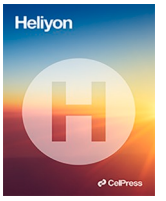
Heliyon 10 (2024) e38729

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Establishment of prognosis model of hepatocellular carcinoma based on prognosis related gene analysis and study on gene regulation mechanism of model

Jingjing Dai [a,*,1], Bo Yang [b,1], Abdusemer Reyimu [c,1], Weiqiang Li [a], Wubi Zhou [c], Xiang Wang [d], Weijie Dai [b,**], Wan Wang [e,***], Jianghong Yan [f], Suxia Hu [g,****]

[a] *Department of Medical Laboratory, The Affiliated Huaian No.1 People's Hospital of Nanjing Medical University, Huai'an, 223300, Jiangsu, People's Republic of China*
[b] *Department of Endoscopy Center, The Affiliated Huaian No.1 People's Hospital of Nanjing Medical University, Huai'an, 223300, Jiangsu, People's Republic of China*
[c] *Department of Pathology, The Affiliated Huaian No.1 People's Hospital of Nanjing Medical University, Huai'an, 223300, Jiangsu, People's Republic of China*
[d] *Department of Paediatrics, The Affiliated Huai'an No.1 People's Hospital of Nanjing Medical University, Huai'an, 223300, Jiangsu, People's Republic of China*
[e] *Department of Reproductive Medicine Center, Anhui No. 2 Provincial People's Hospital, Hefei, 230041, Anhui, People's Republic of China*
[f] *Department of Pediatric Research, Children's Hospital of Hebei Province, Shijiazhuang, 050031, Hebei, People's Republic of China*
[g] *Department of Medical Laboratory, Huainan First People's Hospital, The First Affiliated Hospital of Anhui University of Science and Technology, Huainan, 232007, Anhui, People's Republic of China*

ARTICLE INFO

Keywords:
Hepatocellular carcinoma (HCC)
Module gene
Prognosis
Risk score
Survival

ABSTRACT

Objective: To analyze the expression and mechanism of prognosis related differentially expressed genes (DEGs) in hepatocellular carcinoma (HCC), and establish a prognosis risk model of prognosis related DEGs.
Methods: Transcriptome data and clinical information of 374 HCC samples were downloaded from TCGA data. Kaplan-Meier (KM) survival analysis screened prognostic genes in DEGs. Protein-protein interaction (PPI) network was constructed based on prognostic genes and key genes were screened. Cox regression was used to analyze the key genes and construct the prognostic risk model, calculate the risk score of each patient, divide the patients into low-risk group and high-risk group according to the median risk value, use KM analysis method for survival analysis and draw the survival curve, and use receiver operating characteristic (ROC) curve to evaluate the prognostic risk model, The relationship between risk score and clinicopathological features of HCC patients was analyzed. GEPIA and the human protein atlas (HPA) databases were used for expression verification of model genes. The mirDIP database is used to analyze the regulatory network of model genes. GSCAlite platform is used to analyze the mechanism and drug sensitivity of model genes.

* Corresponding author.
** Corresponding author.
*** Corresponding author.
**** Corresponding author.
E-mail addresses: hayydaijj@njmu.edu.cn (J. Dai), dwj19831016@163.com (W. Dai), wangwan1227@126.com (W. Wang), 1052495833@qq.com (S. Hu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e38729
Received 1 July 2023; Received in revised form 26 September 2024; Accepted 28 September 2024
Available online 30 September 2024

Results: A total of 1987 DEGs were extracted from the transcriptome data of HCC and normal samples, of which 258 were related to the prognosis of HCC ($P < 0.01$). Six key genes (CDK1, CCNA2, BUB1, CDC20, CCNB1 and TOP2A) were screened from the PPI network based on prognostic related genes, and the prognostic risk model was established. Survival analysis showed that the overall survival rate of patients in the high-risk group was significantly lower than that in the low-risk group ($P < 0.01$). The AUC values of 1, 3 and 5 years in the prognostic risk model were 0.716, 0.678 and 0.633. Multivariate Cox regression analysis showed that patient age and patient risk score were independent risk factors for the prognosis of HCC. The model gene is highly expressed in HCC and can promote apoptosis, cell cycle and EMT pathway. In addition, the high expression of model gene produced drug resistance to trametinib, selumetinib and RDEA119 (refametinib).
Conclusion: The prognostic risk model based on six prognostic related DEGs is an independent prognostic factor of HCC, which can effectively predict the survival and prognosis of HCC patients.

1. Introduction

Hepatocellular carcinoma (HCC) is a kind of highly heterogeneous malignant tumor. Different infection history, different personal living environment and individual genetic susceptibility make liver cancer highly heterogeneous [1]. Because of its high mortality, it has become the fourth leading cause of cancer-related death [2]. Although the current treatment of HCC includes multidisciplinary comprehensive treatment such as surgery, radiotherapy, chemotherapy and neoadjuvant therapy, because the early clinical symptoms of HCC are not obvious, most patients are in advanced stage when diagnosed, and their prognosis is still very poor [3]. Early detection, research on new therapeutic targets and effective prediction of the prognosis of HCC patients are important means to improve the treatment and prognosis of HCC.

Due to the strong heterogeneity of HCC, the mechanism of its occurrence and development is not completely clear. Clinically, it is still necessary to explore the markers for HCC [4]. At present, some protein molecules are used as tumor markers in the diagnosis of HCC. For example, alpha-foetoprotein (AFP) is the most widely used tumor marker of HCC in the world. About 70 % of patients with elevated serum AFP. The increase of AFP concentration suggests a poor prognosis [5]. Its sensitivity and specificity were not very satisfactory, especially in the early stage of HCC with small masses, 80 % of patients had no significant increase in serum AFP [6]. Alpha-fetoprotein isoform 3 (AFP-L3) is a type of AFP isoform. AFP-L3 and AFP are widely recognized markers for early diagnosis of HCC [7]. AFP-L3 is unique to HCC cells. AFP-L3 is of great significance in the evaluation of malignancy, therapeutic effect and prognosis of HCC, especially in the early diagnosis of low-level AFP HCC or small HCC [8]. In recent years, des-gamma-carboxyprothrombin (DCP) has also been recommended as a tumor marker for HCC monitoring and has been widely used. Although DCP increased slightly in the serum of patients with liver disease except HCC, the increase of DCP level in the serum of patients with HCC was more significant. DCP plays a certain role in the early screening of AFP negative HCC patients [9]. However, the sensitivity and specificity of commonly used clinical markers in the diagnosis of HCC are still insufficient. Therefore, exploring new biomarkers for early diagnosis and prognosis of HCC is of great significance for its clinical treatment.

In recent years, as a tool for gene expression analysis, high-throughput microarray and sequencing technology have been widely used to identify the changes of genetic information in the process of tumorigenesis [10]. There are abundant tumor related genomes and gene expression profiles in TCGA database, which provides a basis for mining DEGs and gene action mechanism in liver cancer. This study used bioinformatics and TCGA database to screen the DEGs genes with prognostic value for HCC. Through Cox regression analysis, HCC prognostic risk model composed of six DEGs was finally constructed to evaluate the prediction performance of the model and determine the independent prognostic value and clinical relevance of the model, so as to provide reference for the individualized diagnosis of HCC.

2. Materials and methods

2.1. Data acquisition and preprocessing

RNA sequencing data of HCC were downloaded from TCGA (https://portal.gdc.cancer.gov/) database. At the same time, the corresponding clinical information of these patients was collected, including gender, age, lymph node metastasis and tumor stage.

2.2. Differential expression analysis and prognostic gene screening of DEGs

The “limma” package in R software was used to screen DEGs between HCC and normal liver tissues, and the expression profile data were standardized by “edger” package. The screening criteria of differential genes were: discovery error rate (FDR) < 0.01 and $|\log2$ (fold change)$| > 2$. Volcano and heat maps were drawn using “ggplot2″ and “pheatmap” packages for the visualization of differential genes. The relationship between DEGs and overall survival (OS) was determined by survival analysis, and the threshold was set as $P <$ 0.01 to preliminarily screen DEGs with prognostic value for HCC.

2.3. Functional analysis and pathway analysis of differential genes

GO analysis process exists in large-scale function enrichment and is a common research method in function enrichment research. Gene function includes three parts: molecular function (MF), biological process (BP) and cellular component (CC). KEGG pathway exploration is also widely used in bioinformatics analysis, which contains a large amount of data on genome, biological processes and diseases. Using DAVID database, annotation analysis and pathway enrichment analysis were performed on 258 prognostic related DEGs in this study.

2.4. Construction of PPI network and analysis of network modules

258 DEGs obtained from the intersection of Wayne diagrams were uploaded to STRING network analysis to construct protein-protein interaction network (PPI), Take the score greater than 0.9 as the standard. The further screened genes were input into Cytoscape 3.7.2 open source platform for visual analysis of PPI network. Use the molecular complex detection algorithm mcode (molecular complex detection) in Cytoscape to find the core gene module in PPI network. The standard is (degree cutoff = 2,node score cutoff = 0.2,k-core = 2,maxdepth = 100). Then the genes in the core module were enriched and analyzed. We used the cytohubba plug-in to identify key genes using the algorithm of degree and betweenness.

2.5. Construction and evaluation of prognostic risk model

RNA SEQ data were standardized. The standardized expression of DEGs was combined with the survival information of HCC samples. After removing patients without survival time records, a total of 369 HCC samples were included in Cox regression analysis. Cox regression analysis was carried out on the predictive DEGs, and the forest map was drawn to construct the HCC prognosis risk model. The calculation formula of risk score is: riskscore = gene expression 1 × Coef1 + gene expression 2 × Coef2 + … + gene expression n × Coefn (coef: regression coefficient of gene in multivariate Cox regression analysis, N: total number of DEGs related to prognosis). According to the risk score formula, taking the median risk score as the dividing point, each patient with HCC was divided into low-risk and high-risk groups. KM analysis was used to compare the difference of overall survival (OS) rate between high-risk group and low-risk group. The "survivalROC" package of R software was used to draw ROC curve to evaluate the accuracy of HCC prognosis model.

2.6. Correlation analysis of independent prognosis and clinical characteristics of model

Cox regression analysis was performed to confirm the prognostic value of HCC risk score after adjusting for other clinical variables. If the risk score is significantly different from OS in univariate and multivariate Cox analysis, it indicates that the risk score can be used as an independent risk factor. Chi square test was used to compare the distribution of clinicopathological parameters between high-risk group and low-risk group. The risk scores of patients grouped according to clinicopathological and molecular pathological features were compared by *t*-test.

2.7. Differential expression of key genes

GEPIA database (http://gepia.cancer-pku.cn/) is an interactive online website, expression data of 9736 tumors and 8587 normal samples from TCGA and GTEX can be analyzed. The mRNA expression levels of these hub genes in HCC and normal liver tissues were analyzed using GEPIA online tool. HPA (Human Protein Atlas) is an open database for industrial and academic scientists to freely access the data of human proteome research. The immunohistochemical results of key genes with obvious prognostic value were analyzed by HPA online database.

2.8. Validation and functional experiments of proteins related to cell proliferation and apoptosis

Based on the results of biological information screening of proteins and HPA database, this experiment selected protein molecule CDK1 closely related to cell proliferation and apoptosis for immunohistochemical analysis (CDK1, abcam, ab32094, 1:250) of hepatocellular carcinoma, determined their differential expression in cancer tissue and normal tissue, and analyzed their correlation with cell proliferation molecule MKI67 (abcam, ab15580, 1:250). By using relevant protein inhibitors (Ro-3306, abcam, ab141491, 25 μM) or interfering RNA techniques to reduce the expression of target proteins, we analyzed the apoptosis of cells after the reduction of target proteins and further explored their relationship with apoptosis.

2.9. Prediction of model gene miRNA regulatory network

MirDIP database integrates human related target gene information from 30 source databases, which can improve the reliability of prediction results. The one-way search function in mirDIP predicts miRNAs targeting model genes. The 1 % miRNA with the highest confidence is considered to be a potential regulatory miRNA. The miRNA gene regulatory network was visualized by Cytoscape software.

2.10. Pathways involved in model genes and drug sensitivity analysis

GSCAlite (http://bioinfo.life.hust.edu.cn/web/GSCALite/) is a web-based gene set cancer analysis platform. In GSCAlite database, TCGA database tumor data set, GTEx normal control data set and GDSC drug database were integrated. They are well-known cancer-related pathways. Gene set drug resistance analysis was performed from GDSC drug database. Spearman correlation indicates that gene expression is related to drugs. Positive correlation means that the high expression of the gene is resistant to drugs, and vice versa. The Drugbank database (https://go.drugbank.com/) is used to explore the structure of drugs.

3. Results

3.1. Screening and functional enrichment analysis of prognostic related DEGs

The mRNA expression data of 374 HCC tissues and 50 normal tissues were collected from TCGA database, and 1987 DEGs were screened (Fig. 1A–B). Survival analysis of the above 1987 DEGs showed that 258 DEGs were significantly related to prognosis ($P < 0.01$) (Fig. 2A), including 221 up-regulated genes and 37 down-regulated genes (Fig. 2B). GO function and KEGG pathway enrichment analysis were performed on 258 prognostically valuable DEGs. The GO function enrichment results showed that in BP, DEGs mainly involved mitotic division, DNA repair, microtubule based movement, regulation of cell cycle. In CC, DEGs mainly involves chromosome, central region, condensed chromosome kinetochore, midbody, kinesin complex and spin pole. On MF, DEGs mainly involves microtubule, ATP binding, protein binding and chrome binding. KEGG signaling pathway was mainly enriched in cell cycle, oocyte meiosis, progesterone mediated oocyte matching, p53 signaling pathway, HTLV-I infection and FOXO signaling pathway ($P < 0.05$, Fig. 3B).

3.2. Construction of PPI network and screening of key genes

258 DEGs were input into STRING database to construct PPI network, and 1277 interactions among 132 nodes were mapped with 0.9 as the lowest confidence score. The results were visualized using Cytoscape software (Fig. 4A), and the core modules ranked first were selected with MCODE plug-in (Fig. 4B). The module includes 39 nodes and 681 edges. GO analysis is mainly enriched in cell division, cell promotion, microtubule based movement, cell cycle, protein procurement catabolic process, midbody, kinetochore, cytosol, microtubule, spindle pole, ATPase activity, cyclin dependent protein serine/threonine kinase activity (Fig. 5A). KEGG analysis was mainly concentrated in cell cycle, oocyte meiosis, p53 signaling pathway, HTLV-I infection and FOXO signaling pathway (Fig. 5B). Use the degree and betweenness algorithms in the cytohubba plug-in to calculate the top 10 node genes respectively and visualize the PPI network (Fig. 4C and D). Among them, six overlapping genes were further used as HCC model genes (Table 1).

3.3. Construction and evaluation of HCC prognostic risk model

Univariate Cox analysis showed that six genes (CDK1, CCNA2, BUB1, CDC20, CCNB1 and TOP2A) were significantly correlated with the prognosis of HCC patients (Fig. 6A). After further multivariate Cox regression analysis, the prognostic risk model of HCC was formed (Fig. 6B). After integrating 6 genes and weighting their multivariable Cox regression coefficients, the risk score formula is obtained: $(-0.060094715) \times$ Expression of CDK1) $+ (0.0300994 \times$ Expression of CCNA2) $+ (0.193051398 \times$ Expression of BUB1) $+ (0.23640234 \times$ Expression of CDC20) $+ (-0.00465794) \times$ Expression of CCNB1) $+ (-0.102246728) \times$ Expression of TOP2A). The risk values of 369 HCC patients were calculated, and the median risk value was used as the critical value. They were divided into high-risk group (184 cases) and low-risk group (185 cases). HCC patients were sorted by risk score and divided into high and low risk subgroups

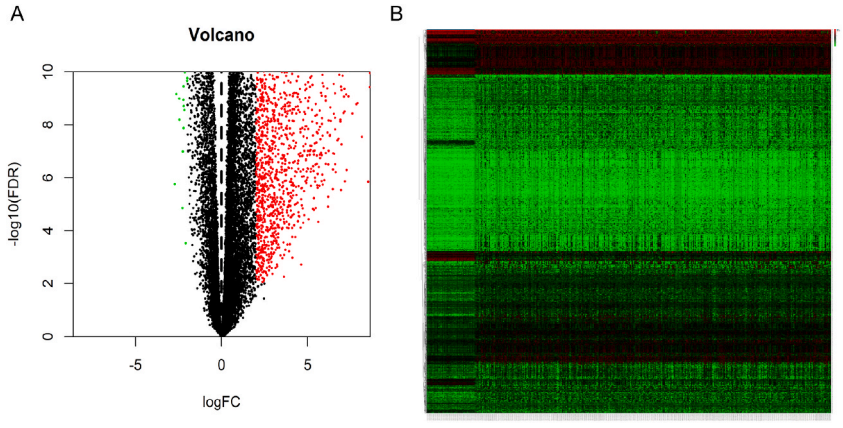
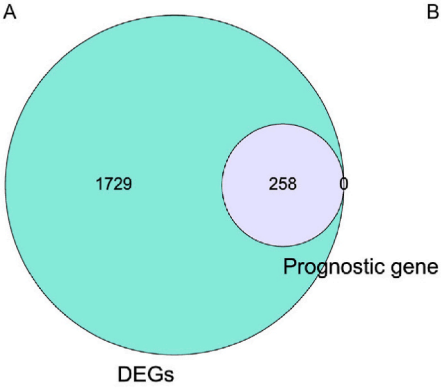


Fig. 1. Screening of DEGs in HCC. (A) Volcano map of DEGs differentially expressed between HCC and normal tissues. (B) The expression heat map corresponding to DEGs.

A



B

DEGS	Prognostic gene
Up-regulated	ADAM12,AGR2,AKR1B15,ANKRD34C,ANLN,AP1M2,ARHGAP11A,ASPM,ATP1A2,ATP6V0D2,AURKA,AURKB,AXDND1,BIRC5,BPIFB4,BUB1,BUB1B,CA9,CAPN11,CBX2,CCDC170,CCNA2,CCNB1,CCNB2,CCNE1,CCNE2,CCNF,CD34,CDC20,CDC25A,CDC25C,CDC45,CDC6,CDC7,CDCA2,CDCA5,CDCA8,CDK1,CDKN2C,CDT1,CELSR3,CENPA,CENPE,CENPF,CENPI,CENPK,CENPL,CENPM,CENPW,CEP55,CHAF1B,CIP2A,CKAP2L,CKMT2,CLEC2L,CLSPN,COL15A1,COL24A1,CREG2,CSPG5,CTHRC1,CTSE,CTSV,DEPDC1,DEPDC1B,DHDH,DIAPH3,DLGAP5,DNAJC8,DPF1,DSCC1,DTL,E2F1,E2F2,E2F7,E2F8,EBF3,ECT2,ELOVL3,EME1,EPHX4,EPS8L3,ERCC6L,ETV4,EXO1,EZH2,FABP6,FABP7,FAM24B,FAM72B,FAM72D,FAM83D,FANCD2,FANCI,FOXM1,G6PD,GABRA3,GINS1,GLDN,GPRIN1,GTSE1,GUCY2C,HASPIN,HELLS,HES2,HJURP,HMMR,HOXD10,HOXD9,IBSP,IQGAP3,KIF11,KIF14,KIF15,KIF18A,KIF18B,KIF20A,KIF23,KIF2C,KIF4A,KIF5A,KIFC1,LHX8,LY6H,MAD2L1,MAMSTR,MAPT,MCM10,MCM2,MELK,MESP2,MKI67,MMP1,MMP10,MTFR2,MUC5AC,MYBL2,MYH7B,MYO18B,N4BP3,NCAPG,NCAPH,NDC80,NEIL3,NEK2,NKX32,NUF2,NXPH4,OIP5,ORC1,OTOG,PAEP,PARPBP,PBK,PCLAF,PLK1,PLK4,PLVAP,POLQ,PPP2R2C,PPY,PRAMEF2,PRC1,PRLHR,PRR11,PRR19,PTPRN,PTTG1,QRFPR,RACGAP1,RAD51,RAD51AP1,RAD54L,RDM1,RECQL4,RGS20,RNFT2,RRM2,SCN4A,SFN,SGO1,SGO2,SHCBP1,SKA1,SKA3,SLC7A11,SOX11,SPATC1L,SPC25,SPP1,STARD6,STC2,STEAP1B,STK39,STMN1,TCF19,TDRD5,TEDC2,TICRR,TINAG,TK1,TLX3,TMC7,TOP2A,TPX2,TRAIP,TRIM54,TRIP13,TRPC6,TTK,UBE2S,UBE2T,UGT1A5,UHRF1,USH1G,VSIG10L,WDHD1,WDR62,XRCC2,ZIC2,ZWINT
Down-regulated	ADH4,ADRA1A,ADRA2B,C7,CCL23,CD5L,CLEC1B,COL25A1,COLEC10,CPEB3,CTNNA3,CYP2C8,CYP4A11,DNASE1L3,ESR1,GHR,GPR182,GYS2,HSD17B13,IGFALS,IL1RL1,KBTBD11,LCAT,MT1HL1,MYOM2,NDST3,NPY1R,NTF3,PZP,RANBP3L,RIPOR3,SCGB3A1,SLC22A1,SOCS2,TPPP2,UROC1,VIPR1

Fig. 2. Identification of prognostic related DEGs in HCC. (A) Venn diagram of the proportion of prognostic genes in deg. (B) Expression of prognostic genes in tumor tissues.

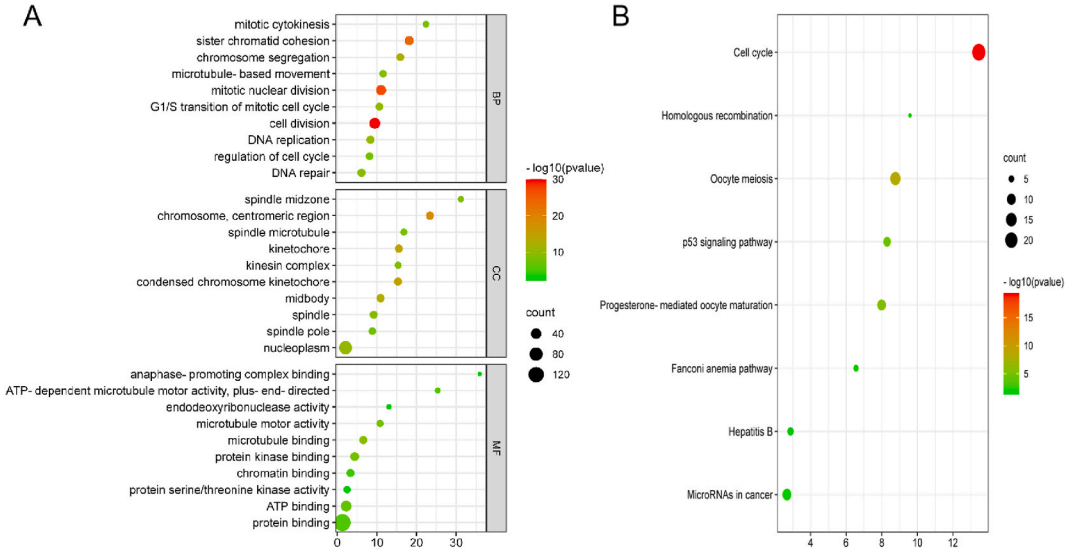


Fig. 3. Functional enrichment analysis of 258 prognostic genes. (A) GO enrichment analysis of prognostic genes. (B) KEGG pathway analysis of prognostic genes. According to the p value from small to large, the color of the circle increases from red to green. The size of the circle indicates the count, that is, the number of genes.

(Fig. 7A). The expression of CDK1, CCNA2, BUB1, CDC20, CCNB1 and TOP2A increased with the increase of HCC risk score (Fig. 7B). In the high-risk subgroup, the death population of HCC is more intensive and the survival time is shorter (Fig. 7C). The KM survival curve showed that the survival rate of patients in the high-risk group was significantly lower than that in the low-risk group (P = 6.33e-07) (Fig. 7D). The time-dependent ROC curve was used to evaluate the predictive performance of the prognostic model for the prognosis of patients at 1, 3 and 5 years. The AUC (area under curve) of ROC curve in 1 year, 3 years and 5 years are 0.716, 0.678 and 0.633 respectively. (Fig. 7E). The above evaluation results show that the risk score model has good sensitivity and specificity in predicting the prognosis of HCC.

3.4. Clinical application of prognostic risk model

The risk scores of 6 model genes were combined with the clinical data (age, gender, stage and grade) of HCC samples. After removing patients with unknown information, a total of 234 HCC samples were included in the analysis. Chi square test analysis showed that there were differences in tumor stage, T stage and survival status distribution between high and low risk groups (Fig. 8A). ROC curve showed that stage (AUC = 0.712), T stage (AUC = 0.699) and riskScore (AUC = 0.698) had good predictive value for the prognosis of HCC. T-test analysis found that there were differences in risk scores of patients grouped according to tumor stage, T-stage and grade (Fig. 8E–G). With the increase of stage, the risk score of patients also showed an upward trend. At the same time, age, gender, tumor stage, TNM stage, grade and risk score were included in Cox regression analysis. Univariate Cox analysis showed that tumor

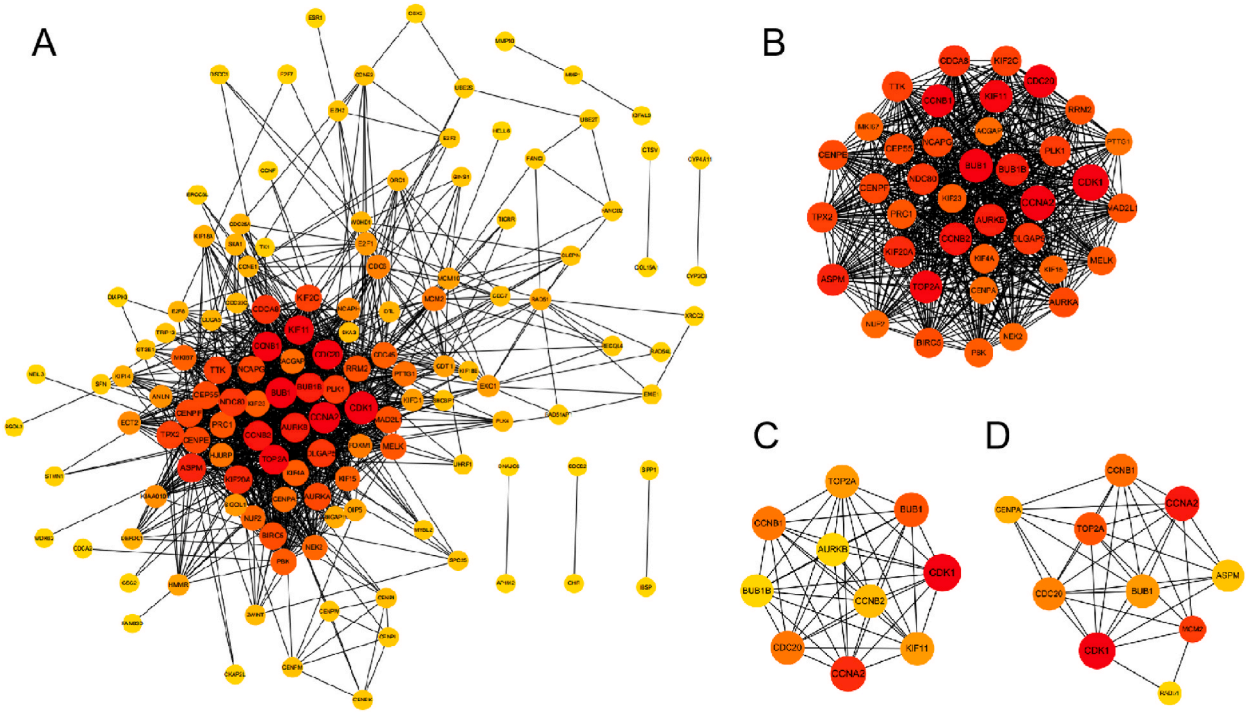


Fig. 4. Prognostic related DEGs were used to construct PPI networks and subnetworks. (A) Cytoscape visualized the genes of PPI network. The node color deepens as the degree value increases. (B) The mcode plug-in filters out the most significant modules in the PPI network. (C) The top ten genes with degree value. The node color deepens as the degree value increases. (D) Genes with betweenness value in the top ten. The node color deepens as the betweenness value increases.

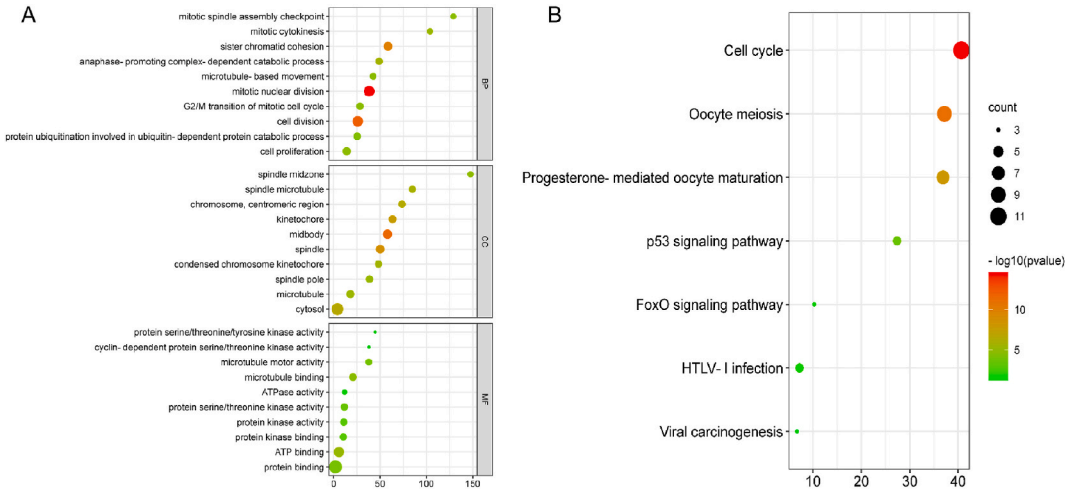


Fig. 5. Functional enrichment analysis of 39 core module genes. (A) GO enrichment analysis of module genes. (B) KEGG pathway analysis of module genes. According to the p value from small to large, the color of the circle increases from red to green. The size of the circle indicates the count, that is, the number of genes.

stage, T stage, M stage and risk score were prognostic risk factors for patients with HCC ($P < 0.05$, Fig. 8C).

Multivariate Cox analysis showed that risk score was a reliable independent prognostic factor for HCC ($P < 0.001$, Fig. 8D).

3.5. Validation of six model genes in bioinformatics database

The mRNA expression of six model genes was verified in GEPIA database. The results showed that CDK1, CCNA2, BUB1, CDC20,

Table 1
Top 10 Central genes of Degree and Betweenness scores in PPI network.

Gene	Expression	Degree	Gene	Expression	Betweenness
CDK1	up	76	**CDK1**	up	2172.34624
CCNA2	up	67	**CCNA2**	up	978.11916
BUB1	up	62	MCM2	up	841.32533
CDC20	up	58	**TOP2A**	up	759.18841
CCNB1	up	58	**CCNB1**	up	696.60976
TOP2A	up	57	**CDC20**	up	691.43916
KIF11	up	57	**BUB1**	up	628.23855
CCNB2	up	56	CENPA	up	620.1354
BUB1B	up	55	ASPM	up	556.57578
AURKB	up	55	RAD51	up	518.17256

Note: Bold Fonts represent overlapping genes in the two algorithms.

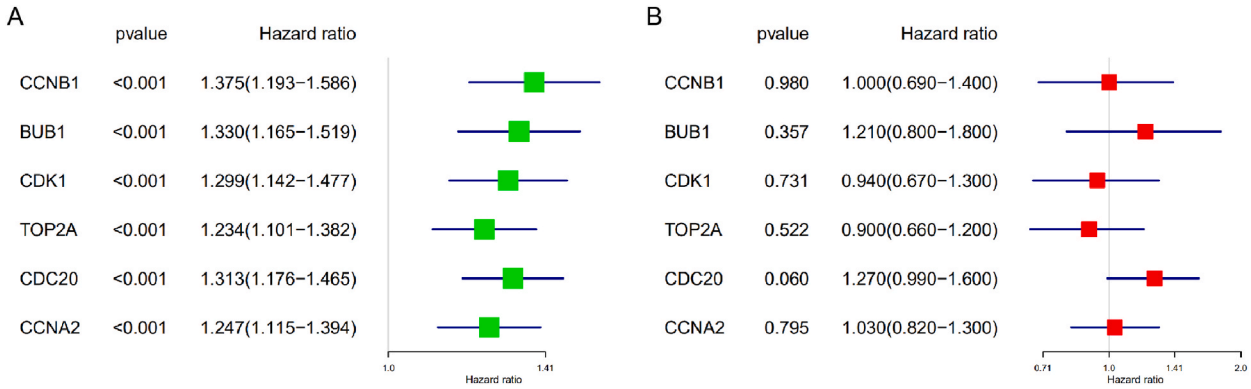


Fig. 6. Univariate and multivariate prognostic analysis of six model genes. (A) Univariate prognostic analysis forest map of model gene. (B) Multivariate prognostic analysis of model genes.

CCNB1 and TOP2A were highly expressed in HCC (Fig. 9). In addition, HPA database was used to show the differences in protein levels of CDK1, CCNA2, CDC20, CCNB1 and TOP2A. Immunohistochemical staining was shown in Fig. 10 (BUB1 data was missing in the database). Even in tissues, CDK1, CCNA2, CDC20, CCNB1 and TOP2A proteins were significantly stained in HCC.

3.6. The expression of CDK1 and MKI67 in HCC

In Fig. 11, the expression of CDK1 and MKI67 in HCC is higher than that in normal adjacent tissues and has statistical significance ($P < 0.05$). MKI67 is known to be a marker of cell proliferation, CDK1 is positively correlated with MKI67(r = ,p =). Therefore, CDK1 is closely related to cell proliferation.

3.7. The functional mechanism of CDK1 in hepatocellular carcinoma cell line HepG2

Further experiments were needed to elucidate the proliferation mechanism of CDK1 in HCC. We conducted cell function tests to illustrate it. As shown in Fig. 12, we inhibited the expression of CDK1 in HepG2 cells by Ro-3306, and detected the apoptosis before and after inhibition. We found that inhibiting CDK1 increased HepG2 apoptosis, which was statistically significant (p =). Therefore, CDK1 has the effect of inhibiting cell apoptosis and enhancing HCC proliferation in HCC.

3.8. Regulatory network and drug sensitivity analysis of model genes

In this study, target miRNAs corresponding to model genes were further collected and analyzed. The number of sources that can predict the regulatory relationship between miRNA and model genes is greater than 5 as the inclusion criteria. MirDIP database analysis showed that there were 66 potential regulatory relationships between the four model genes and 69 miRNAs(Fig. 13). Explore the regulatory mechanism of six model genes on GSCAlite platform. The results showed that CDK1, CCNA2, BUB1, CDC20, CCNB1 and TOP2A could promote apoptosis and cell cycle in most tumors (Fig. 14A). Meanwhile, in lihc, CDK1, CCNA2, BUB1, CDC20, CCNB1 and TOP2A can promote EMT, cell cycle and apoptosis pathway (Fig. 14B). The drug sensitivity results showed that the high expression of CDK1, CCNA2, BUB1 and TOP2A was resistant to trametinib, selumetinib and RDEA119 (refametinib) (Fig. 15A). Finally, the drug structure was visualized (Fig. 15B–D).

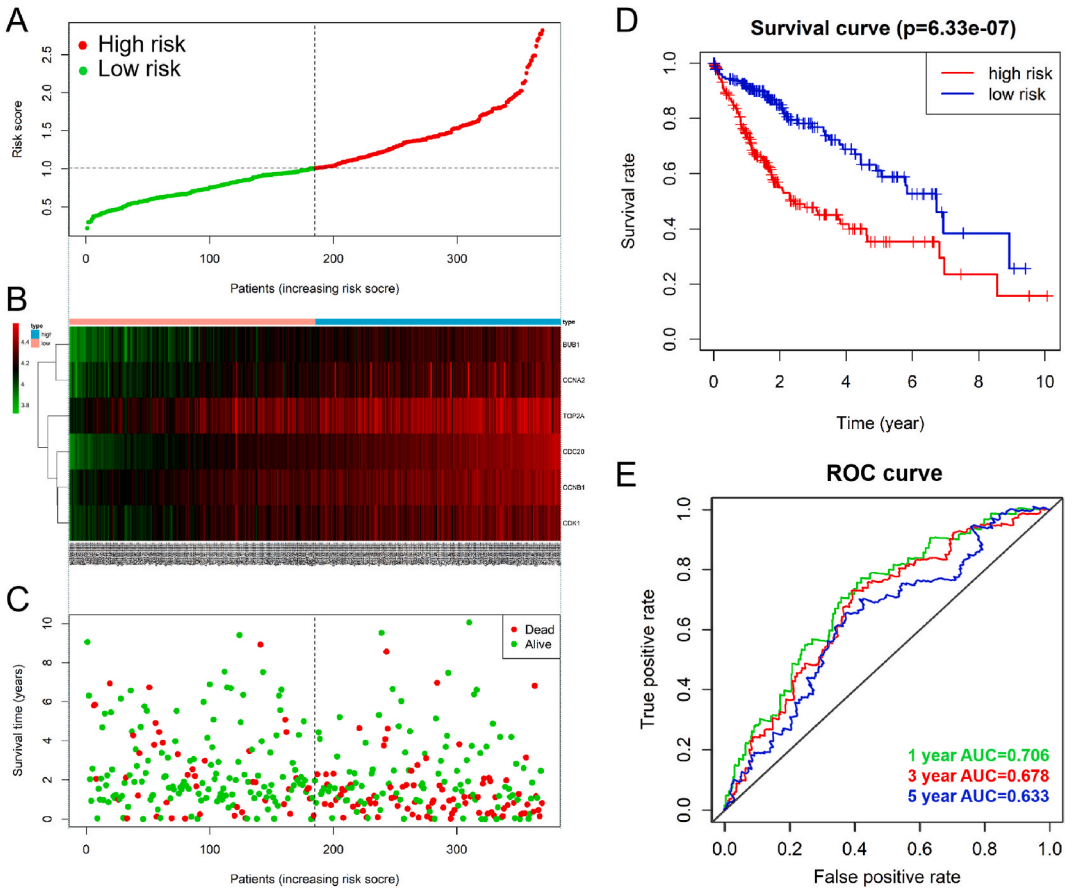


Fig. 7. Performance evaluation of prognostic risk scoring model. (A) Patients' risk score and grouping: red represents high-risk group and green represents low-risk group. (B) Expression heat map of 6 model genes (green indicates low expression and red indicates high expression). (C) Survival status distribution of patients in each group (red represents death and green represents survival). (D) KM survival curve of patients in high and low risk group. (E) Time dependent ROC curve to evaluate the predictive performance of prognostic risk model.

4. Discussion

Hepatocellular carcinoma (HCC) remains high in the mortality of malignant tumors [11]. Because liver cancer usually occurs in the damaged liver, the recurrence rate is high and the prognosis is poor. At present, clinical guidelines can not fully reflect the specific factors affecting the prognosis of patients with liver cancer [12]. The occurrence of liver cancer has experienced a series of complex biological processes, which is the result of the joint action of multiple genes and pathways, which greatly limits the clinical early diagnosis and treatment of tumors [13]. At present, although there are clinical biomarkers such as AFP, AFP-L3% and DCP, comprehensive evaluation is still needed to provide reference for the diagnosis of HCC. Nowadays, the development of next-generation sequencing technology makes it possible to determine the gene map of human cancer. Using genomic information to find tumor prognosis indicators in individuals is a new trend in tumor diagnosis and treatment in the future [14].

Firstly, the transcriptome data of HCC were downloaded from TCGA database, and the differential genes between cancer and normal tissues were screened. Among the 1987 DEGs screened, 258 were related to the prognosis of HCC patients. In addition, these prognostic genes are mainly involved in cell division and cell cycle regulation. In the PPI network constructed based on prognostic genes, the most important network modules were selected. Similarly, this module plays a key role in the regulation of cell cycle. It is speculated that the poor prognosis of patients may be related to the abnormal regulation of cell cycle. Subsequently, six key genes (CDK1, CCNA2, BUB1, CDC20, CCNB1 and TOP2A) were obtained through the degree and betweenness algorithms in the cytohubba plug-in.

Cyclin-dependent kinase 1 (CDK1) plays a key role in eukaryotic cell cycle control by regulating centrosome cycle and mitotic initiation. CDK1 was up-regulated in HCC, which was significantly related to the low overall survival rate [15]. CDK1 promotes the expression of cell cycle genes and is essential in cell division [16], and targeting CDK1 can inhibit the proliferation of HCC cells [17]. Consistent with this study, CDK1 is up-regulated in HCC and can activate cell cycle pathway. By inhibiting its expression (Ro-3306), it can induce enhanced cell apoptosis. Therefore, CDK1 can serve as a potential therapeutic target or diagnostic marker for HCC. The poor prognosis of patients with high expression of CDK1 may be related to the proliferation caused by over activation of cell cycle.

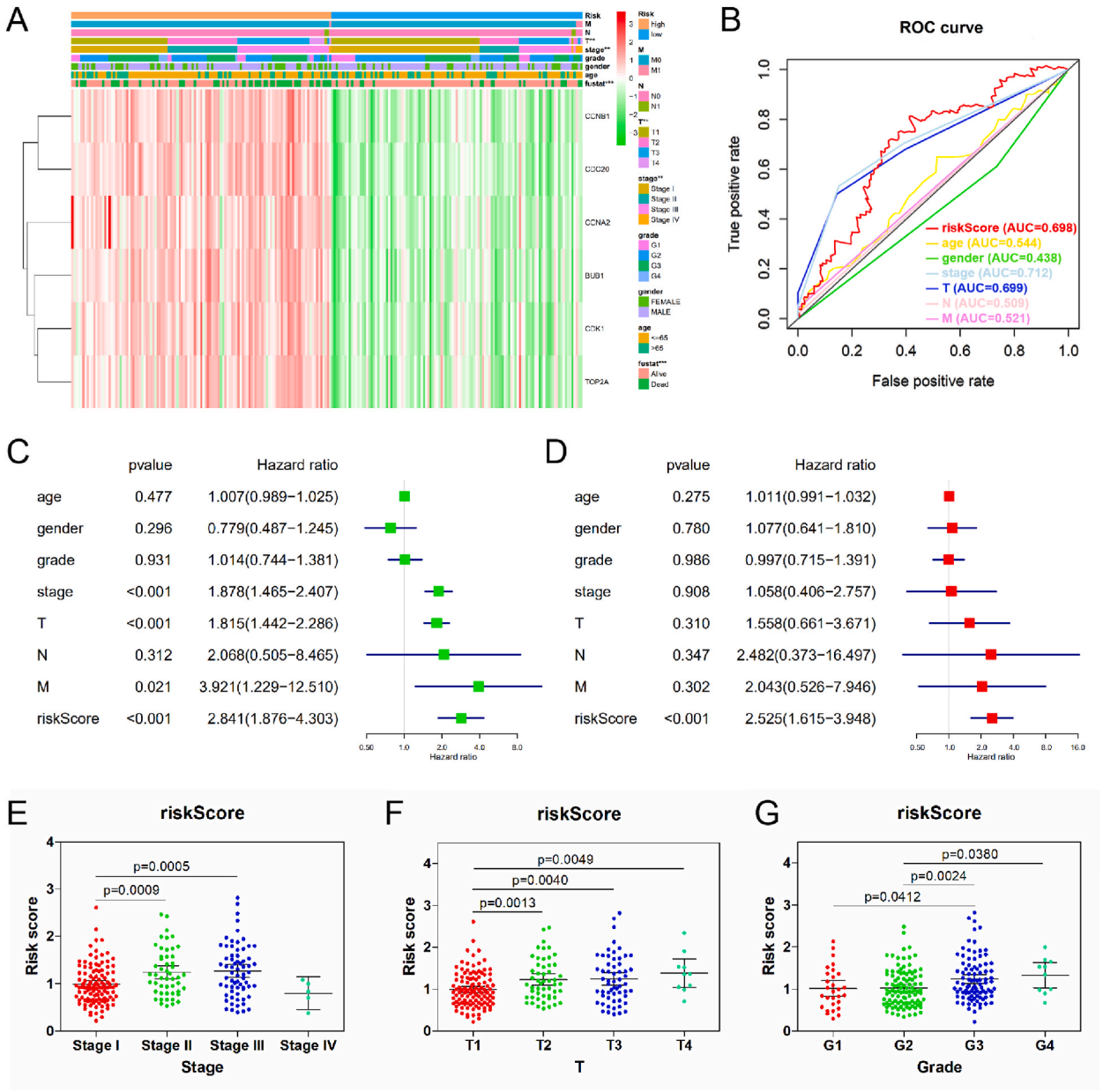


Fig. 8. Verify the accuracy of the prognostic scoring model. (A) HCC risk clinical heat map data set. (B) Time dependent ROC curve to evaluate the predictive performance of clinical factors. (C) Univariate independent prognostic analysis. (D) Multivariate independent prognostic analysis. (E–G) clinical correlation analysis of risk score.

CDK1 is a high-risk gene in the risk model, which can be used as a marker to predict the prognosis of patients.

Cyclin A2 (CCNA2) controls cyclin in G1/s and G2/M transition of cell cycle, and its high expression may promote the formation of HCC [18]. In vitro experiments show that CCNA2 contributes to HCC [19]. CCNA2 inhibition experiment shows its potential as a therapeutic target for HCC [20]. In this study, CCNA2 is highly expressed, which is a high-risk factor in HCC and promotes the cell cycle pathway. Patients with high expression of CCNA2 have poor prognosis and may play a cancer promoting role in HCC.

BUB1 is a serine-threonine kinase, which is essential for spindle assembly checkpoint signal and correct chromosome arrangement [21]. Therefore, BUB1 plays a role in promoting cancer cell proliferation and tumor progression [22]. Mir-490-5p inhibits the proliferation, migration and apoptosis of HCC cells by inhibiting BUB1. Similarly, this study also indicated the promoting effect of BUB1 on cell cycle in HCC, and its high expression was associated with poor prognosis. It can be used as a reliable index to predict the prognosis of HCC.

The up-regulated expression of cell division cycle protein 20 (CDC20) in HCC specimens is related to tumor differentiation and

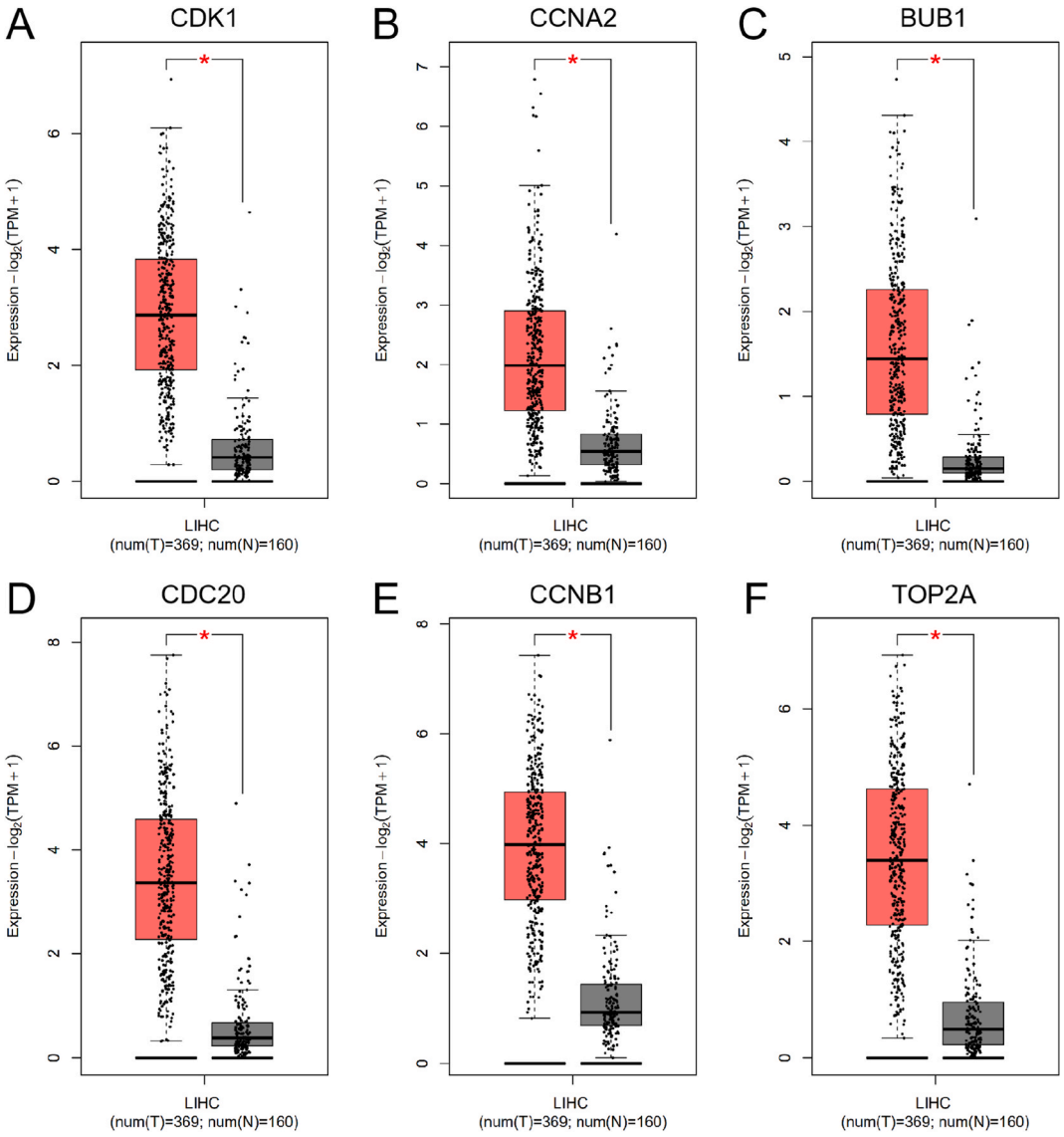


Fig. 9. The expression differences of six model genes in liver cancer and normal tissues were analyzed by GEPIA database. (A) CDK1. (B) CCNA2. (C) BUB1. (D) CDC20. (E) CCNB1. (F) TOP2A. $*p < 0.01$.

metastasis, resulting in low patient survival rate [23]. The up regulation of CDC20 can be used to predict the prognosis of HCC [24]. In this study, CDC20 is also associated with poor prognosis, and can be used as a high-risk gene to predict the prognosis of patients.

G2/mitotic specific cyclin-b1 (CCNB1) is very important to control the cell cycle of G2/M (mitotic) transition phase, and belongs to cyclin family. The up-regulated expression of CCNB1 promotes the proliferation of HCC cells [25]. High expression of CCNB1 predicts poor prognosis of HCC. Knockout of CCNB1 inhibited cell proliferation and invasion in human HCC by targeting CCNB1 [26]. In our study, CCNB1 is not only a high-risk gene, but also can promote the cell cycle, which may be the reason why it plays an important role in cell proliferation. Patients with poor prognosis of HCC have high expression of CCNB1, which may be a potential tumor marker.

DNA toposomerase 2 (TOP2A) is a ribozyme that controls the topological state of DNA. It is very important for the correct division of ion chromosomes during mitosis and meiosis [27]. The up-regulation of TOP2A expression is related to the shortening of survival time and chemoresistance [28]. Consistent with our study, TOP2A is not only highly expressed in HCC, but also associated with cell cycle and poor prognosis. Immunohistochemistry showed that it was higher in HCC than other model genes, and could be used as a potential tumor marker of HCC.

Although a single molecular marker has certain auxiliary diagnostic significance, the combined detection of multiple markers can improve the detection rate of HCC, make up for the deficiency of a single marker, and better evaluate the prognosis of patients. Therefore, we verified the correlation between these six key genes and patient prognosis through Cox analysis. Univariate analysis showed that six key genes were related to the prognosis of patients. At the same time, the comprehensive risk score of six key genes

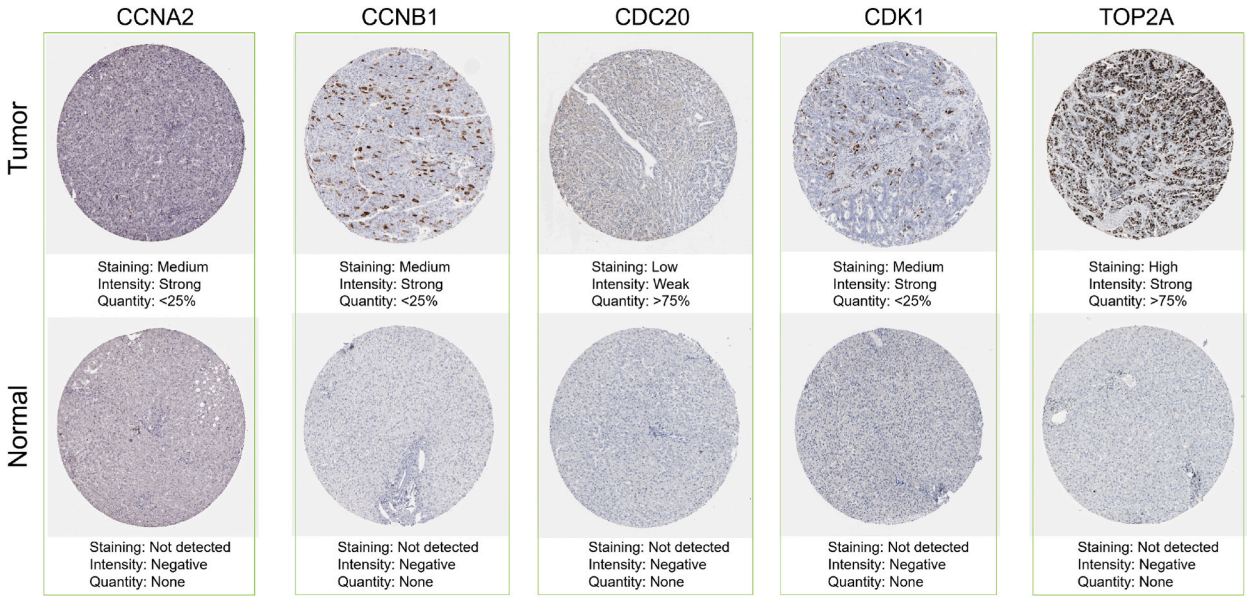


Fig. 10. Immunohistochemistry of model genes in HCC and normal samples based on HPA database. Brown indicates protein expression.

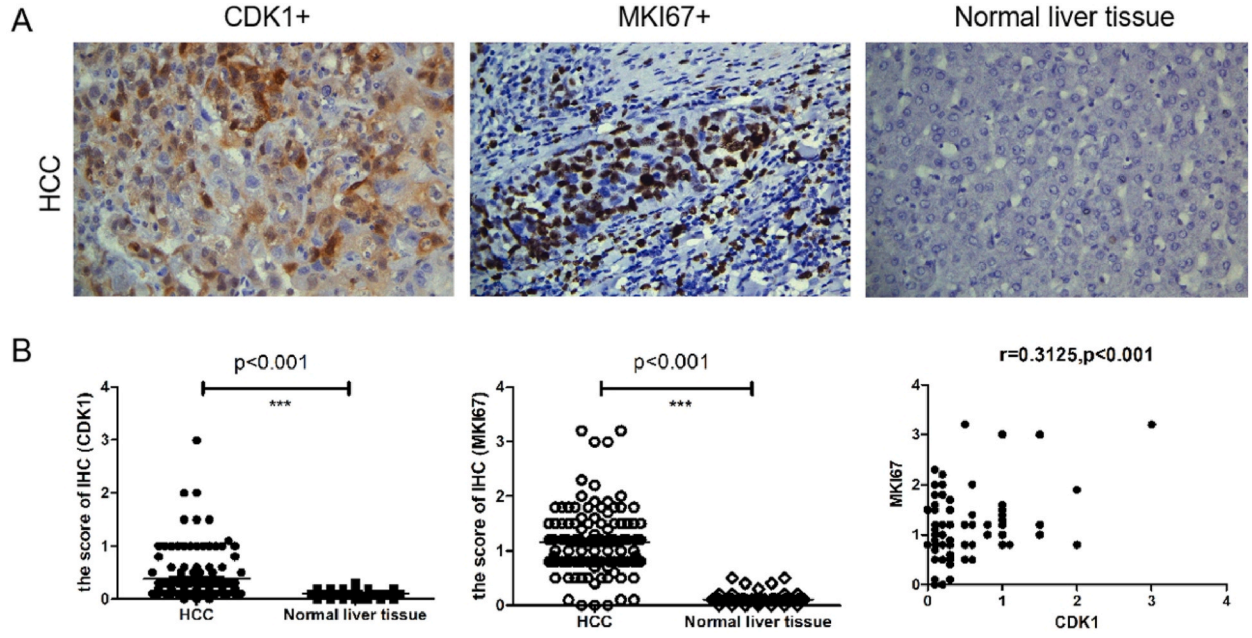


Fig. 11. The expression of CDK1 and MKI67 in HCC. (A) The positive expression of CDK1 and MKI67 in HCC. Negative expression of normal tissue (Magnification, × 400). (B) Statistical analysis of protein expression and Correlation analysis of two protein expressions.

obtained by multivariate Cox analysis has a good predictive ability for patients and prognosis. Then, a prognostic scoring model of HCC was constructed based on six key genes. The prognosis of high-risk group is worse than that of low-risk group. In addition, when using the time-dependent ROC to analyze the prediction performance of the risk score on the prognosis of patients, it was found that the AUCs of 1 year, 3 years and 5 years were 0.706, 0.678 and 0.633 respectively (Fig. 7E), indicating that the risk score has a better prediction on the prognosis of patients in 1 year.

In order to further analyze the correlation between the risk score and the clinical characteristics of patients, we performed chi square test and *t*-test analysis. Results there were significant differences in tumor stage, T stage and survival between high and low risk groups. In addition, with the increase of tumor stage, the risk score of patients also showed an upward trend. At the same time, combined with the patient's clinical information and risk value, univariate and multivariate Cox regression analysis were carried out.

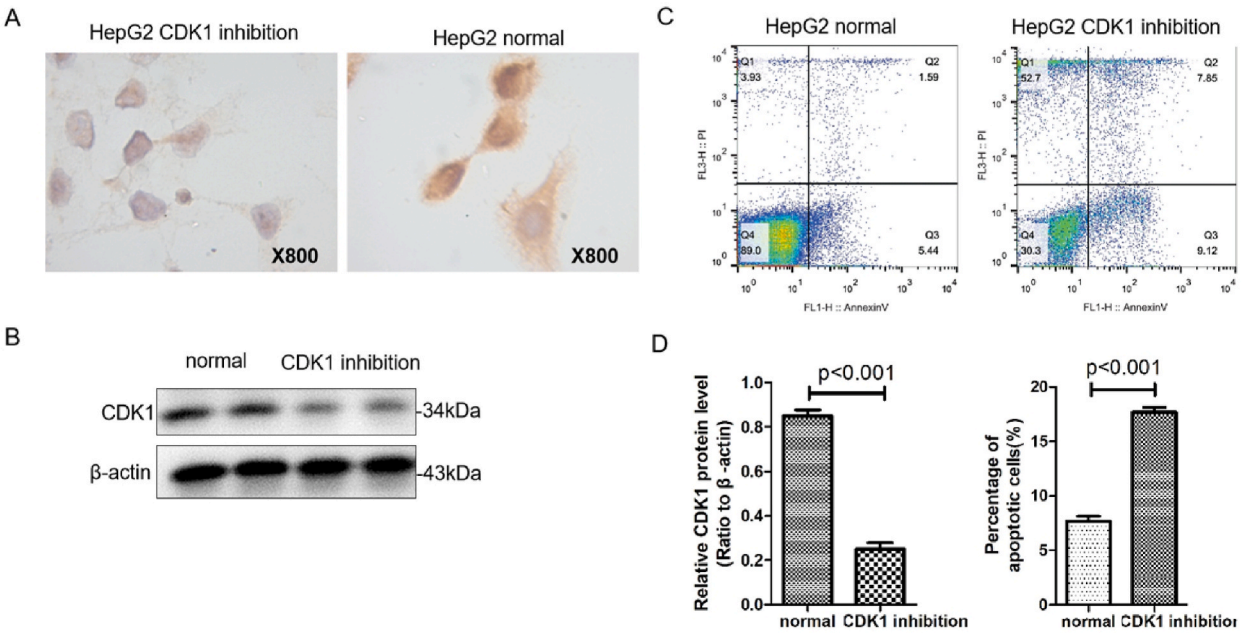


Fig. 12. Apoptosis analysis after CDK1 inhibition. (A) Immunochemical analysis of CDK1 expression in HepG2 cell lines after using CDK1 inhibitors. (B) Western blot analysis of CDK1 expression in HepG2 cell lines after using CDK1 inhibitors. (C) Detection of cell apoptosis through apoptosis detection kit. (D) Statistical analysis of protein and cell apoptosis.

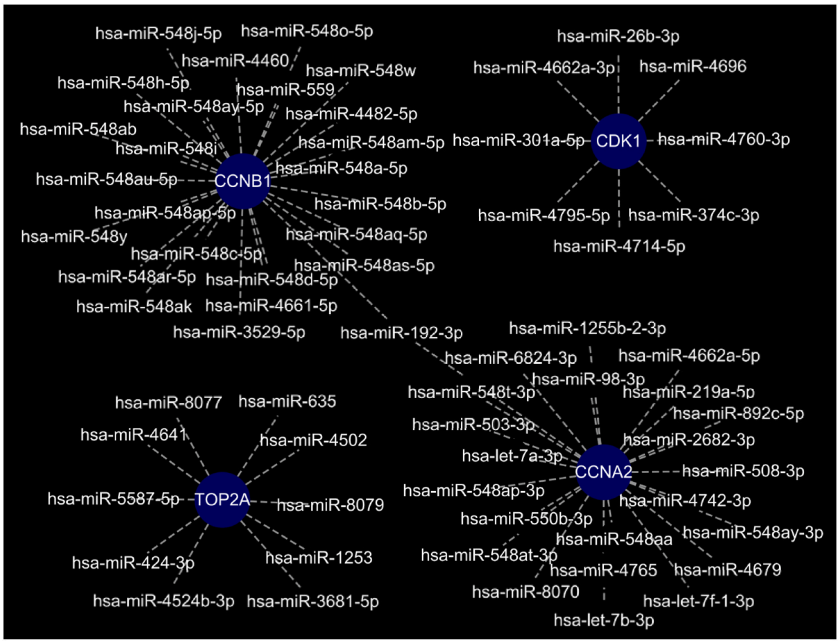


Fig. 13. Potential miRNA regulatory networks of model genes. The blue node represents mRNA, and the miRNA connected with the blue node by dotted line.

It was found that tumor stage, T stage, M stage and risk score were related to the prognosis of HCC, but only risk score was an independent risk factor of liver cancer.

The highlights and significance of this work as compared to the previous studies are as follows. Firstly, we use conventional bioinformatics analysis to identify potential liver cancer biomarkers or therapeutic targets, providing a foundation for research in related fields. Although the research methods are relatively conventional, the differential gene expression results obtained through cross analysis of different chips are also a prerequisite for subsequent related research, and have important exploratory value for the

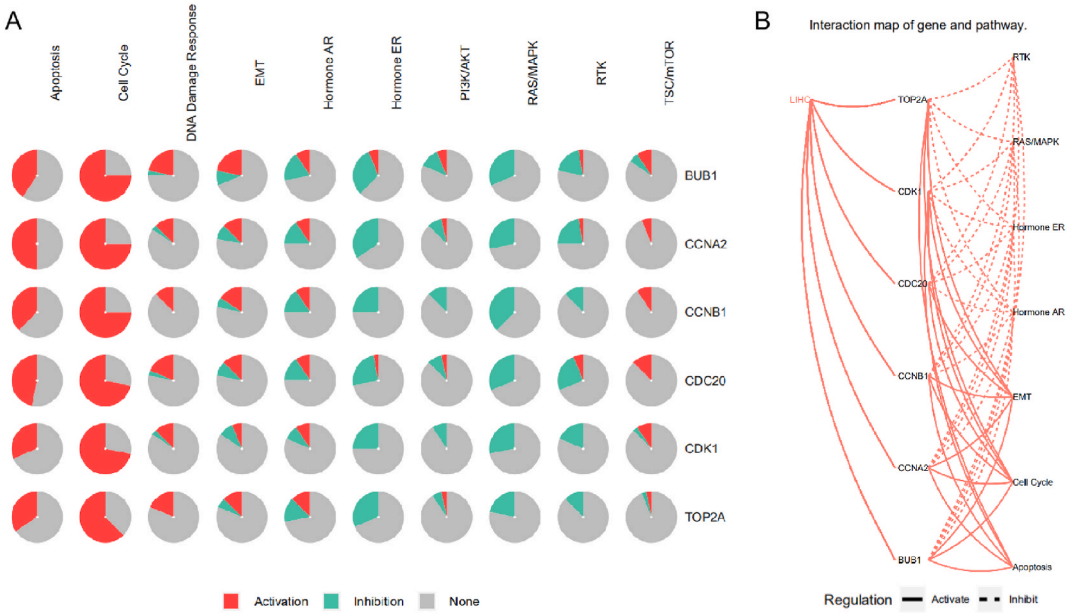


Fig. 14. Pathway prediction analysis of 6 model genes. (A) Global percentage of cancers in which genes affect pathways among 32 cancer types (number of cancer types activated or inhibited/32 * 100 %). (B) Possible pathways of six model genes in LIHC. The solid line indicates the pathway connected with gene promotion, and the dotted line indicates the pathway connected with gene inhibition.

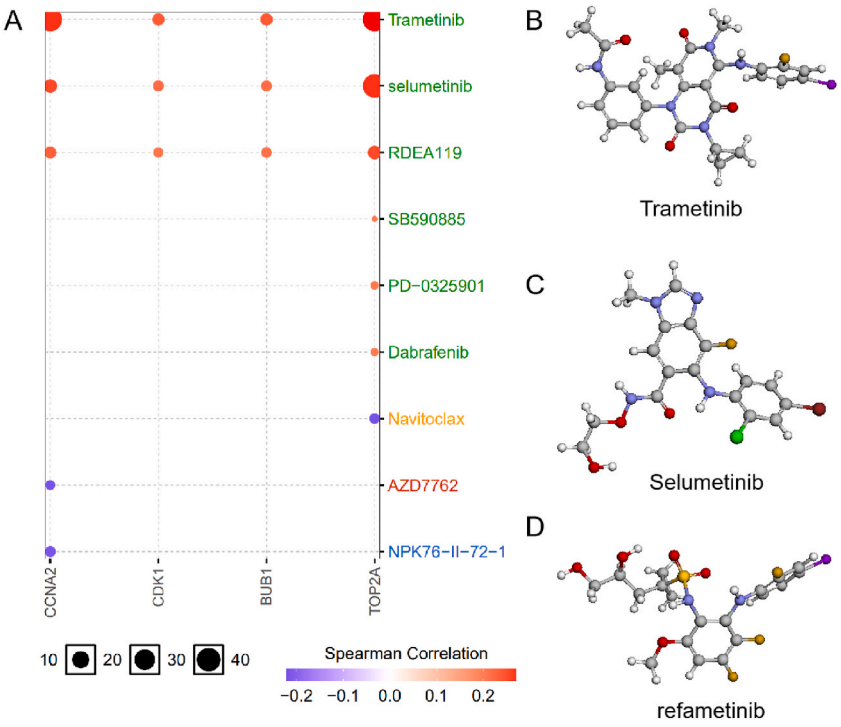


Fig. 15. Drug resistance analysis of model genes. (A) Model gene resistance was analyzed from gdsc drug data. Spearman correlation represents the correlation between gene expression and drugs. Red dots represent positive correlation and purple dots represent negative correlation. Positive correlation means that the high expression of the gene is resistant to the drug. (B–D) The ball-and-stick structures of the drugs were displayed through the Drugbank database.

diagnosis and treatment of hepatocellular carcinoma. Secondly, by combining our own clinical data to verify the feasibility and reliability of the results of this study, it guides further clinical research, such as the clinical application of CDK1 as a therapeutic or diagnostic target for hepatocellular carcinoma. Last but not least, it is important to note that the current high incidence rate of hepatocellular carcinoma and the lack of specific markers for diagnosis and treatment are the key reasons for our study. For example, Normal expression of AFP is frequently detected in HCC, the HCC has progressed to the middle and late stages and timely treatment has been missed.

5. Conclusion

To sum up, this study screened DEGs related to HCC prognosis in TCGA database, conducted functional enrichment analysis on DEGs, and further constructed a HCC prognostic risk model through univariate and multivariate Cox risk regression analysis. The model has high predictive value for the survival rate of HCC. The DEGs found in the study provide a theoretical basis for the study of the pathogenesis and the exploration of diagnosis and treatment targets of HCC.

Funding

This work was supported by the grants from Plan Project of Huainan Science and Technology in 2020 (No.2020069, to Suxia Hu) and General Project of Development Fund for Affiliated Hospital of Xuzhou Medical University in 2022 (No.XYFM202234, to Xiang Wang).

Conflict of interest

The authors declare that they have no competing interests.

Ethics approval

All research experiments involving patient data were approved by the Ethics Committee (approval number: KY-2022-014-01).

Data availability

The data associated with my study has been deposited into a publicly available repository.

Data can be downloaded from TCGA database (https://portal.gdc.cancer.gov/). Protein protein interaction (PPI) network analysis data are available in the string database (https://www.string-db.org/). The expression analysis results of model genes can be obtained from GEPIA database (http://gepia.cancer-pku.cn/). Immunohistochemical data are available from the human protein atlas (HPA) database (https://www.proteinatlas.org/). Gene miRNA regulatory networks are available from the mirdip database (http://ophid.utoronto.ca/mirDIP/). The pathways involved in the model genes and drug sensitivity results were analyzed by GSCALite platform (http://bioinfo.life.hust.edu.cn/web/GSCALite/).

CRediT authorship contribution statement

Jingjing Dai: Writing – review & editing, Writing – original draft, Software, Resources, Project administration, Investigation, Data curation, Conceptualization. **Bo Yang:** Writing – original draft, Investigation, Data curation, Conceptualization. **Abdusemer Reyimu:** Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Weiqiang Li:** Visualization, Software, Resources, Methodology, Investigation, Formal analysis. **Wubi Zhou:** Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **Xiang Wang:** Software, Resources, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Weijie Dai:** Writing – original draft, Visualization, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Wan Wang:** Writing – review & editing, Supervision, Software, Resources, Methodology, Investigation. **Jianghong Yan:** Visualization, Supervision, Investigation, Formal analysis. **Suxia Hu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Suxia Hu reports financial support was provided by Plan Project of Huainan Science and Technology in 2020. Xiang Wang reports financial support was provided by General Project of Development Fund for Affiliated Hospital of Xuzhou Medical University in 2022. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank the team members for their contributions to this paper, and then we will continue to work hard to do

relevant research.

Abbreviations

HCC	hepatocellular carcinoma
TCGA	the Cancer Genome Atlas
DEGs	differentially expressed genes
PPI	protein protein interaction
KM	Kaplan-Meier
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
HPA	human protein atlas
GEPIA	Gene Expression Profiling Interactive Analysis
OS	overall survival
DAVID	database for annotation, visualization and integrated discovery
ROC	reciver operating characteristic
AUC	Area Under Curve
MF	molecular function
CC	cell components
BP	biological process
OS	overall survival

References

[1] A. Alqahtani, Z. Khan, A. Alloghbi, A.T. Said, M. Ashraf, D.M. Hammouda, Hepatocellular carcinoma: molecular mechanisms and targeted therapies, Medicina (Kaunas) 55 (9) (2019).
[2] Z. Chen, H. Xie, M. Hu, et al., Recent progress in treatment of hepatocellular carcinoma, Am. J. Cancer Res. 10 (9) (2020) 2993–3036.
[3] D. Anwanwan, S.K. Singh, S. Singh, V. Saikam, R. Singh, Challenges in liver cancer and possible treatment approaches, Biochim. Biophys. Acta Rev. Canc 1873 (1) (2020) 188314.
[4] R. Dutta, R.I. Mahato, Recent advances in hepatocellular carcinoma therapy, Pharmacol. Ther. 173 (2017) 106–117.
[5] Y. Zheng, M. Zhu, M. Li, Effects of alpha-fetoprotein on the occurrence and progression of hepatocellular carcinoma, J. Cancer Res. Clin. Oncol. 146 (10) (2020) 2439–2446.
[6] P.P. Song, J.F. Xia, Y. Inagaki, et al., Controversies regarding and perspectives on clinical utility of biomarkers in hepatocellular carcinoma, World J. Gastroenterol. 22 (1) (2016) 262–274.
[7] J. Cheng, W. Wang, Y. Zhang, et al., Prognostic role of pre-treatment serum AFP-L3% in hepatocellular carcinoma: systematic review and meta-analysis, PLoS One 9 (1) (2014) e87011.
[8] F. Qi, A. Zhou, L. Yan, et al., The diagnostic value of PIVKA-II, AFP, AFP-L3, CEA, and their combinations in primary and metastatic hepatocellular carcinoma, J. Clin. Lab. Anal. 34 (5) (2020) e23158.
[9] X. Wang, Y. Zhang, N. Yang, et al., Evaluation of the combined application of AFP, AFP-L3%, and DCP for hepatocellular carcinoma diagnosis: a meta-analysis, BioMed Res. Int. 2020 (2020) 5087643.
[10] A. Reyimu, Y. Chen, X. Song, W. Zhou, J. Dai, F. Jiang, Identification of latent biomarkers in connection with progression and prognosis in oral cancer by comprehensive bioinformatics analysis, World J. Surg. Oncol. 19 (1) (2021) 240.
[11] J.J. Harding, D.N. Khalil, G.K. Abou-Alfa, Biomarkers: what role do they play (if any) for diagnosis, prognosis and tumor response prediction for hepatocellular carcinoma? Dig. Dis. Sci. 64 (4) (2019) 918–927.
[12] A.R. Benson, M.I. D'Angelica, D.E. Abbott, et al., NCCN guidelines insights: hepatobiliary cancers, version 1.2017, J Natl Compr Canc Netw 15 (5) (2017) 563–573.
[13] A. Forner, M. Reig, J. Bruix, Hepatocellular carcinoma, Lancet (N. Am. Ed.) 391 (10127) (2018) 1301–1314.
[14] T.Z. Tan, M. Rouanne, K.T. Tan, R.Y. Huang, J.P. Thiery, Molecular subtypes of urothelial bladder cancer: results from a meta-cohort analysis of 2411 tumors, Eur. Urol. 75 (3) (2019) 423–432.
[15] C.X. Wu, X.Q. Wang, S.H. Chok, et al., Blocking CDK1/PDK1/beta-Catenin signaling by CDK1 inhibitor RO3306 increased the efficacy of sorafenib treatment by targeting cancer stem cells in a preclinical model of hepatocellular carcinoma, Theranostics 8 (14) (2018) 3737–3750.
[16] B.D. Landry, C.E. Mapa, H.E. Arsenault, K.E. Poti, J.A. Benanti, Regulation of a transcription factor network by Cdk1 coordinates late cell cycle gene expression, EMBO J. 33 (9) (2014) 1044–1060.
[17] Y. Zhang, W. Huang, Y. Ran, et al., miR-582-5p inhibits proliferation of hepatocellular carcinoma by targeting CDK1 and AKT3, Tumour Biol 36 (11) (2015) 8309–8316.
[18] Y.H. Hung, H.L. Huang, W.C. Chen, et al., Argininosuccinate lyase interacts with cyclin A2 in cytoplasm and modulates growth of liver tumor cells, Oncol. Rep. 37 (2) (2017) 969–978.
[19] F. Yang, J. Gong, G. Wang, P. Chen, L. Yang, Z. Wang, Waltonitone inhibits proliferation of hepatoma cells and tumorigenesis via FXR-miR-22-CCNA2 signaling pathway, Oncotarget 7 (46) (2016) 75165–75175.
[20] F. Yang, Y. Hu, H.X. Liu, Y.J. Wan, MiR-22-silenced cyclin A expression in colon and liver cancer cells is regulated by bile acid receptor, J. Biol. Chem. 290 (10) (2015) 6507–6515.
[21] R.M. de Voer, V.K.A. Geurts, R.D. Weren, et al., Germline mutations in the spindle assembly checkpoint genes BUB1 and BUB3 are risk factors for colorectal cancer, Gastroenterology 145 (3) (2013) 544–547.
[22] K. Takagi, Y. Miki, Y. Shibahara, et al., BUB1 immunolocalization in breast carcinoma: its nuclear localization as a potent prognostic factor of the patients, Horm Cancer 4 (2) (2013) 92–102.
[23] X. Zhang, X. Zhang, X. Li, et al., Connection between CDC20 expression and hepatocellular carcinoma prognosis, Med. Sci. Mon. Int. Med. J. Exp. Clin. Res. 27 (2021) e926760.
[24] L. Zhuang, Z. Yang, Z. Meng, Upregulation of BUB1B, CCNB1, CDC7, CDC20, and MCM3 in tumor tissues predicted worse overall survival and disease-free survival in hepatocellular carcinoma patients, BioMed Res. Int. 2018 (2018) 7897346.

[25] W. Sun, W.W. Dong, L.L. Mao, et al., Overexpression of p42.3 promotes cell growth and tumorigenicity in hepatocellular carcinoma, World J. Gastroenterol. 19 (19) (2013) 2913–2920.
[26] J. Gu, X. Liu, J. Li, Y. He, MicroRNA-144 inhibits cell proliferation, migration and invasion in human hepatocellular carcinoma by targeting CCNB1, Cancer Cell Int. 19 (2019) 15.
[27] M.F. de Resende, S. Vieira, L.T. Chinen, et al., Prognostication of prostate cancer based on TOP2A protein and gene assessment: TOP2A in prostate cancer, J. Transl. Med. 11 (2013) 36.
[28] N. Wong, W. Yeo, W.L. Wong, et al., TOP2A overexpression in hepatocellular carcinoma correlates with early age onset, shorter patients survival and chemoresistance, Int. J. Cancer 124 (3) (2009) 644–652.