Explore potential biomarkers to provide a theoretical basis for the treatment of liver fibrosis with pirfenidone

Hao Li¹,², Xi-Yang Dong¹,², Qin Zhou³, Zhi-Xiang Ding¹, Qing-Hai Wang⁴, De-Hui Li⁴*

¹Department of Infectious Diseases, The First People's Hospital of Changde City, Changde 415000, China. ²Intensive Care Unit, The First People's Hospital of Changde City, Changde 415000, China.

*Hao Li and Xi-Yang Dong are the co-first authors of this paper.

Abstract

Background: Hepatic fibrosis is a common chronic liver disease in clinic, the purpose of our study is to explore potential biomarkers to provide a theoretical basis for the treatment of liver fibrosis with pirfenidone. Methods: We downloaded a gene-sequencing dataset and a single-cell dataset from the GEO database and pirfenidone target genes from three different databases. First, we performed GO, KEGG, and DO analysis on pirfenidone target genes. Then, we grouped the liver tissue sequencing data (GSE162694) in the sequencing data set (N-F0 group and F1-F4 group) and performed gene expression differential analysis on these two groups, weighted gene co-expression network analysis and gene Enrichment analysis. Finally, we intersected the significantly upregulated genes in the F1-F4 group with the pirfenidone target genes and performed PPI network analysis. In order to further explore the expression of both pirfenidone drug target genes and liver fibrosis disease genes (PDLFG) in different immune cells of liver tissue, we used the CD45 + cell data in the GSE136103 data set for further analysis. Results: A subnetwork consisting of CDEC42, HNF4A, BHLHE40, CDDC71L, NR1H3, TNF, MGLL, GPT, SCD and PLIN1 was screened out, and by analysis, we finally identified the SCD as PDLFG. In single-cell sequencing analysis, we found that SCD was highly expressed in M2-polarized macrophages. Conclusion: SCD may be an important target protein to inhibit the progression of liver fibrosis.

Keywords: hepatic fibrosis; pirfenidone target genes; biomarker; GSEA; WGCNA; single cell sequencing analysis

Citation

**Introduction**

Hepatic fibrosis, a persistent and chronic hepatic affliction, is attributed to various etiological factors, including viral infections such as Hepatitis B and C [11], excessive alcohol consumption [2], metabolic anomalies [3], non-alcoholic steatohepatitis [4], drug toxicity [5], and cholestasis [6]. The primary pathogenesis of hepatic fibrosis involves the incessant accumulation of extracellular matrix within the hepatic lobule [7], the transitional activation and proliferation of hepatic stellate cells [8], and alterations in hepatic hemodynamics [9]. These pathophysiological triggers culminate in structural changes in the hepatic lobules.

Pirfenidone, a synthetic pyridone derivative, is clinically employed in the management of idiopathic pulmonary fibrosis and also has a certain role in treating pulmonary fibrosis after COVID-19 infection [10]. Despite the ambiguity surrounding the precise mechanism of pirfenidone, several investigations suggest its potential role in downregulating critical inflammatory mediators such as transforming growth factor-β, tumor necrosis factor-α, and interleukin-4 [11, 12], thereby modulating inflammatory responses and impeding collagen synthesis in tissues. These properties confer a reduction in the extent of tissue fibrosis, achieving the therapeutic objective.

While pirfenidone holds a pivotal role in anti-fibrotic therapy, its application is restricted due to associated toxic side effects such as dermatological reactions, vertigo, nausea, anorexia, gastroesophageal reflux, and hepatic dysfunction [13]. Nonetheless, prior research indicates that adjusting the dosage of pirfenidone can ameliorate these side effects [14].

Our investigation seeks to identify potential biomarkers during the progression of liver fibrosis via the pharmacological target sites of pirfenidone. This endeavor aims to provide a theoretical foundation for the clinical utility of pirfenidone in managing hepatic fibrosis.

**Materials and methods**

**Acquisition of high-throughput sequencing data of hepatic fibrosis and pirfenidone drug targets**

High-throughput sequencing datasets of liver fibrosis and single-cell sequencing datasets were procured from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The GSE162694 dataset encompasses 143 liver tissue samples from nonalcoholic steatohepatitis (NASH) patients, distributed across varying stages of fibrosis as per the NASH grading standards. Pirfenidone target genes were retrieved from three distinct databases (CTD [http://ctdbase.org/], Swiss-Target-Prediction [http://www.swissargettprediction.ch/], and target-net [http://targetnet.scbdd.com/home/index/]) but no associated target genes were identified in the binding database. The GSE162694 dataset facilitated the exploration of gene expression variances between different NASH stages, while the GSE136103 dataset was used to discern differences in cellular and gene expression between healthy and cirrhotic liver tissues.

**Bioinformatic analysis of pirfenidone drug target genes and identification of differential expression genes across NASH stages**

The intersecting pirfenidone target genes from the three databases were consolidated for further analysis. Sample data from GSE162694 were bifurcated into control (normal and F0 liver tissue) and hepatic fibrosis groups (F1, F2, F3, F4 stage liver tissue). GO, KEGG, and DO analyses of pirfenidone target genes were executed using “clusterProfile” and “DO” R packages, while the “edgeR” R package was employed to scrutinize gene expression differences between the control and hepatic fibrosis groups.

**Weight gene co-expression network analysis**

Applying the “WGCNA” R package to the GSE162694 transcriptome sequencing data, we performed an analysis of variance and selected the genes exhibiting the most significant variance differences. Upon selecting an appropriate threshold, we proceeded to construct a lead matrix and a topological overlap matrix to investigate co-expression modules. After establishing correlations between different modules, nine gene co-expression modules were identified, with the magenta and dark red modules demonstrating a high correlation with the two phenotypes.

**Gene expression difference and gene set enrichment analysis between N-F0 and F1-F4 groups**

Calculating gene expression disparities between the N-F0 and F1-F4 groups in the GSE162694 dataset, we performed gene set enrichment analysis (GSEA) on the differential genes. The resultant gene sets underwent further analysis.

**Protein-protein interaction network analysis and pirfenidone drug target genes**

Intersecting the Magenta-deg set and Dark red-deg set with pirfenidone target genes, we observed shared genes, which were then subjected to GO, KEGG, DO analyses. The combined genes were uploaded to the STRING database to obtain protein-protein interaction (PPI) network connections, which were then visualized using Cytoscape.

**Differences in the immune microenvironment between N-F0 and F1-F4 group and single-cell sequencing analysis**

The “ssGSEA” algorithm, implemented in the ‘GSEA’ R package, was used to evaluate and compare the immune microenvironments of N-F0 stage and F1-F4 stage liver tissues. After preprocessing single-cell sequencing data, we performed single-cell annotation and examined differences in the expression of PDL1F among individual cell subpopulations. The “CellChat” R package was utilized to analyze and compare communication between cell subsets in the normal and cirrhotic liver groups.

**Statistical analysis**

All statistical evaluations in this study were conducted using R-software (version 4.2.1). Various methodologies were incorporated to assess the gene expression discrepancies between the N-F0 group and the F1-F4 group, including empirical Bayesian estimation, exact tests, generalized linear models, and quasi-likelihood tests. The weighted gene co-expression network analysis facilitated the assignation of similarly expressed gene modules. To examine the correlations between these modules and the clinical phenotype, Pearson correlation tests were employed. The non-linear dimensionality reduction was utilized for the dimensionality reduction and visualization analysis of single-cell data. The Wilcoxon Rank Sum test was further applied to undertake a differential analysis of highly expressed genes among various cell subpopulations.

**Results**

**Identification and analysis of pirfenidone target genes**

The pirfenidone target genes, sourced from three databases, are enumerated in Figure 1a. Following the removal of duplicates, a total of 118 pirfenidone target genes were identified. Biological process (BP) analysis under Gene Ontology (GO) revealed key pathways, including positive regulation of MAPK cascade, response to xenobiotic stimulus, and regulation of tube diameter. Cellular component (CC) analysis emphasized significant enrichment in integral and intrinsic components of presynaptic and synaptic membranes. Molecular function (MF) analysis highlighted neurotransmitter receptor activity and G protein-coupled amine receptor activity, among others (Figure 1b). These findings suggest that pirfenidone target genes are implicated in cell damage and post-damage response. Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis revealed enrichment in neuroactive ligand-receptor interaction, AGE-RAGE signaling pathway in diabetic complications, and serotonergic synapse, among others (Figure 1c). Disease ontology (DO) analysis pointed towards a connection between pirfenidone...
target genes and various common chronic diseases of vital human organs, including arteriosclerotic cardiovascular disease, pulmonary hypertension, and cerebrovascular disease (Figure 1d).

**Differential gene expression and weighted gene co-expression network analysis**

Differential gene expression between the N-F0 and F1-F4 groups is depicted in Figure 2a (P value < 0.05, log2FC > 1.0). Utilizing weighted gene co-expression network analysis (WGCNA), 26 samples were eliminated based on the agglomerative hierarchical clustering algorithm. The best soft threshold power was determined to be 16 (R² = 0.8520), suggesting a network that closely mirrors the true biological state. The N-F0 and F1-F4 groups were used to represent distinct clinical phenotypes, and the dynamic cut tree and color module results are illustrated in Figure 2e–Figure 2d. Both the magenta and dark modules demonstrated significant differences between the N-F0 and F1-F4 groups.

In order to further investigate the gene-disease association, we analyzed the upregulated genes in the disease group (adj.P.val < 0.05, log2FC > 0.5, 1059 genes) that intersected with the darkened module (19 genes) and magenta module (11 genes) gene-set (Figure 2e–Figure 2f).

**Gene set enrichment analysis, PPI network analysis, and PDLFG**
Gene set enrichment analysis (GSEA) of differentially expressed genes between the N-F0 group and the F1-F4 group is presented in Figure 3a. This analysis revealed significant upregulation of pathways related to Graft-versus-host disease, Allograft rejection, and Malaria in F1-F4 groups, while pathways associated with ascorbate and aldarate metabolism and drug metabolism-cytochrome P450 were significantly upregulated in the N-F0 group.

The protein-protein interaction (PPI) of dark-red deg and magenta-deg are shown in Figure 3b and Figure 3c (confidence score > 0.40). The only target gene of pirfenidone in the dark-red module was the SCD gene, with no corresponding target in the magenta-deg gene set (Figure 3d). Further analysis utilized the MCODE plug-in in Cytoscape to locate and visualize the subnetwork gene group centered on the SCD gene (node = 10, degree = 10) (Figure 3e and Figure 3f). GO, KEGG, and DO analyses were then conducted on these genes (CDC42, HNF4A, BHLHE40, CCDC71L, NR1H3, TNF, MGLL, GPT, SCD, PLIN1). The major activated pathways identified in GO-BP analysis were circadian rhythm and rhythmic process. Due to the limited number of enriched genes, the remaining GO-BP and GO-MF pathways were less enriched and thus considered to have limited significance. No statistically significant gene enrichment was found in GO-CC (Figure 3g). KEGG analysis highlighted the non-alcoholic fatty liver disease, PPAR signaling, and regulation of lipolysis in adipocyte pathways. DO analysis revealed that these genes were associated with diseases such as lysosomal storage disease, lipid storage disease, and non-alcoholic fatty liver disease. These findings suggest a connection between genes in this sub-network and liver lipid metabolism. Finally, SCD was identified as a common gene for pirfenidone target sites and liver fibrosis (PDGF).

Analysis of immune microenvironment and single-cell sequencing

Differences in the immune microenvironment of liver tissue between the N-F0 and the F1-F4 groups are depicted in Figure 4a and Figure 4b. Immune cell markers (IMC) analysis (Figure 4a) revealed higher expression of activated dendritic cells (aDCs), CD8+ T cells, Dendritic cells (DCs), Macrophages, Neutrophils, plasmacytoid Dendritic cells (pDCs), Type 1 T-helper cells, Type 2 T-helper cells, and Tumor-Infiltrating Lymphocytes (TIL) in the F1-F4 group. Analysis of immune function markers (IMF) demonstrated higher expression of APC co-inhibition, Chemokine Receptor (CCR), checkpoint, cytolytic-activity, Human Leukocyte Antigen (HLA), inflammation-promoting, MHC class I, Para-inflammation, T cell co-inhibition, and T cell co-stimulation in the F1-F4 group.
Figure 3 The gene set enrichment analysis between N-F0 group and F1-F4 group and PPI networks. (a) The top5 pathways enriched in N-F0 and F1-F4 respectively. (left pathways enriched in F1-F4 group, right pathways enriched in N-F0 group). (b) Protein-Protein Interaction (PPI) network of intersection genes of Magenta Module and up-regulated in F1-F4 group. (c) Protein-Protein Interaction (PPI) network of intersection genes of darked red Module and up-regulated in F1-F4 group. (d) The venn diagram of genes in Pirfenidone targets and IG. IG: intersection genes of darked red Module and up-regulated in F1-F4 group. (e) Protein-Protein Interaction (PPI) network of all genes shown in Figure 3d. (f) Protein-Protein Interaction (PPI) subnetwork including SCD.
To further explore the expression of the SCD gene in subpopulation cells of cirrhotic liver tissue and normal liver tissue and the differences in the immune environment between these two tissues, CD45+ samples obtained from the GSE136103 data set were used for single-cell sequencing analysis. After processing the samples with R-software, a total of 21658 cells in normal liver tissues and 14997 cells in cirrhotic liver tissues were obtained (Figure 4c). Cell annotation revealed that SCD genes were primarily expressed in Macrophage cell subsets derived from monocyte cells major stimulated by IL-4 (M2 Macrophage, adj.P.val < 0.0001, Average expression of Log2FC > 1.0; Figure 4d–Figure 4e). However, no significant difference was found when comparing the expression of SCD in normal liver tissue and cirrhotic liver tissue (Figure 4f).

To further examine potential differences in pathway activation of macrophage cells between normal liver tissue and cirrhotic liver tissue, hallmark gene sets were downloaded from the Molecular Signature Database (https://www.gsea-msigdb.org/gsea/msigdb) for analysis. As displayed in Figure 4g, the number of communication interactions between macrophages and other CD45+ cells in cirrhotic liver tissue was increased compared to normal liver tissue, although the intensity of interaction was decreased.

**Discussion**

Contemporary clinical treatments for liver fibrosis, such as silymarin and glycyrzrhizic acid preparations, primarily function by mitigating oxidative stress, inhibiting the inflammatory response, and curtailing matrix metalloproteinase activity [15, 16]. Utilizing publicly available high-throughput sequencing data from liver tissue biopsies, this study explored differential gene expression across various stages of non-alcoholic Steatohepatitis (NASH), identifying overlaps with pirfenidone drug targets. Single-cell sequencing datasets were further employed for the comparative analysis of normal and cirrhotic liver tissues.

Pirfenidone, a recognized antifibrotic agent, has a well-established role in the clinical management of idiopathic pulmonary fibrosis. Previous research indicates its potential in treating cardiac fibrosis [17], primarily through modulating the inflammatory response and inhibiting collagen synthesis in cardiac tissues. This mechanism bears resemblance to the etiology of liver fibrosis, often the result of perisinusoidal ischemia and hypoxia. Our Disease Ontology (DO) analysis revealed that pirfenidone target genes were implicated not solely in pulmonary fibrosis but also in atherosclerotic cardiovascular diseases, a frequent precursor to cardiac fibrosis following myocardial ischemia.

Gene Set Enrichment Analysis (GSEA) indicated that the F1-F4 group, in comparison to the N-F0 group, exhibited an enhanced immune response and a concomitant decrease in liver metabolic and synthetic functions during liver fibrosis progression. This aligns with our subsequent liver tissue immune microenvironment analysis. Notably, non-alcoholic liver disease is a prevalent cause of liver fibrosis and cirrhosis. Literature suggests a significantly increased activity of most immune cells in the liver tissue of non-alcoholic steatohepatitis patients as compared to normal liver tissue [18].

Stearoyl-CoA desaturase (SCD) is a key enzyme involved in the conversion of saturated fatty acids into monounsaturated fatty acids, thus maintaining cell membrane stability and physiological function and SCD may induce heart failure by inducing angiotensin II AT1 receptor upregulation and Cardiac Lipid Overload [19]. Existing literature reveals that pharmacological inhibition of SCD-1 expression significantly ameliorates hepatic lipid deposition in high-fat diet-fed mice models of non-alcoholic fatty liver disease [20, 21].

Macrophages are implicated in the pathogenesis of chronic liver injury and represent potential targets for antifibrotic therapy [22]. However, their dual role in promoting and eliminating excessive Extracellular Matrix (ECM) deposition during liver fibrosis has been observed. Macrophages typically exhibit M1 polarization during acute liver inflammatory injury and M2 polarization during anti-inflammatory and reparative processes [23]. Our single-cell sequencing analysis of liver tissue revealed a significantly higher expression of SCD in M2-polarized macrophages compared to other CD45+ cells. While the proportion of SCD expression in M2 cells of normal liver tissue and cirrhotic liver tissue did not significantly differ, its high expression in M2 polarized macrophages, coupled with its role as a pirfenidone drug target gene, warrants further investigation.

Our study was constrained by a limited sample size, preventing a comprehensive analysis of different liver fibrosis stages. Additionally, the difference in pathway activation between cirrhotic liver tissue samples with varying SCD expression levels necessitates further
exploration. Lastly, our reliance on a public database necessitates further experimental validation.

While the therapeutic potential of pirfenidone in hepatic fibrosis is hindered by its potential hepatotoxicity, the feasibility of achieving treatment goals through dosage and frequency modulation remains to be elucidated. Our study aims to furnish a theoretical foundation for the clinical application of pirfenidone in liver fibrosis treatment through the identification of shared genes between liver fibrosis.

References