Biostatistics (2022) **23**, 2, *pp.* 467–484
doi: 10.1093/biostatistics/kxaa035
Advance Access publication on September 19, 2020

Bayesian sparse heritability analysis with high-dimensional neuroimaging phenotypes

YIZE ZHAO*

Department of Biostatistics, Yale University, 300 George Street, New Haven, CT 06511, USA
yize.zhao@yale.edu

TENGFEI LI

Department of Radiology, University of North Carolina at Chapel Hill, 101 Manning Dr, Chapel Hill, NC 27514, USA

HONGTU ZHU

Department of Biostatistics, University of North Carolina at Chapel Hill, 135 Dauer Drive, Chapel Hill, NC 27514, USA

Summary

Heritability analysis plays a central role in quantitative genetics to describe genetic contribution to human complex traits and prioritize downstream analyses under large-scale phenotypes. Existing works largely focus on modeling single phenotype and currently available multivariate phenotypic methods often suffer from scaling and interpretation. In this article, motivated by understanding how genetic underpinning impacts human brain variation, we develop an integrative Bayesian heritability analysis to jointly estimate heritabilities for high-dimensional neuroimaging traits. To induce sparsity and incorporate brain anatomical configuration, we impose hierarchical selection among both regional and local measurements based on brain structural network and voxel dependence. We also use a nonparametric Dirichlet process mixture model to realize grouping among single nucleotide polymorphism-associated phenotypic variations, providing biological plausibility. Through extensive simulations, we show the proposed method outperforms existing ones in heritability estimation and heritable traits selection under various scenarios. We finally apply the method to two large-scale imaging genetics datasets: the Alzheimer's Disease Neuroimaging Initiative and United Kingdom Biobank and show biologically meaningful results.

Keywords: ADNI; Bayesian hierarchical selection; Dirichlet process; Heritability; Imaging genetics; Ising model; UK Biobank.

1. Introduction

The concept of heritability describes the aggregate of genetic signals within human complex traits and has been used to prioritize downstream analyses when phenotypes are in large scale. To estimate heritability (or narrow-sense heritability h^2), twin or family studies was originally required with additive genetic variation

*To whom correspondence should be addressed.

captured by pedigree information (Almasy and Blangero, 1998). Alternatively, using unrelated individual samples, genome-wide complex trait analysis (GCTA) was developed (Yang *and others*, 2010, 2011) to calculate heritability based on all common single nucleotide polymorphisms (SNPs), and becomes popular due to its flexibility and is widely applicable to the emerged large-scale data sources.

Neuroimaging traits on brain structure and function are among the most appealing yet not well-studied phenotypes. Understanding how genetic underpinning impacts human brain variation will provide great insight to elucidate etiological mechanisms in healthy or disease brain. Different from disease diagnosis or other types of quantitative trait, brain measurements are often high-dimensional and correlated. However, most of the previous analytical models (Yang *and others*, 2011; Ge *and others*, 2015) focused on univariate phenotype without consideration of phenotypic correlation. More recently, a few methods have been proposed to estimate heritability for multivariate traits (Ge *and others*, 2016; Luo *and others*, 2019). Though these works showed promising results under low phenotypic dimensions, they are not feasible under high-dimensional phenotypes due to the estimation of gigantic covariance matrix.

In addition, environmental factors and genetic effects combine in distinct ways to explain different brain units (Fjell *and others*, 2015; Roshchupkin *and others*, 2016) and they separately play a dominant role in different parts of the brain (Polderman *and others*, 2015). Though the underlying biological mechanism is still under investigated, we expect the whole brain wise heritability should contain sparsity where a range of environmentally driven brain areas are weakly and insignificantly heritable. Furthermore, since we consider imaging measurements at a fine voxel-level, it is anticipated the variance explained by SNPs are highly correlated among neighborhood voxels, making it spatially clustered and contiguous. Motivated by these considerations, we develop a unified Bayesian nonparametric heritability analysis under mixed effect model for large-scale neuroimaging phenotypes with simultaneous heritable traits selection. Specifically, we impose sparsity for SNP-associated phenotypic variations and encourage grouping effect for the nonzero SNP heritabilities under a nonparametric Dirichlet process ($\mathcal{DP}$) mixture of inverse gamma distributions. Though there is a broad literature in Bayesian variable selection, a large proportion of the works focus on selection among high-dimensional covariates under fixed effect models (Zhang *and others*, 2016; Teng *and others*, 2019). More recent years, selection of random effects or joint selection of mixed effects receives growing attention under application of longitudinal and repeated measurements data (Ibrahim *and others*, 2011; Cai and Bandyopadhyay, 2017). However, the problem of our interest is fundamentally different from these works in the following two aspects: (i) we consider high-dimensional outcomes and (ii) we focus more on estimating and selecting variance components but less on covariates.

Another unique feature for imaging phenotypes is their underlying biological structure. Besides the spatial correlation among neighborhood voxels, brain connectivities among regions of interest (ROIs) are also widespread metrics to capture regional dependence, where ROIs and voxels form a group structure. Group or hierarchical selection as a popular concept for selecting structural covariates has been extensively adopted in regression settings to achieve sparsity between and within feature groups (Zhang *and others*, 2014; Zhao *and others*, 2016). To achieve plausibly hierarchical sparsity among phenotypic heritability, we introduce two nested sets of selection indicator at region level and voxel level, where selection indicators are assigned to individual or grouped variance components. To incorporate region-wise connectivity and voxel-wise dependence, we choose Ising or binary Markov random field priors (Li and Zhang, 2010) to encourage coupling effect on selecting biologically dependent voxels/ROIs.

Our major contributions in this work are several-fold. First, we are among the very first to conduct SNP-based heritability analysis for high-dimensional phenotypes under unrelated individuals. We apply our method to study the integrative genetic contribution on large-scale brain phenotypes under both the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the United Kingdom (UK) biobank datasets. Second, we impose hierarchical sparsity among region-level and voxel-level heritabilities which receives less attention in literature but dramatically helps to improve estimation accuracy and computational efficiency. Third, we incorporate structural connectivity among ROIs and spatial dependence among voxels

within the selection procedure by smoothing the selection indicators, leading to biological plausibility. Lastly, we induce grouping effect among SNP-associated phenotypic variations using Bayesian nonparametric to further improve detection power as well as reduce dimension. Eventually, the whole integrative procedure will help us establish a biologically meaningful and interpretable heritability brain map, which will offer a great potential for future genetically informative endophenotype construction.

The remainder of the article is organized as follows. In Section 2, we present the heritability analysis and our extension to high-dimensional neuroimaging phenotypes. In Section 3, we present our method and corresponding Markov chain Monte Carlo (MCMC) algorithm. We conduct simulation studies to assess the performance of our method and competing ones in Section 4 followed by two real data applications in Section 5. Finally, we conclude with a discussion in Section 6.

2. Model specification

We formulate the heritability analysis for an imaging genetics study via random effect models. Assuming for each subject, we obtain neuroimaging phenotypes across S voxels and genetic variants over P whole genome SNPs. At each voxel s, we have

$$Y(s) = \mathbf{W}V(s) + e(s), \quad s = 1, \ldots, S. \tag{2.1}$$

Here, $Y(s)$ is a mean-centered $N \times 1$ vector of phenotypes across N subjects at voxel s, and $\mathbf{W} = (w_{ij})_{N\times P}$ is the standardized genotype matrix summarized pairwisely under individual SNP data as $w_{ij} = \frac{z_{ij}-f_j}{\sqrt{2f_j(1-f_j)}}$, with z_{ij} the number of copies of the reference allele of subject i and SNP j, and f_j the reference allele frequency for SNP j. The additive genetic effects are modeled as random effects denoted as $V(s)$ which is a vector of P SNP effects, and the corresponding residual effects are denoted as $e(s)$. We also assume $V(s)$ and $e(s)$ are independent across genetic loci and subjects, and $V(s) \sim N(0, \mathbf{I}\tau_u^2(s))$ and $e(s) \sim N(0, \mathbf{I}\sigma^2(s))$, respectively. Here, we focus our attention on the unrelated subject studies without the existence of a shared random residual component. Hence, for each phenotype, $\tau_g^2(s) = P\tau_u^2(s)$ indicates the variance explained by the additive genetic effects from the genome-wide common SNPs, and $\sigma^2(s)$ denotes the variance from subject-specific environment. For the ease of illustration, we do not adjust covariates (e.g., age, gender, sex) at this moment, and we will discuss the general case in the later part.

Based on model (2.1), the phenotypic covariance matrix for neuroimaging voxel s becomes

$$\mathrm{Var}\{Y(s)\} = \tau_u^2(s)\mathbf{W}\mathbf{W}^T + \sigma^2(s)\mathbf{I} = \tau_g^2(s)\mathbf{R} + \sigma^2(s)\mathbf{I} := \mathcal{V}(\tau_g, \sigma), \tag{2.2}$$

where $\mathbf{R} = \mathbf{W}\mathbf{W}^T/P$ is the empirical genetic relationship matrix (GRM) among subjects. To further explore additive genetic heritability effect, we define the heritability for phenotypes s as the proportion of genetic variation to the intersubject phenotypic variation: $h(s) = \frac{\tau_g^2(s)}{\tau_g^2(s)+\sigma^2(s)}$, which is nonnegative over $s = 1, \ldots, S$. Due to the topological configuration of human brain, the heritability estimates across brain locations are highly correlated and their spatial dependence should be carefully modeled. In order to construct a practically meaningful realization for $h(\cdot)$, our attentions land on the following four aspects. First, due to the brain anatomical structure, the genetic impact capturing by $\tau_g^2(\cdot)$ on neighborhood voxels should be highly correlated, leading to localized grouping effect. Second, under such a fine phenotype scale, it is biologically more meaningful to assume a similar magnitude of $\tau_g^2(\cdot)$ upon correlated voxels to reflect the topological smoothness of neurogenetic procedure. Third, besides voxel-wise correlation, it is well known the existence of anatomical connections among ROIs. Given their interactions, phenotypes belong to the connected brain regions are more likely to be coupling on their heritable status. Finally, to improve statistical power and biological interpretability, it is meaningful to assume $\tau_g^2(\cdot)$ over whole

brain voxels contain sparsity with certain proportion of environmentally driven imaging traits impacted marginally by the additive genetic effects. Particularly, given the atlas-based brain parcellations are genetically noninformative, it is unlikely certain atlas-based ROI will be significantly associated with additive genetic variations as a whole. Thus, our sparsity assumption is placed at both within and among ROI levels. Given these considerations, we propose an integrative Bayesian model to simultaneously estimate the heritabilities over whole brain phenotypes and identify heritable brain locations. It is worth noting that to some extent, model (2.1) can be viewed as a imaging-on-scalar random effects regression. However, different from most of the previous imaging-on-scalar models, our focus is on a more biologically plausible representation of the variance components rather than the association between imaging outcome and predictors.

3. BAYESIAN SPARSE HERITABILITY ANALYSIS

3.1. *Phenotype selection*

Our goal is to model heritability function $h(\cdot)$ under brain structural driven sparsity and smoothness. To impose sparsity on genetic effect, we introduce a latent indicator function $\gamma(\cdot)$ over brain with $\gamma(s) \in \{0, 1\}$ indicating the existence of significant genetic effect at voxel s. The total genetic variance $\tau_g^2(s)$ can be further written as

$$\tau_g^2(s) = \gamma(s) \times \eta^2(s); \qquad s = 1, \ldots, S, \tag{3.3}$$

where $\eta^2(s)$ represents the nonzero variance explained by common genetic variants once phenotype measured at voxel s is heritable. In the case of $\gamma(s) = 0$, we have $\tau_g^2(s) = 0$ indicating a negligible genetic effect leading to $h(s) = 0$.

As mentioned previously, the brain anatomical architecture contains inherent hierarchies with ROIs and voxels formed a nested group and within-group structure. Therefore, we first split whole brain latent indicate set $\boldsymbol{\gamma} = \{\gamma(1), \ldots, \gamma(S)\}^T$ into region-level indicators $C = (c_1, \ldots, c_R)^T$ where c_r describes the heritable status of region r, and voxel-level selection indicators $U = (U_1^T, \ldots, U_R^T)$; $U_r = (u_{r1}, \ldots, u_{rS_r})^T$, where S_r is the total number of voxels in region r, and u_{rg} denotes the heritable status for voxel g within region r. We summarize their relationship

$$\boldsymbol{\gamma} = (\mathbf{M} \cdot C) \circ U, \tag{3.4}$$

where $\mathbf{M} = (m_{sr})$ represents the mapping matrix from voxels to ROIs with $m_{sr} = 1$ when voxel s belongs to region r, zero otherwise; and $\circ$ denotes the entry-wise product. Based on (3.4), we can locate each neuroimaging phenotype s to voxel g within region r; and phenotype s will be excluded from the model when at least one of c_r and u_{rg} is zero.

From model formulation perspective, (3.3) and (3.4) essentially achieve the so-called "structured sparsity" (Chen *and others*, 2016) by assuming the active components are from certain structure. Computationally, incorporating group information into selection will improve stochastic signal searching. As we can see from (3.4), at each iteration of the posterior algorithm, for elements within C that have been sampled to zero, we can directly mark all their child-level phenotypes as insignificance. In other words, our modeling strategy allows an intra-model dynamic screening procedure via the value of C during posterior simulation. With $R \ll S$ and the existence of sparsity, we are able to improve the parameter estimation under less computation.

Another advantage of model formulation (3.4) is we can incorporate hierarchical brain structural information in a more explicit way. Particularly, at region level, we extract a graph $\mathcal{G}_1 = (V_1, E_1)$ based on brain structural connectivity, where $V_1 = \{1, \ldots, R\}$ is the set of ROIs and $E_1 = \{(r_1, r_2) : r_1, r_2 \in$

$V_1, r_1 \neq r_2, r_1 \sim r_2\}$ is the set of edges with $r_1 \sim r_2$ indicating a connection in the structural network. Similarly, we construct a graph over voxels within each region $\mathcal{G}_{r2} = (V_{r2}, E_{r2}); r = 1, \ldots, R$, with voxel set $V_{r2} = \{\sum_{g<r} S_g + 1, \ldots, \sum_{g \leq r} S_g\}$ and edge set $E_{r2} = \{(s_1, s_2) : s_1, s_2 \in V_{r2}, s_1 \neq s_2, s_1 \sim s_2\}$ with edges defined by between each voxel and its six nearest adjacent (top, bottom, left, right, front, back) neighbors. Denote the symmetric binary adjacency matrix for $\mathcal{G}_1$ as $\mathbf{A}_1$, and those for $\mathcal{G}_{r2}$ as $\mathbf{A}_{r2}$; $r = 1, \ldots, R$. We then resort to Ising model, a binary-variable graphical model for latent indicators to incorporate structural configuration among brain regions and voxels as

$$p(C) \propto \text{Ising}(C, \mu_1, \phi_1, \mathbf{A}_1); \quad p(U_r) \overset{i.i.d}{\propto} \text{Ising}(U_r, \mu_2, \phi_2, \mathbf{A}_{r2}); r = 1, \ldots, R, \tag{3.5}$$

where vector $V = (v_1, \ldots, v_m) \propto \text{Ising}(\mu, \phi, \mathbf{A}) \propto \exp(\mu \sum v_m + \phi \sum_m \sum_{m'} \mathbf{A}(m, m') I[v_m = v_{m'}])$, with $I[\cdot]$ the indicator function. In models (3.5), parameters $\boldsymbol{\mu} = (\mu_1, \mu_2)$ control the sparsity of selection indicators, i.e., how many regions or voxels will be treated as heritable and included in the model; and parameters $\boldsymbol{\phi} = (\phi_1, \phi_2)$ impact the smoothness of indicator values over graphs, i.e., how strong belief we have to include or exclude the connected regions or voxels simultaneously. In the literature, a number of different strategies can be used to determine the value of sparsity and smoothness parameters (Li and Zhang, 2010; Huang *and others*, 2013). Here, we borrow the idea of Teng *and others* (2019) to first decide a proper range of Ising parameters to avoid phase transition, and then conduct sensitivity analyses under different choices of hyperparameters (see Sections S1 and S4 of the supplementary materials available at *Biostatistics* online). When we ignore the brain structural information, all the elements of $\mathbf{A}_1$ and $\mathbf{A}_{r2}, r = 1 \ldots, R$ will become zero with the second term in the *r.h.s* of each model in (3.5) eliminated, then the prior for C and U will become *i.i.d* Bernoulli distribution, a special case of Ising model without any coupling effect.

3.2. *Heritability estimation*

After imposing sparsity, we further consider the estimation of genetic variance component $\tau_g^2(\cdot)$ and environmental variance component $\sigma^2(\cdot)$. From practical point of view, we assume each individual environmental variance independently follows an Inverse Gamma prior $\sigma^2(s) \sim \text{IG}(a_1, b_1), s = 1, \ldots, S$. In term of the genetic component $\tau_g^2(\cdot)$, as shown in (3.3), the first component $\gamma(\cdot)$ identifies heritable neuroimaging phenotypes with smoothing effect among selection indicators driven by brain structural information given by (3.4) and (3.5). We are now focusing on the second component $\eta^2(\cdot)$, which captures the nonzero genetic effect; and as mentioned previously, it should be spatially clustered and contiguous over brain.

We first assume at each brain location s, $\eta^2(s) \sim \mathcal{F}$ with $\mathcal{F}$ a predefined probability function. To fully achieve a grouping effect by learning from the data, we assign a nonparametric $\mathcal{DP}$ prior on $\mathcal{F}$

$$\mathcal{F} \sim \mathcal{DP}(\mathcal{F}_0, \alpha), \tag{3.6}$$

with base measure $\mathcal{F}_0$ defining the expectation of the random probability on $(0, +\infty)$, and scalar parameter α describing the variance. To conduct posterior inference, among different representations on $\mathcal{DP}$, we choose the infinite mixtures of point masses realization of $\mathcal{F}$ (Sethuraman, 1994), where two separate sets of independent and identically distributed (*i.i.d*) random variables $\{\pi_k'\}_{k=1}^{\infty}$ and $\{\theta_k\}_{k=1}^{\infty}$ are introduced with

$$\pi_k' \mid \mathcal{F}_0, \alpha \sim \text{Beta}(1, \alpha); \qquad \theta_k \mid \mathcal{F}_0, \alpha \sim \mathcal{F}_0; \qquad k = 1, \ldots, \infty. \tag{3.7}$$

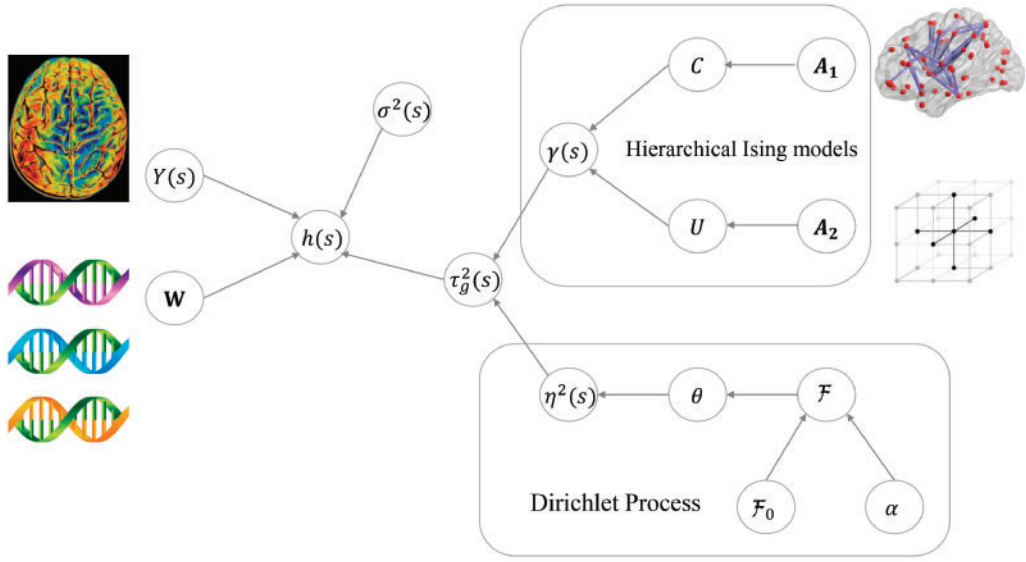

Fig. 1. A demostration of the Bayesian sparse heritability analysis (BSHA) modeling framework.

Then, the probability function $\mathcal{F}$ can be written as

$$\mathcal{F} = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k}, \qquad \text{with} \quad \pi_k = \pi_k' \prod_{h=1}^{k-1} (1 - \pi_h'), \tag{3.8}$$

where δ_θ is a probability measure concentrated at θ. Here, $\{\theta_k\}_{k=1}^{\infty}$ is a sequence of independent draws from $\mathcal{F}_0$ with its elements partitioning the natural number set into clusters with individually identical value; and $\{\pi_k\}_{k=1}^{\infty}$ is the weight parameter constructed via a stick-breaking distribution. With probability one, $\mathcal{F}$ is a discrete distribution consisting of infinite number of point masses.

The discrete nature of $\mathcal{DP}$ induces grouping for the active genetic effect in a nonparametric way. As it is clearly revealed in (3.7) and (3.8)—underweight $\{\pi_k\}_{k=1}^{\infty}$, samples of $\eta^2(\cdot)$ will be obtains directly from discrete distribution $\mathcal{F}$ consisting of components $\{\theta_k\}_{k=1}^{\infty}$, which drawn randomly from the base measure $\mathcal{F}_0 = \text{IG}(a_0, b_0)$. The scalar parameter α further controls the concentration of clustering, and we assign a noninformative hyper-prior G(1, 1) to allow enough flexibility. With the increase of k, π_k decreases exponentially concentrating the sampling on a number of initial components. Therefore, each cluster in our case has the identical value of $\eta^2(\cdot)$, which facilitates our biological insight that contiguous brain locations share similar genetic explanations. At the same time, with one cluster sharing the same genetic effect, the posterior computation is also reduced since we ultimately estimate $\eta^2(\cdot)$ for each cluster instead of individual phenotype.

Overall, our method successfully realizes integrative heritability analysis for large-scale neuroimaging traits with simultaneous sparse group selection of heritable phenotypes, smoothness over genetic effect and incorporation of hierarchical brain structural information. We name our method Bayesian sparse heritability analysis (BSHA) with a demonstration in Figure 1.

3.3. *Posterior inference*

We develop MCMC algorithm to conduct posterior inference for the proposed BSHA. For the ease of computational allocation, we first choose a value K to cover the upper limit of cluster number on $\eta^2(s)$

under $\mathcal{DP}$. As mentioned previously, the stick-breaking weighted format makes the sampling concentrate at initial clusters, and we can start with large K value and run MCMC multiple times to find a proper range. For each $\eta^2(s)$, we introduce a latent indicator $t(s) \in \{1,\ldots,K\}$ as its cluster index, where $t(s)$ should follow a multinomial (MN) distribution with K categories under weight $\boldsymbol{\pi} = (\pi_1,\ldots,\pi_K)$. Given imaging and genetics data, the joint conditional posterior distribution for unknown parameter/hyperparameter set $\boldsymbol{\Theta} = \big(\eta^2(\cdot), \sigma^2(\cdot), t(\cdot), \boldsymbol{\pi}, C, U\big)$ follows

$$\pi\big(\boldsymbol{\Theta} \mid Y(\cdot), \mathbf{R}\big) \propto \pi(Y(\cdot), \mathbf{R} \mid \boldsymbol{\Theta})\pi(\eta^2(\cdot) \mid t(\cdot), \boldsymbol{\pi})\pi(\sigma^2(\cdot))\pi(t(\cdot))\pi(\boldsymbol{\pi})\pi(C, U),$$

under which Metropolis–Hastings (MH) and Gibbs are used for posterior computation following:

Step 0:

Assign random initials to each element of $\boldsymbol{\Theta}$;

Step 1:

For $s = 1,\ldots,S$, if $\gamma(s) = 0$, update $\sigma^2(s)$ from $\mathrm{IG}(a_1 + N/2, b_1 + Y'(s)Y(s)/2)$; else if $\gamma(s) = 1$, sample a proposal $\widetilde{\sigma^2}(s)$ based on $\log(\widetilde{\sigma^2}(s)) = \log(\sigma^2(s)) + \epsilon$ with ϵ a mean zero Normal distribution, then calculate $R_\sigma = \exp\Big[-\frac{1}{2}Y'\{\mathcal{V}(\tau_g, \widetilde{\sigma})^{-1} - \mathcal{V}(\tau_g, \sigma)^{-1}\}Y - \frac{b_1}{\widetilde{\sigma^2}} + \frac{b_1}{\sigma^2}\Big] \frac{|\mathcal{V}(\tau_g,\sigma)|^{\frac{1}{2}}}{|\mathcal{V}(\tau_g,\widetilde{\sigma})|^{\frac{1}{2}}} \cdot \big(\frac{\widetilde{\sigma^2}}{\sigma^2}\big)^{a_1}$, with a probability $\min(1, R_\sigma)$ to accept $\widetilde{\sigma^2}(s)$.

Step 2:

For $s = 1,\ldots,S$, if $\gamma(s) = 0$, sample $\eta^2(s)$ from $\mathrm{IG}(a_0, b_0)$; else if $\gamma(s) = 1$, denote $\eta_k^2 = \{\eta^2(s); z(s) = k\}$ as the unified genetic variance for group $k, k = 1,\ldots,K$. Sample a proposal $\widetilde{\eta_k^2}$ based on $\log(\widetilde{\eta_k^2}) = \log(\eta_k^2) + \epsilon'$ with ϵ' a mean zero Normal distribution, then calculate $R_\eta = \exp\Big[\sum_s I[z(s) = k]\big(-\frac{1}{2}Y'\{\mathcal{V}(\widetilde{\eta_k}, \sigma)^{-1} - \mathcal{V}(\eta_k, \sigma)^{-1}\}Y\big) - \frac{b_0}{\widetilde{\eta_k^2}} + \frac{b_0}{\eta_k^2}\Big] \cdot \frac{|\mathcal{V}(\eta,\sigma)|^{\frac{\sum_s I[z(s)=k]}{2}}}{|\mathcal{V}(\eta,\widetilde{\sigma})|^{\frac{\sum_s I[z(s)=k]}{2}}} \cdot \big(\frac{\widetilde{\eta_k^2}}{\eta_k^2}\big)^{a_0}$, with a probability $\min(1, R_\tau)$ to accept $\widetilde{\tau_k^2}$. Based on the accepted $\eta^2(s)$, we can calculate $\tau_g^2(s)$ based on (3.3).

Step 3:

For $s = 1,\ldots,S$, if $\gamma(s) = 0$, sample $z(s)$ from $\mathrm{MN}(\pi_1,\ldots,\pi_K)$; else if $\gamma(s) = 1$, sample $z(s)$ from $\mathrm{MN}(q_1,\ldots,q_K)$ with $q_k = \frac{\pi_k \exp(-\frac{1}{2}Y'\{\mathcal{V}(\tau_k,\sigma)^{-1}\}Y)|\mathcal{V}(\tau_k,\sigma)|^{\frac{1}{2}}}{\sum_{h=1}^K \pi_h \exp(-\frac{1}{2}Y'\{\mathcal{V}(\tau_h,\sigma)^{-1}\}Y)|\mathcal{V}(\tau_k,\sigma)|^{\frac{1}{2}}}$.

Step 4:

For $k = 1,\ldots,K-1$ sample π_k' from $\mathrm{Beta}(1 + d_k, \alpha + \sum_{j>k} d_j)$ where $d_k = \sum_s I[z(s) = k]$, and $\pi_K = 1$. We then have $\pi_k = \pi_k' \prod_{h<k}(1 - \pi_h')$.

Step 5:

For $r = 1,\ldots,R$, define $l_1(c_r) := \exp(\mu_1 c_k + \phi_1 \sum_{r'} A_1(r, r')I[c_r = c_{r'}]) \prod_{s=1}^m \big[\exp(-\frac{1}{2}Y'\{\mathcal{V}(\tau_g c_r, \sigma)^{-1}\}Y) \mid \mathcal{V}(\tau_g c_r, \sigma) \mid^{\frac{1}{2}}\big]^{m_{rs}}$. Then sample c_r from $\mathrm{Bern}(\frac{l_1(1)}{l_1(1)+l_1(0)})$.

Step 6:

For $r = 1,\ldots,R; s = 1,\ldots,S_r$, define $l_2(u_{rs}) := \exp(\mu_2 u_{rs} + \phi_2 \sum_{s'} A_{r2}(s, s')I[u_{rs} = u_{rs'}]) \exp(-\frac{1}{2}Y'\{\mathcal{V}(\tau_g u_{rs}, \sigma)^{-1}\}Y) \mid \mathcal{V}(\tau_g u_{rs}, \sigma) \mid^{\frac{1}{2}}$. Then sample u_{rs} from $\mathrm{Bern}(\frac{l_2(1)}{l_2(1)+l_2(0)})$.

During implementation, we repeat steps 1–6 iteratively. Eventually, our goal of interest is to estimate brain voxel-wise heritability. So at each iteration of MCMC, we calculate $h(s)$ based on posterior samples of $\sigma^2(s)$ and $\tau_g^2(s)$ for $s = 1, \ldots, S$, and use posterior median as suggested by Xu *and others* (2015) to obtain heritability estimates which are sparse with zero indicating an heritably inactive brain phenotype.

3.4. *Adjusting covariates*

When there are nongenetic covariates $\mathbf{X}_{N\times Q}$, e.g. age, gender, needed to be adjusted, model (2.1) becomes voxel-specific mixed effect model by adding $\mathbf{X}_{N\times Q}$ as fixed effects. To remove the design variables, we adopt a strategy which is similar to that considered by Ge *and others* (2016). Specifically, we introduce $\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ which is a symmetric and idempotent matrix with rank $N - Q$. Matrix $\mathbf{P}$ can be further decomposed as $\mathbf{P} = \mathbf{T}^T\mathbf{T}$ where $\mathbf{T}$ is a $(N - Q) \times N$ matrix satisfying $\mathbf{T}\mathbf{T}^T = \mathbf{I}$ and $\mathbf{T}\mathbf{X}^T = \mathbf{0}$. Eventually, by projecting data via matrix $\mathbf{T}$, we can remove the fixed effect term with the model becomes $\mathbf{T}Y(s) = \mathbf{T}\mathbf{W}V(s) + \mathbf{T}e(s), s = 1, \ldots, S$. Therefore, once we update the phenotype at each brain voxel to $\widetilde{Y}(s) = \mathbf{T}Y(s)$ and GRM to $\widetilde{\mathbf{G}} = \mathbf{T}\mathbf{G}\mathbf{T}^T$, the heritability estimate will follow the same procedure as that for model (2.1).

4. SIMULATION STUDIES

We conduct simulation studies to evaluate the finite sample performance of the proposed method in heritability estimation and heritable phenotypes selection. To mimic real data, we randomly sample 1000 GRM from the ADNI database (ADNI 1/Go/2 phases), and consider a 3D cubic as phenotypes with the total voxel number 64 000($40 \times 40 \times 40$). To construct structural information, we partition the cubic into 64 equally sized adjacent regions with 1000 voxels in each one, and generate a scale-free network within the regions. In terms of heritabilities, we generate $\sigma^2 \sim \text{IG}(5, 10)$ for a low variance case and $\sigma^2 \sim \text{IG}(0.5, 1)$ for a high variance one. For genetic component, we consider the following scenarios for the locations and values.

Scenario 1: We set four regions to be significantly heritable based on network configuration. Within each of them, we randomly assign 100 spatially contiguous voxels to be true signals and denote their index set as $\mathcal{R}_1$. We then generate $\tau_g^2(\mathcal{R}_1) \sim \text{IG}(5, 10)$ or $\tau_g^2(\mathcal{R}_1) \sim \text{IG}(0.5, 1)$ for two variance cases to keep a balanced scale for the two variance components.

Scenario 2: We randomly set four regions to be significantly heritable without consideration of network configuration. Within each them, following Scenario 1, we randomly assign 100 spatially contiguous voxels to be true signals with their index set $\mathcal{R}_2$. We then generate $\tau_g^2(\mathcal{R}_2) \sim \text{IG}(5, 10)$ or $\tau_g^2(\mathcal{R}_2) \sim \text{IG}(0.5, 1)$ for two variance cases.

Scenario 3: We randomly set eight pieces of significantly heritable areas based on network configuration. Within each of them, we randomly assign 50 spatially contiguous voxels to be true signals with their index set $\mathcal{R}_3$. We then generate $\sigma_g^2(\mathcal{R}_3) \sim \text{IG}(5, 10)$ (or $\tau_g^2(\mathcal{R}_3) \sim \text{IG}(0.5, 1)$).

Scenario 4: We follow the signal pattern in scenario 1 to construct $\mathcal{R}_1$. Differently, we directly set heritability $h(\mathcal{R}_1) = 0.7$ for an identically heritable effect within significant voxels, which results in $\tau_g^2(\mathcal{R}_1)$ not following any distribution.

Scenario 5: We follow the signal pattern in scenario 2 to construct $\mathcal{R}_2$, and directly set $h(\mathcal{R}_2) = 0.7$ for the significant heritability.

Scenario 6: We follow the signal pattern in scenario 3 to construct $\mathcal{R}_3$, and directly set $h(\mathcal{R}_3) = 0.7$ for the significant heritability.

To implement BSHA, we assign noninformative priors by using $a_k = b_k = 1$ $(k = 0, 1)$ and $\mu_1 = \mu_2 = 0$. In terms of smoothness hyperparameters, we set $\phi_1 = \phi_2 = 1$ after sensitivity analyses. Since BSHA represents the very first for high-dimensional heritability analysis under unrelated subject

study, we apply three existing univariate phenotypic analysis: GCTA, massively expedited genome-wide heritability analysis (MEGHA) (Ge *and others*, 2015), moment matching method for SNP-based heritability estimation (MMHE) (Ge *and others*, 2017) on the data via pipelines provided by their authors. These three approaches work in a similar fashion by providing a point estimate of heritability and the corresponding p-value. To further assess different components of model construction in BSHA, we also consider three different variations of BSHA as competitors: (i) we replace the Ising priors on selection indicator sets by $i.i.d$ Bernoulli priors with parameter 0.5 and denote the method as BSHA1; (ii) we replace the $\mathcal{DP}$ prior on $\mathcal{F}$ by directly assigning $\mathcal{F} = \text{IG}(a_0, b_0)$ and denote the method as BSHA2; and (iii) we remove the regional sparsity assumption by only inserting voxel-level selection indicator and denote the method as BSHA3. For each scenario under both low and high variance cases, we generate 100 Monte Carlo datasets. The tuning parameter settings for BSHA1, 2, 3 follow those for BSHA, and we conduct MCMC with random initials for 5000 iterations with 2000 burn-in for the Bayesian approaches where the posterior convergence is checked by trace plots as well as Gelman–Rubin method (Gelman *and others*, 1992) on the average estimated heritability over heritable traits, number of selected heritable traits, and logarithm posterior likelihood. We calculate root mean square error (RMSE) for $h(s)$ to assess the heritability estimation, and use sensitivity and specificity to quantify heritable phenotype selection. All the results are summarized in Table 1.

Overall, our proposed BSHA outperforms the competing methods in heritability estimation and heritable phenotypes identification in almost all the settings. Specifically, under low variance scenarios, BSHA achieves a considerably better performance particularly in comparison with existing approaches. This reassures the necessity to consider the joint effect among phenotypes. When variance increases, though the performance of all the methods shows deterioration, BSHA stays as the winner in estimation and selection. We further study the strength of each single component in our model. Among three BSHA variations, BSHA1 disables the property to incorporate structural information compared to BSHA. As we can see from Table 1, when true signals are concentrated, regional network $\mathcal{G}_1$ contributes little to posterior inference indicated by similar performance for BSHA1 between scenarios 1 and 2 (4 and 5), highly likely due to the domination of hierarchical sparsity. However, when signals get scattered, a failure to incorporate network information will lead to significantly worse results showed by the comparison between BSHA1 and BSHA in scenarios 3 and 6. The advantage of nonparameter setting is demonstrated by the performance of BSHA2 which is comparable to that of BSHA when we generate true signals based on parametric model assumption (scenarios 1–3) but gets deteriorated when the assumption no longer holds. Lastly, the hierarchical sparsity construction as assessed by comparing BSHA3 and BSHA significantly impacts the result in our current simulation settings. Without regional sparsity, BSHA3 suffers with the worst selection and estimation performance among BSHA and its variations.

Of note, we also conduct additional simulations to investigate the impact of sample size on the performance of different methods and show the robustness of our method in small sample sizes. See Section S2 of the supplementary material available at *Biostatistics* online.

5. Data applications

5.1. *ADNI*

We focus on the baseline ADNI1 data in our analysis (www.adni-info.org) to illustrate our method in a disease sample. For data quality control, 745 Caucasian subjects are considered after removing the ones with sex check failure, more than 10% missing SNP, and outliers. To construct GRM, we first remove SNPs with (i) more than 5% missing values, (ii) minor allele frequency smaller than 5%, and (iii) Hardy–Weinberg equilibrium p-value $< 1e^{-6}$. A total of 421 823 SNPs are left eventually, and we calculate GRM among subject. In terms of phenotypes, the raw magnetic resonance imaging (MRI) data for each

Table 1. *Simulation results: RMSE (average root mean square error) for heritability estimation, sensitivity (Sens), and specificity (Spec) for heritable phenotype selection under different scenarios and variances. The Monte Carlo standard deviation for RMSE is included in the parentheses.*

		Low variance			High variance		
Scenario	Method	RMSE	Sens	Spec	RMSE	Sens	Spec
1	GCTA	0.184 (7.051e−04)	0.882	0.860	0.188 (7.222e−04)	0.605	0.860
	MEGHA	0.125 (5.932e−04)	0.744	0.950	0.126 (7.348e−04)	0.625	0.950
	MMHE	0.092 (5.724e−04)	0.833	0.897	0.092 (9.802e−04)	0.686	0.897
	BSHA1	0.029 (1.326e−03)	0.908	0.998	0.029 (1.816e−03)	0.815	0.997
	BSHA2	0.022 (1.402e−03)	0.904	0.999	0.023 (1.514e−03)	0.831	0.999
	BSHA3	0.050 (1.363e−03)	0.908	0.996	0.050 (2.737e−03)	0.735	0.996
	BSHA	0.024 (1.463e−03)	0.928	0.999	0.024 (1.954e−03)	0.828	0.999
2	GCTA	0.184 (6.276e−04)	0.881	0.860	0.187 (6.771e−04)	0.605	0.861
	MEGHA	0.125 (6.353e−04)	0.740	0.950	0.126 (5.546e−04)	0.623	0.950
	MMHE	0.092 (5.892e−04)	0.828	0.897	0.092 (1.343e−03)	0.684	0.897
	BSHA1	0.029 (1.347e−03)	0.883	0.997	0.029 (1.463e−03)	0.818	0.997
	BSHA2	0.022 (9.905e−04)	0.904	0.999	0.023 (1.924e−03)	0.855	0.999
	BSHA3	0.051 (1.283e−03)	0.911	0.997	0.050 (2.945e−03)	0.753	0.996
	BSHA	0.024 (1.066e−03)	0.927	0.999	0.024 (1.888e−03)	0.851	0.999
3	GCTA	0.184 (9.772e−04)	0.882	0.860	0.190 (7.866e−04)	0.601	0.860
	MEGHA	0.125 (5.791e−04)	0.739	0.950	0.127 (5.543e−04)	0.614	0.950
	MMHE	0.091 (4.304e−04)	0.832	0.897	0.092 (1.413e−03)	0.678	0.897
	BSHA1	0.043 (1.422e−03)	0.833	0.994	0.044 (1.494e−03)	0.672	0.993
	BSHA2	0.034 (1.058e−03)	0.881	0.999	0.034 (1.697e−03)	0.759	0.998
	BSHA3	0.050 (1.033e−03)	0.867	0.996	0.050 (2.594e−03)	0.774	0.995
	BSHA	0.030 (2.496e−04)	0.904	0.998	0.034 (1.427e−03)	0.798	0.997
4	GCTA	0.184 (5.633e−04)	0.881	0.860	0.189 (1.435e−03)	0.607	0.860
	MEGHA	0.124 (5.961e−04)	0.737	0.950	0.126 (6.291e−04)	0.611	0.950
	MMHE	0.092 (4.537e−03)	0.831	0.897	0.092 (6.832e−03)	0.678	0.897
	BSHA1	0.029 (3.104e−03)	0.885	0.997	0.029 (5.742e−03)	0.811	0.996
	BSHA2	0.025 (1.997e−03)	0.858	0.999	0.025 (5.564e−03)	0.719	0.999
	BSHA3	0.050 (3.684e−03)	0.888	0.997	0.050 (4.646e−03)	0.702	0.995
	BSHA	0.023 (2.062e−03)	0.912	0.999	0.023 (4.953e−03)	0.830	0.999
5	GCTA	0.184 (6.285e−04)	0.892	0.860	0.189 (1.217e−03)	0.591	0.860
	MEGHA	0.125 (5.354e−04)	0.751	0.950	0.127 (4.504e−04)	0.604	0.950
	MMHE	0.091 (3.078e−03)	0.843	0.897	0.092 (6.848e−03)	0.666	0.897
	BSHA1	0.030 (2.332e−03)	0.881	0.997	0.030 (4.913e−03)	0.781	0.996
	BSHA2	0.027 (1.631e−03)	0.897	0.999	0.028 (4.686e−03)	0.717	0.999
	BSHA3	0.050 (2.937e−03)	0.904	0.997	0.050 (4.643e−03)	0.702	0.995
	BSHA	0.024 (1.584e−03)	0.922	0.999	0.024 (5.325e−03)	0.801	0.999
6	GCTA	0.184 (7.298e−04)	0.883	0.860	0.190 (9.152e−04)	0.616	0.860
	MEGHA	0.126 (6.026e−04)	0.739	0.950	0.127 (5.564e−04)	0.618	0.950
	MMHE	0.092 (2.015e−04)	0.831	0.897	0.092 (5.021e−03)	0.682	0.897
	BSHA1	0.042 (1.412e−03)	0.764	0.994	0.043 (2.406e−03)	0.668	0.993
	BSHA2	0.046 (1.415e−03)	0.774	0.999	0.046 (3.099e−03)	0.689	0.998
	BSHA3	0.050 (2.263e−03)	0.878	0.996	0.050 (4.152e−03)	0.678	0.995
	BSHA	0.031 (2.334e−03)	0.907	0.998	0.036 (3.045e−03)	0.792	0.997

subject were collected through 1.5 T MRI scanners and then preprocessed by standard steps (Shen and Davatzikos, 2004). A detailed dataset description, image acquisition, and preprocessing steps can be found in Section S3.1 of the supplementary material available at *Biostatistics* online. To quantify the voxel-wise brain volumetric difference, we generate RAVENS maps using deformation field (Davatzikos *and others*, 2001) for whole brain tissue. Simultaneously, we label 90 ROIs based on Automated Anatomical Labeling (AAL)-90 atlas, and directly adopt the connectivity information via a previous study on Alzheimers patients (Kuceyeski *and others*, 2013). The total number of voxels covered by gray matter and sub-cortical regions from AAL is 144 999. We finally include gender, age, and the first five principle components of GRM as fixed effects. We apply the proposed BSHA and competing ones GCTA and MMHE with implementation settings following those in the simulations.

Our main goal is to prioritize the heritable brain phenotypes and quantify the genetic effect on those brain locations, and we summarize the heritability heatmaps for the significantly heritable brain locations under different methods in Figure 2. It is straightforward to draw the following conclusions from the figure: first, there is a general consistency between BSHA and the other methods in the locations with strongest genetic effects. As shown in the sagitall view, we can see a rough match in the heritably significant regions between BSHA and MMHE/GCTA, which ensures our method works properly. Second, the heritable phenotypes identified by BSHA are spatially much more clustered and contiguous, revealing its biologically appealing property; while the ones selected by MMHE/GCTA are more scattered, lacking meaningful interpretation. Third, MMHE/GCTA have identified a large amount of isolated signals, i.e., heritable phenotypes at a single or very small number of voxels with no other heritable ones nearby. Since our brain imaging measurements are at a very fine scale, such a big interruption at single voxel level makes little biological sense. By using BSHA, we extensively avoid the potential false positives.

We further summarize the total number of selected voxels and the average of corresponding heritabilities under each ROI in Table 2. Here, we list ROIs (a full list of region names is in Table 3 of the supplementary material available at *Biostatistics* online) with their phenotypes selected by at least one of the three methods, along with the number of heritable phenotypes (N_{voxel}) and their average heritability ($\bar{h}^2$). Based on Table 2, we confirm a general agreement among the three methods in identified brain areas. In terms of specific heritable brain regions, as shown in Table 2, BSHA helps detect a number of ROIs which have not been caught by the other two methods. For instance, the volume of middle frontal gyrus (MFG.R) is shown to be highly heritable among schizophrenia patients and siblings in previous studies (Hu *and others*, 2013), and there is also significant activation in MFG.R under working memory-related task which is strongly heritable (Blokland *and others*, 2008). Fusiform gyrus (FFG.L, FFG.R) is a well-known brain area involved in the processing of both faces and words, and its measurements are also shown to be highly heritable in multiple previous studies on different populations (Chouinard-Decorte *and others*, 2014).

5.2. *UK biobank*

The UK Biobank project (https://www.ukbiobank.ac.uk/) is a large prospective cohort study for middle- and older-aged adults from the United Kingdom. Here, we focus on individuals of middle or elderly ages (age range 40–69) with British ancestry. To prepare the data, we construct genetic GRM used the imputed SNP data with a quantity control step followed that in Zhao *and others* (2019). As for the imaging traits, we download the fractional anisotropy (FA) and mean diffusivity (MD) maps from the UK Biobank website and process the FA and MD according to the TBSS-Enigma pipeline (http://enigma.ini.usc.edu/protocols/dti-protocols/#dtiproj). We then skeletonize the registered FA and MD images by projecting the ENIGMA skeleton onto them and further resample the FA and MD skeletons to $2 \times 2 \times 2$ mm^3. We use 46 white matter tracts in brains two lateral hemispheres labeled by the ENIGMA-DTI pipeline (Jahanshad *and others*, 2013), which is widely applied to measure the variation of microstructural integrity. The ROIs location can be found in Figure 3. We do not incorporate

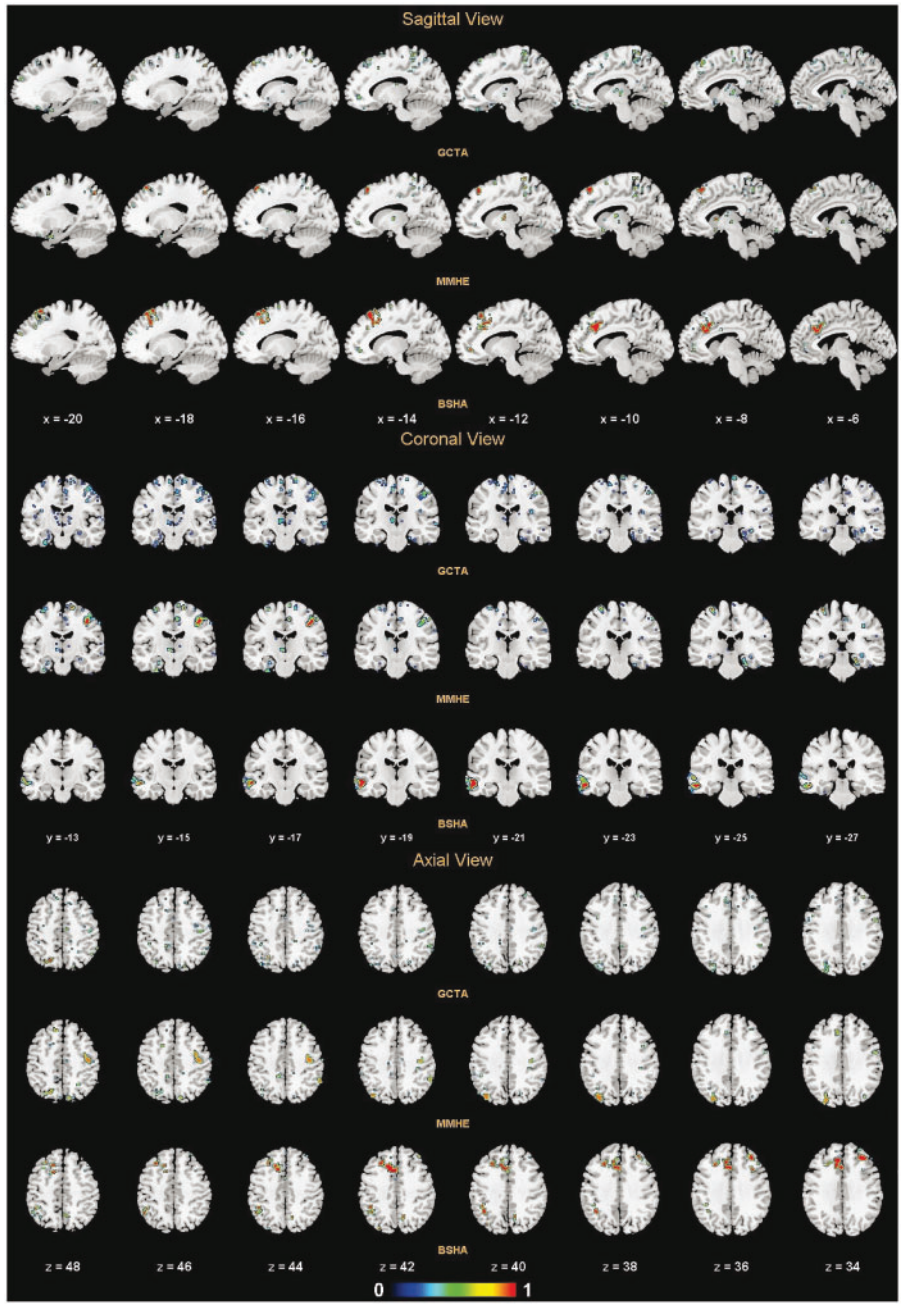

Fig. 2. ADNI data analysis results: estimated heritabilities under GCTA, MMHE, and BSHA over the whole brain. Only significant ones are represented by heatmaps from slides −20 to −6 in sagitall view, slides −13 to −27 in coronal view, and slides 48 to 34 in axial view with a 2-slide skipping.

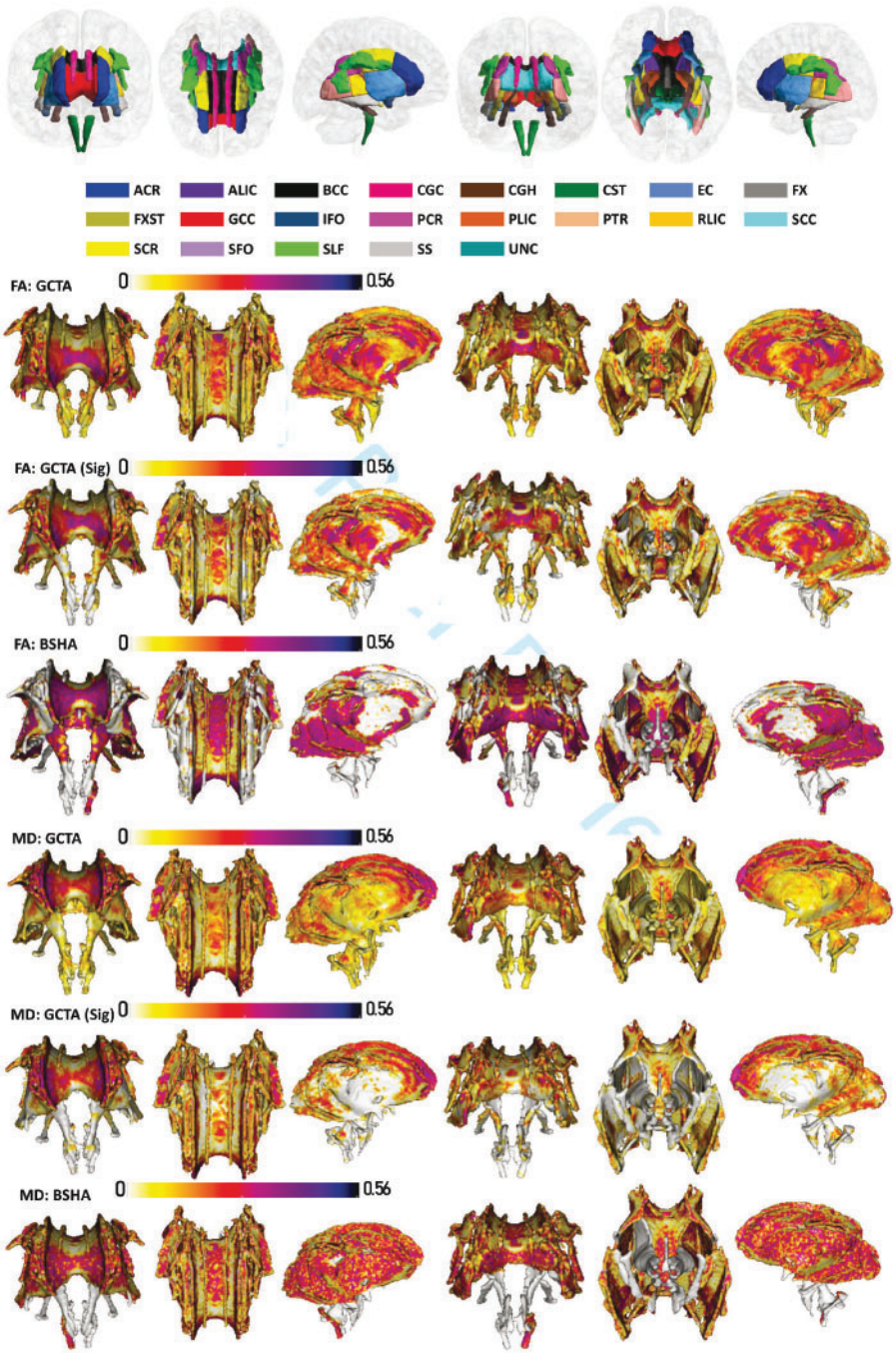

Fig. 3. UK Biobank data analysis results: the top row gives the orientation and location of white matter tracts in the brain; and Rows 2–4 and 5–7 display the estimated heritabilities under GCTA, GCTA (only significant voxels), and BSHA methods for FA and MD skeletons, respectively.

Table 2. *ADNI data analysis results: number of significantly heritable phenotypes (N_{voxel}) and their average heritability ($\bar{h}^2$) for all the ROIs whose phenotypes selected by at least one of the three methods.*

	GCTA		MMHE		BSHA			GCTA		MMHE		BSHA	
	N_{voxel}	$\bar{h}^2$	N_{voxel}	$\bar{h}^2$	N_{voxel}	$\bar{h}^2$		N_{voxel}	$\bar{h}^2$	N_{voxel}	$\bar{h}^2$	N_{voxel}	$\bar{h}^2$
MTG.R	16	0.963	17	0.742	86	0.861	ORBinf.R	1	0.994	2	0.834	0	—
SFGdor.L	20	0.961	26	0.836	57	0.838	ROL.L	1	0.986	2	0.709	0	—
MFG.R	7	0.963	3	0.848	42	0.836	ACG.R	2	0.938	2	0.808	0	—
FFG.R	13	0.961	4	0.718	35	0.849	PHG.L	3	0.966	2	0.868	0	—
SFGmed.L	9	0.893	31	0.820	34	0.853	CAL.R	2	0.943	2	0.836	0	—
FFG.L	6	0.966	1	0.740	32	0.789	LING.R	1	0.994	2	0.676	0	—
PoCG.R	30	0.933	28	0.803	31	0.820	CAU.L	0	—	2	0.886	0	—
MFG.L	14	0.944	3	0.808	30	0.844	ITG.R	12	0.962	2	0.912	0	—
ITG.L	7	0.901	5	0.859	25	0.834	ORBinf.L	9	0.942	1	0.930	0	—
PreCG.L	13	0.945	30	0.784	24	0.863	OLF.L	1	0.861	1	0.908	0	—
ANG.L	4	0.959	9	0.796	17	0.892	ORBsupmed.L	0	—	1	0.635	0	—
SFGmed.R	9	0.971	4	0.835	16	0.912	PCG.L	0	—	1	0.667	0	—
PCUN.R	10	0.930	11	0.747	16	0.784	HIP.R	7	0.973	1	0.816	0	—
MOG.L	36	0.962	24	0.798	13	0.832	SOG.R	2	0.964	1	0.985	0	—
IPL.L	8	0.934	8	0.815	12	0.887	SMG.L	1	0.956	1	0.719	0	—
PreCG.R	27	0.957	84	0.813	10	0.909	MTG.L	4	0.944	1	0.971	0	—
PCUN.L	21	0.954	12	0.848	10	0.822	PCL.L	11	0.951	0	—	0	—
STG.L	0	—	3	0.876	10	0.783	PCL.R	11	0.943	0	—	0	—
SFGdor.R	10	0.963	13	0.812	0	—	IFGoperc.L	8	0.960	0	—	0	—
SPG.L	20	0.883	20	0.782	0	—	IFGoperc.R	4	0.950	0	—	0	—
ANG.R	5	0.954	16	0.840	0	—	IFGtriang.L	4	0.959	0	—	0	—
ROL.R	6	0.949	13	0.773	0	—	SMA.L	4	0.980	0	—	0	—
PoCG.L	15	0.963	13	0.736	0	—	ACG.L	4	0.969	0	—	0	—
SPG.R	9	0.955	12	0.829	0	—	CUN.L	4	0.960	0	—	0	—
INS.L	2	0.974	11	0.776	0	—	ORBmid.L	3	0.929	0	—	0	—
REC.L	4	0.966	10	0.782	0	—	PCG.R	3	0.969	0	—	0	—
INS.R	4	0.958	9	0.812	0	—	LING.L	3	0.979	0	—	0	—
PHG.R	6	0.966	9	0.833	0	—	SMG.R	3	0.971	0	—	0	—
THA.L	4	0.978	9	0.822	0	—	PUT.R	3	0.962	0	—	0	—
MOG.R	15	0.957	8	0.718	0	—	IOG.R	2	0.970	0	—	0	—
IPL.R	8	0.965	8	0.816	0	—	IFGtriang.R	1	0.982	0	—	0	—
DCG.R	6	0.959	4	0.801	0	—	OLF.R	1	0.983	0	—	0	—
CAL.L	2	0.966	4	0.833	0	—	ORBsupmed.R	1	0.978	0	—	0	—
SMA.R	4	0.973	3	0.909	0	—	REC.R	1	0.984	0	—	0	—
THA.R	1	0.949	3	0.906	0	—	HIP.L	1	0.993	0	—	0	—
STG.R	7	0.938	3	0.679	0	—	CUN.R	1	0.990	0	—	0	—
TPOsup.R	0	—	3	0.704	0	—	SOG.L	1	0.973	0	—	0	—
ORBmid.R	1	0.888	2	0.865	0	—	TPOsup.L	1	0.970	0	—	0	—

regional connectivity in this case and consider the six adjacent neighbors for voxel-wise dependency. More detailed dataset description, image acquisition and preprocessing steps are included in Section S3.2 of the supplementary material available at *Biostatistics* online. We finally merge the SNP GRM, imaging phenotypes and nongenetic covariates (age, gender, and the first 40 principle components) together and obtain 16 432 subjects each with 34 646 voxel-level imaging phenotypes for both FA and MD.

We apply BSHA to conduct integrative heritability analysis for each brain phenotype map, and as a comparison, we also perform GCTA on individual measures for both FA and MD. The final results are provided in Figure 3, where we plot the heatmaps of the estimated heritability over whole brain skeleton under GCTA, those over significant voxels under GCTA, and those over BSHA. As we can see, similar to the ADNI data, the following conclusions still hold—there is a general result consistency between GCTA and BSHA, but heritable estimates identified by BSHA are spatially much more clustered, contiguous, and sparse. We further summarize the number and average estimated heritability over heritable voxels in each ROI in Tables 4 and 5 provided in the supplementary material available at *Biostatistics* online. Among the top identified ROIs by BSHA for both FA and MD skeletons, most of them are consistent with previous literature. For instance, the Corpus Callosum (BCC, GCC, and SCC) connects the left and right cerebral hemispheres, enabling communications between them and is the largest white matter structure in the human brain (Luders *and others*, 2010); and the high heritability of FA and MD at the Corpus Callosum, of MD value at corona radiata (PCR, ACR, and SCR), the low (zero) FA heritabilities of the FX and the SCT and the low MD heritabilities of CST, IFO, CGH, FX/ST are also consistent with the results in Zhao *and others* (2019) and Lee *and others* (2015).

6. Discussion

In this article, we fill the gap in existing literature by developing a unified Bayesian heritability analysis for high-dimensional phenotypes with biological structure. Under application to study SNPs effect on large scale brain imaging traits among unrelated subjects, we impose hierarchical selection among phenotypic heritabilities. In addition, our model incorporates the structural connectivity among ROIs and spatial dependence among voxels, leading to smoothness of heritable status by biological information. To further improve the robustness and interpretability of our result, we adopt a nonparametric Bayesian prior to model genetic effect, providing desired grouping effect with reduced computation. We show the superiority of our model construction in simulations and UK biobank application.

Computation is always a challenge for MCMC-based Bayesian method particularly in the presence of high-dimensional data. In our method, the novel model construction allows us to dramatically reduce the computation cost from several layers. The hierarchical sparsity enables to exclude a large number of inactive phenotypes at region level within MCMC and refines the selection at actual scale. By doing so, we considerably scale up the posterior computation and improve the mixing of MCMC. Additional, we use $\mathcal{DP}$ model to induce grouping effect among significant heritabilities, which is another source to reduce computation cost compared with parametric model that updates each heritability parameter one at a time. In practice, to mitigate the computational complexity coming from the inverse of $\mathcal{V}$ matrix which needs to be updated at each MCMC interaction, we could predetermine $\mathcal{V}^{-1}$ which can be transferred to a function of h under a wide range of domain values (say 0.2–0.9 with an accuracy 0.0001), and directly import $\mathcal{V}^{-1}$ when the posterior sample of τ_g^2 and σ^2 fit its support value. By doing so, we only need less than 4 h for ADNI data and less than 2 h for UK biobank applications (Matlab implementation, 2.4 GHz CPU, 64GB Memory, Windows System).

In our model formation, prior (3.3) is equivalent to the well-known point mass mixture prior on $\tau_g^2(s)$ to achieve selection. An interesting extension is to replace it with computationally more efficient continuous shrinkage prior. Currently, most of the works on this area focus on regression models, and more recently, structural based Bayesian shrinkage regression models are also proposed to incorporate prior network information (Chang *and others*, 2018). In our case, we could borrow the idea of structural-driven shrinkage regression but develop relevant prior model on variance component with assumption that inactive genetic effects are small but not exactly zero. Compared with our current model, this new direction may easily solve the computational issue and avoid the potential phase transition problem from Ising prior. Some other extensions of our current analysis include to consider different phenotypes like the

ones summarized from other imaging modalities (functional MRI or positron emission tomography). It is also possible though not readily applicable to conduct heritability analysis for brain functional/structural network.

7. SOFTWARE

Software in the form of MATLAB code is available at Github: https://github.com/yizekaren/Bayesian-sparse-heritability-analysis-with-high-dimensional-neuroimaging-phenotypes.

SUPPLEMENTARY MATERIAL

Supplementary material is available online at http://biostatistics.oxfordjournals.org.

Acknowledgments

This publication was made possible by CTSA Grant Number UL1 TR000142 from the National Center for Advancing Translational Science (NCATS), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NIH. *Conflict of Interest*: None declared.

REFERENCES

ALMASY, L. AND BLANGERO, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *The American Journal of Human Genetics* **62**, 1198–1211.

BLOKLAND, G. A. M., MCMAHON, K. L., HOFFMAN, J., ZHU, G., MEREDITH, M., MARTIN, N. G., THOMPSON, P. M., DE ZUBICARAY, G. I. AND WRIGHT, M. J. (2008). Quantifying the heritability of task-related brain activation and performance during the n-back working memory task: a twin fMRI study. *Biological Psychology* **79**, 70–79.

CAI, B. AND BANDYOPADHYAY, D. (2017). Bayesian semiparametric variable selection with applications to periodontal data. *Statistics in Medicine* **36**, 2251–2264.

CHANG, C., KUNDU, S. AND LONG, Q. (2018). Scalable Bayesian variable selection for structured high-dimensional data. *Biometrics* **74**, 1372–1382.

CHEN, R.-B., CHU, C.-H., YUAN, S. AND WU, Y. N. (2016). Bayesian sparse group selection. *Journal of Computational and Graphical Statistics* **25**, 665–683.

CHOUINARD-DECORTE, F., MCKAY, D. R., REID, A., KHUNDRAKPAM, B., ZHAO, L., KARAMA, S., RIOUX, P., SPROOTEN, E., KNOWLES, E., KENT, J. W. *and others*. (2014). Heritable changes in regional cortical thickness with age. *Brain Imaging and Behavior* **8**, 208–216.

DAVATZIKOS, C., GENC, A., XU, D. AND RESNICK, S. M. (2001). Voxel-based morphometry using the ravens maps: methods and validation using simulated longitudinal atrophy. *NeuroImage* **14**, 1361–1369.

FJELL, A. M., GRYDELAND, H., KROGSRUD, S. K., AMLIEN, I., ROHANI, D. A., FERSCHMANN, L., STORSVE, A. B., TAMNES, C. K., SALA-LLONCH, R., DUE-TØNNESSEN, P. *and others*. (2015). Development and aging of cortical thickness correspond to genetic organization patterns. *Proceedings of the National Academy of Sciences United States of America* **112**, 15462–15467.

GE, T., CHEN, C.-Y., NEALE, B. M., SABUNCU, M. R. AND SMOLLER, J. W. (2017). Phenome-wide heritability analysis of the UK biobank. *PLoS Genetics* **13**, e1006711.

GE, T., NICHOLS, T. E., LEE, P. H., HOLMES, A. J., ROFFMAN, J. L., BUCKNER, R. L., SABUNCU, M. R. AND SMOLLER, J. W. (2015). Massively expedited genome-wide heritability analysis (megha). *Proceedings of the National Academy of Sciences United States of America* **112**, 2479–2484.

GE, T., REUTER, M., WINKLER, A. M., HOLMES, A. J., LEE, P. H., TIRRELL, L. S., ROFFMAN, J. L., BUCKNER, R. L., SMOLLER, J. W. AND SABUNCU, M. R. (2016). Multidimensional heritability analysis of neuroanatomical shape. *Nature Communications* **7**, 13291.

GELMAN, A. and RUBIN, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–472.

HU, M., LI, J., EYLER, L., GUO, X., WEI, Q., TANG, J., LIU, F., HE, Z., LI, L., JIN, H. *and others*. (2013). Decreased left middle temporal gyrus volume in antipsychotic drug-naive, first-episode schizophrenia patients and their healthy unaffected siblings. *Schizophrenia Research* **144**(1-3), 37–42.

HUANG, L., GOLDSMITH, J., REISS, P. T., REICH, D. S. AND CRAINICEANU, C. M. (2013). Bayesian scalar-on-image regression with application to association between intracranial DTI and cognitive outcomes. *NeuroImage* **83**, 210–223.

IBRAHIM, J. G., ZHU, H., GARCIA, R. I. AND GUO, R. (2011). Fixed and random effects selection in mixed effects models. *Biometrics* **67**, 495–503.

JAHANSHAD, N., KOCHUNOV, P. V., SPROOTEN, E., MANDL, R. C., NICHOLS, T. E., ALMASY, L., BLANGERO, J., BROUWER, R. M., CURRAN, J. E., DE ZUBICARAY, G. I. *and others*. (2013). Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: a pilot project of the ENIGMADTI working group. *NeuroImage* **81**, 455–469.

KUCEYESKI, A., MARUTA, J., RELKIN, N. AND RAJ, A. (2013). The network modification (NeMo) tool: elucidating the effect of white matter integrity changes on cortical and subcortical structural connectivity. *Brain Connectivity* **3**, 451–463.

LEE, S. J., STEINER, R. J., LUO, S., NEALE, M. C., STYNER, M., ZHU, H. AND GILMORE, J. H. (2015). Quantitative tract-based white matter heritability in twin neonates. *NeuroImage* **111**, 123–135.

LI, F. AND ZHANG, N. R. (2010). Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *Journal of the American Statistical Association* **105**, 1202–1214.

LUDERS, E., THOMPSON, P. M. AND TOGA, A. W. (2010). The development of the corpus callosum in the healthy human brain. *Journal of Neuroscience* **30**, 10985–10990.

LUO, S., SONG, R., STYNER, M., GILMORE, J. H. AND ZHU, H. (2019). FSEM: functional structural equation models for twin functional data. *Journal of the American Statistical Association* **114**, 344–357.

POLDERMAN, T. J. C., BENYAMIN, B., DE LEEUW, C. A., SULLIVAN, P. F., VAN BOCHOVEN, A., VISSCHER, P. M. AND POSTHUMA, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* **47**, 702–709.

ROSHCHUPKIN, G. V., GUTMAN, B. A., VERNOOIJ, M. W., JAHANSHAD, N., MARTIN, N. G., HOFMAN, A., MCMAHON, K. L., VAN DER LEE, S. J., VAN DUIJN, C. M., DE ZUBICARAY, G. I. *and others*. (2016). Heritability of the shape of subcortical brain structures in the general population. *Nature Communications* **7**, 1–8.

SETHURAMAN, J. (1994). A constructive definition of dirichlet priors. *Statistica Sinica*, **4**, 639–650.

SHEN, D. AND DAVATZIKOS, C. (2004). Measuring temporal morphological changes robustly in brain MR images via 4-dimensional template warping. *NeuroImage* **21**, 1508–1517.

TENG, M., NATHOO, F. S. AND JOHNSON, T. D. (2019). Bayesian analysis of functional magnetic resonance imaging data with spatially varying auto-regressive orders. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **68**, 521–541.

XU, X., GHOSH, M. *and others*. (2015). Bayesian variable selection and estimation for group lasso. *Bayesian Analysis* **10**, 909–936.

YANG, J., BENYAMIN, B., MCEVOY, B. P., GORDON, S., HENDERS, A. K., NYHOLT, D. R., MADDEN, P. A., HEATH, A. C., MARTIN, N. G., MONTGOMERY, G. W. *and others*. (2010). Common SNPS explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565.

YANG, J., LEE, S. H., GODDARD, M. E. AND VISSCHER, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**, 76–82.

ZHANG, L., BALADANDAYUTHAPANI, V., MALLICK, B. K., MANYAM, G. C., THOMPSON, P. A., BONDY, M. L. AND DO, K.-A. (2014). Bayesian hierarchical structured variable selection methods with application to molecular inversion probe studies in breast cancer. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **63**, 595–620.

ZHANG, L., GUINDANI, M., VERSACE, F., ENGELMANN, J.M., AND VANNUCCI, M. (2016). A spatiotemporal nonparametric Bayesian model of multi-subject fMRI data. *The Annals of Applied Statistics* **10**, 638–666.

ZHAO, B., ZHANG, J., IBRAHIM, J. G., LUO, T., SANTELLI, R. C., LI, Y., LI, T., SHAN, Y., ZHU, Z., ZHOU, F. *and others*. (2019). Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). *Molecular Psychiatry*, https://doi.org/10.1038/s41380-019-0569-z.

ZHAO, Y., CHUNG, M., JOHNSON, B. A., MORENO, C. S. AND LONG, Q. (2016). Hierarchical feature selection incorporating known and novel biological information: identifying genomic features related to prostate cancer recurrence. *Journal of the American Statistical Association* **111**, 1427–1439.

[*Received March 23, 2020; revised July 15, 2020; accepted for publication August 11, 2020*]