European Review for Medical and Pharmacological Sciences 2016; 20: 4266-4273

A personalized approach identifies disturbed pathways and key genes in hepatitis C virus-cirrhosis with hepatocellular carcinoma

Y.-L. LI, M.-X. ZHENG, G. WANG

Department of Hepatopancreatobiliary and Spleen Surgery, Baoji Central Hospital, Baoji, Shaanxi Province, P.R. China

Abstract. – OBJECTIVE: This work aimed to identify disturbed pathways in hepatitis C virus (HCV)-cirrhosis with hepatocellular carcinoma (HCC) based on individualized pathway aberrance score (iPAS) method.

MATERIALS AND METHODS: First of all, gene expression data and pathway data were recruited and preprocessed. Next, iPAS method, which contained three steps (gene-level statistics based on average Z algorithm, pathway-level statistics and pathway significant analysis based on Wilcoxon-test), was performed to identify differential pathways in HCV-cirrhosis with HCC. Then, a protein-protein interaction (PPI) network was conducted based on the genes enriched in the differential pathways. Finally, topological analysis of the PPI network combined with cancer genes was conducted to identify hub disease genes.

RESULTS: After a systematic operation by the iPAS method, a total of 34 differential pathways were identified (p-value < 0.01). From the PPI network that was constructed using the 243 genes in the differential pathways, a total of 24 hub genes were obtained by conducting degree centrality, and 4 hub cancer genes (UBC, MAPK1, NOTCH1 and RHOA) were identified. An in-depth analysis indicated that NF-kB is activated and signals survival pathway contained the most cancer genes (number = 7), in which there was a hub cancer gene UBC. In addition, as we set the p-value in ascending order, we found that opioid signaling pathway was the most significant pathway (p = 1.59E-06), and hub cancer gene MAPK1 was enriched in this pathway.

CONCLUSIONS: The altered pathways and several key genes identified by this method were predicted to play important roles in HCV-cirrhosis with HCC and might be potentially novel predictive and prognostic markers for HCV-cirrhosis with HCC.

Key Words:

Hepatitis C virus, Hepatocellular carcinoma, Network, Individualized pathway aberrance score.

Introduction

Hepatocellular carcinoma (HCC) is the most common form of liver cancer[1], and hepatitis C virus (HCV) is a leading cause of HCC[2]. It was indicated that the de-regulation of common reference genes in HCC arose from HCV-infected liver[3]. Also, genetic variation had been reported to have important links to the development and progression of HCC in HCV-cirrhosis patients[4,5]. However, due to the poor clinical outcomes of patients with HCV-cirrhosis who are diagnosed with advanced-stage HCC, improved markers for early detection are needed, so as to reduce time to transplantation and thereby yield improved patient outcomes.

Techniques, such as high-throughput sequencing and gene/protein profiling techniques, have transformed biological research by enabling comprehensive monitoring of a biological system; pathway analysis has become the first choice for extracting and explaining the underlying biology for high-throughput molecular measurements. Khatri et al[6] classified these methods into three types: over-representation analysis (ORA), functional class scoring (FCS) and pathway topology (PT)-based approaches. The traditional method of pathway analysis uses only the most significant genes and discards others, and considers only the number of genes and ignores the magnitude of expression changes, thus resulting in information loss for marginally significant genes[7]. Moreover, the operation of many important pathways is altered during disease initiation and progression, so identifying altered pathways in an individual is important for understanding disease mechanisms and for the future application of custom therapeutic decisions[8]. In other words, existing pathway analysis techniques are

Corresponding Author: Gang Wang, MD; e-mail: aiwangwang2015@sina.cn

mainly focused on discovering altered pathways between normal and disease groups and are not suitable for identifying the pathway aberrance that may occur in an individual sample.

While, there was a brand new pathway analysis method, individualized pathway aberrance score (iPAS) method, to conduct pathway cluster analysis to identify altered pathways[9]. The iPAS was based on the comparison of one disease sample with many accumulated control samples and was suitable to adopt single-layer omics data and expendable to interpret a patient in the context of many published or user-defined pathway gene sets. Compared with pathway recognition algorithm using data integration on genomic models (PARADIGM)[10], this method had more freedom in terms of data and gene sets, as it preferred multi-layered omics data and required predefined functional structure among omics objects. While compared with pathway deregulation score (PDS)[11], this new method did not assume an individual sample belongs to a cohort; it used accumulated control tissue data as a reference.

Therefore, in the present research, we used this brand new method to quantify the aberrance of an individual sample's pathway in HCV-cirrhosis with HCC patients by comparing it with accumulated HCV-cirrhosis without concomitant HCC samples. To achieve this, firstly, we separately recruited and preprocessed the gene expression data and pathway data from the database. Secondly, gene-level statistics, pathway-level statistics and pathway significant analysis were successively conducted to identify differential pathways. Then, protein-protein interaction (PPI) network was constructed based on the genes in differential pathways, and topological analysis of the PPI network that contained cancer genes was performed to identify the hub genes and key pathways. According to personalized pathway analysis, this work could reveal the disturbed pathways involved in HCV-cirrhosis with HCC compared with accumulated HCV-cirrhosis without HCC samples, which might contribute to further understand the underlying mechanisms of HCV-cirrhosis with HCC.

Materials and Methods

Gene Expression Data Recruitment and Preprocessing

To study the molecular change of HCV-cirrhosis with and without concomitant HCC, the gene expression profile of E-GEOD-17967, existed on A-AFFY-37 - Affymetrix GeneChip Human Genome U133A 2.0 [HG-U133A_2] platform, was downloaded from ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/). The data of E-GEOD-17967 was obtained from 63 samples, which contained 16 samples of HCV-cirrhosis with HCC and 47 samples of HCV-cirrhosis without HCC[4]. The HCV-cirrhosis without HCC was defined as control group (we used 'nRef' to refer to the accumulated control samples hereinafter) and the HCV-cirrhosis with HCC was defined as disease group in the following.

Background-corrected signal intensities were determined using the Micro Array Suite 5.0 (MAS 5.0) software (Affymetrix®)[12]. Robust multichip average (RMA) method[13] and quantile-based algorithm[13] were used to normalize the dataset obtained from the ArrayExpress database. Meanwhile, the gene expression value was transformed to a comparable level and a gene-filter package was used to screen the data. Finally, we got 12,493 genes in all.

Pathway Data Recruitment and Preprocessing

The pathways of all human beings were downloaded from Reactome pathway database (http://www.reactome.org/). In general, the larger number of genes enriched in the pathway, the more complex metabolic reactions existed. Pathways with small gene sizes were more easily understood by human experts. Therefore, the pathways whose gene sizes > 100 were removed. Next, we took the intersection of the genes contained in the pathways with genes contained in the gene expression profile, as well as removed the pathway whose intersection value was zero. Finally, we obtained 1,009 pathways, which contained 4,284 genes.

Gene-level Statistics

Firstly, we performed data normalization across the nRef using normalize quantiles function in the Bioconductor package preprocessCore[15]. Secondly, the mean expression value and standard deviation of the genes in the control samples were calculated. Finally, quantile normalization of the individual gene contained in the disease samples was performed by taking the mean expression value and standard deviation of the control samples as reference, so as to obtain the gene-level statistics value of the

individual gene. The formula was defined as following:

$$Z_i = \frac{g_{Di} = \text{mean}\ (N_j}{\text{stdeve}\ (N_j)}$$

Where *mean* (N_j) symbolized mean expression value of the genes in nRef and *stdev* (N_j) symbolized the standard deviation of the genes in nRef, g_{Di} symbolized the expression value of *i*-th gene in the disease samples and Z_i symbolized the standardized expression value of *i*-th gene in the disease samples, where the number of genes belonging to the disease samples was *i*.

Pathway-level Statistics

Average Z method was used to conduct pathway-level statistics after having standardized the gene expression value. This method was a modification of existing pathway analysis techniques, enabling us to test an individual disease sample's pathway aberrance by using the accumulated control data. For each pathway, the gene-level statistics value of all genes was extracted and the average of the gene-level statistics value was defined as the pathway statistics value. The formula for calculating iPAS was defined as following:

$$\text{iPAS} = \frac{\sum_i^n Z_j}{n}$$

A vector $Z = (z_1, z_2 \ldots z_n)$ denoted the expression status of a pathway, where Z_i symbolized the standardized expression value of *i*-th gene, where the number of genes belonging to the pathway was *n*.

Pathway Significant Analysis

Cluster analysis using Average Z was conducted on the disease and control data to identify pathway clusters. First of all, all of the pathway statistics value were tested by *Wilcoxon*-test[16] and the *p*-values were adjusted by false discovery rate (FDR)[17]. The pathways of whose *p*-value < 0.01 were regarded as differential pathways. Then, the differential pathways were extracted to draw a heatmap.

PPI Network

The genes contained in the differential pathways were imported to STRING (http://string-db.org/) to construct the PPI network, and the genome of human sapiens were chosen as a background. Also, text mining was conducted on the HCC disease genes existed in the NCBI database. Disease genes mapped on the PPI network were selected for further analysis.

Centrality Analysis

Centrality analysis is a network analysis method to investigate biological networks, such as gene regulatory, protein interaction and metabolic networks, so as to identify interesting elements of a network[18,19]. Degree centrality is one of the centrality measures and is a simple local measure, which based on the notion of neighborhood. The index is useful in case of static graphs, for situations when we are interested in finding vertices that have the most direct connections to other vertices[20]. Therefore, in the present study, we mainly analyzed the degree centrality of the network by using Cytoscape Version 3.1.0. The genes whose degree ≥ 55 were considered as hub genes.

Results

Differential Pathways Analysis

We obtained the pathway-level statistics value of each pathway contained in the disease and normal samples according to the iPAS method. The *p*-value of each pathway was gained via conducting Wilcoxon-test on the pathway-level statistics value and the *p*-value adjusted by FDR. A total of 34 differential pathways between the disease and control conditions (*p*-value < 0.01) were identified. These differential pathways were selected and drew a heatmap (Figure 1).

PPI Network Construction and Centrality Analysis

There were 243 genes in the 34 disrupted pathways. All of the genes were imported to STRING to construct the PPI network. Finally, a PPI network contained 239 genes (2,945 relationships) was constructed (Figure 2). By text mining in NCBI database, a total of 737 human protein-coding genes related to HCC were mined, and there were 30 disease genes contained in these differential pathways. Network analysis showed that 29 of them could be mapped onto the PPI network. The sub-network of these 29 cancer genes was extracted (Figure 3). Further, by analyzing the degree centrality of the PPI network, twenty-four hub genes were obtained under the threshold value of degree ≥ 55 (Table I), and 4 of them were hub

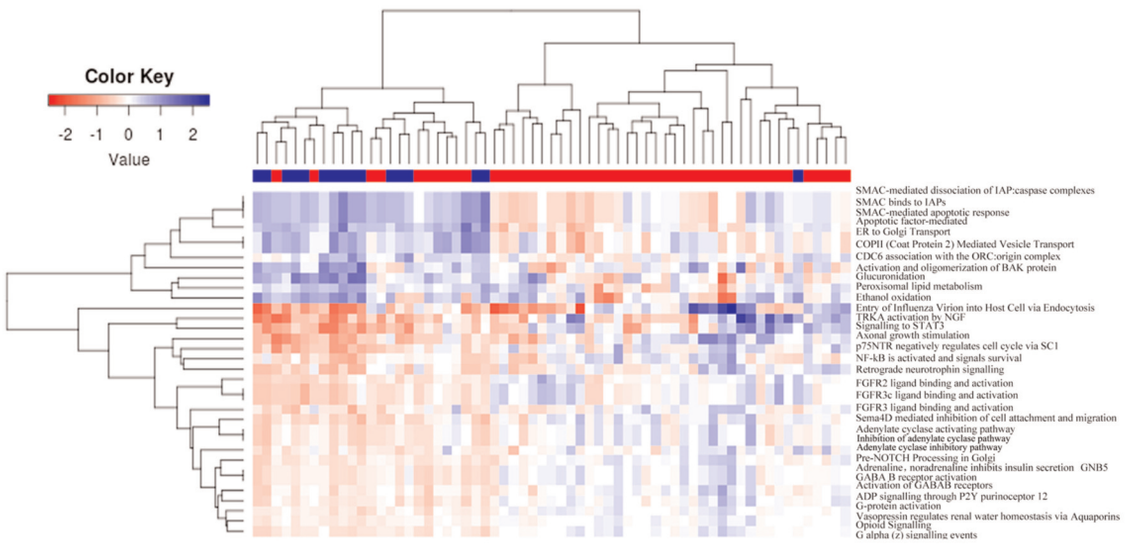

Figure 1. Cluster iPAS of altered pathways in HCV-cirrhosis with HCC. Differential pathways ($p < 0.01$) were clustered according to iPAS. The colors in the heatmap represented the pathway statistics values according to the color scaleplate. The abscissa axes represented pathways and vertical axes represented samples.

cancer genes, including *UBC* (degree = 155), *MAPK1* (degree = 86), *NOTCH1* (degree = 55) and *RHOA* (degree = 55).

Then, a comprehensive analysis was performed on these 29 cancer genes and differential pathways, and the details were shown in Table II. It was obviously that NF-kB is activated and signals survival pathway contained the most cancer genes (number = 7), in which there was a hub cancer gene *UBC*. In addition, setting the *p*-value in ascending order, we found that opioid signaling pathway was the most significant pathway (p = 1.59E-06), and hub cancer gene *MAPK1* was enriched in this pathway.

Discussion

In the present paper, we firstly performed analysis on the pathways via the iPAS method to identify the altered pathways that related to HCV-cirrhosis with HCC. A total of 34 differential pathways which contained 243 genes were identified. Degree centrality analysis of the subnetwork that not only contained pathway genes but also contained cancer genes indicated that there were several hub cancer genes, such as *UBC* and *MAPK1*. In order to further to understand these altered pathways and hub genes, we performed an in-depth discussion on the repre-

Table I. The degree value of the hub genes.

Gene symbol	Degree	Gene symbol	Degree
UBC	155	PRKCA	60
MAPK1	86	ADCY9	57
GNAI3	75	ADCY8	57
PRKACA	74	ADCY2	57
GNAO1	70	GNAT2	56
PRKACB	69	ADCY3	56
PRKACG	68	ADCY6	56
GNAI1	66	GNB2	55
GNAZ	65	RHOA	55
CREB1	64	GNAT1	55
GNB1	64	ADCY7	55
GNAT3	63	NOTCH1	55

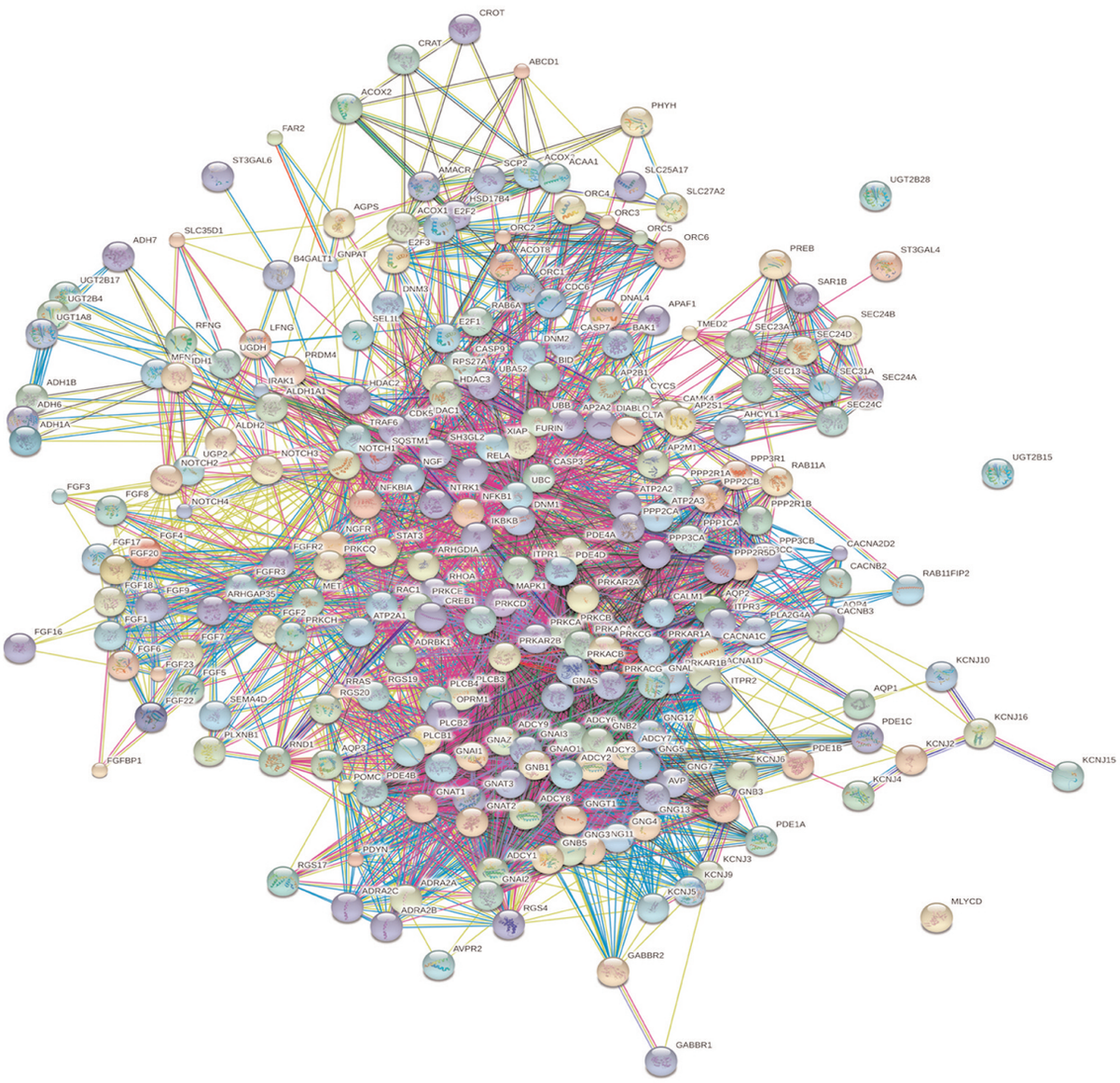

Figure 2. The protein-protein interaction network constructed with the genes enriched in the differential pathways based on the genome of human sapiens.

sentative pathways (NF-kB is activated and signals survival and opioid signaling pathway) and hub genes (*UBC* and *MAPK1*).

Nuclear factor kappa-B (NF-kB) is a transcription factor consisting of NF-kB DNA-binding dimers, and the proteins modulate the activation and function of NF-kB[21]. NF-kB is activated in response to DNA damage and may contribute to the cell death process by inducing the express ion of p53[22], while it had been indicated that there was some association of p53 polymorphisms with the presence and early-onset of HCC. Furthermore, it was reported that the hepatic cancer HA22T/VGH cell line, which constitutively expresses activated NF-kB, was chosen as a model to examine the antitumor activity of curcumin, also in relationship to its possible influences on the activation of the transcription factor and on the expression of the inhibitory of apoptosis proteins and of other NF-kB target genes[23].

UBC is one of two stress-inducible polyubiquitin genes in mammals and is thought to supplement the constitutive *UBA* genes in maintaining cellular ubiquitin (UB) levels during episodes of

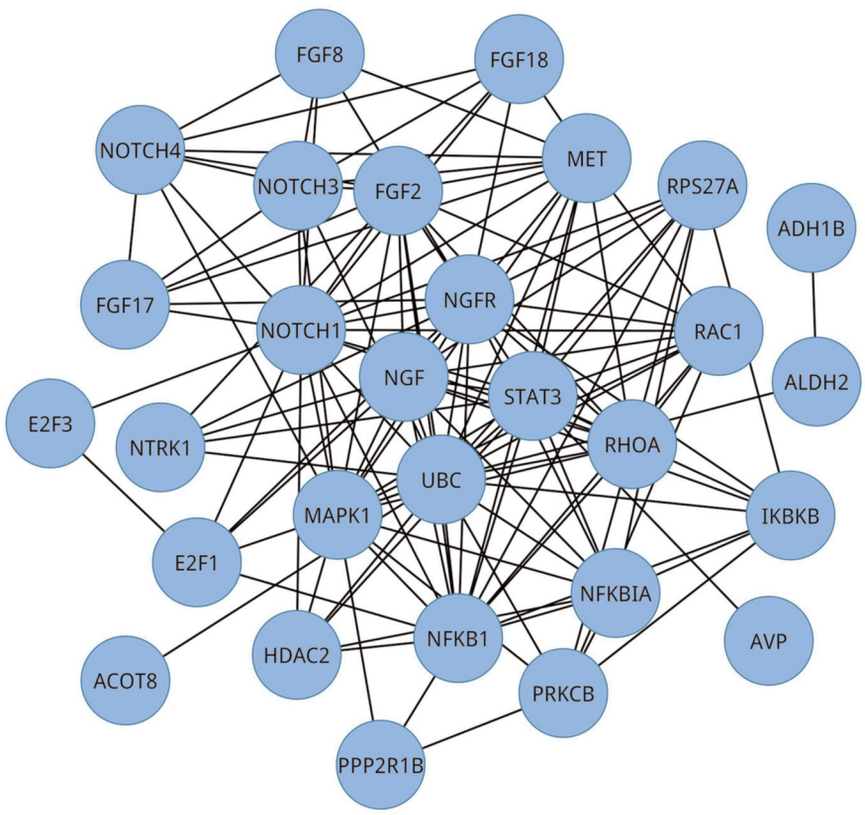

Figure 3. Module A identified from the HCC-related PPI network. Yellow circles represent hub genes.

Table II. The details of disease genes that enriched in the differential pathways.

Pathways term	*p*-value	No.	Gene symbol
NF-kB is activated and signals survival	1.59E-04	7	IKBKB, NFKB1, NFKBIA, NGF, NGFR, RPS27A, UBC
FGFR2 ligand binding and activation	1.59E-04	4	FGF17, FGF18, FGF2, FGF8
FGFR3 ligand binding and activation	1.84E-04	4	FGF17, FGF18, FGF2, FGF8
FGFR3c ligand binding and activation	1.84E-04	4	FGF17, FGF18, FGF2, FGF8
Axonal growth stimulation	8.00E-05	3	NGF, NGFR, RHOA
Ethanol oxidation	8.65E-05	3	ADH1B, ADH1C, ALDH2
p75NTR negatively regulates cell cycle via SC1	1.77E-06	3	HDAC2, NGF, NGFR
Pre-NOTCH Processing in Golgi	1.48E-04	3	NOTCH1, NOTCH3, NOTCH4
Sema4D mediated inhibition of cell attachment and migration	8.65E-05	3	MET, RAC1, RHOA
Signalling to STAT3	9.47E-06	3	NGF, NTRK1, STAT3
CDC6 association with the ORC:origin complex	6.83E-05	2	E2F1, E2F3
Opioid signalling	1.59E-06	2	MAPK1, PPP2R1B
Retrograde neurotrophin signalling	3.62E-06	2	NGF, NTRK1
TRKA activation by NGF	2.46E-04	2	NGF, NTRK1
G alpha (z) signalling events	2.07E-06	1	PRKCB
Peroxisomal lipid metabolism	5.82E-05	1	ACOT8
Vasopressin regulates renal water homeostasis via aquaporins	3.04E-04	1	A1AVP

cellular stress[24]. *UBC* had been identified to be suitable for the normalization of gene expression data among tumor tissues[25]. It was reported that the polyubiquitin gene *UBC* was essential for fetal liver development, cell-cycle progression and stress tolerance in mouse[26]. In this research, *UBC* was with the highest degree in the PPI network, as well as enriched in NF-kB is activated and signals survival pathway, the significant altered pathway. Therefore, we could infer that NF-kB is activated and signals survival pathway was a significant pathway for HCV-cirrhosis with HCC.

Opioids are chemical substances similar to opiates, the active substances found in opium (morphine, codeine, etc.). The opioid action was mediated by the receptors for endogenous opioids; peptides such as the enkephalins, the endorphins or the dynorphins[27]. It was reported that binding of opioids to opioid receptors on immune cells triggered similar second messengers as in neuronal cells and direction of changes of second messengers in immune cells is the same or could be the opposite to that in neuronal cells, depending upon experimental conditions[28]. NF-kB was one of the common transcription factors induced by opioids in neuronal and immune cells[28]. In the presented research, gene *MAPK1* was enriched in opioid signaling pathway. It was indicated that there was a significant increase in *MAPK* expression and functional activity in human HCC, which was due to the important role of the *MAPK* pathway in cellular growth and differentiation, over expression of *MAPK* might be of critical importance to the formation and maintenance of human HCC[30]. Therefore, there might be a significant relationship between the opioid signaling pathway and HCV-cirrhosis with HCC.

In the process of the present research, we also tried to identify the significant pathways via the traditional method, which performed the pathway analysis of the differentially expressed genes (DEGs) based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. However, under the threshold values of $|\log FoldChange| > 2.0$ and p-value < 0.01, no DEGs could be identified. In other words, this traditional method was not suitable for this study. However, by the iPAS method, we successfully identified several pathways in HCV-cirrhosis with HCC, such as NF-kB is activated and signals survival, opioid signaling and pre-NOTCH processing in Golgi pathway, which were good indicators of HCV-cirrhosis with HCC. Therefore, this brand new method was extraordinarily suitable.

Conclusions

Altered pathways (such as NF-kB is activated and signals survival and opioid signaling) and hub genes (such as *UBC* and *MAPK1*) may play important roles in HCV-cirrhosis with concomitant HCC and are potentially novel predictive and prognostic markers for HCV-cirrhosis with HCC.

Acknowledgements

This research was supported by Department of Hepatopancreatobiliary and Spleen Surgery, Baoji Central Hospital. We thank all members of the research group. Meanwhile, we are grateful to Ji'nan Evidence Based Medicine Science-Technology Center for technical support during the data processing and analysis.

Conflict of Interest

The Authors declare that they have no conflict of interests.

Reference

1) Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. Hepatology 2011; 53: 1020-1022.
2) Morgan RL, Baack B, Smith BD, Yartel A, Pitasi M, Falck-Ytter Y. Eradication of hepatitis C virus infection and the development of hepatocellular carcinoma: a meta-analysis of observational studies. Ann Intern Med 2013; 158: 329-337.
3) Waxman S, Wurmbach E. De-regulation of common housekeeping genes in hepatocellular carcinoma. BMC Genomics 2007; 8: 243.
4) Archer KJ, Mas VR, David K, Maluf DG, Bornstein K, Fisher RA. Identifying genes for establishing a multigenic test for hepatocellular carcinoma surveillance in hepatitis C virus-positive cirrhotic patients. Cancer Epidemiol Biomarkers Prev 2009; 18: 2929-2932.
5) Wang Y, Li J, Chen J, Liu L, Peng Z, Ding J, Ding K. From cirrhosis to hepatocellular carcinoma in HCV-infected patients: genes involved in tumor progression. Eur Rev Med Pharmacol Sci 2012; 16: 995-1000.
6) Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 2012; 8: e1002375.
7) Breitling R, Amtmann A, Herzyk P. Iterative Group Analysis (iGA): a simple tool to enhance sensitivity and facilitate interpretation of microarray experiments. BMC Bioinformatics 2004; 5: 34.
8) Emmert-Streib F, Glazko GV. Pathway analysis of expression data: deciphering functional building blocks of complex diseases. PLoS Comput Biol 2011; 7: e1002053.

9) Ahn T, Lee E, Huh N, Park T. Personalized identification of altered pathways in cancer using accumulated normal tissue data. Bioinformatics 2014; 30: i422-429.

10) Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, Haussler D, Stuart JM. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics 2010; 26: i237-245.

11) Drier Y, Sheffer M, Domany E. Pathway-based personalized analysis of cancer. Proc Natl Acad Sci USA 2013; 110: 6388-6393.

12) Pepper SD, Saunders EK, Edwards LE, Wilson CL, Miller CJ. The utility of MAS5 expression summary and detection call algorithms. BMC Bioinformatics 2007; 8: 273.

13) Ma L, Robinson LN, Towle HC. ChREBP*Mlx is the principal mediator of glucose-induced gene expression in the liver. J Biol Chem 2006; 281: 28721-28730.

14) Rifai N, Ridker PM. Proposed cardiovascular risk assessment algorithm using high-sensitivity C-reactive protein and lipid screening. Clin Chem 2001; 47: 28-30.

15) Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. Bioinformatics 2009; 25:415-416.

16) Gehan EA. A generalized wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika 1965; 52: 203-223.

17) Nichols T, Hayasaka S. Controlling the familywise error rate in functional neuroimaging: a comparative review. Stat Methods Med Res 2003; 12: 419-446.

18) Junker BH, Schreiber F. Analysis of biological networks. John Wiley & Sons, 2011.

19) Brandes U, Erlebach T. Network analysis: methodological foundations. Springer Science & Business Media, 2005.

20) Koschützki D, Schreiber F. Centrality analysis methods for biological networks and their application to gene regulatory networks. Gene Regul Syst Bio 2008; 2: 193-201.

21) Mc Guire C, Prinz M, Beyaert R, van Loo G. Nuclear factor kappa B (NF-κB) in multiple sclerosis pathology. Trends Mol Med 2013; 19: 604-613.

22) Aleyasin H, Cregan SP, Iyirhiaro G, O'Hare MJ, Callaghan SM, Slack RS, Park DS. Nuclear factor-κB modulates the p53 response in neurons exposed to DNA damage. J Neurosci 2004; 24: 2963-2973.

23) Notarbartolo M, Poma P, Perri D, Dusonchet L, Cervello M, D'Alessandro N. Antitumor effects of curcumin, alone or in combination with cisplatin or doxorubicin, on human hepatic cancer cells. Analysis of their possible relationship to changes in NF-kB activation levels and in IAP gene expression. Cancer Lett 2005; 224: 53-65.

24) Jentsch S, Seufert W, Hauser HP. Genetic analysis of the ubiquitin system. Biochim Biophys Acta 1991; 1089: 127-139.

25) Cicinnati VR, Shen Q, Sotiropoulos GC, Radtke A, Gerken G, Beckebaum S. Validation of putative reference genes for gene expression studies in human hepatocellular carcinoma using real-time quantitative RT-PCR. BMC Cancer 2008; 8: 350.

26) Ryu KY, Maehr R, Gilchrist CA, Long MA, Bouley DM, Mueller B, Ploegh HL, Kopito RR. The mouse polyubiquitin gene UbC is essential for fetal liver development, cell-cycle progression and stress tolerance. EMBO J 2007; 26: 2693-2706.

27) McNicol ED, Midbari A, Eisenberg E. Opioids for neuropathic pain. Cochrane Database Syst Rev 2013; 8: CD006146.

28) Martin-Kleiner I, Balog T, Gabrilovac J. Signal transduction induced by opioids in immune cells: a review. Neuroimmunomodulation 2006; 13: 1-7.

29) Haddad JJ. On the enigma of pain and hyperalgesia: a molecular perspective. Biochem Biophys Res Commun 2007; 353: 217-224.

30) Gedaly R, Angulo P, Hundley J, Daily MF, Chen C, Evers BM. PKI-587 and sorafenib targeting PI3K/AKT/mTOR and Ras/Raf/MAPK pathways synergistically inhibit HCC cell proliferation. J Surg Res 2012; 176: 542-548.