

## Identification of hub genes involved in the development of hepatocellular carcinoma by transcriptome sequencing

Yongchang Zheng<sup>1,\*</sup>, Junyu Long<sup>1,\*</sup>, Liangcai Wu<sup>1,\*</sup>, Haohai Zhang<sup>1,\*</sup>, Lin Li<sup>2</sup>, Ying Zheng<sup>3</sup>, Anqiang Wang<sup>1</sup>, Jianzhen Lin<sup>1</sup>, Xiaobo Yang<sup>1</sup>, Xinting Sang<sup>1</sup>, Ke Hu<sup>4</sup>, Jie Pan<sup>5</sup> and Haitao Zhao<sup>1</sup>

<sup>1</sup>Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

<sup>2</sup>School of Life Sciences, Center for Synthetic and Systems Biology, Ministry of Education Key Laboratory of Bioinformatics, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Tsinghua University, Beijing, China

<sup>3</sup>State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Science, University of Macau, Macau, China

<sup>4</sup>Department of Radiation Oncology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

<sup>5</sup>Department of Radiology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

\*These authors have contributed equally to this work

Correspondence to: Ke Hu, email: huk@pumch.cn

Jie Pan, email: markpan6885@sina.com

Haitao Zhao, email: Zhaoh@pumch.cn

**Keywords:** hub gene, differentially expressed protein-coding genes, development, hepatocellular carcinoma, transcriptome sequencing

**Received:** April 25, 2017

**Accepted:** June 19, 2017

**Published:** July 22, 2017

**Copyright:** Zheng et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

Hepatocellular carcinoma (HCC) is a leading cause of cancer-related death. The aim of this study was to identify underlying hub genes and dysregulated pathways associated with the development of HCC using bioinformatics analysis. Differentially expressed protein-coding genes were subjected to transcriptome sequencing in 11 pairs of liver cancer tissue and matched adjacent non-cancerous tissue. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed, followed by protein-protein interaction (PPI) network construction. Hub genes were identified via centralities analysis and verified using published datasets. In total, 720 significantly differentially expressed protein-coding genes were identified in the samples, including 335 upregulated genes and 385 downregulated genes. The upregulated genes were significantly enriched in cell adhesion, biological adhesion and cell-cell adhesion GO terms under biological process (BP). Conversely, the downregulated genes were significantly enriched in embryonic organ morphogenesis, embryonic organ development and embryonic morphogenesis. The KEGG pathway analysis showed that the upregulated genes were enriched in ECM-receptor interaction and focal adhesion pathways. Furthermore, the downregulated genes were enriched in the ErbB, VEGF and MAPK signaling pathways. The PPI network and centralities analysis suggested that ITGA2 and 12 alternate genes were significant hub genes. These findings improve current understanding of the molecular

## **mechanisms underlying HCC development and may be helpful in identifying candidate molecular biomarkers for use in diagnosing, treating and monitoring the prognosis of HCC.**

### **INTRODUCTION**

Liver cancer is a leading cause of malignancy and mortality worldwide. Hepatocellular carcinoma (HCC), the most common malignancy in the liver, accounts for 85-90% of tumors derived from liver tissue [1]. Despite significant advances in early diagnosis and interventional therapies, there remains a need for novel management methods for advanced HCC [2].

HCC develops through a complicated biological process that involves several genomic changes and various pathogenic molecular mechanisms [3]. The process is slow and involves genomic alterations gradually changing the phenotype of liver cells to produce cellular intermediates that evolve to become cancer cells. Studies of the regulation of gene expression are helpful for understanding the pathogenesis of HCC [4]. RNA sequencing (RNA-Seq) has become useful for examining genetic changes on the whole-genome scale and produces large quantities of data [5]. A comprehensive systematic study of differentially expressed pathways and protein-coding gene interactions can more accurately identify the biological changes that occur during HCC carcinogenesis. In addition, protein-protein interaction (PPI) networks can be used to identify highly connected hub genes that play a key role in maintaining network structure [6]. Consequently, analyzing RNA-Seq data using these bioinformatics methodologies can help predict molecular pathogenesis and identify effective biomarkers of cancer.

HCC is a highly heterogeneous disease, and diverse changes in gene expression contribute to the progression of this cancer [7]. To obtain complete understanding of the alterations in gene expression that occur during HCC, RNA-Seq has been used to identify many key genes involved in disease progression. However, knowledge of the genetic changes that lead to HCC initiation and progression remains fragmented, and key drivers of carcinogenesis are still unknown, limiting the development of targeted therapy for HCC [8]. Furthermore, the overlap of the most significantly dysregulated genes among multiple studies is very low. Inconsistencies in results are caused by various factors, including measurement errors, small sample sizes and different statistical methods [9]. Therefore, understanding the pathogenesis of this disease remains a major challenge, and many hub genes must still be identified.

In the present study, we identified differentially expressed protein-coding genes from RNA transcriptional profiling performed on 11 paired cancer tissues and adjacent non-cancerous tissues. Then, we conducted GO, KEGG, PPI network and centralities analyses to study and identify changes in pathways and hub genes. The

aim of this study was to improve understanding of HCC carcinogenesis by providing information concerning the genetic changes that occur during disease progression and to uncover the expression of biomarkers with potential use for clinical diagnosis, treatment, and monitoring of disease progression.

### **RESULTS**

#### **Identification of differentially expressed protein-coding genes in HCC and adjacent non-cancerous tissues**

Fifteen pairs of samples from tumor and adjacent tissues were collected from patients enrolled in Peking Union Medical College Hospital from May 2015 to April 2016. All patients provided consent. RNA transcriptional profiling was performed on the tissue samples. Approximately 542.8 million reads were mapped to the human hg38 genome, with a mean of 15.7 M reads per sample (range 4.2-23.5 M) (Supplementary Table 1). Paired samples A2 and T2 were filtered out because of the small amount of data obtained. Multidimensional scaling analysis (MDS) was performed to show that the tumor samples were distinct from the adjacent non-cancerous tissue samples [10]. Four samples (A5, T5, T6, and T14) were found to be potentially misidentified and were excluded from further analysis (Supplementary Figure 1). Hence, transcriptome data from 11 paired samples was used to identify differentially expressed genes. We identified 720 significantly differentially expressed protein-coding genes, including 335 upregulated genes and 385 downregulated genes; a heatmap showing these genes can be found in Figure 1.

#### **Functional characterization of differentially expressed protein-coding genes**

To gain insight into the functional characteristics of the identified protein-coding genes, we conducted GO and KEGG pathway enrichment analyses [11]. The GO analysis results showed that the upregulated protein-coding genes were significantly enriched in the cell adhesion, biological adhesion and cell-cell adhesion categories under biological process (BP). Under molecular function (MF), the genes were enriched in glycosaminoglycan binding, polysaccharide binding and pattern binding. Additionally, GO cellular component (CC) analysis revealed genes significantly enriched in extracellular region part, extracellular region, and extracellular space. For the downregulated protein-coding genes, for BP, there was significant enrichment in embryonic organ

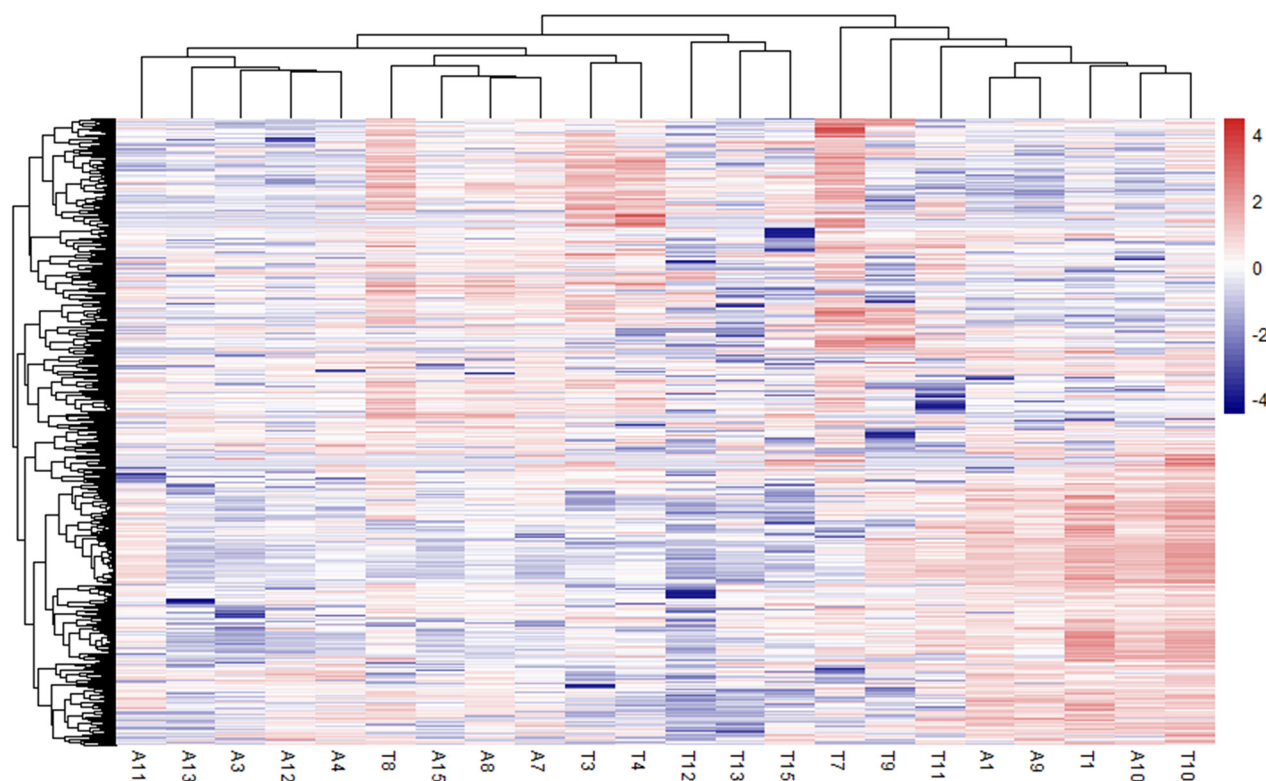
morphogenesis, embryonic organ development and embryonic morphogenesis. For MF, there was enrichment in sequence-specific DNA binding, transcription regulator activity and transcription factor activity. In addition, GO CC analysis revealed genes significantly enriched in cell-cell junction, apical junction complex and apicolateral plasma membrane (Figure 2 and Supplementary Table 2). Moreover, we found upregulated protein-coding genes that were significantly enriched in the ECM-receptor interaction and focal adhesion pathways using KEGG pathway enrichment analysis (Figure 2 and Supplementary Table 3). Meanwhile, downregulated protein-coding genes were enriched in the ErbB, VEGF, and MAPK signaling pathways (Figure 2 and Supplementary Table 3). We then compared these enriched genes with potential HCC driver genes from the Driver DB 2.0 database and identified 15 protein-coding genes that were recorded as HCC driver genes (triangle node) (Figure 2).

### PPI network construction and centralities analysis

PPI network analysis is a powerful tool for recognizing critical hub members among a cluster of molecules. Thus, we conducted PPI network analysis using the STRING database to identify critical

members among our protein-coding genes. The PPI networks for the upregulated and downregulated genes are shown in Figure 3 and Supplementary Figure 2, respectively.

Centralities represent the possibility that a gene is functionally able to keep communicating nodes together in a biological network. There are five types of centralities: degree centrality, betweenness centrality, stress centrality, closeness centrality and clustering coefficient. The five types of centralities were calculated based on the complex network, and we found that the distribution of the top 10 genes was not completely consistent among various centralities analyses. Therefore, the genes that were among the top 10 identified and that were shared more than twice among the five types of centralities were defined as hub genes. Using these criteria, 6 hub genes were obtained for the upregulated protein-coding genes, with the alpha 2 integrin gene (ITGA2), bone morphogenetic protein 4 (BMP4), PLCB1 and PRKG2 identified as the common hub genes across the degree, betweenness and stress centrality analyses. Moreover, SPINK6 and POU3F4 detected in two of the five types of analysis. In addition, 7 hub genes were identified for the downregulated protein-coding genes, with lysine (K)-specific demethylase 6B (KDM6B), GLI1, MYC, CSNK2A1 and POTEF identified as common hub genes across the degree, betweenness and stress centrality



**Figure 1: Heatmap showing significantly differentially expressed protein-coding genes among 11 paired HCC and adjacent non-cancerous tissue.** Rows represent genes, and columns represent samples.

analyses, and HSF1 and SCNN1A were commonly obtained in the degree and stress centrality analyses (Figure 4).

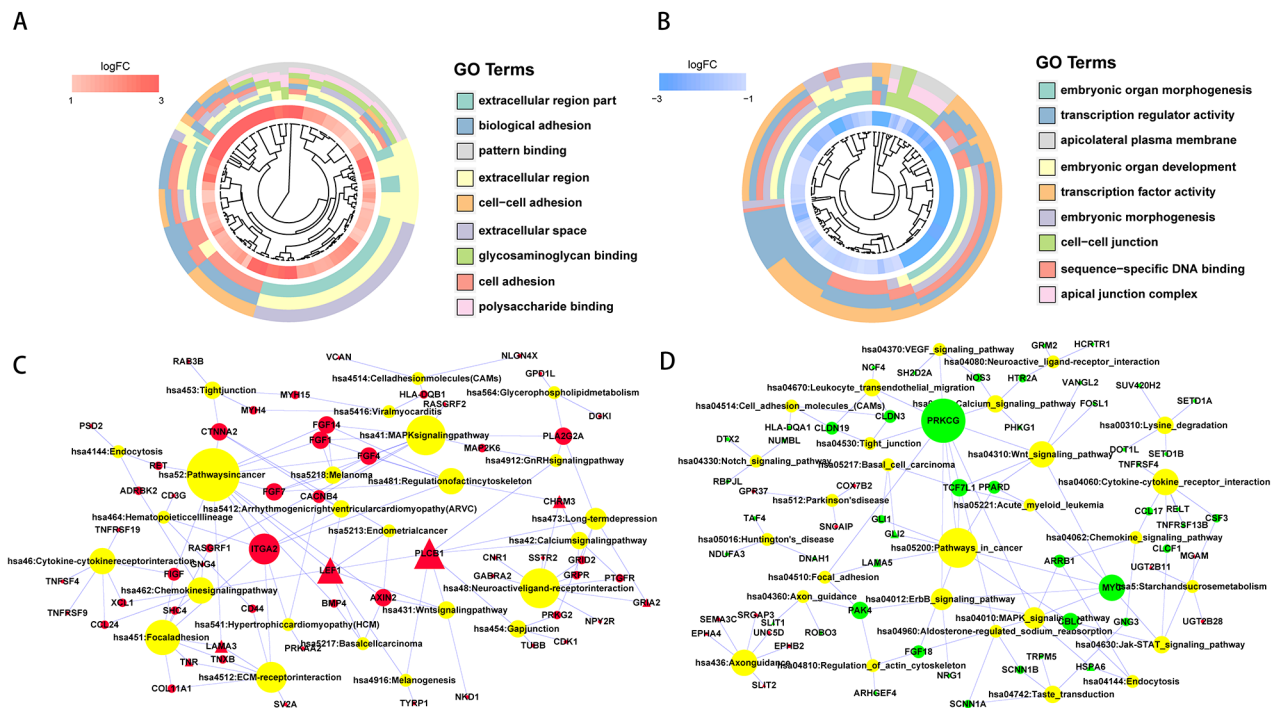
## The role of hub genes in the development of HCC

To further explore the functional roles of the identified critical hub molecules in the development of HCC, we evaluated changes in the expression spectra of the 13 hub genes using one-way ANOVA (Kruskal-Wallis test) in HBV-related HCC (GSE25097) and HCV-related HCC datasets (GSE6764). We did not identify SCNN1A and PLCB1 in the HBV-related HCC dataset or POTEF and MYC in the HCV-related HCC dataset. The transcriptome profiling data in the GSE25097 dataset contain 557 samples, including 6 healthy livers, 40 cirrhotic tissue samples, 243 adjacent non-tumor samples, and 268 early-to-advanced stage HCC samples. Moreover, the GSE6764 dataset contains a total of 75 tissue samples, including 10 healthy liver samples, 13 cirrhosis samples, 17 dysplastic nodules, 18 early HCC samples and 17 advanced HCC samples. Consistent with our results, ITGA2, BMP4 and PLCB1 were significantly upregulated, and KDM6B and MYC were significantly downregulated during HCC oncogenesis

(Figure 5 and Supplementary Figure 4). In contrast to our findings, GLI1, CSNK2A1, POTEF and HSF1 were upregulated and SPINK6 downregulated in the HCC cohort (Supplementary Figures 3 and 4). No significant differences were found in the expression of the other 3 genes (Supplementary Figures 3 and 4). Based on these results, the expression of ITGA2, BMP4 and PLCB1 may contribute to cancer development. ITGA2, BMP4, PLCB1, KDM6B and MYC represent the most likely diagnostic or therapeutic biomarkers associated with HCC.

## DISCUSSION

HCC is a common malignant tumor with a heterogeneous molecular pathogenesis that has not been fully elucidated. Identifying significantly dysregulated genes and pathways associated with HCC carcinogenesis may improve understanding of the molecular pathogenesis underlying HCC development and could identify potential biomarkers for treatment [4]. In this study, GO and KEGG pathway enrichment analysis showed that cell adhesion, biological adhesion and cell-cell adhesion were significant GO terms in upregulated protein-coding genes

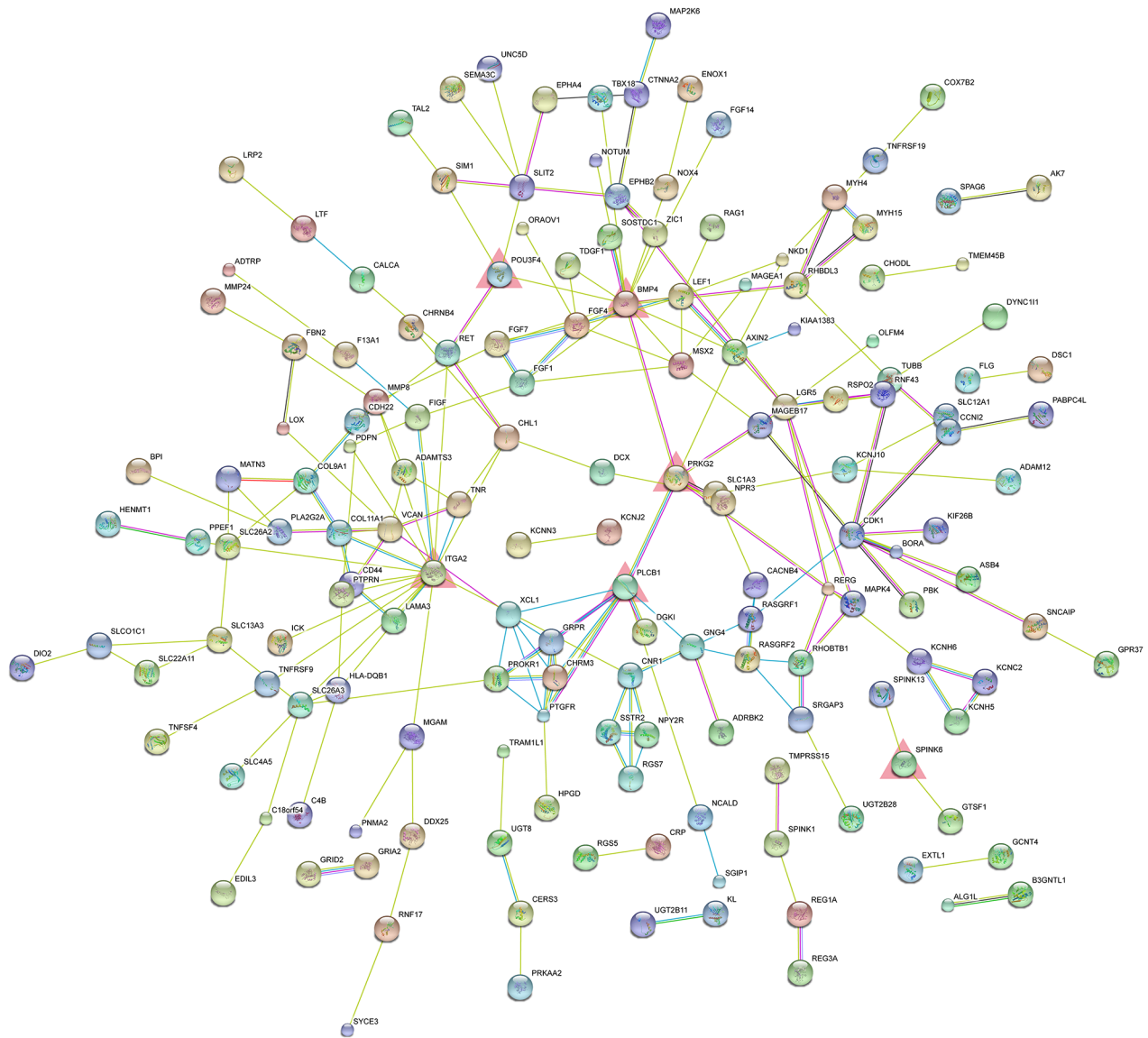


**Figure 2: Functional enrichment analysis of significantly upregulated and downregulated protein-coding genes.** GOcluster plot showing a circular dendrogram of the clustering of the expression spectrum. The inner ring indicates the color-coded logFC. Red represents significantly upregulated (A) and blue represents significantly downregulated (B) protein-coding genes. The outer ring displays the assigned functional terms. KEGG pathway enrichment of significantly upregulated (C) and downregulated (D) protein-coding genes. The red node represents significantly upregulated protein-coding genes. The green node represents significantly downregulated protein-coding gene. The yellow node represents enriched pathway symbols. The triangular node represents HCC driver genes from the Driver DB V2 database. The size of the node represents the number of genes.

and that embryonic organ morphogenesis, embryonic organ development and embryonic morphogenesis were significant GO terms in downregulated protein-coding genes. Moreover, ECM-receptor interactions and focal adhesion pathways were the relevant pathways associated with the upregulated protein-coding genes, and the ErbB, VEGF, and MAPK signaling pathways were the relevant pathways associated with the downregulated protein-coding genes.

Hub genes were identified based on degree centrality, betweenness centrality, stress centrality, closeness centrality and clustering coefficients in the PPI network. The outcomes across these five centralities were integrated to address inconsistent results provided

by the different methods of analysis. A total of 6 hub genes were identified among the upregulated protein-coding genes, and 7 hub genes were obtained among the downregulated protein-coding genes. ITGA2, BMP4, PLCB1, PRKG2, KDM6B, GLI1, MYC, CSNK2A1 and POTEF were common hub genes across three centrality methods. In addition, the expression spectra of the 13 hub genes were verified in HCC tumorigenesis using GEO datasets GSE25097 and GSE6764. ITGA2, BMP4, and PLCB1 were significantly upregulated, and KDM6B and MYC were significantly downregulated during HCC oncogenesis, consistent with our results. In contrast, GLI1, CSNK2A1, POTEF and HSF1 were upregulated and SPINK6 downregulated in the HCC cohort. No

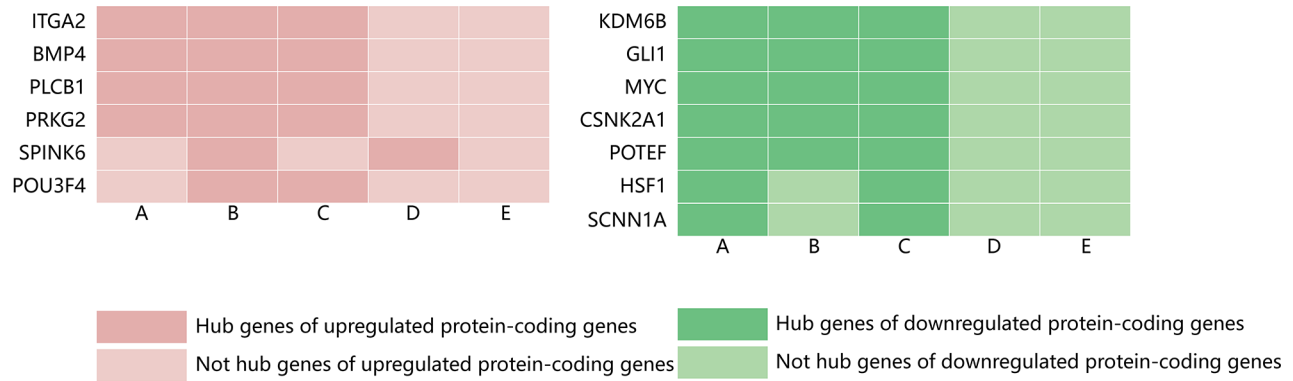


**Figure 3: PPI network of significantly upregulated protein-coding genes. The nodes represent the significantly upregulated protein-coding genes. The edges represent the interaction of significantly upregulated protein-coding genes. The red triangles represent the significantly upregulated hub genes.**

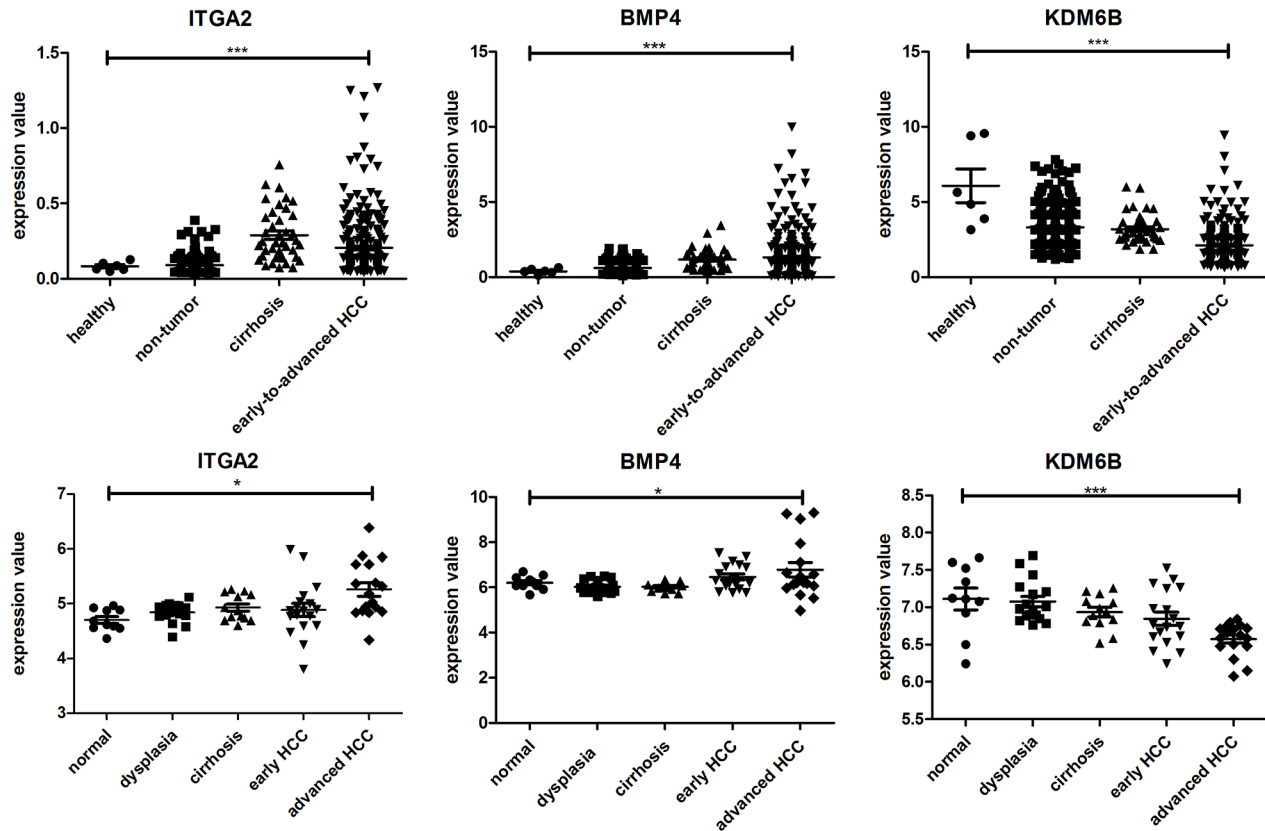
significant differences were found in the expression of the other 3 genes. BMP4 and ITGA2 were identified as critical hub molecules among the upregulated protein-coding genes and had many interactions with their neighbors. KMD6B was considered to be a hub molecule due to its high degree of downregulation when compared to all protein coding genes. Therefore,

we mainly discuss the function of these top degree centrality genes.

KDM6B is a demethylase that acts on histone H3 at lysine 27 (H3K27), which specifically catalyzes the demethylation of H3 lysine-27 tri/di methylation (H3K27me3/2) [12]. KDM6B has been shown to play a key role in promoting transcription elongation associated with related elongation factors and RNA polymerase II



**Figure 4: Distribution of hub genes among the significantly upregulated and downregulated protein-coding genes identified by five types of centrality. (A) Degree centrality; (B) betweenness centrality; (C) stress centrality; (D) closeness centrality; and (E) clustering coefficient.**



**Figure 5: Dynamic expression of ITGA2, BMP4 and KDM6B in HBV-related HCC (up) and HCV-related HCC (down);  $p < 0.01$  (\*),  $p < 0.001$  (\*\*), and  $p < 0.0001$  (\*\*\*).**

to activate gene transcription [13, 14]. Growing evidence suggests that KDM6B regulates cancer development. However, the effect of KDM6B on tumorigenesis is not consistent among different human cancers. Compared with adjacent normal tissues, KDM6B expression was significantly increased in renal cell carcinoma [15]. Overexpression of KDM6B has also been shown to promote invasion-metastasis cascades and induce the expression of mesenchymal genes in breast cancer [16]. In contrast, KDM6B knockdown promotes tumor sphere formation, increases hepatic metastasis and peritoneal dissemination in pancreatic adenocarcinoma, and enhances epithelial-mesenchymal transition and invasiveness in colon cancer cells [17, 18]. Similarly, the knockdown of KDM6B inhibits cell apoptosis and promotes cell growth by reducing the nuclear translocation of FOXO1 in non-small cell lung cancer cells [19]. However, literature focused on KDM6B expression in HCC is lacking. In this study, KDM6B was identified as a hub gene among the downregulated protein-coding genes differentially expressed throughout HCC development.

The integrin family of  $\alpha/\beta$  heterodimeric transmembrane receptors mediates the binding of cells to the extracellular matrix (ECM) and the intracellular signaling events that occur within the ECM [20]. ITGA2 encodes the integrin  $\alpha 2$  subunit of the  $\alpha 2\beta 1$  integrin receptor, which is a cell adhesion molecule [20]. Previous studies have shown that the ITGA2 gene is associated with various types of cancer, including colorectal cancer, prostate cancer, hepatocellular carcinoma, pancreatic cancer, breast cancer, melanoma and ovarian carcinoma. ITGA2 is expressed in colon cancer cell lines and is involved in the proliferation and migration of these cells [21]. Slambrouk *et al.* reported that integrin  $\alpha 2$  subunits interact with the ECM protein collagen I, increasing prostate cancer cell adhesion and cellular invasion [22]. Integrin  $\alpha 2$  subunits also downregulate E-cadherin-mediated cell-cell adhesion architecture, enhancing pancreatic cancer cell invasiveness, migration and proliferation [23]. In addition, integrin  $\alpha 2\beta 1$  inhibits mammalian sterile 20-like kinase 1 kinase (MST1) phosphorylation and activates Yes-associated protein (YAP) oncogenic signaling in HCC [24]. It has been previously shown that ITGA2 is enriched in ECM-receptor interactions and pathways in cancer. However, the impact of ITGA2 on the progression of HCC is still unknown. Therefore, it seems necessary to investigate and clarify the basic biological links between ITGA2 and HCC.

BMP4 belongs to the transforming growth factor (TGF- $\beta$ ) superfamily of extracellular signaling molecules, which is of great relevance both during development and in adult tissues [25, 26]. BMP signaling regulates early liver development and promotes liver bud morphogenesis as well as the migration, proliferation and survival of hepatoblasts

[27, 28]. Interestingly, BMP4 has been shown to participate in human carcinogenesis. Recent findings have revealed that BMP4 is overexpressed in breast cancer and may promote cell invasion and migration by modulating TGF- $\beta$  factor signaling [29]. In addition, BMP4 signaling causes direct overexpression of ID3, a proto-oncogene that contributes to the pathogenesis of human ovarian cancer [30]. BMP4 is upregulated in HCC, and the overexpression of BMP4 promotes the metastasis and proliferation of HCC cells by activating the mitogen-activated protein/extracellular signal-regulated kinase (MEK)/extracellular signal-regulated kinase (ERK) signaling pathway [31]. BMP4 was enriched in KEGG pathways associated with pathways in cancer and basal cell carcinoma in this study. Therefore, BMP4 and its associated pathways may be used as a biological indicator for the development of various types of cancers, including HCC.

Although we identified and verified hub genes important for the development of HCC using comprehensive bioinformatics technology, there are some limitations to the present study. First, this study lacked experimental validation of genes and their functions in HCC carcinogenesis. Second, we cannot rule out the possibility that these key genes may be involved in non-developmental aspects of HCC. Finally, the sample size for the RNA-Seq was small, which may have led to a high rate of false-positive outcomes. Hence, a larger sample size is needed for further bioinformatics analysis, and experimental studies are necessary to validate our results.

In conclusion, we identified several hub genes and systematically presented the biological processes and signaling pathways associated with the development of HCC. Many of these genes were not previously reported but could play important roles in HCC. Further research is required to focus on the clinical application of these genes and pathways for diagnosing, treating and monitoring the prognosis of HCC.

## MATERIALS AND METHODS

### Ethics

This study was approved by the Clinical Research Ethics Committee of Peking Union Medical College Hospital. Written informed consent with a signature was obtained from each patient.

### Clinical samples

We collected tissue samples from patients with liver cancer undergoing surgery at Peking Union Medical College Hospital, Beijing. The samples were collected in pairs, i.e., cancer tissue and adjacent non-cancerous tissue. The collected tissue samples were stored in liquid nitrogen.

## RNA preparation and sequencing

First, 50 mg of tissue was lysed in TRIzol (Invitrogen) to extract RNA following the manufacturer's instructions. Next, ribosomal RNA was depleted using a RiboZero Gold kit (Epicentre Bio-technologies). RNA integrity was assessed with an Agilent Bioanalyzer 2100. An RNA-Seq library was generated with the rRNA-depleted samples using an Illumina standard RNA Sample Prep kit according to the manufacturer's instructions. The library was subsequently sequenced on an Illumina HiSeq2500 as 125-bp paired-ends with approximately 300-bp size selection.

## Published dataset and database

Liver cancer-related RNA-Seq data are available from the NCBI GEO database, including two independent microarray datasets: HBV-related HCC (GSE25097) and HCV-related HCC (GSE6764). HCC-related driver genes were obtained from the Driver DB 2.0 database [32].

## Transcriptome sequencing analysis

Sequencing quality was assessed with FASTQC [33]. After removing adaptor and low-quality reads using cutadapt [34] (-q 10--quality-base=32 -e 0.1 -O 10 -m 50), the clean reads were aligned to human (hg38) GRCh38.p5 ([http://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001405.31](http://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.31)) genome reference sequences using Tophat2 [35] (-a 6 --microexon-search -m 2); bam files were generated and sorted, and duplicate reads were then removed using Samtools [36]. Read counts were tabulated with HT-Seq [37] using "union" mode and the Gencode human v24 GTF file as a reference. edgeR was used to identify the small library size, make MDS plots of the samples, and check reproducibility from replicates to remove small sample sizes and outliers and check for batch effects [38]. Significantly differential gene expression between tumor and adjacent non-cancerous tissue was estimated, with a minimum twofold change and FDR less than 0.01, using the edgeR package from Bioconductor. A heatmap was plotted using the pheatmap package. Cufflinks [39] (v2.21) was also used to estimate the total transcriptional output based on the Gencode gene annotation for human HG38 (version 24) [40].

## Gene ontology and KEGG pathway enrichment analysis

We performed Gene Ontology and KEGG pathway enrichment analysis using The Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.7 (<https://david-d.ncifcrf.gov/home.jsp>) [11]. Unique lists of significantly differentially expressed protein-coding genes and all the expressed genes (FPKM >0 in any sample) were submitted to the web interface as the gene list and background, respectively. Enrichment results were visualized using R and Cytoscape 3.5.0 software [41]. In addition, we

compared these enriched genes with potential HCC driver genes from the Driver DB 2.0 database [32].

## PPI network analysis

Proteins rarely perform their functions independently, and it is therefore important to investigate protein interactions by studying larger functional groups [42]. Genes that were identified as significantly upregulated or downregulated were mapped to the STRING (Search Tool for the Retrieval of Interacting Genes) version 10.0 database (<http://www.string-db.org/>) and used to evaluate PPI information and construct a PPI network [6]. The STRING database covers 9.6 million proteins from 2031 organisms. In the PPI network, each node represents a gene, and the edges stand for interactions between nodes.

## Centralities analysis of the PPI network

Researchers have revealed strong correlations between PPI networks and the functions of protein/gene components [43]. Topological centrality is effective for identifying molecules that may play important roles in significantly perturbed networks [44]. The PPI data were downloaded from the STRING database. Centralities analysis was performed using Cytoscape 3.5.0 software. We presented the centralities of the PPI network on a local (degree and clustering coefficient) and global scale (betweenness, closeness and stress). Genes identified in the top 10 genes and that were shared more than twice among the five types of centralities were defined as hub genes.

## Data submission

Sequence data has been deposited at the European Genome-phenome Archive (EGA) (<https://ega-archive.org>), which is hosted by European Bioinformatics Institute (EBI) and the Centre for Genomic Regulation (CRG), under accession number EGAS00001002526.

## ACKNOWLEDGMENTS

This work was supported by International Science and Technology Cooperation Projects (2015DFA30650 and 2016YFE0107100), the Capital Special Research Project for Clinical Application (Z151100004015170), the Capital Special Research Project for Health Development (2014-2-4012), the Beijing Nature Science Foundation for Young Scholars Project (7164293), the Beijing Municipal Science & Technology Commission (Z171100000417004) and the Program for New Century Excellent Talents in University (NCET-11-0288).

## CONFLICTS OF INTEREST

None of the authors have any conflicts of interest to disclose.



## REFERENCES

1. Kimhofer T, Fye H, Taylor-Robinson S, Thursz M, Holmes E. Proteomic and metabolomic biomarkers for hepatocellular carcinoma: a comprehensive review. *Br J Cancer*. 2015; 112:1141-1156.
2. Shibata C, Otsuka M, Kishikawa T, Ohno M, Yoshikawa T, Takata A, Koike K. Diagnostic and therapeutic application of noncoding RNAs for hepatocellular carcinoma. *World J Hepatol*. 2015; 7:1-6.
3. Whittaker S, Marais R, Zhu AX. The role of signaling pathways in the development and treatment of hepatocellular carcinoma. *Oncogene*. 2010; 29:4989-5005.
4. Wang XW, Hussain SP, Huo TI, Wu CG, Forgues M, Hofseth LJ, Brecht C, Harris CC. Molecular pathogenesis of human hepatocellular carcinoma. *Toxicology*. 2002; 181-182:43-47.
5. Maekawa S, Suzuki A, Sugano S, Suzuki Y. RNA sequencing: from sample preparation to analysis. *Methods Mol Biol*. 2014; 1164:51-65.
6. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015; 43:D447-D452.
7. Jeng KS, Chang CF, Jeng WJ, Sheen IS, Jeng CJ. Heterogeneity of hepatocellular carcinoma contributes to cancer progression. *Crit Rev Oncol Hematol*. 2015; 94:337-347.
8. Nault JC, Zucman-Rossi J. Genetics of hepatocellular carcinoma: the next generation. *J Hepatol*. 2014; 60:224-226.
9. Liang D, Han G, Feng X, Sun J, Duan Y, Lei H. Concerted perturbation observed in a hub network in Alzheimer's disease. *PLoS One*. 2012; 7:e40498.
10. Kadota PO, Hajjiri Z, Finn PW, Perkins DL. Precision subtypes of T cell-mediated rejection identified by molecular profiles. *Front Immunol*. 2015; 6:536.
11. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44-57.
12. Agger K, Cloos PA, Christensen J, Pasini D, Rose S, Rappsilber J, Issaeva I, Canaani E, Salcini AE, Helin K. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature*. 2007; 449:731-734.
13. Chen S, Ma J, Wu F, Xiong LJ, Ma H, Xu W, Lv R, Li X, Villen J, Gygi SP, Liu XS, Shi Y. The histone H3 Lys 27 demethylase JMJD3 regulates gene expression by impacting transcriptional elongation. *Genes Dev*. 2012; 26:1364-1375.
14. Estaras C, Fueyo R, Akizu N, Beltran S, Martinez-Balbas MA. RNA polymerase II progression through H3K27me3-enriched gene bodies requires JMJD3 histone demethylase. *Mol Biol Cell*. 2013; 24:351-360.
15. Shen Y, Guo X, Wang Y, Qiu W, Chang Y, Zhang A, Duan X. Expression and significance of histone H3K27 demethylases in renal cell carcinoma. *BMC Cancer*. 2012; 12:470.
16. Ramadoss S, Chen X, Wang CY. Histone demethylase KDM6B promotes epithelial-mesenchymal transition. *J Biol Chem*. 2012; 287:44508-44517.
17. Yamamoto K, Tateishi K, Kudo Y, Sato T, Yamamoto S, Miyabayashi K, Matsusaka K, Asaoka Y, Ijichi H, Hirata Y, Otsuka M, Nakai Y, Isayama H, et al. Loss of histone demethylase KDM6B enhances aggressiveness of pancreatic cancer through downregulation of C/EBPalpha. *Carcinogenesis*. 2014; 35:2404-2414.
18. Pereira F, Barbachano A, Silva J, Bonilla F, Campbell MJ, Munoz A, Larriba MJ. KDM6B/JMJD3 histone demethylase is induced by vitamin D and modulates its effects in colon cancer cells. *Hum Mol Genet*. 2011; 20:4655-4665.
19. Ma J, Wang N, Zhang Y, Wang C, Ge T, Jin H, Deng X, Huo X, Gu D, Ge Z, Chu W, Jiang L, Qin W. KDM6B elicits cell apoptosis by promoting nuclear translocation of FOXO1 in non-small cell lung cancer. *Cell Physiol Biochem*. 2015; 37:201-213.
20. Gurkan A, Emingil G, Afacan B, Berdeli A, Atilla G. Alpha 2 integrin gene (ITGA2) polymorphism in renal transplant recipients with and without drug induced gingival overgrowth. *Arch Oral Biol*. 2014; 59:283-288.
21. Beaulieu JF. Integrins and human intestinal cell functions. *Front Biosci*. 1999; 4:D310-D321.
22. Van Slambrouck S, Jenkins AR, Romero AE, Steelant WF. Reorganization of the integrin alpha2 subunit controls cell adhesion and cancer cell invasion in prostate cancer. *Int J Oncol*. 2009; 34:1717-1726.
23. Grzesiak JJ, Ho JC, Moossa AR, Bouvet M. The integrin-extracellular matrix axis in pancreatic cancer. *Pancreas*. 2007; 35:293-301.
24. Wong KF, Liu AM, Hong W, Xu Z, Luk JM. Integrin  $\alpha 2 \beta 1$  inhibits MST1 kinase phosphorylation and activates Yes-associated protein oncogenic signaling in hepatocellular carcinoma. *Oncotarget*. 2016; 7:77683-77695. <https://doi.org/10.18632/oncotarget.12760>.
25. Carreira AC, Lojudice FH, Halcsik E, Navarro RD, Sogayar MC, Granjeiro JM. Bone morphogenetic proteins: facts, challenges, and future perspectives. *J Dent Res*. 2014; 93:335-345.
26. Rider CC, Mulloy B. Bone morphogenetic protein and growth differentiation factor cytokine families and their protein antagonists. *Biochem J*. 2010; 429:1-12.
27. Rossi JM, Dunn NR, Hogan BL, Zaret KS. Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. *Genes Dev*. 2001; 15:1998-2009.

28. Zaret KS, Grompe M. Generation and regeneration of cells of the liver and pancreas. *Science*. 2008; 322:1490-1494.
29. Pal A, Huang W, Li X, Toy KA, Nikolovska-Coleska Z, Kleer CG. CCN6 modulates BMP signaling via the Smad-independent TAK1/p38 pathway, acting to suppress metastasis of breast cancer. *Cancer Res*. 2012; 72:4818-4828.
30. Shepherd TG, Theriault BL, Nachtigal MW. Autocrine BMP4 signalling regulates ID3 proto-oncogene expression in human ovarian cancer cells. *Gene*. 2008; 414:95-105.
31. Chiu CY, Kuo KK, Kuo TL, Lee KT, Cheng KH. The activation of MEK/ERK signaling pathway by bone morphogenetic protein 4 to increase hepatocellular carcinoma cell proliferation and migration. *Mol Cancer Res*. 2012; 10:415-427.
32. Chung IF, Chen CY, Su SC, Li CY, Wu KJ, Wang HW, Cheng WC. DriverDBv2: a database for human cancer driver gene research. *Nucleic Acids Res*. 2016; 44:D975-D979.
33. Kroll KW, Mokaram NE, Pelletier AR, Frankhouser DE, Westphal MS, Stump PA, Stump CL, Bundschuh R, Blachly JS, Yan P. Quality control for RNA-Seq (QuaCRS): an integrated quality control pipeline. *Cancer Inform*. 2014; 13:7-14.
34. Chen C, Khaleel SS, Huang H, Wu CH. Software for pre-processing Illumina next-generation sequencing short read sequences. *Source Code Biol Med*. 2014; 9:8.
35. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009; 25:1105-1111.
36. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009; 25:2078-2079.
37. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015; 31:166-169.
38. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26:139-140.
39. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010; 28:511-515.
40. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012; 22:1760-1774.
41. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13:2498-2504.
42. Srihari S, Leong HW. Temporal dynamics of protein complexes in PPI networks: a case study using yeast cell cycle dynamics. *BMC Bioinformatics*. 2012; 13 Suppl 17:S16.
43. Choi JK, Yu U, Yoo OJ, Kim S. Differential coexpression analysis using microarray data and its application to human cancer. *Bioinformatics*. 2005; 21:4348-4355.
44. Prifti E, Zucker JD, Clement K, Henegar C. Interactional and functional centrality in transcriptional co-expression networks. *Bioinformatics*. 2010; 26:3083-3089.