ARTICLES
https://doi.org/10.1038/s41588-022-01036-9
nature genetics

Check for updates

Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores

Omer Weissbrod[1,34] ✉, Masahiro Kanai[2,3,34], Huwenbo Shi[1,4,34], Steven Gazal[1,5,6], Wouter J. Peyrot[1,7], Amit V. Khera[2,8], Yukinori Okada[3,9], The Biobank Japan Project*, Alicia R. Martin[2], Hilary K. Finucane[2,10] and Alkes L. Price[1,2] ✉

Polygenic risk scores suffer reduced accuracy in non-European populations, exacerbating health disparities. We propose PolyPred, a method that improves cross-population polygenic risk scores by combining two predictors: a new predictor that leverages functionally informed fine-mapping to estimate causal effects (instead of tagging effects), addressing linkage disequilibrium differences, and BOLT-LMM, a published predictor. When a large training sample is available in the non-European target population, we propose PolyPred+, which further incorporates the non-European training data. We applied PolyPred to 49 diseases/traits in four UK Biobank populations using UK Biobank British training data, and observed relative improvements versus BOLT-LMM ranging from +7% in south Asians to +32% in Africans, consistent with simulations. We applied PolyPred+ to 23 diseases/traits in UK Biobank east Asians using both UK Biobank British and Biobank Japan training data, and observed improvements of +24% versus BOLT-LMM and +12% versus PolyPred. Summary statistics-based analogs of PolyPred and PolyPred+ attained similar improvements.

Polygenic risk scores (PRSs) can identify individuals at elevated risk of complex diseases, providing opportunities for preventive action[1–6]. However, many studies have shown that PRSs based on European training data attain lower accuracy when applied to populations of non-European ancestry[7–26]. This loss of accuracy is primarily driven by linkage disequilibrium (LD) differences[12–15], allele frequency differences (including population-specific SNPs)[13,14,27] and causal effect-size differences[12–14,28–31], although differences in heritability also play a minor role[13,14,32]. PRSs based on non-European training data do not suffer from these limitations, but are currently limited by much smaller training sample sizes[1,12–15,21,33]. The development of new methods to reduce this gap in cross-population PRS accuracy has the potential to ameliorate health disparities[13].

In the present study, we propose PolyPred, which linearly combines two complementary predictors derived from European training data: (1) PolyFun-pred, a new predictor that circumvents LD differences by applying genome-wide, functionally informed fine-mapping[34,35] to precisely estimate causal effects (instead of tagging effects); and (2) BOLT-LMM[36,37], a published predictor that analyzes all loci jointly and can capture all signals in extremely polygenic loci. BOLT-LMM requires individual-level training data. If individual-level training data are not available, we propose two analogous methods: (1) PolyPred-S, which linearly combines PolyFun-pred with SBayesR[38], and (2) PolyPred-P, which linearly combines PolyFun-pred with PRS-CS[39]. Recommendations for when to use PolyPred, PolyPred-S or PolyPred-P are provided below.

In the special case where a large (for example, $n \geq 50{,}000$) non-European training sample exists from the target population (or a closely related population), we propose PolyPred+, a polygenic prediction method that leverages both European and non-European training data. PolyPred+ linearly combines (1) PolyFun-pred, (2) BOLT-LMM and (3) BOLT-LMM-pop, which is obtained by applying BOLT-LMM to the non-European training data, addressing minor allele frequency (MAF) differences and causal effect-size differences. If individual-level training data are not available, we propose the alternative methods PolyPred-S+ and PolyPred-P+, which replace BOLT-LMM with either SBayesR or PRS-CS, respectively. Recommendations for when to use PolyPred+, PolyPred-S+ or PolyPred-P+ are provided below.

We compared PolyPred and PolyPred+ (and their summary statistics-based analogs) with state-of-the-art polygenic prediction methods via simulations and analyses of 49 diseases and complex traits in four populations from the UK Biobank[40], Biobank Japan[41] and Uganda-APCDR[42,43]. We conclude that PolyPred and its summary statistics-based analogs substantially increase cross-population polygenic prediction accuracy, and that PolyPred+ and its summary statistics-based analogs further increase cross-population prediction accuracy in the special case where non-European training data are available in large sample size.

[1]Epidemiology Department, Harvard School of Public Health, Boston, MA, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan. [4]OMNI Bioinformatics, San Francisco, CA, USA. [5]Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [6]Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [7]Department of Psychiatry, Amsterdam UMC, Vrije Universiteit, Amsterdam, the Netherlands. [8]Verve Therapeutics, Cambridge, MA, USA. [9]Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [10]Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. [34]These authors contributed equally: Omer Weissbrod, Masahiro Kanai, Huwenbo Shi. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: oweissbrod@hsph.harvard.edu; aprice@hsph.harvard.edu

Table 1 | Summary of main methods evaluated

Method	Constituent methods	SNP set	Training data	Fine-mapped effect sizes	Summary statistics	Reference
P + T	-	All (18 million)	Eur	No	Yes	45,46
BOLT-LMM	-	HapMap 3 (1.2 million)	Eur	No	No	36,37
SBayesR	-	HapMap 3 (1.2 million)	Eur	No	Yes	38
PRS-CS	-	HapMap 3 (1.2 million)	Eur	No	Yes	39
PolyPred	PolyFun-pred, BOLT-LMM	All (18 million)	Eur	Yes	No	This work
PolyPred-S	PolyFun-pred, SBayesR	All (18 million)	Eur	Yes	Yes	This work
PolyPred-P	PolyFun-pred, PRS-CS	All (18 million)	Eur	Yes	Yes	This work
PolyPred$^+$	PolyFun-pred, BOLT-LMM, BOLT-LMM-pop	All (18 million)	Eur + target pop	Yes	No	This work
PolyPred-S$^+$	PolyFun-pred, SBayesR, SBayesR-pop	All (18 million)	Eur + target pop	Yes	Yes	This work
PolyPred-P$^+$	PolyFun-pred, PRS-CS, PRS-CS-pop	All (18 million)	Eur + target pop	Yes	Yes	This work

For each method, we report its constituent methods (or '-' for individual methods), the set of SNPs analyzed in model training using UK Biobank training data (and its size when restricted to imputed UK Biobank SNPs with European MAF ≥ 0.1% and INFO score ≥ 0.6), the training data analyzed, whether it incorporates fine-mapped effect sizes (as opposed to tagging effect sizes), whether it can work with summary statistics and the corresponding reference. Eur, European; [Method]-pop, Method applied to training data from non-European target population; target pop, non-European target population.

Results

Overview of methods. PolyPred combines two complementary predictors: PolyFun-pred and BOLT-LMM (Table 1 and Fig. 1a). PolyFun-pred is a new predictor that leverages genome-wide, functionally informed fine-mapping[34,35] to estimate posterior mean causal effects (rather than tagging effects; Supplementary Note) for all SNPs with European MAF ≥ 0.1% (18 million SNPs in the present study) by applying PolyFun + SuSiE[35] to European training data across 2,763 overlapping 3-Mb loci. Leveraging fine-mapped posterior mean causal effects for cross-population polygenic prediction aims to address LD differences between populations. BOLT-LMM[36,37] is a published predictor that estimates posterior mean tagging effects of common SNPs (1.2 million HapMap 3 SNPs[44] in the present study) using European individual-level training data. Combining PolyFun-pred with BOLT-LMM is advantageous because they have complementary advantages: PolyFun-pred estimates causal effects rather than tagging effects. BOLT-LMM estimates tagging effects, but it analyzes all loci jointly and can potentially capture all signals in extremely polygenic loci (Methods).

In the special case where a large training sample is available in the target population (or a closely related population), we propose PolyPred$^+$, which combines three complementary predictors: PolyFun-pred, BOLT-LMM and BOLT-LMM-pop (Table 1 and Fig. 1b); BOLT-LMM-pop refers to application of BOLT-LMM to common SNPs (1.2 million HapMap 3 SNPs in the present study) using training data from the non-European target population, addressing both MAF and causal effect-size differences.

PolyPred computes linear combinations of the estimated effect sizes of their constituent predictors:

$$\widehat{\beta}_i^{\mathrm{PolyPred}(+)} = \sum_j w^j \widehat{\beta}_i^j, \quad (1)$$

where i indexes SNPs, j indexes the constituent predictors (PolyFun-pred and BOLT-LMM for PolyPred; PolyFun-pred, BOLT-LMM and BOLT-LMM-pop for PolyPred$^+$), $\widehat{\beta}_i^{\mathrm{PolyPred}(+)}$ is the PolyPred$^+$ per-allele effect size of SNP i, w^j is the method-specific weight and $\widehat{\beta}_i^j$ is the per-allele effect size of SNP i for method j (or 0 if SNP i was not considered by method j). Predicted phenotypes are computed by applying effect sizes to target genotypes:

$$\hat{y} = \sum_i x_i \widehat{\beta}_i^{\mathrm{PolyPred}(+)}, \quad (2)$$

where $\hat{y}$ is the predicted phenotype of an individual from the target population and x_i is the number of minor alleles of SNP i carried by the individual. The mixing weights w^j in equation (1) are estimated via non-negative least squares regression using a small number of training individuals from the target population (500 in the present study), regressing true phenotypes on a linear combination of the constituent predictors (which are computed as in equation (2)).

PolyPred requires individual-level training data for its BOLT-LMM component. If only summary statistics (and summary LD information) are available, we propose two analogous methods (Table 1): (1) PolyPred-S, which linearly combines PolyFun-pred and SBayesR[38]; and (2) PolyPred-P, which linearly combines PolyFun-pred and PRS-CS[39]. We also propose the analogous methods PolyPred-P$^+$ and PolyPred-S$^+$ (Table 1). Further details of PolyPred and PolyPred$^+$ (and their summary statistics-based analogs) are provided in Methods; we have publicly released open-source software implementing these methods (Code availability).

We evaluate prediction accuracy for each method and target population using relative R^2, defined as the R^2 obtained in the target non-European population (after correcting for covariates and potential confounders; Methods) divided by the R^2 obtained by BOLT-LMM in UK Biobank non-British Europeans (employing the same correction), using the same training data for the numerator and the denominator. This quotient transforms the prediction accuracies from an absolute scale to a scale of relative improvement (versus the BOLT-LMM predictor in the UK Biobank non-British European target population), which is invariant to factors such as training sample size and trait heritability. For disease traits, we additionally evaluated the area under the receiving operating characteristic. We provide further details in Methods. We compare PolyPred and PolyPred$^+$ (and their summary statistics-based analogs) with four published methods: LD pruning + P-value thresholding (P + T)[45,46], BOLT-LMM[36,37], SBayesR[38] and PRS-CS[39] (Table 1).

Our recommendation for which version of PolyPred to use (Table 1) depends on three factors: (1) whether individual-level training data are available; (2) the size and consistency of matched ancestry of the LD reference panel (if individual-level training data are not available); and (3) whether non-European training data are available. Our results for the underlying constituent methods are summarized in Table 2 and our recommendations are summarized in Fig. 2.

Simulations with in-sample LD. We compared PolyPred, PolyPred-S and PolyPred-P with P + T, BOLT-LMM, SBayesR and

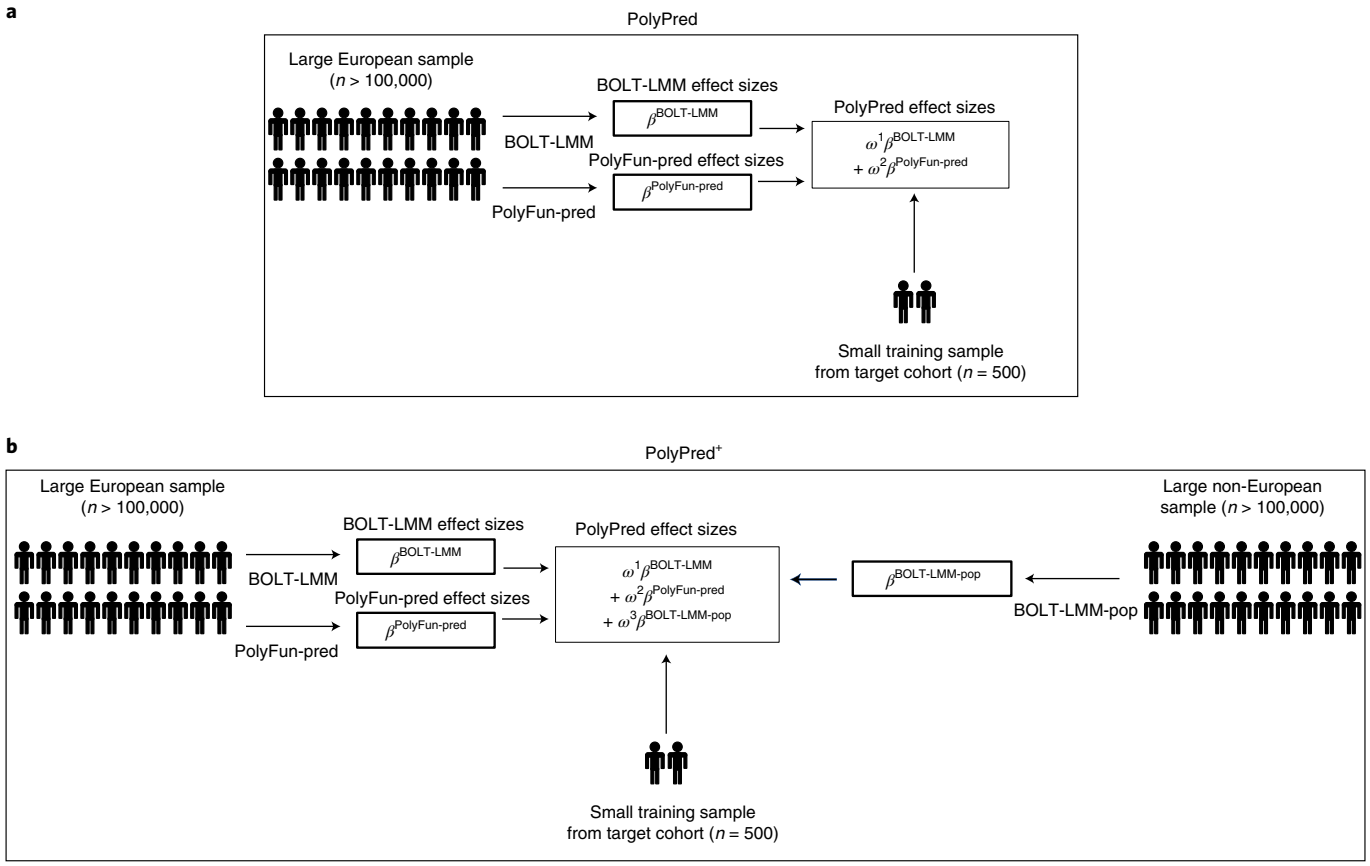


Fig. 1 | Overview of PolyPred and PolyPred+. a, Overview of PolyPred. PolyPred linearly combines the effect sizes of BOLT-LMM ($\beta^{\text{BOLT-LMM}}$) and PolyFun-pred ($\beta^{\text{PolyFun-pred}}$) (trained using European training data). It uses a small training sample from the target population to estimate mixing weights (ω^1, ω^2) for the constituent methods. **b**, Overview of PolyPred+. PolyPred+ linearly combines the effect sizes of BOLT-LMM ($\beta^{\text{BOLT-LMM}}$), PolyFun-pred ($\beta^{\text{PolyFun-pred}}$) (trained using European training data) and BOLT-LMM-pop ($\beta^{\text{BOLT-LMM-pop}}$) (trained using non-European training data from the target population). It uses a small training sample from the target population to estimate mixing weights (ω^1, ω^2, ω^3) for the constituent methods. PolyPred-S and PolyPred-P (respectively, PolyPred-S+ and PolyPred-P+) replace all instances of BOLT-LMM with SBayesR or PRS-CS, respectively.

PRS-CS via simulations, using real genotypes or in-sample LD from the UK Biobank[40]. We trained each method using 337,491 unrelated British-ancestry individuals[40] and computed predictions in four target populations: non-British Europeans, south Asians, east Asians and Africans. We estimated mixing weights for PolyPred, PolyPred-S and PolyPred-P using 500 individuals from the target population. We evaluated prediction accuracy using held-out individuals from each target population that were not included in the training sets: 42,000 non-British Europeans, 7,700 south Asians, 900 east Asians and 6,200 Africans. We computed PRS using 250,963 MAFs ≥ 0.1% SNPs with INFO score ≥ 0.6 on chromosome 22.

Generative trait architectures were specified as follows: we simulated traits with polygenicity (genome-wide proportion of causal SNPs) either 0.1% (less polygenic) or 0.3% (more polygenic) and heritability = 5%. We specified prior causal probabilities for each SNP in proportion to per-SNP heritabilities, which we generated for each SNP based on its British LD, MAF and functional annotations, using the baseline-LF model[47]. For each causal SNP, we sampled ancestry-specific causal effect sizes from a multivariate normal distribution assuming cross-population genetic correlations of 0.8 (refs. [13,30]). Other parameter settings were explored in secondary analyses (see below).

We computed relative R^2 for each method, target population and trait architecture, averaged across 100 simulations. In addition to the simulations with in-sample LD described below, we also performed simulations with reference panel LD (Supplementary Note; see also Table 2). Further details of the simulation framework are provided in Methods.

The simulation results are reported in Fig. 3 and Supplementary Table 1 (see also Table 2). PolyPred was the most accurate method in each target population, with relative improvements versus BOLT-LMM (respectively P values for improvement) ranging from +13% in non-British Europeans ($P < 10^{-16}$) to +65% in Africans ($P < 10^{-16}$) for the less polygenic architecture, and from +2% in non-British Europeans ($P = 0.0001$) to +17% in Africans ($P = 10^{-8}$) for the more polygenic architecture. PolyPred-S and PolyPred-P performed slightly worse than PolyPred, but were substantially and significantly more accurate than their corresponding constituent methods. Among the remaining methods, BOLT-LMM was consistently the most accurate and P + T the least accurate method, far underperforming the other methods (despite its widespread recent use[11,13–18,23,31,48–52]). We note that the higher accuracy of BOLT-LMM versus SBayesR and PRS-CS does not imply that BOLT-LMM is a superior method, because BOLT-LMM analyzes individual-level training data whereas SBayesR and PRS-CS analyze summary statistics.

We additionally performed many secondary analyses to investigate the sensitivity of the results to the simulation parameters, the SNP set and the functional annotations, and to evaluate the computational cost and memory cost of each method (Supplementary Note and Supplementary Tables 1 and 2).

Table 2 | Summary of the relative performance of constituent PRS methods

LD	BOLT -LMM	SBayesR	PRS-CS	Figure(s)/Table(s)
Individual-level data (UKB, $n=337{,}000$)	✓✓	✓	✓	Figs. 3, 4 and 6
In-sample LD (UKB, $n=337{,}000$)	-	✓✓	✓	Figs. 3, 4 and 6
Very large unmatched LD (UKB, $n=337{,}000$)	-	✓	✓✓	Extended Data Fig. 1
Small unmatched LD (1000G, $n=489$)	-	✗	✓✓*	Supplementary Tables 4–6

For each of three constituent PRS methods (BOLT-LMM, SBayesR and PRS-CS), we report its relative performance in prediction in UK Biobank (UKB) non-British Europeans under various settings; we also provide links to the corresponding figure(s)/table(s). ✓✓, the method is significantly more accurate than the second-best method in the same row and combining this method with PolyFun-pred increases prediction accuracy; ✓✓*, the method is significantly more accurate than the second-best method in the same row and combining this method with PolyFun-pred does not increase prediction accuracy; ✓, the method is significantly less accurate than the best method in the same row, but is significantly more accurate than P + T; ✗, the method is not significantly more accurate than P + T; -, the method is not applicable, because it requires individual-level data. For individual-level data, the difference between BOLT-LMM and the second-best method was significant in simulations but non-significant in real-trait analyses. For In-sample LD, the difference between SBayesR and PRS-CS was significant in simulations but non-significant in real-trait analyses. For very large unmatched LD (a probable scenario when analyzing summary statistics from a meta-analysis of many cohorts), we performed real-trait analyses only, because simulations would have required another very large individual-level dataset in addition to UK Biobank (Supplementary Note). For small unmatched LD, we performed both simulations and real-trait analyses but report results of real-trait analyses, which we believe to be most reflective of real-life settings (in simulations, SBayesR was significantly more accurate than PRS-CS; Supplementary Note). 1000G, 1000 Genomes project. Results for non-European target populations from UK Biobank were similar, although some of the differences were not statistically significant due to smaller prediction accuracies and sample sizes. We have facilitated the use of very large LD reference panels for European training data by publicly releasing summary LD information for $n=337{,}000$ British-ancestry UK Biobank samples across 18 million SNPs (Data availability).

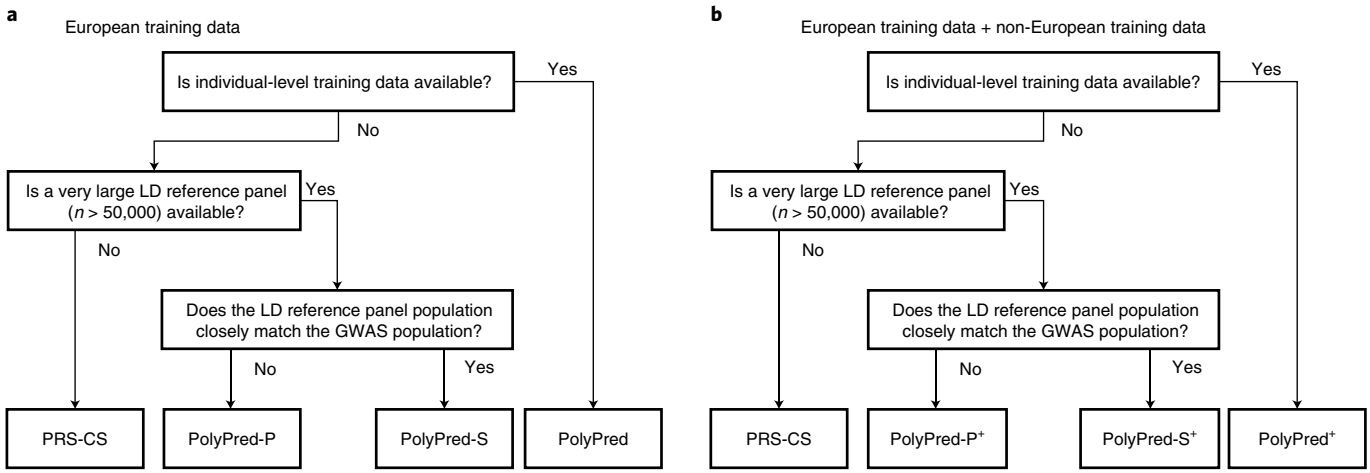


Fig. 2 | Recommendations for the application of PolyPred, PolyPred+ and related methods. a, Flowchart of recommendations when only European training data are available. **b**, Flowchart of recommendations when both European and non-European training data are available. We note that, when working with summary statistics from a meta-analysis of many cohorts, there is typically no LD reference panel that closely matches the GWAS population. Also, it is possible that the answers to the flowchart questions are different for European versus non-European training data, in which case the recommendation would be to use a hybrid method based on the answers to each flowchart in turn (for example, PolyFun-pred + BOLT-LMM + PRS-CS-pop; not listed in Table 1). For both **a** and **b**, we recommend training PolyFun-pred using a very large LD reference panel (for example, $n=337{,}000$ UK Biobank British) with a dense SNP set (for example, 8 million SNPs). We have facilitated this by publicly releasing summary LD information for $n=337{,}000$ British-ancestry UK Biobank samples across 18 million SNPs (Data availability).

We conclude that PolyPred and its summary statistics-based analogs are more accurate than BOLT-LMM, SBayesR, PRS-CS and P + T, with small but significant improvements versus BOLT-LMM in Europeans and substantial improvements in Africans.

PRS in four UK Biobank populations using British training data. We applied PolyPred and its summary statistics-based analogs to 49 diseases and complex traits from the UK Biobank, analyzing four target populations (Methods and Supplementary Table 3). As in our simulations, we used UK Biobank British training data (average $n=325{,}000$) to estimate SNP effect sizes, used 500 additional individuals from the target population to estimate mixing weights, evaluated prediction accuracy using individuals from each of the four target populations that were not included in the training data—42,000 non-British Europeans, 7,700 south Asians, 900 east Asians and 6,200 Africans—and compared PolyPred and its summary statistics-based analogs to P + T, BOLT-LMM, SBayesR and PRS-CS. We meta-analyzed relative R^2 across traits by restricting to seven well-powered, independent complex traits from the UK Biobank[40] ($|r_g| < 0.3$; Methods and Supplementary Table 3) that were also available in Biobank Japan and in Uganda-APCDR (see below). We have publicly released SNP effect sizes used for prediction for each of the four methods (Data availability).

We computed relative-R^2 for each method and target population, and the results are summarized in Fig. 4 and provided in Supplementary Tables 4–6 (see also Table 2). Among the published methods, BOLT-LMM attained the highest prediction accuracy in all target populations (differences between BOLT-LMM and SBayesR were small and not statistically significant). P + T was much less accurate than the other methods (despite its widespread recent use[11,13–18,23,31,48–52]), suffering relative losses of 37–50% versus BOLT-LMM. We thus used BOLT-LMM as a benchmark.

Across all seven methods, PolyPred attained the highest prediction accuracy in each target population. Improvements in average relative R^2 of PolyPred versus BOLT-LMM were equal to +7.5% in non-British Europeans ($P=0.05$), +6.8% in south Asians ($P=0.02$),

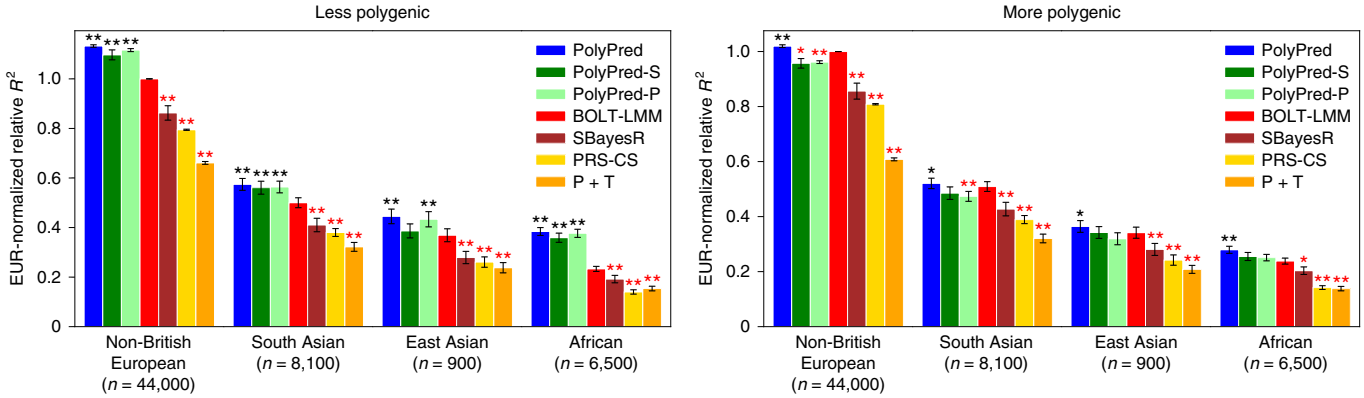


Fig. 3 | Cross-population PRS results for simulated UK Biobank traits using in-sample LD. We report average prediction accuracy (relative R^2; see text) for PRSs trained in UK Biobank British samples (n = 337,000) and applied to four UK Biobank target populations across 100 simulated traits with less polygenic (0.1% of SNPs causal; left panel) or more polygenic (0.3% of SNPs causal; right panel) architectures. Target population sample sizes are indicated in parentheses; PolyPred and its summary statistics-based analogs used 500 additional training samples from each target population to estimate mixing weights. Asterisks above each bar denote statistical significance of the difference versus BOLT-LMM, with black asterisks denoting an advantage and red asterisks a disadvantage ($^*P < 0.05$; $^{**}P < 0.001$). P values were computed using a two-sided Wald's test and were not adjusted for multiple comparisons. Error bars denote s.e. Numerical results, absolute prediction accuracies (R^2) and P values of relative improvements versus BOLT-LMM are reported in Supplementary Table 1.

+11% in east Asians (P = 0.12) and +32% in Africans (P = 0.02). The larger improvement in Africans reflects the larger LD differences versus British training data, due to earlier divergence times[13,14,53]. The lack of statistical significance in east Asians reflects the low power to detect significant differences in very small target samples. PolyPred-S and PolyPred-P were consistently the second and third most accurate methods, respectively, with statistically significant improvements versus their constituent methods. We additionally verified that PolyPred was well calibrated (that is, regressing the true phenotype on the predicted phenotype yielded a slope of 1) in all target populations, whereas the alternative methods were not always well calibrated (Supplementary Tables 4–6 and Supplementary Note). Despite the improvements attained by PolyPred, the reductions in prediction accuracy in non-European populations remained significant ($P < 0.002$), with meta-analyzed absolute $R^2 = 0.17$ in non-British Europeans, 0.11 in south Asians. 0.093 in east Asians and 0.053 in Africans (Methods and Supplementary Tables 4 and 5).

As a secondary analysis, we meta-analyzed the results of each method across three independent diseases: type 2 diabetes, asthma and all autoimmune diseases (Methods); these diseases were not included in our primary meta-analyses due to low heritabilities. PolyPred attained the highest prediction accuracy for each target population and each disease, except for east Asians (where the s.e. was large due to the small sample size) and for type 2 diabetes in non-British Europeans (where BOLT-LMM performed slightly but non-significantly better) (Supplementary Table 4). We performed additional secondary analyses to evaluate the impact of the LD reference panel and the SNP set on prediction accuracy, to evaluate additional methods, and to evaluate the results when modifying the parameters of PolyPred and the other evaluated methods (Supplementary Note and Supplementary Tables 4–7).

We conclude that PolyPred and its summary statistics-based analogs substantially increase cross-population polygenic prediction accuracy versus published methods (with a particularly large improvement in Africans), consistent with simulations. However, there remains a large gap in cross-population polygenic prediction accuracy compared with Europeans.

PRS using ENGAGE meta-analysis training data. We sought to analyze training data consisting of summary statistics for real traits from a meta-analysis of many European cohorts, for which in-sample LD is generally not available. We analyzed 8.1 million meta-analyzed summary statistics from the European Network for Genetic and Genomic Epidemiology (ENGAGE) consortium[54–56] for four traits (body mass index (BMI), waist:hip ratio (adjusted for BMI), total cholesterol and triglycerides; average n = 61,365), and evaluated the prediction accuracy using the same four UK Biobank populations analyzed previously. For each method, we used an LD reference panel based on UK Biobank British individuals; we emphasize that, unlike the other primary analyses, the LD reference panel was mis-specified, because it was not based on in-sample LD.

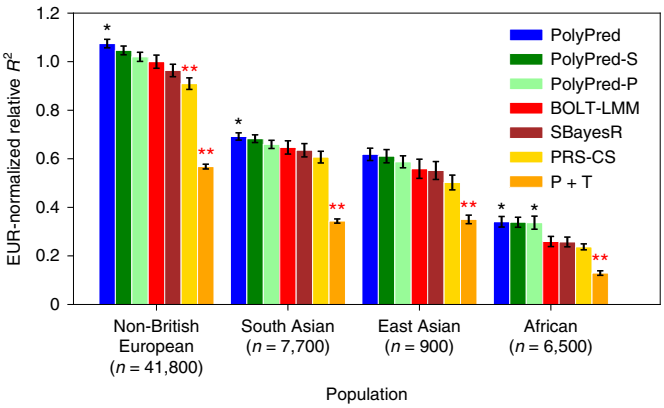


Fig. 4 | Cross-population PRS results for real UK Biobank traits. We report average prediction accuracy (relative R^2; see text), meta-analyzed across seven well-powered, independent traits, for PRSs trained in UK Biobank British samples (average n = 325,000) and applied to four UK Biobank target populations. Target population sample sizes are indicated in parentheses; PolyPred and its summary statistics-based analogs used 500 additional training samples from each target population to estimate mixing weights. Asterisks above each bar denote statistical significance of the difference versus BOLT-LMM, with black asterisks denoting an advantage and red asterisk a disadvantage ($^*P < 0.05$; $^{**}P < 0.001$). P values were computed using a two-sided Wald's test and were not adjusted for multiple comparisons. Error bars denote s.e. Numerical results, results for all 49 traits analyzed, absolute prediction accuracies (R^2) and P values of relative improvements versus BOLT-LMM are reported in Supplementary Tables 4–6.

We excluded methods that require individual-level training data (BOLT-LMM and PolyPred) from this analysis.

The results are summarized in Extended Data Fig. 1 and reported in Supplementary Tables 5 and 8 (see also Table 2). Briefly, PolyPred-P was generally the most accurate method and PRS-CS outperformed SBayesR (with a significant improvement for non-British Europeans and Africans), consistent with a previous study[57] (unlike our analysis of UK Biobank training data, where SBayesR outperformed PRS-CS; Fig. 4). However, differences between similarly performing methods were generally not statistically significant (due to a moderately large s.e.) and thus caution should be exercised in their interpretation; for this reason, we did not perform secondary analyses to further assess differences between methods.

We conclude that PolyPred-P can increase cross-population polygenic prediction accuracy versus published methods when analyzing summary statistics from a meta-analysis of many cohorts.

PRS in Biobank Japan and Uganda-APCDR cohorts. We applied PolyPred and its summary statistics-based analogs to predict 23 diseases and complex traits in Biobank Japan[41] and seven complex traits in Uganda-APCDR, an African-ancestry cohort[42,43] (Methods and Supplementary Table 3). We performed these experiments to avoid training effect sizes and testing predictions in the same cohort, which may produce inflated prediction accuracies[33,58–60]. We again used UK Biobank British training data (average $n = 325{,}000$) to estimate SNP effect sizes and used 500 individuals from the target population to estimate mixing weights. We evaluated prediction accuracy using individuals from each of the two target cohorts that were not included in the training data: 5,000 Biobank Japan individuals and 1,300 Uganda-APCDR individuals. We again compared PolyPred and its summary statistics-based analogs with P + T, BOLT-LMM, SBayesR and PRS-CS. We meta-analyzed relative R^2 across the same seven well-powered, independent complex traits used in the UK Biobank analyses (Supplementary Table 3).

The results are summarized in Fig. 5 and reported in Supplementary Tables 5 and 9. Among the published methods, we again observed that BOLT-LMM attained the highest prediction accuracy in each target population, and that P + T was substantially less accurate than the other methods. Across all seven methods, PolyPred attained the highest prediction accuracy in Biobank Japan and PolyPred-P attained the highest prediction accuracy in Uganda-APCDR (although the difference between PolyPred and PolyPred-P in Uganda-APCDR was not statistically significant). Improvements of PolyPred versus BOLT-LMM in average relative-R^2 = +13% in Biobank Japan ($P = 2 \times 10^{-6}$) and +22% in Uganda-APCDR ($P = 0.26$), similar to our UK Biobank results above. We observed similar improvements for PolyPred-S versus SBayesR and PolyPred-P versus PRS-CS (both of which were statistically significant in Biobank Japan). Prediction accuracy for each method was much smaller in Biobank Japan and Uganda-APCDR (for example, 0.32 and 0.11 for PolyPred; Fig. 5) than in UK Biobank east Asians and UK Biobank Africans (0.62 and 0.34; Fig. 4), probably due to higher SNP heritabilities in the UK Biobank (see below). We also applied PolyPred$^+$ and its summary statistics-based analogs to Biobank Japan, incorporating additional Biobank Japan training data (average $n = 124{,}000$), with the caveat that this analysis involved training and testing in the same cohort (Methods). PolyPred$^+$ attained increased prediction accuracy, with a further +23% improvement versus PolyPred ($P = 0.0004$), with similar results for PolyPred-S$^+$ and PolyPred-P$^+$ (Supplementary Tables 5 and 9).

We performed additional experiments to investigate the above result of decreased prediction accuracy in Biobank Japan versus UK Biobank east Asians. We matched the BOLT-LMM British training sample size to the Biobank Japan training sample size and obtained a relative R^2 in UK Biobank non-British Europeans (using UK Biobank British training samples) +108% larger than in Biobank Japan (using Biobank Japan training samples), consistent with the +104% increase expected from theory[61,62] based on the +67% higher SNP heritabilities in the UK Biobank (Supplementary Table 10 and Supplementary Note). This suggests that differences in SNP heritability due to ancestry or cohort differences may explain most of the differences in prediction accuracies observed between the UK Biobank and Biobank Japan. Further experiments and interpretation are provided in Supplementary Note. We performed six additional secondary analyses to evaluate the sensitivity of the results to various factors (Supplementary Note and Supplementary Tables 5 and 9).

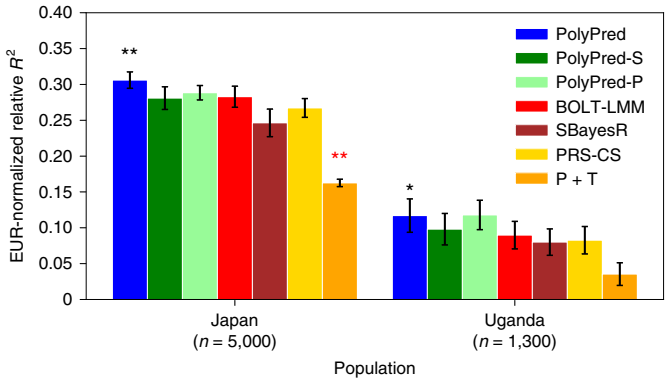


Fig. 5 | Cross-population PRS results for Biobank Japan and Uganda-APCDR traits. We report average prediction accuracy (relative R^2; see text), meta-analyzed across seven well-powered, independent traits, for PRSs trained in UK Biobank British samples (average $n = 325{,}000$) and applied to Biobank Japan and Uganda-APCDR target populations. Target population sample sizes are indicated in parentheses; PolyPred and its summary statistics-based analogs used 500 additional training samples from each target population to estimate mixing weights. Asterisks above each bar denote statistical significance of the difference versus BOLT-LMM, with black asterisks denoting an advantage and red asterisks a disadvantage ($^*P < 0.05$; $^{**}P < 0.001$). *P* values were computed using a two-sided Wald's test and were not adjusted for multiple comparisons. Error bars denote the s.e. Numerical results, results for all 23 traits analyzed, absolute prediction accuracies (R^2) and *P* values of relative improvements versus BOLT-LMM are reported in Supplementary Table 9.

We conclude that PolyPred and its summary statistics-based analogs substantially increase cross-population polygenic prediction accuracy versus published methods when applied to target cohorts different from the training cohort.

PRSs in east Asians using British and Japanese training data. We applied PolyPred$^+$ and its summary statistics-based analogs to predict 23 diseases and complex traits in UK Biobank east Asians using UK Biobank British and Biobank Japan training data (Supplementary Table 3). We performed this experiment to explore the special case where non-European training data are available in large sample size from a population that is genetically similar to the target population, in a cohort that is distinct from the target cohort (previous studies considered only European training data or analyzed non-European training data from the target cohort[11,13–17]). We note that this experiment is still imperfect in that the European training data and non-European target data are from the same cohort (UK Biobank); however, we believe that cohort effects would deflate rather than inflate the relative improvement of PolyPred$^+$ versus other methods, because they would confer an advantage on the European training data but not the non-European training data. We used UK Biobank British training data (average $n = 325{,}000$) and Biobank Japan training data (average $n = 124{,}000$) to estimate SNP effect sizes. We again used 500 individuals from the target population to estimate mixing weights and evaluated prediction

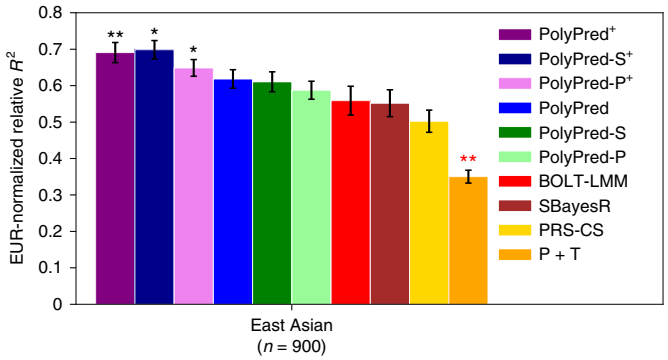


Fig. 6 | Cross-population PRS results for UK Biobank east Asians when incorporating both European and non-European training data. We report average prediction accuracy (relative R^2; see text), meta-analyzed across seven well-powered, independent traits, for PRSs trained in UK Biobank British (average $n = 325{,}000$) and Biobank Japan samples (average $n = 124{,}000$; used by PolyPred+ and its summary statistics-based analogs only) and applied to UK Biobank east Asians. The target population sample size is indicated in parentheses; PolyPred, PolyPred+ and their summary statistics-based analogs used 500 additional training samples from the target population to estimate mixing weights. Asterisks above each bar denote statistical significance of the difference versus BOLT-LMM, with black asterisks denoting an advantage and red asterisks a disadvantage ($^*P < 0.05$; $^{**}P < 0.001$). *P* values were computed using a two-sided Wald's test and were not adjusted for multiple comparisons. Error bars denote the s.e. Numerical results, results for all 23 traits analyzed, absolute prediction accuracies (R^2) and *P* values of relative improvements versus BOLT-LMM are reported in Supplementary Tables 4–6.

accuracy using 900 UK Biobank east Asians who were not included in the training data. We compared PolyPred, PolyPred+ and their summary statistics-based analogs to P + T, BOLT-LMM, SBayesR and PRS-CS (Methods). We meta-analyzed relative R^2 across the same seven well-powered, independent complex traits used in the previous analyses (Supplementary Table 3).

The results are summarized in Fig. 6 and reported in Supplementary Tables 4–6. PolyPred+ attained the highest prediction accuracy, with a +24% improvement versus BOLT-LMM ($P = 0.0009$) and a +12% improvement versus PolyPred ($P = 0.0014$). This implies that incorporating non-European training data can provide a substantial advantage, if it is available in large sample size. Results for PolyPred-S+ (versus SBayesR and PolyPred-S) and PolyPred-P+ (versus PRS-CS and PolyPred-P) were similar. We emphasize that the +12% improvement for PolyPred+ versus PolyPred should be viewed as a lower boundary on the improvement that would be obtained in settings without cohort effects that may confer an advantage on the European training data. We performed additional secondary analyses to evaluate the sensitivity of the results to various factors (Supplementary Note and Supplementary Tables 4–6).

We conclude that PolyPred+ and its summary statistics-based analogs further increase cross-population prediction accuracy in the special case where non-European training data from the target population (or a closely related population) are available in large sample sizes. We emphasize that efforts to assess the benefit of incorporating non-European training data should analyze non-European training data from a cohort that is distinct from the target cohort, otherwise results may be inflated due to cohort effects.

Discussion

We have introduced PolyPred, which improves cross-population polygenic risk prediction by incorporating causal effects in addition to tagging effects, addressing cross-population LD differences. Across seven well-powered independent traits, PolyPred significantly increased prediction accuracy over BOLT-LMM by 32% in UK Biobank Africans and by 13% in Biobank Japan (with similar results versus SBayesR and PRS-CS). In the special case where a large training sample is available in the non-European target population (or a closely related population), we have introduced PolyPred+, which further incorporates the non-European training data, addressing MAF differences and causal effect-size differences. PolyPred+ significantly increased prediction accuracy in UK Biobank east Asians over BOLT-LMM by 24% (and over PolyPred by 12%). PolyPred and PolyPred+ require individual-level training data (for their BOLT-LMM component), but we have also introduced summary statistics-based analogs of PolyPred and PolyPred+ in cases where individual-level training data are not available; specific recommendations are provided in Fig. 2 (see also Table 2). In conclusion, PolyPred and its summary statistics-based analogs substantially improve cross-population polygenic prediction accuracy, ameliorating health disparities[13]. We have publicly released the PRS coefficients for all SNPs and traits analyzed under all evaluated methods (Data availability).

Although we substantially improved cross-population PRS accuracy over the state of the art, prediction accuracy in non-Europeans is still substantially lower compared with Europeans, even within the UK Biobank. There are two reasons for the remaining accuracy gap. First, European sample sizes are still limited, which limits the ability of PolyFun-pred to estimate causal rather than tagging effects. Second, non-European sample sizes are limited, which limits the ability of BOLT-LMM applied to non-European samples to estimate tagging effects. Even with an infinite European training sample, which allows estimating causal effects perfectly (thus addressing LD differences), prediction accuracy could still be higher for Europeans versus non-Europeans due to cross-population genetic correlations <1 (refs. [13,30,63,64]) and different allele frequencies (including population-specific SNPs) (Supplementary Note). Hence our theory and results confirm that larger non-European genome-wide association studies (GWASs) are the best way to further improve PRS accuracy in non-European populations[9,10,12,13,21].

Our work had several limitations, providing opportunities for future work. First, we did not evaluate a setting where the British training data, the non-British training data and the target population were sampled from three different cohorts. Second, PolyPred requires a large number of imputed SNPs (for example, 8.1 million SNPs in the ENGAGE analysis) to perform fine-mapping, motivating the need for large cross-population imputation panels. Third, it could be possible to improve PRS accuracy for admixed individuals by using European effect sizes for European alleles and non-European effect sizes for non-European alleles[16,17]. Fourth, PolyPred and its summary statistics-based analogs were slower than alternative PRS methods (Supplementary Note). Fifth, PolyPred cannot use data from a fixed-effects meta-analysis of GWAS data of different populations (Supplementary Note). Sixth, PolyPred requires a small training sample from the target cohort to maintain calibrated predictions (Supplementary Note). Finally, PolyPred prediction accuracy could in principle be improved if it were possible to decompose its constituent predictors into shared and nonshared components (Supplementary Note). Despite all these limitations, PolyPred and PolyPred+ and their summary statistics-based analogs provide a clear improvement for cross-population polygenic risk prediction.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of

data and code availability are available at https://doi.org/10.1038/s41588-022-01036-9.

Received: 10 January 2021; Accepted: 25 February 2022;
Published online: 7 April 2022

References

1. Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **17**, 392–406 (2016).
2. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
3. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590 (2018).
4. Khera, A. V. et al. Polygenic prediction of weight and obesity trajectories from birth to adulthood. *Cell* **177**, 587–596 (2019).
5. Mavaddat, N. et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am. J. Hum. Genet.* **104**, 21–34 (2019).
6. Li, R., Chen, Y., Ritchie, M. D. & Moore, J. H. Electronic health records and polygenic risk scores for predicting disease risk. *Nat. Rev. Genet.* **21**, 493–502 (2020).
7. Márquez-Luna, C., Loh, P.-R. & South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium & Price, A. L. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet. Epidemiol.* **41**, 811–823 (2017).
8. Grinde, K. E. et al. Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genet. Epidemiol.* **43**, 50–62 (2019).
9. Peterson, R. E. et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell* **179**, 589–603 (2019).
10. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The missing diversity in human genetic studies. *Cell* **177**, 26–31 (2019).
11. Duncan, L. et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, 3328 (2019).
12. Gurdasani, D., Barroso, I., Zeggini, E. & Sandhu, M. S. Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet.* **20**, 520–535 (2019).
13. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
14. Wang, Y. et al. Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nat. Commun.* **11**, 3865 (2020).
15. Amariuta, T. et al. Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. *Nat. Genet.* **52**, 1346–1354 (2020).
16. Marnetto, D. et al. Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. *Nat. Commun.* **11**, 1628 (2020).
17. Bitarello, B. D. & Mathieson, I. Polygenic scores for height in admixed populations. *G3* **10**, 4027–4036 (2020).
18. Chen, M.-H. et al. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* **182**, 1198–1213 (2020).
19. Mahajan, A. et al. Trans-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Preprint at *medRxiv* https://www.medrxiv.org/content/10.1101/2020.09.22.20198937v1 (2020).
20. Cavazos, T. B. & Witte, J. S. Inclusion of variants discovered from diverse populations improves polygenic risk score transferability. *Hum. Genet. Genom. Adv.* **2**, 100017 (2021).
21. Mills, M. C. & Rahal, C. The GWAS diversity monitor tracks diversity by disease in real time. *Nat. Genet.* **52**, 242–243 (2020).
22. Lehmann, B. C., Mackintosh, M., McVean, G. & Holmes, C. C. High trait variability in optimal polygenic prediction strategy within multiple-ancestry cohorts. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/2021.01.15.426781v2 (2021).
23. Ji, Y. et al. Incorporating European GWAS findings improve polygenic risk prediction accuracy of breast cancer among East Asians. *Genet. Epidemiol.* https://doi.org/10.1002/gepi.22382 (2021).
24. Ruan, Y. et al. Improving polygenic prediction in ancestrally diverse populations. Preprint at *medRxiv* https://www.medrxiv.org/content/10.1101/2020.12.27.20248738v2 (2020).
25. Cai, M. et al. A unified framework for cross-population trait prediction by leveraging the genetic correlation of polygenic traits. *Am. J. Hum. Genet.* https://doi.org/10.1016/j.ajhg.2021.03.002 (2021).
26. Huang, Q. Q. et al. Transferability of genetic loci and polygenic scores for cardiometabolic traits in British Pakistanis and Bangladeshis. Preprint at *medRxiv* https://www.medrxiv.org/content/10.1101/2020.12.27.20248738v2 (2021).
27. Durvasula, A. & Lohmueller, K. E. Negative selection on complex traits limits phenotype prediction accuracy between populations. *Am. J. Hum. Genet.* **108**, 620–631 (2021).
28. Coram, M. A., Fang, H., Candille, S. I., Assimes, T. L. & Tang, H. Leveraging multi-ethnic evidence for risk assessment of quantitative traits in minority populations. *Am. J. Hum. Genet.* **101**, 218–226 (2017).
29. Wojcik, G. L. et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).
30. Shi, H. et al. Population-specific causal disease effect sizes in functionally important regions impacted by selection. *Nat. Commun.* **12**, 1098 (2021).
31. Kuchenbaecker, K. et al. The transferability of lipid loci across African, Asian and European cohorts. *Nat. Commun.* **10**, 4330 (2019).
32. Mostafavi, H. et al. Variable prediction accuracy of polygenic scores within an ancestry group. *eLife* **9**, e48376 (2020).
33. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
34. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).
35. Weissbrod, O. et al. Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* **52**, 1355–1363 (2020).
36. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
37. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
38. Lloyd-Jones, L. R. et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* **10**, 5086 (2019).
39. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
40. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
41. Nagai, A. et al. Overview of the BioBank Japan project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
42. Asiki, G. et al. The general population cohort in rural south-western Uganda: a platform for communicable and non-communicable disease studies. *Int. J. Epidemiol.* **42**, 129–141 (2013).
43. Heckerman, D. et al. Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc. Natl Acad. Sci. USA* **113**, 7377–7382 (2016).
44. Duan, S., Zhang, W., Cox, N. J. & Dolan, M. E. FstSNP-HapMap3: a database of SNPs with high population differentiation for HapMap3. *Bioinformation* **3**, 139–141 (2008).
45. Purcell, S. M. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
46. Stahl, E. A. et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat. Genet.* **44**, 483–489 (2012).
47. Gazal, S. et al. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat. Genet.* **50**, 1600–1607 (2018).
48. Lam, M. et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat. Genet.* **51**, 1670–1678 (2019).
49. Nievergelt, C. M. et al. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat. Commun.* **10**, 4558 (2019).
50. Sakaue, S. et al. Trans-biobank analysis with 676,000 individuals elucidates the association of polygenic risk scores of complex traits with human lifespan. *Nat. Med.* **26**, 542–548 (2020).
51. Vuckovic, D. et al. The polygenic and monogenic basis of blood traits and diseases. *Cell* **182**, 1214–1231.e11 (2020).
52. Guo, J. et al. Global genetic differentiation of complex traits shaped by natural selection in humans. *Nat. Commun.* **9**, 1865 (2018).
53. Sved, J. A., McRae, A. F. & Visscher, P. M. Divergence between human populations estimated from linkage disequilibrium. *Am. J. Hum. Genet.* **83**, 737–743 (2008).
54. Budin-Ljøsne, I. et al. Data sharing in large research consortia: experiences and recommendations from ENGAGE. *Eur. J. Hum. Genet.* **22**, 317–321 (2014).
55. Surakka, I. et al. The impact of low-frequency and rare variants on lipid levels. *Nat. Genet.* **47**, 589–597 (2015).
56. Horikoshi, M. et al. Discovery and fine-mapping of glycaemic and obesity-related trait loci using high-density imputation. *PLoS Genet.* **11**, e1005230 (2015).
57. Pain, O. et al. Evaluation of polygenic prediction methodology within a reference-standardized framework. *PLoS Genet.* **17**, e1009021 (2021).
58. Chung, W. et al. Efficient cross-trait penalized regression increases prediction accuracy in large cohorts using secondary phenotypes. *Nat. Commun.* **10**, 569 (2019).

59. Chun, S. et al. Non-parametric polygenic risk prediction via partitioned GWAS summary statistics. *Am. J. Hum. Genet.* **107**, 46–59 (2020).
60. Im, C. et al. Generalizability of 'GWAS hits' in clinical populations: lessons from childhood cancer survivors. *Am. J. Hum. Genet.* **107**, 636–653 (2020).
61. Daetwyler, H. D., Villanueva, B. & Woolliams, J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PloS ONE* **3**, e3395 (2008).
62. Visscher, P. M. & Hill, W. G. The limits of individual identification from sample allele frequencies: theory and statistical analysis. *PLoS Genet.* **5**, e1000628 (2009).
63. Galinsky, K. J. et al. Estimating cross-population genetic correlations of causal effect sizes. *Genet. Epidemiol.* **43**, 180–188 (2019).
64. Brown, B. C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic genetic-correlation estimates from summary statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).

The Biobank Japan Project

Koichi Matsuda[11,12], Yuji Yamanashi[13], Yoichi Furukawa[14], Takayuki Morisaki[15], Yoshinori Murakami[16], Yoichiro Kamatani[12,17], Kaori Muto[18], Akiko Nagai[18], Wataru Obara[19], Ken Yamaji[20], Kazuhisa Takahashi[21], Satoshi Asai[22,23], Yasuo Takahashi[23], Takao Suzuki[24], Nobuaki Sinozaki[24], Hiroki Yamaguchi[25], Shiro Minami[26], Shigeo Murayama[27], Kozo Yoshimori[28], Satoshi Nagayama[29], Daisuke Obata[30], Masahiko Higashiyama[31], Akihide Masumoto[32] and Yukihiro Koretsune[33]

[11]Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [12]Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan. [13]Division of Genetics, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [14]Division of Clinical Genome Research, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [15]Division of Molecular Pathology, IMSUT Hospital Department of Internal Medicine, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [16]Department of Cancer Biology, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [17]Laboratory of Complex Trait Genomics, Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan. [18]Department of Public Policy, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [19]Department of Urology, Iwate Medical University, Iwate, Japan. [20]Department of Internal Medicine and Rheumatology, Juntendo University Graduate School of Medicine, Tokyo, Japan. [21]Department of Respiratory Medicine, Juntendo University Graduate School of Medicine, Tokyo, Japan. [22]Division of Pharmacology, Department of Biomedical Science, Nihon University School of Medicine, Tokyo, Japan. [23]Division of Genomic Epidemiology and Clinical Trials, Clinical Trials Research Center, Nihon University School of Medicine, Tokyo, Japan. [24]Tokushukai Group, Tokyo, Japan. [25]Department of Hematology, Nippon Medical School, Tokyo, Japan. [26]Department of Bioregulation, Nippon Medical School, Kawasaki, Japan. [27]Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan. [28]Fukujuji Hospital, Japan Anti-Tuberculosis Association, Tokyo, Japan. [29]Cancer Institute Hospital of the Japanese Foundation for Cancer Research, Tokyo, Japan. [30]Center for Clinical Research and Advanced Medicine, Shiga University of Medical Science, Shiga, Japan. [31]Department of General Thoracic Surgery, Osaka International Cancer Institute, Osaka, Japan. [32]Iizuka Hospital, Fukuoka, Japan. [33]National Hospital Organization, Osaka National Hospital, Osaka, Japan.

Methods

PolyPred and its summary statistics-based analogs. All methods in this paper use a linear PRS, that is, $\hat{y} = \sum_i x_i \hat{\beta}_i$, where $\hat{y}$ is the PRS of an individual, x_i is the number of minor alleles of SNP i carried by that individual and $\hat{\beta}_i$ is the estimated per-allele causal effect size of SNP i. The methods differ in the way they estimate $\hat{\beta}_i$.

PolyPred and PolyPred$^+$ both combine the methods PolyFun-pred and BOLT-LMM, PolyPred-S and PolyPred-S$^+$ both combine the methods PolyFun-pred and SBayesR, and PolyPred-P and PolyPred-P$^+$ both combine the methods PolyFun-pred and PRS-CS. PolyFun-pred estimates $\hat{\beta}_i$ as the (approximate) posterior mean causal effect size of SNP i, as estimated by PolyFun + SuSiE[35] based on European training data, using 187 functional annotations to specify prior causal probabilities (see below). BOLT-LMM (respectively SBayesR and PRS-CS) estimates tagging effects (Supplementary Note) of HapMap 3 SNPs by applying BOLT-LMM[36,37] (respectively SBayesR[38] and PRS-CS[39]) to European training data. BOLT-LMM (respectively SBayesR) treats the effect of each SNP i as a random effect sampled from a mixture of two (respectively four) zero-mean normal distributions, the variances and mixture weights of which are determined in a data-driven manner. PRS-CS treats the effect of each SNP i as a random effect sampled from a continuous shrinkage prior distribution.

PolyPred and its summary statistics-based analogs compute the effect size of each SNP i that is either in HapMap 3 or has a European MAF ≥ 0.1% and INFO score ≥ 0.6 as a weighted combination of (1) its PolyFun-pred effect size based on European training data and (2) its BOLT-LMM (respectively SBayesR and PRS-CS) effect size based on European training data:

$$\begin{aligned}\widehat{\beta}_i^{\text{PolyPred(-S)}} &= w^{\text{PolyFun-pred}} \times \widehat{\beta}_i^{\text{PolyFun-pred}} \\ &+ w^{\text{BOLT-LMM/SBayesR/PRS-CS}} \times \widehat{\beta}_i^{\text{BOLT-LMM/SBayesR/PRS-CS}},\end{aligned} \tag{3}$$

where $\widehat{\beta}_i^{\text{PolyFun-pred}}$ is the PolyFun-pred approximate posterior mean causal effect size of SNP i based on European training data, $\widehat{\beta}_i^{\text{BOLT-LMM/SBayesR/PRS-CS}}$ is the approximate posterior mean tagging effect size of SNP i based on European training data using the indicated method (setting the effects of SNPs not in HapMap 3 to 0), and $w^{\text{PolyFun-pred}}$ and $w^{\text{BOLT-LMM/SBayesR/PRS-CS}}$ are mixing weights. PolyPred estimates the mixing weights via non-negative least squares estimation (that is, least squares estimation constrained to produce to non-negative estimates) based on training individuals from the target cohort. Specifically, PolyPred (respectively PolyPred-S and PolyPred-P) estimates the mixing weights by computing the PRS corresponding to the PolyFun-pred effect sizes (given by $\hat{y}^{\text{PolyFun-pred}} = \sum_i x_i \hat{\beta}_i^{\text{PolyFun-pred}}$) and the PRS corresponding to the BOLT-LMM (respectively SBayesR and PRS-CS) effect sizes (given by), and then fitting the mixing weights by regressing the true phenotypes y_i of the training individuals in the target cohort on the PolyFun-pred and the BOLT-LMM (respectively SBayesR and PRS-CS) PRSs. The use of non-negative least squares estimation guarantees that the correlation of the predicted phenotype with the true phenotype is at least as large as the smallest of the correlations between each constituent predicted phenotype and the true phenotype.

PolyPred$^+$ and its summary statistics-based analogs compute the effect size of each SNP i that either is in HapMap 3 or has a European MAF ≥ 0.1% and INFO score ≥ 0.6 as a weighted combination of (1) its PolyFun-pred effect size based on European training data, (2) its BOLT-LMM (respectively SBayesR and PRS-CS) effect size based on European training data and (3) its effect size as estimated by applying BOLT-LMM (respectively SBayesR and PRS-CS) to training data from the target population (or a closely related population):

$$\begin{aligned}\widehat{\beta}_i^{\text{PolyPred+}} &= w^{\text{PolyFun-pred}} \times \widehat{\beta}_i^{\text{PolyFun-pred}} \\ &+ w^{\text{BOLT-LMM/SBayesR/PRS-CS}} \times \widehat{\beta}_i^{\text{BOLT-LMM/SBayesR/PRS-CS}} \\ &+ w^{\text{BOLT-LMM/SBayesR/PRS-CS-non-Eur}} \\ &\times \widehat{\beta}_i^{\text{BOLT-LMM/SBayesR/PRS-CS-non-Eur}},\end{aligned} \tag{4}$$

where $\widehat{\beta}_i^{\text{BOLT-LMM/SBayesR/PRS-CS-non-Eur}}$ is the BOLT-LMM (respectively SBayesR or PRS-CS) approximate posterior mean tagging effect of SNP i based on training data from the non-European population (and set to zero for SNPs that are not in HapMap 3) and $w^{\text{BOLT-LMM/SBayesR/PRS-CS-non-Eur}}$ is the mixing weight of $\widehat{\beta}_i^{\text{BOLT-LMM/SBayesR/PRS-CS-non-Eur}}$. The mixing weights are estimated as in PolyPred.

In practice, we apply PolyPred and its summary statistics-based analogs by linearly combining the PolyFun-pred PRS and the BOLT-LMM (or SBayesR or PRS-CS) PRSs (rather than linearly combining the SNP effect sizes). The two procedures are almost mathematically identical, with the only difference being that a linear combination of PRSs can also accommodate an intercept, which explicitly bias-corrects the PRS to the target population.

We applied PolyFun-pred in the same way that we applied PolyFun + SuSiE in our previous work[35]. Briefly, we applied PolyFun-pred across 2,763 overlapping 3-Mb loci (equally spaced starting at chromosome 1, position 0) spanning 18,212,157 European MAF > 0.1% imputed SNPs with INFO score > 0.6 (excluding the human leukocyte antigen (HLA) and two other long-range LD regions)[35], assuming 10 causal SNPs per locus. We used summary statistics computed by BOLT-LMM, based on up to $n = 337{,}491$ unrelated British-ancestry UK Biobank individuals and using summary LD information estimated directly from the target samples. Full details are provided in ref. [35]. We note that the use of BOLT-LMM summary statistics is mathematically equivalent to regressing the target phenotypes on BOLT-LMM off-chromosome PRSs before applying PolyFun + SuSiE[37]. We also note that the use of 3-Mb loci guarantees that, for each SNP, the estimation of its causal effect size takes into account virtually all relevant SNPs that may be in LD with that SNP (because LD in European populations rarely ranges beyond 1 Mb[65]), allowing disentanglement of its causal effect size from its tagging effect size.

PRS methods that include noncommon SNPs (MAF < 5%) may be sensitive to MAF-dependent and LD-dependent architectures[47,66,67]. Previous PRS methods have largely alleviated this concern by discarding noncommon SNPs instead of explicitly modeling their lower per-SNP heritability[33,38,39,58,59,68–73]. In contrast, PolyFun-pred accounts for MAF-dependent and LD-dependent architectures by specifying SNP-specific prior causal probabilities based on the baseline-LF model[47] (Supplementary Table 11). In detail, PolyFun-pred uses 187 overlapping functional annotations from the baseline-LF model (previously described in ref. [35]), including: 10 common MAF bins (MAF ≥ 0.05), 10 LF MAF bins (0.05 > MAF ≥ 0.001); 6 LD-related annotations for common SNPs; 5 LD-related annotations for LF SNPs; 40 binary functional annotations for common SNPs; 7 continuous functional annotations for common SNPs; 40 binary functional annotations for LF SNPs; 3 continuous functional annotations for LF SNPs; and 66 annotations constructed via windows around other annotations[74] (Supplementary Table 11).

Estimating relative R^2 and its s.e. We measured prediction accuracy for each trait via a measure that we call relative R^2, defined via the following computations:

(1) Compute R^2-PRS: the R^2 obtained via a linear predictor that includes PRS, age, sex, age × sex (if the correlation with age was <0.95), UK Biobank assessment center (defined via dummy binary variables), genotyping array, ten principal components (PCs; computed separately for each ancestry; see below) and dilution factor (for biochemical traits only).
(2) Compute R^2-noPRS, defined like R^2-PRS but omitting the PRS.
(3) Compute R^2-PRS-BOLT-EUR, computed by applying BOLT-LMM to UK Biobank non-British Europeans as in step (1).
(4) Compute R^2-noPRS-BOLT-EUR, computed by applying BOLT-LMM but omitting the PRS to non-British Europeans.
(5) Compute relative R^2 as (R^2-PRS − R^2-noPRS)/(R^2-PRS-BOLT-EUR − R^2 − noPRS-BOLT-EUR).

We note that fold improvement in relative R^2 is the same as fold improvement in absolute difference in R^2 (that is, in R^2-PRS − R^2-noPRS), because the denominator (R^2-PRS-BOLT-EUR − R^2-noPRS-BOLT-EUR) is a trait-specific scaling factor.

We computed the s.e. of relative R^2, differences in relative R^2 (for example, versus BOLT-LMM), ancestry-specific regression slopes and the area under the receiver operating curve (for disease traits) via genomic block-jackknife, partitioning the genome into 200 equally sized consecutive loci and omitting each one in turn. In secondary analyses, we computed the s.e. by applying jackknife over individuals from the target population. These analyses yielded much smaller s.e.s in the UK Biobank, suggesting that genomic block-jackknife s.e.s may be conservative, whereas individual-based jackknife estimates maty be anti-conservative. We emphasize that individual-based jackknife explicitly assumes a fixed training set.

We estimated statistics (for example, relative R^2) for meta-analyzed traits via an inverse-variance-weighted average, using weights inversely proportional to the s.e. of the R^2 of BOLT-LMM in the target population (as estimated via genomic block-jackknife). We estimated the s.e. of the meta-analyzed statistics as the square root of the weighted average of the trait-specific sampling variances (obtained via genomic block-jackknife), divided by the square root of the number of traits. We computed P values of differences in relative R^2 versus BOLT-LMM via Wald's test.

We computed the statistical significance of the decrease in R^2 in non-European versus European target samples via Wald's test for the difference in R^2, conservatively estimating the sampling variance of this difference as the sum of the sampling variances of the European R^2 and the non-European R^2 (this is a conservative estimate as long as the R^2 estimates in Europeans and non-Europeans are not negatively correlated, which is extremely unlikely).

Cohorts analyzed. *UK Biobank.* The UK Biobank is a UK-based population cohort[40]. We used v.3 of the imputed genotypes, as described in our previous work[35]. We computed ancestry-specific PCs for UK Biobank Africans, UK Biobank east Asians and UK Biobank south Asians via PLINK v.1.9 (ref. [75]), restricting to SNPs that have ancestry-specific MAF > 5%, missingness < 10% and Hardy–Weinberg equilibrium P value > 10^{-10}, and were LD pruned using the command --indep-pairwise 1000 50 0.05, and restricted to unrelated individuals (kinship coefficient <0.05) from the target ancestry with missingness <10%. We used the UK Biobank-provided PCs for UK Biobank Europeans.

We defined the 'autoimmune disease' trait in the UK Biobank as a union of the following UK Biobank codes: 1154 (irritable bowel syndrome), 1222 (type

1 diabetes), 1224 (thyroid problem), 1225 (hyperthyroidism/thyrotoxicosis), 1226 (hypothyroidism/myxedema), 1256 (acute infective polyneuritis/Guillain–Barré syndrome), 1260 (myasthenia gravis), 1261 (multiple sclerosis), 1313 (ankylosing spondylitis), 1372 (vasculitis), 1377 (polymyalgia), 1378 (Wegener's granulomatosis), 1381 (systemic lupus erythematosus), 1382 (Sjögren's syndrome/sicca syndrome), 1384 (scleroderma/systemic sclerosis), 1437 (myasthenia gravis), 1453 (psoriasis), 1456 (malabsorption/celiac disease), 1461 (inflammatory bowel disease), 1462 (Crohn's disease), 1463 (ulcerative colitis), 1464 (rheumatoid arthritis), 1477 (psoriatic arthropathy), 1522 (Graves' disease), 1661 (vitiligo) and 1667 (alopecia/hair loss).

ENGAGE. ENGAGE is a consortium comprising 24 cohorts that study the impact of genetic variations on medical phenotypes through GWASs[54]. The consortium has performed over 80,000 GWASs using genetic and phenotype samples from >600,000 individuals and made the GWAS summary statistics publicly available[54].

We obtained ENGAGE GWAS summary statistics, representing fixed-effect meta-analyses from 22 studies of European ancestry, for 2 lipid phenotypes[55] (triglyceride (n = 56,267) and total cholesterol (n = 58,327)) and 2 obesity-related phenotypes[56] (BMI (n = 80,938) and BMI-adjusted waist:hip ratio (n = 49,877)). In each ENGAGE study, up to 37.4 million autosomal variants were imputed using the 1000 Genomes project (we used 8.1 million variants which were also imputed in the UK Biobank); phenotypes were adjusted for age, age squared, genotype PCs and other study-/trait-specific covariates, and were inverse-rank normalized; GWASs were performed for each sex separately and combined using fixed-effect meta-analysis; a single genomic control correction was performed for each study before a cross-study meta-analysis[55,56].

Biobank Japan. Biobank Japan (BBJ) is a multi-institutional, hospital-based biobank with DNA and serum samples from approximately 200,000 participants from 12 medical institutions in Japan[41]. The participants are mainly of Japanese ancestry and had been diagnosed with at least 1 of 47 diseases by physicians at the cooperating hospitals. Written informed consent was obtained from all the participants, as approved by the ethics committees of RIKEN Center for Integrative Medical Sciences and the Institute of Medical Sciences at the University of Tokyo.

We genotyped samples with either (1) the Illumina HumanOmniExpressExome BeadChip or (2) a combination of the Illumina HumanOmniExpress and HumanExome BeadChips. We applied standard quality control criteria for both samples and variants as detailed elsewhere[76]. We then pre-phased genotypes with Eagle v.2 (ref. [77]) and imputed dosages with Minimac3 (ref. [78]) using 1000 Genomes project phase 3 (v.5) data (n = 2,504) and Japanese whole-genome sequencing (WGS) data (n = 1,037) as a reference[76]. We computed PCs using EIGENSOFT's smartpca[79].

For phenotypes, we retrieved clinical medical records from the participating hospitals through interviews and a standardized questionnaire. We used 23 diseases and complex traits in BBJ which are also analyzed in UK Biobank (Supplementary Table 3). We normalized quantitative phenotypes via inverse-rank normal transformation as described elsewhere[80]. We defined the 'autoimmune disease' trait in BBJ as a union of Graves' disease and rheumatoid arthritis.

Uganda-APCDR. Uganda-APCDR is a population-based cohort from the general population cohort, Uganda. We retrieved genotype and phenotype data through the APCDR initiative via the European Genome–Phenome Archive (EGA), using EGAD00010000965 to access genotype data. Phenotype data were accessed via sftp from the EGA (reference: DD_PK_050716 gwas_phenotypes_28Oct14.txt). The participants were from nine ethno-linguistic groups in sub-Saharan Africa and had been recruited from the study area located in south-western Uganda in the Kyamulibwa subcounty of the Kalungu district, approximately 120 km from Entebbe town. These ethno-linguistic groups have a diverse population structure with varying degrees of admixture between Eurasian and east African Nilo-Saharan ancestries, which has been extensively characterized elsewhere[81]. The detailed cohort demographics, sample collection and processing have been described previously[42,43].

Briefly, the samples were genotyped using the Illumina HumanOmni 2.5-M BeadChip at the Wellcome Trust Sanger Institute. We used the Ricopili pipeline to conduct pre-imputation quality control and perform phasing and imputation[82]. Briefly, we phased the data using Eagle v.2.3.5 (ref. [77]) and imputed variants using Minimac3 (ref. [78]) in chunks ≥3 Mb. The 1000 Genomes project phase 3 haplotypes[65] were used as the reference panel for phasing and imputation.

As described previously, phenotypes were collected using a standard individual questionnaire, blood samples (laboratory tests) and biophysical measurements (height, weight, waist and hip circumferences, and blood pressure)[42]. We normalized quantitative phenotypes via inverse-rank normal transformation.

UK Biobank simulations. We simulated data based on real genotypes of UK Biobank individuals, using 250,963 MAF ≥ 0.1% SNPs with INFO score ≥ 0.6 on chromosome 22 (including short indels) (Supplementary Note). We trained all methods using 337,491 unrelated British-ancestry individuals[40], and we estimated the mixing weights of PolyPred and its summary statistics-based analogs using up to 1,000 additional individuals from each of the four non-British ancestries. We computed summary statistics by applying linear regression via PLINK v.2.0. We did not evaluate PolyPred+ in the simulations because of the relatively small sample sizes of the UK Biobank non-European populations. We evaluated prediction accuracy via R^2, using held-out individuals who were not included in the training sets and were unrelated to the training set individuals and to each other, using 42,000 non-British Europeans, 7,700 south Asians, 900 east Asians and 6,200 Africans. We computed PRSs by applying PLINK v.2.0 with the --score command, using imputed dosage data (rather than hard-called SNP values). We computed s.e.s via a jackknife over simulations.

We trained BOLT-LMM by applying BOLT-LMM v.2.3.4 to PLINK files of HapMap 3 SNPs (hard-coded from imputed dosages), using the same covariates specified in Estimating relative R^2 and its s.e., and specifying the flag --predBetasFile to report PRS coefficients.

We trained SBayesR using summary statistics from the infinitesimal version of BOLT-LMM (BOLT-LMM-inf[36]), which yielded far superior accuracy versus using summary statistics from the noninfinitesimal version of BOLT-LMM. We ran SBayesR using 10,000 iterations, 4,000 burn-in iterations, using values from 10% of the iterations to compute posterior means and the HapMap 3 LD files published by the SBayesR authors[83]. We attempted to run SBayesR using a mixture of four distributions (using $\pi = [0.95,\ 0.02,\ 0.02,\ 0.01]$ and $\gamma = [0,\ 0.01,\ 0.1,\ 1]$). In case SBayesR failed with these parameters, we iteratively shrank the last entry in the vector γ by 50% until it was $<10^{-6}$, at which point we removed the last mixture component and redefined π such that the first entry was equal to 0.95 and all other entries had the same value such that all values sum to 1.0.

We trained PRS-CS using summary statistics from BOLT-LMM-inf (as in SBayesR) with the parameters $a = 1$, $b = 0.5$, thin = 5, n_iter = 10000, n_burnin = 500, and without specifying the value of φ (corresponding to PRS-CS-auto). We used the UK Biobank LD reference panels made publicly available by the authors of PRS-CS (see below).

We trained P + T by applying PLINK with the command –clump-r2 0.5 –clump-kb 250 and various values of –clump-p1 (following ref. [13]), and using 10,000 randomly selected, unrelated UK Biobank British individuals to compute LD. We estimated LD using 10,000 individuals to balance between runtime and accuracy (noting that P + T is relatively insensitive to the LD reference panel size compared with the other methods evaluated in this manuscript). We used summary statistics based on BOLT-LMM, using marginal effect sizes derived from reported χ^2 values (that is, the square root of χ^2 divided by the square root of the BOLT-LMM effective sample size[35], and multiplied by the sign of the effect size estimated by the infinitesimal version of BOLT-LMM). We used the best value of –clump-p1 (out of the evaluated values 10^{-2}, 10^{-3}, 10^{-4}, 10^{-6} and 5×10^{-8}) based on the target sample phenotypes, which leads to anti-conservative prediction accuracy estimates for P + T.

We used slightly different LD reference panels for PolyFun-pred, SBayesR and PRS-CS, because (1) they use different algorithms to impose sparsity on LD matrices, and different file formats to store them, and (2) we assume that naively running SBayesR or PRS-CS using summary LD from the 18 million SNPs used by PolyFun-pred would be computationally infeasible, based on information provided in the papers describing these methods[38,39]. When modifying the training sample size, we kept the LD reference panel sample size fixed to alleviate computational costs.

Analysis of real data. We performed four sets of analyses: (1) analysis of four UK Biobank populations using UK Biobank British training data; (2) analysis of four UK Biobank populations using ENGAGE meta-analysis training data; (3) analysis of BBJ and Uganda-APCDR cohorts; and (4) analysis of UK Biobank east Asians using UK Biobank British and BBJ training data. In analysis sets (1), (3) and (4), we evaluated PRSs generated by training all methods using unrelated UK Biobank British-ancestry individuals. In analysis set (2), we evaluated PRSs generated by training all methods using summary statistics from 8.1 million meta-analyzed summary statistics from the ENGAGE consortium[54–56]. In a subset of analysis set (3) and in analysis set (4) we additionally evaluated PRSs generated by training BOLT-LMM–BBJ (BOLT-LMM trained on BBJ individuals). In all analysis sets, the individuals in the target populations were unrelated to each other and to the individuals in the training set (when available).

In analysis sets (1), (3) and (4), we selected the seven traits to meta-analyze by first restricting the set of 49 traits analyzed in ref. [35] to traits that are available in BBJ and Uganda-APCDR and are well powered across multiple ancestries, having $h^2 > 0.05$ in UK Biobank non-British Europeans, UK Biobank south Asians and UK Biobank Africans (see below for details on ancestry-specific heritability estimation). We then iteratively greedily selected ranked traits according to their heritability in UK Biobank non-British Europeans (estimated as in ref. [35]), such that no selected trait had $|r_g| < 0.3$ with a previously selected trait.

We computed ancestry-specific SNP heritabilities in each UK Biobank ancestry by applying GCTA[84] to unrelated sets of individuals using hard-called HapMap 3 SNPs (using a random set of 10,000 individuals for non-British Europeans to facilitate the computations). We did not use more advanced methods[85] because of the relatively small sample sizes. We meta-analyzed ancestry-specific SNP heritabilities by averaging the estimated heritabilities and we estimated the meta-analyzed s.e. via the square root of the average sampling variance, divided by the square root of the number of traits.

In analysis sets (1), (3) and (4), we trained all PRS methods on UK Biobank-unrelated British-ancestry individuals (average n = 325) as described in

UK Biobank simulations, but using summary statistics generated by BOLT-LMM when applied to UK Biobank British-ancestry individuals, as described in our previous work[35]. We trained P + T separately for each non-UK Biobank cohort by restricting the set of SNPs considered to be the set of SNPs available in both the UK Biobank and the target cohort. We computed the contribution of PolyFun-pred (respectively BOLT-LMM) toward PolyPred via the ratio of the mixing weight of PolyFun-pred (respectively BOLT-LMM) to the sum of the mixing weights of PolyPred and BOLT-LMM.

In analysis sets (1), (2) and (4), we computed a PRS for each UK Biobank individual using imputed dosage data as described in the UK Biobank simulations. In analysis set (3), we computed a PRS for each individual in BBJ and Uganda-APCDR using imputed dosage data and PLINK v.2.0 (refs. [86,87]).

In secondary analyses of analysis set (1) we also evaluated LDpred[33]. We trained LDpred using HapMap 3 SNPs and two different LD reference panels: 1000 Genomes project[65] and UK10K[88]. We used summary statistics from the infinitesimal version of BOLT-LMM (as in SBayesR) and with default parameters, using the parameter --ldr 400. We used the value of '--F' (corresponding to the assumed proportion of causal SNPs, using all the default-evaluated values) that yielded the best prediction accuracy in the target sample, yielding anti-conservative accuracy estimates as in P + T.

In analysis sets (3) and (4), we trained BOLT-LMM-BBJ, SBayesR-BBJ and PRS-CS-BBJ (BOLT-LMM, SBayesR and PRS-CS, respectively, trained using BBJ training data) (average $n = 124{,}000$). We selected individuals for training these methods as described in our previous work[13], but excluding a random subset of 5,000 individuals who were used for evaluating prediction accuracy. For SBayesR-BBJ, we used a subset of individuals ($n = 50{,}000$) from BBJ to compute in-sample LD, following the recommendations of the authors of SBayesR[38]. For PRS-CS-BBJ, we used the east Asian LD reference panels made publicly available by the authors of PRS-CS (Data availability).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Access to the UK Biobank resource is available via application (http://www.ukbiobank.ac.uk). PRS coefficients generated in the present study are available for public download at http://data.broadinstitute.org/alkesgroup/polypred_results. Summary LD information of $n = 337{,}000$ British-ancestry UK Biobank individuals for 2,763 overlapping 3-Mb loci is available at https://data.broadinstitute.org/alkesgroup/UKBB_LD. Summary LD information of $n = 50{,}000$ UK Biobank individuals for SBayesR is available at https://zenodo.org/record/3350914. Summary LD information used by PRS-CS is available at https://github.com/getian107/PRScs. Baseline-LF v.2.2.UKB annotations and LD scores for UK Biobank SNPs are available at https://data.broadinstitute.org/alkesgroup/LDSCORE/baselineLF_v2.2.UKB.tar.gz. Source data are provided with this paper.

Code availability

PolyPred and PolyPred+ are provided as part of the open-source software package PolyFun, which is freely available at https://doi.org/10.5281/zenodo.6139679 (ref. [89]) and https://github.com/omerwe/polyfun. BOLT-LMM is available at https://data.broadinstitute.org/alkesgroup/BOLT-LMM. SBayesR is available at https://cnsgenomics.com/software/gctb. PRS-CS is available at https://github.com/getian107/PRScs.

References

65. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
66. Zeng, J. et al. Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet.* **50**, 746–753 (2018).
67. Schoech, A. P. et al. Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nat. Commun.* **10**, 790 (2019).
68. Zhang, Q., Privé, F., Vilhjálmsson, B. & Speed, D. Improved genetic prediction of complex traits from individual-level data or summary statistics. *Nat. Commun.* **12**, 4192 (2021).
69. Hu, Y. et al. Leveraging functional annotations in genetic risk prediction for human complex diseases. *PLoS Comput. Biol.* **13**, e1005589 (2017).
70. Márquez-Luna, C. et al. Incorporating functional priors improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. *Nat. Commun.* **12**, 6052 (2021).
71. Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X. & Sham, P. C. Polygenic scores via penalized regression on summary statistics. *Genet. Epidemiol.* **41**, 469–480 (2017).
72. Yang, S. & Zhou, X. Accurate and scalable construction of polygenic scores in large biobank data sets. *Am. J. Hum. Genet.* **106**, 679–693 (2020).
73. Qian, J. et al. A fast and scalable framework for large-scale and ultrahigh-dimensional sparse regression with application to the UK Biobank. *PLoS Genet.* **16**, e1009141 (2020).
74. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
75. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–75 (2007).
76. Akiyama, M. et al. Characterizing rare and low-frequency height-associated variants in the Japanese population. *Nat. Commun.* **10**, 4393 (2019).
77. Loh, P.-R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
78. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
79. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
80. Sakaue, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
81. Gurdasani, D. et al. Uganda genome resource enables insights into population history and genomic discovery in Africa. *Cell* **179**, 984–1002 (2019).
82. Lam, M. et al. RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* **36**, 930–933 (2020).
83. Lloyd-Jones, L. GCTB SBayesR shrunk sparse linkage disequilibrium matrices for HM3 variants, summary statistics and predictors generated from 'Improved polygenic prediction by Bayesian multiple regression on summary statistics' by Lloyd-Jones, Zeng et al. 2019. *Zenodo* https://doi.org/10.5281/ZENODO.3350914 (2019).
84. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
85. Gazal, S., Marquez-Luna, C., Finucane, H. K. & Price, A. L. Reconciling S-LDSC and LDAK functional enrichment estimates. *Nat. Genet.* **51**, 1202–1204 (2019).
86. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
87. Purcell, S. & Chang, C. *PLINK v2.00a3LM* www.cog-genomics.org/plink/2.0/
88. The UK10K Consortium et al.The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
89. Weissbrod, O. Source code for PolyFun. *Zenodo* https://doi.org/10.5281/zenodo.6139679 (2022).

Acknowledgements

We thank A. Schoech and C. Márquez-Luna for helpful discussions. This research was conducted using the UK Biobank resource under application no. 16549 and was funded by the National Institutes of Health (NIH; grant nos. U01 HG009379, U01 HG012009, R37 MH107649, R01 MH101244 and R01 HG006399). M.K. was supported by a Nakajima Foundation Fellowship and the Masason Foundation. W.J.P. was supported by an NWO Veni grant (no. 91619152). A.R.M. was supported by the National Institute of Mental Health (grant no. K99/R00MH117229). H.K.F. was supported by E. and W. Schmidt. A.V.K. was supported by grants (nos. 1K08HG010155 and 1U01HG011719) from the National Human Genome Research Institute and a sponsored research agreement from IBM Research. Y.O. was supported by JSPS KAKENHI (grant nos. 19H01021 and 20K21834) and AMED (grant nos. JP21km0405211, JP21ek0109413, JP21ek0410075, JP21gm4010006 and P21km0405217) and JST Moonshot R&D (grant nos. JPMJMS2021 and JPMJMS2024). Computational analyses were performed on the O2 High-Performance Compute Cluster at Harvard Medical School.

Author contributions

O.W., M.K., H.S. and A.L.P. designed the study. O.W., M.K., H.S. and S.G. analyzed the data. O.W., M.K., H.S. and A.L.P. wrote the manuscript with assistance from S.G., W.J.P., A.V.K, Y.O., A.R.M. and H.K.F.

Competing interests

O.W. is an employee and holds equity in Eleven Therapeutics. H.S. is an employee of Genentech and holds stock in Roche. A.V.K. is an employee and holds equity in Verve Therapeutics, and has served as a scientific advisor to Sanofi, Amgen, Maze Therapeutics, Navitor Pharmaceuticals, Sarepta Therapeutics, Novartis, Silence Therapeutics, Korro Bio, Veritas International, Color Health, Third Rock Ventures, Foresite Labs and Columbia University (NIH); A.V.K. received speaking fees from Illumina, MedGenome, Amgen and the Novartis Institute for Biomedical Research, and also received a sponsored research agreement from IBM Research. All other authors declare no competing interests.

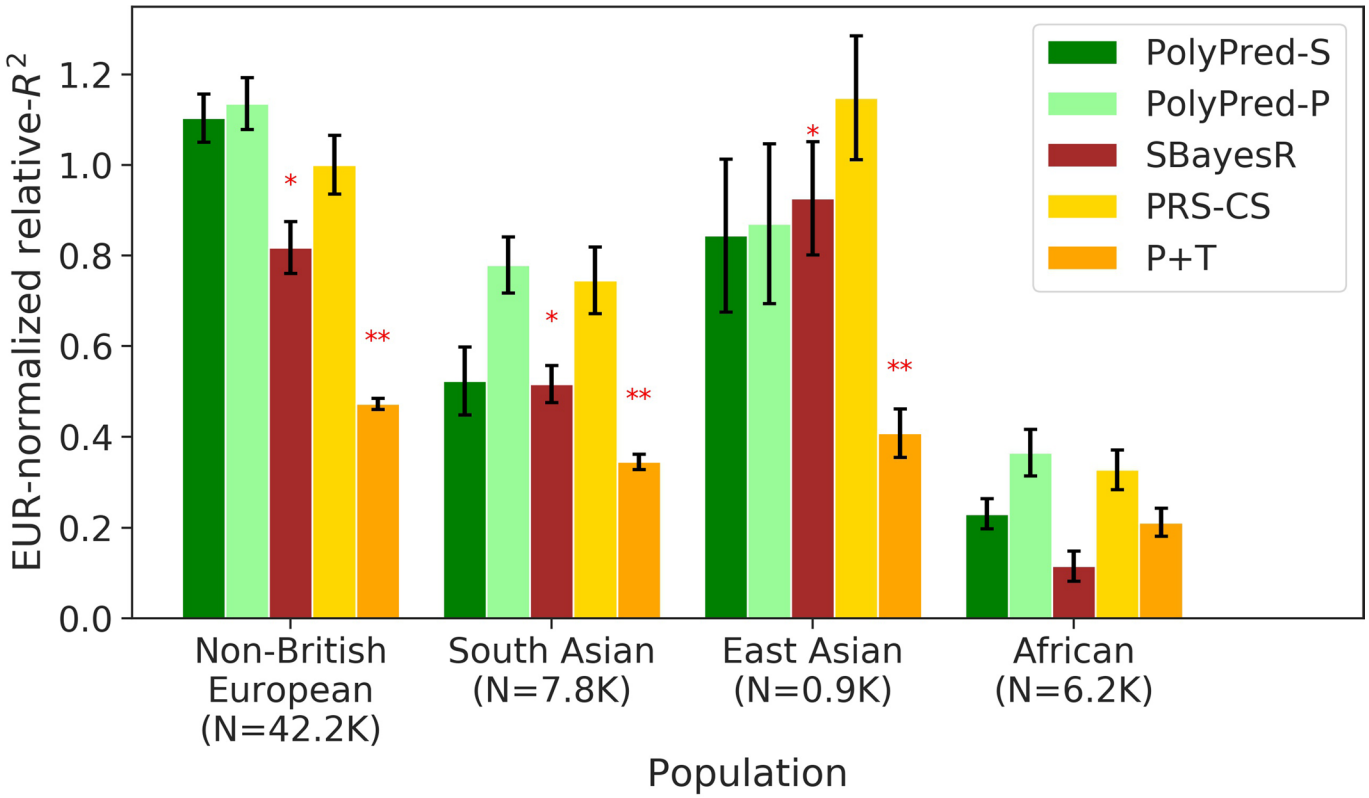
Additional information

Extended data Extended data are available for this paper at https://doi.org/10.1038/s41588-022-01036-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-022-01036-9.

Correspondence and requests for materials should be addressed to Omer Weissbrod or Alkes L. Price.

Peer review file *Nature Genetics* thanks Marylyn Ritchie and Vincent Plagnol for their contribution to the peer review of this work. Peer reviewer reports are available.



Extended Data Fig. 1 | Cross-population PRS results for real UK Biobank traits, using summary statistics from a meta-analysis of many cohorts. We report average prediction accuracy (relative-R^2, but computed with respect to PRS-CS instead of BOLT-LMM; see main text), meta-analyzed across 4 well-powered, approximately independent traits, for PRS trained in European Network for Genetic and Genomic Epidemiology (ENGAGE) samples (average $N = 61{,}365$) and applied to four UK Biobank populations. Target population sample sizes are indicated in parentheses; PolyPred and its summary statistic-based analogues used 500 additional training samples from each target population to estimate mixing weights. Asterisks above each bar denote statistical significance of the difference vs. PRS-CS, with red asterisks denoting a disadvantage (*$P < 0.05$; **$P < 0.001$). P-values were computed using a two-sided Wald test and were not adjusted for multiple comparisons. Errors bars denote standard errors. Numerical results, results for all 4 traits analyzed, absolute prediction accuracies (R^2), and P-values of relative improvements vs. PRS-CS are reported in Supplementary Table 5 and Supplementary Table 8.

nature research

Corresponding author(s): Omer Weissbrod

Last updated by author(s): Feb 18, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We did not collect data for this study.
Data analysis	The software used for data analysis is available in Zenodo: https://doi.org/10.5281/zenodo.6139680

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Access to the UK Biobank resource is available via application (http://www.ukbiobank.ac.uk). PRS coefficients generated in this study are available for public download at http://data.broadinstitute.org/alkesgroup/polypred_results. Summary LD information of N=337K British-ancestry UK Biobank individuals for 2,763 overlapping 3Mb loci is available at: https://data.broadinstitute.org/alkesgroup/UKBB_LD. Summary LD information of N=50K UK Biobank individuals for SBayesR is available at: https://zenodo.org/record/3350914. Summary LD information used by PRS-CS is available at: https://github.com/getian107/PRScs. Baseline-LF v2.2.UKB annotations and LD-scores for UK Biobank SNPs are available at: https://data.broadinstitute.org/alkesgroup/LDSCORE/baselineLF_v2.2.UKB.tar.gz

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We used all the data samples that were available to us using the resources analyzed in this study (UK Biobank, Biobank Japan, Uganda APCDR, ENGAGE consortium)
Data exclusions	We excluded individuals who wished to be omitted from the UK Biobank dataset
Replication	We replicated our results by applying predicting to both UK Biobank, Biobank Japan, and the Uganda-APCDR cohorts. We performed this replication only once, since these were the only datasets available to us.
Randomization	We performed no randomization and analyzed all unrelated UK Biobank individuals. We controlled for non-random allocation by including covariates for UK Biobank assessment center (defined via dummy binary variables), genotyping array, 10 principal components (computed separately for each ancestry; see below), and dilution factor (for biochemical traits only).
Blinding	We did not apply blinding because we only analyzed population cohorts that were already collected

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	We analyzed individual-level data from the UK Biobank, Biobank Japan, Uganda ACPDR, and summary statistics from the ENGAGE consortium. All these datasets have already been thoroughly characterized in the following publications: 1. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203 (2018). 2. Nagai, A. et al. Overview of the BioBank Japan Project: study design and profile. J. Epidemiol. 27, S2–S8 (2017). 3. Asiki, G. et al. The general population cohort in rural south-western Uganda: a platform for communicable and non-communicable disease studies. Int. J. Epidemiol. 42, 129–141 (2013). 4. Budin-Ljøsne, I. et al. Data sharing in large research consortia: experiences and recommendations from ENGAGE. Eur. J. Hum. Genet. 22, 317–321 (2014).
Recruitment	We analyzed large population cohorts data that were recruited in previous studies
Ethics oversight	UK Biobank: Collection of the UK Biobank (UKBB) data was approved by the UKBB's Research Ethics Committee. Approval to use UKBB individual-level in this work was obtained under application #16549. Biobank Japan: All the participants provided written informed consent approved by the ethics committees of RIKEN Center for Integrative Medical Sciences and the Institute of Medical Sciences at the University of Tokyo. Uganda-APCDR: As described previously in Asiki et al 2013, before all survey procedures including interviews, blood tests and sample storage for future use, written consent or assent in conjunction with parental/guardian consent for those less than 18

years of age, are obtained following Uganda National Council of Science and Technology (UNCST) guidelines. Written consent/assent is also obtained from participants on the use of their clinical records for research purposes. All study procedures including material transfer agreements are approved annually by the Uganda Virus Research Institute Science and Ethics Committee and the UNCST. A request to use of these deidentified data for this work (genetic data from EGAD00010000965 for genetic data and phenotype data via sftp with reference: DD_PK_050716 gwas_phenotypes_28Oct14.txt) via a Data Access Application for External Investigators was approved by the Data Access Committee for APCDR via and accessed through the European Genome-Phenome Archive.

ENGAGE: The summary statistics for ENGAGE are anonymized and are publicly available for download.

Note that full information on the approval of the study protocol must also be provided in the manuscript.