livers

MDPI

Article

Advancing Chronic Liver Disease Diagnoses: Targeted Proteomics for the Non-Invasive Detection of Fibrosis

Andrea Villanueva Raisman [1,2], David Kotol [3], Ozlem Altay [1,2], Adil Mardinoglu [1,2], Dila Atak [4], Cihan Yurdaydin [5], Murat Akyildiz [5], Murat Dayangac [6], Hale Kirimlioglu [7], Müjdat Zeybel [5,6] and Fredrik Edfors [1,2,*]

1 SciLifeLab, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden; andrea.villanueva@scilifelab.se (A.V.R.); havva.altay@scilifelab.se (O.A.); adilm@scilifelab.se (A.M.)
2 Department of Protein Science, Division of Systems Biology, School of Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden
3 ProteomEdge AB, 106 91 Stockholm, Sweden
4 Translational Medicine Research Center, Koç University, Istanbul 34010, Turkey; dilatak@ku.edu.tr
5 Department of Gastroenterology and Hepatology, School of Medicine, Koç University, Istanbul 34450, Turkey; cyurdaydin@ku.edu.tr (C.Y.); makyildiz@kuh.ku.edu.tr (M.A.); mzeybel@ku.edu.tr (M.Z.)
6 Department of General Surgery, Faculty of Medicine, Medipol University, Istanbul 34010, Turkey; mdayangac@medipol.edu.tr
7 Department of Pathology, School of Medicine, Acibadem Mehmet Ali Aydinlar University, Istanbul 34752, Turkey
* Correspondence: edfors@kth.se

Abstract: Chronic liver disease poses significant challenges to healthcare systems, which frequently struggle to meet the needs of end-stage liver disease patients. Early detection and management are essential because liver damage and fibrosis are potentially reversible. However, the implementation of population-wide screenings is hindered by the asymptomatic nature of early chronic liver disease, along with the risks and costs associated with traditional diagnostics, such as liver biopsies. This study pioneers the development of innovative, minimally invasive methods capable of improving the outcomes of liver disease patients by identifying liver disease biomarkers using quantification methods with translational potential. A targeted mass spectrometry assay based on stable isotope standard protein epitope signature tags (SIS-PrESTs) was employed for the absolute quantification of 108 proteins in just two microliters of plasma. The plasma profiles were derived from patients of various liver disease stages and etiologies, including healthy controls. A set of potential biomarkers for stratifying liver fibrosis was identified through differential expression analysis and supervised machine learning. These findings offer promising alternatives for improved diagnostics and personalized treatment strategies in liver disease management. Moreover, our approach is fully compatible with existing technologies that facilitate the robust quantification of clinically relevant protein targets via minimally disruptive sampling methods.

Keywords: chronic liver disease (CLD); fibrosis biomarkers; targeted proteomics; plasma proteome profiling; mass spectrometry

Academic Editors: Ralf Weiskirchen and Tilman Sauerbruch

Received: 21 November 2024
Revised: 5 January 2025
Accepted: 8 January 2025
Published: 14 January 2025

Citation: Villanueva Raisman, A.; Kotol, D.; Altay, O.; Mardinoglu, A.; Atak, D.; Yurdaydin, C.; Akyildiz, M.; Dayangac, M.; Kirimlioglu, H.; Zeybel, M.; et al. Advancing Chronic Liver Disease Diagnoses: Targeted Proteomics for the Non-Invasive Detection of Fibrosis. *Livers* **2025**, *5*, 2. https://doi.org/10.3390/livers5010002

1. Introduction

Liver diseases claim approximately 2 million lives globally every year [1], representing a significant clinical and economic challenge to healthcare systems worldwide [2,3]. Most of the cases are attributed to cirrhosis and hepatocellular carcinoma (HCC), followed by a smaller fraction from acute liver failure [4–7]. Chronic liver disease (CLD) is a broad

term, which includes many disease etiologies, such as alcohol-related liver disease (ARLD), metabolic dysfunction-associated steatotic liver disease (MASLD), and chronic viral hepatitis (VHB). The common ground among these disease etiologies is that they tend to progress in sequential stages, which do not necessarily have distinct bounds, making the classification challenging [1,4]. CLD progresses from hepatic steatosis (steatotic liver), through hepatitis (which can also start without steatosis), to fibrosis (excessive buildup of extracellular matrix), and finally to cirrhosis (severe liver scarring) [4]. Normally, fibrogenesis is meant to maintain the integrity of tissue during the wound-healing response [8]. However, when it is progressive or chronic, it may lead to the disruption of the liver parenchymal architecture. This progression is often slow and asymptomatic, delaying the diagnosis until advanced stages, like cirrhosis or HCC, are reached [8–10], which together account for about 3.5% of all global deaths [6].

The liver, the largest internal organ in the body, has a complex functionality [11], including nutrient metabolism, the synthesis of diverse biomolecules (e.g., proteins, hormones), detoxification, fat digestion, red blood cell breakdown and removal, and the immune system front line [2,11,12]. This functional intricacy is achieved by a composite, multicellular tissue, which has not yet been replicated by any artificial systems [13]. Liver transplantation is the only solution for long-term survival.

Despite the high mortality rate of CLD, early and correct management of the disease cannot only stop or decelerate fibrosis progression, but even reverse it, reducing the possibility of decompensated cirrhosis, HCC, and liver failure [4]. With a timely diagnosis, appropriate measures can be taken to hinder disease progression and increase survival chances. These measures can go from lifestyle changes (e.g., alcohol abstinence, increased physical activity) to antiviral therapy for hepatitis B or C, diabetes treatment, and surveillance for HCC [9]. An early diagnosis enables effective interventions that slow liver disease progression and enhance survival.

However, the often slow and symptom-free progression of these diseases makes early detection challenging [14]. Liver fibrosis, a critical structural and functional change in CLD, helps predict the progression from a healthy liver to cirrhosis. While non-invasive methods are helpful, the histological analysis of liver biopsies has traditionally been considered the gold standard for evaluating necro-inflammation and hepatic fibrosis [4,15]. These analyses rely on systematic, semi-quantitative measurements to grade and stage the histologic progression of the disease [16]. Notably, liver biopsies have significant limitations. Besides technical difficulties, it is an expensive and invasive procedure, which poses the risk of life-threatening complications [8,14,15]. Other limitations include that (i) it does not provide insight into the dynamic changes of fibrogenesis, (ii) it only provides information for a comparatively small part of the liver, risking not being representative of the real level of fibrosis, which has a heterogeneous distribution, and (iii) there is inherent variability in the interpretation between pathologists. Not being apt for population-wide screenings, only a small number of patients are identified as being at risk [8]. Furthermore, other clinical complications can emerge as fibrosis progresses, and fibrosis staging does not necessarily reflect these outcomes [16]. This presents a strong case to move to less invasive and more informative tests to improve individual patient outcomes and make population-wide screening more feasible and reduce the clinical and economic burden for healthcare systems [9].

This has led to the move towards minimally invasive procedures, such as blood analysis and imaging [8]. Blood is a key source for molecular analyses in both clinical and research settings [17]. It holds immense potential for biomarker discovery and therapeutic targeting. Currently, some non-invasive methods for the assessment of liver fibrosis are used in patients for first-line assessments [15]. However, biomarker-discovery efforts

have centered on single molecules [14]. This has been an obstacle to their success since the multiplicity of the liver's biochemical, synthetic, and excretory functions cannot be captured by a single biomarker [18].

Mass spectrometry enhances biomarker discovery by allowing for the simultaneous quantitative analysis of multiple targets with high specificity. It is particularly effective in proteomic research, where it is used to gather comprehensive data across different biological systems. Mass spectrometry-based proteomics are suitable to obtain system-wide information, which can be used in hypothesis testing and for differential expression analysis, as well as supervised machine learning. These algorithms can leverage comprehensive data to generate disease-classification models and extract features suitable for a patient's health assessment. This makes mass spectrometry a fitting strategy to be used in healthcare and biomarker discovery. In this context, the integration of targeted proteomics and specifically selected reaction monitoring (SRM) with stable isotope standard protein epitope signature tags (SIS-PrESTs) presents a relevant solution for the highly specific and reproducible quantification of proteins [19]. This high-throughput method consists of using a predefined multiplex panel of SIS-PrESTs, targeting a set of medium-to-high-abundant blood plasma proteins. SIS-PrESTs are recombinant internal protein standards that are spiked in known quantities into patients' blood plasma samples and digested altogether to peptides prior to MS analysis. From the extracted chromatograms, the ratio of areas under the curves from endogenous peptides to SIS-PrEST signals are calculated and converted to absolute protein quantities. The integration of this technology into clinical practice could provide clinicians with a more precise and comprehensive liver health assessment, allowing for early fibrosis and CLD treatment while being minimally invasive and affordable.

2. Materials and Methods

This study is part of the Human Disease Blood Atlas (HDBA), received approval from the Swedish Ethical Review Authority (EPM dnr 2019-00222), and adheres to relevant ethical regulations. Collection procedures align with participants having provided their informed written consent. The Swedish Ethical Review Board granted ethical approval, authorizing a proteomics analysis on previously collected samples.

The study protocol adheres to the ethical guidelines of the 1975 Declaration of Helsinki. For the liver disease sample cohort, collected at Koç University and Medipol University, the study received approval from the Ethics Committees of Koç University in Istanbul. Prior to clinical sample and data collection, patients were informed, and their written consent was taken, ensuring that ethical guidelines were followed across all participating centers (2015.053.IRB1.014, 2016.024.IRB2.005, 2017.139.IRB2.048, 2018.351.IRB1.043, 2022.246.IRB2.040).

2.1. Sample Collection

A total of 277 EDTA plasma aliquots were sourced, and a patient subset was randomly selected from larger cohorts from within the HDBA program. The samples were categorized according to international medical classification standards. After collection and centrifugation, they were stored at $-80\ ^{\circ}C$.

2.2. Cohort Composition and Disease Groups

The obesity cohort (OBES) included patients with a BMI > 35 kg/m^2, with or without type 2 diabetes, not diagnosed with liver disease. Individuals from the liver disease (LIVD) cohort were patients with chronic liver disease chosen based on disease etiology. The cohort is composed of (i) steatotic liver disease (SLD), (ii) other-etiology chronic liver disease (OLD), (iii) VHB, (iv) HCC, and (v) healthy-liver patients. The SLD group includes

patients with ARLD or MASLD, which involve the accumulation of excess fat in liver cells. The OLD group includes other, less common, chronic liver disease etiologies, such as primary biliary cholangitis, drug-induced liver injury, and hepatic venous outflow tract obstruction (Budd–Chiari Syndrome).

Fibrosis stratification data, based on liver histopathological assessments, were available for most individuals (n = 199) from the LIVD cohort, though no biopsies were performed on the healthy-liver or obese patients. The healthy-liver group consisted of individuals between 21 and 54 years of age with a BMI between 19 and 27.5 who did not have any indications of liver disease and who had a normal liver biochemistry, while individuals with abnormal biochemistry or FIB-4 levels higher than 1.33, as well as diabetic and obese individuals, were excluded from this group. On the other hand, the obese patients chosen for the study had no diagnoses of liver disease. For the rest of the LIVD cohort, fibrosis levels ranged from F0 to F4 in the meta-analysis of histological data in viral hepatitis (METAVIR) score and 0 to 6 in the Ishak score. Following the clinicians' stratification indications, Ishak 0 was taken as the equivalent of F0, Ishak 1 and 2 as F1; Ishak 3 as F2, Ishak 4 and 5 as F3, and Ishak 6 as F4. According to the NCI's SEER Registrar Staging Assistant, F0 to 3 indicate no-to-moderate fibrosis, while F4 indicates severe fibrosis and the transition to cirrhosis. Based on this, samples were classified into four fibrosis groups, where F0 was taken as no fibrosis, F1 and F2 as mild fibrosis, F3 as moderate fibrosis, and F4 as severe fibrosis. For the analyses, the healthy-liver and obesity patients were taken as having F0. Most patients from the OLD, VHB, and HCC groups presented severe fibrosis (F4), while the SLD group was more mixed.

2.3. Sample Preparation

A plasma digestion protocol with spiked-in SIS-PrESTs was adopted by the HDBA from Geyer et al. (2016) [20]. Proteins were quantified using SIS-PrESTs mixed in near-healthy levels [19], forming a synthetic heavy-labeled plasma. This mixture was aliquoted into 96-well plates, vacuum dried at 35 °C for 16 h, and stored at −20 °C [21].

For the analysis, patient plasma samples were thawed on ice for an hour and randomized into 96-well plates. Additionally, de-identified in-house plasma was added to empty wells on all plates to serve as biological replicates for intra- and inter-plate quality control. Plasma samples were diluted 10-fold in 1× phosphate-buffered saline (PBS) (Sigma Aldrich, St. Louis, MI, USA). Diluted plasma was mixed for 1:1 volume units with re-suspended SIS-PrESTs, and sodium deoxycholate (SDC) (Sigma Aldrich, St. Louis, MI, USA) and dithiothreitol (DTT) (Sigma Aldrich, St. Louis, MI, USA) were added for a final concentration of 0.66% (w/v) SDC and 10 mM DTT. This mixture was incubated at 37 °C, undergoing a reduction process for an hour, and alkylated with 50 mM chloroacetamide (CAA) (Sigma Aldrich, St. Louis, MI, USA) in the dark for 30 min. Digestion occurred overnight using Trypsin (Thermo Fisher Scientific, Santa Clara, CA, USA) at a 1:50 enzyme-to-substrate ratio and quenched by adding trifluoroacetic acid (TFA) (Sigma Aldrich, St. Louis, MI, USA) to 0.5% (v/v). Half of the sample volume was solid-phase extracted using C18 StageTips packed in-house [22]. Solid phase extraction involved the activation of the C18 matrix (Supelco, Sigma Aldrich, St. Louis, MI, USA) with 100% acetonitrile (ACN), equilibration with 0.1% TFA, sample loading, and subsequently washing and eluting the digested peptides. Eluates were vacuum-dried at 45 °C for 30 min and stored at −20 °C upon analysis. Each plate was reconstituted in Solvent A (3% ACN, 0.1% formic acid (FA)) prior to the analysis.

2.4. LC-SRM/MS

Quantitative analysis and assay development were conducted using an UltiMate 3000 nano-LC system (Thermo Fisher Scientific, Santa Clara, CA, USA) equipped with

an EASY-Spray ion source and coupled to a TSQ Altis mass spectrometer (Thermo Fisher Scientific, Santa Clara, CA, USA). The samples were loaded onto an Acclaim PepMap 100 trap column (75 µm × 2 cm, C18, 3 µm, 100 Å, Thermo Fisher Scientific, Santa Clara, CA, USA) and washed for 0.75 min at a flow rate of 15 µL/min with 1% Solvent B (95% ACN, 0.1% FA, water). Peptides were separated using an analytical PepMap RSLC C18 column (150 µm × 15 cm, 2 µm, 100 Å, Thermo Fisher Scientific, Santa Clara, CA, USA) and a gradient from 1% to 30% Solvent B over 29.25 min at 3 µL/min. The columns were subjected to three 30 s washes with 95% Solvent B, followed by a wash with 1% Solvent B. The columns were re-equilibrated for 1.4 min with 1% Solvent B. Total cycle times, including sample loading, analysis, and column re-equilibration, were 15 min for method development and 35 min for plasma quantification. Column oven temperatures were set at 40 °C for the analytical column and 60 °C for the column oven, with the autosampler temperature held at 10 °C.

Totals of 125 proteins and 294 peptides were targeted (Supplementary Table S1) using a peptide assay library provided by ProteomEdge (ProteomEdge AB, Stockholm, Sweden) for peptide identification and quantification. The library was developed towards the SIS-PrESTs sequences using their proprietary qRePS (quantitative recombinant protein standard) technology. The heavy signals from all SIS-PrEST peptides were used as internal standards for the one-point calibration and calculation of the light-to-heavy peptide ratios and subsequent absolute quantification.

2.5. Data Pre-Processing

The raw data from the mass spectrometer were imported into Skyline (beta version 23.1.1.459) [23] for peak integration and initial quality control. During this process, the areas under the curve (AUCs) for both heavy and light peptides were extracted, enabling relative quantification through the calculation of ratios between the heavy and light signal intensities. Using ratios as a measure of the protein concentration—as opposed to using the AUC of the endogenous protein directly—mitigates the variability commonly encountered in mass spectrometry. Annotated reports were exported from Skyline for downstream bioinformatic analyses in R (version 4.3.1) [24]. These reports indicate the sample, peptide being measured, protein it belongs to, and light-to-heavy peptide ratio, which can later be used for a calculation of the absolute concentration. This is achieved through a simple transformation of the endogenous-to-standard protein ratios into blood plasma concentrations by multiplying the ratios by the known spiked-in amounts of each SIS-PrEST. The LC-SRM/MS data were subjected to quality control using the heavy SIS-PrEST signals for filtering out datapoints with a library dot product (dotp) or a peptide peak found ratio (rdotp) below 0.8. Also, peptides identified in fewer than 50% of the samples were excluded, together with samples where fewer than 80% of the targeted standards were detected.

The in-house biological control samples distributed among the four plates were used to perform a batch effect analysis, based on which the harmonizR function from the HarmonizR package (v 1.0.0) [25] was used to correct the detected batch effects, using Combat_mode = 2. After quality control and batch effect correction, the quantitative data used for the analyses were the ratios of the light-to-heavy precursors. The median of the peptide ratios from a single protein was then used as the final representative measurement for that protein in downstream analyses. Missing values were imputed using the hotdeck() function from the VIM package (v 6.2.2) [26].

2.6. Differential Expression Analysis

The differential expression of proteins was estimated with two-sided *t*-tests. Multiple hypothesis correction was performed using a false discovery rate (FDR) adjustment, with

a significance cutoff of $q < 0.05$. Volcano plots were used to summarize the q-values and expression fold-changes for each of the disease groups compared to the rest.

2.7. Machine Learning and Classification

There were three approaches to generate the classification models. The binary models included (i) classifying patients with any level of fibrosis against those deemed healthy or with no fibrosis (F0) and (ii) classifying patients with a specific level of fibrosis against those deemed healthy or with no fibrosis (F0), while the multi-classification model consisted of classifying patients with different fibrosis levels simultaneously.

The models were generated using the tidymodels package (v 1.1.1) [27]. In every case, the class sizes were balanced. This was determined based on the group with the fewest samples. For instance, in the mild fibrosis model, the subset of data consisted of all samples belonging to patients from the mild liver fibrosis group ($n = 30$) and a same-sized group of samples, which were a random mix of all the F0, healthy-liver, and obesity cohort patients, obtained with the basic sample() function (v 4.3.1) [26]. Then, 70% of this data was used for training and 30% was used for testing, using the initial_split() function in tidymodels. This data size balance and split were the same for all models. Overall fibrosis and each level of fibrosis were classified against the healthy and obesity controls and the samples marked with no fibrosis, always balancing group sizes. For the multiclassification model, patients were put into four groups, which consisted of no fibrosis (healthy, obese and F0), mild fibrosis (F1 and F2), moderate fibrosis (F3), and severe fibrosis (F4). These were simultaneously classified against each other.

The classification models were generated with the selected training sets using a pipeline of tidymodel functions. First, a recipe was created using recipe(), which specifies the data set that will be used and the column with the true classification information. Here, step_corr() is used to eliminate proteins with a correlation of more than 0.8. Then, a model was specified where rand_forest() (trees = 500) was the chosen classification algorithm, set_engine() was used to determine the system to fit the model, with "ranger" as the engine and "permutation" as the mode to determine feature importance, and set_mode() was used to establish that the model should be a "classification" one. After this, a workflow() container is created that uses add_recipe() and add_model() to include the previously formulated recipe and parsnip model. Finally, the fit() [23] function takes the workflow container with all the specifications and uses the chosen data to train the model. Each protein's contribution was retrieved using the vi_model() function from the vip package (v 0.4.1) [28].

Random forest is an algorithm where multiple decision trees are generated during training, so the result of the classification task is the one reached by most trees [29,30]. When generating classification models, random forest estimates each protein's importance for that model. Permutation-based importance measures the decrease in model performance (e.g., accuracy) when the values of a feature are randomly shuffled. Thus, it quantifies the importance of a feature by assessing how much the model's predictive performance deteriorates when the relationship between that feature and the target is disrupted [31]. For all models, the input data were iteratively restricted to a subset of proteins, where the pipeline was run and the proteins with negative permutation importance were eliminated until the model only contained proteins with positive scores. Then, thresholds were used to keep only the topmost important proteins that could maintain at least the same performance as the model with all the positively scoring proteins. These changes were made on the recipe step, using the functions remove_role() and update_role() to eliminate the previous proteins and add the new subset of proteins.

The predict() function from the stats package (v 4.3.1) [24] was used to estimate each of the test sets samples' probabilities of belonging to a given class. Since these samples were not used for training, they can be used to determine model performance. The sensitivity and specificity of the classifications were evaluated with confusion matrices, performed with the confusionMatrix() function from caret (v 6.0.94) [32], with ROC analyses with AUC scores using the roc() function from pROC (v 1.18.5) [33].

2.8. Data Visualization

Graphs were generated with R (version 4.3.1) [24], using the ggplot2 (v 3.5.0) [34], ggpubr (v 0.6.0) [35], factoextra (v 1.7.0) [36], ggstatsplot (v 0.12.2) [37], UpSetR (v 1.4.0) [38], ggrepel (version 0.9.5) [39], and patchwork (v 1.2.0) [40] packages.

2.9. Data Availability

The data supporting the findings of this study are available upon request, subject to a review for validation purposes. We are committed to facilitating access within the limits established by patient consent agreements.

3. Results

This study analyzed a diverse cohort of 277 plasma samples in which 125 target proteins were absolutely quantified with 294 peptides using SIS-PrESTs internal protein standards. The absolute concentration data were processed to select potential biomarkers and classifiers for liver disease patients' stratification (Figure 1a). After stringent quality control, 243 samples and 212 peptides from 108 proteins were selected for further analysis. Measuring more than one peptide per protein is a way to ensure that the protein has indeed been detected. Visual inspection of the data in Skyline and quality control steps make it possible to determine if the measurements of different peptides support each other. After filtering low-quality samples, the liver disease (LIVD) cohort included patients with HCC (n = 54), chronic VHB (n = 59), OLD (n = 28), and SLD (n = 57) who had been stratified by fibrosis levels (METAVIR F0 to F4, Ishak 0 to 6), and healthy-liver controls (n = 31), while the OBES cohort included individuals with a BMI > 35 (n = 14), with or without diabetes, but no diagnosis of liver damage. The proteomic data further supported this assumption, as the profile of the OBES group was markedly distinct from that of the liver disease patients and aligned more closely with the healthy-liver group. Thus, both the healthy-liver and the obese patients were taken as controls (F0). Figure 1b summarizes the age, BMI, and sex distribution for each fibrosis group (healthy, mild-to-moderate, and severe).

3.1. Identification of Fibrosis-Specific Biomarkers Through Differential Expression Analysis

Fibrosis progression proteome profiling was performed through differential expression analyses. Each fibrosis group (mild, moderate, and severe) was compared to the control group. The summary of the proteins up- or downregulated at each fibrosis stage is represented by volcano plots (Figure 2a). Most of the proteins downregulated with mild or moderate fibrosis continue to be downregulated at the next stage of fibrosis (Figure 2b). This analysis revealed a pattern of increasing dysregulation with more advanced fibrosis, with the fold-changes in protein expression becoming more pronounced as the disease progressed. A clear example of both patterns is apolipoprotein M (APOM), a well-known liver-enriched protein, which is specific to hepatocytes and has been associated with high-density lipoproteins (HDLs) according to the Human Protein Atlas (HPA) (v23, www.proteinatlas.org) [41]. On the other hand, while most of the differentially expressed proteins are downregulated, a handful of them show elevated levels in advanced fibrosis. Such is the case for von Willebrand Factor (VWF), a protein involved in blood coagulation, hemostasis, and cell adhesion, for which upregulation has been associated with an

increased risk of arterial thrombosis [42]. Thus, the analysis of protein expression in liver fibrosis reveals key changes in both upregulated and downregulated proteins, shedding light on their diverse roles in disease progression and liver-specific impacts as the fibrosis severity increases. The differentially expressed proteins for the different fibrosis levels are included in the Supplementary Materials (Supplementary Table S2).

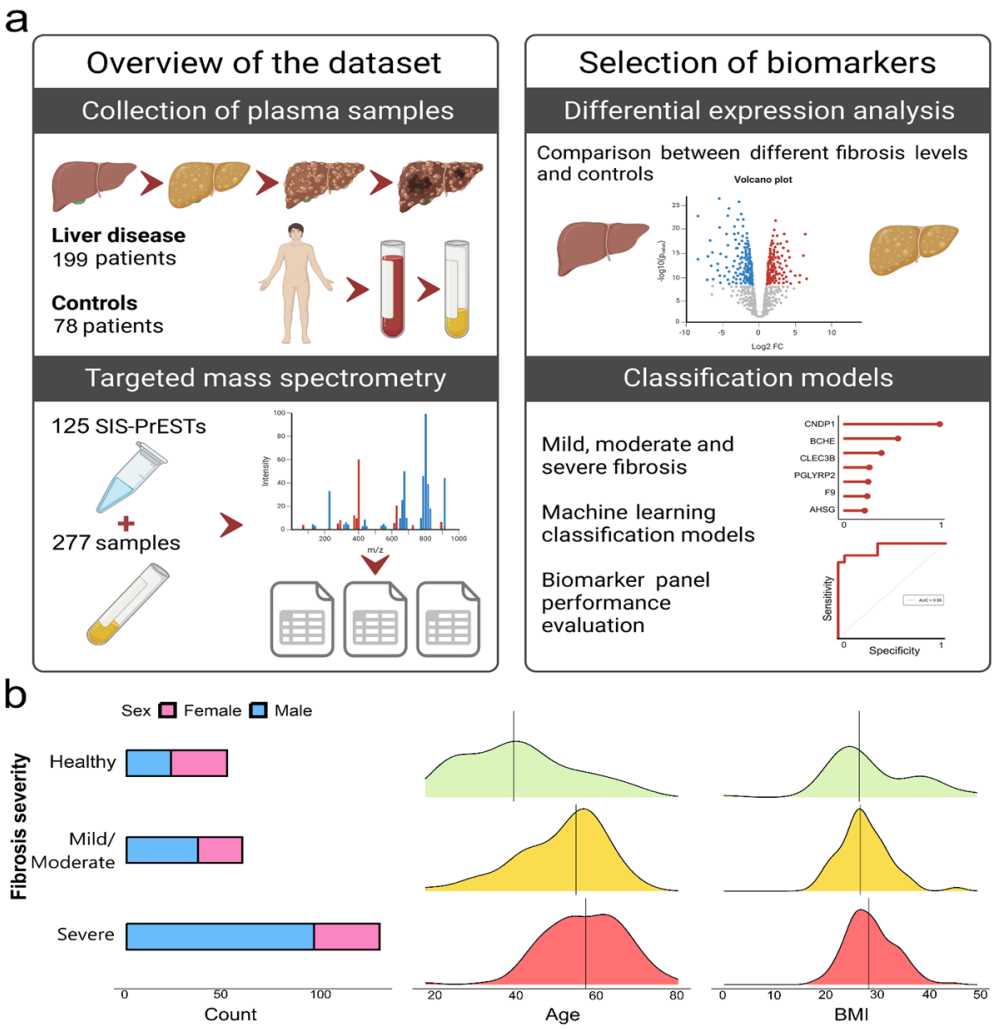

Figure 1. Overview of the study. (**a**) Summarization of the study. Targeted mass spectrometry and SIS-PrESTs were used to profile the plasma proteomes from 277 liver disease and 78 liver-healthy individuals, which were analyzed. Skyline was the first step for data analysis. Peptide intensity spectra were visually inspected, and quantitative reports were exported for downstream bioinformatic analyses. Liver fibrosis samples were compared to the controls using differential expression analysis to find statistical differences, summarized by volcano plots. Important feature selection was performed through classification models, which were generated using a random forest algorithm trained on 70% of the data. Model performance was assessed on the other 30% of the data using confusion matrices and ROC curves (shown). (**b**) Sample distribution of patient samples from Healthy, Mild/Moderate, or Severe Fibrosis.

Figure 2c shows the number of up- and downregulated proteins as the fibrosis severity increases. In total, 84 out of 108 proteins were differentially expressed in at least one level of fibrosis compared to the controls. A functional gene ontology (GO) analysis reveals that the downregulated proteins are involved in wound healing and the wound response, coagulation and fibrinolysis, hemostasis, regulation of body fluid levels, blood pressure homeostasis, immune response, and complement activation, among others (Figure S1a). This is in line with liver disease symptoms, such as coagulopathies, edema, and an increased

risk of sepsis [6,13]. Instead, the upregulated proteins are mainly involved in coagulation, extracellular matrix assembly and organization, and cell–substrate adhesion (Figure S1b). According to the HPA tissue specificity classification, most of the downregulated proteins are enriched or enhanced in the liver. This reflects how CLD progresses regardless of its etiology. The liver fibrosis stratification indicates the level of structural and functional alterations in the liver. Mild fibrosis is essentially asymptomatic, while once fibrosis transitions into cirrhosis and cancer emerges, the liver becomes severely impaired.

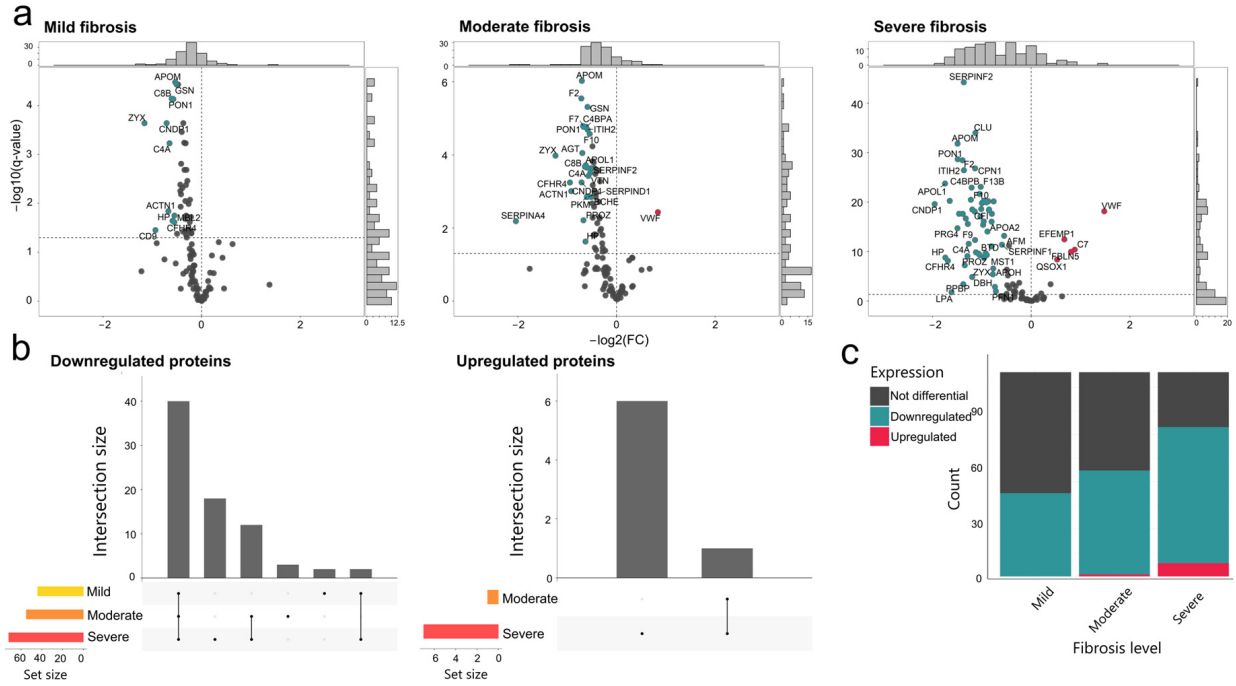

Figure 2. Differential expression analysis. (**a**) Volcano plots summarizing the results for differentially expressed proteins for each fibrosis level against the controls (healthy, obese, and F0), with the fold-change (FC) on the x-axis and the significance (q-value) on the y-axis. Each dot represents a protein. Two-sided t-tests were applied for each protein and a q-value obtained after FDR-correction. The proteins above the established level of significance (q-value > 0.05) and of expression ($\log_2$ (FC) > 0.5) are marked with red (upregulated) or blue (downregulated). (**b**) Upset plots show the number of upregulated and downregulated proteins that overlap between fibrosis levels. The bars on the side show the number of differentially expressed proteins per fibrosis level and, on top, how many of them overlap between fibrosis levels. (**c**) Bar plot with the counts of significantly upregulated (red) and downregulated (blue) proteins for each fibrosis level.

We performed dimensionality reduction using principal component analysis (PCA) to identify proteins driving the protein profile changes. Figure 3a shows the PCA plot (PC1 vs. PC2) for all samples with the distinct clustering of data points, indicating specific proteins that differentiate the samples based on the progression of the disease, particularly in severe cases.

A GO analysis indicated that the most critical proteins in PC1 are mainly involved in the negative regulation of fibrinolysis and positive regulation of wound healing and hemostasis. The four proteins involved in these processes, vitronectin (VTN), coagulation factor II (F2), alpha-2-antiplasmin (SERPINF2), and kininogen 1 (KNG1), are significantly and increasingly downregulated in mild, moderate, and severe fibrosis compared to the controls. Their low levels could lead to increased fibrinolysis and decreased wound healing, which is in line with symptoms, such as excessive bleeding, observed in advanced liver disease patients [43]. The same GO search does not render any significant results for the variables of PC 2. However, the proteins in PC3 are shown to be involved in inflamma-

tion processes—such as neutrophil aggregation and the chronic and acute inflammatory response (Figure 3b). In fact, neutrophil aggregation can cause pseudoleukopenia, a condition in which the neutrophils present in the body migrate towards a specific site, making it look like there is a decrease in the circulating leukocytes, despite the bone marrow's later production of more white blood cells. Liver disorders have been observed to be an underlying cause of this condition [44]. Thus, the most important proteins in the PCA seem to have biological relevance.

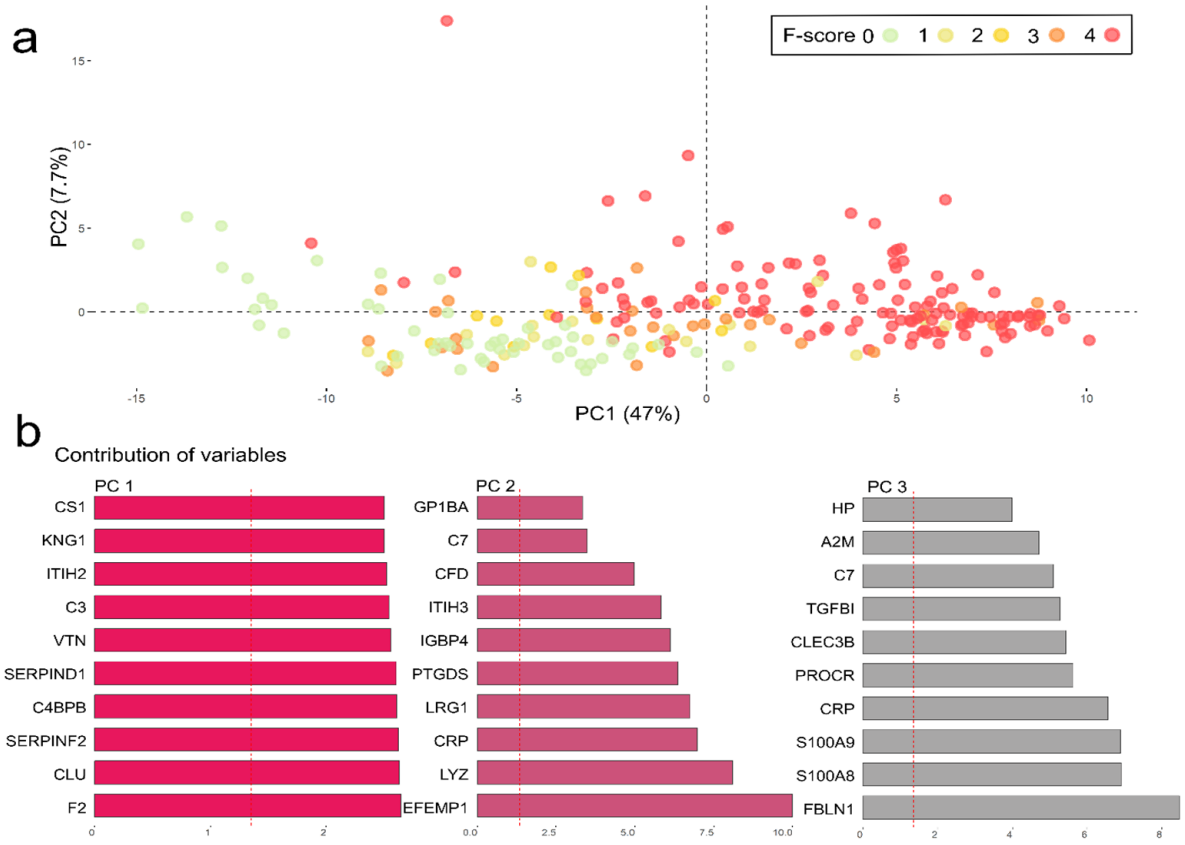

Figure 3. Principal component analysis. (**a**) PCA plot for PC 1 and 2. The data points (n = 243), representing all samples, are colored by fibrosis severity (red = severe fibrosis, orange = moderate fibrosis, yellow = mild fibrosis, green = no fibrosis). (**b**) Top ten contributing proteins for PCs 1, 2, and 3.

3.2. Fibrosis Classification Models

While the PCA provides valuable insights into the overall variance and patterns in the data, it is not specifically tailored to identify biomarkers that are directly associated with fibrosis severity. A supervised machine learning strategy was used to build models to identify the most relevant proteins to classify samples as belonging to the control group or to the mild, moderate, or severe fibrosis patients. All proteins were used as input (n = 108), while a 70:30 split of the samples was used to establish training and testing sets (Figure 4). In each case, the model was trained with a weighted number of samples per group. Thus, the sample size of the smallest group would determine how many samples were to be included from the larger groups to ensure a balanced set. Random forest was used to establish a fibrosis model based on protein levels; a same-sample size mix of all fibrosis levels was classified against the control group of F0, healthy-liver, and obese patients with the goal of separating patients as presenting any level of fibrosis or having a fibrosis-free liver.

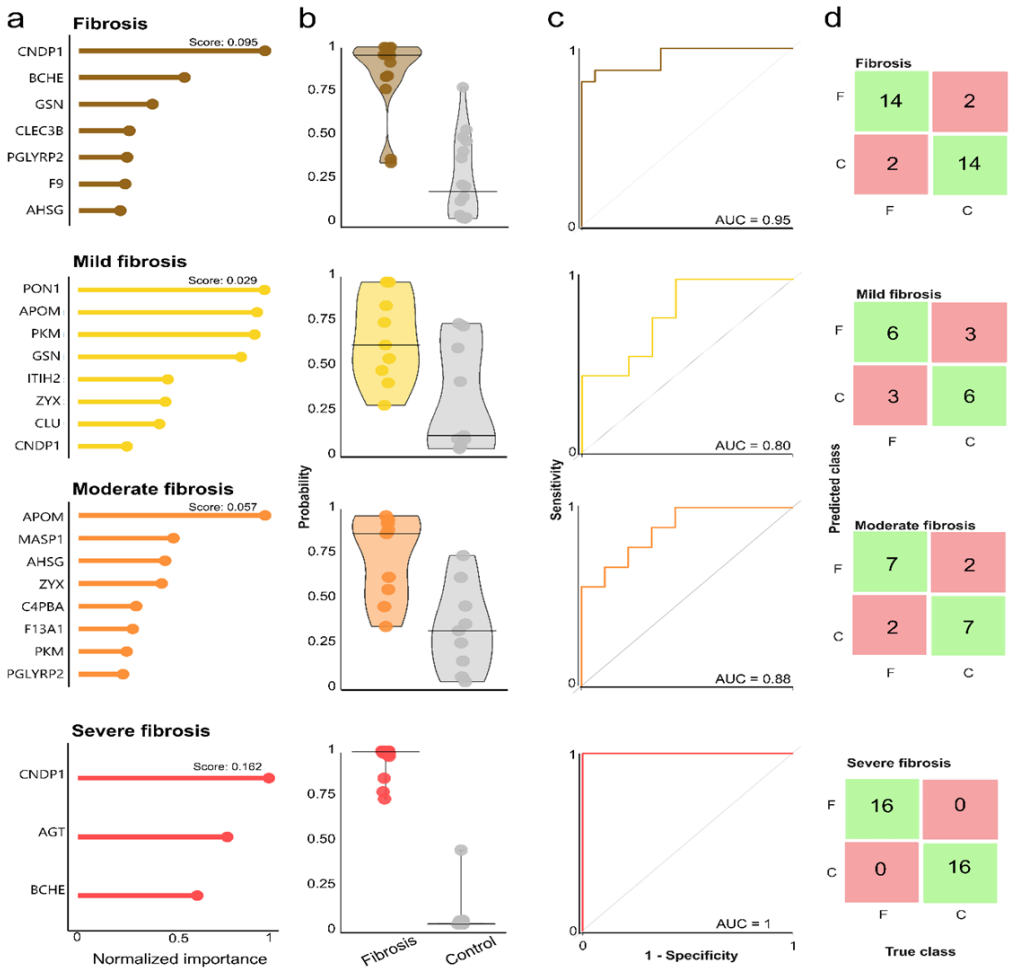

Figure 4. Fibrosis models' selected features and performance. (**a**) Lollipop charts of the selected proteins per model with their normalized importance as a fraction of the top scoring protein, for which the permutation score is shown at the top of each plot. (**b**) Probabilities of general fibrosis or specific fibrosis levels for the samples in the test set. (**c**) ROC curves with AUCs to indicate model performance. (**d**) Confusion matrices with classification results for each model, where F = fibrosis group and C = control group.

The features (i.e., proteins) selected for each classification model are shown in Figure 4a. The objective of these models was to obtain a minimal protein panel that maintained at least the same sensitivity and specificity as previous iterations, which included more proteins. The starting points were models that comprised all the proteins with positive permutation importance, excluding highly correlated proteins (>0.8) for minimally redundant biomarkers (Figure S3 illustrates the protein correlation in this study). The most important proteins for the models include examples, such as (i) carnosine dipeptidase 1 (CNDP1), involved in the metabolism of dipeptides; (ii) paraoxonase 1 (PON1), which hydrolyzes toxic metabolites; (iii) APOM, which is involved in lipid transportation; and (iv) angiotensinogen (AGT), which regulates blood pressure [41]. PON1, APOM, and AGT are enriched in the liver and CNDP1 in the brain. The functions associated with these proteins, with all of their levels seeming to drop as fibrosis progresses, reflect some of the symptoms observed in chronic liver disease patients, such as abnormalities in the regulation of blood pressure and cognitive symptoms related to the accumulation of toxic metabolites (i.e., confusion or drowsiness) [45].

The performance of the classification models was tested on the 30% of the data not used for training. The probabilities of the plasma samples belonging to the disease group for each disease model are shown in Figure 4b. The first model would classify samples as having any level of fibrosis against having no fibrosis with an AUC of 0.95, while the rest could classify specific levels of fibrosis (F1-2, mild; F3, moderate; and F4, severe) against no fibrosis with an AUC between 0.80 and 1 (Figure 4c). For these last three models, the model performance and the distance between the classification probabilities for the disease and the healthy samples increase with the fibrosis severity. The confusion matrices show the number of samples from the test set correctly and incorrectly classified by the model (Figure 4d). The results show high specificity and sensitivity, especially for the general fibrosis and the severe fibrosis models, despite sample size limitations, which should be addressed to further validate the models. On the other hand, the results for the mild fibrosis model were poorer than for the others. This would be in line with the PCA results (Figure 3), where some of the mild fibrosis samples were closer to the controls than to the moderate or severe fibrosis samples.

3.3. Data-Driven Fibrosis Stratification

The models generated up to this point have been binary. While they could be used to determine whether a sample belongs to a liver fibrosis patient, they would not be applicable as a stratification strategy. Therefore, a multiclassification model was generated for this purpose, using the same number of samples for all groups. Figure 5a shows the minimal panel of selected features that enabled the model to maintain at least the same specificity and sensitivity as models including larger sets of proteins. The multiclassification model can classify test set samples as having a specific level of fibrosis with an AUC between 0.62 and 0.91 (Figure 5b). While only 18 of the 36 samples were correctly classified (Figure 5c), the model is placing samples next to the true class. Importantly, no healthy samples were classified as having either moderate or severe fibrosis, and no severe samples were classified as healthy. The most problematic group seemed to be the moderate fibrosis group, where only two of the samples were correctly classified, but most were assigned to another fibrosis class. Figure 5d shows the comparison between the probabilities for the different fibrosis groups of being classified as healthy. The severe fibrosis samples show the smallest probabilities of being classified as healthy, while the opposite is the case for some of the moderate samples. A relevant consideration here, beyond the adequacy of the model, is the potential variability in the biopsy-based clinical stratification itself.

The top proteins selected for this model are (i) CNDP1, involved in the metabolism of dipeptides; (ii) butyrylcholinesterase (BCHE), a liver-enriched protein that can degrade neurotoxins; (iii) gelsolin (GSN), a heart-enhanced protein involved in the modulation of actin; (iv) C-type lectin domain family 3 member B (CLEC3B), an adipose-tissue enhanced protein involved in retinal function; (v) peptidoglycan recognition protein 2 (PGLYRP2), a liver-enriched protein involved in the immune response; (vi) coagulation factor 9 (F9), a liver-enriched protein that participates in blood coagulation and hemostasis; and (vii) alpha 2-HS glycoprotein (AHSG), a liver-enriched associated with the mineral balance [41] (Figures 5e and S2). The seven selected proteins had a significant q-value in the non-parametric Kruskal–Wallis test comparing fibrosis levels (Supplementary Table S3).

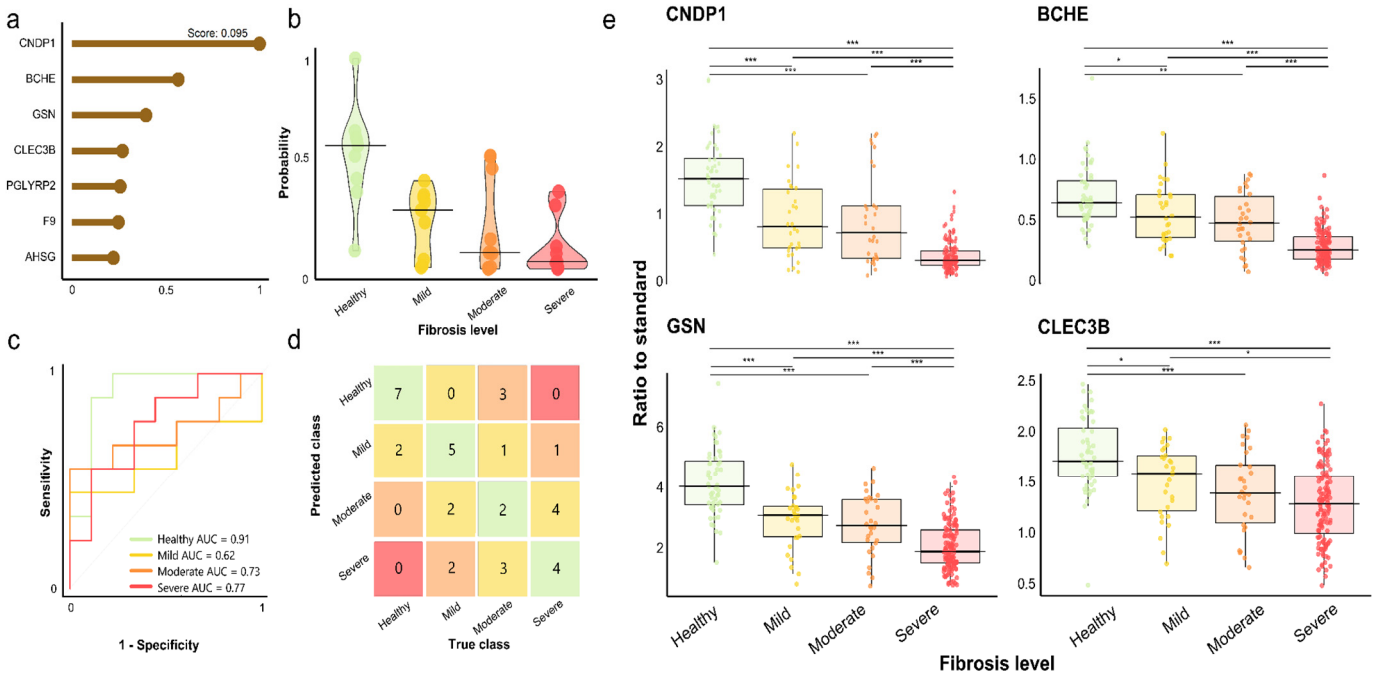

Figure 5. Multiclassification model for fibrosis stratification. (**a**) Lollipop chart of the proteins selected for the model with their normalized importance as a fraction of the top scoring protein for which the permutation importance score is shown at the top of the plot. (**b**) Probabilities for each sample of being classified as healthy by the fibrosis level group. (**c**) ROC curves and AUC show the performance of the model for each fibrosis class. (**d**) Confusion matrix with stratification classification results for the model. (**e**) Boxplots of the topmost important proteins selected for the fibrosis stratification model. The median value and first and third quartiles of the ratio to standard are summarized by the boxplots, and the whiskers mark the minimum and maximum values per group. Samples are grouped by fibrosis level (healthy, n = 52; mild fibrosis, n = 30; moderate fibrosis, n = 30; severe fibrosis, n = 131). The horizontal lines indicate which level pairwise comparisons are significant. p-values were calculated using the post-hoc Wilcox test and FDR-adjusted, and the asterisks indicate the level of significance, where * is significant (q-value < 0.05), ** is very significant (q-value < 0.01), and *** is highly significant (q-value < 0.001).

3.4. Selection of Liver Fibrosis Biomarker Panel

We outline a 20-protein panel (Figure 6) based on the features selected by the random forest models. Most of these proteins show liver specificity, except for (i) CLEC3B, enhanced in the adipose tissue; (ii) GSN, enhanced in the heart muscle; (iii) pyruvate kinase (PKM), enhanced in the skeletal muscle and the tongue; (iv) zyxin (ZYX), with low tissue specificity; (v) coagulation factor XIII A chain (F13A1), enhanced in the adipose tissue and the placenta; and (vi) CNDP1, enriched in the brain, according to the HPA. Following the criteria for the biomarker panel minimization, these proteins all had positive permutation importance and allowed each model to maintain the performance of previous model iterations that included more proteins. Furthermore, the two-sided t-tests, the ANOVA, and the post-hoc tests indicated that these proteins were significantly dysregulated at different levels of fibrosis (Tables S2 and S3). However, it is important to mention that the results of the post-hoc tests performed after the non-parametric Kruskal–Wallis test comparing fibrosis levels were not significant between mild and moderate fibrosis for most proteins, with the exception of MASP1, which was the only protein in the dataset for which this was the case. This lack of significant differences in protein levels could largely explain the problematic classification of patients with mild or moderate liver fibrosis.

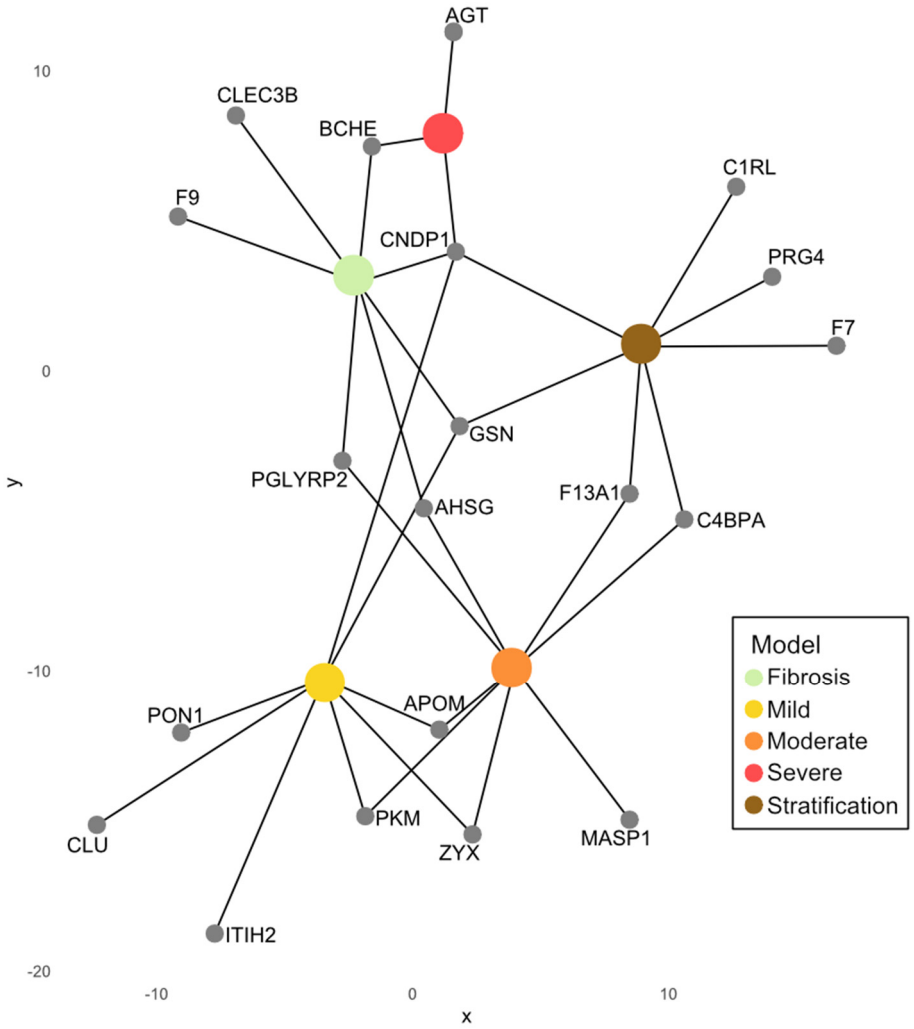

Figure 6. Liver fibrosis protein panel. Network visualization of the proteins selected for each fibrosis model. The distance from the protein to the model node is inversely proportional to the permutation importance score for the model—the longer the distance, the smaller the importance score.

4. Discussion

The high incidence of CLD emphasizes the urgent need for better diagnostic tools to improve patient outcomes, particularly given that many chronic liver diseases can be controlled. The identification and further investigation of the plasma proteins driving the progression of the disease is crucial for developing a less invasive method to assess the presence and severity of liver fibrosis, which could significantly enhance the diagnosis and management of CLD. This study presents an attractive strategy for plasma protein profiling to detect liver fibrosis at an early stage. It advances the search for optimal plasma protein biomarkers by deploying a highly multiplexable targeted mass spectrometry approach that requires small plasma volumes. We analyzed the plasma protein expression of a total of 108 proteins across various liver conditions in 243 samples and obtained reliable quantitative data. This strategy highlights the necessity of the absolute quantification of plasma proteins and facilitates the identification of protein states linked to liver fibrosis progression through differential expression analysis and machine learning techniques.

Our findings demonstrate that targeted mass spectrometry with use of heavy internal protein standards can effectively assess the plasma proteome on a system-wide scale by precisely quantifying medium-to-high-abundance proteins important for liver fibrosis and its progression. This capability is crucial for developing clinical assays that can be used

to monitor liver health, marking a significant step towards screening procedures where high-risk patients can be followed up with a confirmatory assay, such as a liver biopsy.

Previous studies from the HDBA [46,47] have shown the potential of using highly multiplexable methods for plasma proteome profiling to generate pan-disease biomarker panels, while other studies have focused on specific liver diseases [14]. This study is a continuation of the above, harnessing the pan-disease approach to study liver diseases with different etiologies and fibrosis progression comprehensively.

The overall liver fibrosis classification model, which separated F0—healthy and obese patients—from F1 to F4 samples, had high performance (AUC = 0.95). Binary classification models were also generated based on the presence of specific levels of fibrosis. The moderate and severe fibrosis models showed good specificity and sensitivity (AUC = 0.80 and AUC = 0.88), while the severe fibrosis model showed the best performance (AUC = 1). The number of selected features was reduced up to the point where a decrease in model performance was observed. This deterioration of the model performance can also be taken as a reminder of the increased reliability that biomarker panels provide against single biomarkers. At the same time, while these results are promising, the search for biomarkers that can better reflect liver disease progression must continue in order to obtain more reliable panels. Currently available commercial options, such as the LiverEdge panel (ProteomEdge AB, Stockholm, Sweden), facilitate this process.

The seven-feature multiclassification model, which was built to stratify samples showing severe, moderate, mild, or no fibrosis, allowed for the correct stratification of the test set plasma samples with a 0.50 accuracy. Being a four-class model, this is better than random classification, which would have a 0.25 accuracy score. Furthermore, the classification errors were between healthy and mild/moderate or between mild/moderate and severe, but there were no misclassifications between healthy and severe. At the same time, supervised machine learning depends completely on the quality of the training data. This means that the results are also impacted by the inherent variability of the clinical assessment of fibrosis in liver biopsies.

Here, we recognize that it is possible that a subset of the OBES patients could have undiagnosed mild fibrosis or early liver damage, particularly given the known risk of metabolic comorbidities, such as MASLD in obesity. However, given the distinct separation of the control group from the liver disease group, this is unlikely to have significantly impacted the conclusions of the study. Most importantly, treating the OBES group as F0 reduced the probabilities of incorrectly identifying biomarkers that may be confounded by an obesity-related pathophysiology rather than liver fibrosis. On the other hand, while the aspect of disease progression that has been explored here is liver fibrosis, it is not possible to conclude from this study that the selected proteins indicate fibrosis, since they could be indicating other aspects of liver disease progression that accompany it instead. To connect the protein expression to its ultimate cause, it would be necessary to include clinical metadata that accounted for other progression metrics of liver disease, such as steatosis and inflammation [14], as well as a higher number of stratified samples. Overall, the sample size is a limitation of this study. This was especially apparent for the test sets used to evaluate model performance. Prior to any contemplation of the applicability of the biomarker panel in a clinical setting, it would be necessary to validate it using larger, independent patient cohorts.

Another aspect to consider in this study is the characteristics of the targeted proteins. Proteins with medium-to-high-abundant concentrations in blood were analyzed, this being the range in which targeted mass spectrometry delivers more reliable measurements [48]. Almost 60% of the targeted proteins are enriched or enhanced in the liver, according to the HDBA. However, this seemingly biased selection has, at least partly, a biological explana-

tion, since genome-scale metabolic models indicate that the liver is the most metabolically active tissue in the body [41]. Here, as liver disease progresses, most of the targeted proteins that show differential expression are downregulated, reflecting the impairment of liver function. This is in line with Niu et al.'s previous 2019 and 2022 [2,14] mass spectrometry-based liver disease studies, which, through data-independent acquisition (DIA), found that most of the liver-specific proteins were downregulated in liver disease and that most of these had metabolic and bloodstream-related functions. This included proteins, such as (i) prothrombin (F2), which converts fibrinogen to fibrin in the coagulation cascade, (ii) protein S (PROS), an anticoagulant protein, or (iii) apolipoprotein M (APOM), involved in lipid transportation [41]. All of these proteins were also found to be increasingly downregulated as liver fibrosis progressed in this study. On the other hand, proteins involved in cell–extracellular matrix interactions, immunity, and inflammation, which had been found to be upregulated in liver disease [2,14], followed the same tendency in this study. These include examples include (i) quiescin sulfhydryl oxidase 1 (QSOX), which plays a role in cell adhesion and migration, and (ii) complement 7 (C7), which is involved in the innate and adaptive immune response [41]. Both proteins' levels seemed to increase with fibrosis progression in this study.

In summary, targeted proteomics and the use of SIS-PrESTs increase accuracy and reliability in protein identification and quantification. This method addresses issues that could hinder clinical translation, such as the lack of absolute quantification in DIA mass spectrometry. Thus, it presents itself as a more competitive option for proteome profiling in the clinic.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/livers5010002/s1, Figure S1: Overrepresentation analysis, Figure S2: Multiclassification model for fibrosis stratification, Figure S3: Protein expression correlation, Table S1: List of targets, Table S2: De-identified list of samples with diagnose and fibrosis staging, Table S3: Differentially expressed proteins and regulation type, Table S4: Kruskal-Wallis and Wilcox post-hoc tests *q*-values by protein.

Author Contributions: Conceptualization, A.V.R., D.K., O.A., A.M., D.A., C.Y., M.A., M.D., H.K., M.Z. and F.E.; Methodology, A.V.R., D.K. and F.E.; Formal analysis, A.V.R.; Investigation, A.V.R. and F.E.; Resources, D.K., O.A., A.M., C.Y., M.A., M.D., H.K., M.Z. and F.E.; Data curation, A.V.R. and F.E.; Writing—original draft, A.V.R., D.K. and F.E.; Writing—review & editing, A.V.R., D.K., O.A., A.M., D.A., M.Z. and F.E.; Visualization, A.V.R. and F.E.; Supervision, F.E.; Project administration, A.V.R., A.M. and F.E.; Funding acquisition, F.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Koç University (protocol codes 2015.053.IRB1.014: 30 March 2015), (2016.024.IRB2.005: 4 February 2016), (2017.139.IRB2.048: 5 September 2017), (2018.351.IRB1.043: 15 May 2019), (2022.246.IRB2.040: 28 July 2022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting the findings of this study are not publicly available due to ethical restrictions related to patient confidentiality and data protection policies. However, de-identified data may be made available upon reasonable request to the corresponding author, subject to appropriate ethical approvals and data-sharing agreements.

Acknowledgments: We thank the entire staff of the Human Protein Atlas program (www.proteinatlas.org) and Proteome Edge AB (Stockholm, Sweden) for their valuable contributions. We gratefully acknowledge the scientists, services, and facilities of the Koç University Research Center for Translational Medicine (KUTTAM), funded by the Presidency of Turkey, Head of Strategy and Budget.

Conflicts of Interest: DK is employed by ProteomeEdge AB (Stockholm, Sweden). DK and FE are co-founders of ProteomeEdge AB (Stockholm, Sweden).

References

1. Devarbhavi, H.; Asrani, S.K.; Arab, J.P.; Nartey, Y.A.; Pose, E.; Kamath, P.S. Global burden of liver disease: 2023 update. *J. Hepatol.* **2023**, *79*, 516–537. [CrossRef] [PubMed]
2. Niu, L.; Geyer, P.E.; Wewer Albrechtsen, N.J.; Gluud, L.L.; Santos, A.; Doll, S.; Treit, P.V.; Holst, J.J.; Knop, F.K.; Vilsbøll, T.; et al. Plasma proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease. *Mol. Syst. Biol.* **2019**, *15*, e8793. [CrossRef] [PubMed]
3. Underhill, G.H.; Khetani, S.R. Emerging trends in modeling human liver disease in vitro. *APL Bioeng.* **2019**, *3*, 040902. [CrossRef] [PubMed]
4. Chowdhury, A.B.; Mehta, K.J. Liver biopsy for assessment of chronic liver diseases: A synopsis. *Clin. Exp. Med.* **2023**, *23*, 273–285. [CrossRef] [PubMed]
5. Cheemerla, S.; Balakrishnan, M. Global epidemiology of chronic liver disease. *Clin. Liver Dis.* **2021**, *17*, 365–370. [CrossRef]
6. Asrani, S.K.; Devarbhavi, H.; Eaton, J.; Kamath, P.S. Burden of liver diseases in the world. *J. Hepatol.* **2019**, *70*, 151–171. [CrossRef]
7. Mokdad, A.A.; Lopez, A.D.; Shahraz, S.; Lozano, R.; Mokdad, A.H.; Stanaway, J.; Murray, C.; Naghavi, M. Liver cirrhosis mortality in 187 countries between 1980 and 2010: A systematic analysis. *BMC Med.* **2014**, *12*, 145. [CrossRef]
8. Heyens, L.J.; Busschots, D.; Koek, G.H.; Robaeys, G.; Francque, S. Liver fibrosis in non-alcoholic fatty liver disease: From liver biopsy to non-invasive biomarkers in diagnosis and treatment. *Front. Med.* **2021**, *8*, 615978. [CrossRef]
9. Ginès, P.; Castera, L.; Lammert, F.; Graupera, I.; Serra-Burriel, M.; Allen, A.M.; Wai-Sun Wong, V.; Hartmann, P.; Thiele, M.; Caballeria, L.; et al. Population screening for liver fibrosis: Toward early diagnosis and intervention for chronic liver diseases. *Hepatology* **2022**, *75*, 219–228. [CrossRef]
10. Huang, D.Q.; Terrault, N.A.; Tacke, F.; Gluud, L.L.; Arrese, M.; Bugianesi, E.; Loomba, R. Global epidemiology of cirrhosis—Aetiology, trends and predictions. *Nat. Rev. Gastroenterol. Hepatol.* **2023**, *20*, 388–398. [CrossRef]
11. Nagy, P.; Thorgeirsson, S.S.; Grisham, J.W. Organizational principles of the liver. In *The Liver: Biology and Pathobiology*; Arias, I.M., Alter, H.J., Boyer, J.L., Cohen, D.E., Shafritz, D.A., Thorgeirsson, S.S., Wolkoff, A.W., Eds.; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2020; pp. 1–13.
12. Abdel-Misih, S.R.; Bloomston, M. Liver anatomy. *Surg. Clin. N. Am.* **2010**, *90*, 643–653. [CrossRef] [PubMed]
13. Larsen, F.S.; Saliba, F. Liver support systems and liver transplantation in acute liver failure. *Liver Int.* **2023**. [CrossRef] [PubMed]
14. Niu, L.; Thiele, M.; Geyer, P.E.; Rasmussen, D.N.; Webel, H.E.; Santos, A.; Gupta, R.; Meier, F.; Strauss, M.; Kjaergaard, M.; et al. Noninvasive proteomic biomarkers for alcohol-related liver disease. *Nat. Med.* **2022**, *28*, 1277–1287. [CrossRef] [PubMed]
15. European Association for the Study of the Liver. EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. *J. Hepatol.* **2015**, *63*, 237–264. [CrossRef]
16. Everhart, J.E.; Wright, E.C.; Goodman, Z.D.; Dienstag, J.L.; Hoefs, J.C.; Kleiner, D.E.; Ghany, M.G.; Mills, A.S.; Nash, S.R.; Govindarajan, S.; et al. Prognostic value of Ishak fibrosis stage: Findings from the hepatitis C antiviral long-term treatment against cirrhosis trial. *Hepatology* **2010**, *51*, 585–594. [CrossRef]
17. Uhlén, M.; Karlsson, M.J.; Zhong, W.; Tebani, A.; Pou, C.; Mikes, J.; Lakshmikanth, T.; Forsström, B.; Edfors, F.; Odeberg, J.; et al. A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* **2019**, *366*, eaax9198. [CrossRef]
18. Thapa, B.R.; Walia, A. Liver function tests and their interpretation. *Indian J. Pediatr.* **2007**, *74*, 663–671. [CrossRef]
19. Hober, A.; Edfors, F.; Ryaboshapkina, M.; Malmqvist, J.; Rosengren, L.; Percy, A.J.; Lind, L.; Forsström, B.; Uhlén, M.; Oscarsson, J.; et al. Absolute quantification of apolipoproteins following treatment with omega-3 carboxylic acids and fenofibrate using a high precision stable isotope-labeled recombinant protein fragments based SRM assay. *Mol. Cell. Proteom.* **2019**, *18*, 2433–2446. [CrossRef]
20. Geyer, P.E.; Kulak, N.A.; Pichler, G.; Holdt, L.M.; Teupser, D.; Mann, M. Plasma proteome profiling to assess human health and disease. *Cell Syst.* **2016**, *2*, 185–195. [CrossRef]
21. Kotol, D.; Hober, A.; Strandberg, L.; Svensson, A.S.; Uhlén, M.; Edfors, F. Targeted proteomics analysis of plasma proteins using recombinant protein standards for addition only workflows. *Biotechniques* **2021**, *71*, 473–483. [CrossRef]
22. Callesen, A.K.; Mohammed, S.; Bunkenborg, J.; Kruse, T.A.; Cold, S.; Mogensen, O.; dePont Christensen, R.; Vach, W.; Jørgensen, P.E.; Jensen, O.N. Serum protein profiling by miniaturized solid-phase extraction and matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1578–1586. [CrossRef] [PubMed]
23. MacLean, B.; Tomazela, D.M.; Shulman, N.; Chambers, M.; Finney, G.L.; Frewen, B.; Kern, R.; Tabb, D.; Liebler, D.; MacCoss, M.J. Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966–968. [CrossRef] [PubMed]
24. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.

25. Voß, H.; Schlumbohm, S.; Barwikowski, P.; Wurlitzer, M.; Dottermusch, M.; Neumann, P.; Krisp, C. HarmonizR enables data harmonization across independent proteomic datasets with appropriate handling of missing values. *Nat. Commun.* **2022**, *13*, 3523. [CrossRef] [PubMed]
26. Kowarik, A.; Templ, M. Imputation with the R Package VIM. *J. Stat. Softw.* **2016**, *74*, 1–16. [CrossRef]
27. Kuhn, M.; Wickham, H. Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles. Available online: https://www.tidymodels.org (accessed on 10 January 2025).
28. Greenwell, B.M.; Boehmke, B.C.; Gray, B. Variable Importance Plots-An Introduction to the vip Package. *R J.* **2020**, *12*, 343. [CrossRef]
29. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; IEEE: Piscataway, NJ, USA, 1995; Volume 1, pp. 278–282.
30. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
31. Zhu, R.; Zeng, D.; Kosorok, M.R. Reinforcement learning trees. *J. Am. Stat. Assoc.* **2015**, *110*, 1770–1784. [CrossRef]
32. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [CrossRef]
33. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.C.; Müller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [CrossRef]
34. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.
35. Kassambara, A. ggpubr: 'ggplot2' Based Publication Ready Plots. Available online: https://cran.r-project.org/web/packages/ggpubr/index.html (accessed on 10 January 2025).
36. Kassambara, A.; Mundt, F. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package Version 1.0.7. Available online: https://CRAN.R-project.org/package=factoextra (accessed on 10 January 2025).
37. Patil, I. Visualizations with statistical details: The 'ggstatsplot' approach. *J. Open Source Softw.* **2021**, *6*, 3167. [CrossRef]
38. Conway, J.R.; Lex, A.; Gehlenborg, N. UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* **2017**, *33*, 2938–2940. [CrossRef] [PubMed]
39. Slowikowski, K.; Schep, A.; Hughes, S. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R Package Version 0.9. Available online: https://cran.r-project.org/web/packages/ggrepel/ggrepel.pdf (accessed on 10 January 2025).
40. Pedersen, T. Patchwork: The Composer of Plots. R Package Version 1.3.0.9000. Available online: https://github.com/thomasp85/patchwork (accessed on 10 January 2025).
41. Uhlén, M.; Fagerberg, L.; Hallström, B.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; et al. Proteomics. Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419. [CrossRef] [PubMed]
42. Shahidi, M. Thrombosis and von Willebrand factor. *Thromb. Embolism Res. Clin. Pract.* **2017**, *1*, 285–306.
43. Rodríguez-Castro, K.I.; Antonello, A.; Ferrarese, A. Spontaneous bleeding or thrombosis in cirrhosis: What should be feared the most? *World J. Hepatol.* **2015**, *7*, 1818. [CrossRef]
44. Claviez, A.; Horst, H.A.; Santer, R.; Suttorp, M. Neutrophil aggregates in a 13-year-old girl: A rare hematological phenomenon. *Ann. Hematol.* **2003**, *82*, 251–253. [CrossRef]
45. Newton, J.L.; Jones, D.E. Managing systemic symptoms in chronic liver disease. *J. Hepatol.* **2012**, *56*, S46–S55. [CrossRef]
46. Álvez, M.B.; Edfors, F.; von Feilitzen, K.; Zwahlen, M.; Mardinoglu, A.; Edqvist, P.H.; Sjöblom, T.; Lundin, E.; Rameika, N.; Enblad, G.; et al. Next generation pan-cancer blood proteome profiling using proximity extension assay. *Nat. Commun.* **2023**, *14*, 4308. [CrossRef]
47. Kotol, D.; Woessmann, J.; Hober, A.; Álvez, M.B.; Tran Minh, K.H.; Pontén, F.; Fagerberg, L.; Edfors, F. Absolute Quantification of Pan-Cancer Plasma Proteomes Reveals Unique Signature in Multiple Myeloma. *Cancers* **2023**, *15*, 4764. [CrossRef]
48. Tu, C.; Rudnick, P.A.; Martinez, M.Y.; Cheek, K.L.; Stein, S.E.; Slebos, R.J.; Liebler, D.C. Depletion of abundant plasma proteins and limitations of plasma proteomics. *J. Proteome Res.* **2010**, *9*, 4982–4991. [CrossRef]