PROTEOMICS – Clinical Applications

WILEY-VCH

Proteomics
Clinical Applications

RESEARCH ARTICLE

Identification of Pre-Diagnostic Protein Biomarkers for Liver Cirrhosis Based on Prospective Analysis of a Large-Scale Plasma Proteomics in the UK Biobank

Jitian He[1] | Bo Li[2] | Yuping Yan[1] | Yajie Wang[3] | Mingxi Zhang[1] | Renyuan Sun[3] | Baohua Hou[3] | Shanzhou Huang[3] | Limin Zhen[2] | Dongping Wang[1] | Chuanzhao Zhang[3]

[1]Organ Transplant Center, The First Affiliated Hospital of Sun Yat-Sen University, Guangzhou, P. R. China | [2]Digestive System Department, The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou, Guangdong, P. R. China | [3]Department of General Surgery, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, P. R. China

Correspondence: Limin Zhen (zhenlimin2013@126.com) | Dongping Wang (dpwangcn@163.com) | Chuanzhao Zhang (zhangchuanzhao@gdph.org.cn)

Received: 13 January 2026 | **Revised:** 13 July 2026 | **Accepted:** 16 July 2026

Keywords: cirrhosis | proteomics | UK biobank

ABSTRACT

Purpose: Liver fibrosis and cirrhosis represent critical stages in the progression of chronic liver disease, yet their key molecular features remain incompletely understood.

Experimental design: We performed large-scale Olink-based proteomic profiling in over 40,000 participants from the UK Biobank with a median follow-up of 15.6 years to elucidate disease pathophysiology and identify pre-diagnostic biomarkers. Cross-sectional analysis included 66 prevalent cirrhosis cases, and prospective analysis identified 224 incident cirrhosis cases. Machine learning and Mendelian randomization (MR) were applied. An independent cohort was used for validation.

Results: Distinct dysregulated proteins were observed in compensated cirrhosis (CC) and decompensated cirrhosis (DC). In the prospective analysis, 696 proteins were associated with disease onset. A proteomic panel based on these markers achieved an AUC of 0.832 for predicting incident cirrhosis, outperforming established fibrosis scores including FIB-4, APRI, and NFS, and demonstrated robust performance across CC and DC populations. The protein panel showed predictive value (AUC = 0.743) for disease progression in an independent cohort. MR identified 66 proteins with putative causal roles, including 11 potential therapeutic targets.

Conclusions and clinical relevance: These findings provide novel molecular insights into cirrhosis development and support integrated proteomic biomarkers as a discovery and prioritization framework for early risk stratification.

Abbreviations: AI, Artificial intelligence; ALB, Albumin; ALT, Alanine aminotransferase; APRI, Aspartate aminotransferase to platelet ratio index; AST, Aspartate aminotransferase; AUC, Area under the curve; BMI, Body mass index; CC, Compensated cirrhosis; CI, Confidence interval; Cys C, Cystatin C; DC, Decompensated cirrhosis; ECM, Extracellular matrix; ELISA, Enzyme-linked immunosorbent assay; FDR, False discovery rate; FIB-4, Fibrosis-4 index; GWAS, Genome-wide association study; HPA, Human Protein Atlas; HR, Hazard ratio; ICD, International Classification of Diseases; INR, International normalized ratio; IVW, Inverse-variance weighted; KEGG, Kyoto Encyclopedia of Genes and Genomes; KNN, K-nearest neighbours; LC–MS, Liquid chromatography–mass spectrometry; LightGBM, Light gradient boosting machine; MR, Mendelian randomization; NFS, Nonalcoholic fatty liver disease fibrosis score; NPX, Normalized protein expression; OR, Odds ratio; PPI, Protein–protein interaction; pQTL, Protein quantitative trait locus; PSM, Propensity score matching; ROC, Receiver operating characteristic; SNP, Single nucleotide polymorphism; TBIL, Total bilirubin; TDI, Townsend deprivation index; UKB, UK Biobank; UKB-PPP, UK Biobank Pharma Proteomics Project.; UMAP, Uniform manifold approximation and projection.

Jitian He, Bo Li, and Yuping Yan contributed equally to this work.

Statement of Clinical Relevance

Early identification of individuals at risk of liver cirrhosis remains a major clinical challenge, as current tools rely largely on indirect markers and lack sensitivity for early-stage disease. In this study, we leveraged large-scale Olink-based plasma proteomics from the UK Biobank to systematically identify proteins associated with cirrhosis development and constructed a predictive protein panel with improved performance over conventional fibrosis scores. We further conducted preliminary validation in an independent prospective cohort using a clinically accessible assay, supporting the translational potential of the identified markers. These findings provide a framework for prioritizing candidate biomarkers and support the development of strategies for early risk stratification and precision management of chronic liver disease.

1 | Introduction

Liver fibrosis and cirrhosis are critical stages in the progression of chronic liver disease [1]. Chronic inflammation [2], extracellular matrix (ECM) remodeling [3], and immune homeostasis dysregulation [4] contribute to the development of liver fibrosis, cirrhosis, and eventually hepatic failure. While early-stage liver fibrosis could be partially reversible upon removal of etiological factors causing chronic injury, no effective therapy is currently available to facilitate fibrosis resolution or liver regeneration [5, 6]. Moreover, there is a lack of sensitive tools for early diagnosis of liver fibrosis or cirrhosis, as significant changes typically do not appear on ultrasound or blood liver function panels in patients with the compensated stage of the disease [7, 8]. Other clinical indicators have been shown to be closely associated with liver fibrosis, cirrhosis, and liver failure, such as the aspartate aminotransferase to platelet ratio index (APRI), NAFLD fibrosis score (NFS), and the Fibrosis-4 (FIB-4) index [9, 10]. However, most studies focus on a single or a small number of indicators, primarily assessing disease progression in diagnosed patients. It remains unclear whether these potential biomarkers are useful for risk prediction and disease monitoring. Therefore, it is crucial to further explore effective biomarkers for identifying high-risk individuals and for early diagnosis, which would enable prompt intervention for populations at risk for liver fibrosis and cirrhosis.

Recent advances in high-throughput proteomics have enabled the analysis of disease mechanisms and the discovery of biomarkers, paving the way for personalized risk prediction and intervention in human diseases [11–13]. Proteomic technologies differ substantially in the level of molecular information they provide. Mass spectrometry (MS)-based proteomics includes both bottom-up approaches, which infer protein abundance from digested peptides, and top-down approaches, which directly characterize intact proteoforms [14, 15]. Top-down proteomics enables the identification of proteoforms arising from genetic variation, alternative splicing, post-translational modifications, and proteolytic processing, thereby capturing molecular heterogeneity beyond conventional protein-level measurements [14, 16]. Although top-down proteomics provides detailed molecular information, its application in large population cohorts remains challenging due to technical complexity and limited throughput. In contrast, affinity-based platforms such as Olink employ paired antibodies coupled with proximity extension assays to quantify predefined circulating protein targets with high sensitivity and throughput, making them particularly suitable for large population studies, although they generally do not distinguish individual proteoforms [17]. Olink-based proteomics enables scalable measurement of predefined circulating proteins, whereas LC–MS approaches provide broader proteome coverage and proteoform-level information. These complementary technologies provide different perspectives on the circulating proteome and biomarker discovery, and differences between studies should therefore be interpreted in the context of the underlying analytical platform. Several studies have used proteomics to identify blood protein differences between patients with and without cirrhosis [18, 19]. For instance, Forte et al. employed top-down proteomics to identify plasma proteoform signatures in patients at different stages of cirrhosis [19]. Such studies demonstrate the value of proteoform-resolved analyses for revealing molecular heterogeneity associated with liver disease, whereas Olink-based plasma proteomics provides a complementary strategy for large-scale biomarker discovery and risk prediction in population cohorts. However, most of these studies were cross-sectional, with small sample sizes, and did not analyze protein levels before the onset of cirrhosis or include long-term follow-up data. Therefore, large-scale, prospective studies with long-term follow-up based on proteomics are essential for investigating plasma proteins associated with cirrhosis, fibrosis, and liver failure.

In this study, by leveraging population-scale, high-throughput Olink-based proteomics and detailed phenotypic data from the UK Biobank, we aimed to identify proteomic features associated with compensated (CC) and decompensated (DC) cirrhosis and evaluate the impact of proteomic alterations on the prediction of incident cirrhosis. To address these questions, we first identified differentially expressed proteins in patients diagnosed with cirrhosis prior to baseline data collection, offering insights into shared biological pathways and mechanisms underlying the disease. Using machine learning, we then identified key indicators within the proteomic profiles and assessed the predictive performance of top-ranked proteins, both individually and in combination with other biomarkers. Finally, we integrated protein quantitative trait locus (pQTL) and summary genome-wide association study (GWAS) data to determine causal proteins of the diseases and explore potential directions for drug targets.

2 | Results

2.1 | Participant Characteristics

A flowchart of the inclusion strategy is shown in Figure 1. A total of 44,030 participants from the UK Biobank with available baseline proteomic data were included. Of these, 2920 quantified proteins and 61 blood biomarkers that met the quality criteria were used for subsequent analysis (Tables S1 and S2). A detailed baseline characterization of the study population, including demographics, clinical parameters, and diagnoses is provided in Table 1. In the first cohort, we analyzed the proteomic characteristics of 66 patients diagnosed with liver fibrosis and cirrhosis, consisting of 49 with CC and 17 with DC. Controls were matched

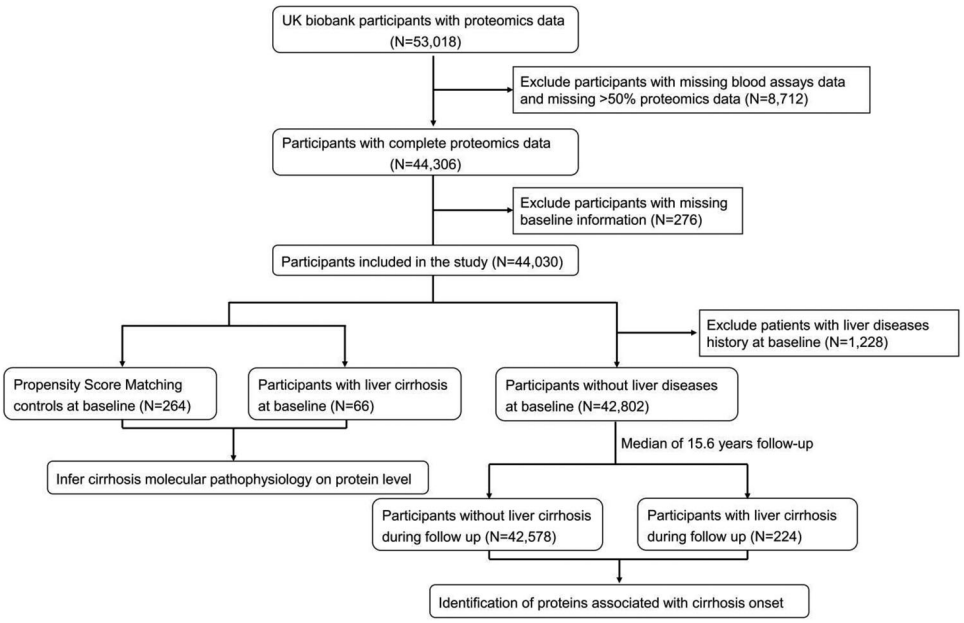

FIGURE 1 | Flow chart of participant inclusion and exclusion. Individuals from UK biobank with missing >50% proteomics data and baseline characteristic were excluded. 66 cases reported liver cirrhosis prior to baseline data collection and 264 matched controls were separately analyzed to infer cirrhosis molecular pathophysiology on protein level. In follow-up cohort, participants with liver diseases at baseline and those with missing blood assay data were excluded, yielding an analytic cohort of 42,802 individuals. After median of 15.6 years follow-up, 224 participants with liver cirrhosis and 42,578 participants without liver cirrhosis were included for subsequent identification of proteins associated with cirrhosis onset.

by demographics using the propensity score matching (PSM) method. In the second cohort, we included 42,802 participants without a history of liver disease at baseline. Over a median follow-up of 15.6 years, 224 participants were diagnosed with incident liver fibrosis and cirrhosis, including 116 with CC and 108 with DC. Compared to cirrhosis-free individuals from both cohorts, patients who developed cirrhosis were older, more likely to be male, and had a higher propensity for smoking and drinking. They also had higher body mass index (BMI), a greater proportion of hypertension and diabetes, and higher Townsend deprivation index (TDI) scores. Laboratory parameters such as ALT, AST, GGT, and fibrosis indices (FIB-4, APRI, and NFS) were elevated in incident cirrhosis cases, with significantly higher levels observed in DC compared to CC.

2.2 | Dysregulated Proteins in Liver Fibrosis and Cirrhosis

To investigate molecular changes associated with liver fibrosis and cirrhosis, we first compared the proteomic profiles between populations with and without a prior diagnosis of liver fibrosis and cirrhosis. A total of 136 proteins were significantly dysregulated (Log_2 |Fold change| $\geq$ 0.5, false discovery rate (FDR)-adjusted $p < 0.05$) in patients with diagnosed cirrhosis, including 124 upregulated and 16 downregulated proteins (Figure 2a; Table S3). More than 50% of the dysregulated proteins were annotated as "secreted" proteins according to the human protein atlas (HPA) [20] (Figure 2b). Patients were further categorized based on complications into CC and DC. Uniform manifold approximation and projection (UMAP) analysis revealed distinct proteomic profiles among etiology-based groups (Figure 2c). A total of 67 dysregulated proteins were common between CC versus controls and DC versus controls, with 10 proteins related to immune response and viral infection shared between both groups (Figure 2d).

A heatmap displaying the significantly dysregulated proteins across different groups is shown in Figure 2e. Proteins associated with ECM organization, including collagen type V (COL5A1), Fibulin-2 (FBLN2), Biglycan (BGN), and Secreted Protein Acidic and Cysteine Rich (SPARC) were notably dysregulated (Figure 2e). Additionally, proteins involved in Rho GTPase and GPCR signaling, such as ARHGEF1, VAV3, and CRKL, were significantly downregulated. A total of 19 liver-enriched proteins were identified based on annotations from the HPA), with four proteins (FABP1, SULT2A1, AGXT, and HAO1) classified as "liver-specific" (Figure 2f). The majority of liver-enriched proteins were upregulated in both CC and DC. These proteins are involved in various liver metabolic processes, including lipid transport (FABP1, SCP2, ACAA1), amino acid catabolism (AGXT, FTCD), and xenobiotic detoxification (SULT2A1, HAO1). Additionally, we identified three proteins potentially associated with fibrosis progression—DPP10, CLSTN2, and EDAR that exhibited consistent trends across controls, CC, and DC (Figure 2g).

Correlation analysis identified proteins significantly associated with fibrosis parameters. Among the top 10 proteins most strongly correlated with the FIB-4 index and APRI, prominent liver-related proteins involved in liver metabolism (AGXT, FABP1, GOT1, etc.) and liver inflammation (KRT18) were observed (Figure 2h). The proteins most correlated with the NFS were linked to cell–ECM interactions (SCD4 and NCAN), platelet and coagulation function (GP1BA and PF4), and immune activation (CD302, EDA2R, and MBL2). To further explore the biological functions of the identified proteins, enrichment analyses for

TABLE 1 | Baseline participant characteristics.

Characteristics	Participants without incident cirrhosis during follow-up (*N* = 42,578)	Participants without incident cirrhosis during follow-up (*N* = 224)			Propensity Score Matching controls at baseline (*N* = 264)	Participants with prevalent cirrhosis at baseline (*N* = 66)		
		Total (*N* = 224)	Compensated Cirrhosis (CC; *N* = 116)	Decompensated Cirrhosis (DC; *N* = 108)		Total (*N* = 66)	Compensated Cirrhosis (CC; *N* = 49)	Decompensated Cirrhosis (DC; *N* = 17)
Age (years)	58 (50-64)	61 (54.75-65)	59 (54-65)	62 (56-66)	59 (52-65)	59 (54-63)	60 (55-64)	57 (52-60)
Male, N (%)	19540 (45.9%)	142 (63.4%)	60 (51.7%)	82 (75.9%)	120 (45.5%)	26 (39.4%)	19 (38.8%)	7 (41.2%)
Ethnicity (British), N (%)	37203 (87.4%)	197 (87.9%)	102 (87.9%)	95 (88%)	213 (80.7%)	53 (80.3%)	40 (81.6%)	13 (76.5%)
BMI (kg/m2)	26.8 (24.2-29.9)	29.8 (26.2-34.3)	29.4 (25.9-32.6)	30.5 (27.2-34.9)	28.1 (25.3-31.9)	28.7 (24.0-33.3)	28.9 (25.8-31.2)	27.1 (22.8-33.3)
Townsend deprivation index	−2.1 (−3.6-0.7)	−0.7 (−2.7-2.5)	−0.7 (−2.7-2.0)	−0.6 (−2.7-2.8)	1.01 (−1.9-4.1)	1.3 (−2.6-4.2)	1.1 (−2.8-4.3)	2.97 (−2.5-3.94)
Current smoker, N (%)	4467 (10.5%)	43 (19.2%)	18 (15.5%)	25 (23.1%)	53 (20.1%)	13 (19.7%)	8 (16.3%)	5 (29.4%)
Daily drinker, N (%)	8592 (20.2%)	62 (27.7%)	25 (21.6%)	37 (34.3%)	49 (18.6%)	11 (16.7%)	10 (20.4%)	1 (5.9%)
Hypertension, N (%)	11753 (27.6%)	106 (47.3%)	55 (47.4%)	51 (47.2%)	133 (50.4%)	34 (51.5%)	24 (49%)	10 (58.8%)
Diabetes, N (%)	1180 (2.8%)	32 (14.3%)	14 (12.1%)	18 (16.7%)	37 (14.0%)	8 (12.1%)	6 (12.2%)	2 (11.8%)
Follow time (years)	15.6 (14.8-16.3)	8.6 (5.1-11.7)	8.5 (5.0-11.8)	8.8 (5.4-11.4)	NA	NA	NA	NA
Laboratory parameters								
Albumin (g/L)	45 (44-47)	44 (43-46)	44 (42-46)	45 (43-46)	45 (43-46)	43 (41-46)	44 (41-46)	42 (38-45)
ALT (U/L)	20 (15-27)	32 (21-51)	30 (18-48)	34 (24-51)	21 (16-29)	27 (19-45)	29 (24-36)	34 (19-48)
AST (U/L)	25 (21-29)	34 (26-49)	33 (25-48)	35 (27-50)	25 (21-30)	31 (25-51)	32 (26-51)	34 (21-45)
GGT (U/L)	27 (19-40)	72 (38-157)	59 (37-136)	92 (43-170)	29 (20-47)	74 (45-178)	46 (21-130)	96 (51-181)
Platelet count (10^9/L)	248 (214-287)	231 (186-276)	256 (202-295)	210 (165-252)	244 (214-277)	210 (172-251)	213 (174-269)	193 (159-224)
Total bilirubin (μmol/L)	8.0 (6.3-10.4)	9 (6.5-11.7)	8.8 (6.3-11.0)	9.1 (6.7-12.4)	8.0 (6.3-10.4)	9.0 (6.4-13.7)	9.0 (6.6-10.7)	12.2 (6.3-20.6)
Direct bilirubin (μmol/L)	1.6 (1.3-2.0)	1.9 (1.4-2.7)	1.8 (1.4-2.5)	1.9 (1.4-2.8)	1.6 (1.4-2.0)	1.9 (1.5-3.4)	1.8 (1.4-2.9)	3.1 (1.7-4.5)
Alkaline phosphatase (U/L)	81 (68-95)	91 (76-112)	93 (78-113)	89 (74-112)	84 (72-98)	106 (84-156)	111 (84-175)	104 (89-132)
Total cholesterol (mmol/L)	1.5 (1.1-2.1)	1.7 (1.2-2.3)	1.6 (1.1-2.2)	1.74 (1.2-2.4)	1.6(1.1-2.4)	1.4(1.0-2.0)	1.53 (1.06-2.12)	1.21 (1.02-1.64)
HbA1C (mmol/mol)	35 (33-38)	37 (34-42)	37 (34-41)	37 (34-44)	37 (34-40)	36 (33-39)	36 (33-39)	33 (31-37)

(Continues)

TABLE 1 | (Continued)

Characteristics	Participants without incident cirrhosis during follow-up (*N* = 42,578)	Participants without incident cirrhosis during follow-up (*N* = 224)			Propensity Score Matching controls at baseline (*N* = 264)	Participants with prevalent cirrhosis at baseline (*N* = 66)		
		Total (*N* = 224)	Compensated Cirrhosis (CC; *N* = 116)	Decompensated Cirrhosis (DC; *N* = 108)		Total (*N* = 66)	Compensated Cirrhosis (CC; *N* = 49)	Decompensated Cirrhosis (DC; *N* = 17)
Fibrosis parameters								
FIB-4 Index	1.1 (0.9-1.4)	1.4 (1.0-2.3)	1.24 (1.0-1.8)	1.7 (1.2-2.6)	1.1 (0.9-1.4)	1.4 (1.1-2.2)	1.4 (1.1-2.0)	1.7 (1.1-2.4)
APRI	0.25 (0.21-0.33)	0.38 (0.26-0.64)	0.34 (0.24-0.5)	0.41 (0.29-0.77)	0.25 (0.21-0.33)	0.43 (0.27-0.82)	0.38 (0.24-0.62)	0.43 (0.27-0.82)
NFS	−0.53 (−1.12-0.19)	−1.24 (−2.22–0.17)	−1.76 (−2.52–0.62)	−0.81 (−1.58-0.11)	−0.53 (−1.18-0.19)	−1.21 (−2.22–0.15)	−1.22 (−2.20–0.29)	−1.20 (−2.29-1.61)
Complications of decompensated cirrhosis								
Hepatic failure	78 (0.2%)	52 (23.2%)	NA	52 (48.1%)	3 (1.1%)	9 (13.6%)	NA	9 (52.9%)
Portal hypertension	44 (0.1%)	72 (32.1%)	NA	72 (66.7%)	1 (0.4%)	13 (19.7%)	NA	13 (76.5%)
Gastro-intestinal haemorrhage	1459 (3.4%)	31 (13.8%)	NA	31 (28.7%)	14 (5.3%)	2 (3.0%)	NA	2 (11.8%)

All summary data are presented as medians (interquartile range) or counts (percentage). Decompensation was defined as the presence of complications including hepatic failure, portal hypertension, or gastro-intestinal haemorrhage on the basis of cirrhosis. The calculation fnor FIB-4 index, APRI, and NFS was described in methods section. BMI, Body Mass Index. FIB-4, fibrosis-4. APRI, Aspartate aminotransferase to platelet ratio index. NFS, Non-invasive Fibrosis Score. ALT, alanine aminotransferase. AST, aspartate aminotransferase. GGT, gamma-glutamyltransferase. HbA1C, glycated haemoglobin. NA, not applicable.

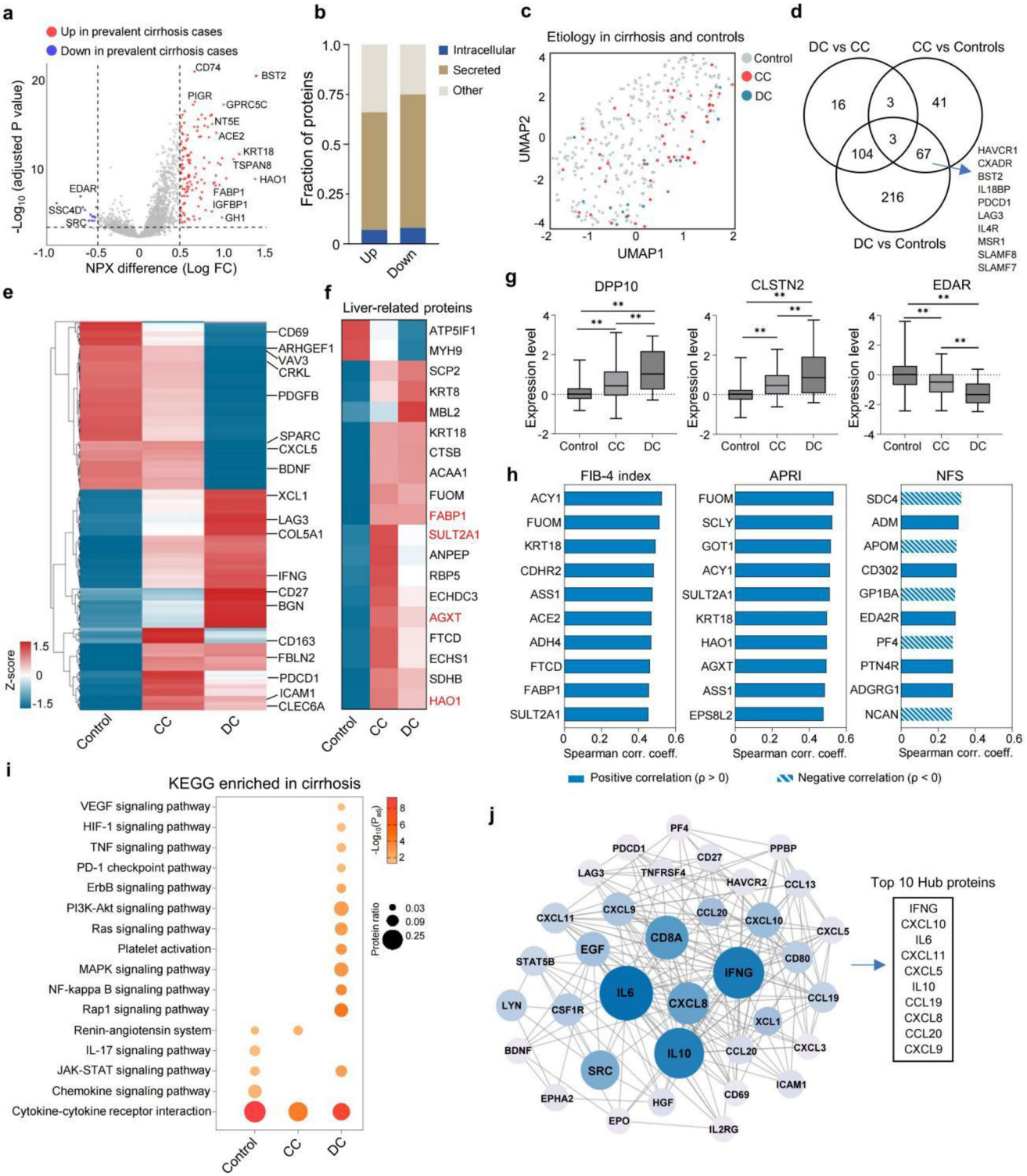

FIGURE 2 | Identification of dysregulated proteins in prevalent cirrhosis cases at baseline. (a) The volcano plot showing the up- and downregulated dysregulated proteins (identified by $\log_2|\text{fold change}| \geq 0.5$; FDR-adjusted p value < 0.05) in prevalent cirrhosis patients. (b) Fraction (%) of up- and downregulated intracellular and secreted proteins according to annotations of the Human Protein Atlas (HPA). (c) A UMAP analysis was performed on proteomics data from matched controls ($n = 264$), compensated (CC; $n = 49$) and decompensated (DC; $n = 17$) cirrhosis, each data point represents a sample in the respective colored group. (d) The number of dysregulated proteins shared among pairwise comparisons across three etiologies. Proteins related to immune response and viral infection overlapped in CC versus controls and DC versus controls were marked in the plot. (e) Hierarchical clustering of all significantly dysregulated proteins in the plasma proteome. Row clustering was based on median log2 intensity after z-score normalization across three groups. (f) Heatmap showing the expression of liver-related proteins among dysregulated proteins according to annotations of the HPA. The liver-specific proteins are highlighted in red. (g) Boxplots showing the significantly dysregulated proteins (DPP10, CLSTN2, and EDAR) that overlapped in DC versus CC, CC versus controls, and DC versus controls. The boxplots represent the interquartile range (IQR), with the horizontal

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were conducted on the dysregulated proteins across the groups (Figure 2i). The cytokine–cytokine receptor interaction pathway was the most highly enriched in both controls and cirrhosis patients. Receptor tyrosine kinase signaling pathways, including MAPK, Ras, PI3K-Akt, ErbB, and VEGF signaling, were notably enriched in the DC group, aligning with the existing literature on signaling pathways involved in fibrogenesis [21–23].

To further investigate potential interactions among proteins, we used the STRING database to analyze the protein–protein interaction (PPI) networks within the pool of dysregulated proteins in cirrhosis (Figure 2j). The top 10 hub proteins were identified using the maximal clique centrality (MCC) method. We found that most of the hub proteins belonged to the chemokine ligand family, indicating a central role of inflammation and chemotaxis in the development of advanced fibrosis.

2.3 | Identifying Proteins Associated With the Onset of Liver Cirrhosis

We then sought to identify cirrhosis-associated proteins within the context of a prospective cohort (Figure 3a). Cumulative time-to-onset distributions for cases revealed that 50% of diagnoses occurred by the eighth year of follow-up (Figure 3b). The majority of cases (>80%) developed DC within the first year after the onset of cirrhosis (Figure 3c). As expected, the survival probability for DC was considerably lower than for CC (Figure 3d). In a multivariable-adjusted model, we identified proteins associated with all incident cirrhosis, CC, and DC, respectively (Figure 3e–h; Table S4). Of the 2920 proteins quantified by the Olink platform, and after applying Bonferroni corrections ($p < 0.05 / 2{,}920$), 696 proteins were significantly associated with all incident cirrhosis (Figure 3f). When CC and DC were considered separately as primary endpoints, we identified 357 and 655 proteins significantly associated with each, respectively (Figure 3g,h; Table S4). Notably, proteins such as C7, THBS2, CEACAM1, and ITGBL1 exhibited the strongest correlations with a higher risk of cirrhosis onset (hazard ratio (HR) > 1). In contrast, proteins like SERPINF2, SERPINC1, and MENT were identified as protective proteins (HR < 1) for cirrhosis.

Clinical parameters such as cystatin C, APRI, NFS, FIB-4 index, apolipoprotein B, LDL-C, triglycerides, and IGF-1 were significantly associated with the risk of incident cirrhosis (Figure 3i). Strong correlations were observed between cirrhosis-associated proteins and clinical parameters (Figure 3j–o). For example, KRT18, a biomarker of hepatocyte apoptosis and a predictive protein for liver injury identified in a longitudinal study [24], showed correlations with higher levels of FIB-4 (Figure 3j), APRI (Figure 3k), and NFS (Figure 3l) as well as an increased risk of cirrhosis in our analysis. A high proportion of proteins was negatively correlated with total bilirubin (TBIL) (Figure 3m), apolipoprotein B (ApoB) (Figure 3n), and cystatin C (Cys C) (Figure 3o) and associated with a higher disease risk.

2.4 | Predictive Value of Plasma Proteins for Cirrhosis Onset

By modeling the cirrhosis risk as an endpoint, we assessed the diagnostic and predictive performance of plasma proteins compared to other clinical biomarkers. Cirrhosis-associated proteins were initially ranked based on their contribution to the prediction task. Among these, plasma GGT1, ADAMTSL2, and ITGBL1 ranked highest in predicting cirrhosis (Figure 4a; Table S5). The predictive accuracy, measured by area under the curve (AUC), increased sharply and then gradually plateaued as more proteins were included. Using a sequential forward selection approach, the top eight proteins (GGT1, ADAMTSL2, ITGBL1, CEACAM1, PLA2G15, C7, GDF15, and ADGRG1) were selected for subsequent analyses.

We examined the temporal trajectories of the eight selected plasma proteins from baseline to cirrhosis onset and compared the protein changes over the same period in individuals free from cirrhosis (Figures 4b–i). Compared to matched controls, the levels of the eight selected plasma proteins decreased more sharply over time in individuals who developed cirrhosis. Significant slope differences were observed for GGT1 ($p = 0.04$) (Figure 4b), ADAMTSL2 ($p = 0.018$) (Figure 4c), ITGBL1 ($p = 0.005$) (Figure 4d), and C7 ($p = 0.010$) (Figure 4g). However, no significant difference in the steepness of the slopes was found for CEACAM1 ($p = 0.113$) (Figure 4e), PLA2G15 ($p = 0.276$) (Figure 4f), GDF15 ($p = 0.092$) (Figure 4h), or ADGRG1 ($p = 0.314$) (Figure 4i).

The predictive accuracy of the selected proteins was evaluated using a 10-fold cross-validation approach (Figure 5a–f; Table S5). For all incident cirrhosis cases, plasma GGT1 alone yielded a modest AUC of 0.782, while ADAMTSL2 and ITGBL1 alone produced AUC values of 0.758 and 0.775, respectively (Figure 5a). When all eight candidate proteins were combined, the AUC increased to 0.832 (Figure 5a), which was considerably higher than the AUC values for fibrosis parameters such as the FIB-4 index (AUC = 0.683), APRI (AUC = 0.709), and NFS (AUC = 0.642). The fibrosis parameters demonstrated poor accuracy in predicting CC (FIB-4 index AUC = 0.594; APRI AUC = 0.635; NFS AUC = 0.517) (Figure 5b), compared to DC (FIB-4 index AUC = 0.732; APRI AUC = 0.741; NFS AUC = 0.682) (Figure 5c). In contrast, the protein panel exhibited strong predictive performance for both CC (AUC = 0.783) (Figure 5b) and DC (AUC = 0.801) (Figure 5c). Additionally, we observed that the earlier the onset of cirrhosis, the higher the predictive accuracy for all parameters (Figure 5d–f). The protein panel achieved an AUC of 0.858 for 3-year incident

line indicating the median. *P* values were derived from Kruskal-Wallis's test. *, $p<0.05$; **, $p<0.01$. (h) Top 10 proteins that correlate with Firbosis-4 (FIB-4) index, aspartate aminotransferase to platelet ratio index (APRI), and NAFLD fibrosis score (NFS), respectively. Associations between plasma proteins and clinical measurements were determined by Spearman correlation. (i) KEGG enrichment analyses showing the enriched pathways in dysregulated proteins among three groups. Dot size represents enriched protein ratio, and color intensity represents significance. (j) A PPI network for dysregulated proteins. The node size represents the modularity level estimated using the MCC method. The top 10 hub proteins are marked in the graph.

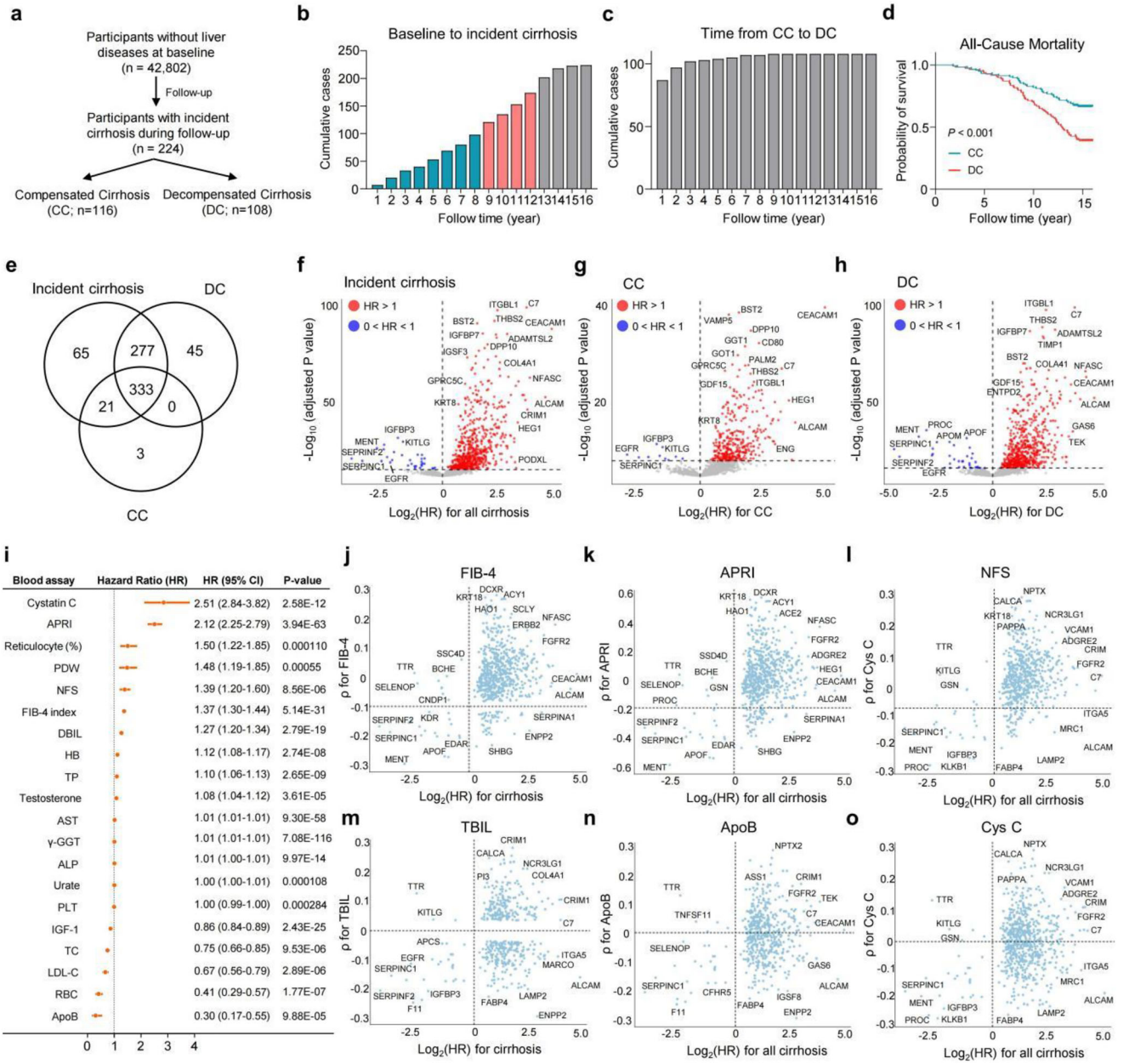

FIGURE 3 | Identification of proteins associated with incident cirrhosis. (a) Schematic diagram of study cohort. 42,802 participants without liver diseases at baseline are included for study and confirmed 224 incident cases including 116 CC and 108 DC during 15.6 median years follow-up. (b, c) Cumulative time-to-onset for cases by incident cirrhosis (b) and CC to DC (c). The number of cases by year of follow-up plotted cumulatively and the year that the proportion of cases diagnosed reached 50% (red) and 80% (grey) demarcated. (d) Overall survival of cirrhosis cases during follow-up comparing CC and DC. (e) The number of cirrhosis-associated proteins shared among pairwise comparisons across three groups. Proteins associated with cirrhosis are identified through Cox proportional hazard regression model adjusted by age, sex, body mass index (BMI), ethnicity, townsend deprivation index (TDI), history of hypertension and diabetes, alcohol intake, and smoking status. (f-h) Volcano plot showing cirrhosis-associated proteins in incident cirrhosis (f), CC (g), and DC (h). The hazard ratio (HR) of proteins was calculated by Cox model above-mentioned. 'Risk' (red) and 'Protective' (blue) proteins were identified by HR > 1 (Log_2 (HR)>0) and HR < 1 (Log_2 (HR) < 0). *P* values were calculated under two-sided tests and adjusted by Bonferroni corrections. (i) Association between clinical parameters and onset of incident cirrhosis. The measure of center and error bars present the HRs and 95% CI of incident cirrhosis risk increase associated with higher parameters level. Clinical parameters with highest and lowest HR are displayed in the graph. (j-o) Shared association between cirrhosis-associated proteins and clinical parameters including FIB-4 index (j), APRI (k), NFS (l), TBIL (m), ApoB (n), and Cys C (o). TBIL, total bilirubin; ApoB, apolipoprotein B; Cys C, cystatin C.

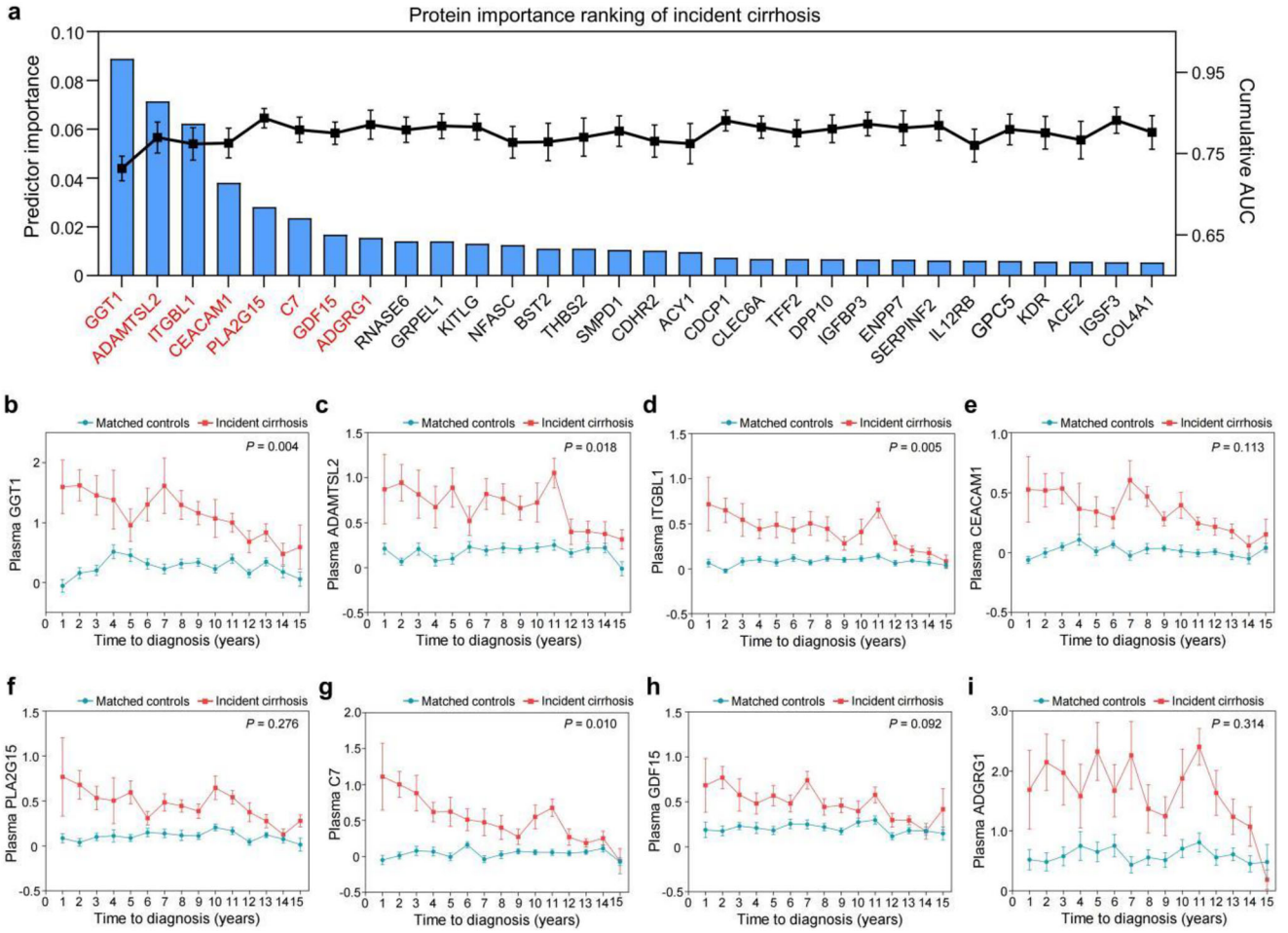

FIGURE 4 | Contribution of proteins to incident cirrhosis prediction. (a) Sequential forward selection was performed on preselected candidate proteins by COX model. The bar chart displays the importance of the ranked proteins based on their contributions to predicting future incident cirrhosis, which was assessed by information gain. The line chart (right axis) illustrates the cumulative AUC values as proteins are sequentially incorporated in each iteration. The top-selected proteins are highlighted in red. Error bars represent standard errors derived from cross-validation. (b-i) Temporal trajectories of plasma GGT1 (b), ADAMTSL2 (c), ITGBL1 (d), CEACAM1 (e), PLA2G15 (f), C7 (g), GDF15 (h), and ADGRG1 (i) expression preceding the cirrhosis diagnosis. Controls were matched cirrhosis cases at each follow-up time point by propensity score matching (PSM) method. Red lines correspond to patients with cirrhosis, and blue lines correspond to matched controls. The dot represents the mean concentrations of selected proteins concerning the corresponding follow-up time for both cases and controls. Error bars represent standard errors. Mann–Kendall trend tests were employed to examine the slope differences of selected plasma protein levels over time for individuals with incident cirrhosis. *P* values were calculated under two-sided tests and multiple comparisons were not applied.

cirrhosis (Figure 5d), and AUC values of 0.803 and 0.818 for 5-year (Figure 5e) and 10-year (Figure 5f) incident cirrhosis, respectively. The γGGT showed the highest predictive accuracy among clinical markers, with an AUC of 0.743, whereas other clinical markers such as ALT, AST, and TBIL exhibited lower AUC values (Figure S1). The combination of γGGT with the proteomic panel resulted in a marginal increase in AUC to 0.759, without a significant improvement. Notably, both γGGT alone and the combined model outperformed other individual clinical biomarkers, indicating limited incremental predictive value from integrating routine clinical parameters with the proteomic panel.

We further divided the participants into high and low expression groups based on baseline protein levels, using the cutoff value determined by the largest Youden index across the follow-up period (Figure 5g–j). Individuals with higher levels of GGT1 ($p < 0.001$; HR = 1.55 [1.46–1.64]) (Figure 5g), ADAMTSL2 ($p < 0.001$; HR = 1.60 [1.50–1.69]) (Figure 5h), ITGBL1 (Figure 5i) ($p < 0.001$; HR = 1.64 [1.55–1.74]), and CEACAM1 (Figure 5j) ($p = 0.049$; HR = 1.06 [1.00–1.13]) exhibited a higher mortality rate than those with lower baseline protein levels.

2.5 | Potential Causal Proteins of Advance Liver Diseases and Assessment of Druggability

To infer causality between the identified proteins and liver fibrosis, cirrhosis, or hepatic failure, we conducted a two-sample MR analysis using pQTL data from the UK Biobank and summary GWAS data for diseases from FinnGen. We identified significant associations between the proteins and advanced liver diseases using the inverse-variance weighting (IVW) method as the primary analytic approach, supplemented by the weighted median, MR-Egger, and weighted mode methods (Table S6). No evidence

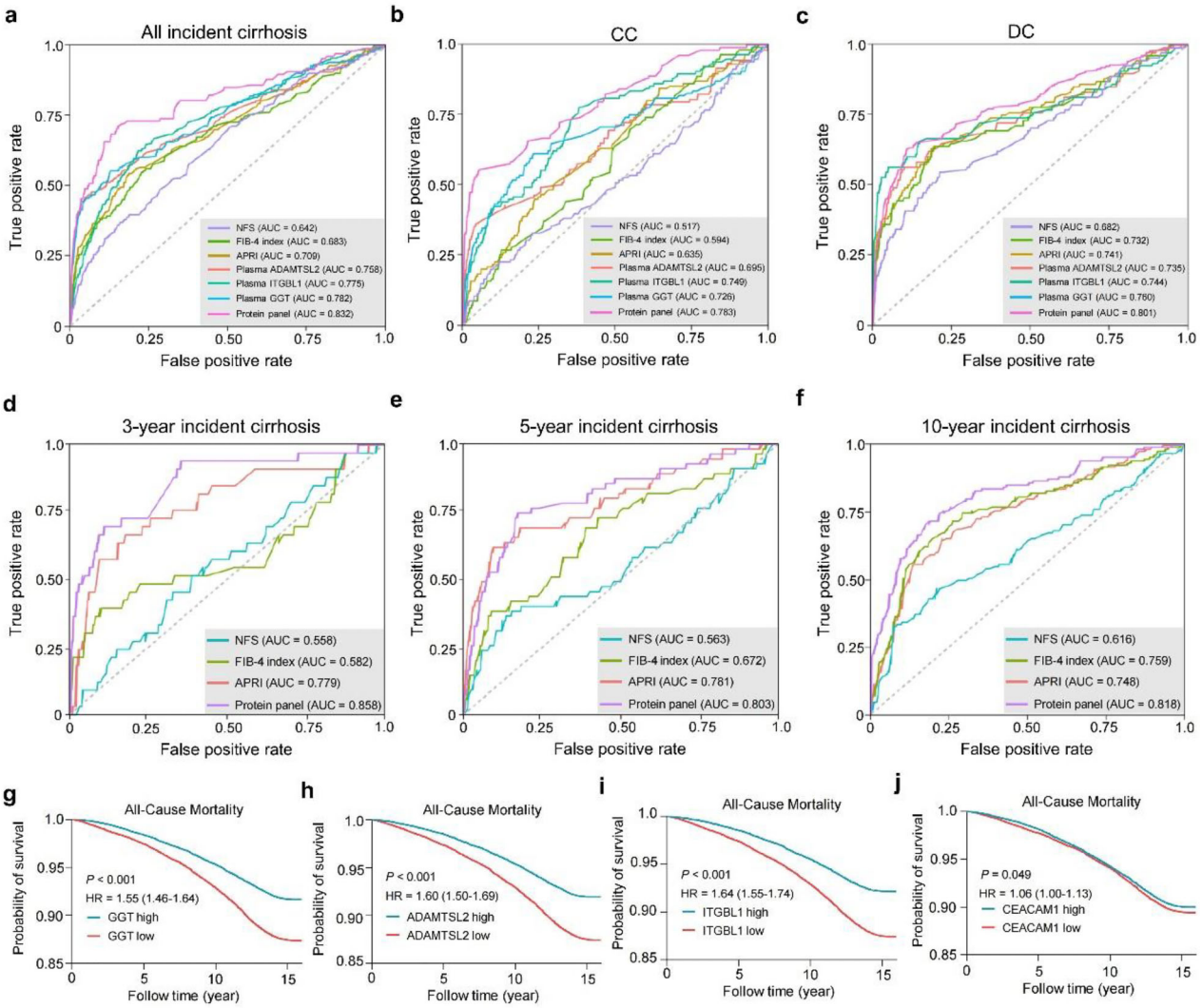

FIGURE 5 | Predictive accuracy of plasma proteins for cirrhosis onset. f, Receiver operating characteristic (ROC) curves illustrate the performance of plasma proteins and fibrosis parameters for predicting all incident cirrhosis (a), CC (b), and DC (c) as well as 3-year (d)/5-year (e)/10-year (f) incident cirrhosis. (g-j) Kaplan–Meier curves present the all-cause mortality over time for individuals with high (red line) and low (blue line) baseline identified plasma protein expression level. The cutoff of high and low protein levels population was determined by the achievement of the largest Youden index. *P* values were calculated under Log-rank tests.

of horizontal pleiotropy or heterogeneity was observed among the instrument variables, and potential causal links driven by influential single-nucleotide polymorphisms (SNPs) were excluded through leave-one-out analyses. The results showed 39 proteins identified as potential causal proteins for hepatic failure and 27 proteins for liver fibrosis/cirrhosis (Figure 6a,b; and Table S6). Two proteins, DCXR and CPXM2, overlapped in the analysis for hepatic failure and liver fibrosis/cirrhosis, suggesting they could play a critical role in the pathogenesis of both disease stages. Importantly, we found that most of the causal proteins overlapped with druggable genes reported by Finan and colleagues (Figure 6c) [25, 26]. The tier 1 and 2 categories include genes targeted by approved therapeutic drugs, biotherapeutic drugs, clinical-phase drug candidates, and drug-like compounds. A total of 11 proteins in tiers 1/2, causally associated with liver fibrosis, cirrhosis, and liver failure, are linked to approved targeted drugs (Figure 5c; Table S7). For example, a drug targeting SERPINC1, Antithrombin-III has been approved as an anticoagulant. Drugs targeting PDCD1, CSF1R, TNFRSF8 (CD30), and EPHA2 have been used to treat cancer [27], graft-versus-host disease [28], Hodgkin's lymphoma [29, 30], leukemia [31], and other diseases. Additionally, ICAM1 is one of the targets for drug with indications to treat multiple sclerosis [32].

2.6 | Independent Validation of the Predictive Model for Cirrhosis Progression

We further performed an independent validation in a small prospective cohort of patients with early-stage fibrosis/cirrhosis (n = 31). We used a compensated cirrhosis cohort due to the long natural history of cirrhosis development in liver disease–free individuals. The primary endpoint was progression to DC. Baseline characteristics of this cohort are summarized in Supplementary Table 8. Selected plasma proteins and clinical blood biomarkers were measured at baseline prior to follow-up. Twelve

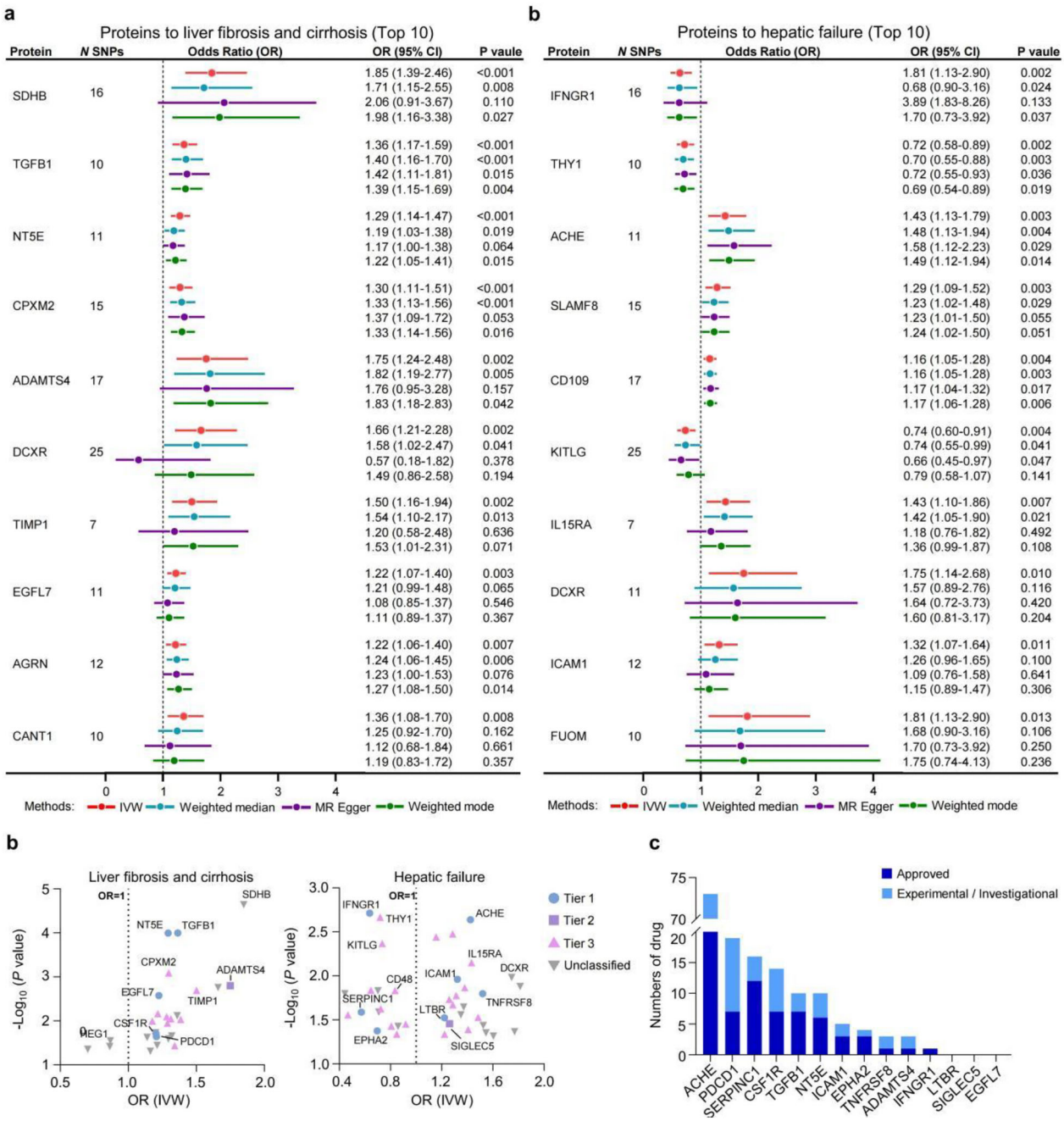

FIGURE 6 | Mendelian randomization (MR) analysis on the associations between cirrhosis and proteins. (a) The IVW method was used as the primary MR analytic approach, complemented by MR–Egger, Weight-median, and Weight mode. The top 10 most significant proteins identified by IVW method in liver fibrosis and cirrhosis and hepatic failure are shown. The dots represent odds ratio (OR) value, and error bars present 95% CIs. Statistical tests were two sided, and FDR correction was applied. (b) A scatter plot showing the enrichment of causal proteins with the druggable genome in liver fibrosis and cirrhosis and hepatic failure. (c) Comparison of the MR findings with drug-target information in DrugBank dataset. The bar plot showing the number of approved and experimental / investigational drugs associated with target proteins.

patients progressed to decompensated cirrhosis during the 1-year follow-up period. The predictive performance of the protein panel was evaluated and compared with conventional clinical parameters. The protein panel demonstrated strong predictive accuracy (AUC = 0.743), while APRI (AUC = 0.579), FIB-4 (AUC = 0.601), and NFS (AUC = 0.539) performed more modestly (Figure S2). These findings further support the potential clinical applicability of the protein panel for predicting cirrhosis progression.

3 | Discussion

Early detection of liver fibrosis and cirrhosis is crucial for preventing the progression to advanced or end-stage liver disease. In this study, we examined plasma proteome changes associated with liver pathophysiology to identify proteins with diagnostic value for detecting liver fibrosis and cirrhosis, using 2920 measured protein analytes from 44,030 UK Biobank participants. In a cohort

of 42,802 participants without liver disease at baseline, we identified 696 proteins linked to the development of incident cirrhosis over more than 15 years of follow-up and found eight proteins with the best predictive performance. Indeed, the protein panel used for detecting incident cirrhosis in this study demonstrates substantially better performance compared to existing best-in-class clinical tests. In an independent prospective cohort, the protein panel also demonstrated predictive value for disease progression. Furthermore, MR analyses confirmed the causal associations of the candidate proteins with cirrhosis. Our study identified pre-diagnostic protein biomarkers for liver fibrosis and cirrhosis, which could potentially enhance the early detection of advanced liver disease.

Advances in high-throughput proteomics offer considerable potential for clinical translation. A subset of protein biomarkers could be measured in a shorter analysis time and at a lower cost by developing targeted proteomics assays [33]. In our study, the cirrhosis-associated protein panel (comprising GGT1, ADAMTSL2, ITGBL1, CEACAM1, PLA2G15, C7, GDF15, and ADGRG1) for predictive models was selected in an unbiased manner based on their contributions to diagnostic performance. These proteins alone showed modest predictive accuracies (AUC 0.7–0.8) and demonstrated desirable predictions (AUC = 0.832) when combined. Classical non-invasive algorithms, such as APRI and FIB-4, are associated with indeterminate-range scores in 30%–50% of patients [34]. A systematic review of 10 noninvasive biomarker panels found that clinically meaningful predictive performance for significant fibrosis was achieved in only 35% of patients with chronic hepatitis C [35]. Given that most routine clinical blood biomarkers remain within normal ranges or show only slight elevation in the early stages of cirrhosis, fibrosis parameters have suboptimal diagnostic accuracy (low AUC values) in the CC population. In the DC population, FIB-4 and APRI demonstrated modest AUC and exhibit comparable accuracy to single protein biomarkers. In contrast, the proteomics model outperformed fibrosis parameters in identifying cirrhosis across all patient populations. Furthermore, changes in GGT1, ADAMTSL2, ITGBL1, and C7 were observed at least 10 years before cirrhosis diagnosis, with their concentrations rising sharply as the onset approached in individuals who developed cirrhosis, compared to those who remained cirrhosis-free. More importantly, the model achieved AUC values of 0.803 and 0.818 for 5 year and 10 year cirrhosis incidence, respectively, demonstrating strong performance for long-term predictions. Our findings highlight the potential of plasma proteomics for accurately detecting cirrhosis at an early stage, which can be used to inform clinical decision-making.

Plasma proteomics primarily identified upregulated proteins in cirrhosis, revealing significant plasma proteome remodeling in both CC and DC. The PPI network for dysregulated proteins highlighted the crucial role of chemokine-related pathways in regulating protein modules throughout the cirrhosis process. Biologically, our findings underscore liver metabolic changes, imbalanced ECM turnover, and immune response in the progression of cirrhosis. While some proteins have already been linked to liver disease, confirming previous studies [19, 23, 36–38], several others such as DPP10, CLSTN2, and EDAR—emerge as promising new circulating biomarkers. Their involvement in ion channel regulation [39], cell adhesion [40], and inflammatory response [41] indicates a key role in liver cirrhosis, warranting further investigation. However, some inconsistencies remain in the conclusions. For example, a top-down proteomics study identified abundant apolipoprotein A-I, haptoglobin, and fibrinogen alpha chain in patients with CC or DC [19]. Additionally, two other studies examining protein biomarkers and pathway perturbations in HBV-related hepatic failure and infection-related acute decompensation of cirrhosis found proteomic signatures, including APOC3, HRG, LGALS3BP, and PLTP that were closely associated with advanced cirrhosis outcomes [36, 37]. In contrast, these proteins were either not detectable or showed less pronounced fold changes in our study. These differences may be attributed to methodological and biological factors. For instance, several studies employed liquid chromatography–mass spectrometry (LC–MS) as the primary proteomic detection method and included proteins derived from liver tissue samples [19, 38]. The Olink platform, which focuses on pre-selected protein panels (e.g., inflammation or cardiovascular markers), may exclude certain proteins like apolipoproteins unless specifically targeted [17]. In contrast, Olink-derived candidates might include novel circulating regulators that fall below the detection thresholds of LC–MS [17]. As a result, non-overlapping dysregulated proteins could reflect platform-specific findings. Nevertheless, the consistent dysregulation of fibrogenic and immune pathways across studies highlights shared pathophysiology in cirrhosis, despite technical and study design variations. To prioritize therapeutic targets and biomarkers for cirrhosis, further integration of multi-platform datasets and functional validation will be crucial.

Moreover, most studies have relied on case-control retrospective comparisons between diagnosed patients and healthy individuals or focused on disease progression from chronic liver disease to advanced liver events. As a result, these studies only identified differentially expressed proteins in chronic liver disease or cirrhosis. In contrast, our study employed a prospective design to investigate the risk transition from a normal liver condition to liver cirrhosis. By utilizing a large-scale longitudinal cohort with proteomic data, our findings validate and extend previous research, revealing important signatures linked to the onset of liver fibrosis and cirrhosis. We observed significant overlap between dysregulated proteins and those associated with cirrhosis, as well as a high degree of similarity in the protein-related pathways. This suggests that the mechanisms linked to these proteins may emerge at earlier stages of liver disease and its progression, including immune dysfunction (characterized by immunodeficiency and increased systemic inflammation) and imbalanced ECM turnover (where collagen formation outweighs its degradation [42]). Therefore, timely recognition and intervention targeting these pathogenic mechanisms could be crucial in preventing disease progression.

Furthermore, we conducted MR analysis to assess the strength of the causal relationship between the associated proteins and the development of liver fibrosis, cirrhosis, and hepatic failure. Our MR findings, which show partial consistency with established target-indication pairs in drug databases [25, 26], provide support for the role of known drug targets in the context of liver fibrosis, cirrhosis, and hepatic failure. For example, SERPINC1 also known as AT-III was linked to a reduced risk of liver fibrosis, cirrhosis, and hepatic failure in both our cohort study and MR analysis. As a coagulation regulator produced by the

liver, AT-III levels decrease with impaired liver function [43], and lower levels of AT-III have been associated with portal vein thrombosis (PVT) in liver cirrhosis patients and liver failure-related mortality [44, 45]. Primary prophylaxis with AT-III has been reported to reduce the risk of PVT in liver cirrhosis patients with portal hypertension [46]. Other causal proteins, such as the tyrosine kinase receptor EPHA2 and the immune checkpoint PDCD1 are currently targeted by drugs used in the treatment of hepatocellular carcinoma [47, 48]. Future studies should focus on further validating the biological functions of these proteins in liver disease development through animal experiments and investigating targeted therapies for these proteins to prevent disease progression in pre-clinical models.

Recent advances in artificial intelligence (AI) have further expanded the potential of large-scale biomedical datasets to improve disease prediction, diagnosis, and personalized management in hepatology. AI-based approaches have shown promising performance in multiple areas of liver disease research, including automated imaging interpretation, biopsy assessment, and prediction of transplant outcomes, demonstrating their ability to handle complex and multidimensional clinical data [49]. In parallel, the integration of AI with omics technologies has opened new avenues for identifying molecular signatures and developing precision medicine strategies in chronic liver diseases. However, current studies applying AI to multi-omics datasets remain relatively limited, particularly at the proteomic level, highlighting an important opportunity for future research [50]. In addition, machine learning models applied to electronic health records have demonstrated high accuracy in identifying patients with cirrhosis, suggesting that AI-driven approaches may facilitate earlier detection and improved disease phenotyping in large populations [51]. In this context, the large-scale proteomic biomarkers identified in the present study may provide a valuable resource for future AI-driven predictive models that integrate clinical, proteomic, and other multi-omics data to improve risk stratification and early detection of liver cirrhosis. Nevertheless, such applications will require further external validation and careful evaluation before clinical implementation.

Several limitations should be considered in our analysis. First, although the UK Biobank provides a comprehensive evaluation of over 2000 circulating proteins, the platform cannot capture the entire human proteome, potentially introducing biases by prioritizing the measurement of secreted proteins. Second, the study cohort predominantly consists of individuals of European ancestry who are middle-aged or older, limiting the generalizability of our findings to other populations. Third, we conducted only forward MR analysis (from protein to diseases) due to a limited number of SNPs reaching genome-wide significance, even after relaxing the p-value threshold to 1×10^{-5}) following clumping. Additionally, the GWAS data used in our analysis were derived from two separate datasets, and because disease information could have been collected prior to baseline protein data, the causal effects observed in our two-sample MR analysis should be interpreted with caution.

Although we performed preliminary validation in an independent prospective cohort, the sample size was relatively small and the follow-up duration was limited. Given the long natural history of cirrhosis development, prospective validation of incident cirrhosis in a liver disease-free population is challenging within a feasible timeframe, therefore, our validation was conducted in individuals with early-stage fibrosis/cirrhosis to assess progression risk. While this design provides supportive evidence for the predictive value of the protein panel in a clinically relevant setting, it may limit direct comparability with the primary prevention context of the UK Biobank analysis. The protein panel identified in this study is not intended to be applied directly as a standalone clinical diagnostic tool in its current form. Instead, it provides a data-driven framework for prioritizing candidate biomarkers and a foundation for developing predictive models of cirrhosis risk. Further validation in larger, population-based cohorts with longer follow-up is warranted.

4 | Conclusion

In conclusion, our study provides a detailed overview of plasma proteomic signatures associated with liver fibrosis, cirrhosis, and hepatic failure. We innovatively identified key protein biomarkers, combining them with clinical indicators, to predict future liver diseases in a large, population-based prospective cohort with long-term follow-up. Notably, several of the identified proteins demonstrated causal associations with liver fibrosis, cirrhosis, and hepatic failure. These findings may offer a novel strategy for screening populations at high risk for liver diseases and could aid in the development of potential therapeutic targets. At present, this work primarily serves as a resource for biomarker discovery and risk stratification, and future studies are needed to validate, refine, and translate these findings into clinically applicable tools.

5 | Methods

5.1 | Study Population

The UK biobank is an ongoing prospective cohort study based on large population, with extensive phenotypic and in-depth proteomic data. Over 500,000 individuals aged 39–70 were successfully recruited during 2006–2010, and their health status was being long-term followed. The study cohort comprised participants who were randomly selected for plasma proteomic analysis at baseline [52]. We excluded participants missing blood assay and >50% proteomics data to gain a 44,030 participants cohort. We design a case-control including 66 cases reported cirrhosis at baseline (10 years prior to baseline data collection) and their matched controls without cirrhosis by using propensity score matching (PSM) analyses. Controls were matched in a ratio of four controls per case by demographic covariates including Age, Sex, BMI, TDI, history of hypertension and diabetes, alcohol intake, and smoking status. Next, we excluded subjects with liver diseases at baseline, yielding a follow-up cohort of 42,802 participants without liver diseases.

5.2 | Disease Definition

The primary outcome was incident cirrhosis, defined by the International Classification of Diseases (ICD)-10 codes (K74, K70.2, K70.3, and K71.7). Decompensation was defined as the presence

of complications including hepatic failure (K72, K70.4), portal hypertension (K76.6), or gastro-intestinal haemorrhage (K92.2) on the basis of cirrhosis. The outcome date for diseases was established from hospital admissions data, primary care reports and death registry records. The follow-up earliest occurrence of until the earliest occurrence date of diagnosis, mortality, or the final available date from hospital records (October 2024). We excluded self-report cases from this paper to ensure the reliability of diagnosis.

5.3 | Plasma Proteomics

Details on the blood sample collection, proteomic data generation, normalization and quality control steps were described in previous studies [52–55]. Plasma proteomic profiling in the UK Biobank was conducted as part of the UK Biobank Pharma Proteomics Project (UKB-PPP). In brief, Blood samples were collected in EDTA tubes and immediately centrifuged at 2500 g for 10 min at 4°C to separate the plasma. The supernatant was subsequently aliquoted and stored at −80°C until analysis. Protein quantification was performed using the Olink Explore Proximity Extension Assay, an antibody-based technology coupled with next-generation sequencing that measures proteins across eight thematic panels, including cardiometabolic, inflammation, neurology and oncology panels. Following centralized preprocessing and quality control procedures implemented by the UK Biobank consortium, measurements for 2923 unique proteins were initially available. To ensure data reliability, stringent quality control procedures were applied, including assessments of inter- and intraplate variation, which remained below 20% and 10%, respectively, for all Olink panels. Protein abundance was reported in Normalized Protein eXpression (NPX) values, a log2-transformed relative quantification generated through normalization to internal controls, followed by batch correction using the standardized UK Biobank/Olink pipeline. Proteins with missing data exceeding 50% were excluded from the final dataset, resulting in the inclusion of 2920 proteins in this study. Baseline blood biomarker data were obtained from the UK Biobank laboratory assays. Blood biochemistry data and haematology measurements were collected at the baseline assessment visit following standardized UK Biobank protocols. In the present study, 61 routinely measured blood biomarkers, including 30 biochemical markers and 31 haematological markers were selected for analysis. All biomarkers included had passed the standard UK Biobank quality control procedures. Missing values for protein and blood assay data were imputed using the K-nearest neighbors (KNN) algorithm.

5.4 | Covariates

We selected demographic factors included age, sex, body mass index (BMI), ethnicity, townsend deprivation index (TDI), history of hypertension and diabetes, alcohol intake, and smoking status. Alcohol intake frequency and smoking status was reported from self-rated questions. Hypertension and diabetes history was obtained according to the first occurrence dates before baseline information collection. We also include blood assays including ALT, AST, and so on. Details of patient characteristics are summarized in Table 1. The formula for calculation of FIB-4 index: FIB-4 = Age (years) * AST (U/L) / [PLT (10ˆ9/L) * ALTˆ(½) (U/L)]; APRI: [AST/40 (U/L) × 100)] / PLT (10ˆ9/L); NFS: -1.765 + 0.037 * Age (years) + 0.094 * BMI (kg/m^2) + 1.13 * Diabetes (yes = 1, no = 0) + 0.99 * (AST/ALT) - 0.013* PLT (10ˆ9/L) - 0.66 * ALB (g/dL). Associations between plasma proteins and clinical measurements were determined by Spearman correlation.

5.5 | Independent Validation Cohort

Peripheral blood was collected from 31 patients with compensated cirrhosis in December 2024. The diagnosis of cirrhosis was determined through comprehensive medical record review, including FibroScan, magnetic resonance elastography, or liver biopsy. Demographic characteristics, clinical data, and laboratory parameters were collected for each participant at the time of blood sampling. All participants were followed for 1 year, and decompensation was defined according to established clinical criteria [56]. Plasma was isolated at baseline and stored at −80°C until analysis. The following ELISA kit were used to measure: ADAMTSL2 (Abebio, AE23855HU), PLA2G15 (Abebio, AE27233HU), C7 (Abebio, AE52986HU), ITGBL1 (Abebio, AE37634HU), GDF15 (Abebio, AE42393HU), ADGRG1 (Abbexa, abx526041), and GGT1 (Signalway Antibody, EK14228). The measured protein levels were then incorporated into the previously established LightGBM model to assess predictive performance for disease progression.

5.6 | Statistical Analysis

5.6.1 | Dysregulated and Cirrhosis-Associated Proteins

Dysregulated proteins were identified using "limma" R package with a cutoff of $\log_2(|\text{fold change}|) > 0.5$ and FDR-adjusted p-value < 0.05. The COX proportional hazards model was performed to identified the associations between plasma NPX value and incident cirrhosis, adjusting for covariates including Age, Sex, BMI, TDI, history of hypertension and diabetes, alcohol intake, and smoking status. No further scaling or transformation of NPX values was applied before differential expression or Cox regression analyses. Bonferroni corrections were applied to evaluate significant associations ($p < 0.05$ / [2,920*2]). Hazard ratios (HRs) values, 95% CIs and p values for the significant proteins were reported.

5.6.2 | Functional Annotation and Enrichment Analysis

Liver-related proteins were annotated according to the HPA, which defines as "liver-enriched" with at least 5-fold higher mRNA levels in liver compared with all other tissues. "Liver-specific" proteins were identified according to previously published studies. Enrichment analysis was performed on the DEPs and significant proteins derived from Cox proportional hazard regressions model after Bonferroni correction. We used "clusterProfiler" R package to uncover the related pathways and biological processes based on KEGG database. Fisher's exact test was employed for the statistical significance.

5.6.3 | Protein Co-Expression Network Analysis

The PPI network of 161 overlapped proteins was constructed using STRING datasets and visualized by Cytoscape [57, 58]. MCC method was utilized to identify hub proteins by the cytoHubba plugin in Cytoscape [59].

5.6.4 | Prediction and Diagnostic Modelling

The prediction model of diseases diagnosis was developed using machine learning algorithm light gradient boosting machine (LightGBM). For the identification of important proteins, a two-step procedure was employed involving initial variable importance ranking followed by sequential forward selection. First, proteins that reached statistical significance after Bonferroni correction in Cox proportional hazards models were introduced into LightGBM classifier. Within this framework, each protein's contribution was quantified using information gain, thereby reflecting its potential in predicting future disease onset. Afterward, a sequential forward selection strategy was implemented. Proteins were added one at a time to the LightGBM classifier in order of decreasing importance. This process continued until the addition of further proteins no longer resulted in a significant improvement in the area under the receiver operating characteristic curve (AUC), as artificially determined by non-significant results in three consecutive DeLong tests.

To benchmark performance, separate models were also developed using clinical-demographic variables (including covariates and blood count and biochemistry biomarkers). An integrated model combining both protein and clinical-demographic features was further constructed to evaluate the additive predictive value of the proteomic data. Comparisons among these models were performed using the DeLong test. Model training and optimization were conducted using a nested leave-one-region-out cross-validation framework. The dataset was divided into 10-folds. In each iteration, nine folds were used as the training set and the remaining fold as an independent test set, with the process repeated until every fold had served as the test set. Within each cross-validation loop, hyperparameters were tuned by randomly splitting the training data into an 80% development subset and a 20% validation subset which helped to further mitigate overfitting. Feature selection was also implemented to significantly reduce the pool of proteins included in the prediction task. The untouched test sets were then used exclusively for evaluating model performance. Given the low incidence of outcome events, we increased the sample weights for these events to address the class imbalance. Finally, to evaluate the robustness and generalizability predictive accuracy of protein panel, we repeated the above analyses among these target populations: CC, DC, and 3-year/5-year/10-year incident cirrhosis.

5.6.5 | Temporal Trajectories Analysis

All incident cases of cirrhosis identified during follow-up were considered as cases, whereas controls were matched to cirrhosis cases in a ratio of four controls per patient case by PSM methods. Controls were matched cirrhosis cases at each follow-up time point by demographic covariant above mentioned. Line chart shows the mean concentrations of selected proteins concerning the corresponding follow-up time for both cases and controls. Mann–Kendall trend tests were employed to examine the monotonic trends in plasma levels over time of those who developed cirrhosis.

5.6.6 | Survival Analysis

The optimal cutoff of high and low protein levels population was determined by maximizing the Youden index from the ROC of incident cirrhosis. we constructed Kaplan–Meier survival curves to visualize the progression of outcome events over time. Log-rank tests were used to compare the statistical significance between two group (high and low baseline protein panel).

5.6.7 | MR

The pQTL mapping in GWAS was conducted in European participants genotyping data from the UK Biobank [52]. To avoid bias from participant overlap, we utilized the summary GWAS data for liver fibrosis/cirrhosis (2653 cases and 485,213 controls) and hepatic failure (1272 cases and 485,213 controls) from the FinnGen R12 [60]. In our primary MR analyses, we used the inverse-variance weighted (IVW) as the primary analytic method to estimate causal effects between proteins and disease outcomes. To avoid confounding from uncorrelated pleiotropy and ensure robustness, we conducted three additional analyses including MR-Egger, weighted-median, and weighted mode [61, 62]. Genetic instruments were derived from genome-wide significant pQTLs ($P < 5 * 10^{-8}$) and were pruned for independence using clumping (1,000 kb distance with a maximum linkage disequilibrium LD r^2 of 0.01). Instrument strength was evaluated using F-statistics, and the Egger intercept test was applied to detect directional pleiotropy [63]. Cochran's Q test assessed heterogeneity among the instruments, while the MR-PRESSO method was employed to identify and correct for potential outlier SNPs driving pleiotropic effects [64]. Additionally, leave-one-out analyses were performed to gauge the influence of individual SNPs on the overall causal estimate. All MR analyses were executed using the "TwoSampleMR" R package (v 0.6.8) [65]. Causal estimates are presented as odds ratios (ORs) with 95% confidence intervals (CIs).

5.6.8 | Evaluation of Druggability

The catalog of 4479 druggable genes was derived from Finan et al. in previously published study [25]. These genes were categorized into three tiers based on their role in drug development. Tier 1 includes 1427 genes targeted by clinical-phase drug candidates, approved biotherapeutics, and small-molecule drugs. Tier 2 consists of 682 genes linked to drug-like compounds or closely related to known drug targets. Tier 3 comprises the remaining 2370 genes, including extracellular or secreted proteins with low similarity to approved targets. We assessed the overlap between causal proteins and different tiers of druggable genes. The DrugBank (https://go.drugbank.com/) database [26] was

employed to perform drug assessment of protein targets related to liver fibrosis, cirrhosis, and hepatic failure after MR analysis. The identified potential drug name, biological function, indications, and drug development details was shown in Table S7.

Author Contributions

Jitian He: Ddata curation, methodology, Writing – original draft, formal analysis, conceptualization, writing revised manuscript & editing. **Bo Li:** Data curation, methodology, writing – original draft, writing revised manuscript & editing. **Yuping Yan:** Data curation, writing – original draft, writing revised manuscript & editing. **Yajie Wang:** Data curation. **Mingxi Zhang:** Conceptualization. **Renyuan Sun:** Conceptualization. **Baohua Hou:** resources. **Shanzhou Huang:** Resources. **Limin Zhen:** Resources, conceptualization, supervision. **Dongping Wang:** Resources, investigation, conceptualization, supervision. **Chuanzhao Zhang:** Resources, investigation, methodology, formal analysis, conceptualization, supervision.

Acknowledgements

We would like to thank all participants and project team members from the UK Biobank

Funding

This study was supported by National Natural Science Foundation of China (82472762).

Ethics Statement

The UK Biobank study received ethical approval from the North West Multicenter Research Ethical Committee (REC reference 11/NW/0382). All participants provided written informed consent. The standard procedures were approved by the Ethics Committee for Medical Research at Guangdong Province people's Hospital (approval no. S2024-643-02).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data used in the present study are available from the UK Biobank (https://www.ukbiobank.ac.uk) with restrictions applied (Application number: 105314). GWAS summary statistics used can be found via the UK Biobank and FinnGen R12.

References

1. C. Villanueva, D. Tripathi, and J. Bosch, "Preventing the Progression of Cirrhosis to Decompensation and Death," *Nature Reviews Gastroenterology & Hepatology* 22 (2025): 265–280, https://doi.org/10.1038/s41575-024-01031-x.

2. M. Ubeda, L. Mun˜oz, M. Borrero, et al., "Critical Role of the Liver in the Induction of Systemic Inflammation in Rats With Preascitic Cirrhosis," *Hepatology* 52 (2010): 2086–2095, https://doi.org/10.1002/hep.23961.

3. T. Zuo, Q. Xie, J. Liu, et al., "Macrophage-Derived Cathepsin S Remodels the Extracellular Matrix to Promote Liver Fibrogenesis," *Gastroenterology* 165 (2023): 746–761 e716, https://doi.org/10.1053/j.gastro.2023.05.039.

4. C. Bernsmeier, E. Triantafyllou, R. Brenig, et al., "CD14(+) CD15(-) HLA-DR(-) Myeloid-derived Suppressor Cells Impair Antimicrobial Responses in Patients With Acute-on-chronic Liver Failure," *Gut* 67 (2018): 1155–1167, https://doi.org/10.1136/gutjnl-2017-314184.

5. J. P. Iredale, R. C. Benyon, J. Pickering, et al., "Mechanisms of Spontaneous Resolution of Rat Liver Fibrosis. Hepatic Stellate Cell Apoptosis and Reduced Hepatic Expression of Metalloproteinase Inhibitors," *Journal of Clinical Investigation* 102 (1998): 538–549, https://doi.org/10.1172/JCI1018.

6. M. Lu, J. Li, T. Zhang, et al., "Serum Biomarkers Indicate Long-term Reduction in Liver Fibrosis in Patients with Sustained Virological Response to Treatment for HCV Infection," *Clinical Gastroenterology and Hepatology* 14 (2016): 1044–1055 e1043, https://doi.org/10.1016/j.cgh.2016.01.009.

7. E. Nilsson, H. Anderson, K. Sargenti, S. Lindgren, and H. Prytz, "Clinical Course and Mortality by Etiology of Liver Cirrhosis in Sweden: A Population Based, Long-term Follow-up Study of 1317 Patients," *Alimentary Pharmacology & Therapeutics* 49 (2019): 1421–1430, https://doi.org/10.1111/apt.15255.

8. P. Ginès, A. Krag, J. G. Abraldes, E. Solà, N. Fabrellas, and P. S. Kamath, "Liver Cirrhosis," *Lancet* 398 (2021): 1359–1376, https://doi.org/10.1016/S0140-6736(21)01374-X.

9. C. Wai, J. K. Greenson, R. J. Fontana, et al., "A Simple Noninvasive Index Can Predict both Significant Fibrosis and Cirrhosis in Patients With Chronic hepatitis C," *Hepatology* 38 (2003): 518–526, https://doi.org/10.1053/jhep.2003.50346.

10. R. K. Sterling, E. Lissen, N. Clumeck, et al., "Development of a Simple Noninvasive Index to Predict Significant Fibrosis in Patients With HIV/HCV Coinfection," *Hepatology* 43 (2006): 1317–1325, https://doi.org/10.1002/hep.21178.

11. Y. Deng, J. You, Y. He, et al., "Atlas of the Plasma Proteome in Health and Disease in 53,026 Adults," *Cell* 188 (2025): 253–271 e257, https://doi.org/10.1016/j.cell.2024.10.045.

12. J. Carrasco-Zanini, M. Pietzner, J. Davitte, et al., "Proteomic Signatures Improve Risk Prediction for Common and Rare Diseases," *Nature Medicine* 30 (2024): 2489–2498, https://doi.org/10.1038/s41591-024-03142-z.

13. L. J. Goeminne, A. Vladimirova, A. Eames, et al., "Plasma Protein-based Organ-specific Aging and Mortality Models Unveil Diseases as Accelerated Aging of Organismal Systems," *Cell Metabolism* 37 (2025): 205–222 e206, https://doi.org/10.1016/j.cmet.2024.10.005.

14. J. C. Tran, L. Zamdborg, D. R. Ahlf, et al., "Mapping Intact Protein Isoforms in Discovery Mode Using Top-Down Proteomics," *Nature* 480 (2011): 254–258, https://doi.org/10.1038/nature10575.

15. A. I. Nesvizhskii and R. Aebersold, "Interpretation of Shotgun Proteomic Data: The Protein Inference Problem," *Molecular & Cellular Proteomics* 4 (2005): 1419–1440, https://doi.org/10.1074/mcp.R500012-MCP200.

16. L. M. Smith and N. L. Kelleher, "Proteoform: A Single Term Describing Protein Complexity," *Nature Methods* 10 (2013): 186–187, https://doi.org/10.1038/nmeth.2369.

17. A. Petrera, C. von Toerne, J. Behler, et al., "Multiplatform Approach for Plasma Proteomics: Complementarity of Olink Proximity Extension Assay Technology to Mass Spectrometry-Based Protein Profiling," *Journal of Proteome Research* 20 (2021): 751–762, https://doi.org/10.1021/acs.jproteome.0c00641.

18. A. J. Sanyal, S. A. Williams, J. E. Lavine, et al., "Defining the Serum Proteomic Signature of Hepatic Steatosis, Inflammation, Ballooning and Fibrosis in Non-alcoholic Fatty Liver Disease," *Journal of Hepatology* 78 (2023): 693–703, https://doi.org/10.1016/j.jhep.2022.11.029.

19. E. Forte, J. M. Sanders, I. Pla, et al., "Top-Down Proteomics Identifies Plasma Proteoform Signatures of Liver Cirrhosis Progression," *Molecular & Cellular Proteomics* 23 (2024): 100876, https://doi.org/10.1016/j.mcpro.2024.100876.

20. M. Uhlén, L. Fagerberg, B. M. Hallström, et al., "Proteomics. Tissue-based Map of the human Proteome," *Science* 347 (2015): 1260419, https://doi.org/10.1126/science.1260419.

21. E. S. Baker, K. E. Burnum-Johnson, J. M. Jacobs, et al., "Advancing the High Throughput Identification of Liver Fibrosis Protein Signatures Using Multiplexed Ion Mobility Spectrometry," *Molecular & Cellular Proteomics* 13 (2014): 1119–1127, https://doi.org/10.1074/mcp.M113.034595.

22. H. Yang, D. Atak, M. Yuan, et al., "Integrative Proteo-transcriptomic Characterization of Advanced Fibrosis in Chronic Liver Disease Across Etiologies," *Cell Reports Medicine* 6 (2025): 101935, https://doi.org/10.1016/j.xcrm.2025.101935.

23. W. Hou, M. Janech, P. M. Sobolesky, et al., "Proteomic Screening of Plasma Identifies Potential Noninvasive Biomarkers Associated With Significant/Advanced Fibrosis in Patients With Nonalcoholic Fatty Liver Disease," *Bioscience Reports* 40 (2020): BSR20190395, https://doi.org/10.1042/BSR20190395.

24. J. W. Dear, J. I. Clarke, B. Francis, et al., "Risk Stratification After Paracetamol Overdose Using Mechanistic Biomarkers: Results From Two Prospective Cohort Studies," *Lancet Gastroenterol Hepatol* 3 (2018): 104–113, https://doi.org/10.1016/S2468-1253(17)30266-2.

25. C. Finan, A. Gaulton, F. A. Kruger, et al., "The Druggable Genome and Support for Target Identification and Validation in Drug Development," *Science Translational Medicine* 9 (2017): eaag1166, https://doi.org/10.1126/scitranslmed.aag1166.

26. C. Knox, M. Wilson, C. Klinger, et al., "DrugBank 6.0: The DrugBank Knowledgebase for 2024," *Nucleic Acids Research* 52 (2024): D1265–D1275, https://doi.org/10.1093/nar/gkad976.

27. M. R. Migden, D. Rischin, C. D. Schmults, et al., "PD-1 Blockade With Cemiplimab in Advanced Cutaneous Squamous-Cell Carcinoma," *New England Journal of Medicine* 379 (2018): 341–351, https://doi.org/10.1056/NEJMoa1805131.

28. C. L. Kitko, M. Arora, Z. DeFilipp, et al., "Axatilimab for Chronic Graft-Versus-Host Disease after Failure of at Least Two Prior Systemic Therapies: Results of a Phase I/II Study," *Journal of Clinical Oncology* 41 (2023): 1864–1875, https://doi.org/10.1200/JCO.22.00958.

29. D. A. Eichenauer, A. Plütschow, S. Kreissl, et al., "Incorporation of brentuximab vedotin Into First-line Treatment of Advanced Classical Hodgkin's Lymphoma: Final Analysis of a Phase 2 Randomised Trial by the German Hodgkin Study Group," *The Lancet Oncology* 18 (2017): 1680–1687, https://doi.org/10.1016/S1470-2045(17)30696-4.

30. J. A. Francisco, C. G. Cerveny, D. L. Meyer, et al., "cAC10-vcMMAE, an Anti-CD30-monomethyl Auristatin E Conjugate With Potent and Selective Antitumor Activity," *Blood* 102 (2003): 1458–1465, https://doi.org/10.1182/blood-2003-01-0039.

31. M. Talpaz, N. P. Shah, H. Kantarjian, et al., "Dasatinib in Imatinib-resistant Philadelphia Chromosome-positive Leukemias," *New England Journal of Medicine* 354 (2006): 2531–2541, https://doi.org/10.1056/NEJMoa055229.

32. E. Pucci, G. Giuliani, A. Solari, et al., "Natalizumab for Relapsing Remitting Multiple Sclerosis," *Cochrane Database of Systematic Reviews* 10 (2011): CD007621, https://doi.org/10.1002/14651858.CD007621.pub2.

33. H. G. Budayeva and D. S. Kirkpatrick, "Monitoring Protein Communities and Their Responses to Therapeutics," *Natural Review Drug Discovery* 19 (2020): 414–426, https://doi.org/10.1038/s41573-020-0063-y.

34. K. Patel and G. Sebastiani, "Limitations of Non-invasive Tests for Assessment of Liver Fibrosis," *JHEP Reports* 2 (2020): 100067, https://doi.org/10.1016/j.jhepr.2020.100067.

35. J. Parkes, I. N. Guha, P. Roderick, and W. Rosenberg, "Performance of Serum Marker Panels for Liver Fibrosis in Chronic hepatitis C," *Journal of Hepatology* 44 (2006): 462–474, https://doi.org/10.1016/j.jhep.2005.10.019.

36. P. Garg, N. Verma, A. Valsan, et al., "Proteomics Guided Biomarker Discovery, Validation, and Pathway Perturbation in Infection-related Acute Decompensation of Cirrhosis," *Clinical Gastroenterology and Hepatology* 24 (2025): 407–420, https://doi.org/10.1016/j.cgh.2025.01.005.

37. Z. Sun, X. Liu, D. Wu, et al., "Circulating Proteomic Panels for Diagnosis and Risk Stratification of Acute-on-chronic Liver Failure in Patients With Viral hepatitis B," *Theranostics* 9 (2019): 1200–1214, https://doi.org/10.7150/thno.31991.

38. L. Niu, M. Thiele, P. E. Geyer, et al., "Noninvasive Proteomic Biomarkers for Alcohol-related Liver Disease," *Nature Medicine* 28 (2022): 1277–1287, https://doi.org/10.1038/s41591-022-01850-y.

39. H. Li, Y. Qu, Y. C. Lu, et al., "DPP10 is an Inactivation Modulatory Protein of Kv4.3 and Kv1.4," *American Journal of Physiology. Cell Physiology* 291 (2006): C966–976, https://doi.org/10.1152/ajpcell.00571.2005.

40. G. Hintsch, A. Zurlinden, V. Meskenaite, et al., "The Calsyntenins–a family of Postsynaptic Membrane Proteins With Distinct Neuronal Expression Patterns," *Molecular and Cellular Neuroscience* 21 (2002): 393–409, https://doi.org/10.1006/mcne.2002.1181.

41. J. Vial, A. Royet, P. Cassier, et al., "The Ectodysplasin Receptor EDAR Acts as a Tumor Suppressor in Melanoma by Conditionally Inducing Cell Death," *Cell Death and Differentiation* 26 (2019): 443–454, https://doi.org/10.1038/s41418-018-0128-1.

42. M. Thiele, S. Johansen, N. S. Gudmann, et al., "Progressive Alcohol-related Liver Fibrosis Is Characterised by Imbalanced Collagen Formation and Degradation," *Alimentary pharmacology & therapeutics* 54 (2021): 1070–1080, https://doi.org/10.1111/apt.16567.

43. A. Tripodi, M. Primignani, V. Chantarangkul, et al., "An Imbalance of Pro- vs Anti-coagulation Factors in Plasma From Patients With Cirrhosis," *Gastroenterology* 137 (2009): 2105–2111, https://doi.org/10.1053/j.gastro.2009.08.045.

44. T. Suda, H. Takatori, T. Hayashi, et al., "Plasma Antithrombin III Levels Can Be a Prognostic Factor in Liver Cirrhosis Patients With Portal Vein Thrombosis," *International Journal of Molecular Sciences* 24 (2023): 7732, https://doi.org/10.3390/ijms24097732.

45. H. Kawanaka, T. Akahoshi, N. Kinjo, et al., "Impact of Antithrombin III Concentrates on Portal Vein Thrombosis After Splenectomy in Patients With Liver Cirrhosis and Hypersplenism," *Annals of Surgery* 251 (2010): 76–83, https://doi.org/10.1097/SLA.0b013e3181bdf8ad.

46. H. Kawanaka, T. Akahoshi, S. Itoh, et al., "Optimizing Risk Stratification in Portal Vein Thrombosis After Splenectomy and Its Primary Prophylaxis With Antithrombin III Concentrates and Danaparoid Sodium in Liver Cirrhosis With Portal Hypertension," *Journal of the American College of Surgeons* 219 (2014): 865–874, https://doi.org/10.1016/j.jamcollsurg.2014.07.939.

47. A. G. Duffy and T. F. Greten, "Liver Cancer: Regorafenib as Second-line Therapy in Hepatocellular Carcinoma," *Nature reviews Gastroenterology & Hepatology* 14 (2017): 141–142, https://doi.org/10.1038/nrgastro.2017.7.

48. R. K. Kelley, "Atezolizumab plus Bevacizumab—A Landmark in Liver Cancer," *New England Journal of Medicine* 382 (2020): 1953–1955, https://doi.org/10.1056/NEJMe2004851.

49. E. Avramidou, D. Todorov, G. Katsanos, et al., "AI Innovations in Liver Transplantation: From Big Data to Better Outcomes," *Livers* 5 (2025): 14, https://doi.org/10.3390/livers5010014.

50. J. S. Obeid, A. Khalifa, B. Xavier, H. Bou-Daher, and D. C. Rockey, "An AI Approach for Identifying Patients with Cirrhosis," *Journal of Clinical Gastroenterology* 57 (2023): 82–88, https://doi.org/10.1097/MCG.0000000000001586.

51. S. Ghosh, X. Zhao, M. Alim, M. Brudno, and M. Bhat, "Artificial Intelligence Applied to 'omics Data in Liver Disease: Towards a Personalised Approach for Diagnosis, Prognosis and Treatment," *Gut* 74 (2025): 295–311, https://doi.org/10.1136/gutjnl-2023-331740.

52. B. B. Sun, J. Chiou, M. Traylor, et al., "Plasma Proteomic Associations With Genetics and Health in the UK Biobank," *Nature* 622 (2023): 329–338, https://doi.org/10.1038/s41586-023-06592-6.

53. G. H. Eldjarn, E. Ferkingstad, S. H. Lund, et al., "Large-scale Plasma Proteomics Comparisons Through Genetics and Disease Associations," *Nature* 622 (2023): 348–358, https://doi.org/10.1038/s41586-023-06563-x.

54. L. Wik, N. Nordberg, J. Broberg, et al., "Proximity Extension Assay in Combination With Next-Generation Sequencing for High-throughput Proteome-wide Analysis," *Molecular & Cellular Proteomics* 20 (2021): 100168, https://doi.org/10.1016/j.mcpro.2021.100168.

55. P. Elliott and T. C. Peakman, "The UK Biobank Sample Handling and Storage Protocol for the Collection, Processing and Archiving of human Blood and Urine," *International Journal of Epidemiology* 37 (2008): 234–244, https://doi.org/10.1093/ije/dym276.

56. G. Garcia-Tsao, J. G. Abraldes, A. Berzigotti, and J. Bosch, "Portal Hypertensive Bleeding in Cirrhosis: Risk Stratification, Diagnosis, and Management: 2016 Practice Guidance by the American Association for the study of liver diseases," *Hepatology* 65 (2017): 310–335, https://doi.org/10.1002/hep.28906.

57. D. Szklarczyk, R. Kirsch, M. Koutrouli, et al., "The STRING Database in 2023: Protein-protein Association Networks and Functional Enrichment Analyses for any Sequenced Genome of Interest," *Nucleic Acids Research* 51 (2023): D638–D646, https://doi.org/10.1093/nar/gkac1000.

58. C. Chin, S. Chen, H. Wu, C. Ho, M. Ko, and C. Lin, "cytoHubba: Identifying Hub Objects and Sub-networks From Complex Interactome," *BMC Systems Biology* 8, no. 4 (2014): S11, https://doi.org/10.1186/1752-0509-8-S4-S11.

59. M. S. Cline, M. Smoot, E. Cerami, et al., "Integration of Biological Networks and Gene Expression Data Using Cytoscape," *Nature Protocols* 2 (2007): 2366–2382, https://doi.org/10.1038/nprot.2007.324.

60. M. I. Kurki, J. Karjalainen, P. Palta, et al., "FinnGen Provides Genetic Insights From a Well-phenotyped Isolated Population," *Nature* 613 (2023): 508–518, https://doi.org/10.1038/s41586-022-05473-8.

61. J. Bowden, G. Davey Smith, P. C. Haycock, and S. Burgess, "Consistent Estimation in Mendelian Randomization With some Invalid Instruments Using a Weighted Median Estimator," *Genetic Epidemiology* 40 (2016): 304–314, https://doi.org/10.1002/gepi.21965.

62. J. Bowden, G. Davey Smith, and S. Burgess, "Mendelian Randomization With Invalid Instruments: Effect Estimation and Bias Detection Through Egger Regression," *International Journal of Epidemiology* 44 (2015): 512–525, https://doi.org/10.1093/ije/dyv080.

63. J. Morrison, N. Knoblauch, J. H. Marcus, M. Stephens, and X. He, "Mendelian Randomization Accounting for Correlated and Uncorrelated Pleiotropic Effects Using Genome-wide Summary Statistics," *Nature Genetics* 52 (2020): 740–747, https://doi.org/10.1038/s41588-020-0631-4.

64. M. Verbanck, C. Y. Chen, B. Neale, and R. Do, "Detection of Widespread Horizontal Pleiotropy in Causal Relationships Inferred From Mendelian Randomization Between Complex Traits and Diseases," *Nature Genetics* 50 (2018): 693–698, https://doi.org/10.1038/s41588-018-0099-7.

65. G. Hemani, J. Zheng, B. Elsworth, et al., "The MR-Base Platform Supports Systematic Causal Inference Across the human Phenome," *Elife* 7 (2018): e34408, https://doi.org/10.7554/eLife.34408.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File 1: prca70050-sup-0001-SuppMat.pdf.

Supporting File 2: prca70050-sup-0002-TableS1-S8.xlsx.