Article

Deep proteome profiling of metabolic dysfunction-associated steatotic liver disease

Check for updates

Felix Boel [1,2], Vyacheslav Akimov[1,2], Mathias Teuchler[1,2], Mike Krogh Terkelsen [1,2], Charlotte Wilhelmina Wernberg [2,3], Frederik Tibert Larsen[1,2], Philip Hallenborg[1], Mette Munk Lauridsen [2,3], Aleksander Krag [2,4], Susanne Mandrup [1,2], Kim Ravnskjær [1,2] & Blagoy Blagoev [1,2] ✉

Abstract

Background Metabolic dysfunction-associated steatotic liver disease (MASLD) affects roughly 1 in 3 adults and is a leading cause of liver transplants and liver related mortality. A deeper understanding of disease pathogenesis is essential to assist in developing blood-based biomarkers.

Methods Here, we use data-independent acquisition mass spectrometry to assess disease-state associated protein profiles in human liver, blood plasma, and white adipose tissue (WAT).

Results In liver, we find that MASLD is associated with an increased abundance of proteins involved in immune response and extracellular matrix (ECM) and a decrease in proteins involved in metabolism. Cell type deconvolution of the proteome indicates liver endothelial and hepatic stellate cells are the main source of ECM rearrangements, and hepatocytes are the major contributor to the changes in liver metabolism. In the blood, profiles of several MASLD-associated proteins correlate with expression in WAT rather than liver and so could serve as suitable liver disease predictors in a multi-protein panel marker. Moreover, our proteomics-based logistic regression models perform better than existing methods for predicting MASLD and liver fibrosis from human blood samples.

Conclusions Our comprehensive proteomic analysis deepens the understanding of liver function and MASLD pathology by elucidating key cellular mechanisms and multi-organ interactions, and demonstrates the robustness of a proteomics-based biomarker panel to enhance diagnosis of MASLD and significant fibrosis.

Plain language summary

Metabolic dysfunction-associated steatotic liver disease (MASLD) is a common condition affecting about 1 in 3 adults. It occurs when there is too much fat in the liver, and is a leading cause of death and people needing liver transplants. To improve early detection, we studied proteins in the liver, blood, and fat of people with MASLD. Our analysis revealed changes in liver proteins and that certain proteins in the blood could serve as early indicators of liver disease. Some of the protein changes in the blood indicated changes in fat tissue rather than the liver. We propose that our blood test could be more accurate than commonly used methods for diagnosing early MASLD and so could enable better, non-invasive ways to detect liver disease.

Chronic liver diseases account for more than 2 million deaths annually. Advanced stages of steatotic liver disease (SLD) gradually impair liver function, which can result in non-reversible end-stage liver disease (ESLD). Liver transplantation is currently the only curative treatment for individuals with ESLD[1]. The most prevalent subtype of SLD, MASLD—formerly known as non-alcoholic fatty liver disease (NAFLD)[2]—affects more than 30% of all adults[3]. Early stages of MASLD are asymptomatic, making preventive screening and disease monitoring a difficult task. Currently, accurate diagnosis relies on a liver biopsy to histologically assess steatosis, lobular

inflammation, and hepatocellular ballooning. Based on the scoring, MASLD become categorized as metabolic dysfunction-associated steatotic liver (MASL) or metabolic dysfunction-associated steatohepatitis (MASH)[4]. In MASH, tissue repair processes lead to an accumulation of tissue scarring (fibrosis), giving rise to decompensated cirrhosis and hepatocellular carcinoma (HCC)[5].

Liver biopsies are invasive, costly, prone to sampling error, and at risk of intraobserver variation[6]. Consequently, many efforts have been made to utilize readily available biochemistry and clinical information to help grade

[1]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M, Denmark. [2]Center for Functional Genomics and Tissue Plasticity (ATLAS), University of Southern Denmark, Odense M, Denmark. [3]Department of Gastroenterology and Hepatology, University Hospital of Southern Denmark, Esbjerg, Denmark. [4]Department of Clinical Research, University of Southern Denmark, Odense M, Denmark. ✉e-mail: bab@bmb.sdu.dk

disease progression and thereby reduce unnecessary biopsies. However, screenings are not standardized and vary depending on the country, research laboratory, and individual health professionals. Moreover, these methods, deriving from correlation-based studies, does not consider interplay with other metabolic disorders and factors of human heterogeneity such as age, gender, ethnicity, genetics, epigenetics, smoking, alcohol consumption, obesity, gut microbiota, and hormonal imbalances[7]. As a result, clinicians face a shortage of tests for precise diagnosis, prognosis, disease monitoring, prediction, and assessing the effectiveness of interventions[8].

Blood sampling is the preferred method in non-invasive diagnostics[9], aiming to measure proteins that end up in circulation either by cell secretion or cell leakage. This makes proteomics an excellent tool in biomarker discovery[10]. To ensure proteins found in the blood are indeed disease-specific, consideration must be given to the biological changes in a tissue-specific context of the proteome[11]. Previous studies have focused on detailed transcriptomic profiling of MASLD along with SomaScan analysis of circulating proteins[12]. However, no studies have taken a purely proteomics-based approach to explore MASLD pathogenesis collectively in human liver and blood samples.

Recent advances in speed and sensitivity of mass spectrometers have unlocked methods for data-independent acquisition (DIA) proteomics[13]. Compared to conventional data-dependent acquisition (DDA), DIA eliminates bias of high abundance peptides, decreases run-to-run variation, improves sample reproducibility, and decreases required input material, all of which are vital attributes when analyzing highly variable human samples from low starting material[14]. Furthermore, DIA avoids the bias towards high abundant peptides and can consistently quantify hundreds even in blood[15]. This makes it an exciting platform for discovery of novel blood biomarkers, and, with increasing instrument speed and automation efforts, this approach holds a potential for high-throughput screening of blood samples in a clinical setting.

In this study, we perform DIA-based deep proteomic profiling of liver, plasma, and adipose tissue samples obtained from patients with ≥class 2 obesity (BMI > 35) at risk of MASLD. We find that MASLD is characterized by an increased abundance of proteins involved in immune activation and extracellular matrix remodeling, and a decrease in metabolic pathways. By integrating these proteomics data with publicly available single-cell transcriptomes, we link the changes to specific liver cell types, implicating endothelial cells and hepatic stellate cells in ECM alterations and hepatocytes in metabolic dysregulation. In parallel, we observe that several MASLD-associated blood proteins reflect expression changes in adipose tissue rather than liver, underscoring the importance of cross-organ interactions. Finally, we develop a multi-protein biomarker panel for identifying MASLD and significant fibrosis in blood samples, demonstrating the potential of a proteomics-based approach to improve non-invasive assessment of liver disease.

Methods
Patient cohort
Study design: samples for proteomics were acquired from individuals participating in the PROMETHEUS study, a prospective interventional case-control study focusing on patients who undergo liver biopsy. PROMETHEUS is a single center study conducted in Denmark. Inclusion and exclusion criteria: eligible participants must be in the age range 18–70 and have a BMI ≥ 35 kg m^{-2}. The study excludes individuals with a limited life expectancy, alcohol consumption above 12 grams daily for women and 24 grams for men, any other chronic liver conditions, or usage of hepatotoxic drugs. Ethics and registration: the regional committee on health research ethics has granted approval for this study (S-20170210), which is registered at OPEN.rsyd.dk (OP-551, Odense Patient data Explorative Network) and at ClinicalTrial.gov (NCT03535142). Full ethical approval and informed written consent from participants were documented before initiation of the study. Data collection: study data include anthropometrics and pharmacological treatment information. Data collection was carried out prospectively and managed using the REDCap (Research Electronic Data

Capture) system hosted at OPEN.rsyd.dk, a secure, web-based platform designed to facilitate data capture for research purposes (https://www.sdu.dk/en/forskning/open). First sample for the current study was collected on June 28th, 2018, and sample collection is still ongoing.

Histological staging
Liver biopsies were evaluated by a single trained radiologist (T.D.C), who was blinded to all other clinical data. The assessment adhered to the NASH Clinical Research Network (NAS-CRN) classification system[16], with scoring for steatosis (0: <5%, 1: 5–33%, 2: 33–66%, 3: >66%), lobular inflammation (0: none, 1: <2 foci per 200x field, 2: 2–4 foci, 3: >4 foci), and hepatocellular ballooning (0: none, 1: moderate, 2: evident). These scores total to the NAS score (0–8). The FLIP (Fatty Liver Inhibition of Progression) algorithm and SAF (Steatosis, Activity, and Fibrosis) scoring system were used to further categorize patients based on the severity of their condition: no-MASLD (steatosis <1), MASL (steatosis ≥1), and MASH (steatosis ≥1, lobular inflammation ≥1, and hepatocellular ballooning ≥1)[17]. Fibrosis was staged using the Kleiner classification system, indicating the extent of tissue scarring (F0: none, F1: either perisinusoidal or portal/periportal fibrosis only, F2: both perisinusoidal and portal/periportal fibrosis, F3: bridging fibrosis, F4: cirrhosis)[16].

Sample acquisition
All liver biopsies, WAT biopsies, and blood samples were collected at the Department of Gastroenterology and Hepatology, Southwest Jutland Hospital in Esbjerg, Denmark as part of the PROMETHEUS study described above. Patients were overnight fasting prior to sampling. Blood samples were collected in the morning prior to liver and WAT biopsies by specially trained hospital lab staff. Two designated lab techs handled blood sampling. Standard biochemical testing was done on the Cobas6000 (Roche). Blood components for the ATLAS biobank were frozen in a −80 °C freezer immediately after aliquoting to secondary tubes. The WAT subcutaneous tissue sampling was performed at the same location as the percutaneous liver biopsy, prior to the collection of liver tissue samples. This was consistently conducted in the midclavicular region of the lower ribcage. WAT was obtained immediately before the liver biopsy and was, as such, collected during the liver biopsy procedure through a 2 cm incision in the liver biopsy needle insertion area. Approximately 2 × 3 cm WAT was removed with a pair of Metzenbaum dissection scissors, cleaned in a sterile saline solution, frozen in liquid nitrogen, and kept at −80 °C. Percutaneous liver biopsies from right liver lobe were acquired under sterile conditions using a 16–18 G Menghini suction needle (Hepafix, Braun, Germany). The liver biopsy was cleaned in a sterile saline solution. At least 1.5 cm was sent for histologic evaluation, and the rest was used for research analysis. Liver biopsies for histologic staging (minimum 1.5 cm) were immediately stored in 4% formalin and embedded in paraffin. All biopsies for proteomics were immediately snap frozen in liquid nitrogen and then stored at −80 °C. Additional data, such as anthropometrics, were collected on the same day as biopsies and blood sampling.

Protein purification and digest of biopsies
For quantitative proteomics, biopsies were sliced in two and homogenized in denaturing buffers (either 8 M guanidine hydrochloride in 25 mM ammonium bicarbonate or 5% SDS) by 3–10 s of sonication (20% intensity) and heated for 5 min at 95 °C. WAT biopsies had an additional centrifugation step at 13,000 rpm for 10 min to separate and avoid carry-on of superfluous lipids into the following steps. Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) was used to measure protein concentrations, and 10 µg of protein (5 µg from each lysate) was combined in a 1:1 volume before diluting 4x with water for a final concentration of 1 M guanidine hydrochloride (and 0.625 % SDS). For protein aggregation capture on microparticles[18] we used 70 % (v/v) acetonitrile (ACN), 40 µg of magnetic HILIC beads (ReSyn Biosciences (Pty) Ltd.) and 30 min incubation at room temperature with agitation on. Bead-bound protein aggregates were retained using a magnet while washing with pure ACN, followed by washing

with 70% (v/v) ethanol. Bead-bound protein aggregates were then resuspended in 100 μL of 50 mM ammonium bicarbonate, reduced with 2 mM DTT for 30 min, and alkylated with 11 mM chloroacetamide for 30 min in the dark. Proteins were proteolytically digested using 1:200 of LysC (Wako) for 1 h at 37 °C, followed by 1:100 of trypsin (Promega) overnight at 37 °C. Samples were acidified (pH 2.0–2.5) using trifluoroacetic acid, loaded onto in-house prepared C18 StageTips, desalted, and eluted again using 45 μL of 60% ACN in 0.5% (v/v) acetic acid. Finally, eluates were vacuum-dried to 2–3 μL in a speed-vac and re-diluted to ~10 μL using 0.5% (v/v) acetic acid.

Preparation of blood plasma
Blood plasma was first diluted in 50 mM ammonium bicarbonate to a final concentration of 1 μg/μL. Next, we prepared 10 μg of protein in a 1:10 dilution of 50 mM ammonium bicarbonate, by reduction, alkylation, protein digest, and desalting on StageTips, as described above.

Data-independent acquisition mass spectrometry
A total of ~800 ng of digested and purified peptide mixture was loaded onto an LC-MS/MS system consisting of an EASY-nLC 1000 nanoflow liquid chromatograph coupled with either a Q Exactive HF-X mass spectrometer or an Orbitrap Exploris 480 mass spectrometer (all from Thermo Fisher Scientific)[19]. For reverse-phase chromatography we used a 20-cm analytical column with an inner diameter of 75 μm packed in-house with ReproSil Pur C18-AQ 1.9 μm resin (Dr Maisch GmbH), paired with an ACN/water (0.5% acetic acid) solvent system at a flow rate of 0.25 μL min^{-1}. For peptide separation of biopsy proteomes, we used a step-gradient from 3 to 8% ACN for 7 min, followed by a slow gradient to 35% ACN for 86 min, a ramp to 45% ACN for 15 min and a plateau of 95% ACN for 5 min. For peptide separation of plasma proteomes, we used a step-gradient from 3 to 8% ACN for 1:22 min:s, followed by a slow gradient to 35% ACN for 16:43 min:s, a ramp to 45% ACN for 2:55 min:s and a plateau of 95% ACN for 5 min. The mass spectrometers were operated by sequential window acquisition of all theoretical fragment ion spectra[19]. For MS1, we used a 350 to 1400 m/z survey scan with a resolution of 120,000, a maximum ion injection time of 45 ms, and a normalized automatic gain control target of 300%. This was followed MS2, fragmentation of precursor ions by higher-collisional dissociation at a collision energy of 28% for each 13 m/z isolation window, and a 1 m/z overlap between sliding windows to sequentially cover the 360.5–1000.5 m/z range by 50 scans. For MS2, resolution was set to 30,000 and the maximum ion injection time to 54 ms. DIA raw files were analyzed in Spectronaut™ version 18.1 (Biognosys) using default settings, and our in-house generated spectral library for WAT and liver, and DirectDIA (library free) for plasma. The report table containing protein identifications and quantitative values were kept for subsequent data analysis in R.

In-house generated spectral library
Briefly, our in-house spectral library was generated using Pulsar in Spectronaut™ (Biognosys) from data files of previous analyses of human liver using high pH fractionation and DDA. In total, our spectral library for liver contains 8678 unique proteins from 139,207 stripped peptide sequences (199,022 modified peptides sequences). While our spectral library for WAT contains 7672 unique proteins from 113,270 stripped peptide sequences (127,551 modified peptides sequences).

Public single-cell RNA-seq data integration and annotation
For deconvolution, three human single-cell RNA-sequencing (scRNA-seq) datasets were retrieved from public repositories GSE115469[20], GSE136103[21], and GSE158723[22]. Seurat (v.4.0.3)[23] was used to process each dataset. Low quality cells were excluded (200 < n < 3000 genes, mitochondrial gene contributions <20%). Moreover, genes expressed in fewer than 50 cells were excluded to remove zero count genes. Following cell exclusion, normalization, scaling, and dimensional reduction were performed. DoubletFinder (v.2.0.3)[24] was employed to predict and remove doublets. For dataset integration, the processed datasets were merged, and the principal component embeddings were corrected using Harmony (v.0.1.0)[25]. CellTypist (v.0.1.4)[26]

was employed for automated cell type annotation of the complete dataset (trained model reference = Immune All Low). Manual correction of cell type annotations was performed to increase annotation resolution for hepatocytes, cholangiocytes, liver sinusoidal endothelial cells, liver endothelial cells, aHSCs, qHSCs, and VSMCs.

Statistics and reproducibility
Software: analysis was performed using the open-source statistics software R, version 4.3.1 (2023-06-16)—"Beagle Scouts". Analysis of differential expression: we first performed quantile normalization of the summarized protein quantifications after log2 transformation to achieve normal distribution. We filtered proteins missing in more than half of all the samples and then performed unsigned network topology analysis, and determined the optimal power, based on the scale free topology model fit and the mean connectivity (Supplementary Fig. S1a). We used a power of 11 to calculate unsigned adjacency, which was then used to calculate the topological similarity and corresponding dissimilarity matrices, for hclust-based dendrogram and gene module calculation, using the flashClust, cutreeDynamic, and labels2colors functions (Supplementary Fig. S1b). We used the WGCNA R package[27] to calculate module eigengenes and correlate eigengenes to patient characteristics by Pearson correlation (Supplementary Fig. S1c). Proteins from Black and Greenyellow gene modules, 348 proteins in total, was kept for further data analysis. Proteins were fitted using robust regression (MSqRob2 integration of the MASS rlm-function). Test statistics and Benjamini–Hochberg (BH) correction of p values for all contrasts of Kleiner fibrosis grade and SAF diagnosis was calculated using the topFeatures-function in MSqRob2[28]. The overlap between statistically significant proteins (BH-adjusted p value < 0.05) for Kleiner fibrosis grade and SAF diagnosis resulted in a total of 234 proteins. Pathway enrichment analyses: we used the ReactomePA R package[29]. Analysis was performed using the enrichPathway function and BH correction. Estimation of cell type contribution: for our deconvolutions, we did a count summarization for each cell type, followed by feature-wise min-max normalization. Transcriptional regulators: we defined transcriptional regulators based on the binding analysis for regulation of transcription tool[30]. Heatmap sorting: specifically for Fig. 4b, we first performed hierarchical clustering, and then subsequently sorted each of the three major protein groups according to their cell type clusters for easier interpretation. The original heatmap can be found in Supplementary Fig. S9. Statistical significance: for analysis of significance for individual proteins shown in the barplots we performed a Wilcoxon signed-rank test and BH correction method. Logistic regression models: we performed 20x five-fold cross-validation logistic regression using random seeds. For each experiment we calculated the AUROC, balanced accuracy, and F1 score and then reported the mean and standard deviation. Other models: for the simpler models we used published models and cutoffs: NAFLD liver fat score[31] (cutoff: >−0.640), hepatic steatosis index[32] (cutoff: >36), fatty liver index[33] (cutoff: ≥60), and APRI[34] (cutoff: ≥0.7). Data visualization: all data visualization were made using ggplot2[35], enrichplot[36], corrplot[37], VennDiagram[38], ComplexHeatmap[39], Pathview[40], and Adobe Illustrator. Sample information: details regarding sample sizes for each tissue and pathology can be found thoroughly described in Table 1. All samples within a given tissue/cohort are considered biologically independent replicates. Plasma samples in the follow-up cohort are not biologically independent from the initial cohort.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results
Study design and data collection
Our cohort consisted of patients with a BMI > 35 (kg m^{-2}) at risk of metabolic disorder. Most study participants were female, with a median age in the late 40s and a nearly 50% estimated body fat (Table 1). As part of the study protocol, all patients had a percutaneous suction needle liver biopsy,

Table 1 | Patient characteristics

Clinical feature	Liver	scWAT	oWAT	Initial plasma	Validation plasma (n = 41)	
	(n = 58)	(n = 58)	(n = 27)	(n = 143)	(n = 28)	(n = 13)
Gender (% female)	65.5	65.5	74.1	78.3	78.6	69.2
Age (years)	47.5 ± (13.6)	46.0 ± (13.5)	45.0 ± (11.9)	46.0 ± (12.3)	54.0 ± (11.8)	47.0 ± (14.7)
BMI (kg m^{-2})	42.6 ± (6.9)	41.5 ± (6.9)	41.1 ± (4.3)	41.1 ± (6.0)	39.8 ± (3.1)	28.8 ± (3.0)
Body fat (%)	49.9 ± (5.7)	49.3 ± (5.6)	NA	48.5 ± (5.6)	47.8 ± (4.8)	37.8 ± (8.6)
Kleiner fibrosis grade						
F0	13	10	6	39	6	5
F1	31	33	16	69	15	8
F2	11	10	3	25	6	0
F3	3	4	1	7	1	0
F4	0	1	1	3	0	0
SAF diagnosis						
No MASLD	12	16	13	39	9	13
MASL	29	30	11	74	14	0
MASH	17	12	3	30	5	0
Steatosis						
0	12	16	13	39	9	13
1	22	27	12	60	11	0
2	17	9	1	28	7	0
3	7	6	1	16	1	0
Hepatocellular ballooning						
0	41	46	24	112	23	13
1	16	12	3	29	4	0
2	1	0	0	2	1	0
Lobular inflammation						
0	13	18	13	49	11	12
1	31	28	11	71	16	1
2	11	12	3	21	1	0
3	3	0	0	2	0	0

Patient data for samples analyzed by proteomics shown as medians ± standard deviation (s.d.) or summarized counts.

which was assessed by an experienced liver pathologist, using NAFLD activity score (NAS) and Kleiner fibrosis grade. To determine their MASLD status, we used the SAF algorithm, which divides patients into no-MASLD, MASL and MASH[17]. For proteomics, we utilized DIA-MS workflow[13] (Fig. 1a) to analyze liver biopsies (n = 58), blood plasma (n = 143), scWAT biopsies (n = 58), and oWAT biopsies (n = 27). Additionally, as a validation cohort, we had 2-year follow-up blood plasma (n = 41) from a subset of the patients, some of whom had undergone bariatric surgery during follow-up. In liver, we quantified 7096 unique proteins with similar proteome coverage in each sample (Supplementary Fig. S2a, b). In scWAT and oWAT, we quantified 6945 and 6979 unique proteins respectively (Supplementary Fig. S2c, d). While in plasma, using short gradient DIA, we quantified 578 unique proteins (Supplementary Fig. S2e).

Proteins involved in the innate immune system, extracellular matrix organization, and metabolism are differentially expressed in the MASLD liver
Robustly calculating differential expression between disease stages require complete data. However, MS-based proteomics is inherently prone to random missingness, and considering human heterogeneity we did not want to impute any missing values. Instead, we used ridge regression to prevent overfitting during statistical analysis of differential expression[28]. To improve model performance, we first performed an unsigned weighted gene

co-expression network analysis (WGCNA)[27] to limit feature size and increase feature similarity in the dataset (Supplementary Fig. S1). Analysis of differential for Kleiner fibrosis grade and SAF diagnosis resulted in 304 and 256 significantly differentially expressed proteins (DEPs) respectively. To improve statistical confidence and ensure biological relevance, we carried out any further analysis using only the 234 overlapping DEPs between Kleiner fibrosis grade and SAF diagnosis (Fig. 1b).

To verify the relationship between the DEPs and MASLD histopathology, we profiled the protein-wise relative expression after sorting patients based on increasing severity of MASLD histopathology. After hierarchical clustering of DEPs, we observed two clusters of proteins pertaining to being either up- or downregulated in expression (Fig. 1c). Using pathway analysis, we found the most significantly downregulated proteins to all originate from various pathways of metabolism, while the upregulated proteins belonged to pathways related to extracellular matrix (ECM) organization and immune response (Fig. 1d, e).

Integration of public single-cell data links MASLD proteome changes to individual cell types
Although single-cell omics play a crucial role in deepening our understanding of disease mechanisms, single-cell proteomics has yet to become a routine methodology. To address this gap, we explored whether single-cell transcriptomic data from human liver could assist in deciphering our bulk
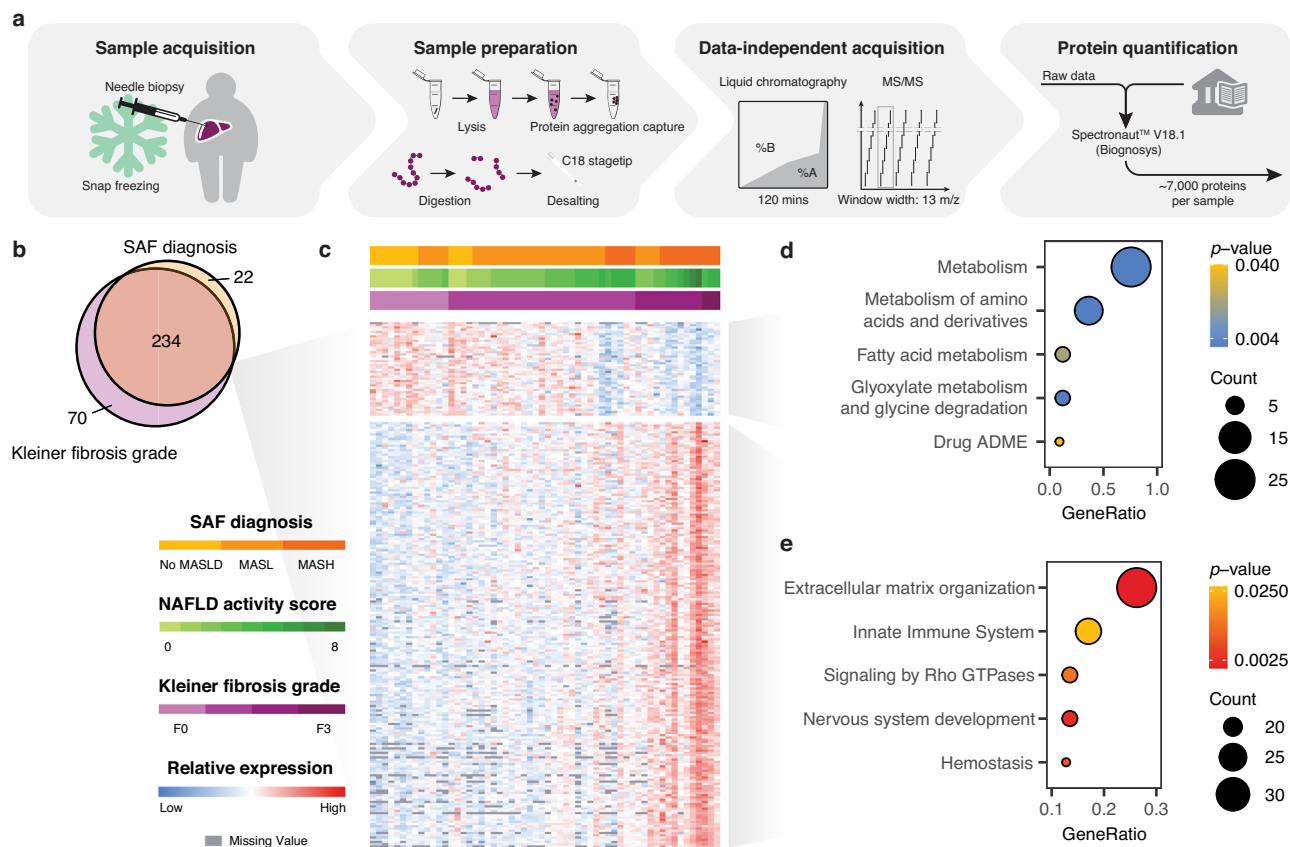
Fig. 1 | Identification of proteins related to MASLD pathogenesis. a Workflow of our data acquisition pipeline for liver samples ($n = 58$ biologically independent samples). b Venn diagram of differentially expressed proteins for Kleiner fibrosis grade and SAF diagnosis. c Heatmap for relative expression of the 234 differentially expressed proteins shared between Kleiner fibrosis grade and SAF diagnosis, shown as the relative expression across patients (columns). Rows (proteins) were sorted based on hierarchical clustering, while columns (patients) were sorted according to increasing (i) Kleiner fibrosis grade, (ii) SAF diagnosis, and (iii) NAFLD activity score. Pathway enrichment analysis based on the Reactome pathway database for proteins downregulated (d) and upregulated (e) with increasing MASLD histopathology. SAF steatosis activity and fibrosis, NAFLD non-alcoholic fatty liver disease, MS mass spectrometry, m/z mass-to-charge ratio, MASLD metabolic dysfunction-associated steatotic liver disease, MASL metabolic dysfunction-associated steatotic liver, MASH metabolic dysfunction-associated steatohepatitis.

proteomics data, aiming to provide insights into the potential cellular origins of the DEPs. To achieve this, we integrated three publicly available datasets[20–22], which we annotated collectively based on cell type-specific gene expression markers and utilized the results to extrapolate a broad overview of the cell type-specific context of the DEPs. While RNA and protein levels generally show some correlation, this analysis should only be used to gauge broader trends in the data. For RNA species that appear cell type-specific, it is reasonable to assume that corresponding protein levels are also derived from that cell type. However, findings for genes expressed in multiple cell types should be interpreted with caution.

Hierarchical clustering based on the predicted cell type-specific contribution towards the expression of the DEPs resulted in five clusters (Fig. 2a and Supplementary Fig. S3). As expected, the largest cluster, the HEP Cluster, consisted of DEPs predicted to be mainly expressed by the hepatocytes. We found the hepatocyte cluster to be the primary contributor of proteins associated with a decrease in metabolism, specifically in amino acid and fatty acid metabolism (Fig. 2b). Among the upregulated DEPs in hepatocytes, we found pathways of MET signaling. The MET receptor tyrosine kinase signals via classical mitogen-activated protein kinase (MAPK) signaling cascades but can, after binding to its ligand, hepatocyte growth factor (HGF), also transactivate other receptors, including the hyaluronan receptor CD44, ICAM1, and integrins[41].

The LEC Cluster consisted of DEPs expressed by liver endothelial cells (LECs), while the HSC Cluster consisted of proteins mainly expressed in quiescent hepatic stellate cells (qHSCs), activated hepatic stellate cells (aHSCs), and vascular smooth muscle cells (VSMCs) (Fig. 2a). High levels of

VSMCs and aHSCs are major hallmarks of MASLD pathogenesis[42]. Pathway analysis showed that DEPs expressed by the main resident cell types—HEP, LEC, and HSC Clusters—largely contribute to the organization of ECM (Fig. 2b). ECM organization is a convoluted process that involves many different regulatory networks. Proteoglycans are a major constituent of ECM and play an important role in cell signaling, e.g., by interacting with growth factors, cytokines, chemokines, and various pathogens. The architecture of the ECM provides many innate properties of tissues, and in MASLD pathogenesis, it is known to play a role in the advancement of fibrosis. In fact, liver stiffness measurements using elastiometry (i.e., Fibroscan) is today one of the best techniques in the non-invasive assessment of liver fibrosis[43].

To further investigate some of the ECM-related sub-categorization, we explored the cell type-specific cellular processes (Fig. 2b). For example, we found laminin interactions primarily in hepatocytes and LECs. Laminins are known to commonly interact with integrins to control cell-to-ECM adhesion. In connection with this, in the IMX Cluster—mainly immune-related cytotoxic T-cells and natural killer (NK) cells—we found pathways of cell-ECM interactions as well as L1CAM interactions and RHO GTPase effectors. L1CAM interactions are known to be involved in cell-cell adhesion and RHO GTPases can regulate cell behavior, cell motility, and cytoskeleton organization[44]. This suggested a finding of proteins that were directly involved in the infiltration of immune cells into the tissue, and indeed, we found many of the proteins involved in the KEGG leukocyte transendothelial migration pathway to be upregulated during MASLD pathogenesis (Fig. 3a and Supplementary Figs. S4 and 5).
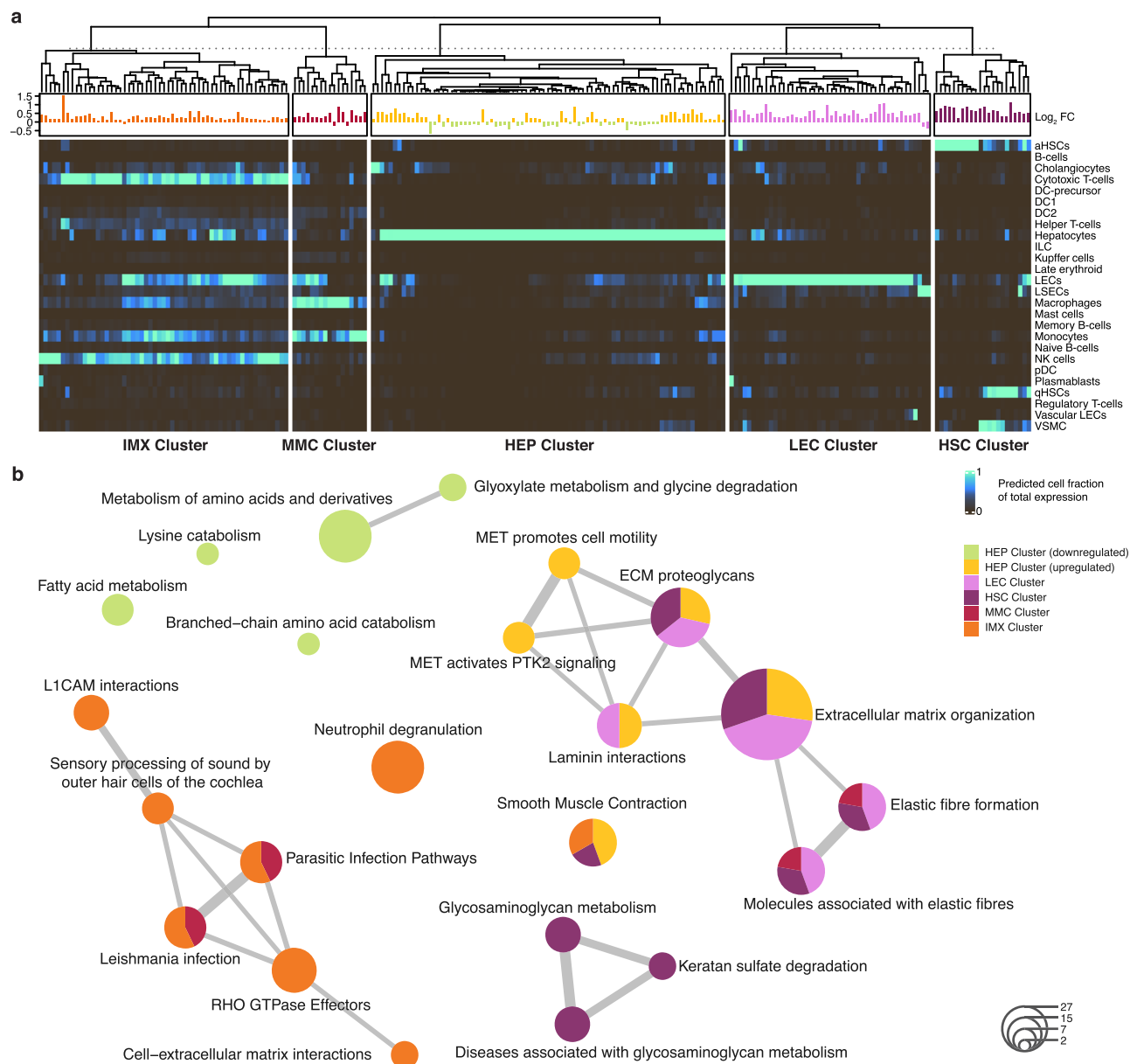
Fig. 2 | Public single-cell data to unravel a cell type specific response in MASLD pathogenesis. a Estimated cell type specific contribution for the total expression of our 234 significantly differentially expressed proteins in MASLD. Estimates are based on aggregated data from three single-cell datasets, collectively annotated, count summarized by cell type, and then min-max normalized features. Each protein's $\log_2$ fold change with increasing SAF diagnosis in expression shown on top (Supplementary Fig. S3). **b** Comparison of enriched pathways for each cluster after calculating the pairwise term similarity matrix. Node sizes show number of proteins in a category. Edge thickness show the overlap between two categories (minimum 20% overlap threshold). IMX immune cell mixture, MMC macrophages and monocytes, HEP hepatocytes, HSC hepatic stellate cell, LEC liver endothelial cells, LSEC liver sinusoidal endothelial cells, aHSCs activated HSCs, qHSCs quiescent HSCs, DC dendritic cell, ILC innate lymphoid cell, NK natural killer, VSMC vascular smooth muscle cell.

In the IMX Cluster—an immune cell mixture—we found a large group of proteins involved in neutrophil degranulation (Fig. 2a). RHO GTPase-activated Rac2 has been shown to promote primary neutrophil degranulation of reactive oxygen species (ROS). Among the DEPs, we also found an upregulation of TP53I3 and GPX3 in the hepatocyte cluster, both of which are known to modulate ROS. Increased ROS is a known hallmark in MASLD pathogenesis, and both TP53I3[45] and GPX3[46] have already been suggested as a drug target and a potential biomarker respectively.

Another interesting observation related to glycosaminoglycan (GAG) metabolism, which is largely contributed by the HSC Cluster (Fig. 2b). GAGs mainly consist of hyaluronan (HA; hyaluronic acid) and various sulfates (dermatan, keratan, heparan, and chondroitin) and act as a lubricating fluid in the ECM. HA is an anionic, nonsulfated glycosaminoglycan which has been suggested in multiple papers as a promising biomarker for MASLD[47]. Production of these GAGs can be activated by transforming growth factor-β (TGF-β) signaling. TGF-β signaling through SMAD2/SMAD3 in hepatocytes has been shown to accumulate free fatty acids, increase ROS production and mediate hepatocyte death[48]. Surprisingly, we also found an upregulation of elastic fibers in the LEC and HSC clusters, which, together with fibrillin microfibrils, can inhibit TGF-β signaling. When we checked for proteins related to TGF-β signaling, we found many of those to be differentially upregulated, but also many inhibitors of TGF-β (Fig. 3b and Supplementary Figs. S6 and 7). This could suggest a negative feedback mechanism or a mitigating response mechanism from nearby cells.
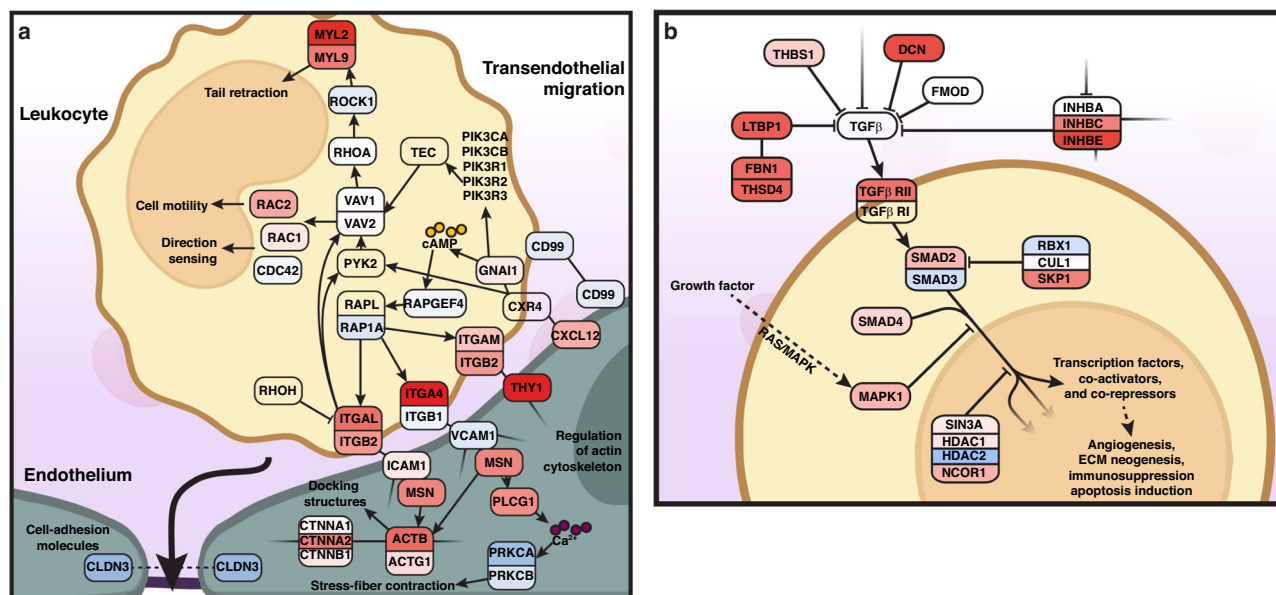
Fig. 3 | Depiction of signaling pathways related to MASLD pathogenesis. Proteins identified in **a** transendothelial migration (hsa04670) and **b** TGFβ signaling (hsa04350). Protein colors indicate change in expression where red is upregulated, blue is downregulated, white is no change, transparent is not identified. Details of full pathways and coloring schemes can be found in **a** Supplementary Figs. S4 and 5 and **b** Supplementary Figs. S6 and 7. ECM extracellular matrix.

Identification of transcriptional regulators that correlate with MASLD DEPs

To investigate the transcriptional regulators[30] involved in MASLD pathogenesis, we performed a correlation analysis of all transcriptional regulators identified in our dataset to various patient characteristics (Fig. 4a and Supplementary Fig. S8). We then kept the transcriptional regulators with an accumulative correlation score above 0.45 across MASLD histopathology (SAF diagnosis, Kleiner fibrosis grade, and NAFLD activity score) and performed hierarchical clustering of their correlation to the DEPs (Fig. 4b). Consolidating our earlier finding (Figs. 1 and 2), the expression patterns of the transcriptional regulators separated the DEPs into three major groups, group A consisting of pathways related to metabolism, group B relating to immune response, and group C relating to ECM organization (Fig. 4c–e). Further highlighting the robustness of the clustering results, we observed a positive correlation for SMAD2 with proteins involved in the immune response and a positive correlation between MYH11[49] and ACTG2[50], both being markers of VSMC. We also noted a positive correlation between the group of SMARC proteins with pathways of immune response, of which SMARCC1 has already been suggested as a putative prognostic marker for hepatocellular carcinoma (HCC) and targeted inhibition of SMARCC1 has been shown to reduce immune infiltration[51].

Among the transcriptional regulators with overall highest positive correlation to the DEPs were aryl hydrocarbon receptor (AHR) and pro-myelocytic leukemia protein (PML) (Fig. 4f, g). AHR is a ligand-activated transcription factor that was recently suggested as a potential drug target in patients with MASLD[52]. However, AHR signaling pathways are not fully understood, and considering most AHR studies are in mouse models, we lack evidence for its potential involvement in human diseases. While PML followed a very similar profile to AHR, it has not yet been described in relation to MASLD and could potentially be a novel transcriptional regulator. PML acts as a tumor suppressor and is frequently found in nuclear bodies and nucleoplasm. However, emerging evidence suggests a multi-faceted role that could make it important in the regulation of metabolism and immune response[53]. Interestingly, it has before been shown that PML can interact with UBE2I and SMAD2, both of which are among the top candidates of transcriptional regulators with the highest overall correlation to MASLD pathogenesis in our data (Fig. 4a).

As a major component of TGF-beta signaling, SMAD2 has already been linked to hepatic fibrosis, while UBE2I has mainly been associated with hepatocellular carcinoma (HCC)[54,55]. UBE2I is an E2 SUMO-conjugating enzyme and displayed a negative correlation with DEPs related to metabolism. Increased UBE2I expression has specifically been associated with autophagy processes and a poor prognosis for patients with HCC[54]. This negative correlation with pathways of metabolism could be a ripple effect from autophagy of dying hepatocytes, thus decreasing overall healthy hepatocyte function within the bulk of the analyzed tissue.

Identification of MASLD-associated proteins in blood plasma

There is an urgent need to find robust, non-invasive biomarkers of MASLD. To test if any of the DEPs identified in the liver could also be found in the blood of the patients, as a result of cellular secretion or cell leakage, we analyzed 184 plasma samples from obese individuals by DIA MS-based proteomics (Table 1). First, we performed a correlation analysis of plasma proteins to MASLD histopathology and then ranked them based on their correlation to Kleiner fibrosis grade and SAF diagnosis (Fig. 5a, b). While there was an overlap between the top 10 most positively and negatively correlating proteins for Kleiner fibrosis grade and SAF diagnosis, they displayed unique signatures. For example, C7 and COLEC11 were specific in advancement of fibrosis but were not as important in determining MASLD. COLEC11 is known to be involved in activating the lectin complement pathway and generally recognizes pathogen cell surface sugars and apoptotic cells, guiding immune cell migration to clear microbes and heal injury[56]. C7 is also part of the complement system, being part of the membrane attack complex on target cells, and was differentially expressed in both significant and severe fibrosis (Fig. 5c). Another protein important for fibrosis was the negatively correlating IGF2 (Fig. 5a, d). A known transcriptional repressor of IGF2 expression is CTCF, which we also identified in our list of transcriptional regulators of MASLD pathogenesis (Supplementary Fig. S9). In relation to IGF2, we also found IGFALS, IGFBP3, IGFBP5, and IGF1 all being among the most downregulated in progression of liver fibrosis (Fig. 5a). IGFBP3 is known to interchangeably carry IGF1 and IGF2 in the circulation until they are released in the target tissue.

When considering the top 10 most positively and negatively correlating proteins in relation to SAF diagnosis, we found ALDOB to be the most positively correlating candidate (Fig. 5b, e). ALDOB deficiency and ALDOB depletion, has been shown to result in increased intrahepatic accumulation of fatty acids and promoting hepatocellular carcinogenesis[57,58]. Interestingly, while IGF2 was most negatively correlating to Kleiner fibrosis grade
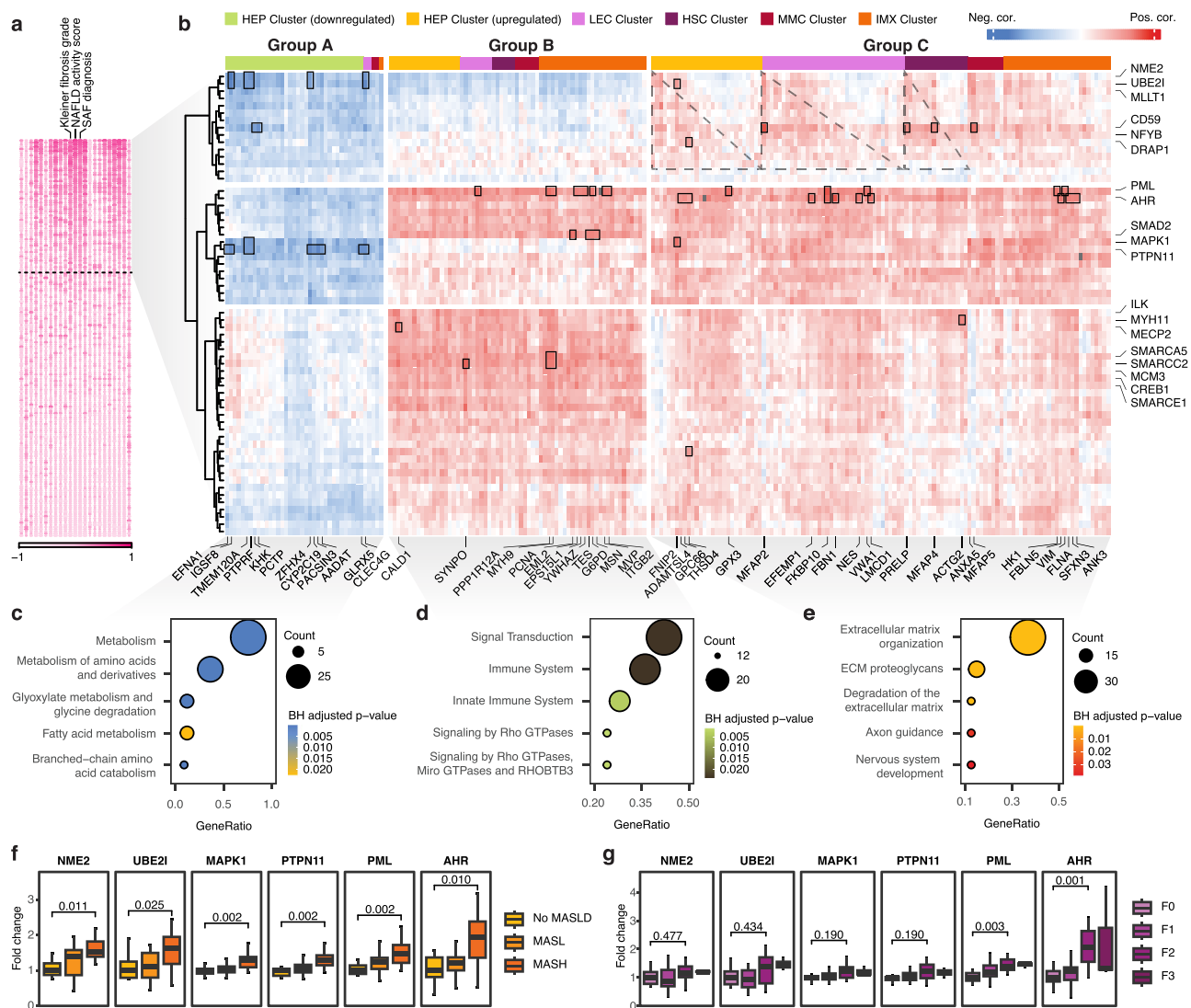
Fig. 4 | The role of transcriptional regulators in progression of MASLD.
a Correlation between all identified transcriptional regulators and patient characteristics, ordered by the sum of correlation to Kleiner fibrosis grade, NAFLD activity score, and SAF diagnosis. Cutoff was arbitrarily selected. **b** Hierarchical clustering of above cutoff transcriptional regulators with their correlation to the 234 significantly differentially expressed proteins of Kleiner fibrosis grade and MASLD. The columns (234 proteins) were first hierarchically clustered (Supplementary Fig. S9a), generating three clusters and then subsequently sorted within each of these three clusters by their cell type cluster (Fig. 2a). Proteins found in both the row and column (perfect correlation) has been grayed out. Squares indicate proteins with a higher correlation and were primarily chosen based on their

correlation to the highest-ranking transcriptional regulators. The full figure containing all protein names can be found in Supplementary Fig. S9b. **c–e** Pathway enrichment analysis based on the Reactome pathway database for proteins in each of the column-wise clusters. **f, g** Fold change in protein expression for a selected subset of transcriptional regulators according to increasing **f** SAF diagnosis and **g** Kleiner fibrosis grade. Boxplot are shown in the style of Tukey (median, hinges: Q1 and Q3, and whiskers: 1.5x of IQRs from Q1 and Q3) and statistical significance calculated using Wilcoxon signed-rank test and BH correction method. SAF steatosis, activity, and fibrosis, NAFLD non-alcoholic fatty liver disease, IMX immune cell mixture, MMC macrophages and monocytes, HEP hepatocytes, HSC hepatic stellate cell, LEC liver endothelial cells, BH Benjamini–Hochberg.

specifically, for SAF diagnosis, it was IGF1 (Fig. 5b, f). In many cancers, including non-islet cell tumor hypoglycemia[59], colorectal adenomas[60] and cancers[61], the molar ratios of IGF2/IGF1, IGF1/IGFBP3, and IGF2/IGFBP3 have already been used in diagnostics.

WAT contributes to the plasma level differences of proteins associated with MASLD

Metabolic disorders are inherently complex, often emerging from inter-organ crosstalk and multifaceted dysfunction across organs[62]. Among the blood plasma proteins correlating to MASLD histopathology (Fig. 5a, b), we find several that were absent in our liver-derived DEPs. This observation highlights the potential cross-organ dynamics related to MASLD. To explore this further, we first checked the expression differences for all blood

plasma proteins against their expression profiles in the liver (Fig. 5g). Here, we found several examples of MASLD correlating proteins showing a discrepancy between blood plasma and liver. To highlight a few, GPX3 was downregulated in blood plasma while being upregulated in liver, and ALDOB was highly upregulated in blood plasma, while showing no changes in the liver (Fig. 5h, i). We also found an inverse relationship for CD163 where CD163 is downregulated in liver but upregulated in blood plasma, corroborating previous findings[63].

Recognizing the pivotal role of adipose tissue in metabolic disorders[64], we investigated if blood plasma proteins that did not correlate with changes in the liver proteome, could instead be explained by changes in the proteome of the adipose tissues. We therefore performed DIA MS-based proteomic analysis of oWAT and scWAT from the same patient cohort and compared
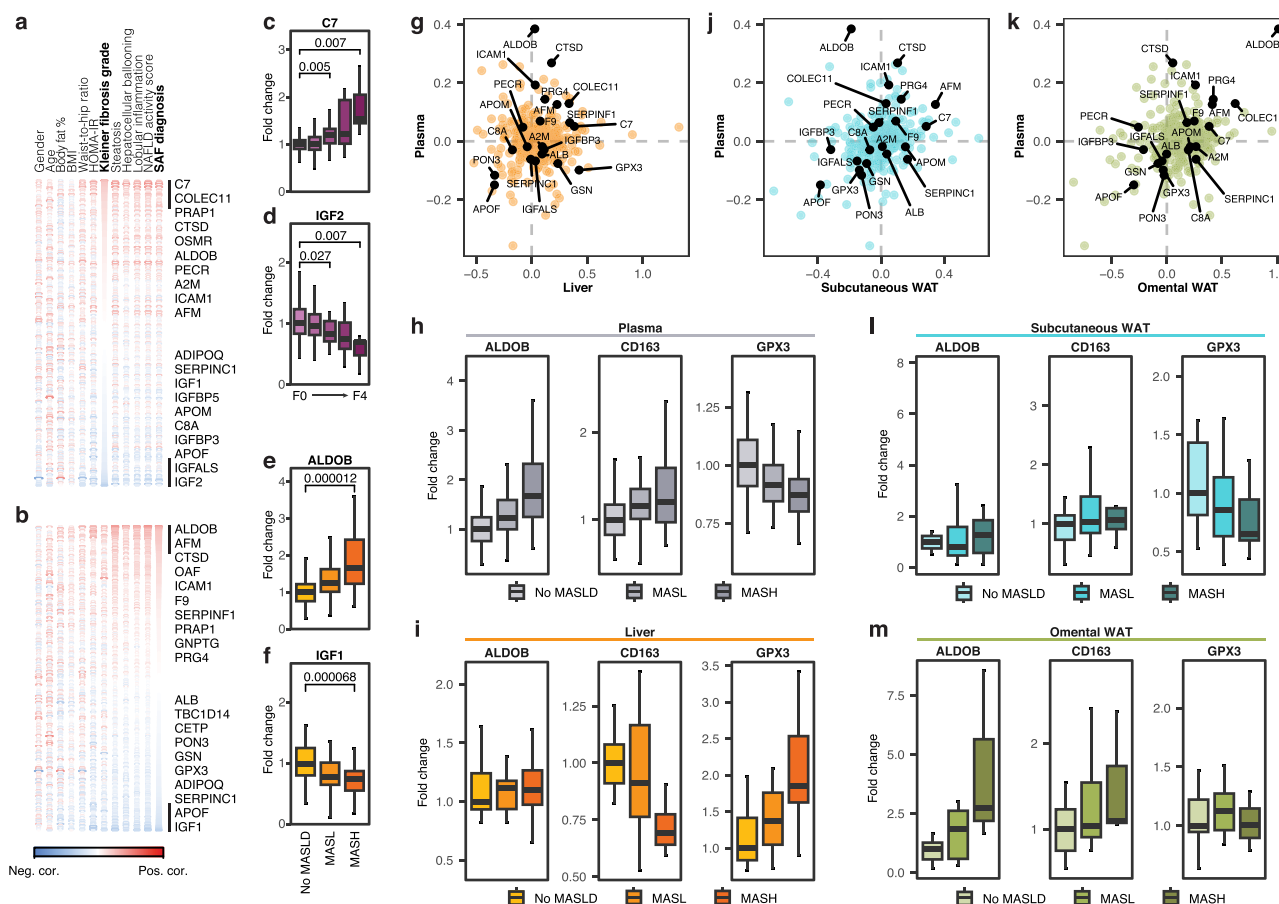
Fig. 5 | MASLD-related profiling of blood plasma. a, b Pearson correlation between all identified blood plasma proteins and patient characteristics (*n* = 184 biologically independent samples), with top 10 most positively and negatively correlated proteins displayed according to increasing **a** Kleiner fibrosis grade and **b** SAF diagnosis. Boxplots of proteins most positively and most negatively correlating with **c, d** Kleiner fibrosis grade and **e, f** SAF diagnosis. Comparisons of the log2 fold change in protein expression across MASLD progression, using all shared proteins between **g** blood plasma and liver, **j** blood plasma and scWAT, and **k** blood plasma and oWAT. Proteins highlighted with black represent those from the top and bottom 10 proteins most correlating to Kleiner fibrosis grade and SAF diagnosis from the Pearson correlations. **h, i, l, m** Boxplots displaying relative fold change in expression for three proteins selected based on no correlation or a negative correlation between blood plasma and liver during progression of MASLD, for each of the four tissues; **h** blood plasma; **i** liver; **l** scWAT; **m** oWAT. Boxplot **c–f** are shown in the style of Tukey (median, hinges: Q1 and Q3, and whiskers: 1.5x of IQRs from Q1 and Q3) and statistical significance calculated using Wilcoxon signed-rank test and BH correction method. SAF steatosis, activity, and fibrosis, NAFLD non-alcoholic fatty liver disease, BMI body mass index, HOMA-IR homeostatic model assessment for insulin resistance, MASLD metabolic dysfunction-associated steatotic liver disease, MASL metabolic dysfunction-associated steatotic liver, MASH metabolic dysfunction-associated steatohepatitis, WAT white adipose tissue.

the change in expression for all blood plasma proteins against the changes observed in adipose depots during MASLD pathogenesis (Fig. 5j, k). Indeed, for some of the MASLD-related proteins that could not be explained by changes in the liver proteome, we could correlate to levels in adipose tissues. For example, ALDOB and CD163 were specifically upregulated in oWAT, matching the profiles observed in blood plasma for these proteins, while difference in expression for GPX3 was better matching the profiles observed in scWAT (Fig. 5l, m). These findings highlight the importance of considering multi-tissue interplay when studying complex diseases such as metabolic disorders.

Predicting MASLD pathogenesis from blood plasma

Strategies relying on a single protein marker have not been very successful in diagnosis of MASLD-related pathologies. We therefore tested whether a proteomics-based model can be used as a "multi-protein panel marker" approach for predicting liver disease progression. For this purpose, we utilized the quantitation we had already attained using short-gradient DIA MS-proteomic analysis in plasma. We generated two 20-protein proteomics models, one designed to predict significant fibrosis (≥F2) and one for MASLD. These two proteomics models are

based on the 10 most positively and 10 most negatively correlating proteins found for Kleiner fibrosis grade and SAF diagnosis, respectively (Fig. 5a, b). We also made a simplified version of the two 20-protein proteomics models consisting of only five proteins each—C7, ALDOB, ICAM1, CTSD, and IGF2 for significant fibrosis, and AFM, APOF, ALDOB, PRG4, and IGF1 for MASLD (Supplementary Fig. S10). These two simplified proteomics models were based on the five proteins—from the 20-protein candidate list—with the highest individual AUROC. We chose simple logistic regression for its ease of interpretability, and stress-tested the performance using 20x five-fold cross-validation. To evaluate model performance, we reported the area under the receiver-operating characteristic (AUROC), balanced accuracy, and F1 score (Fig. 6a–c, e). Our proteomics-based models, the 20-protein panel and the simplified 5-protein version, showed very similar results both performing better than models based on existing frequently used metrics for predicting MASLD and fibrosis, as well as logistic regression models de novo-generated based on the popular FIB-4 and APRI[65].

In addition to our initially screened plasma samples, we also received 2-year follow-up blood plasma from a subset of the cohort (Table 1). From this subset, several patients had drastically reduced their body fat, with 13
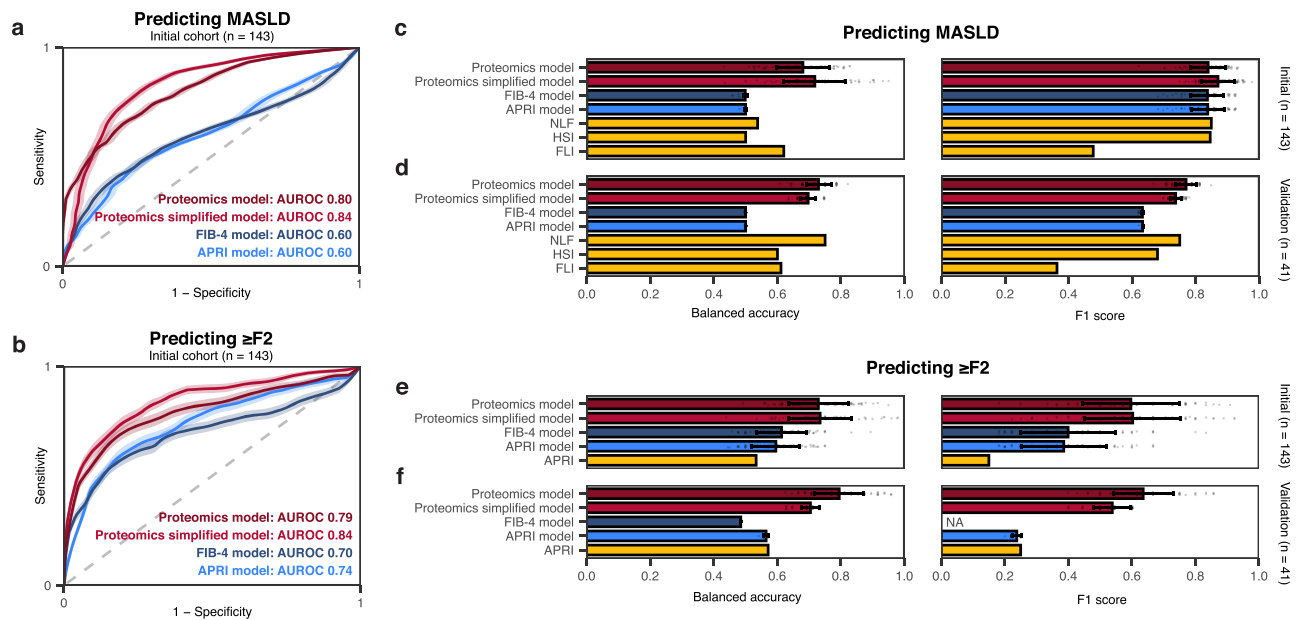
Fig. 6 | Performance evaluation of proteomics biomarkers for MASLD and ≥F2. Performance metrics for the logistic regression models after 20x five-fold cross validation (APRI, FIB-4, 2x Proteomics models one for significant fibrosis (≥F2) and one for MASLD based on proteins from Fig. 5a, b respectively, and 2x Proteomics simplified models one for ≥F2—C7, ALDOB, ICAM1, CTSD, and IGF2—and one for MASLD—AFM, APOF, ALDOB, PRG4, and IGF1). **a–c**, **e** Covers performance metrics of our initial cohort (n = 143 biologically independent samples), and **d**, **f** covers our validation cohort (n = 41 biologically independent samples) consisting in follow-up blood samples. **a**, **c** Model performance on predicting MASLD from our initial cohort and **d** again using the same model and model parameters on our validation cohort. **b**, **e** Model performance on predicting ≥F2 from our initial cohort and **f** again using the same model and model parameters on our validation cohort. **a**, **b** The receiver-operator characteristic (ROC) curve. **c–f** The balanced accuracy (left) and the F1 score (right). Bar plots show the mean across all one-hundred evaluations, dots show performance for each evaluation, and error bars show the standard deviation. **a**, **b** Faded area in the ROC curves show the 95% confidence interval. Coloring for the ROC curves matches the bar plots. MASLD metabolic dysfunction-associated steatotic liver disease, AUROC area under the receiver operating characteristic curve, NAFLD non-alcoholic fatty liver disease, APRI aspartate aminotransferase to platelet ratio index, FIB-4 fibrosis-4.

out of 41 losing more than 10 BMI points due to bariatric surgery. Moreover, 31 out of the 41 patients had changes in their histopathology, of which 18 had improved their liver pathology scoring. We then used these 41 plasma samples as a validation cohort to evaluate the stability of our initial models' performances in response to drastic phenotypic shifts within the initial cohort. We tested each individual model during the 20x five-fold cross validation on this cohort using the exact same parameters. The performance of the simplified 5-protein model was slightly lower compared to the 20-protein panel, however both proteomics-based models again performed similarly or even better compared to the frequently used metrics, which in general had a greater degree of instability (Fig. 6d, f). These results clearly substantiate the notion that such proteomics-based models relying on disease-specific protein panels are indeed very robust and suitable strategy for characterization of complex diseases like MASLD.

Discussion

In this study, we took a proteomics-centric approach to explore MASLD in humans. We found an upregulation of immune response and extracellular matrix organization as key hallmarks of disease pathogenesis, accompanied by a downregulation of metabolism which could likely be a consequence of cellular reprogramming, loss of hepatocyte identity, and cell death[66]. We thoroughly explored the cell type-resolved signaling pathways and transcriptional regulators in MASLD pathogenesis, which we expect to be a valuable resource for future studies aiming to further investigate the cellular mechanisms in MASLD pathogenesis. Importantly, our findings reveal distinct expression patterns across liver, blood plasma, and WAT, highlighting the importance of considering interorgan crosstalk in understanding MASLD. Moreover, our disease-specific proteomics models generally performed better than existing diagnostic methods, offering a more accurate and reliable tool for predicting disease progression. Although rather speculative at this stage, we could envision similar strategies to replace the widely used single marker predictions in the clinics.

Previous studies have predominantly focused on a genomic and transcriptomic understanding of MASLD in humans. Here, we expanded the current understanding of MASLD from a proteomics perspective, while leveraging our findings for improving diagnostics. Interestingly, the difference in performance between our comprehensive 20-protein model and the streamlined 5-protein model was minimal. This suggests that good model performance hinges on its ability to capture essential disease aspects. With further optimization and the adoption of more sensitive quantification methods, we expect that the performance of our simplified models can be improved even further. By integrating comprehensive proteomic profiling across organs, we also uncovered potential crosstalk between liver and adipose tissues, specifically through proteins like ALDOB, CD163, and GPX3, whose differential expression suggests a nuanced mechanism of disease manifestation not previously studied. Indeed, only six of the proteins found in our 20-protein proteomics models, which was purely based on proteins correlating to disease progression, were also found in our list of 234 liver DEPs.

Proteins comprise 75% of all drug targets approved by the FDA and are the primary source of biomarkers[9]. By advancing our understanding of MASLD at a proteomic level, our work lays the foundation for improving diagnostic and therapeutic strategies. The correlation of adipose and plasma proteins with liver pathology not only aids in understanding MASLD mechanisms but also facilitates the development of targeted therapies. Our analysis highlighted proteins involved in the immune response and ECM organization, such as CD163 and GPX3. These proteins represent potential targets for new therapeutic interventions aimed at mitigating the inflammatory and fibrotic pathways in MASLD. Targeting GPX3, which is involved in the regulation of oxidative stress, could help in designing antioxidant therapies to reduce liver damage. Similarly, interventions that modulate CD163 activity could potentially limit the inflammatory responses, offering a therapeutic strategy to slow disease progression. Moreover, the identification of transcriptional regulators such as PML and

Article

AHR that correlate to GPX3 expression adds a layer of complexity and opportunity. Understanding these regulatory mechanisms further, can potentially lead to the precision therapies that target these pathways more effectively. Furthermore, our diagnostic proteomics-based biomarker models pave the way for non-invasive, cost-effective methods that could transform current practices by decreasing reliance on liver biopsies and enhancing early disease detection. They also provide a more nuanced insight for identifying MASLD. Leveraging differential expression patterns in key proteins, could potentially be utilized to predict disease trajectory in patients, which could enable more tailored interventions based on the projected course of the disease.

The strengths of this study include consistency in methodology across tissues and a well-described cohort with histological scoring for all samples. All samples across tissues were prepared using the same sample preparation procedure, processed in the same lab, same instrument setups, and using DIA-MS to vigorously fragment all eluting peptides. Using DIA-MS, we could use remains from the same biopsy that was used for histological scoring which would not have been feasible previously at such a depth. Moreover, conventional DDA would have suffered from high sample variance in human heterogeneity, introducing even more missingness and making the results more difficult to interpret. Being proteomics-based helps the translational potential of our findings in future diagnostic and therapeutic efforts.

A major limitation to our study is the relatively small cohort size combined with cohort biases which can strongly limit the interpretability and applicability of our results to the broader population. Furthermore, we have very few samples from more advanced fibrosis stages, which are typically of greater clinical interest. To address some of these limitations, we took a conservative approach to the statistical analysis and overall prioritized more emphasis on the less-explored early stages of the disease. Lacking comparators for predicting the early disease stages, we decided to also generate de novo logistic regression models optimized for our cohort based on the vastly popular FIB-4 and APRI scores. However, it should be noted that these metrics were not originally designed for detection of early disease stages, and therefore, comparisons to these metrics should be interpreted with some caution. While DIA-MS offers robust quantification and reproducibility, we did not consider post-translational modifications which play a huge role in cellular mechanisms and could be crucial in understanding MASLD pathogenesis. Given the discrepancy between RNA and protein expression levels, it should be expected that our cell type deconvolutions have a margin of error. Although our biomarker panel clearly correlates to disease progression, it does not necessarily reflect liver-specific mechanisms as already highlighted. MASLD is a multifaceted disease and investigating the direct mechanisms is outside the scope of this study. Analyzing multiple tissues adds complexity to the data interpretation, and in a human study like this we cannot derive any direct evidence for interorgan crosstalk. Moreover, we do not consider total abundances between different tissues which could be crucial for explaining the origin of changes observed in the blood proteome. Discrepancies between liver and plasma proteomes in MASLD pathogenesis could also be partly explained by morphological changes in the liver, such as cellular leakage from dying hepatocytes. Finally, mass spectrometry-based equipment is not commonly used in the clinics yet, thus limiting the general applicability of our biomarker panel in its current state.

Future research directions should employ longitudinal studies to monitor the dynamics throughout MASLD progression across relevant organs, and potentially explore even more organs than what we did here. Incorporating single-cell proteomics will also help delineate the cellular origins and functionalities of crucial proteins, providing mechanistic insights into therapeutic targets. Moreover, by integrating multiple omics we can develop a more comprehensive model that captures the intricate biological interplay underlying MASLD. Finally, by further optimizing our proteomics models and improving assay sensitivity, can advance the development of robust cost-effective non-invasive diagnostic tools, to reduce the need for liver biopsies.

In summary, we here provide exploratory proteomics-based study of MASLD, covering over 7000 unique proteins in human liver biopsies from 58 obese patients with various degrees of liver disease progression. Using single-cell transcriptomic information to deconvolute our proteomics dataset, we linked many of the proteins associated with disease progression to individual cell types and the potential transcriptional regulators responsible for their expression changes in the disease states. The analyses uncovered more depth of the mechanisms related to the immune response and the extracellular matrix organization, including HSC differentiation, laminin and integrin interactions, RHO GTPase signaling, cell-ECM interactions, remodeling of LECs to accommodate transendothelial migration of immune cells, production of ROS, GAG metabolism, and regulation of TGF-β signaling. While being a liver-centric study, we nevertheless highlighted the multi-organ effects of MASLD in metabolic dysfunction trough quantitative proteomics analysis of scWAT and oWAT biopsies from the same 58 patient to a similar depth of ~7000 unique proteins in each tissue. Last but not least, we utilized short-gradient DIA MS-proteomic approach for the analysis of 184 blood plasma samples and indicated that our proteomics marker models performed similarly or better than existing frequently used non-invasive methods in diagnosing MASLD. Overall, our findings provide rich and reliable resource, strengthening previous knowledge and bringing new insights that could help understanding normal liver function and MASLD pathology.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[67] partner repository with the dataset identifier PXD051911. Relevant metadata and protein quantitation tables for each tissue has also been added as Supplementary Data for ease of accessibility. Supplementary Data 1–5 (1, meta data and biochemical data; 2, liver protein quantification table; 3, batch-corrected plasma protein quantification table; 4, scWAT protein quantification table; and 5, oWAT protein quantification table).

Code availability

All bioinformatic analyses were performed in the open-source statistics software R, version 4.3.1 (2023-06-16)—"Beagle Scouts". We have made a dedicated GitHub repository with our codes for the predictive proteomics-based models publicly available (https://doi.org/10.5281/zenodo.14904437)[68]. Additional R scripts use existing code examples from popular R packages and have been described in the "Methods" section.

References
1. Friedman, S. L., Neuschwander-Tetri, B. A., Rinella, M. & Sanyal, A. J. Mechanisms of NAFLD development and therapeutic strategies. *Nat. Med.* **24**, 908–922 (2018).
2. De, A., Bhagat, N., Mehta, M., Taneja, S. & Duseja, A. Metabolic dysfunction-associated steatotic liver disease (MASLD) definition is better than MAFLD criteria for lean patients with NAFLD. *J. Hepatol.* **80**, e61–e62 (2024).
3. Younossi, Z. M. et al. The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review. *Hepatology* **77**, 1335–1347 (2023).
4. Hashimoto, E., Taniai, M. & Tokushige, K. Characteristics and diagnosis of NAFLD/NASH. *J. Gastroenterol. Hepatol.* **28**, 64–70 (2013).
5. Hernandez-Gea, V. & Friedman, S. L. Pathogenesis of liver fibrosis. *Annu. Rev. Pathol. Mech. Dis.* **6**, 425–456 (2011).
6. Brunt, E. M. et al. Complexity of ballooned hepatocyte feature recognition: defining a training atlas for artificial intelligence-based imaging in NAFLD. *J. Hepatol.* **76**, 1030–1041 (2022).

7. Eslam, M. et al. MAFLD: a consensus-driven proposed nomenclature for metabolic associated fatty liver disease. *Gastroenterology* **158**, 1999–2014.e1991 (2020).

8. Vali, Y. et al. Biomarkers for staging fibrosis and non-alcoholic steatohepatitis in non-alcoholic fatty liver disease (the LITMUS project): a comparative diagnostic accuracy study. *Lancet Gastroenterol. Hepatol.* **8**, 714–725 (2023).

9. Thiele, M. et al. Opportunities and barriers for omics-based biomarker discovery in steatotic liver diseases. *J. Hepatol.* **81**, 345–359 (2024).

10. Niu, L. et al. Plasma proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease. *Mol. Syst. Biol.* **15**, e8793 (2019).

11. Niu, L. et al. Noninvasive proteomic biomarkers for alcohol-related liver disease. *Nat. Med.* **28**, 1277–1287 (2022).

12. Govaere, O. et al. Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci. Transl. Med.* **12**, eaba4448 (2020).

13. Schubert, O. T. et al. Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nat. Protoc.* **10**, 426–441 (2015).

14. Anjo, S. I., Santa, C. & Manadas, B. SWATH-MS as a tool for biomarker discovery: from basic research to clinical applications. *Proteomics* **17**, 1600278 (2017).

15. Geyer, P. E., Holdt, L. M., Teupser, D. & Mann, M. Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* **13**, 942 (2017).

16. Kleiner, D. E. et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**, 1313–1321 (2005).

17. Bedossa, P. et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* **56**, 1751–1759 (2012).

18. Batth, T. S. et al. Protein aggregation capture on microparticles enables multipurpose proteomics sample preparation. *Mol. Cell. Proteom.* **18**, 1027–1035 (2019).

19. Correia, C. M. et al. Acute deletion of the glucocorticoid receptor in hepatocytes disrupts postprandial lipid metabolism in male mice. *Endocrinology* **164**, bqad128 (2023).

20. MacParland, S. A. et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 1–21 (2018).

21. Ramachandran, P. et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* **575**, 512–518 (2019).

22. Payen, V. L. et al. Single-cell RNA sequencing of human liver reveals hepatic stellate cell heterogeneity. *JHEP Rep.* **3**, 100278 (2021).

23. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).

24. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e324 (2019).

25. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

26. Dominguez Conde, C. et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022).

27. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma.* **9**, 1–13 (2008).

28. Goeminne, L. J., Gevaert, K. & Clement, L. Experimental design and data-analysis in label-free quantitative LC/MS proteomics: a tutorial with MSqRob. *J. Proteom.* **171**, 23–36 (2018).

29. Yu, G. & He, Q.-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479 (2016).

30. Wang, Z. et al. BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. *Bioinformatics* **34**, 2867–2869 (2018).

31. Kotronen, A. et al. Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors. *Gastroenterology* **137**, 865–872 (2009).

32. Lee, J.-H. et al. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. *Dig. Liver Dis.* **42**, 503–508 (2010).

33. Bedogni, G. et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol.* **6**, 1–7 (2006).

34. Lin, Z. H. et al. Performance of the aspartate aminotransferase-to-platelet ratio index for the staging of hepatitis C-related fibrosis: an updated meta-analysis. *Hepatology* **53**, 726–736 (2011).

35. Wickham, H. & Wickham, H. *Data Analysis* (Springer, 2016).

36. Yu, G. enrichplot: Visualization of Functional Enrichment Result. R package: https://www.bioconductor.org/packages/release/bioc/html/enrichplot.html and https://yulab-smu.top/biomedical-knowledge-mining-book (2022).

37. Wei, T. et al. Package 'corrplot'. *Statistician* **56**, e24 (2017).

38. Chen, H. & Boutros, P. C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinforma.* **12**, 1–7 (2011).

39. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

40. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).

41. Maroun, C. R. & Rowlands, T. The Met receptor tyrosine kinase: a key player in oncogenesis and drug resistance. *Pharmacol. Ther.* **142**, 316–338 (2014).

42. Higashi, T., Friedman, S. L. & Hoshida, Y. Hepatic stellate cells as key target in liver fibrosis. *Adv. Drug Deliv. Rev.* **121**, 27–42 (2017).

43. Afdhal, N. H. Fibroscan (transient elastography) for the measurement of liver fibrosis. *Gastroenterol. Hepatol.* **8**, 605 (2012).

44. Rodenburg, W. S. & van Buul, J. D. Rho GTPase signalling networks in cancer cell transendothelial migration. *Vasc. Biol.* **3**, R77–R95 (2021).

45. Feng, J. et al. mTOR: a potential new target in nonalcoholic fatty liver disease. *Int. J. Mol. Sci.* **23**, 9196 (2022).

46. Qi, X. et al. Clinical significance and therapeutic value of glutathione peroxidase 3 (GPx3) in hepatocellular carcinoma. *Oncotarget* **5**, 11103 (2014).

47. Rostami, S. & Parsian, H. Hyaluronic acid: from biochemical characteristics to its clinical translation in assessment of liver fibrosis. *Hepat. Mon.* **13**, e13787 (2013).

48. Fabregat, I. et al. TGF-β signalling and liver disease. *FEBS J.* **283**, 2219–2232 (2016).

49. Dobie, R. et al. Single-cell transcriptomics uncovers zonation of function in the mesenchyme during liver fibrosis. *Cell Rep.* **29**, 1832–1847.e1838 (2019).

50. Vengrenyuk, Y. et al. Cholesterol loading reprograms the microRNA-143/145–myocardin axis to convert aortic smooth muscle cells to a dysfunctional macrophage-like phenotype. *Arterioscler. Thromb. Vasc. Biol.* **35**, 535–546 (2015).

51. Cai, X., Zhou, J., Deng, J. & Chen, Z. Prognostic biomarker SMARCC1 and its association with immune infiltrates in hepatocellular carcinoma. *Cancer Cell Int.* **21**, 701 (2021).

52. Patil, N. Y., Friedman, J. E. & Joshi, A. D. Role of hepatic aryl hydrocarbon receptor in non-alcoholic fatty liver disease. *Receptors* **2**, 1–15 (2023).

53. Hsu, K.-S. & Kao, H.-Y. PML: regulation and multifaceted function beyond tumor suppression. *Cell Biosci.* **8**, 1–21 (2018).

54. Yang, H., Gao, S., Chen, J. & Lou, W. UBE2I promotes metastasis and correlates with poor prognosis in hepatocellular carcinoma. *Cancer Cell Int.* **20**, 1–13 (2020).

55. Xu, F., Liu, C., Zhou, D. & Zhang, L. TGF-β/SMAD pathway and its regulation in hepatic fibrosis. *J. Histochem. Cytochem.* **64**, 157–167 (2016).

56. Selman, L. & Hansen, S. Structure and function of collectin liver 1 (CL-L1) and collectin 11 (CL-11, CL-K1). *Immunobiology* **217**, 851–863 (2012).

57. Simons, N. et al. Patients with aldolase B deficiency are characterized by increased intrahepatic triglyceride content. *J. Clin. Endocrinol. Metab.* **104**, 5056–5064 (2019).

58. Liu, G. et al. Fructose-1, 6-bisphosphate aldolase B depletion promotes hepatocellular carcinogenesis through activating insulin receptor signaling and lipogenesis. *Hepatology* **74**, 3037–3055 (2021).

59. Livingstone, C. IGF2 and cancer. *Endocr. Relat. Cancer* **20**, R321–R339 (2013).

60. Yoon, Y. S., Keum, N., Zhang, X., Cho, E. & Giovannucci, E. L. Circulating levels of IGF-1, IGFBP-3, and IGF-1/IGFBP-3 molar ratio and colorectal adenomas: a meta-analysis. *Cancer Epidemiol.* **39**, 1026–1035 (2015).

61. Probst-Hensch, N. et al. IGF-1, IGF-2 and IGFBP-3 in prediagnostic serum: association with colorectal cancer in a cohort of Chinese men in Shanghai. *Br. J. Cancer* **85**, 1695–1699 (2001).

62. Sanches, J. M., Zhao, L. N., Salehi, A., Wollheim, C. B. & Kaldis, P. Pathophysiology of type 2 diabetes and the impact of altered metabolic interorgan crosstalk. *FEBS J.* **290**, 620–648 (2023).

63. Skytthe, M. K. et al. Obese patients with nonalcoholic fatty liver disease have an increase in soluble plasma CD163 and a concurrent decrease in hepatic expression of CD163. *Gastro Hep Adv.* **2**, 711–720 (2023).

64. Romacho, T., Elsen, M., Röhrborn, D. & Eckel, J. Adipose tissue and its role in organ crosstalk. *Acta Physiol.* **210**, 733–753 (2014).

65. Ooi, G. J. et al. Modified thresholds for fibrosis risk scores in nonalcoholic fatty liver disease are necessary in the obese. *Obes. Surg.* **27**, 115–125 (2017).

66. Loft, A. et al. Liver-fibrosis-activated transcriptional networks govern hepatocyte reprogramming and intra-hepatic communication. *Cell Metab.* **33**, 1685–1700.e1689 (2021).

67. Perez-Riverol, Y. et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).

68. Boel, F. Deep-Proteome-Profiling-MASLD (v1.0.0). *Zenodo* https://doi.org/10.5281/zenodo.14904437 (2025).

Acknowledgements

Author contributions

Conceptualization: F.B. and B.B.; Proteomics laboratory work: F.B., V.A., M.T., and P.H.; Clinical work: C.W.W. and M.M.L.; Cell type deconvolutions data: M.K.T. and F.T.L.; Data analysis and figures: F.B.; Supervised the study: B.B., M.M.L., A.K., K.R., and S.M.; Manuscript preparation: F.B.; All authors provided critical feedback to the manuscript and approved its final version.

Competing interests

The authors declare no competing interests.

Additional information