Differential performance of polygenic prediction across traits and populations depending on genotype discovery approach

Yi-Sian Lin[1], Taotao Tan[1,2], Ying Wang[3,4], Bogdan Pasaniuc[5], Alicia R. Martin[3,4], Elizabeth G. Atkinson[1,6*]

1. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
2. Department of Integrative Physiology, Baylor College of Medicine, Houston, TX 77030, USA
3. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA
4. Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
5. Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
6. Jan and Dan Duncan Neurological Research Institute, Houston, TX 77030, USA

***Corresponding author contact information**:

Baylor College of Medicine

One Baylor Plaza, Taub Research Building Room 621

Houston, TX, USA 77030

Email: elizabeth.atkinson@bcm.edu

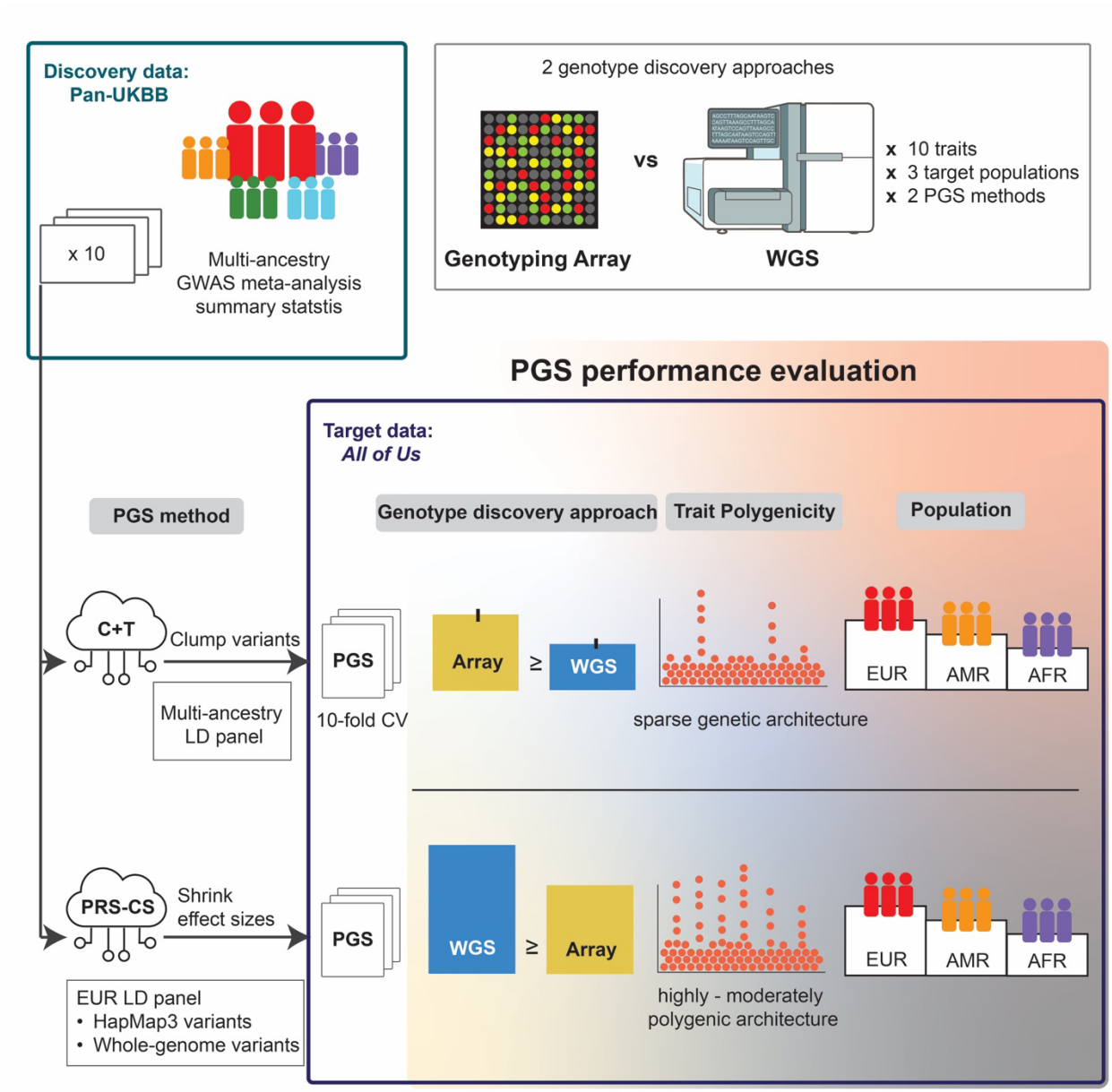
Abstract

Polygenic scores (PGS) are widely used for estimating genetic predisposition to complex traits by aggregating the effects of common variants into a single measure. They hold promise in identifying individuals at increased risk for diseases, allowing earlier screening and interventions. Genotyping arrays, commonly used for PGS computation, are affordable and computationally efficient, while whole-genome sequencing (WGS) offers a comprehensive view of genetic variation. Using the same set of individuals, we compared PGS derived from arrays and WGS across multiple traits to evaluate differences in predictive performance, portability across populations, and computational efficiency. We computed PGS for 10 traits across the spectrum of heritability and polygenicity in the three largest genetic ancestry groups in *All of Us* (European, African American, Admixed American), trained on the multi-ancestry meta-analyses from the Pan-UK Biobank. Using the clumping and thresholding (C+T) method, we found that WGS-based PGS outperformed array-based PRS for highly polygenic traits but showed differentially reduced accuracy for sparse traits in certain populations. This may be attributable to the lower allele frequency observed in clumped variants from WGS compared to arrays. Using the LD-informed PRS-CS method, we observed overall improved prediction performance compared to C+T, with WGS outperforming arrays across most non-cancer traits. In conclusion, while PGS computed using WGS generally provide superior predictive power with PRS-CS, the advantage over arrays is context-dependent, varying by trait, population, and the PGS method. This study provides insights into the complexities and potential advantages of using different genotype discovery approach for polygenic predictions in diverse populations.

Keywords

polygenic score, genotyping array, whole-genome sequencing, biobanks, global data

Graphical abstract



Introduction

Polygenic scores (PGS), also known as polygenic risk scores (PRS), estimate an individual's genetic predisposition to complex traits or diseases. As an emerging tool in precision medicine, PGS holds significant potential for informing early disease detection, risk stratification, prevention, and intervention[1]. PGS is calculated by aggregating the genotypes (i.e., the dosage of minor or risk alleles) from a target cohort, weighted by their corresponding effect sizes derived from relevant genome-wide association studies (GWAS) conducted in a discovery cohort[2]. The prediction accuracy of PGS is influenced by several factors, such as the sample size of the discovery cohort and the genetic similarity between the discovery and target population[3,4]. However, with up to 90% of GWAS for complex traits conducted in populations with predominantly European ancestry[5,6], PGS estimates are often less accurate for individuals of non-European ancestries, which can contribute to health disparities across populations[7].

Genotype data from both discovery and target cohorts are typically obtained using array technologies that detect known genetic polymorphisms. Since 2005, Affymetrix has developed high-throughput microarray genotyping platforms widely used in early GWAS and PGS studies[8,9], which initially assayed between 500,000 and nearly a million SNPs[10]. More recently, Illumina's Omni series arrays expanded this coverage to detect millions of genetic markers[11]. In addition to capturing genetic variation in European populations, arrays designed for global and diverse populations, such as the Global Screening Array[12], Global Diversity Array[13,14], Asian Screening Array[15], and H3 Africa Array[16,17], became available following the completion of the 1000 Genomes Project (1kGP) Phase 3[18] and other diverse reference resources. To allow for more comprehensive genome-wide coverage, imputation is often employed to infer untyped variants, thereby enhancing the accuracy of GWAS and PGS predictions[19].

With the ongoing reduction in costs and advancements in high-throughput sequencing, both whole exome sequencing (WES) and whole genome sequencing (WGS) have become increasingly integrated into population-level genomic studies and clinical applications[20,21]. In contrast to array-based genotyping, sequencing offers several distinct advantages. Unlike genotyping arrays, which are typically designed to assay common variants and require imputation to enhance coverage, sequencing captures all forms of genetic variants ranging from ultra-rare to common alleles[22,23]. Additionally, imputation accuracy for genotyping arrays can impact the predictive performance of PGS[24], particularly in non-European populations. This is due to their inability to reliably impute variants with low frequency, and variants enriched in populations that are not well represented in existing reference panels. WGS enables the direct assessment of low-frequency variants and loci that are missed or poorly imputed in arrays. Therefore, these advantages could potentially lead to more accurate and robust PGS, especially for underrepresented populations and traits influenced by rarer variants. However, WGS remains more expensive, and generates vast amounts of data which requires computationally intensive processing.

Previous studies have demonstrated comparable results between low-coverage sequencing and array-based genotyping in gene discovery[25–27] and PGS for diseases like Parkinson's disease[28]. However, few studies have assessed the impact of different genotype discovery technologies on PGS performance at the biobank scale or considered high-coverage WGS versus arrays. The *All of Us* Research Program[14] (V6), with 95,562 samples with both array and WGS data, provides a unique opportunity to evaluate the predictive accuracy across technologies in multiple ancestral groups and numerous phenotypes. In this study, we benchmarked the performance of PGS derived from genotyping arrays and 30x high-coverage WGS across ten traits and three populations in *All of Us*. To evaluate differential prediction accuracy across diverse populations, we trained 10 PGS using meta-analyzed GWAS summary

statistics from multiple ancestries in the Pan-UK Biobank (Pan-UKBB)[29] and computed PGS on three major genetic ancestry groups in the *All of Us* cohort (**Figure 1**). Through these efforts, we aim to assess the conditions under which WGS-based PGS outperforms array-based PGS and identify trade-offs that may inform the choice of genomic data type for risk prediction in diverse settings.

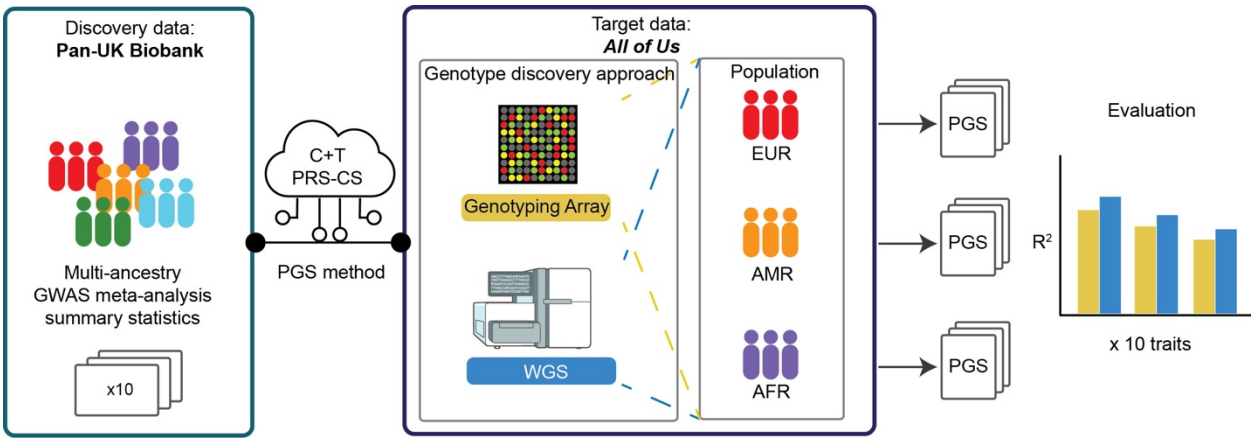


Figure 1: Study design. For each trait in *the All of Us* cohort, we downloaded the corresponding summary statistics from the Pan UKBB to obtain estimated effect sizes. PGS were calculated for each sample in *All of Us* from the EUR, AMR, or AFR populations, using genotypes derived from either array or WGS approaches. Two PGS methods were benchmarked across the entire pipeline: C+T and PRS-CS. The color codes used throughout the paper are as follows: [yellow] array, [blue] WGS, [red] EUR, [orange] AMR, and [purple] AFR.

Results

Selection of traits across a range of genetic architectures

The primary goal of this study is to benchmark the predictive performance from array-based versus WGS-based genotypes. We first selected 95,562 target samples from the *All of Us* V6 dataset, each with both array and WGS genotype data. PGS are expected to exhibit higher predictive accuracy when the ancestry of the discovery and target cohorts is matched[30]. However, since this study aims to develop a generalized PGS model rather than focusing on a single ancestry, we used effect sizes derived from a multi-ancestry GWAS meta-analysis of the Pan-UKBB cohort. For the target samples in *All of Us*, we focused on European (EUR), Latino/Admixed American (AMR), and African/African American (AFR) cohorts based on predicted genetic ancestry (**Figure S1**), as these groups had sample sizes sufficiently large to support multi-trait comparative analyses.

We selected 10 complex traits with distinct genetic architectures (**Table S1**) from the Pan-UKBB and their equivalents in the *All of Us* dataset (**Table S2**): standing height (Height), diastolic blood pressure (DBP), white blood cell count (Leukocyte), red blood cell count (RBC), high-density lipoprotein cholesterol (HDL), total cholesterol (TC), asthma, breast cancer, colorectal cancer, and type 2 diabetes (T2D). Height exhibits the highest heritability among the traits, while cancers show the lowest heritability (**Figure S2**). The polygenicity of the traits was indirectly assessed by the number of GWAS hits, with height displaying thousands of significant loci, whereas binary traits, such as cancer, show only a few significant peaks (**Figure S3**).

WGS does not improve prediction accuracy over array when variants are clumped

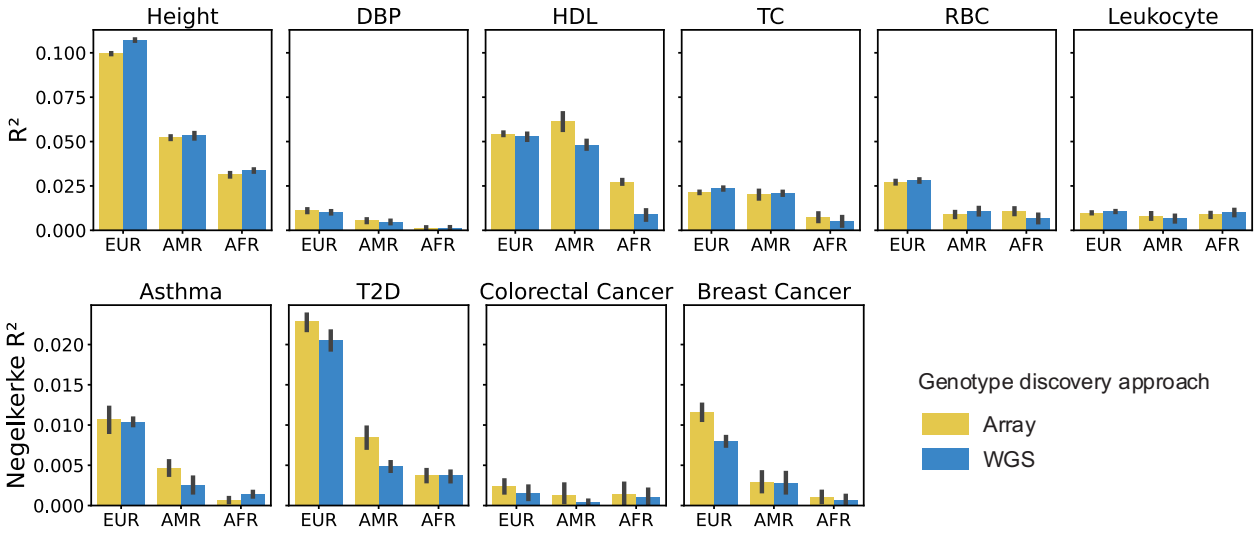
We first assessed the PGS using the classic clumping and thresholding (C+T) method. Notably, the raw average PGS for array-based and WGS-based genotypes were highly correlated across

146 traits ($r > 0.9$ for all traits except for height and TC), suggesting that both platforms captured similar
147 genetic information, resulting in comparable polygenic predictions. This also indicates that the
148 genotype coverage from unimputed array data is sufficient for constructing reliable PGS
149 comparable to those from WGS in the *All of Us* dataset for most of the traits considered.
150 Furthermore, it is noteworthy that EUR populations tend to show smaller and more homogenous
151 raw PGS, whereas AFR populations tend to exhibit larger and more widely spread PGS,
152 particularly for colorectal cancer and breast cancer (**Figure S4**).

153 The predictive accuracy, measured by R^2 or Nagelkerke's R^2, was assessed across different
154 genotype discovery approaches and populations (**Figure 2**, **Table S4** and **5**). Given that EUR
155 individuals constitute a large proportion of both the discovery cohort (95%) and the target cohort
156 (52%), predictive accuracy in EUR was generally higher than in other populations, as expected,
157 regardless of the genotyping approach (ranging from $1.58 \times 10^{-3} \pm 1.28 \times 10^{-3}$ to $1.07 \times 10^{-1} \pm$
158 6.12×10^{-4}). AFR populations generally showed lower performance compared to other groups
159 (ranging from $7.33 \times 10^{-4} \pm 7.53 \times 10^{-4}$ to $3.37 \times 10^{-2} \pm 1.07 \times 10^{-3}$, **Table 1**). Surprisingly, PGS
160 performance from WGS genotypes was suboptimal relative to array genotypes for multiple traits,
161 and only outperformed arrays in all three populations for one trait - height. This discrepancy may
162 be due to height's highly polygenic nature, suggesting that the higher number of variants detected
163 by WGS provides more informative variants for PGS prediction than array genotypes. Additionally,
164 there are a few phenotypes with notable differences in R^2 across genotype discovery approaches.
165 In addition to height, prediction from WGS also performed better for TC, RBC, and leukocyte
166 count in EUR, and for leukocyte count and asthma in AFR. In contrast, we observed that PGS
167 derived from array genotypes performed better for several other traits, particularly HDL in AMR
168 (average array-based R^2=0.061; average WGS-based R^2=0.048; paired t-test $p = 3.43 \times 10^{-4}$)
169 and AFR populations (average array-based R^2=0.027; average WGS-based R^2=0.009; paired t-
170 test $p = 7.7 \times 10^{-7}$), as well as RBC in AFR, T2D in AMR, and breast cancer in EUR (**Table 1** and

171 **Table S6**). Several findings from previous studies may provide insights for interpreting these
172 observations, including aspects of population-enriched architecture, as detailed in **Table S7**.

173

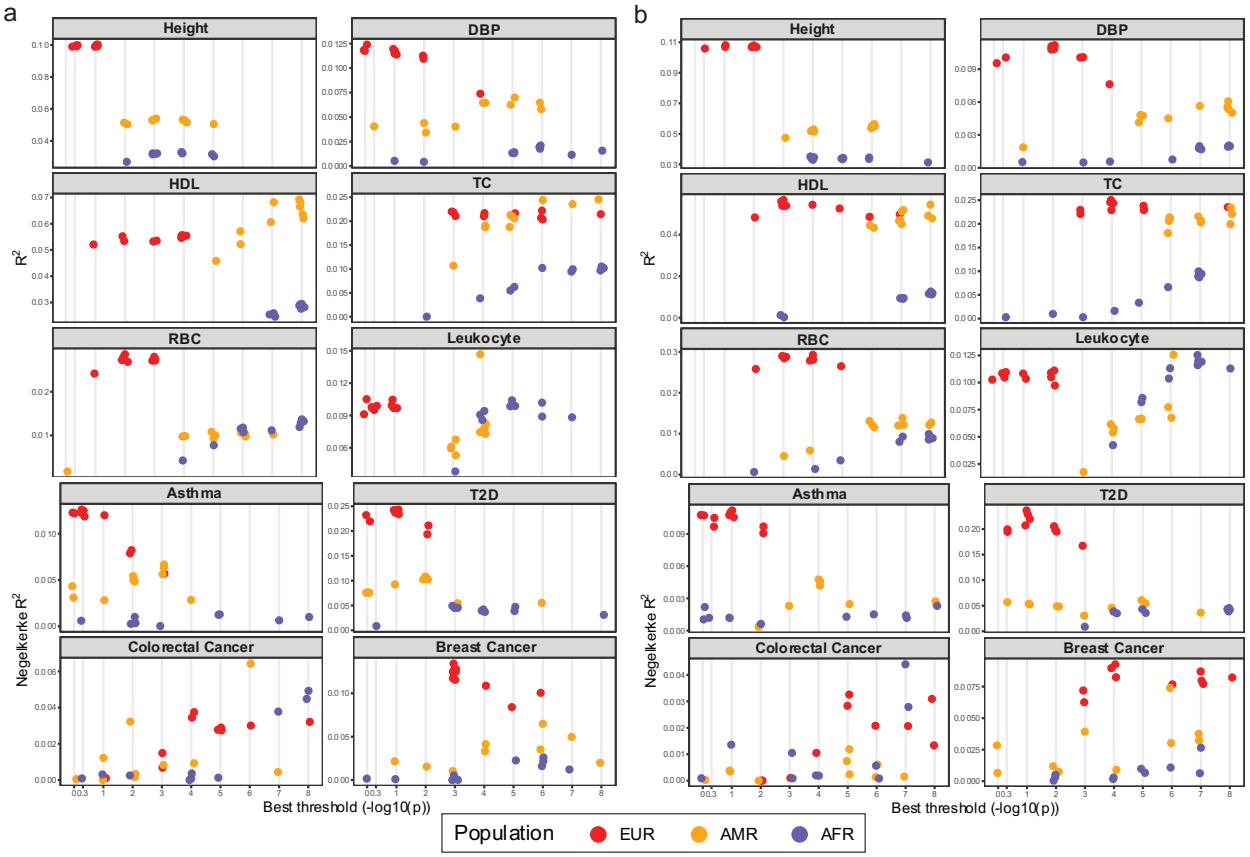


174

175 **Figure 2: Comparison of PGS performance across genotype discovery approaches using**
176 **C+T.** PGS were computed for samples with array- and WGS-genotypes in *All of Us*. R^2 or
177 Nagelkerke's R^2 were determined by adjusting for age, sex, and the first 16 genetic PCs. Error
178 bars indicate 95% confidence intervals.

179 More variants were called from the WGS platform compared to the array (8,996,707 versus
180 975,876 after quality control, QC); however, a higher number of variants did not generally result
181 in better predictive accuracy. We next examined the relationship between the number of variants
182 used in PGS calculation and PGS performance. Based on the optimal *p*-value thresholds
183 determined through 10-fold cross-validation, the number of variants selected for PGS calculation
184 decreased as the thresholds became more stringent (**Figure S5**). We observed that the best PGS
185 performance in EUR populations tended to occur at looser thresholds for most traits, with the
186 number of variants selected positively correlated with improved performance. The greater number
187 of variants selected for PGS calculation in EUR, relative to other populations, may reflect the

188 higher informativity of genetic data from EUR, which is the largest cohort in the training dataset.
189 In contrast, for traits such as HDL, TC, RBC, and leukocyte, fewer variants selected for PGS were
190 associated with higher performance in the AFR and AMR populations (**Figure 3**).



191

192 **Figure 3: Relationship between C+T p-value threshold and prediction accuracy. a.** Scatter
193 plot depicting the best *p*-value thresholds and the corresponding prediction performance derived
194 from array variants. **b.** Scatter plot depicting the best *p*-value thresholds and the corresponding
195 prediction performance derived from WGS variants. Each population is represented by 10 dots,
196 with each dot corresponding to the best PGS from a threshold using 10-fold cross-validation.

197

198 **Allele frequency is higher in array variants than in WGS variants**

199 For common traits, variants with higher allele frequencies are more likely to explain larger portions
200 of genetic variation. We hypothesized that variants with higher allele frequencies contribute to
201 greater explanatory power for PGS because they are more frequently present in samples. To test
202 whether this assumption explains the lower R^2 observed with WGS-based PGS compared to
203 array-based PGS for traits such as HDL and others, we computed the allele frequencies of
204 variants from both array and WGS data, stratified by population. Using HDL as an example, we
205 found that the allele frequencies of 464,726 WGS SNPs were lower than those of 261,583 array
206 SNPs for HDL after LD clumping, regardless of population (**Figure S6a, b**). The allele frequencies
207 of overlapping variants from both genotyping platforms were highly correlated (r > 0.999),
208 indicating strong concordance between array and WGS variant calls (**Figure S6c**). Although the
209 allele frequency patterns were largely consistent across traits (**Figure S7a, b**), these findings
210 provide some support for a hypothesis that lower allele frequencies in WGS variants contribute to
211 poorer predictive accuracy compared to array variants, particularly for less polygenic traits.

212 Using HDL as an example, population comparisons revealed that jointly QC'd variants for all
213 samples in the *All of Us* dataset, after clumping, captured a higher proportion of variants with a
214 minor allele frequency (MAF) < 0.01 in the EUR population alone. Conversely, the allele frequency
215 of clumped variants in the AFR population was higher than in other populations (**Figure S6a, b**).
216 We considered whether further removal of relatively rare variants in AFR would improve polygenic
217 prediction for HDL. To test this, we performed additional population-specific QC by filtering out
218 WGS variants with MAF < 0.01 in AFR, beginning from variants that were already QC filtered
219 jointly in *All of Us*. Using C+T, this approach resulted in an increase in the mean MAF from
220 0.14 ± 0.23 to 0.17 ± 0.24 (**Figure S8a**) and a slight improvement in predictive performance (R^2
221 from 0.009 to 0.01) compared to the full jointly-QC'd variant set (**Figure S8b**), suggesting that the
222 removal of relatively rare variants in AFR modestly reduced noise in polygenic prediction. This

223 finding further suggests that population-specific QC is likely to result in optimal selection of
224 variants for PGS in populations less well represented in large datasets.

225 C+T retains only quasi-independent variants within each LD block after LD clumping; however,
226 this approach may remove informative variants important for PGS calculation. In contrast, PRS-
227 CS does not rely on clumping but instead continuously shrinks effect size estimates in an LD-
228 informed manner to retain as many informative variants as possible[31]. We benchmarked C+T and
229 PRS-CS to determine whether prediction accuracy would differ between the two methods across
230 data generation technologies. For raw PGS using PRS-CS, we again observed a high correlation
231 between PGS derived from array-based genotypes and WGS genotypes across traits ($r > 0.9$ for
232 continuous traits). The range of raw PGS was largest for height, but smaller for binary traits. All
233 populations covered a broad spectrum of PGS values (**Figure S9**). When comparing the
234 predictive accuracy of PGS generated using PRS-CS to those generated using the C+T method,
235 we found that PRS-CS generally outperformed C+T, with performance being approximately two
236 times better for most traits, except for cancers. Additionally, performance using WGS variants was
237 higher than that using array variants for all traits except for cancers and TC. The relative gain in
238 performance using WGS compared to arrays was similar across populations, but smaller in AFR
239 for asthma and T2D (**Figure 4**).

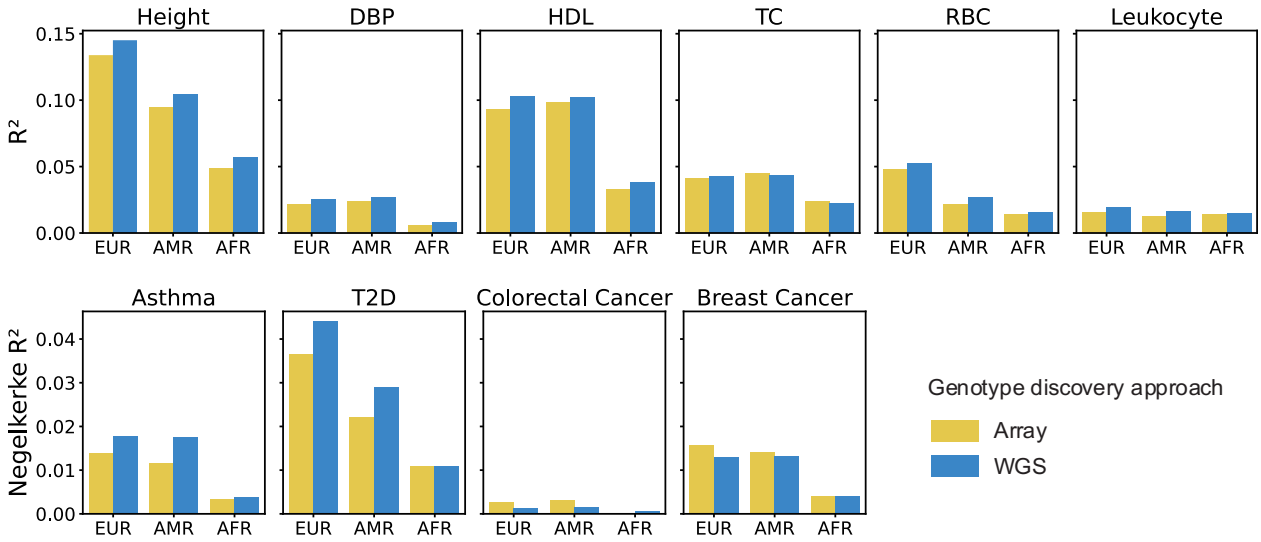


Figure 4: Comparison of PGS performance across genotype discovery approaches using PRS-CS. PGS were computed for samples with array- and WGS-genotypes in *All of Us*. R^2 or Nagelkerke's R^2 were determined by adjusting for age, sex, and the first 16 genetic PCs.

WGS outperforms arrays and achieves maximal accuracy with continuous effect size adjustment rather than LD clumping

Note that the variants selected for PGS computation were restricted to approximately 1.1 million HapMap3 variants, as the default LD panel used in PRS-CS was based on HapMap3 variants. This is reflective of the general implementation of the PRS-CS software. To fully assess the advantage of using WGS genotypes, however, which include around 9 million variants, we additionally computed an EUR-derived LD matrix using samples from the 1kGP and the Human Genome Diversity Project (HGDP) joint call sets, resulting in approximately 7.3 million reference variants. In the *All of Us* dataset, the intersection between the QC'd array data and the HapMap3 LD matrix was approximately 0.3 million variants (~29% of the array variants), while the intersection with the full-scale LD matrix was around 0.7 million variants (~71% of the array

256 variants). For the QC'd WGS data, the intersection with the HapMap3 LD matrix was
257 approximately 0.9 million variants (~10% of the WGS variants), while the intersection with the full-
258 scale LD matrix was around 5.5 million variants (~62% of the WGS variants). Switching from the
259 default LD panel to this new panel, we observed similar levels of predictive accuracies for traits
260 using array genotypes, but lower accuracy using WGS variants in AMR and AFR for multiple traits
261 (**Figure S10**, **Table S8** and **9**). This could be attributable to the higher number of variants in the
262 EUR LD panel, which does not tag well for non-EUR population. However, taken together, LD
263 information from HapMap3 variants is generally sufficient to perform PRS-CS and achieve optimal
264 performance without the computationally demanding need to construct fuller LD resources.

265 **Trade-off between computational efficiency and prediction accuracy**

266 In terms of computational comparisons, C+T is much faster than PRS-CS. Focusing on post-QC
267 computation, C+T generally took less than 20 minutes of cloud computing per trait in the *All of Us*
268 dataset during the processes of clumping and determining the best PGS using 10-fold cross-
269 validation. In contrast, inferring posterior effect sizes with PRS-CS required 9 to 20 hours using a
270 local server outside *All of Us*, and up to 4 weeks when using a full-sized LD matrix with WGS data
271 (**Table S10**). The actual PGS computation in *All of Us* following this step also took 4 to 30 minutes
272 per trait, depending on the data scale, using 80 to 160 compute cores. Once the panel was
273 created, the PGS computation time using the full-scale LD panel upstream is shorter compared
274 to the runtime with the default LD panel. This improvement is due to modifications made to the
275 PRS-CS source code to recognize hg38 variant IDs in a format optimized for Hail genomics
276 analysis in *All of Us*. The default PRS-CS, in contrast, only recognizes hg19 rsIDs.

277 In conclusion, the performance of polygenic prediction varies across genotype discovery
278 approaches, PGS methods, traits, and populations. WGS data, when used with LD-informed
279 methods like PRS-CS, generally provides higher predictive power compared to array data,

280 especially for traits with higher heritability. However, array-derived PGS are more cost-effective
281 and computationally efficient, making them a favorable option for rapid generation, particularly for
282 traits with sparser genetic architectures. This study provides insights for researchers in selecting
283 the most suitable genotype discovery approach and PGS method for prediction, considering the
284 specific traits and populations involved.

285

286 **Table 1. Comparison of polygenic prediction across traits and genotype discovery**
287 **approaches using C+T and PRS-CS.**

Phenotype	Population	R^2			
		C+T		***PRS-CS***	
		Array	**WGS**	**Array**	**WGS**
Height	EUR	9.95E-02 (4.61E-04)	1.07E-01 (6.12E-04)	1.34E-01	1.45E-01
	AMR	5.22E-02 (1.19E-03)	5.34E-02 (2.64E-03)	9.45E-02	1.04E-01
	AFR	3.14E-02 (1.75E-03)	3.37E-02 (1.07E-03)	4.82E-02	5.68E-02
DBP	EUR	1.12E-02 (1.41E-03)	1.02E-02 (1.07E-03)	2.16E-02	2.54E-02
	AMR	5.43E-03 (1.31E-03)	4.78E-03 (1.17E-03)	2.37E-02	2.68E-02
	AFR	1.36E-03 (5.50E-04)	1.37E-03 (6.89E-04)	5.53E-03	8.13E-03
HDL	EUR	5.44E-02 (1.22E-03)	5.28E-02 (2.95E-03)	9.31E-02	1.02E-01
	AMR	6.14E-02 (7.67E-03)	4.81E-02 (3.55E-03)	9.83E-02	1.02E-01
	AFR	2.73E-02 (1.80E-03)	8.86E-03 (4.44E-03)	3.27E-02	3.81E-02
TC	EUR	2.14E-02 (5.89E-04)	2.35E-02 (9.19E-04)	4.08E-02	4.27E-02
	AMR	2.02E-02 (4.24E-03)	2.09E-02 (1.49E-03)	4.46E-02	4.33E-02
	AFR	7.55E-03 (3.54E-03)	5.02E-03 (4.09E-03)	2.34E-02	2.20E-02
RBC	EUR	2.72E-02 (1.18E-03)	2.81E-02 (1.11E-03)	4.75E-02	5.20E-02
	AMR	9.24E-03 (2.68E-03)	1.10E-02 (3.13E-03)	2.11E-02	2.65E-02
	AFR	1.09E-02 (2.86E-03)	6.82E-03 (3.54E-03)	1.43E-02	1.55E-02
Leukocyte	EUR	9.83E-03 (4.16E-04)	1.06E-02 (4.17E-04)	1.55E-02	1.90E-02
	AMR	7.73E-03 (2.61E-03)	6.61E-03 (2.63E-03)	1.22E-02	1.59E-02
	AFR	8.89E-03 (1.89E-03)	1.02E-02 (2.54E-03)	1.36E-02	1.45E-02
Asthma	EUR	1.08E-02 (2.50E-03)	1.04E-02 (6.84E-04)	1.39E-02	1.77E-02
	AMR	4.71E-03 (1.41E-03)	2.55E-03 (1.59E-03)	1.15E-02	1.74E-02
	AFR	6.82E-04 (4.37E-04)	1.39E-03 (5.10E-04)	3.33E-03	3.67E-03
T2D	EUR	2.29E-02 (1.60E-03)	2.05E-02 (1.97E-03)	3.65E-02	4.41E-02
	AMR	8.50E-03 (2.04E-03)	4.88E-03 (9.23E-04)	2.21E-02	2.90E-02
	AFR	3.83E-03 (1.18E-03)	3.74E-03 (1.06E-03)	1.08E-02	1.09E-02
Colorectal Cancer	EUR	2.41E-03 (1.22E-03)	1.58E-03 (1.28E-03)	2.60E-03	1.37E-03
	AMR	1.36E-03 (2.02E-03)	3.81E-04 (3.72E-04)	3.10E-03	1.50E-03
	AFR	1.44E-03 (2.06E-03)	1.08E-03 (1.45E-03)	1.67E-04	4.80E-04
Breast Cancer	EUR	1.17E-02 (1.52E-03)	8.02E-03 (8.83E-04)	1.57E-02	1.30E-02
	AMR	2.93E-03 (1.93E-03)	2.77E-03 (2.07E-03)	1.41E-02	1.31E-02

AFR	1.07E-03 (1.03E-03)	7.33E-04 (7.53E-04)	4.11E-03	3.93E-03

The R^2 columns refer to both R^2 and pseudo-R^2. The values for C+T were averaged from 10 R^2 values. The standard deviations are indicated in parentheses.

Discussion

In this study, we benchmarked the prediction accuracy of PGS generated from array and WGS genotypes across a range of traits and populations. The performance was influenced by several factors, which are discussed below.

Using the C+T method, polygenic prediction generally did not perform better with WGS variants than with array variants, even when the array data were not imputed (**Figure 2**). This can be explained by several factors. First, LD clumping for each trait drastically reduced the number of WGS variants from approximately 9 million to ~467K (~95% reduction), while array variants were clumped from ~1 million to ~262K (~73% reduction) (**Table S3**). Moreover, the overlap of SNPs between discovery GWAS summary statistics and the target genome also restricts the variants included in the PGS calculation. The large reduction in WGS variants may have removed important variants for polygenic prediction, leading to comparable or even lower prediction accuracy than PGS derived from array variants. Second, the field standard of GWAS and PGS methods is designed to model bi-allelic SNPs, which could lead to comparable prediction using array versus WGS. Given that sequencing can capture additional important genetic variations such as INDELs and CNVs that arrays often cannot, incorporating more complex genetic variation could boost PGS performance for WGS and recover some missing heritability. Third, sequencing error rates range from 0.1% to 1% across platforms[32,33]. However, even low error rates can introduce significant errors when large numbers of nucleotides are sequenced[34]. These errors may contribute noise to the PGS, thereby reducing predictive accuracy. Fourth, allele frequencies

311 are lower in WGS-derived clumped variants than in array-derived clumped variants (**Figure S6a**,
312 **b** for HDL). This could differentially impact phenotypes. For example, prediction for HDL, which
313 has a sparse genetic architecture, is significantly affected by lower allele frequencies in WGS
314 variants, while prediction for highly polygenic height benefits from the higher number of variants
315 available in WGS data. Even if the array variants were imputed, however, it is unclear whether
316 this would improve prediction accuracy, given that WGS variants did not show significantly better
317 performance than non-imputed array variants for most traits.

318 On the other hand, WGS-based PGS generally performed better than array-based PGS when
319 using PRS-CS (**Figure 4**). It is possible that if array variants were imputed to the genomic
320 coverage equivalent to WGS, their performance might approach that of WGS-based PGS using
321 this method. The lack of imputation for array genotypes is one limitation of this study due to budget
322 considerations and limitations from leveraging the controlled cloud compute environment of the
323 *All of Us* Research Program. However, given this is the format by which the arrays are centrally
324 released in *All of Us*, this is an accurate capture of the general pipelines deployed in this
325 foundational resource. Further, expanding the variants considered by WGS by computing our own
326 LD matrix (rather than utilizing the provided PRS-CS LD matrix that is restricted to HapMap3
327 SNPs) with 6.2 M more sites did not improve WGS based results compared to the more limited
328 variant set (**Figure S10**). Together, these results suggest that fewer, reliably captured, more
329 common, and well-tagging variants appear to suffice for PGS computation for most traits,
330 particularly those that are less polygenic. This can be illustrated with the theoretical accuracy of
331 PGS, where the expected $R^2 \approx h^2/[1 + M/(h^2 N)]$[35]. With the same sample size N, using WGS
332 over array variants might increase the heritability h^2, but it does not necessarily provide a benefit
333 when a very large number of variants M are involved.

334 As our focus was on benchmarking genotype discovery approaches, in this study we did not
335 investigate the impact of using reference panels with different population compositions on
336 polygenic prediction. It may be optimal to match the population proportions between the reference
337 panel and discovery GWAS[30], or to compute in-sample LD, given that LD patterns differ across
338 populations. Specifically, our GWAS data from Pan-UKBB consists of 95% European samples,
339 however, the proportion of European samples in the 1kGP reference panel used for C+T was only
340 about 20%[18]. It remains to be tested whether prediction accuracy could be improved by reducing
341 the proportion of other populations to 5%, or if maintaining a large reference panel sample size
342 outweighs the advantage of using a proportion-matched reference panel. This approach could
343 potentially provide more accurate LD estimates for our discovery GWAS samples and improve
344 prediction accuracy. This may also partially explain why PRS-CS performed better than C+T, as
345 the reference panel used for PRS-CS was the UK Biobank EUR LD panel, provided by the
346 developers of PRS-CS, from which a high percentage of samples overlap with those in the Pan-
347 UKB discovery data.

348 Traditional SNP arrays have predominantly been based on genomes of European ancestry due
349 to the higher availability of EUR samples in reference databases[36]. In contrast, sequencing does
350 not rely on pre-selected markers and is expected to provide greater benefits for underrepresented
351 groups in genomic studies. Consequently, we anticipated a greater improvement in prediction
352 accuracy using variants from sequencing compared to those from arrays, particularly in non-
353 European populations. However, this trend was not universally observed. We hypothesize that
354 the significantly higher allele frequencies of utilized SNPs stemming from array data as compared
355 to WGS data may explain this result. Additionally, the population-level allele frequencies
356 calculated in *All of Us* were highest in AFR (**Figure S6a**, **b**), suggesting that the Global Diversity
357 Array used in *All of Us* is not biased toward European populations. Ongoing efforts generating

array data for diverse populations would do well to similarly consider capturing variation represented in the target samples.

As has been demonstrated previously[3,37,38], prediction performance was generally better in EUR than in other populations, likely due to the overrepresentation of EUR in the discovery GWAS. Although underrepresented in Pan-UKBB, AMR achieved similar performance to EUR for some traits, while AFR always had low predictive power (**Figure 2** and **Figure 4**). This may be attributed to the larger proportion of European ancestry in the genomes of AMR[39], and the scale of genetic distances from EUR, as has been previously explored[4]. We specifically observed a four-fold reduction in performance for T2D in AFR compared to EUR in this study (**Figure 2** and **Figure 4**), highlighting the ongoing performance disparities of PGS for diverse populations across genotype discovery approaches and PGS methods. Another consideration is that AMR and AFR are admixed populations, with genomes composed of genetic segments from multiple ancestries[40]. Ancestry-informed GWAS methods tailored to admixed populations, such as *Tractor* GWAS[41], may produce summary statistics for polygenic prediction that function better for admixed individuals[30].

Despite the differential advantages of using WGS over arrays for better prediction with PRS-CS, the cost of WGS and arrays remains a key factor in choosing which genotype discovery platform to use for individual researchers and consortia generating their own data. As of 2024, the cost for genotyping arrays ranges from $80 for a screening array to $135 for a global diversity array, while the cost for 30X short-read WGS is about $600. However, the price of WGS has dramatically declined since next-generation sequencing technology came into use in 2007[42]. With the rapid advancement of sequencing technology, WGS is expected to become more affordable and potentially replace arrays as the standard platform in large-scale genomic studies and personal health screening. This development is anticipated to not only accelerate genomic discovery for historically underrepresented populations but also advance PGS method development. The larger

383 and more representative GWAS and reference panels generated from WGS for populations with
384 diverse genetic architectures will enhance prediction accuracy for traits and diseases that have
385 lacked statistical power. This will create many opportunities for genetic prediction models tailored
386 to specific health conditions and populations, facilitating the progress of precision medicine.

387 PGS models vary in performance across training sources such that it would be desirable to
388 compare the performance of array versus WGS genotype using PGS results from other study
389 settings. For example, it would be valuable to compute PGS using the PGS catalog[43] by applying
390 the existing scoring files (with variants, effect alleles, and allelic information) on our target data in
391 *All of Us*, comparing the performance for array versus WGS genotypes. However, the PGC
392 Catalog's software implementation currently only supports imputed array-based genotypes but
393 not WGS data. If the *All of Us* array data is imputed in the future, or if WGS data becomes well-
394 supported by the calculator, we can compare the reproducibility of PGS from our model to existing
395 scores. For future work, it would also be interesting to benchmark PGS models that employ
396 functionally informed fine-mapping methods[44]. Fine-mapping aims to identify causal variants for
397 complex traits[45], and we expect that these methods would be better suited for sequenced
398 genomes than arrays which "tag" the causal effects. This also means that including rarer but
399 functionally mapped variants into PGS calculation as well as GWAS construction using WGS is
400 highly desirable.

401 To conclude, this study evaluates the performance of PGS with variants discovered from arrays
402 and WGS. The advantage of using either genotype discovery approach depends on several
403 factors. Array-PGS exhibit higher predictive accuracy than WGS-PGS for less polygenic traits
404 when variants are clumped. The correlation between the number of variants clumped and
405 prediction performance differs across populations. On the other hand, WGS-PGS outperformed
406 array-PGS for most traits with higher polygenicity when using all variants for prediction, while
407 readjusting effect size estimates based on LD. We hope this study provides researchers with a

408 foundation for selecting an optimal genotype calling platform and polygenic score method,
409 considering diverse populations and traits.

410

411 ## Methods

412 **GWAS summary statistics QC**

413 The multi-ancestry meta-analyzed GWAS summary statistics for six continuous traits—height,
414 DBP, leukocyte, RBC, HDL, and TC—and four binary traits—asthma, breast cancer, colorectal
415 cancer, and T2D—were downloaded from Pan-UKBB. A summary table of the selected
416 phenotypes, including sample sizes by population, is provided in **Table S1**. Each summary
417 statistic was lifted over from hg19 to hg38 to harmonize with the genome build of the target
418 genotypes in *All of Us*. Variants in all summary statistics were filtered to retain only biallelic
419 SNPs. The reference and alternate alleles were swapped when necessary to match the strand
420 orientation in the target dataset, and the effect size signs were reversed accordingly. For
421 quantitative traits, only high-quality samples (denoted with the suffix "hq"), effect sizes, and *P*
422 values were used.

423 **Heritability estimation**

424 Heritability for the 10 Pan-UKBB phenotypes was estimated using LD Score Regression
425 (LDSC)[46]. Given that EUR constitutes 95% of the total samples in the Pan UKBB GWAS meta-
426 analysis, we downloaded 1kGP EUR LD scores as a representative reference from LDSC. To
427 match the SNP IDs to the rsIDs in the EUR LD scores, the GWAS summary statistics were
428 annotated with dbSNP rsIDs v155 using the R package *SNPlocs.Hsapiens.dbSNP155.GRCh37*.
429 The observed-scale heritability was then computed using the formatted summary statistics.

430 ***All of Us* phenotype curation**

431 For the 10 phenotypes with GWAS summary statistics available from pan-UKBB, we curated the
432 corresponding phenotypes in the *All of Us* following the methodology outlined in Tsuo *et al*[47].
433 We curated the quantitative traits based on concept IDs, and binary traits using PheCode map
434 v1.2, which facilitate the mapping of ICD codes into their corresponding phecodes. For
435 quantitative phenotypes, we harmonized their units by adopting the most frequently observed
436 unit as the reference standard. We further removed those individuals with values exceeding 5
437 standard deviations from the mean. Subsequently, we performed inverse-rank normal
438 transformation to ensure normalization.

439 **Target sample selection**

440 Samples with both genotyped array and WGS data were selected. In total, 95,562 individuals
441 had both types of data in *All of Us* V6. Samples were excluded if more than 10% of genotypes
442 were missing. Related samples were removed based on a kinship score > 0.1. Samples
443 predicted to be of EUR (European, composed of Finnish and non-Finnish European), AMR
444 (Latino/admixed American), or AFR (African) ancestry were selected based on an auxiliary
445 genetically predicted ancestry file provided in *All of Us*. Briefly, the ancestry labels were inferred
446 using the prediction probabilities from a random forest classifier, which was trained on the 16-
447 dimenional PCA space of reference HGDP and 1kGP samples and autosomal variants, which
448 obtained from gnomAD[14]. A summary table of sample sizes for the 10 curated phenotypes is
449 provided in **Table S2**.

450 **Target variant QC**

451 The QC criteria were identical for both array and WGS variants in *All of Us* V6. First, only
452 autosomal variants were retained. Variants with a call rate lower than 90% were excluded.
453 Indels were removed, and only common biallelic SNPs with a MAF > 0.01 were retained. QC of
454 the 1,824,517 array variants resulted in 975,876 SNPs, while QC of the 702,574,937 WGS

455 variants resulted in 8,996,707 SNPs. For the analysis of HDL in AFR, we further excluded
456 common SNPs with a MAF < 0.01 in the AFR population to evaluate PGS performance. We did
457 not impute array variants in *All of Us* due to the high computational cost associated with cloud-
458 based analysis. *All of Us* genotype data is released unimputed, such that this is expected to be
459 standard practice for users of this particular dataset. *All of Us* array data is genotyped with the
460 Global Diversity Array, which was designed to integrate highly optimized multiethnic variants
461 with a genome-wide scaffold, incorporating WGS data from over 1,000 individuals of African
462 and/or American ancestry, reflecting the genetic diversity of participants in the *All of Us*
463 Research Program. As such, comparing polygenic prediction performance across ancestral
464 groups using array and WGS data from *All of Us* is particularly relevant.

465 **LD clumping**

466 All 3,202 samples from five continental populations in 1kGP Phase 3 were used as the
467 reference panel for LD calculation in C+T. Samples with first- and second-degree relatives were
468 removed. Reference variants were filtered to retain common biallelic SNPs (MAF > 0.01).
469 Clumping was performed to remove SNPs that were highly correlated based on the LD pattern
470 from the QC'd 1kGP reference panel, retaining only independent significant SNPs. Variants in
471 the ten GWAS summary statistics were clumped using PLINK, with the flags –clump-p1 1, –
472 clump-r2 0.1, and –clump-kb 250. The number of array or WGS variants for each trait after
473 clumping is provided in **Table S3**.

474 **C+T construction**

475 For C+T, variants for the 10 traits were clumped according to the LD in the 1kGP reference
476 panel. PGS for each trait were computed using the following equation:

477 $$PGS_i = \sum_{j=1}^{M} x_{ij}\beta_j$$

478 where X represents the dosage of alternate alleles {0, 1, 2} from the genotype matrix, β denotes
479 the effect size estimates, i indexes the individual, j indexes the variant, and M is the total number
480 of variants used for the calculation (i.e. number of clumped SNPs). The best PGS for each
481 population was determined through 10-fold cross-validation (1 fold for validation and 9 folds for
482 testing) with different sets of variants passing corresponding *p*-value thresholds: 1, 0.5, 1e-1, 1e-
483 2, 1e-3, 1e-4, 1e-5, 1e-6, 1e-7, and 1e-8. The PGS for the best thresholds were then evaluated
484 using R^2 for continuous traits or Nagelkerke's R^2 for binary traits. R^2 is the proportion of variance
485 explained by the full model relative to the null model, calculated using linear regression.
486 Nagelkerke's R^2 was computed via logistic regression by comparing the likelihoods of the null and
487 full models, adjusting for the overall fit to obtain a value indicating the proportion of variance
488 explained by the PGS in the context of the full model. The null model included age, male sex,
489 female sex, and the first 16 principal components as predictors, while the full model included all
490 variables from the null model in addition to PGS.

491 **PRS-CS construction**

492 We prepared the LD reference panel from two sources: 1. A precomputed EUR LD panel
493 constructed using UKBB data, restricted to 1,120,696 HapMap3 variants, which was
494 downloaded from the PRS-CS GitHub repository. 2. A genome-wide EUR LD panel constructed
495 using the 1kGP and HGDP joint call set[48]. To generate this LD panel, we downloaded
496 precomputed EUR LD blocks based on the GRCh38 genome from LDblocks_GRCh38[49]. We
497 then computed the LD matrices for each LD block using PLINK, with a total of 7,330,169 post-
498 QC (bi-allelic SNPs with MAF >0.5%) genome-wide variants from the VCF of 664 unrelated
499 EUR samples in the 1kGP and HGDP joint call set. The LD matrices across 22 chromosomes
500 were saved in HDF5 format, which is compatible with PRS-CS. To run PRS-CS, we did not set
501 the global shrinkage parameter phi and allowed the program to automatically search for the
502 optimal value. After obtaining posterior effect sizes generated from PRS-CS-auto, we computed

503 PGS for each target individual by aggregating the product of their genotypes weighted by the
504 updated effect sizes, as shown in the equation below:

505 $$PGS_i = \sum_{j=1}^{M} x_{ij}\beta_{j,posterior}$$

506 PGS performance was then evaluated using R^2 for continuous traits and Nagelkerke's R^2 for
507 binary traits, as described in the C+T section.

508

509 **Declaration of interests**

510 A.R.M. has received speaker fees from Novartis. Other authors declare no competing interests.

511

512 **Acknowledgements**

513 We gratefully acknowledge *All of Us* participants for their contributions, without whom this
514 research would not have been possible. We also thank the National Institutes of Health's *All of*
515 *Us* Research Program for making available the participant genotype and phenotype data
516 examined in this study. We thank Aishi Ayyanathan for her assistance in generating the QC'd
517 sample list used in this study. This work was supported by NIH grants K01MH121659 and
518 R01HG012869 to E.G.A.

519

520 **Author contributions**

521 Y-S.L. analyzed the data, interpreted the results, and drafted the manuscript. T.T. analyzed the
522 data and interpreted the results. Y.W. curated *All of Us* phenotypes. B.P. and A.R.M. provided

523 critical comments. E.G.A designed and supervised the study and revised the manuscript. All
524 authors approved the final manuscript.

525

526 **Web resources**

527 Pan-UKBB Downloads page, https://pan.ukbb.broadinstitute.org/downloads

528 EUR LD scores, https://zenodo.org/records/8182036

529 1kGP Phase 3 reference panels, https://www.cog-genomics.org/plink/2.0/resources

530 UKBB EUR LD panel for PRS-CS, https://github.com/getian107/PRScs

531

532

533 **Data and code availability**

534 The GWAS summary statistics used in this study were downloaded from the Pan UKBB
535 Downloads page. Individual genotype data were obtained from the *All of Us* Research Program
536 Controlled Tier Dataset V6. EUR LD scores from 1kGP, used for heritability estimation, were
537 downloaded from the zenodo web resources describe above. The 1kGP Phase 3 reference panels
538 used for C+T were downloaded from PLINK resources. The EUR reference LD panel, constructed
539 using UKBB data for PRS-CS, was downloaded from the PRS-CS GitHub repository. All code
540 used for this analysis is available at https://github.com/Atkinson-Lab/PGS_Array_WGS.

541

542

543 **References**

544 1. Konuma, T., and Okada, Y. (2021). Statistical genetics and polygenic risk score for precision
545 medicine. Inflamm. Regen. *41*, 18.

546 2. Choi, S.W., Mak, T.S.-H., and O'Reilly, P.F. (2020). Tutorial: a guide to performing polygenic
547 risk score analyses. Nat. Protoc. *15*, 2759–2772.

548 3. Wang, Y., Tsuo, K., Kanai, M., Neale, B.M., and Martin, A.R. (2022). Challenges and
549 opportunities for developing more generalizable polygenic risk scores. Annu. Rev. Biomed. Data
550 Sci. *5*, 293–320.

551 4. Ding, Y., Hou, K., Xu, Z., Pimplaskar, A., Petter, E., Boulier, K., Privé, F., Vilhjálmsson, B.J.,
552 Olde Loohuis, L.M., and Pasaniuc, B. (2023). Polygenic scoring accuracy varies across the
553 genetic ancestry continuum. Nature *618*, 774–781.

554 5. Fitipaldi, H., and Franks, P.W. (2023). Ethnic, gender and other sociodemographic biases in
555 genome-wide association studies for the most burdensome non-communicable diseases: 2005-
556 2022. Hum. Mol. Genet. *32*, 520–532.

557 6. Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The missing diversity in human genetic
558 studies. Cell *177*, 26–31.

559 7. Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical
560 use of current polygenic risk scores may exacerbate health disparities. Nat. Genet. *51*, 584–
561 591.

562 8. Ikegawa, S. (2012). A short history of the genome-wide association study: where we were
563 and where we are going. Genomics Inform. *10*, 220–225.

564 9. Cross, B., Turner, R., and Pirmohamed, M. (2022). Polygenic risk scores: An overview from
565 bench to bedside for personalised medicine. Front. Genet. *13*, 1000667.

566 10. Yeung, J.M., Sham, P.C., Chan, A.S., and Cherny, S.S. (2008). OpenADAM: an open
567 source genome-wide association data management system for Affymetrix SNP arrays. BMC
568 Genomics *9*, 636.

569 11. Xing, C., Huang, J., Hsu, Y.-H., DeStefano, A.L., Heard-Costa, N.L., Wolf, P.A., Seshadri,
570 S., Kiel, D.P., Cupples, L.A., and Dupuis, J. (2016). Evaluation of power of the Illumina
571 HumanOmni5M-4v1 BeadChip to detect risk variants for human complex diseases. Eur. J. Hum.
572 Genet. *24*, 1029–1034.

573 12. Infinium Global Screening Array-24 Kit | Population-scale genetics.

574 13. Infinium Global Diversity Array-8 Kit | Multiethnic human genotyping.

575 14. All of Us Research Program Genomics Investigators (2024). Genomic data in the All of Us
576 Research Program. Nature *627*, 340–346.

577 15. Infinium Asian Screening Array | Genotyping of East Asian populations.

578 16. content, C. A powerful array for genetic studies focused on African populations.

579 17. H3Africa Consortium, Rotimi, C., Abayomi, A., Abimiku, A., Adabayeri, V.M., Adebamowo,
580 C., Adebiyi, E., Ademola, A.D., Adeyemo, A., Adu, D., et al. (2014). Enabling the genomic
581 revolution in Africa. Science *344*, 1346–1348.

582 18. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P.,
583 Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global
584 reference for human genetic variation. Nature *526*, 68–74.

585 19. Naito, T., and Okada, Y. (2024). Genotype imputation methods for whole and complex
586 genomic regions utilizing deep learning technology. J. Hum. Genet. *69*, 481–486.

587 20. McMahon, A., Lewis, E., Buniello, A., Cerezo, M., Hall, P., Sollis, E., Parkinson, H., Hindorff,
588 L.A., Harris, L.W., and MacArthur, J.A.L. (2021). Sequencing-based genome-wide association
589 studies reporting standards. Cell Genomics *1*,.

590 21. Brlek, P., Bulić, L., Bračić, M., Projić, P., Škaro, V., Shah, N., Shah, P., and Primorac, D.
591 (2024). Implementing whole genome sequencing (WGS) in clinical practice: advantages,
592 challenges, and future perspectives. Cells *13*,.

593 22. Wendl, M.C., and Wilson, R.K. (2009). The theory of discovering rare variants via DNA
594 sequencing. BMC Genomics *10*, 485.

595 23. Wheeler, M.M., Stilp, A.M., Rao, S., Halldórsson, B.V., Beyter, D., Wen, J., Mihkaylova,
596 A.V., McHugh, C.P., Lane, J., Jiang, M.-Z., et al. (2022). Whole genome sequencing identifies
597 structural variants contributing to hematologic traits in the NHLBI TOPMed program. Nat.
598 Commun. *13*, 7592.

599 24. Appadurai, V., Bybjerg-Grauholm, J., Krebs, M.D., Rosengren, A., Buil, A., Ingason, A.,
600 Mors, O., Børglum, A.D., Hougaard, D.M., Nordentoft, M., et al. (2023). Accuracy of haplotype
601 estimation and whole genome imputation affects complex trait analyses in complex biobanks.
602 Commun. Biol. *6*, 101.

603 25. Martin, A.R., Atkinson, E.G., Chapman, S.B., Stevenson, A., Stroud, R.E., Abebe, T.,
604 Akena, D., Alemayehu, M., Ashaba, F.K., Atwoli, L., et al. (2021). Low-coverage sequencing
605 cost-effectively detects known and novel variation in underrepresented populations. Am. J.
606 Hum. Genet. *108*, 656–668.

607 26. Chat, V., Ferguson, R., Morales, L., and Kirchhoff, T. (2021). Ultra Low-Coverage Whole-
608 Genome Sequencing as an Alternative to Genotyping Arrays in Genome-Wide Association
609 Studies. Front. Genet. *12*, 790445.

610 27. Petter, E., Ding, Y., Hou, K., Bhattacharya, A., Gusev, A., Zaitlen, N., and Pasaniuc, B.
611 (2023). Genotype error due to low-coverage sequencing induces uncertainty in polygenic
612 scoring. Am. J. Hum. Genet. *110*, 1319–1329.

613 28. Kim, S., Shin, J.-Y., Kwon, N.-J., Kim, C.-U., Kim, C., Lee, C.S., and Seo, J.-S. (2021).
614 Evaluation of low-pass genome sequencing in polygenic risk score calculation for Parkinson's
615 disease. Hum Genomics *15*, 58.

616 29. Karczewski, K.J., Gupta, R., Kanai, M., Lu, W., Tsuo, K., Wang, Y., Walters, R.K., Turley,
617 P., Callier, S., Baya, N., et al. (2024). Pan-UK Biobank GWAS improves discovery, analysis of
618 genetic architecture, and resolution into ancestry-enriched effects. MedRxiv.

619 30. Wang, Y., Kanai, M., Tan, T., Kamariza, M., Tsuo, K., Yuan, K., Zhou, W., Okada, Y.,
620 BioBank Japan Project, Huang, H., et al. (2023). Polygenic prediction across populations is
621 influenced by ancestry, genetic architecture, and methodology. Cell Genomics *3*, 100408.

622 31. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A., and Smoller, J.W. (2019). Polygenic prediction via
623 Bayesian regression and continuous shrinkage priors. Nat. Commun. *10*, 1776.

624 32. Stoler, N., and Nekrutenko, A. (2021). Sequencing error profiles of Illumina sequencing
625 instruments. NAR Genom. Bioinform. *3*, lqab019.

626 33. Wall, J.D., Tang, L.F., Zerbe, B., Kvale, M.N., Kwok, P.-Y., Schaefer, C., and Risch, N.
627 (2014). Estimating genotype error rates from high-coverage next-generation sequence data.
628 Genome Res. *24*, 1734–1739.

629 34. Sleep, J.A., Schreiber, A.W., and Baumann, U. (2013). Sequencing error correction without
630 a reference genome. BMC Bioinformatics *14*, 367.

631 35. Visscher, P.M., Yang, J., and Goddard, M.E. (2010). A commentary on “common SNPs
632 explain a large proportion of the heritability for human height” by Yang et al. (2010). Twin Res.
633 Hum. Genet. *13*, 517–524.

634 36. Fatumo, S., Chikowore, T., Choudhury, A., Ayub, M., Martin, A.R., and Kuchenbaecker, K.
635 (2022). A roadmap to increase diversity in genomic studies. Nat. Med. *28*, 243–250.

636 37. Kachuri, L., Chatterjee, N., Hirbo, J., Schaid, D.J., Martin, I., Kullo, I.J., Kenny, E.E.,
637 Pasaniuc, B., Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium Methods
638 Working Group, Witte, J.S., et al. (2024). Principles and methods for transferring polygenic risk
639 scores across global populations. Nat. Rev. Genet. *25*, 8–25.

640 38. Kullo, I.J., Lewis, C.M., Inouye, M., Martin, A.R., Ripatti, S., and Chatterjee, N. (2022).
641 Polygenic scores in biomedical research. Nat. Rev. Genet. *23*, 524–532.

642 39. Montinaro, F., Busby, G.B.J., Pascali, V.L., Myers, S., Hellenthal, G., and Capelli, C. (2015).
643 Unravelling the hidden ancestry of American admixed populations. Nat. Commun. *6*, 6596.

644 40. Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D., and Mountain, J.L. (2015). The genetic
645 ancestry of African Americans, Latinos, and European Americans across the United States. Am.
646 J. Hum. Genet. *96*, 37–53.

647 41. Atkinson, E.G., Maihofer, A.X., Kanai, M., Martin, A.R., Karczewski, K.J., Santoro, M.L.,
648 Ulirsch, J.C., Kamatani, Y., Okada, Y., Finucane, H.K., et al. (2021). Tractor uses local ancestry
649 to enable the inclusion of admixed individuals in GWAS and to boost power. Nat. Genet. *53*,
650 195–204.

651 42. Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat. Rev. Genet. *11*,
652 31–46.

653 43. Lambert, S.A., Wingfield, B., Gibson, J.T., Gil, L., Ramachandran, S., Yvon, F., Saverimuttu,
654 S., Tinsley, E., Lewis, E., Ritchie, S.C., et al. (2024). Enhancing the Polygenic Score Catalog
655 with tools for score calculation and ancestry normalization. Nat. Genet. *56*, 1989–1994.

656 44. Weissbrod, O., Kanai, M., Shi, H., Gazal, S., Peyrot, W.J., Khera, A.V., Okada, Y., Biobank
657 Japan Project, Martin, A.R., Finucane, H.K., et al. (2022). Leveraging fine-mapping and
658 multipopulation training data to improve cross-population polygenic risk scores. Nat. Genet. *54*,
659 450–458.

660 45. Pasaniuc, B., and Price, A.L. (2017). Dissecting the genetics of complex traits using
661 summary association statistics. Nat. Rev. Genet. *18*, 117–127.

662 46. Bulik-Sullivan, B., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working
663 Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L., and
664 Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in
665 genome-wide association studies. Nat. Genet. *47*, 291–295.

666 47. Tsuo, K., Shi, Z., Ge, T., Mandla, R., Hou, K., Ding, Y., Pasaniuc, B., Wang, Y., and Martin,
667 A.R. (2024). All of Us diversity and scale improve polygenic prediction contextually with greatest
668 improvements for under-represented populations. BioRxiv.

669 48. Koenig, Z., Yohannes, M.T., Nkambule, L.L., Zhao, X., Goodrich, J.K., Kim, H.A., Wilson,
670 M.W., Tiao, G., Hao, S.P., Sahakian, N., et al. (2024). A harmonized public resource of deeply
671 sequenced diverse human genomes. Genome Res. *34*, 796–809.

672 49. MacDonald, J., Harrison, T., Bammler, T., Mancuso, N., and Lindstroem, S. (2022). An updated
673 map of GRCh38 linkage disequilibrium blocks based on European ancestry data. BioRxiv.