Identification of key genes, pathways and potential therapeutic agents for liver fibrosis using an integrated bioinformatics analysis

Zhu Zhan [1, 2], **Yuhe Chen** [1, 2], **Yuanqin Duan** [1, 2], **Lin Li** [3], **Kenley Mew** [4], **Peng Hu** [1, 2], **Hong Ren** [1, 2], **Mingli Peng** [Corresp. 1, 2]

[1] Key Laboratory of Molecular Biology for Infectious Diseases (Ministry of Education), Chongqing Medical University, Chongqing, China

[2] Department of Infectious Diseases, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China

[3] Department of hepatic disease, Chongqing Traditional Chinese Medicine Hospital, Chongqing, China

[4] Department of foreign language, Chongqing Medical University, Chongqing, China

Corresponding Author: Mingli Peng
Email address: Peng_mingli@hospital.cqmu.edu.cn

Background. Liver fibrosis is often a consequence of chronic liver injury, and has the potential to progress to cirrhosis and liver cancer. Despite being an important human disease, there are currently no approved anti-fibrotic drugs. In this study, we aim to identify the key genes and pathways governing the pathophysiological processes of liver fibrosis, and to screen therapeutic anti-fibrotic agents. **Methods.** Expression profiles were downloaded from the Gene Expression Omnibus (GEO), and differentially expressed genes (DEGs) were identified by R packages (Affy and limma). Gene functional enrichments of each dataset were performed on the DAVID database. Protein–protein interaction (PPI) network was constructed by STRING database and visualized in Cytoscape software. The hub genes were explored by CytoHubba plugin app and validated in another GEO dataset and in a liver fibrosis cell model by quantitative real-time PCR assay. The Connectivity Map L1000 platform was used to identify potential anti-fibrotic agents. **Results.** We integrated 3 fibrosis datasets of different disease etiologies, incorporating a total of 70 severe (F3-F4) and 116 mild (F0-F1) fibrotic tissue samples. Gene functional enrichment analyses revealed that cell cycle was a pathway uniquely enriched in a dataset from those patients infected by hepatitis B virus (HBV), while the immune-inflammatory response was enriched in both the HBV and hepatitis C virus (HCV) datasets, but not in the nonalcoholic fatty liver disease (NAFLD) dataset. There was overlap between these three datasets; 185 total shared DEGs that were enriched for pathways associated with extracellular matrix constitution, platelet-derived growth-factor binding, protein digestion and absorption, focal adhesion, and PI3K-Akt signaling. In the PPI network, 25 hub genes were extracted and deemed to be essential genes for fibrogenesis, and the expression trends were consistent with GSE14323 (an additional dataset) and liver fibrosis cell model, confirming the

relevance of our findings. Among the 10 best matching anti-fibrotic agents, Zosuquidar and its corresponding gene target ABCB1 might be a novel anti-fibrotic agent or therapeutic target, but further work will be needed to verify its utility. **Conclusions.** Through this bioinformatics analysis, we identified that cell cycle is a pathway uniquely enriched in HBV related dataset and the immune-inflammatory response is severe in virus-induced liver fibrosis. Zosuquidar and ABCB1 might be a novel anti-fibrotic agent or target.

Identification of key genes, pathways and potential therapeutic agents for liver fibrosis using an integrated bioinformatics analysis

Zhu Zhan[1, 2], Yuhe Chen[1, 2], Yuanqin Duan[1, 2], Lin Li[3], Kenley Mew[4], Peng Hu[1, 2], Hong Ren[1, 2], Mingli Peng[1, 2]

[1]Key Laboratory of Molecular Biology for Infectious Diseases (Ministry of Education), Chongqing Medical University, Chongqing, China
[2]Department of Infectious Diseases, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China
[3]Department of hepatic disease, Chongqing Traditional Chinese Medicine Hospital, Chongqing, China
[4]Department of foreign language, Chongqing Medical University, Chongqing, China

Corresponding Author:
Mingli Peng[1]
NO. 1, Yixueyuan Road, Yuzhong District, Chongqing, China.
Email address: Peng_Mingli@hospital.cqmu.edu.cn

Abstract

Background. Liver fibrosis is often a consequence of chronic liver injury, and has the potential to progress to cirrhosis and liver cancer. Despite being an important human disease, there are currently no approved anti-fibrotic drugs. In this study, we aim to identify the key genes and pathways governing the pathophysiological processes of liver fibrosis, and to screen therapeutic anti-fibrotic agents.

Methods. Expression profiles were downloaded from the Gene Expression Omnibus (GEO), and differentially expressed genes (DEGs) were identified by R packages (Affy and limma). Gene functional enrichments of each dataset were performed on the DAVID database. Protein–protein interaction (PPI) network was constructed by STRING database and visualized in Cytoscape software. The hub genes were explored by CytoHubba plugin app and validated in another GEO dataset and in a liver fibrosis cell model by quantitative real-time PCR assay. The Connectivity Map L1000 platform was used to identify potential anti-fibrotic agents.

Results. We integrated 3 fibrosis datasets of different disease etiologies, incorporating a total of 70 severe (F3-F4) and 116 mild (F0-F1) fibrotic tissue samples. Gene functional enrichment analyses revealed that cell cycle was a pathway uniquely enriched in a dataset from those patients infected by hepatitis B virus (HBV), while the immune-inflammatory response was enriched in both the HBV and hepatitis C virus (HCV) datasets, but not in the nonalcoholic fatty liver disease (NAFLD) dataset. There was overlap between these three datasets; 185 total shared DEGs that were enriched for pathways associated with extracellular matrix constitution, platelet-derived growth-factor binding, protein digestion and absorption, focal adhesion, and PI3K-Akt signaling. In the PPI network, 25 hub genes were extracted and deemed to be essential genes for fibrogenesis, and the expression trends were consistent with GSE14323 (an additional dataset) and liver fibrosis cell model, confirming the relevance of our findings. Among the 10 best matching anti-fibrotic agents, Zosuquidar and its corresponding gene target ABCB1 might be a novel anti-fibrotic agent or therapeutic target, but further work will be needed to verify its utility.

Conclusions. Through this bioinformatics analysis, we identified that cell cycle is a pathway uniquely enriched in HBV related dataset and immune-inflammatory response is severe in virus-induced liver fibrosis. Zosuquidar and ABCB1 might be a novel anti-fibrotic agent or target.

Key words. liver cirrhosis, bioinformatics, microarray analysis, therapeutics.

Introduction

Hepatic fibrosis is characterized by the pathological accumulation of extracellular matrix (ECM) following chronic liver injury arising from various sources including toxic damage, viral infections, autoimmune conditions, and metabolic or genetic diseases. Patients with advanced liver fibrosis generally have a poor prognosis as they often develop decompensated cirrhosis and hepatocellular carcinoma (Tsochatzis et al. 2014).

Although there have been major experimental advances in the understanding of the molecular mechanisms governing liver fibrosis in recent decades(Lee et al. 2015), due to the complex regulatory network underlying this disease, efforts to develop a successful therapeutic approach are often limited (Weiskirchen et al. 2018). Currently, there are no approved anti-fibrotic drugs (Bottcher & Pinzani 2017). A better understanding of the molecular mechanisms controlling the fibrotic response is thus needed to improve patient outcomes.

High-throughput sequencing technology offers an ideal means of profiling large gene expression datasets in order to gain a comprehensive understanding of the mechanisms underlying fibrosis. There have been many studies in recent years profiling liver fibrosis gene expression microarray data, identifying hundreds of differentially expressed genes (DEGs) associated with this disease (Moylan et al. 2014; Wang et al. 2017). However, these approaches have been limited in their ability to identify the key genes regulating the disease as a whole. Moreover, the results of these studies are often inconsistent due to different etiologies of liver fibrosis. In order to eliminate this variation, we compared fibrosis-related DEGs from multiple different studies using bioinformatics approaches in order to identify key genes and pathways of interest, and to screen for therapeutic agents and novel targets with the potential to treat liver fibrosis.

Materials & Methods

Microarray data.

Four gene expression datasets were downloaded from the Gene Expression Omnibus (GEO) database; three were analyzed to identify DEGs, while one was used for validation. Table 1 summarizes the pertinent information for the selected GEO datasets used in this study. GSE6764 (Wurmbach et al. 2007), GSE49541(Moylan et al. 2014), and GSE84044 (Wang et al. 2017) represent datasets from patients with liver fibrosis arising from hepatitis C virus (HCV), nonalcoholic fatty liver disease (NAFLD), and hepatitis B virus (HBV), respectively. All three of these gene expression profiles were based on the GPL570 platform. GSE49541 and GSE84044 were derived from two liver fibrosis studies in which tissues with severe fibrosis (F3-F4) and mild fibrosis (F0-F1) were selected. GSE6764 and GSE14323 (Mas et al. 2009) (used for validation) were derived from two liver cancer studies, in which cirrhotic (F4) and normal tissues (F0) were selected.

Identification of differentially expressed genes.

Background expression value correction and data normalization were conducted for the raw data in each dataset using an R package (Affy, version 1.52.0). Probes in each data file were then annotated based on the appropriate platform annotation files. Probes without matching gene symbols were removed. In instances where different probes mapped to the same gene, the mean value of all probes mapping to that gene was taken as the final expression value for that gene. Then, the Linear Models for Microarray Analysis R package (limma; version 3.30.11) was applied for differential expression analysis. Those genes with an adjusted P-value <0.05 and absolute value of fold-change (FC) >1.5 were deemed to be the DEGs. DEGs overlapping between datasets were obtained using an online Venn analysis tool (http://bioinformatics.psb.ugent.be/webtools/Venn/).

Gene Ontology and pathway enrichment analyses.

Gene Ontology (GO) is a commonly used bioinformatics tool that provides comprehensive information on gene function of individual genomic products based on defined features. This analysis consists of three facets: molecular functions (MF), biological processes (BP) and cellular components (CC). The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database resource for understanding high-level biological functions and utilities. These analyses and annotations are based on the DAVID database (https://david.ncifcrf.gov/), which provides a comprehensive set of functional annotation tools for investigators to explore and understand the biological meaning underlying particular gene lists. In this study, both GO and KEGG analyses of DEGs were performed with a criterion false discovery rate (FDR) < 0.05.

Protein–protein interaction (PPI) network construction and hub gene analysis.

In order to analyze the connections among the proteins encoded by identified DEGs, DEGs were uploaded to Search Tool for the Retrieval of Interacting Genes (STRING, https://string-db.org/),

a database of known and predicted protein-protein interactions, and the results with a minimum interaction score of 0.4 were visualized in Cytoscape. Furthermore, CytoHubba, a Cytoscape plugin app, providing a user-friendly interface to explore important nodes in biological networks, was utilized with the maximal clique centrality (MCC) method to explore the PPI network for hub genes.

DEGs Validation.

Another dataset GSE14323 was used to confirm the validity and disease relevance of identified DEGs. A heat map of the expression of 25 hub genes was developed using the HemI1.0.3.3 software. Statistical difference analysis between the liver cirrhosis group (LC) and normal control group (NC) was performed via student's t-test using SPSS V20.0. $P < 0.05$ was considered statistically significant. As activation of hepatic stellate cells (HSCs) is considered as a central driver of liver fibrosis, we used a human HSC cell line — LX2 treated with TGF-β1 to represent this activation stage. An expression of 25 hub genes was performed by quantitative real-time PCR assay compared with normal control.

Cell culture and treatment.

The LX2 cell line was purchased from Procell Life Science & Technology (Wuhan, China), cultured with Dulbecco Modified Eagle Medium (DMEM)-high glucose supplemented with 10% fetal bovine serum (FBS) and antibiotics (100 U/mL penicillin-G and 100 μg/mL streptomycin), and incubated at 37°C in 5% CO2 and 95% humidified air. The LX2 cells were seeded in a 10-cm culture dish at a density of $1*10^6$ for 6 hours. After attachment, the LX2 cells were treated with TGF-β1 (R&D systems, Catalog #240-B/CF） at 10 ng/ml concentration or left untreated as normal control for 24 hours. Then RNA and proteins were isolated for further use.

Western blot assay.

LX2 cells total protein were extracted with ice-cold RIPA lysis buffer. Protein concentration was determined using the BCA Protein Assay Kit (Thermo Fisher Scientific, USA). Quantified proteins were separated on SDS-PAGE and transferred onto PVDF membranes (Millipore Corporation, USA). After blocking, membranes were incubated with anti-αSMA (1:20000, ab124964, Abcam, UK) at 4 °C overnight. Then, membranes were washed with TBST and incubated with secondary antibodies for 2 hours at room temperature. The anti-GAPDH (1:1000, CST) was set as internal control. Protein bands were visualized by using ECL equipment (Pierce Chemical, USA).

Quantitative real-time PCR assay.

RNA was extracted from cell line LX2 by Trizol reagent (Takara, Japan) by following the manufacturer's instructions. The cDNAs were synthesized with a commercial kit (Takara, Japan). Gene expressions were measured by real-time PCR with CFX Connect™ Real-Time PCR System (Bio-Rad, USA). GAPDH was used as an internal control and the relative

expression levels of mRNA were calculated using the $2^{-\Delta\Delta Ct}$ method. The primer pairs used in the experiments are listed in Supplementary data 1.

Prediction of therapeutic agents and target genes.

To discover potential anti-fibrotic agents, the identified 185 DEGs were queried using the Connectivity Map online tool (L1000 platform; https://clue.io/l1000-query). This tool compares queried signatures with a gene expression profile database of several cell lines after treatment with approximately 1000 compounds, most of which are FDA approved. Drugs whose signatures were in opposition to the disease signature were assumed to have therapeutic potential.

Results

Identification of 185 conserved DEGs.

As shown in Fig. 1, each dataset was initially analyzed separately to identify DEGs unique to fibrosis of a given origin. 1563 DEGs were identified in GSE6764 (HCV), 243 DEGs in GSE49541 (NAFLD), and 1396 DEGs in GSE84044 (HBV). 185 DEGs overlapped across all three datasets, suggesting that these fibrosis-related DEGs may be conserved regardless of disease etiology. Among these 185 DEGs, 174 were up-regulated while only 11 were down-regulated. Interestingly, although the number of DEGs in NAFLD related dataset is relatively small, 94.7% of the 243 DEGs intersect with other datasets.

Functional enrichment analysis of DEGs.

In order to compare the differences in gene function among these 3 datasets, GO and KEGG analyses were performed on each dataset and top 10 significant GO_BP and KEGG pathways are shown in Table 2 and Table 3, respectively. Unexpectedly, the cell cycle pathway was uniquely enriched in the HBV-related dataset, ranking third among all KEGG pathways for this dataset. When compared with a non-viral fibrosis dataset (GSE49541), those datasets in which fibrosis was of viral origin (GSE6764 and GSE84044) contained DEGs enriched for immune-inflammatory responses, consistent with the distinct role of immunological responses in the initiation and control of local disease in affected individuals.

Next, GO and KEGG analyses were performed on the 185 common DEGs. The GO analysis revealed that most of the proteins encoded by these DEGs were extracellular matrix proteins located in the extracellular space (Fig. 2). The molecular functions (MF) enriched in this dataset were primarily associated with platelet-derived growth-factor binding and extracellular matrix structural constitution, while the enriched biological processes (BP) were primarily those associated with extracellular matrix organization and cell adhesion (Fig. 2). The KEGG analysis revealed that the primary enriched signaling pathways were those associated with ECM-receptor interaction, protein digestion and absorption, focal adhesion, and the PI3K-Akt signaling pathway (Fig. 3). Together, these shared DEGs highlight the central roles of cell-cell adhesion and ECM dysregulation in the development of fibrosis, regardless of the etiological origin of the disease.

PPI network construction and hub gene identification.

To better understand which of these shared DEGs were most likely to be the key genes most essential for the development of fibrosis, a PPI network for these 185 common DEGs was built with 105 nodes and 275 edges. 80 of the 185 DEGs were not included in the PPI network (Fig. 4), as interaction score of these 80 genes were less than 0.4. Among the 105 genes in the PPI network, the top 25 genes according to the MCC method were selected using the CytoHubba plugin and are sequentially ordered as follows: COL1A2, COL1A1, COL6A3, COL3A1, COL5A2, COL5A1, COL4A1, COL4A2, COL4A3, COL4A4, DCN, COL14A1, LUM, COL15A1, THBS2, FBN1, ITGB8, CDH11, ADAMTS2, CTGF, VCAN, PCOLCE2, SPP1,

VWF, CTSK (Fig. 5). These 25 genes were deemed to be the hub genes and were those genes most likely to be essential for fibrogenesis. Most genes encode ECM components, including COL1A2, DCN, and FBN1. Other hub genes play known roles in ECM structural regulation (THBS2, ITGB8, VWF), while some are associated with ECM degradation (ADAMTS2, PCOLCE2, CTSK). This finding is consistent with known fibrogenic mechanisms, and suggests key potential drug targets that are most likely to have effective anti-fibrotic activity when disrupted.

Hub gene validation.

In order to extend and validate our findings in a distinct model of human liver fibrosis, these top 25 hub genes were validated in the GSE14323 dataset, in which liver cirrhotic (LC) and normal control tissues (NC) were selected for analysis. Fig. 6 displays a heatmap of GSE14323 expression profile data. This expression profile was consistent with the overlapping DEGs identified in the initial three datasets, with 23 of these 25 hub DEGs being up-regulated in cirrhotic patients. Statistical analysis of these genes in the validation dataset is shown in Fig. 7. Differences for all the hub genes between the LC and NC groups were statistically significant with the exception of ITGB8.

A cell model of liver fibrosis was also constructed to validate these 25 hub genes. When treated with TGF-β1，the LX2 cells extended more tentacles and expressed more a-SMA protein (One of the markers of hepatic stellate cell activation) (Fig. 8), indicating the cell model was successfully established. Fig. 9 displays statistical analysis of 25 hub genes relative expression to GAPDH, 13 of these 25 hub genes was up-regulated significantly, which is consistent with the trend of GEO datasets in this study. However, 4 genes (LUM, THBS2, ITGB8 and SPP1) was down-regulated in TGF-β1 treated cells. The expression trend of some genes was inconsistent with expectations, likely because TGF-β-induced hepatic stellate cell activation does not represent the entire activated form or that part of the gene is expressed by other cells such as hepatic parenchymal cells and Kupffer cells.

Prediction of potential therapeutic agents and targets.

Given the role of 185 DEGs in fibrogenesis, we next wanted to probe for potential therapeutic compounds that might best be suited to target these genes in order to achieve a beneficial therapeutic outcome. To that end the connectivity map L1000 platform, which compiles gene expression profiles associated with a wide range of therapeutic compounds, was used to search for drugs with the potential for therapeutic repurposing as a means of treating liver fibrosis. The top 10 compounds are shown in Table 4, and when sequentially ordered by median_score are: Prometon, MK-212, Evodiamine, Zosuquidar, CAY-10415, Caffeic-acid, Budesonide, Rilmenidine, Afatinib, Desloratadine. Target genes corresponding to each compound (with the exception of prometon) were also listed. Among these target genes, three (HTR2B, ABCB1 and ALOX5) were significantly up-regulated in HBV and HCV datasets, and the mRNA expression levels in each dataset were listed in Table 5. Together, these compounds and target genes provide

a promising list for researchers or companies interested in conducting pre-clinical research into the mechanisms of and treatments for fibrosis both in vitro and in vivo.

Discussion

Although there have been some bioinformatics analyses investigating liver fibrosis, with some being fairly advanced, in general these studies have limited themselves to a specific cause of liver fibrosis(Chan et al. 2016; Chen et al. 2017; Lou et al. 2017; Qi et al. 2017; Wang et al. 2017). Globally, HBV, HCV, and NAFLD are the most three common causes of liver fibrosis(Altamirano-Barrera et al. 2017). In the present study, we integrated datasets that were focused on these three most common causes of liver fibrosis, and in so doing we were able to identify different and common signaling pathways for the fibrogenesis.

Transition of hepatic stellate cells (HSCs) from a quiescent to an activated state is a sign of the onset of liver fibrosis, and this process is controlled by E-type cyclins (CcnE1, CcnE2) and their associated cyclin-dependent kinase 2 (Cdk2) (Nevzorova et al. 2012; Ohtsubo et al. 1995). According to our KEGG pathway enrichment results, the cell cycle pathway is uniquely enriched in HBV-related fibrosis dataset, with CcnE2 being significantly up-regulated only in this HBV dataset, but not in the HCV or NAFLD datasets. Therapeutic targeting of Cyclin E1 via RNAi has been shown to have robust anti-fibrotic activity in mice (Bangen et al. 2017), and if this technology can be applied clinically in future, we predict that it will be most effective in those patients with chronic hepatitis B.

Both HBV and HCV are non-cytopathic viruses, and liver damage in infected individuals is mainly caused by an inflammatory immune response aimed at eliminating the virus(Guidotti & Chisari 2006), with such persistent inflammation leading to liver fibrosis(Protzer et al. 2012). In contrast to such fibrosis of viral origin, the production of reactive oxygen species (ROS) and resulting oxidative stress is thought to be a critical factor in NAFLD-associated fibrosis. Although NAFLD is always accompanied by an inflammatory reaction with variations in levels of pro-inflammatory cytokines(Cai et al. 2005), the degree of inflammation is typically less severe than that caused by HBV/HCV according to our results.

In this study, we were able to screen for and identify 10 compounds that may have therapeutic activity against liver fibrosis. Among these compounds, evodiamine(Yang et al. 2018), caffic-acid(Alferink et al. 2017; Yang et al. 2017) and budesonide(Silveira & Lindor 2014) have all been shown to be effective in animal models or clinical trials. MK-212(Ebrahimkhani et al. 2011), CAY-10415(Boettcher et al. 2012), afatinib(Liang et al. 2018), and desloratadine(Kennedy et al. 2018) have not been tested in vivo, but other agents targeting similar molecules have the potential to ameliorate liver fibrosis. Prometon, Zosuquidar, and Rilmeidine have not yet been reported to have relationship with fibrosis.

Among the target genes corresponding to these compounds, we found that HTR2B, ABCB1, and ALOX5 were significantly up-regulated in HBV and HCV related liver fibrosis datasets (Data not show). Stimulation of the 5-hydroxytryptamine 2B receptor (HTR2B) on HSCs by serotonin is required to negatively regulate hepatocyte regeneration, and antagonism of HTR2B has been shown to attenuate fibrogenesis and improve liver function in disease models in which fibrosis was pre-established and progressive(Ebrahimkhani et al. 2011). Interestingly, MK-212, an HTR2B agonist, showed a negative liver fibrosis gene expression profile suggesting

potential as an anti-fibrotic agent, although formal experimental testing is needed. Arachidonate 5-lipoxygenase (ALOX5) plays a role in the synthesis of leukotrienes from arachidonic acid, and inhibition of the ALOX5 pathway markedly reduces the number of Kupffer cells in culture and attenuates inflammation and fibrosis in experimental liver disease(Titos et al. 2003). Recently, a clinical study revealed that frequent coffee consumption was inversely correlated with liver stiffness(Alferink et al. 2017), with suggestions that the underlying mechanism may be one related to the inhibition of TGF-β1/Smad3 signaling and the induction of autophagy in HSCs in response to caffeic acid(Yang et al. 2017). As an inhibitor of ALOX5, caffeic acid may thus be able to attenuate liver fibrosis via this ALOX5(Sud'ina et al. 1993) pathway.

ATP Binding Cassette Subfamily B Member 1 (ABCB1), is known for encoding P glycoprotein, which is responsible for decreased drug accumulation in multidrug-resistant cells and often mediates the development of resistance to anticancer drugs, such as Zosuquidar mentioned above. However, there are currently no studies reporting that ABCB1, P glycoprotein or Zosuquidar is associated with liver fibrosis. Some studies have reported P glycoprotein was increased in rat activated HSC(Hannivoort et al. 2008), and its activity was increased by TGF-β (Baello et al. 2014) and endoplasmic reticulum stress(Ledoux et al. 2003), which are considered to be effective activators of HSC. Combining the findings of our research, we infer that ABCB1 might be a novel therapeutic target to liver fibrosis, although this hypothesis need to be verified in further study.

Conclusions

Our study integrated three liver fibrosis datasets, each with fibrosis of a different etiology (HBV, HCV and NAFLD). Through the functional analysis of identified DEGs, we revealed that cell cycle is a pathway uniquely enriched in HBV related dataset and immune-inflammatory response is severe in virus-induced liver fibrosis than non-viral. We further identified 25 key hub genes, the majority of which were linked to ECM regulation, highlighting the central processes common to all causes of fibrogenesis, offering valuable insights into the conserved nature of fibrotic signaling. Based on the 185 DEGs, we were additionally predicted 10 compounds, especially Zosuquidar and corresponding gene target ABCB1, may have anti-fibrotic activity. While further experiments will be needed to validate these findings, this successful compound screening effort suggests that it may be possible to repurpose extant drugs to more readily treat liver fibrosis.

References

Alferink LJM, Fittipaldi J, Kiefte-de Jong JC, Taimr P, Hansen BE, Metselaar HJ, Schoufour JD, Ikram MA, Janssen HLA, Franco OH, and Darwish Murad S. 2017. Coffee and herbal tea consumption is associated with lower liver stiffness in the general population: The Rotterdam study. *J Hepatol* 67:339-348.

Altamirano-Barrera A, Barranco-Fragoso B, and Mã©Ndez-Sã nN. 2017. Management strategies for liver fibrosis. *Annals of Hepatology* 16:48-56.

Baello S, Iqbal M, Bloise E, Javam M, Gibb W, and Matthews SG. 2014. TGF-beta1 regulation of multidrug resistance P-glycoprotein in the developing male blood-brain barrier. *Endocrinology* 155:475-484.

Bangen JM, Hammerich L, Sonntag R, Baues M, Haas U, Lambertz D, Longerich T, Lammers T, Tacke F, Trautwein C, and Liedtke C. 2017. Targeting CCl4 -induced liver fibrosis by RNA interference-mediated inhibition of cyclin E1 in mice. *Hepatology* 66:1242-1257.

Boettcher E, Csako G, Pucino F, Wesley R, and Loomba R. 2012. Meta-analysis: pioglitazone improves liver histology and fibrosis in patients with non-alcoholic steatohepatitis. *Aliment Pharmacol Ther* 35:66-75.

Bottcher K, and Pinzani M. 2017. Pathophysiology of liver fibrosis and the methodological barriers to the development of anti-fibrogenic agents. *Adv Drug Deliv Rev* 121:3-8.

Cai D, Yuan M, Frantz DF, Melendez PA, Hansen L, Lee J, and Shoelson SE. 2005. Local and systemic insulin resistance resulting from hepatic activation of IKK-beta and NF-kappaB. *Nat Med* 11:183-190.

Chan KM, Wu TH, Wu TJ, Chou HS, Yu MC, and Lee WC. 2016. Bioinformatics microarray analysis and identification of gene expression profiles associated with cirrhotic liver. *Kaohsiung J Med Sci* 32:165-176.

Chen W, Zhao W, Yang A, Xu A, Wang H, Cong M, Liu T, Wang P, and You H. 2017. Integrated analysis of microRNA and gene expression profiles reveals a functional regulatory module associated with liver fibrosis. *Gene* 636:87-95.

Ebrahimkhani MR, Oakley F, Murphy LB, Mann J, Moles A, Perugorria MJ, Ellis E, Lakey AF, Burt AD, Douglass A, Wright MC, White SA, Jaffre F, Maroteaux L, and Mann DA. 2011. Stimulating healthy tissue regeneration by targeting the 5-HT(2)B receptor in chronic liver disease. *Nat Med* 17:1668-1673.

Guidotti LG, and Chisari FV. 2006. Immunobiology and pathogenesis of viral hepatitis. *Annu Rev Pathol* 1:23-61.

Hannivoort RA, Dunning S, Vander Borght S, Schroyen B, Woudenberg J, Oakley F, Buist-Homan M, van den Heuvel FA, Geuken M, Geerts A, Roskams T, Faber KN, and Moshage H. 2008. Multidrug resistance-associated proteins are crucial for the viability of activated rat hepatic stellate cells. *Hepatology* 48:624-634.

Kennedy L, Hargrove L, Demieville J, Karstens W, Jones H, DeMorrow S, Meng F, Invernizzi P, Bernuzzi F, Alpini G, Smith S, Akers A, Meadows V, and Francis H. 2018. Blocking H1/H2 histamine receptors inhibits damage/fibrosis in Mdr2(-/-) mice and human cholangiocarcinoma tumorigenesis. *Hepatology*.

Ledoux S, Yang R, Friedlander G, and Laouari D. 2003. Glucose depletion enhances P-glycoprotein expression in hepatoma cells: role of endoplasmic reticulum stress response. *Cancer Res* 63:7284-7290.

Lee YA, Wallace MC, and Friedman SL. 2015. Pathobiology of liver fibrosis: a translational success story. *Gut* 64:830-841.

Liang D, Chen H, Zhao L, Zhang W, Hu J, Liu Z, Zhong P, Wang W, Wang J, and Liang G. 2018. Inhibition of EGFR attenuates fibrosis and stellate cell activation in diet-induced model of nonalcoholic fatty liver disease. *Biochim Biophys Acta* 1864:133-142.

Lou Y, Tian GY, Song Y, Liu YL, Chen YD, Shi JP, and Yang J. 2017. Characterization of

transcriptional modules related to fibrosing-NAFLD progression. *Sci Rep* 7:4748.
Mas VR, Maluf DG, Archer KJ, Yanek K, Kong X, Kulik L, Freise CE, Olthoff KM, Ghobrial RM, McIver P, and Fisher R. 2009. Genes involved in viral carcinogenesis and tumor initiation in hepatitis C virus-induced hepatocellular carcinoma. *Mol Med* 15:85-94.
Moylan CA, Pang H, Dellinger A, Suzuki A, Garrett ME, Guy CD, Murphy SK, Ashley-Koch AE, Choi SS, Michelotti GA, Hampton DD, Chen Y, Tillmann HL, Hauser MA, Abdelmalek MF, and Diehl AM. 2014. Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology* 59:471-482.
Nevzorova YA, Bangen JM, Hu W, Haas U, Weiskirchen R, Gassler N, Huss S, Tacke F, Sicinski P, Trautwein C, and Liedtke C. 2012. Cyclin E1 controls proliferation of hepatic stellate cells and is essential for liver fibrogenesis in mice. *Hepatology* 56:1140-1149.
Ohtsubo M, Theodoras AM, Schumacher J, Roberts JM, and Pagano M. 1995. Human cyclin E, a nuclear protein essential for the G1-to-S phase transition. *Mol Cell Biol* 15:2612-2624.
Protzer U, Maini MK, and Knolle PA. 2012. Living in the liver: hepatic infections. *Nat Rev Immunol* 12:201-213.
Qi S, Wang C, Li C, Wang P, and Liu M. 2017. Candidate genes investigation for severe nonalcoholic fatty liver disease based on bioinformatics analysis. *Medicine (Baltimore)* 96:e7743.
Silveira MG, and Lindor KD. 2014. Obeticholic acid and budesonide for the treatment of primary biliary cirrhosis. *Expert Opinion on Pharmacotherapy* 15:365-372.
Sud'ina GF, Mirzoeva OK, Pushkareva MA, Korshunova GA, Sumbatyan NV, and Varfolomeev SD. 1993. Caffeic acid phenethyl ester as a lipoxygenase inhibitor with antioxidant properties. *FEBS Lett* 329:21-24.
Titos E, Claria J, Planaguma A, Lopez-Parra M, Villamor N, Parrizas M, Carrio A, Miquel R, Jimenez W, Arroyo V, Rivera F, and Rodes J. 2003. Inhibition of 5-lipoxygenase induces cell growth arrest and apoptosis in rat Kupffer cells: implications for liver fibrosis. *FASEB J* 17:1745-1747.
Tsochatzis EA, Bosch J, and Burroughs AK. 2014. Liver cirrhosis. *The Lancet* 383:1749-1761.
Wang M, Gong Q, Zhang J, Chen L, Zhang Z, Lu L, Yu D, Han Y, Zhang D, Chen P, Zhang X, Yuan Z, Huang J, and Zhang X. 2017. Characterization of gene expression profiles in HBV-related liver fibrosis patients and identification of ITGBL1 as a key regulator of fibrogenesis. *Sci Rep* 7:43446.
Weiskirchen R, Weiskirchen S, and Tacke F. 2018. Recent advances in understanding liver fibrosis: bridging basic science and individualized treatment concepts. *F1000Res* 7.
Wurmbach E, Chen YB, Khitrov G, Zhang W, Roayaie S, Schwartz M, Fiel I, Thung S, Mazzaferro V, Bruix J, Bottinger E, Friedman S, Waxman S, and Llovet JM. 2007. Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology* 45:938-947.
Yang D, Li L, Qian S, and Liu L. 2018. Evodiamine ameliorates liver fibrosis in rats via TGF-beta1/Smad signaling pathway. *J Nat Med* 72:145-154.
Yang N, Dang S, Shi J, Wu F, Li M, Zhang X, Li Y, Jia X, and Zhai S. 2017. Caffeic acid phenethyl ester attenuates liver fibrosis via inhibition of TGF-beta1/Smad3 pathway and induction of autophagy pathway. *Biochem Biophys Res Commun* 486:22-28.

Figure 1

Venn diagram of DEGs.

Venn diagram of DEGs from the 3 cohort profile sets (GSE6764, GSE49541, GSE84044), generated using an online tool. Each colored circle represents a different dataset, and areas of overlap indicate shared DEGs. Statistically significant DEGs were defined based on *adj.p* < 0.05 and [FC] > 1.5 as the cut-off criteria.

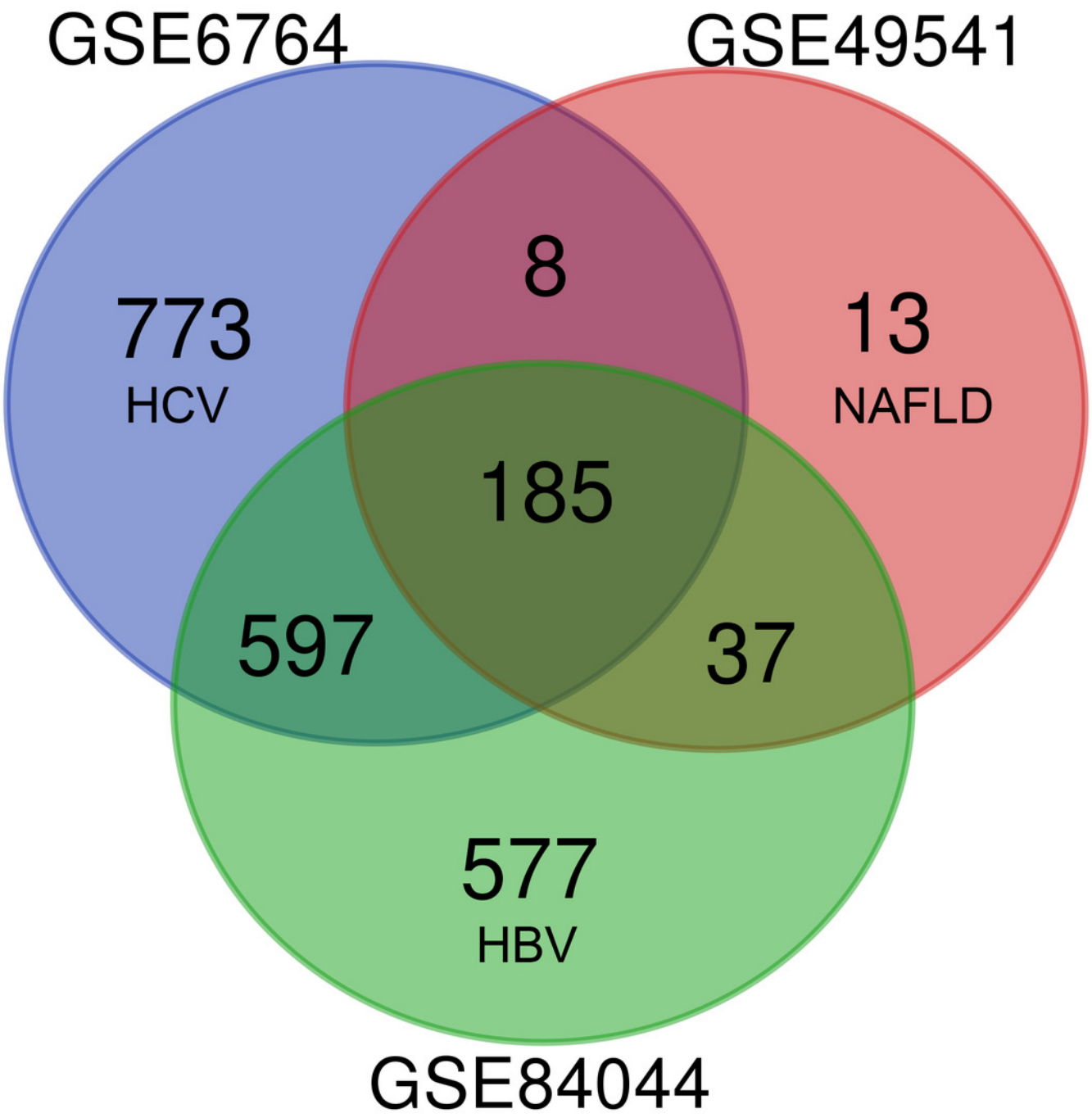
GSE6764
GSE49541
773
HCV
8
13
NAFLD
185
597
37
577
HBV
GSE84044

Figure 2(on next page)

GO enrichment analyses of 158 common DEGs.

The Top 10 terms in each GO category (MF: Molecular function, CC: Cellular components, BP: Biological processes).

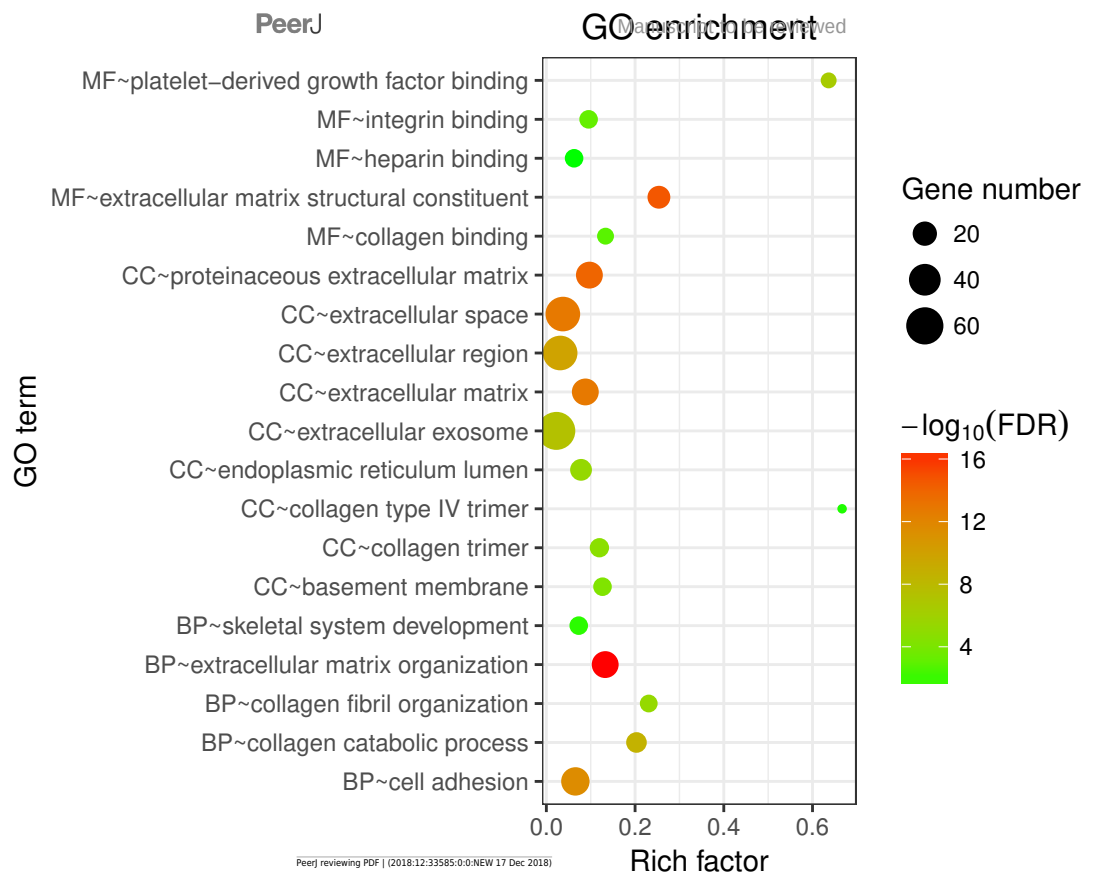

Figure 3(on next page)

KEGG enrichment analyses of 158 common DEGs.

All significant KEGG pathways. GO and KEGG analysis was performed using the DAVID online tool with the cutoff criteria of FDR < 0.05. The color of each bubble represents the FDR for that term, with red representing greater significance. The rich factor refers to the proportion of enriched genes for each term.

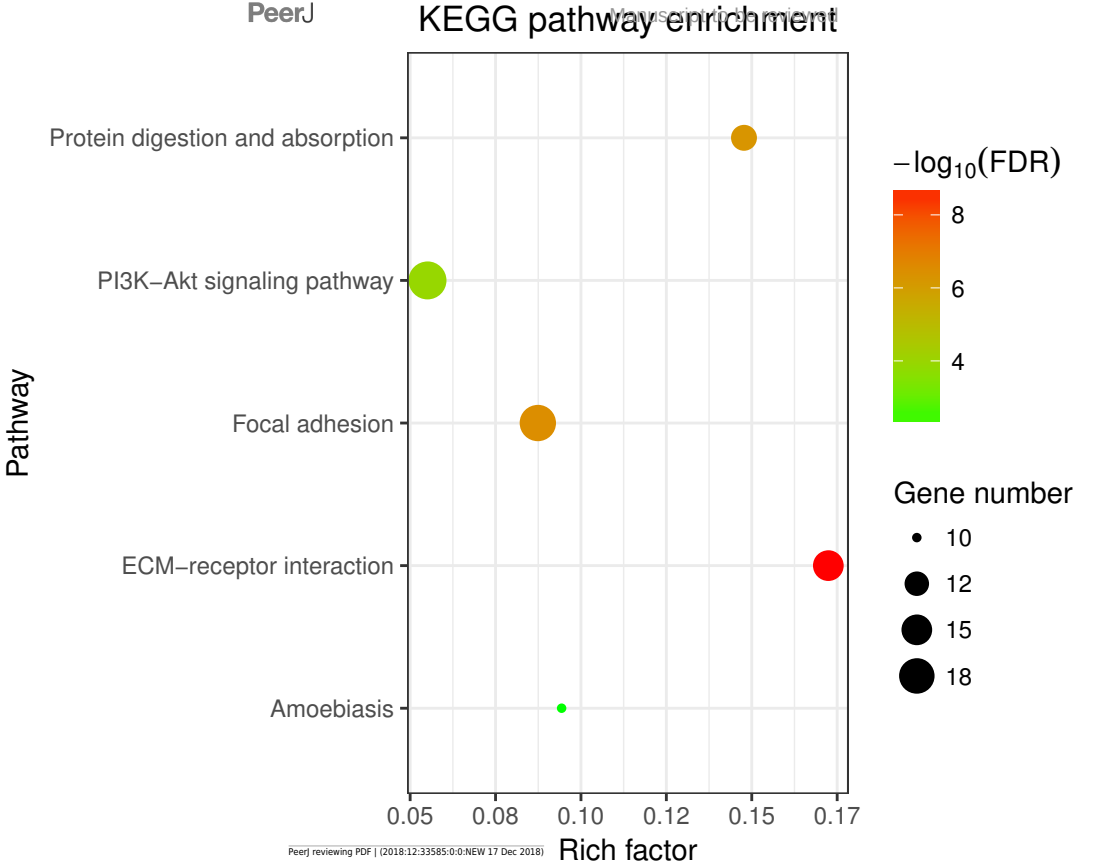

Figure 4(on next page)

Protein–protein interaction (PPI) network complex.

Using the STRING online database, a total of 105 DEGs (101 up-regulated genes shown in Red and 4 down-regulated genes shown in Green) were filtered into a DEG PPI network complex.

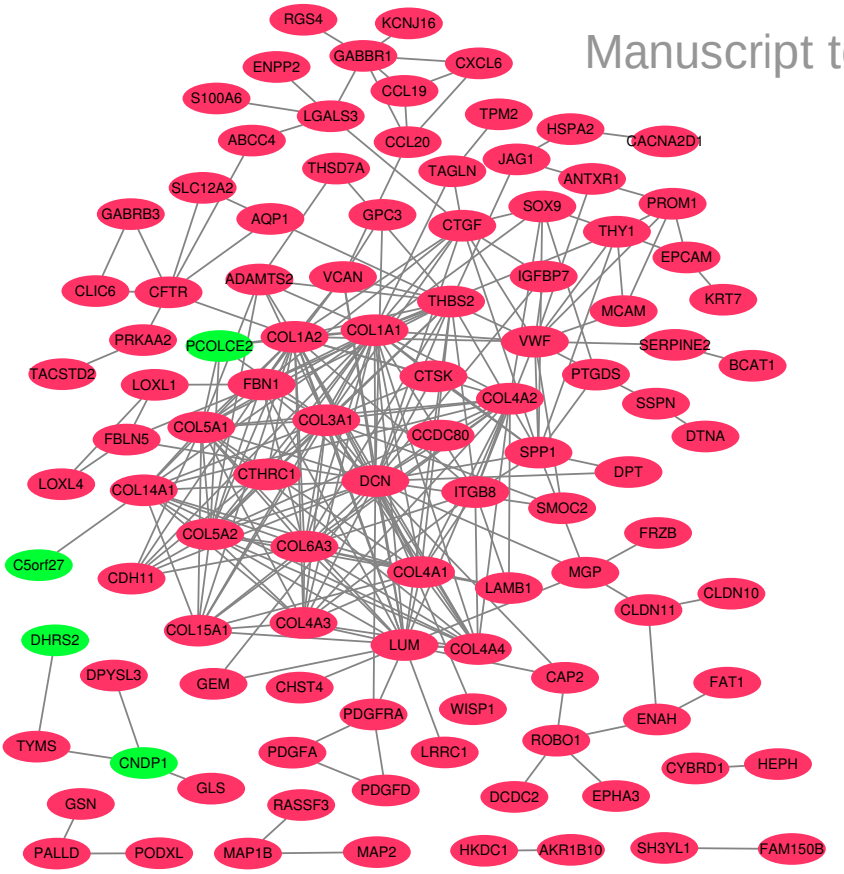
RGS4
KCNJ16
GABBR1
CXCL6
ENPP2
CCL19
S100A6
LGALS3
TPM2
HSPA2
CACNA2D1
ABCC4
CCL20
JAG1
THSD7A
TAGLN
ANTXR1
SLC12A2
SOX9
PROM1
GABRB3
AQP1
GPC3
CTGF
THY1
EPCAM
CLIC6
CFTR
ADAMTS2
VCAN
IGFBP7
KRT7
THBS2
MCAM
PRKAA2
PCOLCE2
COL1A2
COL1A1
VWF
SERPINE2
BCAT1
TACSTD2
LOXL1
FBN1
CTSK
PTGDS
COL4A2
SSPN
COL5A1
COL3A1
DTNA
FBLN5
CCDC80
SPP1
DPT
LOXL4
CTHRC1
DCN
COL14A1
ITGB8
SMOC2
FRZB
COL5A2
COL6A3
C5orf27
CDH11
COL4A1
MGP
LAMB1
CLDN10
CLDN11
COL15A1
COL4A3
DHRS2
LUM
COL4A4
DPYSL3
GEM
CHST4
CAP2
FAT1
WISP1
PDGFRA
ENAH
TYMS
ROBO1
PDGFA
LRRC1
CNDP1
CYBRD1
HEPH
GLS
PDGFD
DCDC2
EPHA3
GSN
RASSF3
PALLD
PODXL
MAP1B
MAP2
HKDC1
AKR1B10
SH3YL1
FAM150B

Figure 5(on next page)

Top 25 hub gene network.

The top 25 genes derived from the MMC method were chosen using the CytoHubba plugin. Advanced ranking is represented by a redder color.

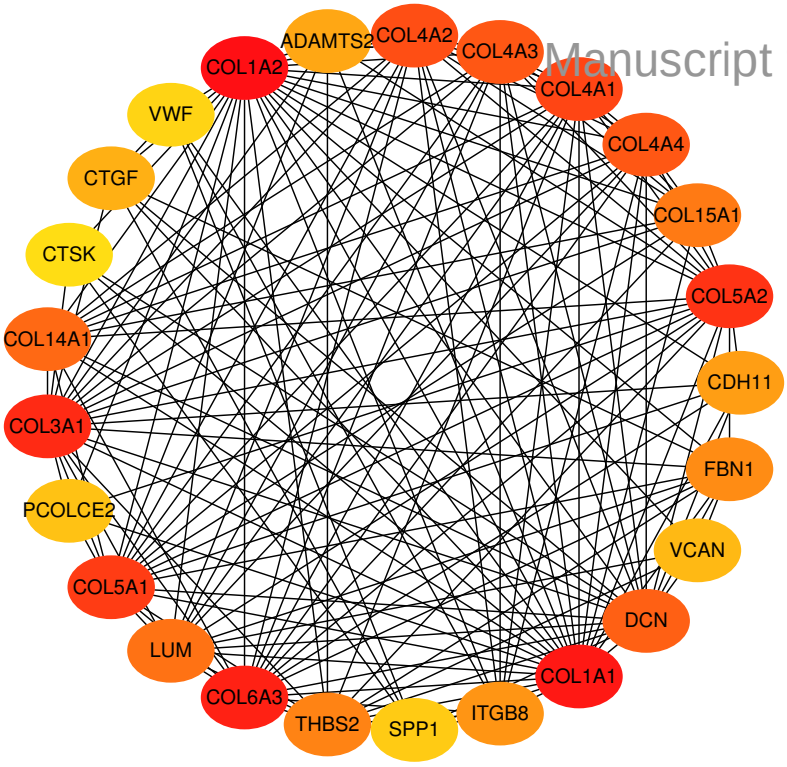
ADAMTS2
COL4A2
COL4A3
COL1A2
COL4A1
VWF
COL4A4
CTGF
COL15A1
CTSK
COL5A2
COL14A1
CDH11
COL3A1
FBN1
PCOLCE2
VCAN
COL5A1
DCN
LUM
COL1A1
COL6A3
ITGB8
THBS2
SPP1

Figure 6

Heatmap of the expression of the 25 hub genes in the GSE14323 validation dataset.

Red colored: up-regulated; Green colored: down-regulated.

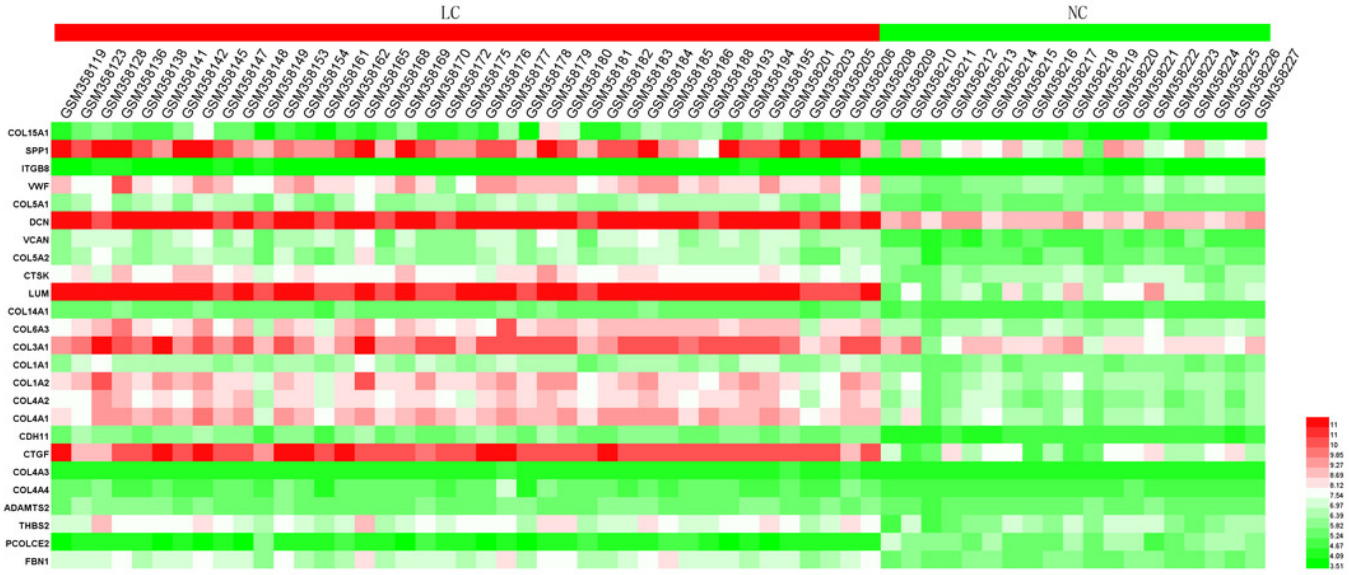

Figure 7 (on next page)

Statistical analysis of the expression of 25 hub genes in GSE14323.

The differences in expression of all hub genes between the liver cirrhosis (LC) group and the normal control (NC) group were statistically significant with the exception of ITGB8. LC: liver cirrhosis; NC: normal control. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

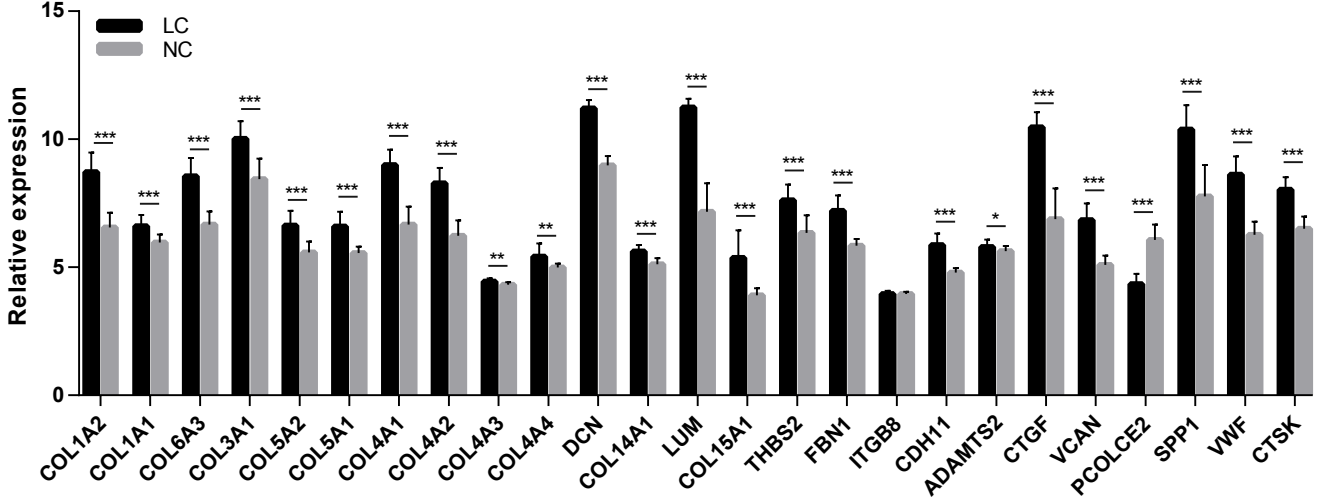
LC
NC
Relative expression
15
10
5
0
***
**
*
COL1A2
COL1A1
COL6A3
COL3A1
COL5A2
COL5A1
COL4A1
COL4A2
COL4A3
COL4A4
DCN
COL14A1
LUM
COL15A1
THBS2
FBN1
ITGB8
CDH11
ADAMTS2
CTGF
VCAN
PCOLCE2
SPP1
VWF
CTSK

Figure 8

Successful construction of liver fibrosis cell model.

After treated with TGF-β1 for 24 hours， the LX2 cells morphology became irregular and extended more ten(A)tacles , and expressed more a-SMA protein (One of the markers of hepatic stellate cell activation) determined by Western Blot (B).

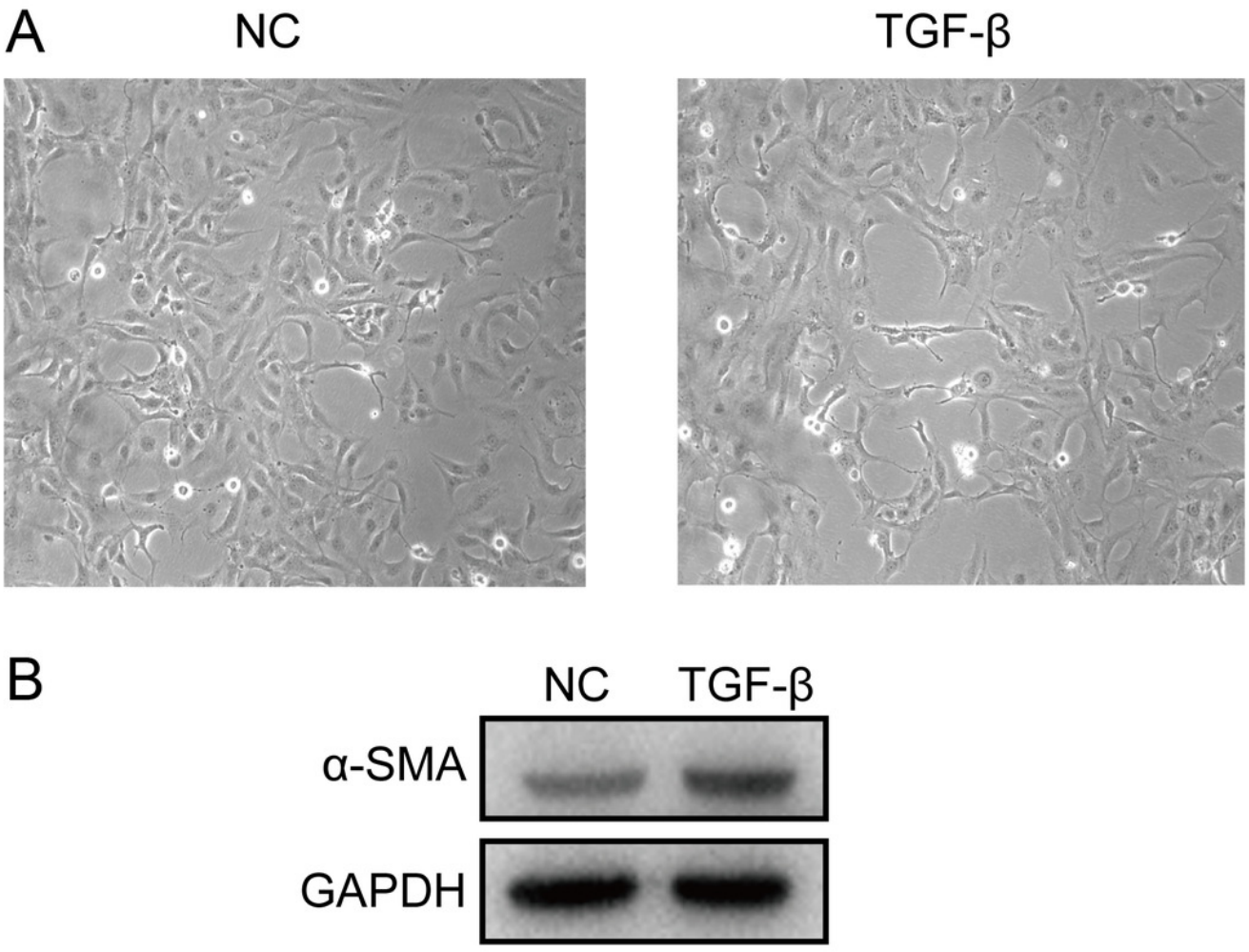

Figure 9(on next page)

mRNA expressions of 25 hub genes in cell model by qPCR.

When LX2 cells were activated by TGF-β1， 13 genes significantly up-regulated which is consistent with the trend of microarray data in this study. 4 genes (LUM, THBS2, ITGB8 and SPP1) was down-regulated.

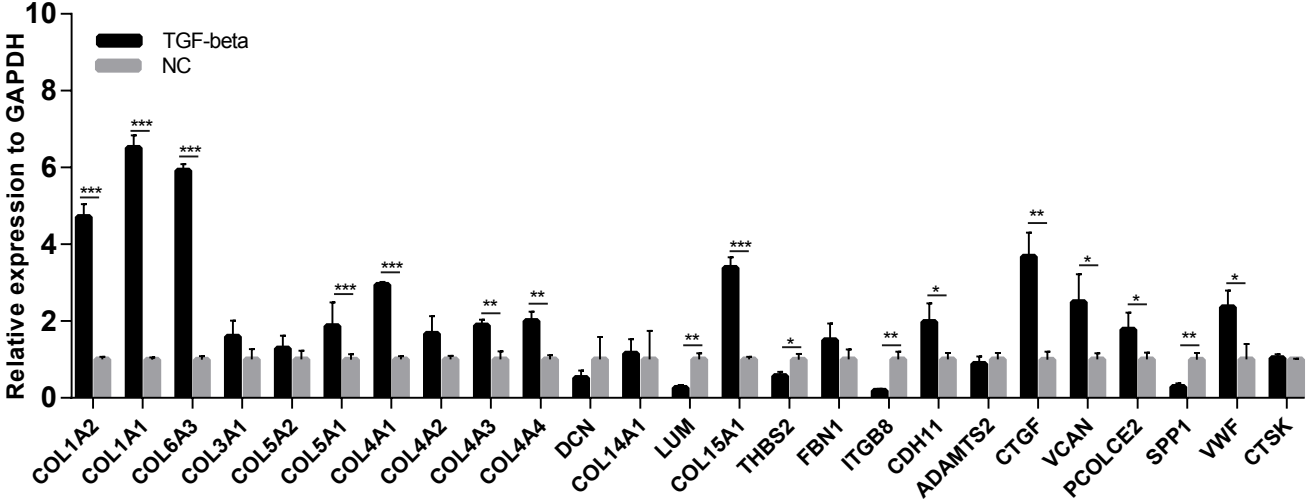
Relative expression to GAPDH
10
8
6
4
2
0
TGF-beta
NC
***
***
***
***
***
**
**
**
***
*
**
*
**
*
*
**
*
COL1A2
COL1A1
COL6A3
COL3A1
COL5A2
COL5A1
COL4A1
COL4A2
COL4A3
COL4A4
DCN
COL14A1
LUM
COL15A1
THBS2
FBN1
ITGB8
CDH11
ADAMTS2
CTGF
VCAN
PCOLCE2
SPP1
VWF
CTSK

Table 1(on next page)

Accession information for datasets downloaded from the GEO database.

GSE6764, GSE49541 and GSE84044 was used for identifying DEGs; GSE14323 was used for validation.

Accession	GPL	Etiology	Sample Size Case/control	Sample Fibrosis Stage	Country	Year
GSE6764	GPL570	HCV	10/13	F4/FO	USA	2007
GSE49541	GPL570	NAFLD	32/40	F3-F4/F0-F1	USA	2013
GSE84044	GPL570	HBV	28/63	F3-F4/F0-F1	China	2016
GSE14323	GPL571	HCV	41/9	F4/F0	USA	2009

1

Table 2(on next page)

Top 10 GO_BP terms in each datasets ordered by FDR.

Count: number of genes enriched in the corresponding pathway; FDR, false discovery rate.

GO ID	Biological process	Count	FDR
GSE6764-HCV			
GO:0030198	extracellular matrix organization	59/196	8.35E-17
GO:0060337	type I interferon signaling pathway	33/64	3.82E-16
GO:0006955	immune response	88/421	7.84E-15
GO:0051607	defense response to virus	47/165	9.18E-12
GO:0060333	interferon-gamma-mediated signaling pathway	30/71	1.73E-11
GO:0007155	cell adhesion	79/459	2.87E-08
GO:0007165	signal transduction	144/1161	9.27E-06
GO:0050900	leukocyte migration	31/122	1.34E-05
GO:0045071	negative regulation of viral genome replication	17/40	2.56E-05
GO:0042493	response to drug	53/304	5.86E-05
GSE49541-NAFLD			
GO:0030198	extracellular matrix organization	29/196	6.55E-18
GO:0007155	cell adhesion	37/459	7.26E-15
GO:0030574	collagen catabolic process	13/64	6.27E-08
GO:0030199	collagen fibril organization	9/39	5.87E-05
GSE84044-HBV			
GO:0006955	immune response	87/421	2.1693E-16
GO:0007155	cell adhesion	81/459	7.1809E-11
GO:0030198	extracellular matrix organization	46/196	2.2322E-09
GO:0070374	positive regulation of ERK1 and ERK2 cascade	38/175	2.6745E-06
GO:0070098	chemokine-mediated signaling pathway	23/71	3.4269E-06
GO:0002250	adaptive immune response	33/148	1.927E-05
GO:0060326	cell chemotaxis	21/65	2.1359E-05
GO:0006954	inflammatory response	58/379	9.3052E-05
GO:0002548	monocyte chemotaxis	16/42	0.00015641
GO:0030574	collagen catabolic process	19/64	0.00052759

1

Table 3(on next page)

Top 10 KEGG pathways in each dataset ordered by FDR.

*Pathway is unique in the corresponding dataset.

KEGG ID	Pathway	Count	FDR
GSE6764-HCV			
hsa04514	Cell adhesion molecules (CAMs)	44/142	2.43E-10
hsa05332	Graft-versus-host disease	18/33	4.53E-07
hsa05330	Allograft rejection	18/37	4.39E-06
hsa04940	Type I diabetes mellitus	19/42	6.01E-06
hsa05416	Viral myocarditis	22/57	8.34E-06
hsa04510	Focal adhesion	46/206	1.08E-05
hsa04512	ECM-receptor interaction	27/87	2.20E-05
hsa04612	Antigen processing and presentation	25/76	2.37E-05
hsa05164	Influenza A	39/174	1.64E-04
hsa05168	Herpes simplex infection	40/183	2.22E-04
GSE49541-NAFLD			
hsa04512	ECM-receptor interaction	17/87	3.17E-10
hsa04510	Focal adhesion	21/206	6.22E-08
hsa04974	Protein digestion and absorption	13/88	1.32E-05
hsa04151	PI3K-Akt signaling pathway	22/345	9.83E-05
hsa05146	Amoebiasis	11/106	0.007509
GSE84044-HBV			
hsa04512	ECM-receptor interaction	28/87	7.34E-07
hsa05323	Rheumatoid arthritis	27/88	5.01E-06
hsa04110	Cell cycle*	32/124	1.4E-05
hsa04672	Intestinal immune network for IgA production	18/47	9.61E-05
hsa05150	Staphylococcus aureus infection	19/54	0.000169
hsa05222	Small cell lung cancer	23/85	0.001041
hsa04151	PI3K-Akt signaling pathway	56/345	0.001159
hsa04510	Focal adhesion	39/206	0.00179
hsa05166	HTLV-I infection	45/256	0.00197
hsa04974	Protein digestion and absorption	23/88	0.001973

1

Table 4(on next page)

Top 10 compounds predicted to have activity against liver fibrosis as predicted via connectivity map.

*Targets were matched to DEGs in selected datasets.

ID	Median_ Score	Name	Description	Target
BRD-K99029477	-93.11	Prometon	photosynthesis inhibitor	
BRD-K19554809	-88.43	MK-212	serotonin receptor agonist	HTR2A, HTR2B*, HTR2C
BRD-A68631409	-85.73	Evodiamine	ATPase inhibitor, TRPV agonist	TRPV1
BRD-K70557564	-85.48	Zosuquidar	P glycoprotein inhibitor, P glycoprotein modulator	ABCB1*, ABCB4
BRD-A61858259	-83.89	CAY-10415	insulin sensitizer	INS
BRD-K09900591	-81.7	Caffeic-acid	lipoxygenase inhibitor, HIV integrase inhibitor, NFkB pathway inhibitor, nitric oxide production inhibitor, PPAR receptor modulator, tumor necrosis factor production inhibitor	ALOX5*, MIF, RELA, TNF
BRD-A82238138	-81.69	Budesonide	glucocorticoid receptor agonist, glucocordicoid receptor antagonist, immunosuppressant	NR3C1
BRD-K52080565	-80.2	Rilmenidine	adrenergic receptor agonist, imidazoline receptor agonist	NISCH
BRD-K66175015	-75.46	Afatinib	EGFR inhibitor, receptor tyrosine protein kinase inhibitor, tyrosine kinase inhibitor	EGFR, ERBB2, ERBB4
BRD-K82357231	-74.73	Desloratadine	histamine receptor antagonist	HRH1

1

Table 5(on next page)

mRNA expression levels of target genes in each dataset.

logFC: log2(Fold change) ; Genes with expression level of |logFC| >0.585 and adj.P.Val<0.05 and were deemed to be the DEGs in this study.

Gene	GSE6764 (HCV)		GSE49541 (NAFLD)		GSE84044 (HBV)	
symbol	logFC	adj.P.Val	logFC	adj.P.Val	logFC	adj.P.Val
HTR2B	1.24	1.63E-02	0.33	2.26E-01	0.74	2.61E-06
ABCB1	0.75	2.24E-02	0.26	1.75E-01	0.76	9.61E-08
ALOX5	0.99	2.88E-02	0.23	2.72E-01	1.06	7.79E-09

1