

## 1 Title

2 Comparing ancestry calibration approaches for a trans-ancestry colorectal cancer polygenic risk score

## 3 Author List

4 Elisabeth A. Rosenthal<sup>1\*</sup>, Li Hsu<sup>2</sup>, Minta Thomas<sup>2</sup>, Ulrike Peters<sup>2</sup>, Christopher Kachulis<sup>3</sup>, Karynne

5 Patterson<sup>4</sup>, Gail P. Jarvik<sup>1,4</sup>

6 \*Correspondence: [erosen@uw.edu](mailto:erosen@uw.edu)

7 Affiliations:

8 1. Division Medical Genetics, School of Medicine, University of Washington, Seattle, WA

9 2. Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA

10 3. Broad Institute of Harvard and MIT, Cambridge, MA

11 4. Department of Genome Sciences, University of Washington, Seattle, WA

12

## 13 Abstract

14 **Background:** Colorectal cancer (CRC) is a complex disease with monogenic, polygenic and environmental  
15 risk factors. Polygenic risk scores (PRS) are being developed to identify high polygenic risk individuals.

16 Due to differences in genetic background, PRS distributions vary by ancestry, necessitating calibration.

17 **Methods:** We compared four calibration methods using the All of Us Research Program Whole Genome  
18 Sequence data for a CRC PRS previously developed in participants of European and East Asian ancestry.

19 The methods contrasted results from linear models with A) the entire data set or an ancestrally diverse

1 training set AND B) covariates including principal components of ancestry or admixture. Calibration with  
2 the training set adjusted the variance in addition to the mean.

3 **Results:** All methods performed similarly within ancestry with OR (95% C.I.) per s.d. change in PRS:  
4 African 1.5 (1.02, 2.08), Admixed American 2.2 (1.27, 3.85), European 1.6 (1.43, 1.89), and Middle  
5 Eastern 1.1 (0.71, 1.63). Using admixture and an ancestrally diverse training set provided distributions  
6 closest to standard Normal with accurate upper tail frequencies.

7 **Conclusion:** Although the PRS is predictive of CRC risk for most ancestries, its performance varies by  
8 ancestry. *Post-hoc* calibration preserves the risk prediction within ancestries. Training a calibration  
9 model on ancestrally diverse participants to adjust both the mean and variance of the PRS, using  
10 admixture as covariates, created standard Normal z-scores. These z-scores can be used to identify  
11 patients at high polygenic risk, and can be incorporated into comprehensive risk scores including other  
12 known risk factors, allowing for more precise risk estimates.

13

## 1 Introduction

2 Colorectal cancer (CRC) is the third most common cancer in the United States (U.S.), and is associated  
3 with the second most deaths from cancer<sup>1</sup>. CRC risk may be attributable to social determinants of  
4 health such as socioeconomic status, health care access, and food security, which are known to be  
5 associated with race<sup>2</sup>. Individuals who self-report as Black or Native American/Alaskan Native have the  
6 highest lifetime risk for CRC as well as higher mortality rates<sup>1,3,4</sup>. Additionally, risk for CRC is affected by  
7 other environmental factors such as diet, smoking and physical activity, as well as genetics<sup>5</sup>. Although  
8 25% of CRC appear to be familial, only 20% of familial CRC is explained by variation at single genes<sup>6-8</sup>.

9 In aggregate, low risk SNPs across the genome also contribute to risk of CRC. Polygenic risk scores (PRS),  
10 which aggregate the effect of genetic variants across the genome, are currently being developed to aid  
11 in identifying individuals at higher genetic risk of developing CRC<sup>9,10</sup>. As these PRS have been shown to  
12 be independent of family history, they may provide orthogonal information when incorporated into a  
13 comprehensive score<sup>11</sup>. Additionally, PRS have been shown to be more predictive at younger ages,  
14 making them useful for identifying patients who can benefit from increased or earlier screening, as  
15 current environmental risk prediction models are optimized for middle-aged patients<sup>12,13</sup>.

16 Development of PRS involves estimation of SNP effects, which are influenced by the genetic ancestry of  
17 the cohorts used to develop the PRS through linkage disequilibrium (LD) across the genome and minor  
18 allele frequency (MAF) at the SNPs<sup>14</sup>. The PRS for CRC used here (PGS catalog PGS003852)<sup>15</sup>, was  
19 developed with participants of European and East Asian ancestry, using Bayesian methods that estimate  
20 SNP effect sizes accounting for LD in both populations, while simultaneously estimating the number of  
21 causal SNPs given the heritability of CRC<sup>9,16</sup>. When participants from multiple ancestries are included,  
22 this method helps delineate the collinearity of SNPs in LD and therefore improves identification of SNPs

1 that are associated with CRC risk in multiple ancestries, reporting a single estimated effect for each SNP  
2 <sup>17,18</sup>.

3 Even when more than one genetic ancestry is used to develop a multi-ancestry PRS, the distribution of  
4 the PRS will differ by genetic ancestry, as LD and MAF vary across genetic ancestries. Therefore, an  
5 individual's PRS must be interpreted within the context of their unique genetic ancestry in order to  
6 estimate their PRS-specific predicted genetic risk. However, genetic ancestry is not strictly categorical,  
7 due to admixture, and can not be adequately determined by observable phenotypes in the clinic, such  
8 as skin color, hair type, or eye shape <sup>19,20</sup>. Several *post-hoc* mechanisms for calibrating the PRS for  
9 genetic ancestry in the clinic setting have been proposed <sup>21,22</sup>. These include using linear models, based  
10 on a training data set of diverse populations, to adjust the mean PRS using principal components (PCs)  
11 of ancestry. Additionally, methods using training models have been adapted to adjust for both the  
12 expected mean and variance of the PRS. One benefit of these approaches is that patients do not need to  
13 be categorized within a genetic ancestry. Furthermore, these approaches can be applied to admixed  
14 individuals, who make up an increasing proportion of U.S. residents.

15 We set out to determine which, if any, of these *post-hoc* PRS calibration methods is preferable in the  
16 context of this multi-ancestry CRC PRS, by analyzing data from the All of Us Research Program (AOU) <sup>23,24</sup>.  
17 The purpose of AOU is to collect survey, electronic health record (EHR), and genotype data on diverse  
18 participants who live in the U.S., for use in broad research and to help reduce inequities in healthcare in  
19 the U.S.. As this data set contains both ancestrally diverse and admixed participants, it is an ideal  
20 biobank to assess *post-hoc* genetic ancestry calibration methods for use in the clinic.

## 1 Subjects and Methods

### 2 Case/Control assignment

3 We updated a previously developed CRC case/control algorithm (see online materials) to define CRC  
4 cases and controls. This algorithm was created using International Classification of Disease (ICD) 9 codes  
5 and Current Procedural Terminology (CPT) codes. The AOU dataset contains a rich database of ICD 9 and  
6 10 codes, CPT codes, and Logical Observation Identifiers Names and Codes (LOINC). As the completeness  
7 of this data varies among participants<sup>25</sup> we attempted to widen our net by creating concept sets based  
8 on the original ICD 9 and CPT codes, incorporating the related ICD 10 codes and LOINC codes. Each  
9 concept set relates to a different table from the word files provided by the algorithm's authors  
10 (Supplemental Tables S2-S10). The algorithm was developed to screen for CRC cases with potentially  
11 monogenic causes of CRC, and therefore excluded individuals with ulcerative colitis (UC) or Crohn's  
12 disease, two disorders associated with CRC. As we are interested in a polygenic component underlying  
13 CRC we did not exclude participants with UC or Crohn's disease. Additionally, we excluded all  
14 participants with a known monogenic (pathogenic or likely pathogenic variant in ClinVar) cause of CRC in  
15 any of *AKT1*, *APC*, *AXIN2*, *BMPR1A*, *CDH1*, *CHEK2*, *CTNNA1*, *EPCAM*, *GALNT12*, *GREM1*, *MLH1*, *MSH2*,  
16 *MSH3*, *MSH6*, *MUTYH*, *NTHL1*, *PDGFRA*, *PIK3CA*, *PMS2*, *POLD1*, *POLE*, *PTEN*, *RPS20*, *SMAD4*, *STK11*, or  
17 *TP53*<sup>24</sup> (*in press*).

18 Potential cases included all participants that had at least one CRC diagnosis in their medical record  
19 (Supplemental table S2) but did not have a known pathogenic variant for CRC or a diagnosis for  
20 monogenic disorders with increased risk of CRC (Supplemental table S3). The cleaned set of cases was  
21 derived from potential cases using the following ordered algorithm:

22 A. They had a surgical procedure related to CRC within a year of diagnosis (Supplemental Table S4)

1 B. If not (A), then they had chemotherapy or radiation (Supplemental Tables S5 and S6) within a  
2 year of CRC diagnosis and they did not have other types of cancers listed in the exclusion table  
3 (Supplemental Table S7)

4 C. If not (A) or (B) and they had at least two CRC diagnosis codes within 2 years of each other and  
5 no other cancers listed in the exclusion table

6 Potential controls were participants with no diagnosis codes for CRC or a monogenic disorder with  
7 increased risk of CRC, and no evidence of a pathogenic variant for CRC. Using guidance from the  
8 algorithm, the screened controls met either of the following criteria:

9 A. They had at least one sigmoidoscopy or colonoscopy and no positive pathology report  
10 (Supplemental Table S8)

11 B. They had at least two instances of a fecal immunochemical test (FIT) or a fecal occult blood test  
12 (FOBT) that were at least 5 years apart and never had a positive lab result from any of these  
13 tests (Supplemental Tables S9 and S10)

14 All other participants that were not assigned case/control status, or who were not excluded by the  
15 algorithm, were left unassigned. Unassigned participants are included in the analysis except when  
16 statistical testing involves case and control status. Some analyses include age and sex. Age is defined as  
17 observational age: age at onset of cases, age at last screening for controls, and age at consent to AOU  
18 for the unassigned participants. In these analyses we further restricted analysis to participants who were  
19 older than the minimum observed age of onset of CRC, 19, as CRC rarely presents so early. Sex is defined  
20 as sex assigned at birth.

## 21 Genetic Ancestry Clustering

22 We calculated 32 genome-wide PCs of ancestry using an unrelated subset of reference global genomes  
23 from 1000 Genomes (1KG) and Human Genome Diversity Project (HGDP) (N = 4151)<sup>26,27</sup>. We then

1 projected the AOU participants onto the PC space. We used EIGMIX to estimate admixture in the AOU  
2 participants for the continental ancestry groups represented by the reference genomes: African (AFR),  
3 Admixed American/Latino (AMR), East Asian (EAS), European (EUR), Middle Eastern (MID) and South  
4 Asian (SAS)<sup>28</sup>. See supplemental methods for details on PC and admixture calculation. We then clustered  
5 the AOU participants into ancestry groups, using the estimated percent ancestries, in order to compare  
6 the performance of the PRS across continental ancestry groups. As the amount of genetic diversity  
7 among different racial and ethnic groups in the United States differs (Supplemental Figures S1 and S2)<sup>28-</sup>  
8 <sup>30</sup>, we used an 80% cutoff to cluster individuals into EAS, EUR and SAS clusters, and a 60% cutoff to  
9 cluster individuals into AFR, AMR and MID clusters. All other participants are referred to as Other (OTH).  
10 A subset of highly admixed participants from OTH, called ADM, contains participants with <50%  
11 estimated ancestry for all continental ancestries considered.

## 12 PRS calculation

13 Details of PRS calculation in AOU and the reference 1KG and HGDP data are given in the supplemental  
14 methods. In brief, we extracted genotype data from WGS using HAIL<sup>31</sup> on the Spark cluster. After  
15 performing quality control, we calculated the sum of genotype effects for each participant using the R  
16 package bigSNPR v. 1.10.8 . The allele effects for each SNP were obtained from the corresponding  
17 author of PMID: 36789420<sup>9</sup> and can also be found in the PGS catalog (PGS003852).

## 18 PRS Calibration

19 We started with three *post-hoc* calibration methods of the raw PRS (RAW) to account for ancestry:  
20 linear model to adjust for the first 5 PCs of ancestry in AOU (PC\_μ), linear model to adjust for the  
21 estimated admixture percentages in AOU (AD\_μ), and a previously published method which adjusts both  
22 mean and standard deviation (s.d.) as a function of PCs using a trained model on a reference dataset  
23 (PC.REF\_μσ)<sup>21,22</sup>. This adjustment estimates the effects of the first 5 PCs of ancestry in the 1KG and

1 HGDP reference genomes on the mean PRS and its variance. Given these estimated effects, the raw PRS  
2 from AOU can be adjusted by both the expected PRS and its expected variance for each participant given  
3 their first 5 PCs of ancestry. We chose to use 5 PCs of ancestry to make comparisons fair with AD\_μ  
4 which has 5 degrees of freedom due to the six global ancestries considered in this analysis. We added in  
5 two more adjustments after observing the distribution of PC.REF\_μσ. We modified the method by using  
6 a random subset (N=4,151) of the unassigned participants from AOU, called AOU.REF, to derive the  
7 prediction model (PC.AOU\_μσ). Additionally, we modified the method, using the same random sample,  
8 AOU.REF, and used the admixture estimates to create the prediction model rather than the first 5 PCs of  
9 ancestry (AD.AOU\_μσ). The properties of the calibration methods are given in Table 3.

10 We kept the patient reference sample size for AOU.REF the same as the global reference sample size  
11 (1KG and HGDP) so that we could make fair comparisons. As the AOU dataset is heavily weighted toward  
12 individuals of EUR ancestry, a simple random sample would also be heavily weighted in this way.

13 Therefore, we attempted to randomly sample from the space of the first 5 PCs. First, we calculated the  
14 geometric median of the first five PCs. Second, we calculated the Euclidean distance to this median for  
15 each participant. We divided the distances into five equal ranges. We then randomly sampled an equal  
16 number of participants from each range, resulting in the following counts of unassigned participants  
17 from each ancestry cluster in the reference patient sample: 1108 AFR; 1459 AMR; 84 EAS; 820 EUR; 72  
18 MID; 577 OTH; 31 SAS.

## 19 Comparing raw and calibrated PRSs

20 We compared the raw PRS and the adjusted PRSs in several ways, for each ancestry cluster. First, we  
21 compared the overall mean and s.d., as this informs how well the adjusted score aligns with z-scores  
22 from a standard Normal distribution. Second, we compared the odds ratio (OR) for a single s.d. change  
23 in the score, with and without adjusting for age and sex, to determine if the PRS is associated with CRC



1 risk in each genetic ancestry cluster. Third, we compared the log(OR) for each quintile compared to the  
2 middle quintile to determine if those at high risk (top quintile) can be differentiated from those at  
3 typical risk (middle quintile) as suggested in <sup>32</sup>. Fourth, we compared the area under the receiver  
4 operating curve (AUC) which compares how well the scores can distinguish between cases and controls,  
5 overall. We used 500 bootstrap iterations, with replacement, to estimate the 95% confidence intervals  
6 (C.I.) for the AUC. We calculated both the unadjusted AUC and the AUC adjusted for age and sex. Finally,  
7 we compared the observed and expected upper 5, 7.5 and 10th percentiles, as these values could be  
8 used to identify patients at higher genetic risk for developing disease, relative to those of similar genetic  
9 ancestry. These patients could then be referred for increased or earlier screening for CRC. Observed  
10 percentiles were defined as the proportion of participants whose adjusted PRS was greater than the  
11 standard normal cutoffs ( $z=1.28$ ,  $1.44$ , or  $1.64$ ). In this case, RAW, PC $_{\mu}$ , and AD $_{\mu}$  were standardized  
12 using the variance from the entire data set. The C.I. for the percentiles was calculated using the R  
13 function `binom.test()`.

## 14 Results

### 15 Case/Control Demographics and Ancestry

16 We excluded 8 CRC affected participants who had penetrant Lynch syndrome or Familial Adenomatous  
17 Polyposis reported in their EHR and 455 participants with a known pathogenic or likely pathogenic  
18 variant for these monogenic disorders <sup>24</sup> (*in press*). There were 668 participants remaining with at least  
19 one CRC code in the EHR, 348 of whom were classified as cases by the algorithm (Table 1, Supplemental  
20 Tables S2 and S3). There were 97,588 participants who had no evidence for CRC. Of these 12,378 had  
21 been screened for CRC with negative results and were assigned as controls. Using the ancestry clustering  
22 rule, the majority of participants clustered with EUR ancestry (45%), followed by AFR (23%), OTH (16%),

1 AMR (9.5%), and MID (4.1)% with EAS and SAS making up < 3% of the sample (Table 1). As the sample  
2 size for EAS and SAS was small, we do not report statistical tests for these ancestries.

3 The mean observed age of unassigned participants was about 10 years younger than that of the cases,  
4 overall (Table 2). The mean censored age of screened controls was similar to that of the cases,  
5 consistent with recommendations for screening in older patients. The observed age for both the cases  
6 and screened controls tended to be younger than their age at consent, indicating that most cases had  
7 CRC before entering the study and most screened controls were screened before entering the study.

8 Overall, males made up 49% of the cases and 40% of the controls. This sex difference may reflect sex  
9 differences in seeking healthcare and/or enrolling in genetic research studies. For AMR, EUR and OTH,  
10 the sex ratio difference was  $\leq 10$  percentage points. The sex ratio differences were  $> 20$  percentage  
11 points for AFR and MID. We do not report exact counts in order to comply with AOU publication rules  
12 for cell counts less than 20.

### 13 PRS Distribution Comparison

14 The distributions of RAW differed by ancestry in both their mean and s.d. (Table\_4, Figure 1). The mean  
15 ranged between 0.05 (MID) and 0.4 (AFR) and the standard deviation (s.d.) ranged between 0.41 (AFR)  
16 and 0.47 (most ancestries). The PC\_μ means were close to zero for all ancestries except MID and OTH.  
17 AD\_μ resulted in means closer to zero, even for MID and OTH. It should be noted that when RAW, PC\_μ  
18 and AD\_μ are scaled to have a s.d. of 1, the scores are scaled (divided by the s.d.), so that the resulting  
19 scaled means are further from zero as the original s.d. is less than one. PC.REF\_μσ resulted in non-zero  
20 means for EAS, EUR, MID and SAS, prompting us to develop PC.AOU\_μσ and AD.AOU\_μσ. PC.AOU\_μσ  
21 performed similarly to PC\_μ and the scaling of the variance results in means further from zero, as  
22 expected. AD.AOU\_μσ performed similarly to AD\_μ. Notably, both AD\_μ and AD.AOU\_μσ resulted in  
23 standard Normal distributions for the highly admixed ADM group. However, AD.AOU\_μσ had a non-

1 zero mean for SAS, likely due to the small number of SAS participants in AOU.REF. Given that we are  
2 unable to perform statistical tests in the SAS cluster due to low sample size, we chose to continue with  
3 these calibrations and make comparisons between RAW, PC\_μ, AD\_μ, PC.AOU\_μσ, and AD.AOU\_μσ.

#### 4 PRS OR

5 Results for the OR per 1 s.d. change in the PRS were similar across calibration methods, within genetic  
6 ancestries (Table 5). The OR estimate was highest for AMR (range 2.1-2.2), was statistically significant,  
7 and had wide confidence intervals. For EUR, the estimated OR ranged between 1.6 and 1.7, was  
8 statistically significant, and had narrow confidence intervals reflecting the large sample size. For OTH,  
9 the estimated OR ranged between 1.5 and 1.6 and was statistically significant. The estimated OR for AFR  
10 ranged between 1.4 and 1.5, and was not significant after adjusting for age and sex. As this lack of  
11 significance could be due to a sex imbalance among the cases and controls, we performed sex stratified  
12 analysis in AFR. Sex stratified analysis in AFR resulted in similar estimated ORs and they were not  
13 statistically significant (data not shown). The estimated OR for MID ranged between 1.05 and 1.12, and  
14 was not statistically significant in any model. Although the ORs indicate that an increase in PRS is  
15 associated with increased risk of CRC, statistical significance for the uppermost quintile compared to the  
16 middle is observed only in EUR, due to sample size (Figures 2-6). For AMR, there were no cases in the  
17 lowest quintile.

#### 18 PRS AUC

19 The estimated AUC and its 95% C.I. are given in Table 6. Overall, these results are similar to those for the  
20 OR, as expected<sup>33</sup>. The estimated AUC is greater than 0.5 for all ancestries and all adjustments.  
21 However, the AUC is not statistically significantly different from 0.5 for both AFR for some calibrations  
22 and when adjusting for age and sex, as well as for MID in all situations. The AUC is highest for AMR  
23 (0.68-0.72), with wide C.I.s. The AUC is similar for EUR and OTH (~0.64). EUR has the narrowest C.I.

1 reflecting the larger sample size. Adjusting the AUC for age and sex does not change the results for AMR,  
2 EUR or OTH.

### 3 Observed Upper Percentiles

4 We compared the observed upper 5, 7.5 and 10<sup>th</sup> percentiles to their expected values for each PRS.  
5 (Figure 7). The 95% C.I.s for RAW are below expected risks for AFR, AMR, MID, and OTH, and above  
6 expected risks for EUR. The 95% C.I.s for PC\_μ are below the estimated risks for AFR, AMR, MID, and  
7 OTH and above the estimated risks for EUR. The 95% C.I.s for PC.AOU\_μσ covers the expected risks for  
8 AFR, covers the expected risk for AMR at the 5th percentile only, are below the expected risks for AMR  
9 (7.5, 10th percentiles), MID, and OTH, and above the estimated risks for EUR. The 95% C.I.s for AD\_μ  
10 cover the expected risks for AMR and OTH, are below the expected risks for AFR and above the expected  
11 risks for EUR. The 95% C.I.s for AD.AOU\_μσ contains the expected risks for AMR, EUR, and MID, and are  
12 just below the expected risks for AFR and OTH. In the highly admixed group, ADM, the 95% C.I.s include  
13 the expected risks for only AD\_μ and AD.AOU\_μσ. We do not comment on the risk assessment for EAS  
14 and SAS due to the large C.I. in these clusters, reflecting their smaller sample sizes.

### 15 Discussion

16 Regardless of the ancestry calibration, the PRS was statistically significantly associated with CRC status  
17 for EUR, AMR, and OTH, was borderline significant for AFR, and was not significant for MID. Additionally,  
18 the strength of the significant effects differed across ancestries. The estimated OR and AUC were  
19 highest for AMR, possibly due to the lack of cases in the lowest quintile of the PRS. In addition, the  
20 confidence interval for AMR was large, as the total sample size was small. The OR estimate was higher in  
21 EUR than in OTH, but they had similar AUC. The sex imbalance between cases and controls in AFR may  
22 contribute to the borderline significance observed in that cluster, but sex stratified analysis did not have  
23 enough power to confirm this, due to low sample size.

1 Power was sufficient in only the large EUR cluster to show that participants with PRS in the upper  
2 quintile were at higher risk of CRC than those at typical risk (the middle quintile). Observed differences  
3 in OR and their large C.I. comparing quintiles in other ancestries is due to the low number of cases  
4 within each quintile, and the overall smaller sample sizes within each quintile. Specifically, We did not  
5 observe an OR for the lowest quintile in AMR, as there were no cases with a PRS in that quintile.

6 Although the predictiveness of the association between the PRS and risk is not affected by genetic  
7 ancestry calibration, within ancestral groups, the resulting distributions are affected, which can  
8 influence downstream clinical decisions or overall risk estimation when incorporating environmental  
9 and lifestyle factors. Adjusting for the first 5 PCs of ancestry resulted in non-zero means for some  
10 ancestries, whereas adjusting for admixture estimates resulted in means close to zero for all ancestries  
11 (except SAS). The difference from zero is exacerbated when scaling the scores to have a s.d. of 1, as the  
12 variance of the raw PRS is less than one. This happens because the scores are divided by the s.d. (which  
13 is less than one), resulting in an inflated mean distance from zero. The AD.AOU\_μσ calibration method  
14 resulted in a standard Normal distribution for the highly admixed cohort, and was close to standard  
15 Normal for all other ancestries (except SAS). Finally, the upper tail frequencies were closely or accurately  
16 estimated by AD.AOU\_μσ only.

17 There are several limitations to this study. The sample sizes for EAS and SAS are too small to make any  
18 inferences about those genetic clusters. Furthermore, the availability of EHR data, and therefore the  
19 number of cases and controls, may be associated with sex and self-identified race or ancestry<sup>25</sup>. This  
20 may affect the power to detect associations in the AFR genetic cluster as well as when comparing the  
21 upper quintiles to the middle quintile for AFR, AMR, and OTH. Additionally, the estimated effect sizes  
22 observed in this study may be an underestimate as cases in AOU data may represent a healthier cohort  
23 than typical CRC cases due to survival bias. This is evidenced by the fact that the diagnosis of CRC  
24 occurred before entry into the study by a mean of 5 years with a maximum of 27 years. This

1 phenomenon of biobank participants being healthier than the population they represent has been  
2 observed previously<sup>34,35</sup>.

3 Another limitation is with the PRS, itself. Although the PRS is multi-ancestry, it was developed using  
4 participants from only two global populations: EUR and EAS. It has been shown that the applicability of  
5 PRS declines with further genetic ancestry distance from the population used to develop a PRS<sup>36</sup>, which  
6 we observe here. Continued development of the PRS, incorporating samples from other ancestral  
7 populations should resolve the differences in predicted accuracy by genetic ancestry. As we used whole  
8 genome sequence data, observed differences in OR across ancestries cannot be attributed to  
9 uncertainty in imputation of genotypes, but rather are due to differences in MAF, LD with causal  
10 variation, and possible effect heterogeneity across ancestries. Interestingly, we observe the highest OR  
11 and AUC for AMR, which appears to cluster between EUR and EAS in PC1xPC2 space. However, this  
12 phenomenon has yet to be analyzed for PRS that are developed in more than one population, as in this  
13 study.

14 Some methodological choices also limit the study. The global populations used to calculate the PCs of  
15 ancestry may be too general and not comprehensive. For example, we defined AFR ancestry as those  
16 that cluster with reference samples from East and West Africa, the Southwest U.S. and the Caribbean,  
17 which encompasses a large swath of diversity. Similarly, the reference samples that made up the MID  
18 cluster were limited to a small region that included samples from Druze, Moabite, Palestinian and  
19 Bedouin populations. We tried to accommodate these differences by allowing for more admixture  
20 among the AFR, AMR and MID clusters. Furthermore, the global ancestry references did not contain  
21 Pacific Islander or Native American samples, so we are unable to analyze the applicability of the PRS or  
22 the *post-hoc* calibration in these populations.

1 Another methodological choice was the size of the reference patient sample. We chose to limit the  
2 sample size to be the same as that in the global reference sample. However, a larger sample size for the  
3 patient reference may have resulted in improved training models for PC.AOU\_μσ and AD.AOU\_μσ for  
4 both the mean and variance corrections. Additionally, the sampling method could be further improved  
5 to ensure that the PC space is adequately sampled and the resulting patient reference sample  
6 represents the patient population of interest. Finally, the availability of millions of SNP genotypes for  
7 both PC projection and PRS calculation might be a barrier. However, the increased availability of WGS  
8 and genotype imputation based on widely available SNP chips should relieve this issue.

9 A comprehensive score which combines a calibrated multi-ancestry PRS with social determinants of  
10 health, and other known risk factors, is the ultimate goal for use in the clinic. We acknowledge that  
11 further development of the multi-ancestry PRS is necessary for equitable risk prediction across genetic  
12 ancestry. However, an improved raw PRS will still be associated with genetic ancestry due to LD and  
13 MAF differences across ancestries. We believe that a *post-hoc* genetic ancestry calibration based on the  
14 AD.AOU\_μσ method will be most effective as it generates distributions closest to standard Normal for all  
15 genetic backgrounds, can be applied to any patient, and does not require the clinician to determine the  
16 patient's ancestry.

17

## 18 Declaration of Interests

19 The authors declare no competing interests.

## 1 Acknowledgments

2 The authors would like to thank Alyna T. Khan for her valuable comments and suggestions. This work  
3 was funded by the Office of the Director at the National Institute of Health, under award notice  
4 1OT2OD002748-01 and by the NHGRI through the grant U01HG008657.

## 5 Web Resources

6 Colorectal Cancer case/control algorithm: [phekb.org/phenotype/colorectal-cancer-crc](http://phekb.org/phenotype/colorectal-cancer-crc)

7 CRC PRS: <https://www.pgscatalog.org/score/PGS003852/>

8 1000 Genomes Data: <https://www.internationalgenome.org/home>

9 <https://www.internationalgenome.org/data-portal/search?q=1000%2Bgenomes>

10 Human Genome Diversity Project: [https://www.internationalgenome.org/data-portal/data-](https://www.internationalgenome.org/data-portal/data-collection/hgdp)  
11 [collection/hgdp](https://www.internationalgenome.org/data-portal/data-collection/hgdp)

12

## 13 Data and code availability

14 Data from the NIH All of Us study are available via institutional data access for researchers who meet  
15 the criteria for access to confidential data. To register as a researcher with All of Us, researchers may  
16 use the following URL and complete the laid out steps: <https://www.researchallofus.org/register/>.

17 Researchers can contact All of Us Researcher Workbench Support at [support@researchallofus.org](mailto:support@researchallofus.org). Code  
18 used in this study is available at the pre-production Researcher Workbench “CRC PRS Analysis  
19 (Interactive notebooks)”.

20



## 1   References

- 2   1. Siegel, R.L., Wagle, N.S., Cercek, A., Smith, R.A., and Jemal, A. (2023). Colorectal cancer statistics,  
3   2023. *CA Cancer J. Clin.* *73*, 233–254.
- 4   2. Bundy, J.D., Mills, K.T., He, H., LaVeist, T.A., Ferdinand, K.C., Chen, J., and He, J. (2023). Social  
5   determinants of health and premature death among adults in the USA from 1999 to 2018: a national  
6   cohort study. *Lancet Public Health* *8*, e422–e431.
- 7   3. Siegel, R.L., Miller, K.D., Goding Sauer, A., Fedewa, S.A., Butterly, L.F., Anderson, J.C., Cercek, A.,  
8   Smith, R.A., and Jemal, A. (2020). Colorectal cancer statistics, 2020. *CA Cancer J. Clin.* *70*, 145–164.
- 9   4. Haverkamp, D., Redwood, D., Roik, E., Vindigni, S., and Thomas, T. (2023). Elevated colorectal cancer  
10   incidence among American Indian/Alaska Native persons in Alaska compared to other populations  
11   worldwide. *Int. J. Circumpolar Health* *82*, 2184749.
- 12   5. Sawicki, T., Ruszkowska, M., Danielewicz, A., Niedźwiedzka, E., Arłukowicz, T., and Przybyłowicz, K.E.  
13   (2021). A Review of Colorectal Cancer in Terms of Epidemiology, Risk Factors, Development, Symptoms  
14   and Diagnosis. *Cancers* *13*,.
- 15   6. Rosenthal, E.A., Shirts, B.H., Amendola, L.M., Horike-Pyne, M., Robertson, P.D., Hisama, F.M., Bennett,  
16   R.L., Dorschner, M.O., Nickerson, D.A., Stanaway, I.B., et al. (2018). Rare loss of function variants in  
17   candidate genes and risk of colorectal cancer. *Hum. Genet.* *137*, 795–806.
- 18   7. Patel, S.G., and Ahnen, D.J. (2012). Familial colon cancer syndromes: an update of a rapidly evolving  
19   field. *Curr. Gastroenterol. Rep.* *14*, 428–438.
- 20   8. Lowery, J.T., Ahnen, D.J., Schroy, P.C., 3rd, Hampel, H., Baxter, N., Boland, C.R., Burt, R.W., Butterly, L.,

- 1 Doerr, M., Doroshenk, M., et al. (2016). Understanding the contribution of family history to colorectal  
2 cancer risk and its clinical implications: A state-of-the-science review. *Cancer* *122*, 2633–2645.
- 3 9. Thomas, M., Su, Y.-R., Rosenthal, E.A., Sakoda, L.C., Schmit, S.L., Timofeeva, M.N., Chen, Z.,  
4 Fernandez-Rozadilla, C., Law, P.J., Murphy, N., et al. (2023). Combining Asian and European genome-  
5 wide association studies of colorectal cancer improves risk prediction across racial and ethnic  
6 populations. *Nat. Commun.* *14*, 6147.
- 7 10. Thomas, M., Sakoda, L.C., Hoffmeister, M., Rosenthal, E.A., Lee, J.K., van Duijnhoven, F.J.B., Platz,  
8 E.A., Wu, A.H., Dampier, C.H., de la Chapelle, A., et al. (2020). Genome-wide Modeling of Polygenic Risk  
9 Score in Colorectal Cancer Risk. *Am. J. Hum. Genet.* *107*, 432–444.
- 10 11. Briggs, S.E.W., Law, P., East, J.E., Wordsworth, S., Dunlop, M., Houlston, R., Hippisley-Cox, J., and  
11 Tomlinson, I. (2022). Integrating genome-wide polygenic risk scores and non-genetic risk to predict  
12 colorectal cancer diagnosis using UK Biobank data: population based cohort study. *BMJ* *379*, e071707.
- 13 12. Archambault, A.N., Su, Y.-R., Jeon, J., Thomas, M., Lin, Y., Conti, D.V., Win, A.K., Sakoda, L.C.,  
14 Lansdorp-Vogelaar, I., Peterse, E.F.P., et al. (2020). Cumulative Burden of Colorectal Cancer-Associated  
15 Genetic Variants Is More Strongly Associated With Early-Onset vs Late-Onset Cancer. *Gastroenterology*  
16 *158*, 1274–1286.e12.
- 17 13. Polygenic Risk Score Task Force of the International Common Disease Alliance (2021). Responsible  
18 use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat. Med.* *27*, 1876–1884.
- 19 14. Choi, S.W., Mak, T.S.-H., and O’Reilly, P.F. (2020). Tutorial: a guide to performing polygenic risk score  
20 analyses. *Nat. Protoc.* *15*, 2759–2772.
- 21 15. Lambert, S.A., Gil, L., Jupp, S., Ritchie, S.C., Xu, Y., Buniello, A., McMahon, A., Abraham, G., Chapman,

- 1 M., Parkinson, H., et al. (2021). The Polygenic Score Catalog as an open database for reproducibility and  
2 systematic evaluation. *Nat. Genet.* *53*, 420–425.
- 3 16. Privé, F., Arbel, J., and Vilhjálmsson, B.J. (2021). LDpred2: better, faster, stronger. *Bioinformatics* *36*,  
4 5424–5431.
- 5 17. Márquez-Luna, C., Loh, P.-R., South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2  
6 Diabetes Consortium, and Price, A.L. (2017). Multiethnic polygenic risk scores improve risk prediction in  
7 diverse populations. *Genet. Epidemiol.* *41*, 811–823.
- 8 18. Privé, F., Arbel, J., Aschard, H., and Vilhjálmsson, B.J. (2022). Identifying and correcting for  
9 misspecifications in GWAS summary statistics and polygenic scores. *HGG Adv* *3*, 100136.
- 10 19. Parra, E.J., Kittles, R.A., and Shriver, M.D. (2004). Implications of correlations between skin color and  
11 genetic ancestry for biomedical research. *Nat. Genet.* *36*, S54–S60.
- 12 20. Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., Fuentes,  
13 M., Pizarro, M., Everardo, P., de Avila, F., et al. (2014). Admixture in Latin America: geographic structure,  
14 phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet.* *10*,  
15 e1004572.
- 16 21. Ge, T., Irvin, M.R., Patki, A., Srinivasasainagendra, V., Lin, Y.-F., Tiwari, H.K., Armstrong, N.D., Benoit,  
17 B., Chen, C.-Y., Choi, K.W., et al. (2022). Development and validation of a trans-ancestry polygenic risk  
18 score for type 2 diabetes in diverse populations. *Genome Med.* *14*, 70.
- 19 22. Khera, A.V., Chaffin, M., Zekavat, S.M., Collins, R.L., Roselli, C., Natarajan, P., Lichtman, J.H.,  
20 D’Onofrio, G., Mattera, J., Dreyer, R., et al. (2019). Whole-Genome Sequencing to Characterize  
21 Monogenic and Polygenic Contributions in Patients Hospitalized With Early-Onset Myocardial Infarction.

- 1 Circulation 139, 1593–1602.
- 2 23. All of Us Research Program Investigators, Denny, J.C., Rutter, J.L., Goldstein, D.B., Philippakis, A.,
- 3 Smoller, J.W., Jenkins, G., and Dishman, E. (2019). The “All of Us” Research Program. *N. Engl. J. Med.*
- 4 381, 668–676.
- 5 24. Bick, A., Metcalf, G., Mayo, K., Lichtenstein, L., Rura, S., Carroll, R., Musick, A., Linder, J., Jordan, I.K.,
- 6 Nagar, S., et al. Genomic Data in the All of Us Research Program. *Nature. In Press.*
- 7 25. Sulieman, L., Cronin, R.M., Carroll, R.J., Natarajan, K., Marginean, K., Mapes, B., Roden, D., Harris, P.,
- 8 and Ramirez, A. (2022). Comparing medical history data derived from electronic health records and
- 9 survey answers in the All of Us Research Program. *J. Am. Med. Inform. Assoc.* 29, 1131–1141.
- 10 26. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M.,
- 11 Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human
- 12 genetic variation. *Nature* 526, 68–74.
- 13 27. Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S.,
- 14 Hallast, P., Kamm, J., et al. (2020). Insights into human genetic variation and population history from 929
- 15 diverse genomes. *Science* 367,.
- 16 28. Zheng, X., and Weir, B.S. (2016). Eigenanalysis of SNP data with an identity by descent interpretation.
- 17 *Theor. Popul. Biol.* 107, 65–76.
- 18 29. Hsu, C.-Y., Yang, W., Parikh, R.V., Anderson, A.H., Chen, T.K., Cohen, D.L., He, J., Mohanty, M.J., Lash,
- 19 J.P., Mills, K.T., et al. (2021). Race, Genetic Ancestry, and Estimating Kidney Function in CKD. *N. Engl. J.*
- 20 *Med.* 385, 1750–1760.
- 21 30. Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D., and Mountain, J.L. (2015). The genetic ancestry of

- 1 African Americans, Latinos, and European Americans across the United States. *Am. J. Hum. Genet.* *96*,  
2 37–53.
- 3 31. Hail Team (2021). hail (Github).
- 4 32. Hingorani, A.D., Gratton, J., Finan, C., Schmidt, A.F., Patel, R., Sofat, R., Kuan, V., Langenberg, C.,  
5 Hemingway, H., Morris, J.K., et al. (2023). Performance of polygenic risk scores in screening, prediction,  
6 and risk stratification: secondary analysis of data in the Polygenic Score Catalog. *BMJ Med* *2*, e000554.
- 7 33. Pepe, M.S., Kerr, K.F., Longton, G., and Wang, Z. (2013). Testing for improvement in prediction  
8 model performance. *Stat. Med.* *32*, 1467–1482.
- 9 34. Fry, A., Littlejohns, T.J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., and Allen, N.E.  
10 (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants  
11 With Those of the General Population. *Am. J. Epidemiol.* *186*, 1026–1034.
- 12 35. Schoeler, T., Speed, D., Porcu, E., Pirastu, N., Pingault, J.-B., and Kutalik, Z. (2023). Participation bias  
13 in the UK Biobank distorts genetic associations and downstream analyses. *Nat Hum Behav.*
- 14 36. Ding, Y., Hou, K., Burch, K.S., Lapinska, S., Privé, F., Vilhjálmsón, B., Sankararaman, S., and Pasaniuc,  
15 B. (2022). Large uncertainty in individual polygenic risk score estimation impacts PRS-based risk  
16 stratification. *Nat. Genet.* *54*, 30–39.

## 17 Figures

18

19 **Figure 1:** Each PRS by ancestry. A) RAW=original PRS. B)  $PC_{\mu}$ =PRS adjusted by the first 5 PCs of ancestry  
20 in AOU. C)  $AD_{\mu}$ =PRS adjusted by admixture estimates in AOU. D)  $PC_{REF_{\mu},\sigma}$ =PRS adjusted using a  
21 model trained on the first 5 PCs of ancestry in the 1KG and HGDP reference dataset. E)  $PC_{AOU_{\mu},\sigma}$ =PRS

1 adjusted using a model trained on the first 5 PCs of ancestry in the subset of unassigned AOU  
2 participants (AOU.REF). F) AD.AOU\_μσ=PRS adjusted using a model trained on admixture in AOU.REF.  
3 AFR=African, AMR=Admixed American, EAS=East Asian, EUR=European, MID=Middle Eastern,  
4 OTH=Other, SAS=South Asian.

5 **Figure 2:** log(OR) by quintiles and calibration for African ancestry. A) RAW=raw PRS; B) PC\_μ=raw PRS  
6 adjusted for first 5 PCs of ancestry; C) AD\_μ=raw PRS adjusted for admixture; D) PC.AOU\_μσ=raw PRS  
7 adjusted using a training model developed using a random subset of unscreened controls (AOU.REF) E)  
8 AD.AOU\_μσ=raw PRS adjusted using a training model developed using AOU.REF and admixture  
9 estimates rather than PCs.

10 **Figure 3:** log(OR) by quintiles and calibration for Admixed American ancestry. A) RAW=raw PRS; B)  
11 PC\_μ=raw PRS adjusted for first 5 PCs of ancestry; C) AD\_μ=raw PRS adjusted for admixture; D)  
12 PC.AOU\_μσ=raw PRS adjusted using a training model developed a using random subset of unscreened  
13 controls (AOU.REF). E) AD.AOU\_μσ=raw PRS adjusted using a training model developed using AOU.REF  
14 and admixture estimates rather than PCs. OR is not calculated in the lowest quintile as there were no  
15 cases with a PRS in this quintile.

16 **Figure 4:** log(OR) by quintiles and calibration for European ancestry. RAW=raw PRS; PC\_μ=raw PRS  
17 adjusted for first 5 PCs of ancestry; AD\_μ=raw PRS adjusted for admixture; PC.AOU\_μσ=raw PRS  
18 adjusted using a training model developed using a random subset of unscreened controls (AOU.REF).  
19 AD.AOU\_μσ=raw PRS adjusted using a training model developed using AOU.REF and admixture  
20 estimates rather than PCs.

21 **Figure 5:** log(OR) by quintiles and calibration for Middle Eastern ancestry. A) RAW=raw PRS; B)  
22 PC\_μ=raw PRS adjusted for first 5 PCs of ancestry; C) AD\_μ=raw PRS adjusted for admixture; D)  
23 PC.AOU\_μσ=raw PRS adjusted using a training model developed using a random subset of unscreened

1 controls (AOU.REF). E) AD.AOU\_ $\mu\sigma$ =raw PRS adjusted using a training model developed using AOU.REF  
2 and admixture estimates rather than PCs.

3 **Figure 6:** log(OR) by quintiles and calibration for Other ancestry. A) RAW=raw PRS; B) PC\_ $\mu$ =raw PRS  
4 adjusted for first 5 PCs of ancestry; C) AD\_ $\mu$ =raw PRS adjusted for admixture; D) PC.AOU\_ $\mu\sigma$ =raw PRS  
5 adjusted using a training model developed using random subset of unscreened controls (AOU.REF.) E)  
6 AD.AOU\_ $\mu\sigma$ =raw PRS adjusted using a training model developed using AOU.REF and admixture  
7 estimates rather than PCs.

8 **Figure 7:** Observed versus expected upper percentiles, by genetic ancestry and PRS. A) African B)  
9 Admixed American C) East Asian D) European E) Middle Eastern F) Other G) South Asian, and H) Highly  
10 Admixed (subset of Other). For each panel, solid colored line is observed upper 10th percentile. Dashed  
11 colored line is observed upper 7.5 percentile. Dotted colored line is observed upper 5th percentile. 95%  
12 C.I. are indicated with error bars. Horizontal black solid, dashed and dotted lines indicate the expected  
13 percentiles, respectively. RAW=raw PRS; PC\_ $\mu$ =raw PRS adjusted for first 5 PCs of ancestry; AD\_ $\mu$ =raw  
14 PRS adjusted for admixture; PC.AOU\_ $\mu\sigma$ =raw PRS adjusted using a training model developed using a  
15 random subset of unscreened controls (AOU.REF). AD.AOU\_ $\mu\sigma$ =raw PRS adjusted using a training model  
16 developed using AOU.REF and admixture estimates rather than PCs.

17

## 1 Tables

2 Table 1: Count (percent) by ancestry and case status. “Cleaned cases” refers to participants assigned  
3 case status according to the CRC phenotyping algorithm. “Screened controls” refers to participants  
4 assigned control status by the CRC phenotyping algorithm. “Unassigned” refers to participants who were  
5 not excluded from analysis and were not assigned to case or control status by the CRC phenotype  
6 algorithm. Ancestries are AFR=African, AMR=Admixed American, EAS=East Asian, EUR=European,  
7 MID=Middle Eastern, OTH=Other, SAS=South Asian. For AMR, EAS and SAS, only total counts are given  
8 as the number of cases was <20, to comply with AOU publication rules.

Ancestry	Cleaned Cases	Screened Controls	Unassigned	N(%)
AFR	<b>42</b>	2012	20296	22350 (23)
AMR				9360 (9.5)
EAS				2065 (2.1)
EUR	<b>204</b>	6625	37102	43931 (45)
MID	<b>23</b>	821	3179	4023 (4.1)
OTH	<b>57</b>	2085	13305	15447 (16)
SAS				760 (0.8)
All	348	12378	85210	97936

9

10



1 Table 2: Mean observed age (min, max) by ancestry within all cleaned cases (age of onset), screened  
2 controls (age of last negative screening) and unassigned participants (age of consent). AFR=African,  
3 AMR=Admixed American, EAS=East Asian, EUR=European, MID=Middle Eastern, OTH=Other, SAS=South  
4 Asian. Age diff = difference between age of consent and observed age. Distributions are not reported for  
5 AMR, EAS and SAS due to AOU publication rules for sample sizes < 20.

<b>Ancestry</b>	<b>Cleaned cases</b>	<b>Screened controls</b>	<b>Unscreened controls</b>
AFR	57(41,83)	57(18,90)	48(18,104)
EUR	62(26,89)	61(18,94)	54(18,103)
MID	65(34,84)	64(18,96)	59(18,101)
OTH	55(19,83)	58(18,88)	46(18,100)
All	60(19,89)	60(18,96)	50(18,104)
Age diff	-5.2 (-27,2.5)	-2.5 (-34, 3.7)	0

6  
7  
8

1 Table 3: Properties of the calibrations methods. The training data included either global reference  
2 samples (1KG and HGDP, N=4,151), all of All of Us participants (AOU, N=97,936) or a reference  
3 ancestrally diverse subset of AOU (AOU.REF, N=4,151). 1KG=1000 Genomes; HGDP=Human Genome  
4 Diversity Project; PC=Principal component of ancestry.

5

Method	Training Data	Ancestry Variables	Variance Adjustment
<b>PC_μ</b>	AOU	PC1-PC5	overall variance
<b>AD_μ</b>	AOU	Admixture estimates	overall variance
<b>PC.REF_μσ</b>	1KG and HGDP	PC1-PC5	linear model on PC1-PC5
<b>PC.AOU_μσ</b>	AOU.REF	PC1-PC5	linear model on PC1-PC5
<b>AD.AOU_μσ</b>	AOU.REF	Admixture estimates	linear model on Admixture estimates

6

1 Table 4: Mean (s.d.) of PRS by ancestry. AFR=African, AMR=Admixed American, EAS=East Asian,  
 2 EUR=European, MID=Middle Eastern, OTH=Other, ADM = Highly Admixed subset of OTH, SAS=South  
 3 Asian. RAW=raw PRS; PC\_μ=raw PRS adjusted for first 5 PCs of ancestry; AD\_μ=raw PRS adjusted for  
 4 admixture; PC.REF\_μσ=raw PRS adjusted using a training model developed using 1KG and HGDP  
 5 reference data. PC.AOU\_μσ=raw PRS adjusted using a training model developed using a random subset  
 6 of unscreened controls (AOU.REF). AD.AOU\_μσ=raw PRS adjusted using a training model developed  
 7 using AOU.REF and admixture estimates rather than PCs.

ANC	N	RAW	PC_μ	AD_μ	PC.REF_μσ	PC.AOU_μσ	AD.AOU_μσ
AFR	22385	0.40(0.41)	0.01(0.41)	0.00(0.41)	0.01(1.01)	0.01(0.99)	-0.01(0.99)
AMR	9382	0.29(0.45)	-0.01(0.45)	0.00(0.45)	0.01(1.01)	-0.03(1.00)	0.00(1.01)
EAS	2069	0.37(0.47)	0.00(0.47)	0.00(0.47)	-0.1(1.03)	-0.01(0.95)	-0.02(0.95)
EUR	44099	0.38(0.47)	0.04(0.47)	0.00(0.47)	0.28(0.99)	0.10(0.97)	0.01(1.00)
MID	4056	0.05(0.47)	-0.25(0.47)	0.03(0.47)	-0.43(0.99)	-0.53(0.97)	0.04(0.93)
OTH	15505	0.27(0.47)	-0.07(0.47)	-0.01(0.46)	-0.03(1.01)	-0.14(1.00)	-0.03(0.99)
ADM	6421	0.30 (0.46)	-0.05 (0.46)	0.0 (0.46)		-0.11 (1.01)	0 (1)
SAS	760	0.25(0.47)	0.18(0.47)	0.01(0.47)	-0.17(1.01)	0.28(0.98)	0.14(1.05)

8

9

1 Table 5: Estimated OR (95% C.I.) per s.d. change in PRS, unadjusted (first row) and adjusted for age and  
 2 sex (second row), by ancestry. AFR=African, AMR=Admixed American, EUR=European, MID=Middle  
 3 Eastern, OTH=Other. RAW=raw PRS; PC\_μ=raw PRS adjusted for first 5 PCs of ancestry; AD\_μ=raw PRS  
 4 adjusted for admixture; PC.AOU\_μσ=raw PRS adjusted using a training model developed using a random  
 5 subset of unscreened controls (AOU.REF). AD.AOU\_μσ=raw PRS adjusted using a training model  
 6 developed using AOU.REF and admixture estimates rather than PCs.

7

Ancestry	RAW	PC_μ	PC.AOU_μσ	AD_μ	AD.AOU_μσ
AFR	1.5(1.02,2.08)	1.5(1.02,2.07)	1.4(1.01,1.93)	1.4(1.01,2.04)	1.4(1.01,1.91)
	1.4(0.99,2.02)	1.4(0.99,2.02)	1.4(0.99,1.88)	1.4(0.98,1.99)	1.4(0.98,1.87)
AMR	2.2(1.27,3.85)	2.2(1.26,3.80)	2.2(1.26,3.72)	2.2(1.29,3.84)	2.2(1.28,3.72)
	2.2(1.24,3.75)	2.1(1.23,3.70)	2.1(1.23,3.61)	2.2(1.26,3.74)	2.1(1.25,3.63)
EUR	1.6(1.43,1.89)	1.6(1.43,1.88)	1.7(1.46,1.95)	1.6(1.43,1.87)	1.7(1.45,1.92)
	1.7(1.44,1.90)	1.7(1.44,1.90)	1.7(1.47,1.96)	1.6(1.44,1.88)	1.7(1.46,1.93)
MID	1.1(0.71,1.63)	1.1(0.72,1.63)	1.1(0.7,1.67)	1.1(0.74,1.67)	1.1(0.71,1.77)
	1.1(0.7,1.60)	1.1(0.7,1.59)	1.1(0.68,1.63)	1.1(0.72,1.63)	1.1(0.69,1.73)
OTH	1.5(1.17,1.98)	1.5(1.16,1.97)	1.5(1.17,1.99)	1.5(1.16,1.97)	1.5(1.17,2)
	1.5(1.18,2.01)	1.5(1.17,2.00)	1.6(1.18,2.02)	1.5(1.18,2)	1.5(1.18,2.04)

8

9

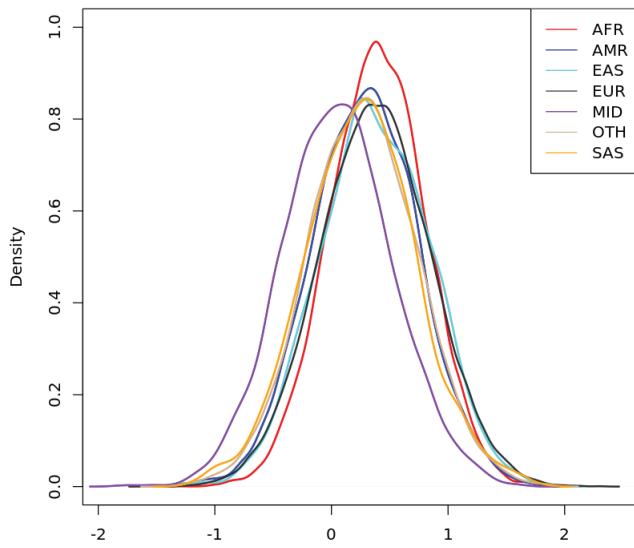
10

1 Table 6: Estimated AUC and 95% C.I. for each PRS, by ancestry, comparing cleaned cases to screened  
 2 controls. For each PRS, the first row contains results for the unadjusted AUC and the second row  
 3 contains results for the AUC adjusted for age and sex. Although the theoretical lower bound for the C.I.  
 4 is 0.5, we include the actual lower value from the bootstrap estimation. AFR=African, AMR=Admixed  
 5 American, EUR=European, MID=Middle Eastern, OTH=Other. RAW=raw PRS; PC\_μ=raw PRS adjusted for  
 6 first 5 PCs of ancestry; AD\_μ=raw PRS adjusted for admixture; PC.AOU\_μσ=raw PRS adjusted using a  
 7 training model developed using a random subset of unscreened controls (AOU.REF). AD.AOU\_μσ=raw  
 8 PRS adjusted using a training model developed using AOU.REF and admixture estimates rather than PCs.

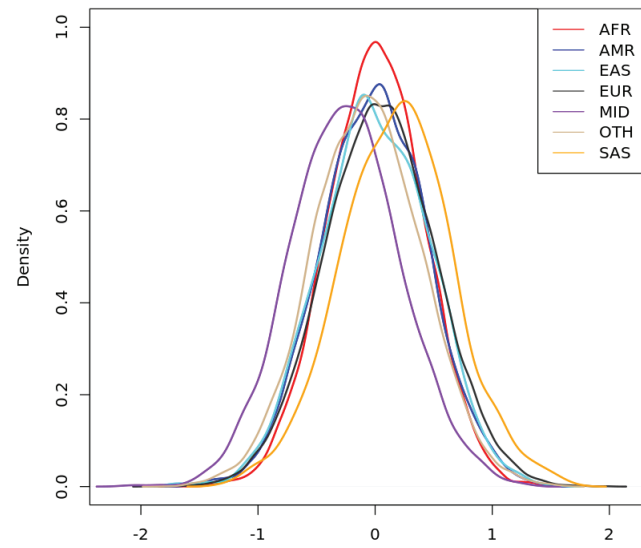
Ancestry	RAW	PC_μ	PC.AOU_μσ	AD_μ	AD.AOU_μσ
AFR	0.6 (0.51, 0.67)	0.6 (0.51, 0.67)	0.6 (0.51, 0.67)	0.6 (0.50, 0.67)	0.6 (0.50, 0.67)
	0.6 (0.50, 0.69)	0.6 (0.49, 0.68)	0.6 (0.50, 0.69)	0.6 (0.50, 0.69)	0.6 (0.50, 0.69)
AMR	0.7 (0.60, 0.83)	0.7 (0.60, 0.83)	0.7 (0.60, 0.82)	0.7 (0.61, 0.83)	0.7 (0.61, 0.83)
	0.7 (0.58, 0.80)	0.7 (0.57, 0.80)	0.7 (0.57, 0.80)	0.7 (0.58, 0.80)	0.7 (0.59, 0.81)
EUR	0.6 (0.59, 0.67)	0.6 (0.59, 0.67)	0.6 (0.59, 0.67)	0.6 (0.59, 0.67)	0.6 (0.59, 0.67)
	0.6 (0.59, 0.68)	0.6 (0.59, 0.68)	0.6 (0.59, 0.68)	0.6 (0.59, 0.67)	0.6 (0.59, 0.68)
MID	0.5 (0.40, 0.67)	0.5 (0.40, 0.67)	0.5 (0.40, 0.67)	0.5 (0.41, 0.67)	0.5 (0.41, 0.67)
	0.5 (0.39, 0.70)	0.5 (0.39, 0.70)	0.5 (0.38, 0.69)	0.6 (0.40, 0.71)	0.6 (0.40, 0.71)
OTH	0.6 (0.56, 0.70)	0.6 (0.56, 0.70)	0.6 (0.56, 0.70)	0.6 (0.56, 0.70)	0.6 (0.56, 0.70)
	0.7 (0.58, 0.71)	0.6 (0.57, 0.71)	0.6 (0.56, 0.70)	0.6 (0.58, 0.71)	0.6 (0.58, 0.71)

9  
 10  
 11

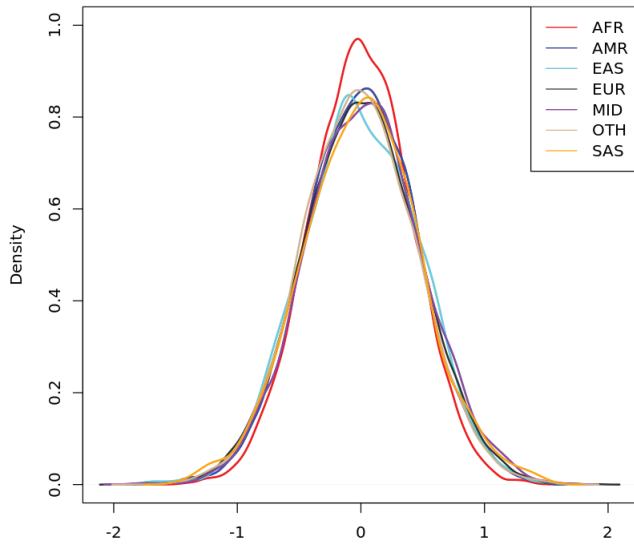
A



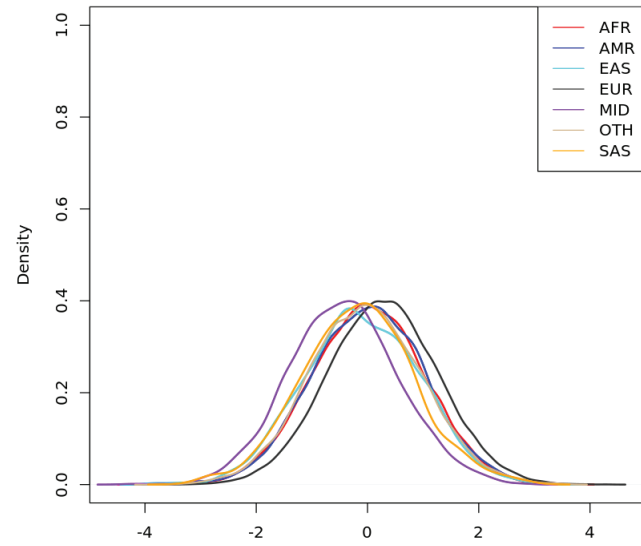
B



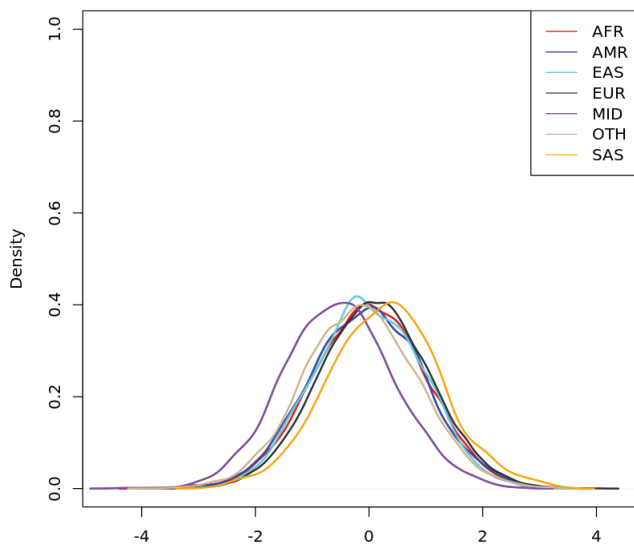
C



D



E



F

