

## Quantifying factors that affect polygenic risk score performance across diverse ancestries and age groups for body mass index

Daniel Hui<sup>1\*</sup>, Brenda Xiao<sup>1\*</sup>, Ozan Dikilitas<sup>2</sup>, Robert R. Freimuth<sup>3</sup>, Marguerite R. Irvin<sup>4</sup>, Gail P. Jarvik<sup>5</sup>, Leah Kottyan<sup>6</sup>, Iftikhar Kullo<sup>7</sup>, Nita A. Limdi<sup>8</sup>, Cong Liu<sup>9</sup>, Yuan Luo<sup>10</sup>, Bahram Namjou<sup>11</sup>, Megan J. Puckelwartz<sup>12</sup>, Daniel Schaid<sup>13</sup>, Hemant Tiwari<sup>14</sup>, Wei-Qi Wei<sup>15</sup>, Shefali Verma<sup>16</sup>, Dokyoon Kim<sup>17</sup>, Marylyn D. Ritchie<sup>18\*\*</sup>

<sup>1</sup>Graduate Program in Genomics and Computational Biology, University of Pennsylvania, Philadelphia, PA, USA

<sup>2</sup>Department of Internal Medicine, Department of Cardiovascular Medicine, Clinician-Investigator Training Program, Mayo Clinic, Rochester MN

<sup>3</sup>Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, USA

<sup>4</sup>Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, United States

<sup>5</sup>Departments of Medicine and Genome Sciences, University of Washington, Seattle WA, USA

<sup>6</sup>Center for Autoimmune Genomics and Etiology, Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA

<sup>7</sup>Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN 55905, USA

<sup>8</sup>Department of Neurology & Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA

<sup>9</sup>Department of Biomedical Informatics, Columbia University, New York, NY, USA

<sup>10</sup>Department of Preventive Medicine (Health and Biomedical Informatics), Northwestern University, Chicago, IL USA

<sup>11</sup>Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA

<sup>12</sup>Department of Pharmacology, Northwestern University, Chicago, IL USA

<sup>13</sup>Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA

<sup>14</sup>Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, United States

<sup>15</sup>Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>16</sup>Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

<sup>17</sup>Department of Biostatistics, Epidemiology and Informatics, Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

<sup>18</sup>Department of Genetics, Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA  
Email: marylyn@penncmedicine.upenn.edu

\*Equal contributions to the manuscript

\*\*Corresponding author

Polygenic risk scores (PRS) have led to enthusiasm for precision medicine. However, it is well documented that PRS do not generalize across groups differing in ancestry or sample characteristics e.g., age. Quantifying performance of PRS across different groups of study participants, using genome-wide association study (GWAS) summary statistics from multiple ancestry groups and sample sizes, and using different linkage disequilibrium (LD) reference panels may clarify which factors are limiting PRS transferability. To evaluate these factors in the PRS generation process, we generated body mass index (BMI) PRS (PRS<sub>BMI</sub>) in the Electronic Medical Records and Genomics (eMERGE) network (N=75,661). Analyses were conducted in two ancestry groups (European and African) and three age ranges (adult, teenagers, and children). For PRS<sub>BMI</sub> calculations, we evaluated five LD reference panels and three sets of GWAS summary statistics of varying sample size and ancestry. PRS<sub>BMI</sub> performance increased for both African and European

ancestry individuals using cross-ancestry GWAS summary statistics compared to European-only summary statistics (6.3% and 3.7% relative  $R^2$  increase, respectively,  $p_{\text{African}}=0.038$ ,  $p_{\text{European}}=6.26 \times 10^{-4}$ ). The effects of LD reference panels were more pronounced in African ancestry study datasets.  $\text{PRS}_{\text{BMI}}$  performance degraded in children;  $R^2$  was less than half of teenagers or adults. The effect of GWAS summary statistics sample size was small when modeled with the other factors. Additionally, the potential of using a PRS generated for one trait to predict risk for comorbid diseases is not well understood especially in the context of cross-ancestry analyses – we explored clinical comorbidities from the electronic health record associated with  $\text{PRS}_{\text{BMI}}$  and identified significant associations with type 2 diabetes and coronary atherosclerosis. In summary, this study quantifies the effects that ancestry, GWAS summary statistic sample size, and LD reference panel have on PRS performance, especially in cross-ancestry and age-specific analyses.

*Keywords: polygenic risk scores (PRS), risk prediction, transferability, diversity*

## Introduction

Polygenic risk scores (PRS) provide individualized genetic estimates of a phenotype by aggregating genetic effects across hundreds or thousands of loci, typically from genome-wide association studies (GWAS). PRS are potentially a powerful source of increased prediction performance, even when combined with family history (1,2). However, in recent years it has become increasingly apparent that performance of PRS is substantially reduced when the ancestry of the individuals in whom prediction is being done differs from the ancestry of the individuals from the GWAS used to generate SNP weights used for PRS construction. For instance, when using GWAS from European ancestry individuals, the prediction accuracy of polygenic scores in individuals of African or Hispanic/Latino ancestry have a relative performance of 25% and 65% compared to performance in European ancestry individuals (3). Additionally, evidence exists suggesting that for some traits, such as adiposity traits, this disparity may be further exacerbated by environmental, demographic, or social risk factors (including age, physical activity, smoking status, and alcohol use (4–7)). For example, differences in the genetic architecture of body mass index (BMI) have been shown to differ between age groups (8–11). Thus, the performance of PRS for BMI is also affected by the age of the individuals used in the GWAS and the study data where the PRS is evaluated (12). Broad-sense heritability estimates for BMI in adults ranges from 40%-90% when estimated in adults of different cohorts even of homogeneous ancestry (13); even if heritability estimates are similar across populations, genetic architecture and enrichment for variants in different functional categories may still differ (14,15).

Several outstanding questions surrounding PRS, especially within the context of adiposity traits and BMI, warrant further investigation. For instance, when cross-ancestry summary statistics (i.e., those including individuals of multiple ancestry groups in the GWAS) are available, can they be used to improve prediction performance in individuals from one or more different ancestry groups? We need a more thorough evaluation of the potential prediction performance gain (or loss) in African ancestry individuals when cross-ancestry GWAS summary statistics are used to estimate the SNP weights. In addition, we need to improve our understanding of the impact of the composition of the linkage disequilibrium (LD) reference panel in combination with cross-ancestry GWAS summary statistics on PRS prediction performance. For prediction of BMI specifically, how does prediction performance differ for individuals in different age groups, especially those who are not adults (i.e., less than age 18)?

Additionally, how much these different variables impact the PRS performance when considered together is important to explore. Developing a deeper understanding of which features (ancestry of individuals in the GWAS, ancestry of the individuals generating the LD references panel, ancestry of the study data, age of the study data) have the greatest impact of PRS performance will help the field develop future studies and strategies around clinical risk prediction with PRS. The degree to which increased GWAS sample size increases prediction performance regardless of these other factors is also important to determine. Finally, there is potential for using a PRS generated for one trait to predict risk for comorbid traits. Understanding how much the different elements of PRS generation affects associations with clinical comorbidities of obesity is of great importance for precision medicine.

We comprehensively investigated the influence of these factors on the performance of PRS using the Electronic Medical Records and Genomics (eMERGE) Network dataset. eMERGE is an NIH funded consortium that combines participants from multiple electronic health record (EHR) linked biobanks (16). In the present study, we included 75,661 individuals of diverse ancestry and age (14% African ancestry, 55% female, and 12% children age < 13). These individuals were from the eMERGE III imputed array dataset (N=83,717) (dbGaP Study: phs001584.v2.p2), estimated European or African ancestry, and had BMI measurements available. For these analyses, we used published BMI GWAS summary statistics from the GIANT (Genetic Investigation of ANthropometric Traits) consortium, an international consortium that primarily studies anthropometric traits, which included participants (max N=339,224, mean N per variant=226,960) from European, African, and Asian ancestry groups (17). We also used summary statistics from a European ancestry BMI GWAS (18) in UK Biobank (UKBB) individuals (N=339,721), which was conducted using both the full sample size of the European ancestry UKBB, as well as after down-sampling to the same number of individuals in the GIANT GWAS. This comparison allowed us to better evaluate whether it was the ancestry composition or the sample size of the dataset where the GWAS summary statistics were derived that affected the results of the PRS performance. We calculated PRS for BMI ( $PRS_{BMI}$ ) across 90 different combinations of analyses (described more in Methods) – six different groupings based on ancestry and age, five different LD reference panels (of varying ancestry and from three different cohorts), and the three mentioned sets of GWAS summary statistics. We then statistically compared the different sets of analyses to see what factors most influence  $PRS_{BMI}$  performance across various groupings of individuals based on ancestry and age. Lastly, we also tested the association of the best performing  $PRS_{BMI}$  with common comorbidities across ancestry groups to identify the clinical relevance of the  $PRS_{BMI}$  in phenotypes derived from an Electronic Health Record (EHR). Investigation of these variables elucidates our understanding of the factors that affect PRS performance and transferability across ancestries and populations, especially within the context of BMI, as well as the potential of using  $PRS_{BMI}$  to predict risk for comorbid disease.

## Methods

### *Overall study design*

The electronic Medical Records and Genomics (eMERGE) network dataset is an NIH funded consortium that combines participants from multiple electronic health record (EHR) linked biobanks. In this study, we included 75,661 individuals with available genetic and phenotypic data. The individuals in the eMERGE dataset include multiple ancestry groups – genetically

inferred ancestry was assigned by the eMERGE consortium (16) – and a large age distribution (14% African ancestry, 19% less than age 18, Figure 1). Briefly, we calculated  $PRS_{BMI}$  for all individuals within each combination where the following elements of the model varied: 1) LD panels that differed in ancestry, 2) GWAS summary statistics with variable ancestry composition, and 3) GWAS summary statistics for two different sample sizes. The details for each of these are provided more below. The data was also split by ancestry and age group, and we statistically compared  $PRS_{BMI}$  performance between all the different groups – in total, 90 sets of  $PRS_{BMI}$  were calculated separately and then compared. We first estimated the effect and significance of each variable (i.e., ancestry of GWAS summary statistics and test data, LD panel ancestry, size of GWAS summary statistics, and age of test individuals) on PRS performance. Next, we estimated how much each variable affects  $PRS_{BMI}$  performance when all are modeled together, and finally we analyzed the potential clinical associations by testing the  $PRS_{BMI}$  for association with common comorbid conditions from the EHR. For the primary results related to LD panel or ancestry of summary statistics and test data, we restricted analyses to adults as the other age groups were limited in sample size. In the following sections, we describe all these elements in more detail.

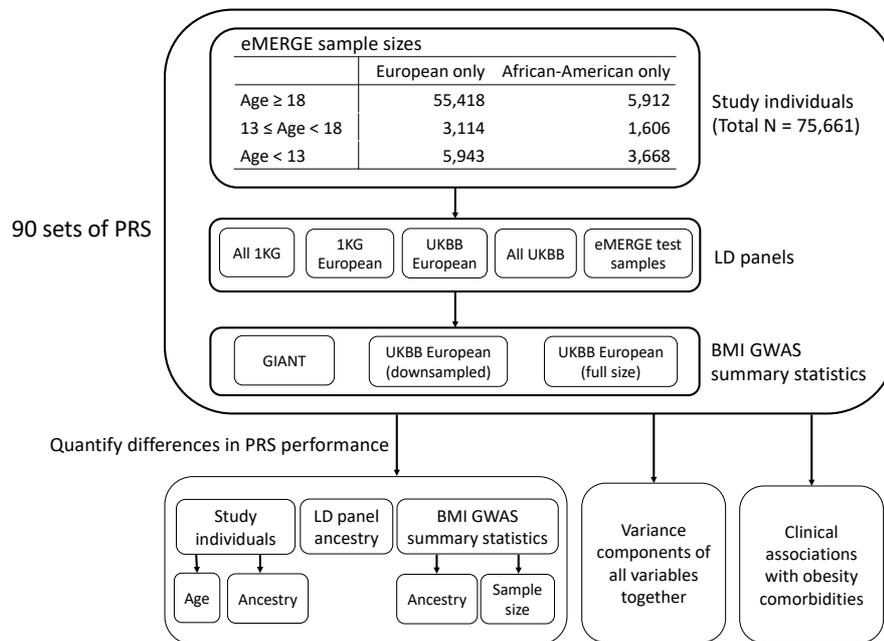


Figure 1. Flowchart of project. Max size of LD panel was 5,000 individuals. UK Biobank (UKBB) European GWAS summary statistics were down-sampled to the mean sample size per variant of GIANT (N=226,960), full size of UKBB European was N=377,921. 1000 Genomes is abbreviated as 1KG.

### **Summary statistics to generate $PRS_{BMI}$**

We obtained published GWAS summary statistics from the GIANT consortium (17) to use as one set of BMI GWAS summary statistics. Up to 322,154 adults of European ancestry, as well as an additional 17,072 adults of non-European ancestry (adults of African, East Asian, and South Asian ancestry), were included in the GIANT GWAS analysis.

For the second set of summary statistics, we performed a GWAS in the individuals of European ancestry from the UK Biobank (UKBB). Individuals were first filtered by low quality

samples (sex mismatch between genetically inferred and self-reported, variant missingness > 5%), relatedness (no 2<sup>nd</sup> degree relatives or higher), and within the White British ancestry subset (with these individuals being defined by UKBB and selected based on self-reports and genetically determined ancestry) (18); a total of 377,921 individuals initially remained. Variants were filtered on imputation quality score (using the INFO metric (19)) > 0.30, and minor allele frequency > 1% within this subset of individuals. In addition, we generated a second set of GWAS summary statistics from the UKBB, where we randomly down-sampled individuals to the sample size in the GIANT GWAS dataset (N=226,960). In each UKBB GWAS, data processing and modeling were performed similarly as in the GIANT GWAS – summary statistics were calculated using linear regression, with age, age<sup>2</sup>, sex, and the first 5 genetic principal components (PCs) included as covariates. BMI, defined as weight in kilograms divided by squared height in meters, was first inverse-rank normal transformed.

After calculation of BMI GWAS summary statistics in each of the two datasets of UKBB individuals of European ancestry, we harmonized variants across all datasets used (UKBB, eMERGE, GIANT, and 1000 Genomes Phase 3). For the remainder of downstream analyses, we kept only those variants that were present in all datasets, and additionally excluded any strand-ambiguous SNPs (alleles A/T or C/G), and retained only biallelic variants; in total, 2,014,457 variants were retained for analyses.

### ***LD reference panels***

Five different LD reference panels were used for each set of PRS<sub>BMI</sub> calculations: 1) all of 1000 Genomes (1KG<sub>All</sub>) (N=2,504), 2) 1000 Genomes European ancestry (1KG<sub>EUR</sub>) (N=503), 3) 5,000 randomly selected European ancestry individuals from the UK Biobank (UKBB<sub>EUR</sub>), 4) 5,000 randomly selected individuals from all of UK Biobank (UKBB<sub>All</sub>), and 5) up to 5,000 randomly selected individuals from the dataset for which PRS<sub>BMI</sub> were being calculated for in the eMERGE dataset (referred to as test data henceforward). These panels were chosen to test for differences in ancestry distribution and sample size on PRS performance.

### ***Statistical methods***

#### ***PRS software***

For each comparison set, PRS<sub>BMI</sub> were calculated using pruning and thresholding method via PRSice v2.1.9 (20). We chose to use PRSice due to the flexibility it provides in choosing external LD panels and allowed us to easily include multi-ancestry LD panels in our analyses. Default parameters were used in all analyses (clumping performed in 250 kb windows using an R<sup>2</sup> of 0.1, p-value step size of 0.00005 between p-values of .0001 up to .10 and step size of .0001 between p-values of .10 up to .50).

#### ***Statistical comparisons***

Incremental R<sup>2</sup> for PRS<sub>BMI</sub> was calculated by subtracting the R<sup>2</sup> using a model with only the covariates from the R<sup>2</sup> of the model using the covariates and the PRS<sub>BMI</sub> (the default option in PRSice). Statistical differences between model performances from different iterations were determined using the Wilcoxon rank-sum test to compare the distributions of the squared residuals generated from the model for all individuals in the iteration; for comparisons between the same set of individuals, the paired Wilcoxon rank-sum test was used. When testing which of

the five LD panels performed the best, we used a Bonferroni-corrected threshold of  $0.05/10 = 0.005$  (ten comparisons between five LD panels). When comparing the best performing  $PRS_{BMI}$  across ancestries and summary statistics using their best LD panel, we used a Bonferroni threshold of  $0.05/25 = 0.002$  (25 comparisons between the five LD panels used).

#### *Proportion of variance explained by each individual variable*

We modeled all evaluated features together in the following linear regression model:

$$R^2 \sim LD\ panel + N_{Sumstat} + Age_{Test} + Ancestry_{Sumstat} + Ancestry_{Test} + Ancestry_{Sumstat} * Ancestry_{Test}$$

Where the *Sumstat* subscript is defined as a set of GWAS summary statistics, and the *Test* subscript is defined as a set of test individuals that PRS prediction is being assessed in. We quantified the variance in  $R^2$  that could be explained by each of these different variables using type II sum of squares from ANOVA. The sum of squares of variables involving ancestry were summed together; an interaction term between summary statistics ancestry and test data ancestry was included to identify whether the ancestry of summary statistics and test data matched.

#### *Association of $PRS_{BMI}$ with comorbidities*

We selected the ten most frequent Phecodes (21) from the EHR data in the eMERGE dataset (which includes obesity as a positive control) to test their association with the  $PRS_{BMI}$ . For each Phecode, individuals were classified as a case for the condition if there was at least one occurrence of the respective Phecode in their EHR record; individuals were classified as a control for that condition if there was no occurrence of the Phecode. This classification is a rule-of-one instance of a Phecode to define case status. For each eMERGE ancestry subgroup, we selected the best performing  $PRS_{BMI}$  i.e., the  $PRS_{BMI}$  with the highest  $R^2$ , and tested the association of the  $PRS_{BMI}$  with these ten clinical conditions using a logistic regression model.  $PRS_{BMI}$  was first mean-centered and standard deviation was set to 1. Sex, age, age<sup>2</sup>, and the first five genetic PCs were included as covariates.

#### *Data visualization*

The ‘ggplot2’ R package was used for plotting, with the ‘geom\_signif’ package used to include significance bars. The association results were plotted using PheWAS-View (22).

## **Results**

### *Effect of LD panel*

For adults of African ancestry, when using the down-sampled UKBB GWAS summary statistics, using either cross-ancestry or African ancestry test data LD panels significantly improved  $PRS_{BMI}$  performance compared to European ancestry LD panels (Figure 2). When using the UKBB summary statistics, the top  $PRS_{BMI}$   $R^2$  was 0.0140 using the test data as LD panel, while the second-best performing LD panel (UKBB European) had an  $R^2$  of 0.0109 ( $p = 4.94 \times 10^{-20}$ ). When using the GIANT summary statistics, the top  $PRS_{BMI}$   $R^2$  was 0.0149 using  $1KG_{All}$  as the reference panel. The  $PRS_{BMI}$  calculated using the best European ancestry panel ( $1KG_{EUR}$ ) resulted in a  $R^2$  of 0.0141, but this difference between these two reference panels was not Bonferroni significant ( $p = 0.037$ ). However, the  $1KG_{All}$  LD panel performed significantly better than the two UKBB LD panels (UKBB<sub>All</sub>:  $R^2 = 0.0134$ ,  $p = 3.65 \times 10^{-5}$ ; UKBB European:  $R^2 = 0.0128$ ,  $p =$

$3.65 \times 10^{-9}$ ). The test data LD panel performed the second-best with an  $R^2$  of 0.0142, and significantly outperformed the UKBB European LD panel ( $p = 4.78 \times 10^{-5}$ ). For adults of European ancestry, we observed more significant differences in performance when using the GIANT summary statistics compared to the down-sampled UKBB summary statistics. The  $1KG_{All}$  LD panel performed the best with a  $R^2$  of 0.0612. It also significantly outperformed all other LD panels ( $1KG_{EUR}$ :  $R^2 = 0.0560$ ,  $p = 5.54 \times 10^{-104}$ ; Test data:  $R^2 = 0.0564$ ,  $p = 6.50 \times 10^{-67}$ ;  $UKBB_{All}$ :  $R^2 = 0.0561$ ,  $8.09 \times 10^{-107}$ ;  $UKBB_{EUR}$ :  $R^2 = 0.0561$ ,  $p = 3.02 \times 10^{-77}$ ). We note that this increase was larger when using the GIANT summary statistics but was still present when using the UKBB summary statistics. When using the UKBB summary statistics, the choice of LD panel had a much smaller impact on prediction performance. While the  $1KG_{All}$  LD panel performed the best, the difference in performance was much less significant between the next best performing LD panel ( $R^2_{1KG_{All}} = 0.0590$ ,  $R^2_{UKBB_{All}} = 0.0583$ ,  $p = 3.48 \times 10^{-4}$ ). The difference between the best and worst performing scores – LD panel using 1KG all versus 1KG European – was also much less significant ( $p = 1.15 \times 10^{-12}$ ). These results suggest that the choice of LD panel particularly matters when calculating  $PRS_{BMI}$  using cross-ancestry GWAS, or for African ancestry individuals when the GWAS summary statistics are derived from European ancestry individuals.

However, we did observe a slight decrease in the impact of the choice of LD panel when using the full UKBB summary statistics for adults; again, the largest differences were observed in adults of African ancestry, but differences in performance across LD panels were not as significant. The test LD panel performed second best with the  $1KG_{EUR}$  LD panel performing best ( $R^2_{Test} = 0.0197$ ,  $R^2_{1KG_{EUR}} = 0.0200$ ,  $p = 0.18$ ). The  $1KG_{All}$  LD panel was the worst performing LD panel with an  $R^2$  of 0.0185, and this difference between the  $1KG_{EUR}$  LD panel was significant after multiple hypothesis correction ( $p = 5.08 \times 10^{-7}$ ).

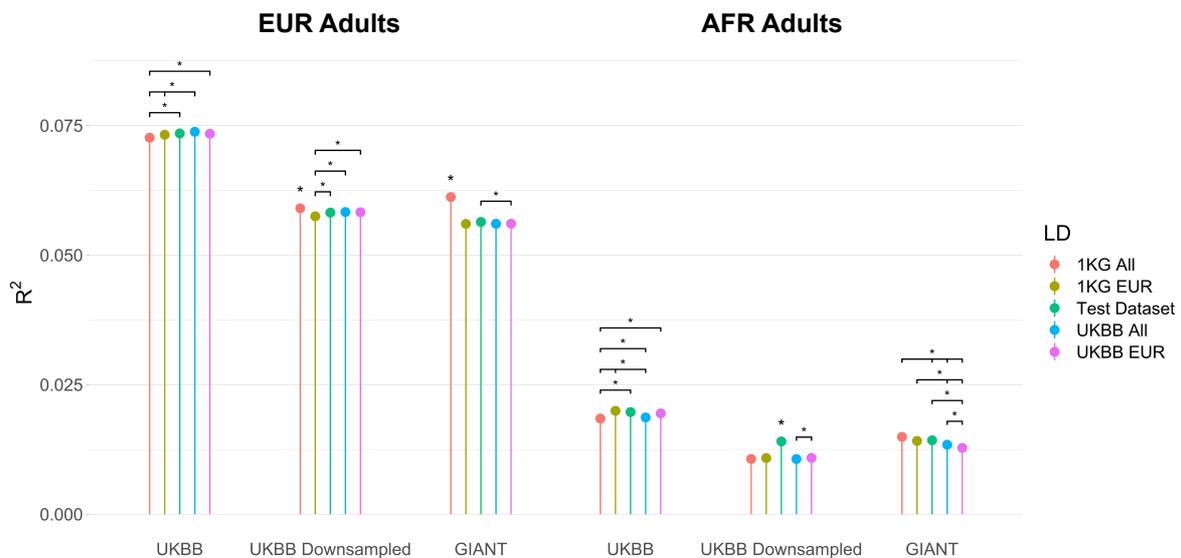


Figure 2. PRS  $R^2$  values across all runs in adults. Asterisks without bars indicate significantly different  $R^2$  values between the other 4 LD panels used. Bars are present for significant differences between specific  $R^2$  values.

### ***Effect of summary statistics and ancestry of test data***

As expected, the  $R^2$  values of the  $PRS_{BMI}$  were significantly higher when calculated for European ancestry adults than adults of African ancestry, even when using the cross-ancestry GIANT summary statistics (Figure 2). When using the GIANT summary statistics, the best performing  $PRS_{BMI}$  in adults of European ancestry had an  $R^2$  of 0.0612, which was significantly higher than the  $R^2$  from the best performing  $PRS_{BMI}$  in African ancestry adults ( $R^2 = 0.0149$ ,  $p < 4.9 \times 10^{-324}$ ).

In African ancestry adults, the  $R^2$  when using the GIANT summary statistics was higher than the  $R^2$  when using the down-sampled UKBB summary statistics with their respective best LD panel (GIANT (1KG<sub>All</sub> LD panel):  $R^2 = 0.0149$ , UKBB (test data LD panel):  $R^2 = 0.0140$ ;  $p = 0.038$ ). This difference was not statistically significant after multiple hypothesis correction. However, the GIANT summary statistics with the 1KG<sub>All</sub> LD panel did significantly outperform the UKBB summary statistics with all other LD panels. When keeping the LD panel constant, the  $PRS_{BMI}$  calculated using the GIANT summary statistics resulted in higher  $R^2$  than using the UKBB summary statistics for all LD panels except for the test data LD panel, and this difference was statistically significant for the 1KG<sub>All</sub> ( $p = 1.55 \times 10^{-33}$ ), 1KG<sub>EUR</sub> ( $p = 6.78 \times 10^{-18}$ ), and UKBB<sub>All</sub> ( $p = 1.28 \times 10^{-15}$ ) LD panels. Somewhat surprisingly, we observed higher  $R^2$  values for European ancestry adults when using the cross-ancestry GIANT summary statistics versus the down-sampled European UKBB summary statistics ( $R^2_{GIANT} = 0.0612$  versus  $R^2_{UKBB} = 0.0590$ ), with this difference being statistically significant ( $p = 6.26 \times 10^{-4}$ ); the best performing LD panel for both set of summary statistics was 1KG<sub>All</sub>.

We also compared prediction performance in all individuals using the full ( $N=377,921$ ) European UKBB GWAS versus the European UKBB GWAS down-sampled to GIANT's sample size ( $N=226,960$ ) (Figure 2, Supplemental Table 1). For consistency, UKBB European individuals were used for the European test ancestry comparisons, and for the African ancestry comparisons the test sets (i.e., African ancestry LD panels) were used as LD panels. Uniformly across test ancestry and age groups, we observed higher and statistically significant increases in  $R^2$ .

### ***Prediction performance across different age groups***

Across different ancestries and summary statistics, we broadly observed similar  $R^2$  values for adults and teenagers, with substantially reduced performance in children (Supplemental Figure 1).  $R^2$  values in children were consistently less than half of that in adults and teenagers, with differences in  $R^2$  values for adults and teenagers being minimal (except in the case of African ancestry individuals using the GIANT summary statistics, with teenagers having more than double the  $R^2$  of adults). Somewhat surprisingly, teenagers consistently had higher  $R^2$  than adults across all analyses, although these differences were much less significant than those compared with children.

### ***Proportion of variance explained by each assessed factor***

While we observed significant differences due to ancestry, age, and number of individuals used to calculate summary statistics, we aimed to quantify the effect of these different variables on  $PRS_{BMI}$  performance when considered together (Table 1). We observed that 89.5% of the variance in  $PRS_{BMI}$   $R^2$  could be explained using these variables, indicating that the majority of the effects of LD panel, ancestry, age, and sample size could be explained through linear

relationships with  $PRS_{BMI} R^2$ . In the context of these comparisons, the ancestry of the summary statistics or test data accounts for 55.1% of the variance explained in  $PRS_{BMI} R^2$ . Choice of LD panel and age of test individuals accounted for similar amounts of variance explained in  $PRS_{BMI} R^2$  (16.5% and 15.9%, respectively), while the number of individuals used to calculate the GWAS summary statistics only accounted for 1.9% of variance explained in  $PRS_{BMI} R^2$ . Per previous sections, while number of individuals used for summary statistics resulted in significant differences in  $PRS_{BMI}$  performance, its overall impact when modeled jointly with all the other factors in the context of these analyses seemed to be small.

Variable	Proportion of explained variance
Ancestry of summary statistics or test data	0.5510
Choice of LD reference panel	0.1650
Age of test individuals	0.1590
N individuals used to calculate summary statistics	0.0195
Residuals (unexplained variance)	0.1050

Table 1. Proportion of variance in  $R^2$  that can be explained by different variables using type II sum of squares from ANOVA.

### ***PRS<sub>BMI</sub> association with comorbid traits***

To determine whether the  $PRS_{BMI}$  was associated with clinical comorbidities, we performed a Phenome-Wide Association Study for ten clinical conditions (Supplemental Table 2, described more in Methods). Here, the  $PRS_{BMI}$  was tested for association with diagnosis codes (Phecodes) to evaluate whether the polygenic background for BMI associates with these clinical diagnoses. The  $PRS_{BMI}$  was significantly associated with several of the most frequent Phecodes in eMERGE, particularly in European adults (Figure 3a). As expected, obesity had the strongest association with  $PRS_{BMI}$  in all ancestry groups ( $p_{EUR} < 4.9 \times 10^{-324}$ ;  $p_{AFR} = 5.17 \times 10^{-8}$ ); this was a positive control. In European ancestry individuals, the best performing  $PRS_{BMI}$  was also significantly positively associated with type 2 diabetes ( $p_{EUR} = 1.04 \times 10^{-102}$ ), essential hypertension ( $p_{EUR} = 7.12 \times 10^{-56}$ ), coronary atherosclerosis ( $p_{EUR} = 3.61 \times 10^{-26}$ ), hyperlipidemia ( $p_{EUR} = 4.38 \times 10^{-16}$ ), depression ( $p_{EUR} = 1.95 \times 10^{-13}$ ), hypercholesterolemia ( $p_{EUR} = 3.64 \times 10^{-15}$ ), asthma ( $p_{EUR} = 3.13 \times 10^{-13}$ ), and diverticulosis ( $p_{EUR} = 0.0017$ ). These associations were less statistically significant in African ancestry individuals, which had much lower sample size, and many associations were no longer significant after Bonferroni correction. Only type 2 diabetes ( $p_{AFR} = 1.2 \times 10^{-5}$ ) and coronary atherosclerosis ( $p_{AFR} = 0.001$ ) were significantly associated with the  $PRS_{BMI}$  in African ancestry adults. We also looked at the prevalence of each condition per PRS quintile for the most significantly associated conditions (Figure 3b). The case prevalence generally increased in higher  $PRS_{BMI}$  quintile groups for conditions significantly associated with the  $PRS_{BMI}$ , a trend matching the results we obtained from the association analysis. Phenotypes with downward trends were not significantly associated with  $PRS_{BMI}$ , and low sample sizes in earlier quintile groups may have contributed to this seemingly decreasing prevalence. We performed similar analyses in teens and children but identified no statistically significant associations (results not shown). The much smaller sample sizes of the Phecodes in these age groups may have also contributed to the lack of statistically significant results – most of these diagnoses are adult-onset conditions.

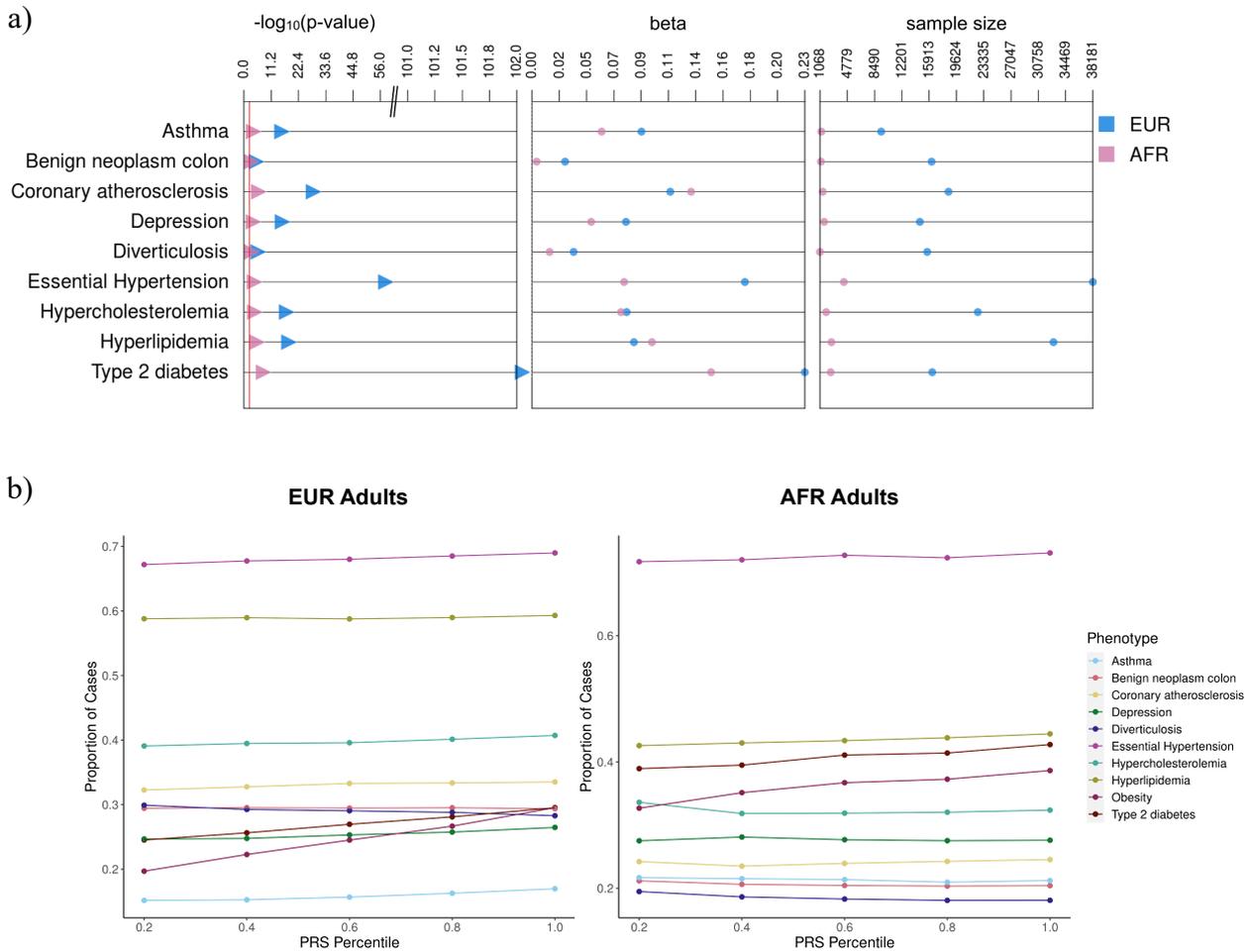


Figure 3. a) Best PRS<sub>BMI</sub> associations with top 9 most prevalent conditions overall in eMERGE adults.

Note the association with obesity is not included in the plot because the p-value in European ancestry individuals was  $p_{EUR} < 4.9 \times 10^{-324}$  which was off the scale of the plot. b) Prevalence plots of significantly associated conditions in eMERGE adults by best performing PRS quintile

### Discussion

Somewhat unintuitively, African ancestry LD panels performed best for African ancestry individuals, regardless of whether European ancestry or cross-ancestry GWAS summary statistics were used. We observed minimal impact of the choice of LD panel when both test data and summary statistics were of European ancestry. These results suggest that as long as either the test data or GWAS summary statistics are of similar ancestry, or the test data and LD panel are of similar ancestry, the difference in PRS performance may be minimal as compared to if all the GWAS summary statistics, test data, and LD panel are all of the same ancestry. We also observed significantly decreased PRS performance in children compared to adults and teens, with the GWAS used in this study being conducted on adult populations.

While the findings in this study highlight many important strategies for performing PRS in different ancestry and age groups, there are limitations that should be addressed in future studies.

First, inclusion of analyses that evaluate how different proportions of non-European ancestry individuals affect the prediction performance of PRS would be useful. The GIANT summary statistics we used in this study are only about 6% non-European ancestry. It may be useful to see how the PRS prediction performance changes in both non-European and European ancestry datasets as a function of the proportion of non-European ancestry samples included in the GWAS. Such analyses may be possible by combining African ancestry individuals from these different datasets. These analyses will be possible once larger datasets that include non-European ancestry cohorts are publicly available or could be tested by analyzing other traits with larger African ancestry GWAS. Future analyses could also include sex-stratified GWAS and comparison sets to evaluate the influence of sex on PRS<sub>BMI</sub> performance. Finally, repeating these types of analyses with different PRS methods would be useful as novel PRS methods are being developed on a regular basis, many of which incorporate ancestry in different ways.

Overall, this study demonstrates the importance of expanding non-European ancestry data resources for PRS, specifically in the generation of GWAS summary statistics and LD reference panels. Failure to do so reduces the impact of PRS in diverse populations and increases the potential for continued health disparities, especially in precision medicine where genetics is being integrated into clinical care.

### **Description of supplemental data**

Supplemental data include one figure and two tables (<https://upenn.box.com/s/7cec5711tjkcyv409vwi7w9t0stkvo12>).

### **Declaration of author competing interests**

Authors have no competing interests to declare.

### **Data and code availability**

Code supporting the current study are available from the corresponding author on request.

### **Acknowledgements**

eMERGE Network (Phase III): This phase of the eMERGE Network was initiated and funded by the NHGRI through the following grants: U01HG8657 (Group Health Cooperative/University of Washington); U01HG8685 (Brigham and Women's Hospital); U01HG8672 (Vanderbilt University Medical Center); U01HG8666 (Cincinnati Children's Hospital Medical Center); U01HG6379 (Mayo Clinic); U01HG8679 (Geisinger Clinic); U01HG8680 (Columbia University Health Sciences); U01HG8684 (Children's Hospital of Philadelphia); U01HG8673 (Northwestern University); U01HG8701 (Vanderbilt University Medical Center serving as the Coordinating Center); U01HG8676 (Partners Healthcare/Broad Institute); and U01HG8664 (Baylor College of Medicine). UK Biobank: All data for this cohort pertained to project 32133 – "Integration of multi-organ imaging phenotypes, clinical phenotypes, and genomic data". MDR would also like to acknowledge NIH AI077505.

### **References**

1. Truong B, Zhou X, Shin J, Li J, van der Werf JHJ, Le TD, et al. Efficient polygenic risk scores for biobank scale data by exploiting phenotypes from inferred relatives. *Nat Commun.* 2020 Jun 17;11(1):3074.
2. Margaux L.A. Hujoel, Po-Ru Loh, Benjamin M. Neale, Alkes L. Price. Incorporating family history of disease improves polygenic risk scores in diverse populations. Available from: <https://www.biorxiv.org/content/10.1101/2021.04.15.439975v1>

3. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019 Apr;51(4):584–91.
4. Rask-Andersen M, Karlsson T, Ek WE, Johansson Å. Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. *PLoS Genet.* 2017 Sep;13(9):e1006977.
5. Robinson MR, English G, Moser G, Lloyd-Jones LR, Triplett MA, Zhu Z, et al. Genotype-covariate interaction effects and the heritability of adult body mass index. *Nat Genet.* 2017 Aug;49(8):1174–81.
6. Sulc J, Mounier N, Günther F, Winkler T, Wood AR, Frayling TM, et al. Quantification of the overall contribution of gene-environment interaction for obesity-related traits. *Nat Commun.* 2020 Mar 13;11(1):1385.
7. Justice AE, Winkler TW, Feitosa MF, Graff M, Fisher VA, Young K, et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun.* 2017 Apr 26;8:14977.
8. Helgeland Ø, Vaudel M, Juliusson PB, Lingaas Holmen O, Juodakis J, Bacelis J, et al. Genome-wide association study reveals dynamic role of genetic variation in infant and early childhood growth. *Nat Commun.* 2019 Oct 1;10(1):4448.
9. Vogelesang S, Bradfield JP, Ahluwalia TS, Curtin JA, Lakka TA, Grarup N, et al. Novel loci for childhood body mass index and shared heritability with adult cardiometabolic traits. *PLoS Genet.* 2020 Oct;16(10):e1008718.
10. Couto Alves A, De Silva NMG, Karhunen V, Sovio U, Das S, Taal HR, et al. GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI. *Sci Adv.* 2019 Sep;5(9):eaaw3095.
11. Choh AC, Lee M, Kent JW, Diego VP, Johnson W, Curran JE, et al. Gene-by-age effects on BMI from birth to adulthood: the Fels Longitudinal Study. *Obes Silver Spring Md.* 2014 Mar;22(3):875–81.
12. Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. *eLife.* 2020 Jan 30;9:e48376.
13. Elks CE, den Hoed M, Zhao JH, Sharp SJ, Wareham NJ, Loos RJF, et al. Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol.* 2012;3:29.
14. Galinsky KJ, Reshef YA, Finucane HK, Loh PR, Zaitlen N, Patterson NJ, et al. Estimating cross-population genetic correlations of causal effect sizes. *Genet Epidemiol.* 2019 Mar;43(2):180–8.
15. Shi H, Gazal S, Kanai M, Koch EM, Schoech AP, Siewert KM, et al. Population-specific causal disease effect sizes in functionally important regions impacted by selection. *Nat Commun.* 2021 Feb 17;12(1):1098.
16. Stanaway IB, Hall TO, Rosenthal EA, Palmer M, Naranbhai V, Knevel R, et al. The eMERGE genotype set of 83,717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype. *Genet Epidemiol.* 2019 Feb;43(1):63–81.
17. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015 Feb 12;518(7538):197–206.
18. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018 Oct;562(7726):203–9.
19. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009 Jun;5(6):e1000529.
20. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinforma Oxf Engl.* 2015 May 1;31(9):1466–8.
21. Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, et al. Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med Inform.* 2019 Nov 29;7(4):e14325.
22. Pendergrass SA, Dudek SM, Crawford DC, Ritchie MD. Visually integrating and exploring high throughput Phenome-Wide Association Study (PheWAS) results using PheWAS-View. *BioData Min.* 2012 Jun 8;5(1):5.