Bioinformatic Identification and Validation of Key Genes and Biological Pathways Involved in Hepatitis B Virus-related Hepatocellular Carcinoma

Guan-Hua Ren
Guangxi Medical University Cancer Hospital
Fan-Biao Mei
Guangxi Medical University Cancer Hospital
Chao-Jun Zhang
Guangxi Medical University Cancer Hospital
Dong-Mei Cai
Guangxi Medical University Cancer Hospital
Long Long
Guangxi Teachers Education University
Ke-Zhi Li
Guangxi Medical University Cancer Hospital
Tian-Ren Huang
Guangxi Medical University Cancer Hospital
Wei Deng (dengwei@gxmu.edu.cn)
Department of Research, Guangxi Medical University Cancer Hospital, Nanning, Guangxi Zhuang Autonomous Region, 530021, People's Republic of China https://orcid.org/0000-0003-0946-5579

Research

Keywords: Hepatocellular carcinoma, Hepatitis B virus, Gene Expression Omnibus, Bioinformatic analysis

Posted Date: November 18th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-106893/v1

Abstract

Objectives: Hepatocellular carcinoma (HCC) is a common malignant tumor severely scathing human health. As we all know, one of the main risk factors of HCC is chronic hepatitis B virus (HBV) infection, which involved in the oncogenesis of HCC through direct and indirect mechanisms such as inflammatory injury, integration into the host genome and interaction with some special target genes. This study aimed to identify these key genes potentially associated with HBV-related HCC by bioinformatics analyses.

Results: Total of 320 DEGs was identified between HBV-related HCC tissue samples and adjacent normal samples. These DGEs were strongly associated with several biological processes, such as retinol metabolism and steroid hormone biosynthesis. A PPI network was constructed and top six hub genes, including CDK1, CCNB1, CDC20, CDKN3, HMMR and MKI67, were determined. GEPIA online tool analysis validated the six key hub genes had the same expression trend as predicted in The Cancer Genome Atlas (TCGA) datasets. The overall survival and disease-free survival reflected that high expression of CDK1, CDC20, HMMR, MKI67 and CCNB1 significantly predicted poor prognosis, whereas CDKN3 expression has no statistical differences in overall survival.

Conclusion: The present study identified key genes and pathways involved in HBV-related HCC, which will improve our understanding of the mechanisms underlying the development and recurrence of HCC. The six identified genes might be potential biomarkers for the diagnosis and treatment of HBV-related HCC.

Introduction

Primary liver cancer is one of the most common malignant tumors with high mortality and seriously affects human health worldwide. Liver cancer ranks as the fifth most frequently-diagnosed cancer in global cases, and it ranks as the second in terms of deaths for male death. Hepatocellular carcinoma (HCC) comprising 75% to 85% cases of primary liver cancer [1]. One of main risk factors of HCC is chronic infection with hepatitis B virus (HBV), responsible for 50~80% of HCC cases worldwide [2]. Prospective cohort studies indicated that HBV confers with 5- to 100-fold increased risk of HCC patient [3]. Active hepatitis B continues to drive most of the global burden of HCC from a public health perspective. As we all know, the process of HBV causing HCC is a complex and multistep process involving many factors. Both direct and indirect mechanisms related to HBV genome and HBV expression products were associated with HCC oncogenesis [4].

In the past two decades, with the development of microarray and high-throughput sequencing technology, as well as the public databases such as Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA), many high-throughput platforms are widely utilized to identify gene expression profiling in clinical oncology research [5, 6]. As for liver cancer, there are numerous studies being performed to explore the differentially expressed genes (DEGs) to date [7], which have provided us with great help in understanding the genetic landscape and driver pathways leading to HCC [8-11]. In those studies, a series of liver cancer biomarkers have been identified, including various oncogenes and tumor suppressor genes [12, 13]. Among that, the discovery of tumor protein p53 (TP53), catenin beta 1 (CTNNB1), Axin-1 (AXIN1) and other key genes were meaningful as the driver mutations evidence for the study of liver cancer mechanism [10]. Unfortunately, few additional DEGs have been validated and clinically practiced, especially for the HBV-related HCC. This is mainly due to the insufficient sample size of the existing studies utilizing the HBV-related HCC tissue in the public datasets, and the

heterogeneity among these studies also increase the difficulty of the integrated bioinformatics analysis between different data sets. Therefore, it is still imperative to investigate the specific biomarkers potentially associated with HBV-related HCC and clarify the underlying mechanism of the initiation and progression of HCC. Furthermore, it is necessary to use the appropriate bioinformatics analysis methods to integrate the latest genomic data derived from different experimental platforms for accurate results.

In the present study, four microarray datasets GSE121248, GSE19665, GSE47197 and GSE55092 from GEO database was integrated and subsequently identify the DEGs by a series of credible bioinformatics analysis methods (Figure 1). The corresponding results that used the gene expression profiles derived from the HBV induced HCC and adjacent liver samples in these datasets might provide new scientific insights into the diagnosis and treatment of HBV-related HCC.

Results

Identification of DEGs in HBV-related HCC

Before DEGs identification, four HBV-related HCC chip expression datasets GSE121248, GSE19665, GSE47197 and GSE55092 were normalized conventionally [14], and the results were shown in Figure 2). After screening the normalized data by the R package limma, 580, 647, 452 and 1043 DEGs (adjusted $P < 0.05$ and |log fold change (FC)| > 1) were found in these four datasets respectively (Table 1). The volcano plot described the distribution of the DEGs expression, shown in Figure 3. Screen the top differential genes by *P*-value from small to large, the top 100 upregulated genes and the top 100 downregulated genes of each dataset were shown in the Figure 4.

After integration and analysis by RRA, a total of 320 DEGs were identified from the four datasets, including 104 upregulated genes and 216 downregulated genes. The top 20 upregulated genes and the top 20 downregulated genes were charted as a heat map and shown in Figure 5.

Functional annotation analysis of DEGs

The GO functional analysis of 320 integrated differential genes was divided into the following three parts: biological process (BP), molecular function (MF) and cell component (CC), and the three parts of the GO results are shown in Figure 6. The top 15 GO enrichment terms associated with the upregulated and downregulated genes were presented in Table 2. The results showed that the DEGs significantly regulating the BP including cell division, mitotic nuclear division, chromosome segregation, sister chromatid cohesion, anaphase-promoting complex-dependent catabolic process, immune response, complement activation, classical pathway, negative regulation of vascular endothelial growth factor signaling pathway and chemokine-mediated signaling pathway ($P < 0.05$). The DEGs significantly regulating the CC involving in spindle, condensed chromosome kinetochore, midbody, nucleus and chromosome, centromeric region, extracellular exosome, membrane attack complex, extracellular space, extracellular region, and insulin-like growth factor ternary complex ($P < 0.05$). The DEGs significantly regulating the MF including protein binding, microtubule binding, protein kinase binding, ATP binding, histone binding, heme binding, iron ion binding, chemokine activity, monooxygenase activity and mannose binding ($P < 0.05$).

Signaling pathway analysis of DEGs

Following the KEGG pathway enrichment analysis by DAVID, the results (Table 3) revealed that the identified DEGs associated with HBV-related HCC were mainly enriched in the signaling pathway of retinol metabolism, steroid hormone biosynthesis, tryptophan metabolism, metabolic pathways, and prion diseases, complement and coagulation cascades ($P < 0.05$).

Protein–protein interaction network of DEGs and key gene selection

By using the STRING online database, we construct a PPI network of 320 identified DEGs, which consisted of 277 nodes and 1,943 edges (data not shown). The top 12 proteins were defined as crucial proteins based on four different centrality parameters (Table 4), and the intersection of these top 12 proteins were shown in Figure 7. Of which, six consistent key genes were identified, that is CDK1, CDC20, CDKN3, HMMR, MKI67 and CCNB1, and all six genes were upregulated in HBV-related HCC samples.

By using MCODE, 14 functional modules were screened from the PPI network and two most significant modules were selected for KEGG pathway enrichment analysis (Figure 8). Results showed that the genes in module 1 (MCODE score= 47.1) were mainly enriched in cell cycle, oocyte meiosis, p53 signaling pathway, small cell lung cancer; and the genes in module 2 (MCODE score= 11.636) were enriched in complement and coagulation cascades, prion diseases, systemic lupus erythematosus (Table 5).

Validation of key genes mRNA expression in GEPIA (TCGA) data base

The mRNA expression of six key genes (i.e. CDK1, CDC20, CDKN3, HMMR, MKI67 and CCNB1) of 369 HCC cancer tissue samples and 160 adjacent normal tissue samples in the GEPIA database were evaluated. The results showed that the expression levels of CDK1, CDC20, CDKN3, HMMR, MKI67 and CCNB1 were higher in HCC cancer tissue than in normal tissue (P 0.05, Figure 9 and Figure 10). These surveys have produced results essentially in agreement with the GEO dataset analysis. Additionally, we compared the expression levels of these key genes between different stages of HCC. There were significant differences of expression in CDK1 (P = 5.76e-05), CDC20 (P = 1.59e-05), CCNB1 (P = 8.7e-05), CDKN3 (P = 0.00354), HMMR (P = 0.000504) and MKI67groups (P = 0.000167) (Pr(>F) <0.05, Figure 11). The expression of key genes increased with the progress in HCC.

The protein expression of key genes in HPA data base

The immunohistochemistry staining of CDK1, CDC20, HMMR, MKI67 and CCNB1 protein were validated upregulated in liver cancer tissues when comparing with normal tissues in the Human Protein Atlas database (Figure 12). The result from the HPA dataset proved the protein expression matched with the mRNA expression in CDK1, CDC20, HMMR, MKI67 and CCNB1. The protein expression data of CDKN3 was absent in the Human Protein Atlas database.

The association of the key genes with the HBV related-HCC survival in GEPIA (TCGA) data base

The associations of the key genes (i.e. CDK1, CDC20, CDKN3, HMMR, MKI67 and CCNB1) with the survival of HCC patients in GEPIA were explored and the results were shown in Figure 13 and Figure 14. Disease-free survival curves are presented in Figure 13, it turns out to be that the high expression of CDK1 (HR =1.7, logrank *P* =0.00057), CDC20 (HR=1.6, logrank *P* =0.0026), HMMR (HR =1.6, logrank *P* =0.0034), CDKN3 (HR =1.5, logrank *P* =0.0074), and MKI67 (HR =1.9, logrank *P* =4.2e-05) and CCNB1 (HR =2, logrank *P* =2.8e-06) mRNA level were associated with the worse DFS in patients with HCC. Overall survival curves are stated in Figure 14, except for CDKN3 (HR = 1.4, logrank *P* =0.058), the high expression of the other five DEGs were also associated with the worse OS in patients with HCC. The result turns out to be that the high expression of CDK1 (HR =2, logrank *P* =0.00017), CDC20 (HR=2.3, logrank *P* =3.8e-06), HMMR (HR =1.7, logrank *P* =0.0031) and MKI67 (HR =1.9, logrank *P* =0.00045) and CCNB1 (HR =2, logrank *P* =0.00015) were significant difference on the survival curve in HCC.

Discussion

In this study, we used the bioinformatics method to analyze four GEO datasets (GSE121248, GSE19665, GSE47197 and GSE55092) and identified 320 DEGS between HBV-related HCC and adjacent liver tissue samples, including 104 upregulated genes and 216 downregulated genes. The results of GO analysis revealed that the DEGs were mainly involved in the functions and pathways of cell division (GO:0051301), chromosome and centromeric region (GO:0000775), protein binding (GO:0005155), inflammatory response (GO:0006954), extracellular exosomes (GO:0070062) and heme binding (GO:0020037), which suggested that these DEGs might promote the occurrence and development of HBV-related HCC by regulating the cell proliferation and cell cycle, Interfering with the host immune system and disturbing the endocrine function. The KEGG pathway analysis revealed that the DEGs were mainly enriched in retinol metabolism, steroid hormone biosynthesis, tryptophan metabolism, metabolic pathways, and prion diseases. KEGG pathway analysis showed that the genes in module 1 and module 2 were enriched in cell cycle, complement and coagulation respectively. All the above results show that the progression in cell cycle has a close relationship with HBV-HCC. These gene functions and the signaling pathways might interact with HBV and corporately affect the development of HCC, especially by regulating the cell cycles and proliferation [15]. Xia et al. found that HBV parvoviral host factors (such as PPARA, CEBPB and RXRA) were upregulated upon infection and enriched in primary human hepatocytes in the G2/M phase, which supports that HBV may promote viral replication and a premalignant phenotype that predisposes infected hepatocytes to subsequent malignant transformation by deregulating the cell cycle [16].

By using the STRING online database to construct key genes coding protein-protein interaction (PPI) network and perform the key genes and modules analysis, we finally identified six key genes by intersected four different centrality parameters, which were CDK1, CCNB1, CDC20, CDKN3, HMMR and MKI67. The expression analysis of these key genes was verified consist with the TCGA database. And the survival analysis showed that the upregulated expression of these six key genes might correlate with worse prognosis of HBV-related HCC. The above validation data demonstrated that the bioinformation analyses and results in our present study were robust and reliable.

CDK1, also known as cyclin-dependent kinase1 or cell division cycle protein 2 homolog, belongs to serine/threonine protein kinase family that can phosphorylate hundreds of proteins [17]. CDK1 plays a key role in regulation of cellular architecture and cell adhesion during cell cycle [18]. Studies have shown that CDK1 over-expression directly reflects the cell cycle progression and the activity of cell proliferation, and CDK1 is a main predictive factor of HCC recurrence together with Ki-67 [19]. Kim et al. found inhibitory phosphorylation of CDK1 was elevated during the period which Hepatitis B virus X protein activates the ATM-Chk2 pathway and delays cell cycle progression[20]. Hu et al. conclude that CDK1 was able to adjusting phosphorylation of SAMHD1, and may contribute the suppression of HBV replication in HCC[21].

CCNB1, also known as CyclinB1, belongs to the highly conserved cyclin family and is significantly over-expressed in various malignant tumors. Over-expression of CCNB1 has been reported in breast [22], colorectal [23], lung [24], thyroid [25], prostate [26], pancreatic [27], stomach [28] and liver [19] cancers. CCNB1 plays most important role in mitotic cycle, in which it regulates the G2-M transition and cell proliferation [29]. Recent studies reported that CCNB1 is a diagnostic markers and potential therapeutic target for HBV-related HCC following surgery [30, 31]. Reports indicated that knockdown of CCNB1 regulated by microRNA-144 significantly inhibited cell proliferation, migration, and invasion in HCC [31]. Besides, CCNB1 is also a promising biomarker for ER+ breast cancer [32] and rhabdomyosarcoma [33].

CCNB1 in partnership with CDK1 could generate the M phase-promoting factor (MPF) activity [34], in which can coordinate the cell cycle progression directly or with enhanced mitochondrial bioenergetics [35]. Furthermore, Wang et al. proved that CCNB1/CDK1-mediated phosphorylation provides efficient bioenergy for G2/M transition in cells and shortens the overall cell-cycle time [29]. Therefore, CCNB1/CDK1 plays an important role in the cell cycle and cell proliferation. Hepatitis B virus X protein (HBx) was able to inhibit the growth of HCC cells and induce G2/M arrest in vitro by apoptosis through sustained activation of cyclin B1-CDK1 kinase [36]. The HBx gene of HBV is the most common open reading frame that may develop into HCC. This study suggested that CCNB1 may be involved in the process of HBV-induced HCC. Zou Y, et al. conducted a multi-omics analysis of CDK1, CCNB1 and CCNB2 and found that they are potential prognostic biomarkers and associated with immune cell infiltration in HCC [37]. The experimental version of the Jin J L, et al study confirmed LINC00346 could regulate LINC00346-miR-199a-3p-CDK1/CCNB1 Axis to influence the progress of HCC[38].

CDC20, also known as cell division cycle 20, is an cell-cycle regulator required for the completion of mitosis and an activator of the anaphase-promoting complex/cyclostome (APC/C) in the cell cycle [39]. CDC20 has been validated to be critically involved in the development and progression of HCC. One study has found that high levels of CDC20 over-expression were positively correlated with gender, tumor differentiation, TNM stage of HCC , however, no statistically significant association was found between high CDC20 and HBV infection [40].

CDKN3, as known as cyclin-dependent kinase inhibitor 3, belongs to the protein phosphatases family and is often increasingly express in human cancer. CDKN3 has a complex dual function in control of the cell cycle in the occurrence and development of HCC. Previous study shown that CDKN3 is up-regulated in HCC and is related to poor clinical outcome of HCC [41]. But Wei et.al observed that the expression levels of CDKN3 was decreased in HCC tumor compared with normal liver tissue [42]. While they also found that the HBV infection ratio was higher in cancer group.

HMMR, as known as the receptor for hyaluronan mediated motility or RHAMM, CD168, is a multifunctional protein for the regulation of cell motility and proliferation acting both inside and outside the cells [43]. Several studies have provided evidence that the HMMR expression level was higher in HCC tissues compared with normal issues [44, 45]. And the upregulation of HMMR is mutually related to HCC progression and prognosis [44]. Liu et al. demonstrated that HMMR might represent a cellular motility factor that can be induced by HBx protein from HBV in HCC via PI3K-Akt-Oct-1 signaling [46].

MKI67, the marker of proliferation Ki-67,which is a traditional proliferation marker for the evaluation of cell proliferation found in the cell nucleus [47]. Multivariate studies have demonstrated that the expression of Ki-67 is an independent prognostic indicator for the prognosis of HCC [48]. and for patients with HCC after resection [49]. The meta-analysis also proved that Ki-67 is a biomarker for clinical deterioration and poor prognosis in HCC [50]. C.Yang et al. found that MKI67 expression is related to transforming growth factor beta 1 and can predict the prognosis of patients with HBV-related HCC. MKI67 expression combined with cirrhosis and BCLC stages could predict the clinical outcomes of HBV-related HCC. Higher levels of MKI67 expression in HBV-related HCC patients lead to a worse OS and DFS [51].

Conclusion

In the present study, we identified six DEGs that might be the potential biomarkers of HBV-related HCC using the bioinformation analysis methods, including CDK1, CCNB1, CDC20, CDKN3, HMMR and MKI67. These six genes might collaborate with HBV infection to regulate the initiation and progression of HCC by primarily disturbing the cell cycle and proliferation process. Additional biological researches and mechanistic studies should be performed to verify our founding and explore possible molecular mechanisms underlying the observed associations between these six DEGs and HBV-related HCC.

Materials And Methods

Microarray data and data preprocessing

The microarray datasets were acquired from the National Center for Biotechnology Information (NCBI) GEO database (http://www.ncbi.nlm.nih.gov/geo/). The dataset was searched by using terms “liver cancer” or “HCC” or “hepatocellular carcinoma” and “HBV” or “hepatitis B virus” in GEO database. The selection criteria were as follows: (a) Inclusion of gene expression data of HBV-related HCC and adjacent normal liver tissue samples or non-tumor tissue samples; (b) Normal liver tissue samples from donors and liver cirrhosis tissue samples were excluded; (c) Arrays contained a minimum of 5 tumor and adjacent normal tissue samples; (d) Inclusion of >5,000 genes in the GEO platform. According to the above screening criteria, four eligible datasets including GSE121248, GSE19665, GSE47197 and GSE55092 were selected for the subsequent analysis in this study. There was a total of 381 samples (185 cancer tissue samples and 196 adjacent normal tissue samples) in these four datasets, and the detail sample information were shown in Table 6. The platform and series matrix files of the four databases were downloaded separately as TXT files. The sample IDs corresponding to the probe name were converted into the international standard gene name and resaved in each matrix files by using a Perl language command.

Data processing and identification of DEGs

The R software package “dplyr” (http://www.bioconductor.org/) was used to process the downloaded files to convert and reject the unqualified data. Subsequently, each gene expression data should normalize by the Normalization Between Arrays function in the R package “limma” (V 3.40.6) (http://www.bioconductor.org/) (Ritchie et al., 2015). All gene expression data were subjected to log2 transformation. Besides, mean values of log2 transformation was used when multiple probe sets were used for one gene. After normalizing the data, we screened the DEGs in each dataset separately by the R package “limma” and saved the result of those DEGs as Excel files. To further next integration analysis, we also saved the list of all gene sorted by log FC in four TXT files. Then we used Robust Rank Aggreg (RRA) R package (https://cran.rstudio.com/bin/windows/contrib/3.5/RobustRankAggreg_1.1.zip) [52] to integrate and analyze the results of four datasets to identify the most significant DEGs. Samples with the adjust *P*-value < 0.05 and |log fold change (FC)| > 1 were considered DEGs. The list of integrated upregulated and downregulated DEGs were exported and saved for subsequent analysis. The R package “Heatmap” and "ggplot2" were utilized to visualize the expression patterns of the results from DEGs analysis.

GO and KEGG pathway enrichment analyses of DEGs

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a tool for systematic analysis of gene functions, linking genomic information with higher order functional information by current knowledge on cellular processes and by standardizing gene annotations, including complex pathways [53]. The Gene Ontology (GO) analysis is a powerful tool for cataloguing gene function and identify characteristic biological attributes by analyzing high-throughput genome data, which is useful for the mining of functional and biological significance from very large datasets [54, 55]. The Database for Annotation, Visualization and Integrated Discovery (DAVID; http://david.ncifcrf.gov) (version 6.8) is a common online program integrate heterogeneous gene annotation resources for gene enrichment and functional annotation analyses [56, 57]. In order to further conduct the function and pathway analysis, the DEGs list of the current study were copied into the DAVID to perform the GO and KEGG enrichment analyses. The results were considered statistically significant when $P < 0.05$.

PPI network construction and module analysis

The Search Tool for the Retrieval of Interacting Genes (STRING; https://string-db.org/) [58] is a public search tool designed to provide a critical assessment and integration of protein–protein interactions (PPI). To find the interactional correlation of the proteins encoded by DEGs, we use the STRING to construct the PPI network (integrated scores >0.4), and the TSV files of the results of PPI networks were downloaded for subsequent analysis. Cytoscape (3.7.1) is an open source software project for integrated models of biomolecular interaction networks [59]. The plug-in Molecular Complex Detection (MCODE) in Cytoscape software [60] was performed to screen key modules of the PPI network. The criteria were set as follows: degree cut-off ≥ 2, node score cut-off ≥ 0.2, K-core ≥ 2, and max depth = 100.

Key genes selection and analysis

Cytohubba (http://apps.cytoscape.org/download/stats/cytohubba/) [61] is an identifying hub objects and sub-networks from complex interactome. It also provides a user-friendly interface to explore important nodes by ranking nodes in their network features. To screen the key genes from PPI network, we analyzed the centrality of nodes of the result of network and calculated node's scores of each gene by the Cytohubba app in Cytoscape software. The key proteins were identified based on four different centrality parameters degree centrality, betweenness centrality, and closeness centrality, Maximal Clique Centrality (MCC). And MCC can capture more essential proteins in the top ranked list in both high-degree and low-degree proteins among them.

Transcriptional expression level and survival analysis of key genes in HCC

Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/), a web-based tool that can deliver fast and customizable functionalities based on the Cancer Genome Atlas and the Genotype-Tissue Expression (GTEx) projects data (Tang et al., 2017), including the RNA sequencing expression data of 9736 tumors and 8587 normal samples. GEPIA provides key interactive and customizable functions including differential expression analysis, profiling plotting, correlation analysis and patient survival analysis. After key genes identified from these datasets, GEPIA was used to validate the selected upregulated and downregulated key genes. The thresholds in expression scatter diagram and boxplot of key genes were set to "1,0.01 and 0.4," respectively, for the$|Log_2FC|$ cutoff, *P* value cutoff and Jitter size. The expression stage plot of key genes was exported with the default parameters "Use major stage = Yes", "Datasets = LIHC" and "Log Scale = Yes".

To further elucidate whether these key genes contributed to the survival of patients with HCC, we analyzed the association of these key genes with overall survival (OS) and disease-free survival (DFS) of HCC, and the staging and survival curves were plotted by GEPIA. The following parameters were set as: Group cutoff = Median, Hazards Ratio= Yes, 95% Confidence Interval = Yes, Axis Units= Mouths.

Revalidation of protein expression of key genes in HCC

The Human Protein Atlas (HPA, http://www.proteinatlas.org) was used for protein detection by immunohistochemistry [62, 63]. HPA have provided many useful visualization and analysis tools for gene expression analysis, which is best at integrating protein information. The expression level of the proteins encoded by key DEGs in this study were validated using the HPA dataset.

Abbreviations

BP	Biological process
CC	Cell component
CCNB1	CyclinB1
CDC20	Cell division cycle 20
CDK1	Cyclin-dependent kinase1
CDKN3	Cyclin-dependent kinase inhibitor 3
DAVID	Database for annotation, visualization and integrated discovery
DEGs	Differentially expressed genes
DFS	Disease-free survival
GEO	Gene Expression Omnibus
GEPIA	Gene Expression Profiling Interactive Analysis
GO	Gene Ontology
HBV	Hepatitis B virus
HBx	Hepatitis B virus X protein
HCC	Hepatocellular carcinoma
HMMR	Receptor for hyaluronan mediated motility
HPA	Human Protein Atlas
KEGG	Kyoto encyclopedia of genes and genomes
MCC	Maximal Clique Centrality
MCODE	Molecular Complex Detection
MF	Molecular function
MKI67	Marker of proliferation Ki67
MPF	M phase-promoting factor
NCBI	National Center for Biotechnology Information
OS	Overall survival
PPI	Protein–protein interactions
RRA	Robust Rank Aggreg
STRING	Search Tool for the Retrieval of Interacting Genes
TCGA	Cancer Genome Atlas

Declarations

Acknowledgments

The authors thank the numerous individuals who participated in this study.

Authors’ contributions

Conception and design of the research: WD, LKZ; the acquisition of data: GHR, DMC; processing the data: GHR, FBM; the visualization of analysis: GHR, LL; the analysis and interpretation of data: GHR, CJZ; drafting the manuscript: GHR; reviewing the manuscript: WD, TRH. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China [grant number 81660561, grant number 81260319]; 2018 Guangxi One Thousand Young and Middle-Aged College and University Backbone Teachers Cultivation Program (To Wei Deng) and the Key Project of GuangXi Natural Science Foundation [grant number 2019GXNSFDA245001]

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing financial interests.

Footnotes

Publisher’s Note

Contributor Information

Wei Deng, Email: dengwei@gxmu.edu.cn

Tian-Ren Huang, Email: tianrenhuang@sina.com

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. Venook AP, Papandreou C, Furuse J, de Guevara LL. The incidence and epidemiology of hepatocellular carcinoma: a global and regional perspective. The oncologist. 2010;15 Suppl 4:5-13.
3. El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. Gastroenterology. 2012;142(6):1264-73 e1.
4. Chaturvedi VK, Singh A, Dubey SK, Hetta HF, John J, Singh MP. Molecular mechanistic insight of hepatitis B virus mediated hepatocellular carcinoma. Microb Pathog. 2019;128:184-94.
5. King HC, Sinha AA. Gene expression profile analysis by DNA microarrays: promise and pitfalls. JAMA. 2001;286(18):2280-8.
6. Ramaswamy S, Golub TR. DNA microarrays in clinical oncology. J Clin Oncol. 2002;20(7):1932-41.
7. Xing T, Yan T, Zhou Q. Identification of key candidate genes and pathways in hepatocellular carcinoma by integrated bioinformatical analysis. Exp Ther Med. 2018;15(6):4932-42.
8. Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nature genetics. 2012;44(6):694-8.
9. Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nature genetics. 2014;46(12):1267-73.
10. Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nature genetics. 2015;47(5):505-11.
11. Chen J, Zaidi S, Rao S, Chen JS, Phan L, Farci P, et al. Analysis of Genomes and Transcriptomes of Hepatocellular Carcinomas Identifies Mutations and Gene Expression Changes in the Transforming Growth Factor-β Pathway. Gastroenterology. 2018;154(1):195-210.
12. Mathew S, Ali A, Abdel-Hafiz H, Fatima K, Suhail M, Archunan G, et al. Biomarkers for virus-induced hepatocellular carcinoma (HCC). Infect Genet Evol. 2014;26:327-39.
13. Shen S, Lin Y, Yuan X, Shen L, Chen J, Chen L, et al. Biomarker MicroRNAs for Diagnosis, Prognosis and Treatment of Hepatocellular Carcinoma: A Functional Survey and Comparison. Sci Rep. 2016;6:38311.
14. Gao X, Chen Y, Chen M, Wang S, Wen X, Zhang S. Identification of key candidate genes and biological pathways in bladder cancer. PeerJ. 2018;6:e6036.
15. Torresi J, Tran BM, Christiansen D, Earnest-Silveira L, Schwab RHM, Vincan E. HBV-related hepatocarcinogenesis: the role of signalling pathways and innovative ex vivo research models. BMC Cancer.

2019;19(1):707.
16. Xia Y, Cheng X, Li Y, Valdez K, Chen W, Liang TJ. Hepatitis B Virus Deregulates the Cell Cycle To Promote Viral Replication and a Premalignant Phenotype. J Virol. 2018;92(19).
17. Gavet O, Pines J. Progressive activation of CyclinB1-Cdk1 coordinates entry to mitosis. Developmental cell. 2010;18(4):533-43.
18. Jones MC, Askari JA, Humphries JD, Humphries MJ. Cell adhesion is regulated by CDK1 during the cell cycle. J Cell Biol. 2018;217(9):3203-18.
19. Ito Y, Takeda T, Sakon M, Monden M, Tsujimoto M, Matsuura N. Expression and prognostic role of cyclin-dependent kinase 1 (cdc2) in hepatocellular carcinoma. Oncology. 2000;59(1):68-74.
20. Kim S, Lee HS, Ji JH, Cho MY, Yoo YS, Park YY, et al. Hepatitis B virus X protein activates the ATM-Chk2 pathway and delays cell cycle progression. The Journal of general virology. 2015;96(8):2242-51.
21. Hu J, Qiao M, Chen Y, Tang H, Zhang W, Tang D, et al. Cyclin E2-CDK2 mediates SAMHD1 phosphorylation to abrogate its restriction of HBV replication in hepatoma cells. FEBS Lett. 2018;592(11):1893-904.
22. Chae SW, Sohn JH, Kim DH, Choi YJ, Park YL, Kim K, et al. Overexpressions of Cyclin B1, cdc2, p16 and p53 in human breast cancer: the clinicopathologic correlations and prognostic implications. Yonsei medical journal. 2011;52(3):445-53.
23. Korenaga D, Takesue F, Yasuda M, Honda M, Nozoe T, Inutsuka S. The relationship between cyclin B1 overexpression and lymph node metastasis in human colorectal cancer. Surgery. 2002;131(1 Suppl):S114-20.
24. Arinaga M, Noguchi T, Takeno S, Chujo M, Miura T, Kimura Y, et al. Clinical implication of cyclin B1 in non-small cell lung cancer. Oncol Rep. 2003;10(5):1381-6.
25. Nar A, Ozen O, Tutuncu NB, Demirhan B. Cyclin A and cyclin B1 overexpression in differentiated thyroid carcinoma. Medical oncology (Northwood, London, England). 2012;29(1):294-300.
26. Mashal RD, Lester S, Corless C, Richie JP, Chandra R, Propert KJ, et al. Expression of cell cycle-regulated proteins in prostate cancer. Cancer research. 1996;56(18):4159-63.
27. Zhou L, Li J, Zhao YP, Cui QC, Zhou WX, Guo JC, et al. The prognostic value of Cyclin B1 in pancreatic cancer. Medical oncology (Northwood, London, England). 2014;31(9):107.
28. Begnami MD, Fregnani JH, Nonogaki S, Soares FA. Evaluation of cell cycle protein expression in gastric cancer: cyclin B1 expression and its prognostic implication. Human pathology. 2010;41(8):1120-7.
29. Wang Z, Fan M, Candas D, Zhang TQ, Qin L, Eldridge A, et al. Cyclin B1/Cdk1 coordinates mitochondrial respiration for cell-cycle G2/M progression. Developmental cell. 2014;29(2):217-32.
30. Weng L, Du J, Zhou Q, Cheng B, Li J, Zhang D, et al. Identification of cyclin B1 and Sec62 as biomarkers for recurrence in patients with HBV-related hepatocellular carcinoma after surgical resection. Molecular cancer. 2012;11:39.
31. Gu J, Liu X, Li J, He Y. MicroRNA-144 inhibits cell proliferation, migration and invasion in human hepatocellular carcinoma by targeting CCNB1. Cancer Cell Int. 2019;19:15.
32. Ding K, Li W, Zou Z, Zou X, Wang C. CCNB1 is a prognostic biomarker for ER+ breast cancer. Med Hypotheses. 2014;83(3):359-64.
33. Li Q, Zhang L, Jiang J, Zhang Y, Wang X, Zhang Q, et al. CDK1 and CCNB1 as potential diagnostic markers of rhabdomyosarcoma: validation following bioinformatics analysis. BMC Med Genomics. 2019;12(1):198.

34. Gautier J, Minshull J, Lohka M, Glotzer M, Hunt T, Maller JL. Cyclin is a component of maturation-promoting factor from Xenopus. Cell. 1990;60(3):487-94.
35. Xie B, Wang S, Jiang N, Li JJ. Cyclin B1/CDK1-regulated mitochondrial bioenergetics in cell cycle progression and tumor resistance. Cancer Lett. 2019;443:56-66.
36. Cheng P, Li Y, Yang L, Wen Y, Shi W, Mao Y, et al. Hepatitis B virus X protein (HBx) induces G2/M arrest and apoptosis through sustained activation of cyclin B1-CDK1 kinase. Oncol Rep. 2009;22(5):1101-7.
37. Zou Y, Ruan S, Jin L, Chen Z, Han H, Zhang Y, et al. CDK1, CCNB1, and CCNB2 are Prognostic Biomarkers and Correlated with Immune Infiltration in Hepatocellular Carcinoma. Med Sci Monit. 2020;26:e925289.
38. Jin J, Xu H, Li W, Xu X, Liu H, Wei F. LINC00346 Acts as a Competing Endogenous RNA Regulating Development of Hepatocellular Carcinoma via Modulating CDK1/CCNB1 Axis. Front Bioeng Biotechnol. 2020;8:54.
39. Kapanidou M, Curtis NL, Bolanos-Garcia VM. Cdc20: At the Crossroads between Chromosome Segregation and Mitotic Exit. Trends in biochemical sciences. 2017;42(3):193-205.
40. Li J, Gao JZ, Du JL, Huang ZX, Wei LX. Increased CDC20 expression is associated with development and progression of hepatocellular carcinoma. Int J Oncol. 2014;45(4):1547-55.
41. Xing C, Xie H, Zhou L, Zhou W, Zhang W, Ding S, et al. Cyclin-dependent kinase inhibitor 3 is overexpressed in hepatocellular carcinoma and promotes tumor cell proliferation. Biochem Biophys Res Commun. 2012;420(1):29-35.
42. Dai W, Miao H, Fang S, Fang T, Chen N, Li M. CDKN3 expression is negatively associated with pathological tumor stage and CDKN3 inhibition promotes cell survival in hepatocellular carcinoma. Mol Med Rep. 2016;14(2):1509-14.
43. Zhang S, Chang MC, Zylka D, Turley S, Harrison R, Turley EA. The hyaluronan receptor RHAMM regulates extracellular-regulated kinase. The Journal of biological chemistry. 1998;273(18):11342-8.
44. He X, Liao W, Li Y, Wang Y, Chen Q, Jin J, et al. Upregulation of hyaluronan-mediated motility receptor in hepatocellular carcinoma predicts poor survival. Oncol Lett. 2015;10(6):3639-46.
45. Yang CW, Su JY, Tsou AP, Chau GY, Liu HL, Chen CH, et al. Integrative genomics based identification of potential human hepatocarcinogenesis-associated cell cycle regulators: RHAMM as an example. Biochem Biophys Res Commun. 2005;330(2):489-97.
46. Liu YC, Lu LF, Li CJ, Sun NK, Guo JY, Huang YH, et al. Hepatitis B Virus X Protein Induces RHAMM-Dependent Motility in Hepatocellular Carcinoma Cells via PI3K-Akt-Oct-1 Signaling. Molecular cancer research : MCR. 2020;18(3):375-89.
47. Scholzen T, Gerdes J. The Ki-67 protein: from the known and the unknown. J Cell Physiol. 2000;182(3):311-22.
48. Cui J, Dong BW, Liang P, Yu XL, Yu DJ. Effect of c-myc, Ki-67, MMP-2 and VEGF expression on prognosis of hepatocellular carcinoma patients undergoing tumor resection. World J Gastroenterol. 2004;10(10):1533-6.
49. King KL, Hwang JJ, Chau GY, Tsay SH, Chi CW, Lee TG, et al. Ki-67 expression as a prognostic marker in patients with hepatocellular carcinoma. Journal of gastroenterology and hepatology. 1998;13(3):273-9.
50. Luo Y, Ren F, Liu Y, Shi Z, Tan Z, Xiong H, et al. Clinicopathological and prognostic significance of high Ki-67 labeling index in hepatocellular carcinoma patients: a meta-analysis. International journal of clinical and experimental medicine. 2015;8(7):10235-47.

51. Yang C, Su H, Liao X, Han C, Yu T, Zhu G, et al. Marker of proliferation Ki-67 expression is associated with transforming growth factor beta 1 and can predict the prognosis of patients with hepatic B virus-related hepatocellular carcinoma. Cancer Manag Res. 2018;10:679-96.
52. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics. 2012;28(4):573-80.
53. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.
54. Gaudet P, Dessimoz C. Gene Ontology: Pitfalls, Biases, and Remedies. Methods Mol Biol. 2017;1446:189-205.
55. du Plessis L, Skunca N, Dessimoz C. The what, where, how and why of gene ontology–a primer for bioinformaticians. Brief Bioinform. 2011;12(6):723-35.
56. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome biology. 2003;4(5):P3.
57. Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. 2007;35(Web Server issue):W169-75.
58. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(Database issue):D447-52.
59. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.
60. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003;4:2.
61. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8 Suppl 4(Suppl 4):S11.
62. Lindskog C. The potential clinical impact of the tissue-based map of the human proteome. Expert review of proteomics. 2015;12(3):213-5.
63. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science (New York, NY). 2015;347(6220):1260419.

Tables

Table 1. Information of DEGs screened from each GEO dataset.

GEO ID	Sample	Number of DEGs	Number of up-regulated genes	Number of down-regulated genes
GSE121248	hepatocellular carcinoma (HBV)	580	167	413
GSE19665	hepatocellular carcinoma (HBV)	647	257	390
GSE47197	hepatocellular carcinoma (HBV)	452	148	304
GSE55092	hepatocellular carcinoma (HBV)	1034	409	634

Abbreviation: GEO, Gene Expression Omnibus. DEGs, Differentially Expressed Genes.

Table2 Top 15 GO enrichment terms associated with the upregulated and downregulated DEGs.

A. Upregulated genes top 15 enriched GO terms.

Category	Term	Count	%	*P*-value
BP	GO:0051301~cell division	23	0.149031	1.24E-17
BP	GO:0007067~mitotic nuclear division	19	0.123113	1.38E-15
BP	GO:0007059~chromosome segregation	9	0.058317	3.36E-09
BP	GO:0007062~sister chromatid cohesion	10	0.064796	4.91E-09
BP	GO:0031145~anaphase-promoting complex-dependent catabolic process	8	0.051837	2.47E-07
CC	GO:0000775~chromosome, centromeric region	10	0.064796	1.41E-11
CC	GO:0005819~spindle	11	0.071276	7.18E-10
CC	GO:0000777~condensed chromosome kinetochore	9	0.058317	1.68E-08
CC	GO:0030496~midbody	10	0.064796	2.37E-08
CC	GO:0005634~nucleus	53	0.34342	3.03E-07
MF	GO:0005515~protein binding	72	0.466533	1.16E-06
MF	GO:0008017~microtubule binding	10	0.064796	2.17E-06
MF	GO:0019901~protein kinase binding	12	0.077755	6.88E-06
MF	GO:0005524~ATP binding	21	0.136072	1.53E-04
MF	GO:0042393~histone binding	5	0.032398	4.62E-03

B. Downregulated genes top 15 enriched GO terms.

Category	Term	Count	%	*P*-value
BP	GO:0006954~inflammatory response	10	0.036544	6.49E-04
BP	GO:0006955~immune response	10	0.036544	6.49E-04

BP	GO:0006958~complement activation, classical pathway	4	0.014618	6.70E-04
BP	GO:1900747~negative regulation of vascular endothelial growth factor signaling pathway	3	0.010963	7.20E-04
BP	GO:0070098~chemokine-mediated signaling pathway	5	0.018272	1.39E-03
CC	GO:0070062~extracellular exosome	55	0.200994	3.67E-10
CC	GO:0005579~membrane attack complex	5	0.018272	5.91E-08
CC	GO:0005615~extracellular space	26	0.095015	2.04E-05
CC	GO:0005576~extracellular region	17	0.062125	8.83E-05
CC	GO:0042567~insulin-like growth factor ternary complex	3	0.010963	6.54E-04
MF	GO:0020037~heme binding	11	0.040199	1.07E-06
MF	GO:0005506~iron ion binding	10	0.036544	3.47E-05
MF	GO:0008009~chemokine activity	6	0.021927	5.06E-05
MF	GO:0004497~monooxygenase activity	5	0.018272	4.52E-04
MF	GO:0005537~mannose binding	4	0.014618	7.25E-04

Abbreviation: BP, Biological Process; MF, Molecular Function; CC, cell component; DEGs, Differentially Expressed Genes.

Table3 KEGG pathway analysis of DEGs associated with HBV-related hepatocellular carcinoma.

Term	Count	%	P Value	Genes
bta00830: Retinol metabolism	8	0.01869	6.14E-05	CYP3A4, CYP2C18, CYP2B6, ADH4, CYP26A1, HSD17B6, CYP1A2, RDH16
bta00140: Steroid hormone biosynthesis	8	0.01869	6.89E-05	CYP3A4, HSD17B2, CYP2C18, CYP2B6, HSD17B6, SRD5A2, CYP1A2, AKR1D1
bta00380: Tryptophan metabolism	7	0.01636	1.65E-04	AADAT, TDO2, IDO2, ACMSD, KMO, CYP1A2, INMT
bta01100: Metabolic pathways	40	0.09346	2.30E-04	CYP3A4, XDH, CNDP1, FOLH1B, HSD17B2, CYP2C18, CYP2B6, COX7B2, ALDOB, KMO, ALDH3A1, TK1, GLS2, TDO2, ASPA, ADH4, GSTZ1, HSD17B6, CDA, ACSL4, HPD, AADAT, ST6GAL2, IDO2, FBP1, ACMSD, CYP26A1, TKT, CYP1A2, MAN1C1, TAT, ACSM3, GBA3, GGT5, SDS, AKR1B10, HAO2, AGXT2, RDH16, AKR1D1
bta05020: Prion diseases	6	0.01402	2.48E-04	C8A, EGR1, C8B, C7, C9, C6
bta04610: Complement and coagulation cascades	8	0.01869	3.63E-04	C8A, C8B, MBL2, C7, C9, KLKB1, C6, F9
bta04115: p53 signaling pathway	7	0.01636	1.74E-03	STEAP3, CCNE2, CCNB1, CDK1, TP53I3, IGF1, IGFBP3
bta04110: Cell cycle	9	0.02103	1.9E-03	CCNE2, CCNB1, CDK1, TTK, BUB1B, CDC20, PTTG1, MCM2
bta04114: Oocyte meiosis	8	0.01869	4.56E-03	CCNE2, CCNB1, CDK1, IGF1, AURKA, CDC20, PTTG1
bta00350: Tyrosine metabolism	5	0.01168	5.72E-03	ADH4, GSTZ1, TAT, ALDH3A1, HPD
bta00360: Phenylalanine metabolism	4	0.00935	6.12E-03	GLYAT, TAT, ALDH3A1, HPD
bta04976: Bile secretion	6	0.01402	7.64E-03	SLCO1B3, SLC22A7, KCNN2, CA2, SLC10A1, SLC22A1
bta05204: Chemical carcinogenesis	6	0.01402	8.62E-03	CYP3A4, CYP2C18, CYP2B6, ADH4, CYP1A2, ALDH3A1
bta00250: Alanine, aspartate and	4	0.00935	2.70E-02	GLS2, ASPA, FOLH1B, AGXT2

glutamate metabolism				
bta04060: Cytokine-cytokine receptor interaction	9	0.02103	4.44E-02	CCL2, CXCL14, CCL20, IL1RAP, LIFR, CCL19, IL7R, CXCL12, GHR

Abbreviation: KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, Differentially Expressed Genes.

Table4 The top 12 proteins ranked based on the node centrality of the PPI network.

Rank	Degree centrality		Betweenness centrality		Closeness centrality		MCC centrality	
	Gene symbol	Expression in HBV-HCC	Gene symbol	Expression in HBV-HCC	Gene symbol	Expression in HBV-HCC	Gene symbol	Expression in HBV-HCC
1	AURKA	up-regulated	IGF1	down-regulated	NDC80	up-regulated	CDK1	up-regulated
2	CDK1	up-regulated	SPP2	down-regulated	MKI67	up-regulated	AURKA	up-regulated
3	BUB1B	up-regulated	CYP3A4	down-regulated	FOXM1	up-regulated	CDC20	up-regulated
4	CDC20	up-regulated	NDC80	up-regulated	HMMR	up-regulated	BUB1B	up-regulated
5	MKI67	up-regulated	FOXM1	up-regulated	HJURP	up-regulated	CCNB1	up-regulated
6	UBE2C	up-regulated	CCL2	down-regulated	PTTG1	up-regulated	UBE2C	up-regulated
7	CCNB1	up-regulated	MKI67	up-regulated	CDK1	up-regulated	TTK	up-regulated
8	CDKN3	up-regulated	HJURP	up-regulated	CDKN3	up-regulated	TOP2A	up-regulated
9	HMMR	up-regulated	LCN2	up-regulated	IGF1	down-regulated	NUSAP1	up-regulated
10	NUSAP1	up-regulated	HMMR	up-regulated	AURKA	up-regulated	CDKN3	up-regulated
11	TOP2A	up-regulated	ANG	down-regulated	CCNB1	up-regulated	ASPM	up-regulated
12	TTK	up-regulated	SPP1	up-regulated	CDC20	up-regulated	DLGAP5	up-regulated

Abbreviation: PPI, Integration of Protein–protein Interactions.

Table 5 KEGG enrichment of genes in the top 2 modules.

Module	Term	Count	%	P value	Genes
1	cfa04110: Cell cycle	8	0.087787	2.36E-08	CCNE2, CCNB1, CDK1, BUB1B, TTK, CDC20, PTTG1, MCM2
	cfa04114: Oocyte meiosis	5	0.054867	2.00E-04	CCNE2, CDK1, CDC20, AURKA, PTTG1
	cfa04115: p53 signaling pathway	3	0.03292	1.41E-02	CCNE2, CCNB1, CDK1
	cfa05222: Small cell lung cancer	3	0.03292	2.23E-02	CCNE2, CKS1B, CKS2
2	cfa04610: Complement and coagulation cascades	6	0.312337	9.64E-11	C8A, MBL2, C8B, KLKB1, C6, F9
	cfa05020: Prion diseases	3	0.156169	2.16E-04	C8A, C8B, C6
	cfa05322: Systemic lupus erythematosus	3	0.156169	2.46E-03	C8A, C8B, C6

Abbreviation: KEGG, Kyoto Encyclopedia of Genes and Genomes.

Table 6. Details of four HBV-related HCC datasets from GEO.

GEO ID	Platform	Tissue type	Normal	Tumor	Country	Time	PMID
GSE121248	GPL 570 [HGU133_Plus_2] Affymetrix Human	hepatocellular carcinoma (HBV)	37	70	Singapore	2018	17975138
GSE19665	GPL 570 [HGU133_Plus_2] Affymetrix Human	hepatocellular carcinoma (HBV)	5	5	Japan	2010	20345479
GSE47197	GPL16699 Agilent-039494 Sure Print G3 Human GE v2	hepatocellular carcinoma (HBV)	63	61	France	2013	/
GSE55092	GPL 570 [HGU133_Plus_2] Affymetrix Human	hepatocellular carcinoma (HBV)	91	49	USA	2014	25141867

Abbreviation: GEO, Gene Expression Omnibus.

Figures

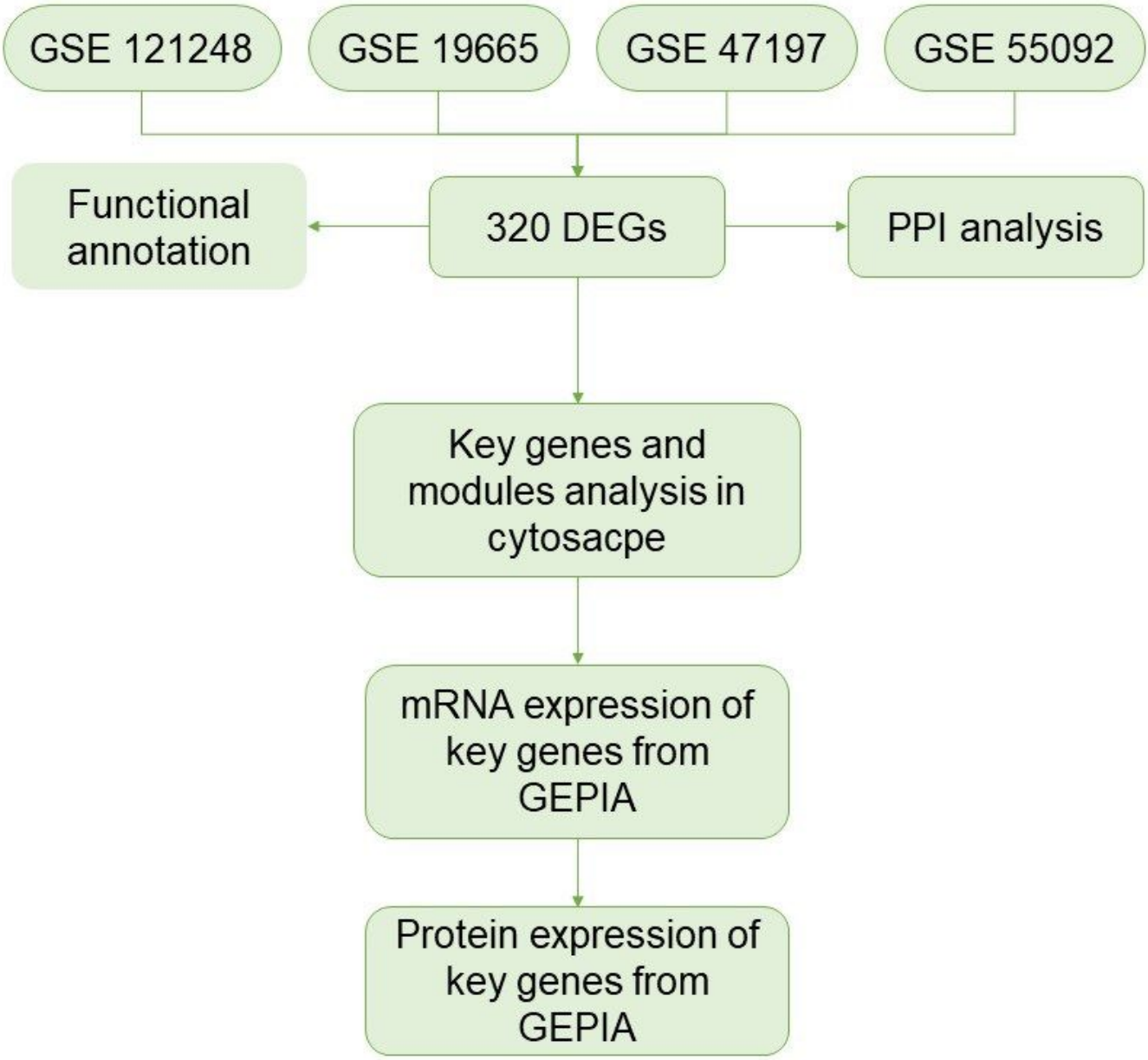

Figure 1

The workflow diagram to identify key genes and pathways associated with HBV-related HCC. Abbreviation: DEGs, differentially expressed genes.

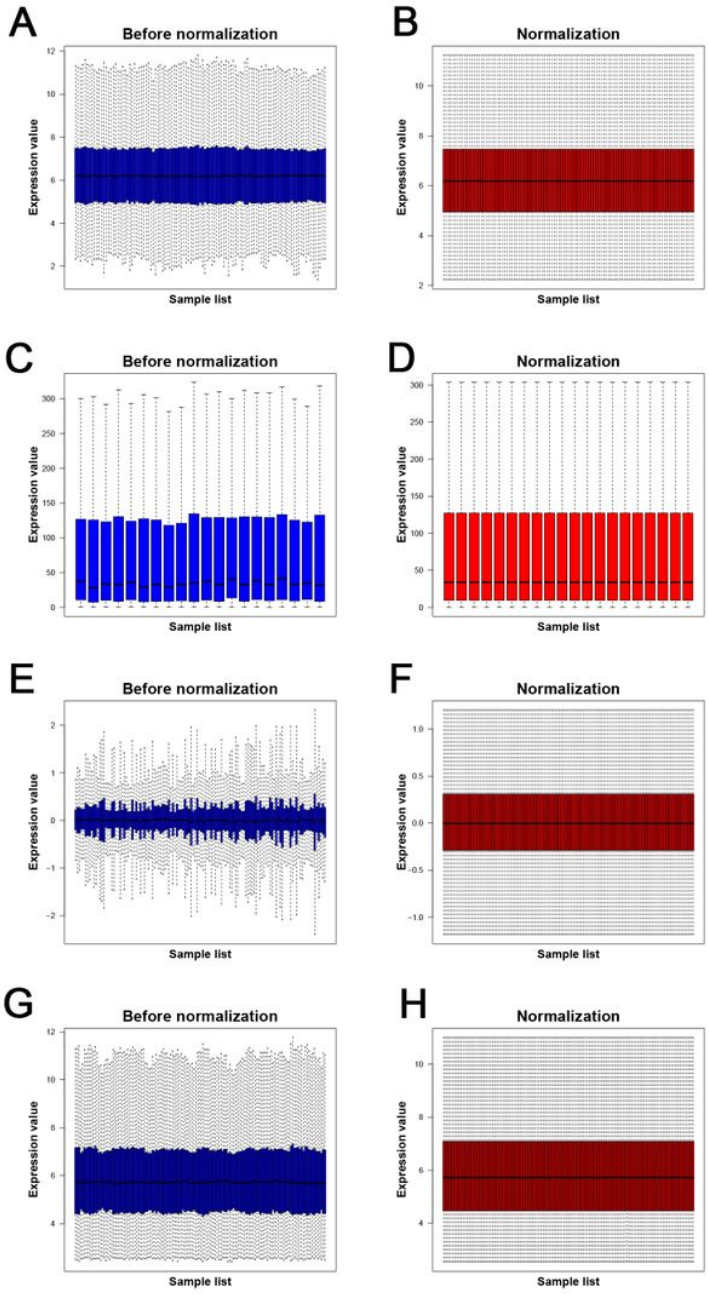

Figure 2

Normalization of gene expression in four GEO datasets. (A–B) Normalization of the data set GSE121248. (C–D) Normalization of the data setGSE19665. (E–F) Normalization of the data set GSE47197. (G–H) Normalization of the data set GSE55092. The blue bars represent data before normalization, and the red bars represent data after normalization.

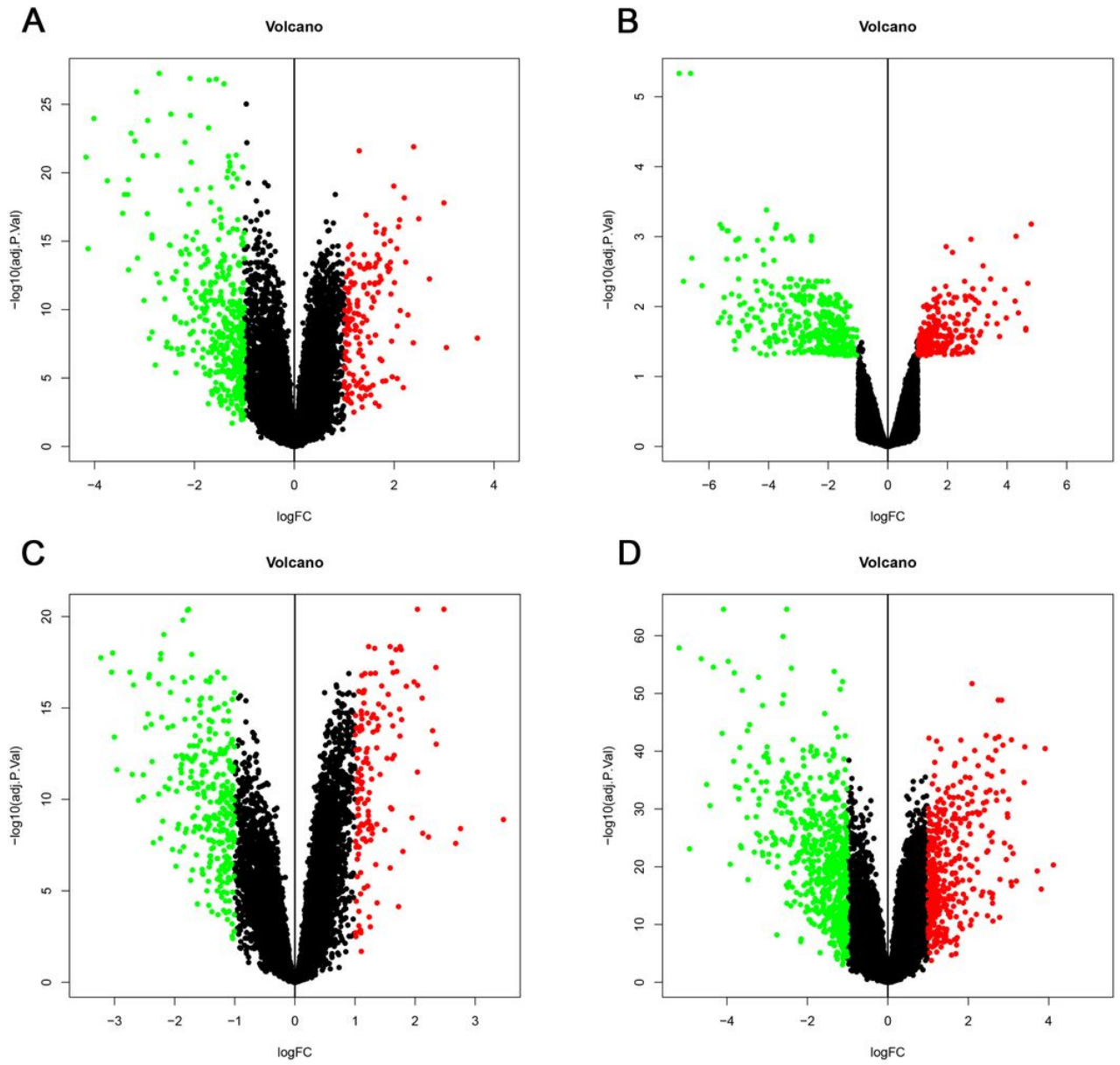

Figure 3

Volcano plots of differential expression analysis for four GEO datasets. (A) GSE121248, (B) GSE19665, (C) GSE47197 and (D) GSE55092. The red dots represent the upregulated genes based on an adjusted P < 0.05 and |log fold change| > 1; the green dots represent the downregulated genes based on an adjusted P < 0.05 and |log fold change| > 1; the black spots represent genes with no significant difference in expression. Abbreviation: DEGs, differentially expressed genes; GEO, Gene Expression Omnibus.

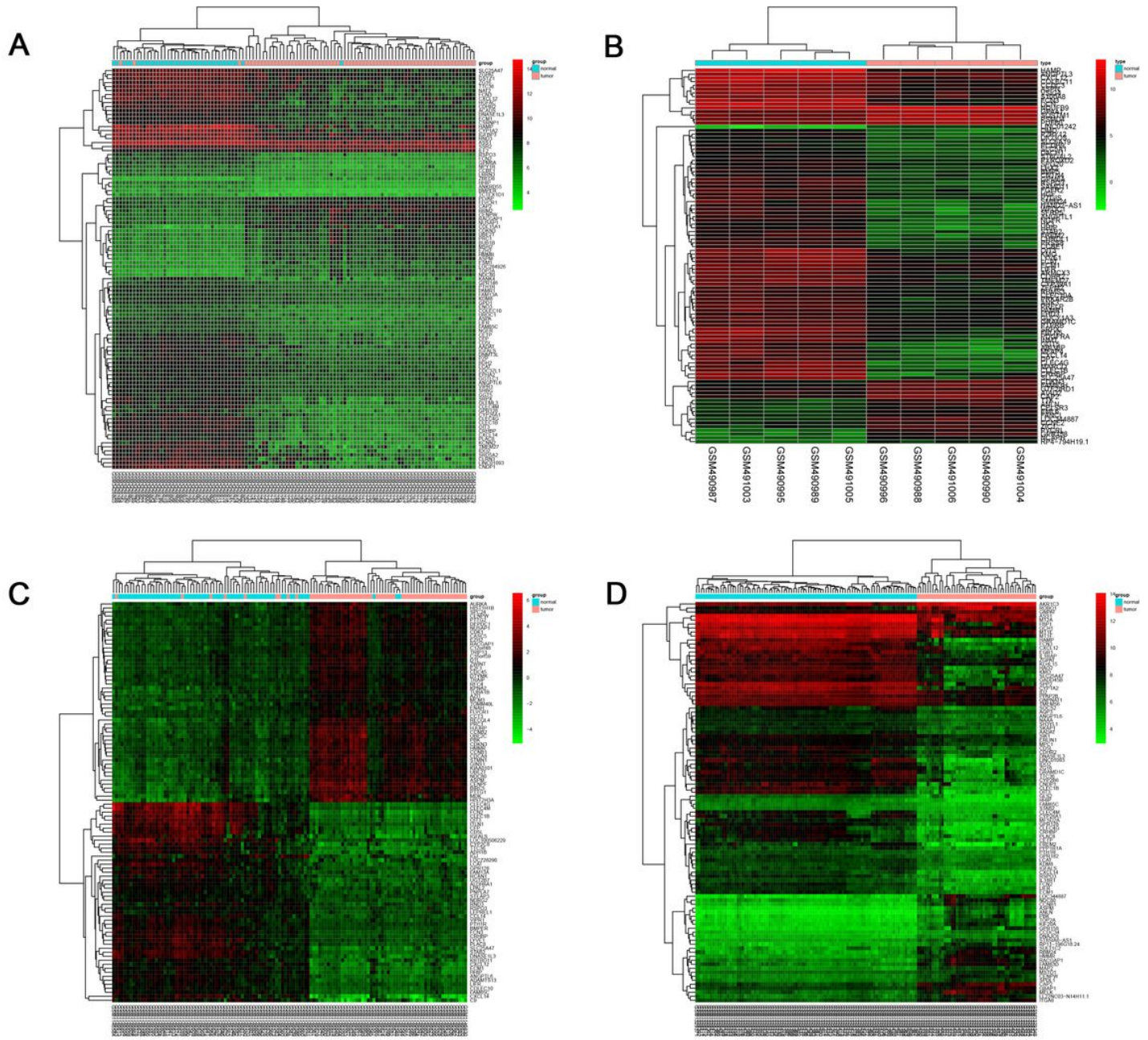

Figure 4

Heatmaps of the top 100 DEGs for four GEO datasets. (A) GSE121248, (B) GSE19665, (C) GSE47197 and (D) GSE55092. Hierarchical clustering heatmap of the 100 representative DEGs screened based on adjust P-value < 0.05 and |log fold change (FC)| > 1. The rows showed that the 100 representative DEGs according adjust P-value, including top 50 upregulated expressed genes, and the last 50 downregulated expressed genes in HBV-related HCC. The abscissa represents the GEO IDs, the ordinate represents the gene name, the red represents log FC > 1, the green represents log FC < -1 and the value in the box represents the log FC value. Abbreviation: DEGs, differentially expressed genes; GEO, Gene Expression Omnibus.

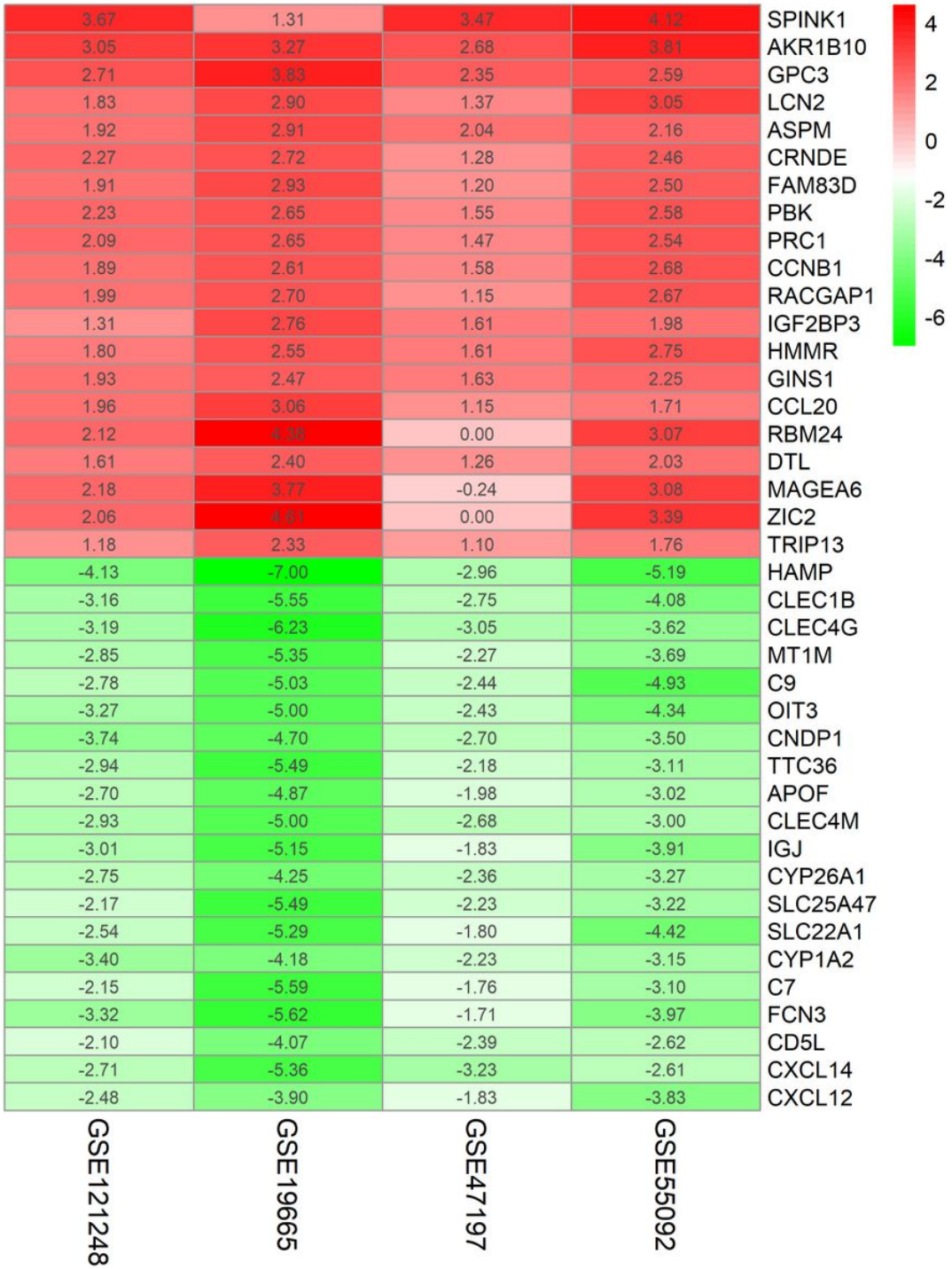

Figure 5

Heatmap of the top 40 DEGs expression in four GEO datasets by RRA method. Hierarchical clustering heatmap of the 40 representative DEGs screened based on adjust P-value < 0.05 and |log fold change (FC)| > 1. The rows showed that the 20 representative DEGs according adjust P-value, including top 20 upregulated expressed genes, and the last 50 downregulated expressed genes in HBV-related HCC. The abscissa represents the GEO IDs, the ordinate represents the gene name, the red represents log FC > 1, the green represents log FC < -1 and the value in the box represents the log FC value. Abbreviation: DEGs, differentially expressed genes; GEO, Gene Expression Omnibus; RRA, Robust Rank Aggreg.

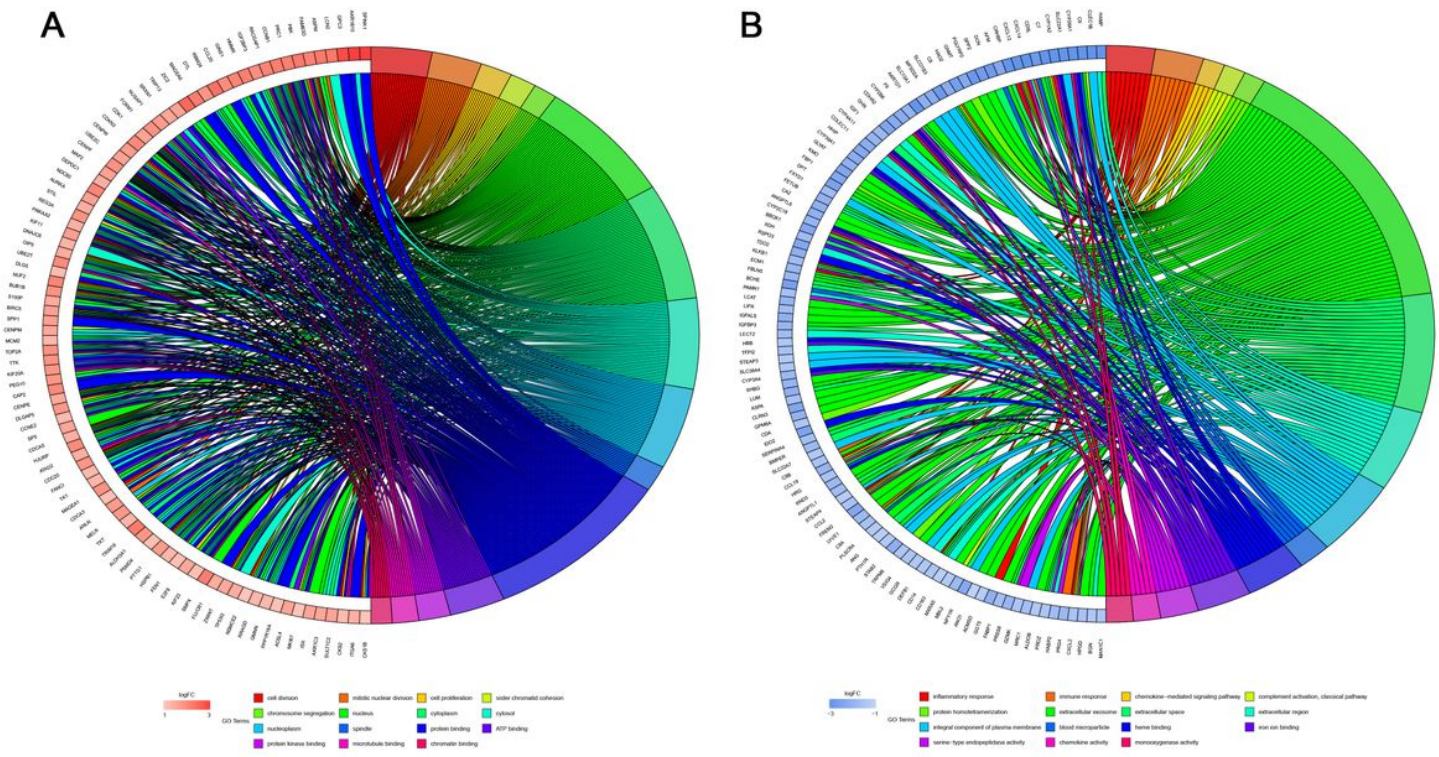

Figure 6

Distribution of 320 DEGs in HBV-related HCC for different GO-enriched functions. (A) Upregulated DEGs. (B) Downregulated DEGs. Abbreviation: DEGs, differentially expressed genes; GEO, Gene Expression Omnibus.

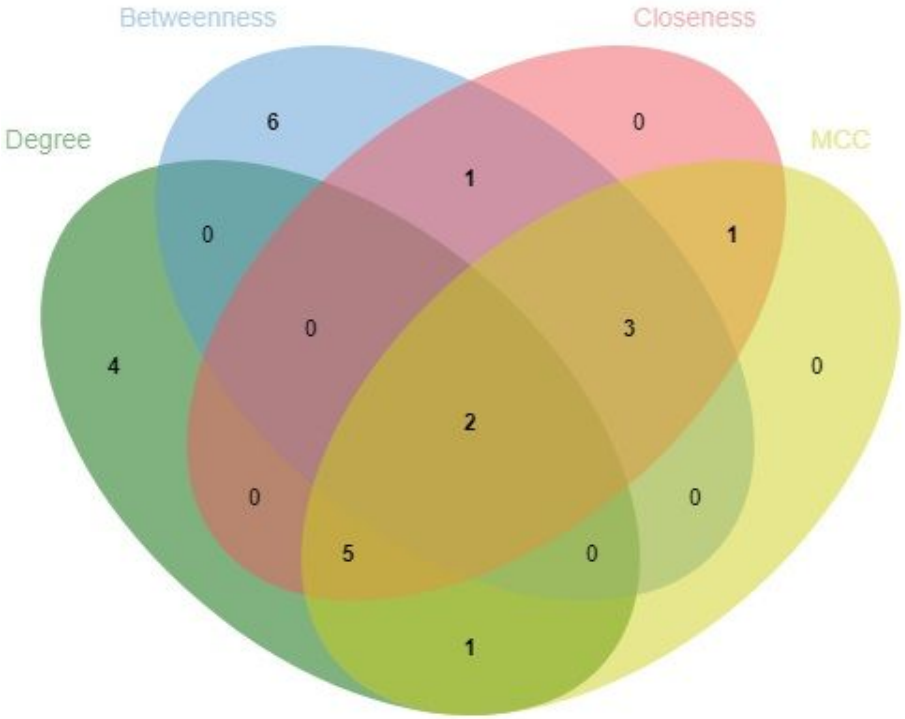

Figure 7

Venn's diagram of top 12 proteins in four centrality parameters. Intersection of top 12 proteins in four centrality parameters from cytohubba in cytoscape software. Abbreviation: MCC, Maximal Clique Centrality.

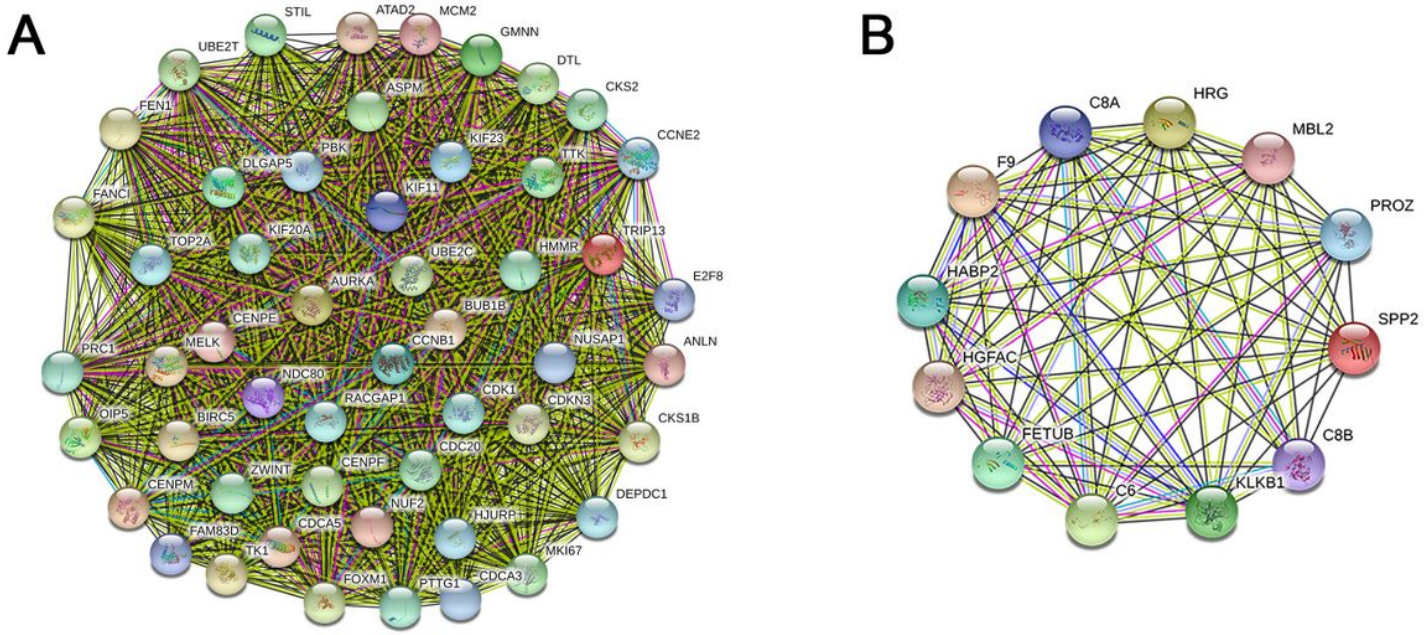

Figure 8

PPI network of two main module from 320 DEGs. (A) PPI network of module 1, (B) PPI network of module 2. Circles represent genes, lines represent interactions between gene-encoded proteins and line colors represent evidence of interactions between proteins. Abbreviation: DEGs, differentially expressed genes; PPI, integration of protein–protein interactions.

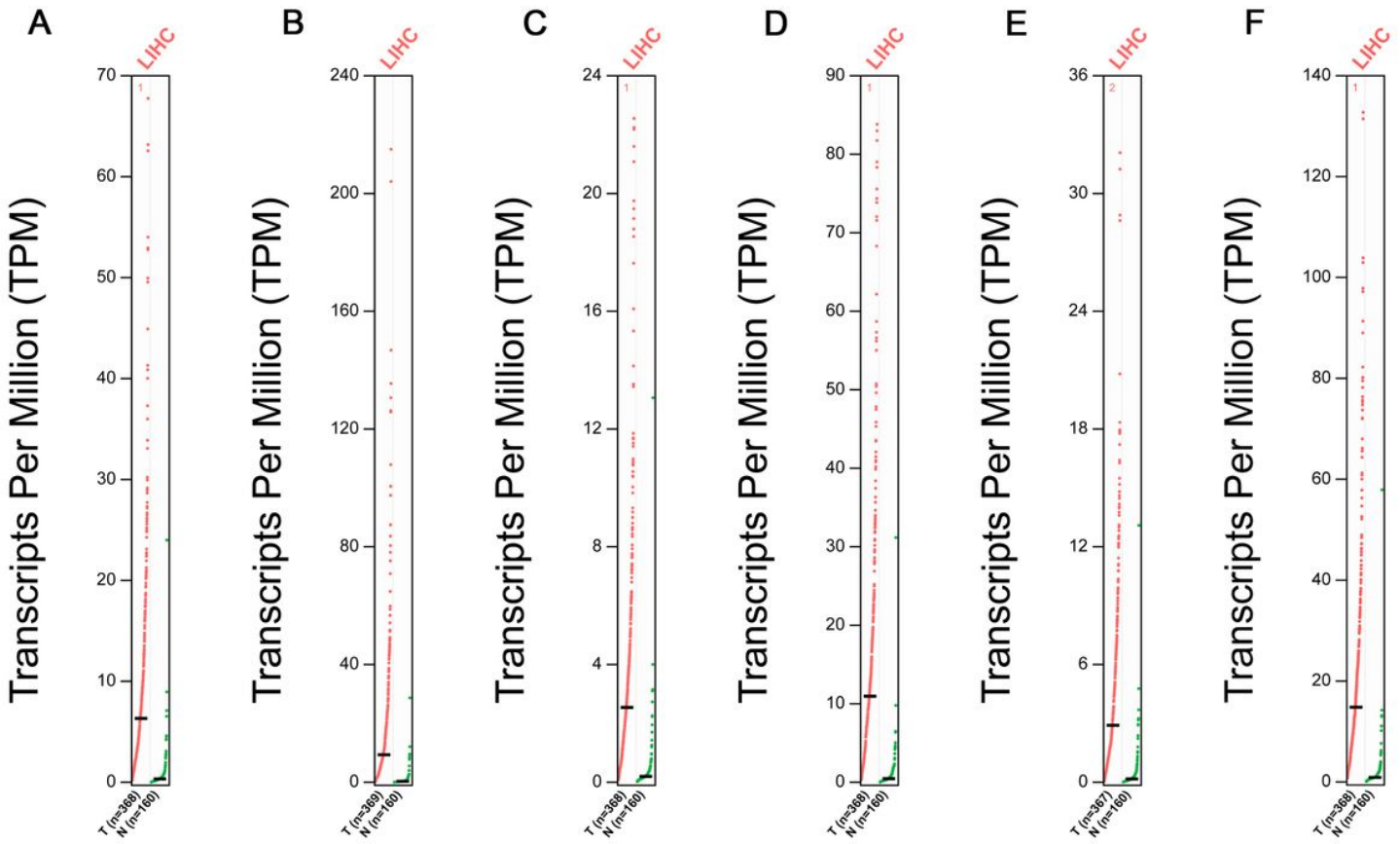

Figure 9

The scatter diagram of expression of six identified Genes in HCC. Validation of (A) CDK1, (B) CDC20, (C) HMMR, (D) CDKN3, (E) MKI67 and (F) CCNB1 in HBV-related HCC based on TCGA dataset from GEPIA. Abbreviation: GEPIA, Gene Expression Profiling Interactive Analysis.

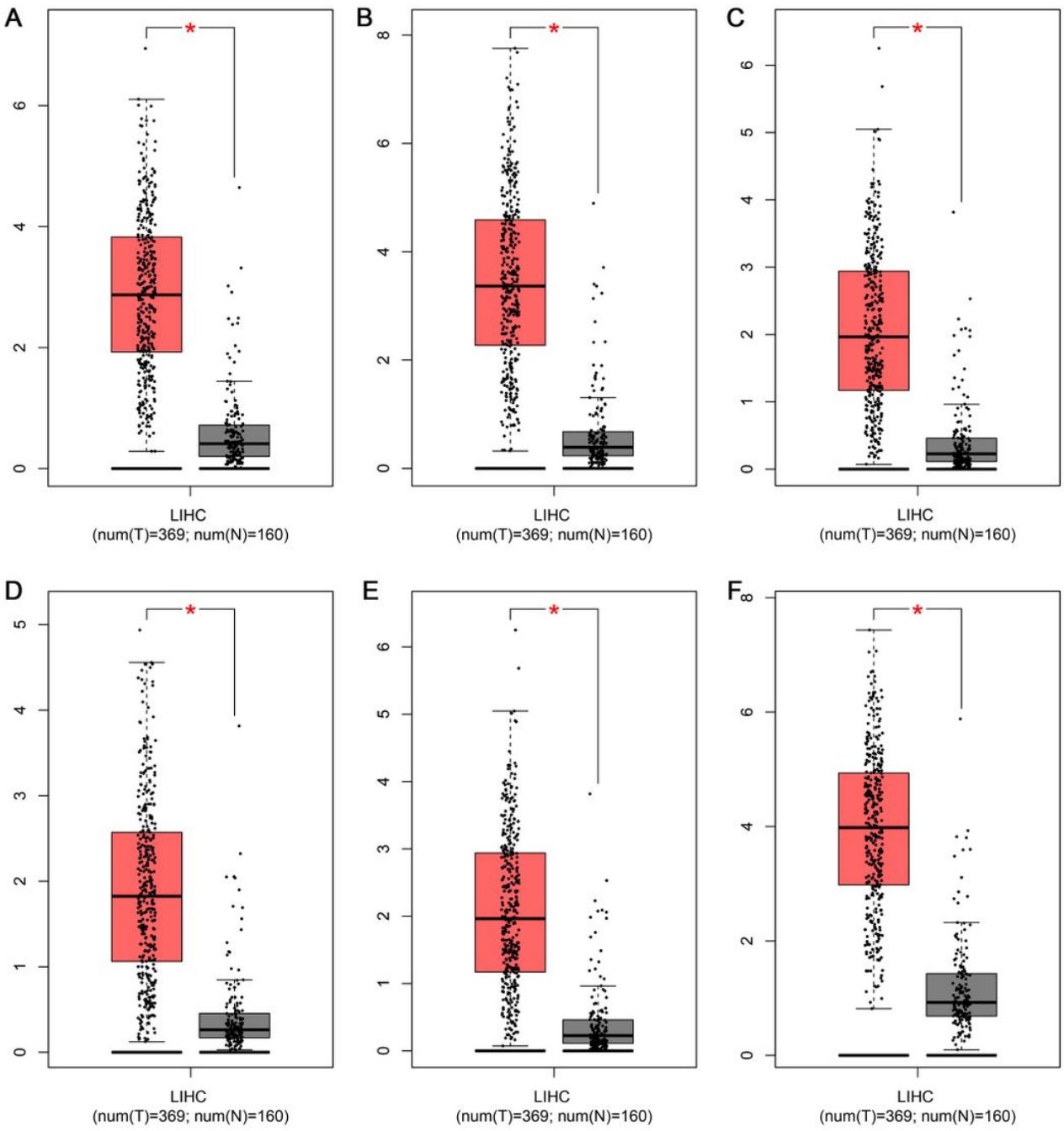

Figure 10

The box plot of expression of six identified hub in HCC patients. Validation of (A) CDK1, (B) CDC20, (C) HMMR, (D) CDKN3, (E) MKI67 and (F) CCNB1 in HBV-related HCC based on TCGA dataset from GEPIA. Red box represents tumor, grey box normal. Y-axis units are -Log2 (TPM + 1). *P<0.05. Abbreviation: GEPIA, Gene Expression Profiling Interactive Analysis.

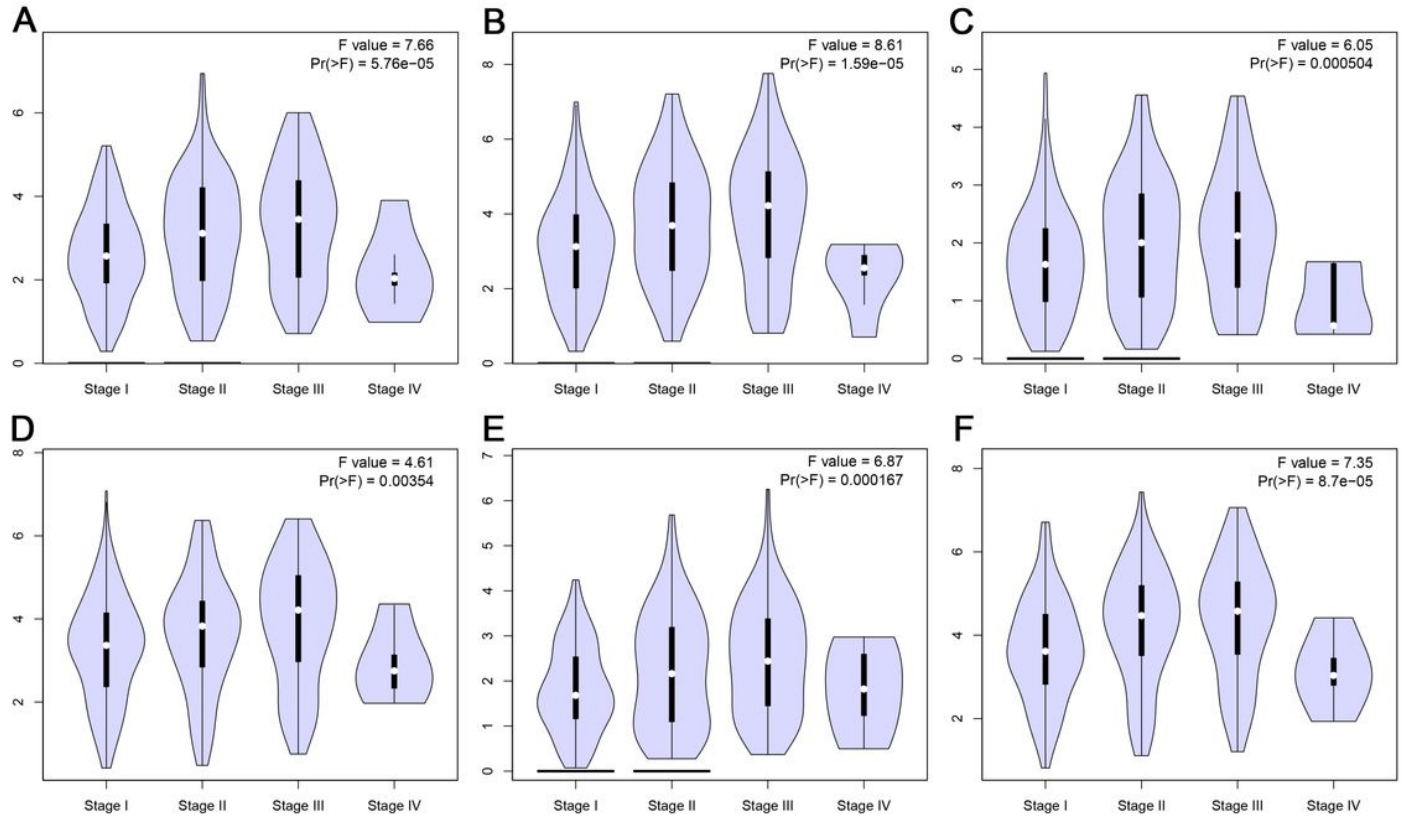

Figure 11

Correlation between six identified hub genes expression and tumor stage in HCC patients. Validation of (A) CDK1, (B) CDC20, (C) HMMR, (D) CDKN3, (E) MKI67 and (F) CCNB1 in HBV-related HCC based on TCGA dataset from GEPIA. The Pr (>F) value is equivalent to the P-value, and Pr (>F) <0.05 is considered statistically significant. Abbreviation: GEPIA, Gene Expression Profiling Interactive Analysis.

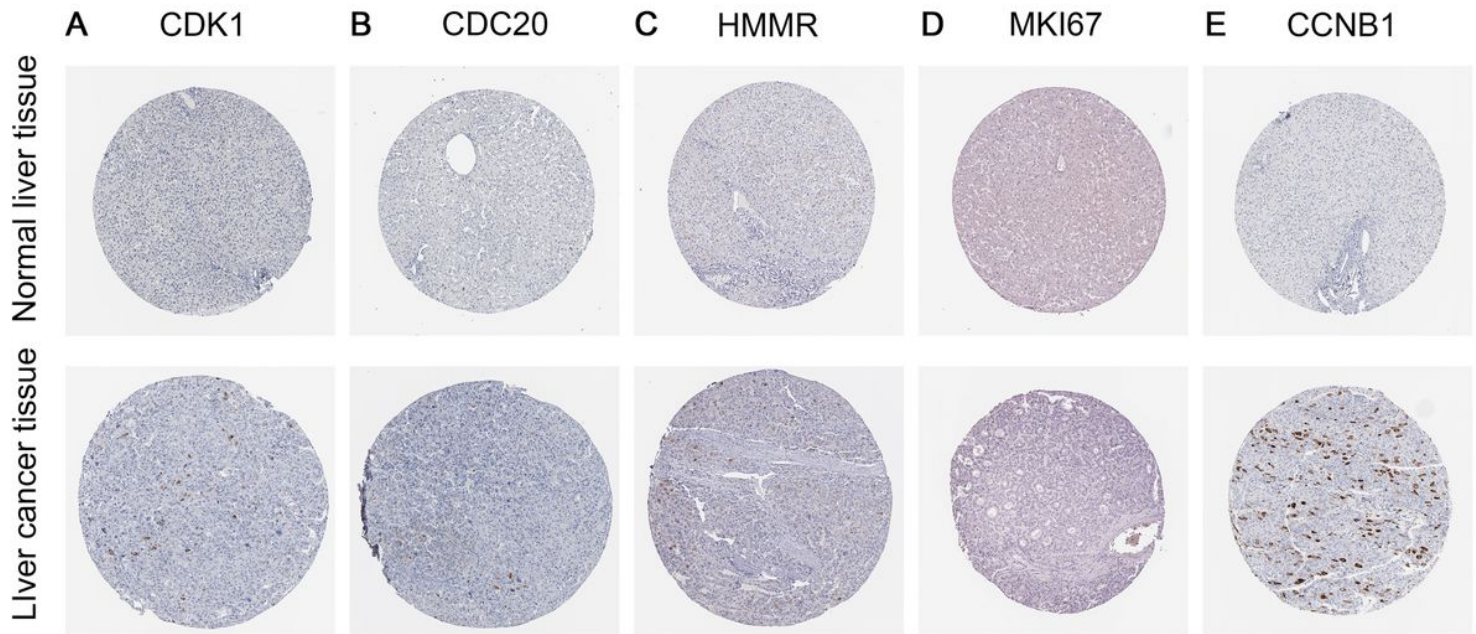

Figure 12

The protein expression of six identified hub in HPA database. The immunohistochemistry staining of (A) CDK1, (B) CDC20, (C) HMMR, (D) CDKN3, (E) MKI67 and (F) CCNB1. Abbreviation: HPA, Human Protein Atlas.

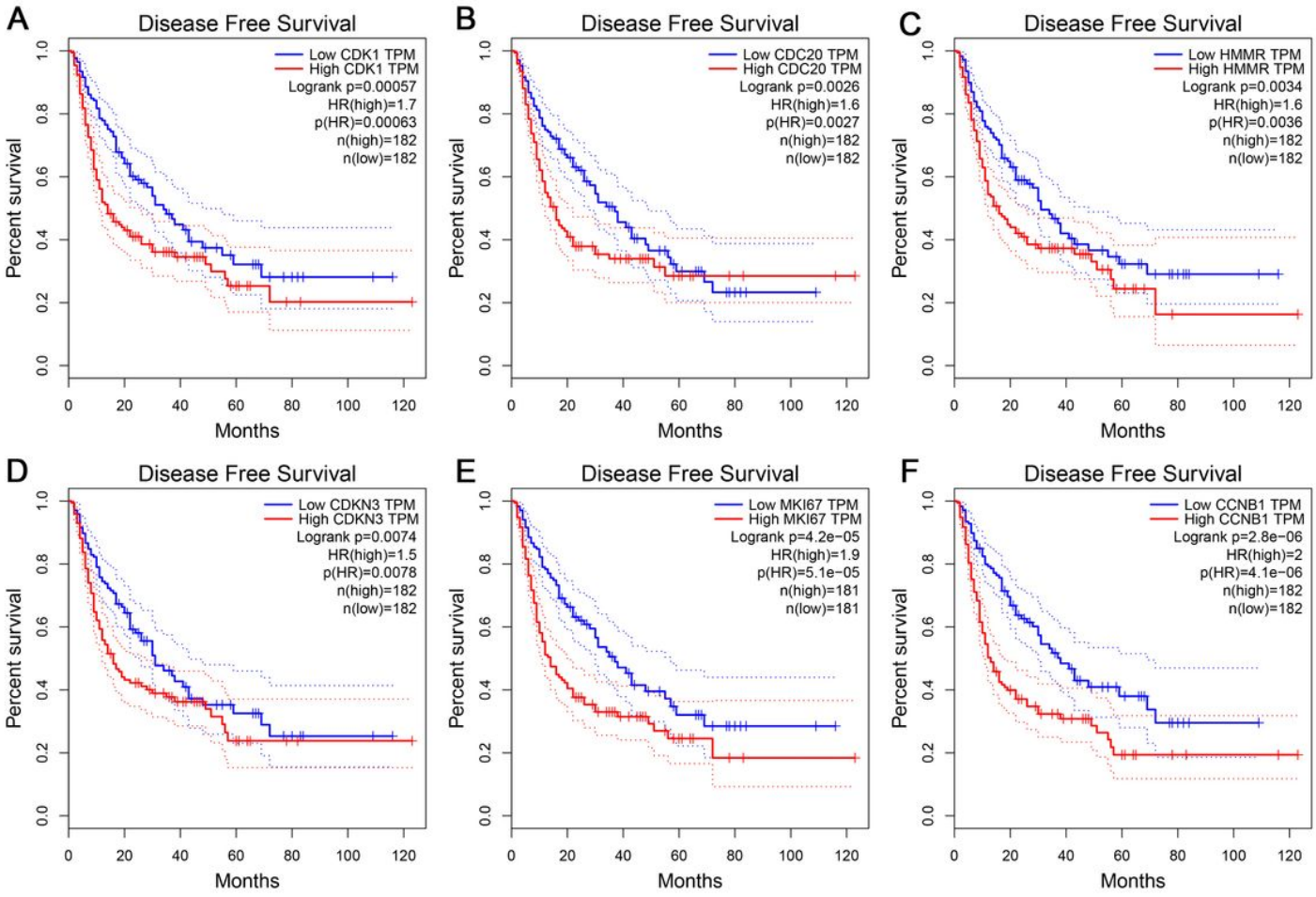

Figure 13

Disease-free survival curves of six identified hub genes in HCC Patients. Disease-free survival analyses of hub genes (A) CDK1, (B) CDC20, (C) HMMR, (D) CDKN3, (E) MKI67 and (F) CCNB1 were performed using GEPIA online platform. $P < 0.05$ was considered statistically significant.

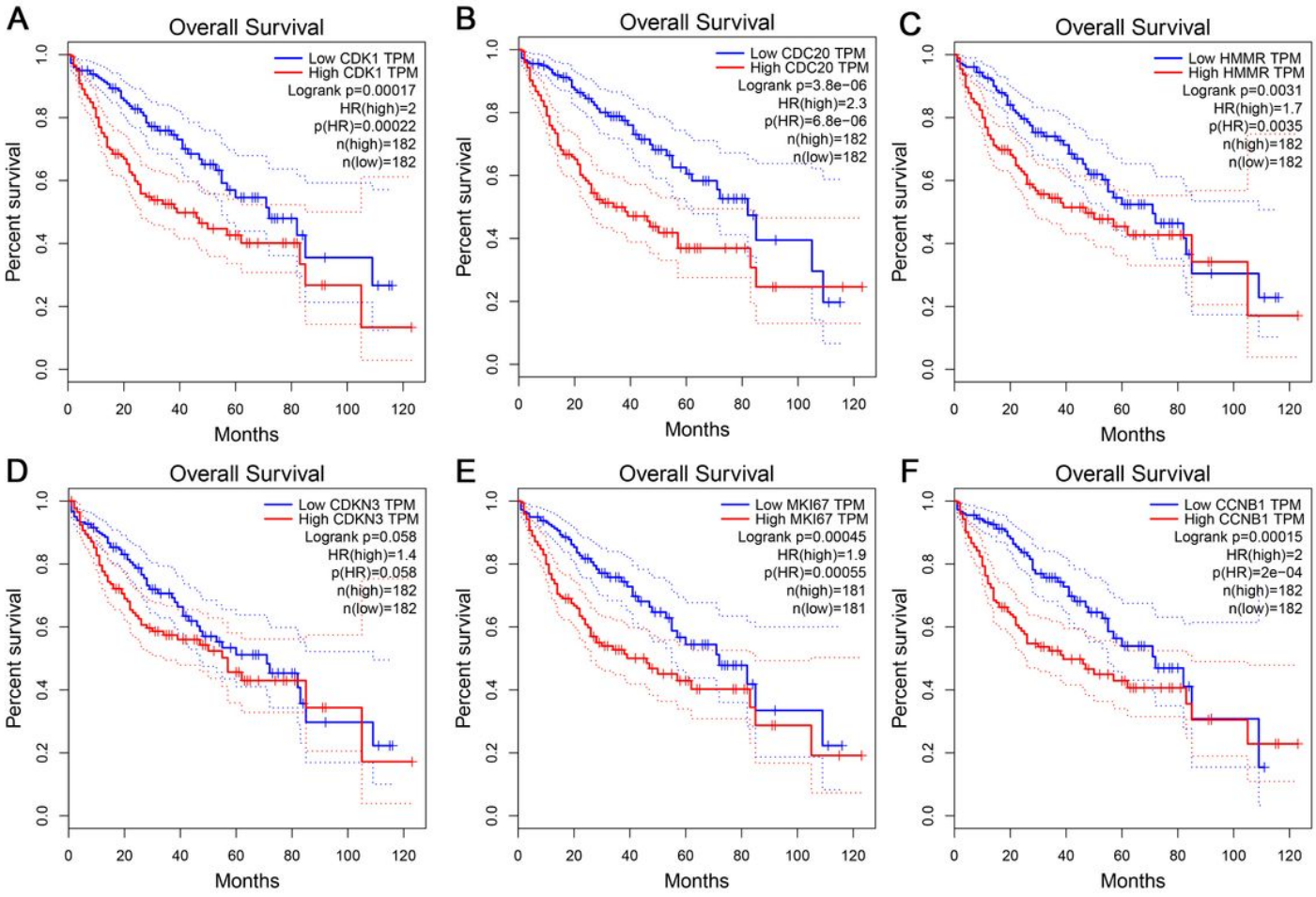

Figure 14

Overall survival curves of six identified hub genes in HCC Patients. Overall survival analyses of hub genes (A) CDK1, (B) CDC20, (C) HMMR, (D) CDKN3, (E) MKI67 and (F) CCNB1 were performed using GEPIA online platform. $P < 0.05$ was considered statistically significant.