Research Article

Based on Integrated Bioinformatics Analysis Identification of Biomarkers in Hepatocellular Carcinoma Patients from Different Regions

Linxin Teng, Kaiyuan Wang, Yu Liu, Yanxia Ma, Weiping Chen, and Lei Bi

School of Preclinical Medicine, Nanjing University of Chinese Medicine, 138 Xianlin Road, Nanjing, Jiangsu 210023, China

Correspondence should be addressed to Lei Bi; bella_nj@163.com

Received 24 July 2019; Accepted 24 September 2019; Published 28 October 2019

Guest Editor: Bin Duan

Copyright © 2019 Linxin Teng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accumulating statistics have shown that liver cancer causes the second highest mortality rate of cancer-related deaths worldwide, of which 80% is hepatocellular carcinoma (HCC). Given the underlying molecular mechanism of HCC pathology is not fully understood yet, identification of reliable predictive biomarkers is more applicable to improve patients’ outcomes. The results of principal component analysis (PCA) showed that the grouped data from 1557 samples in Gene Expression Omnibus (GEO) came from different populations, and the mean tumor purity of tumor tissues was 0.765 through the estimate package in R software. After integrating the differentially expressed genes (DEGs), we finally got 266 genes. Then, the protein-protein interaction (PPI) network was established based on these DEGs, which contained 240 nodes and 1747 edges. FOXM1 was the core gene in module 1 and highly associated with FOXM1 transcription factor network pathway, while FTCD was the core gene in module 2 and was enriched in the metabolism of amino acids and derivatives. The expression levels of hub genes were in line with the Cancer Genome Atlas (TCGA) database. Meanwhile, there were certain correlations among the top ten genes in the up- and downregulated DEGs. Finally, Kaplan–Meier curves and receiver operating characteristic (ROC) curves were plotted for the top five genes in PPI. Apart from CDKN3, the others were closely concerned with overall survival. In this study, we detected the potential biomarkers and their involved biological processes, which would provide a new train of thought for clinical diagnosis and treatment.

1. Introduction

Liver cancer is highly fatal, which causes the second highest death rate of cancer-related mortality worldwide [1,2]. Globally, it is estimated that approximately 80% of liver cancers were HCC [3]. Nobody disputes that this is a public health challenge that needs widespread attention. HCC is a multigene disease caused by the interaction of multiple cancer-promoting and suppressing genes with the microenvironment, and its molecular mechanism is still unclear. Thus, the identification of new potential therapeutic targets is urgently needed.

In recent years, despite the advances in our knowledge of the genetic factors, it is a pity that the death rates are increasing rapidly [4, 5]. If the HCC patients can be diagnosed early, the survival rate may be greatly improved by means of liver resection [6, 7]. However, due to the late diagnosis of most patients with HCC, the physical condition is not good enough to withstand the risk of surgery [8, 9]. What is worse, the survival rate of patients with advanced HCC is further decreased due to the widespread resistance to chemotherapy. Sorafenib, for example, a multikinase inhibitor, is widely used for the treatment of patients with advanced HCC with a long application time [10, 11], but patients invariably develop sorafenib resistance and it only provides limited survival benefit [12]. As a result, we are badly in need of finding new diagnostic and prognostic markers for HCC, which might facilitate early diagnosis and guide treatment decisions to improve patients’ survival and quality of life.

In this study, we downloaded the expression matrix of six datasets from the Gene Expression Omnibus (GEO) database, including 630 adjacent normal and 927 tumor tissues. PCA, tumor purity evaluation, and differential expression gene
(DEG) analysis were performed by using R software. 266 DEGs were finally obtained, consisting of 81 upregulated and 185 downregulated genes. FunRich undertook the entire enrichment analyses in our experiment, while Cytoscape was employed to build a network diagram. We had found that upregulated genes were closely related to mitotic cell cycle. Different from upregulated genes, downregulated genes were enriched in lipid and lipoprotein derivative pathway. To further explore the role of these DEGs, we divided the PPI network into several independent modules. FOXM1 and FTCD were core genes in two separate models with the highest score, respectively. The former is enriched in FOXM1 transcription factor network pathway, while the latter is mainly enriched in the metabolism of amino acid and derivative pathway. Finally, The Cancer Genome Atlas (TCGA) data were used to test our results and predict overall survival related to five hub genes in PPI. We found certain correlations among hub genes, which might reveal potential signaling pathways in HCC. And 4 of 5 hub genes were connected with low overall survival of HCC patients. Undoubtedly, recognition of biomarkers in HCC that plays a key role in disease progression can provide new insights into the development, prognosis, and treatment of HCC.

2. Materials and Methods

2.1. Tumor Purity Estimation and Differential Expression Gene Analysis. All the gene expression profile data originated from the GEO database (https://www.ncbi.nlm.nih.gov/geo). There were huge liver cancer mRNA microarray datasets in the GEO database, and the included datasets need to meet the following conditions: (1) the microarray data were available; (2) they contained at least 100 samples; and (3) they employed tumor and adjacent normal tissues. Therefore, we selected the following datasets: GSE25097, GSE36376, GSE45436, GSE54236, GSE64041, and GSE112790. GEOquery package in R/Bioconductor software (version 3.6.1, https://www.r-project.org) was used to get datasets, which was applied to download gene expression and probe annotation information for the selected datasets. Then, the estimate package was used to estimate tumor purity, while the limma package for data normalization and gene differential expression matrix acquisition. In the differential expression gene analysis, FDR < 0.05 and \( |\log_{2} FC| \geq 1 \) were considered to be significant DEGs, which were visualized by ggplot2 package.

2.2. Enrichment Analysis. We divided the DEGs into two categories and ranked them in descending order of absolute values. Since the six datasets were not from the same platform, we used RRA package to integrate the DEGs. FunRich (version 3.1.3, http://www.funrich.org) is such powerful stand-alone software that we primarily used to perform functional enrichment analysis [13]. Biological process, biological pathway, cellular component, and molecular function can be achieved by FunRich in the present study.

2.3. PPI Network and Module Analysis. The Search Tool for the Retrieval of Interacting Genes database (STRING, https://string-db.org) can provide information on protein interactions, whose data mainly came from structural predictions and literature reports [14]. Combined score \( \geq 0.4 \) was considered as the cutoff value, and the filtered node information was saved locally for subsequent visualization. Then, we used Cytoscape software (version 3.6.0, https://cytoscape.org) to build the protein-protein interaction (PPI) network and one of the plug-in in Cytoscape named Molecular Complex Detection (MCODE) was applied to detect notable modules in this PPI network [15]. As is known to all, network modules, as one of the characteristics of protein networks, may have specific biological significance. The default advanced option parameters (degree cutoff = 2, node score cutoff = 0.2, and k-core = 2) in MCODE already met our requirements, so we did not modify it. Moreover, models with score \( \geq 5 \) were used for further path enrichment analysis, which can help to explore the potential biological functions of DEGs.

2.4. Analysis for Expression Level and Correlation of the Hub Genes. The Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn) is an online website tool that can perform analysis including gene expression analysis and correlation analysis [16]. Data from TCGA and the Genotype-Tissue Expression (GTEx, http://commonfund.nih.gov/GTEx/) were used to apply a standard processing pipeline before being used by GEPIA. Based on the huge amount of data from GEPIA, we used it to demonstrate the expression of hub genes in LIHC tissues and normal ones and then made a boxplot to visualize the results. There are three correlation coefficients (Pearson, Spearman, and Kendall) for users to choose in GEPIA, and any sets given by TCGA and/or GTEx expression data were used to check the relative ratios between two genes.

2.5. Overall Survival Analysis and ROC Curve Analysis of Hub Genes. Kaplan–Meier plotter (KM plotter, http://kmplot.com/analysis/) is a database that can be accessed openly, which is the largest dataset including breast, ovarian, lung, and gastric cancer [17]. This database is rich in gene expression data and overall survival information from TCGA, which we can use to draw survival curves with 95% confidence interval hazard ratio and log-rank P value. Receiver operating characteristic (ROC) curve analysis was employed to verify the diagnostic performance of hub genes, and 3 years was set as the predicted time. Multivariate Cox proportional hazards regression analysis was performed based on hub genes. The risk score for predicting overall survival was calculated as follows: risk score = \( \sum_{i=1}^{n} (\text{coef}_i \times \text{Expr}_i) \), where coef is the regression coefficient and Expr is the expression level of the gene. Then, according to the mean risk score, samples were divided into low- and high-risk groups. Finally, survival analysis and ROC curve analysis of the risk score were performed by using the same method as described above.

3. Results

3.1. Tumor Purity Estimation of Tumor Tissue in Datasets and Identification of DEGs. 630 normal and 927 tumor samples were selected in this study (Table 1). The results of PCA showed...
that the normal control group and the tumor group in the six datasets could be discriminated very well (Figures 1(a)–1(f)). Then, we calculated the tumor purity of 927 liver tumor tissues through the estimate algorithm. As shown in Figures 1(g) and 1(h), the purity of tumors ranged from 0.179 to 0.979 and 55.8%, the purity of tumors ranged from 0.179 to 0.979 and 55.8%, Q\textsubscript{hen} we calculated the tumor purity of 927 liver tumor tissues datasets could be discriminated very well (Figures 1(a)–1(f)).

3.2. Enrichment Analysis of DEGs. DEGs from six independent datasets were integrated and introduced into FunRich for enrichment analysis. The biological processes for upregulated genes were mainly associated with spindle assembly and cell cycle (Figure 3(a) and Supplementary Table S2), while the molecular functions were about protein binding and kinase binding (Figure 3(b) and Supplementary Table S2). In addition, the vital gene of upregulated genes was enriched in methyltransferase activity, energy pathways, and histidine catabolism. Through the biological pathway enrichment analysis, we found that upregulated genes were closely related to mitotic cell cycle, DNA replication, mitotic G1-G1/S phases, and ATM pathway (Figure 4(a) and Supplementary Table S2). Different from upregulated genes, downregulated genes were enriched in lipid and lipoprotein derivative pathway (Figure 4(b) and Supplementary Table S2).

3.3. PPI Network Establishment and Pathway Analysis of Network Module. After introducing the gene list into STRING website, we finally got the information of 240 nodes and 1747 edges (combined score ≥ 0.4). Then, the network diagram was presented by Cytoscape based on the STRING database (Figure 5(a) and Supplementary Table S3). Interestingly, most of the nodes with higher connectivity were upregulated genes, which signified that they would be closer to the center of the circle. Four modules with score ≥ 5 were detected via MCODE (Figures 5(b)–5(e)).

3.4. Expression Level and Correlation of Hub Genes. A total of 419 samples were selected for gene expression level analysis, including 369 tumor tissues and 50 normal liver tissues. As shown in Figures 6(a)–6(e), the expression of the five hub genes in cancer tissues was significantly higher than that in normal ones. Moreover, it turned out by the correlation analysis that the increased expression of these genes was strongly correlated with the decreased expression of FTCD (Figures 6(f)–6(j)), and heatmap of correlation coefficients between hub genes was shown in Figure 6(k).

3.5. Prognostic Factor Detection and Validation. The analysis of 364 HCC patients clearly showed that 4 out of 5 genes with the highest degree in PPI had a significant impact on prognosis (Figure 7). High expression of FOXM1 (HR = 1.68; P = 0.0033), CCNA2 (HR = 1.69; P = 0.0029), AURKA (HR = 1.61; P = 0.0069), CDKN3 (HR = 1.29; P = 0.15), and CDC20 (HR = 2.3; P = 3.4e − 06) indicated worse survival rate in patients with HCC. In ROC curve analysis, FOXM1 (AUC = 0.654, sensitivity = 0.533, and specificity = 0.688),

---

**Table 1:** The gene expression profile data characteristics.

<table>
<thead>
<tr>
<th>Record</th>
<th>Organism</th>
<th>Tissue</th>
<th>Normal (n = 630)</th>
<th>Tumor (n = 927)</th>
<th>Region</th>
<th>Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE25097</td>
<td>Homo sapiens</td>
<td>Liver tumor</td>
<td>243 (38.6%)</td>
<td>268 (28.9%)</td>
<td>Boston</td>
<td>GPL10687—Rosetta/Merck Human RSTA Affymetrix 1.0 microarray, custom CDF expression beadchip</td>
</tr>
<tr>
<td>GSE36376</td>
<td>Homo sapiens</td>
<td>Liver tumor</td>
<td>193 (30.6%)</td>
<td>240 (25.9%)</td>
<td>Seoul</td>
<td>GPL10558—Illumina HumanHT-12 V4.0 expression beadchip</td>
</tr>
<tr>
<td>GSE45436</td>
<td>Homo sapiens</td>
<td>Liver tumor</td>
<td>39 (6.2%)</td>
<td>95 (10.2%)</td>
<td>Taipei</td>
<td>GPL570—[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array</td>
</tr>
<tr>
<td>GSE54236</td>
<td>Homo sapiens</td>
<td>Liver tumor</td>
<td>80 (12.7%)</td>
<td>81 (8.7%)</td>
<td>Modena</td>
<td>GPL570—Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name version)</td>
</tr>
<tr>
<td>GSE64041</td>
<td>Homo sapiens</td>
<td>Liver tumor</td>
<td>60 (9.5%)</td>
<td>60 (6.5%)</td>
<td>Basel</td>
<td>GPL6244—[HuGene-1.0-st] Affymetrix Human Genome 1.0 ST Array [transcript (gene) version]</td>
</tr>
<tr>
<td>GSE112790</td>
<td>Homo sapiens</td>
<td>Liver tumor</td>
<td>15 (2.4%)</td>
<td>183 (19.7%)</td>
<td>Tokyo</td>
<td>GPL570—[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array</td>
</tr>
</tbody>
</table>
Figure 1: Continued.
CCNA2 (AUC = 0.655, sensitivity = 0.81, and specificity = 0.433), AURKA (AUC = 0.621, sensitivity = 0.623, and specificity = 0.584), CDKN3 (AUC = 0.613, sensitivity = 0.941, and specificity = 0.292), and CDC20 (AUC = 0.686, sensitivity = 0.697, and specificity = 0.632) had certain predictive value for the 3-year survival rate of HCC patients. All of the top 20 genes in PPI exhibited a significant prognostic value (Supplementary Table S5). A risk score model was built based on CDC20 (coef = 0.349; \( P = 2.8 \times 10^{-7} \)) and another gene among the top 20 genes, named NUSAP1 (coef = −0.182; \( P = 0.046 \)). Compared with overall survival prediction with a single gene, the risk score model had a higher predictive value (AUC = 0.71, sensitivity = 0.789, and specificity = 0.547) (Figure 8).

4. Discussion

In the present study, we had detected totally 266 DEGs. FOXM1 was the most connected gene in upregulated genes in the PPI network, which had 44 edges. Increasing evidence has suggested that FOXM1 is elevated in many tumors, such as intrahepatic cholangiocarcinoma, oesophageal adenocarcinoma, gastric cancer, cervical cancer, and HCC [18–24]. Since FOXM1 can promote the proliferation and invasion of cancer cells, it may give rise to the poor prognosis and low survival rate of patients with high FOXM1 expression [22, 24–31]. Not only that, it was also found that FOXM1 contributes to tumor angiogenesis in the study of colorectal and gastric cancer [23, 32]. In previous studies, there was a large amount of evidence that FOXM1 directly or indirectly affects the occurrence and development of HCC, which is in line with our results [27–29, 33–39]. In addition, in an in vivo study of HCC, the growth of tumors in mice with FOXM1 deficiency was completely stagnated, suggesting that FOXM1 has the potential to become an independent biomarker of HCC [31]. It is worth mentioning that KIF4A and the expression level of KIF4A is positively correlated with FOXM1. Overexpression of both genes will lead to excessive cell proliferation and promote tumor development [22]. Meanwhile, we observed a significant increase in the expression of KIF4A in our experimental results (\( P = 3.18 \times 10^{-7} \)), which coincided with previous studies [22]. MircoRNA plays an active role in HCC as well. For instance, the expression of microRNA-135a transcribed by FOXM1 can affect the prognosis and survival rate of patients with HCC [40]. Unfortunately, we did not build a competing endogenous RNAs (ceRNA) network in this project to find potential downstream noncoding RNAs for FOXM1, which should be investigated in future study.

FTCD was the core gene in model 2 and had a certain correlation with FOXM1. Additionally, FTCD was found useful to distinguish early HCC from benign tumors, suggesting that it might be a potential marker for HCC early diagnosis [41]. The results of enrichment analysis for FTCD were consistent with prior reports that the decrease in FTCD expression impeded the degradation pathway of histidine, which leads to the poor performance of methotrexate [42]. Therefore, we infer that patients with HCC who were not responding to methotrexate may be associated with abnormal expression of FTCD. Besides, in the available evidence, we found that autoimmune hepatitis has a 0.6% to 0.7% probability of inducing HCC [43, 44]. Interestingly, by reducing the number of circulating autoreactive T cells, the increased expression of FTCD can prevent the progression of autoimmune hepatitis [45]. Thus, low level of FTCD might contribute to high incidence of HCC and serves as a useful biomarker for primary HCC. Further research will be needed to clarify the role of FTCD in tumorigenesis.

In addition to FOXM1, CCNA2, AURKA, CDKN3, and CDC20 can be seen in the forefront of PPI. It is common knowledge that CCNA2, a core cell cycle regulator, plays a critical role with high expression from S
Figure 2: Continued.
phase to early mitosis [46, 47]. It was reported that high expression of CCNA2 might induce hepatocyte nodular proliferation [48]. And the exosome circRNAs, secreted from liver adipocytes, promoted tumor growth by controlling miR-34a level and activating the USP7/CCNA2 signaling pathway [49]. All of the above findings indicate that CCNA2 directly or indirectly influences the development of HCC, which is considered as a vital part in HCC development. High expression of AURKA has previously been detected in different cancer types as well [50–52], which is implicated with the regulation of cell cycle and division [53]. With no exception to HCC, it is also described as an oncoprotein and therapeutic target. Microarray analysis pointed out that AURKA phosphorylated and stabilized hepatoma upregulated protein [54]. Moreover, a research revealed that AURKA can, in turn, give rise to malignant phenotypes of HCC by regulating HIF-1α through activation of AKT and p38-MAPK signaling pathways [55]. As a result, we conjecture that it may function as a cancer-promoting gene. Excessive replication of the centrosome is considered to be a common feature of almost all human cancers [56], and

<table>
<thead>
<tr>
<th>GSE112790</th>
<th>GSE64041</th>
<th>GSE45436</th>
<th>GSE36376</th>
<th>GSE25097</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.60</td>
<td>2.35</td>
<td>3.94</td>
<td>3.90</td>
<td>3.86</td>
</tr>
<tr>
<td>2.49</td>
<td>1.95</td>
<td>2.24</td>
<td>2.37</td>
<td>3.76</td>
</tr>
<tr>
<td>5.05</td>
<td>1.61</td>
<td>2.47</td>
<td>2.38</td>
<td>3.14</td>
</tr>
<tr>
<td>1.66</td>
<td>1.14</td>
<td>1.35</td>
<td>2.17</td>
<td>1.52</td>
</tr>
<tr>
<td>2.86</td>
<td>1.41</td>
<td>1.97</td>
<td>3.34</td>
<td>1.52</td>
</tr>
<tr>
<td>1.75</td>
<td>1.40</td>
<td>1.27</td>
<td>1.59</td>
<td>1.56</td>
</tr>
<tr>
<td>2.24</td>
<td>1.54</td>
<td>1.96</td>
<td>3.11</td>
<td>2.54</td>
</tr>
<tr>
<td>2.26</td>
<td>1.24</td>
<td>1.80</td>
<td>2.74</td>
<td>1.61</td>
</tr>
<tr>
<td>2.02</td>
<td>1.59</td>
<td>1.98</td>
<td>2.72</td>
<td>1.59</td>
</tr>
<tr>
<td>1.89</td>
<td>0.54</td>
<td>2.01</td>
<td>2.31</td>
<td>1.96</td>
</tr>
<tr>
<td>2.22</td>
<td>0.79</td>
<td>1.56</td>
<td>2.53</td>
<td>1.42</td>
</tr>
<tr>
<td>1.85</td>
<td>1.48</td>
<td>1.85</td>
<td>0.98</td>
<td>1.69</td>
</tr>
<tr>
<td>2.90</td>
<td>1.28</td>
<td>1.55</td>
<td>3.39</td>
<td>1.91</td>
</tr>
<tr>
<td>2.73</td>
<td>1.09</td>
<td>0.93</td>
<td>2.82</td>
<td>1.15</td>
</tr>
<tr>
<td>3.51</td>
<td>1.57</td>
<td>1.52</td>
<td>2.39</td>
<td>2.16</td>
</tr>
<tr>
<td>2.34</td>
<td>1.55</td>
<td>2.24</td>
<td>2.56</td>
<td>0.65</td>
</tr>
<tr>
<td>2.39</td>
<td>1.24</td>
<td>1.44</td>
<td>2.44</td>
<td>0.93</td>
</tr>
<tr>
<td>2.41</td>
<td>1.19</td>
<td>1.98</td>
<td>2.70</td>
<td>0.89</td>
</tr>
<tr>
<td>2.00</td>
<td>0.42</td>
<td>1.64</td>
<td>2.34</td>
<td>1.04</td>
</tr>
<tr>
<td>3.65</td>
<td>0.60</td>
<td>1.22</td>
<td>2.66</td>
<td>1.05</td>
</tr>
</tbody>
</table>

![Figure 2: Volcano plot of DEGs and heatmap of the top 10 up- and downregulated genes.](image)
Figure 3: Continued.
Figure 3: Histograms of molecular function and biological process for up- and downregulated genes. (a) Biological process for upregulated genes. (b) Molecular function for upregulated genes. (c) Biological process for downregulated genes. (d) Molecular function for downregulated genes.

Figure 4: Pie charts of biological pathway for (a) upregulated genes and (b) downregulated genes.
Figure 5: PPI network diagrams of DEGs and module network and pathway enrichment analysis of modules. (a) PPI network of DEGs. Red nodes represent upregulated genes and blue nodes represent downregulated genes. (b) Module 1, MCODE score = 40.78; (c) module 2, MCODE score = 11.778; (d) module 3, MCODE score = 6.615; (e) module 4, MCODE score = 5.6. Pathway analysis of (f) module 1 and (g) module 2.
Figure 6: Continued.
Figure 6: Top five hub genes expression level in liver hepatocellular carcinoma (LIHC) and correlation analysis among hub genes. (a–c) Boxplots of expression level of FOXM1, CCNA2, AURKA, CDKN3, and CDC20. (f–j) Scatter plots of correlation analysis of FTCD and the top five hub genes. (k) Heatmap of correlation coefficients among the top ten hub genes in module 1 and module 2.

Figure 7: Continued.
CDKN3 happens to have the ability to prevent this abnormality [57]. Further research indicated CDKN3 seemed to play a role in tumor suppression by CDC2 signaling pathway [58]. However, bioinformatics analysis for identification of molecular target genes in HCC revealed that the relative expression levels of CDKN3 were significantly upregulated in tumor tissues, which proved that our results were not accidental [59]. Therefore, we conjecture that the result may be due to the positive feedback regulation in the tumor microenvironment, which surely requires subsequent experiments to verify. Regrettably, since the expression level of CDKN3 is not associated with the prognosis of patients with HCC in our study, CDKN3 cannot be counted as a candidate biomarker accordingly. CDC20 played a pivotal role in the regulation of chromosome segregation and the timely end of mitosis [60]. Abnormal CDC20 expression has been detected in most human cancers [61–63], and CDC20 knockdown caused mitotic arrest to efficiently kill slip-page-prone and apoptosis-resistant cancer cells [64], supporting an oncogenic role of CDC20. In conclusion, combined with literature reports and our findings, FOXM1, FTCD, CCNA2, AURKA, and CDC20 are very competitive biomarkers of HCC, while whether CDKN3

![Figure 7](image-url)
Research still has the following shortcomings: first of all, our results were only based on GEO and TCGA data analysis and have not been verified; next in importance, there may be distinctions in gene expression among different types of tumors, which will be perfected in future experiments.

5. Conclusion

In conclusion, the integrated bioinformatics analysis was derived from 1557 tumor tissues and adjacent normal tissues in the GEO database. Tumor and normal samples came from different populations and half of the tumor samples have a purity of more than 0.765. 266 genes were eventually identified as candidate HCC biomarkers, which were enriched in signaling pathways closely related to cell proliferation and metabolic function. FOXM1, CCNA2, AURKA, CDKN3, and CDC20 were at the core of these
genes, which opened up new horizons for diagnosis, prognosis, and treatment of HCC patients.

**Data Availability**

The data used to support the findings of this study are available from the corresponding author upon request.

**Conflicts of Interest**

No conflicts of interest exist in the submission of this manuscript.

**Authors’ Contributions**

All authors substantially contributed to the drafting and revising of the article and the final approval of the version to be submitted. Linxin Teng contributed to the writing of the manuscript. Linxin Teng, Kaiyuan Wang, and Yu Liu contributed to data analysis, and Yanxia Ma contributed to data interpretation. Lei Bi provided research direction and manuscript revision. Weiping Chen and Lei Bi provided scientific research fund declaration and fund support. All authors read and approved the manuscript for publication.

**Acknowledgments**

This work was supported by grants from the National Natural Sciences Foundation of China (81873219 and 81774180), National Science Foundation of Jiangsu Province (BK20181425), Young Elite Scientists Sponsorship Program by CACM (CACM-2018-QNRC2-B07), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX19_1221). In addition, this work was also financially supported in part by grants from a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

**Supplementary Materials**

Supplementary Table S1: information for the DEGs. Supplementary Table S2: enrichment analysis of the DEGs. Supplementary Table S3: node information for PPI network. Supplementary Table S4: pathway analysis of network modules. Supplementary Table S5: univariate Cox regression analysis of the prognosis of the top 20 genes in PPI. (Supplementary Materials)

**References**


[34] M. Han, H. Lu, K. Han et al., “XTP8 promotes hepatocellular carcinoma growth by forming a positive feedback loop with FOXM1 oncogene,” *Biochemical and Biophysical Research Communications*, vol. 515, no. 3, pp. 455–461, 2019.


[38] C. Ho, C. Wang, S. Mattu et al., “AKT (v-akt murine thymoma viral oncogene homolog 1) and N-Ras (neuroblastoma ras viral oncogene homolog) coactivation in the mouse liver promotes rapid carcinogenesis by way of mTOR (mammalian target of rapamycin complex 1),” *FOXM1 (forkhead box M1)/SKP2, and c-Myc pathways*, *Hepatology*, vol. 55, no. 3, pp. 833–845, 2012.


