An Ensemble Penalized Regression Method for Multi-ancestry Polygenic Risk Prediction

Jingning Zhang[1,+], Jianan Zhan[2], Jin Jin[3], Cheng Ma[4], Ruzhang Zhao[1], Jared O'Connell[2], Yunxuan Jiang[2], 23andMe Research Team[*], Bertram L. Koelsch[2], Haoyu Zhang[5,6], Nilanjan Chatterjee[1,7,+]

[1] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

[2] 23andMe Inc., Sunnyvale, CA, USA.

[3]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA, USA.

[4] Department of Statistics, University of Michigan, Ann Arbor, MI, USA.

[5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA.

[6] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

[7] Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD, USA.

[*]A list of authors and their affiliations appears at the end of the paper.

[+] Correspondence to: Jingning Zhang (jingningzhang238@gmail.com) and Nilanjan Chatterjee (nilanjan@jhu.edu)

Abstract

Great efforts are being made to develop advanced polygenic risk scores (PRS) to improve the prediction of complex traits and diseases. However, most existing PRS are primarily trained on European ancestry populations, limiting their transferability to non-European populations. In this article, we propose a novel method for generating multi-ancestry Polygenic Risk scOres based on enSemble of PEnalized Regression models (PROSPER). PROSPER integrates genome-wide association studies (GWAS) summary statistics from diverse populations to develop ancestry-specific PRS with improved predictive power for minority populations. The method uses a combination of $\mathcal{L}_1$ (lasso) and $\mathcal{L}_2$ (ridge) penalty functions, a parsimonious specification of the penalty parameters across populations, and an ensemble step to combine PRS generated across different penalty parameters. We evaluate the performance of PROSPER and other existing methods on large-scale simulated and real datasets, including those from 23andMe Inc., the Global Lipids Genetics Consortium, and All of Us. Results show that PROSPER can substantially improve multi-ancestry polygenic prediction compared to alternative methods across a wide variety of genetic architectures. In real data analyses, for example, PROSPER increased out-of-sample prediction R^2 for continuous traits by an average of 70% compared to a state-of-the-art Bayesian method (PRS-CSx) in the African ancestry population. Further, PROSPER is computationally highly scalable for the analysis of large SNP contents and many diverse populations.

Introduction

Tens of thousands of single nucleotide polymorphisms (SNP) have been mapped to human complex traits and diseases through genome-wide association studies (GWAS) [1, 2]. Though each SNP only explains a small fraction of variation of the underlying phenotype, polygenic risk scores (PRS), which aggregate the genetic effects of many loci, can have a substantial ability to predict traits and stratify populations by underlying disease risks [3-12]. However, as existing GWAS to date have been primarily conducted in European ancestry populations (EUR) [13-16], recent studies have consistently shown that the transferability of EUR-derived PRS to non-EUR populations often is suboptimal and in particular poor for African Ancestry populations [17-22].

Despite growing efforts of conducting genetic research on minority populations [23-26], the gap in sample sizes between EUR and non-EUR populations is likely to persist in the foreseeable future. As the performance of PRS largely depends on the sample size of training GWAS [3, 27], using single ancestry methods [28-32] to generate PRS for a minority population, using data from that population alone may not achieve ideal results. To address this issue, researchers have developed methods for generating powerful PRS by borrowing information across diverse ancestry populations [33]. For example, Weighted PRS [34] combines single-ancestry PRS generated from each population using weights that optimize performance for a target population. Bayesian methods have also been proposed that generate improved PRS for each population by jointly modeling the effect-size distribution across populations [35, 36]. Recently, our group proposed a new method named CT-SLEB [22], which extends the clumping and thresholding (CT)

65 [37] method to multi-ancestry settings. The method uses an empirical-Bayes (EB) approach to
66 estimate effect sizes by borrowing information across populations and a super learning model
67 to combine PRSs under different tuning parameters. However, the optimality of the methods
68 depends on many factors, including the ability to account for heterogeneous linkage
69 disequilibrium (LD) structure across populations and the adequacy of the models for underlying
70 effect-size distribution [3, 27]. In general, our extensive simulation studies and data analyses
71 suggest that no method is uniformly the most powerful, and exploration of complementary
72 methods will often be needed to derive the optimal PRS in any given setting [22].
73
74 In this article, we propose a novel method for generating multi-ancestry Polygenic Risk scOres
75 based on an enSemble PEnalized Regression (PROSPER) using GWAS summary statistics and
76 validation datasets across diverse populations. The method incorporates $\mathcal{L}_1$ penalty functions
77 for regularizing SNP effect sizes within each population, an $\mathcal{L}_2$ penalty function for borrowing
78 information across populations, and a flexible but parsimonious specification of the underlying
79 penalty parameters to reduce computational time. Further, instead of selecting a single optimal
80 set of tuning parameters, the method combines PRS generated across different populations and
81 tuning parameters using a final ensemble regression step. We compare the predictive
82 performance of PROSPER with a wide variety of single- and multi-ancestry methods using
83 simulation datasets from our recent study[22] across five populations (EUR, African (AFR), Ad
84 Mixed American (AMR), East Asian (EAS), and South Asian (SAS))[22]. Furthermore, we evaluate
85 these methods using a variety of real datasets from 23andMe Inc. (23andMe), the Global Lipids
86 Genetics Consortium (GLGC) [38], All of Us (AoU) [39], and the UK Biobank study (UKBB) [40]. Results

from these analyses indicate that PROSPER is a highly promising method for generating the most powerful multi-ancestry PRS across diverse types of complex traits. Computationally, PROSPER is also exceptionally scalable compared to other advanced methods.

Results

Method overview

PRSOSPER is a method designed to improve prediction performance for PRS across distinct ancestral populations by borrowing information across ancestries (Figure 1). It can integrate large EUR GWAS with smaller GWAS from non-EUR populations. Ideally, individual-level tuning data are needed for all populations, because the method needs optimal parameters from single-ancestry analysis as an input; however, even when data is only available for a target population, PRSOSPER can still be performed, and the PRS will be optimized and validated towards the target population. The method can account for population-specific genetic variants, allele frequencies, and LD patterns and use computational techniques for penalized regressions for fast implementation.

PROSPER

Assuming a continuous trait, we first consider a standard linear regression model for underlying individual-level data for describing the relationship between trait values and genome-wide

genetic variants across M distinct populations. Let $\mathbf{Y}_i$ denote the $n_i \times 1$ vector of trait values, $\mathbf{X}_i$ denote the $n_i \times p_i$ genotype matrix, $\boldsymbol{\beta}_i$ denote the $p_i \times 1$ vector of SNP effects, and $\boldsymbol{\epsilon}_i$ denote the $n_i \times 1$ vector of random errors for the ith population. We assume underlying linear regression models of the form $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, i = 1, \dots M$; and intend to solve the linear regression system by least square with a combination of $\mathcal{L}_1$ (lasso) [41] and $\mathcal{L}_2$ (ridge) [42] penalties in the form

$$\sum_{1\le i\le M} \frac{1}{n_i}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_i)^T(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_i) + \sum_{1\le i\le M} 2\lambda_i\|\boldsymbol{\beta}_i\|_1^1 + \sum_{1\le i_1<i_2\le M} c_{i_1 i_2} \left\| \boldsymbol{\beta}_{i_1}^{s_{i_1 i_2}} - \boldsymbol{\beta}_{i_2}^{s_{i_1 i_2}} \right\|_2^2$$

where $\lambda_i, i = 1, \dots, M$ are the population-specific tuning parameters associated with the lasso penalty; $\boldsymbol{\beta}_{i_1}^{s_{i_1 i_2}}$ and $\boldsymbol{\beta}_{i_2}^{s_{i_1 i_2}}$ denote the vectors of effect-sizes for SNPs for the i_1-th and i_2-th populations, respectively, restricted to the set of shared SNPs ($s_{i_1 i_2}$) across the pair of the populations; and $c_{i_1 i_2}, 1 \le i_1 < i_2 \le M$ are the tuning parameters associated with the ridge penalty imposing effect-size similarity across pairs of populations.

In the above, the first part, $\sum_{1\le i\le M} 2\lambda_i\|\boldsymbol{\beta}_i\|_1^1$, uses a lasso penalty. Lasso can produce sparse solution [41] and recent PRS studies that have implemented the lasso penalty in the single-ancestry setting have shown its promising performance [29, 30]. The second part, $\sum_{1\le i_1<i_2\le M} c_{i_1 i_2} \left\| \boldsymbol{\beta}_{i_1}^{s_{i_1 i_2}} - \boldsymbol{\beta}_{i_2}^{s_{i_1 i_2}} \right\|_2^2$, uses a ridge penalty. As it has been widely shown that the causal effect sizes of SNPs tend to be correlated across populations [43, 44], we propose to use the ridge penalty to induce genetic similarity across populations. Compared to the fused lasso [45], which uses lasso penalty for the differences, we use ridge penalty instead, which allows a small

difference in SNP effects across populations rather than truncating them to zero. The solutions for population-specific effect size using the combined lasso and ridge penalties can be sparse.

The estimate of $\boldsymbol{\beta}_i, i = 1, \dots, M$ in the above individual-level linear regression systems can be obtained by minimizing the above least square objective function. Following the derivation of lassosum [29], a single-ancestry method for fitting the lasso model to GWAS summary statistics data, we show that the objective function for individual-level data can be approximated using GWAS summary statistics and LD reference matrices by substituting $\frac{1}{n_i}\mathbf{X}_i^T\mathbf{X}_i$ by $\mathbf{R}_i$, where $\mathbf{R}_i$ is the estimated LD matrix based on a reference sample from the i-th population , and $\frac{1}{n_i}\mathbf{X}_i^T\mathbf{y}_i$, by $\mathbf{r}_i$, where $\mathbf{r}_i$ is the GWAS summary statistics in the i-th population. Therefore, the objective function of the summary-level model can be written as

$$\sum_{1\le i\le M}(\boldsymbol{\beta}_i^T(\mathbf{R}_i+\delta_i\mathbf{I})\boldsymbol{\beta}_i - \mathbf{2}\boldsymbol{\beta}_i^T\mathbf{r}_i + 2\lambda_i\|\boldsymbol{\beta}_i\|_1^1) + \sum_{1\le i_1<i_2\le M} c_{i_1 i_2}\left\|\boldsymbol{\beta}_{i_1}^{s_{i_1 i_2}} - \boldsymbol{\beta}_{i_2}^{s_{i_1 i_2}}\right\|_2^2$$

where the additional tuning parameters $\delta_i, i = 1, \dots, M$, are introduced for regularization of the LD matrices across the different populations [30]. For a fixed set of tuning parameters, the above objective function can be solved using fast coordinate descent algorithms [46] by iteratively updating each element of $\beta_i, i = 1, \dots, M$ (see the section of "Obtain PROSPER solution" under "Methods").

Reducing tuning parameters

149 For the selection of tuning parameters, we assume we have access to individual-level data
150 across the different populations which are independent of underlying GWAS from which
151 summary statistics are generated. The above setting involves three sets of tuning parameters,
152 $\{\delta_i\}_{i=1}^M$, $\{\lambda_i\}_{i=1}^M$, and $\{c_{i_1 i_2}\}_{1 \le i_1 < i_2 \le M}$, totaling to the number of $M + M + \frac{M(M-1)}{2}$. As grid search
153 across many combinations of tuning parameter values can be computationally intensive, we
154 propose to reduce the search range by a series of steps. First, we use lassosum2 [30] to analyze
155 GWAS summary statistics and tuning data from each ancestry population by itself and obtain
156 underlying values of optimal tuning parameters, $(\delta_i^0, \lambda_i^0)$ for $i = 1, \dots, M$; if tuning data is only
157 available for the target population, the $(\delta_i^0, \lambda_i^0)$ for other populations can be optimized towards
158 the target population. For fitting PROSPER, we fix $\delta_i = \delta_i^0$ for $i = 1, \dots, M$ as these are
159 essentially used to regularize estimates of population-specific LD matrices. We note that the
160 optimal $\{\lambda_i\}_{i=1}^M$ depend on sample sizes of underlying training GWAS (Supplementary Figure 1),
161 and thus should not be arbitrarily assumed to be equal across all populations. Considering that
162 the optimal tuning parameters associated with the $\mathcal{L}_1$ penalty function from the single-ancestry
163 analyses should reflect the characteristics of GWAS data, which includes underlying sparsity of
164 effect sizes and sample sizes, we propose to specify the $\mathcal{L}_1$-tuning parameters in PROSPER as
165 $\lambda_i = \lambda \lambda_i^0$, i.e. they are determined by the corresponding tuning parameters from the ancestry-
166 specific analysis except for the constant multiplicative factor λ . Finally, for computational
167 feasibility, we further assume that effect sizes across all pairs of populations have a similar
168 degree of homogeneity and thus set all $\{c_{i_1 i_2}\}_{1 \le i_1 < i_2 \le M}$ to be equal to c. We will later discuss
169 this assumption and perform a sensitivity analysis (see Discussion). By using the above
170 assumptions, the objective function to minimize with respect to $\boldsymbol{\beta}_i, i = 1, \dots, M$, becomes

$$\sum_{1\le i\le M}(\boldsymbol{\beta}_i^T(\mathbf{R}_i+\delta_i^0\mathbf{I})\boldsymbol{\beta}_i-\mathbf{2}\boldsymbol{\beta}_i^T\mathbf{r}_i+2\lambda\lambda_i^0\|\boldsymbol{\beta}_i\|_1^1)+\sum_{1\le i_1<i_2\le M}c\left\|\boldsymbol{\beta}_{i_1}^{s_{i_1i_2}}-\boldsymbol{\beta}_{i_2}^{s_{i_1i_2}}\right\|_2^2$$

where λ and c are the only two tuning parameters needed for lasso penalty and genetic similarity penalty, respectively.

Ensemble

Using an ensemble method to combine PRS has been shown to be promising in CT-type methods as opposed to picking an optimal threshold [22, 37]. In general, a specific form of the penalty function, or equivalently a model for prior distribution in the Bayesian framework, may not be able to adequately capture the complex nature of the underlying distribution of the SNPs across diverse populations. We conjecture that when effect size distribution is likely to be mis-specified, an ensemble method, which combines PRS across different values of tuning parameters instead of choosing one optimal set, is likely to improve prediction. Therefore, as a last step, we obtain the final PROSPER model using an ensemble method, super learning [47-49], implemented in the *SuperLearner* R package, to combine PRS generated from various tuning parameter settings and optimized using tuning data from the target population. The super learner we use here was based on three supervised learning algorithms, including lasso [41], ridge [42], and linear regression (see Methods).

Results

Methods comparison on simulated data

193

194 We conducted simulation analyses on continuous traits under various genetic architectures [22]
195 to evaluate the performance of different methods that can be categorized into five groups:
196 single-ancestry methods trained from target GWAS data (single-ancestry method), single-
197 ancestry methods trained from EUR GWAS data (EUR PRS based method), simple multi-ancestry
198 methods by weighting single-ancestry PRS (weighted PRS), recently published multi-ancestry
199 methods (existing multi-ancestry methods), and our proposed method, PROSPER. Single-
200 ancestry methods include CT [37], LDpred2 [31], and lassosum2 [30]. Existing multi-ancestry methods
201 include PRS-CSx [35] and CT-SLEB [22]. All of the methods were implemented using the latest
202 available version of the underlying software. The performance of the methods is evaluated by
203 R^2 measured on validation samples independent of training and tuning datasets. Analyses in
204 this and the following sections are restricted to a total of 2,586,434 SNPs, which are included in
205 either HapMap 3 (HM3) [50] or the Multi-Ethnic Genotyping Arrays (MEGA) chips array [51]. LD
206 reference samples for all five ancestries, EUR, AFR, AMR, EAS, and SAS, in this and the following
207 sections, are from 1000 Genomes Project (Phase 3) [52] (1000G).

208

209 The results (Figure 2, Supplementary Figure 2-5, Supplementary Data 1-5) show that multi-
210 ancestry methods generally exhibit superior performance compared to single-ancestry
211 methods. Weighted PRS generated from methods modeling LD (LDpred2 and lassosum2) can
212 lead to a noticeable improvement in performance (green bars in Figure 2). Notably, PROSPER
213 shows robust performance uniformly across different scenarios. When the sample size of the
214 target non-EUR population is small ($N_{target} = 15\mathrm{K}$) (Figure 2a), PROSPER has comparable good

215 performance with other multi-ancestry methods, such as weighted LDpred2 and PRS-CSx, under
216 a high degree of polygenicity ($p_{causal} = 0.01$). However, under the same sample size setting
217 and lower polygenicity ($p_{causal} = 0.001$ and 5×10^{-4}), PRS-CSx and CT-SLEB outperform
218 PROSPER, with the margin of improvement increasing as the strength of negative selection
219 decreases (strong negative selection in Figure 2a, mild negative selection in Supplementary
220 Figure 2a, and no negative selection in Supplementary Figure 3a). When the sample size of the
221 target population is large ($N_{target} = 80\mathrm{K}$) (Figure 2b, and Supplementary Figure 2-5 b),
222 PROSPER almost uniformly outperforms all other methods, particularly for the AFR population,
223 and weighted LDpred2 remains a close competitor.
224
225 We further compare the computational efficiency of PROSPER in comparison to PRS-CSx, the
226 state-of-the-art Bayesian method available for generating multi-ancestry PRS. We train PRS
227 models for the two methods using simulated data for chromosome 22 using a single core with
228 AMD EPYC 7702 64-Core Processors running at 2.0 GHz. We observe (Supplementary Data 6)
229 that PROSPER is 37 times faster than PRS-CSx (3.0 vs. 111.1 minutes) in a two-ancestry analysis
230 including AFR and EUR; and 88 times faster (6.8 vs. 595.8 minutes) in the analysis of all five
231 ancestries. The memory usage for PRS-CSx is about 2.8 times smaller than PROSPER (0.78 vs.
232 2.24 Gb in two-ancestry analysis, and 0.84 vs. 2.35 Gb in five-ancestry analysis).
233
234 *23andMe data analysis*
235

236 We applied various methods to GWAS summary statistics available from the 23andMe, Inc. to
237 predict two continuous traits, heart metabolic disease burden and height; as well as five binary
238 traits, any cardiovascular disease (any CVD), depression, migraine diagnosis, morning person,
239 and sing back musical note (SBMN). The datasets are available for all five ancestries, African
240 American (AA), Latino, EAS, EUR, and SAS. The methods are tuned and validated on a set of
241 independent individuals of the corresponding ancestry from the 23andMe participant cohort
242 (see the section of "Real data analysis" under "Methods" for data description, and
243 Supplementary Data 7-8 for sample sizes used in training, tuning and validation). In an earlier
244 version of the analysis, we had analyzed the data using an older version LDpred2 in its package
245 of bigsnpr (version 1.8) that was available when the project was initiated. Quality control
246 analysis following comments from one of the reviewers indicated problem with convergence of
247 those results. As we were not able to further update the analysis using the most recent version
248 of the LDpred2 in its package of bigsnpr (version 1.12) due to time constraint of the 23andMe
249 team , we did not report results from LDpred2 and its corresponding EUR and weighted
250 methods in the 23andMe data analysis.
251
252 From the analysis of two continuous traits (Figure 3 and Supplementary Data 9), we observe
253 that lassosum2 and its related methods (EUR lassosum2 and weighted lassosum2) generally
254 perform better than CT and its related methods. On the basis of the advantage of lassosum2,
255 PROSPER further improves the performance, and for most of the settings, outperforms all
256 alternative methods, including PRS-CSx and CT-SLEB. PROSPER demonstrates particularly
257 remarkable improvement for both traits in AA and Latino (26.9 % relative improvement in R^2

258 over the second-best method on average, yellow cells in Supplementary Data 10) (first two
259 panels in Figure 3a-b). For EAS and SAS, PROSPER is slightly better than other methods, except
260 for heart metabolic disease burden of SAS (the last panel in Figure 3a), which has the smallest
261 sample size (~20K).
262
263 The results from the analysis of the binary traits (Figure 4 and Supplementary Data 9) show that
264 PROSPER generally exhibits better performance (7.8% and 12.3% relative improvement in logit-
265 scale variance (see Methods) over CT-SLEB and PRS-CSx, respectively, averaged across
266 populations and traits) (blue and red cells, respectively, in Supplementary Data 10). A similar
267 trend is observed for the analyses of AA and Latino, where PROSPER usually has the best
268 performance (first two panels in Figure 4a-e). In general, no single method can uniformly
269 outperform others. Weighted lassosum2 has outstanding performance for depression (Figure
270 4b), while PROSPER is superior for morning person (Figure 4d). PRS-CSx shows a slight
271 improvement in the analysis of migraine diagnosis for EAS populations (last second panel in
272 Figure 4c), and CT-SLEB performs the best in the analysis of any CVD for SAS population (last
273 panel in Figure 4a).
274
275 *GLGC and AoU data analysis*
276
277 Considering the uncommonly huge sample sizes from 23andMe, we further applied alternative
278 methods for the analysis of two other real datasets, GLGC and AoU. The GWAS summary
279 statistics from GLGC for four blood lipid traits, high-density lipoprotein (HDL), low-density

280 lipoprotein (LDL), log-transformed triglycerides (logTG), and total cholesterol (TC), are publicly
281 downloadable and available for all five ancestries, African/Admixed African, Hispanic, EAS, EUR,
282 and SAS (see Methods for data description, and Supplementary Data 7 for sample sizes).
283 Further, we generated GWAS summary statistics data from the AoU study for two
284 anthropometric traits, body mass index (BMI) and height, for individuals from three ancestries,
285 AFR, EUR, and Latino/Admixed American (see Methods for data description, and
286 Supplementary Data 7 for sample sizes). Both the blood lipid traits and anthropometric traits
287 have corresponding phenotype data available in the UKBB, which we use to perform tuning and
288 validation (see the section of "Real data analysis" under "Methods" for the ancestry
289 composition, and Supplementary Data 8 for sample sizes). Given the limited sample sizes of
290 genetically inferred AMR ancestry individuals in UKBB, we do not report the performance of
291 PRS on AMR individuals in UKBB. In these analyses, we implemented LDpred2 method using the
292 latest version of the software (version 1.12).
293
294 Results from analysis of four blood lipid traits (Figure 5 and Supplementary Data 11) from GLGC
295 and UKBB show that weighted PRS methods substantially outperform alternative methods. In
296 particular, we observe that the weighted lassosum2 outperforms the other two weighted
297 methods. Furthermore, our proposed method, PROSPER, shows improvement over weighted
298 lassosum2 in both AFR and SAS (13.5% and 12.3% relative improvement in R^2, respectively,
299 averaged across traits) (green and orange cells, respectively, in Supplementary Data 12), but
300 not in EAS. Notably, PROSPER outperforms PRS-CSx and CT-SLEB in most scenarios (34.2% and
301 37.7% relative improvement in R^2, respectively, averaged across traits and ancestries) (blue and

302 red cells, respectively, in Supplementary Data 12), with the improvement being particularly
303 remarkable for the AFR population (Figure 5) in which PRS development tends to be the most
304 challenging.
305
306 The results from AoU and UKBB (Figure 6 and Supplementary Data 13) show that PROSPER
307 generates the most predictive PRS for the two analyzed anthropometric traits for the AFR
308 population. It appears that Bayesian and penalized regression methods that explicitly model LD
309 tend to outperform corresponding CT-type methods (CT, EUR CT, and weighted CT) which
310 excluded correlated SNPs. Among weighted methods, both LDpred2 and lassosum2 show major
311 improvement over the corresponding CT method. Further, for both traits, PROSPER shows
312 remarkable improvement over the best of the weighted methods and the two other advanced
313 methods, PRS-CSx and CT-SLEB (91.3% and 76.5% relative improvement in R^2, respectively,
314 averaged across the two traits) (blue and red cells, respectively, in Supplementary Data 14).
315
316 *Gain from PROSPER over lassosum2*
317
318 To investigate whether the additional gain from PROSPER arises from modeling shared effects
319 across populations or from combining PRS with super learning, we further employ a super
320 learning step for lassosum2 (termed as advanced weighted lassosum2) as a point of
321 comparison. The results in simulations (Supplementary Figure 6-10 and Supplementary Data 15)
322 indicate that PROSPER consistently has more advantage than the advanced weighted lassosum2
323 in all scenarios. The results in real data (Supplementary Figure 11-12 and Supplementary Data

16) show that the performance of the two methods depends on traits and ancestries. PROSPER has comparable performance with advanced weighted lassosum2 in AFR; while PROSPER outperforms advanced weighted lassosum2 almost in all scenarios in SAS and EAS. In summary, PROSPER has 41.1% relative improvement in R^2 over advanced weighted lassosum2 on average across all ancestries and all traits in GLGC and AoU. We were not able to perform this analysis in 23andMe due to time constraint of the 23andMe team.

Discussion

In this article, we propose PROSPER as a powerful method that can jointly model GWAS summary statistics from multiple ancestries by an ensemble of penalized regression models to improve the performance of PRS across diverse populations. We show that PROSPER is a uniquely promising method for generating powerful PRS in multi-ancestry settings through extensive simulation studies, analysis of real datasets across a diverse type of complex traits, and considering the most recent developments of alternative methods. Computationally, the method is an order of magnitude faster compared to PRS-CSx [35], an advanced Bayesian method, and comparable to CT-SLEB [22], which derives the underlying PRS in closed forms. We have packaged the algorithm into a command line tool based on the R programming language (https://github.com/Jingning-Zhang/PROSPER).

345 We compare PROSPER with a number of alternative simple and advanced methods using both
346 simulated and real datasets. The simulation results show that PROSPER generally outperforms
347 other existing multi-ancestry methods when the target sample size is large (Figure 2b).
348 However, when the sample size of the target population is small (Figure 2a), no method
349 performed uniformly the best. In this setting, when the degree of polygenicity is the lowest
350 ($p_{causal} = 5 \times 10^{-4}$), CT-SLEB outperforms other methods by a noticeable margin, and
351 PROSPER performs slightly worse than PRS-CSx. Simulations also show that in the scenario of a
352 highly polygenic trait ($p_{causal} = 0.01$), irrespective of sample size, both weighted lassosum2
353 and PROSPER tend to exhibit superiority compared to all other methods. In terms of
354 computational time, PROSPER is an order of magnitude faster than PRS-CSx in a five-ancestry
355 analysis. The memory usage for PRS-CSx is smaller than PROSPER, but both are acceptable
356 (Supplementary Data 6).
357
358 We observe that for the analysis of both continuous and binary traits using 23andMe Inc. data,
359 PROSPER demonstrates a substantial advantage over all other methods for the AA and Latino
360 populations, which have the largest sample sizes among all minority groups. The result is
361 consistent with the superior performance of PROSPER observed in simulation settings when the
362 sample size of the target population is large. However, it is worth noting that even for the two
363 other populations, EAS and SAS, which have much smaller sample sizes, PROSPER still performs
364 the best in half of the settings (the last two panels in Figure 3a-b and Figure 4a-e). For the
365 prediction of blood lipid traits, PROSPER and weighted PRS methods perform noticeable better
366 than other alternative methods. For the analysis of two anthropometric traits using training

367 data from AoU, we observe that methods that explicitly model and account for LD differences
368 (e.g. lassosum2, LDpred2, and their corresponding weighted methods) generally achieve higher
369 predictive accuracy than CT-based methods which discard correlated SNPs. The result is
370 consistent with what we have observed in simulation settings under extreme polygenic
371 architectures as expected for complex traits like height and BMI. In addition, we observe
372 significant improvement in PRS performance using PROSPER over advanced weighted
373 lassosum2 method which is allowed to incorporate a super learning step in lassosum2. This
374 suggests that the additional gain of PROSPER arises from modeling shared effects across
375 populations through the $\mathcal{L}_2$-penalty function.
376
377 PROSPER, while showing promising results in our simulations and real data analyses, does have
378 several limitations. First, when the sample size for the training sample for a target population is
379 small, particularly for traits with low polygenicity, the method may not perform as well as some
380 of the other existing methods (Figure 2a). In this specific scenario where the number of true
381 causal variants is small, a potential reason for suboptimal performance of PROSPER is the bias
382 induced by lasso. This inspires future work of extending PROSPER to adaptive lasso [53] for
383 unbiased estimation and other forms of penalty functions for sparser solutions. Second, the use
384 of a super learning step in PROSPER can lead to poorer performance compared to weighted
385 lassosum2 when the sample size for the tuning dataset is not adequately large. In the analysis
386 of lipid traits for EAS, for example, we observe lower predictive accuracy of PROSPER than
387 weighted lassosum2 (the middle panel in Figure 5b and d). This can be attributed to overfitting
388 in the tuning sample, as the number of tuning samples of EAS origin in the UKBB is only ~1000,

389 while the number of PRSs combined in the super learning step is close to 500. In this scenario,
390 we suggest comparing the performance of the ensemble PRS with that without the ensemble
391 step, as the latter one will be more resilient to overfitting. We conducted simulation analyses to
392 further explore the ideal sample size for tuning (Supplementary Figure 13). Generally, a tuning
393 sample size within the range of 1000-3000 is adequate for continuous traits. Third, we used a
394 constant tuning parameter for the genetic similarity penalty, disregarding varying genetic
395 distances among populations [54]. However, introducing additional tuning parameters could
396 result in both computational challenges and numerical instability. We have investigated this by
397 analyzing GLGC data (see Supplementary Data 17, and Methods), adding an extra tuning
398 parameter to accommodate adaptable distances between the AFR population and others.
399 Results indicate a disproportionate increase in computational load (the last column in
400 Supplementary Data 17) relative to the marginal enhancement in predictive accuracy, and a
401 potential of instability and overfitting (gray cells in Supplementary Data 17). Lastly, the
402 framework is modeled on a standardized genotype scale characterized by strong negative
403 selection; however, there could be diverse genetic architectures in reality. To address this
404 limitation, models could be extended to varying degrees of negative selection by multiplied by
405 exponentiations of allele frequencies, as discussed in Zhang et al. (Nature Genetics, 2023) [22].
406
407 PROSPER and a number of other recent methods have been developed for modeling summary
408 statistics data across discrete populations typically defined by self-reported ancestry
409 information. Increasing sample size for reference sample sizes for various populations well-
410 matched with those providing training datasets can further enhance performance of PROSPER

411 and other methods that explicitly incorporates LD information into modeling. Further, there is
412 an emerging need to consider the underlying continuum of genetic diversity across populations
413 in both the development and implementational of PRS in diverse populations in the future [55].
414 Towards this goal, a recent method called GAUDI [56] has been proposed based on the fused
415 lasso penalty for developing PRS in admixed population using individual-level data. While
416 GAUDI shares similarities with PROSPER in terms of the use of the lasso-penalty function, the
417 two methods are distinct in terms of the specification of tuning parameters and use of the
418 ensemble step. Our model specification of PROSPER makes it easily amendable to handle
419 continuous genetic ancestry data, but further research is needed for scalable implementation
420 of the method with individual-level data and extensive empirical evaluations.
421
422 To conclude, we have proposed PROSPER, a statistically powerful and computationally scalable
423 method for generating multi-ancestry PRS using GWAS summary statistics and additional tuning
424 and validation datasets across diverse populations. While no method is uniformly powerful in
425 all settings, we show that PROSPER is the most robust among a large variety of recent methods
426 proposed across a wide variety of settings. As individual-level data from GWAS of diverse
427 populations becomes increasingly available, PROSPER and other methods will require additional
428 considerations for incorporating continuous genetic ancestry information, both global and local,
429 into the underlying modeling framework.
430

Methods

We confirm that our research complies with all relevant ethical regulations. All individuals from 23andMe included have provided informed consent and answered surveys online according to our human subject protocol reviewed and approved by Ethical & Independent Review Services, a private institutional review board (http://www.eandireview.com). All participants from UK Biobank provided written informed consent (more information is available at https://www.ukbiobank.ac.uk/2018/02/gdpr/). The information of individuals from All of US included in our analyses has been collected according to All of Us Research Program Operational Protocol (https://allofus.nih.gov/sites/default/files/aou_operational_protocol_v1.7_mar_2018.pdf). Detailed consent process of All of Us is described on https://allofus.nih.gov/about/protocol/all-us-consent-process.

Data preparation and formatting in PROSPER. We match SNPs and their alleles in GWAS summary statistics and genotypes of individuals for tuning and validation purposes to that in 1000G reference data (phase 3) [52]. To simplify computing huge-dimensional LD matrix, we use existing LD block information from EUR [29] to divide the whole genome, and assume the blocks to be independent. We use PLINK1.9 [57] with flag --r bin4 to compute the LD matrix within each block in each ancestry for common SNPs (MAF>0.01) either in HM3 [50] or the MEGA [51]. For SNPs not common in all populations, we only model them in the populations where they are common; if a SNP is population-specific that is only common in one population, we model it

only using the lasso penalty without the genetic similarity penalty. The parameter path of the tuning parameter λ for the scale factor in lasso penalty is set to a sequence evenly spaced on a logarithmic scale from $\lambda^{\max} = \min_{1 \le i \le m} \left(\frac{\max_{1 \le k \le p} (|r_{ik}|)}{\lambda_i^0} \right)$ to $\lambda^{\min} = 0.001 \times \lambda^{\max}$ which is set to guarantee non-zero solutions, where r_{ik} is the GWAS summary statistics for the k-th SNP in the i-th population, and λ_i^0 is the underlying values of optimal tuning parameter λ for the i-th population. The parameter path for the tuning parameter c for the genetic similarity penalty is set to a sequence evenly spaced on a quad-root scale from $c^{\min} = 2$ to $c^{\max} = 100$, i.e. seq($c^{\min}$^(1/4), $c^{\max}$^(1/4), length.out = 10)^4 using R command. For all analyses excluding 23andMe, the length of sequences of both parameters are set to be 10, while for the analysis of 23andMe, it is set to be 5 to reduce the computation workload caused by the confidential requirements of the 23andMe dataset.

Obtain PROSPER solution. For M populations, the objective function to minimize for p_i-dimentional vector of SNP effect, $\boldsymbol{\beta}_i, i = 1, \dots, M$, is

$$\mathbf{L}(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m) = \sum_{1 \le i \le M} (\boldsymbol{\beta}_i^T (\mathbf{R}_i + \delta_i \mathbf{I}) \boldsymbol{\beta}_i - \mathbf{2}\boldsymbol{\beta}_i^T \mathbf{r}_i + 2\lambda_i \|\boldsymbol{\beta}_i\|_1^1) + \sum_{1 \le i_1 < i_2 \le M} c_{i_1 i_2} \left\| \boldsymbol{\beta}_{i_1}^{s_{i_1 i_2}} - \boldsymbol{\beta}_{i_2}^{s_{i_1 i_2}} \right\|_2^2$$

where $\mathbf{R}_i$ is an estimate of p_i-by-p_i LD matrix based on a reference sample from the i-th population, $\mathbf{r}_i$ is the p_i-dimentional vector of GWAS summary statistics in the i-th population, $\boldsymbol{\beta}_{i_1}^{s_{i_1 i_2}}$ and $\boldsymbol{\beta}_{i_2}^{s_{i_1 i_2}}$ denote the effect vectors for the SNPs shared across i_1-th and i_2-th populations (the set of SNPs is denoted by $s_{i_1 i_2}$); δ_i, λ_i and $c_{i_1 i_2}$ are tuning parameters as defined in above sections.

This optimization can be solved using coordinate descent algorithms by iteratively updating each element in the vectors. We take derivative for SNP k in i-th population, $k = 1, \dots, p_i$, $i = 1, \dots, M$

$$\frac{\partial \mathbf{L}(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m)}{\partial \beta_{ik}} = 2\left(1 + \delta_i + \sum_{i' \neq i, 1 \leq i' \leq M} c_{ii'}\right)\beta_{ik} + 2\lambda_i \frac{\partial |\beta_{ik}|}{\partial \beta_{ik}}$$

$$- 2\left(r_{ik} - \sum_{k' \neq k, 1 \leq k' \leq p} R_{i,k'k}\beta_{ik'} + \sum_{1 \leq i' \leq M, \text{s.t.} k \in S_{i,i'}} c_{ii'}\beta_{i'k}\right)$$

where β_{ik} denotes the effect for SNP k in $\boldsymbol{\beta}_i$, r_{ik} denotes the summary statistics for SNP k in $\mathbf{r}_i$, and $R_{i,k'k}$ denotes LD between the SNP k and the SNP k' in $\mathbf{R}_i$.

By solving $\frac{\partial \mathbf{L}(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m)}{\partial \beta_{ik}} = 0$ after the (t)-th iteration, we can get the updating rule for the $(t + 1)$-th iteration

$$\beta_{ik}^{(t+1)} = \frac{\text{sign}(u_{ik}) \cdot \max\{0, |u_{ik}| - \lambda_i\}}{1 + \delta_i + \sum_{1 \leq i' \leq M, \text{s.t.} k \in S_{i,i'}} c_{ii'}}$$

where

$$u_{ik} = r_{ik} - \sum_{k' \neq k, 1 \leq k' \leq p} R_{i,k'k}\beta_{ik'}^{(t)} + \sum_{1 \leq i' \leq M, \text{s.t.} k \in S_{i,i'}} c_{ii'}\beta_{i'k}^{(t)}$$

Super learning. After getting PRSs for all populations under all tuning parameter settings, we further apply super learning to combine them to be trained on the tuning samples to get the final PROSPER model and tested on the validation samples. We use the function "*SuperLearner*" implemented in the R package with the same name, and include three linear prediction algorithms: lasso, ridge, and linear regression for continuous outcomes; and two prediction algorithms: lasso and linear regression for binary outcomes. We did not include ridge for binary

492 outcomes due to the unavailability of ridge for binary outcomes in the function. For the
493 included algorithms which have parameters: (1) in lasso, we use 100 values in lambda path
494 calculated in the default setting in glmnet package; (2) in ridge, we use a lambda path of
495 sequence from 1 to 20 incrementing by 0.1. We use Area under the ROC curve (AUC) as the
496 objective function for binary outcomes and thus use the flag "method = method. AUC" in the
497 function.
498
499 **Existing PRS methods.** We compare five groups of PRS methods. The first group is: single-
500 ancestry method, which contains commonly known single-ancestry methods, including CT,
501 LDpred2, and lassosum2, that are trained from the GWAS data from the target population. The
502 second group is: EUR PRS based method, which is the three above single-ancestry methods
503 trained from EUR GWAS data. The third group is: weighted PRS, which uses the weights
504 estimated from a linear regression to combine the PRSs estimated from the corresponding
505 single-ancestry method from all populations. The fourth group is: existing multi-ancestry
506 methods, which includes two recently published and well-performed multi-ancestry methods,
507 PRS-CSx and CT-SLEB. The last group is our proposed PROSPER. For all algorithms that have
508 tuning parameters or weights, the optimal ones are determined based on predictive R^2 or AUC
509 on tuning samples and finally evaluated on validation samples.
510 Below are detailed descriptions of the existing PRS methods used as comparisons in this
511 manuscript. In short, CT and CT-SLEB are methods that use less-dependent genetic variants
512 after a clumping step in models. LDpred2 and PRS-CSx are Bayesian methods that can account
513 for LD among genetic variants. Lassosum2 and our proposed PROSPER are penalized regression

514 methods capable of modeling genome-wide genetic variants and fitting the model in a speedy
515 way. As for the three multi-ancestry methods, CT-SLEB and PRS-CSx model the cross-ancestry
516 genetic correlation using a multivariate Bayesian prior, while our proposed PROSPER uses a
517 ridge penalty to impose effect-size similarity across pairs of populations.

518 ***CT*** is implemented in our analysis by using r^2-cutoff of 0.1 in the clumping step and then
519 thresholding by treating p-value-cutoff as a tuning parameter and being chosen from
520 $5 \times 10^{-8}, 1 \times 10^{-7}, 5 \times 10^{-7}, 1 \times 10^{-6}, \ldots, 5 \times 10^{-1}, 1.0$; p-value are from GWAS summary
521 statistics using Chi-squared test.

522 ***LDpred2*** is a PRS method that uses a spike-and-slab prior on GWAS summary statistics and
523 modeling LD across SNPs. We implement LDpred2 by the function *"snp_ldpred2_grid"* in the R
524 package "bigsnpr" version 1.12. The two tuning parameters in the algorithm include: the
525 proportion of causal SNPs, which is chosen from a sequence of length 21 that are evenly spaced
526 on a logarithmic scale from 10^{-5} to 1; per-SNP heritability, which is chosen from 0.3, 0.7, 1, or
527 1.4 times the total heritability estimated by LD score regression divided by the number of
528 causal SNPs. We fix the additional "sparse" option (for truncating small effects to zero) to
529 FALSE.

530 ***lassosum2*** is a PRS method that uses lasso regression on GWAS summary statistics for a single
531 ancestry. We implement lassosum2 by the function *"snp_lassosum2"* in the R package
532 "bigsnpr" version 1.12. The two tuning parameters in the algorithm include: tuning parameter
533 for the lasso penalty, which is chosen from a sequence of length 30 that are evenly spaced on a
534 logarithmic scale from $0.01 \times \max_{1 \le k \le p}(|r_k|)$ to $\max_{1 \le k \le p}(|r_k|)$; and regularization parameter for LD
535 matrix, which is chosen from c(0.001, 0.01, 0.1, 1).

536 ***EUR PRS*** are the PRSs trained from EUR GWAS using the above single-ancestry methods, CT,
537 LDpred2, and lassosum2, that are then applied to individuals of the target population. There is
538 no need to perform tuning for them because the models have been tuned in EUR tuning
539 samples. When computing scores for EUR PRS based method, we exclude SNPs that are not
540 presented in the validation samples from the target population.
541 ***Weighted PRS*** linearly combines the corresponding single-ancestry method trained from all
542 populations. The weights in the linear combination are estimated by a simple linear regression
543 in the tuning samples from the target population.
544 ***PRS-CSx*** is a Bayesian multi-ancestry PRS method that jointly models GWAS summary statistics
545 and LD structures across multiple populations using a continuous shrinkage prior. It has a
546 further step to linearly combine the posterior effect-sizes estimates for EUR and the target
547 population using weights in a simple linear regression in the tuning samples from the target
548 population. We implement PRS-CSx using their python-based command line tool "PRS-CSx
549 available at https://github.com/getian107/PRScsx. The parameter phi was chosen from the
550 default candidate values, $1, 10^{-2}, 10^{-4}$ and 10^{-6}. Due to the package restriction, the models
551 are fitted with only HM3 SNPs.
552 ***CT-SLEB*** is a multi-ancestry PRS method that starts from clumping and thresholding, then uses
553 Empirical-Bayes (EB) method to estimate the coefficients of PRS, and finally combines PRS by a
554 super learning model. We implement CT-SLEB by codes available at
555 https://github.com/andrewhaoyu/CTSLEB. The three tuning parameters in the algorithm
556 include: r^2-cutoff and base size of the clumping window size used in the clumping step, which
557 are chosen from (0.01, 0.05, 0.1, 0.2, 0.5) and (50kb, 100kb), respectively; and p-value cutoffs

558 for EUR and the target population, which are chosen from $5 \times 10^{-8}, 5 \times 10^{-7}, 5 \times$

559 $10^{-6}, \ldots, 5 \times 10^{-1}$ and 1.0; p-value are from GWAS summary statistics using Chi-squared test.

560

561 **Simulation analysis.** The simulated data were generated by Zhang et al. (Nature Genetics,

562 2023)[22]. In brief summary, the data were simulated under five assumed genetic architecture (as

563 described in the legends of Figure 2, Supplementary Figure 2-5) and three different degrees of

564 polygenicity $p_{causal} = 0.01, 0.001$ and 5×10^{-4}. The sample sizes for GWAS training data are

565 assumed to be n=15,000 and n=80,000 for the four non-EUR target populations; and is fixed at

566 n=100,000 for the EUR population. PRS generated from all methods are tuned in n=10,000

567 samples, and then tested in n=10,000 independent samples in each target population. We

568 randomly repeated the simulation three times, and the reported the average R^2 for all

569 candidate methods.

570

571 **Computational time and memory usage**. The computational time and memory usage of

572 PROSPER and PRS-CSx are compared based on the analysis using simulated data on

573 chromosome 22. The analysis starts from inputting all required data into the algorithms, such as

574 summary statistics and LD reference data, and ends with outputting the final PRS coefficients

575 from the algorithms. PROSPER requires an input of optimal parameters in single-ancestry

576 analysis, so we also include the step of running the single-ancestry analysis, lassosum. The

577 analyses are performed using a single core with AMD EPYC 7702 64-Core Processors running at

578 2.0 GHz. The reported results are averaged over 10 replicates. The sample size for training

579 GWAS summary statistics is n=15,000 for non-EUR populations and n=100,000 for EUR

580 population. The sample size for the tuning dataset is n=10,000 for each population.

581

582 **Real data analysis.** Training GWAS summary statistics are from 23andMe, GLGC, and AoU.

583 Tuning and validation individual-level data are from 23andMe and UKBB. LD reference data are

584 from 1000G. Detailed descriptions of those datasets are listed below.

585 ***1000G Data***. We used samples in five populations, AFR, AMR, EAS, EUR, and SAS from 1000

586 Genomes Project (Phase 3) [52]. The components of the five populations are described in

587 https://useast.ensembl.org/Help/Faq?id=532.

588 ***23andMe Data***. We analyzed two continuous traits, heart metabolic disease burden and height;

589 and five binary traits, any CVD, depression, migraine diagnosis, morning person and SBMN,

590 using GWAS summary statistics obtained from 23andMe Inc.. Data on these seven traits are

591 available for all five populations: AA, EAS, EUR, Latino, and SAS. The LD reference panels used

592 for the five populations, respectively, are unrelated individuals from 1000G of AFR, EAS, EUR,

593 AMR, and SAS origins. The tuning and validation are performed on a set of independent

594 individuals of the corresponding ancestry from 23andMe participant cohort. Please see

595 Supplementary Data 7 for training sample sizes and Supplementary Data 8 for tuning and

596 validation sample sizes. The data we used was preprocessed by Zhang et al. (Nature Genetics,

597 2023)[22], accessible from

598 https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/3NBNCV. The

599 details of the data, including genotyping, quality control, imputation, removing related

600 individuals, ancestry determination, and the preprocessing of GWAS, are described in pages 54-

601 61 in its Supplementary Notes, and Manhattan plots and QQ plots were shown in its
602 Supplementary Figures 9-15. For continuous traits, we evaluate PRS performance by the
603 predictive R^2 of the PRS for residualized trait values obtained from regressing the traits on
604 covariates. For binary traits, we evaluated PRS performance by the AUC by using the roc.binary
605 function in the R package RISCA version 1.0 [58]. To compare the PRS performance for two
606 different methods, we used the relative increase of logit-scale variance. The logit-scale variance
607 of binary traits is converted from AUC by the formula $\sigma^2 = 2\phi^{-1}(AUC)$, where ϕ is the
608 cumulative distribution function of the standard normal distribution.
609 ***GLGC Data***. We analyzed four blood lipid traits, LDL, HDL, logTG and TC, using GWAS summary
610 statistics computed without UKBB samples that are publicly available from GLGC
611 (http://csg.sph.umich.edu/willer/public/glgc-lipids2021/). Detailed information about the
612 design of the study, genotyping, quality control, and GWAS is described in Graham, S. E. *et al*.
613 (2021) [38]. The data we used was preprocessed by Zhang et al. (Nature Genetics, 2023) [22] in
614 pages 61-62 in its Supplementary Notes, and Manhattan plots and QQ plots were shown in its
615 Supplementary Figures 16-19. Data on the four traits are available for all five populations:
616 admixed African or African, EAS, EUR, Hispanic, and SAS. The LD reference panels used for the
617 five populations, respectively, are unrelated individuals from 1000G of AFR, EAS, EUR, AMR,
618 and SAS origins. The tuning and validation are performed on UKBB individuals (as described
619 below) from the same reference ancestry label as the LD reference panel. Please see
620 Supplementary Data 7 for sample sizes and the number of SNPs included in the analysis.
621 ***AoU Data***. We analyzed two anthropometric traits, BMI and height, using GWAS summary
622 statistics trained from AoU. Data for the two traits are available for three ancestries: AFR,

623 Latino/Admixed American, and EUR. The data we used was preprocessed by Zhang et al.
624 (Nature Genetics, 2023) [22], accessible from
625 https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FAWEQK. The
626 details of the data are described in pages 61-62 in its Supplementary Notes, and Manhattan
627 plots and QQ plots were shown in its Supplementary Figures 20-21. The LD reference panel
628 used for the three populations, respectively, are 1000G unrelated individuals of AFR, AMR, and
629 EUR origins. The tuning and validation are performed using UKBB individuals (as described
630 below) from the same reference ancestry label as the LD reference panel. Please see
631 Supplementary Data 7 for sample sizes and the number of SNPs included in the analysis.
632 ***UKBB data.*** We used UKBB data only for tuning and validation purposes. The four blood lipid
633 traits and two anthropometric traits mentioned above have direct measurements in UKBB. The
634 ancestry label of UKBB individuals is determined by genetically predicted ancestry, which are
635 described in pages 62-63 in the Supplementary Notes of the paper from Zhang et al. (Nature
636 Genetics, 2023) [22]. Tuning and validation are based on R^2 of the PRS regressed on the residuals
637 of the phenotypes adjusted by sex, age and PC1-10. Please see Supplementary Data 8 for
638 sample sizes. We note that for PRS we tested in UKBB validation samples, we use the ancestry
639 labels in UKBB (AFR, AMR, EAS, EUR or SAS), instead of ancestry labels in the GWAS training
640 data, to report the R^2 in the figures, result, and discussion sections of this paper.
641
642 **Extra tuning parameter for varying genetic distances.** In the discussion, we investigated adding
643 an extra tuning parameter to accommodate adaptable distances between the AFR population
644 and others. Specifically, the pair-wise c_{ij} follows the formula

$$c_{ij} = \begin{cases} r \times c & \text{if } i \text{ or } j = AFR \\ c & \text{if } i \text{ and } j \neq AFR \end{cases}$$

where r and c are tuning parameters; r takes values from $0.5, 1, 1.5$; and c takes the same sequence of candidate values as described in the first paragraph of Methods.

Data Availability

The PRSs developed for traits in GLGC and AoU will be released through the PGS Catalog (https://www.pgscatalog.org) with publication ID PGP000595 and score IDs PGS004622-PGS004686 upon publication. Simulated genotype data for 600K subjects from five ancestries are available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/COXHAP. GWAS summary level statistics for five ancestries from GLGC are available at http://csg.sph.umich.edu/willer/public/glgc-lipids2021/results/ancestry_specific/. GWAS summary level statistics for three ancestries from AoU are available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FAWEQK. GWAS summary statistics for the 23andMe discovery data set could be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit https://research.23andme.com/collaborate/#dataset-access/ for more information and to apply to access the data. GRCh37 and GRCh38 reference genome data from Phase-3 1000 Genome Project (1000G) are available at https://www.internationalgenome.org/data. Access to UKBB individual level data can be requested from https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access. Supplementary Data files and Source Data files are provided with this paper.

Code Availability

All codes for data analysis, including simulation and real data analysis, are posted through GitHub at https://github.com/Jingning-Zhang/PROSPER_analysis (ref. [59]). Codes, scripts, reference data, and toy example to perform PROSPER are publicly available at https://github.com/Jingning-Zhang/PROSPER (ref. [60]). The majority of our statistical analysis was performed using R 3.6.1 and R 4.0.2, and R packages: bigsnpr_1.12.2, bigstatsr_1.5.12, doMC_1.3.8, iterators_1.0.14, inline_0.3.19, RcppArmadillo_0.12.6.4.0, Rcpp_1.0.11, MASS_7.3-60, glmnet_4.1-8, Matrix_1.6-1.1, SuperLearner_2.0-28.1, gam_1.22-2, foreach_1.5.2, nnls_1.5, caret_6.0-94, lattice_0.21-8, ggplot2_3.4.3, stringr_1.5.0, readr_2.1.4, bigreadr_0.2.5, optparse_1.7.3. We used python 3.8.2 for PRS-CSx. We used PLINK2 for computing PRS.

682

683 References

684 1. Buniello, A. *et al*. The NHGRI-EBI GWAS Catalog of published genome-wide association
685 studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005-D1012
686 (2019).

687 2. Visscher, P. M. *et al*. 10 years of GWAS discovery: biology, function, and translation. *The*
688 *American Journal of Human Genetics* **101**, 5-22 (2017).

689 3. Chatterjee, N. *et al*. Projecting the performance of risk prediction based on polygenic
690 analyses of genome-wide association studies. *Nat. Genet.* **45**, 400-405 (2013).

691 4. Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating polygenic risk
692 prediction models for stratified disease prevention. *Nature Reviews Genetics* **17**, 392
693 (2016).

694 5. Sugrue, L. P. & Desikan, R. S. What are polygenic scores and why are they important?
695 *JAMA* **321**, 1820-1821 (2019).

696 6. Aragam, K. G. & Natarajan, P. Polygenic scores to assess atherosclerotic cardiovascular
697 disease risk: clinical perspectives and basic implications. *Circ. Res.* **126**, 1159-1177 (2020).

698 7. Ma, Y. & Zhou, X. Genetic prediction of complex traits with polygenic scores: a statistical
699 review. *Trends in Genetics* **37**, 995-1011 (2021).

700 8. Lambert, S. A., Abraham, G. & Inouye, M. Towards clinical utility of polygenic risk scores.
701 *Hum. Mol. Genet.* **28**, R133-R142 (2019).

702 9. Wray, N. R. *et al*. From basic science to clinical application of polygenic risk scores: a
703 primer. *JAMA psychiatry* **78**, 101-109 (2021).

704 10. Mavaddat, N. *et al*. Polygenic risk scores for prediction of breast cancer and breast
705 cancer subtypes. *The American Journal of Human Genetics* **104**, 21-34 (2019).

706 11. Dikilitas, O. *et al*. Predictive utility of polygenic risk scores for coronary heart disease in
707 three major racial and ethnic groups. *The American Journal of Human Genetics* **106**, 707-
708 716 (2020).

709 12. Li, R., Chen, Y., Ritchie, M. D. & Moore, J. H. Electronic health records and polygenic risk
710 scores for predicting disease risk. *Nature Reviews Genetics* **21**, 493-502 (2020).

711 13. Fatumo, S. *et al*. A roadmap to increase diversity in genomic studies. *Nat. Med.* **28**, 243-
712 250 (2022).

713 14. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161-164
714 (2016).

715 15. Peterson, R. E. *et al*. Genome-wide association studies in ancestrally diverse
716 populations: opportunities, methods, pitfalls, and recommendations. *Cell* **179**, 589-603
717 (2019).

718 16. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The missing diversity in human genetic
719 studies. *Cell* **177**, 26-31 (2019).

720 17. Martin, A. R. *et al*. Clinical use of current polygenic risk scores may exacerbate health
721 disparities. *Nat. Genet.* **51**, 584-591 (2019).

722 18. Cavazos, T. B. & Witte, J. S. Inclusion of variants discovered from diverse populations
723 improves polygenic risk score transferability. *Human Genetics and Genomics Advances* **2**,
724 100017 (2021).

725 19. Tanigawa, Y. *et al*. Significant sparse polygenic risk scores across 813 traits in UK
726 Biobank. *PLoS Genetics* **18**, e1010105 (2022).

727 20. Duncan, L. *et al*. Analysis of polygenic risk score usage and performance in diverse
728 human populations. *Nature communications* **10**, 1-9 (2019).

729 21. Duncan, L. *et al*. Analysis of polygenic risk score usage and performance in diverse
730 human populations. *Nature communications* **10**, 1-9 (2019).

731 22. Zhang, H. *et al*. A new method for multiancestry polygenic prediction improves
732 performance across diverse populations. *Nat. Genet.* **55**, 1757-1768 (2023).

733 23. Wojcik, G. L. *et al*. Genetic analyses of diverse populations improves discovery for
734 complex traits. *Nature* **570**, 514-518 (2019).

735 24. Mahajan, A. *et al*. Multi-ancestry genetic study of type 2 diabetes highlights the power of
736 diverse populations for discovery and translation. *Nat. Genet.* **54**, 560-572 (2022).

737 25. Bentley, A. R. *et al*. Multi-ancestry genome-wide gene–smoking interaction study of
738 387,272 individuals identifies new loci associated with serum lipids. *Nat. Genet.* **51**, 636-
739 648 (2019).

740 26. Partanen, J. J. *et al*. Leveraging global multi-ancestry meta-analysis in the study of
741 Idiopathic Pulmonary Fibrosis genetics. *Cell Genomics* **2**, 100181 (2022).

742 27. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS genetics* **9**,
743 e1003348 (2013).

744 28. Vilhjálmsson, B. J. *et al*. Modeling linkage disequilibrium increases accuracy of polygenic
745 risk scores. *The american journal of human genetics* **97**, 576-592 (2015).

746 29. Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X. & Sham, P. C. Polygenic scores via
747 penalized regression on summary statistics. *Genet. Epidemiol.* **41**, 469-480 (2017).

748 30. Privé, F., Arbel, J., Aschard, H. & Vilhjálmsson, B. J. Identifying and correcting for
749 misspecifications in GWAS summary statistics and polygenic scores. *Human Genetics and*
750 *Genomics Advances* **3**, 100136 (2022).

751 31. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics*
752 **36**, 5424-5431 (2020).

753 32. Ge, T., Chen, C., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via Bayesian
754 regression and continuous shrinkage priors. *Nature communications* **10**, 1-10 (2019).

755 33. Kachuri, L., Chatterjee, N., Hirbo, J. *et al*. Principles and methods for transferring
756 polygenic risk scores across global populations. *Nature Reviews Genetics*, 1-18 (2023).

757 34. Márquez-Luna, C., Loh, P., South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA
758 Type 2 Diabetes Consortium & Price, A. L. Multiethnic polygenic risk scores improve risk
759 prediction in diverse populations. *Genet. Epidemiol.* **41**, 811-823 (2017).

760 35. Ruan, Y. *et al*. Improving polygenic prediction in ancestrally diverse populations
761 . *Nat. Genet.* **54**, 573-580 (2022).

762 36. Cai, M. *et al*. A unified framework for cross-population trait prediction by leveraging the
763 genetic correlation of polygenic traits. *The American Journal of Human Genetics* **108**, 632-
764 655 (2021).

765 37. Privé, F., Vilhjálmsson, B. J., Aschard, H. & Blum, M. G. Making the most of clumping and
766 thresholding for polygenic scores. *The American Journal of Human Genetics* **105**, 1213-1221
767 (2019).

768 38. Graham, S.E., Clarke, S.L., Wu, KH.H. *et al*. The power of genetic diversity in genome-
769 wide association studies of lipids. *Nature* **600**, 675-679 (2021).

770 39. All of Us Research Program Investigators. The "All of Us" research program. *N. Engl. J.*
771 *Med.* **381**, 668-676 (2019).

772 40. Allen, N. E., Sudlow, C., Peakman, T., Collins, R. & Uk biobank. UK biobank data: come
773 and get it. *Science translational medicine* **6**, 224ed4 (2014).

774 41. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal*
775 *Statistical Society: Series B (Methodological)* **58**, 267-288 (1996).

42. Hoerl, A. E. & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55-67 (1970).

43. Brown, B. C., Ye, C. J., Price, A. L., Zaitlen, N. & Asian Genetic Epidemiology Network Type 2 Diabetes Consortium. Transethnic genetic-correlation estimates from summary statistics. *The American Journal of Human Genetics* **99**, 76-88 (2016).

44. Mishra, A., Malik, R., Hachiya, T. *et al*. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature* **611**, 115-123 (2022).

45. Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. & Knight, K. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 91-108 (2005).

46. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software* **33**, 1 (2010).

47. Van der Laan, M. J., Polley, E. C. & Hubbard, A. E. Super learner. *Statistical applications in genetics and molecular biology* **6** (2007).

48. Polley, E. C. & Van Der Laan, M. J. Super learner in prediction. (2010).

49. Van der Laan, M. J. & Rose, S. in *Targeted learning: causal inference for observational and experimental data* (Springer, 2011).

50. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52 (2010).

51. Bien, S. A. *et al*. Strategies for enriching variant coverage in candidate disease loci on a multiethnic genotyping array. *PloS one* **11**, e0167758 (2016).

52. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

53. Zou, H. The adaptive lasso and its oracle properties. *Journal of the American statistical association* **101**, 1418-1429 (2006).

54. Pritchard, J. K. & Przeworski, M. Linkage disequilibrium in humans: models and data. *The American Journal of Human Genetics* **69**, 1-14 (2001).

55. Ding, Y., Hou, K., Xu, Z. *et al*. Polygenic scoring accuracy varies across the genetic ancestry continuum in all human populations. *Nature* **618**, 774–781 (2023).

56. Sun, Q., Rowland, B.T., Chen, J. *et al*. Improving polygenic risk prediction in admixed populations by explicitly modeling ancestral-specific effects via GAUDI. *Nat Commun* **15**, *1016 (2024).*

808 57. Purcell, S. *et al*. PLINK: a tool set for whole-genome association and population-based
809 linkage analyses. *The American journal of human genetics* **81**, 559-575 (2007).

810 58. Chatton, A. *et al*. G-computation, propensity score-based methods, and targeted
811 maximum likelihood estimator for causal inference with different covariates sets: a
812 comparative simulation study. *Scientific reports* **10**, 1-13 (2020).

813 59. Zhang, J. & Ma, C. An Ensemble Penalized Regression Method for Multi-ancestry
814 Polygenic Risk Prediction. Data analysis codes. *Zenodo*
815 https://doi.org/10.5281/zenodo.10882357 (2024).

816 60. Zhang, J. & Ma, C. An Ensemble Penalized Regression Method for Multi-ancestry
817 Polygenic Risk Prediction. PROSPER. *Zenodo* https://doi.org/10.5281/zenodo.10822288
818 (2024).

819

820

821

Acknowledgements

We would like to thank the research participants and employees of 23andMe, Inc. for making this work possible. Full list of members in 23andMe Research Team can be found in the Supplemental Information. We want to thank Liz Noblin, Melissa J. Francis and Emily Voeglein for helping with the research collaboration agreement with Harvard T.H. Chan School of Public Health, Johns Hopkins Bloomberg School of Public Health and 23andMe, Inc. The analysis utilized the Joint High Performance Computing Exchange at Johns Hopkins Bloomberg School of Public Health. The UK Biobank data was obtained under the UK Biobank resource application 17731. This work was funded by NIH grants: R01 HG010480-01 (J.Zhang, JJ and NC), K99 CA256513-01 (HZ), U01 HG011719 (NC) and K99 HG012223 (JJ). The All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276. In addition, the All of Us Research Program would not be possible without the partnership of its participants.

Author Contributions Statement

J.Zhang and N.C. conceived the project. J.Zhang, J.Zhan, J.J., and H.Z. carried out all data analyses with supervision from N.C.; H.Z. created all simulated data and ran GWAS on simulated training data with the supervision from N.C.; J.Zhan, J.O., Y.J. run GWAS for training data from 23andMe Inc. with the supervision from B.L.K.; R.Z. ran GWAS on AoU training data with the supervision from N.C. and H.Z.; J.Zhang and C.M. developed the PROSPER software; J.Zhang and N.C. drafted the manuscript, and H.Z., J.J. provided comments. All co-authors reviewed and approved the final version of the manuscript.

Competing Interests Statement

J.Zhan, J.O., Y.J. and B.L.K. are employed by and hold stock or stock options in 23andMe, Inc. The remaining authors declare no competing interests.

Figure Legends/Captions

Figure 1: Detailed flowchart of PROSPER.

The analysis of M populations in PROSPER involves three key steps: 1. Separate single-ancestry analysis for all populations $i = 1, \ldots, M$; 2. Joint analysis across populations using penalized regression; 3. Ensemble regression. In step 1, the training GWAS data is used to train lassosum2 models, and the tuning data is used to obtain the optimal tuning parameters in a single-ancestry analysis. In step 2, the training GWAS and the optimal tuning parameter values from step 1 are used to train the joint cross-population penalized regression model, and obtain solution $\boldsymbol{\beta}_{\lambda,c,i}$ for each λ and c. In step 3, the tuning data is used to train the super learning model for the ensemble of PRSs computed from the solutions in step 2, $\mathbf{PRS}_{\lambda,c,i} = \mathbf{X}\boldsymbol{\beta}_{\lambda,c,i}$. The final PRS is computed as $\mathbf{PRS} = \mathbf{X}\left(\sum w_{\lambda,c,i}\boldsymbol{\beta}_{\lambda,c,i}\right)$, where $w_{\lambda,c,i}$ are the weights from the super learning model. Refer to the "Method Overview" section in the main text for a full explanation of all notations in the flowchart.

Figure 2: Performance comparison of alternative methods on simulated data generated with different sample sizes and genetic architectures under strong negative selection and fixed common-SNP heritability.

Data are simulated for continuous phenotype under a strong negative selection model and three different degrees of polygenicity (top panel: $p_{causal} = 0.01$, middle panel: $p_{causal} = 0.001$, and bottom panel: $p_{causal} = 5 \times 10^{-4}$). Common SNP heritability is fixed at 0.4 across all populations, and the correlations in effect sizes for share SNPs between all pairs of

882 populations is fixed at 0.8. The sample sizes for GWAS training data are assumed to be (**a**)
883 n=15,000, and (**b**) n=80,000 for the four non-EUR target populations; and is fixed at n=100,000
884 for the EUR population. PRS generated from all methods are tuned in n=10,000 samples, and
885 then tested in n=10,000 independent samples in each target population. The PRS-CSx package
886 is restricted to SNPs from HM3, whereas other alternative methods use SNPs from either HM3
887 or MEGA. Bars in the figure show the performance of R^2 for each method in each dataset.
888 Colors are described on the right side of the figure. Source data are provided in Supplementary
889 Data 1.
890
891 **Figure 3: Performance comparison of alternative methods for prediction of two continuous**
892 **traits in 23andMe.**
893 We analyzed two continuous traits, (**a**) heart metabolic disease burden and (**b**) height. PRS are
894 trained using 23andMe data that available for five populations: African American, Latino, EAS,
895 EUR, and SAS, and then tuned in an independent set of individuals from 23andMe of the
896 corresponding ancestry. Performance is reported based on adjusted R^2 accounting for sex, age
897 and PC1-5 in a held-out validation sample of individuals from 23andMe of the corresponding
898 ancestry. The ratio of sample sizes for training, tuning and validation is roughly about 7:2:1, and
899 detailed numbers are in Supplementary Data 7-8. The PRS-CSx package is restricted to SNPs
900 from HM3, whereas other alternative methods use SNPs from either HM3 or MEGA. LDpred2
901 and its corresponding EUR and weighted methods are excluded to avoid misinterpretation, as a
902 result of our collaboration restrictions with 23andMe, Inc., preventing us from updating these
903 methods to the latest version of its package. Bars in the figure show the performance of

904 adjusted R^2 for each method in each dataset. Colors are described on the right side of the
905 figure. Source data are provided in Supplementary Data 9.
906
907 **Figure 4: Performance comparison of alternative methods for prediction of five binary traits**
908 **in 23andMe.**
909 We analyzed five binary traits, (**a**) any CVD, (**b**) depression, (**c**) migraine diagnosis, (**d**) morning
910 person and (**e**) SBMN. PRS are trained using 23andMe data that available for five populations:
911 African American, Latino, EAS, EUR, and SAS, and then tuned in an independent set of
912 individuals from 23andMe of the corresponding ancestry. Performance is reported based on
913 adjusted AUC accounting for sex, age, PC1-5 in a held-out validation sample of individuals from
914 23andMe of the corresponding ancestry. The ratio of sample sizes for training, tuning and
915 validation is roughly about 7:2:1, and detailed numbers are in Supplementary Data 7-8. The
916 PRS-CSx package is restricted to SNPs from HM3, whereas other alternative methods use SNPs
917 from either HM3 or MEGA. LDpred2 and its corresponding EUR and weighted methods are
918 excluded to avoid misinterpretation, as a result of our collaboration restrictions with 23andMe,
919 Inc., preventing us from updating these methods to the latest version of its package. Bars in the
920 figure show the performance of adjusted AUC for each method in each dataset. Colors are
921 described on the right side of the figure. Source data are provided in Supplementary Data 9.
922
923 **Figure 5: Performance comparison of alternative methods for prediction of four blood lipid**
924 **traits (GLGC-training and UKBB-tuning/validation)**.

925 We analyzed four blood lipid traits, (**a**) HDL, (**b**) LDL, (**c**) logTG and (**d**) TC. PRS are trained using
926 GLGC data that available for five populations: admixed African or African, East Asian, European,
927 Hispanic, and South, and then tuned in individuals from UKBB of the corresponding ancestry:
928 AFR, EAS, EUR, AMR, and SAS (see the section of "Real data analysis" under "Methods" for
929 ancestry composition). Performance is reported based on adjusted R^2 accounting for sex, age,
930 PC1-10 in a held-out validation sample of individuals from UKBB of the corresponding ancestry.
931 Sample sizes for training, tuning and validation data are in Supplementary Data 7-8. Results for
932 AMR are not included due to the small sample size of genetically inferred AMR ancestry
933 individuals in UKBB. The PRS-CSx package is restricted to SNPs from HM3, whereas other
934 alternative methods use SNPs from either HM3 or MEGA. Bars in the figure show the
935 performance of adjusted R^2 for each method in each dataset. Colors are described on the right
936 side of the figure. Source data are provided in Supplementary Data 11.
937
938 **Figure 6: Performance comparison of alternative methods for prediction of two**
939 **anthropometric traits (AoU-training and UKBB-tuning/validation)**.
940 We analyzed two anthropometric traits, (**a**) BMI and (**b**) height. PRS are trained using AoU data
941 that are available for three populations: African, Latino/Admixed American, and European and
942 then tuned in individuals from UKBB of the corresponding ancestry: AFR, AMR, and EUR (see
943 the section of "Real data analysis" under "Methods" for ancestry composition). Performance is
944 reported based on adjusted R^2 accounting for sex, age, PC1-10 in a held-out validation sample
945 of individuals from UKBB of the corresponding ancestry. Sample sizes for training, tuning and
946 validation data are in Supplementary Data 7-8. Results for AMR are not included due to the

947 small sample size of genetically inferred AMR ancestry individuals in UKBB. The number of SNPs
948 analyzed in AoU analyses is much smaller than other analyses because the GWAS from AoU is
949 on array data only (see Supplementary Data 7 for the number of SNPs). The PRS-CSx package is
950 restricted to SNPs from HM3, whereas other alternative methods use SNPs from either HM3 or
951 MEGA. Bars in the figure show the performance of adjusted R^2 for each method in each dataset.
952 Colors are described on the right side of the figure. Source data are provided in Supplementary
953 Data 13.
954

Figure 1

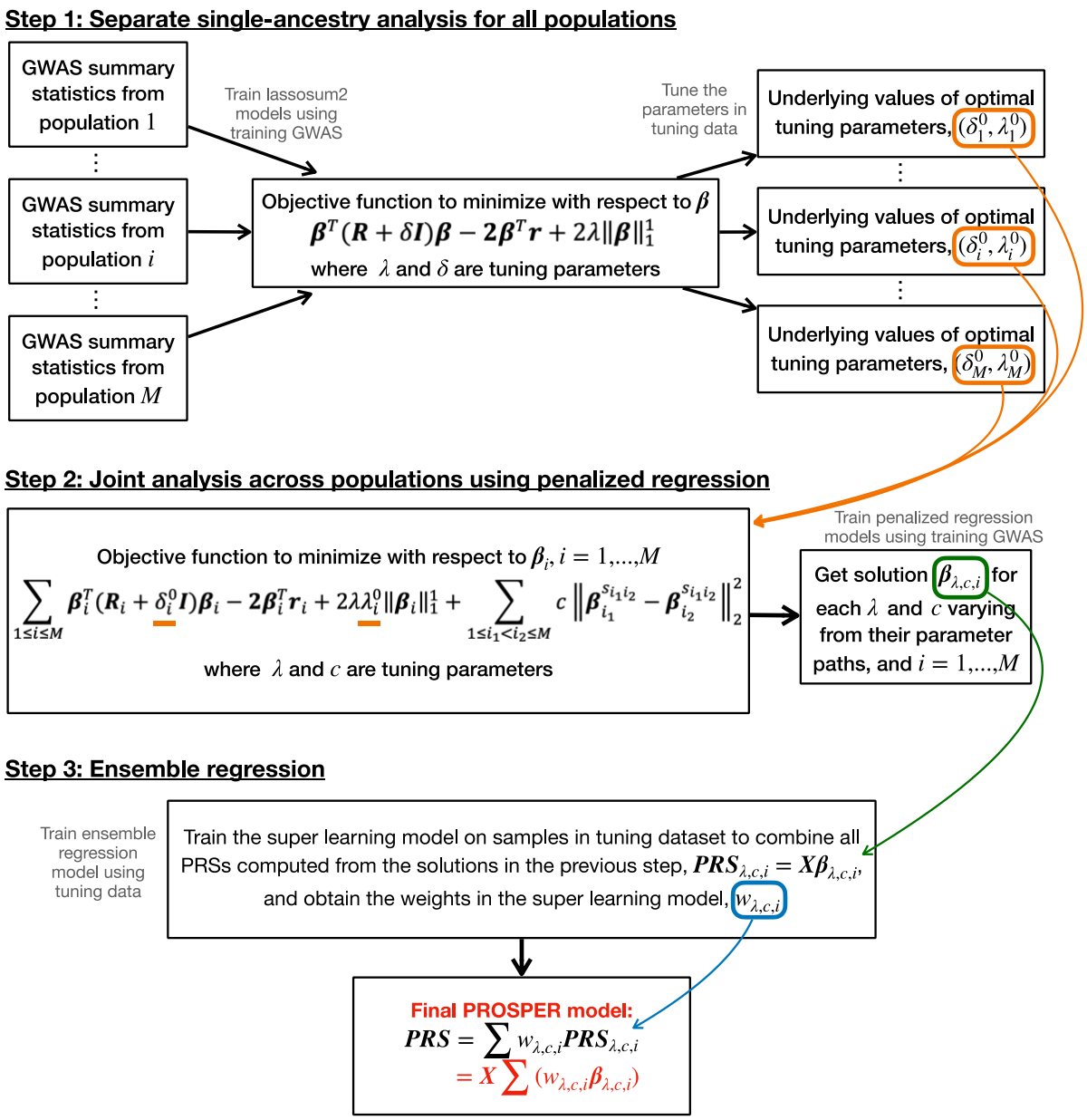

Step 1: Separate single-ancestry analysis for all populations
GWAS summary statistics from population 1
GWAS summary statistics from population i
GWAS summary statistics from population M
Train lassosum2 models using training GWAS
Objective function to minimize with respect to β
$\boldsymbol{\beta}^T(\boldsymbol{R}+\delta\boldsymbol{I})\boldsymbol{\beta}-2\boldsymbol{\beta}^T\boldsymbol{r}+2\lambda\|\boldsymbol{\beta}\|_1^1$
where λ and δ are tuning parameters
Tune the parameters in tuning data
Underlying values of optimal tuning parameters, (δ_1^0,λ_1^0)
Underlying values of optimal tuning parameters, (δ_i^0,λ_i^0)
Underlying values of optimal tuning parameters, (δ_M^0,λ_M^0)
Step 2: Joint analysis across populations using penalized regression
Objective function to minimize with respect to $\boldsymbol{\beta}_i$, $i=1,\ldots,M$
$\sum_{1\le i\le M}\boldsymbol{\beta}_i^T(\boldsymbol{R}_i+\delta_i^0\boldsymbol{I})\boldsymbol{\beta}_i-2\boldsymbol{\beta}_i^T\boldsymbol{r}_i+2\lambda\lambda_i^0\|\boldsymbol{\beta}_i\|_1^1+\sum_{1\le i_1<i_2\le M}c\left\|\boldsymbol{\beta}_{i_1}^{s_{i_1i_2}}-\boldsymbol{\beta}_{i_2}^{s_{i_1i_2}}\right\|_2^2$
where λ and c are tuning parameters
Train penalized regression models using training GWAS
Get solution $\boldsymbol{\beta}_{\lambda,c,i}$ for each λ and c varying from their parameter paths, and $i=1,\ldots,M$
Step 3: Ensemble regression
Train ensemble regression model using tuning data
Train the super learning model on samples in tuning dataset to combine all PRSs computed from the solutions in the previous step, $\boldsymbol{PRS}_{\lambda,c,i}=\boldsymbol{X}\boldsymbol{\beta}_{\lambda,c,i}$, and obtain the weights in the super learning model, $w_{\lambda,c,i}$
Final PROSPER model:
$\boldsymbol{PRS}=\sum w_{\lambda,c,i}\boldsymbol{PRS}_{\lambda,c,i}$
$=\boldsymbol{X}\sum(w_{\lambda,c,i}\boldsymbol{\beta}_{\lambda,c,i})$

Figure 2

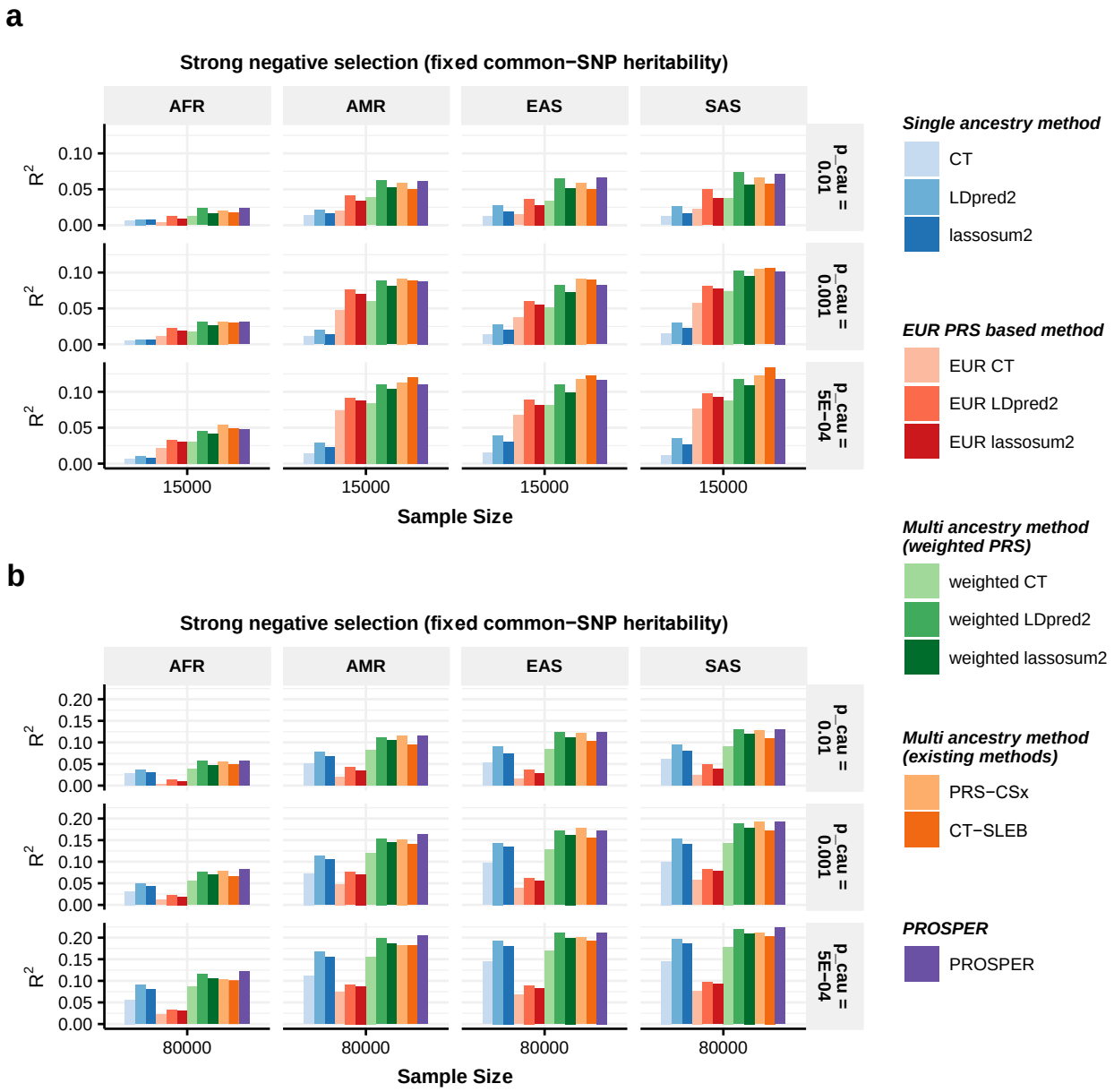

a
Strong negative selection (fixed common−SNP heritability)
AFR
AMR
EAS
SAS
R2
0.10
0.05
0.00
p_cau = 0.01
p_cau = 0.001
p_cau = 5E−04
15000
Sample Size
b
Strong negative selection (fixed common−SNP heritability)
0.20
0.15
0.10
0.05
0.00
80000
Sample Size
Single ancestry method
CT
LDpred2
lassosum2
EUR PRS based method
EUR CT
EUR LDpred2
EUR lassosum2
Multi ancestry method (weighted PRS)
weighted CT
weighted LDpred2
weighted lassosum2
Multi ancestry method (existing methods)
PRS−CSx
CT−SLEB
PROSPER
PROSPER

Figure 3

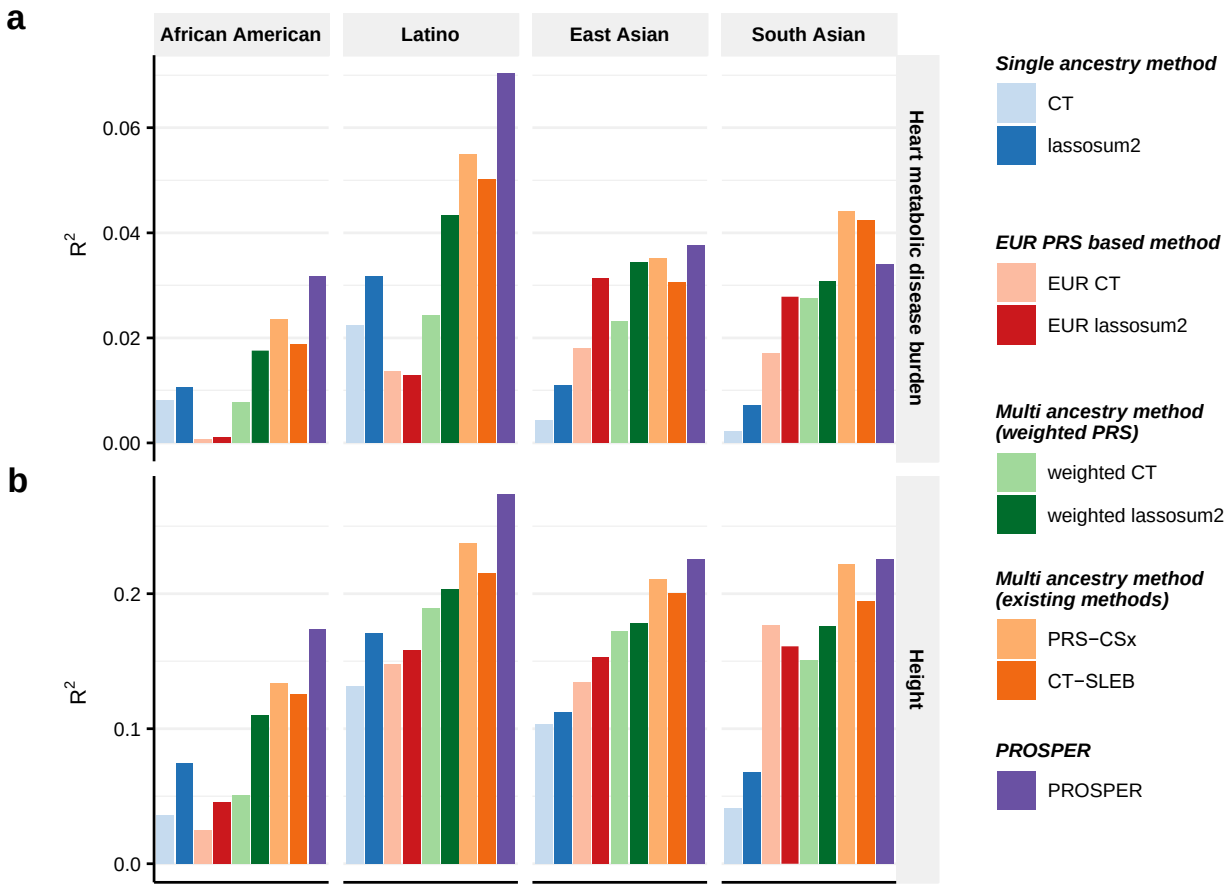

a
b
African American
Latino
East Asian
South Asian
Heart metabolic disease burden
Height
R2
0.00
0.02
0.04
0.06
0.0
0.1
0.2
Single ancestry method
CT
lassosum2
EUR PRS based method
EUR CT
EUR lassosum2
Multi ancestry method (weighted PRS)
weighted CT
weighted lassosum2
Multi ancestry method (existing methods)
PRS−CSx
CT−SLEB
PROSPER
PROSPER

Figure 4

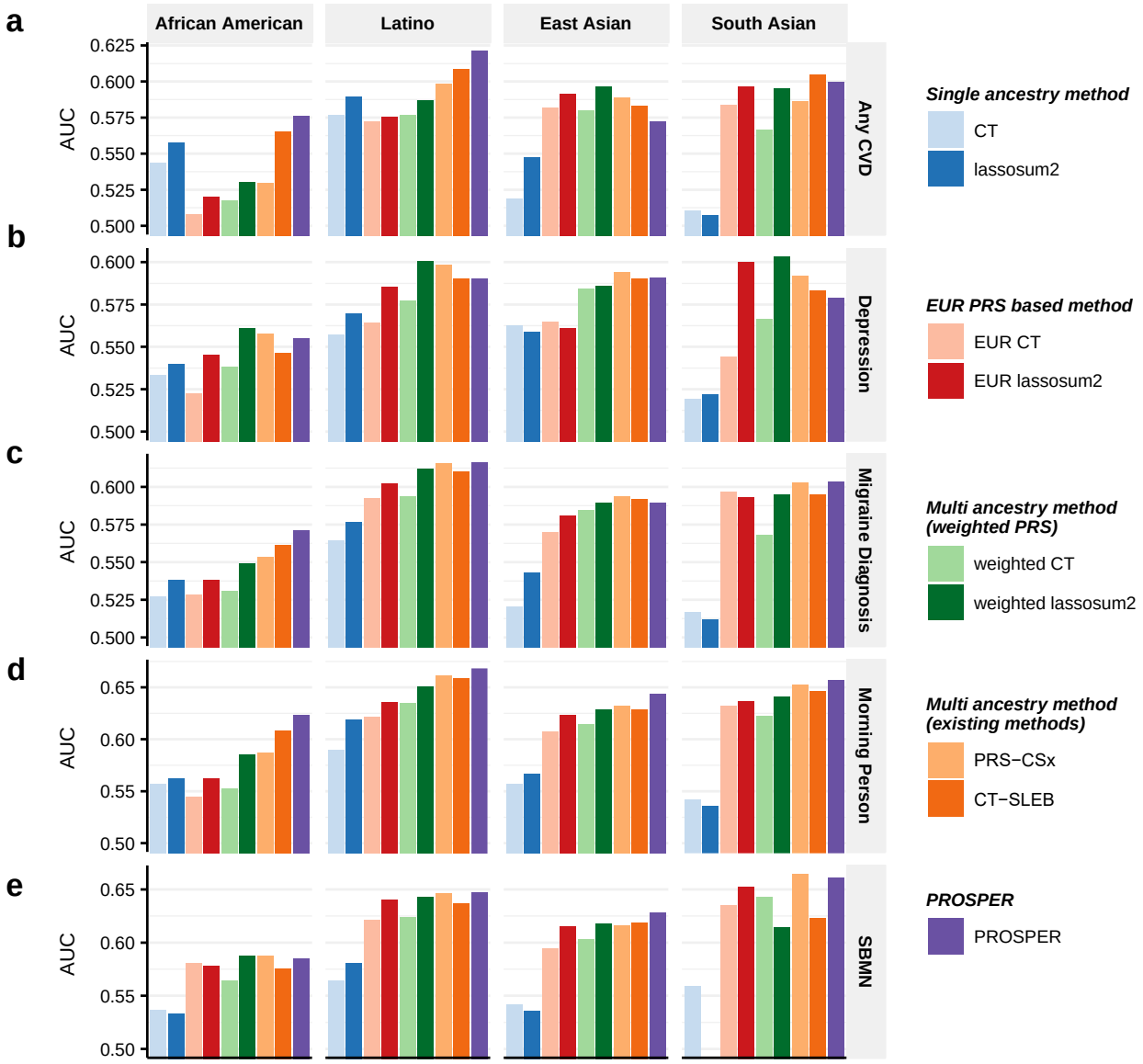

a
b
c
d
e
African American
Latino
East Asian
South Asian
AUC
Any CVD
Depression
Migraine Diagnosis
Morning Person
SBMN
Single ancestry method
CT
lassosum2
EUR PRS based method
EUR CT
EUR lassosum2
Multi ancestry method (weighted PRS)
weighted CT
weighted lassosum2
Multi ancestry method (existing methods)
PRS−CSx
CT−SLEB
PROSPER
PROSPER

Figure 5

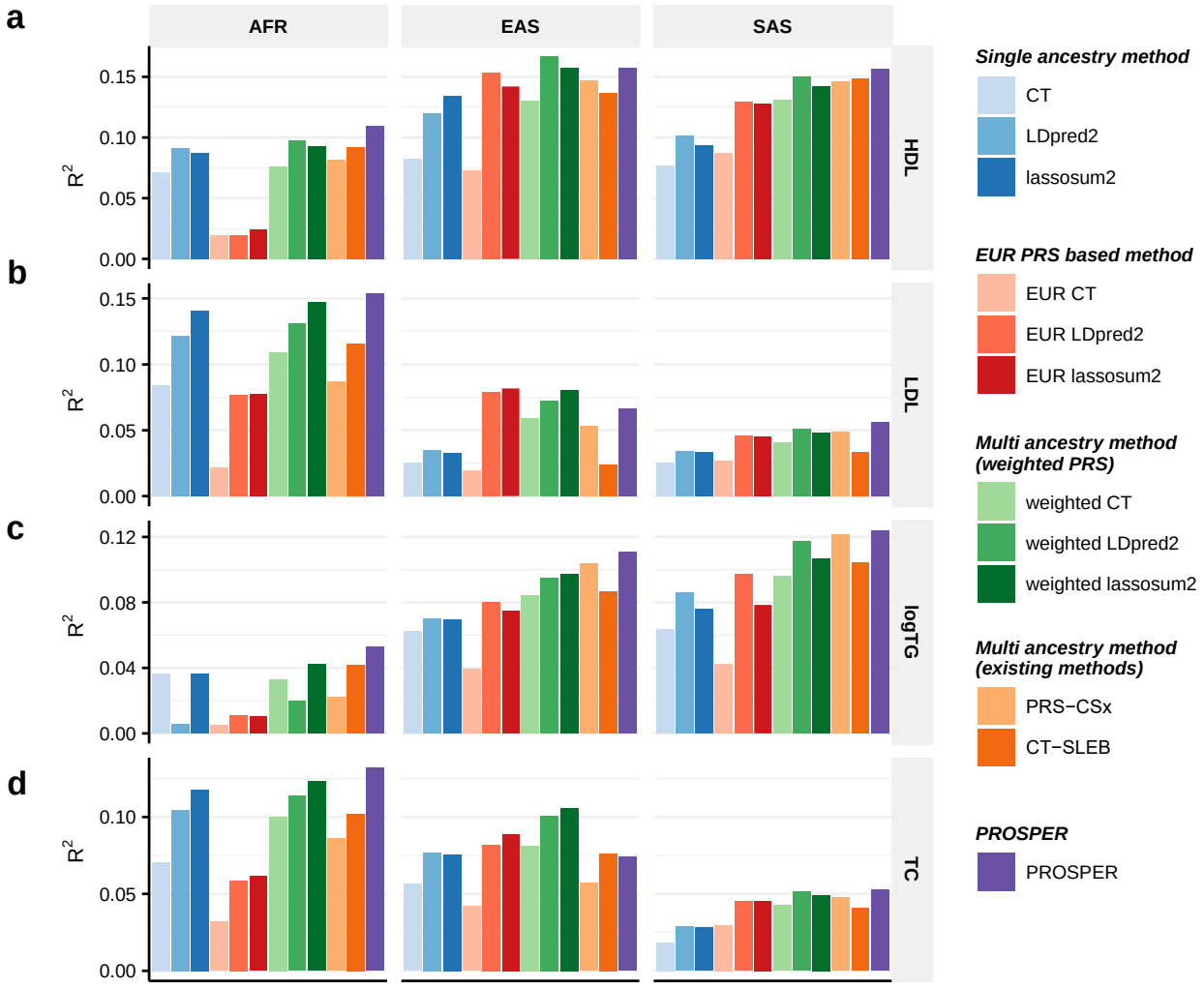
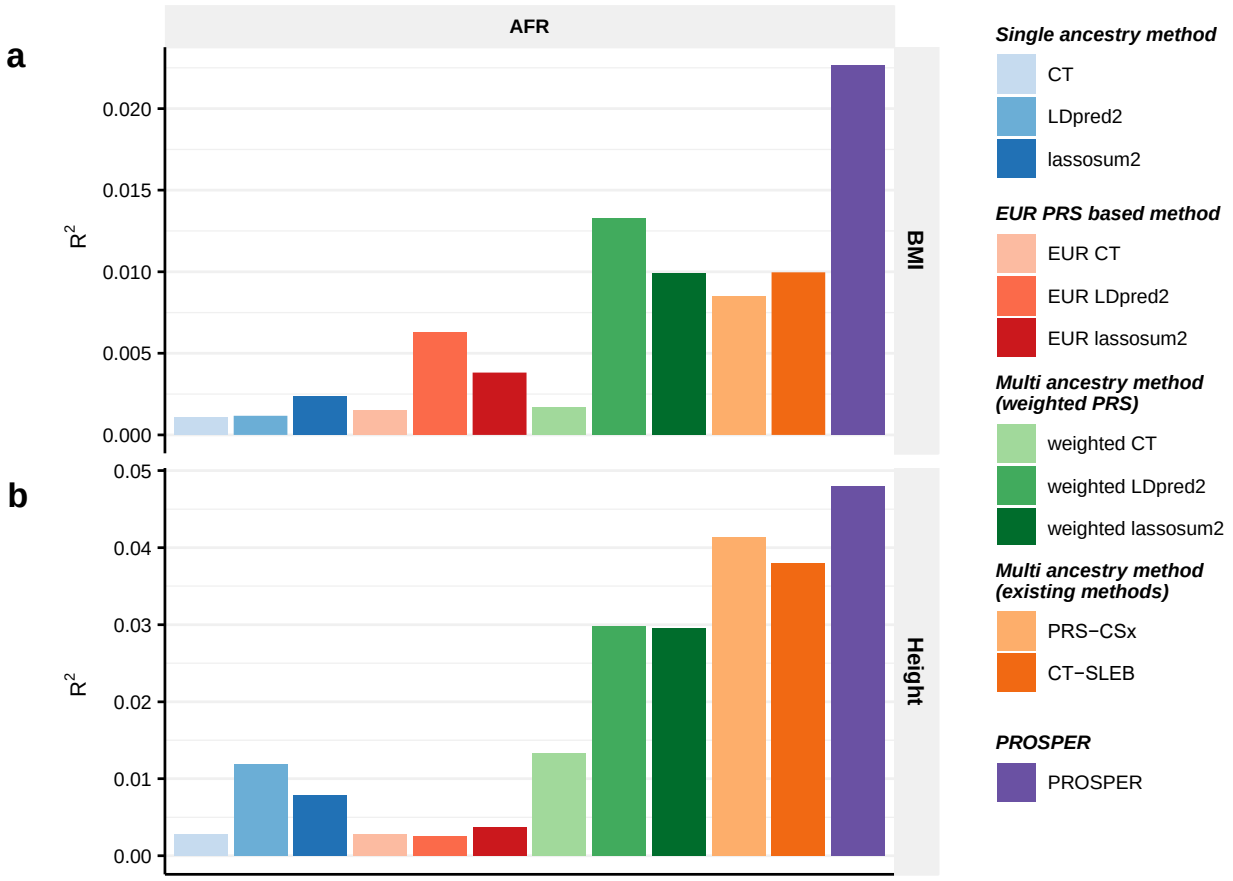

a
b
c
d
AFR
EAS
SAS
HDL
LDL
logTG
TC
R2
0.15
0.10
0.05
0.00
0.12
0.08
0.04
Single ancestry method
CT
LDpred2
lassosum2
EUR PRS based method
EUR CT
EUR LDpred2
EUR lassosum2
Multi ancestry method (weighted PRS)
weighted CT
weighted LDpred2
weighted lassosum2
Multi ancestry method (existing methods)
PRS−CSx
CT−SLEB
PROSPER
PROSPER

Figure 6


AFR
a
b
BMI
Height
R2
0.020
0.015
0.010
0.005
0.000
0.05
0.04
0.03
0.02
0.01
0.00
Single ancestry method
CT
LDpred2
lassosum2
EUR PRS based method
EUR CT
EUR LDpred2
EUR lassosum2
Multi ancestry method (weighted PRS)
weighted CT
weighted LDpred2
weighted lassosum2
Multi ancestry method (existing methods)
PRS−CSx
CT−SLEB
PROSPER
PROSPER

Consortia

23andMe Research Team

Jianan Zhan[2], Jared O'Connell[2], Yunxuan Jiang[2], Bertram L. Koelsch[2]

[2] 23andMe Inc., Sunnyvale, CA, USA.