Multi-ancestry polygenic risk scores for the prediction of type 2 diabetes and complications in diverse ancestries

Alicia Huerta-Chagoya, Joohyun Kim*, Ravi Mandla†, Yingchang Lu†, Ken Suzuki†, Lauren E Petty†, Hong Kiat Ng†, Jaewon Choi†, Simon Lee†, Madhusmita Rout†, Kuang Lin†, Katherine Taylor†, ENSA Genomics Consortium‡, Genes & Health Research Team§, VA Million Veteran Program¶, Carlos A Aguilar-Salinas, Lourdes García-García, Clicerio González-Villalpando, Christopher A Haiman, Young Jin Kim, Soo Heon Kwak, Aaron Leong, Ruth J F Loos, Andres Moreno-Estrada, Andrew P Morris, Lorena Orozco, Jerome I Rotter, Dharambir Sanghera, Teresa Tusie-Luna, Benjamin F Voight, Marijana Vujkovic, Robin G Walters, Tian Ge ||, Alisa K Manning ||, Marie Loh ||, Jennifer E Below ||, Xueling Sim ||, Josep M Mercader ||, Maggie C Y Ng ||, D-PRISM Consortium***

Summary

Background Polygenic risk scores (PRSs) improve prediction of the development of type 2 diabetes over the use of clinical risk factors alone; however, they perform poorly in populations of non-European ancestry, limiting their global clinical utility. We aimed to deliver comprehensive and rigorously tested multi-ancestry PRSs for prediction in type 2 diabetes.

Methods We conducted meta-analyses using data from type 2 diabetes genome-wide association studies (GWAS) across cohorts from five major global ancestries: European, African or African American, Admixed American, South Asian, and East Asian. We used summary statistics from the GWAS to construct single-ancestry PRSs (using the continuous-shrinkage PRS-CS method) and multi-ancestry PRSs (using the PRS-CSx method), and constructed ancestry-specific linkage disequilibrium panels to model pairwise correlations between single-nucleotide polymorphisms in GWAS during PRS construction. Models were validated for association with type 2 diabetes in at least four independent cohorts per ancestry. The effect sizes of PRSs were estimated as the odds ratio (OR) per SD of the PRS, and ORs for individuals at the 90th, 95th, and 97·5th PRS percentiles were compared with the IQR as a reference. We also tested our PRS models for prediction of diabetes incidence with or without additional clinical factors, as well as microvascular complications and comorbidities.

Findings Our analysis used data from 409 959 individuals with type 2 diabetes and 1 983 345 controls: respectively, 359 819 and 1 825 729 indivduals were included in the GWAS dataset, with 10 992 and 31 792 individuals in the training dataset and 39 148 and 125 824 individuals in the validation dataset. The best predictive performance for the single-ancestry PRSs was in European (incremental AUC 0·07–0·14) and East Asian (0·02–0·16) ancestries, whereas prediction was poorer for African or African American (0·02–0·03), Admixed American (0·02–0·04), and South Asian (0·02–0·04) ancestries, correlating with sample sizes in the GWAS. Compared with single-ancestry PRSs, our multi-ancestry PRSs showed higher effect sizes and smaller 95% CIs across all ancestries: OR per SD 1·73 (95% CI 1·67–1·80) in African or African American, 2·82 (2·67–2·97) in Admixed American, 2·45 (2·36–2·54) in East Asian, 2·36 (2·32–2·41) in European, and 2·23 (2·05–2·42) in South Asian ancestries. Individuals in the 97·5th PRS percentile had a 3–7 times increased risk of type 2 diabetes compared with those in the IQR (OR 3·43 [95% CI 2·80–4·21] in African or African American, 7·47 [5·64–9·89] in Admixed American, 6·62 [5·58–7·85] in East Asian, 6·25 [5·72–6·82] in European, and 4·50 [2·70–7·53] in South Asian ancestries). These PRSs were also associated with earlier onset of type 2 diabetes, higher risk of developing microvascular complications, and provide additional predictive value beyond clinical factors. In individuals with type 2 diabetes, the association between multi-ancestry PRSs and risk of microvascular complications and comorbidity was studied in populations of African, Admixed American, and European ancestries and was significant in all three ancestry groups for diabetic retinopathy (ORs per SD 1·28–1·57), diabetic nephropathy (1·25–1·58), proliferative diabetic retinopathy (1·39–2·08), and end-stage diabetic nephropathy (1·44–1·87); PRS was associated with coronary artery disease in the Admixed American ancestry group only (1·16 [95% CI 1·08–1·25]).

Interpretation These validated, publicly available PRSs can improve risk stratification for type 2 diabetes onset and complications across diverse ancestries, supporting their further evaluation in clinical settings.

Funding The National Human Genome Research Institute of the US National Institutes of Health.

Lancet Diabetes Endocrinol 2026

Published **Online** April 27, 2026 https://doi.org/10.1016/S2213-8587(25)00405-X

*Contributed equally

†Contributed equally

‡Members listed in appendix 1 (p 39)

§Members listed in appendix 1 (pp 40–41)

¶Members listed in appendix 1 (pp 42–43)

|| Jointly directed this work

**Members listed in appendix 1 (pp 17–21)

Programs in Metabolism and Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA (A Huerta-Chagoya PhD, R Mandla BA, K Taylor MS, A Leong MD, A K Manning PhD, J M Mercader PhD); **Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA** (A Huerta-Chagoya, R Mandla, A Leong, J M Mercader); **Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA** (A Huerta-Chagoya, R Mandla, A Leong, J M Mercader); **Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA** (J Kim PhD, Y Lu PhD, L E Petty PhD, J E Below PhD, M C Y Ng PhD); **Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA** (R Mandla, B F Voight PhD); **Department of Statistical Genetics, Osaka University, Osaka, Japan** (K Suzuki PhD); **Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo,**

Japan (K Suzuki)**; Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Division of Musculoskeletal and Dermatological Sciences, University of Manchester, Manchester, UK** (K Suzuki, A P Morris PhD)**; Lee Kong Chian School of Medicine, Clinical Sciences Building, Nanyang Technological University, Singapore** (H K Ng PhD, M Loh PhD)**; Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, South Korea** (J Choi MS)**; The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA** (S Lee MD, R J F Loos MD, D Sanghera PhD)**; Department of Pediatrics, Section of Genetics, University of Oklahoma, Oklahoma City, OK, USA** (M Rout PhD)**; Nuffield Department of Population Health, University of Oxford, Oxford, UK** (K Lin PhD, R G Walters PhD)**; Unidad de Investigación de Enfermedades Metabólicas, Research Direction of the Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico** (C A Aguilar-Salinas PhD)**; Tecnológico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Mexico City, Mexico** (C A Aguilar-Salinas)**; Center for Infectious Disease Research, National Institute of Public Health, Cuernavaca, Mexico** (L García-García MD)**; Centro de Estudios en Diabetes, Unidad de Investigacion en Diabetes y Riesgo Cardiovascular, Centro de Investigacion en Salud Poblacional, Instituto Nacional de Salud Pública, Mexico City, Mexico** (C González-Villalpando MD)**; Department of Population and Public Health Sciences, Keck School of Medicine of USC, Los Angeles, CA, USA** (C A Haiman PhD)**; Department of Precision Medicine, Division of Genome Science, National Institute of Health, Cheongju-si, South Korea** (Y J Kim PhD)**; Department of Internal Medicine, Seoul National University College of Medicine and Seoul National University Hospital, Seoul, South Korea** (S H Kwak PhD)**; Department of**

Research in context

Evidence before this study

Dozens of genome-wide association studies (GWAS) have been conducted for type 2 diabetes, generating vast amounts of genetic data used to construct polygenic risk scores (PRSs). These scores aggregate thousands of genetic variants to estimate an individual's genetic predisposition to type 2 diabetes. However, more than 70% of the data underpinning these risk scores originate from populations of primarily European ancestry. This ancestral bias fundamentally restricts the wide clinical applicability of PRSs, as their predictive performance decreases in non-European populations, including those of African or African American, South Asian, and Admixed American descent, in whom the burden of type 2 diabetes is often disproportionately high. Before this work, there was a deficit of robustly developed and extensively validated PRSs for type 2 diabetes that perform consistently across diverse ancestries. We searched the PGS Catalog for English-language publications, published between Oct 15, 2019, and Oct 7, 2024, reporting "type 2 diabetes" PRSs. Excluding process-specific PRSs or PRSs trained with data from the All of Us Research Program, we identified 25 papers detailing 55 PRSs published by our methodological cutoff to benchmark against our PRS models. Of these 25 studies, 12 relied entirely on GWAS from populations of European ancestry. Although 18 studies evaluated the PRSs across multiple ancestries, ten of these cross-ancestry studies still derived the PRSs from European GWAS. Overall, the sample sizes of the GWAS reached up to 1·1 million individuals, yet ancestry diversity was poor.

Added value of this study

This study leveraged the wealth of accumulated genetic data on individuals with type 2 diabetes to construct, train, and validate what are, to our knowledge, the most comprehensive PRSs applicable to five major global ancestry groups (African or African American, Admixed American, East Asian, European, and South Asian). Our multi-ancestry PRSs had better predictive ability for type 2 diabetes risk than our single-ancestry PRSs as well as other previously available PRSs, and could also predict an individual's risk of developing common microvascular complications and provide additional risk information beyond what traditional clinical factors alone can offer.

Implications of all the available evidence

PRSs are powerful tools that can significantly improve risk prediction when used alongside traditional clinical factors. However, to be clinically useful and equitable, they must work reliably for everyone. This study developed and rigorously tested PRSs to accurately identify individuals at high risk of type 2 diabetes and its complications from five major ancestry groups worldwide. By making these scores freely available, we provide researchers and clinicians with a resource that could help to advance genetic risk stratification and make type 2 diabetes prevention strategies more precise and globally fair.

Introduction

Type 2 diabetes is a major health problem. Approximately 589 million people are living with the condition globally, with numbers predicted to increase to 853 million by 2050.[1] Both genetic and environmental factors contribute to susceptibility to type 2 diabetes. Genome-wide association studies (GWAS) have identified 1289 genetic signals for the condition in individuals from diverse ancestries.[2,3] The aggregation of risk alleles in polygenic risk scores (PRSs) for type 2 diabetes[2,4] could provide insights into disease progression and prognosis, and identify people at risk of developing the condition for prioritisation of therapeutic or lifestyle interventions.[5] Several initiatives have tested the use of PRSs in clinical settings,[6,7] and have shown that PRSs for type 2 diabetes are particularly valuable for identifying individuals at risk among those perceived to be at low risk on the basis of standard clinical risk factors.[8]

Type 2 diabetes and related complications disproportionately affect populations of South Asian, African American, and Hispanic or Latin American ancestries compared with those of European ancestry.[9] However, most PRSs are based on GWAS comprising predominantly individuals of European ancestry and have poor performance in individuals of other ancestries, which could further exacerbate health disparities if PRSs were used for disease prediction in diverse populations.[2,10]

Research efforts over the past 7 years have aimed to improve the transferability of PRSs across diverse ancestries. First, the Polygenic Risk Methods Development (PRIMED) consortium and others have developed new methods, such as those incorporating GWAS and linkage disequilibrium data from multiple ancestries, showing improved prediction.[11,12] Second, data from large-scale GWAS of type 2 diabetes in non-European populations have become available, which could improve the power of PRSs in these populations. Third, the emergence of large-scale biobank data, such as from the All of Us Research Program, provides independent samples from diverse ancestries that could be used to develop, train, and validate novel PRSs.[13,14]

As part of the Type 2 Diabetes Global Genomics Initiative (T2DGGI), we published the largest meta-analysis of multi-ancestry GWAS of type 2 diabetes to date, including data from 2·5 million individuals.[3] Here we present results from Diabetes Polygenic Risk Scores in Multiple ancestries (D-PRISM),[2,12] an international consortium to develop PRSs for different types of diabetes and disease progression across the lifespan in diverse ancestries. We aimed to leverage the T2DGGI and the D-PRISM consortia to aggregate data from type 2 diabetes GWAS from five major ancestry groups, and to

train and validate PRS models in independent cohorts to evaluate the performance of these scores in predicting type 2 diabetes, its complications, and comorbidities using a unified pipeline. We tested the optimal PRS models for each ancestry and have made them available to the community for future testing of their clinical utility.

Methods

Study design and participants

The overall study design is detailed in appendix 1 (pp 1–6). We used type 2 diabetes GWAS data from selected cohorts participating in three large consortia: Diabetes Meta-analysis of Trans-ethnic Association Studies,[15] Million Veteran Program,[16] and FinnGen.[17] Cohorts were categorised by genetic similarity to the 1000 Genomes project[18] or to the predominant major global ancestry group according to the country of origin: African or African American, Admixed American, East Asian, European, and South Asian. We leveraged nearly all available genetic datasets with type 2 diabetes phenotype information for the following purposes: to maximise diversity in ancestries, harmonising 125 type 2 diabetes GWAS datasets including up to 2·2 million individuals (appendix 2 tabs 1, 2); to enhance the representation and tagging of the genetic variants contributing to the PRSs by generating new, ancestry-specific linkage disequilibrium reference panels (appendix 2 tab 7); to train PRS models across ancestries (appendix 2 tabs 3, 4); and to thoroughly validate the PRSs in multiple independent, harmonised cohorts (appendix 2 tabs 5, 6) and to assess the association of these PRSs with diabetes-related complications and comorbidities (appendix 2 tabs 13, 14). All studies were approved by local institutional review boards and/or ethics committees (appendix 1 pp 33–38).

Procedures

We used summary statistics from ancestry-specific type 2 diabetes GWAS to construct single-ancestry PRSs, using the continuous-shrinkage PRS-CS[19] method, and multi-ancestry PRSs, using the PRS-CSx[20] method. The PRS-CSx approach uses the increased statistical power from GWAS of five major ancestry groups, which are jointly modelled to maximise the power of variants present in all ancestries as well as incorporating ancestry-specific effects. Ancestry-specific linkage disequilibrium panels were used to model pairwise correlations between single-nucleotide polymorphisms (SNPs) in GWAS during PRS construction. The original linkage disequilibrium panels are based on SNPs derived from the International HapMap Project phase 3 data (HapMap3), which do not tag well in non-European populations. In addition, the original linkage disequilibrium panels used samples from the 1000 Genomes project, which has a limited number of samples, or the UK Biobank, which has only a small amount of data from individuals with non-European ancestry. To improve tagging, we developed new linkage disequilibrium panels with an expanded set of SNPs generated using the Tag(ging) It(erative) of SNPs in multiple populations (TagIt) program[21] for variants with ancestry-specific minor allele frequencies of at least 0·01 in the 1000 Genomes project. We also used more than 8000 in-house samples to compute ancestry-specific pairwise linkage disequilibrium, which enables more accurate linkage disequilibrium modelling (appendix 2 tab 7).

Statistical analysis

We compared the effect of ancestry in GWAS, SNP sets, and linkage disequilibrium sources on PRS performance in the training cohorts. We defined the best models as those that maximised the incremental area under the curve (iAUC) for predicting prevalent type 2 diabetes (appendix 2 tab 8). To calculate the iAUC, we subtracted the AUC of a model without the PRS from the AUC of a full model that included the PRS, sex, age, and genetic principal components. We validated the best-performing models in at least four independent cohorts per ancestry and fitted secondary models with BMI as an additional covariate (appendix 2 tabs 9, 10). We estimated the effect size of PRSs as the odds ratio (OR) per SD of the PRS, and calculated ORs for individuals at the 90th, 95th, and 97·5th PRS percentiles compared with the IQR reference. We used the DeLong test to compare the iAUCs of single-ancestry and multi-ancestry PRSs. We combined the PRS estimates across validation cohorts using fixed-effects meta-analyses by ancestry, weighting each cohort's β coefficient by its inverse variance via the metafor package in R.

In the All of Us version 7 cohort, we compared our best-performing type 2 diabetes PRSs with 55 scores from the Polygenic Score (PGS) Catalog. We defined statistical significance using a Bonferroni-corrected threshold of $p<0·0009$ (0·05 / 55) to account for multiple testing. We also compared different strategies for PRS construction. Although some studies leverage ancestry diversity using inverse-variance-weighted meta-analysis from multi-ancestry GWAS—typically using a single European linkage disequilibrium panel—this approach might not accurately model linkage disequilibrium patterns or tag ancestry-specific variants. To evaluate the PRS-CSx approach against standard meta-analysis, we conducted an inverse-variance-weighted meta-analysis of D-PRISM ancestry-specific GWAS summary statistics and applied PRS-CS using the European 1000 Genomes Project HapMap3 linkage disequilibrium reference panel. We also evaluated PRSs based on the largest meta-analysis of multi-ancestry GWAS of type 2 diabetes, conducted by Suzuki and colleagues,[3] which included a sample size 16% larger than our datasets, as we intentionally held out several cohorts for PRS training and validation. For these PRSs based on data from Suzuki and colleagues, we either applied PRS-CS or

Medicine, Massachusetts General Hospital, Boston, MA, USA (A Leong); **Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA** (A Leong); **Endocrine Division, Massachusetts General Hospital, Boston, MA USA** (A Leong); **The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA** (R J F Loos); **Aging Research Center, Cinvestav Sede Sur, Center for Research and Advanced Studies of the National Polytechnic Institute, Mexico City, Mexico** (A Moreno-Estrada PhD); **NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK** (A P Morris); **Instituto Nacional de Medicina Genómica, Mexico City, Mexico** (L Orozco PhD); **The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA** (J I Rotter MD); **Unidad de Biología Molecular y Medicina Genómica, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México/Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico** (T Tusie-Luna PhD); **Corporal Michael J Crescenz Philadelphia VA Medical Center, Philadelphia, PA, USA** (B F Voight, M Vujkovic PhD); **Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA** (B F Voight); **Institute for Translational Medicine and Therapeutics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA** (B F Voight); **Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA** (M Vujkovic); **Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA** (T Ge PhD); **Clinical and**

Translational Epidemiology Unit, Mongan Institute, Massachusetts General Hospital, Boston, MA, USA (A K Manning); **Department of Medicine, Harvard Medical School, Boston, MA, USA** (A K Manning, J M Mercader); **Department of Epidemiology and Biostatistics, Imperial College London, London, UK** (M Loh); **Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore** (M Loh); **Saw Swee Hock School of Public Health,**

constructed a restricted-to-significance PRS using 1289 distinct type 2 diabetes variants (appendix 2 tab 11).

Type 2 diabetes adversely affects the functions of multiple organs, leading to long-term complications and comorbidities.[23] We used data from the All of Us version 8 cohort to assess the utility of the best-performing multi-ancestry PRSs in identifying individuals with type 2 diabetes who are at risk of developing microvascular complications (diabetic retinopathy, proliferative diabetic retinopathy, diabetic nephropathy, and end-stage diabetic nephropathy) and comorbidities (coronary artery disease and ischaemic stroke). We fitted logistic regression models adjusted for sex, age, and genetic principal components, focusing our subsequent analyses on African or African American, Admixed American, and European ancestries, for which sample sizes in the All of Us cohort were sufficient. To determine if the duration of type 2 diabetes mediated these effects, we conducted two sensitivity analyses: first, we restricted the sample to patients in whom the onset of type 2 diabetes preceded the first record of the complication; second, we adjusted for the estimated duration of type 2 diabetes (appendix 2 tabs 12, 13).

Finally, to better assess the clinical impact of the PRSs, we extracted electronic health record data to implement a comprehensive clinical risk score based on the type 2 diabetes Framingham Risk Score. This score comprised nine components: age, sex, parental history of type 2 diabetes, BMI, systolic blood pressure, and high-density lipoprotein, total cholesterol, triglycerides, and random glucose concentrations. We conducted Kaplan–Meier analysis to estimate diabetes-free survival across PRS tertiles and used Cox proportional hazards models to quantify the predictive value of the PRSs for incident type 2 diabetes, evaluating the score both as a stand-alone predictor and in combination with the nine-component clinical risk score. To evaluate if the PRSs provided predictive value independent of glucose concentration, we fitted Cox models using the lowest

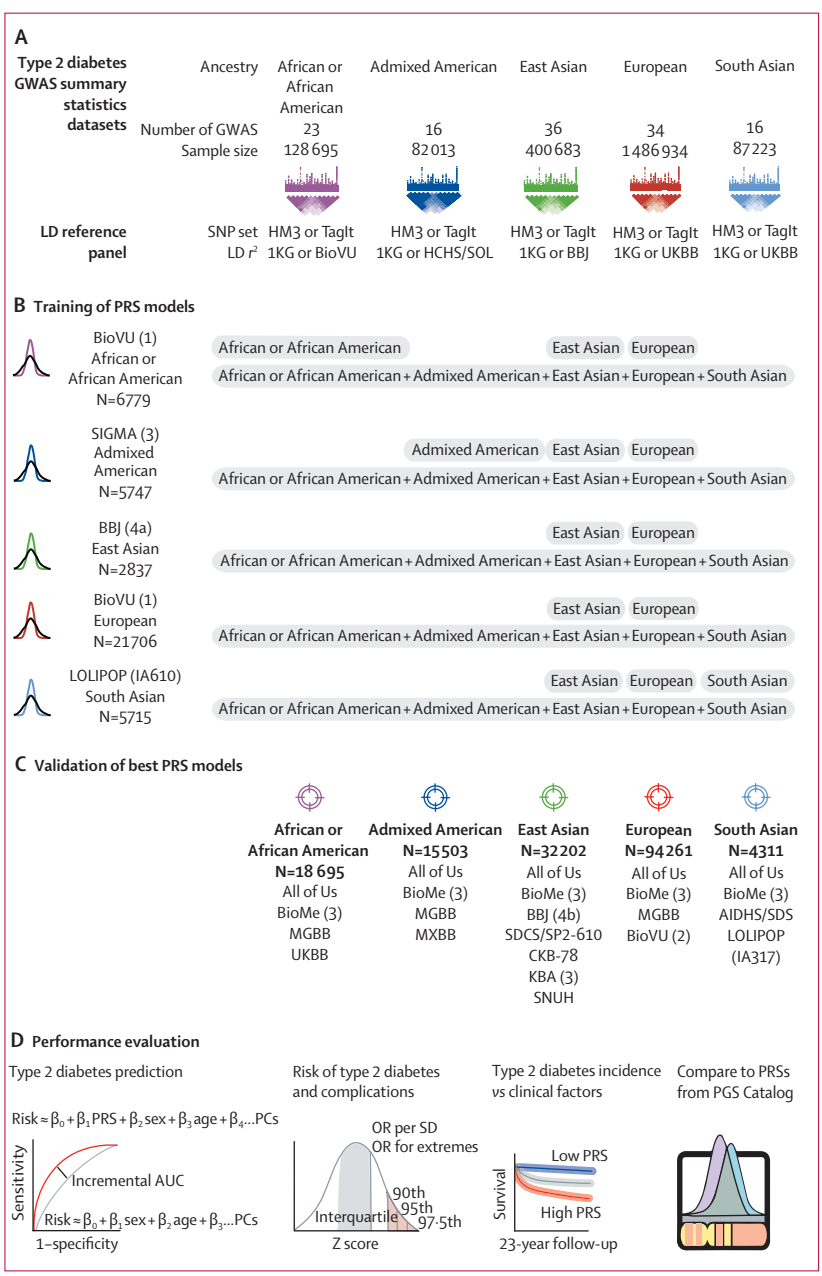
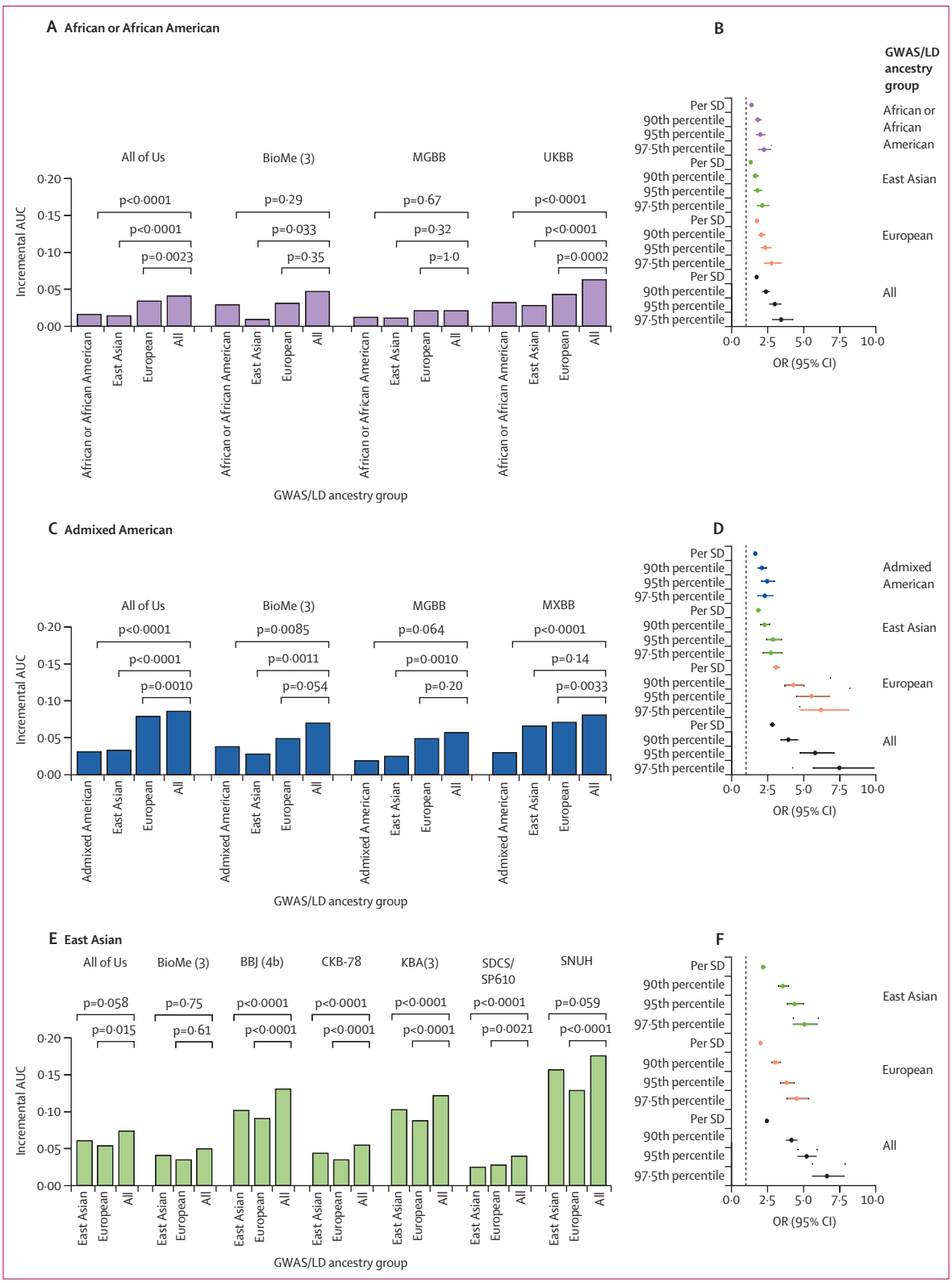


Figure 1: **Overall analysis approach**
(A) Overview of the 125 type 2 diabetes GWAS datasets, SNP sets, and pairwise LD information used to generate five ancestry-specific type 2 diabetes GWAS meta-analyses and 20 LD reference panels for PRSs training (appendix 1 p 1). (B) Independent, ancestry-specific cohorts used to train the PRS models and select the optimal continuous shrinkage prior from five ϕ values (0·01, 0·001, 1×10^{-4}, 1×10^{-5}, and 1×10^{-6}) based on predictive performance. Single-ancestry PRSs used GWAS summary statistics and LD panels matched to the validation ancestry or used data from East Asian or European ancestry datasets. Multi-ancestry PRSs jointly modelled GWAS summary statistics and LD panels from all five ancestry groups. (C) Set of 23 ancestry-specific and independent cohorts used to validate the 18 best-performing PRSs. (D) Evaluation of PRS predictive performance: incremental AUC, calculated as the difference between the AUC of the full model (risk ≈ β_0 + β_1 PRS + β_2 sex + β_3 age + β_4 ... PCs) and the model without the PRS (risk ≈ β_0 + β_1 sex + β_2 age + β_3 ... PCs), and the proportion of variation in type 2 diabetes status explained by the PRS, estimated using the difference between the Nagelkerke's r^2 of the full model (risk ≈ β_0 + β_1 PRS + β_2 sex + β_3 age + β_4 ... PCs) and the model without the PRS (risk ≈ β_0 + β_1 sex + β_2 age + β_3 ... PCs); the OR per SD of the PRS distribution or OR comparing PRS distribution extremes relative to the IQR; diabetes-free survival probabilities over 23 years of follow-up, estimated using Kaplan–Meier curves stratified by PRS tertile; and a comparative benchmarking of the best-performing multi-ancestry PRS against 55 existing type 2 diabetes PRSs from the PGS Catalog (appendix 1 pp 1–4). 1KG=1000 Genomes. AIDHS/SDS=Asian Indian Diabetic Heart Study/Sikh Diabetes Study. AUC=area under the curve. BBJ=BioBank Japan. BioMe=BioMe Biobank. BioVU=Vanderbilt University Medical Center. CKB-78=China Kadoorie Biobank. GWAS=genome-wide association studies. HCHS/SOL=Hispanic Community Health Study/Study of Latinos. HM3=HapMap3. KBA=Korea Biobank Array. LD=linkage disequilibrium. LOLIPOP=London Life Sciences Prospective Population. MGBB=Mass General Brigham Biobank. MXBB=Mexican Biobank. OR=odds ratio. PCs=principal components. PGS=polygenic score. PRS=polygenic risk score. SDCS/SP2-610=Singapore Diabetic Cohort Study and Singapore Prospective Study Program. SIGMA=Slim Initiative for Genomic Medicine in the Americas. SNP=single-nucleotide polymorphism. SNUH=Seoul National University Hospital. TagIt=Tag(ging) It(erative) of SNPs in multiple populations. UKBB=UK Biobank. Numbers in parentheses represent different subset of data used in a cohort.



National University of Singapore and National University Health System, Singapore (X Sim PhD)**; Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, MA, USA** (J M Mercader)

Correspondence to:
Josep M Mercader, Diabetes Unit, Massachusetts General Hospital, Boston, MA 02114, USA
jmercader@mgb.org

or

Maggie C Y Ng, Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA
maggie.ng@vumc.org

See Online for appendices 1 and 2

(Figure 2 continues on next page)

PRS tertile and random glucose concentration of less than 140 mg/dL as the reference group (appendix 2 tabs 14, 15). We conducted all statistical analyses using R version 4.3.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Compared with those based on HapMap3, the new linkage disequilibrium panels generated using TagIt improved the overall SNP coverage of variants from the 1000 Genomes Project by up to two times. Coverage increased by seven times for SNPs with minor allele frequencies of 0·01–0·05 in individuals with African or African American ancestry (appendix 1 pp 6, 7, appendix 2 tab 7). In all ancestry groups, the best-performing PRSs were those constructed using the expanded TagIt set of variants and/or the recomputed in-house pairwise linkage disequilibrium with large sample sizes (appendix 1 p 7, appendix 2 tab 8), suggesting that higher SNP coverage and better linkage disequilibrium modelling improve prediction.

In the ancestry-specific meta-analyses, we included 2 185 548 individuals (359 819 with type 2 diabetes and 1 825 729 controls) from 125 type 2 diabetes GWAS datasets (34 European, 36 East Asian, 23 African or African American, 16 Admixed American, and 16 South Asian; figure 1A, appendix 2 tabs 1, 2). To train PRSs (ie, tuning PRS construction parameters) we used

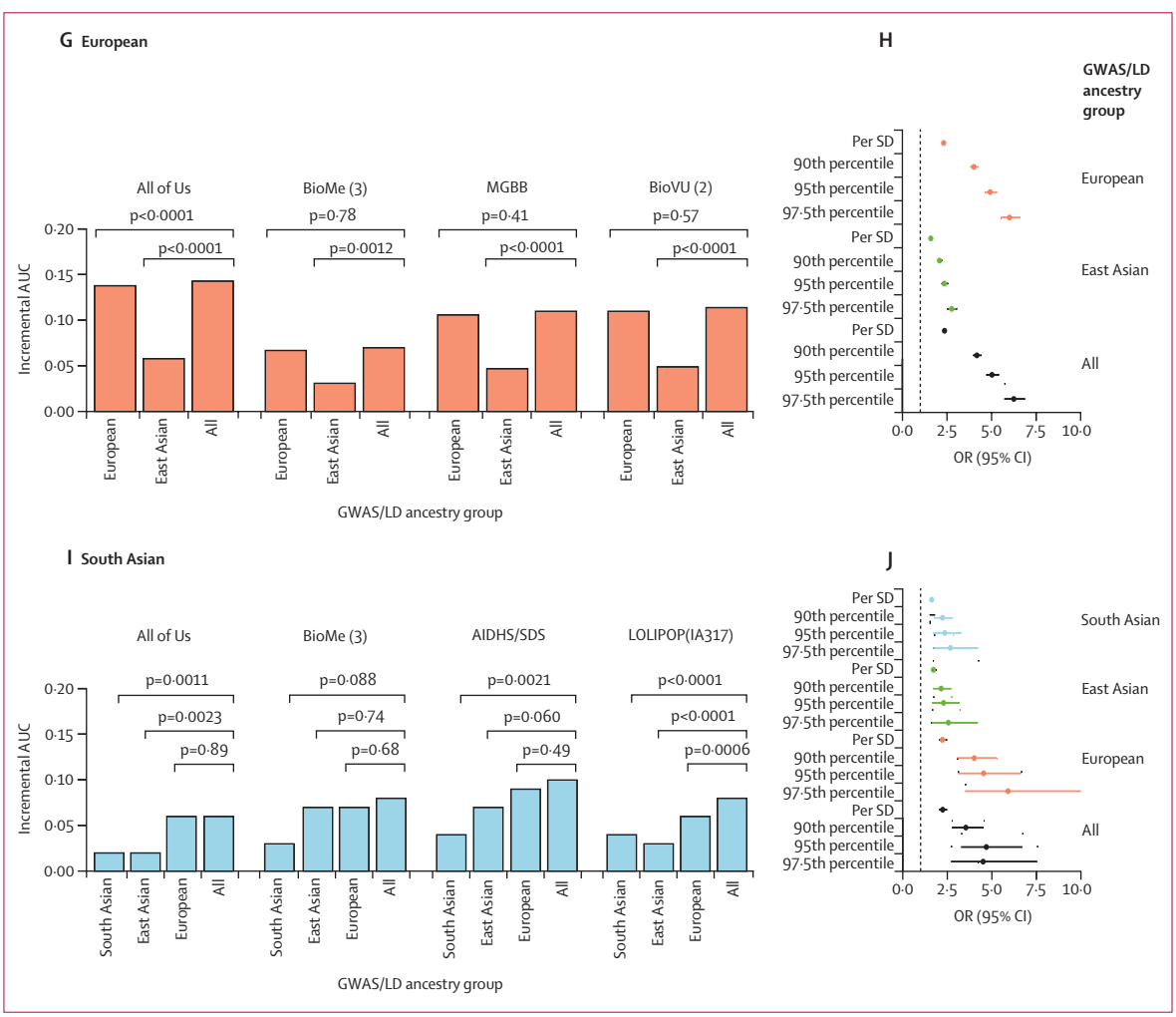


Figure 2: **Performance of the type 2 diabetes PRSs in the validation cohorts across ancestry groups**

Incremental AUC of the type 2 diabetes PRSs in the validation cohorts across ancestry groups: African or African American (A), Admixed American (C), East Asian (E), European (G), and South Asian (I). For each ancestry, the best-performing single-ancestry and multi-ancestry PRSs were evaluated. Each bar represents a single cohort. Odds ratio from the meta-analysis of validation cohorts across ancestry groups: African or African American (B), Admixed American (D), East Asian (F), European (H), and South Asian (J). Points represent the OR per SD of the PRS distribution or the OR comparing different PRS distribution extremes (90th, 95th, and 97·5th percentiles) relative to the IQR. Error bars show the 95% CIs. The dashed lines indicate an OR of 1. All models were adjusted for sex, age, and genetic principal components. AUC=area under the curve. GWAS=genome-wide association studies. LD=linkage disequilibrium. OR=odds ratio. PRS=polygenic risk score.

one independent cohort per ancestry, aggregating data from 42 784 individuals (10 992 with type 2 diabetes and 31 792 controls; figure 1B, appendix 2 tabs 3, 4), and for validation (ie, testing PRS performance) we used at least four cohorts per ancestry, including data from 164 972 individuals (39 148 with type 2 diabetes and 125 824 controls; figure 1C, appendix 2 tabs 5, 6).

We first trained PRSs (appendix 2 tabs 3, 4) using GWAS summary statistics and four linkage disequilibrium panels, each matched to the ancestry of validation cohorts (appendix 2 tabs 5, 6). The performance of the matched single-ancestry PRS was positively correlated with the sample sizes of the GWAS. The best predictive performance was observed in European (incremental AUC 0·07–0·14) and East Asian (0·02–0·16) ancestries, whereas prediction was poorer for African or African American (0·02–0·03), Admixed American (0·02–0·04), and South Asian (0·02–0·04; figure 2, appendix 2 tab 9) ancestries.

We also observed larger effect sizes in European (OR per SD 2·31 [95% CI 2·26–2·35]) and East Asian (2·19 [2·12–2·27]) validation cohorts than in African or African American (1·38 [1·34–1·43]), Admixed American (1·64 [1·57–1·71]), and South Asian (1·61 [1·50–1·74]) cohorts. The PRSs for European and East Asian ancestries also had better power for the identification of individuals at very high risk of developing type 2 diabetes. Individuals at the 90th percentile of the PRS distribution had an approximately 3·5–4 times increased risk of type 2 diabetes compared with those in the IQR in European (OR 4·01 [95% CI 3·82–4·21]) and East Asian (3·58 [3·24–3·94]) ancestries, compared with an approximately two times increased risk in African or African American (1·82 [1·64–2·01]), Admixed American (2·10 [1·84–2·39]), and South Asian (2·22 [1·75–2·81]) ancestries (figure 2, appendix 2 tab 9). We observed similar results regardless of adjustment for BMI (appendix 1 p 8, appendix 2 tab 10).

We tested whether constructing PRSs using GWAS summary statistics from ancestries with larger sample sizes, such as European and East Asian, could improve the prediction of type 2 diabetes risk in ancestries with smaller GWAS sample sizes, despite ancestral differences between the discovery and validation cohorts. Compared with the matched single-ancestry PRSs, the use of European GWAS improved type 2 diabetes risk prediction in African or African American, Admixed American, and South Asian validation cohorts, but gave poorer predictions in East Asian validation cohorts. The PRS based on East Asian GWAS showed a small improvement in prediction of type 2 diabetes risk in Admixed American and South Asian validation cohorts, but poorer performance in African or African American and European validation cohorts (figure 2, appendix 2 tab 9). The effect sizes of the PRSs were consistent with their prediction performance. The PRS using European GWAS was the best-performing single-ancestry PRS for African or African American (OR per SD 1·75 [95% CI 1·67–1·82]), Admixed American (3·07 [2·89–3·27]), South Asian (2·21 [2·03–2·42]), and European (2·31 [2·26–2·35]) cohorts, whereas the PRS using East Asian GWAS had the best performance in East Asian cohorts (2·19 [2·12–2·27]; figure 2, appendix 2 tab 9).

We further applied PRS-CSx[20] to develop multi-ancestry PRSs. Compared with the best single-ancestry PRSs, multi-ancestry PRSs showed higher prediction performance across validation cohorts from all ancestries (AUC 0·63–0·75 and incremental AUC 0·02–0·06 in African or African American, 0·71–0·77 and 0·06–0·09 in Admixed American, 0·72–0·86 and 0·04–0·17 in East Asian, 0·69–0·76 and 0·07–0·14 in European, and 0·67–0·80 and 0·06–0·10 in South Asian cohorts; figure 2, appendix 2 tab 9).

The multi-ancestry PRSs also had higher effect sizes and smaller CIs than the single-ancestry PRSs across all ancestries (OR per SD 1·73 [95% CI 1·67–1·80] in African or African American, 2·82 [2·67–2·97] in Admixed American, 2·45 [2·36–2·54] in East Asian, 2·36 [2·32–2·41] in European, and 2·23 [2·05–2·42] in South Asian ancestries). The improvement was particularly notable for individuals at the extremes of PRS distributions. Individuals in the 97·5th percentile of the multi-ancestry PRSs had a 3–7 times increased risk of type 2 diabetes than those in the IQR (OR 3·43 [95% CI 2·8–4·21] in African or African American, 7·47 [5·64–9·89] in Admixed American, 6·62 [5·58–7·85] in East Asian, 6·25 [5·72–6·82] in European, and 4·50 [2·70–7·53] in South Asian ancestry groups; figure 2, appendix 2 tab 9).

We leveraged data from the All of Us Research Program to compare our best-performing multi-ancestry PRSs against the published type 2 diabetes PRSs from the PGS Catalog.[22] We tested 55 of 147 available PRSs for type 2 diabetes (accessed Oct 7, 2024), after excluding pathway-specific PRSs or those including data from the All of Us cohort in PRS development. The multi-ancestry PRSs showed significantly better predictive performance than those reported previously for the African or African American, Admixed American, and European ancestry groups (incremental AUC for our multi-ancestry PRSs *vs* best incremental AUC from the PGS Catalog: 0·041 *vs* 0·029 in African or African American, 0·086 *vs* 0·073 in Admixed American, and 0·143 *vs* 0·123 in European ancestries; DeLong p<0·0009, Bonferroni correction for 55 tests) but not for the East Asian (0·074 *vs* 0·062) and South Asian (0·061 *vs* 0·068) ancestry groups (DeLong p>0·0009; appendix 1 pp 9–14, appendix 2 tab 11).

To assess the utility of the PRS-CSx method, we constructed a PRS-CS model using summary statistics from the inverse variance-weighted GWAS meta-analysis of our ancestry-specific GWAS summary statistics and tested its performance in the All of Us cohort. The performance of the multi-ancestry PRS-CSx

models was significantly higher than that of the PRS-CS models in the prediction of type 2 diabetes risk for Admixed American, East Asian, and European ancestry groups (incremental AUC for multi-ancestry PRS-CSx *vs* multi-ancestry PRS-CS: 0·086 *vs* 0·077 in Admixed American, 0·074 *vs* 0·061 in East Asian, and 0·143 *vs* 0·126 in European ancestries; DeLong p<0·05) but not for African or African American (0·041 *vs* 0·038) and South Asian (0·061 *vs* 0·060) ancestry groups (DeLong p>0·05) (appendix 1 pp 9–14, appendix 2 tab 11).

We also showed that, despite being based on a smaller sample of GWAS, our multi-ancestry PRS-CSx model showed higher predictive performance than the PRS-CS model for the dataset from Suzuki and colleagues in the Admixed American, European, and East Asian ancestry groups (incremental AUC for our multi-ancestry PRS-CSx *vs* Suzuki and colleagues PRS-CS: 0·086 *vs* 0·077 in Admixed American, 0·074 *vs* 0·059 in East Asian, and 0·143 *vs* 0·128 in European, DeLong p<0·05) and similar predictive performance in the African or African American (0·041 *vs* 0·041) and South Asian (0·061 *vs* 0·068) ancestry groups (DeLong p>0·05; appendix 1 pp 9–14).

Additionally, our multi-ancestry PRS-CSx models showed consistently improved prediction performance in Admixed American (0·086 *vs* 0·054), East Asian (0·074 *vs* 0·042), European (0·143 *vs* 0·090), and South Asian (0·068 *vs* 0·034) ancestry groups (DeLong p<0·05), and similar performance in the African or African American ancestry group (0·041 *vs* 0·038; DeLong p>0·05), when compared with a multi-ancestry PRS that was restricted to the 1289 genome-wide significant variants identified by Suzuki and colleagues (appendix 1 pp 9–14, appendix 2 tab 11).[3]

In individuals with type 2 diabetes, our multi-ancestry PRS-CSx were significantly associated with increased risk of microvascular complications across the three ancestries studied (African or African American, Admixed American, and European). Effect sizes (OR per SD) ranged from 1·28 (95% CI 1·18–1·39; African or African American) to 1·57 (1·43–1·73; Admixed American) for diabetic retinopathy and from 1·25 (1·18–1·32; European) to 1·58 (1·41–1·78; Admixed American) for diabetic nephropathy, and were higher for the more severe forms, proliferative diabetic retinopathy (1·39 [1·17–1·65; African or African American] to 2·08 [1·73–2·51; Admixed American]) and end-stage diabetic nephropathy (1·44 [1·20–1·74; African or African American] to 1·87 [1·56–2·22; Admixed American]; p<0·008, Bonferroni correction for six outcomes). Conversely, the PRS was associated with coronary artery disease in the Admixed American ancestry group only (OR 1·16 [95% CI 1·08–1·25]; figure 3, appendix 2 tabs 12, 13). Additional sensitivity analyses revealed an association between PRS and earlier age of onset of type 2 diabetes: a 1 SD increase in the PRS was associated with an earlier onset of type 2 diabetes of up to 2·3 years across the three genetic ancestries (appendix 1 p 16). Adjusting for the duration of type 2 diabetes, or restricting to individuals in which the onset of diabetes preceded the onset of complications on the basis of electronic health record data, yielded consistent results (appendix 1 pp 3–4, 15–16, appendix 2 tab 13).

Over a median follow-up of 7·4 years (IQR 3·91–9·31; range 0·05–21·96), 2074 of 17 920 individuals developed incident type 2 diabetes and 15 846 remained disease-free (appendix 1 p 4, appendix 2 tabs 14–15). Kaplan–Meier analysis by PRS tertile showed significant, stepwise associations with disease-free survival across ancestries (figure 4; log-rank p<0·0001 for all). Compared with the lowest tertile (T1), the highest tertile (T3) had higher hazard ratios (HR 1·76 [95% CI 1·39–2·23] in African or African American, 3·87 [2·63–5·68] in Admixed American, and 5·26 [4·39–6·29] in European ancestries; appendix 2 tab 14). The PRSs provided predictive value independent of random glucose concentration, especially for individuals with high concentrations (≥140 mg/dL;

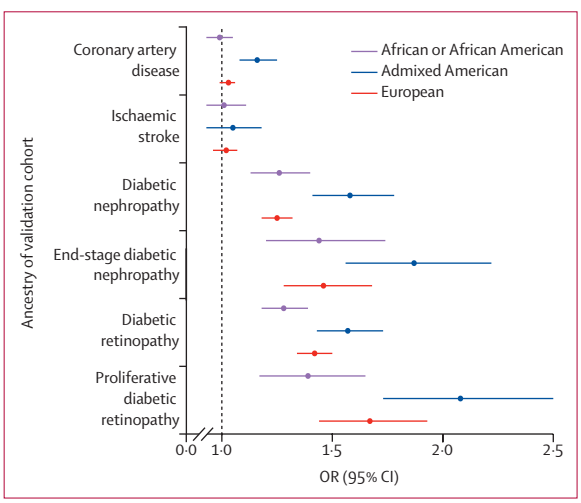


Figure 3: **Association of our multi-ancestry PRS with common complications and comorbidities of diabetes in the All of Us cohort**
ORs per SD increase for six outcomes in African or African American, Admixed American, and European cohorts: coronary artery disease, ischaemic stroke, diabetic nephropathy, end-stage diabetic nephropathy, diabetic retinopathy, and proliferative diabetic retinopathy. All analyses were restricted to individuals with type 2 diabetes. Error bars indicate 95% CIs. All models were adjusted for sex, age, and genetic principal components. OR=odds ratio. PRS=polygenic risk score.

Figure 4: **Kaplan–Meier curves by PRS tertile, stratified by ancestry and prediabetes status**
Diabetes-free survival analyses, without (left) and with (middle and right) stratification by prediabetes (random glucose concentration ≥140 mg/dL) status, by genetic ancestry group: African or African American (A), Admixed American (B), and European (C). Diabetes-free survival is stratified by PRS tertile (coloured lines), whereby increased colour intensity corresponds to higher genetic risk; 95% CIs are shaded. p values were derived from the log-rank test to assess differences in survival distributions across PRS tertiles. PRS=polygenic risk score. T1=tertile 1 (lowest tertile). T2=tertile 2 (middle tertile). T3=tertile 3 (highest tertile).

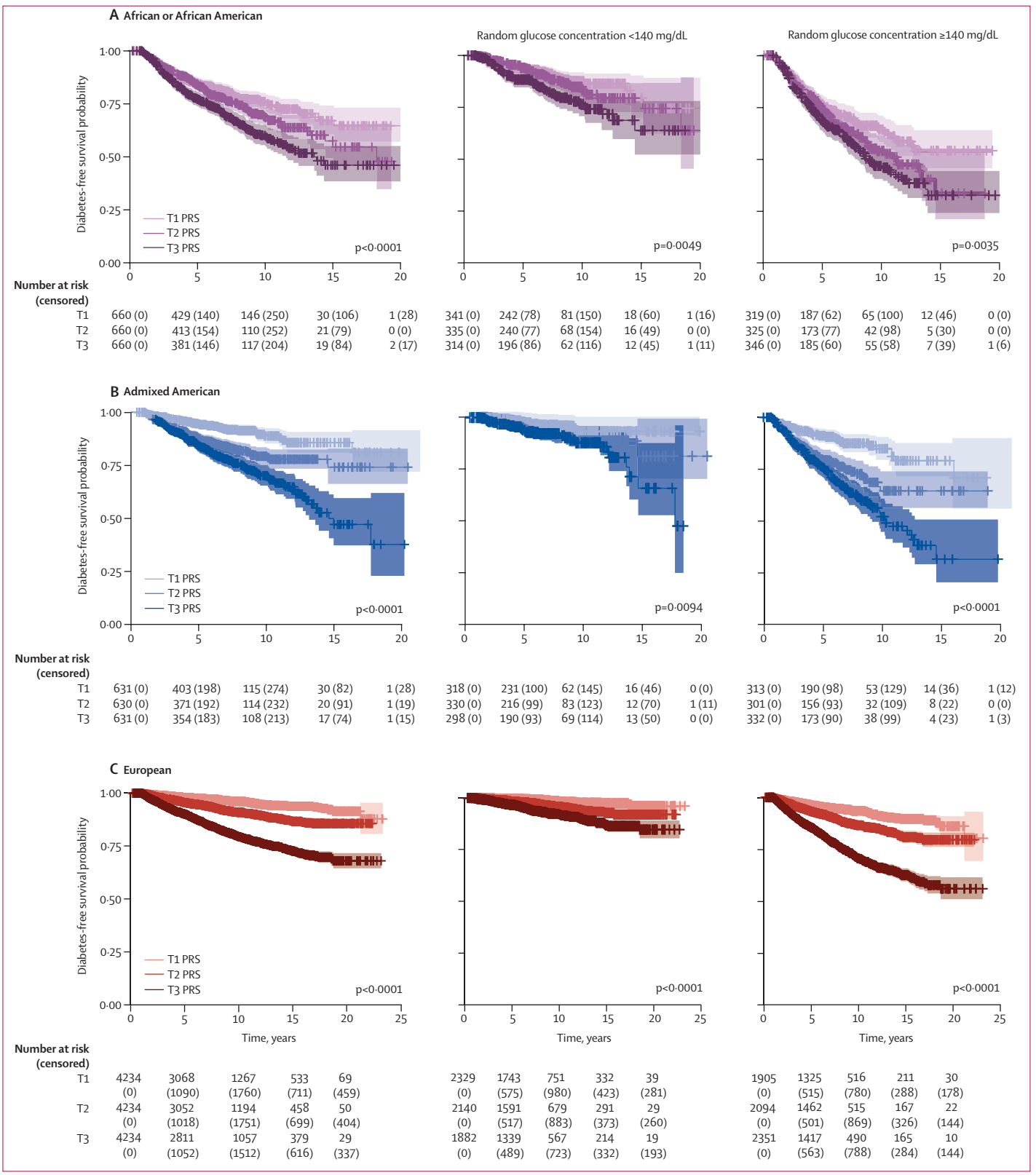

A African or African American
Random glucose concentration <140 mg/dL
Random glucose concentration ≥140 mg/dL
Diabetes-free survival probability
1·00
0·75
0·50
0·25
0·00
T1 PRS
T2 PRS
T3 PRS
p<0·0001
p=0·0049
p=0·0035
0
5
10
15
20
Number at risk (censored)
T1 660 (0) 429 (140) 146 (250) 30 (106) 1 (28) 341 (0) 242 (78) 81 (150) 18 (60) 1 (16) 319 (0) 187 (62) 65 (100) 12 (46) 0 (0)
T2 660 (0) 413 (154) 110 (252) 21 (79) 0 (0) 335 (0) 240 (77) 68 (154) 16 (49) 0 (0) 325 (0) 173 (77) 42 (98) 5 (30) 0 (0)
T3 660 (0) 381 (146) 117 (204) 19 (84) 2 (17) 314 (0) 196 (86) 62 (116) 12 (45) 1 (11) 346 (0) 185 (60) 55 (58) 7 (39) 1 (6)
B Admixed American
Diabetes-free survival probability
1·00
0·75
0·50
0·25
0·00
T1 PRS
T2 PRS
T3 PRS
p<0·0001
p=0·0094
p<0·0001
0
5
10
15
20
Number at risk (censored)
T1 631 (0) 403 (198) 115 (274) 30 (82) 1 (28) 318 (0) 231 (100) 62 (145) 16 (46) 0 (0) 313 (0) 190 (98) 53 (129) 14 (36) 1 (12)
T2 630 (0) 371 (192) 114 (232) 20 (91) 1 (19) 330 (0) 216 (99) 83 (123) 12 (70) 1 (11) 301 (0) 156 (93) 32 (109) 8 (22) 0 (0)
T3 631 (0) 354 (183) 108 (213) 17 (74) 1 (15) 298 (0) 190 (93) 69 (114) 13 (50) 0 (0) 332 (0) 173 (90) 38 (99) 4 (23) 1 (3)
C European
Diabetes-free survival probability
1·00
0·75
0·50
0·25
0·00
T1 PRS
T2 PRS
T3 PRS
p<0·0001
p<0·0001
p<0·0001
0
5
10
15
20
25
Time, years
Number at risk (censored)
T1 4234 (0) 3068 (1090) 1267 (1760) 533 (711) 69 (459) 2329 (0) 1743 (575) 751 (980) 332 (423) 39 (281) 1905 (0) 1325 (515) 516 (780) 211 (288) 30 (178)
T2 4234 (0) 3052 (1018) 1194 (1751) 458 (699) 50 (404) 2140 (0) 1591 (517) 679 (883) 291 (373) 29 (260) 2094 (0) 1462 (501) 515 (869) 167 (326) 22 (144)
T3 4234 (0) 2811 (1052) 1057 (1512) 379 (616) 29 (337) 1882 (0) 1339 (489) 567 (723) 214 (332) 19 (193) 2351 (0) 1417 (563) 490 (788) 165 (284) 10 (144)

figure 4; log-rank p<0·01). The risk of developing type 2 diabetes for individuals with both risk factors (ie, PRS T3 and high random glucose concentrations) was substantially higher than the risk for individuals with either factor alone, with HR 5·35 [95% CI 3·67–7·79) in African or African American, 13·82 [7·24–26·39] in Admixed American, and 22·43 [15·61–32·21] in European ancestry groups (appendix 2 tab 14). Similarly, in all ancestry groups, we observed that individuals who had normoglycaemia but were in the highest PRS tertile had a similar risk to individuals who were in the lowest PRS tertile but had random glucose concentrations of at least 140 mg/dL, a criterion in the definition of prediabetes.

The clinical risk score—including age, sex, parental history of type 2 diabetes, BMI, systolic blood pressure, and high-density lipoprotein, total cholesterol, triglyceride, and random glucose concentrations—had concordance indices of 0·73 for African or African American, 0·77 for Admixed American, and 0·83 for European ancestries (appendix 2 tab 15). After adjusting for clinical risk score, the PRS remained significantly associated with incident risk (HR 1·28 [95% CI 1·14–1·44] in African or African American, 1·71 [1·44–2·03] in Admixed American, and 1·81 [1·70–1·94] in European ancestries). This model, combining the PRS and clinical risk score in a multivariable hazards model, further improved the concordance indices in Admixed American (0·77 *vs* 0·78, DeLong p=0·038) and European (0·83 *vs* 0·85; DeLong p<0·0001) ancestries (appendix 2 tab 15).

Discussion

More than 100 PRSs for type 2 diabetes have been published,[22] yet most have low predictive performance in under-represented populations—partly due to the over-representation of European populations in type 2 diabetes GWAS and differential linkage disequilibrium patterns across ancestries. Populations in which PRSs are less predictive—including African or African American, Admixed American, and South Asian ancestry groups—are disproportionately affected by diabetes and its complications, highlighting the risk of exacerbating health disparities through the use of PRSs derived only from European GWAS.[12] To improve the transferability and predictive accuracy of PRSs across diverse populations, including those under-represented in GWAS, various multi-ancestry PRS methods have been developed, while more GWAS from populations with diverse ancestries are also becoming available.[3,15,16,24,25] Beyond large GWAS sample sizes, multiple large datasets are also essential for training and validating PRSs, and rigorous data aggregation and harmonisation following best practices are crucial to assess the accuracy of PRSs. In this study, we developed what are, to our knowledge, the most comprehensive PRSs for type 2 diabetes across five major global ancestries and conducted extensive evaluations.

As previously described,[11,26] we observed that PRSs have lower predictive accuracy when the validation samples are genetically distant from the discovery GWAS samples. However, because GWAS in under-represented populations have limited power and imprecise variant effect-size estimates, PRSs derived from European GWAS—for which the sample size was at least 12 times larger than GWAS in populations with African or African American, Admixed American, and South Asian ancestries—continue to outperform those derived from ancestry-matched GWAS in the respective populations. By contrast, European PRSs had poorer performance in populations of East Asian ancestry than matched-ancestry PRS, even though the East Asian GWAS dataset was four times smaller than the European GWAS dataset. This finding suggests that non-matched-ancestry PRSs (ie, those based on GWAS from an ancestry different from that of the validation cohort) improve performance only when the GWAS sample size in the non-matched ancestry is orders of magnitude larger than that of the matched-ancestry GWAS, despite the lack of power to capture ancestry-specific genetic effects. However, ancestry-matched GWAS have high accuracy when power is sufficient. The minimum ancestry-specific GWAS sample size for robust PRS prediction remains trait-dependent and ancestry-dependent and warrants further investigation. Regardless, we observe that for all ancestry groups—even European, which has the largest GWAS sample size[15]—the optimal strategy for PRS development is to combine PRSs from multiple ancestries, here using the PRS-CSx method. This approach leverages estimates from variants present in all ancestries while accounting for population-specific or enriched variants, therefore improving the overall accuracy of prediction.

Across all ancestries, multi-ancestry PRSs showed the strongest associations with type 2 diabetes risk at the distribution extremes. In European, East Asian, and Admixed American ancestries, individuals in the top 97·5th percentile had a seven-times higher risk of developing type 2 diabetes than individuals in the IQR. Despite lower representation in terms of GWAS, the multi-ancestry PRS still outperformed matched-ancestry PRSs for African or African American and South Asian ancestries. However, the overall performance was lower for these populations than for those of other ancestry groups, with individuals in the top 97·5th percentile having a 3–4-times higher risk of developing type 2 diabetes than those in the IQR. These results underscore that, despite improvements, a substantial performance gap remains for under-represented populations.

In the All of Us cohort, our multi-ancestry PRSs outperformed previously available type 2 diabetes PRSs, probably because we used, to our knowledge, the largest and most diverse GWAS dataset to date and multi-ancestry-based methods for PRS development. Additionally, we constructed these multi-ancestry PRSs

using a standardised and rigorous approach and tested them extensively across diverse ancestry groups, supporting their broader applicability. For example, the multi-ancestry PRSs outperformed PRSs derived from the largest trans-ancestry meta-analysis.[3] These findings underscore that, in addition to increasing representation in GWAS, jointly modelling GWAS and linkage disequilibrium panels across ancestries can enhance causal variant tagging, thereby improving predictive performance. By contrast, standard methods using multi-ancestry inverse variance-weighted GWAS meta-analysis results rely on a single linkage disequilibrium reference panel, and could fail to capture ancestry-specific linkage disequilibrium patterns, limiting prediction accuracy.

Previous studies have reported significant associations between type 2 diabetes PRSs, proliferative diabetic retinopathy, and end-stage diabetic nephropathy in diverse ancestries.[15,16] We extend these findings, showing that our multi-ancestry PRSs can predict microvascular complications (diabetic retinopathy, diabetic neuropathy, proliferative diabetic retinopathy, and end-stage diabetic nephropathy) in individuals with type 2 diabetes of African or African American, Admixed American, and European ancestries. Notably, these associations are independent of the duration of type 2 diabetes. Future work is needed to assess whether these associations are driven by specific type 2 diabetes subtypes and, if so, whether the associations vary by ancestry.[3,27]

Our work has limitations. First, although constructing multi-ancestry PRSs that include GWAS from diverse ancestries enables the capture of ancestry-specific effects, the SNP effect sizes are still strongly influenced by the largest European cohorts. Second, we acknowledge that using discrete population categories is suboptimal, as this does not fully capture heterogeneous ancestry, particularly in admixed populations. Newer methods that model both local and global ancestry continuously could be powerful,[11,26] but often require individual-level data, limiting their scalability. Our approach, although limited to discrete ancestral groups, enables us to leverage the wealth of data from large-scale GWAS summary statistics across multiple ancestries. As larger biobanks emerge and methodological challenges (eg, the accurate modelling of local ancestry in admixed populations) are solved, adopting continuous genetic ancestry methods or hybrid models will become the optimal choice.

Despite some concerns that PRSs add little value over the use of clinical risk factors, previous work shows that PRSs can help to identify individuals at high risk of developing type 2 diabetes among those who are clinically perceived as being at low risk, (eg, those who are young, lean, or who have sparse clinical data[8]), and can improve prediction compared with family history alone.[28] We found that higher multi-ancestry PRSs were associated with an earlier onset of type 2 diabetes of up to 2·3 years across the three studied ancestries, supporting their utility in identifying individuals at high risk before they accumulate clinical risk factors. Furthermore, our PRSs predicted the development of type 2 diabetes independently of a nine-component clinical risk score, and showed that individuals with the highest PRSs and high random glucose concentrations (≥140 mg/dL) were at higher risk of developing type 2 diabetes (eg, HR 13·82 [95% CI 7·24–26·39] in Admixed American populations) than individuals with normoglycaemia and low PRSs.

Ongoing randomised controlled trials are starting to evaluate PRS implementation in primary care.[7,29] Viable interventions require state-of-the-art type 2 diabetes PRSs, prioritising generalisability, portability across ancestries, and the feasibility of implementation. For example, Lennon and colleagues[6] prioritised a multi-ancestry type 2 diabetes PRS[30] trained using GWAS data from populations of European, East Asian, and African ancestries to propose a high-risk threshold at the 98th percentile. Compared with the remaining individuals not at high risk, this PRS threshold was associated with a 4·44-times increased risk of developing type 2 diabetes in populations of European ancestry and a 2·35-times increased risk in populations of African ancestry. Notably, our multi-ancestry PRSs showed consistently improved predictive performance in all studied ancestries over previously reported PRS. Compared with individuals with average genetic risk, those at the highest genetic risk cutoff (97·5th percentile) have ORs ranging from 3·43 to 7·47 across ancestries. In summary, we deliver a comprehensive and rigorously tested set of multi-ancestry PRSs for type 2 diabetes.

Contributors

JMM and MCYN conceptualised the study. JMM, MCYN, and AH-C supervised the implementation of research plans across teams. AH-C and JK led the data curation. RM, YL, KS, LEP, HKN, JC, SL, MR, KL, and KT curated data and conducted statistical analyses to validate the PRSs in independent cohorts. AKM supervised the analyses. CAA-S, LG-G, CG-V, CAH, YJK, SHK, AL, RJFL, AM-E, APM, LO, JIR, DS, TT-L, BFV, MV, RGW, ML, JEB, and XS provided necessary data and supervised different stages of the analyses. JK and AH-C conducted the initial statistical analyses, with JK conducting additional analyses during revisions. JMM, MCYN, AH-C, JK, and TG guided the interpretation of the results. TG contributed to the design of the methodology and developed PRS-CSx. JMM, MCYN, and AKM contributed to funding acquisition and project administration. JMM, MCYN, AH-C, JK, and TG wrote the initial draft of the manuscript, which was reviewed by JMM and MCYN; AH-C prepared the figures and tables and prepared the final version of the manuscript. JMM and MCYN had full access to all data in the study. JMM and MCYN accessed and verified the data. All authors had final responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

Individual participant data are not available because they are subject to data protection laws and restrictions imposed by the relevant ethics committees to ensure the privacy of the study participants. The developed PRS weights are available without restrictions through the Polygenic Score Catalog (https://www.pgscatalog.org; publication ID PGP000773, score IDs PGS005353–5377). The linkage disequilibrium panels are available upon request. The analysis code is reported at https://github.com/ahuertach/D-PRISM-T2D-PRS.

Acknowledgments

This work is supported by the National Human Genome Research Institute (NHGRI) of the US National Institutes of Health (NIH) grant U01HG011723. AH-C is supported by the American Diabetes Association (ADA) grant 11–23-PDF-35. YL is supported by National Heart, Lung, and Blood Institute (NHBLI) grant R56HL150186; NHLBI grant R01HL158884; and National Institute of Diabetes and Digestive and Kidney Disease (NIDDK) grant R01DK135938. AL is supported by grant 2020096 from the Doris Duke Foundation, the ADA grant 7–22-ICTSPM-23, and National Human Genome Research Institute (NHGRI) grant U01HG011723. APM acknowledges support from the National Institute for Health and Care Research Manchester Biomedical Research Centre (NIHR203308). BFV is grateful for support from NIDDK grants DK138521 and DK126194. AKM is supported by NIDDK grant UMDK078616. JMM is supported by ADA grant 11–22-ICTSPM-16; by NHGRI grant U01HG011723; by the NIDDK under award numbers R01DK137993, R01DK140545, and U01 DK140757; by the Accelerating Medicines Partnership Common Metabolic Diseases award from RFP 6 from the Foundation for the NIH; and by a Medical University of Bialystok grant from the Ministry of Science and Higher Education (Poland). This work is supported by the Novo Nordisk Foundation (NNF21SA0072102). MCYN is supported by NHGRI grant U01HG011723 and by NIDDK grants R01DK066358 and U01DK105556. This study was also supported by a grant from the National Research Foundation of Korea funded by the Korean Ministry of Science and ICT (RS-2023–00262002), and by the Ministry of Food and Drug Safety grant (23212MFDS202) awarded to SHK and JC. JEB and LEP were supported in part by NHLBI grant R01HL142302.

References

1 International Diabetes Federation. IDF Diabetes Atlas, 11th edition. Brussels: International Diabetes Foundation, 2025.
2 Mercader JM, Ng MCY, Manning AK, Rich SS. Predicting diabetes risk in diverse populations: what next? *Lancet Diabetes Endocrinol* 2021; **9:** 808–10.
3 Suzuki K, Hatzikotoulas K, Southam L, et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* 2024; **627:** 347–57.
4 Udler MS, McCarthy MI, Florez JC, Mahajan A. Genetic risk scores for diabetes diagnosis and precision medicine. *Endocr Rev* 2019; **40:** 1500–20.
5 Knowler WC, Fowler SE, Hamman RF, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet* 2009; **374:** 1677–86.
6 Lennon NJ, Kottyan LC, Kachulis C, et al. Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. *Nat Med* 2024; **30:** 480–87.
7 Vassy JL, Brunette CA, Lebo MS, et al. The GenoVA study: equitable implementation of a pragmatic randomized trial of polygenic-risk scoring in primary care. *Am J Hum Genet* 2023; **110:** 1841–52.
8 Mandla R, Schroeder P, Porneala B, et al. Polygenic scores for longitudinal prediction of incident type 2 diabetes in an ancestrally and medically diverse primary care physician network: a patient cohort study. *Genome Med* 2024; **16:** 63.
9 Fortmann AL, Savin KL, Clark TL, Philis-Tsimikas A, Gallo LC. Innovative diabetes interventions in the U.S. Hispanic population. *Diabetes Spectr* 2019; **32:** 295–301.
10 Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 2019; **51:** 584–91.
11 Kachuri L, Chatterjee N, Hirbo J, et al. Principles and methods for transferring polygenic risk scores across global populations. *Nat Rev Genet* 2024; **25:** 8–25.
12 Kullo IJ, Conomos MP, Nelson SC, et al. The PRIMED Consortium: reducing disparities in polygenic risk assessment. *Am J Hum Genet* 2024; **111:** 2594–606.
13 Kidenya BR, Mboowa G. Inclusiveness of the All of Us Research Program improves polygenic risk scores and fosters genomic medicine for all. *Commun Med (Lond)* 2024; **4:** 227.
14 Bick AG, Metcalf GA, Mayo KR, et al. Genomic data in the All of Us Research Program. *Nature* 2024; **627:** 340–46.
15 Mahajan A, Spracklen CN, Zhang W, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet* 2022; **54:** 560–72.
16 Vujkovic M, Keaton JM, Lynch JA, et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet* 2020; **52:** 680–91.
17 Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 2023; **613:** 508–18.
18 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; **526:** 68–74.
19 Ge T, Chen C-Y, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019; **10:** 1776.
20 Ruan Y, Lin Y-F, Feng YA, et al. Improving polygenic prediction in ancestrally diverse populations. *Nat Genet* 2022; **54:** 573–80.
21 Wojcik GL, Fuchsberger C, Taliun D, et al. Imputation-aware tag SNP selection to improve power for large-scale, multi-ethnic association studies. *G3 (Bethesda)* 2018; **8:** 3255–67.
22 Lambert SA, Wingfield B, Gibson JT, et al. Enhancing the Polygenic Score Catalog with tools for score calculation and ancestry normalization. *Nat Genet* 2024; **56:** 1989–94.
23 Farmaki P, Damaskos C, Garmpis N, Garmpi A, Savvanis S, Diamantis E. Complications of the type 2 diabetes mellitus. *Curr Cardiol Rev* 2020; **16:** 249–51.
24 Loh M, Zhang W, Ng HK, et al. Identification of genetic effects underlying type 2 diabetes in South Asian and European populations. *Commun Biol* 2022; **5:** 329.
25 Spracklen CN, Horikoshi M, Kim YJ, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* 2020; **582:** 240–45.
26 Ding Y, Hou K, Xu Z, et al. Polygenic scoring accuracy varies across the genetic ancestry continuum. *Nature* 2023; **618:** 774–81.
27 Hodgson S, Williamson A, Bigossi M, et al. Genetic basis of early onset and progression of type 2 diabetes in South Asians. *Nat Med* 2025; **31:** 323–31.
28 Drzymalla E, Raffield L, Kolor K, et al. Additive value of polygenic risk score to family history for type 2 diabetes prediction: results from the All of Us research database. *Diabetes Care* 2025; **48:** 212–19.
29 Linder JE, Allworth A, Bland HT, et al. Returning integrated genomic risk and clinical recommendations: the eMERGE study. *Genet Med* 2023; **25:** 100006.
30 Ge T, Irvin MR, Patki A, et al. Development and validation of a trans-ancestry polygenic risk score for type 2 diabetes in diverse populations. *Genome Med* 2022; **14:** 70.

THE LANCET
Diabetes & Endocrinology

Supplementary appendix 1

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Huerta-Chagoya A, Kim J, Mandla R, et al. Multi-ancestry polygenic risk scores for the prediction of type 2 diabetes and complications in diverse ancestries. *Lancet Diabetes Endocrinol* 2026; published online April 27. https://doi.org/10.1016/S2213-8587(25)00405-X.

Supplementary material

SUPPLEMENTARY METHODS 1

SUPPLEMENTARY FIGURES 6

FULL LIST OF D-PRISM AUTHORS 17

AUTHOR ACKNOWLEDGMENTS AND FUNDING 21

COHORT ACKNOWLEDGMENTS AND FUNDING 22

ETHICS APPROVAL COMMITTEES/IRBS 33

MEMBERS OF THE ENSA GENOMICS CONSORTIUM 39

GENES & HEALTH RESEARCH TEAM AUTHORSHIP FOR SCIENTIFIC PUBLICATIONS 40

VA MILLION VETERAN PROGRAM CORE ACKNOWLEDGEMENTS FOR PUBLICATIONS 42

Supplementary methods

Type 2 diabetes GWAS meta-analyses used for the construction of the PRSs

We leveraged the T2D GWAS summary statistics from a subset of cohorts participating in three large Consortia: Diabetes Meta-analysis of Trans-ethnic Association Studies (DIAMANTE)[1], the Million Veteran Program (MVP)[2], and the FinnGen Study.[3] An independent subset of cohorts was selected for the development and validation of the PRSs. If a cohort was multi-ancestry, each individual was categorized by genetic similarity to one or more of the five ancestries available in the 1000 Genomes (1KG) Project[4] and/or the Human Genome Diversity Project[5] as reference panels: African/African American (AFR), Admixed American (AMR), East Asian (EAS), European (EUR), and South Asian (SAS). For single-ancestry cohorts, the grouping was based on the country of recruitment. We included 2,185,548 individuals (359,819 T2D cases and 1,825,729 controls) across 125 T2D GWAS to conduct ancestry-specific meta-analyses (figure 1A, appendix 2 tabs 1,2).

Each GWAS tested the association of the genetic variants with T2D adjusted for age, sex, the top genetic principal components (PCs), and cohort-specific covariates. We performed an inverse variance weighted (IVW) fixed-effect meta-analysis for each ancestry group with the METAL software.[6] We then applied quality control to keep biallelic, nonpalindromic SNPs in at least half of the effective sample size with a minor allele frequency (MAF) ≥0·01. For each ancestry-specific T2D GWAS meta-analysis, we intended to include most cohorts to maximize the sample size of the summary statistics for constructing PRSs while leaving out sufficient cohorts for each ancestry to be used for training and validation.

Cohorts for the training and validation of the PRSs

We trained the PRSs in one cohort per ancestry group and validated them in at least four validation cohorts per ancestry. All training and validation cohorts included unrelated individuals and were independent of those included in the GWAS summary statistics to avoid overfitting (figure 1B, C, appendix 2 tabs 3-6). Except for the AoU cohort, for which whole genome sequencing is available, the genotyping of the other cohorts was chip array-based. The genotypes were imputed to the 1KG[4] or the TOPMed r2[7,8] reference panels using the Michigan Imputation server.[9] We applied a separate post-imputation quality control in each cohort and ancestry to keep biallelic nonpalindromic SNPs with an imputation quality of r^2≥0·8 and MAF≥0·005. We excluded the variants not included in the LD reference panels, as explained below, or variants that showed an allelic frequency discordance≥0·2 compared to the 1KG ancestry-specific allelic frequency.

LD reference panels for the construction of the PRSs

To account for the correlation between variants, we constructed customized ancestry-specific LD reference panels using the same scripts used in PRS-CS[10] and PRS-CSx[11] tools. We built four new sets of ancestry-specific LD reference panels using the HapMap3 (HM3) set of variants, similar to the official panel (https://github.com/getian107/PRScsx), or an expanded 1KG set of variants, along with pairwise LD from the 1KG or in-house samples.

First, we identified ancestry-specific LD blocks using LDetect.[12] For each of the five ancestry groups in the 1KG dataset (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#reference), we selected common SNPs with MAF≥0·01 to generate a covariance matrix of variants based on the LD r^2 calculated in PLINK v1·9[13] and derived the boundaries of LD blocks. Second, we generated two sets of reference SNPs. One was based on the HM3 set of variants (https://www.sanger.ac.uk/data/hapmap-3/), similar to the official PRS-CS[10]/PRS-CSx[11] HM3 version. Another expanded set of variants based on 1KG (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#reference) was selected using the Tag(ging) It(erative) of SNPs in multiple populations (TagIt) program.[14] TagIt allows the selection of tag SNPs by leveraging genetic information from multiple diverse ancestries to maximize cross-population coverage. We only included non-palindromic SNPs with MAF≥0·01 in at least one ancestry group for both the HM3 and TagIt SNP lists. Third, the two sets of SNPs were extracted in each ancestry for individuals from the 1KG dataset (347 to 661 individuals per ancestry) and from the larger imputed in-house datasets (including around 10,000 individuals per ancestry). SNPs with low imputation quality (r^2<0·8) were further excluded in the in-house datasets. Last, we calculated the variants' pairwise LD (r^2) using PLINK v1·9 to generate ancestry-specific LD reference panels. In total, for each ancestry, we built four different LD reference panels and used them to construct the PRSs, combining two sets of variants (i.e., HM3 and TagIt) and two sources of LD information (i.e., 1KG and in-house samples) (appendix 2 tab 7).

Training of the PRSs

We used the PRS-CS[10] Bayesian polygenic method to construct single-ancestry PRSs. For each ancestry group, we leveraged the T2D GWAS summary statistics and LD reference panels matching the ancestry of the validation cohort (e.g., a PRS trained using the AMR GWAS and AMR LD reference panel to be validated in AMR cohorts). When applicable, we also modeled non-matched single-ancestry PRSs (e.g., a PRS trained using the EUR GWAS and EUR LD reference panel to be evaluated in AMR validation cohorts). The PRS-CS method returns a single-ancestry posterior variant effect size.

We then used PRS-CSx[11] to construct multi-ancestry PRSs. Instead of meta-analyzing the ancestry-specific GWAS summary statistics, PRS-CSx jointly models the GWAS summary statistics along with their matching LD reference panels, using a shared continuous shrinkage prior, to generate ancestry-specific variant posterior effect sizes in a coupled manner that leverages cross-population genetic architecture. We used these ancestry-specific effect sizes to compute standardized ancestry-specific z-scores, which were then combined in a linear regression model to derive the multi-ancestry posterior variant effect size as follows:

$$y = PRS_{\theta,AFR} + PRS_{\theta,AMR} + PRS_{\theta,EAS} + PRS_{\theta,EUR} + PRS_{\theta,SAS}$$

Where y is the T2D status, and $PRS_{\theta,\text{ ancestry group}}$ is the standardized PRS for a given shrinkage prior (θ) and ancestry.

For both PRS-CS and PRS-CSx, we used the training cohorts to select the optimal continuous shrinkage prior from five phi values (i.e., 0·01, 0·001, $1x10^{-4}$, $1x10^{-5}$, $1x10^{-6}$) based on predictive performance. In total, we constructed 80 PRSs for each of the AFR, AMR, and SAS ancestry groups (i.e., 1 matched-ancestry PRS, 2 non-matched ancestry PRSs, and 1 multi-ancestry PRS × 4 LD panels × 5 phi values = 80 models) and 60 PRSs for each of the EAS and EUR ancestry groups (i.e., 1 matched-ancestry PRS, 1 non-matched ancestry PRS, and 1 multi-ancestry PRS × 4 LD panels × 5 phi values = 60 models), resulting in 360 PRS models overall.

To test the predictive performance, we applied the posterior variant effect sizes for each PRS model to calculate the individual scores in each of the five training cohorts using the --score function in PLINK v1·9.[9] We standardized them to have a mean of zero and unit variance. Then, we fitted two logistic regression models and calculated the area under the receiver operator characteristic curve (AUC) using the "pROC" package[15] in R. One model included the explanatory variables sex, age, and genetic PCs, and a second full model also included the standardized PRS. We also fitted logistic regression models adjusted for body mass index (BMI).

We calculated the incremental AUC (iAUC) by subtracting the AUC of the model without the PRS from the AUC of the full model. We defined the best-trained PRS models as those with the continuous shrinkage prior and LD panel that maximized the iAUC. After the training step, we ended up with four best-trained PRS models for the AFR, AMR, and SAS ancestries (i.e., one matched-ancestry PRS, one EUR non-matched ancestry PRS, one EAS non-matched ancestry PRS, and one multi-ancestry PRS), 3 best-trained PRS models for the EUR ancestry (i.e., one single matched-ancestry PRS, one EAS non-matched ancestry PRS, and one multi-ancestry PRS), and 3 best-trained PRS models for the EAS ancestry (i.e., one single matched-ancestry PRS, one EUR non-matched ancestry PRS, and one multi-ancestry PRS) (appendix 2 tab 8).

Validation of the PRSs

To validate each of the 18 best-trained PRSs, we applied the posterior variant effect sizes in a second set of unrelated samples and independent cohorts from each ancestry group. We calculated the individual scores in each validation cohort using the --score function in PLINK 1·9[13] and standardized them to have a mean of zero and unit variance. For the multi-ancestry PRS, we combined ancestry-specific standardized scores weighted for the trained ancestry-specific effect sizes as follows:

$$y = \beta_{\theta,AFR}\, PRS_{\theta,AFR} + \beta_{\theta,AMR}\, PRS_{\theta,AMR} + \beta_{\theta,EAS}\, PRS_{\theta,EAS} + \beta_{\theta,EUR}\, PRS_{\theta,EUR} + \beta_{\theta,SAS}\, PRS_{\theta,SAS}$$

Where y is T2D status, $\beta_{\theta,\text{ ancestry group}}$ is the regression coefficient for a given shrinkage prior and ancestry in the training cohort, and $PRS_{\theta,\text{ ancestry group}}$ is the standardized PRS for a given shrinkage prior and ancestry in the validation cohort.

To test the predictive performance of the PRS, we calculated i) the iAUC as explained above, ii) the proportion of the variation in the T2D status explained by the PRS using Nagelkerke's r^2, iii) the odds ratio per standard deviation (OR per SD) change in the PRS, and iv) the discrimination capacity at the extremes of the PRS distribution by identifying

the individuals at the top 97·5th percentile, 95th percentile and 90th percentile for comparison with the interquartile range.

We applied the DeLong test to statistically assess the difference between iAUCs of the single ancestry vs. multi-ancestry PRSs. We combined the PRS estimates across validation cohorts using fixed-effects meta-analyses by ancestry, weighting each cohort's beta coefficient by the inverse of its variance, using the "metafor" package[16] in R (appendix 2 tab 9,10).

Comparison of the best-performing multi-ancestry PRSs to previously published T2D PRS

Multiple efforts have been made to improve the portability of PRS to apply to individuals from diverse ancestries. Until the preparation of this work (revised on October 07, 2024), 147 PRSs for T2D had been constructed and made publicly available through the PGS catalog.[17] To compare the performance of our best-trained PRS models using PRS-CSx (which we refer to as 'D-PRISM multi-ancestry PRS-CSx' model), we selected 55 PRSs from the PGS catalog that i) were trained for the overall T2D trait and not for any specific subtype of the disease, ii) were trained considering all types of genetic variants and not any specific set of variants for specific biological mechanisms, and iii) were constructed and trained by leveraging genetic information from cohorts other than the AoU, which we used as a validation cohort in this study.

We downloaded the PRSs and extracted the variants from the AoU cohort[18] (release v7, May 2022). Since the variant missingness rate was below 10% for all PRSs, we included all of them for testing. We calculated the individual scores in each of the five ancestries using the --score function in PLINK v1·9[13] and standardized them to mean zero and unit variance. Then, we tested the performance of the PRSs using the same procedure as for the validation cohorts.

To assess the added value of the PRS-CSx approach over constructing a PRS from meta-analyzed multi-ancestry GWAS results, we performed an IVW meta-analysis of D-PRISM ancestry-specific GWAS summary statistics and applied PRS-CS to the resulting meta-analysis. We used the EUR 1KG HM3 LD reference panel and allowed the shrinkage prior to be automatically learned from the data (which we refer to as 'D-PRISM multi-ancestry PRS-CS' model).

Similarly, we also evaluated the performance of a PRS based on the most statistically powered multi-ancestry T2D GWAS to date from Suzuki et al.[19,] including a 16% larger sample size than this study (i.e., Suzuki et al.: 2,535,601 individuals including 428,452 T2D cases and 2,107,149 controls vs. D-PRISM: 2,185,548 individuals including 359,819 T2D cases and 1,825,729 controls). We first applied the same quality control steps as those used for D-PRISM GWAS summary statistics and then used PRS-CS to construct a PRS using the EUR 1KG HM3 LD reference panel and let the prior shrinkage prior to be automatically learned from the data (which we refer to as 'Suzuki et al., PRS-CS' model).

Additionally, we constructed an rsPRS using the 1,289 distinct T2D variants identified by Suzuki et al.,[19] defined as having an association $p<5 \times 10^{-8}$. For 90 out of 203 palindromic variants, we used proxy variants ($r^2 \geq 0{\cdot}8$ in all the ancestry groups). We excluded 113 variants, as no proxy was available, thereby using 1,176 variants to construct the rsPRS (which we refer to as 'Suzuki et al., rsPRS' model) (appendix 2 tab 11).

Association of PRSs with diabetes complications and comorbidities

We also evaluated the association of the best-performing D-PRISM multi-ancestry PRS-CSx models with T2D-related microvascular complications (i.e., diabetic nephropathy, diabetic retinopathy, end-stage diabetic nephropathy, and proliferative diabetic retinopathy) and comorbidities (i.e., cardiovascular disease and ischemic stroke) in individuals with T2D in the AoU validation cohort. To comply with the AoU policies, we only considered individuals of the AFR, AMR, and EUR ancestry groups, as there were limited sample sizes for the EAS and SAS ancestries (i.e., <30 individuals with diabetes complications). We defined the traits based on International Classification of Diseases (ICD) codes (versions 9 and 10), as previously described (appendix 2 tab 12).[19] We tested the association of each T2D-related traits with the standardized PRSs by fitting logistic regression models adjusted for sex, age, and genetics PCs in each ancestry group, separately (appendix 2 tab 13).

The defining ICD codes for microvascular complications were restricted to those explicitly indicating a diabetic etiology. However, to establish the temporal sequence between diabetes and its microvascular complications, we extracted the date when the microvascular ICD code was assigned and contrasted it with an estimated date of diabetes

onset. To estimate the date of diabetes onset, we leveraged the EHR and survey data to extract the earliest date when a patient met altered glycemic criteria (HbA1C≥6.5% or random glucose≥200 mg/dL at outpatient settings), had an ICD code for T2D, or was prescribed a diabetes-specific medication at outpatient settings (insulin or non-insulin). When available, this estimated date was contrasted with the self-reported date of diagnosis. If the estimated onset date fell outside the self-reported diagnosis window, the evidence was considered insufficient to establish a reliable onset time-point. Once we estimated the date of diabetes onset, we considered a post-diabetes complication when the estimated age at diabetes onset preceded the first record of the relevant outcome by 6 months or more. More than 70% of the individuals were recorded to have a microvascular complication after the estimated age at diabetes onset (79% with DR, 77% with PDR, 85% with DN, and 72% with ESDN).

We performed two additional sensitivity analyses. First, we restricted the analysis to T2D patients where the estimated age at T2D onset preceded the first record of the complication. This analysis yielded consistent results compared to the analysis of the overall sample (Supplementary Figure 10). Second, we adjusted for the estimated duration of T2D, in addition to the other covariates, to assess whether the PRS's effect was mediated by T2D duration (e.g., via earlier onset) or if it represented a direct effect independent of duration. T2D duration was defined as: T2D duration = last age in EHR record - EHR-based estimated age at T2D onset. Although the effect was attenuated, the PRS remained significantly associated with complications, suggesting a direct effect risk independent of disease duration (Supplementary Figure 11; appendix 2 tab 13).

Association of PRSs with incident T2D, alone and in combination with clinical risk factors

To better assess the clinical impact of the PRSs, we extracted EHR data from the All of Us v8 cohort, implemented a comprehensive clinical risk score (CRS) based on the T2D Framingham Risk Score[20] and assessed the effects of the PRS in the context of the CRS.

We extracted EHR data from January 1, 2000, to December 31, 2023, and considered patients who had received primary care data and had at least two records with available glucose measurements at outpatient settings. After excluding patients with a T2D diagnosis before or within 1 year of their first record (the "baseline date"), 17,920 individuals with available CRS data remained for the analysis. Of these, 2,074 developed incident T2D and 15,846 remained non-diabetic during follow-up. Only individuals who met criteria for T2D based on a previously reported algorithm for T2D in the All of Us cohort were considered incident cases.[21] Time to T2D onset was defined as the time until a participant met glycemic criteria for T2D (HbA1C ≥6.5% or random glucose at an outpatient setting ≥ 200 mg/dL), had an ICD code for T2D, or was prescribed a diabetes-specific medication (insulin or non-insulin) at an outpatient setting during the observation period. Individuals were right-censored at the date of the first T2D-defining event or their last clinical record before December 31, 2023.

For this cohort, we implemented an adapted clinical risk score (CRS) based on the Framingham Offspring Study as described in Mandla et al.[22] This CRS comprised 9 components: age, sex, parental history of T2D, body mass index (BMI), systolic blood pressure (SBP), high-density lipoprotein (HDL), total cholesterol, triglycerides, and random glucose. The original Framingham score included fasting glucose and waist circumference, but these were not available in AoU data. Therefore, we omitted waist circumference and used random glucose values as a proxy for fasting glucose. Briefly, for each participant, we extracted the mean values of SBP, HDL, total cholesterol, triglycerides, and random glucose measurements at the outpatient setting between 1 year before and 1 year after their baseline date. No imputation of missing data was performed. The parental history of T2D was extracted from survey data, and we used the median BMI during the follow-up period. The resulting CRS was log-transformed to be normally distributed and standardized.

We evaluated the predictive value of the PRS for T2D across the three ancestries for which we had sufficient sample size: AFR, AMR, and EUR. We performed Kaplan-Meier survival analysis, stratifying participants by tertiles of the PRS. To quantify this risk, Cox proportional hazards models were fitted for each ancestry group. We next investigated whether the PRS provided predictive value independent of random glucose, a primary T2D risk factor. We stratified participants into those with random glucose < 140 mg/dl (i.e., normoglycemia), and those with random glucose ≥140 mg/dl (i.e., one of the criteria to define prediabetes) and compared the effects of each PRS tertile in each group. To formally quantify this combined effect, we fitted Cox models setting the group with the lowest PRS tertile (T1) and random glucose <140 mg/dl as reference. Finally, we also evaluated the added value of the PRS in addition to the CRS. We then fitted a multivariable Cox model including the PRS, the CRS, and genetic principal components. All

models were adjusted for the genetic principal components. Median follow-up time was estimated using the reverse Kaplan-Meier method.

Method's references

1 Mahajan A, Spracklen CN, Zhang W, *et al.* Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet* 2022; 54: 560–72.

2 Vujkovic M, Keaton JM, Lynch JA, *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet* 2020; 52: 680–91.

3 Kurki MI, Karjalainen J, Palta P, *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 2023; 613: 508–18.

4 1000 Genomes Project Consortium, Auton A, Brooks LD, *et al.* A global reference for human genetic variation. *Nature* 2015; 526: 68–74.

5 Bergström A, McCarthy SA, Hui R, *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science* 2020; 367. DOI:10.1126/science.aay5012.

6 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; 26: 2190–1.

7 Kowalski MH, Qian H, Hou Z, *et al.* Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet* 2019; 15: e1008500.

8 Taliun D, Harris DN, Kessler MD, *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 2021; 590: 290–9.

9 Das S, Forer L, Schönherr S, *et al.* Next-generation genotype imputation service and methods. *Nat Genet* 2016; 48: 1284–7.

10 Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019; 10: 1776.

11 Ruan Y, Lin Y-F, Feng Y-CA, *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat Genet* 2022; 54: 573–80.

12 Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* 2016; 32: 283–5.

13 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; 4: 7.

14 Wojcik GL, Fuchsberger C, Taliun D, *et al.* Imputation-Aware Tag SNP Selection To Improve Power for Large-Scale, Multi-ethnic Association Studies. *G3 (Bethesda)* 2018; 8: 3255–67.

15 Robin X, Turck N, Hainard A, *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; 12: 77.

16 Viechtbauer W. Conducting Meta-Analyses in *R* with the metafor Package. *J Stat Softw* 2010; 36. DOI:10.18637/jss.v036.i03.

17 Lambert SA, Wingfield B, Gibson JT, *et al.* Enhancing the Polygenic Score Catalog with tools for score calculation and ancestry normalization. *Nat Genet* 2024; 56: 1989–94.

18 All of Us Research Program Genomics Investigators. Genomic data in the All of Us Research Program. *Nature* 2024; 627: 340–6.

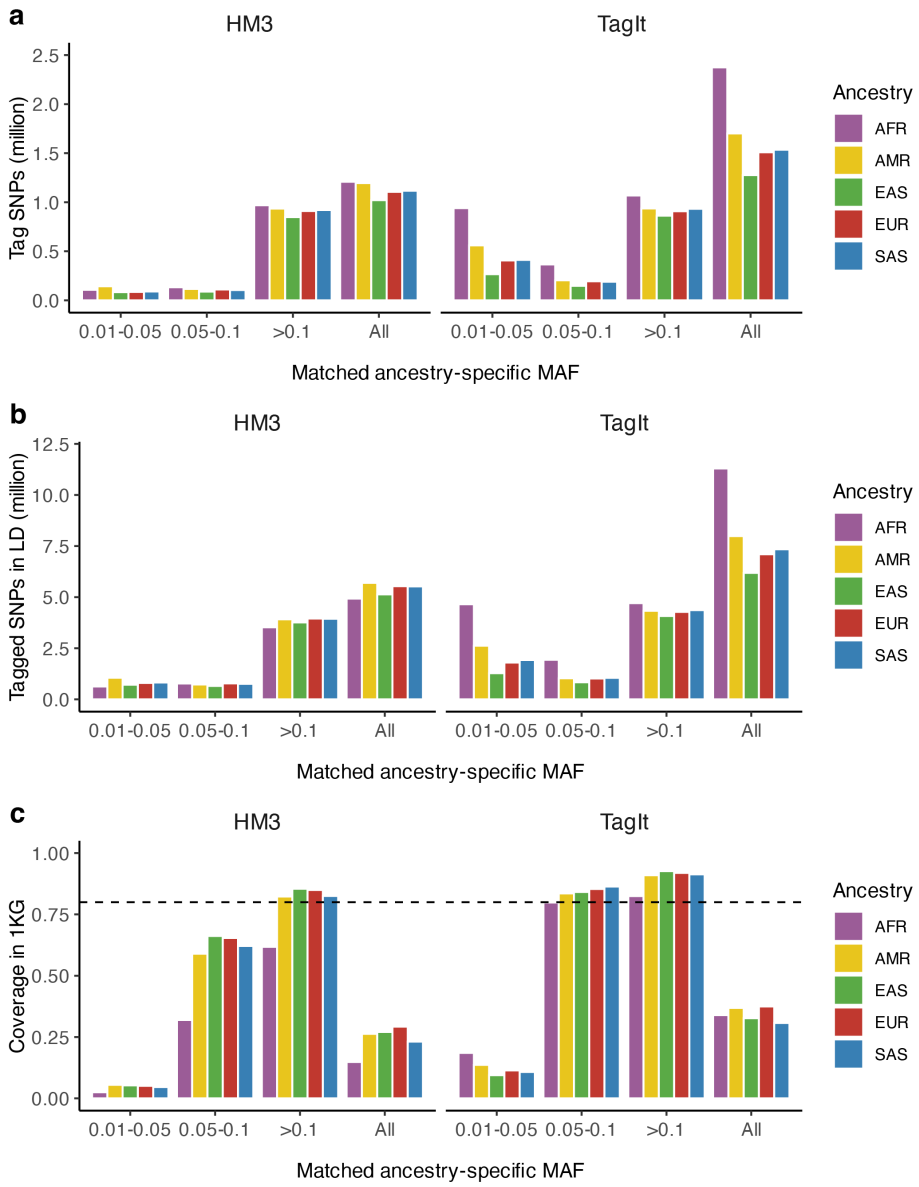
19 Suzuki K, Hatzikotoulas K, Southam L, *et al.* Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* 2024; 627: 347–57.

20 Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med* 2007; 167: 1068–74.

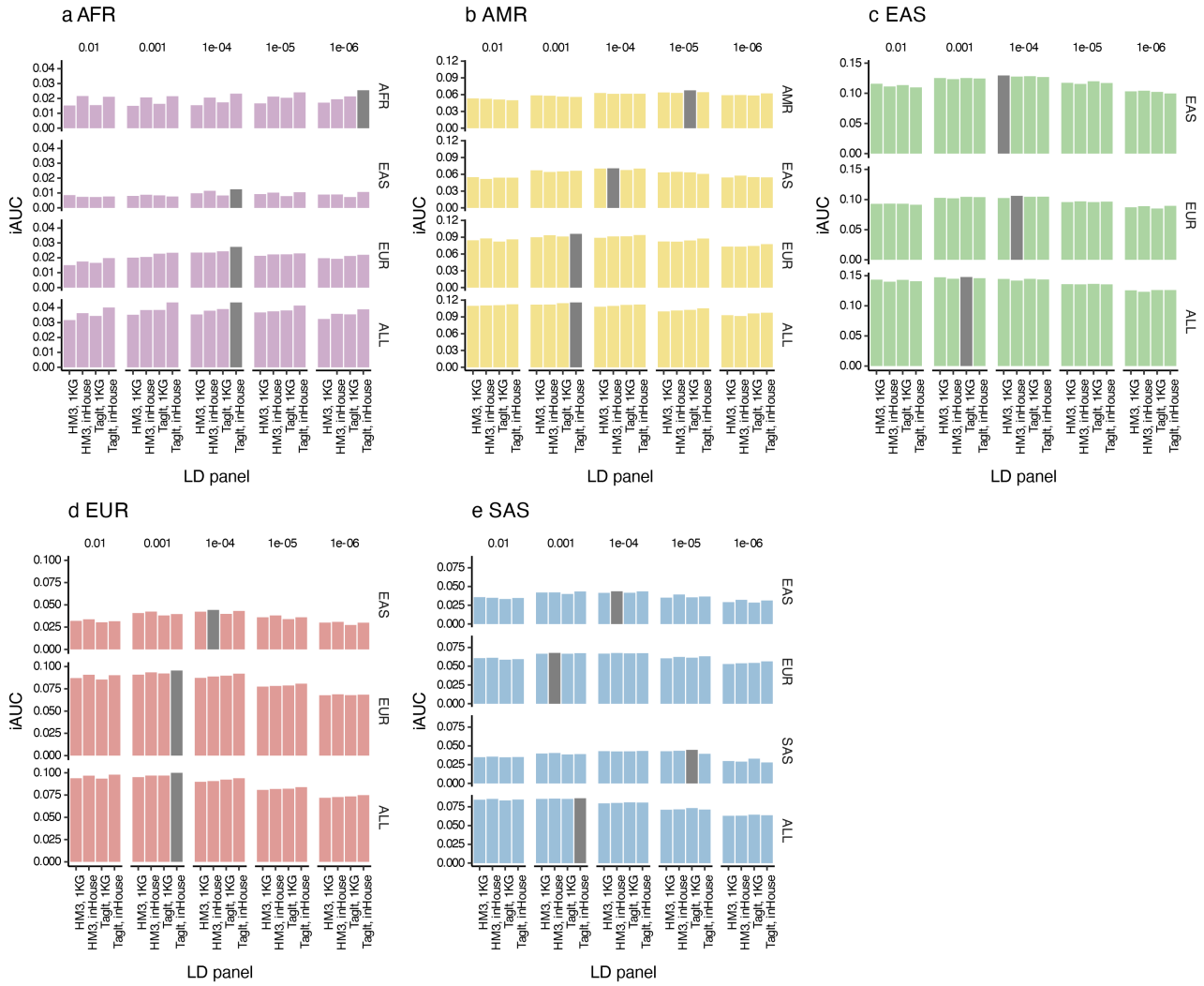
21 Szczerbinski L, Mandla R, Schroeder P, *et al.* Algorithms for the identification of prevalent diabetes in the All of Us Research Program validated using polygenic scores. *Sci Rep* 2024; 14: 26895.

22 Mandla R, Schroeder P, Porneala B, *et al.* Polygenic scores for longitudinal prediction of incident type 2 diabetes in an ancestrally and medically diverse primary care physician network: a patient cohort study. *Genome Med* 2024; 16: 63.

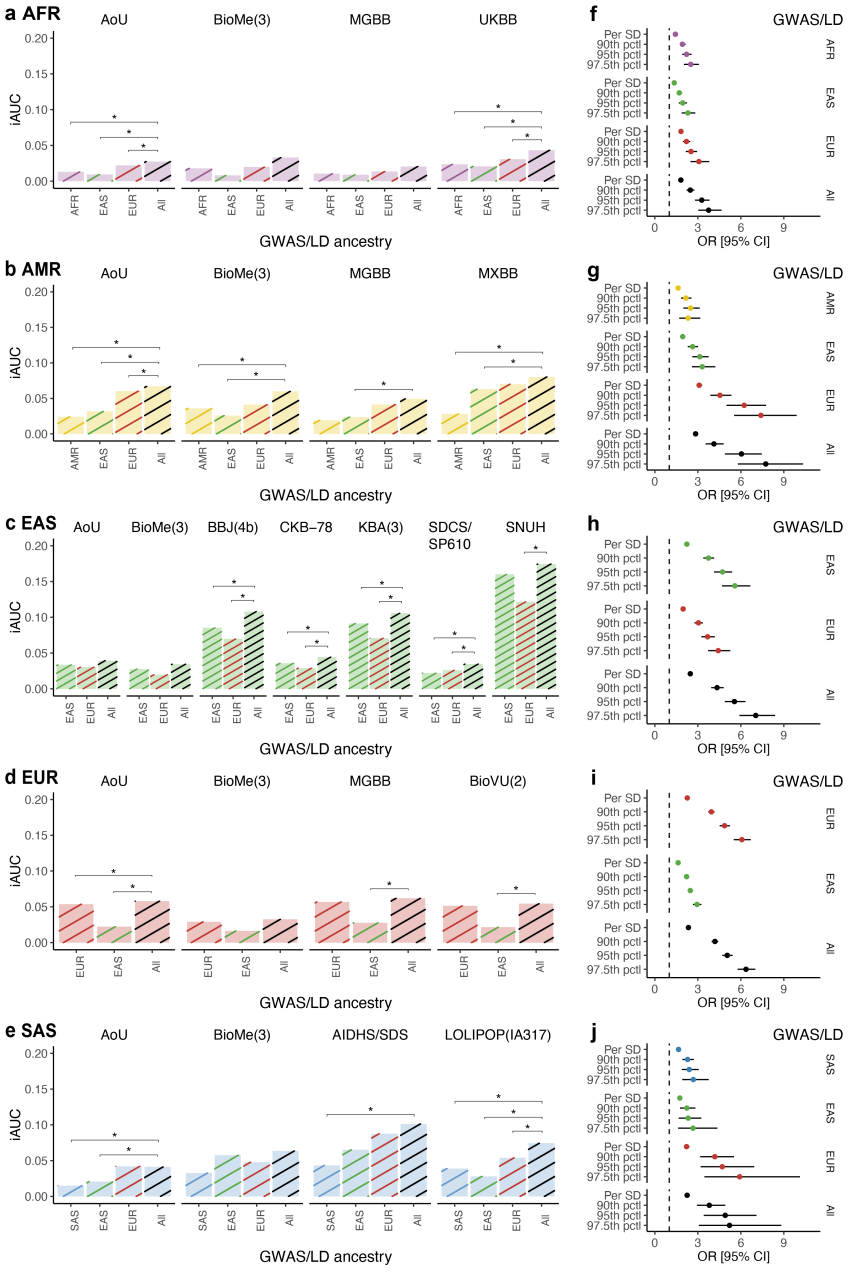
Supplementary Figures



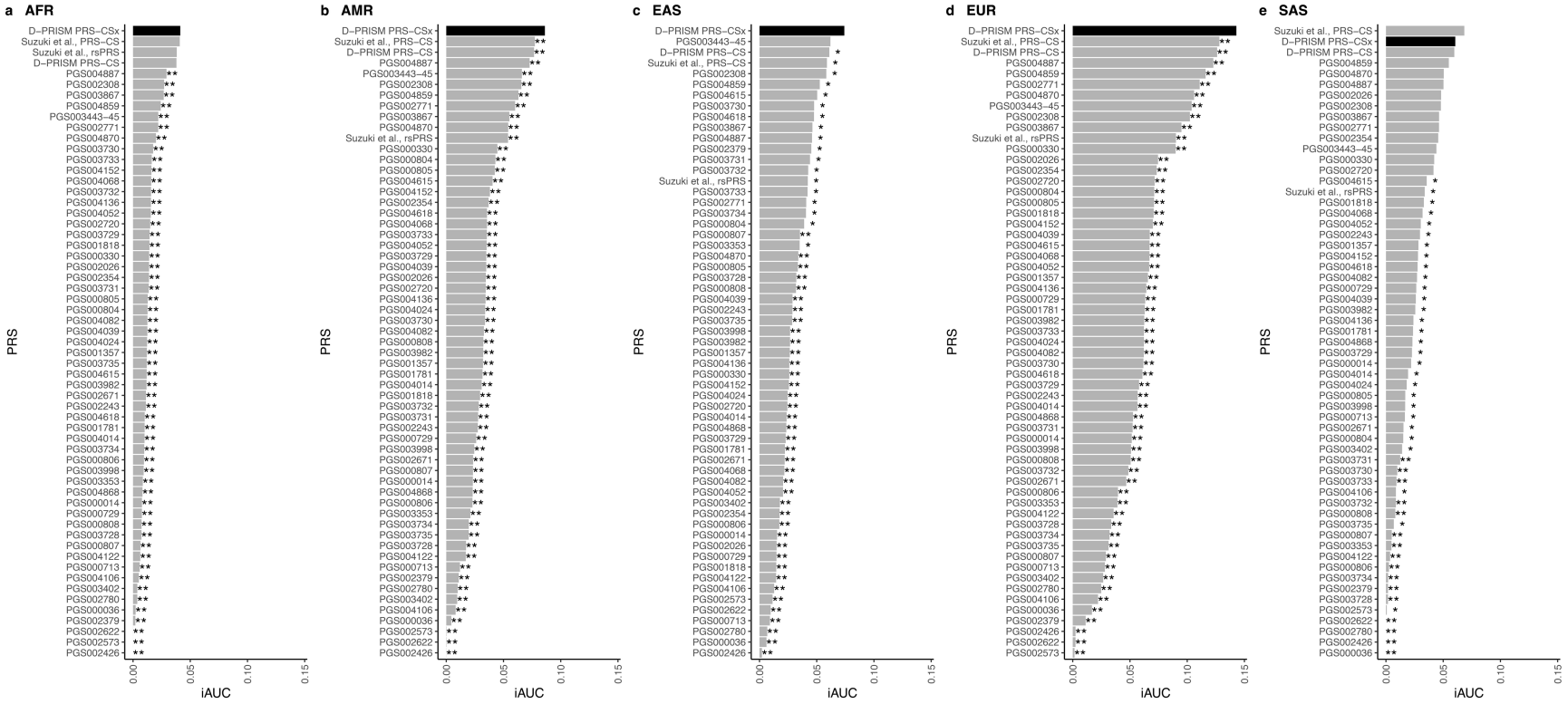
Supplementary Figure 1 | Tag SNPs informativeness across ancestries. We considered two sets of reference SNPs: one was based on the HapMap3 (HM3) set of variants, and the other was selected using the TagIt program. **a,** Number of tag SNPs in the LD reference panels, stratified by minor allele frequency and ancestry (*left*, HM3-based; *right*, TagIt-based), **b,** Number of SNPs being tagged in the LD reference panels, stratified by minor allele frequency and ancestry for a minimum pairwise correlation threshold of r^2>0.8 (*left*, HM3-based; *right*, TagIt-based), **c,** Proportion of SNPs that are either tags or are tagged (r^2>0.8) in the LD reference panels, stratified by minor allele frequency and ancestry in the individuals from 1KG. The dashed line represents 80% coverage.



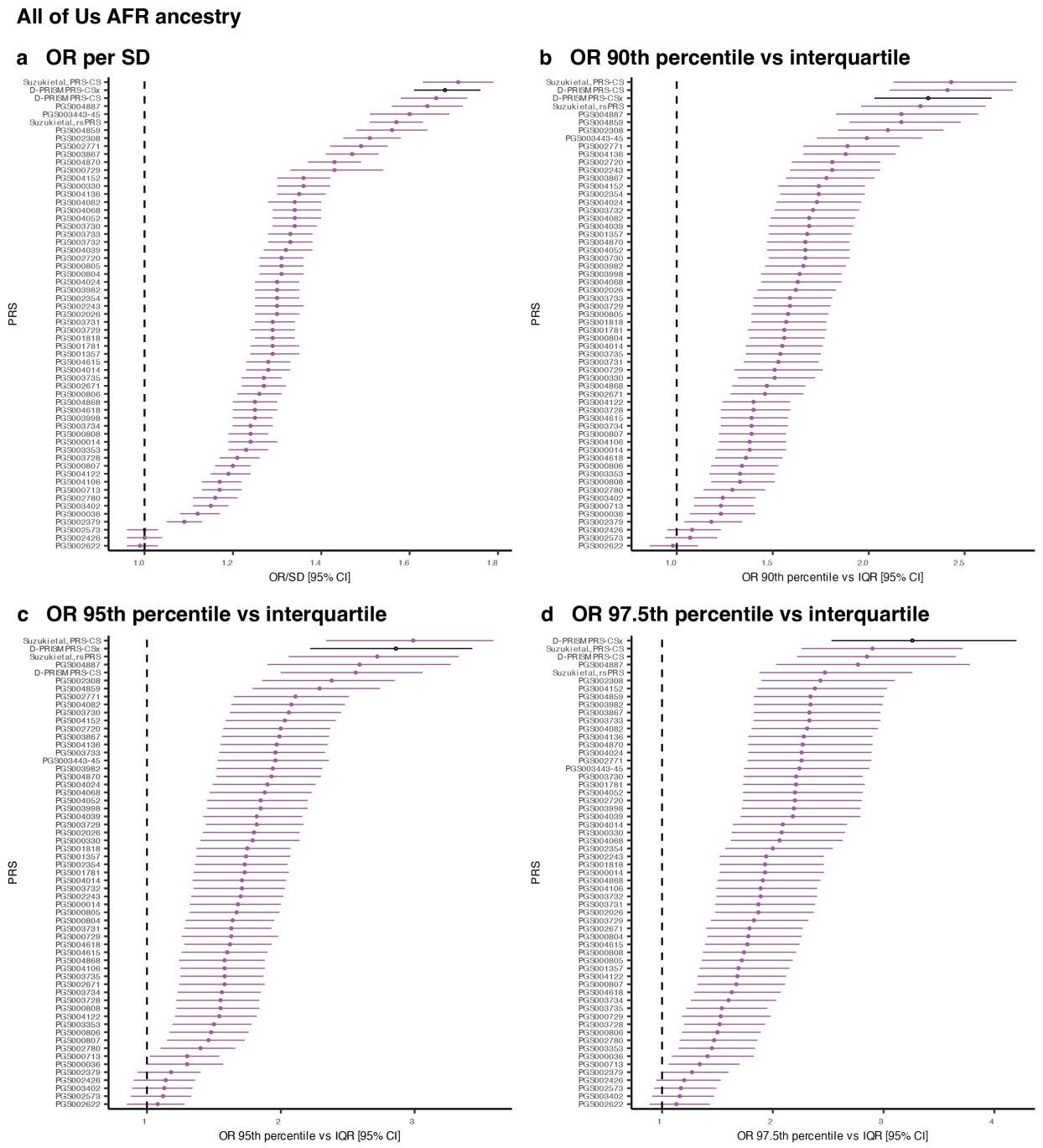
Supplementary Figure 2 | Performance of the T2D PRSs in the training cohorts across ancestry groups. Incremental AUC (iAUC) of the T2D PRS in the training cohorts: **a**, AFR, **b,** AMR, **c,** EAS, **d,** EUR, **e,** SAS. For each ancestry, we trained single-ancestry and multi-ancestry (ALL) PRSs using four LD panels and 5 phi continuous shrinkage priors. Bar colors represent the ancestry group: purple for AFR, yellow for AMR, green for EAS, red for EUR, and blue for SAS. The grey color highlights the best-trained PRS models that maximize the iAUC. All models were adjusted for sex, age, and genetic principal components.



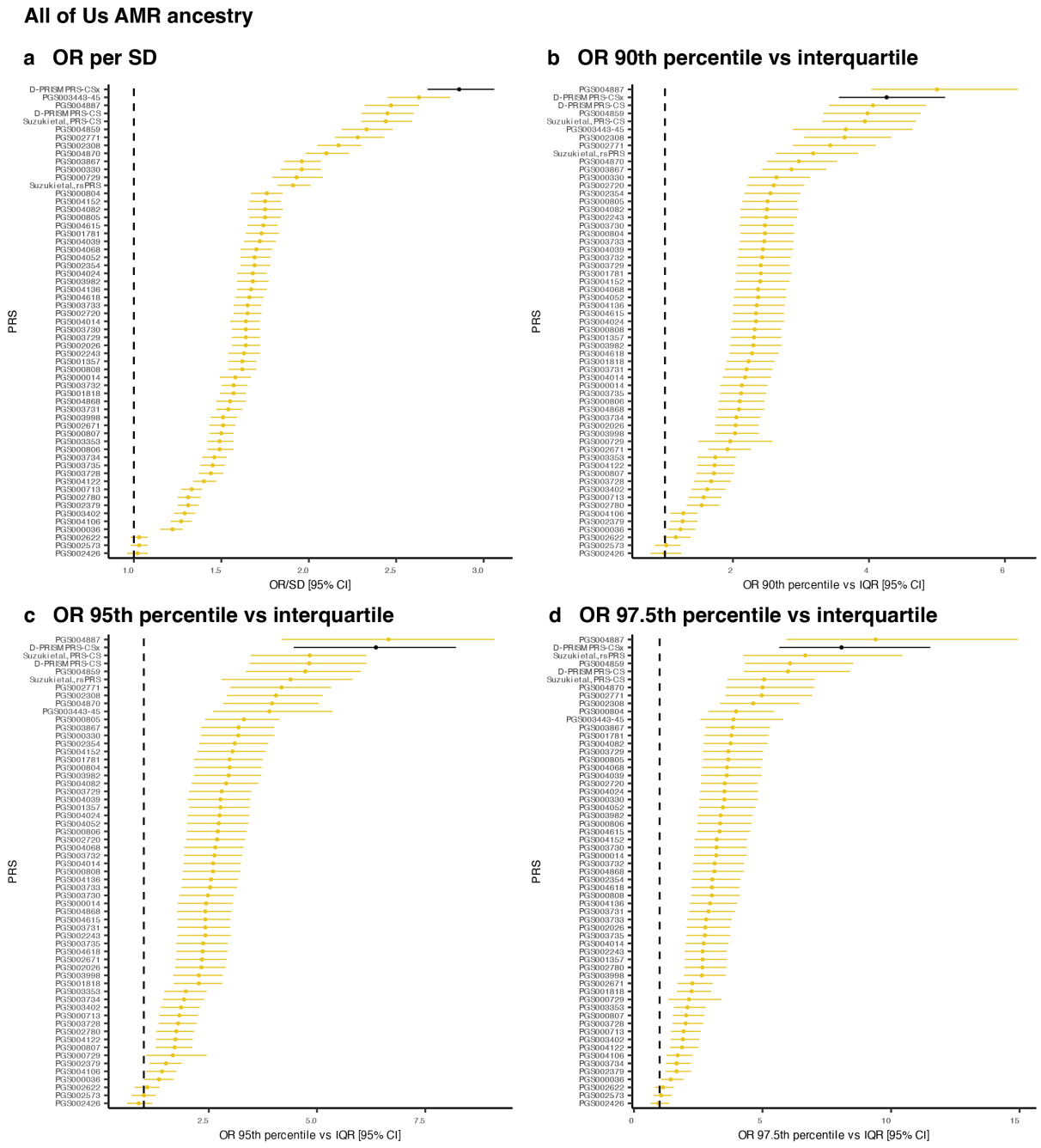
Supplementary Figure 3 | Performance of the T2D PRSs adjusted for BMI in the validation cohorts across ancestry groups. a-e: Incremental AUC (iAUC) of the T2D PRS in the validation cohorts across ancestry groups: **a,** AFR, **b,** AMR, **c,** EAS, **d,** EUR, **e,** SAS. For each ancestry, the best-performing single-ancestry and multi-ancestry (All) PRSs were evaluated. Each bar represents a single cohort. Bar colors represent the ancestry group: purple for AFR, yellow for AMR, green for EAS, red for EUR, and blue for SAS. Line colors represent the ancestry of the T2D GWAS summary statistics and LD panels used to train the PRS, using the same color codes for single-ancestry PRSs, and black for multi-ancestry PRSs. **f-j:** Odds ratio (OR) from the meta-analysis of validation cohorts across ancestry groups: **f,** AFR, **g,** AMR, **h,** EAS, **i,** EUR, **j,** SAS. Points represent the odds ratio per standard deviation of the PRS distribution or the odds ratio comparing different PRS distribution extremes relative to the interquartile range. Error bars show the 95% confidence intervals (95% CI). Point colors represent the ancestry of the T2D GWAS summary statistics and LD panels used to train the PRS. * De Long $p<0.05$. All models were adjusted for sex, age, and genetic principal components



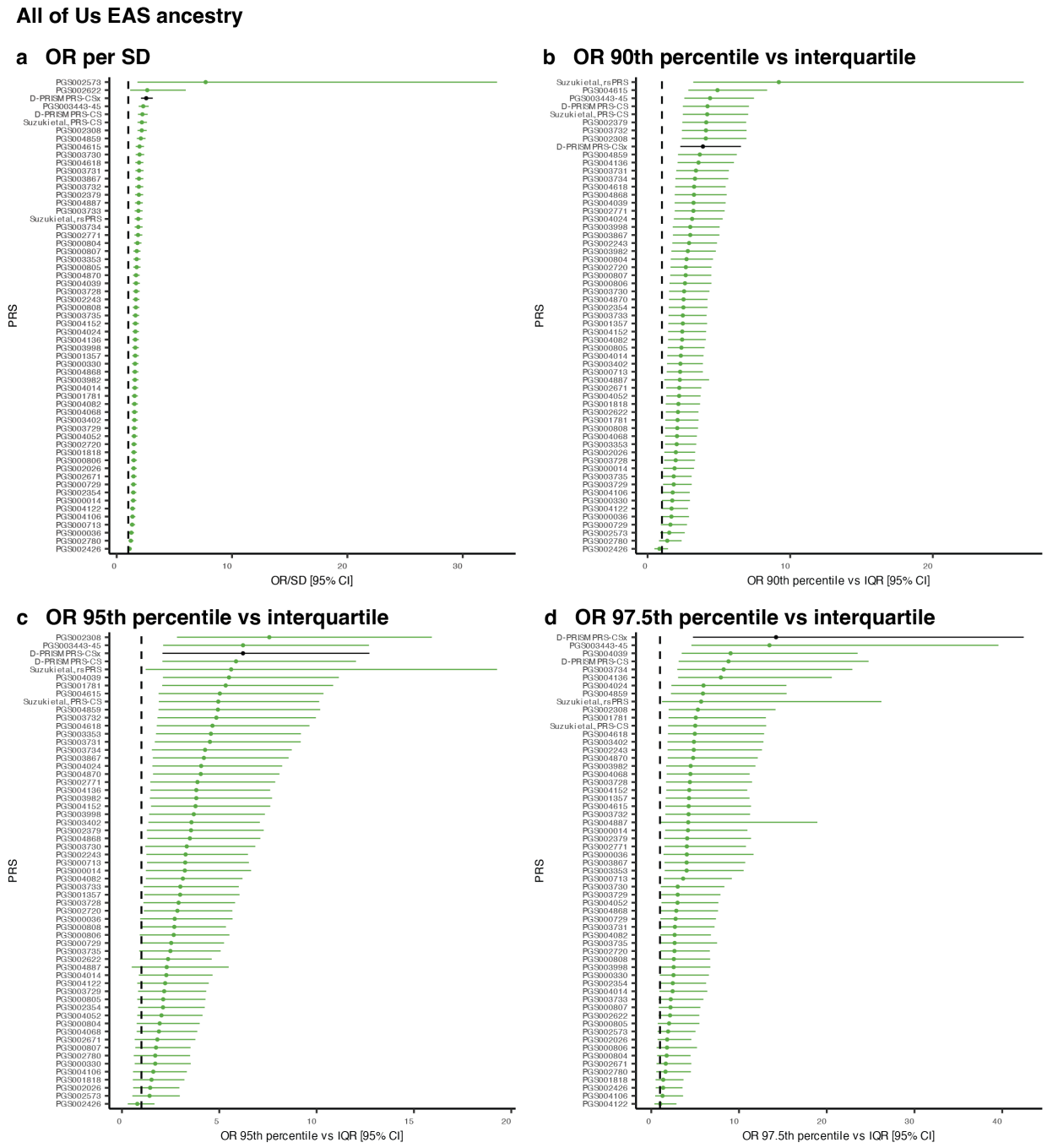
Supplementary Figure 4 | Performance of D-PRISM multi-ancestry PRSs compared to published T2D PRSs from the PGS Catalog and other sources in the All of Us cohort. a-e: Incremental AUC (iAUC) across ancestry groups: **a,** AFR, **b,** AMR, **c,** EAS, **d,** EUR, **e,** SAS. Black bars highlight this study's multi-ancestry PRSs. *DeLong $p<0.05$; **Bonferroni-corrected DeLong $p<9\times10^{-4}$. All models were adjusted for sex, age, and genetic principal components.



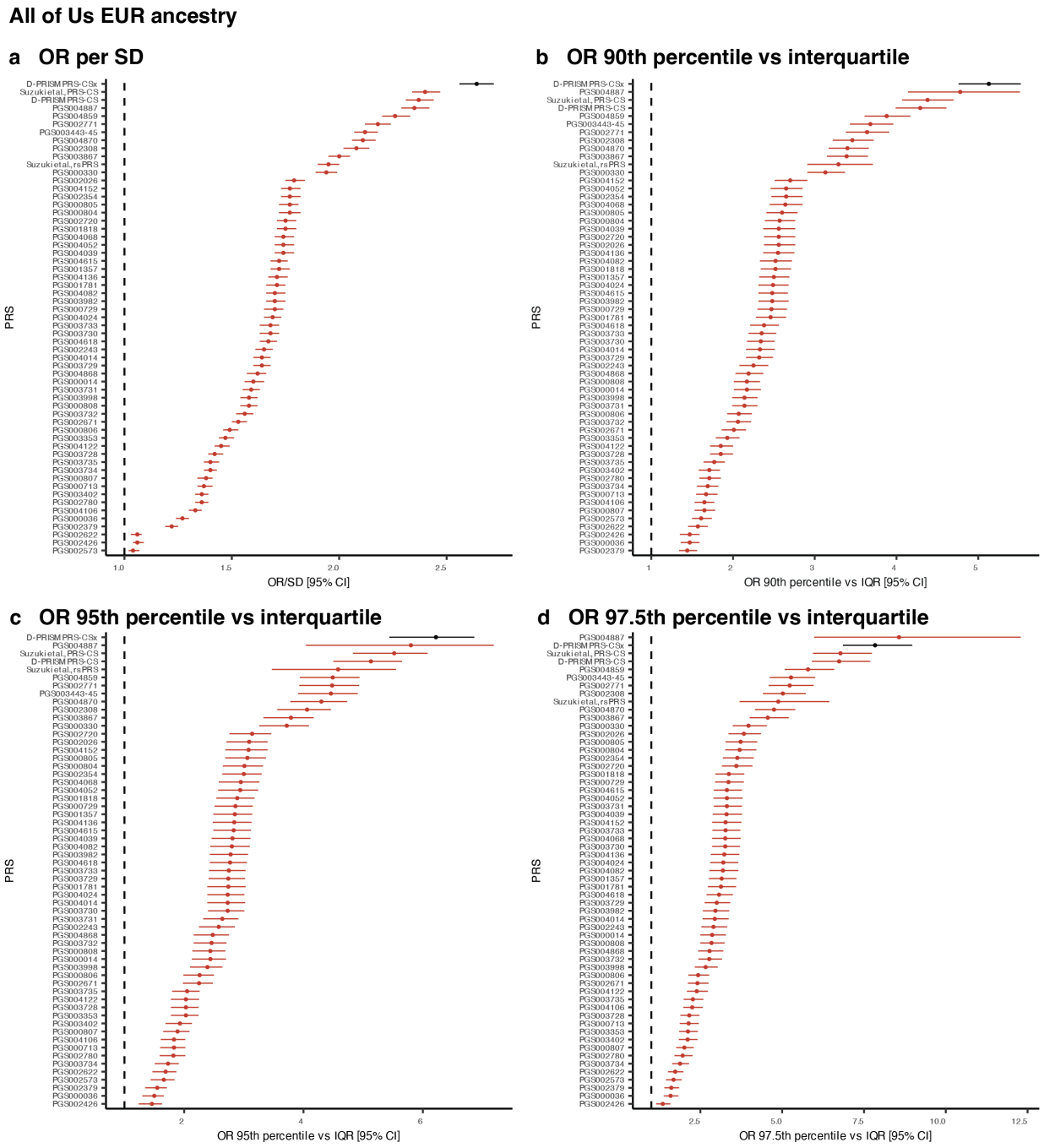
Supplementary Figure 5 | Performance of D-PRISM multi-ancestry PRS-CSx compared to the published T2D PRSs from the PGS Catalog and others in individuals of AFR ancestry from the All of Us validation cohort. a, Odds ratio per standard deviation (OR per SD) of the PRS distribution, **b,** OR comparing the 90th percentile of the PRS relative to the interquartile range, **c,** OR comparing the 95th percentile of the PRS relative to the interquartile range, **d,** OR comparing the 97.5th percentile of the PRS relative to the interquartile range. All models were adjusted for sex, age, and genetic principal components.



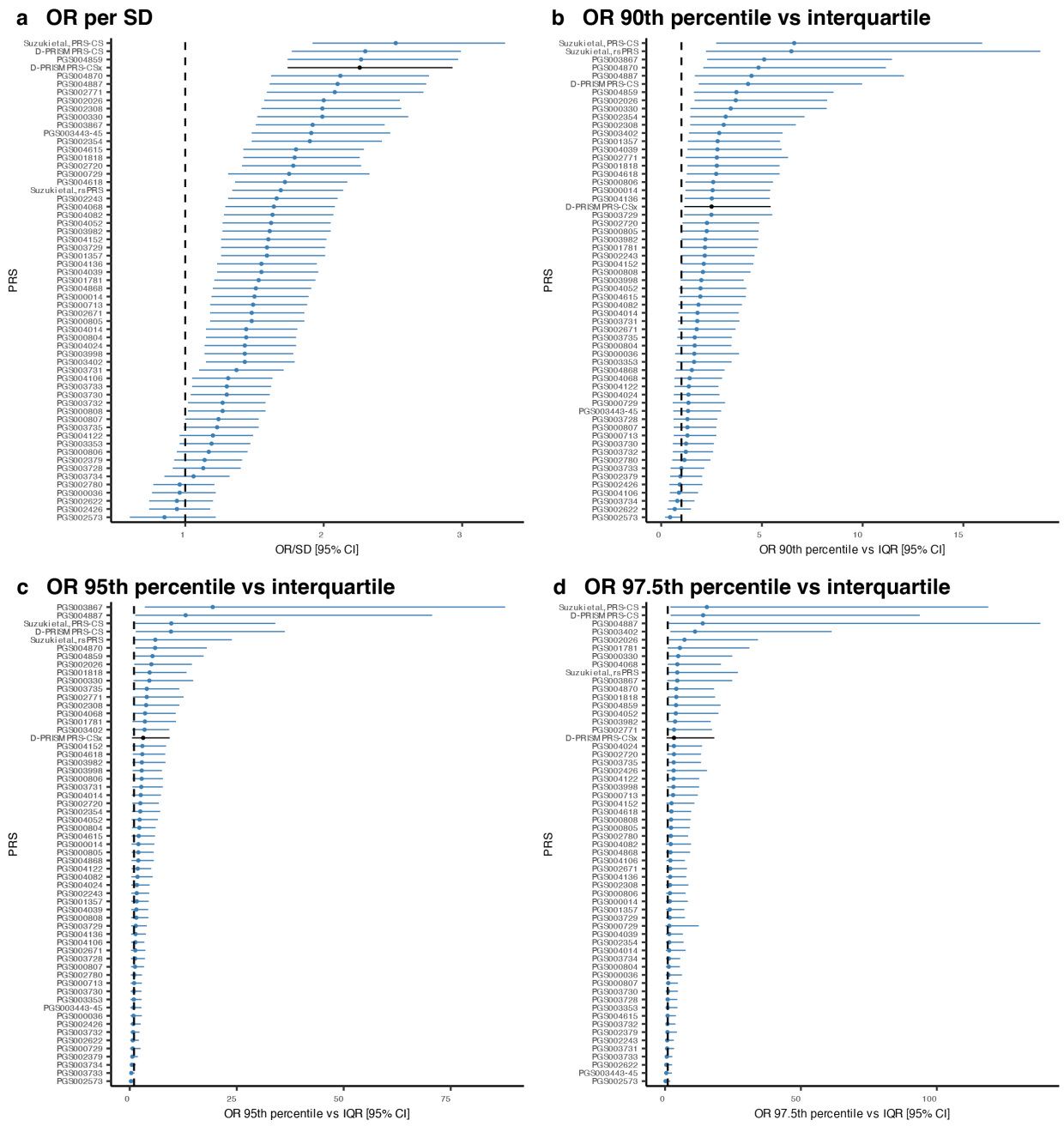
Supplementary Figure 6 | Performance of D-PRISM multi-ancestry PRS-CSx compared to the published T2D PRSs from the PGS Catalog and others in individuals of AMR ancestry from the All of Us validation cohort. a, Odds ratio per standard deviation (OR per SD) of the PRS distribution, **b,** OR comparing the 90th percentile of the PRS relative to the interquartile range, **c,** OR comparing the 95th percentile of the PRS relative to the interquartile range, **d,** OR comparing the 97.5th percentile of the PRS relative to the interquartile range. All models were adjusted for sex, age, and genetic principal components.



Supplementary Figure 7 | Performance of D-PRISM multi-ancestry PRS-CSx compared to the published T2D PRSs from the PGS Catalog and others in individuals of EAS ancestry from the All of Us validation cohort. a, Odds ratio per standard deviation (OR per SD) of the PRS distribution, **b,** OR comparing the 90th percentile of the PRS relative to the interquartile range, **c,** OR comparing the 95th percentile of the PRS relative to the interquartile range, **d,** OR comparing the 97.5th percentile of the PRS relative to the interquartile range. All models were adjusted for sex, age, and genetic principal components.

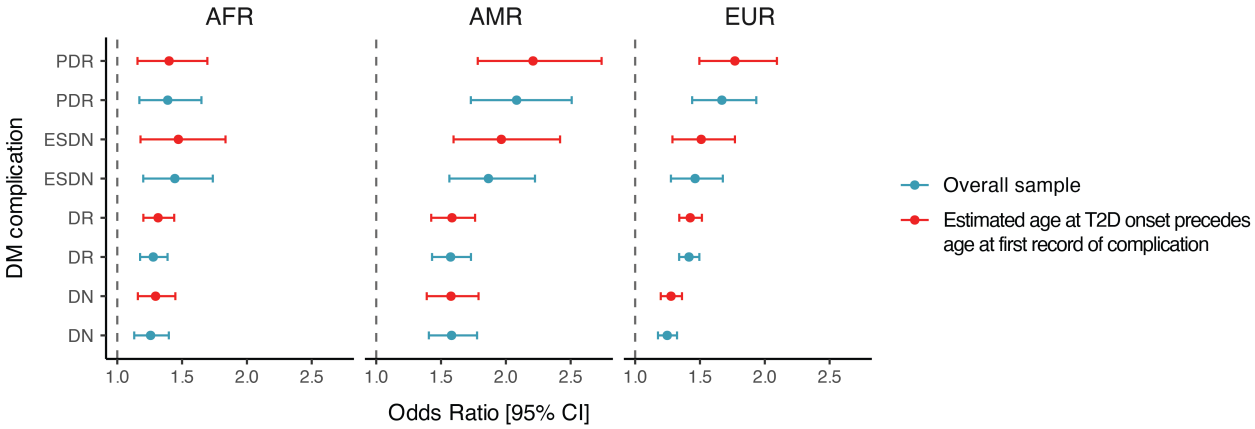


Supplementary Figure 8 | Performance of D-PRISM multi-ancestry PRS-CSx compared to the published T2D PRSs from the PGS Catalog and others in individuals of EUR ancestry from the All of Us validation cohort. a, Odds ratio per standard deviation (OR per SD) of the PRS distribution, **b,** OR comparing the 90th percentile of the PRS relative to the interquartile range, **c,** OR comparing the 95th percentile of the PRS relative to the interquartile range, **d,** OR comparing the 97.5th percentile of the PRS relative to the interquartile range. All models were adjusted for sex, age, and genetic principal components.

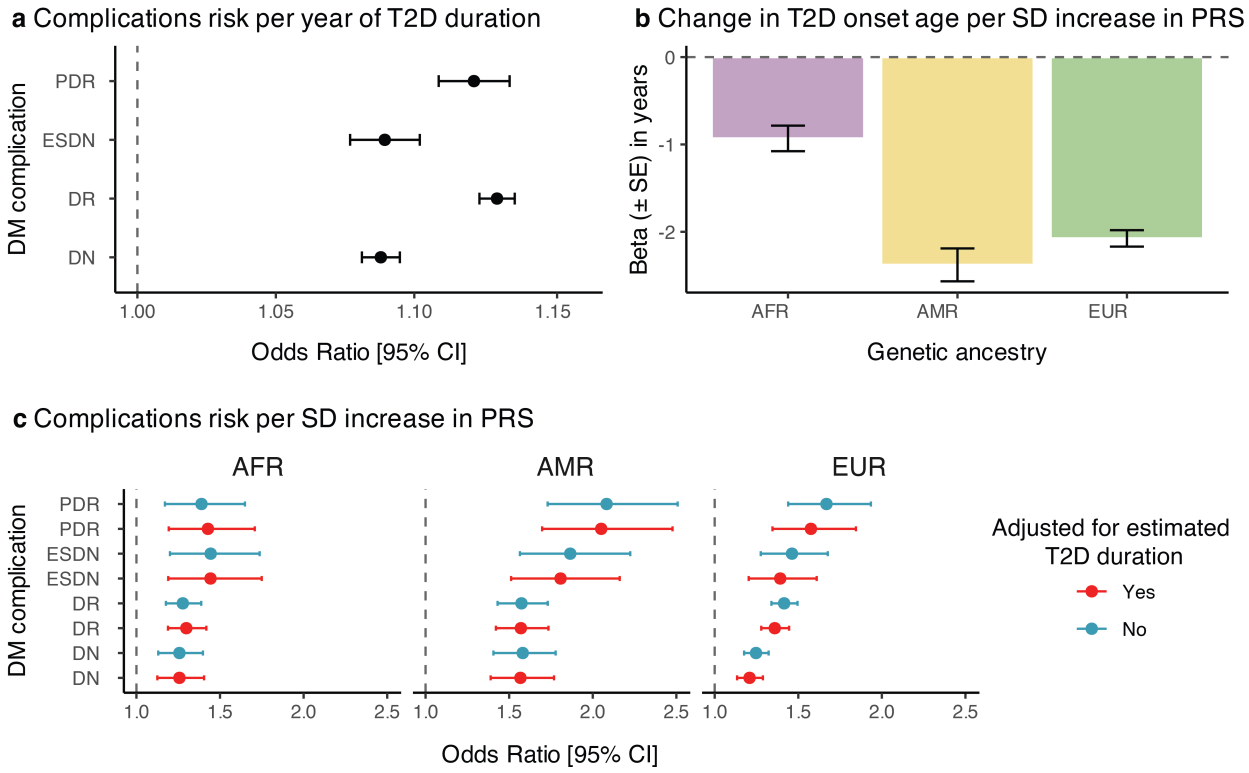


Supplementary Figure 9 | Performance of D-PRISM multi-ancestry PRS-CSx compared to the published T2D PRSs from the PGS Catalog and others in individuals of SAS ancestry from the All of Us validation cohort. a, Odds ratio per standard deviation (OR per SD) of the PRS distribution, **b,** OR comparing the 90th percentile of the PRS relative to the interquartile range, **c,** OR comparing the 95th percentile of the PRS relative to the interquartile range, **d,** OR comparing the 97.5th percentile of the PRS relative to the interquartile range. All models were adjusted for sex, age, and genetic principal components.

Temporal sequence of estimated age at diabetes onset and first record of the microvascular complication based on the All of Us v8 EHR and survey data.				
	% of T2D cases with microvascular complications			
	PDR	ESDN	DR	DN
Estimated age at T2D onset precedes age at first record of complication	77%	72%	79%	85%
Age at first record of complication precedes estimated age at T2D onset	18%	22%	14%	7%
Unavailable estimated age at T2D onset	5%	6%	7%	8%



Supplementary Figure 10 | Sensitivity analysis of the association of D-PRISM Multi-Ancestry Polygenic Risk Score (PRS) with Diabetes Microvascular Complications in the All of Us Research Program Cohort. The plot displays the Odds Ratio (OR) per standard deviation increase in the D-PRISM PRS for four outcomes: proliferative diabetic retinopathy (PDR), end-stage diabetic nephropathy (ESDN), diabetic retinopathy (DR), and diabetic neuropathy (DN). Points are colored to indicate the sample used: red for individuals for whom the estimated diabetes onset preceded the complication EHR first record, and blue for the overall sample. Error bars indicate the 95% confidence intervals (95% CI). All models were adjusted for sex, age, and genetic principal components.



Supplementary Figure 11 | Effect of PRS and T2D duration on risk of microvascular complications. a, Odds Ratio per year of EHR-based estimated T2D duration for Proliferative Diabetic Retinopathy (PDR), End-Stage Diabetic Nephropathy (ESDN), Diabetic Retinopathy (DR), and Diabetic Nephropathy (DN). The plot shows a significant positive association between microvascular complications and T2D duration, with up to a 1.12-fold increased risk per year. All models were restricted to T2D cases and adjusted for sex and genetic principal components. **b,** Change in T2D onset age (years) per one-SD increase in PRS, stratified by genetic ancestry. One-SD increase in the T2D PRS was associated with up to 2.3 years earlier onset of T2D across genetic ancestries. Models were adjusted for sex and genetic principal components. Bars represent the beta coefficient (change in years), and error bars show the 95% confidence intervals (95% CI). **c,** Odds Ratio per one-SD increase in PRS for PDR, ESDN, DR, and DN. All models were adjusted for sex, age, and genetic principal components. Points are colored based on statistical adjustment for T2D duration: red (adjusted for T2D duration) and blue (unadjusted).

Full list of D-PRISM authors

Alicia Huerta-Chagoya *[1,2,3], Joohyun Kim *[4], Ravi Mandla ‡[1,2,3,5], Yingchang Lu ‡[4], Ken Suzuki ‡[6,7,8], Lauren E. Petty ‡[4], Hong Kiat Ng ‡[9], Jaewon Choi ‡[10], Simon Lee ‡[11], Madhusmita Rout ‡[12], Kuang Lin ‡[13], Katherine Taylor ‡[1], Linda S. Adair[14], Adebowale Adeyemo[15], Habibul Ahsan[16], Masato Akiyama[17,18], Ping An[19], Sonia S. Anand[20,21,22], Diane M. Becker[23], Alain G. Bertoni[24], Zheng Bian[25], Lawrence F. Bielak[26], John Blangero[27], Michael Boehnke[28], Erwin P. Bottinger[29,30,11], Donald W. Bowden[31,32,33], Fiona Bragg[13,34], Jennifer A. Brody[35], Thomas A. Buchanan[36], Brian E. Cade[37,38], Jin-Fang Chai[39], John C. Chambers[9], Giriraj R. Chandak[40,41], Li-Ching Chang[42], Kyong-Mi Chang[43,44], Miao-Li Chee[45], Chien-Hsiun Chen[42], Yuan-Tsong Chen[42], Zhengming Chen[13], Yii-Der I. Chen[46], Ji Chen[47], Guanjie Chen[15], Shyh-Huei Chen[48], Wei-Min Chen[49], Ching-Yu Cheng[50,51,45,52], Yoon Shin Cho[53], Hyeok Sun Choi[53], Lee-Ming Chuang[54,55], Miguel Cruz[56], Mary Cushman[57], Swapan K. Das[58], Ralph A. DeFronzo[59], H Janaka deSilva[60], Latchezar Dimitrov[32], Ayo P. Doumatey[15], Shufa Du[14], Qing Duan[61], Ravindranath Duggirala[27], Leslie S. Emery[62], James C. Engert[63], Daniel S. Evans[64], Michele K. Evans[65], Sarah Finer[66], Jose C. Florez[1,67,68], James S. Floyd[35], Myriam Fornage[69], Elizabeth G. Frankel[4], Barry I. Freedman[70], Pauline Genter[71], Hertzel C. Gerstein[20,21,22], Mark O. Goodarzi[72], Penny Gordon-Larsen[14], Mariaelisa Graff[73], Myron Gross[74], Yu Guo[25], Xiuqing Guo[46], Yang Hai[46], Craig L. Hanis[75], MGeoffrey Hayes[76,77,78], Momoko Horikoshi[79], Annie-Green Howard[80], Sarah Hsu[1,3], Willa Hsueh[81], Wei Huang[82], Mengna Huang[83,84], Yi-Jen Hung[85,86], Mi Yeong Hwang[87], Chii-Min Hwu[88,89], Sahoko Ichihara[90], Michiya Igase[91], Eli Ipp[71], Mohammad T. Islam[92], Masato Isono[93], Hye-Mi Jang[87], Farzana Jasmine[16], Jost B. Jonas[94,95,96,97,98], Yoonjung Y. Joo[99,78], Edmond Kabagambe[100], Takashi Kadowaki[7,101], Yoichiro Kamatani[102,103], Fouad R. Kandeel[104], Sharon L.R. Kardia[26], Elizabeth W. Karlson[105], Anuradhani Kasturiratne[106], Norihiro Kato[93], Tomohiro Katsuya[107,108], Varinderpal Kaur[1,2,3], Takahisa Kawaguchi[109], Jacob M. Keaton[32,100], Abel N. Kho[110,111], Chiea-Chuen Khor[112,45,113], Muhammad Kibriya[16], Bong-Jo Kim[87], Woon-Puay Koh[114,39], Katsuhiko Kohara[115], Jaspal S. Kooner[116], Charles Kooperberg[117], Raymond J. Kreienkamp[1,67,2,118], Amel Lamri[20,21], Leslie A. Lange[119], Nanette R. Lee[120], Myung-Shik Lee[121,122], Jung-Jin Lee[123], Donna M. Lehman[59], Liming Li[124], Yun Li[125], Victor JY. Lim[39], Jianjun Liu[112,126], Yongmei Liu[127,24], Simin Liu[128,129,130,131], Jirong Long[100], Tin Louie[62], Xi Luo[132], Jun Lv[124], Julie A. Lynch[133,134], Shiro Maeda[79,135,136], Anubha Mahajan[137], Nisa M. Maruthur[138], Fumihiko Matsuda[109], Mark I. McCarthy[137,139,140], Roberta McKean-Cowdin[141], James B. Meigs[1,67,142], Iona Y. Millwood[13], Karen L. Mohlke[61], Ayesha A. Motala[143], Girish N. Nadkarni[11,29], Jerry L. Nadler[144], Masahiro Nakatochi[145], Mike A. Nalls[146,147,148], Uma Nayak[149], Aude Nicolas[146], Kari E. North[73], Darryl Nousome[141], Yukinori Okada[150,151,152], Sumit Paliwal[153], Ian Pan[84], James S. Pankow[154], Guillaume Paré[155,21,22], Jaehyun Park[4], Kyong Soo Park[156,157], Esteban J. Parra[158], Sanjay R. Patel[159], Mark A. Pereira[154], Patricia A. Peyser[26], Fraser J. Pirie[143], Michael Preuss[11], Michael A. Province[19], Bruce M. Psaty[160], PSKDB PunyaSri[161], Leslie J. Raffel[162], Laura M. Raffield[61], Laura J. Rasmussen-Torvik[163], Susan Redline[164,165,38], Alexander P. Reiner[166,167], Stephen S. Rich[168], Rebecca Rohde[73], Kathryn Roll[46], Rashedeh Roshani[4], Charles N. Rotimi[15], Charumathi Sabanayagam[45,50], Danish Saleheen[169], Kevin Sandow[46], Claudia Schurmann[29,30,11], Mohammad Shahriar[16], Douglas M. Shaw[4], Wayne H-H. Sheu[86,88,89,170,171], Jinxiu Shi[82], Xiao-Ou Shu[100], Megan M. Shuey[4], Moneeza K. Siddiqui[66], Jennifer A. Smith[26,172], Tamar Sofer[173,174,67], Cassandra N. Spracklen[61,175], Adrienne M. Stilp[62], Meng Sun[176], Yasuharu Tabara[109], E-Shyong Tai[126,39,113], Salman M. Tajuddin[65], Atsushi Takahashi[102,177], Fumihiko Takeuchi[93], Jingyi Tan[46], Kent D. Taylor[46], Farook Thameem[178], Lin Tong[16], Fuu-Jen Tsai[179], Philip S. Tsao[180,181], Miriam S. Udler[1,67,2,3], Adan Valladares-Salgado[56], David A. van Heel[182], Rob M. vanDam[39,126], Rohit Varma[183], Maheak Vora[1,2], Niels Wacher-Rodarte[184], Ya-Xing Wang[185], Ellie Wheeler[186,47], Eric A. Whitsel[73,187], Ananda R. Wickremasinghe[106], Genevieve L. Wojcik[188], Tien Y. Wong[45,50,51], Jer-Yuarn Wu[42], Yong-Bing Xiang[189,190], Anny H. Xiang[191], Chittaranjan S. Yajnik[192], Ken Yamamoto[193], Toshimasa Yamauchi[7], Lisa R. Yanek[23], Jie Yao[46], Mitsuhiro Yokota[193], Canqing Yu[124], Jian-Min Yuan[194,195], Salim Yusuf[196,197,198], Eleftheria Zeggini[199,200], Liang Zhang[45], Weihua Zhang[201], Wei Zheng[202], Alan B. Zonderman[65], ENSA Genomics Consortium[203], Genes & Health Research Team[182], VA Million Veteran Program[204], Carlos A. Aguilar-Salinas[205,206], Lourdes García-García[207], Clicerio González-Villalpando[208], Christopher A. Haiman[209], Young Jin Kim[87], Soo Heon Kwak[156], Aaron Leong[1,2,210,3,142,211], Ruth J.F. Loos[11,212], Andres Moreno-Estrada[213], Andrew P. Morris[8,214], Lorena Orozco[215], Jerome I. Rotter[46], Dharambir Sanghera[12], Teresa Tusie-Luna[216], Benjamin F. Voight[43,5,217,218], Marijana Vujkovic[219,43], Robin G. Walters[13], Tian Ge ^[220], Alisa K. Manning ^[221,1,67], Marie Loh ^[9,222,223], Jennifer E. Below ^[4], Xueling Sim ^[39], Josep M. Mercader ^ #[1,67,2,3,224], Maggie C.Y. Ng ^ #[4], D-PRISM Consortium[225]

[1]Programs in Metabolism and Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, 2142, USA, [2]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, 2114, USA, [3]Diabetes Unit, Massachusetts General Hospital, Boston, MA, 2114, USA, [4]Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA, [5]Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA, [6]Department of Statistical Genetics, Osaka

University, Osaka, 565-0871, Japan, [7]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, 113-8654, Japan, [8]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Division of Musculoskeletal and Dermatological Sciences, University of Manchester, Manchester, UK, [9]Lee Kong Chian School of Medicine, Clinical Sciences Building, Nanyang Technological University, Singapore, 308232, Singapore, [10]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, 3080, South Korea, [11]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA, [12]Department of Pediatrics, Section of Genetics, University of Oklahoma, Oklahoma, OK, USA, [13]Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK, [14]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA, [15]Center for Research on Genomics and Global Health, National Human Genome Institute, National Institutes of Health, Bethesda, MD, USA, [16]Biological Sciences Division, Public Health Sciences, The University of Chicago, Chicago, IL, 60637, USA, [17]Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences, Yokohama, 230-0045, Japan, [18]Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, 812-8582, Japan, [19]Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO, USA, [20]Department of Medicine, McMaster University, Hamilton, ON, Canada, [21]Population Health Research Institute, Hamilton, ON, Canada, [22]Hamilton Health Sciences, Hamilton, ON, Canada, [23]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA, [24]Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA, [25]Chinese Academy of Medical Sciences, Beijing, 100730, China, [26]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA, [27]Human Genetics and STDOI, University of Texas Rio Grande Valley, Edinburg, TX, USA, [28]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA, [29]Digital Health Center, Digitl Engineering Faculty of Hasso Plattner Institue and University Potsdam, Potsdam, Germany, [30]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, USA, [31]Center for Diabetes Research, Wake Forest University School of Medicine, Winston-Salem, NC, USA, [32]Center for Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston-Salem, NC, USA, [33]Department of Biochemistry, Wake Forest University School of Medicine, Winston-Salem, NC, USA, [34]Health Data Research UK Oxford, University of Oxford, Oxford, UK, [35]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA, [36]Department of Medicine, Division of Endocrinology and Diabetes, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA, [37]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA, [38]Harvard Medical School, Boston, MA, USA, [39]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, 117549, Singapore, [40]Genomic Research on Complex diseases, CSIR-Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad, India, [41]Lightening Lives, Hyderabad, India, [42]Institute of Biomedical Sciences, Academia Sinica, Taipei, 11529, Taiwan, [43]Corporal Michael J. Crescenz Philadelphia VA Medical Center, Philadelphia, PA, USA, [44]Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104, USA, [45]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, 168751, Singapore, [46]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, 90502, USA, [47]Wellcome Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK, [48]Department of Biostatistics and Data Science, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA, [49]Department of Public Health Sciences and Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA, USA, [50]Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, 169857, Singapore, [51]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, 119228, Singapore, [52]Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, 119228, Singapore, [53]Biomedical Science, Hallym University, Chuncheon, 24252, South Korea, [54]Department of Internal Medicine, Division of Endocrinology and Metabolism, National Taiwan University Hospital, Taipei, Taiwan, [55]College of Medicine, National Taiwan University, Taipei, Taiwan, [56]Unidad de Investigación Médica en Bioquimica, Hospital de Especialidades, Centro Medico Nacional Siglo XXI, IMSS, Mexico City, Mexico, [57]Department of Medicine, University of Vermont, Colchester, VT, USA, [58]Department of Internal Medicine, Section on Endocrinology and Metabolism, Wake Forest University School of Medicine, Winston-Salem, NC, USA, [59]Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA, [60]Department of Medicine, Faculty of Medicine, University of Kelaniya, Sri Lanka, [61]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA, [62]Department of Biostatistics, University of Washington, Seattle, WA, USA, [63]Department of Medicine, McGill University, Montreal, QC, Canada, [64]California

Pacific Medical Center Research Institute, San Francisco, CA, USA, [65]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA, [66]Wolfson Institute of Population Health, Queen Mary University of London, London, UK, [67]Department of Medicine, Harvard Medical School, Boston, MA, 2215, USA, [68]Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, 2114, USA, [69]Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA, [70]Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA, [71]Division of Endocrinology & Metabolism, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, 90502, USA, [72]Department of Medicine, Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, USA, [73]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [74]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, 55454, USA, [75]Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA, [76]Department of Anthropology, Northwestern University, Evanston, IL, USA, [77]Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, [78]Division of Endocrinology, Metabolism, and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, [79]Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Centre for Integrative Medical Sciences, Yokohama, 230-0045, Japan, [80]Department of Biostatistics, Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA, [81]Department of Internal Medicine, Diabetes and Metabolism Research Center, Ohio State University Wexner Medical Center, Columbus, OH, USA, [82]Shanghai-MOST Key Laboratory of Health and Disease Genomics, Shanghai Institute for Biomedical and Pharmaceutical Technologies (SIBPT), Shanghai, 200032, China, [83]Center for Global Cardiometabolic Health, Brown University, Providence, RI, USA, [84]Department of Epidemiology, Brown University School of Public Health, Providence, RI, USA, [85]Division of Endocrine and Metabolism, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan, [86]School of Medicine, National Defense Medical Center, Taipei, Taiwan, [87]Department of Precision Medicine, Division of Genome Science, National Institute of Health, Chungcheongbuk-do, 28159, Republic of Korea, [88]Department of Medicine, Section of Endocrinology and Metabolism, Taipei Veterans General Hospital, Taipei, Taiwan, [89]School of Medicine, National Yang-Ming University, Taipei, Taiwan, [90]Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, 329-0498, Japan, [91]Department of Anti-aging Medicine, Ehime University Graduate School of Medicine, Ehime, 791-0295, Japan, [92]U Chicago Research, Uttara, Dhaka, 1230, Bangladesh, [93]Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, 162-8655, Japan, [94]Rothschild Foundation Hospital, Paris, France, [95]Singapore Eye Research Institute, Singapore National Eye Center, Singapore, [96]Privatpraxis Prof Jonas und Dr. Panda-Jonas, Heidelberg, Germany, [97]Beijing Visual Science and Translational Eye Research Institute (BERI), Beijing Tsinghua Changgung Hospital, Tsinghua Medicine, Tsinghua University, Beijing, China, [98]L V Prasad Eye Institute, Hyderabad, Telangana, India, [99]Department of Health and Biomedical Informatics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, [100]Department of Medicine, Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN, USA, [101]Toranomon Hospital, Tokyo, 105-8470, Japan, [102]Laboratory for Statistical and Translational Genetics, RIKEN Centre for Integrative Medical Sciences, Yokohama, 230-0045, Japan, [103]Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, 108-8639, Japan, [104]Departments of Clinical Diabetes, Endocrinology & Metabolism and Translational Research and Cellular Therapeutics, City of Hope, Duarte, CA, USA, [105]Department of Medicine, Brigham and Women's Hospital, Boston, MA, 2215, USA, [106]Department of Public Health, Faculty of Medicine, University of Kelaniya, Sri Lanka, [107]Department of Clinical Gene Therapy, Osaka University, Osaka, 565-0871, Japan, [108]Department of Geriatric and General Medicine, Graduate School of Medicine, Osaka University, Osaka, 565-0871, Japan, [109]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, 606-8507, Japan, [110]Center for Health Information Partnerships, Institute for Public Health and Medicine, Northwestern UniversityFeinberg School of Medicine, Chicago, IL, USA, [111]Department of Medicine, Division of General Internal Medicine and Geriatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, [112]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, 138672, Singapore, [113]Duke-NUS Medical School, Singapore, 169857, Singapore, [114]Health Services and Systems Research, Duke-NUS Medical School, Singapore, 169857, Singapore, [115]Department of Regional Resource Management, Ehime University Faculty of Collaborative Regional Innovation, Ehime, 790-8577, Japan, [116]Department of Cardiology, Ealing Hosptial, London North West Healthcare NHS Trust, Middlesex, UB1 3HW, UK, [117]Fred Hutchinson Cancer Center, Seattle, WA, USA, [118]Department of Pediatrics, Division of Endocrinology, Boston Children's Hospital, Boston, MA, USA, [119]Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA, [120]Department of Anthropology, Sociology and History, University of San Carlos, Cebu City, 6000, Philippines,

[121]Severance Biomedical Science Institute and Department of Internal Medicine, Yonsei University College of Medicine, Seoul, 3722, South Korea, [122]Soonchunhyang Institute of Medi-bio Science, Cheonan, 31151, South Korea, [123]Translational Medicine and Human Genetics, Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA, [124]Department of Epidemiology and Biostatistics, Peking University Health Science Centre, Beijing, 100191, China, [125]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [126]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, 117549, Singapore, [127]Department of Medicine, Division of Cardiology, Duke University School of Medicine, Durham, NC, USA, [128]Department of Epidemiology and Biostatistics, The Joe Wen School of Population & Pubic Health, UCI, Irvine, CA, USA, [129]Center for Global Cardiometabolic Health and Nutrition, the University of California, Irvine (UCI), Irvine, CA, USA, [130]Division of Cardiology, Department of Medicine, School of Medicine, UCI, USA, [131]Center for Global Cardiometabolic Health, Brown University School of Public Health, Providence, RI, USA, [132]Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston, Houston, TX, USA, [133]Salt Lake City VA Medical Center, Salt Lake City, UT, USA, [134]University of Utah, School of Medicine, Salt Lake City, UT, USA, [135]Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Okinawa, 903-0213, Japan, [136]Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Okinawa, 903-0213, Japan, [137]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK, [138]Department of Medicine and Epidemiology, Johns Hopkins University, Baltimore, MD, 21205, USA, [139]Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK, [140]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK, [141]Department of Preventative Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA, [142]Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, 2114, USA, [143]Department of Diabetes and Endocrinology, Nelson R Mandela School of Medicine, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa, [144]Department of Medicine and Endocrinology Division, University of California Davis School of Medicine, Sacramento, CA, USA, [145]Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, 461-8673, Japan, [146]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA, [147]Center for Alzheimer's and Related Dementias, National Institutes of Health, Bethesda, MD, USA, [148]DataTecnica LLC, Washington, DC, USA, [149]Department of Genome Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA, [150]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Tokyo, 565-0871, Japan, [151]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Osaka, 565-0871, Japan, [152]Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, [153]Genomic Research on Complex diseases, CSIR-Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad, India, [154]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, 55454, USA, [155]Department of Pathology and Molecular Medicine, McMaster University, Hamilton, ON, Canada, [156]Department of Internal Medicine, Seoul National University College of Medicine and Seoul National University Hospital, Seoul, 3080, South Korea, [157]Department of Internal Medicine, Konkuk University Medical Center, Seoul, 3080, South Korea, [158]Department of Anthropology, University of Toronto at Mississauga, Mississauga, Canada, [159]Department of Medicine, Division of Pulmonary, Allergy, Critical Care, and Sleep Medicine, University of Pittsburgh, Pittsburgh, PA, USA, [160]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Systems and Population Health, University of Washington, Seattle, WA, USA, [161]Lightening Lives LLP, Hyderabad, India, [162]Department of Pediatrics, Division of Genetic and Genomic Medicine, UCI Irvine School of Medicine, Irvine, CA, USA, [163]Department of Preventive Medicine, Northwestern University, Chicago, IL, USA, [164]Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA, [165]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA, [166]Department of Epidemiology, University of Washington, Seattle, WA, USA, [167]Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [168]Department of Genome Sciences, University of Virginia, Charlottesville, VA, USA, [169]Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, 19104, USA, [170]Department of Medicine, Division of Endocrinology and Metabolism, Taichung Veterans General Hospital, Taichung, Taiwan, [171]Institutes of Molecular and Genomic Medicine, National Health Research Institutes, Taiwan, [172]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA, [173]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [174]Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA, [175]Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA, 1003, USA, [176]Nuffield Department of Surgical Sciences, University of Oxford, Oxford, OX3 9DU, UK, [177]Department of Genomic Medicine, National Cerebral and Cardiovascular Center,

Osaka, 565-8565, Japan, [178]Department of Biochemistry, Faculty of Medicine, Health Science Center, Kuwait University, Safat, Kuwait, [179]Department of Medical Genetics and Medical Research, China Medical University Hospital, Taichung, 40402, Taiwan, [180]Department of Medicine, Stanford University, Palo Alto, CA, USA, [181]VA Palo Alto Health Care System, Palo Alto, CA, USA, [182]Blizard Institute, Queen Mary University of London, London, UK, [183]Southern California Eye Institute, CHA Hollywood Presbyterian Medical Center, Los Angeles, CA, USA, [184]Unidad de Investigación Médica en Epidemiologia Clinica, Hospital de Especialidades, Centro Medico Nacional Siglo XXI, IMSS, Mexico City, Mexico, [185]Beijing Institute of Ophthalmology, Ophthalmology and Visual Sciences Key Laboratory, Beijing Tongren Hospital, Capital Medical University, Beijing, 100005, China, [186]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, UK, [187]Department of Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [188]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA, [189]State Key Laboratory of Oncogene and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, 200032, China, [190]Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, 200032, China, [191]Department of Research & Evaluation, Division of Biostatistics Research, Kaiser Permanente of Southern California, Pasadena, CA, USA, [192]Diabetes Unit, King Edward Memorial Hospital and Research Centre, Pune, India, [193]Department of Medical Biochemistry, Kurume University School of Medicine, Kurume, 830-0011, Japan, [194]Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, 15232, USA, [195]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, 15232, USA, [196]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada, [197]Population Health Research Institute, Hamilton, ON, Canada, [198]Hamilton Health Sciences, Hamilton, ON, Canada, [199]Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, [200]Technical University of Munich (TUM), TUM University Hospital, TUM School of Medicine and Health, Munich, 81675, Germany, [201]Department of Epidemiology and Biostatistics, Imperial College London, London, W2 1PG, UK, [202]Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, 37203, USA, [203]Instituto Nacional de Salud Pública, Cuernavaca, Mexico, [204]Corporal Michael J Crescenz VA Medical Center, Philadelphia, PA, USA, [205]Unidad de Investigación de Enfermedades Metabólicas, Research Direction of the Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico, [206]Tecnológico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Mexico City, Mexico, [207]Center for Infectious Disease Research, National Institute of Public Health, Cuernavaca, Mexico, [208]Centro de Estudios en Diabetes, Unidad de Investigacion en Diabetes y Riesgo Cardiovascular, Centro de Investigacion en Salud Poblacional, Instituto Nacional de Salud Pública, Mexico City, Mexico, [209]Department of Population and Public Health Sciences, Keck School of Medicine of USC, Los Angeles, CA, USA, [210]Department of Medicine, Massachusetts General Hospital, Boston, MA, 2114, USA, [211]Endocrine Division, Massachusetts General Hospital, Boston, MA, 2114, USA, [212]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA, [213]Aging Research Center, Cinvestav Sede Sur, Center for Research and Advanced Studies of the National Polytechnic Institute, Mexico City, Mexico, [214]NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK, [215]Instituto Nacional de Medicina Genómica, Mexico City, Mexico, [216]Unidad de Biología Molecular y Medicina Genómica, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México/Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico, [217]Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104, USA, [218]Institute for Translational Medicine and Therapeutics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA, [219]Department of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA, [220]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, 2114, USA, [221]Clinical and Translational Epidemiology Unit, Mongan Institute, Massachusetts General Hospital, Boston, MA, 2114, USA, [222]Department of Epidemiology and Biostatistics, Imperial College London, St Mary's Campus, London, UK, [223]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, [224]Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

Author acknowledgments and funding

This work is supported by the National Human Genome Research Institute (NHGRI) of the National Institutes of Health (NIH) grant U01HG011723. AH-C is supported by the American Diabetes Association (ADA) grant 11–23-

PDF-35. YL is supported by NHLBI R56HL150186, NHLBI R01HL158884, and NIDDK R01DK135938. MOG was supported in part by NIH grants from the National Institute of Diabetes and Digestive and Kidney Disease (NIDDK) (P30-DK063491) and from the National Center for Advancing Translational Sciences (NCATS) (UL1TR001420, UL1TR001881) and the Eris M Field Chair in Diabetes Research. SSR is supported by NHGRI U01HG011723. MAN research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging (NIA), NIH, Department of Health and Human Services; project number ZO1 AG000535, as well as the National Institute of Neurological Disorders and Stroke (NINDS). This work utilised the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov). AL is supported by grant 2020096 from the Doris Duke Foundation, the ADA grant 7–22-ICTSPM-23, and NHGRI grant U01HG011723. APM acknowledges support from the Manchester Biomedical Research Centre (NIHR203308). CNS was supported by the NIH (R01DK118011; R01DK136671) and the ADA (11–22-JDFPM-06). JBM is supported by grant UMDK078616. BFV is grateful for support from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grants DK138521 and DK126194. AKM is supported by NIDDK grant UMDK078616. JMM is supported by ADA grant 11–22-ICTSPM-16, by NHGRI grant U01HG011723, by the NIDDK under award numbers R01DK137993 and U01 DK140757, by the Accelerating Medicines Partnership Common Metabolic Diseases award from RFP 6 from the Foundation for the NIH, and by a Medical University of Bialystok grant from the Ministry of Science and Higher Education (Poland). This work is supported by the Novo Nordisk Foundation (NNF21SA0072102). MCYN is supported by NHGRI U01HG011723, NIDDK R01DK066358 and NIDDK U01DK105556. BMP is supported by NHLBI (HL105756). This study was also supported by a grant from the National Research Foundation of Korea funded by the Korean Ministry of Science and ICT (RS-2023–00262002), and by the Ministry of Food and Drug Safety grant (23212MFDS202) awarded to SHK and JC. JEB and LEP were supported in part by NHLBI R01HL142302.

Cohort acknowledgments and funding

Anti-aging Study Cohort (AASC) is supported by the Grant-in-Aid for Scientific Research (20018020, 19659163, 20390185, 23659382, 24390084, 23659352, 25293141, 26670313, 17H04123) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, research grant from the Japan Atherosclerosis Prevention Found, National Cardiovascular Research Grants, and Research Promotion Award from Ehime University.

Asian Indian Diabetic Heart Study/Sikh Diabetes Study (AIDHS/SDS) was started in 2002 by the Fogarty International Center of the National Institute of Health (NIH), NIH/FIC K01 TW006087 (2002-2005) and 2K01 TW006087 (2006-2010) awards. The AIDHS/SDS has been supported by NIH Grants: NHLBI -HV48141, NHLBI R&G Service/R218, NIDDK/R01DK082766, NIH/NHGRI HG-11-009, NIDDK/DK105913, NIDDK/R01DK118427. The authors thank the participants of the study for their important contributions.

All Of Us Research Program (AOU) is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276. In addition, the All of Us Research Program would not be possible without the partnership of its participants.

Atherosclerosis Risk in Communities (ARIC) study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I), R01HL087641, R01HL059367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

BioBank Japan (BBJ). This study was funded by the BioBank Japan project, which is supported by the Ministry of Education, Culture, Sports, Sciences and Technology (MEXT) of Japanese government and the Japan Agency for

Medical Research and Development (AMED, grant ID JP21km0605001). AMED GRIFIN Diabetes Initiative Japan was supported by Japan Agency for Medical Research and Development (JP20km0405202, JP21tm0424218). Scarda was supported by AMED under Grant Number 223fa627011.

Beijing Eye Study (BES) was supported by National Natural Science Foundation of China (grant 81570835).

BioMe Biobank (BIOME) is supported by The Andrea and Charles Bronfman Philanthropies and in part by funding of the NIH (U01HG007417; R56HG010297; X01HL134588). BIOME thanks all participants in the Mount Sinai Biobank, and also thanks all the recruiters who have assisted and continue to assist in data collection and management. BIOME is grateful for the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

Vanderbilt University Medical Center (BIOVU) projects are supported by numerous sources: institutional funding, private agencies, and federal grants. These include NIH funded Shared Instrumentation Grant S10OD017985, S10RR025141, and S10OD025092; CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, and R01HD074711. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN, supported in part by an S10 instrumentation award (1S10OD023680-01).

Bangladesh Population Cohort (BPC) was supported by US National Institute of Environmental Health Sciences Grants P42 ES10349 and P30 ES09089. Cardiometabolic Genome Epidemiology (CAGE-AMAGASKI and CAGE-GWAS) was supported by grants for the Core Research for Evolutional Science and Technology (CREST) from the Japan Science Technology Agency; KAKENHI (Grant-in-Aid for Scientific Research) from the Ministry of Education, Culture, Sports, Science and Technology of Japan; and the Grant and research budget of National Center for Global Health and Medicine (NCGM). CAGE-AMAGASKI thanks Drs. Toshio Ogihara, Yukio Yamori, Akihiro Fujioka, Chikanori Makibayashi, Sekiharu Katsuya, Ken Sugimoto, Kei Kamide, and Ryuichi Morishita and the many physicians of the participating hospitals and medical institutions in Amagasaki Medical Association for their assistance in collecting the DNA samples and accompanying clinical information.

Cardiometabolic Genome Epidemiology Kita-Nagoya Genomic Epidemiology (CAGE-KING) was supported in part by Grants-in-Aid from MEXT (nos. 24390169, 16H05250, 15K19242, 16H06277) as well as by a grant from the Funding Program for Next-Generation World-Leading Researchers (NEXT Program, no. LS056).

Coronary Artery Risk Development in Young Adults (CARDIA) was conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (HHSN268201800005I & HHSN268201800007I), Northwestern University (HHSN268201800003I), University of Minnesota (HHSN268201800006I), and Kaiser Foundation Research Institute (HHSN268201800004I). CARDIA was also partially supported by the Intramural Research Program of the National Institute on Aging (NIA) and an intra-agency agreement between NIA and NHLBI (AG0005). Genotyping was funded as part of the NHLBI Candidate-gene Association Resource (N01-HC-65226) and the NHGRI Gene Environment Association Studies (GENEVA) (U01-HG004729, U01-HG04424, and U01-HG004446).

Cleveland Family Study (CFS) is supported by grants to Case Western Reserve University (NIH HL 46380, M01RR00080) and Brigham and Women's Hospital (K01-HL135405-01, R01-HL113338-04, R35-HL135818-01, 5-R01-HL046380-15 and 5-KL2-RR024990-05).

China Health and Nutrition Survey (CHNS) was supported by: the National Institute for Nutrition and Health, the Chinese Center for Disease Control and Prevention; the National Institutes of Health (R01AG065357, R01HD30880, R01HL108427 and R01DK104371); the Fogarty International Center of the National Institutes of Health (TW009077); the China-Japan Friendship Hospital, the Beijing Municipal Center for Disease Prevention and Control, the China National Health Commission (formerly the Chinese Ministry of Health); the Chinese National Human Genome Center at Shanghai; and the Carolina Population Center (P2CHD050924), The University of North Carolina at Chapel Hill.

Cardiovascular Health Study (CHS). This Cardiovascular Health Study research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, 75N92021D00006; and NHLBI grants U01HL080295, R01HL087652, R01HL103612, R01HL105756, R01HL120393, U01HL130114, and R01HL172803 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

China Kadoorie Biobank (CKB) chiefly acknowledges the participants, project staff, and the China National Centre for Disease Control and Prevention (CDC) and its regional offices. China's National Health Insurance provides electronic linkage to all hospital treatment. Funding sources: Baseline survey and first re-survey - Kadoorie Charitable Foundation, Hong Kong; long-term follow-up - UK Wellcome Trust (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z), National Natural Science Foundation of China (82192901, 82192904, 82192900), and National Key Research and Development Program of China (2016YFC 0900500, 0900501, 0900504, 1303904); DNA extraction and genotyping –GlaxoSmithKline, and the UK Medical Research Council (MC-PC-13049, MC-PC-14135); core funding for the project to the Clinical Trial Service Unit and Epidemiological Studies Unit at Oxford University - British Heart Foundation (CH/1996001/9454), UK MRC (MC-UU-00017/1, MC-UU-12026/2, MC_U137686851), and Cancer Research UK (C16077/A29186, C500/A16896).

Cebu Longitudinal Health and Nutrition Survey (CLHNS) was supported by: US National Institutes of Health grants DK078150, TW005596 and HL085144; pilot funds from RR020649, ES010126, and DK056350; and the Office of Population Studies Foundation.

Diabetic Cohort and Singapore Prospective Study Program (DC/SP2) were supported by the individual research grant and clinician scientist award schemes from the National Medical Research Council (NMRC) and the Biomedical Research Council (BMRC) of Singapore, Ministry of Health, Singapore, and infrastructure funding from the Singapore Ministry of Health (Population Health Metrics and Analytics PHMA), National University of Singapore and National University Health System, Singapore.

Durban Diabetes Study and Durban Diabetes Case Control (DDS/DCC) was supported by: the Wellcome Trust (grant number 098051); the African Partnership for Chronic DiseaseResearch (Medical Research Council UK partnership grant number MR/K013491/1); the National Institute for Health Research Cambridge Biomedical Research Centre (UK); Novo-Nordisk (South Africa); Sanofi-Aventis (South Africa); MSD Pharmaceuticals (Pty) Ltd (Southern Africa); Servier Laboratories (South Africa); South African Sugar Association; and the Victor Daitz Foundation.

deCODE genetics (DECODE) thank the participants in the deCODE study, the staff at deCODE genetics core facilities and the staff at the Research Service Center for their contribution to this work.

Diabetes Gene Discovery Group (DGDG) was supported by Genome Canada, Génome Québec, the Canada Foundation for Innovation, the French Government ("Agence Nationale de la Recherche"), the French Region of "Nord Pas De Calais" ("Contrat de Projets État-Région"), and the charities: "Association Française des Diabétiques", "Programme National de Recherche sur le Diabète" and "Association de Langue Française pour l'Etude du Diabète et des Maladies Métaboliques". This study was also supported in part by a grant from the European Union (Integrated Project EuroDia LSHM-CT-2006-518153in the Framework Programme 6 [FP6] of the European Community). This work was supported by grants from the French National Research Agency (ANR-10-LABX-46 [European Genomics Institute for Diabetes] and ANR-10-EQPX-07-01 [LIGAN-PM]). Case and control recruitment was supported by the Fédération Française des Diabetiques, INSERM, CNAMTS, Centre Hospitalier Universitaire Poitiers, La Fondation de France, and the Endocrinology-Diabetology department of the Corbeil-Essonnes Hospital. C. Petit, J.-P. Riveline, and S. Franc were instrumental in recruitment and S. Brunet, F. Bacot, R. Frechette, V. Catudal, M. Deweirder, F. Allegaert, P. Laflamme, P. Lepage, W. Astle, M. Leboeuf, and S. Leroux provided technical assistance. K. Shazand and N. Foisset provided organizational guidance. The D.E.S.I.R. study, which mostly contributed controls, was supported by CNAMTS, Lilly, Novartis Pharma and Sanofi-Aventis, by INSERM ("Réseaux en Santé Publique, Interactions entre les déterminants de la santé"), by "Association Diabète Risque Vasculaire", "Fédération Française

de Cardiologie", "Fondation de France", ALFEDIAM, ONIVINS, Ardix Medical, Bayer Diagnostics, Becton Dickinson, Cardionics, Merck Santé, Novo Nordisk, Pierre Fabre, Roche, Topcon. The D.E.S.I.R. Study Group: INSERM U780: B. Balkau, P. Ducimetière, E. Eschwège; INSERM U367: F. Alhenc-Gelas; CHU D'Angers: Y. Gallois, A. Girault; Bichat Hospital: F. Fumeron, M. Marre; Medical Examination Services: Alençon, Angers, Caen, Chateauroux, Cholet, Le Mans, and Tours; Research Institute for General Medicine: J. Cogneau; General practitioners of the region; Cross-Regional Institute for Health: C. Born, E. Caces, M. Cailleau, J. G. Moreau, F. Rakotozafy, J. Tichet, S. Vol. DGDG thank M. Deweider and F. Allegaert for the DNA bank management and are sincerely indebted to all study participants.

Diabetes Genetics Initiative (DGI) was supported by the Novartis Institute for BioMedical Research with additional support from The Richard and Susan Smith Family Foundation and American Diabetes Association Pinnacle Program Project Award. The Botnia Study (study subject cohort) was financially supported by the Folkhalsan Research Foundation, the Sigrid Juselius Foundation, Nordic Center of Excellence in Disease Genetics, EU (EXGENESIS), The Academy of Finland, University of Helsinki, Finnish Diabetes Research Foundation, Foundation for Life and Health in Finland, Finnish Medical Society, Helsinki UniversityCentral Hospital Research Foundation, Perklén Foundation, Ollqvist Foundation, Närpes Health Care Foundation, Municipal Heath Care Center and Hospital in Jakobstad and Health Care Centers in Vasa, Närpes and Korsholm. The work in Malmö, Sweden, was also funded by a Linné grant from the Swedish Research Council (349-2006-237). The contribution of the Botnia and Skara research teams is gratefully acknowledged.

Electronic Medical Records and Genomics Network (EMERGE) was initiated and funded by NHGRI through the following grants: U01HG006828 (Cincinnati Children's Hospital Medical Center/Boston Children's Hospital); U01HG006830 (Children's Hospital of Philadelphia); U01HG006389 (Essentia Institute of Rural Health, Marshfield Clinic Research Foundation and Pennsylvania State University); U01HG006382 (Geisinger Clinic); U01HG006375 (Group Health Cooperative/University of Washington); U01HG006379 (Mayo Clinic); U01HG006380 (Icahn School of Medicine at Mount Sinai); U01HG006388 (Northwestern University); U01HG006378 (Vanderbilt University Medical Center); and U01HG006385 (Vanderbilt University Medical Center serving as the Coordinating Center). The Northwestern University Enterprise Data Warehouse was funded in part by a grant from the National Center for Research Resources, UL1RR025741. Part of the dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant UL1 TR000445 from NCATS/NIH. The eMERGE imputed merged Phase I and Phase II dataset was generated by genotyping centers CIDR (U01HG004438) and the Broad Institute (U01HG004424).

European Prospective Investigation into Cancer and Nutrition (EPIC-INTERACT) project (LSHM-CT-2006-037197) is a European-Community funded project under Framework Programme 6. EPIC-INTERACT thank all EPIC participants and staff for their contribution to the study. EPIC-INTERACT thank Nicola Kerrison (MRC Epidemiology Unit, Cambridge) for managing the data for the InterAct Project and staff from the Laboratory Team, Field Epidemiology Team, and Data Functional Group of the MRC Epidemiology Unit in Cambridge, UK, for carrying out sample preparation, DNA provision and quality control, genotyping, and data-handling work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. GWAS summary statistics from the EPIC-InterAct study are available to download from the Dryad Digital Repository (https://doi.org/10.5061/dryad.qnk98sfcg).

Epidemiologic Study of the Screenees for Diabetes Reduction Assessment with Ramipril and Rosiglitazone Medication (EPIDREAM) was funded by a grant from the Canadian Institutes of Health Research University Industry competition with partner funding from the GlaxoSmithKline and Sanofi Aventis Global, Sanofi Aventis Canada, Genome Quebec Innovation Centre, Heart and Stroke Foundation of Canada.

Estonian Biobank (ESTBB) was funded by the Estonian Research Council Grant IUT20-60, IUT24-6, PRG687, and the European Union through the European Regional Development Fund Project No. 2014-2020.4.01.15-0012 GENTRANSMED.

Family Heart Study (FAMHS) was supported by NIH grants R01-HL-087700 and R01-HL-088215 from NHLBI, and R01-DK-089256 and R01-DK-075681 from NIDDK.

Framingham Heart Study (FHS) was conducted and supported by the National Heart, Lung and Blood Institute (NHLBI) in collaboration with Boston University (contracts 75N92019D00031, HHSN268201500001I and N01-HC-25195), and its contract with Affymetrix, Inc for genotyping services (contract number N02-HL-6-4278). The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. FHS was also supported by: NHLBI R01 HL105756, National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) R01 DK078616, U01 DK078616, NIDDK K24 DK080140 and American Diabetes Association Mentor-Based Postdoctoral Fellowship Award #7-09-MN-32 (to J.B.M.); and NIDDK K24 DK110550 (to J.C.F.).

FinnGen study (FINNGEN) is a large-scale genomics initiative that has analyzed over 500,000 Finnish biobank samples and correlated genetic variation with health data to understand disease mechanisms and predispositions. The project is a collaboration between research organisations and biobanks within Finland and international industry partners. We want to acknowledge the participants and investigators of the FinnGen study.

Finland-United States Investigation of NIDDM Genetics (FUSION) was supported by DK093757, DK072193, DK062370, and ZIA-HG000024.

German Chronic Kidney Disease (GCKD) was funded by the German Ministry of Research and Education (Bundesminsterium für Bildung und Forschung, BMBF) and by the Foundation KfH Stiftung Präventivmedizin. Unregistered grants to support the study were provided by Bayer, Fresenius Medical Care and Amgen. Genotyping was supported by Bayer AG.

Genetic Study of Atherosclerosis Risk (GENESTAR) was supported by NIH grants through the National Heart, Lung, and Blood Institute (HL49762, HL58625, HL59684, HL071025, U01HL72518, and HL087698) and the National Institute of Nursing Research (NR0224103) and by M01-RR000052 to the Johns Hopkins General Clinical Research Center.

Genetic Epidemiology Network of Arteriosclerosis (GENOA) was supported by the National Institutes of Health grant numbers U01HL054457, U01HL054464, U01HL054481, R01HL087660, and R01HL119443 from the National Heart, Lung, and Blood Institute. Genotyping was performed at the Mayo Clinic by Stephen Turner, Mariza de Andrade, and Julie Cunningham. GENOA thanks Eric Boerwinkle and Megan Grove from the Human Genetics Center and Institute of Molecular Medicine and Division of Epidemiology, University of Texas Health Science Center, Houston, Texas, USA for their help with genotyping. GENOA also thanks the families that participated in the study.

Genes and Health. Genes & Health is/has recently been core-funded by Wellcome (WT102627, WT210561), the Medical Research Council (UK) (M009017, MR/X009777/1, MR/X009920/1), Higher Education Funding Council for England Catalyst, Barts Charity (845/1796), Health Data Research UK (for London substantive site), and research delivery support from the NHS National Institute for Health Research Clinical Research Network (North Thames). Genes & Health is/has recently been funded by Alnylam Pharmaceuticals, Genomics PLC; and a Life Sciences Industry Consortium of AstraZeneca PLC, Bristol-Myers Squibb Company, GlaxoSmithKline Research and Development Limited, Maze Therapeutics Inc, Merck Sharp & Dohme LLC, Novo Nordisk A/S, Pfizer Inc, Takeda Development Centre Americas Inc. We thank Social Action for Health, Centre of The Cell, members of our Community Advisory Group, and staff who have recruited and collected data from volunteers. We thank the NIHR National Biosample Centre (UK Biocentre), the Social Genetic & Developmental Psychiatry Centre (King's College London), Wellcome Sanger Institute, and Broad Institute for sample processing, genotyping, sequencing and variant annotation. This work uses data provided by patients and collected by the NHS as part of their care and support. This research utilised Queen Mary University of London's Apocrita HPC facility, supported by QMUL Research-IT, http://doi.org/10.5281/zenodo.438045. We thank: Barts Health NHS Trust, NHS Clinical Commissioning Groups (City and Hackney, Waltham Forest, Tower Hamlets, Newham, Redbridge, Havering, Barking and Dagenham), East London NHS Foundation Trust, Bradford Teaching Hospitals NHS Foundation Trust, Public Health England (especially David Wyllie), Discovery Data Service/Endeavour Health Charitable Trust (especially David Stables), Voror Health Technologies Ltd (especially Sophie Don), NHS England (for what was NHS Digital) - for GDPR-compliant data sharing backed by individual written informed consent. Most of all we thank all of the volunteers participating in Genes & Health. A favourable ethical opinion for the main Genes & Health research study was granted by NRES Committee London - South East (reference 14/LO/1240) on 16 Sept 2014. Queen Mary University of London is the Sponsor.

Resource for Genetic Epidemiology on Adult Heath and Aging (GERA) was supported by a grant (RC2 AG033067; PIs Schaefer and Risch) awarded to the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and the UCSF Institute for Human Genetics. The RPGEH was supported by grants from the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, Kaiser Permanente Northern California, and the Kaiser Permanente National and Northern California Community Benefit Programs.

Genetics of Diabetes and Audit Research in Tayside Scotland (GODARTS) was funded by The Wellcome Trust Study Cohort Functional Genomics Grant (2004-2008, 072960/Z/03/Z) and The Wellcome Trust Scottish Health Informatics Programme (SHIP, 2009-2012, 086113/Z/08/Z).

Genetics of Latinos Diabetic Retinopathy (GOLDR) was supported by grants EY14684 and UL1TR000124.

Genetic Overlap Between Metabolic and Psychiatric Traits and Teens of Attica: Genes and Environment (GOMAP-TEENAGE) was funded by the Wellcome Trust (098051) and was also co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. GOMAP-TEENAGE thanks all study participants and their families, as well as all volunteers for their contribution in this study. GOMAP-TEENAGE is grateful to: Georgia Markou, Laiko General Hospital Diabetes Centre; Maria Emetsidou and Panagiota Fotinopoulou, Hippokratio General Hospital Diabetes Centre; Athina Karabela, Dafni Psychiatric Hospital; Eirini Glezou and Marios Mangioros, Dromokaiteio Psychiatric Hospital; Angela Rentari, Harokopio University of Athens; and Danielle Walker, Wellcome Trust Sanger Institute. GOMAP-TEENAGE thanks the Sample Management and Genotyping Facilities staff at the Wellcome Trust Sanger Institute for sample preparation, quality control and genotyping.

Genomic Research Cohort for CCMB Diabetes Study (GRCCDS) comprises of various cohorts that are supported by: Council of Scientific Industrial Research (CSIR); Ministry of Science and Technology, Govt. of India, India; and Wellcome Trust, London, UK. GRCCDS is grateful to the patients and subjects who voluntarily participated in the study, and thankfully acknowledge other researchers who have supported the study.

Health, Aging and Body Composition Study (HABC) was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.

Healthy Aging in Neighborhoods of Diversity Across the Life Span Study (HANDLS) was supported by the Intramural Research Program of the NIH, National Institute on Aging (project Z01-AG000513 and human subjects' protocol 09 AGN248). Data analyses for HANDLS utilized the high-performance computational resources of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://hpc.nih.gov).

Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN268201300001I / N01-HC-65233), University of Miami (HHSN268201300004I / N01-HC-65234), Albert Einstein College of Medicine (HHSN268201300002I / N01-HC-65235), University of Illinois at Chicago (HHSN268201300003I / N01-HC-65236 Northwestern Univ), and San Diego State University (HHSN268201300005I / N01-HC-65237). The following Institutes/Centers/Offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03).

Health Professionals' Follow-Up Study (HPFS) and Nurses Health Study (NHS) acknowledge assistance with data cleaning that was provided by the National Center for Biotechnology Information. Support for collection of datasets

and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" (HHSN268200782096C). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000091.v1.p1 through dbGaP accession number phs000091.v1.p. Mexican American Hypertension and Insulin Resistance (HTNIR) was supported by grant HL0597974.

Hypertension and Insulin Resistance (HTN-IR) was supported by R01-HL067974, R01-HL-55005 and R01-HL 067974.

Howard University Family Study (HUFS) was supported by National Institutes of Health grants S06GM008016-320107 to CNR and S06GM008016-380111 to AA. Participant enrollment was carried out at the Howard University General Clinical Research Center, supported by National Institutes of Health grant 2M01RR010284. Genotyping support was provided by the Coriell Institute for Medical Research. This research was supported by the Intramural Research Program of the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute, the National Institute of Diabetes and Digestive and Kidney Diseases, the Center for Information Technology, and the Office of the Director at the National Institutes of Health (Z01HG200362).

INTERHEART (INTERHEART) was funded by: the Canadian Institutes of Health Research, the Heart and Stroke Foundation of Ontario, and the International Clinical Epidemiology Network (INCLEN); unrestricted grants from several pharmaceutical companies (with major contributions from AstraZeneca, Novartis, Hoechst Marion Roussel [now Aventis], Knoll Pharmaceuticals [now Abbott], Bristol-Myers Squibb, King Pharma, and Sanofi-Synthelabo); and various national bodies in different countries (see Online Appendix at http://image.thelancet.com/extras/04art8001webappendix2.pdf). Funding sources had no involvement in the study design; in the collection, analysis, and interpretation of data; or the writing of the manuscript.

The Jackson Heart Study (JHS) is supported by Contracts HHSN268201800010I, HHSN268201800011I, HHSN268201800012I, HHSN268201800013I, HHSN268201800014I, HHSN268201800015I from the National Heart, Lung, and Blood Institute (NHLBI) with additional support from the National Institute on Minority Health and Health Disparities (NIMHD). This manuscript has been reviewed by JHS for scientific content.

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

Korean Association Resource (KARE) was supported by grants from Korea Centers for Disease Control and Prevention (4845–301, 4851–302, 4851–307) and intramural grants from the Korea National Institute of Health (2016-NI73001-00, 2019-NG-053-00). KARE was performed with bioresources from National Biobank of Korea, the Centers for Disease Control and Prevention, Republic of Korea.

Korea Biobank Array (KBA) Project was supported by an intramural grant from the National Institute of Health, Disease Control Prevention and Control Agency, Republic of Korea (2025-NI-002-00).

Korean Biobank Array from the Korean Genome and Epidemiology (KoGES) Consortium (KBA) was supported by grants from Korea Centers for Disease Control and Prevention (4845–301, 4851–302, 4851–307) and intramural grants from the Korea National Institute of Health (2016-NI73001-00, 2019-NG-053-00). KBA was performed with bioresources from National Biobank of Korea, the Centers for Disease Control and Prevention, Republic of Korea. Genotype data were provided by the Collaborative Genome Program for Fostering New Post-Genome Industry (3000-3031b).

Collaborative Health Research in the Region of Augsburg (KORA) research platform was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ and by the German Center for Diabetes Research (DZD).

Los Angeles Latino Eye Study (LALES) acknowledges funding from NEI grant U10EY011753.

London Life Sciences Prospective Population (LOLIPOP) is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the British Heart Foundation (SP/04/002), the MedicalResearch Council (G0601966, G0700931), the Wellcome Trust (084723/Z/08/Z, 090532 & 098381) the NIHR (RP-PG-0407-10371), the NIHR Official Development Assistance (ODA, award 16/136/68), the European Union FP7 (EpiMigrant, 279143) and H2020 programs (iHealth-T2D, 643774). LOLIPOP acknowledges support of the MRC-PHE Centre for Environment and Health, and the NIHR Health Protection Research Unit on Health Impact of Environmental Hazards. The work was carried out in part at the NIHR/Wellcome Trust Imperial Clinical Research Facility. The views expressed are those of the author(s) and not necessarily those of the Imperial College Healthcare NHS Trust, the NHS, the NIHR or the Department of Health. LOLIPOP thanks the participants and research staff who made the study possible.

Mexican American Study of Coronary Artery Disease (MACAD) was supported by grant R01-HL088457 and R01-HL-60030.

Mexico City T2D study (MC). In Mexico, this work was supported by the Fondo Sectorial de Investigación en Salud y Seguridad Social (SSA/IMSS/ISSSTE-CONACYT) project 150352, Temas Prioritarios de Salud Instituto Mexicano del Seguro Social 2014-FIS/IMSS/PROT/PRIO/14/34 and the Fundación IMSS. We thank Jorge Gutierrez Cuevas, Jaime Gómez Zamudio and Araceli Méndez Padrón for technical support. In Canada, the research was supported by a Canadian Institutes of Health Research (CIHR) operating grant to EJP and also by funding from the Banting and Best Diabetes Centre to EJP. In Canada, computations were performed on the GPC supercomputer at the SciNet HPC Consortium (https://scinethpc.ca/). SciNet is funded by Innovation, Science and Economic Development Canada; the Digital Research Alliance of Canada (https://alliancecan.ca/) ; the Ontario Research Fund: Research Excellence; and the University of Toronto.

Multi-Ethnic Study of Atherosclerosis (MESA). MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institue (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92025D00022, 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92025D00026, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92025D00024, 75N92020D00003, N01-HC-95162, 75N92025D00027, 75N92020D00006, N01-HC-95163, 75N92025D00025, 75N92020D00004, N01-HC-95164, 75N92025D00028, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1TR001881, DK063491, and R01HL105756. The authors thank the MESA participants and the MESA investigators and staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org. This research was also supported by the Mexican-American Coronary Artery Disease (MACAD) National Heart, Lung, and Blood Institute, contracts R01-HL088457, R01-HL-60030; Hypertension and Insulin Resistance (HTN-IR) contracts R01-HL067974, R01-HL-55005, R01-HL 067974, and the Genetics of Latinos Diabetic Retinopathy (GOLDR) Study grant EY14684.

Metabolic Syndrome in Men (METSIM) was supported by the Academy of Finland (contract 124243), the Finnish Heart Foundation, the Finnish Diabetes Foundation, Tekes (contract 1510/31/06), and the Commission of the European Community (HEALTH-F2-2007 201681), and the US National Institutes of Health grants DK093757, DK072193, DK062370, and ZIA-HG000024.

Mexican Biobank (MXBB) was supported by Mexico's CONACYT (Grant number FONCICYT/50/2016; PI Moreno-Estrada), and the Newton Fund through the UK Medical Research Council (Grant number MR/N028937/1; PI Moreno-Estrada) to genetically characterize the population-based cohort derived from the National Health Survey 2000 (ENSA2000). The resulting ENSA Genomics Consortium acknowledges the seminal effort of Dr. Jaime Sepúlveda, the Mexican Ministry of Health, and the National Institute of Public Health, in the design and

implementation of the ENSA2000 survey from which genomic data were generated for the MXB Project. Members of the ENSA Genomics Consortium are also acknowledged for biobank maintenance, sample selection and processing of materials contributed to the MXBB.

Mass General Brigham Biobank (MGBB) acknowledges the Partners HealthCare System for support of the MGB biobank and MGB patients for providing samples, genomic data, and health information data, as well as research support by NIDDK K24 DK110550 (to J.C.F.), K24 DK080140 (to J.B.M.) and NIDDK K23DK114551 (to M.S.U).

Michigan Genomics Initiative (MGI) was supported by NIH research grants HL117626 and HG007022. MGI was supported by internal research funds from the University of Michigan School of Public Health, the University of Michigan Medical School, and the University of Michigan President's Office. MGI are especially grateful to the generosity of all research participants.

VA Million Veteran Program (MVP). We gratefully acknowledge the Veterans who participated in the VA's Million Veteran Program. This research is based on data from the Million Veteran Program, Office of Research and Development, and Veterans Health Administration, with support from MVP000, VA Merit Award #I01-BX003362 (Chang/Tsao) and the Department of Veterans Affairs (VA) Informatics and Computing Infrastructure (VINCI), including data analytics conducted by its Precision Medicine research team, which is funded under the research priority to Put VA Data to Work for Veterans (VA ORD 24-D4V-02). Support was also provided by CSP2012 (JAL), R01DK134575 (MV). JAL, BFV, and MV are also supported by MVP003. This publication does not represent the views of the Department of Veterans Affairs or the United States Government.

The NAGAHAMA Study is a community-based longitudinal study in Japan. Participants in the Nagahama study were recruited from general population of Nagahama, a rural city of 125,000 inhabitants located in central Japan (N = 9,764). Community residents, aged between 30 and 74 years at recruitment, who were living independently without physical impairment or dysfunction, were eligible. All study procedures were approved by the Ethics Committee of Kyoto University Graduate School of Medicine and the Nagahama Municipal Review Board. Written informed consent was obtained from all participants.

Netherlands Epidemiology of Obesity (NEO) thanks all individuals who participated in the study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. NEO thank the study group, Pat van Beelen, Petra Noordijk and Ingeborg de Jonge for the coordination, lab and data management of the study. Genotyping was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. NEO is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area Vascular and Regenerative Medicine. NIDDM-Atherosclerosis Study Hispanic Cohorts (NIDDM) was supported by grant HL055798.

Northwestern University Genetics (NUGENE) was funded by the Northwestern University's Center for Genetic Medicine, Northwestern University, and Northwestern Memorial Hospital. Samples and data used in this study were provided by the NUgeneProject (www.nugene.org). Assistance with phenotype harmonization was provided by the eMERGE Coordinating Center (Grant number U01HG04603). This study was funded through the NIH, NHGRI eMERGE Network (U01HG004609). Funding support for genotyping, which was performed at The Broad Institute, was provided by the NIH (U01HG004424). Assistance with phenotype harmonization and genotype data cleaning was provided by the eMERGE Administrative Coordinating Center (U01HG004603) and the National Center for Biotechnology Information (NCBI). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000237.v1.p1.

Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) was supported by Wellcome Trust Grants (WT098017, WT064890, WT090532), Uppsala University, Uppsala University Hospital, the Swedish Research Council, and the Swedish Heart-Lung Foundation.

Pakistan Risk of Myocardial Infarction Study (PROMIS) was funded by the Wellcome Trust, UK, and Pfizer (genotyping) and was supported through funds available to investigators at the Center for Non-Communicable Diseases, Pakistan, and the University of Cambridge, UK (fieldwork). Biomarker assays in PROMIS have been funded through grants awarded by the National Institutes of Health (RC2HL101834 and RC1TW008485) and the Fogarty International (RC1TW008485).

Prospective Study of Pravastatin in the Elderly at Risk (PROSPER) was supported by an investigator-initiated grant obtained from Bristol-Myers Squibb. Prof. J.W.J. is an Established Clinical Investigator of the Netherlands Heart Foundation (grant 2001 D 032). Support for genotyping was provided by the seventh framework program of the European commission (grant 223004) and by the Netherlands Genomics Initiative (Netherlands Consortium for Healthy Aging grant 050-060-810).

Sea Islands Genetic Network Reasons for Geographic and Racial Differences in Stroke (REGARDS) is supported by cooperative agreement U01 NS041588 co-funded by the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institute on Aging (NIA), National Institutes of Health, Department of Health and Human Service. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NINDS or the NIA. Additional funding was from R01 DK084350 from the National Institutes of Health.

Ragama Health Study (RHS) was supported by a grant from the National Center for Global Health and Medicine (NCGM)

Rotterdam Study (RS) is grateful to the participants and staff involved in the study, and the participating general practitioners and pharmacists. RS is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

San Antonio Family Heart Study (SAFHS) was supported by U01 DK085524, R01 HL113323, P01 HL045222, R01 DK047482, and R01 DK053889. The Veterans Administration Genetic Epidemiology Study (VAGES) study was supported by a Veterans Administration Epidemiologic grant. The Family Investigation of Nephropathy and Diabetes - San Antonio (FIND-SA) study was supported by NIH grant U01 DK57295. The SAMAFS research team acknowledges late Dr. Hanna E. Abboud's contributions to the research activities of the SAMAFS.

Shanghai Breast Cancer Study and Shanghai Women's Health Study (SBCS/SWHS) was supported in part by US National Institutes of Health grants R01CA64277 and R01CA124558, as well as Ingram Professorship and Research Reward funds from the Vanderbilt University School of Medicine. We want to thank participants and research staff of the study, Regina Courtney for plasma and DNA sample preparation, and Hui Cai, Ben Zhang and Jing He for data processing and analyses.

Singapore Chinese Eye Study (SCES) is supported by the National Medical Research Council (NMRC), Singapore (grants 0796/2003, 1176/2008, 1149/2008, STaR/0003/2008, 1249/2010, CG/SERI/2010, CIRG/1371/2013, and CIRG/1417/2015), and Biomedical Research Council (BMRC), Singapore (08/1/35/19/550 and 09/1/35/19/616).

Starr County Health (SCH) was supported by grants from the National Institutes of Health (DK073541, DK085501, HL102830 and DK116378) and funds from the State of Texas. SCH thank the field staff in Starr County for their careful collection of these data and are especially grateful to the participants who so graciously cooperated and gave of their time. Starr County Health Singapore Chinese Health Study (SCHS) was supported by the US National Institutes of Health grants R01DK08072, R01CA144034 and UM1CA182876.

Slim Initiative for Genomic Medicine in the Americas (SIGMA). This work was conducted as part of the Slim Initiative for Genomic Medicine, a joint U.S.-Mexico project funded by the Carlos Slim Health Institute. The UNAM/INCMNSZ diabetes study was supported by Consejo Nacional de Ciencia y Tecnología grants 138826, 128877, CONACyT- SALUD 2009-01-115250, and a grant from Dirección General de Asuntos del Personal Académico, UNAM, IT 214711. The Diabetes in Mexico Study was supported by Consejo Nacional de Ciencia y Tecnología grant 86867 and by Instituto Carlos Slim de la Salud, A.C. The Mexico City Diabetes Study was supported by National Institutes of Health (NIH) grant R01HL24799 and by the Consejo Nacional de Ciencia y Tenologia grants: 2092, M9303, F677-M9407, 251M, and 2005-C01-14502, SALUD 2010-2-151165. The Multiethnic Cohort was supported by NIH grants CA164973, CA054281, and CA063464.

Singapore Malay Eye Study (SIMES) is supported by the National Medical Research Council (NMRC), Singapore (grants 0796/2003, 1176/2008, 1149/2008, STaR/0003/2008, 1249/2010, CG/SERI/2010, CIRG/1371/2013, and CIRG/1417/2015), and Biomedical Research Council (BMRC), Singapore (08/1/35/19/550 and 09/1/35/19/616).

Singapore Indian Eye Study (SINDI) is supported by the National Medical Research Council (NMRC), Singapore (grants 0796/2003, 1176/2008, 1149/2008, STaR/0003/2008, 1249/2010, CG/SERI/2010, CIRG/1371/2013, and CIRG/1417/2015), and Biomedical Research Council (BMRC), Singapore (08/1/35/19/550 and 09/1/35/19/616).

Samsung Medical Center (SMC) was supported by a grant from Samsung Biomedical Research Institute. Genotyping of the patients and control subjects from SMC was conducted by Duk-Hwan Kim in the Dept. of Molecular Cell Biology, Sungkyunkwan University School of Medicine, and was supported by a grant from Samsung Biomedical Research Institute.

Seoul National University Hospital (SNUH) was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare (grant numbers HI15C1595, HI14C0060, HI15C3131).

Taiwan Metabochip Consortium Zhonghua (TAICHI-G) was supported by grants from: the National Health Research Institutes, Taiwan (PH-099-PP-03, PH-100-PP-03, and PH-101-PP- 03); the National Science Council, Taiwan (NSC 101-2314-B-075A-006-MY3, MOST 104-2314- B-075A-006-MY3, MOST 104-2314-B-075A-007, and MOST 105-2314-B-075A-003); and the Taichung Veterans General Hospital, Taiwan (TCVGH-1020101C, TCVGH-1020102D, TCVGH-1023102B, TCVGH-1023107D, TCVGH-1030101C, TCVGH- 1030105D, TCVGH-1033503C, TCVGH-1033102B, TCVGH-1033108D, TCVGH-1040101C, TCVGH-1040102D, TCVGH-1043504C, and TCVGH-1043104B). TAICHI-G was also supported in part by the National Center for Advancing Translational Sciences (CTSI grant UL1TR001881).

Taiwan Type 2 Diabetes (TWT2D) was supported by the GMM Study, Academia Sinica, Taiwan.

Danish T2D Case-Control Study (UCPH) was undertaken by the Novo Nordisk Foundation Center for Basic Metabolic Research, which is an independent Research Center, based at the University of Copenhagen, Denmark and partially funded by an unconditional donation from the Novo Nordisk Foundation (www.cbmr.ku.dk, Grant number NNF18CC0034900). Included study samples were supported by the Danish Research Fund and the National Danish Research Fund (The Vejle Diabetes Biobank), the Velux Foundation, The Danish Medical Research Council and Danish Agency for Science, Technology and Innovation (Health 2006); the Danish Research Council, the Danish Centre for Health Technology Assessment and Novo Nordisk Inc. (Inter99), the Timber Merchant Vilhelm Bang's Foundation and the Danish Heart Foundation (Health 2008), TrygFonden, the Lundbeck Foundation and the Novo Nordisk Foundation (NNF15OC0015896, DanFunD).

UK Biobank (UKBB) analyses were conducted using the UK Biobank resource under applications 236, 9161, and 10035. This research was supported by the British Heart Foundation (grant SP/13/2/30111). Large-scale comprehensive genotyping of UK Biobank for cardiometabolic traits and diseases: UK CardioMetabolic Consortium (UKCMC).

Uppsala Longitudinal Study of Adult Men (ULSAM) was supported by Wellcome Trust Grants (WT098017, WT064890, WT090532), Uppsala University, Uppsala University Hospital, the Swedish Research Council, and the Swedish Heart-Lung Foundation.

Wake Forest School of Medicine (WFSM) was supported by NIH grants K99 DK081350, R01 DK066358, R01 DK053591, R01 DK087914, U01 DK105556, R01 HL56266, R01 DK070941 and in part by the General Clinical Research Center of the Wake Forest School of Medicine grant M01 RR07122. Genotyping services were provided by the Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSC268200782096C.

Women's Health Initiative (WHI). The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, 75N92021D00005.

Wellcome Trust Case Control Consortium (WTCCC) analysis and genotyping was supported by: Wellcome Trust funding 090367, 098381, 090532, 083948, 085475, 101630, and 203141; MRC (G0601261); EU (Framework 7) HEALTH-F4-2007-201413; and NIDDK DK098032 and U01-DK105535.

Ethics approval committees/IRBs

Anti-aging study cohort (AASC). The ethics committees of Ehime University Graduate School of Medicine approved all study procedures. Written informed consent was obtained from all participants.

All Of Us Research Program (AOURP). All research was conducted under the guidelines defined by the All of Us Ethical Conduct of Research Policy.

Atherosclerosis Risk in Communities (ARIC). Institutional Review Board approvals were obtained at all study sites: National Heart, Lung, and Blood Institute, University of North Carolina at Chapel Hill, Wake Forest Baptist Medical Center, University of Mississippi Medical Center, University of Minnesota, and Johns Hopkins University. All participants provided written informed consent.

Biobank Japan (BBJ). All participants provided written informed consent as approved by the ethical committees of the RIKEN Yokohama Institute and the Institute of Medical Science, the University of Tokyo. Ethical approvals of AMED GRIFIN Diabetes Initiative Japan were gained from the Ethics Committees of Osaka University and the University of Tokyo.

Beijing Eye Study (BES). Approval was obtained from the Medical Ethics Committee of the Beijing Tongren Hospital. All participants gave written informed consent.

BioMe Biobank (BIOME). Approval was obtained from the Institutional Review Board at the Icahn School of Medicine at Mount Sinai. All participants provided written informed consent for genomic data sharing.

Vanderbilt University Medical Center's BioVU (BIOVU). Analyses of DIAMANTE data at Vanderbilt University Medical Center are approved under IRB #190891 and analysis of BioVU data are approved under IRBs #210163 and #171279. In all three cases, the data analyzed received non-human subject determinations.

Bangladesh Population Cohort (BPC). The conduct of the BPC was reviewed and approved by Ethical Committees of the Bangladesh Medical Research Council and Institutional Review Boards of the University of Chicago.

Cardiometabolic Genome Epidemiology (CAGE-AMAGASKI and CAKE-GWAS). Approval was obtained from the Institutional Review Boards at the National Center for Global Health and Medicine. All participants provided written informed consent.

Cardiometabolic Genome Epidemiology (CAGE-KING). Approval was obtained from the ethics committees of Aichi Gakuin University, Jichi Medical University, Nagoya University and Kyushu University. All participants provided written informed consent.

Coronary Artery Risk Development in Young Adults (CARDIA). Participating centers (Northwestern University, University of Alabama Birmingham, University of Minnesota, and Kaiser Foundation Research Institute) provided ethics approval for the CARDIA study, and all participants provided written informed consent to participate.

Cleveland Family Study (CFS). Approval was obtained from the Institutional Review Board of Mass General Brigham (formerly Partners HealthCare). Written informed consent was obtained from all participants.

China Health and Nutrition Survey (CHNS). Approval was obtained from the Institutional review Boards at the University of North Carolina at Chapel Hill, the Chinese National Human Genome Center at Shanghai, and the Institute of Nutrition and Food Safety at the China Centers for Disease Control. All participants provided written informed consent.

Cardiovascular Health Study (CHS). Approval was obtained from the Institutional Review Boards at Wake Forest University, University of California, Davis, Johns Hopkins, University of Pittsburgh, and the University of Washington, Seattle. All participants provided written informed consent.

China Kadoorie Biobank (CKB). All participants provided written informed consent. Ethical approval was obtained from Oxford Tropical Research Ethics Committee (OxTREC) and from the Ethical Review Committees of the Chinese Centre for Disease Control and Prevention and the Chinese Academy of Medical Sciences/Peking Union Medical College.

Cebu Longitudinal Health and Nutrition Survey (CLHNS). Written informed consent was obtained from all participants. Study protocols were approved by the University of North Carolina Institutional review Board for the Protection of Human Subjects.

Diabetic Cohort and Singapore Prospective Study Program (DC/SP2). Study protocols were approved by the Singapore General Hospital Ethics Committee, and National University of Singapore Institutional Review Board. All participants provided written informed consent.

Durban Diabetes Study and Durban Diabetes Case Control (DDS/DCC). Approvals were granted by the Biomedical Research Ethics Committee at the University of KwaZulu-Natal and the UK National Research Ethics Service. All participants provided written informed consent.

deCODE genetics (DECODE). The study was approved by the Icelandic National Bioethics Committee (approval no. VSN-16-112) after evaluation by the Icelandic Data Protection Authority. We obtained written informed consent for all participants in this study who donated samples. All data processing complies with the Icelandic Data Protection Authority (no. PV_2017060950ÞS).

Diabetes Gene Discovery Group (DGDG). All participants signed informed consent, and the protocol was approved by the French ethics committee.

Diabetes Genetics Initiative (DGI). The study was approved by the Ethics Committees of the Helsinki University Hospital, Helsinki, Finland, and Lund University, Sweden.Estonian Genome Center of the University of Tartu (EGCUT). All analyses were approved by the Ethics Review Committee of the University of Tartu. All participants provided written informed consent.

Electronic Medical Records and Genomics Network (EMERGE). Approval was obtained from the Institutional Review Boards at Boston Children's Hospital, Children's Hospital of Philadelphia, Cincinnati Children's Hospital Medical Center, Essentia Institute of Rural Health, Geisinger Clinic, Group Health Cooperative, Marshfield Clinic Research Foundation, Mayo Clinic, Icahn School of Medicine at Mount Sinai, Northwestern University, Pennsylvania State University, Vanderbilt University Medical Center, and University of Washington. All participants provided written informed consent.

European Prospective Investigation into Cancer and Nutrition (EPIC-INTERACT). The EPIC-InterAct study was approved by the local ethics committee in the participating countries and the Internal Review Board of the International Agency for Research on Cancer. All participants gave written informed consent. The study was coordinated by the Medical Research Council Epidemiology Unit at the University of Cambridge.

Epidemiologic Study of the Screenees for Diabetes Reduction Assessment with Ramipril and Rosiglitazone Medication (EPIDREAM). All study participants consented to analysis of blood samples. Approval was granted by the Hamilton Integrated Research Ethics Board, at McMaster University, Hamilton, Canada.

Family Heart Study (FAMHS). Approval was obtained from the Institutional Review Board at Washington University, St. Louis. Written informed consent, including consent to participate in genetic studies, was obtained from all participants.

Framingham Heart Study (FHS). Approval was obtained from the Institutional review Board of Boston University Medical Campus. All study participants provided written informed consent.

Finland-United States Investigation of NIDDM Genetics (FUSION). Approval was obtained from the coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa. All participants provided written informed consent.

German Chronic Kidney Disease (GCKD). All participants provided written informed consent. The study was registered in the national registry for clinical studies (DRKS 00003971) and was approved by local ethics committees.

Genetic Study of Atherosclerosis Risk (GENESTAR). Approval was obtained from the Johns Hopkins Medicine Institutional Review Board. All participants gave written informed consent.

Genetic Epidemiology Network of Arteriosclerosis (GENOA). Approval was granted by Institutional Review Boards of the University of Michigan, University of Mississippi Medical Center and Mayo Clinic. Written informed consent was obtained from all participants.

Resource for Genetic Epidemiology on Adult Heath and Aging (GERA). The Institutional Review Boards for Human Subjects Research of both Kaiser Permanente Medical Care Plan (Northern California Region) and the University of California at San Francisco approved the project.

Genetics of Diabetes and Audit Research in Tayside Scotland (GODARTS). Approval was obtained from the Tayside Medical Ethics Committee. Informed consent was obtained for all participants.

Genetics of Latinos Diabetic Retinopathy (GOLDR). Approval was granted by the Institutional Review Board of the Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center.

Genetic Overlap Between Metabolic and Psychiatric Traits and Teens of Attica: Genes and Environment (GOMAP-TEENAGE). Ethical permission for TEENAGE was obtained from the Bioethics Committee of Harokopio University, Athens. Ethical permission for GOMAP was obtained from the Dromokaiteio Scientific Committee, Dromokaiteio Management Committee, Dafni Scientific Committee, Eginitio Scientific Committee and Harokopio Ethics Committee. All participants of GOMAP-TEENAGE gave written informed consent.

Genomic Research Cohort for CCMB Diabetes Study (GRCCDS). Ethics committees of CSIR-Centre for Cellular and Molecular Biology and KEM Hospital and Research Centre approved the project.

Health, Aging and Body Composition Study (HABC). The Institutional Review Boards at the University of Memphis and the University of Pittsburgh granted approval to conduct the Health ABC Study, and all participants provided written informed consent.

Healthy Aging in Neighborhoods of Diversity Across the Life Span Study (HANDLS). Approval was granted by the National Institutes of Health Institutional Review Board (study number 09AGN248). All participants provided written informed consent.

Hispanic Community Health Study/Study of Latinos (HCHS/SOL). Approval was obtained from Institutional Review Boards at the University of North Carolina at Chapel Hill, Albert Einstein College of Medicine, University of Illinois at Chicago, University of Miami, and San Diego State University. All participants provided written informed consent.

Health Professionals' Follow-Up Study (HPFS). Approval was obtained from the Human Research Committee at the Brigham and Women's Hospital. All participants provided written informed consent.

Mexican American Hypertension and Insulin Resistance (HTNIR). Approval was granted by Human Subjects Protection Institutional Review Boards at the University of California at LosAngeles, University of Southern California, Lundquist/LABioMed/Harbor-UCLA and Cedars-Sinai Medical Center.

Howard University Family Study (HUFS). All human participants from the HUFS included in the analyses of this manuscript provided written informed consent prior to enrollment. The HUFS study was approved by the Institutional Review Board at Howard University.

INTERHEART (INTERHEART). All study participants consented to analysis of blood samples. Approval was granted by the Hamilton Integrated Research Ethics Board, at McMaster University, Hamilton, Canada.

Jackson Heart Study (JHS). Approval was obtained from Institutional Review Boards at Jackson State University, Tougaloo College and the University of Mississippi Medical Center. All participants provided written informed consent.

Korean Association Resource (KARE). Approval was granted by the Institutional review Board at the Korean National Institute of Health. All participants provided written informed consent.

Korean Biobank Array from the Korean Genome and Epidemiology (KoGES) Consortium (KBA). Approval was granted by the Institutional Review Board of the Korean National Institute of Health. All participants provided written informed consent.

Collaborative Health Research in the Region of Augsburg (KORA). Approval was granted by the Ethics Committee of the Medical Association of Bavaria (number 06068). All participants provided informed consent.

Los Angeles Latino Eye Study (LALES). Approval was obtained from the Los Angeles County/University of Southern California Institutional Review Board, and Western Institutional Review Board at Southern California Eye Institute. All participants provided written informed consent.

London Life Sciences Prospective Population (LOLIPOP). Approval was obtained from the London-Fulham Research Ethics Committee (ref 07/H0712/150). All participants gave an written informed consent.

Mexican American Study of Coronary Artery Disease (MACAD). Approval was granted by Human Subjects Protection Institutional Review Boards at the University of California at Los Angeles, University of Southern California, Lundquist/LABioMed/Harbor-UCLA and Cedars-Sinai Medical Center.Mexico City (MC). Approval was obtained from Institutional Review Boards at the Ethics and Scientific Commission members and the AUTHORIZATION is issued with registration number R-2011-785-018 and the Conacyt SALUD-2010-02-150352. In Canada, approval was obtained from the Research Ethics Board from the University of Toronto (Protocol 15770).

Multi-Ethnic Study of Atherosclerosis (MESA). Approval was obtained from Institutional Review Boards at the University of Washington, Wake Forest School of Medicine, Northwestern University, University of Minnesota, Columbia University, Johns Hopkins University, Cedars-Sinai Medical Center, and the University of California at Los Angeles.

Metabolic Syndrome in Men (METSIM). Approval was granted by the Ethics Committee of the University of Kuopio and the Kuopio University Hospital. All participants gave written informed consent.

Mass General Brigham Biobank (MGB). The MGB Biobank protocol and informed consent documents are reviewed annually by the Partners-MGB Institutional Review Board (#2009P002312). All patients who participate in the MGB Biobank are consented for their samples to be linked to their identified clinical information. They have also consented for their information to be used for a broad range of research and for their deidentified information to be shared outside of MGB.

Michigan Genomics Initiative (MGI). Approval was granted by the IRBMED Institutional Review Board of the University of Michigan. All participants gave written informed consent.

VA Million Veteran Program (MVP). All participating studies were conducted in compliance with the Declaration of Helsinki and comply with all relevant ethical and local regulatory requirements. Specifically, the contributing genetic association studies were approved by the Department of Veteran's Affairs central IRB.

Nagahama Study (NAGAHAMA). Approval was granted by the ethics committees of Kyoto University Graduate School of Medicine. Written informed consent was obtained from all participants.

Netherlands Epidemiology of Obesity (NEO). Approval was obtained from the Medical Ethics Committee of Leiden University Medical Center. All participants gave written informed consent.

Nurses Health Study (NHS). Approval was obtained from the Human Research Committee at the Brigham and Women's Hospital. All participants provided written informed consent.

NIDDM-Atherosclerosis Study Hispanic Cohorts (NIDDM). Approval was granted by Human Subjects Protection Institutional Review Boards at the University of California at Los Angeles, University of Southern California, City of Hope, Lundquist/LABioMed/Harbor-UCLA and Cedars-Sinai Medical Center.

Northwestern University Genetics (NUGENE). Approval was obtained from Institutional Review Boards at Northwestern University and Vanderbilt University

Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS). Approval was granted by the Ethics Committee of Uppsala University. All participants provided written informed consent.

Pakistan Risk of Myocardial Infarction Study (PROMIS). The study was approved by the Institutional Review Board of the Center for Non-Communicable Diseases Pakistan and by regional Ethical Review Committees in the different centres across Pakistan involved in the study. Institutional Review Boards at the National Institute of Cardiovascular Disorders, Karachi, Punjab Institute of Cardiology, Lahore, and Tabba Heart Institute, Karachi approved the study. All participants provided written informed consent.

Prospective Study of Pravastatin in the Elderly at Risk (PROSPER). Approval was obtained from the Institutional Ethics Review Boards of Cork University (Ireland), Glasgow University (UK) and Leiden University Medical Center (The Netherlands). All participants gave written informed consent.

Sea Islands Genetic Network Reasons for Geographic and Racial Differences in Stroke (REGARDS). The REGARDS study protocol was approved by the institutional review boards of each participating institution, and written informed consents were obtained from all participants.

Ragama Health Study (RHS). Approval was obtained from Institutional Review Boards at the National Center for Global Health and the University of Kelaniya (P38/09/2006). All participants provided written informed consent.
Rotterdam Study (RS). Approval was granted by the Institutional review Board at Erasmus University Medical Center. All participants provided written informed consent.

Shanghai Breast Cancer Study and Shanghai Women's Health Study (SBCS/SWHS). Approval was obtained from Institutional review Boards at Vanderbilt University Medical Center and Shanghai Cancer Institute. A written informed consent form was obtained from all study participants.

Singapore Chinese Eye Study (SCES). The study adhered to the Declaration of Helsinki. Ethical approval was obtained from the SingHealth Institutional Review Board and National University of Singapore Institutional Review Board. Written informed consent was obtained from all participants.

Starr County Health (SCH). All protocols were reviewed and approved by the Institutional Committee for the Protection of Human Subjects (HSC-SPH-02-042). All participants provided written informed consent permitting the collection and sharing of data.

Singapore Chinese Health Study (SCHS). Approval was obtained from the Institutional Review Board at the National University of Singapore. All participants provided written informed consent.

Slim Initiative for Genomic Medicine in the Americas (SIGMA). Approval was obtained from the Institutional Review Board of the Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran. All participants provided written informed consent.

Singapore Malay Eye Study (SIMES). The study adhered to the Declaration of Helsinki. Ethical approval was obtained from the SingHealth Institutional Review Board and National University of Singapore Institutional Review Board. Written informed consent was obtained from all participants.

Singapore Indian Eye Study (SINDI). The study adhered to the Declaration of Helsinki. Ethical approval was obtained from the SingHealth Institutional Review Board and National University of Singapore Institutional Review Board. Written informed consent was obtained from all participants.

Samsung Medical Center (SMC). Approval was obtained from the Institutional Review Board of the Samsung Medical Center (No. 2004-12-005). All participants provided written informed consent.

Seoul National University Hospital (SNUH). The Institutional Review Board of the Biomedical Research Institute at Seoul National University Hospital approved the study protocol (1205–130–411). Written informed consent was obtained from each participant.

Taiwan Metabochip Consortium Zhonghua (TAICHI-G). Approval was granted by Institutional Review Boards at Stanford University School of Medicine, Hudson-Alpha Biotechnology Institute, Lundquist/LABioMed/Harbor-UCLA, Cedars-Sinai Medical Center, Taichung Veterans General Hospital, Taipei Veterans General Hospital, National Health Research Institute, Tri-Service General Hospital, and National Taiwan University Hospital.

Taiwan Type 2 Diabetes (TWT2D). Approval was obtained from Institutional Review Boards at China Medical University Hospital, Chia-Yi Christian Hospital, and National Taiwan University Hospital.

Danish T2D Case-Control Study (UCPH). The studies included in the Danish T2D Case-Control Study (UCPH) were conducted in accordance with the Declaration of Helsinki II and were approved by the local Ethical Committees of Copenhagen County, the Capital Region of Denmark, or the Region of Southern Denmark.

UK Biobank (UKBB). Approval was obtained from the North West Centre for Research Ethics Committee (11/NW/0382).

Uppsala Longitudinal Study of Adult Men (ULSAM). Approval was granted by the Ethics Committee of Uppsala University. All participants provided written informed consent.

Wake Forest School of Medicine (WFSM). Approval was granted by the Institutional Review Board at Wake Forest School of Medicine. All participants provided written informed consent.

Women's Health Initiative (WHI). Approval was granted by the Institutional review Board at the Fred Hutchinson Cancer Research Centre in accordance with the US Department of Health and Human Services regulations at 45 CFR 46 (approval number IR# 3467-EXT). All participants provided written informed consent. Additional written consent to review medical records was obtained. The Fred Hutchinson Cancer Research Centre has an approved FWA on file with the Office for Human Research Protections under assurance number 0001920.

Wellcome Trust Case Control Consortium (WTCCC). Approval for the study was obtained from Peterborough & Fenland Local Research Ethics Committee, National Research Ethics Service, Leeds (East) Research Ethics Committee, South West Multicentre Research Ethics Committee, Tayside Committee on Medical Research Ethics and Oxford Tropical Research Ethics Committee.

Members of the ENSA Genomics Consortium

Last name	Name	Affiliation	Present address
Aguilar-Salinas	Carlos	Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ)	
Barberena-Jonás	Carmina	Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav)	
Canizales-Quintero	Sergio	Instituto Nacional de Salud Pública (INSP)	
Cedillo-Castelán	Vianka	Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav)	
Cruz-Hervert	Luis Pablo	Instituto Nacional de Salud Pública (INSP)	Universidad Nacional Autónoma de México
Delgado-Sánchez	Guadalupe	Instituto Nacional de Salud Pública (INSP)	
Ferreira-Guerrero	Elizabeth	Instituto Nacional de Salud Pública (INSP)	
Ferreyra-Reyes	Leticia	Instituto Nacional de Salud Pública (INSP)	
García-García	Lourdes	Instituto Nacional de Salud Pública (INSP)	
Gutiérrez-López	Cecilia	Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav)	
Hernández-Avila	Juan Eugenio	Instituto Nacional de Salud Pública (INSP)	
Huerta-Chagoya	Alicia	Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ)	Broad Institute of Harvard and MIT
Juárez-Figueroa	Luis	Instituto Nacional de Salud Pública (INSP)	Deceased
Kuri-Morales	Pablo	Secretaría de Salud	Tecnológico de Monterrey
Lazcano-Ponce	Eduardo	Instituto Nacional de Salud Pública (INSP)	
Magis-Rodriguéz	Carlos	Secretaría de Salud	Universidad Nacional Autónoma de México
Mongua-Rodriguéz	Norma	Instituto Nacional de Salud Pública (INSP)	
Moreno-Estrada	Andres	Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav)	
Moreno-Macias	Hortensia	Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ)	Universidad Autónoma Metropolitana
Ordóñez-Sánchez	Maria Luisa	Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ)	
Ortega-Estrada	María de Jesús	Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav)	
Palma-Martínez	María José	Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav)	Universidad Nacional Autónoma de México
Quinto-Cortes	Consuelo	Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav)	
Rodríguez-Guillen	Rosario	Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ)	
Sarti-Gutíerrez	Elsa	Instituto Nacional de Salud Pública (INSP)	Retired
Sepúlveda-Amor	Jaime	Instituto Nacional de Salud Pública (INSP)	University of California San Francisco (UCSF)
Sohail	Mashaal	Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav)	Universidad Nacional Autónoma de México
Tapia-Conyer	Roberto	Secretaría de Salud	Fundación Carlos Slim
Tellez-Vazquez	Norma	Instituto Nacional de Salud Pública (INSP)	
Tusié-Luna	Maria Teresa	Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ)	
Valdespino-Gómez	José Luis	Instituto Nacional de Salud Pública (INSP)	Deceased
Velázquez-Meza	Manuel	Instituto Nacional de Salud Pública (INSP)	
Velázquez-Monroy	Oscar	Secretaría de Salud	Deceased

Genes & Health Research Team authorship for Scientific Publications

Eamonn Maher e.maher@aston.ac.uk Aston University
Shabana Chaudhary s.chaudhary@qmul.ac.uk Blizard Institute, Queen Mary University of London
Joseph Gafton j.gafton@qmul.ac.uk Blizard Institute, Queen Mary University of London
Karen A Hunt k.a.hunt@qmul.ac.uk Blizard Institute, Queen Mary University of London
Shapna Hussain shapna.hussain@qmul.ac.uk Blizard Institute, Queen Mary University of London
Kamrul Islam k.islam@qmul.ac.uk Blizard Institute, Queen Mary University of London
Mohammed Bodrul Mazid m.b.mazid@qmul.ac.uk Blizard Institute, Queen Mary University of London
Elizabeth Owor e.owor@qmul.ac.uk Blizard Institute, Queen Mary University of London
Jessry Russell jessry.russell@qmul.ac.uk Blizard Institute, Queen Mary University of London
Nishat Safa n.safa@qmul.ac.uk Blizard Institute, Queen Mary University of London
John Solly j.solly@qmul.ac.uk Blizard Institute, Queen Mary University of London
Marie Spreckley m.spreckley@qmul.ac.uk Blizard Institute, Queen Mary University of London
David A Van Heel d.vanheel@qmul.ac.uk Blizard Institute, Queen Mary University of London
Jan Whalley j.whalley@qmul.ac.uk Blizard Institute, Queen Mary University of London
Ishevanhu Zengeya i.zengeya@qmul.ac.uk Blizard Institute, Queen Mary University of London
Emily Mantle e.mantle@qmul.ac.uk Blizard Institute, Queen Mary University of London
Shaheen Akhtar shaheen.akhtar@bthft.nhs.uk Bradford Teaching Hospitals NHS Foundation Trust
Samina Ashraf samina.ashraf@bthft.nhs.uk Bradford Teaching Hospitals NHS Foundation Trust
Dan Mason dan.mason@bthft.nhs.uk Bradford Teaching Hospitals NHS Foundation Trust
John Wright john.wright@bthft.nhs.uk Bradford Teaching Hospitals NHS Foundation Trust
Daniel MacArthur d.macarthur@garvan.au.org Garvan Institute
Michael Simpson michael.simpson@kcl.ac.uk King's College London
Richard C Trembath richard.trembath@kcl.ac.uk King's College London
Gerome Breen gerome.breen@kcl.ac.uk Kings College London
Raymond Chung raymond.chung@kcl.ac.uk Kings College London
Sang Hyuck Lee sang_hyuck.lee@kcl.ac.uk Kings College London
Omar Asgar omar.asghar1@nhs.net Manchester University Hospitals
Joanne Harvey joanne.henry@nihr.ac.uk Manchester University Hospitals
Karen Tricker karen.tricker@mft.nhs.uk Manchester University Hospitals
Caroline Winckley caroline.winckley@nihr.ac.uk Manchester University Hospitals
Hanifa Khatun hanifa.khatun@mft.nhs.uk Manchester University Hospitals
Amna Asif amna.asif@mft.nhs.uk Manchester University Hospitals
Claudia Langenberg claudia.langenberg@qmul.ac.uk Precision Healthcare University Research Institute, Queen Mary University of London
Grainne Colligan grainnec@safh.org.uk Social Action for Health (charity)
Ceri Durham cerid@safh.org.uk Social Action for Health (charity)
Bill Newman william.newman@manchester.ac.uk University of Manchester
Ahsan Khan cllrahsan.khan@walthamforest.gov.uk Waltham Forest Council
Hilary Martin hilary.martin@qmul.ac.uk Wellcome Sanger Institute
Teng Heng th13@sanger.ac.uk Wellcome Sanger Institute
Matt Hurles meh@sanger.ac.uk Wellcome Sanger Institute
Vivek Iyer vvi@sanger.ac.uk Wellcome Sanger Institute
Georgios Kalantzis gk18@sanger.ac.uk Wellcome Sanger Institute
Vladimir Ovchinnikov vo3@sanger.ac.uk Wellcome Sanger Institute
Iaroslav Popov ip13@sanger.ac.uk Wellcome Sanger Institute
Klaudia Walter kw8@sanger.ac.uk Wellcome Sanger Institute
Panos Deloukas p.deloukas@qmul.ac.uk William Harvey Research Institute, Queen Mary University of London
David Collier d.j.collier@qmul.ac.uk William Harvey Research Institute, Queen Mary Universityof London
Ana Angel a.cristinaangelgarcia@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Saeed Bidi saeed.bidi@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Fabiola Eto f.eto@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London

Sarah Finer s.finer@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Chris Griffiths c.j.griffiths@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Sam Hodgson s.hodgson@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Benjamin M Jacobs b.jacobs@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Rohini Mathur r.mathur@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Caroline Morton c.morton@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Asma Qureshi asmaa.qureshi@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Stuart Rison s.rison@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Annum Salman a.salman@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Miriam Samuel m.samuel@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Moneeza K Siddiqui moneeza.siddiqui@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Daniel Stow d.stow@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Sabina Yasmin sabina.yasmin@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Julia Zöllner j.zollner@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London
Sheik Dowlut s.dowlut@qmul.ac.uk Wolfson Institute of Population Health, Queen Mary University of London

VA Million Veteran Program
Core Acknowledgements for Publications
October 2025

MVP Program Office

- Sumitra Muralidhar, Ph.D., Program Director
US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420
- Jennifer Moser, Ph.D., Associate Director, Scientific Programs
US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420
- Jennifer E. Deen, B.S., Associate Director, Cohort & Public Relations
US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

MVP Steering Committee

- Co-Chair: Philip S. Tsao, Ph.D.
VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304 -
Co-Chair: Sumitra Muralidhar, Ph.D.
US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420 -
J. Michael Gaziano, M.D., M.P.H.
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Adriana Hung, M.D., M.P.H.,
VA Tennessee Valley Healthcare System, 1310 24th Avenue, South Nashville, TN 37212 -
Dave Oslin, M.D.
Philadelphia VA Medical Center, 3900 Woodland Avenue, Philadelphia, PA 19104 -
Deepak Voora, M.D.
Durham VA Medical Center, 508 Fulton Street, Durham, NC 27705

MVP Co-Principal Investigators

- J. Michael Gaziano, M.D., M.P.H.
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130 -
Philip S. Tsao, Ph.D.
VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304

MVP Core Operations

- Jessica V. Brewer, M.P.H., Director, MVP Cohort Operations
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Mary T. Brophy M.D., M.P.H., Director, VA Central Biorepository
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Kelly Cho, M.P.H, Ph.D., Director, MVP Phenomics
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Lori Churby, B.S., Director, MVP Regulatory Affairs
VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304
- Jacob T. Kean, Ph.D., Acting Director, VA Informatics and Computing Infrastructure (VINCI)

VA Salt Lake City Health Care System, 500 Foothill Drive, Salt Lake City, UT 84148

- Saiju Pyarajan Ph.D., Director, Data and Computational Sciences
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Robert Ringer, Pharm.D., Director, VA Albuquerque Central Biorepository
New Mexico VA Health Care System, 1501 San Pedro Drive SE, Albuquerque, NM 87108
- Luis E. Selva, Ph.D., Director, MVP Biorepository Coordination
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Shahpoor (Alex) Shayan, M.S., Director, MVP PRE Informatics
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Brady Stephens, M.S., Principal Investigator, MVP Information Center
Canandaigua VA Medical Center, 400 Fort Hill Avenue, Canandaigua, NY 14424
- Stacey B. Whitbourne, Ph.D., Director, MVP Cohort Development and Management VA
Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130