Transcriptomic Profiling of Obesity-Related Nonalcoholic Steatohepatitis Reveals a Core Set of Fibrosis-Specific Genes

Glenn S. Gerhard,[1] Christophe Legendre,[2] Christopher D. Still,[3] Xin Chu,[3] Anthony Petrick,[3] and Johanna K. DiStefano[2]

[1]Lewis Katz School of Medicine, Temple University School of Medicine, Philadelphia, Pennsylvania 19140; [2]Translational Genomics Research Institute, Phoenix, Arizona 85004; and [3]Geisinger Obesity Institute, Danville, Pennsylvania 17822

Nonalcoholic steatohepatitis (NASH) is strongly associated with obesity and type 2 diabetes. The molecular factors underlying the development of inflammation and severe fibrosis in NASH remain largely unknown. The purpose of this study was to identify gene expression patterns related to obesity-related NASH inflammation and fibrosis. We performed sequencing-based mRNA profiling analysis of liver samples from individuals with normal histology (n = 24), lobular inflammation (n = 53), or bridging fibrosis, incomplete cirrhosis, or cirrhosis (n = 65). Hepatic expression of a subset of mRNAs was validated using an orthogonal method, analyzed in a hepatic stellate cell line, and used to identify transcriptional patterns shared by other forms of cirrhosis. We observed evidence for differential levels of 3820 and 2980 transcripts in lobular inflammation and advanced fibrosis, respectively, compared with normal histology (false discovery rate ≤0.05), including 176 genes specific to fibrosis. Functional enrichment analysis of these genes revealed participation in pathways involving cytokine-cytokine receptor interaction, PI3K-Akt signaling pathway, focal adhesion, and extracellular matrix-receptor interaction. We identified 34 differentially expressed transcripts in comparisons of lobular inflammation and fibrosis, a proportion of which were also upregulated during activation of hepatic stellate cells. A set of 16 genes from a previous independent study of NASH bridging fibrosis/cirrhosis were replicated, several of which have also been associated with advanced fibrosis/cirrhosis due to hepatitis viruses or alcohol in human patients. Dysregulated mRNA expression is associated with inflammation and fibrosis in NASH. Advanced NASH fibrosis is characterized by distinct set of molecular changes that are shared with other causes of cirrhosis.

Freeform/Key Words: cirrhosis, liver biopsy, liver fibrosis, nonalcoholic steatohepatitis, obesity, RNA-sequencing

Obesity and type 2 diabetes (T2D) are strongly associated with nonalcoholic steatohepatitis (NASH) [1], a clinically severe manifestation of nonalcoholic fatty liver disease (NAFLD). NASH is characterized by substantial liver inflammation that can progress to hepatic fibrosis and cirrhosis [2]. Liver cirrhosis represents the end-stage pathology evolving from a variety of pathogenic mechanisms, including those of viral, chemical, genetic, and metabolic origins,

mediated by a chronic inflammatory state [3]. Cirrhosis contributes to a variety of chronic medical conditions and is also a major risk factor for the development of hepatocellular carcinoma (HCC) [4]. NASH is recognized as the major cause of chronic liver disease [5], and is projected to become the most common indication for liver transplantation [6]. The prevalence of NASH in the setting of extreme obesity ranges from 10% to 58%, depending on diagnostic criteria and distribution of risk factors within specific population cohorts [7].

The molecular factors underlying the development of severe fibrosis, cirrhosis, and HCC within the pathological context of NASH remain largely unknown. Further, the relationships among these different manifestations of NASH-related liver pathologies are poorly understood. Lipotoxicity, oxidative stress, cytokines, and other inflammatory molecules have been implicated in the molecular mechanisms leading to the progression of inflammation to fibrosis in NAFLD/NASH in the setting of risk factors such as obesity [8] and T2D [9, 10]. In addition, cirrhosis that develops as a result of infection with hepatitis B virus (HBV) or hepatitis C virus (HCV) or chronic alcohol abuse [4] shares morphological and histological features with NASH cirrhosis, suggesting the presence of common biological pathways at advanced stages of fibrosis. This observation is supported by evidence showing a HCC gene signature with comparable prognostic significance in patients with HBV, HCV, alcoholic steatohepatitis, or NASH [11].

A number of studies have sought to identify molecular profiles capable of discriminating among different stages on the NAFLD spectrum. For example, gene expression patterns distinguished steatohepatitis from steatosis and normal liver [12], mild fibrosis and septal fibrosis [13], low vs high levels of steatosis [14], and mild vs advanced fibrosis [15]. Most of these studies applied a microarray-based approach, were limited by small sample sizes, and did not specifically address transcriptional differences between lobular inflammation and advanced fibrosis. Delineation of transcriptional patterns specific to fibrosis may provide new insights into the molecular mechanisms that trigger progression from inflammation or prefibrotic states.

We performed sequencing-based mRNA profiling analysis of liver samples from obese individuals with normal histological findings and patients with NAFLD with either lobular inflammation or severe fibrosis, encompassing bridging fibrosis to cirrhosis. We validated hepatic expression of a subset of mRNAs using an orthogonal method, analyzed expression in a hepatic stellate cell line, and identified common transcriptional patterns shared by other forms of cirrhosis. These results provide a comprehensive gene expression profile of lobular inflammation and severe fibrosis and cirrhosis in human NASH, and nominate several key pathways and genes as targets for future studies.

1. Materials and Methods

A. Study Participants

Under the auspices of a standardized protocol, liver biopsies were obtained from white individuals enrolled in the Bariatric Surgery Program at the Geisinger Clinic Center for Nutrition and Weight Management [16]. Details of the study population can be found elsewhere [17–19]. All study participants provided written informed consent for research, which was conducted according to The Code of Ethics of the World Medical Association (Declaration of Helsinki). The institutional review boards of Geisinger Health System, Translational Genomics Research Institute, and Temple University School of Medicine approved the research protocol.

B. RNA Extraction, Sequencing, and Analysis

We extracted total RNA from liver wedge biopsies using the RNeasy Mini Kit (Qiagen Inc., Valencia, CA), and quantified products using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE). We used the Ovation RNA-Seq System V2 (NuGEN, San Carlos, CA) to prepare amplified cDNA from polyA-selected RNA prior to

sequencing; samples were analyzed using the HiSeq 2000 platform (Illumina, San Diego, CA). RNA libraries were sequenced to a depth of 60 M 83 bp paired-end reads. Sequencing data were processed using the Illumina pipeline CASAVA v1.8.4 to generate raw FASTQ reads. During the generation of qseq and FASTQ files for alignment, low quality reads were identified and removed, and indexed reads were identified and grouped accordingly. Filtered reads were aligned against the human genome using the Bowtie program [20]. Aligned RNA-Seq reads and Ensembl GRCh37-74 GTF files were imported into the HTSeq tool [21] to quantify the number of reads per gene. The counts were then used as input to DESeq2 [22] to identify transcripts showing differences in base mean levels among histological categories. We determined statistical significance with the method of Benjamini and Hochberg [23]. The level of statistical significance and degree of fold-change were used to rank transcripts showing the strongest base mean differences between phenotypic categories.

The heatmap.2 function from the gplots R package (http://CRAN.R-project.org/package=gplots) was used to plot heatmaps, which were created using $\log_{10}$-transformed normalized readcounts of the differentially expressed mRNAs obtained from the DESeq2 data analysis. Counts were normalized within DESeq2 [22] using the default normalization method (quantile). Hierarchical data clustering of normalized readcounts was performed using a combination of the Manhattan method and the Ward-linkage clustering method. These methods were performed and applied to cluster samples (represented as columns in the heatmap) and genes (rows). Clustering is shown as dendrograms in each heatmap.

C. Targeted Sequencing Using the Ion AmpliSeq Transcriptome Human Gene Expression Kit

To validate RNA-sequencing results, we analyzed gene expression with an orthogonal method utilizing the real-time PCR transcriptome panel found in the Ion AmpliSeq Transcriptome Human Gene Expression kit (Thermo Fisher Scientific). We selected a subset of 48 individuals from the discovery sample, corresponding to 16 samples from each of the three histological classes (normal, lobular inflammation, and advanced fibrosis). Total RNA was isolated and quantified as described and reverse transcription was performed using random priming. Libraries were sequenced as barcoded-pooled samples on the Ion S5XL NGS platform (Thermo Fisher Scientific).

D. Functional Enrichment Analysis

We used the DAVID Bioinformatics Resources 6.8 Beta with updated Knowledgebase (https://david-d.ncifcrf.gov/) to identify canonical signaling pathways and establish network connections between differentially expressed RNAs. The significance of the association between RNA transcripts and the canonical pathway was assessed using two criteria: (1) the ratio of the number of molecules mapping to the pathway and the total number of molecules involved in the canonical pathway and (2) the Benjamini-Hochberg corrected P-value from the right-tailed Fisher's exact test.

E. Cell Culture

We used LX-2 cells as a model for hepatic stellate cells, which are central to the fibrogenic process in the liver. Cell line authentication was performed using short tandem repeat profiling (Cell Line Genetics, Madison, WI), which confirmed the presence of a single cell line and alleles matching the known DNA fingerprint [24]. LX-2 cells (Merck Millipore, Billerica, MA) were grown in 24-well culture dishes (VWR International, Radnor, PA) containing 0.5 mL DMEM supplemented with 10% fetal bovine serum and maintained at 37°C in a Heracell 5% CO_2 incubator (Thermo Fisher Scientific). Culture medium was replaced the first day after seeding, and then every 72 hours until 80% confluence was reached. Activated LX-2 cells were put into a quiescent state by treating with DMEM, 10% fetal bovine serum, and MDI solution (0.5 mM isobutylmethylxanthine, 1 μM dexamethasone, and 167 nM insulin; Sigma-Aldrich, St. Louis, MO) as described [25]. Phenotypic differences observed with a light microscope and changes in levels of α smooth muscle actin were measured as markers of cell state. RNA was extracted from cells using the RNeasy Mini Kit

according to the recommended protocol (Qiagen Inc.). Gene expression was measured using the TaqMan RNA-to-Ct 1-Step kit (Thermo Fisher Scientific) followed by detection and analysis with the QuantStudio 6 Flex (Thermo Fisher Scientific). All TaqMan probe information is available upon request. Ribosomal protein *S18 (S18)*, actin *β (ACTB)*, and glyceraldehyde 3-phosphate dehydrogenase *(GAPDH)* were included as endogenous controls, although data were normalized only against *S18* because it was the most stable transcript. The -$\Delta\Delta$Ct method was used to estimate fold-change of gene expression between quiescent and activated cells. All assays were performed in triplicate. A two-tailed *t* test was used to determine *P*-values. A *P*-value ≤ 0.05 was considered significant.

2. Results

We applied massively parallel RNA-sequencing to investigate hepatic levels of transcripts from individuals with normal liver histology (n = 24), lobular inflammation (n = 53), and advanced fibrosis (n = 65). Patient characteristics are summarized in Supplemental Table 1. Individuals with histological findings of lobular inflammation or advanced fibrosis had higher BMIs and levels of glucose, insulin, triglycerides, and liver enzymes ($P \leq 0.01$) compared with individuals with normal liver histology. The presence of T2D was greater in patients with fibrosis compared with those with normal histology or inflammation (62% vs ~25%).

A. Transcript Levels Differentiate Normal Liver From Lobular Inflammation and Severe Fibrosis

On average, 58 million paired reads of 83 bp in length were generated for each sample, with 70% of transcripts mapping to the reference genome. Expression levels of 57,905 transcripts were calculated using the HTSeq tool. We applied unsupervised hierarchical clustering to the data in which the final clustering result was based on multiple rounds of resampling using subsets of samples. Clustering revealed strong discrimination between normal samples and those with inflammation or severe fibrosis (Supplemental Fig. 1). Only a single lobular inflammation sample was grouped with the normal histology samples, which was scored as grade 1 inflammation (<2 inflammatory foci per microscopic 200× field).

B. Genes Differentially Expressed in Lobular Inflammation vs Normal Histology

A column-wise hierarchical clustering dendrogram was generated for the normal histology and lobular inflammation samples. As described above, a single sample with a histological phenotype of grade 1 inflammation was grouped with samples showing normal liver histology (Fig. 1). The differentiation between grade 1 and 2 inflammation was less robust, although most grade 1 samples were clustered together, with a minority dispersed throughout the grade 2 inflammation groupings. This may reflect more variability in the histological classification of grade 2 vs grade 1 inflammation.

To identify genes showing differential expression between lobular inflammation and normal liver, we performed DESeq2 analysis of normalized transcript counts. In the pairwise comparison, filtering thresholds included minimum expression level (base mean ≥ 10) and $\log_2$ fold-change (± 1). Using these criteria, 3820 transcripts (Supplemental Table 2) met the statistical significance threshold measured by false discovery rate (Benjamini-Hochberg-corrected *P*-value < 0.05). The mRNAs showing the greatest fold-change and highest expression levels are shown in Table 1. Of the most dysregulated genes, few have been associated with lobular inflammation in humans, although *ATF3* [26, 27] and *CCL20* [28, 29] have been reported to play roles in alcoholic hepatitis.

C. Genes Differentially Expressed in Severe Fibrosis vs Normal Histology

Complete discrimination was seen in the hierarchical clustering heatmap comparing samples with severe fibrosis with those showing normal liver histology (Fig. 2). However, of the three
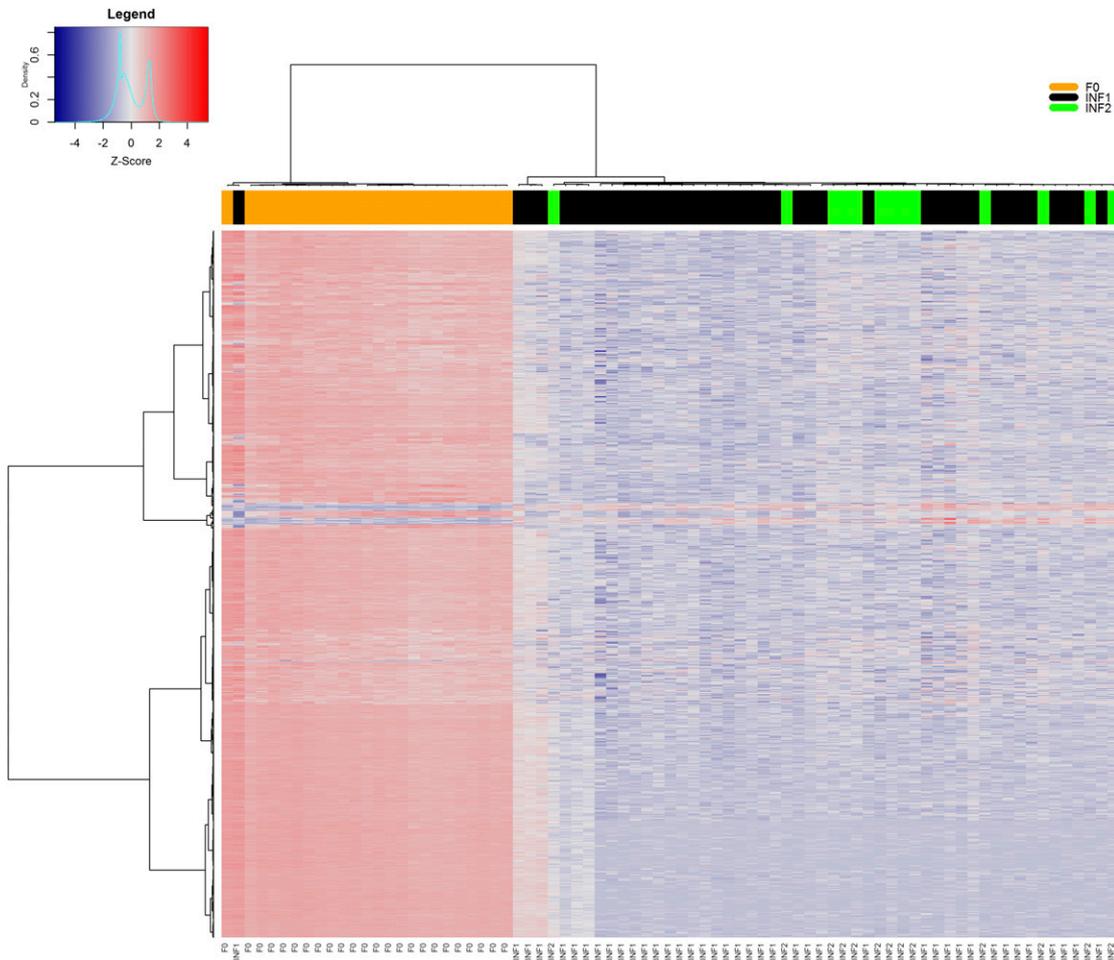
Figure 1. Patterns of gene expression differentiate lobular inflammation from normal liver. The heatmaps show differentially expressed mRNAs in patients with NAFLD with lobular inflammation compared with individuals with no histological evidence of liver damage. Samples were grouped using hierarchical clustering based on similar expression profiles. Heatmap color codes for column labels are indicated on the top right of the heatmap. The title of each label is displayed on the left side of each band. The data are represented by the z score of $\log_2$-normalized readcounts. The color-key legend is shown on the top left of each heatmap: red (*i.e., z* score > 0) indicates overexpression; white indicates no change in gene expression; blue (*i.e., z* score < 0) indicates underexpression. F0, no fibrosis; INF1, mild inflammation; INF2, moderate inflammation.

histological classes of severe fibrosis, bridging fibrosis (F3), incomplete cirrhosis (F3/4), and cirrhosis (F4), incomplete cirrhosis had the most self-clustered samples (13/27), which was surprising given the less formalized criteria for this histological classification.

Using the filtering criteria described above, we identified 2980 transcripts (Supplemental Table 3) showing evidence for differential expression between normal and fibrotic liver; the 20 genes showing the greatest differences in fold-change are shown in Table 2.

D. Levels of Differentially Expressed Transcripts Are Correlated Between RNA-Sequencing and an Orthogonal Platform

We found that $>97\%$ of differentially expressed mRNAs in both lobular inflammation and severe fibrosis were downregulated relative to normal tissue. To validate this distribution, we selected a randomly selected subset of samples (16 samples corresponding to each histological class) for evaluation with an orthogonal method using a real-time PCR transcriptome panel containing over 20,000 mRNAs. In comparisons with normal tissue, we observed statistically

Table 1. Transcripts Showing the Strongest Differences in Hepatic Expression Between Lobular Inflammation and Normal Tissue

Gene ID	Description	Discovery Data Set[a]			Validation Data Set[b]		
		BM[c]	Log$_2$FC[d]	Q Value[e]	BM[c]	Log$_2$FC[d]	Q Value[e]
ASPHD1	Aspartate β-hydroxylase domain containing 1	15.3	−4.83	1.87E-82	6.2	−1.91	3.84E-07
ATF3	Activating transcription factor 3	619.8	1.13	4.26E-04	839.4	1.34	1.87E-02
BTG2	BTG antiproliferation factor 2	436.8	1.12	1.20E-06	348.3	1.25	3.97E-03
C2orf53	Chromosome 2 open reading frame 53	45.2	−4.63	7.53E-18	1.4	−5.45	2.86E-06
CCER1	Coiled-coil glutamate rich protein 1	32.8	−4.29	8.52E-16	51.9	−7.23	2.89E-59
CCL20	C-C motif chemokine ligand 20	40.5	1.34	1.07E-03	27.0	1.45	2.45E-02
COX6B2	Cytochrome C oxidase subunit 6B2	18.8	−4.49	3.43E-31	27.4	−6.08	2.38E-45
CYP7A1	Cytochrome P450 family 7 subfamily A member 1	1072.1	1.89	1.39E-08	271.2	1.74	9.08E-03
FCGR3B	Fc fragment of IgG receptor IIIb	212.6	1.23	2.05E-04	148.7	1.17	1.54E-02
GATA1	GATA binding protein 1	21.0	−4.41	1.97E-35	36.4	−3.84	1.98E-59
GBP1	Guanylate binding protein 1	1659.4	1.22	2.04E-07	463.5	0.70	3.12E-02
GRID2IP	Glutamate receptor, ionotropic, delta 2 (Grid2) interacting protein 1	13.3	−4.57	8.14E-46	9.2	−3.69	3.70E-25
IBTK	Inhibitor of bruton tyrosine kinase	2665.5	1.14	7.25E-32	481.3	0.44	2.85E-03
IFIT3	Interferon induced protein with tetratricopeptide repeats 3	390.3	1.14	1.08E-08	179.9	0.89	4.84E-03
KRTAP10-10	Keratin associated protein 10-10	11.3	−4.54	2.93E-24	45.7	−6.83	1.49E-62
MXD1	MAX dimerization protein 1	1073.7	1.24	1.66E-07	186.3	1.17	1.20E-02
NABP1	Nucleic acid binding protein 1	1101.2	1.23	2.45E-15	114.7	0.77	5.25E-03
NCR2	Natural cytotoxicity triggering receptor 2	14.0	−4.60	1.15E-23	10.3	−6.48	1.30E-18
PKP3	Plakophilin 3	15.5	−4.97	1.46E-46	14.5	−4.31	4.93E-32
RASD1	Ras related dexamethasone induced 1	261.6	1.41	1.68E-07	270.0	1.18	1.03E-02

[a]Expression profiles derived from Illumina RNA-Seq.
[b]Expression profiles derived from AmpliSeq Transcriptome Human Gene Expression kit.
[c]Base mean: mean of normalized counts of all samples.
[d]Log$_2$ fold-change.
[e]Benjamini-Hochberg-corrected P value.

significant (false discovery rate <0.05) evidence for differential expression of 359 (Supplemental Table 4) and 372 (Supplemental Table 5) mRNAs in inflammation and fibrosis compared with normal liver samples, respectively, using the same filtering criteria described above [base mean (≥10) and log$_2$ fold-change (±1)]. As shown in Tables 1 and 2, the majority of differentially expressed genes identified in the RNA-sequencing analysis (discovery data set) were differentially expressed in the AmpliSeq analysis (validation data set).

E. Upregulated Transcripts Specific to Advanced Fibrosis

We next sought to determine whether any genes were uniquely activated in fibrotic samples, relative to normal histology, as opposed to being dysregulated during the progression from inflammation to fibrosis. We first investigated transcriptomic profiles derived from comparisons of inflammation (n = 3820) and normal liver histology, and advanced fibrosis (n = 2980) and normal liver histology. We found 2804 transcripts shared between the two groups, consistent with results from the combined inflammation-fibrosis analysis. However, 176 genes showed differential expression levels in fibrosis, but not in the inflammation samples (Supplemental Table 6).

F. Pathways Dysregulated in Advanced Fibrosis

Pathway analysis of these 176 genes revealed 16 pathways considered to be upregulated and six downregulated pathways (Supplemental Table 7). The top upregulated pathways
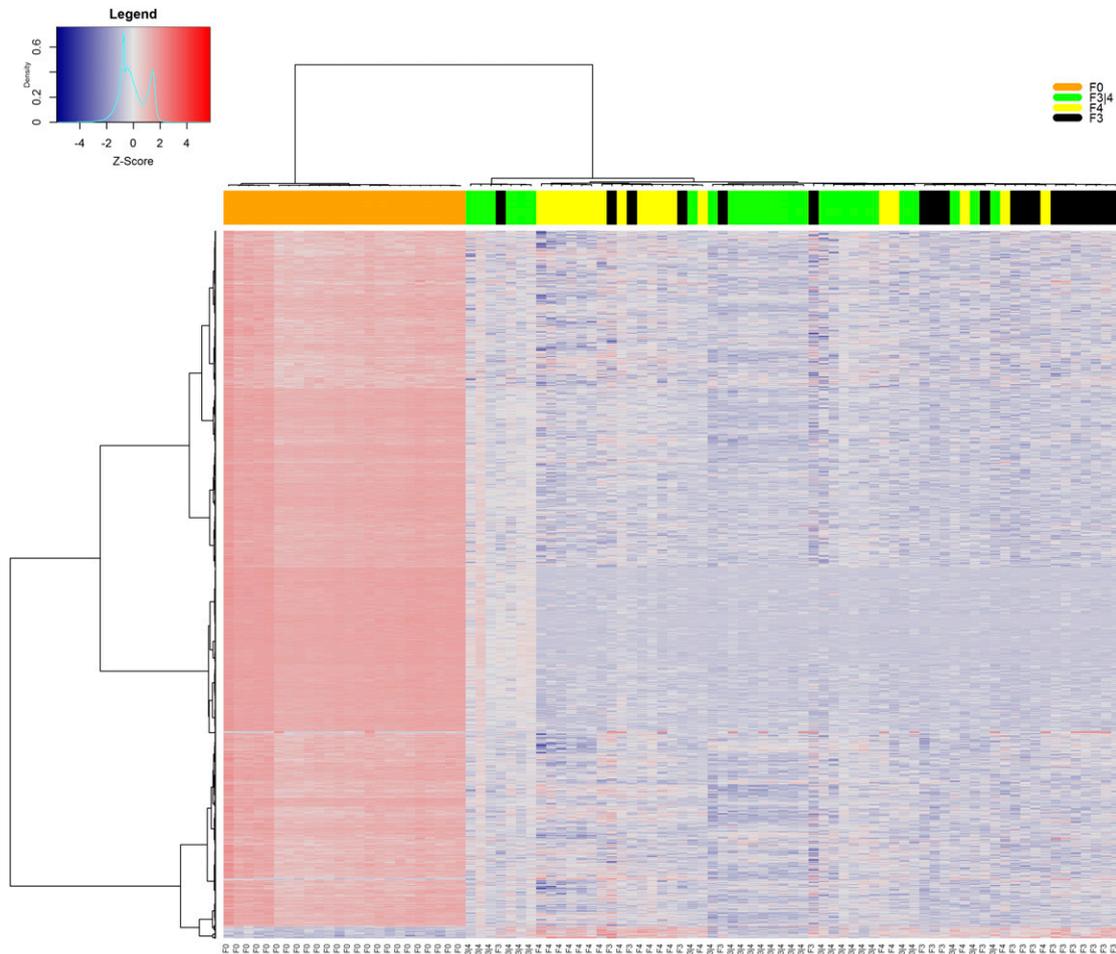
Figure 2. Patterns of gene expression differentiate fibrosis from normal liver. The heatmaps show differentially expressed mRNAs in patients with NAFLD with advanced fibrosis compared with individuals with no histological evidence of liver damage. Samples were grouped using hierarchical clustering based on similar expression profiles. Heatmap color codes for column labels are indicated on the top right of the heatmap. The title of each label is displayed on the left side of each band. The data are represented by the z score of $\log_2$-normalized readcounts. The color-key legend is shown on the top left of each heatmap: red (*i.e.*, z score > 0) indicates overexpression; white indicates no change in gene expression; blue (*i.e.*, z score < 0) indicates underexpression.

included cytokine–cytokine receptor interaction, PI3K-Akt signaling pathway, focal adhesion, extracellular matrix (ECM)-receptor interaction, and amoebiasis.

G. Genes Differentially Expressed in Advanced Fibrosis vs Lobular Inflammation

Although the preceding analysis assessed overlap between differentially expressed genes identified through comparisons of inflammation and fibrotic tissue with normal tissue, these results do not contain information about gene expression differences resulting from comparisons of lobular inflammation with advanced fibrosis. Therefore, we generated a clustering dendrogram comparing these two sample groups (Fig. 3). In contrast to comparisons with normal histology, differentiation of lobular inflammation from severe fibrosis by hierarchical clustering was less discriminatory (Fig. 3). Samples from all subgroups, *i.e.*, lobular inflammation scores of 1 and 2 and bridging fibrosis, incomplete cirrhosis, and cirrhosis were interspersed. Because samples histologically classified as severe fibrosis also manifest various degrees of lobular inflammation, a blurring of discriminatory power in

Table 2. Transcripts Showing the Strongest Differences in Hepatic Expression Between Advanced Fibrosis and Normal Tissue

Gene ID	Description	Discovery Data Set[a]			Validation Data Set[b]		
		BM[c]	Log$_2$FC[d]	Q Value[e]	BM[c]	Log$_2$FC[d]	Q Valu[e]
AKR1B10	Aldo-keto reductase family 1 member B10	182.6	2.49	2.06E-06	125.46	3.61	1.08E-05
C2orf53	Chromosome 2 open reading frame 53	37.1	−7.27	6.63E-57	205.6	−8.33	5.80E-140
CASP14	Caspase 14	56.2	−6.41	3.49E-27	3.8	−3.31	7.06E-10
CCER1	Coiled-coil glutamate-rich protein 1	26.4	−7.07	3.15E-42	17.9	−7.13	3.43E-26
CCL20	C-C motif chemokine ligand 20	65.1	2.14	3.67E-07	184.92	2.39	2.82E-23
COL1A1	Collagen type I α 1 chain	804.5	1.57	4.11E-11	1403.0	1.61	6.57E-05
CYP7A1	Cytochrome P450 family 7 subfamily A member 1	1041.3	1.97	2.01E-12	212.56	2.31	1.37E-08
GOLGA6L1	Golgin A6 family-like 1	20.2	−7.61	2.12E-126	53.4	−9.33	1.17E-32
IGFN1	Immunoglobulin-like and fibronectin type III domain containing 1	168.5	−4.80	6.19E-47	289.4	−1.24	7.64E-03
IL36RN	Interleukin 36 receptor antagonist	14.5	−7.22	9.87E-92	54.0	−7.65	1.15E-58
IL7R	Interleukin 7 receptor	498.1	1.39	9.08E-10	3.9	1.30	3.51E-02
ITGBL1	Integrin subunit β like 1	429.9	1.75	1.23E-12	42.66	2.19	4.25E-05
KLF17	Kruppel like factor 17	13.9	−6.78	1.51E-71	1.4	−5.21	1.27E-06
LUM	Lumican	842.0	1.50	7.83E-16	635.88	1.56	2.43E-12
MUC16	Mucin 16	610.2	−5.26	4.07E-34	0.8	−2.53	1.28E-02
NCAN	Neurocan	25.0	−5.39	1.12E-73	0.4	−2.89	2.18E-02
PLA2G2A	Phospholipase A2 group IIA	1662.8	1.61	3.44E-08	159.15	1.62	1.06E-08
RASD1	Dexamethasone-induced Ras-related protein 1	302.1	1.52	2.82E-06	259.1	1.59	4.07E-03
SOX11	SRY-box 11	26.3	−6.24	5.59E-43	1.0	−3.17	4.94E-04
SPP1	Secreted phosphoprotein 1	201.0	1.61	1.14E-09	869.33	2.04	2.87E-11

[a]Expression profiles based on Illumina RNA-Seq.
[b]Expression profiles based on AmpliSeq Transcriptome Human Gene Expression kit.
[c]Base mean: mean of normalized counts of all samples.
[d]Log$_2$ fold-change.
[e]Benjamini-Hochberg-corrected P value.

hierarchical clustering is not surprising. This is analogous to data reported for HCC, where more robust and stable transcriptional profiling was found using data from a separate group of normal liver samples as controls, rather than cancer plus adjacent noncancer paired liver samples from the same individual [30].

We compared base mean levels between samples with lobular inflammation and those with advanced fibrosis, and identified 34 transcripts showing statistically significant differences in expression levels (Table 3), the majority of which were upregulated with disease severity. Differences in expression levels of several of these genes, including aldo-keto reductase family 1 member B10 (*AKR1B10*), C-C motif chemokine ligand 19 (*CCL19*), and stathmin 2 (*STMN2*), are consistent with findings reported in microarray and quantitative RT-PCR studies [12, 31].

We then plotted normalized counts from the validation data set corresponding to histological status for each of the 34 genes (Supplemental Fig. 2). We identified 21 genes upregulated only in fibrosis, whereas expression levels between inflammation and normal tissue were not significantly different from each other. These results were consistent with findings from the independent comparisons of fibrotic liver and normal liver vs inflammation and normal liver.

H. Genes Involved in Hepatic Fibrosis Are Upregulated in Activated LX-2 Cells

Hepatic stellate cells have been primarily characterized as the main effector cells in liver fibrosis, due to their capacity to transdifferentiate into collagen-producing myofibroblasts
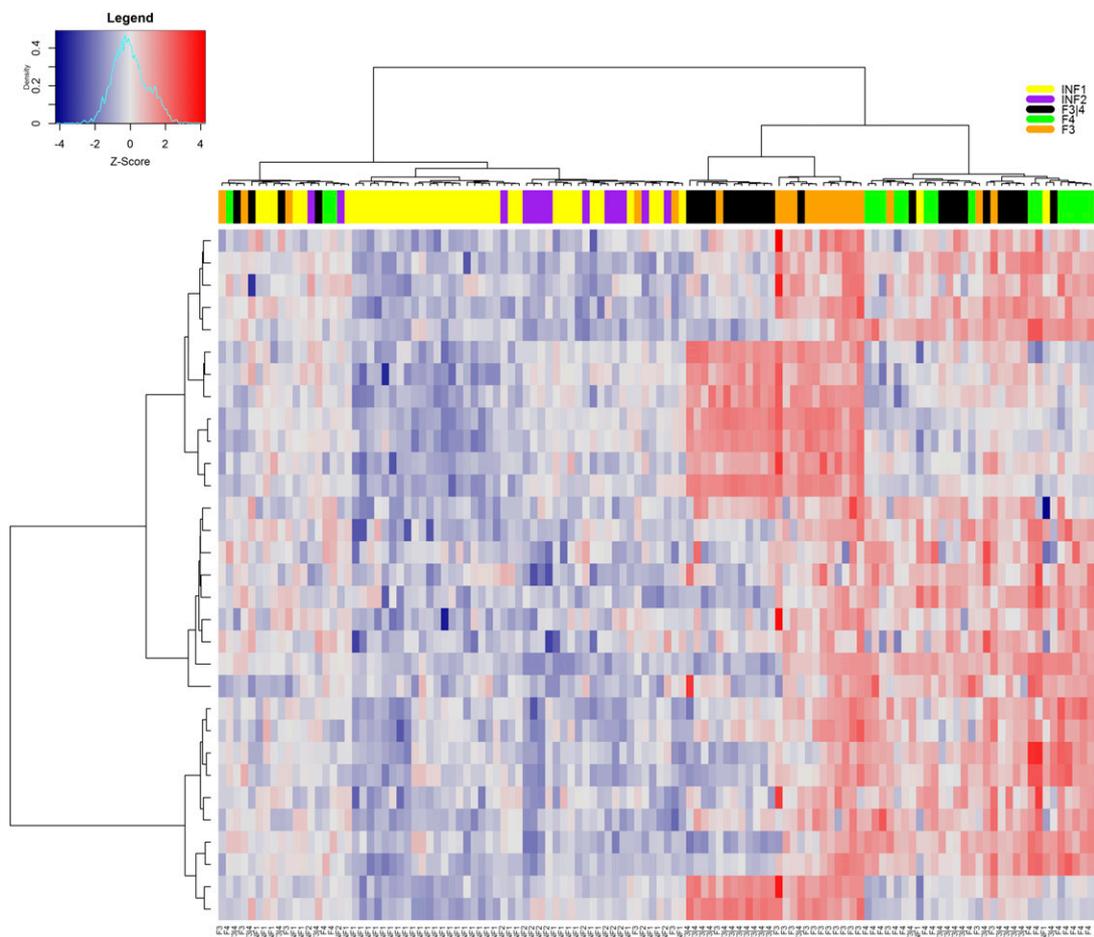
Figure 3. Clustering dendrogram of differentially expressed genes between lobular inflammation and fibrosis. The heatmaps show differentially expressed mRNAs in patients with NAFLD with advanced fibrosis compared with individuals with lobular inflammation. Samples were grouped using hierarchical clustering based on similar expression profiles. Heatmap color codes for column labels are indicated on the top right of the heatmap. The title of each label is displayed on the left side of each band. The data are represented by the z score of $\log_2$-normalized readcounts. The color-key legend is shown on the top left of each heatmap: red (*i.e., z* score > 0) indicates overexpression; white indicates no change in gene expression; blue (*i.e., z* score < 0) indicates underexpression.

[32–34]. The LX-2 cell line is an established model for hepatic stellate cells, retaining key features of cytokine signaling, neuronal gene expression, retinoid metabolism, and fibrogenesis [35]. To determine whether the 34 genes uniquely upregulated in fibrosis were relevant to fibrogenesis in this model, we measured expression levels of these transcripts in quiescent and activated LX-2 cell states. As shown in Table 3, levels of *AKR1B10,* coiled-coil domain containing 80 (*CCDC80*), matrix gla protein (*MGP*), matrix metallopeptidase 7 (*MMP7*), neurotensin (*NTS*), ribosomal protein S28 (*RPS28*), superoxide dismutase 3 (*SOD3*), secreted phosphoprotein 1 (*SPP1*), and ubiquinol-cytochrome C reductase (*UQCR11*) were elevated in activated LX-2 cells compared with those in a quiescent state.

3. Discussion

The high prevalence of NAFLD and NASH, and the corresponding risks for severe fibrosis and cirrhosis, underlie widespread efforts to identify targets for the development of

therapeutic agents. We applied the unbiased approach of RNA sequencing to profile liver transcriptomes of human NASH patients. A large collection of well annotated liver biopsies in obese patients not ascertained for definite or suspected underlying liver disease was used, allowing inclusion of large numbers of samples with normal liver histology and bridging fibrosis, incomplete cirrhosis, or cirrhosis. These results represent a large study to define the transcriptomic profile of severe fibrosis due to NASH in humans.

Unsupervised hierarchical clustering of the RNA-sequencing data distinguished between normal histology and lobular inflammation and normal histology and bridging fibrosis/ cirrhosis, but was less discriminatory in classifying lobular inflammation from severe fibrosis. This may be due to the potentially heterogeneous distribution of inflammatory cell infiltration throughout the liver. A large wedge biopsy that is partitioned for formalin fixation and histological examination may manifest a different level of inflammation than the corresponding portion preserved for RNA sequencing. In addition, the objective and sensitive nature of RNA-sequencing vs the histological classification of lobular inflammation may reveal different levels of sensitivity. RNA-sequencing measures gene expression at the level of the intact tissue, in which the contribution of infiltrating inflammatory cells may be diluted by otherwise normal gene expression from hepatocytes and other liver cell types. Morphological classification may thus be based on relatively few inflammatory cells. In addition, expression of proinflammatory genes may be attenuated in some individuals for reasons not yet known. The loss of discriminatory power may also be due to the coincidence of inflammation with severe fibrosis.

We identified 21 genes upregulated only in fibrosis, with their expression levels between inflammation and normal tissue were not significantly different from each other. These results are consistent with findings from the independent comparisons of fibrotic liver and normal liver vs inflammation and normal liver. Some of these genes, including AE binding protein 1(*AEBP1*), dermatopontin (*DPT*), fibroblast activation protein α (FAP), integrin subunit β like 1 (*ITGBL1*), S100 calcium-binding protein A6 (*S100A6*), *SPP1*, and *STMN2*, have established roles in biological processes relevant to fibrosis (*e.g.*, ECM maintenance, wound healing, or cytoskeletal function) or have been previously associated with NASH fibrosis in mouse models of the disease [36–43].

Activation of hepatic stellate cells into myofibroblastlike cells is a central event underlying the development of liver fibrosis. Immortalized LX-2 cells represent a well-characterized model for studies of human hepatic fibrosis and retain key features of cytokine signaling, neuronal gene expression, retinoid metabolism, and fibrogenesis [35]. Of the 34 genes measured in LX-2 cells, nine showed upregulation in the transition from quiescence to activation, matching trends seen in comparisons of lobular inflammation and advanced fibrosis in NASH patients. Of interest, *AKR1B10* levels were increased ~10-fold in NASH patients compared with healthy individuals [44], and in patients with HCV infection [45] or HCC [46]. Likewise, levels of *MMP7*, a matrix metallopeptidase family member involved with breakdown of ECM during normal physiological and pathophysiological processes, were elevated in patients with idiopathic pulmonary fibrosis [47] and biliary atresia-associated fibrosis [48]. Finally, SPP1 expression was elevated not only in alcoholic liver disease, but also in LX-2 cells following acute alcohol exposure [49]. These results suggest that some of these transcripts may be altered during the transition to a myofibroblastlike phenotype and may contribute to fibrogenic features of this activated state. Differences between liver tissue and LX-2 cell activation for the remaining 23 transcripts may be a reflection of the differences inherent between whole-organ and cell-specific analyses. In sequencing analyses of RNA extracted from liver wedge biopsies, the cellular origin of transcripts is not known, which limits interpretation of RNA-sequencing data. Future studies using single-cell sequencing of hepatic stellate cells, hepatocytes, and other cells of the liver will be important to identify cell type-specific contributions to biological changes that occur during the development of fibrosis and cirrhosis in NASH.

A comprehensive study previously used microarray technology to profile gene expression in human NASH fibrosis [15], although there are several key differences between that work and the present findings. First, the previous study analyzed liver biopsy material obtained

Table 3. Transcripts Showing Differential Levels in Comparisons of Lobular Inflammation and Advanced Fibrosis Histological Samples

Gene ID	Description	Discovery Data Set[a]			Validation Data Set[b]			LX-2 Analysis[c]	
		BM[d]	Log$_2$FC[e]	Q Value[f]	BM[d]	Log$_2$FC[e]	Q Value[f]	Log$_2$FC[e]	P Value[g]
AEBP1	AE binding protein 1	344.3	1.17	3.58E-15	287.0	1.28	4.01E-08	0.08	9.60E-05
AKR1B10	Aldo-keto reductase family 1 member B10	137.0	1.57	6.42E-13	133.9	0.80	4.26E-03	3.11	2.46E-02
AQP1	Aquaporin 1	104.4	1.03	1.38E-15	433.4	1.08	1.52E-05	0.57	1.17E-02
ATP5I	Mitochondrial membrane ATP synthase	90.5	1.03	8.90E-11	1138.7	−0.36	8.85E-03	0.83	3.83E-03
CCDC80	Coiled-coil domain containing 80	153.2	1.03	5.24E-11	6.3	0.73	2.77E-02	2.37	1.42E-04
CCL19	C-C motif chemokine ligand 19	49.6	1.34	2.85E-12	306.8	1.03	6.43E-04	ND	ND
CD52	Cluster of differentiation 52	38.9	1.01	1.26E-10	256.2	0.76	2.27E-03	0.15	4.41E-05
DPT	Dermatopontin	186.5	1.02	4.66E-13	211.1	1.10	2.33E-06	ND	ND
EFEMP1	EGF-containing fibulin-like ECM protein 1	148.6	1.19	6.42E-13	126.5	1.47	4.74E-09	0.67	6.41E-05
FAM173A	Family with sequence similarity 173 member A	5.7	1.04	3.79E-10	121.8	0.15	3.43E-01	0.75	1.82E-03
FAP	Fibroblast activation protein α	18.2	1.00	3.23E-07	16.1	1.27	1.18E-05	0.10	3.59E-03
H1FX	H1 histone family member X	19.2	1.03	2.92E-10	684.7	−0.08	6.93E-01	0.56	1.82E-05
HOXB2	Homeobox B2	11.7	1.14	1.48E-17	6.1	0.54	1.22E-01	0.42	1.86E-05
ITGBL1	Integrin subunit β like 1	386.6	1.05	7.48E-10	196.5	1.78	1.88E-16	0.33	3.09E-05
LAMC3	Laminin subunit γ 3	31.5	1.27	1.38E-15	130.4	1.54	1.24E-11	ND	ND
LTB	Lymphotoxin β	24.1	1.01	4.63E-09	77.7	0.71	1.43E-02	ND	ND
MGP	Matrix gla protein	168.1	1.10	1.31E-14	0.3	0.21	NA	1.97	7.51E-04
MMP7	Matrix metallopeptidase 7	18.4	1.16	6.28E-10	74.3	1.07	3.97E-04	7.51	5.45E-06
MOXD1	Monooxygenase DBH like 1	55.9	1.29	2.24E-13	48.6	1.58	1.66E-09	0.17	6.65E-08
MZB1	Marginal zone B and B1 cell specific protein	16.1	1.06	3.65E-10	62.9	0.77	6.49E-03	ND	ND
NDUFA11	Ubiquinone oxidoreductase subunit A11	47.3	1.18	7.93E-14	1189.9	0.03	8.89E-01	0.76	1.84E-04
NME3	Nucleoside diphosphate kinase 3	26.5	1.16	2.49E-12	144.4	−0.11	6.36E-01	1.20	5.11E-03
NTS	Neurotensin	17.6	1.07	1.19E-07	7.7	0.78	1.66E-02	18.99	8.23E-07
PDZK1IP1	PDZK1-interacting protein 1	5.6	1.08	7.92E-08	23.1	1.21	1.66E-05	ND	ND
PTGDS	Prostaglandin D2 synthase	69.9	1.26	1.06E-13	220.2	1.13	1.23E-05	ND	ND
RPS28	Ribosomal protein S28	215.0	1.07	5.66E-13	5699.7	0.03	9.09E-01	1.10	1.29E-01
S100A6	S100 calcium-binding protein A6	57.9	1.31	6.37E-16	108.9	0.33	3.70E-01	0.26	4.62E-06

(*Continued*)

Table 3. Transcripts Showing Differential Levels in Comparisons of Lobular Inflammation and Advanced Fibrosis Histological Samples (Continued)

Gene ID	Description	Discovery Data Set[a]			Validation Data Set[b]			LX-2 Analysis[c]	
		BM[d]	Log$_2$FC[e]	Q Value[f]	BM[d]	Log$_2$FC[e]	Q Value[f]	Log$_2$FC[e]	P Value[g]
SLC52A2	Solute carrier family 52 member 2	7.6	1.03	1.73E-09	54.0	0.03	9.24E-01	0.86	2.55E-02
SOD3	Superoxide dismutase 3	10.5	1.03	8.51E-09	109.4	1.33	3.96E-09	1.43	7.37E-05
SPP1	Secreted phosphoprotein 1	177.6	1.06	6.52E-09	236.4	0.98	1.36E-03	3.88	3.64E-06
STMN2	Stathmin 2	30.3	1.47	9.72E-12	37.3	1.66	1.23E-09	N/A	N/A
TMSB10	Thymosin β 10	331.4	1.11	1.48E-17	4015.6	0.40	4.96E-02	0.92	1.81E-01
TSPO	Translocator protein	32.0	1.15	6.93E-12	97.5	0.43	1.06E-01	0.50	2.45E-06
UQCR11	Ubiquinol-cytochrome C reductase	90.4	1.03	4.24E-12	1763.6	−0.16	1.63E-01	1.06	2.54E-01

Abbreviations: N/A, not applicable; ND, not detected.
[a]Expression profiles based on Illumina RNA-Seq.
[b]Expression profiles based on AmpliSeq Transcriptome Human Gene Expression kit.
[c]Quantitative PCR analysis comparing quiescent and activated LX-2 cells.
[d]Base mean: mean of normalized counts of all samples.
[e]Log$_2$ fold-change.
[f]Benjamini-Hochberg-corrected P value.
[g]Student t test.

with needle biopsies, upon which histologic analysis was based on more than five portal tracts and 10 mm of tissue. In contrast, the current study used much larger wedge biopsies, providing a better representation of liver parenchyma and minimizing sample-to-sample variability. Second, in the microarray study, patients without fibrosis were grouped with those classified as fibrosis stage 1a, 1b, and 1c, potentially masking expression patterns specific to mild fibrosis. Third, although the severe NAFLD group in the earlier study was similar to the fibrosis cohort described here, the degree of obesity was lower and the percentage of males higher than in the present work. Despite these differences in design, 16 genes showing differential expression in bridging fibrosis and cirrhosis were common between the two studies: complement C7 (*C7*), CD24 molecule (*CD24*), chitinase 3 like 1 (*CHI3L1*), collagen type 1, α 1 chain (*COL1A1*), collagen type 1, α 2 chain (*COL1A2*), collagen type III, α 1 chain (*COL3A1*), doublecortin domain containing 2 (*DCDC2*), EGF containing fibulin like ECM protein 1 (*EFEMP1*), EPH receptor A3 (*EPHA3*), fibulin 5 (*FBLN5*), GTP binding protein overexpressed in skeletal muscle (*GEM*), *ITGBL1*, laminin subunit α 2 (*LAMA2*), lumican (*LUM*), *SPP1*, and versican (*VCAN*).

A number of these genes have been previously implicated in human cirrhosis. For example, *COL1A1* encodes a widely expressed matrix protein that contributes to fibrosis across a range of cirrhosis etiologies [50], whereas upregulation of *COL1A2* in hepatic stellate cells has been implicated in ethanol-induced liver fibrosis [51]. *COL3A1* expression has been associated with HCV and NASH liver fibrosis [52] and *ITGBL1* interacts with TGFB1 (transforming growth factor β 1) to regulate progression of fibrosis in patients with HBV [42]. Mutations in *DCDC2* give rise to neonatal sclerosing cholangitis [53]. Plasma levels of SPP1 have been associated with higher risk of liver fibrosis in NASH [54], alcoholic liver disease [55], chronic HBV infection [56], and chronic HCV infection [57], whereas blood levels of CHI3L1 were able to distinguish between early and advanced stages of liver fibrosis, including cirrhosis, in patients with HBV [58]. Expression of LUM, a proteoglycan that regulates collagen fibril assembly, was increased with NASH severity [59] and correlated with the fractional synthesis rate of collagen in NASH fibrosis [60]. LUM expression was also associated with hepatic fibrosis in both HBV and HCV infection [61]. In addition, a six-gene signature, consisting of *CD24*, *CXCL6*, *EHF*, *ITGBL1*, *LUM*, and *SOX9*, was found to be predictive for cirrhosis risk in

patients with chronic HBV infection [62], suggesting that *CD24*, *ITGBL1*, and *LUM* may be common features of cirrhosis across highly divergent proximate mechanisms.

Pathway analysis identified several aspects of ECM that were dysregulated in advanced fibrosis. NASH fibrosis is considered to be a dynamic process resulting from both qualitatively and quantitatively abnormal ECM deposition [63]. Maintenance of ECM in the liver depends upon the interaction of hepatic stellate cells with hepatocytes, sinusoidal endothelial cells, and immune cells that together regulate the deposition and resorption of ECM. Pathway analysis also indicated that *N*-linked glycosylation and glycoproteins were dysregulated in fibrosis. A number of carbohydrates, including Gal, Galb1-4GlcNAc, 3a, 4b, 3a-Galactotetraose, and GalNAc, are increased with activation of LX-2 cells [64], consistent with these findings.

Efforts to identify molecular pathways leading to fibrosis shared among different organs have been reported [65]. In a comparison of pathways shared between lung and liver fibrosis, only "Cell adhesion_Chemokines and adhesion" overlapped with our results. That aspects of ECM were not found in the dual organ analysis suggests there may be important differences in experimental design that obfuscate the identification of ECM pathways. However, pathway analysis of the core set of 16 NASH advanced fibrosis/cirrhosis genes shared by our study and those of Moylan *et al.* [15] revealed ECM-receptor interaction, focal adhesion, and PI3K-Akt signaling pathway (Supplemental Table 8), further narrowing promising target pathways for therapeutic intervention. That these 16 genes and three pathways were replicated in two completely independent studies of NASH advanced fibrosis/cirrhosis in human patients adds strong evidence of reliability and reproducibility of the findings, and indicates that these candidates hold relevance for human NASH.

Compared with microarray hybridization, RNA-sequencing provides a broader dynamic range through quantification of discrete, digital sequencing read counts, and higher levels of specificity and sensitivity, which enhance detection of differential expression [66]. Our RNA-sequencing results showed a surprising preponderance of downregulated transcripts in inflammation and advanced fibrosis relative to normal histology, prompting the use of an orthogonal PCR-based method for verification. A similar trend was reported for transcriptomic profiling of miRNAs in human NASH and cirrhosis, in which data were generated using RNA-sequencing of liver wedge biopsies [67]. Essentially all miRNA transcripts were downregulated in the progression from normal and fatty liver to steatohepatitis and cirrhosis. In contrast, a study of mRNAs measured using microarray technology in human fatty liver (but not NASH or fibrosis) reported an approximately equal distribution between increased and decreased expression, although a large majority of long noncoding RNAs were downregulated [68]. Another study of miRNA in human NAFLD that did not find significant downregulation also used microarray technology with only ~25% of miRNAs detected [69]. It is unclear from the data of Moylan *et al.* [15] what the overall distribution was for their microarray-based study. We suspect that the greater sensitivity of RNA sequencing, as well as differences in sample collection or preparation, may account for the different distributions. The correlation between RNA-sequencing and AmpliSeq data were substantial but not high, again likely reflecting differences in sensitivity and specificity between the two methodologies.

We acknowledge some limitations of the current study. An overall disadvantage of studies conducted in human populations is the uncontrolled environment, behavioral and genetic heterogeneity, and other confounding variables. Such confounding is often addressed through the use of animal models to approximate the human disease. However, a recent study comparing hepatic gene expression in nine NAFLD mouse models with transcriptional changes in human liver biopsies found little evidence of overlap between species [70]. Even a diet-induced obese mouse model purporting a high level of fidelity to human NASH fibrosis was characterized by increased serum high-density lipoprotein cholesterol levels and decreased triglyceride levels [71], raising concerns about the potential contribution of these starkly different metabolic effects on physiological phenotype.

Because the study design was cross-sectional, inferences about progression cannot be made directly and must therefore be interpreted with caution. Longitudinally obtained

biopsies from a bariatric surgery population are also complicated by metabolic effects resulting from the surgical procedure and the high likelihood of significant weight loss, along with dietary changes, that result in amelioration of disease, and in some cases, even reversal of hepatic fibrosis. Nevertheless, a paired biopsy approach has been reported [72], although only one patient with cirrhosis was analyzed.

In summary, the results obtained in the current study demonstrate that advanced NASH fibrosis is characterized by a distinct set of molecular changes that are shared with other causes of cirrhosis. Future investigations, including validation in independent cohorts and functional characterization of dysregulated genes and pathways, will be important to extend these findings.

Acknowledgments

References and Notes

1. Tilg H, Moschen AR, Roden M. NAFLD and diabetes mellitus. *Nat Rev Gastroenterol Hepatol*. 2017;**14**(1):32–42.
2. Neuschwander-Tetri BA, Caldwell SH. Nonalcoholic steatohepatitis: summary of an AASLD Single Topic Conference. *Hepatology*. 2003;**37**(5):1202–1219.
3. Altamirano-Barrera A, Barranco-Fragoso B, Méndez-Sánchez N. Management strategies for liver fibrosis. *Ann Hepatol*. 2017;**16**(1):48–56.
4. Ghouri YA, Mian I, Rowe JH. Review of hepatocellular carcinoma: epidemiology, etiology, and carcinogenesis. *J Carcinog*. 2017;**16**(1):1.
5. Starley BQ, Calcagno CJ, Harrison SA. Nonalcoholic fatty liver disease and hepatocellular carcinoma: a weighty connection. *Hepatology*. 2010;**51**(5):1820–1832.
6. Wong RJ, Cheung R, Ahmed A. Nonalcoholic steatohepatitis is the most rapidly growing indication for liver transplantation in patients with hepatocellular carcinoma in the U.S. *Hepatology*. 2014;**59**(6):2188–2195.
7. Machado M, Marques-Vidal P, Cortez-Pinto H. Hepatic histology in obese patients undergoing bariatric surgery. *J Hepatol*. 2006;**45**(4):600–606.
8. Wong VW, Wong GL, Choi PC, Chan AW, Li MK, Chan HY, Chim AM, Yu J, Sung JJ, Chan HL. Disease progression of non-alcoholic fatty liver disease: a prospective study with paired liver biopsies at 3 years. *Gut*. 2010;**59**(7):969–974.
9. Campbell PT, Newton CC, Patel AV, Jacobs EJ, Gapstur SM. Diabetes and cause-specific mortality in a prospective cohort of one million U.S. adults. *Diabetes Care*. 2012;**35**(9):1835–1844.
10. Porepa L, Ray JG, Sanchez-Romeu P, Booth GL. Newly diagnosed diabetes mellitus as a risk factor for serious liver disease. *CMAJ*. 2010;**182**(11):E526–E531.
11. Nakagawa S, Wei L, Song WM, Higashi T, Ghoshal S, Kim RS, Bian CB, Yamada S, Sun X, Venkatesh A, Goossens N, Bain G, Lauwers GY, Koh AP, El-Abtah M, Ahmad NB, Hoshida H, Erstad DJ, Gunasekaran G, Lee Y, Yu ML, Chuang WL, Dai CY, Kobayashi M, Kumada H, Beppu T, Baba H, Mahajan M, Nair VD, Lanuti M, Villanueva A, Sangiovanni A, Iavarone M, Colombo M, Llovet JM, Subramanian A, Tager AM, Friedman SL, Baumert TF, Schwarz ME, Chung RT, Tanabe KK, Zhang B, Fuchs BC, Hoshida Y; Precision Liver Cancer Prevention Consortium. Molecular liver cancer prevention in cirrhosis by organ transcriptome analysis and lysophosphatidic acid pathway inhibition. *Cancer Cell*. 2016;**30**(6):879–890.
12. Starmann J, Fälth M, Spindelböck W, Lanz KL, Lackner C, Zatloukal K, Trauner M, Sültmann H. Gene expression profiling unravels cancer-related hepatic molecular signatures in steatohepatitis but not in steatosis. *PLoS One*. 2012;**7**(10):e46584.

13. Asselah T, Bièche I, Laurendeau I, Paradis V, Vidaud D, Degott C, Martinot M, Bedossa P, Valla D, Vidaud M, Marcellin P. Liver gene expression signature of mild fibrosis in patients with chronic hepatitis C. *Gastroenterology*. 2005;**129**(6):2064–2075.

14. Wruck W, Kashofer K, Rehman S, Daskalaki A, Berg D, Gralka E, Jozefczuk J, Drews K, Pandey V, Regenbrecht C, Wierling C, Turano P, Korf U, Zatloukal K, Lehrach H, Westerhoff HV, Adjaye J. Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes. *Sci Data*. 2015;**2**:150068.

15. Moylan CA, Pang H, Dellinger A, Suzuki A, Garrett ME, Guy CD, Murphy SK, Ashley-Koch AE, Choi SS, Michelotti GA, Hampton DD, Chen Y, Tillmann HL, Hauser MA, Abdelmalek MF, Diehl AM. Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology*. 2014;**59**(2):471–482.

16. Wood GC, Chu X, Manney C, Strodel W, Petrick A, Gabrielsen J, Seiler J, Carey D, Argyropoulos G, Benotti P, Still CD, Gerhard GS. An electronic health record-enabled obesity database. *BMC Med Inform Decis Mak*. 2012;**12**(1):45.

17. Leti F, Malenica I, Doshi M, Courtright A, Van Keuren-Jensen K, Legendre C, Still CD, Gerhard GS, DiStefano JK. High-throughput sequencing reveals altered expression of hepatic microRNAs in nonalcoholic fatty liver disease-related fibrosis. *Transl Res*. 2015;**166**(3):304–314.

18. DiStefano JK, Kingsley C, Craig Wood G, Chu X, Argyropoulos G, Still CD, Doné SC, Legendre C, Tembe W, Gerhard GS. Genome-wide analysis of hepatic lipid content in extreme obesity. *Acta Diabetol*. 2015;**52**(2):373–382.

19. Gerhard GS, Benotti P, Wood GC, Chu X, Argyropoulos G, Petrick A, Strodel WE, Gabrielsen JD, Ibele A, Still CD, Kingsley C, DiStefano J. Identification of novel clinical factors associated with hepatic fat accumulation in extreme obesity. *J Obesity*. 2014:368210.

20. 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**(3):R25.

21. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;**31**(2):166–169.

22. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;**15**(12):550.

23. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;**57**:289–300.

24. Weiskirchen R, Weimer J, Meurer SK, Kron A, Seipel B, Vater I, Arnold N, Siebert R, Xu L, Friedman SL, Bergmann C. Genetic characteristics of the human hepatic stellate cell line LX-2. *PLoS One*. 2013; **8**(10):e75692.

25. Wu Y, Liu X, Zhou Q, Huang C, Meng X, Xu F, Li J. Silent information regulator 1 (SIRT1) ameliorates liver fibrosis via promoting activated stellate cell apoptosis and reversion. *Toxicol Appl Pharmacol*. 2015;**289**(2):163–176.

26. Mohammadnia A, Yaqubi M, Fallahi H. Predicting transcription factors in human alcoholic hepatitis from gene regulatory network. *Eur Rev Med Pharmacol Sci*. 2015;**19**(12):2246–2253.

27. Liu J, Wang B, Wang W, Sun M, Li Y, Jia X, Zhai S, Dang S. Computational networks of activating transcription factor 3 gene in Huh7 cell lines and hepatitis C virus-infected Huh7 cell lines. *Mol Med Rep*. 2015;**12**(1):1239–1246.

28. Affò S, Morales-Ibanez O, Rodrigo-Torres D, Altamirano J, Blaya D, Dapito DH, Millán C, Coll M, Caviglia JM, Arroyo V, Caballería J, Schwabe RF, Ginès P, Bataller R, Sancho-Bru P. CCL20 mediates lipopolysaccharide induced liver injury and is a potential driver of inflammation and fibrosis in alcoholic hepatitis. *Gut*. 2014;**63**(11):1782–1792.

29. Gao B, Xu M. Chemokines and alcoholic hepatitis: are chemokines good therapeutic targets? *Gut*. 2014; **63**(11):1683–1684.

30. Makowska Z, Boldanova T, Adametz D, Quagliata L, Vogt JE, Dill MT, Matter MS, Roth V, Terracciano L, Heim MH. Gene expression analysis of biopsy samples reveals critical limitations of transcriptome-based molecular classifications of hepatocellular carcinoma. *J Pathol Clin Res*. 2016;**2**(2):80–92.

31. Bièche I, Asselah T, Laurendeau I, Vidaud D, Degot C, Paradis V, Bedossa P, Valla DC, Marcellin P, Vidaud M. Molecular profiling of early stage liver fibrosis in patients with chronic hepatitis C virus infection. *Virology*. 2005;**332**(1):130–144.

32. Friedman SL. Molecular regulation of hepatic fibrosis, an integrated cellular response to tissue injury. *J Biol Chem*. 2000;**275**(4):2247–2250.

33. Friedman SL, Roll FJ, Boyles J, Arenson DM, Bissell DM. Maintenance of differentiated phenotype of cultured rat hepatic lipocytes by basement membrane matrix. *J Biol Chem*. 1989;**264**(18): 10756–10762.

34. Schuppan D, Ruehl M, Somasundaram R, Hahn EG. Matrix as a modulator of hepatic fibrogenesis. *Semin Liver Dis*. 2001;**21**(3):351–372.

35. Xu L, Hui AY, Albanis E, Arthur MJ, O'Byrne SM, Blaner WS, Mukherjee P, Friedman SL, Eng FJ. Human hepatic stellate cell lines, LX-1 and LX-2: new tools for analysis of hepatic fibrosis. *Gut*. 2005; **54**(1):142–151.

36. Kato A, Okamoto O, Ishikawa K, Sumiyoshi H, Matsuo N, Yoshioka H, Nomizu M, Shimada T, Fujiwara S. Dermatopontin interacts with fibronectin, promotes fibronectin fibril formation, and enhances cell adhesion. *J Biol Chem*. 2011;**286**(17):14861–14869.

37. Li Z, Tang M, Ling B, Liu S, Zheng Y, Nie C, Yuan Z, Zhou L, Guo G, Tong A, Wei Y. Increased expression of S100A6 promotes cell proliferation and migration in human hepatocellular carcinoma. *J Mol Med (Berl)*. 2014;**92**(3):291–303.

38. Lorena D, Darby IA, Gadeau AP, Leen LL, Rittling S, Porto LC, Rosenbaum J, Desmoulière A. Osteopontin expression in normal and fibrotic liver. altered liver healing in osteopontin-deficient mice. *J Hepatol*. 2006;**44**(2):383–390.

39. Paradis V, Dargere D, Bieche Y, Asselah T, Marcellin P, Vidaud M, Bedossa P. SCG10 expression on activation of hepatic stellate cells promotes cell motility through interference with microtubules. *Am J Pathol*. 2010;**177**(4):1791–1797.

40. Schissel SL, Dunsmore SE, Liu X, Shine RW, Perrella MA, Layne MD. Aortic carboxypeptidase-like protein is expressed in fibrotic human lung and its absence protects against bleomycin-induced lung fibrosis. *Am J Pathol*. 2009;**174**(3):818–828.

41. Tumelty KE, Smith BD, Nugent MA, Layne MD. Aortic carboxypeptidase-like protein (ACLP) enhances lung myofibroblast differentiation through transforming growth factor β receptor-dependent and -independent pathways. *J Biol Chem*. 2014;**289**(5):2526–2536.

42. Wang M, Gong Q, Zhang J, Chen L, Zhang Z, Lu L, Yu D, Han Y, Zhang D, Chen P, Zhang X, Yuan Z, Huang J, Zhang X. 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.

43. Williams KH, Viera de Ribeiro AJ, Prakoso E, Veillard AS, Shackel NA, Bu Y, Brooks B, Cavanagh E, Raleigh J, McLennan SV, McCaughan GW, Bachovchin WW, Keane FM, Zekry A, Twigg SM, Gorrell MD. Lower serum fibroblast activation protein shows promise in the exclusion of clinically significant liver fibrosis due to non-alcoholic fatty liver disease in diabetes and obesity. *Diabetes Res Clin Pract*. 2015;**108**(3):466–472.

44. Arendt BM, Comelli EM, Ma DW, Lou W, Teterina A, Kim T, Fung SK, Wong DK, McGilvray I, Fischer SE, Allard JP. Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. *Hepatology*. 2015;**61**(5):1565–1578.

45. Sato S, Genda T, Hirano K, Tsuzura H, Narita Y, Kanemitsu Y, Kikuchi T, Iijima K, Wada R, Ichida T. Up-regulated aldo-keto reductase family 1 member B10 in chronic hepatitis C: association with serum alpha-fetoprotein and hepatocellular carcinoma. *Liver Int*. 2012;**32**(9):1382–1390.

46. Jin J, Liao W, Yao W, Zhu R, Li Y, He S. Aldo-keto reductase family 1 member B 10 mediates liver cancer cell proliferation through sphingosine-1-phosphate. *Sci Rep*. 2016;**6**(1):22746.

47. Bauer Y, White ES, de Bernard S, Cornelisse P, Leconte I, Morganti A, Roux S, Nayler O. MMP-7 is a predictive biomarker of disease progression in patients with idiopathic pulmonary fibrosis. *ERJ Open Res*. 2017;**3**(1):3.

48. Huang CC, Chuang JH, Chou MH, Wu CL, Chen CM, Wang CC, Chen YS, Chen CL, Tai MH. Matrilysin (MMP-7) is a major matrix metalloproteinase upregulated in biliary atresia-associated liver fibrosis. *Mod Pathol*. 2005;**18**(7):941–950.

49. Seth D, Duly A, Kuo PC, McCaughan GW, Haber PS. Osteopontin is an important mediator of alcoholic liver disease via hepatic stellate cell activation. *World J Gastroenterol*. 2014;**20**(36):13088–13104.

50. Lua I, Li Y, Zagory JA, Wang KS, French SW, Sévigny J, Asahina K. Characterization of hepatic stellate cells, portal fibroblasts, and mesothelial cells in normal and fibrotic livers. *J Hepatol*. 2016; **64**(5):1137–1146.

51. Reyes-Gordillo K, Shah R, Arellanes-Robledo J, Hernández-Nazara Z, Rincón-Sánchez AR, Inagaki Y, Rojkind M, Lakshman MR. Mechanisms of action of acetaldehyde in the up-regulation of the human α2(I) collagen gene in hepatic stellate cells: key roles of Ski, SMAD3, SMAD4, and SMAD7. *Am J Pathol*. 2014;**184**(5):1458–1467.

52. Staten NR, Welsh EA, Sidik K, McDonald SA, Dufield DR, Maqsodi B, Ma Y, McMaster GK, Mathews RW, Arch RH, Masferrer JL, Souberbielle BE. Multiplex transcriptional analysis of paraffin-embedded liver needle biopsy from patients with liver fibrosis. *Fibrogenesis Tissue Repair*. 2012;**5**(1):21.

53. Girard M, Bizet AA, Lachaux A, Gonzales E, Filhol E, Collardeau-Frachon S, Jeanpierre C, Henry C, Fabre M, Viremouneix L, Galmiche L, Debray D, Bole-Feysot C, Nitschke P, Pariente D, Guettier C,

Lyonnet S, Heidet L, Bertholet A, Jacquemin E, Henrion-Caude A, Saunier S. DCDC2 mutations cause neonatal sclerosing cholangitis. *Hum Mutat*. 2016;**37**(10):1025–1029.

54. Syn WK, Agboola KM, Swiderska M, Michelotti GA, Liaskou E, Pang H, Xie G, Philips G, Chan IS, Karaca GF, Pereira TA, Chen Y, Mi Z, Kuo PC, Choi SS, Guy CD, Abdelmalek MF, Diehl AM. NKT-associated hedgehog and osteopontin drive fibrogenesis in non-alcoholic fatty liver disease. *Gut*. 2012;**61**(9):1323–1329.

55. Patouraux S, Bonnafous S, Voican CS, Anty R, Saint-Paul MC, Rosenthal-Allieri MA, Agostini H, Njike M, Barri-Ova N, Naveau S, Le Marchand-Brustel Y, Veillon P, Calès P, Perlemuter G, Tran A, Gual P. The osteopontin level in liver, adipose tissue and serum is correlated with fibrosis in patients with alcoholic liver disease. *PLoS One*. 2012;**7**(4):e35612.

56. Zhao L, Li T, Wang Y, Pan Y, Ning H, Hui X, Xie H, Wang J, Han Y, Liu Z, Fan D. Elevated plasma osteopontin level is predictive of cirrhosis in patients with hepatitis B infection. *Int J Clin Pract*. 2008;**62**(7):1056–1062.

57. Huang W, Zhu G, Huang M, Lou G, Liu Y, Wang S. Plasma osteopontin concentration correlates with the severity of hepatic fibrosis and inflammation in HCV-infected subjects. *Clin Chim Acta*. 2010;**411**(9-10): 675–678.

58. Huang H, Wu T, Mao J, Fang Y, Zhang J, Wu L, Zheng S, Lin B, Pan H. CHI3L1 is a liver-enriched, noninvasive biomarker that can be used to stage and diagnose substantial hepatic fibrosis. *OMICS*. 2015;**19**(6):339–345.

59. Charlton M, Viker K, Krishnan A, Sanderson S, Veldt B, Kaalsbeek AJ, Kendrick M, Thompson G, Que F, Swain J, Sarr M. Differential expression of lumican and fatty acid binding protein-1: new insights into the histologic spectrum of nonalcoholic fatty liver disease. *Hepatology*. 2009;**49**(4):1375–1384.

60. Decaris ML, Li KW, Emson CL, Gatmaitan M, Liu S, Wang Y, Nyangau E, Colangelo M, Angel TE, Beysen C, Cui J, Hernandez C, Lazaro L, Brenner DA, Turner SM, Hellerstein MK, Loomba R. Identifying nonalcoholic fatty liver disease patients with active fibrosis by measuring extracellular matrix remodeling rates in tissue and blood. *Hepatology*. 2017;**65**(1):78–88.

61. Bracht T, Schweinsberg V, Trippler M, Kohl M, Ahrens M, Padden J, Naboulsi W, Barkovits K, Megger DA, Eisenacher M, Borchers CH, Schlaak JF, Hoffmann AC, Weber F, Baba HA, Meyer HE, Sitek B. Analysis of disease-associated protein expression using quantitative proteomics—fibulin-5 is expressed in association with hepatic fibrosis. *J Proteome Res*. 2015;**14**(5):2278–2286.

62. Xu MY, Qu Y, Li Z, Li F, Xiao CY, Lu LGA. A 6 gene signature identifies the risk of developing cirrhosis in patients with chronic hepatitis B. *Front Biosci*. 2016;**21**(3):479–486.

63. Ellis EL, Mann DA. Clinical evidence for the regression of liver fibrosis. *J Hepatol*. 2012;**56**(5): 1171–1180.

64. Zhong Y, Qin Y, Dang L, Jia L, Zhang Z, Wu H, Cui J, Bian H, Li Z. Alteration and localization of glycan-binding proteins in human hepatic stellate cells during liver fibrosis. *Proteomics*. 2015;**15**(19):3283–3295.

65. Makarev E, Izumchenko E, Aihara F, Wysocki PT, Zhu Q, Buzdin A, Sidransky D, Zhavoronkov A, Atala A. Common pathway signature in lung and liver fibrosis. *Cell Cycle*. 2016;**15**(13):1667–1673.

66. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One*. 2014;**9**(1):e78644.

67. Guo Y, Xiong Y, Sheng Q, Zhao S, Wattacheril J, Flynn CR. A micro-RNA expression signature for human NAFLD progression. *J Gastroenterol*. 2016;**51**(10):1022–1030.

68. Sun C, Liu X, Yi Z, Xiao X, Yang M, Hu G, Liu H, Liao L, Huang F. Genome-wide analysis of long noncoding RNA expression profiles in patients with non-alcoholic fatty liver disease. *IUBMB Life*. 2015;**67**(11):847–852.

69. Soronen J, Yki-Järvinen H, Zhou Y, Sädevirta S, Sarin AP, Leivonen M, Sevastianova K, Perttilä J, Laurila PP, Sigruener A, Schmitz G, Olkkonen VM. Novel hepatic microRNAs upregulated in human nonalcoholic fatty liver disease. *Physiol Rep*. 2016;**4**(1):4.

70. Teufel A, Itzel T, Erhart W, Brosch M, Wang XY, Kim YO, von Schonfels W, Herrmann A, Bruckner S, Stickel F, Dufour JF, Chavakis T, Hellerbrand C, Spang R, Maass T, Becker T, Schreiber S, Schafmayer C, Schuppan D, Hampe J. Comparison of gene expression patterns between mouse models of non-alcoholic fatty liver disease and liver tissues from patients. *Gastroenterology*. 2016;**151**:513–525.

71. Krishnan A, Abdullah TS, Mounajjed T, Hartono S, McConico A, White T, LeBrasseur N, Lanza I, Nair S, Gores G, Charlton M. A longitudinal study of whole body, tissue, and cellular physiology in a mouse model of fibrosing NASH with high fidelity to the human condition. *Am J Physiol Gastrointest Liver Physiol*. 2017;**312**(6):G666–G680.

72. Lefebvre P, Lalloyer F, Baugé E, Pawlak M, Gheeraert C, Dehondt H, Vanhoutte J, Woitrain E, Hennuyer N, Mazuy C, Bobowski-Gérard M, Zummo FP, Derudas B, Driessen A, Hubens G, Vonghia L, Kwanten WJ, Michielsen P, Vanwolleghem T, Eeckhoute J, Verrijken A, Van Gaal L, Francque S, Staels B. Interspecies NASH disease activity whole-genome profiling identifies a fibrogenic role of PPARα-regulated dermatopontin. *JCI Insight*. 2017;**2**(13):2.