Genome Biology

RESEARCH

Open Access

Genetic associations of circulating plasma proteins with cardiometabolic diseases

Ruixin Zhou[1†], Jun Qiao[1†], Xiuzhen Zhang[2†], Xiaofei Yang[3], Wanzi Hong[4], Liuyang Cai[1], Pengwei Zhang[1], Siim Pauklin[5*], Yining Yang[6,7*], Zhaoyan Xu[4*] and Yuliang Feng[1,8*]

[†]Ruixin Zhou, Jun Qiao and Xiuzhen Zhang contributed equally to this work.

*Correspondence: pauklin@ndorms.ox.ac.uk; yangyn5126@xjrmyy.com; xuzhaoyanccu@gmail.com; fengyl@sustech.edu.cn

[1] Department of Pharmacology, SUSTech Homeostatic Medicine Institute, School of Medicine, Southern University of Science and Technology; Department of Geriatrics, Shenzhen People's Hospital (the First Affiliated Hospital of Southern University of Science and Technology), Shenzhen 518055, Guangdong, China
[4] Department of Cardiology, the First People's Hospital of Foshan (Foshan Hospital Affiliated to Southern University of Science and Technology), School of Medicine, Southern University of Science and Technology, Foshan 528000, Guangdong, China
[5] Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Headington, Oxford, UK
[6] Department of Cardiology, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi 830000, Xinjiang Uygur Autonomous Region, China
Full list of author information is available at the end of the article

Abstract

Background: Cardiovascular disease is the leading cause of death worldwide, and its risk is closely linked to metabolic abnormalities. Through summary-data-based mendelian randomization and colocalization analysis, we investigate the causal relationships between plasma proteins, six cardiovascular diseases (atrial fibrillation, coronary artery disease, heart failure, venous thromboembolism, peripheral artery disease and stroke), and 19 metabolic traits (including anthropometric phenotypes, blood pressure, glycemic phenotypes, inflammatory phenotypes, kidney-related phenotypes, lipidemic phenotypes, and liver-related phenotypes).

Results: We identify 49 proteins genetically associated with cardiovascular diseases, validated across two proteomic platforms. Among them, 35 are also associated with one or more metabolic phenotypes, with six showing evidence of colocalization. These six candidate proteins are classified into three categories based on drug development status, with PCSK9 already successful in therapies for cardiovascular diseases and hypercholesterolemia. DUSP13B, LRIG1, APOH, INHBC, and GUSB also demonstrate high therapeutic potential. Further phenome-wide MR analysis indicates that INHBC, APOH and DUSP13B represent promising therapeutic targets for cardiovascular diseases characterized by metabolic disorders.

Conclusions: Overall, this study revealed causal plasma proteins underlying the onset of cardiovascular diseases and metabolic abnormalities, advancing the understanding of disease mechanisms and facilitating drug discovery.

Keywords: Mendelian randomization, Cardiovascular diseases, Metabolic phenotypes, Proteome, Drug target

Background

Cardiovascular diseases (CVDs) encompass a range of conditions affecting the heart and blood vessels [1], and remain a leading cause of death worldwide, with a notably rising prevalence in developing countries and regions [1, 2]. Key behavioral risk factors, such as elevated blood pressure, high blood sugar, increased cholesterol levels, and obesity, significantly contribute to CVD risk. Addressing these metabolic abnormalities can

mitigate the risk of developing cardiovascular conditions. However, despite available clinical treatments for hypertension, dyslipidemia, and diabetes, there is an urgent need to improve their effectiveness [3–6].

Circulating proteins play a crucial role in the development of CVDs and may offer therapeutic potential [7]. As key metabolic products and signaling molecules, plasma proteins influence both physiological and pathological processes. Proteins closely linked to specific diseases can serve as biomarkers for diagnosis or as targets for pharmaceutical intervention. For instance, research has highlighted the involvement of the plasma protein proprotein convertase subtilisin/kexin type 9 (PCSK9) in lipid metabolism, and the introduction of PCSK9 inhibitors has significantly impacted lipid management and reduced cardiovascular risk [8, 9]. Therefore, targeting plasma proteins presents a promising strategy for treating both cardiovascular and metabolic diseases.

Increasing evidence suggests that genetic data can be used to identify and prioritize new drug targets and therapeutic indications [10]. Mendelian randomization (MR), an approach employing genetic variants as instrumental variables (IVs) to assess the impact of exposure on a specific outcome [11], is progressively being utilized to determine the causal links between diseases and associated proteins or genes [12, 13], facilitating the identification of druggable targets. Recently, Kim et al. conducted multi-omics and multi-trait MR analyses to identify 30 potential therapeutic targets for dyslipidemia, showcasing the utility of MR in drug discovery and development [14].

Here, we applied a proteome-wide summary data-based MR (SMR) analysis [15] and colocalization analysis [16], using the top single nucleotide polymorphism (SNP) from protein quantitative trait loci (pQTL) studies as IVs. Our analysis focused on genome-wide association studies (GWAS) outcomes related to atrial fibrillation (AF), coronary artery disease (CAD), heart failure (HF), venous thromboembolism (VTE), peripheral artery disease (PAD), stroke, and 19 metabolic traits. These analyses allowed us to investigate causal links between cardiometabolic diseases and plasma proteins, highlighting their potential as unified targets for both metabolic dysfunctions and cardiovascular conditions. To further assess the clinical relevance of these candidate proteins, we employed a comprehensive triple-analysis approach: (i) evaluating the repurposing potential of approved drugs and those in clinical trials, (ii) assessing the druggability of potential target proteins, and (iii) investigating the phenome-wide consequences of targeting these proteins. A flow diagram of our study is basically presented in Fig. 1.

Results

Proteome-wide MR and colocalization analysis identified associations between plasma proteins and CVDs

We conducted a proteome-wide MR analysis to assess the association between plasma proteins (with pQTL information) and the risk of six major CVDs. To control for false positives in analyses involving multiple phenotypes, we applied the Benjamin-Hochberg procedure to adjust the P values for false discovery rate (FDR) correction (p.adjust(p, method = 'fdr')) [17]. The number of tests in each analysis was set to n = length(p). After strict FDR correction ($P_{FDR} < 0.05$) and heterogeneity in dependent instruments (HEIDI) test ($P_{HEIDI} > 0.01$) for multiple testing ($P_{multi} < 0.05$), SMR analysis identified 189 proteins that showed causal relationships with the risk of

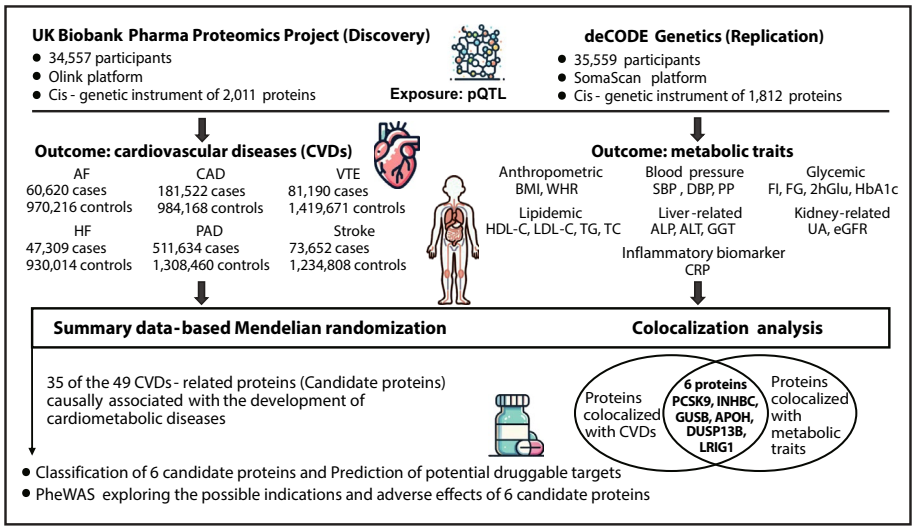

Fig. 1 Overview of this MR study design. After identifying that the top SNP with the strongest association signal in the cis-pQTL study was used as the single genetic instrument, SMR analysis comprehensively investigated the causal association of 2,011 plasma proteins with 6 major CVDs and 19 metabolic phenotypes. Plasma proteome data were obtained from two large-scale pQTL studies, including the UK Biobank Pharma Proteomics Project (UKB-PPP; $N = 34,557$) and the deCODE Health study ($N = 35,559$). Furthermore, we further explored potential therapeutic targets and evaluated the druggability of identified proteins using colocalization analysis. Additionally, a phenome-wide association study (PheWAS) was used to profile the possible side effects and indications of candidate proteins after their development into drugs. AF, atrial fibrillation; CAD, coronary artery disease; HF, heart failure; VTE, venous thromboembolism; PAD, peripheral artery disease; BMI, body mass index; WHR, waist hip ratio; SBP, systolic blood pressure; DBP, diastolic blood pressure; PP, pulse pressure; FI, fasting insulin; FG, fasting glucose; 2hGlu, two-hour glucose; HbA1c, glycated hemoglobin levels; LDL, low-density-lipoprotein; HDL, high-density-lipoprotein; TG, triglyceride; TC, total cholesterol; ALT, alanine aminotransferase; ALP, alkaline phosphatase; GGT, gamma-glutamyl transferase; eGFR, estimated glomerular filtration rate; UA, uric acid; CRP, C-reactive protein; PCSK9, proprotein convertase subtilisin/kexin type 9; INHBC, inhibin beta C chain; GUSB, beta-glucuronidase; APOH, apolipoprotein H; DUSP13B, dual-specificity phosphatase13B; LRIG1, leucine rich repeats and immunoglobulin like domains 1

CVDs in the discovery study (Fig. 2A). These thresholds were applied in the replication study. In the combined analysis with replication study, genetically predicted levels of 52 proteins were significantly associated with CVDs (Fig. 2B, Additional files 1 and 2). Among the proteins showing consistent associations across both studies, 49 exhibited directionally consistent relationships with CVDs. However, three proteins—S100 calcium-binding protein A16 (S100A16), angiopoietin-like 4 (ANGPTL4), and Fc gamma receptor II B (FCGR2B)—did not show consistent associations.

Per SD increase in genetically predicted levels of protein, the odds ratio of CVDs ranged from 0.56 (95% confidence interval [CI], 0.47 to 0.66) for protein S (PROS1) to 2.04 (95% CI, 1.74 to 2.39) for coagulation factor II (F2). It was genetically predicted that higher levels of five proteins were associated with decreased risk of AF, while eleven proteins were related to CAD, eight to VTE, and one to stroke. Conversely, elevated levels of two proteins were associated with a higher risk of AF, ten with CAD, twelve with VTE, and two each with stroke and PAD. Significantly, four proteins—dual-specificity phosphatase 13B (DUSP13B), asialoglycoprotein receptor 1 (ASGR1), PCSK9, and F2—were identified across multiple CVD conditions. A higher abundance of PCSK9 correlated with an increased risk of CAD and PAD, while F2 levels

were causally associated with a greater risk of VTE and stroke. DUSP13B showed an opposite effect between AF and CAD, as did ASGR1 between VTE and CAD.

Three different MR methods, including inverse variance weighted (IVW), MR Egger, and weighted median, were additionally applied to validate the robustness of the CVD-related proteins [18]. SNPs from pQTL in UKB-PPP and deCODE were extracted as IVs. IVW was applied as the main outcome, while MR-Egger and weighted median were used to improve the IVW estimates (Additional files 3 and 4). Among the 49 unique SMR-identified CVD-related proteins, 16 proteins had high support of colocalization analysis (PPH4 $\geq$ 0.8), and 4 proteins had medium support of colocalization analysis (0.8 > PPH4 $\geq$ 0.5) (Fig. 2C, Additional file 5). Four proteins had high support of colocalization with AF, including beta-glucuronidase (GUSB), DUSP13B, spondin 1 (SPON1), and tumor necrosis factor superfamily member 12 (TNFSF12). Seven plasma proteins were strongly colocalized with VTE, including epidermal growth factor-containing fibulin-like extracellular matrix protein 1 (EFEMP1), PROC, serine proteinase inhibitor clade E member 2 (SERPINE2), PROS1, protein phosphatase 1 regulatory inhibitor subunit 14 A (PPP1R14A), glycoprotein 6 (GP6) and chymotrypsin-like elastase family member 2 A (CELA2A; PPH4 $\geq$ 0.8 in discovery study but 0.8 > PPH4 $\geq$ 0.5 in replication). Additionally, PCSK9, hepatocyte growth factor activator (HGFAC), and inhibin beta C chain (INHBC) showed strong genetic colocalization support with CAD, while coagulation Factor XI (F11) and scavenger receptor class A member 5 (SCARA5) were associated with stroke. PCSK9 was the only protein with strong supportive evidence for PAD, which was also associated with CAD. Furthermore, leucine-rich repeats and immunoglobulin-like domains 1 (LRIG1), apolipoprotein H (APOH), thrombospondin 2 (THBS2), and F2, which showed medium support for one of CVDs, were also included in further analysis.

(See figure on next page.)

Fig. 2 SMR and colocalization analysis on the associations between plasma proteins and the risk of six kinds of cardiovascular diseases. **A** Manhattan plot for SMR analysis. Proteins above the dotted line were those with false discovery rate corrected $P < 0.05$. **B** Forest plot of the SMR analysis, displaying the odds ratios (OR) for the associations between plasma proteins and the outcome variables. **C** Colocalization analysis of associations between plasma proteins and five CVDs in both discovery and replication studies. Full names of proteins: DUSP13B, dual-specificity phosphatase13B; FBP1, fructose-1,6-bisphosphatase 1; GUSB, beta-glucuronidase; LRIG1, leucine rich repeats and immunoglobulin like domains 1; SPON1, spondin 1; TNFSF12, tumor necrosis factor superfamily member 12; ANGPTL4, angiopoietin-like 4; APOH, apolipoprotein H; ASGR1, asialoglycoprotein receptor 1; COL6A3, collagen type VI alpha 3 chain; COMT, catechol-O-methyltransferase; GAS6, growth arrest specific 6; GSTT2B, glutathione S-transferase theta 2B; HGFAC, hepatocyte growth factor activator; INHBC, inhibin beta C chain; ITIH3, inter-alpha-trypsin inhibitor heavy chain H3; LAYN, layilin; MIF, macrophage migration inhibitory factor; NADK, NAD kinase; PCSK9, proprotein convertase subtilisin/kexin type 9; PDE5A, phosphodiesterase type 5 A; RARRES2, retinoic acid receptor responder 2; S100A14, S100 calcium binding protein A14; S100A16, S100 calcium binding protein A16; SCARF2, scavenger receptor class F member; TIMD4, T cell immunoglobulin and mucin domain containing 4; VAT1, vesicle amine transport 1; WARS1, tryptophanyl-tRNA synthetase 1; F11, coagulation factor XI; F2, coagulation factor II; SCARA5, scavenger receptor class A member 5; AHSG, alpha 2-HS glycoprotein; ANXA2, annexin A2; APOC1, apolipoprotein C1; C2, complement C2; CACYBP, calcyclin binding protein; CD36, CD36 molecule; CELA2A, chymotrypsin like elastase 2 A; EFEMP1, EGF containing fibulin extracellular matrix protein 1; FCGR2B, Fc gamma receptor IIb; GP6, glycoprotein VI platelet; ITIH1, inter-alpha-trypsin inhibitor heavy chain 1; PPP1R14A, protein phosphatase 1 regulatory inhibitor subunit 14 A; PRDX6, peroxiredoxin 6; PROC, protein C; PROS1, protein S; SERPINE1, serpin family E member 1; SERPINE2, serpin family E member 2; SHBG, sex hormone-binding globulin; THBS2, thrombospondin 2

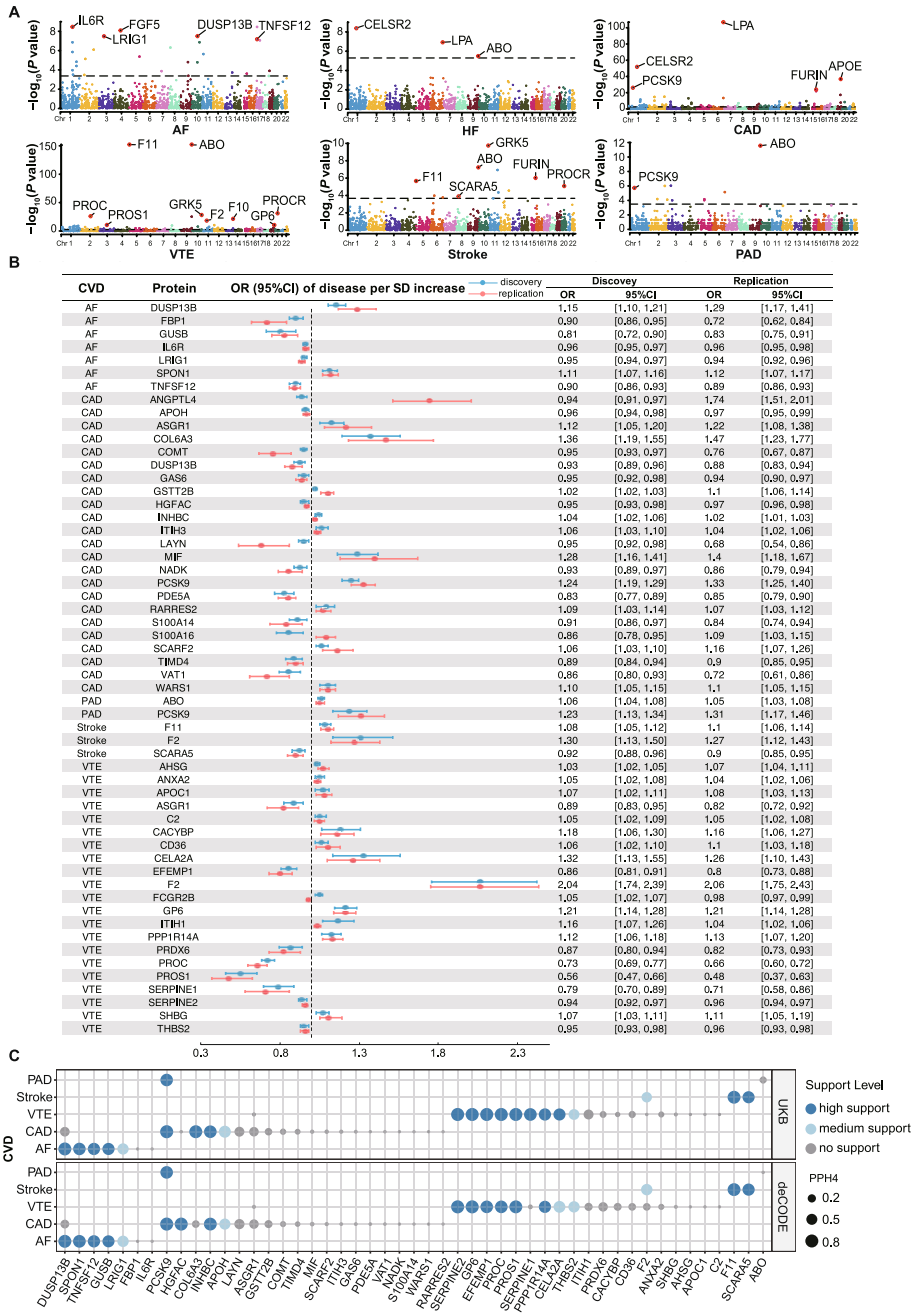

Fig. 2 (See legend on previous page.)

Causal connection and colocalization between CVD-related proteins and metabolic traits

We analyzed the relationships between 49 CVD-related plasma proteins and 19 major traits across 7 categories of metabolic phenotypes using the SMR method in both discovery and replication studies (Additional files 6 and 7). Among these proteins, 35 were causally associated with at least one metabolism-related trait after adjusting for multiple testing ($P_{FDR} < 0.05$, $P_{multi} < 0.05$) and HEIDI test ($P_{HEIDI} > 0.01$). However, no CVD-related proteins survived in SMR analysis with two-hour glucose (2hGlu) and fasting insulin (FI) as outcomes under the condition. The directions of

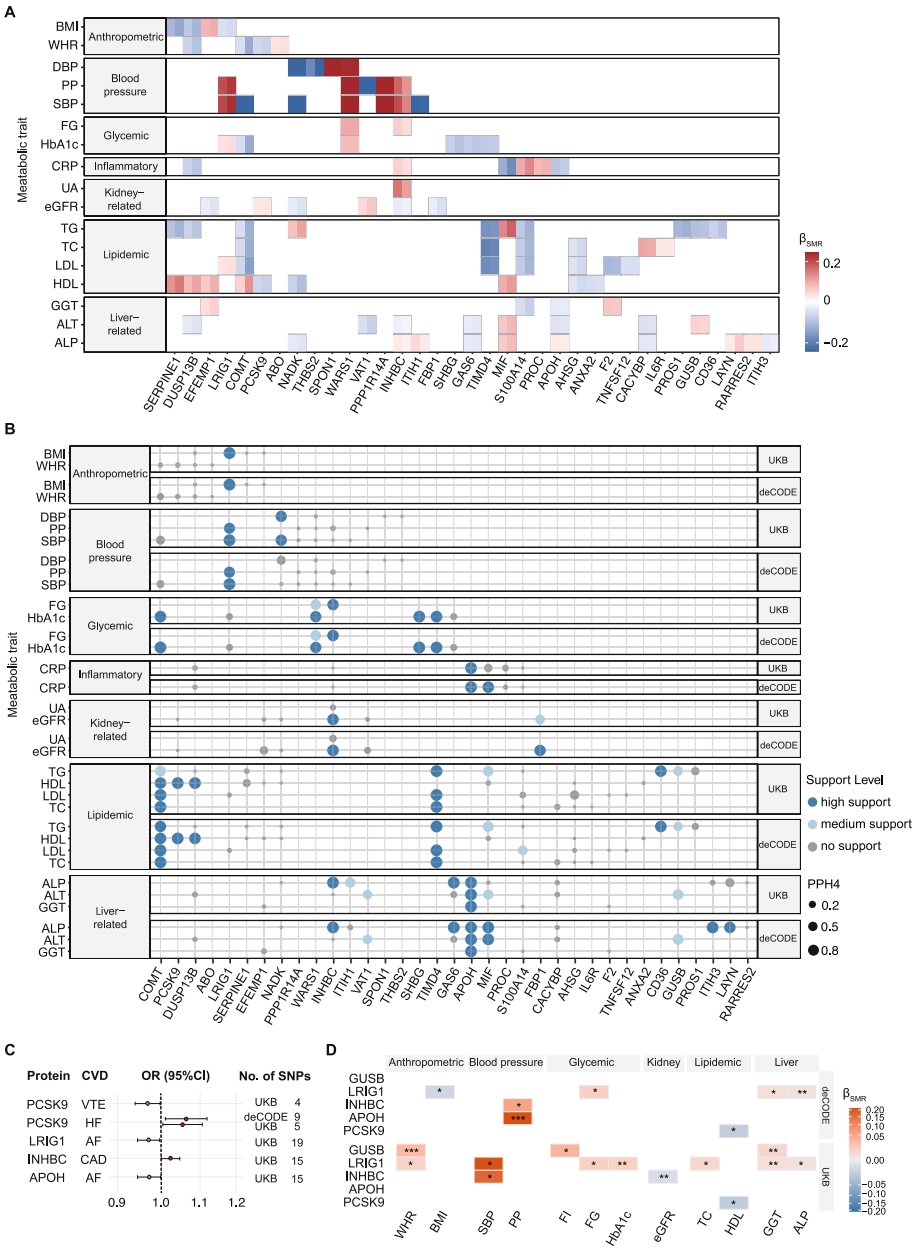

Fig. 3 Multiple analyses on the associations between plasma proteins and the six kinds of metabolic phenotypes. **A** Heatmap summarized beta value in SMR analysis (β_{SMR}) for associations between CVD-related proteins and 17 metabolic phenotypes. Each protein was split vertically: the left half shows the β_{SMR} from the discovery study, and the right half shows the β_{SMR} from the replication study. **B** Colocalization results for protein-metabolic-trait pairs across 17 phenotypes grouped into seven categories, shown for both discovery and replication datasets. Posterior support for a shared causal variant was reported as PPH4 (e.g., high support PPH4 $\geq$ 0.8; medium support 0.5 $\leq$ PPH4 < 0.8). **C** Random-effects IVW MR results for six proteins (PCSK9, LRIG1, INHBC, APOH, GUSB, and DUSP13B) versus cardiovascular outcomes. Only associations meeting $P < 0.05$ were displayed. ORs and their 95% CIs were shown. **D** Random-effects IVW MR results for the same six proteins versus six metabolic traits (anthropometric phenotypes, glycemic phenotypes, liver-related enzyme phenotypes, kidney-related phenotypes, lipidemic phenotypes and blood pressure). Displayed associations satisfy $P < 0.05$. β_{SMR} were presented

the associations between proteins and metabolic traits were summarized in Fig. 3A. For instance, a SD increase in NAD kinase (NADK) was linked with a decrease of −1.30 (95% CI, −1.57 to −1.05) in systolic blood pressure (SBP), while tryptophanyl-tRNA synthetase 1 (WARS1) was associated with an increase of 0.58 (95% CI, 0.31 to 0.85) in SBP.

Among the four proteins that showed overlapping associations with different CVDs, genetically predicted levels of DUSP13B were inversely associated with CAD, body mass index (BMI), and waist hip ratio (WHR) but positively associated with AF and high-density-lipoprotein (HDL). PCSK9 levels were significantly associated with lower HDL and WHR. Elevated levels of F2 were associated with lower low-density lipoprotein (LDL) levels, but also increased the risk of both VTE and stroke. The SMR results for ASGR1 and metabolic traits could not be replicated in the deCODE Health study.

Fifteen proteins were supported by colocalization analysis with at least one metabolic phenotype after overlapping the replication study (Fig. 3B, Additional file 8). In detail, five proteins had high support of genetic colocalization (PPH4 > 0.8) with fasting glucose (FG) or glycated hemoglobin levels (HbA1c) in glycemic traits, which were T cell immunoglobulin and mucin domain containing 4 (TIMD4), sex hormone-binding globulin (SHBG), catechol-O-methyltransferase (COMT), WARS1 and INHBC. Among the five proteins that had strong support evidence of colocalization (PPH4 > 0.8) with lipidemic traits, COMT was supported with all four lipidemic traits, including HDL, LDL, triglyceride (TG) and total cholesterol (TC), with the protein levels positively correlated with HDL and negatively correlated with LDL. TIMD4 displayed the most significant MR result (P_{FDR} = 1.67E-61) and was colocalized with LDL, TG, and TC, whose level was inversely correlated with the three phenotypes. Both DUSP13B and PCSK9 were supported only with HDL-C and cluster of differentiation 36 (CD36) only with TG in strong evidence. Besides, macrophage migration inhibitory factor (MIF) and GUSB had moderate support evidence (0.8 ≥ PPH4 > 0.5) with TG. Six proteins showed evidence of colocalization (PPH4 > 0.5) with liver-related traits. APOH strongly colocalized with all three enzymes, which were alanine aminotransferase (ALT), alkaline phosphatase (ALP), and γ-glutamyl transferase (GGT), and five other proteins (MIF, INHBC, growth arrest-specific 6 (GAS6), inter-alpha-trypsin inhibitor heavy chain H3 (ITIH3), layilin (LAYN)) were associated with one enzyme phenotype. Interestingly, APOH also showed powerful evidence (PPH4 > 0.8) colocalized with C-reactive protein (CRP), an inflammatory biomarker. And the evidence was observed between LRIG1 and SBP or pulse pressure (PP), and also between LRIG1 and BMI, indicating a high probability for a shared causal variant between LRIG1 level and levels of blood pressure and BMI. In kidney-related traits, INHBC and fructose-1,6-bisphosphatase 1 (FBP1) were supported by colocalization with estimated glomerular filtration rate (eGFR).

Integrating the colocalization analysis results, we identified six plasma proteins that had colocalization evidence with both CVDs and metabolic phenotypes, including LRIG1, INHBC, GUSB, APOH, DUSP13B, and PCSK9. Random-effects IVW analysis indicated that levels of PCSK9 were inverse associated with VTE and HDL but positively associated with HF. Moreover, both IVW analysis and colocalization analysis revealed that levels of INHBC had a positive association with CAD and eGFR (Figs. 2C and 3B-D).

Prediction of potential druggable targets

We performed KEGG pathway enrichment using DAVID [19] (EASE ≤ 0.1, Count ≥ 2) for 189 significant proteins in the discovery set (Fig. 4A) and 49 CVD-related proteins

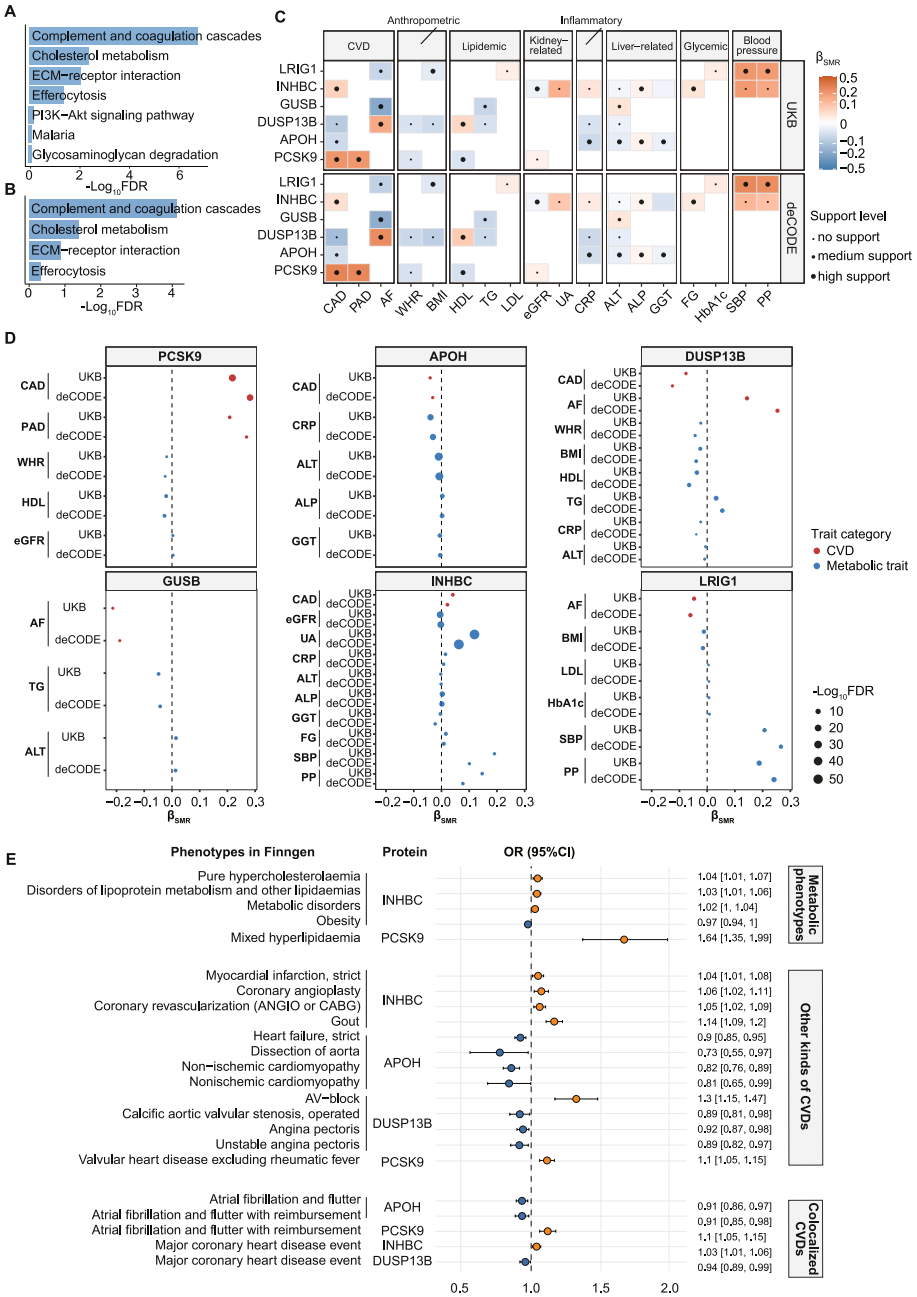

Fig. 4 Phenotypes significantly associated with six candidate proteins. **A**, **B** KEGG pathway enrichment for 189 significant proteins in the discovery set (A) and 49 CVD-related proteins (B). **C** Integrated result of SMR analysis of six colocalized candidate proteins, CVDs and six metabolic phenotypes. The direction of the associations was indicated by β_{SMR}. **D** Significance degree of SMR associations for PCSK9, APOH, DUSP13B, GUSB, INHBC and LRIG1 proteins with CAD, PAD, AF, and metabolic traits. **E** Phenome-wide SMR analysis of prioritized plasma proteins using FinnGen data. Only associations related to metabolism and CVDs that meet the thresholds of $P_{multi} < 0.05$ and $P_{HEIDI} > 0.01$ were displayed

(Fig. 4B). The analysis revealed that complement and coagulation cascades, as well as extracellular matrix (ECM)-receptor interaction, were significantly enriched. These pathways are directly linked to atherosclerosis and vascular stability. Cholesterol metabolism was also enriched as a key pathway, with APOH and PCSK9 as related genes, supporting our findings.

Each candidate protein displayed diverse relationships (Fig. 4C) and colocalization evidence between CVDs and metabolic phenotypes (Fig. 4D). To assess the druggability of these proteins, we comprehensively searched the DrugBank [20], the Drug Gene Interaction Database (DGIdb) [21], and Therapeutic Target Database (TTD) [22] and classified the proteins into three categories (Additional file 9). Category 1 included proteins for which targeted drugs have already been approved or are in clinical trials for treating CVDs or metabolic diseases. For instance, PCSK9 inhibitors like Alirocumab and Evolocumab have been approved for lowering the amount of cholesterol in the blood and decrease the chance of having a heart attack or stroke in selected patients. Category 2 included proteins whose targeted drugs are in clinical trials for diseases other than cardiovascular or metabolic disorders. For example, INHBC is being explored as a potential target for ovarian cancer treatment. Concurrently, GUSB has achieved success in treating mucopolysaccharidosis and periodontal disease, acting as an effective enzyme replacement therapy. APOH has been tested as a therapeutic target for Hughes syndrome (also known as antiphospholipid syndrome), although its efficacy requires further evaluation. Category 3 included proteins with no current clinical applications but showing potential as druggable targets for future therapeutic development, such as DUSP13B and LRIG1.

PheWAS exploring the possible indications and adverse effects of targeted proteins

To explore the possible indications and side effects of the six candidate proteins, a Phenome-wide association study (PheWAS) was conducted across 771 phenotypes from the FinnGen database, with at least 500 cases per phenotype (Additional file 10).

PCSK9 was one of the significant targets supported by PheWAS (Additional file 11). In addition to CAD and PAD (supported with colocalization analysis), the onset risks of multiple CVDs were positively associated with PCSK9 levels, including HF (OR [95%CI]: 1.05 [1.00,1.11]; OR [95%CI]: 1.06 [1.01,1.12]) (Fig. 3C) and AF (OR [95%CI]: 1.1 [1.05,1.15]) (Fig. 4E). Evidence showed that levels of PCSK9 were negatively associated with HDL (Figs. 3D and 4C, D), which also indicated that a high abundance of PCSK9 might contribute to hyperlipidemia (Fig. 4E).

Surprisingly, INHBC levels were causally associated with a reduced risk of CAD (Figs. 3C and 4C-E). Additionally, INHBC was positively linked with myocardial infarction (OR [95%CI]: 1.04 [1.01, 1.08]) and gout (OR [95%CI]: 1.14 [1.09, 1.2]), both of which were accompanied with metabolic disorders (Fig. 4E). However, using INHBC-targeted drugs might carry a poteintial risk of asthma (Additional file 11), with a P_{multi} value < 0.05. Both PheWAS data and IVW analysis indicated that increase levels of APOH may reduce the risk of AF (OR [95%CI]: 0.97 [0.94, 1.00] (Fig. 3C); OR [95%CI]: 0.91 [0.86, 0.97] (Fig. 4E)), HF (OR [95%CI]: 0.9 [0.85, 0.95]) and non-ischemic cardiomyopathy (OR [95%CI]: 0.82 [0.76, 0.89]) (Fig. 4E). SMR also suggested a reduced risk of CAD with higher APOH levels (Fig. 4C). Notably, APOH was also linked with

reduced CRP, ALT and GGT levels (Fig. 4C), but it was positively associated with PP level (Fig. 3D). Among the Category 2 proteins, GUSB showed fewer consistent findings across various analyses regarding its potential as a druggable target for both CVD and metabolic traits.

In addition to being associated with low risks of CAD (OR [95%CI]: 0.94 [0.89, 0.99]), calcific aortic valvular stenosis (OR [95%CI]: 0.89 [0.81, 0.98]) and angina pectoris (OR [95%CI]: 0.92 [0.87, 0.98]) (Fig. 4E), the abundance of DUSP13B was also inversely associated with multiple metabolic traits, including WHR, BMI, TG, CRP and ALT, whereas the abundance was positively associated with AF (Fig. 4C). The negative correlation between LRIG1 and atrial fibrillation and flutter was supported by the IVW analysis (OR [95%CI]: 0.97 [0.94, 1.00]) (Fig. 3C) and SMR analysis (OR [95%CI]: 0.95 [0.94, 0.97]) (Figs. 2B and 4C). Meanwhile, LRIG1 was positively connected with WHR, FG, HbA1c, GGT, ALP, TC, LDL and blood pressure according to prediction results (Fig. 3D), which can be taken as a consideration in drug designs.

Discussion

This study investigated the association between 2,011 plasma proteins and cardiometabolic diseases and evaluated the potential druggable targets. We conducted proteome-wide MR and colocalization analyses to identify causal plasma proteins on CVDs and metabolic traits exploiting genetic variants. Such an approach enhanced causal inference by minimizing biases from confounding and reverse causation. The MR results predicted that 49 unique plasma proteins had causal associations with at least one CVD, of which 35 proteins were also associated with diverse metabolic traits. Six proteins had the support of colocalization with CVDs and metabolic traits at the same time, which were classified as candidate proteins for druggability prioritization.

Our study found that genetically predicted higher levels of LRIG1 and GUSB were inversely associated with AF risks, whereas higher levels of circulating DUSP13B were positively associated with AF risks. Higher levels of circulating PCSK9 and INHBC, and lower levels of APOH and DUSP13B were associated with an increased risk of CAD, with high PCSK9 abundance also associated with the risk of PAD. Besides, the six proteins were associated with different metabolic traits, some of which were indications of the potential targets and a few as adverse effects. PheWAS analysis further revealed the wide range of health benefits and anticipated adverse effects after targeting the six candidate proteins.

Our study corroborated some previously identified associations between plasma proteins and CVDs, such as the associations of AF with IL6R [23], CAD with ASGR1 [24] and COL6A3 [25], and VTE with SHBG [26] and PROC [27]. Some well-studied proteins associated with metabolic diseases or traits were successfully identified, such as SPON1 [28], COMT [29], SHBG [30], and CD36 [31]. Of note, PCSK9, one of the 6 candidate proteins, has already been approved for use or is under clinical trials for hypercholesterolemia and CVD [8, 9, 32], indicating the reliability of data sources and validity of research approaches in this study. However, the study did not pinpoint well-known proteins that have been described in previous studies to be associated with both cardiovascular and metabolic diseases, like tumor necrosis factor-alpha (TNF-α) [33, 34] and insulin-like growth factor 1 receptor (IGF1R) [35], which was non-significant

with multiple testing or unrepeatable after datasets overlapping. However, this multiplex correction strategy was in accord with one of the study's purposes, which was to find plasma proteins that were strongly associated with cardiometabolic disease.

In addition to PCSK9, we observed that five proteins were more likely to be causally related to cardiometabolic diseases than other plasma proteins, including INHBC, APOH, GUSB, DUSP13B and LRIG1. As a member of the TGF-β family, the function of INHBC in regulating inflammatory responses and cell proliferation may be related to the pathogenesis of CVDs, especially atherosclerosis and myocardial infarction with chronic inflammatory processes. Human genetic and transgenic mouse studies have indicated the possibility of INHBC served as a drug target for atherogenic dyslipidemia, CAD, and NAFLD [36], which is concordant to our results. Moreover, other indications and modulatory targets for targeting INHBC were further noticed in this study. The causal associations of INHBC with eGFR and FG suggested a role in blood-glucose homeostasis and renal function; however, the mechanism—potentially involving stimulation of renal cell proliferation in the pathogenesis of diabetic nephropathy—remains unclear [37]. Three types of liver metabolic enzymes were closely related to INHBC levels, which needs to be considered in drug design.

Interestingly, DUSP13B acted as a protective factor in CAD and lipidemic disorders but not in AF in the study. The role of DUSP13B in inflammation regulation, cell proliferation, and signaling transduction [38] may suggest its molecular mechanisms in CVD. Recent findings also suggested that apoptosis of cardiomyocytes induced by reactive oxygen species might be reduced by up-regulating DUSP13 and inactivating the p38 MAPK pathway [39]. Likewise, APOH, associated with cardiometabolic disease, has been found to maintain blood fluidity and prevent nonspecific thrombosis, and it was revealed as a new candidate gene associated with thrombosis [40]. Possibly caused by multi-level of pleiotropy assessment, our colocalization analysis did not capture the association of APOH with lipidemic traits, as well as the inverse relationship between APOH and AF detected by IVW and phenome-wide SMR. By regulating EGFR signal and its related pathways [41], LRIG1 may indirectly affect inflammation and other metabolic regulation-related signaling pathways such as obesity and insulin resistance [42]. Actually, LRIG1, one of the LRIG proteins family members, regulated lipid metabolism via BMP signaling and affect the risk of type 2 diabetes [43], which supported the results in this work. Due to the important role of GUSB in female estrogen metabolism [44] and periodontitis development [45], abnormal activity may lead to enhanced cellular stress and inflammatory responses. GUSB has also been reported as an inherited metabolic disorder factor related to carbohydrate metabolism, leading to functional or structural lesions of the heart [46]. Except PCSK9 and INHBC, the candidate proteins levels were mostly negatively correlated with cardiovascular disease risk, followed by various relationships with blood lipids, blood glucose, liver metabolic enzymes, inflammatory factors and anthropometric phenotypes.

Our study evaluates the potential druggable targets beyond prior work [47, 48] by adopting a phenome-wide, multi-source, and translationally oriented framework. Rather than simply expanding the number of outcomes, we systematically evaluated six cardiovascular diseases and nineteen metabolic traits spanning anthropometry, glycemic control, lipid metabolism, blood pressure, renal and liver function, and

inflammation to capture the cardiometabolic continuum through which circulating proteins may exert pleiotropic and stage-specific effects. To improve robustness and generalizability, we integrated GWAS summary statistics from multiple independent consortia and selected instruments using cis-pQTLs from two large proteomic studies with stringent cross-platform consistency filters. Causal inference was strengthened through triangulation across complementary MR approaches (SMR, IVW, MR-Egger, and weighted median), prioritizing associations that were consistent across methods and datasets. Finally, we extended discovery to target prioritization by conducting PheWAS for prioritized proteins to assess potential on-target benefits and liabilities across diverse clinical phenotypes. This MR-PheWAS strategy provides novel insights that are not attainable from single-disease MR analyses and enhances the clinical relevance of proteome-wide causal inference.

Limitations of this analysis deserve to be noted. First, although our MR analyzed causal plasma proteins from two independent sources to increase the power, it is likely to overlook some weak associations and neglect viable therapeutic targets. However, all of the analyses on causality and colocalization were based on reproducible results from independent datasets of genetic variants from the UKB and deCODE, the bias introduced by the data source was reduced. Secondly, the pQTL data from the UK Biobank and the UKB-based GWAS datasets used as exposures and outcomes in our MR analyses inevitably share some degree of sample overlap, which could potentially introduce bias. Previous methodological work by Minelli et al. has demonstrated, however, that in large-scale biobank settings, such overlap does not materially distort two-sample MR estimates. Nevertheless, to further mitigate this potential concern, we incorporated independent pQTL data from deCODE, thereby substantially reducing sample overlap and strengthening the robustness of our causal inference. Furthermore, to assess the impact of any residual overlap, we additionally performed SMR analyses using UK Biobank-excluded GWAS summary statistics for AF [49] and HF [50]. The results were highly consistent with the primary analyses (Additional file 12), indicating that sample overlap did not materially influence our findings. Thirdly, we used large sample size GWAS data to discover more associated causal proteins, and restricted the scope of analysis to the European population to minimize population structure bias; however, it limits the generalization of our findings to other populations. As larger GWAS datasets from multiple populations become available, the depth of our analysis may be further improved. Finally, although we focused the drug targets on 6 plasma proteins, this does not necessarily mean that the remaining proteins cannot be treated with drugs. Our research aims to narrow the scope of drug development targets and reduce their time and resource costs. For exmple, some approved drug targets for abnormal conditions were not included among our 6 protein candidates, such as the COMT-targeted drug LOMITAPIDE, which is used to treat multiple dyslipidemias (including hypercholesterolemia, type II hyperlipoproteinemia and hyperlipidemia) and CVDs. Similarly, the SHBG-targeted drug LISINOPRIL is used for treating CVDs (including

heart failure, myocardial infarction, arterial disease, stroke and hypertrophy), hypercholesterolemia, hypertension, non-alcoholic fat liver and atherosclerosis.

Conclusions

In conclusion, this study revealed 35 causal proteins for the onset of cardiometabolic diseases and highlighted 6 promising drug targets: PCSK9, INHBC, APOH, GUSB, DUSP13B and LRIG1. These proteins suggest the involvement of inflammation and cell proliferation in the progression of cardiometabolic diseases, providing valuable insights for future drug development.

Methods

Study design

This study consisted of two parts: a proteome-wide summary data-based MR (SMR) [15] analysis that used single-nucleotide polymorphisms (SNPs) as instrument variables (IVs) to identify cardiometabolic-related plasma proteins, a colocalization analysis and a phenome-wide association study (PheWAS) analysis to explore protein targets with the greatest druggable potential.

Data sources for plasma proteins

Plasma proteome data were obtained from two large-scale pQTL studies: the UK Biobank Pharma Proteomics Project (UKB-PPP) [51] and the deCODE Health study [52]. UKB-PPP conducted proteomic analysis on plasma samples from 34,557 participants through the Olink platform and collected data on 2,011 proteins. Similarly, the deCODE Health study collected data on 1,812 proteins from 35,559 participants using the SomaScan platform. Cis-single-nucleotide polymorphisms (cis-SNPs), defined as SNPs within 1 Mb of the transcription start sites (TSS) of the corresponding genes, were selected from pQTL studies. These cis-SNPs were associated with the abundance of plasma proteins at the genome-wide significant level ($P < 5 \times 10^{-8}$) and were used as IVs. In this analysis, the UKB-PPP served as the discovery study and the deCODE Health study was used for replication. To ensure consistency of results across different proteomic analysis platforms, we presented the findings for proteins that overlapped between both studies, focusing on those with shared directional relationships in both the SMR and colocalization analyses.

Genome-wide association study (GWAS) data sources

Six major CVDs were included in our study: atrial fibrillation (AF; N cases = 60,620, N controls = 970,216) [49], heart failure (HF; N cases = 47,309, N controls = 930,014) [50], stroke (N cases = 73,652, N controls = 1,234,808) [53], venous thromboembolism (VTE; N cases = 81,190, N controls = 1,419,671) [54], coronary artery disease (CAD; N cases = 181,522, N controls = 984,168) [55] and peripheral artery disease (PAD; N cases = 12,086, N controls = 499,548) [56]. All participants in the studies are European. In addition to CVDs, our study also included summary-level data of

GWAS for the 19 metabolic traits across seven different categories of metabolic phenotypes. These data were obtained from several large-scale consortia, including the GIANT Consortium (GIANT), UKB, International Consortium of Blood Pressure-Genome Wide Association Studies (ICBP), the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC), the CKDGen Consortium and Global Lipids Genetics Consortium (GLGC). The 19 metabolic traits including the following: body mass index (BMI; $N = 806{,}834$) and waist hip ratio (WHR; $N = 697{,}734$) for anthropometric phenotypes [57]; systolic blood pressure (SBP; $N = 757{,}601$), diastolic blood pressure (DBP; $N = 757{,}601$) and pulse pressure (PP; $N = 757{,}601$) for blood pressure phenotypes [58]; fasting insulin (FI; $N = 151{,}013$), fasting glucose (FG; $N = 200{,}622$), two-hour glucose (2hGlu; $N = 63{,}396$) and glycated hemoglobin levels (HbA1c; $N = 146{,}806$) for glycemic phenotypes [59]; low-density-lipoprotein cholesterol (LDL-C; $N = 1{,}231{,}289$), high-density-lipoprotein cholesterol (HDL-C; $N = 1{,}244{,}580$), triglyceride (TG; $N = 1{,}253{,}277$) and total cholesterol (TC; $N = 1{,}320{,}016$) for lipidemic phenotypes [60]; alanine aminotransferase (ALT; $N = 437{,}267$), alkaline phosphatase (ALP; $N = 437{,}438$) and γ-glutamyl transferase (GGT; $N = 437{,}194$) for liver-related phenotypes [61]; estimated glomerular filtration rate (eGFR; $N = 1{,}004{,}040$) [62] and uric acid (UA; $N = 288{,}649$) [63] for kidney-related phenotypes; and C-reactive protein (CRP; $N = 575{,}531$) [64] as an inflammatory biomarker. The detailed sources of data used in our analyses are listed in Table 1.

Summary-data-based Mendelian randomization analysis

We conducted SMR analysis to integrate summary statistics from GWAS and pQTL studies and detect causal associations between each circulating protein and multiple complex traits, including CVDs and metabolic traits. SMR is an analytical method that combines GWAS data with QTL data to identify quantitative traits with potential causal effects on diseases. For MR, the genetic variant (or multiple variants) used as an IV for a risk factor must meet the following conditions [11]: (i) robustly associated with the exposure phenotype under investigation (relevance assumption); (ii) not associated with any confounding factors (independence assumption); and (iii) influence the outcome solely through the risk factor and not through any direct causal pathway (exclusion restriction assumption). For each plasma protein, we selected the top SNP with the strongest association signal from the cis-pQTL study as the genetic instrument. The odds ratios (ORs) or beta coefficients, along with their respective confidence intervals (CIs), were calculated to quantify the associations between plasma protein levels and the outcomes. These associations were scaled to a one standard deviation (SD) increase in genetically inferred plasma protein levels. The HEIDI test was employed as an instrument to distinguish proteins that were associated with the risk of CVDs or metabolic traits due to genetic variant sharing, rather than genetic linkage. A P value < 0.01 in HEIDI test was considered likely caused by pleiotropy and thus removed from the further analyses. We applied a threshold of $P_{\text{multi}} < 0.05$ as suggestive evidence of statistical significance in SMR using multi-SNPs. To further account for the multiple tests across proteins with cis-SNPs, we established an FDR corrected P-value threshold of < 0.05 as evidence to determine the significant association between the proteins and the risk of diseases or

Table 1 GWAS data sources for cardiovascular diseases and metabolic traits

Category	Phenotype	Abbreviation	PMID	Source	Population	N	N case	N control
Cardiovascular diseases	Atrial fibrillation	AF	30061737	HUNT, UKB, deCODE, MGI, DiscovEHR, the AFGen Consortium	European	1,030,836	60,620	970,216
	Coronary artery disease	CAD	36474045	CDVKP	European	1,165,690	181,522	984,168
	Venous thromboembolism	VTE	36658437	deCODE; FinnGen; CHB-CVDC; DBDS; UKB; Inter-Mountain Healthcare	European	1,500,861	81,190	1,419,671
	Heart failure	HF	31919418	HERMES	European	977,323	47,309	930,014
	Peripheral artery disease	PAD	34601942	UKB	European	511,634	12,086	499,548
	Stroke	Stroke	36180795	FinnGen; UKB; PSI	European	1,308,460	73,652	1,234,808
Anthropometric phenotypes	Body mass index	BMI	30124842	GIANT	European	806,834	NA	NA
	Waist hip ratio	WHR	30124842	GIANT	European	697,734	NA	NA
Blood pressure	Systolic blood pressure	SBP	30224653	UKB, ICBP	European	757,601	NA	NA
	Diastolic blood pressure	DBP	30224653	UKB, ICBP	European	757,601	NA	NA
	Pulse pressure	PP	30224653	UKB, ICBP	European	757,601	NA	NA
Glycemic phenotypes	Fasting insulin	FI	34059833	MAGIC	European	151,013	NA	NA
	Fasting glucose	FG	34059833	MAGIC	European	200,622	NA	NA
	Two-hour glucose	2hGlu	34059833	MAGIC	European	63,396	NA	NA
	Glycated hemoglobin levels	HbA1c	34059833	MAGIC	European	146,806	NA	NA
Lipidemic phenotypes	Low-density-lipoprotein cholesterol	LDL-C	34887591	GLGC	European	1,231,289	NA	NA
	High-density-lipoprotein cholesterol	HDL-C	34887591	GLGC	European	1,244,580	NA	NA
	Triglyceride	TG	34887591	GLGC	European	1,253,277	NA	NA
	Total cholesterol	TC	34887591	GLGC	European	1,320,016	NA	NA
Liver-related phenotypes	Alanine aminotransferase	ALT	33972514	UKB	European	437,267	NA	NA
	Alkaline phosphatase	ALP	33972514	UKB	European	437,438	NA	NA
	γ-glutamyl transferase	GGT	33972514	UKB	European	437,194	NA	NA

Table 1 (Continued)

Category	Phenotype	Abbreviation	PMID	Source	Population	N	N case	N control
Kidney-related phenotypes	Estimated glomerular filtration rate	eGFR	34272381	CKDGen Consortium, UKB	European	1,004,040	NA	NA
	Uric acid	UA	31578528	UKB	European	288,649	NA	NA
Inflammatory biomarker	C-reactive protein	CRP	35459240	UKB	European	575,531	NA	NA

UK Biobank Study, *UKB* deCODE Health Study, *deCODE* the Genetic Investigation of ANthropometric Traits consortium, *GIANT* the Meta-Analyses of Glucose and Insulin-related traits Consortium, *MAGIC* Global Lipids Genetics Consortium, *GLGC* International Consortium of Blood Pressure-Genome Wide Association Studies, *ICBP* the Nord-Trøndelag Health Study, *HUNT* the Michigan Genomics Initiative, *MGI* the Heart Failure Molecular Epidemiology for Therapeutic Targets, *HERMES* the Dutch Parelsnoer Initiative Cerebrovascular Disease Study Group, *PSI* the FinnGen Consortium, *FinnGen* Copenhagen Hospital Biobank Cardiovascular Disease Cohort, *CHB-CVDC* the Danish Blood Donor Study, DBDS

metabolic traits, which helps control the possibility of false rejection of the null hypothesis and corrects for errors when conducting multiple comparisons. For the discovery study, we performed SMR analysis using GWAS data and pQTL summary data from the UKB-PPP. Replication was performed using SMR analysis of cis-pQTL summary data from deCODE and disease GWAS data.

Colocalization analysis

We performed colocalization analysis using 'coloc' R package [16] to assess whether the associations between proteins with CVDs or metabolic traits were driven by a shared causal variant rather than by linkage disequilibrium. The analysis assessed the support for the following five exclusive hypotheses: (1) no association with either trait; (2) association only with trait 1; (3) association only with trait 2; (4) association with both traits, but with distinct causal variants for each; and (5) both traits are associated, driven by a shared causal variant [65]. Posterior probabilities were provided for each hypothesis test (H0, H1, H2, H3, and H4). In our study, we established the prior probabilities that a SNP is associated only with trait 1 (p1) at 1×10^{-4}; the probability of the SNP being associated only with trait 2 (p2) at 1×10^{-4}; and the probability of the SNP being associated with both traits (p12) at 1×10^{-5}. Colocalization was considered to have high support if the posterior probability for shared causal variants (PPH4) was ≥ 0.8. Medium support was defined as $0.5 < \text{PPH4} < 0.8$.

Sensitivity analysis

Three different MR methods random effects IVW, MR Egger and weighted median were applied to fix heterogeneity and horizontal pleiotropy of the variants [18]. Three criteria were applied to determine eligible SNPs: (i) SNPs related to exposure with a significant threshold of $P < 5 \times 10^{-8}$; (ii) The PLINK clumping method was used to identify independent SNPs. SNPs were considered independent if the linkage disequilibrium (LD) threshold was $r^2 < 0.01$ and the distance between SNPs was greater than 5000 kb; (iii) The F-statistics were calculated to verify the strength of each SNP. Only SNPs with F-statistics were greater than ten were considered strong enough to mitigate potential bias.

Phenome-wide association study

We used PheWAS to profile the possible side effects and indications of candidate proteins after their development into drugs. The PheWAS approach has been used to investigate the association between exposure to a set of genetic variants and thousands of phenotypes [66]. GWAS data of diseases from FinnGen were conducted using the Scalable and Accurate Implementation of Generalised Mixed Model (SAIGE) to effectively address imbalances in case–control ratios [67]. A total of 771 phenotypes with more than 500 cases were selected for phenome-MR analysis. To evaluate the impact of variations in the levels of candidate proteins, we calculated ORs and 95% CIs for each of the 771 phenotypes. A *P*-value threshold of 0.05 was set to identify significant associations, and the P_{HEIDI} criterion ($P_{\text{HEIDI}} > 0.01$) was applied to exclude associations showing significant heterogeneity.

Evaluation of druggable targets

We explore the druggability of identified proteins using DrugBank database, the Drug Gene Interaction Database (DGIdb; https://www.dgidb.org/) and Therapeutic Target Database (TTD; https://idrblab.org/ttd/). DrugBank combines detailed drug with comprehensive drug target information, making it a reliable database for drug analysis [20]. DGIdb is an open-source search engine for drug-gene interactions and the druggable genome [21], and TTD systematically assesses targets via established druggability characteristics [22]. We basically divided the identified protein targets into three categories for discussion: (1) Approved drugs or those in clinical trials for treating CVDs or metabolic diseases; (2) Drugs targeted for diseases other than cardiovascular or metabolic diseases; (3) Other potential targets for future drug development.

Pathway enrichment

Information on genomes, biological pathways, and molecular functions was obtained through the Database for Annotation, Visualization, and Integrated Discovery (DAVID) [19]. We set the KEGG pathway enrichment threshold on Expression Analysis Systematic Explorer (EASE) score ≤ 0.1 and gene count ≥ 2 to ensure accurate results.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-026-03962-x.

Additional file 1. SMR analysis of 6 kinds of CVDs and the cis-SNP on plasma proteins from UKB.

Additional file 2. SMR analysis of 6 kinds of CVDs and the cis-SNP on plasma proteins from deCODE.

Additional file 3. Sensitivity analysis of 24 traits and the cis-SNP for 49 CVD-related proteins in the UK Biobank.

Additional file 4. Sensitivity analysis of 24 traits and the cis-SNP for 49 CVD-related proteins in deCODE.

Additional file 5. Colocalization analysis on 49 CVD-related proteins in UKB and deCODE with 6 kinds of CVDs.

Additional file 6. SMR analysis of 19 metabolic phenotypes and the cis-SNP on plasma proteins from UKB.

Additional file 7. SMR analysis of 19 metabolic phenotypes and the cis-SNP on plasma proteins from deCODE.

Additional file 8. Colocalization analysis on 35 proteins related to CVDs and metabolic traits in UKB and deCODE with 19 metabolic phenotypes.

Additional file 9. Pathway enrichment for CVD-related proteins.

Additional file 10. PheWAS data source in FinnGen.

Additional file 11. SMR analysis across 771 phenotypes using FinnGen PheWAS data and cis-SNPs for candidate plasma proteins.

Additional file 12. SMR analysis of UKB excluded GWAS data and the cis-SNP for plasma proteins.

Acknowledgements
We acknowledge the participants and investigators of the studies in this article.

Peer review information
Tim Sands was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions
R.Z., J.Q., X.Z., and Y.F. conceptualized and supervised this project. R.Z. and J.Q. performed the statistical analysis and assisted with the interpretation of results. R.Z., X.Y., L.C., W.H. and P.Z. drafted and revised the manuscript. S.P., Y.Y., Z.X. and Y.F. provided expertise in cardiovascular biology and GWAS summary statistics. All authors discussed the results and commented on the paper.

Funding
This study was supported by Shenzhen Science and Technology Program, Shenzhen, China (Grant No. GJHZ20240218111401002); National Science and Technology Major Project (Grant No. 2023ZD0505902) (To Jin-Song Bian); Natural Science Foundation of China Excellent Young Scientists Fund (Overseas) (Grant No. K241141101); Guangdong Basic and Applied Basic Research Foundation for Distinguished Young Scholars (Grant No. 2024B1515020047); Shenzhen Basic Research General Projects of Shenzhen Science and Technology Innovation Commission (Grant No.

JCYJ20230807093514029; Grant No. JCY20250604144355075); National Natural Science Foundation of China (Grant No. 82470452); Natural Science Foundation of Xinjiang Uygur Autonomous Region (Grant No. 2024D01D15); Shenzhen Medical Research Funds (Grant No. B2502036) (To Y.F.); National Natural Science Foundation of China (Grant no. 82260073); Tianshan Talent Cultivation Program Project of Xinjiang Uygur Autonomous Region (Grant no. 2022TSYCLJ0028) (To Y.Y.); BEIJING medical award foundation (Grant No. YXJL-2025–0106-0190) (To X.Z.); Cancer Research UK (Grant No. C59392/A25064) (To S.P.); National Natural Science Foundation of China Young Student Basic Research Program (PhD candidate) (Grant No. 325B2015) (To J.Q.) and Center for Computational Science and Engineering at Southern University of Science and Technology. The funder had no role in the design, implementation, analysis, interpretation of the data, approval of the manuscript, and decision to submit the manuscript for publication.

Data availability

Comprehensive datasets encompassing genome-wide summary statistics for Atrial Fibrillation (AF), Heart Failure (HF), and Stroke can be accessed via the GWAS Catalog (GCST006414, GCST009541, and GCST90104539). AF UKB-left-out summary statistics are available at https://csg.sph.umich.edu/willer/public/afib2018/, and HF summary statistics minus UK Biobank are accessible via the Cardiovascular Disease Knowledge Portal at https://cvd.hugeamp.org/dinspector.html?dataset=GWAS_HERMES_eu&phenotype=HF. Detailed summary statistics for Coronary Artery Disease (CAD) and Peripheral Arterial Disease (PAD) are available on the Cardiovascular Disease Knowledge Portal (CVDKP), accessible at: https://cvd.hugeamp.org/datasets.html. GWAS statistics on Venous Thromboembolism (VTE) are available through the deCODE genetics at https://www.decode.com/summarydata/. Genome-wide summary statistics for Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Pulse Pressure (PP), Alanine aminotransferase (ALT), Alkaline Phosphatase (ALP), γ-Glutamyl Transferase (GGT), Estimated Glomerular Filtration Rate (eGFR), and Uric Acid (UA) can be retrieved from the GWAS Catalog (GCST006624, GCST006630, GCST006629, GCST90013405, GCST90013406, GCST90013407, GCST90103634, and GCST008971) and the GWAS data on C-Reactive Protein (CRP) is also available at the GWAS catalogue under accession code GCST90029070 (https://www.ebi.ac.uk/gwas/studies/GCST90029070).
GWAS summary statistics on Body Mass Index (BMI) and Waist Hip Ratio (WHR) are obtainable from the Genetic Investigation of Anthropometric Traits consortium (GIANT) at https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files. Additionally, genome-wide summary statistics for glycemic phenotypes, including Fasting Insulin (FI), Fasting Glucose (FG), Two-hour Glucose (2hGlu) and Glycated Hemoglobin levels (HbA1c), and lipidemic phenotypes, including Low-Density-Lipoprotein Cholesterol (LDL-C), High-Density-Lipoprotein Cholesterol (HDL-C), Triglyceride (TG) and Total Cholesterol (TC), are obtained from the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC; https://magicinvestigators.org/) and Global Lipids Genetics Consortium (GLGC; http://csg.sph.umich.edu/willer/public/glgc-lipids2021/), respectively. Information on blood-based cis-pQTL, derived from both deCODE and UKB-PPP, can be found at https://www.decode.com/summarydata/ and https://www.synapse.org/#!Synapse:syn51365303, respectively.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Pharmacology, SUSTech Homeostatic Medicine Institute, School of Medicine, Southern University of Science and Technology; Department of Geriatrics, Shenzhen People's Hospital (the First Affiliated Hospital of Southern University of Science and Technology), Shenzhen 518055, Guangdong, China. [2]Department of Endocrinology, Shenzhen People's Hospital (the First Affiliated Hospital of Southern University of Science and Technology), Shenzhen 518020, Guangdong, China. [3]Translational Medicine Collaborative Innovation Center, Shenzhen People's Hospital (the First Affiliated Hospital of Southern University of Science and Technology), Shenzhen 518020, Guangdong, China. [4]Department of Cardiology, the First People's Hospital of Foshan (Foshan Hospital Affiliated to Southern University of Science and Technology), School of Medicine, Southern University of Science and Technology, Foshan 528000, Guangdong, China. [5]Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Headington, Oxford, UK. [6]Department of Cardiology, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi 830000, Xinjiang Uygur Autonomous Region, China. [7]Xinjiang Key Laboratory of Cardiovascular Homeostasis and Regeneration Research, Urumqi 830000, Xinjiang Uygur Autonomous Region, China. [8]Joint Laboratory of Guangdong-Hong Kong Universities for Vascular Homeostasis and Diseases, Shenzhen 518055, Guangdong, China.

Received: 16 May 2024 Accepted: 15 January 2026
Published online: 28 January 2026

References

1. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392:1736–88. https://doi.org/10.1016/S0140-6736(18)32203-7.
2. Yusuf S, Joseph P, Rangarajan S, Islam S, Mente A, Hystad P, et al. Modifiable risk factors, cardiovascular disease and mortality in 155,722 individuals from 21 high-, middle-, and low-income countries. Lancet. 2020;395:795–808. https://doi.org/10.1016/S0140-6736(19)32008-2.

3. Ettehad D, Emdin CA, Kiran A, Anderson SG, Callender T, Emberson J, et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. Lancet. 2016;387:957–67. https://doi.org/10.1016/S0140-6736(15)01225-8.
4. Holman RR, Sourij H, Califf RM. Cardiovascular outcome trials of glucose-lowering drugs or strategies in type 2 diabetes. Lancet. 2014;383:2008–17. https://doi.org/10.1016/S0140-6736(14)60794-7.
5. Keene D, Price C, Shun-Shin MJ, Francis DP. Effect on cardiovascular risk of high density lipoprotein targeted drug treatments niacin, fibrates, and CETP inhibitors: meta-analysis of randomised controlled trials including 117 411 patients. BMJ. 2014;349:g4379. https://doi.org/10.1136/bmj.g4379.
6. Ray KK, Troquay RPT, Visseren FLJ, Leiter LA, Wright RS, Vikarunnessa S, et al. Long-term efficacy and safety of inclisiran in patients with high cardiovascular risk and elevated LDL cholesterol (ORION-3): results from the 4-year open-label extension of the ORION-1 trial. Lancet Diabetes Endocrinol. 2023;11:109–19. https://doi.org/10.1016/S2213-8587(22)00353-9.
7. Chan MY, Efthymios M, Tan SH, Pickering JW, Troughton R, Pemberton C, et al. Prioritizing candidates of post-myocardial infarction heart failure using plasma proteomics and single-cell transcriptomics. Circulation. 2020;142:1408–21. https://doi.org/10.1161/CIRCULATIONAHA.119.045158.
8. Cupido AJ, Reeskamp LF, Hingorani AD, Finan C, Asselbergs FW, Hovingh GK, et al. Joint genetic inhibition of PCSK9 and CETP and the association with coronary artery disease: a factorial Mendelian randomization study. JAMA Cardiol. 2022;7(9):955–64. https://doi.org/10.1001/jamacardio.2022.2333.
9. Bonaca MP, Nault P, Giugliano RP, Keech AC, Pineda AL, Kanevsky E, et al. Low-density lipoprotein cholesterol lowering with evolocumab and outcomes in patients with peripheral artery disease: insights from the FOURIER trial (further cardiovascular outcomes research with PCSK9 inhibition in subjects with elevated risk). Circulation. 2018;137:338–50. https://doi.org/10.1161/CIRCULATIONAHA.117.032235.
10. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. Nat Rev Drug Discov. 2013;12:581–94. https://doi.org/10.1038/nrd4051.
11. Larsson SC, Butterworth AS, Burgess S. Mendelian randomization for cardiovascular diseases: principles and applications. Eur Heart J. 2023;44:4913–24. https://doi.org/10.1093/eurheartj/ehad736.
12. Yuan S, Xu F, Li X, Chen J, Zheng J, Mantzoros CS, et al. Plasma proteins and onset of type 2 diabetes and diabetic complications: proteome-wide Mendelian randomization and colocalization analyses. Cell Rep Med. 2023;4:101174. https://doi.org/10.1016/j.xcrm.2023.101174.
13. Storm CS, Kia DA, Almramhi MM, Bandres-Ciga S, Finan C, Hingorani AD, et al. Finding genetically-supported drug targets for Parkinson's disease using Mendelian randomization of the druggable genome. Nat Commun. 2021;12:7342. https://doi.org/10.1038/s41467-021-26280-1.
14. Kim MS, Song M, Kim B, Shim I, Kim DS, Natarajan P, et al. Prioritization of therapeutic targets for dyslipidemia using integrative multi-omics and multi-trait analysis. Cell Rep Med. 2023;4:101112. https://doi.org/10.1016/j.xcrm.2023.101112.
15. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016;48:481–7. https://doi.org/10.1038/ng.3538.
16. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10:e1004383. https://doi.org/10.1371/journal.pgen.1004383.
17. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol. 1995;57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
18. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. Epidemiology. 2017;28:30. https://doi.org/10.1097/EDE.0000000000000559.
19. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022;50:W216–21. https://doi.org/10.1093/nar/gkac194.
20. Knox C, Wilson M, Klinger CM, Franklin M, Oler E, Wilson A, et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. Nucleic Acids Res. 2023;52:D1265-75. https://doi.org/10.1093/nar/gkad976.
21. Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, et al. Integration of the drug–gene interaction database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Res. 2020;49:D1144-51. https://doi.org/10.1093/nar/gkaa1084.
22. Zhou Y, Zhang Y, Zhao D, Yu X, Shen X, Zhou Y, et al. TTD: Therapeutic target database describing target druggability information. Nucleic Acids Res. 2023;52:D1465–77. https://doi.org/10.1093/nar/gkad751.
23. Miyazawa K, Ito K, Ito M, Zou Z, Kubota M, Nomura S, et al. Cross-ancestry genome-wide analysis of atrial fibrillation unveils disease biology and enables cardioembolic risk prediction. Nat Genet. 2023;55:187–97. https://doi.org/10.1038/s41588-022-01284-9.
24. Nioi P, Sigurdsson A, Thorleifsson G, Helgason H, Agustsdottir AB, Norddahl GL, et al. Variant ASGR1 associated with a reduced risk of coronary artery disease. N Engl J Med. 2016;374:2131–41. https://doi.org/10.1056/NEJMoa1508419.
25. Savić R, Yang J, Koplev S, An MC, Patel PL, O'Brien RN, et al. Integration of transcriptomes of senescent cell models with multi-tissue patient samples reveals reduced COL6A3 as an inducer of senescence. Cell Rep. 2023;42:113371. https://doi.org/10.1016/j.celrep.2023.113371.
26. Scheres LJJ, van Hylckama Vlieg A, Ballieux BEPB, Fauser BCJM, Rosendaal FR, Middeldorp S, et al. Endogenous sex hormones and risk of venous thromboembolism in young women. J Thromb Haemost. 2019;17:1297–304. https://doi.org/10.1111/jth.14474.
27. Folsom AR, Aleksic N, Wang L, Cushman M, Wu KK, White RH. Protein c, antithrombin, and venous thromboembolism incidence. Arterioscler Thromb Vasc Biol. 2002;22:1018–22. https://doi.org/10.1161/01.ATV.0000017470.08363.AB.

28. Clemitson J-R, Dixon RJ, Haines S, Bingham AJ, Patel BR, Hall L, et al. Genetic dissection of a blood pressure quantitative trait locus on rat chromosome 1 and gene expression analysis identifies SPON1 as a novel candidate hypertension gene. Circ Res. 2007;100:992–9. https://doi.org/10.1161/01.RES.0000261961.41889.9c.
29. Jordan J, Lipp A, Tank J, Schröder C, Stoffels M, Franke G, et al. Catechol-O-methyltransferase and blood pressure in humans. Circulation. 2002;106:460–5. https://doi.org/10.1161/01.CIR.0000022844.50161.3B.
30. Jiang C, Wang Y, Yang W, Yang X. New evidence for the effect of type 2 diabetes and glycemic traits on testosterone levels: a two-sample Mendelian randomization study. Front Endocrinol (Lausanne). 2023;14:1238090. https://doi.org/10.3389/fendo.2023.1238090.
31. Son N-H, Basu D, Samovski D, Pietka TA, Peche VS, Willecke F, et al. Endothelial cell CD36 optimizes tissue fatty acid uptake. J Clin Invest. 2018;128:4329–42. https://doi.org/10.1172/JCI99315.
32. Khan SU, Yedlapati SH, Lone AN, Hao Q, Guyatt G, Delvaux N, et al. PCSK9 inhibitors and ezetimibe with or without statin therapy for cardiovascular risk reduction: a systematic review and network meta-analysis. BMJ. 2022;377:e069116. https://doi.org/10.1136/bmj-2021-069116.
33. Dunlay SM, Weston SA, Redfield MM, Killian JM, Roger VL. Tumor necrosis factor-α and mortality in heart failure. Circulation. 2008;118:625–31. https://doi.org/10.1161/CIRCULATIONAHA.107.759191.
34. Sethi JK, Hotamisligil GS. Metabolic messengers: tumour necrosis factor. Nat Metab. 2021;3:1302–12. https://doi.org/10.1038/s42255-021-00470-z.
35. Sapra G, Tham YK, Cemerlang N, Matsumoto A, Kiriazis H, Bernardo BC, et al. The small-molecule BGP-15 protects against heart failure and atrial fibrillation in mice. Nat Commun. 2014;5:5705. https://doi.org/10.1038/ncomms6705.
36. Loh NY, Rosoff DB, Richmond R, Noordam R, Smith GD, Ray D, et al. Bidirectional mendelian randomization highlights causal relationships between circulating INHBC and multiple cardiometabolic diseases and traits. Diabetes. 2024;73:2084–94. https://doi.org/10.2337/db24-0168.
37. Du XY, Zheng BT, Pang Y, Zhang W, Liu M, Xu XL, et al. The potential mechanism of INHBC and CSF1R in diabetic nephropathy. Eur Rev Med Pharmacol Sci. 2020;24:1970–8.
38. Katagiri C, Masuda K, Nomura M, Tanoue K, Fujita S, Yamashita Y, et al. DUSP13B/TMDP inhibits stress-activated MAPKs and suppresses AP-1-dependent gene expression. Mol Cell Biochem. 2011;352:155–62. https://doi.org/10.1007/s11010-011-0749-x.
39. Luo J, Gao Q, Qiu H, Zhang S, Zou W, Wang P, et al. Myogenin regulates DUSP13 to inhibit apoptosis induced by reactive oxygen species. Front Biosci-Landmark. 2024;29:49. https://doi.org/10.31083/j.fbl2902049.
40. Tang L, Zeng W, Lu X, Wang Q-Y, Liu H, Cheng Z-P, et al. Identification of APOH polymorphisms as common genetic risk factors for venous thrombosis in the Chinese population. J Thromb Haemost. 2014;12:1616–25. https://doi.org/10.1111/jth.12679.
41. Billing O, Holmgren Y, Nosek D, Hedman H, Hemmingsson O. LRIG1 is a conserved EGFR regulator involved in melanoma development, survival and treatment resistance. Oncogene. 2021;40:3707–18. https://doi.org/10.1038/s41388-021-01808-3.
42. Cao S, Pan Y, Tang J, Terker AS, Arroyo Ornelas JP, Jin G, et al. EGFR-mediated activation of adipose tissue macrophages promotes obesity and insulin resistance. Nat Commun. 2022;13:4684. https://doi.org/10.1038/s41467-022-32348-3.
43. Herdenberg C, Mutie PM, Billing O, Abdullah A, Strawbridge RJ, Dahlman I, et al. Lrig proteins regulate lipid metabolism via BMP signaling and affect the risk of type 2 diabetes. Commun Biol. 2021;4:90. https://doi.org/10.1038/s42003-020-01613-w.
44. Hu S, Ding Q, Zhang W, Kang M, Ma J, Zhao L. Gut microbial beta-glucuronidase: a vital regulator in female estrogen metabolism. Gut Microbes. 2023;15:2236749. https://doi.org/10.1080/19490976.2023.2236749.
45. Lietzan AD, Simpson JB, Walton WG, Jariwala PB, Xu Y, Boynton MH, et al. Microbial β-glucuronidases drive human periodontal disease etiology. Sci Adv. 2023;9:eadg3390. https://doi.org/10.1126/sciadv.adg3390.
46. Conte F, Sam J-E, Lefeber DJ, Passier R. Metabolic cardiomyopathies and cardiac defects in inherited disorders of carbohydrate metabolism: a systematic review. Int J Mol Sci. 2023;24:8632. https://doi.org/10.3390/ijms24108632.
47. Lind L, Mazidi M, Clarke R, Bennett DA, Zheng R. Measured and genetically predicted protein levels and cardiovascular diseases in UK biobank and China Kadoorie biobank. Nat Cardiovasc Res. 2024;3:1189–98. https://doi.org/10.1038/s44161-024-00545-6.
48. Fan M, Li N, Huang L, Chen C, Dong X, Gao W. Exploring potential drug targets in multiple cardiovascular diseases: a study based on proteome-wide Mendelian randomization and colocalization analysis. Cardiovasc Ther. 2025;2025(1):5711316. https://doi.org/10.1155/cdr/5711316.
49. Nielsen JB, Thorolfsdottir RB, Fritsche LG, Zhou W, Skov MW, Graham SE, et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. Nat Genet. 2018;50:1234–9. https://doi.org/10.1038/s41588-018-0171-3.
50. Shah S, Henry A, Roselli C, Lin H, Sveinbjörnsson G, Fatemifar G, et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. Nat Commun. 2020;11:163. https://doi.org/10.1038/s41467-019-13690-5.
51. Sun BB, Chiou J, Traylor M, Benner C, Hsu Y-H, Richardson TG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature. 2023;622:329–38. https://doi.org/10.1038/s41586-023-06592-6.
52. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet. 2021;53:1712–21. https://doi.org/10.1038/s41588-021-00978-w.
53. Mishra A, Malik R, Hachiya T, Jürgenson T, Namba S, Posner DC, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. Nature. 2022;611:115–23. https://doi.org/10.1038/s41586-022-05165-3.
54. Ghouse J, Tragante V, Ahlberg G, Rand SA, Jespersen JB, Leinøe EB, et al. Genome-wide meta-analysis identifies 93 risk loci and enables risk prediction equivalent to monogenic forms of venous thromboembolism. Nat Genet. 2023;55:399–409. https://doi.org/10.1038/s41588-022-01286-7.

55. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. Nat Genet. 2022;54:1803–15. https://doi.org/10.1038/s41588-022-01233-6.
56. van Zuydam NR, Stiby A, Abdalla M, Austin E, Dahlström EH, McLachlan S, et al. Genome-wide association study of peripheral artery disease. Circ Genom Precis Med. 2021;14:e002862. https://doi.org/10.1161/CIRCGEN.119.002862.
57. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Hum Mol Genet. 2018;27:3641–9. https://doi.org/10.1093/hmg/ddy271.
58. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet. 2018;50:1412–25. https://doi.org/10.1038/s41588-018-0205-x.
59. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan J, et al. The trans-ancestral genomic architecture of glycemic traits. Nat Genet. 2021;53:840–60. https://doi.org/10.1038/s41588-021-00852-9.
60. Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJ, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. Nature. 2021;600:675–9. https://doi.org/10.1038/s41586-021-04064-3.
61. Pazoki R, Vujkovic M, Elliott J, Evangelou E, Gill D, Ghanbari M, et al. Genetic analysis in European ancestry individuals identifies 517 loci associated with liver enzymes. Nat Commun. 2021;12:2579. https://doi.org/10.1038/s41467-021-22338-2.
62. Stanzick KJ, Li Y, Schlosser P, Gorski M, Wuttke M, Thomas LF, et al. <article-title update="added">Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. Nat Commun. 2021;12(1):4350. https://doi.org/10.1038/s41467-021-24491-0.
63. Tin A, Marten J, Halperin Kuhns VL, Li Y, Wuttke M, Kirsten H, et al. Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. Nat Genet. 2019;51:1459–74. https://doi.org/10.1038/s41588-019-0504-x.
64. Said S, Pazoki R, Karhunen V, Võsa U, Ligthart S, Bodinier B, et al. Genetic analysis of over half a million people characterises C-reactive protein loci. Nat Commun. 2022;13:2198. https://doi.org/10.1038/s41467-022-29650-5.
65. Foley CN, Staley JR, Breen PG, Sun BB, Kirk PDW, Burgess S, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. Nat Commun. 2021;12:764. https://doi.org/10.1038/s41467-020-20885-8.
66. Bastarache L, Denny JC, Roden DM. Phenome-wide association studies. JAMA. 2022;327:75–6. https://doi.org/10.1001/jama.2021.20356.
67. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet. 2018;50:1335–41. https://doi.org/10.1038/s41588-018-0184-y.

Publisher's Note