Identification of key biomarkers in hepatocellular carcinoma induced by non-alcoholic Steatohepatitis or metabolic syndrome via integrated bioinformatics analysis

Bing Wang, Yiqing Zhang, Lin Gai, Yujie He, Hong Qiu*, Ping Li
Qinhua Medical District, Jinling Hospital, Nanjing, Jiangsu 210002, China

#Bing Wang and Yiqing Zhang as co-first authors

ARTICLE INFO

Original paper

Article history:
Received: April 20, 2023
Accepted: June 17, 2023
Published: July 31, 2023

Keywords:
Tumor biomarkers, hepatocellular carcinoma, gene expression profiling, metabolic syndrome, Non-alcoholic fatty liver disease

ABSTRACT

The burden of hepatocellular carcinoma (HCC) is steadily growing because obesity, type 2 diabetes, and nonalcoholic fatty liver disease (NAFLD) are replacing viral- and alcohol-related liver disease as major pathogenic promoters. The current study attempted to identify the key genes and pathways in the non-alcoholic steatohepatitis (NASH) induced development of HCC using integrated bioinformatics analyses. Two gene expression profiling datasets, GSE102079 and GSE164760 were downloaded. Differentially expressed genes (DEGs) from HCC and healthy control samples were screened. Functional enrichment analyses based on Gene Ontology (GO) resource, Kyoto Encyclopedia of Genes and Genomes (KEGG) resource. Then protein-protein interaction (PPI) of these DEGs was visualized by Cytoscape with Search Tool for the Retrieval of Interacting Genes (STRING). Expression and survival analysis of hub genes, methylation and genetic mutation analysis were explored with GEPIA2, UALCAN, GSCA, and TIMER2.0 databases. We identified 158 overlapping genes from the 2 datasets. Up-regulated genes were mainly related to the proliferation, adhesion and metastasis of tumors, while down-regulated genes were mainly related to oxidative stress and energy metabolism. CDKN2A, SPP1, CYP2C9 and CYP4A11 were associated with prognostic performance and were considered the potential crucial genes, which SPP1, CYP2C9 and CYP4A11 were identified as the DNA methylation-driven genes. In different mutation statuses of HCC, gene expression of CDKN2A, SPP1, CYP2C9 and CYP4A11 showed significant differences. CDKN2A and SPP1 were identified as risk genes, while CYP2C9 and CYP4A11 were identified as protective genes, which may affect the transformation of NASH into HCC.

Introduction

Hepatocellular carcinoma (HCC), the predominant form of liver cancer, is a major global health problem. HCC has several known etiologic factors: hepatitis B, hepatitis C, alcohol use, nonalcoholic fatty liver disease (NAFLD)/non-alcoholic steatohepatitis (NASH), and obesity (1,2). Although the incidence of NAFLD-related HCC is lower than that of HCC of other etiologies such as hepatitis C, more people have NAFLD than other liver diseases. The global incidence rate of NAFLD is increasing rapidly, so it has a huge impact on the incidence rate of HCC. In NAFLD, the incidence of HCC in patients with nonalcoholic steatohepatitis (NASH) is relatively high, especially in patients with metabolic syndrome (3). However, the research on the mechanism of NASH to promote the development of HCC is not deep enough.

Several studies showed that whether NASH is associated with cirrhosis nor NASH without cirrhosis, the risk of HCC was increased (4,5). NASH can progress to fibrosis and eventually lead to cirrhosis and its complications, including HCC. The main risk factors of liver cancer are caused by liver cirrhosis (6). However, there was evidence that a considerable number of NAFLD or NASH patients progress to HCC without cirrhosis (7). It is well known that NAFLD is strongly associated with metabolic syndrome, which includes obesity, dyslipidemia, and hypertension, and increases the risk of developing type 2 diabetes Mellitus (T2DM) (8). One hypothesis to explain the development of HCC in NAFLD patients without cirrhosis is that hepatocellular adenoma may lead to malignant transformation in the presence of metabolic syndrome (9). The reviews showed that patients with metabolic syndrome have an 81% increased risk of HCC. This may be related to specific molecular pathways of liver tumorigenesis, such as oxidative stress and reactive oxygen species production, hepatokines and adipokines imbalance (10-12).

A significant problem is the early recognition of NAFLD patients who will develop HCC, where new biomarkers are potential solutions to tackle this issue. In recent decades, the introduction of high-throughput technologies and bioinformatics analysis enabled identifying the pathogenic genes in carcinogenesis and screening potential biomarkers of cancer (13,14). In this research, we analyzed two public datasets to identify differentially expressed genes (DEGs) among healthy controls and metabolic syndrome/NASH-derived HCC. Then the candidate genes were performed comprehensive bioinformatic analysis, including expression analysis, the correlation with clinicopathological features, survival analysis and promoter methylation level, et al. We hope to screen potential genes for metabolic syndrome-related HCC or NASH-re-

* Corresponding author. Email: mapleqh@126.com
Cellular and Molecular Biology, 2023, 69(7): 174-180

Copyright: © 2023 by the C.M.B. Association. All rights reserved.
lated HCC and provide useful insights into the pathogenesis of HCC.

Materials and Methods

Gene expression profile data collection
The Gene Expression Omnibus (GEO) database collects and shares publicly a range of different high-throughput sequencing and microarray-based data sets. In our research, we searched the data sets, which consisted of patients with HCC and healthy controls. Microarray data of GSE102079, which was contributed by Chiyonobu N et al., was based on GPL570 (HG-U133 Plus 2) Affymetrix Human Genome U133 Plus 2.0 Array. Microarray data of GSE164760, which was contributed by Pinyol R et al., was based on GPL13667 (HG-U219) Affymetrix Human Genome U219 Array. Only 152 metabolism-associated HCCs and 14 normal livers in GSE102079, and 53 NASH-associated HCCs and 6 normal livers in GSE164760 were selected for subsequent analyses.

Identification of DEGs
GEO2R, an R-based tool in the web of GEO, enables users in screening DEGs from two groups of samples. We then screened DEGs between HCC and normal samples by applying GEO2R. The threshold for the DEGs was set as adjusted P-value <0.05 and |log2 fold change (FC)| ≥ 1. Then, the raw data in TXT format were checked in Evenn (http://www.ehbio.com/test/venn/#/) (15) to identify overlapping upregulated DEGs and downregulated DEGs in the two datasets.

Enrichment analysis of DEGs
In order to analyze the biological functions and pathways involved in common genes, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed using Metascape (https://metascape.org/), which is a web-based portal designed to provide a comprehensive gene list annotation and analysis resource for experimental biologists (16).

PPI network and module analysis
Protein-protein interaction (PPI) networks can be evaluated by a publicly available online tool, STRING (http://string-db.org/). PPIs of DEGs in STRING were selected with a score (median confidence) >0.4. The STRING app in Cytoscape3.9.1 was applied to examine the potential correlation between these DEGs. Then the hub genes were screened by MCODE. The criteria for the MCODE analysis were as follows: MCODE score >5, degree cutoff = 2, node score cutoff = 0.2, k-score = 2, and max depth = 100.

Expression and survival analysis of core genes
The expression and survival data of hub genes were analyzed by GEPIA2, which is a web-based tool that provides fast and customizable functionality based on data from The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) (17). GEPIA2 has key interactive and customizable analytical functions, including differential expression, gene mapping, correlation, patient survival analysis, similar gene detection, and dimension reduction analyses. Protein expression was analyzed by the Hub Genes in the Human Protein Atlas (HPA). HPA (https://www.proteinatlas.org/) website was used to compare the protein expression of the hub genes between normal endometrial tissue and endometrial cancer tissue with the application of the immunohistochemical (IHC) method. The association of mRNA expression of hub genes with clinicopathological parameters of LIHC patients was analyzed by UALCAN (http://ualcan.path.uab.edu/).

DNA methylation analysis
Correlation analysis between methylation levels and mRNA expression levels, and survival analysis for high and low methylation groups were evaluated through the GSCA database (18). Given that there are several methylation sites in one gene, the database calculated the site most negatively associate with mRNA expression. The UALCAN database (19) was applied to compare the promoter methylation levels between normal control tissues and HCC tissues. The level of DNA methylation was measured by the Beta value, which ranges from 0 (unmethylated) to 1 (fully methylated). P-value < 0.05 was considered to be significant. Functional DNA methylation refers to a significant negative correlation between methylation and gene expression of a particular gene.

Analysis of gene mutation
The cBioportal web platform (https://www.cbioportal.org/) was designed for comprehensive genomic analysis and TIMER2.0 (http://timer.comp-genomics.org/) was used to create a bar plot showing the common gene mutation rate for LIHC.

Results

Identification of DEG in HCC
There were 205 HCC tissues and 20 normal liver tissues in our present study. Via GEO2R online tools, we extracted 2291 and 644 DEGs from GSE102079 and GSE1064760 respectively. Then, we used Evenn online tool (15) to identify the common DEGs in the two datasets. Results showed that a total of 158 common DEGs were detected, including 102 down-regulated genes (logFC< 0) and 56 up-regulated genes (logFC> 0) (Figure 1).

Gene function and pathway enrichment analysis of DEGs
To investigate the biological function of 158 overlapping genes, GO/KEGG analyses were performed by the Metascape database. GO analysis results showed that the up-regulated genes were mainly enriched in BP (biological processes) such as ossification, blood vessel development, vasculature development, blood vessel morphogenesis, cell-cell adhesion; CC (cell component) such as basement membrane, vacuolar lumen, endoplasmic reticulum lumen; and MF (molecular function) such as integrin binding, extracellular matrix structural constituent, protease binding. KEGG results showed that the up-regulated genes were enriched in ECM-receptor interaction, Small cell lung cancer, AGE-RAGE signaling pathway in diabetic complications, Focal adhesion and other signaling pathways (Supplementary Figure 1A).

For down-regulated genes, they were mainly enriched in BP of alpha-amino acid metabolic process, carboxylic acid catabolic process, organic acid catabolic process, and cellular amino acid metabolic process. The enriched CC included a peroxisomal matrix, microbody lumen, blood
To validate different expressions of the above 4 genes between tumor and non-tumor tissue, 369 HCC tissues and 160 normal tissues were compared. The mRNA expression levels of CDKN2A and SPP1 were significantly increased in HCC tissues, while the levels of CYP2C9 and CYP4A11 were significantly decreased (Figure 3A). Additionally, we explored the protein expression of the hub genes on the HPA website and representative images were presented in Figure 3B. By the method of IHC, CDKN2A was medium staining in HCC tissues while not detected in normal tissues by antibody CAB018232; SPP1 was medium staining in HCC tissues while not detected in normal tissues by antibody HPA027541; CYP2C9 was low staining in HCC tissues while high staining in normal tissues by antibody HPA015066. IHC images of CYP4A11 in HCC tissues and normal tissues were not found on the HPA website.

Association between the hub genes expression and clinicopathological parameters in LIHC

The association of mRNA expression of hub genes (CDKN2A, SPP1, CYP2C9 and CYP4A11) with clinicopathological parameters of LIHC patients were analyzed by UALCAN, containing individual cancer stages and tumor grade. It was shown that the mRNA expression of hub genes was significantly correlated with LIHC individual cancer stages. CYP2C9 and CYP4A11 as favorable factors for LIHC patients, the high mRNA expressions of them tended to be in stage 1 or 2, whereas the high mRNA expressions of unfavorable factors of CDKN2A and SPP1 tended to be in stage 3 or 4. In addition, mRNA expres-
sion of the 4 hub genes was significantly related to histologic grades. The high mRNA expressions of CDKN2A and SPP1 tended to be in grade 3 or 4, whereas the high mRNA expressions of CYP2C9 and CYP4A11 tended to be in grade 1 or 2 (Supplementary Figure 3).

**Methylation analysis of 4 hub genes in LIHC**

The GSCA tool was introduced to analyze 4 hub genes methylation and 4 hub genes mRNA expression. As shown in Figure 4A, CDKN2A expression was weakly positively correlated with its methylation level in LIHC, but SPP1, CYP2C9 and CYP4A11 had a significantly strong negative correlation between gene expression and DNA methylation.

Compared with normal samples, the DNA methylation levels of CDKN2A and CYP2C9 were dramatically higher immethylation-level tumor tissues, while the DNA methylation levels of SPP1 and CYP4A11 were dramatically lower in tumor tissues (Figure 4B). We noticed that SPP1, CYP2C9 and CYP4A11 had been identified as DNA methylation-driven genes, and the gene expression value was significantly affected by DNA methylation. In tumor tissue, SPP1 showed high expression levels with low DNA methylation, while CYP2C9 showed high DNA methylation with low expression levels.

Furthermore, survival analysis for DNA methylation levels of 4 hub genes was analyzed. As shown in Fig 5, higher CDKN2A methylation was significantly correlated with a lower overall survival rate ($P=0.012756647$) and disease-specific survival rate ($P=0.006886467$), and higher CYP2C9 methylation was significantly correlated with lower disease-free interval rate ($P=0.01280962$) and progression-free survival rate ($P=0.003875283$). Compared with the high SPP1/CYP4A11 methylation group and the low SPP1/CYP4A11 methylation group, the survival difference between the groups were no significant statistical differences.

**Correlation between 4 hub genes expression level and TP53/CTNNB1/AXIN1 mutation in LIHC**

First, We analyzed gene mutations of 4 hub genes in LIHC/TCGA by inputting the genes into the cBioPortal website and found that the genetic alterations of CDKN2A, SPP1, CYP2C9 and CYP4A11 among 372 LIHC samples were 13%, 4%, 5%, 4% respectively (Supplementary Figure 4). Among the 4 hub genes, CDKN2A was the most frequently altered gene. In particular, deep deletion was identified as the primary type of genetic alteration of CDKN2A in LIHC. Next, the Gene_Mutation module of Timer2.0 compares the 4 hub genes expression between different mutation statuses. In Timer2.0, the commonest gene mutations in LIHC were TP53 mutation (101 of 365 patients), CTNNB1 mutation (95 of 365 patients), and AXIN1 mutation (22 of 365 patients) respectively. As shown in Figure 6A, the expression of CDKN2A in LIHC TP53-mutant tumors ($n=101$) was dramatically higher than in LIHC TP53-wildtype tumors ($n=264$), while the expression of CYP2C9 and CYP4A11 in LIHC TP53-mutant tumors was dramatically lower than in LIHC TP53-wildtype tumors. As shown in Figure 6B, the expression of SPP1 and CYP2C9 in LIHC CTNNB1-mutant tumors ($n=95$) was dramatically higher than in LIHC CTNNB1-wildtype tumors ($n=270$). As shown in Figure 6C, the expression of CDKN2A and SPP1 in LIHC AXIN1-mutant tumors ($n=22$) was dramatically lower than in LIHC AXIN1-wildtype tumors ($n=343$).
HCC is a major public health problem and a leading cause of death worldwide. HCC is characterized by rapid progression, recurrence, and metastasis; it is also associated with a high degree of malignancy and a high mortality rate (20,21). Previously, hepatitis C virus (HCV) was thought to be the leading cause of HCC (22,23), but recent reports that newly diagnosed HCC patients are non-viral HCC (2). NAFLD/NASH, obesity, T2DM, excessive alcohol consumption, and metabolic syndrome are at an increased risk of developing HCC (24-26). The average 5-year survival of HCC patients is generally poor, ranging from 5 to 14%. As for NAFLD-related HCC, the outcomes are in general inferior to HCV-related HCC due to the more advanced stage at diagnosis (27). NAFLD is characterized by fatty denaturation and lipid accumulation in hepatocytes and insulin resistance (28-30). It is the hepatic manifestation of metabolic syndrome and is a spectrum of conditions ranging from benign hepatic steatosis to non-alcoholic steatohepatitis (NASH). NASH, the more aggressive form of NAFLD, could develop into progressive fibrosis and is directly associated with the risk of developing hepatocellular carcinoma (HCC). Therefore, an increased understanding of the underlying mechanisms leading to the risk of HCC induced by NASH is necessary to develop effective prevention and treatment approaches for HCC.

Herein, a series of bioinformatics analyses were performed on two independent gene chip databases (from normal and NASH or metabolic syndrome-related liver cancer tissues), and 158 common DEGs were identified, of which 56 were up-regulated and 102 down-regulated. The results demonstrated that, in up-regulated DEGs, BP terms of GO were mainly enriched in blood vessel development and cell-cell adhesion, and the KEGG analysis results were mainly enriched in ECM-receptor interaction and Focal adhesion. These results indicated that low-expression genes play a key role in oxidative stress and energy metabolism.

We observed that the high expression of CDKN2A and SPP1 was closely related to the decrease of OS and DFS in HCC patients. CDKN2A is a member of the INK4 family, which is an important family of cyclin-dependent kinase inhibitors (CDKIs). CDKN2A gene can produce different transcripts through variable splicing, encoding at least three different proteins, two of which are p16 (INK4) and p14 (ARF), respectively. The high expression of CDKN2A can promote the proliferation of cancer cells, inhibit the apoptosis of cancer cells, induce tumor interstitial angiogenesis, reduce the sensitivity of cancer cells to chemoradiotherapy, and ultimately affect the prognosis of HCC patients (31).

SPP1 gene encodes a secretory phosphorylated glycoprotein, which is highly expressed in many tumors. It belongs to the small integrin-binding ligand N-type glycoprotein (SIBLING) family, which is highly expressed in lymphocytes, endothelial cells, bone cells and a variety of malignant tumor cells (32). Previous studies have illustrated that tumor-driven hypoxia promotes the expression of SPP1, which in turn promotes tumor angiogenesis and immunosuppressive microenvironment (33-35). By modulating epidermal growth factor (EGFR) activation, SPP1 can influence the immune escape and malignant biological activity of tumor cells, and its overexpression enhances HCC development and metastasis (36,37). Guixiong Zhang et al. (38) found that the OS of the low-SPP1-expression group of HCC patients who received anti-angiogenesis combined with immunotherapy after resection was a trend longer than that of the high-SPP1-expression group. SPP1 may be considered a general marker of cancer progression, would be valuable in combination with other biomarkers to guide patient stratification and treatment strategies, and would be an attractive therapeutic target due to its multiple roles in promoting tumor aggressiveness.

CYP2C9 and CYP4A11 were downregulated in HCC tissues and identified as protective genes. CYP2C9 and CYP4A11 are members of the CYP450 gene family, both of which are down-regulated in HCC tissues. CYP2C9 is involved in various biological processes from the synthesis of lipids to drug metabolism and is a drug-metabolizing enzyme gene (DME gene) that regulates cell growth, apoptosis, differentiation, and homoeostasis and is involved in hepatocarcinogenesis (39). CYP2C9 participates in the metabolism of xenobiotics and fatty acids in the liver. Downregulation of CYP2C9 may be a biomarker of HCC (40,41). Hyuk Soo Eun et al showed that CYP4A11 expression is a favorable prognostic factor of HCC and suggest potential predictive diagnostic and prognostic roles of CYP4A11 expression in HCC (42). Reducing cytochrome P450 gene expression has been linked to the aggravation of hepatocellular carcinoma and affects various regulated metabolites.

It has been confirmed that epigenetic aberrance, especially DNA methylation, plays an important role in the progression of carcinoma (43). In the present study, SPP1 was hypomethylated and expressed at a higher level in HCC than in normal tissues, while CYP2C9 was hypermethylated and had low expression. Methylation levels of the CDKN2A and CYP2C9 were closely related to the survival rate of patients with HCC. The higher methylation levels were usually associated with poor survival, which
indicates the prognostic value of CDKN2A and CYP2C9 methylation.

In addition, cBioPortal was used to summarize the possible genetic alterations for 4 hub genes in HCC. We identified deep deletion as the most frequent alteration in the CDKN2A gene in HCC. Approximately 8% of HCCs harbor CDKN2A deletions (44,45). CDKN2A inactivation has been correlated with poor prognosis independently of other traditional factors; in addition, CDKN2A alterations are discerned in more advanced, aggressive cancer (44).

Genetic mutation is one of the most common mechanisms of carcinogenesis. HCC has especially frequent mutations, including TP53, TERT, CTNNB1, AXIN1, CCND1 and FGF19 et al. In our study, gene expression levels of four hub genes showed significant differences in HCC patients with and without mutant TP53, CTNNB1, and AXIN1. TP53 acts as a tumor suppressor in tumors. Genomic aberrations in the p53 pathway are the most frequent abnormalities in diverse cancers and often correlate with high-grade histology (46). HCC patients with TP53 mutations had worse clinical stages and shorter overall survival time compared with patients with wild-type TP53 (47,48). The risk gene (CDKN2A) up was regulated while protective genes (CYP2C9 and CYP4A11) were downregulated in mutated TP53. CTNNB1 and AXIN1 are important components of the canonical Wnt signaling pathway, which regulates cell adhesion, growth, and differentiation. The most important mechanism of β-catenin activation in HCC is the mutations in CTNNB1 and the mutations in AXIN1. However, despite belonging to the same pathway, genetic alterations in CTNNB1 and AXIN1 are mutually exclusive, possibly because they carry opposite roles in terms of pathway activation (49). CTNNB1 is the effector of Wnt signaling while AXIN1 is a negative regulator of Wnt signaling. Studies showed that the gene sets significantly up-regulated in CTNNB1 mutant HCCs were all associated with metabolic pathways (50).

In conclusion, by integrating multiple microarray gene expression profiles, four key genes (CDKN2A, SPP1, CYP2C9, CYP4A11) were identified that appear to be important as potential diagnostic and therapeutic targets in metabolic syndrome/NASH-derived HCC. The GO and KEGG analyses revealed further molecular mechanisms underpinning the regulation of HCC induced by metabolic syndrome. This study had some limitations. All data were downloaded from online databases and analyzed by computer algorithms; further studies including clinical sample analyses, and cell and animal experiments are required to validate the results. Together, these findings may lead to more effective prevention, detection, and treatment of metabolic syndrome/NASH-derived HCC.

Acknowledgements
This work was supported by Jinling Hospital Natural Science Foundation (YYMS2021039).

Conflict of interest
The authors declare that there is no conflict of interest regarding the publication of this paper.

Data availability
The open-access datasets are available through the following URL: GSE102079: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE102079
All data generated or analyzed during this study are available from the corresponding author upon reasonable request.

References
20. Chidambaramanathan-Reghupathy S, Fisher PB, Sarkar D. Hepato-
41. Ye T, Lin L, Cao L, et al. Novel Prognostic Signatures of Hepato-
cellular Carcinoma Based on Metabolic Pathway Phenotypes. Front Oncol 2022; 12: 863266.