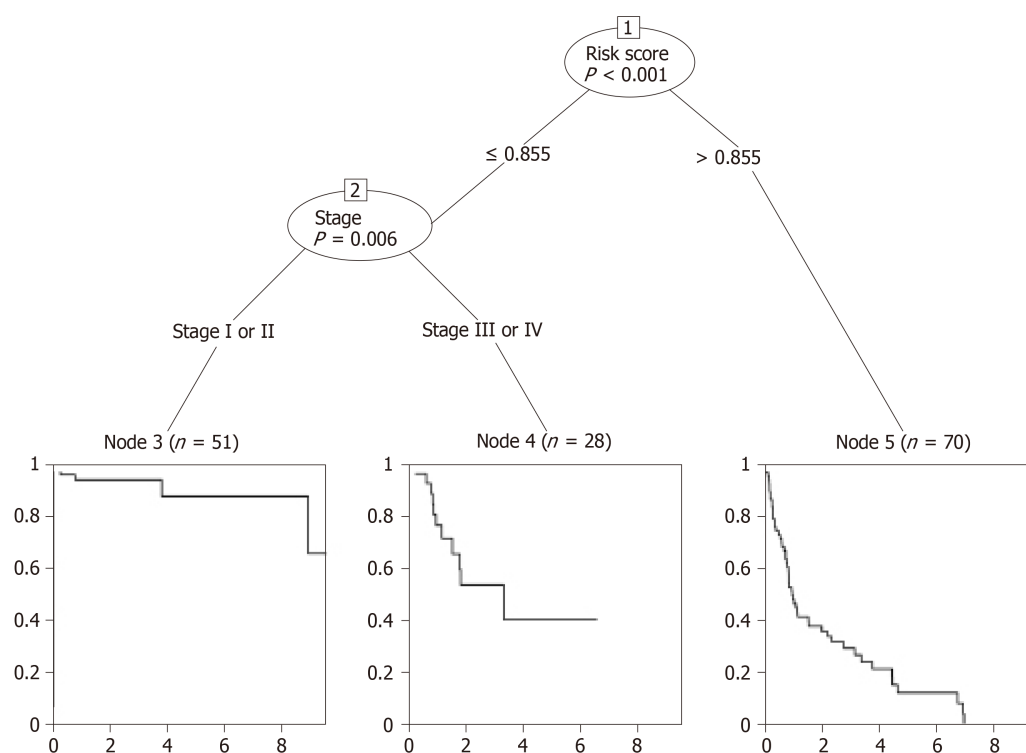
Figure 5 Conditional interference tree for prediction. The nodes of the conditional inference tree were calculated automatically based on the influence power of variables. The first node of the tree was risk_score, while the second was stage. Node 3, 4, or 5 was the corresponding group a certain patient would be assigned to. For example, a stage I patient with a risk_score ≤ 0.855 would be divided into group node 3, which was related to the best overall survival.
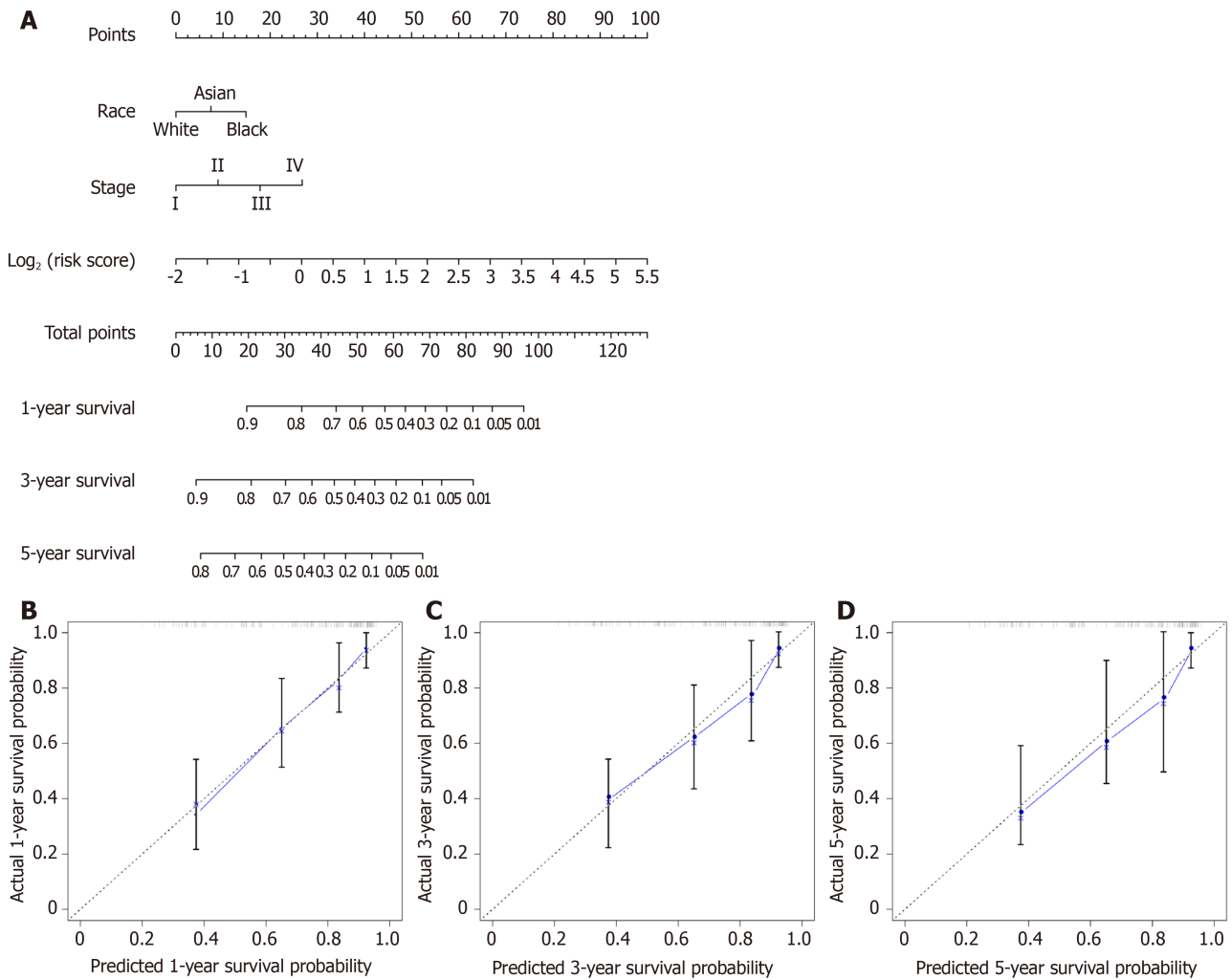
Figure 6 Overall survival nomogram predicting 1-, 3-, and 5-year survival of hepatocellular carcinoma, and calibrated curves to predict the 1-year, 3-years, and 5-year patient survival. A: Overall survival nomogram predicting 1-, 3-, and 5-year survival of hepatocellular carcinoma; B: Calibrated curve to predict the 1-year patient survival; C: Calibrated curve to predict the 3-years patient survival; D: Calibrated curve to predict the 5-years patient survival. In calibrated curves, the X-axis represents the survival prediction by nomogram, and the Y-axis represents the actual observation which showed a good agreement with prediction.

Figure 7 Subgroup analysis regarding the etiology. A-C: Survival curves of risk_score for hepatitis B virus + hepatitis C virus (HBV + HCV) (A), HBV alone (B), and HCV alone (C) infected patients with HCC; D-F: Receiver operating characteristic curves for HBV + HCV (D), HBV alone (E), and HCV alone (F) infected patients with HCC. HBV: Hepatitis B virus; HCV: Hepatitis C virus.
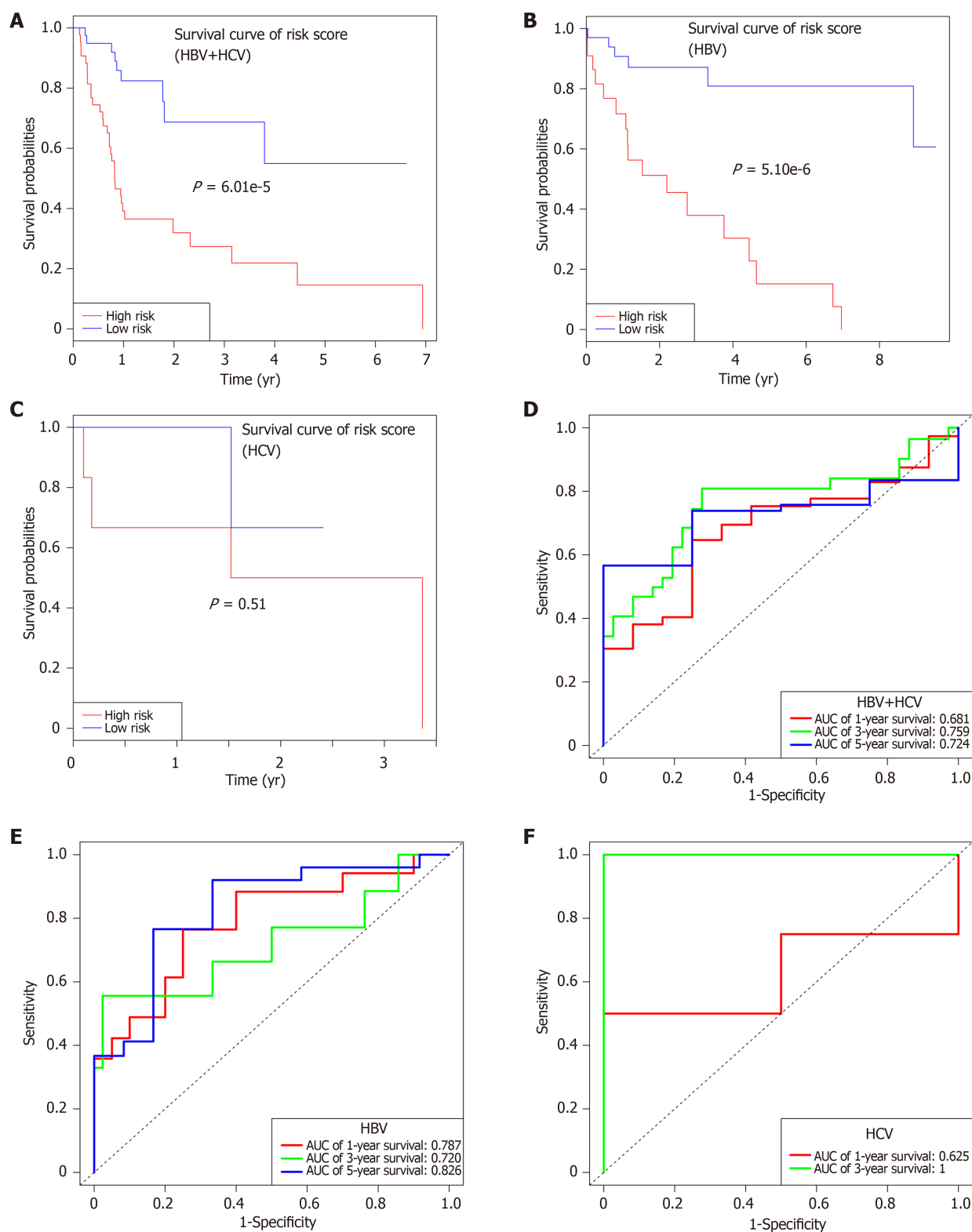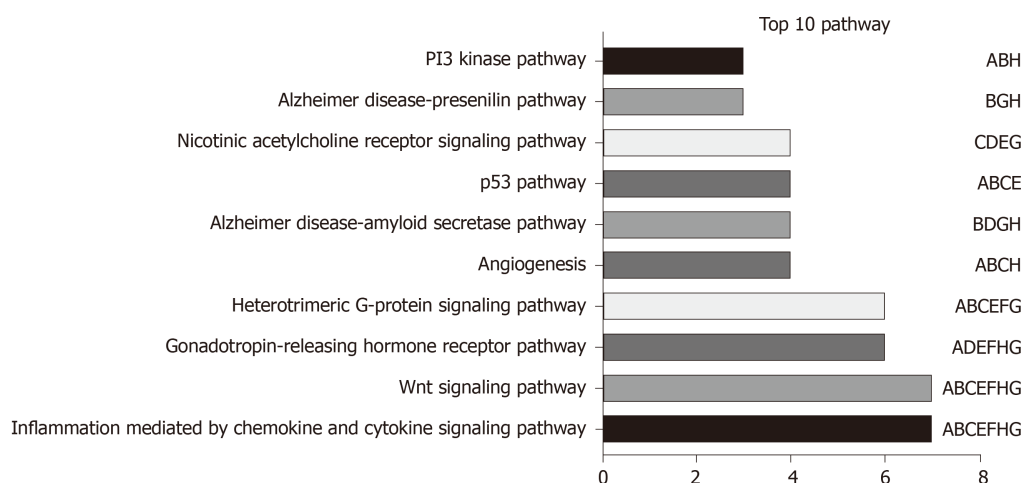
Figure 8 Functional annotations for the eight long non-coding RNAs. The 10 most remarkable pathways enriched based on the co-expressed mRNAs of eight long non-coding RNAs (lncRNAs). Corresponding lncRNAs involved in the deregulated pathways are also presented. A: AC005722.2; B: AC107959.3; C: AL353803.1; D: AL589182.1; E: AP000844.2; F: AP002478.1; G: FLJ36000; H: NPSR1-AS1.

ARTICLE HIGHLIGHTS

Research background
Hepatocellular carcinoma (HCC), one of the most frequent liver cancer subtype, has posed a serious health issue in the world. Hepatitis virus is recognized to be a major factor leading to HCC. Recently, a verity of genetic markers as well as prediction models are proposed to improve HCC treatment. At the same time, numerous statistical techniques are also utilized to mine data in numerous large-scale public databases. Thanks to the well-developed clinical approaches, prognosis models with a higher accuracy and robustness can be established for HCC.

Research motivation
This study aimed to establish a prognosis model on the basis of HCC molecular biomarkers. Typically, long non-coding RNAs (lncRNAs) have been recognized as new predictive factors. Great efforts have been made to establish lncRNA-based HCC models, however, the lncRNA features for hepatitis virus positive HCC (VHCC) are not available at present.

Research objectives
This study aimed to develop a prognostic lncRNA feature for VHCC based on candidate lncRNAs through analyzing data collected from The Cancer Genome Atlas (TCGA) database.

Research methods
Specifically, the least absolute shrinkage and selection operator (LASSO), the most advanced statistic algorithm, was used to establish the prediction model. This approach was carried out on the basis of typical lncRNAs selected according to the expression profiles of lncRNAs collected from the TCGA database. The as-established lncRNA feature was validated, and its clinical applicability was also examined.

Research results
Using LASSO, a risk score system was established to predict the overall survival (OS) for VHCC, which incorporated eight lncRNAs (including AC005722.2, AC107959.3, AL353803.1, AL589182.1, AP000844.2, AP002478.1, FLJ36000, and NPSR1-AS1). Notably, the as-constructed lncRNA feature was helpful in stratifying the risk for VHCC. To extend the model application range, a nomogram was also constructed, which involved both the lncRNA feature and other clinical features. Our results also suggested that the lncRNAs were markedly enriched in the Wnt signaling pathway, angiogenesis, the p53 pathway, and the PI3 kinase pathway.

Research conclusions
The as-constructed signature based on eight lncRNAs displayed a favorable capacity in predicting the prognosis for VHCC patients, which can also contribute to risk stratification and may provide more useful clinical advice for individual patients.

Research perspectives
Our results further support that lncRNAs can serve as possible functional regulating factors for the progression of VHCC. It is the future direction to search for the effective molecular biomarkers and predictive factors for the prognosis of HCC patients.

REFERENCES

1   **Chen QF**, Jia ZY, Yang ZQ, Fan WL, Shi HB. Transarterial Chemoembolization Monotherapy Versus Combined Transarterial Chemoembolization-Microwave Ablation Therapy for Hepatocellular Carcinoma Tumors ≤5 cm: A Propensity Analysis at a Single Center. *Cardiovasc Intervent Radiol* 2017; **40**: 1748-1755 [PMID: 28681222 DOI: 10.1007/s00270-017-1736-8]

2   **El-Serag HB**. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* 2012; **142**: 1264-1273.e1 [PMID: 22537432 DOI: 10.1053/j.gastro.2011.12.061]

3   **de Martel C**, Maucort-Boulch D, Plummer M, Franceschi S. World-wide relative contribution of hepatitis B and C viruses in hepatocellular carcinoma. *Hepatology* 2015; **62**: 1190-1200 [PMID: 26146815 DOI: 10.1002/hep.27969]

4   **Zhou HY**, Luo Y, Chen WD, Gong GZ. Hepatitis B virus mutation may play a role in hepatocellular carcinoma recurrence: A systematic review and meta-regression analysis. *J Gastroenterol Hepatol* 2015; **30**: 977-983 [PMID: 25689418 DOI: 10.1111/jgh.12917]

5   **Tang A**, Hallouch O, Chernyak V, Kamaya A, Sirlin CB. Epidemiology of hepatocellular carcinoma: target population for surveillance and diagnosis. *Abdom Radiol (NY)* 2018; **43**: 13-25 [PMID: 28647765 DOI: 10.1007/s00261-017-1209-1]

6   **Llovet JM**, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M, Gores G. Hepatocellular carcinoma. *Nat Rev Dis Primers* 2016; **2**: 16018 [PMID: 27158749 DOI: 10.1038/nrdp.2016.18]

7   **Yan X**, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD, Zhang Y, Yang L, Shan W, He Q, Fan L, Kandalaft LE, Tanyi JL, Li C, Yuan CX, Zhang D, Yuan H, Hua K, Lu Y, Katsaros D, Huang Q, Montone K, Fan Y, Coukos G, Boyd J, Sood AK, Rebbeck T, Mills GB, Dang CV, Zhang L. Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. *Cancer Cell* 2015; **28**: 529-540 [PMID: 26461095 DOI: 10.1016/j.ccell.2015.09.006]

8   **Ulitsky I**, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell* 2013; **154**: 26-46 [PMID: 23827673 DOI: 10.1016/j.cell.2013.06.020]

9   **Fatica A**, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* 2014; **15**: 7-21 [PMID: 24296535 DOI: 10.1038/nrg3606]

10 **Lanzafame M**, Bianco G, Terracciano LM, Ng CKY, Piscuoglio S. The Role of Long Non-Coding RNAs in Hepatocarcinogenesis. *Int J Mol Sci* 2018; **19**: pii: E682 [PMID: 29495592 DOI: 10.3390/ijms19030682]

11 **Gu JX**, Zhang X, Miao RC, Xiang XH, Fu YN, Zhang JY, Liu C, Qu K. Six-long non-coding RNA signature predicts recurrence-free survival in hepatocellular carcinoma. *World J Gastroenterol* 2019; **25**: 220-232 [PMID: 30670911 DOI: 10.3748/wjg.v25.i2.220]

12 **Wu Y**, Wang PS, Wang BG, Xu L, Fang WX, Che XF, Qu XJ, Liu YP, Li Z. Genomewide identification of a novel six-LncRNA signature to improve prognosis prediction in resectable hepatocellular carcinoma. *Cancer Med* 2018; **7**: 6219-6233 [PMID: 30378276 DOI: 10.1002/cam4.1854]

13 **Hosaka T**, Suzuki F, Kobayashi M, Seko Y, Kawamura Y, Sezaki H, Akuta N, Suzuki Y, Saitoh S, Arase Y, Ikeda K, Kobayashi M, Kumada H. Long-term entecavir treatment reduces hepatocellular carcinoma incidence in patients with hepatitis B virus infection. *Hepatology* 2013; **58**: 98-107 [PMID: 23213040 DOI: 10.1002/hep.26180]

14 **Ogawa E**, Furusyo N, Nomura H, Dohmen K, Higashi N, Takahashi K, Kawano A, Azuma K, Satoh T, Nakamuta M, Koyanagi T, Kato M, Shimoda S, Kajiwara E, Hayashi J; Kyushu University Liver Disease Study (KULDS) Group. Short-term risk of hepatocellular carcinoma after hepatitis C virus eradication following direct-acting anti-viral treatment. *Aliment Pharmacol Ther* 2018; **47**: 104-113 [PMID: 29035002 DOI: 10.1111/apt.14380]

15 **Liu GM**, Huang XY, Shen SL, Hu WJ, Peng BG. Adjuvant antiviral therapy for hepatitis B virus-related hepatocellular carcinoma after curative treatment: A systematic review and meta-analysis. *Hepatol Res* 2016; **46**: 100-110 [PMID: 26331530 DOI: 10.1111/hepr.12584]

16 **Shinn BJ**, Martin A, Coben RM, Conn MI, Prieto J, Kroop H, DiMarino AJ, Hann HW. Persistent risk for new, subsequent new and recurrent hepatocellular carcinoma despite successful anti-hepatitis B virus therapy and tumor ablation: The need for hepatitis B virus cure. *World J Hepatol* 2019; **11**: 65-73 [PMID: 30705719 DOI: 10.4254/wjh.v11.i1.65]

17 **Friedman J**, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010; **33**: 1-22 [PMID: 20808728 DOI: 10.1016/j.jspi.2009.07.020]

18 **Gao J**, Kwan PW, Shi D. Sparse kernel learning with LASSO and Bayesian inference algorithm. *Neural Netw* 2010; **23**: 257-264 [PMID: 19604671 DOI: 10.1016/j.neunet.2009.07.001]

19 **McNeish DM**. Using Lasso for Predictor Selection and to Assuage Overfitting: A Method Long Overlooked in Behavioral Sciences. *Multivariate Behav Res* 2015; **50**: 471-484 [PMID: 26610247 DOI: 10.1080/00273171.2015.1036965]

20 **Dai B**, Ma Y, Yang T, Fan M, Yu R, Su Q, Wang H, Liu F, Yang C, Zhang Y. Synergistic effect of berberine and HMQ1611 impairs cell proliferation and migration by regulating Wnt signaling pathway in hepatocellular carcinoma. *Phytother Res* 2019; **33**: 745-755 [PMID: 30565332 DOI: 10.1002/ptr.6267]

21 **Li B**, Cao Y, Meng G, Qian L, Xu T, Yan C, Luo O, Wang S, Wei J, Ding Y, Yu D. Targeting glutaminase 1 attenuates stemness properties in hepatocellular carcinoma by increasing reactive oxygen species and suppressing Wnt/beta-catenin pathway. *EBioMedicine* 2019; **39**: 239-254 [PMID: 30555042 DOI: 10.1016/j.ebiom.2018.11.063]

22 **Hu P**, Ke C, Guo X, Ren P, Tong Y, Luo S, He Y, Wei Z, Cheng B, Li R, Luo J, Meng Z. Both glypican-3/Wnt/β-catenin signaling pathway and autophagy contributed to the inhibitory effect of curcumin on hepatocellular carcinoma. *Dig Liver Dis* 2019; **51**: 120-126 [PMID: 30001951 DOI: 10.1016/j.dld.2018.06.012]

23 **Tan A**, Li Q, Chen L. CircZFR promotes hepatocellular carcinoma progression through regulating miR-3619-5p/CTNNB1 axis and activating Wnt/β-catenin pathway. *Arch Biochem Biophys* 2019; **661**: 196-202 [PMID: 30468709 DOI: 10.1016/j.abb.2018.11.020]

24 **Huynh H**, Ong R, Goh KY, Lee LY, Puehler F, Scholz A, Politz O, Mumberg D, Ziegelbauer K. Sorafenib/MEK inhibitor combination inhibits tumor growth and the Wnt/βcatenin pathway in xenograft models of hepatocellular carcinoma. *Int J Oncol* 2019; **54**: 1123-1133 [PMID: 30747223 DOI: 10.3892/ijo.2019.4693]

25 **Miura K**, Ohnishi H, Morimoto N, Minami S, Ishioka M, Watanabe S, Tsukui M, Takaoka Y, Nomoto H, Isoda N, Yamamoto H. Ezetimibe suppresses development of liver tumors by inhibiting angiogenesis in mice fed a high-fat diet. *Cancer Sci* 2019; **110**: 771-783 [PMID: 30520543 DOI: 10.1111/cas.13902]

26   **Morse MA**, Sun W, Kim R, He AR, Abada PB, Mynderse M, Finn RS. The Role of Angiogenesis in Hepatocellular Carcinoma. *Clin Cancer Res* 2019; **25**: 912-920 [PMID: 30274981 DOI: 10.1158/1078-0432.CCR-18-1254]

27   **Meng X**, Franklin DA, Dong J, Zhang Y. MDM2-p53 pathway in hepatocellular carcinoma. *Cancer Res* 2014; **74**: 7161-7167 [PMID: 25477334 DOI: 10.1158/0008-5472.CAN-14-1446]

28   **Kong Y**, Zhang L, Huang Y, He T, Zhang L, Zhao X, Zhou X, Zhou D, Yan Y, Zhou J, Xie H, Zhou L, Zheng S, Wang W. Pseudogene PDIA3P1 promotes cell proliferation, migration and invasion, and suppresses apoptosis in hepatocellular carcinoma by regulating the p53 pathway. *Cancer Lett* 2017; **407**: 76-83 [PMID: 28823960 DOI: 10.1016/j.canlet.2017.07.031]

29   **Jiao M**, Nan KJ. Activation of PI3 kinase/Akt/HIF-1α pathway contributes to hypoxia-induced epithelial-mesenchymal transition and chemoresistance in hepatocellular carcinoma. *Int J Oncol* 2012; **40**: 461-468 [PMID: 21922131 DOI: 10.3892/ijo.2011.1197]

30   **Liu W**, Guo TF, Jing ZT, Yang Z, Liu L, Yang YP, Lin X, Tong QY. Hepatitis B virus core protein promotes hepatocarcinogenesis by enhancing Src expression and activating the Src/PI3K/Akt pathway. *FASEB J* 2018; **32**: 3033-3046 [PMID: 29401603 DOI: 10.1096/fj.201701144R]

31   **Zhang Y**, Zhang T, Wu C, Xia Q, Xu D. ASIC1a mediates the drug resistance of human hepatocellular carcinoma via the Ca2+/PI3-kinase/AKT signaling pathway. *Lab Invest* 2017; **97**: 53-69 [PMID: 27918554 DOI: 10.1038/labinvest.2016.127]