LIVER DISEASE

Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis

Olivier Govaere[1], Simon Cockell[2], Dina Tiniakos[1,3], Rachel Queen[2], Ramy Younes[1,4], Michele Vacca[5], Leigh Alexander[6], Federico Ravaioli[1,7], Jeremy Palmer[1], Salvatore Petta[8], Jerome Boursier[9], Chiara Rosso[4], Katherine Johnson[1], Kristy Wonders[1], Christopher P. Day[1], Mattias Ekstedt[10], Matej Orešič[11], Rebecca Darlay[12], Heather J. Cordell[12], Fabio Marra[13], Antonio Vidal-Puig[5], Pierre Bedossa[1,14], Jörn M. Schattenberg[15], Karine Clément[16], Michael Allison[17], Elisabetta Bugianesi[4], Vlad Ratziu[14], Ann K. Daly[1]*, Quentin M. Anstee[1,18]*

The mechanisms that drive nonalcoholic fatty liver disease (NAFLD) remain incompletely understood. This large multicenter study characterized the transcriptional changes that occur in liver tissue across the NAFLD spectrum as disease progresses to cirrhosis to identify potential circulating markers. We performed high-throughput RNA sequencing on a discovery cohort comprising histologically characterized NAFLD samples from 206 patients. Unsupervised clustering stratified NAFLD on the basis of disease activity and fibrosis stage with differences in age, aspartate aminotransferase (AST), type 2 diabetes mellitus, and carriage of *PNPLA3 rs738409*, a genetic variant associated with NAFLD. Relative to early disease, we consistently identified 25 differentially expressed genes as fibrosing steatohepatitis progressed through stages F2 to F4. This 25-gene signature was independently validated by logistic modeling in a separate replication cohort ($n = 175$), and an integrative analysis with publicly available single-cell RNA sequencing data elucidated the likely relative contribution of specific intrahepatic cell populations. Translating these findings to the protein level, SomaScan analysis in more than 300 NAFLD serum samples confirmed that circulating concentrations of proteins AKR1B10 and GDF15 were strongly associated with disease activity and fibrosis stage. Supporting the biological plausibility of these data, in vitro functional studies determined that endoplasmic reticulum stress up-regulated expression of *AKR1B10*, *GDF15*, and *PDGFA*, whereas GDF15 supplementation tempered the inflammatory response in macrophages upon lipid loading and lipopolysaccharide stimulation. This study provides insights into the pathophysiology of progressive fibrosing steatohepatitis, and proof of principle that transcriptomic changes represent potentially tractable and clinically relevant markers of disease progression.

INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD) is an increasingly common progressive disease characterized by excessive hepatic accumulation of triglycerides and reactive lipid species and is strongly associated with the metabolic syndrome, central obesity, type 2 diabetes mellitus (T2DM), hypertension, and dyslipidemia (*1*). With a global increase in sedentary behavior and obesity, the prevalence of NAFLD is rising rapidly and now affects about 25% of the adult population worldwide (*1*). NAFLD is subdivided into "simple" steatosis [nonalcoholic fatty liver (NAFL)] and nonalcoholic steatohepatitis (NASH), defined by the presence of necroinflammation and hepatocyte ballooning. If NASH persists, fibrosis occurs and may progress to cirrhosis and, ultimately, end-stage liver disease (*2*). NAFL was traditionally considered a stable and relatively benign disease state that lacked the capacity to progress. However, recent data from serial biopsy studies have demonstrated that NAFL may transit into NASH and onward to advanced fibrosis (*3*). NAFLD is therefore best considered a dynamic disease with steatohepatitic activity waxing and waning, and fibrosis stage also progressing and regressing subject to the actions of a variety of genetic, epigenetic, and environmental modifiers (*2*). Key management challenges for NAFLD include the lack of effective biomarkers to risk-stratify patients and the lack of approved pharmacological therapies.

To date, efforts to examine the transcriptomic changes that occur as NAFLD progresses have largely used microarray-based techniques and so have lacked the comprehensive approach provided by global RNA sequencing (RNA-seq) (*4*–*10*) or have been confined to

[1]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. [2]Bioinformatics Support Unit, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. [3]Department of Pathology, Aretaieio Hospital, National and Kapodistrian University of Athens, 15771 Athens, Greece. [4]Department of Medical Sciences, Division of Gastro-Hepatology, A.O. Città della Salute e della Scienza di Torino, University of Turin, 10124 Turin, Italy. [5]University of Cambridge Metabolic Research Laboratories, Wellcome-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. [6]SomaLogic Inc., Boulder, CO 80301, USA. [7]Department of Medical and Surgical Sciences, University of Bologna, 40126 Bologna, Italy. [8]Sezione di Gastroenterologia, Dipartimento Biomedico di Medicina Interna e Specialistica, Università di Palermo, 90133 Palermo, Italy. [9]Hepatology Department, Angers University Hospital, 49933 Angers, France. [10]Department of Health, Medicine and Caring Sciences, Linköping University, 581 83 Linköping, Sweden. [11]School of Medical Sciences, Örebro University, 702 81 Örebro, Sweden. [12]Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. [13]Dipartimento di Medicina Sperimentale e Clinica, University of Florence, 50121 Florence, Italy. [14]Assistance Publique–Hôpitaux de Paris, Hôpital Pitié Salpêtrière, Sorbonne University, ICAN (Institute of Cardiometabolism and Nutrition), 75013 Paris, France. [15]Department of Medicine, University Hospital Mainz, 55131 Mainz, Germany. [16]Nutrition and Obesities: Systemic Approaches, INSERM, Sorbonne University, 75006 Paris, France. [17]Liver Unit, Department of Medicine, NIHR Cambridge Biomedical Research Centre, Cambridge University NHS Foundation Trust, Cambridge CB2 0QQ, UK. [18]NIHR Newcastle Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Trust, Newcastle upon Tyne NE7 7DN, UK.
*Corresponding author. Email: a.k.daly@ncl.ac.uk (A.K.D.); quentin.anstee@ncl.ac.uk (Q.M.A.)

relatively small patient cohorts that do not adequately represent the full spectrum of disease from normal liver through NAFL to NASH, exhibiting progressive stages of fibrosis and cirrhosis (*11*–*13*). Consequently, studies have been limited to dichotomous comparisons between mild and advanced disease, which do not provide the granularity needed to fully appreciate the complex transcriptomic changes as NAFLD evolves, and lack independent validation.

We report a comprehensive transcriptomic analysis conducted using RNA-seq technology across the full histological spectrum from healthy controls through to NASH-associated cirrhosis in a large cohort of European patients. Adopting an integrative transcriptomic approach to unravel pathways responsible for the stepwise progression of NAFLD, we identified and validated gene expression signatures associated with early stages of disease, subsequent progression, and specific histological features. Translating our findings from the hepatic transcriptome to the protein level, we further validated selected gene expression changes associated with disease activity and fibrosis stage using immunohistochemistry and measurement of circulating protein.

RESULTS

Unsupervised clustering stratifies NAFLD on the basis of fibrosis and disease activity

The current study incorporates transcriptomic data on 403 individuals, 381 NAFLD samples, and 22 control samples, representing the full histological range from normal liver tissue to NASH-cirrhosis. A total of 206 frozen tissue samples from patients with NAFLD were included in the discovery cohort and processed for high-throughput RNA-seq (Fig. 1). All samples were histologically scored by two expert liver pathologists (D.T. and P.B.) according to the widely accepted, semiquantitative NASH-Clinical Research Network NASH Activity Score (NAS) and the FLIP steatosis (S), activity (A), and fibrosis (F) "SAF" scoring systems (*14*, *15*) and grouped according to histopathological disease grade and stage: NAFL and NASH with different fibrosis stages F0, F1, F2, F3, and F4. Detailed phenotypic description and demographics are reported in Table 1. As indicated by principal components analysis, potential confounding factors including sex, batch, and center were corrected for in the analysis (fig. S1).

Unsupervised clustering based on gene expression stratified the 206 NAFLD samples from the discovery cohort into two distinct groups, annotated as clusters A and B (Fig. 2). Histologically, cluster A was characterized by more advanced fibrosis (Mann-Whitney *U* test, $P = 5.15 \times 10^{-10}$), a higher grade of hepatic ballooning ($P = 1.67 \times 10^{-5}$) and lobular inflammation (Kleiner scoring, 0 to 3, $P = 1.70 \times 10^{-3}$; SAF scoring, 0 to 2, $P = 1.36 \times 10^{-4}$), with no differences in steatosis grade when compared with cluster B ($P = 4.78 \times 10^{-1}$) (Table 1). This translated into a higher number of patients diagnosed with NASH in cluster A (84.62%) compared with cluster B (69.5%; chi-square, $P = 2.11 \times 10^{-2}$). Moreover, when stratifying samples on the basis of a high disease activity using an NAS score ≥ 4 (sum of steatosis, ballooning, and Kleiner inflammation) or an SAF activity score ≥ 2 (sum of ballooning and SAF inflammation), cluster A showed an enrichment compared with cluster B (NAS ≥ 4 83.08% versus 66.67%; chi-square, $P = 1.49 \times 10^{-2}$; SAF activity score ≥ 2, 86.15% versus 68.79%, $P = 8.08 \times 10^{-3}$) (Table 1). Compared with cluster B, patients in cluster A were slightly older (57.08 years versus 52.57 years; Mann-Whitney *U* test, $P = 1.58 \times 10^{-2}$), more overweight [body mass index (BMI) 32.51 versus 30.8; Mann-Whitney *U* test, $P = 1.99 \times 10^{-2}$], more likely to have T2DM (70.77% versus 45.39%; chi-square, $P = 6.90 \times 10^{-4}$), and exhibited higher HbA1c concentrations (52.16 mmol/mol versus 45.8 mmol/mol; Mann-Whitney *U* test, $P = 4.17 \times 10^{-2}$), higher serum aspartate aminotransferase (AST) (53.23 U/liter versus 40.67 U/liter; Mann-Whitney *U* test, $P = 4.62 \times 10^{-4}$), and reduced platelet count (212.53×10^9 versus 237.78×10^9; Mann-Whitney *U* test, $P = 2.05 \times 10^{-2}$) (Table 1). Although carriage of well-described genetic variants that have been associated with NAFLD severity (*16*) (*GCKR* rs1260326, *HSD17B13* rs72613567, *PNPLA3* rs738409, and *TM6SF2* rs58542926) did not confer discrete gene expression profiles within the overall RNA-seq dataset (fig. S2), cluster A showed an increase in carriers for the *PNPLA3* rs738409 polymorphism when compared with cluster B (chi-square, $P = 3.06 \times 10^{-2}$).

Between clusters A and B, we found 1292 differentially expressed genes (DEGs; $q < 0.05$ and fold change > |1.5|) (table S1). KEGG pathway analysis indicated an enrichment in genes correlating to pathways including "extracellular matrix interaction," "focal adhesion," "phosphatidylinositol 3-kinase (PI3K)–Akt signaling," and "Wnt signaling" (fig.S3, A to C). Moreover, cluster A showed an increased expression in cytokine genes such as *CCL2*, *CCL20*, *CCL19*, and *CCL28* (fig. S3B) and also in hepatic progenitor cell marker genes (*TACSD2*/*TROP2*, *EPCAM*, *SOX9*, *KRT19*, *KRT7*, *CD24*, and *JAG1*), suggesting progenitor cell–mediated regeneration. Together, unsupervised clustering stratified NAFLD patient samples not only on the basis of fibrosis stage but also on the basis of disease activity and carriage of the *PNPLA3* rs738409 variant and transcriptionally defined a patient subgroup that showed high Wnt signaling, active tissue remodeling, and progenitor cell–mediated regeneration.

Nested within cluster B were further subgroups that did not show strong differences in histological features alone but were rather characterized by changes in gene expression associated with tissue remodeling and PI3K-Akt signaling pathway, including Mechanistic Target Of Rapamycin Kinase (mTOR) signaling–related genes (for example, *RICTOR* and *LAMTOR4*)

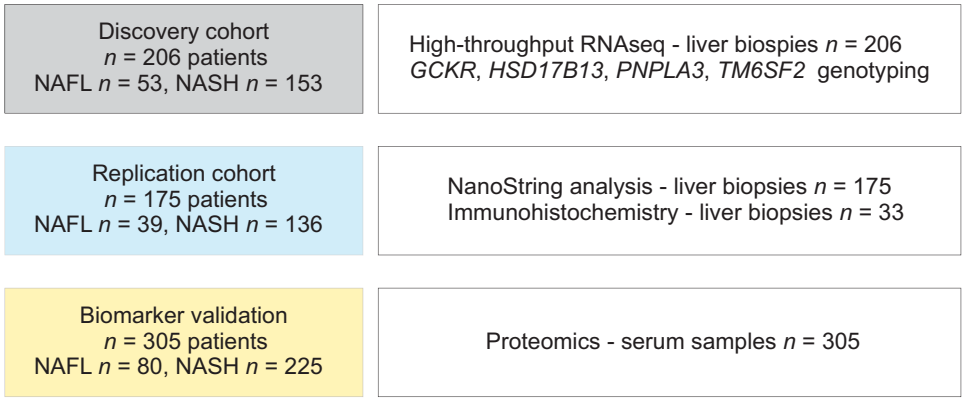

Fig. 1. Experimental study design. A total of 381 NAFLD biopsies and 305 NAFLD serum samples were included in this study. All samples were centrally read by two expert liver pathologists.

Table 1. Clinicopathological characteristics of the unsupervised clusters from the discovery cohort RNA-seq. BMI, body mass index.

Clinical features	Patient (*n*)	Total (*n* = 206)	Cluster A (*n* = 65)	Cluster B (*n* = 141)	*P* value (A vs. B)
Age (mean ± SD)	206	54 (±11.87)	57.08 (±10.11)	52.57 (±12.37)	1.58×10^{-2}
Sex					3.90×10^{-1}
Male	206	123	36	87	
Female		83	29	54	
BMI (mean ± SD)	204	31.34 (±5.04)	32.51 (±4.87)	30.8 (±5.04)	1.99×10^{-2}
T2DM	206				6.90×10^{-4}
No		96	19	77	
Yes		110	46	64	
HBA1C (mmol/mol ± SD)	135	48.06 (±14.54)	52.16 (±17.46)	45.8 (±12.19)	4.17×10^{-2}
Alanine aminotransferase (ALT) (mean ± SD)	204	67.12 (±41.61)	71.05 (±41.39)	65.33 (±41.73)	1.71×10^{-1}
Aspartate aminotransferase (AST) (mean ± SD)	201	44.67 (±23.08)	53.23 (±26.17)	40.67 (±20.39)	4.62×10^{-4}
Platelet ($\times 10^9$)	169	229.56 (±65.97)	212.53 (±63.15)	237.78 (±65.99)	2.05×10^{-2}
Triglycerides (mM)	180	1.95 (±1.41)	2.02 (±1.43)	1.91 (±1.4)	7.40×10^{-1}
Total cholesterol (mM)	178	5.45 (±10.14)	4.65 (±1.25)	5.85 (±12.37)	6.70×10^{-1}
Steatosis grade	206				4.78×10^{-1}
0		0	0	0	
1		60	16	44	
2		73	25	48	
3		73	24	49	
Ballooning	206				1.67×10^{-5}
0		52	10	42	
1		98	23	75	
2		56	32	24	
Kleiner lobular inflammation	206				1.70 × 10–3
0		16	4	12	
1		95	22	73	
2		80	28	52	
3					
SAF lobular inflammation	206				1.36×10^{-4}
0		16	4	12	
1		140	33	107	
2		50	28	22	
Brunt fibrosis stage	206				5.15×10^{-10}
0		38	5	33	
1		47	4	43	
2		53	16	37	
3		54	29	25	
4		14	11	3	
NASH	206				2.11×10^{-2}
No		53	10	43	
Yes		153	55	98	
NAS score ≥4	206				1.49×10^{-2}
No		58	11	47	

continued to next page

Clinical features	Patient (*n*)	Total (*n* = 206)	Cluster A (*n* = 65)	Cluster B (*n* = 141)	*P* value (A vs. B)
Yes		148	54	94	
SAF activity score ≥2	206				8.08×10^{-3}
No		53	9	44	
Yes		153	56	97	
GCKR rs1260326 (CC/CT/TT)	206	49/105/52	12/33/20	37/72/32	3.20×10^{-1}
HSD17B13 rs72613567 (–/–T/TT) [unknown]	188	120/61/7 [18]	33/25/3 [4]	87/36/4 [14]	1.56×10^{-1}
PNPLA3 rs738409 (CC/GC/GG)	206	75/89/42	23/22/20	52/67/22	3.06×10^{-2}
TM6SF2 rs58542926 (CC/CT/TT)	206	156/48/2	46/19/0	110/29/2	2.64×10^{-1}

and differences in nuclear factors including *SP3*, *NR2C2*, and *LCOR* (table S2 and fig. S3D). This suggests that NAFLD is a more heterogeneous condition than has previously been clinically defined.

Transcriptional changes apparent in NAFL are sustained as fibrosing steatohepatitis progresses

To further understand the pathogenesis and progression of NAFLD, we performed supervised clustering of our RNA-seq dataset comparing the different stages of NAFLD with control samples. Principal components analysis showed a distinct separation between the NAFLD cases ($n = 206$) and "healthy obese" controls (histologically normal liver tissue samples obtained from obese patients, $n = 10$) (fig. S4A). Similarly, using publicly available control RNAseq datasets, a gradual shift was observed from healthy nonobese individuals to healthy obese patients and to patients with NAFLD (fig. S4B) (*12*). A total of 2603 DEGs were identified between NAFLD and the control samples and between 1875 and 3578 DEGs when comparing individual stages, showing enrichment for pathways including "metabolic pathways," "ribosome," and "tumor necrosis factor (TNF) signaling" ($q < 0.05$, fig. S4C) with genes in pathways such as "platelet activation," "peroxisome proliferator–activated receptor signaling," and "extracellular matrix" only exhibiting an enrichment at later stages of disease progression such as NASH F3/F4 (fig. S5A).

Considering a priori pathophysiologically relevant candidates (*17*), we found expression of several genes to be increased in early disease (NAFL), including the cellular senescence marker *CDKN1A/p21*, the inflammatory marker *IKBKG/NEMO*, and *CYP7A1*, the rate-limiting enzyme in the classical bile acid synthesis pathway that is subject to farnesoid X receptor (FXR)–mediated regulation (*18*), which peaked as NASH developed and remained elevated but became less so as fibrosis progressed (fig. S5B). These results show that transcriptomic changes related to initiation of inflammation, cellular senescence, and bile secretion begin to occur soon after NAFL inception and before histologically evident steatohepatitis.

Defining DEG sets associated with steatohepatitis and fibrosis

To study mechanisms associated with disease progression, we performed pairwise analyses between disease phenotype categories using the RNA-seq data from the 206 patients with NAFLD in the discovery cohort. To identify modifiers of steatohepatitis, we compared the different NASH stages with NAFL, whereas to discover modifiers of fibrosis, we looked at different fibrosis stages within NASH. Taking NAFL as a baseline, no statistically significant DEGs were identified compared with NASH F0/F1, whereas 50 DEGs, 907 DEGs, and 1369 DEGs were observed when comparing NAFL with NASH F2, NASH F3, or NASH F4, respectively (Fig. 3A and tables S3 to S5). Similarly, when using NASH F0/F1 as a baseline, no genes were differentially expressed compared with NASH F2, whereas comparison with NASH F3 and NASH F4 identified 434 DEGs and 1194 DEGs, respectively, with 393 DEGs in common between these two stages (Fig. 3A and tables S6 and S7). Gene Ontology (GO) annotation analyses of the different comparisons are described in fig. S6.

To identify a core gene set "signature" associated with the progression from NAFL or NASH F0/F1 to more advanced disease, we focused on DEG commonality between the different pairwise comparisons. The two intersections using either NAFL or NASH F0/F1 as a baseline shared 25 DEGs, with 24 of them showing a gradual increase with disease progression (*AKR1B10*, *ANKRD29*, *CCL20*, *CFAP221*, *CLIC6*, *COL1A1*, *COL1A2*, *DTNA*, *DUSP8*, *EPB41L4A*, *FERMT1*, *GDF15*, *HECW1*, *IL32*, *ITGBL1*, *LTBP2*, *PDGFA*, *PPAPDC1A*, *RGS4*, *SCTR*, *STMN2*, *THY1*, *TNFRSF12A*, and *TYMS*) and 1 gene showing a gradual decrease (*HSD17B14*) (Fig. 3B). These 25 genes were all differentially expressed when comparing NASH F2-F4 to NAFL + NASH F0/F1 combined and all apart from *TYMS* (fold change < 1.5) were differentially expressed between unsupervised clusters A and B (Fig. 3, C and D). When we used the widely accepted histological thresholds for high probability of NASH (NAS ≥ 4 or SAF activity ≥ 2) to stratify the discovery cohort, we identified 369 and 320 DEGs, respectively (tables S8 and S9). All 25 genes were present within the DEGs when the cohort was stratified purely by histological activity (inflammation and hepatocyte ballooning) with SAF activity ≥2. However, *ANKRD29*, *GDF15*, and *TYMS* were omitted (and so 22 of the 25 genes were captured) when using the NAS ≥4 threshold, which conflates degree of steatosis and grade of histological activity into a single index.

Findings from the discovery analysis were replicated using NanoString analysis in an independent cohort of 175 NAFLD liver biopsies (table S10). The breakdown of relative expression in the different disease stages compared with the controls in this cohort is presented in fig. S7. When comparing NAFL and NASH F0/F1 versus NASH F2-F4, we found 21 of the initial 25 genes in the gene set to be significantly differentially expressed (*AKR1B10*, *CCL20*, *CFAP221*, *CLIC6*, *COL1A1*, *COL1A2*, *DTNA*, *DUSP8*, *FERMT1*, *GDF15*, *HECW1*, *IL32*, *ITGBL1*, *LTBP2*, *PDGFA*, *PPAPDC1A*, *RGS4*, *SCTR*, *STMN2*, *THY1*, and *TNFRSF12A*) at $P < 0.001$ (fig. S7). Together, these data identified a consistently DEG signature of more advanced disease, associated with disease activity and correlating to the unsupervised clustering.

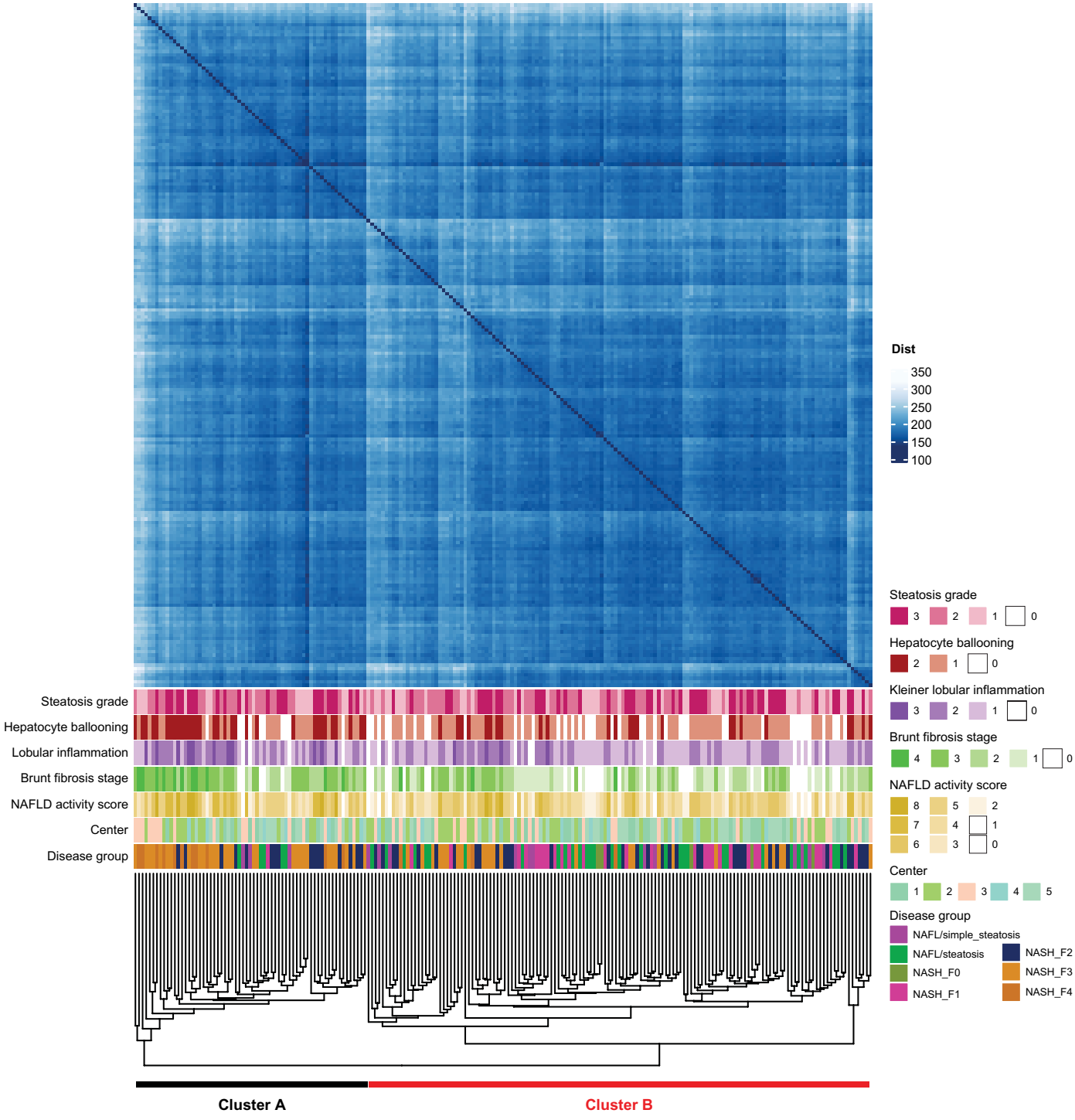

Fig. 2. Unsupervised clustering using RNA-seq from the discovery cohort patients with NAFLD. The distance (Dist) map based on the RNA-seq data ($n = 206$ patients) indicates the distribution of clinicopathological features.

A 25-gene set of DEGs independently predicts features of NAFLD

To investigate further the relationship between the changes in gene expression and the components of the histological phenotype and to dissect apart the collinearity between those features, we performed regression analysis using the RNA-seq data to generate gene signatures linearly associated with the severity of each specific histological feature (steatosis, inflammation, hepatocyte ballooning, and fibrosis) (Fig. 4A). Hierarchal clustering of genes that were significantly (adjusted $P < 0.05$) associated with histological criteria

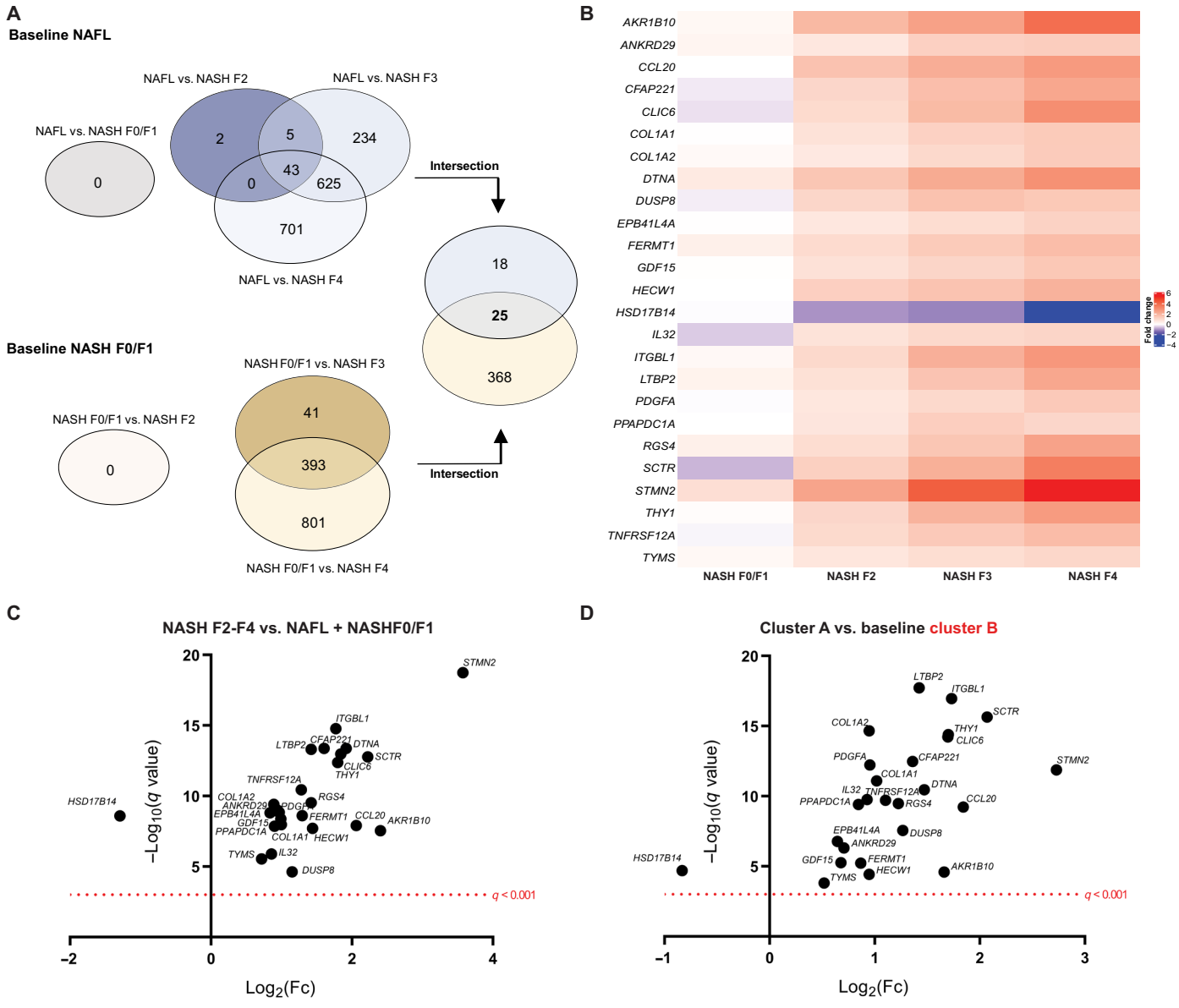

Fig. 3. Gene signatures associated with progressive NAFLD based on RNA-seq data from the discovery cohort ($n = 206$). (**A**) Venn diagram illustrating the number of DEGs identified by pairwise analyses using NAFL or NASH F0/F1 as a baseline to identify modifiers of steatohepatitis and fibrosis. (**B**) Heatmap of the 25-gene signature associated with advanced NAFLD identified by using NAFL or NASH F0/F1 as a baseline. Expression fold change is compared with NAFL. (**C**) Twenty-five–gene signature in the comparison of NASH F2-F4 to baseline NAFL + NASH F0/F1. (**D**) Twenty-five–gene signature in the comparison of unsupervised cluster A versus baseline cluster B. Data are presented using $\log_2$ fold change (Fc) in expression and $-\log_{10}$ of the q values.

revealed a stronger overlap between inflammation, ballooning, and fibrosis than with steatosis (Fig. 4B). Each of the 25 genes correlated strongly with increasing severity of inflammation, ballooning, and fibrosis, with 12 genes showing additional overlap with steatosis (*AKR1B10*, *CCL20*, *COL1A1*, *COL1A2*, *DTNA*, *DUSP8*, *GDF15*, *PDGFA*, *PPAPDC1A*, *STMN2*, *THY1*, and *TNFRSF12A*) (adjusted $P < 0.05$; Fig. 4A and table S11).

To understand the effect of the individual components of the 25-DEG gene set on histological features, we evaluated the predictive value in the NAFLD discovery cohort ($n = 206$) using univariate and multivariate logistic regression analysis. Table S12 summarizes the results from the univariate analyses. In the multivariate models predicting steatohepatitis grade, *HSD17B14* [Odds Ratio (OR) = 0.5049, $P = 0.001$], *PPAPDC1A* (OR = 3.2079, $P < 0.0001$), *SCTR* (OR = 0.6891, $P < 0.05$), and *TNFRSF12A* (OR = 2.7367, $P < 0.0001$) expression were independent factors significantly predictive of NAS ≥4 with an area under the receiver operating characteristic curve (AUROC) of 0.85 (Fig. 4C). In addition, SAF activity score ≥2 was independently predicted by AST concentration (OR = 1.1079, $P < 0.01$) together with *EPB41L4A* (OR = 2.6634, $P < 0.01$), *GDF15* (OR = 1.0253, $P = 0.001$), and *ITGBL1* (OR = 2.022, $P < 0.01$) expression, with an AUROC of 0.86 (Fig. 4D and table S13). The

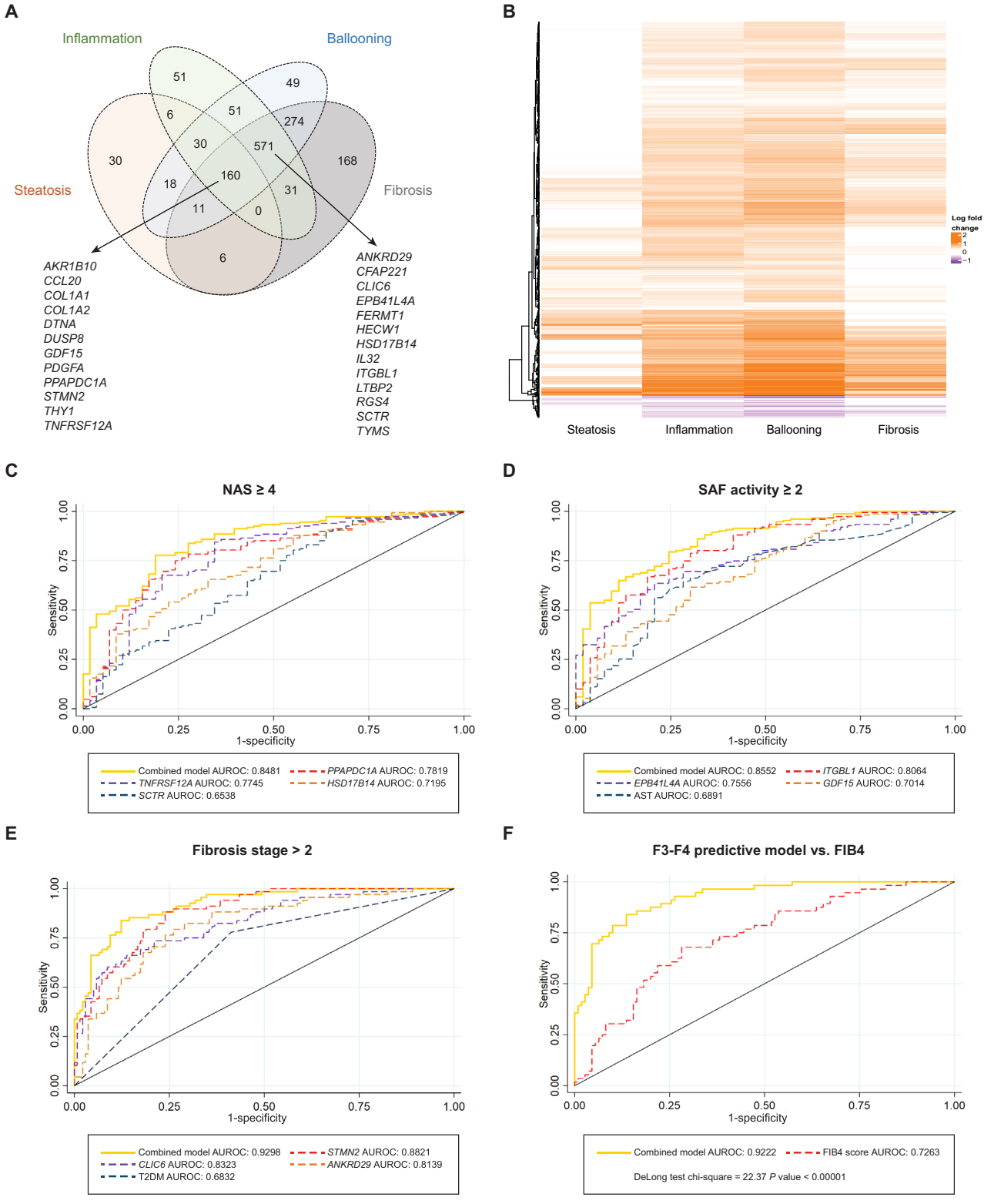

Fig. 4. Predictive modeling of histological features using the 25-gene signature associated with progressive fibrosing steatohepatitis in human NAFLD. (**A**) Venn diagram showing overlap of genes linearly associated with the increase in histological grading (adjusted $P < 0.05$). (**B**) Hierarchical clustering of genes significantly associated (adjusted $P < 0.05$) with at least three histological criteria. Each histological feature was treated as a continuous linear covariate. (**C** to **F**) Association of the 25-gene signature with disease activity and stage in the RNA-seq discovery cohort ($n = 206$) using predictive multivariate models. AUROC curves showing the combined predictive model with the single significant covariates for (C) NAS ≥4, (D) SAF activity ≥2, and (E) advanced fibrosis F3-F4. (F) Comparison of the combined predictive model for advanced fibrosis with the FIB4 score ($n = 153$ patients from the discovery cohort).

presence of advanced fibrosis F3-F4 was independently predicted by T2DM (OR = 2.7481, $P < 0.05$) combined with the expression of *ANKRD29* (OR = 2.5776, $P < 0.01$), *CLIC6* (OR = 2.2211, $P < 0.0001$), and *STMN2* (OR = 1.6397, $P = 0.001$), with an AUROC of 0.93 (Fig. 4E and table S13). This combined model to predict fibrosis had a significantly higher AUROC than the Fibrosis-4 (FIB4) score (*19*) (FIB4 AUROC, 0.73; DeLong test, $P < 0.00001$, Fig. 4F). In addition, the presence of fibrosing steatohepatitis, defined as NAS ≥ 4 and a fibrosis stage ≥F2 (NASH + NAS ≥ 4 + F ≥ 2), was predicted by the expression of *EPB41L4A*, *HSD17B14*, *PPAPDC1A*, and *TNFRSF12A* (table S13). The independent variables predicting the presence of lobular inflammation, hepatocyte ballooning, and portal inflammation are summarized in table S13. The multivariate models established using the discovery RNA-seq data were tested in the replication cohort ($n = 175$). Similar AUROCs were found in both cohorts with no significant differences (Hosmer-Lemeshow chi-square test) for the presence of lobular inflammation, ballooning, portal inflammation, NAS ≥4, SAF activity score ≥2, advanced fibrosis, and NASH + NAS ≥ 4 + F ≥ 2 (Table 2). The relevance of 25-gene signature to NAFLD and associated pathways reported in previous transcriptomics studies is described in table S14.

Integrated single-cell RNA-seq analysis identifies cell clusters on the basis of the 25-gene signature

To explore how the 25-DEG gene set changes within specific cell populations during NAFLD progression, we performed an integrated single-cell RNA-seq (scRNA-seq) analysis by projecting publicly available scRNA-seq data from healthy and cirrhotic liver samples onto the RNA-seq data grouped by fibrosis stage (discovery cohort, $n = 206$) (*20*). We identified epithelial cells, macrophages, mesenchymal cells, endothelial/sinusoidal cells, and lymphocytes within the scRNA-seq data on the basis of the expression of lineage specific markers as annotated in table S15 and fig. S8. Uniform manifold approximation and projection plots with a resolution of 0.7 identified eight distinct clusters within the epithelial cells from the healthy and cirrhotic scRNA-seq data combined (Fig. 5A and fig. S9). One specific cluster within cirrhosis showed concurrent expression of 10 markers from our 25-gene signature, including *AKR1B10*, *ANKRD29*, *CLIC6*, *DTNA*, *GDF15*, *IL32*, *PDGFA*, *RGS4*, *SCTR*, and *TNFRSF12A* (epithelial cluster 5, Fig. 5A). We projected the signatures of the different epithelial cell populations onto the NAFLD RNA-seq data using CIBERSORT and found that the epithelial cluster capturing 10 from our 25-gene signature showed a high expression in advanced NASH, with the strongest enrichment in NASH F4 (Fig. 5B).

Assessing the different macrophage populations, *HSD17B14* expression was observed in one cluster, annotated as macrophage cluster 5, in both healthy and end-stage liver disease (Fig. 5C). To further characterize this cluster, we looked at the expression of *CCR2*, *CD163*, and *TREM2*, markers for macrophages that have been reported to be relevant in fibrosis and NAFLD (*20*–*23*). The *HSD17B14*$^+$ cluster showed expression for *CD163* but was negative for *CCR2* and *TREM2* and was enriched in NAFL within the NAFLD RNA-seq data (Fig. 5D). *TREM2* expression was observed in two clusters within the macrophage populations that also showed positivity for *CD163*. The first cluster, macrophage cluster 0, was enriched in NASH F2, NASH F3, and NASH F4, whereas the second cluster, macrophage cluster 6, was mainly expressed in NASH F3 and NASH F4 (Fig. 5, C and D). Moreover, the *CD163*$^+$/*TREM2*-macrophage population identified only within the cirrhotic scRNA-seq data (macrophage cluster 7) showed an enrichment in NASH F3 and NASH F4 stages (Fig. 5D). *CCR2* expression was only observed in macrophage cluster 8, although this signature was not enriched within the NAFLD RNA-seq data (Fig. 5D). In addition, looking at the expression of *COL1A1*, *COL1A2*, *ITGBL1*, *LTBP2*, *PDGFA*, *PPAPDC1A*, *RGS4*, *STMN2*, and *TNFRSF12A*, we identified different mesenchymal populations that showed distinct enrichment depending on the stage of NAFLD disease (fig. S9). Together, integrated scRNA-seq analysis identified distinct populations on the basis of the expression of our 25-gene signature and suggests that dynamic changes within intrahepatic different macrophage populations occur during NAFLD progression.

Validation of hepatic gene expression in liver tissue by immunohistochemistry

To validate the findings from the integrated RNA-seq/scRNA-seq analysis, we selected three markers on the basis of availability of antibodies for immunohistochemistry staining of 33 formalin-fixed paraffin-embedded (FFPE) liver tissue cases derived from the NAFLD replication cohort (AKR1B10, GDF15, and STMN2). AKR1B10 protein expression was seen focally in hepatocytes in NAFL, showing a cytoplasmic and nuclear immunostaining pattern (fig. S10). In NASH, AKR1B10 positivity was more prominent in ballooned hepatocytes and in hepatocytes located neighboring necroinflammatory foci and periportal/periseptal areas. Additional, weaker immunostaining of sinusoidal lining cells was seen in most of the NAFLD cases. The number of AKR1B10 immunopositive hepatocytes increased with disease stage, peaking at F4. GDF15 staining showed a granular cytoplasmic positivity in the hepatocytes of the NAFLD samples (fig. S10). Moreover, GDF15 expression was focally observed in parenchymal immune cells. Hepatocyte immune-positivity for GDF15 was nonzonal. In addition, STMN2 immunopositivity was seen in macrophages in the portal inflammatory infiltrate with an increasing number of positive cells toward end-stage cirrhosis. Weak STMN2 expression was observed in sinusoidal lining cells in all the stages of NAFLD (fig. S10).

Serum AKR1B10 and GDF15 correlate with disease stage and histological activity score

To determine whether evidence of the hepatic transcriptomic changes could also be detected peripherally, we assessed whether circulating protein concentrations of the 25-DEG gene set accurately reflected histological disease severity as an exemplar for future potential biomarker development. Proteomics analysis was performed on 305 serum samples from patients with histologically characterized NAFLD using SomaScan technology. Thirteen proteins were detectable in the serum samples, reflecting 14 of the 25 genes, with COL1A being the protein for the genes *COL1A1* and *COL1A2*. Assessing the different histological scores, AKR1B10 was the only circulating protein showing a significant increment reflecting the increase in steatosis grade (Kruskal-Wallis test, $P = 9.66 \times 10^{-6}$; Fig. 6 and table S16). Serum AKR1B10, COL1A, and GDF15 concentrations showed significant differences with an increase in the score for ballooning (Kruskal-Wallis test, $P = 3.80 \times 10^{-14}$, $P = 1.02 \times 10^{-2}$, and $P = 2.05 \times 10^{-5}$, respectively), whereas both AKR1B10 and GDF15 were associated with high Kleiner and SAF inflammation scores. Conversely, HECW1 showed a gradual decrease

Table 2. Accuracy and validation of multivariate associations of the 25-gene signature and baseline patient characteristics in predicting histological features in the discovery ($n = 206$) and replication ($n = 175$) cohorts. AIC, Akaike information criterion; BIC, Bayesian information criterion.

		Lobular inflammation > 0		Ballooning > 0		Portal inflammation > 0		NAS ≥ 4	
		Discovery	Replication	Discovery	Replication	Discovery	Replication	Discovery	Replication
Discrimination	AUROC	0.855	0.8071	0.8662	0.7821	0.8786	0.7122	0.8572	0.7791
Calibration	Hosmer-Lemeshow chi-square	13.09	11.82	8.43	13.94	7.83	3.03	8.81	12.33
	P value	0.1089	0.1861	0.3927	0.1833	0.4499	0.9327	0.3587	0.1369
Goodness of fit	AIC	90.1467	98.67291	165.7006	140.5567	169.9194	165.766	176.7047	202.8375
	BIC	103.34	111.2396	182.1922	156.1765	199.6043	194.041	193.1963	218.6615

		SAF ≥ 2		Advance (F3/F4) vs. mild (F0/F1/F2) fibrosis		NASH + NAS ≥ 4 + F ≥ 2	
		Discovery	Replication	Discovery	Replication	Discovery	Replication
Discrimination	AUROC	0.8552	0.8426	0.9298	0.8649	0.9003	0.7541
Calibration	Hosmer-Lemeshow chi-square	8.94	3.54	4.96	9.44	12.93	11.74
	P value	0.3476	0.8962	0.7617	0.3066	0.1145	0.1634
Goodness of fit	AIC	171.5299	138.7547	145.0906	164.3298	174.549	231.4289
	BIC	188.1205	154.463	161.73	180.0673	191.1891	247.2528

with increasing NAS and SAF inflammation scores (Kruskal-Wallis test, $P < 0.05$) (Fig. 6 and table S16). Serum AKR1B10 and GDF15 were significantly increased with the rise in fibrosis stage (Kruskal-Wallis test, $P = 6.22 \times 10^{-13}$ and $P = 3.44 \times 10^{-15}$, respectively), with AKR1B10 showing a 2.19-fold increase when comparing fibrosis stage F4 with F0 and GDF15 a 2.75-fold increase (post hoc corrected, $P < 0.001$; Fig. 6). Furthermore, serum AKR1B10 and GDF15 were significantly increased in patients with NASH (1.9- and 1.35-fold change, respectively), in patients with NAS ≥4 (which conflates steatosis with activity, 2.14- and 1.35-fold change, respectively), and in patients with SAF activity ≥2 (activity of steatohepatitis, 1.91- and 1.38-fold change, respectively) (Mann-Whitney, $P < 0.0001$), whereas COL1A only captured NASH and SAF activity ≥2 (Mann-Whitney, $P < 0.05$) (Fig. 6 and table S16). When the more severe subset of patients who would meet current enrollment criteria for therapeutic trials in NASH were considered (defined as NASH + NAS ≥ 4 + F ≥ 2), a 2.19-fold increase in serum AKR1B10 and a 1.51-fold increase in serum GDF15 were observed (Mann-Whitney, $P < 0.05$). When stratifying patients on the basis of the unsupervised clustering ($n = 59$), among the detectable proteins relating to the 25-gene signature, serum AKR1B10 and GDF15 were significantly increased in patients belonging to cluster A compared with the patients from cluster B (Mann-Whitney, $P < 0.05$) (Fig. 6).

Endoplasmic reticulum stress–induced GDF15 reduces the inflammatory response in vitro

To study the functional basis of the 25 DEGs with respect to established mechanistic processes that underpin NAFLD progression, we first assessed gene expression changes in vitro after Hep G2 cells were exposed to endoplasmic reticulum (ER) stress (tunicamycin or thapsigargin) or lipid loading (oleic and palmitic acid) ($n = 3$ per group) (Fig. 7A). After 24-hour thapsigargin treatment, we observed a significant increase in mRNA of *AKR1B10* (unpaired Student's *t* test, $P < 0.001$), *CCL20* ($P < 0.05$), *DUSP8* ($P < 0.05$), *GDF15* ($P < 0.01$), *PDGFA* ($P < 0.01$), and *TNFRSF12A* ($P < 0.01$), whereas *FERMT1* ($P < 0.05$) and *TYMS* ($P < 0.001$) were reduced. Tunicamycin treatment significantly induced the expression of *AKR1B10* (unpaired Student's *t* test, $P < 0.01$) and *PDGFA* ($P < 0.05$) and reduced the expression of *CCL20* ($P < 0.05$) and *TYMS* ($P < 0.05$). No significant differences in mRNA expression were observed after treatment with palmitic, oleic, or combined palmitic/oleic acid. Western blotting confirmed the increase in AKR1B10 and GDF15 protein expression, together with an increase in the ER stress marker DNA Damage Inducible Transcript 3 (DDIT3/CHOP) in Hep G2 cells after tunicamycin and thapsigargin treatment but not after lipid loading (Fig. 7B).

To investigate the potential role of GDF15 in the inflammatory response, THP-1 monocytes were differentiated into macrophages with or without recombinant human GDF15 for 48 hours, followed by either a 6-hour lipid or lipopolysaccharide (LPS) treatment. Supplementing GDF15 significantly reduced the release of interleukin-6 (IL-6) by the macrophages into the cell culture medium upon addition of palmitic acid (unpaired Student's *t* test, $P < 0.0001$) or palmitic/oleic acid treatment ($P < 0.001$) (Fig. 7C). Furthermore, GDF15 reduced the release of TNFα in untreated cells ($P < 0.01$), in palmitic acid–loaded ($P < 0.01$) and palmitic/oleic acid–loaded ($P < 0.001$) cells, and in cells treated with LPS ($P < 0.0001$) (Fig. 7D). In addition, a reduced release of C-C motif chemokine ligand 2 (CCL2) was observed in LPS-treated ($P < 0.001$) and untreated cells ($P < 0.001$) when conditioned with GDF15, whereas a slight increase was seen when the cells were loaded with lipids ($P < 0.05$) (Fig. 7E). In sum, our functional results showed that ER stress is a strong inducer of several of our core signature genes and that GDF15 tempered the inflammatory response in macrophages in vitro upon lipid loading and LPS stimulation.

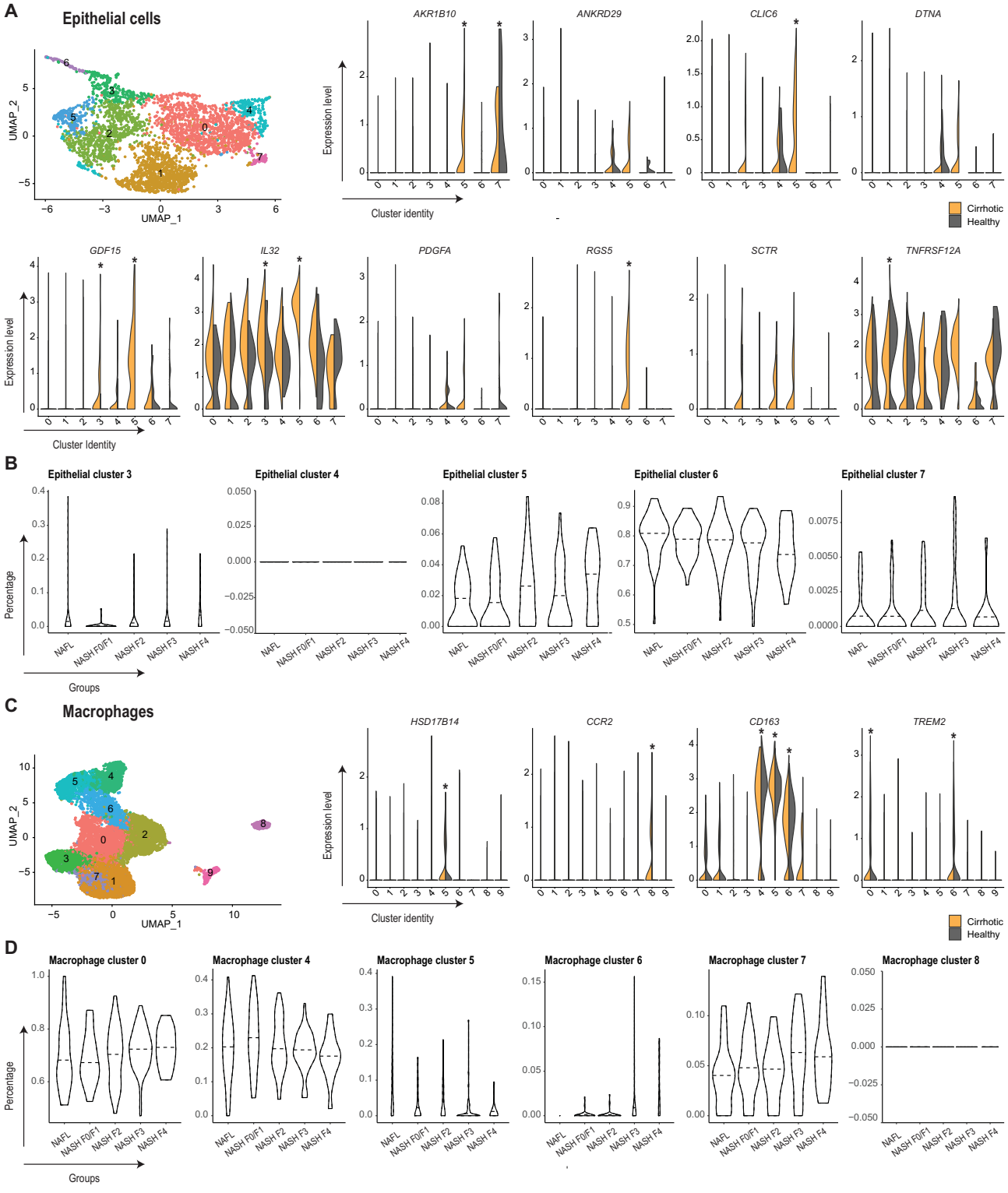

Fig. 5. Integrated scRNA-seq identifies cell populations on the basis of the 25-gene signature associated with progressive fibrosing steatohepatitis in human NAFLD. (**A**) Identification of different cell clusters within the epithelial cells using publicly available scRNA-seq data from healthy and end-stage cirrhotic liver (*20*). Expression of selected genes from our 25-gene signature in these different epithelial cell clusters (*adjusted $P < 0.001$ indicating significantly DEGs identifying the cell cluster). UMAP, uniform manifold approximation and projection. (**B**) CIBERSORT analysis to project cluster signatures onto the discovery RNA-seq dataset from 206 patients with NAFLD. (**C**) Identification of different cell clusters within the macrophage cells using scRNA-seq data from healthy and end-stage cirrhotic liver and visualization of selected genes in those clusters (*adjusted $P < 0.001$ indicating significantly DEGs identifying the cell cluster). (**D**) Projection of selected macrophage population signatures on our discovery RNA-seq data.

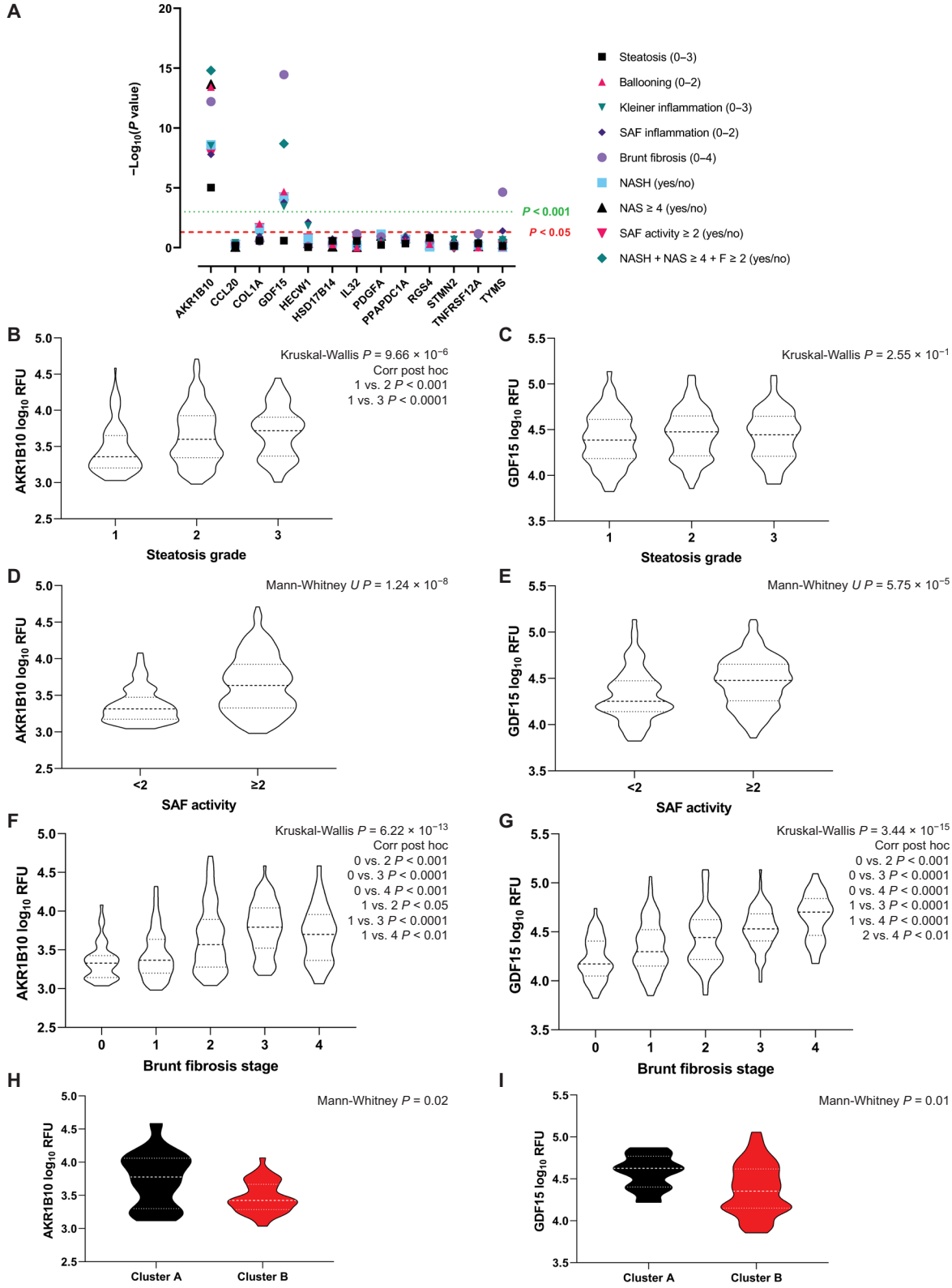

Fig. 6. Proteomics analysis of serum samples from patients with histologically proven NAFLD. Serum samples $n = 305$ are obtained from 305 patients. (**A**) Overview of detectable proteins from our 25-gene signature and their correlation with histopathological features. Correlation of AKR1B10 and GDF15 with steatosis grade (**B** and **C**), SAF activity (**D** and **E**), and Brunt fibrosis stage (**F** and **G**) for $n = 305$ serum samples. (**H** and **I**) $n = 59$ serum samples had a matching biopsy included in the discovery RNA-seq cohort, and circulating AKR1B10 and GDF15 were stratified on the basis of the unsupervised clustering. *P* values based on Kruskal-Wallis test with post hoc Bonferroni correction or Mann-Whitney *U* test. RFU, relative fluorescent units.

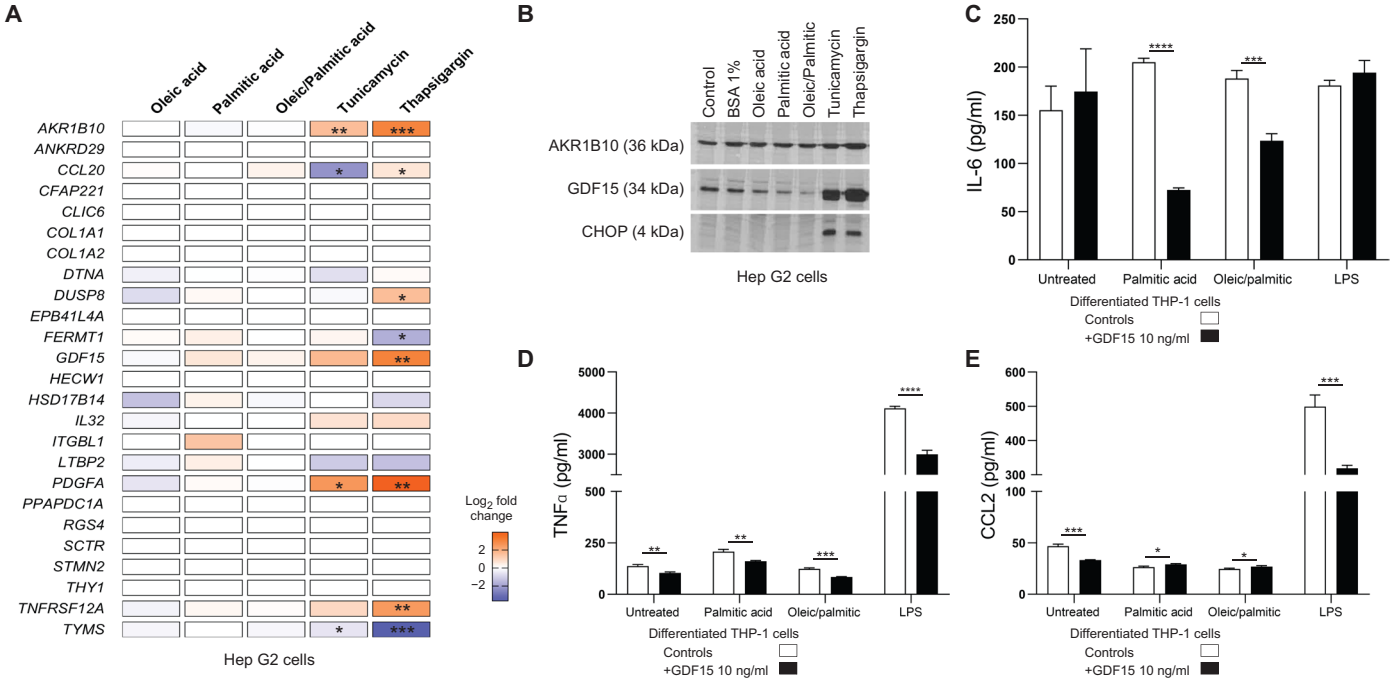

Fig. 7. In vitro functional assessment of the 25-gene signature associated with progressive fibrosing steatohepatitis in human NAFLD. (**A**) Quantitative polymerase chain reaction analysis for the 25-gene signature on Hep G2 cells treated with lipids (oleic, palmitic acid, or combined) or with ER stress inducers tunicamycin or thapsigargin (biological replicates $n = 3$ per group). Data are presented as fold change relative to the control. (**B**) Western blot analysis for AKR1B10, GDF15, and CHOP on treated Hep G2 cells. BSA, bovine serum albumin. (**C** to **E**) Enzyme-linked immunosorbent assay readout for IL-6, TNFA, and CCL2 on differentiated monocyte THP-1 cells with or without GDF15 pretreatment, challenged with palmitic acid, combined oleic/palmitic acid, or lipopolysaccharide (LPS) (biological replicates $n = 3$ per group). Data are presented as means ± SD. Unpaired Student's t test, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

Relevance of *HSD17B14* to advanced NAFLD

Genetic polymorphisms in *HSD17B13* have previously been associated with protection against more advanced steatohepatitis possibly due to modifying retinol metabolism (*24*, *25*). Because *HSD17B14* expression was inversely related to an increased risk of NAS ≥4 using multivariate predictive models in both the discovery and replication cohorts and showed expression in a specific hepatic macrophage population, we explored the role of *HSD17B14* in retinol metabolism. To characterize the oxidoreductase activity of HSD17B14, we used nicotinamide adenine dinucleotide (NAD^+)/reduced form of NAD^+ (NADH) luminescent assays. Recombinant protein HSD17B14 showed a 1.98-fold increase of NAD^+ conversion into NADH in the presence of retinol as a substrate ($P < 0.001$) and a 3.38-fold increase in the presence of the known substrate estradiol (as control) ($P < 0.001$; fig. S11) compared with the enzyme in the presence of NAD^+ alone. Similar results were observed using recombinant protein HSD17B13 as a positive control.

DISCUSSION

The patterns of hepatic gene expression during NAFLD progression provide insights into disease mechanism and may help to identify tractable therapeutic targets. Several previous studies have attempted to address transcriptomic changes in NAFLD; however, although some findings have emerged, many previous studies have been limited by use of expression microarrays that restrict gene coverage, small overall sample sizes that include very few cases with advanced disease, and, frequently, the absence of a replication cohort (*4*–*12*, *26*). Our current comprehensive transcriptomic analysis used a large independent cohort representing the full histological range of NAFLD to detect changes in hepatic gene expression as the disease progresses using RNA-seq, which has a wider dynamic range to detect gene expression changes and sets no a priori restrictions on gene coverage. The results of this study share some commonality with the existing literature, for example, highlighting the relevance of bile acid metabolism and the FXR/CYP7A1 axis in NASH pathogenesis, as well as differential expression of *AKR1B10*, *CCL20*, *COL1A1*, *COL1A2*, *DUSP8*, *IL32*, *ITGBL1*, *STMN2*, *THY1*, and *TYMS*. These data substantially extend knowledge of the transcriptomic profile of NAFL-NASH and provide greater granularity across intermediate grades and stages of disease. Adopting a hypothesis-generating pairwise analysis strategy, we identified a 25-gene set associated with the transition from NAFL to NASH and onward progression to fibrosis and cirrhosis. We also demonstrated that 21 of these 25 genes discriminated mild from advanced disease (NAFL-NASH F0/F1 versus NASH F ≥ 2) in an independent replication cohort and so have transcriptionally defined a key group of high-risk patients who are most likely to progress to advanced disease or experience clinical events and so should be investigated as potential therapeutic targets (*27*).

Only a few of the 25 genes can be considered as true extracellular matrix genes (for example, *COL1A1* and *COL1A2*) or markers for hepatic stellate cell activation and fibrogenesis (*ITGBL1* and *STMN2*) (*28*, *29*). We found that expression of inflammatory genes *GDF15*, *CCL20*, and *IL32* increased with disease progression, although

these ligands/chemokines have been reported to be protective against features of advanced liver disease in animal models (*30*–*32*). In high-fat diet–fed mice, ectopic expression of *Gdf15* has been reported to reduce lipid accumulation by enhancing hepatic fatty acid oxidation, whereas overexpression of *Il32* ameliorates steatosis and inflammation in an NAFLD mouse model (*30*, *32*). Moreover, parenchymal expression of Ccl20 improves hepatic fibrosis through the recruitment of γδ T cells in chronic carbon tetrachloride mouse models (*31*). Similarly, our data suggest that ER stress–induced release of GDF15 may ameliorate the inflammatory response in macrophages.

Other genes such as *AKR1B10* or *HSD17B14* have primarily been described to have a metabolic function (*26*, *33*). AKR1B10 is an aldo/keto reductase that converts retinal into retinol (vitamin A1) and has been reported to be critical for cell survival in vitro through modulation of lipid metabolism and mitochondrial function (*33*). Moreover, AKR1B10 expression is induced by the transcription factor NRF2, which activates protective pathways in response to oxidative stress (*34*). ER stress has also been reported to induce the transcriptional activity of NRF2 to promote cell survival (*35*). Our in vitro data showed that ER stress not only induced the expression of AKR1B10 but also GDF15, which is in line with previous reports (*36*, *37*). This would suggest that some of our core signature genes are expressed by epithelial cells to deal with oxidative or ER stress and to help resolve the inflammatory response in advanced NAFLD. Controversially, ER stress has been described to also induce steatosis in NAFLD, meaning that how hepatocytes deal with lipid-induced stress could actually further induce steatosis (*38*). Although the release of retinol during stellate cell activation has been well documented, its effect in other hepatic cells is less clear (*39*). Recently, it was reported that the polymorphism rs6834314, which is protective against NAFLD, confers a loss of enzymatic activity of HSD17B13 toward retinol (*25*). Our results showed a gradual decrease in expression in macrophages of *HSD17B14*, another member of the 17-beta-hydroxysteroid dehydrogenase family, as NAFLD progresses. This finding could indicate a role for retinol metabolism in hepatic macrophages during the progression of NAFLD. Increased retinol and decreased retinal mean that less conversion of retinal to the biologically active retinoic acid isomers (all-trans and 9-cis retinoic acid) is possible, which may affect nuclear receptor Retinoic Acid Receptor (RAR)/Retinoid X Receptor (RXR) signaling. A number of different enzymes contribute to these processes in addition to AKR1B10, HSD17B13, and HSD17B14, but the large increase in AKR1B10 expression as NAFLD progresses together with the decrease in HSD17B14 would tend to favor increased retinol. Whether a metabolic shift from retinal to retinol is a way for hepatocytes to survive in a background of chronic lipid-induced ER stress or is a means of communicating with stellate cells and macrophages in the hepatic microenvironment is still not clear.

Despite the wide variety in function of the 25 genes, they all showed a strong collinearity with ballooning, inflammation, and fibrosis, suggesting a strong connection between cell damage, inflammation, and tissue scarring. To understand how these genes drive disease, we used predictive multivariate models to identify independent variables associated with specific histological features. The variables predicting fibrosis were different from the variables predicting a high disease activity, either based on the NAS or the SAF activity score, meaning that these variables could be regarded as proper independent drivers of disease features.

From a clinical point of view, there is a clear imperative to develop better noninvasive means to diagnose and stratify patients for treatment or enrollment into clinical trials without the need for a liver biopsy. In recent years, a great effort has been made to identify soluble serum markers to predict the presence of NASH or advanced liver fibrosis, such as collagen turnover biomarkers (*40*). Several of the markers that we detected in our 25-gene signature proved to be candidate serum markers of advanced NAFLD, further demonstrating the robustness of this signature. Previous reports support these findings: Serum AKR1B10 has been shown to correlate with advanced disease in a small Japanese cohort of patients with NAFLD, whereas increased serum GDF15 has been reported to associate with a greater risk of advanced fibrosis in a South Korean study (*41*, *42*). In this study, predictive univariate and multivariate models highlighted the potential of several additional genes as markers for disease activity or fibrosis grading. As a proof of concept, we showed that serum AKR1B10 and GDF15 not only stratified patients with more active steatohepatitis but also discriminated between NAFL-NASH F0/F1 and NASH F $\geq$ 2: the subset of patients with NAFLD who are at greatest risk of future disease progression and would arguably be best targeted for therapy (*27*). Furthermore, serum AKR1B10 and GDF15 were elevated in the unsupervised cluster A, characterized by high disease activity and high fibrosis.

Transcriptomic staging showed features such as senescence, DNA damage, autophagy, and bile secretion/FXR signaling, which one might presume develop late in disease natural history once steatohepatitis was present, were already observed in NAFL. For example, *CYP7A1*, the rate-limiting enzyme in the classical bile acid synthesis pathway that is subject to FXR-mediated regulation, was already up-regulated in NAFL, with expression peaking during the early stages of steatohepatitis (*18*). This possibly suggests that the therapeutic window for response to FXR agonists may range from NAFL onward, although this will require further validation. In contrast, *CCL2* expression was increased in the unsupervised cluster A characterized by high disease activity and concordant expression of extracellular matrix/fibrosis– and hepatic progenitor cell–related genes (*43*). Therapeutic inhibition of CCR2-positive monocyte-derived macrophages has been reported to reduce inflammation and fibrosis in murine NASH and fibrosis in human disease (*21*, *44*). Furthermore, we found that the expression of *TREM2*, a marker for macrophages reported to be involved in hepatic fibrosis, was associated with high disease activity and was most highly expressed in advanced NAFLD (NASH F2-F4) (*20*, *23*). Although *TREM2* is not part of our core gene signature, our integrated scRNA-seq analysis indicated the importance and dynamics of specific macrophage cell populations during NAFLD progression. Different TREM2$^+$ cell clusters, as well as CD163$^+$ clusters, were enriched during different stages of NAFLD. Likewise, we identified different subpopulations of mesenchymal cells on the basis of our core gene signature during NAFLD progression. These findings suggest that therapeutic interventions may be most efficacious if their use is targeted to the specific transcriptomic patterns that occur as NAFLD progresses; serum markers may be useful indicators to identify pathways susceptible to future treatments.

There are some limitations to this study. Although histology remains the optimum approach and accepted reference standard to accurately grade steatohepatitis and stage fibrosis, it is subject to sampling error. Four of the 25 genes were not differently expressed in our replication cohort (*ANKRD29*, *EPB41L4A*, *TYMS*, and *HSD17B14*).

This lack of replication could reflect subtle differences in phenotype between the cohorts, for example, minor differences in histological severity of the piece of biopsy core from which RNA was extracted compared with the part examined by microscopy given that disease is not completely homogeneous within the liver. However, others have previously described an increased *TYMS* expression in advanced NASH, and, using a different approach to assess risk of NAFLD progression, we demonstrated that *HSD17B14* and *EPB41L4A* expression are useful predictors of high NAS and SAF scores, respectively (*7*, *26*). In this study, we observed differences between the unsupervised clusters in carriage of the *PNPLA3* rs738409 variant but not for genotypes *GCKR* rs1260326, *HSD17B13* rs72613567, or *TM6SF2* rs58542926 (*16*). Nevertheless, the carriage of these genotypes did not appear to confer distinct differences in gene expression in the RNA-seq analysis. One possible explanation is that we are only looking within a relatively small NAFLD population, at least compared with GWAS studies. A more likely explanation, however, is that these polymorphisms contribute to the initial susceptibility to NAFLD but do not change the nature of the pathogenic processes that are in play during disease progression. Furthermore, although we confirmed that proteins encoded by many of the mRNAs in our 25-gene signature were detectable in serum and showed changes similar to those at the mRNA level during NAFLD progression, we could not confirm this for all 25 mRNA species. Detection limits of the technique used or limitations in the type of sample available could potentially explain this.

In conclusion, the current study identified a number of changes in gene expression during NAFLD progression, including several that may be of diagnostic and prognostic relevance: For example, small but functionally important changes in gene expression occurred in early NAFL. We also confirmed that genes such as *AKR1B10* and *GDF15* are consistent markers for NAFLD progression. AKR1B10 and HSD17B14 may contribute to retinoic acid homeostasis, which, based on recent findings such as the role for HSD17B13 in NAFLD genetic susceptibility, increasingly seems relevant to NAFLD progression in addition to carcinogenesis (*41*). These transcriptomic data, from a large histologically characterized NAFLD cohort, provide insights into disease pathophysiology, identifying both stable and dynamic differences in gene expression that occur during NAFLD progression.

MATERIALS AND METHODS

Study design

This study aimed to elucidate the pathophysiology of progressive fibrosing steatohepatitis in NAFLD using a transcriptomic analysis approach. A total of 414 NAFLD biopsies and 305 NAFLD serum samples covering the full histological disease spectrum were included in this multicenter study. The discovery cohort of 206 NAFLD samples was processed for RNA-seq, whereas the replication cohort including 175 patients with NAFLD was used for NanoString analysis ($n = 175$) and immunohistochemistry ($n = 33$) (Fig. 1). Detailed phenotypic description and demographics are reported in Table 1 and table S10. Both cohorts were stratified according to histopathological disease grade and stage, including NAFL, NASH F0, NASH F1, NASH F2, NASH F3, and NASH F4. Logistic modeling was used to correlate gene expression with histological features. Potentially tractable and clinically relevant disease biomarkers were tested in a cohort of 305 serum samples. To study the functional basis of core disease-associated gene signatures, we used in vitro cell line models. Additional materials and methods can be found in the Supplementary Materials.

Patient selection

A total of 436 liver biopsy samples (414 NAFLD and 22 controls) from 403 European Caucasian patients were included in this study. In addition, 305 serum samples (including 108 for which contemporaneous transcriptomics data were available) were also used. Cases were derived from the European NAFLD Registry (NCT04442334) (*45*). The discovery cohort comprised 216 snap-frozen biopsy samples from 206 patients diagnosed with NAFLD in France, Germany, Italy, or the United Kingdom and 10 healthy obese control cases without any biochemical or histological evidence of NAFLD from patients undergoing bariatric surgery in France. These patients were selected on the basis of both study participation and ability to isolate sufficient high-quality RNA for sequencing from the liver biopsy. The replication cohort consisted of 220 FFPE and frozen samples from 175 patients with NAFLD (59 FFPE and 116 frozen) diagnosed in France and the United Kingdom and 12 healthy obese control cases (frozen). All samples were centrally scored by two expert liver pathologists (D.T. and P.B.) according to the semiquantitative NASH-Clinical Research Network NAS and the FLIP SAF scoring system (*14*, *15*). Fibrosis was staged from F0 to F4 (cirrhosis). Serum samples collected within 6 months of the biopsy date were available from 305 patients with histologically proven NAFLD diagnosed in France, Germany, Italy, Sweden, or the United Kingdom. Fifty-nine serum samples of the 305 matched with patients and biopsies enrolled in the discovery cohort. Patients with alternate diagnoses and etiologies were excluded, including excessive alcohol intake (30 g per day for males and 20 g per day for females), viral hepatitis, autoimmune liver diseases, and steatogenic medication use. NAFL samples with a fibrosis stage of ≥ 2 were not included in this study. As previously described (*45*), collection and use of samples and clinical data for this study were approved by the relevant local and/or national Ethical Review Committee covering each participating center, with all patients providing informed consent for participation. All participant recruitment and informed consent processes at recruitment centers were conducted in compliance with nationally accepted practice in the respective territory and in accordance with the World Medical Association Declaration of Helsinki 2018.

Bioinformatics

FastQC (v0.11.5) and MultiQC (v1.2dev) were used to establish raw sequencing quality. Alignment to the reference genome (GRCh38, Ensembl release 76) was performed using STAR. Gene-level count tables were produced using HTSeq. Counts were normalized using the trimmed mean of M values method and transformed using limma's voom methodology. Normalized and transformed counts were analyzed for differential expression using linear models as implemented by limma (*46*). Statistical significance of protein-coding genes was determined by an false discovery rate (FDR)–corrected $q < 0.05$ and fold change of $> |1.5|$ (*47*). Confounding effects were corrected for by inclusion as additive effects in the linear model used to determine differential expression. For visualization only, additive effects were subtracted from the expression data using limma's removeBatchEffects function. For each grouped comparison, a correction for batch effect and gender was applied; in the comparisons excluding the controls, an additional correction for center was implemented. *PNPLA3* rs738409, *TM6SF2*

rs58542926, and *GCKR* rs1260326 were determined using the RNA-seq reads, *HSD17B13* rs72613567 single-nucleotide polymorphism (SNP) genotypes, or using TaqMan probes (Applied Biosystems) on DNA from peripheral blood mononuclear cells. If a suitable assay could not be designed, a proxy SNP was chosen (https://ldlink.nci.nih.gov/). Database for Annotation, Visualization and Integrated Discovery (DAVID) annotation tool was used for pathway enrichment (*48*, *49*). Data were visualized with GOplot 1.0.2 and clusterProfiler (*50*, *51*).

Integrated analysis was performed using scRNA-seq data (GSE136103) from healthy liver and end-stage liver disease samples (*20*). Filtering was applied to remove any cells with greater than 30% mitochondrial genes or fewer than 300 total genes. The cells were normalized and clustered using Seurat as described previously (*20*). A clustering resolution of 0.2 was used resulting in 13 clusters, and cell types were annotated on the basis of the expression of specific gene markers. Cell types were then clustered at a higher resolution of 1.2. CIBERSORT was used to determine cell type abundance within the bulk RNA-seq data (*52*).

Statistical analysis

Data were tested for normality using the Kolmogorov-Smirnov/Shapiro-Wilk normality test using IBM SPSS software (IBM Corp). Unpaired Student's *t* test, Mann-Whitney *U* test and Kruskal-Wallis test were performed using IBM SPSS software. Post hoc testing for paired comparisons after the nonparametric Kruskal-Wallis test was adjusted for multiple testing using the Bonferroni correction. Data were visualized using GraphPad Prism 8.0.1 (GraphPad Software Inc.). A $P < 0.05$ was considered significant. Multivariable logistic regression was conducted using Stata software release 15 (StataCorp) on variables that reached $P < 0.1$ at univariate analysis. The final multivariate logistic regression model was built from the set of candidate genes and variables by removing predictors on the basis of their *P* values, in a stepwise manner. Multicollinearity among variables was detected using Spearman's correlation. For each multivariable logistic regression, model discrimination and calibration were reported together with Akaike information criterion and Bayesian information criterion measures for comparing maximum likelihood models. Model discrimination was assessed calculating the AUROC curve (or C statistic), model calibration was determined by Hosmer-Lemeshow technique and the goodness of fit of the models was assessed by Akaike information criterion. We made an a priori decision to consider only models with AUROC C statistic ≥ 0.75 as adequate. To compare effects across nested logistic regression models, the likelihood ratio test was used; to test the equality of the areas under the curve, the DeLong test was used.

SUPPLEMENTARY MATERIALS

stm.sciencemag.org/cgi/content/full/12/572/eaba4448/DC1

Materials and Methods

Fig. S1. PCA plot analysis using RNA-seq data from the discovery cohort.

Fig. S2. Effect of genotype on mRNA expression within the NAFLD discovery cohort.

Fig. S3. Unsupervised clustering of NALD RNA-seq cohort.

Fig. S4. Supervised clustering of NALD RNA-seq cohort.

Fig. S5. Pathview enrichment and candidate gene analysis in the NAFLD-control comparison.

Fig. S6. GO annotation pairwise analysis using the NAFLD discovery cohort.

Fig. S7. Nanostring analysis using the replication cohort.

Fig. S8. Integrated scRNA-seq analysis.

Fig. S9. Expression of 25-gene signature in scRNA-seq cell clusters.

Fig. S10. Immunohistochemistry for AKR1B10, GDF15, and STMN2.

Fig. S11. Enzymatic activity of HSD17B14 and HSD17B13 against estradiol and retinol.

Table S1. Top 250 DEGs comparing unsupervised cluster A with cluster B.

Table S2. Clinicopathological features unsupervised clustering B1 and B2.

Table S3. DEGs comparing NASH F2 to baseline NAFL using discovery cohort RNA-seq data.

Table S4. DEGs comparing NASH F3 to baseline NAFL using discovery cohort RNA-seq data.

Table S5. DEGs comparing NASH F4 to baseline NAFL using discovery cohort RNA-seq data.

Table S6. DEGs comparing NASH F3 to baseline NASH F0/F1 using discovery cohort RNA-seq data.

Table S7. DEGs comparing NASH F4 to baseline NASH F0/F1 using discovery cohort RNA-seq data.

Table S8. DEGs associated with NAS ≥4 based on RNA-seq data.

Table S9. DEGs associated with SAF activity ≥2 based on RNA-seq data.

Table S10. Demographics of discovery, replication, and biomarker cohorts.

Table S11. Correlation of the 25-gene signature with discovery cohort histological features.

Table S12. Logistical univariate analyses of the 25-gene signature.

Table S13. Logistical multivariate analyses of the 25-gene signature.

Table S14. Previous reports on the relevance of 25-gene signature members to NAFLD.

Table S15. Annotation clusters scRNA-seq data.

Table S16. Proteomics analysis of NAFLD serum samples.

Data file S1. Individual-level data from figures.

Reference (*53*).

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. Z. Younossi, Q. M. Anstee, M. Marietti, T. Hardy, L. Henry, M. Eslam, J. George, E. Bugianesi, Global burden of NAFLD and NASH: Trends, predictions, risk factors and prevention. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 11–20 (2018).
2. Q. M. Anstee, H. L. Reeves, E. Kotsiliti, O. Govaere, M. Heikenwalder, From NASH to HCC: Current concepts and future challenges. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 411–428 (2019).
3. S. McPherson, T. Hardy, E. Henderson, A. D. Burt, C. P. Day, Q. M. Anstee, Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: Implications for prognosis and clinical management. *J. Hepatol.* **62**, 1148–1155 (2015).
4. J. Starmann, M. Fälth, W. Spindelböck, K.-L. Lanz, C. Lackner, K. Zatloukal, M. Trauner, H. Sültmann, Gene expression profiling unravels cancer-related hepatic molecular signatures in steatohepatitis but not in steatosis. *PLOS ONE* **7**, e46584 (2012).
5. C. A. Moylan, H. Pang, A. Dellinger, A. Suzuki, M. E. Garrett, C. D. Guy, S. K. Murphy, A. E. Ashley-Koch, S. S. Choi, G. A. Michelotti, D. D. Hampton, Y. Chen, H. L. Tillmann, M. A. Hauser, M. F. Abdelmalek, A. M. Diehl, Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology* **59**, 471–482 (2014).
6. B. M. Arendt, E. M. Comelli, D. W. Ma, W. Lou, A. Teterina, T. Kim, S. K. Fung, D. K. Wong, I. McGilvray, S. E. Fischer, J. P. Allard, Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. *Hepatology* **61**, 1565–1578 (2015).
7. A. Teufel, T. Itzel, W. Erhart, M. Brosch, X. Y. Wang, Y. O. Kim, W. von Schönfels, A. Herrmann, S. Brückner, F. Stickel, J.-F. Dufour, T. Chavakis, C. Hellerbrand, R. Spang, T. Maass, T. Becker, S. Schreiber, C. Schafmayer, D. Schuppan, J. Hampe, Comparison of gene expression patterns between mouse models of nonalcoholic fatty liver disease and liver tissues from patients. *Gastroenterology* **151**, 513–525.e0 (2016).
8. P. Lefebvre, F. Lalloyer, E. Baugé, M. Pawlak, C. Gheeraert, H. Dehondt, J. Vanhoutte, E. Woitrain, N. Hennuyer, C. Mazuy, M. Bobowski-Gérard, F. P. Zummo, B. Derudas, A. Driessen, G. Hubens, L. Vonghia, W. J. Kwanten, P. Michielsen, T. Vanwolleghem, J. Eeckhoute, A. Verrijken, L. Van Gaal, S. Francque, B. Staels, Interspecies NASH disease activity whole-genome profiling identifies a fibrogenic role of PPARα-regulated dermatopontin. *JCI Insight* **2**, e92264 (2017).
9. J. T. Haas, L. Vonghia, D. A. Mogilenko, A. Verrijken, O. Molendi-Coste, S. Fleury, A. Deprince, A. Nikitin, E. Woitrain, L. Ducrocq-Geoffroy, S. Pic, B. Derudas, H. Dehondt, C. Gheeraert, L. Van Gaal, A. Driessen, P. Lefebvre, B. Staels, S. Francque, D. Dombrowicz, Transcriptional network analysis implicates altered hepatic immune function in NASH development and resolution. *Nat. Metab.* **1**, 604–614 (2019).
10. S. Zhao, W. P. Fung-Leung, A. Bittner, K. Ngo, X. Liu, Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLOS ONE* **9**, e78644 (2014).
11. G. S. Gerhard, C. Legendre, C. D. Still, X. Chu, A. Petrick, J. K. DiStefano, Transcriptomic profiling of obesity-related nonalcoholic steatohepatitis reveals a core set of fibrosis-specific genes. *J. Endocr. Soc.* **2**, 710–726 (2018).
12. M. P. Suppli, K. T. G. Rigbolt, S. S. Veidal, S. Heeboll, P. L. Eriksen, M. Demant, J. I. Bagger, J. C. Nielsen, D. Oró, S. W. Thrane, A. Lund, C. Strandberg, M. J. Kønig, T. Vilsboll, N. Vrang, K. L. Thomsen, H. Grønbæk, J. Jelsing, H. H. Hansen, F. K. Knop, Hepatic transcriptome signatures in patients with varying degrees of nonalcoholic fatty liver disease compared with healthy normal-weight individuals. *Am. J. Physiol. Gastrointest. Liver Physiol.* **316**, G462–G472 (2019).

13. S. A. Hoang, A. Oseini, R. E. Feaver, B. K. Cole, A. Asgharpour, R. Vincent, M. Siddiqui, M. J. Lawson, N. C. Day, J. M. Taylor, B. R. Wamhoff, F. Mirshahi, M. J. Contos, M. Idowu, A. J. Sanyal, Gene expression predicts histological severity and reveals distinct molecular profiles of nonalcoholic fatty liver disease. *Sci. Rep.* **9**, 12541 (2019).
14. D. E. Kleiner, E. M. Brunt, M. Van Natta, C. Behling, M. J. Contos, O. W. Cummings, L. D. Ferrell, Y. C. Liu, M. S. Torbenson, A. Unalp-Arida, M. Yeh, A. J. McCullough, A. J. Sanyal, N. Nonalcoholic; Nanolcoholic Steatohepatitis Clinical Research Network, Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**, 1313–1321 (2005).
15. P. Bedossa, C. Poitou, N. Veyrie, J.-L. Bouillot, A. Basdevant, V. Paradis, J. Tordjman, K. Clement, Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* **56**, 1751–1759 (2012).
16. Q. M. Anstee, R. Darlay, S. Cockell, M. Meroni, O. Govaere, D. Tiniakos, A. D. Burt, P. Bedossa, J. Palmer, Y.-L. Liu, G. P. Aithal, M. Allison, H. Yki-Järvinen, M. Vacca, J.-F. Dufour, P. Invernizzi, D. Prati, M. Ekstedt, S. Kechagias, S. Francque, S. Petta, E. Bugianesi, K. Clement, V. Ratziu, J. M. Schattenberg, L. Valenti, C. P. Day, H. J. Cordell, A. K. Daly; EPoS Consortium Investigators, Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically-characterised cohort. *J. Hepatol.* **73**, 505–515 (2020).
17. G. Parthasarathy, X. Revelo, H. Malhi, Pathogenesis of nonalcoholic steatohepatitis: An overview. *Hepatol. Commun.* **4**, 478–492 (2020).
18. J. Y. Chiang, R. Kimmel, C. Weinberger, D. Stroup, Farnesoid X receptor responds to bile acids and represses cholesterol 7α-hydroxylase gene (*CYP7A1*) transcription. *J. Biol. Chem.* **275**, 10918–10924 (2000).
19. J. K. Dyson, S. McPherson, Q. M. Anstee, Non-alcoholic fatty liver disease: Non-invasive investigation and risk stratification. *J. Clin. Pathol.* **66**, 1033–1045 (2013).
20. P. Ramachandran, R. Dobie, J. R. Wilson-Kanamori, E. F. Dora, B. E. P. Henderson, N. T. Luu, J. R. Portman, K. P. Matchett, M. Brice, J. A. Marwick, R. S. Taylor, M. Efremova, R. Vento-Tormo, N. O. Carragher, T. J. Kendall, J. A. Fallowfield, E. M. Harrison, D. J. Mole, S. J. Wigmore, P. N. Newsome, C. J. Weston, J. P. Iredale, F. Tacke, J. W. Pollard, C. P. Ponting, J. C. Marioni, S. A. Teichmann, N. C. Henderson, Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* **575**, 512–518 (2019).
21. O. Krenkel, T. Puengel, O. Govaere, A. T. Abdallah, J. C. Mossanen, M. Kohlhepp, A. Liepelt, E. Lefebvre, T. Luedde, C. Hellerbrand, R. Weiskirchen, T. Longerich, I. G. Costa, Q. M. Anstee, C. Trautwein, F. Tacke, Therapeutic inhibition of inflammatory monocyte recruitment reduces steatohepatitis and liver fibrosis. *Hepatology* **67**, 1270–1283 (2018).
22. C. Rosso, K. Kazankov, R. Younes, S. Esmaili, M. Marietti, M. Sacco, F. Carli, M. Gaggini, F. Salomone, H. J. Møller, M. L. Abate, H. Vilstrup, A. Gastaldelli, J. George, H. Grønbæk, E. Bugianesi, Crosstalk between adipose tissue insulin resistance and liver macrophages in non-alcoholic fatty liver disease. *J. Hepatol.* **71**, 1012–1021 (2019).
23. X. Xiong, H. Kuang, S. Ansari, T. Liu, J. Gong, S. Wang, X.-Y. Zhao, Y. Ji, C. Li, L. Guo, L. Zhou, Z. Chen, P. Leon-Mimila, M. T. Chung, K. Kurabayashi, J. Opp, F. Campos-Pérez, H. Villamil-Ramírez, S. Canizales-Quinteros, R. Lyons, C. N. Lumeng, B. Zhou, L. Qi, A. Huertas-Vazquez, A. J. Lusis, X. Z. S. Xu, S. Li, Y. Yu, J. Z. Li, J. D. Lin, Landscape of intercellular crosstalk in healthy and NASH liver revealed by single-cell secretome gene analysis. *Mol. Cell* **75**, 644–660 e645 (2019).
24. N. S. Abul-Husn, X. Cheng, A. H. Li, Y. Xin, C. Schurmann, P. Stevis, Y. Liu, J. Kozlitina, S. Stender, G. C. Wood, A. N. Stepanchick, M. D. Still, S. McCarthy, C. O'Dushlaine, J. S. Packer, S. Balasubramanian, N. Gosalia, D. Esopi, S. Y. Kim, S. Mukherjee, A. E. Lopez, E. D. Fuller, J. Penn, X. Chu, J. Z. Luo, U. L. Mirshahi, D. J. Carey, C. D. Still, M. D. Feldman, A. Small, S. M. Damrauer, D. J. Rader, B. Zambrowicz, W. Olson, A. J. Murphy, I. B. Borecki, A. R. Shuldiner, J. G. Reid, J. D. Overton, G. D. Yancopoulos, H. H. Hobbs, J. C. Cohen, O. Gottesman, T. M. Teslovich, A. Baras, T. Mirshahi, J. Gromada, F. E. Dewey, A protein-truncating *HSD17B13* variant and protection from chronic liver disease. *N. Engl. J. Med.* **378**, 1096–1106 (2018).
25. Y. Ma, O. V. Belyaeva, P. M. Brown, K. Fujita, K. Valles, S. Karki, Y. S. de Boer, C. Koh, Y. Chen, X. Du, S. K. Handelman, V. Chen, E. K. Speliotes, C. Nestlerode, E. Thomas, D. E. Kleiner, J. M. Zmuda, A. J. Sanyal; Nonalcoholic Steatohepatitis Clinical Research Network, N. Y. Kedishvili, T. J. Liang, Y. Rotman, 17-beta hydroxysteroid dehydrogenase 13 is a hepatic retinol dehydrogenase associated with histological features of nonalcoholic fatty liver disease. *Hepatology* **69**, 150–1519 (2018).
26. M. Ryaboshapkina, M. Hammar, Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Sci. Rep.* **7**, 12361 (2017).
27. M. S. Siddiqui, S. A. Harrison, M. F. Abdelmalek, Q. M. Anstee, P. Bedossa, L. Castera, L. Dimick-Santos, S. L. Friedman, K. Greene, D. E. Kleiner, S. Megnien, B. A. Neuschwander-Tetri, V. Ratziu, E. Schabel, V. Miller, A. J. Sanyal; Liver Forum Case Definitions Working Group, Case definitions for inclusion and analysis of endpoints in clinical trials for nonalcoholic steatohepatitis through the lens of regulatory science. *Hepatology* **67**, 2001–2012 (2018).
28. M. Wang, Q. Gong, J. Zhang, L. Chen, Z. Zhang, L. Lu, D. Yu, Y. Han, D. Zhang, P. Chen, X. Zhang, Z. Yuan, J. Huang, X. Zhang, Characterization of gene expression profiles in HBV-related liver fibrosis patients and identification of ITGBL1 as a key regulator of fibrogenesis. *Sci. Rep.* **7**, 43446 (2017).
29. V. Paradis, D. Dargere, Y. Bieche, T. Asselah, P. Marcellin, M. Vidaud, P. Bedossa, SCG10 expression on activation of hepatic stellate cells promotes cell motility through interference with microtubules. *Am. J. Pathol.* **177**, 1791–1797 (2010).
30. D. Li, H. Zhang, Y. Zhong, Hepatic GDF15 is regulated by CHOP of the unfolded protein response and alleviates NAFLD progression in obese mice. *Biochem. Biophys. Res. Commun.* **498**, 388–394 (2018).
31. L. Hammerich, J. M. Bangen, O. Govaere, H. W. Zimmermann, N. Gassler, S. Huss, C. Liedtke, I. Prinz, S. A. Lira, T. Luedde, T. Roskams, C. Trautwein, F. Heymann, F. Tacke, Chemokine receptor CCR6-dependent accumulation of γδ T cells in injured liver restricts hepatic inflammation and fibrosis. *Hepatology* **59**, 630–642 (2014).
32. D. H. Lee, J. E. Hong, H.-M. Yun, C. J. Hwang, J. H. Park, S. B. Han, D. Y. Yoon, M. J. Song, J. T. Hong, Interleukin-32β ameliorates metabolic disorder and liver damage in mice fed high-fat diet. *Obesity* **23**, 615–622 (2015).
33. C. Wang, R. Yan, D. Luo, K. Watabe, D.-F. Liao, D. Cao, Aldo-keto reductase family 1 member B10 promotes cell survival by regulating lipid synthesis and eliminating carbonyls. *J. Biol. Chem.* **284**, 26742–26748 (2009).
34. L. E. Tebay, H. Robertson, S. T. Durant, S. R. Vitale, T. M. Penning, A. T. Dinkova-Kostova, J. D. Hayes, Mechanisms of activation of the transcription factor Nrf2 by redox stressors, nutrient cues, and energy status and the pathways through which it attenuates degenerative disease. *Free Radic. Biol. Med.* **88**, 108–146 (2015).
35. S. B. Cullinan, D. Zhang, M. Hannink, E. Arvisais, R. J. Kaufman, J. A. Diehl, Nrf2 is a direct PERK substrate and effector of PERK-dependent cell survival. *Mol. Cell. Biol.* **23**, 7198–7209 (2003).
36. M. Parafati, R. J. Kirby, S. Khorasanizadeh, F. Rastinejad, S. Malany, A nonalcoholic fatty liver disease model in human induced pluripotent stem cell-derived hepatocytes, created by endoplasmic reticulum stress-induced steatosis. *Dis. Model. Mech.* **11**, dmm033530 (2018).
37. A. P. Coll, M. Chen, P. Taskar, D. Rimmington, S. Patel, J. A. Tadross, I. Cimino, M. Yang, P. Welsh, S. Virtue, D. A. Goldspink, E. L. Miedzybrodzka, A. R. Konopka, R. R. Esponda, J. T. Huang, Y. C. L. Tung, S. Rodriguez-Cuenca, R. A. Tomaz, H. P. Harding, A. Melvin, G. S. H. Yeo, D. Preiss, A. Vidal-Puig, L. Vallier, K. S. Nair, N. J. Wareham, D. Ron, F. M. Gribble, F. Reimann, N. Sattar, D. B. Savage, B. B. Allan, S. O'Rahilly, GDF15 mediates the effects of metformin on body weight and energy balance. *Nature* **578**, 444–448 (2020).
38. C. Lebeaupin, D. Vallee, Y. Hazari, C. Hetz, E. Chevet, B. Bailly-Maitre, Endoplasmic reticulum stress signalling and the pathogenesis of non-alcoholic fatty liver disease. *J. Hepatol.* **69**, 927–947 (2018).
39. K. Hellemans, I. Grinko, K. Rombouts, D. Schuppan, A. Geerts, All-*trans* and 9-*cis* retinoic acid alter rat hepatic stellate cell phenotype differentially. *Gut* **45**, 134–142 (1999).
40. M. Boyle, D. Tiniakos, J. M. Schattenberg, V. Ratziu, E. Bugianessi, S. Petta, C. P. Oliveira, O. Govaere, R. Younes, S. McPherson, P. Bedossa, M. J. Nielsen, M. Karsdal, D. Leeming, S. Kendrick, Q. M. Anstee, Performance of the PRO-C3 collagen neo-epitope biomarker in non-alcoholic fatty liver disease. *JHEP Rep.* **1**, 188–198 (2019).
41. M. Kanno, K. Kawaguchi, M. Honda, R. Horii, H. Takatori, T. Shimakami, K. Kitamura, K. Arai, T. Yamashita, Y. Sakai, T. Yamashita, E. Mizukoshi, S. Kaneko, Serum aldo–keto reductase family 1 member B10 predicts advanced liver fibrosis and fatal complications of nonalcoholic steatohepatitis. *J. Gastroenterol.* **54**, 549–557 (2019).
42. B. K. Koo, S. H. Um, D. S. Seo, S. K. Joo, J. M. Bae, J. H. Park, M. S. Chang, J. H. Kim, J. Lee, W.-I. Jeong, W. Kim, Growth differentiation factor 15 predicts advanced fibrosis in biopsy-proven non-alcoholic fatty liver disease. *Liver Int.* **38**, 695–705 (2018).
43. O. Govaere, S. Cockell, M. Van Haele, J. Wouters, W. Van Delm, K. Van den Eynde, A. Bianchi, R. van Eijsden, W. Van Steenbergen, D. Monbaliu, F. Nevens, T. Roskams, High-throughput sequencing identifies aetiology-dependent differences in ductular reaction in human chronic liver disease. *J. Pathol.* **248**, 66–76 (2018).
44. S. L. Friedman, V. Ratziu, S. A. Harrison, M. F. Abdelmalek, G. P. Aithal, J. Caballeria, S. Francque, G. Farrell, K. V. Kowdley, A. Craxi, K. Simon, L. Fischer, L. Melchor-Khan, J. Vest, B. L. Wiens, P. Vig, S. Seyedkazemi, Z. Goodman, V. W.-S. Wong, R. Loomba, F. Tacke, A. Sanyal, E. Lefebvre, A randomized, placebo-controlled trial of cenicriviroc for treatment of nonalcoholic steatohepatitis with fibrosis. *Hepatology* **67**, 1754–1767 (2017).
45. T. Hardy, K. Wonders, R. Younes, G. P. Aithal, R. Aller, M. Allison, P. Bedossa, F. Betsou, J. Boursier, M. J. Brosnan, A. Burt, J. Cobbold, H. Cortez-Pinto, C. P. Day, J.-F. Dufour, M. Ekstedt, S. Francque, S. Harrison, L. Miele, P. Nasr, G. Papatheodoridis, S. Petta, D. Tiniakos, R. Torstenson, L. Valenti, A. G. Holleboom, H. Yki-Jarvinen, A. Geier, M. Romero-Gomez, V. Ratziu, E. Bugianesi, J. M. Schattenberg, Q. M. Anstee; LITMUS Consortium, The European NAFLD Registry: A real-world longitudinal cohort study of nonalcoholic fatty liver disease. *Contemp. Clin. Trials* **98**, 106175 (2020).
46. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

47. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9440–9445 (2003).
48. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
49. D. W. Huang, B. T. Sherman, R. A. Lempicki, Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
50. W. Walter, F. Sánchez-Cabo, M. Ricote, GOplot: An R package for visually combining expression data with functional analysis. *Bioinformatics* **31**, 2912–2914 (2015).
51. G. Yu, L.-G. Wang, Y. Han, Q. Y. He, clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
52. A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, A. A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
53. S. A. Williams, M. Kivimaki, C. Langenberg, A. D. Hingorani, J. P. Casas, C. Bouchard, C. Jonasson, M. A. Sarzynski, M. J. Shipley, L. Alexander, J. Ash, T. Bauer, J. Chadwick, G. Datta, R. K. DeLisle, Y. Hagar, M. Hinterberg, R. Ostroff, S. Weiss, P. Ganz, N. J. Wareham, Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* **25**, 1851–1857 (2019).

Acknowledgments: We would like to thank the EPoS (Elucidating Pathways of Steatohepatitis) (epos-nafld.eu) and LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) investigators (litmus-project.eu), the NIHR Newcastle Biomedical Research Centre, the NIHR Cambridge Biomedical Research Centre and the European NAFLD Registry, the Newcastle University Genomics Core Facility, the Newcastle NanoString Core Facility, and the Newcastle Molecular Pathology Node Proximity Laboratory for technical support. **Funding:** This study has been supported by the EPoS consortium funded by the Horizon 2020 Framework Program of the European Union under grant agreement 634413 (to Q.M.A.), the LITMUS consortium funded by the Innovative Medicines Initiative (IMI2) Program of the European Union under grant agreement 777377 (to Q.M.A.). **Author contributions:** Q.M.A. and A.K.D. created the study concept. Q.M.A., A.K.D., and O.G. designed the study and drafted the manuscript. O.G., S.C., and R.Q. performed the bioinformatics analysis. O.G., S.C., and F.R. performed the statistical analysis. Histopathology was done by D.T. and P.B. All authors contributed to data acquisition, analysis, and interpretation and critically revised the manuscript for intellectual content. **Competing interests:** Q.M.A. reports grants from European Commission during the conduct of the study and other grants from AbbVie, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Glympse Bio, Novartis Pharma AG, Pfizer Ltd., and Vertex; reports consultancy for 89Bio, Abbott Laboratories, Acuitas Medical, Allergan/Tobira, Altimmune, AstraZeneca, Axcella, Blade, BMS, BNN Cardio, Celgene, Cirius, CymaBay, EcoR1, E3Bio, Eli Lilly & Company Ltd., Galmed, Genentech, Genfit SA, Gilead, Grunthal, HistoIndex, Indalo, Imperial Innovations, Intercept Pharma Europe Ltd., Inventiva, IQVIA, Janssen, Madrigal, MedImmune, Metacrine, NewGene, NGMBio, North Sea Therapeutics, Novartis, Novo Nordisk A/S, PathAI, Pfizer Ltd., Poxel, ProSciento, Raptor Pharma, Servier, Terns, and Viking Therapeutics; is a speaker for Abbott Laboratories, Allergan/Tobira, BMS, Clinical Care Options, Falk, Fishawack, Genfit SA, Gilead, Integritas Communications, Kenes, and MedScape; and reports royalties from Elsevier Ltd. D.T. reports consultation fees from Intercept Pharmaceuticals Inc., Allergan, Cirius Therapeutics, and an educational grant from Histoindex Pte Ltd. M.A. reports consultancy/advisory with MedImmune/Astra Zeneca and E3Bio, as well as honoraria from Intercept and grant support from GSK and Takeda. M.E. reports personal fees from AbbVie, AstraZeneca, Albireo, Diapharma, and Gilead and nonfinancial support from Echosens (through LITMUS IMI project). K.C. has no personal honoraria but has consultancy and scientific collaboration activity for LNC Therapeutics, Confotherapeutics, and Danone Research. J.M.S. reports grants from Gilead and Boehringer Ingelheim and fees from Gilead, Boehringer Ingelheim, Galmed, Genfit, Intercept, Novartis, Pfizer, and AbbVie outside the submitted work. E.B. reports advisory/consulting for BMS, Genfit SA, Gilead, Intercept, and Novartis. All other authors declare that they have no competing interests. **Data and materials availability:** All data associated with this study are present in the paper or the Supplementary Materials. The RNA-seq data are available in the NCBI GEO repository (accession GSE135251).

Submitted 4 December 2019
Resubmitted 1 July 2020
Accepted 9 November 2020
Published 2 December 2020
10.1126/scitranslmed.aba4448

Citation: O. Govaere, S. Cockell, D. Tiniakos, R. Queen, R. Younes, M. Vacca, L. Alexander, F. Ravaioli, J. Palmer, S. Petta, J. Boursier, C. Rosso, K. Johnson, K. Wonders, C. P. Day, M. Ekstedt, M. Orešič, R. Darlay, H. J. Cordell, F. Marra, A. Vidal-Puig, P. Bedossa, J. M. Schattenberg, K. Clément, M. Allison, E. Bugianesi, V. Ratziu, A. K. Daly, Q. M. Anstee, Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci. Transl. Med.* **12**, eaba4448 (2020).

Science Translational Medicine

Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis

Olivier Govaere, Simon Cockell, Dina Tiniakos, Rachel Queen, Ramy Younes, Michele Vacca, Leigh Alexander, Federico Ravaioli, Jeremy Palmer, Salvatore Petta, Jerome Boursier, Chiara Rosso, Katherine Johnson, Kristy Wonders, Christopher P. Day, Mattias Ekstedt, Matej Oresic, Rebecca Darlay, Heather J. Cordell, Fabio Marra, Antonio Vidal-Puig, Pierre Bedossa, Jörn M. Schattenberg, Karine Clément, Michael Allison, Elisabetta Bugianesi, Vlad Ratziu, Ann K. Daly and Quentin M. Anstee

Sci Transl Med **12**, eaba4448.
DOI: 10.1126/scitranslmed.aba4448

Gene expression during disease progression

Nonalcoholic fatty liver disease (NAFLD) manifests as an array of conditions ranging from hepatic fat buildup to liver inflammation and scarring. Govaere *et al.* detected a signature of 25 genes whose expression in diseased tissue associated with worsening of histologically defined NAFLD severity in two independent cohorts of patients. Protein expression of GDF15, a member of the transforming growth factor–β superfamily, also associated with ballooning, inflammation, and fibrosis in NAFLD serum samples. Together, these studies provide a comprehensive analysis of gene expression during human NAFLD progression.