

1 **Integrative polygenic risk score improves the prediction accuracy of**
2 **complex traits and diseases**

3
4 Buu Truong^{1,2}, Leland E. Hull^{3,4}, Yunfeng Ruan^{1,2}, Qin Qin Huang⁵, Whitney Hornsby^{1,2},
5 Hilary Martin⁵, David A. van Heel⁶, Ying Wang^{1,7,8}, Alicia R. Martin^{7,8}, S. Hong Lee⁹, Pradeep
6 Natarajan^{1,2,4}

7
8 ¹Program in Medical and Population Genetics and the Cardiovascular Disease Initiative,
9 Broad Institute of MIT and Harvard, 415 Main St, Cambridge, MA 02142

10 ²Center for Genomic Medicine and Cardiovascular Research Center, Massachusetts
11 General Hospital, 185 Cambridge Street, Boston, MA, 02114

12 ³Division of General Internal Medicine, 100 Cambridge Street, Massachusetts General
13 Hospital, Boston, MA, 02114

14 ⁴Department of Medicine, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115

15 ⁵Department of Human Genetics, Wellcome Sanger Institute, Cambridge, UK

16 ⁶Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary
17 University of London, London, UK

18 ⁷Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge,
19 MA, USA

20 ⁸Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA,
21 USA

22 ⁹Australian Centre for Precision Health, University of South Australia Cancer Research
23 Institute, University of South Australia, Adelaide, SA, 5000, Australia

24
25 Please address correspondence to:

26 Pradeep Natarajan, MD MMSc
27 185 Cambridge Street, CPZN 3.184, Boston, MA 02114
28 pnatarajan@mgh.harvard.edu
29 617-726-1843

30
31
32 **ABSTRACT**

33
34 Polygenic risk scores (PRS) are an emerging tool to predict the clinical phenotypes and
35 outcomes of individuals. Validation and transferability of existing PRS across independent
36 datasets and diverse ancestries are limited, which hinders the practical utility and
37 exacerbates health disparities. We propose PRSmix, a framework that evaluates and
38 leverages the PRS corpus of a target trait to improve prediction accuracy, and PRSmix+,
39 which incorporates genetically correlated traits to better capture the human genetic
40 architecture. We applied PRSmix to 47 and 32 diseases/traits in European and South Asian
41 ancestries, respectively. PRSmix demonstrated a mean prediction accuracy improvement of
42 1.20-fold (95% CI: [1.10; 1.3]; P-value = 9.17×10^{-5}) and 1.19-fold (95% CI: [1.11; 1.27]; P-
43 value = 1.92×10^{-6}), and PRSmix+ improved the prediction accuracy by 1.72-fold (95% CI:
44 [1.40; 2.04]; P-value = 7.58×10^{-6}) and 1.42-fold (95% CI: [1.25; 1.59]; P-value = 8.01×10^{-7})
45 in European and South Asian ancestries, respectively. Compared to the previously
46 established cross-trait-combination method with scores from pre-defined correlated traits, we
47 demonstrated that our method can improve prediction accuracy for coronary artery disease
48 up to 3.27-fold (95% CI: [2.1; 4.44]; P-value after FDR correction = 2.6×10^{-4}). Our method

49 provides a comprehensive framework to benchmark and leverage the combined power of
50 PRS for maximal performance in a desired target population.

51

52 INTRODUCTION

53

54 Thousands of polygenic risk scores (PRS) have been developed to predict an individual's
55 genetic propensity to diverse phenotypes¹. PRS are generated when risk alleles for distinct
56 phenotypes are weighted by their effect size estimates and summed². Risk alleles included
57 in PRS have traditionally been identified from genome-wide association studies (GWAS)
58 results conducted on a training dataset, which are weighted and aggregated to derive a PRS
59 to predict distinct phenotypes. The association between PRS and the phenotype of interest
60 is subsequently evaluated in a test dataset that is non-overlapping with the training dataset³.

61

62 Most PRS have been developed in specific cohorts that may vary in terms of population
63 demographics, admixture, environment, and SNP availability. Limited validation of many
64 PRS outside of the training datasets and poor transferability of PRS to other populations
65 may limit their clinical utility. However, pooling of data from individual PRS generated and
66 validated in diverse cohorts has the potential to improve the predictive ability of PRS across
67 diverse populations. The Polygenic Score Catalog (PGS Catalog) is a publicly available
68 repository that archives SNP effect sizes for PRS estimation. The SNP effect sizes were
69 developed from various methods (e.g. P+T⁴, LDpred^{5,6}, PRS-CS⁷, etc.) to obtain the highest
70 prediction accuracy in the studied dataset. PRS metadata enables researchers to replicate
71 PRS in independent cohorts and aggregate SNP effects to refine PRS and enhance the
72 accuracy and generalizability in broader populations⁸. However, optimizing PRS
73 performance requires methodological approaches to adjust GWAS estimate effect sizes that
74 take into account correlated SNPs (i.e., linkage disequilibrium) and refine PRS for the target
75 population^{4,5,7,9-12}. Furthermore, numerous scores are often present for single traits with
76 varied validation metrics in non-overlapping cohorts. There is a lack of standardized
77 approaches combining PRS from this growing corpus to enhance prediction accuracy and
78 generalizability while minimizing bias, for a target cohort^{8,11,13}.

79

80 To address these issues, we sought to: 1) validate previously developed PRS in two
81 geographically and ancestrally distinct cohorts, the *All of Us* Research Program (AoU) and
82 the Genes & Health cohort, and 2) present and evaluate new methods for combining
83 previously calculated PRS to maximize performance beyond all best performing published
84 PRS. To better capture the genetic architecture of the outcome traits, we proposed PRSmix,
85 a framework to combine PRS from the same trait with the outcome trait. Previous studies
86 highlighted the effect of pleiotropic information on a trait's genetic architecture^{14,15}.
87 Therefore, we proposed PRSmix+ to additionally combine PRS from other genetically
88 correlated traits to further improve the PRS for a given trait.

89

90 To assess the prediction improvement, we performed PRSmix and PRSmix+ for 47 traits in
91 European ancestry and 32 traits in South Asian ancestry. We evaluated 1) the relative
92 improvement of the proposed framework over the best-performing pre-existing PRS for each
93 trait, 2) the efficient training sample sizes required to improve the PRS, 3) the predictive
94 improvement in 6 groups including anthropometrics, blood counts, cancer, cardiometabolic,
95 biochemistry and other conditions as the prediction accuracies varied in each group, and 4)
96 the clinical utility and pleiotropic effect of the newly built PRS for coronary artery disease.

97 Overall, we show that PRSmix and PRSmix+ significantly improved prediction accuracy. An
98 R package for preprocessing and harmonizing the SNP effects from the PGS Catalog as
99 well as assessing and combining the scores was developed to facilitate the combining of
100 pre-existing PRS scores for both ancestry-specific and cross-ancestry contexts using the
101 totality of published PRS. The development of this framework has the potential to improve
102 precision health by improving the generalizability in the application of PRS¹⁶.

103

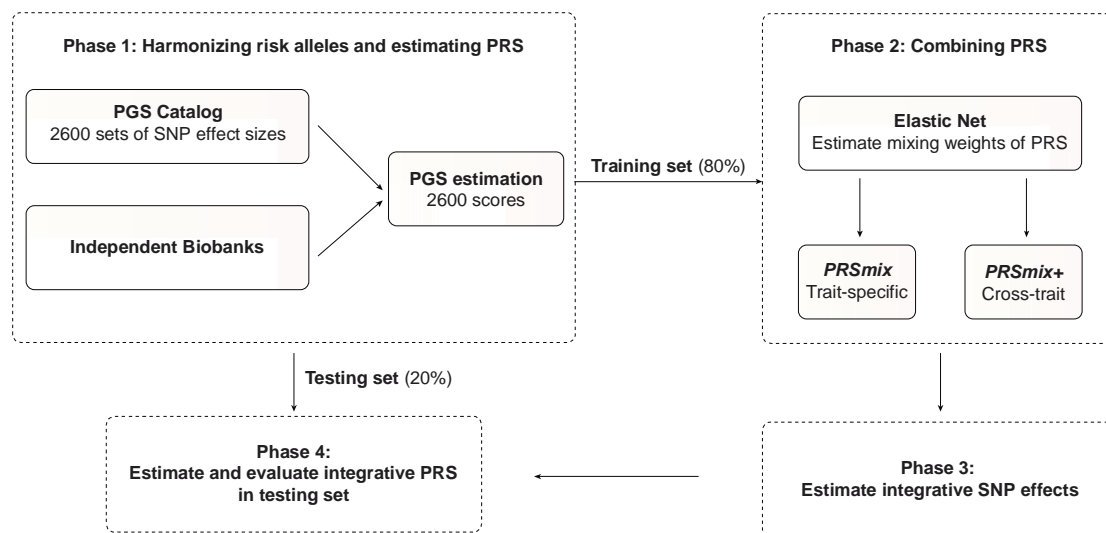
104 RESULTS

105

106 Overview of methods

107

108



109

110 **Figure 1. The framework of the trait-specific and cross-trait PRS integration.** In Phase
111 1, we obtained the SNP effects from the PGS Catalog and then harmonized the effect alleles
112 as the alternative alleles in the independent cohorts. In each independent biobank (*All of Us*,
113 *Genes & Health*), we estimated the PRS and split the data into training (80%) and testing
114 (20%) datasets. In Phase 2, in the training dataset, we trained the Elastic Net model with
115 high-power scores to estimate the mixing weights for the PRSs. The training phase could
116 include PRSs from traits corresponding to outcomes (PRSmix) or all traits (PRSmix+). The
117 training was adjusted for age, sex, and 10 principal components (PCs). In Phase 3, we
118 adjusted the per-allele effect sizes from each single PRS by multiplying with the
119 corresponding mixing weights obtained in the training phase. The final per-allele effect sizes
120 are estimated as the weighted sum of the SNP effects across different single scores. In
121 Phase 4, we evaluated the re-estimated per-allele effect sizes in the testing dataset.

122

123 A single PRS may only reflect genetic effects captured in the discovery dataset of a single
124 study that may be only a part of the total genetic effects underlying the trait of interest.
125 Therefore, we harmonized and combined multiple sets of PRS to establish a new set of
126 scores, which gather information across studies and traits. Our approach leveraged multiple
127 well-powered PRSs to improve prediction accuracy and is detailed in Fig. 1.

128

129 Our combination frameworks leveraged the PGS Catalog¹⁷ as the resource of SNP effects to
130 estimate single PRSs. To avoid overfitting, we used *All of Us* and Genes & Health cohorts
131 (see Methods) due to non-overlapping samples from the original GWAS. We randomly
132 divided the target cohort into a training set (80%) and a testing set (20%). We selected the
133 most common traits from the PGS Catalog which have the highest number of PRS. For the
134 stability of the linear combination, we curated binary traits with a prevalence > 2% in the
135 target cohort. Continuous traits were assessed using partial R^2 which is estimated as the
136 difference between the full model of PRS and covariates (age, sex, and 10 PCs) and the null
137 model of only covariates. For binary traits, the prediction accuracy was converted to liability
138 R^2 with disease prevalence approximated as the prevalence in the corresponding cohort.

139

140 To combine the scores, we employed Elastic Net¹⁸ to construct linear combinations of the
141 PRS. We proposed two combination frameworks: 1) PRSmix combines the scores
142 developed from the same outcome trait, and 2) PRSmix+ combines all the high-power
143 scores across other traits. Trait-specific combinations, PRSmix, can leverage the PRSs
144 developed from different studies and methods to more fully capture the genetic effects
145 underlying the traits. It has also been shown that complex traits are determined by genes
146 with pleiotropic effects¹⁵. Therefore, we additionally proposed a cross-trait combination,
147 PRSmix+, to make use of pleiotropic effects and further improve prediction accuracy.

148

149 First, we evaluated the improvement for each method, defined as the fold-ratio of the method
150 compared to the prediction accuracy of the best single PRS. For a fair comparison with the
151 proposed framework, we selected the best single PRS from the training set and evaluated its
152 performance in the testing set. First, we performed simulations to assess the improvement
153 with various heritabilities and training sample sizes. We estimated the slope of improvement
154 of prediction accuracy by increasing training sample sizes for various heritabilities.

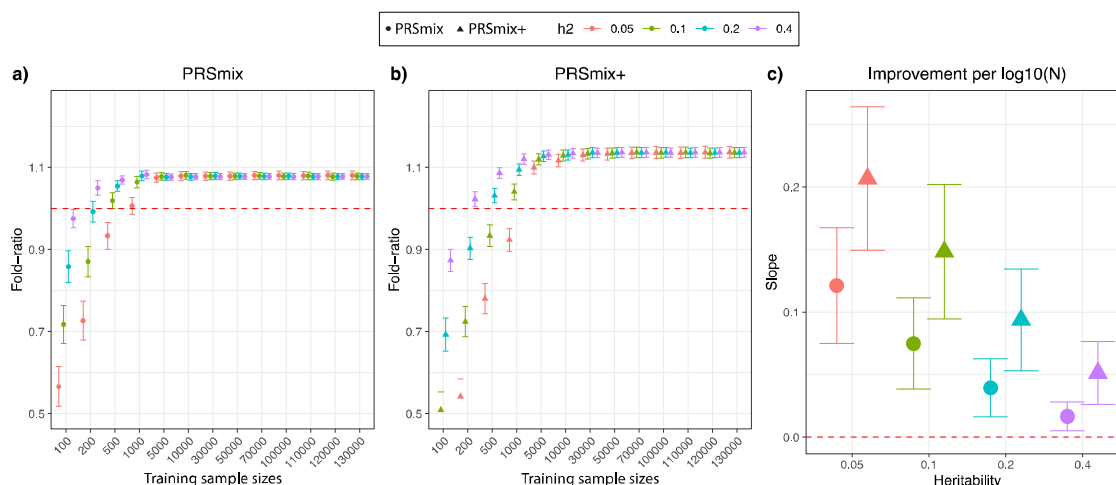
155

156 Next, we applied the proposed frameworks in two distinct cohorts; (1) the *All of Us* program,
157 in which 47 traits were tested in U.S. residents of European ancestry, and (2) the Genes &
158 Health (G&H) cohort, in which 32 traits were tested in British South Asian ancestry
159 (Supplementary Table 1). In each cohort, we compared the improvement of our proposed
160 framework with the single best score from the PGS Catalog. We estimated the averaged
161 fold-ratio as a measure of the improvement of prediction accuracy by our approach,
162 compared to the best single score from PGS Catalog. We also classified the traits into 6
163 categories as anthropometrics, blood counts, cancer, cardiometabolic, biochemistry, and
164 other conditions (Supplementary Table 2 and 3). Cancer traits were not considered in the
165 younger Genes & Health cohort due to their low prevalence (<2%). We then present
166 additional detailed analyses for coronary artery disease focused on clinical utility
167 improvements relative to existing PRS.

168

169 **Simulations were used to evaluate the combination frameworks**

170



171
172

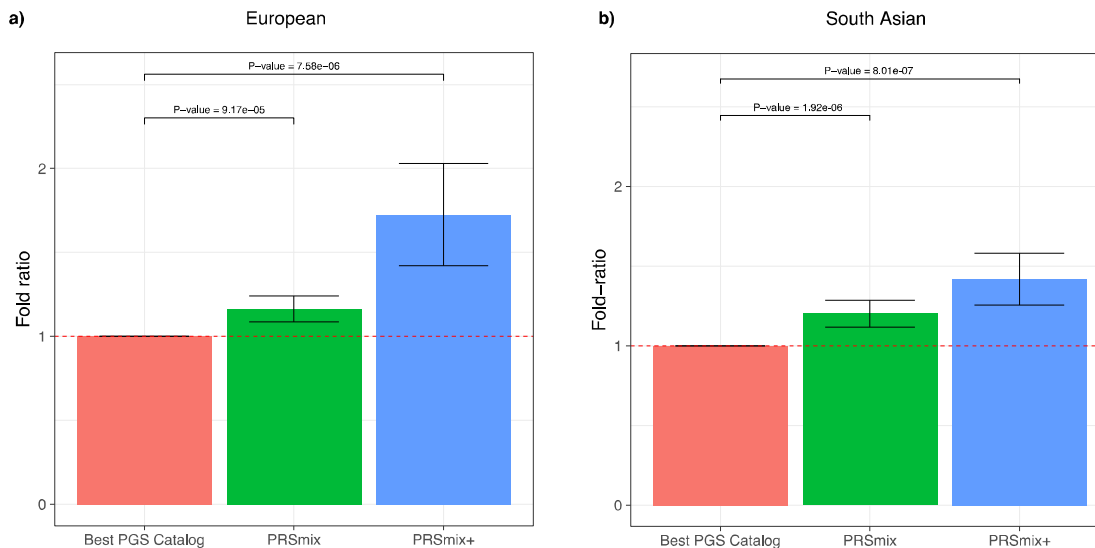
173 **Figure 2. Simulations to demonstrate the predictive improvement of PRSmix and**
 174 **PRSmix+.** The points and triangles represent the mean fold-ratio of R^2 between (a) PRSmix
 175 and (b) PRSmix+, respectively, versus the best single PRS. (c) The improvement per
 176 logarithm with base 10 of sample size for various heritabilities was represented as a slope of
 177 a linear regression of fold-ratio $\sim \log_{10}(N)$. In simulations, the correlation within simulated
 178 trait-specific PRSs was 0.8, and the correlation between trait-specific and correlated PRSs
 179 was 0.4 (see Methods). The whiskers demonstrate confidence intervals across 200
 180 replications. The dashed red lines represent the reference for fold-ratio equal 1 for (a) and
 181 (b), and equal 0 for (c).
 182

183 To compare the performance of PRSmix and PRSmix+ against the best single PRS and
 184 evaluate the sample sizes needed for training the mixing weights, we performed simulations
 185 with real genotypes of European ancestry in the UK Biobank given the large sample sizes
 186 available (Fig. 2). Briefly, we randomly split 7,000 individuals as a testing data set mimicking
 187 the testing size of 20% of real data. In the remaining dataset, we used 200,000 individuals
 188 for GWAS to estimate the SNP effect sizes for PRS calculations. Finally, with the rest of the
 189 data, we randomly selected different sample sizes as the training sample to evaluate the
 190 sample sizes needed to train the mixing weights. To assess the improvement of PRS
 191 performance, we computed the fold-ratio of prediction accuracy R^2 between PRSmix and
 192 PRSmix+ against the best-performing single simulated PRS.
 193

194 Our results showed that the trait-specific combination, PRSmix, showed no improvement
 195 with the training sample smaller than 500 for most of the traits. Our simulations illustrated
 196 that traits with low heritability required a larger sample size to achieve an improvement
 197 compared to traits with high heritability (Fig. 2a and 2b). PRSmix demonstrated a better
 198 performance compared to the best single PRS with training sample sizes from $N_{\text{training}} = 200$
 199 samples for the high heritable trait ($h^2 = 0.4$) to $N_{\text{training}} = 5000$ samples for the low heritable
 200 trait ($h^2 = 0.05$) (Fig. 2a and 2b). We observed that PRSmix demonstrated a saturation of
 201 improvement from $N_{\text{training}} = 10,000$. PRSmix+ demonstrated negligible further improvement
 202 when the training sample size was increased from 30,000 but maintained consistent
 203 improvement relative to PRSmix and the best single PRS. Moreover, we observed that traits
 204 with higher heritability or higher best prediction accuracy of a single PRS demonstrated a
 205 smaller improvement compared to traits with a smaller heritability (Fig. 2c).

206
207
208
209
210

Combining trait-specific PRS improves prediction accuracy (PRSmix)



211
212

Figure 3. Comparison of PRSmix and PRSmix+ versus the best PGS Catalog in European and South Asian ancestries. The relative improvement compared to the best single PRS was assessed in (a) the European ancestry in the *All of US* cohort and (b) South Asian ancestry in the Genes & Health cohort. PRSmix combines trait-specific PRSs and PRSmix+ combines additional PRSs from other traits. The best PGS Catalog score was selected by the best performance trait-specific score in the training sample and evaluated in the testing sample. The prediction accuracy (R^2) was calculated as partial R^2 which is a difference of R^2 between the model with PRS and covariates including age, sex, and 10 PCs versus the base model with only covariates. Prediction accuracy for binary traits was assessed with liability- R^2 where disease prevalence was approximately estimated as a proportion of cases in the testing set. The whiskers reflect the maximum and minimum values within the $1.5 \times$ interquartile range. The bars represent the ratio of prediction accuracy of PRSmix and PRSmix+ versus the best PRS from the PGS Catalog across 47 traits and 32 traits in *All of Us* and Genes and Heath cohorts, respectively, and the whiskers demonstrate 95% confidence intervals. P-values for significance difference of the fold-ratio from 1 using a two-tailed paired t-test. PRS: Polygenic risk scores.

229

230 To determine if a trait-specific combination, namely PRSmix, would improve the accuracy of
231 PRS prediction, we used data from European ancestry participants in the *All of Us* research
232 program who had undergone whole genome sequencing, and Genes & Health participants
233 of South Asian ancestry. We randomly split the independent cohorts into training (80%) and
234 testing sets (20%). The training set was used to train the weights of each PRS, referred as
235 mixing weights, that indicate how much each PRS explain the phenotypic variance in the
236 training set, and the PRS accuracies were evaluated in the testing set (Fig. 1). We curated
237 47 traits and 32 traits in the *All of Us* and Genes & Health cohorts, respectively. For binary
238 traits, we removed traits with a prevalence of smaller than 2% (see Methods, Supplementary

239 Table 1). Traits with the best-performance trait-specific single PRS which showed a lack of
240 power were also removed. Overall, we observed a significant improvement compared to 1
241 using a two-tailed paired t-test with PRSmix. PRSmix significantly improves the prediction
242 accuracy compared to the best PRS estimated from the PGS Catalog. PRSmix improved
243 1.20-fold (95% CI: [1.10; 1.3]; P-value = 9.17×10^{-5}) and 1.19-fold (95% CI: [1.11; 1.27]; P-
244 value = 1.92×10^{-6}) compared to the best PRS from PGS Catalog for European ancestry and
245 South Asian ancestry, respectively.

246

247 In European ancestry, we observed the greatest improvement of PRSmix against the best
248 single PRS for rheumatoid arthritis of 3.36-fold. Furthermore, in South Asian ancestry, we
249 observed that PRSmix of coronary artery disease had the best improvement of 2.32-fold
250 compared to the best-performance single PRS. Details of the prediction accuracy are shown
251 in Supplementary Fig. 1, 2 and Supplementary Table 2, 3. This was consistent with findings
252 in simulations since traits with a lower single PRS performance demonstrated a better
253 improvement with the combination strategy.

254

255 **Cross-trait combination further improved PRS accuracy and highlighted the**
256 **contribution of pleiotropic effects (PRSmix+)**

257

258 We next assessed the contribution of pleiotropic effects from cross-trait PRSs to determine if
259 these would further improve the combination framework (PRSmix+), by including high-power
260 PRSs from within 2600 PRSs in the PGS Catalog. To evaluate the power of PRS and
261 improve computational efficiency, we employed the theoretic power and variance of partial
262 R^2 for continuous traits and liability R^2 for binary traits (see Methods). We observed that
263 PRSmix+ further improved the prediction accuracy compared to the best PGS Catalog in
264 European ancestry (Fig. 3a) and South Asian ancestry (Fig. 3b). We observed an
265 improvement of 1.72-fold (95% CI: [1.40; 2.04]; P-value = 7.58×10^{-6}) and 1.42-fold (95% CI:
266 [1.25; 1.59]; P-value = 8.01×10^{-7}) higher compared to the best PGS Catalog for European
267 ancestry and South Asian ancestry, respectively. PRSmix+ significantly improved the
268 prediction accuracy compared to PRSmix, in both European and South Asian ancestry with
269 1.46-fold (95% CI: [1.17; 1.75]; P-value = 0.002) and 1.19-fold (95% CI: [1.07; 1.32]; P-value
270 = 0.001), respectively (Supplementary Fig. 3).

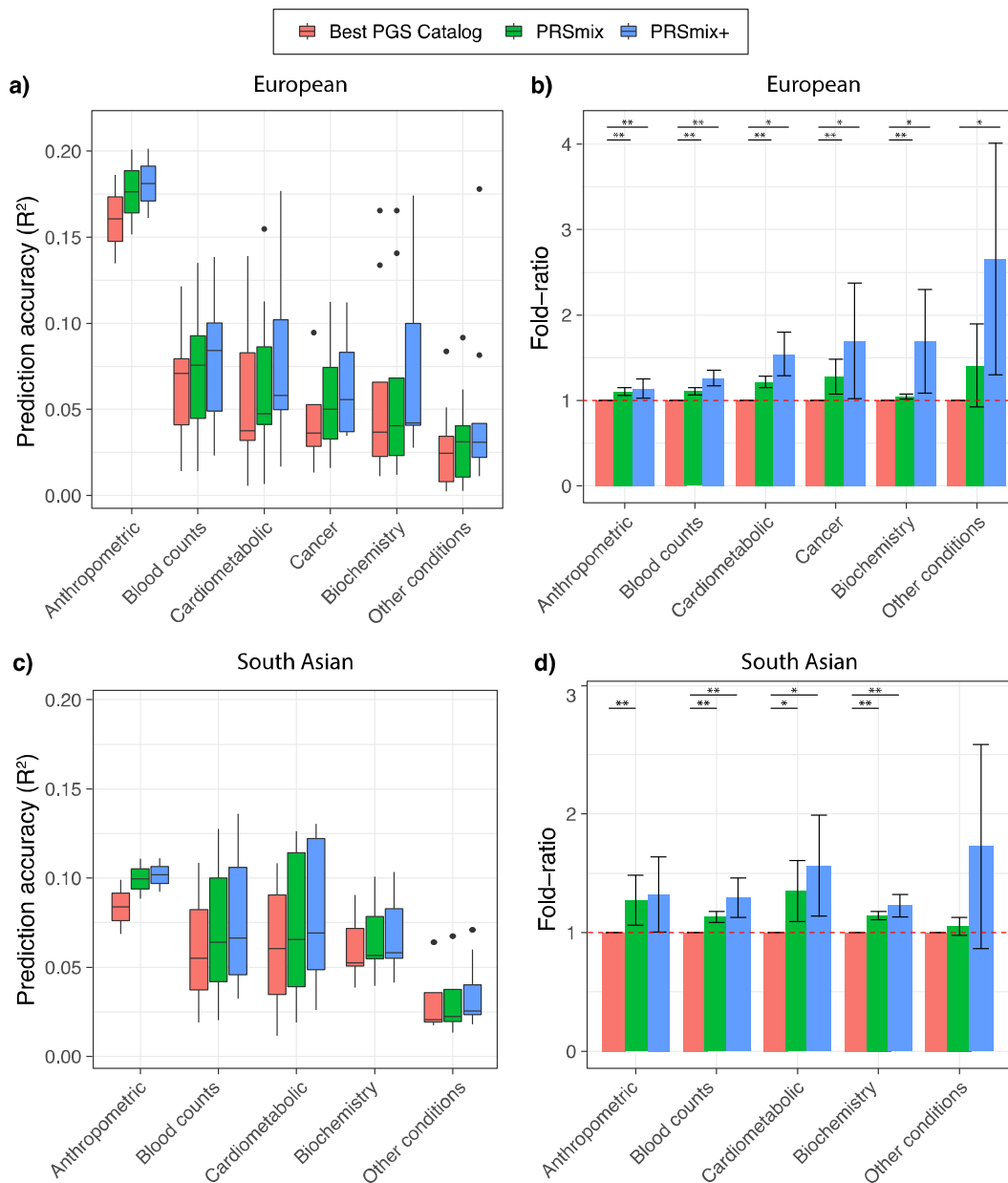
271

272 Consistent with our simulation results, a smaller improvement was observed for traits with a
273 higher baseline prediction accuracy from PGS Catalog (Supplementary Fig. 4), noting that
274 the baseline prediction accuracy depends on the heritability and genetic architecture (i.e.
275 polygenicity). In contrast, more improvement was observed for traits with lower heritability,
276 thus lower prediction accuracy, when comparing the single best PRS (Fig. 1c).

277

278 **Prediction accuracy and predictive improvement across various types of traits**

279



280

281

282

283

284

285

286

287

288

289

290

291

292

Figure 4. Prediction accuracy and improvement across various types of traits in the European and South Asian ancestry. We classified the traits into 6 main categories for European ancestry in the *All of Us* cohort and 5 categories for South Asian ancestry in the Genes & Health cohort due to the low prevalence of cancer traits in Genes & Health. The prediction accuracies, (a) and (c), are estimated as partial R^2 and liability R^2 for continuous traits and binary traits, respectively. The relative improvements, (b) and (d), are estimated as the fold-ratio between the prediction accuracies of PRSmix and PRSmix+ against the best PGS Catalog. The order on the axis followed the decrease in the prediction accuracy of PRSmix+. The boxplots in (a) and (c) show the first to the third quartile of prediction accuracies for 47 traits and 32 traits in European and South Asian ancestries, respectively. The whiskers reflect the maximum and minimum values within the 1.5 x interquartile range

293 for each group. The bars in **(b)** and **(d)** represent the mean prediction accuracy across the
294 traits in that group and the whiskers demonstrate 95% confidence intervals. The red dashed
295 line in (b) and (d) represents the ratio equal to 1 as a reference for comparison with the best
296 PGS Catalog score. The asterisk (*) and (**) indicate P-value < 0.05 and P-value < 0.05 /
297 number of traits in each type with a two-tailed paired t-test, respectively.

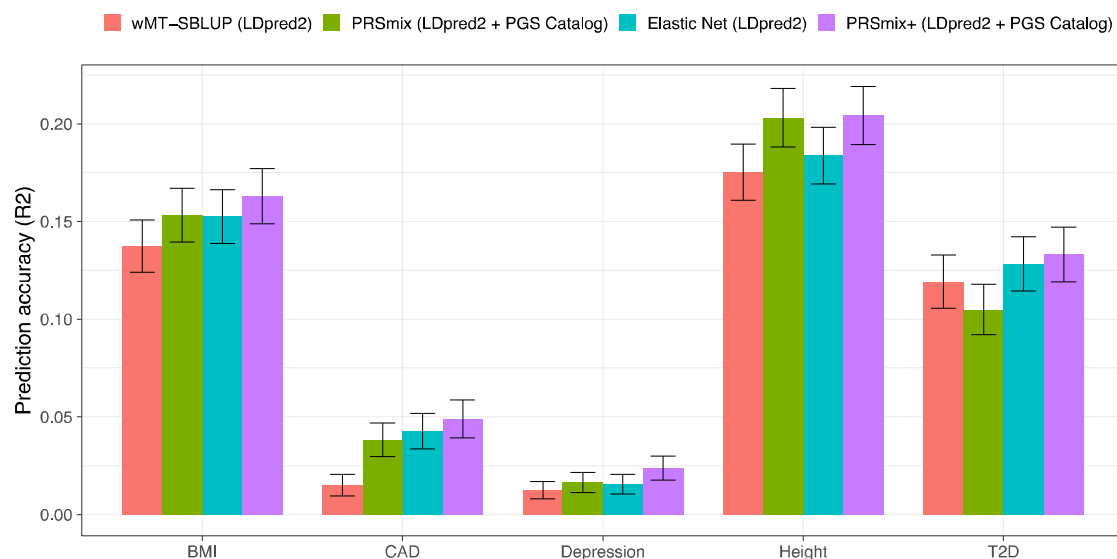
298

299 We next compared PRSmix and PRSmix+ with the best PRS estimated from the PGS
300 Catalog across 6 categories, including anthropometrics, blood counts, cancer,
301 cardiometabolic, biochemistry, and other conditions (see Methods). PRSmix demonstrates a
302 higher prediction accuracy across all types of traits in both European and South Asian
303 ancestries (Fig. 4). We observed a similar trend in the predictive performance of PRSmix+
304 across different types of traits. In European, the smallest improvement with PRSmix+ was in
305 anthropometric traits of 1.14-fold (95% CI: [1.03; 1.25]; P-value = 0.01) while “other
306 conditions” (including depression, asthma, migraine, current smoker, hypothyroid,
307 osteoporosis, glaucoma, rheumatoid arthritis, and gout) obtained the highest mean
308 predictive improvement but also with high variance of 2.66-fold (95% CI: [1.30; 4.01]; P-
309 value = 0.01) (Supplementary Table 4). In South Asian ancestry, the mean predictive
310 improvement was highest in “other conditions” (including asthma, migraine, current smoker,
311 and rheumatoid arthritis) type of 2.10-fold (95% CI: [0.787; 3.405]; P-value = 0.1).
312 Biochemistry demonstrated the smallest improvement of 1.23-fold (95% CI: [1.15; 1.31]; P-
313 value = 5.8×10^{-9}). We note that PRSmix and PRSmix+ improve prediction accuracy for all
314 traits (Supplementary Table 2 and 3). The large variance could be due to the wide range of
315 improvement and the small number of traits in each subtype.

316

317 Comparison with previous combination methods

318



319

320 **Figure 5. Benchmarking previous methods with PRSmix and PRSmix+.** LDpred2-auto
321 was used as the baseline method to input in the methods. 5 traits from Maier et al.¹⁹ and 26
322 publicly available GWAS for European ancestry were curated. The components of each
323 combination method are denoted in parentheses. wMT-SBLUP was conducted with the input
324 of sample sizes from the GWAS summary statistics and heritabilities and genetic correlation

325 between all pairs of traits using LD score regression. PRSmix (LDpred2 + PGS Catalog)
326 combined target trait-specific scores within 26 scores and PGS Catalog. Elastic Net
327 (LDpred2) was performed using Elastic Net with all scores from 26 traits generated with
328 LDpred2-auto. PRSmix+ (LDpred2 + PGS Catalog) was conducted using 26 scores from
329 LDpred2-auto and scores from all traits obtained from PGS Catalog. Partial R² and liability
330 R² were used for continuous traits and binary traits, respectively. The whiskers demonstrate
331 95% confidence intervals of mean prediction accuracy. BMI, Body mass index; CAD,
332 coronary artery disease; T2D, type 2 diabetes. GWAS, genome-wide association study.

333

334 There have been several studies proposed to incorporate multiple traits to improve
335 prediction accuracy of the target trait^{8,19,20}. For example, wMT-SBLUP¹⁹ created a weighted
336 index for correlated PRSs and required the input sample sizes, genetic correlation and
337 heritability across all pairs of traits from GWAS summary statistics to be determined. Krapohl
338 et al.²⁰ and Albinana et al.¹³ combined PRSs using scores estimated from LDpred2⁵. Here
339 we benchmarked PRSmix and PRSmix+ against the previous methods using summary
340 statistics with a pre-defined set of correlated traits to the main outcomes and an extension of
341 scores generated by different methods from PGS Catalog (Fig. 5).

342

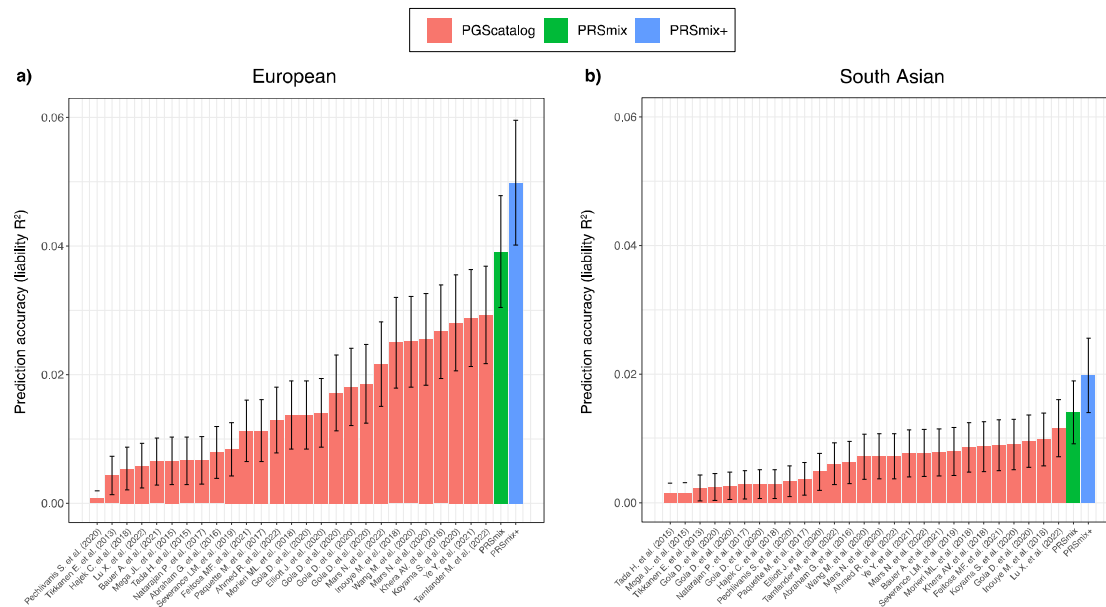
343 We first observed that integrating scores by Elastic Net with scores from pre-defined traits
344 improved prediction accuracy compared to wMT-SBLUP ranging between 1.08-fold (95% CI:
345 [1.03; 1.12]; P-value after FDR correction = 0.36) for T2D and 2.87-fold (95% CI: [1.58;
346 4.15]; P-value = 0.006) for CAD (Supplementary Table 5 and Supplementary Table 6).
347 PRSmix+, with scores from both pre-defined traits and PGS Catalog, demonstrated a
348 consistent boost in prediction accuracy compared to wMT-SBLUP between 1.12-fold (95%
349 CI: [1.02; 1.21]; P-value = 0.016) for T2D and 3.27-fold (95% CI: [2.1; 4.44]; P-value = $2.6 \times$
350 10^{-4}) for CAD. PRSmix+ equipped with both LDpred2-auto and PGS Catalog scores also
351 outperformed the Elastic Net combination of LDpred2 scores best observed with 1.6-fold
352 (95% CI: [1.31; 1.89]; P-value = 1.1×10^{-4}) for depression. Interestingly, height, a highly
353 polygenic trait²¹, demonstrated has the similarly best performance under a trait-specific
354 combination (PRSmix with trait-specific LDpred2-auto and PGS Catalog scores) and
355 PRSmix+ equipped with both LDpred2-auto and PGS Catalog scores (Fig. 5). Employing
356 pleiotropic effects only provided a small improvement with height (Supplementary Table 6).
357 On the other hand, T2D demonstrated that all methods of cross-trait combinations provided
358 a significant improvement over the trait-specific combination (Fig. 5).

359

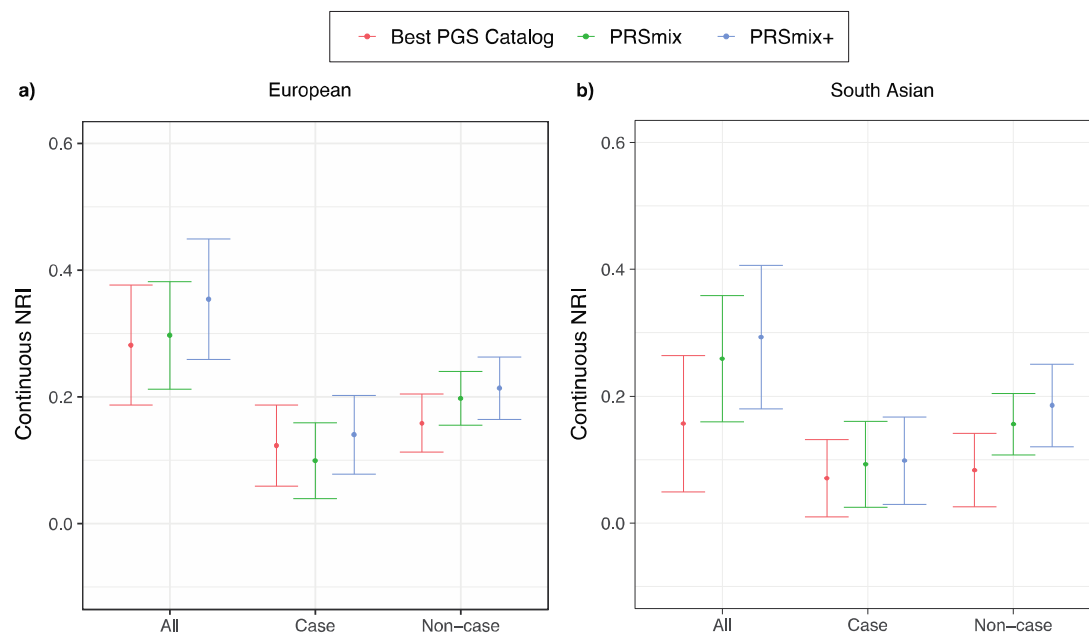
360 **Clinical utility for coronary artery disease**

361

362



363
 364 **Figure 6. Comparison of prediction accuracies with PRSmix, PRSmix+ and CAD PRS**
 365 **from PGS Catalog.** PRSmix was computed as a linear combination of CAD PRS and
 366 PRSmix+ was computed as a linear combination of all significant PRS obtained from the
 367 PGS Catalog. The PRSs were evaluated in the testing set with liability R^2 in the (a)
 368 European ancestry from the All of Us cohort and (b) South Asian ancestry from the Genes &
 369 Health cohort. The bars indicate the mean prediction accuracy and the whiskers show 95%
 370 confidence intervals. CAD, coronary artery disease.
 371



372
 373 **Figure 7. Net reclassification improvement (NRI) for coronary artery disease with the**
 374 **addition of polygenic risk scores to the baseline model in European and South Asian**
 375 **ancestries.** The baseline model for risk prediction includes age, sex, total cholesterol, HDL-
 376 C, systolic blood pressure, BMI, type 2 diabetes, and current smoking status. We compared
 377 the integrative models with PGS Catalog, PRSmix, and PRSmix+ in addition to clinical risk

378 factors versus the baseline model with only factors. The points indicate the mean estimate
379 for continuous NRI and the whiskers indicate 95% confidence intervals estimated from 500
380 bootstraps. HDL-C: High-density lipoprotein; BMI: Body mass index. NRI: Net
381 Reclassification Improvement.

382

383

384 To evaluate the utility of the proposed methods, we assessed the PRSmix and PRSmix+ for
385 coronary artery disease (CAD), which is the leading cause of disability and premature death
386 among adults²²⁻²⁴. The single best CAD PRSs (PRS_{CAD}) s from the PGS Catalog in the
387 training set were from Koyama S. et al²⁵. and Tamlander M. et al.²⁶ in European and South
388 Asian ancestries, respectively (Supplementary Fig. 5 and Supplementary Fig. 6). In the
389 testing set, liability R² with Koyama S et al. for European ancestry was 0.03 (95% CI: [0.03;
390 0.04]; P-value < 2 x 10⁻¹⁶) and with Tamlander M. et al. for South Asian ancestry was 0.006
391 (95% CI: [0.003; 0.009]; P-value = 2.39 x 10⁻⁴) (Fig. 6).

392

393 Subsequently, we assessed the clinical utility of the integrative model with PRS and
394 established clinical risk factors, including age, sex, total cholesterol, HDL-C, systolic blood
395 pressure, BMI, type 2 diabetes, current smoking status versus the traditional model with
396 clinical risk factors. (Fig. 7 and Supplementary Table 7). In European ancestry, the CAD
397 PRSmix+ integrative score improved the continuous net reclassification of 35% (95% CI:
398 [26%; 45%]; P-value < 2 x 10⁻¹⁶) compared to PRSmix (30%; 95% CI: [21%; 38%]; P-value =
399 P-value < 2 x 10⁻¹⁶) and the best PRS from the PGS Catalog (28%; 95% CI: [19%; 38%]; P-
400 value < 2 x 10⁻¹⁶). In South Asian ancestry, the integrated score with PRSmix+ showed
401 significant continuous net reclassification of 27% (95% CI: [16%; 38%]; P-value = 6.07 x 10⁻⁷)
402 compared to PRSmix (15%; 95% CI: [9%; 20%]; P-value = 7.18 x 10⁻⁶) and the best PGS
403 Catalog (7%; 95% CI: [1%; 13%]; P-value = 0.02). Our results also demonstrated an
404 improvement in net reclassification for models without clinical risk factors (Supplementary
405 Table 7).

406

407 We assessed the incremental area under the curve (AUC) between the full model of PRS
408 and covariates and the null model with only covariates (Supplementary Table 8). PRSmix+
409 demonstrated an incremental AUC of 0.02 (95% CI: [0.018; 0.02]; P-value < 2.2x10⁻¹⁶) and
410 0.008 (95% CI: [0.007; 0.009]; P-value < 2.2x10⁻¹⁶) in European and South Asian ancestries,
411 respectively. PRSmix obtained an incremental AUC of 0.016 (95% CI: [0.016; 0.017]; P-
412 value < 2.2x10⁻¹⁶) and 0.006 (95% CI: [0.005; 0.007]; P-value < 2.2x10⁻¹⁶) in European and
413 South Asian ancestries, respectively. The best PGS Catalog had the smallest incremental
414 AUC of 0.012 (95% CI: [0.011; 0.013]; P-value < 2.2x10⁻¹⁶) and 0.003 (95% CI: [0.002; 0.003];
415 P-value < 2.2x10⁻¹⁶) in European and South Asian ancestries, respectively.

416

417 We also compared the risks for individuals in the top decile versus the remaining population
418 (Supplementary Table 9). For European ancestry, an increased risk with OR per 1-SD of the
419 best PGS Catalog, PRSmix and PRSmix+ were 1.43 (95% CI: [1.30-1.57]; P-value < 2.2x10⁻¹⁶),
420 1.60 (95% CI: [1.45-1.76]; P-value < 2.2x10⁻¹⁶) and 1.74 (95% CI = [1.58; 1.91]; P-value
421 < 2.2x10⁻¹⁶), respectively. The top decile of PRSmix+ compared to the remaining population
422 demonstrated an increased risk of OR = 2.53 (95% CI: [1.96; 3.25]; P-value = 8.64 x 10⁻¹³).
423 The top decile for the best PGS Catalog versus the remainder was OR = 1.67 (95% CI:
424 [1.27; 2.19]; P-value = 2 x 10⁻⁴). For South Asian ancestry, an increased risk with OR per 1-
425 SD of the best PGS Catalog, PRSmix and PRSmix+ was 1.24 (95% CI: [1.13; 1.37]; P-value

426 $< 1.52 \times 10^{-16}$), 1.39 (95% CI: [1.33; 1.46]; P-value $< 2.2 \times 10^{-16}$), 1.40 (95% CI: [1.27; 1.55];
427 P-value $< 2.2 \times 10^{-16}$) and 1.50 (95% CI = [1.36; 1.66]; P-value $< 2.2 \times 10^{-16}$), respectively. In
428 South Asian ancestry, PRSmix+ demonstrated an OR of 2.34 (95% CI: [1.79; 3.05]; P-value
429 = 4.22×10^{-10}), and with the best PGS Catalog, OR was 1.73 (95% CI: [1.30; 2.28]; P-value
430 = 1.31×10^{-4}) for the top decile versus the remaining population.

431

432 Moreover, we observed that there is a plateau of improvement for PRSmix from the training
433 size of 5000 in both European and South Asian ancestries (Supplementary Fig. 7), which
434 aligned with our simulations (Fig. 2a and 2b). Our results demonstrated the generalization of
435 our combination methods across diverse ancestries to improve prediction accuracy. With
436 PRSmix+, our empirical result showed that there was a modest improvement with training
437 sample sizes larger than 5,000.

438

439 Finally, we conducted phenome-wide association studies (PheWAS) in *All of Us* between
440 PRS_{CAD} with 1815 phecodes to compare the pleiotropy of PRS and assess the relationship
441 between CAD PRS and disease phenotypes given the inherent use of pleiotropy in
442 development (Supplementary Table 10). As expected, PRSmix+ had a stronger association
443 for coronary atherosclerosis relative to the single best PRS from the PGS Catalog. PRSmix+
444 associations with cardiometabolic risk factors were significantly greater with averaged fold-
445 ratio = 1.10 (95% CI: [1.09-1.12]; P-value with paired T test = 1.07×10^{-25}) and 1.07 (95% CI:
446 [1.05-1.081]; P-value = 4.8×10^{-13}) for circulatory system and endocrine/metabolic system
447 (Supplementary Table 11). The PheWAS result for PRSmix+ aligned with the list of traits
448 from the selected PRS (Supplementary Table 10).

449

450 DISCUSSION

451

452 In this paper, we propose a trait-specific framework (PRSmix), and cross-trait framework
453 (PRSmix+) to leverage the combined power of existing scores. We performed and evaluated
454 our method using the *All of Us* and Genes & Health cohorts showcasing a framework to
455 develop the most optimal PRS for a given trait in a target population leveraging all existing
456 PRS. Across 47 traits in *All of Us* cohort and 32 traits in the Genes & Health cohort with
457 either continuous traits or binary traits with prevalence $> 2\%$, we demonstrated substantial
458 improvement in average prediction R^2 by using a linear combination with Elastic Net. The
459 empiric observations are concordant with simulations. To our knowledge, there has been a
460 number of emerging studies to combine PRS, but there is a limited number of frameworks
461 that comprehensively evaluate, harmonize, and leverage the combination of these
462 scores^{8,13,27}. Our studies permit several conclusions for the development, implementation,
463 and transferability of PRS.

464

465 First, externally derived and validated PRS are generally not the most optimal PRS for a
466 given cohort. Consistent with other risk predictors, recalibration within the ultimate target
467 population improves performance²⁸. By leveraging the PGS Catalog, our work carefully
468 harmonizes the risk alleles to estimate PRS across all scores and provides newly estimated
469 per-allele SNP effects (provided to the PGS Catalog) to assist the interpretability of the
470 models.

471

472 Second, previous studies selected an arbitrary training sample size to estimate the mixing
473 weights, which may lead to a poor power of the combination frameworks and inaccurate

474 estimate of sampling variance¹⁰. We assessed the expected sample sizes to estimate the
475 mixing weights via simulations and real data. Our results demonstrated that while low
476 heritability traits benefit the most, they require a greater training sample size.

477

478 Third, we leveraged all PRS, including those not trained on the primary trait, to
479 systematically optimize PRS for a target cohort. We showed that PRSmix improved the
480 prediction by combining the scores matching the outcome trait. In addition, we showed that
481 PRSmix+ was able to leverage the power of cross-traits, which highlighted the contribution
482 of pleiotropic effects to enhance PRS performance. We leverage prior work demonstrating
483 the effects of pleiotropy on complex traits^{15,29,30}.

484

485 Fourth, we demonstrated that our method outperformed previous methods combining
486 scores. We showed that PRSmix+ outperformed wMT-SBLUP¹⁹ using a limited number of
487 correlated traits. wMT-SBLUP required GWAS's sample sizes, heritability, and genetic
488 correlation between all traits. LDpred2-auto required GWAS summary statistics and
489 initialized heritability and proportion of causal SNPs. Krapohl et al.²⁰ and Abraham et al.⁸
490 proposed to use Elastic Net to combine the scores developed from summary statistics, and
491 correlated traits were selected with prior knowledge. However, these strategies consider
492 scores developed from a particular methods using predefined summary statistics. Our
493 framework utilizes all PRSs available in the PGS Catalog which were optimized for their
494 target traits. Additional summary statistics and PRS scores could be added to further
495 enhance the models. We let our Elastic Net model penalize the component PRSs without the
496 need for prior knowledge. Elastic Net can select PRSs to include and efficiently handle multi-
497 collinearity³¹⁻³³. Furthermore, PRSmix and PRSmix+ only required a set of SNPs effect to
498 estimate the PRSs and estimated the prediction accuracy to the target trait to select the best
499 scores for the combination. Additionally, compared to the preselected traits for stroke by
500 Abraham et al.⁸ we also observed that our method could identify more related risk factors to
501 include compared to previous work conducted on stroke (Supplementary Fig. 8). Therefore,
502 our method is more comprehensive in an unbiased way in terms of choosing the risk factors
503 and traits to include with empirically improved performance.

504

505 Fifth, greater performance is observed even for non-European ancestry groups
506 underrepresented in GWAS and PRS studies. We empirically demonstrate the value of
507 training and incorporating pleiotropy with all available PRS to improve performance,
508 including multiple metrics of clinical utility for CAD prediction in multiple ancestries. In South
509 Asian ancestry, we observed that PRSmix and PRSmix+ demonstrated a significant
510 improvement with the best improvement for CAD. Of note for CAD, the relative
511 improvements in South Asian ancestry were higher than in European ancestry for PRSmix
512 and equivalent for PRSmix+. Transferability of PRS has been shown to improve the clinical
513 utility of PRS in non-European ancestry^{16,34}. Although the prediction accuracy for South
514 Asian ancestry is still limited, our results highlighted the transferability of predictive
515 improvement with PRSmix and PRSmix+ to South Asian ancestry. We anticipate that
516 ongoing and future efforts to improve our understanding of the genetic architecture in non-
517 European ancestries will further improve the transferability of PRS across ancestry.

518

519 Lastly, traits with low heritability or generally low-performing single PRS benefit the most
520 from this approach, especially with PRSmix+, such as migraine in both European and South
521 Asian ancestries. Additionally, our results showed that pleiotropic effects play an important

522 role in understanding and improving prediction accuracies of complex traits. However,
523 anthropometric traits, which are highly polygenic³⁵ and have good predictive performance
524 using the best PGS Catalog, also showed improvement with the combination framework in
525 both European and South Asian ancestries.

526

527 Given that PRSmix+ outperformed PRSmix, one might consider if there is a reason to use
528 PRSmix instead of PRSmix+. We observed that in cases of highly heritable traits or high
529 performance with a single PRS, there was only marginal improvement of PRSmix+ over
530 PRSmix. In this scenario, PRSmix could provide similar predictive performance while being
531 less time-consuming because trait-specific PRS inputs only are required. However, for traits
532 with lower heritability PRSmix+ shows a marked improvement over PRSmix and would be
533 preferred. Wang et al.³⁶ showed that the theoretical prediction accuracy of the target trait
534 using the PRS from the correlated trait is a function of genetic correlation, heritability,
535 number of genetic variants and sample size. Future directions could include defining the
536 minimum parameters required for the performance of the PRSmix+ model to improve on
537 single trait-specific PRS.

538

539 Our work has several limitations. First, the majority of scores from PGS Catalog were
540 developed in European ancestry populations. Further non-European SNP effects will likely
541 improve the single PRS power, which may in turn, also improve the prediction accuracy of
542 our proposed methods. Second, the Elastic Net makes a strong assumption that the
543 outcome trait depends on a linear association with the PRS and covariates. However, a
544 recent study demonstrated there is no statistical significance difference between linear and
545 non-linear combinations for neuropsychiatric disease¹³. Third, we did not validate the mixing
546 weights in an independent cohort. We expect that in the future, there will be emerging large
547 independent biobanks, but prior non-genetic work demonstrates the value of internal
548 calibration for optimal risk prediction. Fourth, we estimated the mixing weights for each
549 single SNP as a mixing weight of the PRS. Future studies could consider linkage
550 disequilibrium between the SNPs and functional annotations of each SNP. Fifth, our
551 frameworks were conducted on binary traits with a prevalence > 2%. Additional combination
552 PRS models are emerging that seek to use preexisting genotypic data from genetically
553 related, but low prevalence conditions, to improve the prediction accuracy of rare
554 conditions¹³. Sixth, the baseline demographic characteristics (i.e., age, sex, social economic
555 status) in the target cohort might limit the validation and transferability of PRS³⁷. Although
556 these factors were considered by using a subset of the target cohorts as training data, it is
557 necessary to have PRS developed on similar baseline characteristics. Lastly, with the
558 expanding of all biobanks, there might be no perfect distinction between the samples
559 deriving PRS and the testing cohort, future studies may consider the potential intersection
560 samples to train the linear combination.

561

562 In conclusion, our framework demonstrates that leveraging different PRS either trait-specific
563 or cross-trait can substantially improve model stability and prediction accuracy beyond all
564 existing PRS for a target population. Importantly, we provide software to achieve this goal in
565 independent cohorts.

566

567 **METHODS**

568

569 **Data**

570

571 *The All of Us Research Program*

572

573 The *All of Us* Research Program is a longitudinal cohort continuously enrolling (starting May
574 2017) U.S. adults ages 18 years and older from across the United States, with an emphasis
575 on promoting inclusion of diverse populations traditionally underrepresented in biomedical
576 research, including gender and sexual minorities, racial and ethnic minorities, and
577 participants with low levels of income and educational attainment.³⁸ Participants in the
578 program can opt-in to providing self-reported data, linking electronic health record data, and
579 providing physical measurement and biospecimen data.³⁹ Details about the *All of Us* study
580 goals and protocols, including survey instrument development,⁴⁰ participant recruitment, data
581 collection, and data linkage and curation were previously described in detail.^{39,41}

582

583 Data can be accessed through the secure *All of Us* Researcher Workbench platform, which
584 is a cloud-based analytic platform that was built on the Terra platform.⁴² Researchers gain
585 access to the platform after they complete a 3-step process including registration,
586 completion of ethics training, and attesting to a data use agreement attestation.⁴³ *All of Us*
587 uses a tiered approach based on what genomic data is accessible through the Controlled
588 Tier, and includes both whole genome sequencing (WGS), genotyping array variant data in
589 multiple formats, as well as variant annotations, access to computed ancestry, and quality
590 reports.⁴⁴ This study includes data on the 98,600 participants with (WGS) data in the *All of*
591 *Us* v6 Curated Data Repository release. Participant data in this data release was collected
592 between May 6, 2018 and April 1, 2021. This project is registered in the *All of Us* program
593 under the workspace name “Polygenic risk score across diverse ancestries and biobanks.”

594

595 *The Genes & Health Biobank*

596

597 Genes & Health is a community-based genetics study enrolling British South Asian, with an
598 emphasis on British Bangladeshi (two-thirds) and British Pakistani (remaining) people, with a
599 goal of recruiting at least 100,000 participants. Currently, over 52,000 participants have
600 enrolled since 2015. All participants have consented for lifelong electronic health record
601 access and genetic analysis. The study was approved by the London South East National
602 Research Ethics Service Committee of the Health Research Authority. 97.4% of participants
603 in Genes & Health are in the lowest two quintiles of the Index of Multiple Deprivation in the
604 United Kingdom. The cohort is broadly representative of the background population with
605 regard to age, but slightly over-sampled with females and those with medical problems since
606 two-thirds of people were recruited in healthcare settings such as General Practitioner
607 surgeries⁴⁵.

608

609 *The Polygenic Score (PGS) Catalog*

610

611 Polygenic risk scores were obtained from the Polygenic Score (PGS) Catalog¹⁷, which is a
612 publicly accessible resource cataloging published PRS, including the metadata. The
613 metadata provides information describing the computational algorithms used to generate the
614 score, and performance metrics to evaluate a PRS¹⁷. At the time of this study, over 2,600
615 PRS were cataloged in the PGS Catalog (version July 18, 2022) designed to predict 538
616 distinct traits.

617

618 **Clinical Outcomes**

619

620 Clinical phenotypes were curated using a combination of electronic health record data, direct
621 physical measurements, and/or self-reported personal medical history data, from the *All of*
622 *Us* v6 Data Release as detailed in Supplementary Table 16. Individuals in the Genes and
623 Health cohort were also curated with similar definitions based on ICD10, SNOMED and
624 operation codes (Supplementary Table 17). Traits with the best performing single trait-
625 specific PRS with power < 0.95 such as hemoglobin, sleep apnea, and depression were
626 removed. Binary traits with a prevalence < 2% were removed.

627

628 **A linear combination of scores**

629

630 We proposed PRSmix to combine PRS of outcome traits and PRSmix+ to combine high-
631 power PRS (defined in the following subsection) from all traits obtained from PGS Catalog.
632 The linear combination was conducted by using an Elastic Net algorithm from the “*glmnet*” R
633 package⁴⁶ (version 4.1) to combine the estimated PRS. First, we randomly split the
634 independent cohorts into 80% of training and 20% testing. The PRS in the training set was
635 standardized with mean 0 and variance 1. Before conducting linear combination, we first
636 evaluated the performance of each individual PRS by their power and P-value (see below).
637 An Elastic Net algorithm was used with 5-fold cross-validation and default parameters to
638 estimate the mixing weights of each PRS. The mixing weights were then divided by the
639 corresponding original standard deviation of the PRS in the training set.

640

$$\hat{\alpha}_i = \hat{\omega}_i / \sigma_i$$

641

642 Where $\hat{\omega}_i$ and σ_i is the mixing weight estimated from the Elastic Net and standard deviation
643 of PRS_{*i*} in the training set, respectively. $\hat{\alpha}_i$ is the adjusted mixing weight for PRS_{*i*}. To derive
644 the per-allele effect sizes from the combination framework, we multiplied the SNP effects
645 with the corresponding adjusted mixing weights:

$$\hat{\gamma}_j = \sum_{i=1}^M \hat{\alpha}_i * \beta_{ji}$$

646 Where $\hat{\gamma}_i$ is the adjusted effect size of SNP_{*j*} and β_{ij} is the original effect sizes of SNP_{*j*} in
647 PRS_{*i*}. We set $\beta_{ji} = 0$ if SNP_{*j*} is not in PRS_{*i*}. The adjusted effect sizes were then utilized to
648 calculate the final PRS.

649

650 The mixing weights for PGS Catalog scores for PRSmix and PRSmix+ in European ancestry
651 are provided in Supplementary Table 12 and Supplementary Table 13, respectively. For
652 South Asian ancestry, the mixing weights for PRSmix and PRSmix+ in European ancestry
653 are provided in Supplementary Table 14 and Supplementary Table 15, respectively.

654

655 **Power and variance of PRS accuracy**

656

657 We selected high-power PRS to conduct the combination by assessing the power and
658 variance of prediction accuracy. The power of PRS can be estimated based on the power of
659 the two-tailed test of association as follow^{3,47}:

660

$$661 \quad 1 - \phi(\phi^{-1}(1 - \alpha/2) - \sqrt{\lambda}) + \phi(\phi^{-1}(\alpha/2) - \sqrt{\lambda}) \quad (1)$$

662

663 where ϕ is the Chi-squared distribution function, α is the significance level, and λ is the non-
664 centrality parameter which can be estimated as

665

$$\lambda = \frac{NR^2}{1-R^2} \quad (2)$$

667

668 where N , R^2 is the sample size and estimated prediction accuracy in the testing set,
669 respectively. R^2 can be estimated as partial R^2 or liability R^2 for continuous traits and binary
670 traits, respectively. Briefly, partial R^2 compared the difference in goodness-of-fit between a
671 full model with PRS and covariates including age, sex, and first 10 PCs, and a null model
672 with only covariates. Additionally, for binary traits, liability R^2 was estimated with the disease
673 prevalence approximated as the prevalence in the samples. The theoretical variance and
674 standard error of R^2 can be estimated as follow⁴⁸⁻⁵⁰:

675

$$se(R^2) = \sqrt{var(R^2)} = \sqrt{\frac{4R^2(1-R^2)(N-2)^2}{(N^2-1)(N+3)}} \quad (3)$$

677

678 Therefore, we can analytically estimate the confidence interval of prediction accuracy for
679 each of the score. We selected high-power scores defined as power > 0.95 with P-value <=
680 0.05 or P-value <= 1.9×10^{-5} (0.05/2600) for the combination with Elastic Net.

681

682 To compare the improvement, for instance between PRSmix and the best PGS Catalog, we
683 estimate the mean fold-ratio of R^2 across different traits with its 95% confidence interval and
684 evaluated the significance difference from 1 using a two-tailed paired t-test.

685

686 Simulations

687

688 We used UK Biobank European ancestry to conduct simulations for trait-specific and cross-
689 trait combinations. Overall, we simulated 7 traits with heritability h^2 equal to 0.05, 0.1, 0.2,
690 and 0.5. We randomly selected $M=1000$ causal SNPs among 1.1 million HapMap3 variants
691 with $INFO > 0.6$, $MAF > 0.01$ and P-value Hardy-Weinberg equilibrium $> 10^{-7}$. We removed
692 individuals with PC1 and PC2 > 3 standard deviation from the mean. We randomly remove
693 one in a pair of related individuals with closer than 2nd degree. The genetic components
694 were simulated as PRSs where PRS1, PRS2, and PRS3 are considered trait-specific scores
695 with genetic correlations are 0.8 and 0.4 for cross-trait scores. PRS4, PRS5 and PRS6 are
696 simulated as pleiotropic effects on the outcome traits with genetic correlation equal to 0.4.
697 The SNP effects for PRSs are simulated by a multivariate normal distribution $MVN(0, \Sigma)$
698 where Σ is the covariance matrix between PRSs. The main diagonal contains the heritability
699 of the traits as h^2 / M and the covariance between PRSs are simulated as $r_g * h^2 / M$ where
700 r_g is the genetic correlation between PRSs (0.8 for trait-specific scores and 0.4 for cross-trait
701 scores). The PRSs of the outcome are estimated by the weighted combination of PRS where
702 the weights follow $U(0,1)$. 7 phenotypes were simulated as $y = g + e, e \sim N(0, 1 - h^2)$ where
703 g is PRS and e is the residuals.

704

705 We split the simulated cohort into 3 data sets for: 1) GWAS 2) training set: training the
706 mixing weights with a linear combination and 3) testing set: testing the combined PRS. We
707 incorporated PRS1, PRS2 and PRS3 to assess the trait-specific PRSmix framework. We

708 combined all 6 single PRS to evaluate the cross-trait PRSmix+ framework. We compared
709 the fold-ratio of the R^2 of the combined PRS to the R^2 of best single PRS to assess the
710 improvement of the combination strategy. To evaluate the improvement across different
711 heritabilities, we estimated the slope of improvement per $\log_{10}(N)$ increase of training
712 sample sizes on the fold-ratio of predictive improvement.

713

714 **Sample and genotyping quality control**

715

716 The AoU data version 5 contains more than 700 million variants from whole genome
717 sequencing³⁹. We curated European ancestry by predicted genetic ancestry with a
718 probability > 90% provided by AoU yielding 48,112 individuals in the AoU. For variant quality
719 control beyond AoU central efforts, we further filtered SNPs to include MAF > 0.001 which
720 retained 12,416,130 SNPs. We performed a similar quality control for imputed genotype data
721 for South Asian ancestry in the Genes & Health cohort with additional criteria of INFO score
722 > 0.6 and genotype missing rate < 5%. Individuals with a missing rate > 5% were removed.
723 Eventually, 44,396 individuals and 8,935,207 SNPs remained in Genes & Health.

724

725 **Assessment of clinical utility**

726

727 We applied PRSmix and PRSmix+ for coronary artery disease as a clinical application. The
728 phenotypic algorithm includes at least one ICD or CPT code below: ICD9 410x, 411x, 412x;
729 ICD10 I22x, I23x, I24.1, I25.2 CPT 92920-92979 (PCI), 33533-33536, 33517-33523, 33510-
730 33516 (CABG) or self-reported personal history of MI or CAD. CAD in Genes and Health
731 cohort was defined with at least one ICD10 I22x, I23x, I24.1, I25 or operation codes K401,
732 K402, K403, K404, K411, K451, K452, K453, K454, K455, K491, K492, K499, K502, K751,
733 K752, K753, K754, K758, K759 or SNOMED codes 1755008, 22298006, 54329005,
734 57054005, 65547006, 70211005, 70422006, 73795002, 233838001, 304914007,
735 401303003, 401314000.

736

737 The category-free NRI was used to evaluate the clinical utility. NRI was calculated by adding
738 the PRS to the baseline logistic model including age, sex, the first 10 principal components,
739 and clinical risk factors. The clinical risk factors include total cholesterol, HDL-C, BMI, type 2
740 diabetes, and current smoking status or model includes only age, sex, and 10 principal
741 components. NRI was calculated as the sum of NRI for cases and NRI for controls:

742

$$NRI = P(up|case) - P(down|case) + P(down|control) - P(up|control)$$

743

744 $P(up|case)$ and $P(down|case)$ estimate the proportion of cases that had higher or lower risk
745 after classification with logistic regression, respectively. The confidence interval for NRI was
746 estimated with 500 bootstraps. We also compared the risk increase between individuals in
747 the top decile of PRS versus those remaining in the population. In addition to liability R^2 to
748 compare the PRS performance, we also used the incremental area under the curve (AUC) to
749 compare the PRS. The incremental AUC was estimated as the difference between the AUC
750 of models with the integrative score versus the model with only clinical variables.

751

752 **wMT-SBLUP and linear combination of LDpred2-auto derived scores**

753

754 *LDpred2-auto*: LDpred2 is a Bayesian method that computes the adjusted SNP effect sizes
755 from GWAS summary statistics. LDpred2 utilizes the SNP effect sizes as prior and
756 incorporates LD between markers to infer the posterior effect sizes. In our work, we
757 implemented LDpred2-auto⁵¹ since this method can infer heritability and the proportion of
758 causal variants. LDpred2-auto was conducted with 800 burn-in iterations and 500 iterations.
759 The proportion of causal variants was initialized between 10^{-4} and 0.9. Furthermore,
760 LDpred2-auto does not require a validation set, the SNP effect sizes were averaged
761 between scores. We used 1,138,726 HapMap3 variants that overlapped with SNPs from
762 whole-genome sequencing data in the All of Us cohort. The LD reference panel developed
763 from European ancestry was provided by the LDpred2-tutorial.

764

765 *wMT-SBLUP*: wMT-SBLUP¹⁹ calculates the mixing weights of PRS using sample sizes from
766 GWAS summary statistics, SNP-heritability and genetic correlation. We compared wMT-
767 SBLUP with our method using 5 traits that were originally assessed with wMT-SBLUP
768 including CAD, T2D, depression, height, and BMI. We curated 26 publicly available GWAS
769 summary statistics (Supplementary Table 18) and performed LDpred2-auto with quality
770 controls suggested by Privé et al^{5,51}. We used LD score regression to estimate SNP-
771 heritability and genetic correlation across 26 traits. For each of the 5 outcome traits, we
772 selected correlated traits with P-value of genetic correlation less than 0.05.

773

774 *Elastic Net for linear combination*: we also implemented linear combination by Elastic Net
775 with the LDpred2-auto-derived PRSs for contributing traits since this strategy was proposed
776 by several works^{8,13,20}. We selected scores with significant variance explained (P-
777 value<0.05) to the outcome trait and conducted Elastic Net using the *glmnet* R package⁴⁶.

778

779 **Phenome-wide association study**

780

781 We obtained the list of 1815 phecodes from the PheWAS website (last accessed December
782 2022)⁵². The phecodes were based on ICD-9 and ICD-10 to classify individuals. PheWAS
783 was conducted on European ancestry only in AoU. For each phecodes as the outcome, we
784 conducted an association analysis using logistic regression on PRS and adjusted for age,
785 sex, and first 10 PCs. The significance threshold for PheWAS was estimated as 2.75×10^{-5}
786 (0.05/1815) after Bonferroni correction.

787

788 **Data availability**

789

790 The PGS Catalog is freely available at <https://www.pgscatalog.org/>. Our new scores are
791 deposited in the PGS Catalog. The All of Us and Genes & Health individual-level data is a
792 controlled access dataset and may be granted at <https://www.researchallofus.org/> and
793 <https://www.genesandhealth.org/>, respectively.

794

795 The weights from the PRSmix and PRSmix+ scores in this manuscript have been returned to
796 the PGS Catalog. The R package to implement PRSmix and PRSmix+ in independent
797 datasets is at <https://github.com/buutrg/PRSmix>.

798

799 *Software/analyses*:

800 Analyses were performed on the AoU Researcher Workbench in Jupyter Notebook 14 using
801 R version 4.0.0 programming language. Results are reported in compliance with the AoU
802 Data and Statistics Dissemination Policy.

803

804 **ACKNOWLEDGEMENT**

805

806 We would like to thank Alkes L. Price for critical comments for this works. L.E.H. is
807 supported by the National Human Genome Research Institute (K08HG012221). P.N. is
808 supported by grants from NHGRI (U01HG011719), NHLBI (R01HL142711, R01HL127564,
809 R01HL151152), and Massachusetts General Hospital (Paul & Phyllis Fireman Endowed
810 Chair in Vascular Medicine). The content is solely the responsibility of the authors and does
811 not necessarily represent the official views of the National Institutes of Health.

812

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

823

824 Genes & Health is/has recently been core-funded by Wellcome (WT102627, WT210561),
825 the Medical Research Council (UK) (M009017), Higher Education Funding Council for
826 England Catalyst, Barts Charity (845/1796), Health Data Research UK (for London
827 substantive site), and research delivery support from the NHS National Institute for Health
828 Research Clinical Research Network (North Thames). Genes & Health is/has recently been
829 funded by Alnylam Pharmaceuticals, Genomics PLC; and a Life Sciences Industry
830 Consortium of Bristol-Myers Squibb Company, GlaxoSmithKline Research and Development
831 Limited, Maze Therapeutics Inc, Merck Sharp & Dohme LLC, Novo Nordisk A/S, Pfizer Inc,
832 Takeda Development Centre Americas Inc.

833

834 We thank Social Action for Health, Centre of The Cell, members of our Community Advisory
835 Group, and staff who have recruited and collected data from volunteers. We thank the NIHR
836 National Biosample Centre (UK Biocentre), the Social Genetic & Developmental Psychiatry
837 Centre (King's College London), Wellcome Sanger Institute, and Broad Institute for sample
838 processing, genotyping, sequencing and variant annotation. We thank: Barts Health NHS
839 Trust, NHS Clinical Commissioning Groups (City and Hackney, Waltham Forest, Tower
840 Hamlets, Newham, Redbridge, Havering, Barking and Dagenham), East London NHS
841 Foundation Trust, Bradford Teaching Hospitals NHS Foundation Trust, Public Health
842 England (especially David Wyllie), Discovery Data Service/Endeavour Health Charitable
843 Trust (especially David Stables), NHS Digital - for GDPR-compliant data sharing backed by
844 individual written informed consent.

845

846 Most of all we thank all of the volunteers participating in the *All of Us* Research Program and
847 Genes & Health.

848

849 **CONFLICT OF INTEREST**

850

851 P.N. reports grants from Allelica, Amgen, Apple, Boston Scientific, Genentech, and Novartis,
852 is a consultant to Allelica, Apple, AstraZeneca, Blackstone Life Sciences, Foresite Labs,
853 HeartFlow, Novartis, Genentech, and GV, scientific advisory board membership to Esperion
854 Therapeutics, Preciseli, and TenSixteen Bio, is a scientific co-founder of TenSixteen Bio,
855 and spousal employment at Vertex Pharmaceuticals, all unrelated to the present work.
856 Others declare no conflict of interest.

857 **REFERENCES**

- 858 1. Catalog, P. G. S. PGS Catalog - the Polygenic Score Catalog.
859 <http://www.pgscatalog.org/>.
- 860 2. Choi, S. W., Mak, T. S.-H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk
861 score analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).
- 862 3. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**,
863 e1003348 (2013).
- 864 4. Choi, S. W. & O'Reilly, P. SA20 - PRSice 2: POLYGENIC RISK SCORE SOFTWARE
865 (UPDATED) AND ITS APPLICATION TO CROSS-TRAIT ANALYSES. *Eur.*
866 *Neuropsychopharmacol.* **29**, S832 (2019).
- 867 5. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics*
868 (2020) doi:10.1093/bioinformatics/btaa1029.
- 869 6. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of
870 Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
- 871 7. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via
872 Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
- 873 8. Abraham, G. *et al.* Genomic risk score offers predictive performance comparable to
874 clinical risk factors for ischaemic stroke. *Nat. Commun.* **10**, 5819 (2019).
- 875 9. Chung, W. *et al.* Efficient cross-trait penalized regression increases prediction accuracy
876 in large cohorts using secondary phenotypes. *Nat. Commun.* **10**, 569 (2019).
- 877 10. Weissbrod, O. *et al.* Leveraging fine-mapping and multipopulation training data to
878 improve cross-population polygenic risk scores. *Nat. Genet.* **54**, 450–458 (2022).
- 879 11. Inouye, M. *et al.* Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults:
880 Implications for Primary Prevention. *J. Am. Coll. Cardiol.* **72**, 1883–1893 (2018).
- 881 12. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat.*
882 *Genet.* **54**, 573–580 (2022).
- 883 13. Albiñana, C. *et al.* Multi-PGS enhances polygenic prediction: weighting 937 polygenic
884 scores. Preprint at <https://doi.org/10.1101/2022.09.14.22279940>.

- 885 14. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex
886 traits. *Nat. Genet.* **51**, 1339–1348 (2019).
- 887 15. Li, C., Yang, C., Gelernter, J. & Zhao, H. Improving genetic risk prediction by leveraging
888 pleiotropy. *Hum. Genet.* **133**, 639–650 (2014).
- 889 16. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health
890 disparities. *Nature Genetics* vol. 51 584–591 Preprint at [https://doi.org/10.1038/s41588-](https://doi.org/10.1038/s41588-019-0379-x)
891 019-0379-x (2019).
- 892 17. Lambert, S. A. *et al.* The Polygenic Score Catalog as an open database for
893 reproducibility and systematic evaluation. *Nat. Genet.* **53**, 420–425 (2021).
- 894 18. Buch, G., Schulz, A., Schmidtman, I., Strauch, K. & Wild, P. S. A systematic review
895 and evaluation of statistical methods for group variable selection. *Stat. Med.* **42**, 331–
896 352 (2023).
- 897 19. Maier, R. M. *et al.* Improving genetic prediction by leveraging genetic correlations
898 among human diseases and traits. *Nat. Commun.* **9**, (2018).
- 899 20. Krapohl, E. *et al.* Multi-polygenic score approach to trait prediction. *Mol. Psychiatry* **23**,
900 1368–1374 (2018).
- 901 21. Yengo, L. *et al.* A saturated map of common genetic variants associated with human
902 height. *Nature* **610**, 704–712 (2022).
- 903 22. Klarin, D. & Natarajan, P. Clinical utility of polygenic risk scores for coronary artery
904 disease. *Nat. Rev. Cardiol.* **19**, 291–301 (2022).
- 905 23. Heart Association Council on Epidemiology, A. Heart disease and stroke statistics—
906 2022 update: a report from the American Heart Association. *Circulation* (2022).
- 907 24. Arnett, D. K. *et al.* 2019 ACC/AHA Guideline on the Primary Prevention of
908 Cardiovascular Disease: A Report of the American College of Cardiology/American
909 Heart Association Task Force on Clinical Practice Guidelines. *Circulation* **140**, e596–
910 e646 (2019).
- 911 25. Koyama, S. *et al.* Population-specific and trans-ancestry genome-wide analyses identify
912 distinct and shared genetic risk loci for coronary artery disease. *Nat. Genet.* **52**, 1169–

- 913 1177 (2020).
- 914 26. Tamlander, M. *et al.* Integration of questionnaire-based risk factors improves polygenic
915 risk scores for human coronary heart disease and type 2 diabetes. *Commun Biol* **5**, 158
916 (2022).
- 917 27. Zhang, H. *et al.* Novel methods for multi-ancestry polygenic prediction and their
918 evaluations in 5.1 million individuals of diverse ancestry. *bioRxiv* (2022)
919 doi:10.1101/2022.03.24.485519.
- 920 28. Sud, M. *et al.* Population-Based Recalibration of the Framingham Risk Score and
921 Pooled Cohort Equations. *J. Am. Coll. Cardiol.* **80**, 1330–1342 (2022).
- 922 29. Carroll, R. J., Bastarache, L. & Denny, J. C. R PheWAS: data analysis and plotting tools
923 for phenome-wide association studies in the R environment. *Bioinformatics* **30**, 2375–
924 2376 (2014).
- 925 30. Bastarache, L., Denny, J. C. & Roden, D. M. Phenome-Wide Association Studies. *JAMA*
926 **327**, 75–76 (2022).
- 927 31. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat.*
928 *Soc. Series B Stat. Methodol.* **67**, 301–320 (2005).
- 929 32. Zhou, D.-X. On grouping effect of elastic net. *Stat. Probab. Lett.* **83**, 2108–2112 (2013).
- 930 33. Zou, H. & Hastie, T. Addendum: Regularization and variable selection via the elastic
931 net. *J. R. Stat. Soc. Series B Stat. Methodol.* **67**, 768–768 (2005).
- 932 34. Wang, M. *et al.* Validation of a Genome-Wide Polygenic Score for Coronary Artery
933 Disease in South Asians. *J. Am. Coll. Cardiol.* **76**, 703–714 (2020).
- 934 35. Lloyd-Jones, L. R. *et al.* Improved polygenic prediction by Bayesian multiple regression
935 on summary statistics. *Nat. Commun.* **10**, 5086 (2019).
- 936 36. Wang, Y., Tsoo, K., Kanai, M., Neale, B. M. & Martin, A. R. Challenges and
937 opportunities for developing more generalizable polygenic risk scores. *Annu. Rev.*
938 *Biomed. Data Sci.* **5**, 293–320 (2022).
- 939 37. Mostafavi, H. *et al.* Variable prediction accuracy of polygenic scores within an ancestry
940 group. *Elife* **9**, (2020).

- 941 38. Mapes, B. M. *et al.* Diversity and inclusion for the All of Us research program: A scoping
942 review. *PLoS One* **15**, e0234962 (2020).
- 943 39. The “All of Us” Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019).
- 944 40. Cronin, R. M. *et al.* Development of the initial surveys for the All of Us Research
945 Program. *Epidemiology* **30**, 597–608 (2019).
- 946 41. All of Us Research Program Protocol. *All of Us Research Program | NIH*
947 <https://allofus.nih.gov/about/all-us-research-program-protocol> (2020).
- 948 42. Pereira, F. Home. *Terra.Bio* <https://terra.bio/> (2020).
- 949 43. Researcher Workbench. <https://www.researchallofus.org/workbench/>.
- 950 44. Data Methods – All of Us Research Hub. [https://www.researchallofus.org/data-](https://www.researchallofus.org/data-tools/methods)
951 [tools/methods](https://www.researchallofus.org/data-tools/methods).
- 952 45. Finer, S. *et al.* Cohort Profile: East London Genes & Health (ELGH), a community-
953 based population genomics and health study in British Bangladeshi and British Pakistani
954 people. *International Journal of Epidemiology* vol. 49 20–21i Preprint at
955 <https://doi.org/10.1093/ije/dyz174> (2020).
- 956 46. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear
957 models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
- 958 47. Lee, S. H., Clark, S. & van der Werf, J. H. J. Estimation of genomic prediction accuracy
959 from reference populations with varying degrees of relationship. *PLoS One* **12**,
960 e0189775 (2017).
- 961 48. Wishart, J., Kondo, T. & Elderton, E. M. The mean and second moment coefficient of
962 the multiple correlation coefficient, in samples from a normal population. *Biometrika* **22**,
963 353 (1931).
- 964 49. Stuart, A., Ord, K. & Arnold, S. *Kendall's Advanced Theory of Statistics, Classical*
965 *Inference and the Linear Model*. (Wiley, 2010).
- 966 50. Momin, M. M., Lee, S., Wray, N. R. & Lee, S. H. Significance tests for R² of out-of-
967 sample prediction using polygenic scores. *Am. J. Hum. Genet.* (2023)
968 [doi:10.1016/j.ajhg.2023.01.004](https://doi.org/10.1016/j.ajhg.2023.01.004).

- 969 51. Privé, F., Arbel, J., Aschard, H. & Vilhjálmsson, B. J. Identifying and correcting for
970 misspecifications in GWAS summary statistics and polygenic scores. *HGG Adv.* **3**,
971 100136 (2022).
- 972 52. Denny, J. C. *et al.* PheWAS: demonstrating the feasibility of a phenome-wide scan to
973 discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010).
- 974