Pathology – Research and Practice 213 (2017) 327–332

Contents lists available at ScienceDirect

Pathology – Research and Practice

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

Original article

Identification of personalized dysregulated pathways in hepatocellular carcinoma

Hong Li [1], Xiumei Jiang [1], Shengjie Zhu, Lihong Sui *

Department of Oncology, Weihai Central Hospital, Weihai, 264400, Shandong, PR China

ARTICLE INFO

Article history:
Received 6 May 2016

Keywords:
Hepatocellular carcinoma
Differentially expressed genes
Personalized dysregulated pathways
Normal distribution analysis

ABSTRACT

Introduction: Hepatocellular carcinoma (HCC) is the most common liver malignancy, and ranks the fifth most prevalent malignant tumors worldwide. In general, HCC are detected until the disease is at an advanced stage and may miss the best chance for treatment. Thus, elucidating the molecular mechanisms is critical to clinical diagnosis and treatment for HCC. The purpose of this study was to identify dysregulated pathways of great potential functional relevance in the progression of HCC.
Materials and methods: Microarray data of 72 pairs of tumor and matched non-tumor surrounding tissues of HCC were transformed to gene expression data. Differentially expressed genes (DEG) between patients and normal controls were identified using Linear Models for Microarray Analysis. Personalized dysregulated pathways were identified using individualized pathway aberrance score module.
Results: 169 differentially expressed genes (DEG) were obtained with $|\log FC| \geq 1.5$ and $P \leq 0.01$. 749 dysregulated pathways were obtained with $P \leq 0.01$ in pathway statistics, and there were 93 DEG overlapped in the dysregulated pathways. After performing normal distribution analysis, 302 pathways with the aberrance probability ≥ 0.5 were identified. By ranking pathway with aberrance probability, the top 20 pathways were obtained. Only three DEGs (TUBA1C, TPR, CDC20) were involved in the top 20 pathways.
Conclusion: These personalized dysregulated pathways and overlapped genes may give new insights into the underlying biological mechanisms in the progression of HCC. Particular attention can be focused on them for further research.

1. Introduction

Hepatocellular carcinoma (HCC) is the most common liver malignancy, and ranks the fifth most prevalent malignant tumors worldwide [1]. Despite recent progress in anticancer, only 7% of patients with advanced HCC can expect to live for 5 years. In general, HCC are detected until the disease is at an advanced stage and may miss the best chance for treatment. Therefore, elucidating the molecular mechanisms is critical to clinical diagnosis and treatment for HCC.

It is generally accepted that the occurrence of HCC is the result of genomic alteration [2,3]. In recent years, microarray technology has been used as an advanced high-throughput strategy for detecting gene expression profiles [4]. Tackels-Horne and colleagues successfully identified 842 over-regulated genes and 343 down-regulated genes involved in the pathogenesis of HCC [5]. Similarly, Chen and co-workers made a cross-study comparison for the tumor tissues and non-tumor tissues, and the results indicated that 1640 genes were significantly differentially expressed in the tumor tissues [6]. In addition, several growth factors, including EGFR, HDGF and IGF, have been proposed to be involved in the progression of HCC [7–9]. A previous study has demonstrated that FOXO-TXNIP pathway played a pivotal role in the inhibition of HCC growth by MK-801 [10].

Previous studies have proved that many critical genes and pathways are dysregulated during cancer initiation and progression [11]. Although there have been numerous studies on HCC pathogenesis, the results are not uniform and share only a small number of potential genes and pathways. The shortcoming highlights the importance of personalized pathway analysis. However, most current pathway analyses are mainly focused on discovering dysregulated pathways between two phenotype groups. Based on the shortcoming, a new approach is proposed to identify dysregulated pathways in an individual case. Our proposed method is based on the comparison of one cancer sample with many accumulated normal samples (we use "nRef" to refer to the accumulated normal

* Corresponding author at: Weihai Central Hospital, No.3, Mishan West Road, Wendeng District, Weihai 264400, Shandong, PR China.
E-mail address: suihonglimed@yeah.net (L. Sui).
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.prp.2017.01.015
0344-0338/

samples). Personalized identification of dysregulated pathways is an important step toward understanding the disease mechanisms. An important clinical aspect of our method is providing a pathway interpretation of a single cancer, even though matched normal data are unavailable. Finally, individualized dysregulated pathways were identified by quantifying the aberrance of an individual sample's pathway by comparing it with accumulated normal samples.

In this study, differentially expressed genes (DEG) and individualized dysregulated pathways were identified. Microarray data of HCC patients were downloaded from ArrayExpress database [12]. The data were then preprocessed, and DEGs were identified using the LIMMA package. Subsequently, significantly dysregulated pathways between normal and cancer groups were identified. Finally, personalized dysregulated pathways were analyzed by using accumulated normal data. Among the 72 patients with HCC, the aberrance probability of each pathway was calculated. This study might provide useful information for exploring critical genes and individualized altered pathways, which can give insights into the diagnosis and treatment of HCC.

2. Materials and methods

2.1. Datasets

2.1.1. Gene expression data and preprocessing

Microarray expression data were downloaded from ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/) using accession number E-GEOD-39791 [13]. The dataset included gene expression data from tumor and matched non-tumor surrounding tissues of 72 HCC patients who underwent surgical resection as the primary treatment. Gene expression data was generated using the HumanHT-12 Version 4.0 Expression Beadchip (Illumina) according to the annotation files.

Linear Models for Microarray Data (LIMMA) was chosen to reprocess gene expression data. To eliminate influences of nonspecific hybridization, the background was corrected using robust multichip average (RMA) [14]. Normalization was performed with quantiles function. Then we used the MAS method to correct perfect match (PM)/mismatch (MM) [15], and the Medianpolish function was used to summarize expression data [16]. Then probes were mapped to gene symbols. The probe was discarded if it could not match any genes. The levels of probes were averaged as the final gene expression value if more than one probe was mapped to a single gene [17].

2.1.2. Pathway data

All Homo sapiens-related biological pathway data were downloaded from Reactome database (http://www.reactome.org/) [18]. In general, pathways with a large number of genes are not easily understood by human experts [19]. Therefore, the pathways with a gene size >100 were filtered out. Meanwhile, the pathways without transformational gene symbols were removed. As a result, a total of 1022 pathways, including 4928 genes, were obtained for further analysis.

2.2. DEG analysis

In the present study, Linear Models for Microarray Analysis (LIMMA) package of R were used to identify DEG between HCC patients and normal controls [20]. The values of $|\log FC| \geq 1.5$ and $P \leq 0.01$ were selected as the cut-off criteria. In addition, the DEG belonging to pathways were ascertained by taking the intersection of identified DEG and all genes in pathways.

2.3. Pathway analysis

Pathway analysis has become the first choice for capturing biologically and clinically relevant information. In our study, an individualized pathway aberrance score (iPAS) model was proposed to identify dysregulated pathways [21]. It makes special use of accumulated normal data to serve as the nRef. Data-processing procedures were described in more detail as follows.

2.3.1. Gene-level statistics

The microarray data of normal samples were normalized using the quantile normalization method in preprocessCore package. For normal controls, the mean and standard deviations of the gene expression level were calculated. For individual tumor cases, quantile normalization [15] was performed after combining the single tumor microarray with all nRef samples. The gene-level statistics of individual tumor samples were standardized using the mean and standard deviation of the reference. The formula is shown as the following:

$$z_i = \frac{g_{\mathrm{Ti}} - mean\,(g_{\mathrm{nRef}})}{stdev\,(g_{\mathrm{nRef}})} \quad (1)$$

where g_{Ti} represents the expression value of i-th gene belonging to the tumor cases, where $mean\,(g_{\mathrm{nRef}})$ and $stdev\,(g_{\mathrm{nRef}})$ represent the mean and standard deviation of the gene expression value of all reference samples, respectively.

2.3.2. Pathway-level statistics

For each specific pathway, the standardized gene-level statistics of all genes were extracted. Then the gene-level statistics for all genes in a pathway were aggregated into a single pathway-level statistic. The model is as follows:

$$\mathrm{iPAS} = \sum \mathrm{zin} \quad (2)$$

where z_i symbolizes the standardized expression value of i-th gene belonging to the pathway, where n symbolizes the number of genes belonging to the pathway.

2.3.3. Dysregulated pathways

Significance test was performed to assess dysregulated pathways associated with HCC. A generalized two-sample wilcoxon test was applied to assess the statistical significance of the pathway-level statistic. The false discovery rate (FDR) was used to correct the significance level [22,23]. The pathway with $P < 0.01$ was considered to be significantly dysregulated between normal and cancer groups. Moreover, a hierarchical clustering algorithm was applied to grouping the dysregulated pathways and the results were visualized with TreeView [24].

In addition, we obtained the common genes between known disease genes and dysregulated pathway genes. The known genes related to liver disease were selected, and a total of 39 genes were downloaded [25]. The pathways, including disease genes, were also ascertained.

2.3.4. Individualized dysregulated pathways

Significance can be obtained against null distribution generated from normal samples on the basis of pathway-level statistics. The statistic of the null distribution was acquired via comparing every normal sample with all nRef samples. Then the p-value of each pathway was obtained after comparing a single cancer case with the null distribution generated above. A value with $P < 0.05$ was considered to indicate a statistically significant difference. Among the 72 patients with HCC, the aberrance probability of each significant pathway ($P < 0.05$) was calculated and ranked. The pathways with higher aberrance probability were considered to be more prone to alterations.