Integrative analysis implicates the significance of m^{6}A in the liver fibrosis of biliary atresia by regulating THY1

Junfeng Wang[1#], Min Du[1,2#], Lingdu Meng[1], Yifan Yang[1], Shiwei He[1], Ye Zhu[1], Xue Ren[1], Meng Wei[1,3], Rui Dong[1], Shan Zheng[1*], Gong Chen[1*]

[1]Department of Pediatric Surgery, Children's Hospital of Fudan University, Shanghai Key Laboratory of Birth Defect, and Key Laboratory of Neonatal Disease, Ministry of Health, Shanghai 201102, P.R. China

[2]Department of Pediatric Gastroenterology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu 610091, P.R. China.

[3]Department of Pediatric Hematology and Oncology, Shanghai Children's Hospital, Shanghai Jiao Tong University, Shanghai 200062, P.R. China.

[#]These authors contributed to the work equally.

[*]Correspondence to*:* Gong Chen or Shan Zheng, Department of Pediatric Surgery, Children's Hospital of Fudan University, Shanghai Key Laboratory of Birth Defect, and Key Laboratory of Neonatal Disease, Ministry of Health, 399 Wan Yuan Road, Shanghai 201102, China. Fax: +86 021 64931901; Tel: +86 021 64932791; Email: chengongzlp@hotmail.com (G.C.), szheng@shmu.edu.cn (S.Z.)

Supplementary Material

Materials and Methods

M^{6}A RNA methylation quantification

The m^{6}A RNA methylation content was determined using a colorimetric quantification kit (ab185912). First, the total RNA was isolated from each sample using Trizol (Takara, 9109). The same amount of RNA (200 ng) was added into each well of a 96-well plate and bounded to the strip wells using an RNA high binding solution. Then, a specific N6-methyladenosine antibody and detection antibody were added into the wells to capture the m^{6}A status of the total RNA. Subsequently, the detected signal was amplified using an enhancer solution and colorimetrically quantified by reading the absorbance in a microplate spectrophotometer (ThermoFisher Scientific, Varioskan LUX) at a wavelength of 450 nm. Finally, the absolute amount of m^{6}A was calculated using a standard curve. The levels of m^{6}A at different groups were visualized via boxplots using the R packages ggpubr. The correlation analyses among m^{6}A and the indicators of liver function were performed using the R package ggcorrplot, and their scatter diagrams were plotted by R package ggplot2.

qRT-PCR

Total RNA isolated from liver tissues or primary BECs was transcribed into cDNA using a reverse transcription kit (Takara, RR047A). qRT-PCR was performed

according to the instructions of TB Green® Premix Ex Taq™ II (Takara, RR820A). Primer sequences of each gene are shown in Table S5. The relative expression level of mRNA was calculated by a $2^{-\Delta\Delta Ct}$ method.

Western blot

Proteins were extracted from liver tissues using the RIPA buffer (Thermo Scientific, 89900) containing inhibitors of proteases and phosphatases (Thermo Scientific, 78440). The total protein concentration was measured using a BCA kit (Beyotime, P0010S). Protein samples (30 ug) were loaded on a SDS-PAGE gel for electrophoresis, and then transferred onto a PVDF membrane through the wet-transfer method. The blotting membrane was then blocked with 5% skim milk for 1 h at room temperature and incubated overnight at 4 °C with the following primary antibodies: METTL3 (Abcam, ab195352), METTL14 (Abcam, ab220030), WTAP (Abcam, ab195380), FTO (Abcam, ab126605), ALKBH5 (Abcam, ab195377), COL1A1 (CST, 72026), α-SMA (CST, 48938), PDGFRB (CST, 3169), THY1 (CST, 13801), MMP2 (40994), β-actin (Proteintech, 60008-1-Ig), and GAPDH (Proteintech, HRP-60004). Next, the membrane was incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies for 1 h at room temperature. Finally, the protein bands were visualized under a chemiluminescence apparatus (BIO-RAD, ChemiDOC™ XRS+) using an ECL kit (Thermo Scientific, 32209).

Co-immunoprecipitation

Human liver lysates (1 mg) were immunoprecipitated with 2 μg of anti-CD90 antibody (ThermoFisher, 14-0909-82) overnight at 4 °C, then incubated with 30 ul PureProteome^{TM} Protein A/G Mix Magnetic Bead (Merck Millipore, LSKMAGAG02) for 4 hours at 4 °C. The beads were captured by PureProteome^{TM} Magnetic Stand (LSKMAGS08) and washed 3 times with IP buffer (ThermoFisher, 87787), re-suspended in 60 ul 1×loading buffer, and heated at 95 °C for 10 min before western blot analysis by using the following antibodies: ITGAX (CST, 45581), ITGB2 (CST, 73663), and THY1 (CST, 13801). Whole liver tissue lysate (30 μg) was used as an input, and Normal Mouse IgG (Merck Millipore, 12-371) was set as the negative control.

Immunohistochemistry staining of liver microarray

Paraffin-embedded liver sections in the microarray were firstly dewaxed and hydrated, followed by antigen retrieval and inactivation of endogenous peroxidases. After antigen blocking with 5% BSA, the sections were incubated overnight at 4 °C with the following primary antibodies: METTL3, METTL14, WTAP, FTO, ALKBH5, COL1A1, α-SMA, PDGFRB, and THY1. Subsequently, the sections were incubated with secondary antibodies conjugated to HRP (Gene tech, GK500710) for 30 min at room temperature. The signals were detected using the DAB substrate and hematoxylin. The IHC images in the liver microarray were captured using a scanner (Pannoramic DESK, 3DHISTECH, Hungary) and browsed using the CaseViewer software (version 2.4, 3DHISTECH). The expression intensity of each protein was

quantified based on the histochemistry score (H-score), which was calculated by the formula: H-SCORE = $\sum$ (pi × i) = (percentage of weak intensity cells × 1) + (percentage of moderate intensity cells × 2) + (percentage of strong intensity cells × 3). In this formula, pi stands for the percentage of positive cells, and i represents the grade of positive cells (negative, point 0; weak yellow, point 1; moderate yellow, point 2; strong yellow, point 3). The H-score in different groups were displayed using boxplots or violin plots.

Immunofluorescence

Multiple immunofluorescence (IF) staining was performed using a tyramide signal amplification kit (Panovue, 10079100020). Liver sections underwent the same pre-processing steps as described in IHC procedures, except that they needn't to be incubated with H_2O_2 to inactivate endogenous peroxidases. According to the kit instructions, the same liver section was treated with 3 cycles of procedures. In each cycle, the liver sections underwent antigen retrieval and antigen blocking, and were incubated overnight with the primary antibodies (COL1A1, α-SMA, PDGFRB, THY1), and then reacted with goat anti-rabbit/mouse IgG for 1 h at room temperature, and colored with a fluorescent dye at different wavelengths (λ 520, λ 570, λ 650) for 15 minutes at room temperature. After the third cycle, the nucleis were stained with DAPI for 15 min at room temperature. Finally, the IF images were captured using a confocal microscope (LEICA, TCS SP8).

In the cell IF procedures, cell membranes were primarily permeated with PBS containing 0.5% Triton. Cells were exposed overnight to the following primary antibody mixtures: COL1A1, α-SMA, MMP2, THY1, MSLN (Santa Cruz, sc-365324), PDGFRA (CST, 3174), CK19 (ThermoFisher, MA5-12663), and DCN (Santa Cruz, sc-73896), then incubated with the following secondary antibody mixtures: anti-mouse (Invitrogen, A21202), anti-mouse (Invitrogen, A31570), anti-rabbit (Invitrogen, A21206), anti-rabbit (Invitrogen, A31572) followed by the nuclear staining with DAPI (Invitrogen, D1306).

RNA isolation, library construction, and MeRIP-seq

Total RNA was isolated from each sample by using Trizol (Takara, 9109). The concentration and purity of the RNA samples were assessed using the NanoPhotometer® spectrophotometer (GE Healthcare) and agarose gel electrophoresis, and the RNA integrity was further checked via Agilent 2100 (Agilent) to confirm the high quality of the samples.

The library construction and high-throughput m^6A sequencing were performed by Beijing Novogene Biotech Co., Ltd. The regular-amount method was adopted for library construction, in which the total RNA obtained from each sample was over 300 ug. PolyA-tailed mRNAs were enriched using Oligo-dT–conjugated magnetic beads and then broken into 100 nt fragments by using the RNA fragmentation buffer (10 mM Tris-HCl [pH 7.0], 10 mM $ZnCl_2$). Subsequently, the fragmented mRNA was divided into two parts. One part was mixed with magnetic beads conjugated with the

m^{6}A antibody (Synaptic System, 202003, 0.5 mg/ml) in the IP buffer (150 mM NaCl, 0.1% Igepal CA-630, 10 mM Tris-HCl [pH 7.4], and RNasin Plus RNase inhibitor) to enrich for the m^{6}A-modified mRNAs and construct the IP library. The other part was used to prepare the input library. Library construction included the following steps: Double-stranded cDNA was synthesized from the mRNA template by using random hexamer primers, and the remaining overhangs were converted into blunt ends by using an exonuclease/polymerase. Then, adenylation was performed on the 3' ends of the DNA fragments, and NEBNext adaptors with hairpin loops were ligated to the fragments for hybridization. Next, the library fragments were purified using the AMPure XP system (Beckman Coulter, Beverly, USA) to preferentially select 150-200 bp cDNA fragments. After adding the USER enzyme (NEB, USA), the adaptor-ligated cDNA was heated at 37 ℃ for 15 min and then at 95 ℃ for 5 min. Subsequently, PCR amplification was performed using Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. Finally, the cDNA library was purified using the AmPure XP system, and the quality of the library was assessed on the Agilent Bioanalyzer 2100 system. Finally, the IP and input libraries were sequenced on an Illumina NovaSeq 6000 platform to generate 150 bp paired-end reads.

Data analysis in MeRIP-seq

The raw reads of fastq format were pre-processed using the Trimmomatic software by removing the reads with a mean quality less than 20 or more than 15% N bases, and

align with the adapter sequence or are below 18 bases. After all the filtering steps, the clean reads were obtained for the subsequent analysis. The clean reads were aligned to the reference genome (GRCH38) by using the Burrows-Wheeler Aligner (BWA) software.[1] After fragment size estimation by using the RSeQC software, the exomePeak R package (version 2.16.0) was used for the m^{6}A peak identification in each IP group with the corresponding input sample serving as a control. The FDR threshold of enrichment was set to 0.05 for all the data sets. The statistics of m^{6}A peak calling are shown in Table S6. The characteristics of m^{6}A-modified peaks, including the chromosome distribution, peak width, fold enrichment, FDR, and peak summit number per peak, were all analyzed. The m^{6}A modification sites, identified using HOMER (version 4.9.1), have the consensus RRACH sequence (where R represents an A or G, A is m^{6}A, and H is a non-G base). PeakAnnotator was used to identify the nearest transcript start sites (TSS) of genes. The distribution of the peak summits on the RNA regions including 5' UTR, TSS, CDS, stop codon, and 3'UTR was analyzed. Differential peak analysis was based on the fold enrichment and *P* value of the peaks in the BA group versus the control group. A significant peak was determined when the absolute value of fold enrichment exceeded 1.25 and *P* value less than 0.05. The differential m^{6}A peaks were visualized using the integrative genomics viewer (IGV). GO and KEGG analyses of the genes corresponding to the differential m^{6}A peaks were performed using clusterProfile R package.[2] The GO and KEGG terms with corrected *P*-values less than 0.05 were considered significant.

Data analysis in RNA-seq

The procedures of RNA isolation and library preparation were the same as those described for MeRIP-seq analysis. The clean reads were aligned to the reference genome (GRCh38) by using the HISAT2 software. FeatureCounts (v1.5.0-p3) was applied to count the number of reads that map to each gene, and then Fragments Per Kilobase of exon model per Million mapped fragments (FPKM) of each gene was calculated to represent the gene expression levels. Additionally, the DESeq2 R package (v1.16.1) was used to determine the differentially expressed genes between the control and BA groups. Genes with a *P*-value less than 0.05 and the absolute value of fold change exceeds 1.5 were assigned as differentially expressed genes (DEGs). The DEGs were displayed on a volcano map. GO and KEGG analyses of the DEGs were also performed using the clusterProfile R package.

Integrated analysis of MeRIP-seq and RNA-seq

To compare the density of m^6A peaks in BA and controls, genes in BA and control groups were divided into 10 parts according to their expression levels, the average number of each part's common m^6A peaks shared by all samples in BA and control groups was calculated. The overlapping genes between the differential m^6A peak-related genes (|fold change| > 1.25 & $P < 0.05$) and differentially expressed genes (DEGs) (|fold change| > 1.5 & $P < 0.05$) in the RNA-seq data was obtained using a Venn diagram. GO analyses were performed by clusterProfile R package [2] (*P* & adjusted *P* value < 0.05) to acquire the functional enrichment of differential m^6A

modified DEGs in BA and controls. According to the m^{6}A and RNA levels, a four-quadrant diagram was drawn, and the correlation between the log2fold change of the RNA expression and the log2fold change of the m^{6}A levels was analyzed.

Identification of m^{6}A methylated target genes

To narrow down the overlapping genes, the single-cell RNA sequencing data in GSE136103, acquired from the liver of 5 healthy controls and 5 cirrhotic patients, was used to visualize the expressions of overlapping genes in the liver cell population. According to the enriched fibrotic pathways in overlapping genes, the genes highly expressed in hepatic stellate cells (HSCs) and myofibroblasts (MFs) were selected for further analysis. To identify the hub genes, protein-protein interaction (PPI) network of the overlapping genes was primarily constructed by the STRING website (https://string-db.org), and the top 5 hub genes were screened by the MCC algorithm in CytoHuba, a tool of Cytoscape, with default parameters. Combing the bulk and single-cell results, we determined hyper-methylated and up-regulated genes *COL1A1* and *THY1* as the m^{6}A modified targets in BA. The m^{6}A peak map of target genes in each sample was drawn using the Integrative Genomics Viewer (IGV) browser.

MeRIP-qPCR

Total RNA was isolated from the liver tissues using Trizol (Takara, 9109). A total of 200 ug RNA was used for MeRIP, and 1 ug RNA was used as the input. Then, 10 mg Dynabeads was coupled to 50 ug m^{6}A antibody (Synaptic System, 202003) and

suspended in 1 ml SB by following the instructions of the Dynabeads® antibody-coupling kit (Invitrogen, 14311D). Subsequently, one-third volume (330 ul) of the m^{6}A-coupled bead suspension was incubated with 200 ug RNA in 1 ml immunoprecipitation buffer (150 mM NaCl, 0.1% NP-40, 10 mM Tris-HCl pH 7.4, and 10 U/ml RNase inhibitor) at 4 ℃ with continuous rotating. The m^{6}A-antibody-RNA-bead complex was captured using the PureProteome™ Magnetic Stand (Merck Millipore, LSKMAGS08), and m^{6}A-immunoprecipitation RNA was further obtained by washing the beads four times with the elution buffer (50 mM pH 7.5 Tris-HCl, 20 mM DTT, 0.15 mM NaCl, 1 mM pH 8.0 EDTA, 0.1% SDS, and 2 U/ml RNase inhibitor). Finally, m^{6}A-immunoprecipitation RNA was concentrated by adding one-tenth volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of 100% ethanol, and the pellet was washed with 75% (vol/vol) ethanol and then dissolved in 15 ul DEPC-treated water. The immunoprecipitated RNA and input RNA were reverse-transcribed into cDNA using a reverse-transcription kit (Takara, RR047A). Finally, qRT-PCR was performed according to the instructions of TB Green® Premix Ex Taq™ II (Takara, RR820A). The sequences of the primers used are shown in Table S7. The housekeeping gene *HPRT1* was set as the internal control since it has no m^{6}A modification.[3] The relative fold of RIP fraction in each input was calculated using the $2^{-(\Delta Ct\ IP-\Delta Ct\ input)}$ method.

Single cell solution preparation

The harvest human liver tissue was immediately preserved in the tissue storage solution (Miltenyi Biote, 130-100-008), then transferred into the digestive solution containing collagenase IV (Gibco, 17104-19, 2 mg/ml) and DNase I (Sigma, DN25-100MG, 0.1 mg/ml), and was cut into pieces with sterile scissors, followed by digestion for 30 min in the 37 ℃ constant temperature shaker. The enzymolysis was terminated by adding DMEM containing 10% FBS, and the suspension was then filtered through the 70 um and 30 um Nylon cell strainers, respectively, and centrifuged at 400 g for 5 min. The cell pellets were re-suspended in red blood cell lysis solution (Miltenyi Biote, 130-094-183) to remove red blood cells. After 2 min incubation at room temperature, the mixture was centrifuged at 300 g for 10 min. The cell pellets were finally re-suspended in DMEM/F12 containing 10% FBS to prepare the single cell solution. The cell number of single cell solution was determined by the Countess® II FL (Invitrogen, AMQAF1000). Dead cells were further removed by a magnetic activated cell sorting (MACS) kit (Miltenyi Biote, 130-090-101) according to the manufacturer's instructions. The magnetically labeled dead cells are retained within the column. The unlabeled living cells ran through and were used for the following experiments.

Human primary hepatic stellate cells isolation

Primary HSCs were isolated from normal liver tissue adjacent to the excised hepatoblastoma. The single cell suspension was re-centrifuged and the cell pellets were re-suspended in GBSS/B with 1% DNAses. Subsequently, density gradient

centrifugation was used to isolate HSCs as previous described.[4] Nycodenz solution (Accurate Chemical, 1002424) was added into the cell suspension, and the cell-Nycodenz mixture was divided into 15 ml Falcon tubes. Then GBSS/B was overlayed on cell-Nycodenz mixture, and centrifuge at 1380 g for 17 min without brake. HSCs were harvested by aspirating the layer between the cell-Nycodenz solution and the overlay with GBSS/B. HSC pellets were re-suspended with GBSS/B and centrifuged at 580 g for 10 min, and the cell pellets were finally re-suspended in DMEM/F12 with 10% FBS and seeded in a 6-well plate.

Human primary THY1$^+$ cells isolation

THY1$^+$ cells were positively sorted from the BA livers by the CD90 MicroBeads (Miltenyi Biote, 130-096-253). In detail, cell pellets were re-suspended in 80 μl of MACS separation buffer (Miltenyi Biote, 130-091-222) per 10^7 total cells, and 20 ul CD90 MicroBeads were added into the suspension per 10^7 total cells, and the mixture was incubated for 15 min at 4 ℃. Next, 1-2 ml of buffer per 10^7 cells was added into the mixture to wash cells and the cell suspension was centrifuged at 300 g for 10 min, and re-suspended in 500 ul of buffer every 10^8 cells. A MACS LS column (Miltenyi Biote, 130-042-401) was placed in the QuadroMACS^{TM} separator (Miltenyi Biote, 130-091-051), and rinsed with 3 ml buffer. The above CD90 magnetic labeling cell suspension was loaded onto the column, and 3 ml buffer was added to repeatedly wash the column for 3 times. Remove column from the separator, and add 5 ml buffer onto the column. Eventually, CD90 labeled cells were flushed out by firmly pushing

the plunger into the column and collected in a new tube. Finally, $THY1^+$ cells were suspended in DMEM/F12 containing 10% FBS and seeded into 6-well plates.

Plasmid construction and lentivirus infection

The LX-2 cell line was purchased from Shanghai FuHeng Biology Co., LTD, and cultured in DMEM medium (Gibco) containing 10% fetal bovine serum (Gibco) and 1% penicillin/streptomycin (Gibco) at 37 ℃ in a humidified 5% CO_2 incubator. METTL3 and THY1 overexpression control (OC) and overexpression (OE) lentivirus, and METTL3 negative control (NC) and 3 koncking down (KD) shRNA plasmids were purchased from Shanghai Genechem Co., LTD. METTL14 OC and OE plasmids were purchased from Shanghai Genomeditech Co., LTD. Plasmid was transformed in Trans5α chemically competent cell, amplified in LB medium containing ampicillin, and extracted using the kit (OMEGA, D6915-03). METTL3 NC and KD lentivirus was produced in HEK 293T cells when transfected the vector and target plasmids using jetPEI® DNA transfection reagent (PolyPlus, 101-10N). Given the lowest level of METTL3 in the LX-2's KD2 group, we chose KD2 to knockdown METTL3 in primary HSCs. The permanent METTL3 disturbed LX-2 cells and primary HSCs were generated by the lentivirus infection, and were continuously selected with 2 ug/ml of puromycin for 2 weeks. The transfection efficacy was determined by western blot.

Cell migration

The migration of LX-2 cells varying METTL3 expressions was detected by transwell devices (Corning, 3422). At the day before the experiment, the cells were starved for 24 h by removing the FBS from the complete medium. During this assay, about 1×10^5 cells resuspended in the 200 ul free DMEM medium were placed on the upper layer of a cell culture insert with permeable membrane, while 650 ul DMEM medium with 20% FBS was placed below the cell permeable membrane. Following an incubation for 48 h, the cells that have migrated through the membrane are stained by the crystal violet, photoed in the light microscope and counted by ImageJ (NIH, version 1.52a). For the cell migration in primary HSCs, cell starvation was not performed, and DMEM/F12 with 10% FBS was added into the upper and lower layers.

Bioinformatic analysis of the public bulk datasets

A total of two bulk microarray datasets (GSE46960, GSE15235) were included for analysis.[5, 6] In GSE46960, RNA expression profiling was performed in the liver biopsy samples from 64 BA patients, 14 patients with other causes of intrahepatic cholestasis, and 7 cases of deceased-donor children. GSE15235 is a microarray dataset from 47 BA patients, and the patients' clinical information such as grades of liver inflammation, stage of fibrosis, and the status of NLS was provided. All the data in both datasets were normalized using the Robust Muti-array Average method for further analysis. The genes encoding fibrotic markers, as well as the m^6A-modified genes were visualized by boxplots. To know about the gene profile of patients with

different THY1 levels, all patients in GSE46960 were divided into high (> median) and low THY1 (≤ median) groups. The DEGs among the two groups were determined by the limma R package with *P* value less than 0.05 and fold change > 1.5, and the GO and KEGG analysis of the DEGs was performed by the clusterProfile R package with both *P* and adjusted *P* value less than 0.05. Similarly, the differential gene profile of BA patients with native liver or not was analyzed by the limma R package, and DEGs with *P* value less than 0.05 and fold change > 1.5 were included for GO analysis. To extract the gene modules that were correlated with clinical phenotypes (NLS), weighted correlation network analysis (WGCNA) R package was performed in the dataset of GSE15235. Firstly, the correlation coefficients of the gene expression matrix were weighed to construct the gene modules with similar expression patterns. Then, the key modules that were highly correlated with the status of NLS were selected for further analysis. Finally, genes with higher correlation with both module and clinical trait (NLS) were identified.

Seurat analysis of the public single-cell dataset

One single-cell RNA sequencing (scRNA-seq) dataset (GSE136103) acquired from 5 healthy and 5 cirrhotic livers was included for analysis.[7] Cirrhotic livers were collected from 2 patients with non-alcoholic fatty liver disease, 2 patients with alcoholic liver disease, and 1 patient with primary biliary cirrhosis. All the scRNA-seq data were processed under the analytical workflow of the R package Seurat (https://satijalab.org/seurat/articles/pbmc3k_tutorial.html).[8] Cells that

expressed fewer than 300 genes, genes expressed in less than 3 cells were primarily removed. Subsequently, cells with the number of detected genes per cell in the range of 500-6000 and less than 30% percentage of mitochondrial genes detected were retained for further analyses. After quality control, the gene expression of individual cells was processed, including data normalization (*NormalizeData*), identification of highly variable 2000 genes (*FindVariableFeatures*), linear dimensional reduction (*ScaleData*), PCA reduction (*RunPCA*), determination dimensionality of the dataset (*ElbowPlot*), and cell clustering with resolution = 0.6 (*FindNeighbors* and *FindClusters*). UMAP was applied for the non-linear dimensional reduction of dataset visualization (*RunUMAP*).

Integration of multiple datasets from each sample

To reduce the batch effect among the samples, we adopted the R package Harmony (https://github.com/immunogenomics/harmony) to merge multiple datasets.[9] First, the *merge* function was used to integrate the RDS file of each sample and establish the merging Seurat object. Then, the merged data was processed under the Seurat analytical workflow of the previous steps. After PCA reduction, the *RunHarmony* function was performed to use Harmony instead of PCA for the dataset dimensional reduction and cell clustering. Finally, the clusters were visualized using UMAP with the default parameters. The *FindAllMarkers* function (wilcox statistical test, logfc.threshold = 0.25, min.pct = 0.25) was used to screen the DEGs in a given cluster when compared with all the other clusters, and DEGs in each cluster were visualized

using heatmaps. According to the DEGs and known cell markers, the cell type annotation of each cluster was defined using the *RenameIdents* function. A total of 60880 cells were obtained. The number of each cell type in every sample is shown in Table S8. VlnPlot and Dotplot were used to visualize specific gene expressions. The cell cluster of interest was extracted using the *subset* function and re-analyzed as described above.

Single-cell gene enrichment analysis

To understand the differential gene profile of THY1 positive (count > 0) and negative (count = 0) cells, the *FindAllMarkers* function in Seurat was applied to acquire the DEGs between these two groups. Then, GO analysis was performed to analyze their enriched pathways by clusterProfile R package with both *P* and adjusted *P* value less than 0.05. Furthermore, GSVA [10] was employed to figure out the enriched pathways at single-cell level, the gene sets in GO biological processes (c5.go.bp.v7.5.1.symbols.gmt) were set as the reference, and the results were visualized by heatmaps.

Deconvolution analysis

To estimate of the abundances of cell types in a mixed cell population, CIBERSORTx [11] was used to decode the cell fractions in GSE15235. The single cell counts matrix of GSE136103 was extracted using Seurat R package. In this table, gene symbols are in column 1, and the cell types are in row 1. To acquire the characteristic gene profile

of each cell type, the DEGs of each cell type were filtered out by the *FindAllMarkers* function in Seurat. The single cell reference data was further prepared by extracting the column 1 with DEGs of cell types from the whole cell counts matrix. Then, a signature matrix file was built from the single cell reference data, and the cell fractions were calculated by the CIBERSORTx analytical tool (https://cibersortx.stanford.edu). Finally, the cell composition was compared between the BA patients with NLS or not.

CellChat analysis of the public single-cell dataset

To infer the interactions of THY1 and other proteins in the cirrhotic circumstance, CellChat (Version 1.1.3) [12] was applied to analyze cell-cell communications in GSE136103. The gene expression data of cirrhotic and healthy groups were extracted from the merged dataset. The CellChat object was created from a data matrix, and the cell information was added into the meta slot of the object using *addMeta*, and cell types were set as the default cell identities using *setIdent*. The database CellChatDB was set as the ligand-receptor interaction database. The cell-cell communication network was constructed, visualized, and systemically analyzed in accordance with the tutorials published in GitHub (https://github.com/sqjin/CellChat). The cirrhotic-specific signaling was further identified by comparing the CellChat objects of cirrhosis and healthy controls.

Statistical analysis

Statistical analysis was performed using R (version 4.0.3, Bell Laboratories). The results are presented as mean ± SD, or median and quartiles. For the normalized data, unpaired 2-tailed *t*-test was performed to analyze the differences between two groups, and one-way ANOVA was adopted to analyze the differences among three groups, where Dunnett's multiple comparison test was adopted to compare several experimental groups with the control group, and Tukey's multiple comparison test was applied to compare every two groups for post-hoc analysis. Pearson test was used to analyze the correlation between two variables. The survival time for NLS in BA patients with different THY1 levels was compared by Kaplan–Meier analysis with a log-rank test. Wilcoxon signed-rank test was employed to compare the differences between two groups in the single-cell RNA sequencing (scRNA-seq) data. Differences were considered significant when $P < 0.05$.

Results

Tables

Table S1. The clinical information of patients for the m^{6}A RNA methylation quantification

Group	NC (n=27)	CS (n=28)	BA (n=63)
Age（Mon）（$\bar{x}$±s）	29.2±23.9****	2.2±0.9	1.9±0.7
Gender			
Male/Female	11/16	22/6	38/25
Laboratory examination（$\bar{x}$±s）			
ALT（U/L）	20.8±9.9****	190.0±177.0	155.9±106.0
AST（U/L）	50.4±22.3****	275.9±190.1	241.6±136.0
ALB（g/L）	41.0±3.7**	36.2±4.3	38.0±3.5
TBIL（umol/L）	7.7±6.8****	150.6±37.1**	178.0±46.5
DBIL（umol/L）	2.0±1.9****	108.6±30.9**	128.4±30.7
ALP（U/L）	215.8±101.4****	639.9±219.0	710.2±243.3
GGT（U/L）	62.2±180.9****	216.0±155.0***	597.1±537.1
TBA（umol/L）	9.7±7.8****	122.8±82.7	121.7±73.3
APRI	0.5±0.4****	1.6±1.1	1.8±1.3
Grades of inflammation			
G1	-	1 (3.6%)	1 (1.6%)
G2	-	14 (50.0%)	21 (33.3%)
G3	-	13 (46.4%)	41 (65.1%)
Stages of fibrosis			
S0	-	1 (3.6%)	0 (0%)
S1	-	13 (46.4%)	0 (0%)
S2	-	14 (50.0%)	21 (33.3%)
S3	-	0 (0%)	29 (46.0%)
S4	-	0 (0%)	13 (20.6%)

$^{**}P$<0.01 vs. BA, $^{***}P$<0.001 vs. BA, $^{****}P$<0.0001 vs. BA. Abbreviations: ALB, albumin; ALP alkaline phosphatase; ALT, alanine aminotransferase; APRI, aspartate aminotransferase-to-platelet ratio index; AST, aspartate aminotransferase; CS, cholestasis; DBIL, direct bilirubin; GGT, gamma-glutamyl transpeptidase; Mon, Month; NC, normal control; TBA, total bile acid; TBIL, total bilirubin.

Table S2. The clinical information of patients for the qPCR assays

Group	NC (n=24)	CS (n=24)	BA (n=24)
Age（Mon）（$\bar{x}$±s）	25.7±20.7****	2.2±0.8	1.8±0.6
Gender			
Male/Female	9/15	19/5	15/9
Laboratory examination（$\bar{x}$±s）			
ALT（U/L）	18.2±10.1**	182.8±171.7	141.8±112.2
AST（U/L）	48.0±25.0****	283.1±180.3	232.4±169.8
ALB（g/L）	38.8±9.2	36.7±5.1	37.4±3.7
TBIL（umol/L）	11.3±22.2****	148.4±40.5	169.6±46.8
DBIL（umol/L）	1.9±1.8****	108.2±31.3	122.8±26.9
ALP（U/L）	202.4±106.4****	626.2±219.1	675.0±244.6
GGT（U/L）	27.5±24.5****	218.8±160.1***	583.4±476.9
TBA（umol/L）	9.0±8.4****	127.4±87.8	108.5±41.5
APRI	0.5±0.5****	1.6±1.0	2.0±1.7
Grades of inflammation			
G1	-	1 (4.2%)	1 (4.2%)
G2	-	14 (58.3%)	9 (37.5%)
G3	-	9 (37.5%)	14 (58.3%)
Stages of fibrosis			
S0	-	1 (4.2%)	0 (0%)
S1	-	11 (45.8%)	0 (0%)
S2	-	12 (50.0%)	9 (37.5%)
S3	-	0 (0%)	10 (41.7%)
S4	-	0 (0%)	5 (20.8%)

**P<0.01 vs. BA, ***P<0.001 vs. BA, ****P<0.0001 vs. BA.

Table S3. The clinical information of patients for the MeRIP-seq and RNA-seq

Group	NC (n=4)	BA (n=4)
Age (Mon) ($\bar{x}$ ±s)	22.5±12.5	16.2±3.5
Gender		
Male/Female	2/2	2/2
Laboratory examination ($\bar{x}$ ±s)		
ALT (U/L)	25.3±13.1**	129.1±43.1
AST (U/L)	60.9±24.2*	173.1±72.3
ALB (g/L)	41.0±2.6	39.7±4.2
TBIL (umol/L)	9.7±7.7***	166.7±24.8
DBIL (umol/L)	5.3±5.0***	139.3±16.6
ALP (U/L)	239.2±81.6**	645.5±145.8
GGT (U/L)	108.4±91.0*	916.6±356.9
TBA (umol/L)	29.8±38.9*	88.2±9.6
APRI	0.4±0.2	0.9±0.6
Grades of inflammation		
G2	-	2 (50.0%)
G3	-	2 (50.0%)
Stages of fibrosis		
S2	-	0 (0%)
S3	-	1 (25.0%)
S4	-	3 (75.0%)

*P<0.05 vs. BA, **P<0.01 vs. BA, ***P<0.001 vs. BA, ****P<0.0001 vs. BA.

Table S4. The clinical information of patients for the IHC staining of liver tissue microarray

Group	NC (n=30)	CS (n=34)	BA (n=93)
Age (Mon) ($\bar{x}$ ±s)	26.9±28.1****	2.1±0.7	1.9±0.7
Gender			
Male/Female	15/15	24/10	48/45
Laboratory examination ($\bar{x}$ ±s)			
ALT (U/L)	26.3±23.8***	244.3±243.1**	147.3±96.4
AST (U/L)	51.0±22.3****	319.6±264.4**	220.4±143.6
ALB (g/L)	41.3±4.0****	36.6±4.7	37.6±3.6
TBIL (umol/L)	12.0±20.9****	158.5±42.3	169.4±49.1
DBIL (umol/L)	2.6±2.5****	117.3±34.6	131.2±36.5
ALP (U/L)	212.2±94.5****	681.4±226.0	666.0±255.4
GGT (U/L)	63.1±176.9****	206.3±182.5***	514.3±412.7
TBA (umol/L)	11.5±9.4****	129.9±90.9	121.2±64.0
APRI	0.5±0.4**	2.0±2.1	1.7±1.6
Grades of inflammation			
G1		0 (0%)	3 (3.2%)
G2	-	21 (61.8%)	31 (33.3%)
G3	-	13 (38.2%)	59 (63.4%)
Stages of fibrosis			
S1		12 (35.3%)	9 (9.7%)
S2	-	10 (29.4%)	50 (53.8%)
S3	-	9 (26.5%)	24 (25.8%)
S4	-	3 (8.8%)	10 (10.8%)

**P<0.01 vs. BA, ***P<0.001 vs. BA, ****P<0.0001 vs. BA.

Table S5. Primer sequences for qRT-PCR

Primers	Base sequence
β-actin-f	CCTGGCACCCAGCACAAT
β-actin-r	GGGCCGGACTCGTCATAC
METTL3-f	AGATGGGGTAGAAAGCCTCCT
METTL3-r	TGGTCAGCATAGGTTACAAGAGT
METTL14-f	CTGAAAGTGCCGACAGCATTGG
METTL14-r	CTCTCCTTCATCCAGATACTTACG
WTAP-f	GCAACAACAGCAGGAGTCTGCA
WTAP-r	CTGCTGGACTTGCTTGAGGTAC
FTO-f	CCAGAACCTGAGGAGAGAATGG
FTO-r	CGATGTCTGTGAGGTCAAACGG
ALKBH5-f	CCAGCTATGCTTCAGATCGCCT
ALKBH5-r	GGTTCTCTTCCTTGTCCATCTCC
THY1-f	ATCGCTCTCCTGCTAACAGTC
THY1-r	CTCGTACTGGATGGGTGAACT

Table S6. The statistics of m^{6}A peak calling in BA and controls

Sample	Fragment length	Count of peak	FRiP	Count of summits
NC31	126	14563	59.67%	19900
NC34	126	10443	53.89%	13759
NC35	133	14442	63.17%	19378
NC36	133	10845	65.75%	14127
BA1346	125	15398	56.49%	21505
BA1359	134	15863	61.05%	21869
BA1427	135	15578	62.05%	21361
BA1434	133	16329	62.65%	22649

Table S7. Primer sequences for MeRIP-qPCR

Primers	Base sequence
HPRT1-f	ACCAGTCAACAGGGGACATAA
HPRT1-r	CTTCGTGGGGTCCTTTTCACC
THY1-f	ATCGCTCTCCTGCTAACAGTC
THY1-r	CTCGTACTGGATGGGTGAACT
COL1A1-f	GATTCCCTGGACCTAAAGGTGC
COL1A1-r	AGCCTCTCCATCTTTGCCAGCA

Table S8. The cell composition of single-cell RNA sequencing data in GSE136103

Cell type	CR1	CR2	CR3	CR4	CR5	HE1	HE2	HE3	HE4	HE5
T	1031	1408	596	2462	1395	742	4557	1723	2755	2427
NK	921	1530	385	638	315	287	3188	1346	2746	1112
Monocyte	360	437	571	400	262	625	610	254	562	636
Periportal LSEC	958	1314	1968	52	44	272	68	1114	351	4
Cholangiocyte	1403	605	262	17	0	164	825	69	43	2
M1 macrophage	159	477	421	325	184	183	513	379	388	505
M2 macrophage	86	284	250	189	131	223	251	402	698	238
Central venous LSEC	126	169	349	8	6	186	72	864	197	3
Naive B	90	265	84	450	316	31	200	259	217	31
HSC	206	86	109	6	5	190	597	643	20	0
NKT	71	133	71	82	33	15	212	68	142	84
Plasma	57	48	117	100	156	87	25	196	84	15
Myofibroblast	95	34	212	4	5	28	85	39	1	5
Hepatocyte	152	84	46	5	1	15	22	20	1	8
pDC	16	27	19	69	61	24	26	46	16	31

Abbreviations: CR: cirrhotic patients; HE: healthy control; LSEC: liver sinusoidal endothelial cell; HSC, hepatic stellate cell

Figures

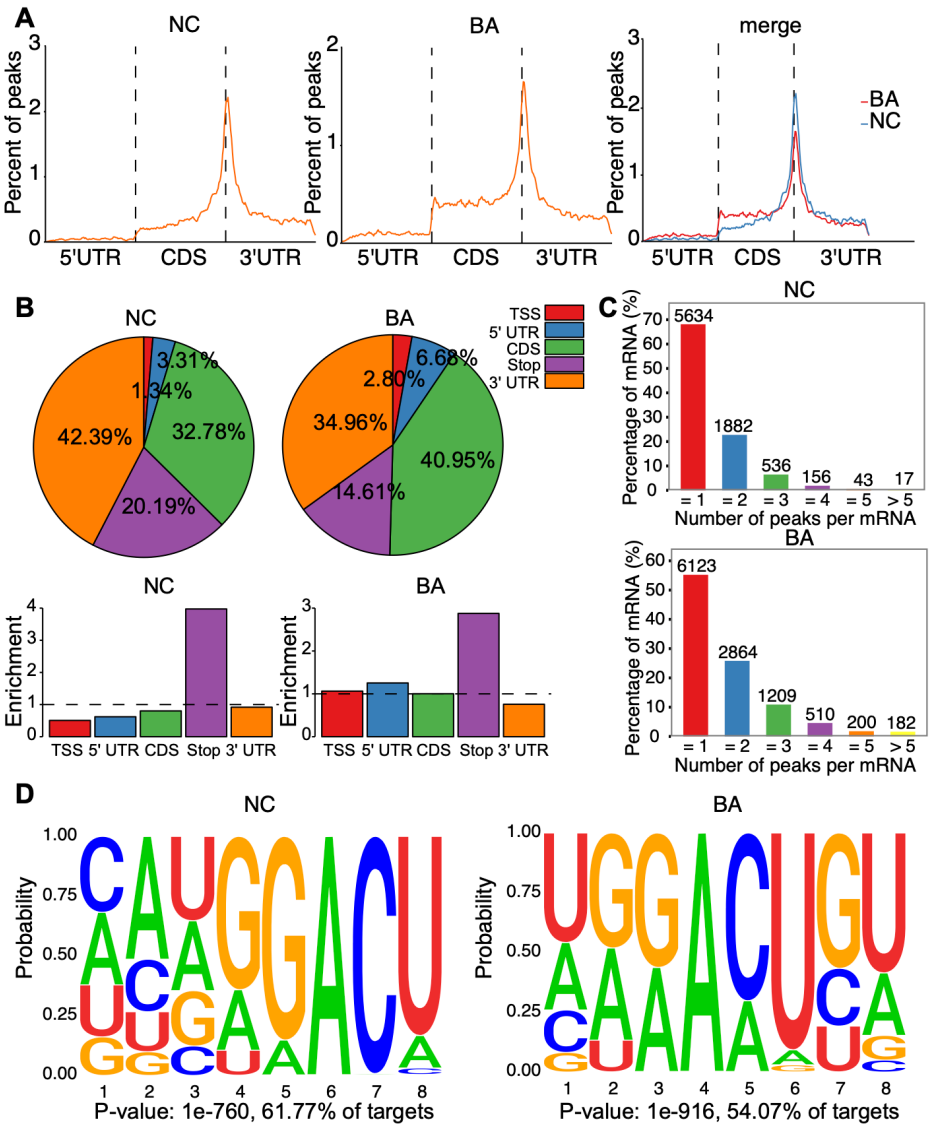

FIGURE S1 The m^{6}A topological patterns in biliary atresia and normal controls. (A) The distribution of m^{6}A peaks on the 3 genic regions (UTR, CDS, and 3' UTR) in BA and NC. (B) The distribution and enrichment fold of m^{6}A peaks on the 5 RNA transcript regions (5' UTR, TSS, CDS, stop, and 3' UTR) in BA and NC. (C) The distribution of m^{6}A peak number on the mRNAs in BA and NC. (D) The most conserved motif (RRACH) in BA and NC. Abbreviations: BA: biliary atresia; NC: normal control.

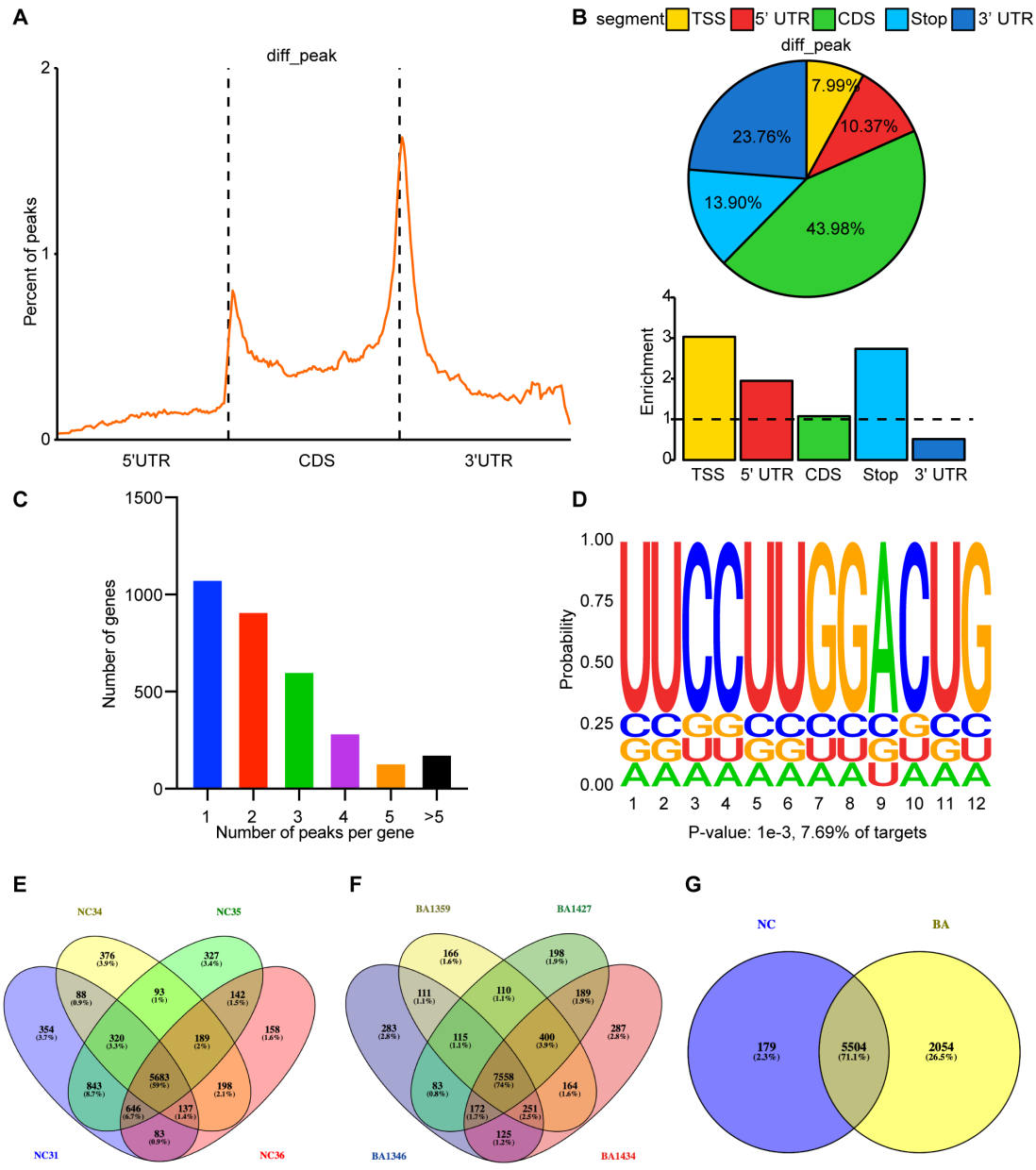

FIGURE S2 The m^{6}A topological patterns of the differential peaks in BA and normal controls. (A) The distribution of the differential m^{6}A peaks on the 3 genic regions (5' UTR, CDS, and 3' UTR). (B) The distribution and enrichment fold of the differential m^{6}A peaks on the 5 RNA transcript regions (5' UTR, TSS, CDS, stop, 3' UTR). (C) The distribution of the differential m^{6}A peak number on genes. (D) The conserved motif (RRACH) of the differential m^{6}A peaks. (E) Venn diagram of m^{6}A peaks in each control sample. (F) Venn diagram of m^{6}A peaks in each BA sample. (G) Venn

diagram of group-conserved m^6A peaks. Abbreviations: BA: biliary atresia; NC: normal control.

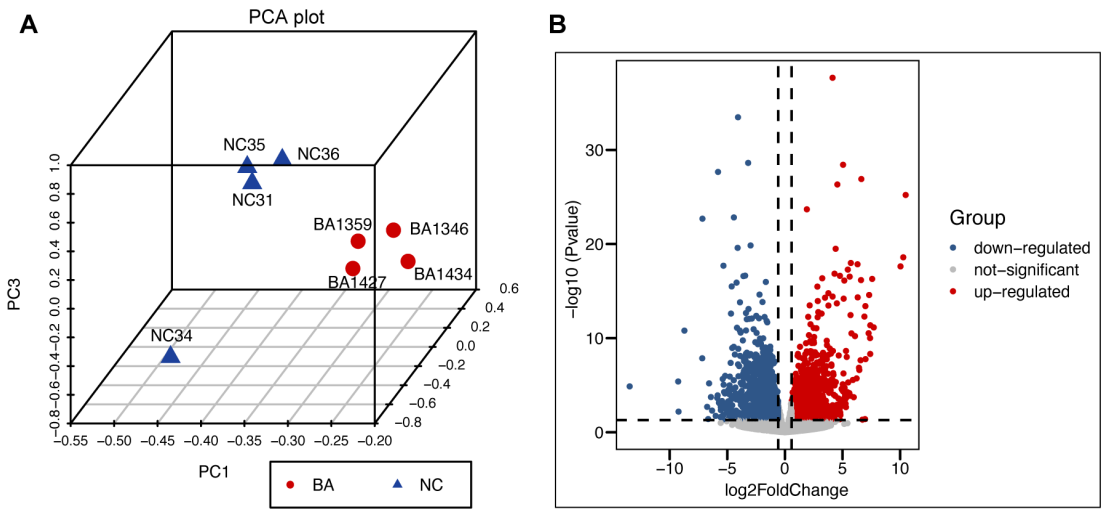

FIGURE S3 PCA and volcano plots of the gene expression levels in BA and controls. (A) PCA plots. (B) Volcano plots. Abbreviations: BA: biliary atresia; NC: normal control.

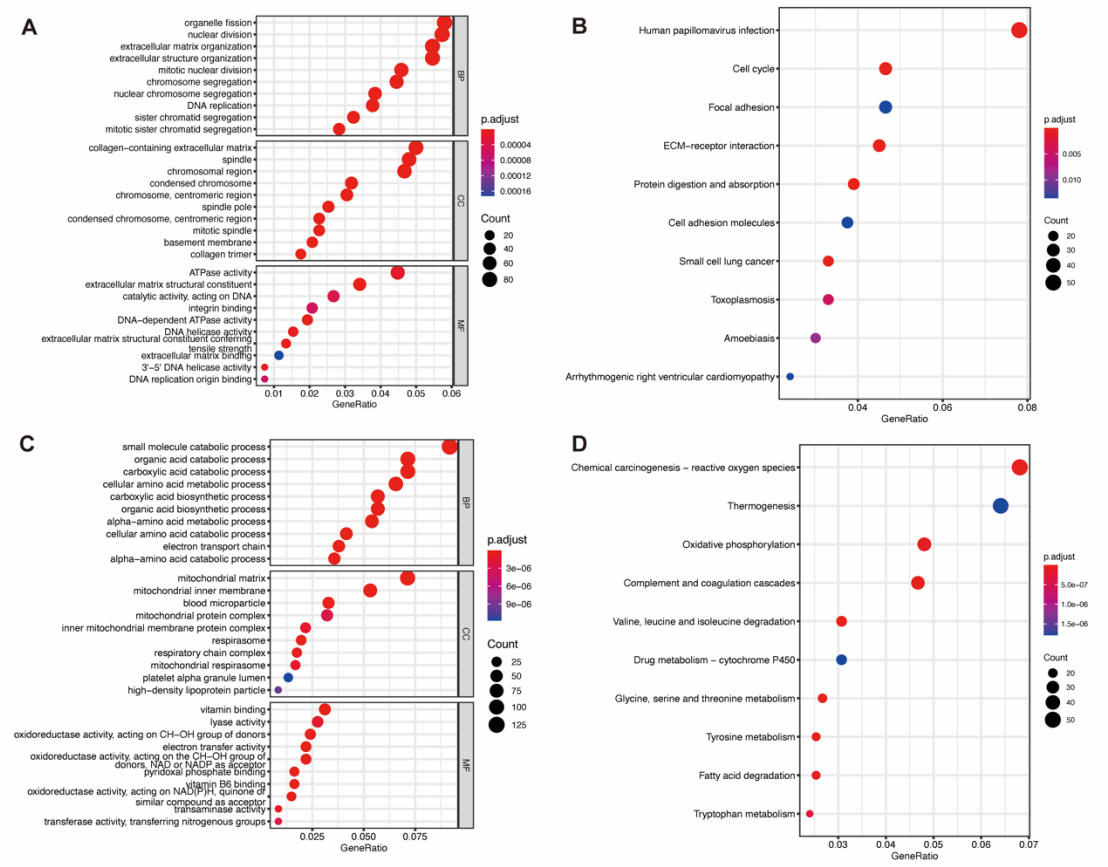

FIGURE S4 Gene functional enrichment analysis of the RNA-seq data from the BA and control groups. (A) GO analyses of the upregulated genes in biliary atresia (BA). (B) KEGG analyses of the upregulated genes in BA. (C) GO analyses of the downregulated genes in BA. (D) KEGG analyses of the downregulated genes in BA.

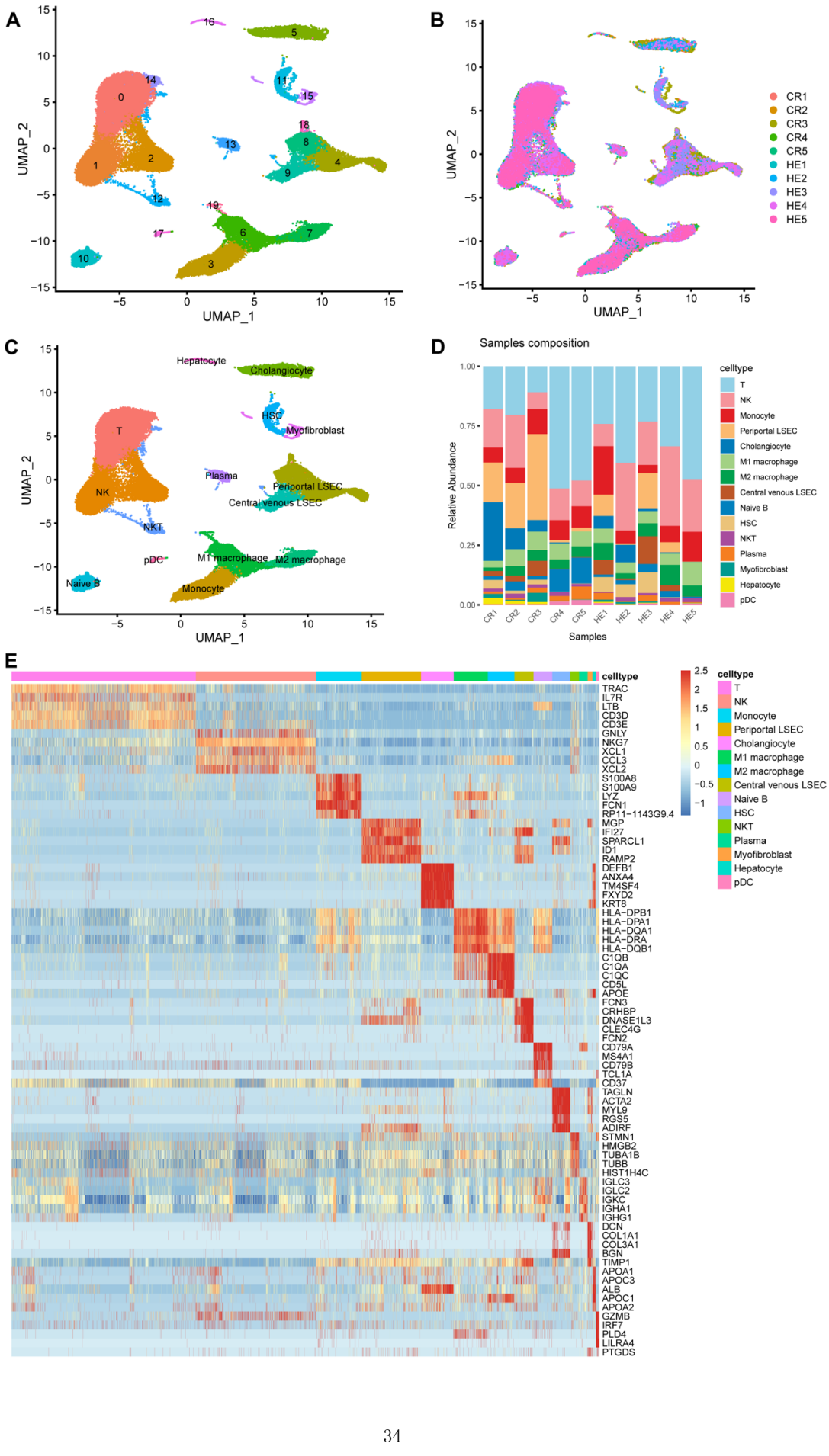
A
B
C
D
E
UMAP_1
UMAP_2
CR1
CR2
CR3
CR4
CR5
HE1
HE2
HE3
HE4
HE5
Hepatocyte
Cholangiocyte
HSC
Myofibroblast
T
NK
Plasma
Periportal LSEC
Central venous LSEC
NKT
pDC
M1 macrophage
M2 macrophage
Naive B
Monocyte
Samples composition
Relative Abundance
Samples
celltype

FIGURE S5 The cell atlas of the single-cell RNA sequencing dataset of GSE136103. (A) UMAP plot of 20 clusters in all cells. (B) UMAP plot of sample source in all cells. (C) UMAP plot of cell types in all cells. (D) The cell composition in each sample. (E) Heatmap of the marker genes in each cell type.

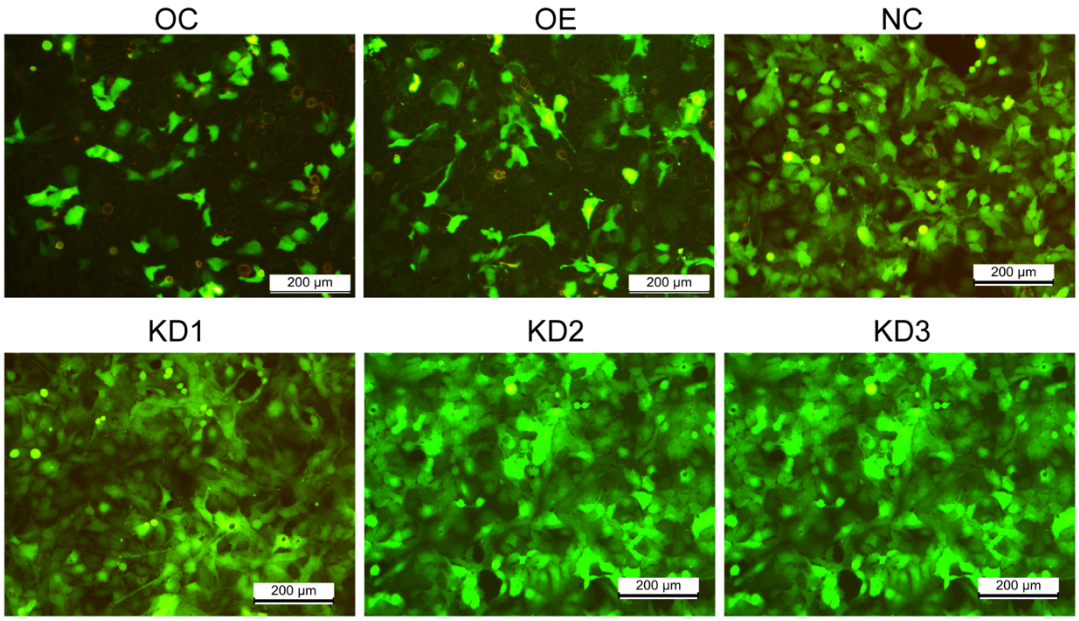

FIGURE S6 The GFP fluorescence of LX-2 cells transfected with METTL3 lentivirus. Abbreviations: OC, overexpression control; OE, overexpression; NC, negative control; KD, knocking down.

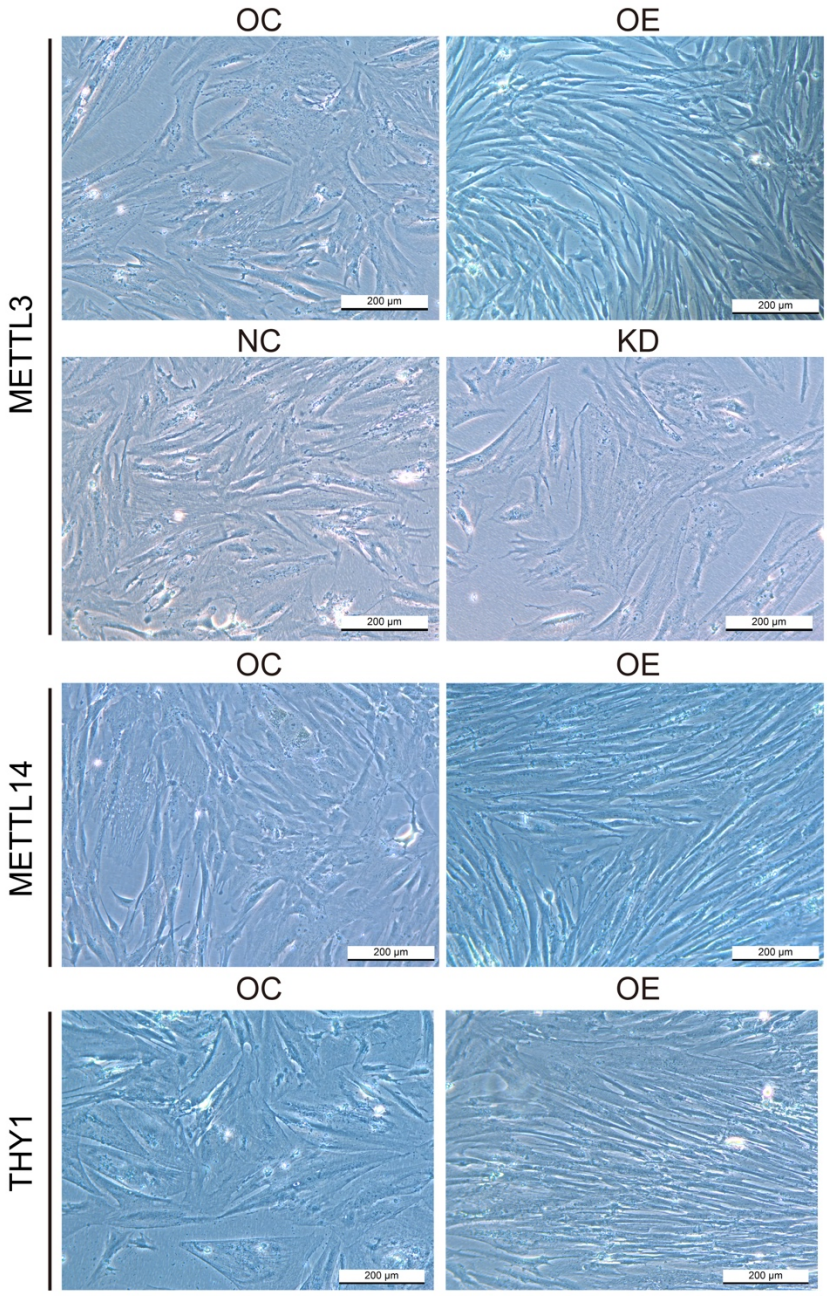

FIGURE S7 The morphology of primary hepatic stellate cells with gene interference under the light microscope. Abbreviations: OC, overexpression control; OE, overexpression; NC, negative control; KD, knocking down.

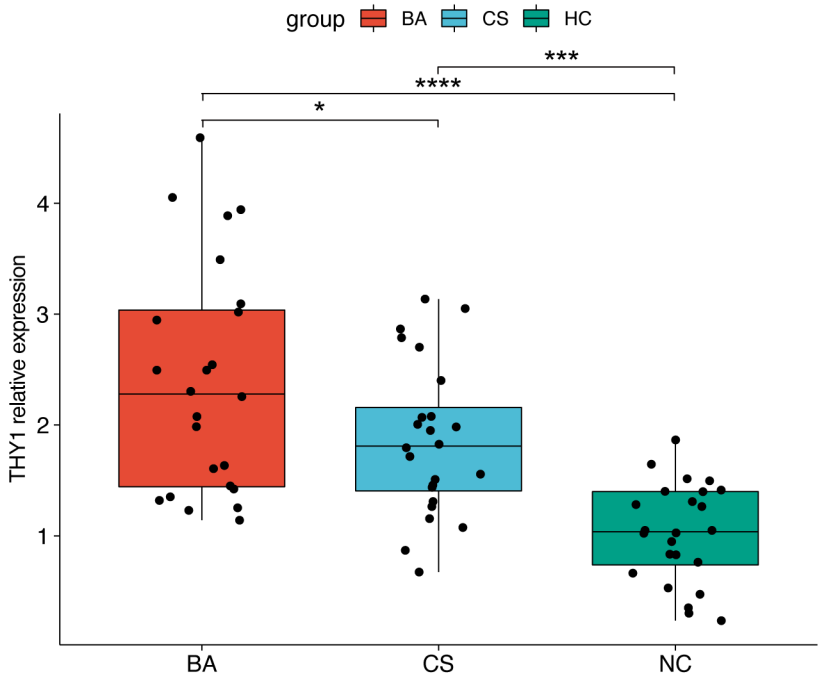

FIGURE S8 The expression levels of THY1 in BA (24 cases) and controls (NC: 24 cases, CS: 24 cases). Abbreviations: BA, biliary atresia; NC, normal control; CS, cholestasis.

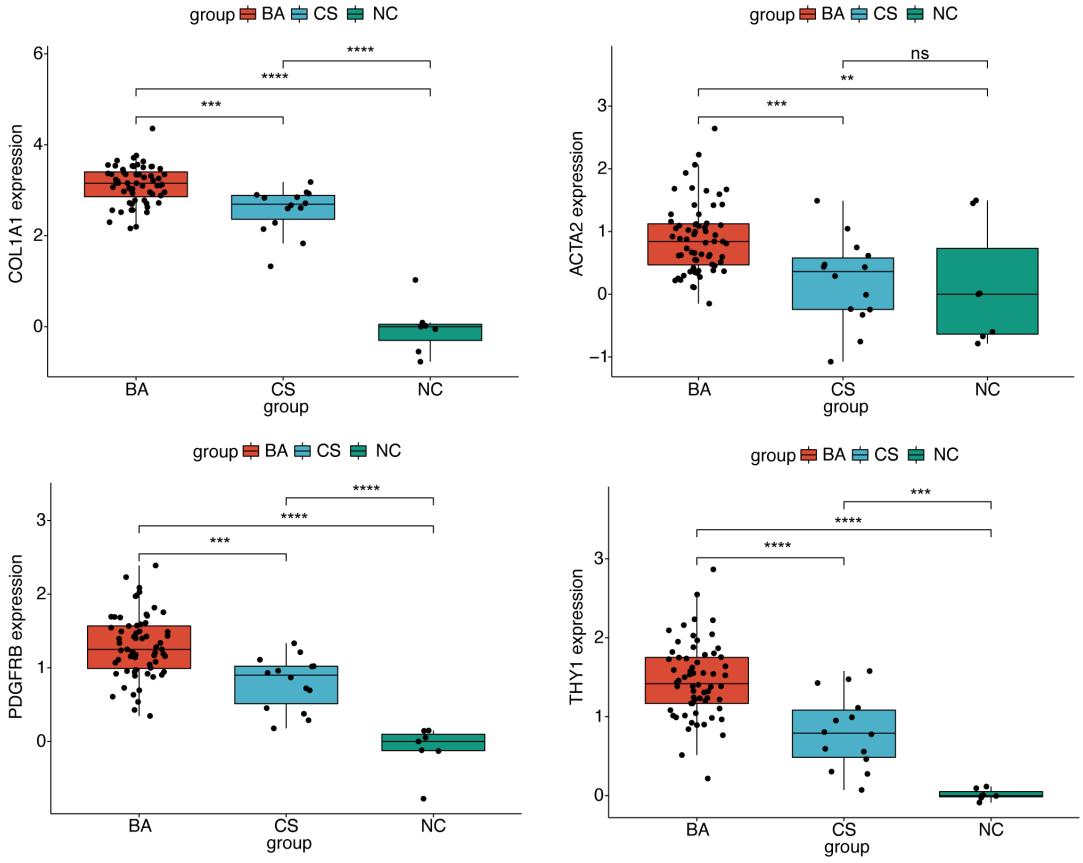

FIGURE S9 The expression levels of fibrotic markers and THY1 in the dataset of GSE46960. Abbreviations: BA, biliary atresia; NC, normal control; CS, cholestasis.

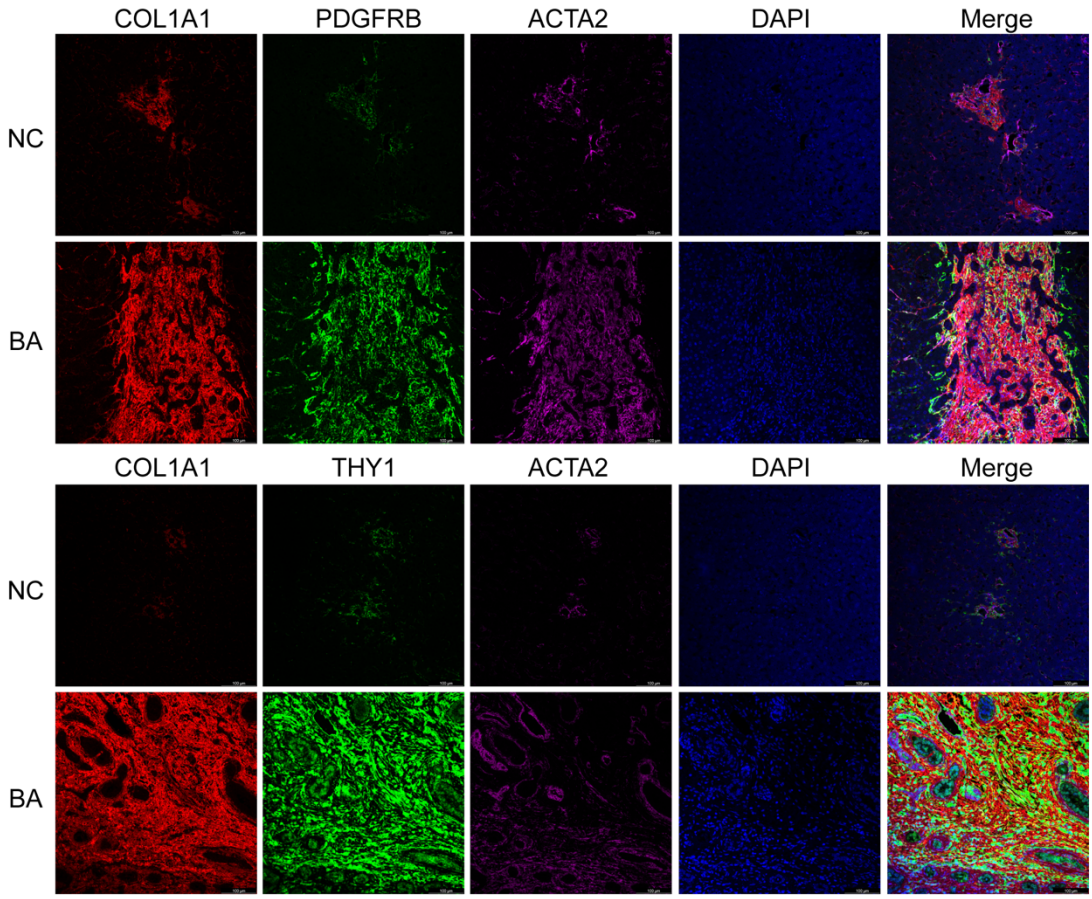

FIGURE S10 Immunofluorescence staining of THY1 and fibrotic markers in the livers of BA and NC. Abbreviations: BA, biliary atresia; NC, normal control.

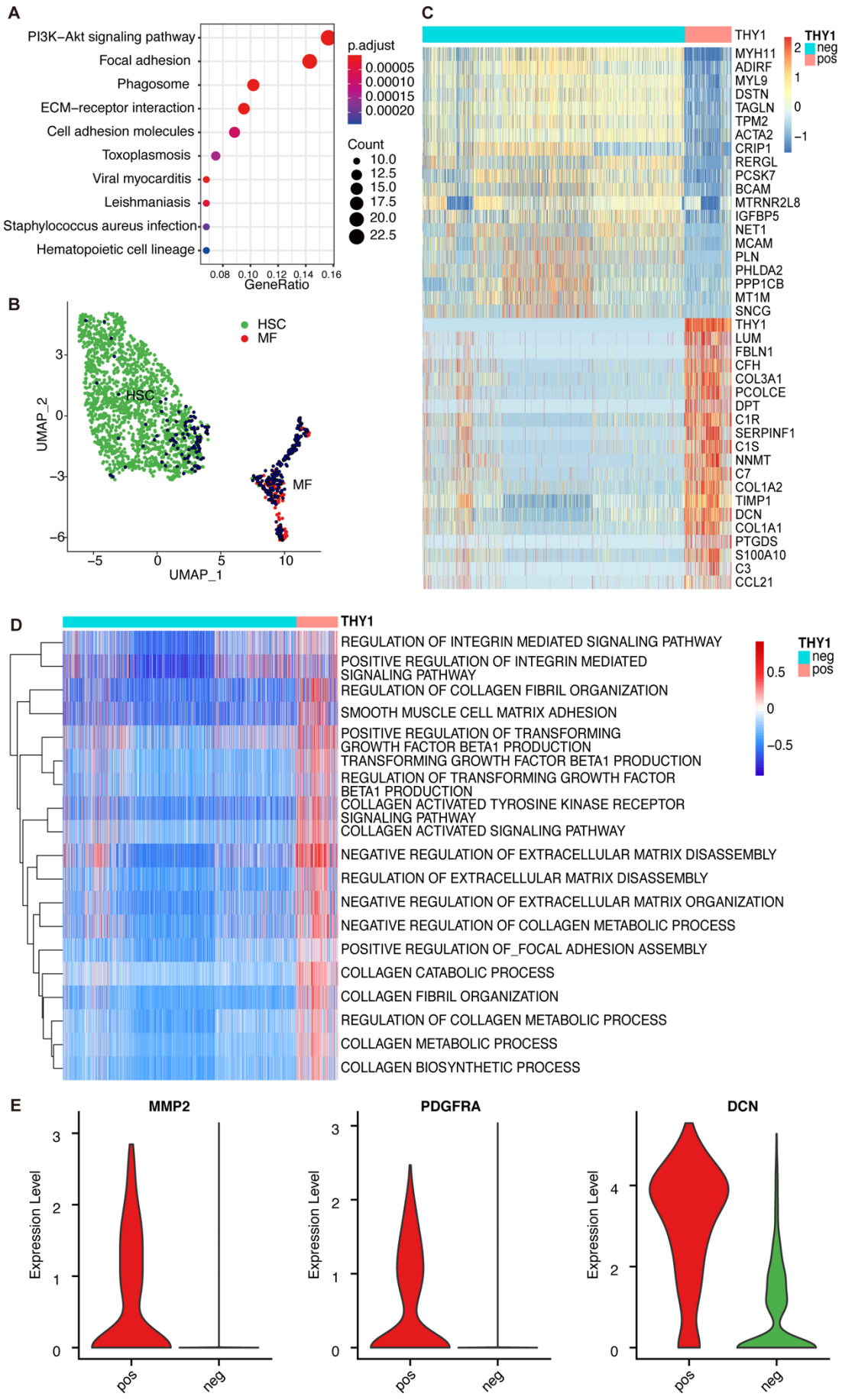
A
PI3K−Akt signaling pathway
Focal adhesion
Phagosome
ECM−receptor interaction
Cell adhesion molecules
Toxoplasmosis
Viral myocarditis
Leishmaniasis
Staphylococcus aureus infection
Hematopoietic cell lineage
0.08 0.10 0.12 0.14 0.16
GeneRatio
p.adjust
0.00005
0.00010
0.00015
0.00020
Count
10.0
12.5
15.0
17.5
20.0
22.5
B
HSC
MF
UMAP_2
UMAP_1
C
THY1
MYH11
ADIRF
MYL9
DSTN
TAGLN
TPM2
ACTA2
CRIP1
RERGL
PCSK7
BCAM
MTRNR2L8
IGFBP5
NET1
MCAM
PLN
PHLDA2
PPP1CB
MT1M
SNCG
THY1
LUM
FBLN1
CFH
COL3A1
PCOLCE
DPT
C1R
SERPINF1
C1S
NNMT
C7
COL1A2
TIMP1
DCN
COL1A1
PTGDS
S100A10
C3
CCL21
neg
pos
D
THY1
REGULATION OF INTEGRIN MEDIATED SIGNALING PATHWAY
POSITIVE REGULATION OF INTEGRIN MEDIATED SIGNALING PATHWAY
REGULATION OF COLLAGEN FIBRIL ORGANIZATION
SMOOTH MUSCLE CELL MATRIX ADHESION
POSITIVE REGULATION OF TRANSFORMING GROWTH FACTOR BETA1 PRODUCTION
TRANSFORMING GROWTH FACTOR BETA1 PRODUCTION
REGULATION OF TRANSFORMING GROWTH FACTOR BETA1 PRODUCTION
COLLAGEN ACTIVATED TYROSINE KINASE RECEPTOR SIGNALING PATHWAY
COLLAGEN ACTIVATED SIGNALING PATHWAY
NEGATIVE REGULATION OF EXTRACELLULAR MATRIX DISASSEMBLY
REGULATION OF EXTRACELLULAR MATRIX DISASSEMBLY
NEGATIVE REGULATION OF EXTRACELLULAR MATRIX ORGANIZATION
NEGATIVE REGULATION OF COLLAGEN METABOLIC PROCESS
POSITIVE REGULATION OF_FOCAL ADHESION ASSEMBLY
COLLAGEN CATABOLIC PROCESS
COLLAGEN FIBRIL ORGANIZATION
REGULATION OF COLLAGEN METABOLIC PROCESS
COLLAGEN METABOLIC PROCESS
COLLAGEN BIOSYNTHETIC PROCESS
THY1
neg
pos
E
MMP2
PDGFRA
DCN
Expression Level
pos
neg

FIGURE S11 THY1 is associated with extracellular matrix organization in the liver fibrosis. (A) KEGG analysis of up-regulated genes in patients with higher THY1 expressions in GSE46960. (B) The distribution of THY1 positive cells (count > 0) in hepatic stellate cells (HSCs) and myofibroblasts (MFs) of GSE136103. (C) Heatmap visualization of differentially expressed genes between THY1 positive (pos) and negative (neg) cells in HSCs and MFs of GSE136103. (D) GSVA analysis of differential pathways between THY1 positive and negative cells in HSCs and MFs of GSE136103. (E) The expressions of MMP2, PDGFRA, DCN in the THY1 positive and negative cells in HSCs and MFs of GSE136103.

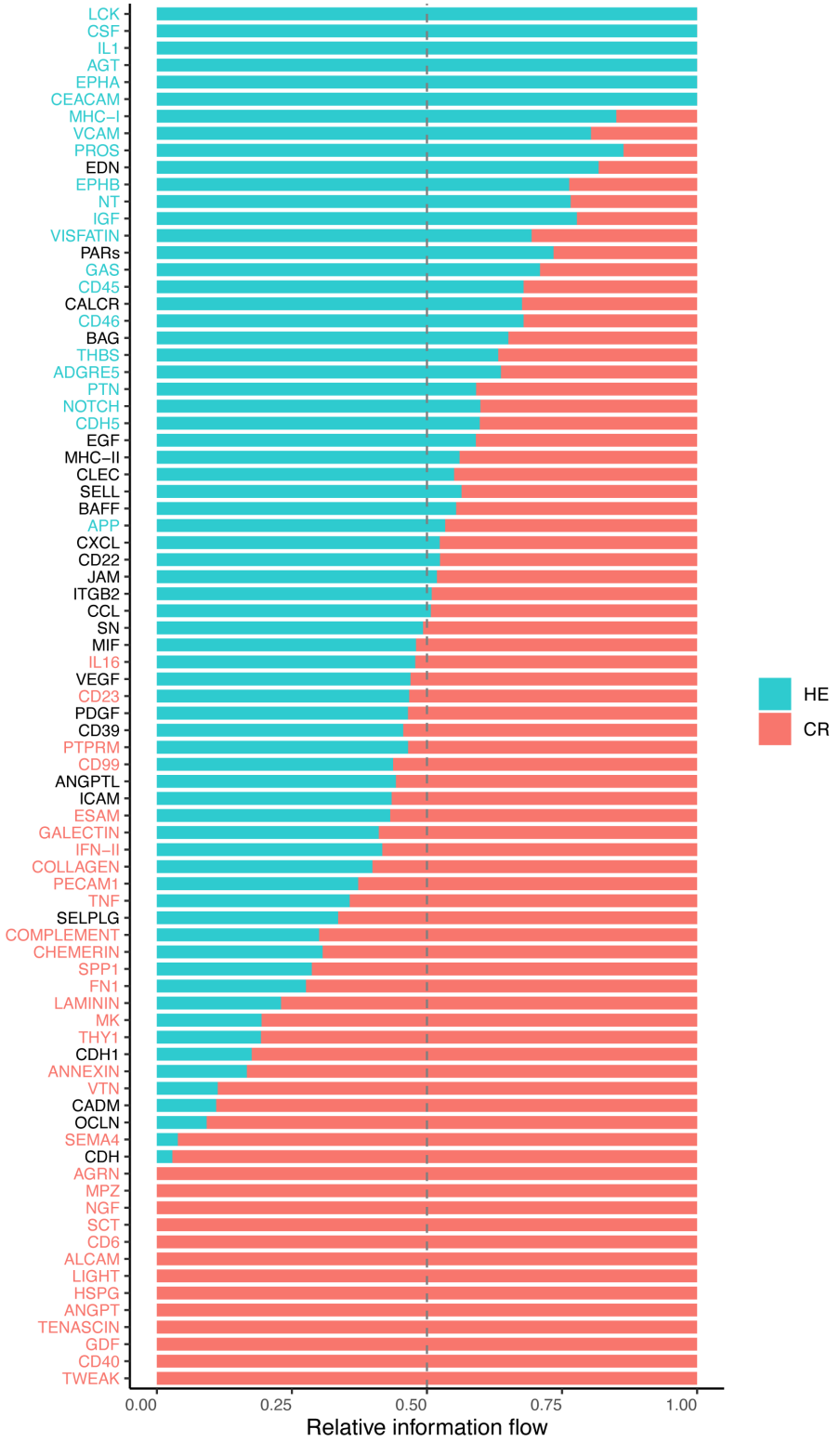

FIGURE S12 CellChat analysis showing the relative information flow of all signaling pathways between healthy controls and cirrhotic patients in the dataset of GSE136103. Abbreviations: CR, cirrhosis; HE, healthy controls.

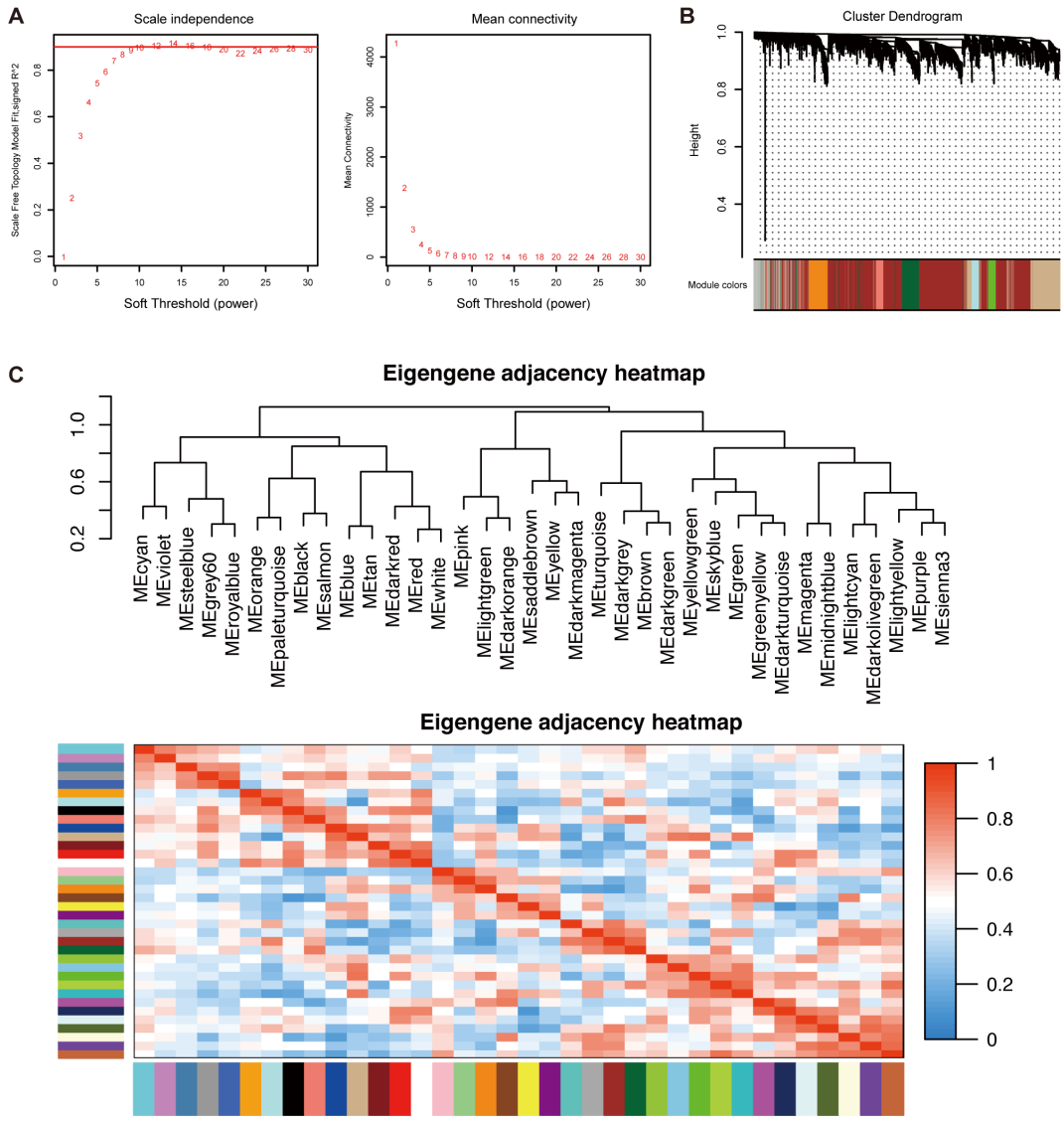

FIGURE S13 WGCNA analysis for the identification of LT/death-related hub genes in the dataset of GSE15235. (A) The R^2 of scale free topology model with different soft thresholds. (B) The mean connectivity with different soft thresholds. (C) The correlation heatmap among 37 modules from 20161 genes in 45 samples.

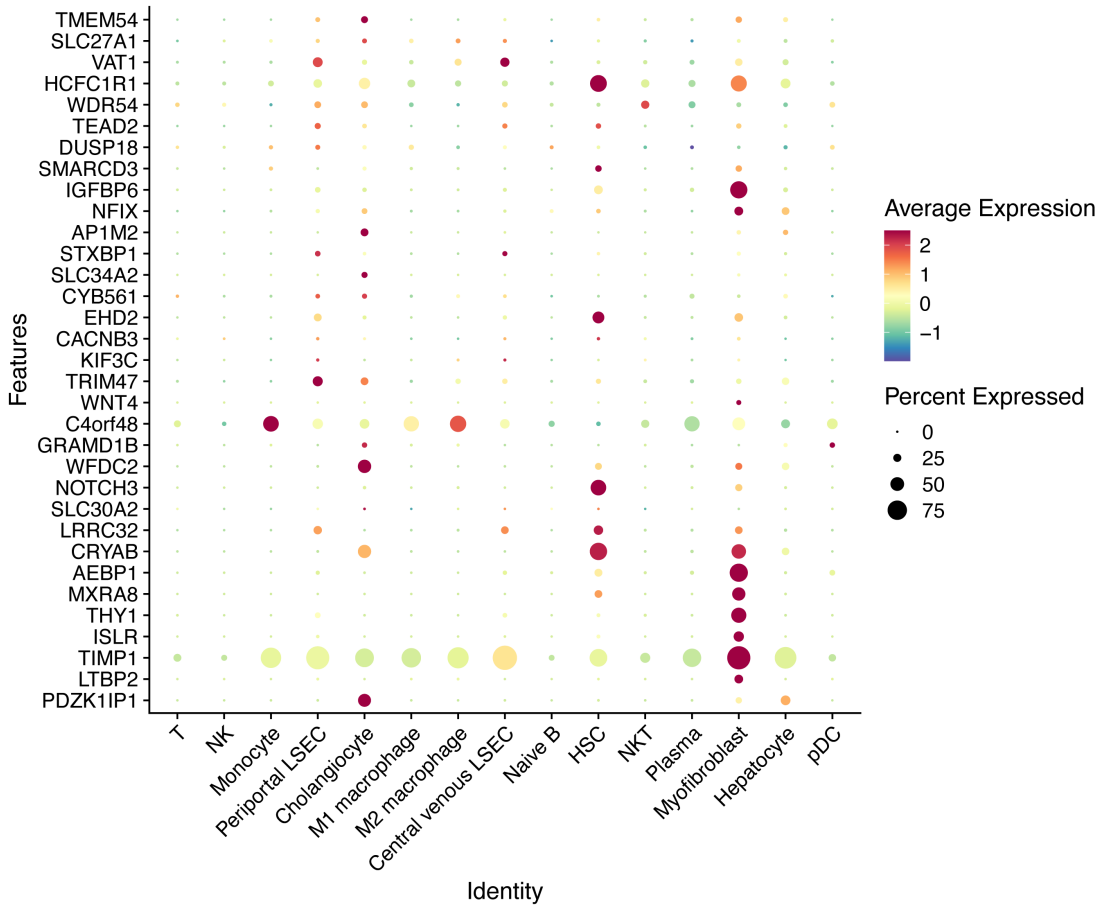

FIGURE S14 Dotplot for the expressions of LT/death-related hub genes in the dataset of GSE136103. Abbreviations: LT: liver transplantation.

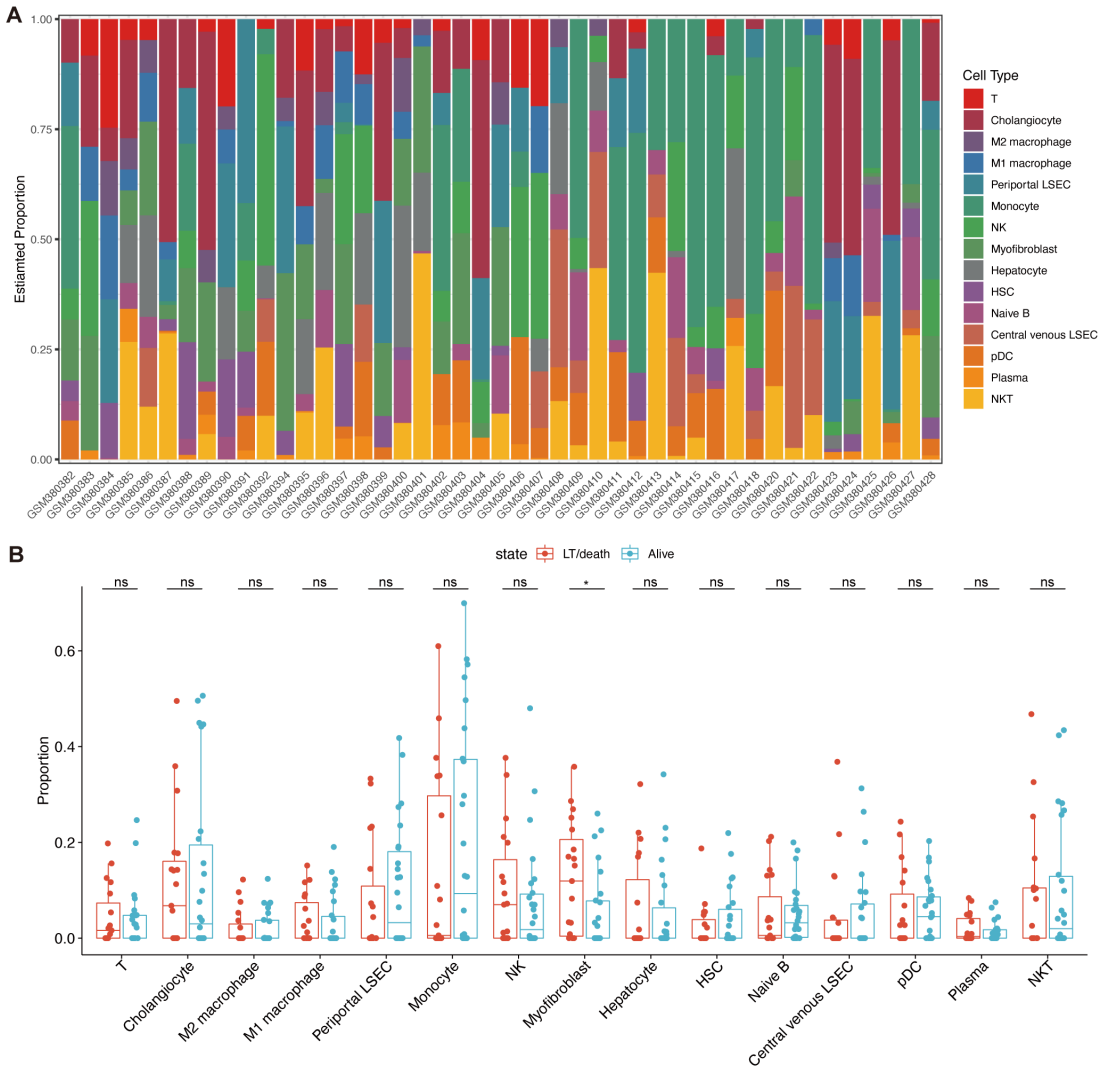

FIGURE S15 Deconvolution analysis demonstrated the estimated cell proportions in each sample of GSE15235 using the single cell reference data GSE136103. (A) The bar plot of the estimated cell proportions in each sample. (B) Comparison of cell proportions in BA patients with native liver survival or LT/death. Abbreviations: BA: biliary atresia; LT: liver transplantation.

References

1. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-1760.
2. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16:284-287.
3. Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, et al. N6-methyladenosine-dependent regulation of messenger RNA stability. Nature. 2014;505:117-120.
4. Mederacke I, Dapito DH, Affo S, Uchinami H, Schwabe RF. High-yield and high-purity isolation of hepatic stellate cells from normal and fibrotic mouse livers. Nat Protoc. 2015;10:305-315.
5. Bessho K, Mourya R, Shivakumar P, Walters S, Magee JC, Rao M, et al. Gene expression signature for biliary atresia and a role for interleukin-8 in pathogenesis of experimental disease. Hepatology. 2014;60:211-223.
6. Moyer K, Kaimal V, Pacheco C, Mourya R, Xu H, Shivakumar P, et al. Staging of biliary atresia at diagnosis by molecular profiling of the liver. Genome Med. 2010;2:33.
7. Ramachandran P, Dobie R, Wilson-Kanamori JR, Dora EF, Henderson BEP, Luu NT, et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. Nature. 2019;575:512-518.
8. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184:3573-3587.e3529.
9. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16:1289-1296.
10. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
11. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37:773-782.
12. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12:1088.