Supplemental Information: Modeling heterogeneity in the genetic architecture of ethnically diverse groups using random effect interaction models

Yogasudha Veturi[*†]**, Gustavo de los Campos**[‡§**]**, Nengjun Yi**[†]**, Wen Huang**[††]**, Ana I. Vazquez**[†‡] **and Brigitte Kühnel**[‡‡]

[*]Department of Genetics, University of Pennsylvania, Philadelphia, PA, 19104, USA

[†]Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, 35205, USA

[‡]Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, 48824, USA

[§]Institute for Quantitative Health Science and Engineering, Michigan State University, East Lansing, MI, 48824, USA

[**]Department of Statistics and Probability, Michigan State University, East Lansing, MI, 48824, USA

[††]Department of Animal Science, Michigan State University, East Lansing, MI, 48824, USA

[‡‡]Department of Molecular Epidemiology, Helmholtz Zentrum München, 85764, Germany

Corresponding author: Yogasudha Veturi, Department of Genetics, University of Pennsylvania, 415 Curie Blvd, Philadelphia, PA 19104 yveturi@upenn.edu

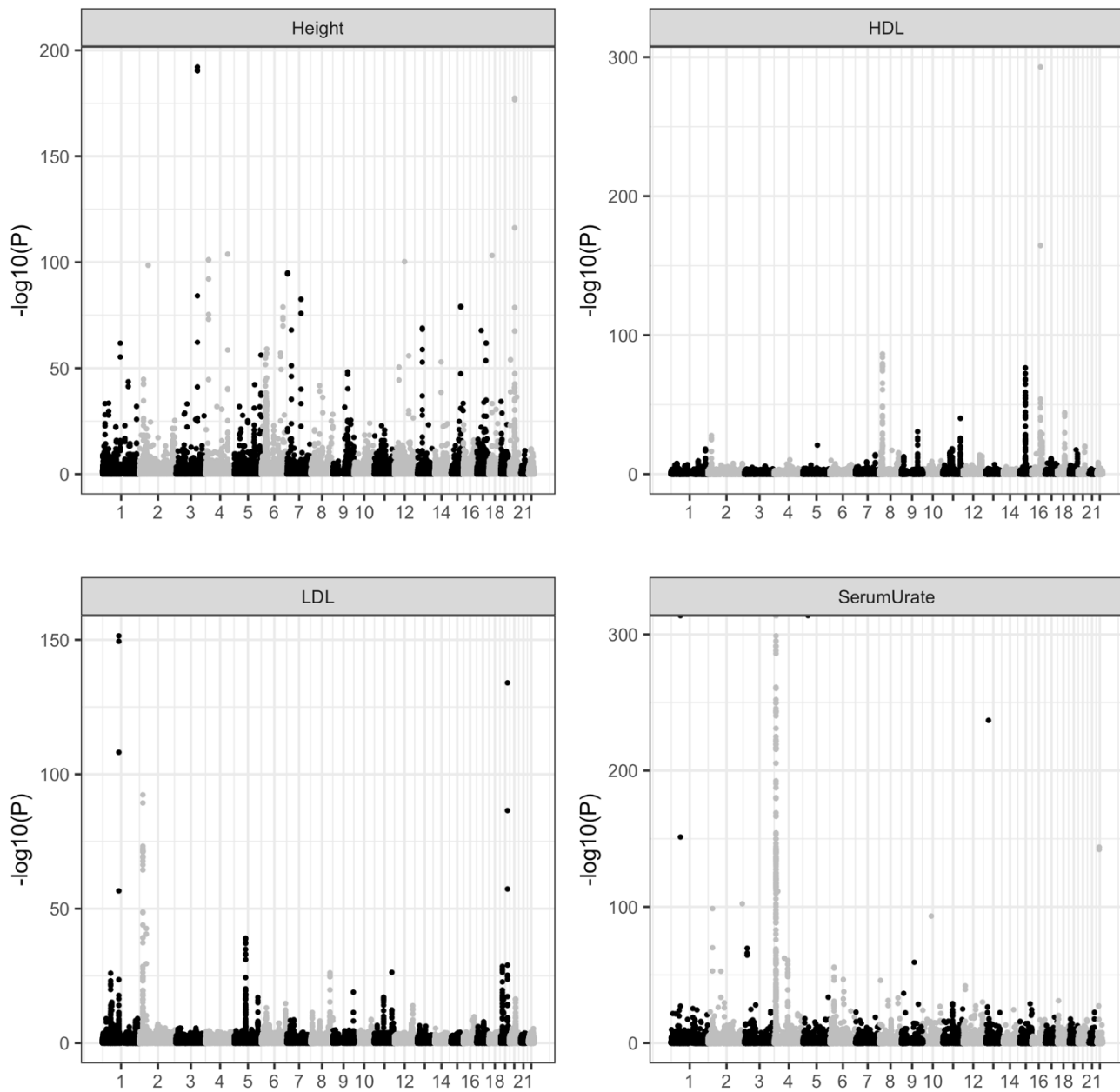# Figures
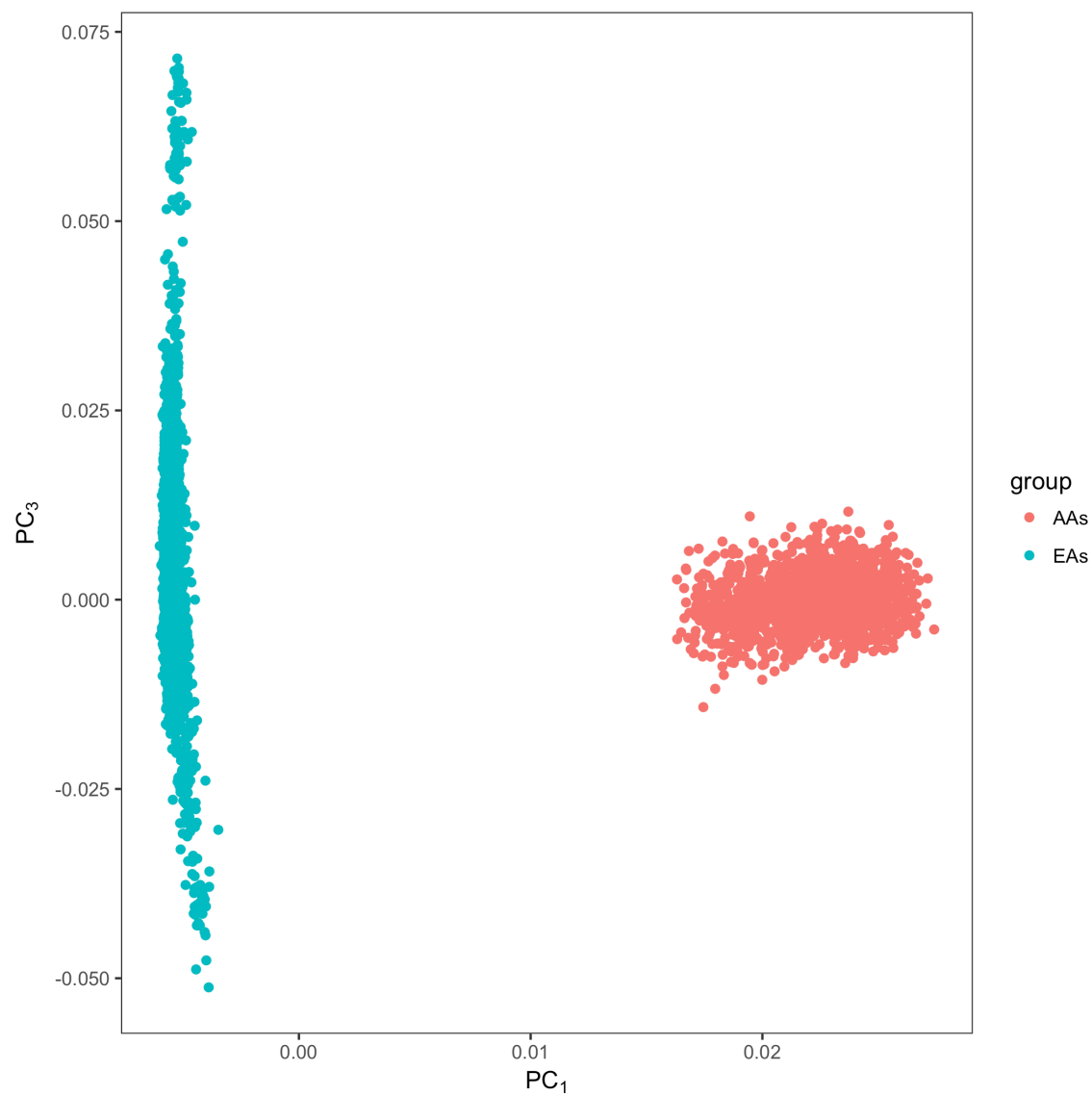
Figure S1. Evidence of clustering among individuals from the ARIC study. Principal component 3 (y-axis) is plotted against principal component 1 (x-axis) for EAs (blue) and AAs (orange), after standard quality control procedures as explained in the manuscript.
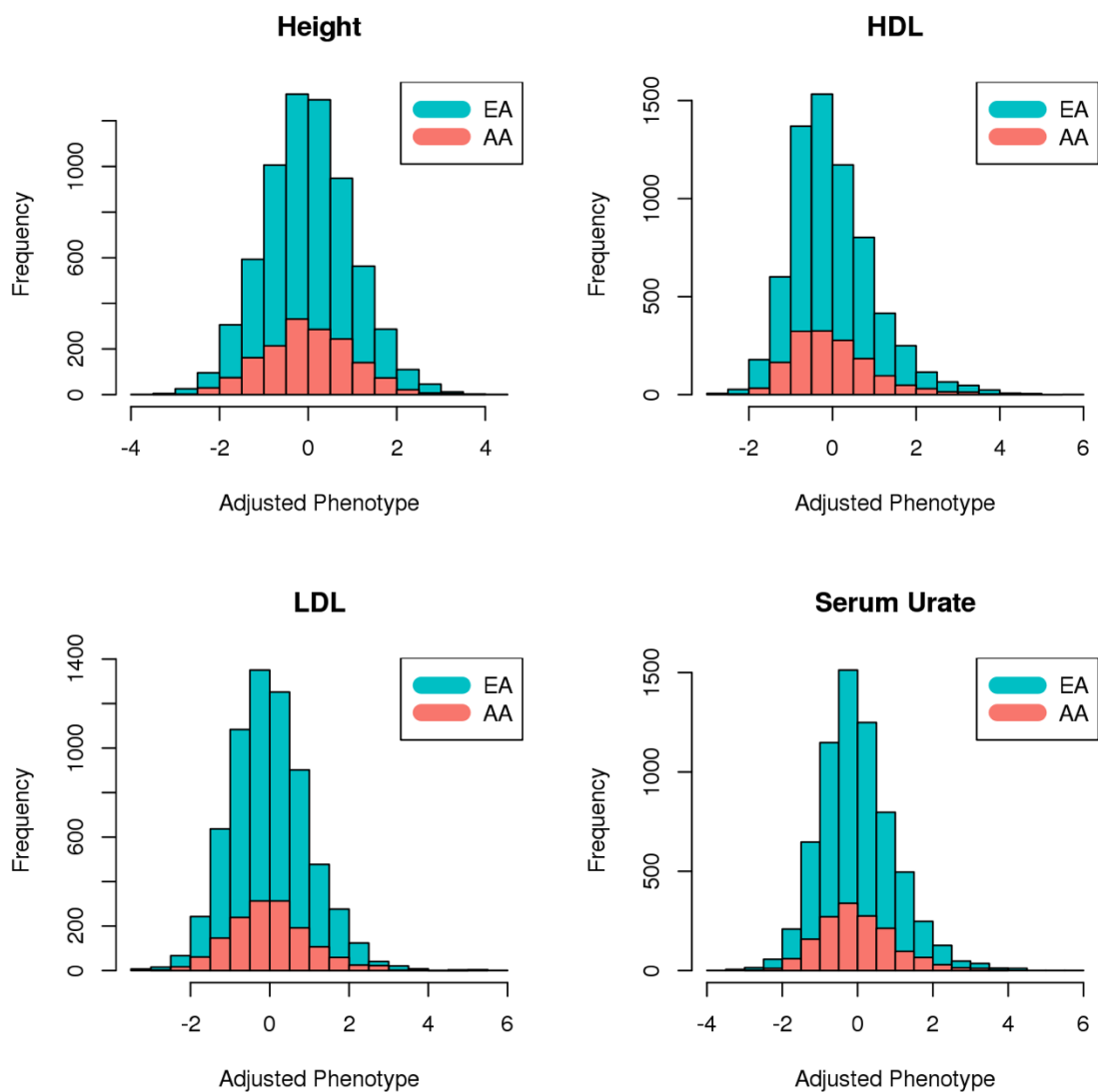
Figure S2. Histograms of adjusted phenotypes for EAs and AAs. In each plot, frequency distributions of (sex-age-ethnic group-adjusted) adjusted phenotypes for EAs are represented in green and those for AAs are represented in orange. Phenotypes are adjusted by age and gender.

Figure S3. Manhattan plots for the full SNP set. Plots of the post-QC SNP set of 828,822 SNPs are shown by trait, where $-\log_{10}$(p-value) (y-axis) is plotted against chromosome number (x-axis). The p-values for the four traits: height (cm), HDL (mmol/L), LDL (mmol/L), and serum urate (mg/dL) were obtained from GWAS conducted on the UK Biobank cohort and previously published GWAS (after excluding the ARIC dataset). The Manhattan plots are truncated at a $-\log_{10}$(p-value) of 60 for the sake of presentation.
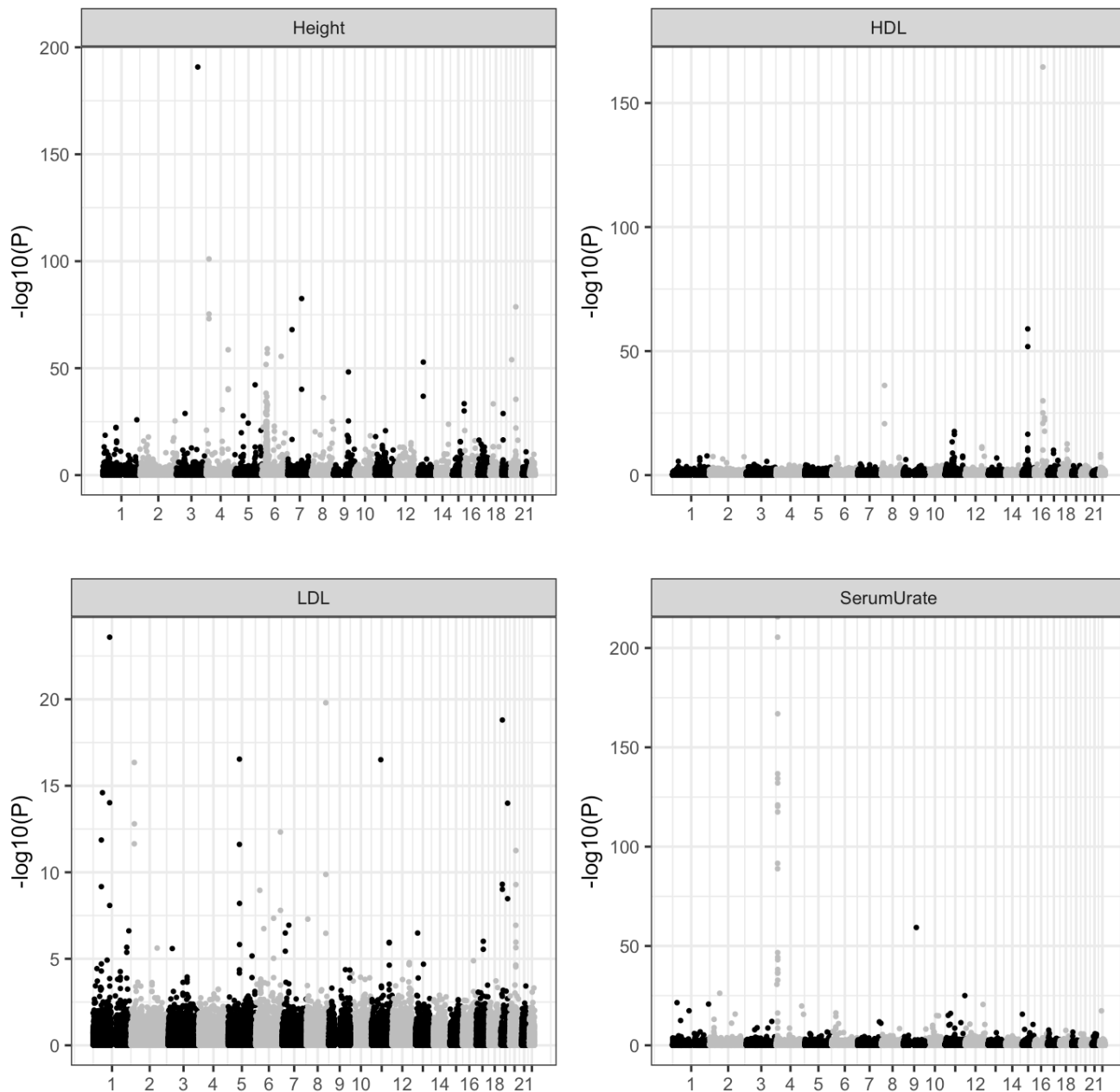
Figure S4. Manhattan plots for a reduced SNP set. Plots of SNPs selected from windows constructed at a –log₁₀(p-value) cutoff of 1 are shown by trait, where –log₁₀(p-value) (y-axis) is plotted against chromosome number (x-axis). The p-values for the four traits: height (cm), HDL (mmol/L), LDL (mmol/L), and serum urate (mg/dL) were obtained from GWAS conducted on the UK Biobank cohort and previously published GWAS (after excluding the ARIC dataset). The Manhattan plots were truncated at a –log₁₀(p-value) of 60 for the sake of presentation.

5

Figure S5. Average estimates of proportion of variance explained by a SNP-set obtained in the second simulation scenario, by prior and number of SNPs used. The simulated heritability was 0.5, bars represent the average estimates over 200 Monte Carlo replicates and the vertical lines gives +/- standard errors. Results for the 1ˢᵗ simulation scenario are presented in Figure 1.
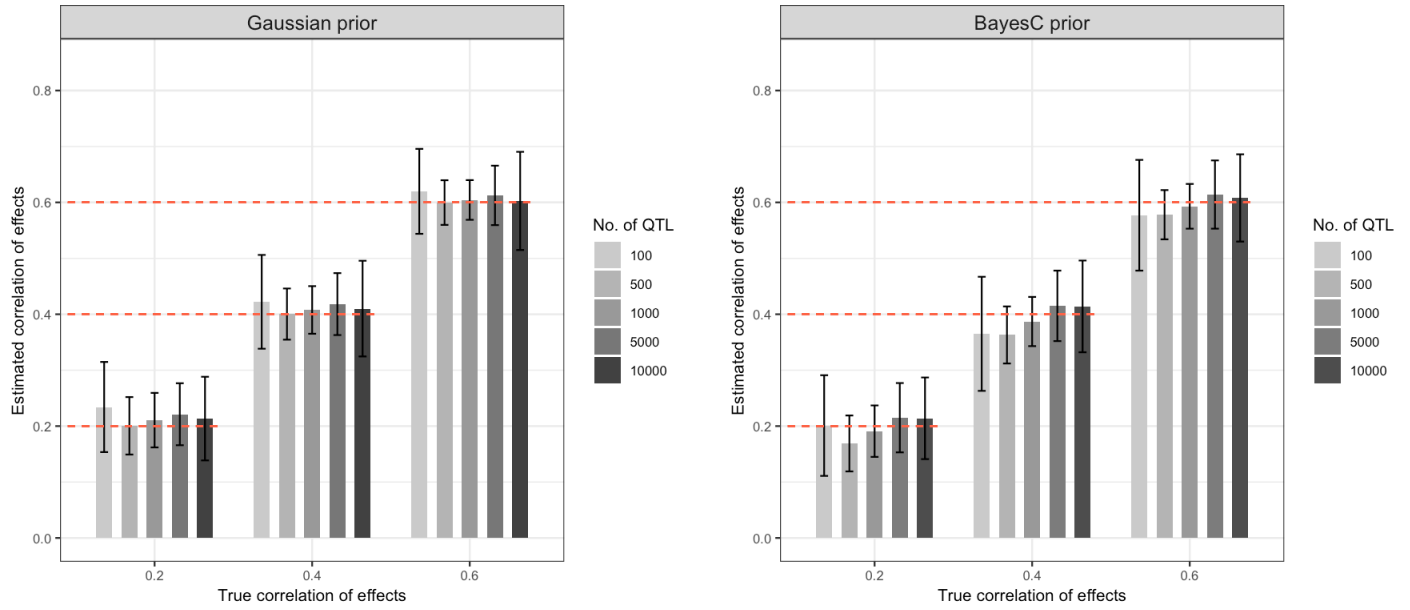
Figure S6. **Average estimates of the correlation of effects in the second simulation scenario by prior and number of SNPs used.** The simulated heritability was 0.5; bars represent the average estimates over 200 Monte Carlo replicates and the vertical lines gives +/- standard errors. Results for the 1st simulation scenario are presented in Figure 2.
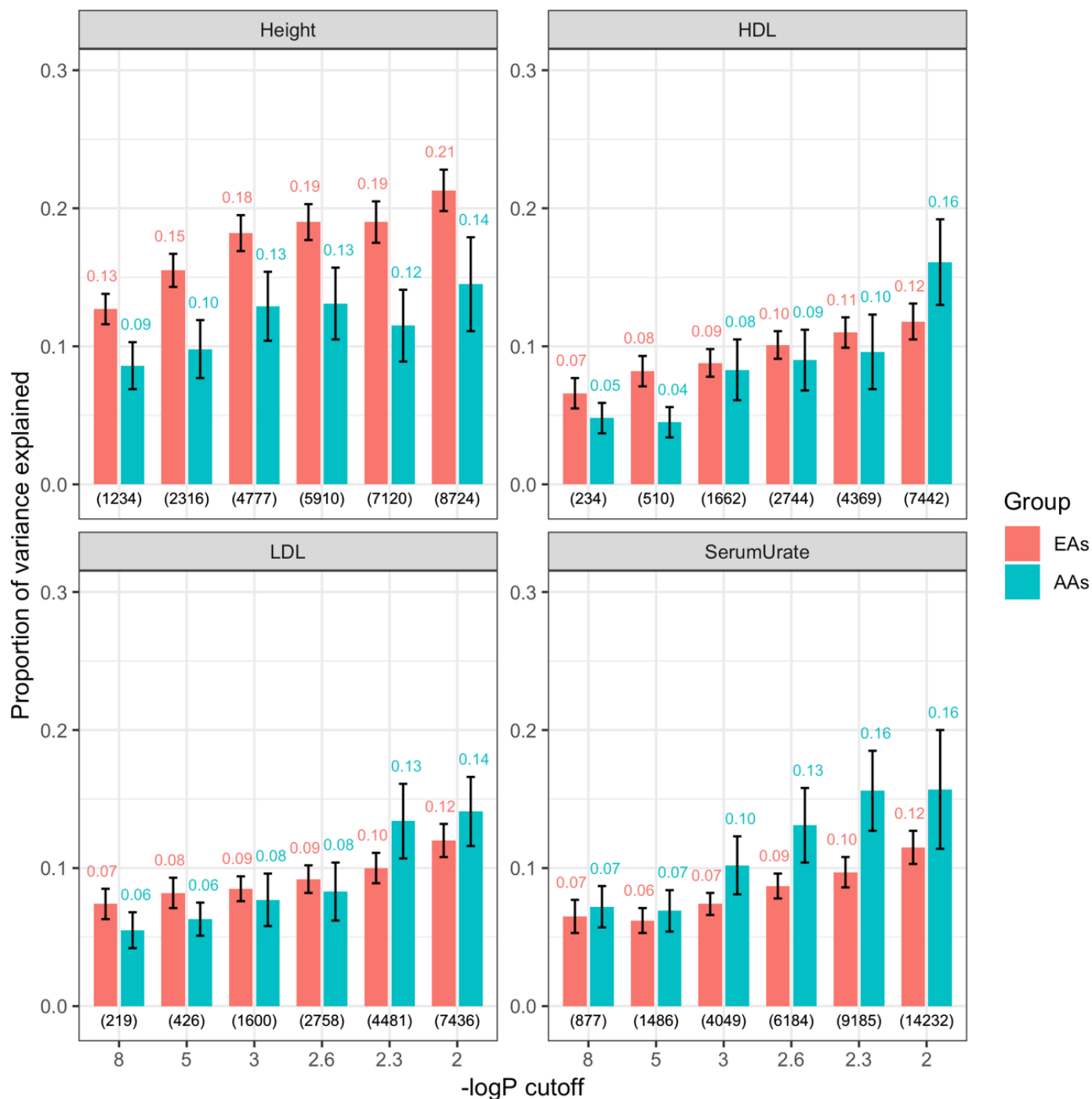
Figure S7. Proportion of variance explained by subsets of SNPs obtained with the Gaussian-interaction model, by trait, ethnicity and SNP set. Estimated (median) proportion of variance explained (y-axis) is plotted by trait, ethnicity and $\log_{10}(p$-value) cutoff used to choose markers from GWAS consortia (excluding ARIC). Numerals above the bars indicate the proportion of variance explained by either ethnic group and the corresponding number of SNPs used for model fitting (in parentheses at the bottom). Vertical lines give estimates of +/- posterior standard deviation.
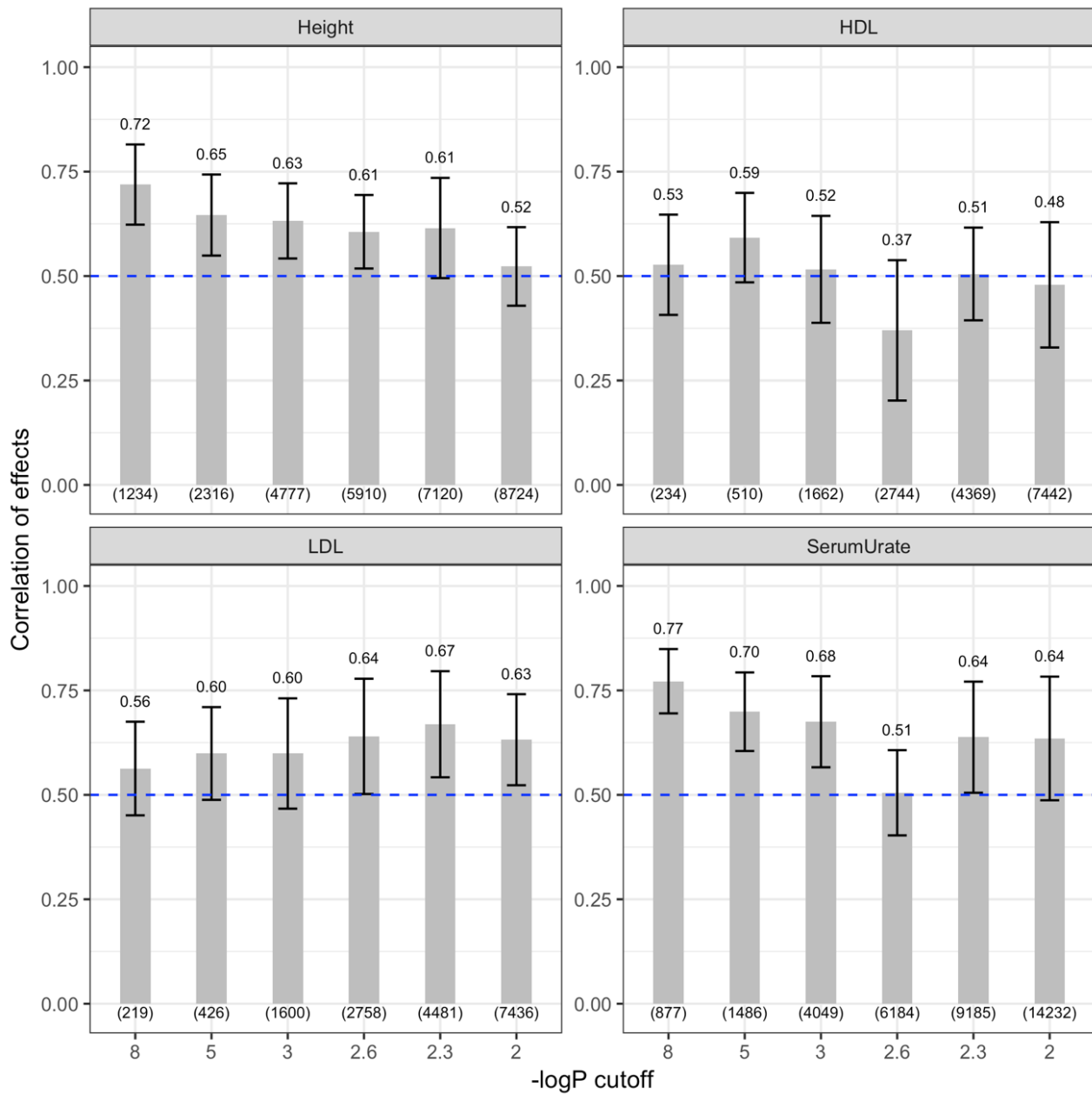
Figure S8. Estimated correlation of effects between African Americans (AAs) and European Americans (EAs) obtained with the Gaussian-interaction model, by trait and SNP set. Estimated correlation of effects between AAs and EAs (y-axis) is plotted by trait using markers selected from GWAS consortia (excluding ARIC). Numerals above the bars indicate the median correlation of effects and the corresponding number of SNPs used for model fitting (in parentheses at the bottom). Vertical lines give estimates of +/- posterior standard deviation.
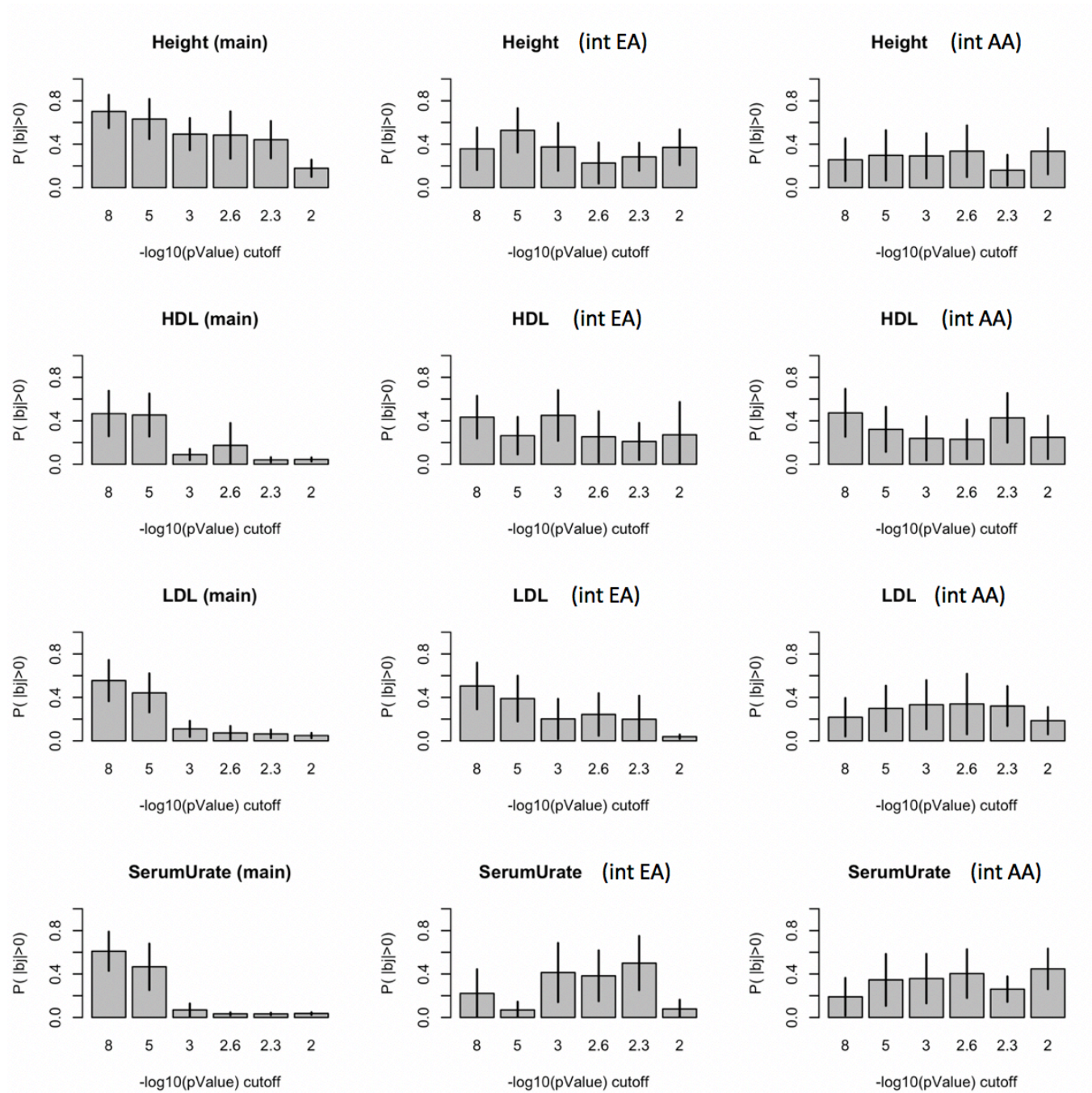
Figure S9. Estimated proportion of non-zero effects between EAs and AAs by trait and SNP set for main effects and interaction effects obtained using Bayes-C interaction model. Estimated proportion of non-zero effects between AAs and EAs (y-axis) is plotted by trait using markers selected from GWAS consortia (excluding ARIC) and 6 different $-\log_{10}(p$-value) cutoffs. The first column corresponds to the main effects whereas the second and third correspond to interaction effects for EAs and AAs, respectively. In each plot, the vertical lines give estimates of +/- posterior standard deviation.

Figure S10. Estimated correlation of effects between African Americans (AAs) and European Americans (EAs) obtained with the BayesC-interaction model, by trait, ethnicity and *randomly chosen* SNP set. Estimated (median) proportion of variance explained (y-axis) is plotted by trait, ethnicity and number of randomly chosen SNPs (x-axis). Vertical lines represent estimates of +/- posterior standard deviation. Numerals above the bars indicate the proportion of variance explained by either ethnic group. Vertical lines give estimates of +/- posterior standard deviation.
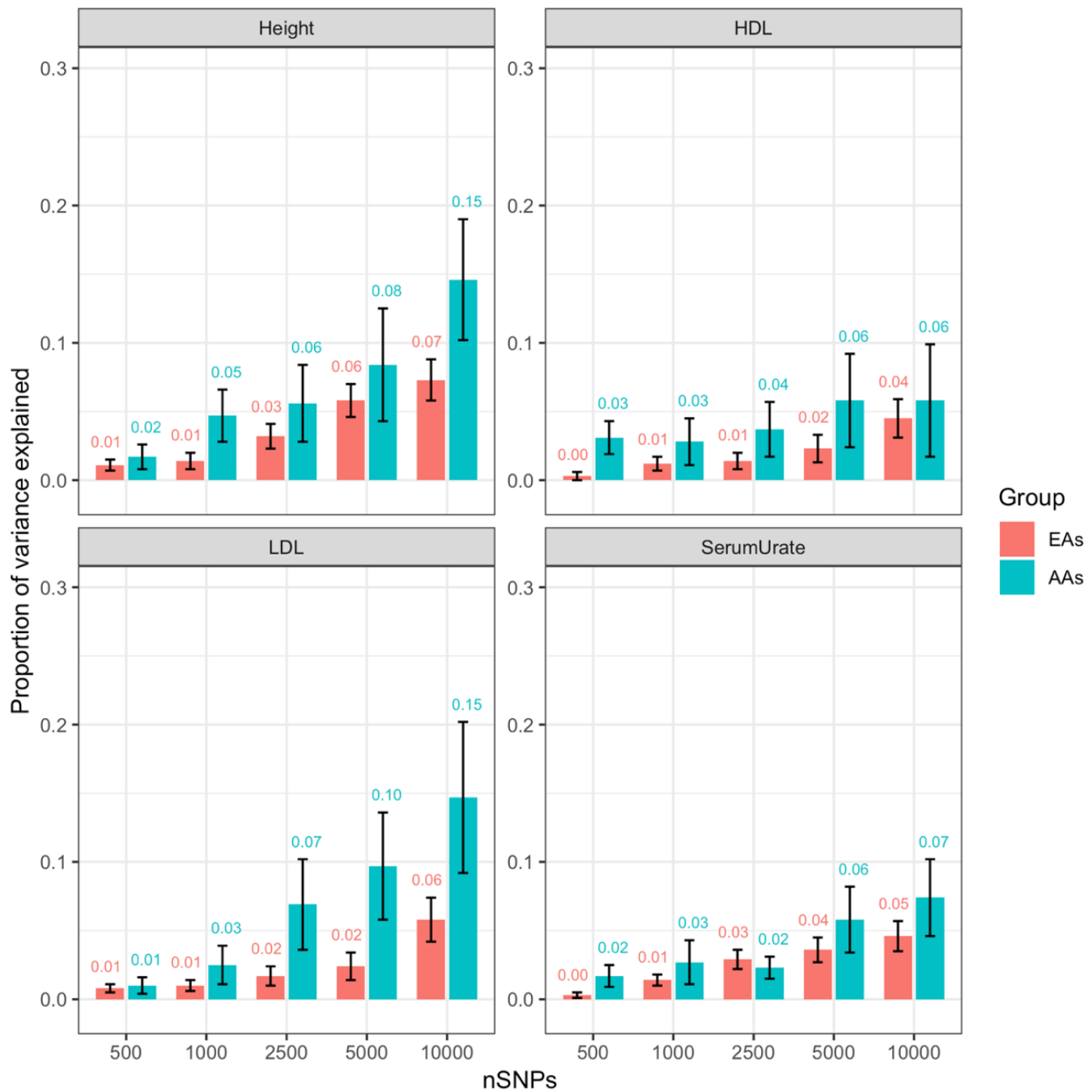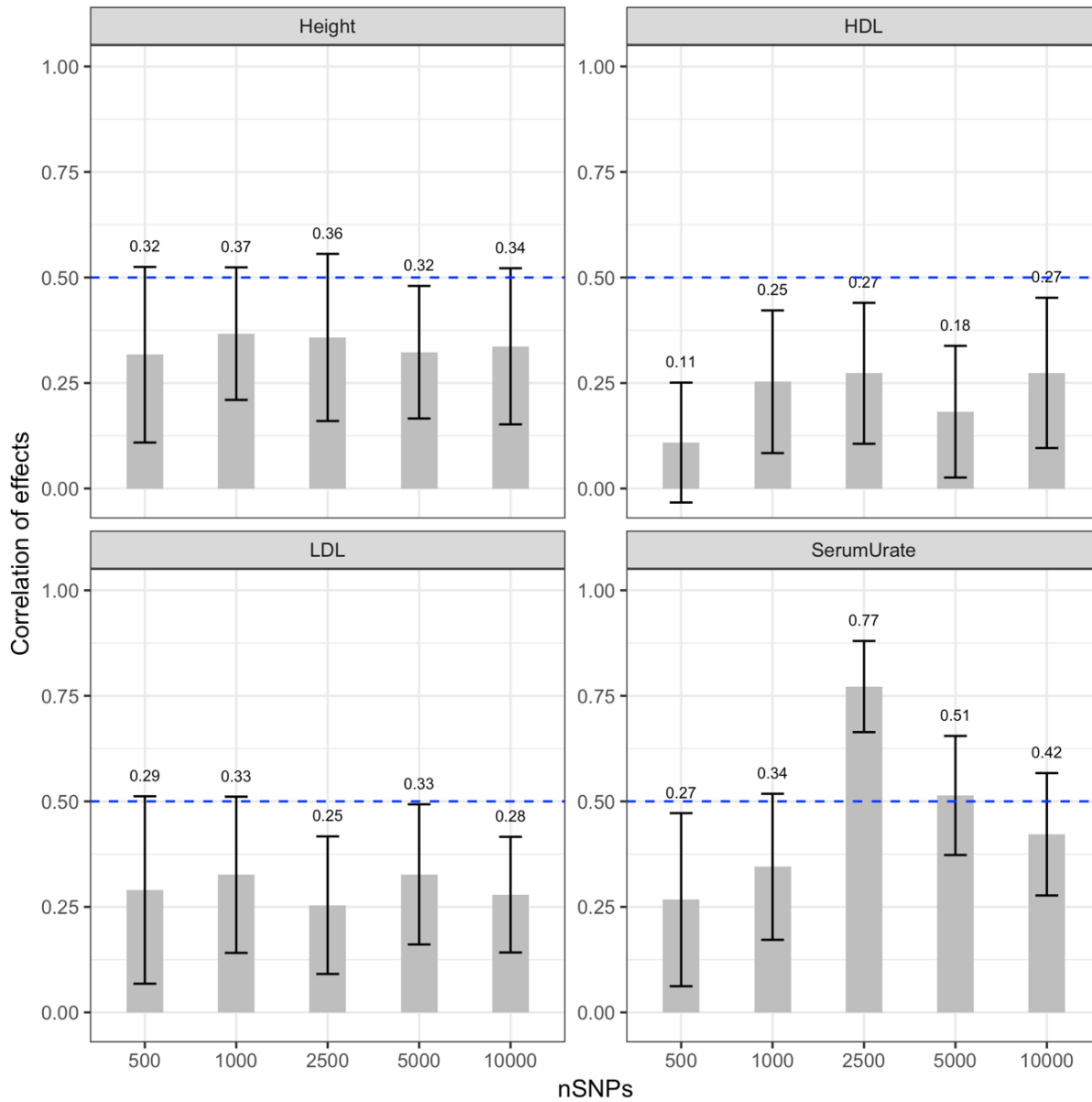
Figure S11. Estimated correlation of effects between African Americans (AAs) and European Americans (EAs) obtained with the BayesC-interaction model, by trait and *randomly chosen* SNP set. Estimated (median) correlation of effects between AAs and EAs (y-axis) is plotted by trait at different numbers of randomly chosen SNPs (x-axis). Vertical lines represent estimates of +/- posterior standard deviation. In each plot, the numerals above the bars indicate the median correlation of effects.

Figure S12. Proportion of variance explained by subsets of SNPs obtained with the BayesC-interaction model, by trait and SNP set and ethnic-group label randomly permuted within EAs. Estimated (median) proportion of variance explained (y-axis) is plotted by trait, permuted EA group label and $\log_{10}(p$-value) cutoff used to choose markers from GWAS consortia (excluding ARIC). Numerals above the bars indicate the proportion of variance explained by either ethnic group and the corresponding number of SNPs used for model fitting (in parentheses at the bottom). Vertical lines give estimates of +/- posterior standard deviation.
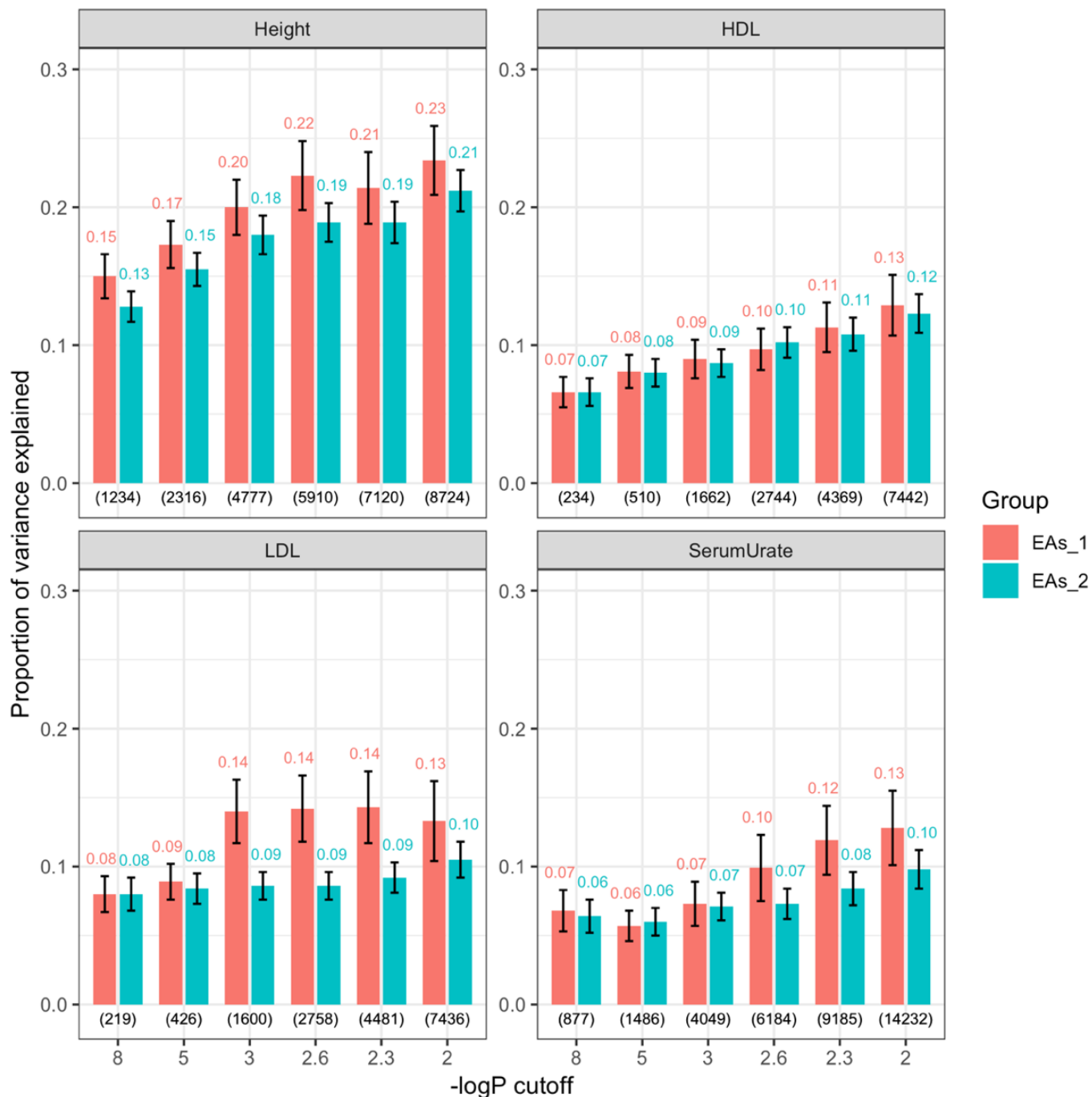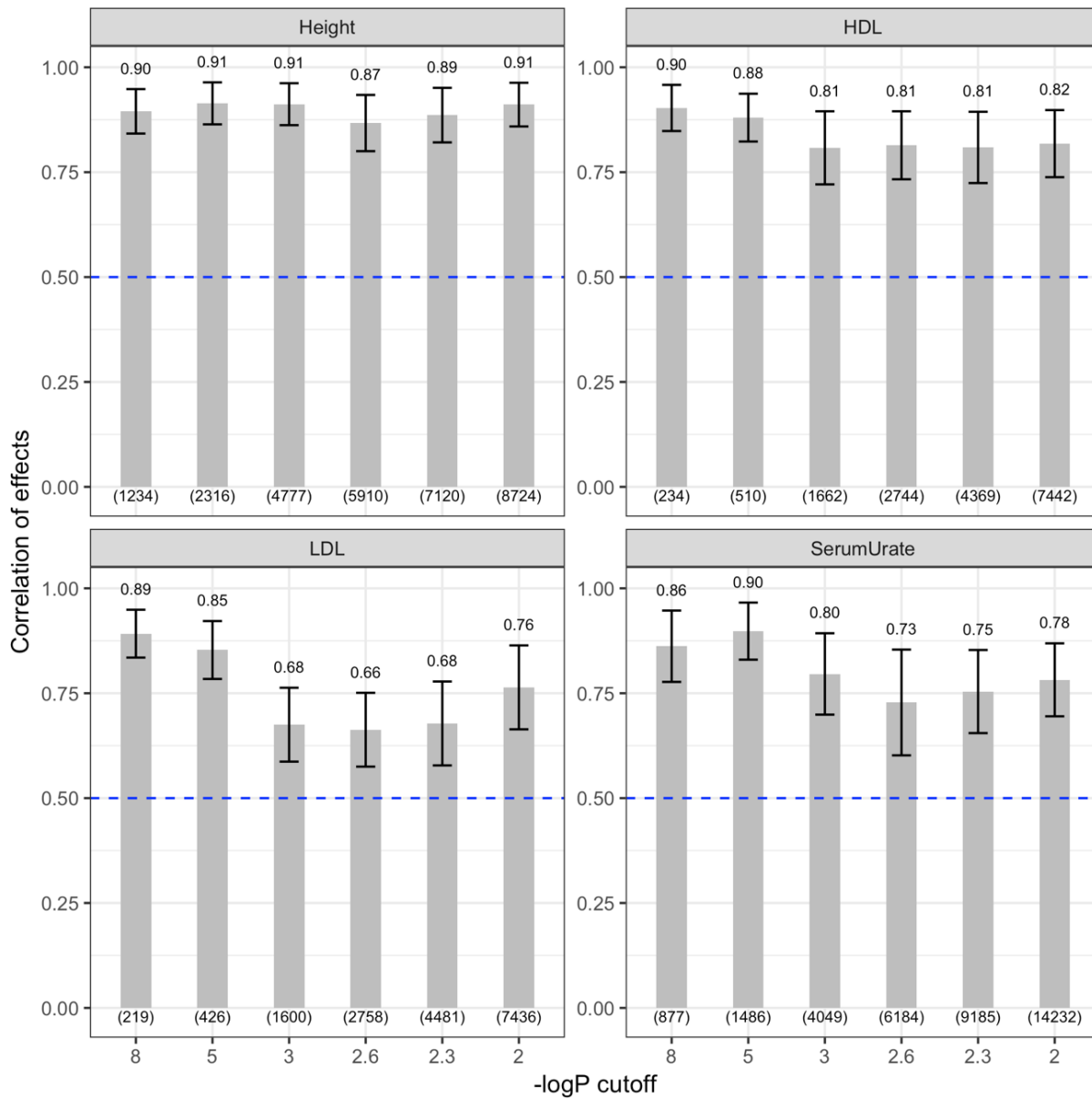
Figure S13. Estimated correlation of effects between African Americans (AAs) and European Americans (EAs) obtained with the BayesC-interaction model, by trait, SNP set and ethnic-group label randomly permuted within EAs. Estimated correlation of effects between AAs and EAs (y-axis) is plotted by trait using markers selected from GWAS consortia (excluding ARIC). Numerals above the bars indicate the median correlation of effects and the corresponding number of SNPs used for model fitting (in parentheses at the bottom). Vertical lines give estimates of +/- posterior standard deviation.

Tables

Table S1. Numbers of windows (in gray) and the corresponding numbers of SNPs chosen for model fitting (in black) for the four traits: Height (cm), HDL (mmol/L), LDL (mmol/L), and serum urate (mg/dL) at different -log$_{10}$(p-value) cutoffs. The p-values for each trait were obtained from previously published GWAS.

-log$_{10}$(p-value)	Height	HDL	LDL	Serum urate
8	259	153	152	708
	1,234	234	219	877
5	649	333	299	1202
	2,316	510	426	1,486
3	1983	1,185	1,151	3,249
	4,777	1,662	1,600	4,049
2.6	2795	1,992	2,036	4,957
	5,910	2,744	2,758	6,184
2.3	3806	3,209	3,324	7,267
	7,120	4,369	4,481	9,185
2	5306	5,510	5,526	11,111
	8,724	7,442	7,436	14,232

Table S2. Estimated proportion of variance explained for EAs and AAs from simulation settings 1 and 2 using BayesC- and Gaussian-interaction models where the true heritability for both groups are: (1) 0.2 for EAs and AAs (2) 0.5 for EAs and 0.2 for AAs (3) 0.2 for EAs and 0.5 for AAs. In parentheses are posterior standard deviations across 50 MC simulations.

Setting	nSNPs	rho	BayesC h2.1 (EA) 0.2	0.5	0.2	BayesC h2.2 (AA) 0.2	0.2	0.5	Gaussian h2.1 (EA) 0.2	0.5	0.2	Gaussian h2.2 (AA) 0.2	0.2	0.5
1	100	0.2	0.206	0.509	0.207	0.205	0.209	0.504	0.21	0.514	0.209	0.21	0.214	0.513
1			(0.009)	(0.01)	(0.011)	(0.018)	(0.02)	(0.018)	(0.01)	(0.009)	(0.01)	(0.018)	(0.02)	(0.016)
1	500	0.2	0.202	0.502	0.203	0.2	0.203	0.5	0.203	0.503	0.204	0.205	0.208	0.503
1			(0.009)	(0.009)	(0.009)	(0.021)	(0.026)	(0.017)	(0.009)	(0.009)	(0.009)	(0.022)	(0.026)	(0.017)
1	1000	0.2	0.203	0.5	0.202	0.2	0.197	0.501	0.204	0.501	0.203	0.206	0.203	0.505
1			(0.011)	(0.009)	(0.012)	(0.031)	(0.034)	(0.023)	(0.011)	(0.009)	(0.012)	(0.031)	(0.034)	(0.023)
1	5000	0.2	0.201	0.501	0.201	0.194	0.204	0.49	0.203	0.502	0.203	0.199	0.208	0.499
1			(0.015)	(0.012)	(0.016)	(0.044)	(0.045)	(0.048)	(0.016)	(0.012)	(0.016)	(0.049)	(0.051)	(0.048)
1	10000	0.2	0.2	0.5	0.201	0.207	0.224	0.485	0.203	0.501	0.204	0.215	0.224	0.498
1			(0.024)	(0.018)	(0.023)	(0.061)	(0.064)	(0.075)	(0.023)	(0.017)	(0.022)	(0.068)	(0.064)	(0.075)
1	100	0.4	0.206	0.504	0.207	0.205	0.21	0.498	0.209	0.511	0.21	0.21	0.214	0.508
1			(0.01)	(0.01)	(0.01)	(0.018)	(0.021)	(0.018)	(0.01)	(0.01)	(0.01)	(0.018)	(0.021)	(0.016)
1	500	0.4	0.202	0.502	0.203	0.2	0.206	0.5	0.203	0.503	0.203	0.205	0.21	0.503
1			(0.009)	(0.009)	(0.009)	(0.021)	(0.024)	(0.017)	(0.009)	(0.009)	(0.009)	(0.022)	(0.024)	(0.017)
1	1000	0.4	0.204	0.5	0.203	0.2	0.202	0.501	0.204	0.501	0.204	0.206	0.206	0.504
1			(0.011)	(0.009)	(0.011)	(0.03)	(0.03)	(0.022)	(0.011)	(0.009)	(0.011)	(0.03)	(0.032)	(0.022)
1	5000	0.4	0.201	0.501	0.202	0.201	0.218	0.489	0.203	0.502	0.204	0.205	0.222	0.497
1			(0.015)	(0.012)	(0.014)	(0.043)	(0.039)	(0.046)	(0.016)	(0.012)	(0.015)	(0.045)	(0.042)	(0.047)
1	10000	0.4	0.199	0.499	0.203	0.219	0.243	0.486	0.203	0.501	0.205	0.223	0.243	0.496
1			(0.024)	(0.018)	(0.022)	(0.059)	(0.053)	(0.075)	(0.023)	(0.018)	(0.021)	(0.064)	(0.058)	(0.075)
1	100	0.6	0.205	0.491	0.207	0.204	0.214	0.48	0.209	0.498	0.213	0.21	0.22	0.489
1			(0.01)	(0.012)	(0.01)	(0.018)	(0.021)	(0.019)	(0.01)	(0.012)	(0.009)	(0.018)	(0.02)	(0.016)
1	500	0.6	0.202	0.499	0.204	0.2	0.218	0.492	0.202	0.5	0.205	0.205	0.222	0.495
1			(0.009)	(0.009)	(0.008)	(0.021)	(0.019)	(0.015)	(0.009)	(0.009)	(0.008)	(0.022)	(0.019)	(0.015)
1	1000	0.6	0.203	0.499	0.206	0.203	0.22	0.495	0.204	0.499	0.207	0.207	0.223	0.498
1			(0.01)	(0.009)	(0.01)	(0.027)	(0.024)	(0.021)	(0.01)	(0.009)	(0.01)	(0.027)	(0.024)	(0.021)
1	5000	0.6	0.202	0.501	0.209	0.213	0.242	0.486	0.203	0.502	0.21	0.214	0.244	0.494
1			(0.015)	(0.012)	(0.013)	(0.034)	(0.028)	(0.046)	(0.015)	(0.012)	(0.014)	(0.041)	(0.032)	(0.046)
1	10000	0.6	0.201	0.499	0.21	0.225	0.263	0.485	0.204	0.501	0.212	0.235	0.272	0.494
1			(0.022)	(0.018)	(0.019)	(0.054)	(0.046)	(0.073)	(0.022)	(0.018)	(0.019)	(0.058)	(0.049)	(0.075)
1	100	0.8	0.203	0.463	0.21	0.203	0.227	0.439	0.208	0.47	0.221	0.21	0.238	0.446
1			(0.009)	(0.012)	(0.009)	(0.016)	(0.02)	(0.017)	(0.009)	(0.011)	(0.01)	(0.014)	(0.02)	(0.016)
1	500	0.8	0.201	0.491	0.214	0.203	0.25	0.459	0.202	0.491	0.217	0.207	0.254	0.462

1			(0.009)	(0.009)	(0.008)	(0.017)	(0.016)	(0.015)	(0.009)	(0.009)	(0.009)	(0.019)	(0.016)	(0.015)
1	1000	0.8	0.203	0.495	0.22	0.21	0.257	0.469	0.205	0.495	0.222	0.213	0.259	0.473
1			(0.01)	(0.009)	(0.009)	(0.022)	(0.019)	(0.021)	(0.01)	(0.009)	(0.009)	(0.022)	(0.018)	(0.022)
1	5000	0.8	0.203	0.5	0.223	0.225	0.278	0.479	0.205	0.501	0.225	0.228	0.281	0.486
1			(0.014)	(0.012)	(0.013)	(0.032)	(0.023)	(0.047)	(0.015)	(0.012)	(0.013)	(0.035)	(0.024)	(0.045)
1	10000	0.8	0.203	0.498	0.223	0.24	0.3	0.484	0.206	0.5	0.225	0.247	0.306	0.491
1			(0.021)	(0.018)	(0.018)	(0.046)	(0.036)	(0.07)	(0.021)	(0.018)	(0.018)	(0.054)	(0.043)	(0.075)
2	100	0.2	0.201	0.502	0.203	0.203	0.21	0.496	0.208	0.514	0.207	0.209	0.213	0.511
2			(0.01)	(0.01)	(0.011)	(0.019)	(0.022)	(0.019)	(0.01)	(0.008)	(0.01)	(0.019)	(0.022)	(0.018)
2	500	0.2	0.201	0.5	0.201	0.201	0.205	0.501	0.202	0.503	0.202	0.205	0.207	0.504
2			(0.01)	(0.009)	(0.01)	(0.024)	(0.025)	(0.018)	(0.011)	(0.009)	(0.011)	(0.024)	(0.025)	(0.019)
2	1000	0.2	0.203	0.5	0.202	0.201	0.195	0.503	0.203	0.501	0.202	0.205	0.199	0.505
2			(0.009)	(0.009)	(0.01)	(0.03)	(0.032)	(0.023)	(0.009)	(0.009)	(0.01)	(0.031)	(0.033)	(0.023)
2	5000	0.2	0.2	0.502	0.2	0.201	0.205	0.496	0.202	0.503	0.202	0.208	0.21	0.504
2			(0.015)	(0.011)	(0.015)	(0.049)	(0.04)	(0.05)	(0.016)	(0.012)	(0.015)	(0.052)	(0.041)	(0.051)
2	10000	0.2	0.198	0.499	0.199	0.216	0.216	0.493	0.201	0.502	0.201	0.22	0.221	0.504
2			(0.02)	(0.019)	(0.019)	(0.066)	(0.064)	(0.086)	(0.019)	(0.018)	(0.019)	(0.072)	(0.055)	(0.087)
2	100	0.4	0.201	0.492	0.206	0.204	0.214	0.488	0.208	0.509	0.208	0.208	0.214	0.507
2			(0.011)	(0.011)	(0.011)	(0.02)	(0.023)	(0.02)	(0.01)	(0.009)	(0.01)	(0.018)	(0.021)	(0.017)
2	500	0.4	0.2	0.498	0.201	0.202	0.21	0.498	0.202	0.502	0.202	0.205	0.208	0.503
2			(0.01)	(0.009)	(0.01)	(0.024)	(0.025)	(0.017)	(0.011)	(0.009)	(0.011)	(0.025)	(0.023)	(0.018)
2	1000	0.4	0.203	0.5	0.203	0.201	0.2	0.501	0.204	0.501	0.203	0.205	0.203	0.504
2			(0.009)	(0.009)	(0.009)	(0.029)	(0.029)	(0.023)	(0.009)	(0.009)	(0.009)	(0.031)	(0.028)	(0.023)
2	5000	0.4	0.2	0.503	0.201	0.209	0.216	0.497	0.202	0.503	0.203	0.213	0.219	0.503
2			(0.015)	(0.011)	(0.014)	(0.048)	(0.034)	(0.049)	(0.016)	(0.012)	(0.015)	(0.048)	(0.036)	(0.05)
2	10000	0.4	0.199	0.499	0.201	0.225	0.238	0.496	0.201	0.5	0.203	0.228	0.234	0.504
2			(0.019)	(0.02)	(0.019)	(0.061)	(0.049)	(0.087)	(0.019)	(0.019)	(0.019)	(0.067)	(0.054)	(0.087)
2	100	0.6	0.202	0.477	0.209	0.205	0.226	0.47	0.208	0.495	0.211	0.208	0.222	0.486
2			(0.011)	(0.012)	(0.012)	(0.019)	(0.023)	(0.021)	(0.01)	(0.013)	(0.01)	(0.018)	(0.019)	(0.021)
2	500	0.6	0.2	0.494	0.203	0.204	0.221	0.489	0.201	0.499	0.204	0.206	0.22	0.495
2			(0.01)	(0.009)	(0.01)	(0.023)	(0.024)	(0.016)	(0.011)	(0.009)	(0.01)	(0.025)	(0.017)	(0.017)
2	1000	0.6	0.204	0.498	0.207	0.203	0.218	0.494	0.204	0.499	0.207	0.207	0.219	0.498
2			(0.009)	(0.009)	(0.008)	(0.027)	(0.023)	(0.022)	(0.009)	(0.009)	(0.008)	(0.028)	(0.022)	(0.022)
2	5000	0.6	0.201	0.502	0.208	0.217	0.242	0.493	0.202	0.502	0.209	0.22	0.241	0.499
2			(0.016)	(0.011)	(0.014)	(0.041)	(0.026)	(0.05)	(0.016)	(0.011)	(0.014)	(0.043)	(0.029)	(0.05)
2	10000	0.6	0.2	0.499	0.208	0.234	0.264	0.497	0.202	0.5	0.21	0.24	0.269	0.502
2			(0.019)	(0.019)	(0.017)	(0.054)	(0.041)	(0.084)	(0.019)	(0.02)	(0.017)	(0.06)	(0.048)	(0.087)

Table S3. Estimated correlation of effects between EAs and AAs from simulation settings 1 and 2 using BayesC- and Gaussian- interaction models, where the true heritability for both groups are: (1) 0.2 for EAs and AAs (2) 0.5 for EAs and 0.2 for AAs (3) 0.2 for EAs and 0.5 for AAs. In parentheses are posterior standard deviations across 50 MC simulations.

Setting	nSNPs	rho	BayesC h2.1(EA)_h2.2 (AA)			Gaussian h2.1(EA)_h2.2 (AA)		
			0.2_0.2	0.2_0.5	0.5_0.2	0.2_0.2	0.2_0.5	0.5_0.2
1	100	0.2	0.226	0.199	0.222	0.248	0.226	0.24
1			(0.083)	(0.12)	(0.092)	(0.061)	(0.09)	(0.068)
1	500	0.2	0.207	0.212	0.194	0.208	0.209	0.194
1			(0.075)	(0.063)	(0.059)	(0.069)	(0.063)	(0.057)
1	1000	0.2	0.223	0.211	0.201	0.223	0.213	0.2
1			(0.08)	(0.066)	(0.062)	(0.078)	(0.065)	(0.06)
1	5000	0.2	0.26	0.211	0.228	0.267	0.224	0.227
1			(0.108)	(0.089)	(0.081)	(0.107)	(0.086)	(0.078)
1	10000	0.2	0.274	0.212	0.238	0.259	0.217	0.233
1			(0.123)	(0.103)	(0.095)	(0.117)	(0.098)	(0.097)
1	100	0.4	0.417	0.389	0.416	0.416	0.384	0.407
1			(0.078)	(0.103)	(0.07)	(0.076)	(0.091)	(0.061)
1	500	0.4	0.4	0.404	0.394	0.393	0.399	0.392
1			(0.074)	(0.057)	(0.053)	(0.073)	(0.057)	(0.054)
1	1000	0.4	0.418	0.402	0.403	0.411	0.402	0.399
1			(0.088)	(0.058)	(0.06)	(0.088)	(0.058)	(0.058)
1	5000	0.4	0.426	0.382	0.417	0.428	0.388	0.412
1			(0.129)	(0.091)	(0.077)	(0.124)	(0.086)	(0.078)
1	10000	0.4	0.398	0.355	0.407	0.393	0.36	0.401
1			(0.153)	(0.115)	(0.098)	(0.138)	(0.108)	(0.102)
1	100	0.6	0.61	0.577	0.605	0.607	0.543	0.557
1			(0.064)	(0.071)	(0.049)	(0.064)	(0.062)	(0.037)
1	500	0.6	0.599	0.573	0.582	0.59	0.563	0.568
1			(0.062)	(0.041)	(0.035)	(0.061)	(0.04)	(0.03)
1	1000	0.6	0.614	0.564	0.584	0.604	0.56	0.572
1			(0.08)	(0.04)	(0.04)	(0.078)	(0.04)	(0.037)
1	5000	0.6	0.585	0.535	0.566	0.595	0.537	0.56
1			(0.102)	(0.068)	(0.055)	(0.11)	(0.072)	(0.052)
1	10000	0.6	0.547	0.507	0.545	0.534	0.498	0.541
1			(0.132)	(0.1)	(0.078)	(0.128)	(0.093)	(0.083)
1	100	0.8	0.803	0.749	0.78	0.791	0.674	0.682
1			(0.041)	(0.039)	(0.028)	(0.037)	(0.043)	(0.028)
1	500	0.8	0.789	0.704	0.724	0.778	0.683	0.68

18

1			(0.046)	(0.03)	(0.02)	(0.046)	(0.032)	(0.021)
1	1000	0.8	0.79	0.69	0.705	0.78	0.68	0.68
1			(0.057)	(0.03)	(0.028)	(0.057)	(0.032)	(0.03)
1	5000	0.8	0.723	0.659	0.657	0.722	0.654	0.647
1			(0.084)	(0.059)	(0.046)	(0.08)	(0.058)	(0.043)
1	10000	0.8	0.664	0.627	0.631	0.653	0.62	0.63
1			(0.107)	(0.084)	(0.063)	(0.105)	(0.074)	(0.072)
2	100	0.2	0.222	0.202	0.218	0.248	0.236	0.24
2			(0.095)	(0.101)	(0.098)	(0.078)	(0.082)	(0.074)
2	500	0.2	0.206	0.18	0.189	0.208	0.203	0.192
2			(0.07)	(0.052)	(0.053)	(0.069)	(0.053)	(0.053)
2	1000	0.2	0.231	0.192	0.2	0.229	0.211	0.203
2			(0.082)	(0.057)	(0.057)	(0.083)	(0.066)	(0.058)
2	5000	0.2	0.272	0.211	0.23	0.267	0.215	0.227
2			(0.112)	(0.081)	(0.081)	(0.098)	(0.076)	(0.076)
2	10000	0.2	0.259	0.217	0.24	0.259	0.212	0.23
2			(0.118)	(0.1)	(0.109)	(0.126)	(0.09)	(0.108)
2	100	0.4	0.41	0.39	0.414	0.416	0.398	0.411
2			(0.094)	(0.09)	(0.088)	(0.091)	(0.084)	(0.079)
2	500	0.4	0.398	0.362	0.386	0.398	0.39	0.391
2			(0.071)	(0.052)	(0.049)	(0.071)	(0.052)	(0.051)
2	1000	0.4	0.424	0.379	0.399	0.419	0.398	0.403
2			(0.084)	(0.053)	(0.053)	(0.087)	(0.055)	(0.055)
2	5000	0.4	0.435	0.38	0.417	0.43	0.384	0.413
2			(0.114)	(0.087)	(0.073)	(0.106)	(0.085)	(0.072)
2	10000	0.4	0.392	0.366	0.398	0.388	0.34	0.395
2			(0.137)	(0.095)	(0.114)	(0.149)	(0.109)	(0.118)
2	100	0.6	0.603	0.579	0.602	0.609	0.556	0.56
2			(0.09)	(0.077)	(0.077)	(0.087)	(0.06)	(0.057)
2	500	0.6	0.597	0.544	0.576	0.596	0.553	0.564
2			(0.067)	(0.045)	(0.038)	(0.069)	(0.039)	(0.031)
2	1000	0.6	0.617	0.548	0.581	0.61	0.556	0.575
2			(0.074)	(0.041)	(0.037)	(0.075)	(0.04)	(0.031)
2	5000	0.6	0.602	0.529	0.565	0.595	0.54	0.557
2			(0.104)	(0.071)	(0.049)	(0.091)	(0.062)	(0.05)
2	10000	0.6	0.537	0.505	0.533	0.525	0.505	0.534
2			(0.138)	(0.088)	(0.097)	(0.142)	(0.088)	(0.1)

APPENDIX A

Variance component estimation using a Metropolis Hastings-within-Gibbs algorithm

The package BGLR can be used to implement the model of Equation 1 using a Gibbs sampler. By default, BGLR assigns scaled-inverse chi-square priors to variance components. When individuals are distantly related, as they are in our case, the scaled-inverse Chi-square prior can have influences on inferences. To overcome this problem, we used as prior a modified Beta distribution, which allows specifying a relative flat prior for variances.

Let $\sigma_.^2$ be a variance parameter (any of the variances entering the model) and K denote an upper-bound on $\sigma_.^2$ so that $\sigma_.^2 \in [0, K]$; therefore, $\tilde{\sigma}_.^2 = \frac{\sigma_.^2}{K} \in [0,1]$. We then assigned $\tilde{\sigma}_.^2$ a Beta prior, that is

$$p(\tilde{\sigma}_.^2) \propto (\tilde{\sigma}_.^2)^{(\alpha_1 - 1)} (1 - \tilde{\sigma}_.^2)^{(\alpha_2 - 1)} 1(0 < \tilde{\sigma}_.^2 < 1)$$

$$(A1)$$

where α_1 and α_2 are prior "shape" hyper-parameters and $1(0 < \tilde{\sigma}_.^2 < 1)$ is an indicator variable that specify the support of the distribution. Parameter values $\alpha_1 = \alpha_2 = 1$ give a uniform distribution on the 0-1 interval for $\tilde{\sigma}_.^2$; more informative distributions can be obtained by increasing the values of these hyper-parameters. In our application we used $\alpha_1 = \alpha_2 = 1.01$ which gives a very weakly informative prior.

According to the Jacobian theorem, the prior distribution of $\sigma_.^2$ is

$$p(\sigma_.^2) = p\left(\tilde{\sigma}_.^2(\sigma_.^2)\right) \left| \frac{d\tilde{\sigma}_.^2(\sigma_.^2)}{d\sigma_.^2} \right| \propto \left(\frac{\sigma_.^2}{K}\right)^{(\alpha_1 - 1)} \left(1 - \frac{\sigma_.^2}{K}\right)^{(\alpha_2 - 1)} 1(0 < \sigma_.^2 < K)$$

$$(A2)$$

When this prior is used, the fully conditional density of the variance parameter does not have a closed form. Therefore, we sampled variance parameters using a Metropolis Hastings algorithm. The rest of the unknowns, effects and intercepts were subsequently sampled from their fully conditional densities as implemented in BGLR.

Metropolis-Hastings Step. This was used to sample the variances of effects as well as the error variances. Assuming a BayesC prior for <u>effects</u> ($b_.$ with $. = 0,1,2$ for main effects and interactions, respectively), the fully-conditional density of the variance parameters when the Beta prior is used can be shown to be

$$p(\frac{\sigma^2_{b.}}{K}|b_., \pi_.) \propto \pi_. \sigma^{2\left(-\frac{p}{2}\right)}_{b.} e^{-\left(\frac{b_.'b_.}{2\sigma^2_{b.}}\right)} \left(\frac{\sigma^2_{b.}}{K}\right)^{(\alpha_1 - 1)} \left(1 - \frac{\sigma^2_{b.}}{K}\right)^{(\alpha_2 - 1)} 1(0 < \sigma^2_{b.} < K)$$

(A3)

This distribution does not have a closed form; therefore, we draw samples for variance parameters using a Metropolis Hastings step (Sorensen and Gianola 2007; Gelman *et al.* 2014) using as target distribution Equation A3 and as proposal distribution

$$Q(\sigma^2_{b.}) = \chi^{-2}\left(p, \frac{b_.'b_.}{p}\right)$$

(A4)

Assuming a Gaussian likelihood, the fully conditional distribution of the <u>error variances</u> can be shown to be

$$p\left(\frac{\sigma^2_.}{K}|\varepsilon_.\right) \propto \sigma^{2\left(-\frac{n_.}{2}\right)}_. e^{-\left(\frac{\varepsilon_.'\varepsilon_.}{2\sigma^2_.}\right)} \left(\frac{\sigma^2_.}{K}\right)^{(\alpha_1 - 1)} \left(1 - \frac{\sigma^2_.}{K}\right)^{(\alpha_2 - 1)} 1(0 < \sigma^2_. < K)$$

(A5)

where $. = 1,2$ define the error variances for groups one and two, respectively.

The above expression gives the target distribution for the error variances. The proposal distribution for the error variances was:

$$Q(\sigma^2_.) = \chi^{-2}\left(n_., \frac{\varepsilon_.'\varepsilon_.}{n_.}\right)$$

(A6)

References:

Gelman A., J. B. Carlin, D. B. Rubin, H. S. Stern, 2014 *Bayesian data analysis*. Chapman and Hall/CRC, Boca Raton, FL.

Sorensen D., D. Gianola, 2007 *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer Science & Business Media, Berlin.

APPENDIX B

Mixture model representation for simulations to detect bias in estimates of average correlation of effects

In this note we show the mathematical representation of the mixture of normal distributions that was used to draw marker effects in the simulations. The phenotypes were simulated from the following model:

$$y_{1i} = \sum X_{1ij}\beta_{1j} + \varepsilon_{1i}; y_{2i} = \sum X_{2ij}\beta_{2j} + \varepsilon_{2i}$$

where $y_{ki}(k = 1,2)$ is the simulated phenotype for the ith individual in the kth group, β_{kj} is the effect of the jth marker $(j = 1,2,\dots p)$ in the kth group that is normally distributed with means 0 and variances $\sigma_{\beta k}^2$ and ε_{ki} is the error term for the kth group that is normally distributed with means 0 and variances σ_k^2.

The mixture models used in the simulations can be represented as:

$$\begin{bmatrix}\beta_{1j} \\ \beta_{2j}\end{bmatrix} \overset{iid}{\sim} \pi_1 MVN\left(0, \begin{matrix}\sigma_{\beta 11}^2 & \sigma_{\beta 12} \\ \sigma_{\beta 12} & \sigma_{\beta 22}^2\end{matrix}\right) + \pi_2 N\left(\beta_{1j}|0, \sigma_{\beta 12}^2\right)I(\beta_{2j} = 0) + \pi_3 N\left(\beta_{2j}|0, \sigma_{\beta 21}^2\right)I(\beta_{1j} = 0) + (1 - \pi_1 - $$

$$\pi_2 - \pi_3)I\left(\beta_{1j} = 0\right)I\left(\beta_{2j} = 0\right)$$

(A7)

where π_1, π_2, π_3 are the proportions of loci that have effects on both traits (π_1), on trait two only (π_2) and on trait one only (π_3). Under this assumption the variance of effects for group 1 and 2 are given by $\sigma_{\beta 1}^2 = \pi_1\sigma_{\beta 11}^2 + \pi_2\sigma_{\beta 12}^2$ and $\sigma_{\beta 2}^2 = \pi_1\sigma_{\beta 22}^2 + \pi_3\sigma_{\beta 21}^2$ and the covariance of effects is $\pi_1\sigma_{\beta 12}$. For the sake of simplicity, we assume all the variance parameters in Equation A7 to be the same, say σ_β^2. Thus $var(\beta_{1j}) = (\pi_1 + \pi_2)\sigma_\beta^2$, $var(\beta_{2j}) = (\pi_1 + \pi_3)\sigma_\beta^2$ and $Cov\left(\beta_{1j}, \beta_{2j}\right) = \pi_1\sigma_{\beta 12}$

Upon centering and scaling the markers so that the average diagonal values of X_1X_1' and X_2X_2' are equal to one, the proportion of variance explained parameters become:

$$v_{g1}^2 = \frac{(\pi_1+\pi_2)\sigma_\beta^2}{(\pi_1+\pi_2)\sigma_\beta^2+\sigma_{\varepsilon 1}^2} \text{ and } v_{g2}^2 = \frac{(\pi_1+\pi_3)\sigma_\beta^2}{(\pi_1+\pi_3)\sigma_\beta^2+\sigma_{\varepsilon 2}^2}$$

(A8)

Likewise, the correlation of effects is given by:

$$Cor\left(\beta_{1j}, \beta_{2j}\right) = \rho = \frac{\pi_1\sigma_{\beta 12}}{\sqrt{(\pi_1+\pi_2)\sigma_\beta^2 \times (\pi_1+\pi_3)\sigma_\beta^2}}$$

(A9)

In both simulation settings we assumed $\sigma_{\varepsilon1}^2 = 1 - v_{g1}^2, \sigma_{\varepsilon2}^2 = 1 - v_{g2}^2$ and $v_{g1}^2 = v_{g2}^2 = 0.5$.

In the **first simulation setting**, we assume $\pi_1 = 1, \pi_2 = 0, \pi_3 = 0$. This reduces the proportion of variance explained and correlation parameters to:

$$v_{g1}^2 = \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_{\varepsilon1}^2}, v_{g2}^2 = \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_{\varepsilon2}^2} \text{ and} \rho = \frac{\sigma_{\beta12}}{\sigma_\beta^2} \tag{A10}$$

In the **second simulation setting**, we assume $\pi_1 = 0.5, \pi_2 = 0.2, \pi_3 = 0.2$. This reduces the proportion of variance explained and correlation parameters to:

$$v_{g1}^2 = \frac{0.7\sigma_\beta^2}{0.7\sigma_\beta^2 + \sigma_{\varepsilon1}^2}, v_{g2}^2 = \frac{0.7\sigma_\beta^2}{0.7\sigma_\beta^2 + \sigma_{\varepsilon2}^2} \text{ and} \rho = \frac{5}{7} \frac{\sigma_{\beta12}}{\sigma_\beta^2} \tag{A11}$$

In both simulations we conducted a total of 200 Monte Carlo (MC) replicates and fit the interaction model (Equation 1) for each MC replicate. Bias and variance of estimates were estimated by comparing estimates (posterior means) with the true population parameters.

APPENDIX C

Bias in effect correlation using estimated effects vs. variance components

In this note, we show that the simple correlation of estimated effects can be a highly biased (towards zero) estimate of the true correlation of effects. Let β_{1j} and β_{2j} be the effects of marker j in groups 1 and 2. The true correlation of effects between the groups is given by $Cor(\beta_{1j}, \beta_{2j}) = \frac{Cov(\beta_{1j}, \beta_{2j})}{\sqrt{Var(\beta_{1j})Var(\beta_{2j})}}$. Likewise, let $\hat{\beta}_{1j}$ and $\hat{\beta}_{2j}$ denote the estimated effects from some unbiased estimator (e.g., Ordinary Least Squares or Maximum Likelihood with sufficiently large sample size) such that $\hat{\beta}_{1j} = \beta_{1j} + \delta_{1j}$ and $\hat{\beta}_{2j} = \beta_{2j} + \delta_{2j}$ where δ_{1j} and δ_{2j} represent estimation errors. Since the groups are disjoint, $cor(\delta_{1j}, \delta_{2j}) = 0$. Now, the correlation of *estimated* effects between the groups is given by $Cor\left(\hat{\beta}_{1j}, \hat{\beta}_{2j}\right) = \frac{Cov(\beta_{1j}, \beta_{2j})}{\sqrt{[Var(\beta_{1j})+Var(\delta_{1j})]\times[Var(\beta_{2j})+Var(\delta_{2j})]}}$. Clearly, $\left|Cor\left(\hat{\beta}_{1j}, \hat{\beta}_{2j}\right)\right| \leq |Cor(\beta_{1j}, \beta_{2j})|$ since $Var(\delta_{kj}) \geq 0, k = 1,2$. This shows that the simple correlation of estimated effects is a biased estimate of the true correlation of effects.