Annals of Hepatology

ORIGINAL ARTICLE

The Official Journal of the Mexican Association of Hepatology, the Latin-American Association for Study of the Liver and the Canadian Association for the Study of the Liver

CUL4B, *NEDD4*, and *UGT1As* involve in the TGF-β signalling in hepatocellular carcinoma

Zhaowei Qu,* Di Li,** Haitao Xu,* Rujia Zhang,*** Bing Li,* Chengming Sun,* Wei Dong,* Yubao Zhang*

* Department of Hepatobiliary and Pancreatic Surgery, The Affiliated Tumor Hospital of Harbin Medical University, Harbin 150081, Heilongjiang Province, China.
** Department of Anesthesiology, The Second Affiliated Hospital of Harbin Medical University, Harbin 150086, Heilongjiang Province, China.
*** Operating Room, The Second Affiliated Hospital of Harbin Medical University, Harbin 150086, Heilongjiang Province, China.

ABSTRACT

Introduction and Aim. TGF-β signalling is involved in pathogenesis and progress of hepatocellular carcinoma (HCC). This bioinformatics study consequently aims to determine the underlying molecular mechanism of TGF-β activation in HCC cells. **Material and methods.** Dataset GSE10393 was downloaded from Gene Expression Omnibus, including 2 Huh-7 (HCC cell line) samples treated by TGF-β (100 pmol/L, 48 h) and 2 untreated samples. Differentially expressed genes (DEGs) were screened using Limma package (false discovery rate < 0.05 and $|\log_2$ fold change$| > 1.5$), and then enrichment analyses of function, pathway, and disease were performed. In addition, protein-protein interaction (PPI) network was constructed based on the PPI data from multiple databases including INACT, MINT, BioGRID, UniProt, BIND, BindingDB, and SPIKE databases. Transcription factor (TF)-DEG pairs (Bonferroni adjusted p-value < 0.01) from ChEA database and DEG-DEG pairs were used to construct TF-DEG regulatory network. Furthermore, TF-pathway-DEG complex network was constructed by integrating DEG-DEG pairs, TF-DEG pairs, and DEG-pathway pairs. **Results.** Totally, 209 DEGs and 30 TFs were identified. The DEGs were significantly enriched in adhesion-related functions. PPI network indicted hub genes such as *CUL4B* and *NEDD4*. According to the TF-DEG regulatory network, the two hub genes were targeted by SMAD2, SMAD3, and HNF4A. Besides, the 11 pathways in TF-pathway-DEG network were mainly enriched by *UGT1A* family and *CYP3A7*, which were predicted to be regulated by SMAD2, SMAD3, SOX2, TP63, and HNF4A. **Conclusions.** TGF-β might influence biological processes of HCC cells via SMAD2/SMAD3-*NEDD4*, HNF4A-*CUL4B/NEDD4*, SOX2/TP63/HNF4A-*CYP3A7*, and SMAD2/SMAD3/SOX2/TP63/HNF4A-*UGT1As* regulatory pathways.

Key words. Hepatocellular carcinoma. Differentially expressed genes. Transcription factors. Protein-protein interaction network. Transcription factor-target regulatory network.

INTRODUCTION

Hepatocellular carcinoma (HCC) is the primary liver cancer that mainly occurs in developing countries.[1] The worldwide incidence of HCC ranks the seventh in women (226,000 cases per year, 6.5% of all cancers) and the fifth in men (523,000 cases per year, 7.9% of all cancers).[2] HCC is the main form of malignant liver cancer. Due to the high occurrence of invasion to intrahepatic large vessels and stomach, the 5-year survival rate for patients with HCC remains poor after major resection.[3]

In recent years, many studies have surveyed the pathogenesis of HCC. For example, Wnt signaling pathway which is associated with cell differentiation, proliferation, apoptosis, motility, and homeostasis is aberrantly regulated in HCC.[4] Cytokines like transforming growth factor β (TGF-β) play important roles in HCC progression and invasion. Reportedly, TGF-β is a multifunctional factor and plays bipartite roles in HCC:[5,6]

- TGF-β suppresses tumor formation at early stage of liver damage.
- TGF-β promotes HCC progression by inducing microenvironment changes.

Moreover, TGF-β stimulation induces epithelial-mesenchymal transitions (EMT) in malignant cancers, promoting cancer cell migration and invasion.[7] Hoshida, *et al*.

Manuscript received: August 06, 2015. **Manuscript accepted:** October 06, 2015.

DOI:10.5604/16652681.1203154

have found that TGF-β treatment (100 pmol/L, 48 h) enhances Wnt signaling via the intracellular pool of free β-catenin in HCC cell line.[8] Besides, transcription factors (TFs) play crucial roles in TGF-β stimulation via regulating gene transcription. For example, the modulation of SMAD family member 7 (SMAD7) expression influences the sensitivity of HCC cell lines (Huh7, FLC-4, HLE and HLF) for cytostatic TGF-β effects, while the knockdown of TF Y-box binding protein-1 (YB-1) reduces TGF-β-induced SMAD7 expression in Huh7 cells.[9] Despite these encouraging findings, the mechanism of TGF-β action in HCC cells has not been clearly illustrated.

This bioinformatics study was conducted to comprehensively determine the underlying molecular mechanism of TGF-β activation in HCC cells by using the gene expression profile up-loaded by Hoshida, *et al*. Consequently, differentially expressed genes (DEGs) between HCC cells with and without TGF-β treatment were identified, DEG functions were investigated, and TF-pathway-DEG complex network was constructed. The results of this study might provide novel directions for further HCC studies and therapeutic targets for HCC treatment.

MATERIAL AND METHODS

Microarray data

Gene expression profile of GSE10393 deposited by Hoshida, *et al.*[8] was downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database. Totally, GSE10393 included 2 Huh-7 (HCC cell line) samples with TGF-β treatment (100 pmol/L, 48 h) and 2 Huh-7 samples without TGF-β treatment. The corresponding platform was GPL3921 [HT_HG-U133A] Affymetrix HT Human Genome U133A Array.

Data preprocessing

Based on the R statistical programming language, Bioconductor provides 1024 open software packages for the analysis of high-throughput genomic data. Therefore, microarray data GSE10393 was preprocessed by using Bioconductor package series (version 3.0, available at: http://www.bioconductor.org/packages/3.0/bioc/),[10] including affy, simpleaffy, gcrma, and genefilter. In this study, affy package was used to read the raw data of microarray, and simpleaffy package was utilized to detect and control the quality of microarray data. Then, gcrma package was used to adjust the background of different chips, normalize the data in different chips, and transfer the expression values into $\log_2$ values. The nsFilter function in genefilter package was utilized to delete the probes with no or little expression value. Moreover, we transformed the expression values at probe level into expression values at gene level based on the annotation files in platform GPL3921.

DEGs screening

Limma package has been widely used to identify the genes which are differentially expressed between two groups. Thus, DEGs between Huh-7 cells with and without TGF-β treatment were identified by using Limma package (version 3.22.7, available at: http://www.bioconductor.org/packages/3.0/bioc/html/limma.html).[11] The raw p-value produced in DEGs screening was adjusted into false discovery rate (FDR) by using Benjamini-Hochberg method.[12] Then, FDR < 0.05 and $|\log_2$ fold change (FC)$| > 1.5$ were used as the cut-off criteria (Figure 1).

Enrichment analyses

Gene Ontology (GO) terms in GO database and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in KEGG database are usually used to describe the biological processes, molecular functions, and subcellular locations of protein.[13] Functional enrichment analysis has been widely used to identify the functions that involve the genes in a gene set, while pathway enrichment analysis is generally utilized to find the pathways that include the genes in a gene set. For the identified DEGs, functional and pathway enrichment analyses were performed by using clusterProfiler package (version 2.0.1, available at: http://www.bioconductor.org/packages/3.0/bioc/html/clusterProfiler.html)[14] based on GO and KEGG databases. Besides, Disease Ontology (DO) terms can provide the biomedical community with consistent, reusable and sustainable descriptions of human disease terms, phenotype characteristics, and related medical vocabulary disease concepts.[15] Similarly, DO term enrichment analysis was conducted by using DOSE package (version 2.4.0, available at: http://www.bioconductor.org/packages/3.0/bioc/html/DOSE.html).[15] For these analyses, FDR < 0.05 was used as the cut-off criterion.

$$FC = \frac{\text{Expression value in Huh-7 cells with TGF-}\beta\text{ treatment}}{\text{Expression value in Huh-7 cells without TGF-}\beta\text{ treatment}}$$

Figure 1. *The calculation formula for fold change of genes. FC: fold change.*

Protein-protein interaction (PPI) network and its modules

Proteins coded by genes often play their biological functions via interacting with other proteins. Thus, PPIs associated with the identified DEGs were extracted from the online databases, including the INDECT project Advanced Image Catalogue Tool (INACT),[16] Molecular INTeraction (MINT),[17] Biological General Repository for Interaction Datasets (BioGRID),[18] Universal Protein (UniProt),[19] Biomolecular Interaction Network Database (BIND),[20] protein-ligand Binding DataBase (BindingDB)[21] and Signaling Pathways Integrated Knowledge Engine (SPIKE)[22] databases. Then, PPI network was conducted and visualized by using Cytoscape software (version 3.2.1, available at: http://cytoscape.org/).[23] Furthermore, the algorithm GLay[24] of clusterMaker plugin[25] in Cytoscape was used to screen modules in the PPI network. The nodes in the PPI network and its modules included both DEGs and other non-DEGs that directly interacted with these DEGs.

TF-DEG regulatory network and its modules

Chromatin Immunoprecipitation (ChIP) Enrichment Analysis (ChEA) database records huge quantities of information about TFs and their targets based on genome-wide experiments like ChIP-chip, ChIP-sequencing, ChIP coupled with pair-end ditag sequencing analysis (ChIP-PET) and DNA adenine methylation identification (Dam-ID).[26] In this study, ChEA database was used to search the TFs which could regulate the identified DEGs. Then, TF enrichment analysis was performed, and the raw p-values were adjusted by using Bonferroni method.[27] For this analysis, the adjusted p-value < 0.01 was set as the cut-off criterion. Based on TF-DEG pairs found in this analysis and DEG-DEG pairs in PPI network, the TF-DEG regulatory network was constructed. Moreover, modules in the TF-DEG regulatory network were also screened out by using clusterMaker plugin[25] in Cytoscape. The nodes in the TF-DEG regulatory network and its modules included only DEGs and TFs.

TF-pathway-DEG complex network

TF-pathway-DEG complex network was constructed by integrating the DEG-DEG pairs from PPI network, TF-DEG pairs from ChEA database, and DEG-pathway pairs from KEGG pathway enrichment analysis. The nodes in the TF-pathway-DEG complex network included TFs, KEGG pathways, and DEGs.

RESULTS

DEGs

A total of 209 DEGs between Huh-7 cells with and without TGF-β treatment were identified, including 109 up-regulated DEGs and 100 down-regulated DEGs.

Enrichment analyses

Totally, 107 GO functions, 11 KEGG pathways, and 2 DO terms were enriched by the DEGs (Table 1). The top 10 enriched functions were mainly associated with cell adhesion and glucose metabolism, while the 11 enriched pathways were mainly associated with drug, vitamin, and saccharides metabolism.

PPI network and its modules

The PPI network included 2,679 nodes and 4,247 interactions, involving 166 DEGs (Figure 2A). In addition, the top 10 nodes with connectivity degree ≥ 120 in PPI network are shown in table 2. Nodes with higher degree and betweenness centrality have closer association with TGF-β activation in HCC. Therefore, the DEGs with high degree and betweenness centrality were defined as hub genes, which might relate to TGF-β stimulation closely. The hub genes included *VCAM1* (vascular cell adhesion molecule 1), *CUL4B* (cullin 4B), UBE2I (ubiquitin-conjugating enzyme E2I), *NEDD4* (neural precursor cell expressed, developmentally down-regulated 4), *POT1* (protection of telomeres 1 homolog), and so on. Furthermore, a total of 35 modules were obtained from the PPI network, among which the biggest module was ClusterID4 (Figure 2B). Module ClusterID4 consisted of 349 genes (including 58 DEGs) and 561 PPIs. The top 10 nodes in module ClusterID4 are shown in table 2. Hub gene *RPL31* (Ribosomal protein L31) in the PPI network was also the most significant hub gene in module ClusterID4.

TF-DEG regulatory network and its modules

A total of 30 TFs were significantly enriched by DEGs, of which the top 10 TFs are shown in table 3. Based on the DEG-DEG pairs from the PPI network and the TF-DEG pairs, the TF-DEG regulatory network was constructed, involving 226 nodes and 1086 interactions (Figure 3). In this network, the top 5 TFs with high connectivity degree were TP63 (transcription factors tumor protein p63), SOX2 (SRY-box 2), HNF4A (hepatocyte nuclear factor 4 a), SMAD3 (SMAD family member 3), and SMAD2 (SMAD family member 2). These 5 TFs

Table 1. The GO functions (top 10), KEGG pathways, and DO terms enriched by DEGs.

Term	Description	Count	Gene symbol	FDR
• GO function				
GO:0007156	Homophilic cell adhesion	23	*PCDHGA12, PCDHGC3, PCDHGC5, PCDHGC4, PCDHGB7, PCDHGB6, PCDHGB5, PCDHGB3, PCDHGB2, PCDHGB1,* etc.	5.91E-19
GO:0044699	Single-organism process	169	*NAMPT, MPHOSPH9, IRF9, TXNIP, LEFTY1, NUDT4, SLC6A14, CSNK1D, CSTA, CTH,* etc.	6.39E-15
GO:0008150	Biological process	183	*NAMPT, MPHOSPH9, IRF9, CEBPD, TXNIP, LEFTY1, NUDT4, SLC6A14, CSNK1D, CSTA,* etc.	8.88E-14
GO:0016337	Cell-cell adhesion	29	*CSTA, SLC7A11, PCDHGA12, APOA1, PCDHGC3, PDGFRA, PCDHGC5, PCDHGC4, PCDHGB7, PCDHGB6,* etc.	1.29E-13
GO:0052695	Cellular glucuronidation	8	*UGT1A10, UGT1A8, UGT1A7, UGT1A6, UGT1A9, UGT1A4, UGT1A1, UGT1A3,* etc.	2.15E-10
GO:0006063	Uronic acid metabolic process	8	*UGT1A10, UGT1A8, UGT1A7, UGT1A6, UGT1A9, UGT1A4, UGT1A1, UGT1A3,* etc.	2.75E-10
GO:0019585	Glucuronate metabolic process	8	*UGT1A10, UGT1A8, UGT1A7, UGT1A6, UGT1A9, UGT1A4, UGT1A1, UGT1A3,* etc.	2.75E-10
GO:0007155	Cell adhesion	37	*CSTA, SLC7A11, PCDHGA12, HABP2, APOA1, IL8, KNG1, RHOB, LGALS3BP, PCDHGC3,* etc.	3.04E-10
GO:0022610	Biological adhesion	37	*CSTA, SLC7A11, PCDHGA12, HABP2, APOA1, IL8, KNG1, RHOB, LGALS3BP, PCDHGC3,* etc.	3.04E-10
GO:0044763	Single-organism cellular process	151	*NAMPT, MPHOSPH9, IRF9, TXNIP, LEFTY1, NUDT4, SLC6A14, CSNK1D, CSTA, CTH,* etc.	3.85E-10
• KEGG pathway				
hsa00053	Ascorbate and aldarate metabolism	9	*UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9*	5.56E-11
hsa00860	Porphyrin and chlorophyll metabolism	10	*EPRS, UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9*	3.73E-10
hsa00040	Pentose and glucuronateinterconversions	9	*UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9*	4.64E-10
hsa00983	Drug metabolism - other enzymes	10	*UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9, CYP3A7*	2.77E-09
hsa00140	Steroid hormone biosynthesis	10	*UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9, CYP3A7*	7.11E-09
hsa00514	Other types of O-glycan biosynthesis	9	*UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9*	1.54E-08
hsa00830	Retinol metabolism	10	*UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9, CYP3A7*	2.29E-08
hsa00980	Metabolism of xenobiotics by cytochrome P450	10	*UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9, CYP3A7*	6.43E-08
hsa00500	Starch and sucrose metabolism	9	*UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9*	6.76E-08
hsa00982	Drug metabolism - cytochrome P450	10	*UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9, CYP3A7*	8.44E-08
hsa01100	Metabolic pathways	25	*ACSM3, ACSS3, ALDH6A1, ASNS, BCAT1, CTH, EPRS, PHGDH, PLA2G4A, PSAT1, SUCLG2,* etc.	1.96E-02
• DO term				
DOLite:537	Ulcerative colitis	7	*LEPR, NAMPT, TXNIP, CD14, HSPA1B, IGF2, IL8*	6.10E-03
DOLite:264	Hypertension	8	*ACSM3, CTH, HTR2C, KNG1, LEPR, ANPEP, LYZ, SGK1*	1.73E-02

GO: gene ontology. KEGG: Kyoto Encyclopedia of Genes and Genomes. DO: disease ontology. DEGs: differentially expressed genes. FDR: false discovery rate, namely, Benjamini-Hochberg adjusted p-value.

Table 2. The top 10 nodes with high degree in PPI network and module ClusterID4.

Gene symbol	Gene name	Average shortest path length	Betweenness centrality	Closeness centrality	Clustering coefficient	Mark	Connectivity degree
• Top 10 nodes in PPI network							
VCAM1	Vascular celladhesionmolecule 1	2.550943	0.219	0.392	8.36E-04	-1	424
CUL4B	Cullin 4B	2.653208	0.147	0.377	2.15E-03	-1	304
UBE2I	Ubiquitin-conjugating enzyme E2I	2.751698	0.153	0.363	6.83E-04	1	271
NEDD4	Neural precursor cell expressed, developmentally down-regulated 4	2.798113	0.141	0.357	4.98E-04	-1	254
H2AFX	H2A histone family, member X	2.771698	0.094	0.361	1.33E-03	1	207
POT1	Protection of telomeres 1 homolog	2.735094	0.114	0.366	1.37E-03	-1	203
HSPA1B	Heat shock 70kDa protein 1B	2.515472	0.090	0.398	5.85E-03	0	170
TUBA1A	Tubulin, alpha 1a	2.933962	0.052	0.341	0	1	128
SFPQ	Splicing factor proline/glutamine-rich	2.675849	0.058	0.374	2.97E-03	1	125
RPL31	Ribosomal protein L31	2.807547	0.035	0.356	6.26E-03	-1	123
• Top 10 nodes in module ClusterID4 of PPI network							
RPL31	Ribosomalprotein L31	2.807547	0.035	0.356	6.26E-03	-1	123
UBC	Ubiquitin C	2.076604	0.282	0.482	4.67E-03	0	108
EPRS	Glutamyl-prolyl-tRNAsynthetase	2.706792	0.031	0.369	9.94E-03	-1	70
PHGDH	Phosphoglyceratedehydrogenase	2.956981	0.017	0.338	0	-1	60
ANXA1	Annexin A1	2.702264	0.025	0.370	1.06E-02	-1	52
ASNS	Asparaginesynthetase (glutamine-hydrolyzing)	2.842642	0.017	0.352	1.77E-02	-1	43
SRRT	Serrate RNA effector molecule	2.992453	0.014	0.334	0	1	43
TTN	Titin	2.880377	0.014	0.347	9.96E-03	-1	38
LGALS3BP	Lectin, galactoside-binding soluble 3 binding protein	2.813962	0.009	0.355	3.68E-02	1	30
PSAT1	Phosphoserine aminotransferase 1	3	0.007	0.333	0	-1	28

In column Mark, -1 represents down-regulated DEG, 1 represents up-regulated DEG, and 0 represents other gene interacted with DEG directly. DEG: differentially expressed gene. PPI: protein-protein interaction.

Table 3. The top 10 TFs significantly enriched by DEGs.

TFs	Bonferroni adjusted p-value	Target DEGs	Degree
SMAD2	3.27E-21	*ANXA1, ANXA8L2, AQP3, BCAT1, BCR, CCPG1, CEP57, CHD9, GLUD2, H2AFX,* etc.	55
SMAD3	3.27E-21	*ANXA1, ANXA8L2, AQP3, BCAT1, BCR, CCPG1, CEP57, CHD9, GLUD2, H2AFX,* etc.	56
SOX2	2.86E-17	*ACSM3, AQP3, ASNS, ASPH, BCR, CALML4, CCDC25, CD14, CDC42EP4, CEBPD,* etc.	90
TP63	7.21E-13	*ADAMTS3, ANPEP, ANXA1, ANXA8, ANXA8L1, ANXA8L2, AQP3, BBX, BICD2, C10ORF2,* etc.	91
SMAD4	1.92E-12	*ANXA1, ANXA8L2, BCAT1, CDC42EP4, CHD9, LIMK2, PHLDA1, PVRL2, SREBF1, THBS1,* etc.	46
ATF3	3.74E-11	*ALDH6A1, ANXA1, ASNS, CSNK1D, CSTA, CTH, DUSP4, DYNLT3, HBP1, LEFTY2,* etc.	42
BACH1	7.57E-11	*BCR, C1R, CCPG1, CEBPD, CSTA, CUL4B, CXCL1, CXCR4, DZIP3, EFEMP1,* etc.	33
FOXA2	3.07E-08	*ALDH6A1, ANPEP, APOA1, BBX, BCR, CCDC86, CD14, CHD9, CXCL2, EXOC7,* etc.	44
RELA	3.50E-07	*ASPH, BCAT1, CCDC25, CDC42EP4, CEBPD, CXCL1, CXCL2, HSPA1A, IL8, LYZ,* etc.	27
HNF4A	1.76E-06	*ACSM3, ACSS3, ALDH6A1, ANPEP, APOA1, AQP3, ASPH, BBX, BCAT1, BCR, CCDC25,* etc.	66

Degree means the degree in TF-DEG regulatory network. TF(s): transcription factor(s). DEG(s): differentially expressed gene(s).

were also the top TFs significantly enriched by DEGs, and thus were defined as hub TFs. Furthermore, a total of 6 modules were extracted from the TF-DEG regulatory network (Table 4). Especially, hub gene *NEDD4* was simultaneously targeted by SMAD2 and SMAD3, and hub genes *CUL4B* and *NEDD4* were both targeted by HNF4A.

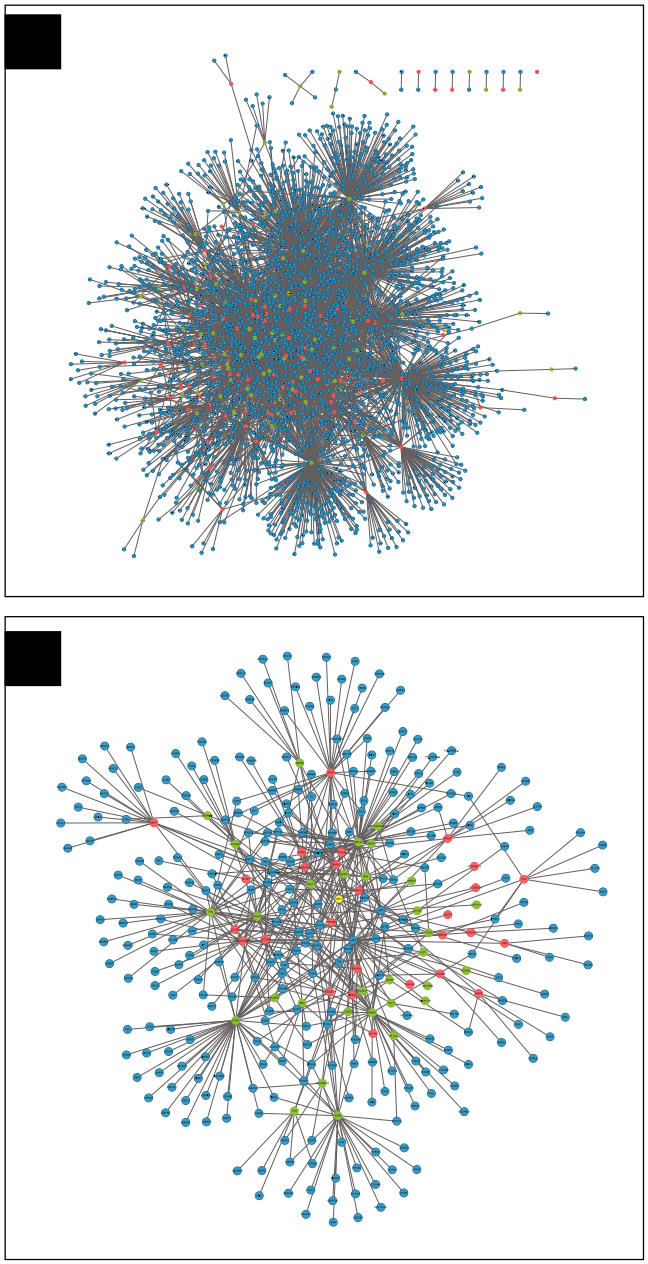

Figure 2. *Protein-protein interaction network and module ClusterID4 of DEGs.* ***A.*** *Protein-protein interaction network.* ***B.*** *Module ClusterID4. Red and green circles represent for the up- and down-regulated genes, respectively. Blue circles indicate the genes that interact with DEGs directly. DEGs:* ***differentially expressed genes.***

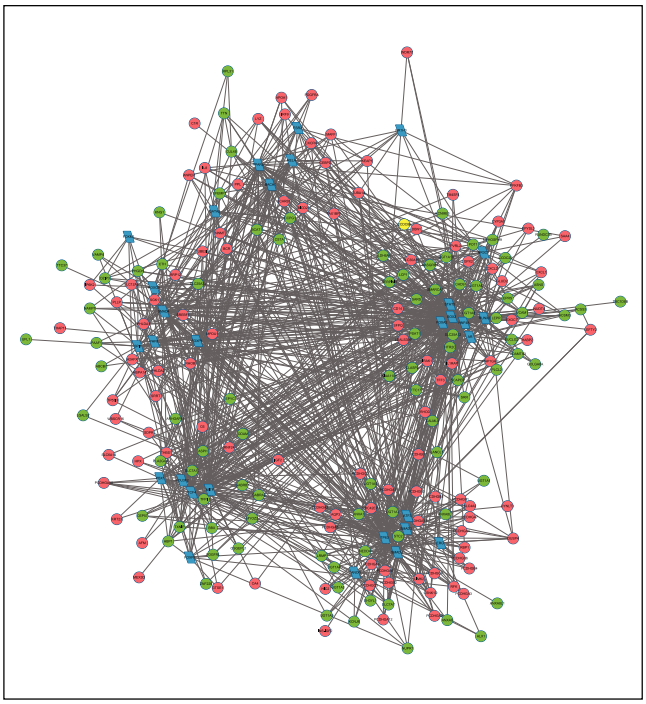

Figure 3. *TF-DEG regulatory network. Red and green circles represent for the up- and down-regulated genes, respectively. Blue circles stand for TFs. TF(s): transcription factor(s). DEG: differentially expressed gene.*

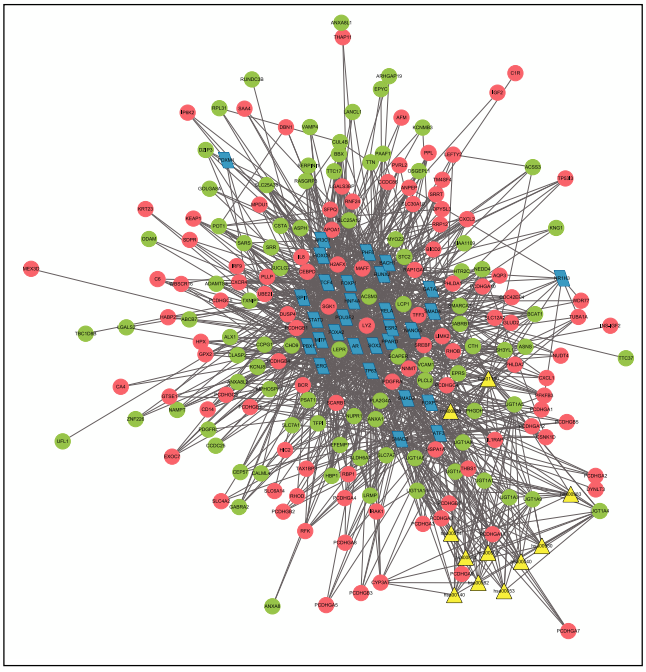

Figure 4. *Complex network of TFs, KEGG pathways, and DEGs. Red and green circles represent for the up- and down-regulated genes, respectively. Blue parallelograms indicate TFs. Yellow triangles stand for the KEGG pathways enriched by DEGs. TFs: transcription factors. KEGG: Kyoto* ***Encyclopedia of Genes and Genomes. DEGs: differentially expressed genes.***

Table 4. DEGs and TFs in the 6 modules of TF-DEG regulatory network.

Cluster	DEGs	TFs	Interactions
1	*DPYSL3, SLC30A10, TM4SF4, CD14, NUDT4, CCDC86, IRAK1, CYP3A7, CXCL2, TFF3, GLUD2, CXCL1, HABP2, PVRL2, GPX2, DBN1, RAP1GAP, LGALS3BP, SFPQ, IL1RAP, LEFTY2, SAA4, EXOC7, SCAPER, GOLGA8A, SERPINI1, TTC17, CCDC25, SUCLG2, EPRS, SRR, SARS, RUNDC3B, RASGRP3, POT1, PLCL2, SLC25A12, ALDH6A1, HTR2C, LEPR, CLASP2, UGT1A10, UGT1A8, PSAT1, ADAMTS3, LCP1, UGT1A6, CALML4, KCNMB3, VCAM1, MPHOSPH9, KIAA1109, TBC1D8B, ASNS, CHD9, ACSS3, SMARCA1, ACSM3*	STAT3, AR, HNF4A, FOXA2, NR3C1, SOX2, RUNX2	189
2	*LANCL1, ANXA8, SH3YL1, ALX1, SLC7A7, UGT1A9, UGT1A1, NEDD4, UGT1A4, LRMP, STC2, NUPR1, KCNJ8, ANXA1, ANXA8L2, ANXA8L1, UGT1A3, UGT1A5, UGT1A7, DUSP4, DYNLT3, PCDHGC5, PCDHGB2, SLC4A2, RFK, RHOD, CSNK1D, AQP3, PCDHGA12, HIC2, RBP1, PCDHGC3, PCDHGA11, LIMK2, CDC42EP4, PCDHGB6, PCDHGB7, PCDHGC4, PCDHGB3, PCDHGB4, PCDHGB5, INS-IGF2, PCDHGA8, PCDHGA5, PCDHGB1, PCDHGA7, PCDHGA4, PCDHGA6, PCDHGA3, PCDHGA2, PCDHGA9, PCDHGA1*	NANOG, SMAD3, ERG, SMAD2, ATF3, TP63	165
3	*HSPA1A, MPDU1, THAP11, PLLP, TP53I3, PHLDA2, SRRT, WBSCR16, SGK1, SLC12A2, RHOB, RRP12, IP6K2, UBE2I, SREBF1, H2AFX, PHLDA1, TTC37, CTH, UFL1, KNG1, LGALS2, DZIP3, PHGDH, ABCB7, PAAF1, VAMP4, SLC25A36, NAMPT*	SPI1, PHF8, FOXP1, FOXM1, HOXC9, SMAD4	67
4	*ARHGAP19, ASPH, BBX, CEP57, EPYC, GABRA2, GABRB1, HBP1, MYOZ2, ODAM, OSGEPL1, PDGFRL, PLA2G4A, SLC7A11, TFPI, TXNIP, ZNF226, AFM, C6, CA4, GTSE1, HPX, IGF2, KRT23, MEX3D, PCDHGA10, RNF24, SDPR, SLC6A14, THBS1*	FOXP2, MITF, PBX1, POU3F2, TCF4	62
5	*CSTA, CUL4B, RPL31, EFEMP1, CCPG1, BCAT1, TTN, CXCR4, IL8, IRF9, MAFF, KEAP1, TAX1BP3, LYZ, PPL, ANPEP, SCARB1, C1R, NNMT, CEBPD, BCR, BICD2, PDGFRA, APOA1*	PPARD, RELA, ESR2, BACH1, GATA2	50
6	*TUBA1A, PFKFB3, WDR77*	NR1H3	3

TF(s): transcription factor(s); DEG(s): differentially expressed gene(s).

TF-pathway-DEG complex network

The complex network of TFs, KEGG pathways, and DEGs was constructed, involving 237 nodes (including 11 pathways, 196 DEGs, and 30 TFs) and 1207 interactions (Figure 4). Especially, the 11 pathways were mainly enriched by down-regulated *UGT1A* (uridine-5'-diphosphate glucuronosyltransferases subgroup family 1A) family and up-regulated *CYP3A7* (cytochrome P450, family 3, subfamily A, member 7), both of which were mainly regulated by TFs SMAD2, SMAD3, SOX2, TP63, and HNF4A.

DISCUSSION

As the main form of malignant liver cancer, HCC is the primary cause of cancer-related death worldwide.[2] In this study, a total of 209 DEGs were identified between the HCC cells treated with and without TGF-β. These DEGs were significantly enriched in cell adhesion-related functions which inhibited tumor invasion and metastasis.[28] This is consistent with the invasion-promoting effects of TGF-β on HCC cells.[29]

Then, the PPI network, TF-DEG regulatory network, and TF-pathway-DEG complex network were constructed. Especially, hub genes *CUL4B* and *NEDD4* in the PPI network were targeted by the significant enriched TFs SMAD2, SMAD3, and HNF4A in the TF-DEG regulatory network. In addition, the 11 pathways in the TF-pathway-DEG complex network were mainly enriched by *UGT1A* family and *CYP3A7,* which were mainly regulated by TFs SMAD2, SMAD3, SOX2, TP63, and HNF4A.

NEDD4 is predicted to be targeted by hub TFs SMAD3 and SMAD2, based on the TF-DEG regulatory network. It is reported that abnormal expression of SMAD3 can reduce hepatocarcinogenesis in a murine HCC model,[30] and the interaction between SMAD2/SMAD3/SMAD4 and PKB/Akt modulates TGF-β signalling during EMT.[31,32] TGF-β1 mediates cell apoptosis in HCC via interacting with SMADs proteins.[33] In addition, *NEDD4-2* can negatively regulate TGF-β signalling via inducing ubiquitin-mediated degradation of TGF-β type I

receptor and SMAD2.[34] In the present study, hub gene *NEDD4* was significantly enriched in cellular response to cytokine stimuli and transmembrane transport, and *NEDD4* was simultaneously targeted by SMAD2 and SMAD3. Therefore, *NEDD4* might play a role in HCC cell response to TGF-β stimulation, and this process is regulated by SMAD2 and SMAD3.

As a regulatory TF targeting hub genes *CUL4A* and *NEDD4*, HNF4A also had a high connectivity degree in the TF-DEG regulatory network. HNF4A, a member of the steroid/thyroid nuclear receptor superfamily, is a key regulator of liver metabolism.[35] Previous study has found that the overexpression of *CUL4A* is related to HCC.[36] In this study, *CUL4A* was significantly enriched in cellular response to stimulus and stress. Therefore, *CUL4B* might contribute to HCC cell response to TGF-β stimulation, and this process is modulated by HNF4A.

In the TF-pathway-DEG complex network, the 11 pathways were enriched by up-regulated DEG *CYP3A7,* which was mainly regulated by TFs SOX2, TP63, and HNF4A. Reportedly, overexpression of SOX2 is associated with a low survival rate of HCC patients, and it promotes cancer cell invasion.[37] TF TP63 is a member of the p53 family,[38] which plays crucial roles in cell cycle arrest, apoptosis, and oncogenesis.[1,39] In addition, CYP (cytochrome P450) family is closely associated with the pathogenesis of HCC.[40] As a member of CYP family, *CYP3A4* plays roles in the metabolism of many drugs and the activation of pro-carcinogens in human liver.[41] In the present study, *CYP3A4* was significantly enriched in cellular response to stimulus and metabolism of drug and xenobiotics. These findings suggested that *CYP3A7* might participate in HCC cell response to TGF-β stimulation, and this process is regulated by SOX2, TP63, and HNF4A.

What's more, all the pathways in the TF-pathway-DEG complex network were enriched by *UGT1A* family *(UGT1As),* which was mainly regulated by TFs SMAD2, SMAD3, SOX2, TP63, and HNF4A. *UGT1A7* (uridine-5'-diphosphate glucuronosyltransferases subgroup family 1A member 7) genetic polymorphisms (UGT1A7*2 and *3 alleles) are significantly associated with HCC risk and onset age.[42] In the present study, *UGT1A7* was enriched in response to stimulus and metabolism of drug, vitamin, and saccharides. Therefore, it's suggested that *UGT1A* might participate in HCC cell response to TGF-β stimulation, and this process is modulated by SMAD2, SMAD3, SOX2, TP63, and HNF4A.

In conclusion, we conducted a comprehensive bioinformatics analysis on the gene expression changes in HCC cells after TGF-β treatment. It' proposed that TGF-β might affect biological processes of HCC cells via SMAD2/SMAD3-*NEDD4*, HNF4A-*CUL4B*/*NEDD4*, SOX2/TP63/HNF4A-*CYP3A7*, and SMAD2/SMAD3/SOX2/TP63/HNF4A-*UGT1As* regulatory pathways. Although further experiments are still needed to validate these hypotheses, the results of this study might provide directions for future researches on HCC invasion and drug design.

ACKNOWLEDGEMENTS

Our study was supported by Scientific Research Fund of Heilongjiang Provincial Education Department (12531357).

CONFLICT OF INTERESTS

All authors declare that they have no conflict of interests to state.

REFERENCES

1. Aravalli RN, Steer CJ, Cressman EN. Molecular mechanisms of hepatocellular carcinoma. *Hepatology* 2008; 48: 2047-63.
2. El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* 2012; 142: 1264-73.
3. Andreou A, Vauthey JN, Cherqui D, Zimmitti G, Ribero D, Truty MJ, Wei SH, et al. Improved long-term survival after major resection for hepatocellular carcinoma: a multicenter analysis based on a new definition of major hepatectomy. *J Gastrointest Surg* 2013; 17: 66-77.
4. Nelson WJ, Nusse R. Convergence of Wnt, β-catenin, and cadherin pathways. *Science* 2004; 303: 1483-7.
5. Giannelli G, Villa E, Lahn M. Transforming growth factor-β as a therapeutic target in hepatocellular carcinoma. *Cancer Res* 2014; 74: 1890-4.
6. Meindl-Beinker NM, Matsuzaki K, Dooley S. TGF-beta signaling in onset and progression of hepatocellular carcinoma. *Dig Dis* 2012; 30: 514-23.
7. Katsuno Y, Lamouille S, Derynck R. TGF-beta signaling and epithelial-mesenchymal transition in cancer progression. *Curr Opin Oncol* 2013; 25: 76-84.
8. Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet J-P, Chiang DY, Villanueva A, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res* 2009; 69: 7385-92.
9. Feng T, Dzieran J, Gu X, Marhenke S, Vogel A, Machida K, Weiss TS, et al. Smad7 regulates compensatory hepatocyte proliferation in damaged mouse liver and positively relates to better clinical outcome in human hepatocellular carcinoma. *Clin Sci (Lond)* 2015; 128: 761-74.
10. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; 5: R80.
11. Smyth GK. Limma: linear models for microarray data, in Bioinformatics and computational biology solutions using R and Bioconductor. Springer; 2005, p. 397-420.
12. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995; 57: 289-300.
13. Tweedie S, Ashburner M, Falls K, Leyland P, Mcquilton P, Marygold S, Millburn G, et al. FlyBase: enhancing Drosophila

gene ontology annotations. *Nucleic Acid Res* 2009; 37: D555-D559.
14. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; 16: 284-7.
15. Yu G, Wang L-G. Disease Ontology Semantic and Enrichment analysis. 2013.
16. Grega M, Bryk D, Napora M, Gusta M, INACT-INDECT advanced image cataloguing tool, in Multimedia Communications, Services and Security. Springer; 2011, p. 28-36.
17. Chatr-Aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G. MINT: the Molecular INTeraction database. *Nucleic Acid Res* 2007; 35: D572-D574.
18. Stark C, Breitkreutz B-J, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, et al. The BioGRID interaction database: 2011 update. *Nucleic Acid Res* 2011; 39: D698-D704.
19. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, et al. UniProt: the universal protein knowledgebase. *Nucleic Acid Res* 2004; 32: D115-D119.
20. Bader GD, Betel D, Hogue CW. BIND: the biomolecular interaction network database. *Nucleic Acid Res* 2003; 31: 248-50.
21. Wassermann AM, Bajorath J. BindingDB and ChEMBL: online compound databases for drug discovery. *Expert Opin Drug Discov* 2011; 6: 683-7.
22. Paz A, Brownstein Z, Ber Y, Bialik S, David E, Sagir D, Ulitsky I, et al. SPIKE: a database of highly curated human signaling pathways. *Nucleic Acid Res* 2011; 39: D793-D799.
23. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; 13: 2498-504.
24. Su G, Kuchinsky A, Morris JH, Meng F. GLay: community structure analysis of biological networks. *Bioinformatics* 2010; 26: 3135-7.
25. Morris JH, Apeltsin L, Newman AM, Baumbach J, Wittkop T, Su G, Bader GD, et al. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC bioinformatics* 2011; 12: 436.
26. Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* 2010; 26: 2438-44.
27. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988; 75: 800-2.
28. Cavallaro U, Christofori G. Cell adhesion in tumor invasion and metastasis: loss of the glue is not enough. *Biochim Biophys Acta* 2001; 1552: 39-45.
29. Fransvea E, Mazzocca A, Antonaci S, Giannelli G. Targeting transforming growth factor (TGF)-betaRI inhibits activation of beta1 integrin and blocks vascular invasion in hepatocellular carcinoma. *Hepatology* 2009; 49: 839-50.
30. Yang Y-A, Zhang G-M, Feigenbaum L, Zhang YE. Smad3 reduces susceptibility to hepatocarcinoma by sensitizing hepatocytes to apoptosis through downregulation of Bcl-2. *Cancer cell* 2006; 9: 445-57.
31. Remy I, Montmarquette A, Michnick SW. PKB/Akt modulates TGF-b signalling through a direct interaction with Smad3. *Nat Cell Biol* 2004; 6: 358-65.
32. Wendt MK, Allington TM, Schiemann WP. Mechanisms of the epithelial-mesenchymal transition by TGF-beta. *Future Oncol* 2009; 5: 1145-68.
33. Ceballos MP, Parody JP, Alvarez MDL, Ingaramo PI, Carnovale CE, Carrillo MC. Interferon-á2b and transforming growth factor-b1 treatments on HCC cell lines: Are Wnt/b-catenin pathway and Smads signaling connected in hepatocellular carcinoma? *Biochem Pharmacol* 2011; 82: 1682-91.
34. Kuratomi G, Komuro A, Goto K, Shinozaki M, Miyazawa K, Miyazono K, Imamura T. NEDD4-2 (neural precursor cell expressed, developmentally down-regulated 4-2) negatively regulates TGF-beta (transforming growth factor-beta) signalling by inducing ubiquitin-mediated degradation of Smad2 and TGF-beta type I receptor. *Biochem J* 2005; 386: 461-70.
35. Watt AJ, Garrison WD, Duncan SA. HNF4: a central regulator of hepatocyte differentiation and function. *Hepatology* 2003; 37: 1249-53.
36. Yasui K, Arii S, Zhao C, Imoto I, Ueda M, Nagai H, Emi M, et al. TFDP1, CUL4A, and CDC16 identified as targets for amplification at 13q34 in hepatocellular carcinomas. *Hepatology* 2002; 35: 1476-84.
37. Sun C, Sun L, Li Y, Kang X, Zhang S, Liu Y. Sox2 expression predicts poor survival of hepatocellular carcinoma patients and it promotes liver cancer cell invasion by activating Slug. *Med Oncol* 2013; 30: 1-10.
38. Yang A, Kaghad M, Wang Y, Gillett E, Fleming MD, Dötsch V, Andrews NC, et al. p63, a p53 Homolog at 3q27–29, Encodes Multiple Products with Transactivating, Death-Inducing, and Dominant-Negative Activities. *Mol Cell* 1998; 2: 305-16.
39. Moll UM, Slade N. p63 and p73: Roles in Development and Tumor Formation 11 National Cancer Institute. *Mol Cancer Res* 2004; 2: 371-86.
40. Tsunedomi R, Iizuka N, Hamamoto Y, Uchimura S, Miyamoto T, Tamesa T, Okada T, et al. Patterns of expression of cytochrome P450 genes in progression of hepatitis C virus-associated hepatocellular carcinoma. *Int J Oncol* 2005; 27: 661-7.
41. Kivist KT, Bookjans G, Fromm MF, Griese EU, Mnzel P, Kroemer HK. Expression of CYP3A4, CYP3A5 and CYP3A7 in human duodenal tissue. *Br J Clin Pharmacol* 1996; 42: 387-9.
42. Tseng CS, Tang KS, Lo HW, Ker CG, Teng HC, Huang CS. UDP-glucuronosyltransferase 1A7 genetic polymorphisms are associated with hepatocellular carcinoma risk and onset age. *Am J Gastroenterol* 2005; 100: 1758-63.

Correspondence and reprint request:
Yubao Zhang, MD
Department of Hepatobiliary and Pancreatic Surgery,
The Affiliated Tumor Hospital of Harbin Medical University,
No.150, Haping Road, Nangang District, Harbin 150081,
Heilongjiang Province, China
Tel.: +86-0451-86298082
Fax: +86-0451-86298082
E-mail: gdywk@hotmail.com