CCNB2, CDC20, AURKA, TOP2A, MELK, NCAPG, KIF20A, UBE2C, PRC1, and ASPM May Be Potential Therapeutic Targets for Hepatocellular Carcinoma Using Integrated Bioinformatic Analysis

Zhiqiang Yang
Xinglang Wu
Junbo Li
Qiang Zheng
Junwei Niu
Shengwei Li

Department of Hepatobiliary Surgery,
Second Affiliated Hospital of Chongqing Medical University, Chongqing, People’s Republic of China

Background: Hepatocellular carcinoma (HCC) is a highly malignant, recurrent and drug-resistant tumor, and patients often lose the opportunity for surgery when they are diagnosed. Abnormal gene expression is closely related to the occurrence of HCC. The aim of the present study was to identify the differentially expressed genes (DEGs) between tumor tissue and non-tumor tissue of HCC samples in order to investigate the mechanisms of liver cancer.

Methods: The gene expression profile (GSE62232, GSE89377, and GSE112790) was downloaded from the Gene Expression Omnibus (GEO) and analyzed using the online tool GEO2R to identify differentially expressed genes (DEGs). Gene Ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed using the Database for Annotation, Visualization and Integrated Discovery. Protein–protein interaction (PPI) of these DEGs was analyzed based on the Search Tool for the Retrieval of Interacting Genes database and visualized by Cytoscape software. In addition, we used the online Kaplan–Meier plotter survival analysis tool to evaluate the prognostic value of hub genes expression. HPA database was used to reveal the differences in protein level of hub genes.

Results: A total of 50 upregulated DEGs and 122 downregulated DEGs were identified. Among them, ten hub genes with a high degree of connectivity were picked out. Overexpression of these hub genes was associated with unfavorable prognosis of HCC.

Conclusion: Our study suggests that CCNB2, CDC20, AURKA, TOP2A, MELK, NCAPG, KIF20A, UBE2C, PRC1, and ASPM were overexpressed in HCC compared with normal liver tissue. Overexpression of these genes was an unfavorable prognostic factor of HCC patients. Further study is needed to explore the value of them in the diagnosis and treatment of HCC.

Keywords: hepatocellular carcinoma, hub genes, expression profiling data, protein–protein interaction network

Introduction

Hepatocellular carcinoma (HCC) ranks sixth among the most common cancers, and its incidence is increasing in almost all countries. Hepatocellular carcinoma accounts for more than 90% of all liver cancers, and its risk factors include hepatitis B virus (HBV) and hepatitis C virus (HCV) infection, alcohol consumption, aflatoxin exposure and metabolic abnormalities. Alpha fetoprotein (AFP) combined with imaging is the most widely used method for screening and diagnosis of HCC.
liver cancer currently. Because of the low sensitivity of serum AFP, the diagnostic sensitivity of all HCC patients is only 59%. Due to the lack of diagnostic accuracy, the American Association for the Study of Liver Diseases and the European Association for Liver Research do not recommend AFP for the diagnosis of liver cancer.\(^3\),\(^4\)

There are many treatment methods for liver cancer, including surgical resection, chemotherapy, TACE, and liver transplantation.\(^5\) However, the rate of recurrence, progression, and metastasis of HCC is still high. Previous studies have found that abnormal gene expression is closely related to the biological aggressiveness of HCC and patient survival. These genes provide new drug targets for relapsed or unresectable HCC, improving the therapeutic effect greatly.\(^6\),\(^7\) Therefore, screening of new prognostic biomarkers and therapeutic targets for HCC patients is an important research topic.

In order to detect the differentially expressed genes (DEGs) between HCC and healthy human liver tissues, we downloaded three gene expression profiles from the GEO database to obtain DEGs between HCC tissues and normal tissue samples. Through gene ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment, the selected DEGs were analyzed for their potential functions. Then, we established and used Cytoscape software to construct a protein–protein interaction (PPI) network composed of these predicted target genes to find hub genes related to HCC. The online database Kaplan–Meier plotter was used for survival analysis of these hub genes. The hub gene immunohistochemical staining of liver cancer tissues and normal tissues was obtained from the HPA database. The purpose of this study is to explore and identify the hub gene and its potential molecular mechanisms related to invasion and metastasis based on a comprehensive bioinformatics analysis.

### Materials and Methods

#### Data Source

In this study, the gene expression profile datasets were obtained from the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) of NCBI. A total of 2241 series about human hepatocellular carcinoma were retrieved from the database. After a careful review, three gene expression profiles (GSE62232, GSE89377, and GSE112790) were selected. Among them, GSE89377 was based on the platform GPL6947 Illuma HumanHT-12 V3.0 expression beadchip, and GSE62232 and GSE112790 were based on platform GPL570[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. All the data are freely available online, and this study does not involve any experiments conducted by any author on humans or animals. Ethical approval was not necessary for this study because public datasets were analyzed.

#### Data Processing of DEGs

GEO2R (www.ncbi.nlm.nih.gov/geo/geo2r/), a web tool performing comparisons based on limma and GEOquery R packages of the Bioconductor project, was used to identify the DEGs between HCC and normal specimens. Genes that met the cutoff criteria, adjusted P<0.05 and |logFC|>1.0, were considered as DEGs. Statistical analysis was carried out for each dataset, and the Venn diagram web tool (bioinformatics.psb.ugent.be/webtools/Venn/) was performed to identify the common DEGs shared by these three GSEs.

#### GO and KEGG Pathway Analysis of DEGs

GO is a common method for annotating genes, gene products and sequences to underlying biological phenomena; gene functions can be classified into biological process (BP), molecular function (MF), and cellular component (CC). KEGG is an integrated database resource for biological interpretation of genome sequences and other high-throughput data. Both analyses were available in the DAVID database (https://david.ncifcrf.gov/), which is a bioinformatics data resource composed of an integrated biology knowledge base and analysis tools to extract meaningful biological information from large quantities of genes and protein collections. P < 0.01 and gene counts >5 were considered statistically significant.

#### PPI Network Construction and Hub Gene Identification

The Search Tool for the Retrieval of Interacting Genes (STRING; www.string-db.org/) was used to evaluate the PPI information. The PPI network served to identify the key genes and Cytoscape software (www.cytoscape.org) was used to draw the network diagram. The topology of the PPI network was analyzed and the extent of the expression of each gene was calculated. Nodes with higher degree of connectivity tend to be more essential in maintaining the stability of the entire network. CytoHubba was used to calculate the degree of each protein node.
Survival Analysis of Hub Genes
The Kaplan–Meier plotter (http://kmplot.com/analysis/) is capable to assess the effect of 54k genes (mRNA, miRNA, protein) on survival in 21 cancer types including breast (n = 6234), ovarian (n = 2190), lung (n = 3452), and gastric (n = 1440) cancer. Sources for the databases include GEO, EGA, and TCGA. The Kaplan–Meier plotter mRNA liver cancer database was applied to evaluate the prognostic values of hub genes in hepatocellular carcinoma patients. Probes of genes were selected based on the “only JetSet best probe set”, and the probe ID required for each gene is shown in Table 1. P <0.01 was considered as a statistically significant result.

Evaluation of Immunohistochemical Staining
To verify the protein expression level of candidate genes in HCC tissues, we used Human Protein Atlas (HPA, https://www.proteinatlas.org/) database to obtain immunohistochemical staining. The flow diagram of bioinformatics analysis is shown in Figure 1.

Results
Identification of DEGs
A total of 3 datasets were retrieved from GEO repository and further subjected for the data analysis according to the inclusion criteria, including GSE62232, GSE89377 and GSE112790. Among them, GSE62232 contained 81 HCC samples and 10 normal samples, GSE89377 contained 40 HCC samples and 13 normal samples, and GSE112790 included 183 HCC specimens and 15 normal liver specimens (Table 2). Based on the criteria of P < 0.05 and |logFC|>1, a total of 1033 DEGs were identified from GSE62232, including 608 upregulated genes and 425 downregulated genes. In gene chip GSE89377, 391 DEGs were identified; 109 genes were upregulated, and 282 genes were downregulated. And from GSE112790, 1337 DEGs including 813 upregulated genes and 560 downregulated genes were identified. All DEGs were identified by comparing HCC samples with normal liver samples. The overlap among the 3 datasets as shown in the Venn diagram (Figure 2). One hundred and seventy-two DEGs were significantly differentially expressed among all three groups, of which 50 were significantly upregulated genes and 122 were downregulated.

Functional Enrichment Analyses of DEGs
To analyze the biological classification of DEGs, functional and pathway enrichment analyses were performed using DAVID (Table 3). The results of GO analysis indicated that DEGs were mainly enriched in cellular component, including extracellular region part, membrane fraction, extracellular region, and extracellular space. BP analysis showed that the DEGs were significantly enriched in oxidation reduction, cell cycle, DNA metabolic process, and cell cycle process. For the molecular function, the DEGs were enriched in electron carrier activity, carbohydrate binding, cofactor binding, and lipid binding. In addition, the results of KEGG pathway analysis showed that DEGs were mainly enriched in cell cycle, DNA replication, fatty acid metabolism, and retinol metabolism.

PPI Network Construction and Hub Gene Identification
The PPI network of DEGs was constructed (Figure 3) and the most significant module was obtained using Cytoscape (Figure 4). A total of 45 nodes were involved in the PPI network. The top ten genes evaluated by connectivity degree in the PPI network were identified (Table 1). The results showed that cyclin B2 (CCNB2) was the most outstanding gene with connectivity degree = 30, followed by cell division cycle 20 (CDC20; degree = 27), aurora kinase A (AURKA; degree = 27), topoisomerase (DNA) II alpha (TOP2A; degree = 27), maternal embryonic leucine zipper kinase (MELK; degree = 26), non-SMC condensin I complex subunit G (NCAPG; degree = 26), kinesin family member 20A (KIF20A; degree = 25), ubiquitin conjugating enzyme E2 C (UBE2C; degree = 25), protein regulator of cytokinesis 1 (PRC1; degree = 25), and...
abnormal spindle microtubule assembly (ASPM; degree = 25). All of these hub genes were upregulated in HCC.

Survival Analysis of Ten Hub Genes
The prognostic information of the 10 hub genes was freely available in Kaplan Meier-plotter. A total of 364 liver cancer patients were available for the analysis of overall survival (OS) and relapse-free survival (RFS). It was found that high expression of these hub genes was associated with unfavorable OS and RFS for HCC patients (Figures 5 and 6).

Differences of Nine Hub Genes in Protein Level Between HCC and Normal Tissues
HPA database was used to exhibit the differences in protein level of CCNB2, CDC20, AURKA, TOP2A, MELK, NCAPG, KIF20A, UBE2C, and PRC1. The results showed

Table 2 Statistics of the Three Microarray Databases Derived from the GEO Database

<table>
<thead>
<tr>
<th>GEO Accession</th>
<th>Contributor</th>
<th>Region</th>
<th>Organism</th>
<th>HCC</th>
<th>Normal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE62232</td>
<td>Zucman Rossi J</td>
<td>France</td>
<td>Homo sapiens</td>
<td>81</td>
<td>10</td>
<td>91</td>
</tr>
<tr>
<td>GSE89377</td>
<td>Eun J</td>
<td>South Korea</td>
<td>Homo sapiens</td>
<td>40</td>
<td>13</td>
<td>53</td>
</tr>
<tr>
<td>GSE112790</td>
<td>Shimada S</td>
<td>Japan</td>
<td>Homo sapiens</td>
<td>183</td>
<td>15</td>
<td>198</td>
</tr>
</tbody>
</table>

Abbreviations: GEO, Gene Expression Omnibus; HCC, hepatocellular carcinoma.
the immunohistochemical staining of CCNB2, CDC20, AURKA, TOP2A, MELK, NCAPG, KIF20A, UBE2C, and PRC1 was negative staining in normal tissues and positive in HCC tissues, demonstrating that these genes were significantly expressed in HCC tissues than in normal liver tissues (Figure 7).

**Discussion**

Hepatocellular carcinoma (HCC) is a common malignant tumor of the digestive system, and its incidence has been on the rise in recent years. The main causes of liver cancer include liver cirrhosis, chronic hepatitis virus infection, alcoholic liver disease and aflatoxin. The most common cause is chronic hepatitis B virus (HBV) infection. The analysis of HCC gene expression profile through bioinformatics helps us better understand the molecular mechanism of liver cancer formation and find new targets for its diagnosis and treatment. In this study, we obtained DEGs between liver cancer tissues and non-cancerous tissues. A total of 172 DEGs genes were identified,

Table 3 Significantly Enriched GO Terms and KEGG Pathways of DEGs

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>Description</th>
<th>Count</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOTERM_BP</td>
<td>GO:0055114</td>
<td>Oxidation reduction</td>
<td>19</td>
<td>5.85E-05</td>
</tr>
<tr>
<td></td>
<td>GO:0007049</td>
<td>Cell cycle</td>
<td>21</td>
<td>7.60E-05</td>
</tr>
<tr>
<td></td>
<td>GO:0006259</td>
<td>DNA metabolic process</td>
<td>15</td>
<td>1.55E-04</td>
</tr>
<tr>
<td></td>
<td>GO:0022402</td>
<td>Cell cycle process</td>
<td>16</td>
<td>1.04E-04</td>
</tr>
<tr>
<td></td>
<td>GO:0044421</td>
<td>Extracellular region part</td>
<td>28</td>
<td>2.36E-04</td>
</tr>
<tr>
<td></td>
<td>GO:0005624</td>
<td>Membrane fraction</td>
<td>17</td>
<td>4.73E-04</td>
</tr>
<tr>
<td></td>
<td>GO:0005576</td>
<td>Extracellular region</td>
<td>36</td>
<td>4.55E-04</td>
</tr>
<tr>
<td></td>
<td>GO:0005615</td>
<td>Extracellular space</td>
<td>22</td>
<td>3.40E-04</td>
</tr>
<tr>
<td></td>
<td>GO:0009055</td>
<td>Electron carrier activity</td>
<td>13</td>
<td>8.68E-07</td>
</tr>
<tr>
<td></td>
<td>GO:0030246</td>
<td>Carbohydrate binding</td>
<td>13</td>
<td>4.65E-06</td>
</tr>
<tr>
<td></td>
<td>GO:0048037</td>
<td>Cofactor binding</td>
<td>10</td>
<td>2.75E-05</td>
</tr>
<tr>
<td></td>
<td>GO:0008289</td>
<td>Lipid binding</td>
<td>13</td>
<td>3.59E-05</td>
</tr>
<tr>
<td></td>
<td>hsa04110</td>
<td>Cell cycle</td>
<td>9</td>
<td>3.41E-04</td>
</tr>
<tr>
<td></td>
<td>hsa03030</td>
<td>DNA replication</td>
<td>5</td>
<td>0.001545772</td>
</tr>
<tr>
<td></td>
<td>hsa00071</td>
<td>Fatty acid metabolism</td>
<td>5</td>
<td>0.0022982</td>
</tr>
<tr>
<td></td>
<td>hsa00830</td>
<td>Retinol metabolism</td>
<td>5</td>
<td>0.006854236</td>
</tr>
</tbody>
</table>

Abbreviations: DEG, differentially expressed gene; GO, Gene Ontology; BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes.
including 122 down-regulated genes and 50 up-regulated genes. Through GO and KEGG enrichment analysis, the interaction between DEGs was explored. Differential genes are mainly enriched in Oxidation reduction, Cell cycle, DNA metabolic process and Fatty acid metabolism. In order to understand the interaction between genes, 10 hub genes were screened through PPI network and module analysis, including CCNB2, CDC20, AURKA, TOP2A, MELK, NCAPG, KIF20A, UBE2C, PRC1, and ASPM.

CCNB2 is synthesized in the G1 phase, and then rapidly decreases in the early, middle and late stages. Abnormal expression of CCNB2 will cause the G2/M checkpoint function to fail in the cell cycle. The function of G2/M checkpoint is to repair damaged DNA and maintain genome integrity. Defects in DNA damage checkpoints can lead to gene mutations and even tumors. In addition, CCNB2 overexpression is associated with bladder cancer, lung cancer and breast cancer.\textsuperscript{12} CCNB2 can inhibit cell proliferation and migration, promote cell apoptosis, and cause S phase arrest of liver cancer cells. Promotes cell proliferation and migration through the CCNB2/PLK1 pathway.\textsuperscript{12} Cell division cycle 20 (CDC20), as one of the regulatory factors of cell cycle checkpoints, is initiated in the late cell cycle and exited in the late mitosis. CDC20 directly binds to APC during mitosis and G1 phase and activates its cyclin ubiquitination activity.\textsuperscript{13} The expression of CDC20 can be significantly inhibited by p53 protein, which inhibits malignant transformation by regulating cell cycle, cell senescence and apoptosis-related genes. AURKA is up-regulated in HCC tissues and is related to pathological staging and distant metastasis. The overexpression of AURKA induces epithelial–
mesenchymal transition (EMT) and cancer stem cell (CSC) behavior through PI3K/AKT pathway, while silencing AURKA inhibits the invasiveness of liver cancer cells. AURKA is involved in cancer metastasis after HCC radiotherapy.  

The TOP2A gene encodes a 170 kDa ribozyme that controls the topological structure of DNA, and is often co-amplified with the HER2 gene in breast and bladder cancer. The expression of TOP2A is related to the advanced histological grade of liver cancer, microvascular invasion and early onset of malignant tumors. The expression of TOP2A has been reported as a predictor of tumor progression, recurrence and poor survival rates in small cell lung cancer, rectal cancer, nasopharyngeal cancer and prostate cancer. MELK inhibits the growth and invasion of HCC cells by inducing apoptosis and mitosis. The overexpression of MELK in liver cancer tissues is closely related to the cell cycle and mitosis-related genes, which are directly regulated as part of the forkhead transcription factor and FoxM1-related cell division program. NCAPG is a mitotic gene required for HCC cell proliferation and migration. NCAPG gene knockout can induce mitosis of HCC cells and inhibit cell growth, proliferation and migration. Tetracycline-induced shRNA knockout NCAPG inhibits the growth of

**Figure 5** Kaplan–Meier overall survival analyses for the top ten hub genes expressed in hepatocellular carcinoma patients.

**Figure 6** Kaplan–Meier relapse-free survival analyses for the top ten hub genes expressed in hepatocellular carcinoma patients.
liver cancer cells. In addition, the overexpression of NCAPG in clinical HCC specimens is related to the recurrence and survival of patients. The overexpression of NCAPG is closely related to the overexpression of CCNB1 (mitotic-specific cyclin B1), which is a regulatory protein involved in mitosis.\textsuperscript{20,21} Kinesin family member 20A (KIF20A) can directly interact with Rab6 small GTPase and participate in the dynamic process of the Golgi apparatus. It is also an important mitotic kinesin required for cytokinesis. KIF20A accumulates in the nucleus during the G2 phase of the cell cycle and promotes normal and pathological hepatocyte proliferation.\textsuperscript{22} KIF20A is the downstream target of Gli2, which is an important transcription factor that regulates the proliferation and growth of human liver cancer cells.\textsuperscript{23} Down-regulation of UBE2C expression has a significant inhibitory effect on the proliferation, migration and invasion of liver cancer cells. In addition, UBE2C can protect cancer cells from autophagy death.\textsuperscript{24} Hepatoma cells with UBE2C gene knockout are more sensitive to chemotherapy drugs such as adriamycin (ADR) and 5-fluorouracil (5-FU). Silencing of UBE2C also increased the sensitivity of liver cancer cells to sorafenib.\textsuperscript{25} The expression and distribution of PRC1 are dynamically regulated by Wnt3a signal.\textsuperscript{26,27} The knockout of PRC1 gene damages the transcriptional activity of transcription factor (TCF), reduces the expression of Wnt target genes, and reduces nuclear β-catenin levels. PRC1 interacts with the β-catenin destruction complex, regulates the membrane isolation of the destruction complex induced by Wnt3a, and

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{representativeproteinexpression.png}
\caption{Representative protein expression of the nine hub genes.}
\end{figure}
promotes the release of β-catenin from the APC complex. In vivo, the high expression of PRC1 is related to the targeted expression of nuclear β-catenin and Wnt. 28 PRC1 plays a carcinogenic effect by promoting tumor proliferation, metastasis and tumor occurrence. Overexpression of ASPM alleviates G2/M cell cycle arrest. NS5A protein inhibits the promoter activity of Aspm gene in a dose-dependent manner. NS5A protein down-regulates the expression of the mitotic spindle protein ASPM and induces abnormal mitotic cell cycles, leading to chromosomal instability and tumorigenesis.29,30

Conclusion
Our bioinformatics analysis identified 172 DEGs between HCC and normal liver tissues based on the gene expression datasets obtained from the GEO database. Among them, ten hub genes might be the core genes of liver cancer, including CCNB2, CDC20, AURKA, TOP2A, MELK, NCAPG, KIF20A, UBE2C, PRC1, and ASPM. All of them were upregulated in liver cancer, and overexpression of these genes was associated with unfavorable clinical outcome in liver cancer patients. Further study is needed to confirm the results of our research.

Ethics Statement
This study was approved by the Ethics committee of Second Affiliated Hospital of Chongqing Medical University. The research was conducted in accordance with the “Declaration of Helsinki” (revised in 2013).

Disclosure
The authors declare that they have no conflicts of interest in this work.

References


