The American Journal of Human Genetics, Volume *108*

Supplemental data

Whole-genome sequencing of African Americans implicates differential genetic architecture in inflammatory bowel disease

Hari K. Somineni, Sini Nagpal, Suresh Venkateswaran, David J. Cutler, David T. Okou, Talin Haritunians, Claire L. Simpson, Ferdouse Begum, Lisa W. Datta, Antonio J. Quiros, Jenifer Seminerio, Emebet Mengesha, Jonathan S. Alexander, Robert N. Baldassano, Sharon Dudley-Brown, Raymond K. Cross, Themistocles Dassopoulos, Lee A. Denson, Tanvi A. Dhere, Heba Iskandar, Gerald W. Dryden, Jason K. Hou, Sunny Z. Hussain, Jeffrey S. Hyams, Kim L. Isaacs, Howard Kader, Michael D. Kappelman, Jeffry Katz, Richard Kellermayer, John F. Kuemmerle, Mark Lazarev, Ellen Li, Peter Mannon, Dedrick E. Moulton, Rodney D. Newberry, Ashish S. Patel, Joel Pekow, Shehzad A. Saeed, John F. Valentine, Ming-Hsi Wang, Jacob L. McCauley, Maria T. Abreu, Traci Jester, Zarela Molle-Rios, Sirish Palle, Ellen J. Scherl, John Kwon, John D. Rioux, Richard H. Duerr, Mark S. Silverberg, Michael E. Zwick, Christine Stevens, Mark J. Daly, Judy H. Cho, Greg Gibson, Dermot P.B. McGovern, Steven R. Brant, and Subra Kugathasan

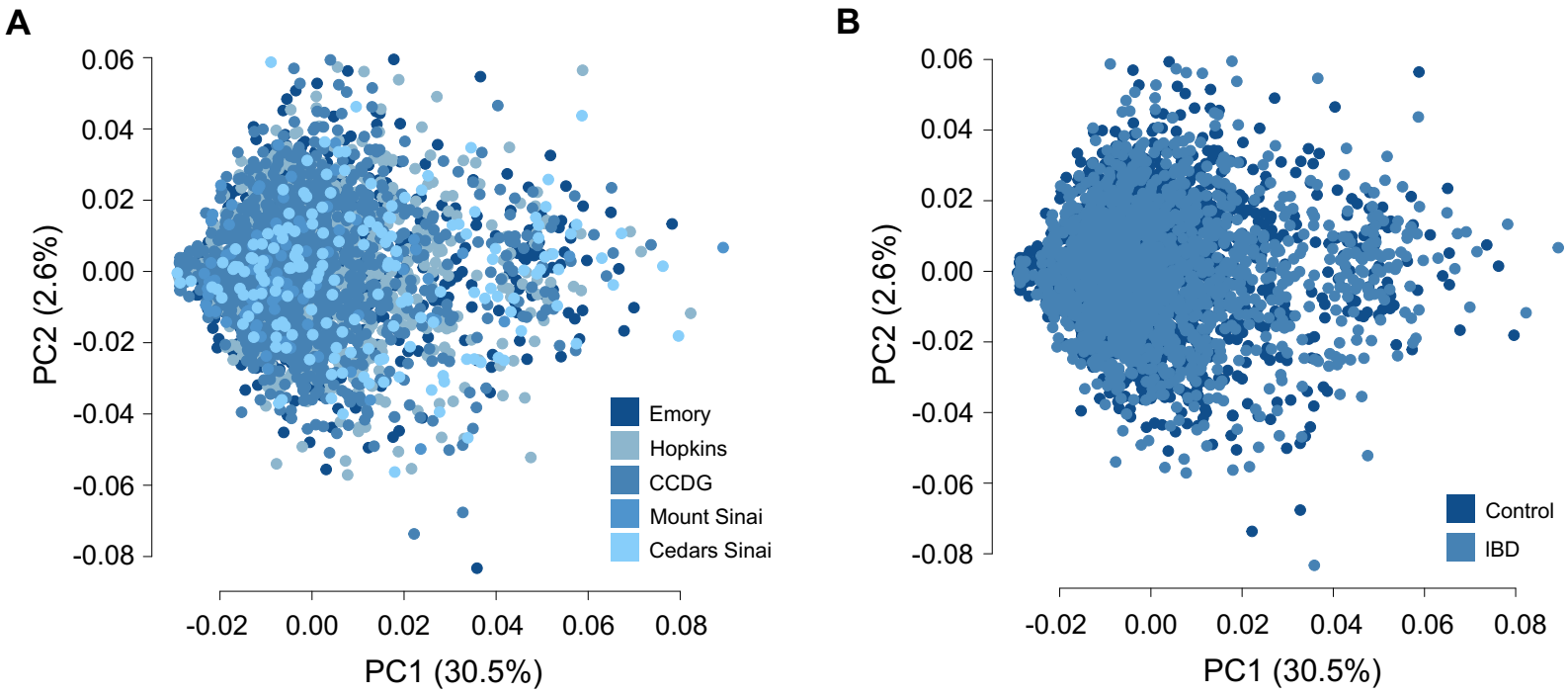

Figure S1. Principal component analysis of discovery whole-genome sequence cohort genotypes. Principal component plots of genetic data for the 3418 African American subjects included in the discovery analyses. Individuals are color coded based on either the sample collection site **(A)** or case-control status **(B)** Plots were based on ~1.4 million, LD-pruned (r^2 < 0.1), high frequency (MAF > 1%) variants.

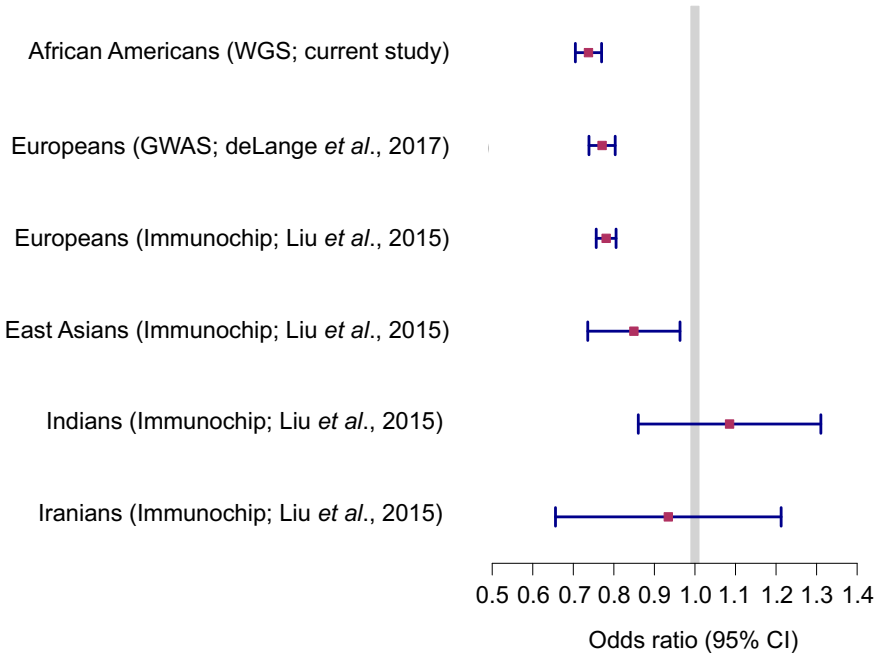

Figure S2. Crohn's disease odds ratio of *PTGER4* locus across divergent populations.
Odds ratio values and 95% confidence intervals (CI) of signal 1 at *PTGER4* locus that reached standard (GWAS) genome-wide significance for Crohn's disease in the current study are shown across divergent populations and attributed to respective studies.

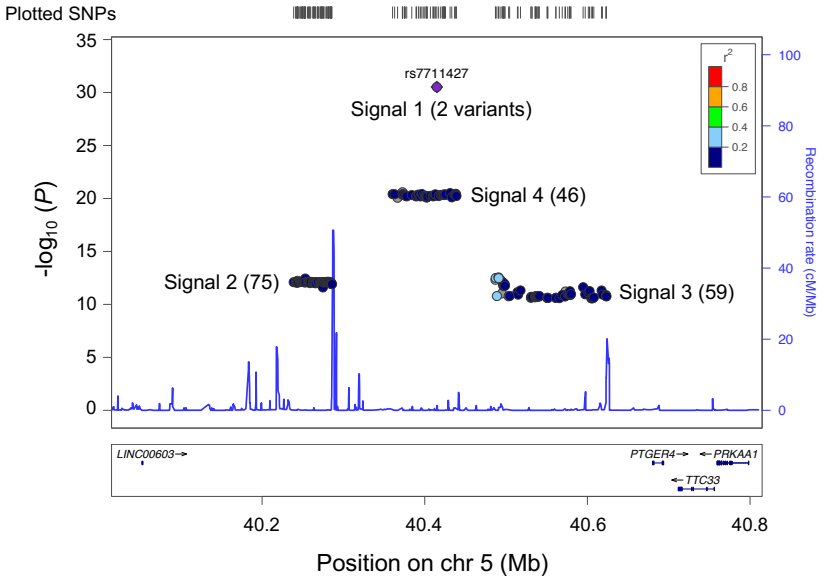

Figure S3. LocusZoom plot of credible variants in *PTGER4* locus fine-mapped recently in European populations.

189 credible variants from Huang *et al*.[24] representing four independent signals within the *PTGER4* locus are shown. The sentinel SNP from signal 1 is shown in *purple* (one of the two credible variants from signal 1 from Huang *et al*.). The remainder of the credible SNPs are color coded based on their pair-wise LD with the sentinel SNP. SNPs with missing LD information are in *gray*. Number of credible variants in each signal is indicated.

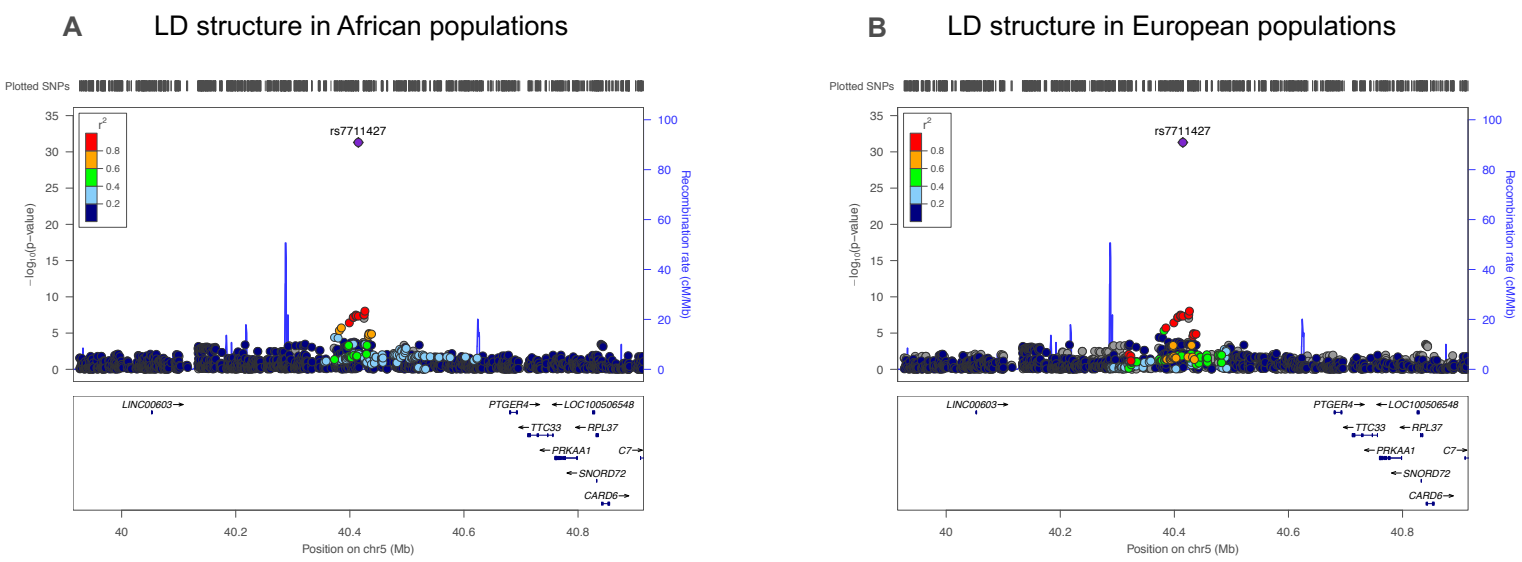

Figure S4. The 22 variant *PTGER4* African American Crohn's disease signal is in LD with the known European signal.

LocusZoom plot of *PTGER4* based on the LD structure in **(A)** African and **(B)** European populations from the 1000 Genomes. rs7711427 – one of the two credible variants from signal 1 in *PTGER4* locus fine-mapped by Huang *et al*.[24] in European populations – was chosen as the tag variant (shown in *purple*). Because this variant failed QC in this study, we used the *P* value obtained in Huang *et al*., for of association with Crohn's disease. *P* values obtained in this study are shown for the remaining variants. Variants are color coded based on their pair-wise LD with the tag SNP. Variants with missing LD information are in *gray*.

Figure S5. QQ plots of single-marker rare variant association analysis.
(A) Inflammatory bowel disease, **(B)** Crohn's disease, and **(C)** ulcerative colitis.

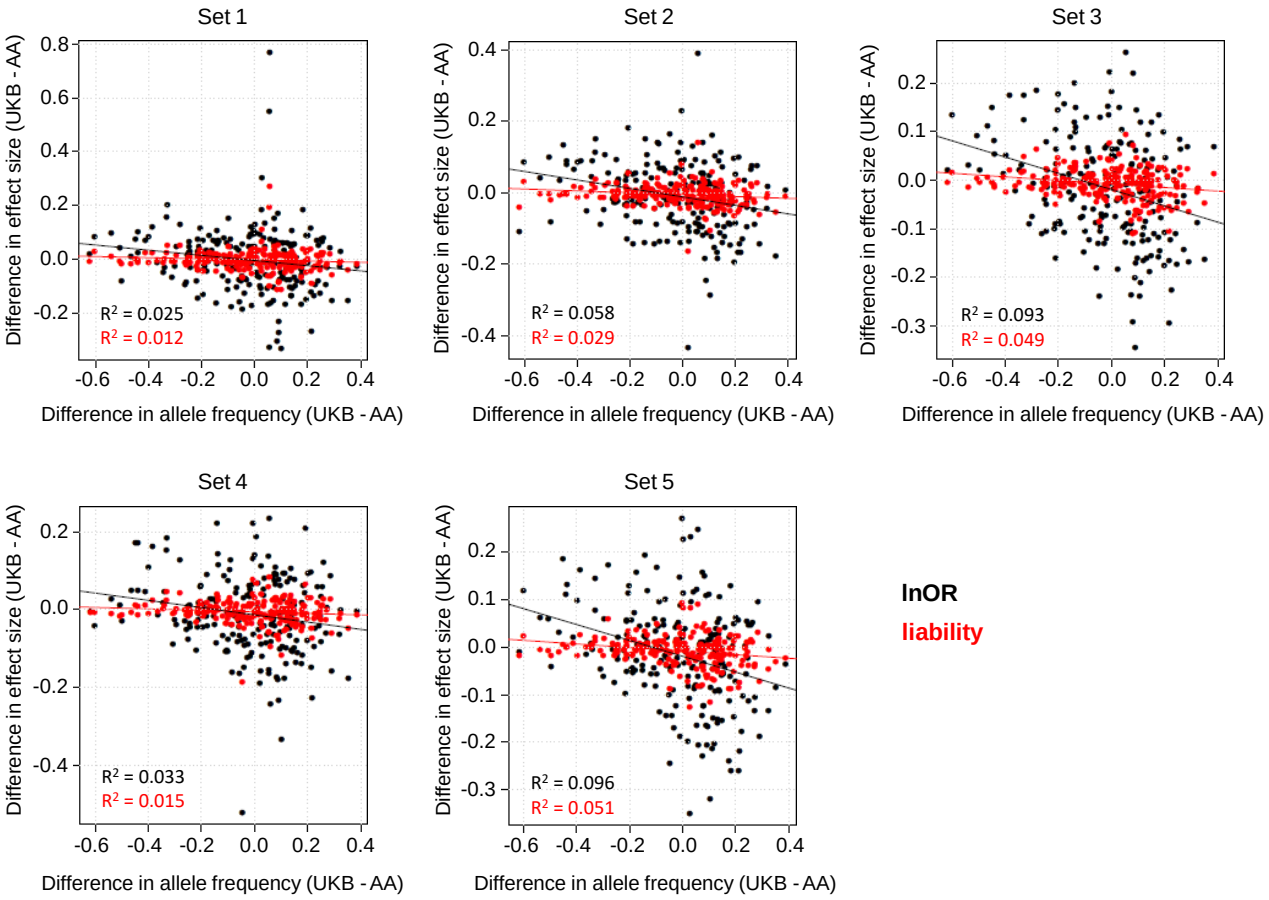

Figure S6. Difference in allele frequencies vs difference in effect sizes in Europeans and African Americans on logit and liability scales.

Differences in allele frequency between the UK Biobank and African American pseudo-controls datasets at known loci (n = 215 sentinel SNPs) were plotted against their estimated effect size differences on lnOR and liability scale across all five discovery sets. Note that the slope of the black regression lines is non-zero reflecting dependence of lnOR estimates on allele frequencies, and in each case the regression R^2 indicates that the differences in allele frequency explain 6.1% (mean R^2 across five datasets) of the differences in effect size estimated on lnOR scale vs 3.1% with liability scale.

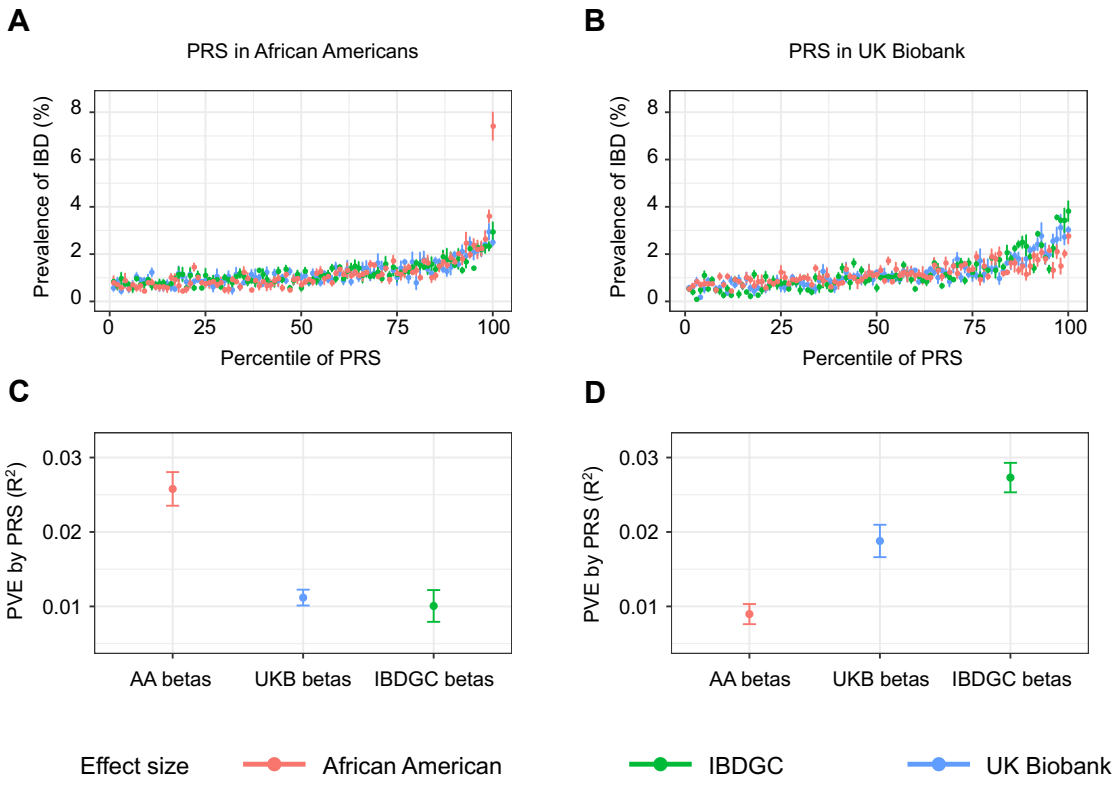

Figure S7. Polygenic risk scores (PRS) in African Americans and Europeans as a function of different discovery ancestry groups.
PRS in African American pseudo-controls cohort and UK Biobank derived using 215 of the known disease SNPs and their effect sizes estimated on lnOR scale from African American pseudo-controls cohort (AA_betas), UK Biobank discovery cohort (UKBB_betas) and European GWAS meta-analysis3 (IBDGC_betas). **(A, B)** prevalence of IBD (%) vs percentile of PRS in each cohort with standard error bars on the prevalence from 5 test sets. **(C, D)** show the mean proportion of variance explained (PVE) by each PRS with error bars from 5 test sets. AA effect estimates are 40%-50% greater than UKB, and PVE with UKB allele frequencies are also 35%-45% greater than corresponding AA estimates.

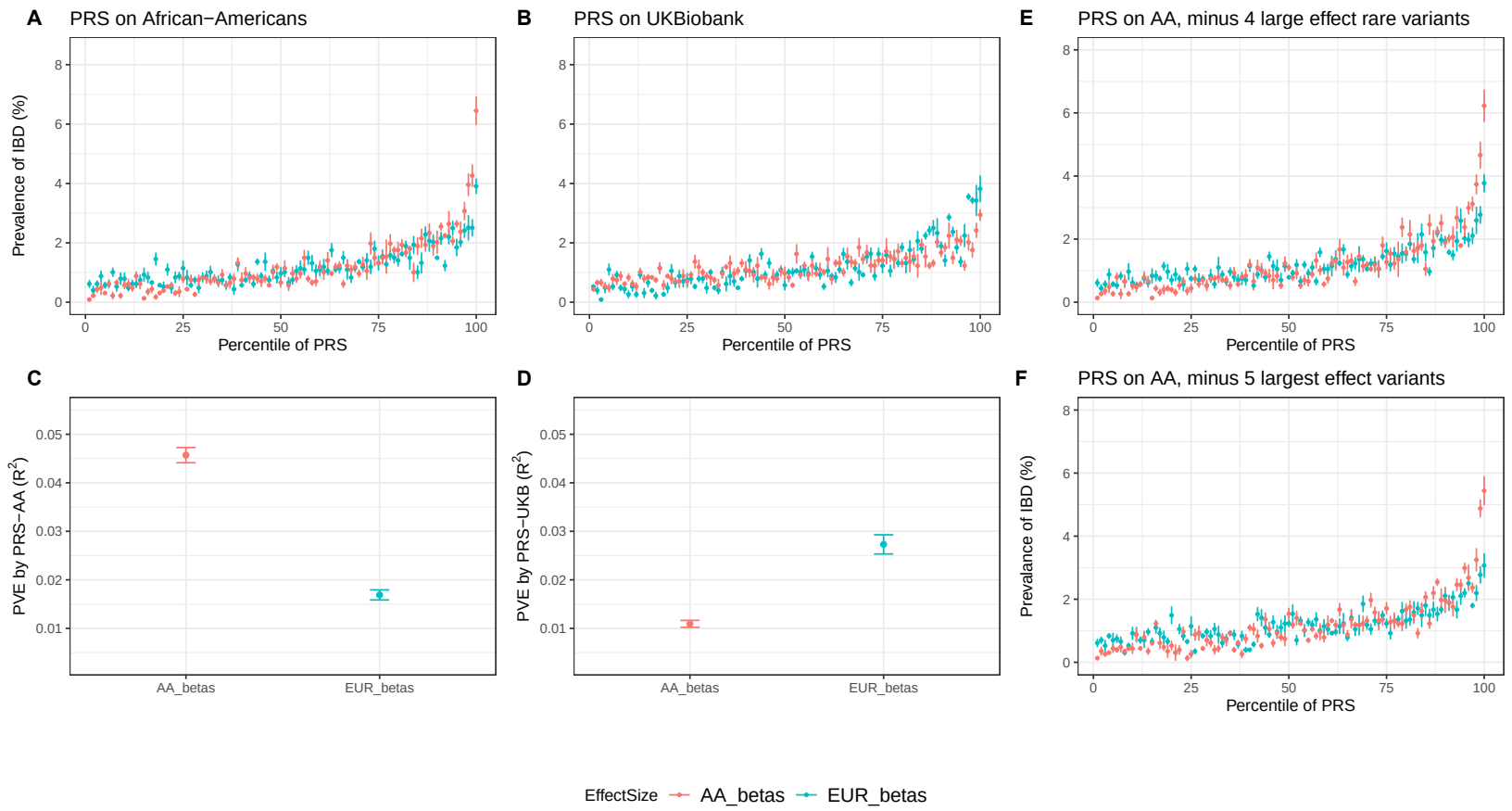

Figure S8. Polygenic risk scores (PRS) in African American pseudo-controls cohort and UK Biobank derived using 215 of the known disease SNPs and their effect sizes estimated on lnOR scale directly from the discovery cohort (AA) or European meta-analysis.

(A, B) prevalence of IBD (%) vs percentile of PRS in each cohort with standard error bars on the prevalence from 5 test sets each involving 532 cases and 45,000 controls. By contrast with Figure 6 and Figure S7, the odds ratios are the same in each of the 5 test sets: inclusion of all 1774 cases provides better fit, but does not allow comparison of the effect of using the direct odds ratios. **(C, D)** show the mean proportion of variance explained (PVE) by each PRS with error bars from 5 test sets. AA effect estimates are more than 2-fold greater than UKB, and PVE with UKB allele frequencies are also more than 2-fold greater than corresponding AA estimates. **(E, F)** Prevalence-risk curves generated with the same weights as in (A, B), but after removal of 4 SNPs with small MAF and large effects (**E**: rs17229679, chr2, OR 1.051, MAF 0.006; rs2066844, chr16, OR 1.276, MAF 0.014; rs11064881, chr12, OR 1.222, MAF 0.016; rs61839660, chr10, OR 1.241, MAF 0.019), or of 5 SNPs with large absolute value of effects (**F**: rs11581607, chr1, OR 0.578, MAF 0.012; rs12199775, chr6, OR 0.638, MAF 0.009; rs11742570, chr5, OR 0.772, MAF 0.362; rs2066844, chr16, OR 1.276, MAF 0.014; rs3764147, chr13, OR 1.316, MAF 0.322).