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

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

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

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

Summary Paragraph

Common human diseases are frequently polygenic in architecture, comprising a large number of risk alleles with small effects spread across the genome[1–3]. Polygenic scores (PGSs) aggregate these alleles into a metric which represents an individual's genetic predisposition to a specific disease. PGSs have shown promise for early risk prediction[4–7], and there is potential to use PGSs to understand disease biology in parallel[8]. Here, we investigate the role plasma protein levels play in cardiometabolic disease risk in a cohort of 3,087 healthy individuals using PGSs. We found PGSs for coronary artery disease (CAD), type 2 diabetes (T2D), chronic kidney disease (CKD), and ischaemic stroke (IS) were associated with levels of 49 plasma proteins. These associations were polygenic in architecture, largely independent of *cis* protein QTLs, and robust to environmental variation. Over a median 7.7 years follow-up, 28 of these plasma proteins were associated with future myocardial infarction (MI) or T2D events, 16 of which were causal mediators between polygenic risk and incident disease. These protein mediators of polygenic disease risk included targets of approved therapies which may have repurposing potential. Our results demonstrate that PGSs can identify proteins with causal roles in disease, and may have utility in drug development.

Main Text

Cardiometabolic diseases have a major polygenic component, which is due to the combination of many thousands of variants across the genome, each exerting small lifelong effects[9–13]. Risk stratification using cardiometabolic PGSs have shown potential clinical utility for disease prevention[14]; however, molecular mediators of polygenic risk and their potential to be modulated to reduce disease risk remains unknown. Variants associated with polygenic traits are spread across many different pathways, exerting their effects through multiple levels of regulation, including gene expression, proteins and their interactions, cell morphology, and higher order physiological processes[15]. Proteins that are pathway-level hubs through which polygenic effects converge, however, could be promising targets for pharmaceutical intervention[16–19].

Here, we demonstrate how PGSs can be used to identify proteins with causal roles in disease aetiology. The INTERVAL cohort comprises approximately 50,000 adult blood donors in England[20,21], of which 3,087 participants have linked electronic hospital records, imputed genome-wide genotypes, and quantitative levels of 3,438 plasma proteins[22] (**Online Methods**, **Supplementary Data 1,2**). The characteristics of the participants are given in **Extended Data Table 1**; and participants with history of any cardiometabolic disease were excluded (**Online Methods, Supplementary Table 1**), reducing the potential for reverse causality in downstream analysis. A schematic of the study is given in **Extended Data Fig. 1**.

To quantify each participant's relative polygenic risk of atrial fibrillation (AF), CAD, CKD, IS, and T2D we applied externally derived genome-wide PGSs comprised of 1.8–3.2 million variants (**Online Methods**). Using PGSs, we identified 49 proteins whose levels differed with respect to polygenic risk at a false discovery rate (FDR) of 5% (**Fig. 1a,b**, **Extended Data Table 2,3, Supplementary Table 2,3**): 31 proteins for the T2D PGS, 11 proteins for the CAD PGS, 1 protein for the IS PGS, and 8 proteins for the CKD PGS. Associations included proteins with established roles in cardiometabolic disease, such as

cystatin-c (CST3) and beta-2-macroglobulin (B2M) which are biomarkers for chronic kidney disease[23], apolipoprotein E (APOE) whose link to coronary artery disease has been extensively studied[24,25], and fructose-1,6-bisphosphatase 1 (FBP1) which plays a key role in glucose regulation and is a target of type 2 diabetes drugs[26]. Associated proteins belonged to multiple non-overlapping pathways (**Supplementary Information**), and many are relatively understudied in the context of their respective diseases (**Extended Data Table 4**) warranting future study.

PGS to protein associations were robust to technical, physiological, and environmental confounding (**Supplementary Information**). We observed directional consistency and strong correlation of effect sizes when utilizing an orthogonal proteomics technology in independent samples (**Extended Data Fig. 2a-c**). Protein levels and PGS to protein associations were also temporally stable over two years of follow-up (**Extended Data Fig. 2c-d**). PGS to protein associations were also robust to circadian and seasonal effects, inclusion of participants with any prevalent cardiometabolic disease, and body mass index (BMI), with the exception of six T2D PGS to protein associations that were partially mediated by BMI (**Extended Data Fig. 2f-g**).

Most PGS to protein associations were not explained by protein quantitative trait loci (pQTLs) but instead were highly polygenic (**Online Methods**; **Fig. 1c**): each protein required a median 12% of the genome to explain its association with a PGS (**Fig. 1c**). Only four associations could be explained by pQTLs, and contributing loci were spread across the genome for the remaining 46 (**Extended Data Fig. 3**). Interestingly, the effects of PGSs and pQTLs on protein levels were largely independent (**Online Methods**, **Supplementary Table 4**), suggesting that polygenic risk can enhance or buffer locus-specific effects on protein levels.

Three possible scenarios could explain a PGS to protein association[27]: (1) the protein plays a causal role in disease, (2) the protein levels are changing in response to disease processes, but are not themselves causal (reverse causality), and (3) the protein levels are correlated with some other causal factor (confounding) (**Fig. 2a**). Utilizing a median of 7.7 years of follow-up in nation-wide electronic hospital records, we examined whether levels of PGS-associated proteins were associated with risk of onset of the respective cardiometabolic disease, then performed mediation analysis[28] to identify the proteins that mediate PGS to disease associations, and thereby play causal roles in disease pathogenesis (**Online Methods**). Limited by the number of incident disease events, we restricted our analyses to CAD and T2D (**Extended Data Fig. 4**).

25 of 31 (81%) of T2D PGS proteins were significantly associated ($P < 0.05$) with increased risk of T2D and 3 of 11 (27%) of CAD PGS proteins were significantly associated with increased risk of incident MI (**Fig. 2b**, **Extended Data Table 2**). There was directional consistency and strong correlation (Pearson correlation: 0.96, $P = 4\times10^{-23}$) between effects of PGSs on protein levels and hazard ratios for protein levels on incident disease risk (**Fig. 2c**). Using mediation analysis, we found that one and 15 proteins were significant mediators between polygenic risk of MI and T2D, respectively, indicating causal roles in disease pathogenesis (**Fig. 2d**).

As polygenic disease risk is itself estimated from population-level data, it is unlikely that any single protein explains polygenic risk. Here, we found that causal protein mediators each explained a median of 6.6% of PGS to disease associations (**Extended Data Table 2**), with the 1 CAD PGS mediator (APOE) explaining 5.4% of CAD polygenic risk to incident MI association, and the 15 T2D PGS mediators explaining 27% of the T2D polygenic risk to incident T2D association. A complementary approach for causal inference,

Mendelian randomisation[29] (**Online Methods**), also supported causal effects on T2D for two proteins (SHBG and CFI) which mediated the T2D PGS to T2D association (**Supplementary Information, Extended Data Fig. 5, Supplementary Table 5, 6**). Notably, only 12 (24%) of the proteins associated with PGSs could be tested with Mendelian randomisation due to lack of *cis* protein quantitative trait loci (pQTLs) as genetic instruments (**Online Methods**) highlighting the complementarity of our PGS-protein association approach.

Finally, to identify druggable targets associated with polygenic disease risk and potential drug repurposing opportunities, we utilised the DrugBank database[30] (**Online Methods**) to find that 18 of the 49 PGS-associated proteins were targeted by 236 drugs (**Extended Data Table 5**, **Supplementary Table 7**). Ten licensed drugs had protein target effects which were consistent with reduction of cardiometabolic disease risk (**Table 1**). These included the well-known T2D drug metformin[31], which reduces liver glucose production by inhibition of FBP1[32], a protein whose levels were elevated in people with high polygenic risk for T2D (**Fig. 1**). Among the other nine licensed drugs, we highlight the potential to repurpose pegvisomant for T2D prevention. Pegvisomant (DB00082) is used to treat acromegaly by blocking the binding of endogenous growth hormone to growth hormone receptor (GHR)[33–35]. We found increased GHR was a causal mediator of polygenic T2D risk and incident T2D (**Fig. 2**) and GHR loss-of-function mutations are associated with lower T2D risk[36] providing additional genetic support for this target. Furthermore, pegvisomant has been shown to improve insulin sensitivity in acromegaly patients[37,38]. Together, these observations suggest pegvisomant is a priority to evaluate for repurposing for T2D prevention (**Table 1**).

Conclusions

Polygenic scores for disease are explicitly constructed to maximise risk prediction, typically without consideration of the underlying biology. However, PGSs also hold considerable promise for identifying molecular pathways in the development and progression of disease[8,27]. Here, we identified plasma proteins significantly associated with PGSs for cardiometabolic disease in a healthy pre-disease cohort. The vast majority of these associations were highly polygenic, revealing an unappreciated role for polygenic effects on protein levels, including for several well-known disease proteins. These proteins were predictive of incident disease, and 16 were mediators of type 2 diabetes or myocardial infarction, suggesting that their modulation is likely to attenuate disease risk. There are multiple licensed drugs for many of these targets. Overall, this study demonstrates the power of polygenic scores to elucidate novel disease biology and their potential to inform development of medicines.

References

1. International Schizophrenia Consortium *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).

2. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (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.* **28**, R133–R142 (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 Review of Precision Medicine and Drug Development* **3**, 279–281 (2018).

8. International Common Disease Alliance. *International Common Disease Alliance Recommendations and White Paper*. (2020).

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

10. 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).

11. 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).

12. 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).

13. 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).

14. Sun, L. *et al.* Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses. *PLoS Med.* **18**, e1003498 (2021).

15. Erdmann, J., Kessler, T., Munoz Venegas, L. & Schunkert, H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc. Res.* **114**, 1241–1257 (2018).

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

17. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022-1034.e6 (2019).

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

19. Ghoussaini, M. *et al.* Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021).

20. 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).

21. 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).

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

23. 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**, 123–133 (2003).

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

25. 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).

26. 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).

27. Holmes, M. V. & Davey Smith, G. Can Mendelian Randomization Shift into Reverse Gear? *Clinical chemistry* vol. 65 363–366 (2019).

28. Imai, K., Keele, L., Tingley, D. & Yamamoto, T. Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *Am. Polit. Sci. Rev.* **105**, 765–789 (2011).

29. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89-98 (2014).

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

31. 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).

32. Hunter, R. W. *et al.* Metformin reduces liver glucose production by inhibition of fructose-1-6-bisphosphatase. *Nat. Med.* **24**, 1395–1406 (2018).

33. Trainer, P. J. *et al.* Treatment of acromegaly with the growth hormone-receptor antagonist pegvisomant. *N. Engl. J. Med.* **342**, 1171–1177 (2000).

34. van der Lely, A. J. *et al.* Long-term treatment of acromegaly with pegvisomant, a growth hormone receptor antagonist. *Lancet* **358**, 1754–1759 (2001).

35. Paisley, A. N. *et al.* Pegvisomant interference in GH assays results in underestimation of GH levels. *Eur. J. Endocrinol.* **156**, 315–319 (2007).

36. 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).

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

38. Rose, D. R. & Clemmons, D. R. Growth hormone receptor antagonist improves insulin resistance in acromegaly. *Growth Horm. IGF Res*. **12**, 418–424 (2002).

39. European Organization for Nuclear Research & OpenAIRE. Zenodo. (2013) doi:10.25495/7GXK-RD71.

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 (BRC-1215-20014). Olink® Proteomics assays were funded by Biogen, Inc. (Cambridge, MA, US). SomaLogic assays were funded by Merck and the NIHR Cambridge BRC (BRC-1215-20014). 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 BRC (BRC-1215-20014). A complete list of the investigators and contributors to the INTERVAL trial is provided in reference [21]. 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.

This work uses data provided by patients and collected by the NHS and Public Health England (PHE) as part of their care and support. Data on Hospital Episode Statistics, mortality and cancer registration was obtained from NHS Digital (data sharing agreement reference: DARS-NIC-156334-711SX).

S.C.R and J.M. were funded by the NIHR Cambridge BRC (BRC-1215-20014). S.A.L. is supported by a Canadian Institutes of Health Research postdoctoral fellowship (MFE-171279). 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. holds a British Heart Foundation Personal Professorship and an 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 author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Data Availability

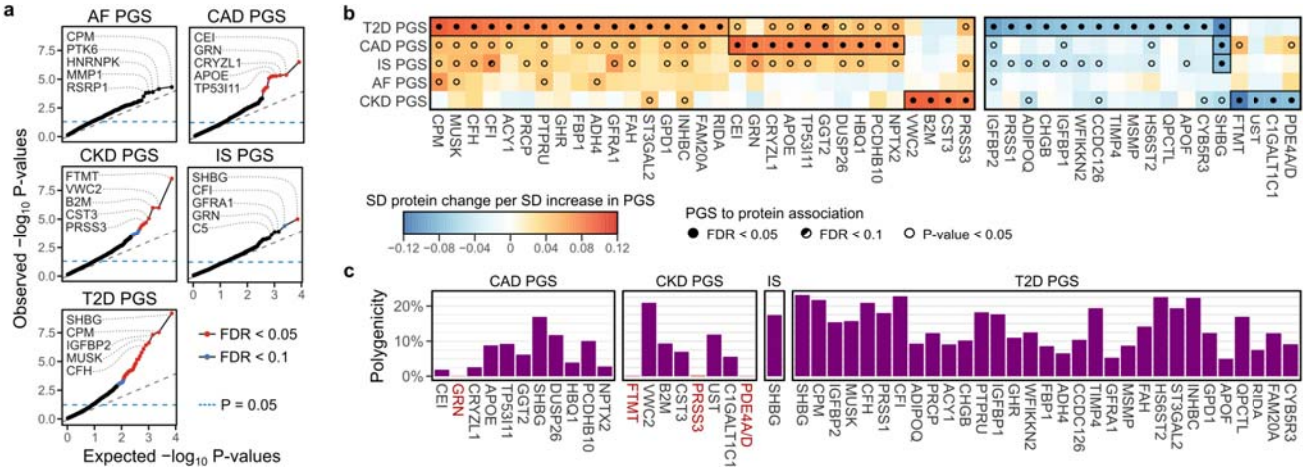
With the exception of electronic hospital records, all data used in this study is publicly available or deposited in a public repository. Genetic data, proteomic data, and basic cohort characteristics for the INTERVAL cohort are available via the European Genotype-phenome Archive (EGA) with study accession EGAS00001002555 (https://www.ebi.ac.uk/ega/studies/EGAS00001002555). Dataset access is subject to approval by a Data Access Committee: these data are not publicly available as they contain potentially identifying and sensitive patient information. Linked electronic hospital records are currently only available to researchers at the University of Cambridge UK, however, may become more widely available in the future. Contact the data access committee for further details. All other data used in this study is publicly available without restriction. The PGS used in this study are available to download through the Polygenic Score Catalog (https://www.pgscatalog.org/) with accession numbers PGS000727 (atrial fibrillation), PGS000018 (coronary artery disease), PGS000728 (chronic kidney disease), PGS000039 (ischaemic stroke), and PGS000729 (type 2 diabetes). GWAS summary statistics used to generate new PGS in this study are available to download through the GWAS Catalog (https://www.ebi.ac.uk/gwas/) with study accessions GCST008065 (chronic kidney disease), GCST007517 (type 2 diabetes), and GCST006414 (atrial fibrillation). Summary statistics for all statistical tests are available in **Supplementary Data 3**. Full pQTL summary statistics published by Sun *et al.* 2018 for all SomaLogic SOMAscan aptamers are available to download from https://www.phpc.cam.ac.uk/ceu/proteins/. A listing of *cis*-pQTLs mapped for this study are provided in **Supplementary Data 4**. GWAS summary statistics used for Mendelian randomisation are available to download through the GWAS Catalog (https://www.ebi.ac.uk/gwas/) with study accessions GCGCST004787 (coronary artery disease), GCST008065 (chronic kidney disease), GCST006906 (ischaemic stroke) and GCST007518 (type 2 diabetes). The DrugBank database is publicly available to download at https://www.drugbank.ca/releases/latest.

Code Availability

Code used to generate the results of this study, along with a detailed list of software and versions, are available on GitHub at https://github.com/sritchie73/cardiometabolic_PGS_plasma_proteome/ which is permanently archived by Zenodo[39] at doi: 10.5281/zenodo.4551565.

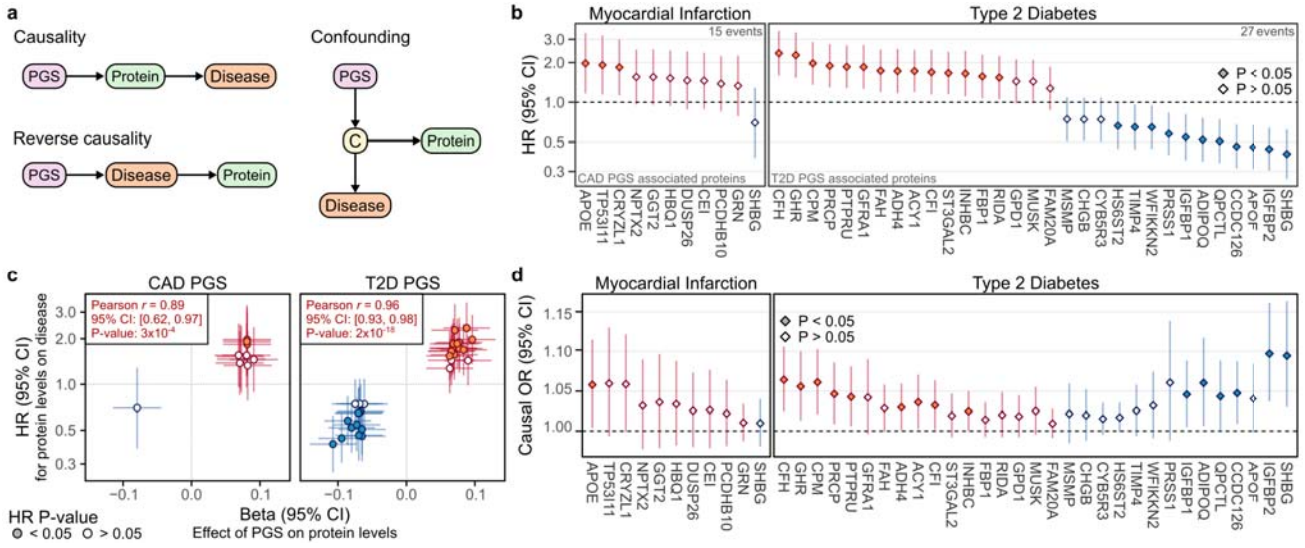
Figures

Figure 1: Proteins associated with polygenic risk for cardiometabolic disease



a) Quantile-quantile plots of P-values for PGS to protein associations across all 3,438 tested proteins. Each plot compares the distribution of observed P-values (y-axes) to the distribution of expected P-values under the null-hypothesis for 3,438 tests (x-axes) on a $-\log_{10}$ scale. Associations were adjusted for age, sex, 10 genotype PCs, sample measurement batch, and time between blood draw and sample processing. Full summary statistics are provided in **Supplementary Data 3a**. **b)** Heatmaps showing the 49 proteins whose levels significantly associated (FDR < 0.05) with at least one PGS. Each heatmap cell shows the standard deviation change in protein levels per standard deviation increase in PGS, estimated linear regression adjusted for age, sex, 10 genotype PCs, sample measurement batch, and time between blood draw and sample processing. Proteins are ordered by PGS from left to right by decreasing association magnitude, positive and negative associations split into separate heatmaps. Point estimates are detailed in **Extended Data Table 2**. Details about each protein are provided in **Extended Data Table 3**. **c)** Barplots showing the proportion of the genome required to explain each PGS to protein association (polygenicity; **Online Methods**). Proteins are ordered from left to right by strength of PGS to protein association. Highlighted in red are PGS to protein associations that were explained by singular variants regulating the protein levels, protein quantitative trait loci (pQTLs), rather than polygenic.

Figure 2: PGS-associated proteins influence 7.7 year risk of myocardial infarction and diabetes



a) Possible models of causality for PGS to protein to disease associations. C: causal disease factor upstream of protein that induces a correlation between protein levels and disease. **b)** Association between PGS-associated proteins with 7.7 year risk of hospitalisation with myocardial infarction and diabetes. There were insufficient events to analyse proteins associated with the IS PGS (N=3 incident disease events) or with the CKD PGS (N=0 incident disease events) (**Extended Data Fig. 4a**). Cox proportional hazard models were fit between protein levels and incident disease using follow-up as time scale and adjusting for age and sex (**Online Methods**). HR: hazard ratio conferred per standard deviation increase in protein levels. 95% CI: 95% confidence interval. See **Extended Data Table 2** for detailed point estimates. **c)** Comparison of effects of PGS on protein levels (x-axes; **Fig. 1b**) to associations between protein levels and incident disease (y-axes; **Fig. 2b**). Points and horizontal bars on the x-axes indicate standard deviation change in protein levels (and 95% confidence interval) per standard deviation increase in respective PGS. Points and vertical bars on the y-axis show hazard ratio (and 95% confidence interval) per standard deviation increase in protein levels. **d)** Estimated causal effect of PGS on disease through each protein in mediation analysis (**Online Methods**). Causal OR: odds ratio for incident disease adjusting for age and sex conferred through each protein per standard deviation increase in PGS. The total odds ratio for MI conferred per standard deviation increase in CAD PGS was 2.94 (95% CI: 1.69–5.31, P-value: 2×10^{-4}). The total odds ratio for T2D conferred per standard deviation increase T2D PGS was 2.00 (95% CI: 1.37–2.96, P-value: 4×10^{-4}). Proteins are ordered from left to right by their hazard ratio in **Fig. 1b**. **b-d**) points in red indicate proteins whose levels increased with PGS, and blue indicates proteins whose levels decreased with PGS.

Tables

Table 1: Drugs whose effects on proteins counteract effects of PGSs on proteins

			PGS associated target			PGS supported use		
Drug ID	Drug Name	Therapeutic uses	Protein	Drug Effect	Pharma	Disease	Phase	Trial Number
DB00331	Metformin	Type 2 diabetes	FBP1	Inhibitor	Yes	T2D	Licensed	-
			GPD1	Inhibitor	-	CAD	4	NCT00361075
DB00396	Progesterone	Female infertility, hormone imbalance	SHBG	Potentiator	-	T2D*	3	NCT00000466
						IS	-	-
DB01088	Iloprost	Pulmonary arterial hypertension	PDE4A	Inducer	-	CKD	3	NCT00345501
			PDE4D	Inducer	-	CAD	3	NCT00000529
DB00675	Tamoxifen	Breast, ovarian, and endometrial cancers	SHBG	Inducer	-	IS	-	-
						T2D*	-	-
DB00082	Pegvisomant	Acromegaly	GHR	Antagonist	Yes	T2D*	2	NCT02023918
						T2D*	2	NCT00494663
DB01026	Ketoconazole	Fungal infections, Cushing's syndrome	SHBG	Ligand	-	CAD	-	-
						IS	-	-
DB12010	Fostamatinib	Immune thrombocytopenic purpura	MUSK	Inhibitor	-	T2D	-	-
DB00131	Adenosine phosphate	Nutritional deficiencies	FBP1	Antagonist	-	T2D	-	-
DB14533	Zinc chloride	Zinc deficiency, intravenous nutrition	APOE	Antagonist	-	CAD*	-	-
DB14548	Zinc sulfate	Intravenous nutrition	APOE	Antagonist	-	CAD*	-	-

List of drugs that reduce the function or levels of proteins whose levels are elevated in participants with high polygenic risk, or increase in function or levels of proteins whose levels are decreased in participants with high polygenic risk (**Supplementary Information**). Drug ID gives the identifier in DrugBank. Columns under the "PGS associated target" heading indicate the PGS-associated protein that interacts with the drug, and the effect of the drug on the listed protein, and "Yes" in the "Pharma" when the pharmacological action of the drug is due to its effect on the protein (as listed in the DrugBank database). Columns under the "PGS supported use" heading indicate the disease(s) whose PGS are associated with the listed protein, and where clinical trials for that drug on that disease have been undertaken, the maximum clinical trial phase reached along with the respective trial number in the National Institute of Health (NIH)'s National Library of Medicine (NLM)'s Clinical Trials database (https://clinicaltrials.gov). A * next to the disease indicates there was evidence supporting a causal effect of the protein on the disease (**Fig. 2d**). See **Supplementary Information** for summary of evidence for each drug.

Online Methods

INTERVAL cohort

INTERVAL is a cohort of approximately 50,000 participants nested within a randomised trial studying the safety of varying frequency of blood donation[20,21]. Participants were blood donors aged 18 years and older (median 44 years of age; 49% women) recruited between June 2012 and June 2014 from 25 centres across England. The collection of their blood samples for research purposes was done using standard protocols and has been extensively described previously[20]. Participants gave informed consent and this study was approved by the National Research Ethics Service (11/EE/0538).

Electronic health records were obtained for all INTERVAL participants from the National Health Service (NHS) hospital episode statistics database (https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics) for all events up to the 8th of February 2020, prior to the onset of the COVID19 pandemic in England. The median and maximum follow-up time were 6.9 years and 7.7 years respectively. The earliest available hospital record for any INTERVAL participant was the 25th March 1999, with maximum retrospective follow-up of 13.6 years. These records came in the form of international classification of diseases 10th revision (ICD-10) codes[40] and were subsequently made available to analysts after summarisation into 301 endpoints using CALIBER rule-based phenotyping algorithms[41] (https://www.caliberresearch.org/portal). ICD-10 codes contributed to each event regardless of whether they coded for primary or non-primary diagnoses in the hospital records.

Genotyping, quality control, and imputation of INTERVAL participants has been described in detail previously[42]. Briefly, participants were genotyped using the Affymetrix UK Biobank Axiom array in 10 batches. Samples were removed if they had sex mismatch, extreme heterozygosity, were of non-European descent, or were duplicate samples. Related samples were removed by excluding one sample from each pair of close relatives (first or second degree; identity-by-descent $\hat{\pi} > 0.187$). Genotyped variants were removed if they were monomorphic, bi-allelic and had Hardy-Weinberg equilibirum p-value $< 5\times10^{-6}$, or call rate < 99%. 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).

Quantification, processing, and quality control of protein levels in INTERVAL using the SOMAscan assays has been described in detail previously[22]. Briefly, relative concentrations of 4,034 SOMAscan aptamers were measured in 3,562 INTERVAL participants in two batches by SomaLogic Inc. (Boulder Colorado, US) using version 3 of the SOMAscan platform. Aptamers were excluded if, in the latest version of the SOMAscan platform, they (1) targeted non-human proteins, (2) have been found to be measuring the fusion construct rather than the target protein, or (3) found to be measuring a contaminant. A curated information sheet for all 4,034 aptamers is provided in **Supplementary Data 1**.

Aptamer concentrations (relative fluorescence units) were natural log transformed then adjusted within each batch for participant age, sex, the first three genetic PCs, and duration between blood draw and sample processing (< 1 day or > 1 day), then the residuals were inverse rank normal transformed. Here, we further adjusted the normalized protein levels used in previous studies for batch number, and filtered to 3,793 high quality aptamers targeting 3,438 proteins after obtaining the latest information about aptamer sensitivity and specificity from SomaLogic. Distributions of aptamer levels and associations with covariates before and after quality control are given in **Supplementary Data 2**.

In total, there were 3,087 INTERVAL participants passing quality control, without prevalent cardiometabolic disease (see below), and with matched genotype, proteomic, and electronic health record data available for the primary analyses.

Prevalent disease exclusion

National Health Service (NHS) Blood and Transplant blood donation eligibility criteria (https://www.blood.co.uk/who-can-give-blood/) meant there were built in exclusions for the INTERVAL cohort for people with a history of major diseases, recent illness, or infection. Specifically for cardiometabolic diseases, blood donation eligibility criteria excluded individuals who had been diagnosed with atrial fibrillation, had a history of any stroke, or a 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. Use of aspirin or other blood thinners to control elevated blood pressure (hypertension) also made people ineligible to donate blood and participate in the INTERVAL cohort. Individuals with type 2 diabetes were ineligible, unless their type 2 diabetes was well controlled by diet alone, did not require regular insulin treatment, and the individual had not required insulin treatment for at least four weeks prior to attempted blood donation. Extended details on blood donation criteria eligibility for specific diseases, medications, and lifestyle factors can be found at https://my.blood.co.uk/knowledgebase.

In addition to intrinsic exclusion due to blood donation eligibility criteria, participants were excluded from analyses if they had any events relating to cardiometabolic disease prior to baseline assessment. Among the 301 CALIBER endpoints, we classified 48 as cardiometabolic disease or having potential to introduce reverse causality by modifying risk for incident AF, CAD, CKD, IS, or T2D (**Supplementary Table 1**). In total 87 participants (2.7%) were excluded, predominantly due to prevalent hypertension (N=57 events; 66% of excluded participants) and prevalent diabetes (N=11 events; 13% of excluded participants); with all others accounting for less than 5% of excluded participants (**Supplementary Table 1**).

Polygenic scores

PGSs were derived in a consistent manner, by linkage-disequilibrium thinning, at an r^2 threshold of 0.9, the latest GWAS summary statistics for each respective disease (**Supplementary Information**). GWAS summary statistics used to derive the AF PGS, CKD PGS, and T2D PGS were those published by Nielsen *et al.* in 2018[9] (GCST006414), Wuttke *et al.* in 2019[10] (GCST008065), and Mahajan *et al.* in 2018[11] (GCST007517), respectively. PGSs for CAD and IS used in this study were our previously published CAD metaGRS[43] and Stroke metaGRS[44]. The CAD PGS was derived from meta-analysis of three PGSs for CAD, including a PGS derived as described above from GWAS summary statistics published by Nikpay *et al.* in 2015[45]. The IS PGS was derived from meta-analysis of PGS for ischaemic stroke and its risk factors, including a PGS derived as described above from GWAS summary statistics for IS published by Malik *et al.* in 2018[12]. The PGSs each comprised 1.75–3.23 million SNPs genome-wide and are available to download through the Polygenic Score Catalog[46] (https://www.pgscatalog.org/) with accession numbers PGS000727 (atrial fibrillation), PGS000018 (coronary artery disease), PGS000728 (chronic kidney disease), PGS000039 (ischaemic stroke), and PGS000729 (type 2 diabetes). All PGSs were derived from GWAS summary statistics including only individuals with European ancestry. See **Supporting Information** and **Extended Data Fig. 4** for details on PGS validation.

Levels of each PGS (sum of dosages × weights) were computed in INTERVAL from probabilistic dosage data using plink (version 2)[47] after mapping PGS variants to those available in the INTERVAL genotype

data (**Supplementary Information**). Levels of each PGS were adjusted for the first 10 principal components (PCs) of the imputed genotype data and standardised to have mean of 0 and standard deviation of 1 prior to downstream statistical analyses.

PGS to protein associations

Each of the five PGSs were tested for association with each of the 3,793 aptamers using linear regression (**Fig 1a,b**, **Extended Data Table 2**). PGS and proteins were adjusted for covariates and normalised prior to model fitting (see above). Linear regression coefficients were averaged where multiple high quality aptamers targeted the same protein (**Supplementary Information**). False discovery rate (FDR) correction was subsequently applied across the 3,438 P-values (one per protein) for each PGS separately. Details on aptamer specificity and sensitivity are given in **Supplementary Table 2** for the 54 aptamers targeting the 49 PGS-associated proteins, and aptamer specific estimates of PGS on protein levels are detailed in **Supplementary Table 3** for the five PGS-associated proteins targeted by more than one aptamer (WFIKKN2, GPD1, IGFBP1, IGFBP2, and SHBG).

Polygenicity of PGS to protein associations

To quantify the polygenicity of PGS to protein associations (**Fig. 1c**, **Extended Data Fig. 3**) we performed a multi-step experiment to determine the proportion of the genome required to explain that association. First, we split the given PGS into separate scores for each of the 1,703 approximately independent LD blocks estimated in Europeans from the 1000 Genomes reference panel by Berisa & Pickrell 2016[48] (https://bitbucket.org/nygcresearch/ldetect-data/src/master/EUR/fourier_ls-all.bed). Next, we tested each of these 1,703 scores for association with the given protein (**Supplementary Data 3e**). Then, we retested the PGS to protein association, progressively removing independent LD blocks, at each step removing the LD block whose score had the strongest association with the protein. From this we quantified the polygenicity (**Fig. 1c**) based on the LD blocks needed to be removed from the given PGS in order to attenuate the PGS to protein association (so that the association P-value became > 0.05, **Supplementary Data 3f**) as the sum of removed LD block sizes / sum of all LD block sizes (*i.e.* proportion of genome removed). **Extended Data Fig. 3** shows the independent LD blocks contributing to the polygenicity of each PGS to protein association.

Independent contributions of PGS and pQTLs to protein levels

Multivariable linear regression models were fit for each protein on PGS levels and pQTL dosages to estimate their independent contributions to protein levels (**Supplementary Table 4**). The pQTLs used for each protein were: (1) conditionally independent pQTLs mapped in INTERVAL and published by Sun *et al.* 2018[22], which included both *cis* (within 1Mb of the encoding gene) and *trans* pQTLs passing the *trans*-significance threshold of $P < 1.5\times10^{-11}$; (2) *trans*-pQTLs with $P < 1.5\times10^{-11}$ (lead variant only) for proteins not published in Sun *et al.* 2018[22] (B2M, DUSP26, and FTMT); and (3) hierarchically significant *cis*-pQTLs (lead variant only) mapped in this study (**Supplementary Data 4**, **Supplementary Information**) for proteins without *cis*-pQTLs passing the trans-pQTL significance threshold above (ACY1, ADIPOQ, APOE, CST3, GPD1, PTPRU, SHBG, and UST).

Incident disease associations

PGSs and protein levels were tested for association with incident disease using Cox proportional hazards models adjusting for age and sex (**Fig. 2b, Extended Data Fig. 4**) using the survival package in R. The timescale used was time from baseline to first event of the relevant disease or to the latest available date in the hospital records (8th February 2020). PGSs and proteins were adjusted for covariates and normalised

prior to model fitting (see above). Cox model coefficients were averaged where multiple high quality aptamers targeted the same protein (**Supplementary Information**).

Incident disease events for AF, CAD, CKD, IS, and T2D were defined as first hospital episode for the closest matching CALIBER phenotype[41] (https://www.caliberresearch.org/portal). Incident AF events were defined as any hospital episode with ICD-10 code I48. Incident IS events were defined as any hospital episode with ICD-10 codes I63 or I69.3. For CAD we analysed incident MI events, defined as any hospital episode with ICD-10 codes I21–I23, I24.1, or I25.2. The closest matching CALIBER phenotype for T2D was for diabetes more broadly, including ICD-10 codes for any hospital episode for type 1 or type 2 diabetes or complications thereof: E10–E14, G59.0, G63.2, H28.0, H36.0, M14.2, N08.3, or O24.0–O24.3, however we note type 1 diabetics are not eligible to donate blood (https://my.blood.co.uk/knowledgebase/) and adult onset of type 1 diabetes is relatively rare compared to type 2 diabetes[49]. closest matching CALIBER phenotype for CKD was for end stage renal disease more broadly, which as defined as any hospital episode with ICD-10 codes N16.5, N18.5, T82.4, T86.1, Y60.2, Y61.2, Y84.1, Z49.1, Z49.2, Z94.0, and Z99.2.

Mediation analysis

Mediation analysis was used to identify causal proteins by identifying the PGS-associated proteins which partially mediate the association of PGS on disease (**Fig. 2d**). This approach uses the counterfactual framework to infer causal effects [28,50,51] and can be adapted to this setting as the arrow of causality between PGS and any associated phenotype can only flow in one direction as the PGS is fixed at conception (i.e. the underlying alleles in each person cannot be modified later in life by protein levels or the development of cardiometabolic disease). Here, we used the natural effects model developed by Vansteelandt *et al.* 2012[52], which is available in the medflex R package[53], to estimate natural indirect effects (effects of PGS on disease through protein levels) on the log odds scale by imputing unobserved counterfactuals. Standard errors were computed using the robust sandwich estimator[54], from which 95% confidence intervals and P-values were calculated. Multiple mediation analysis[55] was performed using the R package mma[56] to quantify the proportion of PGS to disease association mediated by the 15 causal T2D proteins.

Mendelian randomisation

Two-sample Mendelian randomisation[29] was also performed as an orthogonal approach to identify proteins which may play a causal role in disease (**Extended Data Fig. 5**, **Supplementary Table 5**,**6**). PGS-associated proteins were tested provided they had three or more independent by LD ($r^2 < 0.1$) *cis*-pQTLs after mapping pQTL to GWAS summary statistics (**Supplementary Information**), and provided the SomaLogic aptamer(s) did not have similar affinity for or comparable binding to multiple proteins or differential binding to specific isoforms (**Supporting Table 3**, **Supplementary Information**). In total, 12 of the 49 PGS-associated proteins could be tested (24%), substantially higher than the overall measured proteome (497 proteins, 14.5%). GWAS summary statistics were obtained from Nelson *et al.* 2017[13] for coronary artery disease (GCST004787), Wuttke *et al.* 2019[10] for chronic kidney disease (GCST008065), Malik *et al.* 2018[12] for ischaemic stroke (GCST006906) and Mahajan *et al.* 2018[11] for type 2 diabetes (GCST007518). In all cases, we used the GWAS summary statistics for the samples of recent European ancestry. For type 2 diabetes, we used the BMI-adjusted GWAS summary statistics in order to avoid false positive causal estimates arising where pQTLs influence type 2 diabetes risk through BMI rather than through the tested protein (horizontal pleiotropy). We used five different Mendelian Randomisation methods[57–60], each of which make use of information across 3 or more instruments to estimate causal

effects with each method differentially robust to different sources of bias, to obtain a consensus (median) estimate of causal effects of protein levels on disease risk (**Supporting Information**). We considered there to be a significant causal effect where $P < 0.05$ along with no significant evidence that causal effects were due to associations of the pQTLs with some other causal risk factor (horizontal pleiotropy; Egger intercept[60] $P > 0.05$). FDR correction was performed across all tested proteins for each disease separately. Analysis was performed using the R package MendelianRandomization[61]. Colocalisation analysis[62] was also performed where *cis*-pQTL instruments had $P < 1\times10^{-6}$ in the respective GWAS (**Supplementary Table 6**, **Supplementary Information**).

Drug targets

For each PGS-associated protein, a list of drugs that target or interact with the protein was downloaded from the DrugBank database[30] version 5.17 released on the 2nd of July 2020 (https://go.drugbank.com/releases/latest) (**Extended Data Table 5**, **Supplementary Table 7**). To obtain a list of drugs that counteract PGS effects and thus may have potential repurposing opportunities (**Table 1**), we filtered to drugs with approved status and not withdrawn status, then to drugs whose effect on the protein was in the opposite direction to the effect of the PGS on protein levels (e.g. inhibitors where increased PGS was associated with increased protein levels, **Supplementary Information**).

References

40. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems: Instruction manual*. (World Health Organization, 2004).

41. Kuan, V. *et al.* A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *Lancet Digit Health* **1**, e63–e77 (2019).

42. 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).

43. 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).

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

45. 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).

46. Lambert, S. A. *et al.* The Polygenic Score Catalog: an open database for reproducibility and systematic evaluation. *Nat. Genet.* **accepted**, (2021).

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

48. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283–285 (2016).

49. Atkinson, M. A., Eisenbarth, G. S. & Michels, A. W. Type 1 diabetes. *Lancet* **383**, 69–82 (2014).

50. Holland, P. W. Statistics and Causal Inference. *J. Am. Stat. Assoc.* **81**, 945–960 (1986).

51. Imai, K., Keele, L. & Tingley, D. A general approach to causal mediation analysis. *Psychol. Methods* **15**, 309–334 (2010).

52. Vansteelandt, S., Bekaert, M. & Lange, T. Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiol. Method.* **1**, 131–158 (2012).

549 53. Steen, J., Loeys, T., Moerkerke, B. & Vansteelandt, S. medflex : an R package for flexible mediation
550 analysis using natural effect models. *J. Stat. Softw.* **76**, (2017).
551 54. Liang, K.-Y. & Zeger, S. L. Longitudinal data analysis using generalized linear models. *Biometrika*
552 **73**, 13–22 (1986).
553 55. Yu, Q., Fan, Y. & Wu, X. General multiple mediation analysis with an application to explore racial
554 disparities in breast cancer survival. *J. Biom. Biostat.* **5**, 1–9 (2014).
555 56. Yu, Q. & Li, B. mma: an R package for mediation analysis with multiple mediators. *Journal of Open*
556 *Research Software* **5**, (2017).
557 57. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple
558 genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
559 58. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian
560 Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet.*
561 *Epidemiol.* **40**, 304–314 (2016).
562 59. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian
563 randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
564 60. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect
565 estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
566 61. Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian
567 randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).
568 62. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies
569 using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
570 63. Nikpay Majid, Soubeyrand Sebastien, Tahmasbi Rasool & McPherson Ruth. Multiomics Screening
571 Identifies Molecular Biomarkers Causally Associated With the Risk of Coronary Artery Disease.
572 *Circulation: Genomic and Precision Medicine* **13**, e002876 (2020).

64. 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**, 2130–2137 (2005).

65. 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**, 127–133 (2007).

66. Kojima, Y. *et al.* Progranulin expression in advanced human atherosclerotic plaque. *Atherosclerosis* **206**, 102–108 (2009).

67. 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).

68. 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).

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

70. Li, G.-S. *et al.* Do the mutations of C1GALT1C1 gene play important roles in the genetic susceptibility to Chinese IgA nephropathy? *BMC Med. Genet.* **10**, 101 (2009).

71. 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**, 785–792 (2009).

72. 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).

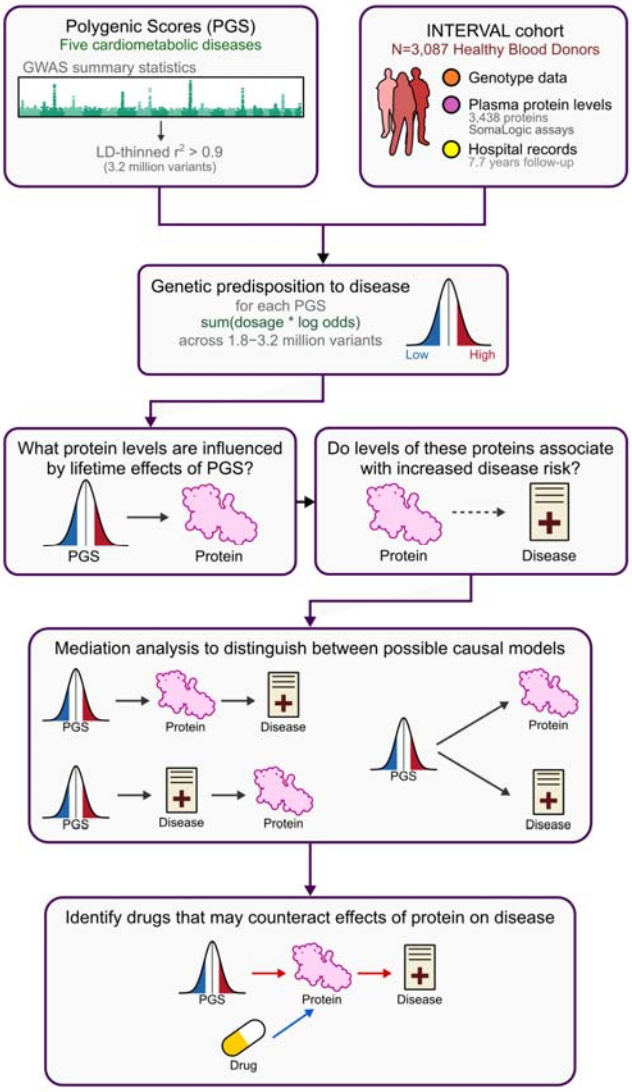
73. Madsen Tracy E. *et al.* Circulating SHBG (Sex Hormone-Binding Globulin) and Risk of Ischemic Stroke. *Stroke* **51**, 1257–1264 (2020).

596 74. Baumeier, C. *et al.* Caloric restriction and intermittent fasting alter hepatic lipid droplet proteome and
597 diacylglycerol species and prevent diabetes in NZO mice. *Biochim. Biophys. Acta* **1851**, 566–576
598 (2015).
599 75. Ngo, D. *et al.* Proteomic profiling reveals novel biomarkers and pathways in type 2 diabetes risk. *JCI*
600 *Insight* (2021) doi:10.1172/jci.insight.144392.
601 76. Spranger, J. *et al.* Adiponectin and protection against type 2 diabetes mellitus. *Lancet* **361**, 226–228
602 (2003).
603 77. Gudmundsdottir, V. *et al.* Circulating Protein Signatures and Causal Candidates for Type 2 Diabetes.
604 *Diabetes* **69**, 1843–1853 (2020).
605 78. Lau, W., Andrew, T. & Maniatis, N. High-Resolution Genetic Maps Identify Multiple Type 2
606 Diabetes Loci at Regulatory Hotspots in African Americans and Europeans. *Am. J. Hum. Genet.* **100**,
607 803–816 (2017).
608 79. Suckale, J. & Solimena, M. The insulin secretory granule as a signaling hub. *Trends Endocrinol.*
609 *Metab.* **21**, 599–609 (2010).
610 80. Kim-Muller, J. Y. *et al.* Aldehyde dehydrogenase 1a3 defines a subset of failing pancreatic β cells in
611 diabetic mice. *Nat. Commun.* **7**, 12631 (2016).
612 81. Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association
613 analysis. *Nat. Genet.* **42**, 579–589 (2010).
614 82. Rajwani, A. *et al.* Increasing circulating IGFBP1 levels improves insulin sensitivity, promotes nitric
615 oxide production, lowers blood pressure, and protects against atherosclerosis. *Diabetes* **61**, 915–924
616 (2012).
617 83. Xu, S., Lind, L., Zhao, L., Lindahl, B. & Venge, P. Plasma prolylcarboxypeptidase (angiotensinase C)
618 is increased in obesity and diabetes mellitus and related to cardiovascular dysfunction. *Clin. Chem.* **58**,
619 1110–1115 (2012).

620 84. Morris, A. P. *et al*. Large-scale association analysis provides insights into the genetic architecture and
621 pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
622 85. Grarup, N., Sandholt, C. H., Hansen, T. & Pedersen, O. Genetic susceptibility to type 2 diabetes and
623 obesity: from genome-wide association studies to rare variants and beyond. *Diabetologia* **57**, 1528–
624 1541 (2014).
625 86. Dwinovan, J., Colella, A. D., Chegeni, N., Chataway, T. K. & Sokoya, E. M. Proteomic analysis
626 reveals downregulation of housekeeping proteins in the diabetic vascular proteome. *Acta Diabetol.* **54**,
627 171–190 (2017).
628 87. Ding, E. L. *et al.* Sex hormone-binding globulin and risk of type 2 diabetes in women and men. *N.*
629 *Engl. J. Med.* **361**, 1152–1163 (2009).
630 88. Lopez, P. H. *et al*. Mice lacking sialyltransferase ST3Gal-II develop late-onset obesity and insulin
631 resistance. *Glycobiology* **27**, 129–139 (2017).
632 89. Kato, N. Insights into the genetic basis of type 2 diabetes. *J. Diabetes Investig*. **4**, 233–244 (2013).

Extended Data

Extended Data Figure 1: Study schematic.



636 **Extended Data Table 1: Cohort characteristics**

	Cohort characteristics
Participants	N=3,087
Women	N=1,528 (49%)
Age (years)	Median: 44.0 (Range: 18.0−75.6, IQR: 30.5−54.7)
Weight (kilograms; kg)	Median: 76.6 (Range: 49.4−177.0, IQR: 66.7−88.0)
Height (meters; m)	Median: 1.73 (Range: 1.07−2.41, IQR: 1.65−1.80)
BMI (kg/m^2)	Median: 25.5 (Range: 13.1−81.5, IQR: 23.1−28.5)

637

638 IQR: interquartile range. Body mass index (BMI) was computed from self-reported height and weight
639 (**Supplementary Information**).

Extended Data Table 2: Point estimates for PGS to protein to disease associations

	Association with PGS for coronary artery disease					Association with incident myocardial infarction (N=15 events)			Causal effect of PGS on disease risk through protein			
Protein	Beta	95% CI	P-value	FDR	Polygenicity	HR	95% CI	P-value	OR	95% CI	P-value	% PGS
APOE	0.081	[0.046, 0.12]	$6x10^{-6}$	0.004	8.8%	1.97	[1.17, 3.33]	0.011	1.058	[1.004, 1.11]	0.034	5.4%
TP53I11	0.081	[0.046, 0.12]	$6x10^{-6}$	0.004	9.2%	1.92	[1.14, 3.22]	0.014	1.060	[0.993, 1.13]	0.078	5.4%
CRYZL1	0.081	[0.046, 0.12]	$6x10^{-6}$	0.004	2.6%	1.84	[1.13, 3.00]	0.015	1.059	[0.999, 1.12]	0.052	5.3%
NPTX2	0.069	[0.034, 0.10]	$1x10^{-4}$	0.040	2.8%	1.55	[0.959, 2.52]	0.073	1.032	[0.977, 1.090]	0.26	3.0%
GGT2	0.081	[0.045, 0.12]	$7x10^{-6}$	0.004	6.2%	1.55	[0.954, 2.52]	0.077	1.036	[0.979, 1.097]	0.22	3.3%
HBQ1	0.071	[0.036, 0.11]	$7x10^{-5}$	0.026	3.9%	1.52	[0.937, 2.47]	0.090	1.034	[0.982, 1.088]	0.21	3.1%
CEI	0.091	[0.056, 0.13]	$4x10^{-7}$	0.001	1.9%	1.46	[0.89, 2.38]	0.14	1.026	[0.978, 1.077]	0.29	2.4%
DUSP26	0.073	[0.037, 0.11]	$6x10^{-5}$	0.024	12%	1.46	[0.88, 2.43]	0.14	1.025	[0.980, 1.073]	0.28	2.3%
PCDHB10	0.070	[0.035, 0.11]	$9x10^{-5}$	0.031	10%	1.38	[0.85, 2.24]	0.19	1.022	[0.981, 1.064]	0.30	2.0%
SHBG	-0.079	[-0.11, -0.044]	$1x10^{-5}$	0.005	17%	0.70	[0.38, 1.29]	0.25	1.010	[0.980, 1.040]	0.53	0.9%
GRN	0.082	[0.047, 0.12]	$5x10^{-6}$	0.004	0.1%	1.33	[0.78, 2.26]	0.29	1.010	[0.987, 1.034]	0.38	1.0%
	Association with PGS for type 2 diabetes					Association with incident diabetes (N=27 events)						
IGFBP2	-0.095	[-0.13, -0.061]	$5x10^{-8}$	$6x10^{-5}$	15%	0.44	[0.31, 0.64]	$2x10^{-5}$	1.097	[1.037, 1.16]	0.001	13%
CFH	0.088	[0.054, 0.12]	$4x10^{-7}$	$3x10^{-4}$	21%	2.35	[1.59, 3.48]	$2x10^{-5}$	1.064	[1.024, 1.11]	0.002	8.8%
CPM	0.096	[0.062, 0.13]	$3x10^{-8}$	$6x10^{-5}$	22%	1.97	[1.36, 2.87]	$4x10^{-4}$	1.061	[1.021, 1.10]	0.003	8.8%
SHBG	-0.11	[-0.14, -0.073]	$8x10^{-10}$	$3x10^{-6}$	23%	0.41	[0.26, 0.62]	$4x10^{-5}$	1.094	[1.030, 1.16]	0.004	13%
GHR	0.071	[0.036, 0.11]	$6x10^{-5}$	0.014	11%	2.28	[1.53, 3.39]	$5x10^{-5}$	1.056	[1.013, 1.10]	0.009	7.9%
CCDC126	-0.070	[-0.10, -0.035]	$7x10^{-5}$	0.014	10%	0.46	[0.32, 0.68]	$8x10^{-5}$	1.047	[1.009, 1.088]	0.016	6.9%
PRCP	0.077	[0.043, 0.11]	$7x10^{-6}$	0.003	12%	1.89	[1.29, 2.76]	$1x10^{-3}$	1.046	[1.008, 1.086]	0.017	6.6%
PTPRU	0.074	[0.040, 0.11]	$2x10^{-5}$	0.006	18%	1.86	[1.28, 2.70]	0.001	1.043	[1.006, 1.081]	0.022	6.1%
IGFBP1	-0.073	[-0.11, -0.038]	$3x10^{-5}$	0.009	18%	0.54	[0.36, 0.81]	0.003	1.046	[1.005, 1.088]	0.027	6.7%
ADIPOQ	-0.080	[-0.11, -0.046]	$4x10^{-6}$	0.002	9.3%	0.52	[0.36, 0.76]	$8x10^{-4}$	1.060	[1.006, 1.12]	0.028	8.8%
ACY1	0.077	[0.043, 0.11]	$8x10^{-6}$	0.003	9.1%	1.72	[1.19, 2.49]	0.004	1.036	[1.002, 1.071]	0.035	5.1%
CFI	0.083	[0.049, 0.12]	$2x10^{-6}$	$1x10^{-3}$	23%	1.69	[1.14, 2.50]	0.009	1.032	[1.002, 1.064]	0.037	4.6%
ADH4	0.070	[0.035, 0.10]	$7x10^{-5}$	0.014	6.5%	1.72	[1.17, 2.54]	0.006	1.030	[1.001, 1.059]	0.043	4.2%
QPCTL	-0.065	[-0.099, -0.030]	$2x10^{-4}$	0.027	17%	0.51	[0.35, 0.74]	$5x10^{-4}$	1.043	[1.000, 1.088]	0.048	6.4%
INHBC	0.066	[0.032, 0.100]	$2x10^{-4}$	0.022	22%	1.65	[1.11, 2.45]	0.013	1.025	[1.000, 1.050]	0.049	3.6%
APOF	-0.065	[-0.099, -0.031]	$2x10^{-4}$	0.026	5.0%	0.46	[0.31, 0.67]	$6x10^{-5}$	1.040	[0.998, 1.085]	0.061	5.9%
GFRA1	0.068	[0.034, 0.10]	$9x10^{-5}$	0.015	5.3%	1.85	[1.26, 2.72]	0.002	1.042	[0.996, 1.090]	0.074	6.0%
PRSS1	-0.086	[-0.12, -0.052]	$1x10^{-6}$	$5x10^{-4}$	18%	0.58	[0.40, 0.84]	0.004	1.061	[0.988, 1.14]	0.10	8.5%
FAH	0.068	[0.033, 0.10]	$1x10^{-4}$	0.018	14%	1.73	[1.19, 2.52]	0.004	1.029	[1.000, 1.058]	0.054	4.1%
ST3GAL2	0.067	[0.032, 0.10]	$1x10^{-4}$	0.021	19%	1.66	[1.15, 2.41]	0.007	1.019	[0.992, 1.047]	0.17	2.8%
FBP1	0.070	[0.036, 0.10]	$7x10^{-5}$	0.014	8.6%	1.57	[1.072, 2.30]	0.021	1.014	[0.992, 1.036]	0.20	2.0%
WFIKKN2	-0.070	[-0.10, -0.036]	$6x10^{-5}$	0.014	12%	0.65	[0.44, 0.947]	0.025	1.032	[0.991, 1.075]	0.11	4.6%
RIDA	0.063	[0.029, 0.097]	$3x10^{-4}$	0.036	7.5%	1.54	[1.054, 2.25]	0.026	1.020	[0.992, 1.049]	0.16	2.9%
TIMP4	-0.069	[-0.10, -0.035]	$8x10^{-5}$	0.015	19%	0.65	[0.44, 0.959]	0.030	1.025	[0.994, 1.058]	0.12	3.7%
HS6ST2	-0.067	[-0.10, -0.033]	$1x10^{-4}$	0.021	23%	0.66	[0.44, 0.991]	0.045	1.017	[0.999, 1.035]	0.070	2.4%
GPD1	0.066	[0.031, 0.10]	$2x10^{-4}$	0.024	12%	1.44	[0.982, 2.11]	0.062	1.018	[0.992, 1.045]	0.18	2.6%
MUSK	0.090	[0.056, 0.12]	$3x10^{-7}$	$2x10^{-4}$	16%	1.44	[0.982, 2.10]	0.062	1.025	[0.996, 1.056]	0.097	3.5%
CHGB	-0.075	[-0.11, -0.041]	$2x10^{-5}$	0.005	10%	0.74	[0.51, 1.081]	0.12	1.020	[0.988, 1.052]	0.23	2.8%
CYB5R3	-0.062	[-0.096, -0.027]	$4x10^{-4}$	0.049	9.1%	0.74	[0.51, 1.086]	0.12	1.015	[0.995, 1.036]	0.13	2.2%
MSMP	-0.068	[-0.10, -0.034]	$1x10^{-4}$	0.017	8.7%	0.74	[0.51, 1.087]	0.13	1.021	[0.985, 1.060]	0.26	3.0%
FAM20A	0.063	[0.029, 0.098]	$3x10^{-4}$	0.036	12%	1.27	[0.87, 1.86]	0.21	1.009	[0.991, 1.028]	0.31	1.3%
	Association with PGS for chronic kidney disease					Association with incident end stage renal disease (N=0 events)						
VWC2	0.094	[0.056, 0.13]	$1x10^{-6}$	0.002	21%							
B2M	0.094	[0.056, 0.13]	$1x10^{-6}$	0.002	9.4%							
CST3	0.085	[0.047, 0.12]	$1x10^{-5}$	0.010	7.0%							
UST	-0.081	[-0.12, -0.042]	$4x10^{-5}$	0.023	12%							
C1GALT1C1	-0.079	[-0.12, -0.041]	$5x10^{-5}$	0.026	5.6%							
FTMT	-0.11	[-0.15, -0.076]	$4x10^{-9}$	$1x10^{-5}$	0.1%							
PRSS3	0.082	[0.043, 0.12]	$3x10^{-5}$	0.020	0.2%							
PDE4D/A	-0.076	[-0.11, -0.038]	$9x10^{-5}$	0.040	0.1%							
	Association with PGS for ischaemic stroke					Association with incident ischaemic stroke (N=3 events)						
SHBG	-0.076	[-0.11, -0.042]	$1x10^{-5}$	0.041	17%							

Beta: standard deviation change in protein levels per standard deviation increase in PGS. 95% CI: 95% confidence interval. HR: hazard ratio for incident disease conferred per standard deviation increase in protein levels. OR: odds ratio for incident disease adjusting for age and sex conferred through each protein per standard deviation increase in PGS. % PGS: Percentage of total effect of PGS on incident disease conferred through protein levels. The total odds ratio for MI conferred per standard deviation increase in CAD PGS was 2.94 (95% CI: 1.69–5.31, P-value: 2×10^{-4}). The total odds ratio for T2D conferred per standard deviation increase T2D PGS was 2.00 (95% CI: 1.37–2.96, P-value: 4×10^{-4}). Point estimates are greyed out where P-value > 0.05

Extended Data Table 3: Information about each PGS associated protein

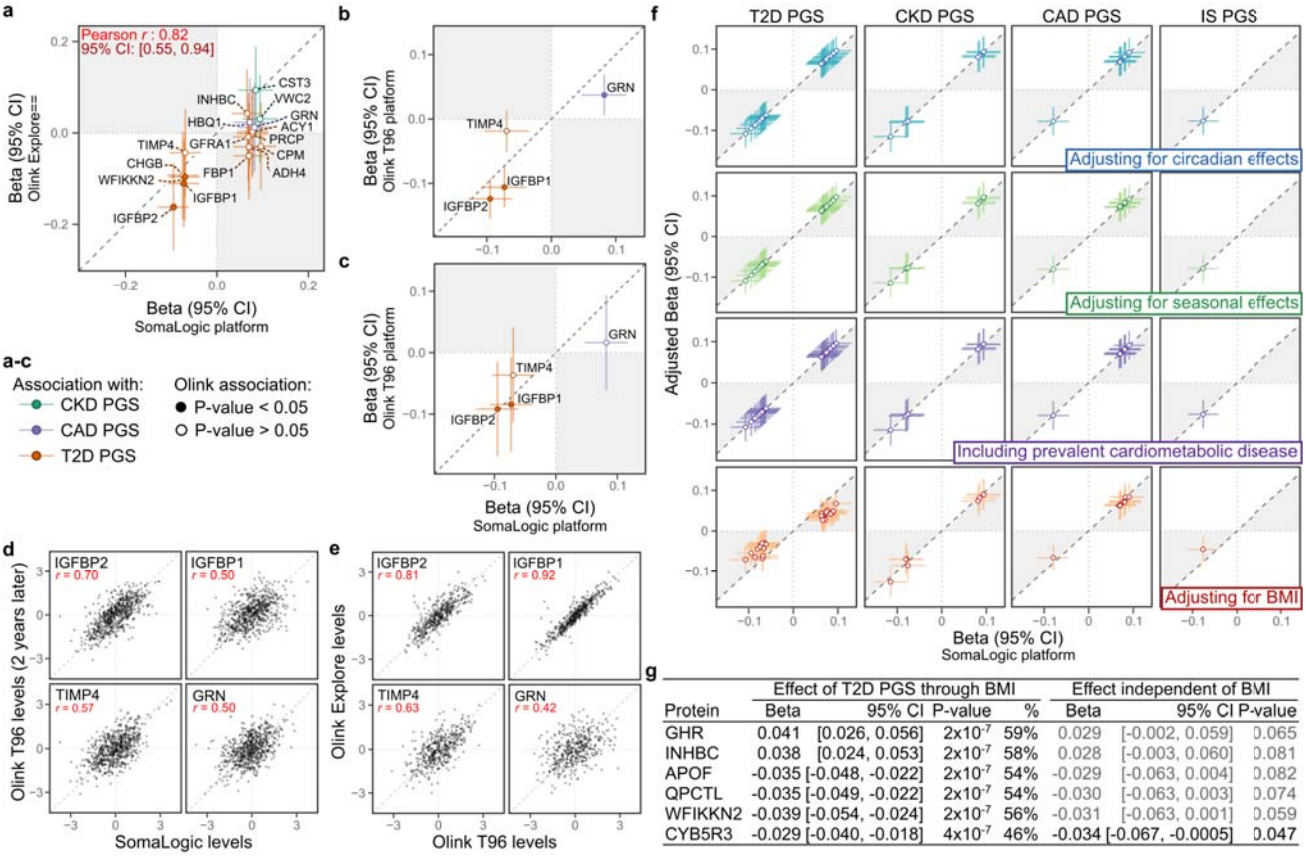
Protein	UniProt	Gene	Chr	Start	PGS	Aptamer	Aptamer target
ACY1	Q03154	*ACY1*	3	52,017,300	T2D	3343-1	Aminoacylase-1
ADH4	P08319	*ADH4*	4	100,044,832	T2D	8325-37	Alcohol dehydrogenase 4
ADIPOQ	Q15848	*ADIPOQ*	3	186,560,463	T2D	3554-24	Adiponectin
APOE*	P02649	*APOE*	19	45,409,039	CAD	2418-55	Apolipoprotein E, isoforms E3 and E4
APOF	Q13790	*APOF*	12	56,754,355	T2D	12370-30	Apolipoprotein F
B2M	P61769	*B2M*	15	45,003,685	CKD	3485-28	Beta-2-microglobulin
C1GALT1C1	Q96EU7	*C1GALT1C1*	X	119,759,529	CKD	5735-54	C1GALT1-specific chaperone 1
CCDC126	Q96EE4	*CCDC126*	7	23,636,998	T2D	6388-21	Coiled-coil domain-containing protein 126
CEI	Q86SI9	*C5orf38*	5	2,752,058	CAD	6378-2	Protein CEI
CFH	P08603	*CFH*	1	196,621,008	T2D	4159-130	Complement factor H
CFI	P05156	*CFI*	4	110,661,848	T2D	2567-5	Complement factor I
CHGB	P05060	*CHGB*	20	5,891,974	T2D	8235-48	Secretogranin-1
CPM	P14384	*CPM*	12	69,244,955	T2D	7768-10	Carboxypeptidase M
CRYZL1	O95825	*CRYZL1*	21	34,961,647	CAD	9207-60	Quinone oxidoreductase-like protein 1
CST3	P01034	*CST3*	20	23,608,534	CKD	2609-59	Cystatin-C
CYB5R3	P00387	*CYB5R3*	22	43,013,846	T2D	7215-18	NADH-cytochrome b5 reductase 3
DUSP26	Q9BV47	*DUSP26*	8	33,448,848	CAD	8967-6	Dual specificity protein phosphatase 26
FAH	P16930	*FAH*	15	80,445,233	T2D	11424-4	Fumarylacetoacetase
FAM20A	Q96MK3	*FAM20A*	17	66,531,257	T2D	6433-57	Pseudokinase FAM20A
FBP1	P09467	*FBP1*	9	97,365,415	T2D	7206-20	Fructose-1,6-bisphosphatase 1
FTMT	Q8N4E7	*FTMT*	5	121,187,650	CKD	8048-9	Ferritin, mitochondrial
GFRA1	P56159	*GFRA1*	10	117,816,436	T2D	3314-74	GDNF family receptor alpha-1
GGT2	P36268	*GGT2*	22	21,562,261	CAD	6334-9	Inactive gamma-glutamyltranspeptidase 2
GHR	P10912	*GHR*	5	42,423,577	T2D	2948-58	Growth hormone receptor
GPD1	P21695	*GPD1*	12	50,497,602	T2D	13697-51 11081-1	Glycerol-3-phosphate dehydrogenase [NAD(+)], cytoplasmic
GRN	P28799	*GRN*	17	42,422,491	CAD	4992-49	Granulins
HBQ1	P09105	*HBQ1*	16	230,333	CAD	7965-25	Hemoglobin subunit theta-1
HS6ST2	Q96MM7	*HS6ST2*	X	131,760,038	T2D	13524-25	Heparan-sulfate 6-O-sulfotransferase 2
IGFBP1	P08833	*IGFBP1*	7	45,927,959	T2D	13741-36 2771-35	Insulin-like growth factor-binding protein 1
IGFBP2	P18065	*IGFBP2*	2	217,498,127	T2D	2570-72 8469-41	Insulin-like growth factor-binding protein 2
INHBC	P55103	*INHBC*	12	57,828,543	T2D	6408-2	Inhibin beta C chain
MSMP	Q1L6U9	*MSMP*	9	35,752,987	T2D	8080-24	Prostate-associated microseminoprotein
MUSK	O15146	*MUSK*	9	113,430,935	T2D	11547-84	Muscle, skeletal receptor tyrosine-protein kinase
NPTX2	P47972	*NPTX2*	7	98,246,597	CAD	6521-35	Neuronal pentraxin-2
PCDHB10	Q9UN67	*PCDHB10*	5	140,571,952	CAD	9963-19	Protocadherin beta-10
PDE4D/A*	Q08499 P27815	*PDE4D* *PDE4A*	5 19	58,264,865 10,527,449	CKD	5255-22	Combined levels of cAMP-specific 3',5'-cyclic phosphodiesterase 4D and 4A
PRCP	P42785	*PRCP*	11	82,535,409	T2D	5722-78	Lysosomal Pro-X carboxypeptidase
PRSS1	P07477	*PRSS1*	7	142,457,319	T2D	3049-61	Trypsin-1
PRSS3	P35030	*PRSS3*	9	33,750,464	CKD	3479-71	Trypsin-3
PTPRU	Q92729	*PTPRU*	1	29,563,028	T2D	8337-65	Receptor-type tyrosine-protein phosphatase U
QPCTL	Q9NXS2	*QPCTL*	19	46,195,741	T2D	8866-53	Glutaminyl-peptide cyclotransferase-like protein
RIDA	P52758	*RIDA*	8	99,114,567	T2D	14636-25	Ribonuclease UK114
SHBG	P04278	*SHBG*	17	7,517,382	CAD, IS, T2D	7909-37 4929-55	Sex hormone-binding globulin
ST3GAL2	Q16842	*ST3GAL2*	16	70,413,338	T2D	6281-51	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 2
TIMP4	Q99727	*TIMP4*	3	12,194,568	T2D	6462-12	Metalloproteinase inhibitor 4
TP53I11	O14683	*TP53I11*	11	44,907,454	CAD	13022-20	Tumor protein p53-inducible protein 11
UST	Q9Y2C2	*UST*	6	149,068,063	CKD	8364-74	Uronyl 2-sulfotransferase
VWC2	Q2TAL6	*VWC2*	7	49,813,257	CKD	11121-56	Brorin
WFIKKN2	Q8TEU8	*WFIKKN2*	17	48,912,011	T2D	3235-50 13408-23	WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 2

Aptamer: Sequence ID for the SomaLogic aptamer(s) targeting the protein. A * next to the protein name indicates the aptamer(s) binds to specific isoforms of the listed protein or binds to multiple proteins; see Aptamer target column. Extended details on aptamer sensitivity and specificity can be found in **Supplementary Table 2.**

Extended Data Table 4: Previous evidence for PGS-associated proteins in disease

Coronary artery disease	
Disease association previously observed	APOE[24,25], CEI[63], GGT2[64,65], GRN[66], SHBG[67–69]
No reported association	CRYZL1, DUSP26, HBQ1, NPTX2, PCDBH10, TP53I11
Chronic kidney disease	
Disease association previously observed	B2M[23], C1GALT1C1[70], CST3[23], PDE4A[71], PDE4D[71], VWC2[72]
No reported association	FTMT, PRSS3, UST
Ischaemic stroke	
Disease association previously observed	SHBG[73]
Type 2 diabetes	
Disease association previously observed	ACY1[74,75], ADIPOQ[75,76], APOF[77], CCDC126[78], CFH[75], CFI[75], CHGB[79], CPM[77], CYB5R3[80], FAH[81], FBP1[26,32], GFRA1[75,77], GHR[36,77], GPD1[77], HS6ST2[77], IGFBP1[82], IGFBP2[77], INHBC[77], MSMP[77], PRCP[77,83], PRSS1[77], PTPRU[77], QPCTL[84,85], RIDA[77,86], SHBG[87], ST3GAL2[88], TIMP4[81,89], WFIKKN2[75,77]
No reported association	MUSK, ADH4, FAM20A

Extended Data Figure 2: Robustness of PGS to protein associations



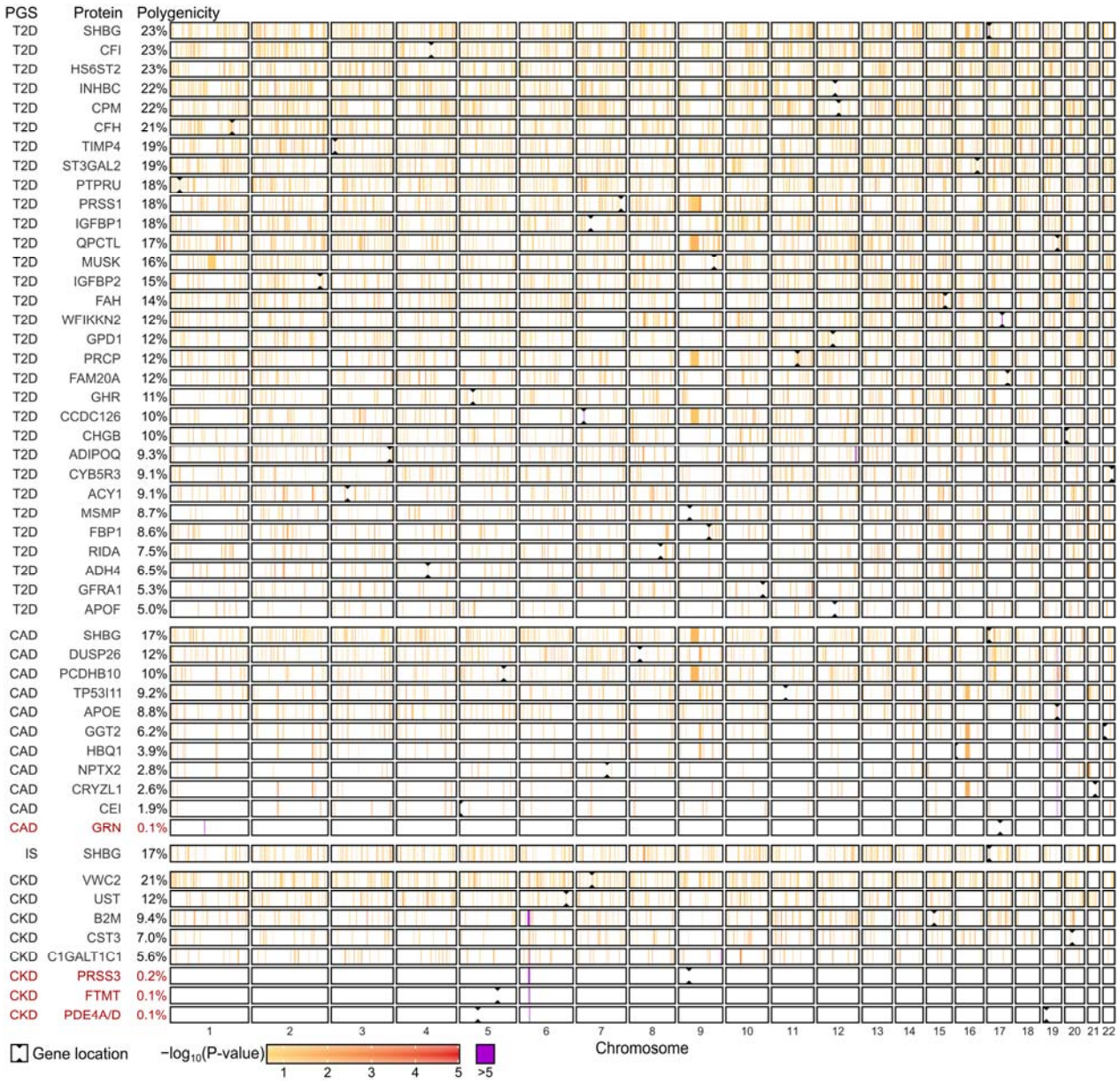
g

	Effect of T2D PGS through BMI				Effect independent of BMI		
Protein	Beta	95% CI	P-value	%	Beta	95% CI	P-value
GHR	0.041	[0.026, 0.056]	$2x10^{-7}$	59%	0.029	[-0.002, 0.059]	0.065
INHBC	0.038	[0.024, 0.053]	$2x10^{-7}$	58%	0.028	[-0.003, 0.060]	0.081
APOF	-0.035	[-0.048, -0.022]	$2x10^{-7}$	54%	-0.029	[-0.063, 0.004]	0.082
QPCTL	-0.035	[-0.049, -0.022]	$2x10^{-7}$	54%	-0.030	[-0.063, 0.003]	0.074
WFIKKN2	-0.039	[-0.054, -0.024]	$2x10^{-7}$	56%	-0.031	[-0.063, 0.001]	0.059
CYB5R3	-0.029	[-0.040, -0.018]	$4x10^{-7}$	46%	-0.034	[-0.067, -0.0005]	0.047

a-c) Robustness of PGS to protein associations to proteomics technology. **c)** Longitudinal stability of PGS to protein associations. **d)** Longitudinal stability of protein levels. **d-e)** Robustness of protein levels to proteomics technology. **f)** Robustness of PGS to protein associations to environmental and physiological confounding. **g)** Mediation of PGS to protein associations through body mass index (BMI) for six proteins associated with PGS for type 2 diabetes. **a)** Compares associations between PGSs and protein levels quantified by SomaLogic SOMAscan aptamers (x-axis; **Fig. 1b**) to associations with protein levels quantified using the Olink Explore platform in 418 independent INTERVAL participants (y-axis) with no prevalent cardiometabolic disease (**Supplementary Information**). In total 1,463 proteins were quantified by the Olink Explore platform, including 907 quantified by the SomaLogic platform, and among these 16 of the 49 PGS-associated proteins. Points correspond to PGS to protein level association beta estimates, and the bars to their 95% confidence intervals. **b)** Compares associations between PGSs and protein levels quantified by SomaLogic SOMAscan aptamers (x-axis; **Fig. 1b**) to associations with protein levels quantified using the Olink T96 platform in 3,848 independent INTERVAL participants. In total 265 proteins were quantified by the Olink T96 platform, including 224 quantified by the SomaLogic platform, and among these, 4 of the 49 PGS-associated proteins. **c)** Compares PGS to protein associations in 646 participants with protein levels quantified by both the SomaLogic platform and Olink T96 platform (from blood samples taken after 2 years of follow-up). **a-c**) share common x-axes and legend. Point estimates for associations between PGS and protein levels assessed by Olink proteomics in each panel are given in **Supplementary Data 3b**. **d)** Compares protein levels quantified by the SomaLogic platform (x-axes) to protein levels quantified by the Olink T96 platform (y-axes) after two years of follow-up in 646 participants. **e)** Compares protein levels quantified by the Olink T96 platform (x-axes) to protein levels

quantified by the Olink Explore platform (y-axes). **f)** Compares PGS to protein associations before (x-axes; **Fig. 1a**) and after (y-axes) adjustment for circadian effects (time of day of blood draw), adjustment for seasonal effects (date of blood draw), when including 87 participants with prevalent cardiometabolic disease, and adjustment for BMI. To capture the potentially non-linear effects of circadian rhythm and season on protein levels both were treated as categorical variables with 10 groups of equal length duration, using the group with the largest sample size as the reference in the model (**Supplementary Information**). Point estimates in sensitivity analyses are given in **Supplementary Data 3c**. **g)** For the six proteins whose association T2D PGS was attenuated ($P > 0.05$; **Extended Data Fig. 2f**) gives, from mediation analysis (**Online Methods**), the estimated effect of T2D PGS on the protein levels through BMI (standard deviation change in protein levels through BMI per standard deviation increase in T2D PGS) and the estimated effect of T2D PGS on protein levels independent of BMI. 95% CI: 95% confidence interval.

Extended Data Figure 3: Polygenicity of PGS to protein associations

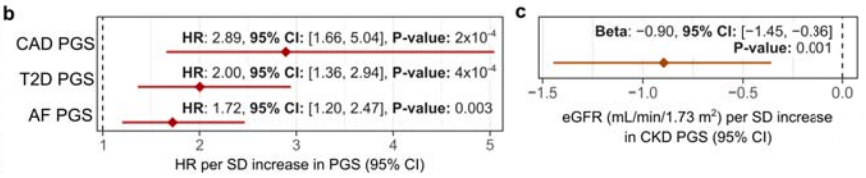


Linkage disequilibrium (LD) blocks contributing to each PGS to protein association (**Online Methods**). Each PGS was partitioned into 1,703 approximately independent LD blocks[48] then tested for association with each protein (**Supplementary Data 3e**). To obtain the set of LD blocks contributing to each PGS to protein association, LD blocks were removed from the PGS in ascending order by association P-value until the PGS to protein association was attenuated ($P > 0.05$; **Supplementary Data 3f**). Here, associations ($-\log_{10}$ P-values) between protein levels and LD blocks contributing to the PGS to protein association are shown. Regions in white contain LD blocks that did not contribute to the PGS to protein association. The total percentage of the genome contributing to the PGS to protein association (polygenicity; **Fig. 1c**) is shown on the right. PGS to protein associations listed in red are those explained by pQTLs (*cis* and/or *trans*) rather than polygenic.

Extended Data Figure 4: Incident disease and PGS validity

a

Incident Disease	Events	Men	Age of onset
Atrial fibrillation	33	25	64.2 (59.2–69.8)
Type 2 Diabetes	27	18	55.5 (47.7–63.3)
Myocardial infarction	15	12	57.2 (58.8–65.3)
Ischaemic stroke	3	1	73.1 (68.8–75.6)
Chronic kidney disease	0	–	–
Any of the above:	74	54	62.1 (53.5–67.7)



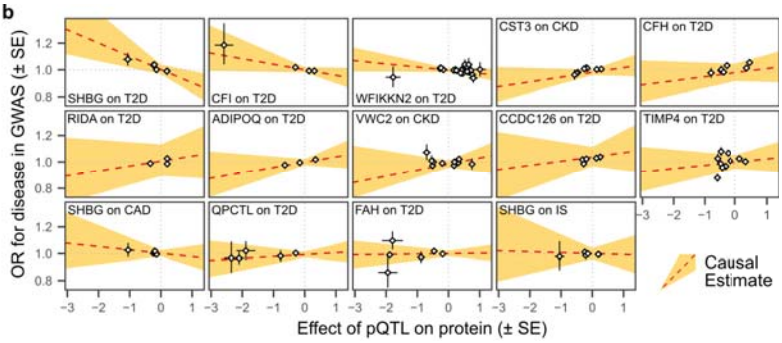
a) Incident disease events over the 7.7 year of follow-up in the 3,087 INTERVAL participants. Endpoint: incident disease definition available in INTERVAL for the relevant PGS, as defined by CALIBER phenotyping algorithms (**Online Methods**). Age of onset: median age of first hospitalisation with the respective endpoint. Numbers in brackets gives the interquartile range. **b)** Hazard ratio (HR) and 95% confidence interval (95% CI) conferred per standard deviation increase of the respective PGS on risk of hospitalisation with the relevant endpoint. CAD PGS was tested for incident myocardial infarction. Hazard ratios were fit using cox proportional hazards models, adjusting for age and sex, and 10 genetic PCs. **c)** Association between PGS for chronic kidney disease with estimated glomerular filtration rate (eGFR), a marker of renal function used in chronic kidney disease diagnosis (**Online Methods**): decreased eGFR is indicative of reduced renal function. EGFR was computed from serum creatinine in 3,307 participants using the CKD-EPI equation (**Supplementary Information**). Association was fit with linear regression adjusting for age and sex, and 10 genetic PCs.

Extended Data Figure 5: Mendelian randomisation analysis

a

		Causal Estimate			Pleiotropy
Protein	Outcome	OR	95% CI	P-value	P-value
SHBG	T2D	0.901*	[0.85, 0.953]	0.001	0.91
CFI	T2D	0.958	[0.916, 0.996]	0.031	0.99
WFIKKN2	T2D	0.976*	[0.954, 0.998]	0.036	0.70
CST3	CKD	1.035*	[0.993, 1.079]	0.10	0.08
CFH	T2D	1.03*	[0.992, 1.064]	0.14	0.52
RIDA	T2D	1.037*	[0.986, 1.086]	0.17	0.92
ADIPOQ	T2D	1.04	[0.983, 1.097]	0.18	0.86
VWC2	CKD	1.046*	[0.967, 1.094]	0.23	0.25
CCDC126	T2D	1.032	[0.969, 1.11]	0.34	0.31
TIMP4	T2D	1.024	[0.982, 1.068]	0.34	0.95
SHBG	CAD	0.975*	[0.923, 1.031]	0.38	0.82
QPCTL	T2D	1.014	[0.975, 1.055]	0.49	0.53
FAH	T2D	1.003*	[0.977, 1.032]	0.83	0.98
SHBG	IS	0.993*	[0.919, 1.066]	0.92	0.97

b



a) Causal effects of protein levels on disease risk estimated through two-sample Mendelian randomisation analysis of pQTL summary statistics and disease GWAS summary statistics (**Online Methods**). OR: consensus estimate of the odds ratio conferred per standard deviation increase in protein levels across five Mendelian randomisation methods (**Supplementary Information**; **Supplementary Table 5**). * Estimated causal effect is directionally consistent with PGS to protein to disease associations in **Fig. 2**. 95% CI: 95% confidence interval. Pleiotropy P-value: P-value for the intercept term in Egger regression, which indicates where $P < 0.05$, confounding of the causal estimate by associations between genetic instruments (*cis*-pQTLs) with multiple disease risk factors (horizontal pleiotropy). Entries are greyed out where $P > 0.05$. **b)** Dose response curves showing the estimated causal effect of changes in protein levels on disease risk for each protein and disease. The slope of the orange dashed line corresponds to the estimated causal effect (Odds Ratio from **Extended Data Fig. 5a**). The yellow ribbon shows the 95% confidence interval for the estimated causal effect (slope), accounting also for the 95% confidence interval for the intercept term in Egger regression. Points on each plot show the *cis*-pQTLs used as genetic instruments for each test (**Supplementary Table 6**). On the x-axes, points show the standard deviation change in protein levels per copy of the minor allele, and horizontal bars indicate the standard error. On the y-axes, points show odds ratio conferred per copy of the minor allele, and vertical bars indicate the standard error.

Extended Data Table 5: PGS-associated drug targets

PGS	Protein	Drugs	Approved	Summary of approved compound usage
T2D	ACY1	3	3	Malnutrition, overdose.
T2D	ADH4	3	2	Female reproductive disorders, infection control.
CAD	APOE	5	5	Zinc deficiency.
CKD	B2M	4	2	Pain management, diarrhoea.
T2D	CFH	5	5	Zinc deficiency, malnutrition, ear and respiratory infections.
T2D	CFI	3	3	Zinc deficiency, malnutrition, ear and respiratory infections.
T2D	CYB5R3	3	3	Fertility and reproductive treatments, zinc deficiency, vitamin B12 deficiency
T2D	FBP1	11	1	Artificial sweetener.
T2D	GHR	3	2	Acromegaly, dwarfism, idiopathic short stature, HIV weight loss.
T2D	GPD1	2	2	Glycaemic control, type 2 diabetes, female reproductive disorders.
T2D	IGFBP1	1	1	Growth failure due to IGF-1 deficiency
T2D	IGFBP2	1	1	Growth failure due to IGF-1 deficiency
T2D	MUSK	1	1	Rheumatoid arthritis, chronic immune thrombocytopenia
CKD	PDE4A	16	10	Respiratory diseases, atopic dermatitis, hypertension, congestive heart failure, bowel disorders.
CKD	PDE4D	23	5	Respiratory diseases, atopic dermatitis, hypertension.
CAD IS T2D	SHBG	68	49	Fertility and reproductive treatments, cancers, mental health, developmental disorders, hypertension, high cholesterol.
T2D	FAH	3	0	-
T2D	PRSS1	101	0	-
CKD	PRSS3	8	0	-

Number of drugs interacting with each PGS-associated protein, the number approved for therapeutic use, and a summary of the therapeutic uses of these compounds. Information was retrieved from DrugBank version 5.17 (**Online Methods**). **Supplementary Table 7** provides details for each of the 236 drugs or compounds targeting any PGS-associated protein.