Bicluster and pathway enrichment analysis related to tumor progression of hepatocellular carcinoma

S.-Y. WANG¹,², L.-Y. FENG¹,², Z.-Q. MENG¹,²

¹Department of Integrative Oncology, Fudan University Shanghai Cancer Center, Shanghai, China
²Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China

Abstract. – OBJECTIVE: Hepatocellular carcinoma is one of the most aggressive cancers with poor prognosis worldwide. Tumor progression remains a significant cause of high mortality in patients with hepatocellular carcinoma. However, the molecular mechanism underlying tumor progression of hepatocellular carcinoma has not been completely unraveled currently. The aim of this study was to gain insight into the molecular mechanisms of tumor progression of hepatocellular carcinoma.

MATERIALS AND METHODS: We performed microarray analysis on 24 tissue specimens obtained at the time of surgical resection or liver transplantation from 24 patients with hepatocellular carcinoma downloaded from the Gene Expression Omnibus database.

RESULTS: Our analysis indicated that several differentially expressed genes might play crucial roles in the progression of hepatocellular carcinoma, such as GADD45G, SPTBN1, CDC27, TP53 and INSIG1. GADD45G and SPTBN1 not only contribute to tumor progression in hepatocellular carcinoma, but also correlate with poor prognosis in esophageal squamous cell carcinoma and pancreatic cancer respectively. Furthermore, we performed pathway enrichment analysis and found enriched pathways, including “Proteasome”, “Alanine, aspartate and glutamate metabolism”, “TGF-beta signaling pathway”, “Wnt signaling pathway”, and so on.

CONCLUSIONS: Our findings confirmed the presence of multiple molecular alterations during tumor progression and indicated the differentially expressed genes might be involved in tumor progression through multiple pathways. Genes GADD45G and SPTBN1 might correlate with poor prognosis in hepatocellular carcinoma as has already been shown for other malignancies of the gastrointestinal tract.

Key Words: Hepatocellular carcinoma, Iterative signature algorithm, Differentially expressed genes, Pathway.

Introduction

Hepatocellular carcinoma (HCC) is one of the most aggressive cancers with poor prognosis in the world, especially in Asia and Africa¹. Tumor progression remains a significant cause of high mortality in patients with HCC. The molecular mechanism underlying tumor progression of HCC has not been completely unraveled currently due to the complexity and heterogeneity of this disease.

With the development of molecular biology techniques, the HCC pathogenesis and progression have been better understood. DNA microarray analysis, which monitors the expression levels of thousands of genes simultaneously, has been used as a global approach to investigate the physiological mechanisms analysis². Genomic expression profiling has been proven to be a useful tool in identifying novel pathological mechanisms in human cancer³.

In recent years, many researchers have identified and described significant genes and pathways in the HCC development. Genes such as GPC3⁴, TERT, STK15, PLA2³, HSP70⁵ and GSTT1⁷ were recognized as potential biomarkers for detection of early HCC. ADAM10⁸, CTHRC1⁹ were identified associated with tumor invasion in HCC. Pathways such as extracellular matrix (ECM) receptor interaction¹⁰, p53 pathway¹¹, wnt/β-Catenin pathway¹², mitogen-activated-protein-kinase (MAPK) pathway¹³, epidermal growth factor receptor (EGFR) were identified associated with HCC development.

Large sets of data, such as expression profiles from many samples, require analytic tools to reduce their complexity¹⁴. The iterative signature algorithm (ISA)¹⁵-¹⁷ is a biclustering method which typically tries to find blocks that are different from the rest of the matrix, for example,
the values covered by the bicluster are all above or below the background. Its input is a matrix and its output is a set of biclusters: blocks of the potentially reordered input matrix, that fulfill some predefined criteria. Since ISA is an unsupervised algorithm, it can conduct very well in finding modules even in the present of noise in the matrix.

The purpose of this study is to analyze the molecular mechanism of tumor progression in HCC using microarray analysis combined with bioinformatics techniques. We sought to identify the differentially expressed genes in the progression and to find enriched pathways for the therapeutic targets in HCC.

Materials and Methods

Microarray Data Selection and Preprocessing
The transcription profile of GSE9843 was downloaded from the Gene Expression Omnibus ( GEO) database (http://www.ncbi.nlm.nih.gov/geo/). These data are based on the Affymetrix GPL570 platform data ( Affymetrix Human Genome U133 plus 2.0 Array). Data from 24 HCC samples was used, in which 9 samples were Barcelona Clinic Liver Cancer (BCLC) stage A, 7 samples were BCLC stage B, and 8 samples were BCLC stage C.

Pathway Data

Analysis of Differentially Expressed Genes
The Limma19 package based on R20 and Bioconductor21 was conducted to identify DEGs in three BCLC stages. Benjamini-Hochberg ( BH) method22 was performed to control False Discovery Rate ( FDR) for multiple testing issue. The original expression datasets from all conditions were processed into expression estimates, and these were used for constructing a linear model.

Bicluster Analysis
The ISA15 is developed to find biclusters(or modules as most of the ISA papers call them) that have correlated rows and columns. More precisely, the rows in the bicluster need to be only correlated across the columns of the bicluster and vice versa. Bicluster analysis with ISA214 package was conducted to find enriched biclusters. This package allows to extract modules from data sets accroding to ISA method.

Gene Ontology (GO)-Enrichment Analysis
BINGO23 is an open-source Java tool to determine which GO terms are significantly overrepresented in a set of genes. We used the BINGO to indentify over-represented GO categories in biological process. Moreover, we performed GO enrichment analysis for each bicluster respectively, and selected the most significant GO term in each bicluster as its GO term annotation.

Pathway-Enrichment Analysis
DAVID24, a high-throughput and integrated data-mining environment, analyzes gene lists derived from high-throughput genomic experiments. We used DAVID to identify over-represented pathways with p values less than 0.05.

Results

Go-Enrichment Analysis of Genes in All Biclusters
For dataset GSE9843, we performed data preprocessing using R and Bioconductor. Total 21 biclusters were obtained after bicluster analysis with ISA2 package. GO enrichment analysis was performed for each bicluster and the most significant GO term in each bicluster was selected as its GO term annotation ( Table I). As seen in Table I, several biological processes were enriched, such as “immune response”2,3,12,14,16,19,20, “RNA processing”5,8, “mRNA metabolic process”6,11,15,17,18, “organic acid catabolic process”7 and so on.

Analysis of DEGs
By analyzing the microarray data from 9 BCLC stage A samples, 7 BCLC stage B samples, and 8 BCLC C stage samples, we identified 1548 DEGs that showed statistically significant differences among the 3 groups. 473 DEGs were identified between BCLC stage A samples and BCLC stage B samples, with 45% (213) up-regulation and 55% (260) down-regulation. 1056 DEGs with 42% (444) up-regulation and 58% (612) down-regulation were identified between
BCLC stage B samples and BCLC stage C samples, and 650 DEGs with 37% (243) up-regulation and 63% (407) down-regulated between BCLC stage A samples and BCLC stage C samples (Figure 1).

Analysis of Overlapping Genes

As shown in Figure 1, there were 9 overlapping genes which expressed differentially in all three type samples. The 9 DEGs include GADD45G, SPTBN1, CDC27, SLC25a27, QD-PD1, two of them are TP52, and two of them are INSIG1. We performed GO enrichment analysis to identify the functional annotation of the 9 DEGs. The results showed 5 enriched biological progresses, including “cellular response to sterol depletion”, “mitotic metaphase/ anaphase transition”, “ER-nuclear sterol response pathway”, “response to sterol deletion”, and “common-partner SMAD protein phosphorylation” (Table II).

Pathway-Enrichment Analysis

A total of 1548 DEGs were further analyzed using DAVID to find the most significant pathways related to the progression of HCC. p value less than 0.05 was chosen as the cut-off for this analysis. Finally, we identified 11 enriched pathways, including “Proteasome”, “Alanine, aspartate and glutamate metabolism”, “TGF-beta signaling pathway”, “Wnt signaling pathway” and so on (Table III).

Discussion

The recent development of cDNA microarray or cDNA chip technology, a high-throughput
method of monitoring gene expression, has made it possible to analyze the expression of thousands of genes at once\textsuperscript{25,26}. To gain insight into the molecular mechanism of tumor progression from BCLC stage A to BCLC stage C, we performed bicluster analysis of the cDNA microarray obtained from GEO on 24 HCC samples, and identified DEGs in three BCLC stages. Furthermore, we used DAVID to identify pathways enriched involving the DEGs.

In our results of bicluster analysis of the gene expression profile, we found that total 21 biclusters were enriched. The “immune response” was the most significant biological process. Immune response is any immune system process that functions in the calibrated response of an organism to a potential internal or invasive threat. As we all know, tumor progression is closely associated with angiogenesis, metastasis and invasion. What prevents the immune response from destroying the tumor? One explanation is that in cancer patients the immune response may not be robust enough\textsuperscript{27}. Tumors need oxygen and nutrients, which are provided by new blood vessels that permeate the tumor mass. Cancer cells activate the angiogenic switch by secreting vascular endothelial growth factor (VEGF)\textsuperscript{28} and acidic and basic fibroblast growth factor (FGF) 1/2. Metastasis and invasion involve primary tumor cells moving out of the tumor mass, invading adjacent tissue and travelling to distant sites. Some mechanisms through which this may occur involve a change in the expression of adhesion molecules\textsuperscript{27}. In a word, tumors make progressions when the immune response may not be robust enough.

The other significant biological process was “mRNA metabolic process”. mRNA metabolism is the chemical reactions and pathways involving mRNA, messenger RNA, which is responsible for carrying the coded genetic ‘message’, transcribed from DNA, to sites of protein assembly at the ribosomes. Altered mRNA metabolism is a feature of many cancers and loss of function of many tumor suppressors regulating cell proliferation, survival, and differentiation results from aberrant mRNA processing, nuclear export, and/or translation\textsuperscript{29}.

The total 9 overlapping genes play an important role in the progression of this disease, especially GADD45G and SPTBN1. GADD45G (growth-arrest and DNA-damage inducible, gamma) is a member of GADD45 family. It is located at the commonly deleted region 9q22 and in-

**Table II.** GO enrichment analysis of the overlapping DEGs.

<table>
<thead>
<tr>
<th>GO ID</th>
<th>Description</th>
<th>(p)-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>71501</td>
<td>Cellular response to sterol depletion</td>
<td>1.0482E-3</td>
<td>3.8574E-2</td>
</tr>
<tr>
<td>7091</td>
<td>Mitotic metaphase/ anaphase transition</td>
<td>1.0482E-3</td>
<td>3.8574E-2</td>
</tr>
<tr>
<td>30967</td>
<td>ER-nuclear sterol response pathway</td>
<td>1.0482E-3</td>
<td>3.8574E-2</td>
</tr>
<tr>
<td>6991</td>
<td>Response to sterol deletion</td>
<td>1.0482E-3</td>
<td>3.8574E-2</td>
</tr>
<tr>
<td>7182</td>
<td>Common-partner SMAD protein phosphorylation</td>
<td>1.0482E-3</td>
<td>3.8574E-2</td>
</tr>
</tbody>
</table>

**Table III.** Pathway enrichment analysis of DEGs in 3 stages.

<table>
<thead>
<tr>
<th>Pathway term</th>
<th>(p)-value</th>
<th>Bonferroni</th>
<th>Benjamini</th>
<th>FDR</th>
<th>Fisher exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa03050:Proteasome</td>
<td>1.8E-8</td>
<td>2.7E-6</td>
<td>2.7E-6</td>
<td>2.1E-5</td>
<td>1.9E-9</td>
</tr>
<tr>
<td>hsa04110: Cell cycle</td>
<td>7.4E-5</td>
<td>1.1E-2</td>
<td>5.7E-3</td>
<td>8.9E-2</td>
<td>2.2E-5</td>
</tr>
<tr>
<td>hsa04114: Oocyte meiosis</td>
<td>5.7E-3</td>
<td>5.9E-1</td>
<td>2.6E-1</td>
<td>6.6E-0</td>
<td>2.1E-3</td>
</tr>
<tr>
<td>hsa00190: Oxidative phosphorylation</td>
<td>9.0E-4</td>
<td>1.2E-1</td>
<td>1.2E-1</td>
<td>1.1E-0</td>
<td>2.3E-4</td>
</tr>
<tr>
<td>hsa00250: Alanine, aspartate and glutamate metabolism</td>
<td>2.5E-3</td>
<td>3.1E-1</td>
<td>3.1E-1</td>
<td>3.0E-0</td>
<td>3.5E-4</td>
</tr>
<tr>
<td>hsa05012: Parkinson’s disease</td>
<td>3.0E-3</td>
<td>3.4E-1</td>
<td>1.9E-1</td>
<td>3.5E-0</td>
<td>8.5E-4</td>
</tr>
<tr>
<td>hsa00520: Amino sugar and nucleotide sugar metabolism</td>
<td>3.2E-3</td>
<td>3.6E-1</td>
<td>1.4E-1</td>
<td>3.8E-0</td>
<td>1.1E-3</td>
</tr>
<tr>
<td>hsa00020: Citrate cycle (TCA cycle)</td>
<td>1.5E-2</td>
<td>9.0E-1</td>
<td>4.3E-1</td>
<td>1.7E-1</td>
<td>2.6E-3</td>
</tr>
<tr>
<td>hsa03320: PPAR signaling pathway 9</td>
<td>1.6E-2</td>
<td>9.2E-1</td>
<td>4.6E-1</td>
<td>1.7E-1</td>
<td>5.2E-3</td>
</tr>
<tr>
<td>hsa04350: TGF-beta signaling pathway</td>
<td>1.9E-2</td>
<td>9.3E-1</td>
<td>4.2E-1</td>
<td>2.1E-1</td>
<td>3.6E-3</td>
</tr>
<tr>
<td>hsa04310: Wnt signaling pathway</td>
<td>3.6E-2</td>
<td>9.9E-1</td>
<td>5.7E-1</td>
<td>3.5E-1</td>
<td>5.8E-3</td>
</tr>
</tbody>
</table>
volved in regulation cell growth and apoptosis. GADD45G mRNA expression is down-regulated in hepatocellular carcinoma, and that GADD45G causes cell cycle arrest at G2/M transition when transfected into Hep-G2 cells (30). Ying et al.31 have demonstrated that GADD45G can act as a functional new-age tumor suppressor but being frequently inactivated epigenetically in multiple tumors. It has been reported that decreased expression and aberrant methylation of GADD45G is associated with tumor progression and poor prognosis in esophageal squamous cell carcinoma.32 Besides, GADD45G induction by androgens requires new protein synthesis arguing that GADD45G is involved in differentiation induced by androgens.33

SPTBN1 (spectrin, beta, non-erythrocytic 1), alias ELF, belongs to spectrin family which has been implicated in transforming growth factor-β (TGF-β) signaling. Kitisin et al.34 have demonstrated that disruption of TGF-β signaling through β-spectrin ELF leads to hepatocellular cancer potentially through cyclin D1 deregulation. Back et al.35 have showed that ELF functions as a critical adaptor protein in TGF-β modulation of angiogenesis as well as cell cycle progression. Besides, loss of ELF in the liver leads the cancer formation by deregulated hepatocyte proliferation and stimulation of angiogenesis in early cancers. In addition, reduced SPTBN1 expression is correlated with shorter survival of pancreatic cancer patients.36 SPTBN1 is not only related to tumor progression but also maybe related to poor prognosis in HCC.

Of the 11 enriched pathways, the most significant pathway was proteasome. The proteasome is a 26S complex localized in the cytoplasm and nucleus, and contains a 20S proteolytic core.37-39 Proteasomes normally perform controlled degradation of proteins and proteins selected for degradation by tagging with a poly ubiquitin chain. In addition, proteasomes are important regulators of several key regulatory proteins including p53, cyclins, CDK inhibitors and NF-kB. The proteasome inhibitor bortezomib has been suggested for treatment of several cancers.40 Moreover, the proteasome pathway is required for cytokine-induced endothelial-leukocyte adhesion molecule expression.41

TGF-β signaling pathway is also a significant pathway. TGF-β signaling pathway is closely associated with gene SPTBN1 which is one of the overlapping genes we discussed above. TGF-β signaling pathway is involved in multiple cellular processes, including cell growth, differentiation, adhesion, migration, and apoptosis. TGF-β acts as a tumor suppressor at the early stages tumor development by inhibiting proliferation and inducing apoptosis, but TGF-β also contributes to tumor progression in carcinogenesis.42 Importantly, inactivation of TGF-β signaling is thought to play a role in the development of a number of cancers.43

Conclusions

We employed microarray analysis combined with bioinformatics methods to analyze the potential mechanism of tumor progression in HCC from BCLC stage A to BCLC stage C. Our analysis indicated several differentially expressed genes might play significant roles in tumor progression, including GADD45G, SPTBN1, CDC27, SLC25a27, QDPS1, TPDP2 and INSIG1. GADD45G and SPTBN1 not only contribute to tumor progression in HCC, but also correlate with poor prognosis in esophageal squamous cell carcinoma and pancreatic cancer respectively, suggesting that these genes might be associated with poor prognosis in other malignancies of the gastrointestinal tract including HCC as well. Furthermore, we performed pathway enrichment analysis on differentially expressed genes and found that some of these genes might play roles in the progression from BCLC stage A to BCLC stage C though pathways of “Proteasome”, “Alanine, aspartate and glutamate metabolism”, “TGF-beta signaling pathway”, “Wnt signaling pathway” and so on. We expect numerous advanced researches in tumor progression of HCC in the coming years based on our study.

Conflict of Interest

The Authors declare that there are no conflicts of interest.

References


1195


