Top-Down Proteomics Identifies Plasma Proteoform Signatures of Liver Cirrhosis Progression

Eleonora Forte[1, 2, #], Jes M. Sanders[2], Indira Pla[1], Vijaya Lakshmi Kanchustambham[1], Michael A. R. Hollas[1], Che-Fan Huang[1], Aniel Sanchez[1], Katrina N. Peterson[1], Rafael D. Melani[1, ^], Alexander Huang[2], Praneet Polineni [2], Julianna M. Doll[2], Zachary Dietch[2], Neil L. Kelleher[1,3,4, *], and Daniela P. Ladner[2, *].

1. Proteomics Center of Excellence, Northwestern University, Evanston, IL, 60208, USA;
2. Northwestern University Transplant Outcomes Research Collaborative (NUTORC), Comprehensive Transplant Center, Feinberg School of Medicine, Northwestern University, Chicago, IL, 60611, USA;
3. Department of Chemistry, Northwestern University, Evanston, IL, 60208, USA;
4. Department of Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, Chicago, IL, 60611, USA.

\#   Current address: Department of Medicine, College of Medicine, University of Illinois Chicago, Chicago. IL 60612, USA

^   Current address: Thermo Fisher Scientific, San Jose, CA 95134, USA

*   Co-last/senior authors

Running Title: Proteoform identification in liver cirrhosis progression
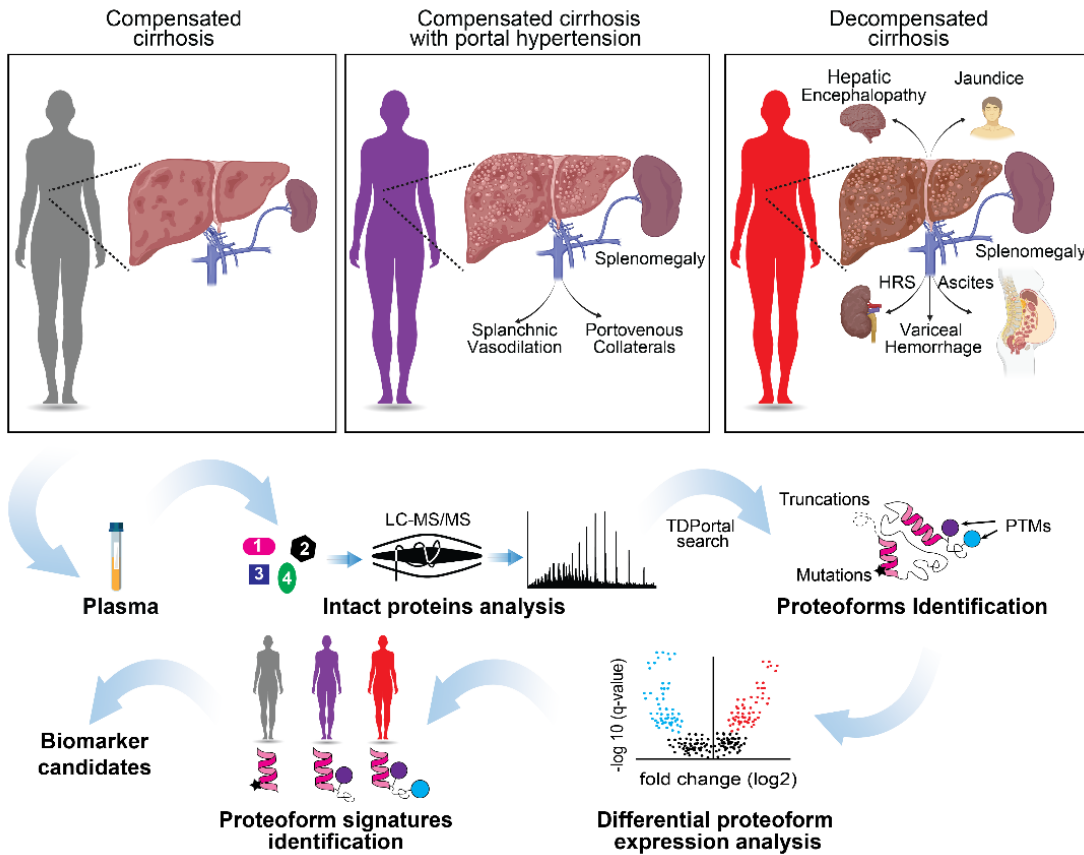
Corresponding Author

Neil L. Kelleher: n-kelleher@northwestern.edu

Conflict-of-interest statement: NLK is involved in entrepreneurial activities in top-down proteomics and consults for Thermo Fisher Scientific. RDM is a current Thermo Fisher Scientific employee. The other authors have declared that no conflict of interest exists.

Graphical Abstract

Highlights:

- Performed a pilot top-down LC-MS/MS analysis to identify proteoforms (PFRs) in the plasma of patients with 3 progressive stages of liver cirrhosis.

- Identified 2867 proteoforms (PFRs) and 209 differentially regulated proteoforms (DRPs) in the different stages of the disease.

- Identified DRP profiles able to potentially distinguish early from late stages of the disease, including 115 liver-derived DRPs.

- Fibrinogen alpha chain, haptoglobin, and Apo A-I are the proteins with the highest number of DRPs and represent potential candidate biomarkers of liver cirrhosis progression.

2

Abbreviations

Abbreviations: BMI (body mass index), WBC (white blood cell count), HGB (hemoglobin), PLT (platelets), HCT (hematocrit), PT (prothrombin time), INR (international normalized ratio), AST (aspartate aminotransferase), ALT (alanine aminotransferase), ALP (alkaline phosphatase), TBILI (total bilirubin), CR (serum creatinine), Na (Sodium), GFR (glomerular filtration rate), MELD-Na (Model for End Stage Liver Disease with Sodium), CAP (controlled attenuation parameter) pHTN (portal hypertension), AIH (autoimmune hepatitis), PBC (primary biliary cholangitis), ETOH (alcohol-associated liver disease), HBV (hepatitis B virus), HCV (hepatitis C virus), MASH (metabolic dysfunction associated steatohepatitis), MASLD (metabolic dysfunction steatotic liver disease), pHTN (portal hypertension), HVPG (Hepatic vein pressure gradients), HE (hepatic encephalopathy), Top-down Proteomics (TDP), Differentially expressed proteoform (DEP), PTM (post-translational modifications), MOD (Modification), aa (amino acids), LC-MS/MS (Liquid chromatography with tandem mass spectrometry), PEPPI-MS (Passively Eluting Proteins from Polyacrylamide gels as Intact species for MS), GO (gene ontology), BP (biological processes), Cluster (C), HDL (high-density lipoprotein), LCAT (lecithin cholesterol acyltransferase), BCA (bicinchoninic acid), IRM (ion routing multipole), PAM (Partitioning Around Medoids), FDR (false discovery rate).

Abstract

Cirrhosis, advanced liver disease, affects 2-5 million Americans. While most patients have compensated cirrhosis and may be fairly asymptomatic, many decompensate and experience life-threatening complications such as gastrointestinal bleeding, confusion (hepatic encephalopathy), and ascites, reducing life expectancy from 12 to less than 2 years. Among patients with compensated cirrhosis, identifying patients at high risk of decompensation is critical to optimize care and reduce morbidity and mortality. Therefore, it is important to preferentially direct them towards specialty care which cannot be provided to all patients with cirrhosis. We used discovery Top-down Proteomics (TDP) to identify differentially expressed proteoforms (DEPs) in the plasma of patients with progressive stages of liver cirrhosis with the ultimate goal to identify candidate biomarkers of disease progression. In this pilot study, we identified 209 DEPs across three stages of cirrhosis (compensated, compensated with portal hypertension, and decompensated), of which 115 derived from proteins enriched in the liver at a transcriptional level and discriminated the three stages of cirrhosis. Enrichment analyses demonstrated DEPs are involved in several metabolic and immunological processes known to be impacted by cirrhosis progression. We have preliminarily defined the plasma proteoform signatures of cirrhosis patients, setting the stage for ongoing discovery and validation of biomarkers for early diagnosis, risk stratification, and disease monitoring.

Introduction

The prevalence of cirrhosis and end-stage liver disease is increasing and pose a significant public health burden, in terms of hospitalizations, cost and mortality [1-5]. In fact, mortality of cirrhosis is comparable to diabetes and pneumonia and surpasses the mortality associated with most cancers [6, 7]. Cirrhosis is the result of liver injury most frequently caused by alcohol use disorder, metabolic dysfunction-associated liver disease [MASLD; previously, non-alcoholic fatty liver disease (NAFLD)], and hepatitis C (HCV), along with other less frequent etiologies (e.g., biliary, hepatitis B). Most patients with cirrhosis have compensated chronic disease, which carries a life expectancy of ~12 years [8]. However, as cirrhosis progresses, intrahepatic structural changes lead to diversion of blood flow to the heart through alternate small blood vessels, rather than the liver. This process of porto-venous collateral formation and splanchnic vasodilation is called portal hypertension (pHTN) [9-11]. As liver disease progresses, pHTN can turn into decompensated cirrhosis, considered as the occurrence of confusion (hepatic encephalopathy), ascites, and gastrointestinal bleeding. Decompensated cirrhosis is associated with significant morbidity, high rates of hospitalizations, and decreased life expectancy of 1.8 years. On average one in ten patients with cirrhosis experience a decompensating event every year [8]. With the increasing prevalence of metabolic syndrome and alcohol use disorder, the two major causes of cirrhosis, cirrhosis-related morbidity and mortality are expected to rise [5, 12].

Early diagnosis, optimal disease management, and identification of patients who are at high risk of decompensation is critical to provide patients with targeted and timely interventions, mitigate disease progression, and decrease morbidity and mortality [13, 14]. While only experimental drugs exist to mitigate disease progression [15, 16], timely specialty care can optimize the disease course. However, with 2-5M adults affected by cirrhosis, and limited resources including specialists, identifying those at highest risk for disease progression is paramount for optimal triage and outcomes. Presently, such diagnostic tools are not widely available. Hepatic vein pressure

gradients (HVPG) have been shown to be a marker of decompensation [17], but the procedure is invasive and may not be available at all centers. Clinical risk scores may hold some prognostic value, but the majority were initially established to predict mortality and not disease progression [18, 19]. To this end, non-invasive and reliable diagnostic tools to identify and predict identify those at highest risk are required. Top-down proteomics (TDP) can be used to develop predictive markers of disease progression by leveraging modified proteins that drive mechanisms of disease. TDP can characterize proteins present in biological systems by determining the molecular composition of all forms of a given protein, so called "proteoforms." Proteoforms (PFRs) arise due to genetic variation in protein coding regions, alternative splicing, and/or post-translational modifications (PTMs) [20]. It is estimated that the human proteome contains >50 million unique PFRs compared to ~20,300 gene-encoded proteins [21]. As a result, by delivering information at the proteoform level, TDP captures the complexity of a biological system better than that of traditional peptide-based proteomics approaches, which only provide data at the protein level. We have recently demonstrated the importance of PFRs as biomarkers in the context of liver transplantation with the identification of PFR signatures of hematopoietic stem cells able to distinguish acute rejection from normal graft function [22]. PFRs specific for cirrhosis, especially those specific to different stages of cirrhosis, have not been described, and their identification could greatly improve the diagnosis and treatment of those at highest risk for progression.

In this study, we applied a TDP approach to identify PFRs present in the plasma of patients with three different stages of cirrhosis [compensated (Stage I)/ compensated + pHTN (Stage II) / decompensated (Stage III)]. We found that patients with different stages of cirrhosis have unique blood proteoform profiles that impact important pathophysiological pathways and cellular processes. Our findings create a framework for applying proteoform-based analysis to biomarker

6

discovery in the cirrhosis population and can provide mechanistic insight to the disrupted processes as the disease progresses.

Experimental Procedures

Human Subject Selection and Plasma Collection. Peripheral blood was collected from 30 subjects with cirrhosis as part of clinical cirrhosis care at Northwestern Medicine who had compensated cirrhosis (Stage I), compensated cirrhosis + pHTN (Stage II), and decompensated cirrhosis (Stage III). Plasma was then isolated and stored at -80°C until protein extraction. Samples from adults (>18 years) with 1) compensated cirrhosis, 2) compensated cirrhosis + pHTN, and 3) decompensated cirrhosis were used. Diagnosis of cirrhosis and stage of cirrhosis was determined by careful medical record review (JMS, PP, DPL), including Fibroscan®, magnetic resonance elastography, or biopsy. Decompensation was defined as the presence of hepatic encephalopathy (HE), ascites, hepatorenal syndrome, spontaneous bacterial peritonitis, and history of variceal gastrointestinal bleeding [23, 24]. pHTN was diagnosed either by imaging, presence of thrombocytopenia (platelets <100), or evidence of esophageal/gastric varices or portal hypertensive gastropathy on surveillance/diagnostic endoscopy [24, 25]. Demographic, clinical characteristics, and laboratory values (complete blood count (CBC), liver enzymes, basic chemistries, and INR) were collected for each subject at the time of lab draw. (Figure S1). The Model for End Stage Liver Disease Sodium (MELD-Na) was calculated for each subject.

Study approval. This study was conducted in compliance with the NIH guidelines for studies involving human subjects. The Northwestern Institutional Review Board (IRB) issued a Waiver of Consent and approved this study under STU00216399.

Plasma Protein Extraction and Fractionation. Protein concentration in the plasma was determined by a bicinchoninic acid (BCA) protein assay kit (Thermo Fisher Scientific) following the manufacturer instructions. Passively Eluting Proteins from Polyacrylamide gels as Intact

7

species for MS (PEPPI-MS) was used to fractionate proteins with a molecular weight lower than 30kDa and prepare them for the mass spectrometry analysis [26]. Briefly, 500 µg of plasma proteins were resuspended in 4X NuPAGE loading buffer (Thermo Fisher Scientific) with 50 mM DTT and boiled at 95°C for 10 minutes. Proteins were fractionated on a NuPAGE 4-12% Bis-Tris gel (Thermo Fisher Scientific) for 10 min at 70 V and 15 minutes at 150 V. The gel was briefly rinsed with deionized water and gel sections with proteins with a molecular weight below 30kDa were excised from the gel and transferred to a Protein LoBind tube (Eppendorf) containing 100 mM ammonium bicarbonate, pH 9 with 0.1% SDS. Proteins were extracted from the gel by crushing it with a pestle and shaking the slurry for 20 minutes at 1400 rpm. Gel debris were removed with 0.45 µm centrifugal filters. The protein fractions were methanol/chloroform/water precipitated by using a slightly modified Wessel and Flügge's method [26, 27]. Briefly, 2 volumes of methanol and 0.5 volume of chloroform were added to 1 volume of the protein sample and the mixture vortexed. 1 volume of water was then added to the solution, and the mixture was vortexed and centrifuged at 20,000 rpm for 20 minutes at 4°C. Without disturbing the interface, most of the upper layer was withdrawn and discarded. Next, 1 mL of methanol was added to the lower phase and the mixture was vortexed and centrifuged at 20,000 rpm for 30 minutes. The supernatant was removed. The precipitated protein pellet was air-dried, resuspended in 30 µL of LC-MS buffer A (5% acetonitrile, 94.8% water, and 0.2% formic acid), and subjected to LC-MS/MS.

LC-MS/MS Analysis. Proteoforms separation was conducted on a Dionex Ultimate 3000 (Thermo Fisher Scientific) by using a trap/column system consisting of: 1) a trap (150 µm I.D. x 2.5 cm length) packed in-house with ReproSil C4, 3 µm particles (Dr. Maisch), 2) a FlowChip column (25 cm in length x 100 µm i.d. obtained from New Objective). For the separation the flow rate was set at 1 µL/min and the following gradient was used: 5% B (5% water, 94.8% acetonitrile, and 0.2% formic acid) from 0 to 10 min., 20% B at 15 min, 55% B at 100 min, 95% B from 103 to 110 min, 5% B at 115 to 120 min.

8

Proteoform identification in liver cirrhosis progression

The LC is in line with the Orbitrap Eclipse (Thermo Fisher Scientific) mass spectrometer operating in "protein mode" with 2 mTorr of N_2 pressure in the ion routing multipole (IRM). Transfer capillary temperature was set at 320°C, ion funnel RF was set at 50%, and a 15 V of source CID was applied. MS[1] spectra were acquired at 120,000 of resolving power (at *m/z* 200), AGC target value of 1000%, 100 ms of maximum injection time, and 1 μscan. Data-dependent top-N-2 sec MS[2] method used 32 NCE for HCD to generate fragmentation spectra acquired at 60,000 resolving power (at *m/z* 200), with target AGC values of 2000%, 600 ms maximum injection time, and 1 μscan. Precursors were quadrupole isolated using a 3 *m/z* isolation window, dynamic exclusion of 60 s duration, and threshold of $1x10^4$ intensity. Each sample was acquired in triplicate for a total of 90 runs.

Experimental Design and Statistical Rationale

Data Analysis for LC-MS/MS. The raw data files were processed with the publicly available workflow on TDPortal (https://portal.nrtdp.northwestern.edu, Code Set 4.0.0) that performs mass inference, searches a database of human proteoforms derived from Swiss-Prot (June 2020) with curated histones, and estimates a conservative, context-dependent 1% false discovery rate (FDR) at the protein, isoform, and proteoform levels [28].

Statistics and Functional enrichment analysis. Data derived from randomized biological & technical replicates were used to assess variability associated with sample processing. From the aggregated data sets a batch effect was observed and corrected with batch standardization (z-score) and statistical analysis was conducted using SAS (Cary, NC) with custom scripts. The appropriate hierarchical linear statistical model was applied for quantitation using SAS PROC MIXED (SAS Institute). Normalized and batch corrected ion intensities from the data files were run through a nested ANOVA process to measure both effect size and significance in differential proteoform expression between the 3 patients' groups. Specific proteoform signatures were then correlated with decompensated outcomes. P-values obtained from the mixed model linear

9

regression were adjusted (Benjamini & Hochberg (BH) algorithm) to control the false discovery rate (FDR) induced by multiple comparison tests. PFRs with adjusted p value < 0.05 and fold changes higher than 1.5 or less than 0.66, were considered differential expressed proteoforms (DEP). A Partitioning Around Medoids (PAM) [29] unsupervised algorithm ('cluster::Pam'Rfunction) was applied to group PFRs based on their expression changes between different disease stages. The number of clusters (k) was manually set to three, as it was the number of clusters that best showed changes in disease progression. To visualize clusters based on differential expressions and changes in specific PFR signatures we created heatmaps using the 'ComplexHeatmap' R package. Clinical biomarkers included in heatmap from Figure 2 were previously standardized across samples using z-score normalization. To know whether the PFR modifications were significantly different between groups of DEPfr, we applied Fisher exact tests followed by an adjustment of p-values (BH). Proteins having at least one deregulated PFR were selected to perform Gene Ontology enrichment analysis based on Biological Process, and the analysis was carried out using 'clusterProfiles' R package.

Results

1.1 Discovery Top-Down Proteomics (TDP) analysis of plasma characterizes blood proteoform profiles in cirrhosis patients

The cohort included 30 subjects—9 (30%) with compensated cirrhosis (Stage I), 10 (33.3%) with compensated cirrhosis + pHTN (Stage II), and 11 (36.7%) with decompensated cirrhosis (Stage III) (Figure 1A). Demographics and clinical data are shown in Table 1. The mean age of patients was 59.7 years (Stage I: 63.7, Stage II: 55.5, Stage III: 61.9). 40% (n=12) of patients were female (Stage I: 55.6%, Stage II 40.0%, Stage III: 27.3%), and 56.7% (n=17) were White. The main etiology of cirrhosis was alcohol use disorder in 60% (n=18). The mean Model for End-stage Liver

10

Proteoform identification in liver cirrhosis progression

Disease – Sodium (MELD-Na) was 11.5 (Stage I: 9.9, Stage II: 10.0, and Stage III: 14.1). Detailed laboratory findings are shown in Figure S1. No clinically significant differences were observed.

We first conducted an untargeted quantitative TDP analysis of 0-30 kDa fractions of plasma (Figure 1A). Within the entire cohort 2867 proteoforms from 99 proteins were captured (Table S1). Modifications of proteoforms mainly included C- and N- terminus truncations (Figure S2). Phosphorylation was the most frequent post-translational modification (PTM) (Figure S2). Other common PTMs included monoacetylation, alpha-amino acetylation, S-nitrosyl-L-cysteine, n-pyruvic acid 2-iminyl-L-valine, monomethylation, monohydroxylation, L-gamma-carboxyglutamic acid, and 2-pyrrolidone-5-carboxylic acid (Figure S2).

1.2 Proteoforms are differentially expressed in stages of cirrhosis and are involved in biologically and clinically relevant processes.

We then separated samples according to their diagnostic stage (I-III) (Figure 1B-1D) and identified a total of 663 differentially expressed proteoforms (DEPs) from 48 proteins. Like the ~3000 total proteoforms, modifications of DEPs mainly included truncations, phosphorylation, and acetylation (Figure S2). Pairwise comparisons between the three stages of cirrhosis showed: 1) 74 upregulated and 83 downregulated DEPs comparing Stage III to Stage II (Figure 1B), 216 upregulated and 79 downregulated DEPs comparing Stage III to Stage I (Figure 1C), and 136 upregulated and 75 downregulated DEPs comparing Stage II to Stage I (Figure 1D). In our differential expression analysis, we identified several proteoforms from hemoglobin subunits. As hemoglobin presence could be due to sample handling and preparation, hemoglobin proteoforms were omitted from further analyses. In the pairwise comparisons shown in Figures S3, S4, and S5, we found that some PFRs from a single protein were upregulated while other PFRs from the

11

same protein were downregulated. For example, we identified a total of 14 DEPs for fibrinogen alpha chain, of which 5 were downregulated and 9 were upregulated in Stage III vs Stage II (Figure S3), 11 DEPs for Apolipoprotein A-I, of which 2 were downregulated and 9 were upregulated in Stage III vs Stage I (Figure S4), and 2 DEPs for ceruloplasmin, of which one was upregulated and one was downregulated in Stage II vs Stage I (Figure S5). We next analyzed the DEP modifications found in the three pairwise comparisons (Figure S6A). DEP modifications in the comparison between Stages III and I included mostly truncations (83%), followed by monohydroxylation (7%), alpha-amino acetylation (6%), and phosphorylation (2%). Both L-gamma-carboxyglutamic acid and 2-pyrrolidone-5 carboxylic acid residues occurred in 1% of the modifications. In the comparison between Stages II and I, the top three modifications were again truncation (75%), monohydroxylation (11%), and alpha-amino acetylation (9%). Phosphorylated, monomethylated, L-gamma-carboxyglutamic acid and 2-pyrrolidone-5 carboxylic acid residues each occurred in 1% of the modifications. Finally, in the comparison between Stages III vs II, we found 63% of PTMs were truncations followed by phosphorylation (11%), alpha-amino acetylation (11%), monohydroxylation (9%), carboxylation of glutamic acid residues (4%), monoacetylation (1%), and pyrrolidone-5 carboxylic acid residues (1%). We then compared the number of each modification in the three sets of DEPs (sets: Stage III vs I, II vs I, III vs II) and found that phosphorylation was the only PTM to be statistically significantly different between the three sets (Figure 6B).

DEPs were then clustered based on their regulation patterns in each stage (Figure 2A-2B). Laboratory test values from clinical evaluation were also included in the clustering to find possible associations between DEPs and clinical markers of disease severity. We also performed gene ontology (GO) based on biological processes (BP) enrichment analyses to correlate PFR profiles with affected biological processes. Cluster 1 (C1) included DEPs that were significantly

12

upregulated in decompensated disease (Stage III). This cluster includes 57 DEPs from 23 proteins. The proteins with the highest number of DEPs in C1 were fibrinogen alpha chain with 15 DEPs, apolipoprotein A-I with 10 DEPs, and insulin-like growth factor-binding protein 4 with 5 DEPs. Pathway enrichment analysis indicated that C1 DEPs derived from proteins involved in several biological processes, including the humoral immune response, apoptosis regulation, complement activation, phagocytosis, and negative regulation of peptidase activity (Figure 2C, Table S2). In addition, C1 DEPs clustered with several clinical markers, including alkaline phosphatase (ALP), which was significantly higher in Stage III (Figure S1), total bilirubin (TBILI), which was significantly increased between Stage II and III, and MELD-Na, which was significantly and appropriately higher in Stage III compared to Stage I (MELD-Na is a marker of disease severity). Other clinical markers included in this cluster were International Normalized Ratio (INR), prothrombin (PT), aspartate aminotransferase (AST), and creatinine (CR), although their increase in Stage III was not statistically significant.

Cluster 2 (C2) included 57 DEPs upregulated in Stage II. The PFRs derived from 9 proteins involved in phospholipid efflux and processes regulated by apolipoproteins (Figure 2C, Table S2). The proteins with the highest number of DEPs were apolipoprotein A-I, transthyretin, and fibrinogen alpha chain with 22, 15, and 7 DEPs, respectively. Clinical markers in this cluster included glomerular filtration rate (GFR), Sodium (Na), alanine aminotransferase (ALT), and albumin. Interestingly, 6 out of 7 albumin PFRs detected in the TDP discovery analysis were found in this cluster together with albumin protein, which was analyzed by routine clinical laboratory testing (Figure S1). Both albumin PFR and protein levels were significantly higher in the plasma of patients with Stage II and lower in those with Stage III. In addition, fold changes obtained for the albumin protein (laboratory results) and albumin PFRs were comparable despite using two distinct approaches, confirming the validity of the TDP discovery analysis.

13

Cluster 3 (C3) included 95 DEPs upregulated in Stage I. These DEPs derived from 24 proteins associated with the immune system biological processes, such as the acute inflammatory response, acute-phase response, and platelet activation. Other processes enriched in C3 were regulation of protein secretions and cell-cell adhesion (Figure 2C, Table S2). In this cluster, haptoglobin was the protein with the highest number of DEPs (29) followed by fibrinogen alpha chain (16). Interestingly, clinical markers related to the immune system, such as white blood cell count (WBC), platelet count (PLT), and total proteins were grouped in C3. However, only the total protein value changed significantly between Stages I vs II and I vs III (Figure S1).

Overall, clusters containing DEPs upregulated in the late stages of the disease (C1 and C2) were mainly from fibrinogen alpha chain, transthyretin, and apolipoprotein A-I proteins (Figure 2B). However, another important group of DEPs from fibrinogen alpha chain (14 PFRs) were upregulated in Stage I and clustered with haptoglobin in C3. In terms of biological processes, DEPs upregulated in the early stage of the disease were involved in cellular defense mechanisms, while those upregulated in the late stages of cirrhosis were associated with cell death. Percentages of DEP modifications differed in the three clusters as shown in Figure S7A. DEPs in C3 had the highest percentage of truncations (90% versus 77% and 54% of DEPs observed in C1 and C2, respectively). Phosphorylation was higher in C2 (10%) than C1 (3%) and not observed in DEPs in C3. In addition, C2 had a higher percentage of alpha-amino acetylated (C1=6%, C2=13%, C3=6%), monohydroxylated (C1=11%, C2=17%, C3=1%), and L-gamma carboxyglutamic acid (C1=0%, C2= 6%, C3=0%) residues. C2 did not contain any 2-pyrrolidone-5-carboxylic acid modifications, which was instead observed in 1% of C1 and C3 DEPs. Monohydroxylation was only observed in C1 and monomethylation in C3. Finally, the comparison of the number of each modification in the three clusters indicated a significant variation in the

14

types of PFR modifications at the level of phosphorylated, monohydroxylated, alpha-amino acetylated, and L-gamma-carboxyglutamic residues (Figure S7B).

1.3   Liver-derived proteoforms may be candidate biomarkers for cirrhosis progression.

Heatmaps were generated for the DEPs identified above, but given that cirrhosis is associated with ongoing damage to hepatocytes, we primarily focused on PFRs derived from proteins that are enriched in the liver tissue (https://www.proteinatlas.org) at a transcriptional level as they may be the most sensitive to differences between stages (Figures S3-S5). Heatmaps of these liver-enriched DEPs were created for each stage of cirrhosis, along with their modifications, including truncations and a variety of PTMs (Figure 3). As the majority of DEPs discriminating the different stages of the disease derived from fibrinogen alpha chain, apolipoprotein A-1, and haptoglobin, we analyzed those PFRs in more detail (Figure 3, Figure S8). Fibrinogen alpha chain PFRs were found in all clusters. Specifically, 14 PFRs were detected in C1, 7 in C2 and 15 in C3. The canonical sequence of fibrinogen alpha chain includes 866 amino acids. All the identified PFRs were truncated forms of the protein missing both the N- and C termini. In C1 (PFRs upregulated in Stage III) PFR sequences were generally longer with an average of 126 amino acids. The average length in the other clusters was 58.86 and 58 for C2 and C3, respectively. This suggests the presence of shorter truncations in early stages of the disease. In terms of PTMs, the majority was found in C1. In this cluster we detected 3 PFRs with monohydroxylation and 3 phosphorylated PFRs. In C2, PTMs included one monohydroxylation and one phosphorylation. C3 (PFRs upregulated in Stage I) contained one PFR with monohydroxylation. Of the 32 apolipoprotein A-I PFRs, 22 were in C2 and 10 in C3. The majority were truncated and had 1) no additional PTM, 2) alpha-amino acetylation, or 3) methionine oxidation. A proline 27 to histidine point mutation was also found in 2 PFRs in C2. Haptoglobin PFRs included both truncations of the canonical protein (P00738-1) and of an isoform that lacks amino acids 38-96 (P00738-2). In addition, we found

15

mutations in 10 truncated PFRs of haptoglobin P00738-2, of which eight had an aspartic acid residue replacing asparagine in position 70, and two with lysine at the place of glutamic acid in position 71. In terms of PTMs, we found that four of the haptoglobin PFRs were acetylated at the N-terminus. Although not described in detail, we also identified DEPs deriving from proteins that are not enriched in the liver tissue at a transcriptional level (Figures S3-S5). Heatmaps of non-liver enriched PFRs are displayed in Figure S9.

To more clearly define PFR signatures that correlated with disease progression, we created profiles of DEPs that could discriminate Stage I (compensated) from Stages II and III of cirrhosis. Comparisons of PFR expression were performed across cirrhosis stages, with compensated disease (Stage I) used as the reference group (Figure 4A). This analysis identified 167 DEPs (from 37 proteins), of which 115 DEPs derived from 20 proteins enriched in the liver at a transcriptional level (Figure 4B). Tables containing detailed information about these 115 DEPs and their modifications are included in Figures S10-12 and Table S1. Apolipoprotein C-I-PFR#16559 was the only DEP found downregulated in Stage III but upregulated in Stages I and II (Figure 4A-4B, and S10, Dark purple square). There were 14 DEPs downregulated in Stage III of the disease compared to Stage I. These included 6 DEPs from fibrinogen alpha chain, 2 from apolipoprotein A-I and serum amyloid A-1 protein, and one from albumin, apolipoprotein A-II, apolipoprotein C-I, and serum amyloid A-2 protein (Figure 4A-4B, Light purple square). Both apolipoprotein A-I PFRs contained oxidized methionine residues at positions 110 and 136 and had N-terminus truncations at two different positions (residues 67 and 35, respectively) (Figure S10). None of the fibrinogen alpha chain PFRs had a PTM, but all contained both C- and N-termini truncations. The average length of truncation was 55.83 amino acids. Three DEPs were specifically downregulated in Stage II disease and derived from serotransferrin, retinol-binding protein 4, and fibrinogen gamma chain (Figure 4A-4B, and S10, Yellow square).

16

54 DEPs were downregulated in both Stages II and III of the disease compared to Stage I (Figure 4A-4B, and S11, Blue square). Haptoglobin was the protein with the highest number of DEPs (29 PFRs), followed by fibrinogen alpha chain (10 PFRs), alpha-2HS-glycoprotein (3 PFRs), and haptoglobin-related protein (2 PFRs). One DEP each was found for albumin, alpha-1-acid glycoprotein 1, angiogenin, apolipoprotein C-I, ceruloplasmin, clusterin, serotransferrin, and serum amyloid A-1 protein. Haptoglobin PFRs (Figure S11) perfectly overlapped with those characterized in C3 (described above). We again observed both C- and N- termini truncations in all 10 fibrinogen alpha chain PFRs at various positions. The average length of truncation was 58.8 amino acids. One fibrinogen alpha chain PFR was hydroxylated at a proline in position 565 (Figure S11).

Twenty DEPs were upregulated in late Stages (II and III) (Figure 4A-4B, and S12, Red square). These originated from apolipoprotein A-I (10 PFRs), fibrinogen alpha chain (5 PFRs), retinol-binding protein 4 (2 PFRs), apolipoprotein C-I (1 PFR), ceruloplasmin (1 PFR), and complement C4-A (1 PFRs). Sixteen DEPs were exclusively upregulated in Stage II of disease. The upregulated DEPs (Figure 4A-4B, and S12, Green square) were from apolipoprotein A-I (13 PFRs), albumin (2 PFRs), and fibrinogen alpha chain (1 PFR). Finally, 7 DEPs were upregulated in Stage III of the disease and included 4 PFRs derived from fibrinogen alpha chain, 2 from apolipoprotein A-I, and one from alpha-1 antitrypsin (Figure 4A-4B, and S12, Orange square). Collectively, 25 of the 27 apolipoprotein A-I PFRs that discriminated early Stage (I) from later Stage disease (II and III) were upregulated in later disease Stages (Figure 4, Figure S12). These apolipoprotein A-I PFRs mostly contained N-terminus truncations at various locations and oxidation of methionine residues in positions 110 and 136. In addition, a proline to histidine point mutation occurred at position 27 in one PFR that was upregulated only in Stage II disease (Figure

17

Proteoform identification in liver cirrhosis progression

S12, green square, PFR#5052090). Notably, all 25 of these PFRs were greater in length compared to those upregulated in Stage I disease (Figure S10). Fibrinogen alpha chain PFRs upregulated in later-stage disease were all truncated at various residues at the N-terminus, but eight of the 10 PFRs had a C-terminal truncation at residue 629. Finally, two fibrinogen alpha chain PFRs were hydroxylated at a proline in position 565 (Figure S12, green and red square), and one was hydroxylated at a serine in position 501 and had a Phospho-proline at position 565 (Figure S12, orange square, PFR#7501917). The average length of truncation of fibrinogen alpha chain PFRs upregulated in later disease stages was higher compared to early-stage disease (Stage III= 137.75, Stages II and III = 73.8, Stage II= 61, Stage I= 55.83-58.8) (Figures 5, S10, S11).

Discussion

Using quantitative discovery TDP, we discovered DEPs in three stages of cirrhosis – compensated (Stage I), compensated + pHTN (Stage II) and decompensated (Stage III) (Figure 5). Our pilot discovery analysis identified 2867 PFRs in total, of which 663 were differentially expressed. DEPs clustered into 3 groups based on their upregulation in the various stages of cirrhosis. Modifications to these DEPs differed depending on the stage of cirrhosis, and DEPs were involved in numerous, biologically relevant processes. 167 DEPs discriminated Stage I vs Stages II and III, of which 115 derived from 20 proteins enriched in the liver at a transcriptional level. Proteoforms originating from fibrinogen alpha chain, apolipoprotein A-I, and haptoglobin were the most abundant and demonstrated varying degrees of modification across disease stages.

Apolipoprotein A-I is a lipoprotein produced by both the liver and intestine and is a key component of high-density lipoprotein (HDL). Its primary function is in reverse cholesterol transport, but it is also an important mediator of immunological and inflammatory processes [30]. Structural changes,

18

mutations, PTMs, and decreased synthesis of apolipoprotein A-I have been associated with multiple disease entities [31]. In the context of liver disease, decreased apolipoprotein A-I levels have been observed in patients with cirrhosis [32-36]. Although apolipoprotein A-I is a component of an established screening tool for liver fibrosis called the "Fibrotest," scarce studies have assessed the ability of apolipoprotein A-I levels to predict decompensation events or death [34-36]. In the most recent of these, Gurbuz et al used bottom-up proteomics to analyze the serum proteome of hepatocellular proteins in patients with cirrhosis versus healthy controls. Of 903 proteins identified, 81 proteins of hepatocellular origin were differentially expressed between groups: 65 downregulated and 16 upregulated in subjects with cirrhosis. Decreased serum levels of apolipoprotein A-I were observed in cirrhosis patients and correlated with a transition from stable to unstable decompensated cirrhosis [36]. In contrast, we observed most apolipoprotein A-I proteoforms were upregulated in later stage disease. Modifications of apolipoprotein A-I proteoforms included truncation, oxidation of two methionine residues, and alpha-amino acetylation. Nearly all the observed proteoforms in later-stage disease (Figure S12) contained more amino acids compared to proteoforms seen in Stage I disease. This could be a result of secretion of apolipoprotein A-I's prepropeptide and/or inappropriate cleavage of the prepropeptide in subjects with more severe disease [31, 37], ultimately leading to dysfunction of apolipoprotein A-I proteins. The hydroxylation of apolipoprotein A-I may exacerbate this as oxidation of HDL can negatively affect cholesterol efflux, anti-apoptotic effects, anti-peroxidation, and other anti-inflammatory actions [38]. More specifically, oxidation of methionine residues within the apolipoprotein A-I protein can interfere with apolipoprotein A-I's interaction with lecithin cholesterol acyltransferase (LCAT) impairing reverse cholesterol transport [39].

Haptoglobin, an acute-phase protein that scavenges cell-free hemoglobin, has also been studied as a biomarker of liver disease severity. Like apolipoprotein A-I, decreased haptoglobin levels are correlated with higher degrees of liver fibrosis [40, 41]. There is also evidence to suggest glycoforms

19

of haptoglobin may be useful in differentiating various stages and etiologies of liver disease [41-47]. However, little is known about the specific effect of cirrhosis progression on haptoglobin function, and there are few to no studies that have investigated haptoglobin and its modifications as a predictor of future decompensation events. We observed DEPs of haptoglobin were relatively upregulated in Stage I disease (Figures 3, 4, S11). All haptoglobin DEPs were identified as a truncation, but review of the amino acid sequences would suggest DEPs derive from mature alpha-1 or alpha-2 haptoglobin chains (www.uniprot.org,[48]). This is additionally supported by the presence of two point mutations at residues 71 and 72, which are known mutations of two alleles of the alpha-1 chain, *alpha-1F* and *alpha-1S* [48]. Interestingly, the alpha-amino acetylation PTM observed in our study was present only on the alpha-1 chain and has not been previously studied in the context of cirrhosis. Alpha-amino acetylation can impact protein half-life, protein folding, protein-protein interactions, and membrane targeting, but it is unknown what effect this specific PTM has on haptoglobin. As such, future investigations should elucidate the importance of this PTM in the setting of cirrhosis as it may impact the production of and function of mature haptoglobin proteins in patients with more severe disease.

Cirrhosis associated coagulopathy has been well described, and subjects with cirrhosis are predisposed to both bleeding and thrombotic events [49]. Patients with stable liver disease generally have normal levels of fibrinogen, but as disease progresses fibrinogen levels steadily decrease [50, 51]. Dysfibrinogenemia also plays an important role in coagulopathy as modifications to fibrinogen subunits can impact both structure and function [52, 53]. In our cohort, fibrinogen DEPs originated most commonly from the alpha chain and were of varying lengths (S10-S12). Given the full-length fibrinogen alpha chain is ~610 amino acids after translation and intracellular processing [54], it's likely these alpha chain DEPs are byproducts of fibrin breakdown based upon their much shorter amino acid sequences (Figure S9). The discrepancies in length of detected DEPs across clusters may be a consequence of complex alterations in fibrin degradation and of

20

the proteinases important for fibrin breakdown (*i.e.,* thrombin, plasmin, neutrophil elastase, and proteinase 3) in various stages of disease. This is consistent with studies that have shown D-dimer, a fibrin degradation product, is elevated in patients with cirrhosis and is associated with increased mortality in patients with later stage disease [55-59]. Further characterization of these specific DEPs can provide a starting point for mechanistic investigations of fibrinolysis in patients with cirrhosis and may present an opportunity for therapeutic intervention, whether that be on the fibrinogen molecule itself or on the proteinases responsible for its degradation.

There are key limitations to our study. First, we are limited by the sample size of patients as this was an early pilot study to identify biomarker candidates. Considering this, a benefit of our proteoform-based analysis is the degree of resolution achieved, which can identify significant differences even in a small cohort study. Still, targeted validation studies in a larger cohort are needed to confirm and expand upon our findings. Second, our cohort consisted mostly of patients with alcohol or MASH-related cirrhosis, as they are the two most common etiologies of cirrhosis in the United States. Future studies will need to include additional patients with other etiologies of disease to ensure generalizability across cirrhosis populations. Finally, it is possible some proteoforms identified in our discovery TDP analysis are a result of electrospray ionization or sample handling during preparation. Specifically, oxidation is a known byproduct of such techniques, and we can assume that at least a few of the proteoforms detected may be a result of sample processing. However, the fact we observed significant differences in the expression of a number of oxidized proteoforms across samples that underwent identical processing makes this less likely and suggests biological processes are driving these differences.

In summary, quantitative TDP of plasma proteins from patients with cirrhosis revealed unique plasma proteoform profiles associated with stages of disease in this pilot study. To our knowledge, this is the first report of such proteoform analysis in patients with cirrhosis. Although we focused

21

on the most abundant DEPs (apolipoprotein A-I, haptoglobin, and fibrinogen alpha chain), there were other PFRs identified in this study that may be candidate biomarkers for disease progression, both individually and in combination with other PFRs. Moving forward, we will perform targeted studies on larger cohorts to validate and further characterize the identified PFRs. The TDP methodology employed here could greatly increase the accuracy of blood-based analyses due to the identification of DEPs that are highly sensitive and specific probes of the underlying biology over time. In a clinical setting, understanding the change in PFRs from one stage of cirrhosis to the next would enable prompt recognition of patients transitioning from compensated to later-stage disease, presenting an opportunity for targeted risk reduction or early intervention in such patients.

Acknowledgments

Data Availability. Raw files and tdReport files can be found in Massive (Accession MSV000094311. The search results in the tdReport format can be viewed by using TDViewer freely available at http://topdownviewer.northwestern.edu. The search results were further analyzed, and figures were generated with R-custom scripts that are available on request.

Supplemental data

Supplementary Figures were placed at the end of this document.

Supplementary Tables can be found here: https://doi.org/10.6084/m9.figshare.26062579

Author Contributions. Eleonora Forte: Conceptualization, Investigation, Validation, Formal analysis, Writing-Original Draft, Visualization, Supervision, Project administration; **Jes M. Sanders:** Writing - Original Draft, Resources; **Indira Pla:** Software, Validation, Formal analysis, Data Curation, Writing- Original Draft, Visualization; **Vijaya Lakshmi Kanchustambham** and **Che-Fan Huang**: Methodology, Investigation, Validation, Writing - Review & Editing; **Michael A. R. Hollas**: Software, Validation, Formal analysis, Data Curation, Writing- Original Draft, Visualization; **Aniel Sanchez**: Investigation, Data Curation, Writing - Review & Editing; **Katrina N. Peterson:** Software, Validation, Formal analysis, Data Curation, Writing - Review & Editing; **Rafael D. Melani:** Conceptualization, Methodology, Writing – Review & Editing; **Alexander Huang:** Formal analysis, Writing – Review & Editing; Praneet Polineni,ˑ Julianna M. Doll, Zachary Dietch: Resources, Data Curation, Writing - Review & Editing, **Neil L. Kelleher:** Conceptualization, Writing – Review & Editing, Funding acquisition; Daniela P. Ladner: Conceptualization, Writing – Review & Editing, Funding acquisition.

References

(1) Mokdad, A. A.; Lopez, A. D.; Shahraz, S.; Lozano, R.; Mokdad, A. H.; Stanaway, J.; Murray, C. J.; Naghavi, M. Liver cirrhosis mortality in 187 countries between 1980 and 2010: a systematic analysis. *BMC Med* **2014**, *12*, 145. DOI: 10.1186/s12916-014-0145-y.

(2) Xu, J.; Murphy, S. L.; Kochanek, K. D.; Bastian, B.; Arias, E. Deaths: Final Data for 2016. *Natl Vital Stat Rep* **2018**, *67* (5), 1-76.

(3) Collaborators, G. B. D. C. The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol* **2020**, *5* (3), 245-266. DOI: 10.1016/S2468-1253(19)30349-8.

(4) Guo, A.; Mazumder, N. R.; Ladner, D. P.; Foraker, R. E. Predicting mortality among patients with liver cirrhosis in electronic health records with machine learning. *PLoS One* **2021**, *16* (8), e0256428. DOI: 10.1371/journal.pone.0256428.

(5) Ladner, D. P.; Gmeiner, M.; Hasjim, B. J.; Mazumder, N.; Kang, R.; Parker, E.; Stephen, J.; Polineni, P.; Chorniy, A.; Zhao, L.; et al. Increasing prevalence of cirrhosis among insured adults

in the United States, 2012-2018. *PLoS One* **2024**, *19* (2), e0298887. DOI: 10.1371/journal.pone.0298887.

(6) Tapper, E. B.; Parikh, N. D. Mortality due to cirrhosis and liver cancer in the United States, 1999-2016: observational study. *BMJ* **2018**, *362*, k2817. DOI: 10.1136/bmj.k2817.

(7) Chung, W.; Jo, C.; Chung, W. J.; Kim, D. J. Liver cirrhosis and cancer: comparison of mortality. *Hepatol Int* **2018**, *12* (3), 269-276. DOI: 10.1007/s12072-018-9850-5.

(8) D'Amico, G.; Garcia-Tsao, G.; Pagliaro, L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *J Hepatol* **2006**, *44* (1), 217-231. DOI: 10.1016/j.jhep.2005.10.013.

(9) Iwakiri, Y. Pathophysiology of portal hypertension. *Clin Liver Dis* **2014**, *18* (2), 281-291. DOI: 10.1016/j.cld.2013.12.001.

(10) Iwakiri, Y.; Groszmann, R. J. The hyperdynamic circulation of chronic liver diseases: from the patient to the molecule. *Hepatology* **2006**, *43* (2 Suppl 1), S121-131. DOI: 10.1002/hep.20993.

(11) Bosch, J. Vascular deterioration in cirrhosis: the big picture. *J Clin Gastroenterol* **2007**, *41 Suppl 3*, S247-253. DOI: 10.1097/MCG.0b013e3181572357.

(12) Cheemerla, S.; Balakrishnan, M. Global Epidemiology of Chronic Liver Disease. *Clin Liver Dis (Hoboken)* **2021**, *17* (5), 365-370. DOI: 10.1002/cld.1061.

(13) Shetty, A.; Jun Yum, J.; Saab, S. The Gastroenterologist's Guide to Preventive Management of Compensated Cirrhosis. *Gastroenterol Hepatol (N Y)* **2019**, *15* (8), 423-430. From NLM PubMed-not-MEDLINE.

(14) European Association for the Study of the Liver. Electronic address, e. e. e.; European Association for the Study of the, L. EASL Clinical Practice Guidelines for the management of patients with decompensated cirrhosis. *J Hepatol* **2018**, *69* (2), 406-460. DOI: 10.1016/j.jhep.2018.03.024 From NLM Medline.

(15) Shan, L.; Wang, F.; Zhai, D.; Meng, X.; Liu, J.; Lv, X. New Drugs for Hepatic Fibrosis. *Front Pharmacol* **2022**, *13*, 874408. DOI: 10.3389/fphar.2022.874408 From NLM PubMed-not-MEDLINE.

(16) Roehlen, N.; Crouchet, E.; Baumert, T. F. Liver Fibrosis: Mechanistic Concepts and Therapeutic Perspectives. *Cells* **2020**, *9* (4). DOI: 10.3390/cells9040875 From NLM Medline.

(17) Ripoll, C.; Groszmann, R.; Garcia-Tsao, G.; Grace, N.; Burroughs, A.; Planas, R.; Escorsell, A.; Garcia-Pagan, J. C.; Makuch, R.; Patch, D.; et al. Hepatic venous pressure gradient predicts clinical decompensation in patients with compensated cirrhosis. *Gastroenterology* **2007**, *133* (2), 481-488. DOI: 10.1053/j.gastro.2007.05.024 From NLM.

(18) Kamath, P. S.; Wiesner, R. H.; Malinchoc, M.; Kremers, W.; Therneau, T. M.; Kosberg, C. L.; D'Amico, G.; Dickson, E. R.; Kim, W. R. A model to predict survival in patients with end-stage liver disease. *Hepatology* **2001**, *33* (2), 464-470. DOI: 10.1053/jhep.2001.22172 From NLM.

(19) Malinchoc, M.; Kamath, P. S.; Gordon, F. D.; Peine, C. J.; Rank, J.; ter Borg, P. C. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology* **2000**, *31* (4), 864-871. DOI: 10.1053/he.2000.5852 From NLM.

(20) Smith, L. M.; Kelleher, N. L.; Consortium for Top Down, P. Proteoform: a single term describing protein complexity. *Nat Methods* **2013**, *10* (3), 186-187. DOI: 10.1038/nmeth.2369.

(21) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; et al. How many human proteoforms are there? *Nat Chem Biol* **2018**, *14* (3), 206-214. DOI: 10.1038/nchembio.2576.

(22) Melani, R. D.; Gerbasi, V. R.; Anderson, L. C.; Sikora, J. W.; Toby, T. K.; Hutton, J. E.; Butcher, D. S.; Negrao, F.; Seckler, H. S.; Srzentic, K.; et al. The Blood Proteoform Atlas: A reference map of proteoforms in human hematopoietic cells. *Science* **2022**, *375* (6579), 411-418. DOI: 10.1126/science.aaz5284.

(23) de Franchis, R.; Bosch, J.; Garcia-Tsao, G.; Reiberger, T.; Ripoll, C.; Baveno, V. I. I. F. Baveno VII - Renewing consensus in portal hypertension. *J Hepatol* **2022**, *76* (4), 959-974. DOI: 10.1016/j.jhep.2021.12.022.

(24) Garcia-Tsao, G.; Abraldes, J. G.; Berzigotti, A.; Bosch, J. Portal hypertensive bleeding in cirrhosis: Risk stratification, diagnosis, and management: 2016 practice guidance by the American Association for the study of liver diseases. *Hepatology* **2017**, *65* (1), 310-335. DOI: 10.1002/hep.28906.

(25) Kaplan, D. E.; Ripoll, C.; Thiele, M.; Fortune, B. E.; Simonetto, D. A.; Garcia-Tsao, G.; Bosch, J. AASLD Practice Guidance on risk stratification and management of portal hypertension and varices in cirrhosis. *Hepatology* **2023**. DOI: 10.1097/HEP.0000000000000647.

(26) Takemori, A.; Butcher, D. S.; Harman, V. M.; Brownridge, P.; Shima, K.; Higo, D.; Ishizaki, J.; Hasegawa, H.; Suzuki, J.; Yamashita, M.; et al. PEPPI-MS: Polyacrylamide-Gel-Based Prefractionation for Analysis of Intact Proteoforms and Protein Complexes by Mass Spectrometry. *J Proteome Res* **2020**, *19* (9), 3779-3791. DOI: 10.1021/acs.jproteome.0c00303.

(27) Wessel, D.; Flugge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **1984**, *138* (1), 141-143. DOI: 10.1016/0003-2697(84)90782-6.

(28) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Shams, D. P.; Thomas, P. M.; Kelleher, N. L. Accurate Estimation of Context-Dependent False Discovery Rates in Top-Down Proteomics. *Mol Cell Proteomics* **2019**, *18* (4), 796-805. DOI: 10.1074/mcp.RA118.000993.

(29) Kaufman, L. a. R., P.J. Partitioning Around Medoids (Program PAM). In *Finding Groups in Data*, 1990; pp 68-125.

(30) Mangaraj, M.; Nanda, R.; Panda, S. Apolipoprotein A-I: A Molecule of Diverse Function. *Indian J Clin Biochem* **2016**, *31* (3), 253-259. DOI: 10.1007/s12291-015-0513-1.

(31) Bhale, A. S.; Venkataraman, K. Leveraging knowledge of HDLs major protein ApoA1: Structure, function, mutations, and potential therapeutics. *Biomed Pharmacother* **2022**, *154*, 113634. DOI: 10.1016/j.biopha.2022.113634.

(32) Poynard, T.; Abella, A.; Pignon, J. P.; Naveau, S.; Leluc, R.; Chaput, J. C. Apolipoprotein AI and alcoholic liver disease. *Hepatology* **1986**, *6* (6), 1391-1395. DOI: 10.1002/hep.1840060628.

(33) Nuno-Gonzalez, P.; Ruiz-Madrigal, B.; Bastidas-Ramirez, B. E.; Martinez-Lopez, E.; Segura, J. E.; Panduro, A. Expression of apolipoprotein AI mRNA in peripheral white blood cells of patients with alcoholic liver disease. *Biochim Biophys Acta* **2005**, *1740* (3), 350-356. DOI: 10.1016/j.bbadis.2004.11.003.

(34) Trieb, M.; Rainer, F.; Stadlbauer, V.; Douschan, P.; Horvath, A.; Binder, L.; Trakaki, A.; Knuplez, E.; Scharnagl, H.; Stojakovic, T.; et al. HDL-related biomarkers are robust predictors of survival in patients with chronic liver failure. *J Hepatol* **2020**, *73* (1), 113-120. DOI: 10.1016/j.jhep.2020.01.026.

(35) Tsai, M. H.; Peng, Y. S.; Chen, Y. C.; Lien, J. M.; Tian, Y. C.; Fang, J. T.; Weng, H. H.; Chen, P. C.; Yang, C. W.; Wu, C. S. Low serum concentration of apolipoprotein A-I is an indicator of poor prognosis in cirrhotic patients with severe sepsis. *J Hepatol* **2009**, *50* (5), 906-915. DOI: 10.1016/j.jhep.2008.12.024.

(36) Gurbuz, B.; Guldiken, N.; Reuken, P.; Fu, L.; Remih, K.; Preisinger, C.; Bruha, R.; Lenicek, M.; Petrtyl, J.; Reissing, J.; et al. Biomarkers of hepatocellular synthesis in patients with decompensated cirrhosis. *Hepatol Int* **2023**, *17* (3), 698-708. DOI: 10.1007/s12072-022-10473-x.

(37) Law, S. W.; Gray, G.; Brewer, H. B., Jr. cDNA cloning of human apoA-I: amino acid sequence of preproapoA-I. *Biochem Biophys Res Commun* **1983**, *112* (1), 257-264. DOI: 10.1016/0006-291x(83)91824-7.

(38) Fadaei, R.; Davies, S. S. Oxidative modification of HDL by lipid aldehydes impacts HDL function. *Arch Biochem Biophys* **2022**, *730*, 109397. DOI: 10.1016/j.abb.2022.109397.

(39) Shao, B.; Cavigiolio, G.; Brot, N.; Oda, M. N.; Heinecke, J. W. Methionine oxidation impairs reverse cholesterol transport by apolipoprotein A-I. *Proc Natl Acad Sci U S A* **2008**, *105* (34), 12224-12229. DOI: 10.1073/pnas.0802025105.

(40) Rossi, E.; Adams, L.; Prins, A.; Bulsara, M.; de Boer, B.; Garas, G.; MacQuillan, G.; Speers, D.; Jeffrey, G. Validation of the FibroTest biochemical markers score in assessing liver fibrosis in hepatitis C patients. *Clin Chem* **2003**, *49* (3), 450-454. DOI: 10.1373/49.3.450.

(41) Munteanu, M.; Pais, R.; Peta, V.; Deckmyn, O.; Moussalli, J.; Ngo, Y.; Rudler, M.; Lebray, P.; Charlotte, F.; Thibault, V.; et al. Long-term prognostic value of the FibroTest in patients with non-alcoholic fatty liver disease, compared to chronic hepatitis C, B, and alcoholic liver disease. *Aliment Pharmacol Ther* **2018**, *48* (10), 1117-1127. DOI: 10.1111/apt.14990.

(42) Pompach, P.; Brnakova, Z.; Sanda, M.; Wu, J.; Edwards, N.; Goldman, R. Site-specific glycoforms of haptoglobin in liver cirrhosis and hepatocellular carcinoma. *Mol Cell Proteomics* **2013**, *12* (5), 1281-1293. DOI: 10.1074/mcp.M112.023259.

(43) Zhu, J.; Lin, Z.; Wu, J.; Yin, H.; Dai, J.; Feng, Z.; Marrero, J.; Lubman, D. M. Analysis of serum haptoglobin fucosylation in hepatocellular carcinoma and liver cirrhosis of different etiologies. *J Proteome Res* **2014**, *13* (6), 2986-2997. DOI: 10.1021/pr500128t.

(44) Zhang, S.; Shang, S.; Li, W.; Qin, X.; Liu, Y. Insights on N-glycosylation of human haptoglobin and its association with cancers. *Glycobiology* **2016**, *26* (7), 684-692. DOI: 10.1093/glycob/cww016.

(45) Shu, H.; Zhang, L.; Chen, Y.; Guo, Y.; Li, L.; Chen, F.; Cao, Z.; Yan, G.; Lu, C.; Liu, C.; et al. Quantification of Intact O-Glycopeptides on Haptoglobin in Sera of Patients With Hepatocellular Carcinoma and Liver Cirrhosis. *Front Chem* **2021**, *9*, 705341. DOI: 10.3389/fchem.2021.705341.

(46) Gutierrez Reyes, C. D.; Huang, Y.; Atashi, M.; Zhang, J.; Zhu, J.; Liu, S.; Parikh, N. D.; Singal, A. G.; Dai, J.; Lubman, D. M.; et al. PRM-MS Quantitative Analysis of Isomeric N-Glycopeptides Derived from Human Serum Haptoglobin of Patients with Cirrhosis and Hepatocellular Carcinoma. *Metabolites* **2021**, *11* (8). DOI: 10.3390/metabo11080563.

(47) Lin, Y.; Zhu, J.; Zhang, J.; Dai, J.; Liu, S.; Arroyo, A.; Rose, M.; Singal, A. G.; Parikh, N. D.; Lubman, D. M. Glycopeptides with Sialyl Lewis Antigen in Serum Haptoglobin as Candidate Biomarkers for Nonalcoholic Steatohepatitis Hepatocellular Carcinoma Using a Higher-Energy Collision-Induced Dissociation Parallel Reaction Monitoring-Mass Spectrometry Method. *ACS Omega* **2022**, *7* (26), 22850-22860. DOI: 10.1021/acsomega.2c02600.

(48) Naryzny, S. N.; Legina, O. K. Haptoglobin as a Biomarker. *Biochem Mosc Suppl B Biomed Chem* **2021**, *15* (3), 184-198. DOI: 10.1134/S1990750821030069.

(49) Islam, R.; Kundu, S.; Jha, S. B.; Rivera, A. P.; Flores Monar, G. V.; Islam, H.; Puttagunta, S. M.; Sange, I. Cirrhosis and Coagulopathy: Mechanisms of Hemostasis Changes in Liver Failure and Their Management. *Cureus* **2022**, *14* (4), e23785. DOI: 10.7759/cureus.23785.

(50) Francis, J. L.; Armstrong, D. J. Acquired dysfibrinogenaemia in liver disease. *J Clin Pathol* **1982**, *35* (6), 667-672. DOI: 10.1136/jcp.35.6.667.

(51) Rizzo, K.; Vella, K.; Zammit, D.; Gatt, P.; Grima, C.; Inguanez, M. B.; Gerada, J.; Ellul, P.; Vassallo, M.; Azzopardi, N.; et al. Fibrinogen measurement in liver disease: validation of the functional fibrinogen thromboelastography assay and a novel mathematical predictive model. *Blood Transfus* **2019**, *17* (3), 237-246. DOI: 10.2450/2018.0105-18.

(52) Martinez, J.; Palascak, J. E.; Kwasniak, D. Abnormal sialic acid content of the dysfibrinogenemia associated with liver disease. *J Clin Invest* **1978**, *61* (2), 535-538. DOI: 10.1172/JCI108964.

(53) de Vries, J. J.; Snoek, C. J. M.; Rijken, D. C.; de Maat, M. P. M. Effects of Post-Translational Modifications of Fibrinogen on Clot Formation, Clot Structure, and Fibrinolysis: A Systematic Review. *Arterioscler Thromb Vasc Biol* **2020**, *40* (3), 554-569. DOI: 10.1161/ATVBAHA.119.313626.

(54) Weisel, J. W.; Litvinov, R. I. Fibrin Formation, Structure and Properties. *Subcell Biochem* **2017**, *82*, 405-456. DOI: 10.1007/978-3-319-49674-0_13.

(55) Li, Y.; Qi, X.; Li, H.; Dai, J.; Deng, H.; Li, J.; Peng, Y.; Liu, X.; Sun, X.; Guo, X. D-dimer level for predicting the in-hospital mortality in liver cirrhosis: A retrospective study. *Exp Ther Med* **2017**, *13* (1), 285-289. DOI: 10.3892/etm.2016.3930.

(56) Spadaro, A.; Tortorella, V.; Morace, C.; Fortiguerra, A.; Composto, P.; Bonfiglio, C.; Alibrandi, A.; Luigiano, C.; De Caro, G.; Ajello, A.; et al. High circulating D-dimers are associated with ascites and hepatocellular carcinoma in liver cirrhosis. *World J Gastroenterol* **2008**, *14* (10), 1549-1552. DOI: 10.3748/wjg.14.1549.

(57) Cioni, G.; Cristani, A.; Mussini, C.; Grandi, S.; Pentore, R.; Zeneroli, M. L.; Tizzanini, W.; Zagni, G.; Ventura, E. Incidence and clinical significance of elevated fibrin(ogen) degradation product and/or D-dimer levels in liver cirrhosis patients. *Ital J Gastroenterol* **1990**, *22* (2), 70-74.

(58) Gram, J.; Duscha, H.; Zurborn, K. H.; Bruhn, H. D. Increased levels of fibrinolysis reaction products (D-dimer) in patients with decompensated alcoholic liver cirrhosis. *Scand J Gastroenterol* **1991**, *26* (11), 1173-1178. DOI: 10.3109/00365529108998610.

(59) Saray, A.; Mesihovic, R.; Gornjakovic, S.; Vanis, N.; Mehmedovic, A.; Nahodovic, K.; Glavas, S.; Papovic, V. Association between high D-dimer plasma levels and ascites in patients with liver cirrhosis. *Med Arch* **2012**, *66* (6), 372-374. DOI: 10.5455/medarh.2012.66.372-374.

Tables

Table 1. Characteristics of the patient cohort. pHTN= portal hypertension, AIH= autoimmune hepatitis, PBC= primary biliary cholangitis, ETOH= alcohol-associated liver disease, HBV= hepatitis B virus, HCV= hepatitis C virus, MASH= metabolic dysfunction associated steatohepatitis, BMI= body mass index, MELD-Na= Model for End-Stage Liver Disease with Sodium.

|	Total (n=30)	Compensated (n=9)	Compensated w/ pHTN (n=10)	Decompensated (n=11)	p-value
Age, mean (SD)	59.73 (12.34)	63.67 (11.18)	55.51 (10.54)	61.89 (14.18)	0.314
Female, *n (%)*	12 (40)	5 (55.6)	4 (40.0)	3 (27.3)	0.612
Race/Ethnicity, *n (%)*					0.206
Asian	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Black or African American	3 (10)	2 (22.2)	1 (10.0)	0 (0.0)	
Hispanic or Latino	10 (33.33)	4 (44.4)	3 (30.0)	3 (27.3)	
other	2 (6.67	0 (0.0)	0 (0.0)	2 (18.2)	
White	17 (56.67)	3 (33.3)	6 (60.0)	8 (72.7)	
Etiology, *n (%)*					<0.01
AIH	3 (10)	0 (0.0)	2 (20.0)	1 (9.1)	
PBC	1 (3.33)	0 (0.0)	0 (0.0)	1 (9.1)	
ETOH	18 (60.0)	5 (55.6)	5 (50.0)	8 (72.7)	
HBV	1 (3.33)	1 (11.1)	0 (0.0)	0 (0.0)	
HCV	8 (26.67)	4 (44.4)	3 (30.0)	1 (9.1)	
MASH	4 (13.33)	1 (11.1)	1 (10.0)	2 (18.2)	
BMI, mean (SD)	29.14 (5.55)	31.14 (6.49)	28.63 (3.21)	27.99 (6.42)	0.336
MELD-Na, mean (median [IQR])	11.47 (4.62)	9.89 (3.18)	10.00 (3.94)	14.09 (5.30)	0.162
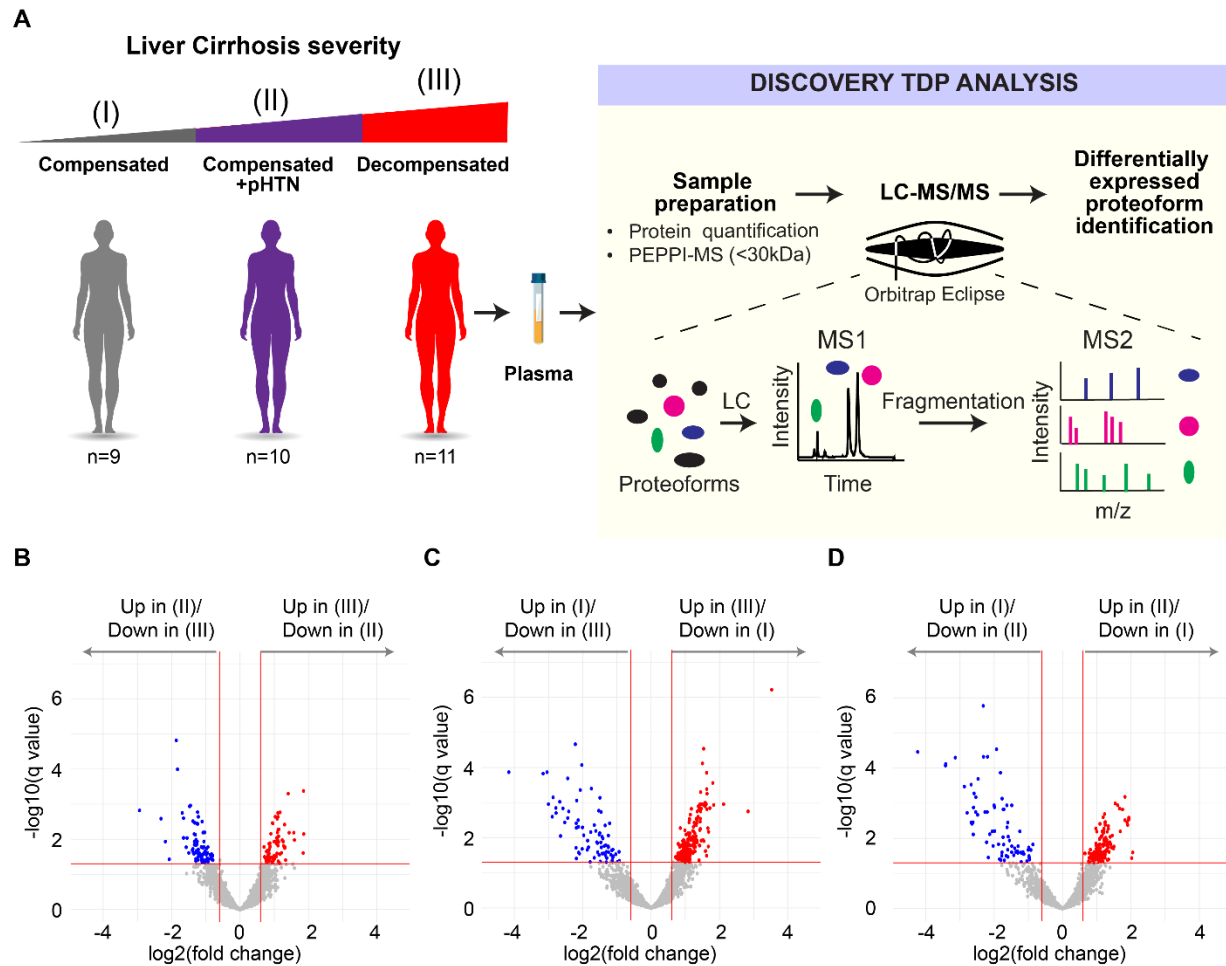
Figures

Figure 1. Workflow used to identify differentially expressed proteoforms (DEPs) in patients with cirrhosis. A) Plasma was extracted from cirrhosis patients with three stages of disease— compensated (I), compensated with portal hypertension (II), and decompensated cirrhosis (III). 30kDa fractions of plasma were then subjected to Top-Down Proteomics (TDP) to identify DEPs across disease stages. **B-D)** Volcano plots showing differentially expressed proteoforms (DEPs) in the plasma from patients with **B)** decompensated cirrhosis (III) vs compensated cirrhosis with portal hypertension (II), **C)** decompensated cirrhosis (III) vs compensated cirrhosis (I), and **D)**

30

compensated cirrhosis with portal hypertension (II) vs compensated cirrhosis (I). A false discovery rate (FDR) threshold of 5% for differential expression was used.

Figure 2. Proteoforms found to be differentially expressed across stages of cirrhosis. A) Cluster distribution of differentially expressed proteoforms (DEPs) and currently used clinical markers of disease severity. Clinical markers highlighted in blue are significantly different between stages. Cluster 1 (C1) included DEPs upregulated in Stage III disease, C2 included DEPs upregulated in stage II disease, and C3 included DEPs upregulated in Stage I disease. **B)** Proteins from which DEPs derived are shown within each cluster. The number of proteoforms of a given

protein is shown in parentheses beside the protein name. For example, in C1 the fibrinogen alpha chain protein had 15 DEPs identified. **C)** Biological process enrichment analyses on proteins identified in each cluster. The level of significance is represented by the color of the bar with scale while the gene ratio is indicated by the diameter of the circle.
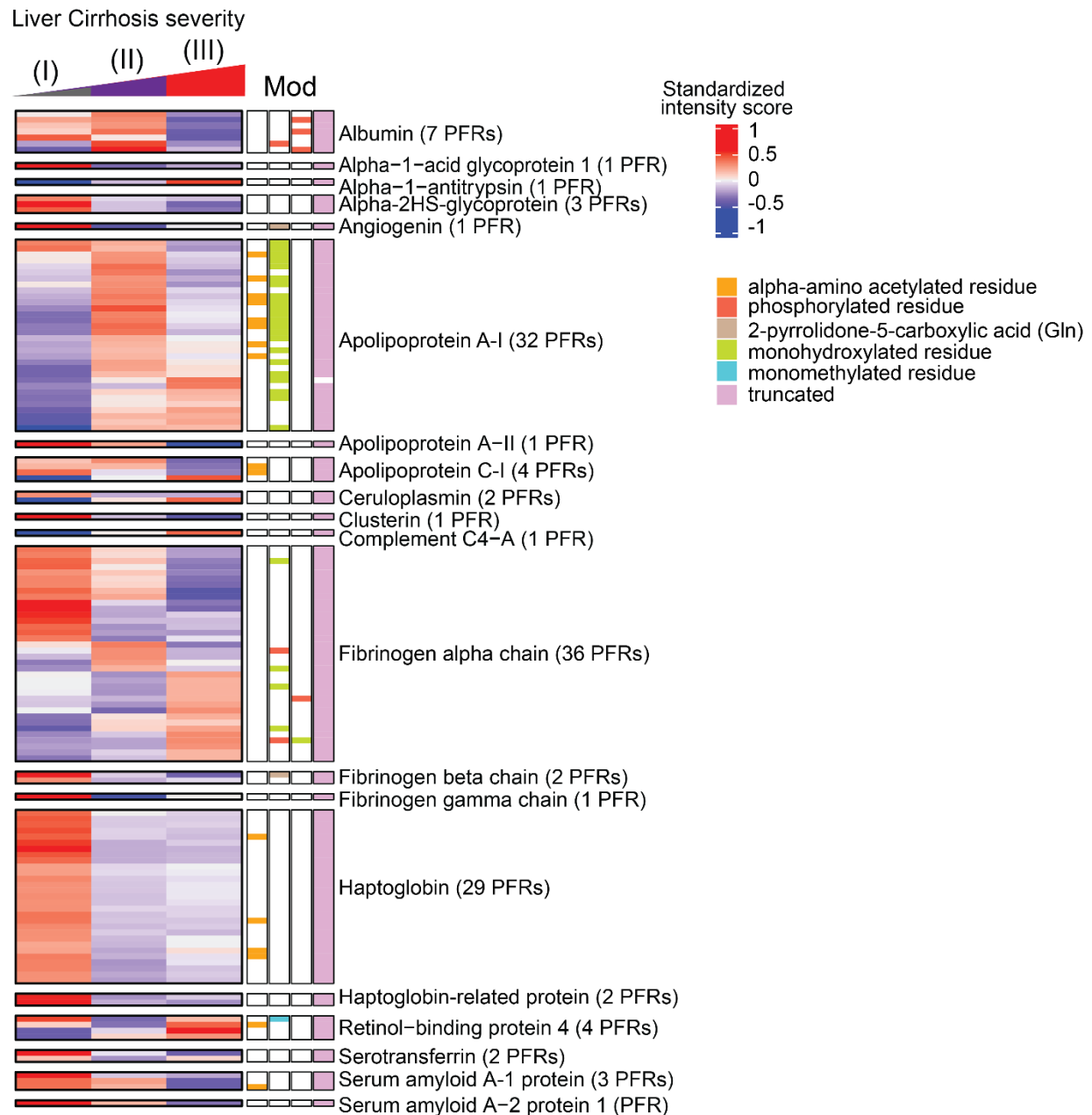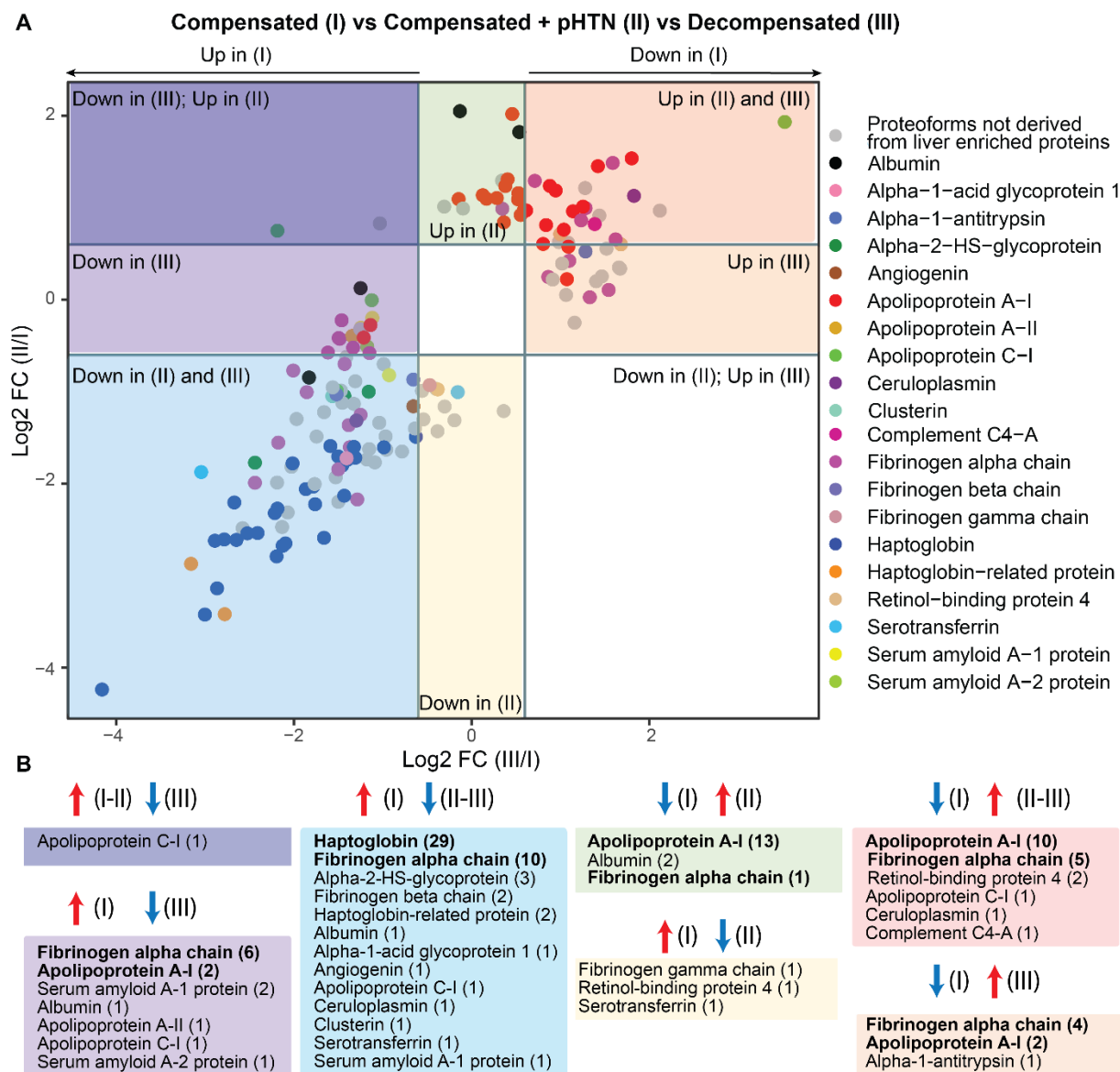

Figure 3. Heatmaps of quantified proteoforms of proteins enriched in the liver at transcriptional level. Associated proteoform modifications are shown and defined by the

figure legend. The proteins and number of identified proteoforms derived from each protein are shown to figure right. Abbreviations: PFR (proteoform), MOD (Modification).

Figure 4. Display of proteoforms stratified by expression in each stage of cirrhosis A) Differentially expressed proteoforms (DEPs) in Stages II and III disease with Stage I disease (compensated) used as the reference. Proteoforms upregulated in Stage II and III (red square), upregulated only in Stage III (orange square), downregulated in Stages II and III (blue square), downregulated in Stage III only (light purple square), downregulated in Stage III but upregulated in Stage II (dark purple square), upregulated only in Stage II (green square), and downregulated

only in Stage II (yellow square). Proteoforms deriving from proteins enriched in liver are highlighted and grouped by color according to the figure legend. **B)** Profiles of proteoforms upregulated in Stages I and II and downregulated in III (dark purple), upregulated in I and downregulated in III (light purple), upregulated in I and downregulated in II and II (light blue), downregulated in I and upregulated in II (green), upregulated in I and downregulated in II (yellow), downregulated in I and upregulated in II and III (red), and downregulated in I and upregulated in III (orange). The proteins from which DEPs derive and the relative number of proteoforms (in parenthesis) are shown. Proteins with the highest number of DEPs for each group are shown in bold.
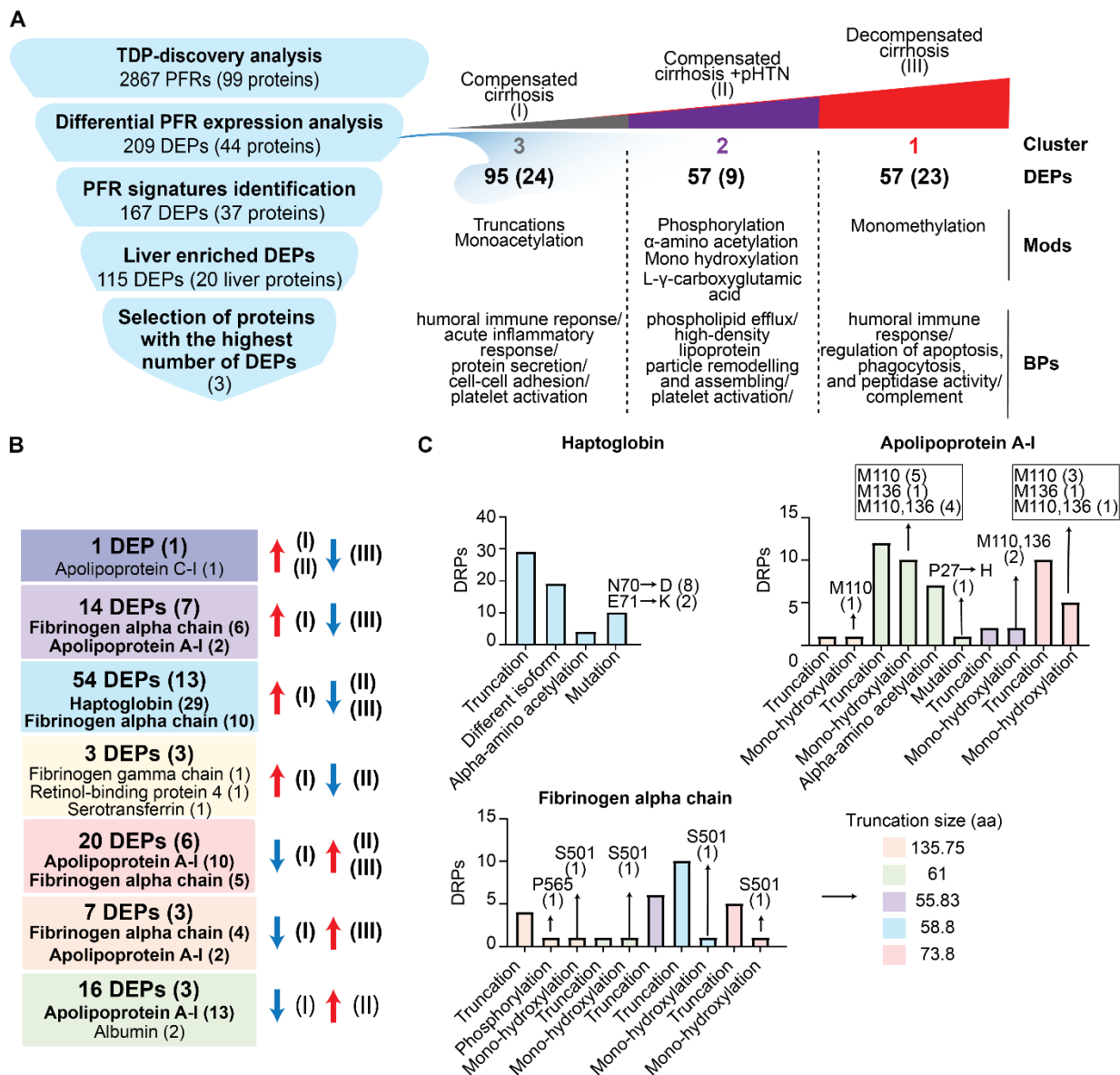
Figure 5. Summary of quantitative TDP analysis of plasma proteoforms in patients with cirrhosis. A) Strategy used to identify potential candidate proteoform biomarkers of liver cirrhosis and summary of the differential PFR expression analysis. Differentially expressed proteoforms (DEPs) were classified in clusters based on their upregulation in 1) compensated cirrhosis (Stage I- Cluster 3), 2) compensated cirrhosis with portal hypertension (Stage II- Cluster 2), and 3) decompensated cirrhosis (Stage III- Cluster 1). For each cluster, the number of upregulated DEPs with their relative proteins (in parenthesis), modifications, and biological processes are reported. **B)** Composition of the identified liver derived DEP signatures. Proteins from which the DEPs derive are shown in parenthesis. Proteins with the highest number of DEPs for each group are also reported and shown in bold. DEPs are grouped by color based on their upregulation or downregulation in disease stages as performed in **Figure 4**. For example, in the red group 20 DEPs deriving from 6 proteins were upregulated in Stages II and III of the disease as compared to Stage I disease. 10 of these DEPs derived from apolipoprotein A-I and 5 derived from fibrinogen alpha chain. **C)** DEPs relative to Haptoglobin, Apolipoprotein A-I, and Fibrinogen alpha chain, which are the 3 plasma proteins with the highest number of detected DEPs, are reported together with their modifications. Colors indicate the signature groups in B and C. Abbreviations: PFR (proteoform), DEP (Differentially expressed proteoform), Mod (Modification) BP (biological process), aa (amino acids).

Supplementary Figures
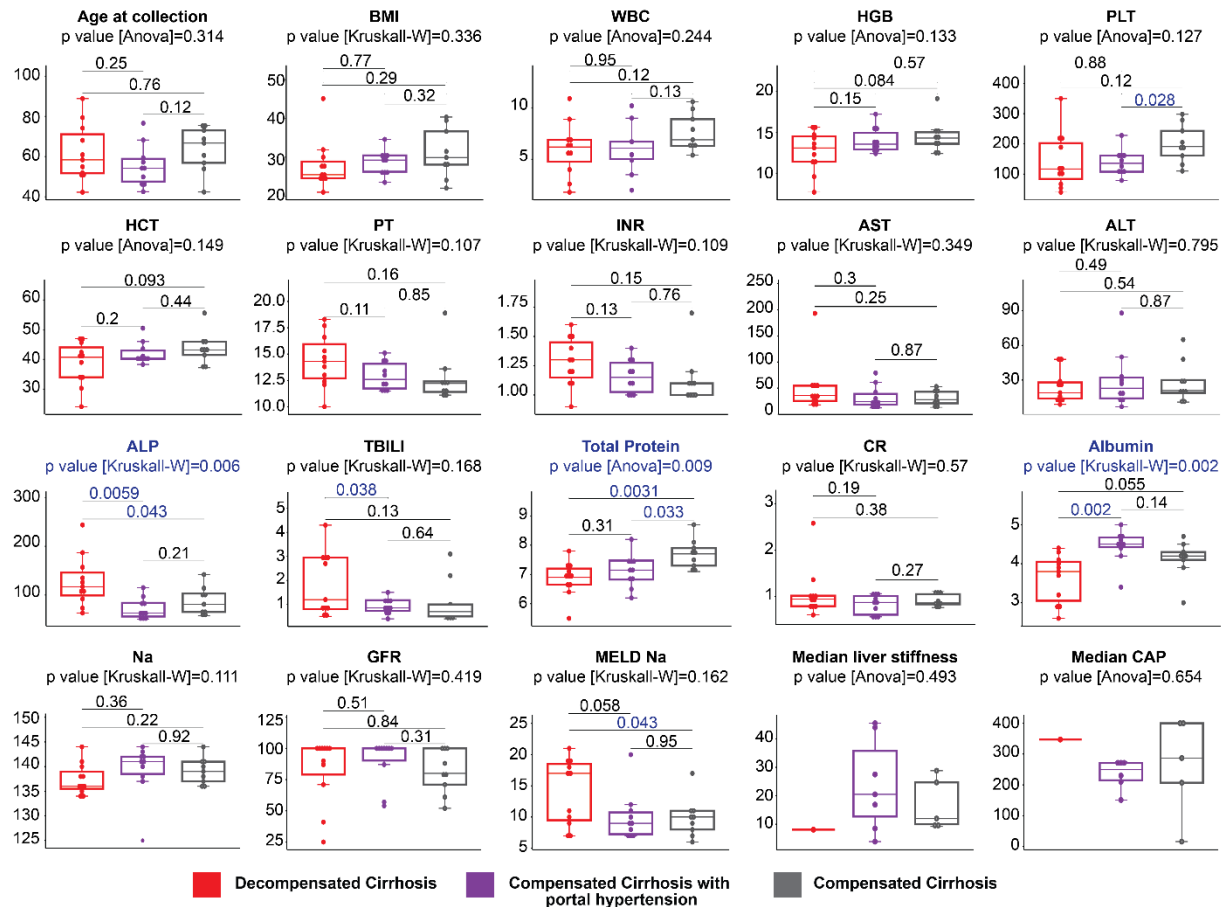
Figure S1. Clinical parameters of patients enrolled in this study. Clinical parameters shown in blue are statistically significant different among groups. Significance (p value <0.05) was calculated by Anova or Kruskall-Wallis (Kruskall-W) tests, depending on the data distribution determined by Shapiro test. Abbreviations: BMI (body mass index), WBC (white blood cell count), HGB (hemoglobin), PLT (platelets), HCT (hematocrit), PT (prothrombin time), INR (international normalized ratio), AST (aspartate aminotransferase), ALT (alanine aminotransferase), ALP (alkaline phosphatase), TBILI (total bilirubin), CR (serum creatinine), Na (Sodium), GFR (glomerular filtration rate), MELD-Na (Model for End Stage Liver Disease Sodium), CAP (controlled attenuation parameter).
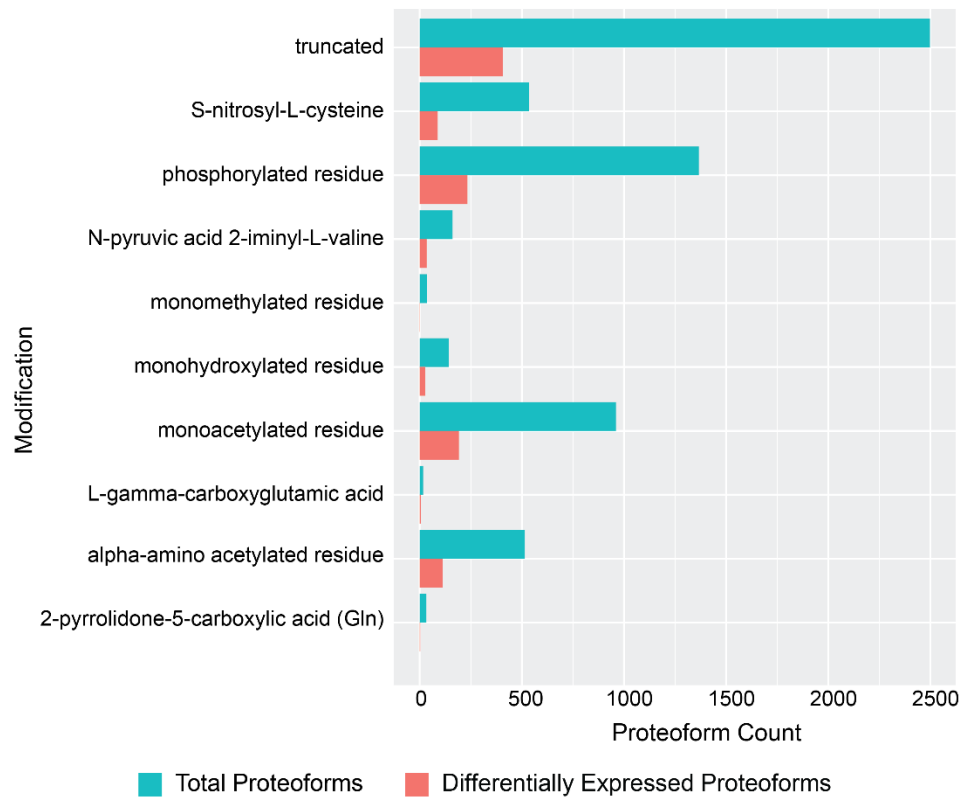
Figure S2. T **Common proteoform modifications detected in the TDP analysis.** The number of truncations and top-10 most common post-translational modifications identified in the total and differentially expressed proteoforms (DEPs) captured in the discovery LC-MS/MS TDP analysis.
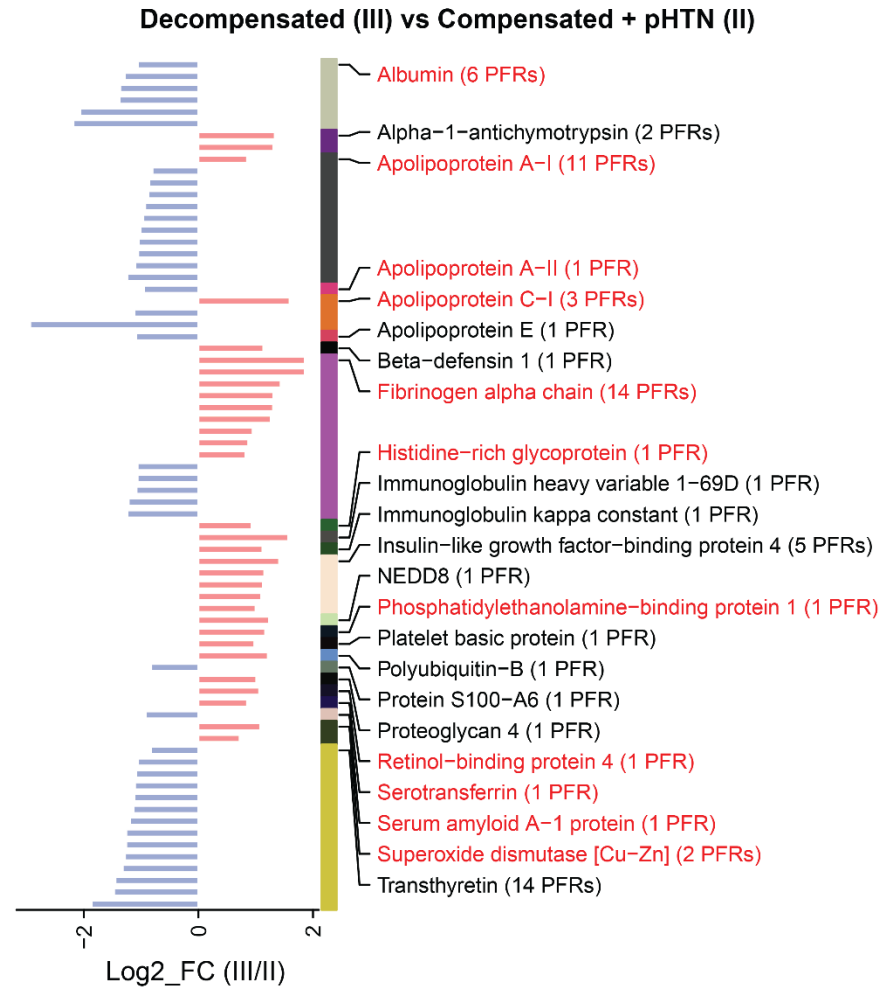
Figure S3. Differentially expressed proteoforms (DEPs) from decompensated (III) vs compensated with portal hypertension (+ pHTN) (II) patients. DEPs are grouped by proteins, and then ordered by fold change. The protein and number of DEPs derived from that protein are shown to figure right. For example, fibrinogen alpha chain had 14 DEPs identified. Proteins highlighted in red are enriched in liver at a transcriptional level. Abbreviations: PFR (proteoform).

Figure S4. . Differentially expressed proteoforms (DEPs) from decompensated (III) vs compensated (I) patients. DEPs are grouped by proteins, and then ordered by fold change. The protein and number of DEPs derived from that protein are shown to figure right. For example, haptoglobin related protein had 2 DEPs identified. Proteins highlighted in red are enriched in liver at a transcriptional level. Abbreviations: PFR (proteoform).
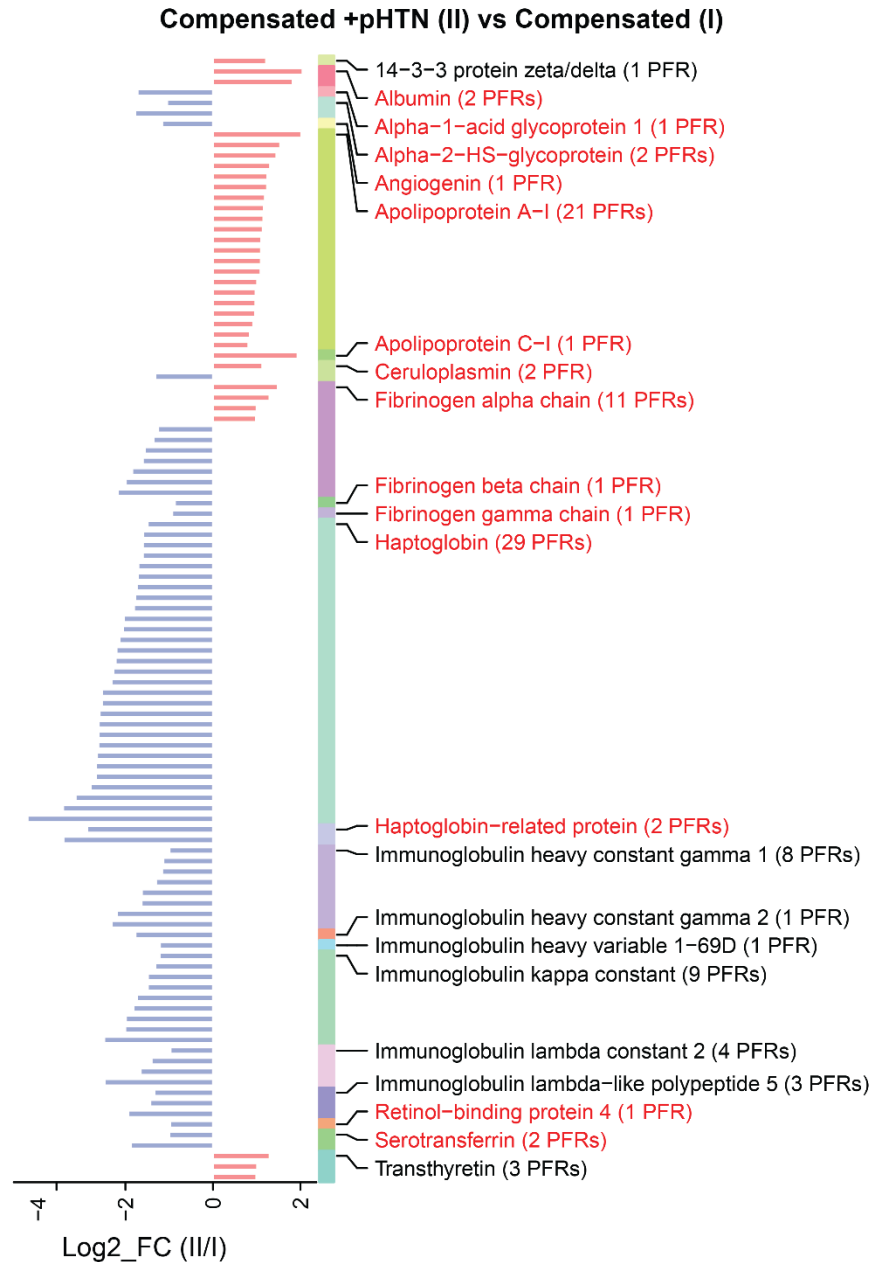
Figure S5. Differentially expressed proteoforms (DEPs) from compensated with portal hypertension (+ pHTN) (II) vs compensated (I) patients. DEPs are grouped by proteins, and then ordered by fold change. The protein and number of DEPs derived from that protein are shown to figure right. For example, haptoglobin related protein had 2 DEPs identified. Proteins highlighted in red are enriched in liver at a transcriptional level. Abbreviations: PFR (proteoform).

Proteoform identification in liver cirrhosis progression
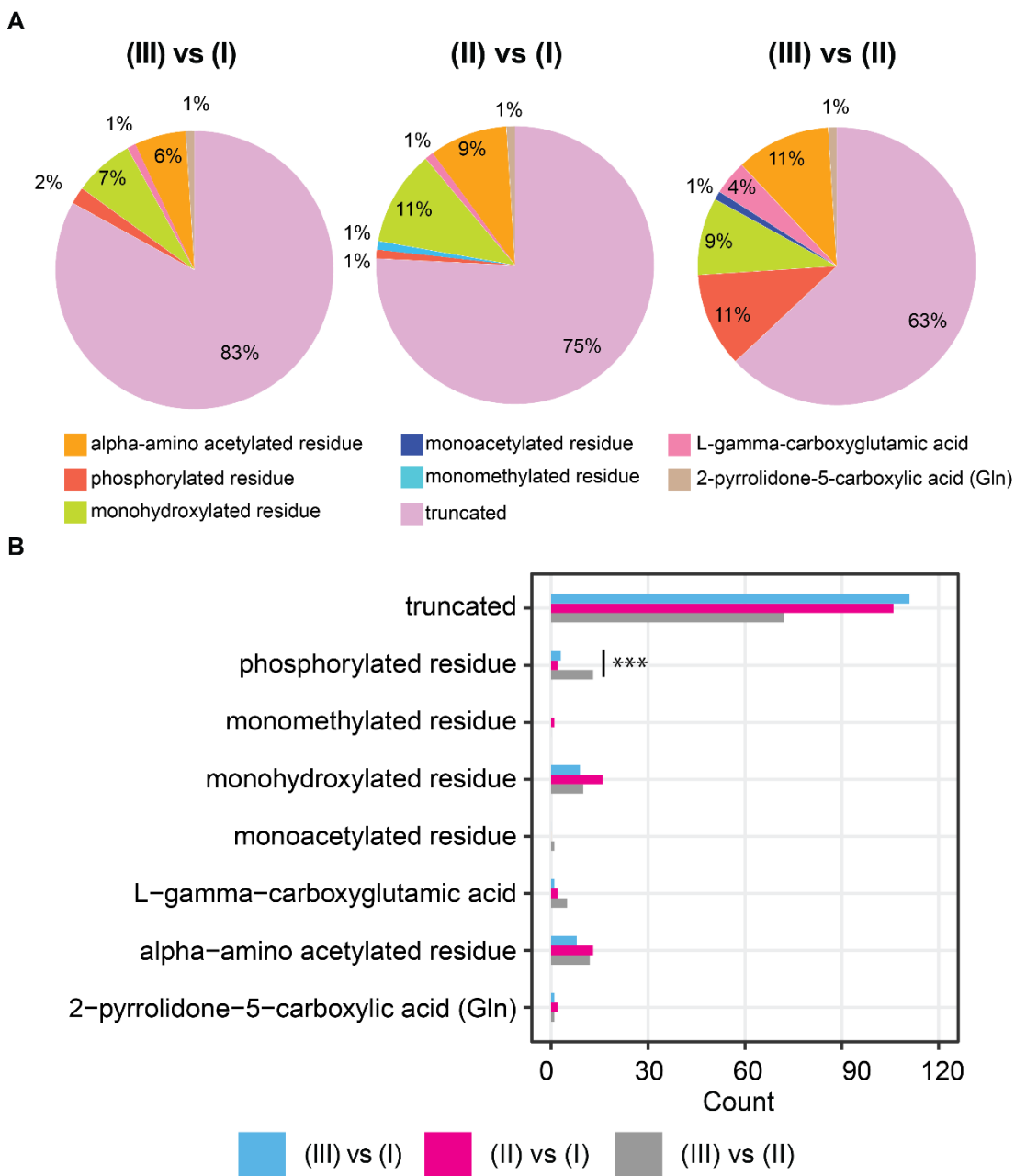
Figure S6. Proteoform modifications in pairwise comparisons of cirrhosis stages. A) Percentages and **B)** absolute number of differentially expressed proteoform (DEP) modifications identified in pairwise comparisons of Stage III vs I, II vs I, and III vs II. For example, 11% of

41

proteoforms differentially expressed between Stages III and II were phosphorylated residues (red portion of pie chart). Statistical significance was calculated with the Fisher exact test (adj. *p*-values: ***<0.001, **<0.01, *<0.05).
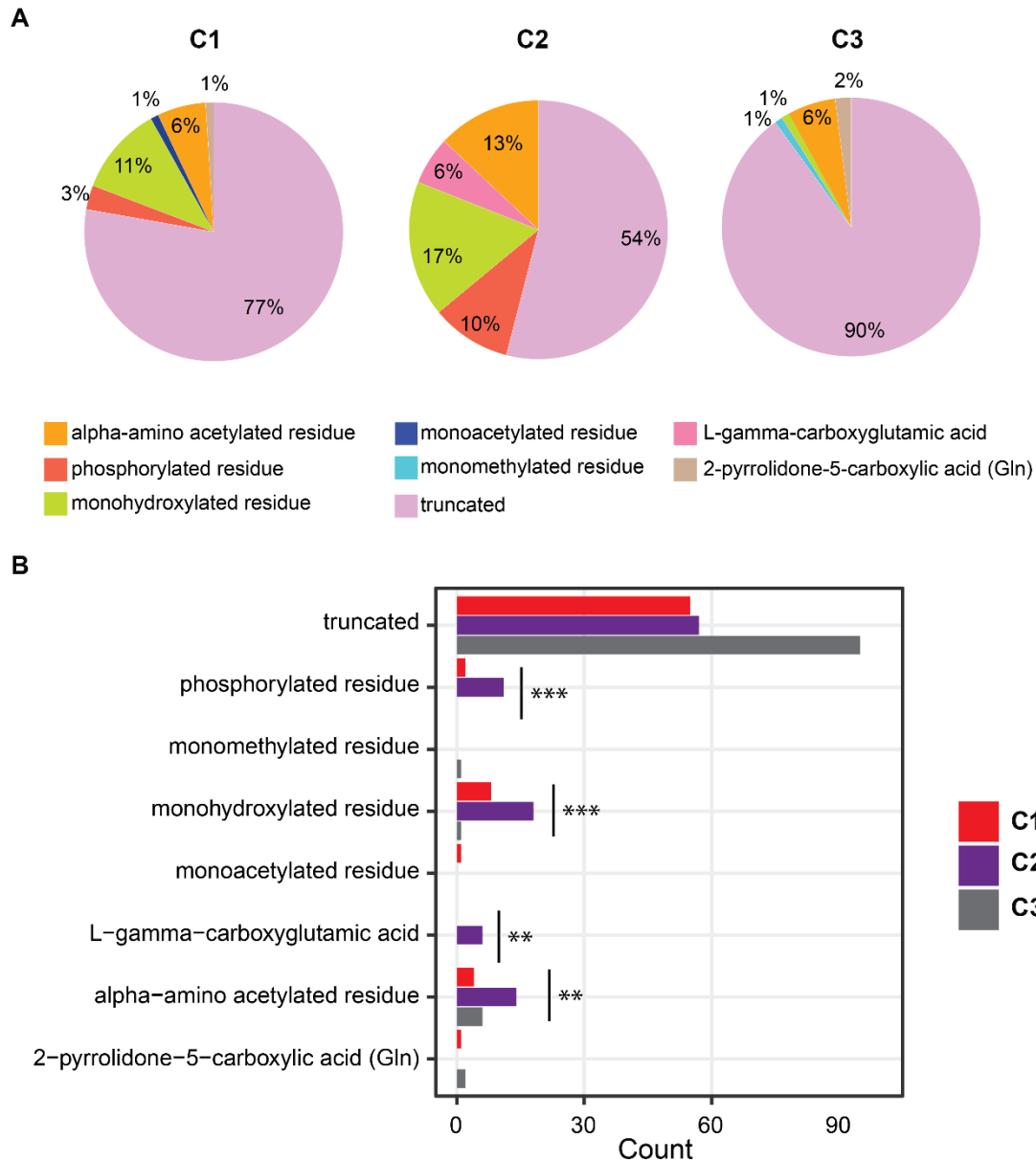

Figure S7. **Proteoform modifications compared across clusters.** **A)** Percentages and **B)** Absolute number of differentially expressed proteoform (DEP) modifications identified in clusters 1 (C1), 2 (C2), and C3. For example, 90% of modifications to DEPs that clustered into C3 were

identified as truncations (light purple portion of the pie chart). Statistical significance was calculated with the Fisher exact test (adj. *p*-values: ***<0.001, **<0.01, *<0.05).

	Compensated cirrhosis (I)	Compensated cirrhosis + pHTN (II)	Decompensated cirrhosis (III)	
	3	**2**	**1**	**Cluster**
Fibrinogen alpha chain (36 DEPs)	**14**	**7**	**15**	**DEPs**
Truncations	14	7	15	
PRF lenght	58	58.86	126	
Phosphorylation	-	1	3	**Mods**
Mono hydroxylation	1	1	3	
Apolipoprotein A-I (32 DEPs)	**-**	**22**	**10**	**DEPs**
Truncations		22	9	
α-amino acetylation		8	-	
Mono hydroxylation		25	6	**Mods**
Mutations		2	-	
P27-->H				
Haptoglobin (29 DEPs)	**29**	**-**	**-**	**DEPs**
Isoforms	2			
Truncations	29			
α-amino acetylation	4			
Mutations	10			**Mods**
E71-->K	2			
N70-->D	8			

Figure S8. **Most common, liver-enriched differentially expressed proteoforms (DEPs) identified in TDP analysis.** DEPs relative to Fibrinogen alpha chain, Apolipoprotein A-I, and Haptoglobin are reported together with their modifications in the 3 clusters established in **Figure 2**. Proteins are shown with the number of DEPs in parentheses. Abbreviations: DEP (Differentially expressed proteoform), Mod (Modification).
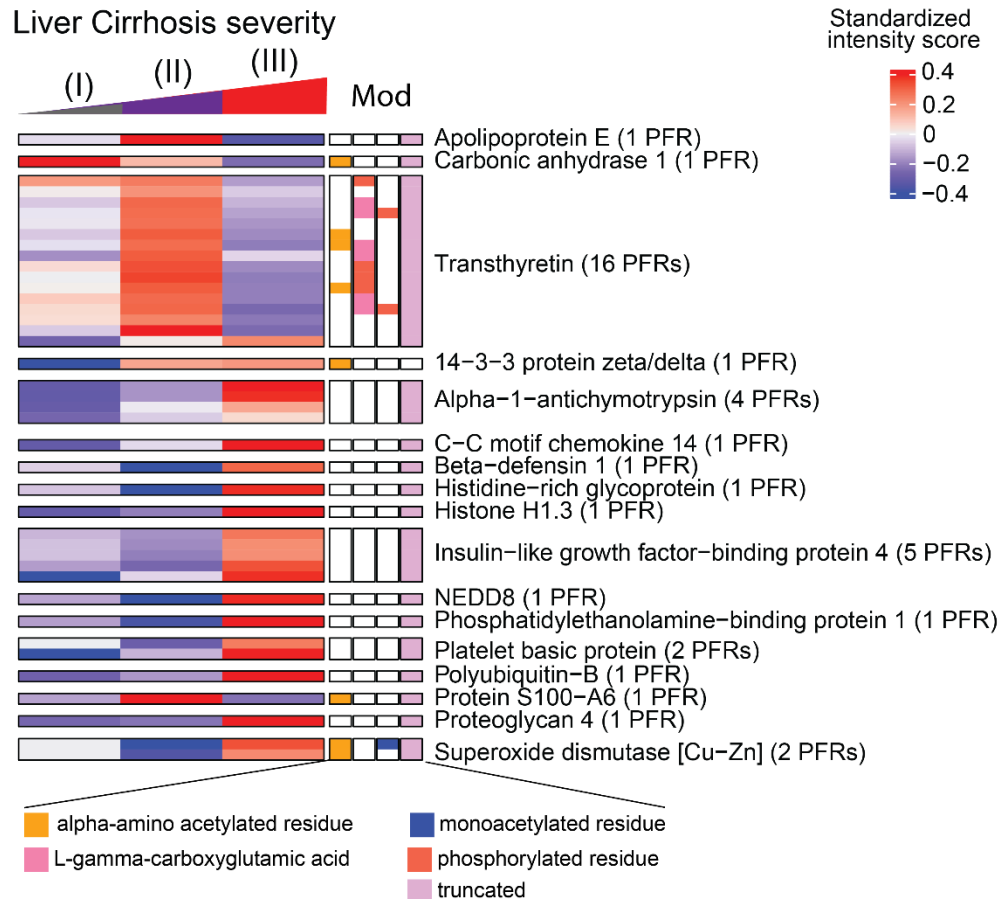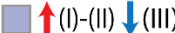
Figure S9. Heatmaps of quantified proteoforms of proteins not enriched in the liver at transcriptional level. Associated proteoform modifications are shown and define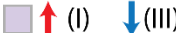d by the figure legend. The proteins and number of identified proteoforms derived from each protein are shown to figure right. Abbreviations: PFR (proteoform), Mod (modification).
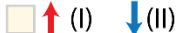
PFR	Protein	Uniprot ID	Sequence	Mutation	Phosphorylation	Mono hydroxylation	Half Cysteine	Methylation	Alpha amino acetylatation
16559	Apolipoprotein C-I	P02654	29-83						
7510185	Albumin	P02768-3	25-277		S-82, 89		C-77, 86		
7033240	Apolipoprotein A-I	P02647	67-267			M-110,136			
7052923	Apolipoprotein A-I	P02647	35-267			M-110,136			
47927	Apolipoprotein A-II	P02652	19-100						
38719	Apolipoprotein C-I	P02654	27-83						yes
18702	Fibrinogen alpha chain	P02671	542-601						
2383017	Fibrinogen alpha chain	P02671	538-604						
6720044	Fibrinogen alpha chain	P02671	561-604						
6720045	Fibrinogen alpha chain	P02671	549-604						
6720672	Fibrinogen alpha chain	P02671	539-604						
7654834	Fibrinogen alpha chain	P02671	557-603						
74009	Serum amyloid A-1 protein	P0DJI8	20-122						
5435552	Serum amyloid A-1 protein	P0DJI8	21-122	G90-->D					yes
215903	Serum amyloid A-2 protein	P0DJI9	20-122	R89-->H					
6700458	Fibrinogen gamma chain	P02679	281-436	M410-->V					
6542874	Retinol-binding protein 4	P02753	19-194	A73-->T				R-139	
215854	Serotransferrin		636-698						

■ ↑(I)-(II) ↓(III)   ■ ↑(I) ↓(III)   ☐ ↑(I) ↓(II)

Figure S10. Individual proteoforms upregulated in early-stage cirrhosis. Differentially expressed proteoforms (DEPs) significantly upregulated in stages (I) and (II) and downregulated in stage (III) (dark purple), upregulated in stage (I) and downregulated in stage (III) (light purple), upregulated in stage (I) and downregulated in stage (II) (yellow)**.** Each proteoform is shown with its unique proteoform number, protein name,Uniprot ID, amino acid sequence relative to the Uniprot ID, any relevant mutations, and presence or absence of post-translational modifications. Colors are based on the proteoform signatures created in **Figure 4**. Abbreviations: PFR (proteoform).

*Note that half Cysteine could be artifact during the identification process.

PFR	Protein	Uniprot ID	Sequence	Mutation	Mono hydroxylation	Half Cysteine	Alpha amino acetylatation	2-pyrrolidone-5-carboxylic acid
7233558	Albumin	P02768-3	25-325			C-77, 86, 325		
7493196	alpha-1acid glycoprotein 1	P02763	165-201					
18967	alpha-2HS glycoprotein	P02765	341 - 367					
56322	alpha-2HS glycoprotein	P02765	275 - 339					
74004	alpha-2HS glycoprotein	P02765	342 - 367					
5931629	Angiogenin	P03950	25-147	Q36-->L		C50,81,105,131		Q25
99695	Apolipoprotein C-I	P02654	29-83				yes	
7411348	Ceruloplasmin	P00450	20-58					
53707	Clusterin	P10909	414-449					
345982	Fibrinogen alpha chain	P02671	542-604		P-565			
6618516	Fibrinogen alpha chain	P02671	551-603					
6700442	Fibrinogen alpha chain	P02671	537-604					
6707159	Fibrinogen alpha chain	P02671	438-519			C-491		
6732079	Fibrinogen alpha chain	P02671	561-603					
6755154	Fibrinogen alpha chain	P02671	549-603					
6788241	Fibrinogen alpha chain	P02671	538-605					
7164381	Fibrinogen alpha chain	P02671	537-599					
7584008	Fibrinogen alpha chain	P02671	549-605					
7587565	Fibrinogen alpha chain	P02671	451-496					
73949	Haptoglobin	P00738-2	19-102					
73950	Haptoglobin	P00738-2	19-101					
73951	Haptoglobin	P00738-2	19-100					
73952	Haptoglobin	P00738-2	22-101					
73953	Haptoglobin	P00738-2	19-102				yes	
379773	Haptoglobin	P00738-1	104-160					
597885	Haptoglobin	P00738-2	18-101					
4927920	Haptoglobin	P00738-2	19-101			C-52, 86		
5055588	Haptoglobin	P00738-2	19-102	E71-->K			yes	
5055594	Haptoglobin	P00738-2	19-101				yes	
5269396	Haptoglobin	P00738-2	19-102	N70-->D		C-52		
5269397	Haptoglobin	P00738-2	19-101	N70-->D		C-90		
5269398	Haptoglobin	P00738-2	19-101	N70-->D		C-86		
5710193	Haptoglobin	P00738-2	19-101	N70-->D		C-52		
5865702	Haptoglobin	P00738-2	19-102	E71-->K		C-52,86	yes	
6372014	Haptoglobin	P00738-2	19-102	N70-->D		C-90		
6476920	Haptoglobin	P00738-1	78-160					
6720057	Haptoglobin	P00738-1	106-160					
6721265	Haptoglobin	P00738-1	78-161			C-86		
6889832	Haptoglobin	P00738-1	78-160			C-86		
7111729	Haptoglobin	P00738-2	22-101	N70-->D		C-90		
7488813	Haptoglobin	P00738-1	106-161					
7488879	Haptoglobin	P00738-2	17-101					
7591703	Haptoglobin	P00738-1	82-160			C-86,111		
7839836	Haptoglobin	P00738-2	22-101	N70-->D		C-52		
7839837	Haptoglobin	P00738-1	75-160			C-86		
7841882	Haptoglobin	P00738-2	19-100	N70-->D		C-52		
7841883	Haptoglobin	P00738-1	19-101			C-86		
7863012	Haptoglobin	P00738-1	97-160			C-111		

⬆ (I) ⬇ (II)-(III)

Figure S11. Individual proteoforms upregulated in Stage I disease. Differentially expressed proteoforms (DEPs) significantly upregulated in stage (I) and downregulated in stages (II) and (III) (blue). Each proteoform is shown with its unique proteoform number, protein name,Uniprot ID, amino acid sequence relative to the Uniprot ID, any relevant mutations, and presence or absence of post-translational modifications. Colors are based on the proteoform signatures created in **Figure 4**. Abbreviations: PFR (proteoform).

*Note that half Cysteine could be artifact during the identification process.

Proteoform identification in liver cirrhosis progression

PFR	Protein	Uniprot ID	Sequence	Mutation	Phosphorylation	Mono hydroxylation	Half Cysteine	Alpha amino acetylatation	
7047846	Albumin	P02768-1	25-157		S-82, 89		C-77		
7841902	Albumin	P02768-1	25-288		S-82				
56323	Apolipoprotein A-I	P02647	19-267					yes	
216392	Apolipoprotein A-I	P02647	25-267			M-110			
5052079	Apolipoprotein A-I	P02647	25-267			M-110		yes	
5052090	Apolipoprotein A-I	P02647	25-266	P27-->H				yes	
5508504	Apolipoprotein A-I	P02647	19-267			M-110,136		yes	↓(I) ↑(II)
6372372	Apolipoprotein A-I	P02647	23-267						
6468740	Apolipoprotein A-I	P02647	9-267			M-110			
6472353	Apolipoprotein A-I	P02647	18-267			M-110,136			
6478105	Apolipoprotein A-I	P02647	12-267			M-110			
6481626	Apolipoprotein A-I	P02647	25-266			M-110		yes	
6481628	Apolipoprotein A-I	P02647	25-266			M-110,136		yes	
6799836	Apolipoprotein A-I	P02647	9-267			M-110,136			
6981680	Apolipoprotein A-I	P02647	25-266			M-136		yes	
345984	Fibrinogen alpha chain	P02671	542-603			P-565			
5052085	Apolipoprotein A-I	P02647	25-267			M-136			
6472160	Apolipoprotein A-I	P02647	23-267			M-110,136			
6472193	Apolipoprotein A-I	P02647	5-267			M-110			
6563625	Apolipoprotein A-I	P02647	5-267						
6571467	Apolipoprotein A-I	P02647	30-267						
6686623	Apolipoprotein A-I	P02647	27-267						
6788235	Apolipoprotein A-I	P02647	6-267						
7013512	Apolipoprotein A-I	P02647	14-267						↓(I) ↑(II) (III)
7072437	Apolipoprotein A-I	P02647	29-267			M-110			
7497747	Apolipoprotein A-I	P02647	27-265			M-110			
4664283	Apolipoprotein C-I	P02654	13-83						
597908	Ceruloplasmin	P00450	910-1065						
74946	Complement C4-A	P0C0L4	680-755						
53596	Fibrinogen alpha chain	P02671	576-629						
74881	Fibrinogen alpha chain	P02671	600-629						
367042	Fibrinogen alpha chain	P02671	548-629			P-565			
6700460	Fibrinogen alpha chain	P02671	459-629						
7811927	Fibrinogen alpha chain	P02671	592-628						
6678845	Retinol-binding protein 4	P02753	19-201				C-192		
6686230	Retinol-binding protein 4	P02753	19-201				C-88		
6372540	Alpha-1-antitrtpsin	P001009	339-418						
6372550	Apolipoprotein A-I	P02647	21-267						
6481673	Apolipoprotein A-I	P02647	1-267			M-110			
597831	Fibrinogen alpha chain	P02671	548-629						↓(I) ↑(III)
6707163	Fibrinogen alpha chain	P02671	444-629						
7179072	Fibrinogen alpha chain	P02671	515-629						
7501917	Fibrinogen alpha chain	P02671	466-629		P-565	S-501			

Figure S12. Individual proteoforms upregulated in late-stage cirrhosis. Differentially expressed proteoforms (DEPs) significantly downregulated in stage (I) and upregulated in stage (II) (green), downregulated in stage (I) and upregulated in stages (II) and (III) (red), downregulated in stage (I) and upregulated in stage (III) (orange). Each proteoform is shown with its unique proteoform number, protein name and Uniprot ID, amino acid sequence relative to the Uniprot ID, and presence or absence of post-translational modifications. Colors are based on the proteoform signatures created in **Figure 4**. Abbreviations: PFR (proteoform).

*Note that half Cysteine could be artifact