

# Identification of novel potential biomarkers in hepatocarcinoma cancer; a transcriptome analysis

## **Mohib kakar**

Beijing Key Laboratory for Separation and Analysis in Biomedicine and Pharmaceutical, School of life Sciences, Beijing Institute of Technology (BIT), Beijing, 100081, PR China

## **Muhammad Mehboob**

CAS Centre for Excellence in Biotic Interaction, College of Life Sciences, University of Chinese Academy of Science, Beijing, 100049, China

## **Muhammad Akram**

Institute for Synthetic Biosystem, School of Chemistry and Chemical Engineering, Beijing Institute of Technology, Beijing 100081, China

## **Imran Iqbal**

Department of Information and Computational Sciences, School of Mathematical Sciences and LMAM, Peking University, Beijing, 100871, China

## **Hafza Ijaz**

CAS Centre for Excellence in Biotic Interaction, College of Life Sciences, University of Chinese Academy of Science, Beijing, 100049, China

## **Ubair Aziz**

Research Centre of Molecular Simulation, National University of Science and Technology, Islamabad

## **Adil Lodhi**

Beijing Key Laboratory for Separation and Analysis in Biomedicine and Pharmaceutical, School of life Sciences, Beijing Institute of Technology (BIT), Beijing, 100081, PR China

## **Madiha Rasheed**

Beijing Key Laboratory for Separation and Analysis in Biomedicine and Pharmaceutical, School of life Sciences, Beijing Institute of Technology (BIT), Beijing, 100081, PR China

## **Yulin Deng**

Beijing Key Laboratory for Separation and Analysis in Biomedicine and Pharmaceutical, School of life Sciences, Beijing Institute of Technology (BIT), Beijing, 100081, PR China

## **Rongji Dai (✉ [dairongji@bit.edu.cn](mailto:dairongji@bit.edu.cn))**

Beijing Key Laboratory for Separation and Analysis in Biomedicine and Pharmaceutical, School of life Sciences, Beijing Institute of Technology (BIT), Beijing, 100081, PR China

**Keywords:** Cyclin, Gene Expression Omnibus; differentially expressed genes, Hepatocellular carcinoma, bioinformatics analysis

**Posted Date:** March 2nd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-154350/v3>

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

[Read Full License](#)

---

# Abstract

## Objective

The goal of this study was **to** understand possible core genes associated with hepatocellular carcinoma (HCC) pathogenesis and prognosis.

## Methods

GEO contains datasets of gene expression, miRNA and methylation patterns of diseased and healthy/control patients. GSE62232 Dataset was selected by employing the server Gene Expression Omnibus. A total of 91 samples were collected, including 81 HCC samples and 10 healthy samples as control. GSE62232 was analyzed through GEO2R, and Functional Enrichment Analysis was performed to extract rational information from a set of DEGs. The Protein-Protein Relationship Networking search method has been used for extracting genes interacting. MCC method was used to calculate the top 10 genes according to their importance. Hub genes in the network were analyzed using GEPIA to estimate the effect of their differential expression on cancer progression.

## Results

We identified the top 10 hub genes through Cytohubba plugin. These genes include Cell Cycle Regulatory Cyclins and Cyclin-dependent proteins CCNA2, CCNB1 and CDK1. The pathogenesis and prognosis of HCC may be directly linked with the aforementioned genes.

## Conclusion

In this analysis, we found critical genes for HCC that showed recommendations for more diagnostic and predictive biomarkers studies that could promote selective molecular therapy for HCC.

## 1. Introduction

Cancer, also known as malignancy, is irregular cell growth. More than 100 different types of cancer exist, in which most common are breast, skin, prostate, lungs, colon, and lymphoma [1]. Cancer is present in human as the most considerable public health concern worldwide, and liver cancer adds greatly to the morbidity and mortality in cancer [2]. Liver cancer (hepatocellular carcinoma) is the fourth leading cause of cancer-related death globally, ranking sixth in prevalence [3, 4]. Hepatocellular carcinoma (HCC) constitutes about 85–90% of all primary malignant liver tumors. Chronic hepatitis B virus (HBV) infection, hepatitis C virus (HCV), smoking, aflatoxin, obesity, chronic liver disease, and type 2 diabetes are the main risk factors [3, 5, 6]. Of these variables, recurrent liver disease is the primary cause of liver cancer [7].

The prevalence of viral infection in HCC cases varies from developed to developing nations, where HBV reflects 60% in developing nations and 23% in developed nations, while HCV infection is responsible for 23% in emerging nations, and 20% of patients in developed nations [8]. Moreover, the highest incidence of HBV is in sub-Saharan Africa, South-eastern Asia, and East Asia, while HCV is high in the USA, Europe, and Japan [7]. The prevalence of non-alcoholic fatty liver disease (NAFLD) also adversely affects individual health causing an increase in obesity and other metabolic disorders [9]. Around 25–30% of patients having a western lifestyle possess more fats in their liver, 2–5% of which have NAFLD, and 1–2% suffer from non-alcoholic steatohepatitis cirrhosis [10].

The World Health Organization (WHO) estimates that in 2030 over one million people are going to die of liver cancer [11]. The key factor that affects HCC mortality is a poor diagnosis, which results in just 18% survival [12] less than the cancers of the breast (77.1%), renal pelvic (74.8%), and myeloma (52.2%). The high risk of recurrence and metastasis of the HCCs also contribute to a shorter life span and poor survival after hepatectomy [13]. Different variables participate in HCC diagnoses, such as cell proliferation, apoptosis, and genes linked to the mTOR pathway [14]. HCC is on a global increase, but early detection and therapy of HCC remain a concern [7]. In developing countries, the HCC prevalence is growing as a consequence of low levels of health and treatment, with a global rate of liver cancer per 100000 people approximately at 9.3 in 2018 [15], as well as poor prognosis [16].

The diverse factors implicated in liver cancer are cellular tumor antigen p53 (TP53), axin-1 (AXIN1), catenin  $\beta$ -1 (CTNNB1), and telomerase reverse transcriptase (TERT) promoters as well as other primary genes for mutation generation, p53 cell cycle system, WNT/ $\beta$ -catenin, oxidative stress, RAS/RAF/MAPK, and PI3K/AKT/MTOR pathways along with other main primary signaling pathways. Liu et al. (2018) used highly efficient microarray technology to screen molecular indicators across all human cancerous tumors, especially for liver cancer, by using Gene Expression Omnibus (GEO) datasets, The Cancer Genome Atlas (TCGA) RNA-sequence and analyzed with the help of bioinformatics methods [17]. Gene chip technology can also reliably represent the molecular expression profile and detect genetic variants correlated with HCC in liver cancer studies [18, 19]. The data, information, knowledge, and wisdom (DIKW) model is widely used in life, including medicine [20–22]. Recently, genome-wide screening has significantly improved the knowledge of the genetic context and pathways that lead to HCC [23–26].

Four core genes and two essential pathways of developing HCC from cirrhosis have been established by GEO dataset using a bioinformatics methodology, including DEG screening and networking of protein-protein interactions (PPIs) [27]. Zhang et al. screened the genes and pathways associated with HCC development and prevalence through a series of bioinformatics observations, such as DEG recognition, functional enrichment analysis, PPI network and module analysis, and weighted network correlation analysis [19]. Zhou et al. identified HCC critical genes and microRNAs through raw data processing by using Gene Ontology (GO), GEO2R, and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway enrichment processing and PPI network creation [28]. Established research, nevertheless, have been of small scale, and the lack of molecular prognostic indexes might have restricted the ability to recognize possible important molecular biomarkers and emerging clinical goals.

Applied bioinformatics research with the current genomic evidence offers an in-depth insight into therapeutic resistance and disease progression processes. This study focuses on the expression profiling of HCC patients compared to healthy ones. GSE62232 dataset (GEO: <https://www.ncbi.nlm.nih.gov/geo/>) has been chosen. GSE62232 was analysed through GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2R>) [29] to evaluate and recalculate the genes that are differentially expressed in healthy and diseased samples. However, new diagnostic and prognostic biomarkers are needed to improving HCC diagnosis and treatment.

## **2. Material And Methods**

### **2.1. Data Collection and Expression Profiling of DEGs**

GEO contains datasets of gene expression, miRNA and methylation patterns of diseased and healthy/control patients [30]. This study was focused on the expression profiling of HCC patients in comparison to healthy ones. A total of 91 samples were collected, including 81 HCC samples and 10 healthy samples as control [31]. GSE62232 was analysed through GEO2R [29] to evaluate and recalculate genes expressed uniquely in healthy and unhealthy samples. Analysis provided a lot of genes whose expression differed in both samples. To minimize the background noise, statistical filters of *p*-value and fold-change value were used.

### **2.2. Functional Enrichment Annotation of DEGs**

Functional enrichment analysis was performed to extract rational information from a set of DEGs. This provided us with the most prominent pathways being affected by the change of gene expression in diseased samples. The database named Database for Annotation, Visualization, and Integrated Discovery (DAVID; <http://david.ncifcrf.gov>) has been used to perform functional enrichment analysis. DAVID annotates the provided set of genes into the most affected processes such as Biological Processes (BP), Molecular Functions (MF), and Cellular Components (CC) [32]. We used GO and KEGG datasets in DAVID to annotate the most affected processes. KEGG is a database of molecular pathways containing detailed information about functional and biological systems [33]. The *p*-value filtering  $\leq 0.05$  was used to exclude the results with a low confidence level.

### **2.3. Protein-Protein Interaction Network Backbone Analysis**

Search Tool for the Retrieval of Interacting Genes (STRING; <http://string-db.org>) has been used to establish PPIs Network [34]. A confidence score of  $\geq 0.07$  was used to scrutinize statistically significant results. The PPIs network was visualized using Cytoscape (open-source platform), in which the most considerable module with a Molecular Complex Detection (MCODE) score of  $> 5$  and node score-cut-off = 0.02 was scanned with the Cytoscape plug-in MCODE.

### **2.4. Hub Gene Identification, Expression and Survival Analysis**

Cytohubba is a Cytoscape module that has been used in the PPI network for identifying primary hub genes. MCC method was used to calculate the top 10 genes according to their importance in the system because it is comparatively recent and is highly recommended [35]. These top 10 genes having the highest MCC score were considered as hub genes. However, the Gene Expression Profiling Interactive Analysis (GEPIA) server provided a correlation of gene expression and their effect on the survival chances in specific cancer types [36]. Hub genes were analysed using GEPIA to calculate the impact of their differential expression on cancer progression.

## 2.5. Functional Annotation of MCODE Cluster Genes

ClueGO plugin was utilized to functionally annotate the top 3 MCODE Cluster genes into the biological process. It classified gene products into crucial biological processes using GO datasets as the reference [37].

## 3. Results

### 3.1. DEGs Identification

Expression profile of genes for GSE62232, titled in NCBI “large-scale gene expression profiling of 81 hepatocellular carcinomas” was obtained from GEO dataset, which was generated on the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array). After getting data of 91 liver samples (81 HCC patients and 10 control) of GSE62232 study, the data was analysed through GEO2R to identify the DEGs of both upregulated and downregulated genes using a  $p$ -value of  $< 0.05$  and  $\log FC > 1.5$  as selection criteria. Overall, 598 genes out of a total 19982 were differentially expressed in HCC samples with 233 upregulated and 365 downregulated genes (**Supplementary file 1**). The DEGs are represented by a volcano plot constructed using Prism ([www.graphpad.com/scientific-software/prism](http://www.graphpad.com/scientific-software/prism)) for a better graphical representation of overall genes (Fig. 1).

### 3.2. Enrichment Analysis of DEGs

In order to explore GO terms and cellular mechanisms affected by these DEGs, both over and down-regulated genes were imported into an online DAVID server to conduct the annotation process. The annotated GO terms were divided into MF, BP, CC ontologies ( $p$ -value  $< 0.05$ , FDR  $< 0.05$ ). The GO BP analysis revealed that the majority of DEGs were enriched into oxidation-reduction (GO:0055114), metabolic (steroids, drug, xenobiotic) and catabolic (exogenous drug, tryptophan), cell division (GO:0051301), response to the drug (GO:0042493) and p450 pathways (GO:0097267). For the GO MF analysis, most of the genes were significantly enriched in oxygen (GO:0019825) and iron (GO:0005506) ion binding, oxidoreductase activity (GO:0016705), monooxygenase activity (GO:0004497), heme-binding (GO:0020037), steroid hydroxylase (GO:0008395), and electron carrier activity (GO:0009055). Concerning the GO CC analysis, commonly DEGs were enriched in extracellular space (GO:0070062, GO:0005615), organelle membrane (GO:0031090), and condensed chromosome kinetochore (GO:0000777) (Table 1).

Table 1

Top 10 GO terms regarding biological process, molecular function, and cellular components after applying p-value < 0.05 and FDR < 0.05 filter

GO Terms	Count	P-value	FDR
Biological Process (BP)			
GO:0055114 ~ oxidation-reduction process	60	9.88E-16	1.78E-12
GO:0019373 ~ epoxygenase P450 pathway	12	5.33E-13	9.45E-10
GO:0042738 ~ exogenous drug catabolic process	8	1.70E-08	3.01E-05
GO:0008202 ~ steroid metabolic process	12	4.80E-08	8.52E-05
GO:0017144 ~ drug metabolic process	10	6.41E-08	1.14E-04
GO:0006805 ~ xenobiotic metabolic process	15	9.23E-08	1.64E-04
GO:0097267 ~ omega-hydroxylase P450 pathway	6	3.03E-06	0.00537
GO:0006569 ~ tryptophan catabolic process	6	3.03E-06	0.00537
GO:0042493 ~ response to drug	26	8.00E-06	0.0142
GO:0051301 ~ cell division	28	1.13E-05	0.02011
Molecular Function (MF)			
GO:0019825 ~ oxygen binding	16	3.59E-12	5.58E-09
GO:0005506 ~ iron ion binding	25	2.89E-11	4.50E-08
GO:0016705 ~ oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	16	7.99E-11	1.24E-07
GO:0004497 ~ monooxygenase activity	16	1.05E-10	1.63E-07
GO:0020037 ~ heme binding	22	7.83E-10	1.22E-06
GO:0008392 ~ arachidonic acid epoxygenase activity	9	3.30E-09	5.14E-06

GO Terms	Count	P-value	FDR
GO:0008395 ~ steroid hydroxylase activity	10	3.61E-08	5.62E-05
GO:0016491 ~ oxidoreductase activity	23	1.56E-07	2.43E-04
GO:0070330 ~ aromatase activity	9	8.28E-07	0.00129
GO:0009055 ~ electron carrier activity	14	2.70E-06	0.0042
GO:0016712 ~ oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen	7	2.80E-06	0.00436
GO:0015171 ~ amino acid transmembrane transporter activity	10	8.99E-06	0.01399
Cellular Components (CC)			
GO:0070062 ~ extracellular exosome	152	1.98E-14	2.72E-11
GO:0031090 ~ organelle membrane	21	6.51E-13	8.96E-10
GO:0005615 ~ extracellular space	80	2.68E-09	3.68E-06
GO:0000777 ~ condensed chromosome kinetochore	13	9.00E-06	0.01238
GO:0072562 ~ blood microparticle	17	1.10E-05	0.01514
GO:0005829 ~ cytosol	138	1.16E-05	0.01594
GO:0005576 ~ extracellular region	78	1.45E-05	0.01989
GO:0030496 ~ midbody	15	2.75E-05	0.03775

The KEGG signaling pathways examination of DEGs by applying the filter of  $p$  value < 0.05 and FDR < 0.05 sorted mainly metabolic pathways such as retinol metabolism (hsa00830), biosynthesis and metabolism of amino acids (hsa01230), antibiotics (hsa01130), and drug metabolism-cytochrome P450 (hsa00982) (Table 2). Some additional signaling pathways i.e., chemical carcinogenesis (hsa05204), cell

cycle (hsa04110), and p53 signaling pathway (hsa04115) were also markedly enriched by DEGs. The complete list of GO terms and KEGG pathways are enlisted in **supplementary file 2**.

Table 2  
Top 10 pathways in which most of the DEGs enriched at p-value < 0.05 and FDR < 0.05 filter

Pathways Ids	Pathways	P-value	FDR
hsa01100	Metabolic pathways	9.19E-10	1.19E-06
hsa05204	Chemical carcinogenesis	6.68E-08	8.63E-05
hsa00830	Retinol metabolism	1.39E-07	1.79E-04
hsa04110	Cell cycle	3.38E-07	4.37E-04
hsa04115	p53 signalling pathway	1.69E-06	0.002182762
hsa01130	Biosynthesis of antibiotics	2.94E-06	0.003797444
hsa00980	Metabolism of xenobiotics by cytochrome P450	5.43E-06	0.007016019
hsa03320	PPAR signalling pathway	1.02E-05	0.01318125
hsa00982	Drug metabolism - cytochrome P450	1.20E-05	0.01544763
hsa00260	Glycine, serine and threonine metabolism	1.59E-05	0.020548859
hsa01230	Biosynthesis of amino acids	2.19E-05	0.028277953

### 3.3. STRING PPI Network analysis and Interrelation between pathways

For better understanding the role of DEGs in HCC development, we constructed co-expression protein networks. The insertion of DEGs list into STRING and application of confidence score of > 0.70 established a PPI network contained 1715 edges and 322 nodes, with each node connected with 11.6 other proteins on average. Different topological parameters for PPI network included network density of 0.040, network heterogeneity 1.335, network centralization 0.180, clustering coefficient 0.523, and characteristic path length 4.689. A complete interaction network is shown in Fig. 2A, in which degree and betweenness topological features were calculated to distribute genes into different size circles and colors. The higher the value of these quantitative terms, the greater the importance in the network [38]. Through this PPI network, top clusters were sorted using Cytoscape plugin MCODE with score > 4, nodes > 4. Cluster 1 was the densest interaction network showing MCODE score of 40.53, followed by cluster 2 and cluster 3 with scores 10.72 and 8.0, respectively (Fig. 2B). Subsequently, several crucial hub genes including cyclin-dependent kinase (CDK1), cyclins (CCNA2, CCNB1, CCNB2), serine/threonine-protein kinase (BUB1), NDC80, BUB1B, NCAPG, MAD2L1, and CDC20 were determined by Cytohubba plugin

(Fig. 2B). These selected hub genes were either involved in the cell cycle and its regulation or the main components of kinetochore-microtubule interaction.

Protein clusters were processed through ClueGO plugin of Cytoscape which suggested that cluster genes were mostly associated with nuclear and cellular division during meiosis and mitosis, chromosome reorganization, deoxyribonucleotide biosynthetic processes, and regulation of ubiquitin-protein ligase activity (Fig. 3).

### **3.4. Survival Analysis through GEPIA and Expression Level of Hub Genes**

GEPIA servers provide a platform for integrated information of gene expression from TCGA and GTEx databases regarding multiple cancer types [36]. To evaluate the overall prognostic importance of hub genes in this study, Kaplan-Meier survival analysis was performed to examine the association between different expression levels of genes and survival time of patients with HCC. The logrank  $p$  value was estimated, where  $p$  value smaller than 0.05 indicates a statistically significant difference in mortality between the high-level and low-level groups. These high and low-level groups of patients were separated based on the median level of expression. The high expression level of CCNB2 (logrank  $p = 0.052$ ) and NDC80 (logrank  $p = 0.013$ ) demonstrated poor prognosis, while HCC patients showing high level of BUB1 (logrank  $p = 0.001$ ), CDK1 (logrank  $p = 0.00017$ ), NCAPG (logrank  $p = 0.00097$ ), BUB1B (logrank  $p = 0.0028$ ), CCNB1 (logrank  $p = 0.00015$ ), CDC20 (logrank  $p = 3.8e-06$ ) and MAD2L1 (logrank  $p = 0.0047$ ) had a higher risk of mortality (Fig. 4). Furthermore, GEPIA boxplot representation of hub genes expression in 369 HCC patients and 160 normal/healthy persons described considerable increase in the level for all hub genes (Fig. 5). These hub genes with high expression and low survival rate in patients suggested their association with the pathophysiology of liver cancer to varying extents and could be potential biomarkers for HCC prognosis to monitor the severity of liver cancer or a therapeutic target.

## **4. Discussion**

Over recent years, tumor initiation and progression in HCC have been extensively researched; still, early diagnosis of HCC is truly a big challenge, because the exact mechanism of HCC inducement needs a full understanding. Also, incidence and cancer-specific mortality worldwide is increasing due to limited therapeutic strategies; therefore, there is an urgent need to identify the potential key genes and mechanisms to precisely predict the HCC onset, progression and to develop novel therapeutic agents. Bioinformatics has become increasingly popular to evaluate changes in the gene expression profile during the initiation and progression of diseases by the integrated microarray analysis, investigating the screened hub genes for cancer diagnosis and therapy.

In this study, we identified a total of 598 DGEs comprising of 233 upregulated and 365 downregulated genes between HCC patients in comparison to healthy ones chosen from the GSE62232 expression profile datasets. Using the DAVID software, GO functional, and KEGG pathway enrichment analyses of the DEGs were performed. The results revealed that the identified DEGs were closely related to various

biological processes and pathways such as metabolic, heme binding, drug detoxification, cell cycle, meiosis, and mitosis, etc. We had also screened out ten hub genes, including cell cycle regulatory cyclins and cyclin-dependent proteins CCNA2, CCNB1 and CDK1. Besides, a PPI network was constructed to analyse the interactional relationships between the DGEs, and survival analyses of hub genes were performed using the GEPIA. These hub genes have been extensively researched in recent years.

Recent genetic evidence has revealed interphase cyclin-dependent kinases (CDKs) are essential for the proliferation of tumor cells. CDK1 belongs to the CDK family, a member of the Ser/Thr protein kinases necessary for cell-cycle progression, triggering cell cycle transitions, namely G1/S and G2/M [39]. Clinical implications of deregulated CDK1 are closely related to HCC tumorigenesis. Research has found that overexpression of CDK1 in HCC is significantly negatively correlated with HCC patients' survival. Zhang et al. [40] suggested that miR-582-5p regulated the progression of HCC by directly targeting the CDK1. Elevated levels of CDK1 was also shown to be directly associated with advanced stage portal vein invasion, increased AFP levels, and poor patient survival in HCC [41]. Wu et al. [39] revealed higher levels of CDK1 in HCC patients than healthy individuals, which was in agreement with the present study's findings.

CCNA2, CCNB1, and CCNB2 genes are all members of the highly conserved cyclin family. In this study, high expression of CCNA2 was closely associated with poor prognosis in HCC patients. CCNA2 protein functions as regulators of the cell cycle by activating cyclin-dependent kinases, and CCNA2 expression in the cell cycle was driven mostly by E2Fs [42]. CCNA2 overexpression has been observed in several types of cancers; also, a study has demonstrated that inhibition of CCNA2 led to the suppression of HCC, cell proliferation, and tumorigenesis [43]. The aberrant expression of CCNA2 is related to reduced survival in patients with HCC and breast cancer [42, 44]. Chai et al. [45] revealed that CCNB1 is highly expressed and associated with the unfavorable prognosis for patients with HCC, consistent with our findings. This suggested the plausibility of CCNB1 as a potential therapeutic target for the treatment of HCC. This conclusion is further proved by a study that CCNB1 knockdown by miR-144 inhibited HCC cell migration, invasion, and proliferation [46]. Previous research [47] has shown that CCNB1 inhibits the growth of cells by inducing cell cycle arrest at the G2/M phase suggesting CCNB1 may be an effective anti-cancer agent in future therapy. Furthermore, CCNB2 was found to be overexpressed in several malignant tumors, and high expression of CCNB2 is associated with poor prognosis in HCC and invasive breast carcinoma [48, 49].

In the case of BUB1, several studies have found high expression of the BUB1 in a variety of human tumors. Serine/threonine-protein kinase BUB1 binds centromeres during mitosis [50], and has been involved in apoptosis and cell cycle [51], as well as in reducing the overall survival (OS) rate of HCC patients. BUB1B was reported to be involved in tumor cell cycle regulation, and overexpression of BUB1B is related to the progression and recurrence of HCC [52]. NDC80 is highly conserved; a core component of the kinetochore-microtubule interaction machinery which is identified as a requirement for proper chromosome segregation. NDC80 has also been associated with the HCC progression [53], and NDC80-knockdown in pancreatic cancer inhibited cell cycle and cell proliferation [54]. NCAPG organizes the

coiling topology of individual chromatids, Liu et al. [55] demonstrated NCAPG functioning as an oncogene in the development of HCC.

CDC20 is an essential cell-cycle regulator, which plays an important role in promoting the onset of anaphase and mitotic exit. The increased expression levels of CDC20 have been linked with the development and progression of HCC [56]. Additionally, research has shown that silencing CDC20 and HPSE expression activated cell apoptosis; thus, targeting inhibition of both CDC20 and HPSE expression is an ideal therapeutic option of HCC [57]. As for MAD2L1, Yun et al. [58] demonstrated that MiR-200c-5p inhibits HCC cell proliferation, migration, and invasion by targeted suppression of MAD2L1, suggesting that the high expression levels of MAD2L1 are associated with poor prognosis of patients with HCC. Moreover, MAD2L1 may potentially be used as a prognostic and therapeutic target in HCC patients.

## **5. Conclusion**

In summary, the purpose of the present study purpose was to screen and verify hub genes that may provide new insights into the development, prognosis, and treatment of HCC. In total, 598 DEGs were identified via integrated bioinformatics analysis, of which ten were identified as hub genes that may be used as biomarkers for the diagnosis and prognostic evaluation of HCC. However, because the results of this study were based on data analysis only, further experimental verification via animal experiments and clinical trials are required to confirm these results.

## **Declarations**

### **Funding/acknowledgement**

We are thankful for the support of the Innovation Foundation of Radiation Application, China, as they have funded this project under the project number (KFZC2018040208).

### **Authors contribution**

Mohib Ullah Kakar, Muhammad Zubair Mehboob and Muhammad Akram generate the idea and perform experiments; Imran Iqbal and Hafza Wajeeha Ijaz perform the analysis. Ubair Aziz, Adil Farooq Lodhi and Madiha Rasheed write the original draft, Yulin Deng and Rongji Dai provide the resources and supervise.

### **Conflict of Interest**

There is no conflict of interest among the authors regarding the publication of this manuscript.

### **Consent for publication**

All co-authors are willing to publish the article Scientific Report

### **Ethics approval**

Not applicable

## Availability of supporting data

Not applicable

## References

1. Ghebreyesus, G. J. & Al-Ansary, L. A. LA, World health organization global health observatory statistic:Geneva 2018 2018. 380 (2018) 1–100.  
<https://gco.iarc.fr/today/data/factsheets/populations/170-colombia-fact-sheets.pdf>.
2. Feng, R. M., Zong, Y. N., Cao, S. M. & Xu, R. H. Current cancer situation in China: Good or bad news from the 2018 Global Cancer Statistics? *Cancer Communications*. **39**,  
<https://doi.org/10.1186/s40880-019-0368-6> (2019).
3. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: A Cancer Journal for Clinicians. **68** (2018)394–424.  
<https://doi.org/10.3322/caac.21492>.
4. Gilsanz, V. & García Hepatocellular carcinoma, Anales de La Real Academia Nacional de Medicina. **105** (1988)387–409. <https://doi.org/10.1056/NEJMra1713263>.
5. Marengo, A., Rosso, C. & Bugianesi, E. Liver cancer: Connections with obesity, fatty liver, and cirrhosis. *Annual Review of Medicine*. **67**, 103–117 <https://doi.org/10.1146/annurev-med-090514-013832> (2016).
6. Phukan, R. K. *et al.* Association of processed food, synergistic effect of alcohol and HBV with Hepatocellular Carcinoma in a high incidence region of India. *Cancer Epidemiology*. **53**, 35–41  
<https://doi.org/10.1016/j.canep.2018.01.005> (2018).
7. Forner, A., Reig, M. & Bruix, J. Hepatocellular carcinoma, The Lancet. **391** (2018)1301–1314.  
[https://doi.org/10.1016/S0140-6736\(18\)30010-2](https://doi.org/10.1016/S0140-6736(18)30010-2).
8. Parkin, D. M. The global health burden of infection-associated cancers in the year 2002. *International Journal of Cancer*. **118**, 3030–3044 <https://doi.org/10.1002/ijc.21731> (2006).
9. Rinella, M. E. Nonalcoholic fatty liver disease a systematic review. *JAMA - Journal of the American Medical Association*. **313**, 2263–2273 <https://doi.org/10.1001/jama.2015.5370> (2015).
10. Mirizzi, A. *et al.* Effects of some food components on non-alcoholic fatty liver disease severity: Results from a cross-sectional study. *Nutrients*. **11**, <https://doi.org/10.3390/nu11112744> (2019).
11. Mathers, C. D. & Loncar, D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine*. **3**, 2011–2030 <https://doi.org/10.1371/journal.pmed.0030442> (2006).
12. Jemal, A. *et al.* Annual Report to the Nation on the Status of Cancer, 1975–2014, Featuring Survival, Journal of the National Cancer Institute. **109**(2017). <https://doi.org/10.1093/jnci/djx030>.
13. Taskinen, M., Karjalainen-Lindsberg, M. L., Nyman, H., Eerola, L. M. & Leppä, S. A high tumor-associated macrophage content predicts favorable outcome in follicular lymphoma patients treated

- with rituximab and cyclophosphamide- doxorubicin-vincristine-prednisone. *Clinical Cancer Research*. **13**, 5784–5789 <https://doi.org/10.1158/1078-0432.CCR-07-0778> (2007).
14. Zhou, L., Huang, Y., Li, J. & Wang, Z. The mTOR pathway is associated with the poor prognosis of human hepatocellular carcinoma. *Med. Oncol.* **27**, 255–261 <https://doi.org/10.1007/s12032-009-9201-4> (2010).
  15. Kulik, L. & El-Serag, H. B. Epidemiology and Management of Hepatocellular Carcinoma. *Gastroenterology*. **156**, 477–491 (2019).
  16. Dimitroulis, D. *et al.* From diagnosis to treatment of hepatocellular carcinoma: An epidemic problem for both developed and developing world. *World Journal of Gastroenterology*. **23**, 5282–5294 <https://doi.org/10.3748/wjg.v23.i29.5282> (2017).
  17. Liu, S. *et al.* Analysis of Transcription Factor-Related Regulatory Networks Based on Bioinformatics Analysis and Validation in Hepatocellular Carcinoma, *BioMed Research International* 2018 (2018). <https://doi.org/10.1155/2018/1431396>.
  18. Li, L., Lei, Q., Zhang, S., Kong, L. & Qin, B. Screening and identification of key biomarkers in hepatocellular carcinoma: Evidence from bioinformatic analysis. *Oncol. Rep.* **38**, 2607–2618 <https://doi.org/10.3892/or.2017.5946> (2017).
  19. Zhang, C. *et al.* The identification of key genes and pathways in hepatocellular carcinoma by bioinformatics analysis of high-throughput data. *Med. Oncol.* **34**, <https://doi.org/10.1007/s12032-017-0963-9> (2017).
  20. Song, Z. *et al.* Processing Optimization of Typed Resources with Synchronized Storage and Computation Adaptation in Fog Computing, *Wireless Communications and Mobile Computing* 2018 (2018). <https://doi.org/10.1155/2018/3794175>.
  21. Song, M., Duan, Y., Huang, T. & Zhan, L. Inter-Edge and Cloud conversion accelerated user-generated content for virtual brand community, *Eurasip Journal on Wireless Communications and Networking* 2020 (2020). <https://doi.org/10.1186/s13638-019-1635-6>.
  22. Duan, Y., Lu, Z., Zhou, Z., Sun, X. & Wu, J. Data Privacy Protection for Edge Computing of Smart City in a DIKW Architecture. *Engineering Applications of Artificial Intelligence*. **81**, 323–335 <https://doi.org/10.1016/j.engappai.2019.03.002> (2019).
  23. Totoki, Y. *et al.* Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nature Genetics*. **46**, 1267–1273 <https://doi.org/10.1038/ng.3126> (2014).
  24. Ally, A. *et al.* Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell*. **169**, 1327–1341 (2017).
  25. Castro, L. S. E. P. W. *et al.* Potential anti-angiogenic, antiproliferative, antioxidant, and anticoagulant activity of anionic polysaccharides, fucans, extracted from brown algae *Lobophora variegata*. *Journal of Applied Phycology*. **27**, 1315–1325 <https://doi.org/10.1007/s10811-014-0424-1> (2015).
  26. Zucman-Rossi, J., Villanueva, A., Nault, J. C. & Llovet, J. M. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. *Gastroenterology*. **149**, 1226–12394 <https://doi.org/10.1053/j.gastro.2015.05.061> (2015).

27. He, B. *et al.* Bioinformatics analysis of key genes and pathways for hepatocellular carcinoma transformed from cirrhosis, *Medicine (United States)*. **96**(2017).  
<https://doi.org/10.1097/MD.00000000000006938>.
28. Zhou, L., Du, Y., Kong, L., Zhang, X. & Chen, Q. Identification of molecular target genes and key pathways in hepatocellular carcinoma by bioinformatics analysis. *OncoTargets and Therapy*. **11**, 1861–1869 <https://doi.org/10.2147/OTT.S156737> (2018).
29. Barrett, T. *et al.* NCBI GEO: Archive for functional genomics data sets - Update, *Nucleic Acids Research*. **41**(2013). <https://doi.org/10.1093/nar/gks1193>.
30. Clough, E. & Barrett, T. The Gene Expression Omnibus database, in: *Methods in Molecular Biology*, 2016: pp.93–110. [https://doi.org/10.1007/978-1-4939-3578-9\\_5](https://doi.org/10.1007/978-1-4939-3578-9_5).
31. Schulze, K. *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nature Genetics*. **47**, 505–511 <https://doi.org/10.1038/ng.3252> (2015).
32. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nature Genetics*. **25**, 25–29 <https://doi.org/10.1038/75556> (2000).
33. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res*. **32**, <https://doi.org/10.1093/nar/gkh063> (2004).
34. Szklarczyk, D. *et al.* C. Von Mering, STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. **43**, D447–D452 <https://doi.org/10.1093/nar/gku1003> (2015).
35. Chin, C. H. *et al.* cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*. **8**, <https://doi.org/10.1186/1752-0509-8-S4-S11> (2014).
36. Tang, Z. *et al.* GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. **45**, W98–W102 <https://doi.org/10.1093/nar/gkx247> (2017).
37. Bindea, G. *et al.* ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. **25**, 1091–1093 <https://doi.org/10.1093/bioinformatics/btp101> (2009).
38. Liu, X. *et al.* Identification of potential key genes associated with the pathogenesis and prognosis of gastric cancer based on integrated bioinformatics analysis. *Frontiers in Genetics*. **9**, <https://doi.org/10.3389/fgene.2018.00265> (2018).
39. Wu, C. X. *et al.* Blocking CDK1/PDK1/ $\beta$ -Catenin signaling by CDK1 inhibitor R03306 increased the efficacy of sorafenib treatment by targeting cancer stem cells in a preclinical model of hepatocellular carcinoma. *Theranostics*. **8**, 3737–3750 <https://doi.org/10.7150/thno.25487> (2018).
40. Zhang, Y. *et al.* miR-582-5p inhibits proliferation of hepatocellular carcinoma by targeting CDK1 and AKT3, *Tumor Biology*. **36** (2015)8309–8316. <https://doi.org/10.1007/s13277-015-3582-0>.
41. Ito, Y. *et al.* Expression and prognostic role of cyclin-dependent kinase 1 (cdc2) in hepatocellular carcinoma. *Oncology*. **59**, 68–74 <https://doi.org/10.1159/000012140> (2000).

42. Gopinathan, L. *et al.* Loss of Cdk2 and cyclin A2 impairs cell proliferation and tumorigenesis. *Cancer Research*. **74**, 3870–3879 <https://doi.org/10.1158/0008-5472.CAN-13-3440> (2014).
43. Yang, F. *et al.* Waltonitone inhibits proliferation of hepatoma cells and tumorigenesis via FXR-miR-22-CCNA2 signaling pathway. *Oncotarget*. **7**, 75165–75175 <https://doi.org/10.18632/oncotarget.12614> (2016).
44. Bukholm, I. R. K., Bukholm, G. & Nesland, J. M. Over-expression of cyclin a is highly associated with early relapse and reduced survival in patients with primary breast carcinomas. *International Journal of Cancer*. **93**, 283–287 <https://doi.org/10.1002/ijc.1311> (2001).
45. Chai, N. *et al.* FOXM1 promotes proliferation in human hepatocellular carcinoma cells by transcriptional activation of CCNB1, *Biochemical and Biophysical Research Communications*. **500**(2018) 924–929. <https://doi.org/10.1016/j.bbrc.2018.04.201>.
46. Gu, J., Liu, X., Li, J. & He, Y. MicroRNA-144 inhibits cell proliferation, migration and invasion in human hepatocellular carcinoma by targeting CCNB1 *11 Medical and Health Sciences 1112 Oncology and Carcinogenesis. Cancer Cell Int*. **19**, <https://doi.org/10.1186/s12935-019-0729-x> (2019).
47. Wang, G. *et al.* Methyl protodioscin induces G2/M cell cycle arrest and apoptosis in HepG2 liver cancer cells. *Cancer Lett*. **241**, 102–109 <https://doi.org/10.1016/j.canlet.2005.10.050> (2006).
48. Li, R. *et al.* Cyclin B2 Overexpression in Human Hepatocellular Carcinoma is Associated with Poor Prognosis, *Archives of Medical Research*. **50** (2019)10–17. <https://doi.org/10.1016/j.arcmed.2019.03.003>.
49. Shubbar, E. *et al.* Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome. *BMC Cancer*. **13**, <https://doi.org/10.1186/1471-2407-13-1> (2013).
50. Xu, B. *et al.* MiR-490-5p Suppresses Cell Proliferation and Invasion by Targeting BUB1 in Hepatocellular Carcinoma Cells. *Pharmacology*. **100**, 269–282 <https://doi.org/10.1159/000477667> (2017).
51. Ricke, R. M., Jeganathan, K. B. & van Deursen, J. M. Bub1 overexpression induces aneuploidy and tumor formation through Aurora B kinase hyperactivation. *Journal of Cell Biology*. **193**, 1049–1064 <https://doi.org/10.1083/jcb.201012035> (2011).
52. Liu, A. W. *et al.* The clinicopathological significance of BUBR1 overexpression in hepatocellular carcinoma. *Journal of Clinical Pathology*. **62**, 1003–1008 <https://doi.org/10.1136/jcp.2009.066944> (2009).
53. Ju, L. L. *et al.* Effect of NDC80 in human hepatocellular carcinoma. *World Journal of Gastroenterology*. **23**, 3675–3683 <https://doi.org/10.3748/wjg.v23.i20.3675> (2017).
54. Meng, Q. C. *et al.* Overexpression of NDC80 is correlated with prognosis of pancreatic cancer and regulates cell proliferation. *American Journal of Cancer Research*. **5**, 1730–1740 (2015).
55. Liu, K. *et al.* Silencing non-smc chromosome-associated polypeptide g inhibits proliferation and induces apoptosis in hepatocellular carcinoma cells. *Canadian Journal of Physiology and Pharmacology*. **96**, 1246–1254 <https://doi.org/10.1139/cjpp-2018-0195> (2018).

56. Li, J., Gao, J. Z., Du, J. L., Huang, Z. X. & Wei, L. X. Increased CDC20 expression is associated with development and progression of hepatocellular carcinoma. *International Journal of Oncology*. **45**, 1547–1555 <https://doi.org/10.3892/ijo.2014.2559> (2014).
57. Liu, M. *et al.* Evaluation of the Antitumor Efficacy of RNAi-Mediated Inhibition of CDC20 and Heparanase in an Orthotopic Liver Tumor Model. *Cancer Biotherapy and Radiopharmaceuticals*. **30**, 233–239 <https://doi.org/10.1089/cbr.2014.1799> (2015).
58. Li, Y., Bai, W. & Zhang, J. MiR-200c-5p suppresses proliferation and metastasis of human hepatocellular carcinoma (HCC) via suppressing MAD2L1, *Biomedicine and Pharmacotherapy*. **92** (2017)1038–1044. <https://doi.org/10.1016/j.biopha.2017.05.092>.

## Figures

### DEGs genes of hepatocellular carcinoma

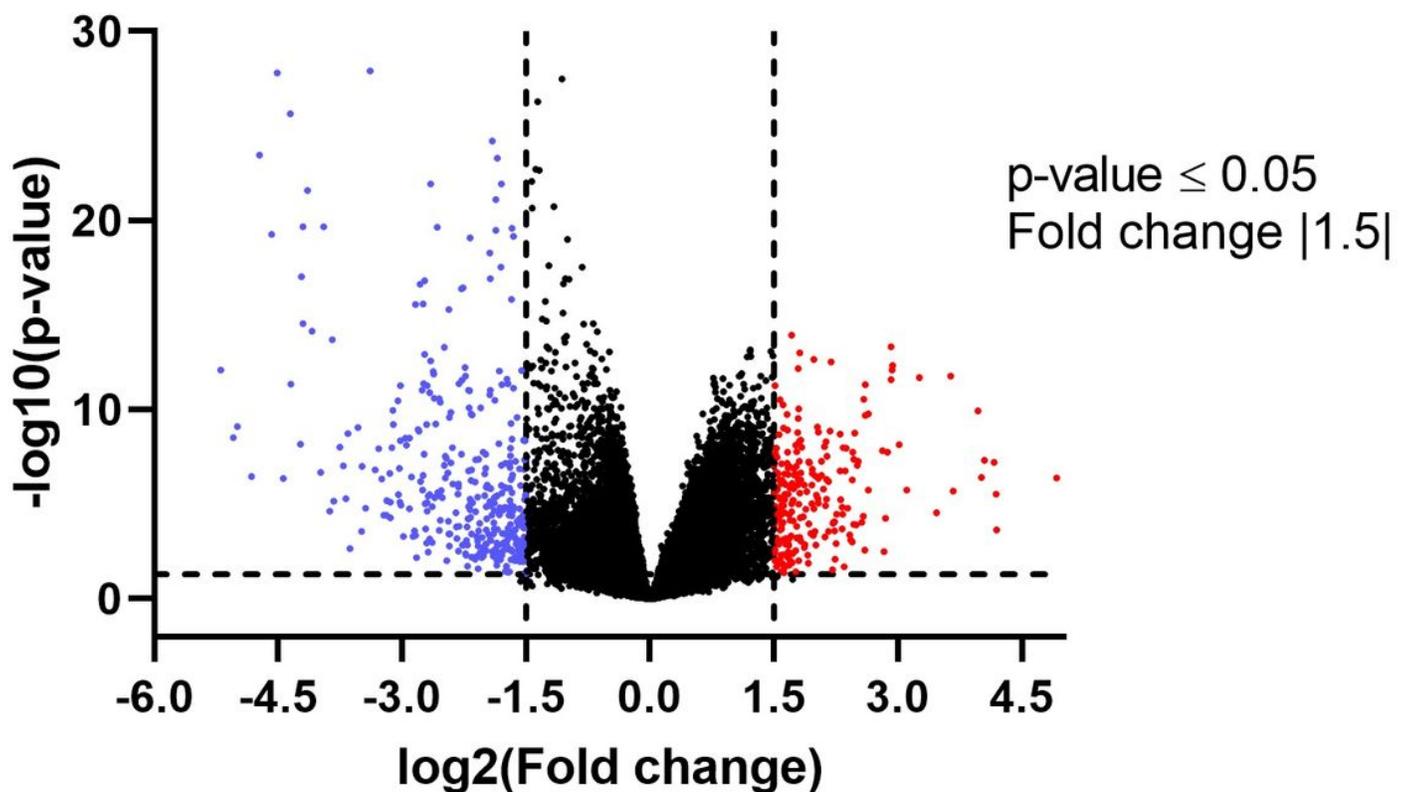
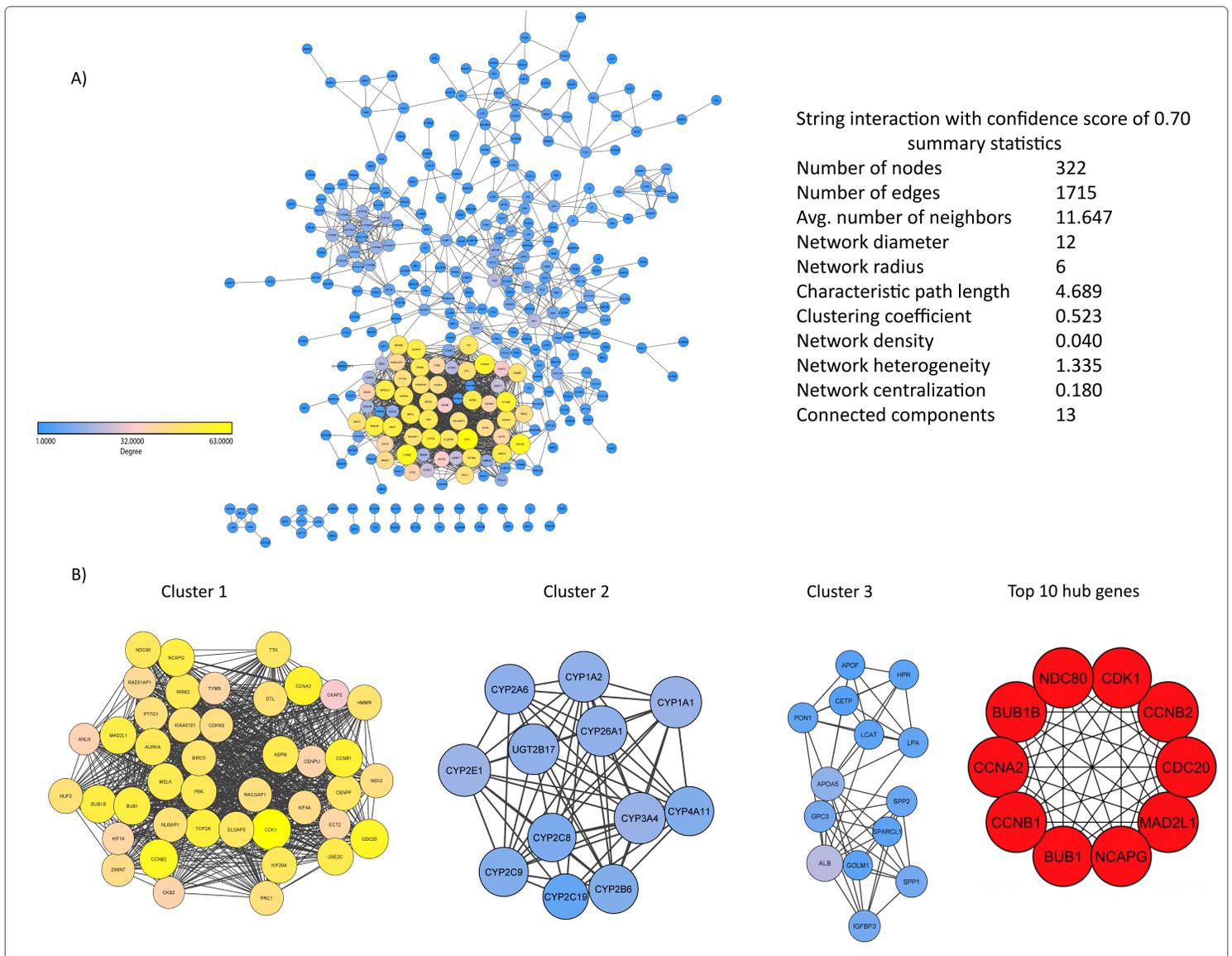


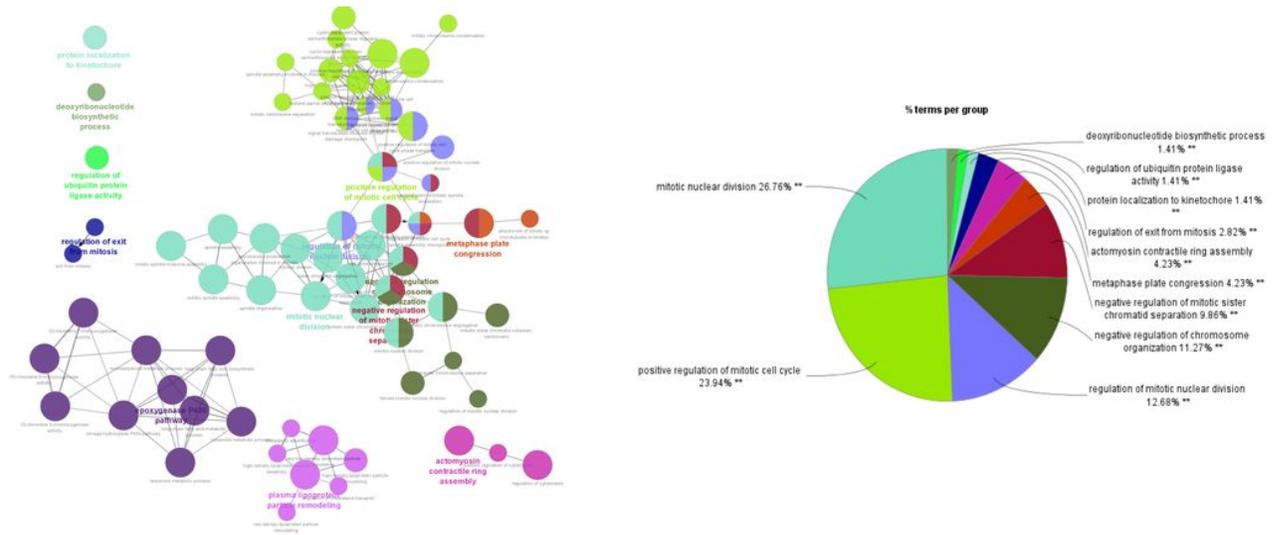
Figure 1

Volcano plot of the DEGs in HCC patients compared with control from the GSE62232 dataset. Each black dot represents one gene. The black and colorful dots above and side of the dotted line corresponds to those genes having p value  $\leq 0.05$  and fold change  $|1.5|$ . Blue color dots are downregulated, and red dots are upregulated genes.



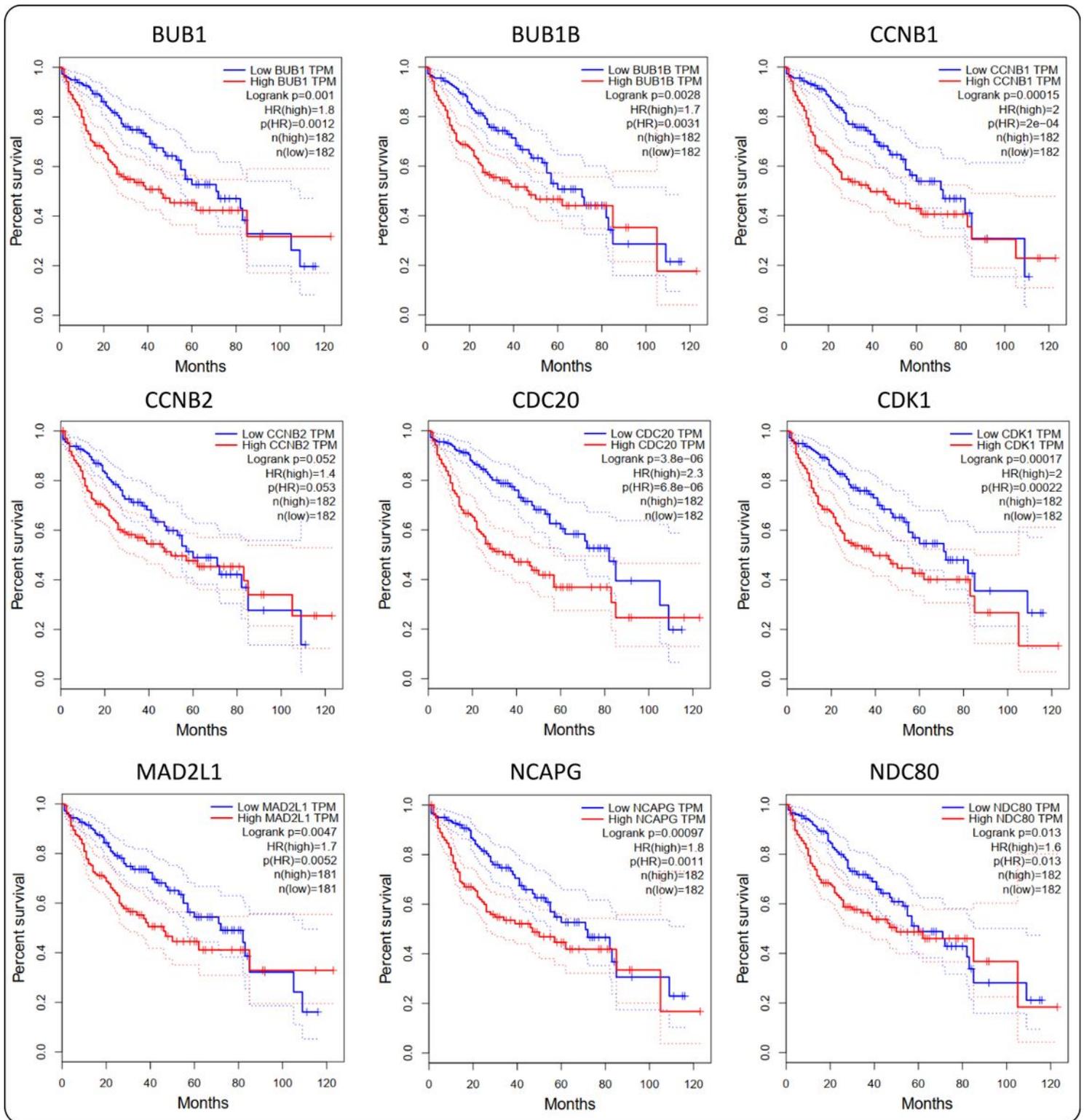
**Figure 2**

A) Using the online STRING tool, a PPI network was developed which was further visualized by Cytoscape software. The size and color map nodes are determined by the degree value, which renders a gradual setting in small size with low degree in blue, large size with a high degree in yellow. B) Top clusters determined by MCODE with score > 4 and node > 4. The top 10 genes derived from the MMC method were chosen using the CytoHubba plugin.



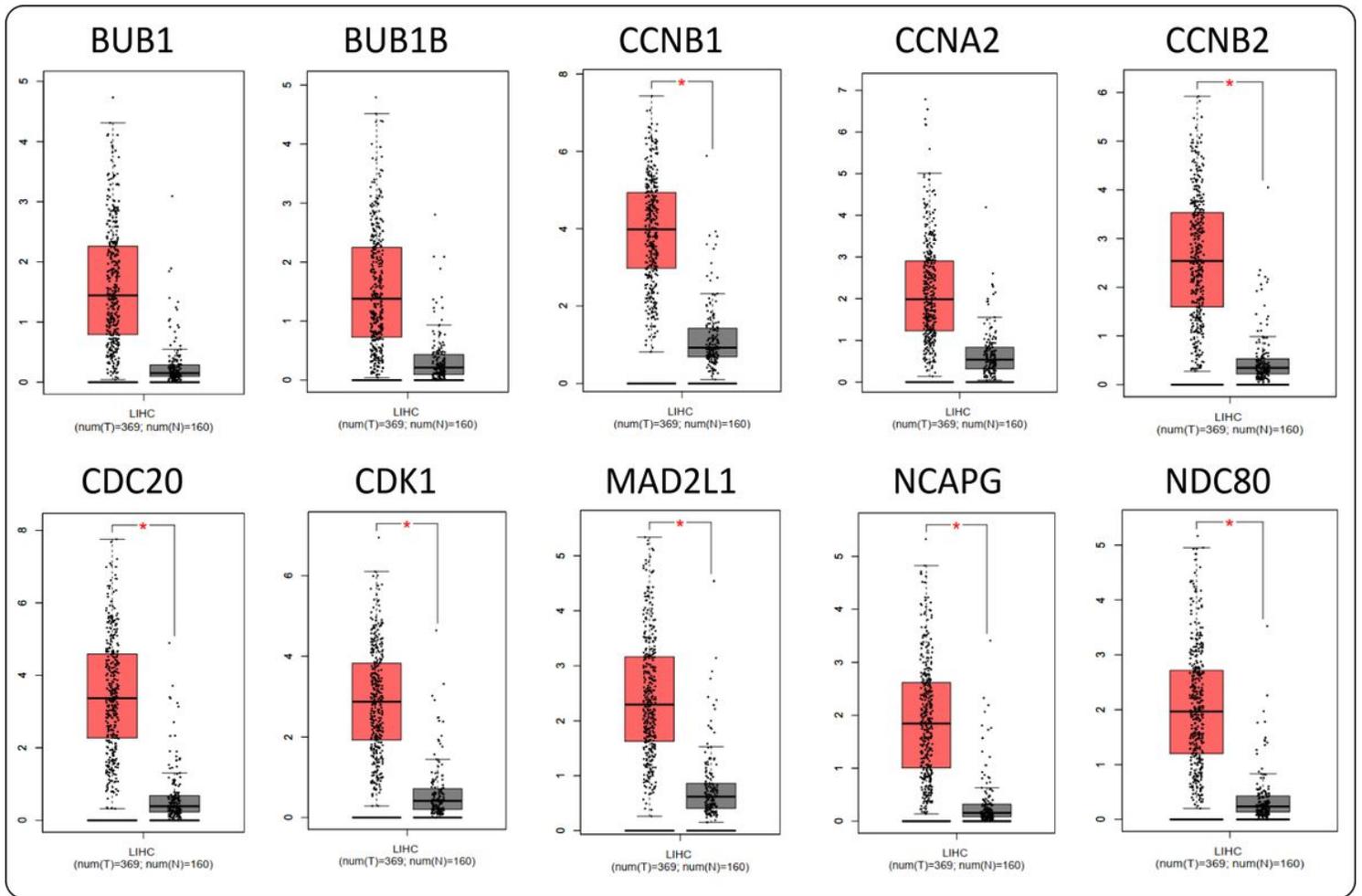
**Figure 3**

GO analysis of 10 hub genes mapped into 11 different biological processes. The group information and percentage are represented by different colors. Major portion of percent was mapped in mitotic nuclear division, positive regulation of mitotic cell cycle, regulation of mitotic nuclear division, negative regulation of chromosome organization, and negative regulation of mitotic sister chromatid separation.



**Figure 4**

Kaplan-Meier survival analysis curve of the hub genes expressed in the HCC. The survival curves were plotted using the GEPIA1 web server. The gene candidates with high expression in the cohorts are shown in red, and the blue line indicates the low-expression cohort, the survival curve is represented in a dotted line, whereas the solid line is the 95% confidence interval. The logrank p value represents the overall significance of analysis, and HR stands for hazard ration; patient number ( $n$ ) = 182.



**Figure 5**

Based on TCGA and GTEx data in GEPIA, the expression level of hub genes was validated in 369 LIHC (liver hepatocellular carcinoma) patients and 160 healthy controls. Interestingly, all hub genes were remarkably overexpressed in LIHC patients.

## Supplementary Files

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

- [supplementaryfile1.xlsx](#)
- [supplementaryfile2.xlsx](#)