Cancer Clinics Journal

ISSN: 2689-9566

Research Article

To Identify Key HCC-Related Genes and Prognostic Models of HBV Infection Based on Protein Interaction Network

*Qingxiu Li and Changzhu Duan**

Chongqing Medical University, China

Abstract

The occurrence of Hepatitis B Virus-induced Hepatocellular Carcinoma (HBV-HCC) is a complex process of multi-gene and multi-step interaction, in which the synergistic effect of various cancer-promoting mechanisms accelerates the disease evolution from inflammation to tumorigenesis. In this study, first, genes encoding human proteins associated with HBV-HCC were screened from the MEDLINE literature. Based on the Protein-Protein Interaction network (PPI) in the database, CluterONE algorithm was used to infer the HBV-HCC-related protein-protein interaction network and identify some Potential HBV-HCC-related Genes (PHHG). Second, the effectiveness of PHHG in classifying cancerous and normal tissues was verified in a public database. Finally, univariate and multivariate Cox regression analyses were performed to identify the best Prognostic Genes (PRGS) for PHHG. CYP2C19 (HR 1.29016, P=0.01894), FLNC (HR 1.2609, P=0.00839), HNRNPC (HR 3.17362, P=0.03754) and UBE3A (HR 0.28233, P=0.03754) were selected. P=0.01461) to establish a prognostic risk score model and validate it in the TCGA cohort. Kaplan-Meier survival analysis showed that patients with high-risk score had significantly worse overall survival (OS, log-scale P=9.04 × 10-7). Finally, a prognostic nomogram including PRG, age, gender, PTNM stage and Grade is constructed. The HBV-HCC related four-gene risk score model may be useful as a prognostic biomarker for HBV-HCC patients, providing targets for their treatment and contributing to individualized survival predictions, which could be of great importance for the study of treatment strategies.

Keywords: Hepatitis B virus; Hepatocellular carcinoma; Protein interaction network; Key genes; Prognostic model

Introduction

Hepatocellular Carcinoma (HCC) is one of the most common malignancies, with the fifth highest incidence and third highest mortality worldwide [1]. HCC is the most influential primary liver cancer, mainly caused by Hepatitis B Virus (HBV), and 50% of HCC patients worldwide are infected by HBV [2]. However, the role of HBV in the development and progression of liver disease remains unclear. While existing studies have shown that the clinical therapeutic targets for HBV-HCC are related to gene expression, autophagy, exosomes, gut microbiota, epigenetic dysregulation and immune mechanisms [2,3]. At present, some completely proven HCC-related driver genes, such as TP53 (R249S) codon 249 third nucleotide is frequently replaced in HBV-HCC patients [4], TERT promoter and CTNNB1 mutations have been observed as the most common somatic genetic changes [5]. The de-regulation of the Met gene suppresses the growth of human HCC cells, and the de-regulation of the *Axl gene* suppresses the metastatic nature of hepatocellular carcinoma through the PI3K/AKT-PAK1 signaling pathway [6]. These genes are of clinical interest as biomarkers in drug clinical trials and pathogenesis studies. But these driver genes are responsible for only a few populations, specific experimental cell lines or animal models. However, the genes and genetic markers associated with HBV-HCC pathogenesis are far from being explored. The occurrence of hepatocellular carcinoma caused by Hepatitis B Virus (HBV-HCC) is a complex process with multi-gene and multi-step interaction. The synergistic effect of various cancer-promoting mechanisms accelerates the disease evolution process from inflammation to tumorigenesis, which is influenced by environmental and genetic factors. Genes with different biological functions can jointly promote the tumorigenesis of HBV-HCC, and each gene plays a moderate or tiny role.

Thus, a comprehensive analysis of potentially relevant genes within the framework of related pathways and/or protein interaction networks may provide numerous crucial insights beyond traditional single-gene analyses. Zhang G et al. [7] identified the genes *PON1*, *AGR2*, *SSR2* and *TMCC1* and constructed a novel prognostic risk model for four ER stress-related genes that accurately predicted survival outcomes in HBV-HCC patients. Hu et al. [8] found that increased PRKRA expression may predict a poorer prognosis for HBV-associated hepatocellular carcinoma and found that a triple combination of blood PRKRA expression, serum AFP and CEA levels may be a non-invasive diagnostic strategy. Data from Zhang C et al. [9] suggest that the C1653T mutation in HBX promotes HCC malignancy by altering the levels of fibrosis, ROS, and some cytokine levels. This mutation could be used as a potential biomarker for screening patients for HCC to determine the efficacy of treatment. Current research focuses on targeted genes, and lacks a network framework to guide and comprehensively explore the underlying genes. However, the expression patterns of multiple genes can serve as excellent molecular biomarkers, allowing for early diagnosis, subgroup classification, risk stratification, prognostic prediction and therapeutic targeting of HBV-HCC.

Citation: Li Q, Duan C. To Identify Key HCC-Related Genes and Prognostic Models of HBV Infection Based on Protein Interaction Network. Cancer Clin J. 2022;3(1):1014.

Publisher Name: Medtext Publications LLC

Manuscript compiled: Nov 22th, 2022

***Corresponding author:** Changzhu Duan, Chongqing Medical University, Yuzhong Medical College Road, Chongqing, China, Tel: +86-023-68485804, E-mail: duanchzhu@cqmu.edu.cn

In this study, we comprehensively collected HBV-HCC related genes from the genetic literature, and constructed PPI to dig out modules corresponding to potential genes. The set of key genes that we identified is of great value for the diagnosis and treatment of HBV-HCC. Furthermore, we combine key genetic features and clinical parameters to develop a novel and promising prognostic nomogram model with higher accurate predictive power than clinical risk factors for HBV-HCC patients.

Materials and Methods

Data and sample acquisition

The validation set data for the identified key genes was downloaded from TCGA database *via* R software package [10], consisting of 145 HBV-associated hepatocellular carcinoma patients and 50 normal samples. All patients without prognostic information were initially excluded. In addition, microarray expression profiles GSE113996 and GSE94660 were downloaded from the GEO database of HBV-associated hepatocellular carcinoma patients. The probes were converted to gene names based on the annotation files provided by the manufacturer, and duplicate probes for the same gene were removed. The value of the median expression is retained for all repeated detections. Since the data were obtained from TCGA and GEO, no ethical committee approval was required for our study.

Identification of related genes

Candidate genes related to HBV-HCC were searched in the PUBMED literature. Epigenetic changes due to the molecular mechanisms of genes play an essential role in the development of HBV-HCC. Therefore, we searched the literature related to HBV-HCC, with terms (HBV-HCC [MeSH]) and (polymorphism [MeSH] or genotype [MeSH] or allele [MeSH]) rather than (tumor [MeSH]). To reduce the number of false positive results, studies in the literature with negative or no significant association were not included. Full reports from selected publications were reviewed to ensure that conclusions were supported by content. As a result, the genes we screened showed significant genetic association with HBV-HCC on the whole genome.

Functional analysis of related genes

The identified genes were subjected to Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis to mine the functional categories of genes with significant HBV-HCC correlation in the database [11,12]. Its statistical principle is to use hypergeometric distribution type to test the significance of a certain functional class in a group of genes (co-expressed or differentially expressed).

Functional classes of genes with significant relevance for experimental purposes, low false positive rates, and targets are obtained using discrete distribution, enrichment analysis, and significance analysis of false positives. The "org.hs.eg.db" package in R language is used to convert HBV-HCC related gene symbols into *Entrez gene* identifiers. The "ClusterProfiler" package is used for gene and gene cluster analysis and gene profile function visualization [10]. The "ClusterProfiler" package provides a gene classification method (groupGO) to classify genes based on specific levels of projection in the GO corpus and "enrichGO" and "enrichKEGG" analyses based on hypergeometric distributions to compute enrichment tests for GO terms and KEGG pathways.

Pathways with p-value Cutoff and q-value Cutoff less than or equal to 0.05 are considered to be significantly enriched. Finally, the "ggplot2" package was combined to draw GO analysis dot plots and KEGG analysis bar charts.

PPI construction of related genes

HIPPIE (Human Integrated Protein-Protein Interaction rEference) is a tool library for Human protein-protein interaction [13]. We downloaded all protein interaction pairs from HIPPIE (V2.2). To construct the most comprehensive library of human protein interaction relations to date, we downloaded protein interaction pairs with scores greater than 900 from the STRING database [14]. We then merge the two interaction group databases by excluding self-interactions and redundant pairs, and use the Stringi package in the R language to transfer protein identifiers to gene symbols. Finally, we extracted genes associated with HBV-HCC from PPI.

Mining and verification of PHHG

The HBV-HCC related PPI subnetwork is imported into Cytoscape platform [15]. CluterONE is an algorithm for detecting potentially overlapping protein complexes in protein-protein interaction networks. It can detect high-density regions containing extra connections and possibly overlapping protein complex models in protein interaction networks. It can detect high-density regions in protein interaction networks that contain additional connections and possibly overlapping protein complex models [16]. Using the CytoCluster plugin, the ClusterONE algorithm was selected, and the minimum number larger than 25 was set as the screening criterion. The minimum density is set to Auto. The higher the centrality, the closer the relationship between the genes, suggesting that each gene also interacts frequently and therefore that these genes are more likely to be involved in the occurrence of HBV-HCC. Finally, we select the gene with the highest PPI network centrality for each protein complex model as PHHG. GO and KEGG analyses were performed using the BiNGO and ClueGO plugins in the Cytoscape software. We performed a differential analysis on the transcriptome data of PHHG in the TCGA database and re-validated it using the related data from the external GEO database.

Construction and evaluation of prognostic risk score model

Firstly, based on the transcriptome data of latent genes of HBV-HCC patients in TCGA database, Kaplan-Meier univariate survival analysis was performed with the survival package in R software to explore the relationship between Overall Survival (OS) and latent genes. After univariate Cox regression analysis, all independent prognostic factors were screened by multivariate Cox regression analysis to construct a prognostic nomogram.

In this study, only patients who completed the follow-up period were selected for survival analysis and then divided into two groups based on the median expression value of the latent gene. PHHG associated with prognosis at $P<0.05$ are considered statistically significant. Subsequently, a prognostic risk score was generated for each patient. All TCGA HBV-HCC patients were divided into high-risk (high-risk score) and low-risk (low-risk score) groups according to their median risk score. Then, a K-M survival curve is constructed to estimate the prognosis of patients with high or low risk scores, and a two-sided log-rank test was used to assess the difference in survival between high and low risk groups. The following formula was used to establish the prognostic risk score model: risk score=expression level of gene 1 × β1+expression level of gene 2 × β2+... + expression

level of gene n × βn; where β is the regression coefficient calculated by the multivariate Cox regression model. Prognostic performance was assessed at 0.5, 1- and 3-years using Harrell's Concordance index (C-index) and transient Receiver Operating Characteristic (ROC) curve scores to assess the predictive accuracy of the prognostic risk score model based on HBV-HCC. R package "survcomp" and "survival". C index and Area Under ROC Curve (AUC) values range from 0.5 to 1, where 1 indicates complete discrimination and 0.5 indicates no discrimination.

Results

Identification of HBV-HCC related genes

By searching PUBMED, we screened the publications on genetic association studies related to HBV-HCC. In this process, a gene is collected whenever it is reported to be significantly associated with HBV-HCC, regardless of whether the gene is directly or indirectly associated with the disease. A detailed list of all genes reported to be associated with HBV-HCC. A total of 1896 HBV-HCC related genes were identified, and these HBV-HCC related genes showed large changes in function, indicating that HBV-HCC tumorigenesis and progression is an extremely complex process.

Biological function enrichment and biochemical pathway of HBV-HCC related genes

We performed functional analysis and identified 1896 HBV-HCC related genes that were significantly enriched in 4251 GO terms. Figure 1A show the top 12 GO items with the lowest P-value. The Biological Process (BP) related to the pathogenesis of HBV-HCC is about the regulation of cytokines and their mediated signaling pathways, and their response to virus invasion. As a non-cellular virus, HBV is widely believed to cause chronic liver injury by attacking the immune system. Therefore, the use of cytokines in the treatment of HBV-HCC has an extremely broad application prospect [17]. Cytoplasmic membranes, RNA polymerase II transcription factors and ubiquitin proteases have been found to be involved in HBV-HCC inhibition [18,19]. In addition, we identified 176 pathways enriched in HBV-

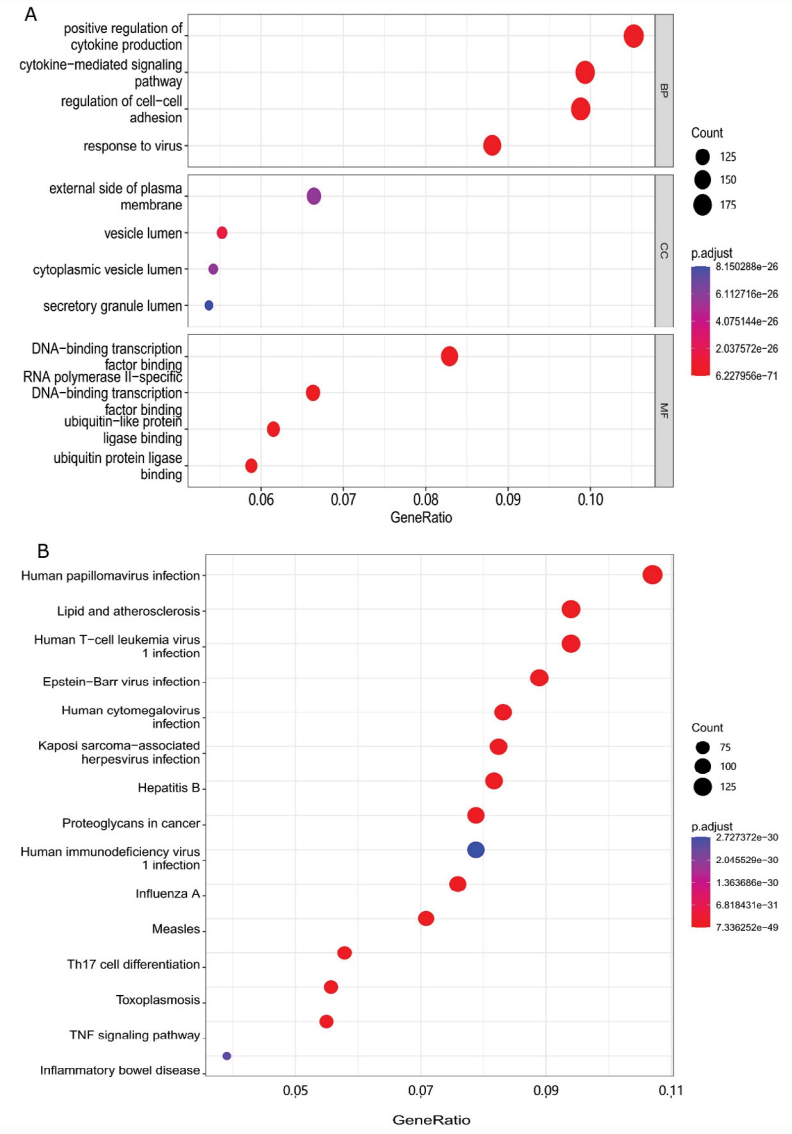

Figure 1: GO and KEGG enrichment analysis of reported HBV-HCC related genes.

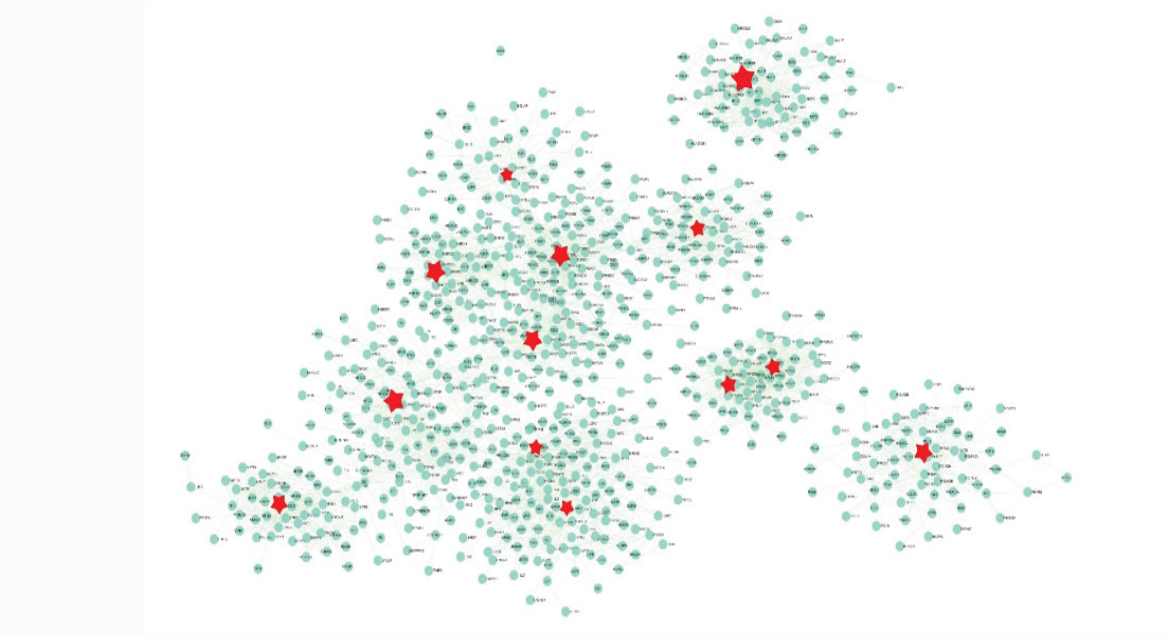

Figure 2: HBV-HCC related gene protein interaction network.

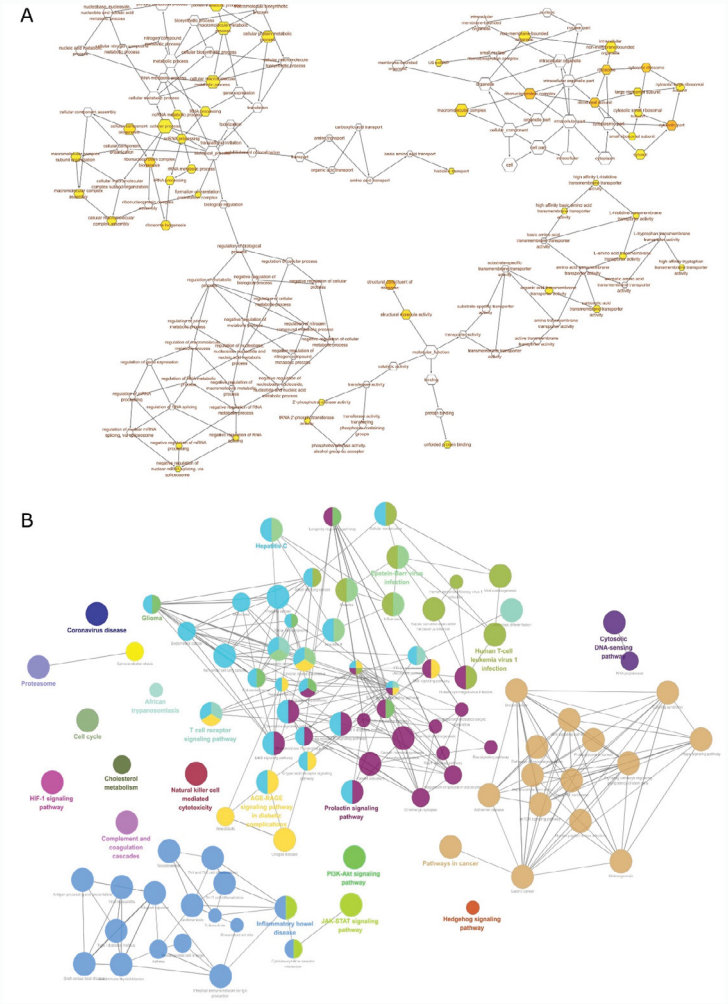

Figure 3: Protein interaction networks were analyzed by GO and KEGG enrichment of HBV-HCC related genes.

HCC related genes. As shown in Figure 1B, the first 15 items with the lowest P-values in KEGG are consistent with previous studies, indicating that TNF signaling pathway and lipid and atherosclerosis pathway are related to HCC signaling pathway [20-23]. There are also several pathways involved in the physiological processes of cells, such as apoptosis, necrosis, adhesion junctions and the cell cycle, which are commonly thought to be associated with tumor cell proliferation [24,25]. Finally, we found that HBV-HCC related genes were enriched in additional viral infections, including Human cytomegalovirus infection, Kaposi sarcoma-associated herpesvirus infection, and Human immunodeficiency virus infection. Thus, it is conjectured that the relevant genes for hepatocellular carcinoma caused by HBV infection are also involved in the pathogenic molecular mechanisms of different diseases, thus further bioinformatics analysis is highly necessary.

Acquisition and validation of PPHG

We have 390,000 protein-protein interaction pairs from the HIPPIE database (last updated April 29, 2022). Meanwhile, human protein interaction pairs with scores greater than 900 were selected from the STRING database. The union of protein pairs in the two databases was used to screen out the PPI of HBV-HCC-related genes, and finally a relatively comprehensive human protein interaction network (HHPPI,) was obtained, consisting of 5043 proteins and 75,170 interaction pairs. The HHPPI network was imported into the Cytoscape software. The mining module is implemented using the CytoCluster plugin. Set the parameters of the module to the minimum gene number threshold of 25, and keep all other parameters as default values. Finally, we selected 13 modules, and we can see in (Figure 2) the PPI consists of 13 modules (red five-pointed stars), each of which contains at least 25 genes. We conducted GO and KEGG functional analysis of the proteins of the 13 modules (Figure 3A and B), and found that the biological functions of each module were correlated with each other. Functional analysis identified functional modules associated with tumorigenesis, such as metabolism of biomolecules, ribosomal subunit binding, high affinity L-histidine transmembrane transporter activity, and negative regulation of mRNA processing. The PI3K-Akt signaling pathway, cytosolic DNA sensing pathway, T-cell receptor signaling pathway and HIF-1 signaling pathway are the main enrichment pathways. In each module, we selected genes with the highest "center", which were seen as potential genes for HBV-HCC-related PPI networks. After deleting the duplicated genes, we obtain 33 PPHGS. We conducted Wilcoxon test on the RNAseq data of PPHG in the TCGA and GSE113996 datasets and found that there were 24 genes with significantly different expressions in normal tissues and cancer tissues, including ADK, BAG3, C2, CDC25B, CLEC12A, CSNK2B, CYP2C19, FLNC, etc. GCG, HNRNPC, ILF3, JMJD6, KAT5, MAPK6, MTR, NXF1, POLR2L, PRKAB2, RAB18, RPGRIP1L, SLC25A3, SRPK2, UBE3A and YWHAH.

Identification of PHHG associated with prognosis

Based on the RNA-Seq data of 145 HBV-HCC patients with PHHG in the TCGA database, we identified nine prognostic genes after univariate Cox regression analysis of 24 candidate genes ($P<0.05$). Subsequent multivariate Cox regression analysis showed that only four genes had significant prognostic value for HBV-HCC patients, which were cytochrome P450 family 2 subfamily C member 19 (CYP2C19, HR 1.29016, P=0.01894), filamin C (FLNC, HR 1.2609, P=0.00839) and heterogeneous nuclear ribonucleoprotein C (HNRNPC, HR 3.17362, P=0.03754) and ubiquitin protein ligase E3A (UBE3A, HR 0.28233, P=0.01461). Then, the differential expression of the above four genes in tumor and normal tissue was further verified in the GSE94660 chip, which consisted of 20 HBV-HCC samples and 20 normal samples. CYP2C19 was significantly down-regulated, but FLNC, HNRNPC and UBE3A were all significantly up-regulated in HBV-HCC tissues (Figure 4A-D). We analyzed the expression of proteins encoded by four genes using clinical specimens of HBV-HCC derived from human protein profiles [26]. It was found that the expression levels of HNRNPC and UBE3A in HBV-HCC tissues were higher than those in normal tissues (Figure 4G and H). Furthermore, the K-M survival curve was constructed to evaluate the association between the expression levels of four prognostic genes and OS, and the results showed that the low expression group of HNRNPC (Log-rank P=0.000214) had a better prognosis (Figure 4K).

Establishment and evaluation of prognostic risk score model

The prognostic risk score model was established by using four genes related to prognosis: risk score=CYP2C19 expression level × 0.2807+FLNC expression level × 0.1372+HNRNPC expression level × 1.0476+UBE3A expression level × 0.5522. Subsequently, we calculated the prognostic risk score for each patient in the TCGA training set. All patients were divided into high-risk (high-risk score) or low-risk (low-risk score) groups, with the median risk score as the cut-off value (Figure 5A). In addition, K-M survival analysis (Figure 5B) showed that patients with high-risk scores had significantly worse OS than patients with low-risk scores (log-rank P=9.04e-07). This model showed decent predictive ability for OS rate at 0.5, 1 and 3 years, with AUC values of 0.757, 0.765 and 0.789, respectively (Figure 5C). In addition, we analyzed the correlation between the prognostic model's risk score (the higher the score, the greater the risk) and various immune cells. As can be seen from Figure 6, Macrophage expression is negatively and significantly correlated with P value of 5.81e-07. Not attractive, but positively correlated with the expression of uncharacterized cells, with P value of 1.5e-05.

Establishment of nomogram model of TCGA

To investigate the coefficient prediction efficiency of this feature, a nomogram model is constructed in the TCGA dataset. The results show that the C-index of 0.783 nomogram provides a quantitative method for accurately predicting 1, 3- and 5-year survival rates (Figure 6A). The overlap between the predicted and actual probabilities of 1, 3- and 5-year survival rates in the calibration curves indicates good agreement (Figure 6B). Meanwhile, Spearman correlation analysis was performed between this model score and the immunization score. According to the results of timer algorithm (Figure 7), the proportion of macrophages in HBV-HCC patients in the low-risk score subgroup moderately increased.

Discussion

With the application and development of genomics, transcriptomics, metabolomics and proteomics in tissue and body fluid samples, a growing number of genes that may be involved in HBV-HCC have been identified [26].

However, the identification and analysis of the underlying genes involved in the biochemical processes associated with HBV-HCC is not yet complete. Therefore, a comprehensive analysis and exploration of HBV-HCC tumorigenesis genes at the level of systematics is increasingly urgent [27]. In this study, first, a fuzzy picture of the molecular mechanism of HBV-HCC tumorigenesis is derived by

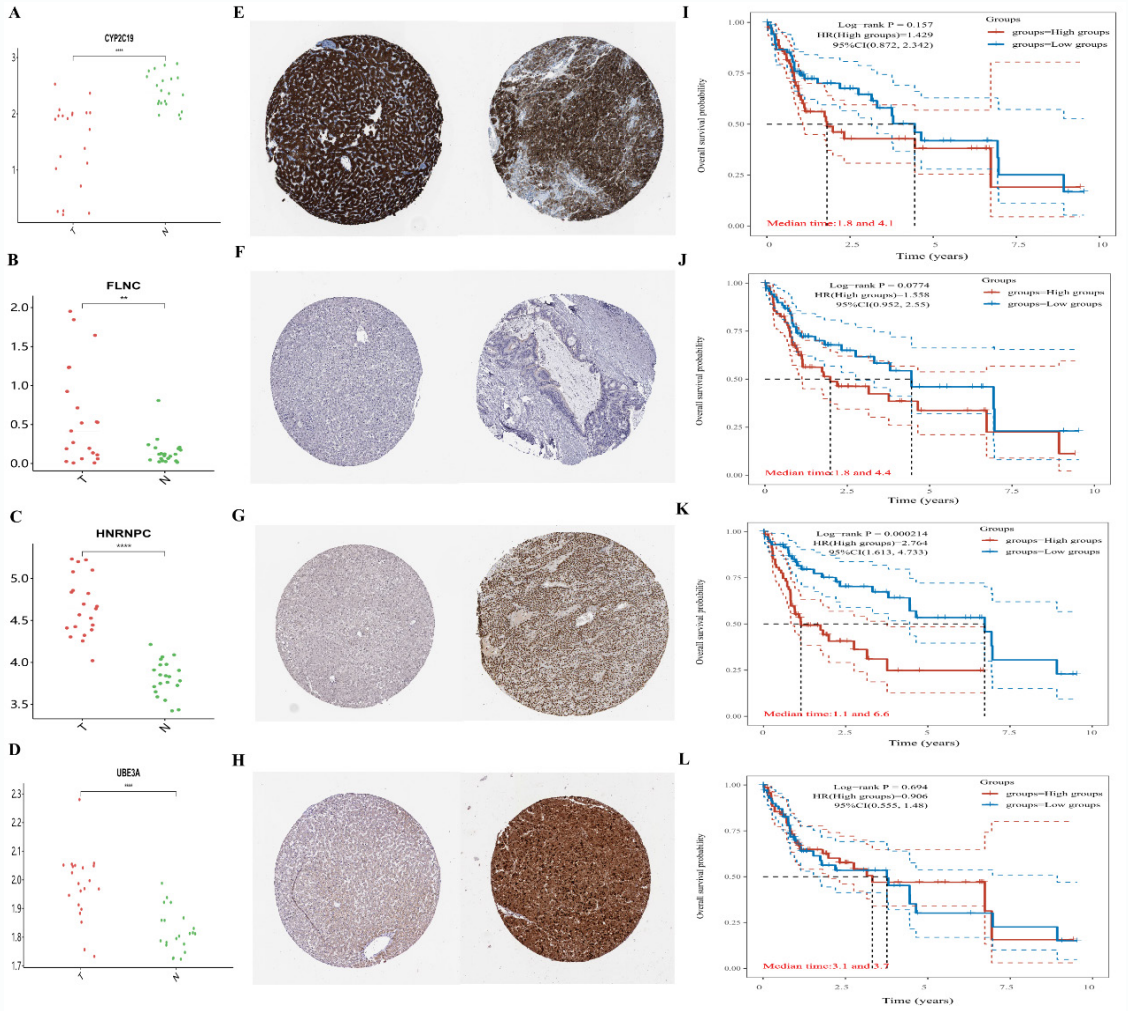

Figure 4: Expression and survival analysis for CYP2C19, FLNC, HNRNPC and UBE3A.

analyzing the reported HBV-HCC related gene sources, performing functional enrichment and pathway crosstalk. Second, 24 potential genes were screened from the PPI network associated with HBV-HCC. Then, two Wilcoxon tests were performed on the 24 potential genes in the TCGA and GEO databases.

According to the study, a number of PPHG have been identified by a small number of people as being involved in the occurrence and development of HCC. For example, data from Yan showed that inhibition of CDC25B protein expression suppresses tumor cell growth and motility in HCC patients [28]. MiR-660-5p interacts with YWHAH through the PI3K/Akt signaling pathway to promote the progression of hepatocellular carcinoma [29].

Tumor necrosis factor-α-induced protein-1 blocks nuclear factor-κB activation in hepatocellular carcinoma by selectively targeting CSNK2B [30]. BAG3 protein is a helper of heat shock protein 70, which regulates the physiological and pathological processes in HCC [31]. The expressions of FLNC, KAT5, HNRNPC RAB18 and ILF3 were significantly increased in HCC tissues [32-37]. RPGRIP1L is a candidate tumor suppressor gene for human hepatocellular carcinoma [38]. There are also some PPHG involved in the occurrence and development of HBV-HCC, and the up-regulation of microRNA499a by targeting MAPK6 can induce the carcinogenesis in hepatitis B virus-associated hepatocellular carcinoma [39]. Tang et al. [40] suggests that aberrant CAR methylation is involved in CYP2C19 regulation in HBV-associated HCC and may play a role in liver tumorigenesis.

According to studies, laboratory diagnostic indicators such as abnormal prothrombin, AFP and AFP alloplastic ratio of HBV-HCC patients are helpful to predict their prognosis to a certain extent [41,42]. However, due to the huge heterogeneity of HBV-HCC disease, it is still necessary to identify new prognostic biomarkers and establish more accurate prognostic models. The combination of prognostic genetic features and traditional clinical parameters may provide better prediction than individual biomarkers. Therefore, in this study, we established a novel four-gene signature (including CYP2C19, FLNC, HNRNPC, and UBE3A) for prognostic prediction of HBV-HCC on the basis of screening 24 potential genes. The prognostic prediction performance was good in the TCGA HBV-HCC cohort, and the survival rate of patients in the high-risk group was significantly lower than that of patients in the low-risk group. The nomogram of the four-gene risk profile combined with conventional clinical parameters such as gender, age, TNM stage and grade also showed good predictability. However, this four-gene risk prediction model also has some limitations. Since our study is mainly based on TCGA data, it is not universal. Efforts should be made to further improve the prediction performance, and further functional experiments are

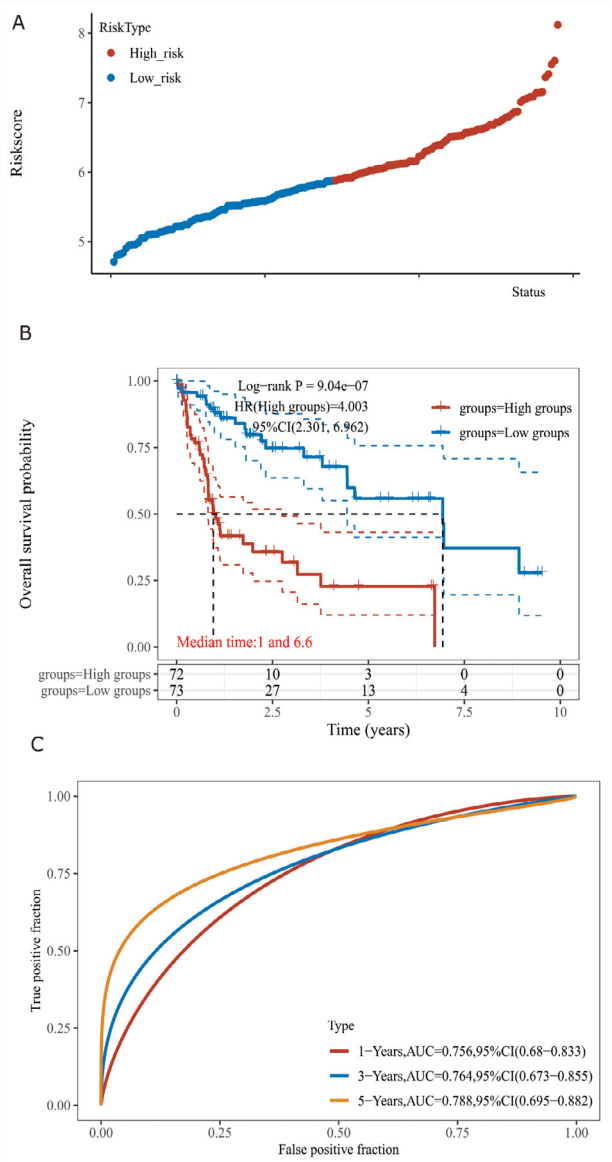

Figure 5: Establishment and evaluation of prognostic risk score model.

needed to elucidate the underlying mechanisms of the four genes.

Conclusion

In this study, we investigated the protein interaction network based on HBV-HCC-related genes through a systems biology framework, constructed HCC pathological PPI network, mined 24 potential HCC genes, and established a novel four-gene signature and nomogram to predict overall survival of HBV-HCC. This comprehensive analysis of the genes involved in HBV-associated hepatocellular carcinoma enhances our understanding of genetic factors and their interaction with pathogenesis, and improves the possibility of identifying potential targets for the diagnosis and treatment of HCC. At the same time, the risk model we constructed in this study can effectively classify HBV-HCC patients and contribute to the clinical decision of individualized immunotherapy.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209-49.
2. Akrami H, Monjezi MR, Ilbeigi S, Amiri F, Fattahi MR. The association between hepatitis b virus mutations and the risk of liver disease and hepatocellular carcinoma. Curr Mol Med. 2022;22(6):514-23.
3. Lee WY, Bachtiar M, Choo CCS, Lee CG. Comprehensive review of Hepatitis B Virus-associated hepatocellular carcinoma research through text mining and big data analytics. Biol Rev Camb Philos Soc. 2019;94(2):353-67.
4. Turner PC, Sylla A, Kuang SY, Marchant CL, Diallo MS, Hall AJ, et al. Absence of TP53 codon 249 mutations in young Guinean children with high aflatoxin exposure. Cancer Epidemiol Biomarkers Prev. 2005;14(8):2053-5.
5. Lee SE, Chang SH, Kim WY, Lim SD, Kim WS, Hwang TS, et al. Frequent somatic TERT promoter mutations and CTNNB1 mutations in hepatocellular carcinoma. Oncotarget. 2016;7(43):69267-75.
6. Schoedel KE, Tyner VZ, Kim TH, Michalopoulos GK, Mars WM. HGF, MET, and matrix-related proteases in hepatocellular carcinoma, fibrolamellar variant, cirrhotic and normal liver. Mod Pathol. 2003;16(1):14-21.
7. Zhang G, Sun J. Endoplasmic reticulum stress-related signature for predicting prognosis and immune features in hepatocellular carcinoma. J Immunol Res. 2022;1366508.
8. Hu YM, Ran R, Yang C, Liu SM. The diagnostic and prognostic implications of PRKRA expression in HBV-related hepatocellular carcinoma. Infect Agent Cancer. 2022;17(1):34.
9. Zhang C, Xie Y, Lai R, Wu J, Guo Z. Nonsynonymous C1653T mutation of hepatitis b virus x gene enhances malignancy of hepatocellular carcinoma cells. J Hepatocell Carcinoma. 2022;9:367-77.
10. Yu G, Wang LG, Han Y, He QY. Cluster profiler: an r package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7.
11. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.
12. Consortium GO. Gene ontology consortium: going forward. Nucleic Acids Res. 2015;43:D1049-56.
13. Lobato AG, Navarro MAA, Schaefer MH. HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. Nucleic Acids Res. 2017;45(D1):D408-14.
14. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Cepas JH, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43:D447-52.
15. Li M, Li D, Tang Y, Wu F, Wang J. Cytocluster: a cytoscape plugin for cluster analysis and visualization of biological networks. Int J Mol Sci. 2017;18(9):1880.
16. Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods. 2012;9(5):471-2.
17. Chen Y, Tian Z. HBV-induced immune imbalance in the development of hcc. Front Immunol. 2019;10:2048.
18. Wan T, Lei Z, Tu B, Wang T, Wang J, Huang F. NEDD4 induces k48-linked degradative ubiquitination of hepatitis b virus x protein and inhibits hbv-associated hcc progression. Front Oncol. 2021;11:625169.
19. Lin X, Li AM, Li YH, Luo RC, Zou YJ, Liu YY, et al. Silencing MYH9 blocks HBx-induced GSK3β ubiquitination and degradation to inhibit tumor stemness in hepatocellular carcinoma. Signal Transduct Target Ther. 2020;5:13.
20. Nakagawa H, Umemura A, Taniguchi K, Burgada JF, Dhar D, Ogata H, et al. ER stress cooperates with hypernutrition to trigger TNF-dependent spontaneous HCC development. Cancer Cell. 2014;26(3):331-43.
21. Yang Y, Qu Y, Lv X, Zhao R, Yu J, Hu S, et al. Sesamol supplementation alleviates nonalcoholic steatohepatitis and atherosclerosis in high-fat, high carbohydrate and

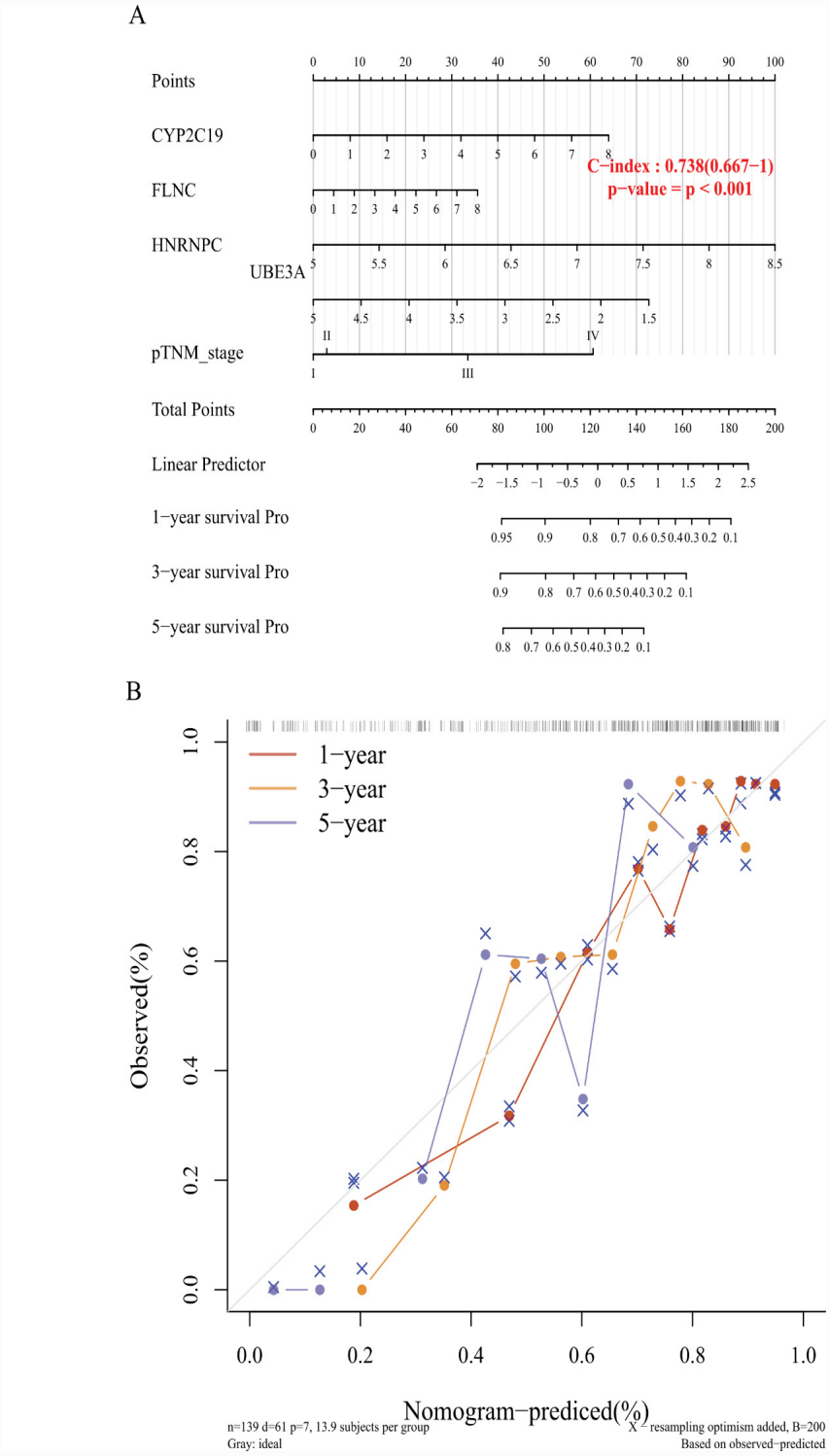

Figure 6: Nomogram of the four-gene feature model.

high-cholesterol diet-fed rats. Food Funct. 2021;12(19):9347-59.

22. Lu M, Wu J, Hao ZW, Shang YK, Xu J, Nan G, et al. Basolateral CD147 induces hepatocyte polarity loss by E-cadherin ubiquitination and degradation in hepatocellular carcinoma progress. Hepatology. 2018;68(1):317-32.

23. Chaturvedi VK, Singh A, Dubey SK, Hetta HF, John J, Singh MP. Molecular mechanistic insight of hepatitis B virus mediated hepatocellular carcinoma. Microb Pathog. 2019;128:184-94.

24. Hu X, Jiang J, Ni C, Xu Q, Ye S, Wu J, et al. HBV integration-mediated cell apoptosis in hepg2.2.15. J Cancer. 2019;10(17):4142-50.

25. Milich DR, Jones J, Hughes J, Maruyama T. Hepatitis B virus infection, the immune response and hepatocellular carcinoma. Ciba Found Symp. 1994;187:113-29; discussion 129-31.

26. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. tissue-based map of the human proteome. Science. 2015;347(6220):1260419.

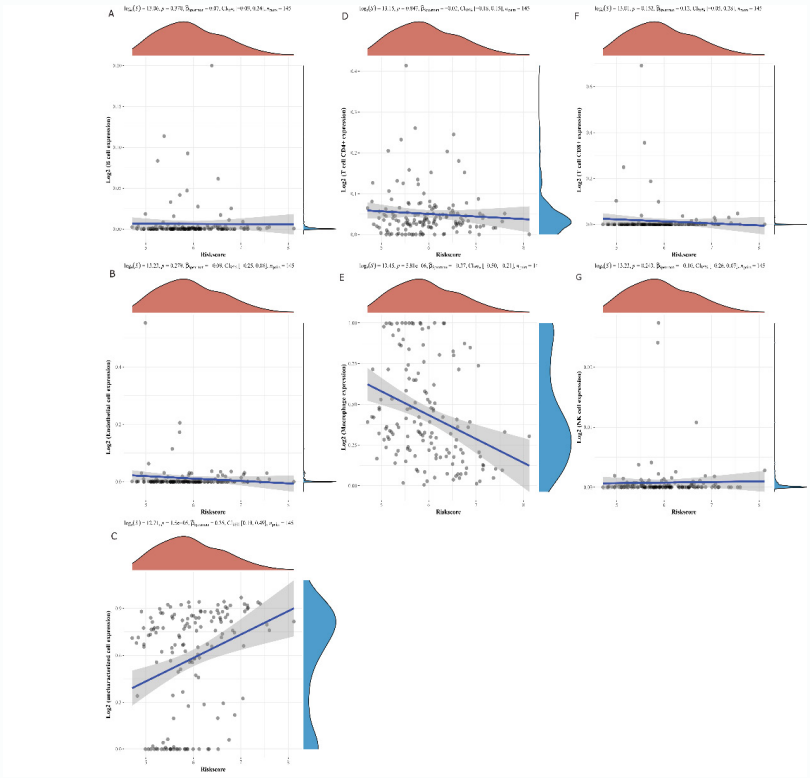

Figure 7: Spearman correlation analysis was performed between model score and immune score.

27. Mitra S, Das S, Chakrabarti J. Systems biology of cancer biomarker detection. Cancer Biomark. 2013;13(4):201-13.

28. Yan X, Chua MS, He J, So SK. Small interfering RNA targeting CDC25B inhibits liver tumor growth in vitro and in vivo. Mol Cancer. 2008;7:19.

29. Wu Y, Zhang Y, Wang F, Ni Q, Li M. MiR-660-5p promotes the progression of hepatocellular carcinoma by interaction with YWHAH via PI3K/Akt signaling pathway. Biochem Biophys Res Commun. 2020;531(4):480-9.

30. Xiao Y, Huang S, Qiu F, Ding X, Sun Y, Wei C, et al. Tumor necrosis factor α-induced protein 1 as a novel tumor suppressor through selective downregulation of CSNK2B blocks nuclear factor-κB activation in hepatocellular carcinoma. EBioMedicine. 2020;51:102603.

31. Xiao H, Cheng S, Tong R, Lv Z, Ding C, Du C, et al. BAG3 regulates epithelial-mesenchymal transition and angiogenesis in human hepatocellular carcinoma. Lab Invest. 2014;94(3):252-61.

32. Gong T, Zhou B, Liu M, Chen X, Huang S, Xu Y, et al. RAB18 promotes proliferation and metastasis in hepatocellular carcinoma. Am J Transl Res. 2019;11(2):1009-19.

33. Yu Z, Chen T, Mo H, Guo C, Liu Q. BRD8, which is negatively regulated by miR-876-3p, promotes the proliferation and apoptosis resistance of hepatocellular carcinoma cells via KAT5. Arch Biochem Biophys. 2020;693:108550.

34. Wan J, Liu H, Yang L, Ma L, Liu J, Ming L. JMJD6 promotes hepatocellular carcinoma carcinogenesis by targeting CDK4. Int J Cancer. 2019;144(10):2489-500.

35. Bo C, Li N, He L, Zhang S, An Y. Long non-coding RNA ILF3-AS1 facilitates hepatocellular carcinoma progression by stabilizing ILF3 mRNA in an m^6A-dependent manner. Hum Cell. 2021;34(6):1843-54.

36. Liu D, Luo X, Xie M, Zhang T, Chen X, Zhang B, et al. HNRNPC downregulation inhibits IL-6/STAT3-mediated HCC metastasis by decreasing HIF1A expression. Cancer Sci. 2022;113(10):3347-61.

37. Qi Y, Xu F, Chen L, Li Y, Xu Z, Zhang Y, et al. Quantitative proteomics reveals FLNC as a potential progression marker for the development of hepatocellular carcinoma. Oncotarget. 2016;7(42):68242-52.

38. Lin YW, Yan MD, Shih YL, Hsieh CB. The basal body gene, RPGRIP1L, is a candidate tumour suppressor gene in human hepatocellular carcinoma. Eur J Cancer. 2009;45(11):2041-9.

39. Xiang Z, Wang S, Xiang, Y. Up-regulated microRNA499a by hepatitis B virus induced hepatocellular carcinogenesis via targeting MAPK6. PLoS One. 2014;9(10):e111410.

40. Tang X, Ge L, Chen Z, Kong S, Liu W, Xu Y, et al. Methylation of the constitutive androstane receptor is involved in the suppression of cyp2c19 in hepatitis b virus-associated hepatocellular carcinoma. Drug Metab Dispos. 2016;44(10):1643-52.

41. Inoue T, Tanaka Y. Novel biomarkers for the management of chronic hepatitis B. Clin Mol Hepatol. 2020;26(3):261-79.

42. Wei RR, Wang CC, Li DJ, Hu XY, Liu K. [Diagnostic value of abnormal prothrombin in hbv-related afp-negative hepatocellular carcinoma]. Sichuan Da Xue Xue Bao Yi Xue Ban. 2020;51(3):411-5.