University of Groningen

Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals

Folkersen, Lasse; Gustafsson, Stefan; Wang, Qin; Hansen, Daniel Hvidberg; Hedman, Åsa K; Schork, Andrew; Page, Karen; Zhernakova, Daria V; Wu, Yang; Peters, James

Published in:
Nature Metabolism

DOI:
[10.1038/s42255-020-00287-2](https://doi.org/10.1038/s42255-020-00287-2)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Folkersen, L., Gustafsson, S., Wang, Q., Hansen, D. H., Hedman, Å. K., Schork, A., Page, K., Zhernakova, D. V., Wu, Y., Peters, J., Eriksson, N., Bergen, S. E., Boutin, T. S., Bretherick, A. D., Enroth, S., Kalnapenkis, A., Gådin, J. R., Suur, B. E., Chen, Y., ... Mälarstig, A. (2020). Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nature Metabolism*, 2(10), 1135-1148. <https://doi.org/10.1038/s42255-020-00287-2>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals

Lasse Folkersen^{1,2,3,66}, Stefan Gustafsson^{3,4,66}, Qin Wang^{3,5,6,66}, Daniel Hvidberg Hansen^{3,7}, Åsa K. Hedman^{1,3,8}, Andrew Schork^{3,9,10}, Karen Page^{3,11}, Daria V. Zhernakova^{3,12}, Yang Wu^{3,13}, James Peters^{3,14,15,16}, Niclas Eriksson^{3,17}, Sarah E. Bergen^{3,18}, Thibaud S. Boutin^{3,19}, Andrew D. Bretherick^{3,19}, Stefan Enroth^{3,20}, Anette Kalnainen^{3,21,22}, Jesper R. Gådén^{1,3}, Bianca E. Suur^{3,23}, Yan Chen^{1,3}, Ljubica Matic^{3,23}, Jeremy D. Gale^{3,24}, Julie Lee^{3,11}, Weidong Zhang^{3,25}, Amira Quazi^{3,11}, Mika Ala-Korpela^{3,5,6,26}, Seung Hoan Choi^{3,27}, Annique Claringbould^{3,12}, John Danesh^{3,14,15,28,29,30,31}, George Davey Smith^{3,32}, Federico de Masi^{3,7}, Sölve Elmståhl^{3,33}, Gunnar Engström^{3,33}, Eric Fauman^{3,34}, Celine Fernandez^{3,33}, Lude Franke^{3,12}, Paul W. Franks^{3,35}, Vilimantas Giedraitis^{3,36}, Chris Haley^{3,19}, Anders Hamsten^{1,3}, Andres Ingason^{3,9}, Åsa Johansson^{3,20}, Peter K. Joshi^{3,37}, Lars Lind^{3,38}, Cecilia M. Lindgren^{3,27,39,40}, Steven Lubitz^{3,27,41}, Tom Palmer^{3,42}, Erin Macdonald-Dunlop^{3,37}, Martin Magnusson^{3,43,44,45}, Olle Melander^{3,33}, Karl Michaelsson^{3,46}, Andrew P. Morris^{3,40,47,48}, Reedik Mägi^{3,21}, Michael W. Nagle^{3,34}, Peter M. Nilsson^{3,33}, Jan Nilsson^{3,33}, Marju Orho-Melander^{3,49}, Ozren Polasek^{3,50}, Bram Prins^{3,14,15}, Erik Pålsson^{3,51}, Ting Qi^{3,13}, Marketa Sjögren^{3,33}, Johan Sundström^{3,52,53}, Praveen Surendran^{3,14,15,28,54}, Urmo Vösa^{3,21}, Thomas Werge^{3,9}, Rasmus Wernersson^{3,7}, Harm-Jan Westra^{3,12}, Jian Yang^{3,13,55,56}, Alexandra Zhernakova^{3,12}, Johan Ärnlöv^{3,57}, Jingyuan Fu^{3,12,58}, J. Gustav Smith^{3,44,59}, Tõnu Esko^{3,21,27}, Caroline Hayward^{3,19}, Ulf Gyllenstein^{3,20}, Mikael Landén^{3,51}, Agneta Siegbahn^{3,60}, James F. Wilson^{3,19,37}, Lars Wallentin^{3,61}, Adam S. Butterworth^{3,14,15,28,29,30}, Michael V. Holmes^{3,62,63,66}, Erik Ingelsson^{3,64,66} and Anders Mälarstig^{1,3,65,66} ✉

Circulating proteins are vital in human health and disease and are frequently used as biomarkers for clinical decision-making or as targets for pharmacological intervention. Here, we map and replicate protein quantitative trait loci (pQTL) for 90 cardiovascular proteins in over 30,000 individuals, resulting in 451 pQTLs for 85 proteins. For each protein, we further perform pathway mapping to obtain *trans*-pQTL gene and regulatory designations. We substantiate these regulatory findings with orthogonal evidence for *trans*-pQTLs using mouse knockdown experiments (*ABCA1* and *TRIB1*) and clinical trial results (chemokine receptors CCR2 and CCR5), with consistent regulation. Finally, we evaluate known drug targets, and suggest new target candidates or repositioning opportunities using Mendelian randomization. This identifies 11 proteins with causal evidence of involvement in human disease that have not previously been targeted, including EGF, IL-16, PAPP, SPON1, F3, ADM, CASP-8, CHI3L1, CXCL16, GDF15 and MMP-12. Taken together, these findings demonstrate the utility of large-scale mapping of the genetics of the proteome and provide a resource for future precision studies of circulating proteins in human health.

Proteins circulating in blood are derived from multiple organs and cell types, and consist of both actively secreted and passively leaked proteins. Plasma proteins are frequently used as biomarkers to diagnose and predict disease and have been of key importance for clinical practice and drug development for many decades.

Circulating proteins are attractive as potential drug targets as they can often be directly perturbed using conventional small molecules or biologics such as monoclonal antibodies¹. However, a prerequisite for successful drug development is efficacy, which is predicated on the drug target playing a causal role in disease. One

approach to investigate causation is Mendelian randomization (MR), which has successfully predicted the outcome of randomized controlled trials for pharmacological targets such as proprotein convertase subtilisin/kexin 9 (PCSK9), Lipoprotein-associated phospholipase A2 (Lp-PLA2) and Niemann-Pick C1 Like 1 (NPC1L1), and is increasingly becoming a standard tool for triaging new drug targets².

Recent technological developments of targeted proteomic methods have enabled hundreds to thousands of circulating proteins to be measured simultaneously in large studies^{3–6}. This has paved the way for studies of genetic regulation of circulating proteins using

A full list of affiliations appears at the end of the paper.

genome-wide association studies (GWAS) for detection of pQTLs, some of which are referenced here^{3,4,7–9}.

Here, we present a genome-wide meta-analysis of 90 cardiovascular-related proteins, many of which are established prognostic biomarkers or drug targets, measured using the Olink proximity extension assay (PEA) cardiovascular (CVD) I panel¹⁰ in 30,931 individuals across 15 studies. The identified pQTLs were combined with other sources of information to suggest new target candidates underpinned by insights into *cis*- and *trans*-regulation of protein levels and to evaluate past and present efforts to therapeutically modify the proteins analysed in the present investigation. We also show that protein-centric polygenic risk scores (PRS) can predict a substantial fraction of inter-individual variability in circulating protein levels, explaining a proportion of disease susceptibility attributable to specific biological pathways.

These are the first results to emerge from the SCALLOP consortium, a collaborative framework for pQTL mapping and biomarker analysis of proteins on the Olink platform (<https://www.olink.com/scallop/>).

Results

Genome-wide meta-analysis of 90 proteins reveals 467 independent genetic loci associated with plasma levels of 85 proteins. Ninety proteins in up to 21,758 participants from 13 cohorts passed quality-control criteria and were available for genome-wide meta-analysis (Supplementary Table 1). We found a total of 401 pQTLs that were significant at a discovery *P*-value threshold conventional for GWAS ($P < 5 \times 10^{-8}$; Fig. 1 and Supplementary Table 2). Conditioning each of these primary pQTLs using the GCTA-COJO software, we identified an additional 144 proximal pQTLs that independently surpassed conventional genome-wide significance ($P < 5 \times 10^{-8}$), termed as secondary pQTLs. We attempted to replicate the primary and secondary pQTLs in two independent studies (9,173 participants) on which the discovery and replication datasets were meta-analysed, leading to 315 primary pQTLs and 136 secondary pQTLs surpassing a Bonferroni-corrected *P* value ($P < 5.6 \times 10^{-10}$). The discovery *P* values were used for pQTLs absent in the replication dataset ($n_{\text{snp}} = 25$; Supplementary Table 2).

Some proteins such as stem cell factor (SCF), RAGE, PAPPA, CTSL1 and MPO showed an association with more than nine primary pQTLs, but most proteins (22 of 85) were associated with two primary pQTLs. We also observed that some proteins were associated with multiple conditionally significant (secondary) pQTLs such as CCL-4 with four secondary signals, implicating complex genetic regulation of circulating CCL-4 at the *CCL-4* locus.

Analysis of *trans*-pQTLs suggests common mechanisms by which genetic variants affect plasma protein levels. A best-guess causal gene for each of the CVD-I *trans*-pQTLs was assigned by a hierarchical approach based on analysis of protein–protein interactions (PPI), literature mining, genomic distance to gene and manual review of literature around the gene, as well as the genomic context of the association signal. In total, 326 primary *trans*-pQTLs were assigned to unique genes and 30 *trans*-pQTLs were assigned more than one gene, with *ABO*, *ST3GAL4*, *JMJD1C*, *SH2B3* and *ZFPM2* showing an association with the levels of five or more CVD-I proteins (Extended Data Fig. 2a,b and Supplementary Table 2). Extending this analysis to pQTLs from the literature expanded the list of genes with five or more protein associations to include *KLKB1*, *GCKR*, *FUT2*, *TRIB1*, *SORT1* and *F12* (Supplementary Table 4).

Gene ontology (GO) analysis of genes assigned to all significant *trans*-pQTLs showed functional enrichment for chemokine binding, glycosaminoglycan binding, receptor binding and G-protein-coupled chemoattractant activity (Fig. 2c). A broader classification of genes assigned to both *cis*- and *trans*-pQTLs (Fig. 2a,b and Supplementary Table 2) using a wider set of

tools (Methods) suggested that transcriptional regulation, post-translational modifications, such as glycation and sialylation, cell-signalling events, protease activity and receptor binding are potential common mechanisms by which *trans*-pQTLs influence circulating protein levels. The default gene calls and paths for the CVD-I *trans*-pQTLs based on PPI and literature mining can be visualized using the SCALLOP CVD-I network tool (Extended Data Fig. 2c), whereas details on the classification of genes are available in the Methods, Supplementary Information 1 and Supplementary Table 3.

Evidence of mRNA expression mediating associations with a third of *cis*-pQTLs. We investigated the overlap of the CVD-I *cis*- and *trans*-pQTLs with expression quantitative trait loci (eQTL) by a combination of approaches and eQTL studies, including direct genetic lookups and colocalization using PrediXcan¹¹, summary data-based MR (SMR) and heterogeneity in dependent instruments (HEIDI) methods¹². For direct lookups, three studies were used: LifeLines-DEEP (whole blood), eQTLGen meta-analysis (whole blood and peripheral blood mononuclear cells) and the Genotype-Tissue Expression (GTEx) project (48 tissue types). Of 545 pQTLs (Supplementary Table 2), eQTL data were available for 434 single-nucleotide polymorphism (SNP)–transcript pairs, including 168 *cis*-pQTLs and 266 *trans*-pQTLs. Of these, 72 (43%) of *cis*-pQTLs had at least one corresponding eQTL (false discovery rate (FDR) < 0.05) in any of the eQTL datasets investigated, implicating 42 of 75 proteins with a *cis*-pQTL. At a more stringent eQTL *P* value $< 5 \times 10^{-8}$, 26% of *cis*-pQTLs had a corresponding eQTL, similar to previous reports^{13–15} (Supplementary Table 5).

Colocalization analysis of CVD-I *cis*-pQTLs and mRNA levels was performed in selected tissues from the GTEx project by first imputing mRNA expression of the CVD-I protein-encoding transcripts using the PrediXcan¹¹ algorithm in one of the SCALLOP CVD-I cohorts (IMPROVE), and then testing imputed mRNA levels for association with CVD-I plasma protein levels using linear regression. Twenty-six of the 90 CVD-I proteins were associated with their corresponding mRNA transcript (FDR < 0.05) in at least one of the 20 GTEx tissues investigated (Extended Data Fig. 3). All 26 proteins were among the 42 proteins found to also be an eQTL by direct lookups. Proteins CCL-4, CD40, CHI3L1, CSTB and IL-6RA were all associated with their corresponding transcript across five or more tissues whereas proteins ST2 and RAGE showed significant association exclusively in lung tissue, and CTSD exclusively in skeletal muscle.

To further investigate if the CVD-I protein pQTLs overlapped with eQTLs, we used the SMR/HEIDI methods¹², with data from the Consortium for the Architecture of Gene Expression (CAGE) study. SMR/HEIDI tests the hypothesis that there is a single variant affecting protein and gene expression (pleiotropy or causality), with the alternative hypothesis being that protein and gene expression are affected by two distinct variants. In total, 125 associations between 96 genes and 54 proteins were identified at an experiment-wise SMR test significance level ($P_{\text{SMR}} < 0.05/8,558$) and a stringent HEIDI test threshold ($P_{\text{HEIDI}} > 0.01$; Supplementary Table 6), of which 23.2% were in *cis*-pQTL regions, such as IL-8 and U-PAR. The 96 genes were located in 74 loci, suggesting that pleiotropic associations between protein and mRNA expression were present for 18.4% of significant and suggestive primary loci using SMR/HEIDI.

A minor proportion of *cis*-acting pQTLs are in high linkage disequilibrium with nonsynonymous coding variants. ‘Pseudo-pQTLs’ caused by epitope effects, that is, differential assay recognition depending on the presence of protein-altering variants, is a theoretical possibility for *cis*-pQTLs and likely dependent on the method of protein quantification^{4,16}. To evaluate the potential for pseudo-pQTLs among the CVD-I pQTLs, we investigated the

Fig. 1 | Chromosomal location of all associations discovered. Cis-pQTLs are shown in red (bold) and trans-pQTLs are shown in blue if they surpassed a conventional GWAS significance threshold of $P < 5 \times 10^{-8}$. The gene annotations refer to the gene closest to the pQTL. A version of this figure with only loci selected according to the criteria for MR is available in Extended Data Fig. 1.

Fig. 2 | Classification of *cis*- and *trans*-pQTL genes. **a**, The GO label of all *cis*-pQTL genes, that is, the protein-encoding genes. **b**, The GO label of all best-guess *trans*-pQTL genes. **c**, Gene-set enrichment analysis of genes assigned to all significant *trans*-pQTLs, showing the top gene sets from the GO set molecular function. MHC, major histocompatibility complex.

Fig. 3 | Clinical trial in humans and knockdown experiment in mice corresponds to *trans*-pQTL effects. **a**, In humans treated with a small-molecule dual-inhibitor of CCR5 and CCR2 (PF-04634817), the induction of MCP-1 and CCL-4 mirrors the observed CVD-I *trans*-pQTLs. Box plot elements are according to standards for box-and-whisker diagrams. **b**, In mice, knockdown of *ABCA1* or *Trib1* resulted in decreased circulating SCF levels mirroring CVD-I *trans*-pQTLs for SCF. SCF levels of individual mice are represented by circles (wild-type (WT) in blue and transgenic (KO) mice in red), and the black line denotes the median level per group. *P* values were calculated using a two-sided *t*-test.

presence of protein-altering variants for sentinel variants or variants in high linkage disequilibrium (LD) with a sentinel variant. Of the 90 proteins, 85 had at least one pQTL, including 12 with only *cis*-pQTLs, 10 with only *trans*-pQTLs and 63 with both *cis*- and *trans*-pQTLs. Of the 170 primary or secondary *cis*-pQTLs for 75 proteins, 20 *cis*-pQTLs for 18 proteins had a sentinel variant in high LD ($R^2 > 0.9$) with a protein-altering variant, which suggests the potential to affect assay performance (Supplementary Table 2).

Orthogonal evidence supports causal gene to protein relationships for a subset of the CVD-I *trans*-pQTLs. Of the 326 *trans*-pQTLs identified, eight were assigned to gene products targeted by compounds or antibodies that have been in clinical development (Supplementary Table 7). Assuming that *trans*-pQTLs represent causal relationships between gene variants and proteins, we hypothesized that the downstream CVD-I proteins associated with CVD-I *trans*-pQTL genes would be modulated on therapeutic modification of the gene product. Support for this

hypothesis was obtained by previous work showing that circulating FABP4 is upregulated after treatment with glitazones (peroxisome proliferator-activated receptor- γ (PPAR- γ) inhibitors)¹⁷, that circulating interleukin (IL)-6 is increased after treatment with tocilizumab¹⁸ (IL-6 receptor inhibitor) and that circulating tumor necrosis factor (TNF)-R2 is decreased after infliximab (TNF- α inhibitor) treatment in patients with Crohn's disease¹⁹, which supports CVD-I *trans*-pQTLs for these proteins. Along these lines, we present new evidence from a clinical trial supporting our observations that a CCR5 variant is a *trans*-pQTL for plasma CCL-4 and a variant in CCR2 is a *trans*-pQTL for plasma MCP-1 (Supplementary Table 2). CCR5 and CCR2 are targeted in combination by the small-molecule dual-inhibitor PF-04634817 (ref. 20). To test whether dual inhibition of CCR5 and CCR2 resulted in a change of circulating CCL-4 and MCP-1, respectively, we measured these proteins in 350 patients with type 2 diabetes (T2D) in a randomized, double-blind, placebo-controlled phase 2 trial evaluating the efficacy of PF-04634817 in diabetic nephropathy (NCT01712061).

Fig. 4 | Main findings of MR analysis. a, Heat map of MR analyses of 38 complex traits. Chapter of indication according to the tenth revision of the International Classification of Diseases (ICD-10) and clinical trial stage are indicated for each target. OGTT, oral glucose tolerance test. **b**, Forest plot showing CVD-I proteins with strong evidence of causality in the MR analysis. β coefficients per standard deviation (s.d.) of the protein and 95% confidence intervals (CI) are shown. Drug development abbreviations: PC: preclinical; Ph1: phase 1; Ph2: phase 2; Ph3: phase 3; post-MA: post-marketing authorization. ICD-10 chapters of disease: A and B: infectious and parasitic; C and D: neoplasms; D: blood and immune; E: endocrine, nutritional and metabolic; F: mental and behavioural; G: nervous system; H: eye, adnexa, ear and mastoid; I: circulatory system; J: respiratory system; K: digestive system; L: skin and subcutaneous tissue; M: musculoskeletal and connective tissue; N: genitourinary; O: pregnancy, childbirth and puerperium; P: perinatal; Q: congenital, deformations and chromosomal; R: clinical and lab findings; S and T: injury and poisoning; U: provisional assignment (new diseases of unknown aetiology); V and Y: external causes; and Z: health status and health services. SLE, systemic lupus erythematosus.

In addition, we measured known or suspected ligands of CCR5 and CCR2, including CCL-3, CCL-5 (RANTES) and CCL-8, and five additional proteins that were present on the Olink CVD-I panel, and for which assays were readily available. We observed a 9.25-fold increase in circulating MCP-1 levels ($P < 0.0001$) and a 2.11-fold increase in circulating CCL-4 levels ($P < 0.0001$) at week 12 (Fig. 3a), as compared with those in the placebo group. An alternative ligand for CCR2, CCL-8, did not change following exposure to PF-04634817, and neither did other CCR5 ligands, such as CCL-5 (RANTES) and CCL-3. Moreover, EN-RAGE, FGF-23, KIM-1,

myoglobin and TNFR-2 were unchanged following PF-04634817 exposure (Extended Data Fig. 4). We conclude that CVD-I *trans*-pQTLs at CCR5 and CCR2 were concordant with the effects of PF-04634817 in humans.

Two of the genes implicated by CVD-I *trans*-pQTLs, *ABCA1* and *TRIB1* for circulating SCF levels, were also investigated in the mouse. Mice with liver-specific or whole-body knockdown of *ABCA1* (ref. 21) and *TRIB1* (ref. 22), respectively, showed lower plasma levels of SCF than those in matched wild-type controls (Fig. 3b), concordant with the human CVD-I *trans*-pQTLs.

Fig. 5 | SNP heritability and variance explained by genetics. a, SNP heritability in the SCALOP consortium discovery cohorts stratified by contributions of major loci (light red) and polygenic effects (dark red). In the independent MDC cohort, additional variability was explained by adding major loci (light blue) and PRS (dark blue). Significance was reported according to the LD score regression (LDSC) algorithm (blue) or a linear regression model (red). **b**, Differences in how protein levels are affected by polygenic (non-genome-wide significant) loci versus major loci are shown for both the SCALOP consortium discovery cohorts as h_{SNP}^2 and for the MDC cohort as variability explained (V.E.).

Mendelian randomization analysis revealed 25 CVD-1 proteins causal for complex traits with strong evidence. To identify potential causal disease pathways indexed by proteins, we conducted an MR analysis of 85 proteins across 38 outcomes; 25 proteins showed strong evidence of causality for at least one disease or phenotype, and an additional 24 proteins showed intermediate evidence of causality (Fig. 4a, Extended Data Fig. 7 and Supplementary Fig. 1). Using open-source information (<https://clinicaltrials.gov/>, <https://www.ebi.ac.uk/chembl/>, <https://www.drugbank.ca/> and <https://www.opentargets.org/>) and the ClariVate Integrity database (<https://integrity.clarivate.com/>), we identified records on past or present clinical drug development programmes for 14 of the 25 proteins, all of which have been in phase 2 trials or later (Supplementary Table 7). Of the 14 proteins, 7 proteins were targeted for an indication different from the phenotype implicated by our MR analysis. Eleven of the 25 proteins have never been targeted in clinical trials, but may provide promising target candidates for indications closely related to the traits in the MR analysis.

Using open-source information (<https://clinicaltrials.gov/>, <https://www.ebi.ac.uk/chembl/>, <https://www.drugbank.ca/> and <https://www.opentargets.org/>) and the ClariVate Integrity database (<https://integrity.clarivate.com/>), we identified records on past or present clinical drug development programmes for 14 of the 25 proteins, all of which have been in phase 2 trials or later (Supplementary Table 7). Of the 14 proteins, 7 proteins were targeted for an indication different from the phenotype implicated by our MR analysis. Eleven of the 25 proteins have never been targeted in clinical trials, but may provide promising target candidates for indications closely related to the traits in the MR analysis.

Fig. 6 | MR using polygenic risk scores. **a**, Association of a PRS with ST2 levels in the independent MDC cohort. **b**, Association of the ST2 PRS with asthma in the UK Biobank. **c**, Association of the ST2 PRS with IBD in the UK Biobank. The ST2 PRS was divided into 11 quantiles, with the middle group (quantile 6) as the reference category. Effect estimates are presented as quantile-specific mean differences (ST2) and odds ratios (ORs; asthma and IBD) relative to the reference category.

Several published MR findings were confirmed, including that *IL-6RA* variants were associated with higher circulating levels of IL-6 and soluble IL-6-RA were associated with lower risk of coronary heart disease, rheumatoid arthritis (RA) and atrial fibrillation but higher risks of atopy, such as asthma and eczema²³. We also replicated previous findings suggesting a causal contribution of IL-1-RA to RA but an inverse causal relationship with cholesterol levels²⁴ and a protective role of genetically higher MMP-12 against stroke^{4,25}.

Some new MR observations included associations between higher levels of CD40 protein and increased risk of RA, higher MMP-12 and increased risk of eczema and higher TRAIL-R2 protein levels and prostate cancer. Further, Dkk-1 has been targeted by a humanized monoclonal antibody (DKN-01) in clinical trials for advanced cancer (NCT01457417 and NCT02375880) and in our study was found to be causally linked to higher risk of bone fractures and lower risk of estimated bone mineral density. In addition, strong evidence for protective roles of PLGF in coronary heart disease, CASP-8 in breast cancer and ST2 in asthma was observed. RAGE was causally linked to several traits, including lower body mass index (BMI) and a corresponding lower risk of T2D, higher total cholesterol and triglycerides and higher risk of prostate cancer and schizophrenia. A small-molecule brain penetrant RAGE inhibitor was tested in a phase 2 trial of Alzheimer's disease (NCT00566397), but was stopped early for futility. We observed no strong signal for Alzheimer's disease (or vascular disease) in our MR analysis. Our findings identify potential target-mediated effects across multiple other complex phenotypes that might manifest in beneficial and/or harmful effects on patients receiving RAGE-modifying therapies.

We also collated observational evidence for 23 of the 50 protein-trait pairs identified as causal in the MR analysis (Supplementary Table 10). The direction of effect inferred from observational studies was concordant with that from MR estimates for 12 pairs.

Heritability analysis and polygenic risk scores demonstrate large differences in genetic architecture. We calculated SNP heritability contributed by the major reported loci (major loci h_{SNP}^2 , any pQTL included in Supplementary Table 2), as well as additional genome-wide SNP heritability (polygenic h_{SNP}^2) for each protein included in the entire SCALOP CVD-I meta-analysis. We observed a large range of different genetic architectures: differences in magnitude of the genetic component (h_{SNP}^2) ranged from 0.01 (EGF) to 0.46 (IL-6RA). Differences in the contribution from non-genome-wide significant SNPs ranged from essentially

monogenic (for example, IL-6RA) to others showing considerable locus heterogeneity with genetic contributions originating entirely from a polygenic background with no single dominating locus (for example, PDGF-B and galanin; Fig. 5).

In addition, we calculated the out-of-sample variance explained in the independent Malmö Diet and Cancer (MDC) study ($n = 4,678$) both for genome-wide significant loci (major loci V.E._{PRS}), as well as additional variance explained by adding PRS (polygenic V.E._{PRS} ; Fig. 5). The protein PRS applied in the MDC study for 11 proteins exceeded 10% of the variance explained (V.E._{PRS}) and the PRS for another 14 proteins exceeded 5% of variance explained, suggesting that the genetic contribution to inter-individual variability of CVD-I protein levels is considerable.

A polygenic risk score for circulating ST2 levels shows a dose-response relationship with asthma. Since circulating ST2 showed strong evidence of causation in asthma and inflammatory bowel disease (IBD) and the polygenic V.E._{PRS} model for ST2 explained nearly 20% of its variance, we attempted to quantify the effect of the ST2 polygenic V.E._{PRS} on circulating ST2 levels in the MDC study, and risk of asthma and IBD in 337,484 unrelated white-British individuals in the UK Biobank. The range of circulating ST2 across 11 categories of the ST2 PRS in MDC was nearly 1.2 standard deviations (Fig. 6a). Corroborating the MR analysis, the ST2 PRS showed a strong negative dose-response relationship with risk of asthma ($P = 1.2 \times 10^{-8}$) and a positive trend for risk of IBD ($P = 0.13$; Fig. 6b,c). Overlaying the linear trends for ST2 levels, asthma and IBD using meta-regression, an increase in the PRS equivalent to one standard deviation higher circulating ST2 corresponded to a 8.6% (95% CI 3.8%, 13.2%; $P = 0.004$) reduction in the relative risk of asthma and a 4.3% (95% CI -3.8%, 13.0%; $P = 0.263$) increase in the relative risk of IBD (Extended Data Fig. 8).

Reverse Mendelian randomization identifies widespread causal relationships, where complex phenotypes affect CVD-I proteins. To investigate whether genetic susceptibility (liability) to complex disease and phenotypes causally alter circulating levels of CVD-I proteins, we also performed MR using 38 complex phenotypes (including continuous risk factors, such as adiposity, and clinical outcomes, such as T2D) as exposures and CVD-I protein levels as outcomes. All CVD-I proteins were causally altered by at least one complex phenotype. BMI and estimated glomerular filtration rate (eGFR) causally affected 32 and 29 of the 85 tested proteins respectively (Fig. 7a, Extended Data Fig. 7 and Supplementary Fig. 2). BMI

Fig. 7 | MR with proteins as the outcome. a, Heat map showing the causal estimates of 38 complex traits on CVD-I protein levels. **b,c**, Correlation between β values for associations between BMI (**b**) or eGFR (**c**) and circulating levels of CVD-I proteins in the IMPROVE cohort and causal estimates from the MR analysis of BMI/eGFR genetic liability on the same CVD-I proteins.

seemed to causally affect protein levels in both positive and negative directions, whereas only REN (renin) was causally decreased with genetically higher eGFR. In an effort to elucidate whether these estimates were recapitulated in simple observational analyses, we compared effect estimates from linear regression analyses of associations of BMI and eGFR with each respective CVD-I protein in one of the participating study cohorts (IMPROVE). The correlation between the observational and MR estimates was high for BMI ($R = 0.78$) and more modest for eGFR ($R = 0.50$; Fig. 7b,c).

Discussion

Using a meta-analysis approach including over 30,000 individuals, we identified and replicated 315 primary and 136 secondary pQTLs for 85 circulating proteins to yield new insights for translational studies and drug development. Our study demonstrates that pQTLs can be harnessed to enhance evaluation of therapeutic hypotheses for protein targets, and to support those hypotheses with basic insights into potential protein regulatory pathways and biomarker strategies. However, we also observed large differences

between proteins in relation to genetic architecture, suggesting that the relative strength to apply these strategies is likely protein dependent.

Our pQTL-based framework was developed to address several key challenges associated with drug development, including (1) mapping of protein regulatory pathways, (2) identification of new target candidates, (3) repositioning of drugs, (4) target-associated safety and (5) matching of target mechanisms to patients by protein biomarkers or genetic PRS (Fig. 8).

The mapping of *trans*-pQTLs, which typically have smaller effects on protein levels (Extended Data Fig. 9), was aided by the large SCALLOP discovery sample size, yielding an average of four independent pQTLs per protein. A causal gene was assigned for each *trans*-pQTL to generate hypotheses that can be further tested using in vitro or in vivo perturbation experiments. The robustness of causal gene assignments for a few selected *trans*-pQTLs was demonstrated using samples from a randomized controlled trial and testing a dual small-molecule inhibitor of the protein products of assigned genes (*CCR5* and *CCR2*) and transgenic mice with liver-specific knockdown of assigned genes (*ABCA1* and *TRIB1*). Although further studies are needed for orthogonal validation of most of the genes assigned from the CVD-I *trans*-pQTLs, several of the implicated genes have previously been identified as regulators of some of the CVD-I proteins, including *CASP1* (ref. ²⁶), *NLR4* (ref. ²⁶) and *GSDMD*²⁷ for IL-18, *FLT1* (ref. ²⁸) for PLGF, *ADAM17* (ref. ²⁹) for TNFR1 and *SLC34A1* (ref. ³⁰) for FGF-23 (Supplementary Table 2).

Further, we attempted to estimate the proportion of pQTLs that were likely driven by effects on mRNA expression, using multiple eQTL approaches and datasets. The lowest estimate was obtained with SMR/HEIDI, suggesting that 18.4% of pQTLs were also eQTLs, whereas direct lookup and colocalization analysis using PrediXcan yielded estimates between 26–29%. We conclude that the majority of pQTLs identified for the CVD-I proteins were not explained by eQTLs.

Clinical-stage targeting with any drug modality was reported for 35 of the 90 proteins on the Olink CVD-I panel (Supplementary Table 7). Our MR analysis identified 11 proteins with causal evidence of involvement in human disease that have not previously been targeted. Among those, four proteins were causal for a disease phenotype and did not show strong evidence of inverse causality with another phenotype (increasing specificity for intended indication), including CHI3L1 and SPON1 for atrial fibrillation and PAPPa for T2D. Strong causal evidence was also identified for proteins targeted in phase 2 or later development. The MR evidence was concordant with drug indications for several protein targets but for some also suggested alternative indications or that monitoring of target-associated safety might be warranted. Monoclonal antibodies that block the CD40 ligand binding to the CD40 protein—a critical element in T cell activation—have been shown to have positive clinical effects in patients with autoimmune diseases, but increased risk of thromboembolism precluded further clinical development³¹. These observations from clinical trials are in line with our findings that genetically lower levels of CD40 are associated with lower risk of RA and higher risk of stroke. There are ongoing efforts to modify CD40 ligand antibodies to retain efficacy while avoiding thromboembolism³¹. However, our results suggest that decreasing circulating CD40 levels may have target-mediated beneficial effects on RA risk, while increasing the risk of ischemic stroke, that is, the increased risk of thromboembolism (manifested as stroke) is an on-target adverse effect. TRAIL-R2 is a key receptor for TRAIL, which has been shown to selectively drive tumour cells into apoptosis. Therefore, considerable effort to agonize TRAIL-R2 for treating cancer has been made in the past number of years³². We demonstrated that increased circulating TRAIL-R2 is protective against prostate cancer, which may suggest that this cancer type should be

Fig. 8 | Protein-trait relationships that support target validation, repositioning and target-mediated safety and new candidates for drug development. For more information, see data presented in Supplementary Table 7. SCZ, schizophrenia. Afib, atrial fibrillation; WHR, waist-hip ratio; eBMD, estimated bone mineral density.

investigated in clinical trials evaluating the efficacy of TRAIL-R2 agonists.

Biomarkers can be broadly classified as generic biomarkers for disease risk or prognosis, or as biomarkers reflecting the activity of specific disease processes or biology. Biomarkers that enable matching of target mechanisms to patient subgroups with greater than average benefit from treatment are enablers of precision medicine. We showed that small-molecule inhibition of *CCR2* and *CCR5* modulated circulating levels of CCL-4 and MCP-1, which may suggest that *trans*-pQTLs can guide selection of exploratory biomarkers to monitor the efficacy of target mechanisms. We also identified multiple complex traits causally affecting circulating protein levels. For example, eGFR and BMI causally influenced over one-third of the CVD-I proteins, suggesting that future biomarker studies should consider these traits as potential confounders. Moreover, the causal phenotype-to-protein associations may represent pathway-related causality to the complex phenotype of interest; or alternatively, ‘reverse causality’, which might pose an opportunity to evaluate implicated proteins as surrogate biomarkers for efficacy in interventional trials³³. We found that higher BMI causally lowered RAGE, while higher circulating levels of RAGE were causally linked to a lower risk of T2D. Thus, developing a hypothetical therapeutic to increase RAGE might represent a mechanism by which it is possible to offset the risk of T2D arising from the global increases in obesity.

Protein-centric PRS may allow stratification of individuals with a genetic propensity for high circulating protein levels. Only 10% of the protein-centric PRS explained 10% or more of the protein variance in the independent replication cohort, including ST2, a prognostic biomarker for heart failure³⁴. ST2 showed evidence of inverse causality in asthma and positive causality in IBD. By constructing a genome-wide PRS for ST2 levels from the MDC study, applying it to the UK Biobank and comparing asthma and IBD prevalence across 11 quantiles of the ST2 PRS, we estimated the magnitude of ST2 increase required to decrease the risk of asthma to similar levels as those in individuals in the highest ST2 PRS category. Such use of PRS for proteins may be expanded to other disease endpoints

and may be of use in precision medicine, to identify patients that will obtain the most benefit from drugs that pharmacologically alter individual proteins.

In conclusion, our findings provide a comprehensive toolbox for evaluation and exploitation of therapeutic hypothesis and precision medicine approaches in complex disease. Such approaches provide an excellent opportunity to rejuvenate the drug development pipeline for new treatments.

Methods

Selection of proteins. Proteins for the Olink PEA CVD-I panel were selected by mining the literature for protein biomarkers associated with cardiovascular risk or prognosis in human observational studies and in animal models and by bringing in protein biomarker suggestions from leading cardiovascular disease researchers¹⁰. The list of proteins curated from these sources was then pruned down based on availability of high-quality antibodies and relative abundance of the proteins in human plasma.

Intra- and inter-plate coefficients of variation (CV) of the CVD-I panel are available from Olink Proteomics AB (<https://www.olink.com/resources-support/document-download-center/>). In addition, we calculated the inter-plate coefficient of variation using data from a pooled plasma sample in one of the participating cohorts (IMPROVE). The mean inter-plate CV averaged across proteins was 16.6% (range 11–26%; Supplementary Table 1).

Cohorts and data collection. Summary statistics from GWAS of Olink CVD-I proteins were obtained from 13 cohorts of European ancestry. The details of all study cohorts are shown in Supplementary Table 9. Together, the cohorts included a total of 21,758 individuals; although the average per-protein sample size was 17,747, as not all proteins passed quality-control criteria in all cohorts. Each cohort provided data imputed to the 1000 Genomes Project phase 3 reference or later or to the Haplotype Reference Consortium (HRC) reference, which resulted in the testing of 21.4 million SNPs. Because imputation schemes varied by cohort, this resulted in an average of 20.3 million SNPs under investigation for each protein.

Each cohort applied quality-control measures for call rate filters, sex mismatch, population outliers, heterozygosity and cryptic relatedness (Supplementary Table 8). Before running the genetic analyses, NPX values of proteins (on the log_e scale) were rank-based, inverse-normal transformed and/or standardized to unit variance, thus avoiding potential Olink batch differences between cohorts. Genetic analyses were conducted using additive model regressions, with adjustment for population structure and study-specific parameters (Supplementary Table 8). Forest plots of cohort-specific effects are available for all significant and suggestive pQTLs using the online tool. Each contributing cohort uploaded the resulting summary statistics in a standardized format using a secure computational cluster provided by NeIC Tryggve (<https://neic.no/tryggve/>). Meta-analysis was performed in duplicate at two different research centres using completely separate bioinformatic pipelines (L.F. and S.G.).

Data cleaning and meta-analysis. A per-protein filtering threshold of >80% of samples above the Olink detection limit was applied to each cohort, leaving data on 90 of the 92 proteins to be analysed. The remaining files had an average of 3% missing samples (per-cohort statistics are available in Supplementary Table 8). Minor allele frequencies (MAFs) were compared with those reported in 1000 Genomes European samples. A per-SNP filter was applied based on imputation quality level (at default setting for respective imputation algorithm) and minor allele count (at least ten alleles per cohort). This resulted in the omission of 10% of the SNPs. Finally, meta-analysis was performed using METAL (2011-03-25)³⁵, applying the inverse-variance-weighted approach (that is, the STDERR option). Throughout the manuscript, *P* values from this test are reported as is, with multiple-testing burden handled through appropriate thresholds. *Cis*-pQTLs were defined as a signal within 1 Mb of the gene encoding the protein, and all other signals were defined as *trans*-pQTLs. See Extended Data Fig. 5 for an overview of the meta-analysis.

Replication analyses. We sought to replicate the findings in the MDC population-based cohort with 4,678 individuals and in the Swedish Mammography Cohort-Clinical (SMC-C; part of the Swedish national research infrastructure SIMPLER described at <https://www.simpler4health.se/>) population-based study of 4,495 women. In MDC, genotypes were imputed to the HRC (unlimited v1.0.1), and data were analysed using linear regression in EPACTS 3.3.0 (linear Wald test). The genotypes in SMC-C were measured using the Illumina Global Screening Array and were imputed up to HRC v1.1 and 1000 Genomes phase 3 (v5), and linear regressions of rank-based inverse-normal transformed protein values adjusting for age, storage time and PC1–15 were performed using PLINK v2 (4 March 2019).

Conditional and joint association analysis. To identify secondary signals at the 401 loci reported in Supplementary Table 2, we performed analyses conditioning on the primary signal using conditional-joint analysis in GCTA (v1.26.0)^{36,37}. The Stanley cohort was chosen as an ancestrally well-matched LD reference cohort.

Meta-analysis summary data were processed with filtering for MAF (0.01) and r^2 (<0.001) to ensure that secondary association signals identified were not driven by LD with the primary signal. See Extended Data Fig. 6 for an overview of signal selection criteria.

Cross-reference of pQTLs with other complex traits. For each pQTL association, we searched PubMed and the European Bioinformatics Institute (EBI) GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) in November 2018 for published SNPs with any complex trait within 10 kb or having an LD of $r^2 \geq 0.85$.

Comparison between eQTLs and pQTLs. To identify eQTLs that corresponded to each pQTL, we used three independent eQTL studies: LifeLines-DEEP³⁸, GTEx³⁹ and eQTLGen⁴⁰. Each SNP–protein pQTL pair was first converted to SNP–gene pairs using Olink platform protein identification and the gene annotation system of Ensembl v91. Then, the significance of eQTLs for those SNP–gene pairs was assessed in three eQTL datasets, using two different cutoffs: (1) a stringent genome-wide significance threshold ($P < 5 \times 10^{-8}$) and (2) a nominal significance of $P < 0.05$.

In the eQTL dataset of LifeLines-DEEP, individual-level whole-blood RNA-sequencing (RNA-seq), protein and genotype data were available. This allowed for a direct comparison of the concordance of blood eQTLs and pQTLs. To do so, we retested eQTL associations for all pQTL pairs, using a previously published pipeline⁴¹. The resulting eQTLs were considered genome-wide significant if they passed the permutation-based FDR level of <0.05 or nominally significant if they had a *P* value <0.05.

In the eQTL datasets of GTEx v7 and eQTLGen, we did not have access to individual-level data. Thus, the comparisons were conducted using publicly available eQTL results. In these datasets, we considered an eQTL to be genome-wide significant if it was within the reported genome-wide significant list and nominally significant if it had a nominal *P* value <0.05. Altogether, if one pQTL pair had at least one significant eQTL effect in any dataset irrespective of allelic direction, it was considered an overlapping pQTL–eQTL pair.

Expression SMR analysis. We performed an SMR and HEIDI analysis⁴² to identify the expression levels of genes that were associated with protein abundance through pleiotropy using pQTL summary statistics from this study and *cis*-eQTL summary data from published studies^{42,43}.

The eQTL summary data used in the SMR analysis were from the CAGE study, comprising 38,624 normalized gene expression probes and ~8 million SNPs from 2,765 blood samples. The eQTL effects were in standard deviation units of expression levels. We excluded the gene probes in the MHC region and included only the gene probes with at least one *cis*-eQTL at $P < 5 \times 10^{-8}$ (a basic assumption of SMR), resulting in 9,538 gene expression probes.

The SMR test uses a SNP instrument (that is, the top associated eQTL) to detect association between two phenotypes (that is, gene and protein in this case). The HEIDI test utilizes LD between the SNP instrument and other SNPs in the *cis*-region to distinguish whether the association identified by the SMR test is driven by a set of shared genetic variants between two traits (pleiotropic or causal model) or distinct sets of variants in LD (linkage model)⁴². Only associations that surpassed the genome-wide significance level of the SMR test ($P_{\text{SMR}} < 0.05/m$, with *m* being the number of SMR tests) and were not rejected by the HEIDI test ($P_{\text{HEIDI}} > 0.01$) were reported as significant.

PrediXcan and transcript-wide association of CVD-I protein levels. Imputation of gene expression was performed in the IMPROVE study. After standard quality control, genotypes were prephased using Eagle2 and then subsequently imputed by minimac4 using the 1000 Genomes reference. A filter using an imputation quality score of RSQ = 0.8 and MAF of 0.01 was set on the imputed genotypes before prediction with PrediXcan, which used 44 tissue models based on GTEx v7.

Using protein data collected on the CVD-I chip in the same individuals, the associations between protein levels in plasma and the predicted expression of their respective coding gene across 20 tissues (from the PrediXcan model) were modelled by a linear model in R. FDRs were estimated based on *Q* values using the qvalue package. In total, 64 genes in 1 to 18 tissues were tested for associations between protein levels and predicted expression. Heat maps were constructed using the pheatmap package for any gene with a significant association (FDR <0.05) in at least one tissue.

Systems biology. Two sets of network analysis were performed: one using the PPI data from the InBio Map (InWeb_InBioMap) and another using significant associations from text mining. These two networks each had 13,033 and 14,635 nodes, respectively, and 147,882 and 193,777 edges, respectively. In both setups, the shortest path between any of the *cis*-gene intermediaries to the protein was identified; altogether, 12,436 pairs were compared. Of the 372 *trans*-pQTL associations reported (Supplementary Table 2), 335 associations had both *cis*-gene intermediaries and plasma protein in the network allowing their analysis. The likelihood of a path arising by chance was calculated by permutation sampling, using 1 million random networks generated with a conserved degree distribution. A new algorithm was developed for de novo random network generation, which

generated random networks with a nearly conserved degree distribution in a feasible time frame (further details are available in Supplementary Information 1).

Assignment of *cis*-intermediary genes. To assign the most plausible causal gene for each of the CVD-I *trans*-pQTLs, we applied a hierarchical approach based on analysis of InWeb_InBioMap PPI, text mining and genomic distance between gene and lead variant at each locus. Results were then manually reviewed by literature, gene expression analysis (<https://www.proteinatlas.org/>) and published pQTLs, which led to the reassignment of 52 genes. The algorithmic gene assignment was overruled or complemented for instances when the assigned gene was different from the gene assigned by multiple previous studies (Supplementary table 4). GO analysis of the most plausible genes was performed using the DAVID bioinformatics tools (<https://david.ncifcrf.gov/>) and the GO MF gene set definition with default settings. The Panther pathway tool (<http://www.pantherdb.org/>), Uniprot (<https://www.uniprot.org/>) and the Human Protein Atlas were used to classify the genes according to basic functional class.

Human in vivo validation of *trans*-pQTLs. PF-04634817 is a competitive dual inhibitor of CCR2 and CCR5 receptors. In the recent B1261007 study (NCT01712061), samples were collected from individuals with diabetic nephropathy and treated with PF-04634817 for 12 weeks. CCL-2 (MCP-1) was measured in serum by enzyme-linked immunosorbent assay (Eurofins). CCL-4 (MIP-1b) and CCL-8 were measured in plasma using Luminex assays (Bio-Rad). CCL-5 (RANTES) was measured in plasma as part of a multi-analyte panel (Myriad Rules Based Medicine).

Mouse in vivo validation of *trans*-pQTLs. Plasma from transgenic and matched control mice were randomized on a PCR plate. The samples included five mice with targeted deletion of hepatocyte *ABCA1* (ref. ²¹) together with five matched control mice, three mice with whole-body *TRIB1* (ref. ²²) knockdown and three controls and four mice with liver-specific knockdown of *TRIB1* and four matched controls. Protein levels of SCF were measured using the Olink PEA Mouse exploratory panel according to the manufacturer's instructions (Olink Proteomics). The plasma levels of SCF were normalized against average protein concentrations using information on an additional 91 proteins. *TRIB1* whole-body and liver-specific mice were analysed jointly as were the respective wild-type controls. The median plasma levels of SCF were compared using the Mann-Whitney U test for unpaired samples.

Mendelian randomization. To study the causal effects of the protein on selected disease outcomes, we performed two-sample MR analyses. We used between-study heterogeneity to guide the instrumental variable (IV) selection. In the presence of between-study heterogeneity ($P_{\text{het}} < 9 \times 10^{-3}$), variants had to surpass a Bonferroni-corrected P -value threshold in discovery ($P < 5.6 \times 10^{-10}$) and show nominal significance ($P < 0.05$) in the replication studies (9,173 individuals), with directionally concordant β coefficients. In the absence of between-study heterogeneity, we included variants showing conventional genome-wide significance ($P < 5 \times 10^{-8}$) in a meta-analysis of the discovery and replication datasets. From these, we created two sets of IVs for each of the 85 proteins with variants reaching multiple-testing-corrected significance in our discovery GWAS: (1) *cis*-IVs, including one or more independent variants (LD $r^2 = 0.001$ within ± 1 Mb of the transcript boundaries of the gene encoding the protein); and (2) *pan*-IVs, including all independent (LD $r^2 = 0$) variants associated with the protein, that is, combining *cis*- and *trans*-pQTLs. The per-allelic β coefficients from the main GWAS analyses were used as weights in the IVs. For the outcomes, we obtained the relevant SNP-to-trait summary statistics from publicly available GWAS as outcomes (Supplementary Table 9). When lead variants from our main GWAS were not available in these summary statistics, we replaced them with proxies (LD $r^2 > 0.85$). For each individual SNP–protein and SNP–outcome association, we generated an IV Wald ratio estimate, with standard errors obtained using the delta method. When the instrument included more than one SNP, summary IV estimates were generated by combining individual-SNP Wald estimates by inverse-variance-weighted fixed-effect meta-analysis. We report associations with a Benjamin–Hochberg FDR $\leq 5\%$, applied separately to summary estimates from *cis*-pQTL and *pan*-pQTL IVs, using pooled estimates for all 38 diseases. We graded the evidence of causality using a framework outlined in Extended Data Fig. 7, with the following categories: strong (*cis*-IV estimate FDR $\leq 5\%$); intermediate (*pan*-IV estimate FDR $\leq 5\%$ with: (1) no heterogeneity between *cis*-IV estimate and *pan*-IV estimate; and (2) no evidence of the MR estimate being unduly influenced by a *trans*-pQTL in leave-one-out analysis); or weak (*pan*-IV estimate FDR $\leq 5\%$ but no *cis*-pQTL IV available, heterogeneity between *cis*- and all-IVs or evidence of undue influence by a *trans*-pQTL). Heterogeneity between *pan*-IV and *cis*-IV estimates were calculated using Cochran's Q tests, with $P < 0.05$ denoting evidence against the null hypothesis and applying a Bonferroni adjustment for multiple testing. MR was conducted in duplicate by two separate analysts, and analyses were performed in Stata (StataCorp) v13.3 using the *mrivests*, *metan* and *multproc* commands and R. Of the 2,437 IV estimates derived using *cis*-pQTL instruments across the 85 proteins and 38 outcome traits, the IV estimates of 50 protein-to-disease associations met

the FDR $\leq 5\%$ (corresponding to an uncorrected $P \leq 1.1 \times 10^{-3}$). Of the 3,044 IV estimates composed using all pQTL instruments, 281 IV estimates met an FDR $\leq 5\%$ (corresponding to $P \leq 4.7 \times 10^{-3}$; Fig. 4a).

Heritability analyses. We estimated the total SNP heritability (h_{SNP}^2) for the plasma level of each protein from the summary statistics of each individual GWAS by summing the contributions from two independent partitions of the SNPs: primary major loci and polygenic background. We defined the variance explained by primary major loci (major loci h_{SNP}^2) as the sum of the estimated variance explained ($2 \times \beta^2 \times f \times (1-f)$, where f is the MAF, and owing to the fact that the phenotypic variance had been standardized across lead SNPs indexing all primary genome-wide significant loci. We used LDSC regression⁴⁴ to estimate the contribution of the polygenic background (polygenic h_{SNP}^2) for each protein, which we defined as the contribution of all loci not indexed by a genome-wide significant lead SNP. LDSC regression is known to perform poorly when genes of large effect are present, as it was derived under the assumption of a simple polygenic genetic architecture⁴⁴. To account for this and avoid double counting the variance explained by major loci through LD surrogates, before estimating the LDSC regression polygenic h_{SNP}^2 , we censored all SNPs within 10 Mb of genome-wide significant lead SNPs for all primary loci.

Polygenic risk score calculation. PRS were derived using the LDpred algorithm⁴⁵, which adjusts the effects of each SNP allele for those of other SNP alleles in LD with it, and also takes into account the likelihood of a given allele to have a true effect according to a user-defined parameter, which we defined as all seven default LDpred settings, with values ranging from 1 to 1×10^{-5} . The algorithm was directed to use HapMap3 SNPs that had MAF > 0.05 , Hardy–Weinberg equilibrium $P > 1 \times 10^{-5}$ and an imputation score > 0.95 . Variance explained in the independent MDC study was tested according to a stepwise model by first including nongenetic covariates, then additional variability explained by adding SNPs from genome-wide significant SNPs (major loci V.E._{PRS}) and finally the additional variability explained by adding the seven LDpred-derived scores as additional covariates (polygenic V.E._{PRS}).

ST2 polygenic risk score for asthma and IBD in the UK Biobank. Before analysis, individuals who were not white British (based on self-reported ancestry in combination with genetic principal-component analysis) in the maximum unrelated subset were filtered out. All biallelic SNPs with MAF $\geq 1\%$ and MaCH RSQ ≥ 0.8 were retained. The PRS weight for ST2, derived as described above (user-defined parameter 0.3), was applied to the 337,484 white-British UK Biobank participants and transformed to a z -score. Associations with asthma and IBD were tested using logistic regression adjusting for age, sex, PC1–10 and genotype batch using either the continuous PRS or the PRS quantile bins as predictors. The UK Biobank protocol has been described previously⁴⁶ and is available at <https://www.ukbiobank.ac.uk/>. The genotype quality control, phasing and imputation were performed centrally as previously described⁴⁷. Outcomes were defined based on self-reported data at baseline and/or the inpatient and death registry, including primary and secondary causes, as well as prevalent and incident disease. For asthma, data were derived by self-reported touchscreen (6,152), self-reported nurse interview (20002) or ICD-10 'J45'. Conflicting self-reported results were set as missing values unless 'J45' was reported. For IBD, data were derived from a nurse interview (20002) or ICD-10 K50–K52.

Meta-regression analysis for ST2 PRS, asthma and IBD. We estimated the associations per quantile and per standard deviation of the weighted PRS for ST2 (MDC study) on the risk of asthma and IBD (UK Biobank) by taking the quantile associations with ST2, asthma and IBD and conducting meta-regression analyses, whereby the dependent variable was the quantile-specific logarithm of the OR and corresponding standard error of asthma or IBD, and the independent variable was the quantile-specific β coefficient for ST2. This was conducted using the 'metareg' package in Stata SE v13.1 (StataCorp). Plots from the meta-regression are presented in Extended Data Fig. 8.

Observational evidence. Observational evidence for the CVD-I proteins showing strong evidence of causality in MR was collated from the literature or by de novo analysis in the IMPROVE cohort (Supplementary Table 10). To identify evidence from the literature, we searched for the protein name or aliases in combination with the implicated trait and disease in PubMed. For clinical outcome traits, only those reported as significant by the paper were included, and directional information is provided in Supplementary Table 10. For quantitative outcome traits, standardized β coefficients and P values are reported.

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

Data availability

The full summary statistics of the Olink CVD-I protein GWAS have been deposited at the SCALLOP CVD-I online resource (www.scallop-consortium.com), allowing access to interactive SCALLOP CVD-I tools and unrestricted download access for

secondary analyses. Additionally, a full copy has been deposited with <https://doi.org/10.5281/zenodo.2615265> for long-term retention, as well as with the GWAS Catalog. A copy of the polygenic scores have been deposited at the Polygenic Score Catalog (PGS) Catalog.

Received: 14 January 2020; Accepted: 2 September 2020;
Published online: 16 October 2020

References

- Chames, P., Van Regenmortel, M., Weiss, E. & Baty, D. Therapeutic antibodies: successes, limitations and hopes for the future. *Br. J. Pharmacol.* **157**, 220–233 (2009).
- Holmes, M. V., Ala-Korpela, M. & Smith, G. D. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat. Rev. Cardiol.* **14**, 577–590 (2017).
- Folkersen, L. et al. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet.* **13**, e1006706 (2017).
- Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
- Williams, S. A. et al. Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* **25**, 1851–1857 (2019).
- Lehallier, B. et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
- Enroth, S., Johansson, A., Enroth, S. B. & Gyllenstein, U. Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cutoffs. *Nat. Commun.* **5**, 4684 (2014).
- Emilsson, V. et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769–773 (2018).
- Melzer, D. et al. A genome-wide association study identifies pQTLs. *PLoS Genet.* **4**, e1000072 (2008).
- Assarsson, E. et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity and excellent scalability. *PLoS ONE* **9**, e95192 (2014).
- Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
- Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
- Sun, W. et al. Common genetic polymorphisms influence blood biomarker measurements in COPD. *PLoS Genet.* **12**, e1006011 (2016).
- Chick, J. M. et al. Defining the consequences of genetic variation on a proteome-wide scale. *Nature* **534**, 500–505 (2016).
- Zhernakova, D. V. et al. Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome. *Nat. Genet.* **50**, 1524–1532 (2018).
- Solomon, T. et al. Identification of common and rare genetic variation associated with plasma protein levels using whole-exome sequencing and mass spectrometry. *Circ. Genom. Precis. Med.* **11**, e002170 (2018).
- Cabre, A. et al. Fatty acid binding protein 4 is increased in metabolic syndrome and with thiazolidinedione treatment in diabetic patients. *Atherosclerosis* **195**, e150–e158 (2007).
- Nishimoto, N. et al. Mechanisms and pathologic significances in increase in serum IL-6 and soluble IL-6 receptor after administration of an anti-IL-6 receptor antibody, tocilizumab, in patients with rheumatoid arthritis and Castleman disease. *Blood* **112**, 3959–3964 (2008).
- Gustot, T. et al. Profile of soluble cytokine receptors in Crohn's disease. *Gut* **54**, 488–495 (2005).
- Gale, J. D. et al. Effect of PF-04634817, an oral CCR2/5 chemokine receptor antagonist, on albuminuria in adults with overt diabetic nephropathy. *Kidney Int. Rep.* **3**, 1316–1327 (2018).
- Bashore, A. C. et al. Targeted deletion of hepatocyte *Abca1* increases plasma HDL (high-density lipoprotein) reverse cholesterol transport via the LDL (low-density lipoprotein) receptor. *Arterioscler. Thromb. Vasc. Biol.* **39**, 1747–1761 (2019).
- Burkhardt, R. et al. *Trib1* is a lipid- and myocardial infarction-associated gene that regulates hepatic lipogenesis and VLDL production in mice. *J. Clin. Invest.* **120**, 4410–4414 (2010).
- Rosa, M. et al. A Mendelian randomization study of IL-6 signaling in cardiovascular diseases, immune-related disorders and longevity. *NPJ Genom. Med.* **4**, 23 (2019).
- Interleukin 1 Genetics Consortium. Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a Mendelian randomisation analysis. *Lancet Diabetes Endocrinol.* **3**, 243–253 (2015).
- Mahdessian, H. et al. Integrative studies implicate matrix metalloproteinase-12 as a culprit gene for large-artery atherosclerotic stroke. *J. Intern. Med.* **282**, 429–444 (2017).
- Kaplanski, G. Interleukin-18: biological properties and role in disease pathogenesis. *Immunol. Rev.* **281**, 138–153 (2018).
- Heilig, R. et al. The Gasdermin-D pore acts as a conduit for IL-1 β secretion in mice. *Eur. J. Immunol.* **48**, 584–592 (2018).
- Autiero, M. et al. Role of PlGF in the intra- and inter-molecular cross talk between the VEGF receptors Flt1 and Flk1. *Nat. Med.* **9**, 936–943 (2003).
- Dri, P. et al. TNF-induced shedding of TNF receptors in human polymorphonuclear leukocytes: role of the 55-kDa TNF receptor and involvement of a membrane-bound and non-matrix metalloproteinase. *J. Immunol.* **165**, 2165–2172 (2000).
- Tenenhouse, H. S. & Sabbagh, Y. Novel phosphate-regulating genes in the pathogenesis of renal phosphate wasting disorders. *Pflug. Arch.* **444**, 317–326 (2002).
- Xie, J. H. et al. Engineering of a novel anti-CD40L domain antibody for treatment of autoimmune diseases. *J. Immunol.* **192**, 4083–4092 (2014).
- de Miguel, D., Lemke, J., Anel, A., Walczak, H. & Martinez-Lostao, L. Onto better TRAILs for cancer treatment. *Cell Death Differ.* **23**, 733–747 (2016).
- Holmes, M. V. & Davey Smith, G. Can Mendelian randomization shift into reverse gear? *Clin. Chem.* **65**, 363–366 (2019).
- McCarthy, C. P. & Januzzi, J. L. Jr Soluble ST2 in heart failure. *Heart Fail. Clin.* **14**, 41–48 (2018).
- Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, S361–S363 (2012).
- Tigchelaar, E. F. et al. Cohort profile: LifeLines-DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open* **5**, e006772 (2015).
- GTEx Consortium Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
- Urmo, V. et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. Preprint at *bioRxiv* <https://doi.org/10.1101/447367> (2018).
- Westra, H. J. et al. Systematic identification of *trans*-eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
- Lloyd-Jones, L. R. et al. The genetic architecture of gene expression in peripheral blood. *Am. J. Hum. Genet.* **100**, 371 (2017).
- McRae, A. F. et al. Identification of 55,000 replicated DNA methylation QTL. *Sci. Rep.* **8**, 17605 (2018).
- Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- Vilhjalmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
- Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
- Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

Acknowledgements

Secure computing was supported by NeIC Tryggve, which is the Nordic collaboration for sensitive data funded by NeIC and ELIXIR nodes of participating countries. Sources of funding for SMCC, part of the national research infrastructure SIMPLER: We acknowledge the national research infrastructure SIMPLER (the Swedish Infrastructure for Medical Population-based Life-course and Environmental Research) for provisioning of facilities and support. SIMPLER receives funding through the Swedish Research Council under grant 2017-00644. This study was also supported by additional grants from the Swedish Research Council (grants 2017-06100, 2015-05997 and 2015-03257), the Swedish Research Council for Health, Working Life and Welfare (FORTE grant 2017-00721) and Stiftelsen Olle Engkvist Byggmästare (grant 2017/49). S.L. was supported by NIH grant 1R01HL139731 and American Heart Association grant 18SFRN34250007. The Orkney Complex Disease Study (ORCADES) was supported by the Chief Scientist Office of the Scottish Government (CZB/4/276 and CZB/4/710), a Royal Society URF to J.F.W., the MRC Human Genetics Unit quinquennial programme 'QTL in Health and Disease', Arthritis Research UK and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). DNA extractions were performed at the Edinburgh Clinical Research Facility. G.D.S. works in the Medical Research Council Integrative Epidemiology Unit at the University of Bristol (MC_UU_00011/1). We would like to acknowledge the invaluable contributions of the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney. M.A.-K. is supported by a Senior Research Fellowship from the National Health and Medical Research Council (NHMRC) of Australia (APP1158958). He also has a research grant from the Sigrid Juselius Foundation, Finland. A.D.B. was supported by a Wellcome PhD training fellowship for clinicians (204979/Z/16/Z), the Edinburgh Clinical Academic Track (ECAT) programme. J.G.S. and the genotyping of MPP-RES was supported by grants from the Swedish Heart-Lung Foundation (2016-0134 and 2016-0315), the Swedish Research Council (2017-02554), the European Research Council (ERC-STG-2015-679242), the Crafoord Foundation, Skåne University Hospital, Scania County, governmental funding of clinical research within the Swedish National

Health Service, a generous donation from the Knut and Alice Wallenberg foundation to the Wallenberg Center for Molecular Medicine in Lund, and funding from the Swedish Research Council (Linnaeus grant Dnr 349-2006-237, Strategic Research Area Exodiab Dnr 2009-1039) and the Swedish Foundation for Strategic Research (Dnr IRC15-0067) to the Lund University Diabetes Center. The study of the LifeLines-DEEP cohort is supported by the Netherlands Heart Foundation CVON grant 2018-27 to J.F. and A.Z., Netherlands Organization for Scientific Research (NWO-VIDI grant 864.13.013 to J.F., 016.178.056 to A.Z., 917.14.374 to L. Franke, VENI grant 194.006 to D.V.Z., Gravitation grant ExosomeNL 024.004.017 to A.Z. and gravitation 024.003.001 to J.F.), European Research Council (ERC starting grant 715772 to A.Z. and 637640 to L. Franke). L. Franke also receives financial support from Oncode Institute. The CROATIA_Vis study was funded by grants from the Medical Research Council (UK), European Commission Framework 6 project EUROSPAN (contract no. LSHG-CT-2006-018947) and the Republic of Croatia Ministry of Science, Education and Sports research grants (108-1080315-0302). We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools, Institute for Anthropological Research in Zagreb and the Croatian Institute for Public Health. The SNP genotyping for the Vis cohort was performed in the core genotyping laboratory of the Clinical Research Facility at the Western General Hospital, Edinburgh, Scotland. C.H. is supported by MRC University Unit Programme Grant MC_UU_00007/10 (QTL in Health and Disease). P.W.F. was supported by a grant from the European Research Council (CoG-2015_681742_NASCENT). P.S. is supported by a Rutherford Fund Fellowship from the Medical Research Council, grant MR/S003746/1. J.D. holds a British Heart Foundation Professorship and a National Institute for Health Research Senior Investigator Award. Participants in the INTERVAL randomized controlled trial were recruited with active collaboration and funding from NHS Blood and Transplant England (www.nhsbt.nhs.uk), the National Institute for Health Research (NIHR), the NIHR BioResource (<http://bioresearch.nihr.ac.uk>) and the NIHR Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust, NIHR Blood and Transplant Research Unit in Donor Health and Genomics (NIHR BTRU-2014-10024), UK Medical Research Council (MR/L003120/1) and British Heart Foundation (SP/09/002, RG/13/13/30194 and RG/18/13/33946). We would like to thank J. Parks at Wake Forest School of Medicine, Winston-Salem, NC, USA, and D. Rader at Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA, for their kind donations of samples from transgenic mice and controls. Estonian Biobank analyses were funded by EU H2020 grant 692145, Estonian Research Council grant PUT1660, European Union through the European Regional Development Fund Project nos. 2014-2020.4.01.15-0012 and 2014-2020.4.01.16-0125, ERA-CVD grant Detectin-HF. Data analyses were carried out in part in the High Performance Computing Center of the University of Tartu. M.V.H. works in a unit that receives funding from the UK Medical Research Council and is supported by a British Heart Foundation Intermediate Clinical

Research Fellowship (FS/18/23/33512) and the National Institute for Health Research Oxford Biomedical Research Centre. This research has been conducted using the UK Biobank Resource under Application no. 13721. Finally, we would like to thank Moving Science for beautiful work on the front-page art proposal (<https://movingscience.dk>). This work was carried out on behalf of the SCALLOP consortium.

Author contributions

L. Folkersen, S.G., Q.W., D.H.H., Å.K.H., D.V.Z., E.F., E.M.D., E.I. and A.M. contributed to meta-analysis. L. Folkersen, Å.K.H., D.V.Z., Y.W., J.R.G., Y.C., A.C., F.M., E.F., L. Franke, T.Q., R.W., H.-J.W., J.Y. and A.M. contributed to functional analysis. L. Folkersen, S.G., Q.W., G.D.S., T.P., T.Q., J.Y., L.W., A.S.B., M.V.H., E.I. and A.M. contributed to MR analysis. S.G., J.P., N.E., S.E.B., T.S.B., A.D.B., St.E., A.K., M.A.-K., S.H.C., J.D., Sö.E., C.F., L. Franke, P.W.F., V.G., C. Haley, A.H., Å.J., P.K.J., L.L., C.M.L., S.L., E.M.D., M.M., A.P.M., R.M., M.W.N., O.P., B.P., E.P., J.S., P.S., U.V., H.-J.W., A.Z., J.Å., J.F., J.G.S., T.E., C. Hayward, U.G., M.L., A. Siegbahn, J.F.W., L.W., A.S.B., E.I. and A.M. contributed to cohort-level analysis. B.E.S., L.M. and A.M. contributed to mouse experiments. K.P., J.D.G., J.L., W.Z., A.Q. and A.M. contributed to clinical trials. L. Folkersen, Å.K.H., A. Schork, J.R.G., F.M., E.F., A.I., T.W. and A.M. contributed to other downstream analysis. L. Folkersen, S.G., Å.K.H., G.E., C.F., O.M., K.M., P.M.N., J.N., M.O.M., M.S. and A.M. contributed to replication analysis. L. Folkersen, S.G., Q.W., M.V.H., E.I. and A.M. contributed to writing. L. Folkersen, S.G., Q.W., A.S.B., M.V.H., E.I. and A.M. contributed to project planning. All authors gave final approval to publish.

Competing interests

J.P. received travel support from Olink AB. R.W., F.M. and D.H.H. are employees of Intomics A/S. A.M., A.Q., E.F., J.D.G., J.L., K.P., M.W.N. and W.Z. are employees of Pfizer Inc. E.I., after this work was completed, became an employee of GSK Inc.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s42255-020-00287-2>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s42255-020-00287-2>.

Correspondence and requests for materials should be addressed to A.M.

Peer review information Primary Handling Editor: Pooja Jha.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

¹Department of Medicine, Karolinska Institute, Solna, Sweden. ²Danish National Genome Center, Copenhagen, Denmark. ³SCALLOP consortium.

⁴Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ⁵Systems Epidemiology, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia. ⁶Computational Medicine, Faculty of Medicine, University of Oulu and Biocenter Oulu, Oulu, Finland. ⁷Intomics A/S, Kgs. Lyngby, Denmark. ⁸Pfizer Worldwide Research, Development and Medical, Cambridge, MA, USA. ⁹Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region, Roskilde, Denmark. ¹⁰Neurogenomics Division, The Translational Genomics Research Institute (TGEN), Phoenix, AZ, USA. ¹¹Early Clinical Development, Pfizer Worldwide Research, Development and Medical, Cambridge, MA, USA. ¹²Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. ¹³Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia. ¹⁴BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ¹⁵Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK. ¹⁶Department of Immunology and Inflammation, Faculty of Medicine, Imperial College London, London, UK. ¹⁷Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden. ¹⁸Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden.

¹⁹MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, Scotland.

²⁰Department of Immunology, Genetics, and Pathology, Biomedical Center, Science for Life Laboratory (SciLifeLab) Uppsala University, Uppsala, Sweden.

²¹Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. ²²Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. ²³Department of Molecular Medicine and Surgery, Karolinska Institute, Solna, Sweden. ²⁴Inflammation and Immunology Research Unit, Pfizer Worldwide Research, Development and Medical, Cambridge, MA, USA. ²⁵Pfizer Global Product Development, Cambridge, MA, USA. ²⁶NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland. ²⁷Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. ²⁸British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK. ²⁹National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK. ³⁰National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge and Cambridge University Hospitals, Cambridge, UK. ³¹Department of Human Genetics, Wellcome Sanger Institute, Hinxton, UK. ³²MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. ³³Department of Clinical Sciences, Lund University, Skåne University Hospital, Malmö, Sweden. ³⁴Internal Medicine Research Unit, Pfizer Worldwide Research, Development and Medical, Cambridge, MA, USA. ³⁵Department of Clinical Sciences, Lund University Diabetes Center, Malmö, Sweden. ³⁶Department of Public Health and Caring Sciences/Geriatrics, Uppsala University, Uppsala, Sweden. ³⁷Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. ³⁸Department of Medical Sciences, Uppsala University, Uppsala, Sweden. ³⁹Big Data Institute at the Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. ⁴⁰Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ⁴¹Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA. ⁴²Department of Mathematics and Statistics, University of Lancaster, Lancaster, UK. ⁴³Department of Cardiology, Skåne University Hospital Malmö, Malmö, Sweden. ⁴⁴Wallenberg Center for Molecular Medicine, Lund University, Lund, Sweden. ⁴⁵North-West University, Hypertension in Africa Research Team (HART),

Potchefstroom, South Africa. ⁴⁶Department of Surgical Sciences, Uppsala University, Uppsala, Sweden. ⁴⁷Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Division of Musculoskeletal and Dermatological Sciences, University of Manchester, Manchester, UK. ⁴⁸Department of Biostatistics, University of Liverpool, Liverpool, UK. ⁴⁹Department of Clinical Sciences, Clinical Research Center, Lund University, Malmö, Sweden. ⁵⁰Faculty of Medicine, University of Split, Split, Croatia. ⁵¹Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, The Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden. ⁵²Department of Medical Sciences, Clinical Epidemiology, Uppsala University, Uppsala, Sweden. ⁵³The George Institute for Global Health, University of New South Wales, Sydney, New South Wales, Australia. ⁵⁴Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ⁵⁵Institute for Advanced Research, Wenzhou Medical University, Wenzhou, China. ⁵⁶School of Life Sciences, Westlake University, Hangzhou, Zhejiang, China. ⁵⁷Department of Neurobiology, Care Sciences and Society (NVS) Division of Family Medicine and Primary Care, Karolinska Institute, Solna, Sweden. ⁵⁸Department of Paediatrics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ⁵⁹Department of Cardiology, Clinical Sciences, Lund University and Skåne University Hospital, Lund, Sweden. ⁶⁰Department of Medical Sciences, Clinical Chemistry, Uppsala University, Uppsala, Sweden. ⁶¹Department of Medical Sciences, Cardiology and Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden. ⁶²Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK. ⁶³Medical Research Council Population Health Research Unit at the University of Oxford, Oxford, UK. ⁶⁴Department of Medicine, Division of Cardiovascular Medicine, Falk Cardiovascular Research Center, Stanford University School of Medicine, Stanford, CA, USA. ⁶⁵Emerging Science & Innovation, Pfizer Worldwide Research, Development and Medical, Cambridge, MA, USA. ⁶⁶These authors contributed equally: Lasse Folkersen, Stefan Gustafsson, Qin Wang, Michael V. Holmes, Erik Ingelsson, Anders Mälarstig. ✉e-mail: anders.malarstig@ki.se

Extended Data Fig. 1 | MR-selected loci. Chromosomal location of all primary associations that were selected as instrument variables for Mendelian Randomization, that is those passing Bonferroni corrected GWAS significance $P < 5.6 \times 10^{-10}$ with replication at nominal $p < 0.05$, or for non-heterogeneous variants ($p < 9 \times 10^{-5}$), surpassing a P -value threshold of $P < 5 \times 10^{-8}$ in the joint discovery and replication meta-analysis.

Extended Data Fig. 2 | Online tools. Illustration of the online interactive tools for visualization of genomic loci, regions and plausible networks (www.scallop-consortium.com). **a.** Illustration of hotspot loci on chromosome 10 (left) and illustration of hotspot loci with independent effects established using COJO analysis (right) **b.** Circular Manhattan plot for TNF-R2. **c.** The pathway implicated by trans-pQTLs for plasma TNF-R2. The network shows the likely path from pQTL to TNF-R2.

Extended Data Fig. 3 | PrediXcan heatmap. Heat map showing PrediXcan associations across tissues for any protein with significant associations between protein and predicted mRNA levels (FDR < 0.05) in at least one tissue. In each cell, numeric labels correspond to the uncorrected P-value from the association of protein with predicted expression levels. The colour palette shows the relative expression level of the gene across tissues in the GTEx resource.

Extended Data Fig. 4 | CCR5-2 trial additional results. Effect of exposure to PF-04634817 on EN-RAGE, FGF-23, KIM-1, myoglobin and TNFR-2. Box plots elements are according to standards for box-and-whisker diagrams.

Extended Data Fig. 5 | Work flows 1. Work flows describing meta analysis.

Extended Data Fig. 6 | Work flows 2. Work flows describing decisions on significance.

c)

Extended Data Fig. 7 | Work flows 3. Work flows describing reasoning behind Mendelian Randomization evidence strength.

Extended Data Fig. 8 | Polygenic risk score effects on complex outcomes. Meta-regression of quantiles of ST2 polygenic risk score and relative risk of asthma (left) and inflammatory bowel disease (right). Values plotted on the x-axis relate to the quantile-specific mean difference in ST2 as compared to the 6th quantile. Values plotted on the y-axis relate to the quantile-specific log odds of disease as compared to the 6th quantile. The red line is the slope derived from the meta-regression across the ST2 quantiles of the PRS on log odds of disease, weighted by the standard error of the log odds.

Extended Data Fig. 9 | Cis and trans comparison. Comparison of absolute effect sizes of all primary cis- and trans loci listed in Supplementary Table 2. Box plots elements are according to standards for box-and-whisker diagrams.

Reporting Summary

Nature Research 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 Research policies, see [Authors & Referees](#) 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

The analysis performed for data collection and pre-processing is performed R (version 3.5.1) and Metal (version 2011-03-25).

Data analysis

Data analysis was performed using a the following bioinformatics tools:

R (version 3.5.1)
Metal (version 2011-03-25)
GCTA (version 1.26.0)
LDpred (version 1.0.6)
InWeb InBioMap (version 2018)
PLINK (version 2, 4 Mar 2019)
EPACTS (version 3.3.0)
PrediXcan (version 2018)
Stata (version 13.3)

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [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 list of figures that have associated raw data
- A description of any restrictions on data availability

The full summary statistics of the Olink CVD-I protein GWAS have been deposited at the SCALLOP-CVD-I online resource, allowing access to interactive SCALLOP-

CVD-I tools and unrestricted download access for secondary analyses. Additionally, a full copy has been deposited at <https://doi.org/10.5281/zenodo.2615265> for long-term retention, as well as with GWAS catalog. A copy of the polygenic scores have been deposited at the PGS catalog.

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](https://www.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	21,758 participants from 13 cohorts. Sample size was not pre-determined by power calculations, since no lower bound on pQTL effects size is known.
Data exclusions	Data exclusions included 1) Gender mismatch, 2) Heterozygosity, and 3) Cryptic relatedness. The exclusion criteria are fully documented in Supplementary Table 8 and were pre-determined before analysis.
Replication	Malmö Diet and Cancer (MDC) population-based cohort with 4,678 individuals and the Swedish Mammography Cohort Clinical (SMCC) population-based study of 4,495 women. A full overview of replication status, including P-values and OR in the replication cohorts are available in Supplementary Table 2. From a total of 401 primary pQTLs in the discovery stage, from which 315 primary pQTLs surpassed a P-value of 5.6×10^{-10} , Bonferroni corrected for testing 90 proteins.
Randomization	The study was not interventional
Blinding	The study was not interventional

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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

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

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Population characteristics that were included as co-variables for each of the 13 cohorts are fully documented in supplementary table 8
Recruitment	The study is a meta-analysis of 13 cohorts. The recruitment details for each are available in supplementary table 8
Ethics oversight	The study is a meta-analysis of 13 cohorts. The recruitment and ethics details for each are available in supplementary table 8

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