Analysis of expression profile data identifies key genes and pathways in hepatocellular carcinoma

XUWEI JIANG and YUQING HAO

Department of General Surgery, Shanghai, Baoshan District Hospital of Integrated Traditional and Western Medicine, Shanghai 201900, P.R. China

Received July 9, 2016; Accepted July 20, 2017

DOI: 10.3892/ol.2017.7534

Abstract. The aims of the present study were to identify key genes and pathways associated with hepatocellular carcinoma (HCC) progression and predict compounds potentially associated with this type of carcinogenesis. The gene expression profile data of the GSE49515 dataset was obtained from the Gene Expression Omnibus database. The limma software package was used to identify the differentially expressed genes (DEGs). Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analyses were performed using the Biological Networks Gene Ontology tool and the Database for Annotation, Visualization and Integrated Discovery, respectively. The Michigan Molecular Interactions database plugin within the Cytoscape software platform was used to perform protein-protein interaction (PPI) network analysis. Chemical-gene interaction data for HCC were obtained from the Comparative Toxicogenomics Database to evaluate the associations between drugs and specific genes. A total of 302 DEGs, including 231 downregulated and 71 upregulated, were identified. Cytokine-cytokine receptor interaction and chemokine signaling were the significantly enriched pathways. Additionally, PPI network analysis indicated a total of 13 highest degree hub nodes, including FBJ murine osteosarcoma viral oncogene homolog (FOS) and DNA damage-inducible transcript 3 protein (DDIT3). Chemical-gene interaction analysis revealed that FUN and FOS were targeted by >500 compounds, while >200 genes were targeted by 2,3,7,8-tetrachlorodibenzo-p-dioxin and benzo(α)pyrene. In conclusion, the present study demonstrated that FOS, DDIT3, the cytokine-cytokine receptor interaction pathway and the chemokine signaling pathway may be key genes and pathways associated with the development of HCC. Furthermore, exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin or benzo(α)pyrene may lead to hepatocarcinogenesis.

Introduction

Hepatocellular carcinoma (HCC), a highly lethal malignancy, has an increasing worldwide incidence and is the third most frequent cause of cancer-associated mortality (1,2). It is estimated that >700,000 new cases are diagnosed each year (3). Infection with hepatitis B or C viruses, alcohol-associated cirrhosis and non-alcoholic steatohepatitis have been described as risk factors (4,5). In addition, the mortality rate almost equals the incidence rate in the majority of countries (4,5). Therefore, more research is required to identify novel effective treatments for HCC and to elucidate the underlying molecular mechanisms of HCC progression.

Extensive studies have been previously conducted and advances have been made in the identification of genes and pathways associated with HCC. Yong et al (6) demonstrated that spalt-like transcription factor 4 is associated with poor prognosis and may serve as a potential target for the therapy of HCC. Downregulation of microRNA-195 (miR-195) and miR-497 may affect molecular pathways associated with cell cycle progression, leading to the abnormal cellular proliferation observed in hepatocarcinogenesis (7). Cillo et al (8) demonstrated that the transcriptional and post-transcriptional deregulation of homeobox A13 may be involved in HCC through mRNA nuclear export of eukaryotic translation initiation factor 4E-dependent transcripts. Revill et al (9), using integrative genomic analysis, revealed that sphingomyelin phosphodiesterase 3 and neurofilament heavy polypeptide are tumor suppressor genes in HCC. Yoshikawa et al (10) reported that suppressor of cytokine signaling 1 and Janus kinase 2 may serve as potential therapeutic targets for the treatment of HCC. In addition, inhibition of the rapidly accelerated fibrosarcoma/mitogen-activated protein kinase (MAPK)/extracellular signal-regulated kinase pathway inhibited tumor angiogenesis and induced apoptosis in a HCC model (11). Xu et al (12) demonstrated that miR-122 induces apoptosis and inhibits cellular proliferation in HCC by directly targeting the Wnt/β-catenin signaling pathway. Furthermore, ubiquitin D, an oncogene located at 6q21.3,
promotes hepatitis B virus-associated HCC progression via the protein kinase B/glycogen synthase kinase 3β, signaling pathway (13). The tyrosine-protein kinase Met pathway is also involved in the pathogenesis of HCC (14). Additional studies indicated that a number of other pathways, including the hedgehog and the GTPase signaling pathways, are associated with the progression of HCC (15,16). Even though a number of genes and pathways have been associated with HCC, the underlying molecular mechanisms of HCC progression have not been completely identified yet. Therefore, further research is required.

In the present study, the expression profile of the GSE49515 dataset was obtained and analyzed, and the differentially expressed genes (DEGs) between HCC and healthy samples were identified. Furthermore, functional enrichment analysis and protein-protein interaction (PPI) network analysis were performed. In contrast to the previous studies of Shi et al (17) and Jiang et al (18), which used the same dataset to identify genes and pathways associated with HCC, the present study also investigated gene and drug interactions using the comparative toxicogenomics database (CTD). The aim of the present study was to identify key genes and pathways associated with HCC progression and investigate potential compounds leading to HCC carcinogenesis.

Materials and methods

Expression profile data. The GSE49515 dataset originally derived from the study by Shi et al (17) was obtained from the Gene Expression Omnibus database (www.ncbi.nlm.nih.gov/geo/). The dataset contains the expression profile of 26 samples of peripheral blood mononuclear cells, including 10 HCC samples, 3 pancreatic carcinoma samples, 3 gastric carcinoma samples and 10 healthy samples. In the present study, the 10 HCC and 10 healthy samples were used to analyze the mRNA expression profile of HCC. These data were analyzed using the GPL570 Affymetrix Human Genome U133 Plus 2.0 Array platform (Affymetrix; Thermo Fisher Scientific, Inc., Waltham, MA, USA).

Data preprocessing. The mRNA expression profile data were preprocessed using the Robust Multi-Array Average algorithm in the Affy software package (19), and subsequently, the expression matrix was generated. If several probes mapped to one gene symbol, then the mean value was set as the final expression value of this gene. A total of 20,108 genes were obtained.

DEG analysis. The limma software package (20) within Bioconductor was used to identify the DEGs in HCC samples compared with healthy samples. The P-values of DEGs were calculated using Student's t-test (21) in the limma software package and adjusted according to the Benjamini-Hochberg (BH) procedure (22). |log_{2}(fold-change)|≥1 and BH_P<0.05 were used as threshold criteria.

Functional enrichment analysis. The Biological Networks Gene Ontology tool (BiNGO) (23) is a Cytoscape plugin used to assess the overrepresentation of Gene Ontology (GO) terms in biological networks. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to assign related gene categories into their associated pathways (24). The Database for Annotation, Visualization and Integrated Discovery (DAVID), an integrated data-mining environment, was used for pathway enrichment analysis (25). GO annotation and KEGG pathway enrichment analysis were performed using BINGO and DAVID. P<0.05 was used as the threshold criterion.

PPI network analysis. The Michigan Molecular Interactions database (MiMI) plugin (26) within the Cytoscape software platform (v2.7.0) (27) was used to perform PPI visual network analysis. The MiMI plugin integrates data from several databases, including the Biomolecular Interaction Network Database, the Biological General Repository for Interaction Datasets, the Human Protein Reference Database, the Molecular Interaction Database, InterPro and SwissProt. The hub nodes were identified by calculating the degree value using the degree-sorted tool in Cytoscape. Highest degree hub nodes indicate key genes with important physiological regulation roles.

Gene and chemical interaction analysis. CTD (28,29) is a publicly available resource that provides manually curated information about chemical-gene interactions and chemical/gene-disease associations derived from microarray data and published literature. In addition, CTD provides chemical-gene interaction information for various diseases in vertebrates and invertebrates (30). The chemical-gene interaction data for HCC were obtained from the CTD database to investigate the therapeutic efficacy of several drugs.

Results

DEG analysis. A total of 302 DEGs, including 231 down-regulated and 71 upregulated, were identified in HCC samples compared with healthy samples.

Functional enrichment analysis. GO and KEGG pathway enrichment analyses were performed. The overrepresented GO terms (adjusted P<1.00x10^{-5}) were mainly associated with drug reactions, immune response and cellular processes associated with stress (Fig. 1). The most significantly enriched pathways were cytokine-cytokine receptor interaction, the chemokine signaling pathway, the Toll-like receptor signaling pathway and the MAPK signaling pathway (Table I).

PPI network analysis. The PPI network analysis was performed using the MiMI Cytoscape plugin. A total of 47 nodes, including 20 overexpressed and 27 underexpressed genes, were identified (Fig. 2). A total of 13 highest degree proteins, including toll-like receptor 1 (TLR1), TLR4, TLR7, TLR8, receptor-interacting serine/threonine kinase 2, FBJ murine osteosarcoma viral oncogene homolog (FOS), FOS-like antigen 2, fas ligand, chemokine ligand 4, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein γ, cyclin-dependent kinase inhibitor 1a, hypoxia inducible factor 1α and DNA-damage-inducible transcript 3
were identified as hub nodes and may potentially be associated with HCC carcinogenesis.

**Gene and chemical interaction analysis.** A total of 264 key genes, 1,030 small molecule compounds and 5,037 small molecule compounds and mRNA interaction association pairs were identified in the CTD database. Subsequently, the number of compounds targeting the same gene and the number of genes targeting the same compound were counted (Fig. 3A and B). A number of genes were targeted by several small molecule compounds. It was observed that >500 compounds were targeted by two genes (FUN, 611; ASLG, 611).

### Table I. Significantly enriched pathways of differentially expressed genes.

<table>
<thead>
<tr>
<th>KEGG term</th>
<th>Count</th>
<th>P-value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa04060: Cytokine-cytokine receptor interaction</td>
<td>23</td>
<td>1.31x10^-10</td>
<td>TNFSF4, CXCL5, IL8, CCR1, CXCL2, TNFRSF17, FASLG, CXCR2, PF4, PF4V1, CCL4, CXCL10, OSM, TNFRSF1A, TNFSF10, CCR6, CCR5, VEGFA, CX3CR1, IFNG, CCR2, IL1B, EGF, IL8, CXCL5, CCR1, CXCL2, PF4, CXCR2, GNG11, PF4V1, CCL4, CXCL10, CCR6, CCR5, CX3CR1, CCR2</td>
</tr>
<tr>
<td>hsa04062: Chemokine signaling pathway</td>
<td>14</td>
<td>1.07x10^-5</td>
<td>FOS, IL8, JUN, TLR1, IL1B, TLR4, TLR7, CCL4, TLR8, CXCL10</td>
</tr>
<tr>
<td>hsa04620: Toll-like receptor signaling pathway</td>
<td>10</td>
<td>3.96x10^-5</td>
<td>FOS, IL8, JUN, TLR1, IL1B, TLR4, TLR7, CCL4, TLR8, CXCL10</td>
</tr>
<tr>
<td>hsa04621: NOD-like receptor signaling pathway</td>
<td>6</td>
<td>3.55x10^-3</td>
<td>NLRC4, IL8, CXCL2, IL1B, RIPK2, CARD6</td>
</tr>
<tr>
<td>hsa04010: MAPK signaling pathway</td>
<td>11</td>
<td>1.29x10^-2</td>
<td>TNFRSF1A, FOS, PLA2G4A, DUSP2, JUN, IL1B, FASLG, EGF, GADD45B, DDIT3, DUSP6</td>
</tr>
<tr>
<td>hsa05120: Epithelial cell signaling in Helicobacter pylori infection</td>
<td>5</td>
<td>2.63x10^-2</td>
<td>IL8, JUN, HBEGF, CXCR2, ATP6V1B2</td>
</tr>
<tr>
<td>hsa05219: Bladder cancer</td>
<td>4</td>
<td>3.24x10^-2</td>
<td>CDKN1A, IL8, VEGFA, EGF</td>
</tr>
</tbody>
</table>

KEGG term indicates the identification number and the name of the KEGG overrepresented pathway. Count represents the number of genes enriched in each KEGG pathway. KEGG, Kyoto Encyclopedia of Genes and Genomes; hsa, Homo sapiens; MAPK, mitogen-activated protein kinase; NOD, nucleotide-binding oligomerization domain.

![Figure 1](image-url) Enriched GO terms for differentially expressed genes. Nodes represent the name of each GO term; the darker the color is, the smaller the adjusted P-value, and the arrows point to the sub-branches. GO, Gene Ontology.
Figure 2. Protein-protein interaction network analysis for differentially expressed genes. Nodes indicate proteins, edges indicate the interaction between nodes, light color indicates overexpressed proteins in HCC and dark color indicates underexpressed proteins in HCC. HCC, hepatocellular carcinoma.

Figure 3. Gene and chemical interaction analysis. (A) The number of genes targeted by compounds. (B) The number of compounds targeted by genes.

**Discussion**

HCC is the most common primary liver malignancy characterized by a multifaceted molecular pathogenesis (31). In the present study, a total of 302 DEGs, including 231 downregulated and 71 upregulated genes, were identified in HCC samples compared with healthy samples. Using KEGG pathway analysis, it was demonstrated that the cytokine-cytokine receptor interaction and chemokine signaling pathways were the major enriched pathways. In addition, a total of 13 highest degree proteins, including FOS and DDIT3, were identified as hub

**FOS**, 521) and >200 genes were targeted by two compounds (tetrachlorodibenzodioxin, 215; benzo(α)pyrene, 197).
nodes in PPI network analysis. Furthermore, >500 compounds were targeted by FUN and FOS, and >200 genes were targeted by 2,3,7,8-tetrachlorodibenzo-p-dioxin and benzo(a)pyrene in the analysis of gene and chemical interactions.

In the present study, FOS was identified as a hub node in the PPI network analysis, and >500 compounds were targeted by FOS in the analysis of gene and chemical interactions. One study using immunohistochemical analysis demonstrated that c-FOS serves a key role in HCC pathogenesis (32). The upregulation of c-FOS and Jun proto-oncogene mediated by protein kinase R promotes HCC proliferation (33). In addition, Fan et al (34) demonstrated that suppression of c-FOS, mediated by miR-193 downregulation, promotes HCC metastasis. miR-101 inhibits the expression of the FOS oncogene in HCC, suppressing hepatocyte growth factor-induced cellular migration and invasion (35). Additionally, Shen et al (36) reported that recombiant adeno-associated virus carrying Vastatin inhibited tumor metastasis and reduced the expression of phosphoenolpyruvate carboxykinase 1, jagged 2 and c-FOS in HCC, inhibiting the cellular metabolism, Notch and activator protein-1 signaling pathways, respectively. This suggests that FOS may serve a critical role in the progression of HCC.

One additional study demonstrated that FUS-DDIT3 and DDIT3 may serve an important role in the induction of a liposarcoma phenotype (37). Kåbjörn Gustafsson et al (38) reported a dual promoting and inhibiting role in the formation of liposarcoma morphology mediated by DDIT3. DDIT3 and lysine acetyltransferase 2A proteins regulate the expression of tumor necrosis factor receptor superfamily, member 10a (TNFRSF10A) and TNFRSF10B in endoplasmic reticulum-associated, stress-induced apoptosis in human lung cancer cells (39). In addition, long non-coding RNA HOXA genes cluster antisense RNA 2 promotes proliferation of gastric cancer via silencing the expression of p21, polo-like kinase 3 and DDIT3 (40). Furthermore, it has also been demonstrated that DDIT3 is associated with the development of several types of cancer, including myxoid liposarcoma and squamous cell carcinoma (41,42). Therefore, DDIT3 serves an important role in several cancer types. In the present study, DDIT3 was overexpressed in HCC samples and indicated to be a hub node in PPI network analysis, suggesting that DDIT3 may be associated with HCC.

A number of studies have also demonstrated that the cytokine-cytokine receptor interaction pathway may be involved in HCC carcinogenesis (43-46). In the study by Hsu et al (47), it was reported that the cytokine-cytokine receptor interaction pathway is associated with HCC. Furthermore, Ryschich et al (48) demonstrated that the chemokine signaling pathway is involved in liver carcinogenesis, while another study indicated its involvement in organ-specific metastatic growth of cancer cells (49). Accordingly, the present study demonstrated that the cytokine-cytokine receptor interaction and chemokine signaling pathways were the major enriched pathways in KEGG pathway analysis. Therefore, these pathways may be involved in the development of HCC.

Analysis of gene and chemical interactions in the present study revealed that >200 genes were targeted by 2,3,7,8-tetrachlorodibenzo-p-dioxin and benzo(a)pyrene. Previously, it has been demonstrated, using in vivo models, that 2,3,7,8-tetrachlorodibenzo-p-dioxin may cause cancer (50). 2,3,7,8-Tetrachlorodibenzo-p-dioxin promotes liver tumor growth via an aryl hydrocarbon and TNF-receptor-1 dependent manner (51). One study reported that benzo(a)pyrene inhibits cell adhesion and promotes cell migration and invasion in HCC (52). Furthermore, a case-control study demonstrated that exposure to benzo(a)pyrene increases the risk of developing HCC (53). Therefore, exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin or benzo(a) pyrene may be associated with hepatocarcinogenesis.

In conclusion, FOS, DDIT3, the cytokine-cytokine receptor interaction pathway and the chemokine signaling pathway may serve critical roles in the development of HCC. Exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin or benzo(a) pyrene may cause hepatocarcinogenesis. However, the lack of experimental verification and the relatively small sample size are major limitations of the present study. Therefore, further research is required to verify the present findings.

References


