



Published in final edited form as:

Neuroimage. 2017 April 01; 149: 305–322. doi:10.1016/j.neuroimage.2017.01.052.

Bayesian longitudinal low-rank regression models for imaging genetic data from longitudinal studies

Zhao-Hua Lu^a, Zakaria Khondker^b, Joseph G. Ibrahim^b, Yue Wang^b, and Hongtu Zhu^{c,*} for the Alzheimer's Disease Neuroimaging Initiative¹

^aDepartment of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA

^bDepartment of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

^cDepartment of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX, USA

Abstract

To perform a joint analysis of multivariate neuroimaging phenotypes and candidate genetic markers obtained from longitudinal studies, we develop a Bayesian longitudinal low-rank regression (L2R2) model. The L2R2 model integrates three key methodologies: a low-rank matrix for approximating the high-dimensional regression coefficient matrices corresponding to the genetic main effects and their interactions with time, penalized splines for characterizing the overall time effect, and a sparse factor analysis model coupled with random effects for capturing within-subject spatio-temporal correlations of longitudinal phenotypes. Posterior computation proceeds via an efficient Markov chain Monte Carlo algorithm. Simulations show that the L2R2 model outperforms several other competing methods. We apply the L2R2 model to investigate the effect of single nucleotide polymorphisms (SNPs) on the top 10 and top 40 previously reported Alzheimer disease-associated genes. We also identify associations between the interactions of these SNPs with patient age and the tissue volumes of 93 regions of interest from patients' brain images obtained from the Alzheimer's Disease Neuroimaging Initiative.

Keywords

Genetic variants; Longitudinal imaging phenotypes; Low-rank regression; Markov chain Monte Carlo; Spatiotemporal correlation

1. Introduction

Many longitudinal neuroimaging studies concomitantly collect genetic and recurrent imaging data to track individual changes in brain structure and function over time. Several neurodegenerative disorders, including Alzheimer disease (AD), are hypothesized to occur from abnormal development of the brain, which may be caused by the additive and/or

*Corresponding author: hzhu5@mdanderson.org (Hongtu Zhu).

¹Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: <http://adni.loni.usc.edu/wp-content/uploads/howtoapply/ADNIAcknowledgementList.pdf>.

interactive effects of various risk genes and environmental risk factors, each contributing small individual effects. Thus, recurrent neuroimaging measures may lead to discoveries of the genetic pathways and the causal genes associated with the specific brain changes underlying such neurodegenerative disorders (Scharinger et al., 2010; Paus, 2010; Peper et al., 2007; Chiang et al., 2011a,b; Saykin et al., 2015).

A standard statistical method used in longitudinal imaging and genetics studies is the massive marginal association (MMA) framework (Li et al., 2013; Zhang et al., 2014; Guillaume et al., 2014; Hibar and et al, 2011; Shen et al., 2010; Bernal-Rusiel et al., 2013; Zhang et al., 2014). This approach repeatedly fits a linear mixed effects model (or generalized estimating equations) for paired imaging phenotypes and genetic markers. Because the MMA framework entails numerous comparisons, it can detect only phenotype-marker pairs with extremely strong signals.

Several attempts have been made to more precisely investigate the effect of multiple genotypes on longitudinal phenotypes. Chen and Wang (2011) proposed functional mixed-effect models with penalized splines and varying coefficients, but they focused on small number of predictors and number of response variables in a low-dimensional setting. Wang et al. (2011) used a sparse multitask regression to examine the association between genetic markers and longitudinal neuroimaging phenotypes. However, their model focused on subjects with the same number of repeated measures and ignored the spatiotemporal correlations of imaging phenotypes. Therefore, the multitask regression model may lead to loss of statistical power to detect phenotype-marker pairs with moderate to weak signals. Vounou et al. (2011) and Silver et al. (2012) proposed that a sparse reduced-rank regression model using penalized regression can detect the main genetic effects on longitudinal phenotypes. They, however, did not account for the spatiotemporal association among the longitudinal phenotypes, which is important for estimation and prediction accuracy. Moreover, none of these studies explored SNP-age interactions, which can reveal dynamic genetic effects on phenotypes.

Several important statistical concerns are associated with the joint analysis of neuroimaging phenotypes and a set of candidate genotypes obtained from longitudinal imaging and genetic studies. First, the number of regression coefficients can be much larger than the sample size, denoted as N . Specifically, let d and p be the dimension of the responses and the number of covariates, respectively. Fitting a multivariate linear mixed effects model usually requires estimating a $d \times p$ matrix of regression coefficients, which can be much larger than N , even for moderately high d and p . Second, as illustrated in Fig. 1a, to improve prediction accuracy (Breiman and Friedman, 1997), it is critically important to account for unstructured, within-subject spatial correlations among multivariate neuroimaging phenotypes. Third, as illustrated in Fig. 2a, to improve both estimation and prediction accuracy, it is also important to account for within-subject temporal correlation. Fourth, as shown in Fig. 2b,c, the temporal growth pattern varies across regions of interest (ROIs) in the brain. Accounting for the overall longitudinal change of ROIs is required to increase the detection power of the genetic effects. Fifth, as shown in Fig. 2b, the genetic effects on ROI volumes can vary across time.

Here, we have developed a Bayesian longitudinal low-rank regression (L2R2) model for the joint analysis of high-dimensional longitudinal responses and covariates. We integrated multiple robust methods to explicitly address the new challenges described previously. Our study has four major methodological contributions that were previously undescribed:

1. To the best of our knowledge, L2R2 is the first model of its kind for jointly analyzing high-dimensional longitudinal responses and covariates, although several approaches have been used for high-dimensional responses and covariates in cross-sectional studies (Rothman et al., 2010; Vounou et al., 2010; Zhu et al., 2014). The L2R2 model also provides a set of standard inference tools (e.g. standard deviation) for determining various unknown parameters. Zipunnikov et al. (2014) proposed a functional principle components analysis for high dimensional ($> 10,000$) longitudinal responses, where the intercepts and slopes of time for all voxels were modeled by a few basis functions. However, their methods cannot handle high-dimensional responses and covariates simultaneously because the dimension of the covariance matrix that requires time-consuming spectral decomposition equals the product of the dimensions of the responses and covariates.
2. The temporal growth patterns of high dimensional responses were characterized by a penalized spline method. Although the spline method has been widely used in other studies (Eilers and Marx, 1996; Ruppert et al., 2003; Wood, 2006; Wu and Zhang, 2006; Lang and Brezger, 2004; Guo, 2002; Morris and Carroll, 2006; Greven et al., 2010), most of these studies focused on univariate responses. We included a new set of coefficients for the high-dimensional gene-age interactions to model the genetic effects on longitudinal trajectories of responses, which were different from previously reported coefficients corresponding to the main genetic effects that were time-invariant (Marttinen et al., 2014; Zhu et al., 2014). A low-rank regression model not only reduces the number of unknown coefficients, but also captures the low-rank structure of the regression coefficient matrix.
3. To efficiently model the correlation structure and increase the detection power of association between response and covariates, we used a sparse factor model (Bhattacharya and Dunson, 2011) to capture the spatial correlation and the structured random effects to model the temporal correlation of the longitudinal response variables.
4. We have developed a downloadable L2R2 package, which is available at <http://odin.mdacc.tmc.edu/bigs2/>.

2. Methods: L2R2 model formulation

Let $y_{ik}(t_{ij})$ be the j -th observation at time t_{ij} of the k th neuroimaging measurement (e.g. volume of ROI) for the i th subject for $i = 1, \dots, n, j = 1, \dots, m_i$ and $k = 1, \dots, d, \mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ be a $p \times 1$ vector of genetic predictors (e.g. $p/2$ genetic markers and $p/2$ gene-age interactions), and $N = \sum_{i=1}^n m_i$ be the total number of observations. To address the

challenges of concomitantly analyzing high-dimensional longitudinal responses and covariates, we proposed the following L2R2 model: For each k , we have

$$y_{ik}(t) = \mathbf{x}_i^T \boldsymbol{\beta}_k + \mu_k(t) + \mathbf{w}_{2,i}(t)^T \boldsymbol{\gamma}_{2,k} + \mathbf{z}_i(t)^T \mathbf{b}_{ik} + \varepsilon_{ik}(t), \quad (1)$$

where $\boldsymbol{\beta}_k$ is a $p \times 1$ vector of coefficients for the genetic predictors, $\mu_k(t)$ is a nonparametric function of t that characterizes the overall trajectories of the k th measurement across all subjects, $\mathbf{w}_{2,i}(t) = (w_{2,i1}(t), \dots, w_{2,iq_2}(t))^T$ is a $q_2 \times 1$ vector of time-variant or time-invariant prognostic factors (e.g. age, gender), and $\boldsymbol{\gamma}_{2,k}$ is a $q_2 \times 1$ vector of coefficients for the prognostic factors. Moreover, \mathbf{b}_{ik} is a $p_b \times 1$ vector of latent random effects that characterize the within-subject correlation of the i th subject for the k th measurement, and $\mathbf{z}_i(t)$ is the related $p_b \times 1$ vector of covariates characterizing the correlation structure, and $\varepsilon_{ik}(t)$ is an error term. Let $\mathbf{b}_i = (\mathbf{b}_{i1}, \dots, \mathbf{b}_{id})$ such that

$$\text{vec}(\mathbf{b}_i) \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b) \text{ and } \varepsilon_i(t) \sim N_d(\mathbf{0}, \boldsymbol{\Sigma}_e = \boldsymbol{\Theta}^{-1}),$$

where $\mathbf{0}$ is a vector of zero of appropriate dimension, and $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\Sigma}_e$ are $p_b d \times p_b d$ and $d \times d$ covariance matrices, respectively. We assumed that \mathbf{b}_i and $\boldsymbol{\varepsilon}_i(t) = (\varepsilon_{i1}(t), \dots, \varepsilon_{id}(t))^T$ are independent for all i and t , \mathbf{b}_i and $\mathbf{b}_{i'}$ are independent for $i \neq i'$, and $\boldsymbol{\varepsilon}_i(t)$ and $\boldsymbol{\varepsilon}_{i'}(t')$ are independent for $i \neq i'$ or $t \neq t'$.

To address the challenges described above, we first incorporated the gene-age interactions into \mathbf{x}_i to model the dynamic genetic effects on individual brain development trajectories. The individual effect of a risk gene on brain development may vary across time. We used data published by the Alzheimer's Disease Neuroimaging Initiative (ADNI) to show the estimated trajectories with local polynomial regression (LOESS, Cleveland and Devlin, 1988) of subjects with and without the minor alleles for two selected phenotype-single nucleotide polymorphism (SNP) pairs (Fig. 2b,c). There was a strong interaction effect of the SNP rs769451 and age on the right hippocampus. Ignoring such interaction in model (1) may inflate errors, decrease the detection power of important main genetic effects, and miss the age-dependent genetic effects on imaging phenotypes. In contrast, the interaction effect of SNP rs439401 and age on the left hippocampus was small. We need to efficiently distinguish the important interaction effects from others with negligible interaction effects.

We considered the $p \times d$ coefficient matrix $\mathbf{B} = [\boldsymbol{\beta}_1 \dots \boldsymbol{\beta}_d]$ that characterizes the time-dependent and main genetic effects on the longitudinal phenotypes across all subjects and time points. Given the large d and p , the number of elements in \mathbf{B} could be much larger than that of N . Moreover, as shown in Fig. 1a,b, the elements in $\mathbf{y}_i(t) = (y_{i1}(t), \dots, y_{id}(t))^T$ and \mathbf{x}_i were highly correlated because of the spatial pattern of brain structure and the linkage disequilibrium structure of SNPs, respectively. Therefore, the estimate of \mathbf{B} was unstable and often exhibited horizontal- and vertical-banded structures (Fig. 1c). These structures were more apparent after setting unimportant (i.e. insignificant) elements to zero (Fig. 1d).

Applying the characteristics of the coefficient matrix, we introduced a low-rank model for \mathbf{B} as follows:

$$\mathbf{B} = \mathbf{U} \mathbf{\Delta} \mathbf{V}^T = \sum_{l=1}^r \mathbf{B}_l = \sum_{l=1}^r \delta_l \mathbf{u}_l \mathbf{v}_l^T, \quad (2)$$

where r is the rank or number of layers of \mathbf{B} ; $\mathbf{B}_l = \delta_l \mathbf{u}_l \mathbf{v}_l^T$ is the l th layer for $l = 1, \dots, r$; $\mathbf{\Delta} = \text{diag}(\delta_1, \dots, \delta_r)$; $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_r]$ is a $p \times r$ matrix; and $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_r]$ is a $d \times r$ matrix. Moreover, \mathbf{u}_l and \mathbf{v}_l may be viewed as the coefficient of the linear combination of $\mathbf{y}_l(t)$ and \mathbf{x}_l , respectively. Rather than investigating the correlation between all phenotype-SNP pairs, we examined the relation between important linear combinations of similar phenotypes and those of similar SNPs. Because only a few sets of genetic variants were expected to be associated with each longitudinal phenotype, a relatively small $r \ll \min(p, d)$ and a sparse \mathbf{U} were considered sufficient to represent the important structure of \mathbf{B} . Model (2) considerably reduced the number of unknown parameters of \mathbf{B} from pd to $(d + p + 1)r$ and increased the power for detecting the important associations.

We used $\mu_k(t)$ to characterize the overall trajectory for the k th neuroimaging measure without consideration of any effects of genomic variables and prognostic factors. The solid curves in Fig. 2b,c depict the LOESS estimates of the development trajectories of the right and left hippocampal volumes, respectively. We used a penalized spline model (Eilers and Marx, 1996; Ruppert et al., 2003; Lang and Brezger, 2004) to model $\mu_k(t)$ with a polynomial of degree s given by

$$\mu_k(t) = \gamma_{k,0} t^0 + \dots + \gamma_{k,s} t^s + \sum_{m=1}^{q_1-s-1} \gamma_{k,s+m} B_m(t; \mathbf{T}_0) = \mathbf{w}_1(t)^T \boldsymbol{\gamma}_{1,k}, \quad (3)$$

where $B_m(t; \mathbf{T}_0)$ is a basis function of t ; \mathbf{T}_0 is a vector of predetermined knots over the range of the t_{ij} 's (e.g. the sample percentiles); $\boldsymbol{\gamma}_{1,k} = (\gamma_{k,0}, \dots, \gamma_{k,q_1-1})^T$; and $\mathbf{w}_1(t) = (1, t, \dots, t^s, B_1(t; \mathbf{T}_0), \dots, B_{q_1-s-1}(t; \mathbf{T}_0))^T$. Various spline basis functions may be considered for $B_m(t; \mathbf{T}_0)$, such as the B-spline basis (Eilers and Marx, 1996), truncated power basis (Ruppert et al., 2003), and wavelet basis (Morris and Carroll, 2006). Using $\mu_k(t)$ and the basis functions to flexibly capture the overall trajectories of the longitudinal ROIs measurements increases the detection power of the SNPs and SNP-age interactions and improves the prediction accuracy of responses at later time points.

Another issue in longitudinal studies that must be accounted for is the within-subject correlation among repeated measurements of each phenotype (Verbeke and Molenberghs, 2009; Hyun et al., 2016). Various correlation structures can be formulated through $\mathbf{z}_k(t)^T \mathbf{b}_{ik}$ and $\boldsymbol{\Sigma}_b$. We examined the trajectories of all phenotypes across all subjects for the longitudinal data set obtained from the ADNI. The within-subject association mainly occurred from the subject-specific intercept. The trajectories of the left lateral ventricle volume across all subjects are depicted in Fig. 2a. To generate these trajectories, we used a

univariate random effect for each ROI by setting $p_b = 1$, $\mathbf{z}_A(t) = 1$, and $\boldsymbol{\Sigma}_b = \tau_b \mathbf{I}_d$ where \mathbf{I}_d is a $d \times d$ identity matrix. In this case, the random effects of different ROIs were assumed to be independent.

The correlation structure among the multivariate phenotypes was modeled with another component in (1). The covariance matrix of phenotypic measurements across all subjects and all time points is depicted in Fig. 1a. Phenotypic variables usually exhibit group correlation structures; therefore, we used latent factors to capture these correlation structures. Specifically, we considered an exploratory method of using the sparse factor model (Bhattacharya and Dunson, 2011) for $\boldsymbol{\varepsilon}_i(t)$ as follows:

$$\boldsymbol{\varepsilon}_i(t) = \boldsymbol{\Lambda} \boldsymbol{\eta}_i(t) + \boldsymbol{\xi}_i(t), \quad (4)$$

where $\boldsymbol{\Lambda}$ is a $d \times \infty$ factor loading matrix, $\boldsymbol{\eta}_i(t) \sim N_{\infty}(\mathbf{0}, \mathbf{I}_{\infty})$, and $\boldsymbol{\xi}_i(t) \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\xi})$ with $\boldsymbol{\Sigma}_{\xi} = \text{diag}(\sigma_{\xi,1}^2, \dots, \sigma_{\xi,d}^2)$. The structure of $\boldsymbol{\Lambda}$ and the number of factors were directly gleaned from the data.

The L2R2 in (1) with (2) – (4) incorporated can be rewritten in the matrix form

$$\mathbf{Y} = \mathbf{X} \mathbf{U} \boldsymbol{\Delta} \mathbf{V}^T + \mathbf{W} \boldsymbol{\Gamma} + \mathbf{Z} \mathbf{b} + \mathbf{E}. \quad (5)$$

The notational definitions, prior distributions of the parameters and the technical details of the posterior computation were provided in the appendix.

3. Simulation study

3.1. Simulation setup

We used simulations to examine the finite-sample performance of the L2R2 model. We simulated \mathbf{Y} according to model (5). We considered four cases with various dimensions and priors: (i) $p = 50$, $d = 50$; (ii) $p = 100$, $d = 100$; (iii) $p = 200$, $d = 100$; and (iv) $p = 400$, $d = 100$. The dimensions of these cases were comparable with the ADNI data set used in this study and with other related neuroimaging and genetic studies (Wang et al., 2011; Zhu et al., 2014; Marttinen et al., 2014). We chose the first $n = 100$ subjects from the ADNI data set and selected the top $p/2$ SNPs from the top 40 genes reported by AlzGene database (www.alzgene.org) as of June 10, 2010. Each subject was observed m_i times to replicate the ADNI data set, resulting in $N = 422$ records. We used the $p/2$ SNPs and their interactions with age to form the \mathbf{X} in model (5). The \mathbf{W} in (5) contains $q = 15$ columns, including the intercept, the standardized intracranial volume, sex, education, handedness, and the basis function of age. The basis function consists of age, age², age³, and the B-spline basis given the j th percentile, $j = 0, 10, 20, \dots, 100$. The elements of \mathbf{b} were independently generated from a $N(0, 1)$ distribution.

The following parameters in (5) were used in the simulation: The low-rank coefficient matrix \mathbf{B} was generated with the true rank $r_0 = 3$ in a moderately sparse case with 65% zero elements and in an extremely sparse case with 95% zero elements. Specifically, we set $\mathbf{B} = \mathbf{U} \mathbf{V}$, where $\mathbf{U} = (u_{jl})$, $\mathbf{\Sigma}_U = \text{diag}(\delta_{jj}) = \text{diag}(100, 80, 60)$, and $\mathbf{V} = (v_{jk})$ as $p \times 3$, 3×3 , and $3 \times d$ matrices, respectively. Moreover, we generated all elements u_{jl} and v_{jk} independently from a $\mathcal{N}(0, 1) \times \text{binomial}(1, p_1)$ distribution and then normalized the columns of \mathbf{U} and \mathbf{V} to have zero mean and unit variance. The value of p_1 was tuned so that approximately 65% of the elements of \mathbf{B} were zeros. Each element of $\mathbf{\Gamma}$ was independently generated as a $\gamma_{jk} \sim \mathcal{N}(0, 1) \times \text{binomial}(1, 0.5)$. We simulated $\mathbf{e}_e(t) \sim \mathcal{N}_d(0, \mathbf{\Sigma}_e)$, where the precision matrix $\mathbf{\Sigma}_e$ was first generated by a $d \times d$ matrix $\mathbf{A} = (a_{jj'})$ with $a_{jj} = 1$ and $a_{jj'} = \text{uniform}(0, 1) \times \text{binomial}(1, p_2)$ for $j \neq j'$, setting $\mathbf{\Sigma}_e = \mathbf{A} \mathbf{A}^T$, and standardizing $\mathbf{\Sigma}_e$ into a correlation matrix. The value of p_2 was tuned so that approximately 20% of the elements in $\mathbf{\Sigma}_e$ were zeros, yielding the mean of the absolute correlations of $\mathbf{\Sigma}_e$ at approximately 0.40.

For each case, 100 simulated data sets were generated. For each simulated data set, we ran the Gibbs sampler for 10,000 iterations after 5000 burn-in iterations. We chose noninformative priors for the hyperparameters and set $a_0 = b_0 = c_0 = d_0 = e_0 = f_0 = 10^{-6}$. For the covariance parameter $\mathbf{\Sigma}_e$, we chose moderately informative priors to impose the positive-definite constraint.

Besides modeling the trajectories of the response variables to improve detection power of the genetic variables, the splines in \mathbf{W} could also be used to predict the response variables at later time points. We calculated the mean of the response variables $\mathbf{X}_{pred} \mathbf{U} \mathbf{V}^T + \mathbf{W}_{pred} \mathbf{\Gamma} + \mathbf{Z}_{pred} \mathbf{b}$ for every the subjects at one year after their last observation, where \mathbf{X}_{pred} , \mathbf{W}_{pred} , and \mathbf{Z}_{pred} contain the corresponding predictors for these subjects at the later time points. The prediction was based on $\mathbf{X}_{pred} \hat{\mathbf{U}} \hat{\mathbf{V}}^T + \mathbf{W}_{pred} \hat{\mathbf{\Gamma}} + \mathbf{Z}_{pred} \hat{\mathbf{b}}$, where the hat represents the posterior means of the parameters and random effects. We compared the means and predicted means to check the prediction accuracy using the splines in \mathbf{W} .

3.2. Comparison with four other state-of-the-art methods

We compared the L2R2 model with four state-of-the-art methods that were developed to establish the association between high-dimensional responses and predictors. First, we considered MMA analysis between the high-dimensional response variables and predictors. Each phenotype was regressed on the covariate matrix \mathbf{W} and the residue was associated with each SNP in \mathbf{X} with regression. We used the lme4 R package (Bates et al., 2015) for MMA.

Second, we considered the group-sparse multitask regression and feature-selection (G-SMuRFS, Wang et al., 2011) method, which is a representation of associating high dimensional imaging and genetic variables with regularization methods. The longitudinal volumes of ROIs were used as the response variables, and the \mathbf{X} and \mathbf{W} in L2R2 were combined as the matrix of predictors in G-SMuRFS. The spline basis in \mathbf{W} modeled the overall longitudinal trajectories of the ROI volumes, and the SNP-age interactions in \mathbf{X} represented the SNP effects on the longitudinal trajectories of ROIs. The large matrix of coefficients was estimated with penalty functions and optimization. Each covariate, SNP and

SNP-age interaction formed a group in the penalty function. We used $\exp(-7)$, $\exp(-6)$, \dots , $\exp(2)$ as possible values of the tuning parameters in the G-SMuRFS method and performed a five-fold cross validation to choose the tuning parameters. G-SMuRFS did not make any assumption on the distribution and correlation structure of the response variables. In our longitudinal study, G-SMuRFS method did not explicitly account for the within-subject temporal correlations and spatial correlations. Also, G-SMuRFS did not use dimension reduction for the coefficient matrix like the low-rank representation of \mathbf{B} in L2R2.

Third, we considered the Bayesian canonical correlation analysis (BCCA Klami et al., 2013) to obtain linear combinations of the response variables \mathbf{Y} and the predictor \mathbf{X} . The coefficients of the linear combinations were conceptually similar to \mathbf{U} and \mathbf{V} in (5), and we denoted their estimates by $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$, respectively. The coefficient matrix \mathbf{B} was estimated by $\tilde{\mathbf{U}} \tilde{\mathbf{V}}^T$. Gamma mixture multivariate normal distributions were assigned to \mathbf{U} and \mathbf{V} , and the correlation structure of the response variables and the SNPs were characterized by a factor model. We used the default setting of the companion R package and used eight components in the BCCA method. The R package cannot run a continuous predictor; therefore, we excluded the SNP-age interactions in \mathbf{X} . Also, the package does not process covariates like \mathbf{W} . Therefore, we regressed \mathbf{Y} on \mathbf{W} and applied the BCCA method to the residue and \mathbf{X} . Unlike the L2R2 model, the BCCA method does not account for within-subject structural correlation or the interactions between SNPs and age.

Fourth, we considered the Bayesian reduced-rank regression (BRRR) method developed by Marttinen et al. (2014). Because the BRRR software package does not allow continuous predictors, we excluded the SNP-age interaction terms in \mathbf{X} . We also set the rank to three and ran all of the other program parameters at their default settings. The BRRR method can process \mathbf{W} directly, so we did not need to regress \mathbf{Y} on \mathbf{W} beforehand. In contrast with the L2R2 model, the BRRR method does not account for SNP-age interactions or within-subject structural correlation.

3.3. Evaluation

To evaluate the finite-sample classification performances of all five of the methods, we calculated the sensitivity and specificity scores of selecting the nonzero elements in the coefficient matrix \mathbf{B} for each method. The following criteria were used for the selection of the coefficients in each method. As the G-SMuRFS method does not provide standard error estimate for each coefficient, the importance of the association between each pair of ROI and SNP/SNP-age interaction was determined by the absolute values of the coefficient estimates. For all other methods, the coefficients were selected if the absolute values of their standardized estimates (coefficient estimate over its standard error commonly used to account for the estimation uncertainty) were larger than a threshold. The sensitivity and specificity scores for a given threshold T_0 were defined as

$$\text{Se}(T_0) = \frac{\text{TP}(T_0)}{\text{TP}(T_0) + \text{FN}(T_0)} \text{ and } \text{Sp}(T_0) = \frac{\text{TN}(T_0)}{\text{TN}(T_0) + \text{FP}(T_0)},$$

where $TP(T_0)$, $FN(T_0)$, $TN(T_0)$, and $FP(T_0)$ were the numbers of true positives, false negatives, true negatives, and false positives, respectively. Changing the threshold T_0 produced different sensitivity and specificity scores, generating the receiver operating characteristic (ROC) curves. In each ROC curve, the sensitivity was plotted against one minus specificity.

3.4. Results

The ROC curves of all four methods and the L2R2 model are depicted in Fig. 3. ROC curves that are closer to the upper-left corner are more likely to identify true positives and control for false positives.

In all simulation settings, the L2R2 model outperformed the four competing methods because the L2R2 model was designed to specifically address the challenges described above. Among the four previously developed methods, the G-SMuRFS method had the best performance. However, the G-SMuRFS method did not characterize the correlation structure among the response variables like the L2R2 model did. Therefore, the G-SMuRFS method potentially increased the number of false positives. Consequently, when the number of phenotypic variables increased, the difference between the L2R2 model and the G-SMuRFS method for $d = 100$ was larger in terms of main effects and interactions than that for $d = 50$.

The BRRR and BCCA methods cannot process SNP-age interactions because of the limitations of their companion R packages. Therefore, they increase modeling error and reduce the power of detecting important main genetic effects in the presence of the SNP-age interactions. These limitations are worsened as the number of SNPs increases. Although the model setup of the BRRR method was similar to that of the L2R2 model, the prior setting for the BRRR method was more subtle and appeared to be more suitable for small numbers of responses and predictors. Thus, the BRRR method is very sensitive to the number of responses and predictors. The MMA method performed the worst, as it did not account for the correlation structure of responses and predictors or the spatio-temporal correlation structure of the longitudinal phenotypes.

Selected spline functions in the true coefficients (Γ) and their corresponding estimates obtained from the L2R2 model and the G-SMuRFS method are shown in Fig. 4a–b. It is apparent from these splines that the bias of the L2R2 model was smaller than that of the G-SMuRFS method. The estimated trajectories with smaller bias enabled us to better predict the response variables of later time points. A scatter plot of the true means at later time points versus the predicted means across all subjects and response variables in a randomly selected data set is shown in Fig. 4c. The true and predicted means were close to each other, indicating that model (5) can accurately predict longitudinal responses.

We examined the Markov chain Monte Carlo (MCMC) convergence through calculation of Gelman and Rubin's shrink factors for all the parameter estimates obtained from three MCMC chains with different start values of parameters (Fig. 4d). The MCMC chains appeared to converge after 2000 iterations, as the shrink factors for all the parameters were less than 1.2. Generating 4000 MCMC samples after discarding 4000 burn-in samples for the L2R2 model to fit the four cases (i.e. $p=50, d=50$; $p=100, d=100$; $p=200, d=100$; and

$p=400$, $d=100$) took approximately 6, 14, 16, and 22 min, respectively, with a single core of an Intel Xeon E5520 CPU. In comparison, the G-SMuRFS method took 8, 18, 30, and 90 s to fit the four cases, respectively. However, the MCMC samples of the L2R2 model allowed us to make statistical inferences on all parameter estimates, whereas this could not be performed with the G-SMuRFS method. It is important to note that statistical inferences made under a high-dimensional setting require some resampling and splitting procedures (Meinshausen and Bühlmann, 2010). In this case, the G-SMuRFS method would require processing of thousands of subsamples, which may be computationally demanding.

4. Real data analysis

4.1. ADNI data description

The ADNI was launched in 2003 as a 5-year public-private partnership by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations. The primary goal of the ADNI is to test whether serial magnetic resonance imaging (MRI), positron emission tomography, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment and early AD. Determination of sensitive and specific markers of very early AD progression may aid researchers and clinicians in developing new treatments and monitoring their effectiveness, as well as reduce the time and cost of clinical trials.

4.2. Preprocessing

The MRI data were collected across a variety of MRI scanners with individualized protocols for each scanner to obtain standard T1-weighted images, volumetric 3-dimensional sagittal magnetization prepared gradient-echo sequences, or equivalent images with various resolutions. The typical imaging protocol included inversion time=1000 ms, repetition time=2400 ms, flip angle=8°, and field of view=24 cm with a $256 \times 256 \times 170$ acquisition matrix in the x -, y -, and z -dimensions, yielding a voxel size of $1.25 \times 1.26 \times 1.2 \text{ mm}^3$. Standard steps such as anterior commissure and posterior commissure correction, skull stripping, cerebellum removing, intensity inhomogeneity correction, segmentation, and registration (Shen and Davatzikos, 2004) were used to preprocess the MRI data. We then carried out automatic regional labeling for the template and transferring the labels after the deformable registration of subject images. After labeling 93 ROIs, we computed the ROI volumes for each subject.

The Human 610-Quad BeadChip (Illumina, Inc., San Diego, CA) was used to genotype the subjects whose images in the ADNI database we analyzed, resulting in a set of 620,901 SNP and copy number variation markers. Because the Apolipoprotein E (APOE) SNPs rs429358 and rs7412 are not included in the Human 610-Quad Bead-Chip, they were genotyped separately. These two SNPs define a three allele haplotype, namely the $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$ variants, and the presence of each variant was available in the ADNI database for all subjects. We used EIGENSTRAT software (package 3.0) to calculate the population stratification coefficients of all subjects. To reduce population stratification effects, we used

images from only white subjects who had at least one imaging sample available (749 of 818 subjects in the ADNI data set).

We also performed quality control on this initial set of genotypes (Wang et al., 2011). To input the missing genotypes into our analysis, we used MACH4 software (version 1.0.16) with default parameters to infer the haplotype phase. We also included the APOE- ϵ_4 variant, coded as the number of observed ϵ_4 variants. We removed SNPs with more than 5% missing values and entered the mode for the missing SNPs. In the final quality-controlled genotype data, we removed SNPs with a minor allele frequency smaller than 0.1 and a Hardy-Weinberg p -value $< 10^{-6}$.

4.3. Data analysis

The aim of our analysis was to apply the L2R2 model to establish an association between the SNPs in the top AD-risk genes reported by AlzGene (<http://www.alzgene.org/>) and the volumes of 93 brain ROIs collected by the ADNI. We used the L2R2 model to carry out formal statistical inferences, such as the identification of significant SNPs and SNP-age interactions. In this data analysis, we included data from $n = 749$ subjects and $N = 2817$ MRI measurements, resulting in an unbalanced data set. Among the subjects, 41 had only one observation and another 67 had only two observations. A random effect was used to account for the subject-specific intercept. For the age effect, we used a quadratic penalized spline with 11 knots that were based on the percentiles of standardized age. We also included intracranial volume, sex, education, and handedness as covariates in \mathbf{W} to adjust the ROI volumes.

We considered two sets of top AD-risk genes. First, we used the 114 SNPs in the top 10 genes reported by the AlzGene database. After we performed quality control to the data set, 87 SNPs, APOE- ϵ_4 , and their interaction with age were included in the model (1). Second, we selected the 1224 SNPs in the top 40 AD-risk genes by the AlzGene database. After applying quality control to the data, we included 1072 SNPs and their interactions with age in our analysis. A map of the linkage disequilibrium among the 1072 SNPs revealed a clear clustering pattern of SNPs (Fig. 1b). Specifically, the correlations among SNPs within each gene were large (> 0.7), whereas those among SNPs in different genes were relatively small.

We fitted the L2R2 model (1) to the ADNI data as follows: To determine the rank of \mathbf{B} , the L2R2 model was run up to $r = 10$ layers. By comparing the five different selection criteria in Zhu et al. (2014), we chose $r = 3$ layers as the optimal rank for the final data analysis. We ran the Gibbs sampler for 20,000 iterations after 20,000 burn-in iterations. For the G-SMuRFS method, we used the same \mathbf{Y} matrix, combined the \mathbf{W} and \mathbf{X} matrices into a single predictor matrix, and then used the five-fold cross validation to choose the optimal penalty.

On the basis of the MCMC samples, we calculated the posterior median and median absolute deviations of \mathbf{U} , \mathbf{V} , and \mathbf{B} , and used $1.4826 \times$ median absolute deviations to compute the robust standard errors for each element of \mathbf{B} . To determine the important coefficients, we used the standardized estimates \mathbf{B}_{std} , which were the absolute values of the estimates over their standard errors. To validate the low-rank structure of \mathbf{B} , we used the results of the top 10 genes and set the elements of \mathbf{B} to zero if their absolute standardized

estimates were larger than a threshold (Fig. 5). The sparsely distributed points along the horizontal and vertical directions indicated that the low-rank model was suitable for characterizing the association between the neuroimaging and genetic variables in the ADNI data set. We used a pragmatic approach to calculate the threshold for the Bayesian estimates, which resembled the Bonferroni correction in a frequentist approach. The threshold was the $F(1-0.025/(pr/2+dr+r))$, and F was the quantile function of standard normal distribution, where the denominator was the total number of parameters in the low-rank structure \mathbf{U}_{int} , and \mathbf{V} . \mathbf{U}_{int} was the matrix of coefficients of interaction terms. The thresholds for the top 10 genes and top 40 genes were 3.912 and 4.339, respectively. We referred the SNPs that passed the thresholds as significant SNPs though the meaning is different from frequentist significance.

4.4. Results

We determined the longitudinal trajectories of the neuroimaging ROI volumes, the effects of the significant SNPs on such longitudinal trajectories, and the dynamic genetic effects on the trajectories. We estimated the trajectories of all standardized ROI volumes on the basis of $\hat{\Gamma}$ and \mathbf{W} (Fig. 6). ROIs with decreasing volumes occurred in many regions of the brain, including the left and middle temporal gyri, the left and superior temporal gyri, and the left and right amygdala. ROIs with increasing volumes occurred in the hollow areas of the brain and areas of white matter and included the left and right lateral ventricles, the left and right frontal lobe white matter, and the left and right temporal lobes. Moreover, the trajectories of most ROIs demonstrated structural symmetry.

Because hippocampal atrophy and ventricular enlargement are consistent findings in patients with AD and mild cognitive impairment (Apostolova et al., 2012), we depicted the effects of some SNPs on the longitudinal trajectories of the left and right lateral ventricles and the left and right hippocampi (Fig. 7). Let A and a denote the major and minor alleles of a specific SNP, respectively. For each ROI, we depicted the mean trajectory associated with a selected SNP (black solid curves) and the mean trajectory corresponding to the three allelic combinations (i.e. AA , Aa , and aa) of the SNP (Fig. 7). The minor allele of SNP rs10501608 increased the rate of hippocampal atrophy, whereas the SNP rs1354106 minor allele had the opposite effect. Moreover, the minor allele of SNP rs10501608 increased the rate of ventricular enlargement, yet the SNP rs10501604 minor allele had the opposite effect.

We plotted the primary SNP-age interaction effects, whose absolute standardized estimates were larger than the predetermined threshold, on the longitudinal trajectories corresponding to the top 10 genes (upper row) and the top 40 genes (lower row) (Fig. 8). The plots revealed a generally symmetric pattern for the left and right hemispheres, which indicated the genetic effects on the longitudinal change of ROI volumes were symmetric. Some asymmetric longitudinal genetic effects also occurred. For example, in the analysis of the SNPs from the top 40 genes, only the left nucleus accumbens and the right subthalamic nucleus were associated with SNP rs2273684, whereas their contralateral counterparts were not substantially correlated with any SNP. With exception of the left and right lateral ventricles, the genetic effects on most ROIs in the left and right hemispheres occurred in the same direction. Because both the number of parameters and the threshold for significance

increased with the number of top genes, the number of significant SNP-age interactions decreased with the number of top genes. However, the effects of significant SNP-age interaction detected for the top 40 genes were more substantial.

The significant SNP-age interaction effects of the ROIs depicted in Fig. 8 formed SNP-specific networks that we visualized with BrainNet Viewer (Xia et al., 2013) in Fig. 9 and Fig. 10. Among all the networks, each SNP-specific network demonstrated spatially adjacent patterns. Some SNPs were associated with many ROIs, although other SNPs were associated only with ROIs in a specific brain area. The networks could be considerably different for different SNPs in a gene. In each network the sizes of genetic effects on various ROIs was relatively homogeneous, whereas the those across different networks may be substantially different. The most significant SNPs were close to the ventricular system. Some ROIs in the frontal lobe were also associated with SNPs that exhibited small growth effects, but ROIs near the prefrontal cortex were not strongly associated with the SNPs in the top 10 and top 40 genes.

We used MMA based on linear mixed effects models where each SNP, its interaction with age, and the covariates in \mathbf{W} were regressed on each ROI volume. The p-values of the significant coefficients are smaller than $0.05/(pr/2 + dr + r)$ (Bonferroni correction with the number of parameter in the L2R2), even though the true number of regression parameter is much larger (pd). The number of significant coefficients is smaller if the true number of parameters is used for correction. We compared the important estimated associations with L2R2. The two ROIs identified in the case with top 10 genes were the left and right lateral ventricles. Twelve ROIs were found in the case with top 40 genes, ten of which were found by L2R2. The number of significant associations identified by MMA were much less than that of L2R2. L2R2 provided more power in identifying important pairs of ROIs and SNP-age interactions and revealed more SNPs that may alter the longitudinal trajectories of ROIs.

The results of G-SMuRFS were not directly comparable to those of L2R2 and MMA because G-SMuRFS does not provide standard error estimates or a cut-off for all pairs of association as in L2R2 and MMA. The importance of the SNPs and SNP-age interactions was ranked by the sum of the absolute values of the estimated coefficients across all ROIs. Nonetheless, we found the top ROIs and SNP-age interactions, and some of them overlap with those from L2R2 in the case with top 40 genes.

We built the networks of ROIs with two criteria to study the structure of the longitudinal volumes of ROIs. Let \mathbf{B}_{int} be the part of \mathbf{B} corresponding to the SNP-age interactions, and $\mathbf{B}_{int,std}$ be the standardized \mathbf{B}_{int} . Firstly, we selected the ROIs corresponding to the large diagonal elements of $\mathbf{B}_{int,std}^T \mathbf{B}_{int,std}$ to form a network. The change of each ROI in the network was heavily associated with the SNPs. The network was a collection of important ROIs which were affected by the time-dependent SNP effects. Secondly, we constructed networks based on the absolute values of every columns of the standardized \mathbf{V}_{std} which is the standardized \mathbf{V} . Each network was an important combination of the longitudinal ROI measurements, and was strongly associated with the joint effect of a weighted combination of SNPs and age-SNP interactions quantified by the corresponding column of \mathbf{U} . The top 10 ROIs in the networks were listed in Table 1 and Table 2 for the top 10 and 40 genes,

respectively. Such networks were mostly symmetric for the left and right hemispheres, which also illustrated the genetic effects on the longitudinal volumes of ROIs are largely symmetric.

The imaging phenotypes provided quantitative measurements of the morphometric changes of the ROIs, which may characterize the underlying degenerative process of AD and capture information mirroring patients' diagnostic status. We used a generalized linear-mixed model with logit link to determine the associations between longitudinal diagnostic status and the ROI volume. We found that 70 of 93 ROIs and 69 ROI and age interactions were significantly correlated with longitudinal diagnostic status after applying the false discovery rate correction with the significant level 0.05. Tissues with ROIs that correlated with longitudinal diagnostic status included the hippocampus, cingulate cortex, amygdala, and lateral ventricles, in addition to many other areas of the brain. Moreover, most of the ROIs identified by the L2R2 model in Figure 8 were among the 69 ROIs that demonstrated age-dependent interactions. The SNPs used in the L2R2 model were previously reported in a genome-wide association study of the diagnostic status of patients with AD. Our findings show that the longitudinal trajectories of many ROIs are highly associated with the longitudinal patterns of diagnostic status; therefore, these ROIs may serve as consistent surrogates for measuring AD progression.

The L2R2 model elucidated some ROI-SNP pairs that were reported in previous genome-wide association studies on the basis of diagnostic status, cognitive scores, and imaging measurements. For example, the SNP rs1354106, which occurs in the *CD33* gene, was associated with a slower declining rate of the AD assessment scale cognitive score (Sherva et al., 2014). Biffi et al. (2010) reported that SNP rs1408077, which occurs in the *CR1* gene, was associated with disease status. Moreover, the SNP rs1408077 was related to the entorhinal cortex thickness (Biffi et al., 2010) and hippocampal atrophy (Lazaris et al., 2015). The SNP rs6088662, which occurs in the *PRNP* gene, was associated with reduced hippocampal volume and cognitive performance in healthy individuals (Li et al., 2016).

Multivariate imaging measurements may provide additional information compared with univariate measurements, such as diagnostic status. Some ROIs that were not associated with diagnostic status in the generalized linear mixed model were associated with certain SNPs in the L2R2 model, such as the left and right thalamus, right medial frontal gyrus, and right lateral front-orbital gyrus. More ROI measurements yielded additional variation in the SNPs compared with that of diagnostic status. Thus, integrating multivariate imaging measurements with genetic variables and diagnostic status may provide a clearer picture of the biological processes occurring in AD (Zhao and Castellanos, 2016).

4.5. Random effects

We examined the usefulness of incorporating random effects into our model by investigating the distribution of the estimated random effects over all subjects for each ROI. We generated histograms and quantile-quantile plots of the random effects obtained from two randomly selected ROIs from the L2R2 model that were fitted to the top 10 genes (Fig. 11). The variances of the estimated random intercepts were substantial compared with the scale of the response variables, which coincided with the subject-specific intercepts of the neuroimaging

measurements (Fig. 2a). Moreover, these distributions did not severely deviate from the normal distribution.

To statistically test the presence of the subject-specific random effects \mathbf{Zb} , we refitted the data with a restricted L2R2 model without the presence of \mathbf{Zb} and then compared the log likelihood ratio of the full and restricted L2R2 models. The inclusion of random effects decreased the log likelihood function by 14,000 (or 5%), and the number of parameters increased by 93. Therefore, incorporating random effects into the L2R2 model was useful for characterizing the correlation structure of the imaging phenotypes and improved the detection of the association between imaging phenotypes and genetic variants.

5. Discussion

We have developed a Bayesian L2R2 model to determine the association between longitudinal imaging responses and covariates with applications in imaging genetic data. We used this model to approximate the large association matrix. We combined a sparse latent factor model and random effects to flexibly capture the complex spatiotemporal correlation structure. We incorporated splines to capture the effect of aging and combined a traditional coefficient estimation with a low-rank approach. The L2R2 model dramatically reduced the number of parameters to be sampled and tested, leading to a remarkably faster sampling scheme and efficient inference. The simulation studies demonstrated that the L2R2 model has higher power in detecting important time-dependent and main genetic effects. Age-dependent genetic effects are useful for characterizing age-related degenerative disorders like AD. When strong genetic effects modify changes in brain morphology over time, it is critically important to account for SNP-age interaction effects and to improve the detection of time-invariant main genetic effects. Although our findings confirmed the important role of well-known genes, such as APOE- $\epsilon 4$, in the pathology of AD, they also elucidated other potential candidate genes that warrant further investigation.

The L2R2 model effectively established associations between phenotypic markers present in neuroimaging data and the presence of high-risk AD alleles. However, many considerations still merit further research. First, it is important to consider the joint effect of genetic markers and environmental factors on high-dimensional imaging phenotypes (Thomas, 2010). Second, it is of great interest to incorporate rare variant genetic markers (Bansal et al., 2010) with the L2R2 model. Third, the key features of the L2R2 model may be adapted to more complex data structures (e.g. twin and family sequencing studies) and other parametric and semiparametric models. Fourth, the L2R2 model may be extended to combine different imaging phenotypes calculated from other imaging modalities (e.g. diffusion tensor imaging, functional MRI, and electroencephalography) in concomitant imaging and genetic studies. Fifth, group structures among SNPs (i.e. gene-based grouping) could be incorporated to improve the efficiency of modeling-correlated SNPs by choosing group priors for \mathbf{U} .

One important feature of L2R2 is the inclusion of SNP-age interactions to model the genetic effects on the longitudinal change of ROIs. The interactions can also be viewed as the age-dependent genetic effects. The total age-varying genetic effects of the j th SNP x_{ij} is $\beta_j +$

β_{2jij} , which assumes that the effect is a linear function of age. This assumption can be relaxed by modeling the effects as more flexible functions of age, e.g., polynomial functions and nonparametric functions, which is known as the varying-coefficient models (Hastie and Tibshirani, 1993; Fan and Zhang, 1999, 2008). Varying-coefficient models have been used to study the time-dependent genetic effects and gene-environment interactions (Gong and Zou, 2012; Wu and Cui, 2013; Li et al., 2015). Compared to L2R2, these studies only modeled univariate response variables. Extending L2R2 to incorporate more flexible varying genetic effects is helpful to better understand the genetic impact on the complex diseases at different age and under various states of environment factors.

It is worth noting that in Fig. 2b, the volume of right hippocampal formation appeared to increase in the three-year visit for subjects with minor allele of rs768451. This phenomenon is less expected in studies of Alzheimer's disease but it may be explained by the following reasons. We included healthy subjects, MCI and AD patients from the ADNI data set in our analysis. The volumes for the healthy subjects may be relatively stable. Moreover, several studies in the literature (Erickson et al., 2011; Leavitt et al., 2014; Niemann et al., 2014) reported that exercises were associated with the increase of hippocampal volume in older adults. Also, the number of subjects in the three-year visit was smaller than those in the previous visits, and the average of the hippocampal volumes for these subjects was larger than the average for all subjects at the initial visit. The dropout during the last visit may be informative. These factors warrant further investigation in the future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This material was based upon work partially supported by the NSF grant DMS-1127914 to the Statistical and Applied Mathematical Science Institute. The research of Dr. Zhu was partially supported by NSF grants SES-1357666 and DMS-1407655, NIH grant MH086633, and a grant from Cancer Prevention Research Institute of Texas.

References

- Apostolova LG, Green AE, Babakchianian S, Hwang KS, Chou YY, Toga AW, Thompson PM. Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment (mci), and alzheimer disease. *Alzheimer Disease & Associated Disorders*. 2012; 26:17–27. [PubMed: 22343374]
- Bansal V, Libiger O, Torkamani A, Schork NJ. Statistical analysis strategies for association studies involving rare variants. *Nature Reviews Genetics*. 2010; 11:773–785.
- Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. 2015; 67:1–48. DOI: 10.18637/jss.v067.i01
- Bernal-Rusiel J, Greve D, Reuter M, Fischl B, Sabuncu MR. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. *NeuroImage*. 2013; 66:249–260. [PubMed: 23123680]
- Bhattacharya A, Dunson DB. Sparse bayesian infinite factor models. *Biometrika*. 2011; 98:291–306. [PubMed: 23049129]

- Biffi A, Anderson CD, Desikan RS, Sabuncu M, Cortellini L, Schmansky N, Salat D, Rosand J. Genetic variation and neuroimaging measures in Alzheimer Disease. *Archives of neurology*. 2010; 67:677–685. DOI: 10.1001/archneurol.2010.108 [PubMed: 20558387]
- Breiman L, Friedman J. Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society*. 1997; 59:3–54.
- Chen H, Wang Y. A penalized spline approach to functional mixed effects models analysis. *Biometrics*. 2011; 67:861–870. [PubMed: 21155747]
- Chiang MC, Barysheva M, Toga AW, Medland SE, Hansell NK, James MR, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM. Bdnf gene effects on brain circuitry replicated in 455 twins. *NeuroImage*. 2011a; 55:448–454. [PubMed: 21195196]
- Chiang MC, McMahon KL, de Zubicaray GI, Martin NG, Hickie I, Toga AW, Wright MJ, Thompson PM. Genetics of white matter development: A dti study of 705 twins and their siblings aged 12 to 29. *NeuroImage*. 2011b; 54:2308–2317. [PubMed: 20950689]
- Cleveland WS, Devlin SJ. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*. 1988; 83:596–610. DOI: 10.1080/01621459.1988.10478639
- Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science*. 1996; 11:89–121. DOI: 10.1214/ss/1038425655
- Erickson KI, Voss MW, Prakash RS, Basak C, Szabo A, Chaddock L, Kim JS, Heo S, Alves H, White SM, Wojcicki TR, Mailey E, Vieira VJ, Martin SA, Pence BD, Woods JA, McAuley E, Kramer AF. Exercise training increases size of hippocampus and improves memory. *Proceedings of the National Academy of Sciences*. 2011; 108:3017–3022. DOI: 10.1073/pnas.1015950108
- Fan J, Zhang W. Statistical estimation in varying coefficient models. *Ann Statist*. 1999; 27:1491–1518. DOI: 10.1214/aos/1017939139
- Fan J, Zhang W. Statistical methods with varying coefficient models. *Stat Interface*. 2008; 1:179–195. [PubMed: 18978950]
- Gong Y, Zou F. Varying coefficient models for mapping quantitative trait loci using recombinant inbred intercrosses. *Genetics*. 2012; 190:475–486. DOI: 10.1534/genetics.111.132522 [PubMed: 22345613]
- Greven S, Crainiceanu C, Caffo B, Reich D. Longitudinal functional principal component analysis. *Electron J Statist*. 2010; 4:1022–1054.
- Guillaume B, Hua X, Thompson PM, Waldorp L, Nichols TE. Fast and accurate modelling of longitudinal and repeated measures neuroimaging data. *NeuroImage*. 2014; 94:287–302. [PubMed: 24650594]
- Guo W. Functional mixed effects models. *Biometrics*. 2002; 58:121–128. [PubMed: 11890306]
- Hastie TJ, Tibshirani RJ. Varying-coefficient models. *J Roy Statist Soc B*. 1993; 55:757–796.
- Hibar D, et al. Voxelwise gene-wide association study (vgenewas): Multivariate gene-based association testing in 731 elderly subjects. *Neuroimage*. 2011; 56:1875–1891. [PubMed: 21497199]
- Hyun JW, Li Y, Huang C, Styner M, Lin W, Zhu H, et al. Alzheimer’s Disease Neuroimaging Initiative. Stgp: Spatio-temporal gaussian process models for longitudinal neuroimaging data. *NeuroImage*. 2016; 134:550–562. [PubMed: 27103140]
- Khondker Z, Zhu H, Chu H, Lin W, Ibrahim J. The bayesian covariance lasso. *Statistics and its Interface*. 2013; 6:243–259. [PubMed: 24551316]
- Klami A, Virtanen S, Kaski S. Bayesian canonical correlation analysis. *The Journal of Machine Learning Research*. 2013; 14:965–1003.
- Lang S, Brezger A. Bayesian P-Splines. *Journal of Computational and Graphical Statistics*. 2004; 13:183–212.
- Lazaris A, Hwang KS, Goukasian N, Ramirez LM, Eastman J, Blanken AE, Teng E, Gyls K, Cole G, Saykin AJ, Shaw LM, Trojanowski JQ, Jagust WJ, Weiner MW, Apostolova LG. Alzheimer risk genes modulate the relationship between plasma apoE and cortical PiB binding. *Neurology: Genetics*. 2015; 1doi: 10.1212/NXG.0000000000000022
- Leavitt VM, Ciriugliaro C, Cohen A, Farag A, Brooks M, Wecht JM, Wylie GR, Chiaravalloti ND, DeLuca J, Sumowski JF. Aerobic exercise increases hippocampal volume and improves memory in

multiple sclerosis: Preliminary findings. *Neurocase*. 2014; 20:695–697. DOI: 10.1080/13554794.2013.841951 [PubMed: 24090098]

- Li J, Wang Z, Li R, Wu R. Bayesian group lasso for nonparametric varying-coefficient models with application to functional genome-wide association studies. *Ann Appl Stat*. 2015; 9:640–664. <http://dx.doi.org/10.1214/15-AOAS808>. DOI: 10.1214/15-AOAS808 [PubMed: 26478762]
- Li M, Luo Xj, Landén M, Bergen SE, Hultman CM, Li X, Zhang W, Yao YG, Zhang C, Liu J, Mattheisen M, Cichon S, Mühleisen TW, Degenhardt FA, Nöthen MM, Schulze TG, Grigoriou-Serbanescu M, Li H, Fuller CK, Chen C, Dong Q, Chen C, Jamain S, Leboyer M, Bellivier F, Etain B, Kahn JP, Henry C, Preisig M, Katalik Z, Castelao E, Wright A, Mitchell PB, Fullerton JM, Schofield PR, Montgomery GW, Medland SE, Gordon SD, Martin NG, Group TSBS, Rietschel M, Liu C, Kleinman JE, Hyde TM, Weinberger DR, Su B. Consortium M. Impact of a cis-associated gene expression SNP on chromosome 20q11.22 on bipolar disorder susceptibility, hippocampal structure and cognitive performance. *The British Journal of Psychiatry*. 2016; 208:128–137. DOI: 10.1192/bjp.bp.114.156976 [PubMed: 26338991]
- Li Y, Gilmore JH, Shen D, Styner M, Lin W, Zhu H. Multiscale adaptive generalized estimating equations for longitudinal neuroimaging data. *NeuroImage*. 2013; 72:91–105. [PubMed: 23357075]
- Martinen P, Pirinen M, Sarin AP, Gillberg J, Kettunen J, Surakka I, Kangas AJ, Soininen P, O'Reilly P, Kaakinen M, Khnen M, Lehtimki T, Ala-Korpela M, Raitakari OT, Salomaa V, Jrvelin MR, Ripatti S, Kaski S. Assessing multivariate gene-metabolome associations with rare variants using Bayesian reduced rank regression. *Bioinformatics*. 2014; 30:2026–2034. DOI: 10.1093/bioinformatics/btu140 [PubMed: 24665129]
- Meinshausen N, Bühlmann P. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2010; 72:417–473.
- Morris JS, Carroll RJ. Wavelet-based functional mixed models. *J R Stat Soc Ser B Stat Methodol*. 2006; 68:179–199.
- Niemann C, Godde B, Voelcker-Rehage C. Not only cardiovascular, but also coordinative exercise increases hippocampal volume in older adults. *Frontiers in Aging Neuroscience*. 2014; 6:170. doi: 10.3389/fnagi.2014.00170 [PubMed: 25165446]
- Paus T. Population neuroscience: Why and how. *Human Brain Mapping*. 2010; 31:891–903. [PubMed: 20496380]
- Peper JS, Brouwer RM, Boomsma DI, Kahn RS, Pol HEH. Genetic influences on human brain structure: A review of brain imaging studies in twins. *Human Brain Mapping*. 2007; 28:464–473. [PubMed: 17415783]
- Rothman AJ, Levina E, Zhu J. Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics*. 2010; 19:947–962. [PubMed: 24963268]
- Ruppert, D., Wand, MP., Carroll, RJ. *Semiparametric regression*. Cambridge University Press; Cambridge: 2003.
- Saykin AJ, Shen L, Yao X, Kim S, Nho K, Risacher SL, Ramanan VK, Foroud TM, Faber KM, Sarwar N, Munsie LM, Hu X, Soares HD, Potkin SG, MTP, Kauwe JS, Kaddurah-Daouk R, Green RC, Toga AW, Weiner MW. Alzheimer's Disease Neuroimaging Initiative. Genetic studies of quantitative mci and ad phenotypes in adni: Progress, opportunities, and plans. *Alzheimer's & Dementia*. 2015; 11:792–814.
- Scharinger C, Rabl U, Sitte HH, Pezawas L. Imaging genetics of mood disorders. *NeuroImage*. 2010; 53:810–821. [PubMed: 20156570]
- Shen DG, Davatzikos C. Measuring temporal morphological changes robustly in brain mr images via 4-dimensional template warping. *NeuroImage*. 2004; 21:1508–1517. [PubMed: 15050575]
- Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West J, Foroud TM, Pankratz ND, Moore JH, Sloan SD, Huentelman MJ, Craig DW, DeChairo BM, Potkin SG, Jack CR, Weiner MW, Saykin AJ. ADNI. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in mci and ad: A study of the adni cohort. *Neuroimaging*. 2010; 53:1051–1063. [PubMed: 20100581]

- Sherva R, Tripodis Y, Bennett DA, Chibnik LB, Crane PK, de Jager PL, Farrer LA, Saykin AJ, Shulman JM, Naj A, Green RC. Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimer's & Dementia*. 2014; 10:45–52. DOI: 10.1016/j.jalz.2013.01.008
- Silver M, Janousova E, Hue X, Thompson P, Montana G. Identification of gene pathways implicated in alzheimer's disease using longitudinal imaging phenotypes with sparse regression. *Neuroimage*. 2012; 63:1681–1694. [PubMed: 22982105]
- Thomas D. Gene–environment-wide association studies: emerging approaches. *Nature Reviews Genetics*. 2010; 11:259–272.
- Verbeke, G., Molenberghs, G. Springer Series in Statistics. Springer; New York: 2009. Linear Mixed Models for Longitudinal Data.
- Vounou M, Janousova E, Wolz R, Stein JL, Thompson PM, Rueckert D, Montana G. ADNI. Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in alzheimer's disease. *NeuroImage*. 2011; 60:700–716. [PubMed: 22209813]
- Vounou M, Nichols TE, Montana G. the Alzheimer's Disease Neuroimaging Initiative. Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *Neuroimage*. 2010; 53:1147–1159. [PubMed: 20624472]
- Wang H, Nie F, Huang H, Kim S, Nho K, Risacher SL, Saykin AJ, Shen L. Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the adni cohort. *Bioinformatics*. 2011; 28:229–237. [PubMed: 22155867]
- Wood, SN. Texts in Statistical Science Series. Chapman & Hall/CRC; Boca Raton, FL: 2006. Generalized additive models: An introduction with R.
- Wu C, Cui Y. A novel method for identifying non-linear gene–environment interactions in case–control association studies. *Human Genetics*. 2013; 132:1413–1425. <http://dx.doi.org/10.1007/s00439-013-1350-z>. DOI: 10.1007/s00439-013-1350-z [PubMed: 23974428]
- Wu, HL., Zhang, JT. Nonparametric Regression Methods for Longitudinal Data Analysis. John Wiley & Sons, Inc; Hoboken, New Jersey: 2006.
- Xia M, Wang J, He Y. Brainnet viewer: A network visualization tool for human brain connectomics. *PLoS ONE*. 2013; 8:1–15. DOI: 10.1371/journal.pone.0068910
- Zhang Y, Xu Z, Shen X, Pan W. Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *NeuroImage*. 2014; 96:309–325. DOI: 10.1016/j.neuroimage.2014.03.061 [PubMed: 24704269]
- Zhao Y, Castellanos FX. Annual research review: Discovery science strategies in studies of the pathophysiology of child and adolescent psychiatric disorders - promises and limitations. *Journal of Child Psychology and Psychiatry*. 2016; 57:421–439. DOI: 10.1111/jcpp.12503 [PubMed: 26732133]
- Zhu H, Khondker Z, Lu Z, Ibrahim JG. Bayesian Generalized Low Rank Regression Models for Neuroimaging Phenotypes and Genetic Markers. *Journal of the American Statistical Association*. 2014; 109:977–990. DOI: 10.1080/01621459.2014.923775
- Zipunnikov V, Greven S, Shou H, Caffo BS, Reich DS, Crainiceanu CM. Longitudinal high-dimensional principal components analysis with application to diffusion tensor imaging of multiple sclerosis. *Ann Appl Stat*. 2014; 8:2175–2202. DOI: 10.1214/14-AOAS748 [PubMed: 25663955]

7. Appendix: technical details

7.1. Matrix formulation of L2R2

The L2R2 can be rewritten in the matrix form (5)

$$Y = XU\Delta V^T + WT + Zb + E,$$

where $\mathbf{y}_i(t) = (y_{i1}(t), \dots, y_{id}(t))^T$, $\mathbf{y}_i = (\mathbf{y}_i(t_{i1}), \dots, \mathbf{y}_i(t_{im_i}))^T$, $\mathbf{Y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T$, $\mathbf{1}_{m_i}$ be a $m_i \times 1$ vector of ones, and $\mathbf{t}_{m_i} = (t_{i1}, \dots, t_{im_i})^T$. Let \mathbf{x}_i^* be the SNPs for the i th subject, $\mathbf{X}_{main} = (\mathbf{x}_1^* \otimes \mathbf{1}_{m_1}^T, \dots, \mathbf{x}_n^* \otimes \mathbf{1}_{m_n}^T)^T$, $\mathbf{X}_{int} = (\mathbf{x}_1^* \otimes \mathbf{t}_{m_1}^T, \dots, \mathbf{x}_n^* \otimes \mathbf{t}_{m_n}^T)^T$, and $\mathbf{X} = (\mathbf{X}_{main}, \mathbf{X}_{int})$. Moreover, let $\mathbf{Z}_i = (\mathbf{z}_i(t_{i1}), \dots, \mathbf{z}_i(t_{im_i}))^T$, \mathbf{Z} be an $N \times np_b$ block diagonal matrix with diagonal elements $\mathbf{Z}_1, \dots, \mathbf{Z}_n$, and $\mathbf{b} = (\mathbf{b}_{ik})$ be an $np_b \times d$ block matrix with elements \mathbf{b}_{ik} . We defined $\mathbf{w}_i(t) = (\mathbf{w}_1(t)^T, \mathbf{w}_2(t)^T)^T$, $\mathbf{W}_i = (\mathbf{w}_i(1), \dots, \mathbf{w}_i(t_{im_i}))^T$, and $\mathbf{W} = (\mathbf{W}_1^T, \dots, \mathbf{W}_n^T)^T$. Let $\mathbf{e}_i(t) = (\mathbf{e}_{i1}(t), \dots, \mathbf{e}_{id}(t))^T$, $\mathbf{E}_i = (\mathbf{e}_i(t_{i1}), \dots, \mathbf{e}_i(t_{im_i}))^T$, and $\mathbf{E} = (\mathbf{E}_1^T, \dots, \mathbf{E}_n^T)^T$. Consequently, \mathbf{Y} is an $N \times d$ matrix of imaging measurements; \mathbf{X} is an $N \times p$ matrix of genetic predictors; \mathbf{W} is an $N \times q$ matrix of covariates with fixed effects; \mathbf{Z} is an $N \times np_b$ matrix of covariates with random effects; and \mathbf{E} is an $N \times d$ matrix of error terms, where $q = q_1 + q_2$. $\mathbf{\Gamma}$ is a $q \times d$ matrix and its first q_1 row consists of $\mathbf{\Gamma}_1 = (\boldsymbol{\gamma}_{1,1}, \dots, \boldsymbol{\gamma}_{1,d})$, and its last q_2 columns consist of $\mathbf{\Gamma}_2 = (\boldsymbol{\gamma}_{2,1}, \dots, \boldsymbol{\gamma}_{2,d})$. In model (5), we primarily made statistical inferences from \mathbf{B} (equivalently \mathbf{U} , \mathbf{V}) and $\mathbf{\Gamma}$.

7.2. Priors

We consider priors on the elements of \mathbf{B} . Let $\text{Ga}(a, b)$ be a gamma distribution with scale a and shape b . We choose L_2 priors on the parameters at each layer \mathbf{B}_l as follows:

$$\delta_l \sim N(0, \tau_\delta^{-1}), \quad \tau_\delta \sim \text{Ga}(a_0, b_0), \quad \mathbf{u}_l \sim N_p(0, p^{-1}I_p), \quad \text{and} \quad \mathbf{v}_l \sim N_d(0, d^{-1}I_d),$$

where a_0 and b_0 are prespecified hyperparameters. Although, we have used the same precision parameter for each element of \mathbf{u}_l , group information can be incorporated by choosing separate precision parameters for each group. The number of predictors p (or d) is included in the hyperprior of \mathbf{u}_l (or \mathbf{v}_l) to have a positive definite covariance matrix of high dimensional \mathbf{u}_l (or \mathbf{v}_l). Moreover, this data driven approach for the priors of \mathbf{u}_l and \mathbf{v}_l requires no additional hyperparameters to choose, because we standardize all predictors to have zero mean and unit variance, a single prior suffices for all elements of \mathbf{u}_l . Moreover, because we rescale all of the responses, we use the same dispersion for all components of \mathbf{v}_l . Since we focus on exploiting the potential two-way correlations among the estimated coefficients, we choose the L_2 priors, which tend to borrow strength from correlated neighbors and force the coefficients towards each other to produce two highly correlated coefficients. Moreover, the posterior computations are simpler and faster under the L_2 priors.

We consider priors on the elements of $\mathbf{\Gamma} = (\boldsymbol{\gamma}_{jk})$. We also choose the L_2 prior on the $\boldsymbol{\gamma}_{jks}$ as follows:

$$\boldsymbol{\gamma}_{jk} \sim N(0, \tau_\gamma^{-1}) \quad \text{and} \quad \tau_\gamma \sim \text{Ga}(c_0, d_0),$$

where c_0 and d_0 are hyperparameters. For the subject-specific random coefficients, we also choose independent and identically distributed normal priors as

$$\tau_b \sim \text{Ga}(e_0, f_0),$$

where e_0 and f_0 are hyperparameters.

For the sparse factor model (4), we assign the multiplicative gamma process shrinkage prior (Bhattacharya and Dunson, 2011) on Λ to automatically determine the dimension of the factors needed to characterize the error covariance structure. The shrinkage prior increasingly shrinks the factor loadings towards zero with the column index and avoids identifiability of the order of the factor in $\boldsymbol{\eta}_k(t)$. Specifically, these priors are summarized as follows:

$$\begin{aligned}\Lambda &= \{\lambda_{kh}\}, k=1, \dots, d, h=1, \dots, \infty, \\ \lambda_{kh} | \phi_{kh}, \tau_{\lambda h} &\sim N(0, \phi_{kh}^{-1} \tau_{\lambda h}^{-1}), \phi_{kh} \sim \text{Ga}(v/2, v/2), \tau_{\lambda h} = \prod_{l=1}^h \psi_l, \\ \psi_1 &\sim \text{Ga}(a_1, 1), \psi_l \sim \text{Ga}(a_2, 1), l \geq 2, \sigma_{\xi, k}^{-2} \sim \text{Ga}(a_{\sigma k}, b_{\sigma k}),\end{aligned}$$

where $v, a_1, a_2, a_{\sigma k}$, and $b_{\sigma k}$ are prefixed hyperparameters, $\tau_{\lambda h}$ is a global shrinkage parameter for the h -th column, and the ϕ_{kh} s are local shrinkage parameters for the elements in the h -th column. When $a_2 > 1$, the $\tau_{\lambda h}$ s increase stochastically with the column index h , which shrinks the elements of the loading matrix to zero as the column index progresses and determines the effective dimension of Λ .

7.3. Posterior computation

The joint posterior for the L2R2 model with the above priors can be written as

$$\begin{aligned}p(\mathbf{U}, \Delta, \mathbf{V}, \Gamma, \mathbf{b}, \Theta | \text{data}) &\propto \tau_{\delta}^{a_0 + \frac{1}{2}r - 1} \tau_{\gamma}^{c_0 + \frac{1}{2}qd - 1} \tau_b^{e_0 + \frac{1}{2}np_b d - 1} e^{-b_0 \tau_{\delta} - d_0 \tau_{\gamma} - f_0 \tau_b} \\ &\exp \left\{ -\frac{1}{2} \sum_{l=1}^r (\tau_{\delta} \delta_l^2 + d \mathbf{v}_l^T \mathbf{v}_l + p \mathbf{u}_l^T \mathbf{u}_l) \right\} + \exp \left\{ -\frac{1}{2} \text{Tr}(\tau_{\gamma} \Gamma \Gamma^T + \tau_b \mathbf{b} \mathbf{b}^T) \right\} \\ &\exp \left\{ -\frac{1}{2} \text{Tr}(\mathbf{Y} - \mathbf{X} \mathbf{U} \Delta \mathbf{V} - \mathbf{W} \Gamma - \mathbf{Z} \mathbf{b}) \Theta (\mathbf{Y} - \mathbf{X} \mathbf{U} \Delta \mathbf{V} - \mathbf{W} \Gamma - \mathbf{Z} \mathbf{b})^T \right\},\end{aligned}\quad (6)$$

where $\text{Tr}(\cdot)$ is the trace of the matrix.

We propose a straightforward Gibbs sampler for posterior computation, which converges rapidly. Starting from the initiation step, the Gibbs sampler at each iteration proceeds as follows:

1. For $l = 1, \dots, r$, update \mathbf{u}_l from the full conditional distributions

$$p(\mathbf{u}_l | -) \sim N_p(\delta_l \sum_{\mathbf{u}_i} \mathbf{X}^T \mathbf{Y}_{B,l} \Theta \mathbf{v}_l, \sum_{\mathbf{u}_i}),$$

where $\mathbf{Y}_{B,l} = \mathbf{Y} - \mathbf{X} \sum_{l' \neq l} \mathbf{B}_{l'} - \mathbf{W} \Gamma - \mathbf{Z} \mathbf{b}$ and $\sum_{\mathbf{u}_i} = \{p \mathbf{I}_p + \delta_l^2 (\mathbf{v}_l^T \Theta \mathbf{v}_l) \mathbf{X}^T \mathbf{X}\}^{-1}$.

2. Update \mathbf{v}_l from

$$p(\mathbf{v}_l|-) \sim N_d(\delta_l \sum_{\mathbf{v}_l} \Theta \mathbf{Y}_{B,l}^T \mathbf{X} \mathbf{u}_l, \sum_{\mathbf{v}_l}),$$

$$\text{where } \sum_{\mathbf{v}_l} = \{d \mathbf{I}_d + \delta_l^2 (\mathbf{u}_l^T \mathbf{X}^T \mathbf{X} \mathbf{u}_l) \Theta\}^{-1}.$$

3. Update δ_l from

$$p(\delta_l|-) \sim N(\sigma_{\delta_l}^2 \mathbf{u}_l^T \mathbf{X}^T \mathbf{Y}_{B,l} \Theta \mathbf{v}_l, \sigma_{\delta_l}^2),$$

$$\text{where } \sigma_{\delta_l}^2 = \{\tau_\delta + (\mathbf{v}_l \Theta^T \mathbf{v}_l) (\mathbf{u}_l^T \mathbf{X}^T \mathbf{X} \mathbf{u}_l)\}^{-1}.$$

4. Update τ_δ from

$$p(\tau_\delta|-) \sim \text{Ga}(a_0 + 0.5r, b_0 + 0.5 \sum_{l=1}^r \delta_l^2).$$

5. Update $\boldsymbol{\gamma}_k$ (i.e. the k th column of Γ) from

$$p(\boldsymbol{\gamma}_k|-) \sim N_d(\sum_{\boldsymbol{\gamma}_k} \mathbf{W}^T \{\mathbf{y}_{\Gamma,k} \theta_{kk} + (\mathbf{Y}_{\Gamma,-k} - \mathbf{W} \Gamma_{-k}) \boldsymbol{\theta}_k, \sum_{\boldsymbol{\gamma}_k}),$$

where $\boldsymbol{\Sigma}_{\Gamma,k} = \{\theta_{kk} \mathbf{W}^T \mathbf{W} + \tau_\gamma \mathbf{I}_q\}^{-1}$. $\mathbf{y}_{\Gamma,k}$ is the k th column of $\mathbf{Y}_\Gamma = \mathbf{Y} - \mathbf{X}\mathbf{B} - \mathbf{Z}\mathbf{b}$; $\mathbf{Y}_{\Gamma,-k}$ is the matrix after dropping the k th column of \mathbf{Y}_Γ ; θ_{kk} is the element for the k th row and k th column of Θ ; $\boldsymbol{\theta}_k$ is the k th column of Θ after dropping θ_{kk} ; and Θ_{-kk} is the matrix after dropping k th row and k th column of Θ .

6. Update τ_γ from

$$p(\tau_\gamma|-) \sim \text{Ga}(c_0 + 0.5qd, d_0 + 0.5 \sum_{j=1}^q \sum_{k=1}^d \gamma_{jk}^2).$$

7. Let $\tilde{\mathbf{b}}_k$ be the k th column of \mathbf{b} . Update $\tilde{\mathbf{b}}_k$ from

$$p(\tilde{\mathbf{b}}_k|-) \sim N(\sum_{b_k} \mathbf{Z}^T \{\mathbf{y}_{b,k} - (\mathbf{Y}_{b,-k} - \mathbf{Z} \mathbf{b}_{-k}) \boldsymbol{\theta}_k\}, \sum_{b_k}^2),$$

where $\boldsymbol{\Sigma}_{b_k} = (\theta_{kk} \mathbf{Z}^T \mathbf{Z} + \tau_b \mathbf{I}_{pp_b})^{-1}$. $\mathbf{y}_{b,k}$ is the k th column of $\mathbf{Y}_b = \mathbf{Y} - \mathbf{X}\mathbf{B} - \mathbf{W}\Gamma$, and $\mathbf{Y}_{b,-k}$ is the matrix after dropping the k th column of \mathbf{Y}_b .

8. Update τ_b from

$$p(\tau_b|-) \sim \text{Ga}(e_0 + 0.5np_b d, f_0 + 0.5 \sum_{i=1}^n \sum_{j=1}^{p_b} \sum_{k=1}^d b_{ijk}^2).$$

9. Let k_* be a integer large enough to approximate the number of factors in the factor model (4), and $\varepsilon_{it} = \Lambda^* \boldsymbol{\eta}_{it}^* + \boldsymbol{\xi}_{it}$ be the corresponding approximation. Update $\Lambda_{k^*}^*$, which is the k th row of Λ^* , from its full conditional distribution

$$p(\Lambda_{k^*}^* | -) \sim \mathcal{N}((\sigma_{\xi,k}^{-2} \boldsymbol{\Omega}^T \boldsymbol{\Omega} + \mathbf{D}_k^{-1})^{-1} \boldsymbol{\Omega}^T \sigma_{\xi,k}^{-2} \mathbf{E}_k, (\sigma_{\xi,k}^{-2} \boldsymbol{\Omega}^T \boldsymbol{\Omega} + \mathbf{D}_k^{-1})^{-1}),$$

where $\boldsymbol{\Omega} = (\boldsymbol{\eta}_{11}, \dots, \boldsymbol{\eta}_{1m_1}, \dots, \boldsymbol{\eta}_{nm_p})^T$, \mathbf{E}_k is the k th column of $\mathbf{E} = \mathbf{Y} - \mathbf{X}\mathbf{B} - \mathbf{W}\mathbf{T} - \mathbf{Z}\mathbf{b}$, and $D_k = \text{diag}(\phi_{k1}^{-1} \tau_{\lambda 1}^{-1}, \dots, \phi_{kk_*}^{-1} \tau_{\lambda k_*}^{-1})$ for $k = 1, \dots, d$.

10. Update ϕ_{kh} from its full conditional distribution

$$p(\phi_{kh} | -) \sim \text{Ga}\left(\frac{v+1}{2}, \frac{v + \lambda_{kh}^2 \tau_{\lambda h}}{2}\right).$$

11. Update $\sigma_{\xi,k}^{-2}$, $k = 1, \dots, d$, from its full conditional distribution

$$p(\sigma_{\xi,k}^{-2} | -) \sim \text{Ga}\left(a_{\sigma k} + \frac{n}{2}, b_{\sigma k} + \frac{1}{2} \sum_{i=1}^n \sum_{t=1}^{m_i} (\varepsilon_{itk} - \Lambda_{k^*}^{*T} \boldsymbol{\eta}_{it}^*)^2\right),$$

where ε_{itk} is the element in \mathbf{E} for the i th subject at time t and k th response variable.

12. Update ψ_1 from its full conditional distribution

$$p(\psi_1 | -) \sim \text{Ga}\left(a_1 + \frac{1}{2} dk_*, 1 + \frac{1}{2} \sum_{g=1}^{k_*} \tau_{\lambda g}^{(1)} \sum_{k=1}^d \phi_{kg} \lambda_{kg}^2\right),$$

where $\tau_{\lambda g}^{(1)} = \prod_{t=2}^g \psi_t$ for $h = 1, \dots, k_*$.

13. Update ψ_h , $h = 2$ from its full conditional distribution

$$p(\psi_h | -) \sim \text{Ga}\left(a_2 + \frac{1}{2} d(k_* - h + 1), 1 + \frac{1}{2} \sum_{g=h}^{k_*} \tau_{\lambda g}^{(h)} \sum_{k=1}^d \phi_{kg} \lambda_{kg}^2\right),$$

where $\tau_{\lambda g}^{(h)} = \prod_{t=1, t \neq h}^g \psi_t$ for $h = 1, \dots, k_*$.

14. Update $\boldsymbol{\eta}_{it}$ for $t = 1, \dots, m_i$, $i = 1, \dots, n$ from the conditionally independent posteriors

$$p(\boldsymbol{\eta}_{it} | -) \sim \mathcal{N}((I_{k_*} + \Lambda^T \sum_{\xi}^{-1} \Lambda)^{-1} \Lambda_{k_*}^T \sum_{\xi}^{-1} \varepsilon_{it}, (I_{k_*} + \Lambda^T \sum_{\xi}^{-1} \Lambda)^{-1}),$$

where ε_{it} is the row of \mathbf{E} corresponding to the i th subject at time t .

7.4. Full conditional distributions of L2R2

7.4.1. Full conditional distribution for

From equation (6), we can write

$$\begin{aligned} p(\delta_l|-) &\propto \exp\left(-\frac{1}{2}\tau_\delta\delta_l^2\right) \text{etr}\left\{\left(\mathbf{Y}_{B,l}-\delta_l\mathbf{X}\mathbf{u}_l\mathbf{v}_l\right)\Theta\left(\mathbf{Y}_{B,l}-\delta_l\mathbf{X}\mathbf{u}_l\mathbf{v}_l\right)^\top\right\} \\ &\propto \exp\left(\delta_l\mathbf{u}_l^\top\mathbf{X}^\top\mathbf{Y}_{B,l}\Theta\mathbf{v}_l-\frac{1}{2}\delta_l^2\left(\tau_\delta+\mathbf{v}_l\Theta^\top\mathbf{v}_l\right)\left(\mathbf{u}_l^\top\mathbf{X}^\top\mathbf{X}\mathbf{u}_l\right)\right); \\ p(\tau_\delta|-) &\propto \tau_\delta^{a_0+\frac{1}{2}r-1} \exp\left(-\left(b_0+\frac{1}{2}\sum_{l=1}^r\tau_\delta\delta_l^2\right)\right). \end{aligned}$$

This implies

$$p(\delta_l|-) \sim N\left(\sigma_{\delta_l}^2\mathbf{u}_l^\top\mathbf{X}^\top\mathbf{Y}_{B,l}\Theta\mathbf{v}_l, \sigma_{\delta_l}^2\right) \text{ and } p(\tau_\delta|-) \sim \text{Ga}\left(a_0+\frac{1}{2}r, b_0+\frac{1}{2}\sum_{l=1}^r\delta_l^2\right),$$

where $\sigma_{\delta_l}^2 = \left\{T_\delta + (\mathbf{v}_l\Theta^\top\mathbf{v}_l)(\mathbf{u}_l^\top\mathbf{X}^\top\mathbf{X}\mathbf{u}_l)\right\}^{-1}$ and $\mathbf{Y}_{B,l} = \mathbf{Y} - \mathbf{W}\Gamma - \mathbf{Z}\mathbf{b} - \sum_{l' \neq l} \delta_{l'}\mathbf{X}\mathbf{u}_{l'}\mathbf{v}_{l'}^\top$.

7.4.2. Full conditional distribution for U

From equation (6), we can write

$$\begin{aligned} p(\mathbf{u}_l|-) &\propto \exp\left(-\frac{1}{2}p\mathbf{u}_l^\top\mathbf{u}_l\right) \text{etr}\left\{\left(\mathbf{Y}_{B,l}-\delta_l\mathbf{X}\mathbf{u}_l\mathbf{v}_l\right)\Theta\left(\mathbf{Y}_{B,l}-\delta_l\mathbf{X}\mathbf{u}_l\mathbf{v}_l\right)^\top\right\} \\ &\propto \text{etr}\left(\mathbf{u}_l^\top\delta_l\mathbf{X}^\top\mathbf{Y}_{B,l}-\frac{1}{2}\mathbf{u}_l^\top\left(p\mathbf{I}_p+\delta_l^2\mathbf{v}_l\Theta^\top\mathbf{v}_l\mathbf{X}^\top\mathbf{X}\right)\mathbf{u}_l\right). \end{aligned}$$

This implies

$$p(\mathbf{u}_l|-) \sim N_p\left(\delta_l\sum_{\mathbf{u}_l}\mathbf{X}^\top\mathbf{Y}_{B,l}\Theta\mathbf{v}_l, \sum_{\mathbf{u}_l}\right),$$

where $\sum_{\mathbf{u}_l} = \left\{p\mathbf{I}_p + \delta_l^2(\mathbf{v}_l\Theta^\top\mathbf{v}_l)\mathbf{X}^\top\mathbf{X}\right\}^{-1}$.

7.4.3. Full conditional distribution for V

From equation (6), we can write

$$\begin{aligned} p(\mathbf{v}_l|-) &\propto \exp\left(-\frac{1}{2}\mathbf{v}_l^\top(d\mathbf{I}_d)\mathbf{v}_l\right) \text{etr}\left(\left(\mathbf{Y}_{B,l}-\delta_l\mathbf{X}\mathbf{u}_l\mathbf{v}_l\right)\Theta\left(\mathbf{Y}_{B,l}-\delta_l\mathbf{X}\mathbf{u}_l\mathbf{v}_l\right)^\top\right) \\ &\propto \exp\left(\mathbf{v}_l^\top\delta_l\Theta\mathbf{Y}_{B,l}^\top\mathbf{X}\mathbf{u}_l-\frac{1}{2}\mathbf{v}_l^\top\left(p\mathbf{I}_p+\delta_l^2\mathbf{u}_l^\top\mathbf{X}^\top\mathbf{X}\mathbf{u}_l\mathbf{v}_l\mathbf{X}^\top\mathbf{X}\right)\mathbf{v}_l\right). \end{aligned}$$

This gives

$$p(\mathbf{v}_l|-) \sim N_d(\delta_l \sum_{\mathbf{v}_l} \Theta \mathbf{Y}_{B,l}^T \mathbf{X} \mathbf{u}_l, \sum_{\mathbf{v}_l}),$$

$$\text{where } \sum_{\mathbf{v}_l} = \{d \mathbf{I}_d + \delta_l^2 (\mathbf{u}_l^T \mathbf{X}^T \mathbf{X} \mathbf{u}_l) \Theta\}^{-1}.$$

7.4.4. Sampling Γ by columns

Sampling the coefficient matrix one element at a time would be time consuming and less attractive in high-dimensional settings. Another approach is to convert the whole matrix into a vector and sample the vector at once, but this requires a covariance matrix with dimension $qd \times qd$, which can be quite large and requires considerable memory making it infeasible. We propose a columnwise sampling scheme, which allows for computationally efficient sampling with a feasible dimension of the conditional covariance matrix (Khondker et al., 2013). Let $\mathbf{Y}_\Gamma = \mathbf{Y} - \mathbf{X} \mathbf{U}$, $\mathbf{V} - \mathbf{Z} \mathbf{b}$, we can write $\mathbf{Y}_\Gamma = \mathbf{W} \Gamma + \mathbf{E}$. For $k = 1, \dots, d$, we can reorder and partition

$$\mathbf{Y}_\Gamma = (\mathbf{y}_{\Gamma,k} \mathbf{Y}_{\Gamma,-k}), \quad \Gamma = (\boldsymbol{\gamma}_k \Gamma_{-k}), \quad \text{and } \Theta = \begin{pmatrix} \theta_{kk} & \boldsymbol{\theta}_k \\ \boldsymbol{\theta}_k^T & \Theta_{-kk} \end{pmatrix}.$$

In the above partition, $\mathbf{y}_{\Gamma,k}$ is the k th column of \mathbf{Y}_Γ , and $\mathbf{Y}_{\Gamma,-k}$ and Γ_{-k} are the matrices after dropping the k th column of \mathbf{Y}_Γ and Γ , respectively. θ_{kk} is the element for the k th row and k th column of Θ ; $\boldsymbol{\theta}_k$ is the k th column of Θ after dropping θ_{kk} ; and Θ_{-kk} is the matrix after dropping the k th row and k th column of Θ . We can write

$$\begin{aligned} p(\boldsymbol{\gamma}_k|-) &\propto \text{etr} \left\{ -\frac{1}{2} (\mathbf{Y}_\Gamma - \mathbf{W} \Gamma)^T (\mathbf{Y}_\Gamma - \mathbf{W} \Gamma) \Theta \right\} \exp \left(-\frac{1}{2} \tau_\gamma \boldsymbol{\gamma}_k^T \mathbf{I}_q \boldsymbol{\gamma}_k \right) \\ &\propto \text{etr} \left(-\frac{1}{2} (\mathbf{y}_{\Gamma,k} - \mathbf{W} \boldsymbol{\gamma}_k)^T (\mathbf{y}_{\Gamma,k} - \mathbf{W} \boldsymbol{\gamma}_k) \theta_{kk} \right) \times \text{etr} \left((\mathbf{y}_{\Gamma,k} - \mathbf{W} \boldsymbol{\gamma}_k)^T (\mathbf{Y}_{\Gamma,-k} - \mathbf{W} \Gamma_{-k}) \boldsymbol{\theta}_k \right) \exp \left(-\frac{1}{2} \boldsymbol{\gamma}_k^T (\tau_\gamma \mathbf{I}_q) \boldsymbol{\gamma}_k \right) \\ &\propto \exp \left(\boldsymbol{\gamma}_k^T \mathbf{W}^T \{ \mathbf{y}_{\Gamma,k} \theta_{kk} + (\mathbf{Y}_{\Gamma,-k} - \mathbf{W} \Gamma_{-k}) \boldsymbol{\theta}_k \} - \frac{1}{2} \boldsymbol{\gamma}_k^T (\theta_{kk} \mathbf{W}^T \mathbf{W} + (\tau_\gamma \mathbf{I}_q)) \boldsymbol{\gamma}_k \right). \end{aligned}$$

This gives us

$$p(\boldsymbol{\gamma}_k|-) \sim N_d \left(\sum_{\boldsymbol{\gamma}_k} \mathbf{W}^T \{ \mathbf{y}_{\Gamma,k} \theta_{kk} + (\mathbf{Y}_{\Gamma,-k} - \mathbf{W} \Gamma_{-k}) \boldsymbol{\theta}_k \}, \sum_{\boldsymbol{\gamma}_k} \right),$$

$$\text{where } \boldsymbol{\Sigma}_{\boldsymbol{\gamma}_k} = \{ \theta_{kk} \mathbf{W}^T \mathbf{W} + (\tau_\gamma \mathbf{I}_q) \}^{-1}.$$

7.4.5. Sampling \mathbf{b}

Given that $\text{vec}(\mathbf{b}_j) \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b)$ with $\boldsymbol{\Sigma}_b = \tau_b \mathbf{I}_d$, the full conditionals for \mathbf{b} can be derived in a similar way to Γ . Let $\tilde{\mathbf{b}}_k$ be the k th column of \mathbf{b} . Update $\tilde{\mathbf{b}}_k$ from

$$p(\tilde{\mathbf{b}}_k | -) \sim N\left(\sum_{b_k} \mathbf{Z}^T \{\mathbf{y}_{b,k} - (\mathbf{Y}_{b,-k} - \mathbf{Z}\mathbf{b}_{-k})\boldsymbol{\theta}_k\}, \sum_{b_k}^2\right),$$

where $\boldsymbol{\Sigma}_{b_k} = (\boldsymbol{\theta}_k \mathbf{Z}^T \mathbf{Z} + \tau_b \mathbf{I}_{np_b})^{-1}$. $\mathbf{y}_{b,k}$ is the k th column of $\mathbf{Y}_b = \mathbf{Y} - \mathbf{X}\mathbf{B} - \mathbf{W}\boldsymbol{\Gamma}$; and $\mathbf{Y}_{b,-k}$ and \mathbf{b}_{-k} is the matrix after dropping the k th column of \mathbf{Y}_b and \mathbf{b} , respectively.

Full conditionals for all other parameters are straightforward.

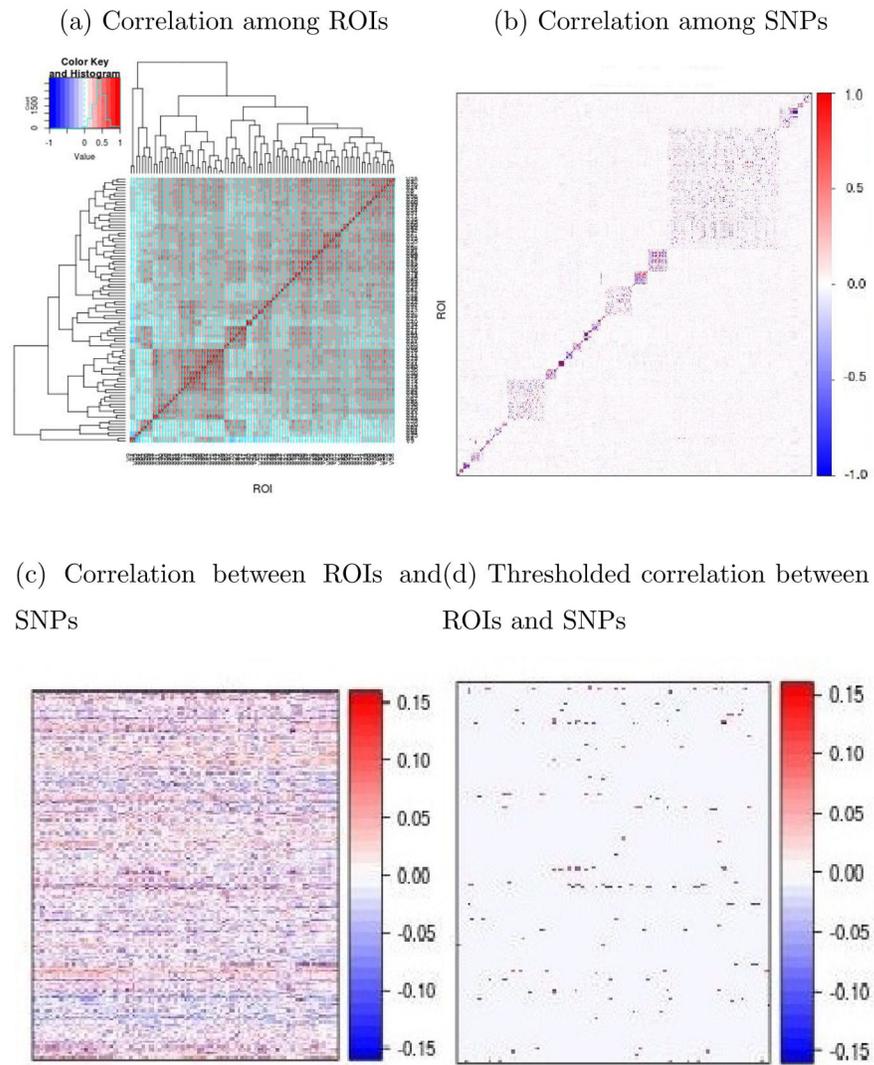


Fig. 1. The characteristics of the ROIs and SNPs from the ADNI data set and their association.

(a) Lateral ventricle left (b) rs769451, hippocam- (c) rs439401, hippocam-
 pal formation right pal formation left

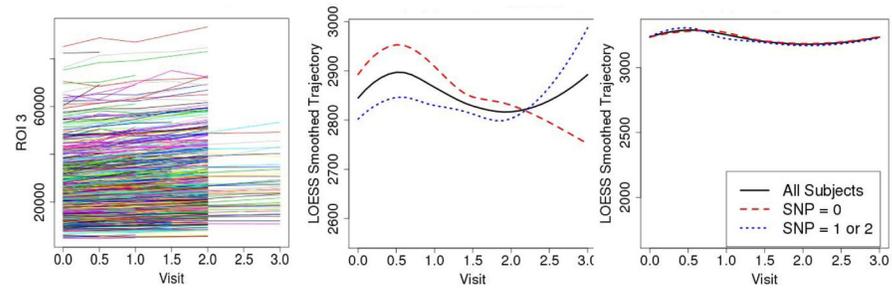


Fig. 2. The longitudinal characteristics of the ROI volumes in the ADNI data set. (a) The trajectories of the volumes of left lateral ventricle for all subjects. (b) and (c) The LOESS estimates of two volume trajectories based on the subjects with different SNP alleles.

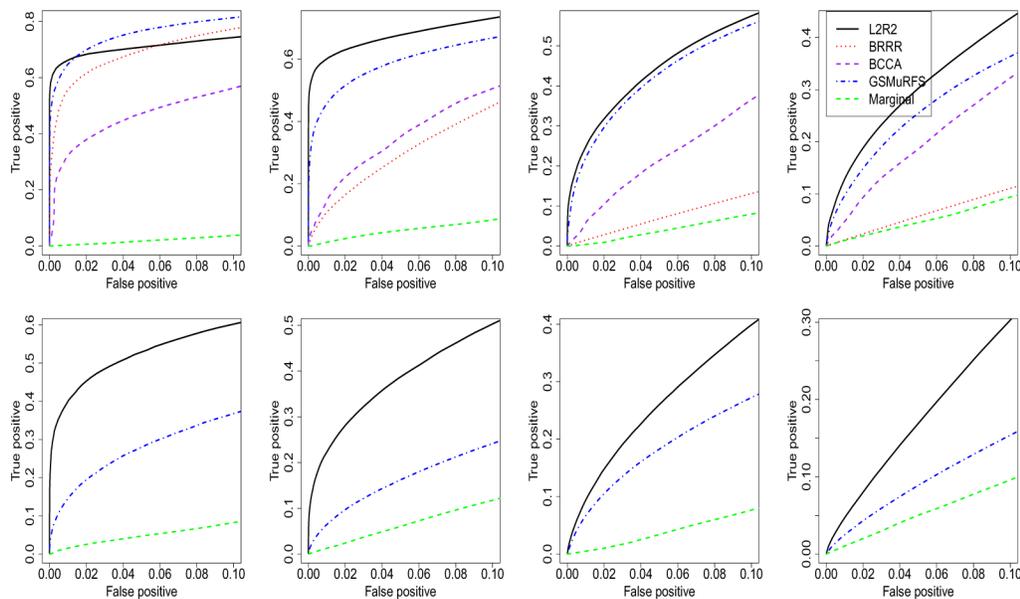


Fig. 3. Simulation results depicting the estimated ROC curves of the true positive rates and false positive rates for **B**. The four columns correspond to the settings (i) $p = 50, d = 50$; (ii) $p = 100, d = 100$; (iii) $p = 200, d = 100$; (iv) $p = 400, d = 100$. The first row shows the coefficients of the SNPs and the second row shows the coefficients of the interactions between the SNPs and subject age. The lines represent the L2R2 (black solid), BRRR (red dotted), BCCA (purple dashed), G-SMuRFS (blue dot-dash), and MMA (green dashed).

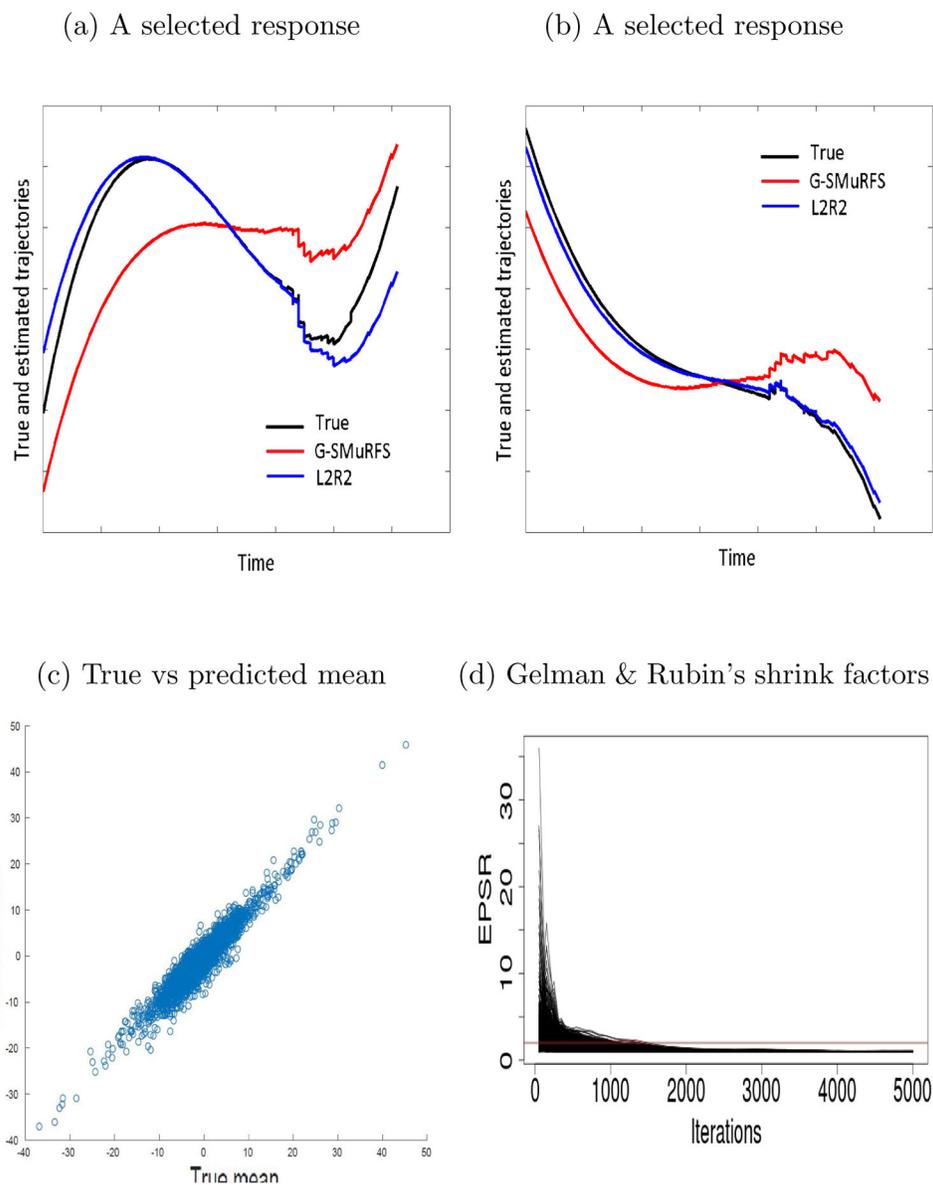


Fig. 4. Simulation results from one randomly selected replication with $p = 50$, $d = 50$, and moderately sparse \mathbf{B} . (a) and (b) The true and estimated splines for a randomly selected standardized ROI volume. (c) The scatterplot showing the true mean versus the predicted mean at one year after the subject's last observation using model (5). (d) The Gelman and Rubin's shrink factors at different iterations for all the parameters in \mathbf{B} .

(a) Thresholded main genetic effect (b) Thresholded SNP-age effect

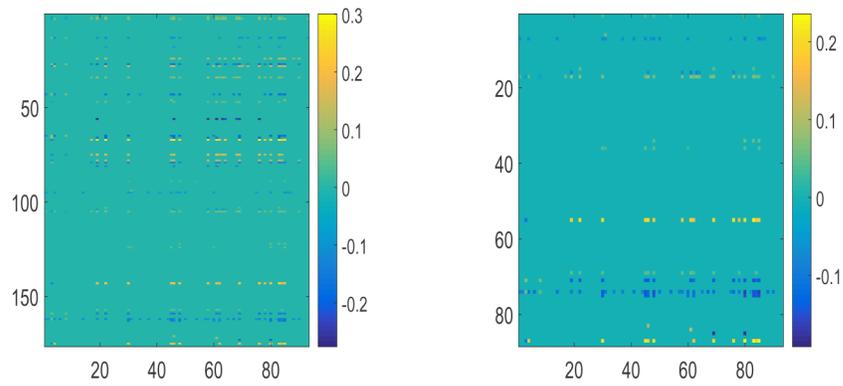


Fig. 5.

The thresholded estimates of the main genetic effects and the SNP-age interaction effects on ROIs based on the ADNI data set with the top 10 genes.

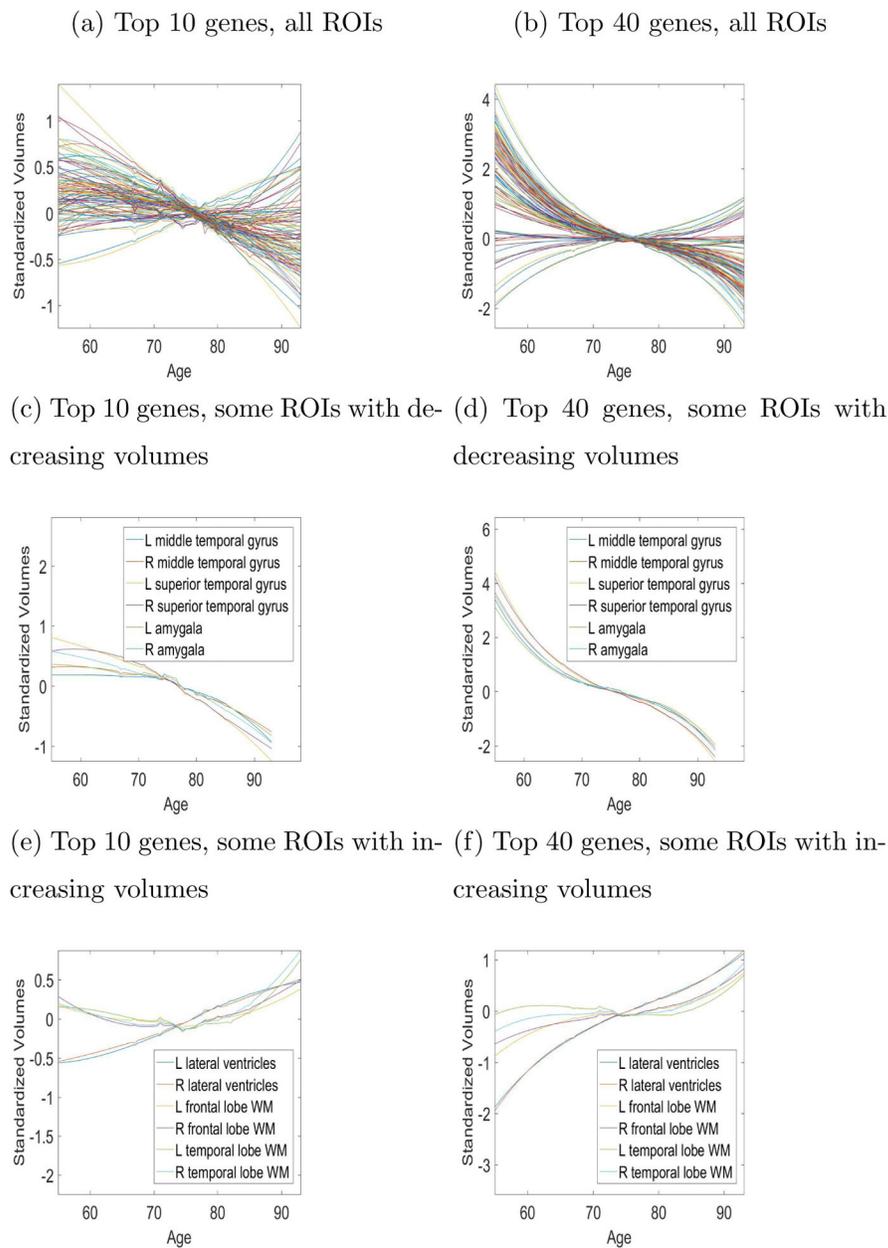
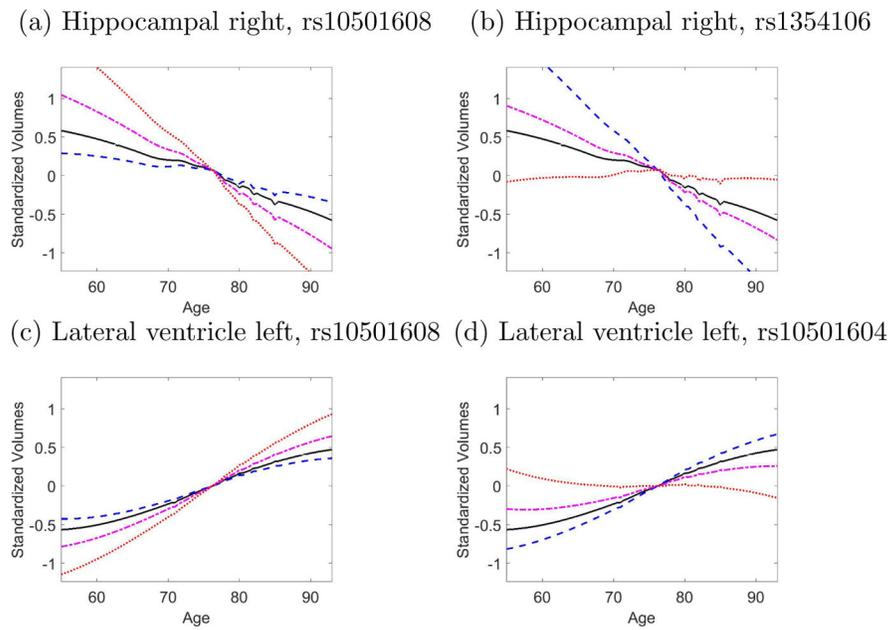


Fig. 6. The estimated spline functions for (a) and (b) all the ROIs, (c) and (d) selected ROIs with declining volumes, and (e) and (f) selected ROIs with increasing volumes based on the SNPs from the ADNI data set with the top 10 genes and top 40 genes, respectively.

**Fig. 7.**

The effects of SNP alleles on the trajectories of two ROIs in the ADNI analysis. Black solid line represents the overall effects characterized by the basis functions and coefficients in Γ . The blue dashed, magenta dot-dash, and red dotted lines are the trajectories of the ROIs altered by the SNP with AA, Aa, and aa, respectively. 'A' represents the major allele and 'a' represents the minor allele of a SNP.

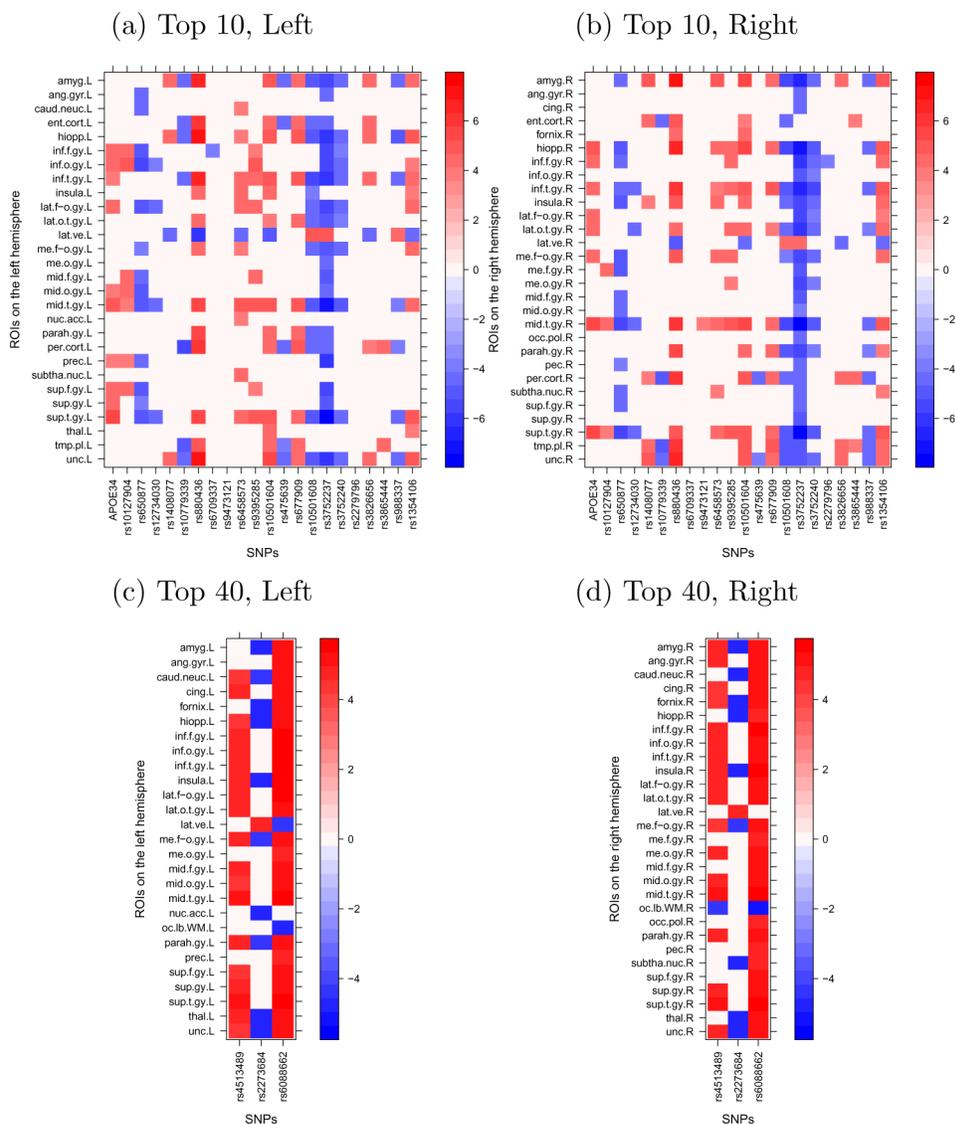


Fig. 8. Heatmaps of the standardized estimates of the SNP-age interaction coefficients in the (a) and (c) left and (b) and (d) right brain hemispheres. (a) and (b) represent the estimates with the top 10 genes, and (c) and (d) represent the estimates with the top 40 genes. Standardized estimates with absolute value smaller than the threshold were set to zero.

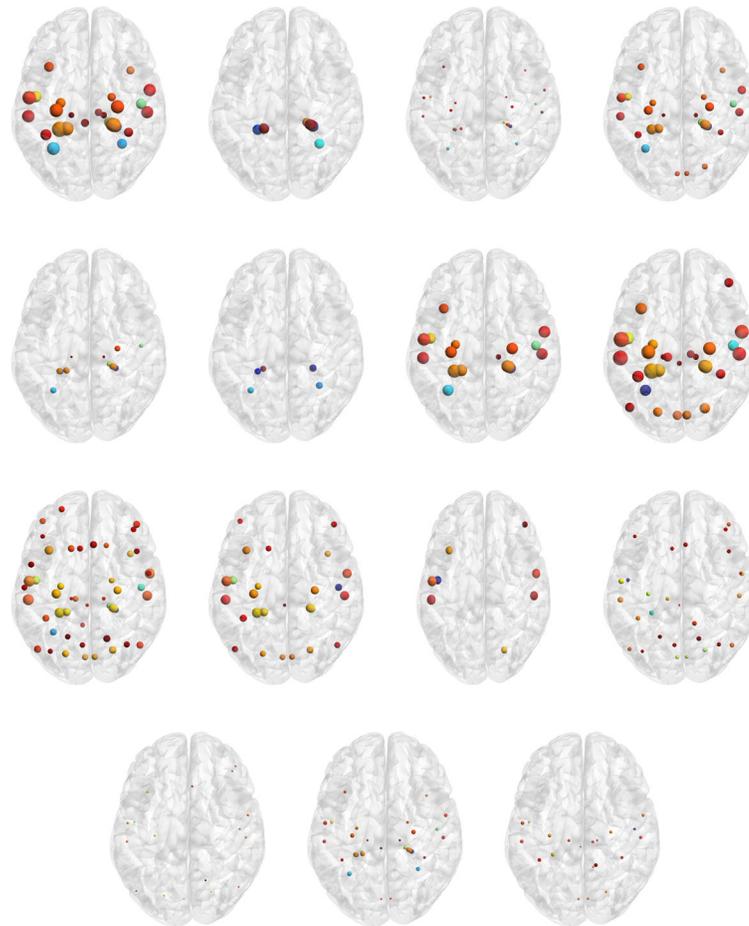


Fig. 9. The SNP-specific networks of ROIs based on the ADNI data set with the top 10 genes. The first row depicts the locations of ROIs that are correlated with the SNPs rs10501604, rs475639, rs677909, and rs10501608 in gene *PICALM*; the second row shows those of rs3826656, rs3865444, rs988337, rs1354106 in gene *CD33*; the third row shows those of rs3752237 and rs3752240 in gene *ABCA7* and rs12734030 and rs650877 in gene *CRI*; the last row shows those of APOE34 (*APOE*), rs880436 (*BINI*) and rs6458573 (*CD2AP*). The sizes of the dots represent the absolute magnitudes of the regression coefficients. Warmer and cooler colors represent ROIs located in superior and inferior anatomical sites of the brain, respectively.

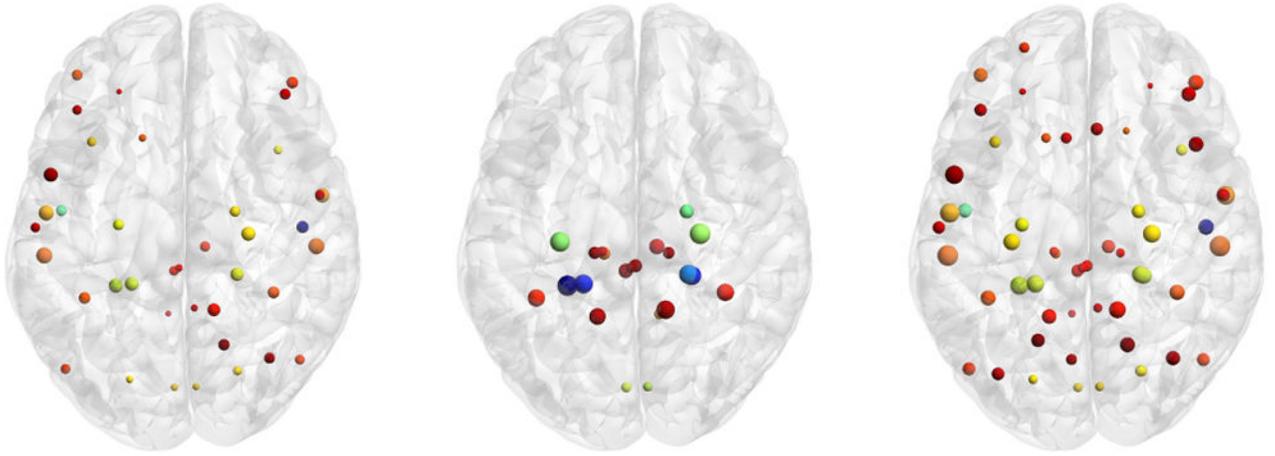


Fig. 10.

The SNP-specific networks of ROIs based on the ADNI data set with the top 40 genes, which show the locations of ROIs that are correlated with SNPs rs4513489 (*CCR2*), rs2273684 (*PRNP*) and rs6088662 (*PRNP*). The sizes of the dots represent the absolute magnitudes of the regression coefficients. Warmer and cooler colors represent ROIs located in superior and inferior anatomical sites of the brain, respectively.

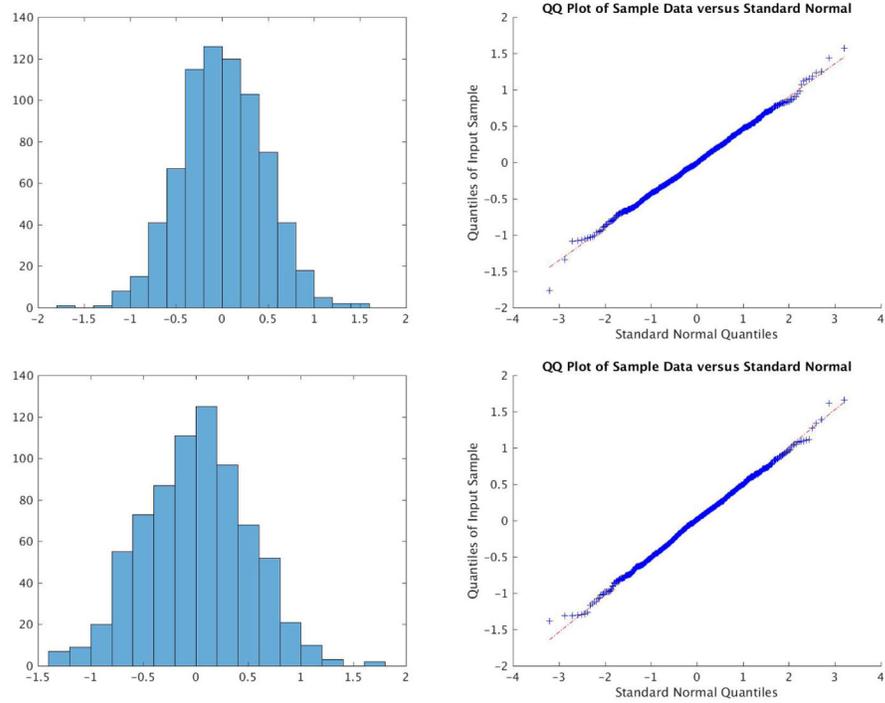


Fig. 11. Histograms and QQ plots of the posterior means of the random effects for two random selected ROIs over all subjects in the analysis of ADNI data set with the top 10 genes.

Table 1

Top-ranked ROIs based on the diagonal of $B_{bin}^T B_{bin}$ and columns of V based on the top 10 genes.

$B_{int,std}^T B_{int,std}$	V_1	V_2	V_3
sup.t.gy.R	mid.t.gy.R	caud.neuc.L	prec.L
sup.t.gy.L	mid.t.gy.L	thal.L	per.cort.L
mid.t.gy.L	sup.t.gy.R	caud.neuc.R	tmp.pl.L
mid.t.gy.R	hiopp.R	glob.pal.L	per.cort.R
inf.t.gy.L	amyg.R	post.limb.R	ent.cort.R
parah.gy.L	unc.L	putamen.L	mid.f.gy.R
inf.f.gy.R	unc.R	nuc.acc.R	ling.gy.R
lat.f-o.gy.L	inf.t.gy.R	post.limb.L	sup.f.gy.L
inf.f.gy.L	sup.t.gy.L	insula.R	pec.R
unc.R	inf.t.gy.L	subtha.nuc.L	sup.p.lb.L

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Top-ranked ROIs based on the diagonal of $B_{bin}^T B_{bin}$ and columns of V based on the top 40 genes.

$B_{int,std}^T B_{int,std}$	V_1	V_2	V_3
insula.R	hiopp.R	per.cort.R	glob.pal.R
insula.L	hiopp.L	sup.p.lb.L	post.limb.R
amyg.R	inf.t.gy.R	per.cort.L	thal.L
hiopp.L	sup.t.gy.R	tmp.pl.R	caud.neuc.R
sup.t.gy.R	unc.L	sup.p.lb.R	caud.neuc.L
unc.L	mid.t.gy.R	me.f.gy.L	nuc.acc.R
unc.R	insula.R	lat.ve.R	ant.caps.R
mid.t.gy.L	amyg.R	pstc.gy.L	glob.pal.L
hiopp.R	mid.t.gy.L	prec.L	subtha.nuc.L
mid.t.gy.R	inf.o.gy.L	tmp.pl.L	subtha.nuc.R