Comprehensive Analysis of the Expression and Prognosis for Lipid Metabolism-Related Genes in Hepatocellular Carcinoma

Wen-Jie Fan1  Hao Ding2  Xiang-Xun Chen1  Lin Yang1

1 Department of Radiology, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, People’s Republic of China
2 Department of Gastroenterology, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, People’s Republic of China

Address for correspondence: Hao Ding, PhD, Department of Gastroenterology, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China (e-mail: dh198917@126.com).

South Asian J Cancer

Abstract

Background  This study aimed to screen potential key genes associated with lipid metabolism and to evaluate their expressions and prognosis values in hepatocellular carcinoma (HCC).

Methods  Data sets GSE6764, GSE14520, and GSE112790 were used to identify the common differentially expressed genes (DEGs). Protein–protein interaction (PPI) network was constructed by STRING database. Hub genes in PPI network were identified and subjected to functional enrichment analysis to screen lipid metabolism-related genes. The expressions of selected genes and their associations with prognosis were analyzed using UALCAN, The Human Protein Atlas, and Kaplan–Meier plotter databases. The transcriptional factor (TF)-gene regulatory network was constructed using NetworkAnalyst.

Results  A total of 331 common DEGs including 106 upregulated and 225 downregulated genes were identified. PPI network analysis showed that 76 genes with high degrees were identified as hub genes, among which 14 genes were lipid metabolism-related genes. PON1, CYP2C9, and SPP1 were found to be the independent prognostic markers. Key TFs with close interactions with these prognostic genes, including HINFP, SRF, YY1, and NR3C1, were identified from the TF-gene regulatory network.

Conclusion  This study presented evidence for the prognostic capabilities of lipid metabolism-related genes in HCC, and newly identified HINFP and NR3C1 as potential biomarkers for HCC.

Introduction

Hepatocellular carcinoma (HCC), the most common type of liver cancer, ranks as the fifth most common cancer and the third leading cause of cancer-related mortality worldwide, with an age-adjusted incidence of 10.1 cases per million per year.1,2 The medium survival time of patients with HCC following diagnosis is estimated between 6 and 20 months if not intervened.3 Though many efforts have been done, the
molecular mechanisms underlying the initiation and progression of HCC remains largely undetermined, and the therapeutic effect is far from satisfactory due to multiple reasons including but not limited to postsurgical recurrence and drug resistance. Therefore, exploring the underlying mechanisms and identifying effective prognostic biomarkers for the development of novel diagnostic and treatment approaches has become an urgent mandate.

Alterations in cellular metabolism are recognized as one of the 10 hallmarks of cancer, which allow tumor cells to survive in unfavorable environments and to maintain their homeostasis. Compared with other metabolic alterations such as glucose or glutamine metabolism, alterations in lipid metabolism in HCC have received less attention. However, recent studies have reported that fatty acid (FA) synthesis-related genes including ACLY, FASN, SCD, and SREBP1 were upregulated, whereas FA oxidation-related genes including ACAA1/2, ACADSB, and HADH were downregulated in patients with HCC. Moreover, pharmacological or genetic modulation of FA metabolic enzymes FASN, ACADS, and ACADL significantly affected tumor growth, invasion, and metastasis, indicating the importance of lipid metabolic alteration in HCC carcinogenesis. Despite these advances, studies concerning lipid metabolism in HCC remain quite limited. In the present study, we used integrated bioinformatics to screen crucial genes associated with lipid metabolism and to evaluate their expressions and prognosis values in HCC, thus to deepen the understanding of pathogenesis and to provide a novel insight into HCC-interventional strategies.

Methods

Gene Expression Data

The gene expression profiles of GSE6764, GSE14520, and GSE112790 were downloaded from Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). The GSE6764 is based on the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array) and includes 10 normal liver tissue specimens and 35 HCC tissue specimens. GSE14520 is based on the GPL3921 platform (Affymetrix HT Human Genome U133A Array) and includes 220 nontumor tissue specimens and 225 tumor tissue specimens. GSE112790 is based on the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array) and includes 15 nontumor tissue specimens and 183 tumor tissue specimens. The t-test and Benjamini–Hochberg method were used to calculate the p-value and false discovery rate, respectively. Genes showing altered expression with adjusted p-value < 0.05 and |log₂FC (fold change)| > 1 were identified as differentially expressed genes (DEGs). Then, the common DEGs aggregated in these three data sets were obtained using VENNY 2.1.0 (https://bioinfogp.cnb.csic.es/tools/venny/).

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Enrichment Analyses

The common DEGs from GSE6764, GSE14520, and GSE112790 were further subjected to Gene Ontology (GO) biological process and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses using the online STRING database (http://string-db.org). Meanwhile, the enrichment analyses were also performed for up- and downregulated DEGs and the identified hub genes. The cutoff value for significant GO and KEGG enrichment analyses was set as p < 0.05.

Protein–Protein Interaction Network

The aggregated DEGs were mapped using the STRING database (http://string-db.org) to evaluate the protein–protein interaction (PPI) information. Then, the PPI network was visualized through Cytoscape software V3.5.1. In the PPI network, a protein serves as a node and the degree of a node represents the number of nodes linked to it. In the present study, all the nodes with degree ≥ 1 were reserved and genes displaying a node degree of ≥ 15 were identified as hub genes.

Overall Survival Analysis

The Kaplan–Meier plotter (http://kmplot.com/analysis/) is a comprehensive online platform that can assess the effect of 54k genes on survival in 21 cancer types. The platform incorporates public ribonucleic acid-sequencing (RNA-seq) and microarray data obtained from the GEO, European Genome-phenome Archive, and The Cancer Genome Atlas (TCGA) databases. In the present study, patient samples were stratified into high- and low-expression groups according to the median expressions of the selected genes and assessed by a Kaplan–Meier survival plot with hazard ratio, 95% confidence intervals, and log rank p-value.

Validation of the Expression Levels of DEGs

UALCAN (http://ualcan.path.uab.edu) is a comprehensive, publicly available, and interactive online portal for analyzing cancer transcriptome data, which includes RNA-seq and clinical information of 31 cancer types from TCGA. The Human Protein Atlas (HPA) (http://www.proteinatlas.org) is an interactive open-access database that provides the protein expression profiles for a variety of human proteins. In the present study, the gene and protein expressions of the selected genes in HCC were validated using UALCAN and HPA database, respectively.

Construction of Transcriptional Factor-Gene Regulatory Network

The transcriptional factors (TFs) of selected genes were explored using the online tool NetworkAnalyst (http://www.networkanalyst.ca/faces/home.xhtml). The TF-gene regulatory network was then constructed and visualized using Cytoscape software V3.5.1.

Results

Gene Expression Analysis

As shown in Fig. 1A, a total of 1,253 (550 upregulated and 703 downregulated), 852 (350 upregulated and 502 downregulated), and 1,100 (531 upregulated and 569 downregulated) DEGs were identified from the GSE6764, GSE14520, and GSE112790, respectively. Next, 331 common DEGs aggregated in these three data sets were obtained using VENNY 2.1.0.
Among the common DEGs, 106 DEGs were upregulated and 225 DEGs were downregulated.

Then, the GO and KEGG pathway enrichment analyses were performed. GO enrichment analysis revealed that the top enriched biological processes in the upregulated 106 DEGs and downregulated 225 DEGs were mainly related to cell mitosis/cell cycle and material metabolism, respectively. As for the KEGG pathway analysis, the top 3 pathways enriched by the upregulated genes were related to cell cycle, deoxyribonucleic acid (DNA) replication, and extracellular matrix-receptor interaction; while the top 10 pathways enriched by the downregulated genes were mostly related to material metabolism (Fig. 1B and C).

**PPI Network Construction and Identification of Hub Genes Associated with Lipid Metabolism**

The online STRING database was used to construct the PPI network of proteins encoded by the 331 common DEGs. As shown in Fig. 2, a total of 299 nodes (proteins) and 1,854 edges (interactions) were identified. Of the 299 proteins, 95 were upregulated and 204 were downregulated. Seventy-six genes with degree $\geq 15$ were identified as hub genes (Table 1). Then, GO biological process analysis for the 76 genes was performed using the STRING database and showed that 14 genes (ACSL1, PON1, HAO2, CYP1A2, LPA, FABP1, ACYL, ACA1, SPP1, ALDH8A1, CYP2C9, PCK1, CYP2B6, TTR) were involved in lipid metabolic process.
Identification and Expression Validation of Lipid Metabolism-Related Genes with Prognostic Value in HCC

Overall survival analysis was performed using the online platform Kaplan–Meier plotter. A total of 364 HCC patients were stratified into high- and low-expression groups according to the median expressions of the selected genes. As shown in Fig. 3, high expressions of ACSL1, PON1, HA02, LPA, ALDH8A1, CYP2C9, PCK1, and TTR were significantly associated with a better prognosis (p < 0.05), while high expression of SPP1 was significantly associated with a worse prognosis (p < 0.05). In addition, no significant correlation with prognosis was observed for the expressions of CYP1A2, FABP1, ACLY, ACA1, and CYP2B6.
The prognosis-related genes were further included in the subsequent multivariate analysis. It was concluded that PON1, CYP2C9, and SPP1 were independent prognostic markers for overall survival of patients with HCC (►Fig. 4A). We then assessed the expression levels of these prognostic genes in HCC and normal tissues using UALCAN and HPA databases and found that PON1 and CYP2C9 were lowly expressed and SPP1 was highly expressed in HCC tissues, both at the messenger RNA and protein levels (►Fig. 4B and C).

**TF Analysis for the Prognostic Genes Associated with Lipid Metabolism**

A TF-gene regulatory network was constructed for the prognostic genes associated with lipid metabolism, which included 27 interaction pairs among the 3 prognostic genes and 23 TFs. HINFP, SRF, YY1, and NR3C1 were predicted to be the key TFs regulating the expressions of these three prognostic genes (►Fig. 5A). Furthermore, the gene expressions of HINFP, SRF, YY1, and NR3C1 were analyzed using UALCAN database and found to be dysregulated in HCC (►Fig. 5B). Survival analysis revealed that the expression level of each gene was associated with HCC patient survival (►Fig. 5C).

**Discussion**

With the rapid development of high-throughput sequencing technologies, bioinformatics has become an important component of biomedical research and been widely used in the...
field of oncology. Public databases provide a large number of clinical information and gene expression data and serve as a tremendous resource for biomedical researchers. In the present study, we integrated expression data from three GEO data sets and screened out 331 common DEGs in HCC by bioinformatic analysis. GO and KEGG pathway-enrichment analyses were performed and showed that the upregulated DEGs were mainly related to cell division, while the downregulated DEGs were mainly related to material metabolism. These findings were consistent with a previous study by Yan et al. Furthermore, the PPI network of the aggregated DEGs was constructed using the STRING database and identified a total of 76 hub genes with a cutoff degree of ≥ 15.

In recent years, lipid metabolism has attracted increasing attention in the field of cancer research. The lipid-rich environment is considered to promote the proliferation and metastasis of tumor cells. Consistently, tumor cells usually uptake larger amount of lipids accompanied by enhanced lipogenesis compared with normal cells. Furthermore, alterations in lipid metabolic process were observed both in FVB and C57B6 strains of Mdr2-knockout

---

**Fig. 4** Expression profile and prognostic value of lipid metabolism-related genes in hepatocellular carcinoma (HCC). (A) Risk ratio forest plot showing the prognostic value of the selected genes. (B) Validation of the messenger ribonucleic acid (mRNA) expression levels of the prognostic genes using the UALCAN database. (C) Validation of the protein expression levels of the prognostic genes using The Human Protein Atlas (HPA) database.
mice, a model of inflammation-mediated HCC. While in the mouse model of nonalcoholic steatohepatitis-driven HCC, acylcarnitine was found to be accumulated in liver tumor tissues due to the suppression of FA oxidation, which further enabled HCC cells to acquire stem cell properties. These findings highlight the importance of lipid metabolic process in hepatocarcinogenesis and suggest a great potential in the development of novel diagnostic and treatment approaches in HCC. In this study, we performed GO biological process analysis for the 76 hub genes using the STRING database and found that 14 genes, namely ACSL1, PON1, HAO2, CYP1A2, LPA, FABP1, ACILY, ACAAI, SPP1, ALDH8A1, CYP2C9, PCK1, CYP2B6, and TTR, were enriched in the GO term of “lipid metabolic process.” Of these 14 lipid metabolism-related genes, PON1, CYP2C9, and SPP1 were significantly associated with overall survival and were identified to be independent prognostic markers for HCC, which were in concordance with previous studies. Notably, two recent studies have obtained different results regarding the key lipid metabolism-related genes in HCC. Wang et al have identified ME1, MED10, and MED22 as the lipid metabolism-related genes with prognostic value in HCC. 22 While Zhu et al have screened out a 6-gene signature consisting of 6 lipid metabolism-related genes (FM03, SLC11A1, RNF10, KCNH2, ME1, and ZIC2) as an independent prognostic factor for HCC. 23 The reasons for these discrepancies may be as followed. First, the patient data were obtained from different public sources (Wang et al: TCGA liver cancer data [370 samples]; Zhu et al: TCGA liver cancer data [370 samples];)}
Role of lipid metabolism-related genes in HCC  Fan et al.

TCGA liver cancer data [342 samples] + GSE15654 data set; the present study: GSE6764 + GSE14520 + GSE112790.

Second, different analysis strategies were used. Third, different conclusions are often drawn due to the molecular heterogeneity of tumors, even if the included tumor samples are almost the same. Future studies with larger sample sizes and more accurate classification are needed to address this question.

We further confirmed the gene and protein expressions of these three prognostic genes using the UALCAN and HPA databases and showed consistent results with microarray data of GSE6764, GSE14520, and GSE112790. To explore the underlying molecular mechanisms regulating these prognostic genes, the TF-gene regulatory network was further constructed and identified HINFP, SRF, YY1, and NR3C1 as key TFs. SRF is a transcription factor of the MADS-box family that governs fundamental biological processes such as cell migration, cell growth, cytoskeletal organization, and differentiation.24,25 Accumulating evidence has suggested that SRF plays an essential role in triggering the formation and inducing epithelial to mesenchymal transition with resistance to sorafenib in HCC.25,26 YY1 is a member of the GLI-Kruppel class of zinc finger DNA-binding proteins and has been shown to promote tumorigenicity, angiogenesis, metastasis, and inhibit apoptosis of cancer cells in HCC.27–29 As for NR3C1, there was only one study reporting that the hydroxymethylation of NR3C1, the glucocorticoid receptor, was increased in mice with chronic alcohol consumption, which may create a carcinogenic environment.30 However, no study concerning the role of HINFP in HCC has been reported until now. To our knowledge, this is the first study demonstrating the abnormal expressions of NR3C1 and HINFP in tumor tissues and their associations with the prognosis in HCC.

Conclusion

This study systematically analyzed the differential gene expression pattern in HCC and identified PON1, CYP2C9, and SPP1 as the prognostic genes associated with lipid metabolism and HINFP, SRF, YY1, and NR3C1 as their key TFs. These findings further supported the essential roles of PON1, CYP2C9, SPP1, SRF, and YY1 in HCC, and suggested HINFP and NR3C1 as new potential biomarkers for the prognosis of HCC. Given the limitations of this study, further studies are needed to verify the reliability of our results and elucidate the function of these potential prognostic genes.

Funding

None.

Conflict of Interest

The authors declare that there are no competing financial interests.

Acknowledgments

None.

References

Role of lipid metabolism-related genes in HCC

Fan et al.

28 Yang W, Li Z, Qin R, et al. YY1 promotes endothelial cell-dependent tumor angiogenesis in hepatocellular carcinoma by transcriptionally activating VEGFA. Front Oncol 2019;9:1187