

Pathobiochemical signatures of cholestatic liver disease in bile duct ligated mice

**Kerstin Abshagen^{1¶}, Matthias König^{2¶}, Andreas Hoppe², Isabell Müller¹, Matthias Ebert⁴, Honglei
Weng⁴, Herrmann-Georg Holzhütter², Ulrich M. Zanger³, Johannes Bode⁵, Brigitte Vollmar¹, Maria
Thomas³, Steven Dooley⁴**

¹Institute for Experimental Surgery, University Medicine Rostock, 18057 Rostock, Germany

²Institute for Biochemistry, Computational Systems Biochemistry Group, Charité University Medicine Berlin,
10117 Berlin, Germany

³Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, 70376 Stuttgart, and University of
Tuebingen, Tuebingen, Germany

⁴Department of Medicine II, Section Molecular Hepatology, Medical Faculty Mannheim, Heidelberg
University, 68167 Mannheim, Germany

⁵Department for Gastroenterology, Hepatology and Infectiology, Heinrich-Heine University of Düsseldorf,
40225 Düsseldorf

¶These authors contributed equally to this work

*Correspondence should be addressed to:

Kerstin Abshagen, PhD

Institute for Experimental Surgery

Rostock University Medical Center

Schillingallee 69a

18057 Rostock

Germany

phone: +49 381 494 2503

fax: +49 381 494 2502

E-mail: kerstin.abshagen@uni-rostock.de

1 Email addresses:

2 matthias.koenig@charite.de

3 andreas.hoppe@charite.de

4 Maria.Thomas@ikp-stuttgart.de

5 isabell-mueller86@web.de

6 matthias.ebert@umm.de

7 honglei.weng@medma.uni-heidelberg.de

8 hergo@charite.de

9 Uli.Zanger@ikp-stuttgart.de

10 Johannes.Bode@med.uni-duesseldorf.de

11 brigitte.vollmar@med.uni-rostock.de

12 Steven.Dooley@medma.uni-heidelberg.de

13

14 **Running title:** Cellular and molecular signatures of cholestasis

Abstract

Background: Disrupted bile secretion leads to liver damage characterized by inflammation, fibrosis, eventually cirrhosis, and hepatocellular cancer. As obstructive cholestasis often progresses insidiously, markers for the diagnosis and staging of the disease are urgently needed. To this end we compiled a comprehensive data set of serum markers, histological parameters and transcript profiles at 8 time points after bile duct ligation (BDL) in mice.

Results: Statistical analysis of the more than 6,000 data points revealed distinct temporal phases of disease. Putting time-courses of biochemical, histochemical and mRNA transcripts (=factors) together, time-series correlation analysis defined 6 distinct clusters related to different phases of disease progression. The count of CTGF-positive cells provided the most reliable overall measure for disease progression at the histological level, bilirubin at biochemical level, and metalloproteinase inhibitor 1 (Timp1) on transcript level. Prominent molecular events exhibited by strong transcript peaks are found for the transcriptional regulator Nr0b2 (Shp) and 1,25-dihydroxyvitamin D(3) 24-hydroxylase (Cyp24a1) at 6 hours. Based on these clusters we constructed a decision tree to identify factor combinations that can be used as markers for specific time intervals of disease progression. Best prediction of onset of disease is achieved by fibronectin (Fn1), the initial phase by (Cyp1a2), transition to the perpetuation phase by collagen α -1 I (Col1a1), and transition to the progression phase by interleukin 17-a (Il17a), with early and late progression separated by Col1a1. Notably, these predictions remained stable even for randomly chosen small sub-sets of factors from the clusters.

Conclusion: Our explorative study demonstrates the existence of a reproducible and well-coordinated temporal gene-expression program underlying disease progression in BLD cholestasis. To further elucidate the regulatory network behind this expression program additional studies are needed which have to include knowledge on transcription factor activation due to accumulation of bile salts and shared transcriptions factor binding motifs of genes belonging to the same transcript cluster.

Keywords: liver injury, mouse, systems biology, fibrosis, cell proliferation, bile duct ligation, cholestasis, morphological profiling, VirtualLiver Network

Background

Cholestatic liver diseases are caused by an impaired flow of the bile from the liver to the duodenum. The main component of the bile are bile salts, strong detergents required for the emulgaion of lipids in the gut and for the extraction of lipids from the apical membrane of hepatocytes into the bile fluid. Moreover, the bile fluid, contains numerous endogenous products (e.g. bilirubin) and potentially toxic compounds cleared by the liver. Hence, accumulation of bile compounds in the cholestatic liver causes unspecific cell damage that initiates a cascade of inflammatory events: Necrosis of hepatocytes and cholangiocytes, activation of macrophages releasing pro-inflammatory cytokines and chemokines, neutrophil infiltration, cholangiocyte and hepatocyte proliferation, stellate cell activation with progressive fibrosis causing secondary biliary cirrhosis, and ultimately leading to liver failure [1].

Multiple pathologies may represent the primary trigger of impaired bile flow like defects in the export of the bile from hepatocytes to the extracellular space (hepatocellular cholestasis) or obstruction of bile ducts by gall stones or local tumor impingement (extrahepatic cholestasis) [2]. Among the most common causes of cholestatic liver disease in the adult population are primary biliary cirrhosis (PBC) and primary sclerosing cholangitis (PSC), while biliary atresia and Alagille syndrome are commonly found in the pediatric population [3-5]. One of the main causes of cholestasis is biliary tract obstruction, which is a common mechanism of hepatic injury in a variety of clinical settings, including obstructing neoplasms, post-operative bile duct injury, biliary atresia, sclerosing cholangitis, and primary biliary cirrhosis. The corresponding experimental model to induce obstructive cholestatic injury in mice and rats is surgical bile duct ligation (BDL) [6, 7], which results in stereotypical histopathological changes of cholestasis. The BDL experimental model has been well described and evaluated in rats and mice and is widely used to study cholestatic liver injury and fibrogenesis [8].

Chronic liver diseases (CLD) like cholestasis represent with characteristic temporal morphologic, biochemical and molecular changes in liver and serum. For instance, in the BDL model of cholestasis an early phase of acute hepatocyte injury is followed by a proliferative response, up-regulation of pro-inflammatory cytokines and pro-fibrotic metabolic enzymes, and subsequent development of fibrosis after around 7 days [7, 8]. Such alterations can be exploited to identify biomarkers reporting on specific stages

of disease progression and severity of the disease. Hitherto semi-quantitative morphological scoring is the standard technique for grading in CLD. However, nowadays the availability of high-throughput technologies enables to flank the histological assessment of injured tissue with a comprehensive molecular profiling of gene transcripts, gene products (proteins) and metabolites. Such analyses not only provide a more detailed characterization and thus more refined staging of disease progression, they also lead to a deeper understanding of the molecular networks governing the histological and physical alterations observed at higher scales of tissue organization. The identification of key processes triggering the transition between different phases of disease progression requires mathematical analyses which take into account multiple parallel processes, process dynamics, and experimental data from different levels of cellular organization.

In this study we explicitly wanted to systemize existing and newly acquired knowledge on the molecular biomarkers of cholestasis and to analyze the time-dependent disease progression following BDL in mice in a comprehensive manner. Our approach sets itself apart from existing studies which either described the time course of a limited number of selected parameters after BDL [7, 8] or provided gene expression profiling for a limited number of time points, thereby missing the acute damage after BDL in the first 24h and long-term effects after 7 days [9]. None of these preceding studies applied predictive models based on the acquired time course data. Therefore, a central aim of our study was to identify molecular markers for the temporal progression of BDL cholestasis by correlating high-accuracy image data and transcriptional profiles of preselected targets with pathobiochemical markers of obstructive cholestasis. We collected from 8 different time points after BDL more than 6,000 experimental data points, comprising immunohistochemistry, biochemistry and molecular profiling measures. Statistical methods were applied to unravel robust interrelations in this large-scale data set and to find clusters of parameters corresponding to characteristic time profiles in cholestatic injury induced by BDL. We correlated level and timing of pathophysiological events with transcriptional changes in order to define molecular markers, and developed predictive decision trees on a subset of biomarkers for the assessment of the different disease phases occurring during the development of cholestasis.

Materials and Methods

Ethic Statement

All experiments were approved by the local government Landesamt für Landwirtschaft, Lebensmittelsicherheit und Fischerei Mecklenburg-Vorpommern (LALLF M-V/TSD/7221.3-1.2-049/09) and performed in accordance with the German legislation on protection of animals and the National Institutes of “Health Guide for the Care and Use of Laboratory Animals” (Institute of Laboratory Animal Resources, National Research Council; NIH publication 86-23 revised 1985).

Mice

Male C57BL/6J (Charles River Laboratories, Sulzfeld, Germany) at 8–10 weeks of age with a body weight of 23–26 g were kept on water and standard laboratory chow ad libitum.

Surgical procedure and experimental groups

Mice were anesthetized by breathing isoflurane (1.5 vol%). BDL was performed after midline laparotomy. The common bile duct was ligated three times with 5-0 silk and transected between the two most distal ligations. Sham operation was performed similarly, except for ligation and transection of the bile duct (0 h, n=5). All surgical procedures were performed under aseptic conditions. Animals were allowed to recover from anesthesia and surgery under a red warming lamp and were held in single cages until subsequent experiments followed at postoperative hours 6, 12, 18 and 30 (n=5 animals per time point), and at 2, 5 and 14 days after BDL (n=5 animals per time point). Sham-operated animals without BDL served as controls (n=5). To analyze the regenerative response in regard to proliferation of different cell types, 5-bromo-2-deoxyuridine (BrdU; 50 mg/kg bw ip) was injected 1 h prior to harvest of liver tissue. BrdU incorporation into DNA was analyzed by immunohistochemistry. To obtain blood and liver samples, mice were killed at the indicated time points. Liver tissue was partially embedded in paraffin for morphology analysis and snap frozen for molecular biology and biochemistry analyses. In addition, liver tissue served for the parallel Taqman qRT-PCR using microfluidic Fluidigm Biomark™ platform (Fluidigm, CA, USA).

Hematological measurements and plasma enzyme levels

Animals were anesthetized and exsanguinated by puncture of the vena cava inferior. Red blood cell and blood platelets count, hemoglobin, and hematocrit were assessed with an automated cell counter

(Sysmex KX-21, Sysmex). Plasma activities of alanine aminotransferase (ALT), aspartate aminotransferase (AST) and glutamate dehydrogenase (GLDH) were measured spectrophotometrically.

Assays

EDTA plasma served for the analysis of albumin as a parameter of liver function, which was determined with a commercially available enzyme-linked immunosorbent assay kit in accordance with the manufacturer's instructions (Assaypro, MO, USA).

Histopathology and image analysis

Liver tissue samples were fixed in formalin for 2 to 3 days and embedded in paraffin. 5 µm sections were stained with hematoxylin and eosin (H&E) for routine examination and quantification of bile infarcts. Sirius red staining served for quantification of collagen deposition. All samples from a series of experiments were stained simultaneously and evaluated in a blinded manner. For histomorphometric analysis, images of 20 random low power fields (x10 magnification, Olympus BX 51, Hamburg, Germany) were acquired with a Color View II FW camera (Color View, Munich, Germany) and evaluated using an image analysis system (Adobe Photoshop, Adobe Systems, Uxbridge, UK). Fibrosis deposition was quantified as a percentage of Sirius red stained area compared with the total section area. The surfaces of large centrilobular veins and large portal tracts were excluded from this calculation. Bile infarcts were quantified in H&E-stained sections in a similar manner and the percentage of the focal necrosis surface to the whole liver section area was assessed.

Immunohistochemistry and image analysis

For analyzing DNA-incorporated BrdU in liver cells, 4 µm sections collected on poly-L-lysine-coated glass slides were incubated with monoclonal mouse anti-BrdU antibody (1:50; Dako Cytomation, Hamburg, Germany) overnight at 4°C, followed by horseradish-peroxidase (HRP)-conjugated goat anti-mouse immunoglobulin (LSAB kit plus; Dako). Sites of peroxidase-binding were detected by 3,3'-diaminobenzidine (Dako). Sections were counterstained with hemalaun. BrdU-positive hepatocellular nuclei were counted in a blinded manner within 30 consecutive high power fields (HPF) (x40 objective, numerical aperture 0.65) and are given as cells/mm². In analogy, BrdU-expressing non-parenchymal cells were assessed and also given as cells/mm².

To specify the proliferative response of non-parenchymal cells upon BDL, we performed double immunohistochemistry of BrdU with specific markers for different liver cells: F4-80/BrdU for Kupffer cells and SM22 α /BrdU for BEC. For each protocol, the immune-staining procedure for the specific marker was conducted after the BrdU staining protocol. Resident liver macrophages were analyzed using the F4-80 antigen (1:10; Serotec, Oxford, UK). Overnight incubation (4°C) with the first antibody (polyclonal rat anti-F4-80) was followed by alkaline-phosphatase (AP) conjugated mouse anti-rat immunoglobulin (1:200; Santa Cruz Biotechnology, Santa Cruz, CA, USA). The sites of AP-binding were detected using the chromogen fuchsin (Dako).

BECs and oval cells were detected by overnight incubation (4°C) with a polyclonal rabbit anti-SM22 α antibody (1:100; Abcam, Cambridge, UK) followed by AP conjugated goat anti-rabbit immunoglobulin as secondary antibody (1:100; Dako). The sites of AP-binding were detected by Permanent Red (Dako).

Slides were viewed under a light microscope (Olympus BX 51) and the number of BrdU-positive cells co-expressing F4-80 or SM22 α were counted in a blinded manner within 30 consecutive high power fields (HPF) (x40 objective, numerical aperture 0.65) and are given as cells/mm².

Antibodies for detection of α -SMA in HSCs and of S100a4-positive cells were from DAKO (M0851 and A5114, 1:500 and 1:200 dilution, respectively). CTGF antibody was from Santa Cruz (sc-1439, 1:200 dilution). Sections were de-paraffinized in serial ethanol dilutions. After a PBS wash, sections were transferred into 10 mM sodium citrate buffer (pH 6.0) and antigen unmasking was performed in a microwave. After cooling down, sections were incubated in peroxidase blocking reagent (Dako) for 1 h and with primary antibodies overnight at 4°C. EnVision peroxidase (Dako) was applied for 1 h at room temperature after a PBS wash next day. Sections were developed with diaminobenzidine for 5 minutes. The number of α -SMA-, CTGF- and S100a4-positive cells was quantified under a Leica light microscope (x20) by counting three fields.

High-throughput quantitative Taqman RT-PCR analysis

Total RNA was isolated from the liver tissue samples using RNeasy Mini Kit including on column genomic DNA digestion with RNase free DNase Set (Qiagen, Hilden, Germany). RNA was reverse transcribed to cDNA with TaqMan Reverse Transcription Reagents (Applied Biosystems, Darmstadt, Germany). For quantitative real-time PCR, we used the Fluidigm's Biomark high-throughput qPCR chip platform

(Fluidigm Corporation, San Francisco, CA, USA) with pre-designed gene expression assays from Applied Biosystems, according to the manufacturer's instructions [10]. Data were analyzed using the ddCt method and expression values were normalized to the expression levels of the Gapdh gene.

Statistical data analysis

Dimension reduction: A one-way analysis of variance (ANOVA) was applied to reduce the data set to the parameter subset showing significant ($p_{adj} < 0.05$) up- or down-regulation during the time course. Multiple testing correction was performed using the Holm procedure [11]. To specifically test for the initial changes a two-tailed unpaired t-test (Welch correction for nonhomogeneity of variance) was performed for all factors between the classes 0h and 6h.

Correlation analysis: Correlation between two factor time courses was calculated via Y^{S3} , a modified correlation coefficient based similarity measure for clustering time-course gene expression data [12]. The correlation $Y_{i,j}^{S3}$ between two factors i and j is the linear combination of three terms: i) a classical correlation part based on Spearman correlation $S_{i,j}^*$, ii) $A_{i,j}^{**}$ accounting for the similarity in changes between the two time courses, and iii) $M_{i,j}^*$ comparing the location of the minimum and maximum values of the time course (see [12] and Supplement 2, correlation analysis for definitions). $S_{i,j}^*$ is calculated on individual data points, $A_{i,j}^{**}$ and $M_{i,j}^*$ on the mean time courses averaged over the repeats per time point. For the calculation of the similarity of changes within time courses, $A_{i,j}^{**}$, we used Spearman (S) correlation which is more robust against outliers as Pearson (R) correlation:

$$A_{i,j}^{**} = (S(d_i, d_j) + 1)/2$$

$$Y_{i,j}^{S3} = \omega_1 S_{i,j}^* + \omega_2 A_{i,j}^{**} + \omega_3 M_{i,j}^*$$

In the analysis the following weights were used: $\omega_1 = 0.5$, $\omega_2 = 0.3$, $\omega_3 = 0.2$. All reported correlations are $\bar{Y}_{i,j}^{S3}$ values in the interval [-1,1]:

$$\bar{Y}_{i,j}^{S3} = 2(Y_{i,j}^{S3} - 0.5)$$

Cluster analysis of the correlation matrix used complete-linkage hierarchical clustering with Euclidian distance measurement. This combination of complete-linkage with Y^S provided the best enrichments on gene-expression time-series in a recent comparisons of methods [13, 14]. The number of clusters $N_c = 6$, was selected as maximum number of clusters so that all clusters contain more than one factor.

Normalization of factors was performed separately for each factor f_k for all time points $i = 1, \dots, N_t$ and repeats $r = 1, \dots, N_r$ with $N_t = 8$ and $N_r = 5$ via

$$\bar{f}_k(t_{i,r}) = \frac{(f_k(t_{i,r}) - \langle f_k \rangle)}{\max(f_k) - \min(f_k)}$$

Decision Trees: For the prediction of distinct time points of disease progression a regression tree with the mean normalized factor data of the 6 clusters as predictor variables and the log transformed time points \tilde{t}_i as dependent variable was fitted based on recursive partitioning using *rpart* [15]. Logarithmic transformation was performed to obtain approximately equal-distant time points.

$$\tilde{t}_i = \log(t_i + 1)$$

The regression tree was fitted using the complete trainings set ($N_s = N_t N_r = 40$), with the minimum number of observations in a node for which a split was computed being 6, the minimum number of observations in a terminal node as 2, and the complexity parameter $c_p = 0.01$. The splitting criterion deciding which predictor variable gives the best split for nodes in the regression tree was $S_T - (S_L + S_R)$, with $S_T = \sum(\tilde{t}_i - \langle \tilde{t} \rangle)^2$ the sum of squares for node T, and S_T and S_L the sums of squares for the left and right child. A leave-one-out approach was used to test the robustness of the predicted time classes and predictive performance: For each sample ($N_s = 40$ mice), the regression tree was generated under the exclusion of data from the sample, with subsequent prediction on the left out test data (see Supplement 2, decision trees).

The predictive capacity of the regression tree was evaluated using all single combinations of individual factors from the clusters (88572), and a random subset of 10000 two factor combinations from each cluster. Predictions for a given combination of factors $\langle f_1 \rangle, \dots, \langle f_6 \rangle$ from the 6 clusters were scored using the root mean square distance on log scale d , with the best combination of factors minimizing d

$$d(\langle f_1 \rangle, \dots, \langle f_6 \rangle) = \frac{1}{N_s} \sqrt{\sum_{i=1}^{N_s} (t_i^{pre} - t_i^{exp})^2}$$

All computations were performed in R with source code, data and the full analysis available in Supplement S2 and from <https://github.com/matthiaskoenig/bdl-analysis>.

Results and Discussion

Temporal changes of biochemical, cellular and histochemical markers after BDL

In mice, BDL over 14 days induces time dependently progressing stages of a secondary biliary CLD. The first week after BDL begins with an acute cholestatic injury associated with necroinflammation, followed by a chronic injury stage, comprising hepatitis and liver fibrosis. Macroscopically, marked dilation of the gallbladder and formation of bilioma are found, associated with weight loss and a mortality rate of 10% in the first week due to bile leakage and rupture of the gallbladder [16]. Pathophysiologically, BDL interferes with glandular liver function and hepatobiliary transport, which comprises its detoxification and secretion function including bile formation. Obstruction of the bile duct leads to afflux of newly generated bile fluid. The main components of the bile, bile acids and phospholipids, induce toxicity and damage towards hepatocytes and cholangiocytes, therewith initiating the disease process. Rapidly after BDL, mice develop obstructive jaundice and cholestasis, as displayed by markedly elevated serum transaminases and bilirubin levels (Figure 1), macroscopically evident from yellow ears and urine. Within the first 30 h, there is a massive release of liver enzymes, like ALT and GLDH, reflecting the initial hepatocyte damage as initial pathophysiological event in the process of BDL-induced liver fibrosis (Figures 1A, B).

The plasma levels of diagnostic liver enzymes remained elevated over several days, but then dropped to reach after 14 days values that were slightly above those of sham operated mice. Concomitantly, liver detoxification capacity was deteriorated, as indicated by the rise of total bilirubin, a classical plasma marker of cholestasis (Figure 1C). Notably the plasma level of albumin, an important parameter for the evaluation of liver functionality remained relatively constant over the time course of 14 d (Figure 1D). The systemic blood cell count showed constant levels of erythrocytes and platelets up to day 5 (Table 1). In contrast, leukocytes decreased by 50% during the first two days, reflecting intrahepatic cell entrapment, and recovered to values of sham operated animals within the subsequent observation period (Table 1). During progression of fibrosis, red blood cells, hemoglobin and hematocrit slightly decreased.

In consequence of intrahepatic toxic bile accumulation, progressive development of confluent bile lakes is a hallmark of cholestasis. Histological quantification of bile infarcts, defined as clusters of injured hepatocytes, revealed a steady rise of infarct areas until day 14 after BDL (Figure 2A). The typical

appearance of liver tissue at representative time points after BDL using H&E staining is depicted in Figure 2B. Further histopathological changes of the livers after BDL included enlargement of portal tracts, accompanied by dilation of bile canaliculi and proliferation of BECs and oval cells (Figure 3A), resulting in formation of artificial bile ductules (Figure 2C), a cellular response termed 'ductular reaction' [7, 17]. However, recent data from lineage tracing experiments indicated that these cells, however, do not contribute to the population of ECM producing/fibrogenic cells, which in the BDL model is largely consisting of HSCs [18].

The inflammatory response resulting from chronic hepatocyte injury resulted in the accumulation of immune cells in the liver, among others, T cells, macrophages and dendritic cells, which were mainly found within and around bile infarct areas (Figure 2C, asterisk) [19]. The inflammatory and fibrogenic response is initiated by resident liver cells, primarily liver macrophages (Kupffer cells) and activated hepatic stellate cells (HSCs) secreting a wide range of cytokines and chemokines which determine the quality and quantity of inflammatory and consequently fibrotic responses [20, 21]. Upon parenchymal damage, quiescent HSCs undergo phenotypical changes to myofibroblasts (MFBs). The most prominent role of MFBs is extracellular matrix (ECM) production and reorganization, as reflected by, among others, increased synthesis of α -SMA, type I collagen and TIMPs. A marked increase of the number of α -SMA- and S100a4-positive cells, as measured by immunohistochemistry, was observed after BDL (Figures 3B, C). Migration of MFBs to the site of injury and their contractility contribute to liver scarring. This is accompanied by parenchymal cell proliferation, which began to rise at day 2 as regenerative response but decreased at day 14 (Figure 3D). With a slightly faster response as compared to HSCs, starting at 30 h upon BDL, Kupffer cells started to proliferate (Figure 3E). The overall hepatic proliferative response as analyzed by immunohistochemistry was confirmed by elevated mRNA expression of Ki67 (Figure 4A).

Between days 5 and 14 after BDL, periportal alterations were associated with fibrotic changes. As demonstrated by Sirius red staining, extensive fibrosis, characterized by a several-fold increase of collagen deposition (Figure 4B), including bridging, was observed 5 days after BDL (Figure 4C). We further stained for connective tissue growth factor (CTGF), a prominent fibrogenic cytokine and enhancer of TGF- β effects [22]. Appearance of CTGF-positive cells started as early as 12 h upon BDL and their count increased continuously (Figure 3F).

Time phases of disease progression after BDL

For the identification of distinct phases in disease progression induced by BDL time-resolved transcriptomics profiles of three different gene panels related to hepatocyte metabolism, fibrogenesis, and inflammation using the Fluidigm platform were measured in parallel to the biochemical and histochemical markers (Figure 5, Supplement 2 explorative data analysis). The selection of representative genes for ADME- (absorption, distribution, metabolism, and excretion) (Figure 5A) fibrogenesis- (Figure 5B), and inflammation-related genes (Figure 5C) was hereby made based on state-of the art knowledge.

In a first step, a one-way analysis of variance (ANOVA) was applied to reduce the complete data set of biochemical, histochemical and transcript data consisting of 153 factors to a subset of those factors showing significant ($p_{adj} < 0.05$) changes during the time course of disease progression (Supplement 2, dimension reduction). This resulted in a subset of 90 factors comprising two biochemical markers (bilirubin, GLDH), five histochemical markers (BEC, NHC, Kupffer cells, Sirius red), three immunostainings by antibodies (CTGF, α -SMA, S100a4) and 80 genes (14/47 ADME, 22/46 fibrosis, 44/47 inflammation panel). Many of the ADME genes and fibrosis genes were filtered out, whereas almost all of the inflammation panel was retained. The top significant factors were: Cyp1a2, serum bilirubin, Il10rb, Tgfb1, Ccl2, Cd86, Ccr2, and Mrc1. Within this subset, a bivariate time-dependent correlation analysis (Methods and Supplement 2 correlation analysis) was performed for all pairs of factors to identify those factors displaying similar temporal profiles (Figure 7), with the top correlations for biochemical, histological and immunostaining factors depicted in Figure 8. Based on the obtained correlation matrix a hierarchical cluster analysis was applied resulting in 6 different clusters comprising between 2 - 61 factors with distinct time courses attaining their maximum at different time points (see Figure 9). RNA levels of commonly regulated genes form highly correlated clusters, e.g., (i) cytokines and growth factors, such as Il6 and Tgfb (lower left, blue), (ii) fibrosis-related genes, such as Coll1a1 and Timp1 (bottom right, cyan), (iii) ADME-related genes, such as isoforms of Cyp24a1 and Nr0b2 (red, top right). The found clusters comprise both 'classical' biochemical and histochemical factors and genes characteristic for a specific phase of disease progression.

In what follows, the biological significance of the correlations between different types of factors are discussed (Figure 8), followed by analysis of the time course clusters (Figure 9) and predictive models based on these clusters (Figure 10).

Correlations between transcripts and non-transcript factors

The time course of each of the ‘classical’ factors contained in the ANOVA-subset can be correlated with the expression time course of at least one gene (Figure 8A). Only for GLDH and Sirius Red the correlations are weak. Notably, all top correlations to genes come either from cluster 4 or cluster 1. Bilirubin, bile infarcts and the immunostainings (α -SMA, CTGF and S100a4) all have high correlations among each other, so do the BrdU positive BEC, Kupffer and NHC measurements (Figure 8B). GLDH and Sirius red do not show high correlation with any other classical factor. In the following the top correlated factors with the classical biochemical, histological and immunostaining factors are discussed in the context of different aspects of the disease process (Figure 8C). Main points are summarized in Figure 11 bringing the different aspects, phases and markers together.

Initial response

Immediately after BDL, there is a massive release of liver enzymes up to day 5, followed by a drop down to almost values of sham-operated livers (Figures 1A, B). GLDH like ALT shows a strong initial increase, but in contrast to the latter increases further up to 18h, before it decreases gradually. GLDH is highly correlated with the other members of cluster 3 (Figure 8C, Figure 9C) the initially up-regulated transcripts Fn1 (fibronectin, Figure 6G) and Sult1a1 (Sulfotransferase 1A1).

Initial molecular events with strong transcript peaks are observed at 6 h for the members of cluster 2, the transcriptional regulator Nr0b2 (small heterodimer partner, SHP, Figure 6D) and Cyp24a1 (mitochondrial 1,25-dihydroxyvitamin D3 24-hydroxylase, see Figure 6B). Nr0b2 is suggested to be associated to cirrhosis and hepatic tumors [23], supported by the fact that cholestatic liver fibrosis induced by BDL was increased in SHP-/- mice [24]. Additionally up-regulated transcripts at 6h (Supplement 2, t-test for initial phase) are Tnfrsf1a (Figure 6L), Il6st (Interleukin-6 receptor subunit beta), Osmr, Cd14, Cxcl1/2, Timp1 and Hmox1 (heme oxygenase), in line with reported marked increase in heme oxygenase activity following BDL in rats [25]. Marked initial down-regulation was observed among others for Cdh2 and Pde4a and the main enzyme

of bile acid synthesis Cyp7a1 (cholesterol-7- α -hydroxylase), which can be seen as a fast and straightforward response to cholestasis. As underlying mechanism, activation of the JNK/c-Jun pathway has been reported [26].

Macroscopic organ damage

Necroinflammation was caused by the BDL-induced intrahepatic toxic bile accumulation with individual liver cell death and progressive development of confluent bile infarct areas documented by H&E staining in Figure 2B. The total area of infarcts increases steadily with relatively high variance (Figure 2A). Bilirubin (Figure 1C) showed the highest correlation with bile infarcts followed by the immunostainings CTGF (Figure 3F) and α -SMA (Figure 3B), whereas Gsta2 (Glutathione S-transferase A2), Gstm1 (glutathione-S-transferase mu 1, Figure 6C) and Timp1 (Metalloproteinase inhibitor 1) showed the highest positive correlation with the bile infarcts. Timp1 is a metalloproteinase inhibitor that functions by forming one to one complexes with target metalloproteinases, such as collagenases. In contrast, Cyp1a2 (Cytochrome P450 1A2, Figure 6A) and Cyp2e1 (Cytochrome P450 2E1) were highly anti-correlated to the bile infarct area, with Cyp1a2 decreasing continuously after BDL in line with observations in rats [27].

Loss of liver function

Liver function after BDL was measured using albumin (Figure 1D) and bilirubin (Figure 1C) levels. Surprisingly, albumin synthesis is maintained relatively constant over the observation period of 14 days, and was consequently filtered out via ANOVA. Bilirubin on the other hand increased continuously after BDL. The highest positive correlation with serum bilirubin levels are observed for the transcripts Timp1, Cd14. Ccl2 (chemokine C-C motif ligand 2), a soluble biomarker for hepatic fibrosis in NAFLD [28], and Ccl3 (Figure 8C). Notably, a very high negative correlation was found between bilirubin and Slc10a1 (Sodium/bile acid cotransporter), the hepatic sodium bile acid uptake system. The decrease in Slc10a1, encoding the Na⁺-taurocholate co-transporting polypeptide, which transports bile acids, has been shown to protect hepatocytes from cholestasis-induced injury [29].

Hepatic cell proliferative response

During disease progression various hepatic cell types start proliferating, underlined by co-immunostaining with BrdU and specific markers for hepatocytes (Figure 3D), Kupffer cells (KC, Figure 3E), and BECs (Figure 3A), and indirectly reflected by the marked up-regulation of Ki67 mRNA (Figure 4A). As compensatory activity of the liver to restore the damaged parenchyma, hepatocytes proliferate, monitored by the parameter BrdU-positive hepatocytes, with a marked increase between 30h and 2d (Figure 3D). The observed time course is very similar to the increase in KC, BECs and NHCs, resulting in high correlation within this group (Figure 8B). In addition, S100a4 positive cells are highly correlated with the proliferation time course of these three cell types.

Kupffer cells (liver macrophages) infiltrate the liver tissue, are activated, and proliferate starting at 30 h in our time course to reach a maximum at day 2 and to decrease again thereafter. They have the highest correlation to the transcript Mki67 (antigen Ki-67), a known proliferation marker, followed by the transcripts Birc5 (Baculoviral IAP repeat-containing protein 5, survivin) and Notch1, a transmembrane receptor involved in developmental processes (Figure 8C). Yes-associated protein has been reported to regulate the hepatic response after bile duct ligation via modulation of survivin [30], underlining the role of survivin in hepatic tissue restructuring.

Biliary epithelial cells (BEC) also start proliferating after BDL, with the main increase between days 2 and 5, with a minor increase already observed after 30 h. Interestingly, the highest correlations to transcripts are all negative, namely Cyp2c37 (Cytochrome P450 2C37), Slc10a1, Cyp2e1, and Cyp2c29. Among the top correlated transcripts to the classical factors (Figure 8A) Cyp2c37 and Slc10a1 are two very interesting candidates being, with Cyp1a2 and Ppara, the only factors from time course cluster 1 and having high negative correlations.

Increase in fibrogenic cells

In accordance to the proliferative activity of HSCs, immunohistochemical analyses demonstrate the gradual rise in CTGF- (Figure 3F), α -SMA- (Figure 3B) and S100a4-expressing cells (Figure 3C), reflecting activated HSC, activated KC and trans-differentiated hepatocytes. CTGF and α -SMA are highly correlated

1 to each other (respective top correlation Figure 8C), with S100a4 having strong correlation to both of them
2 (Figure 8B).

3 CTGF is a highly pro-fibrogenic protein expressed by HSCs, BECs and hepatocytes [31, 32] and
4 mediates extracellular matrix inducing properties. Levels of CTGF have been reported significantly up-
5 regulated in experimental liver fibrogenesis and human chronic liver disease patients of various etiologies
6 [33, 34]. CTGF-positive cell number is the best candidate to monitor the disease progress among the
7 selected biochemical, histological and immunostaining parameters, showing a steady increase with
8 comparatively little variance (among top ANOVA results, $p_{adj}=7.9E-10$). This is consistent with data from
9 other studies, which observed a correlation of increased CTGF levels with histological fibrosis stages [35,
10 36] suggesting CTGF as a valuable diagnostic target, since it can be measured in patients' blood and have
11 a potential application in the follow-up of patients suffering from chronic liver diseases [37]. The highest
12 transcript correlation with CTGF shows *Tgfb2* (cytokine TGF- β 2, Figures 6I) followed by *Pdgfb* (platelet-
13 derived growth factor subunit B). TGF- β is the major stimulus for CTGF expression in hepatocytes [38], and
14 elevated levels of *Tgfb2* were reported for bile duct ligated rats [39]. *Pdgfb* has been reported up-regulated
15 in liver inflammation and fibrosis [40]. Additionally, there are considerable correlations to several other
16 genes such as *Tgfb2*, encoding the transforming growth factor β receptor 2 [41], *Cd14*, *Cxcl5*, *Ccr2*, and
17 *Timp1*.

18 α -SMA-positive cells, a marker of HSC activation, increase steadily during disease progression (Figure
19 3B), highly correlated to CTGF. Consequently, the top transcript correlations are the very similar: *Tgfb2*,
20 *Cxcl4*, *Timp1*, *Tnc*, and *Pdgfb*. Notably, α -SMA staining, but also CTGT and S100a, are highly negatively
21 correlated to *Cyp1a2* (cytochrome P450 1A2, Figure 6A), known to be lowered in liver cirrhosis mediated
22 by AhR [42], and to *Ppara* (Figure 8A).

23 S100a4-positive cells show a steady rise until day 2, after which it stays elevated (Figure 3C). S100a4-
24 positive cells are also a good marker for disease progression with a similar time course than CTGF (Figure
25 3B), but larger variation than CTGF from 18h on. Many transcripts are highly correlated with S100a4, e.g.
26 *Pdgfb*, *Birc5*, *Tgfb2* or *Notch1*.

Fibrosis

The progression of fibrogenesis was histomorphologically characterized by excessive deposition of extracellular matrix visible by Sirius red staining of liver slices (Figure 4B, C). Surprisingly, Sirius red did not have high correlations to other factors (Figure 8ABC), mainly due to the high variability in measurements from 0 h to 2 d and was in this study not a very reliable predictor for fibrogenesis. After day 2 a strong increase in Sirius red was observed. Both Col1a1 (fibrillar collagen 1 α 1, Figure 6E) and Col3a1 (fibrillar collagen 3 α 1, Figure 6F) transcripts, which predominantly exist in fibrotic livers, showed up-regulation beginning 30h after BDL and a continuous increase with severity of liver fibrosis up to 14 days. Among the peptide mediators, Tgfb1 and Tgfb2 (Tgf- β isoforms 1 and 2) expression was increasing after 2d, confirming its postulated role as fibrogenic master cytokine [43]. Tgfb, encoding the cytokine TGF- β is well known to correspond with the fibrotic process in a positive feedback loop [44]. Associated with induction of fibrogenesis-related genes (Figure 5B), which particularly are representative for HSC activation, the dynamics of the inflammation gene signature (Figure 5C) nicely matches with the increase in the number of proliferating Kupffer cells (F4-80/BrdU staining values) observed from day 2 onwards (Figure 3E). Very low expression levels were present immediately after BDL, except for the chemokines Cxcl1 und 2. Starting at time points between 2 and 5 d after BDL, most cytokines and chemokines in the list were strongly upregulated until day 14. During the perpetuation phase (18h-2d), paracrine and autocrine cytokines amplify hepatic inflammation and HSC activation resulting in continued ECM remodeling, being characterized by the enhanced mRNA expression of both fibrillar collagen1 α 1 and 3 α 1 (Figures 6E and F).

Markers of disease progression

The main focus of this study was to detect factors and factor combinations which characterize particular stages of the disease process best. Here, we discuss in more detail the biological significance of the 6 time course clusters.

Cluster 1 (Figure 9A) decreases continuously over time, with no classical factor included in c1. Most of the members (8/11) are from the ADME chip, with exception of Cyp2e1, Rarres1 and Egfr coming from the fibrosis chip. All members of cluster have very high significance in the ANOVA, with exception of Rarres1 ($p_{adj} = 0.036$). The ADME genes of c1 decrease during the late initial, perpetuation and progression phase.

Top correlations with the cluster mean time course are in order of decreasing correlation Cyp2c37, Cyp2e1, Cyp2e29, Ugt1a1, Cyp1a2 (Figure 6A), Rarres1 and Slc10a1 containing many enzymes of the cytochrome P450 system. The down-regulation of Ugt1a1 (UDP-glucuronosyl-transferase 1A), the main enzyme for conjugation of bilirubin, and Slc10a1, encoding the Na⁺-taurocholate co-transporting polypeptide, which transports bile acids, are protective against the increased concentration of conjugated bilirubin in hepatocytes after BDL.

Cluster 2 (Figure 9B) consists of strong transcript peaks at 6h by the transcriptional regulator Nr0b2 (SHP, Figure 6D) and Cyp24a1 (Figure 6B), both probes of the ADME chip. Nr0b2 ($p_{adj} = 1.30E^{-7}$) and Cyp24a1 ($p_{adj} = 9.88E^{-3}$) both have very high ANOVA significance. Other transcripts also show an increase in the initial phase at 6h like for instance the members of cluster 3, but none of them to baseline during disease progression at 18h up to 14d. This makes Nr0b2 and Cyp24a1 interesting targets for the detection of very acute cholestasis.

Cluster 3 (Figure 9C) is characterized by an increase in the initial phase up to 18h with subsequent decrease during disease progression up to 14d. The cluster consists of the biochemical factor GLDH, the fibrosis transcript Fn1 (Fibronectin, Figure 6G), and the ADME gene Sulf1a1 (Sulfotransferase 1A1), making this an interesting cluster combining various factor types.

Cluster 4 (Figure 4) shows continuous increase starting in the initial phase, lasting throughout disease progression up to 14d. Consequently, members of the cluster are good candidates for continuous disease progression (fibrosis), among the top candidates the bilirubin, CTGF and α -SMA. Cluster 4 is the largest cluster containing 61/90 significant factors of the ANOVA. Notably, most of the classical markers are contained in cluster 4: bilirubin, CTGF α -SMA, S100a4, bile infarcts and BEC. The transcripts in c4 come all either the inflammation chip or the fibrosis chip, with exception of Gsta2 and Gstm1 (ADME chip).

The top transcript correlation with the cluster mean shows Timp1, followed by Ccr2, and Tgfbr2 with a large number of transcripts showing high correlation to the cluster mean.

Cluster 5 (Figure E) has a highly similar time course than c4, but only increasing from baseline after around 30h compared to the continuous increase of c4 starting already in the initial phase. Cluster 5 contains the classical markers NHC, Kupffer cells and Sirius red. Top correlating transcripts are Gdl2 and Cyp7a1,

1 which shows a strong increase starting around 30h, despite the strong down-regulation in the initial phase
2 (see above). Notably, the interleukins Il2, Il17a (interleukin-17A, Figure 6H) and Il28b (interleukin 28 β ,
3 Figure 6K) are members of c5, which are secreted proteins, likely to be detectable in blood, thus
4 representing a candidate diagnostic marker. Il17a (interleukin-17A, Figure 6H), plays a pivotal role in
5 cholestatic liver fibrosis in mice by activation of both the KCs and HSCs [45].

6 Cluster 6 (Figure 9F) is characterized by an initial decrease followed by an increase in the late initial
7 phase at 12h up to 2d with subsequent decrease during disease progression at 5d and 14d. Both cluster
8 members Cdh2 (cadherin 2) and Bad1 are part of the fibrosis chip. Cluster 6 shows a similar up and down
9 than c3, but the increase starts later, the decrease starts later and the maximum transcript values are
10 consequently shifted to a later time point.

11 The large majority of histologic parameters, cell count observations, as well as most genes related to
12 inflammation and fibrogenesis increased with disease progression either in a continuous manner starting in
13 the perpetuation phase after ~1 day or at the latest beginning at day 5 (cluster 4 and 5). Based on the strong
14 increase between day 2 and 5 in cluster 5 (but also in cluster 4), we conclude that transition from day 2 to
15 day 5 can be easily monitored, also in clinical practice, as serum bilirubin is among the top correlations of
16 cluster 4, and the interleukins among cluster 5 (Il2, Il17a, Il28b, Figure 6H and K) as well as the growth
17 factors in cluster 4 (Pdgfb, Tgfb1, Tgfb2, Figure 6I, Hgf) and contributors to the extracellular matrix among
18 cluster 4 (Sparc, Col3a1, Col1a1, Figures 6F and E).

19 Cluster 4 can be seen representative for disease progression due to the continuous increase starting
20 already in the initial phase making the count of CTGF-positive cells the most reliable overall measure for
21 disease progression at the histological level, bilirubin at biochemical level, and metalloproteinase inhibitor 1
22 (Timp1) on transcript level.

23 Interestingly no histological, biochemical or immunostaining based factors were found in the clusters
24 c1, c2 and c6. The transcripts in these clusters provide unique time course information which cannot be
25 captured with the subset of analyzed histological markers, thereby providing crucial information for the initial
26 and perpetuation phase, not attainable via c3, c4 and c5.

Decision trees for disease progression

We next asked the question, which of the analyzed factors are best suited as markers for particular stages of the disease process. To this end we constructed a decision tree based on the time courses of the normalized clusters (Figure 9). This regression tree allows the prediction of a specific time interval of the disease process based on the combination of time courses of factors from the clusters (Figure 10). The algorithm used for the construction of the decision tree aims at avoiding over-fitting of the data by balancing the number of knots against robustness of the classification tested by cross-validation. Consequently, the decision tree assigns a pattern of cluster to time intervals of disease progression rather than the discrete experimental time points, resulting in 6 time classes. Interestingly, mainly time points in the late initial phase and perpetuation phase (12h, 18h, 30h, 48h) were merged into time classes, whereas the control (0h), early initial phase (6h), and progression phase (5d, 14d) remained almost unchanged. The robustness of the predicted time classes and prediction performance were tested with a leave-one-out approach, resulting in reproducible time classes and good prediction performance on the left out test data (Supplement 2, decision tree).

The resulting regression tree exploits time course information from cluster c1, c3, c4, and c5, whereas clusters c2 and c6 are not used. Cluster 4 was found to possess the strongest discriminating power performing the important split between early phase after BLD (classes 0h, 6h and 14h with range 0h-21.8h) and the later perpetuation and progression phase (classes 34h, 6d, 14d, with range 21.8h-14d): If the mean value of the factors within this cluster is smaller than -0.12 at a measured time point this time point of BLD induced disease is classified as being not larger than 21.8 hours. The more detailed assignment of the respective time interval requires to interrogate additional clusters: The value of c3 decides between control and initial phase with subsequent splitting based on c1 in early and late initial phase. Analogue, the value of c5 decides between perpetuation and progression phase with subsequent splitting in early and late progression phase based on c4. Note that the values of cluster 4 appear twice in the decision tree, owing to the monotonous increase after BDL.

The predictive performance of the regression tree on the mean cluster data is depicted in Figure 10B (blue), providing the information which experimental time points were classified in which classes. All

1 samples of control, 6h, and 14d are assigned to their respective time classes 0h, 6h and 14d, whereas
2 neighboring time points are combined in classes 14h, 34h and 6d. In addition, we evaluated the decision
3 tree based on a subsets of factors from the time course clusters using either a single factor or two factors
4 randomly chosen from each cluster and their values were used for prediction (Figure 10B, single and double
5 factors). Even with only one factor selected from c1, c3, c4, and c5 surprisingly consistent classifications
6 were achieved. As expected, the quality of predictions was improved by increasing the number of factors.
7 The control, early initial phase and late progression phase can be predicted with high accuracy, whereas
8 the intermediate phase shows more misclassifications in case of the single and double factors.

9 The best performing decision tree based on a single transcript from each cluster (Figure 10A) splits on
10 Col1a1 (Collagen alpha-1(I) chain, c4), Fn1 (Fibronectin, c3), Cyp1a2 (Cytochrome P450 1A2, c1), and
11 Il17a (Interleukin-17A, c5), all important factors involved in disease progression (discussed above). The
12 best performing decision tree based on all factors, i.e. histological, biochemical, immunostaining factors and
13 transcripts, resulted in a highly similar tree, with the single change of using S100a4 instead of Col1a1 for c4
14 splitting, and providing GLDH as alternative factor to Fn1 to perform the c3 split. By selecting a subset of
15 good performing transcripts very good prediction performance already on a small subset of factors can be
16 achieved (Figure 10B red). Notably, our approach allows the prediction of progression after BDL from an
17 arbitrary subset of measured parameters.

18 Individual variability

19 We observed a large variability of many analyzed factors, when comparing individual mice of the same
20 time points. For example at day 5, the infarct area varies between 0.9% and 12% (Figure 2A) and the
21 collagen deposition area measured by Sirius red between 0.8%-5.9% (Figure 4B). Similar high variability
22 can be observed in the proliferative response (BEC Figure 3A, Kupffer cells Figure 3E) or expression of
23 collagen (Col1a1 Figure 6E, Col3a1 Figure 6F), to name a few. These large variation in parameters during
24 the perpetuation and especially progression phase are an intriguing finding, considering the homogeneity
25 of the experimental system (see also heatmap of time courses in Supplement 2). A possible explanation
26 could be a different individual pace of disease progression due to variations in susceptibility to the damage

1 induced by BDL, for instance due to different genetic background. As a consequence heterogenous time
2 courses develop, with highly affected mice showing strong signs of fibrosis earlier.

3 Another hypothesis are variable routes of disease progression: One route is characterized by a strong
4 increase in necrotic tissue and a weaker activation of HSC and lower expression change of inflammation
5 factors. The other route is characterized by strong activation of fibrogenesis factors, and finally macroscopic
6 collagen tissue. Both routes are similarly connected to loss of liver function, however, the histopathological
7 display is quite different: the former contains a large amount of necrotic tissue while the latter contains large
8 fibrotic tissue. Such alternative disease routes could be of far-reaching importance for an individualized
9 therapy, as obviously the medical interventions avoiding necrosis differ from interventions reducing
10 overshooting fibrosis. The design of the study, which included the sacrifice of mice after a specific time, did
11 not allow to answer whether the alternate developments, shown macroscopically in later time points, can
12 also be observed in earlier time points.

13 Conclusion

14 The time-resolved analysis of a wide range of parameters in bile duct ligated mice has shown that many
15 factors share the pattern of an increase throughout disease progression (Figure 9). Particularly, pronounced
16 changes were observed during the transition from perpetuation to progression phase 2 to 5 days after BDL,
17 characterized by strong increase of parameters, such as Il17A, Il2, Il28b or Il13. This information has strong
18 clinical relevance, as it indicates a switching point, and human homologs of the respective interleukins are
19 top candidates to be used as clinical markers.

20 Our experiments also found previously unknown molecular events which are probably elements of a
21 transcriptional program. For instance, there are strong transcript peaks for SHP (Nr0b2) at 6h, which now
22 need a focused analysis to delineate the chain of molecular interactions causing it and the functional
23 consequences for disease onset and progression.

24 Based on time course correlation analysis we found a distinct number of time course patterns following
25 BDL. Based on these results, we propose a decision tree from a subset of parameters, such as in Figure 10,
26 as a promising tool for assessment of disease progression. Therefore, as next step, suitability of the
27 parameters selected from the mouse model need to be confirmed for human patients as in [46].

Many of the measured parameters display a large variability, which maybe the major reason why translation of a set of such parameters into diagnostic approaches has not proven sufficient robust for predictions in human patients with chronic liver diseases. The here suggested approach of pooling information from factors falling in the same time course classes could be a possible solution for more robust predictors.

Taken together, the detailed time-resolved transcriptional profiling of liver homogenates following BDL revealed a coordinated response, resulting in an up-regulation of an inflammatory response along with activation of metabolically active genes, which can be explained by physiological recovery and adaptation of the mice to the bile acid exposure.

Availability of supporting data

The data sets supporting the results of this article are included within the article and its additional files.

Dataset S1. Measurements in bile duct ligated mice.

Dataset S2. Statistical analysis.

List of Abbreviations

CTGF: connective tissue growth factor; BDL: bile duct ligation; SHP: small heterodimer partner; CLD: chronic liver disease; HSC: hepatic stellate cells; KC: Kupffer cells; ADME: absorption, distribution, metabolism, and excretion; ALT: alanine aminotransferase; GLDH: glutamate dehydrogenase; EDTA: ethylenediaminetetraacetic acid; H&E: hematoxylin and eosin;

Competing Interests

The authors declare that they have no conflicts of interest.

Author Contributions

Conceived and designed the experiments: KA AH MT IM BV SD. Performed the experiments: KA MT IM HW. Analyzed the data: KA MK AH MT IM ME HW HH UMZ BV SD. Wrote the paper: KA MK AH MT ME HW HH UMZ BV SD. All authors read and approved the final manuscript.

Acknowledgments

The authors kindly thank Dorothea Frenz, Berit Blendow, Maren Nerowski, Doris Butzlaff and Eva Lorbeer-Rehfeldt (Institute for Experimental Surgery, University of Rostock) and Igor Liebermann (Institute of Clinical Pharmacology, Stuttgart) for their excellent technical assistance. The authors thank Pablo Jaskowiak for helpful discussions. The study was supported by grants from the German Research Foundation to Kerstin Abshagen (AB 453/1-1), by the BMBF program “Virtual Liver” (Grants 0315755, 0315764) and by the Robert Bosch Foundation, Stuttgart, Germany. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Wang H, Vohra BPS, Zhang Y, Heuckeroth RO: **Transcriptional profiling after bile duct ligation identifies PAI-1 as a contributor to cholestatic injury in mice.** *Hepatology* 2005, **42**(5):1099-1108.
2. Woolbright BL, Antoine DJ, Jenkins RE, Bajt ML, Park BK, Jaeschke H: **Plasma biomarkers of liver injury and inflammation demonstrate a lack of apoptosis during obstructive cholestasis in mice.** *Toxicol Appl Pharmacol* 2013, **273**(3):524-531.
3. Lindblad L, Lundholm K, Schersten T: **Bile acid concentrations in systemic and portal serum in presumably normal man and in cholestatic and cirrhotic conditions.** *Scand J Gastroenterol* 1977, **12**(4):395-400.
4. Trauner M, Jansen PL: **Molecular pathogenesis of cholestasis:** Springer Science & Business Media; 2004.
5. Nguyen KD, Sundaram V, Ayoub WS: **Atypical causes of cholestasis.** *World J Gastroenterol* 2014, **20**(28):9418-9426.
6. Tag CG, Sauer-Lehnen S, Weiskirchen S, Borkham-Kamphorst E, Tolba RH, Tacke F, Weiskirchen R: **Bile duct ligation in mice: induction of inflammatory liver injury and fibrosis by obstructive cholestasis.** *J Vis Exp* 2015(96).
7. Georgiev P, Jochum W, Heinrich S, Jang JH, Nocito A, Dahm F, Clavien P-A: **Characterization of time-related changes after experimental bile duct ligation.** *Br J Surg* 2008, **95**(5):646-656.
8. Huss S, Schmitz J, Goltz D, Fischer H-P, Büttner R, Weiskirchen R: **Development and evaluation of an open source Delphi-based software for morphometric quantification of liver fibrosis.** *Fibrogenesis Tissue Repair* 2010, **3**(1):10.

9. Tanaka A, Tsuneyama K, Mikami M, Uegaki S, Aiso M, Takikawa H: **Gene expression profiling in whole liver of bile duct ligated rats: VEGF-A expression is up-regulated in hepatocytes adjacent to the portal tracts.** *J Gastroenterol Hepatol* 2007, **22**(11):1993-2000.
10. Spurgeon SL, Jones RC, Ramakrishnan R: **High throughput gene expression measurement with real time PCR in a microfluidic dynamic array.** *PLoS One* 2008, **3**(2):e1662.
11. Holm S: **A simple sequentially rejective multiple test procedure.** *Scandinavian journal of statistics* 1979.
12. Son YS, Baek J: **A modified correlation coefficient based similarity measure for clustering time-course gene expression data.** *Pattern Recognition Letters* 2008, **29**(3):232-242.
13. Jaskowiak PA, Campello RJGB, Costa IG: **On the selection of appropriate distances for gene expression data clustering.** *BMC Bioinformatics* 2014, **15** Suppl 2:S2.
14. Jaskowiak PA, Campello RJGB, Costa IG: **Proximity measures for clustering gene expression microarray data: a validation methodology and a comparative analysis.** *IEEE/ACM Trans Comput Biol Bioinform* 2013, **10**(4):845-857.
15. Breiman L, Friedman J, Stone CJ, Olshen RA: **Classification and regression trees:** CRC press; 1984.
16. Geerts AM, Vanheule E, Praet M, Van Vlierberghe H, De Vos M, Colle I: **Comparison of three research models of portal hypertension in mice: macroscopic, histological and portal pressure evaluation.** *Int J Exp Pathol* 2008, **89**(4):251-263.
17. Kisseleva T, Brenner DA: **Anti-fibrogenic strategies and the regression of fibrosis.** *Best Pract Res Clin Gastroenterol* 2011, **25**(2):305-317.
18. Mederacke I, Hsu CC, Troeger JS, Huebener P, Mu X, Dapito DH, Pradere J-P, Schwabe RF: **Fate tracing reveals hepatic stellate cells as dominant contributors to liver fibrosis independent of its aetiology.** *Nat Commun* 2013, **4**:2823.
19. Holt AP, Salmon M, Buckley CD, Adams DH: **Immune interactions in hepatic fibrosis.** *Clin Liver Dis* 2008, **12**(4):861-882, x.
20. Wasmuth HE, Lammert F, Zaldivar MM, Weiskirchen R, Hellerbrand C, Scholten D, Berres M-L, Zimmermann H, Streetz KL, Tacke F *et al*: **Antifibrotic effects of CXCL9 and its receptor CXCR3 in livers of mice and humans.** *Gastroenterology* 2009, **137**(1):309-319, 319.e301-303.
21. Heinrichs D, Berres M-L, Nellen A, Fischer P, Scholten D, Trautwein C, Wasmuth HE, Sahin H: **The chemokine CCL3 promotes experimental liver fibrosis in mice.** *PLoS One* 2013, **8**(6):e66106.
22. Leask A, Abraham DJ: **All in the CCN family: essential extracellular signaling modulators emerge from the bunker.** *J Cell Sci* 2006, **119**(Pt 23):4803-4810.
23. Smalling RL, Delker DA, Zhang Y, Nieto N, McGuinness MS, Liu S, Friedman SL, Hagedorn CH, Wang L: **Genome-wide transcriptome analysis identifies novel gene signatures implicated in human chronic liver disease.** *Am J Physiol Gastrointest Liver Physiol* 2013, **305**(5):G364-G374.
24. Zhang Y, Xu N, Xu J, Kong B, Copple B, Guo GL, Wang L: **E2F1 is a novel fibrogenic gene that regulates cholestatic liver fibrosis through the Egr-1/SHP/EID1 network.** *Hepatology* 2014, **60**(3):919-930.
25. Schacter BA, Joseph E, Firneisz G: **Effect of cholestasis produced by bile duct ligation on hepatic heme and hemoprotein metabolism in rats.** *Gastroenterology* 1983, **84**(2):227-235.
26. Gupta S, Stravitz RT, Dent P, Hylemon PB: **Down-regulation of cholesterol 7alpha-hydroxylase (CYP7A1) gene expression by bile acids in primary rat hepatocytes is mediated by the c-Jun N-terminal kinase pathway.** *J Biol Chem* 2001, **276**(19):15816-15822.
27. Fukushima S, Okuno H, Shibatani N, Nakahashi Y, Seki T, Okazaki K: **Effect of biliary obstruction and internal biliary drainage on hepatic cytochrome P450 isozymes in rats.** *World J Gastroenterol* 2008, **14**(16):2556-2560.
28. Page S, Birerdinc A, Estep M, Stepanova M, Afendy A, Petricoin E, Younossi Z, Chandhoke V, Baranova A: **Knowledge-based identification of soluble biomarkers: hepatic fibrosis in NAFLD as an example.** *PLoS One* 2013, **8**(2):e56009.
29. Eipel C, Menschikow E, Sigal M, Kuhla A, Abshagen K, Vollmar B: **Hepatoprotection in bile duct ligated mice mediated by darbepoetin- α is not caused by changes in hepatobiliary transporter expression.** *Int J Clin Exp Pathol* 2013, **6**(1):80-90.

- 1 30. Bai H, Zhang N, Xu Y, Chen Q, Khan M, Potter JJ, Nayar SK, Cornish T, Alpini G, Bronk S *et al*: **Yes-associated protein regulates the hepatic response after bile duct ligation**. *Hepatology* 2012, **56**(3):1097-1107.
- 2
- 3 31. Paradis V, Dargere D, Vidaud M, De Gouville AC, Huet S, Martinez V, Gauthier JM, Ba N, Sobesky R, Ratziu V *et al*: **Expression of connective tissue growth factor in experimental rat and human liver fibrosis**. *Hepatology* 1999, **30**(4):968-976.
- 4
- 5 32. Sedlacek N, Jia JD, Bauer M, Herbst H, Ruehl M, Hahn EG, Schuppan D: **Proliferating bile duct epithelial cells are a major source of connective tissue growth factor in rat biliary fibrosis**. *Am J Pathol* 2001, **158**(4):1239-1244.
- 6
- 7 33. Rachfal AW, Brigstock DR: **Connective tissue growth factor (CTGF/CCN2) in hepatic fibrosis**. *Hepatol Res* 2003, **26**(1):1-9.
- 8
- 9 34. Dendooven A, Gerritsen KG, Nguyen TQ, Kok RJ, Goldschmeding R: **Connective tissue growth factor (CTGF/CCN2) ELISA: a novel tool for monitoring fibrosis**. *Biomarkers* 2011, **16**(4):289-301.
- 10
- 11 35. Hayashi N, Kakimura T, Soma Y, Grotendorst GR, Tamaki K, Harada M, Igarashi A: **Connective tissue growth factor is directly related to liver fibrosis**. *Hepatogastroenterology* 2002, **49**(43):133-135.
- 12
- 13 36. Colak Y, Senates E, Coskunpinar E, Oltulu YM, Zemheri E, Ozturk O, Doganay L, Mesci B, Yilmaz Y, Enc FY *et al*: **Concentrations of connective tissue growth factor in patients with nonalcoholic fatty liver disease: association with liver fibrosis**. *Dis Markers* 2012, **33**(2):77-83.
- 14
- 15 37. Gressner OA, Fang M, Li H, Lu LG, Gressner AM, Gao CF: **Connective tissue growth factor (CTGF/CCN2) in serum is an indicator of fibrogenic progression and malignant transformation in patients with chronic hepatitis B infection**. *Clin Chim Acta* 2013, **421**:126-131.
- 16
- 17 38. Liu Y, Liu H, Meyer C, Li J, Nadalin S, Königsrainer A, Weng H, Dooley S, ten Dijke P: **Transforming growth factor- β (TGF- β)-mediated connective tissue growth factor (CTGF) expression in hepatic stellate cells requires Stat3 signaling activation**. *J Biol Chem* 2013, **288**(42):30708-30719.
- 18
- 19 39. Schierwagen R, Leeming DJ, Klein S, Granzow M, Nielsen MJ, Sauerbruch T, Krag A, Karsdal MA, Trebicka J: **Serum markers of the extracellular matrix remodeling reflect antifibrotic therapy in bile-duct ligated rats**. *Front Physiol* 2013, **4**:195.
- 20
- 21 40. Malizia G, Brunt EM, Peters MG, Rizzo A, Broekelmann TJ, McDonald JA: **Growth factor and procollagen type I gene expression in human liver disease**. *Gastroenterology* 1995, **108**(1):145-156.
- 22
- 23 41. Yang L, Inokuchi S, Roh YS, Song J, Loomba R, Park EJ, Seki E: **Transforming growth factor- β signaling in hepatocytes promotes hepatic fibrosis and carcinogenesis in mice with hepatocyte-specific deletion of TAK1**. *Gastroenterology* 2013, **144**(5):1042-1054.e1044.
- 24
- 25 42. Floreani M, De Martin S, Gabbia D, Barbierato M, Nassi A, Mescoli C, Orlando R, Bova S, Angeli P, Gola E *et al*: **Severe liver cirrhosis markedly reduces AhR-mediated induction of cytochrome P450 in rats by decreasing the transcription of target genes**. *PLoS One* 2013, **8**(4):e61983.
- 26
- 27 43. Gressner AM, Weiskirchen R, Breitkopf K, Dooley S: **Roles of TGF- β in hepatic fibrosis**. *Front Biosci* 2002, **7**:d793-d807.
- 28
- 29 44. Inagaki Y, Okazaki I: **Emerging insights into Transforming growth factor β Smad signal in hepatic fibrogenesis**. *Gut* 2007, **56**(2):284-292.
- 30
- 31 45. Hara M, Kono H, Furuya S, Hirayama K, Tsuchiya M, Fujii H: **Interleukin-17A plays a pivotal role in cholestatic liver fibrosis in mice**. *J Surg Res* 2013, **183**(2):574-582.
- 32
- 33 46. Gadd VL, Skoien R, Powell EE, Fagan KJ, Winterford C, Horsfall L, Irvine K, Clouston AD: **The portal inflammatory infiltrate and ductular reaction in human nonalcoholic fatty liver disease**. *Hepatology* 2014, **59**(4):1393-1405.
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52

1 Tables

2 **Table 1.** Systemic blood cell count of sham-operated mice (S) and mice which underwent BDL. Values
3 are given as means \pm SEM.

	erythrocytes	platelets	leukocytes	hemoglobin	hematocrit
S	8.4 \pm 0.1	1177 \pm 60	7.5 \pm 0.3	7.9 \pm 0.1	44.8 \pm 0.7
6h	8.1 \pm 0.1	1061 \pm 39	4.2 \pm 0.5	7.7 \pm 0.1	42.8 \pm 0.7
12h	8.2 \pm 0.1	1036 \pm 47	4.5 \pm 0.4	7.7 \pm 0.1	43.0 \pm 0.6
18h	8.7 \pm 0.3	856 \pm 110	4.1 \pm 0.2	8.1 \pm 0.3	45.6 \pm 1.9
30h	8.5 \pm 0.5	1071 \pm 100	5.8 \pm 0.9	7.9 \pm 0.5	44.9 \pm 2.8
2d	8.7 \pm 0.2	1117 \pm 65	4.7 \pm 1.3	6.5 \pm 1.7	45.9 \pm 0.9
5d	8.7 \pm 0.3	1295 \pm 107	7.6 \pm 1.2	7.8 \pm 0.3	46.5 \pm 1.6
14d	7.6 \pm 1.4	1362 \pm 58	7.4 \pm 1.1	6.6 \pm 0.2	38.4 \pm 1.3

4

Figure legends

Figure 1: Analysis of liver injury and-function. Plasma activities of alanine aminotransferase (ALT) (**A**) and glutamate dehydrogenase (GLDH) (**B**) and concentrations of plasma bilirubin (**C**) and albumin (**D**) at multiple time points after BDL. Values are given in means \pm SEM of five independent experiments per time point.

Figure 2: Quantification of bile infarcts in H&E-stained liver sections at multiple time points after BDL (A). Values are given in means \pm SEM of five independent experiments per time point. Representative H&E stainings of paraffin-embedded liver sections for each time point after BDL (**B**; arrows indicate bile lakes; magnification x10) with higher magnifications (x40) in (**C**), displaying cellular infiltrates (asterisk) and formation of artificial bile ductules (arrowhead).

Figure 3: Analysis of the proliferative and cellular response at multiple time points after BDL. Quantitative immunohistochemical analysis of BrdU-positive biliary epithelial cells (**A**), liver cells positive for α -SMA (**B**) and S100a4 (**C**), BrdU-positive hepatocytes (**D**) and Kupffer cells (**E**) and CTGF-positive cells (**F**). Values are given in means \pm SEM of five independent experiments per time point. Corresponding representative immunohistochemical stainings are shown in the right panel (magnifications x40).

Figure 4: Analysis of proliferation and extracellular matrix accumulation. mRNA quantification of the proliferation marker Ki67 (**A**) by Fluidigm real-time PCR. Values are given in means \pm SEM of five independent experiments per time point. Quantitative analysis of extracellular matrix deposition (**B**) and representative histological images (**C**; magnification x10) of Sirius red-positive areas at multiple time points after BDL. Values are given in means \pm SEM of five independent experiments per time point.

Figure 5: Heat maps displaying gene expression pattern at multiple time points after BDL. Gene expression relative to Gapdh gene, obtained from Fluidigm qPCR, are shown as fold changes to sham mice (0 h) and are displayed in log₂ scale. Red colour indicates up-regulation (2), blue down-regulation (-2) and

white transcription levels about 1. **(A)** selected ADME genes, **(B)** selected fibrogenesis genes, and **(C)** selected cytokine genes.

Figure 6: mRNA quantification of different selected genes by Fluidigm real-time PCR displayed in log₂ scale. **(A)** Cyp1a2, **(B)** Cyp24a1, **(C)** Gstm1, **(D)** Nr0b2, **(E)** Col1a1, **(F)** Col3a1, **(G)**, Fn1, **(H)** Mmp10, **(I)** Tgfb2, **(J)** Il2, **(K)** Il28b, **(L)** Tnfrsf1a. Values are given in means \pm SEM of five independent experiments per time point.

Figure 7: Correlation matrix of factors. Correlation matrix (YS3) of the subset of significantly changed factors determined by ANOVA. Positive correlation depicted in blue, negative correlation in red according to color key. Side dendrogram depicts the hierarchical clustering results with the 6 time course clusters marked in the color sidebar (see Figure 9 for the time courses corresponding to clusters c1-c6). Histological factors marked with H, immunostainings with A, and biochemical factors with B. The list of full names is provided in Supplement 2, gene probes.

Figure 8: Histological (H), biochemical (B), and immunostaining (A) correlations. All correlations are YS3 correlation between the respective factors with numerical values provided in Supplement 2. **(A)** Correlation between histological, biochemical and antibody factors and gene transcripts. Only genes with at least one YS3 correlation of $\text{abs(YS3)} \geq 0.6$ are shown. Positive correlation in blue, negative correlation in red analog to Figure 7, with the area of circles corresponding to the correlation coefficients. **(B)** Correlation between histological, biochemical and antibody factors. **(C)** Highest absolute correlations between histological, biochemical and antibody factors and all other ANOVA filtered factors. Data sorted from left to right by absolute value of correlation. Color and size of the filled pie corresponds to the respective correlation value.

Figure 9: Time course clusters in BDL. Time course clusters based on YS3 correlation with hierarchical clustering based on complete linkage corresponding to clusters in Figure 8. Mean cluster time course (averaged over all factors and repeats) depicted in blue, normalized factor representatives of cluster in grey.

Shaded grey area corresponding to SD of all mean factor time courses. Top correlations based on YS3 between the factors in the cluster and the mean cluster time course are shown above the time course. Histological factors marked with H, immunostainings with A, and biochemical factors with B. Cluster members are fully enumerated in the figure for all classes with exception of cluster 4 with the members and respective correlation to mean cluster time course being: Timp1 (0.94), bilirubin (B 0.92), Ccr2 (0.92), CTGF (A 0.91), Tgfb2 (0.89), α -SMA (A 0.89), Ccl5 (0.88), Tgfb1 (0.88), Ccl3 (0.87), Tnc (0.87), Cd14 (0.87), Ccl2 (0.86), Cd86 (0.86), Pdgfb (0.86), Col1a1 (0.86), Cxcl3 (0.86), Ccl4 (0.85), Cxcl5 (0.85), Il10ra (0.85), Col3a1 (0.85), Il10rb (0.84), Ccl7 (0.82), Cd69 (0.82), Ifnar1 (0.82), Tnf (0.82), Osm (0.81), Sparc (0.8), Il6 (0.8), Tnfrsf1b (0.8), Cxcr2 (0.78), Il1b (0.78), Timp2 (0.77), Ifnar2 (0.77), Ccr5 (0.77), Il10 (0.76), Osmr (0.75), Gsta2 (0.74), Il4 (0.71), Ifng (0.71), Ccl8 (0.71), Hgf (0.7), Bak1 (0.7), Mrc1 (0.69), Tgfb2 (0.69), Ccr3 (0.68), Actb (0.68), S100a4 (A 0.66), Il13 (0.66), Met (0.66), bile infarcts (H 0.65), Il6st (0.63), Tnfrsf1a (0.63), Mki67 (0.62), Birc5 (0.6), Ctgf (0.58), BEC (H 0.56), Bax (0.56), Notch1 (0.54), Cxcr1 (0.51), Gstm1 (0.45), Cdh1 (0.42)

Figure 10: Decision tree for disease progression. (A) Regression tree for the prediction of time phases after BDL based on mean cluster time courses. The best gene representatives from the clusters are depicted above the decision rules. The regression tree classifies the data in 6 time classes 0h, 6h, 14h, 24h, 6d, 14d with information on mean time value, range, and number and percentage of samples falling given in the respective class. The best tree based on genes, histological, biochemical and antibody factors is highly similar to the best gene tree, with the single change of using S100a4 instead of Col1a1 for the decision on cluster c4 and allowing GLDH as equally good alternative to Fn1 in c3. (B) Predictive performance of decision tree. The predictive performance of the regression tree was evaluated using all single factor combinations from the individual clusters (white), a random sample (N=10000) of two factors from each cluster (gray), the best gene representative tree (red), and the mean cluster data (blue, trainings data).

Figure 11: Outline of the disease process. Each box is dedicated to a specific disease aspect (first line) which is represented by a commonly known marker (second line) or several markers. Below (in small font) the genes are shown whose expression is correlated to the factor above.