Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identification and Validation of Real Hub Genes in Hepatocellular Carcinoma Based on Weighted Gene Co-expression Network Analysis

Zhigang Zhao (dorisjiang667@163.com)
Wuhan Fourth Hospital
Yu Qiao
Jianghan University
Xin Wang
Wuhan Fourth Hospital
Jun Hu
Wuhan Fourth Hospital
Yurong Mao
Jianghan University
Fahu Yuan
Jianghan University

Research Article

Keywords: bioinformatics analysis, Gene Expression Omnibus (GEO), hepatocellular carcinoma, the Cancer Genome Atlas (TCGA), the Human Protein Atlas database (HPA), Weighted Gene Co-expression Network Analysis (WGCNA)

Posted Date: March 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1458636/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Background: Hepatocellular Carcinoma (HCC) is one of the most common liver malignancies in the world. With highly invasive biological characteristics and lack of obvious clinical manifestations, hepatocellular Carcinoma usually has a poor prognosis and ranks fourth in cancer mortality. The aetiology and exact molecular mechanism of primary hepatocellular carcinoma are still unclear. This work aims to help identify biomarkers of early HCC diagnosis or prognosis.

Results: In this study, Expression data and clinical information of HTSEQ-Counts were downloaded from The Cancer Genome Atlas (TCGA) database, and Gene Expression map GSE121248 was downloaded from Gene Expression Omnibus (GEO). By differentially expressed genes (DEGs) and Weighted Gene co-expression Network Analysis（WGCNA）searched for modules in the two databases that had the same effect on the biological characteristics of HCC, and extracted the module genes with the highest positive correlation with HCC from two databases, and finally obtained overlapping genes. Then, we performed functional enrichment analysis on the overlapping genes to understand their potential biological functions. The top ten hub genes were screened according to MCC through the String database and Cytoscape software. By survival analysis, high expression of CDK1, CCNA2, CDC20, KIF11, DLGAP5, KIF20A, ASPM, CEP55, TPX2 was associated with poorer overall survival (OS) of HCC patients. The DFS curve was plotted using the online website GEPIA2. Finally, based on the enrichment of these genes in the KEGG pathway, real hub genes were screened out, which were CDK1, CCNA2 and CDC20 respectively.

Conclusions: High expression of these three genes was negatively correlated with survival time in HCC, and the expression of CDK1, CCNA2 and CDC20 were significantly higher in tumour tissues of HCC patients than in normal liver tissues as verified again by the HPA database. All in all, this provides a new feasible target for early and accurate diagnosis of HCC, clinical diagnosis, treatment and prognosis.

Introduction

Hepatocellular carcinoma (HCC) is currently the fifth most prevalent and fourth most deadly disease in the world [1]. It increases year by year, and the incidence of males is much higher than that of females, and the incidence of males and females is 2.8:1 [2]. Globally, East Asia and Africa have the highest incidence of HCC, but the incidence and mortality of HCC are also increasing in Europe and the United States. Hepatitis B virus (HBV) infection accounts for 60% in Asia, and cirrhosis is the strongest risk factor [3]. According to the SEER database of The National Cancer Institute, the average five-year survival rate of HCC patients in the United States is 19.6%, but for advanced and metastatic diseases, the survival rate may be as low as 2.5% [1]. Currently, there are surgical treatments and non-surgical treatments for HCC. Among them, surgical treatment includes partial liver resection and liver transplantation, which are suitable for early patients. Non-surgical treatment includes tumour radiotherapy, chemotherapy, biotherapy, immunotherapy, etc., both surgical and non-surgical treatment has a significant effect on early patients, but the treatment effect on advanced patients is not good. Therefore, how to improve the early diagnosis rate of HCC patients is the focus of our attention. The current outlook for HCC has

mentioned that screening with novel serum biomarkers may help in early patient diagnosis and more effective selection of appropriate treatments for the disease, thus effectively improving patient survival cycles, survival rates, etc [3].

In recent years, with the rapid development of high-throughput measurement technologies, bioinformatics analysis of expression profiles has been widely used to find potential targets for cancer and biomarkers for patient prognostic analysis [4]. Differential gene expression analysis between different clinical phenotypes is a common method to search for potential biomarkers of diseases. Weighted gene co-expression network analysis allows the construction of co-expressed gene modules by finding coordinately expressed gene modules and analysing the relationship between different modules and specific features [5].

In our study, overlapping genes were obtained by using a combination of differential gene analysis and WGCNA on HCC samples from the TCGA and GEO databases. The real hub genes were obtained by further analysis of biological functions and clinical characteristics of overlapping genes. The final results are helpful for us to further understand the mechanism of occurrence and development of HCC and provide new molecular targets for early detection and treatment of HCC. The flow chart of this research is shown in **Fig. 1**.

Results

Identification of Differentially Expressed Genes

Data information of HCC patients from different platforms was analyzed using the Limma software package and set to $|logFC| \geq 1$ and adjusted p-value < 0.05. A total of, 2703 DEGs were screened from HCC and normal control samples from THE TCGA-LIHC database, among which 1023 genes were up-regulated and 1680 genes were down-regulated in tumour tissues (**Figure 2A**). A total of 580 DEGs were screened from the GSE121248 dataset, of which 167 were up-regulated genes and 413 were down-regulated genes (**Figure 2B**). In addition, heat maps of DEGs of two data sets were constructed using the R language, as shown in **Figure 3A, B.**

Weighted Gene Co-Expression Network Construction

The data sets of TCGA-LIHC and GSE121248 were used to construct the sample clustering tree of HCC and normal samples through WGCNA data packets, and the soft power threshold β was selected as 3 and 10, respectively. To ensure that the fitting index (R2) and average connectivity of the scale-free topological model reaches a stable state, as shown in **Fig. 4A and Fig. 5A**. Then, gene modules were detected based on the TOM matrix, and genes with similar expression patterns were placed into different modules through average linkage clustering. Hierarchical clustering analysis is shown in **Figs. 4B and 5B**. Finally,

after assigning colours to each module, 9 modules in TCGA-LIHC (Fig. 4C) and 8 modules in GSE121248 (Fig. 5C) were identified respectively.

Identification of DEGs Overlapping Genes in TCGA-LIHC and GSE121248 Dataset

Subsequently, hierarchical cluster analysis was further evaluated to construct the correlation heat maps between different modules and clinical features. The results showed that the yellow color in TCGA-LIHC and the turquoise color in GSE121248 had the highest correlation with tumor tissues (Fig. 6A, **B**) (yellow module: R = 0.47, p = 5e-25; Turquoise module: r = 0.72, P = 3e-18). The yellow module and the turquoise module selected above were subjected to the next step of the analysis, and the scatter plots of individual genes within the modules were obtained by calculating the gene significance (GS) and module membership (MM) analysis within the modules (Fig. 7A, **B**). Finally, the "VennDiagram" software package was used to intersection the DEGs of TCGA-LIHC and GSE121248 with the selected yellow and turquoise module genes to obtain the Venn diagram. The final 2,703 DEGs from TCGA and 580 DEGs from GEO, 73 co-expressed genes were extracted from the yellow module of TCGA-LIHC and the turquoise module of GSE121248, respectively, with 1235 co-expressed genes and 734 co-expressed genes (Fig. 8).

GO and KEGG Functional Enrichment Analysis of the 73 Overlapping Genes

The "clusterProfiler" R language package was used to enrich the functions of 73 overlapping genes, to map their GO and KEGG pathway bubbles diagram (**Figure 9A, B**), to predict the potential biological functions and related pathways of these 73 genes. GO enrichment analysis showed that these genes were significantly enriched in the nuclear division, organelle fission, spindle, chromosomal region, microtubule-binding and tubulin binding ($p<0.05$). In the KEGG enrichment pathway, these genes are significantly enriched in the cell cycle, DNA replication, Cellular senescence, p53 signalling pathway, Viral carcinogenesis and mismatch repair, nucleotide excision repair ($P<0.05$).

Identification of Hub Genes from Overlapping Genes

To further select hub genes in the overlapping genes obtained by WGCNA analysis, 73 genes were imported into the STRING 11.5 database in this study to construct a PPI network (**Figure 10A**). The results were then imported into the visualization software Cytoscape (**Figure 10B**), and the MCC algorithm of CytoHubba plug-in was used to screen out the top 10 nodes (**Figure 10C**). These 10 genes are hub genes, namely CDK1, TOP2A, DLGAP5, CDC20, ASPM, CEP55, TPX2, KIF11, KIF20A and CCNA2.

Correlation Analysis of Hub Genes expression and survival

To verify the authenticity of the 10 screened hub genes, we standardized the target genes based on TCGA-LIHC data set through the TPM method, and finally drew the expression point map of Hub genes in normal tissues and liver cancer tissues (**Figure 11**). Overall Survival (OS) of these 10 genes was analyzed

and Kaplan-Meier curves were drawn by "Survival" and "SurvMiner" R language package (**Figure 12**). Disease-free survival (DFS) curves of 10 genes were plotted using the online website GEPIA2 (**Figure 13**), and their Log-rank P values were obtained.

Validation of The Real Hub Genes Based on HPA

By analyzing the difference in expression of Hub genes in normal and hepatocellular carcinoma tissues, combined with the OS curve and DFS curve of Hub genes. We found that the expression levels of CDK1, DLGAP5, CDC20, ASPM, CEP55, TPX2, KIF11, KIF20A and CCNA2 were significantly correlated with OS and DFS in HCC patients ($P < 0.05$). Combined with the distribution of these 9 genes in KEGG enrichment pathways (**Table 1**), it can be concluded that the pathways enriched by CDK1, CDC20 and CCNA2 account for the highest proportion and are the real hub genes we need to find. It is mainly involved in the cell cycle, DNA replication, Oocyte meiosis and other related pathways. The immunohistochemical images of real hub genes were obtained from the HPA database (Fig. 14), and the expression of real hub genes in different tissue samples could be directly reflected by immunohistochemistry.

Discussion

In the current clinical diagnosis and treatment, the continuous increase of serum alpha-fetoprotein (AFP) is often used as an important indicator of liver cancer, but the increase of AFP is also common in some other liver diseases, such as cirrhosis [6], indicating that the specificity and sensitivity of AFP for the diagnosis of liver cancer is not enough. Therefore, biomarkers with higher specificity and sensitivity are needed to diagnose early HCC and effectively judge the future survival of HCC patients. At present, many new and more effective biomarkers are being discovered and applied in clinics. For example, silencing DUSP21 significantly inhibited cell proliferation, colony formation, and tumorigenicity of HCC cells in vivo. DUSP21 plays a significant role in maintaining hepatocellular carcinoma cell proliferation, therefore it could be a potential therapeutic target for liver cancer [7]. High expression of cholesterol esterase (SOAT1) was associated with the worst prognostic risk. By inhibiting the cholesterol esterification enzyme SOAT1, cholesterol levels on the cytoplasmic membrane can be reduced, effectively inhibiting tumour cell proliferation and migration [8]. Therefore, in our study, DEGs and WGCNA were used for the biological analysis of HCC in TCGA and GEO databases. As a result, 73 genes were obtained; PPI and Cytoscape were used for further analysis of Overlapping genes to obtain 10 Hub genes. Namely, CDK1, TOP2A, DLGAP5, CDC20, ASPM, CEP55, TPX2, KIF11, KIF20A, CCNA2. Combining the enrichment of OS, DFS and KEGG pathways of 10 genes, ultimately we found that high expression of CDK1, CDC20 and CCNA2 was significantly associated with OS in HCC patients.

Cyclin Dependent Kinase 1 (CDK1) is the protein-encoding gene for a protein that is a member of the Serine/threonine-protein Kinase family. The protein is a catalytic subunit of a highly conserved protein kinase complex called the M-phase promoting factor (MPF), which is required for the G1/S and G2/M phase transitions of the eukaryotic cell cycle. It was found that miR-582-5p was down-regulated in hepatocellular carcinoma tissues and up-regulation of miR-582-5p inhibited cell proliferation and blocked

the cell cycle in G0/G1 phase. miR-582-5p directly inhibited the expression of CDK1 and AKT3, and indirectly inhibited the expression of cyclinD1, thereby regulating HCC [9]. Experiments have also confirmed that dihydroartemisinin (DHA) can reduce the expression level of CDK1, thereby inhibiting the proliferation of liver cancer cells [10]. At present, CDK1 is highly expressed in various tumour tissues to varying degrees, and it has been proved that CDK1 can play an important role in lung cancer [11], breast cancer [12], pancreatic cancer [13], colorectal cancer [14] and so on by regulating the cell cycle. In the present study, high CDK1 expression was strongly associated with poor prognosis in HCC patients. KEGG pathway also showed that CDK1 was mainly involved in cell cycle, Oocyte meiosis, Cellular Senescence, P53 signalling Pathway and other signalling pathways. Studies have found that CCNB1/CDK1 can mediate mitochondria, coordinate ATP output, and provide biological energy for cell G2/M transformation, thus interrupting the cell replication cycle [15, 16]. Therefore, CDK1 plays an essential role in the cell cycle and cell proliferation.

Anaphase Promoting Complex (APC) was activated by cell division cycle 20 (CDC20) and CDH1 to form two different E3 ubiquitin ligase complexes sub-complexes, APC^{Cdc20} and APC^{Cdh1}. CDC20 is an oncogenic factor and CDH1 is a tumour suppressor factor [17]. It is currently believed that the APC^{Cdc20} complex can disrupt the ubiquitination of downstream cell cycle regulators securin and cyclin B [18] and regulate the activity of downstream pluripotent related transcription factors SOX2 to promote the invasion and renewal of glioma stem cells [19]. The degree of CDC20 expression has now been found to correlate significantly with tumour prognosis, with higher CDC20 expression associated with lower 5-year survival rates for patients. Studies have shown that the expression level of CDC20 increases in non-small cell lung cancer (NSCLC), resulting in the more significant pleural invasion of cancer cells, while the growth rate of lung cancer cells is significantly slowed down after the expression level of CDC20 is down-regulated [20]. In our study, the high expression of CDC20 also predicted the poor prognosis of HCC patients, and CDC20 was mainly enriched in cell cycle, oocyte meiosis and other pathways. A related study found that CDC20 was overexpressed in 68% of hepatocellular carcinoma tissues [21], that overall survival was lower in HCC patients with high CDC20 expression, and that CDC20 was predictive of prognosis in HCC patients [22]. In addition, numerous studies have found that anti-mitotic agents such as paclitaxel can inhibit APC^{Cdc20}, and CDC20 gene knockout can lead to a checkpoint-independent mitotic block, which can effectively kill cancer cells [23, 24]. All of these suggest that CDC20 is a potential target for the treatment of HCC.

The CCNA2 gene encodes a protein that belongs to a family of highly conserved cell cycle proteins, whose members function to regulate the cell cycle. CCNA2 binds and activates cyclin-dependent kinase 2, thereby facilitating the G1/S and G2/M transitions. Studies have shown that p53-P21-DREAM-CDE /CHR pathway can down-regulate the CCNA2 gene, thereby blocking the G2/M cell cycle and controlling the growth of cancer cells [25]. This article also found that high CCNA2 expression was significantly associated with the prognosis of HCC patients and that the proliferation of hepatocellular carcinoma cells could be inhibited by drugs that inhibit the FXR-miR-22-CCNA2 signalling pathway [26]. The current studies have found that CCNA2 is highly expressed in all types of tumours. For example, E2F1 can

promote the proliferation of triple-negative breast cancer (TNBC) through up-regulation of CCNA2 [27]. miR-545 can inhibit proliferation by inhibiting CCNA2 expression in osteosarcoma cells [28]. In addition, the high expression of CCNA2 has been reported in lung cancer [29], gastric cancer [30], ovarian cancer [31] and esophageal cancer [32].

Multiple bioinformatics methods were used to jointly screen out potential targets for HCC. This study indicated that the high expression of CDK1, CDC20 and CCNA2 was negatively correlated with the survival rate of HCC patients, and these targets are mainly involved in regulating the cell cycle and thus controlling the proliferation of tumour cells. These newly discovered potential targets may provide new ideas and new research directions for the treatment of HCC and have important clinical significance. However, our study also had many limitations. First, the sample size we selected was not large enough for a specific analysis of subtypes in HCC, which may lead to inaccurate hub genes screening and insufficient coverage. Secondly, the reliability of Hub genes has not been verified experimentally in this paper, and the specific mechanism of hub genes regulating HCC is not fully understood, which needs further exploration in the future.

Conclusion

In summary, our combined analysis of the WGCNA and TCGA and GEO databases has uncovered CDK1, CDC20 and CCNA2 as three potential targets that may be able to diagnose, analyse prognosis and even treat liver cancer. We also further validated the expression of real hub genes in HCC tissues and normal tissues through tissue expression differences and immunohistochemistry. We also analysed that CDK1, CDC20 and CCNA2 control tumour growth mainly by regulating the cell cycle, but we lack further experimental validation. These findings may provide ideas for our subsequent studies and provide a reference for the more complex study of hepatocellular carcinoma to investigate this complex mechanism of hepatocellular carcinoma treatment.

Methods

Data Collection and Data Processing

The Cancer Genome Atlas (TCGA) [33] database downloaded 424 HTSeq-Counts expression data on HCC and clinical information (https://portal.gdc.cancer.gov/), They included 374 cancerous tissue samples and 50 normal tissue samples. The genetic transformation was performed using Perl language, and, 19600 genes were analyzed. Downloaded from GEO database (https://www.ncbi.nlm.nih.gov/geo/) on human liver cancer cell gene expression profile, select sample than 50 sets of data, eventually, take GSE121248 [34] as the data source, It was provided by Hui KM based on the GPL570 platform ([HG-U133_PLus_2] Affymetrix Human Genome U133 Plus 2.0 Array) and included 37 para cancer tissue samples and 70 tumour tissue samples. Probes were converted to gene symbols according to the annotation file, and duplicate probes for the same genes were removed by determining the median

expression values of all corresponding probes. The results resulted in a total of 21,654 genes screened for subsequent analysis.

Differentially Expressed Genes Identification

Significance analysis of differential gene expressions (DEGs) between tumour tissue and normal tissue samples was performed using the Limma package in R language [35], with set criteria of $|logFC| \geq 1$ and adjusted p-value < 0.05. And two R language packages, "pheatmap" and "ggplot2", were used to create the heatmap and volcano map.

Weighted Gene Co-expression Network Analysis and Overlapping Genes Construction

WGCNA is an analytical method for analysing the gene expression profiles of multiple samples. It allows clustering of genes with similar expression patterns and analysis of the relationship between modules and specific traits or phenotypes. Distances between different transcripts were calculated using Pearson correlation coefficients. In our study, the "WGCNA" [5] software package of R language was used to implement Weighted Gene co-expression Network Analysis (WGCNA) and construct a weighted co-expression network for gene expression data of HCC. The weighted adjacency matrix is transformed into a topological overlap matrix (TOM) to detect network interactions and a hierarchical clustering approach is used to build a clustering tree structure for the TOM matrix. Eventually, genes with similar expression patterns are combined into the same module. The "VennDiagram" R language package [36] was used to draw the Venn diagram of TCGA and modules in GSE121248 that are positively correlated with tumour tissues.

GO and KEGG Function Enrichment Analysis of Gene Modules

GO enrichment analysis can effectively predict the biological characteristics of overlapping genes, and KEGG can effectively analyze the pathways of gene function. Functional annotation of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) based on R language package clusterProfiler [37] pathway enrichment analysis. Adjusted p-value < 0.05 was considered to have statistical significance.

Identification of Hub Genes in Functional Modules

The intersection of TCGA-LIHC and tumour tissue positively related modules in GSE121248 is obtained to obtain overlapping genes, which is imported into STRING 11.5 (https://cn.string-db.org/) [38] to construct Protein-Protein Interaction (PPI) network diagram. Set confidence score ≥ 0.90 and import the results into Cytoscape 3.9.0 (http://www.cytoscape.org) to build network visualization images. The CytoHubba plugin in Cytoscape was used to calculate each of its nodes, where the maximum cluster centrality (MCC) algorithm was considered to be the most efficient way to find the central node, and the top 10 genes with the largest weight coefficients were obtained for subsequent analysis [39].

Hub Genes Clinical Trait Validation and Survival Analysis

To better verify the relationship between hub genes and HCC patients, the Transcripts Per Million (TPM) was used as the standardized value for sample difference analysis, and the expression point map of hub genes in normal tissues, as well as tumour tissues, were plotted. To obtain the relationship between hub genes and clinical characteristics, a survival R language package was used to analyze the clinical information of HCC patients downloaded from TCGA with hub genes. The R package "Survival" and "SurvMiner" used in survival analysis need to perform the log-rank test and set p-value < 0.05. Kaplan-Meier curve was plotted by expression spectrum, and overall survival (OS) curves were plotted for HCC patients by dividing them into two groups based on the median expression values of the hub genes. The online tool Gene Expression Profilings Interactive Analysis (http://gepia.cancer-pku.cn/) [40] was used to plot disease-free survival (DFS). The genes related to the survival of HCC patients were obtained, and then compared with the genes contained in the top 10 pathways of KEGG enrichment analysis, and finally, 3 real hub genes were identified for subsequent analysis.

Validation of Hub Genes Protein Expression by the Human Protein Atlas Database

The Human Protein Atlas database (HPA) (https://www.proteinatlas.org/), which is based on proteomics, transcriptome and systems biology data to tissue, cells, organs such as mapping [41]. Not only tumour tissue but also normal tissue protein expression were included. We used the HPA database to validate the protein expression of real hub genes in normal and HCC tissues based on the immunohistochemical (IHC) method.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

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

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

YQ and XW designed the study. YQ, JH and YM acquired and analyzed the data. YQ, XW, JH and FY discussed the results. YQ, XW, JH, FY and ZZ wrote the manuscript. FY and ZZ reviewed and supervised the work. All authors reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Authors' information

[1] School of Medicine, Jianghan University, Wuhan Hubei, China. [2] Department of Spine Surgery, Wuhan Fourth Hospital, Wuhan Hubei, China

References

1. Chidambaranathan-Reghupaty S, Fisher P, Sarkar D. Hepatocellular carcinoma (HCC): Epidemiology, etiology and molecular classification[J]. Advances in cancer research, 2021, 149: 1-61.
2. Kulik L, El-Serag H. Epidemiology and Management of Hepatocellular Carcinoma[J]. Gastroenterology, 2019, 156(2): 477-491.e1.
3. Hepatocellular carcinoma[J]. Nature reviews. Disease primers, 2021, 7(1): 7.
4. Li N, Zhan X. Identification of clinical trait-related lncRNA and mRNA biomarkers with weighted gene co-expression network analysis as useful tool for personalized medicine in ovarian cancer[J]. The EPMA journal, 2019, 10(3): 273-290.
5. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis[J]. BMC bioinformatics, 2008, 9: 559.
6. Li J, Ling W, Lu Q, et al. Liver Stiffness and Serum Alpha-Fetoprotein in Discriminating Small Hepatocellular Carcinoma from Cirrhotic Nodule[J]. Ultrasound quarterly, 2016, 32(4): 319-326.
7. Deng Q, Li K, Chen H, et al. RNA interference against cancer/testis genes identifies dual specificity phosphatase 21 as a potential therapeutic target in human hepatocellular carcinoma[J]. Hepatology (Baltimore, Md.), 2014, 59(2): 518-30.
8. Jiang Y, Sun A, Zhao Y, et al. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma[J]. Nature, 2019, 567(7747): 257-261.
9. Zhang Y, Huang W, Ran Y, et al. miR-582-5p inhibits proliferation of hepatocellular carcinoma by targeting CDK1 and AKT3[J]. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine, 2015, 36(11): 8309-16.
10. Hao L, Li S, Peng Q, et al. CDK1Anti-malarial drug dihydroartemisinin downregulates the expression levels of and in liver cancer[J]. Oncology letters, 2021, 22(3): 653.

11. Shi Q, Zhou Z, Ye N, et al. MiR-181a inhibits non-small cell lung cancer cell proliferation by targeting CDK1[J]. Cancer biomarkers : section A of Disease markers, 2017, 20(4): 539-546.
12. Qian J, Gao J, Sun X, et al. KIAA1429 acts as an oncogenic factor in breast cancer by regulating CDK1 in an N6-methyladenosine-independent manner[J]. Oncogene, 2019, 38(33): 6123-6141.
13. Huang J, Chen P, Liu K, et al. CDK1/2/5 inhibition overcomes IFNG-mediated adaptive immune resistance in pancreatic cancer[J]. Gut, 2021, 70(5): 890-899.
14. Zhang P, Kawakami H, Liu W, et al. Targeting CDK1 and MEK/ERK Overcomes Apoptotic Resistance in BRAF-Mutant Human Colorectal Cancer[J]. Molecular cancer research : MCR, 2018, 16(3): 378-389.
15. Xie B, Wang S, Jiang N, et al. Cyclin B1/CDK1-regulated mitochondrial bioenergetics in cell cycle progression and tumor resistance[J]. Cancer letters, 2019, 443: 56-66.
16. Wang Z, Fan M, Candas D, et al. Cyclin B1/Cdk1 coordinates mitochondrial respiration for cell-cycle G2/M progression[J]. Developmental cell, 2014, 29(2): 217-32.
17. Wang L, Zhang J, Wan L, et al. Targeting Cdc20 as a novel cancer therapeutic strategy[J]. Pharmacology & therapeutics, 2015, 151: 141-51.
18. Holland A, Cleveland D. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis[J]. Nature reviews. Molecular cell biology, 2009, 10(7): 478-87.
19. Mao D, Gujar A, Mahlokozera T, et al. A CDC20-APC/SOX2 Signaling Axis Regulates Human Glioblastoma Stem-like Cells[J]. Cell reports, 2015, 11(11): 1809-21.
20. Kato T, Daigo Y, Aragaki M, et al. Overexpression of CDC20 predicts poor prognosis in primary non-small cell lung cancer patients[J]. Journal of surgical oncology, 2012, 106(4): 423-30.
21. Li J, Gao J, Du J, et al. Increased CDC20 expression is associated with development and progression of hepatocellular carcinoma[J]. International journal of oncology, 2014, 45(4): 1547-55.
22. Zhang X, Zhang X, Li X, et al. Connection Between CDC20 Expression and Hepatocellular Carcinoma Prognosis[J]. Medical science monitor : international medical journal of experimental and clinical research, 2021, 27: e926760.
23. Huang H, Shi J, Orth J, et al. Evidence that mitotic exit is a better cancer therapeutic target than spindle assembly[J]. Cancer cell, 2009, 16(4): 347-58.
24. Wang Z, Wan L, Zhong J, et al. Cdc20: a potential novel therapeutic target for cancer treatment[J]. Current pharmaceutical design, 2013, 19(18): 3210-4.
25. Fischer M, Quaas M, Steiner L, et al. The p53-p21-DREAM-CDE/CHR pathway regulates G2/M cell cycle genes[J]. Nucleic acids research, 2016, 44(1): 164-74.
26. Yang F, Gong J, Wang G, et al. Waltonitone inhibits proliferation of hepatoma cells and tumorigenesis via FXR-miR-22-CCNA2 signaling pathway[J]. Oncotarget, 2016, 7(46): 75165-75175.
27. Lu Y, Su F, Yang H, et al. E2F1 transcriptionally regulates CCNA2 expression to promote triple negative breast cancer tumorigenicity[J]. Cancer biomarkers : section A of Disease markers, 2021.
28. Li L, Kong X, Zang M, et al. Hsa_circ_0003732 promotes osteosarcoma cells proliferation via miR-545/CCNA2 axis[J]. Bioscience reports, 2020, 40(6).

29. Gao M, Kong W, Huang Z, et al. Identification of Key Genes Related to Lung Squamous Cell Carcinoma Using Bioinformatics Analysis[J]. International journal of molecular sciences, 2020, 21(8).
30. Lee Y, Lee C, Oh S, et al. CCNA2Pharmacogenomic Analysis Reveals as a Predictive Biomarker of Sensitivity to Polo-Like Kinase I Inhibitor in Gastric Cancer[J]. Cancers, 2020, 12(6).
31. Guo F, Zhang K, Li M, et al. miR-508-3p suppresses the development of ovarian carcinoma by targeting CCNA2 and MMP7[J]. International journal of oncology, 2020, 57(1): 264-276.
32. Wang H, Fu L, Wei D, et al. MiR-29c-3p Suppresses the Migration, Invasion and Cell Cycle in Esophageal Carcinoma via CCNA2/p53 Axis[J]. Frontiers in bioengineering and biotechnology, 2020, 8: 75.
33. Colaprico A, Silva T, Olsen C, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data[J]. Nucleic acids research, 2016, 44(8): e71.
34. Wang S, Ooi L, Hui K. Identification and validation of a novel gene signature associated with the recurrence of human hepatocellular carcinoma[J]. Clinical cancer research : an official journal of the American Association for Cancer Research, 2007, 13(21): 6275-83.
35. Ritchie M, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. Nucleic acids research, 2015, 43(7): e47.
36. Chen H, Boutros P. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R[J]. BMC bioinformatics, 2011, 12: 35.
37. Yu G, Wang L, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. Omics : a journal of integrative biology, 2012, 16(5): 284-7.
38. Szklarczyk D, Gable A, Nastou K, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets[J]. Nucleic acids research, 2021, 49: D605-D612.
39. Chin C, Chen S, Wu H, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome[J]. BMC systems biology, 2014: S11.
40. Tang Z, Kang B, Li C, et al. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis[J]. Nucleic acids research, 2019, 47: W556-W560.
41. Thul P, Lindskog C. The human protein atlas: A spatial map of the human proteome[J]. Protein science : a publication of the Protein Society, 2018, 27(1): 233-244.

Table

Tables 1 is available in the Supplementary Files section

Figures

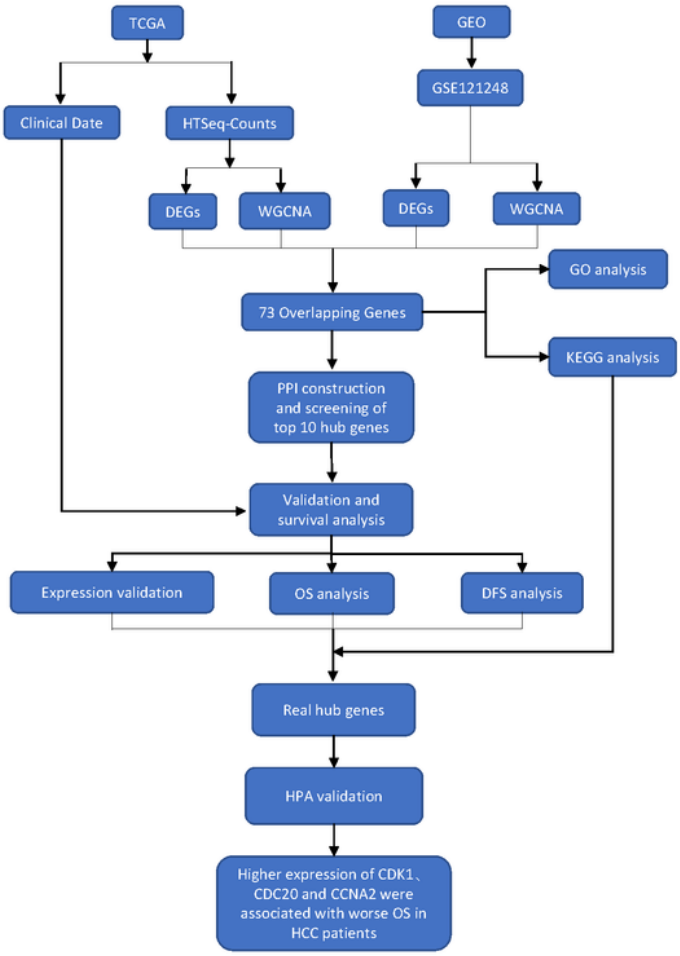

Figure 1

The flow chart of present study.

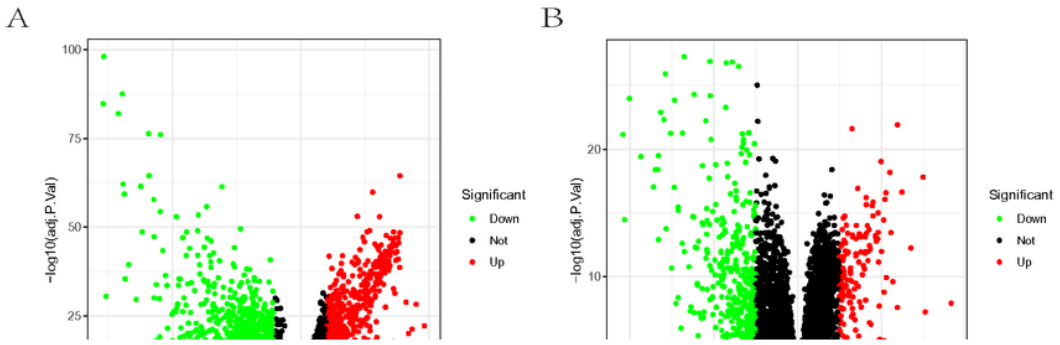

Figure 2

The volcano plot of DEGs between HCC and normal tissue. (A)Volcano plot of DEGs in the TCGA-LIHC dataset. (B) Volcano plot of DEGs in the GSE121248 dataset.

Figure 3

The heatmaps of DEGs between HCC and normal tissues. (A)The heatmap of DEGs obtained from TCGA-LIHC dataset.

Figure 4

Weighted gene co-expression modules construction and selection from TCGA-LIHC.(A) The left graph shows the scale free fit index for soft-threshold powers. The right graph displays the mean connectivity

for soft-threshold powers. (B) Hierarchical clustering analyses in TCGA-LIHC. (C)Dendrogram of all co-expression genes modules were clustered based on a dissimilarity measure(1-TOM).

Figure 5

Weighted gene co-expression modules construction and selection from GES121248. (A) The left graph shows the scale free fit index for soft-threshold powers. The right graph displays the mean connectivity for soft-threshold powers. (B) Hierarchical clustering analyses in GES121248. (C)Dendrogram of all co-expression genes modules were clustered based on a dissimilarity measure(1-TOM).

Figure 6

The heatmaps of correlation between clinical traits and the modular genes.(A)The heatmap of traits and modular genes constructed from TCGA-LIHC. (B)The heatmap of traits and modular genes

Figure 7

MM and GS correlation analysis scatterplots.（A）Scatterplot of MM and GS correlation analysis of yellow module gene in TCGA-LIHC.（B）Scatterplot of MM and GS correlation analysis of turquoise module gene inGSE121248.

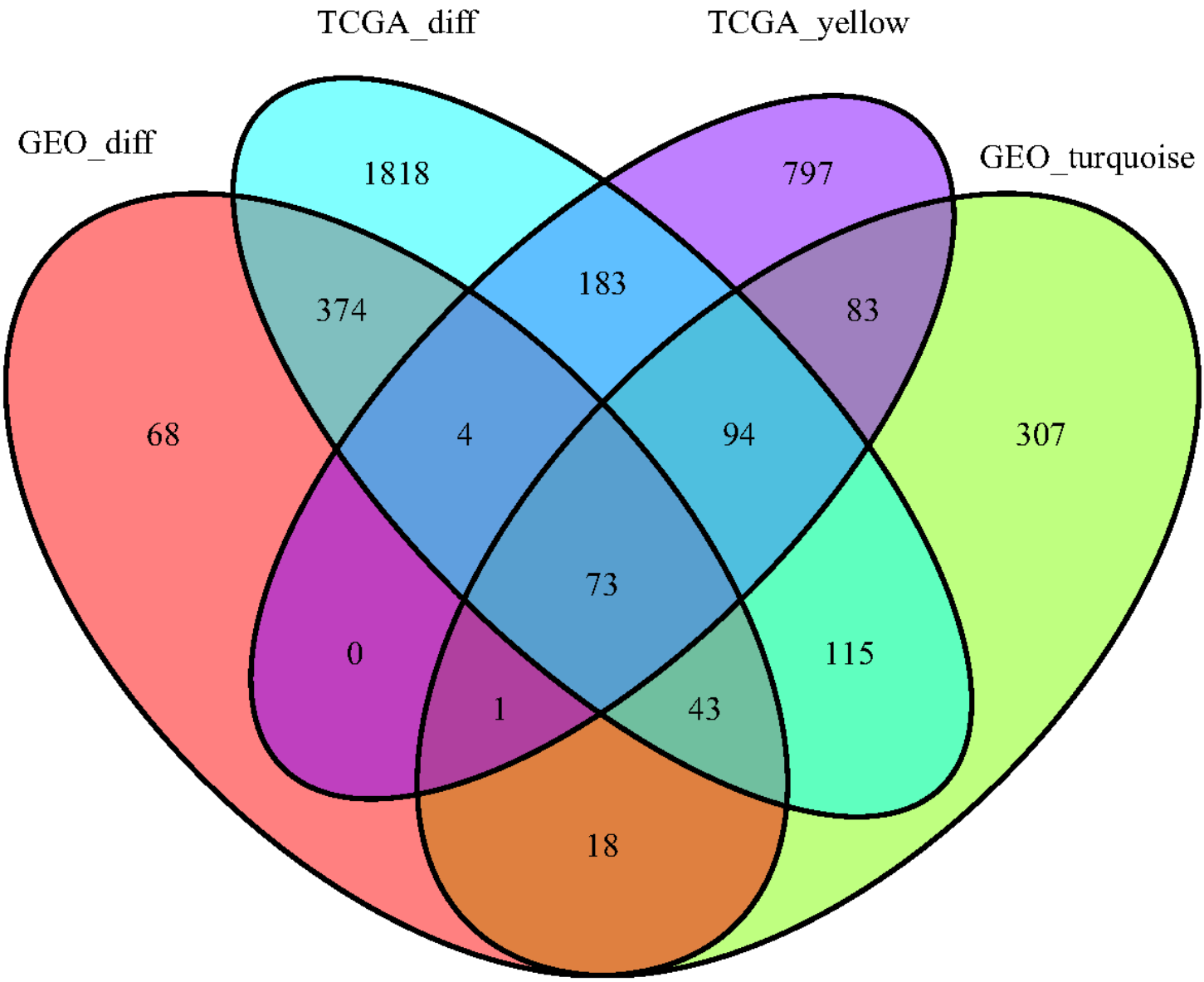

Figure 8

The Venn diagram of DEGs and co-expression genes. As a result, 73 overlapping genes were obtained from the intersection of DEGs and co-expression.

Figure 9

Functional enrichment analysis of the 73 Overlapping Genes.(A)GO enrichment analysis.(B)KEGG pathway enrichment analysis.

The size of the bubble represents the number of enriched genes and the color of the bubble represents the p-value.

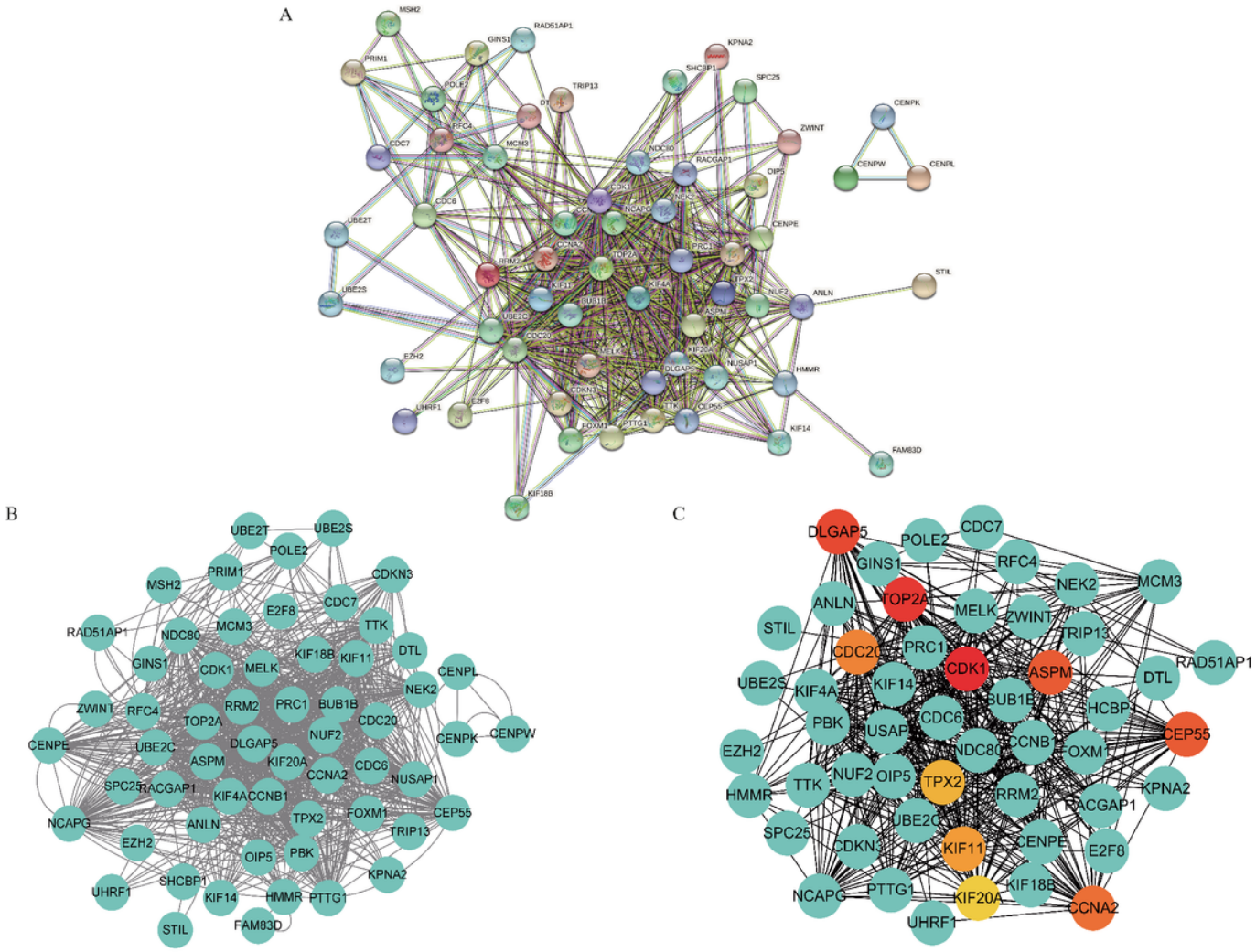

Figure 10

Construction of PPI network and screening of Hub genes based on CytoHubba. (A) Construct overlapping genes PPI network based on STRING. (B)PPI network of overlapping genes was constructed based on Cytoscape. (C)The top 10 genes with the highest MCC score were screened from PPI.

Figure 11

The Expression Level of Hub Genes in HCC and normal tissues was verified based on TCGA-LICH. (A)ASPM, (B)CCNA2, (C)CDC20, (D)CDK1, (E)CEP55, (F)DLGAP5, (G)KIF11, (H)KIF20A, (I)TOP2A, (J)TPX2.

Figure 12

Validation of Overall Survival (OS) Curve of Hub Genes. (A)ASPM, (B)CCNA2, (C)CDC20, (D)CDK1, (E)CEP55, (F)DLGAP5, (G)KIF11, (H)KIF20A, (I)TOP2A, (J)TPX2.

Figure 13

Disease-free survival analysis of Hub genes. (A)ASPM, (B)CCNA2, (C)CDC20, (D)CDK1, (E)CEP55, (F)DLGAP5, (G)KIF11, (H)KIF20A, (I)TOP2A, (J)TPX2.

Figure 14

Immunohistochemistry of the real hub genes in tumor(top) and normal tissues(base) from the human protein atlas(HPA) datebase. (A) CDK1, (B)CDC20, (C)CCNA2. represents liver tumor tissues. represents normal tissues.

Supplementary Files

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

- Table1.pdf