Screening and Identification of Key Genes in Hepatitis B Virus-Related Hepatocellular Carcinoma Through an Integrated Bioinformatics Approach

Thanh Thi Nguyen1*, Nga Phuong Vu1*, Trung Sy Tran2, Hieu Van Hoang3, Hoai Thi-Thu Bui1,4

1Department of Tropical and Infectious Diseases, E Hospital, Hanoi, Vietnam. 2Vietnam Military Medical University, Hanoi, Vietnam. 3Military Institute of Traditional Medicine, Hanoi, Vietnam. 4School of Medicine and Pharmacy, Vietnam National University, Hanoi, Vietnam.

Abstract

Objective: Primary liver cancer is one of the main causes of cancer mortality globally, with hepatocellular carcinoma (HCC) being the most frequent type. Chronic hepatitis B virus (HBV) infection is leading cause of HCC. This study aimed to identify significant genes for predicting prognosis in HBV-associated HCC.

Methods: The GSE121248 gene expression profile was obtained from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) for HBV-associated HCC were identified by analyzing this expression profile. Enrichment analyses were performed to discover the role of DEGs in biological processes, cell components, molecular functions, and pathways. Then, protein-protein interaction (PPI) was constructed and 5 hub genes were identified. Finally, survival analysis was conducted to validate the prognostic value of these genes. Results: A total of 20188 official gene symbols were found, and 119 DEGs were identified between HBV-associated HCC and normal liver tissues. The PPI network identified CCNB1, CDK1, TOP2A, RACGAP1, and ASPM as hub genes. Kaplan-Meier curves showed that the high expression of the hub genes had significantly lower survival. Conclusion: CCNB1, CDK1, TOP2A, RACGAP1, and ASPM could be potential prognostic biomarkers and therapeutic targets for HBV-associated HCC.

Keywords: Hepatitis B virus- Hepatocellular carcinoma- Bioinformatics- Hub genes- GEO
outcome of HCC patients.

Gene profiling and signatures, which can swiftly find differentially expressed genes (DEGs), have substantially expedited cancer research during the last few decades. Several research have looked at the prognostic implications of array-based genes from HCC, however, few have found gene signatures that indicate poor prognosis for HBV-associated HCC. Thus, the aim of this study was to identify significant genes for predicting prognosis in HBV-associated HCC utilizing publicly available data and integrated bioinformatics tools.

Materials and Methods

Data collection

For this research, we chose Gene Expression Omnibus (GEO), a publicly available collection of gene/microarray profiles. This public database has been used to conduct bioinformatics research on many types of cancers, such as breast cancers [9, 10], lung cancers [11, 12], gastrointestinal cancers [13, 14], and bone cancers [15, 16].

The GSE121248 gene expression profile was obtained from the GEO database repository. This dataset used the platform GPL570 (HG-U133 Plus 2) Affymetrix Human Genome U133 Plus 2.0 Array for the mRNA expression profiling. A total of 70 HCC samples and 37 non-tumor tissue samples were collected from chronic HBV-associated HCC and their adjacent normal tissues in the dataset [17].

Identification of differentially expressed genes (DEGs)

First, the dataset was annotated and standardized by quantiles. Then, the limma package of the R software was used to screen the DEGs. The threshold criterion of DEGs was $|\log FC|>2$ and $p<0.05$. The Principal Components Analysis (PCA) was visualized by using the FactoMineR and factoextra packages. The M-versus-A (MA) plot, volcano plot, and heatmap of DEGs were plotted using the ggpubr and pheatmap packages.

Enrichment analysis

Functional analysis of the DEGs was conducted to further investigate the biological processes and signal pathways in which the DEGs may be engaged. Using the clusterProfiler package, Gene Ontology (GO) analysis was performed to demonstrate the dominating function of the DEGs from molecular function and biological process. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was used to determine the relationship between DEGs and signal pathways. In this enrichment analysis, all of the findings were visualized, and a statistically significant difference was defined as $P<0.05$.

Protein-protein interaction (PPI) and module analysis

The STRING platform was used to conduct PPI network analysis for differential genes [18]. The PPI of the DEGs was built using the STRING platform and analyzed using the Cytoscape program. Cytoscape is an open-source bioinformatics application for visualizing gene and protein molecular interaction networks. Moreover, the Cytoscape plug-in Cytohubba was used to identify the hub genes that are strongly associated with HBV-associated HCC [19].

Survival Analysis

The hub genes’ prognostic value was evaluated using the GEPIA database. GEPIA, an interactive web server, can perform interactive and configurable tasks such as differential expression analysis, profile charting, correlation analysis, patient survival analysis, related gene recognition, and dimensionality reduction analysis [20].

Results

Identification of DEGs

A total of 20188 official gene symbols were found, and each gene’s expression was determined. 119 DEGs were identified between HBV-associated HCC and normal liver tissues using the established criteria, comprising 30 up-regulated genes and 89 down-regulated genes. Figure 1 depicts the volcano map of all genes. The top 5 up-regulated genes are CAP2, TOP2A, RACGAP1, ASPM and COL15A1, the top 5 down-regulated genes are CXCL14, IGFALS, CLEC1B, HHIP and CDHR2. Figure 2 shows the cluster heatmap of the DEGs.

GO and KEGG pathway enrichment analysis

To do GO functional annotation and KEGG pathway enrichment analysis, we utilized the clusterProfiler R package. Figure 3 illustrates the findings of the substantial enrichment analysis. GO analysis results showed that changes in the biological process of DEGs were significantly enriched in regulation of hormone levels, response to xenobiotic stimulus, steroid metabolic process, and cellular hormone metabolic process (Figure 3A). Changes in cell component were mainly enriched in collagen-containing extracellular matrix, blood microparticle, midbody, collagen trimer, and immunoglobulin complex (Figure 3B). Figure 3C showed the changes in molecular function of DEGs, which were mainly enriched in oxidoreductase activity, monoxygenase activity, iron ion binding, heme binding, and tetrapyrrole binding. The results of KEGG pathway analysis (Figure 3D) revealed that DEGs were mostly enriched in retinol metabolism, drug metabolism - cytochrome P450, p53 signaling pathway, metabolism of xenobiotics by cytochrome P450, and steroid hormone biosynthesis.

Construction of PPI Network and identifying of hub genes

The STRING database generated a PPI network with 79 nodes and 212 edges. The PPI was then displayed in Cytoscape to investigate the functional relationships between DEGs (Figure 4A). Following that, as demonstrated in Figure 4B, the top five hub genes found in the Cytohubba plug-in using the Maximal Clique Centrality algorithm were CCNB1, CDK1, TOP2A, RACGAP1, and ASPM.
Discussion

Despite significant advances in clinical management and pathophysiology prediction for HCC in past few years, the high rate of HCC-specific death remains a significant problem [21]. Chronic HBV infection serves as the most common cause of HCC. In endemic locations, particularly in underdeveloped nations, HBV infection is mostly transmitted vertically. As a result, the typical age of HBV carriers developing HCC is younger than for other...
etologies. Furthermore, the majority of patients acquire liver fibrosis and cirrhosis as a result of immunological responses during HBV infection, which promotes the development of HCC. We hoped to get new insights into the molecular mechanisms underpinning HBV-associated HCC formation and progression by using bioinformatics analyses.

This research discovered 119 DEGs including 30 up-regulated genes and 89 down-regulated genes, these DEGs are involved in retinol metabolism, drug metabolism - cytochrome P450, p53 signaling pathway, xenobiotic metabolism via cytochrome P450, and steroid hormone biosynthesis. Furthermore, five hub genes (CCNB1, CDK1, TOP2A, RACGAP1, and ASPM) were verified and found to be related to poorer outcomes in HCC populations with HBV infection.

CCNB1 encoding cyclin B1, a regulatory protein that is important in the mitosis process, is a key member of the conserved cyclin B family [22]. The abnormal activity of this gene can disorganize the cell cycle and cell proliferation, resulting in triggering oncogenesis and cancer development [23, 24]. CCNB1 was shown that substantially related to development, proliferation, migration, and invasion in HCC tumors [25]. Peng et al. [26] indicated that the miR-16/cyclin-B1 axis is the target of zingiberene in the regulation of the growth and invasion of liver tumor cells. Another study also suggests that cyclin B1 may possibly become a biomarker and therapeutic target for HBV-associated HCC [27].

CDK1 is a serine-threonine protein kinase encoding cyclin dependent kinase 1. The abnormal expression of cyclin dependent kinase 1 strongly correlates with carcinogenesis due to its important role in cell mitosis. The antimalarial medicine dihydroartemisinin suppresses the growth of liver cancer cells by lowering CDK1 and CCNB1 expression levels [28]. Wu et al. [29] demonstrated that anti-CDK1 therapy can improve sorafenib anti-tumor response in HCC patient-derived xenograft tumor models by inhibiting CDK1/PDK1/β-Catenin signaling.

TOP2A encodes DNA topoisomerase II alpha, an enzyme that regulates and modifies the topologic states of DNA during transcription. This gene is a target for various anticancer drugs, and a number of mutations in this gene have been linked to the development of treatment resistance [30-32]. TOP2A expression level was shown to be higher in HBV-associated HCC tissues and resulting in a worse prognosis. The findings were consistent with earlier research. Furthermore, the expression of TOP2A in HCC tumors seemed to be associated with the presence of serum HbsAg [33], this might explain why TOP2A was one of the hub genes in HBV-associated HCC.

RACGAP1 encodes a GTPase-activating protein that is part of the centralspindlin complex that regulates...
Screening and Identification of Key Genes in Hepatitis B Virus-Related Hepatocellular Cytokinesis, Cell Proliferation, and Differentiation. RACGAP1 has been found as a prognostic biomarker for early HCC identification as well as a therapeutic target for liver cancer and liver cancer stem cells [34]. Wang et al. [35] discovered that the pseudogene RACGAP1P acts as a ceRNA to promote the RACGAP1/Rho/ERK signaling axis, leading to early relapse in HCC. Especially, in patients with HBV-associated HCC, high RACGAP1 expression was linked with an increased risk of mortality [36].

ASPM is the human ortholog of the Drosophila melanogaster abnormal spindle gene, which is required for...
proliferation and mitotic spindle activity in embryonic neuroblasts. It has been shown to play a role in tumor development. The overexpression of the gene was proven to be a genetic marker indicating HCC’s increased invasive/metastatic potential, a greater likelihood of early tumor recurrence independent of p53 mutation status or tumor stage, and hence a poor prognosis [37]. The N6-methyladenosine modification of ASPM mRNA mediated by METTL3 promoted its expression in liver HCC [38]. ASPM also promotes tumor growth by antagonizing autophagy-mediated Dvl2 degradation and boosting Wnt/-catenin signaling [39].

This study had several following limitations. First, we only used a gene expression profile from the GEO database; this dataset just covers mRNA expression and does not include microRNA, lncRNA, or circRNA expression. This maybe narrows the identification of other potential biomarkers in HBV HCC. Furthermore, additional experiments, both in vivo and in vitro, are required to demonstrate the predictive usefulness of these hub genes.

In conclusion, in the present research, 119 DEGs were found in HBV-associated HCC. Five hub genes, CCNB1, CDK1, TOP2A, RACGAP1, and ASPM, have been identified as potential prognostic biomarkers and therapeutic targets for HBV-associated HCC.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors declare no conflict of interest.

References


