Identification of key genes and pathways in hepatocellular carcinoma
A preliminary bioinformatics analysis

Min Wu, PhD\textsuperscript{a}, Zhaobo Liu, PhD\textsuperscript{a}, Aijing Zhang, PhD\textsuperscript{b,*}, Ning Li, PhD\textsuperscript{a,b,*}

Abstract
Background: Hepatocellular carcinoma (HCC) is one of the most prevalent cancers worldwide. However, the precise mechanisms of the development and progression of HCC remain unclear. The present study attempted to identify and functionally analyze the differentially expressed genes between HCC and cirrhotic tissues by using comprehensive bioinformatics analyses.

Methods: The GSE63898 gene expression profile was downloaded from the Gene Expression Omnibus (GEO) and analyzed using the online tool GEO2R to identify differentially expressed genes (DEGs). Gene ontology (GO) functional analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the DEGs were performed in DAVID. The STRING database was used to evaluate the interactions of DEGs and to construct a protein-protein interaction (PPI) network using Cytoscape software. Hub genes were selected using the cytoHubba plugin and were validated with the cBioPortal database.

Results: A total of 301 DEGs were identified between HCC and cirrhotic tissues. The GO analysis results showed that these DEGs were significantly enriched in certain biological processes including negative regulation of growth and cell chemotaxis. Several significant pathways, including the p53 signaling pathway, were identified as being closely associated with these DEGs. The top 12 hub genes were screened and included TTK, NCPAG, TOP2A, CCNB1, CDK1, PRC1, RRM2, UBE2C, ZWINT, CDKN3, AURKA, and RACGAP1. The cBioPortal analysis found that alterations in hub genes could result in significantly reduced disease-free survival in HCC.

Conclusion: The present study identified a series of key genes and pathways that may be involved in the tumorigenicity and progression of HCC, providing a new understanding of the underlying molecular mechanisms of carcinogenesis in HCC.

Abbreviations: AURKA = aurora kinase A, BLCL = Barcelona Clinic Liver Cancer, BP = biological processes, CC = cellular components, CCNB1 = Cyclin B1, CDK1 = Cyclin-dependent kinase 1, CDKN3 = Cyclin-dependent kinase inhibitor 3, DEG = differentially expressed gene, DFS = disease free survival, GEO = gene expression omnibus, GO = gene ontology, HCC = Hepatocellular carcinoma, KEGG = Kyoto encyclopedia of genes and genomes, MCC = maximal clique centrality, MF = molecular functions, NCPAG = non-SMC condensin I complex, subunit G, OS = overall survival, PPI = protein-protein interaction, PRCC1 = protein regulator of cytokinesis 1, RACGAP1 = Rac GTPase activating protein 1, RRM2 = Ribonucleotide reductase M2, SATB1 = special AT-rich sequence-binding protein 1, STRING = search tool for the retrieval of interacting genes/proteins, TOP2A = topoisomerase II alpha, TTK = TK protein kinase, UBE2C = ubiquitin-conjugating enzyme E2C, ZWINT = ZW10 interactor.

Keywords: computational biology, genes, hepatocellular carcinoma

1. Introduction

Hepatocellular carcinoma (HCC) is one of the most prevalent cancers worldwide, especially in developing areas such as China and Southeast Asia\textsuperscript{11} It has been estimated that HCC patients in China account for over 50\% of all cases around the world per year.\textsuperscript{12,13} However, in Europe and North America, the HCC incidence rates over the last few decades have not been high.\textsuperscript{14,15} It has been reported that the incidence of HCC in Romania is 5.3/100,000 for both genders.\textsuperscript{16} The major risk factors for the malignant transformation of HCC include chronic hepatitis B or C virus infection, alcoholic liver disease and liver cirrhosis.

Accumulating evidence has revealed that the molecular pathogenesis of HCC may be closely associated with environmental influences and genetic factors, such as tumor suppressor gene inactivation, oncogene activation and gene mutation.\textsuperscript{15,17} However, the precise mechanisms of the development and progression of HCC remain unclear. HCC is difficult to diagnose early and difficult to treat, and it also has a poor prognosis and high recurrence rate. Therefore, it is urgent to identify potentially useful molecular biomarkers to precisely predict HCC onset and
and construct a PPI network. All of the default parameters were used in this study to evaluate their functional associations.

The STRING database, (https://string-db.org/) is a database of known and predicted protein-protein interactions, including direct (physical) and indirect (functional) associations. To better illustrate the interactions and functions of the DEGs, the STRING database was used in this study to evaluate their functional associations and construct a PPI network. All of the default parameters were used. Then, the PPI network was visualized with Cytoscape 3.6.1, an open-access tool for creating integrated models of biomolecular interaction networks. The key DEGs were selected using the maximal clique centrality (MCC) algorithm, and the cytoHubba plugin, a Cytoscape plugin, was used to determine the hub proteins or genes in the PPI network. The top 12 key DEGs were selected as hub genes.

2.5. Analysis of hub genes using the cBioPortal for cancer genomics

To analyze the integrative relationships of the hub genes and their clinical characteristics in HCC, the cBioPortal for Cancer Genomics (http://www.cbioportal.org/) was used, which is an open-access resource for exploring and analyzing genetic alterations across samples from multidimensional studies. The analyses of genomic mutations and survival prognosis in the selected TCGA datasets could be performed in the cBioPortal according to the instructions. In this study, patients with HCC (except for intrahepatic cholangiocarcinomas and fibrolamellar liver cancer) in the liver hepatocellular carcinoma dataset (TCGA, Provisional), were selected for analysis in the present study. (To reviewer #3)

3. Results

3.1. Identification of DEGs

A total of 434 probe set IDs were found to be differentially expressed between HCC and cirrhotic tissues with thresholds of adjusted \( P < 0.05 \) and \( |\log_{2}FC| > 1.5 \). The probes were matched to gene symbols using the Affymetrix database. Finally, 301 DEGs between HCC and liver cirrhosis were screened and used for further analysis.

3.2. GO and KEGG pathway enrichment analyses

To further investigate the functions and mechanisms of the DEGs, GO and KEGG enrichment analyses was performed in DAVID. The GO analysis results demonstrated that the DEGs were significantly enriched for the negative regulation of growth, cell chemotaxis, cell adhesion, cellular response to cadmium ion, and inflammatory response terms in the biological processes category. In addition, in the cell component analysis, it was revealed that these DEGs were predominantly involved in the extracellular region, extracellular space, extracellular matrix, extracellular exosomes and blood microparticles. Moreover, in the molecular function analysis, these genes were mainly associated with heparin binding, chemokine activity, immunoglobulin receptor binding, heme binding and oxygen binding. Furthermore, the KEGG pathway enrichment analysis indicated that the DEGs were mainly involved in mineral absorption (hsa04978), the p53 signaling pathway (hsa04151), caffeine metabolism (hsa00232), tryptophan metabolism (hsa00380) and arachidonic acid metabolism (hsa00590); these results are comprehensively summarized in Table 1. The results obtained for the GO and KEGG enrichment analyses are shown in Figure 1.

3.3. PPI network construction and analysis

Based on the STRING database, a PPI network of DEGs was constructed and visualized, as shown in Figure 2. A total of 269 nodes and 851 edges were mapped in the PPI network, with a local clustering coefficient of 0.48 and a PPI enrichment.
and a total of 301 DEGs between HCC and cirrhotic tissues were screened. Functional analysis showed that these DEGs were robustly associated with various biological processes, such as cell adhesion, inflammatory responses, cell chemotaxis and the negative regulation of growth, most of which are closely related to the genesis and progression of cancer. In addition, the enriched KEGG pathways of DEGs were mainly involved in p53 signaling, mineral absorption, cell cycle progression, metabolism, pathways related to proteoglycans in cancer, and cytokine-cytokine receptor interactions. Moreover, a PPI network with the DEGs was constructed, and 12 hub genes, including TTK, NCAPG, TOP2A, CCNB1, CDK1, PRC1, RRM2, UBE2C, ZWINT, CDKN3, AURKA and RACGAP1, were identified as the key genes in HCC.

TTK encodes a dual specificity protein kinase with the ability to phosphorylate serine, threonine, and tyrosine. Studies have established that the TTK protein kinase is most likely associated with cell proliferation and is essential for the accurate segregation of chromosomes in cellular mitosis. The alteration of this protein may result in aberrant mitotic spindles, and tumorigenesis may then occur. TTK has been found to be overexpressed in various tumors, such as pancreatic cancer, breast cancer and HCC.[13–15] Thus, TTK may be a favorable prognostic biomarker and a therapeutic target in cancer. NCAPG encodes the G subunit of the non-SMC condensin I complex, which is responsible for the conversion of interphase chromatin into mitotic-like condensed chromosomes during cell division. NCAPG has been found to be overexpressed in HCC compared with adjacent normal tissue and is closely associated with poor overall survival in HCC.[16]

TOP2A encodes DNA topoisomerase II alpha, a nuclear enzyme that is involved in altering topologic states of DNA during transcription and replication. It has been reported that overexpression of TOP2A is significantly associated with low survival and tumor metastasis in many types of tumors.[17–19] Cyclin B1 (CCNB1), a regulatory protein, plays an important role in controlling the G2/M transition phase during mitosis. CCNB1 upregulation has been reported to be a significant prognostic marker for poor outcome in HCC.[20,21]

Cyclin-dependent kinase 1 (CDK1), a member of the Ser/Thr protein kinase family, plays an essential role in the control of the eukaryotic cell cycle by modulating the centrosome cycle. CDK1 has been extensively investigated in ovarian cancer and colorectal cancer.[22,23] However, little is known about the role of CDK1 in HCC carcinogenesis. Protein regulator of cytokinesis 1 (PRC1), a microtubule-associated protein, has been shown to be involved in apoptosis inhibition and carcinogenic progression and to promote early recurrence in HCC.[24] Ribonucleotide reductase regulatory subunit M2 (RRM2) plays a role in catalyzing the biosynthesis of deoxyribonucleotides from ribonucleotides. RRM2 has been shown to be a prognostic biomarker and therapeutic target for breast cancer, prostate cancer and gastric adenocarcinoma. However, RRM2 has rarely been researched in HCC.

Ubiquitin-conjugating enzyme E2C (UBE2C), a member of the E2 ubiquitin-conjugating enzyme family, has been reported to be involved in processes such as the destruction of mitotic cyclins, cell cycle progression and even cancer progression. UBE2C has been shown to promote cell proliferation and cancer invasion in prostate carcinoma and breast cancer.[25] UBE2C is also overexpressed in hepatocellular carcinoma.[26] Thus, UBE2C may be a potential biomarker for cancer diagnosis and may be used to predict cancer prognosis and sensitivity to cancer treatment.

ZW10 interacting kinetochore protein (ZWINT) is a protein that is involved in kinetochore formation and spindle checkpoint

### Table 1

The altered Kyoto encyclopedia of genes and genomes (KEGG) pathways in HCC.

| Term             | Description                       | Count | P value |
|------------------|-----------------------------------|-------|---------|
| hsa04978         | Mineral absorption                | 7     | 5.71E-04|
| hsa04115         | p53 signaling pathway             | 8     | 7.87E-04|
| hsa00232         | Caffeine metabolism               | 3     | 0.04965 |
| hsa00380         | Tryptophan metabolism             | 5     | 0.12685 |
| hsa00550         | Arachidonic acid metabolism       | 6     | 0.03136 |
| hsa05205         | Proteoglycans in cancer           | 11    | 0.0608  |
| hsa00830         | Retinol metabolism                | 6     | 0.1698  |
| hsa04110         | Cell cycle                        | 8     | 0.0384  |
| hsa00051         | Fructose and mannose metabolism   | 4     | 0.03627 |
| hsa04060         | Cytokine-cytokine receptor interaction | 11   | 0.04865 |
| hsa04512         | EDM-receptor interaction          | 6     | 0.04865 |
| hsa04064         | NF-kappa B signaling pathway      | 6     | 0.051763|
| hsa04151         | PI3K-Akt signaling pathway        | 14    | 0.06544 |
| hsa04062         | Chemokine signaling pathway       | 9     | 0.07257 |
| hsa0982          | Drug metabolism - cytochrome P450  | 5     |         |
| hsa04610         | Complement and coagulation cascades| 5     |         |

HCC = hepatocellular carcinoma.

P value < 1.0E−16. The hub genes selected from the PPI network using the maximal clique centrality (MCC) algorithm and cytoHubba plugin are shown in Figure 3. The top 12 hub genes were TTK protein kinase (TTK), non-SMC condensin I complex, subunit G (NCAPG), topoisomerase (DNA) II alpha (TOP2A), Cyclin B1 (CCNB1), Cyclin-dependent kinase 1 (CDK1), Protein regulator of cytokinesis 1 (PRC1), Ribonucleotide reductase M2 (RRM2), Ubiquitin-conjugating enzyme E2C (UBE2C), ZW10 protein kinase with the ability to phosphorylate serine, threonine, and tyrosine. Studies have established that the TTK protein kinase is most likely associated with cell proliferation and is essential for the accurate segregation of chromosomes in cellular mitosis. The alteration of this protein may result in aberrant mitotic spindles, and tumorigenesis may then occur. TTK has been found to be overexpressed in various tumors, such as pancreatic cancer, breast cancer and HCC.[13–15] Thus, TTK may be a favorable prognostic biomarker and a therapeutic target in cancer. NCAPG encodes the G subunit of the non-SMC condensin I complex, which is responsible for the conversion of interphase chromatin into mitotic-like condensed chromosomes during cell division. NCAPG has been found to be overexpressed in HCC compared with adjacent normal tissue and is closely associated with poor overall survival in HCC.[16]

TOP2A encodes DNA topoisomerase II alpha, a nuclear enzyme that is involved in altering topologic states of DNA during transcription and replication. It has been reported that overexpression of TOP2A is significantly associated with low survival and tumor metastasis in many types of tumors.[17–19] Cyclin B1 (CCNB1), a regulatory protein, plays an important role in controlling the G2/M transition phase during mitosis. CCNB1 upregulation has been reported to be a significant prognostic marker for poor outcome in HCC.[20,21]

Cyclin-dependent kinase 1 (CDK1), a member of the Ser/Thr protein kinase family, plays an essential role in the control of the eukaryotic cell cycle by modulating the centrosome cycle. CDK1 has been extensively investigated in ovarian cancer and colorectal cancer.[22,23] However, little is known about the role of CDK1 in HCC carcinogenesis. Protein regulator of cytokinesis 1 (PRC1), a microtubule-associated protein, has been shown to be involved in apoptosis inhibition and carcinogenic progression and to promote early recurrence in HCC.[24] Ribonucleotide reductase regulatory subunit M2 (RRM2) plays a role in catalyzing the biosynthesis of deoxyribonucleotides from ribonucleotides. RRM2 has been shown to be a prognostic biomarker and therapeutic target for breast cancer, prostate cancer and gastric adenocarcinoma. However, RRM2 has rarely been researched in HCC.

Ubiquitin-conjugating enzyme E2C (UBE2C), a member of the E2 ubiquitin-conjugating enzyme family, has been reported to be involved in processes such as the destruction of mitotic cyclins, cell cycle progression and even cancer progression. UBE2C has been shown to promote cell proliferation and cancer invasion in prostate carcinoma and breast cancer.[25] UBE2C is also overexpressed in hepatocellular carcinoma.[26] Thus, UBE2C may be a potential biomarker for cancer diagnosis and may be used to predict cancer prognosis and sensitivity to cancer treatment. ZW10 interacting kinetochore protein (ZWINT) is a protein that is involved in kinetochore formation and spindle checkpoint.
Figure 1. Results of GO analysis and KEGG pathway enrichment analysis of the differentially expressed genes (DEGs). The x-axis indicates the functional annotations or significantly enriched KEGG pathways. The y-axis indicates the –log10 (P value). The left y-axis indicates the number of enriched genes. GO = Gene ontology. KEGG = Kyoto encyclopedia of genes and genomes.
activity through interactions with ZW10. It has been demonstrated that ZWINT is notably overexpressed in different human tumors, such as gastric, colon, breast and liver cancer tumors.\[27\] Thus, ZWINT might contribute to the tumorigenicity of hepatocellular carcinoma by upregulating cell proliferation.\[28\] Cyclin-dependent kinase inhibitor 3 (CDKN3), a member of the dual-specificity protein phosphatase family, plays a vital role in cell cycle regulation by dephosphorylating CDK2 kinase. CDKN3 has been demonstrated to be overexpressed and mutated in several types of cancers, such as gastric cancer, breast cancer and hepatocellular carcinoma.\[29,30\] Knockdown of CDKN3 may inhibit migration and proliferation in human cancer and promote apoptosis of cancer cells.\[31,32\] AURKA encodes a cell cycle-regulated kinase that plays a role in various mitotic events during chromosome segregation. It has been reported that AURKA was a target gene of beta-catenin in several tumors such as gastric cancer, multiple myeloma disease, hepatocellular carcinoma, esophageal squamous cell carcinoma.\[33–37\] Moreover, it had been demonstrated that AURKA might further participate in tumor development and metastasis by directly phosphorylating beta-catenin in esophageal squamous cell carcinoma.\[37\] Beta-catenin is a multifunctional protein that plays a critical role in the Wnt signaling pathway and in E-cadherin-mediated intercellular adhesion.\[38\] Moreover, nuclear beta-catenin may activate the transcription of target genes, such as CD44, cyclin D1 and c-myc.\[38,39\] Aberrant localization of
beta-catenin has been found to correlate with the overexpression of special AT-rich sequence-binding protein 1 (SATB1) and with APC gene alteration in colorectal cancer.\(^{[40,41]}\) Activated AKT may also promote the upregulation of beta-catenin in human epidermoid carcinoma cells.\(^{[42]}\) Rac GTPase activating protein 1 (RACGAP1), a component of the centralspindlin complex, plays a key role in controlling cell growth, differentiation and cytokinesis. RACGAP1 had been reported to serve as a metastatic driver in uterine carcinosarcoma and to increase tumor malignant potential in colorectal cancer.\(^{[43]}\) RACGAP1 may be a useful biomarker to monitor the early recurrence of HCC.\(^{[44]}\)

To further explore the relationships between the hub genes and HCC, a provisional TCGA cohort of HCC was used to evaluate the correlations between the hub genes and survival prognosis. The cBioPortal analysis found that mutations or alterations in the hub genes could result in significantly reduced disease-free survival and poorer clinical outcomes in HCC.

5. Conclusion
The present study attempted to identify and functionally analyze the differentially expressed genes between HCC and cirrhotic tissues by using comprehensive bioinformatics analyses and unveiled a series of key genes and pathways that may be involved in the tumorigenicity and progression of hepatocellular carcinoma. The cBioPortal analysis also demonstrated that mutations in the screened key genes could result in reduced disease-free survival in hepatocellular carcinoma. However, more experimental studies need to be carried out in the future to further validate these bioinformatics analysis results for hepatocellular carcinoma.
Acknowledgments
We appreciate all contributors to the GEO, cBioPortal, DAVID and STRING databases for providing open-access resources for exploring cancer genomics data.

Author contributions
Conceptualization: Zhaobo Liu, Ning Li.
Formal analysis: Min Wu.
Project administration: Ning Li.
Writing – original draft: Aiyi Zhang.

References
[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018;68:7–30.
[2] Petrick JL, Braunlin M, Laversanne M, et al. International trends in liver cancer incidence, overall and by histologic subtype, 1978–2007. Int J Cancer 2016;139:1534–45.
[3] Guo P, Huang ZL, Yu P, et al. Trends in cancer mortality in China: an update. Ann Oncol 2012;23:2755–62.
[4] Turdean S, Gurzu S, Turcu M, et al. Global trends and predictions in hepatocellular carcinoma mortality. J Hepatol 2017;67:302–9.
[5] Roman S, Ferro NA, Morelonula LE, et al. Hepatitis B Virus Genotyp H and environmental factors associated to the low prevalence of hepatocellular carcinoma in Mexico. J Cancer Ther 2013;4:367–76.
[6] Tang R, Liu H, Yuan Y, et al. Genetic factors associated with risk of metabolic syndrome and hepatocellular carcinoma. Oncotarget 2017;8:35403–11.
[7] Chen CF, Hsu EC, Lin KT, et al. Overlapping high-resolution copy number alterations in cancer genomes identified putative cancer genes in hepatocellular carcinoma. Hepatology 2010;52:1690–701.
[8] Mzik M, Chmelerova M, John S, et al. Aberrant methylation of tumour suppressor genes WT1, GATA3 and PAX5 in hepatocellular carcinoma. Clin Chem Laborat Med 2016;54:1971–80.
[9] Zhen L, Shijie N, Shuijun Z. Tumor PHD2 expression is correlated with clinical features and prognosis of patients with HCC receiving liver resection. Medicine 2014;93:e179.
[10] Jianjiong G, Bülent Arman A, Ugur D, et al. Integrative analysis of complex cancer genomics and clinical profiles using the eBioPortal. Sci Signal 2013;6:pl1.
[11] Best J, Bechmann L, Dechêne A, et al. Serological diagnosis of early HCC and environmental factors associated to the low prevalence of hepatocellular carcinoma in Mexico. J Cancer Ther 2013;4:367–76.
[12] Jianjiong G, Bülent Arman A, Ugur D, et al. Integrative analysis of complex cancer genomics and clinical pro...