

1 **A Multi-Ancestry Polygenic Risk Score for Coronary Heart Disease Based on an Ancestrally Diverse**
2 **Genome-Wide Association Study and Population-Specific Optimization**

3

4 *Johanna L. Smith, Ph.D. ^{*1}, Catherine Tcheandjieu, DVM, Ph.D. ^{*2,3,4,5}, Ozan Dikilitas, M.D. ¹, Kruthika Iyer,*
5 *Ph.D. ⁶, Kazuo Miyazawa, M.D., Ph.D. ⁷, Austin Hilliard, Ph.D. ^{4,6}, Julie Lynch, Ph.D. ⁸, Jerome I. Rotter, M.D.,*
6 *Ph.D. ⁹, Yii-Der Ida Chen, Ph.D. ⁹, Wayne Huey-Herng Sheu ^{10,11,12}, Kyong-Mi Chang, M.D. ¹³, Stavroula*
7 *Kanoni, Ph.D. ¹⁴, Phil Tsao, Ph.D. ^{4,15}, Kaoru Ito, M.D., Ph.D. ⁷, Matthew Kosel, B.S. ¹⁶, Shoa L. Clarke,*
8 *Ph.D. ^{4,15}, Daniel J. Schaid, Ph.D. ¹⁶, Themistocles L. Assimes, Ph.D. ^{#15}, Iftikhar J. Kullo, M.D. ^{#1}*

9

10 ¹ Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, USA.

11 ² Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA,
12 USA.

13 ³ Gladstone Institute of Data Science and Biotechnology, Gladstone Institutes, San Francisco, CA, USA.

14 ⁴ VA Palo Alto Health Care System, Palo Alto, CA, USA.

15 ⁵ Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine,
16 Stanford, CA, USA.

17 ⁶ Stanford University School of Medicine, Palo Alto, CA, USA.

18 ⁷ Riken Ctr. for Integrative Medical Sciences, Yokohama City, Japan.

19 ⁸ Salt Lake City VA Met CTR., Salt Lake City, UT, USA.

20 ⁹ Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Lundquist
21 Institute at Harbor-UCLA Medical Center, Torrance, CA, USA.

22 ¹⁰ Institute of Molecular and Genomic Medicine, National Health Research Institutes, Taiwan.

23 ¹¹ Division of Endocrinology and Metabolism, Department of Internal Medicine, Taipei Veterans General
24 Hospital, Taipei, Taiwan.

25 ¹² Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans
26 General Hospital, Taichung, Taiwan.

27 ¹³ Corporal Michael J Crescenz VA Medical Ctr. Philadelphia, PA, USA.

28 ¹⁴ Queen Mary University of London, Cambridge, UK.

29 ¹⁵ Stanford University, Stanford, CA, USA.

30 ¹⁶ Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA.

31

32 * Co-first Authors

33 # Corresponding Authors

34

35 Iftikhar J. Kullo, 200 First St SW, Rochester MN, 55905, Kullo.Iftikhar@mayo.edu

36 Themistocles Assimes, Stanford University, Stanford, CA, USA.

37 **Abstract (250 words)**

38 **Background:** Predictive performance of polygenic risk scores (PRS) varies across populations. To facilitate
39 equitable clinical use, we developed PRS for coronary heart disease (PRS_{CHD}) for 5 genetic ancestry
40 groups.

41 **Methods:** We derived ancestry-specific and multi-ancestry PRS_{CHD} based on pruning and thresholding
42 (PRS_{P+T}) and continuous shrinkage priors (PRS_{CSx}) applied on summary statistics from the largest multi-
43 ancestry genome-wide meta-analysis for CHD to date, including 1.1 million participants from 5
44 continental populations. Following training and optimization of PRS_{CHD} in the Million Veteran Program,
45 we evaluated predictive performance of the best performing PRS_{CHD} in 176,988 individuals across 9
46 cohorts of diverse genetic ancestry.

47 **Results:** Multi-ancestry PRS_{P+T} outperformed ancestry specific PRS_{P+T} across a range of tuning values. In
48 training stage, for all ancestry groups, PRS_{CSx} performed better than PRS_{P+T} and multi-ancestry PRS
49 outperformed ancestry-specific PRS. In independent validation cohorts, the selected multi-ancestry
50 PRS_{P+T} demonstrated the strongest association with CHD in individuals of South Asian (SAS) and
51 European (EUR) ancestry (OR per 1SD[95% CI]; 2.75[2.41-3.14], 1.65[1.59-1.72]), followed by East Asian
52 (EAS) (1.56[1.50-1.61]), Hispanic/Latino (HIS) (1.38[1.24-1.54]), and weakest in African (AFR) ancestry
53 (1.16[1.11-1.21]). The selected multi-ancestry PRS_{CSx} showed stronger association with CHD in
54 comparison within each ancestry group where the association was strongest in SAS (2.67[2.38-3.00]) and
55 EUR (1.65[1.59-1.71]), progressively decreasing in EAS (1.59[1.54-1.64]), HIS (1.51[1.35-1.69]), and
56 lowest in AFR (1.20[1.15-1.26]).

57 **Conclusions:** Utilizing diverse summary statistics from a large multi-ancestry genome-wide meta-analysis
58 led to improved performance of PRS_{CHD} in most ancestry groups compared to single-ancestry methods.
59 Improvement of predictive performance was limited, specifically in AFR and HIS, despite use of one of
60 the largest and most diverse set of training and validation cohorts to date. This highlights the need for
61 larger GWAS datasets of AFR and HIS individuals to enhance performance of PRS_{CHD}.

62 Introduction

63 Coronary heart disease (CHD) is a leading cause of death in the United States (U.S.) and
64 worldwide ¹. CHD has an estimated heritability of 40-60% and the majority of the heritable risk is
65 attributable to a polygenic component, i.e., the aggregation of modest effects across many genetic
66 variants ². Polygenic risk scores (PRS) capture a proportion of that heritability and are typically
67 constructed by summing the products of the effect-size and the number of risk alleles at associated loci
68 ^{3,4}. PRS for CHD have evolved over the last decade as progressively larger genome wide association
69 studies (GWAS) have been reported ⁵⁻⁸. These PRS have been evaluated in several studies and are
70 associated with incident CHD independent of conventional risk factors such as hypertension,
71 hypercholesterolemia, diabetes, and smoking as well as family history of CHD ⁸⁻¹⁰.

72 Most PRS for CHD have been developed, optimized, and validated in cohorts consisting largely of
73 individuals of European (EUR) ancestry (here and throughout the manuscript ‘ancestry’ refers to genetic
74 ancestry) ¹¹⁻¹⁴. Furthermore, the portability of these PRS to non-EUR groups is impacted by differences in
75 allele frequencies (AF), effect sizes, and linkage disequilibrium (LD) patterns across ancestry groups,
76 typically resulting in reduced predictive performance as studied populations diverge in these factors; an
77 observation most notable between EUR and African (AFR) ancestry populations ^{6,11,15}. We previously
78 observed significantly lower performance of several EUR-derived PRS for CHD in AFR ancestry individuals
79 ^{16,17}. To prevent exacerbation of health disparities in the context of genomic medicine, there is a need to
80 improve performance of PRS for CHD for non-EUR populations.

81 In this study, we leveraged a large scale, ancestrally diverse genome-wide meta-analysis for CHD
82 to construct PRS for CHD optimized for EUR, AFR, Hispanic/Latino (HIS), East Asian (EAS), and South Asian
83 (SAS) ancestries. To this end, we utilized two PRS derivation methods, pruning and thresholding (P+T)
84 and the continuous shrinkage prior based PRS-CSx ^{8,18}. We assessed the performance of the multi-
85 ancestry PRS in individuals with diverse ancestry belonging to 8 independent validation cohorts. Finally, a
86 PRS was selected for clinical implementation in the electronic Medical Records and Genomics (eMERGE)
87 network phase IV study in which PRS-informed risk profiles for several common conditions are being
88 returned to participants ¹⁹.

89

90 Methods

91 GWAS Summary Statistics for PRS Development

92 We developed PRS using both ancestry-specific and multi-ancestry meta-analysis summary
93 statistics from a large-scale multi-ancestry GWAS for CHD including 1.1 million diverse participants with
94 243,392 CHD cases ¹⁷. This diverse meta-analysis included 17,202 AFR, 6,378 HIS, 29,319 EAS, and
95 190,776 EUR individuals with CHD belonging to four cohorts including the Million Veteran Program
96 (MVP), the UK Biobank (UKBB), CARDIoGRAMplusC4D Consortium (2015 release), and Biobank Japan
97 (BBJ) (Figure 1) ^{17,20-22}.

98 We used two distinct methods to construct PRS, namely, pruning and thresholding (P+T) and the
99 continuous shrinkage prior based PRS-CSx ^{8,18}. Ancestry-specific PRS were defined from ancestry-specific

100 GWAS summary statistics (i.e., EUR specific summary statistics were used to develop a EUR specific PRS),
101 and multi-ancestry PRS were defined as PRS derived from multi-ancestry summary statistics. These PRS
102 were then trained and optimized in a separate set of individuals from the MVP and externally validated
103 in several diverse cohorts including the Atherosclerosis Risk in Communities (ARIC)²³, Multi-Ethnic Study
104 of Atherosclerosis (MESA)²⁴, Cardiovascular Health Study (CHS)²⁵, Women's Health Initiative (WHI)²⁶,
105 eMERGE Phases I-III genotyped cohort²⁷, Biobank Japan (BBJ)²⁸, Osaka Acute Coronary Insufficiency
106 (OACIS) study²⁹, the TAICHI Consortium³⁰, and individuals of SAS ancestry from the UKBB³¹ (Table S1;
107 Supplemental File 1).

108 **Pruning and Thresholding (P+T)**

109 We derived two independent sets of PRS (ancestry-specific and multi-ancestry PRS) in two sequential
110 steps: First, we excluded from the base GWAS summary statistics, correlated single nucleotide variants
111 (SNVs) by LD pruning, applying 4 different R^2 thresholding values (0.2, 0.5, 0.8, and 0.9) and 2 different
112 window distances (250kb and 500kb) within which these R^2 were applied. LD pruning for ancestry-
113 specific PRS was performed based on reference panels comprised of 4,000 participants from each
114 respective ancestry (EUR, AFR, HIS, and ASN), selected among MVP participants included in the large-
115 scale GWAS for CHD. The LD pruning for the multi-ancestry PRS was performed on the full subset of
116 16,000 individuals from EUR, AFR, HIS, and ASN as the reference panel. This step generated 8 ancestry-
117 specific summary statistics and 8 multi-ancestry summary statistics for PRS development. Second, for
118 each newly generated summary statistic from step 1, we applied 16 different p -value thresholds (5×10^{-08} ,
119 1×10^{-04} , 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1) (Figure S1; Supplemental
120 File 1). These led to 128 summary statistics within each ancestry, which were used to train the ancestry-
121 specific PRS. Similarly, we obtained 128 multi-ancestry-based summary statistics to train the multi-
122 ancestry PRS (PRS_{P+T}).

123 **Continuous shrinkage (PRS-CSx)**

124 We applied a continuous shrinkage method, PRS-CSx (PRS_{CSx}), on the effect sizes of a subset of 1.4 million
125 well curated HapMap SNVs on each ancestry-specific summary statistic. To identify the optimal shrinkage
126 parameter, we applied 4 different global shrinkage phi parameters (1, $1e^{-02}$, $1e^{-04}$, and $1e^{-06}$). LD
127 reference panels used were EUR, AFR, AMR and EAS from the 1000 Genomes project. The multi-ancestry
128 PRS were constructed from the meta-analysis of ancestry-specific summary statistics obtained after
129 applying the global shrinkage phi. For each ancestry, 4 ancestry-specific newly derived summary statistics
130 were obtained to train ancestry-specific PRS and 4 newly derived multi-ancestry summary statistics were
131 obtained for train the multi-ancestry PRS (Figure S2; Supplemental File 1). A total of 12 ancestry-specific
132 PRS (one for each global shrinkage parameter value used for each ancestry group and 4 multi-ancestry
133 PRS) were chosen for further development (Figure S3; Supplemental File 1).

134 **PRS Training**

135 Following the construction of the ancestry-specific and multi-ancestry PRS_{P+T} and PRS_{CSx} across a range of
136 training specifications, we proceeded to assess their performance in an independent set of prevalent
137 cases and controls from the MVP (Figure 1B, PRS Training) using multivariable logistic regression with
138 adjustment for age at CHD event for cases and age at the last visit in the electronic health record (EHR)
139 for controls, year of birth, sex, and the first 5 principal components (PCs). We compared parameter

140 training on the multi-ancestry reference panel set versus population-specific reference panel. Ancestry-
141 specific PRS were evaluated in the corresponding ancestry, whereas the multi-ancestry PRS were
142 evaluated in each ancestry. PRS with the highest observed odds ratio (OR) for CHD per 1 standard
143 deviation (SD) increase were deemed to have the optimal training parameter values across ancestry
144 populations and subsequently advanced for validation.

145 **PRS Validation in the Million Veteran Program and Additional External Cohorts**

146 Ancestry-specific and multi-ancestry PRS_{P+T} and PRS_{CSX} trained for each genetic ancestry group were
147 validated in an independent cohort from the MVP and several additional diverse cohorts (Figure 1C,
148 Diverse Cohorts for PRS Validation). The MVP validation cohort was restricted to incident cases of CHD
149 occurring after enrollment, and random controls, in a ratio of 1:10 (Figure 1C) as previously described
150 ¹⁷. Four prospective cohorts, namely ARIC, MESA, CHS, WHI, a subset of the UKBB comprised of
151 individuals of SAS ancestry, and additionally eMERGE Phases I-III, contributed CHD incident cases and
152 controls of EUR, AFR, HIS, and SAS ancestry for PRS validation. Validation for EAS ancestry included
153 individuals from multiple case-control studies, namely Han Chinese participants from Taiwan as a part of
154 the TAICHI consortium, as well as Japanese participants from the BBJ and OACIS studies who were not
155 part of the multi-ancestry discovery GWAS ³⁰.

156 Within MVP, we used diagnosis and procedure codes to identify individuals with any clinical
157 manifestation of CHD as previously described (Supplemental File 1) ¹⁷. This definition included both
158 'hard' (e.g., myocardial infarction, revascularization) and 'soft' outcomes (e.g., angina, non-invasive study
159 positive for ischemia). In the 4 external validation NHLBI cohorts and the eMERGE cohort, cases were
160 restricted to myocardial infarction and revascularization. Prevalent cases were defined as all other cases
161 meeting diagnosis/procedure code criteria at the time of enrollment. Additional study details are
162 included in Supplemental File 1.

163 We calculated OR per 1-SD increase in PRS using multivariable logistic regression across all
164 validation cohorts. The dbGaP, eMERGE, and UKBB cohorts were adjusted for genetic ancestry using a
165 continuous correction further defined in the Supplemental File 1 (Figure S4). The two EAS case-control
166 studies were meta-analyzed using a fixed effect inverse-variance weighted model ³². For all external
167 validation cohorts, we additionally estimated OR for CHD for participants in the top 5% of PRS
168 distribution compared to the rest, as well as area under the curve (AUC) discrimination statistic.
169 Calibration was also assessed using the calibration function in the rms package in R to assess portability
170 to cohorts that were not available for meta-analysis (i.e., the non-EAS cohorts) (Figure S5, Supplemental
171 File 2) ^{33,34}.

172 **Results**

173 **PRS Training**

174 **Pruning and Thresholding (P+T)**

175 Performance of the ancestry-specific and multi-ancestry PRS_{P+T} in each population is shown in Figure 2.
176 The multi-ancestry PRS_{P+T} systematically outperformed ancestry-specific PRS_{P+T} with noticeably higher OR
177 per SD except for the HIS ancestry group where the performance was similar (Figure 2, Supplemental
178 Figure S2). The multi-ancestry PRS_{P+T}, performed best in HIS population, followed by the ASN population

179 (1.78 and 1.73 OR per SD, respectively) (Supplemental File 2). Prediction performance of the PRS_{P+T} for
180 each ancestry was optimal at different p -value thresholds (Figure 2, Supplemental Figure S2). The multi-
181 ancestry PRS_{P+T} performed best at $R^2 \leq 0.8$ with LD blocks of 250 kb, p -value threshold of 0.01 for AFR,
182 0.03 for EUR, and 0.30 for HIS. However, the differences between these PRS and the PRS optimized at
183 $R^2 \leq 0.8$ and a p -value = 0.01 were marginal, and the multi-ancestry PRS with a p -value threshold of
184 0.01 was chosen for validation in additional external cohorts.

185 **Continuous shrinkage (PRS-CSx)**

186 The performances of the 12 ancestry-specific PRS_{CSx} and 3 multi-ancestry PRS_{CSx} built using EUR, AFR,
187 HIS, and EAS summary statistics at various global shrinkage ϕ values for tuning ($1e^{-02}$, $1e^{-04}$, and
188 $1e^{-06}$) are shown in Figure 2. For all ancestry groups, $\phi = 1e^{-02}$ resulted in the best predictive
189 performance for PRS_{CSx} and the multi-ancestry PRS outperformed ancestry-specific PRS at this ϕ value.
190 For the EUR population, both the EUR-derived PRS and the multi-ancestry PRS performed similarly, but
191 ASN and HIS populations performed best with the EUR-derived PRS, while the AFR population performed
192 best with the multi-ancestry PRS (Figure 2, Supplemental File 2). Overall, the multi-ancestry PRS_{CSx} for the
193 ASN population resulted in the highest OR per/SD increase followed by EUR and HIS populations where
194 the strength of association was similar, and lowest in the AFR ancestry.

195

196 **PRS Validation**

197 **Million Veteran Program**

198 Ancestry specific PRS_{P+T} predictive performance (OR per 1 -SD increase) for EUR (1.52), AFR (1.19), and
199 HIS (1.81) was compared to the ancestry-specific PRS_{CSx} performance for EUR (1.66), AFR (1.15), HIS
200 (1.42), and ASN (1.32) (Figure 2; Supplemental File 2). This was also compared to the multi-ancestry-
201 based methods using the same PRS training, i.e., the multi-ancestry PRS_{P+T} for EUR (1.57), AFR (1.22), HIS
202 (1.78), and ASN (1.73), as well as PRS_{CSx} for EUR (1.98), AFR (1.23), HIS (1.94), and ASN (2.06) (Figure 2;
203 Supplemental File 2). Of all the methods assessed at this step, the best performing methods tended to
204 be the multi-ancestry PRS_{CSx} and multi-ancestry PRS_{P+T}. However, there were overlapping confidence
205 intervals (CIs) with some single ancestry methods and the single-ancestry PRS_{CSx} for EUR performed well
206 in other ancestries, so we decided to further assess the three methods (Figure 2).

207 We advanced the ancestry optimized PRS_{P+T} and PRS_{CSx}, for validation in an independent set of
208 incident cases and matching controls in ancestry groups of EUR, AFR, HIS, EAS, and SAS individuals.
209 Predictive performances of the multi-ancestry PRS were assessed within each ancestry group in
210 reference to a previously reported genome-wide PRS (i.e., PRS_{metaGRS}¹⁰) constructed using a cohort of
211 predominantly of EUR ancestry (Figure 3)¹⁷. In this independent validation cohort, the multi-ancestry
212 PRS_{P+T} and PRS_{CSx} had a higher predictive performance compared to metaGRS (Figure 3). The multi-
213 ancestry PRS_{CSx} had a relative increase in the estimated OR per 1-SD of 12% and 23% in reference to
214 PRS_{P+T} and PRS_{metaGRS}, respectively, averaged across all three genetic ancestries.

215

216 **Additional External Validation Cohorts**

217 The best performing PRS_{P+T} were further validated in several additional cohort and case-control studies
218 of CHD including EUR, AFR, HIS, EAS, and SAS participants (Table 1). ORs for ancestry-specific and multi-
219 ancestry PRS_{P+T} ranged from 1.16 in AFR to 2.75 in SAS and were comparable to published reports,
220 despite inclusion of the diverse meta-analysis of GWAS (Supplemental File 2)^{6,17,35,36}. All populations had
221 OR estimates for the top 5% vs the rest of the population ≥ 2.16 for PRS_{P+T} except for AFR (1.68).

222 The two best performing PRS_{CSX} in the training dataset, a EUR-tuned PRS and a multi-ancestry
223 PRS, both with a tuning global phi value of $1e^{-02}$, demonstrated similar performances in our validation
224 cohorts (Table 1, Table S2; Supplemental File 1) as the multi-ancestry PRS marginally outperformed the
225 EUR-tuned PRS in all but the AFR and HIS cohorts. Point estimates of the OR for subjects in the top 5th
226 percentile of scores compared to the remaining participants shifted trend compared to those observed
227 for the ORs per 1-SD for AFR, HIS, and SAS populations, but these differences were in the context of
228 mostly overlapping 95% confidence intervals. When comparing the multi-ancestry PRS_{P+T} to PRS_{CSX}, the
229 point estimates of ORs were similar but higher for the multi-ancestry PRS_{CSX} for EUR, AFR, HIS, and EAS
230 populations. The OR per 1-SD was lower for the multi-ancestry PRS_{CSX} for the SAS population (Table 1).

231

232 Discussion

233 Using summary statistics from the largest multi-ancestry GWAS meta-analysis for CHD to date and 9
234 independent validation cohorts, cumulatively comprised of 1.1 million diverse participants including
235 nearly a quarter of a million CHD cases of EUR, AFR, HIS, EAS, and SAS descent¹⁷, we developed, trained,
236 and validated multi-ancestry and ancestry-specific PRS models to address the gap in predictive
237 performance that currently exists between EUR and non-EUR ancestries.

238 We observed that the use of summary statistics from a multi-ancestry GWAS meta-analysis, in
239 comparison to the use of ancestry-specific summary statistics, improved PRS performance in majority of
240 the ancestry groups. PRS that leveraged shared information between ancestries to estimate SNP weights
241 (i.e., PRS_{CSX}) modestly outperformed the P+T method (i.e., PRS_{P+T}). Based on the multi-ancestry informed
242 PRS_{CSX}, individuals in the high-genetic risk group (i.e., top 5% of the PRS distribution) compared to the
243 remaining participants in the respective ancestry group (EUR, AFR, HIS, EAS, and SAS), had 2.5-fold, 1.7-
244 fold, 2.5-fold, 2.3-fold, and 5-fold increased risk of CHD, respectively. These results collectively highlight
245 complementary effects of integrating summary statistics from multiple ancestries and the use of PRS
246 derivation methods that leverage shared information and LD diversity between ancestry groups to
247 improve polygenic risk prediction for CHD.

248 Although remarkable progress has been achieved to date in both genomic discovery and
249 polygenic risk prediction among EUR cohorts^{5,7-10,37-39}, similar progress has not occurred among non-EUR
250 populations due to their underrepresentation in genomic studies¹¹⁻¹⁴. In recent years, the number of
251 large-scale multi-ancestry GWAS and polygenic risk prediction studies have increased with the
252 establishment of ancestrally diverse biobanks and collaborations efforts^{17,18,30,40-43}. Several multi-
253 ancestry genomic studies, including for glycemic, hematologic and lipid traits as well as disease
254 phenotypes such as type 2 diabetes and CHD, have increased the number of discovered loci, and
255 improved fine-mapping and cross-population polygenic risk prediction with inclusion of non-EUR

256 participants^{17,40-42,44}. Our findings are consistent with these results in that integration of summary
257 statistics from several distinct ancestry groups improved predictive performance of PRS for all ancestries,
258 including EUR descent. One possible explanation for these observations is identification of potential
259 causal variants that are more likely to be shared between ancestries but are obscured by population-
260 specific LD patterns^{14,45}. Another likely contributing factor to improved PRS performance is reduced
261 noise in SNV effect size estimates resulting from both weighted average of population-level estimates
262 and increased total sample size^{46,47}.

263 Despite the use of the largest ancestrally diverse cohort available to date, the improvement in
264 the predictive performance of PRS_{CHD} was limited in individuals of AFR ancestry compared to other
265 ancestry groups. Prior reports investigating portability of PRS between populations noted that prediction
266 performance across a range of traits and phenotypes^{6,11,15,16,48,49} decayed with increasing genetic
267 distance between study cohorts. Among the continental ancestry groups included in this study, AFR is
268 the most genetically distant population from EUR and hence the modest increase in prediction
269 performance with a multi-ancestry PRS_{CHD} compared to the ancestry-specific counterpart. A recent
270 report showed similar heritability for CHD in the major continental ancestry groups but absence of two
271 common haplotypes at the 9p21 locus in AFR individuals, which corresponds to the largest effect locus in
272 EUR ancestry individuals¹⁷. These findings suggest potentially a larger role of ancestry-specific causal
273 variants in individuals of African origin with regards to heritability for CHD.

274 Although the strength of association of PRS with CHD varied between ancestry groups, it is
275 important to consider epidemiological differences in CHD risk across these populations. In clinical
276 practice, primary prevention guidelines for CHD use absolute risk estimates for clinical decision making,
277 such as 10-year or lifetime risk of a CHD event⁵⁰. Individuals are typically classified into different risk
278 groups (e.g., low, borderline, intermediate, high risk) with a correlating intensity of pursued preventive
279 measures. In the United States, African American and South Asian populations have substantially higher
280 atherosclerotic cardiovascular disease (ASCVD) related mortality rates compared to non-Hispanic whites
281^{1,51}. Therefore, in a future risk model for ASCVD similar to the pooled cohort equation⁵², incorporation of
282 a PRS for CHD with a narrower risk gradient in African Americans, compared to a much wider gradient in
283 non-Hispanic whites, could have more impact on re-classification into a higher risk group as we have
284 previously shown⁶.

285 Implementation of PRS in the clinical setting has begun for CHD, including at Mayo Clinic, where
286 a PRS for CHD is available in the clinical setting, based on the results of the MIGENES clinical trial⁵³. The
287 eMERGE Network, in its phase IV study is returning risk assessments to participants for 11 common
288 conditions, including CHD¹⁹. The multi-ancestry PRS_{p+T} for CHD validated in this study¹⁹ will be returned
289 to eMERGE participants. One of the major challenges in the clinical use of PRS include variable
290 performance between genetic ancestry populations^{11,15}. Developing robust PRS for diverse ancestry
291 groups is crucial to avoid worsening existing health disparities¹¹ and a National Institute of Health (NIH)
292 funded initiative is addressing this as a priority⁵⁴. The active recruitment and inclusion of diverse
293 participants and continued development of novel PRS methods that target improvement of cross-
294 population prediction using a variety of approaches (e.g., incorporation of local ancestry⁵⁵, weighting by
295 trans-ancestry genetic correlation⁵⁶, and informing by fine-mapping and functional annotation^{57,58}) will

296 be necessary for equitable implementation of PRS. Consequently, we anticipate that PRS for CHD will
297 continue to evolve and improve over time.

298

299 **Study Limitations**

300 Despite the large and diverse composition of our study, the external validation for the SAS ancestry was
301 limited to a single cohort with a modest number of cases, reducing the precision of the associated risk
302 estimates. We were not able to include smoking status or family history in the models as the data was
303 not available for all cohorts, and this may have affected the strength of the association of PRS with CHD
304 in our analyses.

305

306 **Conclusions**

307 We demonstrated that incorporation of summary statistics from diverse genetic ancestry groups, as
308 opposed to individual ancestry groups alone, and leveraging shared information between these
309 populations, led to improved performance of PRS_{CHD} in majority of the ancestry groups. Despite
310 utilization of one of the largest and most ancestrally diverse set of training and validation cohorts to
311 date, the gain in predictive performance for AFR was limited. Ongoing work is needed to narrow the
312 persistent performance gap for AFR ancestry individuals. Increasing AFR representation at each stage of
313 PRS development is necessary to lessen performance disparities, and such efforts should be a priority for
314 the community of genomics researchers.

315

316 **Acknowledgements**

317 We acknowledge the investigators and participants of the electronic Medical Records and Genomics
318 (eMERGE) Network. Infrastructure for the CHARGE Consortium is supported in part by the National
319 Heart, Lung, and Blood Institute (NHLBI) grant R01HL105756. This work was also supported in part by
320 the National Institutes of Health, National Heart, Lung, Long and Blood Institute (NHLBI) contract
321 1R01HL151855, R01HL146860, and the National Institute of Diabetes and Digestive and Kidney Diseases
322 contract UM1DK078616.

323

324 **Sources of Funding**

325 This work was supported by grants from the Polygenic Risk Methods in Diverse Populations (PRIMED)
326 Consortium through the National Human Genome Research Institute (NHGRI): grant U01 HG11710, the
327 electronic Medical Records and Genomics (eMERGE) Network funded by the NHGRI: grant U01
328 HG06379, a National Heart, Lung, and Blood: grant K24 HL137010, the Clinical Genome Resource
329 (ClinGEN) funded by the NHGRI: grant HG09650, and R35 GM140487.

330

331 **Disclosures**

332 **Conflict of Interest.** The authors declare that they have no conflict of interest.

333 **Human and Animal Rights and Informed Consent.** This article used data from previously published
334 human studies.

335 References

- 336 1. Tsao CW, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, Baker-Smith CM, Beaton AZ,
337 Boehme AK, Buxton AE, et al. Heart Disease and Stroke Statistics—2023 Update: A Report From
338 the American Heart Association. *Circulation*. 2023;147. doi: 10.1161/cir.0000000000001123
- 339 2. Kullo IJ, Ding K. Mechanisms of Disease: the genetic basis of coronary heart disease. *Nature*
340 *Clinical Practice Cardiovascular Medicine*. 2007;4:558-569. doi: 10.1038/ncpcardio0982
- 341 3. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics*.
342 2015;31:1466-1468. doi: <https://doi.org/10.1093/bioinformatics/btu848>
- 343 4. Kullo IJ, Lewis CM, Inouye M, Martin AR, Ripatti S, Chatterjee N. Polygenic scores in biomedical
344 research. *Nature Reviews Genetics*. 2022. doi: 10.1038/s41576-022-00470-z
- 345 5. Tikkanen E, Havulinna AS, Palotie A, Salomaa V, Ripatti S. Genetic Risk Prediction and a 2-Stage
346 Risk Screening Strategy for Coronary Heart Disease. *Arteriosclerosis, Thrombosis, and Vascular*
347 *Biology*. 2013;33:2261-2266. doi: 10.1161/atvbaha.112.301120
- 348 6. Dikilitas O, Schaid DJ, Tcheandjieu C, Clarke SL, Assimes TL, Kullo IJ. Use of Polygenic Risk Scores
349 for Coronary Heart Disease in Ancestrally Diverse Populations. *Current Cardiology Reports*.
350 2022;24:1169-1177. doi: 10.1007/s11886-022-01734-0
- 351 7. O'Sullivan JW, Raghavan S, Marquez-Luna C, Luzum JA, Damrauer SM, Ashley EA, O'Donnell CJ,
352 Willer CJ, Natarajan P. Polygenic Risk Scores for Cardiovascular Disease: A Scientific Statement
353 From the American Heart Association. *Circulation*. 2022;146. doi:
354 10.1161/cir.0000000000001077
- 355 8. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA,
356 Ellinor PT, et al. Genome-wide polygenic scores for common diseases identify individuals with
357 risk equivalent to monogenic mutations. *Nature Genetics*. 2018;50:1219-1224. doi:
358 10.1038/s41588-018-0183-z
- 359 9. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, Guiducci C, Perola
360 M, Jula A, Sinisalo J, et al. A multilocus genetic risk score for coronary heart disease: case-control
361 and prospective cohort analyses. *The Lancet*. 2010;376:1393-1400.
- 362 10. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, Lai FY, Kaptoge S,
363 Brozynska M, Wang T, et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000
364 Adults: Implications for Primary Prevention. *J American Coll Cardiol*. 2018;72:1883-1893.
- 365 11. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic
366 risk scores may exacerbate health disparities. *Nature Genetics*. 2019;51:584-591. doi:
367 10.1038/s41588-019-0379-x
- 368 12. Manolio TA. Using the Data We Have: Improving Diversity in Genomic Research. *The American*
369 *Journal of Human Genetics*. 2019;105:233-236. doi: 10.1016/j.ajhg.2019.07.008
- 370 13. Clarke SL, Assimes TL, Tcheandjieu C. The Propagation of Racial Disparities in Cardiovascular
371 Genomics Research. *Circulation: Genomic and Precision Medicine*. 2021;14. doi:
372 10.1161/circgen.121.003178
- 373 14. Gurdasani D, Barroso I, Zeggini E, Sandhu MS. Genomics of disease risk in globally diverse
374 populations. *Nature Reviews Genetics*. 2019;20:520-535. doi: 10.1038/s41576-019-0144-0
- 375 15. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD,
376 Kenny EE. Human Demographic History Impacts Genetic Risk Prediction across Diverse
377 Populations. *The American Journal of Human Genetics*. 2017;100:635-649. doi:
378 10.1016/j.ajhg.2017.03.004

- 379 16. Dikilitas O, Schaid DJ, Kosel ML, Carroll RJ, Chute CG, Denny JC, Fedotov A, Feng Q, Hakonarson
380 H, Jarvik GP, et al. Predictive Utility of Polygenic Risk Scores for Coronary Heart Disease in Three
381 Major Racial and Ethnic Groups. *The American Journal of Human Genetics*. 2020;106:707-716.
382 doi: 10.1016/j.ajhg.2020.04.002
- 383 17. Tcheandjieu C, Zhu X, Hilliard AT, Clarke SL, Napolioni V, Ma S, Lee KM, Fang H, Chen F, Lu Y, et al.
384 Large-scale genome-wide association study of coronary artery disease in genetically diverse
385 populations. *Nature Medicine*. 2022;28:1679-1692. doi: 10.1038/s41591-022-01891-3
- 386 18. Ge T, Irvin MR, Patki A, Srinivasasainagendra V, Lin Y-F, Tiwari HK, Armstrong ND, Benoit B, Chen
387 C-Y, Choi KW, et al. Development and validation of a trans-ancestry polygenic risk score for type 2
388 diabetes in diverse populations. *Genome Medicine*. 2022;14. doi: 10.1186/s13073-022-01074-2
- 389 19. Linder J, Allworth A, Bland ST, Caraballo PJ, Chisholm R, Clayton EW, Crosslin D, Dikilitas O,
390 DiVietro A, Esplin ED, et al. Returning integrated genomic risk and clinical recommendations: the
391 eMERGE study. *Genetics in Medicine*. 2023. doi: <https://doi.org/10.1016/j.gim.2023.100006>
- 392 20. Van Der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an Expanded View on
393 the Genetic Architecture of Coronary Artery Disease. *Circulation Research*. 2018;122:433-443.
394 doi: 10.1161/circresaha.117.312086
- 395 21. Ishigaki K, Akiyama M, Kanai M, Takahashi A, Kawakami E, Sugishita H, Sakaue S, Matoba N, Low
396 S-K, Okada Y, et al. Large-scale genome-wide association study in a Japanese population
397 identifies novel susceptibility loci across different diseases. *Nature Genetics*. 2020;52:669-679.
398 doi: 10.1038/s41588-020-0640-3
- 399 22. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP,
400 Hopewell JC, et al. A comprehensive 1000 Genomes-based genome-wide association meta-
401 analysis of coronary artery disease. *Nature Genetics*. 2015;47:1121-1130. doi: 10.1038/ng.3396
- 402 23. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and
403 objectives. *American Journal of Epidemiology*. 1989;129:687-702.
- 404 24. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, R. JD, Kronmal
405 R, Liu K, et al. Multi-Ethnic Study of Atherosclerosis: Objectives and Design. *American Journal of
406 Epidemiology*. 2002;156:871-881. doi: <https://doi.org/10.1093/aje/kwf113>
- 407 25. (CHS) MftCHSRG, Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH,
408 Manolio TA, Mittelmark MB, et al. The cardiovascular health study: Design and rationale. *Annals
409 of Epidemiology*. 1991;1:263-276. doi: [https://doi.org/10.1016/1047-2797\(91\)90005-W](https://doi.org/10.1016/1047-2797(91)90005-W)
- 410 26. Group TWsHIS. Design of the Women's Health Initiative Clinical Trial and Observational Study.
411 *Controlled Clinical Trials*. 1998;19:61-109. doi: [https://doi.org/10.1016/S0197-2456\(97\)00078-0](https://doi.org/10.1016/S0197-2456(97)00078-0)
- 412 27. Stanaway IB, Hall TO, Rosenthal EA, Palmer M, Naranbhai V, Knevel R, Namjou-Kahles B, Carroll
413 RJ, Kiryluk K, Gordon AS, et al. The eMERGE genotype set of 83,717 subjects imputed to
414 ~40 million variants genome wide and association with the herpes zoster medical record
415 phenotype. *Genetic Epidemiology*. 2019;43:63-81. doi: 10.1002/gepi.22167
- 416 28. Nagai A, Hirata M, Kamatani Y, Muto K, Matsuda K, Kiyohara Y, Ninomiya T, Tamakoshi A,
417 Yamagata Z, Mushiroda T, et al. Overview of the Biobank Japan Project: Study design and profile.
418 *Journal of Epidemiology*. 2017;27:S2-S8. doi: 10.1016/j.je.2016.12.005
- 419 29. Kurotobi T, Sato H, Kinjo K, Nakatani D, Mizuno H, Shimizu M, Imai K, Hori M, Group O. Reduced
420 Collateral Circulation to the Infarct-Related Artery in Elderly Patients with Acute Myocardial
421 Infarction. *J American Coll Cardiol*. 2004;44:28-34. doi: doi:10.1016/j.jacc.2003.11.066
- 422 30. Assimes TL, Lee IT, Juang J-M, Guo X, Wang T-D, Kim ET, Lee W-J, Absher D, Chiu Y-F, Hsu C-C, et
423 al. Genetics of Coronary Artery Disease in Taiwan: A CardiometaboChip Study by the Taichi
424 Consortium. *PLOS ONE*. 2016;11:e0138014. doi: 10.1371/journal.pone.0138014

- 425 31. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray
426 M, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of
427 Complex Diseases of Middle and Old Age. *PLOS Medicine*. 2015;12:e1001779. doi:
428 10.1371/journal.pmed.1001779
- 429 32. Evangelou E, Ioannidis JPA. Meta-analysis methods for genome-wide association studies and
430 beyond. *Nature Reviews Genetics*. 2013;14:379-389. doi: 10.1038/nrg3472
- 431 33. Harrell Jr. FE. rms: Regression Modeling Strategies. R package version 6.3-0. 2022.
- 432 34. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles
433 heel of predictive analytics. *BMC Medicine*. 2019;17. doi: 10.1186/s12916-019-1466-7
- 434 35. Mars N, Kerminen S, Feng Y-CA, Kanai M, Lall K, Thomas LF, Skogholt AH, dellaBriotta Parolo P,
435 Project TBJ, FinnGen, et al. Genome-wide risk prediction of common diseases across ancestries
436 in one million people. *Cell Genomics*. 2022;2.
- 437 36. Wang M, Menon R, MSanghamitra M, Patel AP, Chaffin M, Tanneeru D, Deshmukh M, Matthew
438 O, Apte S, Devanboo CS, et al. Validation of a Genome-Wide Polygenic Score for Coronary Artery
439 Disease in South Asians. *Journal of the American College of Cardiology*. 2020;76:703-714. doi:
440 <https://doi.org/10.1016/j.jacc.2020.06.024>
- 441 37. Tada H, Melander O, Louie JZ, Catanese JJ, Rowland CM, Devlin JJ, Kathiresan S, Shiffman D. Risk
442 prediction by genetic risk scores for coronary heart disease is independent of self-reported
443 family history. *European Heart Journal*. 2016;37:561-567. doi: 10.1093/eurheartj/ehv462
- 444 38. Ding K, Bailey KR, Kullo IJ. Genotype-informed estimation of risk of coronary heart disease based
445 on genome-wide association data linked to the electronic medical record. *BMC Cardiovascular*
446 *Disorders*. 2011;11:66. doi: 10.1186/1471-2261-11-66
- 447 39. Abraham G, Havulinna AS, Bhalala OG, Byars SG, De Livera AM, Yetukuri L, Tikkanen E, Perola M,
448 Schunkert H, Sijbrands EJ, et al. Genomic prediction of coronary heart disease. *European Heart*
449 *Journal*. 2016;37:3267-3278. doi: 10.1093/eurheartj/ehw450
- 450 40. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, Yu GZ, Rueger S, Speidel L, Kim
451 YJ, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse
452 populations for discovery and translation. *Nature Genetics*. 2022;54:560-572. doi:
453 10.1038/s41588-022-01058-3
- 454 41. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan JA, Willems SM, Wu Y, Zhang X,
455 Horikoshi M, et al. The trans-ancestral genomic architecture of glycemic traits. *Nature Genetics*.
456 2021;53:840-860. doi: 10.1038/s41588-021-00852-9
- 457 42. Chen M-H, Raffield LM, Mousas A, Sakaue S, Huffman JE, Moscati A, Trivedi B, Jiang T, Akbari P,
458 Vuckovic D, et al. Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals
459 from 5 Global Populations. *Cell*. 2020;182:1198-1213.e1114. doi: 10.1016/j.cell.2020.06.045
- 460 43. Lu X, Liu Z, Cui Q, Liu F, Li J, Niu X, Shen C, Hu D, Huang K, Chen J, et al. A polygenic risk score
461 improves risk stratification of coronary artery disease: a large-scale prospective Chinese cohort
462 study. *European Heart Journal*. 2022;43:1702-1711. doi:
463 <https://doi.org/10.1093/eurheartj/ehac093>
- 464 44. Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJM, Ramdas S, Surakka I, Ntalla I, Vedantam S,
465 Winkler TW, et al. The power of genetic diversity in genome-wide association studies of lipids.
466 *Nature*. 2021;600:675-679. doi: 10.1038/s41586-021-04064-3
- 467 45. Evans DM, Cardon LR. A Comparison of Linkage Disequilibrium Patterns and Estimated
468 Population Recombination Rates across Multiple Populations. *The American Journal of Human*
469 *Genetics*. 2005;76:681-687. doi: 10.1086/429274

- 470 46. Cavazos TB, Witte JS. Inclusion of variants discovered from diverse populations improves
471 polygenic risk score transferability. *HGG Adv*. 2021;2:100017. doi:
472 <https://doi.org/10.1016/j.xhgg.2020.100017>
- 473 47. Zhang Y, Qi G, Park J-H, Chatterjee N. Estimation of complex effect-size distributions using
474 summary-level statistics from genome-wide association studies across 32 complex traits. *Nature*
475 *Genetics*. 2018;50:1318-1326. doi: 10.1038/s41588-018-0193-x
- 476 48. Privé F, Aschard H, Carmi S, Folkersen L, Hoggart C, O'Reilly PF, Vilhjálmsón BJ. Portability of 245
477 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the
478 same cohort. *The American Journal of Human Genetics*. 2022;109:12-23. doi:
479 [10.1016/j.ajhg.2021.11.008](https://doi.org/10.1016/j.ajhg.2021.11.008)
- 480 49. Fahed AC, Aragam KG, Hindy G, Chen Y-DI, Chaudhary K, Dobbyn A, Krumholz HM, Sheu WHH,
481 Rich SS, Rotter JI, et al. Transethnic Transferability of a Genome-Wide Polygenic Score for
482 Coronary Artery Disease. *Circulation: Genomic and Precision Medicine*. 2021;14. doi:
483 [10.1161/circgen.120.003092](https://doi.org/10.1161/circgen.120.003092)
- 484 50. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, Himmelfarb CD,
485 Khera AV, Lloyd-Jones D, McEvoy JW, et al. 2019 ACC/AHA Guideline on the Primary Prevention
486 of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart
487 Association Task Force on Clinical Practice Guidelines. *Journal of the American College of*
488 *Cardiology*. 2019;74:177-232. doi: <https://doi.org/10.1016/j.jacc.2019.03.010>
- 489 51. Volgman AS, Palaniappan LS, Aggarwal NT, Gupta M, Khandelwal A, Krishnan AV, Lichtman JH,
490 Mehta LS, Patel HN, Shah KS, et al. Atherosclerotic Cardiovascular Disease in South Asians in the
491 United States: Epidemiology, Risk Factors, and Treatments: A Scientific Statement From the
492 American Heart Association. *Circulation*. 2018;138:CIR.0000000000. doi:
493 [10.1161/cir.0000000000000580](https://doi.org/10.1161/cir.0000000000000580)
- 494 52. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, Greenland P, Lackland
495 DT, Levy D, O'Donnell CJ, et al. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular
496 Risk. *Circulation*. 2014;129:S49-S73. doi: 10.1161/01.cir.0000437741.48606.98
- 497 53. Kullo IJ, Jouni H, Olson JE, Montori VM, Bailey KR. Design of a randomized controlled trial of
498 disclosing genomic risk of coronary heart disease: the Myocardial Infarction Genes (MI-GENES)
499 study. *BMC Medical Genomics*. 2015;8. doi: 10.1186/s12920-015-0122-0
- 500 54. The Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium.
501 <https://primedconsortium.org/>. 2021.
- 502 55. Atkinson EG, Maihofer AX, Kanai M, Martin AR, Karczewski KJ, Santoro ML, Ulirsch JC, Kamatani
503 Y, Okada Y, Finucane HK, et al. Tractor uses local ancestry to enable the inclusion of admixed
504 individuals in GWAS and to boost power. *Nature Genetics*. 2021;53:195-204. doi:
505 [10.1038/s41588-020-00766-y](https://doi.org/10.1038/s41588-020-00766-y)
- 506 56. Cai M, Xiao J, Zhang S, Wan X, Zhao H, Chen G, Yang C. A unified framework for cross-population
507 trait prediction by leveraging the genetic correlation of polygenic traits. *The American Journal of*
508 *Human Genetics*. 2021;108:632-655. doi: 10.1016/j.ajhg.2021.03.002
- 509 57. Weissbrod O, Kanai M, Shi H, Gazal S, Peyrot WJ, Khera AV, Okada Y, Matsuda K, Yamanashi Y,
510 Furukawa Y, et al. Leveraging fine-mapping and multipopulation training data to improve cross-
511 population polygenic risk scores. *Nature Genetics*. 2022;54:450-458. doi: 10.1038/s41588-022-
512 01036-9
- 513 58. Amariuta T, Ishigaki K, Sugishita H, Ohta T, Koido M, Dey KK, Matsuda K, Murakami Y, Price AL,
514 Kawakami E, et al. Improving the trans-ancestry portability of polygenic risk scores by prioritizing

515 variants in predicted cell-type-specific regulatory elements. *Nature Genetics*. 2020;52:1346-
516 1354. doi: 10.1038/s41588-020-00740-8

517

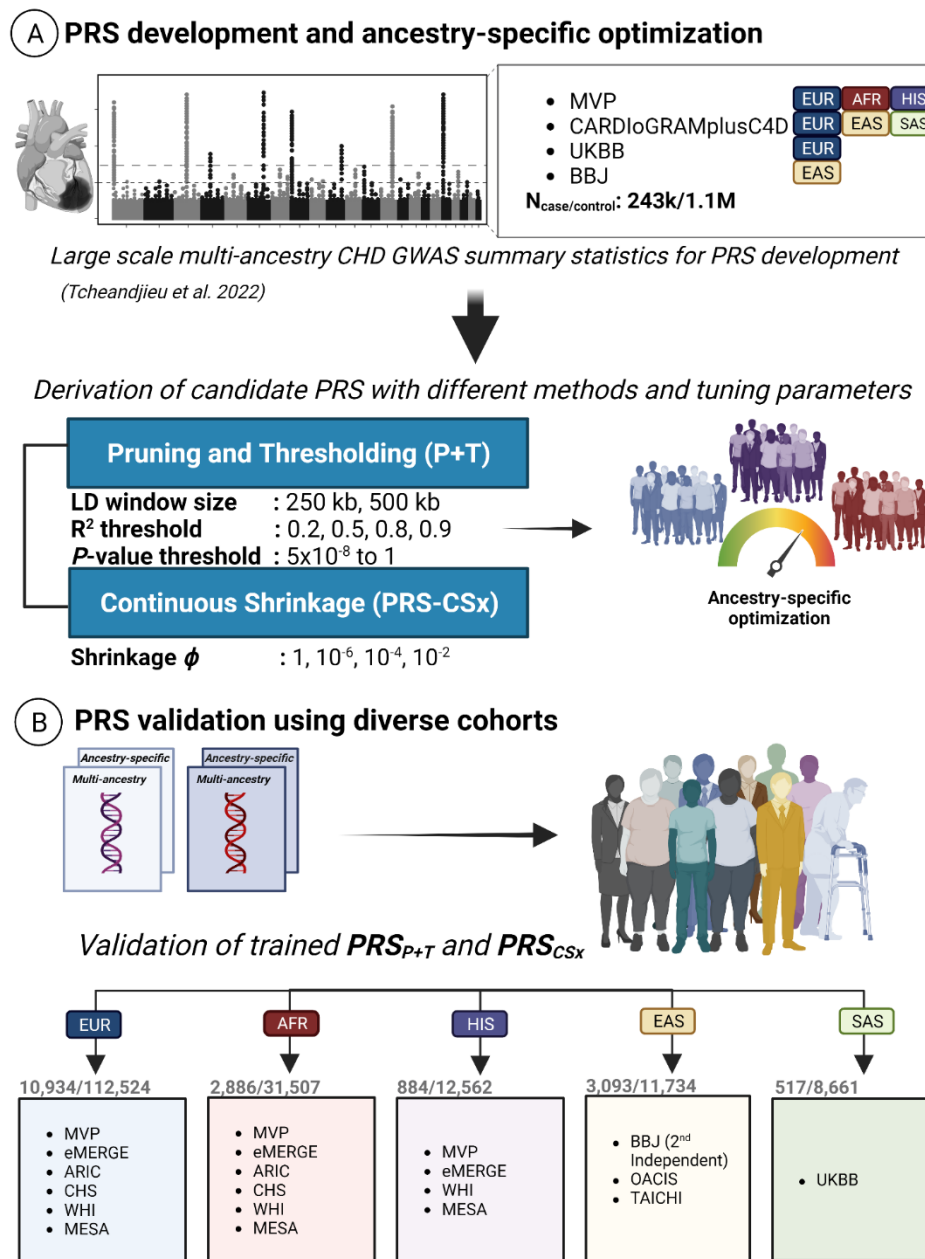
518

519 **Table 1. Odds Ratios for incident CHD for multi-ancestry PRS_{P+T} and PRS_{CSx} in diverse ancestry cohorts.**

Ancestry	Age ^a (mean±SD ^b)	Cases/Controls	Method	AUC ^c	OR ^d (95% CI ^e) per 1 SD	P-value	OR (95% CI) Top 5% vs Rest	P-value
EUR	52.4 ± 15.5	4,970/47,732	PRS _{P+T}	0.773	1.65 (1.59-1.72)	3.00E-159	2.30 (2.07-2.56)	5.71E-55
			PRS _{CSx}	0.774	1.65 (1.59-1.71)	5.19E-171	2.48 (2.23-2.77)	5.88E-61
AFR	53.4 ± 14.7	1,359/15,649	PRS _{P+T}	0.735	1.16 (1.11-1.21)	2.86E-12	1.68 (1.39-2.03)	5.51E-08
			PRS _{CSx}	0.736	1.20 (1.15-1.26)	7.46E-14	1.74 (1.41-2.15)	3.03E-07
HIS	54.8 ± 14.3	314/5,824	PRS _{P+T}	0.699	1.38 (1.24-1.54)	5.89E-09	2.16 (1.47-3.19)	1.02E-04
			PRS _{CSx}	0.706	1.51 (1.35-1.69)	7.48E-13	2.57 (1.77-3.73)	7.98E-07
EAS	65.4 ± 12.9	6,321/16,430	PRS _{P+T}	0.748	1.56 (1.50-1.61)	2.97E-146	2.47 (2.10-2.90)	1.24E-39
			PRS _{CSx}	0.762	1.59 (1.54-1.64)	2.41E-160	2.34 (2.06-2.66)	1.78E-28
SAS	53.2 ± 8.4	517/8,661	PRS _{P+T}	0.786	2.75 (2.41-3.14)	9.44E-52	3.95 (3.03-5.15)	3.07E-24
			PRS _{CSx}	0.803	2.67 (2.38-3.00)	1.48E-63	4.92 (3.81-6.35)	3.90E-34

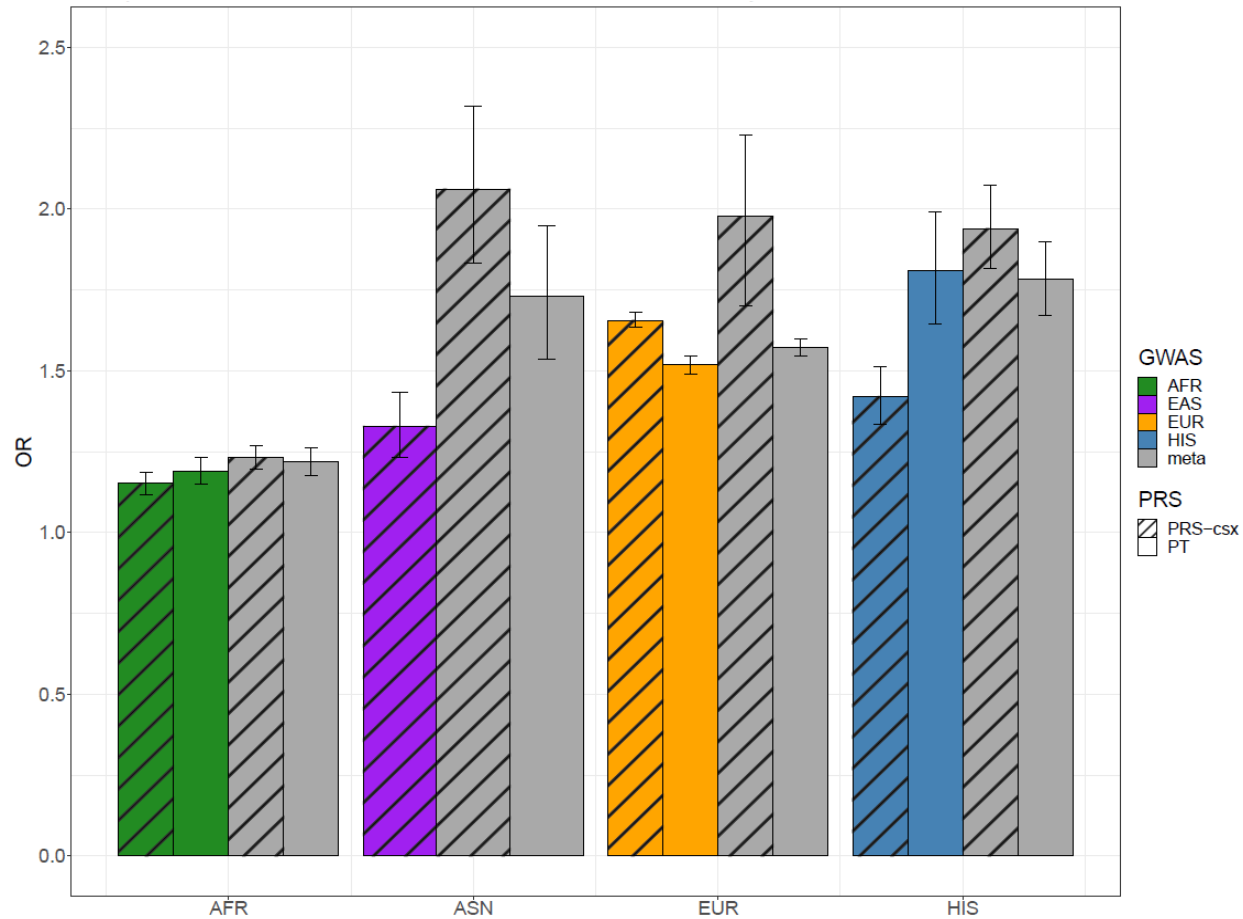
520 ^a Age- Age at enrollment
521 ^b SD- Standard Deviation
522 ^c AUC-Area under the Curve
523 ^d OR- Odds Ratio
524 ^e CI- Confidence Interval

525 **Figures**



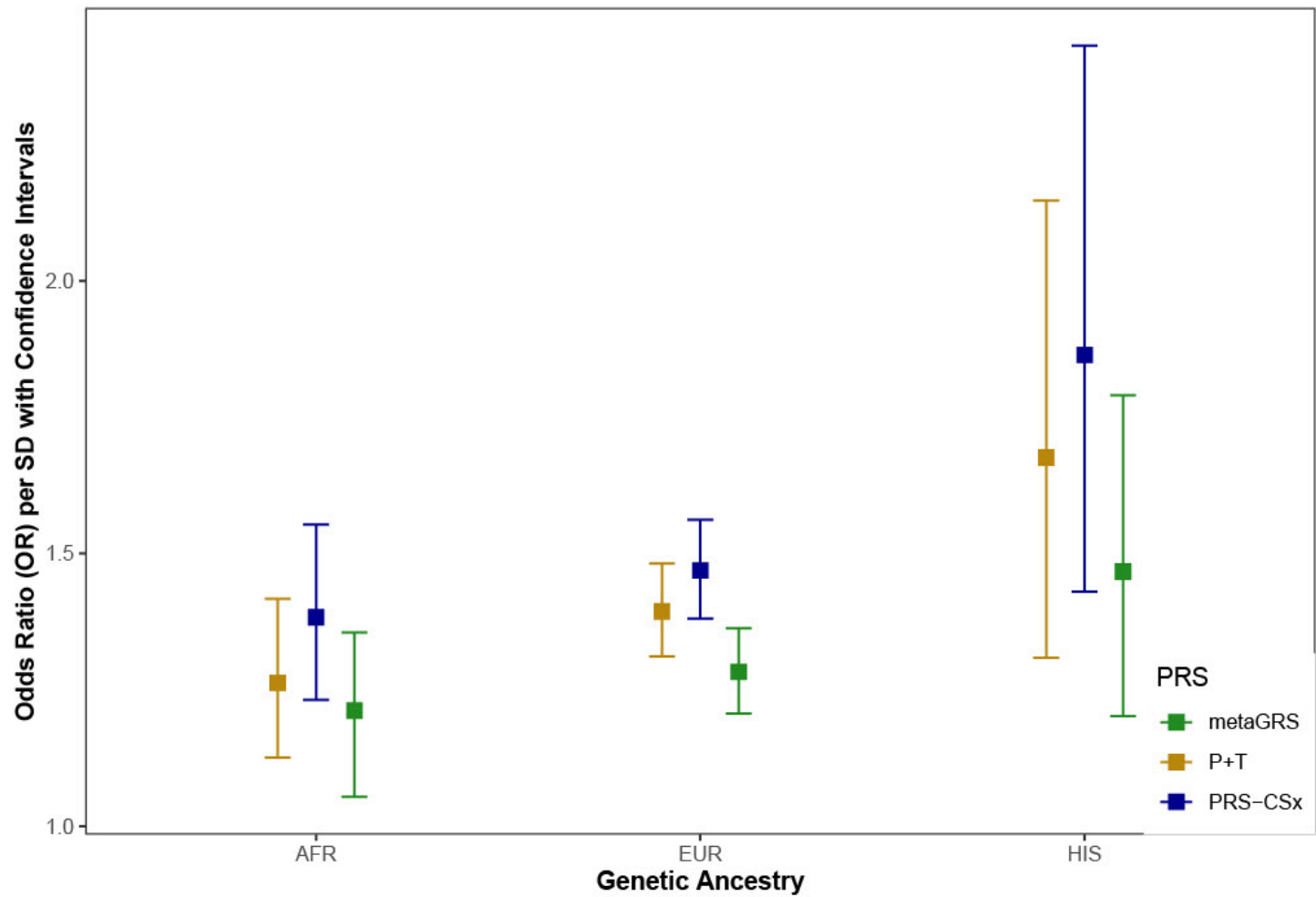
526

527 **Figure 1.** Polygenic Risk Score development using independent MVP cohorts of diverse ancestry.



529

530 **Figure 2.** Performance of PRS-CSx (solid bars) or P+T (dashed bars) across genetic ancestry groups when utilizing the diverse MVP training cohort. The colors
 531 represent the GWAS summary statistics used to construct the PRS (green for AFR, purple for EAS, orange for EUR, and grey for the multi-ancestry meta-
 532 analysis). The Odds Ratios (ORs) per 1 standard deviation (SD) increase with confidence intervals (CIs) in the PRS are represented on the Y-axis and the
 533 populations on which the PRS is trained are on the X-axis.



534

535 **Figure 3.** Comparison of a prior PRS (metaGRS) and two new PRS using multi-ancestry summary statistics for the prediction of coronary heart
 536 disease (CHD) using the ancestrally diverse training cohort of the MVP. Odds Ratios (ORs) per standard deviation (SD) with confidence intervals
 537 (CIs) are shown for each genetic ancestry group as determined in the methods as a result of metaGRS, P+T, and PRS-CSx PRS methods being
 538 performed on the MVP training cohort.