Pathology – Research and Practice 213 (2017) 316–321

Contents lists available at ScienceDirect

Pathology – Research and Practice

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

Aberrant expression of cell cycle and material metabolism related genes contributes to hepatocellular carcinoma occurrence

Hongxian Yan[a,*,1], Zhaohui Li[b,1], Quan Shen[a], Qian Wang[c], Jianguo Tian[a], Qingfeng Jiang[a], Linbo Gao[d]

[a] Department of Hepatobiliary Surgery, Henan Provincial People's Hospital, Zhengzhou, Henan 650000, PR China
[b] Secondary Department of General Surgery, Luoyang Central Hospital Affiliated to Zhengzhou University, Luoyang, Henan 471003, PR China
[c] Department of Hepatobiliary Surgery, Henan Cancer Hospital, Zhengzhou, Henan 650000, PR China
[d] Laboratory of Molecular and Translational Medicine, West China Institute of Women and Children's Health, Key Laboratory of Obstetric & Gynecologic and Pediatric Diseases and Birth Defects of Ministry of Education, West China Second University Hospital, Sichuan University, Chengdu, Sichuan 610041, PR China

ARTICLE INFO

Article history:
Received 28 September 2015

Keywords:
Hepatocellular carcinoma
Differentially expressed genes
Functional enrichment analysis
Pathway enrichment analysis
Protein-protein interaction

ABSTRACT

This study aims to deepen our understanding of the molecular mechanism underlying the occurrence of hepatocellular carcinoma (HCC). We first downloaded a gene expression profile dataset GSE29721 (10 HCC and 10 control samples) from Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/). Differentially expressed genes (DEGs) were identified by the paired *t*-test using limma package. Pathway and functional enrichment analyses were performed with DAVID tools. Transcription factors were annotated with TRANSFAC database and tumor associated genes (TAGs) were annotated with TAG and TSGene databases. Protein-protein interaction (PPI) network was conducted using STRING online tool and function module was further identified with BioNet package. Totally, 527 up-regulated DEGs and 587 down-regulated DEGs were identified. GO functional and KEGG pathway enrichment analyses showed that the up-regulated DEGs were mainly related to cell division and cell cycle, while the down-regulated DEGs were largely related to material metabolism, especially secondary metabolism. Proteins encoded by DEGs *CDK1, BUB1, CDC20, NCAPG, NDC80, CDCA8, MAD2L1, CCNB1, CCNA2 and BIRC5* were hub genes with high degrees in the PPI network; further module analysis detected a subnetwork consisting of 55 proteins, such as CYP2B6, ACAA1, BHMT and ALDH2. Taken together, aberrant expression of cell cycle related genes (e.g., *CDK1, CCNA2, CCNB1, BUB1, MAD2L1* and *CDC20*) and material metabolism related genes (e.g., *CYP2B6, ACAA1, BHMT and ALDH2*) may contribute to HCC occurrence.

1. Introduction

Hepatocellular carcinoma (HCC) is the most common form of liver cancer accounting for 90% of all primary liver tumors, with an survival of 6–20 months if not intervened [1]. It is the fifth most common cancer and the third most common cause of death from cancer worldwide [2]. Strikingly, most HCC cases (>80%) occur in sub-Saharan Africa and Eastern Asia, and most astonishingly, China has 50% of HCC cases globally [3]. HCC brings heavy financial burden to both individuals and the society and this burden is expected to increase in the coming years [4]. Unfortunately, the management of patients with HCC has not dramatically changed [5].

Chronic hepatitis B virus (HBV) and chronic hepatitis C virus (HCV) infection are the primary causes of liver cancer, which account for approx. 75–80% of all HCC cases [6,7]. Environmental factors, such as alcoholic, smoking and aflatoxin exposure are important risk factors for HCC [8–10]. Recently, Milgrom et al. claimed that non-alcoholic steatohepatitis would become the first cause of HCC in America due to the increase of obesity-related liver disease [11]. In addition, genetic alterations have also been implicated in HCC pathogenesis, such as mutations in *PIK3CA* [12], *ARID2* [13], *TP53* and *ARID1A* [14], SNP polymorphism in ADAMTS5 [15], overexpression of *SALL4* [16] and *TFIIB* [17]. However, the mechanisms underlying hepatocellular tumorigenesis and progression remain poorly understood.

* Corresponding author at: Department of Hepatobiliary Surgery, Henan Provincial People's Hospital, Weiwu Road, NO. 7, Zhengzhou City, Henan 650000, PR China.
E-mail address: yanhxianh@163.com (H. Yan).
[1] Hongxian Yan and Zhaohui Li contributed equally to this work.

http://dx.doi.org/10.1016/j.prp.2017.01.019
0344-0338/

In the present study, we reanalyzed a gene expression profile dataset GSE29721 deposited in GEO by Stefanska et al. [18] using bioinformatics tools, with the attempt to deepen our understanding of HCC pathogenesis.

2. Materials and methods

As the paper did not involve any human or animal's study, the ethical approval or patients gave informed consent were not required.

2.1. Source of gene expression profile data

The gene expression profile dataset of GSE29721 was downloaded from the GEO (Gene Expression Omnibus) database (http://www.ncbi.nlm.nih.gov/geo/), which is annotated based on Affymetrix Human Genome U133 Plus 2.0 Array platform. It was deposited in GEO by Stefanska et al. in June, 2011 [18]. A total of 20 cancerous and normal adjacent normal tissue samples were obtained from patients with HCC, respectively, including 10 samples of normal adjacent tissues (control group) and 10 samples of cancerous tissues (tumor group) were used for this analysis.

2.2. Identification of differentially expressed genes

The raw data of GSE29721 dataset were preprocessed using AFFY package in R/Bioconductor software [19]. The detailed preprocessing protocols included background correction, quantile normalization and probe summarization using Robust Multi-array Average (RMA) package [20] with defaulted parameters. Then the gene expression matrix was obtained. DEGs between the tumor and control samples were identified by the paired *t*-test using limma package in R language with cutoffs of $p < 0.05$ and $|\log_2 FC$ (fold change)$|>1$ [21]. Meanwhile, DEGs between the tumor and control samples in the TGCA data were also identified by the paired *t*-test using the same cutoffs.

2.3. Functional and pathway enrichment analyses of DEGs

Both Gene Ontology (GO) [22] enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) [23] pathway enrichment analysis were performed to investigate the functions and potential roles of the identified DEGs in the pathogenesis of HCC using the online tool Database for Annotation, Visualization and Integrated Discovery (DAVID, v6.7, https://david.ncifcrf.gov/) [24]. Meanwhile, the enrichment analyses were performed independently in up-regulated and down-regulated DEGs. *P* value <0.05 was set as the cut-off for enrichment analysis.

2.4. Identification of tumor associated genes

With reference to the database TRANSFAC (http://www.gene-regulation.com/index2), DEGs that were predicted to be transcription factors were identified [25], and the resulting DEGs were finally identified to be transcription factors only when they were also involved in the GO term (Transcription activity). In addition, tumor associated genes (TAG) including oncogenes and tumor suppressor genes were also identified based on the Tumor Suppressor Gene (TSGene) database (http://bioinfo.mc.vanderbilt.edu/TSGene/) [26,27].

2.5. Protein-protein interaction (PPI) network construction and module detection

DEGs were mapped to the STRING (http://string-db.org/cgi/input.pl?UserId=OKHJlLJy0nly&sessionId=pJlFY0Zvk3VZ&input_page_show_search=on) [28] database and protein-protein interaction pairs with combined score larger than 0.9 were selected. Then, the PPI network of the DEGs was visualized using Cytoscape software [29]. Only the PPI pairs that have got experimental verification, co-expression analysis or record in database were chosen for the PPI network construction. Finally, the interaction network was analyzed, and the degree of each protein (namely, the number of other proteins that were connected to a certain protein), was calculated. Furthermore, functional modules were further identified from the PPI network using the BioNet package [30]. The proteins with high degree in the PPI network and its module were considered as the hub nodes. FDR (false discovery rate) <0.0005 was set as the threshold for this analysis.

2.6. Validation of hub proteins

Finally, the gene expression data of liver hepatocellular carcinoma that were available to the public in The Cancer Genome Atlas (TCGA) database (http://cancergenome.nih.gov/) were also downloaded (upgraded to May 5th, 2016) to validate the expression of the hub genes identified in the PPI network and the module. These data had been preprocessed, which were collected from 374 tumor samples and 50 normal samples.

3. Results

3.1. DEGs analysis

A total of 1114 DEGs were identified. Among the 1114 DEGs, 527 DEGs were up-regulated, such as *CDK1, CCNA2, CCNB1, BUB1, MAD2L1* and *CDC20,* and 587 DEGs were down-regulated, such as *CYP2B6*, *ACAA1*, *BHMT*, *ADH1A, ADH1B, ADH4* and *ALDH2*.

3.2. Functional and pathway enrichment analyses of DEGs

According to the GO functional enrichment analysis, the upregulated DEGs and the downregulated were enriched in 293 and 416 GO terms respectively. The top five GO terms were mainly related to cell mitosis and cell cycle, and the down-regulated GO terms were mostly related to material metabolism (Table 1).

According to the KEGG pathway enrichment analysis, the upregulated and downregulated DEGs were enriched in 38 and 11 pathways respectively. The top three pathways enriched by the upregulated genes were related to cell cycle, DNA replication and oocyte meiosis respectively, and the top five pathways enriched by the upregulated genes were all related to metabolism (Table 2).

3.3. Prediction of tumor associated genes

Among the downregulated DEGs, 15 genes were transcriptional factors, and 33 genes were TAGs, including 4 oncogenes, 24 tumor suppressor genes and 5 genes with unidentified role in the development of tumor (Table 3). Noticeably, no upregulated genes were identified as TFs or TAGs here.

3.4. PPI network construction and module detection

Based on the PPI pairs in database STRING, we constructed an interaction network of proteins encoded by the DEGs (Fig. 1). The top 10 proteins with connection degree $\geq$60 were CDK1, BUB1, CDC20, NCAPG, NDC80, CDCA8, MAD2L1, CCNB1, CCNA2 and BIRC5, respectively. Strikingly, all the 10 DEGs were up-regulated. A module consisting of 55 proteins were further identified, including those encoded by DEGs *CYP2B6*, *ACAA1*, *BHMT*, *ADH1A*, *ADH1B*, *ADH4* and *ALDH2* (Fig. 2).