Transcriptomic signatures of progressive and regressive liver fibrosis and portal hypertension

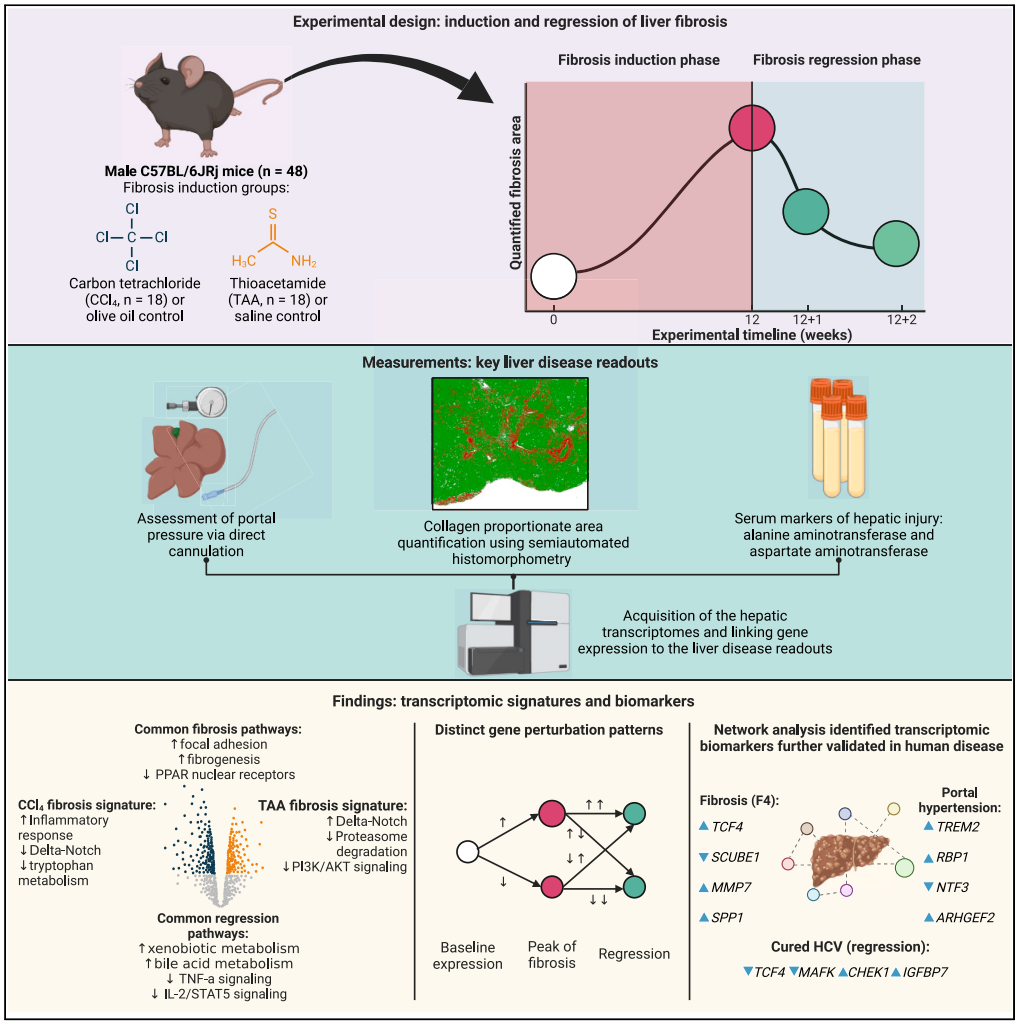
Oleksandr Petrenko, Philipp Königshofer, Ksenia Brusilovskaya, Benedikt S Hofer, Katharina Bareiner, Benedikt Simbrunner, Frank Jühling, Thomas F. Baumert, Joachim Lupberger, Michael Trauner, et al.

iScience

CellPress
OPEN ACCESS

Article

Transcriptomic signatures of progressive and regressive liver fibrosis and portal hypertension



Oleksandr Petrenko, Philipp Königshofer, Ksenia Brusilovskaya, ..., Laura P.M.H. de Rooij, Philipp Schwabl, Thomas Reiberger

thomas.reiberger@meduniwien.ac.at

Highlights

Early-phase liver fibrosis regression displays a robust transcriptomic signature

Genes linked to fibrosis progression show several dynamic patterns in regression

Network analysis revealed gene modules linked to portal hemodynamics and fibrosis

Transcriptomic markers *TCF4*/*TREM2*/*CD74* are linked to advanced human liver disease

Petrenko et al., iScience 27, 109301
March 15, 2024
https://doi.org/10.1016/j.isci.2024.109301

iScience

CellPress
OPEN ACCESS

Article

Transcriptomic signatures of progressive and regressive liver fibrosis and portal hypertension

Oleksandr Petrenko,[1,2,3] Philipp Königshofer,[1,2,3] Ksenia Brusilovskaya,[1,2,3] Benedikt S. Hofer,[1,2,3] Katharina Bareiner,[1] Benedikt Simbrunner,[1,2,3] Frank Jühling,[4] Thomas F. Baumert,[4,5,6] Joachim Lupberger,[4,5,6] Michael Trauner,[1] Stefan G. Kauschke,[7] Larissa Pfisterer,[7] Eric Simon,[8] André F. Rendeiro,[3] Laura P.M.H. de Rooij,[3] Philipp Schwabl,[1,2,3] and Thomas Reiberger[1,2,3,9,*]

SUMMARY

Persistent liver injury triggers a fibrogenic program that causes pathologic remodeling of the hepatic microenvironment (i.e., liver fibrosis) and portal hypertension. The dynamics of gene regulation during liver disease progression and early regression remain understudied. Here, we generated hepatic transcriptome profiles in two well-established liver disease models at peak fibrosis and during spontaneous regression after the removal of the inducing agents. We linked the dynamics of key disease readouts, such as portal pressure, collagen area, and transaminase levels, to differentially expressed genes, enabling the identification of transcriptomic signatures of progressive vs. regressive liver fibrosis and portal hypertension. These candidate biomarkers (e.g., *Tcf4*, *Mmp7*, *Trem2*, *Spp1*, *Scube1*, *Islr*) were validated in RNA sequencing datasets of patients with cirrhosis and portal hypertension, and those cured from hepatitis C infection. Finally, deconvolution identified major cell types and suggested an association of macrophage and portal hepatocyte signatures with portal hypertension and fibrosis area.

INTRODUCTION

Liver fibrogenesis is a complex process characterized by functional alterations in multiple cell types and excessive extracellular matrix (ECM) turnover in response to ongoing hepatic injury,[1] which can ultimately result in cirrhosis and portal hypertension (PH).[2] Despite significant progress in understanding the pathomechanisms contributing to liver fibrosis,[3] the molecular drivers of PH and fibrosis regression still remain elusive. Investigation of hepatic gene expression patterns (i.e., transcriptomic signatures) associated with PH severity and liver fibrosis regression will not only enhance our comprehension of liver disease but also holds promise for the development of novel therapeutics for patients with liver cirrhosis and PH.

Persisting hepatic necroinflammation,[4] vascular remodeling characterized by capillarization of liver sinusoidal endothelial cells (LSECs),[5] macrophage activation,[6] and transdifferentiation of hepatic stellate cells (HSCs) to myofibroblasts[7] are hallmarks of liver fibrosis. Although it is well-known that the removal of the primary etiologic factor of liver disease enables patients to achieve fibrosis regression[8] and amelioration of PH,[9] resulting in the regeneration of hepatic function, there has been no systematic analysis of the of hepatic gene expression dynamics associated with these processes in a controlled setting.

Nevertheless, the dynamics of hepatic gene expression associated with liver fibrosis regression and PH severity have yet to be systematically analyzed in a controlled setting, such as in widely used mouse models of carbon tetrachloride (CCl_4)- or thioacetamide (TAA)-induced liver disease.

Mechanistic insights on the molecular drivers of PH suggest that endothelial dysfunction,[10] pathologic angiogenesis,[11] and abnormal HSC-LSEC crosstalk[12] are critical factors. However, obtaining high-quality human liver biopsy material for research remains challenging,

[1]Division of Gastroenterology and Hepatology, Department of Internal Medicine III, Medical University of Vienna, Vienna 1090, Austria
[2]Christian Doppler Laboratory for Portal Hypertension and Liver Fibrosis, Medical University of Vienna, Vienna 1090, Austria
[3]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna 1090, Austria
[4]Université de Strasbourg, Inserm, Institut de Recherche sur les Maladies Virales et Hepatiques UMR_S1110, Strasbourg 67000, France
[5]Service d'hépato-gastroentérologie, Hôpitaux Universitaires de Strasbourg, Strasbourg 67000, France
[6]Institut Universitaire de France (IUF), 75005 Paris, France
[7]Department of CardioMetabolic Diseases Research, Boehringer Ingelheim Pharma GmbH & Co.KG, 88397 Biberach an der Riss, Germany
[8]Global Computational Biology and Digital Sciences, Boehringer Ingelheim Pharma GmbH & Co.KG, 88397 Biberach an der Riss, Germany
[9]Lead contact
*Correspondence: thomas.reiberger@meduniwien.ac.at
https://doi.org/10.1016/j.isci.2024.109301

thus limiting the application of high-throughput technologies in studying the dynamics of liver disease progression and regression in human patients in an unbiased fashion.

Animal models provide a unique setting to investigate main molecular pathways involved in liver fibrosis regression and PH severity. To this end, we generated and analyzed transcriptome profiles in two well-established mouse models of advanced fibrosis induced by CCl_4 and TAA.[13,14] Our study design included spontaneous regression after removing the inducing agent for up to two weeks, allowing us to identify molecular markers linked to the dynamics of advanced disease and early-phase resolution, correlate them to key liver disease surrogates, and discover related regulatory transcriptional factors (TFs).

RESULTS

Disease characteristics of the parallel fibrosis progression and regression in animal models

Key liver disease parameters were assessed at the same time point when liver samples for RNA sequencing (RNA-seq) were harvested, specifically after 12 weeks of vehicle administration for healthy control (HC) or 12 weeks of toxin exposure inducing peak fibrosis (cirrhosis, CIR), and then after one week (R1) or for two weeks (R2) of spontaneous fibrosis regression (Figures 1A and 1B). Portal pressure (PP), liver fibrosis (collagen proportionate area, CPA), and hepatic injury (reflected by alanine [ALT] and aspartate [AST] aminotransferase blood levels) were significantly increased in CIR animals (i.e., at peak induction) of both models induced by CCl_4 or TAA, and subsequently decreased during R1 and R2 regression (Figure 1C). Fibrosis induction resulted in a significant increase in PP in induced animals. We observed CIR_{TAA} median level of 7.47 mmHg, and CIR_{CCl4} animals reached the highest values with a median of 10.41 mmHg (Table 1). Median CPA levels were below 1% in control animals and were increased 5-fold (CIR_{TAA}: 3.92%) and 20-fold (CIR_{CCl4}: 16.81%) in diseased groups (Figures 1C and 1D). Both ALT and AST levels were elevated at peak fibrosis, however, at lower intensities in TAA than in CCl_4. Both the CPA and PH decreased during the early regression phase. The decrease in PP was not significant at R1 in the TAA model, improving only in $R2_{TAA}$. The levels of transaminases (ALT, AST) significantly decreased in both models already at the R1 time point. The course of these key liver disease readouts indicates that the models and time points are well suited to study gene regulation at peak fibrosis and PH and during liver disease regression, which was subsequently confirmed by unsupervised principal-component analyses (PCAs).

Immune and matrix remodeling-related genes are driving transcriptomic variance in fibrosis

We used PCA to assess group similarity and identify genes defining these groups. Whole transcriptomes of the sequenced animal livers were provided as an input, and the first three components adequately distinguished CIR and HC while grouping R1+R2 together (Figure S2). Principal components (PCs) 1 and 2 together explained 52.3% of the variation of the dataset (Figure 2A). Still, some CIR_{CCl4} animals appeared to be grouped with R1+R2, suggesting that—even in a controlled environment—some heterogeneity in the transcriptomes of peak fibrosis in the CIR_{CCl4} model can be expected. Following PCA, the top 10% of the genes explaining variance in components 1 and 2 were retained (n = 22) (Figure 2B).

We found matrix metalloproteinase-7 (*Mmp7*)—that is reportedly involved in ECM remodeling and cell adhesion[15]—to be upregulated at peak fibrogenesis and with maintained expression throughout the regression timeline, except in some $R1+R2_{TAA}$ animals (Figure 2B). In the CIR cluster, genes *Cd63*, *Ly6d*, and *Cdkn1a*, involved in immune cell activation, were all associated with peak fibrosis. The stearoyl-coenzyme A desaturase 4 (*Scd4*), Abhydrolase Domain Containing 1 (*Abhd2*), and *Mia2* genes were overexpressed in CIR_{TAA} but not to the same magnitude in CIR_{CCl4}. In contrast, Indolethylamine N-Methyltransferase (*Inmt*), responsible for xenobiotic degradation, showed reduced expression in CIR_{TAA}, but not CIR_{CCl4}, returning to control levels already after one week of regression. Two solute carrier genes, *Slco1a1* and *Slc1a2*, were strongly downregulated by fibrosis induction and not completely restored after two weeks for both models: *Slco1a1* is a transporter of organic anions previously described as dramatically downregulated in ethanol-induced liver damage.[16] *Slc1a2* is involved in glutamate transport, which was previously reported to be downregulated in CCl_4 cirrhosis model and identified as a hub gene in fibrosis.[17] Consequently, we performed differential expression analysis to characterize each model independently. It resulted in 1,547 significantly differentially expressed (DE) genes between CIR_{CCl4} and HC_{CCl4}, and 2,614 genes between CIR_{TAA} and HC_{TAA} (Data S1-1 and S1-2). In these groups, upregulation of ECM-related genes in both models was observed, with *Cd63* and *Ccdc80* overlapping among top upregulated genes (Figure 2C). Downregulation in both cases included cytochrome family genes (prominently *Cyp4a12a*, *Cyp4a12b*, *Cyp2j5*), and solute carriers with a less well-defined role in liver fibrosis (*Slc22a30*, *Slc22a28*).

Notably, *Cyp2j5*, involved in arachidonic acid metabolism, was not previously discussed in the fibrogenesis context but was reported as downregulated in a nonalcoholic fatty liver disease (NAFLD) mouse model.[18]

Following these findings, we assessed ECM-related markers among DE genes. In both CIR groups, collagens *Col5a2*, *Col1a1*, and Fibrillin 1 (*Fbn1*) were highly upregulated—as previously reported for both early and advanced fibrosis.[19,20] Fibrillin 2 (*Fbn2*) also was among these markers, but its function is less understood. Notably, *Fbn2* was downregulated in acute CCl_4 injury,[21] but, in line with our findings, it was upregulated in 6-week CCl_4 administration in *Pxdn*-deficient mice.[22] *Mmp12*, a macrophage-derived enzyme involved in the degradation of elastic fibers, and IL13, mediating HSC activation,[23] were upregulated in both models. CIR_{CCl4} had a more prominent upregulation of *Timp1*, hinting at more advanced liver injury.[24] Upregulation of *Scube1* might be of compensatory or protective nature since its downregulation has been linked to endothelial damage severity in pulmonary hypertension and fibrosis.[25,26] Top downregulated genes included *Capn8* and *Col27a1*, indicating degradation of the homeostatic matrix, and Klkb1, providing a functional link to coagulation (Figures S4A and S4B).

Even though both $R1_{CCl4}$ and $R1_{TAA}$ had already a large set of DE genes compared to the fibrosis peak, there were no DE genes between R2 and R1 for the CCl_4 model. For TAA, one gene was significantly upregulated in R2 compared to R1 (*Slc1a2*), and two were downregulated

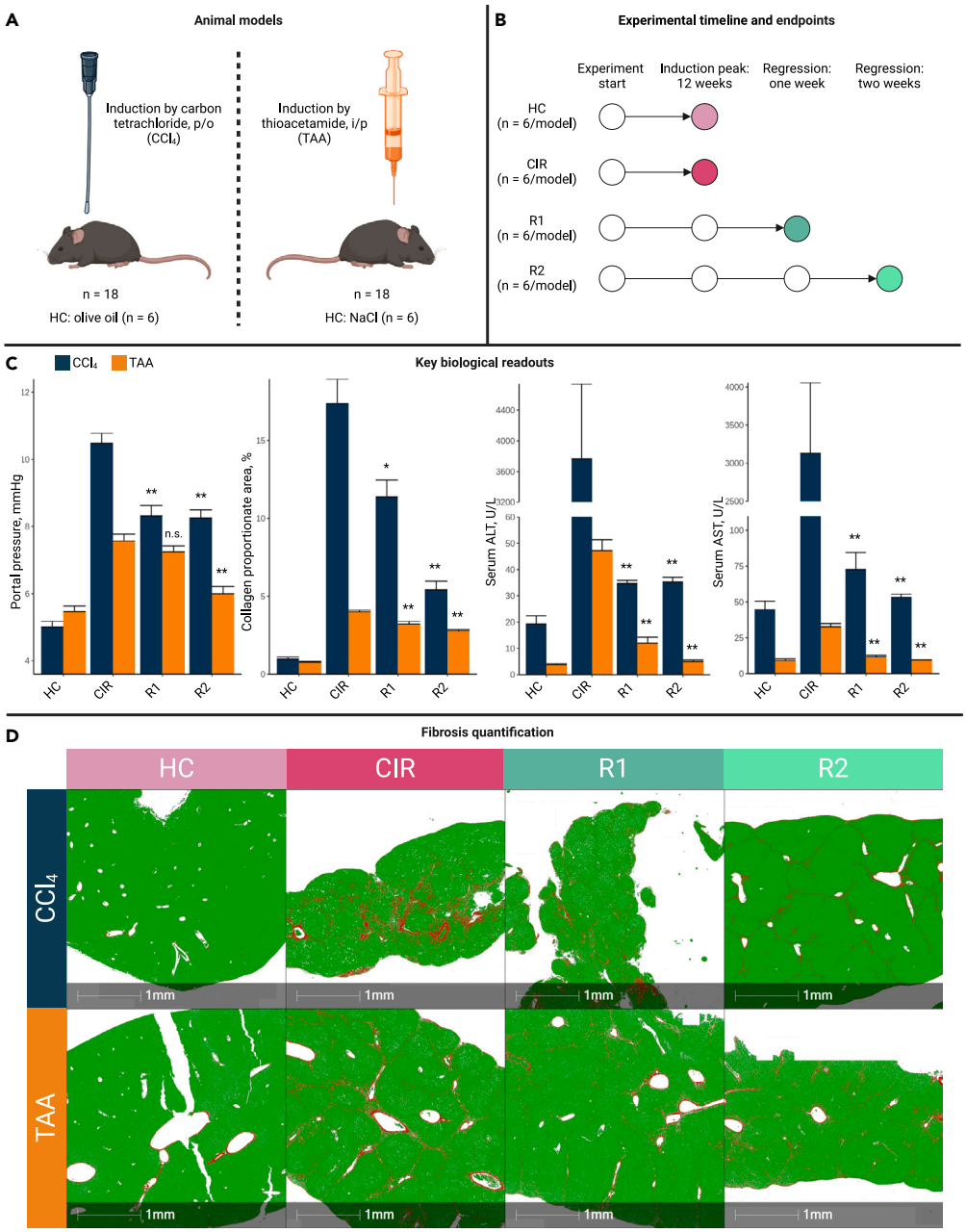


Figure 1. Experimental design and key liver disease readouts

(A) Two different murine fibrosis models were used to study their hepatic transcriptome during peak fibrosis and fibrosis regression.

(B) Experimental timelines are shown. Portal pressure was measured, blood was sampled, and liver tissue was harvested at the respective endpoints.

(C) Comparison of key liver disease readouts in the two models. Results of the Wilcoxon rank-sum test for the differences in the respective parameters after one (R1) and two (R2) weeks of regression to the peak (CIR) time point are indicated (n.s. = non-significant; *$p < 0.05$; **$p < 0.01$; $n = 6$ per group of each model). Data are represented as mean ± SEM.

(D) Illustration of representative Picrosirius Red staining for quantification of collagen proportionate area are shown. Red = collagen, green = fast green tissue counterstain.

(*Cdhr2*, *Dpf1*). *Dpf1* negatively regulates transcription, and *Cdhr2* is involved in contact inhibition and regulation of cell proliferation, suggesting that their functions in fibrosis are of protective nature that is engaged in long-term regression. Among other matrix-related genes, *Vtn*, a target of Mmp2, was strongly upregulated in $R1{+}R2_{TAA}$ (Figures S4C and S4D). In both models' R1+R2, a substantial upregulation of cytochrome subunits was observed with a total of 17 genes in TAA and six overlapping in CCl_4. Most of them were involved in phase I of the

Table 1. Characteristics of the animal cohort

Model	CCl_4	CCl_4	CCl_4	CCl_4	TAA	TAA	TAA	TAA
Parameter/Group	HC	CIR	R1	R2	HC	CIR	R1	R2
n	6	6	6	6	6	6	6	6
PP, mmHg median	5.20	10.41	8.24	8.23	5.56	7.47	7.28	6.09
p value	N/A	*< 0.01*	*< 0.01*	*< 0.01*	N/A	*< 0.01*	0.48	*< 0.01*
CPA, % median	0.82	16.81	10.26	5.10	0.76	3.92	3.27	2.77
p value	N/A	*< 0.01*	*< 0.05*	*< 0.01*	N/A	*< 0.01*	*< 0.01*	*< 0.01*
ALT, U/L median	16	4672	34	36	3.5	45.5	10.5	5.5
p value	N/A	*< 0.01*	*< 0.01*	*< 0.01*	N/A	*< 0.01*	*< 0.01*	*< 0.01*
AST, U/L median	46	3586	58	56	8.5	34	12	9
p value	N/A	*< 0.01*	*< 0.01*	*< 0.01*	N/A	*< 0.01*	*< 0.01*	*< 0.01*

The provided p values resulted from the Wilcoxon rank-sum test result for each group and model for the following comparisons: for CIR: CIR vs. HC; for R1: R1 vs. CIR; for R2: R2 vs. CIR. p values below the significance threshold (<0.05) are italicized. HC = healthy control. CIR = peak of fibrosis. R1 = one week of regression. R1 = two weeks of regression. CCl4 = carbon tetrachloride. TAA = thioacetamide.

P450 pathway, primary bile acid synthesis, retinol, and steroid metabolism. *Cyp1a1*, involved in xenobiotic metabolism, was slightly downregulated in both models' regression. *Timp1* (CCl4), *Itga3*, *Itga6*, and *Itgav* (TAA) were significantly suppressed compared with the fibrosis peak. *Col5a3*, however, still showed increased expression in $R1{+}R2_{TAA}$, suggesting that HSC activity was maintained two weeks after the removal of the injury agent. *Slc1a2* was upregulated in both models after being suppressed at peak fibrosis (Figure 2D, Data S1-3 and S1-4).

Together, these results indicate that both models display an upregulation of classical markers of liver fibrosis, which are then partially again downregulated in disease regression. Other less-studied DE genes were also detected, but their role in liver fibrosis should be further clarified. Considering the strong grouping of the liver transcriptomes of animals during fibrosis regression on PCA and the DE results, we kept R1+R2 (R) merged for the following analyses.

CCl_4 and TAA induce model-specific signatures and distinct recovery patterns

We compared matching groups in two models to identify further functional alterations introduced by CCl_4 and TAA. From DE analysis, 30% (n = 959) of genes overlapped when comparing CIR_{CCl4} and CIR_{TAA} to their respective HC groups (Figure 3A). The upregulation overlap included pathways involved in focal adhesion, cell-cycle control, prostaglandin signaling, fibrosis, and macrophage regulation (Figure 3B). The most robustly downregulated pathway was PPAR (peroxisome proliferator-activated receptor) signaling, followed by the P450 pathway, steroid, fatty acid, and amino acid metabolism. Our findings aligned with the current understanding of the PPAR role in advanced chronic liver disease, as pan-PPAR agonists have shown promising results on fibrosis resolution both in experimental studies and in phase 2b clinical trials.[27,28]

Model-unique genes suggested that CIR_{CCl4} had a more substantial downregulation of the PPAR pathway than CIR_{TAA}, and a more prominent signature of metabolic inhibition (Figures S5A and S5B). Notably, cholesterol biosynthesis was preserved in CIR_{CCl4}. The delta-Notch pathway, involved in portal fibrosis and tissue remodeling in cirrhosis,[29] was inversely regulated: it was upregulated in TAA peak induction but downregulated in CCl_4. In differential testing of the CIR groups, we found that *Pdzk1*, *Ccl9*, and *Ppp1r42* were among the highest expressed markers of CIR_{CCl4}, while *Ets2*, *Slc39a10*, and *Fem1b* were associated with CIR_{TAA} (Figure S15A and Data S1-5).

The regression overlap of both models included 8.1% genes (n = 101) (Figure 3A). The upregulation of xenobiotic metabolism, inflammatory regulation, and bile acid metabolism was observed together with the downregulation of classical proinflammatory pathways (nuclear factor κB [NF-κB], KRAS, IL6-STAT3) (Figures 3B and 3C). Unlike TAA, R_{CCl4} showed a significant upregulation of the WNT signaling. A restoration of PPAR and metabolic signaling was observed in CCl_4, while for TAA the highest pathways were related to oxidative phosphorylation (Figures S5C and S5D). We detected significant upregulation of *Ppargc1a* (coactivator of *Pparg*) and PPARbeta/delta genes and their downstream mediators *Fabp1*, *Acaa1a*, *Ilk*, *Scd1*, and related cytochromes (Data S1-3) in fibrosis regression.

Top R_{TAA}-specific genes included *Thbs4* and *Dmbt1* (Figures S15B and S15C). The latter was previously reported as highly expressed in liver regeneration in rats.[30] *Cd163* and *Hhip* were specific to R_{CCl4}. In previous studies, Hhip was shown as an HSC-derived gene and a member of the Sonic hedgehog (Shh) pathway that is upregulated when HSC activation is prevented[31] (Data S1-6).

These functional overlaps revealed that both models express preserved fibrogenesis pathways, followed by the downregulation of inflammatory pathways in regression and upregulating of metabolic and oxidative phosphorylation pathways (Data S2). The identified model-specific differences in delta-Notch, PPAR, Wnt, and Shh pathways should be considered when testing therapeutic candidates.

As an alternative strategy to differential gene expression testing, we investigated how genes are perturbed by phenotype change toward fibrosis or regression. Both models were merged for these experiments, and expression in the HC or CIR group was considered a baseline to calculate cirrhosis and regression scores (see "STAR Methods"). These scores represented each specific gene's expression alteration toward CIR or R accordingly. We found that most of the genes were altered by cirrhosis toward upregulation and in regression toward downregulation (Figure 3G)—indicating injury-driven gene upregulation in peak fibrosis. In genes impacted by both perturbations, a strong negative

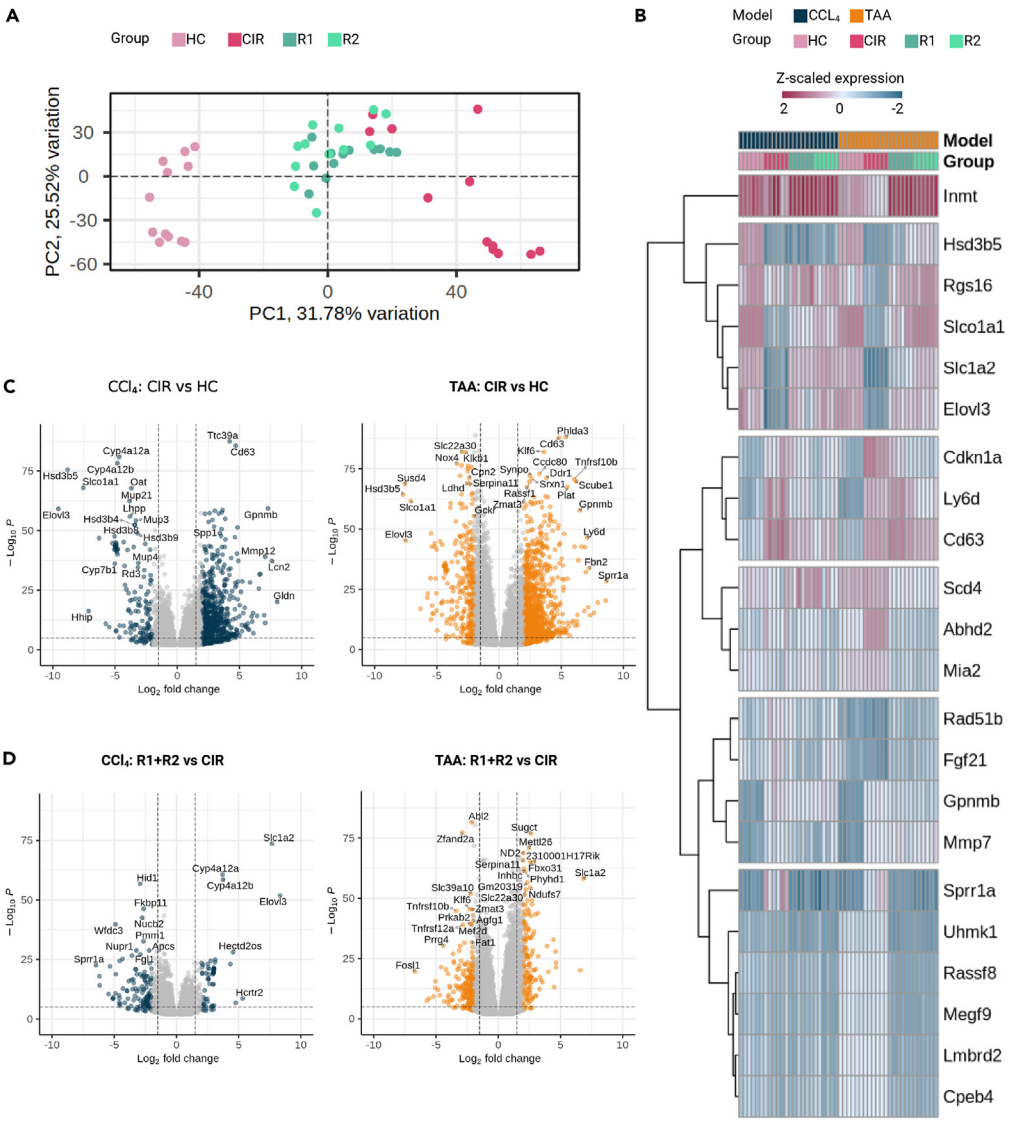


Figure 2. Transcriptomic characterization of genes associated with peak fibrosis and regressive fibrosis

(A) PCA projection demonstrates a clear separation of HC and mostly of CIR, while the R1 and R2 groups cluster together.

(B) Genes with 10% highest loading score for principal components 1 (PC1) and 2 (PC2). Unsupervised gene clustering denotes similar expression patterns.

(C) Volcano plots for most differentially expressed genes between animals at peak fibrosis induction (CIR) and healthy control animals (HC) are shown separately for the CCl4 and TAA models.

(D) Volcano plots for most differentially expressed genes between animals with regressive cirrhosis (combined R1+R2) vs. peak fibrosis induction (CIR) are shown separately for both models. Thresholds: log2FC > 1.5, $p_{adjusted}$ <0.01. PC = principal component.

See also Figures S2–S4, and Data S1.

correlation between the scores was detected (Figure S7). Next, functional analysis was performed with genes ranked with the 20% highest absolute scores. Genes with the highest cirrhosis score showed involvement in adhesion, fibrosis, and integrin-related signaling (Figure 3H). Similar to DE analysis findings, pathways like PPAR, P450, steroid biosynthesis, and oxidative phosphorylation were upregulated among genes with the highest regression score (Figure 3I). The aforementioned pathways were mostly reversely overlapping in the lowest 20%-scored genes (Figure S8).

Pseudotemporal dynamics reveals gene trajectories in disease progression and regression

By jointly leveraging the gene expression data from the models, we endeavored to gain insights into the evolution of disease and recovery. The aim of this analysis was to identify genes perturbed by these processes and the direction of changes over experimental endpoints—to which we refer as pseudotemporal dynamics. We defined all our study groups as pseudotemporal points. We investigated how the

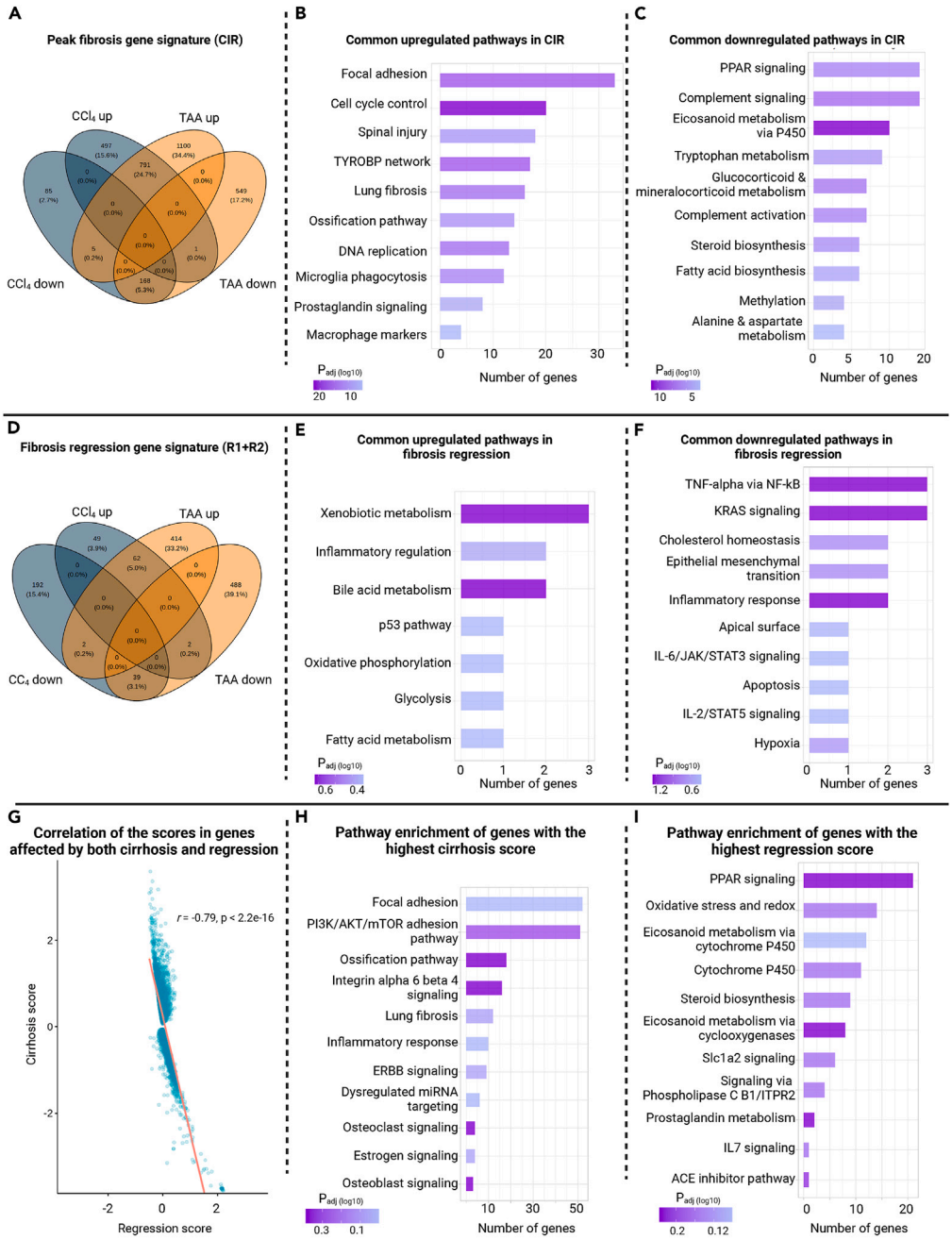


Figure 3. Functional annotation of genes involved in CCl4 and TAA fibrosis progression and regression

(A–C) Venn diagram representing overlaps in up- and downregulated genes (number and proportion of genes are shown) associated with fibrosis induction in the two fibrosis models. The functional analysis displays molecular signaling pathways for both CCl4/TAA models that are (B) up- or (C) downregulated in peak fibrosis induction (CIR).

(D) Venn diagram displaying overlaps in genes (number and proportion of genes are shown) linked to fibrosis regression in the two models.

(E and F) (E) Functional upregulation and (F) downregulation of molecular signaling pathways for both CCl4/TAA models during fibrosis regression. A hypergeometric test with MSigDB annotation was applied for the genes overlapped in regression.

(G) The scores in genes perturbed by both conditions demonstrate a strong negative correlation (Pearson's r is shown). The scores were scaled for visualization.

(H) Signaling pathways from genes with the 20% highest score for CIR are shown.

(I) Pathways from genes with the 20% highest regression score (CIR vs. R1+R2) are shown.

See also Figures S5–S8, and Data S2.

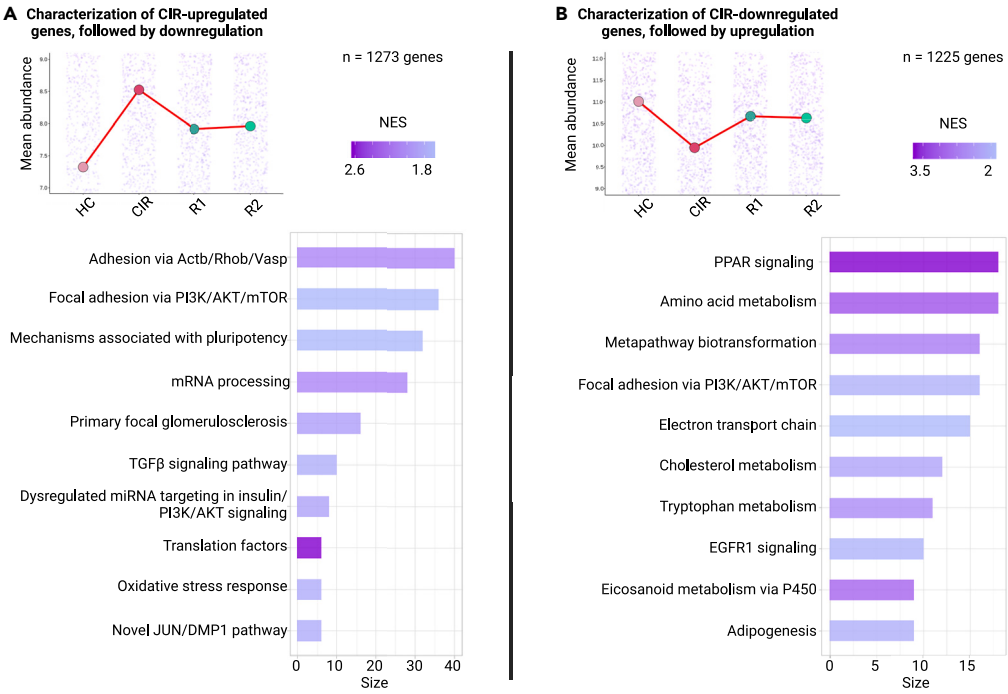


Figure 4. Dynamic signature patterns following liver disease course using pseudotemporal analysis

(A) Cluster 1: pathway analysis of genes upregulated at peak fibrosis followed by subsequent downregulated (vs. CIR) and approximation to baseline (HC) levels in regression.

(B) Cluster 2: pathways analysis of genes downregulated at peak fibrosis followed by subsequent upregulation (vs. CIR) and approximation to baseline (HC) levels in regression. NES = normalized enrichment score.

See also Figures S9–S11, and Data S3.

expression of the same genes changes between the non-diseased state (i.e., the HC group), the peak of induction (CIR), and one to two weeks without the damaging factor (R1+R2). The genes changing in the same direction were then clustered, and the following patterns were observed (Figure S9).

- (Cluster 1) upregulated in fibrosis, then downregulated in regression (1,273 genes);
- (Cluster 2) downregulated in fibrosis, upregulated in regression (n = 1,225);
- (Cluster 3) upregulated in fibrosis, upregulated in regression (n = 261);
- (Cluster 4) downregulated in fibrosis, downregulated in regression (n = 241).

On the pathway level, clusters 1 and 2 were mostly well in line with our findings from DE genes for CIR vs. HC and R vs. CIR accordingly, as well as with top-ranked genes for perturbation signatures. The most prominent pathways in cluster 1 were cell adhesion, followed by progenitor cell signaling, and profibrotic pathways, including transforming growth factor β (TGF-β) and Dmp1 signaling. Cluster 2, in turn, had PPAR as the most prioritized pathway, followed by amino acid and cholesterol metabolism, oxidative phosphorylation, and the P450 pathway (Figures 4A and 4B).

Cluster 3's pathways were similar to those described in CIR and cluster 1, with cell adhesion via *Itga8* and laminins playing the most prominent role. Another finding was persisting upregulation of matrix metalloproteinases, mainly *Timp3*, *Timp2*, and *Mmp2*—which suggests their role in sustained matrix remodeling even in the regression stages (Figure S10A).

Cluster 4 was characterized by a slight downregulation in both fibrosis and regression. Its genes were not strongly grouped in pathways, with highest (yet not significant) ranked ones belonging to proteasome degradation and oxidative damage. The macrophage marker *Cd163* was found among cluster genes, which is known to be downregulated in the proinflammatory microenvironment. Additional findings included solute carriers transporting nucleic bases, prostaglandins, and zinc (*Slc35e3*, *Slco2a1*, *Slc39a4*), and *Scarb2*, one of the scavenger receptors on the LSEC membrane (Figure S10B).

Cluster 2 had the most extensive overlap with DE genes for all groups (n = 473 genes), followed by cluster 1 (n = 316). The most considerable overlap of cluster 3 was with genes DE for both CIR (n = 95), and cluster 4 contained most DE genes of R_{TAA} (Figure S11). These response patterns emphasized the plethora of molecular alterations brought by fibrogenesis and its disease-associated dynamics (Data S3-1–S3-4). Aside from impacting ECM composition, it also suggests the involvement of hepatic non-parenchymal cells, reflected by their hallmark genes.

Co-expression analysis and targeted learning lead to discovery of transcriptional biomarkers

Two independent methods were applied to narrow down the list of DE genes, and biomarker candidates were identified in the overlap of their results. We selected the first one, weighted co-expression network analysis (WGCNA), as an established unsupervised method for finding genes co-expressing with external parameters. The second one, targeted minimum loss-based estimation (TMLE), is a machine learning-based approach to estimating gene expression impact on selected liver disease parameters while controlling for other confounders. In both cases, we aimed to identify markers related to the key biological liver disease severity readouts PP and CPA as continuous values and using cirrhosis and regression as categorical parameters. This biomarkers search was model agnostic and performed using merged regression groups.

WGCNA discovered a total of 19 co-expressed gene modules strongly linked to our external readouts (Figure 5A). Modules 2 (M2), M4, M5, and M18-19 were strongly correlating with regression. PP, CPA, and liver injury markers shared M6-M8 and M14. Unique modules, such as M18 for regression and M12 for PP dynamics, indicated group of genes with high association probability to these mechanisms.

PP, CPA, and cirrhosis shared the correlation direction of such significant modules; the shared injury signature was also for ALT and AST gene modules.

The key pathways in the gene set co-expressed with regression were Myc targets or unfolded protein response (Figure S12). The top markers included genes involved in vascular signaling (e.g., *Vegfa*, *Angptl2*) and matrix-related genes (e.g., *Adam33*, *Adamtsl2*, *Igfbp7*) (Data S4-1).

The highest-scored genes associated with PP were related to pathways involved in epithelial-mesenchymal transition, coagulation, and xenobiotic metabolism. It also contained HSC activation state markers such as *Timp3*, *Wnt5a*, *Vim*, and *Pdgfb*. Identification of both HSC and sinusoidal markers (*Pecam1*, *Vwf*, *Cd9*) among highly significant genes for PP in these animal models supports previous reports on the involvement of the vascular microenvironment in the development of hemodynamic alterations.[32–34]

Furthermore, both cirrhosis- and CPA-linked genes comprised a vast number of matrix turnover genes, as well as angiogenesis genes (e.g., *Adam8*, *Itgav*, *Vegfa*, *Pdgfa*, *Vcan*). In addition to mesenchymal-epithelial transition, angiogenesis was among the high-scored pathways for both gene sets.

ALT and AST share a common signature, including E2F targets, known as a pro-apoptotic pathway in the liver, xenobiotic metabolism, adipogenesis, and bile acid metabolism. Among the highest-ranked genes were *Prtn3*, a monocyte marker; *Vsig8*, a regulator of T cell activation; and *S100a3*, a calcium-binding protein regulating gene expression in hepatocytes and a potential treatment target in hepatocellular carcinoma.[35]

TMLE identified that Aldo-keto reductase family 1 member C18 (*Akr1c18*) and Nucleobindin 2 (*Nucb2*) were linked to higher PP (Figure S13A). *Cyp2u1*, on the opposite, was downregulated in CIR animals with the highest level of PP (Figure 5B). A cluster of genes involved in matrix turnover showed an association with CPA and cirrhosis: *Spire2*, *Pak6*, *Tinag*, and HSC-annotated markers *Fhl2*, *Tinag*, *Tgfb2*, and *Shc2* (Figure S13B). In line with WGCNA, the TMLE prioritized an insulin-like growth factor-binding protein 7, *Igfbp7*, linked to regression (Data S4-2).

Using overlaps of these two methods, we identified biomarker candidates for PP (n = 78 genes), CPA (n = 23), cirrhosis (n = 100), and regression (n = 105) (Figure 5C, Data S4-3). The overlapping genes for PP were classified as scavenger receptors, enzymes, and members of RAS, PI3K, and FYN (proto-oncogene tyrosine-protein kinase Fyn) pathways (Figure S14A). The CPA-linked markers included genes coding proteoglycans, elastic fiber, scavenger receptors, and members of the RUNX2 pathway (Figure S14B). The set of genes overlapping in cirrhosis consisted of ECM turnover-related genes, chemokine receptors, elastic fiber, and matrix metalloproteinases (Figure S14C). Notably, regression genes were also involved in the turnover of elastic fibers; additionally, members of Tie2, PTK2, and apoptotic pathways were found (Figure S14D).

Transcriptional regulators in fibrosis progression and regression and their signaling niches

To identify TFs involved in the perturbation of the identified transcriptomic signatures of fibrosis progression and regression, we used differential co-expression of the annotated mouse TFs based on their known target genes (see "STAR Methods").

A total of 58 TFs were identified as significantly involved in our dataset (Data S5-1). We examined their expression and discovered that three of them (*Hmga1*, *Tcf4*, *Mafk*) are significantly upregulated in peak fibrosis ($p < 0.01$), while significant downregulation of *Klf15*, *Ppara*, *Xbp1*, *Creb3l3*, *Foxo4*, *Nfe2l1*, *Nr1d1*, *Nr1i2m*, and *Junb* was detected in CIR. When exploring potential functional links, *Creb3l3* seems to be implicated in anti-inflammatory control since its knockout aggravated liver fibrosis in mouse models.[36] Notably, the downregulation of Foxo4 was previously described in the context of proinflammatory and profibrogenic microenvironment.[37] The identified TFs belonging to the nuclear receptor family included *Nr1i2*, *Nr1h2*, *Nr1d1*, *Rxrb*, and *Ppara*, and their functions were previously implicated in liver disease.[38]

A protein-protein interaction network was constructed using significant TFs and all DE genes between the study groups to perform their functional annotation (Figure S15A). To have a broad perspective of their downstream, we applied the "guilt by association" principle[39] and identified genes directly connected to the TFs, or their 1st-degree neighbors. Functional annotation of the resulted network showed that the identified TFs are involved in the regulation of several immune pathways, with NF-κB and TGF-β among the most prioritized ones (Figure S15B). Other findings included Wnt/β-catenin, Notch, Hedgehog, adipogenesis, and junction pathways, thus complementing our previous findings from the DE analysis.

Next, we aimed to narrow down further the list of significant TFs relevant for regulating the biomarker candidates identified within WGCNA-TMLE overlaps. Their associations were scored by the expression changes in the biomarker genes, and the weakest links were discarded. It allowed us to select 22 TFs impacting levels of PP and CPA, cirrhosis, and regression (Data S5-2). In the following experiment, we integrate these findings via network analysis.

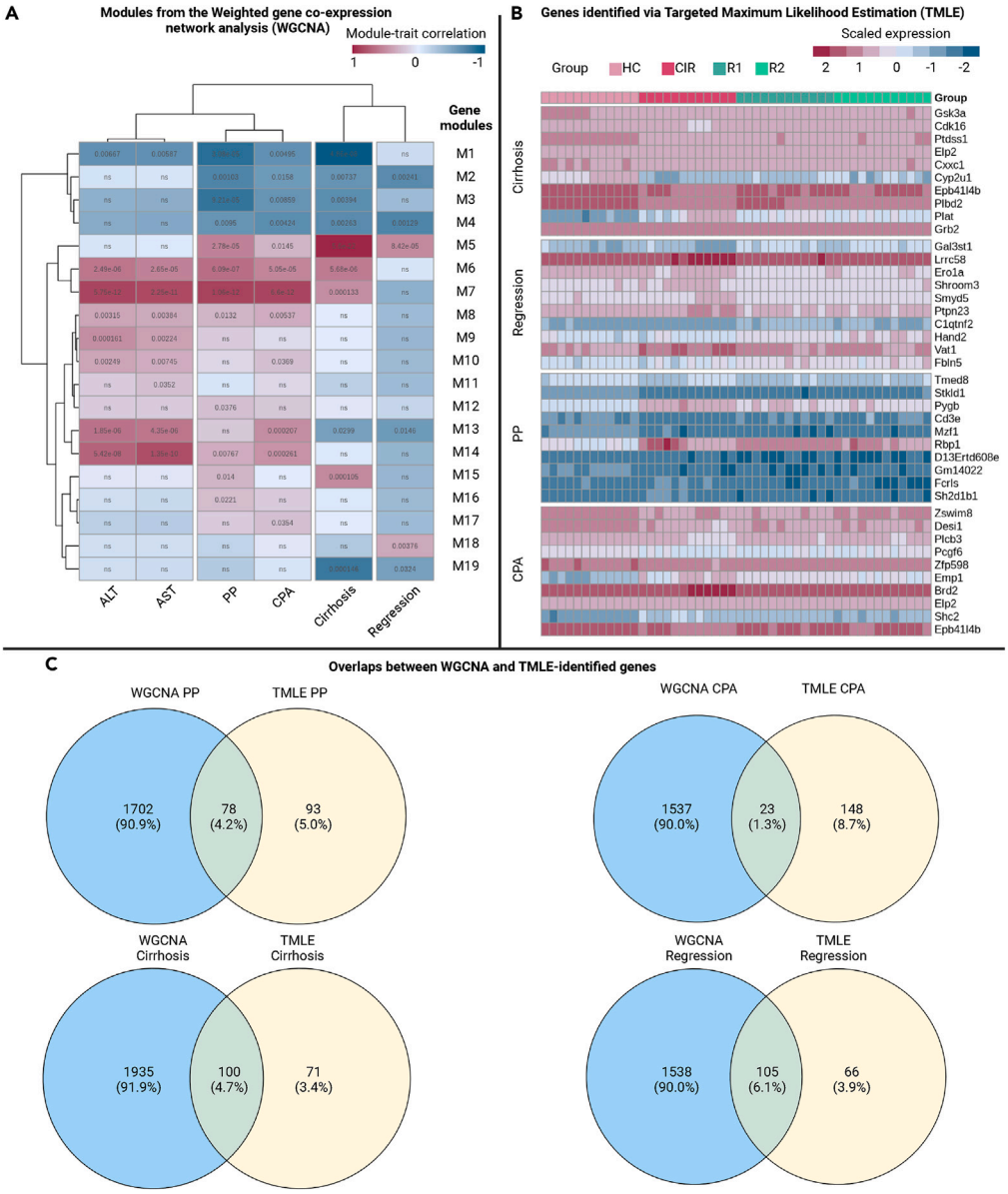


Figure 5. Linking gene signatures to key surrogate parameters of hallmark biological readouts

(A) Gene modules, identified via unsupervised weighted co-expression network analysis (WGCNA), correlating with key biological readouts in the merged dataset.

(B) Supervised clustering of top 40 markers associated with targeted minimum loss-based estimation (TMLE).

(C) Prioritization via overlaps between TMLE markers for each biological readout and respective WGCNA module. PP = portal pressure. CPA = collagen proportionate area.

See also Figures S12–S14, and Data S4.

Analysis of multilayer network results in prioritized biomarkers and their functional links

In previous experiments, sets of biomarker gene candidates and their transcriptional regulators were identified. We assumed that analysis of their biological network could allow us to further select the most informative genes for validation. For this analysis, a three-layer network was constructed: a gene regulatory layer (n = 22 genes) consisting of TFs; a gene co-expression layer (n = 299), including biomarkers, overlapped from the described discovery approaches; and a protein-protein interaction layer (n = 751), including database-mined downstream and upstream targets for the biomarker candidates (Figure 6A).

For prioritization, we employed biology-based scores (perturbation scores that were calculated earlier), several network-based scores, and a machine learning algorithm (see "STAR Methods"). Using combinations of these scores, the network cores (up to n = 25 genes) were

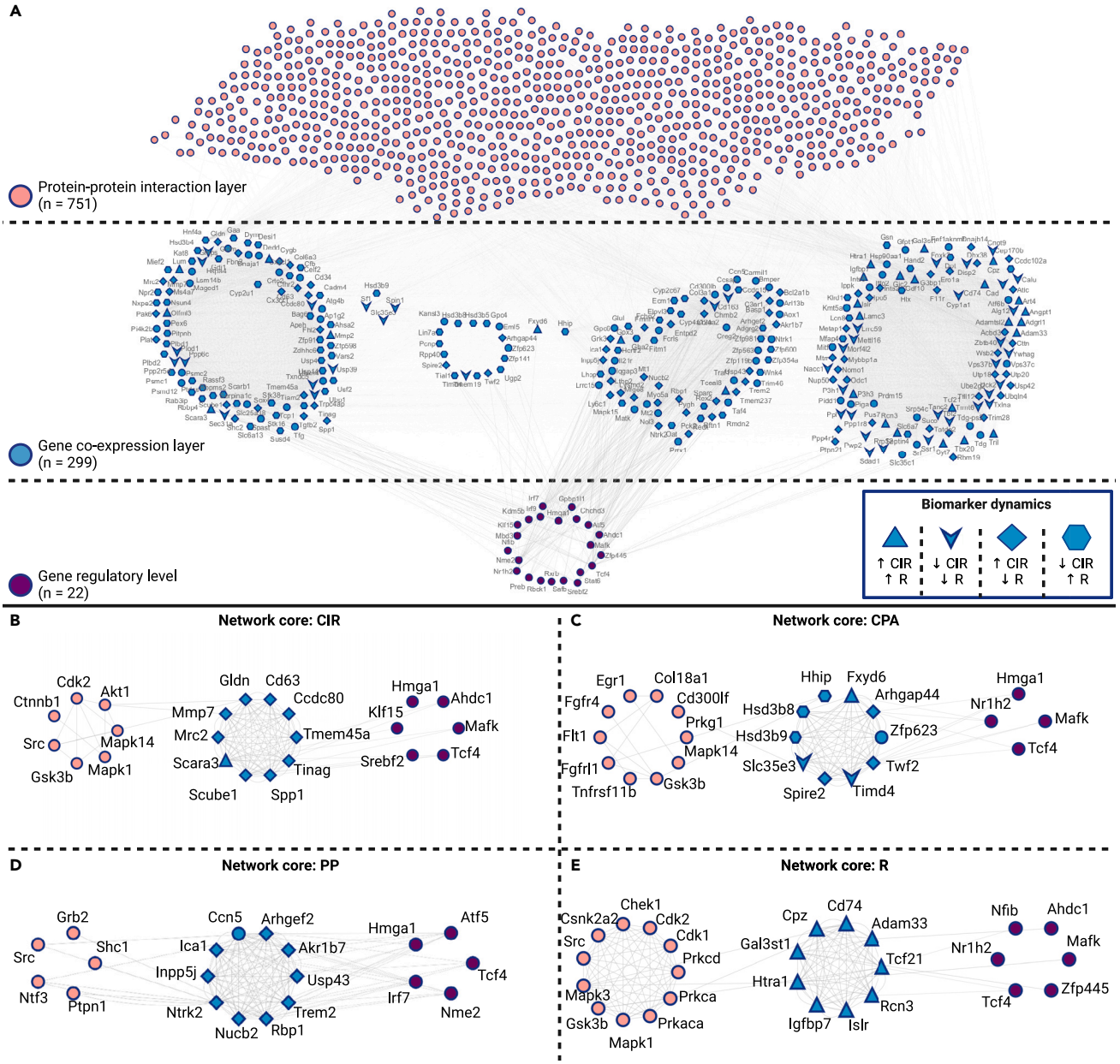


Figure 6. Network analysis prioritizes biomarkers and shows their functional links to liver disease

(A) Multilayer network constructed from selected transcriptional factors (purple, bottom layer), co-expressed biomarkers (blue, middle layer), and their protein-protein interactions (orange, top layer).

(B–E) Biomarkers were scored to identify network cores with the feature's strongest associations. Node shapes indicate gene dynamics between cirrhotic and regression groups. CIR = cirrhosis, CPA = collagen proportionate area, PP = portal pressure, R = fibrosis regression.

See also Figure S15 and Data S5.

identified for each disease "phenotype" category: cirrhosis, CPA, PP, and regression (Figures 6B–6E). We present these genes, scored by several unbiased approaches with the introduction of the biology-based metric, as the prioritized biomarkers and subject them to further validation using available human datasets.

Prioritized biomarkers predict disease severity in validation on human RNA-seq datasets

We aimed to evaluate the prioritized biomarkers in available human hepatic RNA-seq datasets obtained from liver biopsies (Figure 7). First, differential expression analysis identified significant biomarker candidates in these datasets. Then, naive Bayes-aided

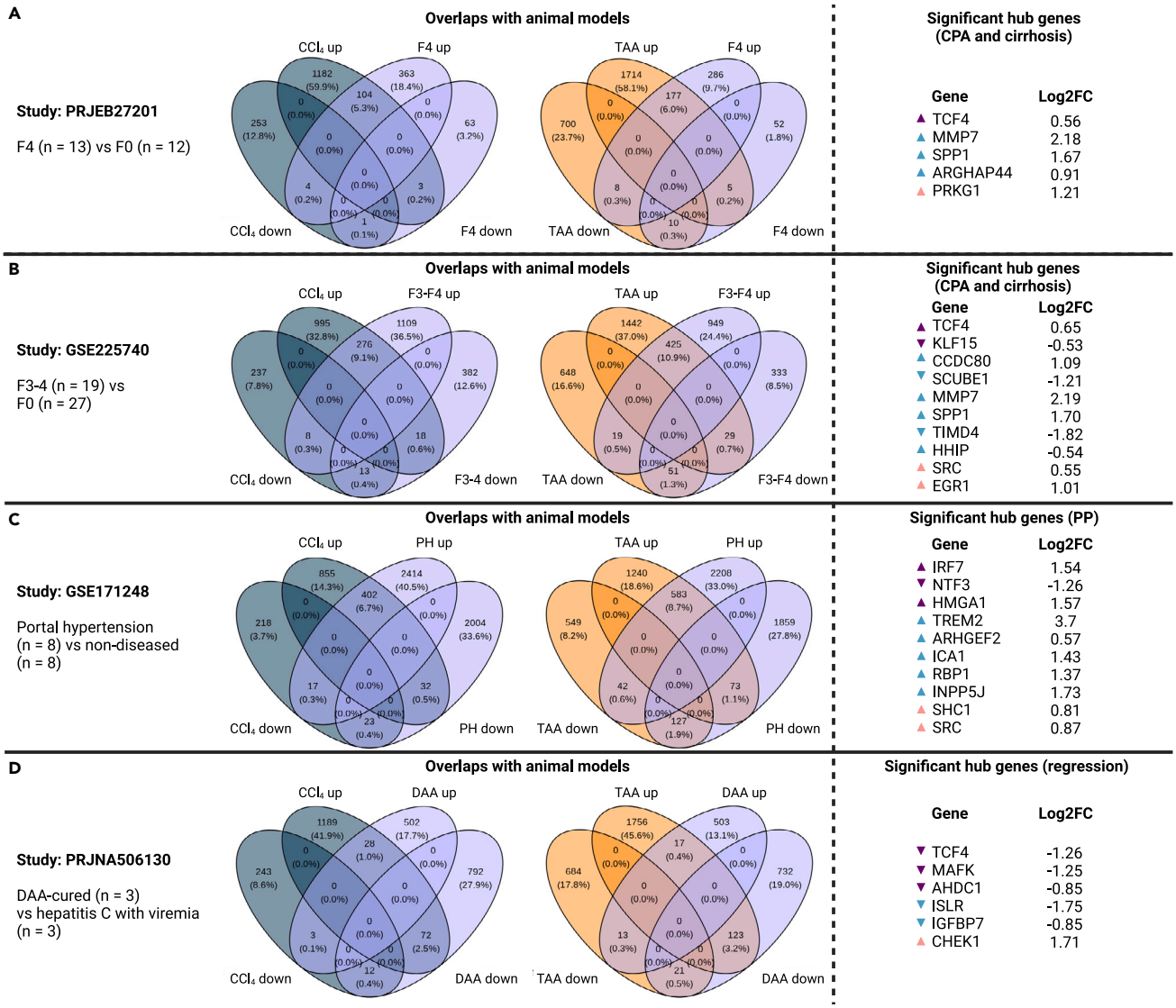


Figure 7. Transcriptomic overlaps and validation of the prioritized biomarkers in human liver disease datasets

The comparisons were performed between (A) patients with histological cirrhosis (F4) and non-diseased; (B) patients with advanced fibrosis/cirrhosis (F3-F4) and non-diseased; (C) patients with portal hypertension and non-diseased; (D) patients with hepatitis C virus cured with direct-acting antiviral therapy and those prior treatment start, accordingly. The right panel indicates detection of the prioritized hub genes. PH = portal hypertension. CPA = collagen proportionate area. PP = portal pressure. DAA = direct-acting antiviral therapy. TAA = thioacetamide. CCl4 = carbon tetrachloride.

See also Table S1.

feature selection was applied to find the best-performing gene sets, which were then utilized to predict disease severity with random forest.

First, we explored expression of the CPA and cirrhosis core genes between patients with advanced fibrosis or cirrhosis compared to patients without histological signs of fibrosis. In the first analysis, comparing F4 (cirrhosis) and F0 (no fibrosis), we identified 538 DEGs. From them, 104 were also upregulated in the CCl_4 model, and 177 – in TAA (Figure 7A). We found five hub genes in these overlaps, with *MMP7*, *SPP1*, and *ARGHAP44* as the most robust upregulated biomarkers.

The second validation used dataset of patients with F3-4 and F0 histologic stages. This increased the DEGs overlaps more than twice (276 commonly upregulated for CCl_4, 425 for TAA). Notably, *MMP7* was again the most upregulated cirrhosis biomarker (Figure 7B). In both cases, *MMP7* was aligned with the regulation pattern in the animal models. The set of biomarkers including T cell immunoglobulin and mucin domain containing 4 (*TIMD4*), *CCDC80*, *SPP1*, and *SCUBE1*, linked to immune regulation and cell adhesion, was also consistent between human data and animal models.

To explore hub genes related to PP, we investigated transcriptomes of patients with cirrhotic PH and non-diseased controls. Despite the modest number of available patient datasets, comparison showed 402 commonly upregulated genes with CCl_4 and 583 with TAA. PP biomarkers among DE genes included *TREM2*, *ARHGEF2*, *ICA1*, *RBP1*, and *INPP5J*. (Figure 7C).

Finally, we investigated our regression-linked hub genes in transcriptomes of patients cured from hepatitis C virus (HCV) infection with direct-acting antiviral (DAA) therapy compared to patients prior to such treatment. We found that the biggest overlaps with the animal models were inversed: 72 (CCl4) and 123 (TAA) upregulated genes were found downregulated in DAA-cured patients. Despite being in a prolonged-phase regression unlike our animal models, *CHEK1* was also found upregulated. Other two biomarkers, *ISLR* (immunoglobulin super family containing leucine-rich repeat) and *IGFBP7*, were downregulated, possibly being involved only in early-phase response, as in the animal models they were upregulated both during the peak induction and regression (Figure 7D).

Notably, all comparisons also showed TFs we predicted among DEGs. TCF4, a member of Wnt pathway, was significant in human datasets with fibrosis and post-DAA treatment. Other prioritized TFs, significant in both models and human data, were MAFK, AHDC1, HMGA1, NTF3, and KLF1.

The machine learning-based feature selection prioritized gene sets, ranging from four to eight genes in size. ISLR, the regression biomarker, was included in all best-performing sets. In prediction, the highest score was achieved with ISLR (R biomarker), HMGA1 (prioritized TF linked to PP), IRF7 (TF for PP), and MMP7 (CPA and cirrhosis biomarker), reaching the accuracy of the random forest of 0.827 (Table S1).

These findings strongly underline the potential clinical value of the prioritized biomarkers inferred using established animal models. It was not possible to access their predictive value in a DAA-treated cohort due to its size. Hence further validation with appropriate patient cohort composition (including available PP and CPA readouts, with expected regression groups) is, however, required.

Cell type deconvolution suggests the contribution of hepatocyte, endothelial, HSC, and macrophage signatures in fibrosis and its regression

Deciphering variations in cell composition across study conditions may give valuable insights into disease pathophysiology. However, due to the complexity of sample preparation and equipment requirements, single-cell methods have yet to become broadly used in the hepatology field. Here, we applied supervised deconvolution to identify the presence of cell type signatures. Two datasets for signature extraction were selected: scRNA-seq of mononuclear macrophages from mouse liver tissue treated for four weeks with CCl_4 ($scCCl_4$)[3] and single-nuclei RNA sequencing (snRNA-seq) of frozen mouse liver tissue treated with TCDD (snTCDD)[40] (n = 24) (Figure S16). Deconvolution analysis resulted in the macrophage signatures from the sorted scRNA-seq dataset and both parenchymal and non-parenchymal cells from the $scCCl_4$ dataset. The two groups of identified Kupffer cell signatures had various functional markers: (1) cells expressing *Cd63* and *Lgals3bp* and (2) cells expressing *Vsig4* and *Marco* (Figure S17A). Among other cell types, HSC, and hepatocytes (with periportal markers *Ass1*, *Pck1*, *Slc7a2*, and *Hal*) had the most prevalent signature (Figure S17B).

The most robust signatures were found for hepatocytes, macrophages, and endothelial cells (source: snTCDD) and HSC ($scCCl_4$) (Figure 8A). In CIR, hepatocyte signatures were suppressed as compared to HC and regression, and the macrophage signature showed a reversed trend. Notably, HSC signatures showed upregulation in CIR and R1, which partially restored in R2, likely suggesting the dynamics of a specific HSC subpopulation. The endothelial signature was upregulated in CIR_{TAA}.

We performed linear regression to investigate whether there is a linear relationship between these cell scores in each sample with PP and CPA values on the other side. The CPA in the complete TAA dataset positively correlated with all cell scores (Figures 8B–8E). However, only macrophage and hepatocyte scores were linked to CPA in CCl_4. PP significantly correlated with all cell scores in HC_{TAA} and CIR_{TAA}, but this relationship did not follow with the inclusion of regression groups.

We further explored the expression of genes with the highest cirrhosis score and biomarker candidates for cirrhosis and CPA on a single-cell resolution in the aforementioned datasets. Cholangiocytes, macrophages including Kupffer cells, and dendritic cells showed the most robust signature with high cirrhosis score (Figures S18 and S19). Cirrhosis markers were downregulated in hepatocyte populations, and several CPA markers were upregulated in HSC and portal fibroblasts (Figure S18).

These findings suggest that the functional state of macrophages, non-parenchymal cells, and hepatocytes are linked to portal hemodynamics and CPA.

DISCUSSION

In this study, we profiled the hepatic transcriptomes of widely used animal models of liver fibrosis (CCl_4 and TAA) at key experimental time points of peak fibrosis induction and during one or two weeks of spontaneous disease regression. We focused on the early-phase regression to uncover molecular mechanisms for potential application in *in vivo* and clinical studies and meet an unfulfilled demand in biomarkers linked to the effectiveness of therapy or intervention.

Liver disease readouts showed that the CCl_4 model had more pronounced PH, liver injury, and fibrosis over the same time course. Yet, both models showed robust transcriptomic features of liver fibrosis such as matrix turnover, with collagens *Col5a2*, *Col1a1*, and Fibrillin 1 being highly upregulated; cell adhesion, and macrophage activation. The common downregulation involved metabolic pathways and PPAR, for which pharmacologic agonists are currently being investigated in clinical trials.[28,41]

We also described models' unique signatures and noted differential regulation in delta-Notch, Wnt, and Shh pathways, which could be considered for testing therapeutic strategies targeting their down- or upstream signaling. Despite having less pronounced liver disease

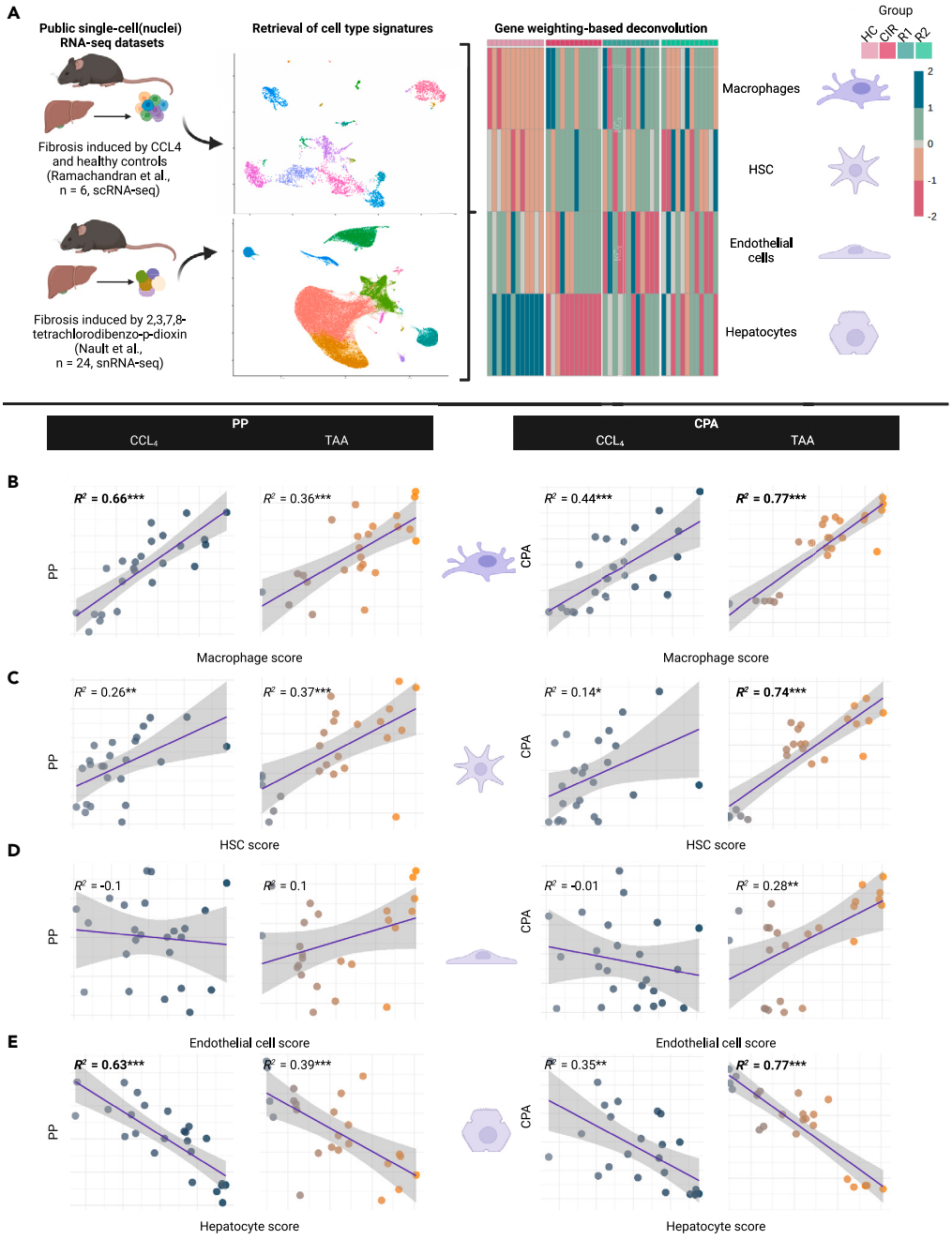


Figure 8. Deconvolution analysis shows cell-specific signatures in mouse RNA-seq datasets and their relationship to PP and CPA

(A) Cell type deconvolution of our study dataset using published single-cell datasets. The heatmap illustrates deconvoluted cell types with the most robust signatures.

(B–E) We found a linear correlation of the respective cell type score to PP and CPA. Cases shown in bold are those in which we consider cell score contributing to parameter variance (R2 > 0.6). The color indicates the respective cell type's low (gray) or high (blue for CCl4, orange for TAA) score value. HSC = hepatic stellate cells. PP = portal pressure. CPA = collagen proportionate area. $p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

See also Figures S16–S19.

readouts, CIR_{TAA} demonstrated transcriptional signs of integrin-regulated cell adhesion, which is notable in the context of the ongoing clinical trials with specific integrin inhibitors involving patients with biliary fibrosis.[42]

Cessation of liver injury for one week already resulted in significant upregulation of previously suppressed pathways in fibrosis, such as cytochrome family proteins, bile acids, and amino acid pathways (Figure S5). There were, nevertheless, only a few genes DE between one

and two weeks of regression. However, disease remnants' signature was still present, such as *Col5a3* and *Fbxo31* (indicators of the HSC activation) among upregulated genes in TAA. These findings support that, in future studies, regression periods should be split between early (cessation-related) and delayed regression to better distinguish a rather long-term fibrosis regression signature from tissue repair and metabolic recovery processes.

Solute carriers, cytochrome, and metabolic enzyme genes represented a large group perturbed by the fibrosis induction (e.g., *Cyp4a12a*, *Cyp2j5*, *Slc1a2*). Some of them were among high-ranked potential biomarkers, such as *Akr1c18* and Slc9a5 (associated with PP), or *Slc35c1* and *Cyp1a1* (regression). Further mechanistic studies are required to determine whether their pharmacological activation may relieve the metabolic phenotype following experimental fibrosis (suppression of bile acid, amino acid, and eicosanoid turnover).

We revealed that DE genes could be combined in four dynamic patterns, with either upregulation or downregulation in fibrosis and then up- or downregulation in regression. Notably, the groups following double upregulation or downregulation across the two models were smaller; this might indicate both a long-term effect of fibrogenesis or shifts in tissue homeostasis that would require therapeutic support to recover. A time-series experiment with transcriptomic profiling of liver biopsies in experimental fibrosis should be considered to provide a definitive answer.

Two independent methods (WGCNA and TMLE) revealed genes strongly linked to the liver disease readouts (i.e., PP, CPA, ALT, AST) or the fibrosis/regression status. For example, *Nucb2*, not previously studied in the PH context but found linked in our study, was involved in the development of experimental arterial hypertension in obesity;[43] *Igfbp7*—found upregulated in TAA fibrosis and downregulated in cured HCV infection—was previously reported as an activator of HSC, expressing its receptor Ifg1r, via *Smad* 2/3 system.[44,45] Moreover, we identified that *Chek1* was involved in early regression response and persists as upregulated through both weeks. This marker was already proposed as a hub gene in NAFLD and its low expression associated with better survival in hepatocellular carcinoma patients, yet, its mechanistic role remained elusive.[46] We validated *Igfbp7* and *Chek1* as transcriptional biomarker candidates in the human dataset after HCV cure.

Overlapping results of these two methods prioritized biomarker candidates having both well-described and understudied roles in liver disease.

We also found that, from more than 1,300 known transcriptional regulators in mice, 58 were perturbed in our models. Next to the nuclear receptors, we identified as significant those regulators whose role in liver diseases is primarily unknown, such as *Creb3l3*, *Akap8l*, *Tcf4*, and *Usf2*. Some TFs were included in the regulation of most of the biomarker candidates (*Hmga1*, *Tcf4*, *Mafk*), thus potentially revealing them as key regulators of transcriptional changes in cirrhotic PH.

Analysis of the multi-layer network, constructed from the prioritized biomarkers, their regulators, and database-mined interaction targets, led to identifying a few dozen hub genes, which were then further validated in human RNA-seq datasets, showing a profound presence in human cirrhotic PH, and achieving promising disease severity prediction in all datasets, with better accuracy in cirrhotic PH. Despite the availability of the RNA-seq data from human patients stratified according to the fibrosis severity, missing PP and liver injury readouts limited our opportunities for human validation. Nevertheless, we identified that TFs, such as *TCF4*, and candidate biomarkers such as *TREM2*, *MMP7*, *SCUBE1*, and *SPP1* are highly significant in both animal models and human data, linked to the severity of fibrosis. The transcriptional biomarker candidates should be validated in following studies with available histological and hemodynamic readouts, focusing on those we validated in the presented datasets. Profiling of hepatic transcriptomes from a large patient cohort with cured liver disease, such as after sustained virologic response to antiviral therapy in hepatitis C, will shed light on the clinical value of the regression markers identified in this study. Another yet-unexplored area is the dynamics of these markers in alcohol-related liver disease, especially in the context of alcohol abstinence, and metabolic dysfunction-associated liver disease.

Finally, cell type deconvolution analysis allowed the identification of signatures of macrophages, non-parenchymal cells, and hepatocytes and their relation to the liver disease readouts. Despite the availability of human scRNA-seq data, employing reference from a similar phenotype to experimental conditions is required to extract accurate cell signatures and thus achieve robust prediction. To further investigate cell contribution and improve accurate cell score prediction, RNA-seq can be paired with FACS for the cell types we could robustly deconvolute (e.g., Kupffer macrophages and endothelial cells) or with low sample-size scRNA-seq for a reference deconvolution.

To summarize, we characterized the hepatic transcriptomes of two commonly used liver fibrosis models, including during a well-defined regression phase, and thereby identified their functional similarities and differences. We identified transcripts with the potential to be developed into prognostic biomarkers or considered for therapeutic applications. Our transcriptomic dataset allowed us to select candidate biomarkers strongly associated with fibrosis progression and regression or key liver disease severity readouts, such as PP or ALT levels. The computational analysis provided insights into their regulatory networks and prioritized genes suitable for screening and assessment in human cohorts. Further mechanistic studies and validation experiments using human liver tissues of well-characterized patient cohorts are warranted to confirm our experimental findings.

Limitations of the study

In both mouse models studied here, the induction of fibrosis was caused by administration of a respective compound (CCl_4 of TAA). These models do not represent etiologies of diseases such as cholestatic and rare liver diseases. It is very likely that exploring bile duct ligation models in a similar manner will allow to reveal distinct set of markers, which could be of value in relation to cholestatic liver disease. The absence of publicly available human liver RNA-seq datasets during early regression phase (e.g., initiation of alcohol abstinence, hepatitis C treatment initiation) did not allow to investigate our early regression signatures in this human context, hence leaving opportunities for follow-up studies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Mouse model of liver fibrosis
- METHOD DETAILS
  - Key liver disease readouts
  - Transcriptomic assessment with RNA-sequencing
  - Differential expression (DE) analysis
  - Gene set analysis
  - Pseudotemporal dynamics and perturbation scores
  - Biomarkers discovery with TMLE and WGCNA
  - Identification of transcriptional regulators and their scoring
  - Identification of hub genes via multi-layer network analysis
  - Validation in human datasets
  - Cell type deconvolution
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109301.

ACKNOWLEDGMENTS

The research group of T.R. (O.P., P.K., K.Br., B.S.H., A.L.L., A.L., K.Ba., B.S., P.S.) and the study were supported by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development, the Christian Doppler Research Association, and Boehringer Ingelheim. A.F.R. is supported by Angelini Ventures S.p.A. Rome, Italy. The help of Alexander L. Lein and Ariton Ljoki with the *in vivo* experiments is thankfully acknowledged. The authors appreciate the technical assistance of Kerstin Zinober and Martha Seif in this study. The Biomedical Sequencing Facility staff is acknowledged for generating NGS data and related support. Separate figure elements were created with BioRender.com.

AUTHOR CONTRIBUTIONS

Conceptualization, T.R., P.S., O.P., P.K., S.G.K., L.P., and M.T.; methodology, O.P., P.K., K.Br., B.S.H., K.Ba., B.S., L.P.M.H.d.R., A.F.R., and E.S.; investigation, O.P., P.K., K.Br., B.S.H., K.Ba., B.S., F.J., T.F.B., and J.L.; software, O.P.; visualization, O.P.; writing – original draft, O.P. and T.R.; writing – review & editing, all authors; funding acquisition, T.R.; resources, O.P., P.K., F.J., T.F.B., and J.L.; supervision, T.R., L.P.M.H.d.R., and A.F.R.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 11, 2023
Revised: December 10, 2023
Accepted: February 16, 2024
Published: February 20, 2024

REFERENCES

1. Kisseleva, T., and Brenner, D. (2021). Molecular and cellular mechanisms of liver fibrosis and its regression. Nat. Rev. Gastroenterol. Hepatol. *18*, 151–166. https://doi.org/10.1038/s41575-020-00372-7.
2. Ginès, P., Krag, A., Abraldes, J.G., Solà, E., Fabrellas, N., and Kamath, P.S. (2021). Liver cirrhosis. Lancet *398*, 1359–1376. https://doi.org/10.1016/S0140-6736(21)01374-X.
3. Ramachandran, P., Dobie, R., Wilson-Kanamori, J.R., Dora, E.F., Henderson, B.E.P., Luu, N.T., Portman, J.R., Matchett, K.P., Brice, M., Marwick, J.A., et al. (2019). Resolving the fibrotic niche of human liver cirrhosis at single-cell level. Nature *575*, 512–518. https://doi.org/10.1038/s41586-019-1631-3.
4. Thanapirom, K., Suksawatamnuay, S., Tanpowpong, N., Chaopathomkul, B., Sriphoosanaphan, S., Thaimai, P., Srisoonthorn, N., Treeprasertsuk, S., and Komolmit, P. (2022). Non-invasive tests for liver fibrosis assessment in patients with chronic liver diseases: a prospective study. Sci. Rep. *12*, 4913. https://doi.org/10.1038/s41598-022-08955-x.
5. Shetty, S., Lalor, P.F., and Adams, D.H. (2018). Liver sinusoidal endothelial

cells - gatekeepers of hepatic immunity. Nat. Rev. Gastroenterol. Hepatol. *15*, 555–567. https://doi.org/10.1038/s41575-018-0020-y.

6. Wen, Y., Lambrecht, J., Ju, C., and Tacke, F. (2021). Hepatic macrophages in liver homeostasis and diseases-diversity, plasticity and therapeutic opportunities. Cell. Mol. Immunol. *18*, 45–56. https://doi.org/10.1038/s41423-020-00558-8.
7. Higashi, T., Friedman, S.L., and Hoshida, Y. (2017). Hepatic stellate cells as key target in liver fibrosis. Adv. Drug Deliv. Rev. *121*, 27–42. https://doi.org/10.1016/j.addr.2017.05.007.
8. Marcellin, P., Gane, E., Buti, M., Afdhal, N., Sievert, W., Jacobson, I.M., Washington, M.K., Germanidis, G., Flaherty, J.F., Aguilar Schall, R., et al. (2013). Regression of cirrhosis during treatment with tenofovir disoproxil fumarate for chronic hepatitis B: a 5-year open-label follow-up study. Lancet *381*, 468–475. https://doi.org/10.1016/S0140-6736(12)61425-1.
9. Mandorfer, M., Kozbial, K., Schwabl, P., Freissmuth, C., Schwarzer, R., Stern, R., Chromy, D., Stättermayer, A.F., Reiberger, T., Beinhardt, S., et al. (2016). Sustained virologic response to interferon-free therapies ameliorates HCV-induced portal hypertension. J. Hepatol. *65*, 692–699. https://doi.org/10.1016/j.jhep.2016.05.027.
10. Ortega-Ribera, M., Gibert-Ramos, A., Abad-Jordà, L., Magaz, M., Téllez, L., Paule, L., Castillo, E., Pastó, R., de Souza Basso, B., Olivas, P., et al. (2023). Increased sinusoidal pressure impairs liver endothelial mechanosensing, uncovering novel biomarkers of portal hypertension. JHEP Rep. *5*, 100722. https://doi.org/10.1016/j.jhepr.2023.100722.
11. Fernandez, M. (2015). Molecular pathophysiology of portal hypertension. Hepatology *61*, 1406–1415. https://doi.org/10.1002/hep.27343.
12. McConnell, M., and Iwakiri, Y. (2018). Biology of portal hypertension. Hepatol. Int. *12*, 11–23. https://doi.org/10.1007/s12072-017-9826-x.
13. Bao, Y.-L., Wang, L., Pan, H.-T., Zhang, T.-R., Chen, Y.-H., Xu, S.-J., Mao, X.-L., and Li, S.-W. (2021). Animal and organoid models of liver fibrosis. Front. Physiol. *12*, 666138. https://doi.org/10.3389/fphys.2021.666138.
14. Delire, B., Stärkel, P., and Leclercq, I. (2015). Animal models for fibrotic liver diseases: What we have, what we need, and what is under development. J. Clin. Transl. Hepatol. *3*, 53–66. https://doi.org/10.14218/JCTH.2014.00035.
15. Chen, P., Abacherli, L.E., Nadler, S.T., Wang, Y., Li, Q., and Parks, W.C. (2009). MMP7 shedding of syndecan-1 facilitates re-epithelialization by affecting α2β1 integrin activation. PLoS One *4*, e6565. https://doi.org/10.1371/journal.pone.0006565.
16. Jiang, L., Chu, H., Gao, B., Lang, S., Wang, Y., Duan, Y., and Schnabl, B. (2020). Transcriptomic profiling identifies novel hepatic and intestinal genes following chronic plus binge ethanol feeding in mice. Dig. Dis. Sci. *65*, 3592–3604. https://doi.org/10.1007/s10620-020-06461-6.
17. Hyun, J., Park, J., Wang, S., Kim, J., Lee, H.-H., Seo, Y.-S., and Jung, Y. (2016). MicroRNA expression profiling in CCl_4-induced liver fibrosis of Mus musculus. Int. J. Mol. Sci. *17*, 961. https://doi.org/10.3390/ijms17060961.
18. Schuck, R.N., Zha, W., Edin, M.L., Gruzdev, A., Vendrov, K.C., Miller, T.M., Xu, Z., Lih, F.B., DeGraff, L.M., Tomer, K.B., et al. (2014). The cytochrome P450 epoxygenase pathway regulates the hepatic inflammatory response in fatty liver disease. PLoS One *9*, e110162. https://doi.org/10.1371/journal.pone.0110162.
19. Karsdal, M.A., Daniels, S.J., Holm Nielsen, S., Bager, C., Rasmussen, D.G.K., Loomba, R., Surabattula, R., Villesen, I.F., Luo, Y., Shevell, D., et al. (2020). Collagen biology and non-invasive biomarkers of liver fibrosis. Liver Int. *40*, 736–750. https://doi.org/10.1111/liv.14390.
20. Lorena, D., Darby, I.A., Reinhardt, D.P., Sapin, V., Rosenbaum, J., and Desmoulière, A. (2004). Fibrillin-1 expression in normal and fibrotic rat liver and in cultured hepatic fibroblastic cells: modulation by mechanical stress and role in cell adhesion. Lab. Invest. *84*, 203–212. https://doi.org/10.1038/labinvest.3700023.
21. Klaas, M., Kangur, T., Viil, J., Mäemets-Allas, K., Minajeva, A., Vadi, K., Antsov, M., Lapidus, N., Järvekülg, M., and Jaks, V. (2016). The alterations in the extracellular matrix composition guide the repair of damaged liver tissue. Sci. Rep. *6*, 27398. https://doi.org/10.1038/srep27398.
22. Sojoodi, M., Erstad, D.J., Barrett, S., Salloum, S., Zhu, S., Qian, T., Gale, E., Jordan, V.C., Wang, Y., Li, S., et al. (2022). Peroxidasin deficiency re-programs macrophages toward pro-fibrolysis function and promotes collagen resolution in liver. J. Hepatol. *77*, S36. https://doi.org/10.1016/s0168-8278(22)00483-4.
23. Naim, A., Pan, Q., and Baig, M.S. (2017). Matrix Metalloproteinases (MMPs) in Liver Diseases. J. Clin. Exp. Hepatol. *7*, 367–372. https://doi.org/10.1016/j.jceh.2017.09.004.
24. Thiele, N.D., Wirth, J.W., Steins, D., Koop, A.C., Ittrich, H., Lohse, A.W., and Kluwe, J. (2017). TIMP-1 is upregulated, but not essential in hepatic fibrogenesis and carcinogenesis in mice. Sci. Rep. *7*, 714. https://doi.org/10.1038/s41598-017-00671-1.
25. Sun, W., Tang, Y., Tai, Y.-Y., Handen, A., Zhao, J., Speyer, G., Al Aaraj, Y., Watson, A., Romanelli, M.E., Sembrat, J., et al. (2020). SCUBE1 controls BMPR2-relevant pulmonary endothelial function: Implications for diagnostic marker development in pulmonary arterial hypertension. JACC. Basic Transl. Sci. *5*, 1073–1092. https://doi.org/10.1016/j.jacbts.2020.08.010.
26. Kökény, G., Calvier, L., and Hansmann, G. (2021). PPARγ and TGFβ-major regulators of metabolism, inflammation, and fibrosis in the lungs and kidneys. Int. J. Mol. Sci. *22*, 10431. https://doi.org/10.3390/ijms221910431.
27. Boyer-Diaz, Z., Aristu-Zabalza, P., Andrés-Rozas, M., Robert, C., Ortega-Ribera, M., Fernández-Iglesias, A., Broqua, P., Junien, J.-L., Wettstein, G., Bosch, J., and Gracia-Sancho, J. (2021). Pan-PPAR agonist lanifibranor improves portal hypertension and hepatic fibrosis in experimental advanced chronic liver disease. J. Hepatol. *74*, 1188–1199. https://doi.org/10.1016/j.jhep.2020.11.045.
28. Francque, S.M., Bedossa, P., Ratziu, V., Anstee, Q.M., Bugianesi, E., Sanyal, A.J., Loomba, R., Harrison, S.A., Balabanska, R., Mateva, L., et al. (2021). A randomized, controlled trial of the pan-PPAR agonist lanifibranor in NASH. N. Engl. J. Med. *385*, 1547–1558. https://doi.org/10.1056/NEJMoa2036205.
29. Geisler, F., and Strazzabosco, M. (2015). Emerging roles of Notch signaling in liver disease. Hepatology *61*, 382–392. https://doi.org/10.1002/hep.27268.
30. Bisgaard, H.C., Holmskov, U., Santoni-Rugiu, E., Nagy, P., Nielsen, O., Ott, P., Hage, E., Dalhoff, K., Rasmussen, L.J., and Tygstrup, N. (2002). Heterogeneity of ductular reactions in adult rat and human liver revealed by novel expression of deleted in malignant brain tumor 1. Am. J. Pathol. *161*, 1187–1198. https://doi.org/10.1016/S0002-9440(10)64395-7.
31. Lian, N., Jiang, Y., Zhang, F., Jin, H., Lu, C., Wu, X., Lu, Y., and Zheng, S. (2015). Curcumin regulates cell fate and metabolism by inhibiting hedgehog signaling in hepatic stellate cells. Lab. Invest. *95*, 790–803. https://doi.org/10.1038/labinvest.2015.59.
32. Reiberger, T., Angermayr, B., Schwabl, P., Rohr-Udilova, N., Mitterhauser, M., Gangl, A., and Peck-Radosavljevic, M. (2009). Sorafenib attenuates the portal hypertensive syndrome in partial portal vein ligated rats. J. Hepatol. *51*, 865–873. https://doi.org/10.1016/j.jhep.2009.06.024.
33. Schwabl, P., Brusilovskaya, K., Supper, P., Bauer, D., Königshofer, P., Riedl, F., Hayden, H., Fuchs, C.D., Stift, J., Oberhuber, G., et al. (2018). The soluble guanylate cyclase stimulator riociguat reduces fibrogenesis and portal pressure in cirrhotic rats. Sci. Rep. *8*, 9372. https://doi.org/10.1038/s41598-018-27656-y.
34. Van Steenkiste, C., Ribera, J., Geerts, A., Pauta, M., Tugues, S., Casteleyn, C., Libbrecht, L., Olievier, K., Schroyen, B., Reynaert, H., et al. (2011). Inhibition of placental growth factor activity reduces the severity of fibrosis, inflammation, and portal hypertension in cirrhotic mice. Hepatology *53*, 1629–1640. https://doi.org/10.1002/hep.24238.
35. Tao, R., Wang, Z.-F., Qiu, W., He, Y.-F., Yan, W.-Q., Sun, W.-Y., and Li, H.-J. (2017). Role of S100A3 in human hepatocellular carcinoma and the anticancer effect of sodium cantharidinate. Exp. Ther. Med. *13*, 2812–2818. https://doi.org/10.3892/etm.2017.4294.
36. Li, G., Zhang, J., Jiang, Q., Liu, B., and Xu, K. (2020). CREBH knockout accelerates hepatic fibrosis in mouse models of diet-induced nonalcoholic fatty liver disease. Life Sci. *254*, 117795. https://doi.org/10.1016/j.lfs.2020.117795.
37. Pan, X., Zhang, Y., Kim, H.-G., Liangpunsakul, S., and Dong, X.C. (2017). FOXO transcription factors protect against the diet-induced fatty liver disease. Sci. Rep. *7*, 44597. https://doi.org/10.1038/srep44597.
38. Königshofer, P., Brusilovskaya, K., Petrenko, O., Hofer, B.S., Schwabl, P., Trauner, M., and Reiberger, T. (2021). Nuclear receptors in liver fibrosis. Biochim. Biophys. Acta, Mol. Basis Dis. *1867*, 166235. https://doi.org/10.1016/j.bbadis.2021.166235.
39. Saito, K., Hirai, M.Y., and Yonekura-Sakakibara, K. (2008). Decoding genes with coexpression networks and metabolomics - "majority report by precogs". Trends Plant Sci. *13*, 36–43. https://doi.org/10.1016/j.tplants.2007.10.006.
40. Nault, R., Fader, K.A., Bhattacharya, S., and Zacharewski, T.R. (2021). Single-Nuclei RNA Sequencing Assessment of the Hepatic Effects of 2,3,7,8-Tetrachlorodibenzo-p-dioxin. Cell. Mol. Gastroenterol. Hepatol. *11*, 147–159. https://doi.org/10.1016/j.jcmgh.2020.07.012.
41. Puengel, T., Liu, H., Guillot, A., Heymann, F., Tacke, F., and Peiseler, M. (2022). Nuclear receptors linking metabolism, inflammation, and fibrosis in nonalcoholic fatty liver disease. Int. J. Mol. Sci. *23*, 2668. https://doi.org/10.3390/ijms23052668.
42. (2020). Phase 2a Evaluation of Safety Tolerability, and Pharmacokinetics of PLN-74809 in Patients with Primary Sclerosing Cholangitis (PSC). https://classic.clinicaltrials.gov/ct2/show/NCT04480840.

43. Mori, Y., Shimizu, H., Kushima, H., Hiromura, M., Terasaki, M., Tanaka, M., Osaki, A., and Hirano, T. (2017). Increased blood pressure in nesfatin/nuclebindin-2-transgenic mice. Hypertens. Res. *40*, 861–867. https://doi.org/10.1038/hr.2017.61.
44. Zhang, Y., Zhang, Q.-Q., Guo, X.-H., Zhang, H.-Y., and Liu, L.-X. (2014). IGFBPrP1 induces liver fibrosis by inducing hepatic stellate cell activation and hepatocyte apoptosis via Smad2/3 signaling. World J. Gastroenterol. *20*, 6523–6533. https://doi.org/10.3748/wjg.v20.i21.6523.
45. Adamek, A., and Kasprzak, A. (2018). Insulin-like growth factor (IGF) system in liver diseases. Int. J. Mol. Sci. *19*, 1308. https://doi.org/10.3390/ijms19051308.
46. Liu, J., Lin, B., Chen, Z., Deng, M., Wang, Y., Wang, J., Chen, L., Zhang, Z., Xiao, X., Chen, C., and Song, Y. (2020). Identification of key pathways and genes in nonalcoholic fatty liver disease using bioinformatics analysis. Arch. Med. Sci. *16*, 374–385. https://doi.org/10.5114/aoms.2020.93343.
47. Ramnath, D., Irvine, K.M., Lukowski, S.W., Horsfall, L.U., Loh, Z., Clouston, A.D., Patel, P.J., Fagan, K.J., Iyer, A., Lampe, G., et al. (2018). Hepatic expression profiling identifies steatosis-independent and steatosis-driven advanced fibrosis genes. JCI Insight *3*, e120274. https://doi.org/10.1172/jci.insight.120274.
48. Powell, N.R., Liang, T., Ipe, J., Cao, S., Skaar, T.C., Desta, Z., Qian, H.-R., Ebert, P.J., Chen, Y., Thomas, M.K., and Chalasani, N. (2023). Clinically important alterations in pharmacogene expression in histologically severe nonalcoholic fatty liver disease. Nat. Commun. *14*, 1474. https://doi.org/10.1038/s41467-023-37209-1.
49. Hernández-Gea, V., Campreciós, G., Betancourt, F., Pérez-Campuzano, V., Seijo, S., Díaz, A., Gallego-Durán, R., Olivas, P., Orts, L., Magaz, M., et al. (2021). Co-expression gene network analysis reveals novel regulatory pathways involved in porto-sinusoidal vascular disease. J. Hepatol. *75*, 924–934. https://doi.org/10.1016/j.jhep.2021.05.014.
50. Hamdane, N., Jühling, F., Crouchet, E., El Saghire, H., Thumann, C., Oudot, M.A., Bandiera, S., Saviano, A., Ponsolles, C., Roca Suarez, A.A., et al. (2019). HCV-induced epigenetic changes associated with liver cancer risk persist after sustained virologic response. Gastroenterology *156*, 2313–2329.e7. https://doi.org/10.1053/j.gastro.2019.02.038.
51. Nault, R., Saha, S., Bhattacharya, S., Sinha, S., Maiti, T., and Zacharewski, T. (2023). Single-cell transcriptomics shows dose-dependent disruption of hepatic zonation by TCDD in mice. Toxicol. Sci. *191*, 135–148. https://doi.org/10.1093/toxsci/kfac109.
52. Pagès, H., Carlson, M., Falcon, S., and Li, N. (2023). AnnotationDbi: Manipulation of SQLite-Based Annotations in Bioconductor. https://doi.org/10.18129/B9.Bioc.AnnotationDbi.
53. Blighe, K., and Lun, A. (2023). PCAtools: PCAtools: Everything Principal Components Analysis. https://doi.org/10.18129/B9.Bioc.PCAtools.
54. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Preprint at bioRxiv. https://doi.org/10.1101/002832.
55. Stephens, M. (2017). False discovery rates: a new deal. Biostatistics *18*, 275–294. https://doi.org/10.1093/biostatistics/kxw041.
56. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2016). Fast gene set enrichment analysis. Preprint at bioRxiv. https://doi.org/10.1101/060012.
57. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97. https://doi.org/10.1093/nar/gkw377.
58. Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenter, D.N., Hanspers, K., A Miller, R., Digles, D., Lopes, E.N., Ehrhart, F., et al. (2021). WikiPathways: connecting communities. Nucleic Acids Res. *49*, D613–D621. https://doi.org/10.1093/nar/gkaa1024.
59. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.
60. Conesa, A., Nueda, M.J., Ferrer, A., and Talón, M. (2006). maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. Bioinformatics *22*, 1096–1102. https://doi.org/10.1093/bioinformatics/btl056.
61. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559. https://doi.org/10.1186/1471-2105-9-559.
62. Hejazi, N.S., Cai, W., and Hubbard, A.E. (2017). biotmle: Targeted Learning for Biomarker Discovery. J. Open Source Softw. *2*, 295. https://doi.org/10.21105/joss.00295.
63. Oliveira de Biagi, C.A., Jr., Nociti, R.P., Brotto, D.B., Funicheli, B.O., Cássia Ruy, P.d., Bianchi Ximenez, J.P., Alves Figueiredo, D.L., and Araújo Silva, W., Jr. (2021). CeTF: an R/Bioconductor package for transcription factor co-expression networks using regulatory impact factors (RIF) and partial correlation and information (PCIT) analysis. BMC Genom. *22*, 624. https://doi.org/10.1186/s12864-021-07918-2.
64. Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. PLoS One *5*, e12776. https://doi.org/10.1371/journal.pone.0012776.
65. Türei, D., Valdeolivas, A., Gul, L., Palacio-Escat, N., Klein, M., Ivanova, O., Ölbei, M., Gábor, A., Theis, F., Módos, D., et al. (2021). Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. Mol. Syst. Biol. *17*, e9923. https://doi.org/10.15252/msb.20209923.
66. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504. https://doi.org/10.1101/gr.1239303.
67. Chin, C.-H., Chen, S.-H., Wu, H.-H., Ho, C.-W., Ko, M.-T., and Lin, C.-Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst. Biol. *8* (*Suppl 4*), S11. https://doi.org/10.1186/1752-0509-8-S4-S11.
68. Valdeolivas, A., Tichit, L., Navarro, C., Perrin, S., Odelin, G., Levy, N., Cau, P., Remy, E., and Baudot, A. (2019). Random walk with restart on multiplex and heterogeneous biological networks. Bioinformatics *35*, 497–505. https://doi.org/10.1093/bioinformatics/bty637.
69. Kuhn, M. (2008). Building Predictive Models inRUsing thecaretPackage. J. Stat. Software *28*. https://doi.org/10.18637/jss.v028.i05.
70. Wang, X., Park, J., Susztak, K., Zhang, N.R., and Li, M. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat. Commun. *10*, 380. https://doi.org/10.1038/s41467-018-08023-x.
71. (2023). The Lawson-Hanson Algorithm for Non-negative Least Squares (NNLS) Comprehensive R Archive Network (CRAN). https://cran.r-project.org/web/packages/nnls/index.html.
72. Königshofer, P., Hofer, B.S., Brusilovskaya, K., Simbrunner, B., Petrenko, O., Wöran, K., Herac, M., Stift, J., Lampichler, K., Timelthaler, G., et al. (2022). Distinct structural and dynamic components of portal hypertension in different animal models and human liver disease etiologies. Hepatology *75*, 610–622. https://doi.org/10.1002/hep.32220.
73. Königshofer, P., Brusilovskaya, K., Schwabl, P., Podesser, B.K., Trauner, M., and Reiberger, T. (2018). Invasive hemodynamic characterization of the portal-hypertensive syndrome in cirrhotic rats. J. Vis. Exp. *1*, 57261. https://doi.org/10.3791/57261.
74. (2012). Mus musculus Genome Assembly GRCm38 NCBI. https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000001635.20/.
75. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinf. *14*, 128. https://doi.org/10.1186/1471-2105-14-128.
76. Bembom, O., Petersen, M.L., Rhee, S.-Y., Fessel, W.J., Sinisi, S.E., Shafer, R.W., and van der Laan, M.J. (2009). Biomarker discovery using targeted maximum-likelihood estimation: application to the treatment of antiretroviral-resistant HIV infection. Stat. Med. *28*, 152–172. https://doi.org/10.1002/sim.3414.
77. Reverter, A., Hudson, N.J., Nagaraj, S.H., Pérez-Enciso, M., and Dalrymple, B.P. (2010). Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. Bioinformatics *26*, 896–904. https://doi.org/10.1093/bioinformatics/btq051.
78. Hammelman, J., Patel, T., Closser, M., Wichterle, H., and Gifford, D. (2021). Ranking reprogramming factors for directed differentiation. Preprint at bioRxiv. https://doi.org/10.1101/2021.05.14.444080.
79. Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. *49*, D605–D612. https://doi.org/10.1093/nar/gkaa1074.
80. Lin, C.-Y., Chin, C.-H., Wu, H.-H., Chen, S.-H., Ho, C.-W., and Ko, M.-T. (2008). Hubba: hub objects analyzer–a framework of interactome hubs identification for network biology. Nucleic Acids Res. *36*, W438–W443. https://doi.org/10.1093/nar/gkn257.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Carbon tetrachloride, anhydrous, ≥99.5%	Sigma-Aldrich	289116
Thioacetamid reagent grade, 98%	Sigma-Aldrich	172502
Deposited data		
Raw RNA-sequencing data	Original data	ENA: PRJEB63606
RNA-sequencing counts	Original data	ArrayExpress: E-MTAB-13804
Human RNA-seq validation dataset (F4 and F0)[47]	Ramnath et al.[47]	ENA: PRJEB27201
Human RNA-seq validation dataset (F3-4 and F0)[48]	Powell et al.[48]	GEO: GSE225740
Human RNA-seq validation dataset (cirrhotic portal hypertension)[49]	Hernández-Gea et al.[49]	GEO: GSE171248
Human RNA-seq validation dataset (cured from hepatitis C virus)[50]	Hamdane et al.[50]	ENA: PRJNA506130
Mouse single-cell RNA-seq validation dataset[3]	Ramachandran et al.[3]	GEO: GSE136103
Mouse single-nuclei RNA-seq validation dataset[51]	Nault et al.[51]	GEO: GSE148339
Experimental Models: Organisms/Strains		
Mouse: C57BL/6JRj: WT	Janvier Labs	SC-C57J-M
Software and Algorithms		
Scripts for data analysis	Original scripts	https://doi.org/10.5281/zenodo.10230613
AnnotationDbi[52]	Pagès et al.[52]	1.60
PCAtools[53]	Blighe and Lun[53]	2.10
DESeq2[54]	Love et al.[54]	1.38.1
ashr[55]	Stephens[55]	2.2-54
fgsea[56]	Korotkevich et al.[56]	1.24.0
enrichR[57]	Kuleshov et al.[57]	3.2
WikiPathways[58]	Martens et al.[58]	2019, mouse
MSigDB[59]	Liberzon et al.[59]	Hallmark and Reactome databases
maSigPro[60]	Conesa et al.[60]	1.70.0
WGCNA[61]	Langfelder and Horvath[61]	1.71
biotmle[62]	Hejazi et al.[62]	1.22
CeTF[63]	Oliveira de Biagi et al.[63]	1.9.0
GENIE3[64]	Huynh-Thu et al.[64]	1.24.0
OmnipathR[65]	Türei et al.[65]	3.2.0
Cytoscape[66]	Shannon et al.[66]	3.9.1
CytoHubba[67]	Chin et al.[67]	0.1
RandomWalkRestartMH[68]	Valdeolivas et al.[68]	1.18.0
Caret[69]	Kuhn[69]	6.0-94
MuSiC[70]	Wang et al.[70]	1.0.0
NNLS[71]	nnls: The Lawson-Hanson algorithm for non-negative least squares (NNLS) Comprehensive R Archive Network (CRAN)[71]	N/A

RESOURCE AVAILABILITY

Lead contact

Additional requests of information and resources related to the study should be addressed and will be fulfilled by the lead contact, Thomas Reiberger (thomas.reiberger@meduniwien.ac.at).

Materials availability

No new unique reagents were generated in this study.

Data and code availability

- The generated mouse RNA-sequencing data have been reposited at ENA and are publicly available as of the date of publication. Accession identifications are listed in the key resources table. This study used the following publicly available datasets: PRJEB27201, GSE225740, GSE171248, and PRJNA506130 (human transcriptomic datasets); GSE136103 and GSE148339 (mouse transcriptomic datasets).
- All original code has been uploaded at Zenodo repository, with DOI as indicated in the key resources table. It is publicly available as of the date of publication.
- Any additional information required to repeat the analysis of the data reported in the paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Mouse model of liver fibrosis

All experimental animal procedures were approved by the Austrian Federal Ministry of Education, Science and Research (BMBWF-V/3b/2020-0.398.007). Male C57BL/6JRj mice were housed and handled according to the standards of care and maintenance of the Center for Biomedical Research at the Medical University of Vienna. Animals were fed standard autoclaved rodent chow and had water access *ad libitum*. Fibrosis induction initiated in 10-week-old mice by the administration of carbon tetrachloride (CCl_4) or thioacetamide (TAA). CCl_4 was gavaged as a 20% v/v solution in olive oil at a dose of 2 mL/kg body weight three times a week. TAA was injected intraperitoneally in 0.9% saline at a dose of 150 mg/kg thrice a week. The dose was gradually increased over the first week for CCl_4 administration and over the first two weeks for TAA administration. The total time of fibrosis induction was 12 weeks, followed by one or two weeks of regression. A total of 48 C57BL/6JRj male mice with complete datasets on key liver disease readouts were selected for this study. The experimental cohort was divided into two different fibrosis models (CCl_4 and TAA) of 24 animals each (Figure 1A). Each model comprised four groups: healthy control with vehicle treatment (hereafter HC), positive control at the peak of the 12-week induction by CCl_4 or TAA respectively (CIR), one week of regression post-induction (R1), and two weeks of regression (R2) upon cessation of the stimulus (Figure 1B). Each group included six animals.

METHOD DETAILS

Key liver disease readouts

The following biological parameters were assessed in each animal: portal pressure (PP), collagen proportionate area (CPA), serum levels of alanine aminotransferase (ALT) and aspartate aminotransferase (AST) (Table 1).

The fibrosis assessment was performed by digital semiautomated histomorphometry of the relative fibrotic area on full liver lobe scans as previously described.[72] Briefly, the left lateral lobe of the liver was harvested, formalin-fixed, and paraffin-embedded. The 4 μm sections were stained by picrosirius red with fast green counterstain and scanned with Olympus VS200. The whole-slide images were then analysed using HALO (V3.3.2541.184, Indica Labs), and the percentage of collagen-positive staining (% CPA) was quantified with the Area Quantification module. PP was measured by direct cannulation of the portal vein under anaesthesia (ketamine 100mg/kg, xylazine 7.5mg/kg), using our established hemodynamic setup.[73] Serum levels of ALT and AST were measured as established surrogates of liver injury.

Transcriptomic assessment with RNA-sequencing

Left and right median liver lobes were harvested and immediately snap-frozen. RNA extraction from the right median lobe was performed using the RNeasy Mini Kit (Qiagen, Germany). RNA concentration and purity were accessed with NanoDrop One (Thermo Fisher Scientific, MA, United States) and submitted to the Biomedical Sequencing Facility of the CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences for further processing. The quality of the RNA was assessed with 2100 Bioanalyzer (Agilent, CA, USA), and RNA integrity $\geq$ 7.0 was used as the threshold (median integrity number: 7.65). The library preparation was done with a TruSeq Stranded mRNA LT sample preparation kit (Illumina, CA, USA). Samples were diluted, pooled into next-generation sequencing (NGS) libraries in equimolar amounts, and sequenced on Illumina HiSeq 4000 Systems (single-end, 50bp). The pipeline for base calling, QC, and demultiplexing was based on Picard tools (version 2.19.2) (Figure S1). The alignment was performed to Mus musculus GRCm38/mm10 Reference Genome Assembly.[74] The quality controls included sequencing depth evaluation, mapping percentage, and outlier detection with RNAseqQC (v. 0.1.4). Genes with no variance between samples, and those detected in fewer than 10% of the samples, were removed from the analysis. Gene annotation was performed with AnnotationDbi (v. 1.60). Dimensionality reduction was performed with PCAtools (2.10).

RNA-sequencing produced an average of 27 487 574 transcript reads per animal and a reference mapping rate of 98.2%. The annotation of reads resulted in 20 869 identified genes that were utilized in the subsequent analysis.

Differential expression (DE) analysis

DE was performed using the DESeq2 package with a detection limit of 2 counts[54] (v. 1.38.1). Count normalization with variance stabilizing transformation was performed. The Wald test was conducted with a design formula of ~ Model + Group for the pairwise analysis with ashr shrinkage (v. 2.2-54). The absolute $\log_2$FC > 1.5 was used as a cutoff to determine the direction of differential expression. The multiple-group comparison was performed with the likelihood ratio test (LRT) in the reduced (intercept-only; ~ 1) model. An adjusted p-value cutoff of < 0.01 was applied for both types of analysis to determine significant genes.

The Wald test was applied to characterize the models, compare groups, and validate the findings. The LRT test output provided a list of all genes significantly perturbed among all study groups, and this was utilized in pseudotemporal analysis to define the perturbation scores (described in own section).

For targeted analysis of genes involved in matrix turnover among identified DE features, genes included in the Reactome pathway R-HSA-1474244 were utilized.

Gene set analysis

We assessed gene sets using fgsea[56] (v. 1.24.0) and enrichR[57] (v. 3.2). For fgsea, genes were ranked based on the formula: rank = $\frac{\log_2 FC}{pvalue}$ and gene sets of WikiPathways[58] (2019, mouse) were used. In gene lists where such ranking was not possible (e.g., biomarker modules, network nodes), we used a hypergeometric test of EnrichR[75] with the MSigDB[59] hallmark and Reactome pathways, indicating the database in figures legend. Only pathways with a q-value $\leq$ 0.25 are reported. Usage of another threshold, or subsetting the top-ranked pathways, is noted in figure legends.

Pseudotemporal dynamics and perturbation scores

DE genes from the LRT test were used for pseudotemporal analysis. Study groups were defined as pseudo time points, and 3000 genes with the highest significance were clustered using DEGreport (v. 1.35.0). This allowed the identification of groups of genes with similar dynamics in the direction of fibrosis and regression.

Perturbation scores were calculated to determine the impact of cirrhosis and regression on gene expression. The cirrhosis score was defined as the impact of cirrhosis with HC as a baseline, and the regression score accounted for CIR as a baseline towards regression. First, stepwise regression fit was performed with maSigPro[60] (v. 1.70.0) for genes differentially expressed (DE) with the LRT test, resulting in goodness to fit (R^2) for each gene. Benjamini-Hochberg correction was used with a threshold of $p_{adjusted} < 0.01$. Next, perturbation scores were calculated based on the formula:

$$Perturbation = mean_{perturbed} - mean_{baseline} * R^2$$

where $mean_{perturbed}$ denotes the mean expression level in cirrhosis or R1+R2. Pearson correlation was used to evaluate the relation between the cirrhosis and regression scores, and pathway analysis was performed. These scores were used for later hub gene identification.

Biomarkers discovery with TMLE and WGCNA

To identify potential biomarkers associated with liver disease readouts, cirrhosis, and regression, we employed two statistical approaches. First, CCl_4 and TAA datasets were merged and vst-normalized for input for WGCNA[61] (v. 1.71) for gene-readout correlation analysis. The network was constructed using a signed type with a defined minimal module size of 10 genes and an optimal soft threshold of 5. All modules with a trait correlation strength > 0.6 and p < 0.01 were considered for further evaluation. We ranked genes separately for each readout (gene significance score). Finally, the upper and lower 10% genes were subsetted by this rank, given that both the strongest positive and negative gene significance indicate a strong association with a trait.

Second, we used the biotmle[62] (1.22) toolkit as a more conserved approach[76] to rank genes based on their association with liver disease readouts and study groups, accounting for other covariates. A cutoff of an adjusted p-value < 0.01 was applied, and genes in the upper 10% by the TMLE score were selected.

The candidate genes from both methods were overlapped to identify biomarkers robust to different sources of bias and confounding, resulting in four sets (for PP, CPA, CIR, and R1+R2). The candidate genes in these overlaps were considered potential biomarkers and were used in future analysis.

Identification of transcriptional regulators and their scoring

To identify transcriptional regulators and their scoring, the CeTF package[63] (v. 1.9.0) was used with the Reverter algorithm,[77] and an adjusted p-value < 0.01 was the threshold. This approach constructs a co-expression network between DE genes and transcriptional factors (TFs) in two conditions of interest (e.g., HC and CIR) and compares the co-expression correlation in these two conditions. TFs with the most differential co-expression are prioritized, allowing the identification of relevant TFs based on their target genes, even if these TFs are not differentially

expressed themselves. The curated list of mouse transcription TFs[78] and DE genes identified in the previous analysis with the Wald test were used for that.

We used GENIE3[64] to narrow down the identified TFs further, focusing only on those relevant to the animals' phenotype by scoring the TFs' interaction with biomarker overlaps for PP, CPA, CIR, and R1+R2. The Extra-Trees algorithm was preferred to the random forest to avoid subsampling bias and to include all possible TFs-genes combinations. The background interactions were defined as TFs-genes interactions with GENIE3 weight < 0.05 and removed.

Identification of hub genes via multi-layer network analysis

We constructed a multi-layer network for hub gene identification, consisting of the following layers (bottom-top):

- Gene regulatory network (GRN) – prioritized TFs served as network nodes;
- Gene co-expression network (GCN) – biomarkers from overlaps as nodes;
- Protein-protein interaction network (PPIN) – inferred targets of the GCN layer as nodes.

The GRN-GCN inter-layer edges were based on TFs-genes interactions identified with GENIE3 analysis. The GCN layer had intra-layer edges, to obtain which we performed co-expression analysis using WGCNA's similarity over topological overlaps with signed type and power = 5.

To construct the PPIN, we first used OmnipathR[65] (v. 3.2.0), and for each gene of the GCN the interaction target was identified using platform's *omnipath*, *kinaseextra*, *pathwayextra*, *ligrecextra*, and enzyme-substrate databases. These target proteins served as PPIN nodes, and interactions were added as GCN-PPIN edges. Secondly, we added intra-layer PPIN interactions from STRINGdb[79] (v. 2.10.0), with the experimental score threshold ≥ 0.7 without the addition of new proteins. When the experimental score was unavailable, a STRINGdb confidence score of ≥ 0.7 was applied.

The resulting multi-layer network was imported into Cytoscape (v. 3.9.1) for visualization and scoring with CytoHubba.[67]

We then identified four network cores (NC), for each biomarker sets: PP, CPA, CIR, and R1+R2. For each of them, the GCN subset with respective biomarkers was used together with all GRN and PPIN nodes connected to them, including intra-layer connections of PPIN. Genes for each NC were prioritized via scoring based on a combination of biological knowledge, the double screening method,[80] and machine learning.

For biological knowledge, we used previously calculated perturbation scores (cirrhosis and regression scores) and gene significance for the PP and CPA.

In the double screening method, the maximum neighbourhood component (MNC) and density of maximum neighbourhood component (DMNC) scores were selected. The reasoning behind that was that MNC is a global network score that measures the size of the largest connected component that a node belongs to, while DMNC is a local network score that measures the density of the subnetwork consisting of a node and its neighbours. We used either one of them (as described in the formulas), or two where applicable. By using MNC as a pre-filtering step, the most essential and highly connected nodes in the network were prioritized. DMNC was used to further refine the selection of the genes most relevant to the nodes mentioned above.

Random walk with restart (RandomWalkRestartMH, v. 1.18.0) was prioritized from machine learning methods. It initiated on seed nodes, represented as TFs (GRN layer); then, it iteratively transposed to their neighbours (in the GCN layer), until, eventually, reaching the last node (in the GCN or PPIN layers), or resetting and starting a new walk, hence ranking the nodes with the walk probability. We set the probability of restarting to 0.7, τ (the likelihood of reset in a layer other than GRN) to 0, and used the above-mentioned inter- and intra-layer connections as weights.

Consequently, we identified NC with the following formula:

$$NC_{var} = GRN_{score(var)} \nabla GCN_{score(var)} \nabla PPIN_{score(var)} \nabla Interlayer_{score(var)}$$

where *del* operator indicates joining operation with preserving inter-layer edges, and *var* denotes one of four biomarker sets. The scores were calculated as following:

$$GRN_{score(var)} = argmax_5(V_{DMNC})$$

$$GRN_{score(var)} = \text{argmax}_5\left(\frac{V_{\text{perturbation}}}{V_{\text{DMNC}}}\right)$$

$$PPIN_{score(var)} = argmax_{10}(V_{MNC}) \rightarrow argmax_5(V_{DMNC})$$

$$Interlayer_{score(var)} = argmax_{10}(VRWR) \rightarrow argmax_5(V_{DMNC})$$

where V_x = set of vertices and their edges for this *var* obtained using one of the indicated scoring approaches; *DMNC* = density of maximum neighbourhood component; *MNC* = maximum neighbourhood component; *RWR* = random walk with restart; $argmax_n$ = function retrieving n components of the list based on the indicated scoring; arrow = chain operation.

If several genes from the $argmax_n$ set had identical scores, all of them were included in the result. If less than five TFs were involved in GRN layer for a biomarker set, all of them were included in the result.

Validation in human datasets

For validation, human RNA-seq data from the following studies were utilized:

- PRJEB27201: liver biopsies from patients with F4 and F0 (n = 25).[47]
- GSE225740: liver biopsies from patients with steatohepatitis and non-diseased (n = 74); patients with F3-4 and F0 (n = 46).[48]
- GSE171248: liver biopsies from patients with cirrhotic portal hypertension and non-diseased (n = 16).[49]
- PRJNA506130: liver biopsies from patients cured from hepatitis C infection (HCV, n = 3) and patients prior treatment (n = 3).[50]

In all scenarios, we investigated the presence of the identified hub genes among DE genes. The Wald test was applied for differential expression with an adjusted p-value threshold of ≤ 0.01 (except for PRJNA506130 due to the cohort size limitations, where we considered non-adjusted p-value threshold ≤ 0.05 and, additionally, Wald statistics threshold of $\geq |1.5|$). Using Caret[69] (v. 6.0-94), machine learning algorithms were employed to validate the hub genes, identifying those with the potential for predicting disease severity. Namely, naïve Bayes was preferred for feature selection, considering the involvement of various transcriptional factors in their expression and the assumption that both their regulation and targets follow conditional independence. The candidate biomarker combinations were scored for the classification task (prediction of the above-mentioned disease severity in each dataset), with five repeats of 10 cross-validations applied. Given the human datasets' modest size and overfitting risks, the random forest in Leave-One-Out cross-validation (mtry = 1) was selected. For each classification task, results included the best set of genes, the accuracy of the random forest-based model, and the Kappa statistic.

Cell type deconvolution

We utilized MuSiC[70] and NNLS[71] algorithms to estimate the cell type proportions in the study dataset for cell type deconvolution. The reference objects were constructed from single-cell RNA-seq (scRNAseq) of CD45-positive cells in healthy and 4-week CCl_4 mice models[3] (n = 2) and whole liver single-nuclei RNA-seq (snRNA-seq) of models induced with 2,3,7,8-Tetrachlorodibenzo-p-dioxin[51] (TCDD, n = 24). Only significant DE markers of each cell cluster, identified with Seurat's FindAllMarkers (parameters: absolute logFC threshold = 1.5, test = Wilcoxon, min.pct = 0.25), were used during deconvolution. MuSiC or NNLS were prioritized depending on their ability to detect cell signatures as non-zero values in all samples. The deconvoluted cell type proportions were compared between groups and models using the Wilcoxon rank-sum test. We used linear regression models with the formula *parameter* ~ *predicted cell score* to investigate the association between continuous readouts (PP and CPA) and the predicted scores.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed in the GNU R environment. The Wilcoxon rank-sum test was used to identify significant differences in the respective experimental measurements in n = 48 animals, and the data were represented as mean ± SEM. The whole-slide images were then analyzed using HALO software, and the percentage of collagen-positive staining was quantified with the Area Quantification module. DESeq2 package was utilized for differential expression analysis of RNA-sequencing data. The Wald test was conducted with a ~ Model + Group design formula for the pairwise analysis with ashr shrinkage. The absolute log2FC > 1.5 was used as a cutoff to determine the direction of differential expression. The multiple-group comparison was performed with the likelihood ratio test (LRT) in the reduced (intercept-only; ~ 1) model. An adjusted p-value cutoff of < 0.01 was applied for both types of analysis to determine significant genes. Statistical analysis of gene sets was performed with fgsea and enrichR, with q-value ≤ 0.25. For validation experiments with human data, where the Wald test was applied for differential expression with an adjusted p-value threshold of ≤ 0.01 (except for PRJNA506130 due to the cohort size limitations, where we considered non-adjusted p-value threshold ≤ 0.05 and, additionally, Wald statistics threshold of $\geq |1.5|$). For cell type deconvolution, MuSiC and NNLS algorithms were employed using the Wilcoxon test from the Seurat package with absolute logFC threshold = 1.5 and min.pct = 0.25.