Experimental and Molecular Pathology 98 (2015) 540–545

Contents lists available at ScienceDirect

Experimental and Molecular Pathology

journal homepage: www.elsevier.com/locate/yexmp

Identification of dysfunctional biological pathways and their synergistic mechanism in hepatocellular carcinoma process

Jun Liang *, Jing Lv, Zimin Liu

Department of Oncology, the Affiliated Hospital of Qingdao University, Qingdao, Shandong Province, China

ARTICLE INFO

Article history:
Received 12 February 2015
Accepted 20 March 2015
Available online 21 March 2015

Keywords:
Hepatocellular carcinoma
Transcriptome profiling
Differentially expressed genes
Pathway enrichment
Synergistic pathways

ABSTRACT

Background: Hepatocellular carcinoma (HCC) is a lethal and prevalent cancer worldwide. This study was conducted to investigate dysfunctional pathways and their synergistic mechanism in the HCC process.
Methods: We downloaded transcriptome profiling data (GSE25097) from the Gene Expression Omnibus (GEO) database, including 6 healthy liver samples, 40 cirrhosis samples, 243 adjacent non-tumor samples, and 268 HCC samples. Robust Multi-Array (RMA) in R software was employed to preprocess the downloaded dataset, and Student's t-test (FDR less than 0.001) was performed to identify the differentially expressed genes (DEGs) between 4 sample groups. Then, pathway enrichment analysis (FDR less than 0.05) based on iSubpathwayMiner was performed. Furthermore, we performed collaborative analysis on these pathways through calculating the Jaccard index, and crosstalk networks were constructed and visualized by Cytoscape.
Results: Totally, 4617, 9517, and 12,479 DEGs were identified between healthy liver and cirrhosis samples, cirrhosis and adjacent non-tumor samples, and adjacent non-tumor and HCC samples, respectively. Furthermore, a total of 26 crosstalks involving 13 pathways, 78 crosstalks involving 54 pathways, and 86 crosstalks involving 52 pathways were identified through the DEGs between healthy liver and cirrhosis samples, cirrhosis and adjacent non-tumor samples, and adjacent non-tumor and HCC samples, respectively. Moreover, 5 dysfunctional pathways were found to co-exist in the three processes of HCC. Among them, 3 dysfunctional pathways have collaborative relationship, including *Staphylococcus aureus* infection, leishmaniasis, and Chagas disease.
Conclusions: In this study, dysfunctional pathways in the HCC process and crosstalks between these pathways were investigated for the first time, providing new insight into the potential mechanisms of HCC.

1. Introduction

As one of the most lethal and prevalent cancers worldwide, hepatocellular carcinoma (HCC) is caused by various risk factors, such as chronic hepatitis B and C infections and prolonged exposure to hepatocarcinogens (El–Serag and Rudolph, 2007; Huang et al., 1999; Kew, 2010). The development of HCC is regarded as a multi-step process that results from genetic mutations (Kudo, 2009; Matsui et al., 2011). The HCC process, in which well differentiated liver cells developed into less differentiated cells, is accompanied by significant changes in the genetic and morphological properties of liver cells (Lazarevich et al., 2004).

The HCC process is characterized by extinction of tissue-specific gene expression, loss of morphological properties of epithelial cells, acceleration of proliferation, increased infiltration, and metastasis (Lazarevich et al., 2004). Over the past decades, a variety of signaling pathways involved in the control of liver cell function and proliferation have been studied (Assaf et al., 2012; Song et al., 2013; Strey et al., 2003). However, the molecular mechanism of the HCC process and liver cell differentiation remains obscure. Analysis revealed that the development of HCC was a dynamic three-step process: from healthy to cirrhosis, from cirrhosis to adjacent non-tumor, and then from adjacent non-tumor to HCC. During the HCC process, a variety of genes are involved (Pogribny and Rusyn, 2012; Thorgeirsson and Grisham, 2002), while their roles are poorly understood.

In the present study, bioinformatic analysis on transcriptome profiling data was performed to identify the significantly differentially expressed genes (DEGs) between the three steps in the HCC process. Furthermore, pathway enrichment analysis of DEGs was performed, and synergistic networks were constructed, consisting of the enriched biological pathways and their crosstalks. Moreover, the roles of DEGs in the HCC process were elucidated by using bioinformatic methods, the pathogenesis of HCC was investigated, and theoretical guidance for HCC treatment was provided.

Abbreviations: HCC, hepatocellular carcinoma; GEO, Gene Expression Omnibus; DEGs, differentially expressed genes; RMA, Robust Multi-Array; FDR, false discovery rate; MRSA, methicillin-resistant *S. aureus*.

* Corresponding author at: Department of Oncology, the Affiliated Hospital of Qingdao University, 16 Jiangsu Road, Qingdao, Shandong Province, China.
E-mail addresses: wushenglili@hotmail.com (J. Liang), jinglv1211@yeah.net (J. Lv), ziminliu201311@yeah.net (Z. Liu).

http://dx.doi.org/10.1016/j.yexmp.2015.03.028
0014-4800/

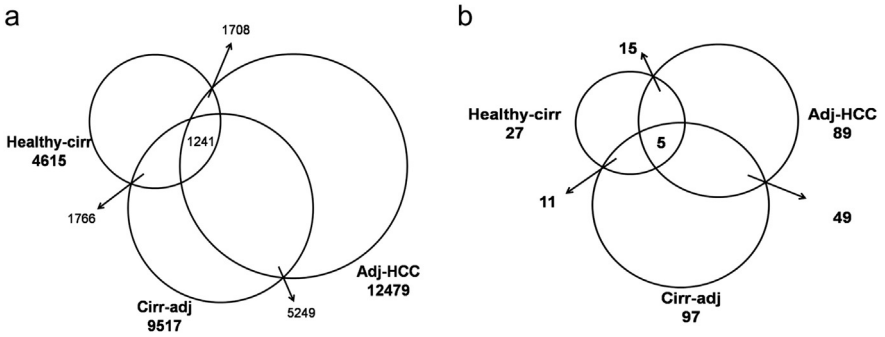

Fig. 1. The overlapped DEGs and dysfunctional pathways between the three steps of HCC process. Note: a. DEGs. DEGs: differentially expressed genes; Healthy: healthy liver samples; cirr: cirrhosis samples; adj: adjacent non-tumor samples; HCC: hepatocellular carcinoma samples; 4615, DEGs between healthy liver and cirrhosis samples; 9517, DEGs between cirrhosis and adjacent non-tumor samples; 12,479, DEGs between adjacent non-tumor and HCC samples. b. Dysfunctional pathways. Healthy: healthy liver samples; cirr: cirrhosis samples; adj: adjacent non-tumor samples; HCC: hepatocellular carcinoma samples; 27, dysfunctional pathways between healthy liver and cirrhosis samples; 89, dysfunctional pathways between cirrhosis and adjacent non-tumor samples; 97, dysfunctional pathways between adjacent non-tumor and HCC samples.

2. Materials and methods

2.1. Transcriptome profiling data

In order to study the mechanism of HCC, we downloaded transcriptome profiling data (GSE25097) (Lamb et al., 2011) from the National Center of Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2007). Dataset GSE25097 contains a total of 557 samples, including 6 healthy liver samples, 40 cirrhosis samples, 243 adjacent non-tumor samples, and 268 HCC samples. A Platform Rosetta/Merck Human RSTA Affymetrix 1.0 microarray, Custom CDF was employed for analysis.

2.2. Data preprocessing

We used Robust Multi-Array (RMA) (Irizarry et al., 2003) in R software to preprocess the downloaded dataset. In this process, the probe IDs were converted into gene symbols, followed by background correction and $\log_2$ transformation. For multiple probes corresponding to the same gene symbols, expression values were averaged to get the final expression value.

2.3. Identification of DEGs

Student's t-test was employed to screen DEGs between three contrast groups: (1) healthy liver and cirrhosis; (2) cirrhosis and adjacent non-tumor; and (3) adjacent non-tumor and HCC. The p-value of each gene was calculated and adjusted by the BH (Benjamin & Hochberg) multiple test correction method (Abbas et al., 2013). The threshold of adjusted p-value, namely false discovery rate (FDR) (Storey and Tibshirani, 2003), was set at 0.001. Under this criterion, DEGs with the fold change value greater than or equal to 2 were retained for subsequent analysis.

Table 1
Pathway enrichment analysis of the up-regulated and down-regulated DEGs between healthy liver and cirrhosis samples (top 5 pathways).

DEGs	Pathway ID	Pathway name	False discovery rate (FDR)
Up-regulated	path:03010	Ribosome	0
	path:00190	Oxidative phosphorylation	5.09E−08
	path:03050	Proteasome	1.00E−06
	path:04145	Phagosome	1.54E−05
	path:05010	Alzheimer's disease	0.000152
Down-regulated	path:04910	Insulin signaling pathway	0.002911
	path:04141	Protein processing in endoplasmic reticulum	0.002911
	path:04520	Adherens junction	0.002911
	path:00670	One carbon pool by folate	0.010015
	path:05213	Endometrial cancer	0.01385

2.4. Pathway enrichment analysis

In order to identify biological processes involved in the three steps of the HCC process, we used package iSubpathwayMiner (Li et al., 2009) to perform the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis of the screened DEGs. Assuming that KEGG pathways include a total of N genes, the number of DEGs is n, and the number of genes involved in a specific pathway is m, then, the probability of overlap of k genes, namely FDR, was calculated using the following formula:

$$\mathrm{FDR} = 1 - \sum_{0}^{k-1} \frac{C_k^m C_{n-k}^{N-m}}{C_n^N}.$$

Only the KEGG pathways with FDR less than 0.05 were selected.

2.5. Synergistic analysis of the enriched pathways

To recognize the synergistic relationships between enriched pathways, we constructed synergistic networks based on similarity among pathways. The Jaccard index (Evgeny et al.) was employed to measure the similarity among pathways. The intersection of genes in pathways A and B is denoted as # intersection (path A, path B), the union of genes in pathways A and B is denoted as # union (path A, path B), and the formula of the Jaccard index was illustrated as follows:

$$\text{Jaccard index (path A, path B)} = \frac{\#\text{ intersection (path A, path B)}}{\#\text{ union (path A, path B)}}.$$

Pathways fitting the following two conditions were regarded as pathways with crosstalk: (1) the Jaccard index is greater than or equal to 0.2; and (2) there exists at least one DEG overlapped in the two pathways. Based on these two requirements, we constructed a crosstalk network among the enriched dysfunctional pathways. In addition, crosstalk networks were visualized by Cytoscape software (Smoot et al., 2011).

3. Results

3.1. Data preprocessing and DEG identification

After the data preprocessing based on RMA in R software, a total of 20,158 genes were obtained from 37,582 raw probes. By utilizing Student's t test (FDR less than 0.001), significantly regulated genes in the development of HCC were identified. (1) A total of 4615 DEGs between healthy liver and cirrhosis samples were identified, including