Supplementary material

Table of Contents

Penn Medicine BioBank Team and Contributions

PMBB Leadership Team

Daniel J. Rader, M.D., Marylyn D. Ritchie, Ph.D.

Contribution: All authors contributed to securing funding, study design and oversight. All authors reviewed the final version of the manuscript.

Patient Recruitment and Regulatory Oversight

JoEllen Weaver, Nawar Naseer, Ph.D., M.P.H., Giorgio Sirugo, M.D., P.h.D., Afiya Poindexter, Yi-An Ko, Ph.D., Kyle P. Nerz

Contributions: JW manages patient recruitment and regulatory oversight of study. NN manages participant engagement, assists with regulatory oversight, and researcher access. GS assists with researcher access. AP, YK, KPN perform recruitment and enrollment of study participants.

Lab Operations

JoEllen Weaver, Meghan Livingstone, Fred Vadivieso, Stephanie DerOhannessian, Teo Tran, Julia Stephanowski, Salma Santos, Ned Haubein, P.h.D., Joseph Dunn

Contribution: JW, ML, FV, SD conduct oversight of lab operations. ML, FV, AK, SD, TT, JS, SS perform sample processing. NH, JD are responsible for sample tracking and the laboratory information management system.

Clinical Informatics

Anurag Verma, Ph.D., Colleen Morse Kripke, M.S. DPT, MSA, Marjorie Risman, M.S., Renae Judy, B.S., Colin Wollack, M.S.

Contribution: All authors contributed to the development and validation of clinical phenotypes used to identify study subjects and (when applicable) controls.

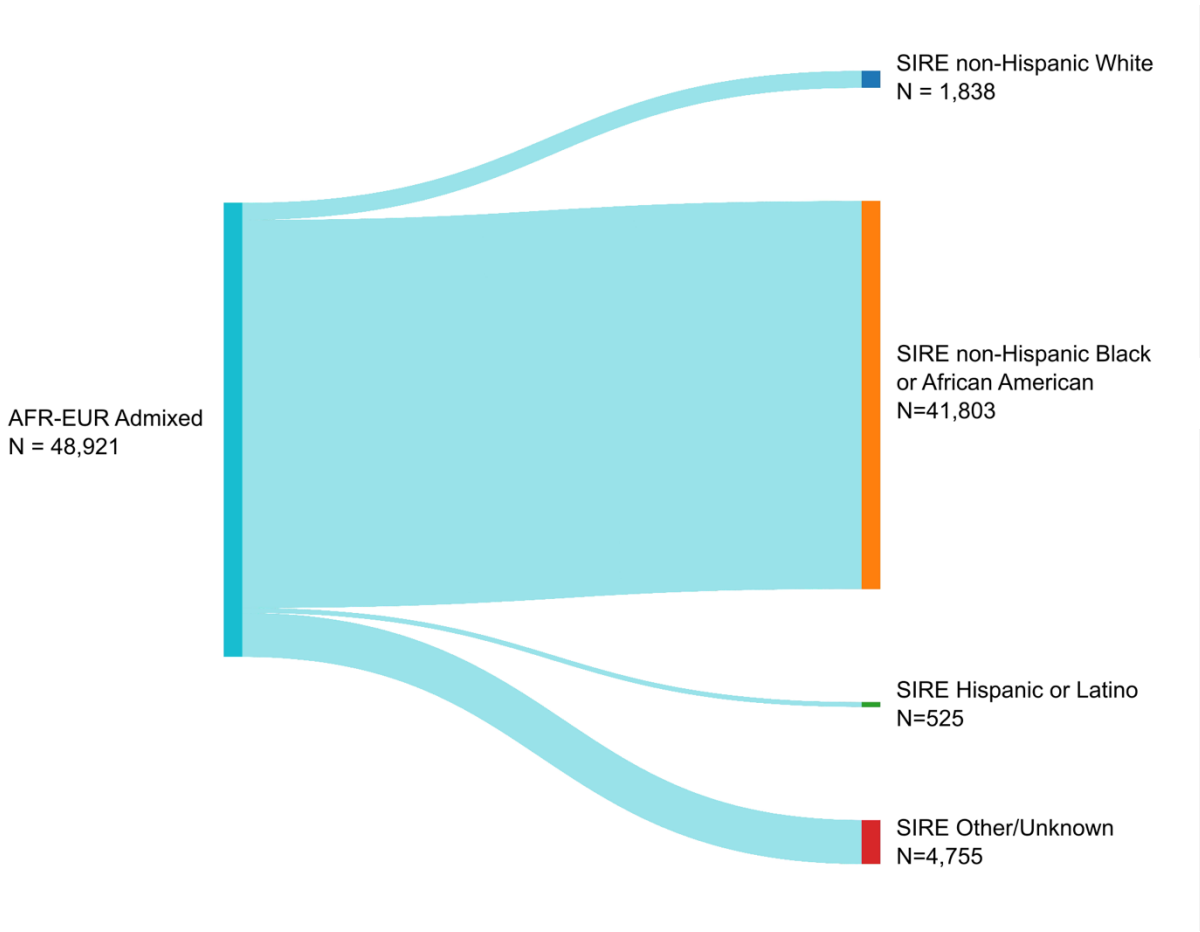
Genome Informatics

Anurag Verma Ph.D., Shefali S. Verma, Ph.D., Scott Damrauer, M.D., Yuki Bradford, M.S., Scott Dudek, M.S., Theodore Drivas, M.D., Ph.D.,

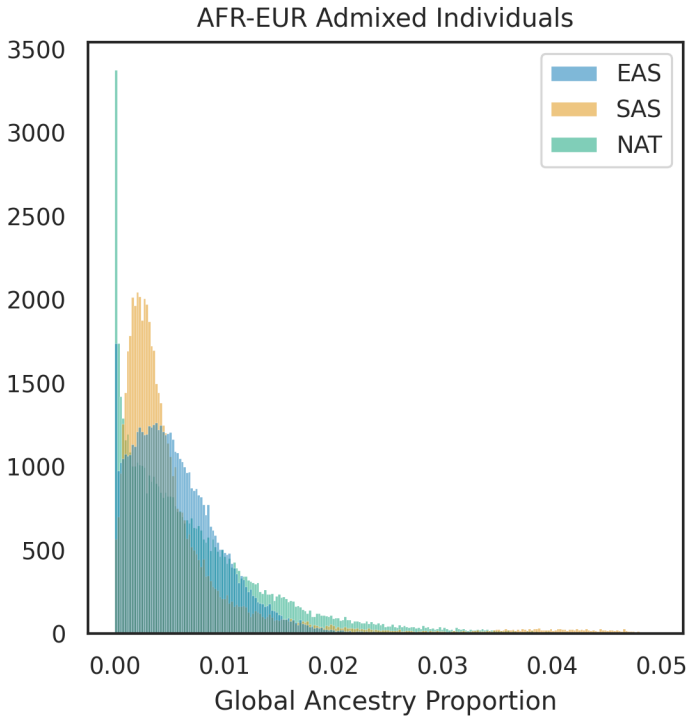
Contribution: AV, SSV, and SD are responsible for the analysis, design, and infrastructure needed to quality control genotype and exome data. YB performs the analysis. TD and AV provides variant and gene annotations and their functional interpretation of variants.

For PMBB, please use: biobank@pennmedicine.upenn.edu

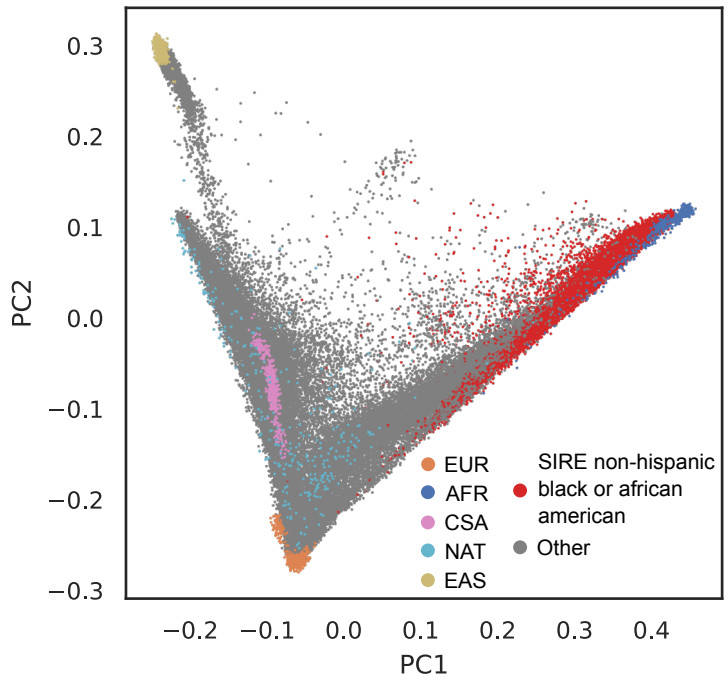
For Regeneron, please use: RGCcollaborations@regeneron.com



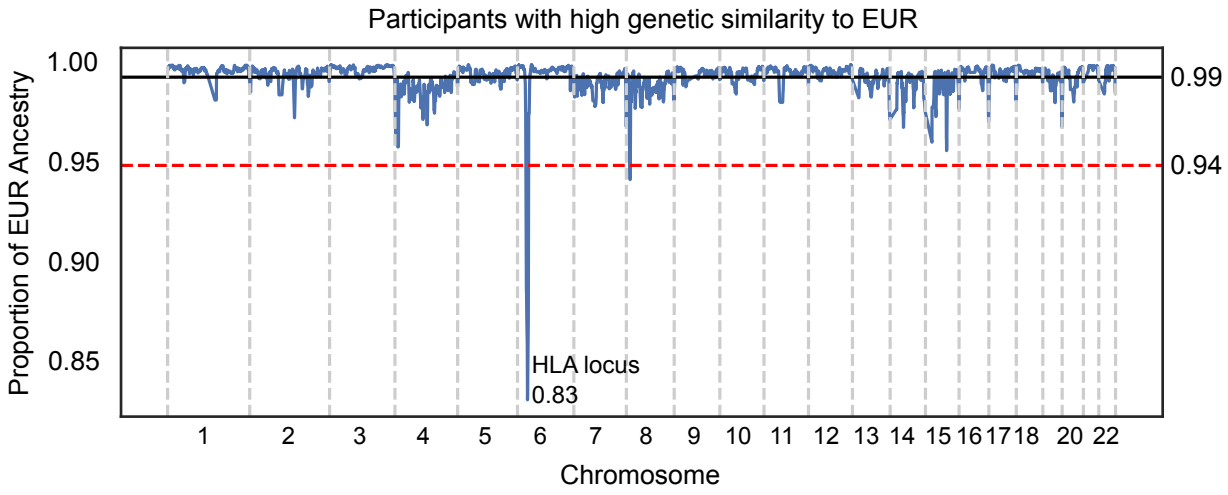
Supplementary Fig. 1. Overlap of genetically inferred AFR-EUR Admixed individuals with SIRE categories*. A Sankey plot of AFR-EUR Admixed individuals and their self-reported race and ethnicity.*



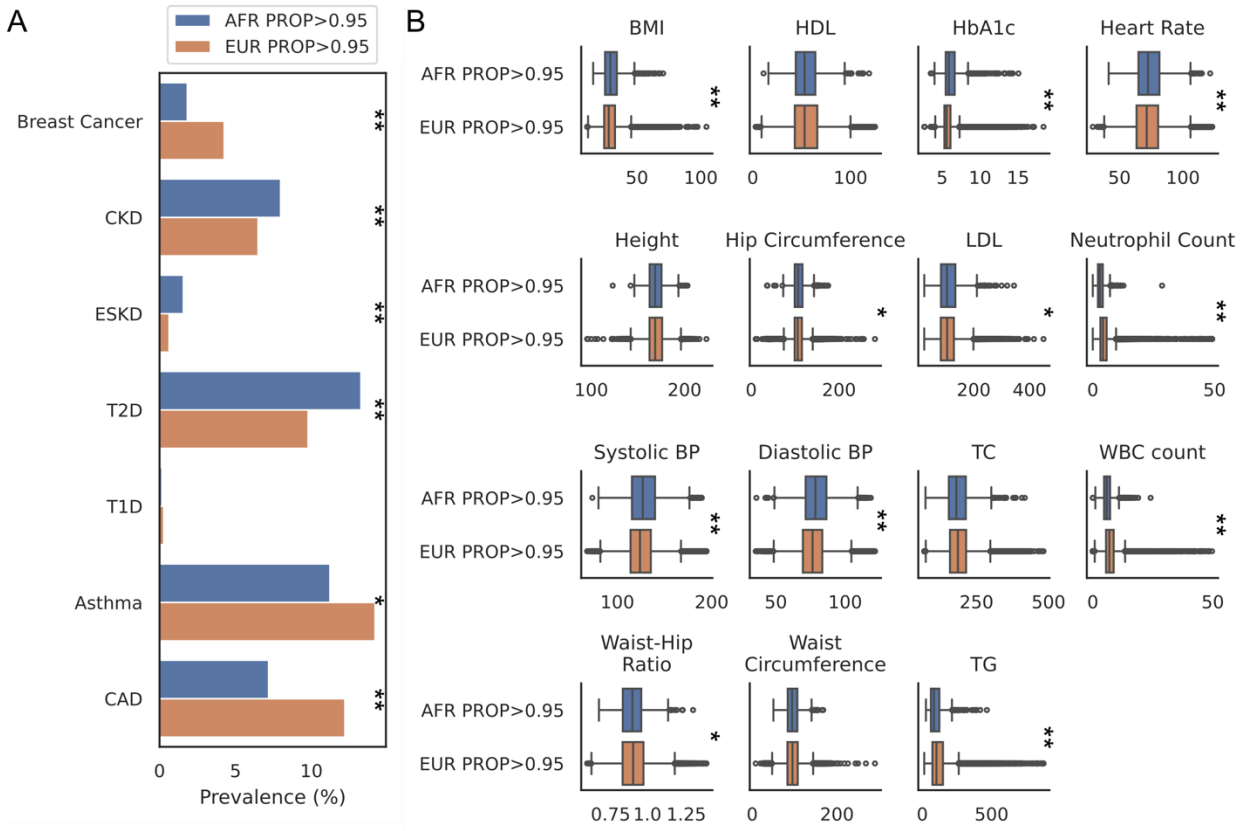
Supplementary Figure 2: Proportion of EAS, SAS, and NAT global ancestry among individuals defined as being AFR-EUR Admixed.



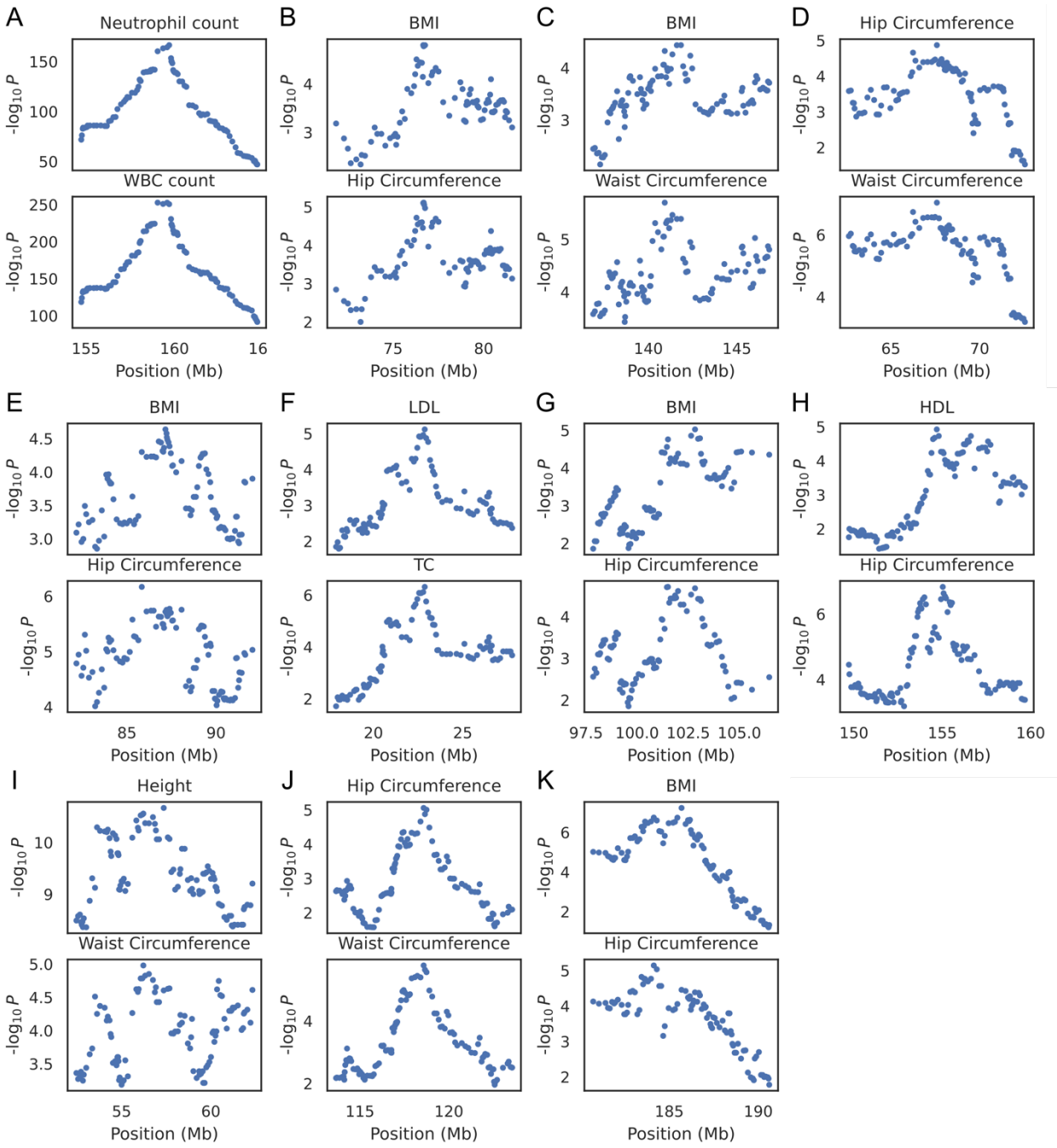
Supplementary Fig. 3. Overlap of SIRE non-Hispanic Black or African American participants in AoU and reference populations from 1000G*. AoU participants plotted in PC space, along with reference populations from 1000G. AoU participants who are not self-report as being non-Hispanic Black or African American colored grey.*



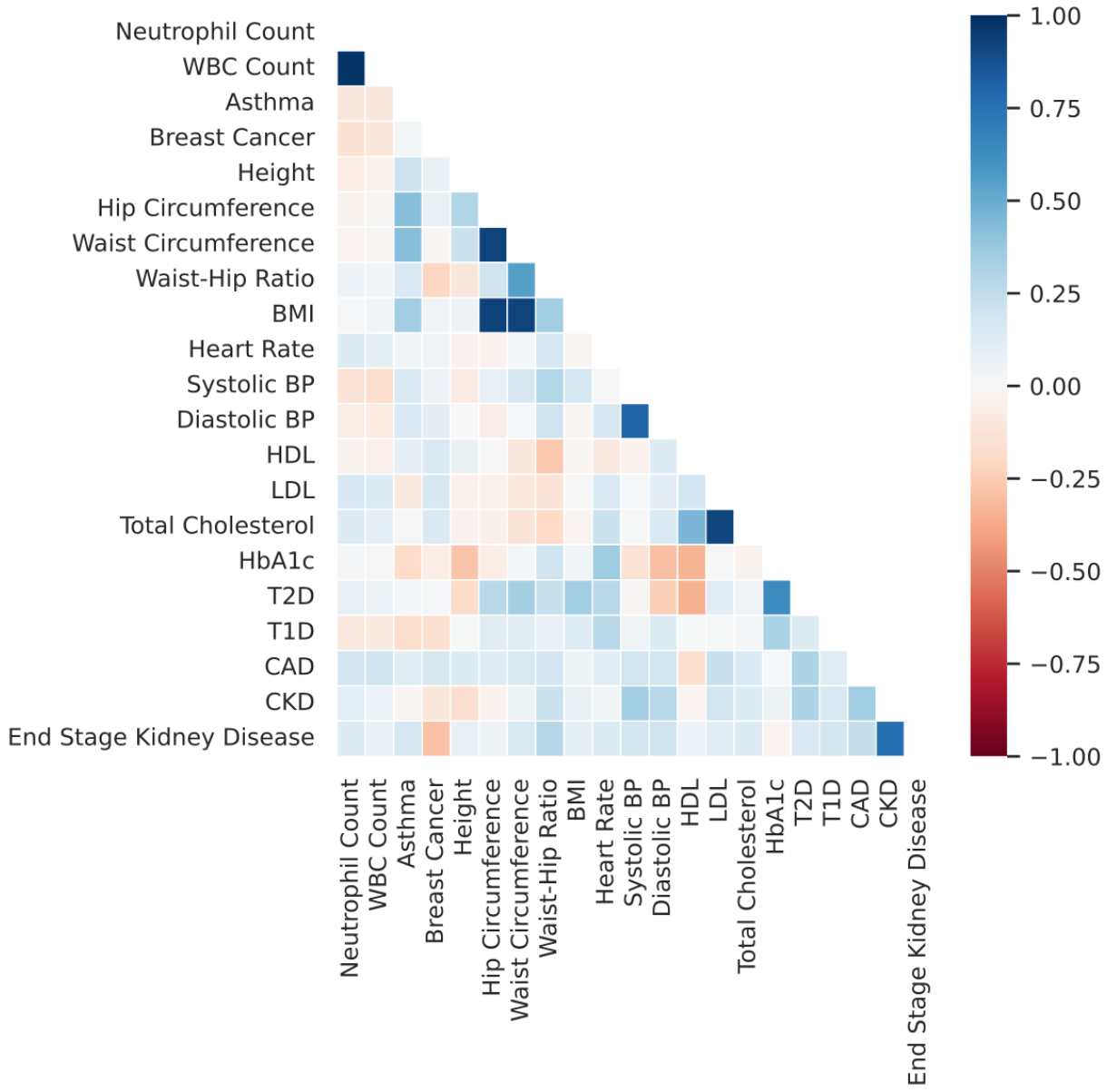
Supplementary Fig. 4. Deviations in local EUR ancestry frequencies among participants with high genetic similarity to EUR*. Proportion of local EUR ancestry across the genome in unrelated, participants in All of Us with high genetic similarity to EUR. Individuals were selected as the 48,921 participants with the smallest Euclidean distance to EUR in PC space, matching the sample size of the AFR-EUR Admixed cohort. The solid black line represents the genome-wide average proportion of local EUR ancestry, and the red dashed lines represent the P<0.05 cutoff of 4.2 SD away from the mean.*



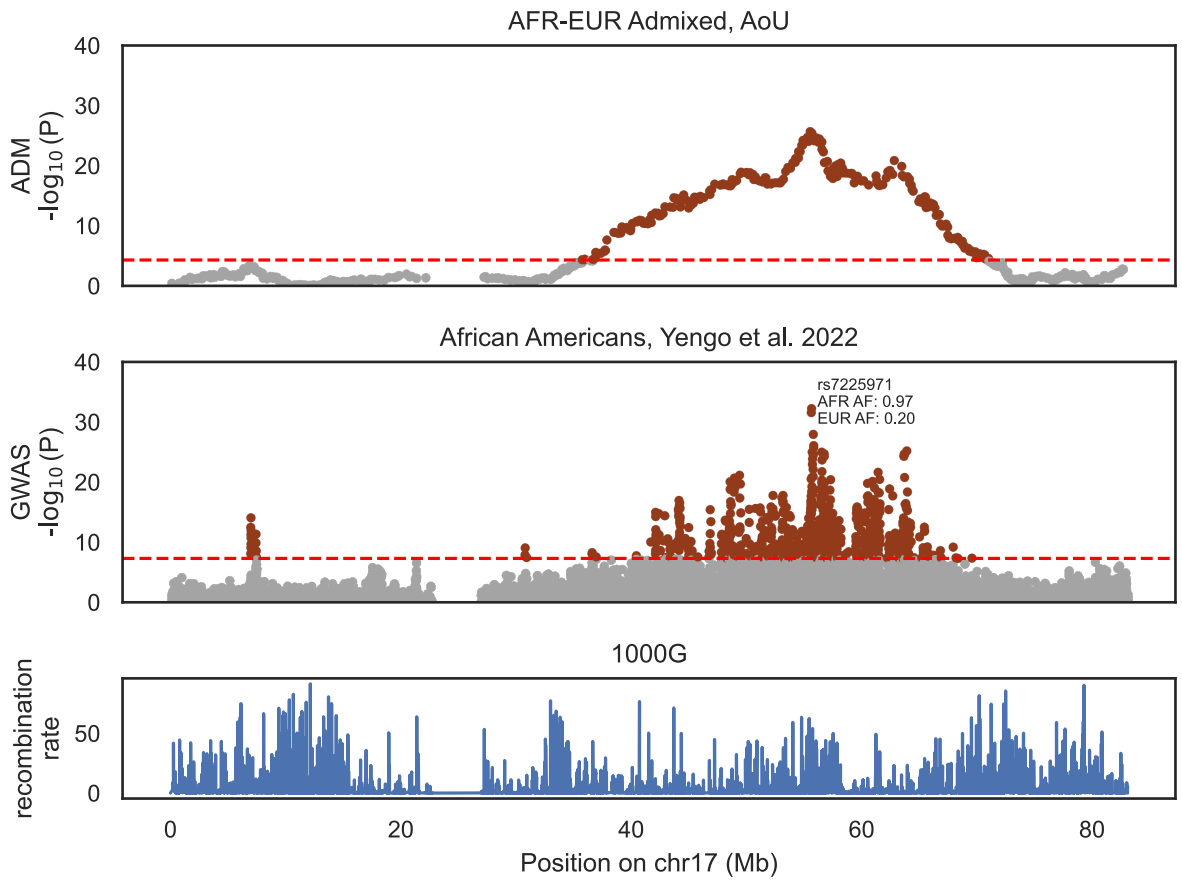
Supplementary Fig. 5. Differences in trait distributions between individuals with genetic similarity to AFR and EUR populations. *(A) Comparison of prevalence rates between individuals with >0.95 global ancestry proportions for either AFR or EUR. P-values were estimated using a two-sided chi2 test. (B) Comparison of quantitative trait distributions between individuals with >0.95 global ancestry proportions for either AFR or EUR. P-values were estimated using a two-sided Welch's t-test. Exact p-values are provided in* ***Supplementary Table 4****. **P<1×10^{-3}. *P<0.05*



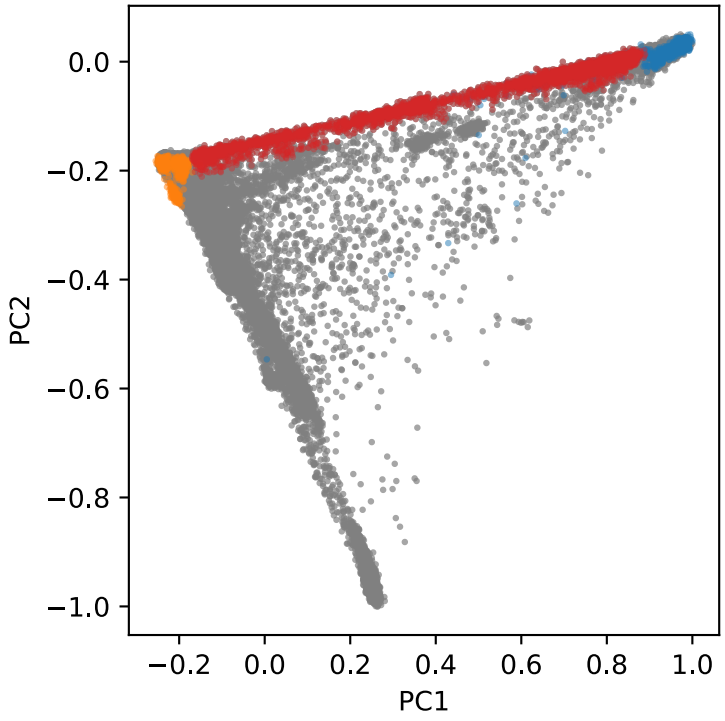
Supplementary Figure 6. Visual comparison of loci with evidence of colocalizing ADM associations. *Scatter plots of all 11 loci with strong evidence of colocalization (PP.H4>0.8) between two separate traits.*



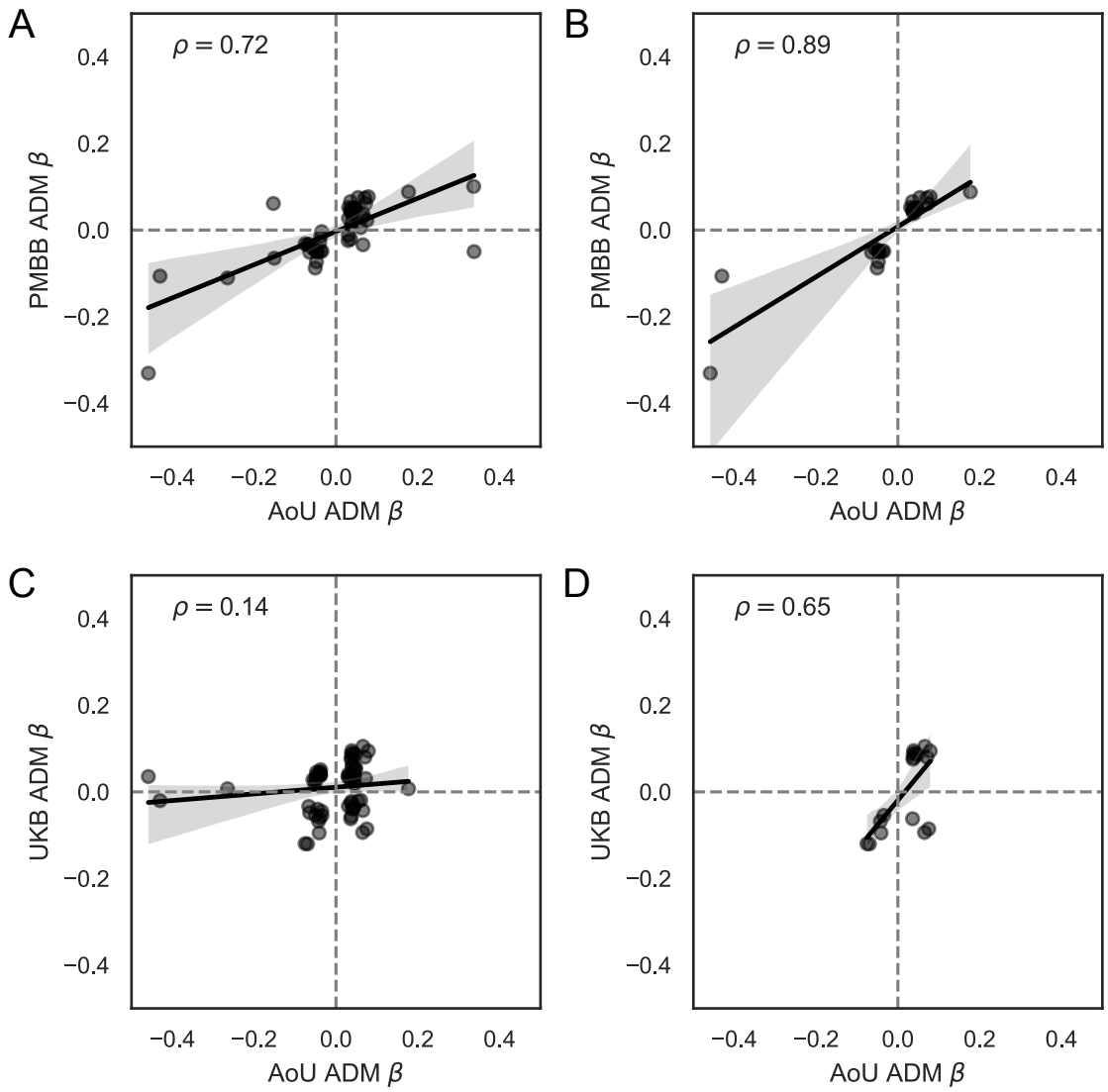
Supplementary Fig. 7. Correlation of effect sizes across ADM-associated loci*. Plotted is a heatmap of Pearson R correlation values between effect sizes per trait pair across all independent loci with at least one genome-wide significant ADM association.*



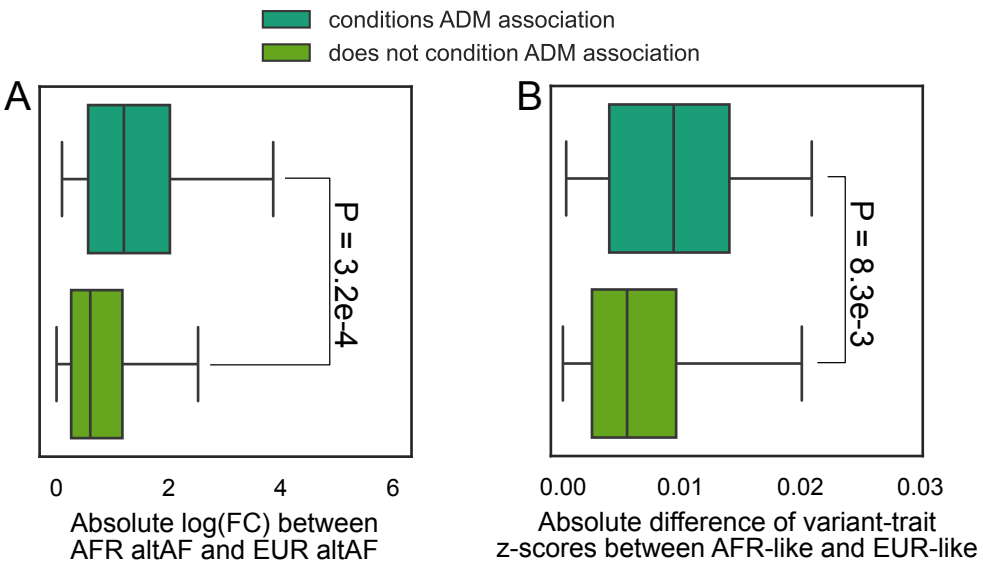
Supplementary Fig. 8. Strong ADM association with height on chr17*. Manhattan plots of the ADM association with height on chr17 (upper panel) and published GWAS from the Yengo et al. 2022 African American height meta-analysis (middle panel). Red points indicate genome-wide significant associations, and the dashed red line indicates the respective genome-wide significance cutoff for ADM (5e-5) and GWAS (5e-8). Allele frequencies (AF) listed are for the G allele of rs7225971 from 1000G populations. The lower panel indicates the recombination rate based on data from 1000G.*



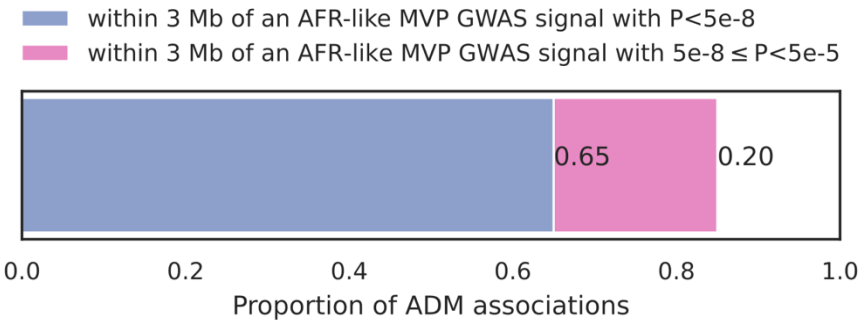
Supplementary Figure 9. UK Biobank participants plotted in PC space. *Red dots highlight genetically inferred UK Biobank two-way admixed AFR/EUR individuals, gray dots denote the entire UK Biobank population, orange dots denote 1000G EUR individuals, and blue dots denote 1000G AFR individuals.*



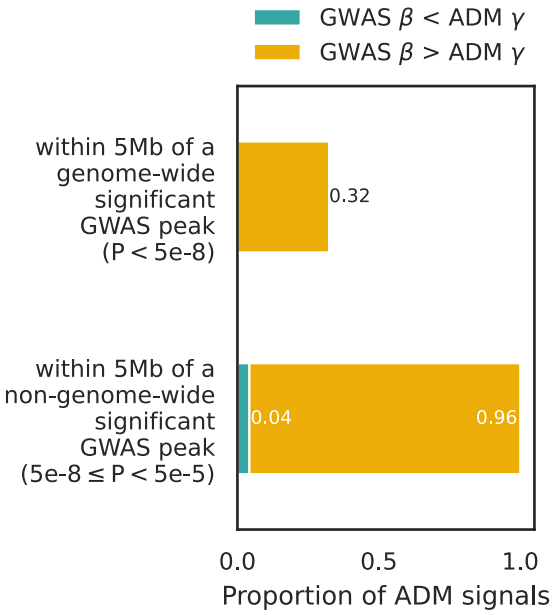
Supplementary Fig. 10. Comparison of admixture mapping results between All of Us, UK Biobank, and Penn Medicine Biobank. *(A) and (B) show scatterplots, with Pearson correlations, of admixture mapping (ADM) association effect sizes between All of Us (AoU) and Penn Medicine Biobank (PMBB), with (A) including all genome-wide significant ADM associations in AoU and (B) including a subset of ADM associations that are also nominally associated in PMBB with P<0.05. (C) and (D) show the same analysis, comparing AoU ADM effect sizes to those estimated in UK Biobank (UKBB).*



Supplementary Fig. 11. Alternative comparisons of variants which condition ADM associations*. (A) Comparison of absolute log fold-change (FC) for the same alternative allele frequencies (altAF) between AFR and EUR populations from 1000G for variants which condition and do not condition ADM associations. (B) Comparison of absolute difference in z-scores for variant-trait associations between AFR-like and EUR-like populations for variants which condition and do not condition ADM associations. Boxplots for "conditions ADM associations" contains the 36 variants which condition an ADM association, while boxplots for "does not condition ADM association" contains the 1,208 variants which do not condition an ADM association. Boxplots include median, upper and lower quartiles, and 1.5x interquartile ranges. P-values were calculated using a two-sided Mann-Whitney U test.*



Supplementary Figure 12: Overlap of MVP GWAS with ADM associations*. Stacked bar plot of the proportion of ADM associations within 3 Mb of a GWAS association in AFR-like individuals within the Million Veteran's Program (MVP). Proportions are out of the number of ADM associations with a corresponding phenotype in MVP (48 total ADM associations).*



Supplementary Figure 13: Summary of ADM within 5Mb of an in-sample GWAS peak. *Summary of the proportion of ADM signals within 5 Mb of a variant associated with the same trait in AFR-EUR ADM individuals, stratified by the P-value of each variant and whether that effect size of the variant in GWAS is larger than the ADM effect size.*