Biostatistics, 2024, **25**, 1195–1209
https://doi.org/10.1093/biostatistics/kxae008
Advance Access publication March 17, 2024

Bayesian mixed model inference for genetic association under related samples with brain network phenotype

Xinyuan Tian[1,†], Yiting Wang[1,†], Selena Wang[1], Yi Zhao[2], Yize Zhao[1,*]

[1]Department of Biostatistics, Yale University, 60 College St, New Haven, CT 06520, United States
[2]Department of Biostatistics and Health Data Science, Indiana University, 410 W. 10th St, Indianapolis, IN 46202, United States

[*]Corresponding author: Department of Biostatistics, Yale University, 60 College St, New Haven, CT 06520, United States
Email: yize.zhao@yale.edu
[†]Equally contributed.

SUMMARY

Genetic association studies for brain connectivity phenotypes have gained prominence due to advances in noninvasive imaging techniques and quantitative genetics. Brain connectivity traits, characterized by network configurations and unique biological structures, present distinct challenges compared to other quantitative phenotypes. Furthermore, the presence of sample relatedness in the most imaging genetics studies limits the feasibility of adopting existing network-response modeling. In this article, we fill this gap by proposing a Bayesian network-response mixed-effect model that considers a network-variate phenotype and incorporates population structures including pedigrees and unknown sample relatedness. To accommodate the inherent topological architecture associated with the genetic contributions to the phenotype, we model the effect components via a set of effect network configurations and impose an inter-network sparsity and intra-network shrinkage to dissect the phenotypic network configurations affected by the risk genetic variant. A Markov chain Monte Carlo (MCMC) algorithm is further developed to facilitate uncertainty quantification. We evaluate the performance of our model through extensive simulations. By further applying the method to study, the genetic bases for brain structural connectivity using data from the Human Connectome Project with excessive family structures, we obtain plausible and interpretable results. Beyond brain connectivity genetic studies, our proposed model also provides a general linear mixed-effect regression framework for network-variate outcomes.

KEYWORDS: brain connectivity; genome-wide association studies; imaging genetics; mixed effects; network-response model; sample relatedness.

1. INTRODUCTION

Brain imaging genetics, aiming to uncover the genetic basis of brain structure and function, has provided an unprecedented opportunity to understand the molecular support for different neurobiological processes. By leveraging imaging quantitative traits as endophenotypes that reflect underlying neurological etiologies, we gain a deeper understanding of the risk biomarkers implicated in both disease outcomes and normal trajectory of development and aging.

Received: May 15, 2023. **Revised:** January 22, 2024. **Accepted:** February 19, 2024

Brain connectivity, encoding the relations between distinct units or nodes within a nervous system, has played an essential role in disclosing the brain neuronal interactions and reflecting correspondence with behavior. Depending on the aspect of characterization, brain connectivity can be summarized by anatomical links capturing the white matter fiber tracts known as structural connectivity, or statistical dependence between functional time courses known as functional connectivity. Converging evidence indicates brain connectivity is heritable, and can offer distinct genetic underpinnings compared with other neuroimaging traits (Zhao et al., 2021; Elliott et al., 2018). This underscores the significance of studying the genetic contributions to connectivity patterns. From an analytical perspective, structural and functional connectivity can be viewed as an indirect graph with all the nodes over the brain as the vertex set and the corresponding connections as the edge set. By extracting single edges as univariate phenotypes, most of the current genome-wide association studies (GWAS) were performed separately on each brain connection (Zhao et al., 2021; Jahanshad et al., 2013; Elsheikh et al., 2020). However, such analyses overlook the biological interdependence and graphical structure inherent in brain network topography, which can raise concerns regarding biological plausibility and interpretability, as our data application demonstrates.

On the other hand, as the study of brain connectivity gains increasing interest, network-variate modeling has emerged as an advanced analytical framework capable of accommodating the underlying dependence and brain topological architectures. In contrast to marginal and univariate analyses, network-variate modeling directly handles the (weighted) adjacency matrix of connectivity, enabling an explicit characterization of the biological structure. Depending on the objectives of the study, the network-variate can serve three distinct roles. Firstly, it can be employed solely to describe the neurobiological profiles of the brain using different types of graphical modeling techniques in light of topological assumptions (Wang and Guo, 2020; Zhang et al., 2020). Secondly, when associated with a behavioral outcome, the network-variate can be treated as a predictor, involving specific matrix/tensor operations such as outer products (Wang et al., 2021) to transform the predictive component into a linear term (Zhao et al., 2022). Finally, to investigate the impact of covariates or exposures on the variation of connectivity, the network-variate can be treated as an outcome in a network-response regression. In this case, the coefficient parameters reveal a matrix or tensor format and can be further decomposed to elucidate the latent effect mechanisms (Zhang et al., 2023; Zhao et al., 2023; Kong et al., 2019). It is evident that the last category could shed light on genetic association analyses involving connectivity or network-variate phenotypes.

From a study design perspective, sample relatedness is highly prevalent and almost unavoidable in quantitative genetics studies. Such relatedness could be induced by recruitment from the same family or pedigree, or unknown or uncertain relationships including distant levels of unknown common ancestry (Eu-Ahsunthornwattana et al., 2014). Failure to account for potential sample structures within GWAS can lead to spurious results (Helgason et al., 2005), emphasizing the necessity for appropriate correction methods. One common approach to account for sample structures is to include a random effect component to incorporate known or unknown relatedness. Building on linear mixed-effect models (LMMs), various numerical implementation approaches proposed in recent years to characterize genetic associations accommodating population substructure and potential sample relatedness (Kang et al., 2010; Zhou and Stephens, 2012). However, most of these approaches are designed for univariate phenotypes or vector-variate multivariate phenotypes, and there is currently no existing framework that adequately considers or readily applies to network-variate phenotypes.

To address the above limitations, we propose a **B**ayesian **N**etwork-phenotype **M**ixed **E**ffect model (BNME) to perform genetic association analyses with brain connectivity phenotype. Within this unified modeling framework, we simultaneously characterize genetic contributions and identify affected phenotypic network components, while quantifying their uncertainty. To leverage the biological knowledge that brain connectivity operates via network configurations, our approach assumes that risk genetic variants influence network alternations by acting upon specific network

configurations that are to be uncovered. By imposing shrinkage and sparsity priors on the effect parameters, we can map out the genetically targeted brain network configurations that play a critical role in guiding future intervention strategies. In contrast to the existing works on network-response genetic association analyses, our proposed method incorporates pedigree information or unknown sample structures, ensuring the reliability and validity of the findings. In our data application, we apply the BNME model to study the genetic bases of brain structural connectivity using data from the Human Connectome Project (HCP), accommodating the extensive family structures among the subjects. Lastly, despite the proposed model being motivated by brain connectivity genetic studies, it can be readily extended to perform general network- or matrix-response mixed effects modeling. To the best of our knowledge, this work is among the very first to develop such a modeling framework, which directly fulfills an urgent need to capture multi-source of random variability for a growing collection of network data in epidemiology and social studies.

The remainder of the article is organized as follows. In Section 2, we describe the proposed LMM with a network response (Section 2.1), the prior specifications (Section 2.2), the posterior inference procedure (Section 2.3), and the covariates effect adjustment (Section 2.4). We conduct simulation studies in Section 3, followed by an application to HCP brain connectivity genetics data in Section 4. In the end, we conclude the article with a discussion in Section 5.

2. MATERIALS AND METHODS

2.1. Linear mixed-effect model with a network phenotype

We first describe the problem setting in the context of GWAS with genetic correlation, though the model formulation represents a general network-response mixed-effect model that can be extended to other applications. Assume the study includes N subjects with known pedigree structure or unknown relationship. For subject i ($i = 1, \ldots, N$), let z_i denote the genotype of interest which is encoded as 0, 1 or 2 according to the number of copies for the tested allele, $\mathbf{x}_i \in \mathbb{R}^{P\times 1}$ represents a set of covariates, and $\mathbf{A}_i \in \mathbb{R}^{V\times V}$ denotes the network phenotype summarized by a graphical matrix. With $\mathbf{A}_i$ stacked across all the subjects, we have the network phenotype array $\mathcal{A} \in \mathbb{R}^{N\times V\times V}$. Specifically in the application of brain connectivity studies, with images processed under a common brain atlas with V nodes, both structural and functional connectivity can be viewed as an indirect graph across vertex set $\{1, \ldots, V\}$. Thus, $\mathbf{A}_i$ becomes a symmetric matrix to summarize brain connectivity for each subject with diagonal elements to be zero, and its (v, v')th entry $a_{ivv'}$, $0 < v \neq v' \leq V$ represents the connection between nodes v and v' characterizing either the white matter fiber tracts (structural connectivity) or statistical dependence of functional time course (functional connectivity). We adopt continuous metrics to measure structural and functional connections. After normalizing the genetic variant and each phenotypic connection, we propose the following genetic association model for the indirect network response:

$$\mathbf{A}_i = \boldsymbol{\Theta} z_i - \mathrm{Hol}[\boldsymbol{\Theta} z_i] + \mathbf{B}_i + \mathbf{E}_i. \tag{2.1}$$

Here, $\boldsymbol{\Theta} \in \mathbb{R}^{V\times V}$ is the symmetric coefficient matrix to capture the genetic effect on the network phenotype, $\mathrm{Hol}[\cdot]$ is the operation to hollow out the diagonal elements to form a diagonal matrix, $\mathbf{B}_i \in \mathbb{R}^{V\times V}$ is the hollow symmetric random polygenic effect matrix, and $\mathbf{E}_i \in \mathbb{R}^{V\times V}$ is the hollow symmetric random error matrix characterizing the environmental effects. To demonstrate the main idea, we include only the genetic fixed effect at this moment, and we will extend the model to include covariates afterwards. Model (2.1) can be viewed as an extension of the traditional linear mixed-effect model for genetic association accommodating sample relatedness. In addition to a matrix-variate phenotype, we design both mean and variance components to maintain their original functions while satisfying the symmetric and hollow structure of the indirect network as shown in the right-hand side of model (2.1). Specifically, for the genetic and environmental effect matrix, by stacking each of them across all the subjects, we have the random effect tensor $\mathcal{B} \in \mathbb{R}^{N\times V\times V}$ and

residual error tensor $\mathcal{E} \in \mathbb{R}^{N\times V\times V}$ with

$$\text{vec}(\mathcal{B}) \sim \text{N}\left\{\begin{pmatrix}0\\ \vdots\\ 0\\ \vdots\\ 0\end{pmatrix}, \text{Diag}\begin{pmatrix}\sigma^{(a)}_{11}\\ \vdots\\ \sigma^{(a)}_{vv'}\\ \vdots\\ \sigma^{(a)}_{VV}\end{pmatrix} \otimes 2\Lambda\right\}, \quad \text{vec}(\mathcal{E}) \sim \text{N}\left\{\begin{pmatrix}0\\ \vdots\\ 0\\ \vdots\\ 0\end{pmatrix}, \text{Diag}\begin{pmatrix}\sigma^{(e)}_{11}\\ \vdots\\ \sigma^{(e)}_{vv'}\\ \vdots\\ \sigma^{(e)}_{VV}\end{pmatrix} \otimes \text{I}_N\right\},$$

where Diag$(\cdot)$ constructs the diagonal matrix formed by the inside vector, $\sigma^{(a)}_{vv'}$ and $\sigma^{(a)}_{vv'}$ represent the additive genetic variance and random environmental variance, $\text{I}_N \in \mathbb{R}^{N\times N}$ is the identify matrix, and $\Lambda \in \mathbb{R}^{N\times N}$ is the kinship matrix estimated by pedigree information for known family structures or genotypic relationship for unknown relatedness (Eu-Ahsunthornwattana et al., 2014). To maintain the symmetric hollow structure of $\mathbf{B}_i$ and $\mathbf{E}_i$, we further specified that $\mathcal{B}_{ivv'} = \mathcal{B}_{iv'v}$ and $\mathcal{E}_{ivv'} = \mathcal{E}_{iv'v}$ when $v \neq v'$, and set $\mathcal{B}_{ivv'}$ and $\mathcal{E}_{ivv'}$ to 0 for $v = v'$. By proposing so, we can show the phenotypic variance of each connection $\text{Var}(\mathbf{a}_{vv'}) = 2\sigma^{(a)}_{vv'}\Lambda + \sigma^{(e)}_{vv'}\mathbf{I}_N$, $\mathbf{a}_{vv'} = (a_{1vv'}, \ldots, a_{Nvv'})^T$, $v \neq v'$, consistent with the existing literature (Kang et al., 2010).

Given the size of the commonly used brain atlas can be large with V in the range of 200–1000, directly performing estimation on model (2.1) is not ideal under a high-dimensional parameter space. More importantly, considering the primary interest in investigating the genetic association with brain network architectures, the topological structure cannot be plausibly reflected by ignoring the dependence within the genetic coefficient matrix. To address so, we adopt the following two-dimensional Tucker decomposing under a symmetry constrain for the coefficient matrix

$$\mathbf{\Theta} = \sum_{h=1}^{H} \eta_h \boldsymbol{\theta}_h \otimes \boldsymbol{\theta}_h, \tag{2.2}$$

where $\otimes$ represents the outer product, and $\boldsymbol{\theta}_h = (\theta_{h1}, \ldots, \theta_{hV})^T, h = 1, \ldots, H$ are column coefficient vectors. Under this representation, each outer product $\boldsymbol{\theta}_h \otimes \boldsymbol{\theta}_h$ forms a clique graph with nodes corresponding to the nonzero elements of $\boldsymbol{\theta}_h$ fully connected. We show that the decomposition structure in equation (2.2) is uniquely determined with detailed proof provided in supplementary material S.2. Additionally, from a neurobiological perspective, each $\boldsymbol{\theta}_h \otimes \boldsymbol{\theta}_h$ describes an effect network component adjusted by a weight parameter η_h. Combining models (2.1) and (2.2), we allow the genetic variant to deliver its impact on the phenotype via a series of signaling network architectures.

2.2. Prior specifications

We consider a fully Bayesian paradigm to estimate and perform inference for the proposed network-response LMM. For the fixed genetic effect component, we anticipate the genetic impact is sparse across the brain as shown by the existing empirical studies (Zhao et al., 2021). Therefore, we assign the following combination of point mass mixture prior and shrinkage prior

$$\eta_h \sim (1-\tau_h)\delta_0 + \tau_h \text{N}(0, \omega); \quad \theta_{hv} \sim \mathcal{L}(v), \quad h = 1, \ldots, H; v = 1, \ldots, V. \tag{2.3}$$

Here, τ_h is the latent selection indicator to determine whether a network configuration is significantly affected by the genotype as a whole. When $\tau_h = 1$, the weight parameter η_h is generated from a noninformative Normal prior with a large variance parameter ω; otherwise, we assign η_h to a point mass at zero denoted by δ_0 to remove the whole component from the model. In real practice, with the number of effect component H unknown, such a specification of sparsity could efficiently assist the determination of the number of associated phenotypic network configurations

during the learning process. As shown in our numerical studies, by imposing a conservative value to H, our model can still correctly uncover the signaling network phenotypes. To specify priors for latent indicators τ_h, one can either impose a noninformative Bernoulli distribution for each of the elements, or resort to a more informative prior by incorporating additional biological structure (Li and Zhang, 2010). For the coefficients, we assign a Laplace prior $\mathcal{L}(\nu)$ with a scale parameter ν to shrink the noise effect to a close to zero value. To further facilitate a straightforward posterior computation, following Park and Casella (2008), we represent each Laplace prior by a scale mixture of Normals for each $h = 1, \ldots, H$

$$\boldsymbol{\theta}_h \sim \mathrm{N}(\mathbf{0}, \mathcal{D}_h); \quad \mathcal{D}_h = \mathrm{Diag}(\sigma_{h1}, \ldots, \sigma_{hV}); \quad \sigma_{h\nu} \sim \frac{\nu^2}{2} \exp(-\frac{\nu^2 \sigma_{h\nu}}{2}) d\sigma_{h\nu}, \nu = 1, \ldots, V. \tag{2.4}$$

Combining priors (2.3) and (2.4), we characterize the phenotypic signals in a hierarchical way with an inter-group sparsity to induce the selection of a phenotypic network as a whole and an intra-group shrinkage to identify the signaling phenotypic network configuration within each selected component. In contrast to the existing sparse group selection or shrinkage models that primarily focus on group structural covariates, the current work emphasizes the network-variate outcome, which captures the associations between covariates and latent topological hierarchies. Additionally, we opt for shrinkage priors for individual coefficients instead of point mass mixture priors, driven by computational considerations that result in lower computational costs for shrinkage priors. However, it is important to note that the Laplace prior can be readily replaced with spike-and-slab types of priors or other graphical priors (Chang et al., 2018; Stingo et al., 2011) to impose sharp sparsity or incorporate spatial information. For the genetic and environmental variances $\sigma^{(a)}_{\nu\nu'}$ and $\sigma^{(e)}_{\nu\nu'}$, we consider two types of prior distribution. For the first one, we assign each $\sigma^{(a)}_{\nu\nu'}$ and $\sigma^{(e)}_{\nu\nu'}$ an Inverse Gamma distribution $\mathrm{IG}(\alpha_0, \beta_0)$. This specification, while not directly accounting for the correlations among $\sigma^{(a)}_{\nu\nu'}$ (or $\sigma^{(e)}_{\nu\nu'}$), $0 < \nu \neq \nu' \leq V$, across brain anatomy, offers a significant reduction of computational demands in posterior inference. Alternatively, in line with Zhao et al. (2022), we assume each $\sigma^{(a)}_{\nu\nu'}$ and $\sigma^{(e)}_{\nu\nu'}$ follows a predefined probability function $G^{(a)}$ and $G^{(e)}$, respectively; and we assign a nonparametric Dirichlet process (DP) prior for $G^{(a)}$ and $G^{(e)}$. Such a modeling strategy has been adopted in the previous brain imaging studies to impose spatial smoothness (Li et al., 2015; Zhao et al., 2022). The discrete nature of DP facilitates a clustering effect on contiguous brain locations, thereby allowing them to share the same parameter value. In our numerical studies shown in Section 3, we comprehensively compare the model performance under these two variance prior implementations and conclude the consistency of their results. Therefore, we primarily focus on the computationally more efficient IG priors in the following sections and name our model **B**ayesian **N**etwork-phenotype **M**ixed **E**ffect model (BNME). We refer to the DP version as $\mathrm{BNME}_{\mathrm{DP}}$ and detail its implementations in supplementary materials. Finally, for the tuning parameters including the number of informative network configurations H and scale parameter ν, we consider a grid search of them and choose the optimal values using the Bayesian information criterion (BIC). Our numerical experience suggests that this strategy is effective in practical applications.

2.3. Posterior likelihood and inference for BNME

To perform posterior inference for the proposed BNME model, we first develop the posterior likelihood for the collection of unknown parameters denoted as $\boldsymbol{\zeta} = \Big[\big\{ \boldsymbol{\theta}_h, \eta_h, \tau_h, \big(\sigma_{h\nu}\big)_{\nu=1}^{V} \big\}_{h=1}^{H}, \big(\sigma^{(a)}_{\nu\nu'}\big)_{\nu,\nu'=1}^{V}, \big(\sigma^{(e)}_{\nu\nu'}\big)_{\nu,\nu'=1}^{V}, \nu \Big]$. Based on the observed data $\mathcal{O} = (\mathbf{A}_i, z_i, \Lambda; i = 1, \ldots, N)$, the joint posterior distribution follows:

$$\pi(\boldsymbol{\zeta} \mid \mathcal{O}) \propto \prod_i \pi\left(\mathbf{A}_i, z_i, \Lambda \mid \{\boldsymbol{\theta}_h, \eta_h, \tau_h\}_{h=1}^H, \left(\sigma_{vv'}^{(a)}\right)_{v,v'=1}^V, \left(\sigma_{vv'}^{(e)}\right)_{v,v'=1}^V\right)$$
$$\times \prod_v \left\{ \prod_h \pi(\boldsymbol{\theta}_{hv} \mid \sigma_{hv}, v)\pi(\sigma_{hv}) \prod_{v'} \pi(\sigma_{vv'}^{(a)})\pi(\sigma_{vv'}^{(e)}) \right\} \prod_h \left\{ \pi(\tau_h)\pi(\eta_h|\tau_h) \right\}, \tag{2.5}$$

which combines the conditional observed data likelihood with prior distributions. Given uncertainty quantification is an essential component for genetic association analyses, instead of pursuing point estimates via optimization algorithms, we develop a Markov chain Monte Carlo sampling algorithm for posterior inference based on a combination of Gibbs samplers and Metropolis–Hastings (MH) updates. Under random initialization, we cycle through the following steps:

- For $h = 1 \ldots H$, $v = 1 \ldots V$, $v' = 1 \ldots V$, denote the (v, v')th entry of matrix $\{\mathbf{A}_i - \sum_{h' \neq h} \eta_{h'} \boldsymbol{\theta}_{h'} \otimes \boldsymbol{\theta}_{h'} z_i\}$ as $\widetilde{a}_{ivv'h}$, and define $\widetilde{\boldsymbol{a}}_{vv'h} = (\widetilde{a}_{1vv'h}, \ldots, \widetilde{a}_{Nvv'h})^T$, $\boldsymbol{z} = (z_1, \ldots, z_N)^T$. Sample θ_{hv} from $\mathrm{N}(\mu_{\theta_{hv}}, \sigma_{\theta_{hv}})$ with $\sigma_{\theta_{hv}} = \left(\sum_{v' \neq v} \eta_h^2 \theta_{hv'}^2 \boldsymbol{z}^T \left(2\sigma_{vv'}^{(a)} \Lambda + \sigma_{vv'}^{(e)} \mathbf{I}\right)^{-1} \boldsymbol{z} + \sigma_{hv}^{-1} \right)^{-1}$ and $\mu_{\theta_{hv}} = \sum_{v' \neq v} \widetilde{\boldsymbol{a}}_{vv'h}^T \left(2\sigma_{vv'}^{(a)} \Lambda + \sigma_{vv'}^{(e)} \mathbf{I}\right)^{-1} \boldsymbol{z} \eta_h \theta_{hv'} \sigma_{\theta_{hv}}$
- For $h = 1 \ldots H, v = 1 \ldots V$, sample σ_{hv}^{-1} from an Inverse Normal distribution $\mathrm{IN}(\frac{v}{|\theta_{hv}|}, v^2)$.
- For $h = 1 \ldots H$, when $\tau_h = 0$, set η_h to be zero. Otherwise, denote the (v, v')th entry of matrix $\boldsymbol{\theta}_h \otimes \boldsymbol{\theta}_h z_i$ as $q_{ivv'h}$, and $\boldsymbol{q_{vv'h}} = (q_{1vv'h}, \ldots, q_{Nvv'h})^T$. Update between network configuration coefficient η_h from their corresponding posterior Normal distribution $\mathrm{N}(\mu_{\eta_h}, \sigma_{\eta_h})$ with $\sigma_{\eta_h} = \left(\sum_{v<v'} \boldsymbol{q_{vv'h}}^T (2\sigma_{vv'}^{(a)} \Lambda + \sigma_{vv'}^{(e)} \mathbf{I})^{-1} \boldsymbol{q_{vv'h}} + \omega^{-1} \right)^{-1}$, and $\mu_{\eta_h} = \sum_{v<v'} \widetilde{\boldsymbol{a}}_{vv'h}^T (2\sigma_{vv'}^{(a)} \Lambda + \sigma_{vv'}^{(e)} \mathbf{I})^{-1} \boldsymbol{q_{vv'h}} \sigma_{\eta_h}$.
- For $h = 1 \ldots H$, define $l_0 := \frac{C}{\omega} \exp(-\frac{1}{2}(\frac{C\eta_h}{\omega})^2)$ and $l_1 := \frac{1}{\omega} \exp(-\frac{1}{2}(\frac{\eta_h}{\omega})^2)$ with C a large constant. We then update the selection indicators τ_h following the posterior Bernoulli distributions $\mathrm{Bern}(\frac{l_1}{l_0 + l_1})$.
- For $v = 1 \ldots V$, $v' = 1 \ldots V$, update $\sigma_{vv'}^{(a)}$ by sampling a proposed value $\sigma_{vv'}^{(a)p}$ from a random walk proposal distribution $\mathrm{N}(\sigma_{vv'}^{(a)}, \rho_1^2)$, and setting $\sigma_{vv'}^{(a)} = \sigma_{vv'}^{(a)p}$ with probability $\min\{1, R_1\} I\{\sigma_{vv'}^{(a)p} > 0\}$, where $R_1 = \frac{\pi(\sigma_{vv'}^{(a)p} | \mathcal{O}, \{\eta_h, \tau_h, \boldsymbol{\theta}_h\}_{h=1}^H, \sigma_{-v,-v'}^{(a)}, \sigma^{(e)})}{\pi(\sigma_{vv'}^{(a)} | \mathcal{O}, \{\eta_h, \tau_h, \boldsymbol{\theta}_h\}_{h=1}^H, \sigma_{-v,-v'}^{(a)}, \sigma^{(e)})}$, with $\pi(\sigma_{vv'}^{(a)} \mid \mathcal{O}, \{\eta_h, \tau_h, \boldsymbol{\theta}_h\}_{h=1}^H, \sigma_{-v,-v'}^{(a)}, \sigma^{(e)}) \propto \mathcal{L}(\mathcal{O} \mid \boldsymbol{\zeta}) \{\sigma_{vv'}^{(a)}\}^{-\alpha_0 - 1} \exp(-\frac{\beta_0}{\sigma_{vv'}^{(a)}})$ the full conditional.
- For $v = 1 \ldots V$, $v' = 1 \ldots V$, update $\sigma_{vv'}^{(e)}$ by sampling a proposed value $\sigma_{vv'}^{(e)p}$ from a random walk proposal distribution $\mathrm{N}(\sigma_{vv'}^{(e)}, \rho_2^2)$ and setting $\sigma_{vv'}^{(e)} = \sigma_{vv'}^{(e)p}$ with probability $\min\{1, R_2\} I\{\sigma_{vv'}^{(e)p} > 0\}$, where $R_2 = \frac{\pi(\sigma_{vv'}^{(e)p} | \mathcal{O}, \{\eta_h, \tau_h, \boldsymbol{\theta}_h\}_{h=1}^H, \sigma^{(a)}, \sigma_{-v,-v'}^{(e)})}{\pi(\sigma_{vv'}^{(e)} | \mathcal{O}, \{\eta_h, \tau_h, \boldsymbol{\theta}_h\}_{h=1}^H, \sigma^{(a)}, \sigma_{-v,-v'}^{(e)})}$, with $\pi(\sigma_{vv'}^{(e)} \mid \mathcal{O}, \{\eta_h, \tau_h, \boldsymbol{\theta}_h\}_{h=1}^H, \sigma^{(a)}, \sigma_{-v,-v'}^{(e)}) \propto \mathcal{L}(\mathcal{O} \mid \boldsymbol{\zeta}) \{\sigma_{vv'}^{(e)}\}^{-\alpha_0 - 1} \exp(-\frac{\beta_0}{\sigma_{vv'}^{(e)}})$ the full conditional.

Based on the posterior samples, the convergence of the algorithm is examined by trace plots and GR method (Gelman and Rubin, 1992). To characterize the genetic impact and dissect the associated signaling brain network configurations, we first determined the overall phenotypic network configurations linked with the genetic variant based on a 0.5 cutoff of the posterior mean for each τ_h. This cutoff is adopted in light of the median probability model (Hastie et al., 2004). Under a conservative H, most of the risk genetic variants are associated with less than H brain connectivity network configurations. When none of the elements in $\{\tau_h\}_{h=1}^H$ surpasses the cutoff, the genetic

variant is considered a noise variant, indicating that it does not have a significant impact on any component of the network phenotype. For the selected network configurations with τ_h larger than the cutoff, the genetic effect over network structures is captured by the posterior mean of θ_h. Despite that a Laplace prior does not impose strict sparsity, we can determine the specific brain network configurations that are most relevant to the genetic impact by extracting the elements from θ_h with 95% posterior credible interval excluding zero. Eventually, our model could provide estimation and inference for the risk genetic factors and their most influencing phenotypic topological elements.

2.4. Covariates adjustment

In genetic association studies, one may need to adjust for additional covariates, such as demographics and genetic principle components. Denote the covariate matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)^T \in \mathbb{R}^{N\times P}$. By including the covariates, model (2.1) takes the following compact representation

$$\mathcal{A} = \mathcal{S} \times_1 \boldsymbol{X} + (\boldsymbol{\Theta} - \text{Hol}[\boldsymbol{\Theta}]) \times_1 \mathbf{z} + \mathcal{B} + \mathcal{E}, \tag{2.6}$$

with $\times_1$ representing the 1-mode product, and $\mathcal{S} \in \mathbb{R}^{P\times V\times V}$ the coefficient tensor for the covariates which is symmetric at the horizontal slice $\mathcal{S}_{p,:,:}$. In practice, $\mathcal{S}$ can be considered as nuisance parameters. The canonical way is to assign simple conjugate priors for $\mathcal{S}$, which in our case are element-wise Gaussian priors, and then perform inference within MCMC. We denote our model under such an implementation as BNME_{adj} and provide detailed prior settings and posterior algorithm in the supplementary materials.

Alternatively, we can remove the nuisance parameters from (2.6) through a projection approach to reduce the parameter space and computational cost, which is in line with existing works on multivariate outcomes (Ge et al., 2016; Zhao et al., 2022). Specifically, we define a projection matrix $\boldsymbol{W} = \mathbf{I}_N - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$. Clearly, $\boldsymbol{W}$ is symmetric and idempotent matrix with a rank $N - P$, and this further indicates that $\boldsymbol{W}$ can be decomposed as $\boldsymbol{W} = \boldsymbol{U}^T\boldsymbol{U}$, where matrix $\boldsymbol{U} \in \mathbb{R}^{(N-P)\times N}$ and satisfies $\boldsymbol{U}\boldsymbol{U}^T = \mathbf{I}_{N-P}$ and $\boldsymbol{U}\boldsymbol{X} = \mathbf{0}$. Through matrix $\boldsymbol{U}$, the data can be projected from the N dimensional space onto an $N - P$ dimensional subspace. This facilitates an efficient way to remove the nuisance covariate effects by multiplying $\boldsymbol{U}$ to both sides in (2.6) that becomes

$$\mathcal{A} \times_1 \boldsymbol{U} = (\boldsymbol{\Theta} - \text{Hol}[\boldsymbol{\Theta}]) \times_1 (\boldsymbol{U}\mathbf{z}) + \mathcal{B} \times_1 \boldsymbol{U} + \mathcal{E} \times_1 \boldsymbol{U}. \tag{2.7}$$

Model (2.7) indicates that by replacing the connectivity array $\mathcal{A}$ with $\widetilde{\mathcal{A}} = \mathcal{A} \times_1 \boldsymbol{U} \in \mathbb{R}^{(N-P)\times V\times V}$, genotype $\mathbf{z}$ with $\widetilde{\mathbf{z}} = \boldsymbol{U}\mathbf{z}$ and the kinship matrix Λ with $\widetilde{\Lambda} = \boldsymbol{U}\Lambda\boldsymbol{U}^T$, the joint posterior distribution will follow the same structure as (2.5). Hence, all sampling procedures can be adapted accordingly. In the following numerical studies, we also confirm that our model complemented under this projection approach achieves consistent results with BNME_{adj}.

3. SIMULATION STUDIES

We carry out simulation studies to evaluate the performance of BNME to uncover genetic signals and the associated phenotypic network configurations under related samples. To mimic the data dimension in our data application, we assign sample sizes $N = 100$ and 500 with brain connectivity generated under a brain atlas with $V = 50$. We consider two scenarios on the phenotypic network configurations that are highly impacted by the genetic factor. In the first scenario, we generate a single phenotypic network configuration that is linked with the genetic variant, and we set $\eta_1 = 1$. In the second scenario, we create a more challenging setting by generating three network configurations with the associated weight parameter η_h equals 0.7, 0.3 and 0, respectively. The third network configuration is not linked to the genetic variant, allowing us to evaluate the performance of our model in detecting the true number of signaling phenotypic components. For both scenarios, we consider a range of sparsity levels for each $\boldsymbol{\theta}_h$ by imposing 50%, 90% and 100% of the elements

within the vector to be zero to define the genetically associated network configurations. As shown in Web Figure 1, we provide the signal patterns upon the whole network phenotype under 50% and 90% sparsity levels for the second scenario assembled across network configurations. Of note, when sparsity level is 100%, the genotype does not impact any of the phenotypic structures, facilitating a test on a noise genetic variant. For the genetic and environmental effects, we first generate a kinship matrix Λ with diagonal entries to be 1 and off-diagonal entries ranging from $(0, 1)$, and consider two scenarios for their variance components. In the first scenario, we set $\sigma^{(a)}_{vv'}$ to be 1.5 and $\sigma^{(e)}_{vv'}$ to be 1 with effects across brain locations to be independent. In the second scenario, we evaluate the robustness of our methods by imposing brain spatial correlation among the effect elements. Specifically, we simulate a random correlation matrix $\Omega_{V\times V}$ and generate $\text{vec}(\mathcal{B})$ and $\text{vec}(\mathcal{E})$ from Normal distributions with covariance matrices $3\Omega \otimes \Delta$ and $\Omega \otimes \text{I}_N$, respectively. Finally, for the fix effects, we sample the genotype for each subject from $\{0, 1, 2\}$, and add three different types of covariates including one generated from a Bernoulli distribution $\text{Bern}(0.5)$, one from a Uniform distribution $\mathcal{U}(-0.5, 0.5)$, and one from a Normal distribution $\mathcal{N}(-0.5, 0.5)$. Each of the fixed effect coefficients are generated from $\mathcal{N}(0.3, 0.5)$ and fixed for all the settings. Overall, we consider 24 settings with different sample sizes and phenotypic signal patterns, and we generated 200 Monte Carlo datasets for each setting.

We implement the proposed BNME along with two variations BNME_{DP} and BNME_{adj}. To assess the robustness of the models, we set $H = 3$ which is larger than the actual number of the associated phenotypic network configurations for both scenarios. We also set $\alpha_0 = \beta_0 = 0.01$, and determine ν by a grid search from $(0.5, 0.8, 1)$ based on BIC. The MCMC algorithm is performed for 5000 iterations after 2000 burn-in, and both trace plots and GR value indicate a convergence. For the competing methods, given there is no existing regression approach that can accommodate a network outcome with mixed effects, we extract unique edges from the phenotype matrix. With each of the upper diagonal elements of $\mathbf{A}_i$ as a phenotypic trait, we implement a linear mixed-effect model (LMM) using the `lme4` package in `R`, linear mixed-effects kinship model (LMEKIN) using the `coxme` package and one of the most popular GWAS pipelines for related samples Genome-wide Efficient Mixed Model Association (GEMMA) (Zhou and Stephens, 2012). To evaluate both estimation and feature selection, we consider the following performance metrics: (a) root mean predicted square error (RMSE) of $\boldsymbol{\Theta}$, (b) sensitivity (Sen_e) and specificity (Spe_e) for distinguishing signaling phenotypic elements captured by the nonzero elements in $\boldsymbol{\Theta}$, and (c) specificity (Spe_g) for identifying noise genetic variant when sparsity level is 100%. The simulation results are summarized in Tables 1 and 2 separated by variance generation scenarios.

Based on the results, we conclude that our proposed BNME along with BNME_{DP} and BNME_{adj} demonstrate excellent performance in uncovering genetic effects, identifying associated phenotypic network configurations, and distinguishing noise genetic variants. Specifically, the proposed methods exhibit significantly smaller RMSEs compared to alternative methods indicating higher estimation accuracy. Our methods also achieve over 90% phenotypic sensitivity and specificity across all the simulation settings, and genotypic specificity when the sparsity level is 100%, indicating their ability to uncover the associated phenotypic networks for the risk genotype and distinguish the noise genetic variant. When comparing different settings, we consistently observe improvements in performance metrics for all methods as the sample size increases. Interestingly, the correlation of effect components across brain spatial locations appears to have minimal influence on the results. As anticipated, a higher sparsity level aids in signal identification for all the methods. Notably, when sparsity reaches 100% with no associated phenotypic connections, given that our methods allow to exclude the noise phenotypic component entirely, it successfully detects this situation as evidenced by a close to one Spe_g. Moreover, as more phenotypic network configurations are impacted, including a noise network configuration, we observe a notable decrease in the accuracy of phenotypic feature selection for all competing methods. However, our methods maintain their superior performance, indicating robustness and ability to uncover the true signaling phenotypic

Table 1. Simulation results for all the methods when random effects and random errors are independent under different settings range from sample sizes, sparsity levels and phenotypic network configurations. The results are summarized over 200 MC datasets and the standard deviations are included in the parenthesis.

			$N = 100$				$N = 500$			
# Sub	Sparsity	Model	RMSE	Spe_e	Sen_e	Spe_g	RMSE	Spe_e	Sen_e	Spe_g
		BNME	0.13 (0.05)	0.96 (0.04)	1.00 (0.00)	–	0.04 (0.02)	0.97 (0.03)	1.00 (0.00)	–
		LMM	0.71 (0.22)	0.94 (0.12)	0.86 (0.10)	–	0.32 (0.06)	0.95 (0.05)	0.98 (0.01)	–
	50%	LMEKIN	0.25 (0.10)	0.94 (0.14)	0.93 (0.05)	–	0.24 (0.07)	0.95 (0.05)	0.98 (0.03)	–
		GEMMA	0.25 (0.13)	0.94 (0.14)	0.93 (0.03)	–	0.20 (0.04)	0.95 (0.05)	0.99 (0.00)	–
		$BNME_{adj}$	0.07 (0.02)	0.95 (0.03)	1.00 (0.00)	–	0.03 (0.01)	0.98 (0.03)	1.00 (0.00)	–
		$BNME_{DP}$	0.17 (0.12)	0.94 (0.05)	1.00 (0.00)	–	0.08 (0.06)	0.97 (0.03)	1.00 (0.00)	–
		BNME	0.53 (0.14)	0.99 (0.01)	1.00 (0.00)	–	0.34 (0.30)	0.99 (0.01)	1.00 (0.00)	–
		LMM	0.58 (0.02)	0.94 (0.01)	0.99 (0.02)	–	0.35 (0.26)	0.95 (0.04)	1.00 (0.00)	–
1	**90%**	LMEKIN	0.26 (0.07)	0.93 (0.03)	0.99 (0.01)	–	0.28 (0.06)	0.95 (0.05)	0.99 (0.00)	–
		GEMMA	0.23 (0.09)	0.95 (0.06)	0.99 (0.01)	–	0.22 (0.03)	0.95 (0.02)	0.99 (0.00)	–
		$BNME_{adj}$	0.56 (0.16)	0.99 (0.01)	1.00 (0.00)	–	0.35 (0.28)	0.99 (0.01)	1.00 (0.00)	–
		$BNME_{DP}$	0.57 (0.12)	0.96 (0.02)	1.00 (0.01)	–	0.38 (0.12)	0.99 (0.01)	1.00 (0.00)	–
		BNME	0.01 (0.02)	1.00 (0.01)	–	0.96	0.01 (0.01)	1.00 (0.01)	–	1.00
		LMM	0.58 (0.02)	0.95 (0.01)	–	0.00	0.25 (0.01)	0.95 (0.01)	–	0.00
	100%	LMEKIN	0.22 (0.02)	0.95 (0.02)	–	0.00	0.25 (0.01)	0.95 (0.01)	–	0.02
		GEMMA	0.22 (0.01)	0.95 (0.01)	–	0.00	0.25 (0.01)	0.94 (0.01)	–	0.00
		$BNME_{adj}$	0.02 (0.02)	0.99 (0.01)	–	0.95	0.00 (0.00)	1.00 (0.00)	–	1.00
		$BNME_{DP}$	0.03 (0.03)	0.99 (0.01)	–	0.89	0.01 (0.01)	0.95 (0.01)	–	0.97
		BNME	0.14 (0.03)	0.90 (0.04)	0.92 (0.08)	–	0.09 (0.03)	0.93 (0.05)	0.88 (0.08)	-
		LMM	0.71 (0.23)	0.90 (0.12)	0.62 (0.12)	–	0.32 (0.06)	0.95 (0.05)	0.87 (0.06)	–
	50%	LMEKIN	0.39 (0.11)	0.93 (0.14)	0.70 (0.18)	–	0.23 (0.06)	0.93 (0.04)	0.90 (0.03)	–
		GEMMA	0.39 (0.09)	0.94 (0.03)	0.69 (0.19)	–	0.23 (0.06)	0.95 (0.05)	0.94 (0.04)	–
		$BNME_{adj}$	0.13 (0.04)	0.90 (0.05)	0.89 (0.07)	–	0.08 (0.02)	0.95 (0.05)	0.90 (0.08)	–
		$BNME_{DP}$	0.20 (0.08)	0.96 (0.06)	0.94 (0.10)	–	0.09 (0.03)	0.95 (0.05)	0.91 (0.06)	–
		BNME	0.21 (0.06)	0.99 (0.01)	0.95 (0.09)	–	0.12 (0.20)	0.99 (0.02)	1.00 (0.00)	–
		LMM	0.71 (0.23)	0.94 (0.12)	0.73 (0.13)	–	0.25 (0.01)	0.95 (0.01)	0.97 (0.03)	–
3	**90%**	LMEKIN	0.23 (0.10)	0.95 (0.09)	0.85 (0.09)	–	0.21 (0.05)	0.96 (0.02)	0.95 (0.05)	–
		GEMMA	0.21 (0.06)	0.96 (0.07)	0.85 (0.09)	–	0.21 (0.05)	0.95 (0.01)	0.99 (0.01)	–
		$BNME_{adj}$	0.23 (0.06)	0.99 (0.01)	0.96 (0.08)	–	0.25 (0.28)	0.99 (0.01)	1.00 (0.00)	–
		$BNME_{DP}$	0.23 (0.09)	0.95 (0.05)	0.90 (0.10)	–	0.13 (0.08)	0.98 (0.02)	1.00 (0.00)	–
		BNME	0.01 (0.01)	1.00 (0.00)	–	1.00	0.02 (0.04)	0.98 (0.02)	–	0.90
		LMM	0.58 (0.02)	0.94 (0.01)	–	0.00	0.25 (0.01)	0.94 (0.01)	–	0.00
	100%	LMEKIN	0.22 (0.09)	0.95 (0.03)	–	0.00	0.21 (0.15)	0.95 (0.04)	–	0.03
		GEMMA	0.23 (0.13)	0.95 (0.01)	–	0.01	0.24 (0.01)	0.94 (0.01)	–	0.01
		$BNME_{adj}$	0.02 (0.02)	0.99 (0.00)	–	1.00	0.01 (0.00)	0.99 (0.01)	–	0.95
		$BNME_{DP}$	0.08 (0.06)	0.98 (0.02)	–	0.93	0.05 (0.05)	0.99 (0.01)	–	0.93

*Phenotypic sensitivity does not exist at a 100% sparse level with no connection associated with the genotype.

network configurations even under a misspecified network configuration number H. In the comparison among competing methods, both LMEKIN and GEMMA demonstrate similar performance, surpassing the traditional LMM. Their performance in the presence of a noise genotype suggests a high risk of false positives when considering GWAS under a network phenotype.Finally, the performance between BNME and the variations $BNME_{DP}$ and $BNME_{adj}$ are highly consistent, including the scenarios with spatially correlated effect components (Table 2). This suggests that the prior independence assumption for variance components across brain locations brings a negligible impact on the model performance, and the application of a projection approach for covariate

Table 2. Simulation results for all the methods when random effects and random errors are correlated under different settings range from sample sizes, sparsity levels and phenotypic network configurations. The results are summarized over 200 MC datasets and the standard deviations are included in the parenthesis.

			N=100				N=500			
# Sub	Sparsity	Model	RMSE	Spe_e	Sen_e	Spe_g	RMSE	Spe_e	Sen_e	Spe_g
1	50%	BNME	0.13 (0.05)	0.96 (0.03)	1.00 (0.00)	-	0.04 (0.02)	0.98 (0.04)	1.00 (0.00)	–
		LMM	0.68 (0.24)	0.94 (0.13)	0.88 (0.01)	-	0.30 (0.07)	0.95 (0.05)	0.99 (0.01)	–
		LMEKIN	0.49 (0.12)	0.88 (0.14)	0.90 (0.11)	–	0.29 (0.10)	0.91 (0.07)	0.95 (0.05)	–
		GEMMA	0.50 (0.09)	0.86 (0.10)	0.90 (0.12)	-	0.33 (0.10)	0.89 (0.10)	0.90 (0.06)	–
		$BNME_{adj}$	0.08 (0.02)	0.94 (0.05)	1.00 (0.00)	-	0.03 (0.01)	0.98 (0.03)	1.00 (0.00)	–
		$BNME_{DP}$	0.13 (0.05)	0.95 (0.02)	0.98 (0.03)	-	0.05 (0.03)	0.99 (0.06)	1.00 (0.00)	–
	90%	BNME	0.56 (0.16)	0.99 (0.01)	1.00 (0.00)	–	0.35 (0.28)	0.99 (0.01)	1.00 (0.00)	–
		LMM	0.60 (0.02)	0.95 (0.01)	0.99 (0.02)	–	0.35 (0.07)	0.95 (0.05)	1.00 (0.00)	–
		LMEKIN	0.26 (0.05)	0.94 (0.03)	0.99 (0.00)	–	0.29 (0.05)	0.95 (0.03)	0.99 (0.01)	–
		GEMMA	0.26 (0.10)	0.95 (0.03)	0.99 (0.01)	–	0.27 (0.06)	0.97 (0.02)	0.99 (0.01)	–
		$BNME_{adj}$	0.35 (0.14)	0.99 (0.01)	1.00 (0.00)	–	0.27 (0.21)	0.92 (0.02)	1.00 (0.00)	–
		$BNME_{DP}$	0.60 (0.10)	0.97 (0.05)	0.99 (0.01)	–	0.43 (0.22)	0.95 (0.02)	0.99 (0.00)	–
	100%	BNME	0.01 (0.02)	0.99 (0.00)	–	1.00	0.01 (0.01)	1.00 (0.00)	–	1.00
		LMM	0.54 (0.02)	0.95 (0.01)	–	0.00	0.24 (0.01)	0.95 (0.01)	–	0.00
		LMEKIN	0.20 (0.01)	0.95 (0.02)	–	0.03	0.23 (0.03)	0.94 (0.01)	–	0.02
		GEMMA	0.24 (0.02)	0.95 (0.01)	–	0.00	0.24 (0.01)	0.94 (0.01)	–	0.01
		$BNME_{adj}$	0.03 (0.03)	0.99 (0.00)	–	0.99	0.01 (0.01)	1.00 (0.00)	–	1.00
		$BNME_{DP}$	0.03 (0.05)	0.99 (0.01)	–	0.98	0.02 (0.03)	1.00 (0.00)	–	1.00
3	50%	BNME	0.15 (0.03)	0.90 (0.01)	0.90 (0.02)	–	0.09 (0.03)	0.95 (0.06)	0.89 (0.08)	–
		LMM	0.77 (0.29)	0.85 (0.10)	0.81 (0.11)	–	0.30 (0.07)	0.96 (0.05)	0.88 (0.09)	–
		LMEKIN	0.36 (0.08)	0.94 (0.10)	0.70 (0.12)	-	0.23 (0.10)	0.91 (0.03)	0.90 (0.06)	–
		GEMMA	0.40 (0.10)	0.95 (0.06)	0.69 (0.10)	–	0.24 (0.09)	0.91 (0.01)	0.93 (0.10)	–
		$BNME_{adj}$	0.12 (0.02)	0.90 (0.05)	0.91 (0.07)	-	0.07 (0.02)	0.93 (0.07)	0.91 (0.08)	-
		$BNME_{DP}$	0.13 (0.05)	0.91 (0.02)	0.91 (0.03)	-	0.10 (0.05)	0.92 (0.06)	0.89 (0.05)	-
	90%	BNME	0.23 (0.06)	0.99(0.01)	0.96 (0.08)	–	0.15 (0.11)	0.99 (0.01)	0.98 (0.04)	–
		LMM	0.68 (0.24)	0.76 (0.14)	0.94 (0.13)	–	0.24 (0.08)	0.93 (0.01)	0.97 (0.05)	–
		LMEKIN	0.27 (0.08)	0.96 (0.08)	0.84 (0.10)	–	0.23 (0.04)	0.96 (0.03)	0.96 (0.03)	–
		GEMMA	0.26 (0.08)	0.98 (0.03)	0.88 (0.06)	–	0.23 (0.08)	0.96 (0.02)	0.97 (0.06)	–
		$BNME_{adj}$	0.23 (0.12)	0.99 (0.01)	0.96 (0.09)	–	0.31 (0.16)	0.99 (0.02)	1.00 (0.00)	–
		$BNME_{DP}$	0.25 (0.03)	0.99 (0.02)	0.96 (0.10)	–	0.16 (0.04)	0.99 (0.01)	0.99 (0.03)	–
	100%	BNME	0.01 (0.01)	1.00 (0.00)	–	1.00	0.01 (0.01)	1.00 (0.00)	–	1.00
		LMM	0.53 (0.02)	0.95 (0.01)	–	0.00	0.54 (0.01)	0.95 (0.00)	–	0.00
		LMEKIN	0.20 (0.05)	0.95 (0.01)	–	0.01	0.23 (0.02)	0.95 (0.02)	–	0.00
		GEMMA	0.20 (0.08)	0.96 (0.02)	–	0.00	0.21 (0.01)	0.96 (0.01)	–	0.02
		$BNME_{adj}$	0.02 (0.02)	0.98 (0.01)	–	0.95	0.01 (0.01)	1.00 (0.00)	–	1.00
		$BNME_{DP}$	0.02 (0.01)	0.99 (0.01)	–	0.98	0.01 (0.01)	1.00 (0.00)	–	1.00

*Phenotypic sensitivity does not exist at the 100% sparse level with no connection associated with the genotype.

adjustment in BNME is validated. From a computational standpoint, we advocate for BNME, considering that $BNME_{DP}$ and $BNME_{adj}$ require approximately 18% and 30% more posterior computational time than BNME, respectively.

4. REAL DATA APPLICATION

4.1. Imaging genetics data for HCP

We implement our model to the Human Connectome Project (HCP) data. HCP is a landmark study that has collected a rich set of imaging, behavioral and genetic data. In the current analyses,

Table 3. Significant genetic variants and their associated phenotypic structural network configurations, along with cis-eQTL results obtained from UKBEC brain database.

	eQTL			Phenotypic network configurations	
SNP	Chromosome	p-value	Regulated genes	# Association	Macroscale systems
rs2465095	2	9.30E-03	THSD7B	91	Subcortical, parietal lobe
rs1918367	2	3.50E-02	GALNT13	91	Subcortical, parietal lobe
rs4725467	7	2.20E-02	GALNTL5	325	Subcortical, temporal lobe
rs10760611	9	5.00E-03	ASB6	20	Subcortical
rs4948428	10	2.50E-02	TMEM26	6	Subcortical
rs1537969	13	5.50E-02	SGCG	22	Subcortical, temporal lobe
rs9928439	16	2.50E-02	SLC38A8	91	Subcortical, temporal lobe
rs6563992	16	1.30E-03	ATP2C2	15	Subcortical
rs58090793	16	3.30E-03	ZDHHC7	3	Frontal lobe

we adopt the WU-Minn HCP minimally processed S1200 release that includes over 1,000 young adults aged 22 to 37 years. For each subject, both T1 magnetic resonance imaging (MRI) and diffusion MRI (dMRI) are available, allowing the construction of brain structural connectivity to capture the white matter fiber tracts connecting different brain regions. Specifically, based on the minimally prepossessed dMRI and T1 data from ConnectomeDB, we first generate the whole-brain tractography for each subject, and perform the anatomical parcellation via Desikan-Killiany (DK) atlas (Desikan et al., 2006) including 68 cortical surface regions and 19 subcortical regions. To extract the streamlines linking each pair of ROIs, a series of steps including dilation of each gray matter ROI to incorporate white matter regions, separation of the streamlines connecting several ROIs into parts, and removing obvious outlier streamlines are conducted. Subsequently, the mean fractional anisotropy (FA) value along streamlines is used to evaluate the strength of structural connections. Eventually, we construct brain structural connectivity for 1,065 subjects. Comprehensive details are available elsewhere on HCP neuroimaging protocols (Van Essen et al., 2013) and our tractography pipeline (Zhao et al., 2023).

The young adult participants in HCP were also genotyped by Illumina's MultiEthnic Global Array (MEGA) Chip and three specialized neuroimaging chips: Psych, NeuroX, and Immunochip. After standard data quality by excluding subjects with more than 10% missing SNPs or sex check failure, 1,010 subjects with both genotypes and phenotypes are included in our analyses. For the genetic variants, to mitigate computational cost, we focus on the 1,860 SNPs that were identified in the previous study to highly associate with brain structural network (Zhao et al., 2023). However, unlike the previous analyses that didn't accommodate the sample relatedness, we consider family structure after creating the kinship matrix for 149 pairs of genetically-confirmed monozygotic twins (298 participants), 94 pairs of genetically-confirmed dizygotic twins (188 participants) and their non-twin siblings (524 participants). All the model implementations closely follow the simulation studies, and we account for age, gender, and the top ten genetic principal components. The computational cost for each model is around 15 h under Yale High-Performance Computing (one CPU core, 3GB RAM) and we apply parallel computing across the models. A demonstration of the model convergence is provided in supplementary material S.4.

4.2. Analysis results

Our goal is to identify risk genetic markers and their associated brain connectivity phenotypic components. Based on the posterior samples of η, we identify nine risk SNPs as shown in Table 3. After mapping those SNPs to the genes they belong to, we identify five unique gene variants including THSD7B, LINC01503, LOC105373693, CDH13 and SLC38A8. Among them, THSD7B and CDH13 have been considered to play an essential role in the development of the central nervous system and neural connectivity (Wang et al., 2011; Polanco et al., 2021). Particularly,

THSD7B has also been shown to be associated with intellectual disability (Lyons-Warren et al., 2022); and CDH13 is related to various psychiatric disorders including ADHD and substance abuse (Rivero et al., 2013; Treutlein and Rietschel, 2011). To evaluate the neurogenetic processes of the selected genetic variants, we further perform a brain tissue-specific expression quantitative trait loci (eQTL) analysis via the UK Brain Expression Consortium (UKBEC) (Ramasamy et al., 2014). The consortium generated genotype and exon-specific expression data for 134 neuropathologically healthy subjects under ten different brain tissues, which allows us to evaluate each identified genetic variant on its alteration of tissue-specific and cross-tissue gene expressions within 100kb of the SNP. Table 3 Column 3 shows the cross-tissue cis-effect p-value calculated in their BRAINEAC web server, and the regulated genes for each risk SNP. The small p-values of cross-tissue eQTLs reflect the molecular regulation through gene expression over different brain areas, aligning with the circular nature of brain network phenotypes.

We further investigate the associated brain network configurations and phenotypic components for each of the identified genetic signals. Visualization of each genetically associated brain network component is displayed in Figure 1, where the color of connections indicates the effect size of genetic association. Additionally, we summarize the macroscale structures involved in network configurations for each identified SNP and add it to Table 3. Our analysis reveals that cross-hemispheric connections and inter-subcortical connections account for the largest proportion of all the signaling connections. This finding agrees with the previous literature, which has consistently demonstrated that genetic effects lead to alterations in white matter fiber tracts across brain hemispheres and subcortical structures (Jahanshad et al., 2013; Zhong et al., 2021).

Finally, we also implement GEMMA to the HCP data. Given that GEMMA is applied on each brain connection individually, we adjust p-value to 1.34×10^{-5} accounting for the 3741 unique connections among 87 ROIs. As a result, GEMMA identifies a total of 36 SNPs that exhibit significant associations with at least one brain connection. To assess the agreement in the top selected genetic variants between the two approaches, we map the top 36 selected SNPs from each method to their associated cytogenetic bands (Clark and Pazdernik, 2016) and examine the overlap in signals. Eventually, there are ten cytogenetic bands that encompass the genetic signals identified by both BNME and GEMMA. This indicates a certain degree of consistency in the genetic signals identified by the two methods, which lends support to the plausibility and reliability of our results. The detailed results are provided in the supplementary materials.

Furthermore, we also visualize the number of associated brain connections for each of the top selected SNPs under both methods respectively in Figure 2. It is evident that, in contrast to BNME, which dissects a phenotypic network configuration architecture for each genetic variant, the phenotypic signals identified under GEMMA appear to be extremely sparse and scattered. This result indicates that the majority of the SNPs identified under GEMMA are associated with a single brain connection, raising questions regarding the biological interpretability and meaningfulness of the observed genetic associations.

5. DISCUSSION

In this article, we present a Bayesian network-response mixed-effect model that addresses the challenges of genetic association studies in brain connectivity. Our model is specifically designed to capture the genetic contributions to phenotypic network configurations while accounting for family structures and unknown sample relatedness. To accommodate the biological architecture in the network phenotype, we consider the genetic variant influences the phenotype via a set of unknown network configurations, where the targeted phenotypic networks are uncovered through a hierarchical selection procedure. Through posterior inference, we quantify the uncertainty associated with determining a risk genetic variant and its impact on the network phenotype. Extensive simulations demonstrate the superiority of our method in estimating genetic effects and identifying relevant phenotypic elements with signaling capabilities. By applying the proposed method to the

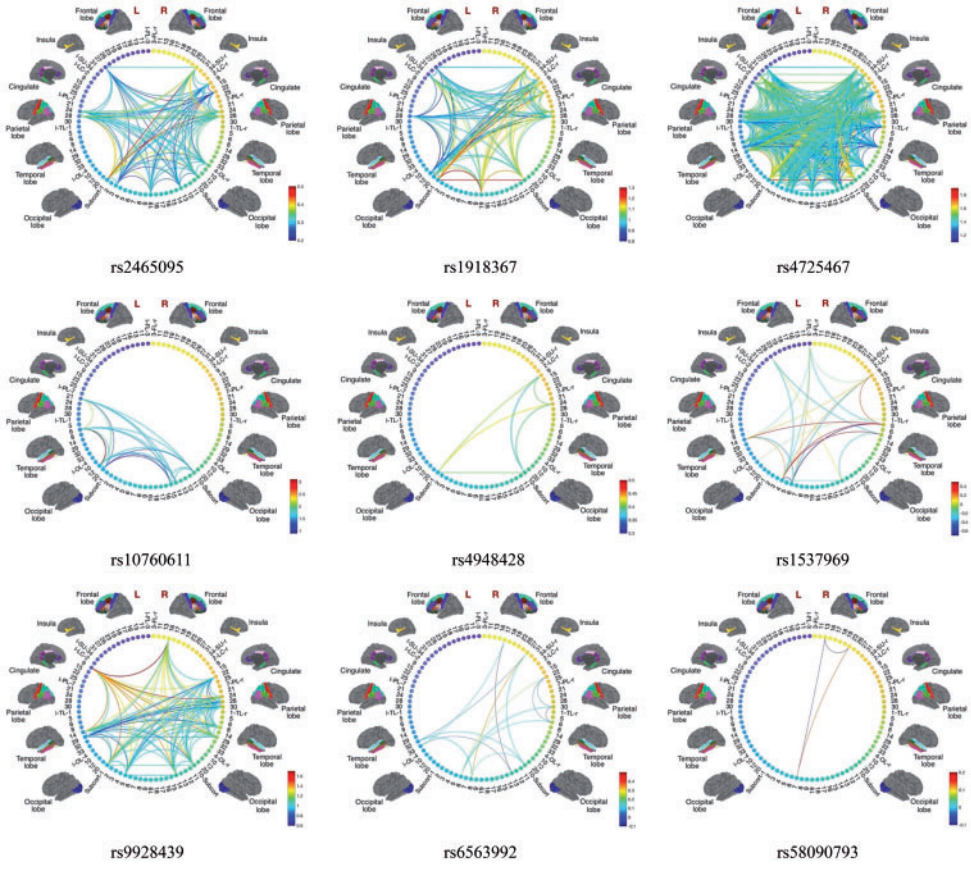

Figure 1. The identified risk genetic variants under the BNME model and their associated brain network configurations.

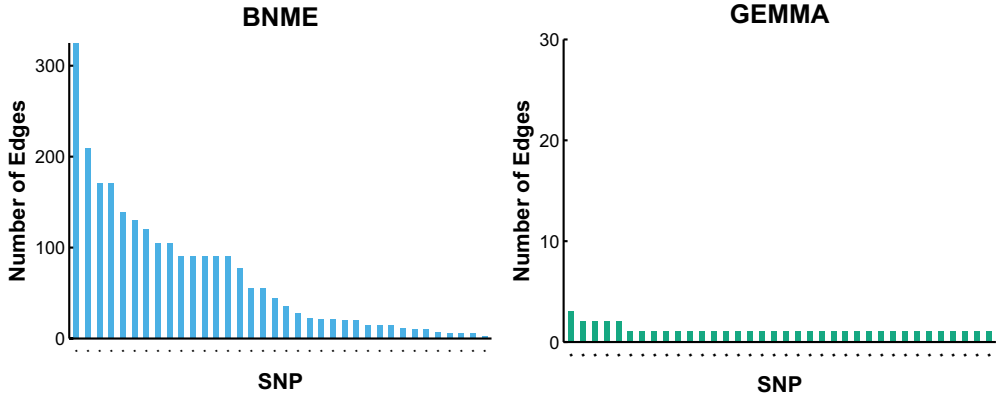

Figure 2. The number of the highly associated phenotypic connections for each of the top selected genetic risk variants obtained by BNME and GEMMA, respectively.

HCP cohort with excessive family structures, we obtain biologically interpretable results that shed light on uncovering the genetic underpinnings of brain structural connectivity.

In addition to the current application to brain connectivity genetics studies, the proposed BNME model provides a fundamental framework for mixed-effect models involving network- or

matrix-variate outcomes. As data collection in epidemiology and social studies becomes more complex, there is a growing need to analyze network-related or matrix-structured outcomes arising from related samples caused by pedigree or repeated measurements. By extending the random effect tensor $\mathcal{B}$ to include an additional dimension corresponding to random slopes, along with the associated variance–covariance component, we can effectively capture more intricate sources of variation and address diverse modeling requirements.

Our current model formulation employs a decomposition of the effect matrix into a series of weighted outer products. This design choice aligns well with the biological assumptions inherent in our application and facilitates the interpretation of results. However, in cases where prior knowledge suggests alternative association structures, such as a modular structure, one can easily modify the model (2.2) by adopting a different decomposition approach, such as a stochastic block model. Moreover, our proposed model can be readily extended to perform heritability analyses for network phenotypes. As a fundamental quantitative genetic analysis, the existing heritability analyses only consider scalar- or vector-variate phenotypes. By adapting our model to this future direction, we could contribute to filling this literature gap and provide valuable insights into the heritability of network-related traits.

DATA AVAILABILITY

Implementation of BNME is available at https://github.com/xt83/Bayesian_mixed_model_inference_for_genetic_association_under_related_samples.

FUNDING

This work was partially supported by the National Institutes of Health grants R01MH126970, RF1AG068191, R01MH126970 and RF1AG081413.

Conflict of interest statement: None declared.

SUPPLEMENTARY MATERIAL

Supplementary material is available online at *Biostatistics Journal* online.

REFERENCES

Chang C, Kundu S, Long Q. Scalable Bayesian variable selection for structured high-dimensional data. Biometrics, 2018:74(4):1372–1382.

Clark DP, Pazdernik NJ. Chapter 8-Genomics and Gene Expression. In: Clark DP, Pazdernik NJ, editors. Biotechnology. 2nd ed. Boston: Academic Cell; 2016.

Desikan RS, Sgonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. NeuroImage. 2006:31:968 – 980.

Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G, Marchini J, Smith SM. Genome-wide association studies of brain imaging phenotypes in UK biobank. Nature. 2018:562(7726):210–216.

Elsheikh SSM, Chimusa ER, Mulder NJ, Crimi A. Genome-wide association study of brain connectivity changes for Alzheimer's disease. Sci Rep. 2020:10(1):1433.

Eu-Ahsunthornwattana J, Miller EN, Michaela F. Wellcome Trust Case Control Consortium, Jeronimo SMB, Blackwell JM, Cordell HJ. Comparison of methods to account for relatedness in genome-wide association studies with family-based data. PLOS Genet.2014:10(7):e1004445.

Ge T, Reuter M, Winkler AM, Holmes AJ, Lee PH, Tirrell LS, Roffman JL, Buckner RL, Smoller JW, Sabuncu MR. Multidimensional heritability analysis of neuroanatomical shape. Nat Commun. 2016:7:13291.

Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci. 1992:7(4):457–472.

Hastie T, Tibshirani R, Friedman J. Optimal predictive model selection. J R Stat Soc: Ser B. 2004:66(2):209–233.

Helgason A, Yngvadottir B, Hrafnkelsson B, Gulcher J, Stefánsson, K. An icelandic example of the impact of population structure on association studies. Nat Genet. 2005:37(1):90–95.

Jahanshad N, Rajagopalan P, Hua X, Hibar DP, Nir TM, Toga AW, Jack Jr CR, Saykin AJ, Green, RC, Weiner MW, et al. Genome-wide scan of healthy human connectome discovers spon1 gene variant influencing dementia severity. Proc Nat Acad Sci. 2013:110(12):4768–4773.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-y, Freimer NB, Sabatti C, Eskin E. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2010:42(4): 348–354.

Kong D, An B, Zhang J, Zhu H. L2RM: Low-rank Linear Regression Models for High-dimensional Matrix Responses. J Am Stat Assoc. 2020:115(529):403–424.

Li F, Zhang NR. Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. J Am Stat Assoc. 2010:105(491):1202–1214.

Li, Fan, Zhang, Tingting, Wang, Quanli, Gonzalez, Marlen Z., Maresh, Erin L, Coan JA. Spatial Bayesian variable selection and grouping for high-dimensional scalar-on-image regression. Ann Appl Stat. 2015:9(2):687–713.

Lyons-Warren AM, Wangler MF, Wan YW. Cluster analysis of short sensory profile data reveals sensory-based subgroups in autism spectrum disorder. Int J Molec Sci. 2022:23(21):13030.

Park T, Casella G. The Bayesian lasso. J Am Stat Assoc. 2008:103(482):681–686.

Polanco J, Reyes-Vigil F, Weisberg SD, Dhimitruka I, Brusés JL. Differential spatiotemporal expression of type i and type ii cadherins associated with the segmentation of the central nervous system and formation of brain nuclei in the developing mouse. Front Molec Neurosci. 2021:14:633719.

Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, De T, Robert UK, Brain Expression Consortium, John H, Mina R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci. 2014:17(10):1418–1428.

Rivero O, Sich S, Popp S, Schmitt A, Franke B, Lesch, K-P. Impact of the adhd-susceptibility gene cdh13 on development and function of brain networks. Eur Neuropsychopharmacol 2013:23(6):492–507.

Stingo FC, Chen YA, Tadesse, MG, Vannucci M. Incorporating biological information into linear models: A Bayesian approach to the selection of pathways and genes. Ann Appl Stat. 2011:5(3):1202–1214.
Stingo FC, Chen YA, Tadesse, MG, Vannucci M. Incorporating biological information into linear models: A Bayesian approach to the selection of pathways and genes. Ann Appl Stat. 2011:5(3).

Treutlein J, Rietschel M. Genome-wide association studies of alcohol dependence and substance use disorders. *Curr Psychiatry Rep*. 2011:13:147–155.

Van Essen DC, Smith SM, Barch DM, Behrens, TEJ, Yacoub E, Ugurbil K, for the WU-Minn HCP Consortium. The wu-minn human connectome project: an overview. Neuroimage. 2013:80:62–79.

Wang KS, Liu X, Zhang Q, Pan Y, Aragam N, Zeng M. A meta-analysis of two genome-wide association studies identifies 3 new loci for alcohol dependence. J Psychiatric Res. 2011:45(11):1419–1425.

Wang L, Lin FV, Cole M, Zhang Z. Learning clique subgraphs in structural brain network classification with application to crystallized cognition. NeuroImage. 2021:225:117493.

Wang Y, Guo Y. 2020. Locus: a novel decomposition method for brain network connectivity matrices using low-rank structure with uniform sparsity. arXiv:2008.08915.

Zhang J, Sun WW, Li L. Mixed-effect time-varying network model and application in brain connectivity analysis. J Am Stat Assoc. 2020:115(532):2022–2036.

Zhang J, Sun WW, Li L. Generalized connectivity matrix response regression with applications in brain connectivity studies. Comput Graph Stat. 2023:32(1), 252–262.

Zhao B, Li T, Yang Y, Wang X, Luo T, Shan Y, Zhu Z, Xiong D, Hauberg ME, Bendl J, et al. Common genetic variation influencing human white matter microstructure. Science. 2021:372(6548):eabf3736.

Zhao Y, Chang C, Zhang J, Zhang Z. Genetic underpinnings of brain structural connectome for young adults. J Am Stat Assoc. 2023: 118(543):1473–1487.
Zhao Y, Chang C, Zhang J, Zhang Z. Genetic underpinnings of brain structural connectome for young adults. J Am Stat Assoc. 2023:1–15.

Zhao Y, Li T, Zhu H. Bayesian sparse heritability analysis with high-dimensional neuroimaging phenotypes. *Biostatistics* 2022:23(2):467–484.

Zhong S, Wei L, Zhao C, Yang, L, Di Z, Francks C, Gong G. Interhemispheric relationship of genetic influence on human brain connectivity. Cereb Cortex 2021:31(1):77–88.

Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012:44(7):821–824.