

Low and differential polygenic score generalizability among African populations due largely to genetic diversity

Lerato Majara,^{1,2,6} Allan Kalungi,^{1,3,4,5} Nastassja Koen,^{1,6,7} Kristin Tsuo,^{8,9,10,11} Ying Wang,^{8,9,10} Rahul Gupta,^{8,9,10,11} Lethukuthula L. Nkambule,^{8,9,10} Heather Zar,¹² Dan J. Stein,^{6,7} Eugene Kinyanda,⁵ Elizabeth G. Atkinson,¹³ and Alicia R. Martin^{8,9,10,14,*}

Summary

African populations are vastly underrepresented in genetic studies but have the most genetic variation and face wide-ranging environmental exposures globally. Because systematic evaluations of genetic prediction had not yet been conducted in ancestries that span African diversity, we calculated polygenic risk scores (PRSs) in simulations across Africa and in empirical data from South Africa, Uganda, and the United Kingdom to better understand the generalizability of genetic studies. PRS accuracy improves with ancestry-matched discovery cohorts more than from ancestry-mismatched studies. Within ancestrally and ethnically diverse South African individuals, we find that PRS accuracy is low for all traits but varies across groups. Differences in African ancestries contribute more to variability in PRS accuracy than other large cohort differences considered between individuals in the United Kingdom versus Uganda. We computed PRS in African ancestry populations using existing European-only versus ancestrally diverse genetic studies; the increased diversity produced the largest accuracy gains for hemoglobin concentration and white blood cell count, reflecting large-effect ancestry-enriched variants in genes known to influence sickle cell anemia and the allergic response, respectively. Differences in PRS accuracy across African ancestries originating from diverse regions are as large as across out-of-Africa continental ancestries, requiring commensurate nuance.

Introduction

Genome-wide association studies (GWASs) have yielded important biological insights into the heritable basis of many complex traits and diseases.¹ However, the vast majority of studies have been conducted in populations of European descent, potentially limiting generalizability across diverse populations.^{2–6} Genome-wide significant SNP associations with phenotypes spanning a wide range of genetic architectures have consistently replicated across populations in both direction and effect size, with few examples of heterogeneous effect sizes.^{7–9} However, previous studies that have compared the association between genetically predicted versus measured phenotypes in diverse populations using polygenic risk scores (PRSs) have found that PRS accuracy decreases with increasing genetic distance between the GWAS discovery and PRS target cohorts.^{4,10,11} This seeming paradox highlights that while

variant-level associations consistently replicate across populations, genome-wide aggregate measures are more predictive but less generalizable.¹² Since the earliest applications of PRS in human genetics, these concepts—coupled with Eurocentric study biases—have resulted in PRSs that are most accurate in European ancestry populations and least accurate in African ancestry populations.¹³ These study biases and phenomena continue to replicate a decade later, with several-fold differences in prediction accuracy of many traits between European and non-European ancestry populations.⁴

Quantifying PRS generalizability within and among African populations requires considerable nuance, as they represent the most genetically diverse populations globally, with more than a million more genetic variants per person than out-of-Africa populations.¹⁴ Populations collected even within the same geographic regions of Africa have complex demographic histories with

¹Global Initiative for Neuropsychiatric Genetics Education in Research (GINGER), Harvard T.H. Chan School of Public Health, Department of Epidemiology, Boston, MA, USA; ²MRC Human Genetics Research Unit, Division of Human Genetics, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Observatory 7925, South Africa; ³Department of Psychiatry, College of Health Sciences, Makerere University, Kampala, Uganda; ⁴Department of Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa; ⁵Mental Health Project, Medical Research Council/Uganda Virus Research Institute (MRC/UVRI) & London School of Hygiene and Tropical Medicine (LSHTM), Uganda Research Unit, Entebbe, Uganda; ⁶Department of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, South Africa; ⁷South African Medical Research Council (SAMRC) Unit on Risk and Resilience in Mental Disorders, Cape Town, South Africa; ⁸Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA; ⁹Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; ¹⁰Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; ¹¹Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA 02115, USA; ¹²Department of Paediatrics and Child Health, Red Cross Children's Hospital and Medical Research Council Unit on Child and Adolescent Health, University of Cape Town, Cape Town, South Africa; ¹³Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

¹⁴Lead contact

*Correspondence: armartin@broadinstitute.org

<https://doi.org/10.1016/j.xhgg.2023.100184>.

© 2023 The Authors. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



complicated patterns of admixture and population structure.^{15–18} Further, African ancestry populations experience vastly different environments within versus outside continental Africa as well as more locally among diverse communities, countries, and regions of Africa. These differences provide unique epidemiological opportunities to query the impacts of vastly differing environments on PRS accuracy. Previous empirical analyses and theoretical work fundamentally informs how demographic history and environmental variation interplay to produce PRS heterogeneity in traditionally underserved populations.^{19–23}

The inclusion of African ancestry participants in large-scale genetic studies is uniquely important for many reasons. They have the lowest life expectancies globally,^{24,25} receive the lowest access to and quality of medical care in the United States,²⁶ and are the most underserved by genetic technologies.^{6,27} A more nuanced understanding of PRS transferability will critically inform which populations are currently the most underserved and thus where building genetic studies and resources will have the biggest benefits globally.

There are also clear benefits to including African populations in statistical genetics efforts. Because humans originated in Africa, populations from Africa have the most genetic diversity among global populations,^{14,28,29} such that more genotype-phenotype associations are expected in Africa than can be found elsewhere. African American individuals have been shown to contribute disproportionately to GWAS findings,² making up 2.8% of GWAS participants but contributing 7% of trait associations. African ancestry populations also have shorter blocks of linkage disequilibrium, which improves resolution to fine-map causal variants.³⁰ PRS accuracy is lowest in African ancestry populations due to GWAS study biases,⁴ but when GWASs include these and other diverse populations, PRS predict traits such as schizophrenia more accurately across all populations compared with single-ancestry GWASs.³¹

In this study, we have investigated how PRSs generalize within and among diverse African populations in simulations and with empirical genotype-phenotype data for dozens of quantitative traits. We first simulated genetic effects and computed genetic risk prediction accuracy using data from two African datasets: the African Genome Variation Project (AGVP) and the Africa Wits-INDEPTH partnership for Genomic Studies (AWI-Gen) project. We then calculated PRS using publicly available GWAS summary statistics from predominantly European ancestry populations to (1) quantify PRS accuracy for five physical and psychosocial traits among populations in the Drakenstein Child Health Study (DCHS) of South Africa, a birth cohort study; and (2) compare PRS accuracy for 34 quantitative traits across the Ugandan General Population Cohort (GPC) versus ancestrally diverse UK Biobank participants. Our results highlight the disproportionate benefits of genetic studies in diverse African populations to improve trait prediction. Further, while PRSs hold promise as biomarkers in precision medicine, a critical prerequisite

is equitable accuracy in diverse populations to avoid exacerbating existing health disparities.

Materials and methods

Genetic and phenotypic data

Total counts of individuals by population and/or study are shown in [Table S2](#).

1000 Genomes Project

1000 Genomes Project data from the phase 3 integrated call set were accessed and used as a reference panel and for phasing and imputation.¹⁴

Human Genome Diversity Project

Genotype data for samples from the Human Genome Diversity Project (HGDP) was publicly available on the Illumina HumanHap650K GWAS array on hg18.³² We lifted over the genotype data to the hg19 genome build using [hail \(http://hail.is\)](http://hail.is).

African Genome Variation Project

As described previously,³³ the AGVP data consist of dense genotype data from 1,481 individuals from 18 ethnolinguistic groups from Eastern, Western, and Southern Africa when including the Luhya and Yoruba from the 1000 Genomes Project.¹⁴ When accessed from the European Genome-Phenome Archive (EGA:E-GAD00010001047), “Ethiopian” is the provided population label encompassing the Oromo, Amhara, and Somali groups. After collapsing these groups and counting the 1000 Genomes data separately, 1,307 individuals from 14 populations are uniquely represented in the AGVP, and 2,504 individuals from 26 populations are represented in the 1000 Genomes Project data (661 individuals from seven populations are in the AFR super population grouping).

Africa Wits-INDEPTH partnership for genomic studies

AWI-Gen is a study that investigates the relationship between genetics and the environment in causing cardiometabolic disease in sub-Saharan Africa with study participants from Burkina Faso, Ghana, Kenya, and South Africa.^{34,35} It is a partnership between the University of Witwatersrand (Wits) in Johannesburg, South and the International Network for the Demographic Evaluation of Populations and Their Health (INDEPTH). Ethical approval was obtained from the Human Research Ethics Committee of the University of the Witwatersrand (Protocol Number: M121029 and from the institutions of the respective centres that are in the international network.

Drakenstein Child Health Study in South Africa

The DCHS is an ongoing, multidisciplinary population-based birth cohort study in the Drakenstein area in Paarl (outside Cape Town, South Africa), that obtained ethical approval from the Faculty of Health Sciences Research Ethics Committee at the University of Cape Town (401/2009) and the Western Cape Provincial Research committee.^{36–38} After providing informed consent, pregnant women were enrolled during their second trimester (20–28 weeks’ gestation); maternal-child dyads were then followed through childbirth and longitudinally thereafter. Enrollment occurred from March 2012 to March 2015 at two primary health care clinics: TC Newman (serving a predominantly mixed ancestry population) and Mbekweni (serving a predominantly Black African population). Women were eligible to participate in the DCHS if they attended one of the study clinics, were at least 18 years of age, and intended to remain residing in the study area. Ancestral diversity computed using PCA with genetic data is shown in [Figure S5](#) with corresponding self-reported ethnicity (“Mixed” versus “Black/African”).

Uganda General Population Cohort

The rural Uganda GPC of MRC/UVRI & LSHTM Uganda Research Unit was set up in 1989 initially to monitor the HIV epidemic among adults, children, and adolescents, after obtaining ethical approval the Uganda Virus Research Institute Science and Ethics committee and the Ugandan National Council of Science and Technology. Its mandate has since expanded to include other medical conditions.³⁹ The “original GPC” is located in the sub-county of Kyamulibwa in rural southwestern Uganda with activities having recently been expanded to the neighboring two peri-urban townships of Lwabenge and Lukaya. The “original GPC” includes about 10,000 adults and about 10,000 children and adolescents. In 2011, genotype data were generated on more than 5,000 adult participants from nine ethnolinguistic groups using the Illumina HumanOmni2.5 BeadChip at the Sanger Wellcome Trust Institute.^{39,40}

UK Biobank

The UK Biobank enrolled 500,000 people aged between 40 and 69 years in 2006–2010 from across the country, as described previously.⁴¹ A more detailed description of the cohort is available on their website: <https://www.ukbiobank.ac.uk/>. We analyzed phenotypes that overlapped with those studied in the Uganda GPC.

Ancestry analysis in the UK Biobank

As described previously,⁴¹ the UK Biobank consists of approximately 500,000 participants of primarily European ancestry who have thousands of measured or reported phenotypes. To assess polygenic score accuracy across diverse ancestries, we identified populations of ancestral groups at two levels: (1) among continental groups, and (2) among regions in Africa. To define continental ancestries, we first combined reference data from the 1000 Genomes Project and HGDP. We combined these reference datasets into continental ancestries according to their corresponding meta-data (Table S5). We then ran PCA on unrelated individuals from the reference dataset. To partition individuals in the UK Biobank based on their continental ancestry, we used the PC loadings from the reference dataset to project UK Biobank individuals into the same PC space. We trained a random forest classifier given continental ancestry meta-data (AFR = African, AMR = admixed American, CSA = Central/South Asian, EAS = East Asian, EUR = European, and MID = Middle Eastern) based on the top six PCs from the reference training data. We applied this random forest to the projected UK Biobank PCA data and assigned initial ancestries if the random forest probability was >50% (similar results obtained for $p > 0.9$), otherwise individuals were dropped from further analysis.

Next, we further partitioned African ancestry individuals using the same random forest approach as above but without further probability thresholding using African ancestry reference data from AGVP, HGDP, and the 1000 Genomes Project. We partitioned these reference data into UN regional codes with an additional region for Ethiopian populations given their unique population history and collapsing in AGVP data (Admixed, Central, East, Ethiopia, South, and West Africa), as shown in Table S5. PCA with reference data at the continental and subcontinental level within Africa are shown in Figures S10 and S11.

Phasing and imputation

We used the Ricopili pipeline to conduct pre-imputation quality control (QC) and perform phasing and imputation for AGVP and the Uganda GPC.⁴² This pipeline was also used on the DCHS data, as described previously.⁴³ Briefly, we phased the data

using Eagle 2.3.5 and imputed variants using minimac3 in chunks ≥ 3 Mb. The 1000 Genomes phase 3 haplotypes were used as the reference panel for phasing and imputation. For the AGVP, we used strict best guess genotypes where a variant was called if it had a probability of $p > 0.8$ and a missing rate less than 0.01 and MAF >5%. Then, variants with MAF <0.001 were excluded from the dataset. For Uganda GPC, we used combined best guess genotypes where a variant was called if it had a probability $p > 0.8$ or set to missing otherwise. Then, SNPs were filtered to keep sites with missingness <0.01 and MAF >0.05. We used genotype dosages when computed PRS.

PCA

Only SNPs with high imputation quality (INFO >0.8) were considered for principal-component analysis. We computed the first 20 principal components using plink with the `-pca` flag for autosomal SNPs MAF >0.05 and individual missingness <0.05.

Simulation setup

We used two independent simulation strategies for two African datasets: AGVP and AWI-Gen. The choice of simulation strategy was informed by the sample size of each dataset. We used the simulation strategy previously used by Scutari et al.¹¹ for AGVP, and the infinitesimal model for simulations in AWI-Gen. To test the PRS prediction accuracy within and across African populations, we simulated four quantitative traits while varying heritabilities ($h^2 = 0.1, 0.2, 0.4$ and 0.8) for both AGVP and AWI-Gen as follows.

AGVP

We randomly assigned an effect size to 5, 20, 100, 2,000, 10,000, and 50,000 causal variants, respectively. We then calculated an individual's “true” polygenic risk as the sum of all causal effects using the `-score` flag in PLINK v1.07B.⁴⁴ True polygenic scores were standardized to a mean of zero and standard deviation of 1. To account for the contribution of environmental risk factors, we assigned environmental effects from a normal random distribution (mean = 0 and SD = 1). The phenotype was generated according to its heritability as the weighted sum of the true polygenic risk and a random environmental effect as below:

$$\text{phenotype} = h^2 \times \text{true polygenic risk} + (1 - h^2) \times \text{environmental effect}$$

We then conducted GWAS for the simulated phenotypes by splitting the AGVP dataset into three groups broadly representing the three geographical areas from which samples were obtained: East ($n = 589$), West ($n = 517$), and South Africa ($n = 186$, Figure 1). To allow for the quantification of PRS prediction accuracy across the geographical regions, each group was further split into discovery and target cohorts. The size of the target cohorts was maintained at $n = 186$ across all groups, while the discovery cohort consisted of all remaining individuals (East $n = 403$, West $n = 331$, and no South Africans). We conducted a linear regression for all the simulated traits for the East and West discovery datasets, controlling for the first 20 principal components.

AWI-Gen

We assigned genetic effects to variants based on their minor allele frequency. The effects were calculated based on the relationship between effect size and minor allele frequency as shown by Schoech et al.⁴⁵ The “true” individual's polygenic risk was calculated in the same way as it was for AGVP, so was the environmental risk factor and the phenotypes. To be able to conduct GWAS, we split AWI-Gen into discovery and target sets such that each discovery set

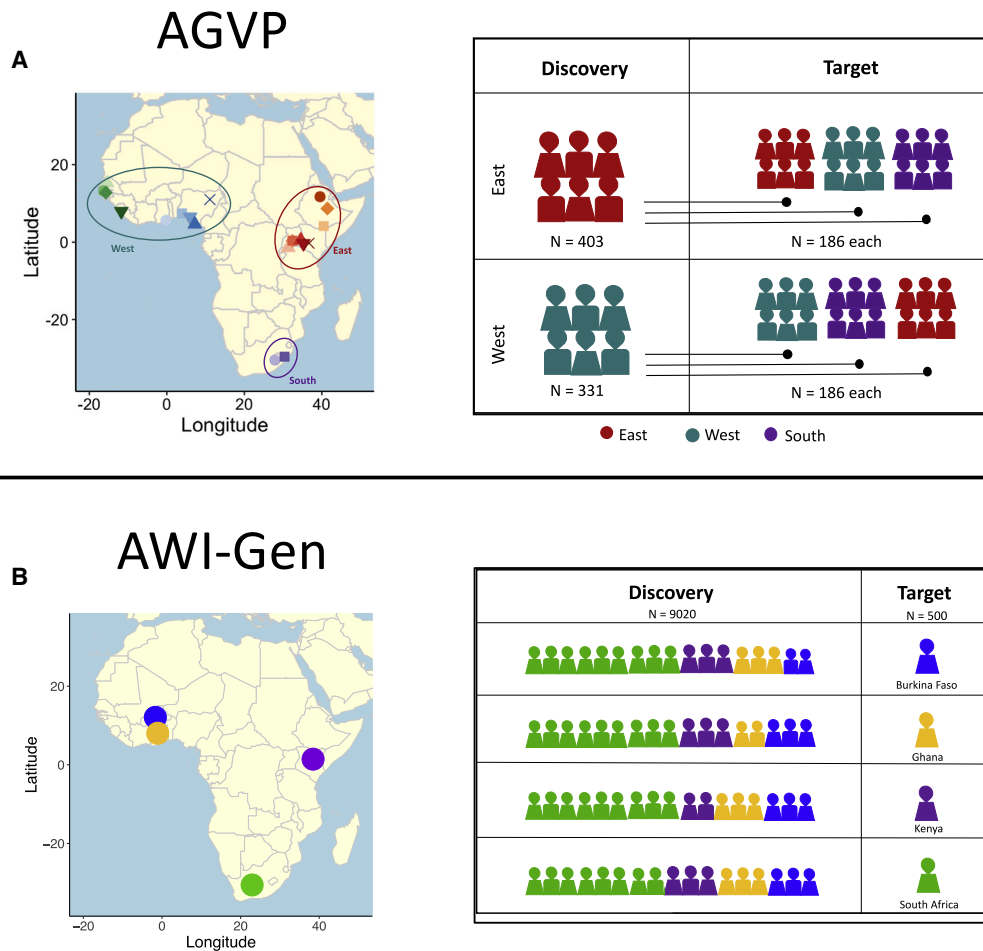


Figure 1. Simulation strategy overview

(A) We used AGVP for simulations in West, East, and South African populations that were grouped based on the United Nations geoscheme groupings. Each group was divided into discovery and target subgroups. GWAS discovery cohorts included East ($n = 403$) and West ($n = 331$) African individuals, which were independent of each target cohort ($n = 186$ individuals per region). South African individuals were excluded from the discovery population due to the limited total sample size (two populations and 186 individuals total).

(B) We used AWI-Gen for simulations in Burkina Faso ($n = 1,703$), Ghana ($n = 1,661$), Kenya ($n = 1,701$), and South Africa ($n = 4,455$). For these simulations we withheld 500 individuals from each of the groups, which were used as the target cohort. The GWAS discovery cohort included the 9,020 individuals who were not in the target cohort. Each figure represents roughly 500 individuals. BF, Burkina Faso; SA, South Africa.

had 9,020 samples and each target 500 (Figure 1). For each discovery-target split, we alternately withheld 500 samples from one of the four countries (Burkina Faso, Ghana, Kenya, and South Africa). We then conducted GWAS for each of the discovery datasets.

For each discovery cohort in AGVP and AWI-Gen, we obtained independent SNP sets by clumping SNPs from corresponding summary statistics files with an r^2 value greater than 0.1 using in-sample linkage disequilibrium (LD) and within 500 kb of each other in PLINK v1.07, were obtained for each discovery cohort. The effect sizes from these SNP sets were used as weights to compute PRSs for all corresponding target datasets for a range of p values (5e-08, 1e-06, 1e-04, 1e-03, 1e-02, 0.05, 0.1, 0.2, 0.5, and all). PRS was calculated as the sum of all SNPs multiplied by their effect sizes.

Heritability estimation

For the first set of heritability estimation analyses, we relied on heritability estimates of 34 quantitative traits computed previously for the Ugandan GPC dataset.⁴⁶ For the UK Biobank, we computed heritability estimates for the same traits using LD score regression with

the default model (i.e., without any functional annotations)⁴⁷ and using population-matched LD score references from European populations downloaded from the authors' website (<https://data.broadinstitute.org/alkesgroup/LDSCORE/>). Due to the difference in study design and heritability estimation methods used for UK Biobank and Uganda GPC, we could not directly compare heritability estimates between the two cohorts. For more comparable estimates, we computed heritability estimates across the 34 quantitative traits in both the Ugandan GPC and UK Biobank using a randomized multi-component Haseman-Elston estimator (RHE-mc⁴⁸) over unrelated individuals. This method offers improved power over summary statistic methods (e.g., LD score regression), which show very large standard errors with small sample sizes, with improved computational tractability when operating at biobank scale over GREML-based approaches.

To improve comparability across cohorts, we used parallel QC approaches in both cohorts after restricting to unrelated individuals ($N = 2,234$ in GPC). Namely, in UK Biobank, we filtered genotypes to those outside the MHC region that were defined with

MAF ≥ 0.01 for which we did not observe significant deviation from Hardy-Weinberg equilibrium ($p_{\text{hwe}} \geq 1e-7$) and only used genotypes that passed these criteria across all ancestry groups. For the Ugandan GPC analysis, we applied the same filters as above (i.e., variants outside MHC with MAF ≥ 0.01 and without significant deviation from Hardy-Weinberg equilibrium). We also restricted the analysis to SNPs with imputation INFO score > 0.3 to match the approach taken for the GWAS conducted by Gurdasani et al.,⁴⁶ resulting in 3,627,507 SNPs passing QC.

To account for differences in heritability as a function of LD and MAF, we performed multi-component analyses with a 4×2 grid of LD and MAF bins defined in Ugandan GPC and per-population in UK Biobank. LD scores in the Ugandan GPC were computed in LD score regression using all imputed SNPs with MAF > 0.005 from unrelated samples, while those in the UK Biobank were computed using SNPs with INFO > 0.8 , MAC > 20 with subsequent covariate correction for age, sex, age*sex, age², age²*sex, and the first 20 genotype PCs in each population. LD score bins in both cohorts were computed as membership in quartiles of the LD score distribution. MAF bins in both cohorts were defined as MAF ≤ 0.05 and MAF > 0.05 .

We included standard GWAS covariates as fixed-effects covariates for heritability estimation in both Ugandan GPC and UK Biobank, namely age, sex, age*sex, age², age²*sex, and the first 20 genotype PCs in each cohort and in each population in the UK Biobank. We ran RHE-mc with 50 random vectors and 100 jackknife blocks in the Uganda GPC and among UK Biobank non-EUR populations; and used 10 random vectors with UK Biobank European samples due to high computational complexity.

Polygenic score calculation

Pruning and thresholding

All PRSs were calculated in plink2 or in hail using custom scripts. For pruning and thresholding approaches, all clumping was done in plink2 using an LD threshold of $r^2 = 0.1$ and a window size of 500 kb with discovery cohort population-specific reference panels. We calculated PRS using plink2 with the `-score` and `-q-score-range` flags for AGVP simulations and DCHS. We wrote custom scripts in hail (<http://hail.is>) to calculate PRS in the Uganda GPC and UK Biobank data due to the larger sample sizes (see [web resources](#)). For imputed genotypes, we used SNP dosages in PRS calculations. We computed 10 PRSs for each analysis using the following p-value thresholds: 1, 0.5, 0.2, 0.1, 0.05, 0.01, 1e-3, 1e-4, 1e-6, 5e-8. The PRS that explained the most phenotypic variance is shown in most figures.

We calculated PRS accuracy for continuous traits computed with custom scripts in R ([Web resources](#)). For AGVP simulations and DCHS (because all participants were mothers of a similar age), we included the first 10 PCs as covariates when computing the partial R^2 specifically attributable to the PRS. For Uganda GPC data, we included age, sex, and the first 10 PCs when computing partial R^2 of the PRS. For consistency with the GWAS that were run in UK Biobank previously⁴⁹ and here with a holdout target set, we included, age, sex, age², age*sex, age²*sex, and the first 10 PCs as covariates when computing the PRS partial R^2 . (The UK Biobank European GWAS included 20 PCs, but fewer were used here due to the particularly small sample sizes of some other target ancestry groups, [Table S5](#), coupled with minimal population structure observed in PCs lower than PC10.) As described in [Table S5](#), we included up to 351,194 European ancestry participants in a GWAS, withholding up to 9,947 European ancestry participants as a target cohort as well as up to the following numbers of partic-

ipants with corresponding ancestries: 8,426 African, 1,099 Admixed American, 10,084 Central/South Asian, 2,753 East Asian, and 1,553 Middle Eastern individuals. Other participants in the UK Biobank but not included in these analyses had either second-degree relatives or closer with participants included in analysis or were ancestry outliers.

PRS accuracy evaluation (incremental R^2)

To evaluate prediction accuracy, we calculated incremental R^2 . Specifically, we compared two models:

H₀: Phenotype \sim covariates.

H₁: Phenotype \sim PRS + covariates.

The incremental R^2 calculates the change in R^2 between H₁ and H₀, indicating the change in model accuracy attributable to the PRS. We used adjusted R^2 , which ensures that the model that includes PRS does not outperform the model without PRS simply because more terms were included. All error bars show 95% confidence intervals calculated from bootstrapping. Specifically, for each iteration of 100 bootstrap replicates, we resampled with replacement each individual's full set of phenotypes, covariates, and polygenic scores, then ran the same models described above. The 95% confidence interval was determined by the 2.5% and 97.5% quantiles.

Relative comparisons of PRS across populations

We compared PRS accuracy across populations by computing relative accuracies (RAs) with respect to a baseline European ancestry PRS. For pruning and thresholding PRS, we computed RA as the ratio between the maximum R^2 in the population of interest versus the maximum R^2 in the European baseline comparison (i.e., for the same phenotype). Across traits, we computed median absolute deviation (MAD), i.e., the median of the absolute deviations from the median.

PRS-CS versus pruning and thresholding

We compared the prediction accuracy of the pruning and thresholding method to that of PRS-CS, a Bayesian method that has been shown to improve PRS prediction accuracy across diverse populations.⁵⁰ To do this, we applied PRS-CS-auto to generate scores for the same 34 quantitative traits that were evaluated using pruning and thresholding. We maintained the same discovery cohort from the UK Biobank, i.e., 351,194 European ancestry individuals and evaluated prediction accuracy in two target cohorts: (1) continental ancestry groups from the UK Biobank comprising 9,947 European ancestry holdout sample, as well as $\sim 24,000$ non-European ancestry individuals, and (2) the Ugandan GPC. We used European ancestry from the UK Biobank as the reference panel. Relative accuracy was calculated as the ratio between the maximum R^2 for pruning and thresholding or R^2 for PRS-CS versus the maximum R^2 in the European population for each trait.

Observed versus predicted PRS accuracy

To evaluate the efficacy of PRS accuracy given the varying genetic architecture of the quantitative traits we assessed here, we compared the PRS accuracy we observed with the accuracy that would be predicted from theoretical models.^{51,52} We calculated the predicted PRS accuracy for the European ancestry individuals from the UK Biobank according to the Daetwyler equation below, where E is the predicted accuracy, h^2 is the heritability estimates, M is the number of independent SNPs (i.e., total number of trait-associated SNPs from LD clumping with p-value < 1), and N is the sample size ($\sim 350,000$ individuals).

$$E(R^2) \approx \frac{h_M^2}{1 + \frac{M}{N h_M^2}}$$

Meta-analysis

We used plink2 to conduct inverse variance-weighted meta-analysis across GWAS summary statistics with the `-meta-analysis` option.

LD reference panels and clumping

All PRS calculations required an LD panel for clumping. Our analyses used in-sample LD where feasible and reference panel data as a proxy with ancestry matching from the 1000 Genomes Project phase 3 data when individual-level data were unavailable. We weighted the ancestral representation of each population per trait matching at the continental level. We matched individuals as follows.

Cohort	1000 Genomes phase 3 reference data
BBJ	East Asian (EAS)
UK Biobank	European (EUR)
Uganda Genome Resource (UGR)	African (AFR)
PAGE	Proportional weighting of AFR, EAS, AMR (depending on trait, see Table S6 description for more detail)

We then used the maximal number of individuals available when weighting proportionally to construct this reference panel. For example, in the meta-analysis of height across the UK Biobank, Biobank Japan (BBJ), and Population Architecture Using Genomics and Epidemiology (PAGE) cohorts, UK Biobank has the largest sample size in the discovery cohort ($n = 350,353$), so all Europeans from 1000 Genomes were included in the reference panel ($n = 503$), then a random sampling of EAS, AFR, and AMR individuals were included proportionally to the overall diversity of the discovery cohorts in the meta-analysis.

Results

Our study uses both simulation-based and empirical approaches to evaluate the generalizability of PRS across diverse African ancestry populations. Abbreviations are in [Table S1](#), and a summary of datasets used in this study is shown in [Table S2](#).

Simulated generalizability within and across diverse African populations

We used two separate simulation strategies for AGVP and AWI-Gen depending on their sample sizes ([Figure 1](#)). Given the limited sample size of the AGVP dataset, we opted to use the strategy previously used by Scutari et al.¹¹ We simulated several quantitative traits with varying numbers of causal variants ($n = 5; 20; 100; 2,000; 10,000; \text{ and } 50,000$) and heritability rates ($h^2 = 0.1, 0.2, 0.4, \text{ and } 0.8$), then conducted independent GWASs for each scenario in East and West African ancestry populations ([materials and methods, Figures S1–S4](#)). We calculated the prediction accuracy for PRSs derived from the GWAS summary statistics considering 10 different p-value thresholds within and

across independent target populations from East, West, and South Africa. In general, ancestry-matched results with the sparsest and most heritable genetic architectures produced the highest prediction accuracy. As expected, prediction accuracy was highest with trait $h^2 = 0.8$ and fewer than 100 causal variants ([Figure 2A](#)), as indicated by the highest R^2 and the identification of genome-wide significant associations. Conversely, when the number of causal variants exceeded 100, prediction accuracy was negligible ([Figure S4](#)), as evidenced by no variants meeting the genome-wide significance threshold (i.e., $p < 5e-08$). Prediction accuracy was highest with 5 and 20 causal variants ([Figure 2A](#)). The within-ancestry prediction at p-value threshold $< 5e-08$ and five causal variants were as follows: $R^2 = 0.86$, $p = 1.74 \times 10^{-74}$ for East discovery - East target scores; $R^2 = 0.85$, $p = 9.9e-74$ for West discovery - West target scores. We observed lower prediction accuracy with ancestry-mismatched discovery versus target cohorts at five causal variants and p-value threshold = $1e-6$ ($R^2 = 0.66$, $p = 1.79e-42$ for West discovery - West target scores, compared with $R^2 = 0.53$, $p = 1.29e-74$ for East discovery - West target scores). The scores in the South target sample were comparable when using East- or West-derived summary statistics ($R^2 = 0.86$, $p = 5.19e-84$ for West-derived summary statistics, and $R^2 = 0.86$, $p = 1.35e-83$ for East-derived summary statistics).

For AWI-Gen, we used the commonly used infinitesimal simulation strategy for quantitative traits. We simulated quantitative traits by assigning genetic variant effects based on their minor allele frequency in accordance with the relationship between effects and minor allele frequency established by Schoech et al.⁴⁵ We varied the trait heritability rates similar to the analysis done with AGVP and conducted GWASs for each trait ([materials and methods, Figure 1](#)). We calculated the prediction accuracy for PRS derived from the GWAS summary statistics considering 10 different p-value thresholds, as before, across independent target populations from Burkina Faso, Ghana, Kenya, and South Africa. Across heritability rates and target datasets, the PRS prediction accuracy was low and had confidence intervals that included zero ([Figure 2B](#)). The lack of PRS transferability in AGVP and AWI-Gen for traits with a polygenic architecture using two independent simulation strategies highlights that large-scale genetic studies in African populations are required to accurately predict phenotypes using genetic data and facilitate a better understanding of how PRS might transfer across African populations given the genetic diversity on the continent.

PRS accuracies in South African populations

While our simulations have shown that PRSs generalize poorly across Africa due to substantial genetic diversity and differences across the continent, there is also considerable genetic and environmental diversity within regions and countries. We quantified PRS accuracy for a range of measured phenotypes in mothers genotyped in the

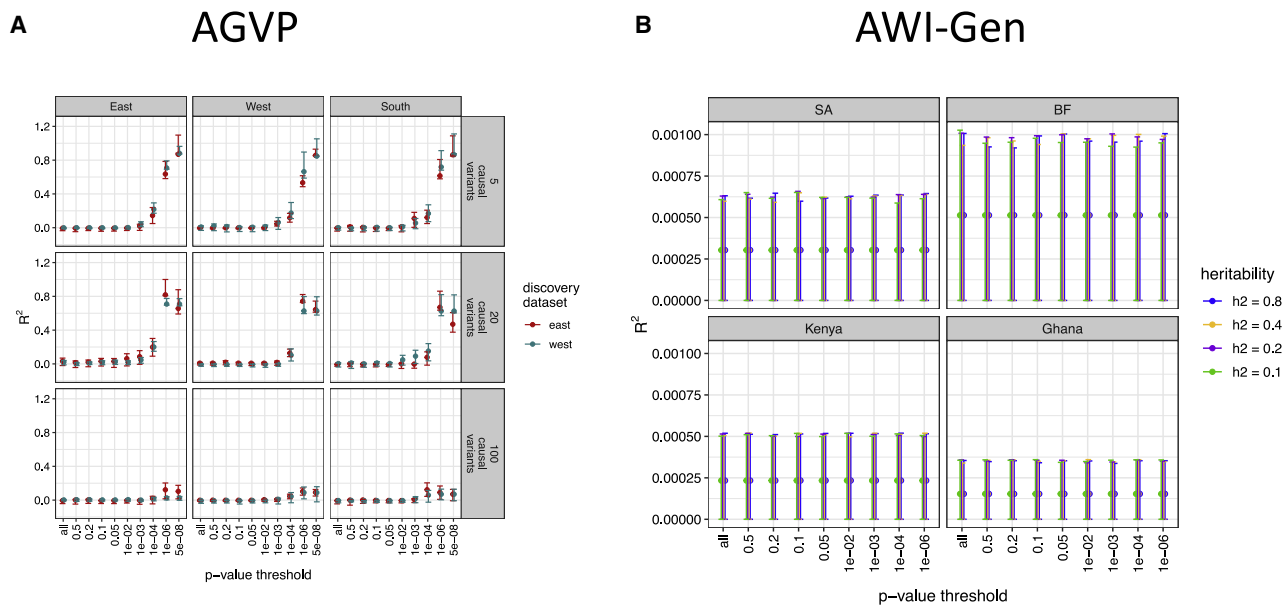


Figure 2. Simulated GWAS and polygenic scores indicate differential prediction accuracy across diverse regions of Africa
 (A) Predictive accuracy of the simulated quantitative trait in AGVP at the heritability of 0.8. The predictive accuracy was calculated for six categories of causal variants for the West and East discovery cohorts, across 10 p-value thresholds. Only the top three categories are shown here, the rest can be found in [Figures S1–S4](#).
 (B) Predictive accuracy of simulated quantitative traits in AWI-Gen for various trait heritability rates across 10 p-value thresholds. The error bars represent the lower and upper limits of 95% confidence interval.

DCHS cohort in South Africa, including several sociodemographic, physical/biomedical, and psychosocial risk traits ([Table S3](#)). The DCHS cohort consists of participants with multiple ancestry groups that include an admixed population with ancestry from multiple continents as well as a population with almost exclusively African population. These ancestry groups correlate with self-reported “Mixed” and “Black/African” ethnicities, respectively ([Figure S5](#)). We computed PRS for maternal height, depression, psychological distress, alcohol consumption, and smoking in DCHS overall, by ethnic group, and by ancestry within the Mixed ethnic group ([materials and methods](#)).

Across all genetically predicted phenotypes, only height was significantly predicted ([Figure S6](#)). We predicted height more accurately in the Mixed versus Black/African ethnic groups ($R^2 = 0.099$, 95% bootstrapped CI = 0.012–0.18, $p = 1.5e-7$ versus $R^2 = 0.021$, 95% CI = -0.031 to 0.043, $p = 5.27e-3$, respectively). We also expect that PRS accuracy increases with decreasing African ancestry within the Mixed ethnic group as has been shown previously in admixed African populations⁵³; we find suggestive evidence consistent with this trend when partitioning the Mixed group into two bins along PC1 ($R^2 = 0.091$, 95% CI = -0.04 to 0.17, $p = 6.4e-4$ in lower half of PC1 with more African ancestry versus $R^2 = 0.12$, 95% CI = $-9.0e-4$ to 0.21, $p = 5.7e-5$ with more out-of-Africa ancestry), although small sample sizes limit definitive comparisons ($n = 137$ in each PC1 bin). Our results are consistent with variable prediction accuracy among diverse African ancestry groups within South Africa and insignificant pre-

diction in African populations for all but the most heritable and accurately predicted traits elsewhere. Notwithstanding these findings, the sample used for these analyses is relatively small and does not represent the larger South African population and some of the traits are greatly impacted by pregnancy, for example, pregnant women are less likely to drink and smoke than the general public. In addition, in contrast to the discovery datasets that include males and females, DCHS is a female-only cohort that has both a lower and narrower age range, which could impact PRS accuracy for the traits where age plays a key role.

Variable phenotypic and genetic similarities across the Uganda GPC and UK Biobank

Lower phenotypic correlations in the Uganda GPC suggest higher contributing environmental effects

We next investigated phenotypic similarities within and across the Uganda GPC and UK Biobank participants because these are two of the largest cohorts with dozens of traits measured in African ancestry individuals. We first considered overall cohort differences between these cohorts: the Uganda GPC enrolled participants using a house-to-house study design and generated genetic data on 5,000 adults from rural villages in southwestern Uganda,³⁹ while the UK Biobank enrolled 500,000 people aged between 40 and 69 years in 2006–2010 from across the country ([materials and methods](#)⁴¹). Previous studies have reported higher rates of infectious diseases (e.g., HIV, hepatitis B and C) in the Uganda GPC than would be expected in the UK Biobank.³⁹ There are many

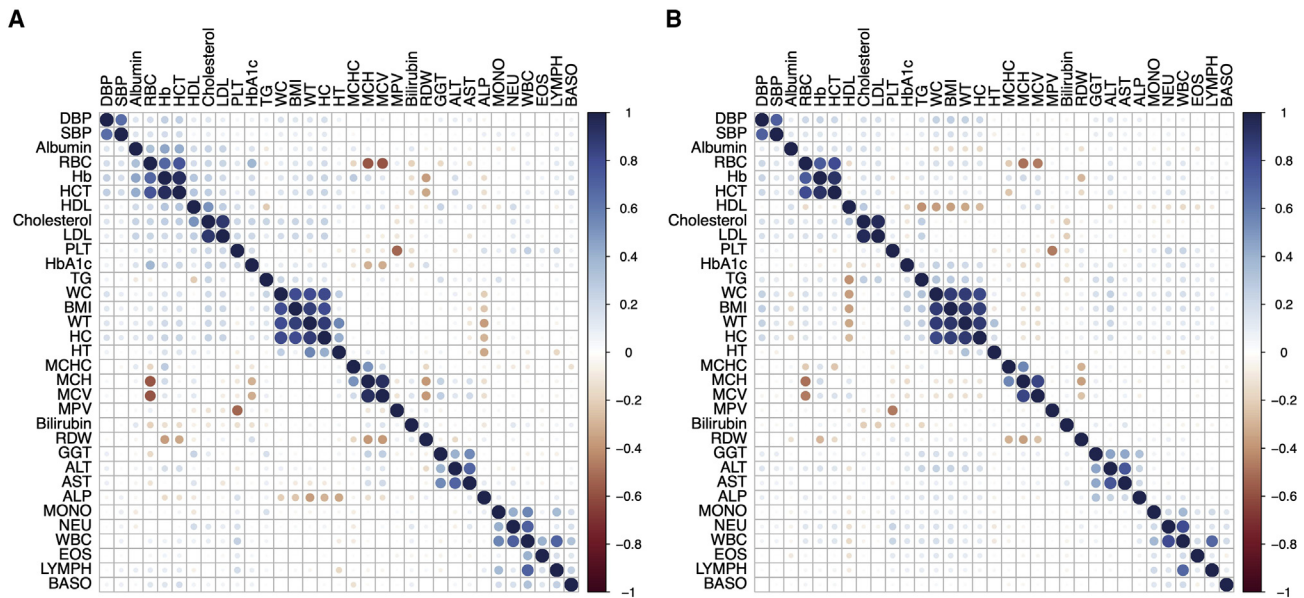


Figure 3. Phenotypic correlations among 33 quantitative traits measured in the Uganda GPC data and the UK Biobank

(A) Phenotypic correlations measured in traits in the Uganda GPC among unrelated individuals.

(B) Phenotypic correlations in the unrelated UK Biobank European ancestry individuals. (A and B) Phenotypes were mean centered and adjusted for age and sex within each cohort prior to correlation analysis. The order of each phenotype correlation is determined by hierarchical clustering in the Uganda GPC.

additional potential environmental explanations for mean shifts in phenotypes, such as dietary, food security, and age differences contributing to considerable BMI differences across cohorts ($\mu = 21.3$ and $\sigma = 3.8$ in Uganda GPC versus $\mu = 27.4$ and $\sigma = 4.8$ in the UK Biobank, $p < 2.2e-16$). To quantify comparisons while controlling for demographic differences for each of the 34 quantitative traits measured in both cohorts, we first mean centered each phenotype and regressed out the effects of age and sex within each cohort. Next, we then compared the distributions and variances of each phenotype across cohorts via Kolmogorov-Smirnov and F-tests, respectively (Table S4). Given the large sample sizes, all K-S tests were significantly different, with several phenotypes showing distributional and variance differences of considerable magnitude (Figure S7 and Table S4, e.g., Bilirubin, BASO, HbA1c, ALP, EOS, TG, and NEU).

We next analyzed how similar the relationships are between phenotypes across datasets. Similar trends emerge overall, with distances across variance-covariance matrices for these cohorts showing evidence of significant correlation (Mantel test Z statistic = 0.73, $p < 1e-4$). The correlations among phenotypes are slightly higher overall in the Uganda GPC than in UK Biobank, both among related and unrelated individuals (Figures 3B and S8). These findings are expected because of shared genetics and/or shared household environments contributing to more similar phenotypes.⁵⁴ More specifically, we see consistent correlations among combinations of phenotypes including SBP and DBP; RBC, Hb, and HCT; Cholesterol and LDL; WC, BMI, WT, and HC; MCHC, MCH, and MCV; GGT, ALT, AST, and ALP; and MONO, NEU, and WBC with high over-

all correlations across these datasets for these traits (Figures 3A and 3B, see abbreviations in Table S1). Some pairs of traits, however, have significantly different correlations across datasets. The largest difference in phenotypic correlations across datasets is between ALP and WT ($\rho = 0.11$, $p < 2.2e-16$ in UK Biobank versus $\rho = -0.36$, $p < 2.2e-16$ in Uganda GPC).

Our next goal was to compare trait heritability estimates in the UK Biobank versus Uganda GPC data (materials and methods); however, the sample size and study design differences between these cohorts limited comparability using standard scalable approaches. Specifically, the household design of Uganda GPC included smaller sample sizes with more relatives in which family-based heritability estimates are most appropriate, whereas the large sample size and volunteer design in the UK Biobank makes SNP-based heritability estimates from unrelated individuals more appropriate. Figure S9 compares heritability estimates across traits in the UK Biobank versus Uganda GPC using these disparate approaches.⁴⁶ As expected from the differences in the methods, study designs, and sample sizes, we find higher but noisier estimates in Uganda GPC for most traits, consistent with expectation from family-based versus unrelated heritability estimates across these two studies. While all of these factors fundamentally limit comparability of heritability estimates across these cohorts, we have also estimated heritability in both cohorts in unrelated individuals with consistent methodology using multi-component Haseman-Elston regression implemented in RHE-mc to improve comparability.⁴⁸ These results showed higher heritability estimates in the Uganda GPC dataset that were not significantly correlated with

heritability estimates from any ancestry group in the Pan-UK Biobank Project, consistent with a wide range of differences influencing these phenotypes across cohorts (Table S7, Figure S10). With these heritability estimates, we also estimated observed versus predicted PRS accuracy and find that predicted R^2 tends to be higher than observed R^2 (Figure S11).

African genetic risk predictions from European ancestry GWAS data are remarkably inaccurate

To understand baseline trans-ancestry PRS accuracy using a typical approach, we predicted 32 traits in the Uganda GPC using GWAS summary statistics from the UK Biobank European ancestry individuals. While several traits were significantly predicted across ancestries, prediction accuracy was low for most traits (Figure S12); the most accurate PRS was for MPV ($R^2 = 0.036$, 95% CI = 0.0069–0.063, $p = 5.73e-7$), while the average variance explained across all traits was less than 1% (mean $R^2 = 0.007$). To assess the relative effects of ancestry versus cohort differences on decreases in prediction accuracy across populations, we next withheld 10,000 European ancestry individuals from UK Biobank for use as a target cohort, reran all GWASs, then used individuals with diverse continental ancestries in the UK Biobank as target populations (EUR = Europeans withheld from the GWAS, AMR = admixed American, MID = Middle Eastern, CSA = Central/South Asian, EAS = East Asian, and AFR = African, Figure S13), subcontinental African ancestries in the UK Biobank (Ethiopian, Admixed, South, East, West African ancestries, Figure S14), as well as the Uganda GPC (Figure 4A, Table S5).

Among continental ancestries, we computed R^2 and 95% CIs for each trait (Figure S15), then computed median RA compared with Europeans and MAD across all traits (materials and methods). We predict these traits most accurately in EUR (RA = 1, MAD = 0), followed by AMR (RA = 0.784, MAD = 0.023), MID (RA = 0.643, MAD = 0.034), CSA (RA = 0.621, MAD = 0.031), EAS (RA = 0.477, MAD = 0.024), and AFR (RA = 0.219, MAD = 0.014) (Figure 4A). Because different PRS methodologies can improve overall prediction accuracy for some traits, we also compared our results using pruning and thresholding with PRS-CS; as described previously, different PRS methods may perform better for some phenotypes than others, but do not generally improve the relative loss of accuracy^{19,55} (Figure S16). We next compared prediction accuracy within African ancestry populations. Because some PRS accuracy estimates were noisy due to small sample sizes in UK Biobank Africans (especially Ethiopian and South African ancestry individuals, Table S5), we restricted analyses to those traits predicted with a 95% CI < 0.08. Among these traits, we predicted most accurately those with Ethiopian ancestry (RA = 0.511, MAD = 0.059), followed by recently admixed individuals with West African and European ancestry (RA = 0.276, MAD = 0.016), East African ancestry (RA = 0.193, MAD = 0.023), West African ancestry (RA = 0.150, MAD = 0.012), and South African ancestry (RA = 0.083,

MAD = 0.014) (Figure 4A). These results track with genetic distance as measured by F_{ST} (Table S8) and population history; the highest prediction accuracy identified in Ethiopians is expected given closer genetic proximity to European populations relative to other Africans due to back-to-Africa migrations influencing population structure there.^{17,56,57} The lowest prediction accuracy is in populations with southern African ancestry, consistent also with higher genetic divergence from European populations and more genetic diversity overall.^{16,18,58}

Next, we quantified the proportion of loss of prediction accuracy (LOA, calculated as $(1 - RA) * 100\%$) due to MAF and LD in the subcontinental African ancestry groups in the UK Biobank. As expected, LOA followed an inverse trend to prediction accuracy, i.e., LOA increased with genetic distance between the discovery and target cohort (Figure S17). LOA was lowest in the Ethiopian group (median LOA = 26.22) and highest in the West group (median LOA = 40.11).

Lower prediction accuracy across ancestries than across cohorts

To compare prediction accuracy among similar ancestry participants from different cohorts, we next computed PRSs for 34 traits using GWAS summary statistics from UK Biobank Europeans in two target populations: UK Biobank participants with East African ancestry versus Uganda GPC. As expected, prediction accuracy in these populations is very low across all traits in both cohorts and only slightly higher in the UK East African ancestry individuals than in the Uganda GPC individuals (mean $R^2 = 0.017$, SD = 0.013 versus mean $R^2 = 0.012$, SD = 0.010, respectively, Figure S18). Across traits, the differences in PRS accuracy across cohorts but within the same ancestry (Figure S18A) are much smaller than the differences across ancestries but within the UK Biobank (Figure 4A, left and middle panels), indicating that ancestry has a larger impact on genetic risk prediction than cross-cohort differences analyzed here. Smaller effects on genetic prediction accuracy differences across cohorts may be attributable to environmental differences, such as higher rates of malnutrition and infectious diseases previously reported in Uganda and in the GPC.^{39,59}

Improved African genetic risk prediction accuracy with multi-ethnic GWAS summary statistics

We next maintained the target populations but varied the discovery cohort to determine how more diverse GWAS impacts PRS accuracy for these phenotypes in diverse populations. Specifically, we computed PRS accuracy in diverse target populations in the UK Biobank (Table S5) using one of two discovery cohorts: the UK Biobank European-only cohort versus diverse discovery cohorts combined via meta-analysis (Table S6). Meta-analyzed GWAS summary statistics come from several cohorts, including the UK Biobank, BBJ,⁶⁰ PAGE Consortium,⁶¹ and Uganda Genome Resource (UGR).⁴⁶ For each trait, discovery cohort, and target cohort combination, we normalized the PRS R^2

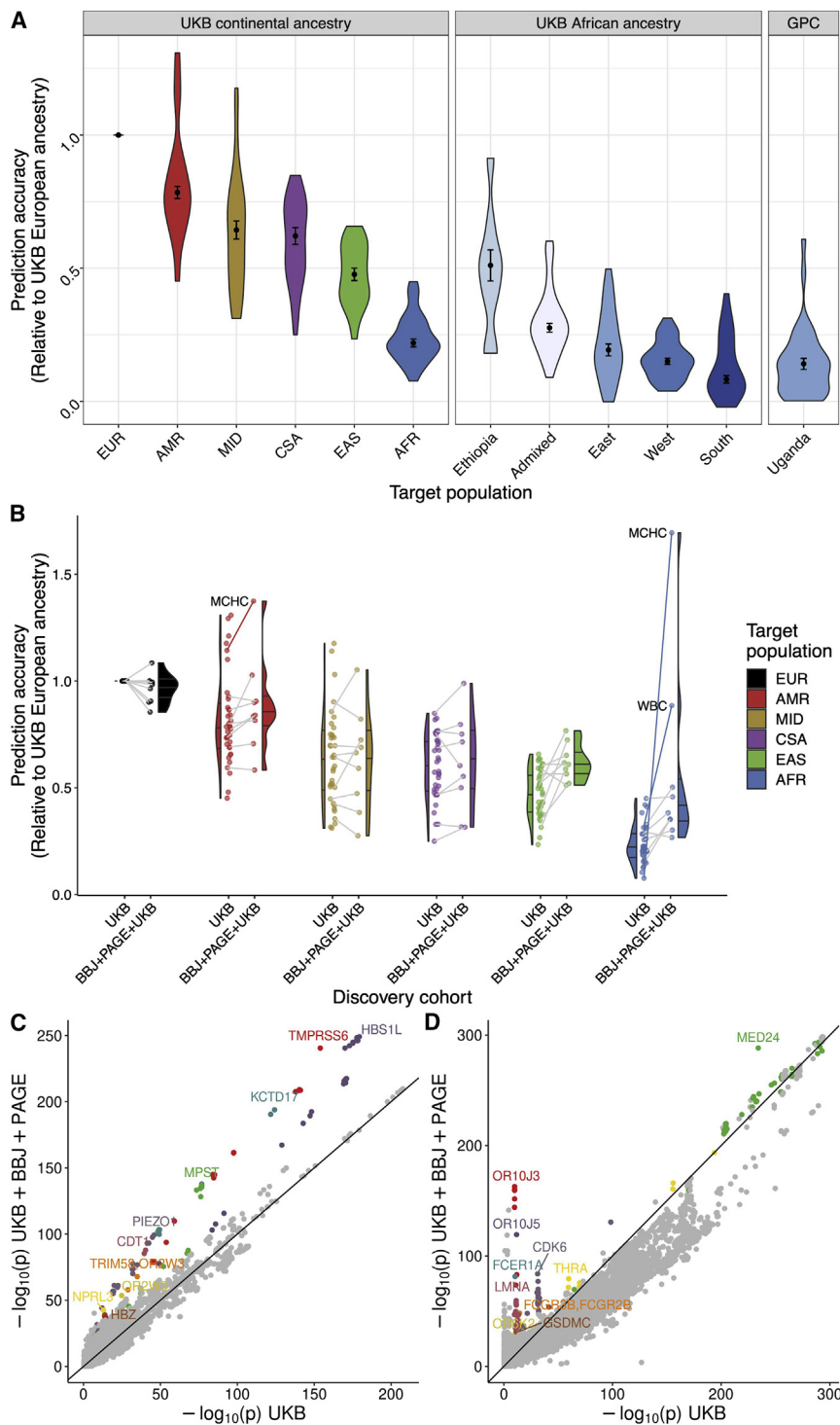


Figure 4. PRS accuracy and corresponding genetic variant contributions for up to 34 traits within and across diverse ancestries

(A) PRS accuracy relative to European ancestry individuals in diverse target ancestries. Discovery data consisted of GWAS summary statistics from UK Biobank (UKB) European ancestry data. Target data consisted of globally diverse continental ancestries (including withheld European target individuals) and regional African ancestry participants from UKB, or unrelated individuals from the Uganda GPC cohort. Traits were filtered to those with a 95% confidence interval range in PRS accuracy < 0.08 .

(B) PRS accuracy from a homogeneous versus multi-ancestry discovery dataset. GWAS discovery data consisted of summary statistics from UKB European ancestry data only or from the meta-analysis of UKB, BioBank Japan (BBJ), and Population Architecture using Genomics and Epidemiology (PAGE). Target populations are from the UKB. Lines connect the 10 traits available in both discovery cohorts to indicate how accuracy changed for the same trait in the UKB only versus meta-analyzed discovery data, while half violin plots show the distribution across all phenotypes in each discovery cohort. When lines are missing, the trait is absent in PAGE. Trait outliers are labeled in text and with solid lines. (A and B) Relative PRS accuracies are compared to the maximum for each trait in target samples withheld from discovery consisting of UKB European ancestry individuals. To simplify comparisons, only the polygenic scores with the highest prediction accuracy are shown here. Colors in these two panels correspond to the same continental ancestries.

(C and D) Trait-specific genetic outlier plots. QQ-like plot showing p values in UKB only versus multi-cohort meta-analysis of UKB, BBJ, and PAGE. The 10 regions that are genome-wide significant in both dataset and show the most significant differences are colored and labeled for (C) MCHC, and (D) WBC.

values from the p-value threshold that explained the maximum phenotypic variance with respect to the prediction accuracy in the European target cohort using UK Biobank summary statistics only, then computed RAs as before.

We find that prediction accuracy improves the most across populations when using a discovery cohort consisting of GWAS summary statistics meta-analyzed across the UK Biobank, BBJ, and PAGE cohorts (Figure 4B). To deter-

mine whether the improvement in prediction accuracy was due to the increase in sample size or the diversification of the GWAS discovery, we compared prediction accuracy across three discovery cohorts: 100,000 EUR individuals from GWAS summary statistics acquired from Martin et al., 2019a¹⁰, 350,000 EUR individuals, and multi-ancestry GWAS comprising UK Biobank, BBJ, and PAGE (Figure S19). We observe that the increase in discovery sample size from 100,000 to 350,000 EUR improves

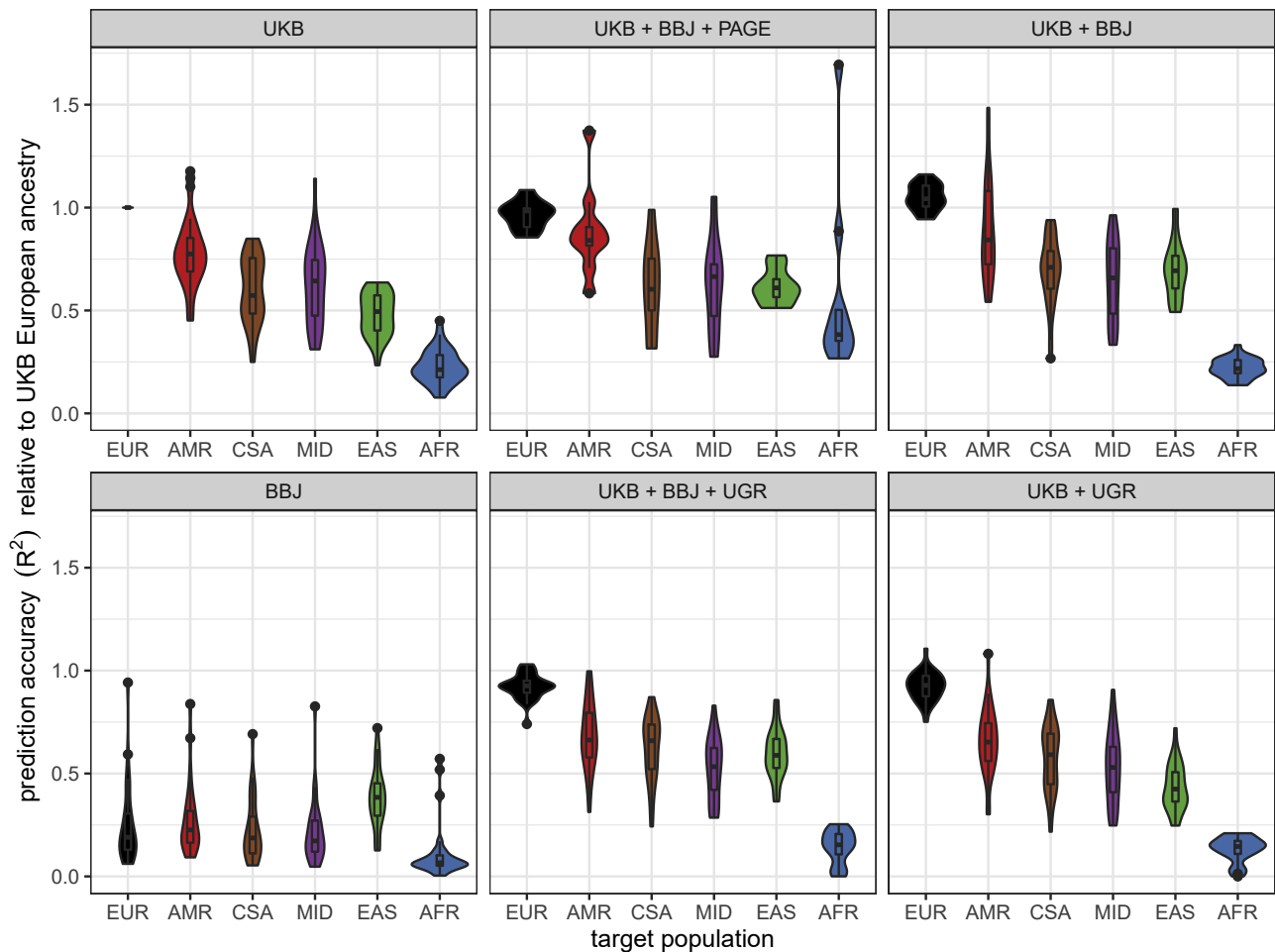


Figure 5. Relative PRS accuracy using the same target individuals and varying discovery cohorts

All relative comparisons are with respect to accuracy in withheld EUR when predicting with UKB European GWAS summary statistics alone as the discovery cohort.

prediction accuracy differentially across populations (Figure S19A). When comparing prediction accuracy across the three discovery cohorts, the results show that increasing the sample size improves prediction accuracy across all ancestries, but more so for the EUR population. The multi-ancestry discovery cohort seemed to improve prediction accuracy in the non-EUR populations more than the increase in sample size in general, with the largest improvement in prediction accuracy observed for BMI in AMR and EAS populations and MCHC and WBC for the AFR population (Figure S19B).

Surprisingly, meta-analyzing the UGR data with UK Biobank did not improve prediction accuracy for any population and most notably decreased accuracy in African ancestry target populations (discovery UK Biobank median $RA = 0.22$, UGR + UK Biobank median $RA = 0.15$, Figures 5 and S20). We hypothesize that (1) the relatively small sample size of UGR adds more noise than signal as indicated by the large error bars, and (2) the difference in effect sizes between UGR and UK Biobank, particularly for the less polygenic traits such as LDL (Figure S21) contributes to the noise. When predicting traits using the UK Biobank,

BBJ, and PAGE meta-analysis as a discovery cohort, we find that prediction accuracy increases most for the AMR, EAS, and AFR target populations, which more closely resemble the ancestry patterns of PAGE and BBJ (Figure 4B). The meta-analysis conflates two factors that are known to improve prediction accuracy: increase in sample size and diversity in the discovery cohort. To determine which of these factors drove the gains in prediction accuracy in Figure 4B, we compared the prediction accuracy from 100K EUR individuals from UK Biobank, downsampled to match the size of BBJ, to that of the 350k EUR (for 17 overlapping phenotypes) and the multi-ancestry discovery (for five overlapping phenotypes). This comparison indicates that indeed increasing the discovery sample size generally improves prediction accuracy; however, it is the inclusion of diverse samples in the discovery cohort that improves prediction accuracy, especially for the populations represented in that cohort (Figure S19). These findings are consistent with ancestry-matched discovery data disproportionately improving prediction accuracy in the corresponding target population.^{4,8,31}

Large-effect population-enriched genetic variants drive heterogeneity in polygenic score accuracy for blood panel traits

We find that PRS accuracy improvements from higher diversity in the discovery cohorts vary across traits, with the largest increases seen in MCHC and WBC particularly in AMR and AFR populations (Figure 4B). We searched for specific genetic loci that could explain this pattern by comparing the significance of genetic associations in UK Biobank alone versus the meta-analysis of UK Biobank, BBJ, and PAGE (Table S6). For MCHC and WBC in particular, the genetic variants contributing to these improved PRSs consist of several well-known population-enriched variants (Figures 4C and 4D). For example, genetic variants that disproportionately explain population-specific risk for MCHC include variants previously associated with hemoglobin concentration, including rs9399137 upstream of *HBS1L* and *MYB* in a study of sickle cell anemia ($p = 5.24e-249$ and $\beta = 0.0783$ in the meta-analysis),⁶² rs855791 in *TMPRSS6* ($p = 3.49e-241$, $\beta = 0.0692$),^{63,64} and rs551118 upstream of *PIEZO1* and *CDT1* ($p = 5.18e-100$, $\beta = -0.0451$)⁶⁵ (Table S9). Associations with WBC tend to show more population-enriched associations as shown in the meta-analysis (Figure 4D), including rs3936197 in *MED24* ($p = 5.18e-289$, $\beta = -0.0772$), rs58650325 near the high affinity immunoglobulin (IgE) receptor *FCER1A* that initiates the allergic response ($1.57e-163$, $\beta = -0.097$, also close to *OR10J3*), and rs11533993 in *CDK6* ($p = 1.55e-84$, $\beta = -0.0799$). Thus, genetic architecture and population genetic considerations are important to bear in mind when considering the generalizability of polygenic scores.

Discussion

PRSs have been proposed as genetic biomarkers for use in preventive medicine,^{66,67} but are currently limited by low accuracy across populations especially in African ancestry populations.^{4,6} Through simulations and empirical work, this study has enabled unique insights into PRS transferability within and among diverse continental African populations as well as among African ancestry populations living in considerably different environments. Simulations will continue to play a crucial role in understanding and mitigating biases, but the small sample size of existing genetic studies in African populations have limited the simulation designs that are even possible with realistic population structure across the African continent in this study. The AGVP dataset used for the first set of simulations was too small to use a typical infinitesimal simulation strategy. As a result, we simulated phenotypes with variants with large effects—a scenario that is inconsistent with the genetic architecture of most polygenic traits. While the simulation done with the AWI-Gen dataset represents a scenario that is more realistic for complex traits, the findings re-emphasize that the paucity of large genetic samples in non-European ancestry populations limits

simulation designs. Future studies could simulate new individuals from observed allele frequencies or from larger scale genetic datasets as they are made available. Despite these limitations, the simulation work done here provides a framework for simulation designs within the current sample size confines and what can be expected from these simulations in African populations.

We demonstrate looming challenges for applying current PRS in African ancestry populations—because relatively few genetic studies have been conducted in African populations coupled with the lack of out-of-Africa population bottlenecks, PRS accuracy is low but widely variable. Differences in PRS accuracy across diverse African ancestries from different regions can be larger than across out-of-Africa continents. This is particularly problematic, as widely used algorithms that guide health decisions already have ingrained racial biases,⁶⁸ warning of compounding challenges with implementation. We demonstrate that there are clear steps the field can take to work against these biases. Specifically, including ancestrally diverse populations in GWASs at considerably larger sample sizes, discovery cohorts improve accuracy for all populations and especially underrepresented populations more than conducting similarly sized studies with only European ancestry cohorts.

Another advantage of using GWASs from globally diverse populations to compute PRS is the routine inclusion of population-enriched variants. Clear examples such as African-enriched variants in *APOL1* and *G6PD* have been shown to contribute to especially high risk of chronic kidney disease and to missed diabetes diagnosis, respectively.^{69,70} These examples highlight the importance of studying diverse populations to predict genetic risk of disease equitably by aggregating variants across the spectrum of allele frequencies and effect sizes in different populations. Relevant to the traits studied in genetic analyses here, hematological differences such as anemia are more common in lower income countries in Africa and in African ancestry populations elsewhere compared with European ancestry populations in high-income countries, particularly among older individuals. These hematological differences potentially arise in part due to genetic variation as well as the higher prevalence of infectious diseases and pathogens, poorer nutritional status, and altitude.^{71,72} Here, we show that variants influencing risk of beta thalassemia disproportionately increase PRS accuracy for hemoglobin variation particularly in African ancestry populations. The inclusion of population-enriched variants in PRS could eliminate genetic justifications for race-based medicine, which problematically reinforces implicit racial biases by overemphasizing the link between genetics and race despite the fact that there is more genetic variation within than between ancestral populations.⁷³ However, for this to be possible, genetic data would have to be available for all populations at scale—an ideal that is still a ways off.

In addition to reduced PRS accuracy with ancestral distance from GWAS cohorts, genetic nurture, social genetic,

and environmental effects can also contribute to low portability of PRS across populations,^{23,74} with some interventions modulating health along PRS strata.⁷⁵ In this study, however, ancestry appears to have a larger effect on portability than cohort differences overall. An important distinction when comparing the magnitude of these and other non-genetic effects in other studies is that the traits most accurately genetically predicted here were primarily anthropometric and blood panel traits. When analyzing traits with more sociodemographic influences in increasingly diverse populations, population stratification, confounding, and study design considerations are thornier issues.^{22,76,77} PRS accuracy comparisons across ancestrally similar but environmentally diverse populations are especially important for medically actionable traits. For example, particularly low PRS portability for triglycerides (TG) from European to the Uganda GPC resulted at least in part from effect size heterogeneity that has previously been connected to pleiotropic and gene * environment effects; specifically, most non-transferable genome-wide significant associations with TG showed pleiotropic associations with BMI in European but not Ugandan individuals.⁷⁸

While PRSs currently have limited portability, increased diversity in genetic studies is already decreasing prediction accuracy gaps across populations.^{31,78,79} This is consistent with causal genetic effects tending to be similar across populations but with LD and allele frequency differences modifying marginal effect size estimates.^{4,7,8} This is also consistent with trans-ethnic genetic correlations tending to be close to or not significantly different from 1.^{80,81} The most rapid path to closing gaps in PRS transferability is to increase the inclusion of GWAS participants from populations most divergent from those already routinely studied. As empirically demonstrated here, when comparing PRS accuracy calculated from diverse cohort meta-analysis versus data from Europeans only, large-scale GWASs with diverse African populations will rapidly reduce portability gaps across global populations because they have the most genetic diversity, most rapid linkage disequilibrium decay, and highest genetic divergence from the best studied populations. Major efforts under way, such as the Human Hereditary and Health in Africa Initiative, PAGE, All of Us, and NeuroGAP programs,^{61,82–85} are especially promising for rectifying current PRS gaps and missed scientific opportunities by increasing inclusion of diverse African participants.

Beyond expanding on diversity by increasing the number of study participants in large-scale studies, it is equally important to diversify researchers working on genomics studies. Currently, the vast majority of researchers in genomics studies are of European ancestry,^{86–88} paralleling the over-representation of European ancestry individuals in genomic studies. The exclusion of African researchers leads to the disparity in research leadership and reduced scientific output from African researchers.⁸⁹ Efforts such as the Global Initiative for Neuropsychiatric Genetics Educa-

tion and Research (GINGER) program,⁹⁰ which provides mentorship and training for early-career investigators on the African continent (particularly in Uganda, Kenya, Ethiopia, and South Africa, including several of this study's authors), are important in moving toward a more inclusive and representative research community.

Conclusion

Previous studies that have examined PRS accuracy across globally diverse ancestry groups have demonstrated that accuracy is lowest in African ancestry samples. However, the extent to which this accuracy varies within African ancestry populations has not been previously investigated. Our findings that prediction accuracy varies by African ancestry populations is a clear reflection of the vast genetic diversity of the continent. It is therefore critically important to create well-powered GWASs that reflect the full range of diversity within Africa.

Data and code availability

All data used in this study are publicly available. Data from the African Genome Variation Project was accessed by combining EGAD00010001045, EGAD00010001046, EGAD00010001049, EGAD00010001050, EGAD00010001051, EGAD00010001052, EGAD00010001053, EGAD00010001054, EGAD00010001055, EGAD00010001056, EGAD00010001057, and EGAD00010001058. The Drakenstein Child Health Study is committed to the principle of data sharing. De-identified data will be made available to requesting researchers as appropriate. Requests for collaborations to undertake data analysis are welcome. Uganda GPC genetic data used in this paper were accessed through EGAD00010000965 and phenotype data were accessed via sftp from EGA (reference: DD_PK_050716 gwas_phenotypes_28Oct14.txt). We accessed data from the UK Biobank with application 31,063. BioBank Japan summary statistics were accessed from <http://jenger.riken.jp/en/result>. GWAS summary statistics for the Population Architecture using Genomics and Epidemiology (PAGE) study were accessed through the NHGRI-EBI GWAS Catalog (<https://www.ebi.ac.uk/gwas/downloads/summary-statistics>).

All code used in analysis is available here: https://github.com/armartin/africa_prs.

Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.xhgg.2023.100184>.

Acknowledgments

We thank Lori Chibnik, Bizu Gelaye, Kristi Post, and Courtney White for facilitating the GINGER program and making this work possible. This work was supported by funding from the National Institutes of Health (K99/R00MH117229 to A.R.M.; K01MH121659 and

T32MH017119 to E.G.A.). UK Biobank analyses were conducted via application 31063. The DCHS is funded by the Bill and Melinda Gates Foundation (OPP1017641). Additional support for H.J.Z. and D.J.S., and for the research reported in this publication, was provided by the South African Medical Research Council (SAMRC). The SAMRC provided additional support through its Division of Research Capacity Development under the National Health Scholarship program from funding received from the Public Health Enhancement Fund/South African National Department of Health. The views and opinions expressed are those of the authors and do not necessarily represent the official views of the SAMRC. We thank the mothers and their children for participating in the DCHS and the study staff, the clinical and administrative staff of the Western Cape Government Health Department at Paarl Hospital, and at the clinics for support of the study. We also thank all research participants in the UK Biobank, BioBank Japan, PAGE study, UGR and Uganda GPC, and AGVP studies.

Declaration of interests

The authors declare no competing interests.

Received: October 10, 2022

Accepted: February 4, 2023

Web resources

European Genomic-Phenome Archive, <https://ega-archive.org>.

Drakenstein Child Health Study, <http://www.paediatrics.uct.ac.za/scah/dchs>.

BioBank Japan summary statistics, <http://jenger.riken.jp/en/result>.

Population Architecture using Genomics and Epidemiology (PAGE), <https://www.ebi.ac.uk/gwas/downloads/summary-statistics>.

References

- Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* *101*, 5–22.
- Morales, J., Welter, D., Bowler, E.H., Cerezo, M., Harris, L.W., McMahon, A.C., Hall, P., Junkins, H.A., Milano, A., Hastings, E., et al. (2018). A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol.* *19*, 21.
- Popejoy, A.B., and Fullerton, S.M. (2016). Genomics is failing on diversity. *Nature* *538*, 161–164.
- Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* *51*, 584–591.
- Manrai, A.K., Funke, B.H., Rehm, H.L., Olesen, M.S., Maron, B.A., Szolovits, P., Margulies, D.M., Loscalzo, J., and Kohane, I.S. (2016). Genetic misdiagnoses and the potential for health disparities. *N. Engl. J. Med.* *375*, 655–665.
- Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The missing diversity in human genetic studies. *Cell* *177*, 26–31.
- Chen, J., Spracklen, C.N., Marenne, G., Varshney, A., Corbin, L.J., Luan, J., Willems, S.M., Wu, Y., Zhang, X., Horikoshi, M., et al. (2021). The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet.* *53*, 840–860.
- Lam, M., Chen, C.-Y., Li, Z., Martin, A.R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B.C., et al. (2019). Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat. Genet.* *51*, 1670–1678.
- Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* *47*, 979–986.
- Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., Daly, M.J., Bustamante, C.D., and Kenny, E.E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* *100*, 635–649.
- Scutari, M., Mackay, I., and Balding, D. (2016). Using genetic distance to infer the accuracy of genomic prediction. *PLoS Genet.* *12*, e1006288.
- Martin, A.R., Daly, M.J., Robinson, E.B., Hyman, S.E., and Neale, B.M. (2019). Predicting polygenic risk of psychiatric disorders. *Biol. Psychiatr.* *86*, 97–109.
- International Schizophrenia Consortium, Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* *460*, 748–752.
- 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. *Nature* *526*, 68–74.
- Uren, C., Kim, M., Martin, A.R., Bobo, D., Gignoux, C.R., van Helden, P.D., Möller, M., Hoal, E.G., and Henn, B.M. (2016). Fine-scale human population structure in southern Africa reflects ecogeographic boundaries. *Genetics* *204*, 303–314.
- Busby, G.B., Band, G., Si Le, Q., Jallow, M., Bougama, E., Mangano, V.D., Amenga-Etego, L.N., Enimil, A., Apinjoh, T., Ndila, C.M., et al. (2016). Admixture into and within sub-saharan Africa. *Elife* *5*, e15266. <https://doi.org/10.7554/eLife.15266>.
- Pagani, L., Schiffels, S., Gurdasani, D., Danecsek, P., Scally, A., Chen, Y., Xue, Y., Haber, M., Ekong, R., Oljira, T., et al. (2015). Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. *Am. J. Hum. Genet.* *96*, 986–991.
- Choudhury, A., Aron, S., Botigué, L.R., Sengupta, D., Botha, G., Bensellak, T., Wells, G., Kumuthini, J., Shriner, D., Fakim, Y.J., et al. (2020). High-depth African genomes inform human migration and health. *Nature* *586*, 741–748.
- Wang, Y., Guo, J., Ni, G., Yang, J., Visscher, P.M., and Yengo, L. (2020). Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nat. Commun.* *11*, 3865.
- de Vlaming, R., Okbay, A., Rietveld, C.A., Johannesson, M., Magnusson, P.K.E., Uitterlinden, A.G., van Rooij, F.J.A., Hofman, A., Groenen, P.J.F., Thurik, A.R., and Koellinger, P.D. (2017). Meta-GWAS accuracy and power (MetaGAP) calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies. *PLoS Genet.* *13*, e1006495.
- Wray, N.R., Yang, J., Hayes, B.J., Price, A.L., Goddard, M.E., and Visscher, P.M. (2013). Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* *14*, 507–515.

22. Zaidi, A.A., and Mathieson, I. (2020). Demographic History Impacts Stratification in Polygenic Scores. <https://doi.org/10.1101/2020.07.20.212530>.
23. Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J.K., and Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *Elife* 9, e48376. <https://doi.org/10.7554/eLife.48376>.
24. Hero, J.O., Zaslavsky, A.M., and Blendon, R.J. (2017). The United States leads other Nations in differences by income in perceptions of health and health care. *Health Aff.* 36, 1032–1040.
25. Roser, M., Ortiz-Ospina, E., and Ritchie, H. (2013). Life expectancy. *Our World in Data*.
26. U.S. Department of Health & Human Services, and Agency for Healthcare Research and Quality (2017). 2016 National Healthcare Quality and Disparities Report.
27. Martin, A.R., Teferra, S., Möller, M., Hoal, E.G., and Daly, M.J. (2018). The critical needs and challenges for genetic architecture studies in Africa. *Curr. Opin. Genet. Dev.* 53, 113–120.
28. Henn, B.M., Cavalli-Sforza, L.L., and Feldman, M.W. (2012). The great human expansion. *Proc. Natl. Acad. Sci. USA* 109, 17758–17764.
29. Campbell, M.C., and Tishkoff, S.A. (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genom. Hum. Genet.* 9, 403–433.
30. Genovese, G., Friedman, D.J., Ross, M.D., Lecordier, L., Uzureau, P., Freedman, B.I., Bowden, D.W., Langefeld, C.D., Oleksyk, T.K., Uscinski Knob, A.L., et al. (2010). Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* 329, 841–845.
31. Bigdeli, T.B., Genovese, G., Georgakopoulos, P., Meyers, J.L., Peterson, R.E., Iyegbe, C.O., Medeiros, H., Valderrama, J., Achtyes, E.D., Kotov, R., et al. (2020). Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Mol. Psychiatr.* 25, 2455–2467. <https://doi.org/10.1038/s41380-019-0517-y>.
32. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., and Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319, 1100–1104.
33. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., et al. (2015). The African genome variation project shapes medical genetics in Africa. *Nature* 517, 327–332.
34. Tamburini, F.B., Maghini, D., Oduaran, O.H., Brewster, R., Hulley, M.R., Sahibdeen, V., Norris, S.A., Tollman, S., Kahn, K., Wagner, R.G., et al. (2022). Short- and long-read metagenomics of urban and rural South African gut microbiomes reveal a transitional composition and undescribed taxa. *Nat. Commun.* 13, 926.
35. Ramsay, M., Crowther, N., Tambo, E., Agongo, G., Baloyi, V., Dikotope, S., Gómez-Olivé, X., Jaff, N., Sorgho, H., Wagner, R., et al. (2016). H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Glob. Health Epidemiol. Genom.* 1, e20.
36. Zar, H.J., Pellowski, J.A., Cohen, S., Barnett, W., Vanker, A., Koen, N., and Stein, D.J. (2019). Maternal health and birth outcomes in a South African birth cohort study. *PLoS One* 14, e0222399.
37. Stein, D.J., Koen, N., Donald, K.A., Adnams, C.M., Koopowitz, S., Lund, C., Marais, A., Myers, B., Roos, A., Sorsdahl, K., et al. (2015). Investigating the psychosocial determinants of child health in Africa: the Drakenstein child health study. *J. Neurosci. Methods* 252, 27–35.
38. Zar, H.J., Barnett, W., Myer, L., Stein, D.J., and Nicol, M.P. (2015). Investigating the early-life determinants of illness in Africa: the Drakenstein child health study. *Thorax* 70, 592–594.
39. Asiki, G., Murphy, G., Nakiyingi-Miir, J., Seeley, J., Nsubuga, R.N., Karabarinde, A., Waswa, L., Biraro, S., Kasamba, I., Pomilla, C., et al. (2013). The general population cohort in rural south-western Uganda: a platform for communicable and non-communicable disease studies. *Int. J. Epidemiol.* 42, 129–141.
40. Heckerman, D., Gurdasani, D., Kadie, C., Pomilla, C., Carstensen, T., Martin, H., Ekoru, K., Nsubuga, R.N., Ssenyomo, G., Kamali, A., et al. (2016). Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc. Natl. Acad. Sci. USA* 113, 7377–7382.
41. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O’Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209.
42. Lam, M., Awasthi, S., Watson, H.J., Goldstein, J., Panagiotaropoulou, G., Trubetskoy, V., Karlsson, R., Frei, O., Fan, C.-C., De Witte, W., et al. (2020). RICOPILI: rapid imputation for CONsortias PipeLine. *Bioinformatics* 36, 930–933.
43. Duncan, L.E., Ratanatharathorn, A., Aiello, A.E., Almlí, L.M., Amstatter, A.B., Ashley-Koch, A.E., Baker, D.G., Beckham, J.C., Bierut, L.J., Bisson, J., et al. (2018). Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol. Psychiatr.* 23, 666–673.
44. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Giga-Science* 4, 7.
45. Schoech, A.P., Jordan, D.M., Loh, P.-R., Gazal, S., O’Connor, L.J., Balick, D.J., Palamara, P.F., Finucane, H.K., Sunyaev, S.R., and Price, A.L. (2019). Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nat. Commun.* 10, 790.
46. Gurdasani, D., Carstensen, T., Fatumo, S., Chen, G., Franklin, C.S., Prado-Martinez, J., Bouman, H., Abascal, F., Haber, M., Tachmazidou, I., et al. (2019). Uganda genome resource enables insights into population history and genomic discovery in Africa. *Cell* 179, 984–1002.e36.
47. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47, 291–295.
48. Pazokitoroudi, A., Wu, Y., Burch, K.S., Hou, K., Zhou, A., Pasa-niuc, B., and Sankaraman, S. (2020). Efficient variance components analysis across millions of genomes. *Nat. Commun.* 11, 4020.
49. Howrigan, D. (2017). Details and Considerations of the UK Biobank GWAS.

50. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A., and Smoller, J.W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* *10*, 1776.
51. Daetwyler, H.D., Villanueva, B., and Woolliams, J.A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* *3*, e3395.
52. Wray, N.R., Kempner, K.E., Hayes, B.J., Goddard, M.E., and Visscher, P.M. (2019). Complex trait prediction from genome data: contrasting EBV in livestock to PRS in humans: genomic prediction. *Genetics* *211*, 1131–1141.
53. Bitarello, B.D., and Mathieson, I. (2020). Polygenic scores for height in admixed populations. *G3* *10*, 4027–4036.
54. Kong, A., Thorleifsson, G., Frigge, M.L., Vilhjalmsdottir, B.J., Young, A.I., Thorgeirsson, T.E., Benonisdottir, S., Oddsson, A., Halldorsson, B.V., Masson, G., et al. (2018). The nature of nurture: effects of parental genotypes. *Science* *359*, 424–428.
55. Wang, Y., Namba, S., Lopera, E., Kerminen, S., Tsuo, K., Läll, K., Kanai, M., Zhou, W., Wu, K.H., Favé, M.J., and Bhatta, L. (2023). Global Biobank analyses provide lessons for developing polygenic risk scores across diverse cohorts. *Cell Genom.* *3*, 100241.
56. Hodgson, J.A., Mulligan, C.J., Al-Meerri, A., and Raaum, R.L. (2014). Early back-to-Africa migration into the horn of Africa. *PLoS Genet.* *10*, e1004393.
57. Henn, B.M., Botigué, L.R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J.K., Fadhlouzi-Zid, K., Zalloua, P.A., Moreno-Estrada, A., Bertranpetit, J., et al. (2012). Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet.* *8*, e1002397.
58. Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A., et al. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc. Natl. Acad. Sci. USA* *108*, 5154–5162.
59. Nalwanga, D., Musiime, V., Kizito, S., Kiggundu, J.B., Batte, A., Musoke, P., and Tumwine, J.K. (2020). Mortality among children under five years admitted for routine care of severe acute malnutrition: a prospective cohort study from Kampala, Uganda. *BMC Pediatr.* *20*, 182.
60. Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroda, T., et al. (2017). Overview of the BioBank Japan project: study design and profile. *J. Epidemiol.* *27*, S2–S8.
61. Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. *Nature* *570*, 514–518. <https://doi.org/10.1038/s41586-019-1310-4>.
62. Lettre, G., Sankaran, V.G., Bezerra, M.A.C., Araújo, A.S., Uda, M., Sanna, S., Cao, A., Schlessinger, D., Costa, F.F., Hirschhorn, J.N., and Orkin, S.H. (2008). DNA polymorphisms at the BCL11A, HBS1L-MYB, and β -globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc. Natl. Acad. Sci. USA* *105*, 11869–11874.
63. Chambers, J.C., Zhang, W., Li, Y., Sehmi, J., Wass, M.N., Zabaneh, D., Hoggart, C., Baye, H., McCarthy, M.I., Peltonen, L., et al. (2009). Genome-wide association study identifies variants in TM6RS6 associated with hemoglobin levels. *Nat. Genet.* *41*, 1170–1172.
64. Benyamin, B., Ferreira, M.A.R., Willemsen, G., Gordon, S., Middelberg, R.P.S., McEvoy, B.P., Hottenga, J.-J., Henders, A.K., Campbell, M.J., Wallace, L., et al. (2009). Common variants in TM6RS6 are associated with iron status and erythrocyte volume. *Nat. Genet.* *41*, 1173–1175.
65. Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A., et al. (2016). The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* *167*, 1415–1429.e19.
66. Knowles, J.W., and Ashley, E.A. (2018). Cardiovascular disease: the rise of the genetic risk score. *PLoS Med.* *15*, e1002546.
67. Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., and Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* *50*, 1219–1224.
68. Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science* *366*, 447–453.
69. Rotimi, C.N., Bentley, A.R., Doumatey, A.P., Chen, G., Shriner, D., and Adeyemo, A. (2017). The genomic landscape of African populations in health and disease. *Hum. Mol. Genet.* *26*, R225–R236.
70. Wheeler, E., Liu, C.T., Mf, H., Hievert, M.F., Strawbridge, R., and Podmore, C. (2018). Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: a transethnic genome-wide meta-analysis. *PLoS medicine*. <https://doi.org/10.1530/ey.15.13.14>.
71. Mugisha, J.O., Seeley, J., and Kuper, H. (2016). Population based haematology reference ranges for old people in rural South-West Uganda. *BMC Res. Notes* *9*, 433.
72. Mugisha, J.O., Baisley, K., Asiki, G., Seeley, J., and Kuper, H. (2013). Prevalence, types, risk factors and clinical correlates of anaemia in older people in a rural Ugandan population. *PLoS One* *8*, e78394.
73. Cerdeña, J.P., Plaisime, M.V., and Tsai, J. (2020). From race-based to race-conscious medicine: how anti-racist uprisings call us to act. *Lancet* *396*, 1125–1128.
74. He, Y., Lakhani, C.M., Manrai, A.K., and Patel, C.J. (2019). Poly-exposure and poly-genomic scores implicate prominent roles of non-genetic and demographic factors in four common diseases in the UK. *Cold Spring Harbor Lab.*, 833632. <https://doi.org/10.1101/833632>.
75. Barcellos, S.H., Carvalho, L.S., and Turley, P. (2018). Education can reduce health differences related to genetic risk of obesity. *Proc. Natl. Acad. Sci. USA* *115*, E9765–E9772.
76. Novembre, J., and Barton, N.H. (2018). Tread lightly interpreting polygenic tests of selection. *Genetics* *208*, 1351–1355.
77. Kerminen, S., Martin, A.R., Koskela, J., Ruotsalainen, S.E., Havulinna, A.S., Surakka, I., Palotie, A., Perola, M., Salomaa, V., Daly, M.J., et al. (2019). Geographic variation and bias in the polygenic scores of complex diseases and traits in Finland. *Am. J. Hum. Genet.* *104*, 1169–1181.
78. Kuchenbaecker, K., Telkar, N., Reiker, T., Walters, R.G., Lin, K., Eriksson, A., Gurdasani, D., Gilly, A., Southam, L., Tsafantakis, E., et al. (2019). The transferability of lipid loci across African, Asian and European cohorts. *Nat. Commun.* *10*, 4330.
79. Graham, S.E., Clarke, S.L., Wu, K.-H.H., Kanoni, S., Zajac, G.J.M., Ramdas, S., Surakka, I., Ntalla, I., Vedantam, S., Winkler, T.W., et al. (2021). The power of genetic diversity in genome-wide association studies of lipids. *Nature* *600*, 675–679.

80. Brown, B.C., Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye, C.J., Price, A.L., and Zaitlen, N. (2016). Transethnic genetic-correlation estimates from summary statistics. *Am. J. Hum. Genet.* *99*, 76–88.
81. Shi, H., Burch, K.S., Johnson, R., Freund, M.K., Kichaev, G., Mancuso, N., Manuel, A.M., Dong, N., and Pasaniuc, B. (2020). Localizing components of shared transethnic genetic architecture of complex traits from GWAS summary data. *Am. J. Hum. Genet.* *106*, 805–817.
82. Hindorff, L.A., Bonham, V.L., Brody, L.C., Ginoza, M.E.C., Hutter, C.M., Manolio, T.A., and Green, E.D. (2018). Prioritizing diversity in human genomics research. *Nat. Rev. Genet.* *19*, 175–185.
83. Mulder, N., Abimiku, A., Adebamowo, S.N., de Vries, J., Matimba, A., Olowoyo, P., Ramsay, M., Skelton, M., and Stein, D.J. (2018). H3Africa: current perspectives. *Pharmgenomics Pers. Med.* *11*, 59–66.
84. Stevenson, A., Akena, D., Stroud, R.E., Atwoli, L., Campbell, M.M., Chibnik, L.B., Kwobah, E., Kariuki, S.M., Martin, A.R., de Menil, V., et al. (2019). Neuropsychiatric genetics of African populations-psychosis (NeuroGAP-Psychosis): a case-control study protocol and GWAS in Ethiopia, Kenya, South Africa and Uganda. *BMJ Open* *9*, e025469.
85. All of Us Research Program Investigators, Denny, J.C., Rutter, J.L., Goldstein, D.B., Philippakis, A., Smoller, J.W., Jenkins, G., and Dishman, E. (2019). The “all of us” research program. *N. Engl. J. Med.* *381*, 668–676.
86. Ginther, D.K., Schaffer, W.T., Schnell, J., Masimore, B., Liu, F., Haak, L.L., and Kington, R. (2011). Race, ethnicity, and NIH research awards. *Science* *333*, 1015–1019.
87. Hoppe, T.A., Litovitz, A., Willis, K.A., Meseroll, R.A., Perkins, M.J., Hutchins, B.I., Davis, A.F., Lauer, M.S., Valantine, H.A., Anderson, J.M., and Santangelo, G.M. (2019). Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. *Sci. Adv.* *5*, eaaw7238.
88. Hamrick, K. (2019). Women, Minorities, and Persons with Disabilities in Science and Engineering: 2019 (National Science Foundation, National Center for Science and Engineering Statistics (NCSES)), pp. 19–304. Alexandria, VA, Special Report NSF.
89. Bentley, A.R., Callier, S.L., and Rotimi, C.N. (2020). Evaluating the promise of inclusion of African ancestry populations in genomics. *NPJ Genom. Med.* *5*, 5. <https://doi.org/10.1038/s41525-019-0111-x>.
90. van der Merwe, C., Mwesiga, E.K., McGregor, N.W., Ejigu, A., Tilahun, A.W., Kalungi, A., Akimana, B., Dubale, B.W., Omari, F., Mmochi, J., et al. (2018). Advancing neuropsychiatric genetics training and collaboration in Africa. *Lancet. Glob. Health* *6*, e246–e247.

HGGA, Volume 4

Supplemental information

**Low and differential polygenic score
generalizability among African
populations due largely to genetic diversity**

Lerato Majara, Allan Kalungi, Nastassja Koen, Kristin Tsuo, Ying Wang, Rahul Gupta, Lethukuthula L. Nkambule, Heather Zar, Dan J. Stein, Eugene Kinyanda, Elizabeth G. Atkinson, and Alicia R. Martin

Supplementary Information

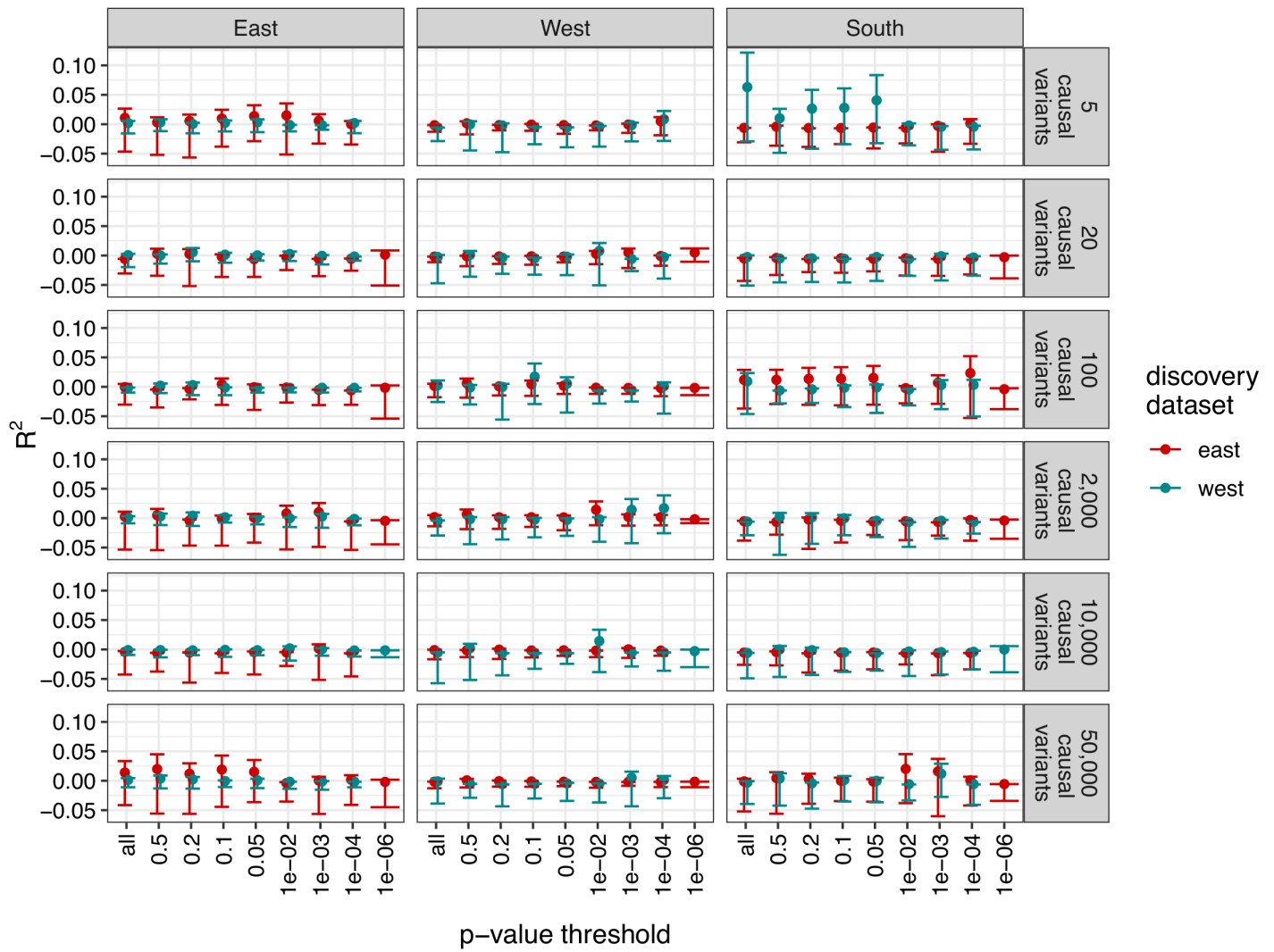


Figure S1 - Simulated polygenic score accuracies across regions of Africa using AGVP genotype data when $h^2=0.1$, similar to Figure 2.

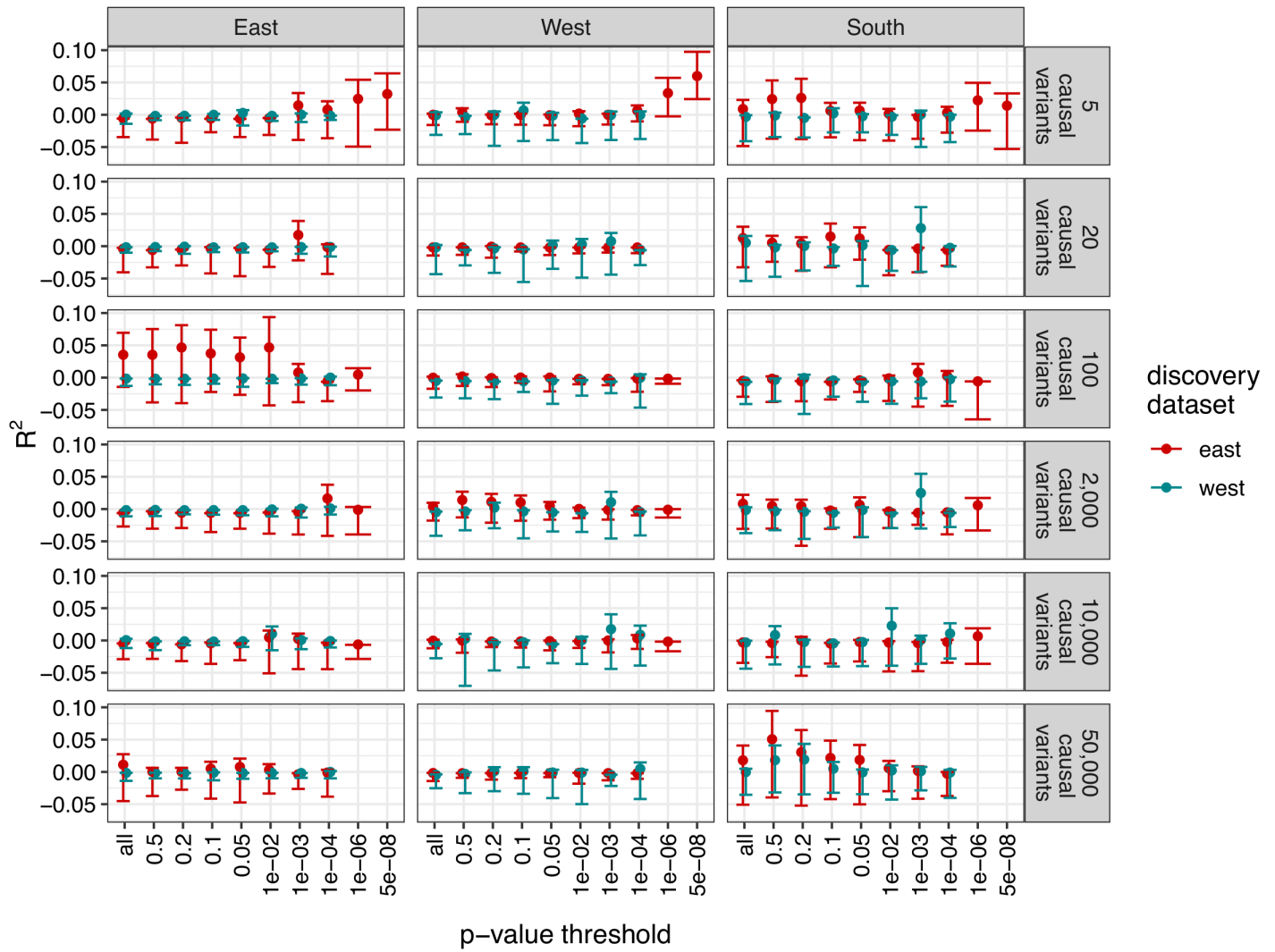


Figure S2 - Simulated polygenic score accuracies across regions of Africa using AGVP genotype data when $h^2=0.2$, similar to Figure 2.

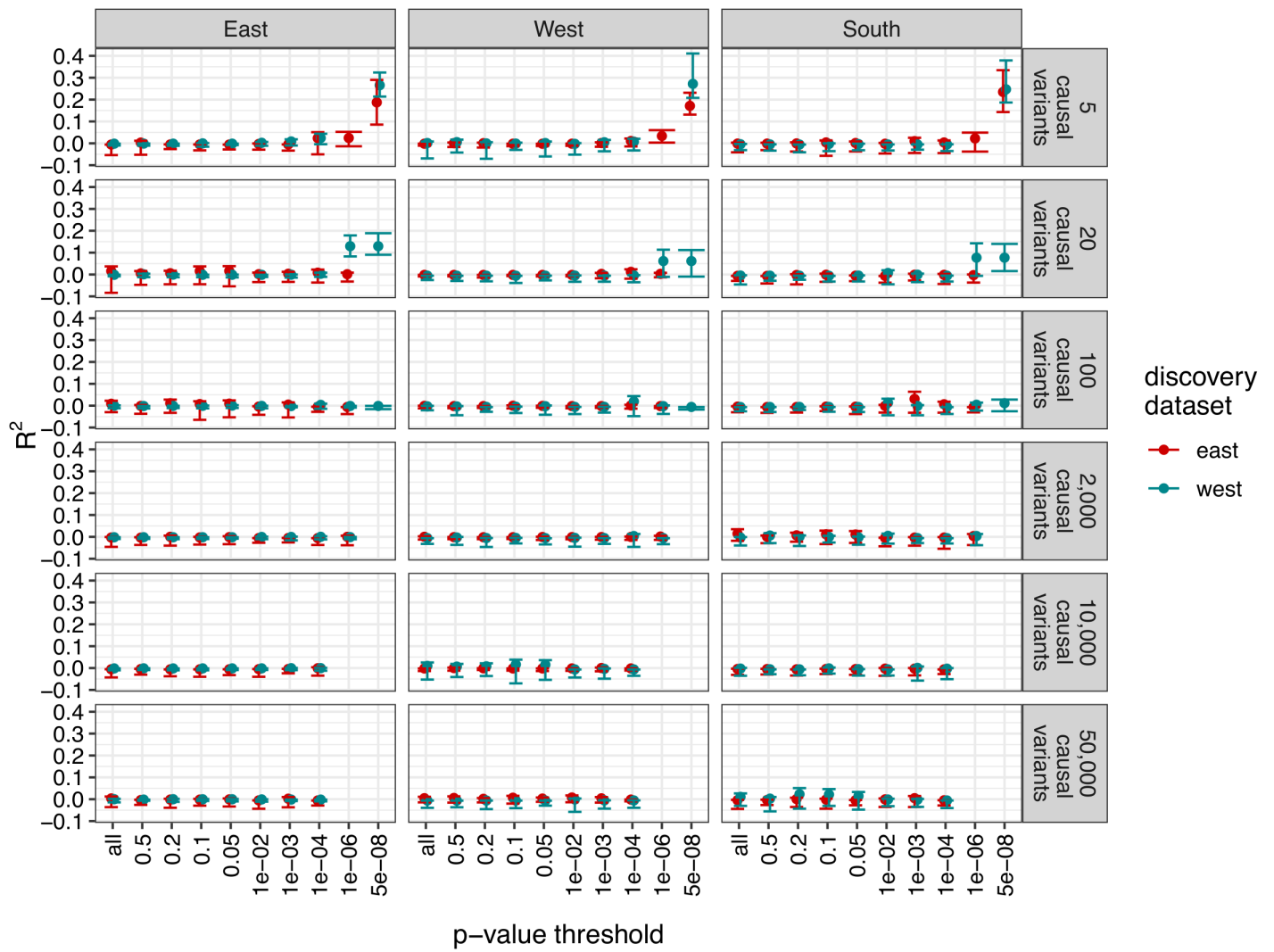


Figure S3 - Simulated polygenic score accuracies across regions of Africa using AGVP genotype data when $h^2=0.4$, similar to Figure 2.

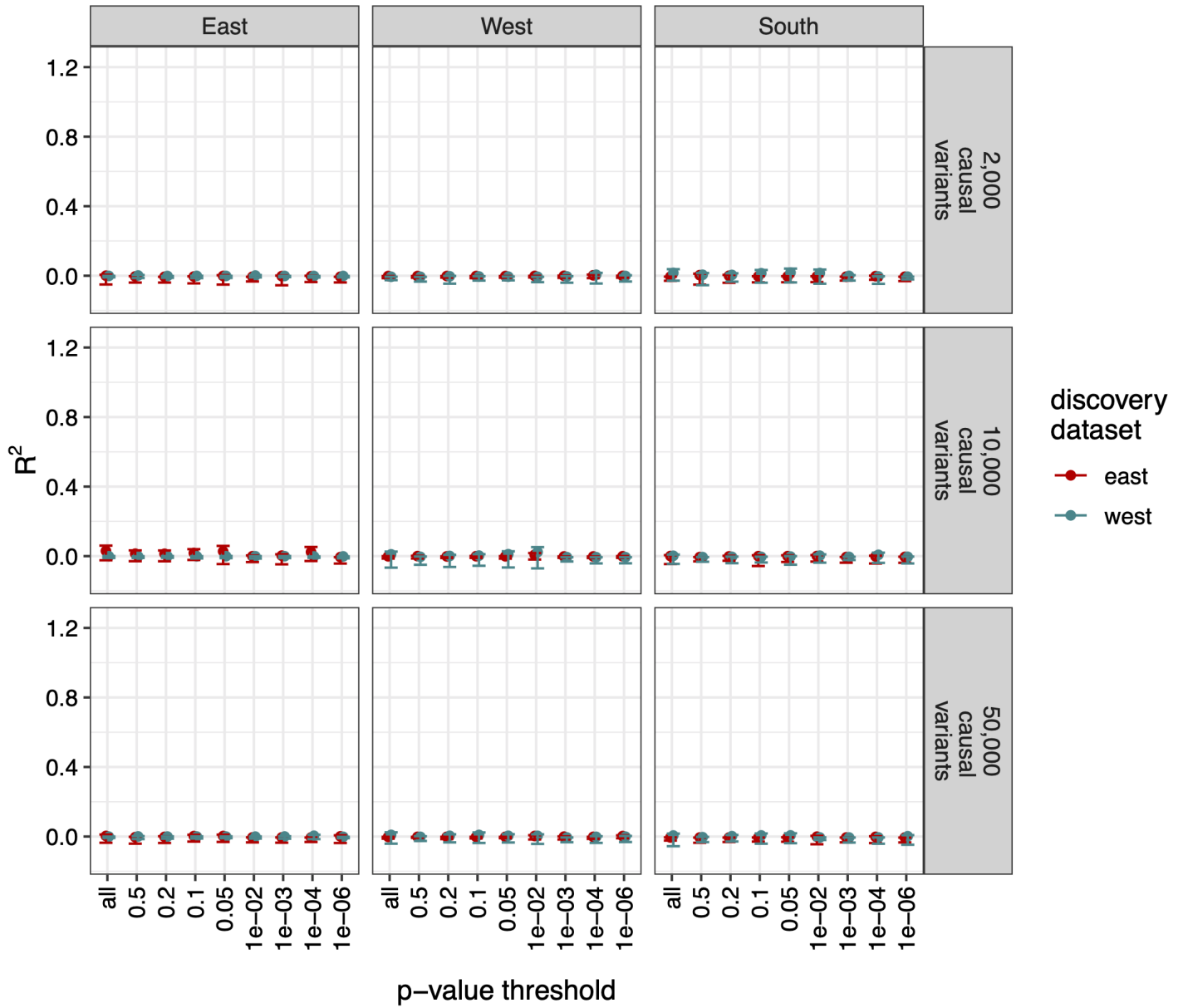


Figure S4 - Simulated polygenic score accuracies across regions of Africa using AGVP genotype data when $h^2=0.8$, similar to Figure 2. For this degree of polygenicity, no variants were associated with the trait at a p -value less than $5e-08$ hence this is not shown in the figure.

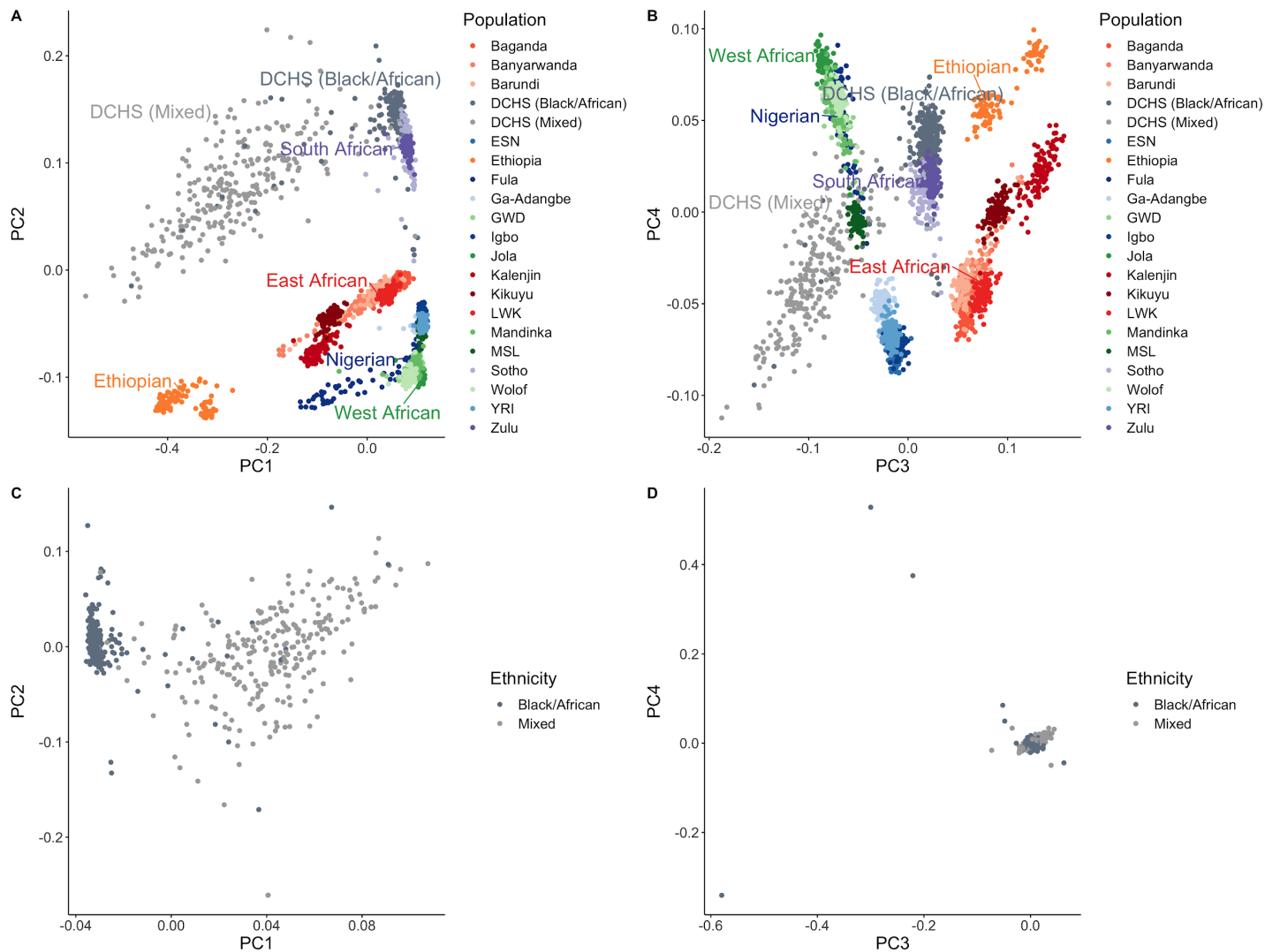


Figure S5 - Ancestry and ethnicity within DCHS and compared to African reference populations from AGVP and the 1000 Genomes Project. A-B) PCs in DCHS Black/African and Mixed ethnicities compared to reference data. C-D) PCs within DCHS without reference data. A, C) PC1 and PC2. B, D) PC3 and PC4.

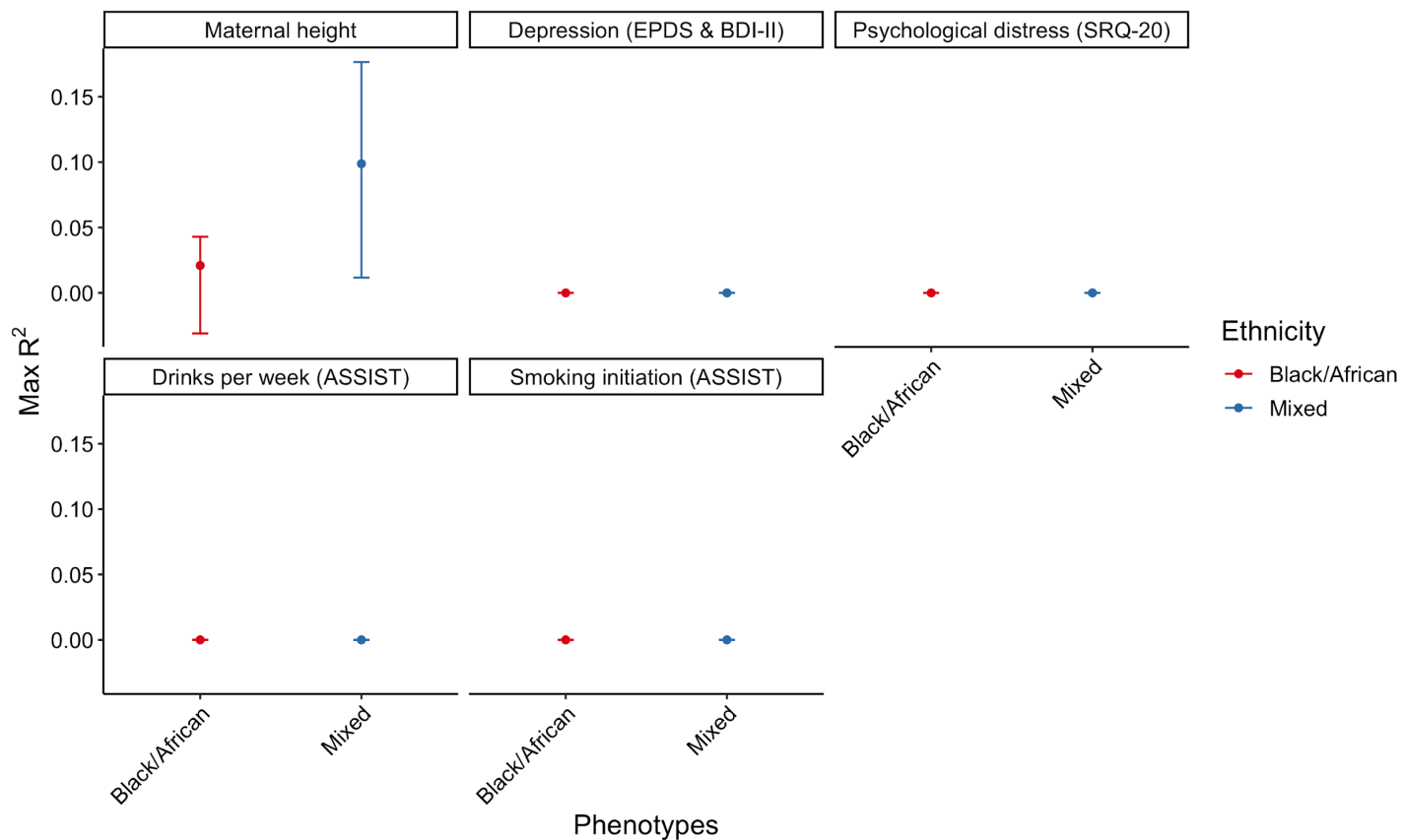


Figure S6 - PRS accuracy by ethnicity in the DCHS cohort for 6 different measured phenotypes. Max R² shows specifically the PRS threshold that explains the most phenotypic variation. Only height is significantly predicted. Abbreviations as follows: EPDS = Edinburgh Postnatal Depression Scale, BDI-II = Beck Depression Inventory II, ASSIST = Alcohol, Smoking and Substance Involvement Screening Test.

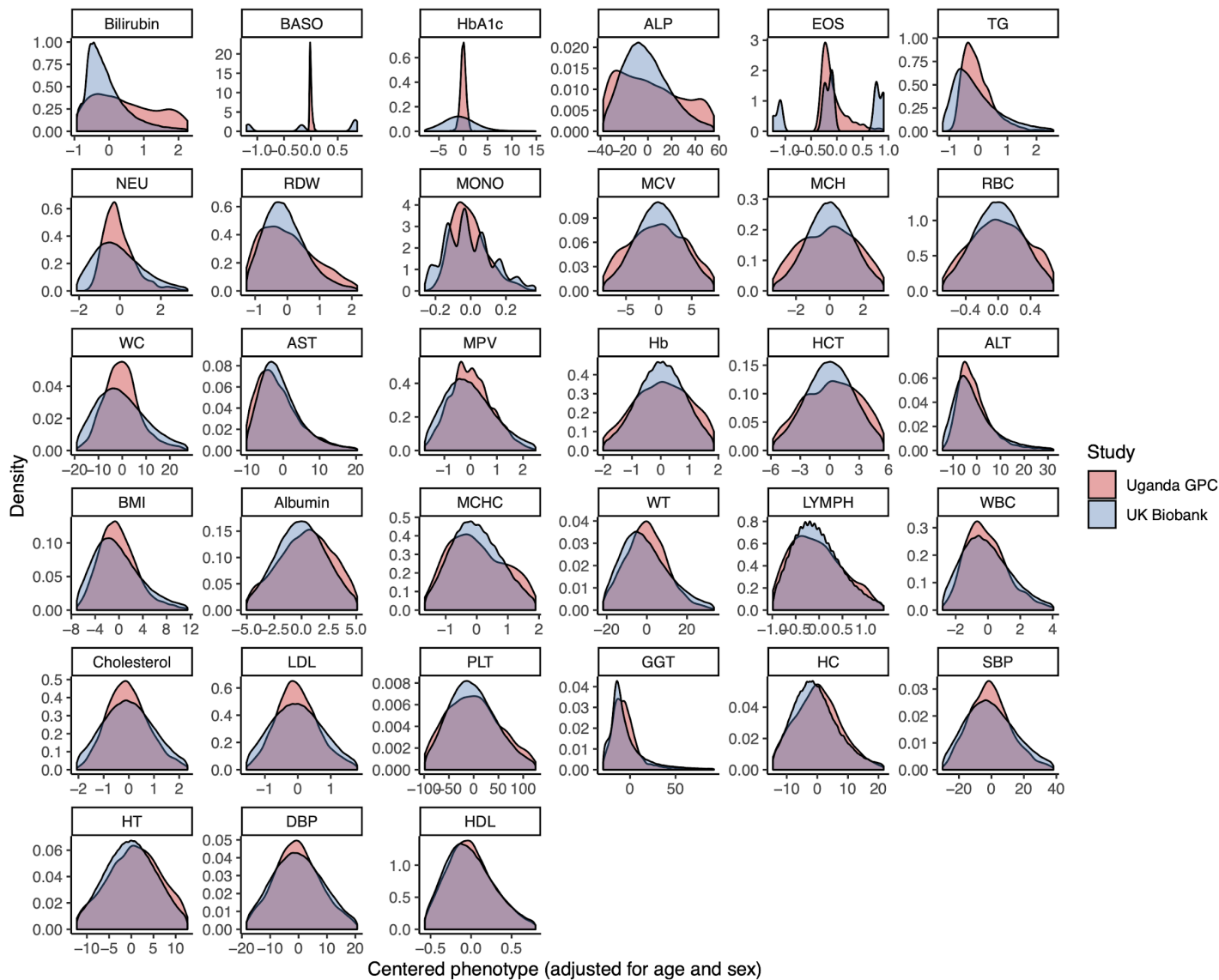


Figure S7 - Comparison of phenotypic distributions between the Uganda GPC and UK Biobank cohorts after mean centering phenotypes and regressing out effects of age and sex. The middle 95th percentile of the data are shown for display purposes such that extreme outliers are removed. Phenotypes are ordered by distributional difference as estimated by the Kolmogorov-Smirnov test statistic.

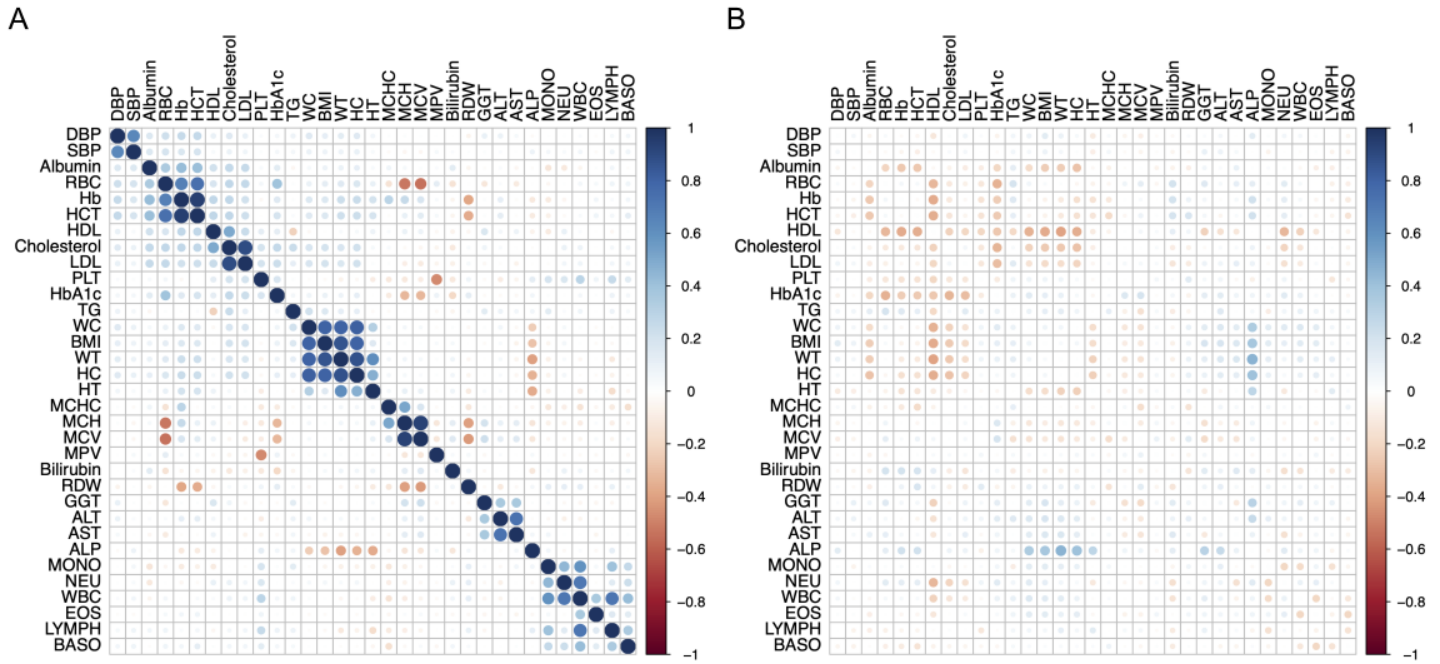


Figure S8 - Phenotype correlations among 34 quantitative traits measured in the Uganda GPC data versus UK Biobank data. Analysis conducted as in Figure 3. A) Phenotypic correlation matrix when relatives are included in the Uganda GPC. B) Differences between phenotypic correlation matrices among unrelated individuals in the UK Biobank - Uganda GPC.

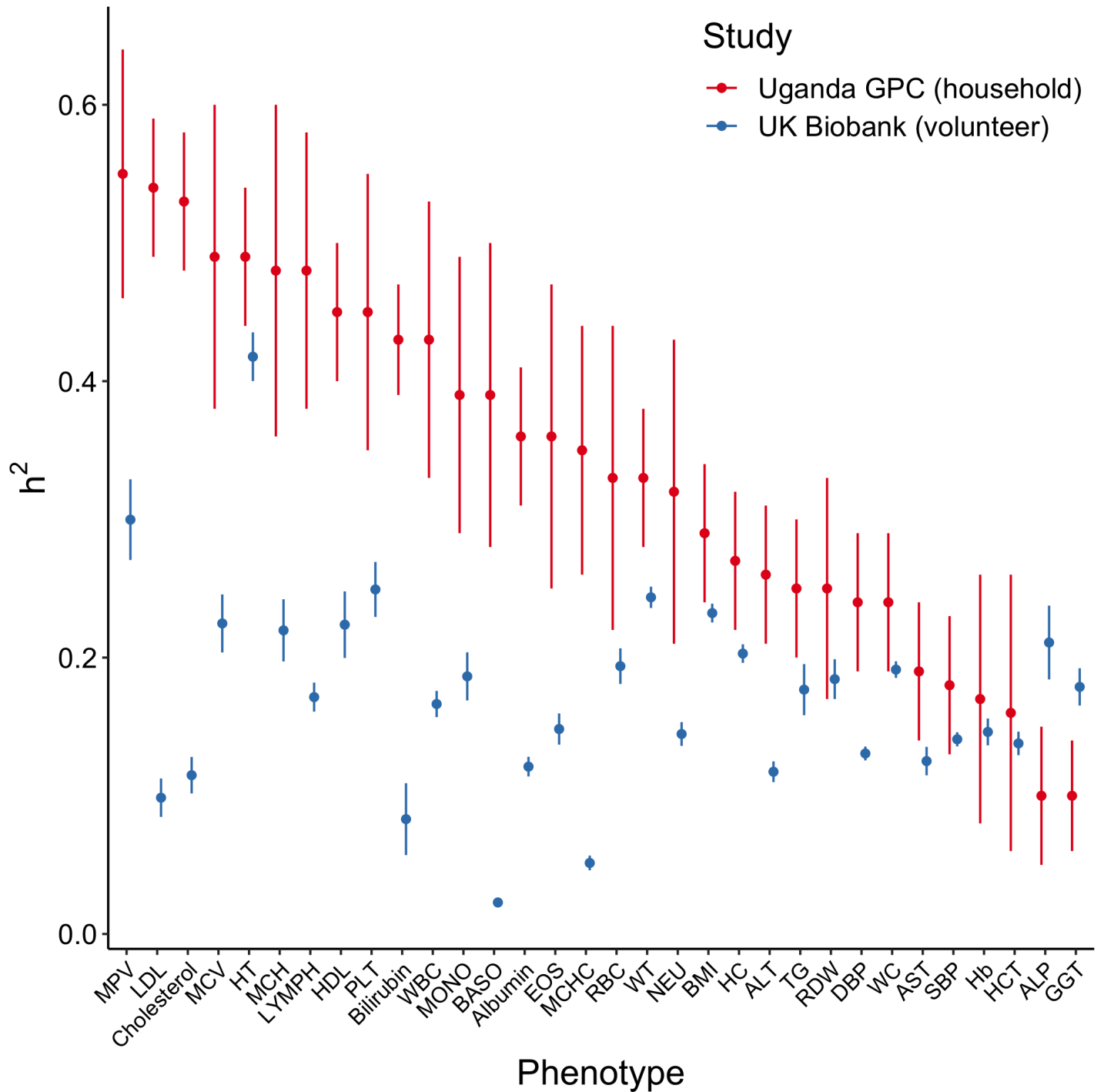


Figure S9 - Trait heritability comparison between UK Biobank and the Uganda GPC. Note that we estimated heritability in the UK Biobank individuals using LD score regression from unrelated individuals, whereas a mixed model approach with multiple random effects were used previously in Uganda GPC due to the complex household and geographically diverse study design, as described previously ([Gurdasani et al. 2019](#); [Asiki et al. 2013](#); [Heckerman et al. 2016](#)). Given these differences, the higher but noisier heritability estimates in Uganda GPC are as expected.

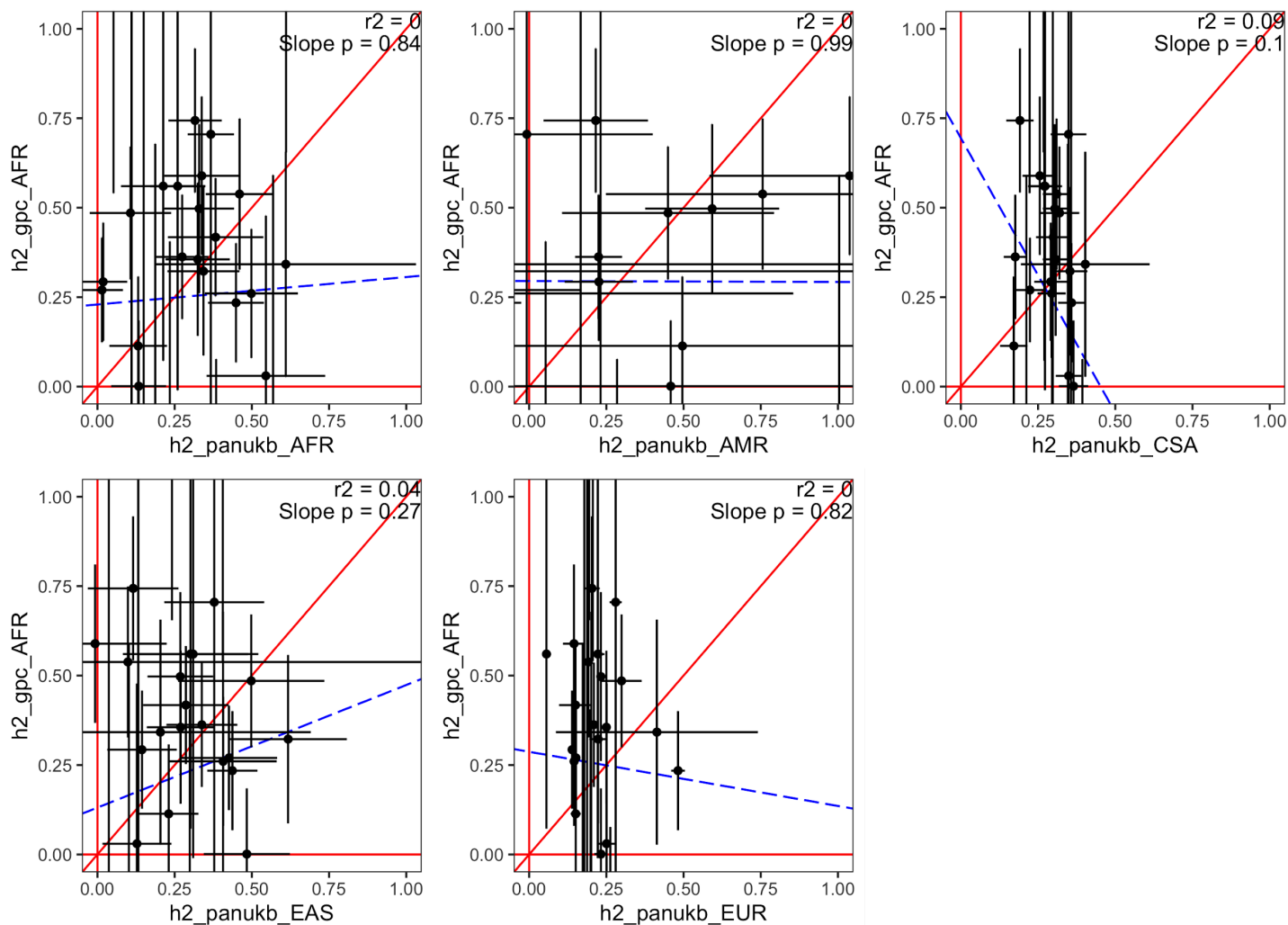


Figure S10 - Trait heritability comparison between UK Biobank and the Uganda GPC using Haseman-Elston regression. We included comparisons with each ancestry group analyzed in the Pan-UK Biobank project. The correlation and corresponding significance, respectively, of the heritability point estimates between Uganda GPC and Pan-UK Biobank for each ancestry group are shown in the top right corner of each panel. The red diagonal line shows $y=x$, whereas the blue dashed line shows the fitted simple linear regression line. The plot is zoomed in to (0,1) on both axes.

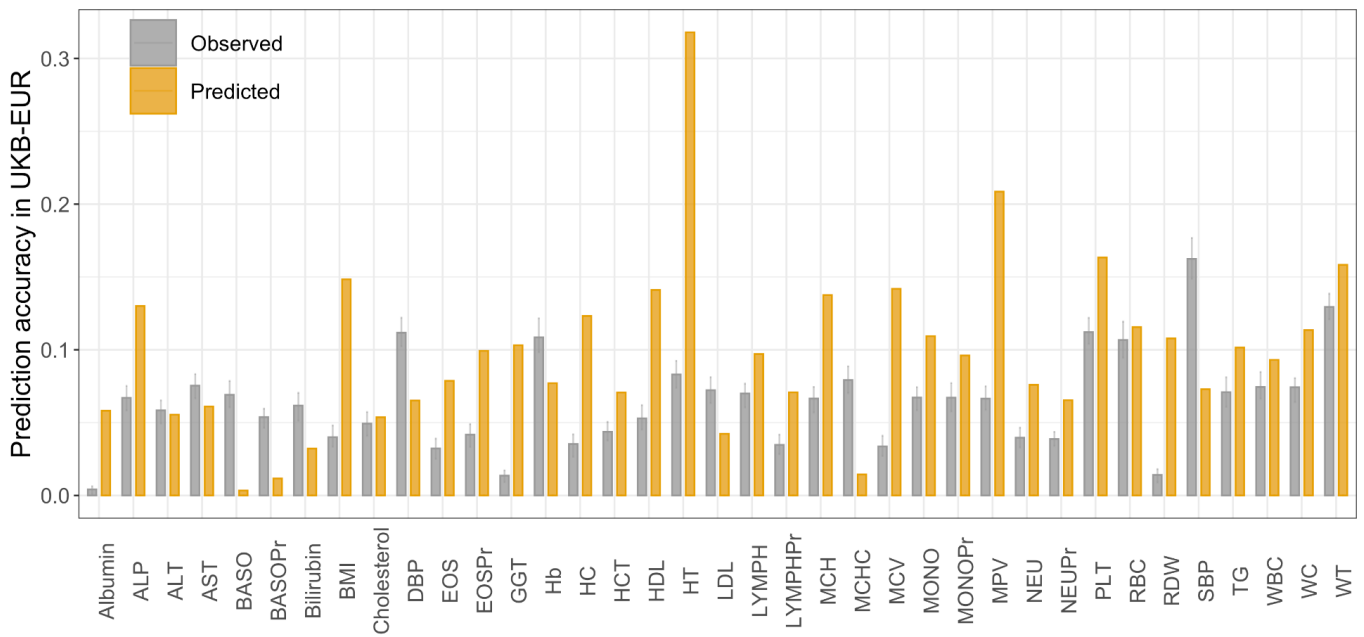


Figure S11 - Observed versus predicted PRS accuracy among traits in UK Biobank European ancestry individuals. Predicted accuracy is calculated via theoretical equations described previously ([Daetwyler et al. 2008](#); [Wray et al. 2013](#)) using estimated trait heritability estimates from the UK Biobank and the total number of trait-associated SNPs from LD clumping with p -value < 1 .

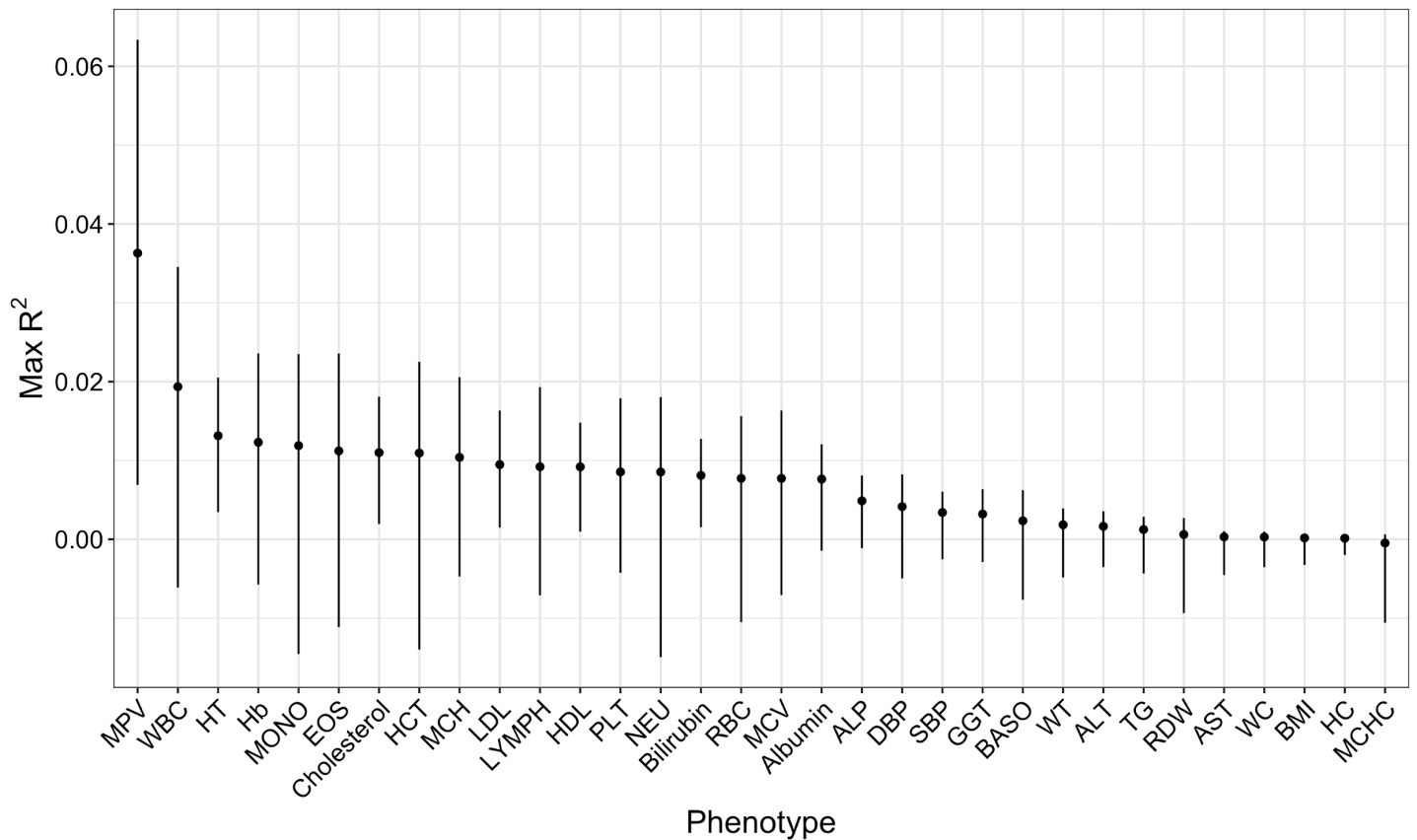


Figure S12 - PRS accuracy for 32 traits in unrelated Uganda APCDR individuals calculated using GWAS summary statistics from UK Biobank European ancestry individuals. The PRS from the p-value threshold with the highest accuracy of 10 thresholds computed is shown.

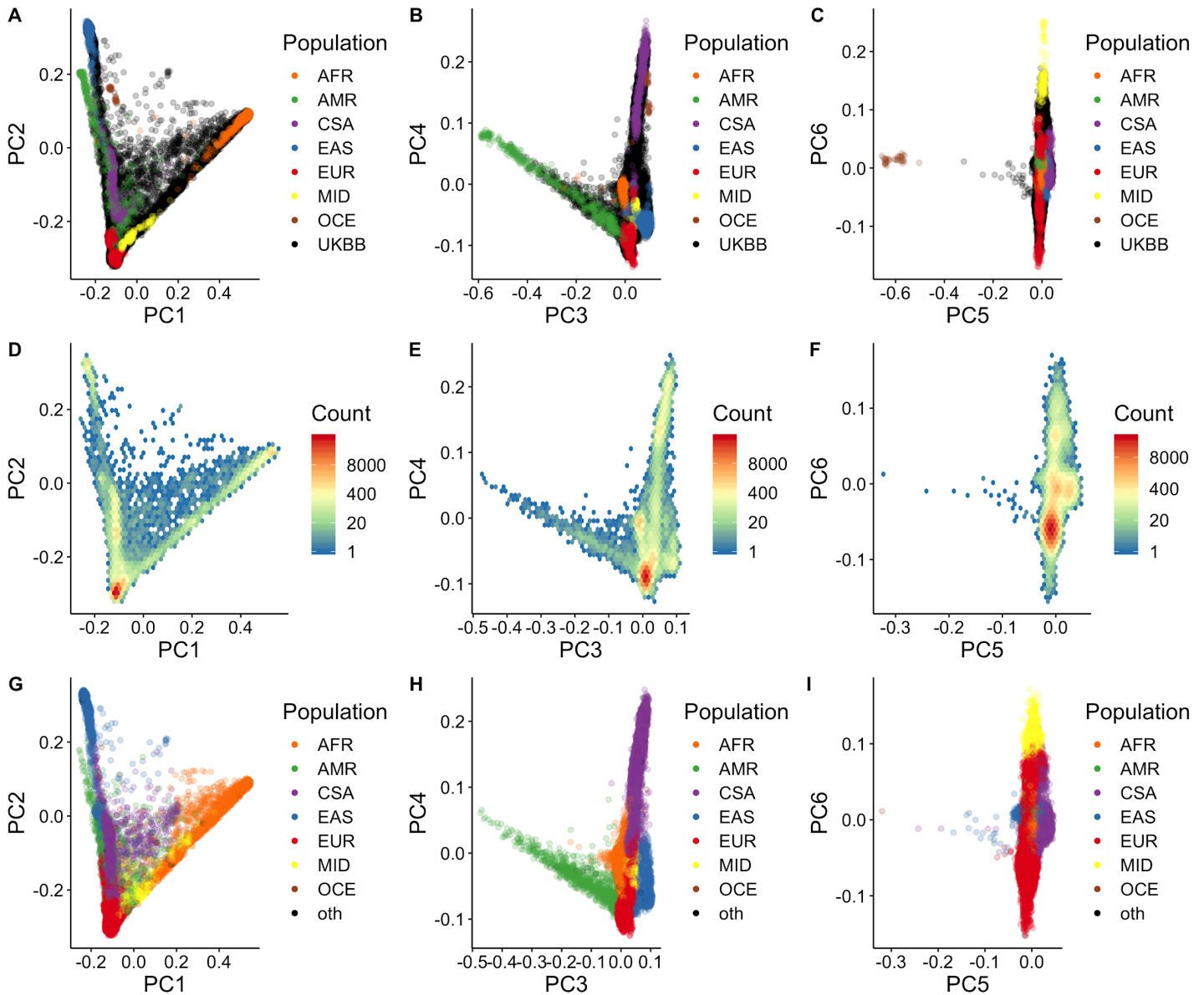


Figure S13 - Continental ancestries in the UK Biobank data with reference data from the 1000 Genomes Project and Human Genome Diversity Panel (HGDP). A-C) Principal components analysis biplots (PCs 1-6) with loadings defined by reference data from 1000 Genomes and HGDP with colors corresponding to continental ancestry meta-data from these projects. UK Biobank data (UKBB) is projected into the same PC space and colored in black. D-F) Density of UK Biobank data (excluding reference panels) in PC1-6. G-I) Continental ancestry assignments in the UK Biobank using a random forest trained on meta-data from 1000 Genomes and HGDP (excluding reference panels). “Oth” are individuals whose ancestry was not confidently assigned to any ancestry group.

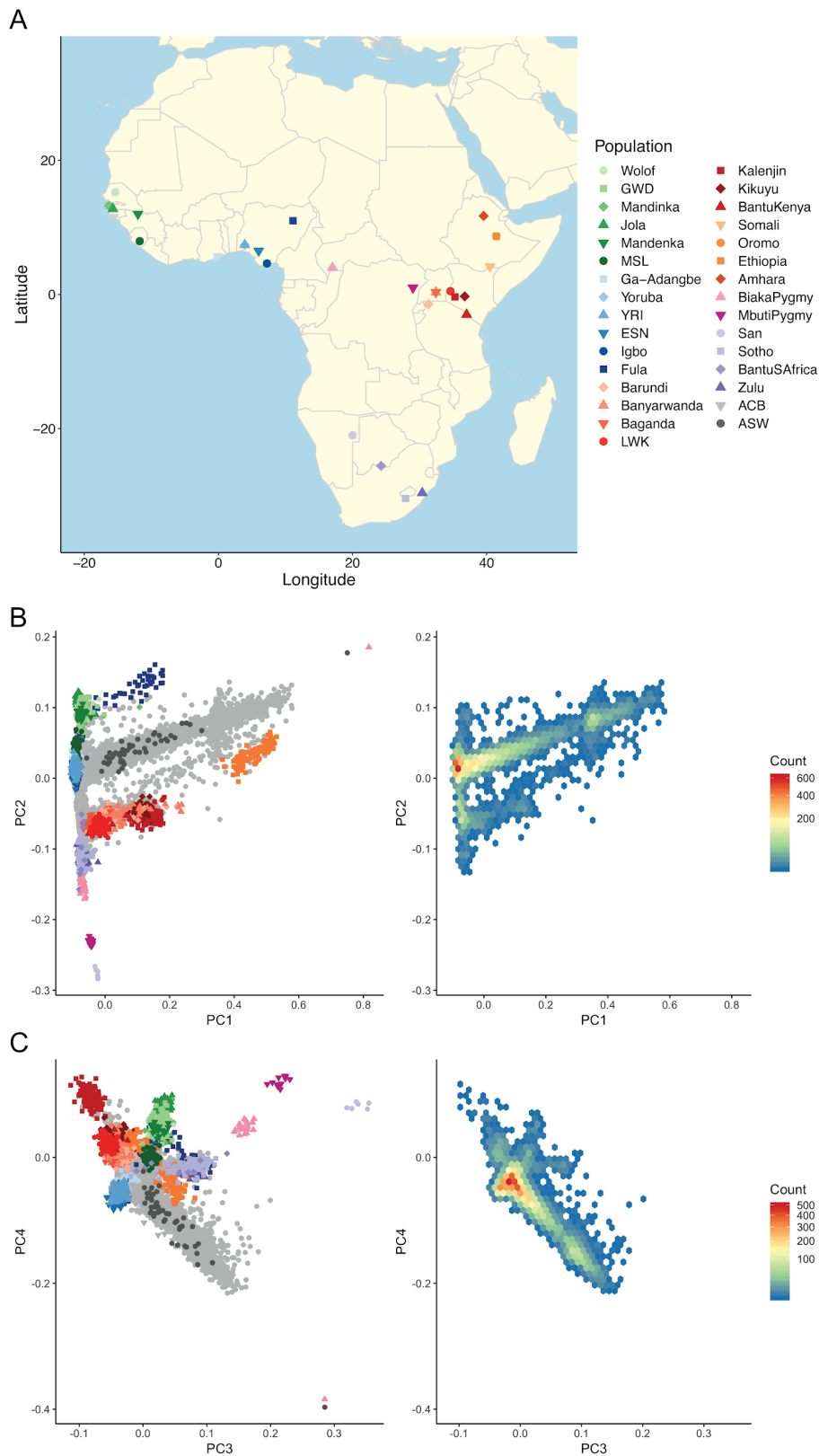


Figure S14 - African subcontinental ancestry in the UK Biobank. A) Coordinates of African reference panel data from the African Genome Variation Project, 1000 Genomes Project, and Human Genome Diversity Panel. B) PC1-2 of UK Biobank and/or reference data. C) PC3-4 of UK Biobank and/or reference data. B-C) Colors and shapes correspond to A), while grey points in the left plots are UK Biobank projected PC coordinates. Right plots show density of African ancestry UK Biobank individuals.

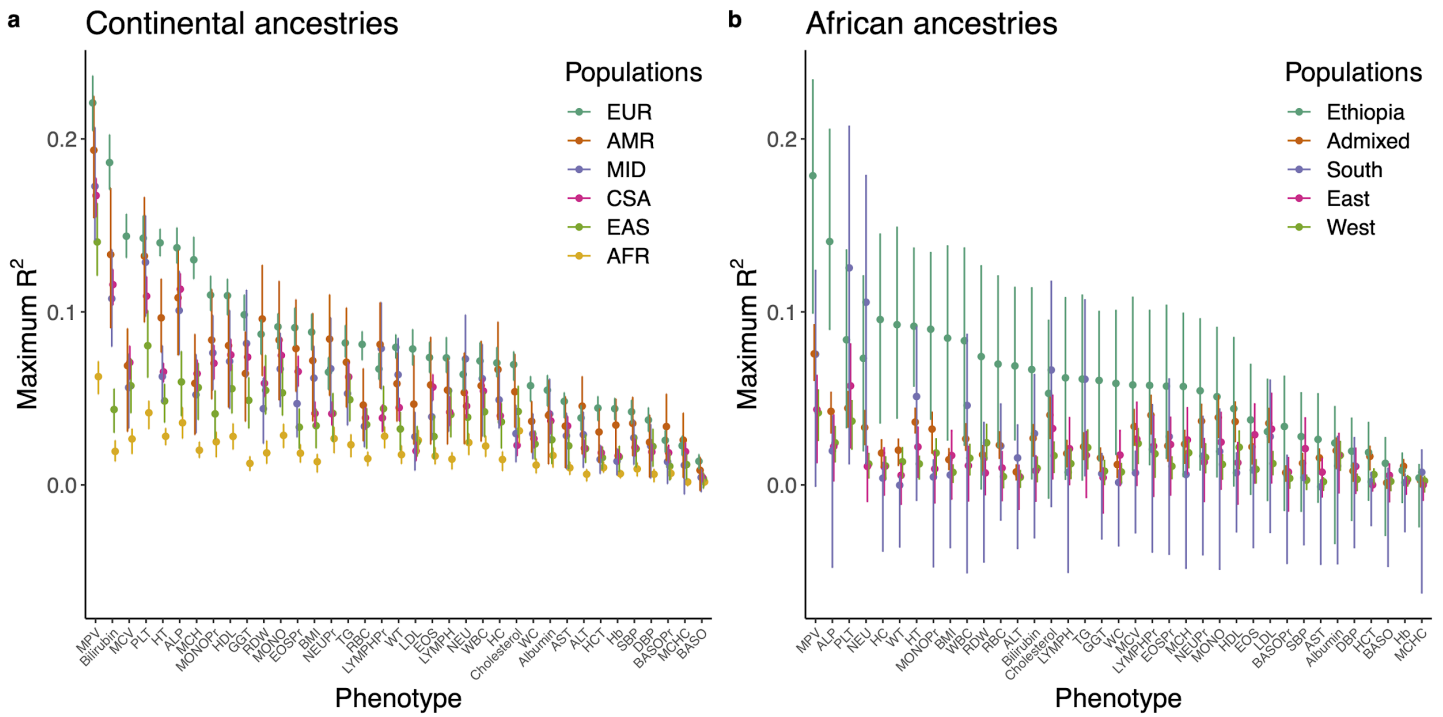


Figure S15 - Relative polygenic prediction accuracy across traits in the UK Biobank in target individuals. Panels shown different A) continental ancestries, and B) African ancestries. All individuals in B) are from the “AFR” population in A). Only one PRS is shown from each trait, which includes the maximum R^2 threshold.

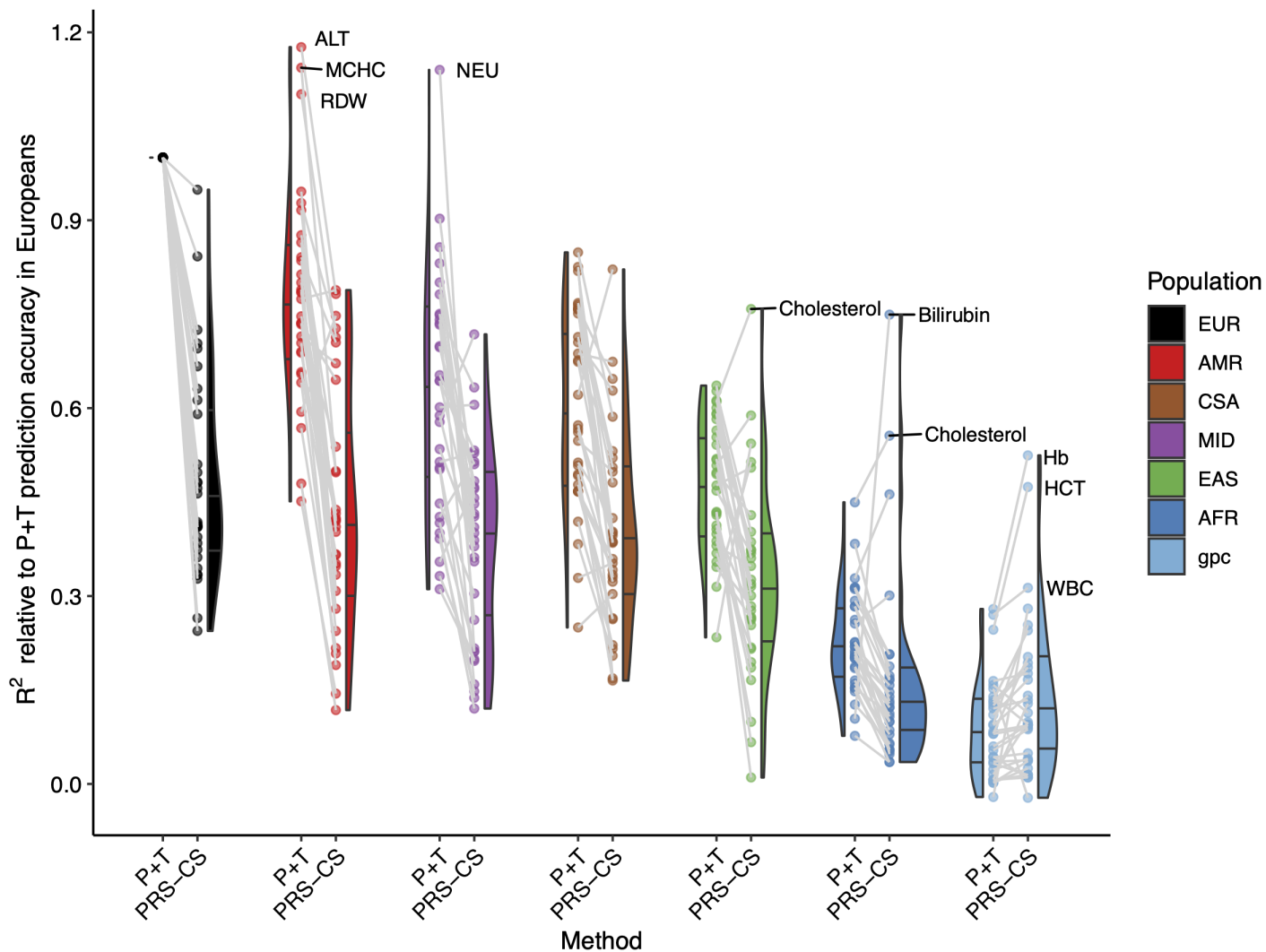


Figure S16 - Relative PRS accuracy across populations using pruning and thresholding (P+T) versus PRS-CS methods. PRS accuracy across populations and methods is normalized with respect to P+T accuracy in European ancestry individuals. Trait abbreviations are described in **Table S1**. gpc = Uganda General Population Cohort. P+T tends to outperform PRS-CS for these traits, which are mostly blood panel traits with sparse genetic architectures. A few traits with particularly high R^2 values are described.

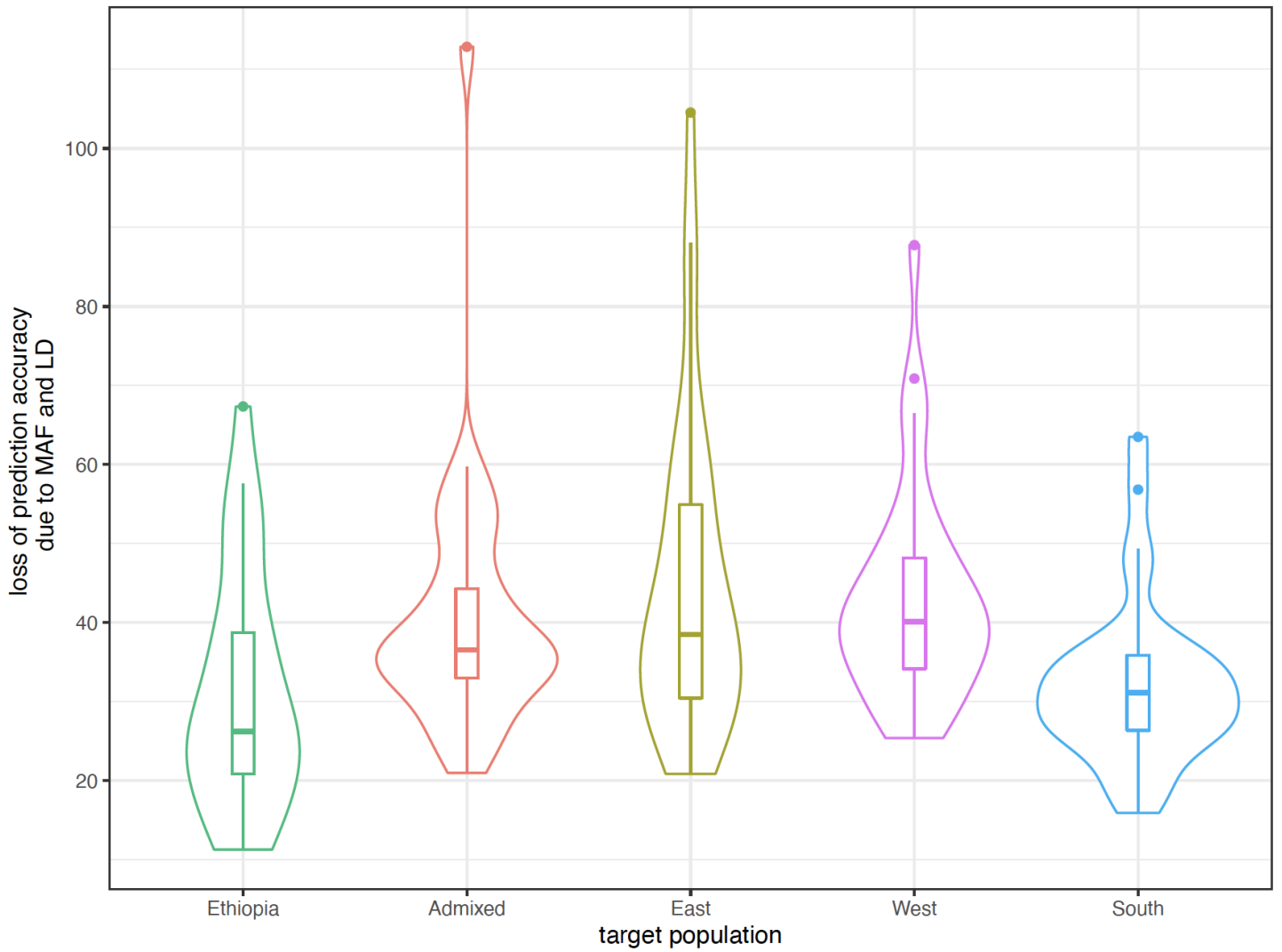


Figure S17 - Loss of accuracy across 32 traits due to linkage disequilibrium (LD) and minor allele frequency (MAF) differences between the UK Biobank European discovery and African ancestry target individuals.

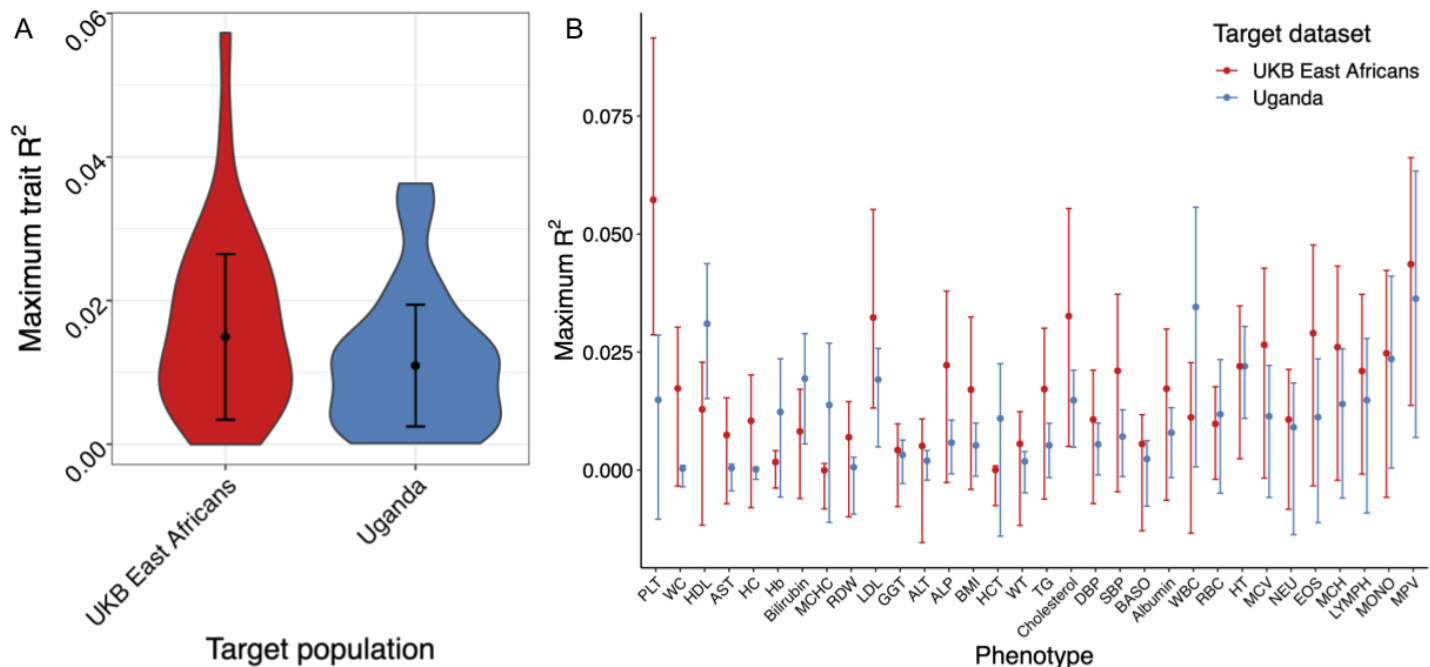


Figure S18 - Comparison of PRS accuracy for 32 traits across cohorts in individuals with East African ancestry. Includes unrelated participants in UK Biobank (N=728 with East African ancestry) versus Uganda GPC (N=2,247), using the same summary statistics from UK Biobank European ancestry individuals (N=351,194). A) Summary of accuracy distributions across 32 traits. B) PRS accuracy in each trait. Traits are ordered by absolute difference in prediction accuracy among cohorts despite similar ancestries.

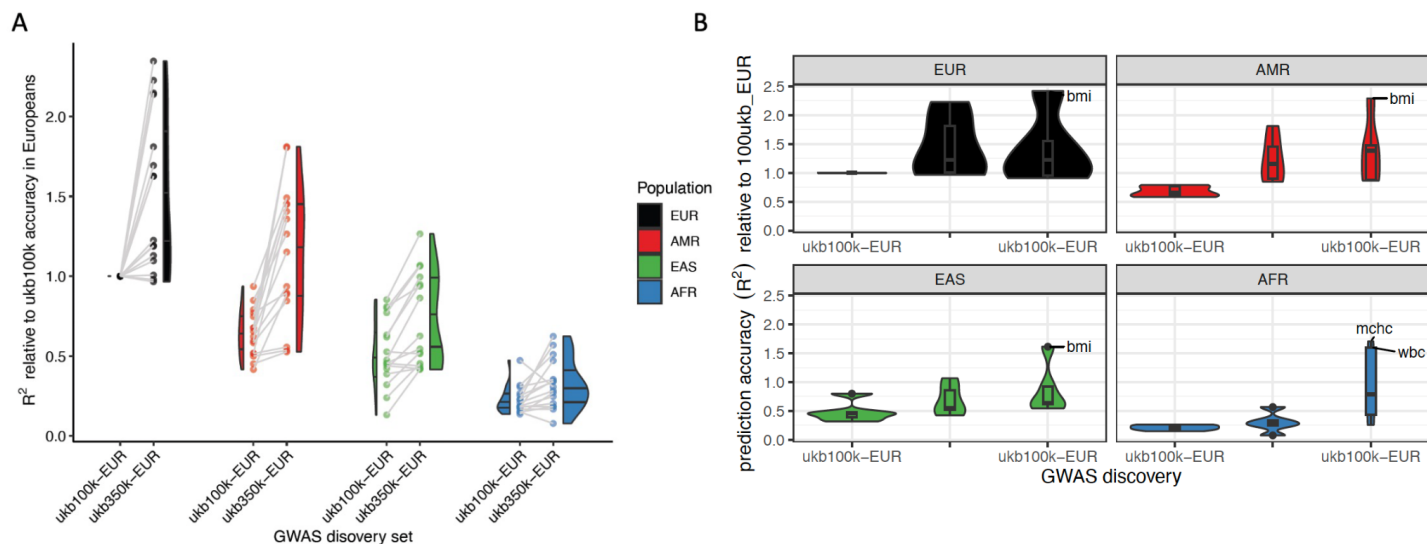


Figure S19 - A comparison of relative PRS accuracy across populations when using different GWAS discovery cohorts. A) A comparison of PRS accuracy across populations when the GWAS discovery set is 350,000 EUR individuals relative to a GWAS discovery set of 100,000 EUR individuals from UK Biobank. B) A comparison of PRS accuracy across populations when the GWAS discovery set is 350,000 EUR individuals

and multi-ancestry relative to 100,000 EUR individuals from UK Biobank. The multi-ancestry discovery set comprised 350,000 EUR from UKB, ~50,000 from PAGE and ~160,000 from BBJ (see Table S6). This comparison was done for five traits that were common between analyses, i.e. height (HT), body mass index (BMI), mean corpuscular hemoglobin concentration (MCHC), platelet count (PLT) and white blood cell count (WBC).

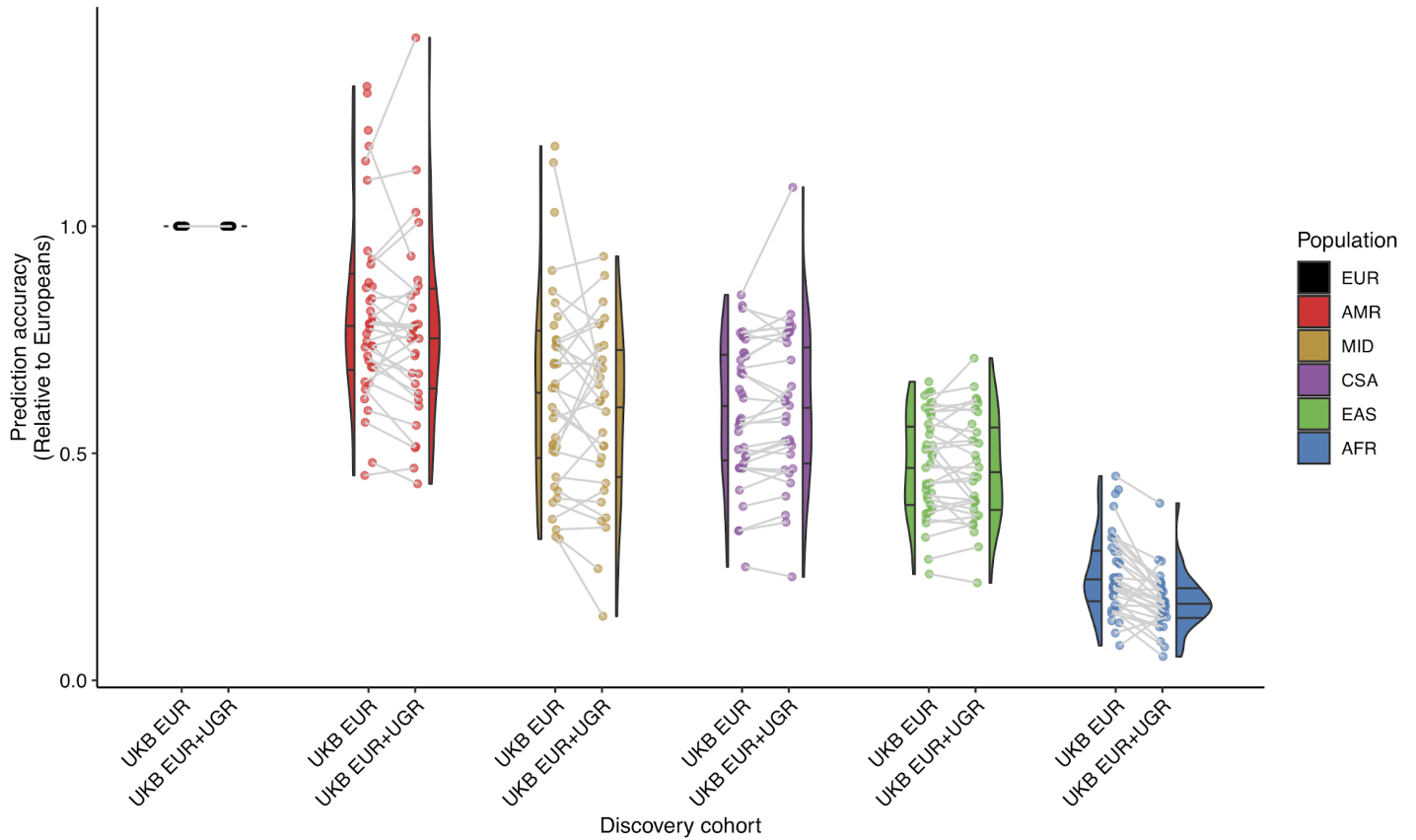


Figure S20 - Polygenic score accuracy in ancestrally diverse target populations from the UK Biobank versus meta-analysis combining UK Biobank with the Uganda Genome Resource (UGR) summary statistics.

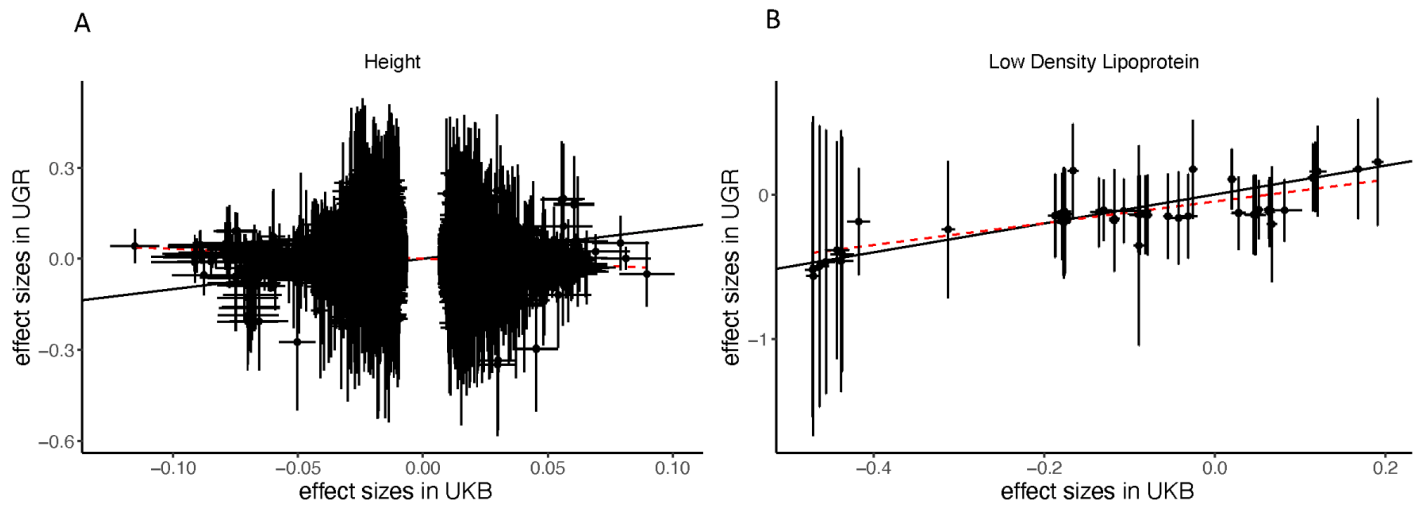


Figure S21 - Regression of SNP effects from GWAS in Uganda Genome Resource (UGR) compared to UK Biobank. A. Regression of SNP effects for SNPs with genome-wide significance in UK Biobank. There were no genome-wide significant SNPs in UGR. B. Regression of SNP effects for low density lipoprotein for SNPs that are genome-wide significant in both UGR and UK Biobank. The dashed red line is the regression line. The black line is slope =1, where $y = x$.

Tables

Table S1 - Abbreviations used throughout this study.

Table S2 - Overview of genetic and phenotypic data used throughout this study.

Table S3 - DCHS phenotype descriptions and corresponding GWAS used for prediction. Abbreviations as follows: EPDS = Edinburgh Postnatal Depression Scale, BDI = Beck Depression Index, ASSIST = Alcohol, Smoking and Substance Involvement Screening Test.

Table S4 - Uganda GPC phenotypes and corresponding GWAS discovery cohorts, including UKB, BBJ, and PAGE. Kolmogorov-Smirnov (K-S) and F-tests were used to compare overall distributions and variances, respectively. CI=confidence interval. 95% upper and lower CI bounds for the F-test are shown. K-S test statistics provided the ordering of panels in **Figure S6**.

Table S5 - Counts of UK Biobank individuals assigned to each ancestry group at the continental level globally as well as regional levels within continental African ancestry.

Table S6 - Meta-analysis discovery cohort summaries across phenotypes. Number of individuals with genotype and phenotype data included in each GWAS discovery cohort analyzed. When constructing PRS from each meta-analysis combination for each phenotype, the fraction of individuals in the meta-analysis from each continental ancestry group was used to generate a weighted LD reference panel in the 1000 Genomes Project by matching with super population metadata. For this purpose, ancestry groups in GWAS were matched to 1000 Genomes super populations as follows: UKB was matched to EUR, BBJ was matched to EAS, UGR was matched to AFR, and PAGE was matched proportionally to AFR, AMR, and EAS by ancestry composition reported for each trait. For PAGE specifically, because the 1000 Genomes Project lacks Native American reference panels, Hispanic/Latino and Native American populations were summed and matched to 1000 Genomes AMR. The relatively small numbers of Native Hawaiian and Other were not matched to any group in 1000 Genomes for constructing LD reference panels.

Table S7 - Heritability estimates in the Uganda GPC and Pan-UK Biobank data.

Table S8 - Population-enriched variants with outsized impacts on genetic prediction accuracy for traits. Lead SNPs come from associations that were genome-wide significant in both the meta-analysis and in UKB only ($p < 5e-8$). The top ten associated loci grouped by nearest gene were identified by modeling the linear relationship between $-\log_{10}(p\text{-values})$ in the meta-analysis versus UKB only, then filtering to only those variants above the 99% prediction interval (i.e. more significant in the meta-analysis than expected by a linear relationship), and then sorted by the difference in $-\log_{10}(p\text{-values})$ between the meta-analysis versus UKB alone.