

Polygenic score informed by genome-wide association studies of multiple ancestries and related traits improves risk prediction for coronary artery disease

Aniruddh P. Patel, MD^{1-5*}; Minxian Wang, PhD^{6*}; Yunfeng Ruan, PhD^{2,3}; Satoshi Koyama, MD, PhD^{2,3,7}; Shoa L. Clarke, MD, PhD^{8,9}; Xiong Yang, PhD⁶; Catherine Tcheandjieu, PhD¹⁰; Saaket Agrawal, BS^{2,3,11}; Akl C. Fahed, MD, MPH¹⁻⁵; Patrick T. Ellinor, MD, PhD¹⁻⁵; Genes & Health Research Team; the Million Veteran Program; Phillip S. Tsao, PhD^{8,9}; Yan V. Sun, PhD¹²; Kelly Cho, PhD⁷; Peter W. F. Wilson, MD¹²; Themistocles L. Assimes MD, PhD^{8,9}; David A. van Heel MD, PhD¹³; Adam S. Butterworth, PhD¹⁴; Krishna G. Aragam, MD¹⁻⁵; Pradeep Natarajan, MD, MMSc^{1-5#}; Amit V. Khera, MD, MSc^{2-4,15#}

Affiliations:

¹Division of Cardiology, Department of Medicine, Massachusetts General Hospital, Boston, MA

²Center for Genomic Medicine, Department of Medicine, Massachusetts General Hospital, Boston, MA

³Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA

⁴Department of Medicine, Harvard Medical School, Boston, MA

⁵Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA

⁶CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing, China

⁷Veteran Affairs Boston Healthcare System, Boston, MA

⁸Stanford University School of Medicine, Palo Alto, CA

⁹Veterans Affairs Palo Alto Healthcare System, Palo Alto, CA

¹⁰Gladstone Institutes, San Francisco, CA

¹¹Feinberg School of Medicine, Northwestern University, Chicago, IL

¹²Veteran Affairs Atlanta Healthcare System, Decatur, GA

¹³Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK

¹⁴British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

¹⁵Verve Therapeutics, Boston, MA

*,#These authors contributed equally

Please address correspondence to:

Minxian Wang, PhD

E-mail: wangmx@big.ac.cn

OR

Amit V. Khera, MD, MSc

Center for Genomic Medicine, Massachusetts General Hospital

185 Cambridge Street | CPZN 6.256, Boston, MA 02114

Tel: 617-643-3388

E-mail: avkhera@mgh.harvard.edu

Abstract:

Accurate stratification of coronary artery disease (CAD) risk remains a critical need. A new polygenic score (GPS_{Mult}) incorporates CAD genome-wide association data across five ancestries (>269,000 cases, >1,178,000 controls) with genetic association data for ten CAD risk factors. GPS_{Mult} associates with an OR/SD 2.14, (95%CI:2.10-2.19, $P<0.001$) for prevalent CAD and HR/SD 1.73 (95%CI 1.70-1.76, $P<0.001$) for incident CAD. When compared with the previously published GPS_{2018} in external datasets, GPS_{Mult} demonstrated 73%, 46%, and 113% increase in effect size for individuals of African, European, and South Asian ancestry, respectively, and significantly outperformed recently published CAD polygenic scores. GPS_{Mult} identifies individuals with CAD risk extremes, including the top 3% of the population at equivalent risk for a new CAD event as those with prior CAD having a second event. Integrating GPS_{Mult} with the Pooled Cohort Equations results in 7.0% [95%CI:5.9%-8.2%, $P<0.001$] net reclassification improvement at the 7.5% threshold. Large-scale integration genetic association data for CAD and related traits from diverse populations meaningfully improves polygenic risk prediction.

INTRODUCTION:

Coronary artery disease (CAD) remains the leading cause of death worldwide and identification of at-risk individuals remains a critical public health need.¹ If identified early, at-risk individuals can benefit from more efficiently targeted lifestyle interventions and cholesterol-lowering medications toward lifelong risk mitigation.² However, commonly used clinical risk estimators for CAD were optimized for use in middle-aged adult populations in historical cohort studies and consequently underperform in younger populations or individuals of non-European ancestries.^{3–6} As CAD is a heritable disease, leveraging the increasing amount of widely available genetic data offers additional opportunities to significantly enhance CAD risk prediction across all groups early in life, particularly at the extremes of the risk distribution.^{7,8}

Polygenic scores integrate data from genome-wide association studies into a single quantitative and predictive metric of inherited risk.^{9–12} Several studies to date have stratified individuals into substantial gradients of CAD risk based on their polygenic score beyond their clinical risk factor profiles.^{13–17} Given this potential, polygenic scores are now already being deployed clinically across some biobanks and returned through direct-to-consumer testing platforms.^{18,19} The past decade has seen numerous advances in score development, however, there still remains room for improvement in their performance, particularly among individuals of non-European ancestry.²⁰ Simulation studies suggest that larger sample sizes of GWAS have the potential to more accurately estimate the effect size associated with each SNP to improve scores for CAD.²¹ Polygenic scores integrating GWAS data from individuals of diverse ancestries in addition to that of the target population show relative improvement in predictive accuracy compared with methods only utilizing GWAS data from a single ancestry source.^{22–26} Furthermore, the principles of genetic correlation suggest benefit in incorporating information from GWAS of related traits to refine polygenic prediction in the trait of interest.^{27–34}

Alongside considerable enthusiasm for polygenic scores to enable a new era of preventive clinical medicine is recognition of several key unmet needs before polygenic scores can be more widely implemented. First, polygenic scores have reduced predictive performance in individuals of non-European ancestry.³⁵ This largely stems from relative underrepresentation of other ancestries in prior GWAS discovery cohorts. Recent efforts have focused on conducting GWAS in larger and more ancestrally diverse populations and designing methods leveraging ancestry-specific linkage disequilibrium patterns to help improve score performance.^{25,36–42} Second, although available scores associate strongly with prevalent disease, they perform less well in predicting incident disease, which would offer more clinical utility.¹⁴ Finally, most risk prediction models to date are based either on genetic or clinical risk factors, but better integration of these modalities and estimation of a clinically actionable risk estimate is needed.^{43–45}

To address these needs, we used information from five-fold larger and more ancestrally diverse GWAS compilation compared to prior efforts along with methods leveraging commonalities in mechanistic pathways to develop a new polygenic risk score for CAD.

RESULTS:

Summary statistics from GWAS for CAD, other atherosclerotic diseases, and their risk factors across over 1.2 million individuals from multi-ancestry cohorts were aggregated to design polygenic risk scores for CAD (Figure 1, Supplementary Table 1). These scores were trained within the UK Biobank cohort in 116,649 individuals of European ancestry and

then validated in the remaining independent study population of 325,991 individuals (54.3% female, 7281 African, 1,464 East Asian, 308,264 European, and 8,982 South Asian ancestry) (Supplementary Table 2).⁴⁶ The participants in the training and validation cohorts are independent from the individuals analyzed in the previously conducted GWAS from which summary statistics were obtained. A total of 58 ancestry- and trait-specific scores were included in the GPS training analysis, with 32 scores significantly contributing to overall prediction after optimization of score selection and weighting through logistic regression (Figure 2A and 2B).

Association of GPS_{Mult} with prevalent disease in UK Biobank

The resulting best performing score (GPS_{Mult}) demonstrated a strong association with prevalent CAD, with significant improvement from previously published scores. Among 308,264 European ancestry individuals in the hold-out validation dataset, GPS_{Mult} was associated with an odds ratio per standard deviation increase (OR/SD) of 2.14 (95%CI:2.10-2.19) in a model adjusted for age, sex, genotyping array, and the first ten principal components of genetic ancestry, with significant improvement over from prior published scores from the Polygenic Score Catalog without UK Biobank participants in discovery data (Supplementary Table 3).⁴⁷ This corresponded to a Nagelkerke R^2 of 0.07 and a logit liability R^2 of 0.19 (Supplementary Figure 1). After adjusting for correlated clinical risk factors including systolic and diastolic blood pressure, LDL cholesterol, HDL cholesterol, triglycerides, diabetes, body mass index, and chronic kidney disease, this risk estimate was only modestly attenuated to an OR/SD 2.07 (95% CI 2.02-2.13) (Supplementary Table 4). The associations between GPS_{Mult} and CAD were largely consistent across studied subgroups, but some evidence of heterogeneity was found when restricting to men (OR/SD 2.20, 95% CI 2.15-2.26, $p < 0.001$) when compared with women (OR/SD 1.94, 95% CI 1.86-2.03, $p < 0.001$), with p -heterogeneity < 0.001 (Supplementary Figure 2). Additionally, the association between GPS_{Mult} and CAD was stronger in younger individuals ages 40-54 years (OR/SD 2.17, 95%CI 2.04-2.31, $p < 0.001$) and 55-64 years (OR/SD 2.18, 95%CI 2.11-2.25, $p < 0.001$), when compared with older individuals ages 65-75 years (OR/SD 2.08, 95%CI 2.01-2.15, $p < 0.001$), consistent with recent studies (Supplementary Figure 2).^{7,48-50}

GPS_{Mult} showed stronger association with CAD risk when compared with the previously published GPS₂₀₁₈¹⁴ in direct comparison using the same group of individuals for validation. Among individuals with CAD, the median percentile of GPS_{Mult} is significantly higher than that of the GPS₂₀₁₈, 75 (IQR 50 - 91) vs 69 (IQR 43 - 88) (Figure 3A). Among individuals of European ancestry, individuals in the bottom and top centile of the polygenic score had a 0.8% and 12.3% prevalence of CAD, respectively, with GPS₂₀₁₈, compared with 0.7% and 16.3% prevalence of CAD with and GPS_{Mult} (Figure 3B). Given improved stratification with this newly developed polygenic score, both tails of the score distribution were associated with a greater magnitude of risk when compared with GPS₂₀₁₈. With the GPS₂₀₁₈, the top 8.3%, 3.1%, and 1.4% of the population had 3-fold, 4-fold, and 5-fold greater odds for CAD relative to the middle quintile of the population, respectively, whereas with the GPS_{Mult}, the top 20%, 9.6%, and 4.9% of the population had 3-fold, 4-fold, and 5-fold greater odds for CAD relative to the middle quintile of the population, respectively (Figure 3C, Supplementary Table 5). Conversely, with the GPS₂₀₁₈, the bottom 1.7%, 0.5%, and 0.1% of the population had 1/3, 1/4, and 1/5 the odds for CAD relative to the middle quintile of the population, respectively, whereas with the GPS_{Mult}, the bottom 13.9%, 1.7%, and 0.2% of the population had 1/3, 1/4, and 1/5 odds for CAD relative to the middle quintile of the population, respectively (Figure 3D).

Validation of GPS_{Mult} in external cohorts

GPS_{Mult} was also strongly associated with prevalent CAD in external cohorts, with significant improvement from prior published scores. Published polygenic scores for CAD from the Polygenic Score Catalog and GPS_{Mult} were calculated in an identical group of individuals to facilitate direct comparison within individuals of African and European Ancestry in Million Veteran Program⁵¹ and South Asian ancestry in Genes & Health⁵² (Figure 4, Supplementary Tables 6-7). For each group, these individuals were not included in published GWAS summary statistics^{39,53} used for GPS_{Mult} derivation. Among 33,096 individuals of African ancestry in the Million Veterans Program, GPS_{Mult} was associated with an OR/SD of 1.24 (95% CI 1.21-1.29, $P < 0.001$) for CAD in a model adjusted for age, sex, genotyping array, and the first ten principal components of genetic ancestry, corresponding in a 73% relative improvement in effect size compared with GPS₂₀₁₈ and 39% improvement when compared with the recently published PRS₂₀₂₂,⁹ respectively ($P = 0.008$). Similarly, among 124,467 individuals of European ancestry in the Million Veteran Program, GPS_{Mult} was associated with an OR/SD of 1.72 (95% CI 1.69-1.75, $P < 0.001$), corresponding in a 46% and 13.6% relative improvement in effect size compared with GPS₂₀₁₈ and PRS₂₀₂₂,⁹ respectively ($P < 0.001$). Additionally, among 27,990 individuals of South Asian ancestry in Genes and Health, GPS_{Mult} was associated with an OR/SD of 1.83 (95% CI 1.69-1.99, $P < 0.02$), corresponding to a 113% ($P < 0.001$) and 29% ($P = 0.02$) relative improvement in effect size compared with GPS₂₀₁₈ and PRS₂₀₂₂, respectively (Figure 4).

Association of GPS_{Mult} with incident disease in UK Biobank

The GPS_{Mult} was predictive of incident CAD events over median [interquartile range] 12.0 [11.2-12.7] years of follow-up across all four ancestral groups in the UK Biobank. Across the entire UK Biobank study validation population without prior CAD, individuals in the bottom centile of the GPS_{Mult} had a 1.1% incidence of CAD while individuals in the top centile had a 11.7% incidence of CAD. Overall, GPS_{Mult} was associated with a hazard ratio per standard deviation (HR/SD) of 1.73 (95% CI 1.70-1.76, $P < 0.001$), compared with HR 1.49 (95% CI 1.47-1.52, $p < 0.001$) found with GPS₂₀₁₈. When stratified by ancestry, risk estimates remained consistent across individuals of East Asian (HR/SD 1.72, 95% CI 1.13-2.60, $P = 0.011$), European (HR/SD 1.75, 95% CI 1.71-1.78, $P < 0.001$), and South Asian (HR/SD 1.62, 95% CI 1.49-1.77, $P < 0.001$) ancestry, but score performance was weakest among individuals of African (HR/SD 1.25, 95% CI 1.07-1.46, $p = 0.004$) ancestry (Figure 5A). Across all individuals in the UK Biobank validation dataset, GPS_{Mult} demonstrated 38% relative improvement in effect size compared with GPS₂₀₁₈. Of this, 26% improvement resulted from larger sample size of the primary CARDIOGRAMplusC4D GWAS (excluding UK Biobank participants), 9% improvement from incorporation of multi-ancestry CAD summary statistics, and 3% improvement from leveraging genetic commonalities with CAD risk factors to refine score weighting (Figure 5B). Incorporation of multi-ancestry and multi-trait genetic data resulted in greater relative gains in incident disease prediction for individuals in each ancestry, with improved relative effect sizes of 143%, 71%, 38%, and 23% for individuals of African, East Asian, European, and South Asian ancestry, respectively, compared to GPS₂₀₁₈ performance in those groups. These also translated into significant gains in prediction by GPS_{Mult} relative to the initial GPS₂₀₁₈ performance in European ancestry, now with improved prediction in African ancestry (relative effect size 0.55 increased from 0.23) (Figure 5B) and performance surpassing the reference score in

East Asian ancestry (relative effect size 1.37, increased from 0.80) and South Asian ancestry (relative effect size 1.19, increased from 0.97).

Assessment of disease risk in the extremes of the GPS_{Mult} distribution

Additionally, we hypothesized that the GPS_{Mult} could identify individuals in the clinically relevant extreme tails of its distribution. Current cardiovascular disease prevention guidelines recommend statin therapy for individuals with prior coronary artery disease, peripheral artery disease, ischemic stroke, diabetes mellitus, or severe hypercholesterolemia (LDL \geq 190 mg/dL) to help mitigate their high risk of cardiovascular disease and mortality.² In the high end of GPS_{Mult} , we sought to identify individuals with genetic risk of equivalent magnitude to that of individuals with these clear indications for statin therapy. In prospective analyses of individuals without prior CAD, those within the top 3 percentiles of GPS_{Mult} had equivalent disease risk of incident CAD as the recurrent event risk for an individual who had a CAD event prior to enrollment, adjusting for age and sex (Supplementary Figure 3A). Furthermore, individuals without peripheral artery disease (PAD) in the top 8% of polygenic score distribution had incident CAD risk equivalent to individuals with prior PAD; individuals without diabetes in the top 21% of polygenic score distribution had incident CAD risk equivalent to individuals with prior diabetes; and individuals without severe hypercholesterolemia (estimated untreated LDL cholesterol \geq 190 mg/dL) in the top 29% of polygenic score distribution had incident CAD risk equivalent to individuals with prior hypercholesterolemia (Supplementary Figures 3B-D). Conversely, in the low end of the GPS_{Mult} distribution, individuals in the bottom 5 percentiles were associated with a significant reduction in incident CAD risk (HR 0.27, 95%CI 0.21-0.35, $P < 0.001$) when compared with the middle quintile (40-59%). When comparing individuals who smoke and are in the bottom 5 percentiles of GPS_{Mult} with non-smokers in the middle quintile, the associated reduction in the absolute incidence of CAD offsets approximately 60 pack-years of smoking. Furthermore, individuals in the 5-9th percentiles of GPS_{Mult} had a significant reduction in CAD risk (HR 0.55, 95%CI 0.49-0.62, $P < 0.001$) when compared with the middle quintile. These individuals experienced comparable risk reduction as those individuals carrying variants in *PCSK9* associated lifelong low levels of LDL cholesterol (Supplementary Figure 4).^{54,55}

Modeling of GPS_{Mult} with clinical risk predictors

A risk prediction approach integrating clinical and genetic risk using the American College of Cardiology/American Heart Association Pooled Cohort Equations (PCE),⁵ GPS_{Mult} , and their interaction in a single model was used to predict 10-year CAD risk estimates in the UK Biobank validation population. Accounting for the interaction between the polygenic score and clinical risk estimate improves performance beyond the simple addition of the two, with lower GPS_{Mult} weighting with higher PCE estimates (effect size - 0.60, $P_{interaction} < 0.001$). This combined model effectively improved risk prediction when compared with PCE alone. When binned by different PCE estimates, this model demonstrated striking stratification of CAD incidence across the GPS_{Mult} distribution, with significant differences observed in ancestry-based subgroups (Figure 6A). The gradient in risk predicted by this model from top to bottom centile was largest in South Asian ancestry individuals with high PCE risk (5.1% to 29.1%), compared with European ancestry individuals (2.6% to 20.6%). When compared with the PCE risk estimate incorporating clinical risk factors alone, integration of the PCE with GPS_{Mult} contributed to significantly higher discrimination and predictive performance across the entire tested population. First,

discrimination was assessed in Cox regression models including various covariables using Harrell's C-Statistic. A gradient in improvement was seen using models using age and sex alone (C-statistic 0.710, 95%CI 0.706 - 0.715), PCE which is inclusive of age and sex (C-statistic 0.739, 95% CI 0.735-0.744), and the model integrating PCE, GPS_{Mult} and their interaction term (C-statistic 0.763, 95%CI 0.759-0.768) (Figure 6B). Similar improvements in C-statistic were observed for models tested in subgroups stratified by ancestry (Supplementary Table 8). Second, categorized net reclassification improvement (NRI) was calculated across the entire study population using a threshold of 7.5% (NRI 0.075) of the predicted 10-year risk of CAD, which is the clinically accepted estimated risk threshold for recommending initiation of statin therapy for prevention of CAD. The risk model combining PCE and GPS_{Mult} resulted in significant improvements in the categorical net reclassification index (NRI = 7.0%, +8.1% for incident cases and -1.1% for non-cases), with GPS_{Mult} resulting in greater up classification of risk largely in individuals who go on to develop disease (Figure 6C). Third, when compared with established risk enhancing factors for CAD risk, categorization within the top 10 percentiles of the GPS_{Mult} distribution corresponded to a significantly higher net reclassification over the use of PCE estimate alone (3.7%) as compared to other risk enhancers like elevated lipoprotein(a) (with NRI 1.3%) (Supplementary Figure 5). Similar results in NRI were observed across other ancestries (Supplementary Table 9). Additionally, similar trends in predictive performance, discrimination, and reclassification were observed with integration of the QRISK score with GPS_{Mult} (Supplementary Tables 8-9).

Association of GPS_{Mult} with recurrent disease in UK Biobank

In addition to first events, the GPS_{Mult} predicted recurrent CAD events in individuals with prior CAD. GPS_{Mult} was associated with a HR/SD of 1.13 (95% CI 1.08-1.18, P<0.001), comparable to prior studies.⁵⁶ Although a significantly less pronounced effect estimate as compared to prediction of first CAD event, the predictive performance of GPS_{Mult} in this context was comparable to that of diastolic blood pressure (HR 1.11, 95%CI 1.06-1.16, P<0.001) and glycated hemoglobin (HR 1.07, 95%CI 1.02-1.12, P<0.001) (Supplementary Figure 6).

DISCUSSION:

A new polygenic score for CAD incorporating multi-ancestry summary statistics from GWAS for CAD and related risk factor traits on a large scale demonstrated significantly improved performance when compared to prior published scores. External validation in fully independent datasets derived from the Million Veterans Program and the Genes & Health studies, confirming enhanced prediction compared to previously published polygenic scores across all studied ancestries. The enhanced predictive capacity of this score was particularly pronounced in the extremes of the score distribution, enabling—in some cases—identification of healthy individuals with risk of CAD equivalent to those with pre-existing disease. When added to risk scores used in current clinical practice, GPS_{Mult} significantly improved discrimination and reclassification relevant to clinically important decision thresholds, such as decision to initiate statin therapy.

This work builds on prior studies in providing a framework for generating the best possible polygenic score for any trait, within the limitations of available GWAS with finite sample sizes and under-representation of diverse populations. The GPS_{Mult} incorporates CAD summary statistics from large non-European ancestry biobanks leading to a total CAD GWAS summary statistics of over 269,000 cases and over 1,178,000 controls, including many-fold larger representation of individuals of non-European ancestries than previously published efforts.^{39,52,57–59} This results in substantial improvements in prediction for individuals of East and South Asian ancestry, reflecting greater representation of summary statistics from Biobank Japan and Genes & Health. However, the majority of improvement in effect size is attributable to use of summary statistics from the largest CAD GWAS to date (CARDIOGRAMplusC4D consortium, excluding UK Biobank participants), particularly in European ancestry individuals.⁹ The modest improvements in prediction observed among individuals of African ancestry are likely due to underrepresentation of this group in GWASs to date.³⁵ Due to smaller haplotype blocks observed in individuals of African ancestry, a 4- to 7-fold larger GWAS representation is needed to yield comparable prediction gains.⁶⁰ The additional incorporation of genetic associations with CAD-related risk factors across ancestries into calculating GPS_{Mult} significantly improves prediction beyond using summary statistics from CAD GWAS alone, with impact most notable in individuals of non-European ancestry. This may potentially be due to greater representation of these ancestries in the discovery GWAS for CAD risk factor traits.^{61–63} With these additions, the phenotypic variance explained by GPS_{Mult} for CAD calculated as R^2 on the logit-liability scale was 0.19. Although this estimate remains below the estimated SNP heritability for CAD of 0.4 - 0.6, it surpasses the phenotypic variance explained of 0.14 by the largest component GWAS from the CARDIOGRAMplusC4D consortium.^{39,64,65}

Improvements in polygenic score performance can help better facilitate clinical decision making. Prospective trials are already underway returning polygenic risk information to patients,^{66,67} and medical societies have begun to provide provisional guidance on their use.⁶⁸ Furthering these goals, GPS_{Mult} is able to better identify individuals at the highest risk for developing incident CAD to potentially guide early preventive interventions.^{69,70} Building on prior work advocating for use of polygenic scores as a risk-enhancing factor to guide decision making regarding statin therapy in individuals at borderline or intermediate CAD risk, the current work more strongly supports use in primary screening across the population to target interventions.⁷¹ Current cardiovascular prevention guidelines recommend statin initiation for individuals solely based on having any of the following conditions as they portend a high risk of a new atherosclerotic cardiovascular disease event: prior CAD, ischemic stroke, PAD, diabetes, or severe hypercholesterolemia.² This score identified 3% of

the population with equivalent risk for a new CAD event as the risk for a recurrent CAD event in individuals who have had prior disease. Similarly, the top 8%, 21%, and 29% of the GPS_{Mult} distribution—despite having no known CAD—had equivalent risk of incident CAD as individuals with prior peripheral artery disease, diabetes mellitus, and severe hypercholesterolemia, respectively. Because all three of these designations are currently clinical indications for statin therapy, a high GPS_{Mult} could be employed to identify additional individuals for cholesterol-lowering therapies as an adjunct to current guidelines. Furthermore, given the GPS_{Mult} 's ability to identify these individuals with the highest propensity for developing CAD, these scores could be employed to enrich for high genetic risk individuals in CAD prevention trials to maximize event rates and minimize drug trial costs.⁷² The GPS_{Mult} could also be employed to identify the individuals with the highest risk of recurrent events for targeted, otherwise costly therapies which have been shown to be beneficial in this population.^{73,74} Additionally, GPS_{Mult} also identifies individuals in the lower end of genetic risk who are seemingly protected from CAD with similar risk reduction as that of carriers of variants in the *PCSK9* gene leading to lifelong reductions on low-density lipoprotein cholesterol.^{54,55}

Furthermore, a risk model incorporating polygenic risk with the PCE estimated risk is applied to individuals across different ancestries to demonstrate improved predictive performance.^{43,44} This improved performance illustrates the potential for an integrated absolute risk prediction model.^{43–45,75} For example, this model is particularly useful in differentiating risk in the high-risk South Asian ancestry population, where traditional clinical risk estimators often fail to capture the increased risk associated with this ancestry.⁴ The integration of the GPS_{Mult} with PCE builds on prior efforts which demonstrated improvement in model discrimination by now showing nearly identical improvement in C-statistic (0.03) in between models incorporating i) age and sex, ii) PCE alone, and iii) combined genetic and clinical risk across the population.^{7,76,77} However measures of C-statistic alone are not optimal or fully comprehensive in evaluating models that predict future risk.⁷⁸ GPS_{Mult} demonstrates nearly three-fold greater net reclassification of CAD cases/noncases when added to the PCE 10-year risk assessment to guide statin initiation as compared with established 'risk enhancing factors'.⁷⁹ Further work is needed to incorporate additional risk factors. To aid in future model calibration efforts, there is a need for population-level disease incidence and mortality data disaggregated by ancestral sub-groups.⁶⁷

These results should be interpreted within the context of limitations. Polygenic scores were developed and validated in individuals of European ancestry and then externally validated in non-European ancestry populations, and this may be partially contributing to poorer predictive performance in these groups. These results underscore the need for larger and more representative GWAS studies. UK Biobank participants were recruited at age 40–69 years, raising the possibility of survivorship or selection bias that limits generalizability to younger patients, however recent studies have demonstrated reliable performance of GPS in younger age groups.⁸ All UK Biobank disease endpoints were similarly ascertained through participant self-report, diagnosis codes from inpatient admissions, national procedure, and death registries. Participants in research studies tend to be healthier than the general population⁸⁰ — recalibration of disease risk models for a given target population may be needed prior to clinical deployment.

In conclusion, incorporating GWAS data for CAD and related traits from multiple ancestries on a large-scale leads to significantly improved performance of GPS_{Mult} in external validation among diverse ancestry populations when compared with previously published scores. This approach is generalizable to all traits and results in a polygenic score

that is able to better identify individuals at the highest and lowest ends of risk, significantly reclassifies risk beyond clinical risk estimators, and has the potential to advance clinical decision making.

METHODS:

Data availability:

All data are made available from the UK Biobank to researchers from universities and other institutions with genuine research inquiries following institutional review board and UK Biobank approval. This research was conducted using the UK Biobank resource under Application Number 7089 and secondary data use was approved by the Mass General Brigham institutional review board. Summary statistics from Biobank Japan are available at <http://jenger.riken.jp/en/result>. Summary statistics for the Coronary ARtery Disease Genomewide Replication and Meta-analysis plus The Coronary Artery Disease Genetics consortium (CARDIoGRAMplusC4D) study are available at <http://www.cardiogramplusc4d.org>. Summary statistics from FinnGen are available at https://www.finnngen.fi/en/access_results. Summary statistics from Genes & Health are available at <https://www.genesandhealth.org/research/scientific-data-downloads>. Summary statistics from the Million Veteran Program are available in dbGaP (accession number [phs001672](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15672)). The full GPS_{Mult} weights will be made available in the Polygenic Score Catalog.

Study populations:

The UK Biobank is a prospective cohort study that enrolled over 500,000 individuals between the ages of 40 and 69 years between 2006 and 2010.^{46,81} A detailed questionnaire completed by UK Biobank participants at enrolment assessed self-report of ancestry, lifestyle factors, including smoking. Anthropometric measurements including body-mass index were measured at the initial enrollment visit. Biomarkers including serum lipid concentrations and renal function markers were assessed at time of enrolment as part of the study protocol. Diagnoses of peripheral artery disease (PAD), diabetes, and hypertension were determined based on self-report or hospitalization records confirming a clinical diagnosis, as previously described.^{4,82}

Participants within the Million Veteran Program were recruited from more than 75 Veteran Affairs Medical Centers nationwide since 2011, with >885,000 individuals currently enrolled.⁵¹ Each participant has consented to linkage to their electronic medical record, wherein ICD9/10 diagnosis codes, Current Procedural Terminology (CPT) codes, clinical laboratory measurements, and reports of diagnostic imaging modalities are available. Participants were also asked to complete baseline and lifestyle questionnaires to further augment data contained in the electronic health record.

Genes & Health is a UK-based cohort of over 48,000 British Pakistani and Bangladeshi individuals recruited and consented for lifelong electronic health record access and genetic analysis.⁵² Medical records are linked to ICD10, OPCS, and SNOMED diagnosis and procedural codes across inpatient and hospital settings as well as clinical laboratory measurements, and a baseline questionnaire.

Clinical endpoints:

Ascertainment of CAD at enrollment in the UK Biobank was based on self-report, hospitalization records or death registry confirming diagnosis of myocardial infarction or its acute complications, or a coronary revascularization procedure (coronary artery bypass graft surgery or percutaneous angioplasty/stent placement), as previously described.^{82,83} The

earliest date at which the diagnosis was ascertained was considered as the diagnosis date. For individuals with CAD prior to enrollment, recurrence of CAD was determined based on diagnosis of a myocardial infarction or revascularization in the follow-up period, as previously described.⁸⁴

Within the Million Veteran Program, ICD9, ICD10, and CPT codes from both inpatient and outpatient encounters were used to curate and classify CAD cases based on having a myocardial infarction or undergoing revascularization, identified as subjects with at least two codes (of any category) that occurred on distinct dates within a 12 month window, as previously described.³⁹ Incident cases were identified as those with the first of the two qualifying codes occurring after enrollment. The remaining CAD cases, including through self-report, were considered prevalent.

In the Genes & Health study, ICD10 and SNOMED codes from the linked electronic health record were used to classify CAD cases defined as myocardial infarction or revascularization based on first diagnosis date, as described elsewhere.⁵³ Prevalent cases were defined as events prior to enrollment while events occurring after enrollment were designated as incident disease.

GPS construction:

Summary statistics from recent CAD GWAS studies (Genes & Health, FinnGen, Million Veterans Program, Biobank Japan, and CARDIOGRAMplusC4D excluding UK Biobank samples) conducted in individuals of diverse ancestries were used to determine primary CAD score weights (Supplementary Table 1).^{9,39,52,57,58} UK Biobank participants were not included among these discovery cohorts to preserve them as an independent hold-out dataset for training and validation of the GPS_{Mult} (Supplementary Table 2). Ancestry-specific linkage disequilibrium reference panels were extracted from the 1000 Genomes Project phase 3 data to match with the ancestry for the discovery GWAS, and only unrelated samples were used.⁸⁵ GPS_{Mult} construction was comprised in a two layer process, with layer 1 consisting of combining multiple polygenic scores derived from different ancestry-specific GWAS data for each trait, and layer 2 consisting of combining the multi-ancestry CAD polygenic score with similarly constructed multi-ancestry CAD-related trait scores predicting CAD (Figure 1) to generate GPS_{Mult} .

Separate GPS were constructed for each ancestry-stratified CAD GWAS using the LDpred2 method, which is a Bayesian approach to calculate a posterior mean effect for all variants based on an effect size in the prior GWAS and subsequent shrinkage based on linkage disequilibrium.⁸⁶ Only HapMap3 variants – a set of >1.4 million variants compiled by the International HapMap Project which capture common patterns of variation in a variety of human populations – were included for score calculation.⁸⁷ The default parameters used in the LDpred2 method included the proportion of variants to be causal (cut-offs of $p=1.0 \times 10^{-04}$, 1.8×10^{-04} , 3.2×10^{-04} , 5.6×10^{-04} , 1.0×10^{-03} , 1.8×10^{-03} , 3.2×10^{-03} , 5.6×10^{-03} , 1.0×10^{-02} , 1.8×10^{-02} , 3.2×10^{-02} , 5.6×10^{-02} , 1.0×10^{-01} , 1.8×10^{-01} , 3.2×10^{-01} , 5.6×10^{-01} and 1), the scale of heritability ($s=0.7, 1$ and 1.4), and whether or not a sparse LD matrix was applied.^{14,86,88} Combinations of these parameters resulted in 102 candidate GPSs for each set of ancestry-stratified GWAS summary statistics. The best GPS was selected among these candidates by assessing their performance in predicting prevalent CAD in an independent 116,649 individuals of White British ancestry from UK Biobank (this data set was used in all the score selection procedures thereafter and same group of individuals used to train previously published score GPS_{2018}).¹⁴ For selecting the best combination of CAD GPS scores from each ancestry-specific CAD GWAS for mixing, the discriminative capacities (Akaike information criterion,

AIC) of these GPS combinations for predicting CAD were assessed using the stepAIC function from R MASS package.⁸⁹ A logistic regression model was used to estimate the mixing weights for each individual ancestry-specific GPS. These GPSs were then linearly combined together into a single CAD_{GPS} score (layer 1, Figure 1). Similar procedures were followed for other atherosclerotic diseases (ischemic stroke, PAD)^{61,90} and risk factor traits – LDL cholesterol, HDL cholesterol, triglycerides^{62,79}, diabetes⁹¹, systolic blood pressure⁹², diastolic blood pressure, chronic kidney disease⁶³, body-mass index⁹³ (Supplementary Table 1, Figure 1).

Then, these multi-ancestry, trait-specific GPSs were linearly combined with the multi-ancestry CAD GPS (from layer 1) to generate the final GPS_{Mult} (layer 2). Just as for layer 1, the discriminative capacities (AIC) of these GPS combinations for predicting CAD were assessed to identify the best combination of trait-level scores for mixing.⁸⁹ A logistic regression model was used to estimate the mixing weights for each individual ancestry-specific GPS using the stepAIC function as described above. These GPSs were then linearly combined together into a single GPS_{Mult} score (layer 2, Figure 1). Of 58 GWAS- and ancestry-specific GPS that went through layers 1 and 2 of selection and mixing, 32 contributed to the final GPS_{Mult}, incorporating GWAS summary statistics from multiple ancestries and multiple CAD-related traits (Figure 2). LDPred2 parameters selected for each score, whether the score survived after feature selection, and mixing weights from layers 1 and 2 are listed in Supplementary Table 1.

GPS validation:

The GPS_{Mult} was compared to previously published polygenic scores for CAD. The variant effect sizes were downloaded from PGS Catalog and calculated in the same UK Biobank validation dataset of 308,264 European ancestry individuals for direct comparison.^{13–15,47,59,70,75,76,94–105} See Supplementary Table 3-5 for score accession numbers and performance metrics. The validation datasets were composed of UK Biobank participants separate from those used to train the GPS_{Mult}. These individuals underwent genotyping using the UK BiLEVE Axiom Array or UK Biobank Axiom Array, containing over 800,000 variants spanning the genome.⁴⁶ Imputation was performed using the Haplotype Reference Consortium resource, the UK10K panel, and the 1000 Genomes panel.^{85,106,107} We identified a subset of 488,243 participants with genotyping array data. After additional exclusion of 45,602 individuals for high heterozygosity or genotype missing rates, discordant reported versus genotypic sex, putative sex chromosome aneuploidy, excess relatedness (third-degree relative or closer), withdrawal of informed consent derived centrally, or unreported ancestry and 116,650 individuals used for score training, 325,991 individuals (54.3% female, 2.2% African, 0.4% East Asian, 92.0% European, and 2.7% South Asian) were included in the validation cohort for subsequent analyses.

Among Million Veteran Program participants, 157,563 individuals not included in the previously published CAD GWAS³⁹ were included and comprised of 33,096 (21%) individuals of African ancestry and 124,467 (79%) individuals of European ancestry (Supplementary Table 2). Individuals were genotyped using the Affymetrix Axiom array and imputed to the TOPMed reference panel. Variants and sample quality control were previously described.¹⁰⁸

Within the Genes & Health study, individuals not included in the previously published CAD GWAS⁵³ were included and comprised 16,874 participants of South Asian ancestry (Supplementary Table 2). These individuals underwent genotyping using the Illumina Infinium Global Screening Array v3 and imputed using the GenomeAsia pilot reference

panel. Variants with low call rate (<0.99), rare variants with minor allele frequency (MAF) $<1\%$, and variants that failed the Hardy–Weinberg test ($p < 1 \times 10^{-6}$) in a subset of samples with low level of autozygosity were removed.

Across all cohorts, individuals were analyzed in distinct self-identified groups of African, East Asian, European, and South Asian ancestries. The generated polygenic scores were residualized for the first ten principal components of genetic ancestry and then scaled to a mean of 0 and standard deviation of 1 for each ancestral group.

Statistical analysis:

Comparison of baseline characteristics between individuals with high or average genetic risk based on polygenic score was performed with the Chi-squared test for categorical variables, analysis of variance (ANOVA) for a subset of continuous variables with normal distributions, and Mann-Whitney U test for continuous variables with nonparametric distributions. Individuals with a given magnitude of increased risk were identified by comparing progressively higher percentile cut-offs to the middle quintile population in a logistic regression model predicting disease status and adjusted for baseline model covariates. Individuals were next binned into 100 groupings according to percentile of the GPS_{Mult} and the unadjusted prevalence of CAD within each bin was determined.

Risk for prevalent disease was calculated using logistic regression models, including baseline model covariates defined as enrollment age, sex, genotyping array, and the first 10 principal components of genetic ancestry. Risk for incident CAD was calculated using Cox proportional-hazards regression models, including baseline model covariates. The proportion of phenotypic variance explained by the polygenic score or risk factor of interest on the observed scale was calculated using the Nagelkerke's pseudo- R^2 metric— where R^2 was calculated for the full model inclusive of the variable of interest plus the baseline model covariates minus R^2 for the baseline model covariates alone. The proportion of phenotypic variance explained on the liability scale was similarly calculated using the logit liability R^2 metric, as described elsewhere.¹⁰⁹

To determine the polygenic risk equivalent of a CAD event comparable to risk experienced by those with prior CAD, a model was constructed comparing three groups and monitored for a CAD event in the follow-up period: individuals with prior CAD, individuals without prior CAD in different groupings of the top distribution of GPS_{Mult} (high GPS_{Mult}) and the remaining individuals without prior CAD. Sequentially lower percentile cut-offs for this high GPS_{Mult} group were tested to find the grouping with equivalent risk increase for CAD as those with prior CAD. This analysis was repeated for diabetes mellitus, PAD, and severe hypercholesterolemia (LDL cholesterol ≥ 190 mg/dL). In the lower tail of GPS_{Mult} , the risk for incident CAD was calculated in individuals in the bottom 5 percentiles or 5-9th percentiles of GPS_{Mult} relative to those in the middle quintile, using Cox proportional hazards regression models including baseline model covariates. The prevalence of CAD among individuals in the bottom 5 percentiles of GPS_{Mult} was calculated, stratified by 20 pack-years smoking increments and compared with the prevalence of CAD in non-smokers in the middle 40-59 percentiles to estimate equivalent offset risk.

Cox proportional hazards models were used to estimate hazard ratios for incident CAD in the UK Biobank, with covariates of the first 10 principal components. In model 1, only age and sex were modeled with the covariates. In model 2, only the clinical risk estimator – ACC/AHA Pooled Cohort Equations (PCE)⁵ or QRISK3⁶ – was modeled with the covariates. In model 3, GPS_{Mult} , clinical risk estimator, and the interaction term of GPS_{Mult} with the clinical risk estimator, and the first 10 principal components of genetic ancestry are modeled.

The 10-year incidence of CAD for individuals grouped by GPS_{Mult} percentile and stratified by ancestry group was quantified using model 3 standardized to four PCE risk levels (mean 10-year risk of atherosclerotic cardiovascular disease as low (<5%), borderline (5 to <7.5%), intermediate (≥ 7.5 to <20%), and high ($\geq 20\%$)) and the means of each of the covariates. The discrimination of each of these predictive models was assessed using Harrell's C-statistic. The improvement in predictive performance of the addition of the GPS_{Mult} to the PCE or QRISK3 was evaluated using continuous and categorized net reclassification improvement (NRI), with a risk probability threshold of 7.5% obtained with Kaplan-Meier estimates for a period of 10 years and 95% confidence intervals obtained from 100-fold bootstrapping. All statistical analyses were performed with the use of R software, versions 3.5 and 3.6 (R Project for Statistical Computing).

ACKNOWLEDGEMENTS:

This work was supported by the KL2/Catalyst Medical Research Investigator Training award from Harvard Catalyst (to A.P.P. and K.G.A.); the Sarnoff Cardiovascular Research Foundation Fellowship (to S.A.); grants 1K08HL153937 (to K.G.A.), 1K08HL161448 (to A.C.F.), R01HL1427 (to P.N.), R01HL148565 (to P.N.), R01HL148050 (to P.N.), 1R01HL092577 (to P.T.E.), 1R01HL157635 (to P.T.E.), and 1R01HL157635 (to P.T.E.) from the National Heart, Lung, and Blood Institute; grant 862032 (to K.A.) and grants 18SFRN34110082 (to P.T.E.), 17IFUNP3384001 (to .G.A.) from the American Heart Association; grant MAESTRIA 965286 from the European Union (to P.T.E.); grants 1K08HG010155 and 1U01HG011719 from the National Human Genome Research Institute (to A.P.P., P.N., and A.V.K.); a Hassenfeld Scholar Award from Massachusetts General Hospital (to P.N. and A.V.K.); a Merkin Institute Fellowship from the Broad Institute of MIT and Harvard (to A.V.K.). This research has been conducted using the UK Biobank Resource, and we thank the volunteers participating. This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by Veterans Administration awards I01-01BX003362, I01-BX004821 (P.S.T), I01-BX003340 (P.W.F.W) and VA HSR RES 13-457 (VA Informatics and Computing Infrastructure). The content of this manuscript does not represent the views of the Department of Veterans Affairs or the United States Government. 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 Astra Zeneca 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. 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), 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.

DISCLOSURES:

S.A. has served as a scientific advisor to Third Rock Ventures. A.C.F. is a co-founder of Goodpath and reports a grant from Abbott Vascular. P.T.E. receives sponsored research support from Bayer AG and IBM Research; he has also served on advisory boards or consulted for Bayer AG, MyoKardia and Novartis. P.N. reports investigator-initiated grants from Amgen, Apple, AstraZeneca, Boston Scientific, and Novartis, personal fees from Apple, AstraZeneca, Blackstone Life Sciences, Foresite Labs, Novartis, Roche / Genentech, is a co-founder of TenSixteen Bio, is a scientific advisory board member of Esperion Therapeutics, geneXwell, and TenSixteen Bio, and spousal employment at Vertex Pharmaceuticals, all unrelated to the present work. A.V.K. is an employee of Verve Therapeutics; has served as a scientific advisor to Amgen, Novartis, Silence Therapeutics, Korro Bio, Veritas International, Color Health, Third Rock Ventures, Illumina, Ambry, and Foresite Labs; holds equity in Verve Therapeutics, Color Health, and Foresite Labs; and is listed as a co-inventor on patent applications related to assessment and mitigation of risk associated with perturbations in body fat distribution.

REFERENCES:

1. Roth, G. A. *et al.* Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**, 1736–1788 (2018).
2. Arnett, D. K. *et al.* 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* **140**, (2019).
3. DeFilippis, A. P. *et al.* An Analysis of Calibration and Discrimination Among Multiple Cardiovascular Risk Scores in a Modern Multiethnic Cohort. *Ann Intern Med* **162**, 266–275 (2015).
4. Patel, A. P., Wang, M., Kartoun, U., Ng, K. & Khera, A. V. Quantifying and Understanding the Higher Risk of Atherosclerotic Cardiovascular Disease Among South Asian Individuals. *Circulation* **144**, 410–422 (2021).
5. Goff David C. *et al.* 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk. *Circulation* **129**, S49–S73 (2014).
6. Hippisley-Cox, J., Coupland, C. & Brindle, P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* **357**, j2099 (2017).
7. Khan, S. S. *et al.* Predictive Utility of a Validated Polygenic Risk Score for Long-Term Risk of Coronary Heart Disease in Young and Middle-Aged Adults. *Circulation* **146**, 587–596 (2022).
8. Emdin, C. A. *et al.* Polygenic Score Assessed in Young Adulthood and Onset of Subclinical Atherosclerosis and Coronary Heart Disease. *J Am Coll Cardiol* **80**, 280–282 (2022).
9. Aragam, K. G. *et al.* Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet* **54**, 1803–1815 (2022).
10. Samani, N. J. *et al.* Genomewide Association Analysis of Coronary Artery Disease. *New England Journal of Medicine* **357**, 443–453 (2007).
11. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* **43**, 333–338 (2011).
12. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* **45**, 25–33 (2013).
13. Inouye, M. *et al.* Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults. *Journal of the American College of Cardiology* **72**, 1883–1893 (2018).
14. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
15. Wang, M. *et al.* Validation of a Genome-Wide Polygenic Score for Coronary Artery Disease in South Asians. *J Am Coll Cardiol* **76**, 703–714 (2020).
16. Hindy, G. *et al.* Genome-Wide Polygenic Score, Clinical Risk Factors, and Long-Term Trajectories of Coronary Artery Disease. *Arterioscler Thromb Vasc Biol* **40**, 2738–2746 (2020).
17. Sun, L. *et al.* Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses. *PLoS Med* **18**, e1003498 (2021).
18. Blout Zawatsky, C. L. *et al.* Returning actionable genomic results in a research biobank: Analytic validity, clinical implementation, and resource utilization. *Am J Hum Genet* **108**, 2224–2237 (2021).

19. Karlson, E. W., Boutin, N. T., Hoffnagle, A. G. & Allen, N. L. Building the Partners HealthCare Biobank at Partners Personalized Medicine: Informed Consent, Return of Research Results, Recruitment Lessons and Operational Considerations. *J Pers Med* **6**, (2016).
20. Patel, A. P. & Khera, A. V. Advances and Applications of Polygenic Scores for Coronary Artery Disease. *Annual Review of Medicine* **74**, 141–154 (2023).
21. Zhang, Y., Qi, G., Park, J.-H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat Genet* **50**, 1318–1326 (2018).
22. Márquez-Luna, C., Loh, P.-R., South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium & Price, A. L. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol* **41**, 811–823 (2017).
23. Koyama, S. *et al.* Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat Genet* **52**, 1169–1177 (2020).
24. Kurniansyah, N. *et al.* A multi-ethnic polygenic risk score is associated with hypertension prevalence and progression throughout adulthood. *Nat Commun* **13**, 3549 (2022).
25. Weissbrod, O. *et al.* Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nat Genet* 1–9 (2022) doi:10.1038/s41588-022-01036-9.
26. Ge, T. *et al.* Development and validation of a trans-ancestry polygenic risk score for type 2 diabetes in diverse populations. *Genome Medicine* **14**, 70 (2022).
27. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236–1241 (2015).
28. Krapohl, E. *et al.* Multi-polygenic score approach to trait prediction. *Mol Psychiatry* **23**, 1368–1374 (2018).
29. Ramirez, J. *et al.* Prediction of Coronary Artery Disease and Major Adverse Cardiovascular Events Using Clinical and Genetic Risk Scores for Cardiovascular Risk Factors. *Circ Genom Precis Med* **15**, e003441 (2022).
30. Sinnott-Armstrong, N. *et al.* Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat Genet* **53**, 185–194 (2021).
31. Neumann, A. *et al.* Combined polygenic risk scores of different psychiatric traits predict general and specific psychopathology in childhood. *J Child Psychol Psychiatry* **63**, 636–645 (2022).
32. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* **50**, 229–237 (2018).
33. Ning, Z., Pawitan, Y. & Shen, X. High-definition likelihood inference of genetic correlations across human complex traits. *Nat Genet* **52**, 859–864 (2020).
34. Maier, R. M. *et al.* Improving genetic prediction by leveraging genetic correlations among human diseases and traits. *Nat Commun* **9**, 989 (2018).
35. Martin, A. R. *et al.* Current clinical use of polygenic scores will risk exacerbating health disparities. *Nat Genet* **51**, 584–591 (2019).
36. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat Genet* **54**, 573–580 (2022).
37. Cai, M. *et al.* A unified framework for cross-population trait prediction by leveraging the genetic correlation of polygenic traits. *Am J Hum Genet* **108**, 632–655 (2021).
38. Gurdasani, D. *et al.* Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. *Cell* **179**, 984–1002.e36 (2019).

39. Tcheandjieu, C. *et al.* Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat Med* **1–13** (2022) doi:10.1038/s41591-022-01891-3.
40. All of Us Research Program Investigators *et al.* The ‘All of Us’ Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019).
41. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program | Nature. <https://www.nature.com/articles/s41586-021-03205-y>.
42. Wall, J. D. *et al.* The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* **576**, 106–111 (2019).
43. Riveros-Mckay, F. *et al.* Integrated Polygenic Tool Substantially Enhances Coronary Artery Disease Prediction. *Circulation: Genomic and Precision Medicine* **14**, e003304 (2021).
44. Weale, M. E. *et al.* Validation of an Integrated Risk Tool, Including Polygenic Risk Score, for Atherosclerotic Cardiovascular Disease in Multiple Ethnicities and Ancestries. *Am J Cardiol* **148**, 157–164 (2021).
45. Mars, N. *et al.* Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med* **26**, 549–557 (2020).
46. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
47. Lambert, S. A. *et al.* The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat Genet* **53**, 420–425 (2021).
48. Manikpurage, H. D. *et al.* Polygenic Risk Score for Coronary Artery Disease Improves the Prediction of Early-Onset Myocardial Infarction and Mortality in Men. *Circ Genom Precis Med* **14**, e003452 (2021).
49. Wells, Q. S. *et al.* Polygenic Risk Score to Identify Subclinical Coronary Heart Disease Risk in Young Adults. *Circ Genom Precis Med* **14**, e003341 (2021).
50. Neumann, J. T. *et al.* Prognostic Value of a Polygenic Risk Score for Coronary Heart Disease in Individuals Aged 70 Years and Older. *Circ Genom Precis Med* **15**, e003429 (2022).
51. Gaziano, J. M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *Journal of Clinical Epidemiology* **70**, 214–223 (2016).
52. Finer, S. *et al.* Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *Int J Epidemiol* **49**, 20–21i (2020).
53. Huang, Q. Q. *et al.* Transferability of genetic loci and polygenic scores for cardiometabolic traits in British Pakistani and Bangladeshi individuals. *Nat Commun* **13**, 4664 (2022).
54. Dron, J. S. *et al.* Association of Rare Protein-Truncating DNA Variants in APOB or PCSK9 With Low-density Lipoprotein Cholesterol Level and Risk of Coronary Heart Disease. *JAMA Cardiology* (2023) doi:10.1001/jamacardio.2022.5271.
55. Cohen, J. C., Boerwinkle, E., Mosley, T. H. & Hobbs, H. H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
56. Howe, L. J. *et al.* Polygenic risk scores for coronary artery disease and subsequent event risk amongst established cases. *Hum Mol Genet* **29**, 1388–1395 (2020).
57. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415–1424 (2021).

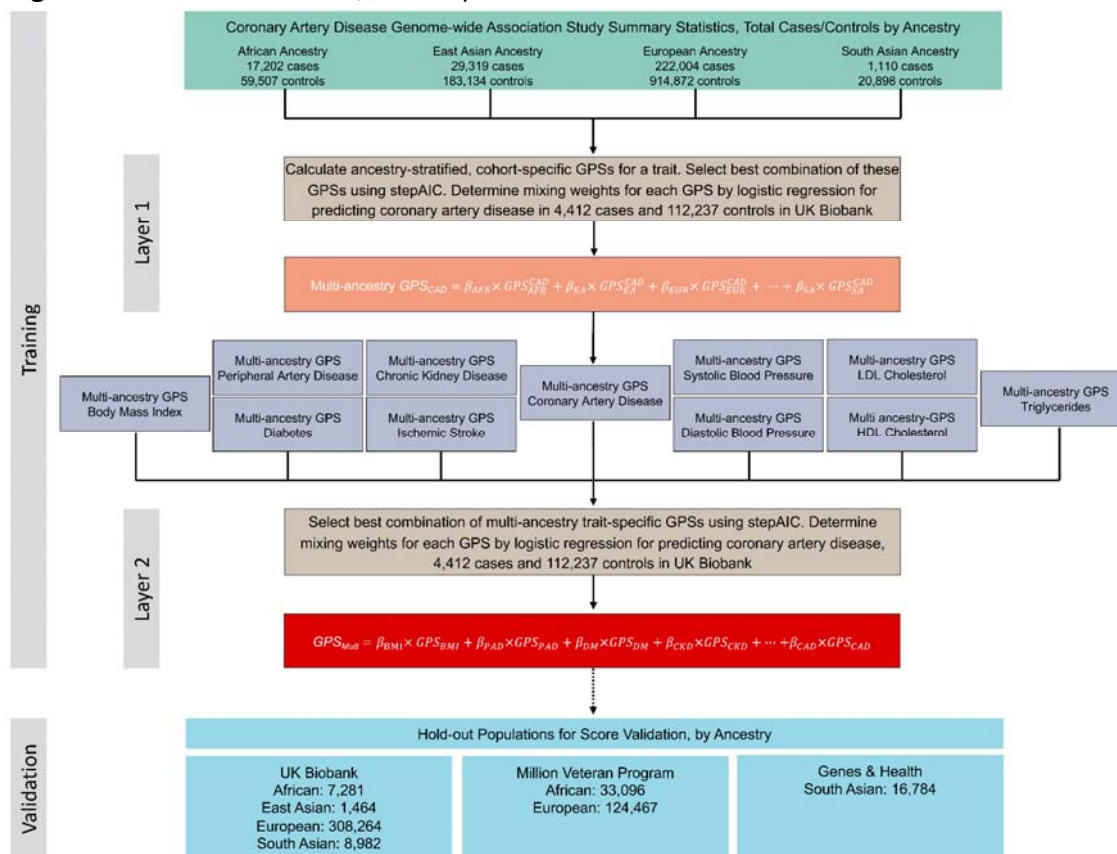
58. Locke, A. E. *et al.* Exome sequencing of Finnish isolates enhances rare-variant association power. *Nature* **572**, 323–328 (2019).
59. Lu, X. *et al.* A polygenic risk score improves risk stratification of coronary artery disease: a large-scale prospective Chinese cohort study. *Eur Heart J* **43**, 1702–1711 (2022).
60. Zhang, H. *et al.* Novel Methods for Multi-ancestry Polygenic Prediction and their Evaluations in 3.7 Million Individuals of Diverse Ancestry. 2022.03.24.485519 Preprint at <https://doi.org/10.1101/2022.03.24.485519> (2022).
61. Klarin, D. *et al.* Genome-wide Association Study of Peripheral Artery Disease in the Million Veteran Program. *Nat Med* **25**, 1274–1279 (2019).
62. Klarin, D. *et al.* Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet* **50**, 1514–1523 (2018).
63. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet* **51**, 957–972 (2019).
64. Wienke, A., Holm, N. V., Skytthe, A. & Yashin, A. I. The heritability of mortality due to heart diseases: a correlated frailty model applied to Danish twins. *Twin Res* **4**, 266–274 (2001).
65. Zdravkovic, S. *et al.* Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med* **252**, 247–254 (2002).
66. McCarty, C. A. *et al.* The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics* **4**, 13 (2011).
67. Hao, L. *et al.* Development of a clinical polygenic risk score assay and reporting workflow. *Nat Med* **28**, 1006–1013 (2022).
68. O'Sullivan, J. W. *et al.* Polygenic Risk Scores for Cardiovascular Disease: A Scientific Statement From the American Heart Association. *Circulation* **146**, e93–e118 (2022).
69. Khera, A. V. *et al.* Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *New England Journal of Medicine* **375**, 2349–2358 (2016).
70. Natarajan, P. *et al.* Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. *Circulation* **135**, 2091–2101 (2017).
71. Aragam, K. G. *et al.* Limitations of Contemporary Guidelines for Managing Patients at High Genetic Risk of Coronary Artery Disease. *J Am Coll Cardiol* **75**, 2769–2780 (2020).
72. Fahed, A. C., Philippakis, A. A. & Khera, A. V. The potential of polygenic scores to improve cost and efficiency of clinical trials. *Nat Commun* **13**, 2922 (2022).
73. Marston, N. A. *et al.* Predicting Benefit From Evolocumab Therapy in Patients With Atherosclerotic Disease Using a Genetic Risk Score: Results From the FOURIER Trial. *Circulation* **141**, 616–623 (2020).
74. Damask, A. *et al.* Patients With High Genome-Wide Polygenic Risk Scores for Coronary Artery Disease May Receive Greater Clinical Benefit From Alirocumab Treatment in the ODYSSEY OUTCOMES Trial. *Circulation* **141**, 624–636 (2020).
75. Tamlander, M. *et al.* Integration of questionnaire-based risk factors improves polygenic risk scores for human coronary heart disease and type 2 diabetes. *Commun Biol* **5**, 158 (2022).
76. Elliott, J. *et al.* Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. *JAMA* **323**, 636–645 (2020).
77. Mosley, J. D. *et al.* Predictive Accuracy of a Polygenic Risk Score Compared With a Clinical Risk Score for Incident Coronary Heart Disease. *JAMA* **323**, 627–635 (2020).
78. Cook, N. R. Use and misuse of the receiver operating characteristic curve in risk

- prediction. *Circulation* **115**, 928–935 (2007).
79. Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).
 80. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* **186**, 1026–1034 (2017).
 81. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
 82. Patel, A. P. *et al.* Association of Rare Pathogenic DNA Variants for Familial Hypercholesterolemia, Hereditary Breast and Ovarian Cancer Syndrome, and Lynch Syndrome With Disease Risk in Adults According to Family History. *JAMA Netw Open* **3**, e203959 (2020).
 83. Fahed, A. C. *et al.* Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun* **11**, 3635 (2020).
 84. Patel, A. P. *et al.* Lp(a) (Lipoprotein[a]) Concentrations and Incident Atherosclerotic Cardiovascular Disease: New Insights From a Large National Biobank. *Arterioscler Thromb Vasc Biol* **41**, 465–474 (2021).
 85. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
 86. Privé, F., Arbel, J. & Vilhjálmsón, B. J. LDpred2: better, faster, stronger. *Bioinformatics* btaa1029 (2020) doi:10.1093/bioinformatics/btaa1029.
 87. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
 88. Vilhjálmsón, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet* **97**, 576–592 (2015).
 89. Zhang, Z. Variable selection with stepwise and best subset approaches. *Ann Transl Med* **4**, 136 (2016).
 90. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* **50**, 524–537 (2018).
 91. Vujkovic, M. *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* **52**, 680–691 (2020).
 92. Giri, A. *et al.* Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat Genet* **51**, 51–62 (2019).
 93. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
 94. Mega, J. *et al.* Genetic Risk, Coronary Heart Disease Events, and the Clinical Benefit of Statin Therapy. *Lancet* **385**, 2264–2271 (2015).
 95. Abraham, G. *et al.* Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nat Commun* **10**, 5819 (2019).
 96. Ripatti, S. *et al.* A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* **376**, 1393–1400 (2010).
 97. Tikkanen, E., Havulinna, A. S., Palotie, A., Salomaa, V. & Ripatti, S. Genetic risk prediction and a 2-stage risk screening strategy for coronary heart disease. *Arterioscler Thromb Vasc Biol* **33**, 2261–2266 (2013).
 98. Tada, H. *et al.* Risk prediction by genetic risk scores for coronary heart disease is

- independent of self-reported family history. *Eur Heart J* **37**, 561–567 (2016).
99. Paquette, M. *et al.* Polygenic risk score predicts prevalence of cardiovascular disease in patients with familial hypercholesterolemia. *J Clin Lipidol* **11**, 725-732.e5 (2017).
 100. Hajek, C. *et al.* Coronary Heart Disease Genetic Risk Score Predicts Cardiovascular Disease Risk in Men, Not Women. *Circ Genom Precis Med* **11**, e002324 (2018).
 101. Pechlivanis, S. *et al.* Risk prediction for coronary heart disease by a genetic risk score - results from the Heinz Nixdorf Recall study. *BMC Med Genet* **21**, 178 (2020).
 102. Gola, D. *et al.* Population Bias in Polygenic Risk Prediction Models for Coronary Artery Disease. *Circ Genom Precis Med* **13**, e002932 (2020).
 103. Bauer, A. *et al.* Comparison of genetic risk prediction models to improve prediction of coronary heart disease in two large cohorts of the MONICA/KORA study. *Genet Epidemiol* **45**, 633–650 (2021).
 104. Ye, Y. *et al.* Interactions Between Enhanced Polygenic Risk Scores and Lifestyle for Cardiovascular Disease, Diabetes, and Lipid Levels. *Circ Genom Precis Med* **14**, e003128 (2021).
 105. Mars, N. *et al.* Genome-wide risk prediction of common diseases across ancestries in one million people. *Cell Genom* **2**, None (2022).
 106. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279–1283 (2016).
 107. Walter, K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
 108. Hunter-Zinck, H. *et al.* Genotyping Array Design and Data Quality Control in the Million Veteran Program. *The American Journal of Human Genetics* **106**, 535–548 (2020).
 109. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol* **36**, 214–224 (2012).
 110. Kurki, M. I. *et al.* FinnGen: Unique genetic insights from combining isolated population and national health register data. 2022.03.03.22271360 Preprint at <https://doi.org/10.1101/2022.03.03.22271360> (2022).
 111. Zhou, W. *et al.* Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases. 2021.11.19.21266436 Preprint at <https://doi.org/10.1101/2021.11.19.21266436> (2021).
 112. Spracklen, C. N. *et al.* Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* **582**, 240–245 (2020).
 113. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* **50**, 1412–1425 (2018).

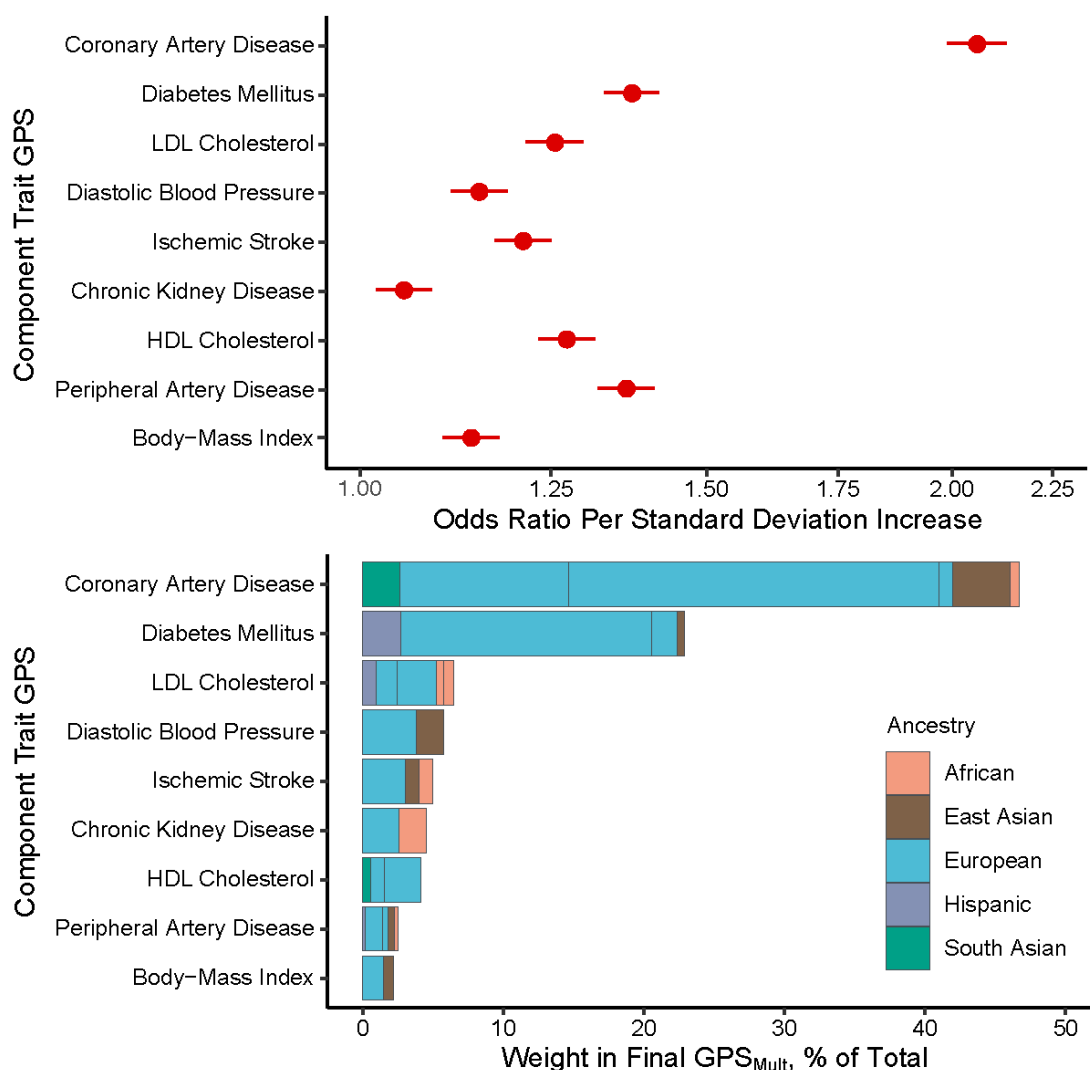
FIGURES AND TABLES:

Figure 1: Overview of GPS_{Mult} development



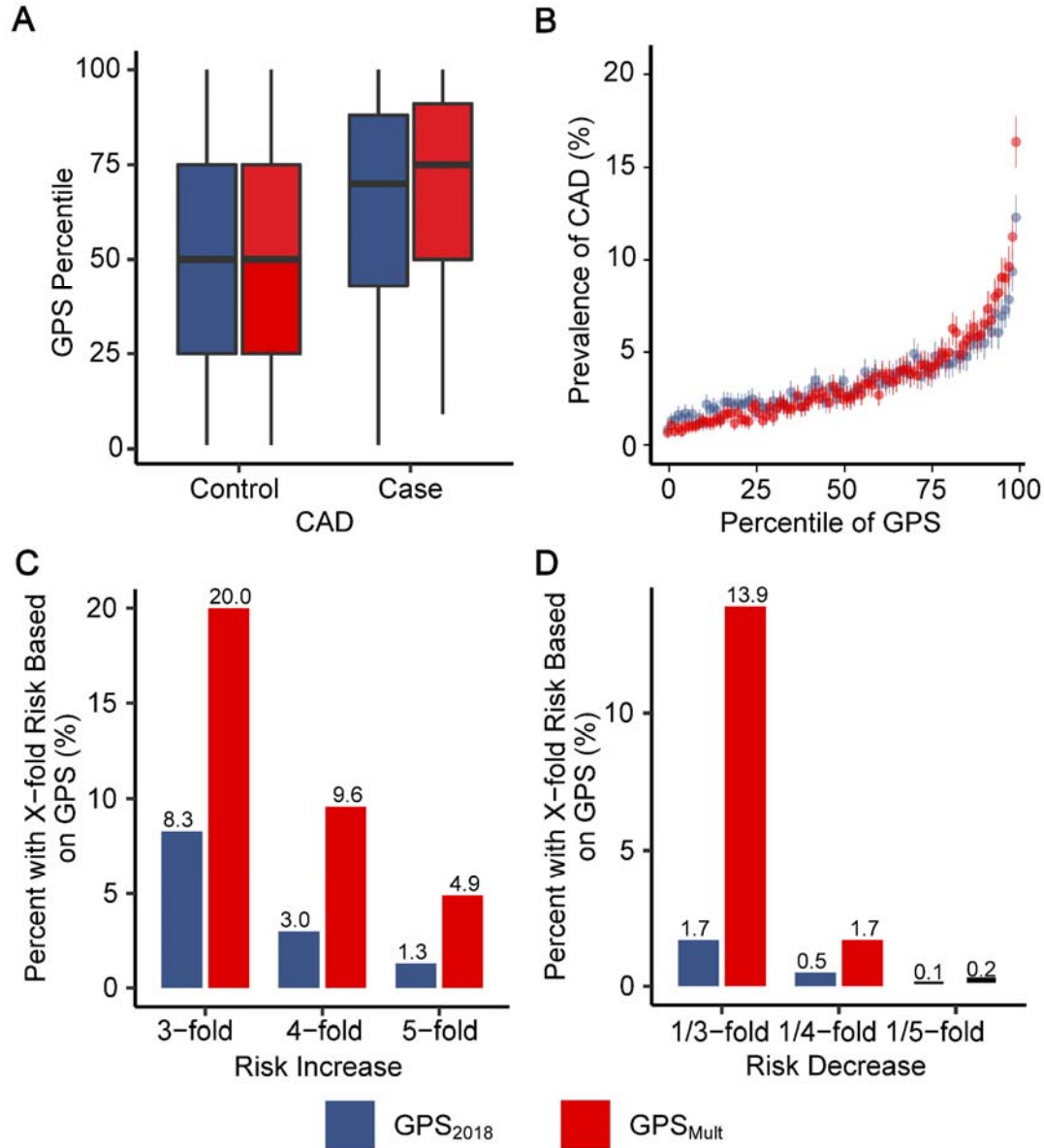
Polygenic scores for coronary artery disease (CAD) were constructed using ancestry-stratified, cohort-specific summary statistics from CAD and CAD-related traits, resulting in 58 GPS across all traits and ancestries. For each source trait (e.g., CAD) the best performing combination of ancestry-stratified, cohort-specific GPS was determined based on ability to predict CAD, selected using stepAIC, and their optimal mixing weights () determined using logistic regression in 116,649 individuals of European ancestry in the UK Biobank training dataset. The selected GPSs were linearly combined using these mixing weights to yield multi-ancestry scores predicting CAD from each source trait (layer 1). The best performing combination of multi-ancestry, trait-specific GPSs in predicting CAD was determined using stepAIC, and their optimal mixing weights () were determined using logistic regression in 116,649 individuals of European ancestry in the UK Biobank training dataset. The selected GPSs were linearly combined using these mixing weights to yield GPS_{Mult} (layer 2). GPS_{Mult} was validated with prediction of CAD in the UK Biobank and externally validated in Million Veteran Program and Genes & Health Studies in hold-out populations not included in score training. Ancestries: AFR – African; EA – East Asian; EUR – European; SA – South Asian. Source GWAS traits: CAD^{9,39,53,57,110}, body mass index (BMI)^{57,93}, ischemic stroke^{57,90,111}, diabetes mellitus (DM)^{91,111,112}, peripheral artery disease (PAD)^{57,61,110}, chronic kidney disease (CKD)^{57,63}, systolic blood pressure (SBP)^{57,113}, diastolic blood pressure (DBP)^{57,113}, low-density lipoprotein cholesterol (LDL)^{57,62,79}, high-density lipoprotein cholesterol (HDL)^{57,62,79}, triglycerides (TG)^{57,62,79}.

Figure 2: Trait-specific component polygenic score performance and ancestry-specific polygenic score composition of GPS_{Mult}



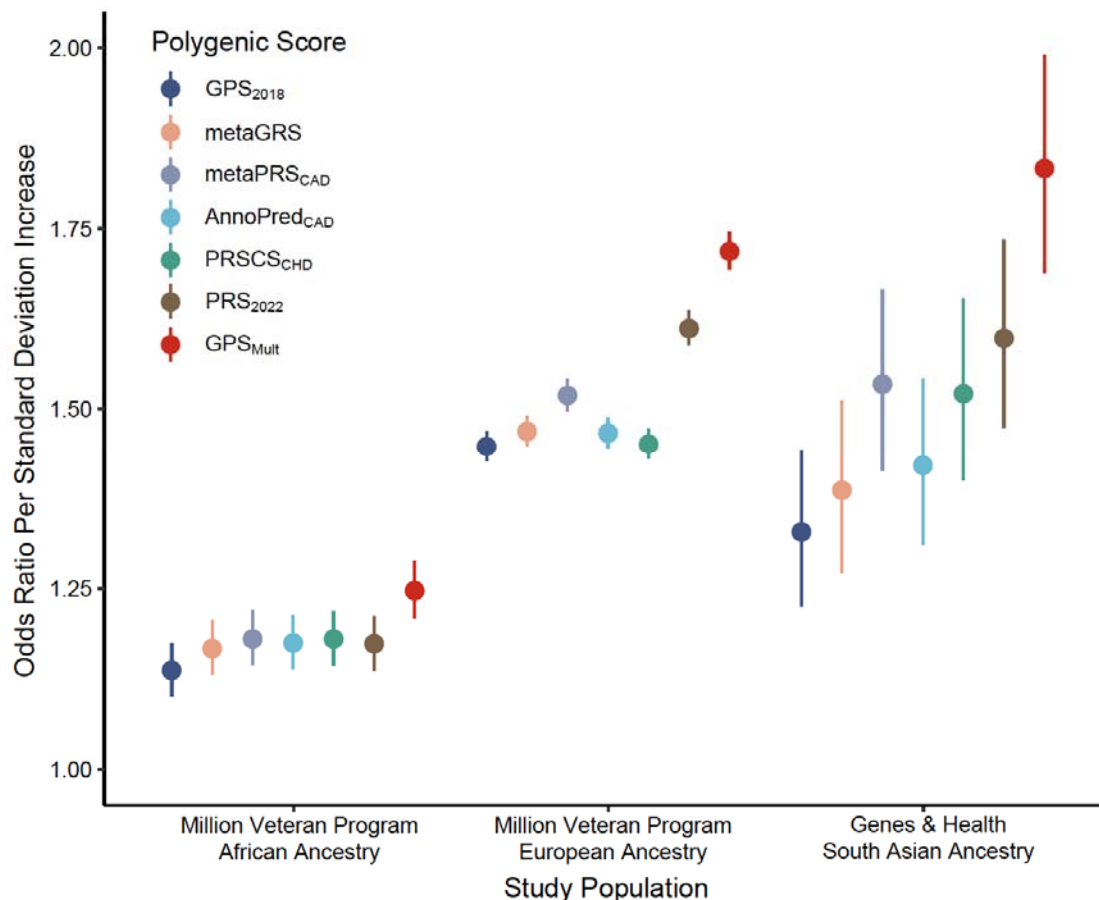
A: The odds ratios for prevalent coronary artery disease (CAD) risk per standard deviation increase of the multi-ancestry, trait-specific layer 1 GPSs were assessed in logistic regression models adjusted for age, sex, genotyping array, and the first ten principal components of ancestry in the same training group of 116,649 UK Biobank European ancestry individuals. B: The contributing weights of each of the ancestry-specific GWAS-based GPS to each of the trait-based layer 1 polygenic scores, color groupings by ancestry of source GWAS, and normalized to 100% to reflect composition in overall GPS_{Mult} . Of 58 ancestry- and trait-specific scores that were included in the GPS training analysis, 32 scores significantly contributed to overall prediction in GPS_{Mult} after optimization of score selection with stepAIC and weighting through logistic regression. GPS: genomewide polygenic score; LDL: low-density lipoprotein; HDL: high-density lipoprotein

Figure 3: Improvements in polygenic prediction of prevalent coronary artery disease prediction



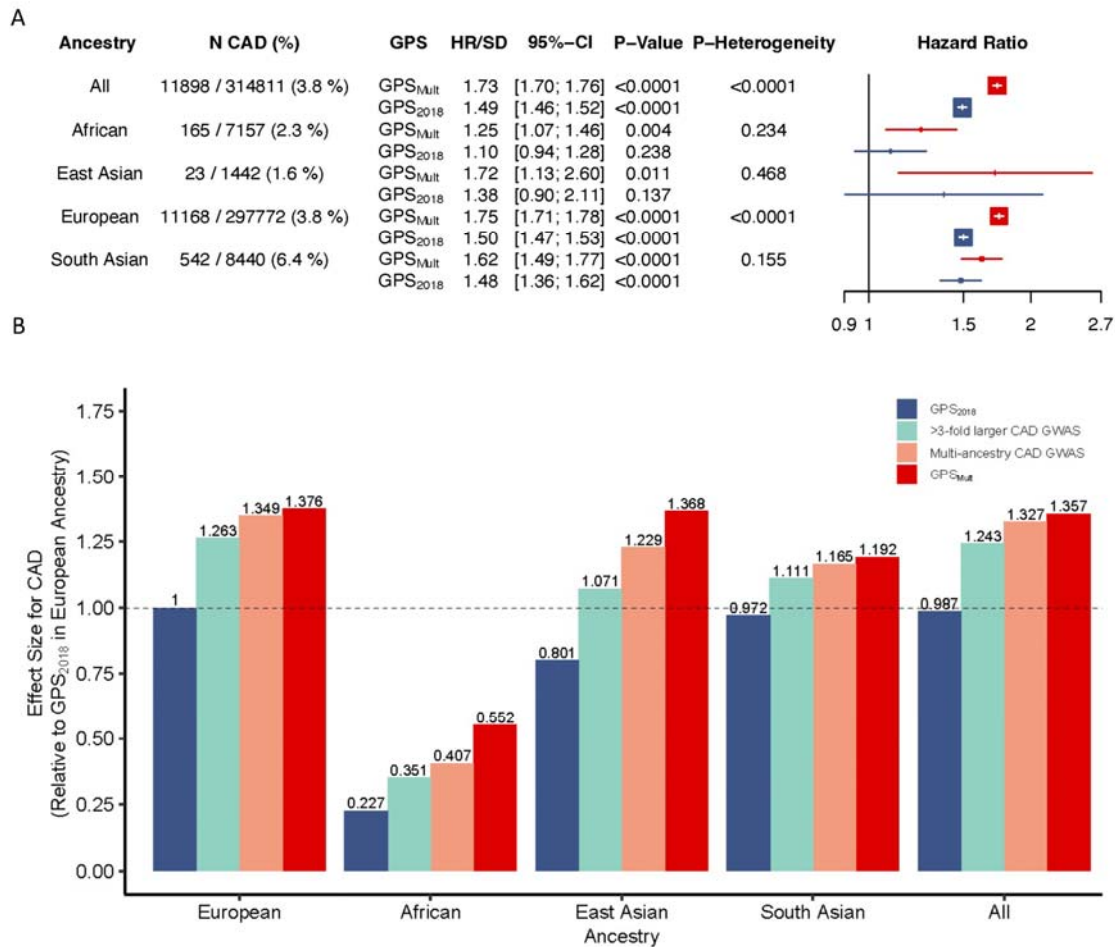
A: Distributions of GPS₂₀₁₈ and GPS_{Mult} percentiles across the UK Biobank validation dataset. B: Prevalence of CAD with 95% CI according to 100 groups of the UK Biobank validation dataset binned according to the percentile of the GPS₂₀₁₈ and GPS_{Mult}. C: Proportion of UK Biobank validation population with 3, 4, and 5-fold increased risk for CAD versus the middle quintile of the population, stratified by GPS. Odds ratio assessed in a logistic regression model adjusted for age, sex, genotyping array, and the first ten principal components of ancestry. D: Proportion of UK Biobank testing population with 1/3, 1/4, and 1/5 risk for CAD versus the middle quintile of the population, stratified by GPS. Odds ratio assessed in a logistic regression model adjusted for age, sex, genotyping array, and the first ten principal components of ancestry. GPS: Genome-wide polygenic score; CAD: coronary artery disease.

Figure 4: External validation of GPS_{Mult} and benchmarking against published polygenic scores for coronary artery disease across multiple ancestries in Million Veteran Program and Genes & Health studies.



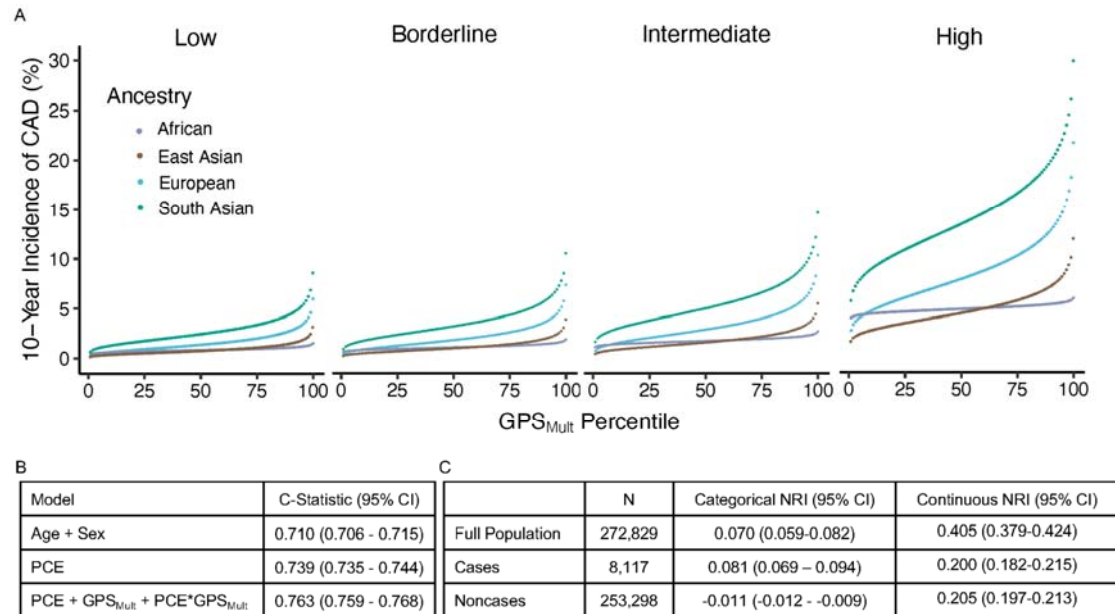
The odds ratio for prevalent coronary artery disease (CAD) risk per standard deviation increase of the polygenic score was assessed in a logistic regression model adjusted for age, sex, genotyping array, and the first ten principal components of ancestry in the same group of individuals per cohort: African ancestry individuals in Million Veteran Program; European ancestry individuals in Million Veteran Program; South Asian ancestry individuals in Genes & Health, using high-performing published scores from the Polygenic Score Catalog (GPS₂₀₁₈¹⁴, metaGRS¹³, metaPRS_{CAD}²³, AnnoPred_{CAD}¹⁰⁴, PRSCS_{CHD}⁷⁵, and PRS₂₀₂₂⁹) and current GPS_{Mult}⁴⁷. Results for these and remaining CAD polygenic scores published in the Polygenic Score Catalog are available in Supplementary Tables 6-7. GPS: Genome-wide polygenic score

Figure 5: Incident CAD prediction by GPS_{Mult} stratified by ancestry



A: Adjusted hazards ratio per standard deviation of the polygenic score with corresponding 95% CIs and P values for coronary artery disease (CAD) by ancestry, stratified by iteration of the version of the polygenic score, calculated from Cox proportional-hazards regression models adjusted for age, sex, genotyping array, and the first ten principal components of ancestry in the UK Biobank validation dataset. GPS₂₀₁₈ corresponds to previously published polygenic score for CAD.¹⁴ B: The score effect sizes relative to the effect size of GPS₂₀₁₈ in European ancestry individuals. >3-fold larger CAD GWAS designates metrics for polygenic score generated using summary statistics from the most recent Coronary ARtery Disease Genomewide Replication and Meta-analysis plus The Coronary Artery Disease Genetics consortium (CARDIOGRAMplusC4D) excluding the UK Biobank, of largely European ancestry. Multi-ancestry CAD GWAS refers to the polygenic score generated by combining ancestry-specific polygenic scores generated using GWAS summary statistics from CARDIOGRAMplusC4D, Genes & Health, Biobank Japan, Million Veteran Program, and FinnGEN biobanks in layer 1. GPS_{Mult} designates polygenic score for CAD designed with summary statistics from multiple ancestries and multiple CAD-related traits in layer 2. *Designates the reference group for calculating relative gain. GPS: Genome-wide polygenic score; CAD: coronary artery disease; GWAS: genome-wide association study.

Figure 6: Discrimination and reclassification by a model integrating polygenic and clinical risk for incident CAD



A: Cumulative incidence of coronary artery disease (CAD) over 10 years predicted by modeling GPS_{Mult}, AHA/ACC Pooled Cohort Equations (PCE) 10-year risk estimate, and their interaction in the UK Biobank validation dataset binned according to the percentile of the GPS_{Mult}, grouped by risk categories of the PCE (mean 10-year risk of atherosclerotic cardiovascular disease as low (<5%), borderline (5 to <7.5%), intermediate (≥7.5 to <20%), and high (≥20%)), and stratified by ancestry. B: C-statistics are based on 10-year follow-up events from Cox regression models of listed variables. PCE includes age and sex variables in its risk estimation. C: The improvement in the predictive performance of the addition of the GPS_{Mult} to the PCE was evaluated using continuous and categorised net reclassification improvement (NRI), with a risk probabilities threshold of 7.5% obtained with Kaplan-Meier estimates for a period of 10 years and confidence intervals (95%) obtained from 100-fold bootstrapping. GPS: genome-wide polygenic score.