Bioinformatics analysis of key biomarkers and potential molecular mechanisms in hepatocellular carcinoma induced by hepatitis B virus

Zhe Li, MD, Jingyong Xu, PhD, Hongyuan Cui, PhD, Jinghai Song, PhD, Jian Chen, MD, Junmin Wei, MD

Abstract

Background: Hepatocellular carcinoma (HCC) accounts for up to 90% of all primary hepatic malignancies; it is the sixth most common cancer and the second most common cause of cancer mortality worldwide. Numerous studies have shown that hepatitis B virus and its products, HBV integration, and mutation can induce HCC. However, the molecular mechanisms underpinning the regulation of HCC induced by HBV remain unclear.

Methods: We downloaded 2 gene expression profiling datasets, of HBV and of HCC induced by HBV, from the gene expression omnibus (GEO) database. Differentially expressed genes (DEGs) between HCC and HBV were identified to explore any predisposing changes in gene expression associated with HCC. DEGs between HCC and adjacent healthy tissues were investigated to identify genes that may play a key role in HCC. Any overlapping genes among these DEGs were included in our bioinformatics analysis. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of overlapping genes were performed using the Metascape online database; the protein–protein interaction (PPI) network was analyzed using the STRING online database; and we obtained the hub genes of the PPI network using Cytoscape software. An overall survival (OS) analysis of hub genes was performed using km-plotter and the gene expression profiling interactive analysis (GEPIA) online database. The expression levels of hub genes were determined using the TCGA and GEPIA databases. Finally, the relationships between hub genes and tumors were analyzed using the comparative toxicogenomics database (CTD).

Results: We identified 113 overlapping genes from the 2 datasets. Using functional and pathway analyses, we found that the overlapping genes were mainly related to the AMPK signaling pathway and cellular responses to cadmium ions. C8A, SPP2, KLKB1, PROZ, C6, FETUB, MBL2, HGFAC, C8B, and ANGPTL3 were identified as hub genes and C8A, SPP2, PROZ, C6, HGFAC, and C8B were found to be significant for survival.

Conclusion: The DEGs re-analyzed between HCC and hepatitis B enable a systematic understanding of the molecular mechanisms of HCC reliant on hepatitis B virus.

Abbreviations: BP = biological processes, CC = cellular components, CTD = comparative toxicogenomics database, DEGs = differentially expressed genes, GEPIA = gene expression profiling interactive analysis, GO = gene ontology, HCC = Hepatocellular carcinoma, KEGG = Kyoto Encyclopedia of Genes and Genomes, MF = molecular functions, OS = overall survival, PPI = protein–protein interaction.

Keywords: bioinformatical analysis, differentially expressed genes, hepatitis B virus, hepatocellular carcinoma

1. Introduction

Hepatocellular carcinoma (HCC) is the fifth most common cancer and the third most common cause of cancer-related death and primary hepatic malignancy in the world. HCC results in more than 600,000 deaths per year because it invades the liver vasculature and undergoes metastasis.\(^1\) Numerous studies have identified risk factors contributing to the development of HCC; one of the most common risk factors is chronic viral hepatitis, especially hepatitis B.\(^1\)\(^,\)\(^3\) It has been reported that patients with early stage HCC with small, solitary tumors and preserved liver function have a high 10-year overall survival (OS) following resection, liver transplantation, or ablation.\(^3\)\(^,\)\(^5\) However, most patients who were diagnosed with HCC at an advanced stage, when it is not suitable for curative therapies, have a low overall survival (OS).\(^1\)\(^,\)\(^4\)\(^,\)\(^5\) Thus, it is necessary to identify appropriate molecular biomarkers for diagnosis of HCC at an early stage and potential targets for therapies.

In recent years, bioinformatics analysis has been commonly used to screen and identify key biomarkers and potential molecular mechanisms of some cancers due to advances in microarrays based on high-throughput platforms, which could
lead to improvements in 10-year OS and prognosis. The gene expression omnibus (GEO) online database is a public repository available worldwide for high-throughput microarray and next-generation sequence functional genomic data sets submitted by the research community. We downloaded and re-analyzed 2 original microarray datasets, GSE44074 and GSE55092, from the GEO database and found biomarkers and disease mechanisms that could prove valuable for future research.

In our study, we searched for DEGs using GEO2R tools. The Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene ontology (GO) analyses of overlapping genes among the 2 datasets were performed using the Metascape database. PPI networks were generated using the STRING online database and analyzed via Cytoscape. Cytoscape was also used to search for hub genes. The 10-year OS associated with hub genes were analyzed using the gene expression profiling interactive analysis (GEPIA) database. The identification of hub genes associated with malignancy was performed using the CTD database. Finally, microRNA (miRNA)-gene pairs of the hub genes were screened using the TargetScan database. The results of this will help us understand the molecular mechanisms underlying HCC induced by hepatitis B virus and improve early detection.

2. Material and methods

2.1. Access to public data

The GEO database is a publicly accessible, functional genomics data repository supporting MIAME-compliant data submissions. We obtained the gene expression profiles GSE44074 (GPL13536, Kanazawa Univ. Human Liver chip 10k) and GSE55092 (GPL570, [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array). The probes were transformed into homologous gene symbols using the platform’s annotation information. GSE44074 contained 17 HCC specimens (hepatitis B virus-related) and 36 chronic hepatitis B specimens, while GSE55092 contained 39 specimens (hepatitis B virus-related) and 81 chronic hepatitis B specimens.

2.2. Identification of DEGs

GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) is a web application for identifying DEGs by comparing samples from GEO series. In our study, DEGs between HCC and hepatitis B were searched for using GEO2R. The cut-off P value was .05 and |logFC| was more than 1.

2.3. GO and KEGG analysis of overlapping genes

GO is a comprehensive source of digital data relating to the functions of genes, spanning 3 aspects of biology: BP (biological processes), CC (cellular components), and MF (molecular functions). KEGG is a database for both the functional interpretation and the practical application of genomic information, which can be used to assign functional meanings to genes and genomes, both at the molecular level and above. Metascape (http://metascape.org/) is a widely used, web-based tool with functional annotations of DEGs. In our study, the GO and KEGG analyses of overlapping genes were performed using Metascape (P < .05).

2.4. Construction and analysis of the PPI network

STRING (http://string-db.org) is an online database that can predict and generate a protein–protein interaction (PPI) network after overlapping genes have been imported. Cytoscape is open-source software used to analyze PPI networks and search for hub genes. In our study, the PPI network was generated by STRING and analyzed using Cytoscape.

2.5. Survival analysis

GEPIA (http://gepia.cancer-pku.cn/) is an online tool based on the TCGA and GTEx databases and can provide key interactive and customizable functions, such as differential expression analysis, correlation analysis, patient survival analysis, profile plotting, and dimension reduction analysis. In our study, the effects of hub genes on the 10-year OS of patients with HCC were analyzed using GEPIA (P < .05).

2.6. Identification of hub genes associated with malignancy

The comparative toxicogenomics database (CTD) (http://ctdbase.org/) is a web-based tool that provides information about interactions between chemicals and gene products and their connection to diseases. The relationships between hub genes and tumors were analyzed using CTD.

2.7. Identification of miRNA-gene pairs

TargetScan (www.targetscan.org) is a web-based database that predicts the biological targets of miRNAs. In our study, the miRNA-gene pairs of the hub genes were screened using TargetScan.

2.8. Ethical approval

All data in this study were obtained from open, public databases, therefore ethical approval was not necessary.

3. Results

3.1. Identification of DEGs

The screening of GSE44074 identified 394 DEGs, 247 genes which were up-regulated and 147 which were down-regulated (Fig. 1A). Similarly, we obtained 1809 DEGs from GSE55092, 1021 genes which were up-regulated and 788 which were down-regulated (Fig. 1C). There were 113 overlapping genes between the two datasets (Fig. 2A). The heat maps of the datasets are shown in Figure 1B and D.

3.2. Functional and pathway enrichment analysis

GO analysis of the overlapping genes showed they were mainly enriched in cellular amino acid catabolic processes, vitamin binding, terpenoid metabolic processes, animal organ regeneration, monocarboxylic acid metabolic processes, responses to bacteria, complement activation lectin pathway, cellular responses to cadmium ions, peptidase inhibitor activity, responses to acid chemical, toxin biosynthetic processes, cellular responses to potassium ion starvation, cellular responses to interferon-gamma, DNA topological changes, cofactor metabolic processes, blood microparticles, and protein activation cascade (Fig. 2E).

KEGG analysis of overlapping genes showed they were mainly enriched in the Citrate cycle (TCA cycle), AMPK signaling pathway, Leukocyte transendothelial migration, Caffeine metabolism, Chemical carcinogenesis, Glycine, Serine and
threonine metabolism, Phenylalanine metabolism, and Complement and coagulation cascades (Fig. 2F).

3.3. PPI network construction and searching for hub genes

The PPI network of DEGs was constructed using the STRING database (Fig. 2B). The top ten hub genes were C8A, SPP2, KLKB1, PROZ, C6, FETUB, MBL2, HGFAC, C8B, and ANGPTL3 (Fig. 2C,D). All hub genes were down-regulated.

3.4. Survival analysis

The OS rate analysis of the hub genes showed that C8A, SPP2, PROZ, C6, HGFAC, and C8B were significant for survival (Fig. 3). All of these genes were positively associated with the OS of patients with HCC ($P < .05$ Table 1).

3.5. Identification of hub genes

The identification of hub genes associated with malignancy was performed using the CTD database. The results are shown in Figure 4.

3.6. Prediction of miRNAs that regulate hub genes

The miRNAs that regulate hub genes were screened using TargetScan (Table 2).

4. Discussion

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. There are many ways in which HCC is treated, such as radiotherapy, embolization, and chemotherapy. However, conventional treatment regimens cannot significantly prolong the median OS of patients with advanced HCC, and there is a high frequency of tumor recurrence following these treatments. Surgery at an early stage is the most effective method for the successful treatment of HCC. Recently, biotherapies such as molecule-targeted therapy and immunotherapy have attracted great attention with respect to their use in standardized treatments for HCC. Thus, it is necessary to identify appropriate molecular biomarkers to enable diagnosis at an early stage as well as potential targets for HCC therapies.

HBV infection is the most common chronic viral infection in the world and is a historically challenging disease that primarily affects populations from developing or low-income countries. The most recent research indicates that approximately two billion people have been infected with HBV and more than 600,000 individuals worldwide die each year of HBV-related diseases, such as chronic hepatitis B, liver cirrhosis, and hepatocellular carcinoma. Numerous studies have shown that HBV
infection is a major risk factor for the development of HCC, with 50% of newly diagnosed HCC cases attributed to HBV infection because of the direct and indirect oncogenic effects of the virus. Surgery at an early stage could improve the long-term survival of patients. However, the early diagnosis of HCC in patients with HBV is challenging. Thus, it is critical to develop ways to enable the early detection of HCC by exploring the underlying molecular mechanisms linking HCC with HBV and to improve the survival rate and prognosis of patients in this situation.

In our study, we obtained two gene expression file datasets, GSE44074 and GSE55092, from the GEO database; we identified 113 overlapping genes between these datasets. Based on the GO analysis of the overlapping genes, monocarboxylic acid metabolic processes had the highest enrichment score; these processes are associated with aerobic glycolysis and lactate efflux. The enhancement of aerobic glycolysis and lactate efflux could provide tumor cells with a metabolic advantage and an invasive phenotype in the acidic microenvironment of a tumor. Cellular amino acid catabolic processes are associated with the metabolism of amino acids. Glutamine, glutamate, proline, and glycine are metabolic regulators involved in supporting cancer cell growth and metastasis. It has been reported that “vitamin binding” is associated with HCC, such as the vitamin B12 (B12)-binding protein, haptocorrin, which could be a potential biomarker in patients with HCC. Animal organ regeneration is associated with tissue remodeling and integrating growth, processes which underlie cancer metastasis. Peptidase inhibitor activity impacts tumor progression. For example, secretory leukocyte protease inhibitor (SLPI) could inhibit elastase, serine proteases, and matrix metalloproteinases, which could influence tumor invasion. Cellular responses to interferon-gamma (IFN-gamma) play a key role in innate and adaptive immune responses, and it has been shown that IFN-gamma could prevent the development of primary and transplanted tumors. Finally, the blood microparticle and protein activation cascade is associated with abnormalities in laboratory coagulation tests of patients with cancer because of tumor-shed procoagulant microparticles.

For the KEGG analysis of overlapping genes, chemical carcinogenesis had the highest enrichment score, which could induce malignant tumors. The AMPK signaling pathway is involved in the process of tumor invasion and migration in the liver, because liver kinase B1 (also known as serine/threonine kinase 11) phosphorylates and activates the AMPK signaling pathway. The citrate cycle (TCA cycle) is a central pathway...
for oxidative phosphorylation in cells. It has been reported that cancer cells with deregulated oncogene and tumor suppressor expression rely on the TCA cycle for energy production and macromolecule synthesis.[30] Leukocyte transendothelial migration could breach structural and cellular boundaries, which could promote cancer cell invasion and metastasis.[31] Finally, glycine, serine, and threonine metabolism and phenylalanine metabolism, together, provide essential precursors for the synthesis of proteins that are crucial to cancer cell growth. Moreover, serine could regulate tumor homeostasis by affecting cellular antioxidative capacity.[32,33]

Protein–protein interaction (PPI) analysis was conducted using the STRING online database and provided detailed connections between overlapping genes. In our study, the top ten hub genes were C8A, SPP2, PROZ, C6, HGFAC, and C8B. C6 (complement C6), C8A (complement C8 alpha chain), and C8B (complement C8 beta chain) encode components of the complement cascade. The complement system is part of the innate immune system, which is generally recognized to be a protective mechanism against the formation of tumors in humans.[34] SPP2 (secreted phosphoprotein 2) encodes a member of the cystatin superfamily. KLKB1 (kallikrein B1) encodes a glycoprotein that participates in the surface-dependent activation of blood coagulation, fibrinolysis, kinin generation, and inflammation. It has been reported that KLKB1 may function as a tumor suppressor gene in lung cancers.[35] PROZ (protein Z) encodes a vitamin K-dependent glycoprotein that is synthesized in the liver and secreted into the plasma; it is associated with oral squamous cell carcinoma and pancreatic cancer.[36,37] FETUB (fetuin B) encodes a member of the fetuin family and may potentially be useful in the treatment of diseases characterized by excess angiogenesis, such as cancer.[38]

Table 1
A summary of hepatocellular carcinoma microarray datasets from different GEO datasets.

| Series   | Platform   | Affymetrix Gene Chip | Samples |
|----------|------------|----------------------|---------|
| 1        | GSE44074   | GPL13536             | Kanazawa Univ. Human Liver chip 10k |
| 2        | GSE55062   | GPL570               | HG-U133_Plus_2 Affymetrix Human Genome U133 Plus 2.0 Array |

Figure 3. Kaplan–Meier curves of 6 common DEGs (C8A, SPP2, PROZ, C6, HGFAC, and C8B). These genes were all correlated with OS in patients with hepatocellular carcinoma. Data were analyzed using the GEPIA online database. Patients with expression above and below the median are indicated by the red and black lines, respectively. (HR = hazard ratio). (A) C8A (B) SPP2 (C) PROZ (D) C6 (E) HGFAC (F) C8B.
MBL2 is an important element in the innate immune system, which plays a key role in the immune surveillance against malignancies. Genetic variations in MBL2 could influence the risk of developing cancer. HGFAC (HGF activator) encodes a member of the peptidase S1 protein family. HGFAC has been shown to be involved in liver and kidney cancers, and the low expression of HGFAC is significantly associated with short OS time in patients with liver cancer.[39] ANGPTL3 (angiopoietin-like 3) encodes a member of a family of secreted proteins that function in angiogenesis; this family of proteins plays an important mechanistic role in cancer initiation, progression, and eventual prognosis.[40]

In conclusion, we obtained genes that were overlapping between HCC and hepatitis B from two datasets, and which may play an important role in the development of HCC. These new biomarkers, such as HGFAC, could be used to diagnose...
HCC at an early stage. The GO and KEGG analyses revealed further molecular mechanisms underpinning the regulation of HCC induced by HBV. Together, these findings may lead to more effective prevention, detection, and treatment of HCC.

Author contributions
Conceptualization: Hongyuan Cui
Data curation: Jinghai Song
Formal analysis: Jian Chen
Investigation: Junmin Wei
Methodology: Hongyuan Cui, Zhe Li
Writing – original draft: Zhe Li, Jingyong Xu
Writing – review and editing: Zhe Li

References
[1] Jiang JF, Lao YC, Yuan BH, et al. Treatment of hepatocellular carcinoma with portal vein tumor thrombus: advances and challenges. Oncotarget 2017;8:33911–21.
[2] Hartke J, Johnson M, Ghabril M. The diagnosis and treatment of hepatocellular carcinoma. Semin Diagn Pathol 2017;34:153–9.
[3] Hemming AW, Berumen J, Mekeel K. Hepatitis B and hepatocellular carcinoma. Clin Liver Dis 2016;20:703–20.
[4] Grandhi MS, Kim AK, Ronnekleiv-Kelly SM, et al. Hepatocellular carcinoma: from diagnosis to treatment. Surg Oncol 2016;25:74–85.
[5] Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. Lancet 2012;379:1245–55.
[6] Rocken C, Cärli-McGrath S. Pathology and pathogenesis of hepatocellular carcinoma. Dig Dis 2003;19:269–78.
[7] Guan YS, Ahmad Al-Shatouri M, He Q, et al. Hepatocellular carcinoma: carcinogenesis, establishment, progression, and therapies. Biomed Res Int 2014;2014:706142.
[8] Xu JM. Current treatment in advanced hepatocellular carcinoma and prospects for immuno-oncology therapy. Zhonghua Zhong Liu Za Zhi 2017;39:561–5.
[9] Tabrizian P, Jibara G, Shragger B, et al. Recurrence of hepatocellular cancer after resection: patterns, treatments, and prognosis. Ann Surg 2018;268:947–55.
[10] Bruix J, Gores GJ, Mazzaferro V. Hepatocellular carcinoma: clinical frontiers and perspectives. Gut 2014;63:844–53.
[11] Zhu ZX, Huang J, Liao Z, et al. Treatment strategy for hepatocellular carcinoma in China: radiofrequency ablation versus liver resection. Jpn J Clin Oncol 2016;46:4075–80.
[12] Li S, Yang F, Ren X. Immunotherapy for hepatocellular carcinoma. Drug Discov Ther 2015;9:363–71.
[13] Prieto J, Melero I, Sangro B. Immunological landscape and immunotherapy of hepatocellular carcinoma. Nat Rev Gastroenterol Hepatol 2015;12:681–700.
[14] Zampino R, Boemio A, Sagnelli C, et al. Hepatitis B virus burden in developing countries. World J Gastroenterol 2015;21:11941–53.
[15] Trepo C, Chan HL, Lok A. Hepatitis B virus infection. Lancet 2014;384:2053–63.
[16] Sunbul M. Hepatitis B virus genotypes: global distribution and clinical importance. World J Gastroenterol 2014;20:5427–34.
[17] Van Hees S, Michielsen P, Vanwolleghe T. Circulating predictive and diagnostic biomarkers for hepatitis B virus-associated hepatocellular carcinoma. World J Gastroenterol 2016;22:8271–82.
[18] Li X, Yu X, Dai D, et al. The altered glucose metabolism in tumor and a tumor acidic microenvironment associated with extracellular matrix metalloproteinase inducer and monocarboxylate transporters. Oncotarget 2016;7:23141–55.
[19] Zhang L, Han J. Branched-chain amino acid transaminase 1 (BCAT1) promotes the growth of breast cancer cells through improving mTOR-mediated mitochondrial biogenesis and function. Biochem Biophys Res Commun 2017;486:224–31.
[20] Porporato PE, Payen VL, Rasetl B, et al. Metabolic changes associated with tumor metastasis, part 2: Mitochondria, lipid and amino acid metabolism. Cell Mol Life Sci 2016;73:1349–63.
[21] Li Z, Zhang H. Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. Cell Mol Life Sci 2016;73:177–92.
[22] Simonsen K, Rode A, Nicoll A, et al. Vitamin B1(2) and its binding proteins in hepatocellular carcinoma and chronic liver diseases. Scand J Gastroenterol 2014;49:1096–102.
[23] Trusolino L, Bertotti A, Comoglio PM. MET signalling: principles and functions in development, organ regeneration and cancer. Nat Rev Mol Cell Biol 2010;11:834–48.
[24] Wen J, Nikitakis NG, Chausapat R, et al. Secretory leukocyte protease inhibitor (SLPI) expression and tumor invasion in oral squamous cell carcinoma. Am J Pathol 2011;178:2866–78.
[25] Ikeda H, Old LJ, Schreiber RD. The roles of IFN gamma in protection against tumor development and cancer immunoediting. Cytokine Growth Factor Rev 2002;13:95–109.
[26] Falanga A, Russo L, Milesi V, et al. Mechanisms and risk factors of thrombosis in cancer. Crit Rev Oncol Hematol 2017;118:79–83.
[27] Poirier MC. Linking DNA adduct formation and human cancer risk in chemical carcinogenesis. Environ Mol Mutagen 2016;57:499–507.
[28] Oliveira PA, Colaco A, Chaves R, et al. Chemical carcinogenesis. An Acad Bras Cienc 2007;79:593.
[29] Li N, Huang D, Lu N, et al. Role of the LKB1/AMPK pathway in tumor and metastasis in human cancer. J Carcinog 2017;39:561.
[30] Anderson NM, Mucka P, Kern JG, et al. The emerging role and targetability of the TCA cycle in cancer metabolism. Protein Cell 2018;9:216–37.
[31] Schaefer A, Hordijk PL. Cell-stiffness-induced mechanosignaling - a key driver of leukocyte transendothelial migration. J Cell Sci 2015;128:2221–30.

[32] Amelio I, Cutruzzola F, Antonov A, et al. Serine and glycine metabolism in cancer. Trends Biochem Sci 2014;39:191–8.

[33] Locasale JW. Serine, glycine and one-carbon units: cancer metabolism in full circle. Nat Rev Cancer 2013;13:572–83.

[34] Mamidi S, Hone S, Kirschfink M. The complement system in cancer: Ambivalence between tumour destruction and promotion. Immunobiology 2017;222:45–54.

[35] Wong J, Sia YY, Misso NL, et al. Effects of the demethylating agent, 5-azacytidine, on expression of the kallikrein-kinin genes in carcinoma cells of the lung and pleura. Patholog Res Int 2011;2011:167046.

[36] Wu X, Zhang ZX, Chen XY, et al. A panel of three biomarkers identified by iTRAQ for the early diagnosis of pancreatic cancer. Proteomics Clin Appl 2019;13:e1800195.

[37] Vylliotis A, Yapijakis C, Nkenke E, et al. Effect of thrombosis-related gene polymorphisms upon oral cancer: a regression analysis. Anticancer Res 2013;33:4033–9.

[38] Lee C, Bongcam-Rudloff E, Sollner C, et al. Type 3 cystatins; fetuins, kininogen and histidine-rich glycoprotein. Front Biosci (Landmark Ed) 2009;14:2911–22.

[39] Yin L, Mu Y, Lin Y, et al. HGFAC expression decreased in liver cancer and its low expression correlated with DNA hypermethylation and poor prognosis. J Cell Biochem 2019;120:9692–9.

[40] Jiang S, Qu H, Zhu N, et al. ANGPTL3: a novel biomarker and promising therapeutic target. J Drug Target 2019;27:876–84.