Integrative analysis of the plasma proteome and polygenic risk of cardiometabolic diseases

Scott C. Ritchie[1-5,*], Yingying Liu[6,7], Samuel A. Lambert[1-3,8], Shu Mei Teo[1,2], Petar Scepanovic[1-3], Jonathan Marten[1,3,5], Sohail Zahid[9,10], Mark Chaffin[9], Gad Abraham[1,2,11], Willem H. Ouwehand[4,12-15], David J. Roberts[13,15,16], Nicholas A. Watkins[13], Emanuele Di Angelantonio[3-5,8,14,15], Nicole Soranzo[4,14,15], Stephen Burgess[17], Brian G. Drew[7,18], Sekar Kathiresan[19], Anna C. Calkin[6,18], Amit V. Khera[9,10,20,21], John Danesh[3-5,8,14,15], Adam S. Butterworth[3-5,8,14,15], Michael Inouye[1-5,8,11,22,*]

[1]Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[2]Cambridge Baker Systems Genomics Initiative, Baker Heart & Diabetes Institute, Melbourne, Victoria, Australia
[3]British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[4]British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK
[5]National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge and Cambridge University Hospitals, Cambridge, UK
[6]Lipid Metabolism & Cardiometabolic Disease Laboratory, Baker Heart & Diabetes Institute, Melbourne, Victoria, Australia
[7]Molecular Metabolism & Ageing Laboratory, Baker Heart & Diabetes Institute, Melbourne, Victoria, Australia
[8]Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK
[9]Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[10]Department of Medicine, Harvard Medical School, Boston, MA, USA
[11]Department of Clinical Pathology, University of Melbourne, Parkville, Victoria, Australia
[12]Department of Haematology, University of Cambridge, Cambridge, UK
[13]National Health Service (NHS) Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK
[14]Department of Human Genetics, Wellcome Sanger Institute, Hinxton, UK
[15]National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK
[16]National Institute for Health Research Oxford Biomedical Research Centre, University of Oxford and John Radcliffe Hospital, Oxford, UK
[17]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
[18]Central Clinical School, Monash University, Melbourne, Victoria, Australia
[19]Verve Therapeutics, Cambridge, MA USA
[20]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
[21]Division of Cardiology, Massachusetts General Hospital, Boston, MA, USA
[22]The Alan Turing Institute, London, UK

* Corresponding authors: mi336@medschl.cam.ac.uk (MI) or sr827@medschl.cam.ac.uk (SCR)

Abstract

Polygenic risk scores (PRSs) capture the genetic architecture of common diseases by aggregating genome-wide genetic variation into a single score that reflects an individual's disease risk. These present new opportunities to identify molecular pathways involved in disease pathogenesis. We performed association analysis between PRSs and 3,442 plasma proteins in 3,175 healthy individuals, identifying 48 proteins whose levels associated with PRSs for coronary artery disease, chronic kidney disease, or type 2 diabetes. Integrative analyses of human and mouse data to characterise these associations revealed a role for polygenic effects on several well-known causal disease proteins and identified promising novel targets for future follow-up. We found implicated PRS-associated genes were responsive to dietary intervention in mice and showed strong evidence of druggability in humans, consistent with PRS-associated proteins having therapeutic potential. Overall, our study provides a framework for polygenic association studies, demonstrating the power of PRSs to unravel novel disease biology.

Introduction

Common human diseases have been shown to be largely polygenic in architecture[1–3]. Polygenic risk scores (PRSs) capture this architecture by combining the effects of genome-wide genetic variation on risk of disease into a single score representing the aggregate of an individual's genetic predisposition to disease. While PRSs have shown promise for risk prediction and early stratification[4–7], there is parallel interest to utilise PRSs to dissect underlying disease biology. A key area of investigation is identification of the convergent pathways and biomolecules, which are perturbed by these aggregate polygenic effects in asymptomatic individuals, influencing the development and progression of disease over decades of exposure[8–10]. Therapeutic targeting of these convergent pathways represents a promising strategy for disease prevention by disrupting causal pathways that may promote pathogenesis.

With its central role in homeostasis, the plasma proteome largely comprises secreted proteins with roles in intercellular communication, tissue remodelling, vascular and endothelial function, metabolism, and immune response[11–13]. Many plasma proteins play a causal role in disease processes and are thus considered important drug targets[14]. Previous genome-wide association studies (GWAS) have identified quantitative trait loci (QTL) associated with the plasma proteome and have demonstrated causal effects of plasma proteins levels on various diseases, particularly cardiovascular and metabolic aetiologies[15–17]. However, protein QTL (pQTL) GWAS do not quantify the total burden of polygenic effects across all disease risk alleles within an individual, therefore may not identify the molecular and pathway-level bottlenecks through which polygenic effects could be operating.

Here, we use integrative analyses to characterise polygenic disease associations with the plasma proteome. First, we perform a polygenic association study utilising matched genotype and proteomic data in a cohort of 3,175 healthy participants to identify proteins in circulation whose levels are associated with polygenic risk of five cardiometabolic diseases: coronary artery disease (CAD), atrial fibrillation, type 2 diabetes (T2D), chronic kidney disease (CKD), and stroke. To explore the biological implications of PRS to protein associations, we used both computational and experimental approaches. First, we investigated the overlap of pQTLs and Mendelian randomisation- effects on their respective disease. Second, we used mouse models to identify associations between tissue-specific gene expression with cardiometabolic phenotypes as well as their response to dietary intervention. Lastly, we investigate potential opportunities for drug repurposing using PRS associations of protein targets with therapeutics on the market or in clinical trials.

Results

INTERVAL is a cohort of approximately 50,000 adult blood donors recruited from 25 centres across England between 2012–2014 (**Methods**)[18,19] with built-in exclusion at recruitment of people who had relevant baseline diseases (i.e., heart disease, stroke, diabetes, atrial fibrillation). These conditions (as well as infection or other recent illness) are routine exclusions from blood donation in the UK, thereby avoiding potential bias due to prevalent disease in genomic-proteomic

associations (**Methods**). In this study, we analysed 3,175 individuals with matched genotypes and quantitative SomaLogic SOMAscan profiles of 3,442 plasma proteins (**Table S1**)[15]. A workflow of the study is given in **Figure 1**. For each individual in INTERVAL, we calculated standardised PRS levels for five cardiometabolic diseases: CAD, T2D, CKD, atrial fibrillation, and stroke (**Methods**, **Table S2**).

For polygenic association analyses, we regressed normalised protein levels on each PRS (**Methods**, **Table S3**). Quantile-quantile (QQ) plots showed substantial enrichment of low P-values for CKD, CAD, and T2D polygenic risk scores, with weak or no enrichment for atrial fibrillation and stroke (**Figure 2A**). Across the five cardiometabolic diseases, we identified 49 PRS-protein associations for 48 unique proteins at a false discovery rate (FDR) threshold of 5%: 7 proteins for CKD PRS, 11 proteins for CAD PRS, and 31 proteins for T2D PRS (**Table 1**, **Figure 2B**, **Table S4**). A further 26 associations (6 proteins for CKD PRS, 20 for T2D PRS) showed suggestive significance at FDR threshold 10% (**Table S5**). Protein associations were largely specific to each PRS, with the exception of sex hormone-binding globulin (SHBG) which showed independent, inverse associations with T2D and CAD PRS (**Figure S1**). The polygenic association analysis identified proteins with established roles in their respective disease, such as cystatin-c (CST3) with CKD[20,21], apolipoprotein E (APOE) with CAD[22,23], and sex hormone binding globulin (SHBG) with T2D[24], as well as those with very little or no known role (**Table 1**).

To test the robustness of PRS to protein associations using an orthogonal technology, we utilised Olink quantitative protein assays in a largely separate set of 4,811 genotyped individuals from INTERVAL (n=692 also with SomaLogic protein levels) (**Methods**)[25]. These allowed us to independently test four proteins with significant PRS associations. Effect size estimates were directionally consistent and strongly correlated between the SomaLogic and Olink platforms (r = 0.86, **Figure S2**), and all but one PRS association (T2D PRS to TIMP4) remained nominally significant ($P < 0.05$) (**Figure S2**). PRS to protein associations were robust to circadian and seasonal effects, and with the exception of the T2D PRS to protein associations, robust to adjustment for body mass index (BMI) (**Figure S3**). Six of 31 (19%) T2D PRS to protein associations were attenuated ($P > 0.05$) by BMI adjustment (**Table S6**, **Figure S3**), and in multivariable models, both the T2D PRS and each of these protein remained independently associated with BMI (**Table S6**), indicating these associations between the T2D PRS and protein levels are mediated by BMI.

To disentangle whether PRS to protein associations were attributable by polygenic effects or individual locus-level effects (e.g. pQTLs in linkage disequilibrium with variants composing the PRS), two parallel analyses were performed for each PRS to protein association (**Methods**): (1) adjusting the association for each of the protein's known pQTLs[15], and (2) retesting the association after removing from the PRS all genetic variants within 1Mb of the protein or any known pQTL. Of the 48 proteins associated with any PRS, 28 had one or more pQTLs (58%) in a previous pQTL study in INTERVAL[15], a substantially higher rate than across all SOMAscan protein aptamers (41%). All but two PRS to protein associations (CKD PRS to PDE4D/PDE4A and CKD PRS to PRSS3) remained significant when adjusting for potential pQTLs effects (**Figure S4**, **Table S7**). These results indicate the observed PRS to protein associations were, in the majority of cases, truly polygenic. Further, these results suggest that polygenic risk could enhance or buffer the locus-specific effects on protein levels of disease-associated variants overlapping with pQTLs.

To triangulate the biological implications of PRS to protein associations, we performed a series of follow-up analyses: (1) assessment of the consistency of T2D PRS associated proteins in a separate Icelandic study of proteomics and T2D[26]; (2) Mendelian randomisation to test for causal relationships between protein levels and the PRS associated disease; (3) tissue-specific gene expression analysis from multiple mouse models including dietary intervention experiments.

Our findings were strongly consistent with a recent study of SomaLogic proteins in serum and incident T2D in the AGES-Reykjavik cohort[26], providing support for both the T2D PRS associations and the temporality of polygenic associations with proteins which precede T2D incidence. Twenty-three of our 31 T2D PRS associated proteins (74%) were associated with prevalent T2D, and 17 (55%) were associated with 5-year risk of incident T2D (**Table S8**; Fisher's exact test P-value = 5×10^{-19}). Of our 12 novel protein associations with the T2D PRS (**Table 1**), 10 (83%) and 7 (58%) proteins were associated with prevalent or incident T2D, respectively.

To investigate the overlap between proteins associated with polygenic disease risk and those that could be tested for causal effects on disease, we used two-sample Mendelian randomisation[27] with publicly available pQTL and disease GWAS summary statistics (**Methods**). Of the 48 PRS-associated proteins, only 14 (29%) had sufficient independent genetic instruments to test for causal effects (**Methods, Table S9**). Among these, three proteins showed significant evidence of causal effects on disease risk, all of which were for T2D (**Figure 3**, **Table S10**, **Figure S5**). We found a strong and consistent inverse causal estimate for the known causal effect of decreased sex hormone binding globulin (SHBG) on increased T2D risk[24] (**Figure 3**, **Table S10**). Consistent with the Mendelian randomisation analysis in the AGES-Reykjavik cohort[26], we found an inverse causal estimate for WFIKKN2 on T2D risk. We additionally found an inverse causal estimate for complement factor I (CFI) on T2D risk. For SHBG and WFIKKN2, but not for CFI, the direction of the estimated causal effects of protein levels on T2D risk were consistent with the directions of association between T2D PRS levels with the proteins' levels (**Figure 3**, **Figure 2B**). This suggests that T2D polygenic risk influences T2D pathogenesis by lowering SHBG and WFIKKN2 levels.

To investigate the tissue-specific role of PRS-associated proteins in cardiometabolic disease, we utilised two mouse reference datasets (**Methods**): (1) the hybrid mouse diversity panel[28] (HMDP; n=706 mice from 100 inbred strains) fed a standard chow diet, and (2) an F2 cross of the inbred ApoE$^{-/-}$ C57BL/6J and C3H/HeJ strains (n=334) fed a western diet (high fat and high cholesterol) to enrich for cardiovascular disease traits[29]. In total, 45 of 48 PRS-associated proteins had mouse orthologs[30] (**Table S11**), for which we observed widespread correlations between tissue-specific gene expression with a variety of cardiometabolic traits, including lipid dysregulation, insulin resistance, and atherosclerosis (**Figure 4**). Among the proteins with no previous known link to cardiometabolic disease (**Table 1**), we observed particularly strong correlations with relevant cardiometabolic traits for tissue-specific expression of DUSP26, TP53I11, CRYZL1, CPM, INHBC, IGFBP2, and MUSK (**Figure 4**). Across all PRS-associated proteins, adipose tissue expression most commonly associated with cardiometabolic traits for proteins associated with PRSs for CKD and CAD (**Figure 4B**). In contrast, for proteins associated with the T2D PRS, trait-associated gene expression was widely distributed across tissues, with the strongest associations in liver (**Figure 4**). Our subsequent dietary intervention experiments in C57BL/6J mice (**Methods**), revealed significant differences in tissue-specific gene expression between standard chow, high fat, and western diets after 12 weeks for the proteins CRYZL1 and FAH (**Figure S6**), with the strongest

effects between western diet and standard chow. These data suggest that these proteins may play a role in lipid metabolism and the setting of dyslipidaemia.

Finally, to identify druggable targets associated with polygenic disease risk and potential drug repurposing opportunities, we intersected PRS-associated proteins with the DrugBank database[31]. We found 17 of the 48 PRS associated proteins were targets for existing drugs (**Table 2**; **Table S12**). We identified seven drugs whose effects on these proteins were consistent with the PRS associations (**Table 3**). This included the well-known T2D medication metformin[32] as well as two drugs for T2D, Managlinat dialanetil and MB-07803, that have completed phase II clinical trials. Metformin is a GPD1 inhibitor, and Managlinat dialanetil and MB-07803 are FBP1 antagonists, consistent with the associations between the T2D PRS and both GPD1 and FBP1 (**Table 1, Figure 2**). Our PRS analyses suggested multiple on-market drugs may be repurposing opportunities for T2D or CKD (**Table 3**). One of these, Pegvisomant, has completed a phase II clinical trial for modulating insulin sensitivity in subjects with pre-diabetes (NCT02023918). Pegvisomant is used to treat acromegaly by blocking the binding of endogenous growth hormone to growth hormone receptor (GHR)[33–35], and has been shown to improve insulin sensitivity in acromegaly patients[36,37]. The association between the T2D PRS and GHR levels was consistent with repurposing, suggesting Pegvisomant may reduce T2D risk by inhibiting GHR function. Further, individuals with GHR loss-of-function mutations have been found to be at decreased T2D risk[38]. A further three drugs may be potential T2D or CKD repurposing opportunities: (1) Iloprost, a drug for treating pulmonary arterial hypertension[39], may also reduce CKD risk by inducing the function of PDE4A and PDE4D[40,41]; (2) Adenosine phosphate, a supplement used to treat dietary shortage or imbalance, may also reduce T2D risk by inhibiting FBP1 thereby increasing gluconeogenesis independently of insulin[42]; and (3) Fostamatinib, a drug used to treat immune thrombocytopenic purpura[43], may also reduce T2D risk by inhibiting the signalling activity of MUSK[44].

Discussion

PRSs for disease are explicitly constructed to maximise risk prediction, irrespective of the underlying biology of the individual genetic variants composing the PRS. However, in capturing the maximal aggregate disease-associated variance, PRSs also hold considerable promise for identifying molecular pathways in the development and progression of the diseases they predict[8,45].

In a series of computational analyses, we identified 48 plasma proteins that were significantly associated with predicted lifetime risk for CAD, CKD, and T2D. These proteins included both well-known disease associated proteins as well as novel associations, i.e., proteins with no previous known link to the associated disease. PRS to protein associations were robust to technical, physiological, and seasonal confounding. PRS to protein associations were also largely independent of any pQTLs for the given protein, indicating the associations between PRSs and protein levels were truly polygenic.

In several cases our results were consistent with, and provided support for, proteins with previously limited evidence for disease association. Here, we highlight several examples including: (i) GRN (granulin) has been suggested to play a role in atherosclerotic plaque formation[46], (ii) plasma PRCP (lysosomal pro-X carboxypeptidase) levels have been shown to be elevated in individuals with T2D[47], (iii) recessive deletion in PDE4D (cAMP-specific 3',5'-cyclic phosphodiesterase 4D) has been associated with CKD[48]; (iv) ACY1 (aminoacylase-1) levels have been shown to be higher in hepatic lipid droplets in diabetes susceptible mice compared to controls[49], and (v) elevated IGFBP1 (insulin-like growth factor-binding protein 1) levels have been shown to increase insulin sensitivity in mice[50].

Our analyses into the biological implications of PRS to protein associations revealed an unappreciated role for polygenic effects on several well-known disease proteins and highlighted several novel protein associations that represent promising targets for follow-up studies. Mendelian randomisation analysis revealed that SHBG and WFIKKN2 likely lie on causal pathways from polygenic risk to T2D, with increased PRS levels leading to decreased SHBG and WFIKKN2, levels which themselves are causal for increased T2D risk. Mouse tissue-specific gene expression analyses revealed DUSP26, TP53I11, CRYZL1, CPM, INHBC, IGFBP2, and MUSK as promising novel CAD and T2D candidates for future experimental follow-up. These mouse data also revealed that many correlations between tissue-specific gene expression and cardiometabolic traits were stronger in mice fed a high fat and high cholesterol diet. Subsequent dietary intervention experiments in mice revealed particularly strong effects of lipid dietary intervention on tissue-specific gene expression of orthologs for human proteins CRYZL1 and FAH. Finally, overlaying our PRS to protein associations with drug target information revealed seventeen proteins were targets for existing drugs, and revealed promising potential drug repurposing opportunities for on-market drugs Fostamatinib and Adenosine phosphate, via MUSK and FBP1 respectively, for treating or preventing T2D; as well as the on-market drug Iloprost, which targets PDE4A and PDE4D, for treating or preventing CKD.

Overall, our study provides a framework for polygenic association studies, and demonstrates the power of polygenic scores to unravel novel disease biology through integration with information on intermediary molecular pathways preceding disease onset.

Methods

INTERVAL cohort data quantification, processing, and quality control

INTERVAL is a cohort of approximately 50,000 participants nested within a randomised trial studying the safety of varying frequency of blood donation, led by Cambridge University in collaboration with National Health Service (NHS) Blood and Transplant, the national blood service of England[18,19]. Participants were blood donors aged 18 years and older recruited between June 2012 and June 2014 from each of the 25 fixed NHSBT centres across England. Upon joining the study, participants completed an online questionnaire about their demographic and lifestyle

including their age, sex, weight, height, alcohol intake and smoking habits, and diet. The collection of blood samples for research purposes has been extensively described previously[18]. Briefly, blood samples were collected in 6-ml EDTA tubes using standard venepuncture protocols. These tubes were inverted three times then transferred to the UK Biocentre (Stockport, UK) at ambient temperature for processing. Plasma was extracted by centrifugation into two 0.8-ml aliquots then stored at −80 °C before use. Participants gave informed consent and this study was approved by the National Research Ethics Service (11/EE/0538).

Although prevalent disease and medication usage information was not specifically collected or recorded for this cohort, NHSBT criteria (https://www.blood.co.uk/who-can-give-blood/) meant people with a history of major diseases, recent illness, or infection were very unlikely to be included in this study[18]. Specifically for the five diseases studied here, blood donation criteria excluded individuals diagnosed with atrial fibrillation, with a history of any stroke, or history of major heart disease (including heart failure, coronary thrombosis, myocardial infarction, cardiomyopathy, ischaemic heart disease, and arrhythmia, or surgery for a non-congenital heart conditions). Individuals with hypercholesterolaemia and hypertension were not excluded provided their respective conditions were well controlled, unless taking regular aspirin or other blood thinners. An extended list of specific medications and eligibility criteria for those with hypertension or hypercholesterolaemia can be found at https://my.blood.co.uk/knowledgebase. Blood donation criteria did not exclude people with type 2 diabetes provided it was well controlled by diet alone, did not require regular insulin treatment, and the individual had not required insulin treatment in the four weeks prior to attempted blood donation. Blood donation criteria did not exclude people with pre-diabetes or those with a history of gestational diabetes provided they did not require treatment at the time of donation. Although NHSBT criteria do not explicitly exclude those with chronic kidney disease, people with CKD were unlikely to be eligible blood donors because they frequently have comorbidities that would exclude them from donation.

Body mass index (BMI) was calculated from self-reported weight and height. Self-reported weight ranged from 49 kg to 177 kg (mean 78.6 kg) and self-reported height ranged from 1.07 m to 2.41 m (mean 1.73 m). Computed BMI ranged from 13.1 to 81.5 (mean 26.4). Sixteen samples with outlier BMI were removed by excluding samples with self-reported weight < 50 kg or > 160 kg; weights outside the range for blood donation eligibility criteria, and samples with self-reported height < 1.47 m or > 2.1 m; clinical thresholds for dwarfism and gigantism respectively. After outlier exclusion the range of participant BMI was 17.0–55.8 (mean 26.3).

Quantification of protein levels using the SomaLogic SOMAscan arrays in INTERVAL has previously been described in detail[15]. Briefly, the relative concentration of 3,617 human plasma proteins and protein complexes targeted by 4,034 aptamers on the SOMAscan array[51,52] were measured in 3,562 INTERVAL participants in two batches (n=2,731 and n=831) from 150-µl aliquots of plasma by SomaLogic Inc. (Boulder Colorado, US). Quality control was performed by SomaLogic using calibrator samples and control aptamers to remove systematic variability in hybridization and within-run and between-run technical variability. Within each batch, raw aptamer levels (relative fluorescence units) were natural log transformed then adjusted for age, sex, the first three genetic principal components and duration between blood draw and sample processing (1 day vs >1 day), and the residuals were inverse normal rank transformed. For this study we did not exclude protein aptamers with greater than 20% coefficient of variation in either batch. We subsequently filtered the data to 3,793 high quality aptamers targeting 3,442 proteins after obtaining

the latest information about aptamer sensitivity and specificity from SomaLogic. In total 241 aptamers were excluded which targeting non-human proteins, or which since the original quantification in INTERVAL had been (1) deprecated by SomaLogic, (2) found to be measuring the fusion construct rather than the target protein, or (3) measuring a common contaminant. Information about each aptamers targeted protein were extracted from the UniProtKB page associated with each aptamer's associated UniProt IDs supplied by SomaLogic and from the NCBI gene webpage corresponding to the Entrez IDs associated with each UniProt ID. Mapping between UniProt IDs was performed with Ensembl biomart, or by searching the primary gene symbol listed on the UniProt page where biomart returned no Entrez identifiers. Information was manually verified where there was discordance between information provided by SomaLogic or obtained from UniProt or NCBI gene.

Quantification of protein levels using Olink proximity extension assays[25] was performed in 5,000 INTERVAL participants using three Olink panels (Olink Bioscience, Uppsala, Sweden), each quantifying 92 proteins: their "inflammation panel", "cardiovascular II panel", and "cardiovascular III panel". Participants were selected to be at least 50 years old at baseline with limited overlap with participants with SomaLogic SOMAscan array data. There was limited overlap in protein content between the significant PRS associated proteins (quantified by the SOMAscan assay), occurring only for four proteins quantified by the cardiovascular III panel. For the cardiovascular III panel, normalised protein levels on $\log_2$ scale (NPX) were regressed on age, sex, sample measurement plate, time from blood draw to sample processing (number of days), and season (categorical: spring, summer, autumn, winter), then inverse rank normal transformed. After quality control there were 4,811 INTERVAL participants with protein measurements on the Olink cardiovascular III panel including 692 participants with SOMAscan assay data.

Genotyping of INTERVAL participants has been described in detail previously[53]. In total 48,813 samples were genotyped using the Affymetrix UK Biobank Axiom array in ten batches. Samples failing the following QC were removed: those with sex mismatch, extreme heterozygosity, were of non-European descent, or were duplicate samples. Related samples were removed by excluding on participant from each pair of close relatives (first or second degree; identity-by-descent $\hat{\pi} > 0.187$). Variants were filtered to a set of 655,966 high-quality autosomal variants for imputation. Variants were excluded if they were monomorphic, if they were bi-allelic and had HWE p-value $< 5\times10^{-6}$, or if they had call rate < 99% across batches or < 75% across all batches. SHAPEIT3 was used to phase variants, then imputation to the UK10K/1000 Genomes panel was performed using the Sanger Imputation Server (https://imputation.sanger.ac.uk) resulting in 87,696,888 imputed variants for 43,059 samples. In total there were 3,175 participants with matched genotype and proteomic data.

Polygenic risk scores

The PRS used for CAD was our previously published CAD metaGRS[54]; a polygenic score comprising 1.75 million variants derived from a meta-analysis of three PRSs for CAD in UK Biobank. The variants and weights composing the CAD metaGRS are available from the polygenic score (PGS) Catalog[55] (https://www.pgscatalog.org/) with identifier PGS000018. Briefly, the three meta-analysed CAD PRSs were: (1) our previously published PRS[56] comprising 46,000 metabochip variants and their log odds for CAD in the 2013 CARDIoGRAMplusC4D consortium GWAS meta-analysis[57] (PGS000012); (2) a PRS comprising 202 variants whose association with CAD in the 2015 CARDIoGRAMplusC4D consortium GWAS meta-analysis[58] were significant at a false

discovery rate (FDR) < 0.05; and (3) a genome-wide PRS derived from the same summary statistics[58] linkage disequilibrium thinned at r^2=0.9 threshold in UK Biobank (version 2 genotype data, imputed to the HRC panel only).

PRSs for CKD, T2D, atrial fibrillation, and stroke were derived from summary statistics from their respective largest available GWAS by filtering to variants that overlapped with a set of 2.3 million variants genome-wide derived by linkage disequilibrium thinning (r^2 < 0.9) the high-confidence (imputation INFO score > 0.4), common (MAF > 1%), unambiguous SNPs (A/T and G/C SNPs excluded) in the UK Biobank version 3 genotype data[59,60] (imputed to the 1000 Genomes, UK10K, and haplotype reference consortium (HRC)[61] panels). GWAS summary statistics used for CKD, T2D, atrial fibrillation, and stroke were those published by Wuttke *et al.* in 2019[62], Mahajan *et al.* in 2018[63], Nielsen *et al.* in 2018[64], and Malik *et al.* in 2018[65], respectively. For stroke, we used the GWAS summary statistics for any type of stroke rather than those for specific subtypes. In all cases, we used the GWAS summary statistics for the samples of recent European ancestry rather than the GWAS summary statistics from trans-ancestry meta-analyses. The resulting PRSs each comprised 1.96–2.21 million SNPs with weights corresponding to their log odds ratio for the trait in the GWAS summary statistics (**Table S2**). The variants and weights composing each PRS are available to download through Figshare at https://dx.doi.org/10.6084/m9.figshare.11369103. This PRS derivation pipeline was designed to ensure consistency across PRSs in terms of genome-wide polygenic architecture and derivation from the most powerful to date GWAS for each disease. The resulting PRSs were strongly correlated with published PRSs shown to predict the corresponding disease: Pearson correlation r = 0.23 with the 1,168 variant atrial fibrillation PRS developed by Weng *et al.* 2018[66] (PGS000035), r = −0.21 with the 149 variant estimated glomerular filtration rate (eGFR) PRS developed by and shown to predict CKD by Wuttke *et al.* 2019[62] (CKD is diagnosed by low eGFR), r = 0.80 with the 171,129 variant T2D PRS developed by Mahajan *et al.* 2018[63] and validated by Udler *et al.* 2019[67] (PGS000036), r = 0.61 with the 3.2 million variant meta PRS for stroke (meta-analysis of stroke subtype and risk factor PRSs) developed by Abraham *et al.* 2019[68].

Levels of each PRSs in each INTERVAL participant were calculated using the linear scoring function in plink version 2.00[69] which summed, for each variant in each PRS, the number of copies of its effect allele multiplied by the weight (*i.e.* log odds of disease) in the PRS. In the case of missing genotypes, the frequency of the effect allele in INTERVAL was used in its place. For each PRS, these total sums were subsequently standardised to have mean of 0 and standard deviation 1 across all INTERVAL participants. Variants with complementary alleles (A/T and G/C variants) were excluded to avoid incorrect effect allele matching due to strand ambiguity. Variants with INFO < 0.3 were removed. Where there were duplicate variants the one with the highest INFO score was kept. In total, 54,069,889 variants passed QC for PRS level calculation. Nearly all variants (>99%) in each PRS passed QC in INTERVAL (**Table S2**).

Polygenic association scan

The levels of each PRS were subsequently tested for association with the normalised levels of each of the 3,793 SomaLogic aptamers (targeting 3,442 plasma proteins) in the 3,175 INTERVAL participants with matched genotype and proteomic data (**Table 1**, **Figure 2**, **Table S3**, **Table S4**, **Table S5**). Linear regression models were fit between each protein aptamer and each PRS, with the protein levels as the response and PRS as the variable, adjusting for sample measurement batch and the first 10 genotype PCs as covariates. As detailed above, prior to model fitting, aptamer levels were log transformed and adjusted for age, sex, time between sample processing, and the first three

genotype PCs then inverse rank normalised. To obtain associations between each PRS and protein, linear regression beta coefficients, 95% confidence intervals, and P-values were averaged where there were multiple aptamers targeting a single protein. For each PRS, multiple testing correction of P-values across the 3,442 proteins was performed using the Benjamini-Hochberg false discovery rate (FDR) method[70], and we considered any association between a PRS and protein to be significant where FDR < 0.05.

Multivariable models were fit to estimate the independent associations of the CAD PRS and T2D PRS on the aptamers for GGT2, P5I11, and SHBG (**Figure S1**) that were associated with both PRSs in the univariate analysis at either FDR < 0.05 (**Table S4**) or FDR < 0.1 (**Table S5**). Models were adjusted as described above for the univariate associations.

Sensitivity analysis were performed by additionally adjusting the linear regression models to BMI, circadian effects, and seasonal effects separately (**Figure S3**, **Table S6**). To capture the potentially non-linear effects of circadian rhythm and season on protein levels, both were treated as a categorical variables by grouping samples into 10 equal duration bins, with the largest sample size bin acting as the reference group in the model. To model circadian effects, the time of day of sample draw was split into 10 bins, each 73.5 minutes in length, with sample sizes ranging from 63 to 417 (median: 274, interquartile range: 245–328) samples. Bin six, covering 2:10pm–3:25pm with 417 samples, was used as the reference group. Samples with no time of blood draw recorded (n=480; 15%) were excluded when adjusting for circadian effects. To model seasonal effects, the date of blood draw was split into 10 bins, each 50 days in length, with sample sizes ranging from 136 to 484 (median: 291, interquartile range: 283-381) samples. Bin three, covering the 23rd September 2012 – 13th November 2012 with 484 samples, was used as the reference group.

Multivariable models were fit to estimate the independent associations of BMI on the T2D PRS and protein levels (**Table S6**) for each protein whose PRS to protein association was attenuated by BMI adjustment (nominal P-value > 0.05 after BMI adjustment; proteins INHBC, WFIKKN2, APOF, QPCTL, GHR, CCDC126, CFB, LCN1, and CYB5R3) to test whether BMI mediated the association between the T2D PRS and each protein. Models were adjusted for the same covariates as described above for the univariate PRS to protein associations.

Associations with matched protein levels found on the Olink panels (**Figure S2**) were similarly performed by univariate regression of each protein on the PRS levels, adjusting for the first 10 genotype PCs as covariates. As detailed above, protein levels were adjusted for additional covariates prior to normalisation. Proteins quantified on the Olink neurology panel (CPM, VCW2, GFRA1) were adjusted for age, sex, 11 genotype PCs, and season prior to inverse rank normalisation. Proteins quantified on the Olink cardiovascular III panel were adjusted for age, sex, season, sample measurement plate, and time from blood draw to sample processing prior to inverse rank normalisation.

Genetic architecture of polygenic protein levels

To investigate the contribution of single variant effects to polygenic protein level associations two parallel analyses were performed for each significant PRS to protein association.

First, multivariable linear regression models were fit to estimate the independent contributions to the protein's levels of the PRS levels and dosages of any pQTLs previously identified in INTERVAL by Sun B *et al.*[15] (**Figure S4A**, **Table S7**). In total 28 of the 48 PRS associated proteins at FDR < 0.05, and 40 of the 72 PRS associated proteins at FDR < 0.1, included proteins

with any pQTLs. Probabilistic dosages for each conditionally independent pQTL were extracted directly from the BGEN files using plink version 2[69].

Second, for each PRS-protein association we refit the univariate linear regression models after re-calculating PRS levels excluding any variants with potential pQTL effects (**Figure S4B**). Variants were excluded from the PRS if they were in *cis* with the protein (within 1MB of the corresponding gene's transcription start site) or were within 1MB of any previously identified[15] pQTLs. Variants within high complexity genomic regions (coordinates on GRCh37: chromosome 5 44,000,000–51,500,000; chromosome 6 25,000,000–33,500,000; chromosome 8 8,000,000–12,000,000; and chromosome 11 45,000,000–57,000,000) were also excluded if they overlapped with the 1MB windows above.

Mendelian randomisation analyses

For each PRS to protein association with FDR < 0.05 in the PRS to proteome scan (**Table 1**), the protein was tested for a causal effect on the respective disease using two-sample Mendelian Randomisation provided (1) at least three independent genetic instruments could be identified for the test, (2) at least one instrument was a *cis*-pQTL, (3) the aptamer designed to target the protein did not have significant cross-reactivity with a different protein or isoform (PDE4D and APOE excluded; see **Table S4** for details). In total 15 tests satisfied these criteria (**Figure 3**, **Figure S5**, **Table S10**, **Table S9**).

Genetic instruments for each test were identified through a multi-step procedure designed to maximise the number of instruments. In summary, for each test we used (1) the conditionally independent *cis*- and *trans*-pQTLs previously identified in INTERVAL by Sun *et al.* 2018[15] for the protein, as well as (2) independent-by-linkage *cis*-pQTLs that did not meet genome-wide significance in that study. Genetic instruments were determine for each test separately to accommodate differences in overlap between the variants present in the pQTL and GWAS summary statistics. A detailed description of this procedure for each test follows.

First, the candidate variant set was restricted to those present in both the pQTL and GWAS summary statistics. Variants with complementary alleles (A/T or G/C SNPs) whose effect allele could not be matched between studies due to strand ambiguity were excluded. Effect sizes and standard errors for each variant on each protein's levels were obtained from the summary statistics published by Sun *et al.* 2018[15]. Log odds and standard errors for each variant for CAD, CKD, and T2D were obtained from summary statistics published by Nelson *et al.* 2017[71], by Wuttke *et al.* 2019[62], and by Mahajan *et al.* 2018[63]. In all cases, we used the GWAS summary statistics for the samples of recent European ancestry rather than the GWAS summary statistics from trans-ancestry meta-analyses. For T2D, we used the BMI-adjusted GWAS summary statistics in order to avoid reduce horizontal pleiotropy in the causal estimates.

Second, any conditionally independent pQTLs previously identified in INTERVAL for the protein by Sun *et al.* 2018[15] were then carried forward as candidate genetic instruments if present in the candidate variant set. If a pQTL could not be mapped to the GWAS summary statistics then the proxy variant in highest linkage disequilibrium was taken forward as a genetic instrument provided its r^2 with the pQTL was greater than 0.8 and its distance to the pQTL was less than 250KB. Where there were multiple aptamers targeting the same protein the pQTL effects of the instrument were averaged across the aptamers.

Third, *cis*-pQTLs in linkage equilibrium that did not meet the *trans*-pQTL significance threshold in Sun *et al.* were identified in a step-forward procedure. To identify additional *cis*-pQTLs we applied a hierarchical multiple testing correction procedure that has been shown to control the false discovery rate at 5% in expression quantitative trait loci (eQTL) studies[72,73] to the pQTL summary statistics from Sun *et al.* 2018[15]. For each aptamer, P-values within each 1MB of any gene encoding the targeted protein or protein complex were first Bonferroni corrected for the number of tests within that *cis* window(s) to obtain locally corrected P-values. The smallest P-value was then taken for each of the 3,793 aptamers and FDR correction was applied to obtain a single globally corrected P-value for each aptamer. Globally corrected P-values were filtered at FDR < 0.05, then the maximum corresponding locally corrected P-value of these lead SNPs (Bonferroni < 0.0124) was used as a significance threshold to identify *cis*-pQTLs from all *cis* variants. Where there were multiple aptamers targeting the same protein we required the *cis* variant to be significant for both aptamers to be carried forward as a candidate genetic instrument. A step-forward procedure was then applied to identify a subset of independent by linkage *cis*-pQTLs to add to the candidate instrument set for each causal test. First, the set of additional *cis*-pQTLs were filtered to include only variants with unambiguous alleles that mapped to the GWAS summary statistics. Next, these were filtered to exclude any variants in linkage disequilibrium ($r^2 > 0.05$) with any of the protein's *cis*-pQTLs identified by Sun *et al.* already in the candidate instrument set. If any *cis*-pQTLs remained, the one with the smallest P-value was added to the candidate instrument set. The LD-filtering and addition to the candidate instrument set of the remaining variant with the smallest P-value was reapplied until no variants remained. Linkage disequilibrium between variants was calculated directly from the probabilistic genotype data in the BGEN files using QCTOOL version 2[74] and LDstore[75].

After additional *cis*-pQTL mapping, 36 of the 48 PRS-associated proteins had any pQTLs and 24 of 48 had at least one *cis*-pQTL. After excluding APOE and PDE4D and applying the step-forward linkage disequilibrium pruning approach to select independent instruments, there were 15 protein to disease pairs with at least one *cis*-pQTL and at least three pQTLs total, which met the criteria for downstream Mendelian Randomisation analysis. **Table S9** details the genetic instruments for these 15 tests (14 proteins; SHBG was tested for a causal effect on both CAD and T2D).

Estimates of the causal effects of protein levels on log odds of disease for the 15 protein to disease pairs were calculated by five Mendelian randomisation methods using the MendelianRandomization R package[27] (**Table S10**). (1) The inverse-variance weighted (IVW) estimator, which takes the average causal estimate across all instruments: the average of their ratios of disease log odds over effect on protein levels multiplied by their standard errors[76]. It assumes all instruments are valid: the variant is associated with both the protein and disease, and that the variant has only a direct effect on that protein's levels; it can only effect the disease by modifying the protein's levels[76]. (2) The simple median estimator takes the median causal effect across instruments and is robust to outliers and up to 50% invalid instruments[77]. (3) The weighted median estimator, which weights each instrument by its inverse variance when calculating the median[77]. (4) The weighted mode estimator, which estimates the average causal effect from the largest set of instruments with consistent causal estimates[78]. (5) The MR-Egger estimator which detects and corrects for bias arising from horizontal pleiotropy; where instruments are affecting other risk factors that may explain their observed association with the disease[79]. To evaluate and summarise evidence for causal effects we took the median causal effect and P-value across all five estimators. We considered there to be significant evidence of a causal effect (**Figure 3**) if the median P-value across all the methods was < 0.05 and

there was no significant evidence of horizontal pleiotropy (P > 0.05 for the MR-Egger intercept term).

We excluded a causal effect for GRN on CAD, as the significant causal estimates for the inverse variance weighted, weighted median, weighted mode, and MR-Egger methods (**Table S10**) were driven by a single outlier *trans*-pQTL, rs646776. This is a well-known *trans*-pQTL for GRN at the *SORT1* locus which modifies CAD risk through altering plasma low density lipoprotein (LDL) cholesterol levels by regulating hepatic *SORT1* expression[80,81]. The lack of support for a causal effect from GRN's *cis*-pQTL (**Table S9**, **Figure S5**) along with the significant MR-Egger intercept (**Table S10**), suggests GRN lies downstream of this process, but not on the causal pathway to CAD.

Tissue specific gene expression to cardiometabolic trait associations in mouse reference datasets

Mouse orthologs for the genes encoding the 48 PRS associated proteins were identified using the Mouse Genome Informatics (MGI) database[30] (http://www.informatics.jax.org/). In total, 45 of the 48 proteins had mouse orthologs (**Table S11**). Correlations between the tissue-specific expression of these proteins and cardiometabolic traits in mice were explored in two publicly available mouse datasets: (1) the hybrid mouse diversity panel (**Figure 4A**), and (2) an F2 cross of two common mouse models of cardiometabolic disease (**Figure 4B**).

The hybrid mouse diversity panel (**Figure 4A**) is a collection of >100 genetically characterised inbred mouse strains (median: n=6 per strain) that in this study were fed a standard chow diet and sacrificed at 16 weeks of age[28]. Gene to trait correlations were searched for and downloaded from the publicly available systems genetics resource[82] (https://systems.genetics.ucla.edu). Gene to trait biweight midcorrelation coefficients and P-values were publicly available for gene expression data from adipose and liver tissues where the P-value was < 0.05 and the absolute value of the correlation coefficient was > 0.1.

Data from the panel of F2 crosses between C57BL/6J ApoE-/- and C3H/HeJ (**Figure 4B**) mice was acquired from a publicly available dataset spanning 334 individual offspring that were fed a western diet from 8 weeks to 16 weeks of age, and sacrificed at 24 weeks of age[29]. Quality controlled and normalised phenotype data[83–85] and gene expression data[29] from adipose, liver, and muscle tissues were obtained from Sage BioNetworks at https://www.synapse.org/#!Synapse:syn4497. We calculated biweight midcorrelation coefficients and P-values between gene expression levels and selected cardiometabolic traits using the WGCNA R package[86], filtering out those with $P \geq 0.05$ or absolute correlation coefficient ≤ 0.01 for consistency with the data available in HMDP.

In the heatmap (**Figure 4**) correlation coefficients and P-values were averaged where there were multiple microarray transcripts for a single gene (**Table S11**), with missing values (correlation coefficient ≤ 0.01 or $P \geq 0.05$) given a correlation coefficient of 0 and P-value of 0.525.

Tissue-specific gene expression following dietary intervention in C57BL/6J mice

Animal studies were approved by the Alfred Research Alliance Animal Ethics Committee. C57BL/6J mice (Jackson Laboratories) were bred at the Precinct Animal Centre in Melbourne, Australia, and randomly allocated by cage to three groups. Each group was either fed (1) a chow diet, (2) a high fat diet (SF04-001, Specialty Feeds, WA, Australia), or (3) a Western diet (SF00-219, Specialty Feeds, WA, Australia) from approximately 8 weeks of age for 12 weeks (**Figure S6**). Mice had access to food and water *ad libitum* and were housed at 22°C on a 12 hour light/dark cycle. Cages were changed weekly. At study end, mice were fasted for 4-6 hours and blood and tissues collected.

RNA was isolated from tissue as previously described[87]. Mouse tissues were homogenised in RNAzol, and RNA isolated using chloroform followed by isopropanol precipitation. Pellets were washed 2-3 times using 75% ethanol and resuspended in molecular grade water. cDNA was generated from RNA using M-MLV reverse transcriptase (Invitrogen) according to the manufacturer's instructions. qPCR was performed as described previously[88]. Briefly, qPCR was carried out on 10ng of cDNA using iTaq Universal SYBR Green Supermix (BioRad) and performed on a QuantStudio 7 Flex (ThermoFisher Scientific) real time detection system. mRNA expression is reported relative to control (chow diet) and quantified using the ΔΔCt after standardisation to a house keeping gene; *Rplp0* expression in muscle (*tibialis anterior*) and liver, and cyclophilin A (*Ppia*) in subcutaneous white adipose tissue. Primer sequences are available upon request.

One-way analysis of variance (anova) testing was performed to test for difference in mean expression between dietary intervention groups for each gene and tissue. Specifically, for the expression of each gene in each tissue, the one-way anova tested for a departure from the null hypothesis of mean(chow) = mean(high fat diet) = mean(western diet). A $\log_2$ transform was applied to the expression (ΔΔCt) of each gene when fitting the anova model.

Acknowledgements

Participants in the INTERVAL randomised controlled trial were recruited with the active collaboration of NHS Blood and Transplant England (www.nhsbt.nhs.uk), which has supported field work and other elements of the trial. DNA extraction and genotyping was co-funded by the National Institute for Health Research (NIHR), the NIHR BioResource (http://bioresource.nihr.ac.uk) and the NIHR (Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust). Olink® Proteomics assays were funded by Biogen, Inc. (Cambridge, MA, US). SomaLogic assays were funded by Merck and the NIHR (Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust). The academic coordinating centre for INTERVAL was supported by core funding from: NIHR Blood and Transplant Research Unit in Donor Health and Genomics (NIHR BTRU-2014-10024), UK Medical Research Council (MR/L003120/1), British Heart Foundation (SP/09/002; RG/13/13/30194; RG/18/13/33946) and the NIHR (Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust). A complete list of the investigators and contributors to the INTERVAL trial is provided in reference [19]. The academic coordinating centre would like to thank blood donor centre staff and blood donors for participating in the INTERVAL trial.

This work was supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome. This study was also supported by the Victorian Government's Operational Infrastructure Support (OIS) program.

S.C.R and J.M. were funded by the NIHR (Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust). G.A. was supported by a National Health and Medical Research Council of Australia (NHMRC) Early Career Fellowship (no. 1090462). A.V.K. was supported by grants from the National Human Genome Research Institute (award numbers 1K08HG010155 and 5UM1HG008895), an institutional grant from the Broad Institute of MIT and Harvard (variant2function), and a Hassenfeld Scholar Award from Massachusetts General Hospital. J.D. was funded by the NIHR (Senior Investigator Award).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The views expressed in this manuscript are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

References

1. International Schizophrenia Consortium *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–52 (2009).

2. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–92 (2015).

3. Khera, A. V *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).

4. Lambert, S. A., Abraham, G. & Inouye, M. Towards clinical utility of polygenic risk scores. *Hum. Mol. Genet.* (2019).

5. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590 (2018).

6. Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **17**, 392–406 (2016).

7. McCarthy, M. I. & Mahajan, A. The value of genetic risk scores in precision medicine for diabetes. *Expert Rev. Precis. Med. Drug Dev.* **3**, 279–281 (2018).

8. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).

9. Schadt, E. E. Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218–223 (2009).

10. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).

11. Stastna, M. & Van Eyk, J. E. Secreted proteins as a fundamental source for biomarker discovery. *Proteomics* **12**, 722–35 (2012).

12. Schwenk, J. M. *et al.* The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J. Proteome Res.* **16**, 4299–4310 (2017).

13. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

14. Finan, C. *et al.* The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, (2017).

15. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).

16. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).

17. Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769–773 (2018).

18. Moore, C. *et al.* The INTERVAL trial to determine whether intervals between blood donations can be safely and acceptably decreased to optimise blood supply: study protocol for a randomised controlled trial. *Trials* **15**, 363 (2014).

19. Di Angelantonio, E. *et al.* Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors. *Lancet* **390**, 2360–2371 (2017).

20. Jovanović, D., Krstivojević, P., Obradović, I., Durdević, V. & Dukanović, L. Serum cystatin C and beta2-microglobulin as markers of glomerular filtration rate. *Ren. Fail.* **25**, 123–33 (2003).

21. Levey, A. S., Inker, L. A. & Coresh, J. GFR estimation: from physiology to public health. *Am. J. Kidney Dis.* **63**, 820–34 (2014).

22. Sofat, R. *et al.* Circulating Apolipoprotein E Concentration and Cardiovascular Disease Risk: Meta-analysis of Results from Three Studies. *PLoS Med.* **13**, e1002146 (2016).

23. Rasmussen, K. L. Plasma levels of apolipoprotein E, APOE genotype and risk of dementia and ischemic heart disease: A review. *Atherosclerosis* **255**, 145–155 (2016).

24. Ding, E. L. *et al.* Sex hormone-binding globulin and risk of type 2 diabetes in women and men. *N. Engl. J. Med.* **361**, 1152–63 (2009).

25. Lundberg, M., Eriksson, A., Tran, B., Assarsson, E. & Fredriksson, S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res.* **39**, e102 (2011).

26. Gudmundsdottir, V. *et al.* Deep serum proteomics reveal biomarkers and causal candidates for type 2 diabetes. *bioRxiv* 633297 (2019).

27. Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).

28. Lusis, A. J. *et al.* The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. *J. Lipid Res.* **57**, 925–42 (2016).

29. Yang, X. *et al.* Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **16**, 995–1004 (2006).

30. Bult, C. J. *et al.* Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* **47**, D801–D806 (2019).

31. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

32. Knowler, W. C. *et al.* Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med.* **346**, 393–403 (2002).

33. Trainer, P. J. *et al.* Treatment of acromegaly with the growth hormone-receptor antagonist pegvisomant. *N. Engl. J. Med.* **342**, 1171–7 (2000).

34. van der Lely, A. J. *et al.* Long-term treatment of acromegaly with pegvisomant, a growth hormone receptor antagonist. *Lancet (London, England)* **358**, 1754–9 (2001).

35. Paisley, A. N. *et al.* Pegvisomant interference in GH assays results in underestimation of GH levels. *Eur. J. Endocrinol.* **156**, 315–9 (2007).

36. Ferraù, F., Albani, A., Ciresi, A., Giordano, C. & Cannavò, S. Diabetes Secondary to Acromegaly: Physiopathology, Clinical Features and Effects of Treatment. *Front. Endocrinol. (Lausanne).* **9**, 358 (2018).

37. Rose, D. R. & Clemmons, D. R. Growth hormone receptor antagonist improves insulin resistance in acromegaly. *Growth Horm. IGF Res.* **12**, 418–24 (2002).

38. Guevara-Aguirre, J. *et al.* Growth hormone receptor deficiency is associated with a major reduction in pro-aging signaling, cancer, and diabetes in humans. *Sci. Transl. Med.* **3**, 70ra13

(2011).

39. Olschewski, H. *et al.* Inhaled iloprost for severe pulmonary hypertension. *N. Engl. J. Med.* **347**, 322–9 (2002).

40. Ghofrani, H. A. *et al.* Amplification of the pulmonary vasodilatory response to inhaled iloprost by subthreshold phosphodiesterase types 3 and 4 inhibition in severe pulmonary hypertension. *Crit. Care Med.* **30**, 2489–92 (2002).

41. Grant, P. G., Mannarino, A. F. & Colman, R. W. cAMP-mediated phosphorylation of the low-Km cAMP phosphodiesterase markedly stimulates its catalytic activity. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9071–5 (1988).

42. Hebeisen, P. *et al.* Allosteric FBPase inhibitors gain 10(5) times in potency when simultaneously binding two neighboring AMP sites. *Bioorg. Med. Chem. Lett.* **18**, 4708–12 (2008).

43. Bussel, J. *et al.* Fostamatinib for the treatment of adult persistent and chronic immune thrombocytopenia: Results of two phase 3, randomized, placebo-controlled trials. *Am. J. Hematol.* **93**, 921–930 (2018).

44. Rolf, M. G. *et al.* In vitro pharmacological profiling of R406 identifies molecular targets underlying the clinical effects of fostamatinib. *Pharmacol. Res. Perspect.* **3**, e00175 (2015).

45. Barton, N. H., Etheridge, A. M. & Véber, A. The infinitesimal model: Definition, derivation, and implications. *Theor. Popul. Biol.* **118**, 50–73 (2017).

46. Kojima, Y. *et al.* Progranulin expression in advanced human atherosclerotic plaque. *Atherosclerosis* **206**, 102–8 (2009).

47. Xu, S., Lind, L., Zhao, L., Lindahl, B. & Venge, P. Plasma prolylcarboxypeptidase (angiotensinase C) is increased in obesity and diabetes mellitus and related to cardiovascular dysfunction. *Clin. Chem.* **58**, 1110–5 (2012).

48. Yoshida, T. *et al.* Association of gene polymorphisms with chronic kidney disease in high- or low-risk subjects defined by conventional risk factors. *Int. J. Mol. Med.* **23**, 785–92 (2009).

49. Baumeier, C. *et al.* Caloric restriction and intermittent fasting alter hepatic lipid droplet proteome and diacylglycerol species and prevent diabetes in NZO mice. *Biochim. Biophys. Acta* **1851**, 566–76 (2015).

50. Rajwani, A. *et al.* Increasing circulating IGFBP1 levels improves insulin sensitivity, promotes nitric oxide production, lowers blood pressure, and protects against atherosclerosis. *Diabetes* **61**, 915–24 (2012).

51. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **5**, e15004 (2010).

52. Sattlecker, M. *et al.* Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. *Alzheimer's Dement. J. Alzheimer's Assoc.* **10**, 724–734 (2014).

53. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429.e19 (2016).

54. Inouye, M. *et al.* Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *J. Am. Coll. Cardiol.* **72**, 1883–1893 (2018).

55. Lambert, S. A. *et al.* The Polygenic Score (PGS) Catalog: a database of published PGS to enable reproducibility and uniform evaluation. www.pgscatalog.org (2019).

56. Abraham, G. *et al.* Genomic prediction of coronary heart disease. *Eur. Heart J.* **37**, 3267–3278 (2016).

57. Deloukas, P. *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45**, 25–33 (2013).

58. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).

59. 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).

60. Bycroft, C. *et al. Genome-wide genetic data on ~500,000 UK Biobank participants*. *bioRxiv* (2017).

61. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).

62. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.* **51**, 957–972 (2019).

63. Mahajan, A. *et al.* Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat. Genet.* **50**, 559–571 (2018).

64. Nielsen, J. B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234–1239 (2018).

65. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).

66. Weng, L.-C. *et al.* Genetic Predisposition, Clinical Risk Factor Burden, and Lifetime Risk of Atrial Fibrillation. *Circulation* **137**, 1027–1038 (2018).

67. Udler, M. S., McCarthy, M. I., Florez, J. C. & Mahajan, A. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. *Endocr. Rev.* **40**, 1500–1520 (2019).

68. Abraham, G. *et al.* Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nat. Commun.* **10**, 5819 (2019).

69. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

70. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

71. Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391 (2017).

72. Peterson, C. B., Bogomolov, M., Benjamini, Y. & Sabatti, C. TreeQTL: hierarchical error control for eQTL findings. *Bioinformatics* **32**, 2556–8 (2016).

73. Huang, Q. Q., Ritchie, S. C., Brozynska, M. & Inouye, M. Power, false discovery rate and Winner's Curse in eQTL studies. *Nucleic Acids Res.* **46**, e133 (2018).

74. Band, G. & Marchini, J. BGEN: a binary file format for imputed genotype and haplotype data. *bioRxiv* 308296 (2018).

75. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am. J. Hum. Genet.* **101**, 539–551 (2017).

76. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–65 (2013).

77. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* **40**, 304–14 (2016).

78. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).

79. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–25 (2015).

80. Carrasquillo, M. M. *et al.* Genome-wide screen identifies rs646776 near sortilin as a regulator of progranulin levels in human plasma. *Am. J. Hum. Genet.* **87**, 890–7 (2010).

81. Musunuru, K. *et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714–9 (2010).

82. van Nas, A. *et al.* The systems genetics resource: a web application to mine global data for complex disease traits. *Front. Genet.* **4**, 84 (2013).

83. Ghazalpour, A. *et al.* Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet.* **2**, 1182–1192 (2006).

84. Estrada-Smith, D. *et al.* Dissection of multigenic obesity traits in congenic mouse strains. *Mamm. Genome* **15**, 14–22 (2004).

85. Meng, H. *et al.* Identification of Abcc6 as the major causal gene for dystrophic cardiac calcification in mice through integrative genomics. *Proc. Natl. Acad. Sci.* **104**, 4530–4535 (2007).

86. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

87. Parker, B. L. *et al.* An integrative systems genetic analysis of mammalian lipid metabolism. *Nature* **567**, 187–193 (2019).

88. Bond, S. T. *et al.* The E3 ligase MARCH5 is a PPARγ target gene that regulates mitochondria and metabolism in adipocytes. *Am. J. Physiol. Endocrinol. Metab.* **316**, E293–E304 (2019).

89. Foster, M. C., Yang, Q., Hwang, S.-J., Hoffmann, U. & Fox, C. S. Heritability and genome-wide association analysis of renal sinus fat accumulation in the Framingham Heart Study. *BMC Med. Genet.* **12**, 148 (2011).

90. Ruttmann, E. *et al.* Gamma-glutamyltransferase as a risk factor for cardiovascular disease mortality: an epidemiological investigation in a cohort of 163,944 Austrian adults. *Circulation* **112**, 2130–7 (2005).

91. Lee, D. S. *et al.* Gamma glutamyl transferase and metabolic syndrome, cardiovascular disease, and mortality risk: the Framingham Heart Study. *Arterioscler. Thromb. Vasc. Biol.* **27**, 127–33 (2007).

92. Pugeat, M. *et al.* Interrelations between sex hormone-binding globulin (SHBG), plasma lipoproteins and cardiovascular risk. *J. Steroid Biochem. Mol. Biol.* **53**, 567–572 (1995).

93. Sutton-Tyrrell, K. *et al.* Sex Hormone–Binding Globulin and the Free Androgen Index Are Related to Cardiovascular Risk Factors in Multiethnic Premenopausal and Perimenopausal Women Enrolled in the Study of Women Across the Nation (SWAN). *Circulation* **111**, 1242–1249 (2005).

94. Liu, P. Y., Death, A. K. & Handelsman, D. J. Androgens and Cardiovascular Disease. *Endocr. Rev.* **24**, 313–340 (2003).

95. Pickup, J. C. Inflammation and activated innate immunity in the pathogenesis of type 2 diabetes. *Diabetes Care* **27**, 813–23 (2004).

96. Flyvbjerg, A. The role of the complement system in diabetic nephropathy. *Nat. Rev. Nephrol.* **13**, 311–318 (2017).

97. Spranger, J. *et al.* Adiponectin and protection against type 2 diabetes mellitus. *Lancet (London, England)* **361**, 226–8 (2003).

98. Suckale, J. & Solimena, M. The insulin secretory granule as a signaling hub. *Trends Endocrinol. Metab.* **21**, 599–609 (2010).

99. Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–89 (2010).

100. van Poelje, P. D., Dang, Q. & Erion, M. D. Fructose-1,6-bisphosphatase as a therapeutic target for type 2 diabetes. *Drug Discov. Today Ther. Strateg.* **4**, 103–109 (2007).

101. Kato, N. Insights into the genetic basis of type 2 diabetes. *J. Diabetes Investig.* **4**, 233–44 (2013).

102. Lopez, P. H. *et al.* Mice lacking sialyltransferase ST3Gal-II develop late-onset obesity and insulin resistance. *Glycobiology* **27**, 129–139 (2017).

103. Dwinovan, J., Colella, A. D., Chegeni, N., Chataway, T. K. & Sokoya, E. M. Proteomic analysis reveals downregulation of housekeeping proteins in the diabetic vascular proteome. *Acta Diabetol.* **54**, 171–190 (2017).

104. Lau, W., Andrew, T. & Maniatis, N. High-Resolution Genetic Maps Identify Multiple Type 2 Diabetes Loci at Regulatory Hotspots in African Americans and Europeans. *Am. J. Hum. Genet.* **100**, 803–816 (2017).

105. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–90 (2012).

106. Grarup, N., Sandholt, C. H., Hansen, T. & Pedersen, O. Genetic susceptibility to type 2 diabetes and obesity: from genome-wide association studies to rare variants and beyond. *Diabetologia* **57**, 1528–41 (2014).

Table Captions

Table 1: Plasma proteins significantly associated with PRSs for cardiometabolic disease. PRS to protein associations were considered significant where their false discovery rate (FDR) adjusted P-value < 0.05 in linear regression model adjusting for age, sex, 10 genotype principal components (PCs), and technical covariates (**Methods**). Protein: proteins are labelled by their encoding gene. The entry for PDE4D/PDE4A indicates the association is with the combined levels of both proteins PDE4D and PDE4A as the associated aptamer binds both proteins with similar affinity (see **Table S4** for details). UniProt: the UniProtKB identifier for the protein. Full Protein Name: the full name of the protein as listed on UniProt. Beta: standard deviation change in the protein's levels associated with a one standard deviation increase in the respective PRS levels. 95% CI: 95% confidence interval of the beta. Pval: P-value in the linear regression model. FDR: false discovery-rate adjusted p-value across all proteins. Beta coefficients, 95% confidence intervals, and P-values were averaged where there were multiple aptamers measuring the protein's levels (GPD1, IGFBP1, IGFBP2, SHBG, and WFIKKN2). Literature: summary of any literature we could identify relating to the association between the protein and the PRS disease. A "-" indicates we could not identify any relevant prior literature for the association. Further details for these proteins and their respective aptamers are provided in **Table S4**. PRS to protein associations passing suggestive FDR significance (FDR < 0.1) are detailed in **Table S5**. Summary statistics for all PRS to protein and PRS to aptamer associations are provided in **Table S3.**

Table 2: Drug target information for proteins associated with PRSs for cardiometabolic diseases. PRS associated proteins targeted by any drug or compound listed in DrugBank (https://www.drugbank.ca/). Protein: protein associated with the PRS in the table subheading. Proteins are labelled by their encoding gene. #D: total number of drugs or compounds which bind to or otherwise interact with the listed protein. #A: number of those drugs that are or have been approved for use in any jurisdiction (in the "approved" but not "withdrawn" group in DrugBank). Summary of approved compound usage: summary of diseases the compounds targeting the protein have been approved to treat, or any other use they have been approved for. RT: are any drugs targeting the protein approved or undergoing clinical trials for the disease associated with the PRS ("yes" or "no"). **Table S12** provides details for each of the 184 drugs or compounds targeting any PRS associated protein.

Table 3: Drugs whose effects on proteins were consistent with PRS associations. Drugs that inhibit the function or levels (inhibitors and antagonists) of proteins whose levels were positively associated with PRS levels, and drugs that increase the function or levels (inducers and agonists) of proteins whose levels were inversely associated with PRS levels. Drug: the name of the drug. Phase: maximum phase the drug has reached in clinical trials or whether the drug is commercially available on the market. Indications: conditions the drug is used to treat. Target: the PRS associated protein to which the drug binds or interacts with. A "*" next to the protein name indicates the drugs pharmacological effect on the indicated disease is partially or wholly due to its effect on the PRS associated protein. Effect: the drug's effect on the PRS associated protein. PRS RT: maximum phase the drag has reached in clinical trials or whether the drug is commercially available on the market for the disease associated with the PRS. A "-" indicates no clinical trials could be identified for the disease associated with the PRS. NCT ID: identifier of the relevant clinical trial on the

National Institute of Health (NIH)'s National Library of Medicine (NLM)'s Clinical Trials database (https://clinicaltrials.gov).

Tables

Table 1: Plasma proteins significantly associated with PRSs for cardiometabolic disease.

Protein	UniProt	Full protein name	Beta	95% CI	Pval	FDR	Literature
Chronic Kidney Disease PRS							
FTMT	Q8N4E7	Ferritin, mitochondrial	-0.12	[-0.15, -0.078]	4×10^{-9}	1×10^{-5}	-
VWC2	Q2TAL6	Brorin	0.096	[0.057, 0.13]	1×10^{-6}	0.001	Nearby GWAS signal[89]
B2M	P61769	Beta-2-microglobulin	0.096	[0.057, 0.13]	1×10^{-6}	0.001	Disease biomarker[20]
CST3	P01034	Cystatin-C	0.090	[0.051, 0.13]	5×10^{-6}	0.004	Disease biomarker[20,21]
PRSS3	P35030	Trypsin-3	0.083	[0.044, 0.12]	3×10^{-5}	0.019	-
PDE4D/ PDE4A	Q08499/ P27815	cAMP-specific 3',5'-cyclic phosphodiesterase 4D / and 4A	-0.081	[-0.12, -0.042]	4×10^{-5}	0.021	Some association evidence[48] / -
UST	Q9Y2C2	Uronyl 2-sulfotransferase	-0.080	[-0.12, -0.041]	5×10^{-5}	0.024	-
Coronary Artery Disease PRS							
CEI	Q86SI9	Protein CEI	0.095	[0.059, 0.13]	2×10^{-7}	7×10^{-4}	-
GRN	P28799	Granulins	0.085	[0.049, 0.12]	3×10^{-6}	0.004	Experimental support[46]
APOE	P02649	Apolipoprotein E	0.084	[0.049, 0.12]	4×10^{-6}	0.004	Well-known association[22,23]
CRYZL1	O95825	Quinone oxidoreductase-like protein 1	0.084	[0.048, 0.12]	5×10^{-6}	0.004	-
TP53I11	O14683	Tumor protein p53-inducible protein 11	0.083	[0.047, 0.12]	6×10^{-6}	0.004	-
GGT2	P36268	Inactive glutathione hydrolase 2	0.082	[0.046, 0.12]	8×10^{-6}	0.005	Well-known association[90,91]
SHBG	P04278	Sex hormone-binding globulin	-0.079	[-0.11, -0.043]	2×10^{-5}	0.008	Well-known association[92–94]
DUSP26	Q9BV47	Dual specificity protein phosphatase 26	0.077	[0.041, 0.11]	3×10^{-5}	0.011	-
PCDHB10	Q9UN67	Protocadherin beta-10	0.076	[0.040, 0.11]	4×10^{-5}	0.014	-
HBQ1	P09105	Hemoglobin subunit theta-1	0.072	[0.037, 0.11]	8×10^{-5}	0.027	-
NPTX2	P47972	Neuronal pentraxin-2	0.070	[0.035, 0.11]	1×10^{-4}	0.046	-
Type 2 Diabetes PRS							
SHBG	P04278	Sex hormone-binding globulin	-0.11	[-0.14, -0.073]	1×10^{-9}	4×10^{-6}	Causal role[24]
CPM	P14384	Carboxypeptidase M	0.095	[0.061, 0.13]	7×10^{-8}	1×10^{-4}	-
IGFBP2	P18065	Insulin-like growth factor-binding protein 2	-0.097	[-0.13, -0.062]	2×10^{-7}	2×10^{-4}	-
MUSK	O15146	Muscle, skeletal receptor tyrosine-protein kinase	0.090	[0.055, 0.12]	4×10^{-7}	3×10^{-4}	-
PRSS1	P07477	Trypsin-1	-0.089	[-0.12, -0.054]	7×10^{-7}	5×10^{-4}	-
CFH	P08603	Complement factor H	0.087	[0.052, 0.12]	1×10^{-6}	6×10^{-4}	Relevant pathway[95,96]
CFI	P05156	Complement factor I	0.084	[0.049, 0.12]	3×10^{-6}	0.001	Relevant pathway[95,96]
PRCP	P42785	Lysosomal Pro-X carboxypeptidase	0.079	[0.045, 0.11]	6×10^{-6}	0.003	Some association evidence[47]
PTPRU	Q92729	Receptor-type tyrosine-protein phosphatase U	0.079	[0.044, 0.11]	9×10^{-6}	0.004	-
ACY1	Q03154	Aminoacylase-1	0.077	[0.043, 0.11]	1×10^{-5}	0.004	Experimental support[49]
ADIPOQ	Q15848	Adiponectin	-0.077	[-0.11, -0.042]	2×10^{-5}	0.005	Well-known association[97]
CHGB	P05060	Secretogranin-1	-0.074	[-0.11, -0.040]	3×10^{-5}	0.007	Relevant pathway[98]
IGFBP1	P08833	Insulin-like growth factor-binding protein 1	-0.074	[-0.11, -0.039]	4×10^{-5}	0.011	Experimental support[50]
GHR	P10912	Growth hormone receptor	0.071	[0.036, 0.11]	6×10^{-5}	0.016	Some association evidence[38]
FAH	P16930	Fumarylacetoacetase	0.070	[0.035, 0.11]	8×10^{-5}	0.019	Nearby GWAS signal[99]
GFRA1	P56159	GDNF family receptor alpha-1	0.069	[0.034, 0.10]	1×10^{-4}	0.020	-
WFIKKN2	Q8TEU8	WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 2	-0.070	[-0.10, -0.035]	1×10^{-4}	0.020	-
FBP1	P09467	Fructose-1,6-bisphosphatase 1	0.069	[0.034, 0.10]	1×10^{-4}	0.020	Well-known association[100]
TIMP4	Q99727	Metalloproteinase inhibitor 4	-0.069	[-0.10, -0.034]	1×10^{-4}	0.020	Nearby GWAS signal[99,101]
ADH4	P08319	Alcohol dehydrogenase 4	0.069	[0.034, 0.10]	1×10^{-4}	0.020	-
INHBC	P55103	Inhibin beta C chain	0.068	[0.033, 0.10]	1×10^{-4}	0.022	-
FAM20A	Q96MK3	Pseudokinase FAM20A	0.067	[0.033, 0.10]	2×10^{-4}	0.024	-
ST3GAL2	Q16842	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 2	0.067	[0.032, 0.10]	2×10^{-4}	0.027	Experimental support[102]
APOF	Q13790	Apolipoprotein F	-0.066	[-0.10, -0.031]	2×10^{-4}	0.027	-
RIDA	P52758	2-iminobutanoate/2-iminopropanoate deaminase	0.066	[0.031, 0.10]	2×10^{-4}	0.027	Some association evidence[103]
CCDC126	Q96EE4	Coiled-coil domain-containing protein 126	-0.066	[-0.10, -0.031]	2×10^{-4}	0.027	Nearby GWAS signal[104]
HS6ST2	Q96MM7	Heparan-sulfate 6-O-sulfotransferase 2	-0.066	[-0.10, -0.031]	2×10^{-4}	0.027	-
GPD1	P21695	Glycerol-3-phosphate dehydrogenase [NAD(+)], cytoplasmic	0.067	[0.032, 0.10]	3×10^{-4}	0.033	-
MSMP	Q1L6U9	Prostate-associated microseminoprotein	-0.065	[-0.10, -0.030]	3×10^{-4}	0.033	-
QPCTL	Q9NXS2	Glutaminyl-peptide cyclotransferase-like protein	-0.065	[-0.10, -0.030]	3×10^{-4}	0.033	Nearby GWAS signal[105,106]
PPP2R3A	Q06190	Serine/threonine-protein phosphatase 2A regulatory subunit B'' subunit alpha	0.064	[0.029, 0.099]	4×10^{-4}	0.040	-

Table 2: Drug target information for proteins associated with PRSs for cardiometabolic diseases

Protein	UniProt	#D	#A	Summary of approved compound usage	RT
Chronic Kidney Disease PRS					
B2M	P61769	3	1	Nutritional deficiencies.	No
PDE4D	Q08499	24	6	Respiratory diseases, skin conditions, hypertension.	Yes
PRSS3	P35030	8	0	-	No
Coronary Artery Disease PRS					
APOE	P02649	4	4	Nutritional deficiencies.	Yes
SHBG	P04278	28	18	Fertility and reproductive treatments, cancers, mental health, developmental disorders, hypertension, high cholesterol.	Yes
Type 2 Diabetes PRS					
ACY1	Q03154	3	3	Nutritional deficiencies, adipose atrophy, cancer, overdose.	Yes
ADH4	P08319	1	0	-	No
CFH	P08603	4	4	Nutritional deficiencies.	Yes
CFI	P05156	4	4	Nutritional deficiencies.	Yes
FAH	P16930	3	0	-	No
FBP1	P09467	11	1	Nutritional deficiencies, asthma.	Yes
GHR	P10912	3	2	Acromegaly, dwarfism, idiopathic short stature, HIV weight loss.	Yes
GPD1	P21695	2	2	Glycaemic control, type 2 diabetes.	Yes
MUSK	O15146	1	1	Chronic immune thrombocytopenia	No
PRSS1	P07477	99	0	-	No
RIDA	P52758	1	1	Food preservative.	No
SHBG	P04278	28	18	Fertility and reproductive treatments, cancers, mental health, developmental disorders, hypertension, high cholesterol.	Yes

Table 3: Drugs whose effects on proteins were consistent with PRS associations

Drug	Phase	Indications	Target	Effect	PRS RT	NCT ID
Chronic Kidney Disease PRS						
Iloprost	Market	Pulmonary arterial hypertension	PDE4A, PDE4D	Inducer	-	-
Type 2 Diabetes PRS						
Metformin	Market	Type 2 diabetes	GPD1	Inhibitor	Market	-
Pegvisomant	Market	Acromegaly	GHR*	Antagonist	II	NCT02023918
MB-07803	II	Type 2 diabetes	FBP1	Antagonist	II	NCT00458016
Managlinat dialanetil	II	Type 2 diabetes	FBP1	Antagonist	II	NCT00290940
Adenosine phosphate	Market	Dietary supplement	FBP1	Antagonist	-	-
Fostamatinib	Market	Immune thrombocytopenic purpura	MUSK	Inhibitor	-	-

Figure Captions

Figure 1: Study Overview. A) Study question **B)** Study design.

Figure 2: Plasma proteins associated with polygenic risk scores for cardiometabolic diseases. A) Quantile-Quantile plots of association P-values across all 3,442 proteins for each PRS (**Methods**). Each panel compares the distribution of observed P-values (y-axes) to the distribution of expected P-values under the null-hypothesis for all tests (x-axes) on a $-\log_{10}$ scale. The top five proteins for each PRS are labelled. **B)** Heatmaps showing the 48 proteins whose levels were significantly associated (FDR < 0.05) with at least one PRS. Each heatmap cell is coloured by the beta-coefficient in a linear regression of the corresponding protein levels on the corresponding PRS levels, which indicates the standard deviation change in the protein's level associated with a standard deviation increase in the PRS levels. Proteins are ordered by PRS from left to right by decreasing association magnitude, positive and negative PRS to protein associations split into separate heatmaps. **A** and **B)** Associations were adjusted for age, sex, 10 genotype PCs, sample measurement batch, and time between blood draw and sample processing (**Methods**). Associations were averaged where there were multiple aptamers measuring the same protein (GPD1, IGFBP2, IGFBP1, WFIKKN2, and SHBG in panel **B**). Association summary statistics across all proteins and aptamers for each PRS are given in **Table S3**. Details for the significant PRS to protein associations are given in **Table 1** and **Table S4**. PRS to protein associations with FDR < 0.1 are detailed in **Table S5**.

Figure 3: Causal proteins for disease suggested by Mendelian randomisation analysis. Dose response curves show the estimated causal effect of changes in protein levels on disease risk for each given protein. Proteins shown are those associated with any PRS that also had significant evidence of a causal effect in Mendelian randomisation analysis (median p-value < 0.05 across five causal estimators, **Methods**, **Table S10**). Points on each dose response curve show the standard deviation (SD) change in protein levels (x-axes) and odds ratio for T2D (y-axes) associated with each copy of the minor allele of each genetic instrument used to estimate the causal effect (**Methods**). Horizontal and vertical bars centred on each point show the respective standard errors. Associated changes in protein levels were obtained from the pQTL summary statistics derived in INTERVAL and published by Sun *et al.* 2018. Associated SD changes in protein levels were averaged where there were multiple aptamers measuring the same protein (SHBG, WFIKKN2). Associated odds ratios for T2D (adjusted for BMI) for each genetic instrument were obtained from GWAS summary statistics published by Mahajan *et al.* 2018. Genetic instruments for each test are detailed in **Table S9** and their selection is detailed in the **Methods**. All genetic instruments for these proteins were *cis*-pQTLs. The orange dashed line and yellow ribbon show the dose response curve and 95% confidence interval respectively of the median causal estimate across five Mendelian randomisation methods, each which each make different assumptions about instrument validity and perform well under different conditions (**Methods**). The slope of the orange dashed line corresponds to the estimated causal effect: the odds ratio conferred per standard deviation change in the protein levels. This median odds ratio and the median p-value across the five Mendelian randomisation methods are annotated on each plot. For SHBG and CFI the odds ratio is given for a standard deviation decrease in protein levels, while for WFIKKN2 the odds ratio is given for a standard deviation increase. **Figure S5** shows the dose-response curves for all 15 tested protein to disease pairs with all causal estimates for all five MR methods.

Figure 4: Tissue-specific gene expression and cardiometabolic traits in mice. Heatmaps of the biweight midcorrelation coefficients between tissue-specific expression of PRS-associated proteins with selected cardiometabolic traits in the (**A**) the hybrid mouse diversity panel: 706 male mice from 100 well-characterised inbred mouse strains (median n=6 per strain) fed a standard chow diet and (**B**) a mouse model of cardiometabolic disease; 334 mice from an F2 cross of the inbred ApoE$^{-/-}$ C57BL/6J and C3H/HeJ strains, fed a western diet to enrich for differentiation of cardiovascular disease traits. In both **A** and **B** cells are coloured where the correlation P-value < 0.05 and the absolute biweight midcorrelation coefficient > 0.1. Proteins are labelled by their corresponding mouse ortholog (detailed in **Table S11**). Reference heatmaps are given above the x axis labels in each panel showing which PRS levels each protein was significantly correlated with in INTERVAL as well as the direction of this correlation (**Table 1**, **Figure 2**). Proteins are ordered from left to right within each PRS group by decreasing order of average correlation magnitude across traits and datasets. Protein names are in bold where they have been subject of dietary intervention experimental follow-up (**Figure S6**).

Figures

Figure 1: Study Overview

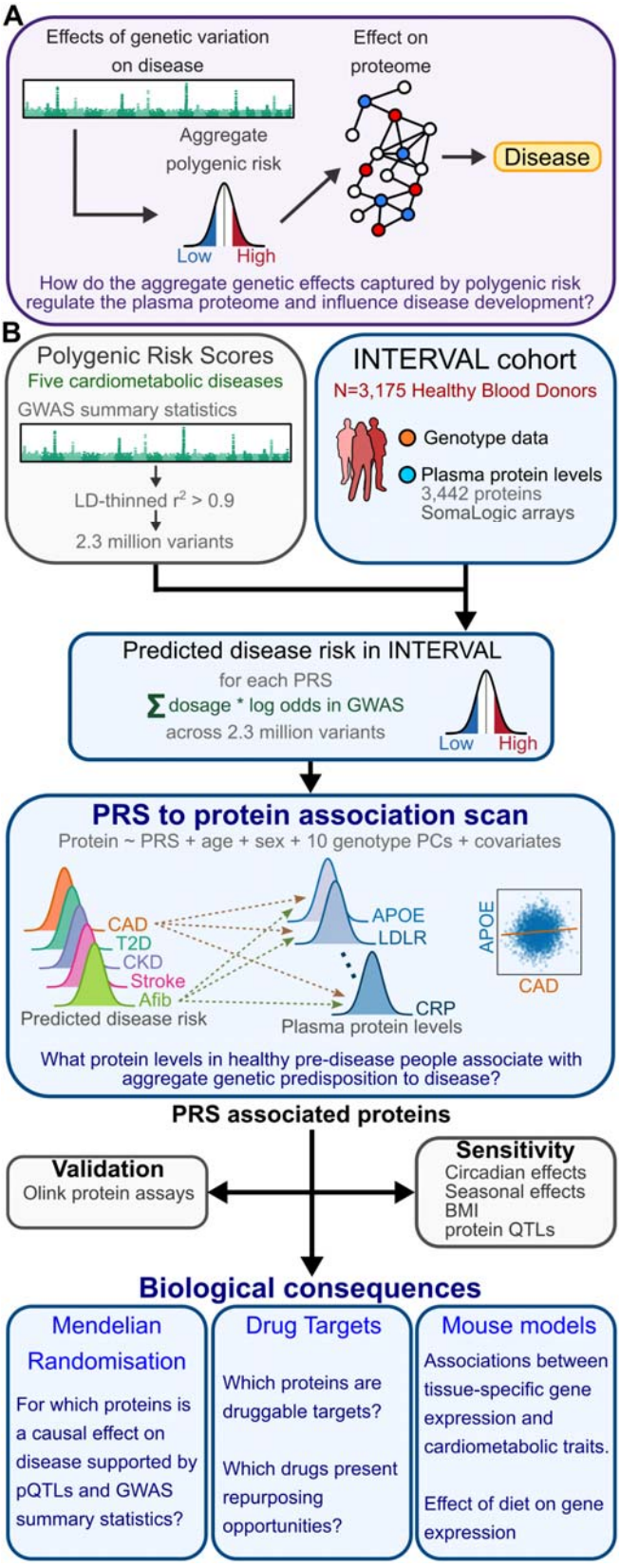


Figure 2: Plasma proteins associated with polygenic risk scores for cardiometabolic diseases.

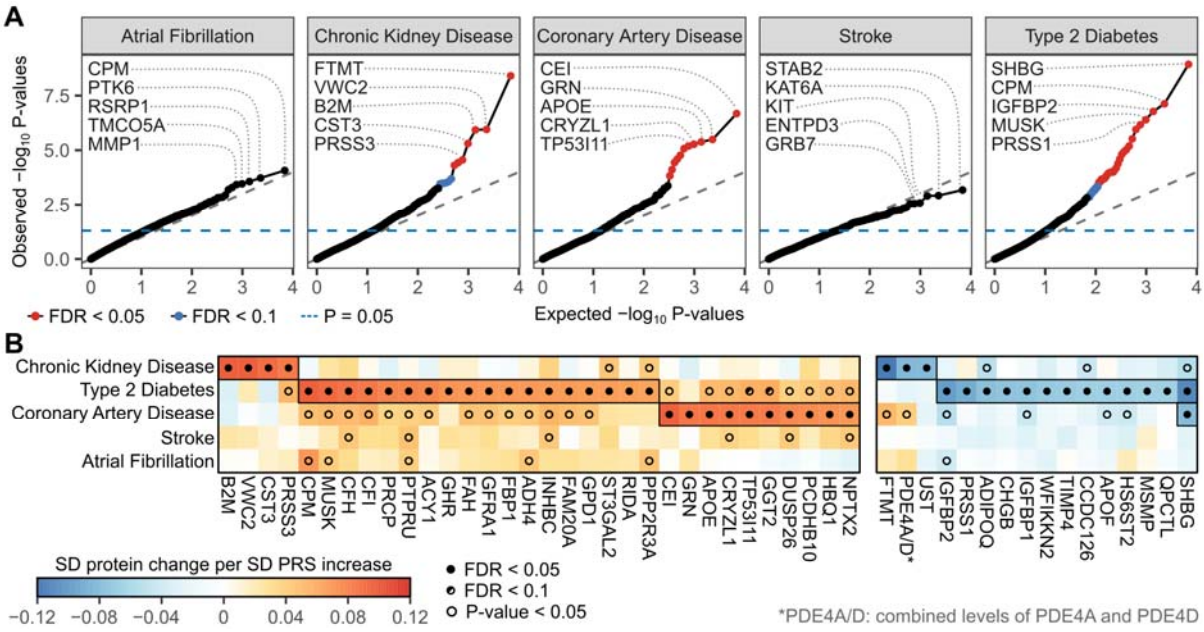


Figure 3: Causal proteins for disease suggested by Mendelian randomisation analysis.

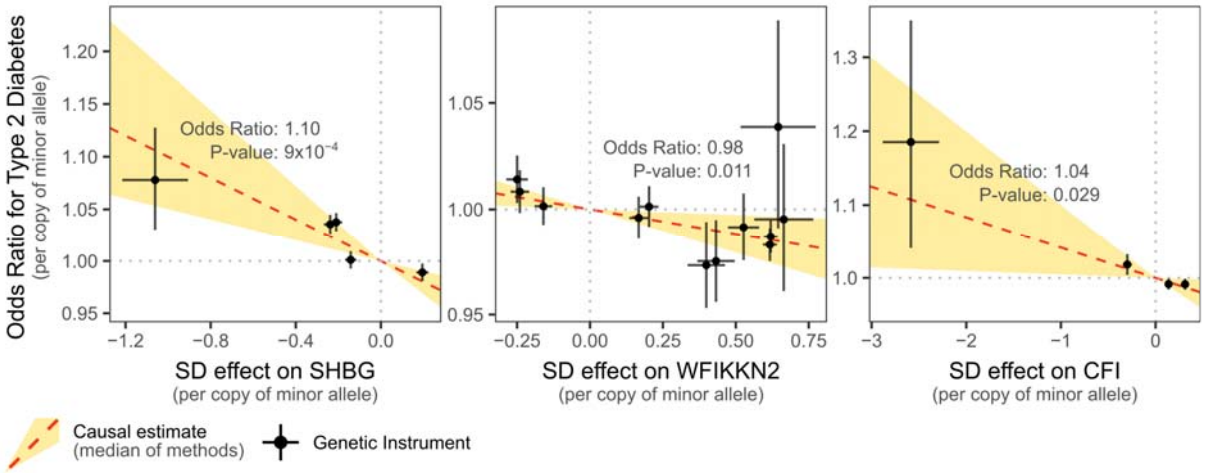
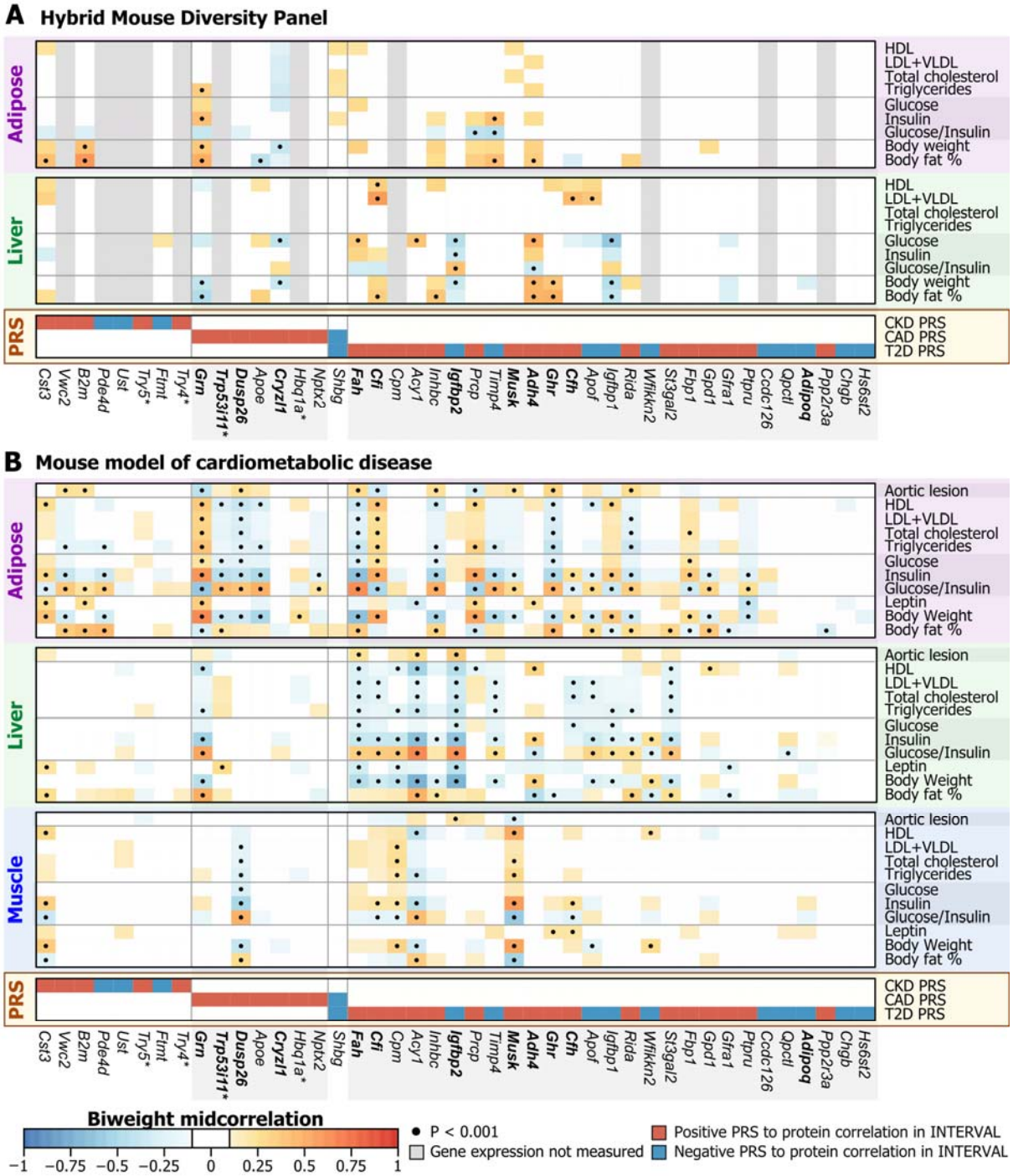


Figure 4: Tissue-specific gene expression and cardiometabolic traits in mice



**Try4* and *Try5* are the mouse orthologs of human gene *PRSS3*. *Trp53i11* is the mouse ortholog of human gene *TP53I11*. *Hbq1a* is the mouse ortholog of the human gene *HBQ1*.