Genetics: Early Online, published on February 22, 2019 as 10.1534/genetics.119.301909

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, Germany

IRB number for ARIC dataset is 15-745; r050661

Study accession number for ARIC dataset: phs000280.v1.p1.

Running title: Cross-ethnic group effect heterogeneity

Keywords: population structure, GWAS, random effect interactions, Bayesian spike slab, effect heterogeneity

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

ABSTRACT. In humans, most genome-wide association studies have been conducted using data from Caucasians and many of the reported findings have not replicated in other populations. This lack of replication may be due to statistical issues (small sample size, confounding) or perhaps more fundamentally to differences in the genetic architecture of traits between ethnically diverse subpopulations. What aspects of the genetic architecture of traits vary between subpopulations and how can this be quantified? We consider studying effect heterogeneity using Bayesian random-effect interaction models. The proposed methodology can be applied using shrinkage and variable selection methods and produces useful information about effect-heterogeneity in the form of whole-genome summaries (e.g., the proportions of variance of a complex trait explained by a set of SNPs and the average correlation of effects) as well as SNP-specific attributes. Using simulations, we show that the proposed methodology yields (nearly) unbiased estimates when the sample size is not too small relative to the number of SNPs used. Subsequently, we used the methodology for the analyses of four complex human traits (standing height, high-density lipoprotein, low-density lipoprotein, and serum urate levels) in European-Americans (EAs) and African-Americans (AAs). The estimated correlations of effects between the two subpopulations were well below unity for all the traits, ranging from 0.73 to 0.50. The extent of effect heterogeneity varied between traits and SNP-sets. Height showed less differences in SNP effects between

AAs and EAs whereas HDL, a trait highly influenced by lifestyle, exhibited greater extent of effect heterogeneity. For all the traits, we observed substantial variability in effect heterogeneity across SNPs, suggesting that effect heterogeneity varies between regions of the genome.

Population structure is a pervasive feature in plant, animal, and human populations (Gaggiotti *et al.* 2009; Pfenninger *et al.* 2011; Puckett *et al.* 2014). In population genetics, differentiation between subpopulations is often measured by comparing allele frequencies, e.g., using the 'F-statistic' (Malécot 1947; Wright 1949 and Cockerham 1969). In genome-wide association studies (GWAS), population differentiation is predominantly viewed as a confounder (Astle and Balding 2009) that can lead to spurious associations (Lander and Schork 1994; Deng 2001; Marchini *et al.* 2004; Liu *et al.* 2011). To address this problem a variety of methods have been proposed (Price *et al.* 2010). However, rather than a confounder, population stratification can act as an effect-modifier, leading to heterogeneity in the genetic architecture of traits.

The evolutionary dynamics involved in the processes that lead to population structure can result in subpopulations with heterogeneity in allele frequencies and linkage disequilibrium (LD) patterns (Gabriel 2002). Moreover, in some instances, ethnic background correlates with environmental exposures (e.g., diet, income, lifestyle) and this can lead to genotype-by-environment interactions. All these differences between ethnic groups can induce heterogeneity in the genetic architecture of traits (de los Campos and Sorensen 2014). Quantifying the extent of effect-heterogeneity between ethnically diverse groups is relevant across disciplines and can shed light on whether results obtained in one group are expected to replicate in others. This is particularly important when we consider that the vast majority of GWAS have been conducted using data from Caucasians and that results reported from these studies do not always replicate in other

populations, which may indicate differences in genetic architectures between ethnic groups (Greene *et al.* 2009a; Kraft *et al.* 2009; Ng *et al.* 2014).

Several studies have demonstrated (or alluded to) effect heterogeneity between ethnic groups (Ntzani *et al.* 2012; de Candia *et al.* 2013; Li and Keating 2014; Brown *et al.* 2016). Most of these studies measured effect heterogeneity by estimating the average correlation of marker effects between two or more ethnically diverse groups.

One may attempt to estimate effect correlations by quantifying the average correlation of estimated effects from GWAS conducted in different ethnic groups. However, estimation errors make the simple correlation of estimates of effects a seriously biased (towards zero) estimate of the correlation of (true) effects (see Appendix C). To overcome this problem, several studies have used multivariate Gaussian random regression models. Such methods have been considered in both animal and plant breeding (Wei and Werf 1994; García-Cortés and Toro 2006; Karoui *et al.* 2012; Olson *et al.* 2012; Christensen *et al.* 2014; Lehermeier *et al.* 2015) as well as in human genetics (e.g. de Candia et al., 2013; Lee et al., 2012). Another approach estimates the correlation of effects using an extension of the LD-score regression (Brown *et al.* 2016).

The methods described above provide whole-genome summaries such as SNP-heritability and average correlation of effects. However, they don't shed light on how effect heterogeneity may vary across regions of the genome or between SNP sets. Moreover, the random regression methods commonly used to estimate effect-correlations assume that SNP effects follow Gaussian distributions. This assumption does not contemplate the possibility that some SNPs may have no effect in one or more than one group. To overcome this limitation, we consider modeling effect heterogeneity using a Bayesian random-effect interaction model that decomposes SNP effects into main and interaction components. Unlike previously used methods, the proposed approach

can be applied with both shrinkage and variable selection priors (e.g., Ishwaran and Rao 2005, Park and Casella 2008) and offers both whole-genome and SNP-specific measures of effect-heterogeneity.

Using simulations, we show that the proposed method yields nearly unbiased estimates when sample size (n) is not too small relative to the number of markers (p) used. Subsequently, we applied the proposed methodology to data from the ARIC (multi-ethnic Atherosclerosis Risk in Communities) study to quantify effect-heterogeneity between European and African ancestries (hereinafter referred to as European-Americans (EAs) and African-Americans (AAs), respectively). These subpopulations have important differences in allele frequencies, LD decay (Shifman 2003) and cultural and socio-economic factors that are linked to environmental exposures.

Our results show that for the four traits there is varying extent of effect-heterogeneity (the correlation of effects was highest for height and lower for lipid traits). Moreover, we show that for HDL, LDL and serum urate there is a great deal of variability in effect heterogeneity across the genome.

Materials and Methods

Meuwissen et al. (2001) proposed to predict complex traits by regressing phenotypes on whole-genome panels of SNPs. Their model was developed with reference to a homogeneous population. Here, following (de los Campos *et al.* 2015b), we consider extending the whole genome regression model by including random-effect interactions between markers and groups. Considering two groups, the regression of phenotypes ($\boldsymbol{y}_k = \{y_{k1}, \ldots, y_{kn_k}\}$, where $k = 1,2$ indexes groups and n_k denotes the number of individuals in the k^{th} group) on p markers (e.g., SNPs) can be represented as follows:

$$\begin{bmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{bmatrix}=\begin{bmatrix}1\mu_1\\ 1\mu_2\end{bmatrix}+\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{bmatrix}\boldsymbol{b}_0+\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{0}\end{bmatrix}\boldsymbol{b}_1+\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{X}_2\end{bmatrix}\boldsymbol{b}_2+\begin{bmatrix}\boldsymbol{\varepsilon}_1\\ \boldsymbol{\varepsilon}_2\end{bmatrix} \quad (1)$$

where μ_1 and μ_2 are group-specific intercepts, $\boldsymbol{b}_0=\{b_{0j}\}_{j=1}^{p}$ is a vector of 'main effects', $\boldsymbol{b}_1=\{b_{1j}\}_{j=1}^{p}$ and $\boldsymbol{b}_2=\{b_{2j}\}_{j=1}^{p}$ are group-specific interactions and $\boldsymbol{\varepsilon}_1=\{\varepsilon_{1i}\}_{i=1}^{n_1}$ and $\boldsymbol{\varepsilon}_2=\{\varepsilon_{2i}\}_{i=1}^{n_2}$ are error terms. In our models, we assume uncorrelated IID Gaussian errors with group-specific variances that is $\varepsilon_{1i}\underset{\sim}{iid}N(0,\sigma_1^2)$ and $\varepsilon_{2i}\underset{\sim}{iid}N(0,\sigma_2^2)$.

Marker effects in groups 1 and 2 are defined by the sum of the main and group-specific terms, that is, $\beta_{1j}=b_{0j}+b_{1j}$ and $\beta_{2j}=b_{0j}+b_{2j}$, respectively. Since the number of markers is usually large relative to sample size, we treat both main and interaction effects as random. Depending on the distribution assigned to SNP effects, the model can induce variable selection, shrinkage, or a combination of both (Ishwaran and Rao 2005; Gianola *et al.* 2009; de los Campos *et al.* 2013). To illustrate, we considered two priors for main and interaction effects: a Gaussian distribution and a prior with a point of mass at zero and a Gaussian slab, also known as BayesC (Habier *et al.* 2011).

In the **Gaussian setting** we assign independent Normal priors with null mean and with different variances for the main and interaction effects, that is

$$b_{0j}\underset{\sim}{iid}N\left(0,\sigma_{b_0}^2\right),\ b_{1j}\underset{\sim}{iid}N\left(0,\sigma_{b_1}^2\right)\text{ and }b_{2j}\underset{\sim}{iid}N\left(0,\sigma_{b_2}^2\right).$$

Above, $\sigma_{b_0}^2$, $\sigma_{b_1}^2$ and $\sigma_{b_2}^2$ represent the prior variances of the main and interaction effects, respectively.

For the **Spike-Slab prior** we adopt the assumptions of the BayesC model (Habier *et al.* 2011), with set-specific variances and proportions of non-zero effects, that is

$$b_{0j}\underset{\sim}{iid}p\left(\pi_0,\tilde{\sigma}_{b_0}^2\right),\ b_{1j}\underset{\sim}{iid}p\left(\pi_1,\tilde{\sigma}_{b_1}^2\right)\text{ and }b_{2j}\underset{\sim}{iid}p\left(\pi_2,\tilde{\sigma}_{b_2}^2\right),$$

where $p\left(b_{.j}|\pi_{.},\tilde{\sigma}_{b_{.}}^2\right)$ is a mixture distribution of the form $p\left(b_{.j}|\pi_{.},\tilde{\sigma}_{b_{.}}^2\right) = (1-\pi_{.})1\left(b_{.j}=0\right) + \pi_{.}N\left(0,\tilde{\sigma}_{b_{.}}^2\right)$, for *"."=0,1,2*. Here, $\pi_{.}$ represents the proportion of non-null effects.

Hyper-parameters. In the Gaussian model the hyper-parameters are the error variance and the three variances of effects, that is $\Omega = \left\{\sigma_1^2, \sigma_2^2, \sigma_{b_0}^2, \sigma_{b_1}^2, \sigma_{b_2}^2\right\}$. In BayesC the hyper-parameters also include the proportion of non-null effects; therefore:$\Omega = \left\{\sigma_1^2, \sigma_2^2, \tilde{\sigma}_{b_0}^2, \tilde{\sigma}_{b_1}^2, \tilde{\sigma}_{b_2}^2, \pi_0, \pi_1, \pi_2\right\}$. These parameters control the extent of shrinkage and variable selection and how the architecture of effects may vary between groups. We treat these hyper-parameters as unknown and therefore assign prior distributions to them. For variance parameters, the conjugate prior is the scaled-inverse chi-squared. However, this prior can have some influence on inference. Therefore, instead we use a prior for variance parameters that is a transformation of the beta distribution (Appendix A). For the proportion of non-zero effects $\{\pi_0, \pi_1, \pi_2\}$ we use independent identical beta priors. This allows us to accommodate different effect distributions for different traits and sets of SNPs. Further details about this are provided in the **Analyses of Complex Human Traits** section below.

The models described above can be used to estimate several parameters that are descriptive of the trait architecture. Whole-genome summaries of the trait architecture and of effect heterogeneity include *the proportion of variance explained by SNPs* (or genomic heritability, e.g., de los Campos *et al.* 2015) in each of the ethnic groups, the *average correlation of effects* and the *average proportions of non-zero effects* (either main effects, interaction terms or total effects). Samples from the posterior distribution can also be used to estimate SNP-specific parameters such as the posterior correlation of a SNP effect, $\rho_j = Cor\left(\beta_{1j}, \beta_{2j}\right)$.

Genomic variance and the average correlation of effects were estimated using the methods described by Lehermeier et al. (2017). Briefly, at each iteration of an MCMC algorithm, we

used the samples of the main and interaction effects to form marker effects ($\beta_{1j(s)} = b_{0j(s)} + b_{1j(s)}$ and $\beta_{2j(s)} = b_{0j(s)} + b_{2j(s)}$ where, *s=1,…,N* is an index for the N MCMC samples collected) to obtain samples from the posterior distribution of the correlation of effects $\rho_s = Cor\left(\beta_{1j(s)}, \beta_{2j(s)}\right)$, here $Cor(\ \ )$ represents Pearson's product moment correlation. Likewise, at each iteration of the sampler, genomic values can be obtained from $\boldsymbol{u}_{1(s)} = \boldsymbol{X}_1\boldsymbol{\beta}_{1(s)}$ and $\boldsymbol{u}_{2(s)} = \boldsymbol{X}_2\boldsymbol{\beta}_{2(s)}$. Therefore, a sample for the posterior distribution of the genomic variances for each group was computed as $\sigma^2_{g1(s)} = (n_1 - 1)^{-1}\sum_i \left(u_{1i(s)} - \overline{u}_{1(s)}\right)^2$ and $\sigma^2_{g2(s)} = (n_2 - 1)^{-1}\sum_i \left(u_{2i(s)} - \overline{u}_{2(s)}\right)^2$, where $\overline{u}_{.(s)} = n_. \sum_i \ u_{.i(s)}$. Finally, samples from the posterior distribution of the proportion of variance of the trait explained by a SNP-set were obtained using: $v^2_{g1(s)} = \frac{\sigma^2_{g1(s)}}{\sigma^2_{g1(s)}+\sigma^2_{1(s)}}$ and $v^2_{g2(s)} = \frac{\sigma^2_{g2(s)}}{\sigma^2_{g2(s)}+\sigma^2_{2(s)}}$.

Data

Our simulation and real data analyses were based on data from the ARIC study. ARIC is a prospective epidemiologic study sponsored by the National Heart, Lung, and Blood Institute (NHLBI) conducted in four U.S. communities to study the causes of atherosclerosis and other cardiovascular risk factors such as blood lipids, lipoprotein cholesterols, and apolipoproteins. It has a total sample size of 15,792 (9,584 EA and 3,107 AA) men and women aged 45-64. A total of 13,113 individuals were genotyped using an Affymetrix array with a total of 934,940 SNPs. Genotype and phenotype data from the ARIC study was acquired through the dbGaP (IRB number 15-745; r050661 and study accession number phs000280.v1.p1).

Genotypes

We retained SNPs that had minor allele frequency higher than 1% in at least one of the two ethnic groups, had higher than 95% calling rate, and were mapped to one of the 23 human chromosomes. After QC, we retained 828,822 SNPs. Individuals with a missing rate >5% in their genotypes were removed. Individuals were classified as EA or AA based on self-reported ethnicity (also confirmed from principal components analyses, Figure S1). We removed individuals that had within-group genomic relationships higher than 0.075; this ensured that we retained enough number of distantly related individuals. The final data sets comprised only distantly related individuals including 6,627 EAs and 1,601 AAs.

Simulations

We simulated phenotypes using genotype data from the ARIC study from 6,627 EAs and 1,601 AAs. Phenotypes were simulated under an additive genetic model with a heritability of 0.5 for both groups. We further considered scenarios with the number of markers (n) varying from 100 to 10,000 and the true correlations of effects between groups varying from 0.2 to 0.8. In a first simulation setting we assumed that all the markers had effects on both groups. In a second setting, we assumed that 50% of the loci had effects on both groups, 20% had effects on EAs but not on AAs, 20% had effects on AAs but not on EAs and 10% had no effects on either group (non-causal variants). These simulations were conducted for 200 Monte Carlo replicates per setting. Finally, we considered an additional scenario where heritability was lower in both groups (0.2) or lower in one of the groups (0.2 in EA and 0.5 in AA and vice versa). Further details of the simulation are given in Appendix B.

Analyses of Complex Human Traits

We considered four complex phenotypes: height (cm), HDL (mmol/L) and LDL (mmol/L) cholesterol and serum urate (mg/dL). Individuals with height < 147 cm, LDL > 10 mmol/L and serum urate > 15 mg/dL were removed. We did not identify clear outliers for HDL. Transformation of the traits was not considered necessary (Figure S2). Phenotypes were pre-corrected for ethnicity, age, sex and the first five marker-derived principal components.

Models were fitted to subsets of SNPs selected based on single-marker regression (GWAS) p-values derived from independent data that did not include ARIC. For height, GWAS p-values were derived from the full release of the UK Biobank. For HDL and LDL p-values were from the Global Lipids Genetics Consortium (GLGC) computed after excluding data from ARIC. Finally, for serum urate, p-values were from the Global Urate Genetics Consortium (GUGC), also derived without using data from ARIC. The simple ranking of markers based on association *p*-values would lead to sets of highly redundant markers, i.e., markers in high LD (see Figure S3). To avoid this, we designed a windows-based selection algorithm where a window was defined as a set of consecutive SNPs that exceeded a given $-\log_{10}(p\text{-value})$ cutoff (this was done on a per-trait basis). Windows were made on a per-trait basis at $-\log_{10}(p\text{-value})$ cutoffs of 2, 2.3, 2.6, 3, 5, and 8 (Table S1). SNPs that cleared a given $-\log_{10}(p\text{-value})$ cutoff were termed "significant" at that cutoff (See Figure S4).

We fitted the interaction model to each of the four traits and each of the SNP-sets above described. For sensitivity analyses, we also fitted the same models to randomly chosen sets of SNPs (of sizes 500, 1,000, 2,500, 5,000, and 10,000 SNPs, respectively). Finally, for further sensitivity testing, we repeated the analysis with the EA ethnic group label randomly permuted.

Software

Models were fit using a modified version of the BGLR (Pérez and de los Campos 2014) R package (available at: https://github.com/gdlc/BGLR-R and at https://cran.r-project.org/web/packages/BGLR/index.html) that implements a weakly informative prior for variance parameters based on a transformation of the beta distribution (de los Campos *et al.* 2009) described above. We ran the Monte Carlo Markov Chain algorithm for 45,000 iterations; the first 15,000 iterations were discarded as burn-in and the remaining samples were thinned at a thinning interval of 5.

Hyper-parameters. BGLR assigns a beta prior to the proportion of non-zero effects, we choose the shape parameters of the beta prior to be equal to 1, which gives a uniform prior in the 0-1 interval. For variance parameters we devised a prior that is a modified version of the Beta prior (see Appendix A) and used shape parameters equal to 1.01 to obtain an almost uniform prior for variance parameters within the interval [0,K] where K was twice the variance of the phenotype.

Data availability

Supplemental material has been uploaded to FigShare. File S1 contains supplementary figures, tables and Appendices. The IRB number for ARIC dataset is 15-745; r050661 and the study accession number for ARIC dataset: phs000280.v1.p1.

Results

Simulations

In both simulation settings, the *proportion of variance explained by a SNP-set* was estimated with almost no bias using both Gaussian and BayesC priors (see Figures 1 and S5 for the first and second simulation scenarios, respectively). The standard errors were higher for AAs as compared to EAs, which was expected given that the sample size was smaller for AAs. As one would expect, the standard errors also increased with the number-of-loci/sample-size ratio. Using the BayesC prior, the estimates of proportion of variance explained by a SNP-set were mildly biased across all values of true effect-correlation when the number of QTL was greater than 10,000. There was a mild to moderate bias in the group with smaller sample size when the true proportion of variance explained in this group went from high (0.8) to low (0.2). (see Figure 2 and Table S2).

FIGURE 1

Estimates of *effect correlations* were also nearly unbiased (see Figures 2 and S6), especially when the true proportion of variance explained was high in group with smaller sample size. However, the standard errors were very large, particularly when the correlations were low. In scenarios involving less than 100 or more than 5,000 QTL, we observed small biases. (Figure 2 and Figure S6). The average standard error of the estimated correlation was high with the smallest (100) and the largest (10,000) numbers of QTL and lower for scenarios in between. With low simulated trait heritability (0.2) and small group sample size (AAs), we observed an upward (downward) bias when the simulated correlation was low (high) (see Table S3).

FIGURE 2

Analyses of Complex Human Traits

Since our simulations revealed that an n/p ratio of at least 1/3 results in nearly unbiased estimates of proportion of variance explained by a SNP-set, we fit our model to subsets of markers instead of using whole-genome data (see methods for a description of how these subsets were obtained). Figure 3 shows the estimated *proportion of variance explained by a SNP-set* obtained using the BayesC prior, by trait, ethnicity and the set of SNPs used. (The results obtained with the Gaussian prior are displayed in Figure S7). As expected, the estimated proportion of variance explained by a SNP-set increased with the number of SNPs used. Interestingly, this parameter was systematically higher in EAs than in AAs for height and HDL, and the order was reversed in other traits (LDL and serum urate). However, the credibility intervals between both ethnic groups overlapped for all traits except height. The estimated proportion of variance explained by a SNP-set obtained with the Gaussian prior were similar to the ones found with the BayesC prior (see Figure S7) for all traits except serum urate, which yielded larger estimates for AAs than those obtained using the BayesC prior.

FIGURE 3

The estimated average correlation of effects (Figure 4) ranged from 0.711 (for height with the SNP-set obtained with $-\log_{10}(p\text{-value})$ cutoff of 8) to 0.500 (for HDL with the SNP-set obtained with a $-\log_{10}(p\text{-value})$ cutoff of 2.3). Overall the correlation of effects was highest for height and serum urate and lowest for LDL and HDL. In all traits except HDL, the correlation of effects tended to decrease as more SNPs were added in the model; however, the confidence regions for the different SNP sets overlapped. The estimated correlation of effects with the Gaussian prior for marker effects was similar to those obtained using the BayesC prior, with subtle differences between the two priors for height, HDL and LDL (Figure S8).

FIGURE 4

Figure 5 shows the estimated proportion of non-zero SNP effects obtained with the BayesC prior, by trait, ethnic group and SNP-set. For both groups, the proportions of non-zero effects were high at large $-\log_{10}(p\text{-value})$ cutoffs and decreased as the number of markers included in the model increased. For height, the proportions of non-zero effects were similar between EAs and AAs. However, for LDL (and serum urate to a lesser extent) the decrease in the proportion of non-zero effects was stronger in EAs. Figure S9 displays the proportion of non-zero main and interaction effects. The proportion of non-zero main effects decreased as the number of SNPs increased and the proportion of non-zero interaction effects tended to remain constant (except for the LDL-interactions for EAs). Interestingly, the proportion of non-zero effects dropped very fast with the number of SNPs for HDL, LDL and serum urate but not for height.

FIGURE 5

Figures 3-5 (and the corresponding supplementary figure S9), correspond to overall summaries (proportion of variance explained by a SNP-set, average correlation of effects, proportion of non-zero effects). However, the models used also render SNP-specific summaries. Figure 6 shows the posterior mean of the correlation of effects between ethnic groups for individual SNPs by trait for the SNP set obtained using a $-\log_{10}(p\text{-value})$ cutoff of 2. We had no SNP with negative posterior correlation of effect. For height, the posterior correlation of individual-SNP effects ranged from 0.4-0.8. However, for HDL, LDL and serum urate, there was more variability among SNPs, with several SNPs having posterior correlation of effects greater than 0.8 and many with posterior correlation of effects smaller than 0.4.

Figure S10 and S11 correspond to the proportion of variance explained and average correlation of effects between EAs and AAs from randomly chosen sets of markers. The estimates of

both, proportion of variance explained by a SNP-set (for EAs and AAs) as well as effect correlations, are far lower than those obtained using GWAS-selected markers (Figure 3 and S7).

Figures S12 and S13 correspond to the estimates of proportion of variance explained and the average correlation of effects from GWAS-selected markers by randomly dividing the EAs into two groups such that the sample size of one of two groups is same as that of the AA dataset. As expected, the estimates of the proportion of variance explained are similar within EAs and the estimates of effect correlation are much higher within EAs than between EAs and AAs across all traits (in particular, the effect correlation estimates are >0.90 for height across all SNP sets).

FIGURE 6

Discussion

Genome-Wide Association (GWA) studies have been conducted predominantly in Caucasian populations (Haga 2010; Rosenberg *et al.* 2010). Although more recent works have recognized inclusion of diverse ethnic groups, especially African Americans (e.g. Brant et al., 2017; Park et al., 2017; Taylor et al., 2016) the total number of GWAS studies for African Americans is still fairly low compared to populations of European ancestry (Peprah *et al.* 2015) and replication of signals in African American populations is much less common (Marigorta and Navarro 2013). Moreover, the associations reported to be strong in Caucasians have been weaker (or even non-significant) in other ethnic groups (Gudbjartsson *et al.* 2007; Omori *et al.* 2008; Yamada *et al.* 2009; Barnholtz-Sloan *et al.* 2011; Tsai *et al.* 2014; Prasad *et al.* 2017) and some studies have reported effects with opposite sign in different populations (Lewis *et al.* 2008; Yamada *et al.* 2009). More recent studies have also confirmed the presence of genetic heterogeneity between ethnic groups for various traits (Brown *et al.* 2016; de Vlaming *et al.* 2017; Zhou *et al.* 2018).

While some of these differences could be attributed to small sample size (some well-powered studies have shown strong overlaps in GWAS-significant variants between Europeans and other ethnic groups (Franceschini *et al.* 2013; Okada *et al.* 2014)), there is substantial evidence supporting effect-heterogeneity. Understanding the reasons that underlie these differences and quantifying the degree of similarity in the architecture of a trait across populations represents an important research goal.

In humans, Shi *et al.* 2017 estimated local correlations between traits using individual-level data while Brown et al. 2016 considered quantifying the average correlation of effects between populations using summary-based association statistics. Their approach extended LD score regression (Bulik-Sullivan *et al.* 2015) to multiple ethnic groups and has the advantage that it can be used with summary statistics. However, some authors have questioned the assumptions of the LD score regression method (Speed *et al.* 2018) and accurate estimation requires using several thousands of SNPs. Thus, the method is not well-suited for studying effect similarity within genomic regions, something that the method proposed here can achieve. Also, the LD-score method requires access to good quality external reference panels (especially for non-European populations) to construct population LD matrices, assumes normal distribution of marker effects (infinitesimal model) and is not directly applicable to admixed populations (or populations that have long range LD).

In this study, we propose to study ethnic differences in the architecture of traits using a random effect Bayesian interaction model. The proposed approach can be used to estimate whole-genome summaries such as (a) the proportion of variance explained by SNPs, (b) the average effect correlation, (c) proportion of non-zero effects, as well as finer features of the trait architecture (e.g. SNP-specific correlation of effects). Similar approaches have been considered in ani-

mal and plant breeding (e.g., Christensen et al., 2014; García-Cortés and Toro, 2006; Lehermeier et al., 2015; Wei and Werf, 1994) and in human genetics (e.g., de Candia et al., 2013; Lee et al., 2012) for the analysis of data from heterogeneous populations. However, previous studies were based on Gaussian assumptions and only offered whole-genome summaries of the trait architecture. The approach presented in this study is more flexible in that it can be used with both shrinkage and variable selection priors (Ishwaran and Rao 2005; Park and Casella 2008) and can be used to infer not only whole-genome features but also regional and SNP-specific features of the trait architecture.

We evaluated the proposed methodology under two different priors (Gaussian and BayesC) using simulations and applied it to real human data to study the genetic architecture of four traits (Height, HDL, LDL, Serum urate) in EAs and AAs. Our simulation study (based on real EA and AA genotypes from ARIC) revealed that both Gaussian and BayesC priors yield nearly unbiased estimates of proportion of variance explained by a SNP-set and of effect correlations. We observed mild to moderate upward (downward) bias in low (high) effect correlations when the proportion of variance explained was low (0.2), the sample size was small (<2000) and the number of considered SNPs was large relative to sample size. Given the small sample size available for our real data analyses, and considering our simulation results, we applied the proposed methodology to subsets of SNPs pre-selected using GWAS results obtained from other (independent) data sets.

From our real data analyses, we observed similar proportions of variance explained with both BayesC and Gaussian priors for marker effects (Figure 3 and S7). The Gaussian prior is a special case of the BayesC prior, thus BayesC is more flexible. Whether these two methods will render different estimates would critically depend on the trait architecture.

With the exception of height, the average proportions of variance explained across all marker sets were similar between EAs and AAs. For height, the average proportion of variance explained was greater among EAs than among AAs. This is likely due to the fact that the SNPs used for the analysis of height were selected using GWAS results entirely based on data from Caucasians (UK Biobank); the same trend was not observed for other traits perhaps because there was some mixture in ethnicity in the other GWAS consortia from which markers were chosen (GLGC, GUGC). When we fit similar models using randomly chosen markers (Figure S10), we observed that the proportion of variance explained by randomly selected markers was smaller than that explained by regression on markers selected from GWAS results for both EAs and AAs. This showed that indeed, selection based on GWAS results leads to more informative markers in both populations.

Our analyses also revealed important differences in correlation of effects between traits. The **estimated correlation of effects** ranged from 0.482-0.728, indicating the presence of genetic heterogeneity across all four traits, even for strongly associated markers (Figure 4). For height, the correlation of effects was highest when using SNPs that had the smallest GWAS p-value (likely SNPs with relatively large effect and not very extreme allele frequency), suggesting that the correlation of effects may be lower for SNPs with small effects and extreme allele frequencies. Another possible explanation for effect heterogeneity could also be the tagging differences between EAs and AAs, especially in the polygenic tail for a given trait.

Height had higher correlation of effects between EAs and AAs than serum urate and lipid traits, suggesting that height may have a more similar genetic architecture between EAs and AAs than the other traits (especially than the lipid traits). Furthermore, we found differences in the estimated proportion of non-zero effects between EAs and AAs for HDL, LDL, and serum urate

but not for height, reinforcing that the genetic architecture of height may be more similar between EAs and AAs in comparison to the other three traits (Figure 5 and S9).

The proportion of non-zero effects markedly decreased with the -logP value; this is expected since relaxing the threshold used to pre-select SNPs is likely to lead to the inclusion of SNPs with no effect. This was particularly clear for lipid traits. This trend is largely driven by the proportion of non-zero *main* effects for both ethnic groups (i.e. effects common to both ethnic groups – Figure S9). Finally, we also observed greater variability in posterior correlation of effects among lipid traits and serum urate in comparison to height (Figure 6).

Genetic-by-environmental interaction (G×E) could lead to effect-heterogeneity in additive models. Indeed, if ethnicity correlates with lifestyle, diet, income and other factors that may induce G×E, then SNP effects can become population-specific. Interestingly, the three traits that are more affected by diet and lifestyle (LDL, HDL and serum urate) showed stronger evidence of effect-heterogeneity than height. Likewise, unaccounted epistasis, coupled with differences in allele frequencies, may also lead to effect-heterogeneity in additive models. Indeed, some authors (Mackay and Moore 2014) have argued that the epistasis may be responsible for the majority of the small-effect additive effect affecting complex traits, and previous studies have attributed the non-replication of genetic associations in different populations to epistasis (Greene *et al.* 2009b). Thus, epistatic gene action can also have a role in explaining differences in the allelic substitution effects of SNPs and can consequently induce effect-heterogeneity.

In **conclusion**, we have proposed a versatile methodology based on random-effects interactions that can apply non-Gaussian priors to marker effects for quantifying the extent of effect heterogeneity between ethnically diverse groups using a combination of variable selection and shrinkage. This proposed approach can yield estimates of proportions of variance explained by a

SNP-set, average correlation of effects, proportion of non-zero effects as well as SNP-specific attributes in genomic regions of interest. According to our simulations, the methodology renders nearly unbiased estimates provided that the n/p ratio is not much smaller than 1/3. Of the traits considered in our study, effect heterogeneity was lower for height than for traits influenced by lifestyle. We postulate that differences in allele frequency and in LD patterns, together with epistasis and G×E can contribute to effect heterogeneity between AAs and EAs.

Acknowledgments

GDLC and YV acknowledge financial support from NIH grants GM R01099992 and GM R01101219. The authors acknowledge valuable comments provided by Drs. Sadeep Shrestha, Trudy Mackay, Edward Buckler, Marguerite Irvin and Nianjun Liu. The authors also acknowledge the help of Cristen Willer and Sebanti Sengupta for providing us summary statistics for HDL and LDL from the GLGC consortium after excluding the ARIC cohort as well as Christian Gieger and Jürgen Riegel for providing summary-statistics for serum urate from the GUGC consortium after excluding the ARIC cohort.

Literature Cited

Astle W., Balding D. J., 2009 Population Structure and Cryptic Relatedness in Genetic Association Studies. Stat. Sci. 24: 451–471.

Barnholtz-Sloan J. S., Raska P., Rebbeck T. R., Millikan R. C., 2011 Replication of GWAS "Hits" by Race for Breast and Prostate Cancers in European Americans and African Americans. Front. Genet. 2: 37.

Brant S. R., Okou D. T., Simpson C. L., Cutler D. J., Haritunians T., *et al.*, 2017 Genome-Wide Association Study Identifies African-Specific Susceptibility Loci in African Americans With Inflammatory Bowel Disease. Gastroenterology 152: 206–217.e2.

Brown B. C., Ye C. J., Price A. L., Zaitlen N., Zaitlen N., 2016 Transethnic Genetic-Correlation Estimates from Summary Statistics. Am. J. Hum. Genet. 99: 76–88.

Bulik-Sullivan B. K., Loh P.-R., Finucane H. K., Ripke S., Yang J., *et al.*, 2015 LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47: 291–295.

Christensen O. F., Madsen P., Nielsen B., Su G., 2014 Genomic evaluation of both purebred and crossbred performances. Genet. Sel. Evol. 46: 23.

Cockerham C., 1969 Variance of gene frequencies. Evolution (N. Y).: 72–84.

de Candia T. R., Lee S. H., Yang J., Browning B. L., Gejman P. V., *et al.*, 2013 Additive Genetic Variation in Schizophrenia Risk Is Shared by Populations of African and European Descent. Am. J. Hum. Genet. 93: 463–470.

de los Campos G., Naya H., Gianola D., Crossa J., Legarra A., *et al.*, 2009 Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics 182: 375–85.

de los Campos G., Hickey J. ., Pong-Wong R., Daetwyler H. ., 2013 Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193: 327–345.

de los Campos G., Sorensen D., 2014 On the genomic analysis of data from structured populations. J. Anim. Breed. Genet. 131: 163–4.

de los Campos G., Sorensen D., Gianola D., 2015a Genomic heritability: what is it? PLoS Genet. 11: e1005048.

de los Campos G., Veturi Y., Vazquez A. I., Lehermeier C., Pérez-Rodríguez P., 2015b Incorporating Genetic Heterogeneity in Whole-Genome Regressions Using Interactions. J. Agric. Biol. Environ. Stat. 20: 467–490.

Deng H. W., 2001 Population admixture may appear to mask, change or reverse genetic effects of genes underlying complex traits. Genetics 159: 1319–23.

Franceschini N., Fox E., Zhang Z., Edwards T. L., Nalls M. A., *et al.*, 2013 Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. Am. J. Hum. Genet. 93: 545–54.

Gabriel S. B., 2002 The Structure of Haplotype Blocks in the Human Genome. Science (80-.). 296: 2225–2229.

Gaggiotti O. E., Bekkevold D., Jørgensen H. B. H., Foll M., Carvalho G. R., *et al.*, 2009 Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. Evolution 63: 2939–51.

García-Cortés L., Toro M., 2006 Multibreed analysis by splitting the breeding values. Genet. Sel.

Gianola D., de los Campos G., Hill W. G., Manfredi E., Fernando R., 2009 Additive genetic variability and the Bayesian alphabet. Genetics 183: 347–63.

Greene C. S., Penrod N. M., Williams S. M., Moore J. H., 2009a Failure to replicate a genetic association may provide important clues about genetic architecture. PLoS One 4: e5639.

Greene C. S., Penrod N. M., Williams S. M., Moore J. H., 2009b Failure to Replicate a Genetic Association May Provide Important Clues About Genetic Architecture (TIA Sorensen, Ed.). PLoS One 4: e5639.

Gudbjartsson D. F., Arnar D. O., Helgadottir A., Gretarsdottir S., Holm H., *et al.*, 2007 Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature 448: 353–7.

Habier D., Fernando R. L., Kizilkaya K., Garrick D. J., 2011 Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics 12: 186.

Haga S. B., 2010 Impact of limited population diversity of genome-wide association studies. Genet. Med. 12: 81–4.

Ishwaran H., Rao J. S., 2005 Spike and slab variable selection: Frequentist and Bayesian strategies. Ann. Stat. 33: 730–773.

Karoui S., Carabaño M. J., Díaz C., Legarra A., VanRaden P., *et al.*, 2012 Joint genomic evaluation of French dairy cattle breeds using multiple-trait models. Genet. Sel. Evol. 44: 39.

Kraft P., Zeggini E., Ioannidis J. P. A., 2009 Replication in genome-wide association studies.

Stat. Sci. 24: 561–573.

Lander E. S., Schork N. J., 1994 Genetic dissection of complex traits. Science 265: 2037–48.

Lee S. H., Yang J., Goddard M. E., Visscher P. M., Wray N. R., 2012 Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics 28: 2540–2.

Lehermeier C., Schön C.-C., de los Campos G., 2015 Assessment of Genetic Heterogeneity in Structured Plant Populations Using Multivariate Whole-Genome Regression Models. Genetics 201: 323-337.

Lehermeier C., de los Campos G., Wimmer V., Schön C.-C., 2017 Genomic variance estimates: With or without disequilibrium covariances? J. Anim. Breed. Genet. 134: 232–241.

Li Y. R., Keating B. J., 2014 Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. Genome Med. 6: 91.

Liu N., Zhao H., Patki A., Limdi N. A., Allison D. B., 2011 Controlling Population Structure in Human Genetic Association Studies with Samples of Unrelated Individuals. Stat. Interface 4: 317–326.

Mackay T. F., Moore J. H., 2014 Why epistasis is important for tackling complex human disease genetics. Genome Med. 6: 124.

Malécot G., 1947 *Les Mathématiques de l'hérédité*. Masson, Paris.

Marchini J., Cardon L. R., Phillips M. S., Donnelly P., 2004 The effects of human population

structure on large genetic association studies. Nat. Genet. 36: 512–7.

Marigorta U. M., Navarro A., 2013 High Trans-ethnic Replicability of GWAS Results Implies Common Causal Variants (SM Williams, Ed.). PLoS Genet. 9: e1003566.

Meuwissen T. H., Hayes B. J., Goddard M. E., 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–29.

Ng M. C. Y., Shriner D., Chen B. H., Li J., Chen W.-M., *et al.*, 2014 Meta-Analysis of Genome-Wide Association Studies in African Americans Provides Insights into the Genetic Architecture of Type 2 Diabetes (E Zeggini, Ed.). PLoS Genet. 10: e1004517.

Ntzani E. E., Liberopoulos G., Manolio T. A., Ioannidis J. P. A., 2012 Consistency of genome-wide associations across major ancestral groups. Hum. Genet. 131: 1057–1071.

Okada Y., Wu D., Trynka G., Raj T., Terao C., *et al.*, 2014 Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506: 376–81.

Olson K. M., VanRaden P. M., Tooker M. E., 2012 Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 95: 5378–5383.

Omori S., Tanaka Y., Takahashi A., Hirose H., Kashiwagi A., *et al.*, 2008 Association of CDKAL1, IGF2BP2, CDKN2A/B, HHEX, SLC30A8, and KCNJ11 with susceptibility to type 2 diabetes in a Japanese population. Diabetes 57: 791–5.

Park T., Casella G., 2008 The Bayesian Lasso. J. Am. Stat. Assoc. 103: 681–686.

Park S. L., Cheng I., Haiman C. A., 2017 Genome-wide association studies of cancer in diverse

populations. Cancer Epidemiol. Biomarkers Prev. 27(4): 405-417

Peprah E., Xu H., Tekola-Ayele F., Royal C. D., 2015 Genome-wide association studies in Africans and African Americans: expanding the framework of the genomics of human traits and disease. Public Health Genomics 18: 40–51.

Pérez P., de los Campos G., 2014 Genome-wide regression and prediction with the BGLR statistical package. Genetics 198: 483–95.

Pfenninger M., Salinger M., Haun T., Feldmeyer B., 2011 Factors and processes shaping the population structure and distribution of genetic variation across the species range of the freshwater snail radix balthica (Pulmonata, Basommatophora). BMC Evol. Biol. 11: 135.

Prasad S., Bhatia T., Kukshal P., Nimgaonkar V. L., Deshpande S. N., *et al.*, 2017 Attempts to replicate genetic associations with schizophrenia in a cohort from north India. npj Schizophr. 3: 28.

Price A. L., Zaitlen N. A., Reich D., Patterson N., 2010 New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. 11: 459–463.

Puckett E. E., Kristensen T. V, Wilton C. M., Lyda S. B., Noyce K. V, *et al.*, 2014 Influence of drift and admixture on population structure of American black bears (Ursus americanus) in the Central Interior Highlands, USA, 50 years after translocation. Mol. Ecol. 23: 2414–27.

Rosenberg N. A., Huang L., Jewett E. M., Szpiech Z. A., Jankovic I., *et al.*, 2010 Genome-wide association studies in diverse populations. Nat. Rev. Genet. 11: 356–66.

Shi H., Mancuso N., Spendlove S., Pasaniuc B., 2017 Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. Am. J. Hum. Genet. 101: 737–751.

Shifman S., 2003 Linkage disequilibrium patterns of the human genome across populations. Hum. Mol. Genet. 12: 771–776.

Speed D., Balding D. J. 2018 Exposing flaws in S-LDSC; reply to Gazal et al. Biorxiv **doi:** https://doi.org/10.1101/280784

Taylor J. Y., Schwander K., Kardia S. L. R., Arnett D., Liang J., *et al.*, 2016 A Genome-wide study of blood pressure in African Americans accounting for gene-smoking interaction. Sci. Rep. 6: 18812.

Tsai E. A., Grochowski C. M., Loomes K. M., Bessho K., Hakonarson H., *et al.*, 2014 Replication of a GWAS signal in a Caucasian population implicates ADD3 in susceptibility to biliary atresia. Hum. Genet. 133: 235–43.

Vlaming R. de, Okbay A., Rietveld C. A., Johannesson M., Magnusson P. K. E., *et al.*, 2017 Meta-GWAS Accuracy and Power (MetaGAP) Calculator Shows that Hiding Heritability Is Partially Due to Imperfect Genetic Correlations across Studies (J Marchini, Ed.). PLOS Genet. 13: e1006495.

Wei M., Werf J. Van der, 1994 Maximizing genetic response in crossbreds using both purebred and crossbred information. Anim. Prod.

Wright S., 1949 The Genetical Structure of Populations. Ann. Eugen. 15: 323–354.

Yamada H., Penney K. L., Takahashi H., Katoh T., Yamano Y., *et al.*, 2009 Replication of prostate cancer risk loci in a Japanese case-control association study. J. Natl. Cancer Inst. 101: 1330–6.

Zhou X., Cheung C.-L., Karasugi T., Karppinen J., Samartzis D., *et al.*, 2018 Trans-ethnic polygenic analysis supports genetic overlaps of lumbar disc degeneration with height, body mass index, and bone mineral density. biorxiv.

Figures

Figure 1. **Average estimates of** proportion of variance explained by a SNP-set **obtained in the first 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 2nd simulation scenario are presented in Figure S5.

Figure 2. **Average estimates of the correlation of effects in the first 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 2nd simulation scenario are presented in Figure S6.

Figure 3. Proportion of variance explained by subsets of SNPs obtained with the BayesC-interaction model, by trait, ethnicity and SNP set. Estimated (median) proportion of variance explained by a SNP-set (y-axis) is plotted by trait, ethnicity and $\log_{10}(p\text{-value})$ cutoff used to

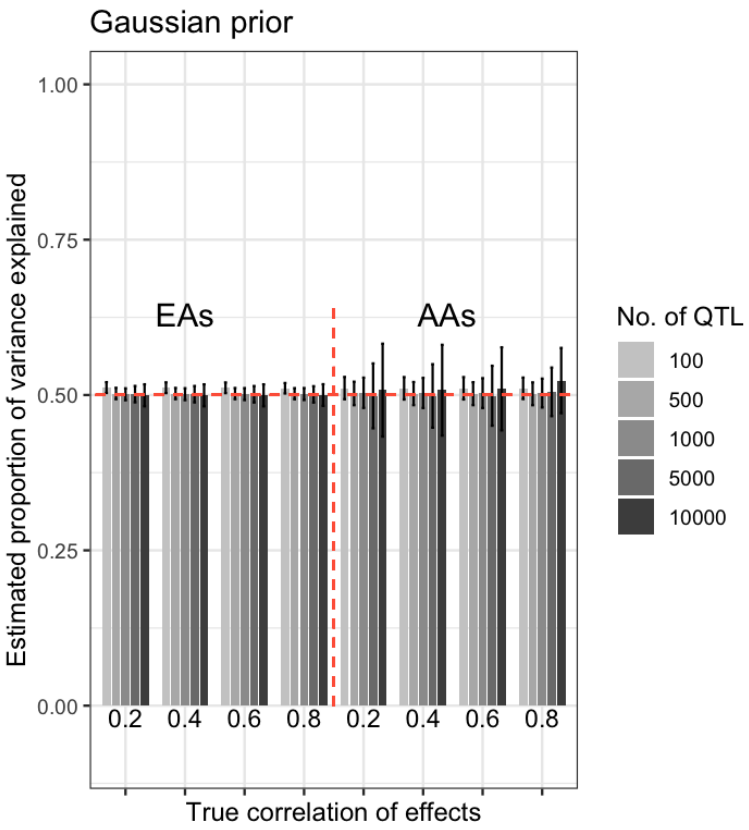
choose markers from GWAS consortia (excluding ARIC). Numerals above the bars indicate the proportion of variance explained by either ethnic group and the corresponding number of SNPs used for model fitting (in parentheses at the bottom). Vertical lines give estimates of +/- posterior standard deviation.

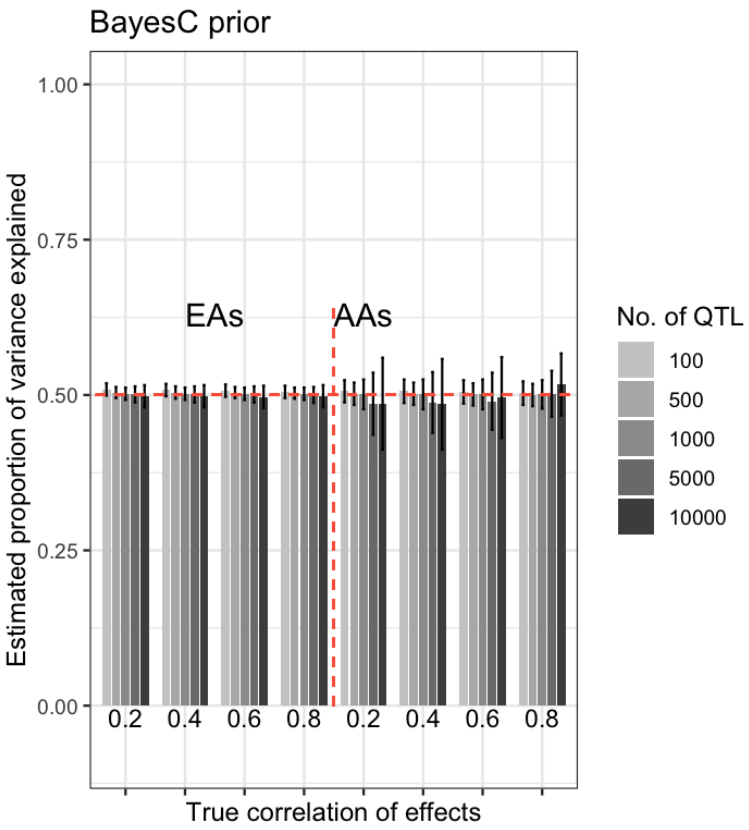
Figure 4. Estimated correlation of effects between African Americans (AAs) and European Americans (EAs) obtained with the BayesC-interaction model, by trait and SNP set. Estimated correlation of effects between AAs and EAs (y-axis) is plotted by trait using markers selected from GWAS consortia (excluding ARIC). In each plot, the numerals above the bars indicate the median correlation of effects and the number of SNPs used for model fitting (in parentheses at the bottom). Vertical lines give estimates of +/- posterior standard deviation.

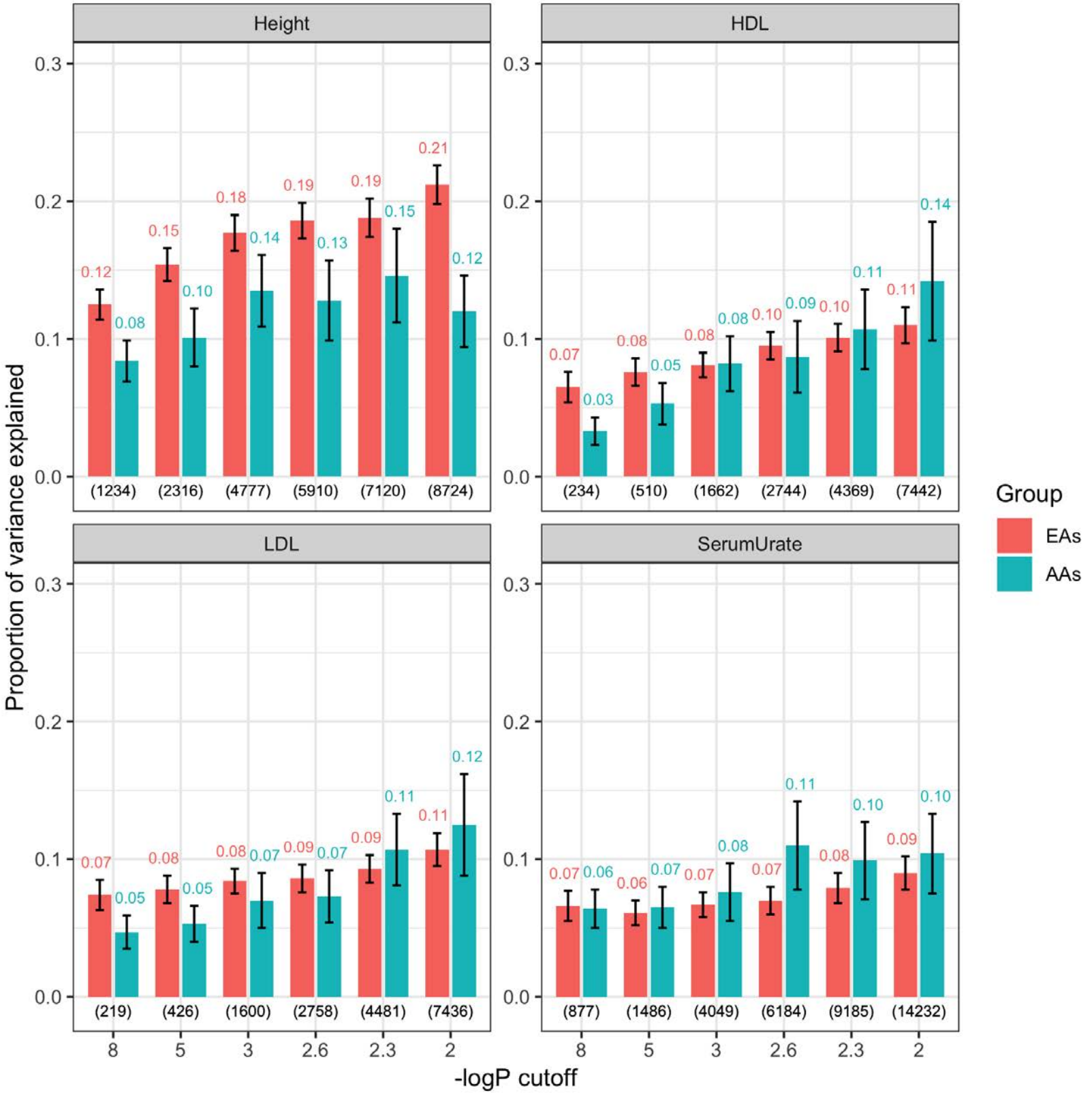
Figure 5. Estimated proportion of non-zero effects between African Americans (AAs) and European Americans (EAs) obtained with the BayesC-interaction model, by trait and SNP set. Estimated proportion of non-zero effects between AAs and EAs (y-axis) is plotted by trait using markers selected from GWAS consortia (excluding ARIC) at 6 different $-\log_{10}(p\text{-value})$ cutoffs. In each plot, the numerals above the bars indicate the proportion of non-zero effects obtained using 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.

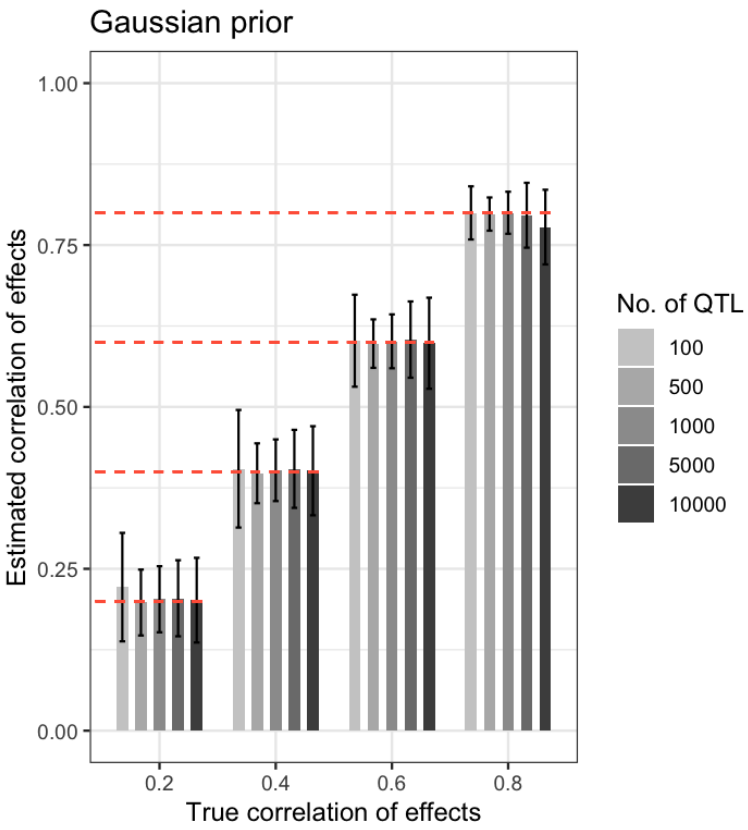
Figure 6. Posterior correlation of individual SNP effect between African Americans (AAs) and European Americans (EAs), by trait for SNPs that clear a $-\log_{10}(p\text{-value})$ of 2. Plots are

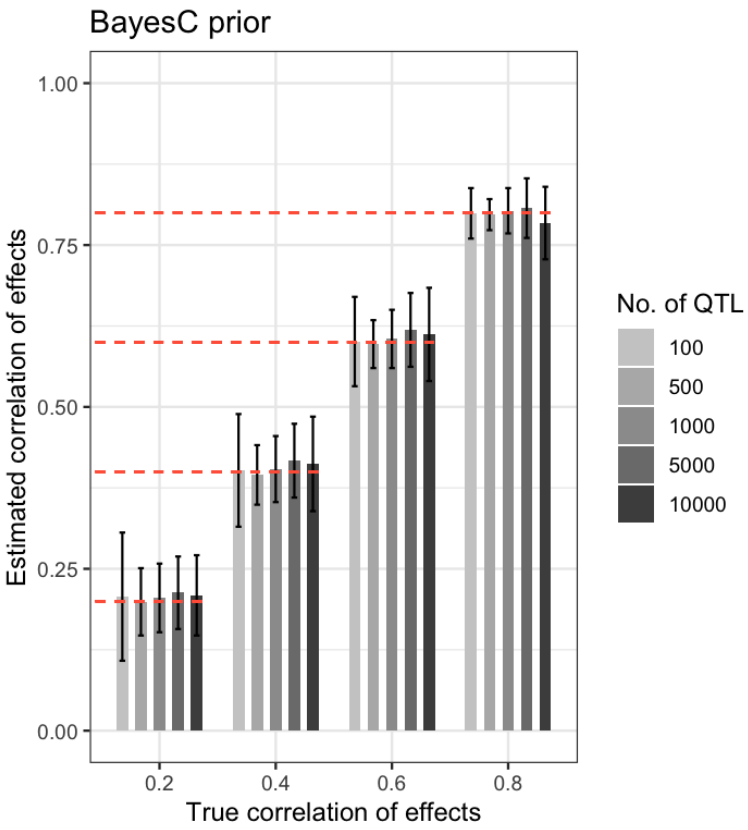
categorized by trait and in each plot, the estimated effect correlation of individual SNP effects (y-axis) is plotted against chromosome number (x-axis).

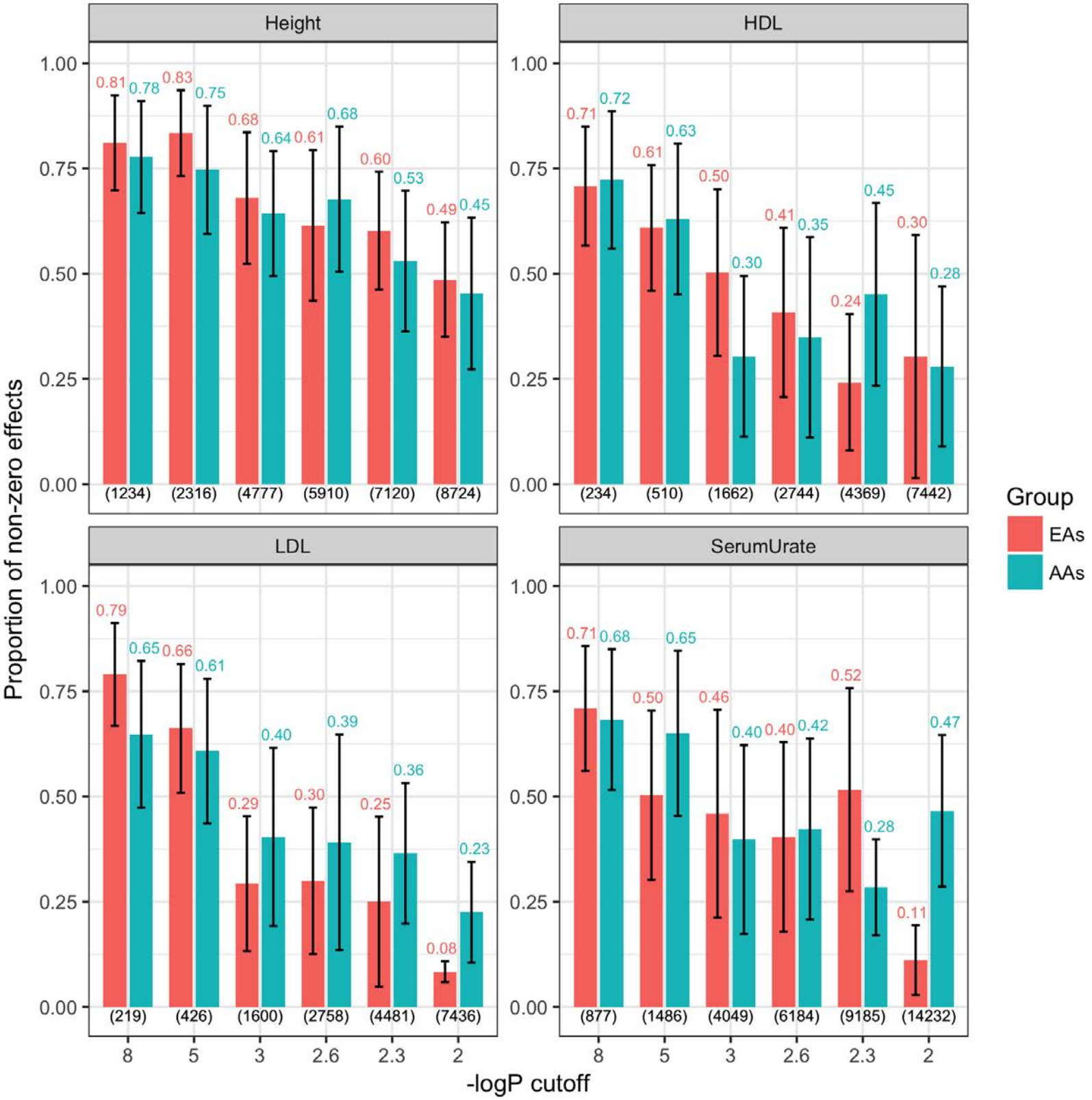

Gaussian prior
Estimated proportion of variance explained
1.00
0.75
0.50
0.25
0.00
EAs
AAs
0.2
0.4
0.6
0.8
0.2
0.4
0.6
0.8
True correlation of effects
No. of QTL
100
500
1000
5000
10000

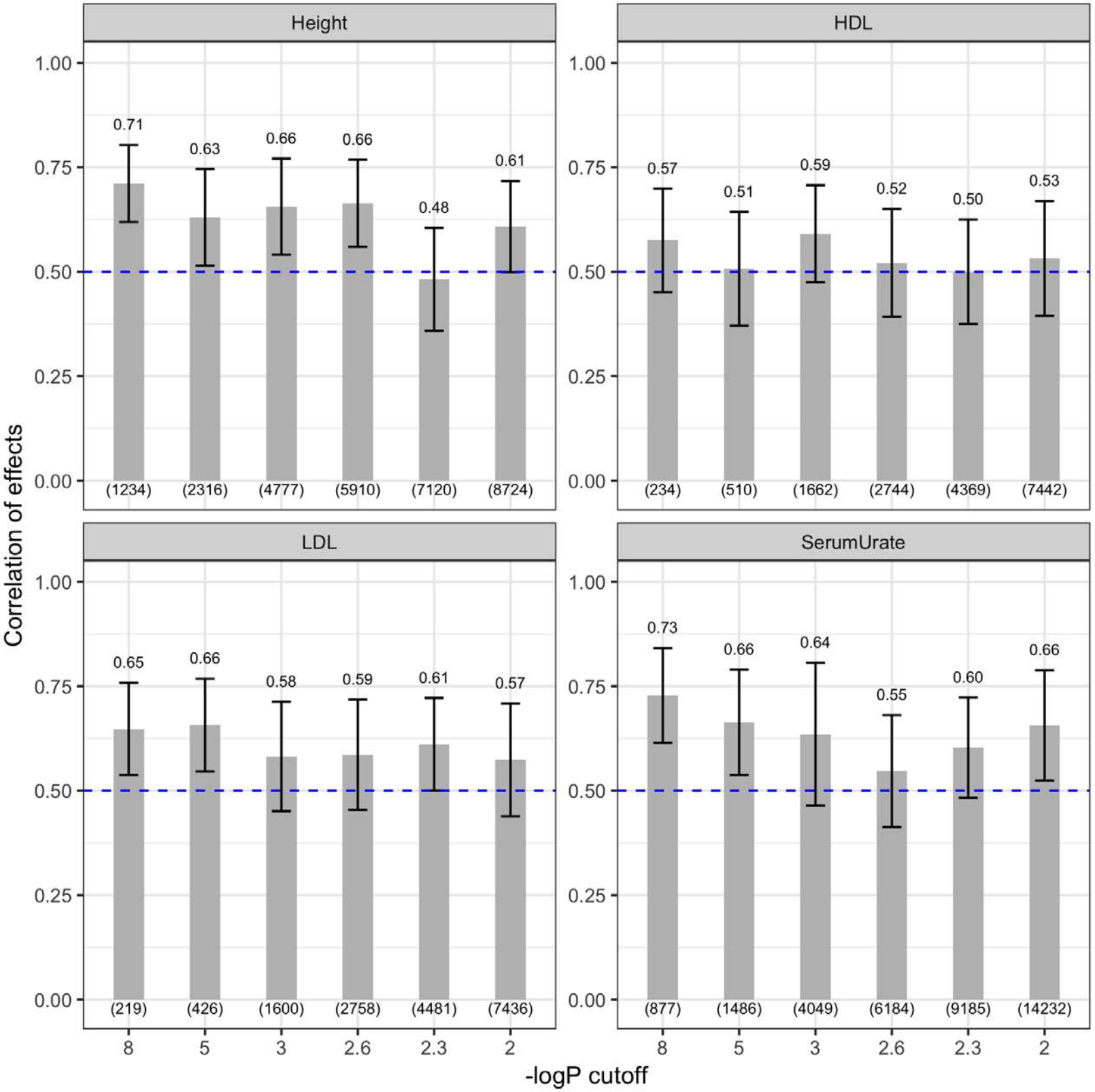

BayesC prior
Estimated proportion of variance explained
1.00
0.75
0.50
0.25
0.00
EAs
AAs
0.2
0.4
0.6
0.8
0.2
0.4
0.6
0.8
True correlation of effects
No. of QTL
100
500
1000
5000
10000

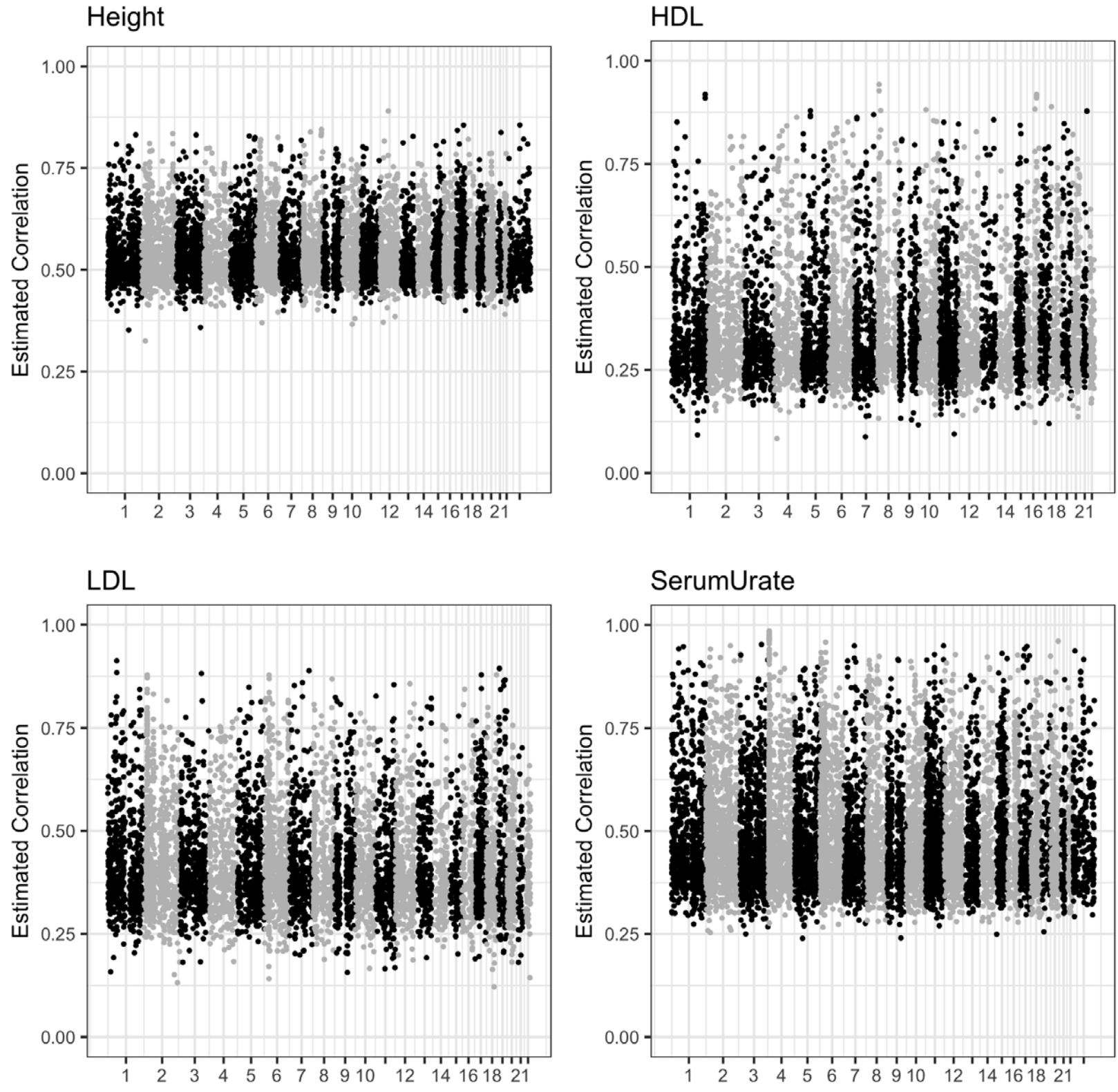

Height
HDL
LDL
SerumUrate
Proportion of variance explained
0.0
0.1
0.2
0.3
0.12
0.08
0.15
0.10
0.18
0.14
0.19
0.13
0.19
0.15
0.21
0.12
(1234)
(2316)
(4777)
(5910)
(7120)
(8724)
0.07
0.03
0.08
0.05
0.08
0.08
0.10
0.09
0.10
0.11
0.11
0.14
(234)
(510)
(1662)
(2744)
(4369)
(7442)
0.07
0.05
0.08
0.05
0.08
0.07
0.09
0.07
0.09
0.11
0.11
0.12
(219)
(426)
(1600)
(2758)
(4481)
(7436)
0.07
0.06
0.06
0.07
0.07
0.08
0.07
0.11
0.08
0.10
0.09
0.10
(877)
(1486)
(4049)
(6184)
(9185)
(14232)
8
5
3
2.6
2.3
2
-logP cutoff
Group
EAs
AAs


Gaussian prior
1.00
0.75
0.50
0.25
0.00
Estimated correlation of effects
0.2
0.4
0.6
0.8
True correlation of effects
No. of QTL
100
500
1000
5000
10000


BayesC prior
1.00
0.75
0.50
0.25
0.00
Estimated correlation of effects
0.2
0.4
0.6
0.8
True correlation of effects
No. of QTL
100
500
1000
5000
10000


Height
HDL
LDL
SerumUrate
Proportion of non-zero effects
1.00
0.75
0.50
0.25
0.00
0.81 0.78
0.83 0.75
0.68 0.64
0.61 0.68
0.60 0.53
0.49 0.45
(1234) (2316) (4777) (5910) (7120) (8724)
0.71 0.72
0.61 0.63
0.50 0.30
0.41 0.35
0.24 0.45
0.30 0.28
(234) (510) (1662) (2744) (4369) (7442)
0.79 0.65
0.66 0.61
0.29 0.40
0.30 0.39
0.25 0.36
0.08 0.23
(219) (426) (1600) (2758) (4481) (7436)
0.71 0.68
0.50 0.65
0.46 0.40
0.40 0.42
0.52 0.28
0.11 0.47
(877) (1486) (4049) (6184) (9185) (14232)
8 5 3 2.6 2.3 2
-logP cutoff
Group
EAs
AAs


Height
HDL
LDL
SerumUrate
Correlation of effects
1.00
0.75
0.50
0.25
0.00
0.71
0.63
0.66
0.66
0.48
0.61
(1234)
(2316)
(4777)
(5910)
(7120)
(8724)
0.57
0.51
0.59
0.52
0.50
0.53
(234)
(510)
(1662)
(2744)
(4369)
(7442)
0.65
0.66
0.58
0.59
0.61
0.57
(219)
(426)
(1600)
(2758)
(4481)
(7436)
0.73
0.66
0.64
0.55
0.60
0.66
(877)
(1486)
(4049)
(6184)
(9185)
(14232)
8
5
3
2.6
2.3
2
-logP cutoff


Height
HDL
LDL
SerumUrate
Estimated Correlation
1.00
0.75
0.50
0.25
0.00
1 2 3 4 5 6 7 8 9 10 12 14 16 18 21