A Bioinformatics-Based Screening and Analysis of Key Genes in Hepatocellular Carcinoma

Yanfei Wei
Department of Physiology, Guangxi University of Chinese Medicine, Nanning, Guangxi, China, 530200
Shengshan Wang
Department of Physiology, Guangxi University of Chinese Medicine, Nanning, Guangxi, China, 530200
Wanjun Chen
Department of Physiology, Guangxi University of Chinese Medicine, Nanning, Guangxi, China, 530200
Yuning Lin
Department of Physiology, Guangxi University of Chinese Medicine, Nanning, Guangxi, China, 530200
Qi Zhang
Department of Physiology, Guangxi University of Chinese Medicine, Nanning, Guangxi, China, 530200
Delun Huang (huangdelun0621@foxmail.com)
GuangXi University of Chinese Medicine https://orcid.org/0000-0002-7449-0386

Research

Keywords: hepatocellular carcinoma, bioinformatics, Key gene

Posted Date: December 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-131655/v1

A Bioinformatics-Based Screening and Analysis of Key Genes in Hepatocellular Carcinoma

YanFei Wei[*], ShengshanWang[*], Wanjun Chen, Yuning Lin, Qi Zhang, Delun Huang[#]

Department of Physiology, Guangxi University of Chinese Medicine, Nanning, Guangxi, China, 530200

[*#]the authors contributed equally to this work.

Keywords: hepatocellular carcinoma; bioinformatics; Key gene

Address correspondence:

Delun Huang, Ph.D

Department of Physiology, Guangxi University of Chinese Medicine, Nanning, Guangxi 530001, China

Phone: 0771-4733794

Email: huangdelun0621@foxmail.com

Abstract

Hepatocellular carcinoma (HCC) is considered as the leading killer disease in the world. So far most of the diagnosis of HCC is mainly established on imaging and biopsy. As sequencing technology is developing quite fast, and it has already been widely applied in the medical area, such as cancer diagnosis. In this article, GSE121248, GSE76427 and GSE60502 datasets were chosen to analyze and screen key genes which could affect the development of liver cancer through the bioinformatics method. The results showed up regulated genes mainly reside in cell division, nucleus, protein binding pathway, and down regulated genes are mostly located in the Oxidation-reduction process, Extracellular region, Heme binding, Metabolic pathway. Secondly, hub gene analysis indicated there were twelve critical hub genes found: RFC4, RACGAP1, CCNB2, CDC20, UBE2C, PTTG1, AURKA, PRC1, NCAPG, CDKN3, TOP2A, KIF20A, AURKA and CDKN3. By applying bioinformatic measures, the genes associated with hepatocellular carcinoma can be efficiently analyzed, that would provide invaluable information for translational studies.

Introduction

Hepatocellular carcinoma (HCC) is the second leading cause of cancer-related deaths in the world[1]. HCC is characterized by high mortality rates, high metastasis and high invasiveness[2]. It is generally believed that the occurrence of HCC is associated with viral infection (hepatitis C virus, hepatitis B virus), long-term alcohol consumption, Aflatoxin

infection, water contamination, nitrous acid substances, primary biliary cirrhosis and non-alcoholic fatty liver, etc[3]. The treatment of early HCC mainly includes surgical resection, liver transplantation, percutaneous radiofrequency ablation, and probably more than 50% of patients have a five-year survival rate. Because the pathogenesis of HCC involves multiple signaling pathways, and the molecular mechanisms of malignant progression are unclear. Almost 80% of advanced-stage liver cancer patients have mainly received radiotherapy and chemotherapy treatment. Therefore, identifying effective early tumor diagnostic markers and targeted therapeutic targets is of great clinical significance [4].

The development of high-throughput gene chip and sequencing technology has helped rapid the study of the gene expression profile of liver cancer, facilitating the discovery of gene and gene expression changes of liver cancer tissues and cells subjected to specific conditions. Due to the low reliability of positive results from a single microarray data analysis, in this study, we analyzed three microarray datasets from the GEO database, each of them includes HCC samples and non-tumor containing liver tissue samples. With the aid of GEO2R, differential expressed genes (DEG) between HCC and normal samples were analyzed. Applying several bioinformatics tools such as DAVID and String, the DEG's biological functions, involved signaling pathways and interactions were worked out. Finally, 12 key genes were selected to be candidate biomarkers for HCC. This study provides a theoretical basis for the clinical screening of molecular markers and drug targets for the development and progression of hepatocellular carcinoma (Fig.1).

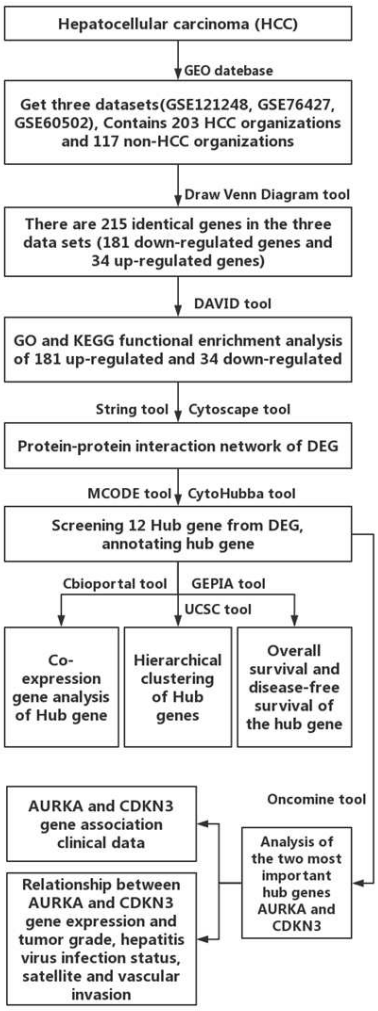


Fig. 1 The schematic of this study is intended to describe bioinformatics screening and analysis of key genes for HCC.

1. Methods and materials

1.1 Obtaining Microarray Data The National Center for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) is the largest public repository of high-throughput gene expression data[5]. To expand the sample size and improve the reliability of screening, three datasets of Homo tissue samples with both hepatic carcinoma and normal hepatic tissues included were selected. In our case, GSE121248[6], GSE76427[7] and GSE60502[8] datasets were screened out from the GEO database platform GPL570 ([HG-U133_Plus_2], Affymetrix Human Genome U133 Plus 2.0 Array), PL10558

(Illumina HumanHT-12 V4.0 expression bead chip) and GPL96 ([HG-U133A] Affymetrix Human Genome U133A Array) respectively. GSE121248 contained 70 HCC samples and 37 adjacent non-tumor hepatic tissue samples, GSE76427 contained 115 HCC samples and 52 adjacent non-tumor hepatic tissue samples, while GSE60502 contained 18 HCC samples and 18 adjacent non-tumor hepatic tissue samples.

1.2 Identification of DEG The GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) is an R-based web application that enables to analyze GEO data between HCC samples and non-cancer samples. The blank genome in each data set was removed, and genes with shared IDs were also eliminated. A value of |logFC| (fold change)>1 and P-value<0.05 was taken to be statistically significant. The Draw Venn Diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) tool was then used to select upregulated and downregulated genes common to all three datasets.

1.3 GO and KEGG pathway enrichment analysis of DEG DAVID (https://david.ncifcrf.gov/) (version 6.8)[9] is a web-based online bioinformatics resource which serves mainly to functionally interpret a large number of genes and protein. The standard annotations for GO[10] biological functions include Biological Processes (BP), Molecular Functions (MF) and Cellular Components (CC). KEGG[11] is a database resource used particularly for genome sequencing and other high-throughput, large-scale molecular information processing. Combining both GO and KEGG functions of the DAVID online tool, DEG were analyzed with $p<0.05$ taken to be statistically significant.

1.4 Protein-Protein Interaction (PPI) Network Analysis The String (https://string-db.org/)

(version 11.0) [12]database is a database that searches for interactions between known proteins and predicted proteins. In this experiment, multiple protein input forms were selected and DEG data were imported, with Homo sapiens selected as the organism form. Setting parameters: medium confidence=0.4, and default other parameters. Default all parameters, then export the PPI data in TSV format. Cytoscape (version 3.7.1)[13] is a standard application for bio-network analysis and visualization. The TSV text saved by String was imported into cytoscape to create a PPI visualization network diagram. Using the MCODE (version 1.5.1)[14] plugin in the Cytoscape database, setting parameters: k-core=2, Node score cutoff=0.2, degree cutoff=2, and default other parameters. key genes of the PPI network were filtered, and the closest set of key genes was selected. To increase the sensitivity and specificity, we use MCC to discover featured nodes, We selected the Maximal Clique Centrality algorithm (MCC) of the cytohubba plugin [15] in the Cytoscape tool to screen the top Hub genes.

1.5 Hub gene selection and analysis cBioPortal for Cancer Genomics (http://cbioportal.org)[16] is a web resource for exploring, visualizing, and analyzing multidimensional cancer genomics data. This platform was used to analyze the 12 selected key genes and co-expressed gene networks. The Gene Expression Profiling Interactive Analysis (GEPIA) online database (http://gepia.cancer-pku.cn/) is a web server with customizable features; Tumor and normal samples in the GEPIA database[17] were derived from The Cancer Genome Atlas (TCGA) and Genotype Tissue Expression (GTEx) programs. GEPIA was used to analyze the overall survival levels and expression levels in disease-free survival. Oncomine (https://www.oncomine.org/)[18] was used to analyze DEG expression in

hepatic cancer tissues and non-hepatocarcinoma tissues, as well as AURKA and CDKN3 expression, tumor grade, hepatitis virus infection status, satellite and vascular invasion relationships.

2 results

2.1 Article ideas and structure

2.2 DEG screening results The GSE121248, GSE76427, and GSE60502 datasets were screened out from the GEO database, 1153, 554, 1728 DEGs was found respectively. Finally, an online Venn Diagram tool yielded 215 co-expression DEGs containing 34 common upregulated genes and 181 downregulated genes from the three datasets. (Tab. S1-3, Fig. 2).

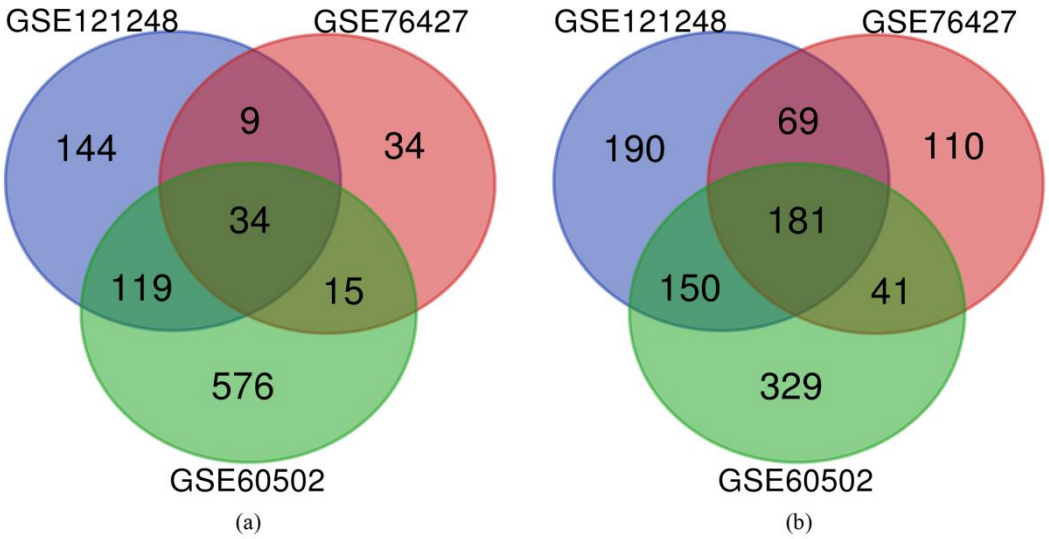


Fig. 2 215 DEG obtained from the three datasets. **(a)** 34 common upregulated genes. **(b)** 181 common downregulated genes. The different colors represent different datasets, and each intersecting area represents a gene shared between datasets. DEGs with |logFC|(fold change)>1 and adj.p-value<0.05 were selected.

2.3 Analysis of GO enrichment results The upregulated genes and downregulated genes were input separately into the DAVID Database for GO and KEGG functional enrichment analysis. GO analysis revealed the BP, CC and MF of the first ten upregulated and

downregulated genes (Fig. 3, Tab. S4). The biological processes that were significantly enriched by upregulated genes include cell division, mitotic nuclear division and cell proliferation. In terms of cellular components, the upregulated genes reside mainly in the nucleus, cytoplasm and nucleoplasm. The molecular functions that were most heavily involved were protein binding and ATP binding. For downregulated genes, the biological processes most commonly affected were the oxidation-reduction process and immune response, while heme binding and oxidoreductase activity were the main molecular functions implicated. Downregulated genes were commonly found in the extracellular region and extracellular exosome.

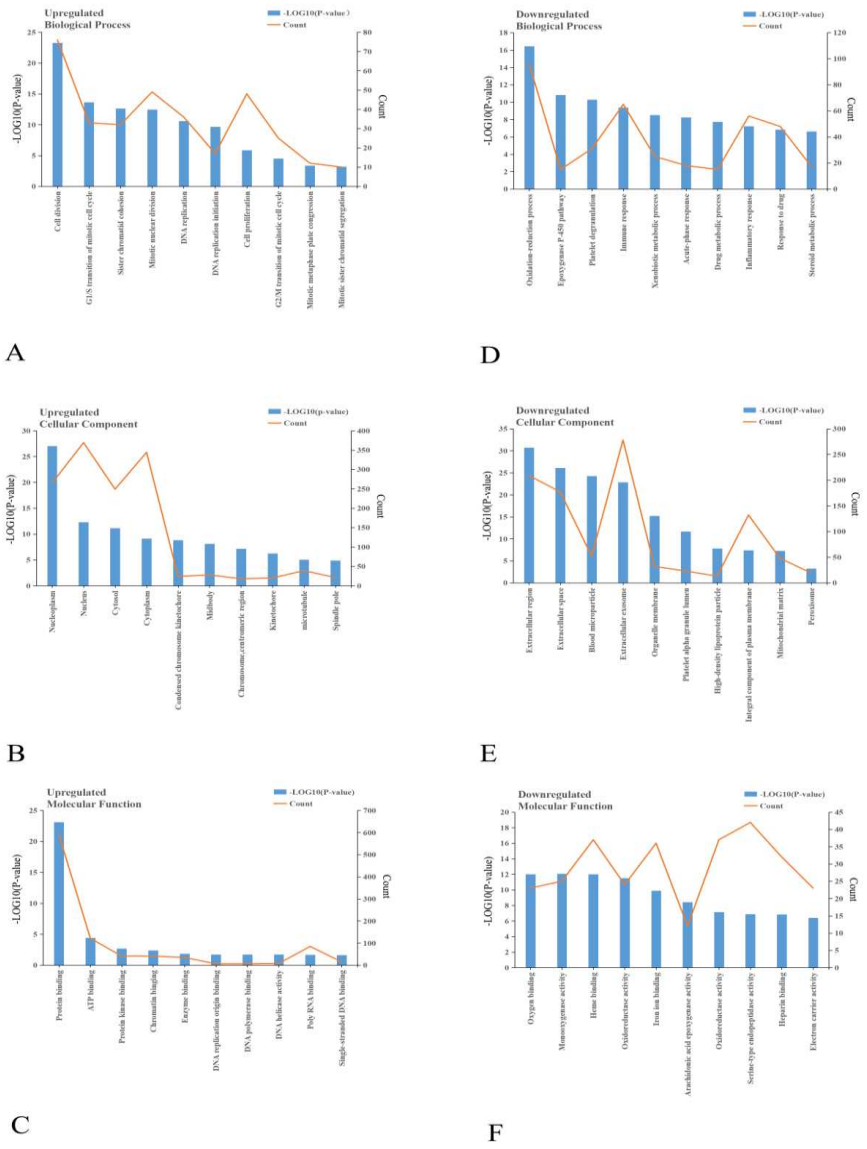


Fig. 3 Results of GO enrichment analysis of differential genes. GO=Gene ontology. The X-axis refers to the function of gene, the left-sided Y-axis represents the p-value (-log10), and the right-sided Y-axis represents the number of genes enriched. Figures A, B, and C represent the GO enrichment results of upregulated genes; Figures D, E, and F represent the GO enrichment results of downregulated genes.

2.4 Analysis of KEGG enrichment results Using KEGG, the pathways of the first ten upregulated and downregulated genes were mapped out (Fig. 4, Tab. S5). Upregulated genes have been shown to affect the cell cycle and DNA replication while downregulated genes affected metabolic pathways and chemical carcinogenesis.

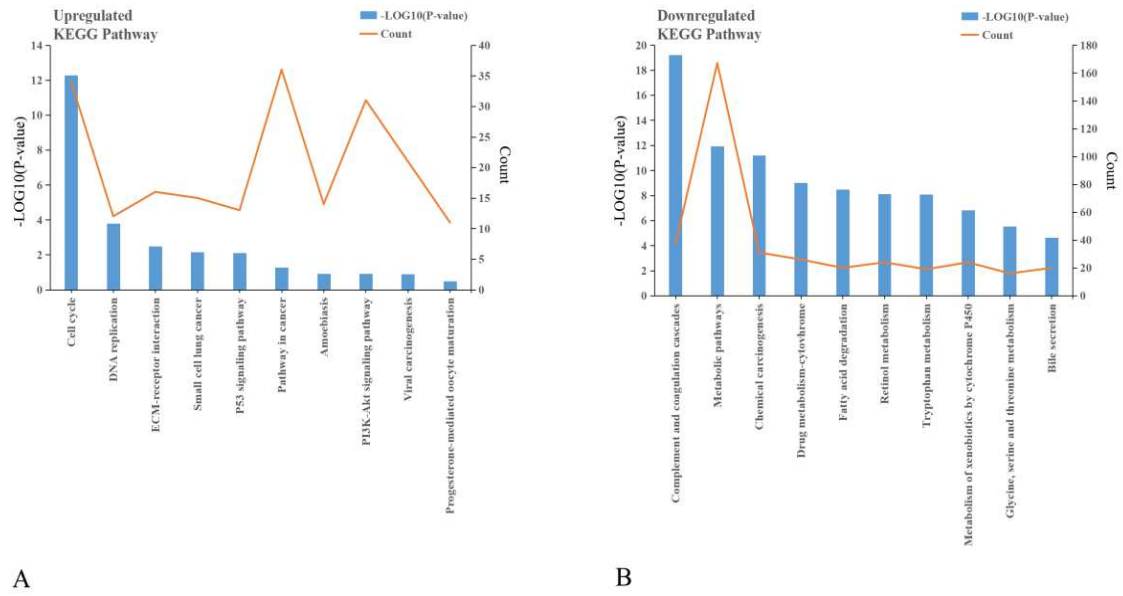


Fig. 4 Results of KEGG enrichment analysis of differential genes. KEGG=Kyoto encyclopedia of genes and genomes. The X-axis refers to the function of the gene, the left-sided Y-axis represents p-value (-log10), and the right-sided Y-axis represents the number of enriched genes. Figure A represents the KEGG enrichment result of upregulated genes; Figure B represents the KEGG enrichment result of downregulated genes.

2.5 Screening results of Hub genes DEGs were imported into the String database. Homo sapiens analysis was selected and the interaction score was set to 0.4 to obtain a protein-protein interaction network (Fig. S1). Using Cytoscape software with MCODE and cytoHubba plugin, the most critical groups in the PPI network were screened out (Fig. 5A), and top 12 key genes were filtered: AURKA> CDKN3> TOP2A> CCNB2> CDC20 > PTTG1> RFC4> UBE2C>RACGAP1> KIF20A> NCAPG> PRC1 (Fig. 5B). According to the MCC algorithm, AURKA and CDKN3 were the two most critical genes. The names, abbreviations and functions of these central genes are shown in Tab S6.

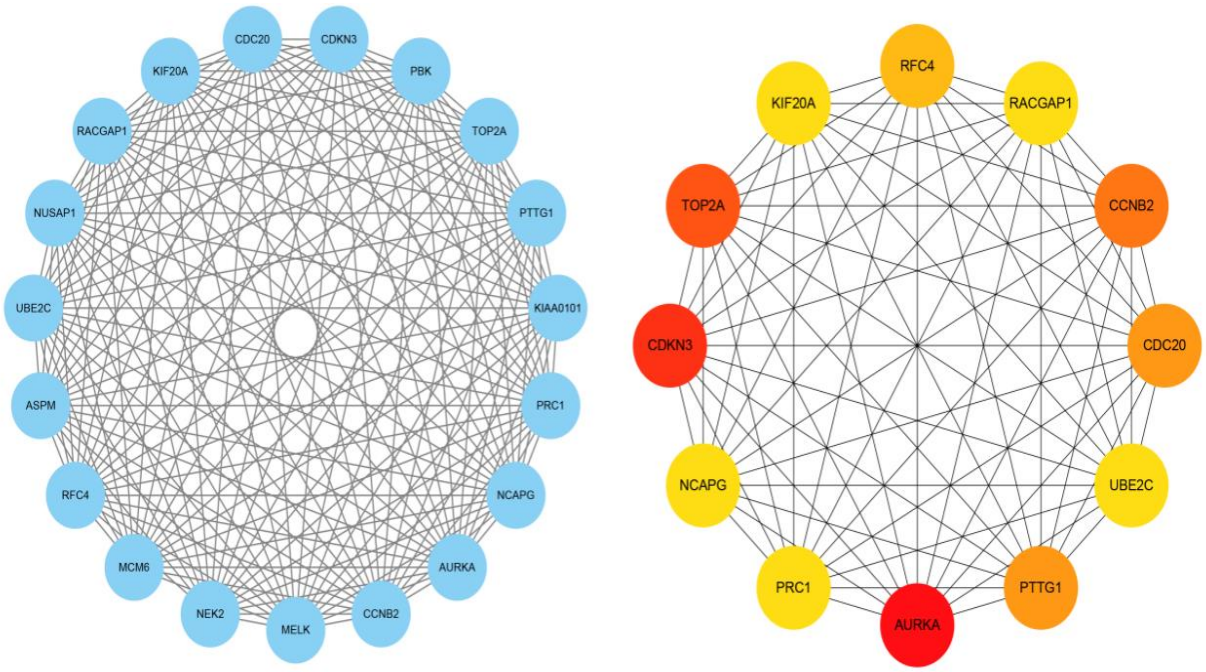
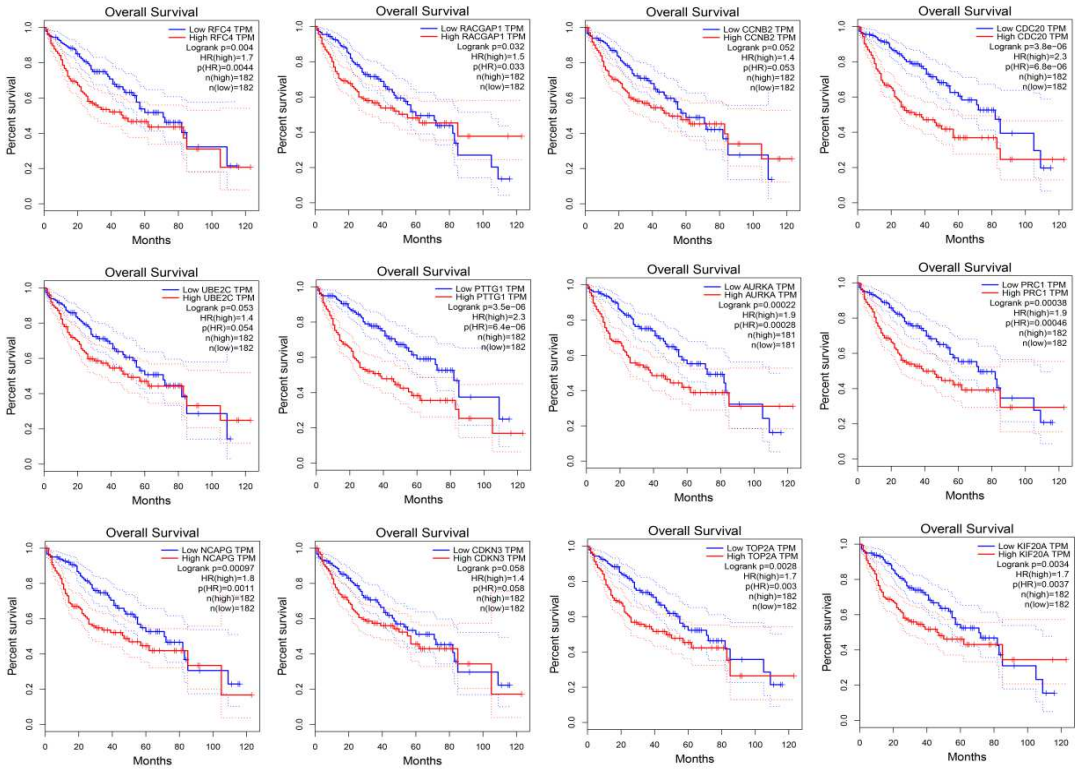
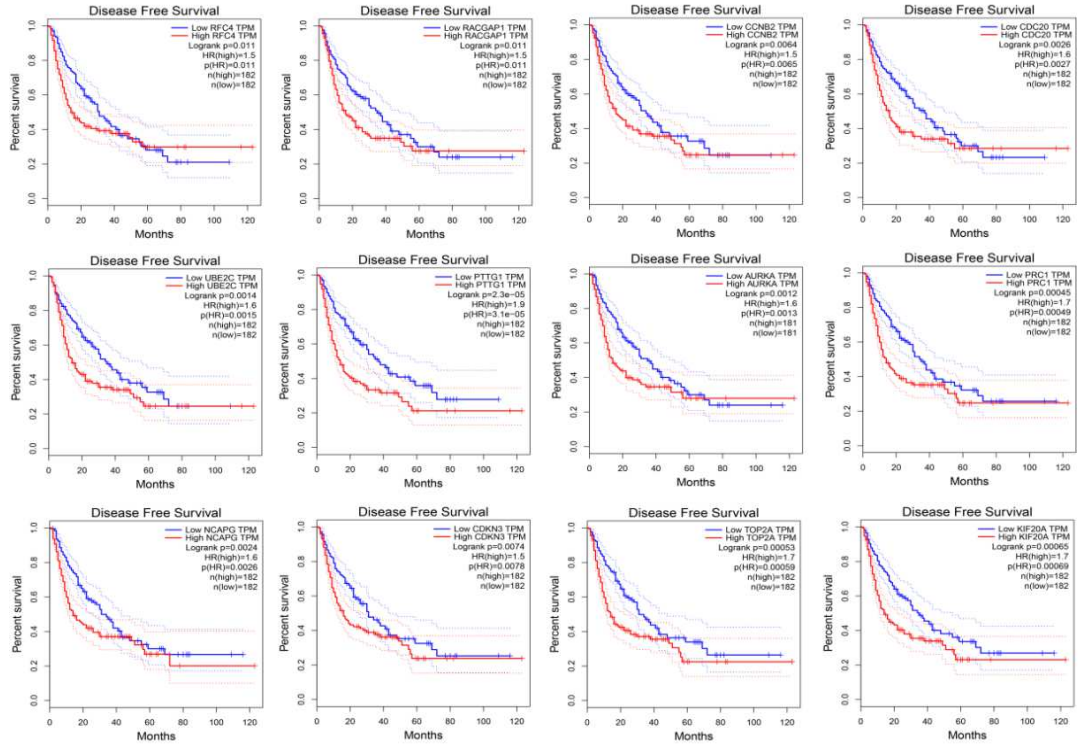


Fig. 5 MCODE and CytoHubba visualization network. **(A)** The first score gene clusters was screened by Cytoscape's MCODE plugins. All the genes were upregulated genes, with a network of 19 nodes and 170 edges. **(B)** Using the MCC algorithm of Cytoscape's CytoHubba plugin, the top 12 Hub genes were selected. All the genes were upregulated genes and the Hub gene network had 12 nodes and 60 edges. The red oval represents a DEG with high PPI score. The darker the red colour, the higher the PPI score. The edge represents protein-protein association. Among them, AURKA and CDKN3 were the two highest-scoring genes.

2.6 Functional analysis of Hub gene Kaplan-Meier curve was used to analyze the overall survival and disease-free survival of the central gene. The expression of AURKA, CDC20, RFC4, KIF20A, RACGAP1, TOP2A, PTTG1, PRC1, NCAPG in HCC patients showed poor overall survival, $p<0.05$, while CCNB2, UBE2C, CDKN3 expression in patients was not significant in overall survival analysis, $p>0.05$. Nonetheless, HCC patients with AURKA, CDC20, RFC4, KIF20A, CDKN3, RACGAP1, TOP2A, CCNB2, UBE2C, PTTG1, PRC1, NCAPG changes showed poorer disease-free survival, $p< 0.05$ (Fig. 6A and B).



A



B

Fig. 6 (A) Using the GEPIA online platform for overall survival analysis, patients' AURKA, CDC20, RFC4, KIF20A, RACGAP1, TOP2A, PTTG1, PRC1 and NCAPG expression were significantly different, while patients' CCNB2, UBE2C and CDKN3 expression were not significant in the overall survival analysis. The difference of $p < 0.05$ was considered statistically significant. (B) Using the GEPIA online platform for Disease Free

Survival analysis, patients' AURKA, CDC20, RFC4, KIF20A, CDKN3, RACGAP1, TOP2A, CCNB2, UBE2C, PTTG1, PRC1, and NCAPG were significantly differently expressed, and the difference of $p<0.05$ was considered statistically significant. significance.

2.7 AURKA and CDKN3 Functional Analysis

Among these genes, AURKA and CDKN3 were the two most critical genes in the PPI network, suggesting that they may play an important role in the development or progression of HCC. Based on data from GEPIA online platform, we noted that HCC patients with genomic alterations in AURKA and CDKN3 showed reduced overall and disease-free survival, as shown in Fig.7and Fig.8, AURKA (total survival P = 0.00022, disease-free survival P = 0.0012), CDKN3 (total survival P = 0.058, disease-free survival P = 0.0074). Oncomine analysis of cancer vs. normal tissue showed that AURKA and CDKN3 were significantly overexpressed in HCC in the different datasets (Fig. 7A and B).

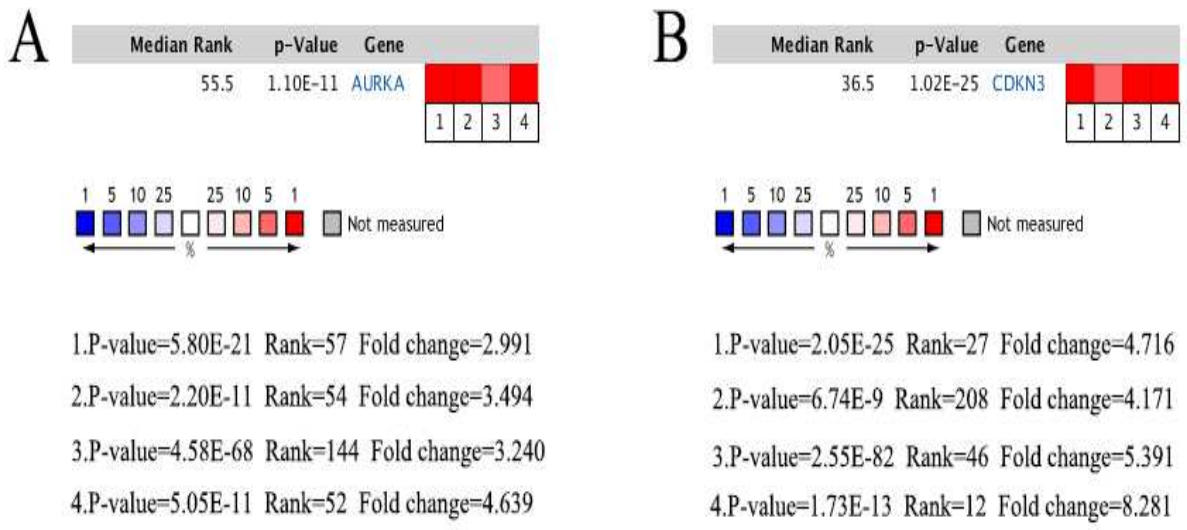


Fig. 7 Oncomine analysis of **(A)** AURKA and **(B)** CDKN3 in hepatic cancer tissues and normal tissues. Heat map of AURKA and CDKN3 gene expression in hepatocellular carcinoma clinical samples and normal tissues.1. Hepatocellular Carcinoma vs. Normal Chen Liver, Mol Biol Cell, 2002; 2. Hepatocellular Carcinoma vs. Normal Roessler Liver, Cancer Res, 2010; 3. Hepatocellular Carcinoma vs. Normal Roessler Liver 2, Cancer Res, 2010;4. Hepatocellular Carcinoma vs. Normal Wurmbach Liver, Hepatology, 2007.

Discussion

In this study, three datasets were selected from the GEO database, which included HCC tissue samples and non-HCC tissue samples. Following the GEO2R analysis of differential genes, 215 DEG were identified, including 34 upregulated genes and 181 downregulated genes. GO enrichment and KEGG functional analysis were performed on these genes. Upregulated genes mainly affected cell division, nucleus, protein binding, and pathway in cancer. Downregulated genes were mainly involved in the oxidation-reduction process, extracellular region, serine-type endopeptidase activity, and metabolic pathways. Besides, a PPI network for the DEG was constructed, and the 12 Hub genes in the differential expression interaction network, including AURKA, CDC20, RFC4, KIF20A, CDKN3, RACGAP1, TOP2A, CCNB2, UBE2C, PTTG1, PRC1 and NCAPG, were screened by the MCODE and MCC algorithm in CytoHubba. Among them, AURKA and CDKN3 were identified as the most important genes.

TOP2A is one of the nuclear matrix components. It has been found that overexpression of TOP2A is closely related to breast cancer, prostate cancer and liver cancer[19, 20]. In normal tissues, PTTG1 is expressed in low amounts or not at all in normal tissue. But in malignant tumors such as adrenocortical carcinoma, prostate cancer and oral squamous cell carcinoma, PTTG1 is abnormally highly expressed and plays a role in tumor cell proliferation[21, 22]. CCNB2 is highly expressed in various malignant tumors such as lung cancer, breast cancer and gastric cancer, and is significantly associated with poor clinical outcomes and prognosis[23, 24]. PRC1 is phosphorylated by CDK1/CyclinB1 protein kinase during cell mitosis and controls the formation of its multimers by regulating its phosphorylation and

dephosphorylation, which in turn affects the formation of the central region of the mitotic spindle[25]. The study found that that PRC1 promotes the early recurrence of hepatocellular carcinoma through the wnt/β-catenin signaling pathway[26]. UBE2C is involved in the regulation of cell mitosis by mediating the ubiquitination of a ubiquitin ligase E3 - APC/C, which plays an important role in regulating cell cycle progression[27]. KIF20A is a member of the kinesin superfamily and is involved in cell mitosis. In recent years, it has been found that KIF20A is abnormally expressed in some tumor cells and tissues, and its abnormal expression is closely related to the formation and development of tumors[28, 29]. RACGAP1 has also been shown to be overexpressed in malignant tumor tissues such as hepatocellular carcinoma, breast cancer and meningioma. Overexpression of RACGAP1 can be used as a marker to evaluate the invasiveness of tumors [30, 31]. NuSAP1 is a microtubule protein that plays an important role in spindle assembly and is an important regulatory molecule in the normal cell cycle. The mRNA and protein levels of NuSAP1 are tightly regulated and exhibit cyclical changes during cell cycle progression[32]. RFC4 is abnormally regulated in various cancers. In a variety of tumors such as colon cancer, breast cancer and liver cancer, silencing RFC4 blocks cells in the S phase, preventing them from subsequent mitosis and proliferation, therefore resulting in a decrease in tumor cell proliferation[33-35]. CDC20 acts as a cell cycle regulator and also plays an important role in human tumors. In numerous tumor types, high expression of CDC20 was observed and this was associated with poor prognosis especially in lung cancer, bladder cancer and liver cancer[36, 37]. NCAPG is a cell cycle-associated gene that affects primary hepatoma cells by influencing mitosis[38].

AURKA is overexpressed in a variety of human cancers, including gastric cancer, breast cancer, bladder cancer, head and neck squamous cell carcinoma and HCC[39-41]. And it plays an important role in the mitotic cell cycle by participating in centrosome replication, isolation and maturation[42-44]. It was demonstrated that the overexpression or inhibition of AURKA significantly opposed or promoted the anticancer effects of cinobufagin in Huh-7 cells, respectively[45]. Genetic variations in the gene encoding AURKA may be a significant predictor of early HCC occurrence and a reliable biomarker for disease progression[46]. Additionally, study results suggested that AURKA contributed to metastasis of irradiated residual HCC through facilitating epithelial-mesenchymal transition (EMT) and cancer stem cell (CSC) properties, suggesting the potential clinical application of AURKA inhibitors in radiotherapy for patients with HCC[47]. The study showed that Aurora-A protein is upregulated in HCC tissues and significantly correlated with recurrence-free and overall survival of patients.

CDKN3 is a cell cycle regulatory protein that interacts directly or indirectly with Cyclin-dependent protein kinase (CDK) and cyclin-dependent kinase inhibitors [48, 49]. CDKN3 exerts different functions in different tumor types. For example, it inhibits cell proliferation in glioblastoma, chronic myeloid leukemia and neuroblastoma[50-52]. It is also a tumor suppressor gene, whereby high expression in liver cancer, lung cancer, ovarian cancer, cervical cancer, kidney cancer, prostate cancer, and breast cancer promotes cell proliferation and facilitates the development of cancer[53-56]. CDKN3 expression may reduce the survival of tumor cells and alter the sensitivity to therapeutic agents via the AKT/P53/P21 signaling pathway. Therefore, CDKN3 may be involved in tumor differentiation and self-renewal[57].

In the PPI network, AURKA and CDKN3 are directly linked to genes such as TOP2A and PTTG1, suggesting its important role in HCC. We evaluated the expression of AURKA and CDKN3 in terms of overall and disease-free survival through GEPIA and Oncomine software, which was reduced when there were genetic alterations in AURKA and CDKN3. Although the overall survival analysis of CDKN3 was not statistically significant in this study, some clinical studies have shown that overexpression of CDKN3 is significantly associated with reduced survival times[58]. It had been reported that the overexpression of CDKN3 also had a significant effect on the overall survival of HCC patients[59]. We speculate that this may be because the survival analysis of the GEPIA tool was based on the relationship between gene mutation and prognosis, but gene overexpression is usually caused by mutation or amplification. Besides, the data included in each database is inconsistent and there will be deviations. Therefore, overexpression of CDKN3 in HCC may come from gene amplification rather than mutation, and further research is needed to confirm our hypothesis.

In summary, this study was designed to identify DEGs that may be involved in the pathogenesis or progression of HCC. A total of 215 DEGs and 12 Hub genes were identified and can be considered diagnostic biomarkers for HCC. However, further research is needed to elucidate the biological functions of these genes in HCC.

Authors' contributions

Shengshan Wang, Yuning Lin, and Wanjun Chen for acquisition of data, analysis and interpretation of data, statistical analysis and drafting of the manuscript; Delun Huang, Yanfei Wei and Qi Zhang for study concept and design, analysis and interpretation of data, obtained

funding. All authors read and approved the final manuscript.

Availability of data and material

The data showed in this article are included in published articles or available from the corresponding author on reasonable request.

Funding

The authors gratefully acknowledge financial support by the National Natural Science Foundation of China (No. 81760757, 81560690), Guangxi Natural Science Foundation Major Project (No. 2017GXNSFAA198152), youth innovation team of Guangxi University of traditional Chinese medicine (2015QT001), Foundation for Guangxi Youth basic ability improving (2020KY07028).

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for Publication

Not applicable.

Competing interests

The authors declare no financial or another competing interest.

Acknowledgements

The authors would like to thank the staff members in physiology department for their work

and support in our scientific work.

References

1. Ferlay, J., et al., Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer, 2015. **136**(5): p. E359-86DOI: 10.1002/ijc.29210.
2. Forner, A., J.M. Llovet, and J. Bruix, Hepatocellular carcinoma. Lancet, 2012. **379**(9822): p. 1245-55DOI: 10.1016/S0140-6736(11)61347-0.
3. Ringelhan, M., J.A. McKeating, and U. Protzer, Viral hepatitis and liver cancer. Philos Trans R Soc Lond B Biol Sci, 2017. **372**(1732)DOI: 10.1098/rstb.2016.0274.
4. Lohitesh, K., R. Chowdhury, and S. Mukherjee, Resistance a major hindrance to chemotherapy in hepatocellular carcinoma: an insight. Cancer Cell Int, 2018. **18**: p. 44DOI: 10.1186/s12935-018-0538-7.
5. Edgar, R., M. Domrachev, and A.E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res, 2002. **30**(1): p. 207-10DOI: 10.1093/nar/30.1.207.
6. Wang, S.M., L.L. Ooi, and K.M. Hui, Identification and validation of a novel gene signature associated with the recurrence of human hepatocellular carcinoma. Clin Cancer Res, 2007. **13**(21): p. 6275-83DOI: 10.1158/1078-0432.CCR-06-2236.
7. Grinchuk, O.V., et al., Tumor-adjacent tissue co-expression profile analysis reveals pro-oncogenic ribosomal gene signature for prognosis of resectable hepatocellular carcinoma. Mol Oncol, 2018. **12**(1): p. 89-113DOI: 10.1002/1878-0261.12153.
8. Wang, Y.H., et al., Plasmalemmal Vesicle Associated Protein (PLVAP) as a therapeutic target for treatment of hepatocellular carcinoma. BMC Cancer, 2014. **14**: p. 815DOI: 10.1186/1471-2407-14-815.
9. Fugelli, P., [General medicine in the 1990's]. Tidsskr Nor Laegeforen, 1990. **110**(17): p. 2268-72.
10. Sanger, D.J. and B. Zivkovic, Discriminative stimulus properties of chlordiazepoxide and zolpidem. Agonist and antagonist effects of CGS 9896 and ZK 91296. Neuropharmacology, 1987. **26**(5): p. 499-505.
11. Kanehisa, M., The KEGG database. Novartis Found Symp, 2002. **247**: p. 91-101; discussion 101-3, 119-28, 244-52.
12. Horne, L., Nutrition and dietetics in the West Indies. Hum Nutr Appl Nutr, 1986. **40**(4): p. 243-52.
13. Shannon, P., et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res, 2003. **13**(11): p. 2498-504DOI: 10.1101/gr.1239303.
14. Bader, G.D. and C.W. Hogue, An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics, 2003. **4**: p. 2.
15. Chin, C.H., et al., cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol, 2014. **8 Suppl 4**: p. S11DOI: 10.1186/1752-0509-8-S4-S11.
16. Sauer, G. and D. Scholz, [On the purification and characterization of 2,3-PGase of red blood cells]. Folia Haematol Int Mag Klin Morphol Blutforsch, 1965. **83**(4): p. 377-82.
17. Seki, M., [Effect of histamine on the kinetics of transfused lymphocytes]. Saishin Igaku, 1971. **26**(8): p. 1495-501.
18. Foote, S.L., et al., Electrophysiological evidence for the involvement of the locus coeruleus in alerting, orienting, and attending. Prog Brain Res, 1991. **88**: p. 521-32.
19. Liu, T., et al., Mutual regulation of MDM4 and TOP2A in cancer cell proliferation. Mol Oncol, 2019.

13(5): p. 1047-1058DOI: 10.1002/1878-0261.12457.
20. Hao, W., L. Zheng, and P. Xin-xing, Curcumol promotes the apoptosis of hepatocellular carcinoma HepG2 cells by regulating PI3K/AKT pathway. Central South Pharmacy, 2019. **17**(01): p. 11-14.
21. Demeure, M.J., et al., PTTG1 overexpression in adrenocortical cancer is associated with poor survival and represents a potential therapeutic target. Surgery, 2013. **154**(6): p. 1405-16; discussion 1416DOI: 10.1016/j.surg.2013.06.058.
22. Cao, X.L., et al., [Downregulation of PTTG1 expression inhibits the proliferation and invasiveness and promotes the apoptosis of human prostate cancer LNCaP-AI cells]. Zhonghua Nan Ke Xue, 2017. **23**(7): p. 589-597.
23. Albulescu, R., Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome. Biomark Med, 2013. **7**(2): p. 203.
24. Shi, Q., et al., ISL1, a novel regulator of CCNB1, CCNB2 and c-MYC genes, promotes gastric cancer cell proliferation and tumor growth. Oncotarget, 2016. **7**(24): p. 36489-36500DOI: 10.18632/oncotarget.9269.
25. Hoffmann, I.M., et al., Acute renal failure in cystic fibrosis: association with inhaled tobramycin therapy. Pediatr Pulmonol, 2002. **34**(5): p. 375-7DOI: 10.1002/ppul.10185.
26. Chen, J., et al., The microtubule-associated protein PRC1 promotes early recurrence of hepatocellular carcinoma in association with the Wnt/beta-catenin signalling pathway. Gut, 2016. **65**(9): p. 1522-34DOI: 10.1136/gutjnl-2015-310625.
27. Han, Q., et al., MicroRNA-196a post-transcriptionally upregulates the UBE2C proto-oncogene and promotes cell proliferation in breast cancer. Oncol Rep, 2015. **34**(2): p. 877-83DOI: 10.3892/or.2015.4049.
28. Stangel, D., et al., Kif20a inhibition reduces migration and invasion of pancreatic cancer cells. J Surg Res, 2015. **197**(1): p. 91-100DOI: 10.1016/j.jss.2015.03.070.
29. Lu, M., et al., Aberrant KIF20A expression might independently predict poor overall survival and recurrence-free survival of hepatocellular carcinoma. IUBMB Life, 2018. **70**(4): p. 328-335DOI: 10.1002/iub.1726.
30. Neubauer, E., et al., Comparative evaluation of three proliferation markers, Ki-67, TOP2A, and RacGAP1, in bronchopulmonary neuroendocrine neoplasms: Issues and prospects. Oncotarget, 2016. **7**(27): p. 41959-41973DOI: 10.18632/oncotarget.9747.
31. Wang, S.M., L.L. Ooi, and K.M. Hui, Upregulation of Rac GTPase-activating protein 1 is significantly associated with the early recurrence of human hepatocellular carcinoma. Clin Cancer Res, 2011. **17**(18): p. 6040-51DOI: 10.1158/1078-0432.CCR-11-0557.
32. Wang, Y., et al., Downregulation of nucleolar and spindle-associated protein 1 expression suppresses liver cancer cell function. Exp Ther Med, 2019. **17**(4): p. 2969-2978DOI: 10.3892/etm.2019.7314.
33. Wang, X.C., et al., Genome-wide RNAi Screening Identifies RFC4 as a Factor That Mediates Radioresistance in Colorectal Cancer by Facilitating Nonhomologous End Joining Repair. Clin Cancer Res, 2019DOI: 10.1158/1078-0432.CCR-18-3735.
34. Chen, Z., et al., Identification of Potential Key Genes for Hepatitis B Virus-Associated Hepatocellular Carcinoma by Bioinformatics Analysis. J Comput Biol, 2019. **26**(5): p. 485-494DOI: 10.1089/cmb.2018.0244.
35. Koleck, T.A., et al., An exploratory study of host polymorphisms in genes that clinically characterize breast cancer tumors and pretreatment cognitive performance in breast cancer survivors. Breast Cancer (Dove Med Press), 2017. **9**: p. 95-110DOI: 10.2147/BCTT.S123785.

36. Kim, Y., et al., MAD2 and CDC20 are upregulated in high-grade squamous intraepithelial lesions and squamous cell carcinomas of the uterine cervix. Int J Gynecol Pathol, 2014. **33**(5): p. 517-23DOI: 10.1097/PGP.0000000000000082.
37. Zhuang, L., Z. Yang, and Z. Meng, Upregulation of BUB1B, CCNB1, CDC7, CDC20, and MCM3 in Tumor Tissues Predicted Worse Overall Survival and Disease-Free Survival in Hepatocellular Carcinoma Patients. Biomed Res Int, 2018. **2018**: p. 7897346DOI: 10.1155/2018/7897346.
38. Arai, T., et al., Regulation of NCAPG by miR-99a-3p (passenger strand) inhibits cancer cell aggressiveness and is involved in CRPC. Cancer Med, 2018. **7**(5): p. 1988-2002DOI: 10.1002/cam4.1455.
39. Lee, J.W., et al., Combined Aurora Kinase A (AURKA) and WEE1 Inhibition Demonstrates Synergistic Antitumor Effect in Squamous Cell Carcinoma of the Head and Neck. Clin Cancer Res, 2019. **25**(11): p. 3430-3442DOI: 10.1158/1078-0432.CCR-18-0440.
40. Guo, M., et al., Increased AURKA promotes cell proliferation and predicts poor prognosis in bladder cancer. BMC Syst Biol, 2018. **12**(Suppl 7): p. 118DOI: 10.1186/s12918-018-0634-2.
41. Wang, L., et al., Gossypin inhibits gastric cancer growth by direct targeting of AURKA and RSK2. Phytother Res, 2019. **33**(3): p. 640-650DOI: 10.1002/ptr.6253.
42. Treekitkarnmongkol, W., et al., Aurora kinase-A overexpression in mouse mammary epithelium induces mammary adenocarcinomas harboring genetic alterations shared with human breast cancer. Carcinogenesis, 2016. **37**(12): p. 1180-1189DOI: 10.1093/carcin/bgw097.
43. Woo, J.K., et al., Daurinol Enhances the Efficacy of Radiotherapy in Lung Cancer via Suppression of Aurora Kinase A/B Expression. Mol Cancer Ther, 2015. **14**(7): p. 1693-704DOI: 10.1158/1535-7163.MCT-14-0960.
44. Goos, J.A., et al., Aurora kinase A (AURKA) expression in colorectal cancer liver metastasis is associated with poor prognosis. Br J Cancer, 2013. **109**(9): p. 2445-52DOI: 10.1038/bjc.2013.608.
45. Zhao, L., et al., The anticancer effects of cinobufagin on hepatocellular carcinoma Huh7 cells are associated with activation of the p73 signaling pathway. Mol Med Rep, 2019. **19**(5): p. 4119-4128DOI: 10.3892/mmr.2019.10108.
46. Wang, B., et al., Variations in the AURKA Gene: Biomarkers for the Development and Progression of Hepatocellular Carcinoma. Int J Med Sci, 2018. **15**(2): p. 170-175DOI: 10.7150/ijms.22513.
47. Chen, C., et al., AURKA promotes cancer metastasis by regulating epithelial-mesenchymal transition and cancer stem cell properties in hepatocellular carcinoma. Biochem Biophys Res Commun, 2017. **486**(2): p. 514-520DOI: 10.1016/j.bbrc.2017.03.075.
48. Mangini, N.S., et al., Palbociclib: A Novel Cyclin-Dependent Kinase Inhibitor for Hormone Receptor-Positive Advanced Breast Cancer. Ann Pharmacother, 2015. **49**(11): p. 1252-60DOI: 10.1177/1060028015602273.
49. Diallo, A. and C. Prigent, [The serine/threonine kinases that control cell cycle progression as therapeutic targets]. Bull Cancer, 2011. **98**(11): p. 1335-45DOI: 10.1684/bdc.2011.1467.
50. Li, H., et al., KAP regulates ROCK2 and Cdk2 in an RNA-activated glioblastoma invasion pathway. Oncogene, 2015. **34**(11): p. 1432-41DOI: 10.1038/onc.2014.49.
51. Poon, R.Y. and T. Hunter, Dephosphorylation of Cdk2 Thr160 by the cyclin-dependent kinase-interacting phosphatase KAP in the absence of cyclin. Science, 1995. **270**(5233): p. 90-3DOI: 10.1126/science.270.5233.90.
52. Cress, W.D., P. Yu, and J. Wu, Expression and alternative splicing of the cyclin-dependent kinase inhibitor-3 gene in human cancer. Int J Biochem Cell Biol, 2017. **91**(Pt B): p. 98-101DOI:

10.1016/j.biocel.2017.05.013.
53. Wang, L., et al., Cyclin-dependent kinase inhibitor 3 (CDKN3) novel cell cycle computational network between human non-malignancy associated hepatitis/cirrhosis and hepatocellular carcinoma (HCC) transformation. Cell Prolif, 2011. **44**(3): p. 291-9DOI: 10.1111/j.1365-2184.2011.00752.x.
54. Fan, C., et al., Overexpression of major CDKN3 transcripts is associated with poor survival in lung adenocarcinoma. Br J Cancer, 2015. **113**(12): p. 1735-43DOI: 10.1038/bjc.2015.378.
55. Li, T., et al., CDKN3 is an independent prognostic factor and promotes ovarian carcinoma cell proliferation in ovarian cancer. Oncol Rep, 2014. **31**(4): p. 1825-31DOI: 10.3892/or.2014.3045.
56. Berumen, J., A.M. Espinosa, and I. Medina, Targeting CDKN3 in cervical cancer. Expert Opin Ther Targets, 2014. **18**(10): p. 1149-62DOI: 10.1517/14728222.2014.941808.
57. Dai, W., et al., CDKN3 expression is negatively associated with pathological tumor stage and CDKN3 inhibition promotes cell survival in hepatocellular carcinoma. Mol Med Rep, 2016. **14**(2): p. 1509-14DOI: 10.3892/mmr.2016.5410.
58. Xing, C., et al., Cyclin-dependent kinase inhibitor 3 is overexpressed in hepatocellular carcinoma and promotes tumor cell proliferation. Biochem Biophys Res Commun, 2012. **420**(1): p. 29-35DOI: 10.1016/j.bbrc.2012.02.107.
59. Wu, M., et al., Analysis of potential key genes in very early hepatocellular carcinoma. World J Surg Oncol, 2019. **17**(1): p. 77DOI: 10.1186/s12957-019-1616-6.

Figures

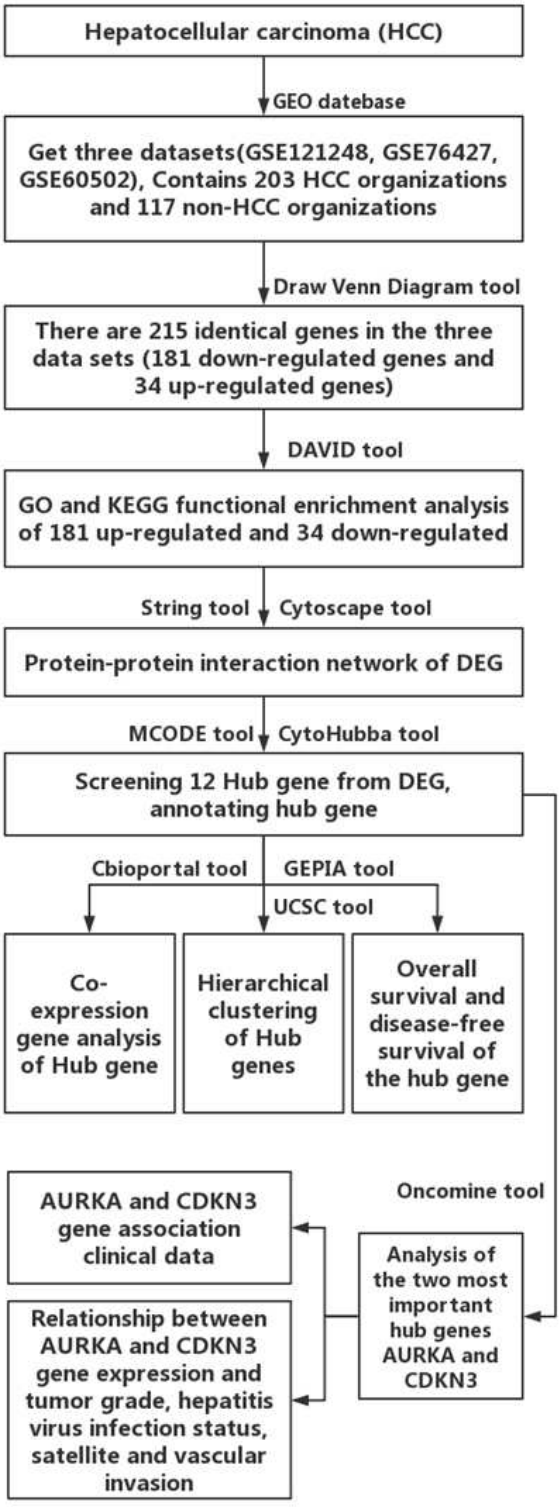


Figure 1

The schematic of this study is intended to describe bioinformatics screening and analysis of key genes for HCC.

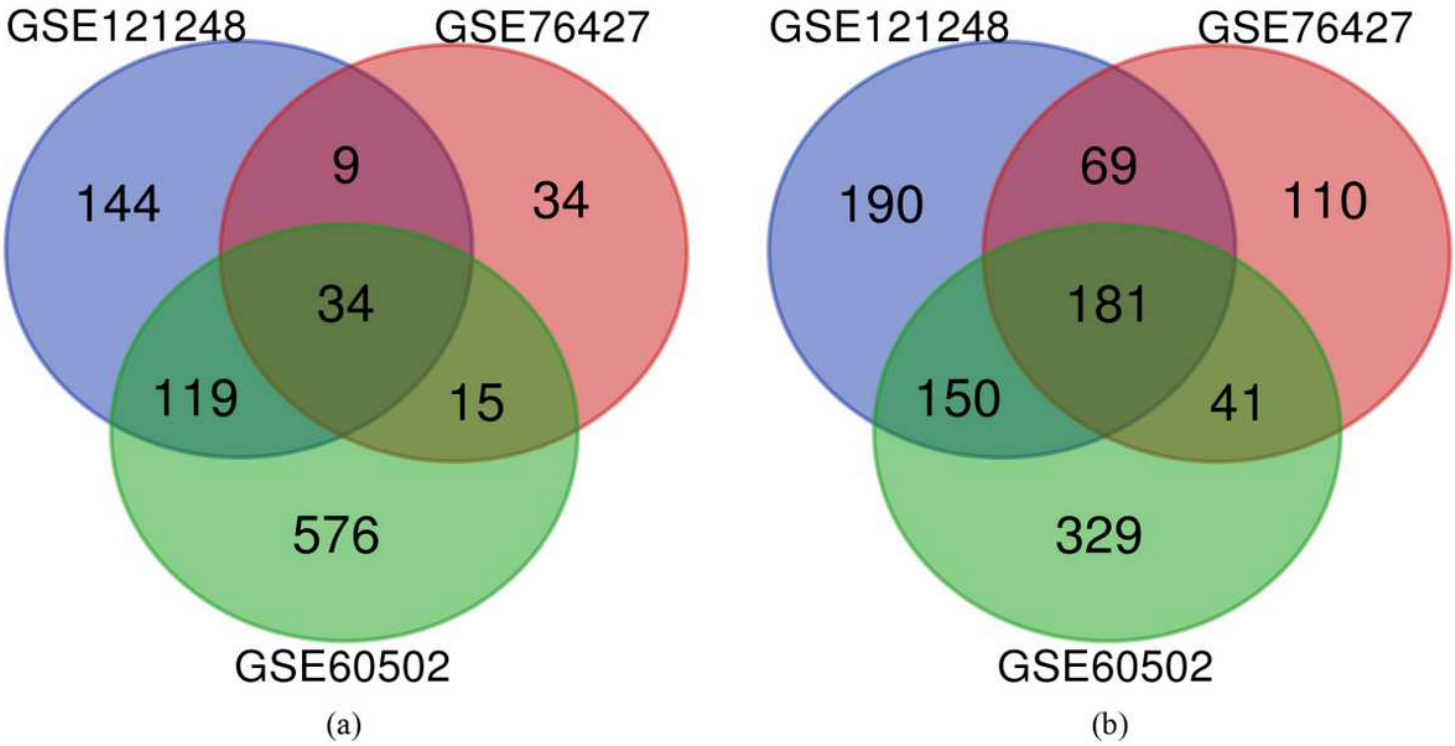


Figure 2

215 DEG obtained from the three datasets. (a) 34 common upregulated genes. (b) 181 common downregulated genes. The different colors represent different datasets, and each intersecting area represents a gene shared between datasets. DEGs with |logFC|(fold change)>1 and adj.p-value<0.05 were selected.

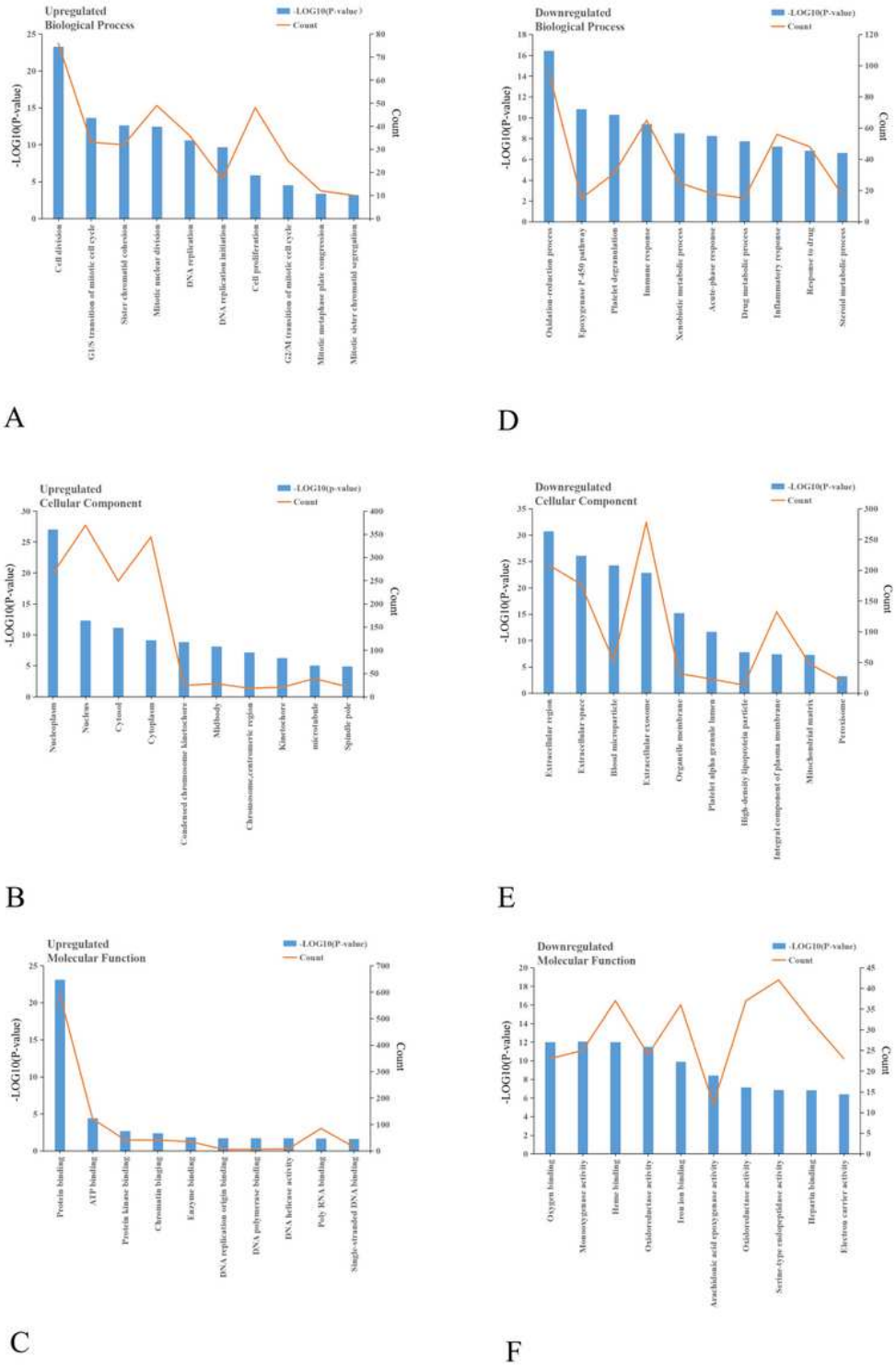


Figure 3

Results of GO enrichment analysis of differential genes. GO=Gene ontology. The X-axis refers to the function of gene, the left-sided Y-axis represents the p-value (-log10), and the right-sided Y-axis represents the number of genes enriched. Figures A, B, and C represent the GO enrichment results of upregulated genes; Figures D, E, and F represent the GO enrichment results of downregulated genes.

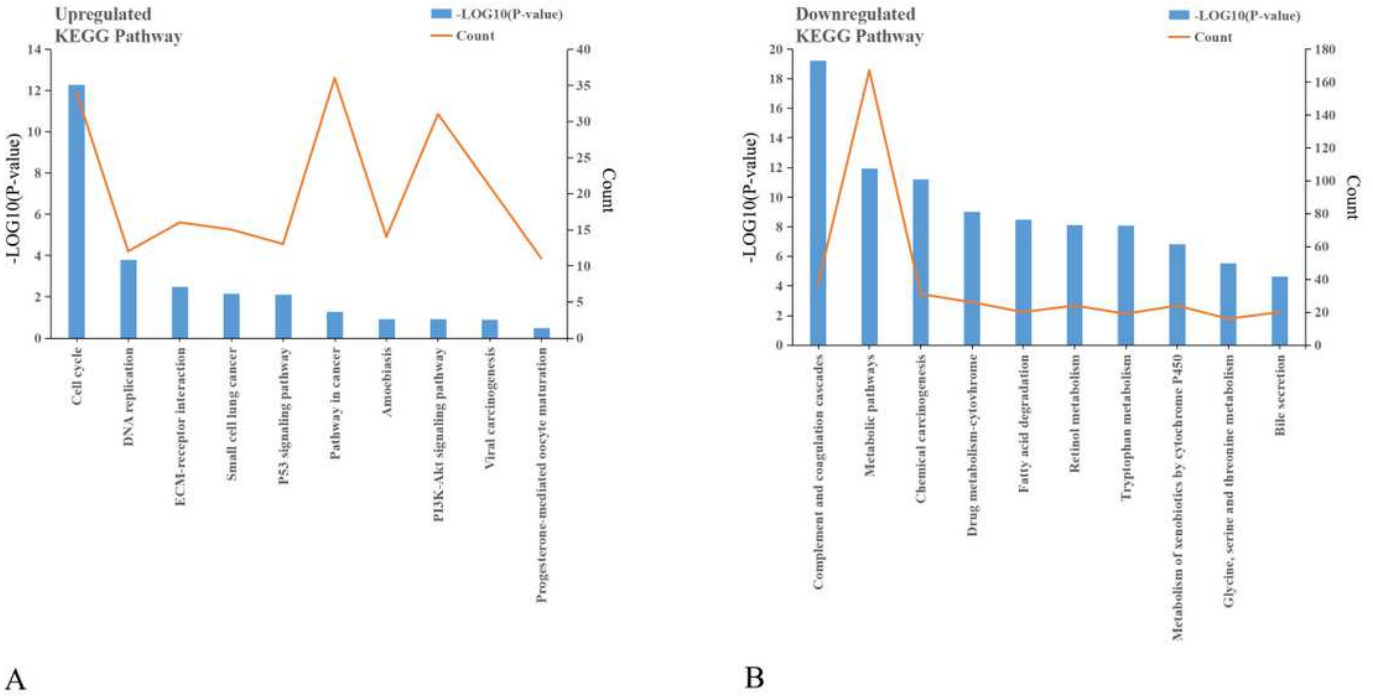


Figure 4

Results of KEGG enrichment analysis of differential genes. KEGG=Kyoto encyclopedia of genes and genomes. The X-axis refers to the function of the gene, the left-sided Y-axis represents p-value (-log10), and the right-sided Y-axis represents the number of enriched genes. Figure A represents the KEGG enrichment result of upregulated genes; Figure B represents the KEGG enrichment result of downregulated genes.

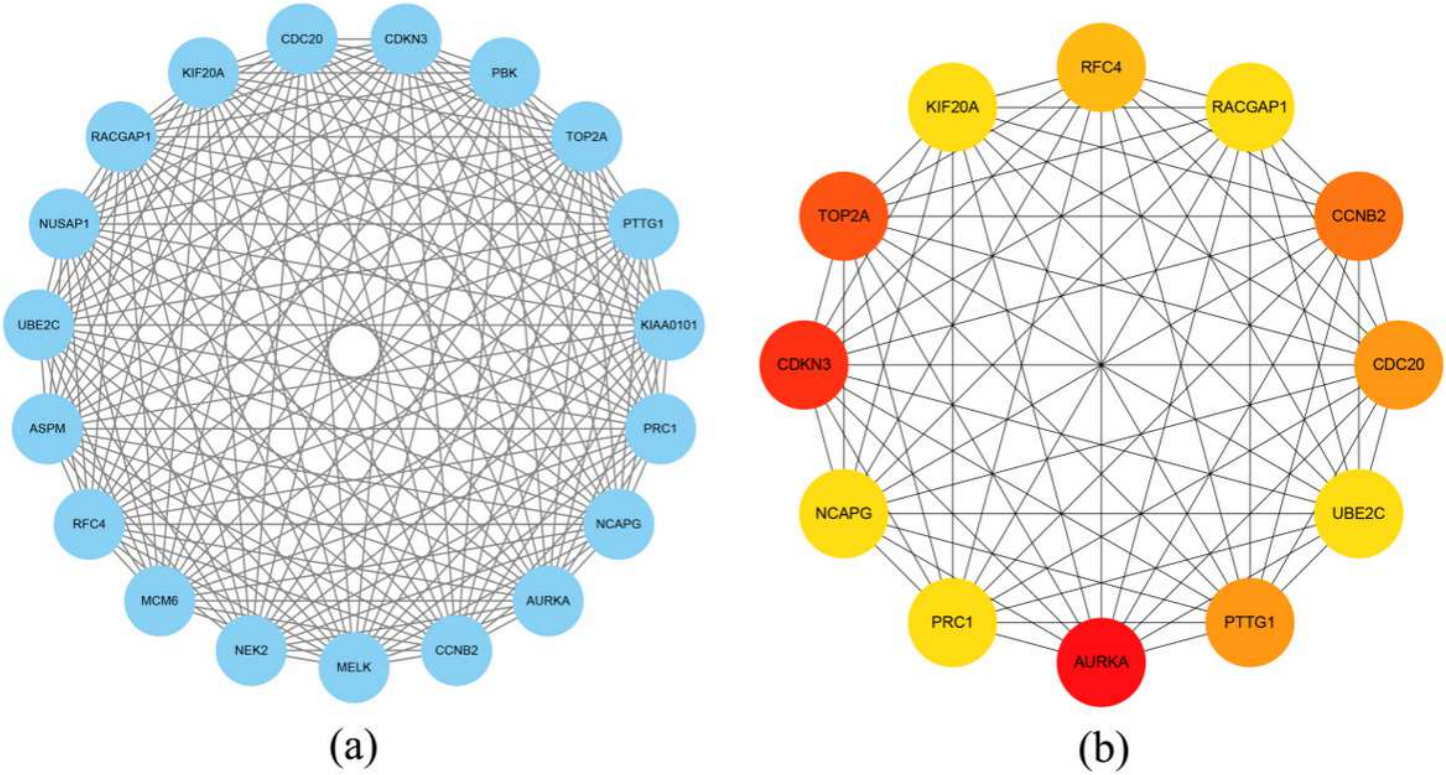


Figure 5

MCODE and CytoHubba visualization network. (A) The first score gene clusters was screened by Cytoscape's MCODE plugins. All the genes were upregulated genes, with a network of 19 nodes and 170 edges. (B) Using the MCC algorithm of Cytoscape's CytoHubba plugin, the top 12 Hub genes were selected. All the genes were upregulated genes and the Hub gene network had 12 nodes and 60 edges. The red oval represents a DEG with high PPI score. The darker the red colour, the higher the PPI score. The edge represents protein-protein association. Among them, AURKA and CDKN3 were the two highest-scoring genes.

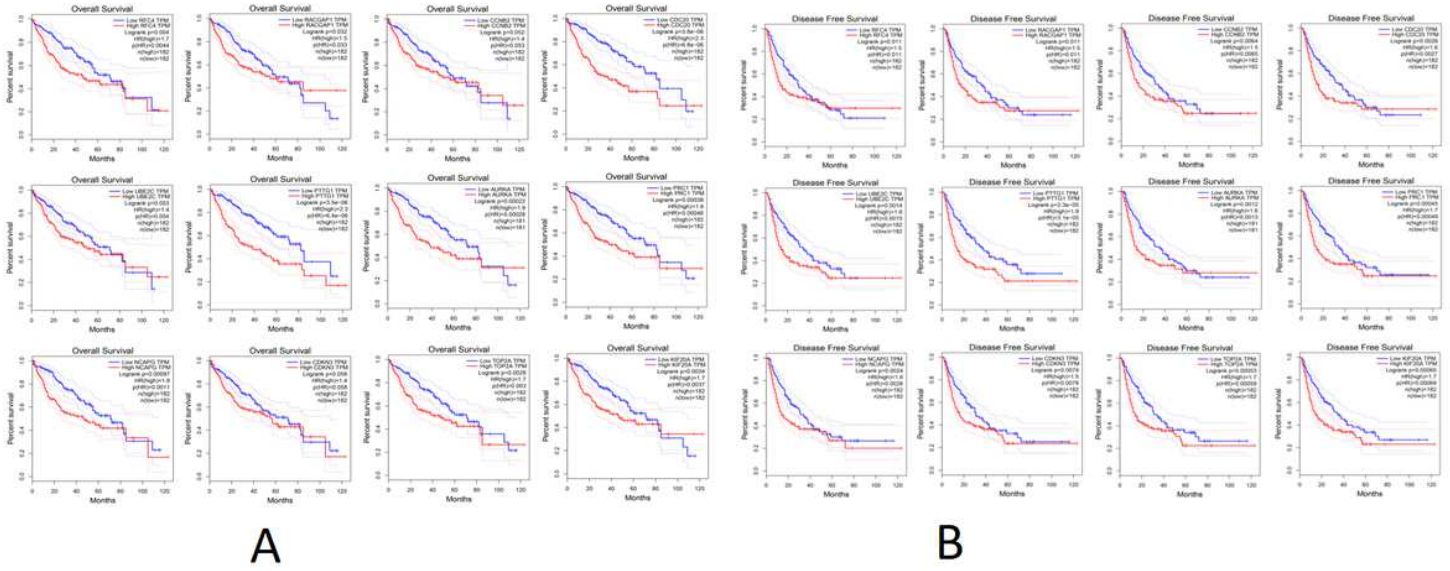


Figure 6

(A) Using the GEPIA online platform for overall survival analysis, patients' AURKA, CDC20, RFC4, KIF20A, RACGAP1, TOP2A, PTTG1, PRC1 and NCAPG expression were significantly different, while patients' CCNB2, UBE2C and CDKN3 expression were not significant in the overall survival analysis. The difference of $p < 0.05$ was considered statistically significant. (B) Using the GEPIA online platform for Disease Free Survival analysis, patients' AURKA, CDC20, RFC4, KIF20A, CDKN3, RACGAP1, TOP2A, CCNB2, UBE2C, PTTG1, PRC1, and NCAPG were significantly differently expressed, and the difference of $p<0.05$ was considered statistically significant. significance.

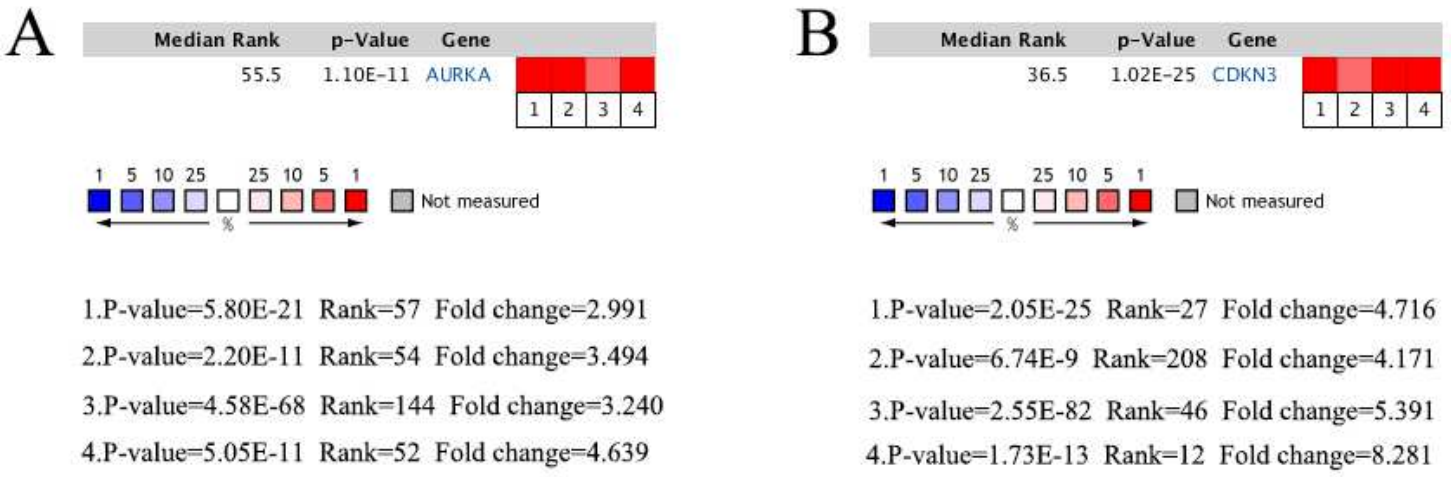


Figure 7

Oncomine analysis of (A) AURKA and (B) CDKN3 in hepatic cancer tissues and normal tissues. Heat map of AURKA and CDKN3 gene expression in hepatocellular carcinoma clinical samples and normal tissues.1. Hepatocellular Carcinoma vs. Normal Chen Liver, Mol Biol Cell, 2002; 2. Hepatocellular Carcinoma vs. Normal Roessler Liver, Cancer Res, 2010; 3. Hepatocellular Carcinoma vs. Normal Roessler Liver 2, Cancer Res, 2010;4. Hepatocellular Carcinoma vs. Normal Wurmbach Liver, Hepatology, 2007.

Supplementary Files

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

- Supplementary20200721.docx