nature cardiovascular research

Letter

https://doi.org/10.1038/s44161-024-00545-6

Measured and genetically predicted protein levels and cardiovascular diseases in UK Biobank and China Kadoorie Biobank

Received: 8 January 2024

Accepted: 28 August 2024

Published online: 25 September 2024

Check for updates

Lars Lind[1], Mohsen Mazidi[2], Robert Clarke[2], Derrick A. Bennett[2] & Rui Zheng[1]

Several large-scale studies have measured plasma levels of the proteome in individuals with cardiovascular diseases (CVDs)[1–7]. However, since the majority of such proteins are interrelated[2], it is difficult for observational studies to distinguish which proteins are likely to be of etiological relevance. Here we evaluate whether plasma levels of 2,919 proteins measured in 52,164 UK Biobank participants are associated with incident myocardial infarction, ischemic stroke or heart failure. Of those proteins, 126 were associated with all three CVD outcomes and 118 were associated with at least one CVD in the China Kadoorie Biobank. Mendelian randomization and colocalization analyses indicated that genetically determined levels of 47 and 18 proteins, respectively, were associated with CVDs, including FGF5, PROCR and FURIN. While the majority of protein–CVD observational associations were noncausal, these three proteins showed evidence to support potential causality and are therefore promising targets for drug treatment for CVD outcomes.

Several studies have measured protein plasma levels in individuals with cardiovascular disease (CVD) outcomes[1–7]. However, since the majority of proteins are interrelated[2], it is difficult for observational studies to identify proteins of etiological relevance. One way to address this problem is to use instrumental variables to investigate whether genetic loci regulating plasma levels of individual proteins also are related to different CVD outcomes. Here, we used observational and genetic analyses to assess the associations of up to 2,919 proteins with risk of myocardial infarction (MI), ischemic stroke (IS) or heart failure (HF) or a combination of all such diseases in the UK Biobank (UKB) to discover novel targets for drug treatment. Overall, 636 of those proteins were related to any CVD outcome and 126 were associated with all three CVDs. Of those, 118 were replicated to be associated with at least one CVD outcome in the China Kadoorie Biobank (CKB). Mendelian randomization (MR) and colocalization analyses indicated that genetically determined levels of 47 and 18 proteins, respectively, were associated with CVDs, including fibroblast growth factor 5 (FGF5), protein C receptor (PROCR) and FURIN. While the majority of protein–CVD observational associations were noncausal, these three proteins showed evidence of potential causality and are therefore promising targets for drug treatment for CVD outcomes.

For the observational part, we used plasma levels of 2,919 proteins measured in 52,164 UKB participants. Selected baseline characteristics of this sample are provided in Extended Data Table 1. In this sample, 1,345 experienced a first MI, 934 a first IS and 1,971 received a diagnosis of HF for the first time during a median follow-up of 12.6 years. In the discovery subsample (a random two-thirds of the sample), 1,105 proteins were significantly associated with any CVD outcome at a false discovery rate (FDR) of <0.05. When these proteins were evaluated in the internal validation subsample (the remaining one-third) in a similar fashion, 636 proteins showed an FDR of <0.05. The top five most strongly associated proteins were ADM, PGF, SHISA5, WFDC2 and PLAUR. The hazard ratios (HRs) were in the 2.5–3.5 range for a 1 s.d. change in protein levels (Supplementary Table 1).

[1]Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [2]Nuffield Department of Population Health, University of Oxford, Oxford, UK.
e-mail: rui.zheng@uu.se

https://doi.org/10.1038/s44161-024-00545-6

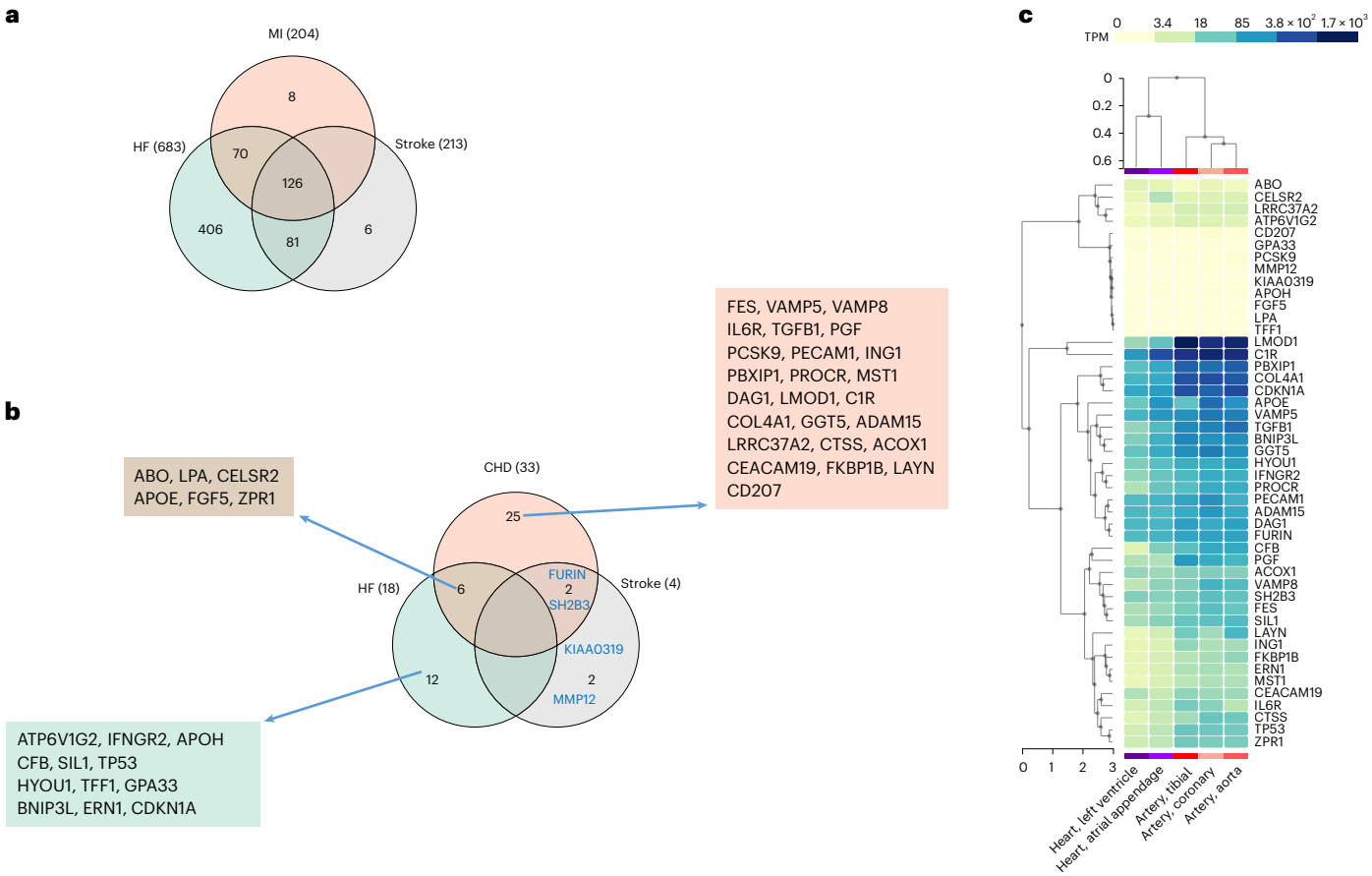


Fig. 1 | Numbers of proteins associated with the three CVD outcomes and gene expression level look-up. a, A Venn diagram visualizing the overlap of measured protein levels associated with incident CVDs by Cox proportional hazards regression after Bonferroni correction ($P < 0.000017$, two sided). Only the numbers of proteins are given. For details, see Supplementary Tables 6–8. **b**, A Venn diagram visualizing the overlap of genetically predicted levels of proteins associated with the three CVDs (CHD, IS and HF) on MR analysis (Wald ratio method). Only proteins with an FDR <0.05 (two sided) are shown. **c**, The expression levels of genes coding proteins found to be significant by MR analysis were looked up in human heart and artery tissues on the GTEx Portal. TPM, transcripts per million.

In a pathway enrichment analysis of the genes for those proteins, pathways related to inflammation and extracellular matrix dominated. In addition, regulation of insulin-like growth factor (IGF) transport, posttranslational protein phosphorylation, integrin surface interactions and growth factors were among the enriched pathways (Supplementary Table 2).

In the analysis of the complete sample, 204 proteins were related to MI (Fig. 1a) at $P < 0.000017$ (Bonferroni adjusted). The top five proteins associated with MI were PLAUR, EDA2R, WFDC2, IGFBP4 and GDF15 (Supplementary Table 3). Likewise, 213 proteins were related to IS (Supplementary Table 4) with the top five hits being CKAP4, WFDC2, NPC2, IGFBP4 and NEFL, while 683 were related to HF with the top five hits being ADM, COL18A1, ACVRL1, CRIM1 and EDN1 (Supplementary Table 5). An overview is provided in Supplementary Tables 6 and 7.

A total of 126 of the proteins were related to all three CVD outcomes in UKB. Of those, 118 were related to any of the CVDs in the replication phase in CKB, when the same proteins were related to incident cases of the three CVDs in 1,937 participants (Extended Data Fig. 1 and Supplementary Table 8). Overall, 87 of the proteins were related to more than one of the three CVD traits and 31 were related to all three traits at $P < 0.05$.

As demonstrated in Figs. 1b and 2 and Supplementary Tables 9–11, genetically predicted levels of 33 proteins were related to coronary heart disease (CHD) (at an FDR <0.05) using MR analysis. The top five most strongly associated proteins were LPA, Cadherin EGF LAG seven-pass G-type receptor 2 (CELSR2), APOE, FES and VAMP5. Four proteins were related to IS (MMP12, KIAA0319, FURIN and SH2B3), while 18 were related to HF (the top five being ABO, LPA, ATP6V1G2, IFNGR2 and CELSR2). CELSR2, ABO, LPA, APOE, FGF5 and ZPR1 were related to both CHD and HF, while SH2B3 and FURIN were shared between CHD and stroke on MR analysis. However, not all of these 47 proteins have strong expressions in human heart or artery tissues (Fig. 1c).

Sensitivity analyses using multiple *cis*-single nucelotide polymorphisms (SNPs) on MR analyses showed similar estimates as the Wald ratio estimates of the sentinel *cis*-quantitative trait loci (QTL) SNP for all 47 proteins except for GGT5, ERN1 and CFB (Supplementary Table 12). However, no significant ($P < 0.05$) results were found by any sensitivity analysis for these three proteins. The MR–Steiger test inferred that all of the causal associations were oriented from proteins to the CVD outcomes (Supplementary Table 13).

Using colocalization analysis, 10 of the 47 proteins found in the MR analysis to be linked to a CVD showed a high (>80%) posterior probability (PP) for hypothesis H4, indicating strong evidence to support shared causal variants between proteins and CVD outcomes (CELSR2–CHD and CELSR2–HF; FGF5–CHD; TGFB1–CHD; FES–CHD; FURIN–stroke and FURIN–CHD; PECAM1–CHD; PROCR–CHD; APOE-CHD; PGF–CHD; ATP6V1G2–HF), while eight showed a moderate PP (50–80%) suggesting medium evidence to support colocalization (Fig. 3).

Only 4 of the 47 proteins were significantly associated with CVD on the observational and MR analyses for MI (PGF, LAYN, FURIN and

https://doi.org/10.1038/s44161-024-00545-6

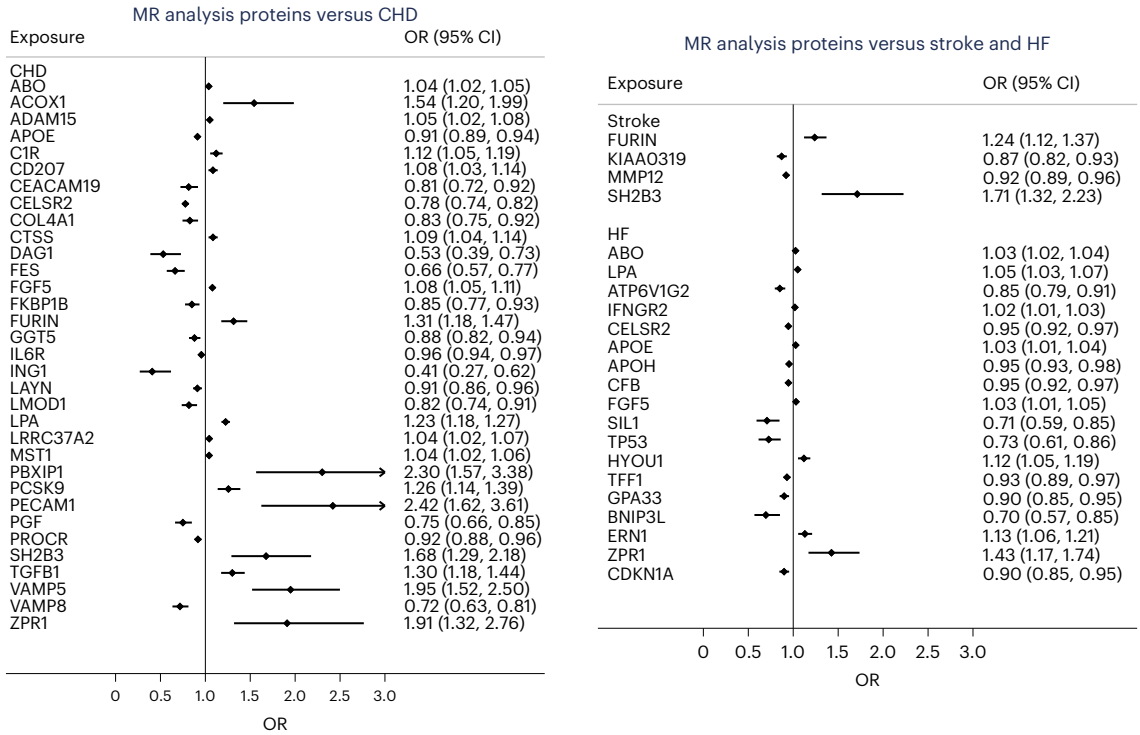


Fig. 2 | The associations of genetically predicted levels of proteins with the three CVDs. The MR Wald ratio method was used. The results are sorted by the outcome, and only associations with FDR <0.05 (two sided) are shown. The MR estimates are presented corresponding to the change of logarithm of the odds of CVD outcomes per one NPX unit. The diamonds and error bars represent the beta coefficients and 95% CIs, respectively. OR, odds ratio.

PCSK9). Of those, only FURIN and PCSK9 showed directionally consistent associations on the observational and MR analyses. In total, 6 out of 18 proteins were related to HF on the MR analyses and observational analyses, and only FGF5 and HYOU1 were directionally concordant between both analyses. Of the four proteins that were related to stroke on the MR analysis, only FURIN and MMP12 were also significantly associated with stroke on the observational analyses.

Overall, 30 out of 47 proteins of interest on the MR analyses were identified as potential drug targets, and 10 of those had tier 1 priority, including CTSS, FKBP1B, IL6R, PCSK9, PGF, TGFB1, ERN1, GPA33, IFNGR2 and TP53 (Supplementary Table 14). FGF5, PROCR and FURIN were ranked as tier 3 druggable proteins.

Of the 47 proteins previously linked to at least one of the CVDs on the primary MR analysis, only five were also related to intima-media thickness (IMT) (CEACAM19, PCSK9, APOE, FGF5 and LAYN) (Supplementary Table 15). The corresponding number of proteins was eight for carotid plaque (CELSR2, KIAA0319, DAG1, PGF, FES, TP53, LPA and APOE).

In the prospective investigation of obesity and metabolism (POEM) study, plasma levels of 20 out of the 47 proteins of interest were measured. All of those 20 proteins were related to at least one of the markers of subclinical CVD, using $P < 0.05$ (Fig. 4). Plasma levels of COL4A1, CTSS, DAG1, FGF5, FURIN, LAYN, PCSK9 and PGF were related to several of the subclinical markers in the expected direction.

Previous observational studies using the proximity extension assay technique identified GDF15 and TRAILR2 in one study[8] and GDF15 and WFDC2 in another study to be associated with all three major CVD outcomes[2]. The present much larger study identified 126 proteins levels linked to all three CVD outcomes independently of major confounders, such as age, sex, body mass index (BMI) and glomerular filtration rate[9]. Almost all of those 126 proteins were related to any CVD in the replication phase in CKB and more than half were related to two or three of the CVD. Less than 6% of the proteins related to any CVD in the observational analyses were significant on the MR analysis, suggesting that most observational associations were not causal. This is consistent with recent findings for CHD in the CKB study[10].

CELSR2 is involved in contact-mediated cell adhesion. *CELSR2* forms a cluster at the 1p13.3 locus with two other genes (*PSRC1* and *SORT1*). A previous MR study indicated a causal role of this cluster for CHD[11]. According to Genotype-Tissue Expression (GTEx), the *cis*-pQTL is expressed in the coronary arteries and aorta. However, a previous study has demonstrated that it is rather a noncoding SNP that creates a transcription factor binding site altering the expression of the *SORT1* gene in the liver. SORT1 in mouse liver had a profound effect on hepatic very-low-density lipoprotein production and thereby low-density lipoprotein (LDL) cholesterol[12].

PROCR plays a key role in protein C activation and regulation of blood coagulation[13]. A previous MR study suggested a causal role of PROCR in the development of CHD[14]. Currently, two drugs have been registered for targeting PROCR in clinical trials for non-CVD outcomes.

On both observational and MR analyses, FURIN was associated with both MI and IS, which was further supported by colocalization analysis. FURIN is a peptidase known to be involved in the cleavage of precursor proteins to active proteins[15]. According to GTEx, the *cis*-pQTL used as instrument for FURIN is expressed in different arteries, including coronary arteries. A *cis*-pQTL SNP has previously been linked to CHD in genome-wide association studies (GWAS)[16], in addition to blood pressure[17] and metabolic syndrome[18]. FURIN is upregulated in the immune cells of human atherosclerotic plaques[19]. Plasma levels of FURIN were also related to several important markers of subclinical CVD in the POEM sample. Plasma FURIN levels have previously been associated with blood pressure, lipids, fasting glucose, obesity and prevalent MI[20], and was strongly associated with ischemic heart disease (IHD) in the CKB study[10]. According to the database DrugBank, FURIN is currently being explored experimentally as a drug target, but no clinical trials were registered (Supplementary Table 14).

https://doi.org/10.1038/s44161-024-00545-6

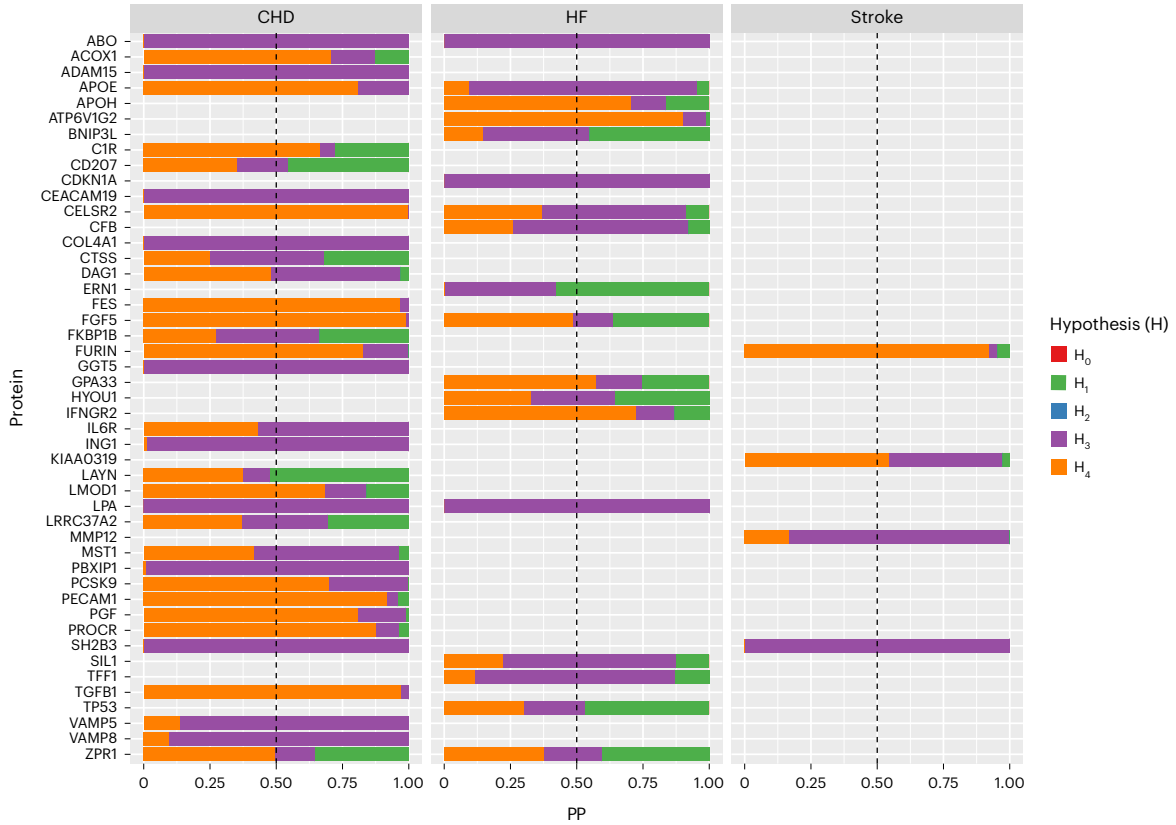


Fig. 3 | Colocalization analysis of the association of proteins of interest with CVD, including CHD, HF and stroke. H_0, no traits have a genetic association; H_1, only protein has a genetic association; H_2, only CVD has a genetic association; H_3, protein and CVD both have an association but with different causal variants; and H_4, protein and CVD are associated with the same causal variant. This Bayesian-based method resulted in five PP to assess the support of each of the five hypotheses. Strong support of colocalization was defined if the $PP.H_4 > 0.8$, whereas medium support of colocalization was defined as $0.5 < PP.H_4 \leq 0.8$.

FGF5 is involved in cell signaling. FGF5 was causally linked to IMT and plasma levels of FGF5 in the POEM study were related to a poor carotid artery distensibility. The *cis*-pQTL SNP used as genetic instrument for FGF5 has been linked to CHD[16] and to blood pressure in previous studies[21,22].Thus, FGF5 is probably a protein of interest for several CVDs, since hypertension is causally related to all major CVDs[23].

Only one sentinel *cis*-pQTL of each protein was used to minimize the risk of horizontal pleiotropy[24], using the approach adopted in UKB[25]. We additionally conducted sensitivity analyses using multiple independent *cis*-pQTLs, but that did only marginally change the MR estimates for most of the proteins.

Only a few proteins identified by the observational analyses showed a significant MR estimate, being directionally consistent with the MR analysis and a high probability of colocalization. The reasons for the discrepant results are probably multifactorial: first, that observational analyses are susceptible to several biases, such as residual confounding and, second, reverse causation. One such example is the N-terminal peptide proBNP and IMT association[6]. Third, the MR estimate is generally interpreted as a life-long effect of the exposure on the outcome[26], which differs from the limited-time effect estimated by the observational design. Fourth, a complicated phenomenon–developmental compensation could occur if a SNP influences the protein levels during fetal or early postnatal development[27].

MR can be biased by different, but correlated (in linkage disequilibrium (LD)), causal variants of the exposure and outcome where the latter variant influences the outcome via an alternative pathway[27]. Less than half of the associations found significant by MR showed evidence of colocalization, which is not unusual[28]. For the rest of proteins, the *cis*-pQTLs may be different from the causal variants of CVD or there existed more than one causal variant. The Bayesian colocalization method is also sensitive to the set of SNPs used and the choice of prior probabilities[29].

The main strength of the present study is the use of the large UKB for both observational and MR analyses. Another strength is the replication in the external CKB. While CKB has protein measurements in a substantial number of incident cases of IHD, it does currently have protein measurements in only a limited number of incident stroke and HF cases. Although the power to reproduce the protein findings in UKB is not optimal, it is reassuring to note that almost all protein findings in UKB were related to at least one of the three CVD outcomes in CKB. Another strength of the present study is that only *cis*-genetic signals for a great number of proteins were used on the MR analyses and colocalization analysis. In addition, we investigated plasma levels for some of the important proteins with several markers of subclinical CVD in POEM to assess the potential mechanisms underlying their associations with CVD outcomes.

A limitation is that most of the outcome GWAS studies mainly were conducted in individuals of European descent and therefore the results require validation in other populations. Another limitation is that not all proteins have an identified strong *cis*-SNP. The *cis*-pQTLs used by us explained between 0.2% and 68% of protein-level variance, with a median value of 5% (Extended Data Fig. 2). In a post hoc power calculation (Supplementary Tables 9–11), the power varied greatly for the different proteins and thus false negative findings could occur owing to limited power for some proteins.

Overall, while several hundred proteins were related to CVD, genetically predicted levels of only 47 proteins were linked to CHD,

https://doi.org/10.1038/s44161-024-00545-6

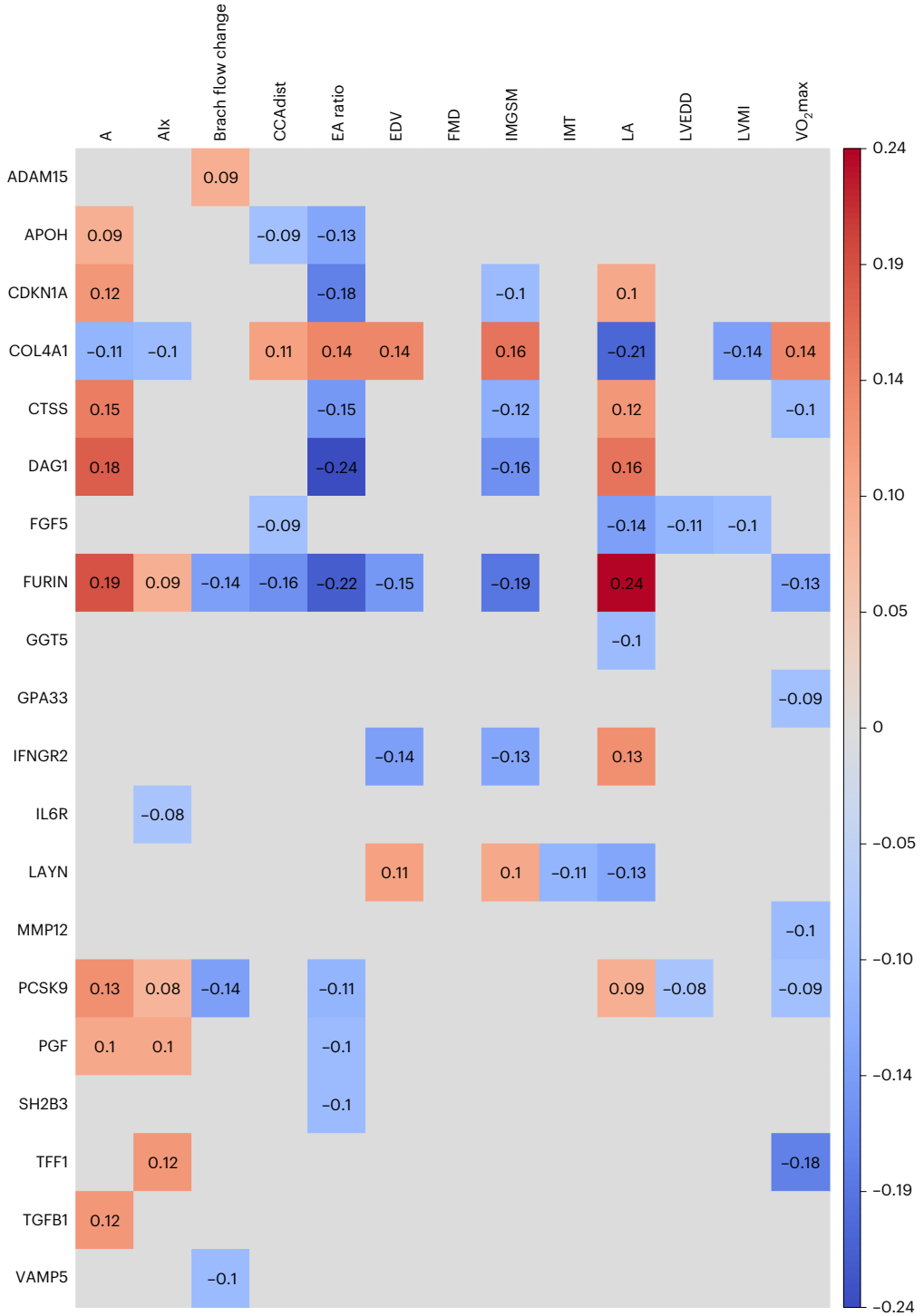


Fig. 4 | Heat map showing cross-sectional relationships between plasma levels of 20 proteins of interest and subclinical markers of CVDs. Multivariable linear regression was used and the regression coefficients are shown for relationships with $P < 0.05$ (two sided). A gray square indicates $P \geq 0.05$. A, transmitral A-wave at echocardiography; AIx, augmentation index at radial artery pulse wave analysis; Brach flow change, the increase in brachial artery blood flow following 5 min of blood flow arrest; CCAdist, carotid artery distensibility by ultrasound; EA ratio, left ventricular diastolic function index; EDV, acetylcholine-mediated increase in forearm blood flow; FMD, brachial artery FMD; IMGSM, echolucency of the carotid artery intima–media complex; VO_2max, maximal oxygen consumption at an exercise test. For complete protein names, see Supplementary Table 16.

HF or stroke, showing that the vast majority of protein–CVD associations found in observational settings were probably not causal. On the basis of both observational and genetic analyses, several proteins, including FGF5, PROCR and FURIN, are potentially important targets for drug development for CVD prevention.

Methods

UKB population sample

The UKB is a large, multicenter, prospective cohort study conducted across the UK. In 2006–2010, over 500,000 individuals aged 40–69 years underwent physical measurements, and blood samples

were stored for later analysis of genes and biomarkers. The present study used data from the 52,164 individuals with valid proteomics data. The UKB study was approved by the UK North West Multicentre Research Ethics Committee (application no. 90143) and the Swedish Ethical Review Authority (no. 2023-00148-01), and all participants provided written informed consent.

All participants in UKB had plasma levels of glucose, LDL and high-density lipoprotein (HDL) cholesterol and creatinine measured by a Beckman Coulter AU5800 using standard methods. Blood pressure was measured twice in the sitting position with an automated Omron device. Estimated glomerular filtration rate (eGFR) was calculated by the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula[30].

Ethnicity was categorized into four groups: White, Black, Asian and other. The Townsend social deprivation index was used as a marker of socioeconomic status. Smoking status was categorized as never, previous or current cigarette smoking.

All disease outcomes were coded using the 10th edition of the International Classification of Diseases (ICD-10 codes) for first fatal or non-fatal MI (ICD-10 code I21), IS (I63) or HF (I50). These rather narrow definitions of the three CVD outcomes might misclassify some cases as controls, but our previous validation study[31] indicated that expanding the code range is likely to result in misclassification of a substantial number of cases since additional codes had additional errors. Moreover, minimizing misclassification of cases had a higher priority than minimizing misclassification of controls in this study design.

Plasma levels of 2,923 proteins were measured by the Olink EXPLORE assay (OLINK), which involves a proximity extension assay technique. All assays underwent an extensive quality control evaluation[25]. According to information from the manufacturer, the mean intra-assay and interassay coefficients of variation observed are in the 8–17% range. Assays for four proteins (GLIPR1, NPM1, PCOLCE and CST1) failed quality control in >40,000 individuals, hence the final analyses were limited to 2,919 proteins.

CKB population sample

CKB is a population-based prospective study of 512,000 Chinese adults aged 30–79 years who were recruited from ten regions (five rural and five urban) in China during 2004–2008. Data were collected on smoking, medical history and education using interview-administered questionnaires. Physical measurements, including BMI and blood pressure, were measured in study clinics. Details of the study design and baseline characteristics have been previously reported[32]. Ethical approval was obtained from the Oxford Tropical Research Ethics Committee, the ethical review committees of the Chinese Center for Disease Control and Prevention, Chinese Academy of Medical Sciences and the institutional review board at Peking University. The Chinese Ministry of Health approved the study at the start in 2004 (including export of data and plasma samples to Oxford), and also approved electronic linkage to health insurance records in 2011. All participants provided written informed consent. Information on incident cases of IHD, IS and HF were obtained from death and disease registers and from the National Health Insurance system for hospital admissions.

Plasma proteins were measured using the Olink EXPLORE panel in stored plasma samples that were collected at baseline in CKB using the same analytical platform as in UKB, with the same number of proteins analyzed.

CKB was used as an external replication dataset for findings in UKB. The sample used in CKB was derived from a nested case–cohort study with 1,937 incident cases of non-fatal or fatal MI or death from IHD (ICD-10 codes: I20, I22–I25) during a 12 year follow-up and a subcohort of 2,001 participants[10]. The mean age was 64 and 51 years and the proportion of women was 38% and 61% in the IHD and subcohort group, respectively. The replication analyses for the present study included 1,937 first incident cases of IHD, 224 cases of IS and 24 cases of HF. For observational analyses of proteins with CVD in CKB, Cox regression models were used to estimate adjusted HRs (and 95% confidence intervals (CIs)) using the Prentice pseudo-partial likelihood for case–cohort designs. All analyses were stratified on the basis of sex and study area, and adjusted sequentially in different models for additional CVD risk factors.

Two-sample MR

Selection of *cis*-pQTLs as genetic instruments. The genetic instruments were obtained from the UKB Pharma Proteomics Project, which is one of the largest genomic studies on protein levels released so far[25]. In this study, pQTLs of 2,923 plasma proteins were discovered in 54,219 UKB participants. We only used the sentinel *cis*-pQTL SNP of each protein (with $P < 1.7 \times 10^{-11}$) in the combined cohort as genetic instruments if available to limit the chance of horizontal pleiotropy. For proteins with the sentinel *cis*-pQTL being multiallelic variants, we used the meta-analyzed statistics of the discovery ($n = 34,557$) and replication ($n = 13,358$) cohorts if applicable. For proteins with the sentinel *cis*-QTLs being multiallelic in both the combined and the discovery-replication cohorts, we used the SNP within a 1 Mb window of the protein-coding gene that has the lowest P value (and required $P < 5 \times 10^{-8}$ and F-statistic > 10) and is biallelic in the combined cohort. A detailed description of the combined and discovery-replication cohorts can be found in ref. 25. In total, 1,826 *cis*-pQTL SNPs were retained as instruments (Supplementary Table 16).

Data sources for outcomes. We obtained summary-level data for the association of the pQTLs with the three CVD outcomes and carotid artery IMT and carotid plaque from several sources as detailed in Supplementary Table 17.

For CHD, the effect size estimates were obtained from a two-stage meta-analysis study from the CARDIoGRAMplusC4D Consortium, which involved 60,801 CHD cases (mainly MI) and 123,504 controls of European and South and East Asian ancestries[33]. The GWAS summary data of IS was obtained from a multiancestry meta-analysis study of 514,791 individuals, including 60,341 cases and 454,450 controls (MEGASTROKE)[34]. The SNP–HF association estimates were obtained from a multiancestry GWAS meta-analysis including 115,150 cases and 1,550,330 controls[35].

Summary GWAS results for carotid artery IMT in 71,128 participants and carotid artery plaques in 21,540 out of 48,434 participants examined were obtained from a GWAS of European ancestry populations[36]. Appropriate ethical approvals were obtained from the different samples contributing with data to the different GWAS. The present two-sample MR study was approved by the Swedish Ethical Review Authority.

Cross-sectional analysis in the POEM study

The POEM study is a population-based cohort of residents of Uppsala City in Sweden, who were aged 50 years ($n = 502$) and included 50% females. All individuals included were of European descent. The data were collected between 2010 and 2016 and further details of data collection and cohort characteristics have been previously reported[37]. The study was approved by the ethics committee of Uppsala University (2009/057), and all participants provided written informed consent.

All participants were investigated after fasting overnight, except the exercise test being performed on another day within a week in the nonfasting state. Blood samples were collected, and plasma samples were stored at −80 °C for later analysis of proteins. A large number of physical tests were performed to obtain information on subclinical markers of CVDs.

Bicycle exercise test with gas exchange. Using a bicycle ergonometer, a maximal exercise test with gas exchange measurements was performed (Jaeger Oxygen Pro, Vyaire medical). Blood pressure, heart

rate, vital capacity, rate of oxygen consumption (VO_2) and VCO_2 were measured at rest, and thereafter the participants were asked to work until exhaustion. The work load was increased by 10 W min^{-1}, starting at 30 W for women and 50 W for men. The maximal VO_2 during the last minute of work was recorded.

Blood pressure. Blood pressure was measured manually by a mercury sphygmomanometer in supine position after 30 min of rest. By applying radial pulse wave recordings (see below), central blood pressure was calculated by the software included in the commercial device.

The invasive forearm technique. Forearm blood flow was measured by venous occlusion plethysmography (Elektromedicin). After evaluation of resting forearm blood flow, local intra-arterial drug infusions were given in the brachial artery during 5 min for each dose. At the end of each infusion step, forearm blood flow was evaluated. The infused dosages were 25 and 50 μg min^{-1} for acetylcholine (Clin-Alpha) to evaluate endothelium-dependent vasodilation (EDV) in forearm resistance vessels and 5 and 10 μg min^{-1} for SNP (Nitropress, Abbot) to evaluate endothelium-independent vasodilation (EIDV). The two drugs were given in random order between subjects.

EDV in forearm resistance vessels was defined as forearm blood flow during infusion of either 25 or 50 μg min^{-1} of acetylcholine minus resting forearm blood flow divided by resting forearm blood flow (denoted EDV 25 μg and EDV 50 μg). EIDV in forearm resistance vessels was defined as forearm blood flow during infusion of either 5 or 10 μg min^{-1} of SNP minus resting forearm blood flow divided by resting forearm blood flow (denoted EIDV 5 μg or EIDV 10 μg).

The brachial artery ultrasound technique. The brachial artery diameter was assessed by external B-mode ultrasound imaging 2–3 cm above the elbow (Acuson XP128 with a 10 MHz linear transducer, Acuson Mountain View). Following measurement of the brachial artery diameter at rest, a blood flow increase was induced by inflation of a pneumatic cuff placed around the forearm to a pressure at least 50 mmHg above systolic blood pressure for 5 min and then a suddenly release of the cuff. Flow-mediated vasodilation (FMD) was defined as the maximal brachial artery diameter recorded between 30 and 90 s following cuff release minus diameter at rest divided by the diameter at rest. Blood flow velocity was recorded by Doppler at rest and in the very early hyperemic period (within 5–10 s). From the Doppler velocity time integral recording, heart rate and the diameter of the vessel, the blood flow was calculated (in millilitres per minute). The blood flow increase during the induced hyperemia was defined as blood flow during hyperemia minus resting blood flow divided by resting blood flow.

Pulse wave analysis. A micromanometer tipped probe (Sphygmocor, Pulse Wave Medical Ltd.) was applied to the surface of the skin overlying the radial artery and the peripheral radial pulse wave was continuously recorded. The mean values of around ten pulse waves were used for analyses. On the basis of transfer functions included in the device, aortic systolic and diastolic blood pressures were calculated from the radial recordings. In addition, the augmentation index (AIx), the ratio between the first reflected wave and the systolic peak, was also measured.

Carotid artery compliance. The diameter of the common carotid artery of the right side 1–2 cm proximal of the bifurcation was measured by ultrasound M-mode at its maximal diameter in systole and the minimal diameter in diastole (Acuson XP128 with a 10 MHz linear transducer, Mountain View). The distensibility of the carotid artery was calculated as the change in diameter maximum to minimum in relation to the minimal diameter in diastole divided by the central pulse pressure obtained by pulse wave analysis.

Carotid artery ultrasound evaluation. The carotid artery was assessed by external B-mode ultrasound imaging (Acuson XP128 with a 10 MHz linear transducer, Mountain View). The IMT was evaluated in the far wall in the common carotid artery 1–2 cm proximal to the bulb.

The images were digitized and imported into the artery measurement software (Artery Measurement Software) automated software for dedicated analysis of IMT and the gray scale median of the intima-media complex. A 10 mm segment with good image quality was chosen for IMT analysis from the carotid artery. The value obtained is the mean of around 100 discrete measurements over the 10 mm segment. The given value for carotid artery IMT is the mean value from both sides.

A region of interest was placed manually around the intima-media segment that was evaluated for IMT, and the program calculates the echogenicity (gray scale) of the intima-media complex from analysis of the individual pixels within the region of interest on a scale from 0 (black) to 256 (white). The blood was used as the reference for black, and the adventitia was the reference for white. The grayscale median value given is the mean value from both sides.

Echocardiography and Doppler. A comprehensive two-dimensional and Doppler echocardiography was performed with an Acuson XP124 cardiac ultrasound unit (Acuson) A 2.5 MHz transducer was used for the majority of the examinations.

Left ventricular dimensions were measured with M-mode on-line from the parasternal projections, using a leading-edge to leading-edge convention. Measurements included left atrial diameter (LA), interventricular septal thickness, posterior wall thickness, left ventricular diameter in end diastole (LVEDD) and left ventricular diameter in end systole.

Left ventricular mass (LVM) was determined from the Penn conversion. The LVM was then indexed for height$^{2.7}$ to obtain the left ventricular mass index (LVMI).

The left ventricular diastolic filling pattern of the mitral inflow was obtained from the apical transducer position with the pulsed Doppler sample volume between the tips of the mitral leaflets during diastole. The peak velocity of the early rapid filling wave (E wave) and the peak velocity of atrial filling (A-wave) were recorded, and the E to A ratio (E/A) was calculated.

Means and standard deviations for the markers of subclinical CVD in POEM are given in Supplementary Table 18.

Statistical analysis

Observational analyses. Cox proportional hazards models were used to evaluate whether plasma protein levels were related to incident CVD (combined endpoint of MI, IS or HF) in the primary analysis. Covariates included age, sex, ethnicity, the Townsend deprivation index, smoking (never, previous or current), systolic blood pressure, LDL and HDL cholesterol, diabetes, BMI and eGFR. eGFR was estimated from plasma creatinine using the updated CKD-EPI formula[30].

One model was run for each protein. In total, 1,832 individuals with prevalent CVD at baseline were excluded from the analyses. The proportional hazard assumption was checked by visual inspection of Kaplan–Meier curves for the proteins of interest. Proteins were analyzed on a *z*-scale, and the HR was expressed as a 1 s.d. change of the normalized protein expression (NPX) unit defined by the manufacturer. These analyses were first performed in a random sample of two-thirds of the complete sample (discovery sample). The proteins showing an FDR Benjamini–Hochberg-adjusted *P* value (shortly as FDR) <0.05 were then evaluated in a similar fashion in the remaining one-third of the complete sample (validation sample). If the associations of these proteins with any one CVD outcome had an FDR <0.05 in the validation sample, such protein–CVD associations were defined as significant. Likewise, we evaluated associations separately with each of the three CVD outcomes (MI, IS and HF). Prevalent cases of the evaluated outcomes were excluded from the respective analyses

(1,407 prevalent cases of MI, 196 of IS and 485 of HF). To maximize the power in this part of the study, we used the complete sample, but now required a P value <0.000017 (Bonferroni correction) to be statistically significant.

In the replication phase of the 126 proteins related to all three CVD outcomes in UKB, data from CKB were used. Consistent with analyses in UKB, Cox proportional hazards models were used to regress each of those 126 proteins separately with each of the three CVD outcomes. Age, sex, study site, fasting time, education, smoking, physical activity, systolic blood pressure, diabetes, apolipoprotein B/apolipoprotein A ratio, BMI and history of chronic kidney disease were used as covariates. To assess replication, a nominal $P < 0.05$ was regarded as statistically significant.

MR analyses. For proteins with *cis*-pQTLs statistics meta-analyzed by those of the discovery and replication cohorts (see ref. 25 for details), the MR instrument statistics were generated via a linear fixed-effects model (R package 'metafor')[38] where the weighted estimation with inverse-variance weights was used.

We used the two-sample MR design for the primary analyses (proteins versus the three CVD outcomes). Since only a single *cis*-pQTL SNP was used as the instrument in each MR analysis, the Wald ratio method was applied to obtain the MR estimate using the R package TwoSampleMR[39]. The MR estimates were presented corresponding to the change of logarithm of the odds of binary outcomes or the original unit of continuous outcome (IMT) per one NPX unit, except when stated otherwise. P values were adjusted within each outcome using FDR controlling procedure, and <0.05 was deemed statistically significant. The MR analysis of proteins of interest (that is, passing FDR-adjusted $P < 0.05$ threshold in the primary MR analysis) with carotid artery IMT and carotid plaque was conducted in the same manner as the primary analysis. However, this analysis was regarded as exploratory and supportive, and therefore $P < 0.05$ was regarded as statistically significant.

We conducted a post hoc power calculation using the R code provided in the work by Burgess[40] for the three CVD outcomes. During the calculation, R^2 (variance of exposure explained by the genetic instrument) was estimated using the formula $R^2 = \text{beta}^2 \times 2 \times \text{EAF} \times (1 - \text{EAF})$ where beta represents the effect estimate of the SNP–exposure association and EAF is the effect allele frequency[41].

MR sensitivity analyses. We used the MR–Steiger test to examine whether the estimated effect was correctly oriented from proteins to CVD outcomes. For the proteins of interest, we further performed sensitivity analyses including MR inverse-variance weights, MR weighted median, MR-weighted mode and MR Egger. In brief, independent (low LD, defined as $R^2 < 0.01$, clumping window >10 kb) GWAS-significant ($P < 5\times10^{-8}$) SNPs at the *cis*-loci (within 1 Mb) of each protein-coding gene with F-statistic >10 were selected as instruments. Horizontal pleiotropy (P value of MR-Egger intercept) and heterogeneity (Cochrane's Q value) of the instruments were examined by the relevant functions integrated in the R package TwoSampleMR. A total of 13 of 47 proteins of interest had a single independent *cis*-SNP at the defined window, and therefore no sensitivity analyses were conducted for these proteins.

Colocalization analysis. To identify whether the MR associations of proteins of interest with CVD outcomes were driven by LD, we further performed colocalization analysis using the R package Coloc[29]. The analysis was based on the enumeration method to gauge the support for five exclusive hypotheses regarding two potentially related traits (T1 and T2) in a predefined genomic region: (1) H_0: no traits have a genetic association; (2) H_1: only T1 has a genetic association; (3) H_2: only T2 has a genetic association; (4) H_3: T1 and T2 both have an association but with different causal variants; and (5) H_4: T1 and T2 are associated with the same causal variant. This Bayesian-based method results in five PP to assess the support of each of the five hypotheses. In this analysis, three prior probabilities were set as follows: p1 (a SNP is associated with T1) = 1×10^{-4}; p2 (a SNP is associated with T2) = 1×10^{-4}; for p12 (a SNP is associated with both traits), we used a more conservative value, 5×10^{-6}, as suggested previously for improved robustness[42]. The sensitivity of colocalization inference to variations in p12 was further examined by visualization. Strong support of colocalization was defined if the PP.H_4 >0.8, whereas medium support of colocalization was defined as 0.5 < PP.H_4 ≤ 0.8.

Pathway enrichment and GTEx analyses. To determine whether CVD-associated proteins were enriched in specific pathways, an overrepresentation analysis was conducted at Reactome[43] for all proteins associated with the combined CVD endpoint in the observational setting. For the proteins of interest found on MR analysis, the expression level (transcript per million) of their coding genes in human heart and artery tissues was further looked up on the GTEx Portal.

Protein druggability. For proteins of interest, we sought evidence for druggability tiers using the approach previously reported by Finan et al.[44], where tier 1 was defined as proteins targeted by approved drugs and drugs in clinical development, tier 2 as proteins closely related to drug targets or with associated drug-like compounds and tier 3 as extracellular proteins and members of key drug target families. For each protein, detailed targeting drug information was curated from DrugBank, clinicaltrials.gov and ChEMBL databases wherever available.

Cross-sectional evaluation in the POEM study. Plasma levels of the proteins found to be of interest in the MR analyses were inverse-normalized rank-transformed and were thereafter associated with each of the subclinical markers of CVD in the POEM sample using linear regression models, including sex as a covariate (age was the same in all subjects). $P < 0.05$ was considered statistically significant in this supportive analysis.

All the observational analyses were conducted in STATA (version 16.1), while the MR and colocalization analyses were conducted in R (version 4.1.0). All statistical tests were two tailed.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The UKB and its data are an open research resource, available following submission of a research plan at https://www.ukbiobank.ac.uk. The CKB observational data that support the findings of this study are available to bona fide researchers on application under the CKB Open Access Data Policy (www.ckbiobank.org). Summary-level GWAS data are available for proteins at https://metabolomips.org/ukbbpgwas/, for CHD at http://www.cardiogramplusc4d.org/, for IS at https://www.megastroke.org/, for HF at https://www.ebi.ac.uk/gwas/studies/GCST90162626 and for ultrasound-measured carotid artery IMT and carotid plaques at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000930.v6.p1 (accession no. phs000930.v6.p1). Data supporting the findings of the POEM study are provided in the article and related files. Raw data are not publicly available owing to Swedish law, as they contain sensitive personal information, but could be obtained from the POEM study following a request to lars.lind@medsci.uu.se. Other online databases used are Reactome (https://reactome.org/, version 86 on 03/11/2023), GTEx Portal (https://gtexportal.org/home, dbGaP accession no. phs000424.v8.p2), DrugBank (https://go.drugbank.com/), clinical trials (https://clinicaltrials.gov/) and ChEMBL (https://www.ebi.ac.uk/chembl/). Source data are provided with this paper.

Code availability

The code used for the main analyses in this work is available via Zenodo at https://doi.org/10.5281/zenodo.13347408 (ref. 45).

References

1. Stenemo, M. et al. Circulating proteins as predictors of incident heart failure in the elderly. *Eur. J. Heart Fail.* **20**, 55–62 (2018).
2. Lind, L. et al. Large-scale plasma protein profiling of incident myocardial infarction, ischemic stroke, and heart failure. *J. Am. Heart Assoc.* **10**, e023330 (2021).
3. Lind, L., Sundstrom, J., Stenemo, M., Hagstrom, E. & Arnlov, J. Discovery of new biomarkers for atrial fibrillation using a custom-made proteomics chip. *Heart* **103**, 377–382 (2017).
4. Lind, L. et al. Discovery of new risk markers for ischemic stroke using a novel targeted proteomics chip. *Stroke* **46**, 3340–3347 (2015).
5. Lind, L. et al. Use of a proximity extension assay proteomics chip to discover new biomarkers for human atherosclerosis. *Atherosclerosis* **242**, 205–210 (2015).
6. Lind, L. et al. Plasma protein profile of carotid artery atherosclerosis and atherosclerotic outcomes: meta-analyses and mendelian randomization analyses. *Arterioscler. Thromb. Vasc. Biol.* **41**, 1777–1788 (2021).
7. Lind, L. et al. The plasma protein profile and cardiovascular risk differ between intima-media thickness of the common carotid artery and the bulb: a meta-analysis and a longitudinal evaluation. *Atherosclerosis* **295**, 25–30 (2020).
8. Lind, L., Arnlov, J. & Sundstrom, J. Plasma protein profile of incident myocardial infarction, ischemic stroke, and heart failure in 2 cohorts. *J. Am. Heart Assoc.* **10**, e017900 (2021).
9. Lind, L. et al. Longitudinal effects of aging on plasma proteins levels in older adults - associations with kidney function and hemoglobin levels. *PLoS ONE* **14**, e0212060 (2019).
10. Mazidi, M. et al. Plasma proteomics to identify drug targets for ischemic heart disease. *J. Am. Coll. Cardiol.* **82**, 1906–1920 (2023).
11. Castillo-Avila, R. G. et al. Association between genetic variants of CELSR2-PSRC1-SORT1 and cardiovascular diseases: a systematic review and meta-analysis. *J. Cardiovasc. Dev. Dis.* **10**, 91 (2023).
12. Musunuru, K. et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714–719 (2010).
13. Rao, L. V. M., Esmon, C. T. & Pendurthi, U. R. Endothelial cell protein C receptor: a multiliganded and multifunctional receptor. *Blood* **124**, 1553–1562 (2014).
14. Schooling, C. M. & Zhong, Y. Plasma levels of the anti-coagulation protein C and the risk of ischaemic heart disease. A Mendelian randomisation study. *Thromb. Haemostasis.* **117**, 262–268 (2017).
15. Thomas, G. Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Nat. Rev. Mol. Cell Biol.* **3**, 753–766 (2002).
16. van der Harst, P. & Verweij, N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ. Res.* **122**, 433–443 (2018).
17. Surendran, P. et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **48**, 1151–1161 (2016).
18. Ueyama, C. et al. Association of FURIN and ZPR1 polymorphisms with metabolic syndrome. *Biomed Rep.* **3**, 641–647 (2015).
19. Turpeinen, H. et al. Proprotein convertases in human atherosclerotic plaques: the overexpression of FURIN and its substrate cytokines BAFF and APRIL. *Atherosclerosis.* **219**, 799–806 (2011).
20. Ferkingstad, E. et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**, 1712–1721 (2021).
21. International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
22. Jeong, H., Jin, H. S., Kim, S. S. & Shin, D. Identifying interactions between dietary sodium, potassium, sodium-potassium ratios, and FGF5 rs16998073 variants and their associated risk for hypertension in Korean adults. *Nutrients* **12**, 2121 (2020).
23. Lind, L., Ingelsson, M., Sundstrom, J. & Arnlov, J. Impact of risk factors for major cardiovascular diseases: a comparison of life-time observational and Mendelian randomisation findings. *Open Heart* **8**, e001735 (2021).
24. Swerdlow, D. I. et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int. J. Epidemiol.* **45**, 1600–1616 (2016).
25. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
26. Smith, G. D., Holmes, M. V., Davies, N. M. & Ebrahim, S. Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues. *Eur. J. Epidemiol.* **35**, 99–111 (2020).
27. Smith, G. D. Mendelian randomization for strengthening causal inference in observational studies: application to gene × environment interactions. *Perspect. Psychol. Sci.* **5**, 527–545 (2010).
28. Yuan, S. et al. Plasma proteins and onset of type 2 diabetes and diabetic complications: proteome-wide Mendelian randomization and colocalization analyses. *Cell Rep. Med.* **4**, 101174 (2023).
29. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
30. Inker, L. A. et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *N. Engl. J. Med.* **385**, 1737–1749 (2021).
31. Ingelsson, E., Arnlov, J., Sundstrom, J. & Lind, L. The validity of a diagnosis of heart failure in a hospital discharge register. *Eur. J. Heart Fail.* **7**, 787–791 (2005).
32. Chen, Z. et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int. J. Epidemiol.* **40**, 1652–1666 (2011).
33. The CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
34. Malik, R. et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
35. Levin, M. G. et al. Genome-wide association and multi-trait analyses characterize the common genetic architecture of heart failure. *Nat. Commun.* **13**, 6914 (2022).
36. Franceschini, N. et al. GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nat. Commun.* **9**, 5141 (2018).
37. Lind, L., Strand, R., Michaelsson, K., Ahlstrom, H. & Kullberg, J. Voxel-wise study of cohort associations in whole-body MRI: application in metabolic syndrome and its components. *Radiology* **294**, 559–567 (2020).
38. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J Stat Softw.* **36**, 1–48 (2010).
39. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
40. Burgess, S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int. J. Epidemiol.* **43**, 922–929 (2014).

41. Levin, M. G. et al. Genetics of height and risk of atrial fibrillation: a Mendelian randomization study. *PLoS Med.* **17**, e1003288 (2020).
42. Wallace, C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLoS Genet.* **16**, e1008720 (2020).
43. Gillespie, M. et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687–D692 (2022).
44. Finan, C. et al. The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, eaag1166 (2017).
45. Zheng, R & Lind, L. Statistical code. *Zenodo* https://doi.org/10.5281/zenodo.13347408 (2024).

Acknowledgements

This research has been conducted using the UKB resource under application no. 90143. We acknowledge the members of the CKB Study Group. The POEM study received a grant from the Swedish Heart-Lung Foundation (L.L., no. 20090237) and also received funding from the University Hospital of Uppsala, Sweden (L.L., no. 110-2023).

Author contributions

L.L. and R.Z. designed the project. L.L. applied for the data from UKB and is the principal investigator of the POEM study. R.C. contributed to acquisition of the CKB data. L.L., M.M., D.A.B. and R.Z. performed statistical analyses and interpretation of the data. L.L. wrote the main paper text. L.L. and R.Z. contributed to the visualization of the data. All authors contributed to critical revision of the paper, read and approved the final version for publication, including the authorship list.

Funding

Open access funding provided by Uppsala University.

Competing interests

The authors declare no competing interests.

Additional information

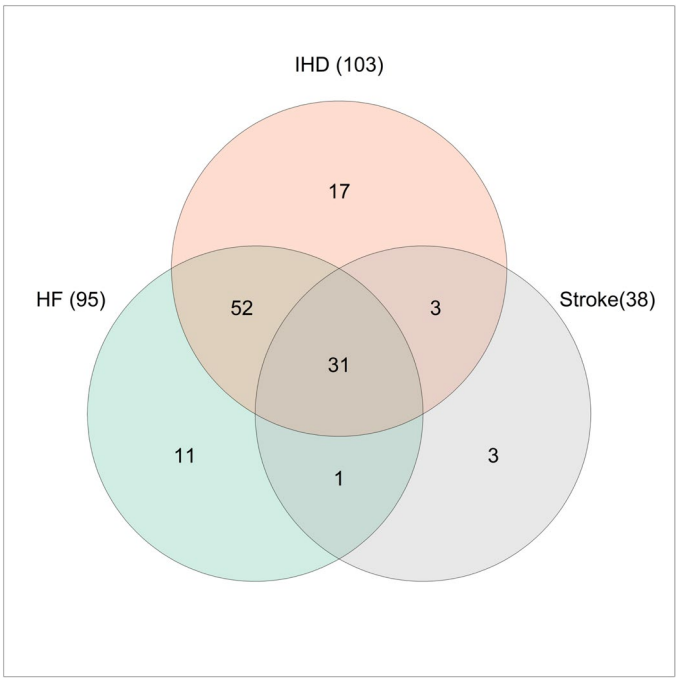
Extended data is available for this paper at https://doi.org/10.1038/s44161-024-00545-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44161-024-00545-6.

Correspondence and requests for materials should be addressed to Rui Zheng.

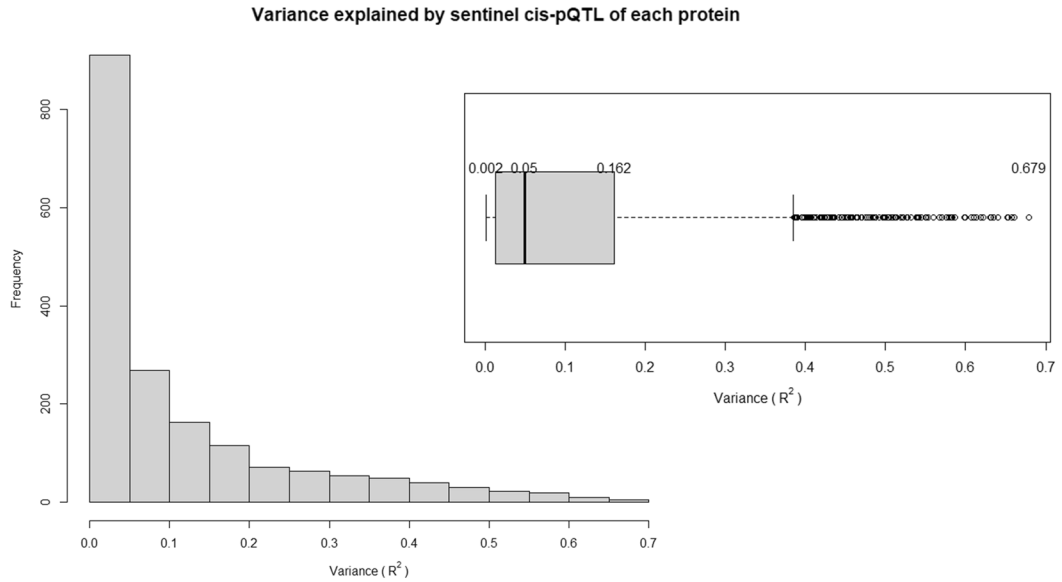
Peer review information *Nature Cardiovascular Research* thanks Peter Ganz and Aroon Hingorani for their contribution to the peer review of this work.

https://doi.org/10.1038/s44161-024-00545-6



Extended Data Fig. 1 | Numbers of proteins of interest associated with three CVD outcomes in CKB. Venn diagram visualizing the overlap of measured protein levels significantly (nominal $P < 0.05$) associated with incident cardiovascular diseases (ischemic heart disease (IHD), ischemic stroke; and heart failure, HF) in the replication phase in China Kadoorie Biobank (CKB) by Cox proportional hazards regression. Only the numbers of proteins are given. Source data can be found in Supplementary Table 8.

https://doi.org/10.1038/s44161-024-00545-6



Extended Data Fig. 2 | Variance (R^2) of plasma levels of proteins explained by the sentinel cis-pQTL SNPs used as the genetic instrument on the MR analysis. Source data can be found in Supplementary Table 13. The minima is the extreme of the lower whisker (0.002); the extreme of the upper whisker is 0.385; the maxima is rightmost point (0.679); the solid vertical line inside the box represent the median (0.05); the lower and upper hinges of the box represent the first and third quartiles (0.013 and 0.162, respectively).

https://doi.org/10.1038/s44161-024-00545-6

Extended Data Table 1 | Basic characteristics of the UK Biobank participants with protein measurements

Characteristic	
Age	56.9 (8.0)
Proportion of females (%)	55
Prevalent diabetes (%)	4.8
Glucose (mmol/l)	4.6 (1.3)
Systolic BP (mmHg)	139.8 (19.6)
Diastolic BP (mmHg)	82.3 (10.6)
LDL-cholesterol (mmol/l)	3.5 (0.9)
HDL-cholesterol (mmol/l)	1.4 (0.3)
Triglycerides (mmol/l)	1.3 (1.0)
Townsend index	-1.3 (3.0)
eGFR (ml/min/BSA)	83.6 (15.5)
BMI (kg/m^2)	27.4 (4.8)
Smoking (%)	
Never	55.4
Former	34.2
Current	10.4
Statin use (%)	15.6
Ethnicity (%)	
White	94.5
Black	1.9
Asian	2.5
Other	1.1

Means and SD or proportions are given (n=52,164).

nature portfolio

Corresponding author(s): Rui Zheng

Last updated by author(s): Aug 19, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No specific softwares were used.
Data analysis	All the analyses for the observational part were conducted in STATA (version 16.1); the MR and colocalization were conducted in R (version 4.1.0). All core functions used for data analysis are integrated in the respective packages. *R packages: Metafor, version 3.4.0; TwoSampleMR, version 0.5.6; Coloc, version 5.1.0.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

I. The UK Biobank and its data is an open research resource available following submission of a research plan at https://www.ukbiobank.ac.uk.
II. The CKB observational data that support the findings of this study are available to bona fide researchers on application under the China Kadoorie Biobank Open Access Data Policy (www.ckbiobank.org).
III. Summary-level GWAS data of
• proteins available at https://metabolomips.org/ukbbpgwas/.
• coronary heart disease (CHD) available at http://www.cardiogramplusc4d.org/.
• ischemic stroke available at https://www.megastroke.org/.
• heart failure (HF) available at https://www.ebi.ac.uk/gwas/studies/GCST90162626.
• ultrasound-measured carotid artery intima-media thickness (IMT) and carotid plaques available at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000930.v6.p1; accession phs000930.v6.p1.
IV. Data supporting the findings of the POEM study are provided in the article and related files. Raw data are not publicly available due to Swedish law, as they contain sensitive personal information, but could be obtained from the POEM study following a request to: lars.lind@medsci.uu.se).
V. Other online datasets
*Reactome (https://reactome.org/), version 86 on 03/11/2023 for pathway enrichment analysis.
*GTEx Portal (https://gtexportal.org/home, dbGaP accession number phs000424.v8.p2) for the look-up of expression levels of protein-coding genes in several human tissues.
*DrugBank (https://go.drugbank.com/), https://clinicaltrials.gov/ and ChEMBL (https://www.ebi.ac.uk/chembl/) databases for druggability look-up.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Both sexes were included in the cohort populations to represent the general population. Sex information was collected in each cohort. Data have been collected stratified on sex to maximize statistical efficiency. No information of gender was collected.
Reporting on race, ethnicity, or other socially relevant groupings	This study used data from multiple cohorts with multiple ancestries. Detailed information can be found in the cited references.
Population characteristics	The UKB study: UKB is a large, multi-center, prospective cohort study conducted across the UK. Over 500,000 individuals aged 40–69years were included. The CKB study: CKB is a population-based prospective study of 512,000 Chinese adults aged 30-79 years. The POEM study: inhabitants of the city of Uppsala, Sweden, all aged 50 years (n=502).
Recruitment	The UKB study: In 2006–10, over 500,000 individuals aged 40–69years underwent physical measurements, and blood samples were stored for later analysis of genes and biomarkers. The present study used data from the 52,164 individuals with valid proteomics data. Detailed information can be found at https://www.ukbiobank.ac.uk. The CKB study: Participants were recruited from 10 regions (5 rural and 5 urban) in China during 2004-2008. Detailed description can be found from a previous work: Chen Z et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol. 2011;40(6):1652-66. The POEM study: Information from our previous publication "Relationships between three different tests to evaluate endothelium-dependent vasodilation and cardiovascular risk in a middle-aged sample" by Lars Lind (DOI: 10.1097/HJH.0b013e3283619d50): The individuals were invited in a random order from the register of inhabitants in the city 1 month following their 50th birthday. No exclusion criteria were applied except that the individuals needed to have a Swedish identification number.
Ethics oversight	The UK Biobank study was approved by the UK North West Multi-Centre Research Ethics Committee (Application Nr. 90143) and the Swedish Ethical Review Authority (Nr. 2023-00148-01). Ethical approval for CKB was obtained from the Oxford Tropical Research Ethics Committee, the Ethical Review Committees of the Chinese Center for Disease Control and Prevention, Chinese Academy of Medical Sciences, and the Institutional Review Board (IRB) at Peking University.

The POEM study was approved by the Ethics committee of Uppsala University (2009/057) .

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For the observational analysis, 52,164 individuals from the UK Biobank and 502 individuals from POEM study, and 3938 participants from CKB study. Post-hoc power calculation was conducted for the MR analysis.
Data exclusions	Protein data failed quality control were excluded and data of individuals with prevalent CVDs were excluded in the observational analysis. Proteins with no cis-pQTLs or no rsIDs and pQTLs being multiallelic genetic variants were excluded in the MR and colocalization analyses.
Replication	The CKB study was used as an independent external replication of the observed protein-CVD associations. Only one replication was conducted. A total of 126 of the proteins were related to all three CVD outcomes in UK Biobank and of those, 118 were related to any of the CVD traits in the replication phase in CKB.
Randomization	Only observational design was used for our study and thus randomization was not required.
Blinding	Irrelevant. All information was collected before proteins were analyzed. Protein instrumental analysis was automated.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*