

DISEASES AND DISORDERS

Cross-trait genomic modeling reveals the polygenic architecture and systemic impact of MASLD

Mingyi Du^{1,2,3†}, Huangbo Yuan^{1,2,3†}, Tianhao Wu^{1,2,3}, Yanfeng Jiang^{1,2,3}, Chen Suo^{2,4,5}, Li Jin^{2,3}, Tiejun Zhang^{2,4,5*}, Zhenqiu Liu^{1,2,3*}, Xingdong Chen^{2,3,6*}

Metabolic dysfunction–associated steatotic liver disease (MASLD) is a globally prevalent disease, yet its genetic architecture remains incompletely characterized. We integrated genome-wide association study data from multiple cohorts totaling nearly 3 million individuals of European ancestry and applied cross-trait genomic modeling of hepatic fat and seven cardiometabolic traits to construct an MASLD-specific polygenic architecture. We identified 128 risk variants across 100 loci and prioritized 55 effector genes, including established (e.g., *PNPLA3* and *TM6SF2*) and previously unreported candidates (e.g., *NRXN3* and *FRMD5*). A phenome-wide scan of the MASLD polygenic risk score revealed broad associations spanning hepatic, cardiometabolic, renal, endocrine, and neuropsychiatric systems. Using a two-step, proteome-wide Mendelian randomization across >4900 plasma proteins, we identified potential mediators linking MASLD to disease. Validation in population-based cohort pinpointed seven proteins (e.g., *FURIN*, aldehyde dehydrogenase 2, and apolipoprotein M) mediating up to 50.6% of the cardiometabolic risk attributable to MASLD. Our findings delineate the polygenic architecture of MASLD, highlight its multisystem consequences, and nominate translational biomarkers for precision prevention.

INTRODUCTION

Metabolic dysfunction–associated steatotic liver disease (MASLD), previously known as nonalcoholic fatty liver disease (NAFLD), is a globally prevalent metabolic disorder with increasing prevalence (1–3). Ample evidence suggests that MASLD is not merely a liver-centered disorder but a multisystem condition that contributes to cardiometabolic, renal, and endocrine diseases (4, 5). Biologically, MASLD is a complex trait (6, 7), defined by the coexistence of hepatic steatosis and systemic cardiometabolic dysfunction. However, most genome-wide association studies (GWASs) to date have focused either on hepatic fat content, quantified by magnetic resonance imaging proton density fat fraction (MRI-PDFF), or on binary NAFLD phenotypes derived from electronic health records (8–11). While these studies have provided important genetic insights into hepatic fat accumulation and NAFLD (12–14), they fall short of capturing the broader biological architecture of MASLD, which integrates both liver-specific and systemic metabolic components. To advance mechanistic understanding and inform precision prevention, genetic studies tailored specifically to MASLD are therefore critically needed.

Unraveling the genetic architecture of MASLD may also clarify its biological links to extrahepatic complications and facilitate the development of targeted interventions (15, 16). Although epidemiological studies have consistently shown MASLD to be a multisystem disorder (1, 17–19), the molecular mechanisms underlying these systemic

effects remain largely unresolved. Circulating proteins in plasma may represent promising intermediaries in the MASLD-disease continuum (20–23), given their accessibility, regulatory potential, and causal links to disease processes. However, few studies have systematically investigated the mediating role of plasma proteins in translating MASLD-related genetic risk into downstream health outcomes.

To address these gaps, we integrated genetic data across hepatic and cardiometabolic traits to derive an MASLD-specific polygenic architecture. We then examined its systemic associations and identified candidate protein mediators using Mendelian randomization (MR) and observational validation. This framework represents a step toward resolving the biological complexity of MASLD and translating genetic insights into clinically actionable strategies.

RESULTS

Study design

The overall study design is shown in Fig. 1. We conducted a comprehensive integrative genomic analysis of MASLD using GWAS summary statistics from large-scale cohorts totaling nearly 3 million individuals of European ancestry. To construct an MASLD-specific genetic architecture, we implemented a two-step cross-trait modeling approach that jointly incorporated genetic signals from hepatic fat (MRI-PDFF) and major cardiometabolic risk factors (CMRFs) (4), including adiposity, blood lipids, glycemic markers, and blood pressure. This framework enabled the identification of pleiotropic loci and the prioritization of candidate genes and pathways potentially involved in MASLD pathogenesis. To evaluate the clinical relevance of the resulting MASLD polygenic risk score (PRS), we performed a phenome-wide association study (PheWAS) to systematically map its associations across a broad range of disease outcomes in the UK Biobank (UKB) cohort. To investigate potential mediating mechanisms, we further integrated proteomic data through a proteome-wide MR analysis, complemented by prospective cohort validation, to identify circulating proteins that may link MASLD genetic risk to multisystem disease burden.

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

¹Human Phenome Institute, Research and Innovation Center, Shanghai Pudong Hospital, Fudan University, Shanghai, China. ²Fudan University Taizhou Institute of Health Sciences, Taizhou, China. ³State Key Laboratory of Genetics and Development of Complex Phenotypes, Human Phenome Institute, Zhangjiang Fudan International Innovation Center, Fudan University, Shanghai, China. ⁴Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, Shanghai, China. ⁵Department of Epidemiology, School of Public Health, Fudan University, Shanghai, China. ⁶National Clinical Research Center for Aging and Medicine, Huashan Hospital, Fudan University, China.

*Corresponding author. Email: tjzhang@shmu.edu.cn (T.Z.); zhenqiu@fudan.edu.cn (Z.L.); xingdongchen@fudan.edu.cn (X.C.)

†These authors contributed equally to this work.

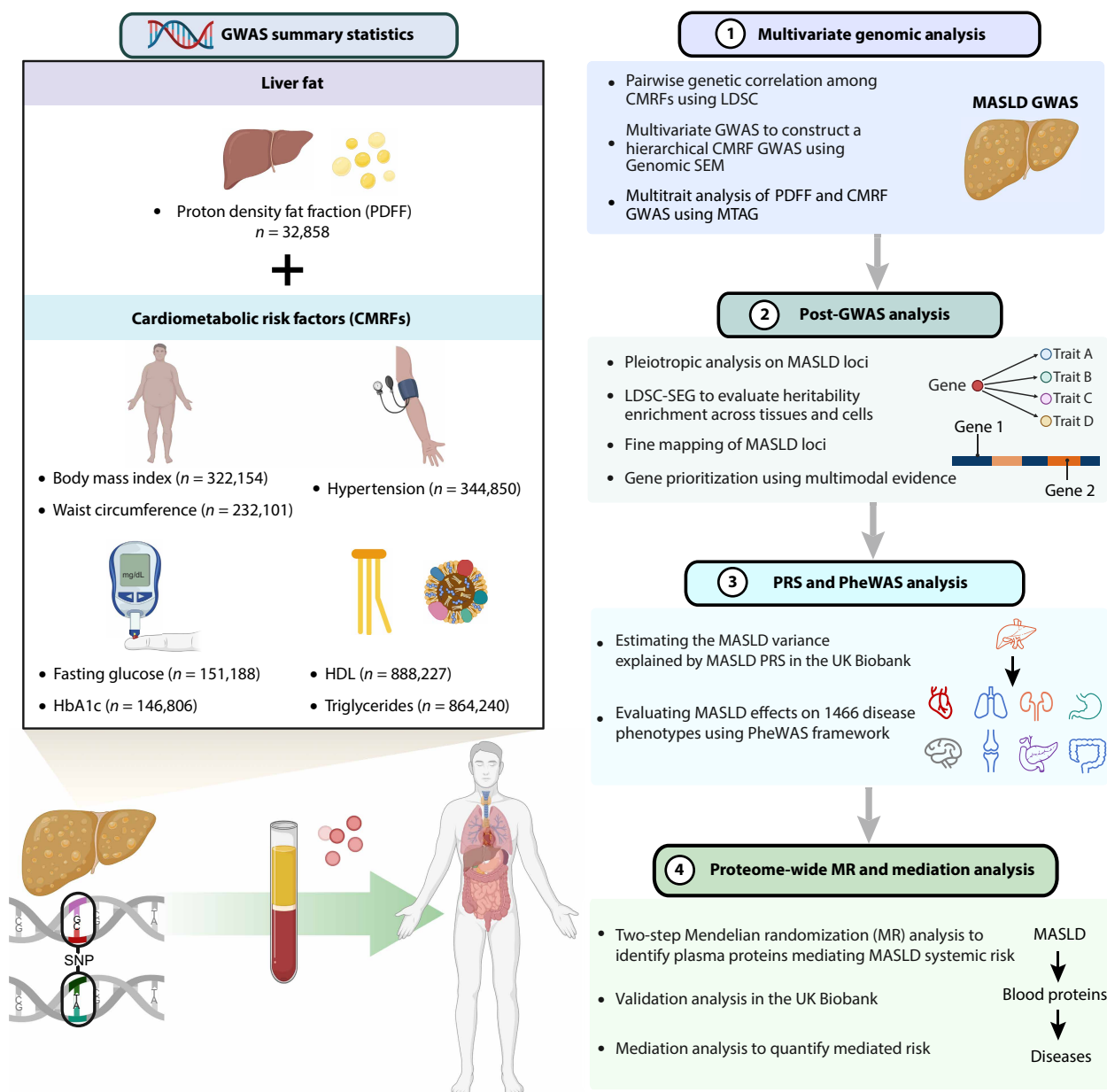


Fig. 1. Overview of study design. Schematic summary of the integrative framework used to dissect the genetic architecture and systemic consequences of MASLD, including multitrait genetic modeling, effector gene prioritization, PRS construction, PheWAS, and proteome-wide MR analysis. Created in BioRender [J. Du (2025); <https://BioRender.com/6x8e7o8>].

Structure equation modeling of cardiometabolic trait GWAS

To capture the shared genetic architecture underlying MASLD-related CMRFs, we first constructed a composite CMRF GWAS using genomic structural equation modeling (Genomic SEM). We compiled publicly available GWAS summary statistics of European ancestry for seven MASLD-related CMRFs: body mass index (BMI), waist circumference (WC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), glycosylated hemoglobin (HbA1c), fasting glucose (FG), and hypertension (HTN), encompassing approximately 2.95 million individuals (table S1). All datasets were drawn from nonoverlapping cohorts independent of the UKB to

minimize bias due to sample overlap. Standardized quality control procedures were applied to each GWAS (table S2).

Univariate linkage disequilibrium score regression (LDSC) analyses confirmed robust polygenicity across traits, with moderate single-nucleotide polymorphism (SNP)-based heritability estimates (h^2 range: 0.11 to 0.13) and minimal inflation from population stratification (table S3). Pairwise genetic correlations ranged from moderate to high, with the strongest observed between BMI and WC ($r_g = 0.89$) (Fig. 2A). To assess the suitability of the genetic correlation matrix for latent factor analysis, we conducted a Kaiser-Meyer-Olkin (KMO) test, which yielded a value of 0.683, indicating adequate factorability

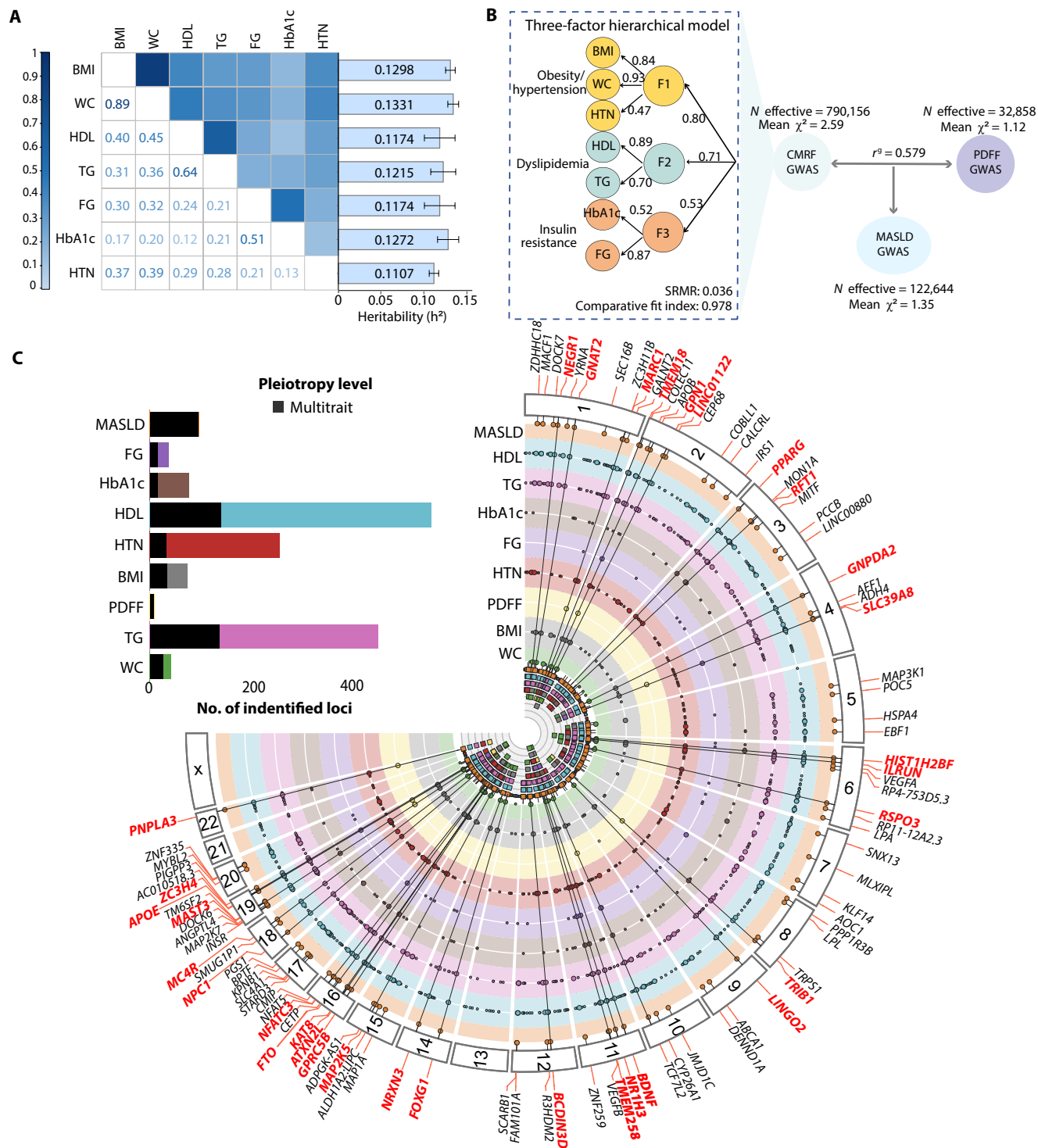


Fig. 2. Heritability and shared genetic architecture of MASLD-related traits. (A) Heatmap of pairwise genetic correlations (r_g) among seven MASLD-related CMRFs: BMI, WC, HDL-C, TG, FG, HbA1c, and HTN. Right panel displays SNP-based heritability (h^2) estimates and standard errors for each trait, estimated via LDSC. (B) Three-factor Genomic SEM model and multitrait analysis of GWAS (MTAG)-based multitrait GWAS of MASLD. Path diagram illustrates standardized loadings of a hierarchical model in which CMRFs load onto three latent factors. (C) Cross-trait pleiotropy landscape and circular Manhattan plot. Top left bar plot shows the number of genome-wide significant loci per trait, with black bars indicating loci shared across multiple traits and colored bars denoting trait-specific signals. Circular Manhattan plot depicts genome-wide associations for MASLD, PDFF, and seven CMRFs. Each concentric track represents one trait. MASLD loci shared by five or more traits are highlighted in red and linked by black lines, indicating hotspots of cross-trait pleiotropy.

(table S4). Parallel analysis based on eigenvalue distributions supported the extraction of three latent factors (fig. S1). Accordingly, we performed exploratory factor analysis (EFA) from odd-numbered chromosomes using promax rotation and evaluated one- to three-factor solutions. Three-factor model explained 69.3% of the total variance across the seven CMRFs (table S5), informing the subsequent confirmatory factor analysis (CFA) via Genomic SEM. A single-factor model from the CFA showed suboptimal fit [comparative fit index (CFI) = 0.900; standardized root mean square residual (SRMR) = 0.106], suggesting inadequate capture of genetic heterogeneity (table S6). In contrast, a hierarchical three-factor model provided better fit ($\chi^2 = 242.44$; CFI = 0.978; SRMR = 0.036), clustering traits into biologically coherent domains: F1 (obesity and HTN), F2 (dyslipidemia), and F3 (insulin resistance) (Fig. 2B and table S7).

We then constructed a composite CMRF GWAS using multivariate Genomic SEM, which estimates SNP effects on shared genetic liability across traits (figs. S2 and S3). The effective sample size (n_{eff}) of the CMRF was 790,156 with mean χ^2 of 2.59, and SNP-based heritability was 0.11 (Fig. 2B and table S8). The resulting CMRF GWAS demonstrated moderate to strong genetic correlations with individual component traits (median $r_g = 0.68$), supporting its validity as a genetically informative composite phenotype (fig. S4).

Multitrait analysis of GWAS-based genome-wide analysis of MASLD

We obtained GWAS summary statistics for liver PDFF in 32,858 individuals (24). Univariate LDSC analysis indicated moderate polygenicity, with a mean χ^2 of 1.12 and SNP-based heritability (h^2) of 0.17 (Fig. 2B and table S3). To construct a genetic model that captures the integrated genetic architecture of hepatic fat and cardiometabolic traits, we applied multitrait analysis of GWAS (MTAG) to jointly analyze the PDFF and composite CMRF GWAS. This cross-trait meta-analysis yielded a composite PDFF-based GWAS (hereafter referred to as the MASLD GWAS) with an increased effective sample size of 122,644 and boosted association strength (mean χ^2 increased from 1.12 to 1.35) while maintaining a maximum false discovery rate (FDR) of 0.18 (Fig. 2B, table S8, and figs. S5 and S6). The genetic architecture of MASLD was polygenic, evidenced by an elevated mean χ^2 value (1.35) and a genomic inflation factor ($\lambda_{\text{GC}} = 1.06$), with minimal confounding from population stratification (LDSC regression intercept < 1) (table S8).

We identified 128 genome-wide significant and independent variants associated with MASLD, mapping to 100 loci, markedly exceeding the number of signals (<20) reported in earlier NAFLD-related GWAS (Fig. 2C and table S9) (8, 9). Of them, 98 loci exhibited pleiotropic associations with at least two MASLD traits, and 34 loci exhibited high pleiotropy, each being associated with five or more MASLD traits (Fig. 2C and table S10). These highly pleiotropic regions included well-established MASLD susceptibility genes such as *PNPLA3*, *APOE*, *TRIB1*, and *GPAM* (8). Beyond the expected overlap with cardiometabolic traits, most lead variants from the MASLD GWAS also exhibited pleiotropic associations with additional biological traits, such as estimated glomerular filtration rate and circulating vitamin D levels, based on previously published GWAS curated in the National Human Genome Research Institute-European Bioinformatics Institute (NHGRI-EBI) GWAS Catalog (table S9).

Functional and cellular context of MASLD risk

To provide a holistic view of MASLD genetic architecture, we conducted integrative enrichment analyses across gene sets, regulatory

elements, and tissue/cell-type-specific expression profiles. Multi-marker Analysis of GenoMic Annotation (MAGMA)-based gene set enrichment analysis identified 37 pathways significantly associated with MASLD after Bonferroni correction (fig. S7 and tables S11 and S12), with top signals related to lipid localization, triglyceride metabolism, and lipoprotein remodeling, highlighting lipid dysregulation as a core feature of MASLD pathogenesis. Stratified LDSC revealed 13 categories with significant heritability enrichment, including conserved elements, human promoters, histone H3 lysine 9 acetylation, and super enhancers (fig. S8 and table S13), reflecting a complex regulatory landscape. Tissue/cell-type enrichment analysis via LDSC-specifically expressed genes (SEG) demonstrated significant enrichment in liver tissue and hepatocytes, as expected, as well as in visceral adipose tissue (omentum), supporting the role of adipose-liver interactions (fig. S9 and table S14). Notably, enrichment was also observed in brain regions such as the cerebral cortex, hippocampus, and limbic system, suggesting possible neuroendocrine involvement in MASLD-related systemic regulation (tables S14 and S15).

Effector gene prediction for MASLD

We prioritized effector genes for MASLD using FLAMES (Fine-mapped Locus Annotation and Effector gene Scoring) framework based on fine-mapped results of the identified 100 loci. Using SuSiEx, we fine mapped 88 credible sets across autosomes, encompassing 925 putative causal variants, including 29 with high posterior inclusion probability (PIP > 0.80) (Fig. 3A and table S16). Most fine-mapped variants resided in noncoding regulatory regions, particularly intronic (49.8%) and intergenic (27.8%) sites (Fig. 3B).

We mapped the 925 causal variants to 1777 candidate genes and calculated FLAMES scores for each using an integrative framework combining SNP-to-gene annotations and convergence-based evidence [e.g., polygenic priority scores (PoPS)] (tables S17 and S18). We then selected 88 genes with the highest FLAMES score within each fine-mapped region and prioritized 55 with a cumulative precision > 0.75 (Fig. 3C and table S19). Top-ranked genes included *NEGR1*, *FTO*, *BDNF*, *GNPDA2*, and *MC4R*, known regulators of adiposity (25–29), as well as *TRIB1*, *APOB*, *APOE*, *CETP*, and *LPL*, which are involved in lipoprotein metabolism (30–33). Canonical MASLD genes such as *PNPLA3*, *TM6SF2*, and *GPAM* were also recovered (9), supporting the robustness of the prioritization framework. In addition, we noted a few less-characterized candidate genes such as *MYO1F* and *FRMD5*, whose roles in MASLD pathogenesis remain largely unexplored and merit further functional investigation.

Tissue expression profiling based on Genotype-Tissue Expression (GTEx) data showed that prioritized genes were most strongly up-regulated in liver and visceral adipose tissue (omentum) (Fig. 3D, fig. S10, and tables S20 and S21), aligning with MASLD's central pathophysiology (34). Pathway enrichment analysis of prioritized genes using FUMA highlighted lipid homeostasis, sterol transport, and triglyceride metabolism as key biological processes (Fig. 3E and table S22). Enriched molecular functions included lipid binding and sterol transfer activity, while cellular components were dominated by terms related to lipoprotein particles and protein-lipid complexes. These findings underscore lipid metabolic regulation as a key axis linking genetic susceptibility to MASLD.

Systemic effects of MASLD genetic liability

To evaluate the clinical relevance of MASLD genetic risk, we constructed an MASLD PRS based on the MASLD GWAS summary

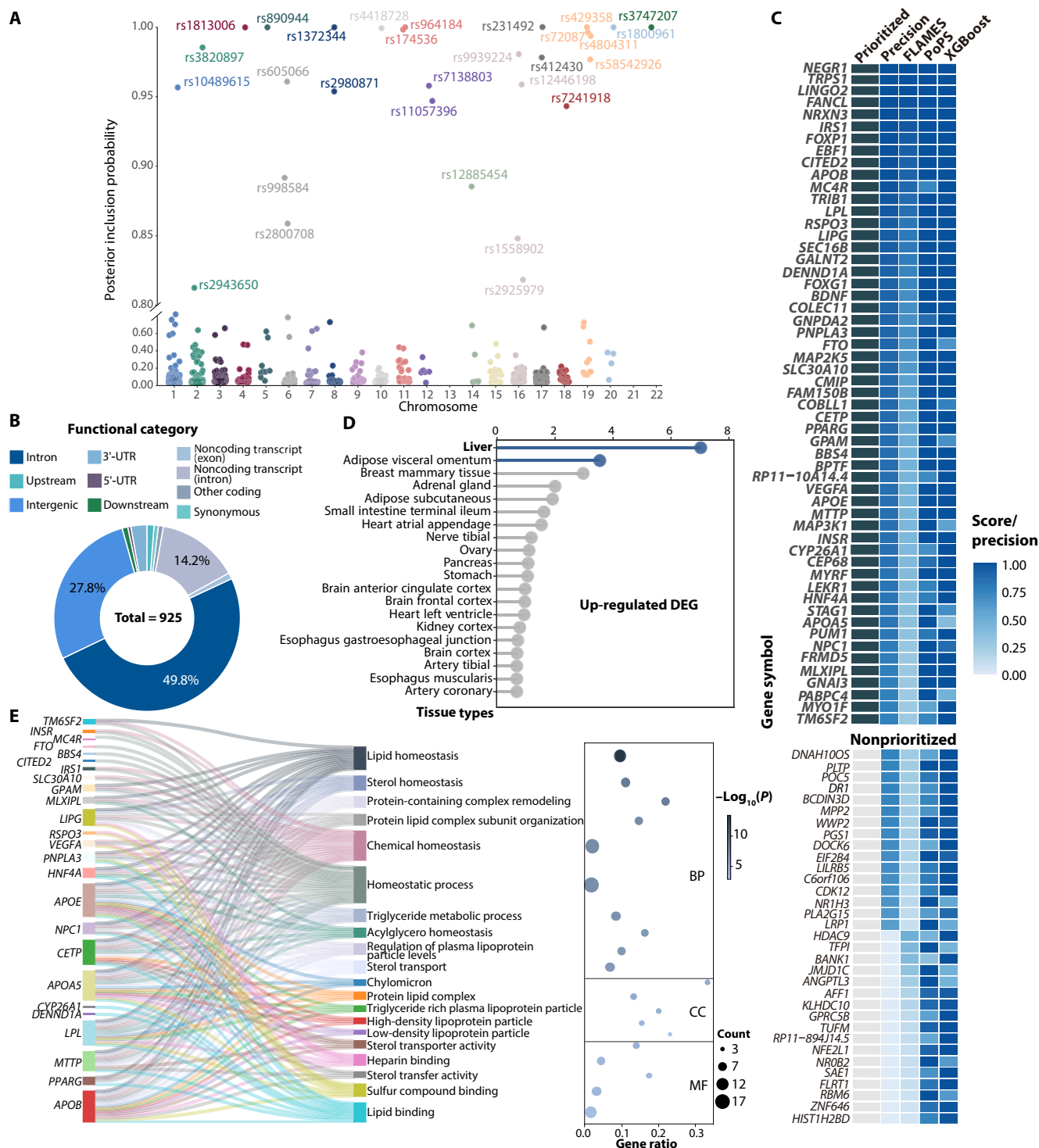


Fig. 3. Fine mapping and functional characterization of MASLD-associated loci. (A) Genomic distribution of fine-mapped SNPs across 22 autosomes, with PIP derived from SuSiEx. Variants with PIP > 0.8 are labeled. (B) Functional annotation of 925 credible set variants. (C) Multimodal effector gene prioritization integrating scores from XGBoost, PoPS, and FLAMES. Heatmap displays standardized scores across methods and cumulative precision. Genes with cumulative precision > 0.75 were retained, yielding 55 high-confidence candidates. (D) Tissue-specific expression enrichment of prioritized genes based on GTEx v8 transcriptomes. Differentially expressed gene (DEG) analysis revealed significant up-regulation in liver and visceral adipose tissue (omentum). (E) Gene Ontology (GO) enrichment analysis of prioritized genes. Left: Sankey diagram linking representative genes to top enriched GO terms spanning biological process (BP), cellular component (CC), and molecular function (MF) categories. Right: Bubble plot showing gene ratios and FDR-adjusted P values for selected enriched terms.

statistics and assessed its performance in the UKB cohort. For comparison, a PDF-based PRS was also derived from univariate PDF GWAS. MASLD cases were defined as individuals with a fatty liver index (FLI) ≥ 60 (fig. S11) (35, 36) plus ≥ 1 cardiometabolic trait, with exclusions for excessive alcohol use, chronic viral hepatitis, and Mendelian liver diseases (Materials and Methods). We assessed liver fibrosis risk in the MASLD cohort by comparing fibrosis-4 (FIB-4) index (37), the aspartate aminotransferase-to-platelet ratio index (APRI) (38), and the Forns score (39) between cases and controls, with the Forns index indicating a significantly higher fibrosis proportion among cases (3.4% versus 1.6%; fig. S12). Among 373,622 European-ancestry participants (51,754 MASLD cases), the MASLD PRS was robustly associated with MASLD risk in a dose-response manner [odds ratio (OR) per SD = 1.11, 95% confidence interval (CI): 1.10 to 1.12, $P < 2.05 \times 10^{-103}$] and explained an incremental 0.24% of variance (95% CI: 0.20 to 0.29%), outperforming the PDF PRS (Fig. 4, A to C, and tables S23 and S24).

To characterize the broader disease impact of MASLD genetic liability, we performed a PheWAS using the MASLD PRS. Significant associations were observed across multiple organ systems (Fig. 4D and table S25). Beyond expected positive associations with hepatic and cardiometabolic conditions (e.g., liver cirrhosis and angina pectoris), the MASLD PRS was also positively associated with a broad range of disorders, including digestive, renal, endocrine, musculoskeletal, and dermatological conditions, some of which have been reported in previous studies (40–42). Conversely, a negative association with dementia and Alzheimer's disease was observed, although this finding has been inconsistently reported in prior epidemiological studies (43, 44). We also identified associations not well characterized in the literatures, such as positive associations with iron deficiency anemias and erythematous dermatosis. In non-European participants, associations did not reach statistical significance for most diseases (table S26). These results underscore the multisystemic impact of MASLD genetic liability and highlight phenotypic domains that may warrant further mechanistic investigation.

Two-step proteome-wide MR analysis

To elucidate molecular mechanisms linking MASLD to multisystem diseases, we implemented a two-step proteome-wide MR framework (Fig. 5A). In step 1, we performed MR to assess the causal effects of MASLD on 4907 plasma proteins using data from the deCODE proteomic GWAS (45) ($n = 35,559$). Step 2 assessed the impact of MASLD-driven proteins on phenotypes identified in the MASLD PRS-PheWAS and had available GWAS summary data (table S27). We further validated these associations in the UKB cohort and quantified mediation effects of proteins in the MASLD-disease pathway.

The step 1 MR analysis revealed 514 proteins significantly influenced by MASLD (hereafter “MASLD-driven proteins”) based on inverse-variance weighted (IVW) method, with no evidence of weak instruments or horizontal pleiotropy (table S28 and fig. S13). Known MASLD-associated proteins such as apolipoprotein E (apoE), GCKR, and alcohol dehydrogenase family members were noted. In addition, proteins such as PHGDH, DCUN1D5, and CRYZL1 showed strong positive effects, while apolipoprotein M (apoM) and MENT were negatively associated with MASLD genetic liability.

In step 2, we used cis-protein quantitative trait loci (pQTLs) from a meta-curated proteomic resource (table S29) to assess the causal effects of MASLD-driven proteins on clinical outcomes. A total of 231 proteins with available cis-pQTLs and 58 diseases with

GWAS summary statistics was included, resulting in 13,398 protein-phenotype pairs tested by MR. After stringent filtering, 82 associations remained significant, of which 47 associations involving 25 unique proteins were retained on the basis of directional consistency between MR estimates and PheWAS results (fig. S14 and tables S30 to S32). Proteins such as FURIN, aldehyde dehydrogenase 2 (ALDH2), SERPINA1, and GSS were linked to multiple endpoints, including liver disease, type 2 diabetes (T2D), and cardiovascular conditions (Fig. 5B).

To validate these findings in a real-world population, we examined protein-disease associations in 52,633 UKB participants with baseline plasma protein data (46). Of the 25 MR-implicated proteins from the deCODE proteomic GWAS, 16 were measurable in UKB due to platform differences. Cox proportional hazard models, adjusted for demographic and lifestyle factors, confirmed 13 protein-disease pairs involving seven proteins [FURIN, ALDH2, RARRES1, lymphotoxin- α (LTA), apoM, thialysine N-epsilon-acetyltransferase (SAT2), and enoyl-CoA hydratase, mitochondrial (ECHS1)] (Fig. 5C and table S33). These proteins were linked to incident cardiometabolic and endocrine outcomes, including coronary atherosclerosis, angina, HTN, T2D, hypothyroidism, and hypoglycemia. Kaplan-Meier curves showed clear dose-response pattern for all associations, except for LTA–type 1 diabetes pair, with higher protein levels predicting greater disease incidence (Fig. 5D and table S34).

Mediation analysis quantified the proportion of the MASLD effect on disease risk explained by protein intermediates (Fig. 5E). FURIN (also known as PCSK3), an important mammalian proprotein convertase (47), emerged as a major mediator linking MASLD to cardiovascular diseases, with the mediated proportion ranging from 12.0% for angina pectoris to 50.6% for essential HTN. Other proteins also showed modest but significant mediating effects: ALDH2, apoM, RARRES1, and SAT2 mediated MASLD-associated risks of essential HTN, hypothyroidism, and T2D, with proportion mediated ranging from 2.2 to 10.0% (table S35). One pair (ECHS1–hypoglycemia) did not reach statistical significance despite consistent associations in MR and cohort analyses.

Genetic relationships between MASLD effector genes and plasma protein mediators

We next asked whether the 55 MASLD-prioritized genes exert regulatory effects on circulating levels of the 25 protein mediators. Using Bayesian colocalization between MASLD GWAS loci corresponding to specific effector gene and plasma-protein GWAS (± 250 -kb windows), we tested whether each gene-protein pair shared a causal variant. Across all pairs, 13 loci showed strong evidence of colocalization [posterior probability for hypothesis 4 (PP.H4) > 0.8] spanning multiple proteins (fig. S15 and table S36). Several canonical MASLD genes displayed broad cross-protein signals; for example, *APOE* colocalized with apoM, CETP, endothelial lipase (LIPG), and HSPA1B. Regional locus plots further highlighted pleiotropic colocalization signals at representative loci, including *APOE* (rs429358), *PNPLA3* (rs3747207), and *FTO* (rs62048402) (figs. S16 to S18), supporting shared genetic regulation between these genomic locus and plasma proteins.

To further delineate regulatory links between MASLD effector genes and circulating proteins, we performed two-sample MR using liver expression quantitative trait loci (eQTLs) as instruments and the 25 candidate proteins as outcomes. Among 21 MASLD-prioritized genes with valid instruments, 22 gene-protein pairs showed nominal

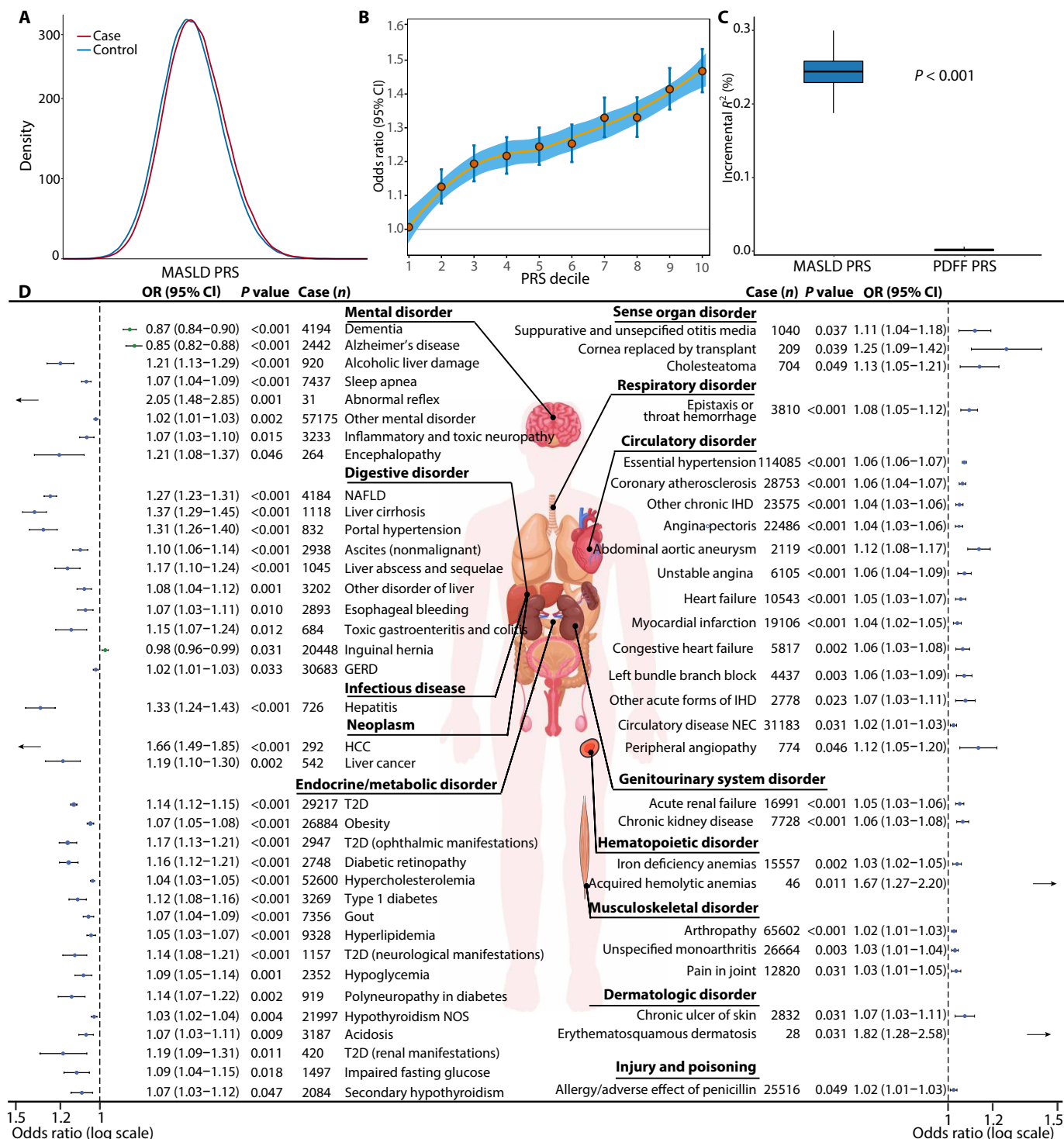


Fig. 4. Predictive performance and phenotype-wide associations of MASLD PRS in the UKB. (A) Distribution of MASLD-PRS among MASLD cases and controls in the UKB cohort. (B) Association between MASLD-PRS deciles and MASLD risk. Points show decile-specific estimates; the curve is a spline fit with a blue shaded band. (C) Comparison of incremental variance explained (ΔR^2) by MASLD-PRS and PDFF-PRS. Box represents the ΔR^2 from bootstrap iteration, with error bars indicating the 95% CIs derived from 1000 bootstrap replicates using the percentile method. (D) PheWAS of MASLD-PRS across 13 disease categories in individuals of European ancestry. Forest plot displays significant associations after FDR correction, with ORs, 95% CIs, case counts, and corresponding P values annotated for each phenotype. HCC, hepatocellular carcinoma; GERD, gastroesophageal reflux disease; IHD, ischemic heart disease; NEC, not elsewhere classified; NOS, not otherwise specified.

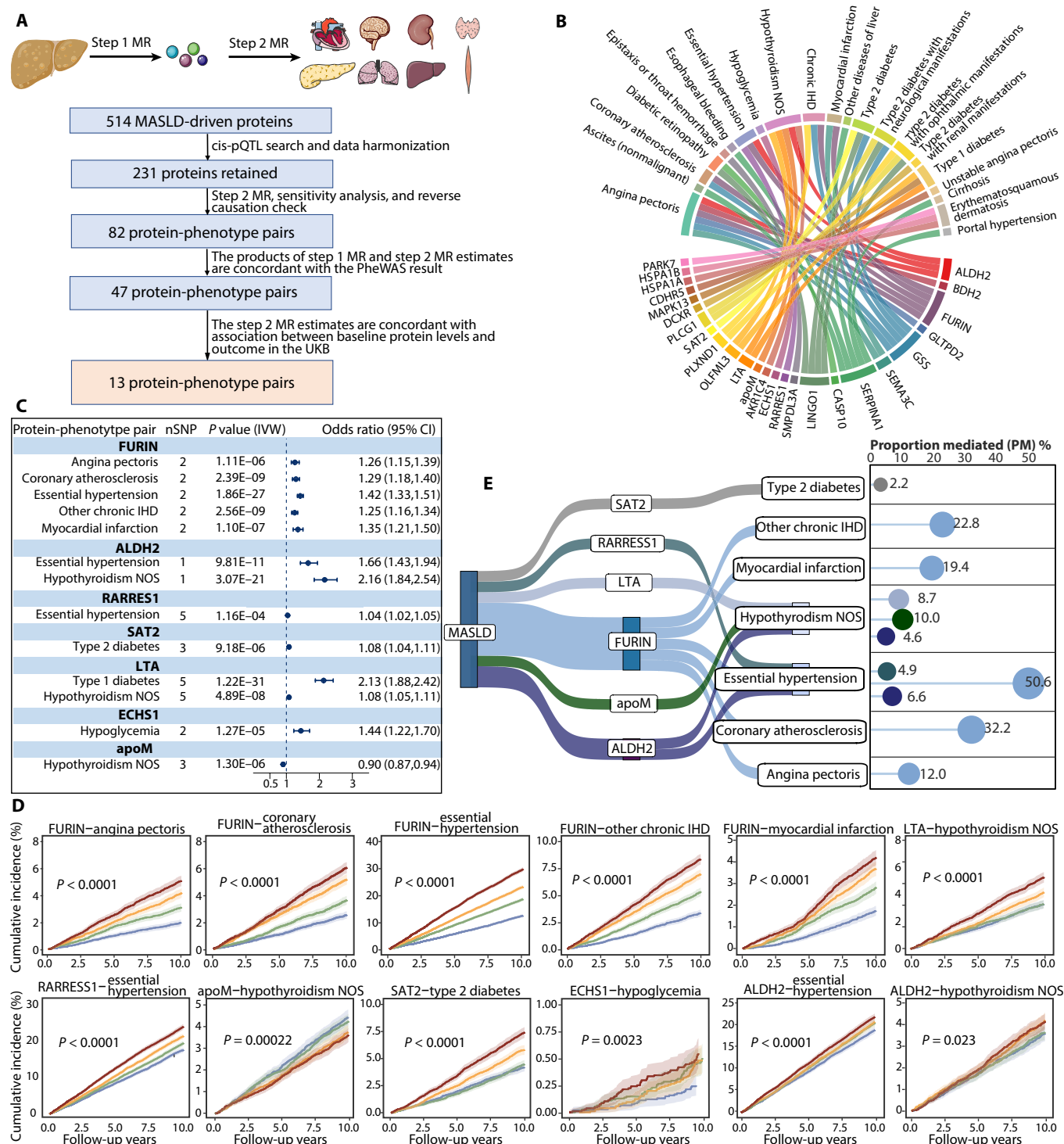


Fig. 5. Two-step MR framework and cohort validation identify plasma proteins mediating MASLD-related disease risk. (A) Overview of the two-step MR framework linking MASLD genetic liability to clinical outcomes via circulating proteins. (B) Circos plot illustrating 47 protein-phenotype pairs with concordant MR and PheWAS effects. (C) Summary of 13 high-confidence protein-phenotype pairs supported by both two-step MR and observational validation in UKB. (D) Cumulative risk curves stratified by baseline protein quartiles (Q1 to Q4) in the UKB participants. Curves represent cumulative incidence across quartiles; shaded areas denote 95% CIs. (E) Sankey diagram and scatter plot summarizing protein-mediated pathways from genetically predicted MASLD to multiple clinical outcomes in a two-step MR framework. The bubble denotes the proportion of MASLD-associated disease risk explained by each protein. nSNP, number of SNP.

causal effects, involving 10 distinct genes (table S37). Several genes (e.g., *CETP* and *COLEC11*) were linked to multiple protein targets, indicating potential pleiotropic regulation of plasma proteins. Notably, pairs such as *CETP*-apoM, *COLEC11*-BDH2, and *COLEC11*-SMPDL3A also exhibited strong colocalization (PPH4 > 0.80; fig. S15), providing convergent genetic support for shared regulatory architecture between MASLD loci and plasma protein abundance.

DISCUSSION

MASLD is a common and multisystem condition. Despite its growing burden, a comprehensive understanding of the genetic and molecular mechanisms underlying MASLD is still lacking. To address these challenges, we developed an integrative cross-trait genomic modeling framework to systematically dissect the polygenic basis and systemic impact of MASLD. This integrative framework yielded four key insights. First, we developed a high-powered MASLD GWAS by jointly modeling hepatic fat and cardiometabolic traits, identifying more than 100 risk loci and improving locus discovery beyond previous GWASs. Second, we revealed widespread pleiotropy at MASLD loci, implicating shared genetic underpinnings across hepatic, cardiovascular, and endocrine traits. Third, we validated an MASLD PRS that not only predicted MASLD risk but also captured its systemic disease burden across organ systems. Fourth, by integrating proteome-wide MR with longitudinal validation, we identified circulating proteins that mediate MASLD's impact on cardiometabolic and endocrine outcomes, providing proteomic insights into the MASLD systemic risk.

Our study addresses several persistent methodological challenges in MASLD genomic discovery. On the one hand, traditional MASLD GWASs, relying either on imaging-derived hepatic fat or clinical diagnostic codes (8, 9, 11), are limited by modest sample sizes, diagnostic heterogeneity, and potential misclassification, especially in capturing subclinical or metabolically driven cases. On the other hand, MASLD is not a single disease entity but a complex syndrome characterized by heterogeneous hepatic and extrahepatic manifestations (1). This syndromic nature challenges traditional, liver-centric definitions and highlights the limitations of prior studies in capturing its full systemic complexity. To overcome these limitations, we implemented a cross-trait genomic modeling strategy that jointly leverages genetically correlated hepatic and cardiometabolic traits to construct a refined phenotype framework. This approach enhances statistical power, mitigates misclassification bias, and better captures the multisystem complexity of MASLD (4). This strategy also enabled robust genetic discovery and functional downstream analyses (48, 49), including pleiotropy mapping and prioritization of causal genes and biological pathways, thereby deepening our understanding of MASLD as a multisystem disorder. By integrating hepatic fat with cardiometabolic traits, our approach redefines MASLD from a genomic perspective, offering a more biologically coherent disease model.

We uncovered three latent cardiometabolic factors, obesity-HTN, dyslipidemia, and insulin resistance, reflecting the interconnected systemic profile of MASLD (5). Many loci showed high pleiotropy—more than 30 influenced five or more MASLD-related traits. Key loci such as *PNPLA3*, *FTO*, *TRIB1*, *APOE*, and *GPAM* exhibited consistent effects across hepatic and extrahepatic domains, underscoring the polygenic and multisystem nature of MASLD. Through integrative fine mapping, we prioritized 55 effector genes enriched in pathways central to lipid metabolism, lipoprotein remodeling, and hepatic

function. Established MASLD genes (*PNPLA3*, *APOE*, *GPAM*, and *TM6SF2*) were recovered alongside upstream regulators of adiposity and insulin (*FTO*, *LPL*, *BDNF*, *NEGR1*, *MC4R*, *IRS1*, and *INSR*), reinforcing the role of adipose and glucose overflow in hepatic lipid accumulation. Expression enrichment in liver and adipose tissue supported MASLD's core pathogenic axis: cross-talk between hepatic steatosis and impaired adipose expandability (50). Some genes (e.g., *EBF1* and *COLEC11*) also participate in immune and inflammatory regulation (51, 52), suggesting shared pathways between metabolic and immune dysfunction in MASLD (53). These findings offer mechanistic insights into MASLD pathogenesis and underscore the need for future studies to dissect the coordinated roles of lipid metabolism, adipose and glucose dysfunction, and immune signaling in disease initiation and progression.

We demonstrated the translational relevance of MASLD genetics through genome-wide analyses. The MASLD PRS, derived from our cross-trait GWAS, was associated with a wide spectrum of clinical outcomes: hepatic (e.g., cirrhosis and liver cancer), cardiometabolic (e.g., myocardial infarction and T2D), renal (e.g., chronic kidney disease), and endocrine diseases (e.g., gout and hypothyroidism). Notably, the MASLD PRS outperformed a PheWAS-derived PRS in predicting MASLD status and associated comorbidities, highlighting the added value of integrating systemic components in genetic modeling. Associations with musculoskeletal, hematologic, and dermatologic conditions—some rarely reported previously—warrant further prospective cohort study validation and mechanistic investigation. A negative association with Alzheimer's disease was observed, echoing prior conflicting epidemiologic findings (44, 54). These findings not only reinforce the systemic nature of MASLD but also highlight the utility of cross-trait genetic approaches in uncovering unexpected disease links, offering potential avenues for mechanistic research and risk stratification in precision medicine.

To explore potential mechanisms linking MASLD to systemic disease, we implemented a two-step proteome-wide MR framework. We identified more than 500 proteins influenced by MASLD and tested their downstream effects on MASLD-related outcomes. After stringent filtering, we identified 47 high-confidence protein-phenotype pairs, of which 13 were validated in the UKB. Notably, *FURIN* emerged as a key mediator of cardiovascular risk, explaining up to 32.2% of MASLD's effect on coronary atherosclerosis, consistent with its known role in vascular remodeling (55, 56). Other validated proteins included *ALDH2* (essential HTN), *SAT2* (T2D), and apoM (hypothyroidism), offering mechanistic insights and potential biomarker targets. Most associations showed dose-response patterns in Kaplan-Meier curves, with higher protein levels predicting increased disease risk. Mediation analysis confirmed partial but significant roles for these proteins, reinforcing their clinical and mechanistic relevance. Our integrative colocalization and gene-protein MR analyses further delineate the regulatory landscape linking MASLD effector genes to circulating proteins. *APOE* exhibited pleiotropic colocalization with several lipoprotein-related proteins (apoM, *CETP*, and *LIPG*), whereas *PNPLA3* and *FTO* also showed shared causal signals with multiple proteins, consistent with their established functions in hepatic lipid metabolism and systemic metabolic regulation. Notably, most colocalized pairs were located on different chromosomes, suggesting trans-regulatory effects rather than local cis-mechanisms. Thus, colocalization likely reflects coordinated genetic regulation, complementing MR evidence of causal links between gene expression and protein abundance. This proteogenomic framework enhances our

understanding of MASLD's biological cascade, from genetic risk to circulating proteins to disease onset, bridging molecular intermediates and clinical endpoints. It also highlights the potential of plasma proteins as biomarkers and therapeutic targets, enabling protein-informed risk stratification and tailored intervention.

Several limitations should be acknowledged. First, our genetic analysis focused on common variants; rare or structural variants were not captured. Second, causal gene prioritization was based on statistical integration and requires experimental validation. Third, our analyses were largely restricted to individuals of European ancestry, potentially limiting generalizability. The diminished performance of PRSs trained in European cohorts when applied to non-European groups likely reflects differences in allele frequencies, linkage disequilibrium (LD) patterns, and GWAS sample sizes and may be further amplified by environmental and sociocultural factors (57, 58). Developing equitable, ancestry-diverse models will require expanding GWAS in underrepresented populations and conducting trans-ethnic meta-analyses so that MASLD PRSs can be deployed broadly without worsening health disparities. Fourth, plasma pQTLs may not fully capture disease-relevant regulation in specific tissues—notably liver and adipose, which are central to MASLD pathophysiology—underscoring the need for integrative analyses that combine tissue-resolved eQTLs, pQTLs, and single-cell multiomics to enhance biological specificity and interpretability. Last, although our results genetically nominate *FURIN* and other proteins as key mediators of MASLD-related systemic outcomes, we acknowledge that direct mechanistic validation is still essential to establish causality. Targeted perturbation studies (e.g., viral or nanoparticle-mediated modulation of gene expression and protein abundance) in relevant MASLD and cardiovascular models will be required to rigorously confirm these causal links and to fully assess their therapeutic potential. This mechanistic validation represents an essential next step, as genetic inference alone, while powerful, cannot fully resolve the specific biological context or the precise cellular and temporal processes involved. Thus, the analytical framework established here provides a solid foundation and crucial prioritization tool for subsequent experimental validation and translational development.

In conclusion, our study presents a comprehensive framework for dissecting MASLD as a genetically complex and multisystem disease. By leveraging cross-trait genomic modeling integrated with proteomic and phenomic data, we enhance locus discovery, uncover systemic disease pathways, and identify circulating mediators with translational relevance. These findings support a shift from liver-centric models toward holistic, precision-guided strategies for MASLD prevention and treatment, addressing the full spectrum of metabolic dysfunction. Our work establishes a foundational resource for the field, demonstrating how integrative genomics can decode the biological complexity of MASLD and guide future efforts in mechanistic research, biomarker discovery, and therapeutic development.

MATERIALS AND METHODS

GWAS population

We investigated eight traits relevant to MASLD, including (i) liver PDFF and (ii) seven MASLD-related CMRFs: BMI, WC, HbA1c, FG, TG, HDL-C, and HTN. PDFF is the current gold-standard imaging biomarker for hepatic fat quantification, reflecting intrahepatic triglyceride accumulation (59). The seven CMRFs were selected on the basis of established MASLD diagnostic criteria (4). GWAS summary

statistics for liver PDFF were obtained from the UKB (24), while those for the seven MASLD-related CMRFs were obtained from the largest available European-ancestry studies that did not include UKB participants.

GWAS quality control

We performed stringent quality control on GWAS summary statistics using EasyQC (v.23.8) and MungeSumstats (v.1.14.1), with the 1000 Genomes Project Phase 3 European panel as the reference (60). SNPs were excluded if they were (i) monomorphic or had minor allele frequency (MAF) < 0.5%; (ii) missing, invalid, or had ambiguous alleles; (iii) poorly imputed (INFO < 0.9, if available); (iv) duplicated or multiallelic; or (v) absent from or inconsistent with the reference panel. All datasets were harmonized to the GRCh37 (hg19) genome build.

Genetic correlation and SNP-based heritability

We applied LDSC (61) to evaluate genomic inflation and estimate SNP-based heritability. Intercepts near 1 and low attenuation ratios indicated that observed inflation was due to polygenicity rather than population stratification. Pairwise genetic correlations were computed across MASLD-related traits and between the composite CMRF construct and individual traits, validating the latent phenotype. Notably, the HDL-C GWAS was reverse coded to align risk directions. Analyses used precomputed LD scores from ~1.3 million high-quality HapMap3 SNPs in the European 1000 Genomes panel. SNPs were filtered for MAF > 0.01 and INFO > 0.9. For binary traits, total sample size ($N_{\text{cases}} + N_{\text{controls}}$) was used.

KMO test and parallel analysis

We used the KMO test to assess sampling adequacy for factor analysis; values > 0.5 were considered acceptable (48). Analyses were performed using the psych R package (v.2.4.12). We conducted parallel analysis (62) to determine the optimal number of factors to retain for subsequent EFA. Observed eigenvalues from the genetic correlation matrix were compared against those from Monte Carlo-simulated noise matrices generated using the sampling covariance matrix. Both matrices were derived from the LDSC function in Genomic SEM (v.0.0.5) (63) based on GWAS summary statistics for the seven MASLD traits. The number of retained factors was determined where observed eigenvalues exceeded simulated ones.

EFA and CFA

EFA was conducted using the factanal function (R stats v.4.0.5) on traits from odd-numbered chromosomes. Guided by parallel analysis results, we fitted models with one to three latent factors, applying promax rotation to account for factor correlations. Factors explaining ≥15% of the total variance were retained, following prior studies (48, 49). CFA was then performed using Genomic SEM on even chromosomes to validate EFA-derived structure. Indicators with loadings ≥ 0.3 were retained. CFA models were specified under unit variance identification and estimated using diagonally weighted least squares. Model fit was evaluated using standard indices: χ^2 , CFI (with ≥0.95 indicating good fit), akaike information criterion, and SRMR (with <0.08 indicating good fit) (64).

Multivariate Genomic SEM for CMRF GWAS

We applied Genomic SEM to jointly model seven MASLD-related CMRF traits. Stage 1 used cross-trait LDSC to estimate the genetic

covariance matrix from univariate GWASs. Stage 2 incorporated SNP associations into a structural equation model to estimate a multivariate GWAS, reflecting shared genetic architecture across traits. A hierarchical factor model, scaled using unit loading identification, demonstrated superior fit and was selected for downstream analyses. The final model included 2,202,333 SNPs with MAF > 1%. Effective sample size (N_{eff}) was calculated using the method of Mallard *et al.* (65).

Construction of MASLD GWAS using MTAG

To enhance power for MASLD locus discovery, we applied MTAG (66) to integrate the liver PDFF GWAS with the CMRF latent trait GWAS. MTAG models the shared variance-covariance structure of SNP effects across traits to increase statistical power and reduce mean squared error. Harmonization steps included aligning alleles, excluding multiallelic variants, and retaining palindromic SNPs after strand unification to GRCh37. The resulting MASLD GWAS retained the meta-analytic format of the PDFF GWAS and showed modest inflation in false positives (MaxFDR reported). The effective size was calculated on the basis of the increase in the mean χ^2 statistic using the formula

$$N_{\text{eff(GWAS-MTAG)}} = N_{\text{eff(GWAS-PDFF)}} \times \left(\frac{\chi^2_{\text{GWAS-MTAG}} - 1}{\chi^2_{\text{GWAS-PDFF}} - 1} \right)$$

Independent significant lead SNPs were identified using PLINK 1.9 (67) (--clump) with the European 1000 Genomes Phase 3 reference panel ($P < 5 \times 10^{-8}$, LD $r^2 < 0.001$, 100-kb window). Nearby lead SNPs (<250 kb) were merged into a single locus, represented by the variant with the lowest P value. Loci were annotated using FUMA (v.1.8.0) (68) and cross referenced with the NHGRI-EBI GWAS Catalog (69). The human leukocyte antigen region was excluded because of its complex LD structure.

Pleiotropy analysis

To identify pleiotropic loci, we evaluated overlap between MASLD-associated lead SNPs and those from PDFF or CMRF GWAS. MASLD loci within 1 Mb of any PDFF or CMRF lead SNP were considered shared and annotated with a common locus ID. Fuji plot [v.1.0.3; Kanai *et al.* (70)] was used to visualize trait-specific and pleiotropic loci. For pleiotropic regions, candidate effector genes were nominated using integrative evidence from Open Targets Genetics (71) and FUMA annotations; otherwise, the nearest protein-coding gene was reported.

Gene-based and gene set analyses

Gene-based and gene set analyses were performed using MAGMA (72), as implemented in FUMA. Gene-based tests aggregated SNP associations to 18,894 protein-coding genes using an SNP-wise mean model, accounting for LD structure. Gene set enrichment tested 17,023 curated and Gene Ontology (GO) gene sets from the Molecular Signatures Database (MSigDB; v.2023.1) (73), spanning biological processes, molecular functions, and cellular components. Enrichment was assessed via competitive testing with Bonferroni correction.

Functional, cell- and tissue-type enrichments

To characterize functional relevance of MASLD-associated loci, we applied stratified LDSC (S-LDSC) (74) to partition SNP-based heritability across 84 genomic annotations. Enrichment was defined as the proportion of heritability relative to the proportion of SNPs in

each annotation, with significance threshold $P < 5.95 \times 10^{-4}$ (0.05/84). Tissue- and cell-type-specific enrichments were assessed using LDSC-SEG (75, 76) across 205 expression datasets from GTEx and Franke lab (77) and 489 epigenomic annotations from Roadmap Epigenomics (78) and the Encyclopedia of DNA Elements project (79). All analyses used precomputed European LD scores from the 1000 Genomes Phase 3 reference.

Gene prioritization with FLAMES

To systematically prioritize effector genes underlying MASLD-associated genetic loci, we applied FLAMES (80), a supervised machine-learning framework that integrates locus-specific functional SNP-to-gene evidence and genome-wide gene convergence signals. We assumed that each true SNP-phenotype association reflects a single causal gene through which the SNP exerts its effect on MASLD. This gene is referred to as the effector gene of the corresponding SNP. FLAMES integrates two evidence streams: (i) SNP-to-gene functional annotations (e.g., eQTLs and enhancers) and (ii) PoPS (81), reflecting genome-wide convergence of gene-level functional features.

Fine mapping of causal variants

Significant MASLD loci were fine mapped using SuSiEx (82) with European UKB data as reference. We defined 500-kb windows centered on lead SNPs and assumed one causal variant per region to reduce false positive rate (80). For each locus, 95% credible sets were constructed on the basis of PIPs.

Functional annotations of candidate effector genes

For each credible set, SNPs were linked to candidate genes using FLAMES annotate, which incorporates both SNP-specific (e.g., eQTLs) and regional (e.g., enhancer-gene links) locus-to-gene annotations. These locus-based gene annotations were used as inputs for XGBoost classifier to generate locus-based SNP-to-gene annotation scores, reflecting the strength of combined functional evidence linking the gene to the credible set. Locus-based annotation scores were computed by multiplying the PIP of each SNP by the corresponding gene annotation score. For regional annotations, the score was calculated as the product of the SNP's PIP and the annotation strength for the region-gene pair. The window of genes annotated to a locus is specified as with default 750 kb of the fine-mapped lead SNP.

PoPS integration and scoring to identify effector genes

Gene-level PoPS were computed on the basis of MASLD MAGMA z -scores using a weighted model of gene expression, coexpression, pathways, and protein-protein interactions. FLAMES combined locus-based scores and PoPS via XGBoost to generate a unified raw score per gene, linearly scaled within each locus. To evaluate confidence, FLAMES estimates cumulative precision via benchmarking against curated gene-trait pairs (e.g., exome-wide association study-validated loci). Specifically, a series of thresholds was evaluated on held-out datasets to identify the combination of raw and scaled score cutoffs that maximized recall while maintaining a minimum of 75% precision. A polynomial regression was then fitted to model the relationship between the scaled FLAMES score and cumulative precision, enabling FLAMES to report precision estimates for all gene predictions.

Pathway and tissue specificity of effector genes

We performed gene set enrichment analysis using the GENE2FUNC module in FUMA. Hypergeometric tests evaluated overrepresentation in curated gene sets from MSigDB (73), WikiPathways (83),

and the GWAS Catalog (69), spanning canonical pathways and GO terms.

Tissue-specific expression was evaluated using GTEx v8 transcriptomic data (84). Prioritized genes were tested for enrichment in tissue-specific differentially expressed gene (DEG) sets, precomputed via pairwise *t* tests (Bonferroni-adjusted $P \leq 0.05$ and $|\log_2FC| \geq 0.58$). Up-regulated, down-regulated, and two-sided DEG sets were tested separately using hypergeometric tests, with Bonferroni correction applied. Background genes were restricted to those with average expression > 1 transcripts per million in at least one tissue. We visualized the tissue-wide expression patterns of prioritized genes using a heatmap displaying $\log_2$ -transformed average expression values across tissue labels.

UKB cohort

UKB is a large-scale, prospective cohort comprising more than 500,000 participants aged 40 to 69 years, recruited from 22 centers across the United Kingdom between 2006 and 2010 (85). Genotyping was performed using either the Applied Biosystems UK BiLEVE Axiom Array or the closely related UKB Axiom Array (86). Following stringent quality control, genotype data were imputed using the Haplotype Reference Consortium, UK10K, and 1000 Genomes Phase 3 reference panels. We excluded participants based on the following criteria: (i) low genotype quality or missing data, (ii) discordant genetic and reported sex, (iii) non-white ethnicity, (iv) relatedness, and (v) participation in the UKB imaging substudy used in the liver PDF GWAS, to avoid overlap with the discovery cohort. After exclusions, 373,835 participants remained for downstream PRS analyses. The UKB has approval from the North West Multi-Centre Research Ethics Committee (ref: 11/NW/0382) and obtained written informed consent from all participants before the study.

MASLD case ascertainment

As the MASLD phenotype is not directly available in the UKB, we defined MASLD status based on a multisociety Delphi consensus statement (4). We excluded individuals reporting alcohol intake > 20 g/day (women) or > 30 g/day (men) (4). Daily alcohol intake (in grams of pure alcohol) was calculated on the basis of baseline questionnaire responses, following the approach used in our previous study (87). We also excluded individuals with chronic viral hepatitis (ICD-10 B18.0 to B18.2 and B18.8 to B18.9) based on hospital records, as well as those with Mendelian conditions including Wilson disease, hereditary fructose intolerance, familial partial lipodystrophy, glycogen storage disorders, abetalipoproteinemia, and α_1 -antitrypsin deficiency. Hepatic steatosis was assessed using the FLI (36), a widely validated noninvasive measure (88). An FLI score ≥ 60 was used as the threshold for likely steatosis. MASLD was defined by the presence of (i) hepatic steatosis, (ii) ≥ 1 CMRF, and (iii) nonexcessive alcohol consumption. On the basis of these criteria, a total of 51,754 individuals in the UKB cohort was identified as MASLD cases.

To characterize fibrosis risk in MASLD, we computed three widely used noninvasive scores: the FIB-4 (37), APRI (38), and the Forns score (39). Participants with incomplete laboratory data were excluded. To limit outlier influence, biomarker values > 3 SD from the mean were winsorized to that threshold rather than removed. Advanced-fibrosis risk was classified as low, intermediate, or high using established cutoffs: FIB-4 (1.30, 2.67), APRI (0.5, 1.5), and Forns (4.2, 6.9) (88).

PRS analysis

We constructed an MASLD PRS using the clumping and thresholding (C + T) approach (89). Independent genome-wide significant SNPs were first selected from the MASLD GWAS through LD-based clumping. The PRS was calculated as the weighted sum of risk alleles, with weights derived from GWAS effect sizes. Association between PRS and MASLD status was assessed by logistic regression across PRS deciles, using the lowest decile as reference. All models adjusted for age, sex, and the top 10 genetic principal components. To evaluate predictive performance, we compared two logistic regression models: a baseline model including age, sex, and principal components and a full model including PRS. Incremental coefficient of determination (R^2) was computed as the difference in Nagelkerke's pseudo- R^2 between the two models, with 95% CIs estimated via 1000 bootstrap replicates.

PheWAS of MASLD

We conducted a PheWAS to assess the associations between the MASLD PRS and a wide range of clinical outcomes in the European-ancestry UKB cohort ($n = 373,835$). Health outcomes were derived from inpatient hospital records, death registries, and primary care data, harmonized into phecodes. To ensure sufficient statistical power, we excluded outcomes with fewer than 20 cases, resulting in 1466 phenotypes included in the primary analysis. For each phenotype, we fit logistic regression models with the MASLD PRS as the independent variable, adjusting for age, sex, and the top 10 genetic principal components. To account for multiple testing, we applied the Benjamini-Hochberg FDR procedure, with an FDR-adjusted $P < 0.05$ considered statistically significant. To evaluate the generalizability of findings across ancestries, we additionally performed PRS-PheWAS in the non-white UKB cohort ($n = 78,248$), encompassing 1250 phenotypes.

Two-step proteomic-wide MR analyses

Step 1: MASLD as exposure, plasma proteins as outcomes

To identify circulating proteins potentially regulated by MASLD, we conducted two-sample MR using MASLD as the exposure and plasma protein levels as outcomes. Genetic instruments were derived from our MASLD GWAS, selecting genome-wide significant and independent SNPs. Protein GWAS summary statistics were obtained from the deCODE study, which measured 4907 plasma proteins in 35,559 individuals of European ancestry using the SomaScan v4 platform (SomaLogic) (45). Proteins were defined as MASLD driven if they met all the following criteria: (i) IVW MR $P < 1.02 \times 10^{-5}$ (Bonferroni correction for 4907 tests); (ii) consistent effect direction across IVW, MR-Egger, weighted median, and weighted mode methods; (iii) all instruments had F-statistics > 10 ; and (iv) no evidence of directional horizontal pleiotropy (MR-Egger intercept $P > 0.05$). Applying these stringent filters, we identified 514 proteins as putative downstream consequences of MASLD.

Step 2: Plasma proteins as exposure, disease phenotypes as outcomes

To assess the downstream effects of MASLD-driven proteins, we conducted a forward two-sample MR analysis linking plasma proteins to MASLD-associated clinical outcomes. We first curated a high-quality cis-pQTL database by harmonizing publicly available summary statistics from seven large-scale proteogenomic studies (45, 46, 90–95). For each study, we selected cis-pQTLs (within ± 1 Mb of the transcription start site) with the largest F-statistics to ensure strong instruments and reduce weak instrument bias.

To exclude proteins that might causally influence MASLD (i.e., reverse causation), we first performed reverse MR using cis-pQTLs as instruments and MASLD as the outcome. Proteins with significant evidence of reverse causality (IVW or Wald ratio $P < 2.22 \times 10^{-4}$ after Bonferroni correction) were excluded ($n = 4$). Subsequently, we conducted a forward two-sample MR analysis using cis-pQTLs for the remaining 227 proteins as exposures and 58 MASLD-associated clinical outcomes (identified via PheWAS) as outcomes. Outcome GWAS summary statistics were sourced from either VA Million Veteran Program (MVP) (96) and FinnGen R10 (97), where available. For each protein-phenotype pair, IVW (or Wald ratio if only one SNP) was used as the primary estimator, and Bonferroni correction was applied on the basis of the number of proteins tested per outcome cohort.

To ensure robustness, we retained only results with consistent effect directions across all MR methods (IVW, MR-Egger, weighted median, and weighted mode). Instrumental variables associated with multiple proteins in *cis* pattern were excluded to minimize pleiotropic bias. For models with ≥ 3 instruments, horizontal pleiotropy was assessed using the MR-Egger intercept test. In total, 82 protein-phenotype pairs passed all criteria for statistical significance, robustness, and pleiotropy control. The MR analyses were performed using TwoSampleMR (v.0.6.10) R package.

Directional concordance check

To further validate the biological relevance of the 82 significant protein-phenotype associations, we assessed directional concordance across the MASLD-protein-phenotype axis. Specifically, we evaluated whether (i) the effect of MASLD liability on protein levels (from step 1 MR) and (ii) the effect of protein levels on clinical outcomes (from step 2 MR) jointly predicted the direction of the MASLD-phenotype association observed in the PheWAS.

Directional concordance was defined as cases where the product of the two MR-derived effect estimates aligned with the direction of the corresponding PRS-phenotype association. This consistency supports a coherent biological cascade in which MASLD alters protein expression, subsequently contributing to disease development. A total of 47 protein-phenotype pairs (involving 25 unique proteins) satisfied this criterion, indicating potential mechanistic relevance.

Validation in the UKB cohort

To evaluate the clinical relevance of MR-identified protein-phenotype associations, we analyzed 52,633 UKB participants with baseline plasma protein profiles measured using the Olink Explore 3072 platform. Among the 47 genetically predicted protein-phenotype pairs, 27 (involving 16 unique proteins) were available for validation. Incident disease outcomes were defined using ICD-10 codes; prevalent cases and participants with missing event dates were excluded.

We applied multivariable Cox proportional hazard models, adjusting for age, sex, ethnicity, household income, smoking, alcohol intake, BMI, and the top 10 genetic principal components. Associations were considered validated if the hazard ratio was statistically significant ($P < 0.05$) and directionally consistent with step 2 MR estimates. Thirteen protein-phenotype associations were validated, involving nine outcomes and seven unique proteins. Kaplan-Meier survival curves were plotted by quartiles of baseline protein levels, and log-rank tests were applied to assess statistical significance.

Mediation analysis

To quantify the extent to which these proteins mediate the relationship between MASLD and clinical outcomes, we performed mediation analysis for the 13 validated pairs using a product-of-coefficients approach implemented in the mediation R package (v.4.5.0). Linear regression was used for the MASLD-protein association and logistic regression for protein-outcome associations, adjusting for the same covariates. The indirect effects and proportion mediated were estimated using 1000 bootstrap samples.

Integrative colocalization and MR analysis for gene-protein pairs

To determine whether MASLD-prioritized loci causally influence circulating proteins identified by proteome-wide MR, we performed an analysis that jointly applied colocalization and MR. We first evaluated whether MASLD-associated loci and the 25 candidate proteins from proteome-wide MR are driven by the same causal variant. Using the coloc R package (v.5.2.3), which estimates the posterior probability that two traits share an underlying genetic signal, we analyzed ± 250 -kb windows centered on MASLD lead SNPs with default priors ($p_1 = 1 \times 10^{-4}$ and $p_2 = 1 \times 10^{-4}$ for trait associations; $p_{12} = 1 \times 10^{-5}$ for a shared association). A $PPH_4 > 0.80$ was considered strong evidence of colocalization. For such loci, we identified MASLD signals colocalizing with multiple proteins and depicted regional patterns using LocusZoomr (v.0.3.8).

We next evaluated regulatory effects of MASLD-prioritized gene expression on candidate plasma proteins using two-sample MR, treating liver eQTLs as exposures and plasma-protein GWAS as outcomes. Liver eQTL summary data were obtained from Broadway *et al.* ($n = 1183$) (98). We derived independent instruments per gene ($LD r^2 < 0.01$; association $P < 1 \times 10^{-5}$; F-statistic > 10) and estimated causal effects of gene expression on circulating abundance for the 25 proteins nominated by proteome-wide MR. Associations were deemed nominally significant at $P < 0.05$ in light of the moderate eQTL sample size.

Supplementary Materials

The PDF file includes:

Figs. S1 to S18

Legends for tables S1 to S37

References

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S37

REFERENCES

1. G. Targher, C. D. Byrne, H. Tilg, MASLD: A systemic metabolic disorder with cardiovascular and malignant complications. *Gut* **73**, 691–702 (2024).
2. Z. Liu, J. Huang, L. Dai, H. Yuan, Y. Jiang, C. Suo, L. Jin, T. Zhang, X. Chen, Steatotic liver disease prevalence in China: A population-based study and meta-analysis of 17.4 million individuals. *Aliment. Pharmacol. Ther.* **61**, 1110–1122 (2025).
3. B. P. Lee, J. L. Dodge, N. A. Terrault, National prevalence estimates for steatotic liver disease and subclassifications using consensus nomenclature. *Hepatology* **79**, 666–673 (2024).
4. M. E. Rinella, J. V. Lazarus, V. Ratzliff, S. M. Francque, A. J. Sanyal, F. Kanwal, D. Romero, M. F. Abdelmalek, Q. M. Anstee, J. P. Arab, M. Arrese, R. Bataller, U. Beuers, J. Boursier, E. Bugianesi, C. D. Byrne, G. E. Castro Narro, A. Chowdhury, H. Cortez-Pinto, D. R. Cryer, K. Cusi, M. El-Kassas, S. Klein, W. Eskridge, J. Fan, S. Gawrieh, C. D. Guy, S. A. Harrison, S. U. Kim, B. G. Koot, M. Korenjak, K. V. Kowdley, F. Lacaille, R. Loomba, R. Mitchell-Thain, T. R. Morgan, E. E. Powell, M. Roden, M. Romero-Gómez, M. Silva, S. P. Singh, S. C. Sookoian, C. W. Spearman, D. Tiniakos, L. Valenti, M. B. Vos, V. W. Wong, S. Xanthakos,

- Y. Yilmaz, Z. Younossi, A. Hobbs, M. Villota-Rivas, P. N. Newsome, A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *Hepatology* **78**, 1966–1986 (2023).
5. C. D. Byrne, A. Armandi, V. Pellegrinelli, A. Vidal-Puig, E. Bugianesi, Metabolic dysfunction-associated steatotic liver disease: A condition of heterogeneous metabolic risk factors, mechanisms and comorbidities requiring holistic treatment. *Nat. Rev. Gastroenterol. Hepatol.* **22**, 314–328 (2025).
6. L. V. C. Valenti, V. Moretti, Implications of the evolving knowledge of the genetic architecture of MASLD. *Nat. Rev. Gastroenterol. Hepatol.* **21**, 5–6 (2024).
7. S. Sookoian, Y. Rotman, L. Valenti, Genetics of metabolic dysfunction-associated steatotic liver disease: The state of the art update. *Clin. Gastroenterol. Hepatol.* **22**, 2177–2187.e3 (2024).
8. G. Sveinbjornsson, M. O. Ulfarsson, R. B. Thorolfsson, B. A. Jonsson, E. Einarsson, G. Gunnlaugsson, S. Rognvaldsson, D. O. Arnar, M. Baldvinsson, R. G. Bjarnason, T. Eiriksdottir, C. Eirikstrup, E. Ferkingstad, H. Helgason, A. Helgadóttir, L. Hindhede, G. Hjorleifsson, D. Jones, K. U. Knowlton, S. H. Lund, P. Melsted, K. Norland, I. Olafsson, S. Olafsson, G. R. Oskarsson, S. R. Ostrowski, O. B. Pedersen, A. S. Snaebjarnarson, E. Sigurdsson, V. Steinthorsdottir, M. Schwinn, G. Thorgeirsson, G. Thorleifsson, I. Jonsdottir, H. Bundgaard, G. R. Nashko, V. Gudnason, I. C. Rulifson, T. Rafnar, G. L. Norddahl, U. Thorsteinsdottir, P. Sulem, D. F. Gudbjartsson, H. Holm, K. Stefansson, Multiomics study of nonalcoholic fatty liver disease. *Nat. Genet.* **54**, 1652–1663 (2022).
9. Y. Chen, X. Du, A. Kuppa, M. F. Feitosa, L. F. Bielak, J. R. O'Connell, S. K. Musani, X. Guo, B. Kahali, V. L. Chen, A. V. Smith, K. A. Ryan, G. Eirisdottir, M. A. Allison, D. W. Bowden, M. J. Budoff, J. J. Carr, Y. I. Chen, K. D. Taylor, A. Oliveri, A. Correa, B. F. Crudup, S. L. R. Kardia, T. H. Mosley Jr., J. M. Norris, J. G. Terry, J. I. Rotter, L. E. Wagenknecht, B. D. Halligan, K. A. Young, J. E. Hokanson, G. R. Nashko, V. Gudnason, M. A. Province, P. A. Peyser, N. D. Palmer, E. K. Speliotes, Genome-wide association meta-analysis identifies 17 loci associated with nonalcoholic fatty liver disease. *Nat. Genet.* **55**, 1640–1650 (2023).
10. Y. Li, E. H. van den Berg, A. Kurilshikov, D. V. Zernakova, R. Gacesa, S. Hu, E. A. Lopera-Maya, A. Zernakova, V. E. de Meijer, S. Sanna, R. P. F. Dullaart, H. Blokzijl, E. A. M. Festen, J. Fu, R. K. Weersma, Genome-wide studies reveal genetic risk factors for hepatic fat content. *Genomics Proteomics Bioinformatics* **22**, qzae031 (2024).
11. N. Ghodsi, E. Abner, C. A. Emdin, É. Gobeil, N. Taba, M. E. Haas, N. Perrot, H. D. Manikpurage, É. Gagnon, J. Bourgault, A. St-Amand, C. Couture, P. L. Mitchell, Y. Bossé, P. Mathieu, M. C. Vohl, A. Tchernof, S. Thériault, A. V. Khera, T. Esko, B. J. Arsenault, Electronic health record-based genome-wide meta-analysis provides insights on the genetic architecture of non-alcoholic fatty liver disease. *Cell Rep. Med.* **2**, 100437 (2021).
12. F. Salomone, R. M. Pipitone, M. Longo, F. Malvestiti, A. M. Amorini, A. Distefano, E. Casirati, E. Ciociola, N. Iraci, L. Leggio, R. Zito, N. Vicario, C. Saoca, G. Pennisi, D. Cabibi, G. Lazzarino, A. L. Fracanzani, P. Dongiovanni, L. Valenti, S. Petta, G. L. Volfi, S. Grimaudo, SIRT5 rs12216101 T>G variant is associated with liver damage and mitochondrial dysfunction in patients with non-alcoholic fatty liver disease. *J. Hepatol.* **80**, 10–19 (2024).
13. D. Lindén, S. Romeo, Therapeutic opportunities for the treatment of NASH with genetically validated targets. *J. Hepatol.* **79**, 1056–1064 (2023).
14. L. Verschuren, A. L. Mak, A. van Koppen, S. Özsezen, S. Difrancesco, M. P. M. Caspers, J. Snabel, D. van der Meer, A. M. van Dijk, E. B. Rashu, P. Nabilou, M. P. Werge, K. van Son, R. Kleemann, A. J. Kiliaan, E. J. Hazebroek, A. Boonstra, W. P. Brouwer, M. Doukas, S. Gupta, C. Kluit, M. Nieuwdorp, J. Verheij, L. L. Gluud, A. G. Holleboom, M. E. Tushuizen, R. Hanemaaijer, Development of a novel non-invasive biomarker panel for hepatic fibrosis in MASLD. *Nat. Commun.* **15**, 4564 (2024).
15. O. Jamialahmadi, A. De Vincentis, F. Tavaglione, F. Malvestiti, R. Li-Gao, R. M. Mancina, M. Alvarez, K. Gelev, S. Maurotti, U. Vespasiani-Gentilucci, F. R. Rosendaal, J. Kozlitina, P. Pajukanta, F. Pattou, L. Valenti, S. Romeo, Partitioned polygenic risk scores identify distinct types of metabolic dysfunction-associated steatotic liver disease. *Nat. Med.* **30**, 3614–3623 (2024).
16. V. L. Chen, G. F. Brady, Recent advances in MASLD genetics: Insights into disease mechanisms and the next frontiers in clinical application. *Hepatol. Commun.* **9**, e0618 (2025).
17. A. L. Hutchison, F. Tavaglione, S. Romeo, M. Charlton, Endocrine aspects of metabolic dysfunction-associated steatotic liver disease (MASLD): Beyond insulin resistance. *J. Hepatol.* **79**, 1524–1541 (2023).
18. A. Mantovani, C. D. Byrne, E. Bonora, G. Targher, Nonalcoholic fatty liver disease and risk of incident type 2 diabetes: A meta-analysis. *Diabetes Care* **41**, 372–382 (2018).
19. A. Mantovani, G. Petracca, G. Beatrice, A. Csermely, A. Lonardo, J. M. Schattenberg, H. Tilg, C. D. Byrne, G. Targher, Non-alcoholic fatty liver disease and risk of incident chronic kidney disease: An updated meta-analysis. *Gut* **71**, 156–162 (2022).
20. Y. J. Abozaid, I. Ayada, L. A. van Kleef, C. L. Valleria, Q. Pan, W. P. Brouwer, M. A. Ikram, J. Van Meurs, R. J. de Knecht, M. Ghanbari, Plasma proteomic signature of fatty liver disease: The Rotterdam study. *Hepatology* **78**, 284–294 (2023).
21. S. Mia, G. Siokatas, R. Sidiropoulou, M. Hoffman, K. Fragkiadakis, E. Markopoulou, M. I. Elesawy, R. Roy, S. Blair, Y. Kuwabara, E. Rapushi, D. Chaudhuri, C. A. Makarewich, E. Gao, W. J. Koch, J. D. Schilling, J. D. Molkentin, M. Marketou, K. Drosatos, Hepato-cardiac interorgan communication controls cardiac hypertrophy via combined endocrine-autocrine FGF21 signaling. *Cell Rep. Med.* **6**, 102125 (2025).
22. M. J. Watt, P. M. Miotto, W. De Nardo, M. K. Montgomery, The liver as an endocrine organ-linking NAFLD and insulin resistance. *Endocr. Rev.* **40**, 1367–1393 (2019).
23. L. D'Erasmus, M. Di Martino, T. Neufeld, T. J. Fraum, C. J. Kang, K. H. Burks, A. Di Costanzo, I. Minicocci, S. Bini, M. Maranghi, G. Pigna, G. Labbadia, J. Zheng, D. Fierro, A. Montali, F. Ceci, C. Catalano, N. O. Davidson, G. Lucisano, A. Nicolucci, M. Arca, N. O. Stitzel, ANGPTL3 deficiency and risk of hepatic steatosis. *Circulation* **148**, 1479–1489 (2023).
24. Y. Liu, N. Basti, B. Whitcher, J. D. Bell, E. P. Sorokin, N. van Bruggen, E. L. Thomas, M. Cule, Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning. *eLife* **10**, e65554 (2021).
25. E. L. Dennis, N. Jahanshad, M. N. Braskie, N. M. Warstadt, D. P. Hibar, O. Kohnim, T. M. Nir, K. L. McMahon, G. I. de Zubicaray, G. W. Montgomery, N. G. Martin, A. W. Toga, M. J. Wright, P. M. Thompson, Obesity gene NEGR1 associated with white matter integrity in healthy young adults. *Neuroimage* **102**, 548–557 (2014).
26. C. E. Elks, R. J. Loos, S. J. Sharp, C. Langenberg, S. M. Ring, N. J. Timpson, A. R. Ness, G. Davey Smith, D. B. Dunger, N. J. Wareham, K. K. Ong, Genetic markers of adult obesity risk are associated with greater early infancy weight gain and growth. *PLoS Med.* **7**, e1000284 (2010).
27. T. M. Frayling, N. J. Timpson, M. N. Weedon, E. Zeggini, R. M. Freathy, C. M. Lindgren, J. R. Perry, K. S. Elliott, H. Lango, N. W. Rayner, B. Shields, L. W. Harries, J. C. Barrett, S. Ellard, C. J. Groves, B. Knight, A. M. Patch, A. R. Ness, S. Ebrahim, D. A. Lawlor, S. M. Ring, Y. Ben-Shlomo, M. R. Jarvelin, U. Sovio, A. J. Bennett, D. Melzer, L. Ferrucci, R. J. Loos, I. Barroso, N. J. Wareham, F. Karpe, K. R. Owen, L. R. Cardon, M. Walker, G. A. Hitman, C. N. Palmer, A. S. Doney, A. D. Morris, G. D. Smith, A. T. Hattersley, M. I. McCarthy, A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007).
28. G. Thorleifsson, G. B. Walters, D. F. Gudbjartsson, V. Steinthorsdottir, P. Sulem, A. Helgadóttir, U. Styrkarsdottir, S. Gretarsdottir, S. Thorlacius, I. Jonsdottir, T. Jonsdottir, E. J. Olafsdottir, G. H. Olafsdottir, T. Jonsson, F. Jonsson, K. Borch-Johnsen, T. Hansen, G. Andersen, T. Jorgensen, T. Lauritzen, K. K. Aben, A. L. Verbeek, N. Roeleveld, E. Kampman, L. R. Yanek, L. C. Becker, L. Tryggvadottir, T. Rafnar, D. M. Becker, J. Gulcher, L. A. Kiemeny, O. Pedersen, A. Kong, U. Thorsteinsdottir, K. Stefansson, Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18–24 (2009).
29. C. J. Willer, E. K. Speliotes, R. J. Loos, S. Li, C. M. Lindgren, I. M. Heid, S. I. Berndt, A. L. Elliott, A. U. Jackson, C. Lamina, G. Lettre, N. Lim, H. N. Lyon, S. A. McCarrroll, K. Papadakis, L. Qi, J. C. Randall, R. M. Roccasecca, S. Sanna, P. Scheet, M. N. Weedon, E. Wheeler, J. H. Zhao, L. C. Jacobs, I. Prokopenko, N. Soranzo, T. Tanaka, N. J. Timpson, P. Almgren, A. Bennett, R. N. Bergman, S. A. Bingham, L. L. Bonnycastle, M. Brown, N. P. Burtt, P. Chines, L. Coin, F. S. Collins, J. M. Connell, C. Cooper, G. D. Smith, E. M. Dennison, P. Deodhar, P. Elliott, M. R. Erdos, K. Estrada, D. M. Evans, L. Gianniny, C. Gieger, C. J. Gillson, C. Guiducci, R. Hackett, D. Hadley, A. S. Hall, A. S. Havulinna, J. Hebebrand, A. Hofman, B. Isomaa, K. B. Jacobs, T. Johnson, P. Jousilahti, Z. Jovanovic, K. T. Khaw, P. Kraft, M. Kuokkanen, J. Kuusisto, J. Laitinen, E. G. Lakatta, J. Luan, R. N. Luben, M. Mangino, W. L. McArdle, T. Meitinger, A. Mulas, P. B. Munroe, N. Narisu, A. R. Ness, K. Northstone, S. O'Rahilly, C. Purmann, M. G. Rees, M. Ridderstråle, S. M. Ring, F. Rivadeneira, A. Ruokonen, M. S. Sandhu, J. Saramies, L. J. Scott, A. Scuteri, K. Silander, M. A. Sims, K. Song, J. Stephens, S. Stevens, H. M. Stringham, Y. C. Tung, T. T. Valle, C. M. Van Duijn, K. S. Vimalaswaran, P. Vollenweider, G. Waeber, C. Wallace, R. M. Watanabe, D. M. Waterworth, N. Watkins, J. C. Wittenman, E. Zeggini, G. Zhai, M. C. Zillikens, D. Altschuler, M. J. Caulfield, S. J. Chanock, I. S. Farooqi, L. Ferrucci, J. M. Guralnik, A. T. Hattersley, F. B. Hu, M. R. Jarvelin, M. Laakso, V. Mooser, K. K. Ong, W. H. Ouwehand, V. Salomaa, N. J. Samani, T. D. Spector, T. Tuomi, J. Tuomilehto, M. Uda, A. G. Uitterlinden, N. J. Wareham, P. Deloukas, T. M. Frayling, L. C. Groop, R. B. Hayes, D. J. Hunter, K. L. Mohlke, L. Peltonen, D. Schlessinger, D. P. Strachan, H. E. Wichmann, M. I. McCarthy, M. Boehnke, I. Barroso, G. R. Abecasis, J. N. Hirschhorn, Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).
30. A. Varbo, M. Benn, A. Tybjaerg-Hansen, P. Grande, B. G. Nordestgaard, TRIB1 and GSKR polymorphisms, lipid levels, and risk of ischemic heart disease in the general population. *Arterioscler. Thromb. Vasc. Biol.* **31**, 451–457 (2011).
31. Y. A. Khalil, J. P. Rabès, C. Boileau, M. Varret, APOE gene variants in primary dyslipidemia. *Atherosclerosis* **328**, 11–22 (2021).
32. A. J. Cupido, L. F. Rees, A. D. Hingorani, C. Finan, F. W. Asselbergs, G. K. Hovingh, A. F. Schmidt, Joint genetic inhibition of PCSK9 and CETP and the association with coronary artery disease: A factorial mendelian randomization study. *JAMA Cardiol.* **7**, 955–964 (2022).
33. Y. Li, P. P. He, D. W. Zhang, X. L. Zheng, F. S. Cayabyab, W. D. Yin, C. K. Tang, Lipoprotein lipase: From gene to atherosclerosis. *Atherosclerosis* **237**, 597–608 (2014).

34. J. T. Haas, S. Francque, B. Staels, Pathophysiology and mechanisms of nonalcoholic fatty liver disease. *Annu. Rev. Physiol.* **78**, 181–205 (2016).
35. L. Crudele, C. De Matteis, F. Novelli, E. Di Buduo, S. Petruzzelli, A. De Giorgi, G. Antonica, E. Berardi, A. Moschetta, Fatty liver index (FLI) is the best score to predict MASLD with 50% lower cut-off value in women than in men. *Biol. Sex Differ.* **15**, 43 (2024).
36. G. Bedogni, S. Bellentani, L. Miglioli, F. Masutti, M. Passalacqua, A. Castiglione, C. Tiribelli, The fatty liver index: A simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol.* **6**, 33 (2006).
37. R. K. Sterling, E. Lissen, N. Clumeck, R. Sola, M. C. Correa, J. Montaner, M. S. Sulkowski, F. J. Torriani, D. T. Dieterich, D. L. Thomas, D. Messinger, M. Nelson, Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* **43**, 1317–1325 (2006).
38. C. T. Wai, J. K. Greenon, R. J. Fontana, J. D. Kalbfleisch, J. A. Marrero, H. S. Conjeevaram, A. S. Lok, A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* **38**, 518–526 (2003).
39. X. Forns, S. Ampurdanès, J. M. Llovet, J. Aponte, L. Quintó, E. Martínez-Bauer, M. Bruguera, J. M. Sánchez-Tapias, J. Rodés, Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. *Hepatology* **36**, 986–992 (2002).
40. Y. Liang, H. Chen, Y. Liu, X. Hou, L. Wei, Y. Bao, C. Yang, G. Zong, J. Wu, W. Jia, Association of MAFLD with diabetes, chronic kidney disease, and cardiovascular disease: A 4.6-year cohort study in China. *J. Clin. Endocrinol. Metab.* **107**, 88–97 (2022).
41. P. Qiu, J. Du, C. Zhang, M. Li, H. Li, C. Chen, Increased risk of reflux esophagitis in non-obese individuals with nonalcoholic fatty liver disease: A cross-sectional study. *Ann. Med.* **55**, 2294933 (2023).
42. Y. Lu, J. Zhang, H. Li, T. Li, Association of non-alcoholic fatty liver disease with self-reported osteoarthritis among the US adults. *Arthritis Res. Ther.* **26**, 40 (2024).
43. H. Huang, Z. Liu, J. Xie, C. Xu, NAFLD does not increase the risk of incident dementia: A prospective study and meta-analysis. *J. Psychiatr. Res.* **161**, 435–440 (2023).
44. X. Bao, L. Kang, S. Yin, G. Engström, L. Wang, W. Xu, B. Xu, X. Zhang, X. Zhang, Association of MAFLD and MASLD with all-cause and cause-specific dementia: A prospective cohort study. *Alzheimer's Res. Ther.* **16**, 318 (2024).
45. E. Ferkingstad, P. Sulem, B. A. Atlason, G. Sveinbjornsson, M. I. Magnusson, E. L. Styrismisdottir, K. Gunnarsdottir, A. Helgason, A. Oddsson, B. V. Halldorsson, B. O. Jonsson, F. Zink, G. H. Halldorsson, G. Masson, G. A. Arnadottir, H. Katrinardottir, K. Juliusson, M. K. Magnusson, O. T. Magnusson, R. Fridriksdottir, S. Saevarsdottir, S. A. Gudjonsson, S. N. Stacey, S. Rognvaldsson, T. Eiriksdottir, T. A. Olafsdottir, V. Steinthorsdottir, V. Tragante, M. O. Ulfarsson, H. Stefansson, I. Jonsdottir, H. Holm, T. Rafnar, P. Melsted, J. Saemundsdottir, G. L. Norddahl, S. H. Lund, D. F. Gudbjartsson, U. Thorsteinsdottir, K. Stefansson, Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**, 1712–1721 (2021).
46. B. B. Sun, J. Chiou, M. Traylor, C. Benner, Y. H. Hsu, T. G. Richardson, P. Surendran, A. Mahajan, C. Robins, S. G. Vasequez-Grinnell, L. Hou, E. M. Kvistad, O. S. Burren, J. Davitte, K. L. Ferber, C. E. Gillies, Å. K. Hedman, S. Hu, T. Lin, R. Mikkilineni, R. K. Pendergrass, C. Pickering, B. Prins, D. Baird, C. Y. Chen, L. D. Ward, A. M. Deaton, S. Welsh, C. M. Willis, N. Lehner, M. Arnold, M. A. Wörheide, K. Suhre, G. Kastenmüller, A. Sethi, M. Cule, A. Raj, L. Burkitt-Gray, E. Melamud, M. H. Black, E. B. Fauman, J. M. M. Howson, H. M. Kang, M. I. McCarthy, P. Nioi, S. Petrovski, R. A. Scott, E. N. Smith, S. Szalma, D. M. Waterworth, L. J. Mitnau, J. D. Zsustakowski, B. W. Gibson, M. R. Miller, C. D. Whelan, Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
47. Y. Zhang, X. Gao, X. Bai, S. Yao, Y. Z. Chang, G. Gao, The emerging role of furin in neurodegenerative and neuropsychiatric diseases. *Transl. Neurodegener.* **11**, 39 (2022).
48. S. Park, S. Kim, B. Kim, D. S. Kim, J. Kim, Y. Ahn, H. Kim, M. Song, I. Shim, S. H. Jung, C. Cho, S. Lim, S. Hong, H. Jo, A. C. Fahed, P. Natarajan, P. T. Ellinor, A. Torkamani, W. Y. Park, T. Y. Yu, W. Myung, H. H. Won, Multivariate genomic analysis of 5 million people elucidates the genetic architecture of shared components of the metabolic syndrome. *Nat. Genet.* **56**, 2380–2391 (2024).
49. R. Karlsson Linnér, T. T. Mallard, P. B. Barr, S. Sanchez-Roige, J. W. Madole, M. N. Driver, H. E. Poore, R. de Vlaming, A. D. Grotzinger, J. J. Tielbeek, E. C. Johnson, M. Liu, S. B. Rosenthal, T. Ideker, H. Zhou, R. L. Kember, J. A. Pasman, K. J. H. Verweij, D. J. Liu, S. Vrieze, H. R. Kranzler, J. Gelernter, K. M. Harris, E. M. Tucker-Drob, I. D. Waldman, A. A. Palmer, K. P. Harden, P. D. Koellinger, D. M. Dick, Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nat. Neurosci.* **24**, 1367–1376 (2021).
50. Y. Duan, Y. Yang, S. Zhao, Y. Bai, W. Yao, X. Gao, J. Yin, Crosstalk in extrahepatic and hepatic system in NAFLD/NASH. *Liver Int.* **44**, 1856–1871 (2024).
51. K. Y. Wu, B. Cao, W. B. Chen, W. Wu, S. Zhao, X. Y. Min, J. Yang, J. Han, X. Dong, N. Wang, Y. Wu, P. Garred, S. H. Sacks, W. Zhou, K. Li, Collectin 11 has a pivotal role in host defense against kidney and bladder infection in mice. *Kidney Int.* **105**, 524–539 (2024).
52. R. Nechanitzky, D. Akbas, S. Scherer, I. Györy, T. Hoyle, S. Ramamoorthy, A. Diefenbach, R. Grosschedl, Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. *Nat. Immunol.* **14**, 867–875 (2013).
53. K. Sawada, H. Chung, S. Softic, M. E. Moreno-Fernandez, S. Divanovic, The bidirectional immune crosstalk in metabolic dysfunction-associated steatotic liver disease. *Cell Metab.* **35**, 1852–1871 (2023).
54. T. Xiao, L. A. van Kleef, M. K. Ikram, R. J. de Kneft, M. A. Ikram, Association of nonalcoholic fatty liver disease and fibrosis with incident dementia and cognition: The Rotterdam study. *Neurology* **99**, e565–e573 (2022).
55. L. Lind, M. Mazidi, R. Clarke, D. A. Bennett, R. Zheng, Measured and genetically predicted protein levels and cardiovascular diseases in UK Biobank and China Kadoorie Biobank. *Nat. Cardiovasc. Res.* **3**, 1189–1198 (2024).
56. A. Mishra, R. Malik, T. Hachiya, T. Jürgenson, S. Namba, D. C. Posner, F. K. Kamanu, M. Koido, Q. Le Grand, M. Shi, Y. He, M. K. Georgakis, I. Caro, K. Krebs, Y. C. Liaw, F. C. Vaura, K. Lin, B. S. Winsvold, V. Srinivasasainagendra, L. Parodi, H. J. Bae, G. Chauhan, M. R. Chong, L. Tomppo, R. Akinyemi, G. V. Roshchupkin, N. Habib, Y. H. Jee, J. Q. Thomassen, V. Abedi, J. Cárcel-Márquez, M. Nygaard, H. L. Leonard, C. Yang, E. Yonova-Doing, M. J. Knol, A. J. Lewis, R. L. Judy, T. Ago, P. Amouyel, N. D. Armstrong, M. K. Bakker, T. M. Bartz, D. A. Bennett, J. C. Bis, C. Bordes, S. Børte, A. Cain, P. M. Ridker, K. Cho, Z. Chen, C. Cruchaga, J. W. Cole, P. L. de Jager, R. de Cid, M. Endres, L. E. Ferreira, M. I. Geerlings, N. C. Gasca, V. Gudnason, J. Hata, J. He, A. K. Heath, Y. L. Ho, A. S. Havulinna, J. C. Hopewell, H. I. Hyacinth, M. Inouye, M. A. Jacob, C. E. Jeon, C. Jern, M. Kamouchi, K. L. Keene, T. Kitazono, S. J. Kittner, T. Konuma, A. Kumar, P. Lacaze, L. J. Launer, K. J. Lee, K. Lepik, J. Li, L. Li, A. Manichaikul, H. S. Markus, N. A. Marston, T. Meitinger, B. D. Mitchell, F. A. Montellano, T. Morisaki, T. H. Mosley, M. A. Nalls, B. G. Nordestgaard, M. J. O'Donnell, Y. Okada, N. C. Onland-Moret, B. Ombigale, A. Peters, B. M. Psaty, S. S. Rich, J. Rosand, M. S. Sabatine, R. L. Sacco, D. Saleheen, E. C. Sandset, V. Salomaa, M. Sargurupremraj, M. Sasaki, C. L. Satizabal, C. O. Schmidt, A. Shimizu, N. L. Smith, K. L. Sloane, Y. Sutoh, Y. V. Sun, K. Tanno, S. Tiedt, T. Tatlisumak, N. P. Torres-Aguila, H. K. Tiwari, D. A. Trégouët, S. Trompet, A. M. Tuladhar, A. Tybjaerg-Hansen, N. Hansen, R. Vibo, S. C. Verma, K. L. Wiggins, P. Wennberg, D. Woo, P. W. F. Wilson, H. Xu, Q. Yang, K. Yoon, I. Y. Millwood, C. Gieger, T. Ninomiya, H. J. Grabe, J. W. Jukema, I. L. Rissanen, D. Strbian, Y. J. Kim, P. H. Chen, E. Mayerhofer, J. M. M. Howson, M. R. Irvin, H. Adams, S. Wassertheil-Smoller, K. Christensen, M. A. Ikram, T. Rundek, B. B. Worrall, G. M. Lathrop, M. Riaz, E. M. Simonsick, J. Körv, P. H. C. França, R. Zand, K. Prasad, R. Frikke-Schmidt, F. E. de Leeuw, T. Liman, K. G. Haeusler, Y. M. Ruigrok, P. U. Heuschmann, W. T. Longstreth, K. J. Jung, L. Bastarache, G. Paré, S. M. Damrauer, D. I. Chasman, J. I. Rotter, C. D. Anderson, J. A. Zwart, T. J. Niiranen, M. Fornage, Y. P. Liaw, S. Seshadri, I. Fernández-Cadenas, R. G. Walters, C. T. Ruff, M. O. Owolabi, J. E. Huffman, L. Milani, Y. Kamatani, M. Dichgans, S. Debette, Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature* **611**, 115–123 (2022).
57. R. E. Peterson, K. Kuchenbaecker, R. K. Walters, C. Y. Chen, A. B. Popejoy, S. Periyasamy, M. Lam, C. Iyegbe, R. J. Strawbridge, L. Brick, C. E. Carey, A. R. Martin, J. L. Meyers, J. Su, J. Chen, A. C. Edwards, A. Kalungi, N. Koen, L. Majara, E. Schwarz, J. W. Smoller, E. A. Stahl, P. F. Sullivan, E. Vassos, B. Mowry, M. L. Prieto, A. Cuellar-Barboza, T. B. Bigdeli, H. J. Edenberg, H. Huang, L. E. Duncan, Genome-wide association studies in ancestrally diverse populations: Opportunities, methods, pitfalls, and recommendations. *Cell* **179**, 589–603 (2019).
58. L. Kachuri, N. Chatterjee, J. Hirbo, D. J. Schaid, I. Martin, I. J. Kullo, E. E. Kenny, B. Pasienc, P. L. Auer, M. P. Conomos, D. V. Conti, Y. Ding, Y. Wang, H. Zhang, Y. Zhang, J. S. Witte, T. Ge, G., Principles and methods for transferring polygenic risk scores across global populations. *Nat. Rev. Genet.* **25**, 8–25 (2024).
59. T. Xia, M. Du, H. Li, Y. Wang, J. Zha, T. Wu, S. Ju, Association between liver MRI proton density fat fraction and liver disease risk. *Radiology* **309**, e231007 (2023).
60. A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
61. B. K. Bulik-Sullivan, P. R. Loh, H. K. Finucane, S. Ripke, J. Yang, N. Patterson, M. J. Daly, A. L. Price, B. M. Neale, LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
62. J. L. Horn, A rationale and test for the number of factors in factor analysis. *Psychometrika* **30**, 179–185 (1965).
63. A. D. Grotzinger, M. Rhemtulla, R. de Vlaming, S. J. Ritchie, T. T. Mallard, W. D. Hill, H. F. Ip, R. E. Marioni, A. M. McIntosh, I. J. Deary, P. D. Koellinger, K. P. Harden, M. G. Nivard, E. M. Tucker-Drob, Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
64. M. G. Wolf, D. McNeish, dynamic : An R package for deriving dynamic fit index cutoffs for factor analysis. *Multivariate Behav. Res.* **58**, 189–194 (2023).
65. T. T. Mallard, R. K. Linnér, A. D. Grotzinger, S. Sanchez-Roige, J. Seidlitz, A. Okbay, R. de Vlaming, S. F. W. Meddens, A. A. Palmer, L. K. Davis, E. M. Tucker-Drob, K. S. Kendler, M. C. Keller, P. D. Koellinger, K. P. Harden, Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *Cell Genom.* **2**, 100140 (2022).
66. P. Turley, R. K. Walters, O. Maghazian, A. Okbay, J. J. Lee, M. A. Fontana, T. A. Nguyen-Viet, R. Wedow, M. Zacher, N. A. Furlotte, P. Magnusson, S. Oskarsson, M. Johannesson,

- P. M. Visscher, D. Laibson, D. Cesarini, B. M. Neale, D. J. Benjamin, Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
67. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
 68. K. Watanabe, E. Taskesen, A. van Bochoven, D. Posthuma, Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
 69. E. Solis, A. Mosaku, A. Abid, A. Buniello, M. Cerezo, L. Gil, T. Groza, O. Günes, P. Hall, J. Hayhurst, A. Ibrahim, Y. Ji, S. John, E. Lewis, J. A. L. MacArthur, A. McMahon, D. Osumi-Sutherland, K. Panoutsopoulou, S. Pendlington, S. Ramachandran, R. Stefancsik, J. Stewart, P. Whetzel, R. Wilson, L. Hindorf, F. Cunningham, S. A. Lambert, M. Inouye, H. Parkinson, L. W. Harris, The NHGRI-EBI GWAS Catalog: Knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–D985 (2023).
 70. M. Kanai, M. Akiyama, A. Takahashi, N. Matoba, Y. Momozawa, M. Ikeda, N. Iwata, S. Ikegawa, M. Hirata, K. Matsuda, M. Kubo, Y. Okada, Y. Kamatani, Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
 71. D. Ochoa, A. Hercules, M. Carmona, D. Suveges, J. Baker, C. Malangone, I. Lopez, A. Miranda, C. Cruz-Castillo, L. Fumis, M. Bernal-Llinares, K. Tsukanov, H. Cornu, K. Tsigiris, O. Razuvayevskaya, A. Buniello, J. Schwartzentruber, M. Karim, B. Ariano, R. E. Martinez Osorio, J. Ferrer, X. Ge, S. Machlitt-Northen, A. Gonzalez-Uriarte, S. Saha, S. Tirunagari, C. Mehta, J. M. Roldán-Romero, S. Horswell, S. Young, M. Ghousaini, D. G. Hulcoop, I. Dunham, E. M. McDonagh, The next-generation open targets platform: Reimagined, redesigned, rebuilt. *Nucleic Acids Res.* **51**, D1353–D1359 (2023).
 72. C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
 73. A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, P. Tamayo, The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
 74. H. K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P. R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, F. R. Day, S. Purcell, E. Stahl, S. Lindstrom, J. R. Perry, Y. Okada, S. Raychaudhuri, M. J. Daly, N. Patterson, B. M. Neale, A. L. Price, Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
 75. H. K. Finucane, Y. A. Reshef, V. Anttila, K. Slowikowski, A. Gusev, A. Byrnes, S. Gazal, P. R. Loh, C. Lareau, N. Shores, G. Genovese, A. Saunders, E. Macosko, S. Pollack, J. R. B. Perry, J. D. Buenrostro, B. E. Bernstein, S. Raychaudhuri, S. McCarroll, B. M. Neale, A. L. Price, Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
 76. R. S. Fehrmann, J. M. Karjalainen, M. Krajewska, H. J. Westra, D. Maloney, A. Simeonov, T. H. Pers, J. N. Hirschhorn, R. C. Jansen, E. A. Schultes, H. H. van Haagen, E. G. de Vries, G. J. te Meerman, C. Wijmenga, M. A. van Vugt, L. Franke, Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
 77. T. H. Pers, J. M. Karjalainen, Y. Chan, H. J. Westra, A. R. Wood, J. Yang, J. C. Lui, S. Vedantam, S. Gustafsson, T. Esko, T. Frayling, E. K. Speliotes, M. Boehnke, S. Raychaudhuri, R. S. Fehrmann, J. N. Hirschhorn, L. Franke, Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
 78. A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y. C. Wu, A. R. Pfening, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shores, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K. H. Farh, S. Feizi, R. Karlic, A. R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthal, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L. H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
 79. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
 80. M. Schipper, C. A. de Leeuw, B. Maciel, D. P. Wightman, N. Hubers, D. I. Boomsma, M. C. O'Donovan, D. Posthuma, Prioritizing effector genes at trait-associated loci using multimodal evidence. *Nat. Genet.* **57**, 323–333 (2025).
 81. E. M. Weeks, J. C. Ulirsch, N. Y. Cheng, B. L. Trippie, R. S. Fine, J. Miao, T. A. Patwardhan, M. Kanai, J. Nasser, C. P. Fulco, K. C. Tashman, F. Aguet, T. Li, J. Ordoñez-Montanes, C. S. Smillie, M. Biton, A. K. Shalek, A. N. Ananthakrishnan, R. J. Xavier, A. Regev, R. M. Gupta, K. Lage, K. G. Ardlie, J. N. Hirschhorn, E. S. Lander, J. M. Engreitz, H. K. Finucane, Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat. Genet.* **55**, 1267–1276 (2023).
 82. K. Yuan, R. J. Longchamps, A. F. Pardiñas, M. Yu, T. T. Chen, S. C. Lin, Y. Chen, M. Lam, R. Liu, Y. Xia, Z. Guo, W. Shi, C. Shen, M. J. Daly, B. M. Neale, Y. A. Feng, Y. F. Lin, C. Y. Chen, M. C. O'Donovan, T. Ge, H. Huang, Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases. *Nat. Genet.* **56**, 1841–1850 (2024).
 83. M. Martens, A. Ammar, A. Riutta, A. Waagmeester, D. N. Slenter, K. Hanspers, R. A. Miller, D. Digles, E. N. Lopes, F. Ehrhart, L. J. Dupuis, L. A. Winckers, S. L. Coort, E. L. Willighagen, C. T. Evelo, A. R. Pico, M. Kutmon, WikiPathways: Connecting communities. *Nucleic Acids Res.* **49**, D613–d621 (2021).
 84. GTEx Consortium, The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
 85. C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, R. Collins, UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
 86. C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, J. Marchini, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
 87. Z. Liu, C. Song, C. Suo, H. Fan, T. Zhang, L. Jin, X. Chen, Alcohol consumption and hepatocellular carcinoma: Novel insights from a prospective cohort study and nonlinear Mendelian randomization analysis. *BMC Med.* **20**, 413 (2022).
 88. A. De Vincentis, F. Tavaglione, O. Jamilahmadi, A. Picardi, R. Antonelli Incalzi, L. Valenti, S. Romeo, U. Vespasiani-Gentilucci, A polygenic risk score to refine risk stratification and prediction for severe liver disease by clinical fibrosis scores. *Clin. Gastroenterol. Hepatol.* **20**, 658–673 (2022).
 89. F. Privé, B. J. Vilhjálmsson, H. Aschard, M. G. B. Blum, Making the most of clumping and thresholding for polygenic scores. *Am. J. Hum. Genet.* **105**, 1213–1221 (2019).
 90. B. B. Sun, J. C. Maranville, J. E. Peters, D. Stacey, J. R. Staley, J. Blackshaw, S. Burgess, T. Jiang, E. Paige, P. Surendran, C. Oliver-Williams, M. A. Kamat, B. P. Prins, S. K. Wilcox, E. S. Zimmerman, A. Chi, N. Bansal, S. L. Spain, A. M. Wood, N. W. Morrell, J. R. Bradley, N. Janjic, D. J. Roberts, W. H. Ouwehand, J. A. Todd, N. Soranzo, K. Suhre, D. S. Paul, C. S. Fox, R. M. Plenge, J. Danesh, H. Runz, A. S. Butterworth, Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
 91. M. Pietzner, E. Wheeler, J. Carrasco-Zanini, A. Cortes, M. Koprulu, M. A. Wörheide, E. Oerton, J. Cook, I. D. Stewart, N. D. Kerrison, J. Luan, J. Raffler, M. Arnold, W. Arlt, S. O'Rahilly, G. Kastenmüller, E. R. Gamazon, A. D. Hingorani, R. A. Scott, N. J. Wareham, C. Langenberg, Mapping the proteo-genomic convergence of human diseases. *Science* **374**, eabj1541 (2021).
 92. K. Suhre, M. Arnold, A. M. Bhagwat, R. J. Cotton, R. Engelke, H. Raffler, H. Sarwath, G. Thareja, A. Wahl, R. K. DeLisle, L. Gold, M. Pezer, G. Lauc, M. A. El-Din Selim, D. O. Mook-Kanamori, E. K. Al-Dous, Y. A. Mohamoud, J. Malek, K. Strauch, H. Grallert, A. Peters, G. Kastenmüller, C. Gieger, J. Graumann, Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).
 93. L. Folkersen, S. Gustafsson, Q. Wang, D. H. Hansen, Å. K. Hedman, A. Schork, K. Page, D. V. Zernakova, Y. Wu, J. Peters, N. Eriksson, S. E. Bergen, T. S. Boutin, A. D. Bretherick, S. Enroth, A. Kalnapenkis, J. R. Gådin, B. E. Suur, Y. Chen, L. Matic, J. D. Gale, J. Lee, W. Zhang, A. Quazi, M. Ala-Korpela, S. H. Choi, A. Claringbould, J. Danesh, G. D. Smith, F. de Masi, S. Elmståhl, G. Engström, E. Fauman, C. Fernandez, L. Franke, P. W. Franks, V. Giedraitis, C. Haley, A. Hamsten, A. Ingason, Å. Johansson, P. K. Joshi, L. Lind, C. M. Lindgren, S. Lubitz, T. Palmer, E. Macdonald-Dunlop, M. Magnusson, O. Melander, K. Michaelsson, A. P. Morris, R. Mägi, M. W. Nagle, P. M. Nilsson, J. Nilsson, M. Orho-Meland, O. Polasek, B. Prins, E. Pålsson, T. Qi, M. Sjögren, J. Sundström, P. Surendran, U. Vösa, T. Werge, R. Wernersson, H. J. Westra, J. Yang, A. Zernakova, J. Ärnlöv, J. Fu, J. G. Smith, T. Esko, C. Hayward, U. Gyllenstein, M. Landén, A. Siegbahn, J. F. Wilson, L. Wallentin, A. S. Butterworth, M. V. Holmes, E. Ingelsson, A. Mälarstig, Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* **2**, 1135–1148 (2020).
 94. J. H. Zhao, D. Stacey, N. Eriksson, E. Macdonald-Dunlop, Å. K. Hedman, A. Kalnapenkis, S. Enroth, D. Cozzetto, J. Digby-Bell, J. Marten, L. Folkersen, C. Herder, L. Jonsson, S. E. Bergen, C. Gieger, E. J. Needham, P. Surendran, D. S. Paul, O. Polasek, B. Thorand, H. Grallert, M. Roden, U. Vösa, T. Esko, C. Hayward, Å. Johansson, U. Gyllenstein, N. Powell, O. Hansson, N. Mattsson-Carlsson, P. K. Joshi, J. Danesh, L. Padyukov, L. Klarenskog, M. Landén, J. F. Wilson, A. Siegbahn, L. Wallentin, A. Mälarstig, A. S. Butterworth, J. E. Peters, Genetics of circulating inflammatory proteins identifies drivers of immune-mediated disease risk and therapeutic targets. *Nat. Immunol.* **24**, 1540–1551 (2023).
 95. A. Gudjonsson, V. Gudmundsdottir, G. T. Axelsson, E. F. Gudmundsson, B. G. Jonsson, L. J. Launer, J. R. Lamb, L. L. Jennings, T. Aspelund, V. Emilsson, V. Gudnason, A genome-wide association study of serum proteins reveals shared loci with common diseases. *Nat. Commun.* **13**, 480 (2022).
 96. A. Verma, J. E. Huffman, A. Rodriguez, M. Conery, M. Liu, Y. L. Ho, Y. Kim, D. A. Heise, L. Guare, V. A. Panickan, H. Garcon, F. Linares, L. Costa, I. Goethert, R. Tipton, J. Honerlaw, L. Davies, S. Whitbourne, J. Cohen, D. C. Posner, R. Sangar, M. Murray, X. Wang,

- D. R. Dochtermann, P. Devineni, Y. Shi, T. N. Nandi, T. L. Assimes, C. A. Brunette, R. J. Carroll, R. Clifford, S. Duval, J. Gelernter, A. Hung, S. K. Iyengar, J. Joseph, R. Kember, H. Kranzler, C. M. Kripke, D. Levey, S. W. Luoh, V. C. Merritt, C. Overstreet, J. D. Deak, S. F. A. Grant, R. Polimanti, P. Roussos, G. Shakt, Y. V. Sun, N. Tsao, S. Venkatesh, G. Voloudakis, A. Justice, E. Begoli, R. Ramoni, G. Tourassi, S. Pyarajan, P. Tsao, C. J. O'Donnell, S. Muralidhar, J. Moser, J. P. Casas, A. G. Bick, W. Zhou, T. Cai, B. F. Voight, K. Cho, J. M. Gaziano, R. K. Madduri, S. Damrauer, K. P. Liao, Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* **385**, eadj1182 (2024).
97. M. I. Kurki, J. Karjalainen, P. Palta, T. P. Sipilä, K. Kristiansson, K. M. Donner, M. P. Reeve, H. Laivuori, M. Aavikko, M. A. Kaunisto, A. Loukola, E. Lahtela, H. Mattsson, P. Laiho, P. D. B. Parolo, A. A. Lehisto, M. Kanai, N. Mars, J. Rämö, T. Kiiskinen, H. O. Heyne, K. Veerapen, S. Rüeger, S. Lemmelä, W. Zhou, S. Ruotsalainen, K. Pärn, T. Hiekkalinna, S. Koskelainen, T. Paajanen, V. Llorens, J. Gracia-Tabuenca, H. Siirtola, K. Reis, A. G. Elnahas, B. Sun, C. N. Foley, K. Aalto-Setälä, K. Alasoo, M. Arvas, K. Auro, S. Biswas, A. Bizaki-Vallaskangas, O. Carpen, C. Y. Chen, O. A. Dada, Z. Ding, M. G. Ehm, K. Eklund, M. Färkkilä, H. Finucane, A. Ganna, A. Ghazal, R. R. Graham, E. M. Green, A. Hakanen, M. Hautalahti, Å. K. Hedman, M. Hiltunen, R. Hinttala, I. Hovatta, X. Hu, A. Huertas-Vazquez, L. Huilaja, J. Hunkapiller, H. Jacob, J. N. Jensen, H. Joensuu, S. John, V. Julkunen, M. Jung, J. Junttila, K. Kaarniranta, M. Kähönen, R. Kajanne, L. Kallio, R. Kälviäinen, J. Kaprio, N. Kerimov, J. Kettunen, E. Kilpeläinen, T. Kilpi, K. Klinger, V. M. Kosma, T. Kuopio, V. Kurra, T. Laisk, J. Laukkanen, N. Lawless, A. Liu, S. Longerich, R. Mägi, J. Mäkelä, A. Mäkitie, A. Malarstig, A. Mannermaa, J. Maranville, A. Matakidou, T. Meretoja, S. V. Mozaffari, M. E. K. Niemä, M. Niemä, T. Niiranen, O. D. CJ, M. E. Obeidat, G. Okafo, H. M. Ollila, A. Palomäki, T. Palotie, J. Partanen, D. S. Paul, M. Pelkonen, R. K. Pendergrass, S. Petrovski, A. Pitkäranta, A. Platt, D. Pulford, E. Punkka, P. Pussinen, N. Raghavan, F. Rahimov, D. Rajpal, N. A. Renaud, B. Riley-Gillis, R. Rodosthenous, E. Saarentaus, A. Salminen, E. Salminen, V. Salomaa, J. Schleutker, R. Serpi, H. Y. Shen, R. Siegel, K. Silander, S. Siltanen, S. Soini, H. Soininen, J. H. Sul, I. Tachmazidou, K. Tasanen, P. Tienari, S. Toppila-Salmi, T. Tukiainen, T. Tuomi, J. A. Turunen, J. C. Ulirsch, F. Vaura, P. Virolainen, J. Waring, D. Waterworth, R. Yang, M. Nelis, A. Reigo, A. Metspalu, L. Milani, T. Esko, C. Fox, A. S. Havulinna, M. Perola, S. Ripatti, A. Jalanko, T. Laitinen, T. P. Mäkelä, R. Plenge, M. McCarthy, H. Runz, M. J. Daly, A. Palotie, FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
98. K. A. Broadaway, S. M. Brotman, J. D. Rosen, K. W. Currin, A. A. Alkhawaja, A. S. Etheridge, F. Wright, P. Gallins, D. Jima, Y.-h. Zhou, M. I. Love, F. Innocenti, K. L. Mohlke, Liver eQTL meta-analysis illuminates potential molecular mechanisms of cardiometabolic traits. *Am. J. Hum. Genet.* **111**, 1899–1913 (2024).

Acknowledgments

Funding: This work was funded by the Noncommunicable Chronic Diseases-National Science and Technology Major Project grant 2023ZD0510000 (X.C. and C.S.), the National Key Research and Development Program of China grant 2022YFC3400700 (Z.L.), the National Natural Science Foundation of China grants 82204125 and 82473700 (Z.L. and X.C.), and the Shanghai Rising-star Program grant 24QA2701000 (Z.L.). **Author contributions:** Conceptualization: Z.L., M.D., X.C., and H.Y. Methodology: M.D., Z.L., and T.W. Formal analysis: M.D. and Z.L. Investigation: M.D., Z.L., and X.C. Resources: M.D., H.Y., and C.S. Data curation: C.S. Writing—original draft: M.D. and Z.L. Writing—review and editing: Z.L., Y.J., X.C., T.Z., H.Y., and L.J. Visualization: M.D. and Z.L. Supervision: Z.L., L.J., T.Z., and X.C. Project administration: Z.L. and X.C. Funding acquisition: Z.L., X.C., and C.S. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The MASLD GWAS summary statistics were generated using a cross-trait approach integrating MRI-PDFF and cardiometabolic risk factors (see “Construction of MASLD GWAS using MTAG” in Materials and Methods) and are available via the NHGRI-EBI GWAS Catalog (GCP ID: GCP001525; accession: GCST90728570). All data and code needed to evaluate and reproduce the results in the paper are present in the paper and/or the Supplementary Materials. No new physical materials were generated. The individual-level phenotypic and genetic data for UKB (www.ukbiobank.ac.uk/) can be accessed through the application. The UKB received ethical approval from the research ethics committee (REC reference for UKB 11/NW/0382), and participants provided written informed consent. The GWAS summary statistics from the following consortia and biobanks are publicly available at the corresponding URL: NHGRI-EBI GWAS Catalog (www.ebi.ac.uk/gwas/), GIANT (https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files), GLGC (<https://csg.sph.umich.edu/willer/public/glgc-lipids2021/>), MAGIC (<https://magicinvestigators.org/>), and FinnGen (www.finnngen.fi/en/access_results). The GWAS summary statistics of the MVP cohort were obtained from GWAS catalog (www.ebi.ac.uk/gwas/publications/39024449). Plasma proteome GWAS was obtained from deCODE (www.deCODE.com/summarydata/; <https://doi.org/10.1038/s41588-021-00978-w>). Files required to run partition heritability with S-LDSC can be found at <https://zenodo.org/records/7768714>. MASLD GWAS summary data are available at GWAS catalog (GCP ID: GCP001525 and GCST90728570).

Submitted 18 August 2025

Accepted 9 January 2026

Published 13 February 2026

10.1126/sciadv.aeb5665