Supporting document

MATRISOME GENE-BASED SUBCLASSIFICATION OF PATIENTS WITH LIVER FIBROSIS IDENTIFIES CLINICAL AND MOLECULAR HETEROGENEITIES

Wei Chen[1, 2, 5, #], Yameng Sun[3, 4, #], Shuyan Chen[3, 4], Xiaodong Ge[5], Wen Zhang[3, 4], Ning Zhang[3, 4], Xiaoning Wu[3, 4], Zhuolun Song[5], Hui Han[5], Romain Desert[5], Xuzhen Yan[1, 2], Aiting Yang[1, 2], Sukanta Das[5], Dipti Athavale[5], Natalia Nieto[5, 6, *], Hong You[3, 4, *]

[1] Beijing Clinical Research Institute, No. 95 Yong'an Road, Xicheng District, Beijing 100050, China

[2] Experimental and Translational Research Center, Beijing Friendship Hospital, Capital Medical University, No. 95 Yong'an Road, Xicheng District, Beijing 100050, China

[3] Liver Research Center, Beijing Friendship Hospital, Capital Medical University, No. 95 Yong'an Road, Xicheng District, Beijing 100050, China

[4] Beijing Key Laboratory of Translational Medicine in Liver Cirrhosis, National Clinical Research Center of Digestive Diseases, No. 95 Yong'an Road, Xicheng District, Beijing 100050, China

[5] Department of Pathology, University of Illinois at Chicago, 840 S. Wood St., Suite 130 CSN, MC 847, Chicago, IL 60612, USA

[6] Department of Medicine, Division of Gastroenterology and Hepatology, University of Illinois at Chicago, 840 S. Wood St., Suite 1020N, MC 787, Chicago, IL 60612, USA

[#] These authors contributed equally.

***Corresponding Authors:**

Hong You, Liver Research Center, Beijing Friendship Hospital, Capital Medical University, No. 95 Yong'an Road, Xicheng District, Beijing 100050, China. E-mail: youhong30@sina.com. Phone: +86 (010) 6313-9019

Natalia Nieto, Department of Pathology, University of Illinois at Chicago, 840 S. Wood St., Suite 130 CSN, MC 847, Chicago, IL 60612, USA. E-mail: nnieto@uic.edu. Phone: +1 (312) 996-7316

Financial support statement: This work was supported by the National Natural Science Foundation of China (82170613 [to W.C.], 81970524 [to H.Y.], 82130018 [to H.Y.], 81800534 [to W.C.]), National Science and Technology Major Project (2018ZX10302204) (to H.Y.), Beijing Talent Fund (2018000021469G202) (to W.C.), and US National Institute of Diabetes and Digestive and Kidney Diseases (R01 DK111677) (to N. N.).

Author's contributions: Study design: H.Y., N.N., W.C.; project supervision: H.Y., N.N.; liver biopsy assessment and clinical data collection: Y.S., S.C., X.W., H.Y.; database search: W.C., R.D.; data analysis: W.C.; transcriptome sequencing: W.C., Y.S., S.C.; single-cell RNA sequencing: X.G., H.H., Z.S., R.D. S.D., D.A., N.N.; animal experiments: W.C., W.Z., N.Z., X.G., X.Y., A.Y.; manuscript writing: W.C.; critical revision of manuscript: N.N., H.Y., Y.S.; manuscript editing: all authors.

SUPPLEMENTARY MATERIAL AND METHODS

Public transcriptomic datasets of liver fibrosis

Publicly available gene expression profiles of human liver fibrosis were retrieved from Gene Expression Omnibus (GEO) using the search terms “fibrosis”, “cirrhosis”, “fibrotic” or “cirrhotic”. After manual retrieval, a total of 9 transcriptomic profiles comprising 892 subjects were used. These transcriptomic datasets were generated from frozen liver specimens or formalin-fixed and paraffin-embedded (FFPE) liver samples, with different etiologies (HBV, HCV, ALD, NASH). Detailed information of the datasets can be found in **Table S1**.

BJFSH cohort

The BJFSH cohort contains 54 FFPE liver biopsies from 28 patients retrospectively obtained from our prospective HBV-related fibrosis/cirrhosis cohort studies (NCT01938781, NCT01938820). The patients were not coinfected with HCV or human immunodeficiency virus, or diagnosed with any other chronic liver disease or severe systematic disorder. All patients were treatment-naïve (0W) when enrolled and then received continual antiviral therapy for 78 and/or 260 weeks (78W and/or 260W). Detailed information of antiviral therapy and timepoint of liver biopsy for each patient is included in **Table S2**. A total of 10 patients in the BJFSH cohort were biopsied at baseline and at 78 and 260 weeks of antiviral treatment. Demographic characteristics and clinical parameters of patients at baseline and at 78 and 260 weeks of antiviral treatment are in **Table S3**.

Liver fibrosis and resolution mouse models

C57BL/6J mice (male, 8 weeks) were intraperitoneally injected with 12.5% of carbon tetrachloride (CCl_4) (Innochem, Beijing, China) in mineral oil (1/7, v/v), at a dose of 0.01ml/g twice a week for 6 weeks (defined as ‘P6’); control mice received equal volume of mineral oil for 6 weeks (defined as ‘P0’). P6 mice subsequently underwent spontaneous recovery for 1, 3 or 6 weeks (defined as ‘R1’, ‘R3’, ‘R6’) after cessation of CCl_4 intoxication. All mice (n=5/group) were housed and bred in a pathogen-free

laboratory animal facility with appropriate temperature (23±2°C), 12-hour light-dark cycle, and standard chow and water *ad libitum*.

Histological evaluation

FFPE liver biopsies were sectioned into 5 μm slices and stained with reticulin for standard histological assessment by two pathologists, who were blinded to the experimental conditions, time-point of the biopsy, or other clinical details. Discordant cases were reviewed again to achieve consensus. Liver fibrosis was assessed as regression when meeting one of the following criteria: 1) Ishak score decreased ⩾1 after treatment; 2) Ishak score decreasing＝0 but post-treatment liver biopsy exhibited predominantly regressive changes according to the progressive, indeterminate and predominately regressive (PIR) scoring system (1). Mouse liver sections (4 μm) were stained with picrosirius red and 10 images with a 10x objective were randomly acquired from four liver lobes except the caudate lobe. Collagen proportionate area (CPA) was measured and averaged using Image-Pro Plus software (Version 6.0, Media Cybernetics, Rockville, MD).

Immunohistochemistry

FFPE slides from human liver biopsies were dewaxed, rehydrated, antigen retrieved, blocked with goat serum, and incubated overnight at 4°C with a primary antibody against ACTA2 (dilution 1:250, ab5694, Abcam, Cambridge, UK), followed by incubation for 1 hour at room temperature with secondary antibody (PV-6001, ZSGB-BIO, Beijing, China). Pictures of immunostained ACTA2 were captured with a 3Dhistech Panoramic Scanner (3Dhistech, Budapest, Hungary).

Measurement of liver stiffness and serum biochemical parameters

Liver stiffness was measured by transient elastography (Fibroscan, Echosens, France), which was considered reliable when the interquartile range-to-liver stiffness ratio was ⩽30% in at least 10 valid measurements, and a success rate ⩾60% was observed. Serum biochemical parameters including white blood cell (WBC), platelet (PLT),

alanine transaminase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), glutamyl transpeptidase (GGT), albumin (ALB), total bilirubin (TBIL), and alpha fetoprotein (AFP) were measured according to standard protocols. Roche COBAS RTaqMan HBV test (Roche, Indianapolis, IN) based on real-time Taqman polymerase chain reaction assay (lower limit of quantification = 20 IU/mL), was used to measure serum HBV DNA levels.

Bulk RNA-seq of human FFPE liver tissues and data analysis

Bulk RNA-seq of human FFPE liver tissues was carried out by Shanghai NextCODE Co., Ltd (Shanghai, China) as previously reported (2). Briefly, total RNA from paraffin blocks was extracted using Allprep RNA FFPE Kit (QIAGEN, Valencia, CA). The amount and quality of extracted RNA were determined by NanoDrop 2000 Spectrophotometer (Thermo Fisher, Wilmington, DE) and Agilent Bioanalyzer 2100 system (Agilent, Santa Clara, CA), respectively. All RNA samples have a DV200 (percentage of RNA fragments >200 nucleotides fragment distribution value) $\geqslant$30%. The TruSeq RNA Access Library Prep Kit (Illumina, San Diego, CA), optimized to provide reproducible results of RNA from FFPE samples, was used to prepare the cDNA library. Paired-end 150-bp sequencing of the libraries was perfomed on the Illumina NovaSeq platform (Illumina). Sequencing reads were trimmed by Trimmomaticm (3) and aligned to the human reference genome hg19 by STAR RNA-seq aligner (4). Abundance of annotated genes was estimated by RSEM software (5).

Bulk RNA-seq of mouse frozen liver tissues and data analysis

Total RNA of mouse frozen liver tissue was isolated using an RNA simple Total RNA kit (Tiangen, Beijing, China) per the manufacturer's protocol. RNA concentration and quality were measured using the same methods than in the RNA samples from human FFPE liver tissues. Next, Poly(A) mRNAs were enriched using magnetic oligo (dT) beads (Invitrogen, Carlsbad, CA), followed by RNA-seq library preparation using the NEBNext Ultra RNA Library Prep Kit (New England Biolabs, Hitchin, UK) according to the manufacturer's instructions. A 125 bp paired-end run was performed on the Illumina

HiSeq2500 platform at Biomarker Technologies Co., Ltd (Beijing, China). Clean reads were obtained by Perl script, and then mapped to the mouse reference genome GRCm38 using Bowtie2 (6) and HISAT (7) algorithms. Gene expression levels were measured using FPKM values by Cufflinks software (8).

Determination and definition of liver fibrosis-specific matrisome genes (LFMGs)

The matrisome gene set previously analyzed (9) was also included in this study. GSE84044 (10), GSE14323 (11), GSE49541 (12), GSE103580 (13) and GSE130970 (14) were selected as derivation datasets for the identification of differentially expressed matrisome genes (DEMGs) between non-fibrotic (mild) and fibrotic (advanced) liver samples, since pathological diagnoses were determined based on liver biopsies and both non-fibrotic (mild) and fibrotic (advanced) samples were included. Gene expression levels detected by microarray were normalized using the robust multichip average algorithm (15), and DEMGs were analyzed using *Limma* (16) *R* package. Read counts from RNA sequencing (RNA-seq) data were scaled into fragments per kilobase per million (FPKM) or transcripts per million (TPM). DEMGs were analyzed using *edgeR* (17) *R* package. Adjusted $p<0.05$ and FC>1.5 were set as statistically significant criterion. Due to batch effect, platform diversity, or dissimilarity of tissue dissection, DEMGs overlapped among at least 4 out of the 5 datasets were analyzed using *UpSetR* (18) *R* package and defined as a LFMG signature.

Validation of the LFMG signature

First, an independent dataset (GSE149601), containing 140 non-cirrhotic and 55 cirrhotic patients diagnosed by liver biopsy or Fibroscan >12.5 kPa, was chosen as validation dataset, to validate the expression of LFMGs using unpaired Student's *t* test. A $p<0.05$ and FC>1.5 were set as statistically significant criterion. The diagnostic ability of the LFMG signature as a combined signature in liver fibrosis was evaluated using unsupervised clustering methods including uniform manifold approximation and projection (UMAP) and hierarchical clustering (HCL) performed by *umap* and *pheatmap R* packages, respectively.

Second, we used the GSE152329 (19) dataset with liver transcriptomic profiles from CCl_4- or mineral oil-injected hybrid mouse diversity panel (HMDP) strains consisting of approximately 200 well-characterized inbred strains of mice (one mouse per strain) and the liver transcriptomic gene expression profile generated from our newly-built liver fibrosis and resolution mouse models (C57BL/6J genetic background) to evaluate the expression pattern of the LFMG signature. The significance of differential expression of LFMGs between CCl_4- and mineral oil-injected mice was determined using unpaired Student's *t* test. A $p<0.05$ and FC >1.5 were considered statistically significant.

Third, we performed a cross-sectional comparison among different etiologies to illustrate whether the LFMG signature is independent of any etiology. The derivation datasets were separated into two groups: the first contained GSE84044 (HBV), GSE49541 (NASH) and GSE14323 (HCV); the second contained GSE130970 (NASH) and GSE149601 (HCV). Datasets from the two groups were generated by microarray and bulk RNA-seq, respectively. Gene expression profiles in each group were merged after normalization to the average level of each LFMG within one profile. Then, unsupervised clustering methods (UMAP and HCL) were performed, based on the LFMG signature expression level, to verify whether the LFMG signature could discriminate fibrotic and non-fibrotic samples with varying etiologies in the merged dataset.

Last, patients without fibrosis from GSE84044 (HBV), GSE130970 (NASH), GSE48452 (NASH) and GSE103580 (ALD) were used to analyze whether the LFMG signature expression was sensitive to the response to acute or chronic insult prior to fibrosis.

Functional specification of LFMGs

The Search Tool for Retrieval of Interacting Genes (STRING, version 11.0) (20) was used to predict internal interactions of LFMGs with a default confidence score >0.4,

followed by visualization of the LFMG regulatory network using Cytoscape (21). Functional interpretation of the LFMGs was performed and visualized using the *ClusterProfiler* (22) *R* package, based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. An adjusted $p<0.05$ was set as the cutoff criterion for statistical significance, which was determined by Fisher's exact test followed by the Benjamini-Hochberg correction.

Soft clustering analysis of LFMGs

Given that the GSE84044 and GSE130970 datasets included liver fibrosis patients with different pathological stages, as METAVIR scores ranged from F0 to F4, both datasets were used for soft clustering analysis, to assess the sustained increased or decreased expression patterns of LFMGs along with liver fibrogenesis. After the normalized LFMG expression level was averaged and $\log_2$-transformed among samples, soft clustering was implemented with the fuzzy c-means (FCM) clustering algorithm embedded in *Mfuzz* (23) *R* package with parameters $c=8$ and $m=1.25$. The cluster number was determined once its increase would not add a new cluster but instead split a previous cluster into two.

Molecular subclassification of liver fibrosis patients

Subclassification of liver fibrosis patients was carried out by the *pheatmap R* package. HCL analysis with the average linkage method and "euclidean" or "canberra" as a distance metric was employed to subclassify, without supervision, liver fibrosis patients naïve to treatment, on anti-HBV therapy for 78 or 260 weeks, and HBV, HCV or NASH-related liver fibrosis patients from GSE84044, GSE130970, GSE193080, GSE193066 and GSE48452, and all patients with identical fibrosis stage from each of the aforementioned profiles, based on expression of the LFMG signature. Patients were subclassified as "$LFMG^{High}$", when LFMGs relatively exhibited a fibrosis-biased expression pattern (higher expression), and as "$LFMG^{Low}$", when LFMGs relatively exhibited a non-fibrosis-biased expression pattern (lower expression).

Gene set enrichment analysis (GSEA)

GSEA was carried out with the c2 KEGG gene sets from the MSigDB as previously described (9) to interpret the biological functions of the identified LFMGs between $LFMG^{Low}$ and $LFMG^{High}$ patients, which was performed using the *gseKEGG* function from the *clusterProfiler* (22) *R* package. The number of permutations was set to 1,000 and a $p<0.05$ was considered as statistically significant. The normalized enrichment score (NES) is the enrichment score for the gene set after normalization of the analyzed gene sets; a positive NES indicates correlation with the first group and a negative NES indicates correlation with the second group.

Estimation of tissue infiltrating immune cells and stromal cells

The *MCPcounter* (24) *R* package is designed to recapitulate inter-sample stromal and immune cell populations based on highly specific transcriptomic markers. We used this method to quantify and compare immune cell infiltration or stromal cells, including fibroblasts, endothelial cells, monocytic lineage, neutrophils, lymphocytes ($CD8^+$, T, B, NK or cytotoxic lymphocytes) and myeloid dendritic cells between $LFMG^{Low}$ and $LFMG^{High}$ patients. We systematically compared liver infiltrating immune cell and stromal cell populations between $LFMG^{Low}$ and $LFMG^{High}$ patients with or without equal fibrosis stage. The summary effect size for each cell population was estimated using an inverse variance model as previously reported (25), which was done and visualized using Review Manager 5.4 software. The difference between $LFMG^{Low}$ and $LFMG^{High}$ patients was recognized as statistically significant when the *p* value in the meta-analysis was <0.05.

In vivo isolation of mouse liver nonparenchymal cells (NPCs)

Mouse liver NPCs were isolated from mineral oil-injected control, CCl_4-injected (peak fibrosis) and fibrosis resolution (1 week recovery from CCl_4) as previously described with minor modifications (26). Liver cell suspensions were obtained by ETGA buffer perfusion and subsequent collagenase D (0.08 U/ml, Roche, Indianapolis, IN) perfusion for 15 min (3 ml/min) via the inferior vena cava (IVC). To obtain a single-cell

NPC suspension, livers were explanted into a sterile Petri dish and further digested with Pronase E (0.5 mg/ml, Millipore Sigma, Burlington, MA) and DNase-I (2%, Roche, Indianapolis, IN) for 25 min at 37°C and spun down at 50*g* for 3 min to remove hepatocytes. Isolated NPCs were purified by density gradient with Percoll (GE Healthcare, Chicago, IL) at 35% and 70%. Supernatants were pelleted at 400*g* for 3 min and resuspended in cold fluorescence-activated cell sorting (FACS) buffer. Propidium Iodide (Biolegend, San Diego, CA) was added and sorting was performed to isolate single NPCs from debris, doublets and dead cells.

Sample preparation and cDNA library construction for single cell RNA sequencing (scRNA-seq)

Cell samples were handled strictly per the protocol of 10x Genomics Single Cell 3, Reagent Kits v3 (10x Genomics, Pleasanton, CA) as previously reported (27). In brief, sorted live NPCs from the mineral oil, peak and resolution groups were pelleted and resuspend to a concentration of 1000 cells/μl. The percentage of viable cells in each sample was greater than 80 by Trypan blue (Gibco, Waltham, MA) staining and automated cell counts. Single-cell barcoded cDNA was synthesized by reverse transcription using RT Master Mix (10x Genomics) and 10x Genomics Single Cell B Chip, followed by PCR amplification, to yield sufficient mass for library construction. Enzymatic fragmentation and size selection including end-repair, A-tailing, adaptor-ligation, and sample indexing PCR were used to optimize the cDNA amplicon size and produce Illumina-ready sequencing libraries.

ScRNA-seq and data analysis

Sequencing was run on the HiSeq 4000 system (Illumina) at University of Illinois at Urbana-Champaign DNA Sequencing Laboratory. Cell Ranger (10x Genomics, version 3.1.0) was used for sample demultiplexing, alignment (reference genome, mm10), filtering, and gene-level unique molecular identifier (UMI) counting. *Seurat v.3* (28) was used to perform downstream analysis filtering the expression matrices to ensure high-quality scRNA-seq data: cells with fewer than 300 genes and greater than 5% of total

UMIs mapping to the mitochondrial genome were filtered; genes found in less than 10% of cells, housekeeping genes and mitochondrial genes were excluded; clusters with fewer cells were also excluded for downstream analysis. "LogNormalData" in *Seurat v.3* was used to normalize the filtered gene-barcode matrix. Preliminary cell clustering was conducted based on principal component analysis (PCA)-reduced data for clustering analysis using *FindClusters* function in *Seurat v.3* (parameter resolution was set to 0.8). Cell types of total cells were annotated to known biological types according to the known marker genes listed in **Table S10**. Annotated hepatic stellate cells (HSCs), infiltrating and resident macrophages, and endothelial cells (ECs) were further re-integrated and re-clustered by *Seurat v.3* based on specific marker genes listed in **Table S10**. Cell clusters were visualized by UMAP plot. Expression of the LFMG signature in all NPCs from mineral oil, peak and resolution were visualized by the *ComplexHeatmap* (29) *R* package. The *FindMarkers* function in *Seurat v.3* was used to determine differentially expressed genes (DEGs) with thresholds of $p<0.05$ and FC>2. KEGG pathway enrichment analysis of top 1,000 genes with the highest abundance was conducted using the *ClusterProfiler* (22) *R* package. Moreover, a publicly available human scRNA-seq data (GSE136103) generated from 6 healthy livers and 4 cirrhotic livers ($CD45^-$ cells) (30, 31) was used and explored strictly per the analysis pipeline of mouse scRNA-data, in order to verify whether the scRNA-seq results from CCl_4-induced fibrosis in mice was comparable with patients with liver fibrosis.

Cell type deconvolution using BayesPrism

We used *BayesPrism*, a newly developed Bayesian model by Chu et al (32), to jointly infer the posterior distribution of cell type fractions from human bulk RNA-seq data using scRNA-seq (GSE136103) reference as prior information. Cell types of interest were deconvolved and compared between $LFMG^{Low}$ and $LFMG^{High}$ patients, with or without equal fibrosis stages. Specific *R* codes for cell type deconvolution using *BayesPrism* were included in the figshare platform (DOI: 10.6084/m9.figshare.22002707). Meta-analysis of cell type proportions between

$LFMG^{Low}$ and $LFMG^{High}$ patients with or without equal fibrosis stage was performed as mentioned above. A $p<0.05$ was considered as statistically significant.

SUPPLEMENTARY REFERENCES

1. Sun Y, Zhou J, Wang L, Wu X, Chen Y, Piao H, Lu L, et al. New classification of liver biopsy assessment for fibrosis in chronic hepatitis B patients before and after treatment. Hepatology 2017;65:1438-1450.

2. Li R, Yang Z, Shao F, Cheng H, Wen Y, Sun S, Guo W, et al. Multi-omics profiling of primary small cell carcinoma of the esophagus reveals RB1 disruption and additional molecular subtypes. Nat Commun 2021;12:3785.

3. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114-2120.

4. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21.

5. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011;12:323.

6. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10:R25.

7. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;12:357-360.

8. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010;28:511-515.

9. Chen W, Desert R, Ge X, Han H, Song Z, Das S, Athavale D, et al. The Matrisome Genes From Hepatitis B-Related Hepatocellular Carcinoma Unveiled. Hepatol Commun 2021;5:1571-1585.

10. Wang M, Gong Q, Zhang J, Chen L, Zhang Z, Lu L, Yu D, et al. Characterization of gene expression profiles in HBV-related liver fibrosis patients and identification of ITGBL1 as a key regulator of fibrogenesis. Sci Rep 2017;7:43446.

11. Mas VR, Maluf DG, Archer KJ, Yanek K, Kong X, Kulik L, Freise CE, et al. Genes involved in viral carcinogenesis and tumor initiation in hepatitis C virus-induced hepatocellular carcinoma. Mol Med 2009;15:85-94.

12. Moylan CA, Pang H, Dellinger A, Suzuki A, Garrett ME, Guy CD, Murphy SK, et al. Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. Hepatology 2014;59:471-482.

13. **Trepo E**, **Goossens N**, Fujiwara N, Song WM, Colaprico A, Marot A, Spahr L, et al. Combination of Gene Expression Signature and Model for End-Stage Liver Disease Score Predicts Survival of Patients With Severe Alcoholic Hepatitis. Gastroenterology 2018;154:965-975.

14. Hoang SA, Oseini A, Feaver RE, Cole BK, Asgharpour A, Vincent R, Siddiqui M, et al. Gene Expression Predicts Histological Severity and Reveals Distinct Molecular Profiles of Nonalcoholic Fatty Liver Disease. Sci Rep 2019;9:12541.

15. Han ES, Wu Y, McCarter R, Nelson JF, Richardson A, Hilsenbeck SG. Reproducibility, sources of variability, pooling, and sample size: important considerations for the design of high-density oligonucleotide array experiments. J Gerontol A Biol Sci Med Sci 2004;59:306-315.

16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.

17. **Robinson MD**, **McCarthy DJ**, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139-140.

18. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 2017;33:2938-2940.

19. **Tuominen I**, **Fuqua BK,** Pan C, Renaud N, Wroblewski K, Civelek M, Clerkin K, et al. The Genetic Architecture of Carbon Tetrachloride-Induced Liver Fibrosis in Mice. Cell Mol Gastroenterol Hepatol 2021;11:199-220.

20. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019;47:D607-D613.

21. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, et al.

Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498-2504.

22. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284-287.

23. Kumar L, M EF. Mfuzz: a software package for soft clustering of microarray data. Bioinformation 2007;2:5-7.

24. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol 2016;17:218.

25. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. PLoS Med 2008;5:e184.

26. Cabral F, Miller CM, Kudrna KM, Hass BE, Daubendiek JG, Kellar BM, Harris EN. Purification of Hepatocytes and Sinusoidal Endothelial Cells from Mouse Liver Perfusion. J Vis Exp 2018;132:56993.

27. Su T, Yang Y, Lai S, Jeong J, Jung Y, McConnell M, Utsumi T, et al. Single-Cell Transcriptomics Reveals Zone-Specific Alterations of Liver Sinusoidal Endothelial Cells in Cirrhosis. Cell Mol Gastroenterol Hepatol 2021;11:1139-1161.

28. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 2018;36:411-420.

29. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847-2849.

30. Ramachandran P, Dobie R, Wilson-Kanamori JR, Dora EF, Henderson BEP, Luu NT, Portman JR, et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. Nature 2019;575:512-518.

31. **Dobie R**, **Wilson-Kanamori JR**, Henderson BEP, Smith JR, Matchett KP, Portman JR, Wallenborg K, et al. Single-Cell Transcriptomics Uncovers Zonation of Function in the Mesenchyme during Liver Fibrosis. Cell Rep 2019;29:1832-1847 e1838.

32. Chu T, Wang Z, Pe'er D, Danko CG. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell

RNA sequencing in oncology. Nat Cancer 2022;3:505-517.

Author names in bold designate shared co-first authorship

SUPPLEMENTARY FIGURES

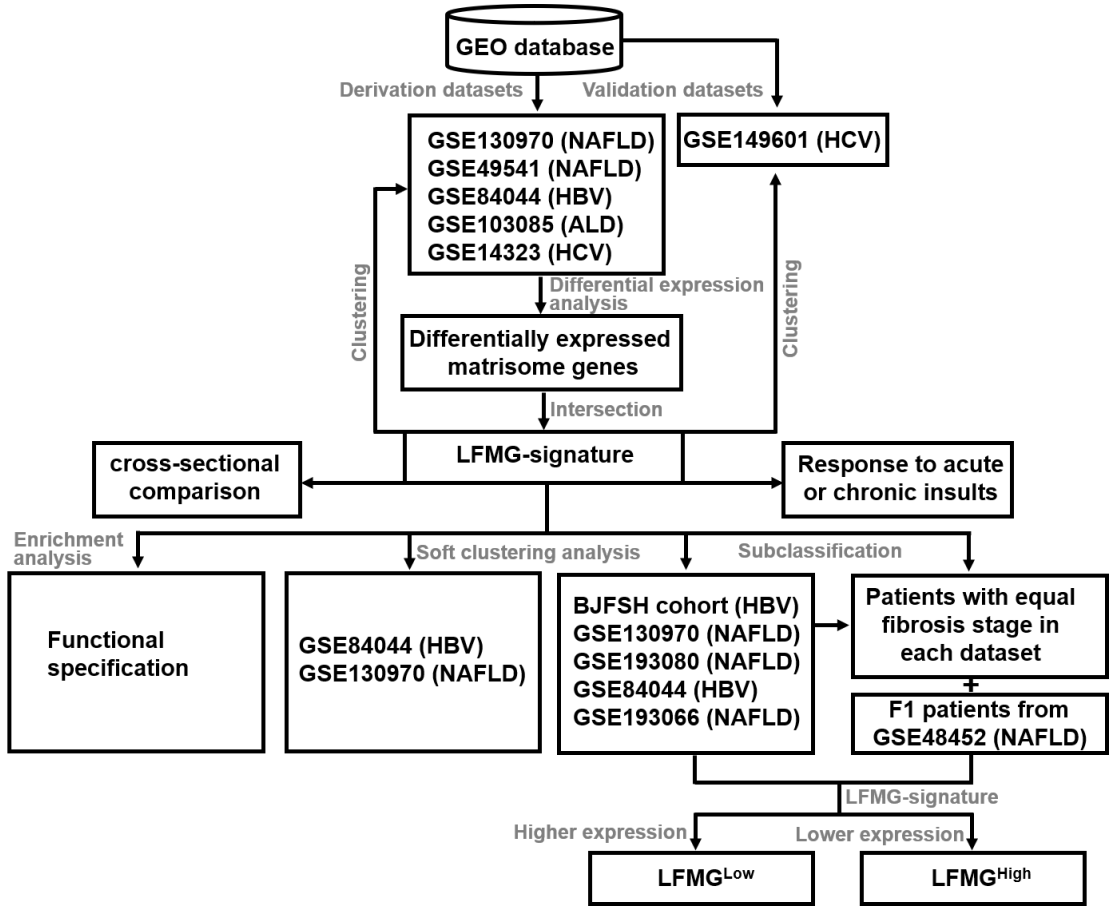

Figure S1. Flowchart for LFMG signature identification and molecular subclassification of liver fibrosis patients.

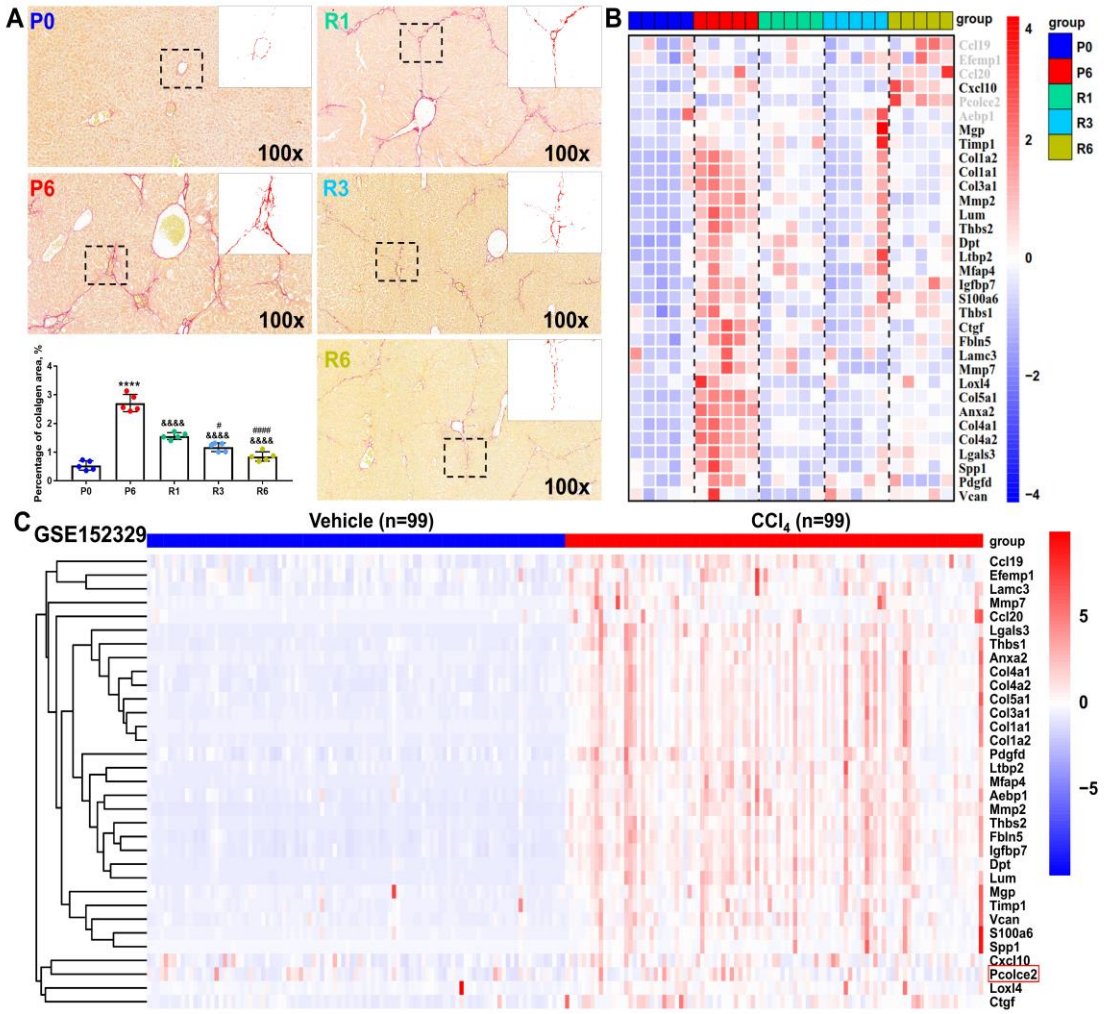

Figure S2. LFMGs expression in CCl_4-induced liver fibrosis and regression mice.

(A) Sirius red staining and quantitation of positive area of collagenous fibers in liver sections from CCl_4-induced liver fibrosis and regression in mice (our newly established mouse models). Zoomed images of positive areas of collagenous fibers are shown on the top right corner of each image. Comparison among the three groups were performed using one-way ANOVA followed by Tukey's multiple comparisons test or Kruskal-Wallis test. $^{****}p<0.0001$ vs P0; $^{\&\&\&\&}p<0.0001$ vs P6; $^{\#}p<0.05$ and $^{\#\#\#\#}p<0.01$ vs R1; n=5/group. **(B)** Heatmap of the LFMGs expression in each group of mice. Expression levels are scaled as a distribution with mean=0 and SD=1. The darker the blue, the lower the expression; the darker the red, the higher the expression. LFMGs

not significantly dysregulated in P6 compared to P0 group are highlighted in grey. **(C)** Heatmap of LFMGs expression in livers from the CCl_4- or mineral oil-injected hybrid mouse diversity panel consisting of ~100 inbred strains of mice (n=99 for each group) from the publicly available dataset GSE152329. Except for *Pcolce2* (p>0.05, highlighted by red rectangle), the other LFMGs were significantly upregulated in CCl_4-treated mice (p<0.05 and FC>1.5). Expression levels of LFMGs in the heatmap are scaled as a distribution with mean=0 and SD=1. The darker the blue, the lower the expression; the darker the red, the higher the expression. Since there are no orthologous mouse genes corresponding to human *CLEC4M* and *CXCL6*, the other 33 LFMGs were analyzed in all mice studies in our present study.

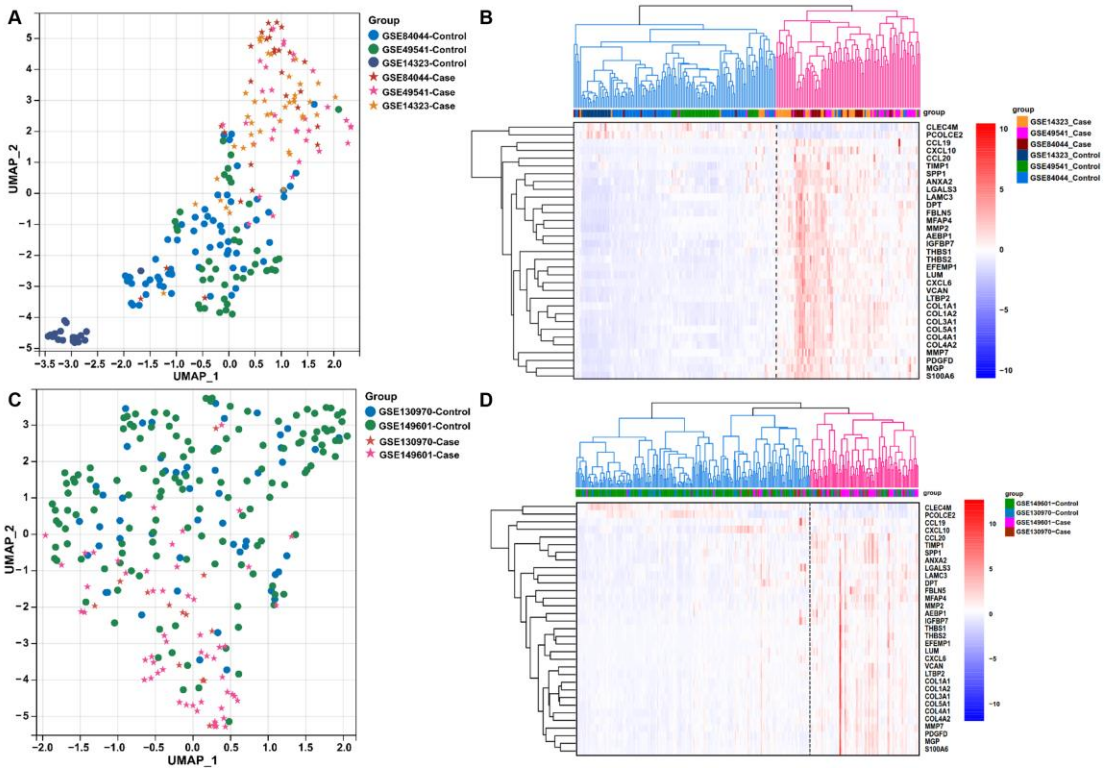

Figure S3. Cross-sectional comparison of the LFMGs expression pattern among liver fibrosis patients with different etiologies. UMAP plot and unsupervised HCL clustering of merged healthy (mild fibrosis) and fibrotic (advanced) liver samples from **(A** and **B)** GSE84044 (HBV), GSE49541 (NAFLD) and GSE14323 (HCV) datasets or **(C** and **D)** GSE130970 (NAFLD) and GSE149601 (HCV) datasets. Grouped liver samples in the UMAP plot are color- and shape-coded. Liver samples in HCL heatmaps are also color-coded. Expression level of LFMGs in heatmap are scaled as a distribution with mean=0 and SD=1. The darker the blue, the lower the expression; the darker the red, the higher the expression.

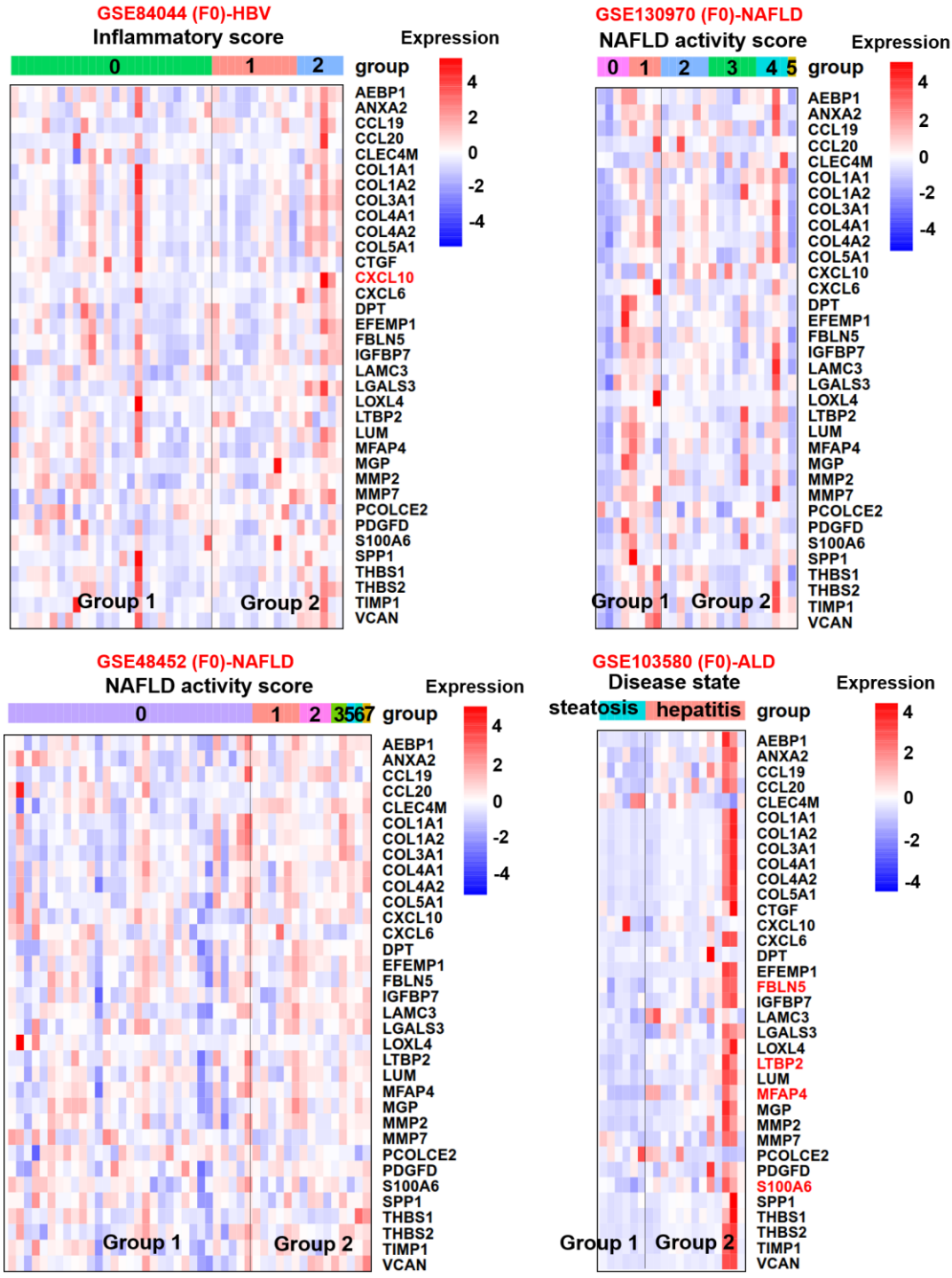

Figure S4. LFMG signature expression along with liver disease progression prior to fibrosis. Non-fibrotic (F0) patients from GSE84044 (HBV), GSE130970 (NAFLD), GSE48452 (NAFLD) and GSE103580 datasets were used to analyze the response of the LFMG signature to disease progression preceding fibrosis. The LFMG signature expression between Group 1 and 2 patients along with inflammation or NAFLD activity

progression was compared. Unpaired student's *t* test or Mann–Whitney U test were performed. A LFMG expression significantly different between groups was highlighted in red ($p<0.05$ and FC>1.5). Expression of LFMGs in heatmap are scaled as a distribution with mean=0 and SD=1. The darker the blue, the lower the expression; the darker the red, the higher the expression.

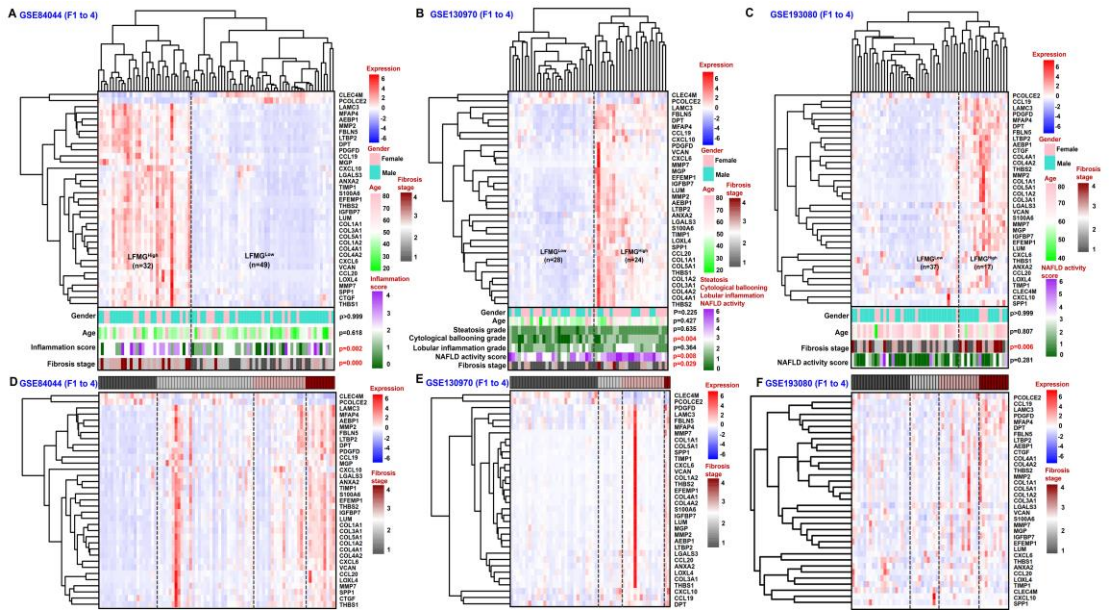

Figure S5. Subclassification of liver fibrosis patients from the GSE84044, GSE130970 and GSE193080 datasets. The LFMG signature subclassified liver fibrosis patients from **(A)** GSE84044 (HBV), **(B)** GSE130970 (NAFLD) and **(C)** GSE193080 (NAFLD) into $LFMG^{High}$ and $LFMG^{Low}$ subgroups. Gender, age, inflammation score, fibrosis stage, steatosis grade, ballooning grade and NAFLD activity score were color-coded and compared between $LFMG^{High}$ and $LFMG^{Low}$ patients. A $p<0.05$ was considered as statistically significant. Relationship between fibrosis stage and the LFMG signature expression pattern in liver fibrosis patients from **(D)** GSE84044, **(E)** GSE130970 and **(F)** GSE193080 were visualized using heatmaps. Fibrosis stage was color-coded. LFMG signature expression is scaled as a distribution with mean=0 and SD=1. The darker the blue, the lower the expression; the darker the red, the higher the expression.

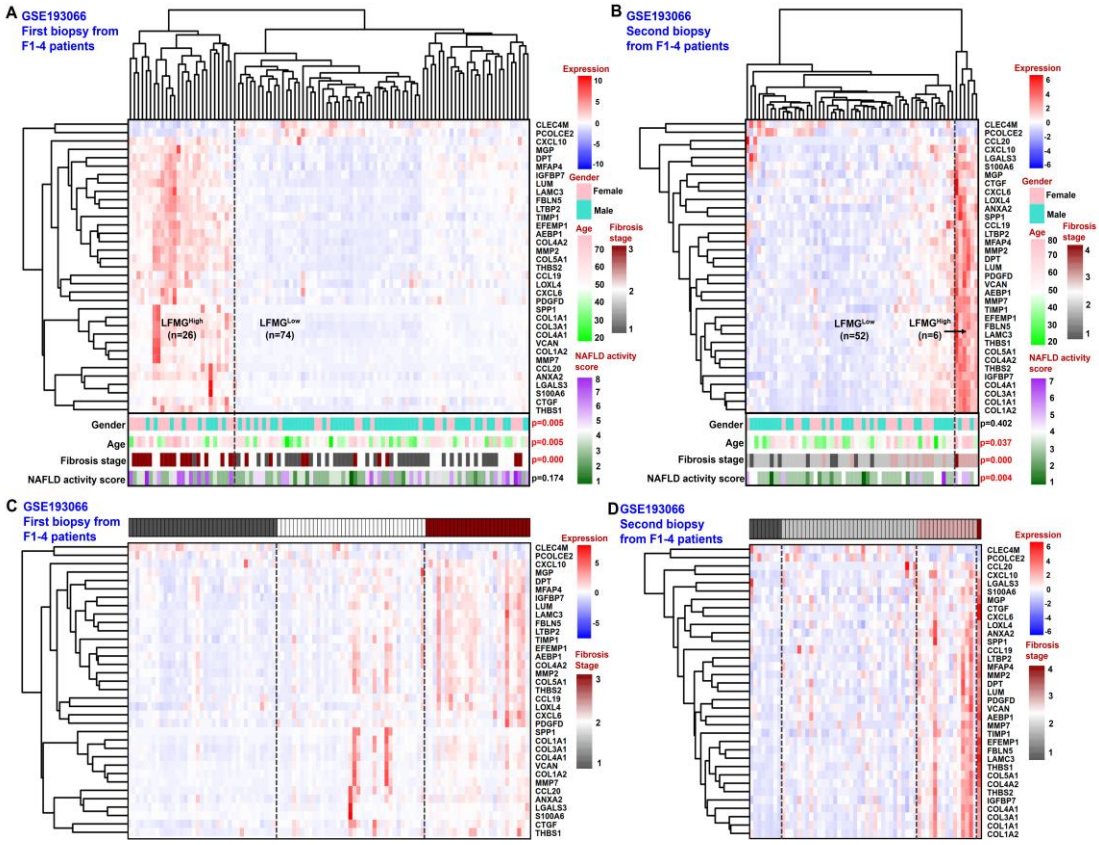

Figure S6. Subclassification of liver fibrosis patients from the GSE193066 dataset. The LFMG signature subclassified liver fibrosis patients undergoing **(A)** first or **(B)** second biopsy from GSE193066 (NAFLD) into $LFMG^{Low}$ and $LFMG^{High}$ subgroups. Gender, age, fibrosis stage and NAFLD activity score were color-coded and compared between $LFMG^{High}$ and $LFMG^{Low}$ patients. A $p<0.05$ was considered as statistically significant. Relationship between fibrosis stage and the LFMG signature expression pattern in liver fibrosis patients undergoing **(C)** first or **(D)** second biopsy from GSE193066 (NAFLD) were visualized using heatmaps. Fibrosis stage was color-coded. LFMG signature expression is scaled as a distribution with mean=0 and SD=1. The darker the blue, the lower the expression; the darker the red, the higher the expression.

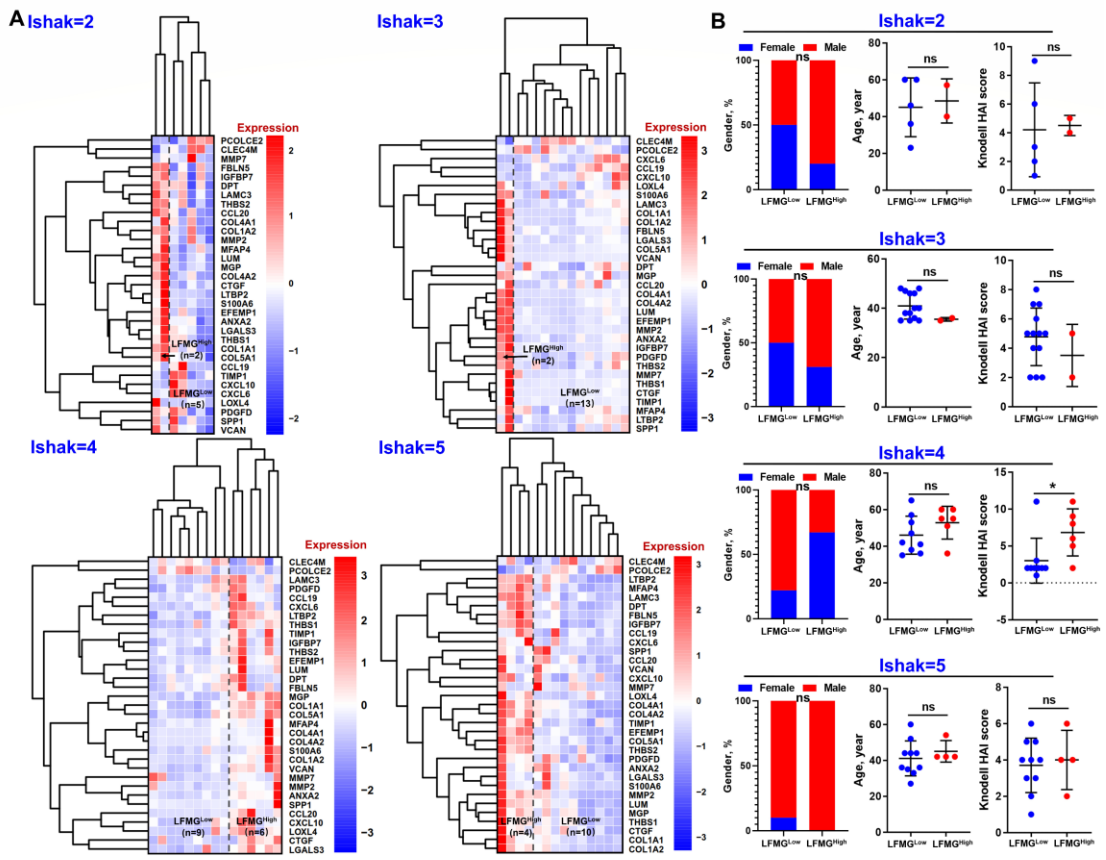

Figure S7. Subclassification of liver fibrosis patients with identical fibrosis stage from the BJFSH cohort. (A) The LFMG signature subclassified liver fibrosis patients with equal fibrosis stage from Ishak 2 to 5 into LFMG^{Low} and LFMG^{High} subgroups. The LFMG signature expression is scaled as a distribution with mean=0 and SD=1. The darker the blue, the lower the expression; the darker the red, the higher the expression. (**B**) Gender, age and Knodell score were compared between LFMG^{High} and LFMG^{Low} patients. *p<0.05; ns, not significant.

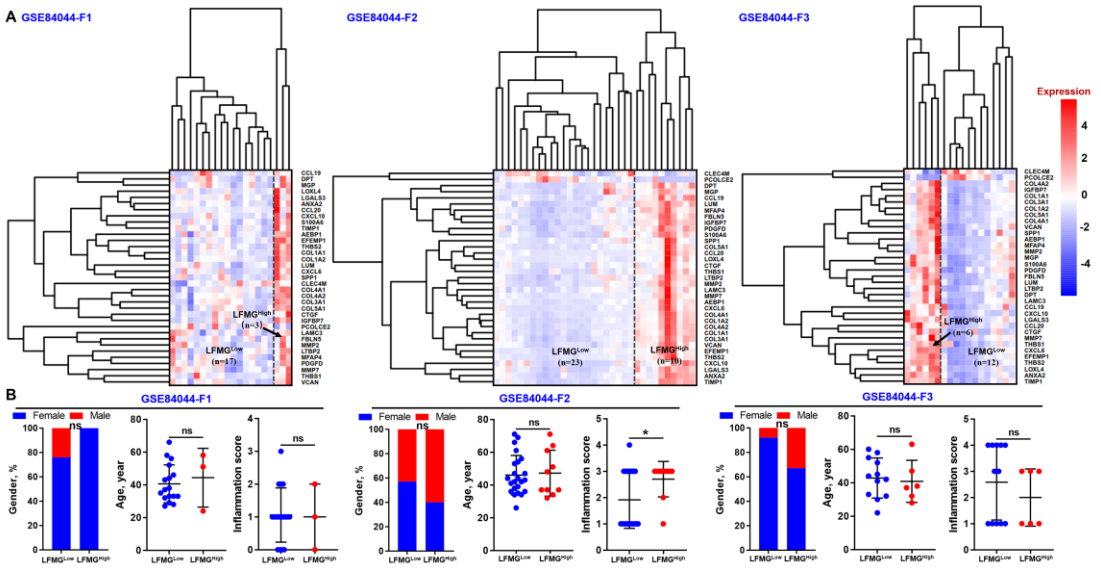

Figure S8. Subclassification of liver fibrosis patients with identical fibrosis stage from GSE84044. (A) LFMG signature subclassified liver fibrosis patients with equal fibrosis stage from METAVIR F1 to 3 into $LFMG^{Low}$ and $LFMG^{High}$ subgroups. The LFMG signature expression is scaled as a distribution with mean=0 and SD=1. The darker the blue, the lower the expression; the darker the red, the higher the expression. **(B)** Gender, age and inflammation score were compared between $LFMG^{High}$ and $LFMG^{Low}$ patients. *p<0.05; ns, not significant.

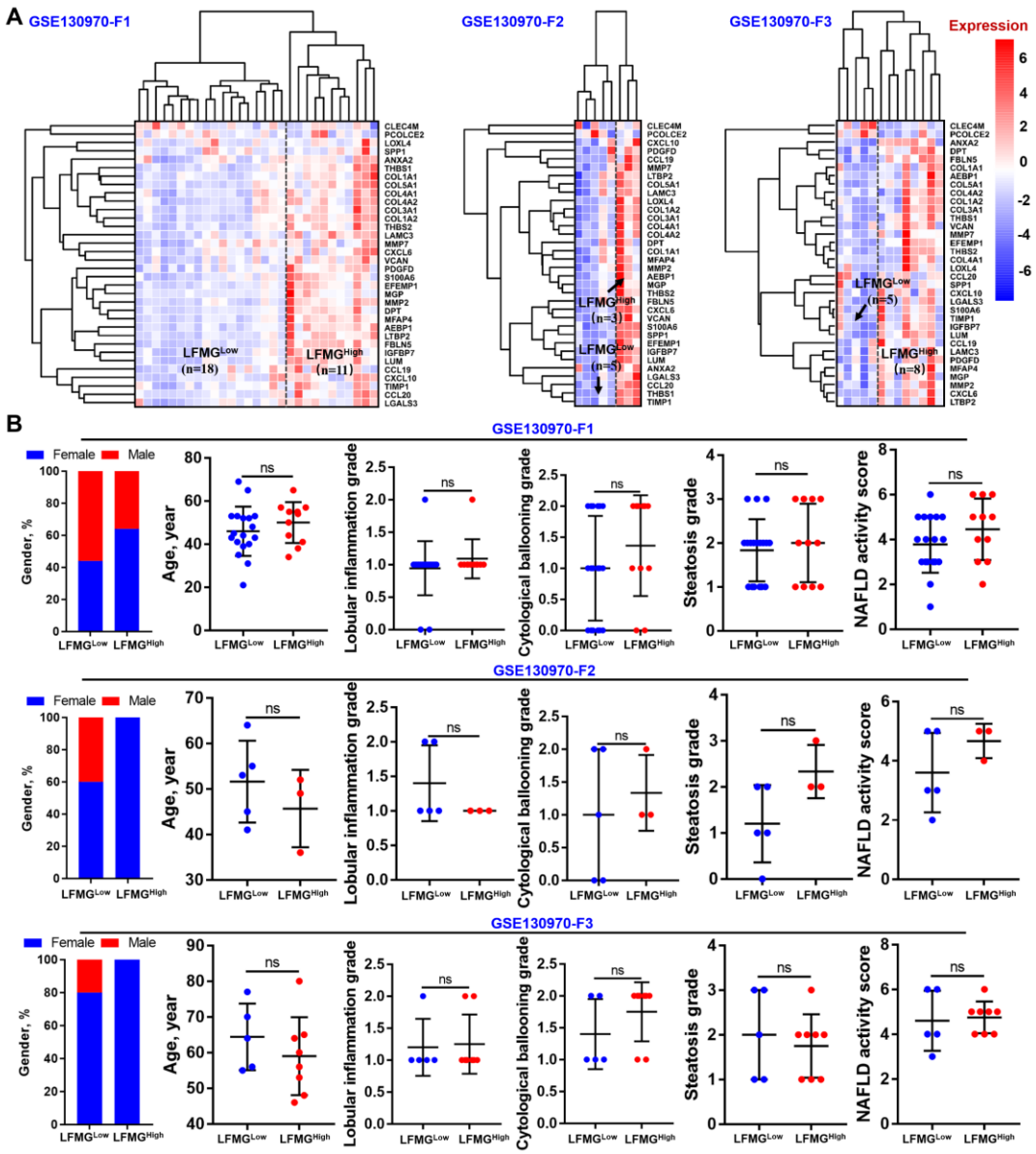

Figure S9. Subclassification of liver fibrosis patients with identical fibrosis stage from GSE130970. (A) The LFMG signature subclassified liver fibrosis patients with equal fibrosis stage from METAVIR F1 to 3 into $LFMG^{Low}$ and $LFMG^{High}$ subgroups. The LFMG signature expression is scaled as a distribution with mean=0 and SD=1. The darker the blue, the lower the expression; the darker the red, the higher the expression. **(B)** Gender, age, inflammation grade, ballooning grade, steatosis grade and NAFLD activity score were compared between $LFMG^{High}$ and $LFMG^{Low}$ patients. ns, not significant.

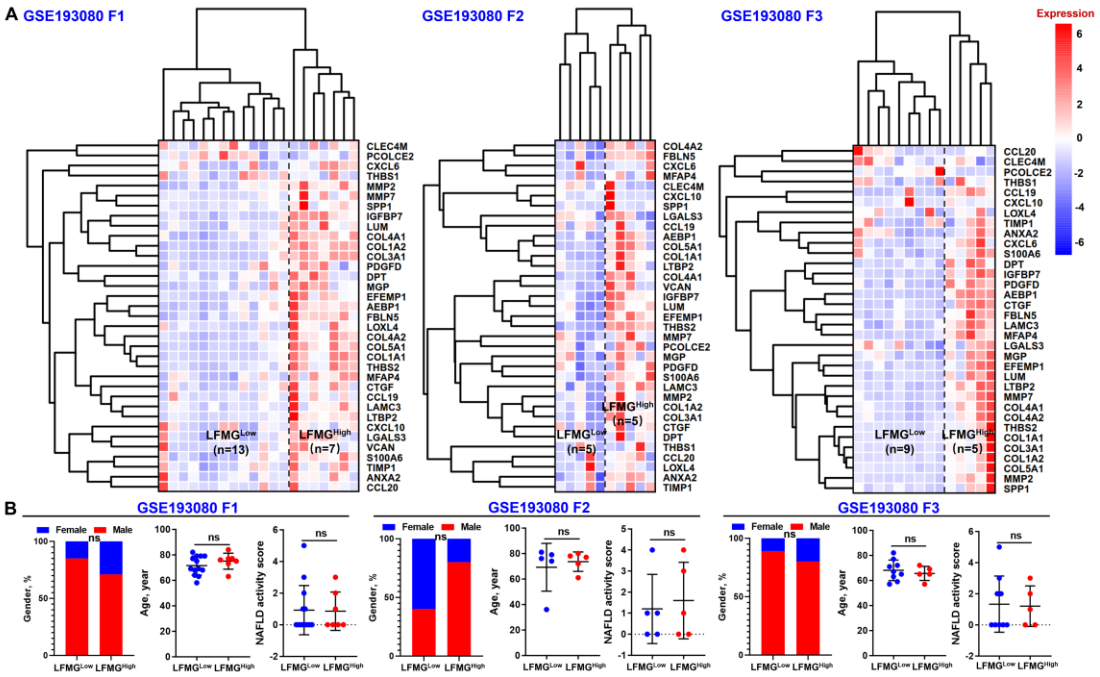

Figure S10. Subclassification of liver fibrosis patients with identical fibrosis stage from GSE193080. (A) The LFMG signature subclassified liver fibrosis patients with equal fibrosis stage from METAVIR F1 to 3 into LFMG^{Low} and LFMG^{High} subgroups. The LFMG signature expression is scaled as a distribution with mean=0 and SD=1. The darker the blue, the lower the expression. **(B)** Gender, age and NAFLD activity score were compared between LFMG^{High} and LFMG^{Low} patients. ns, not significant.

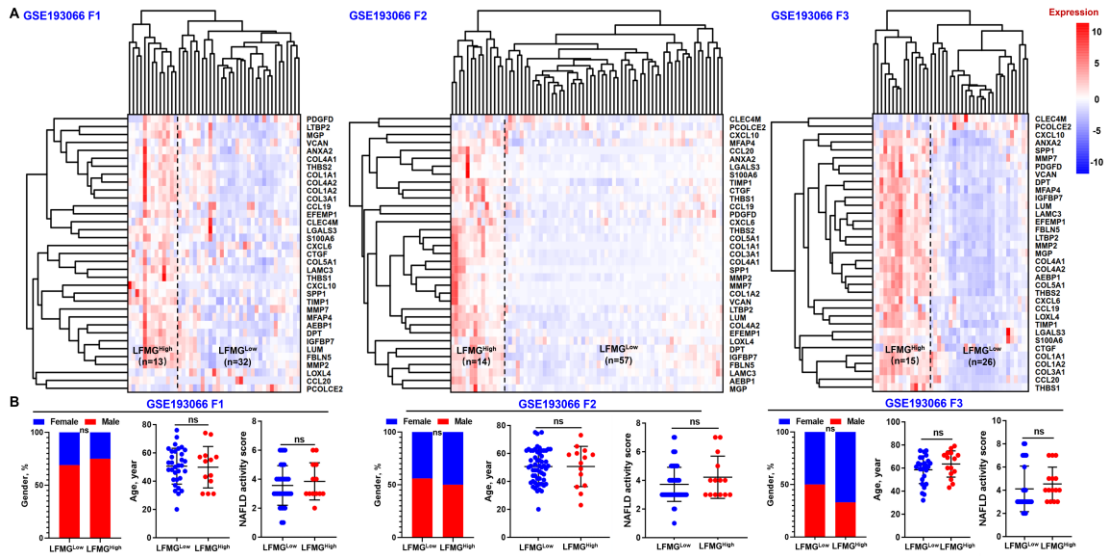

Figure S11. Subclassification of liver fibrosis patients with identical fibrosis stage from GSE193066. (A) The LFMG signature subclassified liver fibrosis patients with equal fibrosis stage from METAVIR F1 to 3 into $LFMG^{Low}$ and $LFMG^{High}$ subgroups. The LFMG signature expression is scaled as a distribution with mean=0 and SD=1. The darker the blue, the lower the expression; the darker the red, the higher the expression. **(B)** Gender, age and NAFLD activity score were compared between $LFMG^{High}$ and $LFMG^{Low}$ patients. ns, not significant.

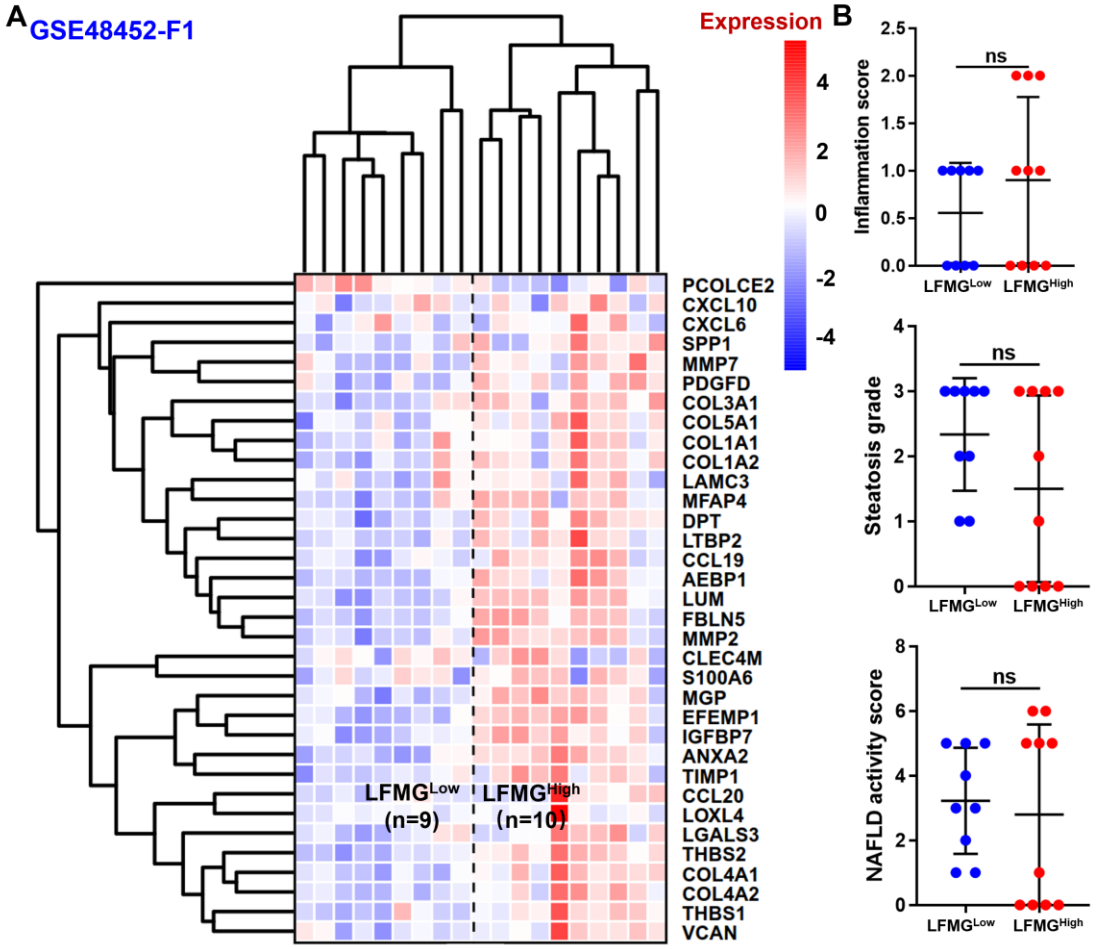

Figure S12. Subclassification of liver fibrosis patients with identical fibrosis stage from GSE48452. (A) The LFMG signature subclassified liver fibrosis patients with equal fibrosis stage (METAVIR F1) into $LFMG^{Low}$ and $LFMG^{High}$ subgroups. The LFMG signature expression is scaled as a distribution with mean=0 and SD=1. The darker the blue, the lower the expression; the darker the red, the higher the expression. **(B)** Inflammation score, steatosis grade and NAFLD activity score were compared between $LFMG^{High}$ and $LFMG^{Low}$ patients. ns, not significant.

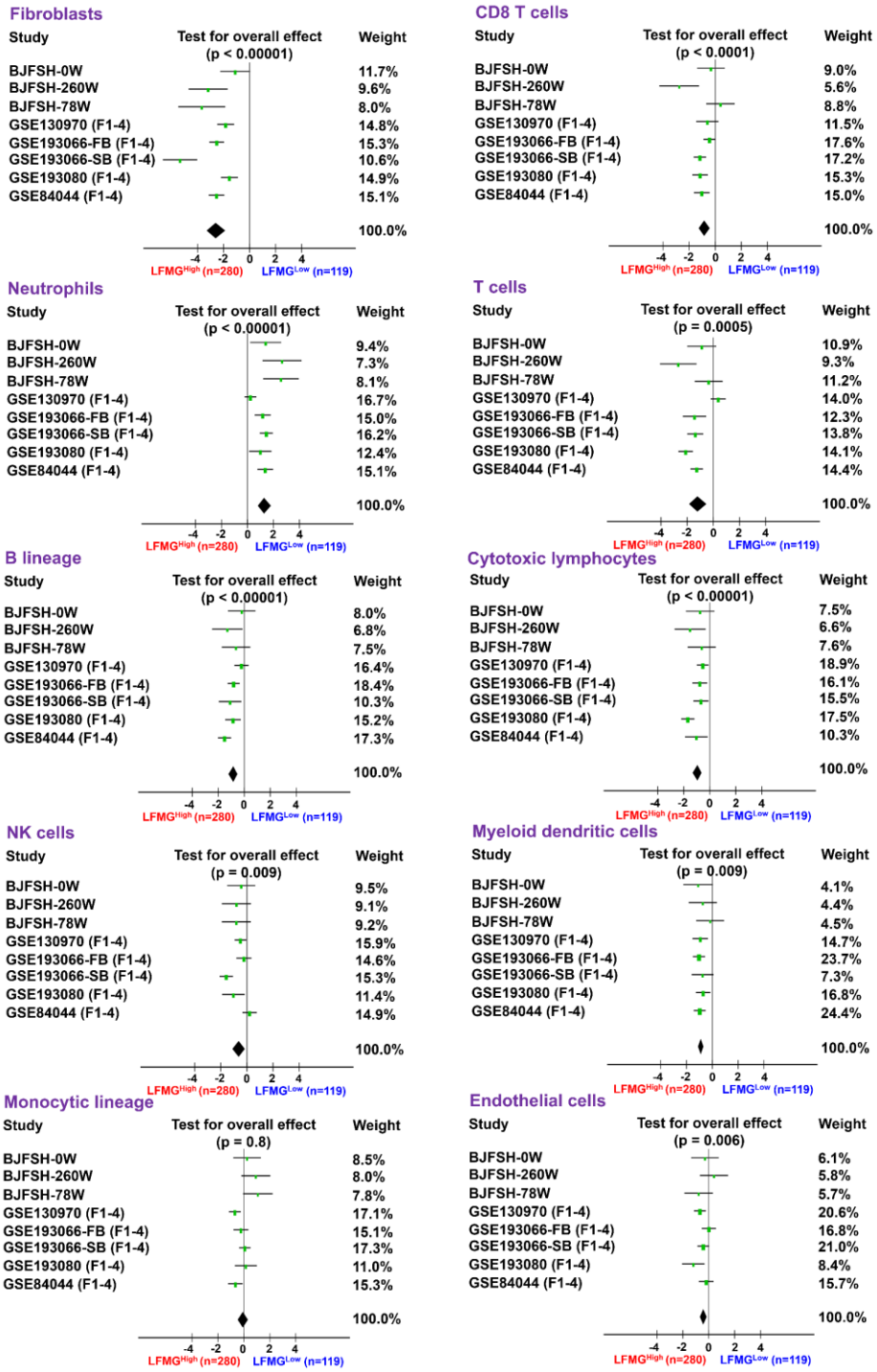

Figure S13. Comparison of liver infiltrating immune cells and fibroblasts between subgroups of patients from the BJFSH cohort (0W, 78W and 260W), GSE130970, GSE193066 (first and second liver biopsy), GSE193080 or GSE84044 datasets. Forest plots show comparisons of cell abundance (fibroblasts, neutrophils, B lineage, NK cells, monocytic lineage, CD8 T cells, T cells, cytotoxic

lymphocytes, myeloid dendritic cells and endothelial cells) between $LFMG^{Low}$ and $LFMG^{High}$ patients with varying fibrosis stages. Whiskers in the forest plots indicate the confidence interval of cell abundance. The overall effect among different datasets was assessed using Hedges' adjusted *g* based on a random effect model. A $p<0.05$ was considered as statistically significant.

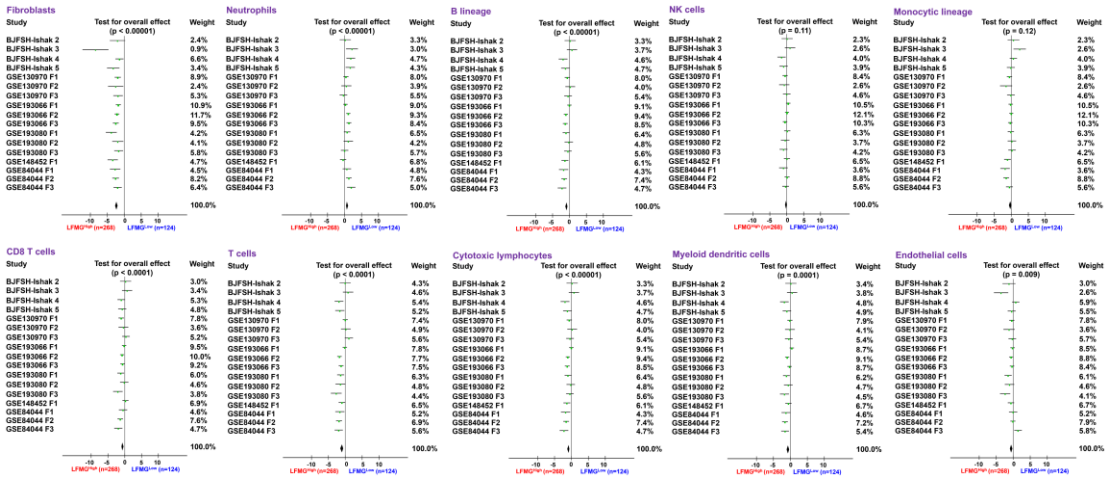

Figure S14. Comparisons of liver infiltrating immune cells and fibroblasts between subgroups of patients with equal fibrosis stage from the BJFSH cohort and public datasets. Forest plots showed the comparisons of cell abundance (fibroblasts, neutrophils, B lineage, NK cells, monocytic lineage, CD8 T cells, T cells, cytotoxic lymphocytes, myeloid dendritic cells and endothelial cells) between $LFMG^{Low}$ and $LFMG^{High}$ patients with equal fibrosis stage. Whiskers in the forest plots indicate the confidence interval of cell abundance. The overall effect among different datasets was assessed using Hedges' adjusted *g* based on a random effect model. A $p<0.05$ was considered as statistically significant.

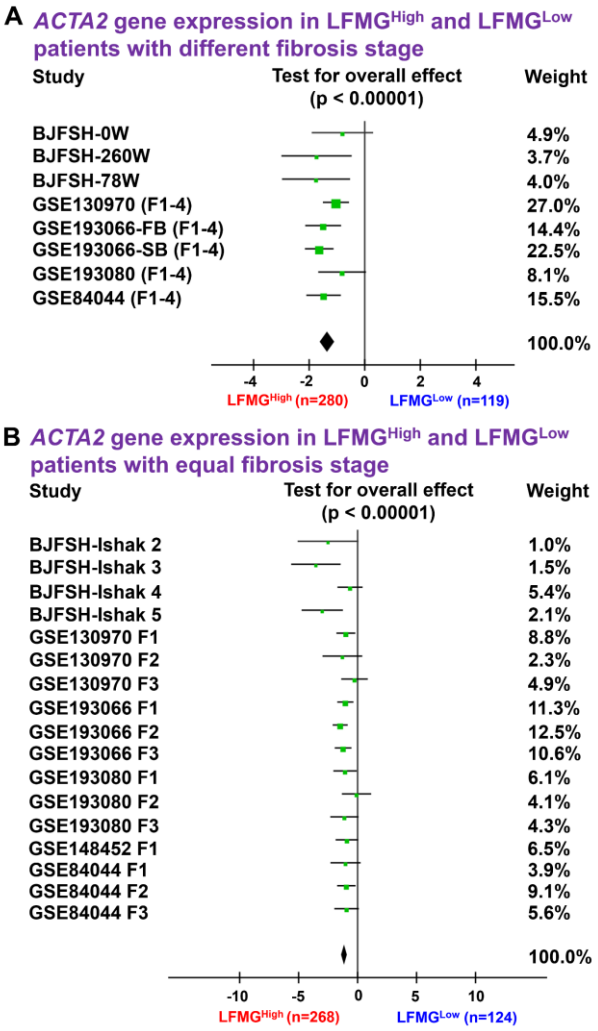

Figure S15. Comparison of *ACTA2* gene expression between LFMG^{Low} and LFMG^{High} patients with or without identical fibrosis stage from the BJFSH cohort and public datasets. Forest plots showed the comparison of *ACTA2* expression between LFMG^{Low} and LFMG^{High} patients with **(A)** varying or **(B)** equal fibrosis stages. Whiskers in the forest plots indicate the confidence interval of *ACTA2* expression. The overall effect among different datasets was assessed using Hedges' adjusted *g* based on a random effect model. A $p<0.05$ was considered as statistically significant.

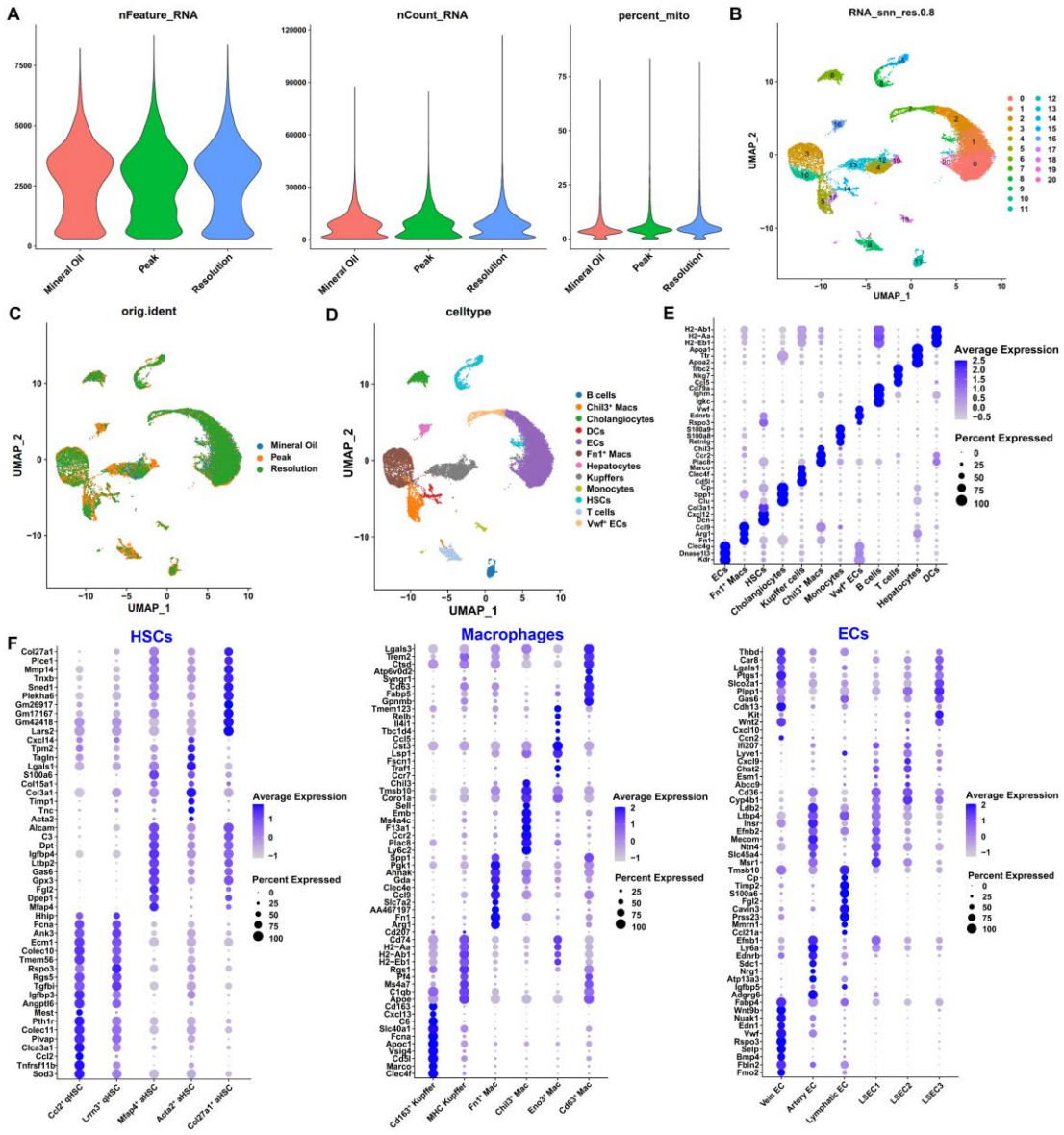

Figure S16. Characterization of scRNA-seq from liver NPCs isolated from mice at peak fibrosis and resolution. (A) Violin plot illustrating the number of genes (nFeature_RNA), UMIs (nCount_RNA) and the percentage of mitochondrial genes (percent_mito) in each cell from mineral oil (control), peak liver fibrosis and resolution in mice, marked by red, green and light blue, respectively. **(B)** UMAP plot showing cell clusters color-coded according to gene expression characteristics. A total of 20 cell clusters were identified. **(C)** UMAP plot showing spatial distribution of color-coded cells from mineral oil, peak and resolution mice models marked by red, green and light blue,

respectively. **(D)** UMAP plot showing annotated cell types based on canonical markers. A total of 12 cell types were identified. **(E)** Dot plot showing the top 3 marker genes of each cell type in **(D)**. **(F)** Dot plot showing the top marker genes for re-clustered cell types of HSCs, macrophages and endothelial cells. Percentage of cells with normalized expression level for marker genes is reflected by circle size; color intensity reflects average expression level across all cells within each cluster.

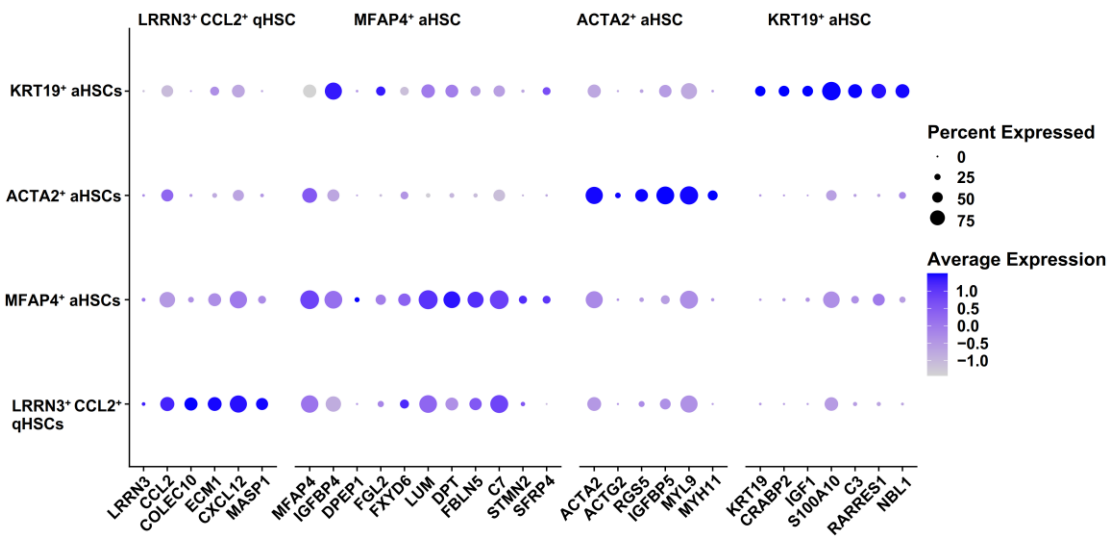

Figure S17. Top marker genes for re-clustered cell types of HSCs (publicly available GSE136103). Dot plot showing top marker genes for re-clustered cell types of HSCs. Percentage of cells with normalized expression level for marker genes is reflected by circle size; color intensity reflects average expression level across all cells within each cluster.

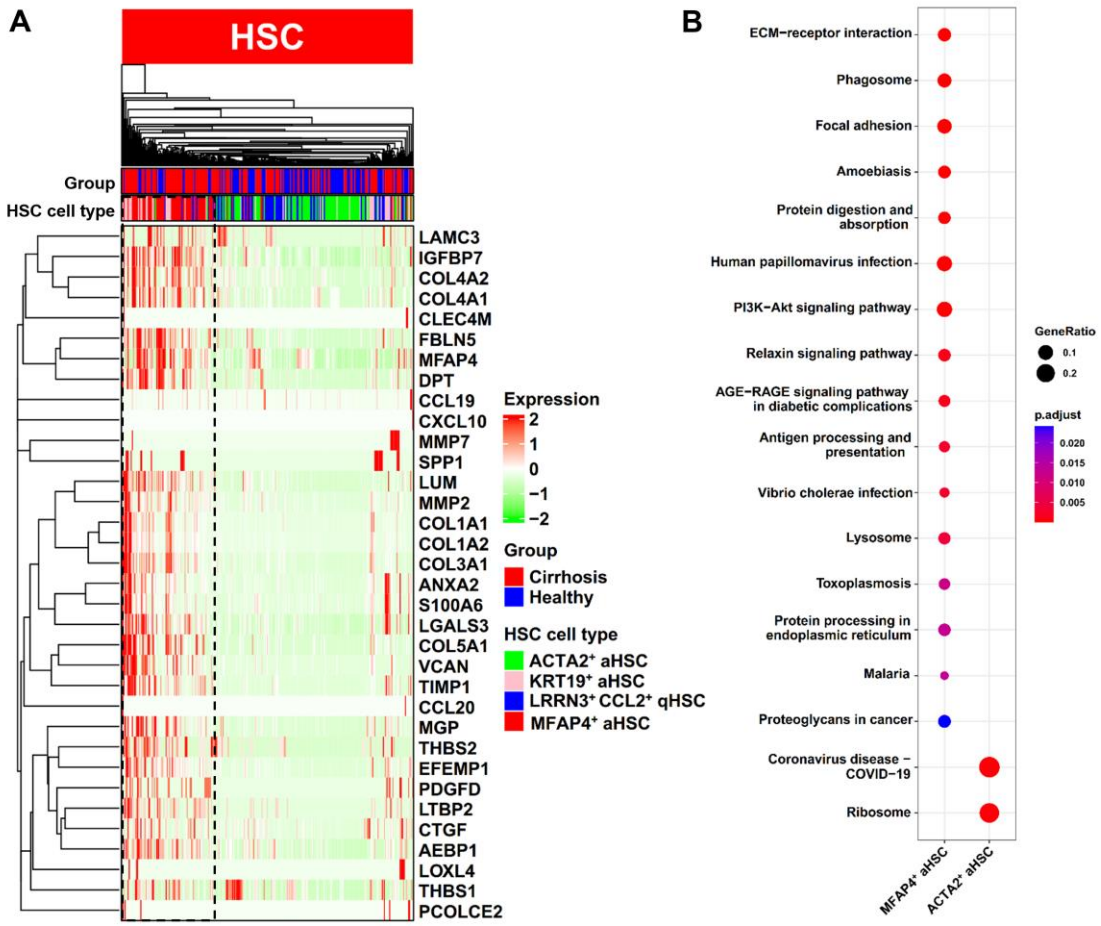

Figure S18. The LFMG signature expression in re-clustered HSC subtypes identified from the GSE136103 dataset and functional analysis of MFAP4⁺ aHSCs.

(A) The heatmap shows the LFMG signature expression in 4 re-clustered HSC subtypes. Group (sample) and HSC cell types were color-coded. The LFMG signature expression is scaled as a distribution with mean=0 and SD=1. The darker the green, the lower the expression; the darker the red, the higher the expression. **(B)** Significantly enriched KEGG pathways related to MFAP4⁺ and ACTA2⁺ aHSCs based on the functional enrichment analysis of top 1,000 genes with the highest abundance in each cell type.

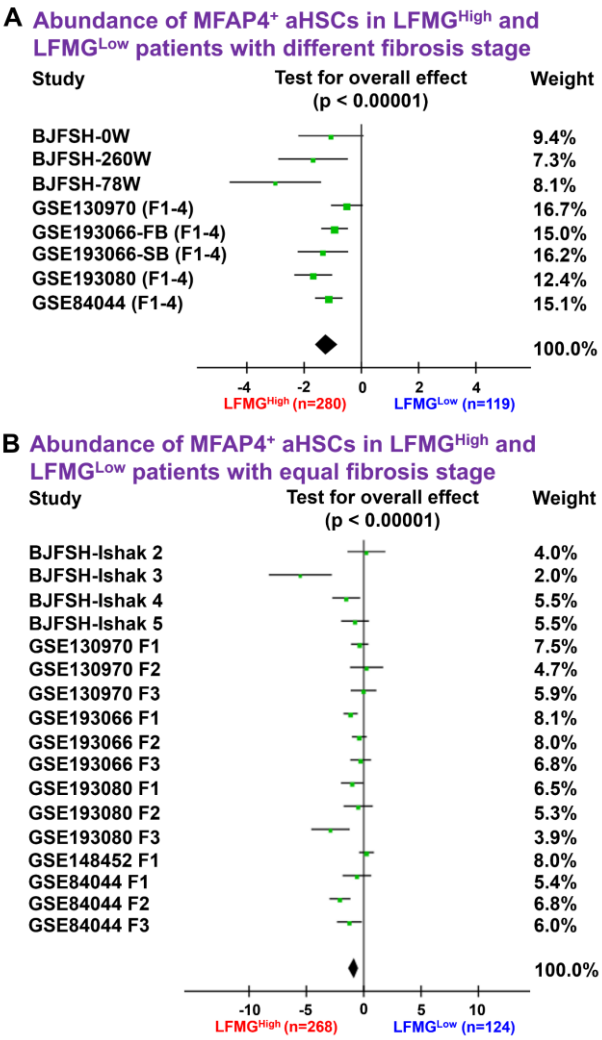

Figure S19. Comparison of MFAP4$^+$ aHSC proportion between LFMG^{Low} and LFMG^{High} patients with or without identical fibrosis stages. Forest plots showed the comparison of the MFAP4$^+$ aHSC percentage between LFMG^{Low} and LFMG^{High} patients with **(A)** varying or **(B)** equal fibrosis stages. MFAP4$^+$ aHSC proportion was deconvolved from bulk RNA-seq data using GSE136103 scRNAseq data as reference information. Whiskers in the forest plots indicate the confidence interval of the MFAP4$^+$ aHSC percentage. The overall effect among different datasets was assessed using Hedges' adjusted *g* based on a random effects model. A $p<0.05$ was considered as statistically significant.

SUPPLEMENTARY TABLES

Table S1. Publicly available gene expression profiles of liver fibrosis from GEO used in this study.

GEO ID	Etiology	Sample size	Sample resource	Platform	Year
GSE84044	HBV	F0=43 F1=20 F2=33 F3=18 F4=10	Frozen liver biopsy	GPL570	2016
GSE14323	HCV	Normal=19 LC=41	Liver transplantation	GPL571	2009
GSE49541	NAFLD	F0-1=40 F3-4=32	Frozen liver biopsy	GPL570	2013
GSE103580	ALD	Non-LC=19 LC=67	FFPE liver biopsy	GPL13667	2018
GSE130970	NAFLD	F0=25 F1=29 F2=8 F3=14 F4=2	Frozen liver biopsy	GPL16791	2019
GSE149601	HCV	Non-LC=140 LC=55	Liver biopsy	GPL20301	2020
GSE193066	NAFLD	F1=45 F2=71 F3=41 F4=1	Liver biopsy	GPL18573	2022
GSE193080	NAFLD	F1=20 F2=10 F3=14 F4=10	FFPE tumor-adjacent liver tissue	GPL18573	2022
GSE48452	NAFLD	F0=46 F1=19	Liver biopsy	GPL11532	2013

***Abbreviations*:** ALD, alcohol-associated liver disease; FFPE, formalin-fixed and paraffin-embedded; GEO, Gene Expression Omnibus; HBV, hepatitis B virus; HCV, hepatitis C virus; LC, liver cirrhosis; FFPE, formalin-fixed and paraffin-embedded; NAFLD, non-alcoholic fatty liver disease.

Table S2. Detailed information of antiviral therapy and timepoint of liver biopsy patients from the BJFSH cohort.

Patient ID	Age	Gender	Treatment	Timepoint of liver biopsy		
				Baseline	78 weeks	260 weeks
1	38	M	ETT	Yes	Yes	Yes
2	35	M	ETT	Yes	Yes	Yes
3	42	M	ETT	Yes	Yes	Yes
4	36	F	ETT	Yes	Yes	Yes
5	46	M	ETT	Yes	Yes	Yes
6	40	F	ETT	Yes	Yes	Yes
7	60	F	ETT	Yes	Yes	Yes
8	44	M	ETN	Yes	Yes	Yes
9	60	M	ETN	Yes	Yes	Yes
10	36	M	ETN	Yes	Yes	Yes
11	55	M	ETV	Yes	Yes	-
12	35	M	ETV	Yes	Yes	-
13	57	M	ETT	Yes	Yes	-
14	48	F	ETT	Yes	-	Yes
15	47	M	ETT	Yes	-	Yes
16	36	M	ETT	-	Yes	Yes
17	23	F	ETT	Yes	-	-
18	51	F	ETV	-	-	Yes
19	53	F	ETT	-	-	Yes
20	41	M	ETN	-	-	Yes
21	52	M	ETN	-	-	Yes
22	42	M	ETV	-	-	Yes
23	35	M	ETV	-	-	Yes
24	44	F	ETN	-	-	Yes
25	65	M	ETN	-	-	Yes
26	54	M	ETN	-	-	Yes
27	33	M	ETT	-	Yes	-
28	27	M	ETT	-	Yes	-

***Abbreviations*:** ETN, combination of entecavir and alpha interferon therapy; ETT, combination of entecavir and alpha thymosin; ETV, entecavir; F, female; M, male. “Yes” represents patient that underwent liver biopsy at the indicated timepoint and “-” represents patient that did not undergo liver biopsy at the indicated timepoint.

Table S3. Clinical characteristics of HBV-related liver fibrosis patients pre- or post-treatment.

	Baseline (n=16)	78 weeks (n=16)	260 weeks (n=22)
Age, year	43.9±10.4	42.5±10.3	45.7±8.8
Gender	M=11	M=13	M=15
WBC, 10^9/L	4.7±0.9	4.0±1.1	5.7±1.2*##
PLT, 10^9L	160.0 (147.5, 182.0)	128.0 (105.5, 151.5)	182.5 (161.5, 205.8)##
ALT, U/L	63.5 (48.7, 106.5)	28.5 (20.8, 33.0)**	21.0 (16.3, 26.0)**
AST, U/L	46.5 (35.3, 65.8)	26.0 (28.4, 20.4)**	20.5 (16.6, 24.5)**
ALP, U/L	71.5 (67.0, 98.0)	70.5 (63.8, 80.0)	78.0 (70.3, 86.8)
GGT, U/L	35.9 (26.8, 63.5)	28.0 (24.0, 29.5)	20.5 (15.5, 42.5)
ALB, g/L	41.5±4.5	42.7±4.3	45.6±4.0*
TBIL, μmol/L	17.4±6.8	13.7±4.4	17.2±7.6
AFP, ng/ml	9.5 (3.6, 17.8)	3.8 (2.7, 6.5)	2.5 (1.9, 3.8)**
LSM, kPa	11.8 (7.1, 14.7)	7.8 (6.5, 11.1)	6.1 (5.1, 11.4)
Log (HBV DNA)	7.3 (6.8, 8.0)	0.0 (0.0, 1.3)**	0.0 (0.0, 0.0)**
Ishak score, n (%)			
1	0 (0.0)	1 (6.3)	0 (0.0)
2	3 (18.8)	0 (0.0)	4 (18.2)
3	7 (43.8)	7 (43.8)	1 (4.5)
4	4 (25.0)	3 (18.8)	8 (36.4)
5	1 (6.3)	5 (31.3)	8 (36.4)
6	1 (6.3)	0 (0.0)	1 (4.5)
PIR, n (%)			**
0	3 (18.8)	1 (6.3)	4 (18.2)
1	1 (6.3)	8 (50.0)	9 (40.9)
2	1 (6.3)	1 (6.3)	8 (36.4)
3	11 (68.8)	6 (37.5)	1 (4.5)

Note: continuous variables were expressed as median (interquartile range, IQR) or mean ± standard deviation (SD). One-way ANOVA followed by Tukey's multiple comparisons test or Kruskal-Wallis test was used to assess changes among three groups. Categorical variables were presented as count (percentage) and compared by Chi-square trend test between any two groups. *p<0.05 and **p<0.01 (vs baseline) and ##p<0.01 (vs 78 weeks) were considered as statistically significant. ***Abbreviations***: ALB, albumin; AFP, alpha fetoprotein; ALP, alkaline phosphatase; ALT, alanine transaminase; AST, aspartate aminotransferase; GGT, glutamyl transpeptidase; HBV, hepatitis B virus; LSM, liver stiffness measurement; M, male; PLT, platelet; PIR, progressive, indeterminate, and predominately regressive; TBIL, total bilirubin; WBC, white blood cell.

Table S4. Clinical characteristics of regressive HBV-related liver fibrosis patients with three consecutive liver biopsies.

	Baseline (n=4)	78 weeks (n=4)	260 weeks (n=4)
Age, year	39.5±4.7	-	-
Gender	M=2	-	-
WBC, 10^9/L	4.9±0.5	3.7±1.3	5.3±0.2
PLT, 10^9L	150.5±33.2	146.0±46.8	205.0±17.3
ALT, U/L	112.0 (86.2, 335.0)	21.0 (19.0, 22.8)	15.0 (10.8, 19.3)*
AST, U/L	68.5 (55.0, 154.3)	23.3 (18.9, 25.8)	16.0 (15.7, 16.5)*
ALP, U/L	81.5 (65.2, 98.0)	65.0 (64.5, 67.3)	73.5 (66.5, 79.8)
GGT, U/L	43.4 (36.5, 84.8)	25.5 (22.5, 27.3)	18.0 (14.0, 23.5)*
ALB, g/L	42.1±5.0	39.3±4.7	46.2±2.9
TBIL, μmol/L	20.4 (14.9, 25.3)	11.4 (6.5, 17.4)*	15.2 (12.1, 19.4)
AFP, ng/ml	7.5 (5.7, 21.9)	2.3 (1.6, 3.8)	1.9 (1.3, 2.7)*
LSM, kPa	15.3±4.5	8.7±3.7	4.5±0.6*
Log (HBV DNA)	5.8±2.2	0.0±0.0*	0.0±0.0*
Ishak score, n (%)			
1	0 (0.0)	0 (0.0)	0 (0.0)
2	0 (0.0)	0 (0.0)	2 (50.0)
3	2 (50.0)	3 (75.0)	0 (0.0)
4	1 (25.0)	0 (0.0)	1 (25.0)
5	0 (0.0)	1 (25.0)	1 (25.0)
6	1 (25.0)	0 (0.0)	0 (0.0)
PIR, n (%)		*	**
0	0 (0.0)	0 (0.0)	2 (50.0)
1	0 (0.0)	3 (75.0)	2 (50.0)
2	1 (25.0)	1 (25.0)	0 (0.0)
3	3 (75.0)	0 (0.0)	0 (0.0)

Note: continuous variables were expressed as median (interquartile range, IQR) or mean ± standard deviation (SD). Matched one-way ANOVA test followed by Tukey's multiple comparisons test or Friedman test was used to assess the changes among the matched three groups. Categorical variables were presented as counts (percentages) and compared by Chi-square trend test between any two groups. *p<0.05 and **p<0.01 (vs 0 weeks) were considered as statistically significant. ***Abbreviations***: ALB, albumin; AFP, alpha fetoprotein; ALP, alkaline phosphatase; ALT, alanine transaminase; AST, aspartate aminotransferase; GGT, glutamyl transpeptidase; HBV, hepatitis B virus; LSM, liver stiffness measurement; M, male; PLT, platelet; PIR, progressive, indeterminate, and predominately regressive; TBIL, total bilirubin; WBC, white blood cell.

Table S5. Clinical characteristics of non-regressive HBV-related liver fibrosis patients with three consecutive liver biopsies.

	Baseline (n=6)	78 weeks (n=6)	260 weeks (n=6)
Age, year	46.5±10.9	-	-
Gender	M=5	-	-
WBC, 10^9/L	5.3±1.1	4.0±0.8	5.7±1.4
PLT, 10^9L	154.2±22.1	119.8±33.1	169.5±22.6
ALT, U/L	55.5 (42.8, 57.0)	33.0 (30.0, 59.3)	23.5 (17.3, 31.3)
AST, U/L	37.5 (36.3, 41.8)	31.3 (28.4, 46.5)	23.9 (23.2, 27.0)
ALP, U/L	103.5 (75.0, 114.8)	78.5 (67.5, 89.5)	83.5 (78.8, 101.0)
GGT, U/L	28.0 (18.8, 29.8)	28.5 (27.3, 45.5)	19.5 (14.3, 21.8)
ALB, g/L	40.7±4.5	42.0±3.5	48.6±2.8*#
TBIL, μmol/L	14.4 (11.4, 22.3)	13.9 (12.3, 16.1)	18.1 (16.9, 25.8)*
AFP, ng/ml	13.2 (5.6, 15.9)	5.3 (3.2, 9.9)	3.7 (2.4, 7.2)
LSM, kPa	13.3±4.9	9.8±4.9	5.2±1.9**
Log (HBV DNA)	6.9 (6.3, 7.4)	0.0 (0.0, 1.2)	0.0 (0.0, 0.0)*
Ishak score, n (%)			
1	0 (0.0)	1 (16.7)	0 (0.0)
2	1 (16.7)	0 (0.0)	1 (16.7)
3	2 (33.3)	2 (33.3)	0 (0.0)
4	2 (33.3)	1 (16.7)	1 (16.7)
5	1 (16.7)	2 (33.3)	4 (66.6)
6	0 (0.0)	0 (0.0)	0 (0.0)
PIR, n (%)			
0	1 (16.7)	1 (16.7)	1 (16.7)
1	1 (16.7)	1 (16.7)	2 (33.3)
2	0 (25.0)	0 (25.0)	3 (50.0)
3	4 (66.6)	4 (66.6)	0 (0.0)

Note: continuous variables were expressed as median (interquartile range, IQR) or mean ± standard deviation (SD). Matched one-way ANOVA test followed by Tukey's multiple comparisons test or Friedman test was used to assess the changes among the matched three groups. Categorical variables were presented as counts (percentages) and compared by Chi-square trend test between any two groups. *p<0.05 and **p<0.01 (vs 0 weeks) and $^{\#}p$<0.05 (vs 78 weeks) were considered as statistically significant. ***Abbreviations***: ALB, albumin; AFP, alpha fetoprotein; ALP, alkaline phosphatase; ALT, alanine transaminase; AST, aspartate aminotransferase; GGT, glutamyl transpeptidase; HBV, hepatitis B virus; LSM, liver stiffness measurement; M, male; PLT, platelet; PIR, progressive, indeterminate, and predominately regressive; TBIL, total bilirubin; WBC, white blood cell.

Table S6. Liver fibrosis-related matrisome genes (LFMGs).

Gene symbol	Entrez ID	Description	Matrisome category
COL5A1↑	1289	Collagen, type V, alpha 1	Collagens
COL4A2↑	1284	Collagen, type IV, alpha 2	Collagens
COL3A1↑	1281	Collagen, type III, alpha 1	Collagens
COL1A1↑	1277	Collagen, type I, alpha 1	Collagens
COL1A2↑	1278	Collagen, type I, alpha 2	Collagens
COL4A1↑	1282	Collagen, type IV, alpha 1	Collagens
LTBP2↑	4053	Latent transforming growth factor beta binding protein 2	ECM Glycoproteins
LAMC3↑	10319	Laminin, gamma 3	ECM Glycoproteins
CTGF↑	1490	Connective tissue growth factor	ECM Glycoproteins
MFAP4↑	4239	Microfibrillar-associated protein 4	ECM Glycoproteins
SPP1↑	6696	Secreted phosphoprotein 1	ECM Glycoproteins
IGFBP7↑	3490	Insulin-like growth factor binding protein 7	ECM Glycoproteins
THBS1↑	7057	Thrombospondin 1	ECM Glycoproteins
FBLN5↑	10516	Fibulin 5	ECM Glycoproteins
DPT↑	1805	Dermatopontin	ECM Glycoproteins
AEBP1↑	165	AE binding protein 1	ECM Glycoproteins
MGP↑	4256	Matrix Gla protein	ECM Glycoproteins
THBS2↑	7058	Thrombospondin 2	ECM Glycoproteins
EFEMP1↑	2202	EGF-containing fibulin-like extracellular matrix protein 1	ECM Glycoproteins
PCOLCE2↓	26577	Procollagen C-endopeptidase enhancer 2	ECM Glycoproteins
LOXL4↑	84171	Lysyl oxidase-like 4	ECM Regulators
MMP2↑	4313	Matrix metallopeptidase 2	ECM Regulators
MMP7↑	4316	Matrix metallopeptidase 7	ECM Regulators
TIMP1↑	7076	TIMP metallopeptidase inhibitor 1	ECM Regulators
ANXA2↑	302	Annexin A2	ECM-affiliated Proteins
LGALS3↑	3958	Lectin, galactoside-binding, soluble, 3	ECM-affiliated Proteins
CLEC4M↓	10332	C-type lectin domain family 4, member M	ECM-affiliated Proteins
VCAN↑	1462	Versican	Proteoglycans
LUM↑	4060	Lumican	Proteoglycans
S100A6↑	6277	S100 calcium binding protein A6	Secreted Factors
PDGFD↑	80310	Platelet derived growth factor D	Secreted Factors
CCL19↑	6363	Chemokine (C-C motif) ligand 19	Secreted Factors
CXCL10↑	3627	Chemokine (C-X-C motif) ligand 10	Secreted Factors

CCL20↑	6364	Chemokine (C-C motif) ligand 20	Secreted Factors
CXCL6↑	6372	Chemokine (C-X-C motif) ligand 6	Secreted Factors

Note: upward arrow indicates upregulation; downward arrow indicates downregulation.

Table S7. Clinical characteristics of $LFMG^{Low}$ and $LFMG^{High}$ patients at baseline.

	$LFMG^{Low}$ (n=11)	**$LFMG^{High}$ (n=5)**	**P value**
Age, year	43.8±11.1	44±9.9	0.975
Gender	M=8	M=3	0.999
WBC, 10^9/L	176.1±57.7	148.8±24.4	0.368
PLT, 10^9L	154.2±22.1	119.8±33.1	0.336
ALT, U/L	57.0 (316.0, 29.0)	84.0 (971.0, 39.0)	0.807
AST, U/L	50.0 (228.0, 27.0)	37.0 (395.0, 30.8)	0.913
ALP, U/L	75.7±16.8	99.6±24.6	0.038*
GGT, U/L	27.0 (78.0, 15.0)	45.0 (204.0, 30.0)	0.084
ALB, g/L	42.2±3.9	39.8±5.7	0.331
TBIL, μmol/L	17.2±6.8	17.6±7.5	0.923
AFP, ng/ml	5.4 (49.1, 2.8)	11.4 (59.8, 5.8)	0.320
Log (HBV DNA)	7.4 (8.2, 5.7)	7.2 (8.2, 3.9)	0.339

Note: continuous variables were expressed as median (interquartile range, IQR) or mean ± standard deviation (SD). Categorical variables were presented as counts (percentages) and compared by Chi-square test. *p<0.05 was considered as statistically significant. ***Abbreviations***: ALB, albumin; AFP, alpha fetoprotein; ALP, alkaline phosphatase; ALT, alanine transaminase; AST, aspartate aminotransferase; GGT, glutamyl transpeptidase; HBV, hepatitis B virus; M, male; PLT, platelet; TBIL, total bilirubin; WBC, white blood cell.

Table S8. Clinical characteristics of $LFMG^{Low}$ and $LFMG^{High}$ patients after 78 weeks of treatment.

	$LFMG^{Low}$ (n=11)	$LFMG^{High}$ (n=5)	P value
Age, year	41.1±10.0	45.6±11.3	0.436
Gender	M=10	M=3	0.214
WBC, 10^9/L	4.0±1.0	4.2±1.4	0.739
PLT, 10^9L	124.0±41.8	166.6±65.9	0.135
ALT, U/L	28.0 (36.0, 18.0)	30.0 (73.0, 13.0)	0.251
AST, U/L	25.7 (33.0, 17.0)	27.0 (89.0, 13.0)	0.334
ALP, U/L	74.4±11.2	73.8±22.9	0.947
GGT, U/L	28.0 (31.0, 16.0)	36.0 (215.0, 15.0)	0.562
ALB, g/L	43.5±3.6	41.0±5.6	0.303
TBIL, μmol/L	14.2±4.2	12.6±4.9	0.531
AFP, ng/ml	4.6±2.9	5.0±4.0	0.812
Log (HBV DNA)	0.0 (1.6, 0.0)	0.0 (2.4, 0.0)	0.358

Note: continuous variables were expressed as median (interquartile range, IQR) or mean ± standard deviation (SD). Categorical variables were presented as counts (percentages) and compared by Chi-square test. ***Abbreviations***: ALB, albumin; AFP, alpha fetoprotein; ALP, alkaline phosphatase; ALT, alanine transaminase; AST, aspartate aminotransferase; GGT, glutamyl transpeptidase; HBV, hepatitis B virus; M, male; PLT, platelet; TBIL, total bilirubin; WBC, white blood cell.

Table S9. Clinical characteristics of LFMG^{Low} and LFMG^{High} patients after 260 weeks of treatment.

	LFMG^{Low} (n=18)	LFMG^{High} (n=4)	P value
Age, year	45.7±9.5	45.5±5.7	0.965
Gender	M=11	M=4	0.263
WBC, 10^9/L	5.3±1.7	5.7±1.3	0.992
PLT, 10^9L	179.0±56.4	160.0±35.9	0.222
ALT, U/L	19.8±4.9	29.8±10.7	0.047*
AST, U/L	22.7±4.5	26.0±8.3	0.069
ALP, U/L	74.8±30.2	97.0±26.1	0.080
GGT, U/L	20.0 (89.0, 10.0)	107.5 (183.0, 18.0)	0.041*
ALB, g/L	42.5±3.1	46.0±4.3	0.838
TBIL, μmol/L	12.2±2.9	17.6±1.7	0.902
AFP, ng/ml	2.7±0.9	5.4±3.4	0.058
Log (HBV DNA)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.999

Note: continuous variables were expressed as median (interquartile range, IQR) or mean ± standard deviation (SD). Categorical variables were presented as counts (percentages) and compared by Chi-square test. *p<0.05 was considered as statistically significant. ***Abbreviations***: ALB, albumin; AFP, alpha fetoprotein; ALP, alkaline phosphatase; ALT, alanine transaminase; AST, aspartate aminotransferase; GGT, glutamyl transpeptidase; HBV, hepatitis B virus; M, male; PLT, platelet; TBIL, total bilirubin; WBC, white blood cell.

Table S10. Marker genes for preliminary cell clustering and re-clustering in the scRNA-seq analysis.

Cell clusters	Marker genes										
T cells	*Cd3d*										
B cells	*Cd79a*		*Ms4a1*								
Kupffer cells	*Cd163*	*Cd68*	*Marco*	*C1qa*	*Adgre1*	*Mrc1*					
Monocytes	*Mmp9*	*Lcn2*	*Ngp*								
HSCs	*Pdgfrb*	*Pdgfra*	*Dcn*	*Acta2*	*Col3a1*						
ECs	*Vwf*	*Pecam1*	*Cdh5*	*Icam2*							
DCs	*Flt3*	*Cd209a*	*Cst3*								
Chil3$^+$ Macs	*Lyz2*	*Chil3*	*Ms4a6c*								
Fn1$^+$ Macrophages	*Fn1*	*Arg1*	*Ccl6*								
Cholangiocytes	*Ankrd1*	*Anxa5*	*Atp1b1*	*Tm4sf4*	*Anxa4*						
Hepatocytes	*Apoa1*	*Fabp1*	*Rbp4*	*Mup20*	*Hamp*						
Cd163$^+$ Kupffer cells	*Cd163*	*Marco*	*C1qa*	*C1qc*	*Clec4f*	*Timd4*					
MHC Kupffer cells	*H2-Aa*	*H2-Ab1*	*Rgs1*								
Fn1$^+$ Macrophages	*Fn1*	*Arg1*	*Ccl6*								
Chil3$^+$ Macrophages	*Lyz2*	*Chil3*	*Ms4a6c*	*Ly6c2*	*F13a1*	*Ifitm6*	*S100a4*	*Hp*			
Eno3$^+$ Macrophages	*Eno3*	*Dusp16*	*Ear2*	*Spn*	*Gngt2*	*Ace*	*Hes1*				
Cd63$^+$ Macrophages	*Cd63*	*Fabp5*	*Ms4a7*								
Vein ECs	*Pecam1*	*Vwf*	*Bgn*	*Cyt11*	*Cpe*	*Wnt9b*	*Thbd*	*Ramp3*	*Jpt1*	*Entpd1*	
Artery ECs	*Dll4*	*Efnb2*	*Plac8*	*Bmx*	*Sdc1*	*Adgrg6*	*Cavin3*	*Nrg1*	*Lmo7*		
Lymphatic ECs	*Meox1*	*Tbx1*	*Dtx1*	*Prox1*	*Pdpn*	*Mmrn1*	*Fxyd6*	*Fabp4*	*Ccl21a*	*Gng11*	
LSEC1	*Msr1*	*Ntn4*	*Acer2*	*Adam23*	*Ltbp4*	*Itga9*	*Cd36*	*Glul*			
LSEC2	*Lyve1*	*Cyp4b1*	*Ctsl*	*Aass*	*Bok*	*Flt4*	*Ccnd1*	*Clec1b*			
LSEC3	*Wnt2*	*Kit*	*Lgals1*	*Plpp1*	*Gatm*	*Plxnc1*	*Rab3b*	*Gas6*			

Lrrn3^{+} qHSCs	*Gem*	*Dcn*	*Rspo3*	*Tgfbi*	*Reln*	*Lrrn3*	*Rgs5*	*Ecm1*	*Abcc9*	*Tmem56*	*Fcna*
Ccl2^{+} qHSCs	*Lsamp*	*Rgs4*	*Gcgr*	*Ccl2*	*Clstn2*	*Nmnat2*	*Vtn*	*Disp2*			
Acta2^{+} aHSCs	*Acta2*	*Col3a1*	*Tnc*	*Timp1*	*Col15a1*	*S100a6*	*Tagln*	*Lgals1*	*Cxcl14*		
Col27a1^{+} aHSC	*Col27a1*	*Plekha6*	*Sned1*	*Plce1*	*Tnxb*	*Slc6a7*	*Fam184a*	*Coro6*	*Cyp46a1*	*Hoxd4*	*Tnnt3*
Mfap4^{+} aHSCs	*Igfbp4*	*Mfap4*	*Serpinf1*	*Lgi2*	*Dpep1*	*Fgl2*	*Fxyd6*	*Fndc5*	*Gpx3*	*Gxylt2*	