

UC Santa Cruz

UC Santa Cruz Previously Published Works

Title

Polygenic risk score portability for common diseases across genetically diverse populations.

Permalink

<https://escholarship.org/uc/item/6mw5n990>

Journal

Human Genomics, 18(1)

Authors

Moreno-Grau, Sonia

Vernekar, Manvi

Lopez-Pineda, Arturo

et al.

Publication Date

2024-09-02

DOI

10.1186/s40246-024-00664-y

Peer reviewed

RESEARCH

Open Access



Polygenic risk score portability for common diseases across genetically diverse populations

Sonia Moreno-Grau^{1,2}, Manvi Vernekar³, Arturo Lopez-Pineda^{1,4,5}, Daniel Mas-Montserrat¹, Míriam Barrabés¹, Consuelo D. Quinto-Cortés¹, Babak Moatamed¹, Ming Ta Michael Lee¹, Zhenning Yu³, Kensuke Numakura³, Yuta Matsuda³, Jeffrey D. Wall¹, Alexander G. Ioannidis^{1,2,6}, Nicholas Katsanis¹, Tomohiro Takano^{3,7*} and Carlos D. Bustamante^{1,2*}

Abstract

Background Polygenic risk scores (PRS) derived from European individuals have reduced portability across global populations, limiting their clinical implementation at worldwide scale. Here, we investigate the performance of a wide range of PRS models across four ancestry groups (Africans, Europeans, East Asians, and South Asians) for 14 conditions of high-medical interest.

Methods To select the best-performing model per trait, we first compared PRS performances for publicly available scores, and constructed new models using different methods (LDpred2, PRS-CSx and SNPnet). We used 285 K European individuals from the UK Biobank (UKBB) for training and 18 K, including diverse ancestries, for testing. We then evaluated PRS portability for the best models in Europeans and compared their accuracies with respect to the best PRS per ancestry. Finally, we validated the selected PRS models using an independent set of 8,417 individuals from Biobank of the Americas-Genomelink (BbofA-GL); and performed a PRS-Phewas.

Results We confirmed a decay in PRS performances relative to Europeans when the evaluation was conducted using the best-PRS model for Europeans (51.3% for South Asians, 46.6% for East Asians and 39.4% for Africans). We observed an improvement in the PRS performances when specifically selecting ancestry specific PRS models (phenotype variance increase: 1.62 for Africans, 1.40 for South Asians and 0.96 for East Asians). Additionally, when we selected the optimal model conditional on ancestry for CAD, HDL-C and LDL-C, hypertension, hypothyroidism and T2D, PRS performance for studied populations was more comparable to what was observed in Europeans. Finally, we were able to independently validate tested models for Europeans, and conducted a PRS-Phewas, identifying cross-trait interplay between cardiometabolic conditions, and between immune-mediated components.

Conclusion Our work comprehensively evaluated PRS accuracy across a wide range of phenotypes, reducing the uncertainty with respect to which PRS model to choose and in which ancestry group. This evaluation has let us

*Correspondence:

Tomohiro Takano
tomo@genomelink.io
Carlos D. Bustamante
carlos.bustamante@galatea.bio

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

identify specific conditions where implementing risk-prioritization strategies could have practical utility across diverse ancestral groups, contributing to democratizing the implementation of PRS.

Background

Polygenic risk scores (PRS) predict complex disease susceptibility and traits based on genetic data [1]. PRSs have been identified as a promising tool to be implemented from genomic medicine to population-based screenings [2]. However, most of PRS are impacted by a lack of portability, i.e. the variation in predictive performance metrics across populations, which limits their utility.

The portability of PRS depends on genetic similarity between training and target cohorts. Several studies have demonstrated a decay in PRS accuracy with genetic distance from the training set, at continental and sub-continental level [1, 3–6]. Therefore, most European-derived PRS are not likely to predict as well in individuals of other ancestries. The underlying reasons explaining these disparities are commonly attributed to differences in causal variants, linkage disequilibrium (LD) patterns, allele frequencies, and effect sizes across populations. Recent studies have indicated that using PRS derived from multi-ancestry data increases the performance in diverse populations [7]. Despite that, lack of precise PRS models hinders risk stratification and leads to exacerbation of health disparities in implementation of the PRSs across different societies [1].

PRS portability might be improved by reducing the genetic distance from target to training data, which might be promoted by increasing population diversity in large biobanks. According to the GWAS catalog on January 2024, ~78% of the individuals included were of European ancestry, followed by ~11% of Asian ancestry, and ~4% from other minorities (i.e. African, Hispanic/Latino, middle Eastern, Native American and Oceanian ancestries) (<https://www.ebi.ac.uk/gwas/docs/ancestry-data>). Although active recruitment of non-European individuals in genetic studies is greatly increasing [8–10], large biobanks and cohorts are still far from proportionally representing global populations.

Although generalizing a single PRS model useful for all populations supposes the gold standard, it is becoming more established that optimal PRS methods are context specific [11]. Nowadays, a large number of PRS per trait have been developed with different methodologies, scarcely validated, and mostly derived from European individuals. Scarce information is available about their comparative performances in a harmonized scenario across populations. Altogether, it complicates the decision-making with respect to which is the most optimal PRS model for a specific condition and ancestry to be implemented at scale.

Here, we investigate the variability of a wide range of PRS models across four ancestry groups (Africans, Europeans, East Asians and South Asians) for 14 conditions of high-medical interest (i.e. cardiometabolic, digestive, and inflammatory conditions, lipid, and Vitamin D levels). With that objective, we first selected the most optimal PRS model for Europeans from a set of 166 PRS available from PGS Catalog [12] or de novo constructed using three different PRS methods (LDpred2 [13, 14], PRS-CSx [15] and SNPnet [16]) in the UKBB data. Then, to test whether the portability problem might be reduced when choosing the best-performing model in each ancestral group, we compare PRS performances between the best PRS for Europeans and the best PRS per ancestry. Finally, we validated the optimal PRS models per condition in a set of European individuals from the Biobank of the Americas – GenomeLink (BbofA-GL) and performed ancestry-specific PRS-Phewas per condition.

Methods

Datasets and phenotypes

UKBB cohort

UKBB is a prospective cohort collected from multiple sites across the United Kingdom, which includes over 500,000 participants. Information on genotyping and quality control has previously been described¹⁸. UK Biobank received ethical approval from the NHS National Research Ethics Service North West (11/NW/0382). This research has been conducted under Application Number 89,006.

UKBB phenotypes were extracted through ICD10 codes (See Supplementary Table 1). We explored the following conditions: cardiometabolic (atrial fibrillation [AF], coronary artery disease [CAD], hypertension, type 1 diabetes [T1D], type 2 diabetes [T2D] and hypothyroidism), digestive (celiac disease [CD], hemorrhoids and ulcerative colitis [UC]), and inflammatory conditions (gout), lipid levels (high-density lipoproteins cholesterol [HDL-C]; low-density lipoproteins cholesterol [LDL-C] and triglycerides [TG]), and Vitamin D levels. Briefly, we excluded all individuals with T2D diagnoses from T1D study composition (both case, and controls), and likewise the opposite. Lipid and Vitamin D levels were dichotomized (case vs. control) according to a specific cut-off, i.e. LDL > 190 mg/dL or 4.9 mmol/L; HDL > 60 mg/dL or 1.5 mmol/L; Triglycerides > 200 mg/dL or 2.3 mmol/L; and Vitamin D > 30 nmol/L or 12 ng/mL. Control groups per each condition were defined as participants not diagnosed with the tested condition. The final number of cases and the sample demographics for both the training

and testing sets are described in the Supplementary Tables 2 and 3.

The biobank of the americas-genomelink (BbofA-GL)

The BbofA-GL cohort includes participants drawn from the customer base of Genomelink (genomelink.io), which offers a direct-to-consumer genetic traits platform with more than 500,000 users globally. All participants included in the analyses provided informed consent and answered surveys online according to our human subjects protocol, which was reviewed and approved by WCG IRB (<https://www.wcgirb.com/>) under IRB tracking ID 2020–1332.

BbofA-GL phenotypes were extracted from self-administered questionnaires. The online survey included specific questions about the 14 tested phenotypes and questions about age, sex, weight, and height. Supplementary Table 4 shows the online questionnaire. Data were collected over a period of six months, from August 2022 to February 2023. Only the initial response of each participant was included in the study if genotype information was available. Case-control groups were created as previously described. We observed an increase in the % of women collected by BbofA-GL, which we explain by using a self-reported questionnaire.

Genotype data preparation and quality control

UKBB cohort: Training and testing samples

We used the UK Biobank dataset (directly genotyped data from release version 2, and the imputed dataset v3) as part of our training and testing samples. Briefly, we removed individuals from our analyses based on the following criteria reported by the UK Biobank in the sample QC file, “ukb_sqc.txt”: (1) individuals showing putative sex chromosome aneuploidy; (2) marked as outliers for heterozygosity and missing rates; (3) presenting an excess of relatives in the dataset (>10); and we kept individuals used in the principal component analysis (PCA). This process generated a set of 406,640 unrelated individuals.

Then, with the objective of estimating genetic ancestry, we performed a principal component analysis (PCA) using the directly genotyped UKBB version. This analysis was restricted to variants particularly used by the UK Biobank team for internal PCA, extracted from the variant QC file “ukb_snp_qc.txt”. PCA was conducted after merging the dataset with the 1000G reference panel, using `-pca` option of Plink2 software [17] (<https://www.cog-genomics.org/plink/2.0/>). We then combined the information derived from self-reported ethnicity (UK Biobank field ID: 21000) and principal component analysis (thresholds on PC1 and PC2), to generate the 4 subpopulations for this study: European ($n=381,199$ individuals), African (8,111), East Asian (2,365), and South Asian (SAS, 6,596). The SAS group self-identified as

Pakistani, Bangladeshi, Indian or Asian_or_Asian_British (Supplementary Fig. 1). Among them, we used 75% of the directly genotyped data for PRS training and 25% of the imputed data for further PRS evaluation in European individuals, and 100% of the data for evaluation in non-Europeans. In the UK Biobank imputed version (v3), we only used bi-allelic variants with imputation information score ≥ 0.7 , MAF ≥ 0.01 , and variant call rate $\geq 97\%$ ($N=8,962,898$).

BbofA-GL: validation Sample

This cohort includes data from nine independent genotyping arrays. Genotype-level data for each array were processed by applying identical quality control and imputation procedures, as previously described [18]. Briefly, variants with a call rate of $<97\%$ and palindromic markers (A/T, G/C, MAF >0.4) were excluded. We performed an exact test for Hardy–Weinberg equilibrium for individuals of the largest ancestral group ($p < 1 \times 10^{-12}$, globally). Individual quality control (QC) includes matching between gender identification and chromosomal sex (when it is possible), and no excess ancestry-adjusted heterozygosity. Samples genetically related to other individuals in the cohort and duplicates were detected and removed using the King algorithm (`-make-king`, `king estimate > 0.177`; Plink2 [17]). PCA was performed to identify global ancestry per individual using 1000 Genomes Project data as reference populations (See Supplementary Fig. 2). Further information about the number of markers and samples per genotyping array pre- and post-QC is available in Supplementary Table 5. Imputation was carried out using 1000 Genomes Project data as a reference panel with Beagle [19]. Next, we generated a merged dataset combining imputed genotypes (MAF >0.01 ; imputation quality $R^2 >0.30$) from available datasets. Imputed makers with call rate >0.97 in the merged data were selected for downstream analysis ($N=9,605,881$). The final sample size for this dataset was 8,394 individuals of European ancestry with all available phenotypes.

Polygenic risk score calculation

We used the `--score` function from Plink2 [19] to calculate PRS. Briefly, for each individual j , the PRS score takes the form $PRS_j = \sum \beta_i G_{ij}$; where β_i is a weight equal to the log odds ratios for the variant i , and G_{ij} is the number of risk alleles (2,1,0) of variant i in subject j . The weights for each variant were obtained from: (a) publicly available models; and (b) new models trained using three different PRS methods (LDpred2, PRS-CSx and SNPnet). We specifically followed previous recommendations⁸ to ensure the coding allele in the target data was the effective allele reported in the summary statistic. PRS-scores were standardized to PRS Z-scores for downstream analyses.

PRS Models publicly available

We downloaded a total of 402 PRS models for 14 phenotypes (Supplementary Table 6) from the PGS-Catalog database [12] (<https://www.pgscatalog.org/>) on April 27, 2023: 14 for AE, 10 for CD, 35 for CAD, 9 for gout, 3 for hemorrhoids, 33 for HDL, 50 for hypertension, 15 for hypothyroidism, 55 for LDL, 34 for TG, 9 for T1D, and 88 for T2D, 3 for UC and 44 for vitamin D. With the goals of generating automatic downloads and harmonizing data, we used custom versions of the “*download_scorefiles.py*” and the “*combine_scorefile.py*” scripts from PGS Catalog Utilities (<https://pypi.org/project/pgscatalog-utils/>). The studied PRS models were downloaded in GRCh37 genome assembly. To avoid overfitting of the results, we excluded from the European analysis all the downloaded PGS-Catalog models using GWAS summary statistics derived from the UKBB European set or trained with them. In the case of the analysis conducted in non-European individuals, we just limited the number of available scores to those specifically using summary statistics or training with the UKBB non-European set. This lets us have a total of 97 and 400 publicly available models available for testing in the UKBB sample, respectively.

PRS model training

New PRS models were trained on 285,900 European individuals from UKBB, when we used LDpred2 and SNPnet. When GWAS summary statistics were needed (LDpred2 and PRS-CSx) we downloaded available summary statistics from FinnGen [20], version DF9 (<https://r9.finnngen.fi/>), and from Biobank of Japan [21] (<https://pheweb.jp/>) for the studied phenotypes. Software conditions are here-with detailed:

Batch screening iterative lasso (BASIL, or SNPnet in its R implementation) [16]. SNPnet applies a multiple regression method that takes advantage of a LASSO solver and optimizes for datasets that are too big to fit into the memory. BASIL identifies the set of SNPs that most efficiently discriminate between cases and controls. The training was performed with the following parameters: n lambda=200; alpha=1; ncores=20. Additional parameters were set up to the default.

PRS Continuous Shrinkage (PRS-CSx) [15]. PRS-CSx places a continuous shrinkage (CS) prior on SNP effect sizes, which is robust to varying genetic architectures. Here, we applied the default “auto” version of PRS-CSx that obtains weights through the Gibbs sampling algorithm, using the specific reference panels per specific ancestry (i.e., European samples from the 1000 Genomes Project (“EUR reference”) for individuals of European ancestry). To construct multi-ancestry models, we used the “meta” flag, to get the combined SNP effect sizes across populations using an inverse-variance-weighted

meta-analysis of the population-specific posterior effect size estimates.

LDpred2 (bigsnpr) [14]. LDpred [13] is a method that infers the posterior mean effect size of each marker by using a prior on effect sizes and LD information from an external reference panel (the default value is from HapMap 3). A Gibbs sampler is used to estimate the residualized marginal effect for each SNP. In this work, we used LDpred2, which extends LDpred by auto-tuning the best-performing hyper-parameters within the inner loop of the algorithm. We used SNPs that are overlapped in HapMap 3 variants and UKBB genotypes (Supplementary Table 7).

Statistical analysis - evaluation of model performance

To assess the performance of the PRS models, we evaluated: (i) area under curve (AUC); (ii) percentage of the variance explained by the PRS (%Variance); (iii) effect and significance of the PRS after multiple testing correction (Bonferroni correction); and (iv) case enrichment across PRS percentiles; using individuals that were not included in the training sets from the UKBB and the BbofA-GL datasets. AUC per each model was estimated using “predict” and “roc” functions from the pROC package in R software for models including (a) covariables only and (b) a full model with covariables and the PRS. The included covariates in the analysis were: age, sex and the first four principal components (PCs) resulting from ancestry analysis. Array type was also used as a covariate to control the impact of the 9 different array types in the BbofA-GL dataset. The %Variance explained by the PRS was calculated as the difference between the Full model and the covariates only. The PRS effect on each specific phenotype was estimated from a generalized linear model (glm) for a binomial outcome for: a) an unadjusted model (PRS only), for the full model (PRS and covariables); and (c) for the full model also including BMI, using R. To account for multiple testing, we used a Bonferroni threshold of $p < 2.95 \times 10^{-4}$ (0.05/166) for the results of the analysis with European ancestry individuals and $p < 1.24 \times 10^{-4}$ (0.05/402) for all models. Finally, disease risk was estimated between intermediate PRS percentiles (40–60%) and PRS percentile (0–10%, 10–20%, 20–30%, 30–40%, 60–70%, 70–80%, 80–90%, and 90–100%) by applying a glm as previously described. Phenotypes with number of cases < 25, were excluded from downstream analysis. We used the Ggplot package from R to plot results.

PRS-Phewas

PRS-phenome-wide association analyses (PRS-Phewas) for each selected PRS were conducted in the UK Biobank testing sample across different ancestries. All the phenotypes studied in this analysis were binary. The association between the 14 PRS scores and each phenotype,

excluding the PRS-condition, was assessed by applying a glm for a binary outcome, for the previously described models, the PRS model only, the full model and the full model considering BMI. This study was performed using R. All estimates correspond to 1 SD change of the PRS. A conservative Bonferroni correction for multiple testing that assumes uncorrelated traits yields a p-value threshold of $p < 2.74 \times 10^{-4}$ (0.05/182). We used Ggplot in R to plot results.

Results

With the objective of evaluating PRS portability across different ancestries and comparing best-performing PRS model in each ancestral group, we have tested publicly available PRS models and constructed new models using three different PRS methods (LDpred2, PRS-CSx and SNPnet) for 14 high-interest medical conditions in two independent cohorts including individuals from different ancestries (see flowchart in Fig. 1). These scores were trained within 285,900 European individuals and tested in 6,595 individuals of African ancestry, 2,473 of East Asian ancestry, 95,300 of European ancestry, and 8,111 of South Asian ancestry from the UK Biobank (See principal component analysis for ancestry in Supplementary Fig. 2). We then validated them in 8,417 European individuals from the BbofA-GL cohort (See Supplementary Fig. 3). Further description about the demographic

information for UKBB and BbofA-GL can be found in Supplementary Tables 2 and 3.

Performance and validation of the optimal PRS models in European-ancestry cohorts

To select the best-performing model per condition, i.e., that exhibiting the highest % of variance explaining the phenotype; we assessed a total of 166 PRS models, including both publicly available and de-novo trained ones, for cardiac, metabolic, digestive, and inflammatory conditions, as well as lipid and vitamin D lab levels. A total of 105 PRSs were significantly associated with their corresponding phenotype ($p < 2.74 \times 10^{-4}$) and showed the expected effect direction in the UKBB European set (Supplementary Table 8). The best-performing models per condition were able to explain between 1% for hemorrhoids to ~22% for celiac disease of disease heritability. We observed effects doubling the risk of developing the condition for UC, CD, T1D, and LDL-C in the testing sample (Table 1). UC was the condition showing the strongest effect, with an odds ratio per standard deviation increase (OR/SD) of 3.17 (95% confidence interval (CI) 2.97–3.38) in a model adjusted for age, sex, and principal components for genetic ancestry. The best-performing PRS were then validated in the BbofA-GL cohort. The selected models were significantly replicated in this

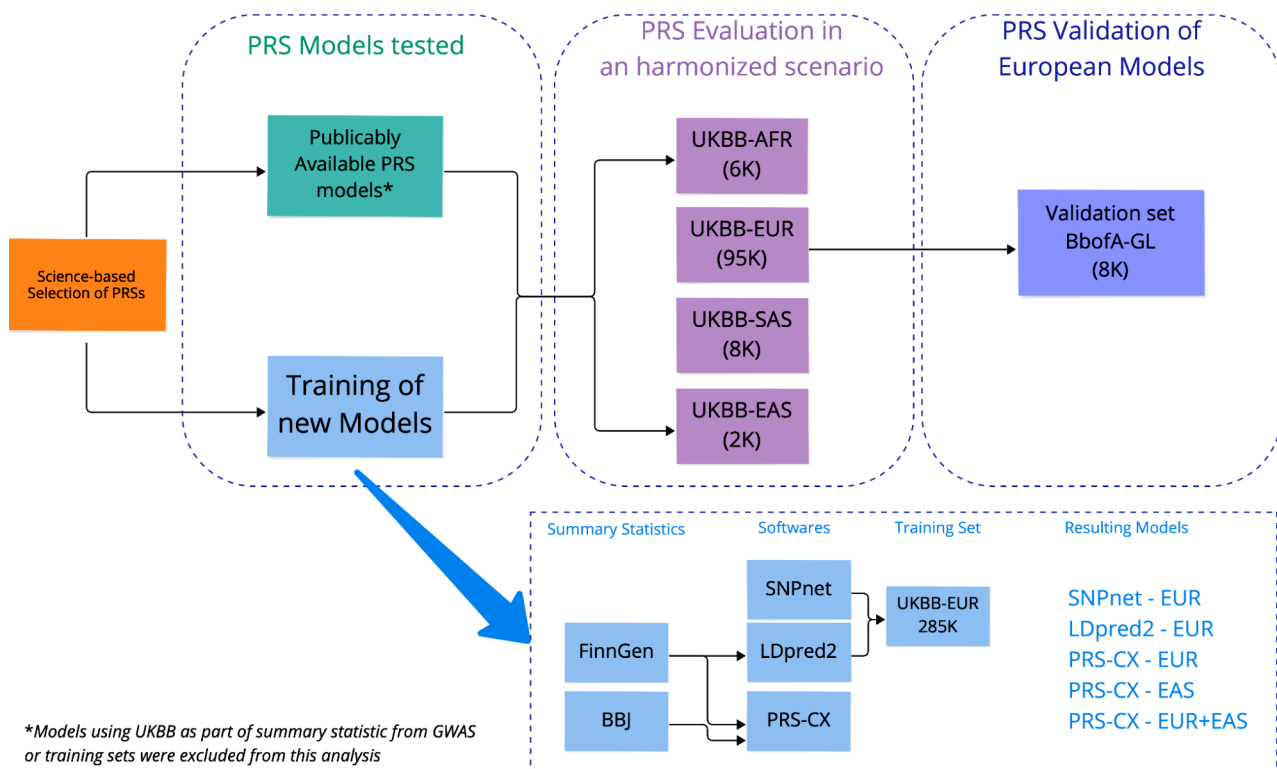


Fig. 1 Flow chart of the strategy applied in this study

Table 1 Best-performing polygenic risk score for 14 medical conditions in europeans

Phenotype	Model	UKBB Cohort			BbofA-GL Cohort		
		AUC	%Var.	OR (95% C.I.)	AUC	%Var.	OR (95% C.I.)
AF	LDpred2	0.75	1.60	1.41 (1.37–1.45)	0.71	1.63	1.44 (1.28–1.63)
CAD	PRSice (PGS003446)	0.77	3.99	1.85 (1.81–1.90)	0.82	1.45	1.60 (1.42–1.82)
CD	SnpNet	0.80	21.85	2.84 (2.64–3.06)	0.67	4.93	1.52 (1.35–1.71)
Gout	SnpNet	0.79	1.80	1.50 (1.43–1.57)	0.73	2.78	1.52 (1.35–1.71)
HDL-C	SnpNet	0.77	7.17	1.98 (1.95–2.01)	0.60	0.02	1.02 (0.94–1.11)
Hemorrhoids	LDpred2	0.55	0.93	1.10 (1.06–1.14)	0.56	0.73	1.10 (1.04–1.15)
Hypertension	SnpNet	0.72	1.98	1.43 (1.41–1.45)	0.72	1.90	1.42 (1.35–1.50)
Hypothyroidism	SnpNet	0.73	3.91	1.64 (1.60–1.69)	0.70	3.05	1.50 (1.41–1.60)
LDL-C	PRS-CSx ^a	0.69	10.64	2.02 (1.95–2.08)	0.65	2.52	1.35 (1.26–1.46)
T1D	GWS variants (PGS000023)	0.77	18.38	2.74 (2.41–3.13)	0.77	16.00	2.31 (2.00–2.68)
T2D	PRS-CS (PGS003103)	0.72	6.93	1.91 (1.86–1.96)	0.75	5.92	1.93 (1.80–2.07)
TG	SnpNet	0.69	7.90	1.79 (1.76–1.82)	0.68	5.88	1.64 (1.54–1.75)
UC	LDpred2 (PGS002066)	0.78	20.60	3.17 (2.97–3.38)	0.57	0.50	1.11 (0.97–1.28)
Vitamin D	SBayesR (PGS000882)	0.67	8.12	0.60 (0.58–0.61)	0.63	1.97	0.82 (0.78–0.86)

Note^aTransethnic model, %Var percentage of variance, OR odds ratio, 95% C.I 95% confidence interval

independent cohort for all the phenotypes, except for HDL and UC (Table 1).

To estimate what proportion of the population may benefit from PRS assessment, we evaluated case enrichment in the top decile compared to the two central deciles (40–60%). We observed that the top decile was significantly associated with T1D, T2D, CAD, hypertension, hypothyroidism, CD, TG measurement, gout, and Vitamin D, in both the UKBB sample and the BbofA-GL cohort (Fig. 2, and Supplementary Table 9). Among these conditions, individuals in the top decile had at least two-fold increased risk for T1D, T2D, hypothyroidism, CAD, TG measurement and CD in both datasets (Supplementary Fig. 4). When we studied this enrichment exclusively in the UKBB we found at least 2-fold increased risk in an additional four conditions. This disparity could be caused by the different phenotyping strategies used in both datasets, as UKBB data uses ICD10 phecodes and BbofA-GL self-reported questionnaires.

PRS accuracies in groups with non-European ancestries

We proceed to evaluate the portability of the optimal PRS models in Europeans across other populations. We observed a decay in the percentage of variance explained for the European-selected PRS models in non-Europeans (Supplementary Fig. 5). The relative predictive performance for the averaged 14 phenotypes compared to prediction performance in Europeans was 51.3% for South Asians, 46.6% for East Asians and 39.4% for Africans.

Considering these results, we modified our approach to select the best-performing model for each condition and ancestral group. We observed a subtle increase in the

phenotypic variance explained by the PRS with respect to the observed in the European selected models of 1.62 for Africans, 1.40 for South Asians and 0.96 for East Asians (Table 2). All the scores were significantly associated ($p < 1.24 \times 10^{-4}$) with their respective conditions, except for: hemorrhoids in all the ancestries; and AF, hypothyroidism and gout for East Asian ancestry individuals, and Vitamin D in both Asian groups, EAS, and SAS (Table 2). Further information about the performance of the whole set of tested models is provided in Supplementary Table 10.

Our ancestry-specific selection increased the case enrichment in the top decile of the PRS distributions in non-European populations. Conditions improved the most from this approach includes hypertension in African individuals, HDL-C, LDL-C and T2D in East Asians, and CAD, hypothyroidism, and LDL-C in South Asians (Fig. 3). A significant two-fold increased risk was observed between the 90th and the intermediate percentiles for HDL levels (>60 mg/dL) across all ancestries (i.e., OR(CI95) P value - $AFR_{HDL} = 2.20 (1.82-2.67) 9.48 \times 10^{-16}$; $SAS_{HDL} = 3.33 (2.61-4.25) 6.13 \times 10^{-22}$, $EAS_{HDL} = 2.50 (1.71-3.68) 2.67 \times 10^{-6}$). Those associations were also observed for CAD, high LDL levels (>190 mg/dL) in African, and South Asian groups (i.e., OR(CI95) P value. $AFR_{CAD} = 4.80 (3.52-6.61) 1.78 \times 10^{-22}$; $SAS_{CAD} = 3.92 (3.08-5.00) 2.16 \times 10^{-28}$; $AFR_{LDL} = 2.20 (1.46-3.30) 1.53 \times 10^{-04}$; $SAS_{LDL} = 2.65 (1.69-4.20) 2.38 \times 10^{-5}$); (See Supplementary Table 9). Similar pattern was found for T2D in East Asians, and high TG levels (>200 mg/dL) and hypothyroidism in South Asian ancestry. PRS percentile effects across

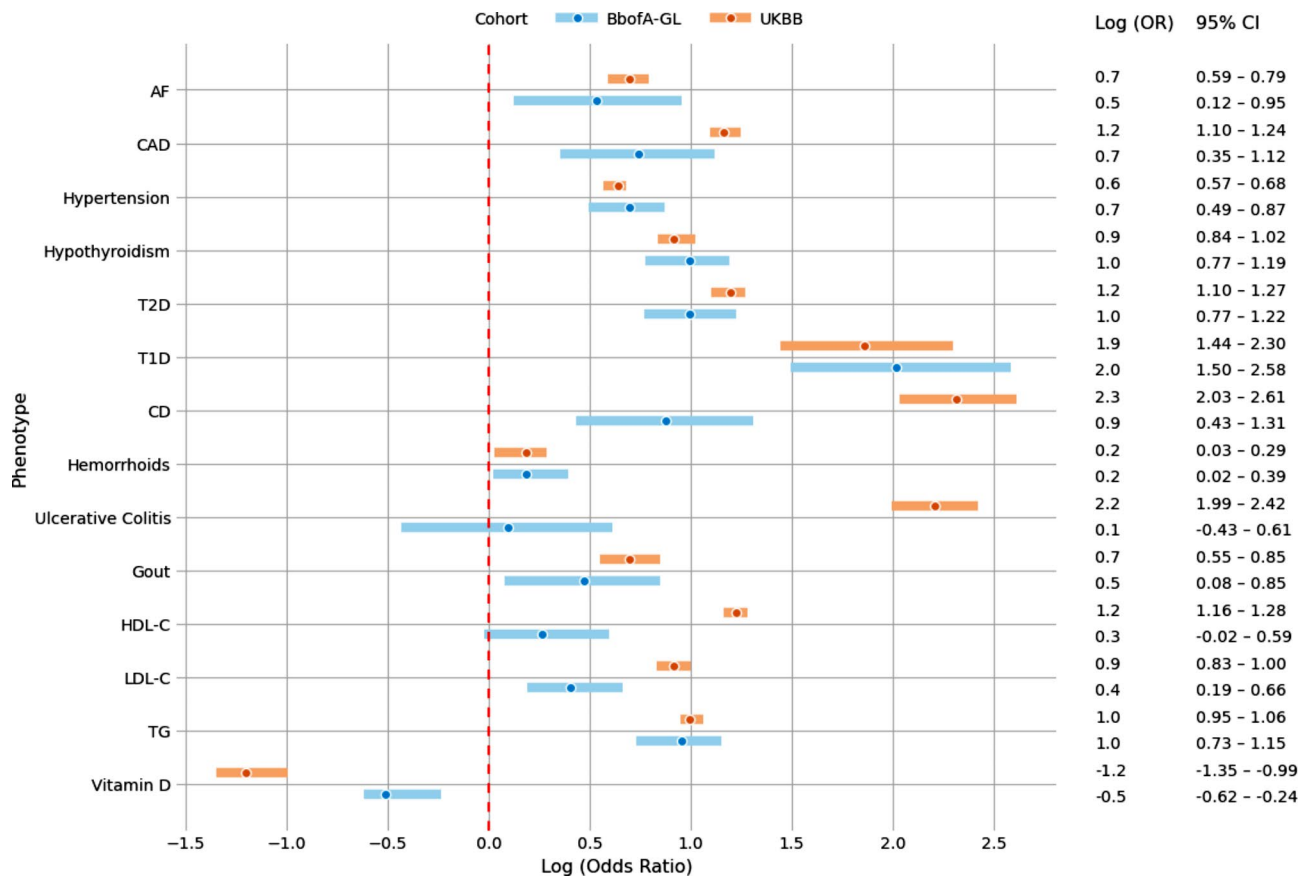


Fig. 2 Effect of the 90th PRS percentile compared to the intermediate percentile for fourteen medical conditions in the European set of individuals from the UKBB and BbofA

different ancestries are shown in Supplementary Tables 9 and Supplementary Fig. 6.

PRS-PheWAS across complex phenotypes

To assess whether our selected PRSs were associated with other medically relevant phenotypes, we performed a PRS-PheWAS across the conditions explored in this study in UK Biobank. We found 65 phenotype associations (43 and 22 for increasing and decreasing disease risk, respectively) across 14 medically relevant phenotypes (Fig. 4, and Supplementary Table 11).

In particular, we identified two groups of conditions with cross-trait interrelationships, those presenting a cardiometabolic or an autoimmune component. Polygenic risk for CAD was associated with traits related to well-known risk factors, i.e., hypertension, altered lipid levels, and insulin resistance syndrome. PRSs for AF, CAD, hypertension, hypothyroidism, T2D and Lipid levels were highly associated with the risk of being diagnosed with other cardiometabolic conditions, highlighting the biological interplay among them. We did not find an association between T1D-PRS and the risk of suffering T2D, or the opposite (Supplementary Table 11). T1D-PRS was associated with CD, hypothyroidism, and UC,

i.e., diseases well-known to present an autoimmune component. Low vitamin D PRS was significantly associated with the risk of developing conditions from the cardio-metabolic axis, i.e., CAD, hypertension, T2D, and high levels of LDL and TG. These trends were similar to those observed for African, East Asian, and South Asian ancestry individuals (Supplementary Table 12), although most of the associations did not reach statistical significance, due to the decreased sample size (and thus decreased power) in non-European populations.

Discussion

In this study, we have conducted a wide assessment of PRS portability and performance across different ancestry groups, and we have contributed to independently validate PRS models. Our data supports the poor portability of European derived models to other ancestries, which is in line with prior findings [1, 3–6]. Conversely, PRS performances relative to Europeans improved when we specifically selected the optimal PRS for a given condition and a given population ancestry, reinforcing the idea that optimal PRS methods are trait and context specific [11].

Table 2 Comparison of best-performing PRS per ancestry for 14 medical conditions in non-european populations

Pop.	Phenotype	Best_Model_per_Ancestry					Best_Model Europeans				
		Model	AUC	Var.	OR (95% C.I.)	P value	Model	AUC	Var.	OR (95% C.I.)	P-value
AFR	AF	PGS003461 (GWS)	0.77	0.96	1.31(1.11–1.53)	9.44×10^{-4}	LDpred2	0.76	0.06	1.08(0.93–1.27)	3.16×10^{-1}
AFR	CAD	PGS000337 (P+T)	0.79	5.79	2.14(1.94–2.37)	1.25×10^{-51}	PGS003446 (P+T)	0.77	4.29	2.09(1.86–2.36)	6.85×10^{-35}
AFR	gout	PGS000711 (SnpNet)	0.77	1.81	1.55(1.31–1.84)	3.40×10^{-7}	SnpNet	0.77	1.52	1.31(1.11–1.54)	1.51×10^{-3}
AFR	HDL-C	PGS001954(bigstatsr)	0.70	4.52	1.58(1.50–1.67)	1.52×10^{-63}	SnpNet	0.70	4.11	1.53(1.45–1.62)	1.37×10^{-55}
AFR	Hemorrhoids	PGS002058 (LDPred2)	0.56	1.10	1.11(1.00–1.24)	5.83×10^{-2}	LDpred2	0.55	0.12	0.95(0.85–1.06)	3.78×10^{-1}
AFR	Hypertension	PGS002047 (LDPred2)	0.72	0.51	1.20(1.15–1.27)	1.12×10^{-13}	SnpNet	0.72	0.35	1.16(1.10–1.22)	1.05×10^{-8}
AFR	Hypothyroidism	PGS000965 (SnpNet)	0.72	1.57	1.31(1.15–1.50)	4.22×10^{-5}	SnpNet	0.72	1.40	1.33(1.17–1.52)	1.40×10^{-5}
AFR	LDL-C	PGS001933 (bigstatsr)	0.68	10.57	1.97(1.70–2.29)	3.94×10^{-19}	PRS-CS*	0.66	8.58	1.82(1.57–2.13)	1.17×10^{-14}
AFR	T2D	PGS000712 (SnpNet)	0.69	0.79	1.23(1.16–1.31)	1.35×10^{-10}	PGS003103 (PRS-CS)	0.68	0.52	1.21(1.13–1.30)	3.44×10^{-8}
AFR	TG	PGS003401 (GWS)	0.69	1.96	1.40(1.28–1.53)	1.69×10^{-13}	SnpNet	0.69	1.49	1.34(1.23–1.46)	5.84×10^{-11}
AFR	Vitamin D	PGS003194 (lassosum)	0.61	0.81	0.88(0.83–0.93)	2.68×10^{-5}	PGS000882 (SBayesR)	0.60	0.50	0.90(0.84–0.95)	2.41×10^{-4}
EAS	AF	PGS001841 (bigstatsr)	0.83	1.49	1.53(1.10–2.12)	1.19×10^{-2}	LDpred2	0.81	0.12	1.07(0.77–1.50)	6.80×10^{-1}
EAS	CAD	PGS003356 (LDPred)	0.80	3.22	1.73(1.39–2.15)	1.16×10^{-6}	PGS003446 (P+T)	0.8	2.48	1.74(1.39–2.18)	1.72×10^{-6}
EAS	gout	PGS002030 (LDPred2)	0.83	1.00	1.50(1.09–2.05)	1.19×10^{-2}	SnpNet	0.83	0.56	1.29(0.93–1.79)	1.26×10^{-1}
EAS	HDL-C	PGS001954 (bigstatsr)	0.76	5.78	1.89(1.70–2.11)	4.36×10^{-32}	SnpNet	0.76	5.32	1.76(1.59–1.96)	1.05×10^{-26}
EAS	Hemorrhoids	PGS002058 (LDPred2)	0.59	1.58	1.16(0.97–1.40)	1.10×10^{-1}	LDpred2	0.58	0.44	1.11(0.93–1.33)	2.53×10^{-1}
EAS	Hypertension	PGS002998 (PRS-CS)	0.75	1.56	1.44(1.29–1.60)	2.08×10^{-11}	SnpNet	0.74	0.58	1.22(1.10–1.35)	1.81×10^{-4}
EAS	Hypothyroidism	SnpNet	0.70	4.36	1.57(1.23–2.00)	2.65×10^{-4}	SnpNet	0.70	4.36	1.57(1.23–2.00)	2.65×10^{-4}
EAS	LDL-C	PRS-CSx ^a	0.69	11.14	2.20(1.69–2.89)	9.27×10^{-9}	PRS-CS*	0.69	11.14	2.20(1.69–2.89)	9.27×10^{-9}
EAS	T2D	PGS000712 (SnpNet)	0.76	1.62	1.42(1.22–1.66)	7.26×10^{-6}	PGS003103 (PRS-CS)	0.74	0.37	1.24(1.07–1.45)	4.81×10^{-3}
EAS	TG	PRS-CSx_EAS	0.67	4.00	1.49(1.35–1.66)	2.89×10^{-14}	SnpNet	0.66	3.60	1.47(1.32–1.63)	5.93×10^{-13}
EAS	Vitamin D	PGS003198 (PRS-CS)	0.63	1.21	0.82(0.73–0.93)	1.30×10^{-3}	PGS000882 (SBayesR)	0.63	1.09	0.81(0.72–0.92)	6.60×10^{-4}
SAS	AF	PGS001356 (AnnoPred)	0.75	2.20	1.53(1.32–1.78)	1.60×10^{-8}	LDpred2	0.75	1.57	1.38(1.19–1.61)	2.72×10^{-5}
SAS	CAD	PGS003356 (LDPred)	0.79	4.97	2.04(1.89–2.20)	6.36×10^{-76}	PGS003446 (P+T)	0.77	3.57	1.89(1.75–2.04)	1.24×10^{-59}
SAS	gout	PGS001248 (SnpNet)	0.75	1.24	1.45(1.23–1.71)	8.43×10^{-6}	SnpNet	0.74	0.29	1.19(1.01–1.40)	3.73×10^{-2}
SAS	HDL-C	PGS001954 (bigstatsr)	0.77	6.80	2.13(1.97–2.30)	2.64×10^{-78}	SnpNet	0.76	5.64	1.95(1.80–2.10)	1.22×10^{-63}
SAS	Hemorrhoids	PGS001024 (SnpNet)	0.58	0.28	1.14(1.01–1.27)	2.84×10^{-2}	LDpred2	0.57	-0.07	1.10(0.98–1.23)	1.10×10^{-1}
SAS	Hypertension	PGS002765 (PRS-CS)	0.74	1.76	1.46(1.38–1.54)	6.16×10^{-40}	SnpNet	0.73	0.83	1.30(1.23–1.37)	5.80×10^{-21}
SAS	Hypothyroidism	PGS000965 (SnpNet)	0.73	3.24	1.64(1.49–1.81)	5.79×10^{-23}	SnpNet	0.72	2.25	1.46(1.33–1.61)	8.20×10^{-15}
SAS	LDL-C	PGS000688 (SnpNet)	0.67	7.58	1.76(1.52–2.03)	1.80×10^{-14}	PRS-CSx	0.66	5.71	1.60(1.38–1.85)	6.31×10^{-10}
SAS	T2D	PGS002308 (PRS-CSx)	0.67	2.28	1.43(1.34–1.52)	1.83×10^{-29}	PGS003103 (PRS-CS)	0.66	1.15	1.30(1.23–1.39)	1.32×10^{-16}
SAS	TG	PGS001979 (bigstatsr)	0.68	8.40	1.79(1.69–1.91)	4.94×10^{-78}	SnpNet	0.67	7.38	1.70(1.60–1.81)	1.58×10^{-67}
SAS	UC	LDpred2	0.68	10.39	1.78(1.46–2.17)	9.56×10^{-9}	PGS002066 (bigstatsr)	0.63	4.92	1.41(1.16–1.72)	7.01×10^{-4}
SAS	Vitamin D	PGS000882 (SBayesR)	0.59	0.46	0.90(0.86–0.96)	3.94×10^{-4}	PGS000882 (SBayesR)	0.59	0.46	0.90(0.86–0.96)	3.94×10^{-4}

Notes PRS-CS^a: Transethnic model; Models available from PGS Catalog are shown with their PGS nomenclature. Models trained by our own are shown with the name of the training software. Phenotypes with number of cases < 25 were not analyzed (i.e. CD, T1D, UC)

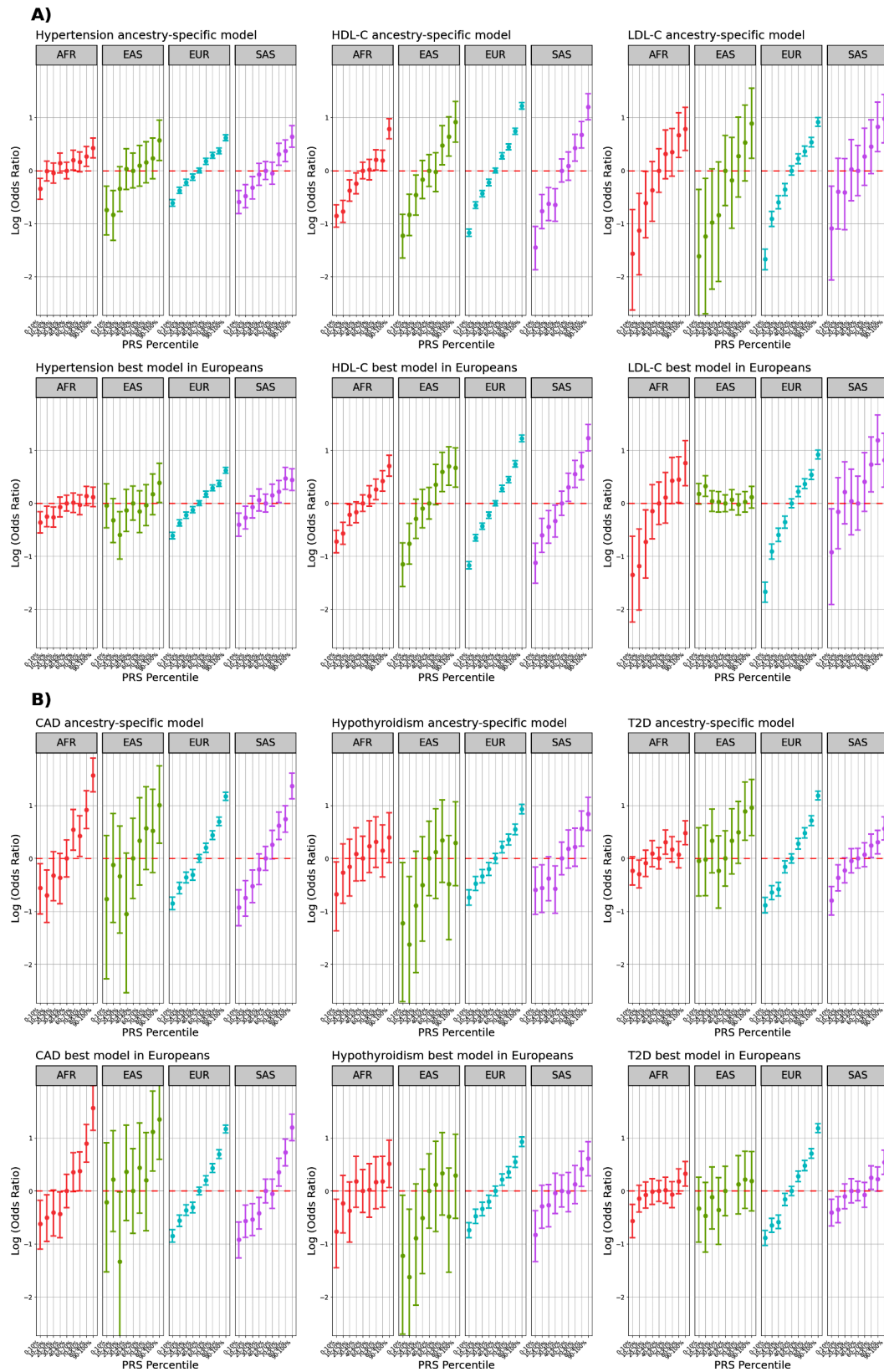


Fig. 3 Effect of PRS percentile for hypertension, HDL-C, LDL-C, CAD, hypothyroidism, T2D across diverse ancestries. Red dashed line indicates OR=1

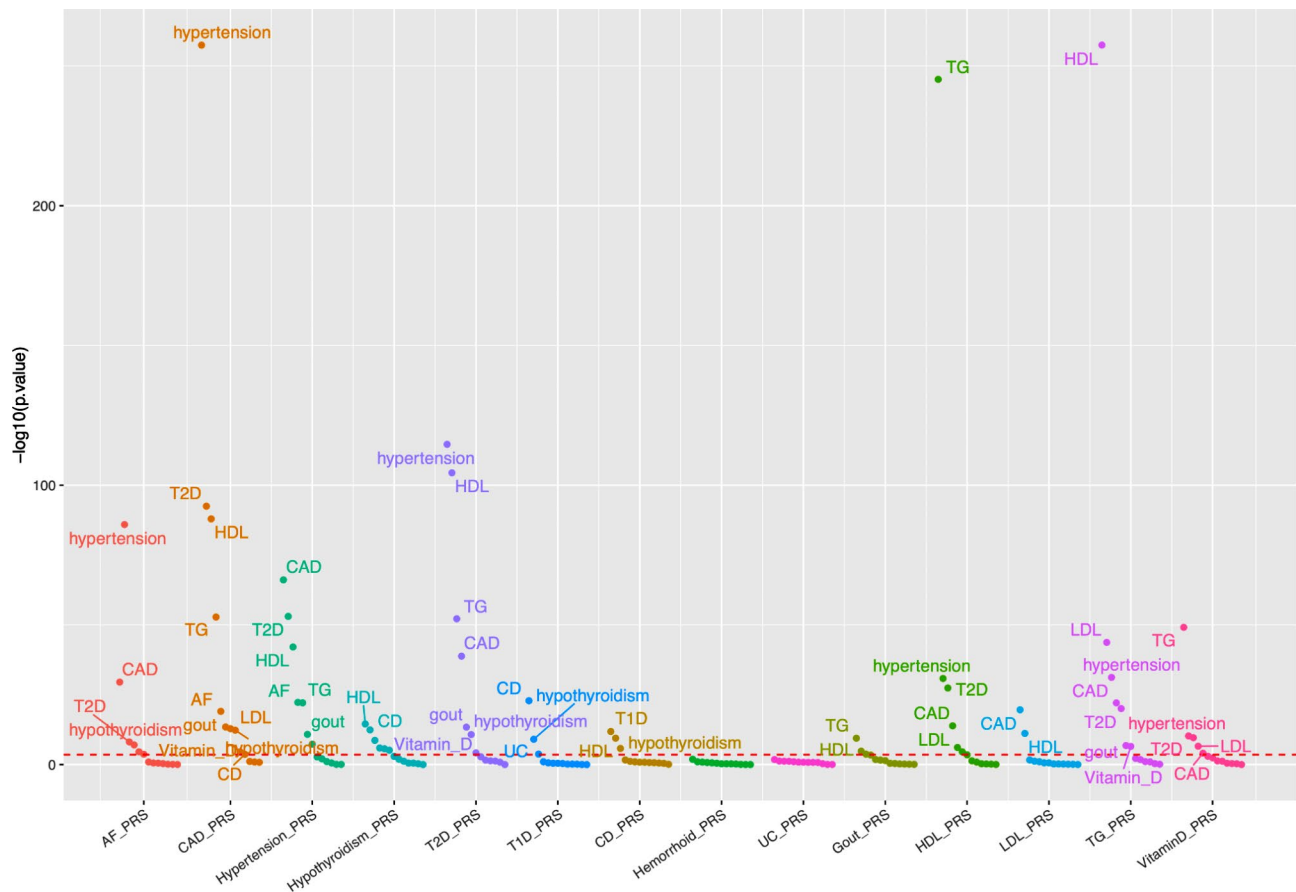


Fig. 4 PRS-phenome wide association for fourteen medical conditions in European individuals. Phenotypes are grouped according to the testing PRS. Red dashed line indicates the Bonferroni correction (2.74×10^{-4})

Our exploration of the prediction accuracy of PRS models in non-European ancestry cohorts has let us identify specific conditions (i.e., CAD, hypertension, hypothyroidism, HDL-C, LDL-C levels and T2D) where the application of PRSs in diverse populations showed more comparable performance to those observed in European groups. Specifically, we found that the PRS scores were more predictive when the selection of the appropriate PRS model for each particular individual was ancestry specific, which necessarily requires ancestry estimation prior to PRS estimation. It should be noted that the top selected models for CAD [22, 23], were generated with genome-wide summary statistics derived from transethnic studies, which agrees with studies demonstrating the increased PRS predictability in multi-ethnic samples [11]. Although our approach does not solve the transferability problem impacting PRSs, we provide an alternative strategy to identify those conditions and ancestries where PRSs could work efficiently.

It should be noted that the gold standard solution to solve portability problems involves increasing the diversity of participants included and analyzed in genetic studies [10, 24]. Training of new PRS models beyond

European ancestries might lead to improved accuracy in genetic effect estimation particularly for variants with higher frequencies in non-European populations. It might also reduce the genetic distance from target and training samples, promoting PRS portability. In parallel, the generation of a new era of methods able to integrate ancestry deconvolution in PRS estimations might also support the generalization of PRS models.

We acknowledge some limitations and future improvements in our strategy. First, we observed disparities in the PRS effect sizes between UKBB and BbofA-GL cohorts for some conditions. In these studies, phenotypes were obtained using different phenotyping strategies - while ICD10 codes were extracted from UKBB, BbofA-GL administered self-reported questionnaires. Although the use of self-reported questionnaires does not offer the same level of granularity as clinical grade information, it is quickly becoming a common ascertainment strategy. In fact, direct-to-consumer genetic companies, which usually use this strategy, collectively account for millions of research participants and actively contribute to genetic studies [25]. Second, in this study we did not train and evaluate PRS predictability in admixed individuals.

Non-European individuals in the cohort are relatively small, which was not sufficient for a deeper analysis in this study. Genomes from admixed individuals consist of haplotypes from more than one ancestry group. It represents an extra-layer of complexity and presents technical limitations in the PRS field, despite offering the opportunity to yield novel genetic findings [6, 26]. Most PRS methods depend on LD reference panels, and computational complexities in representing LD for admixed individuals hamper PRS modeling. Third, the generation of PRS models working for different populations is limited by the availability of combining multiple cohorts with diverse ancestries. Future studies will benefit greatly from the inclusion of more diverse datasets.

While we observed statistically significant improvements for some scores, the generalizability of these improvements to external populations of the same ancestries remains a challenge. Different scores may perform better in other datasets, highlighting the complexity of context-specific score selection. Our future work will focus on expanding validation efforts to include diverse external cohorts to better understand and generalize these findings.

Conclusions

Overall, this study comprehensively evaluates PRS accuracy across a wide range of phenotypes for continental ancestries. Large variability of PRS performances across ancestral groups was observed when the evaluation was conducted using the best-PRS model for Europeans. We observed an improvement in the PRS performances when specifically selecting ancestry specific PRS models, reducing the uncertainty with respect to which PRS model to choose and in which ancestry group, and highlighting the benefit of incorporating ancestry inference in routine PRS evaluation. This approach has let us identify specific conditions with potential to be applied to different ancestries. Our work enables future studies to establish the best practices to evaluate PRS models, democratize the implementation of PRS models across populations and their application to translational research.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40246-024-00664-y>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6

Supplementary Material 7

Supplementary Material 8

Acknowledgements

The authors would like to thank all the participants for being a part of this study.

Author contributions

SMG, MV, ALP designed the study. SMG, ALP, and DMM performed analysis of data. DMM, CQC, BM, MBT, MTML, ZY, KN, YM, JDW, AI, NK, TT, and CDB provided interpretation of the results. SMG, JDW NK and CDB drafted the manuscript. All authors contributed critically, read, revised, and approved the final version.

Funding

This project received partial funding from participating institution. This research is based on results obtained from project JPNP19001, subsidized by the New Energy and Industrial Technology Development Organization (NEDO).

Data availability

The BBoFA-GL cohort that supports the findings of this study is available for qualified researchers at non-profit institutions upon entering into an agreement with Genomelink. All information will be shared subject to the above criterion upon request via info@genomelink.io. We have posted our scripts to replicate our work in the following Github repository: <https://github.com/GalateaBio/PRS-manuscript2024>.

Declarations

Ethics approval and consent to participate

This study was approved by the institutional review board (IRB) at WCG IRB (<https://www.wcgirb.com/>) under IRB tracking number protocol number 20201332.

Competing interests

AI, ALP, BM, CDB, CQC, DMM, MBT, JDW, MTML, NK, and SMG are employees of or consultants to Galatea Bio. MV, ZY, KN, YM, and TT are employees of Genomelink. CDB, IA, and NK are founders and shareholders of Galatea Bio stock. CDB, KN, YM, and TT are shareholders of Genomelink stock. The remaining authors declare that there is no conflict of interest regarding the publication of this article.

Author details

¹Galatea Bio, Inc, 14350 Commerce Way, Miami Lakes, FL 33146, USA

²Department of Biomedical Data Science, Stanford University School of Medicine, 1265 Welch Road, Stanford, CA 94305, USA

³Genomelink, Inc, 2150 Shattuck Avenue, Berkeley, CA 94704, USA

⁴Amphora Health. Batallon Independencia 80, Morelia, Michoacan 58260, Mexico

⁵Escuela Nacional de Estudios Superiores, Unidad Morelia, Universidad Nacional Autonoma de México, Antigua Carretera a Pátzcuaro No. 8701, Col. Ex Hacienda de San José de la Huerta, Morelia, Michoacán C.P. 58190, Mexico

⁶University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA

⁷Japan: Awakens Japan K.K. (Japanese subsidiary of Genomelink, Inc.), 2-11-3, Meguro, Meguro-ku, 1530063 Tokyo, Japan

Received: 15 June 2024 / Accepted: 19 August 2024

Published online: 02 September 2024

References

1. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019;51:584–91.

2. Wray NR, Lin T, Austin J, McGrath JJ, Hickie IB, Murray GK, et al. From Basic Science to Clinical Application of Polygenic Risk scores: a primer. *JAMA Psychiatry*. American Medical Association; 2021. pp. 101–9.
3. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet*. 2017;100:635–49.
4. Privé F, Aschard H, Carmi S, Folkersen L, Hoggart C, O'Reilly PF, et al. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *Am J Hum Genet*. 2022;109:12–23.
5. Bitarello BD, Mathieson I. Polygenic scores for height in admixed populations. *G3 genes*. *Genomes Genet*. 2020;10:4027–36.
6. Ding Y, Hou K, Xu Z, Pimplaskar A, Petter E, Boulier K, et al. Polygenic scoring accuracy varies across the genetic ancestry continuum. *Nature*. 2023;618:774–81.
7. Márquez-Luna C, Loh PR, Price AL, Kooner JS, Saleheen D, Sim X, et al. Multi-ethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol*. 2017;41:811–23.
8. Wall JD, Stawiski EW, Ratan A, Kim HL, Kim C, Gupta R, et al. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*. 2019;576:106–11.
9. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature*. 2019;570:514–8.
10. Martin AR, Stroud RE, Abebe T, Akena D, Alemayehu M, Atwoli L et al. Increasing diversity in genomics requires investment in equitable partnerships and capacity building. *Nat Genet Nat Res*; 2022. p. 740–5.
11. Wang Y, Kanai M, Tan T, Kamariza M, Tsuo K, Yuan K, et al. Polygenic prediction across populations is influenced by ancestry, genetic architecture, and methodology. *Cell Genomics*. 2023;3:100408.
12. Lambert SA, Gil L, Jupp S, Ritchie SC, Xu Y, Buniello A et al. The polygenic score catalog as an open database for reproducibility and systematic evaluation. *Nat Genet Nat Res*; 2021. p. 420–5.
13. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet*. 2015;97:576–92.
14. Privé F, Arbel J, Vilhjálmsson BJ. LDpred2: better, faster, stronger. *Bioinformatics*. 2020;36:5424–31.
15. Ruan Y, Lin YF, Feng YCA, Chen CY, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. *Nat Genet*. 2022;54:573–80.
16. Qian J, Tanigawa Y, Du W, Aguirre M, Chang C, Tibshirani R, et al. A fast and scalable framework for large-scale and ultrahigh-dimensional sparse regression with application to the UK biobank. *PLoS Genet*. 2020;16:e1009141.
17. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
18. Lopez-Pineda A, Vernekar M, Moreno-Grau S, Rojas-Muñoz A, Moatamed B, Lee MTM et al. Validating and automating learning of cardiometabolic polygenic risk scores from direct-to-consumer genetic and phenotypic data: implications for scaling precision health research. *Hum Genomics*. 2022;16.
19. Howrigan DP, Simonson MA, Keller MC. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics*. 2011;12:460.
20. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023;613:508–18.
21. Nagai A, Hirata M, Kamatani Y, Muto K, Matsuda K, Kiyohara Y, et al. Overview of the BioBank Japan Project: study design and profile. *J Epidemiol*. 2017;27:S2–8.
22. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet*. 2022;54:1803–15.
23. Koyama S, Ito K, Terao C, Akiyama M, Horikoshi M, Momozawa Y, et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat Genet*. 2020;52:1169–77.
24. Fatumo S, Chikowore T, Choudhury A, Ayub M, Martin AR, Kuchenbaecker K. A roadmap to increase diversity in genomic studies. *Nat. Med. Nature Research*; 2022. pp. 243–50.
25. Shelton JF, Shastri AJ, Ye C, Weldon CH, Filshtein-Sonmez T, Coker D, et al. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat Genet*. 2021;53:801–8.
26. Gurdasani D, Barroso I, Zeggini E, Sandhu MS. Genomics of disease risk in globally diverse populations. *Nat Rev Genet Nat Publishing Group*; 2019. p. 520–35.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.