

Identification and Analysis of the Key Genes associated with the Development from Liver Cirrhosis to Hepatocellular Carcinoma Based on Bioinformatics Analysis

Tianyu Li

Hebei Medical University <https://orcid.org/0000-0003-2706-2685>

Fang Ding

Hebei Medical University

Huimin Yan (✉ yanhm2538@163.com)

Hebei Medical University <https://orcid.org/0000-0003-3158-2536>

Primary research

Keywords: Hepatocellular carcinoma, Liver cirrhosis, Bioinformatics analysis, Differentially expressed genes, Prognosis, Survival analysis

Posted Date: February 24th, 2020

DOI: <https://doi.org/10.21203/rs.2.24361/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Hepatocellular carcinoma (HCC) is the most frequent primary liver tumor, and one of the most common malignant cancer with poor prognosis. Liver cirrhosis is the major risk factor for HCC. The aim of this study was to identify potential key genes associated with the development from liver cirrhosis to HCC and explore their potential mechanisms.

Methods: Four microarray datasets GSE17548, GSE63898, GSE25097 and GSE89377 were downloaded from the Gene Expression Omnibus database. A protein-protein interaction (PPI) network was constructed using the STRING database, and potential hub genes were screened using MCODE plug-in in Cytoscape software. The Oncomine database was used to verify the expression of differential genes in cirrhosis and HCC. In order to further verify those hub genes, the hierarchical cluster between normal and HCC tissues was constructed using the UCSC Cancer Genomics Browser. The UALCAN database was used to verify the difference of hub genes in normal and HCC tissues and in different tumor grades. Finally, the cBioPortal online platform was used to analyze the association between the expression of hub genes and prognosis in HCC.

Results: A total of 360 DEGs, including 280 downregulated and 80 upregulated genes, were identified. Gene ontology enrichment (GO) analysis showed that these DEGs were mainly enriched in monooxygenase activity, cofactor binding, and oxidoreductase activity (acting on CH-OH group of donors). The mainly enriched pathways were complement and coagulation cascades, prion diseases, and arachidonic acid metabolism. By extracting key modules from the PPI network, 16 hub genes were screened out. In the hierarchical cluster of hub genes between normal and HCC tissues, the results showed that the expression level of 16 hub genes in HCC tissues was significantly higher than that in normal tissues. In addition, expression level of the hub genes was significantly associated with the tumor grades. The survival analysis showed that six hub DEGs, including KIF20A, HMMR, RRM2, TPX2, TTK and UBE2C, were closely associated with the poor prognosis of HCC.

Conclusion: Our study discovered six novel potential genes associated with the development from liver cirrhosis to HCC. These key genes may be used as prognostic biomarkers and molecular therapeutic targets for HCC.

Background

Liver cancer is the sixth most common cancer and the fourth leading cause of cancer-related death worldwide in 2018, with about 841,000 new cases and 782,000 deaths annually^[1]. Hepatocellular carcinoma (HCC) is the main type of primary liver cancer, which accounts for 75%-85% of all cases^[2]. Despite the continuous development of novel treatment strategies, the 5-year survival rate of HCC is still very low, especially in advanced HCC^[3,4]. The main risk factors for HCC are chronic infection with hepatitis B virus or hepatitis C virus, heavy alcohol intake, and obesity^[5,6]. In the process from chronic inflammation to HCC, liver cirrhosis is recognized as a major step. The incidence of HCC is markedly

increased in the cirrhotic state compared with non-cirrhotic state, irrespective of the etiology of liver disease^[7]. Therefore, it is important to understand the precise mechanism involved into the development and progression from liver cirrhosis to HCC.

The study of differentially expressed genes (DEGs) in different states of the disease can infer the intrinsic relationship between genes and disease occurrence. The development of high-throughput microarray technology has provided an efficient tool to analyze gene expression profiles, which helps us better understand the general genetic changes and potential mechanisms of tumorigenesis^[8,9,10]. As more and more resultant data are updated in public databases, bioinformatics analysis basing on these data has been widely applied to identify novel targets associated with cancer, including HCC. Many studies are focused on exploring the difference of gene expression between HCC tissues and adjacent non-tumor tissues, regardless with or without cirrhosis. However, only a few studies have made a comprehensive comparison between HCC and liver cirrhosis^[11,12].

The aim of this study was to screen DEGs between liver cirrhosis and HCC tissue samples, and identify the key genes and pathways associated with HCC. In order to avoid the limitation due to the application of a single dataset, four microarray datasets were used. Furthermore, the potential value of the key genes in the prognosis of HCC was evaluated. The study will provide the help to understand the underlying molecular mechanisms of HCC carcinogenesis and progression.

Materials And Methods

Microarray data

Four gene expression profile datasets GSE17548, GSE25097, GSE63898, and GSE89377 were obtained from the NCBI GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). The data contained 552 HCC and 241 liver cirrhosis tissue samples. The information of samples included in each dataset were listed in Table 1. Perl was used to combine four gene expression datasets, and gene probes were converted into genesymbol. When one gene corresponds to multiple probes, the average value was taken. R package "sva" (<https://bioconductor.org/packages/release/bioc/html/sva.html>) was used to normalize the merged chip data.

Table 1 The sample size included in four Gene Expression Profile datasets

| Dataset ID | HCC | LC | total | Platform |
|------------|-----|-----|-------|---|
| GSE17548 | 17 | 20 | 37 | GPL570 [HG-U133 Plus 2] Affymetrix Human Genome U133 Plus 2.0 Array |
| GSE63898 | 227 | 169 | 396 | GPL13667 [HG-U219] Affymetrix Human Genome U219 Array |
| GSE25097 | 268 | 40 | 308 | GPL10687 Rosetta/Merck Human RSTA Affymetrix 1.0 microarray, Custom CDF |
| GSE89377 | 40 | 12 | 52 | GPL6947 Illumina HumanHT-12 V3.0 expression beadchip |

Identification of DEGs

The DEGs of the four gene expression datasets between HCC and liver cirrhosis samples were analyzed. The \log_2 fold-change (FC) and adjusted P -values were calculated by R package "limma"^[13]. Genes that fulfilled the criteria of $|\log_2FC| \geq 1$ and adjusted $P < 0.05$ were considered statistically significant and termed DEGs.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of DEGs

ClusterProfiler^[14] is an R software package that not only automates the process of biological term classification and gene cluster enrichment analysis, but also provides a visualization module for displaying analysis results. In the present study, the R package "ClusterProfiler" was used to identify and visualize the GO analysis and KEGG pathways enriched by DEGs.

PPI network construction and module analysis

Search Tool for the Retrieval of Interacting Genes Database (STRING) (<https://www.string-db.org/>) was used to assess protein-protein interaction (PPI) information. Confidence score > 0.7 was set as significant. In order to explore the relationship between DEGs, we converted the results visually by using Cytoscape^[15] V3.7.1. The most significant module in the PPI networks was identified using MCODE^[16]. The criteria for selection were as follows: MCODE scores > 5 , degree cut-off=2, node score cut-off=0.2, max depth=100 and K-score=2. Subsequently, the KEGG and GO analyses for genes in those modules were performed using "clusterProfiler" R package.

Hub gene selection and expression analysis of hub genes in TCGA

The most significant gene module was selected as the hub genes for analysis, with the MCODE score ≥ 15 . The biologic process of hub genes was analyzed by using BiNGO^[17] plug-in of Cytoscape. The difference between cirrhosis and HCC genes expression was analyzed using the Oncomine database (<https://www.oncomine.com>). To further identify key genes, normal and tumor tissues were compared based on TCGA database. Hierarchical clustering of hub genes was constructed using UCSC Cancer Genomics Browser (<http://genome-cancer.ucsc.edu>). UALCAN (<http://ualcan.path.uab.edu>) analysis was used to estimate the association between hub gene expression levels and tumour grades in the TCGA liver cancer datasets.

Kaplan–Meier survival analysis for hub genes in TCGA

The cBioPortal (<https://www.cbioportal.org>) is a comprehensive open network platform based on TCGA database, which integrates data mining, data integration and visualization. OncoPrint was used to identify and display recurrent amplifications and deletions. The prognosis values of the hub genes were analyzed by constructing Kaplan-Meier curves.

Results

DEGs between HCC and liver cirrhosis

A total of 360 DEGs, including 80 upregulated genes and 280 downregulated genes, were obtained (Table S1). A volcano plot of the DEGs was shown in Figure 1, and a heat map was shown in Figure S1.

GO and KEGG analyses of DEGs

GO enrichment analysis showed that the DEGs were mainly enriched in cofactor binding, glycosaminoglycan binding, extracellular matrix structural constituent, oxidoreductase activity, iron ion binding, and peptidase inhibitor activity (Figure 2 and Table 2). KEGG pathway analysis showed that the DEGs mainly concentrated in completion and coagulation cascades, prion diseases, chemical carcinogenesis, and viral protein interaction with cytokine and cytokine receptor (Figure 3 and Table 3).

Table 2 The top ten ID of Gene Ontology analysis

| ID | Description | Count | P value | Gene |
|------------|---|-------|----------|---|
| GO:0004497 | monooxygenase activity | 14 | 1.08E-09 | DBH, CH25H, CYP39A1, KMO, CYP1A2, CYP2C9, CYP26A1, SQLE, CYP2C8, CYP2C18, CYP3A4, CYP7A1, PTGIS, CYP4A11 |
| GO:0048037 | cofactor binding | 30 | 2.40E-09 | DBH, AADAT, CYP39A1, TP53I3, PTGS2, KMO, CYP1A2, CYP26A1, STEAP4, ETFDH, ZSWIM6, XDH, SQLE, ADH4, GCH1, PHGDH, GOT1, CYP3A4, HPGD, HMOX1, AMBP, CYP7A1, PTGIS, LCN2, CTH, TDO2, GRHPR, HAO1, SDS, CYP4A11 |
| GO:0016614 | oxidoreductase activity, acting on CH-OH group of donors | 15 | 1.17E-08 | LDHB, AKR7A3, AKR1D1, HAO2, ADH4, PHGDH, HPGD, ALDH3A1, AKR1B10, HSD17B2, RDH16, GRHPR, DHRS2, HAO1, HSD11B1 |
| GO:0005201 | extracellular matrix structural constituent | 17 | 1.24E-08 | SRPX, FGL2, THBS1, LAMA2, DPT, MFAP4, ASPN, FBLN5, LUM, MXRA5, COL15A1, LAMA3, EFEMP1, DCN, LTBP1, COL1A1, AEBP1 |
| GO:0005539 | glycosaminoglycan binding | 19 | 1.25E-08 | RSPO3, STAB2, ADAMTS1, THBS1, CXCL6, HMMR, ENG, LPA, PTPRC, LXN, MMP7, MDK, DCN, PRNP, THBS4, F11, PCOLCE2, APOA5, SERPINC1 |
| GO:0016705 | oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen | 15 | 7.15E-08 | DBH, CH25H, CYP39A1, PTGS2, KMO, CYP1A2, CYP26A1, SQLE, CYP2C8, CYP2C18, CYP3A4, HMOX1, CYP7A1, PTGIS, CYP4A11 |
| GO:0005506 | iron ion binding | 15 | 1.39E-07 | CH25H, CYP39A1, RRM2, CYP1A2, CYP2C9, CYP26A1, XDH, CYP2C8, CYP2C18, CYP3A4, CYP7A1, PTGIS, LCN2, MFAP4, CYP4A11 |
| GO:0016616 | oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor | 13 | 1.76E-07 | LDHB, AKR7A3, AKR1D1, ADH4, PHGDH, HPGD, ALDH3A1, AKR1B10, HSD17B2, RDH16, GRHPR, DHRS2, HSD11B1 |
| GO:0042379 | chemokine receptor binding | 9 | 2.54E-07 | CXCL14, CXCL12, CXCL6, CCL19, CCL2, DEFB1, CCL5, CXCL9, CCL21 |
| GO:0030414 | peptidase inhibitor activity | 15 | 4.32E-07 | LPA, SPINT2, LXN, SPP2, SERPINE1, TIMP2, AMBP, SPINK1, SERPINC1, FETUB, AHSG, ITIH1, SLPI, SERPINA3, GPC3 |

Table 3 The top ten ID of KEGG pathway analysis

| ID | Description | Count | pvalue | Gene |
|----------|---|-------|-------------|--|
| hsa04610 | Complement and coagulation cascades | 15 | 5.55E-09 | C7, THBD, F3, C1QA, C9, SERPINE1, VSIG4, C1QB, F11, C1QC, KLKB1, C4BPB, SERPINC1, C8B, PLG |
| hsa05020 | Prion diseases | 10 | 1.59E-08 | C7, FYN, C1QA, CCL5, PRNP, C9, C1QB, C1QC, EGR1, C8B |
| hsa00590 | Arachidonic acid metabolism | 9 | 4.06E-05 | GGT5, PTGS2, CYP2C9, PTGDS, CYP2C8, PTGIS, EPHX2, CYP4A11, GPX2 |
| hsa05204 | Chemical carcinogenesis | 10 | 6.15E-05 | NAT2, PTGS2, CYP1A2, CYP2C9, ADH4, CYP2C8, CYP2C18, CYP3A4, ALDH3A1, HSD11B1 |
| hsa00120 | Primary bile acid biosynthesis | 5 | 6.28E-05 | CH25H, CYP39A1, AKR1D1, CYP7A1, CYP8B1 |
| hsa00830 | Retinol metabolism | 9 | 6.69E-05 | CYP1A2, CYP2C9, CYP26A1, ADH4, CYP2C8, CYP2C18, CYP3A4, RDH16, CYP4A11 |
| hsa04115 | p53 signaling pathway | 9 | 0.00011864 | THBS1, CCNB1, GADD45B, CCNB2, TP53I3, RRM2, IGF1, GADD45G, SERPINE1 |
| hsa03320 | PPAR signaling pathway | 9 | 0.000181015 | GK, PCK1, ACSL1, CYP7A1, APOA5, HMGCS2, SLC27A2, CYP8B1, CYP4A11 |
| hsa00340 | Histidine metabolism | 5 | 0.000299521 | UROC1, CNDP1, HAL, ALDH3A1, FTCD |
| hsa04061 | Viral protein interaction with cytokine and cytokine receptor | 10 | 0.000327133 | CXCL14, CXCL12, CXCL6, CCL19, CCL2, CXCL1, CXCL2, CCL5, CXCL9, CCL21 |

PPI network construction and module analysis

To identify the core genes and crucial gene modules involved in HCC from the interaction level, Cytoscape software and STRING database were used. The combined score higher than 0.7 in PPI was used for constructing the PPI networks (Figure S2). A total of 359 DEGs of the 360 commonly altered DEGs were filtered into the DEG PPI network complex, which contained 359 nodes and 627 edges (Figure 4A). In addition, two significant modules (modules 1 and 2) with a score ≥ 8 were screened out via MCODE. The functional analyses of genes involved in those modules (Figure 4B and 4C) were analyzed using 'clusterProfiler', the results are shown in Table S2 and Table S3. In total, 16 genes were identified as hub genes, including CDKN3, ASPM, UBE2C, CENPF, TOP2A, TPX2, CCNB2, TTK, AURKA, NUSAP1, CCNB1, CCNA2, RRM2, KIF20A, NDC80 and HMMR. The names, abbreviations and functions for these hub genes were shown in Table 4. The biological process of the hub genes was analyzed and visualized using BiNGO and the result is shown in Figure 5, which mainly enriched in mitotic cell cycle, cell cycle phase and, M phase and nuclear division.

Table 4 Functional introduction of s 16 hub genes

| Gene symbol | Full name | Function |
|-------------|--|---|
| CDKN3 | cyclin dependent kinase inhibitor 3 | CDKN3 is a negative regulator of cell proliferation. |
| ASPM | abnormal spindle microtubule assembly | ASPM regulates mitotic duration and pathways through G1 restriction points, as well as mitotic spindle regulation and neural development. |
| UBE2C | ubiquitin conjugating enzyme E2 C | UBE2C is involved in mitotic cyclin B degradation, which promotes the transition from the M phase to the G1 phase of the cell cycle. |
| CENPF | centromere protein F | CENPF is an essential nuclear protein associated with the centromere-kinetochore complex and plays a critical role in chromosome segregation during mitosis. |
| TOP2A | DNA topoisomerase II alpha | TOP2A accumulates on chromatin during mitosis and targets mitotic centromere during prophase. |
| TPX2 | Targeting protein for Xklp2 | TPX2 is the key to spindle formation and peripheral microtubule nucleation. It is a proliferation-related protein involved in the regulation of cell mitosis and regulated by cell cycle. |
| CCNB2 | cyclin B2 | CCNB2 is mainly associated with the golgi region and binds to transforming growth factor RII. |
| TTK | threonine and tyrosine kinase | TTK protein kinase is most likely associated with cell proliferation and is essential for the accurate segregation of chromosomes in cellular mitosis. |
| AURKA | aurora kinase A | AURKA plays a critical role in centrosome duplication and separation, spindle assembly, maturation, chromosomal alignment, spindle assembly checkpoint, and cytokinesis. |
| NUSAP1 | nucleolar and spindle associated protein 1 | NUSAP1 is a nucleolar-spindle-associated protein that plays a role in spindle microtubule organization |
| CCNB1 | cyclin B1 | CCNB1, a regulatory protein involved in mitosis, is an important modulator of cell cycle progression. |
| CCNA2 | cyclin A2 | CCNA2 Encodes a protein that binds and activates cyclin-dependent kinase 2 and promotes transition through G1/S and G2/M. |
| RRM2 | ribonucleotide reductase regulatory subunit M2 | RRM2 plays an essential role in ribonucleotide reduction. |
| KIF20A | kinesin family member 20A | KIF20A is essential for cell cycle mitosis and its accumulation promotes the proliferation of both normal and pathological cells. |
| NDC80 | nuclear division cycle 80 | NDC80 is a kinetochore complex component that can organize and stabilize microtubule-kinetochore interactions. |
| HMMR | hyaluronan mediated motility receptor | HMMR is a multifunctional oncogenic protein, the overexpression of which is transforming and essential for maintaining H-ras-mediated transformation. |

The expression levels of hub genes in TCGA

The analysis of Wurbach liver dataset from the Oncomine database showed that the mRNA levels of 15 genes except CENPF in HCC were significantly higher than those of liver cirrhosis (Figure 6). The expression levels of 16 hub genes were also compared between HCC and normal tissues using the UCSC Cancer Genomics Browser, revealing that these hub genes were highly expressed in most HCC samples (Figure 7). These results were further confirmed using UALCAN database (Figure 8). Further subgroup analysis demonstrated that the expression levels of the 16 hub genes were significantly associated with tumor grades (Figure 9).

Kaplan–Meier survival analysis for hub genes in TCGA

OncoPrint showed that the genetic alterations of 16 hub genes were found in 174 of 371 HCC patients (47%) (Figure 10A). The proportion of alterations for individual genes varied between 4% and 21%, and ASPM had the highest level of amplification in HCC. (Figure 10C). Subsequently, the prognostic analyses of the 16 hub genes were performed in the HCC datasets of the cBioPortal online platform. The results showed that HCC patients with CCNB1, CCNB2, HMMR, KIF20A, NDC80, RRM2, TPX2, TTK and UBE2C alteration had worse overall survival (Figure 11A), whereas the patients with KIF20A, RRM2, TPX2, TTK, UBE2C, HMMR, ASPM and NUSAP1 alteration exhibited worse disease-free survival (Figure 11B).

Discussion

The occurrence of HCC is an extremely complicated process. Liver cirrhosis is present in 80%-90% of HCC patients and represents a relevant risk for the development of HCC^[18]. However, the molecular mechanisms underlying the progression from liver cirrhosis to HCC remains unclear. In the present study, we performed a series of bioinformatics analysis to screen key genes and pathways closely related to HCC using four GEO databases. Although there were two studies identifying DEGs between HCC and cirrhosis^[11,12], their studies only used one dataset, and contained small sample sizes. The present study utilized multi-chip joint analysis to provide more reliable and accurate assessment. Our study identified a total of 360 differential genes, including 80 up-regulated and 280 down-regulated genes. Furthermore, GO enrichment analysis indicated that aberrant DEGs were predominantly involved into cofactor binding, glycosaminoglycan binding, extra cellular matrix structural constituent, oxidoreductase activity (acting on CH- OH group of donors), oxidoreductase activity (acting on paired donors, with incorporation or reduction of molecular oxygen), iron ion binding, peptidase inhibitor activity. KEGG pathway analysis revealed that DEGs were enriched in multiple pathways, including complement and coagulation cascade pathway, prion diseases, chemical carcinogenesis, and viral protein. interaction with cytokine and cytokine receptor p53 signaling pathway. The results suggest that these pathways may be important HCC-associated oncogenic pathways.

According to the centrality of nodes in the PPI network, we identified the crucial DEGs, including CCNB1, CCNB2, CCNA2, CDKN3, UBE2C, RRM2, KIF20A, TPX2, TTK, AURKA, CENPF, ASPM, TOP2A, NUSAP1, NDC80 and HMMR. The expression level of these crucial genes was further validated based on TCGA

data, and the results showed that HCC patients had significantly higher expression than the patients with liver cirrhosis. Similar results were found when comparing HCC patients and adjacent normal tissues. Furthermore, the expression of these genes was positively correlated with tumor grades. Survival analysis based on cBioPortal database showed that high expression of KIF20A, HMMR, RRM2, TPX2, TTK and UBE2C was associated with worse overall survival and disease-free survival. The results suggest that these genes may be involved in HCC pathogenesis, and provide clues to explore the molecular mechanism for future investigation.

KIF20A is a member of the kinesin family protein, which participates in spindle assembly during mitosis^[19,20]. Numerous studies have shown that KIF20A promotes the proliferation, invasion and migration of cancer cells, and the high expression of KIF20A is significantly related to the occurrence, migration and prognosis of various tumors, such as glioma^[21], gastric cancer^[22], ovarian cancer^[23], and pancreatic cancer^[24]. The prognostic value of KIF20A for HCC is also demonstrated^[25]. Mechanistically, KIF20A was found to serve as a novel downstream target of glioma-associated oncogene 2 (Gli2), a major transcriptional regulator of hedgehog (Hh) signaling. Gli2 could directly activate the transcription of FoxM1 in response to Hh signaling, which in turn increase KIF20A expression by activating FoxM1-MMB complex^[26]. Therefore, KIF20A could have potential as therapeutic targets for HCC.

RRM2 is essential enzyme in DNA synthesis and replication. Many studies have shown that high expression of RRM2 significantly promotes the growth, invasion and resistance of cancer cells^[27]. There is no direct evidence to demonstrate the role of RRM2 in the pathogenesis and progression of HCC. However, a study by Ricardo-Lax *et al.* revealed that RRM2 is essential for HBV replication^[28]. HBV induced RRM2 expression by exploiting the Chk1-E2F1 axis of the DNA damage response pathway. Considering chronic HBV infection is a key factor for HCC, inhibition of RRM2 may have a therapeutic value for HBV-related HCC.

TPX2 level is elevated in many cancers and has been proposed as biomarkers and effectors of cancer progression. TPX2 can disrupt DNA damage responses and promote cancer pathology through the regulation of Ser-139-phosphorylated Histone 2AX signals^[29]. In HCC cells, TPX2 expression is associated with proliferation, apoptosis and EMT^[30]. Recent research suggests that TPX2 expression is positively correlated with MMP2 and MMP9 in HCC tissues^[31]. The down-regulation of TPX2 can result in inactivation of AKT signaling and decrease the expression of MMP2 and MMP9, which reduces the migration and invasion ability of HCC cells.

TTK, a dual-specificity protein kinase, redirects several key proteins to kinetochores and controls mitotic spindle checkpoint^[32]. It has been found to be aberrantly overexpressed in a wide range of human tumors. In HCC, increased TTK expression contributes to HCC tumorigenesis via promoting cell proliferation and migration. Mechanistic studies reveal that TTK stimulates the malignancy of HCC cells through the activation of Akt/mTOR and MDM2/p53 signaling pathways^[33].

UBE2C encodes a member of the E2 ubiquitin-conjugating enzyme family, and is required for the destruction of mitotic cyclins and cell cycle progression. UBE2C is nearly undetectable in normal tissues, but it is upregulated in some human cancers, such as lung^[34], colon^[35], breast^[36], nasopharyngeal carcinoma^[37]. A recent study^[38] has shown that UBE2C is a transcription target of FOXM1, a master regulator of cell cycle progression. The transcriptional activation of UBE2C by FOXM1 leads to increased level of UBE2C protein, and thereby contributes to the loss of G2/M checkpoint control and cell proliferation. Cells overexpressing UBE2C display the mitotic spindle checkpoint inactivation and lose genomic stability^[39].

HMMR is a multifunctional oncogenic protein participating in cell division event. The high expression of HMMR may be related to cancer progression and prognosis. In breast cancer, it promotes cancer cell migration and invasion^[40]. However, few studies have reported on HCC. The results of our study showed that HMMR was overexpressed in HCC patients, and increased expression was significantly associated with overall survival and disease-free survival.

In addition to the genes we have discussed above, five genes was found to be associated only with overall survival or disease-free survival. CCNB1 and CCNB2 belong to the cyclin family, which can form complexes with cyclin-dependent kinases to regulate cell-cycle progression^[41]. Abnormalities of CCNB1 and CCNB2 could lead to the development of malignant tumors. NDC80 is a core component of the outer kinetochore and a mitotic regulator. It has been shown that NDC80 overexpression contributes to HCC progression via the inhibition of apoptosis and cell cycle arrest^[42]. ASPM is essential for mitotic spindle function during cell replication. Overexpressed ASPM has been found to be associated with tumor progression, early tumor recurrence, and poor prognosis in human HCC^[43]. NUSAP1 is a microtubule-binding protein implicated in spindle stability and chromosome segregation. Many studies implicate a crucial role for NUSAP1 in regulating mitotic processes^[44,45]. The overexpression of NUSAP1 is often associated with tumor recurrence and metastasis.

Among the other five genes, CCNA2 also belongs to the cyclin family. However, different with CCNB1 and CCNB2, we found that CCNA2 expression was upregulated in HCC patients, but not associated with survival. CDKN3 is a cell-cycle regulator, but is displays tumor-suppressive or oncogenic role depending on the molecular background of different cancer types. In HCC, CDKN3 is commonly overexpressed and associated with poor outcome^[46]. Both CENPF and TOP2A play important roles in chromosomal segregation during mitosis^[47]. CENPF has been reported to interact synergistically with FOXM1 to promote tumor growth^[48]. It has been reported that the overexpression of CENPF and TOP2A is associated with poor prognosis in cancers, including HCC^[49,50]. AURKA has been confirmed as an oncogene in cancer development, which promotes tumor development by promoting a variety of biological functions, including cell proliferation, migration, invasion, EMT and cancer stem cell behaviors^[51,52].

Conclusion

In the present study, we performed an integrated bioinformatics analysis to identify key genes and pathways involved in the development from liver cirrhosis to HCC. Sixteen genes were considered as hub genes. The survival analysis showed that the overexpression of 6 hub genes (KIF20A, HMMR, RRM2, TPX2, TTK and UBE2C) was closely associated with the poor prognosis of HCC. Our results provide new insight into understanding the molecular mechanism underlying HCC carcinogenesis. Further experimental validation should be carried out to confirm these findings in the future.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TYL and HMY contributed to the study concept and design, the acquisition, analysis, and interpretation of data, and the drafting of the manuscript. DF contributed to the data collections . All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Founding

Not applicable.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394-424.
2. Petrick JL, Kelly SP, Altekruse SF, McGlynn KA, Rosenberg PS. Future of hepatocellular carcinoma incidence in the United States forecast through 2030. *J Clin Oncol.* 2016;34(15):1787-1794.
3. Sarveazad A, Agah S, Babahajian A, Amini N, Bahardoust M. Predictors of 5 year survival rate in hepatocellular carcinoma patients. *J Res Med Sci.* 2019;24:86.
4. Massarweh NN, El-Serag HB. Epidemiology of hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *Cancer Control.* 2017;24(3):1073274817729245.
5. Marengo A, Rosso C, Bugianesi E. Liver Cancer: Connections with Obesity, Fatty Liver, and Cirrhosis. *Annu Rev Med.* 2016;67:103-117.
6. Ghouri YA, Mian I, Rowe JH. Review of hepatocellular carcinoma: Epidemiology, etiology, and carcinogenesis. *J Carcinog.* 2017;16:1.
7. Tarao K, Nozaki A, Ikeda T, Sato A, Komatsu H, Komatsu T, Taguri M, Tanaka K. Real impact of liver cirrhosis on the development of hepatocellular carcinoma in various liver diseases - meta-analytic assessment. *Cancer Med.* 2019;8(3):1054-106.
8. Brazma A, Vilo J. Gene expression data analysis. *FEBS Lett.* 2000;480(1):17-24.
9. Sobek J, Bartscherer K, Jacob A, Hoheisel JD, Angenendt P. Microarray technology as a universal tool for high-throughput analysis of biological systems. *Comb Chem High Throughput Screen.* 2006;9(5):365-380.
10. Chen CF, Hsu EC, Lin KT, Tu PH, Chang HW, Lin CH, Chen YJ, Gu DL, Lin CH, Wu JY, Chen YT, Hsu MT, Jou YS. Overlapping high-resolution copy number alterations in cancer genomes identified putative cancer genes in hepatocellular carcinoma. *Hepatology.* 2010;52:1690–701.
11. He B, Yin J, Gong S, Gu J, Xiao J, Shi W, Ding W, He Y. Bioinformatics analysis of key genes and pathways for hepatocellular carcinoma transformed from cirrhosis. *Medicine (Baltimore).* 2017;96(25):e6938.
12. Wu M, Liu Z, Li X, Zhang A, Lin D, Li N. Analysis of potential key genes in very early hepatocellular carcinoma. *World J Surg Oncol.* 2019;17(1):77.
13. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
14. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284-287.
15. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-2504.

16. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *Bmc Bioinformatics*. 2003;4:2.
17. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*. 2005;21(16):3448-3449.
18. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology*. 2007;132:2557–2576.
19. Yu Y, Feng YM. The role of kinesin family proteins in tumorigenesis and progression: potential biomarkers and molecular targets for cancer therapy. *Cancer*. 2010;116(22):5150-5160.
20. Cesario JM, Jang JK, Redding B, Shah N, Rahman T, McKim KS. Kinesin 6 family member Subito participates in mitotic spindle assembly and interacts with mitotic regulators. *J Cell Sci*. 2006;119(Pt 22):4770-4780.
21. Saito K, Ohta S, Kawakami Y, Yoshida K, Toda M. Functional analysis of KIF20A, a potential immunotherapeutic target for glioma. *J Neurooncol*. 2017;132(1):63-74.
22. Sheng Y, Wang W, Hong B, Jiang X, Sun R, Yan Q, Zhang S, Lu M, Wang S, Zhang Z, Lin W, Li Y. Upregulation of KIF20A correlates with poor prognosis in gastric cancer. *Cancer Manag Res*. 2018;10:6205-6216.
23. Li H, Zhang W, Sun X, Chen J, Li Y, Niu C, Xu B, Zhang Y. Overexpression of kinesin family member 20A is associated with unfavorable clinical outcome and tumor progression in epithelial ovarian cancer. *Cancer Manag Res*. 2018;10:3433-3450.
24. Stangel D, Erkan M, Buchholz M, Gress T, Michalski C, Raulefs S, Friess H, Kleeff J. Kif20a inhibition reduces migration and invasion of pancreatic cancer cells. *J Surg Res*. 2015;197(1):91-100.
25. Lu M, Huang X, Chen Y, Fu Y, Xu C, Xiang W, Li C, Zhang S, Yu C. Aberrant KIF20A expression might independently predict poor overall survival and recurrence-free survival of hepatocellular carcinoma. *IUBMB Life*. 2018;70(4):328-335.
26. Chao Shi, Dengliang Huang, Nonghua Lu, Chen D, Zhang M, Yan Y, Deng L, Lu Q, Lu H, Luo S. Aberrantly activated Gli2-KIF20A axis is crucial for growth of hepatocellular carcinoma and predicts poor prognosis. *Oncotarget*. 2016;7(18):26206-26219.
27. Zhang K, Hu S, Wu J, Chen L, Lu J, Wang X, Liu X, Zhou B, Yen Y. Overexpression of RRM2 decreases thrombospondin-1 and increases VEGF production in human cancer cells in vitro and in vivo: implication of RRM2 in angiogenesis. *Mol Cancer*. 2009;8:11. <https://doi.org/10.1186/1476-4598-8-11>
28. Ricardo-Lax I, Ramanan V, Michailidis E, Shamia T, Reuven N, Rice CM, Shlomai A, Shaul Y. Hepatitis B virus induces RNR-R2 expression via DNA damage response activation. *J Hepatol*. 2015;63(4):789-796.
29. Neumayer G, Helfricht A, Shim SY, Le HT, Lundin C, Belzil C, Chansard M, Yu Y, Lees-Miller SP, Gruss OJ, van Attikum H, Helleday T, Nguyen MD. Targeting Protein for Xenopus Kinesin-like Protein 2 (TPX2) Regulates γ -Histone 2AX (γ -H2AX) Levels upon Ionizing Radiation. *J Biol Chem*. 2012;287(50):42206-42222.

30. Liang B, Jia C, Huang Y, He H, Li J, Liao H, Liu X, Liu X, Bai X, Yang D. TPX2 Level Correlates with Hepatocellular Carcinoma Cell Proliferation, Apoptosis, and EMT. *Dig Dis Sci*. 2015;60(8):2360-2372.
31. Liu Q, Tu K, Zhang H, Zheng X, Yao Y, Liu Q. TPX2 as a novel prognostic biomarker for hepatocellular carcinoma. *Hepatol Res*. 2015;45(8):906-918.
32. Anthony Tighe, Oliver Staples, Stephen Taylor. Mps1 kinase activity restrains anaphase during an unperturbed mitosis and targets Mad2 to kinetochores. *J Cell Biol*. 2008;181(6):893-901.
33. Liu X, Liao W, Yuan Q, Ou Y, Huang J. TTK activates Akt and promotes proliferation and migration of hepatocellular carcinoma cells. *Oncotarget*. 2015;6(33):34309-34320.
34. Perrotta I, Bruno L, Maltese L, Russo E, Donato A, Donato G. Immunohistochemical Analysis of the Ubiquitin-conjugating Enzyme UbcH10 in Lung Cancer. *J Histochem Cytochem*. 2012;60(5):359-365.
35. Fujita T, Ikeda H, Taira N, Hatoh S, Naito M, Doihara H. Overexpression of UbcH10 alternates the cell cycle profile and accelerate the tumor proliferation in colon cancer. *BMC Cancer*. 2009;9:87.
36. Berlingieri M.T, Pallante P, Sboner A, Barbareschi M, Bianco M, Ferraro A, Mansueto G, Borbone E, Guerriero E, Troncone G, Fusco A. UbcH10 is overexpressed in malignant breast carcinomas. *Eur J Cancer*. 2007;43:2729–2735.
37. Shen Z, Jiang X, Zeng C, Zheng S, Luo B, Zeng Y, Ding R, Jiang H, He Q, Guo J, Jie W. High expression of ubiquitin-conjugating enzyme 2c (UBE2C) correlates with nasopharyngeal carcinoma progression. *BMC Cancer*. 2013;13:192.
38. Nicolau-Neto P, Palumbo A, De Martino M, Esposito F, de Almeida Simão T, Fusco A, Nasciutti LE, Meireles Da Costa N, Ribeiro Pinto LF. UBE2C Is a Transcriptional Target of the Cell Cycle Regulator FOXM1. *Genes (Basel)*. 2018;9(4):188.
39. Hao Z, Zhang H, Cowell J. Ubiquitin-conjugating enzyme UBE2C: molecular biology, role in tumorigenesis, and potential as a biomarker. *Tumour Biol*. 2012;33(3):723-730.
40. Schwertfeger KL, Cowman MK, Telmer PG, Turley EA, McCarthy JB. Hyaluronan, Inflammation, and Breast Cancer Progression. *Front Immunol*. 2015; 6: 236.
41. Li J, Qian WP, Sun QY. Cyclins regulating oocyte meiotic cell cycle progression[†]. *Biol Reprod*. 2019;101(5):878-881.
42. Ju LL, Chen L, Li JH. Effect of NDC80 in human hepatocellular carcinoma. *World J Gastroenterol*. 2017;23(20):3675-3683.
43. Lin SY, Pan HW, Liu SH, Jeng YM, Hu FC, Peng SY, Lai PL, Hsu HC. ASPM is a novel marker for vascular invasion, early recurrence, and poor prognosis of hepatocellular carcinoma. *Clin Cancer Res*. 2008;14(15):4814-4820.
44. Tim Raemaekers, Katharina Ribbeck, Joël Beaudouin, Annaert W, Van Camp M, Stockmans I, Smets N, Bouillon R, Ellenberg J, Carmeliet G. NuSAP, a novel microtubule-associated protein involved in mitotic spindle organization. *J Cell Biol*. 2003;162(6):1017-1029.
45. Ribbeck K, Raemaekers T, Carmeliet G, Mattaj IW. A role for NuSAP in linking microtubules to mitotic chromosomes. *Curr Biol*. 2007;17(3):230-236.

46. Xing C, Xie H, Lin Z, Zhou W, Zhang W, Ding S, Wei B, Yu X, Su R, Zheng S. Cyclin-dependent kinase inhibitor 3 is overexpressed in hepatocellular carcinoma and promotes tumor cell proliferation. *Biochem Biophys Res Commun.* 2012;420(1):29–35.
47. Shahid M, Lee MY, Piplani H, Andres AM, Zhou B, Yeon A, Kim M, Kim HL, Kim J. Centromere protein F (CENPF), a microtubule binding protein, modulates cancer metabolism by regulating pyruvate kinase M2 phosphorylation signaling. *Cell Cycle.* 2018;17(24):2802-2818.
48. Aytes A, Mitrofanova A, Lefebvre C, Mitrofanova A, Lefebvre C, Alvarez MJ, Castillo-Martin M, Zheng T, Eastham JA, Gopalan A, Pienta KJ, Shen MM1, Califano A, Abate-Shen C. Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell.* 2014;25(5):638-651.
49. Yang X, Miao BS, Wei CY, Dong RZ, Gao PT, Zhang XY, Lu JC, Gao C, Wang XY, Sun HC, Zhou J, Fan J, Ke AW, Shi GM, Cai JB. Lymphoid-specific helicase promotes the growth and invasion of hepatocellular carcinoma by transcriptional regulation of centromere protein F expression. *Cancer Sci.* 2019;110(7):2133-2144.
50. Panvichian R, Tantiwetrueangdet A, Angkathunyakul N, Leelaudomlipi S. TOP2A amplification and overexpression in hepatocellular carcinoma tissues. *Biomed Res Int.* 2015:381602.
51. Wu J, Yang L, Shan Y, Cai C, Wang S, Zhang H. AURKA promotes cell migration and invasion of head and neck squamous cell carcinoma through regulation of the AURKA/Akt/FAK signaling pathway. *Oncol Lett.* 2016;11(3):1889-1894.
52. Yang L, Zhou Q, Chen X, Su L, Liu B, Zhang H. Activation of the FAK/PI3K pathway is crucial for AURKA-induced epithelial-mesenchymal transition in laryngeal cancer. *Oncol Rep.* 2016;36(2):819-826.

Figures

Volcano

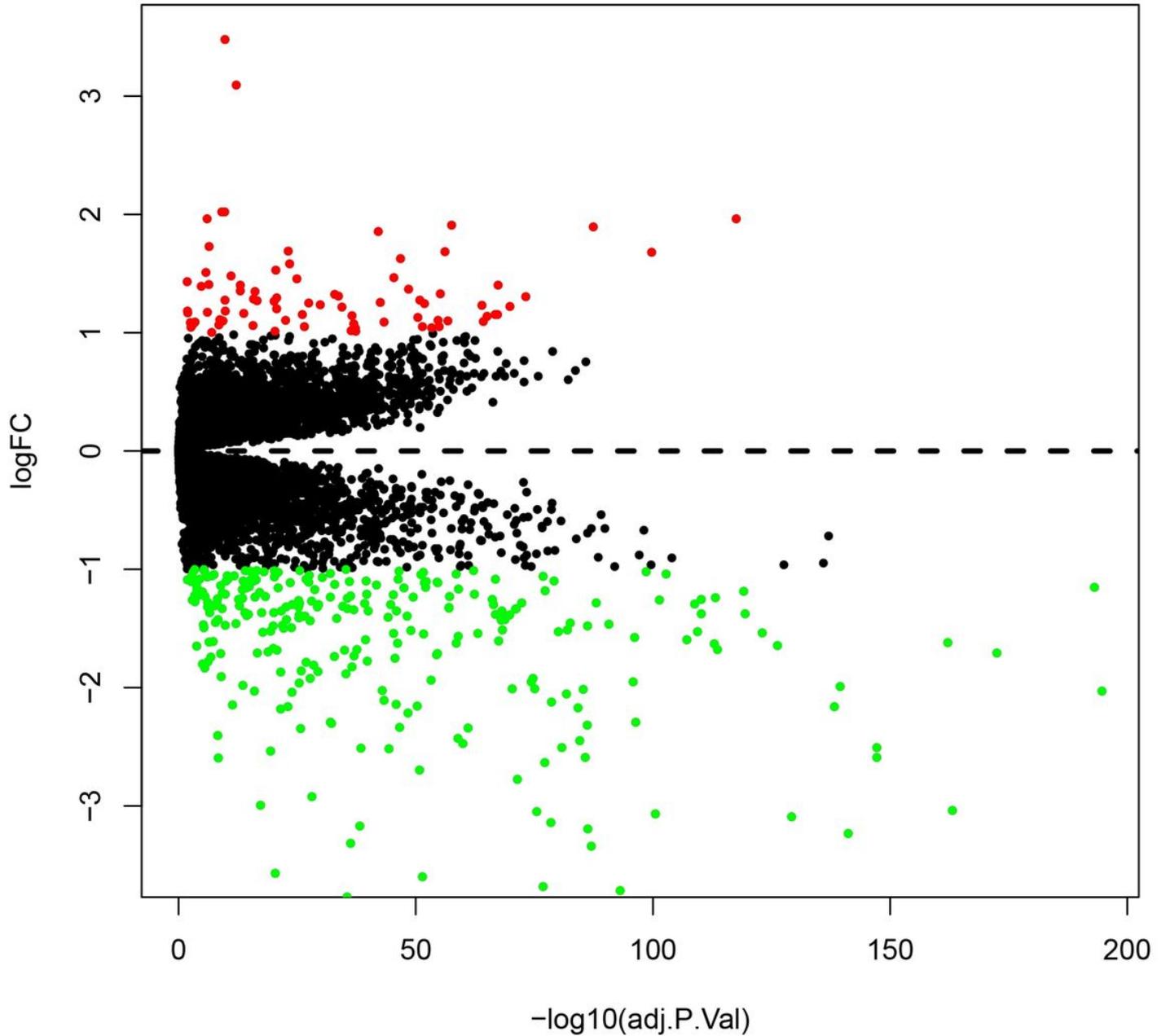


Figure 1

Volcano plot of DEGs between HCC and liver cirrhosis. The DEGs were selected with a $|\log_2\text{FC}| \geq 1$ and adjusted $P < 0.05$ among the expression profiling sets. The upper-right red dots represented upregulated genes, and the upper-left green dots represented down-regulated genes.

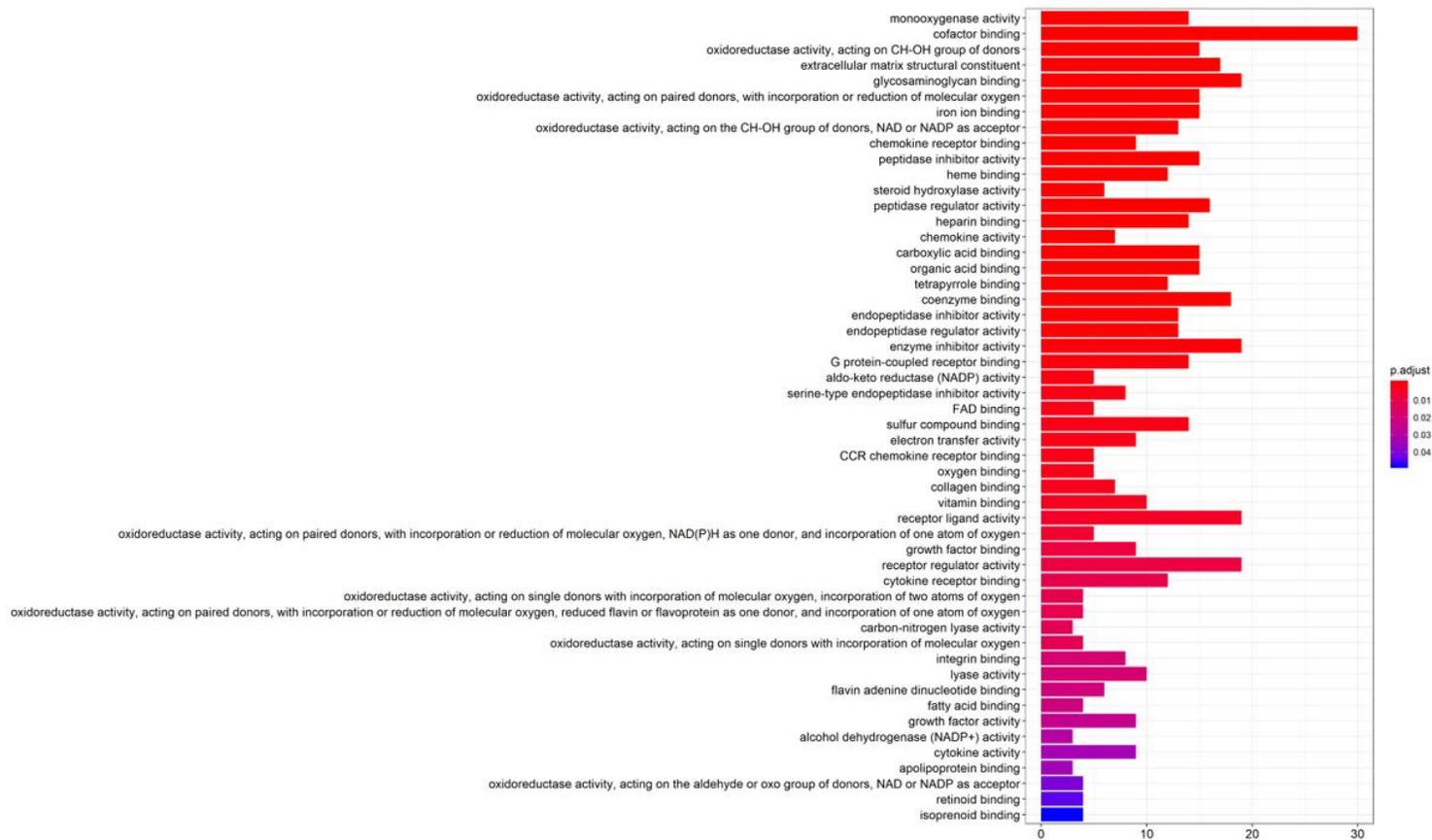


Figure 2

GO analysis of DEGs.

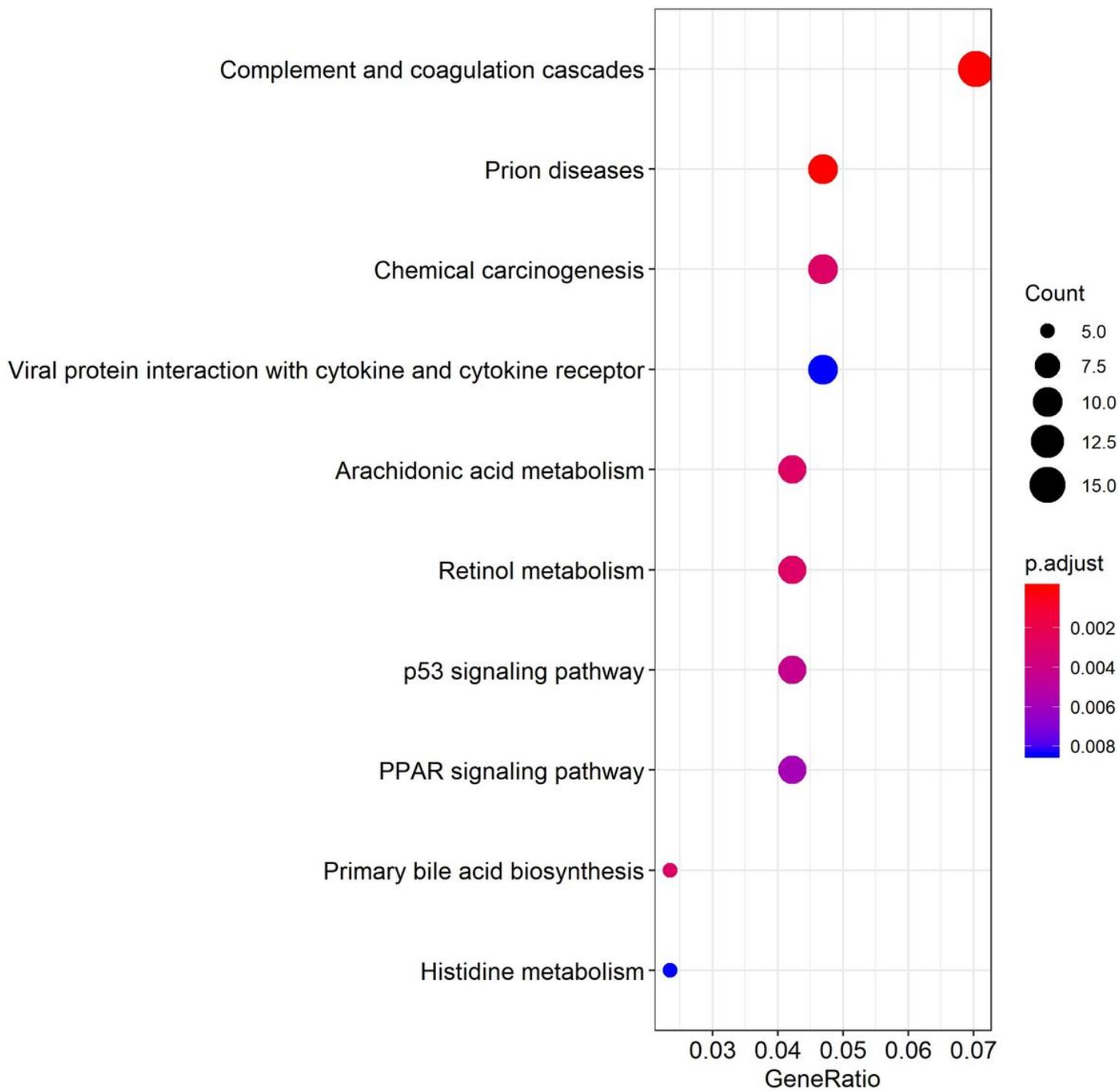


Figure 3

KEGG pathway of DEGs.

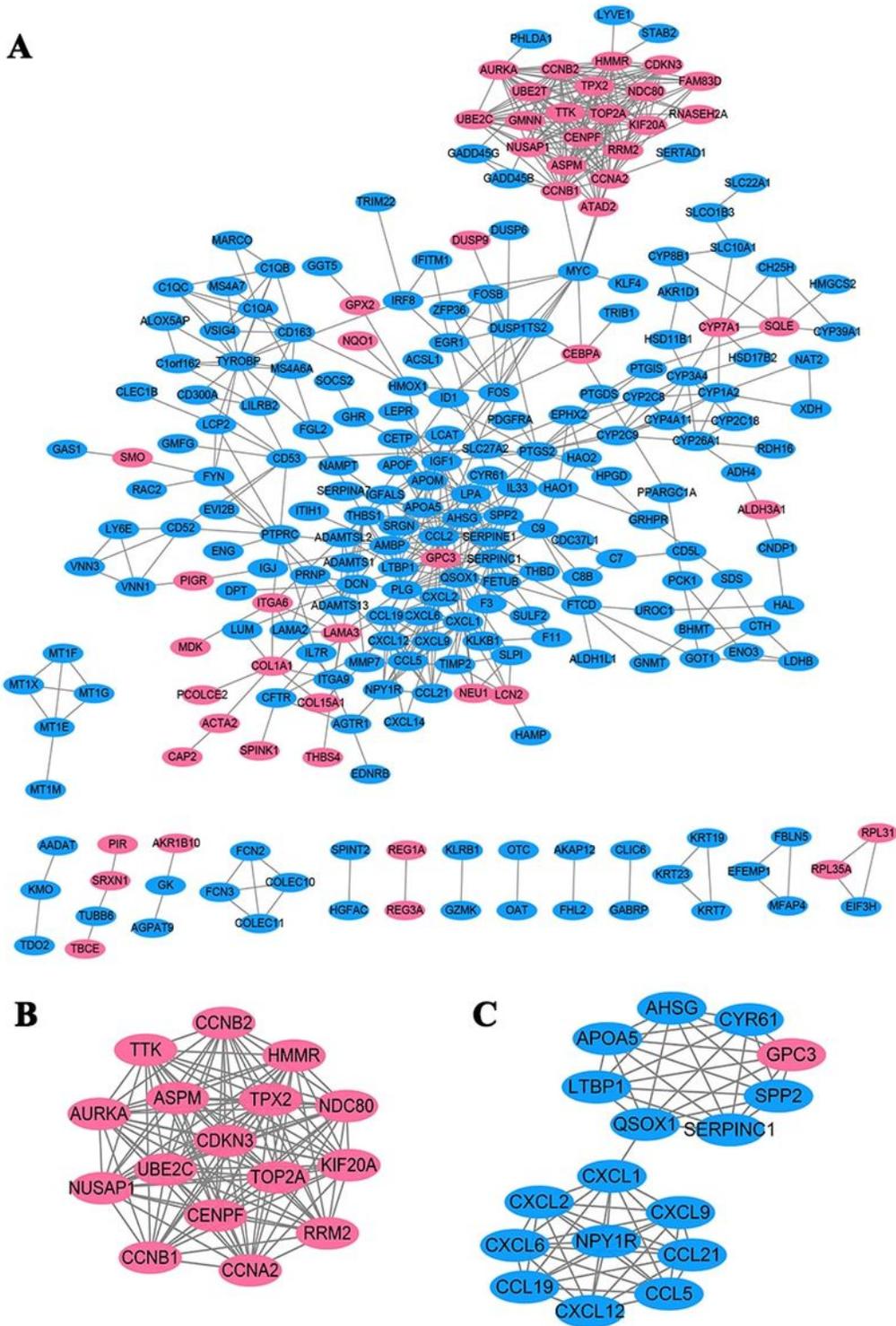


Figure 4

The PPI networks were constructed using Cytoscape. (A) The PPI network of 360 DEGs. Up-regulated genes were marked in light red, down-regulated genes were marked in blue. (B) The most significant module was obtained from PPI network with 16 nodes and 120 edges. (C) The secondary modules of PPI network.

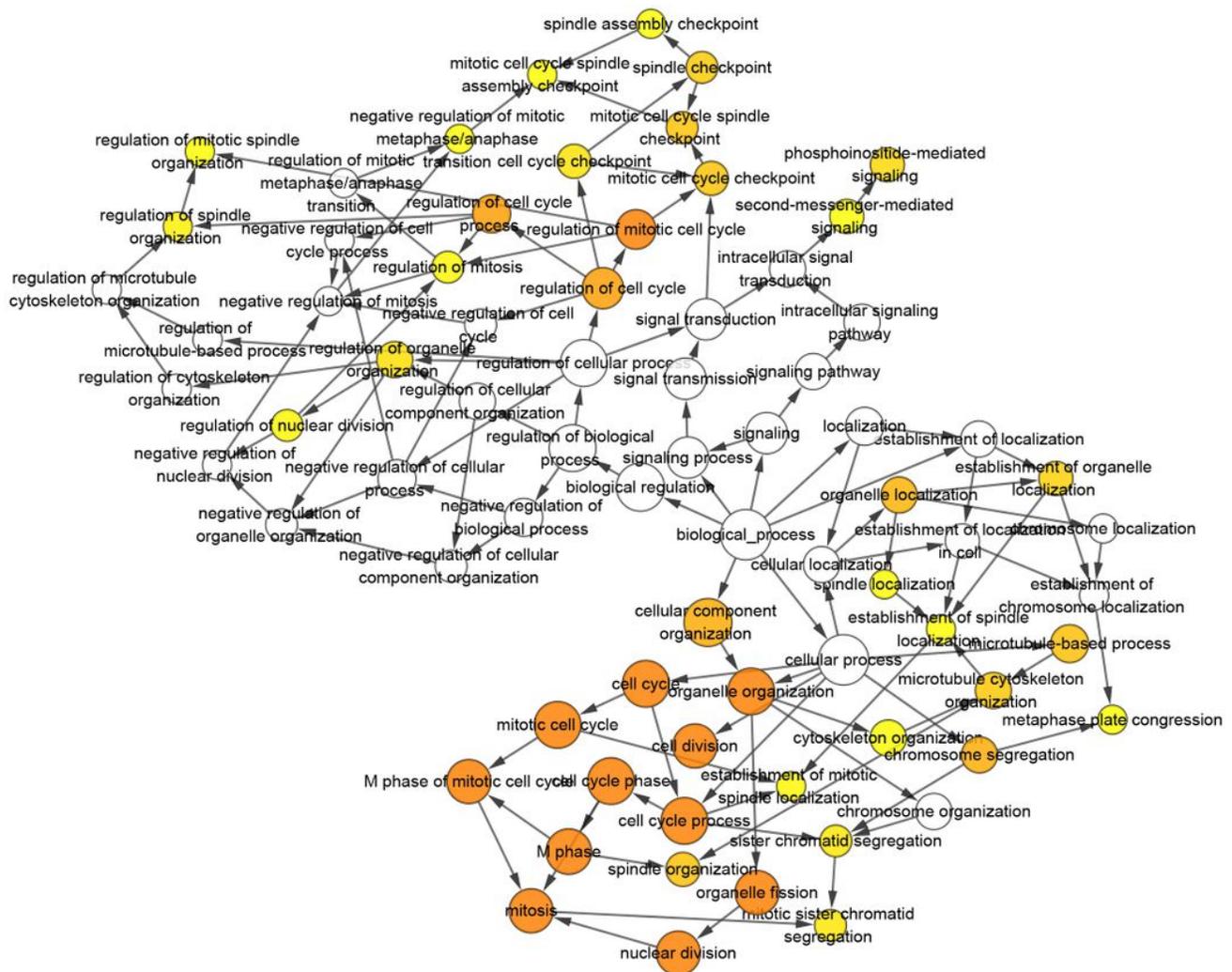


Figure 5

The biological process analysis of hub genes was constructed using BiNGO. The color depth of nodes refers to the corrected P-value of ontologies. The size of nodes refers to the numbers of genes that are involved in the ontologies. $P < 0.001$ was considered statistically significant.

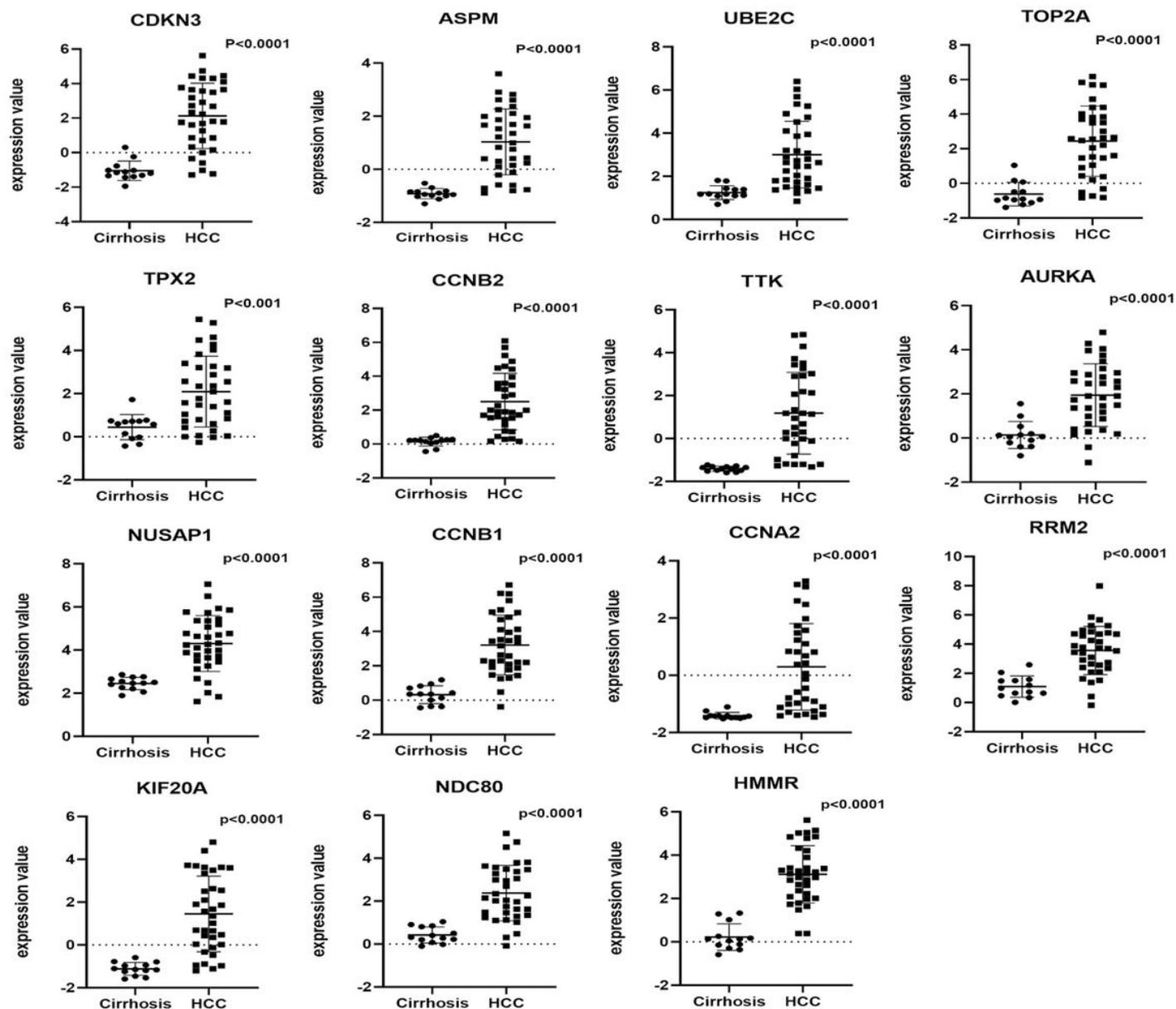


Figure 6

Comparison of hub genes expression between HCC and cirrhosis tissues. mRNA expression of hub genes in HCC and cirrhosis tissues using the Wurmbach Liver dataset based on oncomine database. $P < 0.001$ was considered statistically significant.

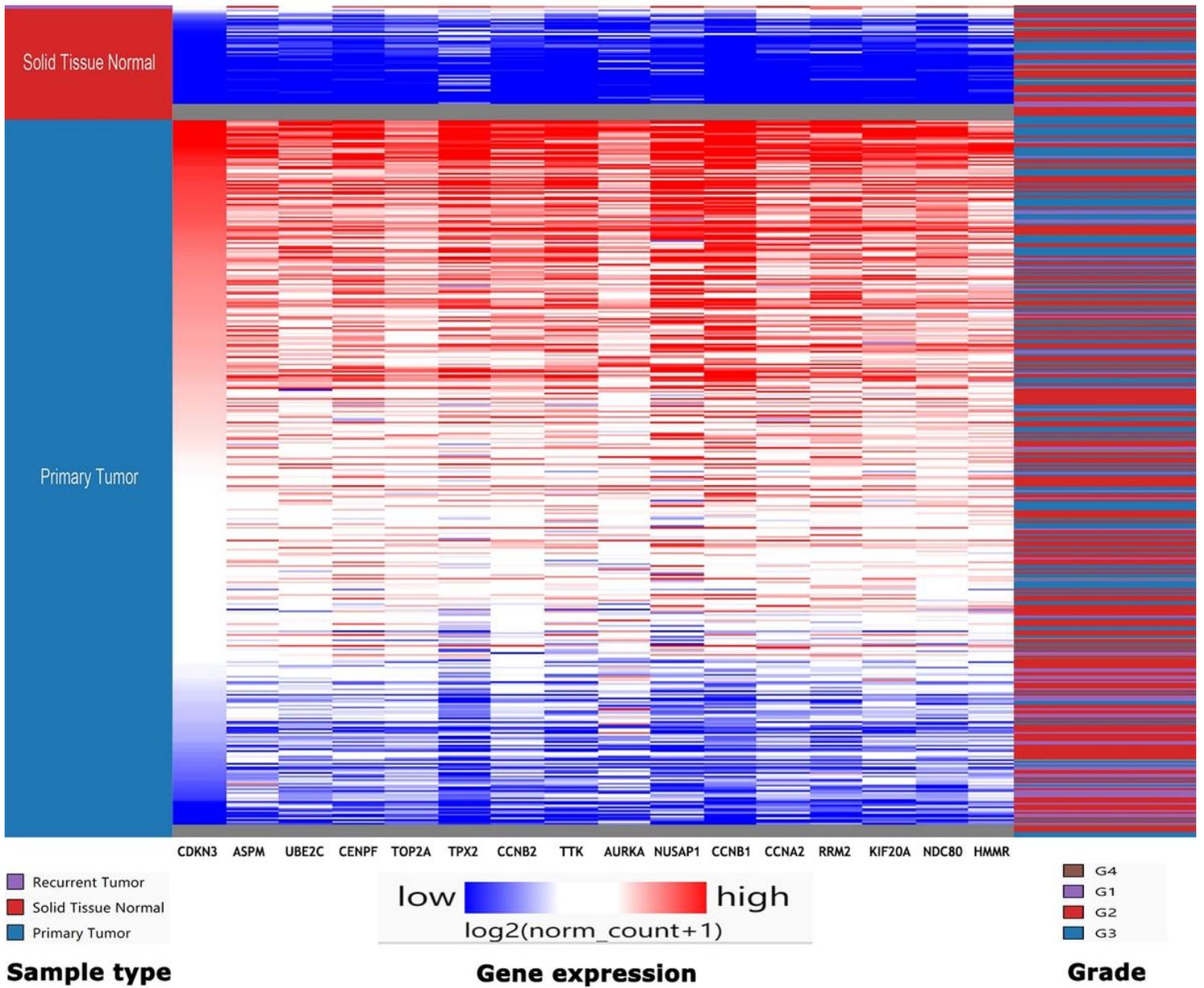


Figure 7

Hierarchical clustering of hub genes was constructed using UCSC. Up-regulation of genes was marked in red and down-regulation of genes was marked in blue.

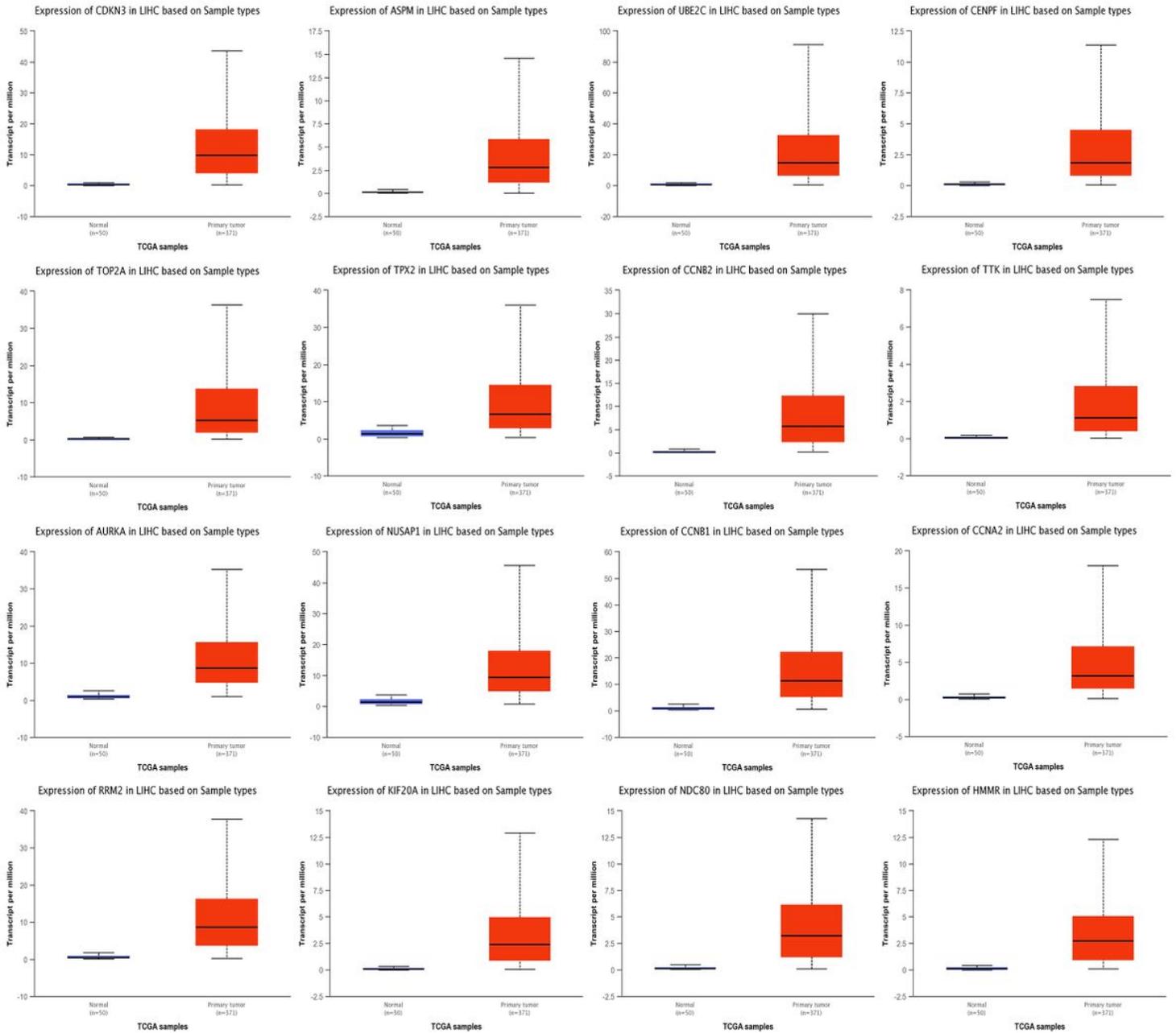


Figure 8

The expression of hub genes in HCC and normal tissues was constructed using UALCAN online database. All of hub genes are different. $P < 0.01$ was considered statistically significant.

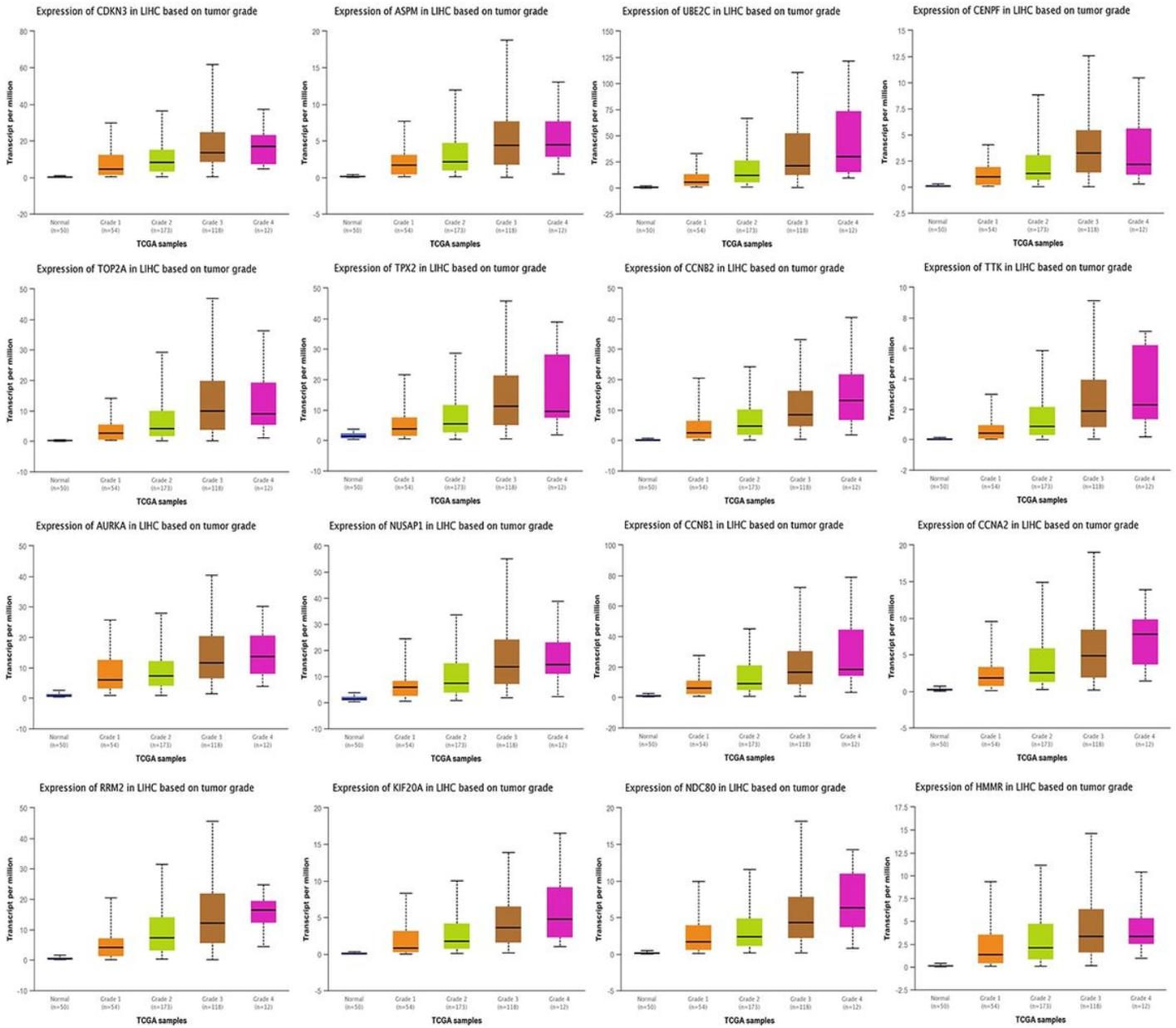


Figure 9

The expression of hub genes in different tumor grades from HCC and normal tissues was constructed using UALCAN online database.

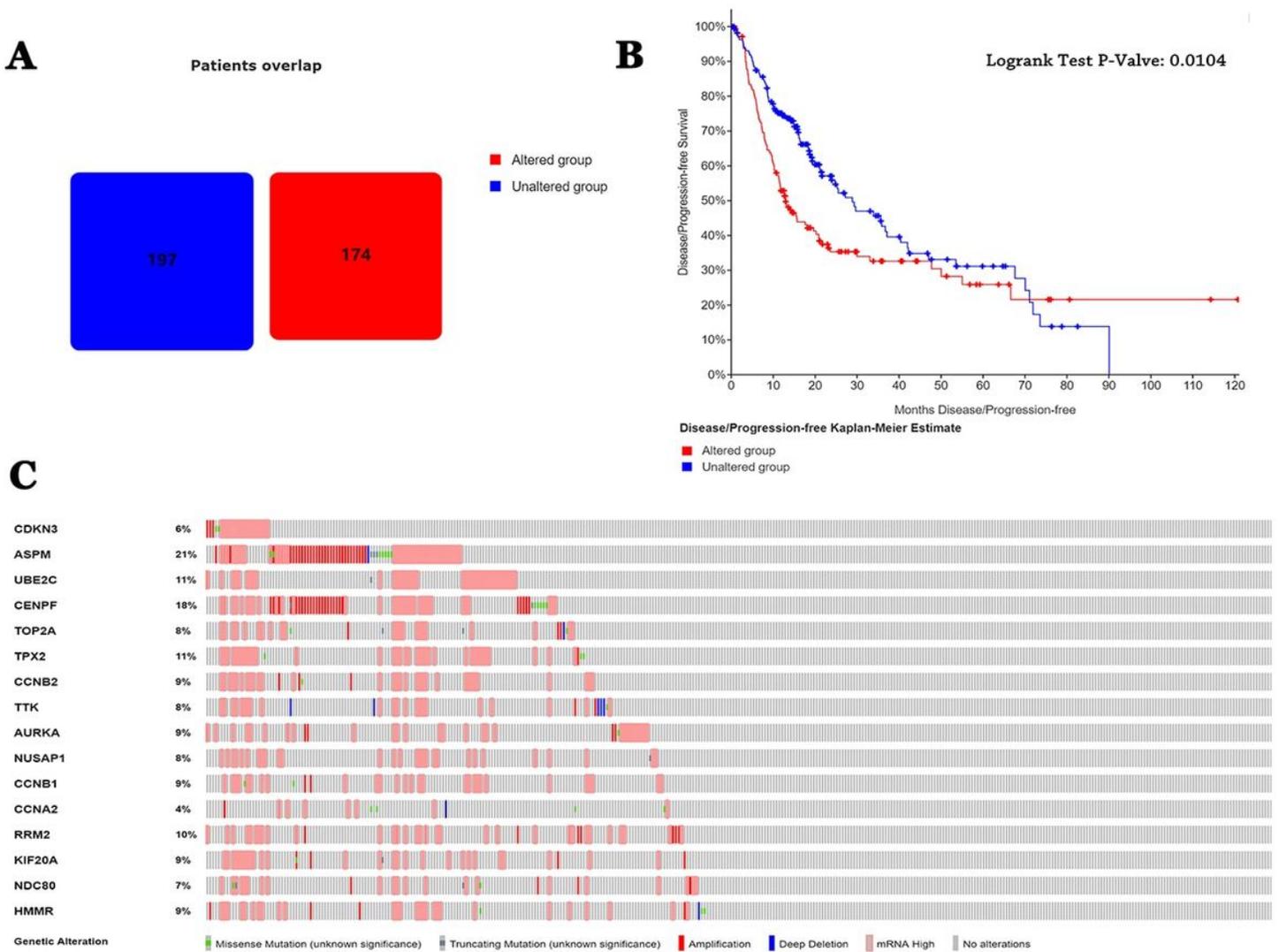


Figure 10

(A) Patients overlap. Genetic alterations were found in 174 of 371 HCC patients (47%). (B) Disease-free survival Kaplan-Meier estimate between groups with and without alterations. $P < 0.05$ was considered statistically significant. (C) Genetic alterations. Red represents amplification, blue represents deep deletion. The threshold was defined as z-score ± 2.0 from the TCGA RNA Seq V2 data. This OncoPrint was conducted by cBioPortal.

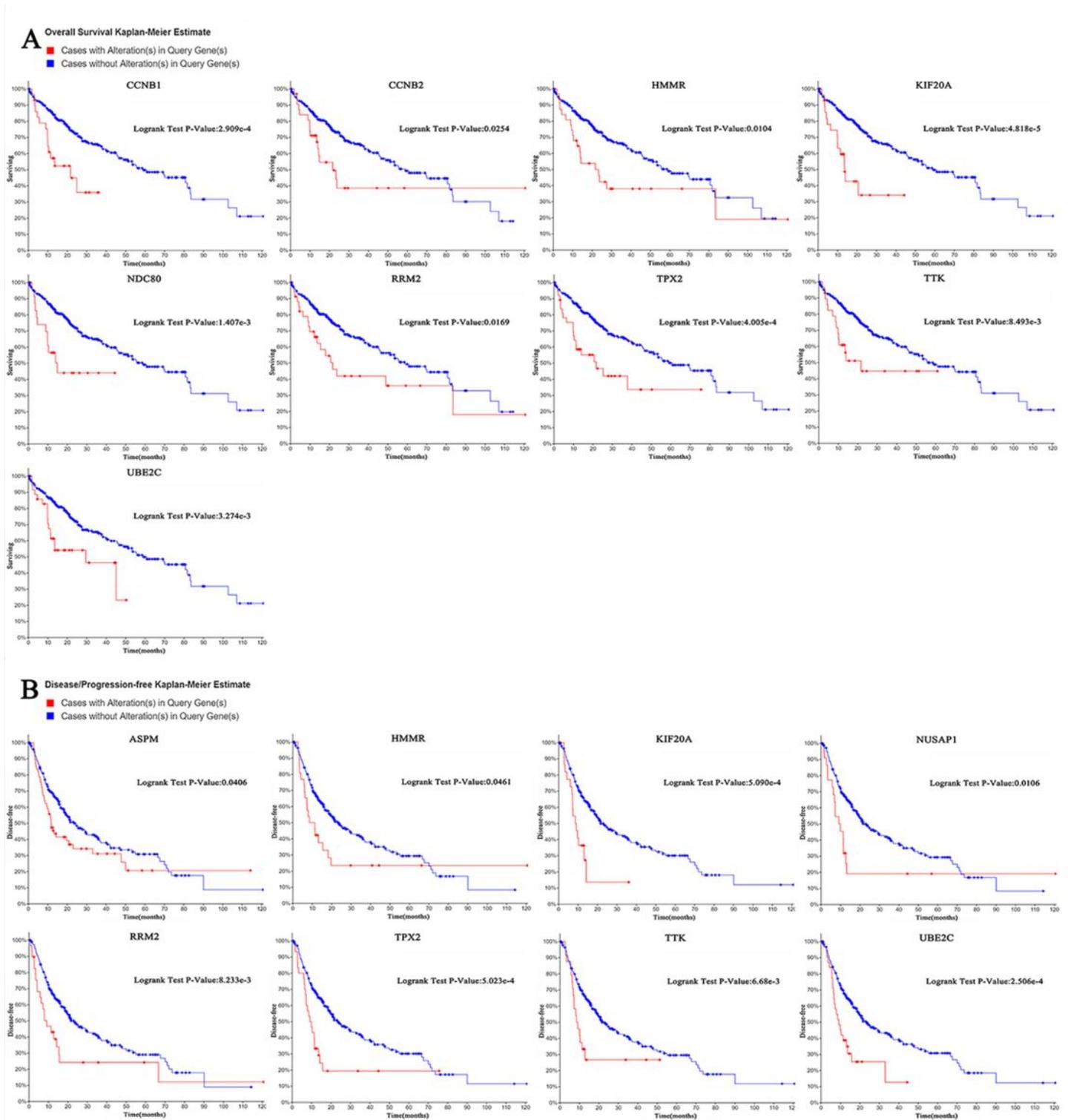


Figure 11

(A) Overall survival and (B) Disease-free survival analyses of hub genes were performed using cBioPortal online platform. $P < 0.05$ was considered statistically significant.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS2string0.7.tif](#)
- [TableS3module2GO.txt](#)
- [TableS1diff.xls](#)
- [TableS2module1GO.txt](#)
- [FigS1heatmap.tif](#)