Therapeutic protein targets for cardiometabolic diseases: Mendelian randomization integrating plasma proteomes with genome-wide association data

Ruixin Zhou
Southern University of Science and Technology
Jun Qiao
Southern University of Science and Technology
Liuyang Cai
Southern University of Science and Technology
Siim Pauklin
siim.pauklin@ndorms.ox.ac.uk
University of Oxford
Yuliang Feng
Southern University of Science and Technology

Research Article

Keywords: Mendelian randomization, cardiovascular diseases, metabolic phenotypes, proteome; drug target

Posted Date: May 30th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4431488/v1

Additional Declarations: No competing interests reported.

Title

Therapeutic protein targets for cardiometabolic diseases: Mendelian randomization integrating plasma proteomes with genome-wide association data

Authors

Ruixin Zhou[1#], Jun Qiao[1#], Liuyang Cai[1], Siim Pauklin[2*], Yuliang Feng[1*]

Affiliations

[1] Department of Pharmacology, School of Medicine, Southern University of Science and Technology, Shenzhen, China.

[2] Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Headington, Oxford, UK.

[#]These authors contributed equally to this work.

*Correspondence to:

Yuliang Feng MD, PhD

Associate Professor

Department of Pharmacology, School of Medicine; Southern University of Science and Technology, Shenzhen, Guangdong, 518055, China.

Tel: +86 (755)-88012564

Email: fengyl@sustech.edu.cn

Siim Pauklin PhD

Associate Professor and CRUK Career Development Fellow

Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Headington, Oxford OX3 7LD, UK.

Tel: +44 (0)-1865226492

Email: siim.pauklin@ndorms.ox.ac.uk

Abstract

Cardiovascular disease (CVD) is the leading cause of death worldwide, and its risk is inseparable from metabolic abnormalities. Through summary-data-based Mendelian randomization and colocalization analysis, we investigated the causal relationships between plasma proteins, 6 cardiovascular diseases, and 19 metabolic phenotypes. We identified 49 proteins genetically associated with CVDs, validated across two platforms, with 35 associated with one or more metabolic phenotypes and six having support of colocalization. These six candidate proteins were classified into three categories based on the utilization of drugs that are currently approved or in clinical trial phases, with PCSK9 already successful in drug development for CVDs and hypercholesterolemia. DUSP13B, LRIG1, APOH, INHBC, and GUSB also showed high drug potential. Further phenome-wide Mendelian randomization analysis indicated no potential side effects from targeting PCSK9 and APOH. This study revealed causal proteins for the onset of cardiovascular diseases and metabolic abnormalities, which contributed to understanding the molecular mechanisms underlying disease pathogenesis and the development of related drugs.

Keywords: Mendelian randomization; cardiovascular diseases; metabolic phenotypes; proteome; drug target.

Teaser

In addition to PCSK9, DUSP13B, LRIG1, APOH, INHBC, and GUSB were also genetically causally associated with metabolic traits and cardiovascular diseases and provide insights into druggable targets.

MAIN TEXT

Introduction

Cardiovascular disease (CVD) encompasses a range of conditions affecting the heart and blood vessels (*1*), which stands as a leading cause of death globally, with a notably rising prevalence in developing countries and regions (*1, 2*). Behavioral risk factors, including elevated blood pressure, high blood sugar, increased cholesterol levels, and obesity, significantly contribute to CVD risk. Addressing these metabolic abnormalities or treating associated metabolic diseases can mitigate the risk of developing cardiovascular conditions. While current clinical treatments for hypertension, dyslipidemia, and diabetes have been applied to CVD, there is a critical need to enhance their effectiveness (*3-7*).

Previous studies have reported that circulating proteins play a crucial role in the onset of cardiovascular diseases (CVD) and may possess therapeutic potential (*8*). As key metabolic products and signaling molecules, plasma proteins participate in physiological and pathological processes within the body. Proteins closely associated with specific diseases can serve as biomarkers for diagnosis or as targets for pharmaceutical intervention. Notably, research has highlighted the involvement of the plasma protein PCSK9 in lipid metabolism, and the introduction of PCSK9 inhibitors has significantly impacted lipid management and reduced cardiovascular risk (*9, 10*). Therefore, targeting plasma proteins offers a promising strategy for developing treatments for both cardiovascular and metabolic diseases.

Increasing evidence suggests that genetic data can potentially be used to identify and prioritize new drug targets and therapeutic indications (*11*). Mendelian randomization (MR), an approach employing genetic variants as instrumental variables to assess the impact of exposure on a specific outcome (*12*), is progressively being utilized to determine the causal links between diseases and associated proteins or genes (*13, 14*), facilitating the identification of druggable targets. Recently, Kim et al. conducted multi-omics and multi-trait analyses through MR to identify 30 potential therapeutic targets for dyslipidemia, showcasing the approach's utility in drug discovery and development (*15*).

Here, we applied a proteome-wide summary data-based MR (SMR) analysis and colocalization analysis, leveraging the top single nucleotide polymorphism (SNP) from protein quantitative trait loci (pQTL) studies as instruments. We analyzed the outcomes from genome-wide association studies (GWAS) on atrial fibrillation (AF), coronary artery disease (CAD), heart failure (HF), venous thromboembolism (VTE), peripheral artery disease (PAD), Stroke, and 19 kinds of metabolic phenotypes to ascertain causal links between cardiometabolic diseases and plasma proteins, highlighting the potential of proteins as unified targets for addressing metabolic dysfunctions and cardiovascular conditions. To further explore the clinical relevance of these candidate proteins, we assessed their druggability through a comprehensive triple-analysis approach: (i) exploring the repurposing of approved drugs and drugs in phases of clinical trials; (ii)

evaluating the druggability of potential target proteins; (iii) revealing phenome-wide consequential effects. The flow diagram of our study is basically displayed in Figure 1.

Results

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

The proteome-wide MR analysis estimated the association between plasma proteins with pQTL information and the risk of six major CVDs. After strict FDR correction ($P_{FDR} < 0.05$) and HEIDI test ($P_{HEIDI} > 0.01$) for multiple testing, SMR analysis identified 189 proteins that showed causal relationships with the risk of CVDs in the discovery study (Figure 2a). In the combined analysis with replication study data, genetically predicted levels of 52 proteins were significantly associated with CVDs (Figure 2b, Table S1, S2). Consistent directional associations for 49 proteins with CVDs across both studies were observed, excluding S100 calcium-binding protein A16 (S100A16), angiopoietin-like 4 (ANGPTL4) and Fc gamma receptor II B (FCGR2B), within which 4 proteins causally related to two different CVDs. Per SD increase in genetically predicted levels of protein, the odds ratio of cardiovascular diseases 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), proprotein convertase subtilisin/kexin type 9 (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.

Among the 49 unique SMR-identified CVDs-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) (Figure 2c, Table S5). 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, which were 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 14A (PPP1R14A), glycoprotein 6 (GP6) and chymotrypsin-like elastase family member 2A (CELA2A; PPH4 $\geq$ 0.8 in discovery study but 0.8> PPH4 $\geq$ 0.5 in replication). We also found that PCSK9, hepatocyte growth factor activator (HGFAC), and inhibin beta C chain (INHBC) had high support of genetic colocalization with CAD, and coagulation Factor XI (F11) and scavenger receptor class A member 5 (SCARA5) with stroke. The only protein that had strong

supportive evidence with PAD was PCSK9, which can also be found in CAD. Besides, leucine-rich repeats and immunoglobulin-like domains 1 (LRIG1), apolipoprotein H (APOH), thrombospondin 2 (THBS2), and F2 that had medium support with one of CVDs were also used in further analysis.

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

The relationships between 49 CVDs-related plasma proteins and 19 major phenotypes in 7 kinds of diverse metabolic traits were analyzed and displayed in the SMR method (Table S3, S4). Among the 49 unique CVDs-related proteins, 35 proteins were causally associated with at least one metabolism-related trait after adjusting for multiple testing ($P_{FDR} < 0.05$) and HEIDI test ($P_{HEIDI} > 0.01$). No CVDs-related proteins survived SMR analysis with two-hour glucose (2hGlu) and fasting insulin (FI) as outcomes under multiple constraints. The directions of the associations between proteins and metabolic traits were shown in Figure 3a overall. For instance, per SD increase, the changes in systolic blood pressure (SBP) ranged from -1.30 (95% CI, -1.57 to -1.05) for NAD kinase (NADK) to 0.58 (95% CI, 0.31 to 0.85) for tryptophanyl-tRNA synthetase 1 (WARS1). In the four overlapped CVDs-related proteins, genetically predicted levels of DUSP13B had an inverse association with CAD, body mass index (BMI), and waist hip ratio (WHR) but a positive association with AF and high-density-lipoprotein (HDL). Levels of PCSK9 were significantly associated with inverse levels of HDL and WHR. However, while elevated levels of F2 were associated with lower levels of low-density-lipoprotein (LDL), they also increased the risk of VTE and stroke. The SMR results for ASGR1 and metabolic phenotypes were not replicable in the deCODE Health study.

Fifteen proteins were supported by colocalization analysis with at least one metabolic phenotype after overlapping the replication study (Figure 3b, Table S6). 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 other five 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).

Prediction of potential druggable targets

Integrating the colocalization analysis results, we identified six plasma proteins that had colocalization evidence with both cardiovascular diseases and metabolic phenotypes, including LRIG1, INHBC, GUSB, APOH, DUSP13B, and PCSK9 (Figure 4). It is noticeable that decreased levels of PCSK9 were genetically associated with a reduced risk of CAD and PAD, and with increased HDL levels in vivo, with strong evidence of colocalization. Meanwhile, increased levels of APOH were causally associated with a decreased risk of CAD, lower levels of CRP, and reductions in two liver function-related enzymes: GGT and ALT. In addition, APOH was enriched in multiple pathways such as the cholesterol metabolism and regulation of leukocyte chemotaxis, following pathway enrichment analysis of 35 causal cardiometabolic proteins. Similarly, GUSB was enriched in metabolism of carbohydrates and Neutrophil degranulation (Table S8). Increased GUSB levels were correlated with a lower risk of AF and TG but with higher levels of ALT and GGT. Reduced INHBC levels may lead to a lower risk of CAD, reduced levels of ALP and FG, but elevated levels of eGFR. While high levels of LRIG1 are associated with reduced risk of AF and obesity, they are also highly associated with increased blood pressure. Besides, increased levels of DUSP13B protein were associated with increased HDL levels and higher AF risk.

To evaluate the drug development and prioritize the druggability of the 6 protein targets, we comprehensively searched the ChEMBL database (*16*), the Drug Gene Interaction Database (DGIdb) (*17*), and Therapeutic Target Database (TTD) (*18*) and classified the proteins into three categories. PCSK9 was classified into category 1 whose targeted-drugs has been approved or in clinical trials to treat CVDs or metabolic diseases. Actually, PCSK9 inhibitors have been approved for lowering cholesterol levels in familial hypercholesterolemia (FH) and are also explored in clinical trials for their potential to provide cardiovascular benefits to patients at risk of CVDs. INHBC, GUSB, and APOH were classified into category 2, which were targeted for diseases other than cardiovascular and metabolic diseases. In recent clinical trials, INHBC has been explored as a target for ovarian cancer treatment. Concurrently, GUSB has achieved success in treating mucopolysaccharidosis and periodontal disease, serving as an effective enzyme replacement therapy. Additionally, APOH as a therapeutic target was tried for Hughes syndrome (also known as antiphospholipid syndrome), although its efficacy requires

further evaluation. DUSP13B and LRIG1 were category 3 that acted as druggable potential targets.

PheWAS exploring the possible indications and adverse effects of targeted proteins

To explore the possible indications and unwanted effects of the 6 proteins, a Phenome-wide association study (PheWAS) was conducted across phenotypes (783 phenotype data from Lee UK Biobank) with at least 500 cases. Candidate proteins all passed significant correction and heterogeneity testing in phenome-wide MR, except for GUSB (Figure 4, Table S7). PCSK9 was one of the significant targets supported by PheWAS. In addition to CAD and PAD, the onset risks of multiple cardiovascular diseases are positively associated with PCSK9 levels, and no noticeable adverse effect was identified (no traits with the negative beta). The utilization of drugs that increase levels of APOH may reduce the risk of hypercholesterolemia (OR [95% CI]: 0.93 [0.90, 0.96]) and gout (OR [95% CI]: 0.84 [0.76, 0.92]). The negative correlation between LRIG1 and atrial fibrillation and flutter was supported by the PheWAS analysis (OR [95% CI]: 0.93 [0.90, 0.96]). Reduced levels of INHBC were causally associated with lower risks of CAD, myocardial infarction and gout but with side effects of asthma (OR [95% CI]: 0.94 [0.91, 0.97]) at an FDR <0.05 threshold level. In addition to being associated with high levels of HDL, the abundance of DUSP13B was inversely associated with multiple diseases, including hypothyroidism (OR [95% CI]: 0.83 [0.76, 0.92]), cataract (OR [95% CI]: 0.83 [0.77, 0.91]), hypertension (OR [95% CI]: 0.91 [0.87, 0.96]), angina pectoris (OR [95% CI]: 0.83 [0.76, 0.91]), coronary atherosclerosis (OR [95% CI]: 0.84 [0.77, 0.91])and ischemic heart disease (IHD, also known as CAD; OR [95% CI]: 0.85 [0.80, 0.91]), whereas the abundance was positively associated with AF. Interestingly, 80% of PCSK9-associated phenotypes overlapped with DUSP13B-associated phenotypes under the threshold of FDR <0.05.

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 improved 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 cardiovascular disease, of which 35 proteins were also associated with diverse metabolic traits. Six proteins had the support of colocalization with cardiovascular diseases 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. We also found that genetically predicted higher levels of circulating PCSK9 and INHBC and lower levels of APOH were associated with an increased risk of CAD, with PCSK9 also associated with a high risk of PAD. Besides, the six proteins are 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 (i.e., PCSK9, APOH, LRIG1, GUSB, DUSP13B, and INHBC).

Our study corroborated some previously identified associations between plasma proteins and cardiovascular disease, such as the associations of AF with IL6R (*19*), CAD with ASGR1 (*20*) and COL6A3 (*21*), and VTE with SHBG (*22*) and PROC (*23*). Some well-studied proteins associated with metabolic diseases or traits were successfully identified, such as SPON1 (*24*), COMT (*25*), SHBG (*26*), and CD36 (*27*). Of note, PCSK9, one of the 6 candidate proteins, has already been approved for use or is under clinical trials for hypercholesterolemia and cardiovascular disease (*9, 10, 28*), 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) (*29, 30*) and insulin-like growth factor 1 receptor (IGF1R) (*31*), 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 are 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, LRIG1, and DUSP13B. Our study was concordant with an epidemiological study, establishing an inverse relationship between the abundance of INHBC and eGFR levels (*32*). The causal correlation of INHBC with eGFR and FG suggests that it may play a role in blood glucose homeostasis and normal renal function, although the mechanism by which it promotes renal cell proliferation in the pathogenesis of diabetic nephropathy has not yet been elucidated (*33*) 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 cardiovascular diseases, especially atherosclerosis and myocardial infarction with chronic inflammatory processes. Likewise, the role of DUSP13B in inflammation regulation, cell proliferation, and signaling transduction (*34*) may suggest its molecular mechanisms in cardiovascular disease. Due to the important role of GUSB in female estrogen metabolism (*35*) and periodontitis development(*36*), 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 (*37*). Another interesting protein, 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 (*38*). Elevated APOH levels due to increased hepatic synthesis are strongly associated with MS changes and vascular disease risk in type 2 diabetes patients,

proposing the idea of APOH as a clinical marker of cardiovascular disease risk (*39*). Possibly caused by pleiotropy assessment, our colocalization analysis did not capture the association of APOH with lipidemic traits. LRIG1 has been reported to inhibit the proliferation of tumor cells (*40*) and the regulation of tissue homeostasis (*41*), which is one of the important pathogenesis mechanisms of cardiovascular diseases. By regulating EGFR signal and its related pathways (*42*), LRIG1 may indirectly affect inflammation and other metabolic regulation-related signaling pathways such as obesity and insulin resistance (*43*). The candidate proteins except PCSK9 are mostly related to the inflammatory response and the regulation of human metabolism, verified by pathway enrichment, and may therefore be appropriately considered as potential targets for the treatment of cardiometabolic disease.

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, we used large sample size GWASs 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. Third, 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. Lastly, some approved drug targets for abnormal conditions are 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 cardiovascular disease, and the SHBG-targeted drug LISINOPRIL, which is used in cardiovascular disease (including heart failure, myocardial infarction, arterial disease, stroke and hypertrophy), hypercholesterolemia, hypertension, non-alcoholic fat liver and atherosclerosis.

In conclusion, this study revealed 35 causal proteins for the onset of cardiometabolic diseases and provided 6 promising previously unknown targets for drug development, including PCSK9, INHBC, APOH, GUSB, LRIG1 and DUSP13B, which suggest the roles of inflammation and cell proliferation in cardiometabolic progression. Drug repurposing targeting INHBC, APOH and GUSB need to be verified in future trials.

Materials and Methods

Study design

This study consisted of two parts: a proteome-wide summary data-based MR (SMR) analysis that used single-nucleotide polymorphisms (SNPs) as instrument variables 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 protein quantitative trait loci (pQTL) studies: the UK Biobank Pharma Proteomics Project (UKB-PPP) (*44*) and the deCODE Health study (*45*). UKB-PPP conducted proteomic analysis on plasma samples from 34,557 participants through the Olink platform and collected data on 2,011 proteins. Likewise, cis data on 1,812 proteins were collected from the deCODE Health study, where 35,559 participants were involved using the SomaScan platform. Cis-single-nucleotide polymorphisms (cis-SNPs), defined as SNPs within 1Mb from the transcription start sites (TSS) of the gene, were selected from protein quantitative trait loci studies that associated with the abundance of plasma proteins at the genome-wide significant level ($P < 5 \times 10^{-8}$) and used as instrumental variables. In the SMR analysis, the UKB-PPP served as the discovery study and the deCODE Health study as the replication study. We presented the results for the overlapping proteins with shared directional relationships in SMR and colocalization analyses from both studies to ensure consistency of results across different proteomic analyzing platforms.

Genome-wide association study (GWAS) data sources

Six major cardiovascular diseases (CVDs) were included in our study, which were atrial fibrillation (AF; N cases = 60,620, N controls = 970,216), heart failure (HF; N cases = 47,309, N controls = 930,014), Stroke (N cases = 73,652, N controls = 1,234,808), venous thromboembolism (VTE; N cases = 81,190, N controls = 1,419,671), coronary artery disease (CAD; N cases = 181,522, N controls = 984,168) and peripheral artery disease (PAD; N cases = 12,086, N controls = 499,548). All participants in the studies are European. In addition to cardiovascular diseases, our study also included summary-level data of GWAS for the 19 metabolic phenotypes across seven kinds of different metabolic traits, which were available in 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 nineteen metabolic phenotypes included body mass index (BMI; N = 806,834) and waist hip ratio (WHR; N = 697,734) for anthropometric traits; systolic blood pressure (SBP; N = 757,601), diastolic blood pressure (DBP; N = 757,601) and pulse pressure (PP; N = 757,601) for blood pressure traits; 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 traits; 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 traits; alanine aminotransferase (ALT; N = 437,267), alkaline phosphatase (ALP; N = 437,438) and γ-glutamyl transferase (GGT; N = 437,194) for liver-related traits;

estimated glomerular filtration rate (eGFR; N = 1,004,040) and uric acid (UA; N = 288,649) for kidney-related traits; and C-reactive protein (CRP; N = 575,531) as an inflammatory biomarker. The detailed sources of data our analyses based on were 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 causality associations between each circulating protein and multiple complex traits, including CVDs and metabolic traits. SMR is a research method that combines GWAS data with quantitative trait loci (QTL) data to identify quantitative traits with potential causal effects on diseases. The genetic variant (or multiple genetic variants) used as an instrumental variable for a risk factor in Mendelian randomization (MR) must meet the following conditions (*12*): (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, the top SNP with the strongest association signal in cis-pQTL study was used as the single genetic instrument in the primary analysis. The odds ratios (ORs) or beta coefficients, along with their respective confidence intervals (CIs), quantifying the associations between plasma protein levels and the outcomes under study were calculated, and the associations were scaled to one standard deviation (SD) elevation in genetically inferred levels of circulating proteins. The heterogeneity in dependent instruments (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 (*46*). The association with *P* value in HEIDI test < 0.01 was considered likely caused by pleiotropy and thus removed from the further analyses. We applied a threshold of $P_{multi} < 0.05$ 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 a false discovery rate (FDR) corrected P-value threshold of < 0.05 as evidence to determine the significant association between the proteins and the risk of diseases or metabolic traits, which helps control the possibility of false rejection of the null hypothesis and corrects for errors when conducting multiple comparisons. We used SMR analysis of disease GWAS data and pQTL summary data in UKBB as discovery study and SMR analysis of cis-pQTL summary data in deCODE and disease GWAS data as replication.

Colocalization analysis

We performed colocalization analysis using 'coloc' R package (*47*) to assess whether identified associations between proteins with CVDs and metabolic traits were consistent with a shared causal variant instead of being driven 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 (*48*). 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 $1 \times 10{-}4$; the probability of the SNP being associated only with trait 2 (p2) at $1 \times 10{-}4$; and the probability of the SNP being associated with both traits (p12) at $1 \times 10{-}5$. Colocalization was considered to have high support if the posterior probability for shared causal variants (PPH4) was ≥0.8. Medium support for colocalization was defined as $0.5 < PPH4 < 0.8$.

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 (*49*). GWASs data of diseases from the UK Biobank were conducted using the Scalable and Accurate Implementation of Generalised Mixed Model (SAIGE) to effectively address imbalances in case-control ratios (*50*). In the Lee UKBB dataset (https://www.leelabsg.org/resources), 783 phenotypes with more than 500 cases were chosen for phenome-MR analysis. The PheCODE schema was employed to define phenotypes for the analysis. We calculated the ORs and 95%CIs to evaluate the impact of variations in the levels of candidate proteins on the 783 phenotypes under investigation. Subsequently, to conduct a power estimation, we adjusted the P-values for multiple comparisons by applying the false discovery rate (FDR) correction, setting a P-value threshold of 0.05, and employed a P_{HEIDI} criterion greater than 0.01 to exclude associations displaying significant heterogeneity.

Evaluation of druggable targets

We explore the druggability of identified proteins using ChEMBL database, the Drug Gene Interaction Database (DGIdb; https://www.dgidb.org/) and Therapeutic Target Database (TTD; https://idrblab.org/ttd/). ChEMBL is a manually curated database of bioactive molecules with drug-like properties, which describes indications and mechanisms of drugs (*16*). DGIdb is an open-source search engine for drug-gene interactions and the druggable genome (*17*), and TTD systematically assesses targets via established druggability characteristics (*18*). We basically divided the identified candidate protein targets into three categories for discussion after integrating the information: 1) drugs that have been approved or in clinical trials to treat CVDs or metabolic diseases; 2) drugs targeted for diseases other than cardiovascular or metabolic diseases; 3) druggable potential targets.

Pathway enrichment

Information on genomes, biological pathways, molecular functions and drugs was obtained through METASCAPE (https://www.metascape.org/), which contains databases such as the Gene Ontology Consortium (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), WIKIPathways (WIKI) and others (*51*). We adjusted the parameters (Min Overlap=3, *P* Value Cutoff=0.01, Min Enrichment=1.5) to acquire accurate

pathway enrichment outcomes and Gene Ontology (GO) annotations for proteins associated with both CVDs and metabolic traits.

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* **392**, 1736-1788 (2018).
2. S. Yusuf, P. Joseph, S. Rangarajan, S. Islam, A. Mente, P. Hystad, M. Brauer, V. R. Kutty, R. Gupta, A. Wielgosz, K. F. AlHabib, A. Dans, P. Lopez-Jaramillo, A. Avezum, F. Lanas, A. Oguz, I. M. Kruger, R. Diaz, K. Yusoff, P. Mony, J. Chifamba, K. Yeates, R. Kelishadi, A. Yusufali, R. Khatib, O. Rahman, K. Zatonska, R. Iqbal, L. Wei, H. Bo, A. Rosengren, M. Kaur, V. Mohan, S. A. Lear, K. K. Teo, D. Leong, M. O'Donnell, M. McKee, G. Dagenais, Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. *Lancet* **395**, 795-808 (2020).
3. F. Taylor, M. D. Huffman, A. F. Macedo, T. H. Moore, M. Burke, G. Davey Smith, K. Ward, S. Ebrahim, Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev* **2013**, Cd004816 (2013).
4. D. Ettehad, C. A. Emdin, A. Kiran, S. G. Anderson, T. Callender, J. Emberson, J. Chalmers, A. Rodgers, K. Rahimi, Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. *Lancet* **387**, 957-967 (2016).
5. R. R. Holman, H. Sourij, R. M. Califf, Cardiovascular outcome trials of glucose-lowering drugs or strategies in type 2 diabetes. *Lancet* **383**, 2008-2017 (2014).
6. D. Keene, C. Price, M. J. Shun-Shin, D. P. Francis, 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* **349**, g4379 (2014).
7. K. K. Ray, R. P. T. Troquay, F. L. J. Visseren, L. A. Leiter, R. Scott Wright, S. Vikarunnessa, Z. Talloczy, X. Zang, P. Maheux, A. Lesogor, U. Landmesser, 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* **11**, 109-119 (2023).
8. M. Y. Chan, M. Efthymios, S. H. Tan, J. W. Pickering, R. Troughton, C. Pemberton, H. H. Ho, J. F. Prabath, C. L. Drum, L. H. Ling, W. M. Soo, S. C. Chai, A. Fong, Y. Y. Oon, J. P. Loh, C. H. Lee, R. S. Y. Foo, M. A. Ackers-Johnson, A. Pilbrow, A. M. Richards, Prioritizing Candidates of Post-Myocardial Infarction Heart Failure Using Plasma Proteomics and Single-Cell Transcriptomics. *Circulation* **142**, 1408-1421 (2020).
9. M. P. Bonaca, P. Nault, R. P. Giugliano, A. C. Keech, A. L. Pineda, E. Kanevsky, J. Kuder, S. A. Murphy, J. W. Jukema, B. S. Lewis, L. Tokgozoglu, R. Somaratne, P. S. Sever, T. R. Pedersen, M. S. Sabatine, 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* **137**, 338-350 (2018).
10. A. J. Cupido, L. F. Reeskamp, 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).
11. R. M. Plenge, E. M. Scolnick, D. Altshuler, Validating therapeutic targets through human genetics. *Nat Rev Drug Discov* **12**, 581-594 (2013).
12. S. C. Larsson, A. S. Butterworth, S. Burgess, Mendelian randomization for cardiovascular diseases: principles and applications. *Eur Heart J* **44**, 4913-4924 (2023).
13. S. Yuan, F. Xu, X. Li, J. Chen, J. Zheng, C. S. Mantzoros, S. C. Larsson, Plasma proteins and onset of type 2 diabetes and diabetic complications: Proteome-wide Mendelian randomization and colocalization analyses. *Cell Rep Med* **4**, 101174 (2023).
14. C. S. Storm, D. A. Kia, M. M. Almramhi, S. Bandres-Ciga, C. Finan, A. D. Hingorani, N. W. Wood, Finding genetically-supported drug targets for Parkinson's disease using Mendelian randomization of the druggable genome. *Nat Commun* **12**, 7342 (2021).
15. M. S. Kim, M. Song, B. Kim, I. Shim, D. S. Kim, P. Natarajan, R. Do, H. H. Won, Prioritization of therapeutic targets for dyslipidemia using integrative multi-omics and multi-trait analysis. *Cell Rep Med* **4**, 101112 (2023).
16. B. Zdrazil, E. Felix, F. Hunter, E. J. Manners, J. Blackshaw, S. Corbett, M. de Veij, H. Ioannidis, D. M. Lopez, J. F. Mosquera, M. P. Magarinos, N. Bosc, R. Arcila, T. Kizilören, A. Gaulton, A. P. Bento, M. F. Adasme, P. Monecke, G. A. Landrum, A. R. Leach, The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res* **52**, D1180-d1192 (2024).
17. S. L. Freshour, S. Kiwala, K. C. Cotto, A. C. Coffman, J. F. McMichael, J. J. Song, M. Griffith, O. L. Griffith, A. H. Wagner, Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res* **49**, D1144-d1151 (2021).
18. Y. Zhou, Y. Zhang, D. Zhao, X. Yu, X. Shen, Y. Zhou, S. Wang, Y. Qiu, Y. Chen, F. Zhu, TTD: Therapeutic Target Database describing target druggability information. *Nucleic Acids Res* **52**, D1465-d1477 (2024).
19. K. Miyazawa, K. Ito, M. Ito, Z. Zou, M. Kubota, S. Nomura, H. Matsunaga, S. Koyama, H. Ieki, M. Akiyama, Y. Koike, R. Kurosawa, H. Yoshida, K. Ozaki, Y. Onouchi, A. Takahashi, K. Matsuda, Y. Murakami, H. Aburatani, M. Kubo, Y. Momozawa, C. Terao, S. Oki, H. Akazawa, Y. Kamatani, I. Komuro, Cross-ancestry genome-wide analysis of atrial fibrillation unveils disease biology and enables cardioembolic risk prediction. *Nat Genet* **55**, 187-197 (2023).
20. P. Nioi, A. Sigurdsson, G. Thorleifsson, H. Helgason, A. B. Agustsdottir, G. L. Norddahl, A. Helgadottir, A. Magnusdottir, A. Jonasdottir, S. Gretarsdottir, I. Jonsdottir, V. Steinthorsdottir, T. Rafnar, D. W. Swinkels, T. E. Galesloot, N. Grarup, T. Jørgensen, H. Vestergaard, T. Hansen, T. Lauritzen, A. Linneberg, N. Friedrich, N. T. Krarup, M. Fenger, U. Abildgaard, P. R. Hansen, A. M. Galløe, P. S. Braund, C. P. Nelson, A. S. Hall, M. J. Williams, A. M. van Rij, G. T. Jones, R. S. Patel, A. I. Levey, S. Hayek, S. H. Shah, M. Reilly, G. I. Eyjolfsson, O. Sigurdardottir, I. Olafsson, L. A. Kiemeney, A. A.

Quyyumi, D. J. Rader, W. E. Kraus, N. J. Samani, O. Pedersen, G. Thorgeirsson, G. Masson, H. Holm, D. Gudbjartsson, P. Sulem, U. Thorsteinsdottir, K. Stefansson, Variant ASGR1 Associated with a Reduced Risk of Coronary Artery Disease. *N Engl J Med* **374**, 2131-2141 (2016).
21. R. Savić, J. Yang, S. Koplev, M. C. An, P. L. Patel, R. N. O'Brien, B. N. Dubose, T. Dodatko, E. Rogatsky, K. Sukhavasi, R. Ermel, A. Ruusalepp, S. M. Houten, J. C. Kovacic, A. F. Stewart, C. B. Yohn, E. E. Schadt, R. M. Laberge, J. L. M. Björkegren, Z. Tu, C. Argmann, Integration of transcriptomes of senescent cell models with multi-tissue patient samples reveals reduced COL6A3 as an inducer of senescence. *Cell Rep* **42**, 113371 (2023).
22. L. J. J. Scheres, A. van Hylckama Vlieg, B. Ballieux, B. Fauser, F. R. Rosendaal, S. Middeldorp, S. C. Cannegieter, Endogenous sex hormones and risk of venous thromboembolism in young women. *J Thromb Haemost* **17**, 1297-1304 (2019).
23. A. R. Folsom, N. Aleksic, L. Wang, M. Cushman, K. K. Wu, R. H. White, Protein C, antithrombin, and venous thromboembolism incidence: a prospective population-based study. *Arterioscler Thromb Vasc Biol* **22**, 1018-1022 (2002).
24. J. R. Clemitson, R. J. Dixon, S. Haines, A. J. Bingham, B. R. Patel, L. Hall, M. Lo, J. Sassard, F. J. Charchar, N. J. Samani, 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* **100**, 992-999 (2007).
25. J. Jordan, A. Lipp, J. Tank, C. Schröder, M. Stoffels, G. Franke, A. Diedrich, G. Arnold, D. S. Goldstein, A. M. Sharma, F. C. Luft, Catechol-o-methyltransferase and blood pressure in humans. *Circulation* **106**, 460-465 (2002).
26. C. Jiang, Y. Wang, W. Yang, X. Yang, New evidence for the effect of type 2 diabetes and glycemic traits on testosterone levels: a two-sample Mendelian randomization study. *Front Endocrinol (Lausanne)* **14**, 1238090 (2023).
27. N. H. Son, D. Basu, D. Samovski, T. A. Pietka, V. S. Peche, F. Willecke, X. Fang, S. Q. Yu, D. Scerbo, H. R. Chang, F. Sun, S. Bagdasarov, K. Drosatos, S. T. Yeh, A. E. Mullick, K. I. Shoghi, N. Gumaste, K. Kim, L. A. Huggins, T. Lhakhang, N. A. Abumrad, I. J. Goldberg, Endothelial cell CD36 optimizes tissue fatty acid uptake. *J Clin Invest* **128**, 4329-4342 (2018).
28. S. U. Khan, S. H. Yedlapati, A. N. Lone, Q. Hao, G. Guyatt, N. Delvaux, G. E. T. Bekkering, P. O. Vandvik, I. B. Riaz, S. Li, B. Aertgeerts, N. Rodondi, PCSK9 inhibitors and ezetimibe with or without statin therapy for cardiovascular risk reduction: a systematic review and network meta-analysis. *Bmj* **377**, e069116 (2022).
29. S. M. Dunlay, S. A. Weston, M. M. Redfield, J. M. Killian, V. L. Roger, Tumor necrosis factor-alpha and mortality in heart failure: a community study. *Circulation* **118**, 625-631 (2008).
30. J. K. Sethi, G. S. Hotamisligil, Metabolic Messengers: tumour necrosis factor. *Nat Metab* **3**, 1302-1312 (2021).
31. G. Sapra, Y. K. Tham, N. Cemerlang, A. Matsumoto, H. Kiriazis, B. C. Bernardo, D. C. Henstridge, J. Y. Ooi, L. Pretorius, E. J. Boey, L. Lim, J. Sadoshima, P. J. Meikle, N. A. Mellet, E. A. Woodcock, S. Marasco, T. Ueyama, X. J. Du, M. A. Febbraio, J. R.

609 McMullen, The small-molecule BGP-15 protects against heart failure and atrial fibrillation
610 in mice. *Nat Commun* **5**, 5705 (2014).
611 32. P. Schlosser, J. Zhang, H. Liu, A. L. Surapaneni, E. P. Rhee, D. E. Arking, B. Yu, E.
612 Boerwinkle, P. A. Welling, N. Chatterjee, K. Susztak, J. Coresh, M. E. Grams,
613 Transcriptome- and proteome-wide association studies nominate determinants of kidney
614 function and damage. *Genome Biol* **24**, 150 (2023).
615 33. X. Y. Du, B. T. Zheng, Y. Pang, W. Zhang, M. Liu, X. L. Xu, S. J. Zhou, The potential
616 mechanism of INHBC and CSF1R in diabetic nephropathy. *Eur Rev Med Pharmacol Sci*
617 **24**, 1970-1978 (2020).
618 34. C. Katagiri, K. Masuda, M. Nomura, K. Tanoue, S. Fujita, Y. Yamashita, R. Katakura, K.
619 Shiiba, E. Nomura, M. Sato, N. Tanuma, H. Shima, DUSP13B/TMDP inhibits stress-
620 activated MAPKs and suppresses AP-1-dependent gene expression. *Mol Cell Biochem*
621 **352**, 155-162 (2011).
622 35. S. Hu, Q. Ding, W. Zhang, M. Kang, J. Ma, L. Zhao, Gut microbial beta-glucuronidase: a
623 vital regulator in female estrogen metabolism. *Gut Microbes* **15**, 2236749 (2023).
624 36. A. D. Lietzan, J. B. Simpson, W. G. Walton, P. B. Jariwala, Y. Xu, M. H. Boynton, J. Liu,
625 M. R. Redinbo, Microbial β-glucuronidases drive human periodontal disease etiology. *Sci*
626 *Adv* **9**, eadg3390 (2023).
627 37. F. Conte, J. E. Sam, D. J. Lefeber, R. Passier, Metabolic Cardiomyopathies and Cardiac
628 Defects in Inherited Disorders of Carbohydrate Metabolism: A Systematic Review. *Int J*
629 *Mol Sci* **24**, (2023).
630 38. L. Tang, W. Zeng, X. Lu, Q. Y. Wang, H. Liu, Z. P. Cheng, Y. Y. Wu, B. Hu, X. R. Jian,
631 T. Guo, H. F. Wang, Y. Hu, Identification of APOH polymorphisms as common genetic
632 risk factors for venous thrombosis in the Chinese population. *J Thromb Haemost* **12**,
633 1616-1625 (2014).
634 39. A. Castro, I. Lázaro, D. M. Selva, E. Céspedes, J. Girona, NúriaPlana, M. Guardiola, A.
635 Cabré, R. Simó, L. Masana, APOH is increased in the plasma and liver of type 2 diabetic
636 patients with metabolic syndrome. *Atherosclerosis* **209**, 201-205 (2010).
637 40. Q. Li, B. Liu, H. P. Chao, Y. Ji, Y. Lu, R. Mehmood, C. Jeter, T. Chen, J. R. Moore, W.
638 Li, C. Liu, K. Rycaj, A. Tracz, J. Kirk, T. Calhoun-Davis, J. Xiong, Q. Deng, J. Huang, B.
639 A. Foster, A. Gokhale, X. Chen, D. G. Tang, Author Correction: LRIG1 is a pleiotropic
640 androgen receptor-regulated feedback tumor suppressor in prostate cancer. *Nat Commun*
641 **11**, 2895 (2020).
642 41. T. Nakamura, J. Hamuro, M. Takaishi, S. Simmons, K. Maruyama, A. Zaffalon, A. J.
643 Bentley, S. Kawasaki, M. Nagata-Takaoka, N. J. Fullwood, S. Itami, S. Sano, M. Ishii, Y.
644 Barrandon, S. Kinoshita, LRIG1 inhibits STAT3-dependent inflammation to maintain
645 corneal homeostasis. *J Clin Invest* **124**, 385-397 (2014).
646 42. O. Billing, Y. Holmgren, D. Nosek, H. Hedman, O. Hemmingsson, LRIG1 is a conserved
647 EGFR regulator involved in melanoma development, survival and treatment resistance.
648 *Oncogene* **40**, 3707-3718 (2021).
649 43. S. Cao, Y. Pan, J. Tang, A. S. Terker, J. P. Arroyo Ornelas, G. N. Jin, Y. Wang, A. Niu,
650 X. Fan, S. Wang, R. C. Harris, M. Z. Zhang, EGFR-mediated activation of adipose tissue
651 macrophages promotes obesity and insulin resistance. *Nat Commun* **13**, 4684 (2022).

44. C. J. Sun B.B., Traylor M., Benner C., Hsu Y.-H., Richardson T.G., Surendran P., Mahajan A., Robins C., Vasquez-Grinnell S.G., et al., Genetic regulation of the human plasma proteome in 54,306 UK Biobank participants. *bioRxiv*. **doi: 10.1101/2022.06.17.496443.**, (2022).
45. E. Ferkingstad, P. Sulem, B. A. Atlason, G. Sveinbjornsson, M. I. Magnusson, E. L. Styrmisdottir, K. Gunnarsdottir, A. Helgason, A. Oddsson, B. V. Halldorsson, B. O. Jensson, 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. Y. Wu, J. Zeng, F. Zhang, Z. Zhu, T. Qi, Z. Zheng, L. R. Lloyd-Jones, R. E. Marioni, N. G. Martin, G. W. Montgomery, I. J. Deary, N. R. Wray, P. M. Visscher, A. F. McRae, J. Yang, Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun* **9**, 918 (2018).
47. C. Giambartolomei, D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, V. Plagnol, Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
48. C. N. Foley, J. R. Staley, P. G. Breen, B. B. Sun, P. D. W. Kirk, S. Burgess, J. M. M. Howson, A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun* **12**, 764 (2021).
49. L. Bastarache, J. C. Denny, D. M. Roden, Phenome-Wide Association Studies. *Jama* **327**, 75-76 (2022).
50. W. Zhou, J. B. Nielsen, L. G. Fritsche, R. Dey, M. E. Gabrielsen, B. N. Wolford, J. LeFaive, P. VandeHaar, S. A. Gagliano, A. Gifford, L. A. Bastarache, W. Q. Wei, J. C. Denny, M. Lin, K. Hveem, H. M. Kang, G. R. Abecasis, C. J. Willer, S. Lee, Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335-1341 (2018).
51. Y. Zhou, B. Zhou, L. Pache, M. Chang, A. H. Khodabakhshi, O. Tanaseichuk, C. Benner, S. K. Chanda, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).

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

Funding: This study was supported by the 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. 24050000763), Shenzhen Pengcheng Peacock Plan, Shenzhen Basic Research General Projects of Shenzhen Science and Technology Innovation Commission (Grant no. JCYJ20230807093514029) (To Y.F.), 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.

Author contributions: R.Z., J.Q., and Y.F. conceptualized and supervised this project. R.Z. and J.Q. performed the main analyses and drafted the manuscript. R.Z. and J.Q. performed the statistical analysis and assisted with the interpretation of results. L.C. drafted the manuscript. S.S. and Y.F. provided expertise in cardiovascular biology and GWAS summary statistics. All authors discussed the results and commented on the paper.

Competing interests: All authors declare that they have no competing interests.

Data and materials availability: Comprehensive datasets encompassing genome-wide summary statistics for Atrial Fibrillation (AF), Heart Failure (HF), and Stroke can be accessed via the GWAS Catalog (GCST90104539, GCST009541, and GCST90104539). Detailed summary statistics for Coronary Artery Disease (CAD) and Peripheral Arterial Disease (PAD) are available on the Cardiovascular Disease Knowledge Portal (CVDKP), accessible at website: 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 UK Biobank: https://www.ebi.ac.uk/gwas/publications/35459240. 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. Besides, 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/downloads/) 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.

Figures and Tables

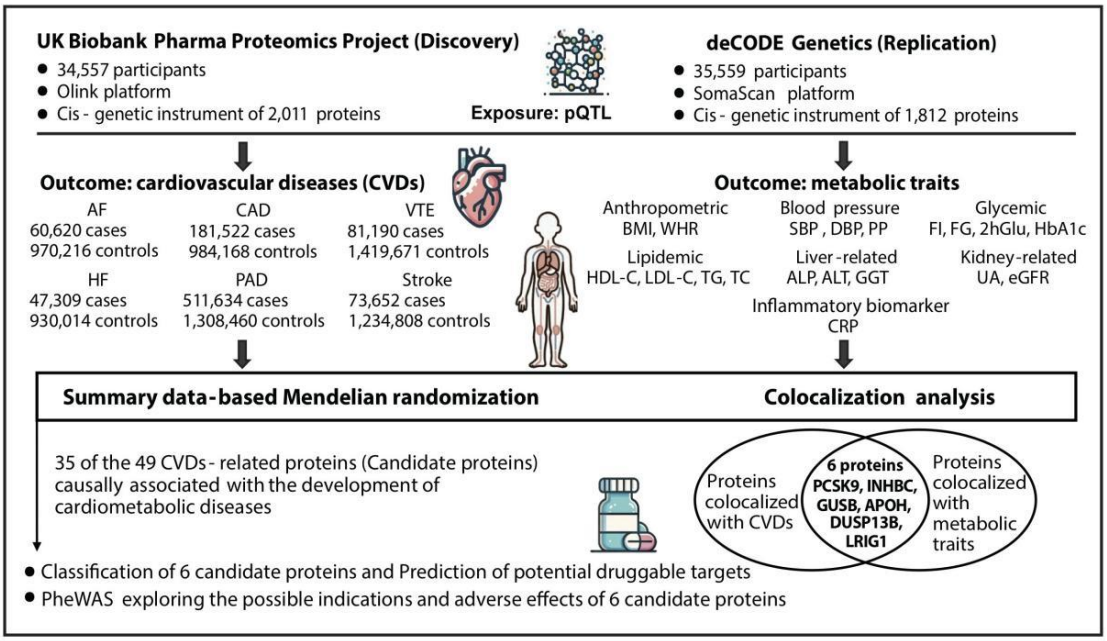


Fig. 1. Overview of this MR study design.

After discovering that the top SNP with the strongest association signal in cis-pQTL study was used as the single genetic instrument, summary-data-based mendelian randomization (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 protein quantitative trait loci (pQTL) studies, including the UK Biobank Pharma Proteomics Project (UKB-PPP; N = 34,557) and the deCODE Health study (N = 35,559). Furthermore, we pursued the exploration of therapeutic targets, and evaluated 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, total cholesterol; ALP, total cholesterol; GGT, total cholesterol; 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.

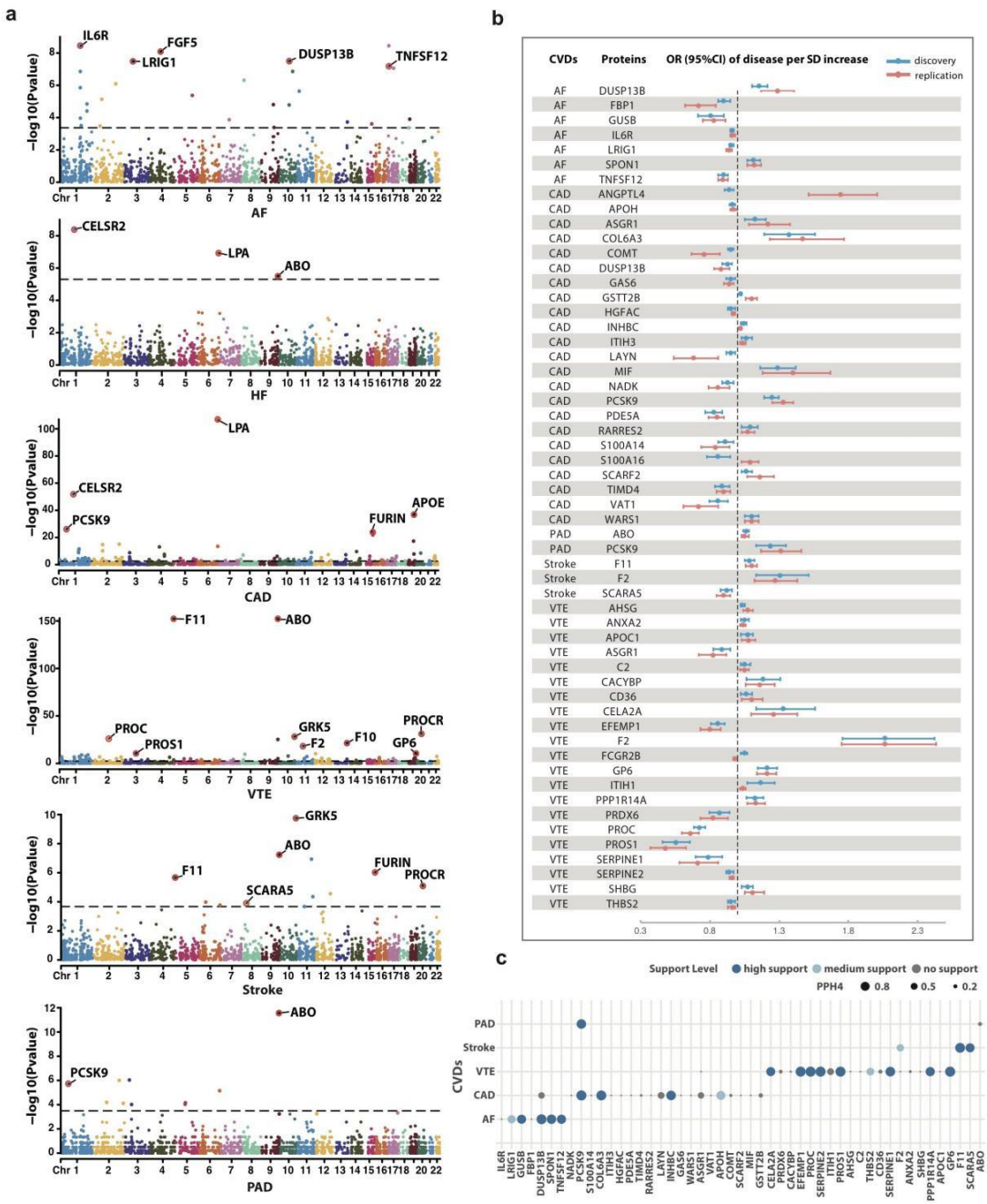


Fig. 2. SMR and colocalization analysis on the associations between plasma proteins and the risk of six kinds of cardiovascular diseases.

(A) Manhattan plot for SMR analysis. Above the dotted line are proteins with corrected *P* value < 0.05 in false discovery rate. **(B)** Forest plot of SMR analysis. OR, odds ratio. **(C)** Colocalization analysis. Full name 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 5A; 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 2A; 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 14A; 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.

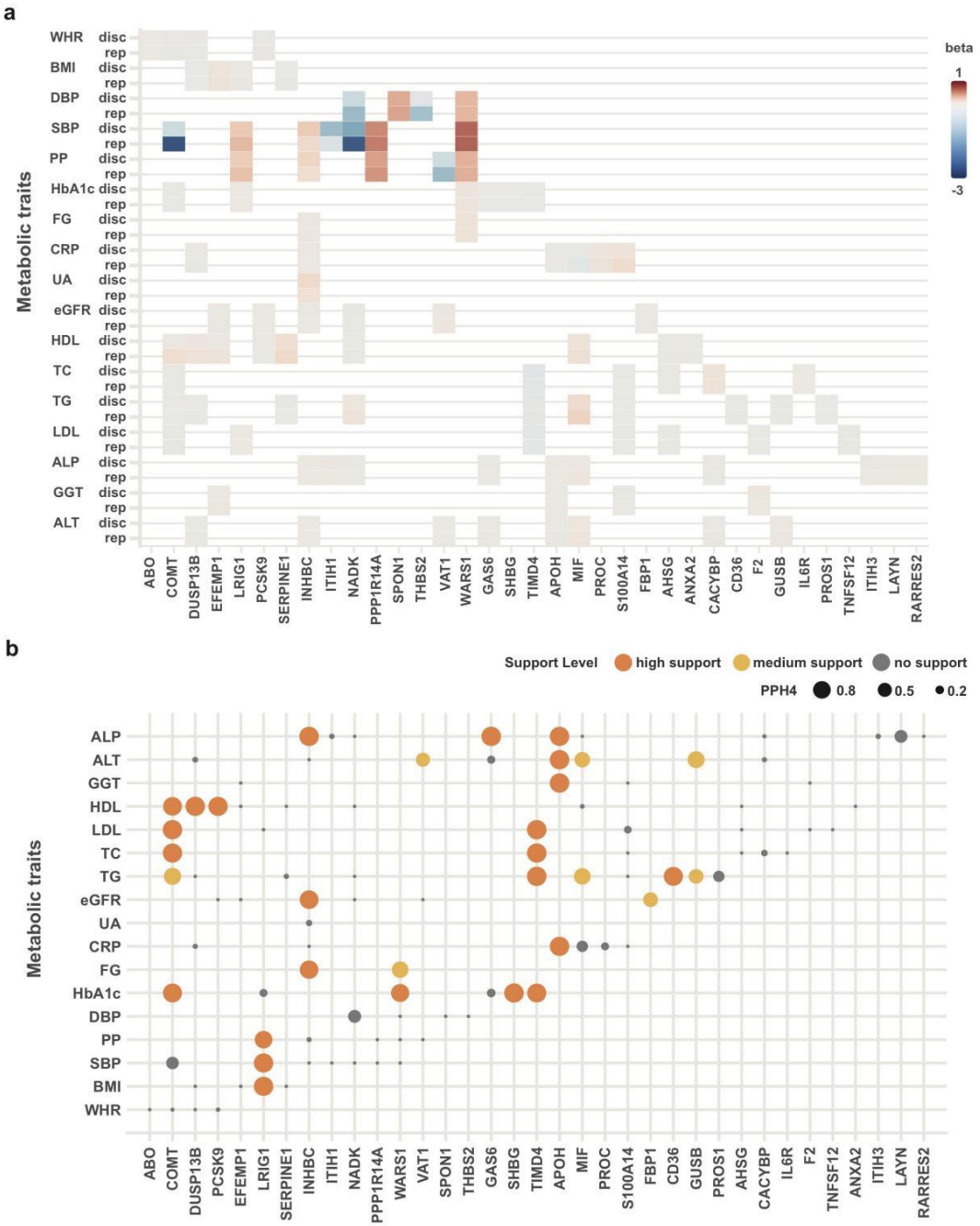
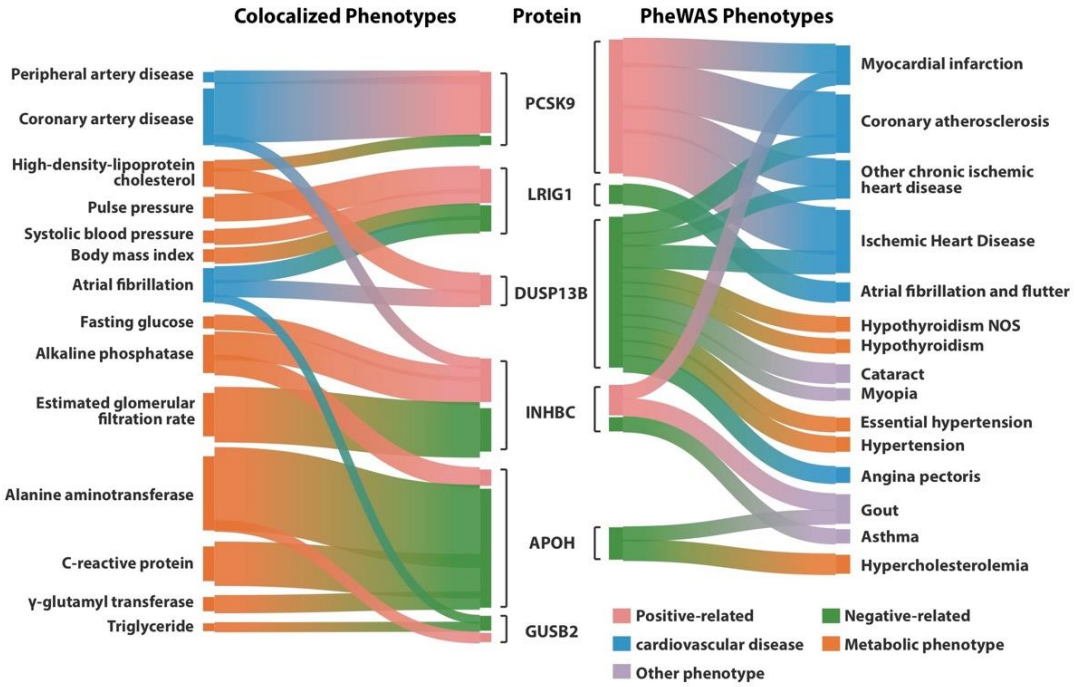


Fig. 3. SMR and colocalization analysis on the associations between plasma proteins and the 17 kinds of metabolic phenotypes.

(A) Heatmap of SMR analysis on the association between CVDs-related proteins and 17 kinds of metabolic phenotypes. **(B)** Colocalization analysis.



796
797 **Fig. 4. Phenotypes significantly associated with candidate proteins.**
798 The left side of the six candidate proteins represents cardiovascular diseases and
799 metabolic phenotypes that with the support of colocalization, and the right side
800 represents the phenotypes that are significantly associated with proteins identified by
801 PheWAS. Thicker lines indicate a higher significance degree.

802

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; InterMountain 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
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

Table 1. GWAS data sources for cardiovascular diseases and metabolic traits.

Note: 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.

Supplementary Materials

Tables S1 to S8 (Table S1. SMR analysis of 6 kinds of CVDs and the cis-SNP on plasma proteins from UKB. Table S2. SMR analysis of 6 kinds of CVDs and the cis-SNP on plasma proteins from deCODE. Table S3. SMR analysis of 19 metabolic phenotypes and the cis-SNP on plasma proteins from UKB. Table S4. SMR analysis of 19 metabolic phenotypes and the cis-SNP on plasma proteins from deCODE. Table S5. Colocalization analysis on 49 CVDs-related proteins in UKB and deCODE with 6 kinds of CVDs. Table S6. Colocalization analysis on 35 proteins related to CVDs and metabolic traits in UKB and deCODE with 19 metabolic phenotypes. Table S7. Sample size and correlated direction of phenotypes significantly associated with candidate proteins. Table S8. Pathway enrichment for CVDs-related proteins.)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableSupplementary.xlsx