This article was downloaded by: [Fondren Library, Rice University]
On: 30 September 2013, At: 06:43
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

An Integrative Bayesian Modeling Approach to Imaging Genetics

Francesco C. Stingo [a] , Michele Guindani [a] , Marina Vannucci [b] & Vince D. Calhoun [c]
[a] The University of Texas MD Anderson Cancer Center , Houston , TX , 77230
[b] Department of Statistics , Rice University , Houston , TX , 77005
[c] Departments of Electrical and Computer Engineering (primary), Neurosciences, Computer Science, and Psychiatry , University of New Mexico , Albuquerque , NM , 87131
Accepted author version posted online: 09 Jun 2013.Published online: 27 Sep 2013.

To cite this article: Francesco C. Stingo , Michele Guindani , Marina Vannucci & Vince D. Calhoun (2013) An Integrative Bayesian Modeling Approach to Imaging Genetics, Journal of the American Statistical Association, 108:503, 876-891, DOI: 10.1080/01621459.2013.804409

To link to this article: http://dx.doi.org/10.1080/01621459.2013.804409

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

An Integrative Bayesian Modeling Approach to Imaging Genetics

Francesco C. STINGO, Michele GUINDANI, Marina VANNUCCI, and Vince D. CALHOUN

In this article we present a Bayesian hierarchical modeling approach for imaging genetics, where the interest lies in linking brain connectivity across multiple individuals to their genetic information. We have available data from a functional magnetic resonance imaging (fMRI) study on schizophrenia. Our goals are to identify brain regions of interest (ROIs) with discriminating activation patterns between schizophrenic patients and healthy controls, and to relate the ROIs' activations with available genetic information from single nucleotide polymorphisms (SNPs) on the subjects. For this task, we develop a hierarchical mixture model that includes several innovative characteristics: it incorporates the selection of ROIs that discriminate the subjects into separate groups; it allows the mixture components to depend on selected covariates; it includes prior models that capture structural dependencies among the ROIs. Applied to the schizophrenia dataset, the model leads to the simultaneous selection of a set of discriminatory ROIs and the relevant SNPs, together with the reconstruction of the correlation structure of the selected regions. To the best of our knowledge, our work represents the first attempt at a rigorous modeling strategy for imaging genetics data that incorporates all such features.

KEY WORDS: Bayesian hierarchical model; Functional magnetic resonance imaging; Markov random field; Neuroimaging; Single-nucleotide polymorphism; Variable selection.

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

1. INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a common tool for detecting changes in neuronal activity. It measures blood oxygenation level-dependent (BOLD) contrast that depends on changes in the regional cerebral blood-flow (rCBF). fMRI has become very popular in the neuroimaging field due to its relatively low invasiveness, absence of radiation exposure, and relatively wide applicability.

Statistical methods play an important role in the analysis of fMRI data and have generated a growing literature (see e.g., Lazar 2008; Lindquist 2008, for reviews on methods). Dimension reduction techniques, such as principal component analysis (PCA) and independent component analysis (ICA), and clustering algorithms are routinely applied to imaging data as a way of mapping connectivity. In the fMRI literature, connectivity refers to parts of the brain that show similarities and/or that interact with each other. In particular, anatomical connectivity deals with how different brain regions are physically connected; functional connectivity is defined as the association, or correlation, between fMRI time series of distinct voxels or regions; while effective connectivity is the directed influence of one brain region on others (Friston 1994). Because of the high-dimensionality of the data, that is, the large number of voxels, studies often perform region-based analyses, by looking at specific regions of interest (ROIs), or by dividing the entire brain into ROIs, for example, by parcellating the brain into anatomical regions (see, e.g., Tzourio-Mazoyer et al. 2002). Bayesian approaches have recently found successful applications in the fMRI field (Smith et al. 2003; Woolrich et al. 2004, 2009; Penny, Trujillo-Barreto, and Friston 2005; Bowman et al. 2008; Guo, Bowman, and Kilts 2008). Also, nonparametric Bayesian methods that cluster brain regions on the basis of their connectivity patterns have been proposed (Jbabdi, Woolrich, and Behrens 2009). Compared to other inferential approaches and algorithmic procedures commonly used in the analysis of fMRI data, Bayesian methods allow for a direct assessment of the uncertainties in the parameters' estimates and, perhaps more importantly, the incorporation of prior knowledge, such as on spatial correlation among voxels and/or ROIs, into the model.

In this article we consider a problem of *imaging genetics*, where structural and functional neuroimaging is applied to study subjects carrying genetic risk variants that relate to a psychiatric disorder. Our work, in particular, is motivated by a dataset on subjects diagnosed with schizophrenia and healthy controls, collected as part of the Mind Clinical Imaging Consortium (MCIC; Chen et al. 2012), where we have available measurements on fMRI scans and single nucleotide polymorphism (SNP) allele frequencies on all participants. The fMRI data were collected during a sensorimotor task, a block-design motor response to auditory stimulation. The resulting images were realigned, normalized, and spatially smoothed as customary, to remove most nontask-related sources of variability from the data (Ashby 2011; Chen et al. 2012). Here, the fMRI information is further summarized in individual contrast images of ROI-based summary statistics, as follows. For each participant, we fit a multiple regression incorporating regressors of the stimulus and its temporal derivative plus an intercept term. The resulting coefficient estimates can be used to build individual synthetic brain maps capturing the stimulus effect at each voxel. The maps are then superimposed to a fixed template atlas, for example, the MNI space automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al. 2002), and additionally segmented

Francesco C. Stingo, The University of Texas MD Anderson Cancer Center, Houston, TX 77230 (E-mail: *FStingo@mdanderson.org*). Michele Guindani, The University of Texas MD Anderson Cancer Center, Houston, TX 77230 (E-mail: *mguindani@mdanderson.org*). Marina Vannucci is Professor, Department of Statistics, Rice University, Houston, TX 77005 (E-mail: *marina@rice.edu*). Vince D. Calhoun, Departments of Electrical and Computer Engineering (primary), Neurosciences, Computer Science, and Psychiatry, University of New Mexico, Albuquerque, NM 87131 (E-mail: *vcalhoun@unm.edu*). Marina Vannucci's research was partially funded by NIH/NHLBI P01-HL082798 and NSF/DMS 1007871. Vince Calhoun's research was partially funded by NIH R01EB005846 and 5P20RR021938.

© 2013 American Statistical Association
Journal of the American Statistical Association
September 2013, Vol. 108, No. 503, Applications and Case Studies
DOI: 10.1080/01621459.2013.804409

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

into automatically labeled ROIs. Summary statistics, as the median or maximum intensity values, can be computed over the voxels included in each region (Kim et al. 2005; Etzel, Gazzola, and Keysers 2009). In addition to the fMRI data, we also have available measurements on SNP allele frequencies on all subjects. We use the genetic data as potential covariates that may affect brain function, in both healthy controls and patients with schizophrenia.

The goal of our analysis is to detect local activations, that is, areas of increased BOLD signal in response to a stimulus, by selecting a subset of ROIs that explain the observed activation patterns, and to identify SNPs that might be relevant to explain such activations. Understanding how connectivity varies between schizophrenic and control subjects is of utmost importance for diagnostic purposes and therapeutic interventions. The ability to link the imaging and genetic components in the participants' subgroups could lead to improved diagnostic tools and therapies, as it is now generally recognized that brain connectivity is affected by genetic characteristics (Colantuoni et al. 2008; Liu et al. 2009). Indeed, Chen et al. (2012) found evidence of a significant correlation between fMRI and SNP data. These authors use a two-step procedure involving a combination of PCA and ICA techniques, whereas we propose to study the association between spatial patterns and genetic variations across individuals within the coherent probabilistic framework offered by Bayesian hierarchical models.

We develop a hierarchical mixture model that includes several innovative characteristics. First, it incorporates the selection of ROIs that discriminate the subjects between schizophrenic patients and healthy controls, allowing for a direct assessment of the uncertainties in the estimates of the model selection parameters. Second, it allows the group-specific distributions to depend on selected covariates, that is, the SNP data. In this sense, our proposed model is integrative, in that it combines the observed brain activation patterns with the subjects' specific genetic information. Third, it incorporates prior knowledge via network models that capture known dependencies among the ROIs. More specifically, it employs spatially defined selection process priors that capture available knowledge on connectivity among regions of the brain, so that regions having the same activation patterns are more likely to be selected together. Furthermore, our hierarchical formulation accounts for additional correlation among selected ROIs that may not be captured by the network prior. Applied to the schizophrenia dataset, the model allows the simultaneous selection of a set of discriminatory ROIs and the relevant SNPs, together with the reconstruction of the dependence structure of the selected regions. To the best of our knowledge, our work represents the first attempt at a rigorous modeling strategy for imaging genetics data that incorporates all such characteristics.

The remainder of the article is organized as follows: In Section 2, we introduce our modeling framework and its major components. We describe posterior inference and prediction in Section 3. In Section 4, we first assess performances of our proposed model on simulated data and then investigate results on data from our case study on schizophrenia. We conclude the article with some remarks in Section 5.

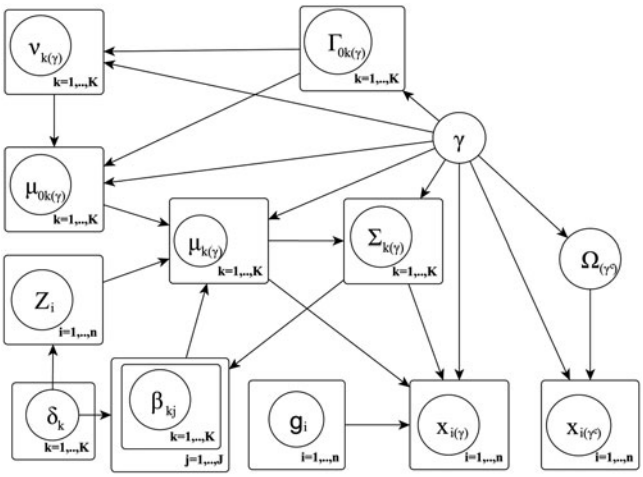

Figure 1. Graphical formulation of the proposed probabilistic model, as described in Section 2.

2. HIERARCHICAL BAYESIAN MODEL

In its most general formulation, we depict our proposed model as a hierarchical mixture model including selection of discriminating features (e.g., ROIs), mixture components that depend on selected covariates (e.g., SNP), and network priors that capture structural dependencies among the features. The graphical formulation of the model is illustrated in Figure 1 and its major components are described below. We also summarize the hierarchical formulation of our full model in Figure 2 at the end of this section.

2.1 A Mixture Model With Feature Selection

We represent the ROI-based summaries of BOLD signal intensity as measurements $\{x_{ij}, i = 1, \ldots, n, j = 1, \ldots, p\}$ on a set of p features (the anatomical ROIs) on n subjects. We envision that some of the features could discriminate the n subjects into, say, K separate groups (e.g., schizophrenia cases and healthy controls). Therefore, we introduce a latent binary vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)$ such that $\gamma_j = 1$ if the jth feature is discriminatory and $\gamma_j = 0$ otherwise. By employing a discriminant analysis framework, we model the data as a mixture model of the general type

$$f_k(x_{ij}|\gamma_j) = (1 - \gamma_j)\, f_0(x_{ij}; \theta_{0j}) + \gamma_j\, f(x_{ij}; \theta_{kj}), \qquad (1)$$

$k = 1, \ldots, K$, where $f_0(x_{ij}; \theta_{0j})$ describes the distribution of the "null" model for the nondiscriminatory features, while $f(x_{ij}; \theta_{kj})$ is the distribution of the measurements on the discriminatory features for subjects in group k. In this article, we assume Gaussian distributions for the mixture components, that is, $f_0(x_{ij}; \theta_{0j}) = N(0, \sigma^2_{0j})$ and $f(x_{ij}; \theta_{kj}) = N(\mu_{kj}, \sigma^2_{kj})$. Without loss of generality, we also assume that the measurement data are centered across all subjects within each feature.

Our model formulation follows an approach to feature selection in mixture models that was introduced by Tadesse, Sha, and Vannucci (2005) and Raftery and Dean (2006). These authors formulated clustering in terms of a finite mixture of Gaussian distributions with an unknown number of components and

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

then introduced latent variables to identify discriminating features. Kim, Tadesse, and Vannucci (2006) proposed an alternative modeling approach that uses infinite mixture models via Dirichlet process priors. Also, Hoff (2006) adopted a mixture of Gaussian distributions where different clusters were identified by mean shifts and where discriminating features were identified via the calculation of Bayes factors. Stingo, Vannucci, and Downey (2012) extended the approach by Tadesse, Sha, and Vannucci (2005) and Raftery and Dean (2006) to the discriminant analysis framework. Building a feature selection mechanism into mixture models is a more challenging problem than in general linear settings, where the latent indicator vector $\boldsymbol{\gamma}$ is used to induce mixture priors on the regression coefficients. In mixture models, only the observed elements of the matrix $\mathbf{X}$ guide the selection and $\boldsymbol{\gamma}$ is used to index the contribution of the different features to the likelihood terms of the model, as in our formulation (1).

Let us now denote features indexed by $\gamma_j = 1$ as $\mathbf{X}_{(\gamma)}$, and those indexed by $\gamma_j = 0$ as $\mathbf{X}_{(\gamma^c)}$. While the former set defines a mixture distribution across the n samples, the latter favors one multivariate normal distribution across all samples. Following the finite mixture model formulation by Tadesse, Sha, and Vannucci (2005), we can write our model for sample i as

$$\begin{aligned} \mathbf{x}_{i(\gamma^c)}|\cdot &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{(\gamma^c)}) \\ \mathbf{x}_{i(\gamma)}|g_i = k, \cdot &\sim \mathcal{N}(\boldsymbol{\mu}_{k(\gamma)}, \boldsymbol{\Sigma}_{k(\gamma)}), \end{aligned} \tag{2}$$

with $g_i = k$ if the ith sample belongs to group k. In the supervised setting, also known as discriminant analysis, in addition to the observed vectors $\mathbf{x}_i$'s, the number of groups K and the classification labels g_i's are also available and the aim is to derive a classification rule that will assign further cases to their correct groups (see Section 3.1). Here we assume diagonal variance-covariance matrices, that is, $\boldsymbol{\Sigma}_{k(\gamma)} = \text{Diag}(\sigma^2_{k1}, \ldots, \sigma^2_{kp_\gamma})$, with p_γ the number of nonzero elements in the vector $\boldsymbol{\gamma}$, and $\boldsymbol{\Omega}_{(\gamma^c)} = \text{Diag}(\sigma^2_{01}, \ldots, \sigma^2_{0(p-p_\gamma)})$, and then impose inverse-gamma priors on the variance components, $\sigma^2_{kj} \sim \text{IG}(\sigma^2_{kj}; a_k, b_k)$, $k = 0, 1, \ldots, K$. Even though we make this simplifying assumption at this stage of the hierarchy, our proposed model is still able to capture structural dependencies via the specification of the prior model for the mean components that we describe in Section 2.3.2.

2.2 Covariate-Dependent Characterization of the Mixture Components

We allow the mixture components of model (2) to depend on a set of covariates. Let us denote with $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iR})^T$ the set of available covariates for the ith individual. We model the means of the discriminating components as subject-specific parameters

$$\boldsymbol{\mu}_{ik(\gamma)} = \boldsymbol{\mu}_{0k(\gamma)} + \boldsymbol{\beta}^T_{k(\gamma)} \mathbf{Z}_i, \quad k = 1, \ldots, K, \tag{3}$$

where $\boldsymbol{\mu}_{0k(\gamma)}$ is a baseline process that captures brain connectivity (described in detail in Section 2.3.2) and $\boldsymbol{\beta}_{k(\gamma)}$ is a $R \times p_\gamma$ matrix of coefficients describing the effect of the covariates on the observed measurements. More in detail, our model formulation uses component-specific parameters that determine how covariates, and other relevant spatial characteristics, affect the observed measurements $\mathbf{x}_{i(\gamma)}$, on the n subjects, given the selected features. In this respect, the classification of the n subjects in K groups is driven by the subjects' covariates. In particular, in our application to the schizophrenia dataset the model relates the subjects' brain activity to available information on genetic covariates.

We want to allow different covariates to affect the individual mixture components. For this we introduce spike and slab priors on $\boldsymbol{\beta}_{k(\gamma)}$. First, we define a binary latent indicator δ_{rk}, $r = 1, \ldots, R$, such that

$$\delta_{rk} = \begin{cases} 1 & \text{with probability } w_{rk} \\ 0 & \text{with probability } 1 - w_{rk}. \end{cases}$$

If $\delta_{rk} = 1$, then the rth covariate is considered relevant to explain the observed measurements in the kth mixture component. In this case, we allow the corresponding vector $\boldsymbol{\beta}_{rk(\gamma)}$ to be sampled from a multivariate normal prior distribution. Otherwise, the rth covariate does not affect the response data and the corresponding regression coefficient vector $\boldsymbol{\beta}_{rk(\gamma)}$ is set to 0 for component k. We choose a conjugate setting and write the prior on the rth row vector $\boldsymbol{\beta}_{rk(\gamma)}$, for the kth component, as

$$\boldsymbol{\beta}_{rk(\gamma)} \sim (1 - \delta_{rk})\mathcal{I}_0(\boldsymbol{\beta}_{rk(\gamma)}) + \delta_{rk}\mathcal{N}(\mathbf{b}_{0k(\gamma)}, h\boldsymbol{\Sigma}_{k(\gamma)}), \tag{4}$$

where $\mathcal{I}_0(\boldsymbol{\beta}_{rk(\gamma)})$ is a vector of point masses at zero and $h > 0$ some suitably large-scale parameter to be chosen. Large values of h correspond to a prior well spread out over the parameters space, and typically encourage the selection of relatively large effects. We refer to the articles by Smith and Kohn (1996), Chipman et al. (2001), and O'Hara and Sillanpaa (2009) for discussions on the choice of this parameter. A Bernoulli prior on δ_{rk}, with parameter w_{rk}, completes our selection prior model on the covariates,

$$P(\boldsymbol{\delta}|\mathbf{w}) = \prod_{k=1}^{K}\prod_{r=1}^{R} w_{rk}^{\delta_{rk}}(1 - w_{rk})^{1-\delta_{rk}}.$$

In our applications, we fix the hyperparameters w_{rk}. Alternatively, one can place a Beta hyperprior on these parameters.

Notice that our setting allows individual covariates to have differential effects $(\boldsymbol{\beta}_{r1(\gamma)}, \ldots, \boldsymbol{\beta}_{rK(\gamma)})$ on the selected features. In the application to the schizophrenia dataset, our integrative model allows the selection of SNPs that are implicated in the differential activation patterns observed in patients and healthy controls. Thus, a SNP can be correlated to a discriminatory ROI for subjects in one group, and hence help understanding the activation patterns for that group, and not be correlated to the same discriminatory ROI in the other group. This may happen if the selected ROI shows heterogeneous activation levels, explained by the selected SNP, in one group but not in the other. Our model also allows a SNP to be associated to differential activations in both groups.

2.3 Networks for Structural Dependencies

We still need to specify priors on the feature selection vector $\boldsymbol{\gamma}$ and on the baseline process $\boldsymbol{\mu}_{0k(\gamma)}$. For those parameters, we employ prior models that capture information on dependence structures.

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

2.3.1 A Markov Random Field for Feature Selection. The selection of the discriminatory features in our model (2) is driven by the latent variables γ_j's. Here we impose a spatial process on γ_j that takes into account network dependencies among the features, so that dependent features are more likely to be selected together. The network structure is determined a priori on the basis of available information. In particular, in our application to the schizophrenia dataset knowledge is available on the anatomical arrangement of the ROIs. We therefore construct a symmetric matrix S that captures spatial connectivity, with element $\{s_{ij} = 1\}$ if ROI i and ROI j are directly connected in the network and with $\{s_{ij} = 0\}$ otherwise. Although we do not pursue this option here, we notice that one could extend the definition of the matrix S to allow for entries $s_{ij} \in \mathbb{R}^+$, that is, continuous weights, to incorporate knowledge on physical distance between features, similarly to what is routinely done in "auto-logistic models" (Besag 1974) and spatial conditional (CAR) and simultaneous (SAR) autoregressive models (Banerjee, Gelfand, and Carlin 2003).

We model spatial dependencies via an Ising prior model on the γ_j's. This is a type of Markov random field (MRF), where the distribution of a set of random variables follows Markov properties that can be described by an undirected graph. Smith et al. (2003) and Penny, Trujillo-Barreto, and Friston (2005) first used MRF models as prior distributions for the coefficients of voxel-based generalized linear models for fMRI data. Ising models have also recently found useful applications in genomics, to capture regulatory networks (see e.g., Li and Zhang 2010; Stingo et al. 2011; Telesca et al. 2012). In a MRF, variables are represented by nodes and relations between them by edges. The Ising model is characterized by the following probabilities

$$P(\gamma_j|\gamma_i, i \in N_j) = \frac{\exp(\gamma_j F(\gamma_j))}{1 + \exp(F(\gamma_j))}, \tag{5}$$

where $F(\gamma_j) = e + f \sum_{i \in N_j}(2\gamma_i - 1)$ and N_j is the set of direct neighbors of feature j in the network. The parameter e controls the sparsity of the model, while higher values of f encourage neighboring features to take on the same γ_j value. The global MRF distribution for $\boldsymbol{\gamma}$ can be defined up to its normalizing constant as

$$P(\boldsymbol{\gamma}|e, f) \propto \exp\left(e\mathbf{1}_p^T\boldsymbol{\gamma} + f\boldsymbol{\gamma}^T\mathbf{S}\boldsymbol{\gamma}\right), \tag{6}$$

where $\mathbf{1}_p$ is the unit vector of dimension p and $\mathbf{S}$ is the matrix capturing a priori likely connections among the nodes, as described above. Note that if a feature does not have any neighbor, then its prior distribution reduces to an independent Bernoulli, with parameter $\exp(e)/[1 + \exp(e)]$, a prior often adopted in the Bayesian variable selection literature.

Following the articles by Li and Zhang (2010) and Stingo, Vannucci, and Downey (2012), we treat e and f as fixed hyperparameters. Although the parameterization is somewhat arbitrary, some care is needed in deciding whether to put a prior distribution on f. In particular, allowing f to vary can lead to a phase transition problem, that is, the expected number of variables equal to 1 can increase massively for small increments of f. This problem can happen because Equation (5) can only increase as a function of the number of γ_i's equal to 1. In variable selection, especially when p is large, phase transition leads to a drastic change in the proportion of included variables. An empirical estimate of the phase transition value can be obtained using the algorithm proposed by Propp and Wilson (1996) and the values of e and f can then be chosen accordingly. In this article, we first set e to a small value that reflects our belief in a sparse model. This value can be chosen based on the logistic transformation of e in the case of no neighbors, as a lower bound of the prior probability of feature selection. As for f, any value of f below the phase transition point can be considered a good choice, with values closer to the phase transition point leading to higher prior probabilities of selection for those nodes whose neighbors are already selected, particularly in a sparse network.

2.3.2 Component-Specific Dependencies. We complete our model with a prior specification on the baseline process $\boldsymbol{\mu}_{0k(\gamma)}$ in model (3). This component is a random effect capturing relevant characteristics of the selected features that affect the observed measurements $\mathbf{x}_{i(\gamma)}$, on the n subjects, and that are not explained by the effects of the selected covariates. In our application to the schizophrenic dataset, we use this baseline process to capture additional effects on activations among selected ROIs that are not explained by the genetic covariates. In particular, we look at $\boldsymbol{\mu}_{0k(\gamma)}$ as a way to capture general relationships among selected features. Correlation among distant ROIs is in fact a well-described phenomenon in the fMRI literature, in addition to the dependence based on proximity, which, in our model, is captured by the MRF prior on $\boldsymbol{\gamma}$ (Friston 1994).

We assume $\boldsymbol{\mu}_{0k(\gamma)}$ a realization of a multivariate normal distribution,

$$\boldsymbol{\mu}_{0k(\gamma)} \sim N_{p_\gamma}(\boldsymbol{\nu}_{k(\gamma)}, h_1\boldsymbol{\Gamma}_{0k(\gamma)}), \quad k = 1, \ldots, K, \tag{7}$$

for some fixed-scale parameter $h_1 > 0$, and assume the precision matrix $\boldsymbol{\Gamma}^{-1}_{0k(\gamma)}$ to come from a Wishart distribution, that is, $\Gamma_{0k(\gamma)}|\delta_k, \mathbf{Q} \sim IW(d_k, \mathbf{Q})$. Alternatively, a more general scaled inverse Wishart can be used (Gelman and Hill 2007, sec. 13.3). Notice how we define, again, component-specific parameters. This allows us to estimate component-specific dependence structures. A normal distribution on $\boldsymbol{\nu}_{k(\gamma)}$ completes the prior specification.

Figure 2 summarizes the hierarchical formulation of our full model.

3. POSTERIOR INFERENCE

For posterior inference, our primary interest is in the selection of the discriminating features and of the covariates that affect the observed measurements, as captured by the selection vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$. We also want to do inference on the dependence structure among the selected features. Here we design a MCMC algorithm that explores the model space for configurations with high posterior probability. These algorithms are commonly used in Bayesian variable selection settings, and have been successfully employed in genomic applications with very large number of variables (George and McCulloch 1993, 1997; Brown, Vannucci, and Fearn 1998; Tadesse, Sha, and Vannucci 2005; Stingo et al. 2011). Clearly, with large dimensions, exploring the posterior space is a challenging problem. A typical strategy relies on exploiting the sparsity of the model, that is, the belief that most of the variables are not related to the underlying biological process. Such algorithm allows then to explore the posterior space in an effective way, quickly finding the most probable

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

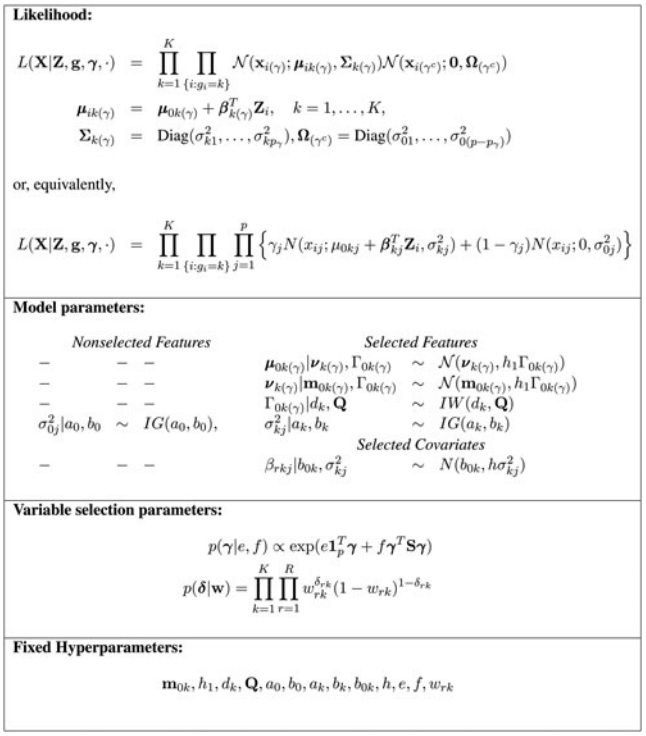

Figure 2. Hierarchical formulation of the proposed probabilistic model. $\mathbf{X}$ denotes the $n \times p$ matrix of ROI-based measurements, $\mathbf{Z}$ is the $n \times R$ matrix of covariates, $\mathbf{g}$ is the vector of indicators of the individual group memberships, and $\boldsymbol{\gamma}$ is the latent binary vector of discriminatory features.

configurations, that is, those corresponding to sets of variables with high posterior probability, while spending less time in regions with low posterior probability.

As it is typical in the implementation of MCMC methods for variable selection, to speed up the convergence of the MCMC we integrate out some of the model parameters from the posterior distribution, specifically the variance components in model (2), the regression coefficients in model (3) for selected covariates, and the hyperparameters of the distribution (7) of the model intercepts. Therefore, we focus on the marginal posterior distribution of $(\boldsymbol{\gamma}, \boldsymbol{\delta}_k, \boldsymbol{\mu}_{0k(\gamma)})$. We detail the MCMC implementation in the Appendix. Here, we succinctly describe our MCMC algorithm by the following three steps:

1. A Metropolis-Hastings step on $\boldsymbol{\gamma}$: This step is based on the marginal posterior distribution of $\boldsymbol{\gamma}$ conditioned upon $\boldsymbol{\delta}$ and the $\boldsymbol{\mu}_{0k(\gamma)}$'s. It consists of randomly choosing between changing the value of a single γ_j, from 0 to 1 or from 1 to 0, and swapping two γ_j's (with opposite values). This step may imply a change in the dimensionality of the matrices $\boldsymbol{\Gamma}_{0k(\gamma)}$, which have been integrated out, while the parameter space of $\boldsymbol{\gamma}$, that is, $\{0,1\}^p$, remains constant.
2. A Metropolis-Hastings step for $\boldsymbol{\delta}_k$: Similarly to the previous step, for each of the K components we randomly choose between changing the value of a single δ_{rk}, from 0 to 1 or from 1 to 0, and swapping two δ_{rk}'s with opposite values.
3. A random walk Metropolis-Hastings step on the $\boldsymbol{\mu}_{0k(\gamma)}$'s: Given the selected features, we update $\boldsymbol{\mu}_{0k(\gamma)}$ by proposing new values for all μ_{0kj} included in the model at that particular iteration, according to $\mu_{0kj}^{\text{New}} = \mu_{0kj}^{\text{Old}} + \epsilon$, with $\epsilon \sim N(0, v^2)$ for $j = 1,\ldots,p_\gamma$ and $k = 1,\ldots K$, for some fixed choice of the proposal parameter v (for details on this step, see the Appendix).

Posterior inference can be performed based on the MCMC output by calculating the marginal posterior probabilities of the individual γ_j's and δ_{rk}'s. A simple strategy is to compute Monte Carlo estimates by counting the number of appearances of each feature/covariate across the visited models. Important features and related covariates can be selected as those with highest posterior probabilities. Samples from the posterior distribution of $\boldsymbol{\mu}_{0k(\gamma)}$ can then be used to infer correlation between selected features for each of the K groups.

Our model also allows inference on the variance components in model (2) and the regression coefficients of the selected covariates in model (4). More specifically, post-MCMC estimates of those parameters can be obtained by sampling $(\boldsymbol{\beta}_{kj(\delta_k)}, \sigma_{kj}^2)$ jointly from

$$p\left(\sigma_{kj}^2|\mathbf{X},\mathbf{Z},\boldsymbol{\gamma},\boldsymbol{\delta}_k,\boldsymbol{\mu}_{0k(\gamma)}\right) \sim \mathrm{IG}(a_k', b_k'),$$

with $a_k' = a_k + n_k/2$ and

$$b_k' = b_k + \left(\mathbf{x}_{jk} - \mathbf{1}_{n_k}\mu_{0jk}\right)^T\left(h^{-1}\mathbf{I}_{n_k} + \mathbf{Z}_{k(\delta_k)}\mathbf{Z}_{k(\delta_k)}^T\right)^{-1} \times \left(\mathbf{x}_{jk} - \mathbf{1}_{n_k}\mu_{0jk}\right)/2,$$

and where n_k is the frequency of the kth group. Then,

$$p\left(\boldsymbol{\beta}_{kj(\delta_k)}|\mathbf{X},\mathbf{Z},\boldsymbol{\gamma},\boldsymbol{\delta}_k,\boldsymbol{\mu}_{0k(\gamma)},\sigma_{kj}^2\right) \sim N\left(\boldsymbol{V}_k\,\mathbf{Z}_{k(\delta_k)}^T(\mathbf{x}_{jk} - \mathbf{1}_n\mu_{0kj}), \sigma_{kj}^2\boldsymbol{V}_k\right),$$

with $\boldsymbol{V}_k = (\mathbf{Z}_{k(\delta_k)}^T\mathbf{Z}_{k(\delta_k)} + h^{-1}\mathbf{I}_{|\delta_k|})^{-1}$ and $|\boldsymbol{\delta}_k| = \sum_{r=1}^{R}\delta_{rk}$. The μ_{0kj}'s can be either set to a fixed value, like the posterior Monte Carlo mean, or integrated out in a Monte Carlo fashion via composition sampling (see Banerjee, Gelfand, and Carlin 2003). Uncertainty on these parameters can be assessed via Bayesian credible intervals (CIs). Notice that, since our setting allows individual covariates to have differential effects on the selected features, some of those effects may not be significant in the final inference (e.g., if the 95% Bayesian CIs of the corresponding β coefficients contain 0).

3.1 Prediction

Our modeling formulation allows prediction of new samples based on the selected features and covariates. The predictive distribution of a new observation $\mathbf{x}^f|\mathbf{z}^f$ is used to classify the new sample into one of the K possible groups. Here $\mathbf{x}^f$ is the $p \times 1$ vector of the observed features and $\mathbf{z}^f$ is the $R \times 1$ vector of the observed covariates for a new sample. Given a subset of selected ROIs, indicated by $\boldsymbol{\gamma}^*$, and component-specific subsets of selected SNP, indicated by $\boldsymbol{\delta}_k^*$'s, the predictive distribution of the new observation $\mathbf{x}_{(\gamma^*)}^f|\mathbf{z}_{(\delta_k^*)}^f$ for class k is

$$p_k\left(\mathbf{x}_{(\gamma^*)}^f\middle|\mathbf{z}_{(\delta_k^*)}^f\right) = \frac{1}{(2\pi)^{n_k/2}}|\mathbf{V}_k^*|^{-\frac{1}{2}}\left|\mathbf{z}_{(\delta_k^*)}^{fT}\mathbf{z}_{(\delta_k^*)}^f + \mathbf{V}_k^{*-1}\right|^{-\frac{1}{2}}\frac{(b_k')^{a_k'}}{\left(b_k^f\right)^{a_k^f}}\frac{\Gamma\left(a_k^f\right)}{\Gamma(a_k')}, \quad (8)$$

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

where

$$
\begin{aligned}
a_k^f &= a_k' + 1/2,\\
\mathbf{V}_k^* &= \left(\mathbf{Z}_{k(\delta_k^*)}^T \mathbf{Z}_{k(\delta_k^*)} + h^{-1}\mathbf{I}_{|\delta_k|}\right)^{-1},\\
b_k^f &= b_k + (\mathbf{x}_{jk(\gamma^*)} - \mathbf{1}_{n_k}\mu_{0jk})^T \mathbf{Z}_{k(\delta_k^*)}\mathbf{V}_k^*\\
&\quad \times \left(\mathbf{V}_k^{-1} - \mathbf{V}_k^{-1}\left(\mathbf{Z}_{k(\delta_k^*)}^T \mathbf{Z}_{k(\delta_k^*)} + \mathbf{V}_k^{-1}\right)^{-1}\mathbf{V}_k^{-1}\right)\\
&\quad \times \mathbf{V}_k^* \mathbf{Z}_{k(\delta_k^*)}^T \left(\mathbf{x}_{jk(\gamma^*)} - \mathbf{1}_{n_k}\mu_{0jk}\right)/2 + \frac{\left(\mathbf{x}_{jk(\gamma^*)}^f - \mu_{0jk}\right)^2}{2\left(1 + \mathbf{z}_{(\delta_k^*)}^f \mathbf{V}_k \mathbf{z}_{(\delta_k^*)}^{fT}\right)}.
\end{aligned}
$$

In model (8) the μ_{0kj}'s are assumed to be set to a fixed value, for example the posterior Monte Carlo mean or median. Alternatively, one can perform a Monte Carlo integration of the μ_{0kj}'s using the values sampled within the MCMC algorithm. In our experience, we have not noticed any significant difference between the two approaches.

The probability that a future observation, given the observed data, belongs to the group k is then

$$\pi_k(g^f|\mathbf{X}, \mathbf{Z}) = p(g^f = k|\mathbf{x}^f, \mathbf{X}, \mathbf{z}^f, \mathbf{Z}),$$

where g^f is the group indicator of the new observation. By estimating the probability $\pi_k = P(g_i = k)$ that one observation comes from group k as $\hat{\pi}_k = n_k/n$, the previous distribution can be written in closed form as

$$\pi_g(g^f|\mathbf{X}, \mathbf{Z}) = \frac{p_k\left(\mathbf{x}_{(\gamma^*)}^f|\mathbf{z}_{(\delta_k^*)}^f\right)\hat{\pi}_k}{\sum_{i=1}^K p_i\left(\mathbf{x}_{(\gamma^*)}^f|\mathbf{z}_{(\delta_k^*)}^f\right)\hat{\pi}_i}, \tag{9}$$

with $p_k(\mathbf{x}_{(\gamma^*)}^f|\mathbf{z}_{(\delta_k^*)}^f)$ the predictive distribution defined in model (8), and the new sample can be classified based on this distribution, for example by assigning it to the group that has the highest posterior probability. The plug-in estimate $\hat{\pi}_k = n_k/n$ can be formally justified as an approximation under a noninformative Dirichlet prior distribution for π and training data exchangeable with future data, meaning that observations from training and validation sets arise in the same proportions from the groups (see e.g., Fearn, Brown, and Besbeas 2002).

4. APPLICATIONS

4.1 Simulation Studies

We investigate the performance of our model using simulated data. We consider simulated scenarios mimicking the characteristics of the real data that motivated the development of the model, where some of the ROIs appear to be highly correlated. We focus on situations where most of the measurements are noisy and test the ability of our method to discover relevant features in the presence of a large amount of noise. The SNP data record the number of copies of the minor allele at each locus for each individual, as a sequence of {0,1,2} values indicating the three possible genotypes at each SNP: major allele homozygote, heterozygote, and minor allele homozygote (see e.g., Servin and Stephens 2007, for a description of this additive coding). In our simulations, we considered the same dataset used in the experimental data described in Section 4.2. This choice was made to preserve realistic patterns of correlation (also known as "linkage disequilibrium") across multiple SNPs. The simulation comprised a total of $R = 50$ covariates (SNP), only two of which were used to generate the measurements (activation profiles), as described below.

We generated a sample of 200 observations from a mixture of $K = 2$ multivariate normal densities, induced by four variables (features), as

$$
\begin{aligned}
\mathbf{x}_i \sim\ & I_{[1\le i\le n_1]}\mathcal{N}_4\left(\boldsymbol{\mu}_{01} + \mathbf{B}_1^T\mathbf{Z}_i, \boldsymbol{\Sigma}_1\right)\\
& + I_{[n_1 < i \le 200]}\mathcal{N}_4\left(\boldsymbol{\mu}_{02} + \mathbf{B}_2^T\mathbf{Z}_i, \boldsymbol{\Sigma}_2\right),
\end{aligned}
$$

with $\mathbf{x}_i = (x_{i1}, \ldots, x_{i4})^T$, for $i = 1, \ldots, 200$, and where $I_{[.]}$ is the indicator function. The first $n_1 = 150$ samples were from the first component of the mixture, the last $n_2 = 50$ from the second. We then randomly divided the samples into a training set of 100 observations and a validation set of the same size. We set the elements of the 4×1 vector $\boldsymbol{\mu}_{01}$ to 0.8 and those of $\boldsymbol{\mu}_{02}$ to -0.8. The (2×4) regression coefficient matrix $\mathbf{B}_k = (\boldsymbol{\beta}_{k1}, \ldots, \boldsymbol{\beta}_{k4})$ determines the effects of the true covariates on the simulated activation profiles. We set $\mathbf{B}_1 = 0.8 \cdot 1_{2\times 4}$ and $\mathbf{B}_2 = 0.8 \cdot 1_{2\times 4}$. The covariance structure among the relevant features was chosen by setting the off-diagonal elements of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ to 0.5 whereas the diagonal elements were set to 1. Note that the data-generating process differs from our proposed model, where the correlation structure among the features is modeled at a hidden level via the baseline component. In addition to the four relevant features, we generated 100 noisy ones from a multivariate normal distribution centered at zero, with variances equal to 1 and off-diagonal elements of the covariance matrix equal to 0.1.

Our simulation comprises a total of 104 features and 50 covariates. We aim at finding the four discriminating features and the two covariates that truly relate to the response measurements. We are also interested in capturing the correlation structure among selected features. Our full prior model and the related hyperparameters are summarized in Figure 1. We report the results obtained by choosing, when possible, hyperparameters that lead to weakly informative prior distributions. In particular, we specified the priors on σ_{0j}^2 and σ_{kj}^2 by setting $a_0 = a_k = 3$, the minimum integer value such that the variance is defined, and $b_0 = b_k = 0.1$. Using the same rationale, we set $d_k = 3$, the minimum value such that the expectation of $\boldsymbol{\Sigma}_k$ exists, and $\mathbf{Q} = c\,\mathbf{I}_p$ with $c = 0.1$. As for the β vectors of regression coefficients, we set the prior mean to $b_{01} = b_{02} = 0$. Similarly, we set $m_{10} = m_{20} = 0$. We then set h to 4 and h_1 to 1, to obtain fairly flat priors over the region where the data are defined; larger values of these hyperparameters would encourage the selection of only very large effects whereas smaller values would encourage the selection of smaller effects. As for the feature selection indicator, $\boldsymbol{\gamma}$, we set $e = -3$, which corresponds to setting the expected proportion of features a priori included in the model to 5% of the total number of available ones. In this simulation study, we did not use any network structure on the prior distribution of $\boldsymbol{\gamma}$, which correspond to set $f = 0$. This is also equivalent to assuming $p(\boldsymbol{\gamma})$ as a product of independent Bernoulli distributions with expected value equal to 0.05. Finally, regarding the prior on the covariate selection indicator $\boldsymbol{\delta}$, we set $w_{rk} = 0.05$, which corresponds to setting the proportion of covariates expected a priori in the model to 5%. Parameters e and w_r influence the sparsity of the model.

We ran three MCMC samplers for 30,000 iterations with the first 1000 discarded as burn-in. As starting points of the chain,

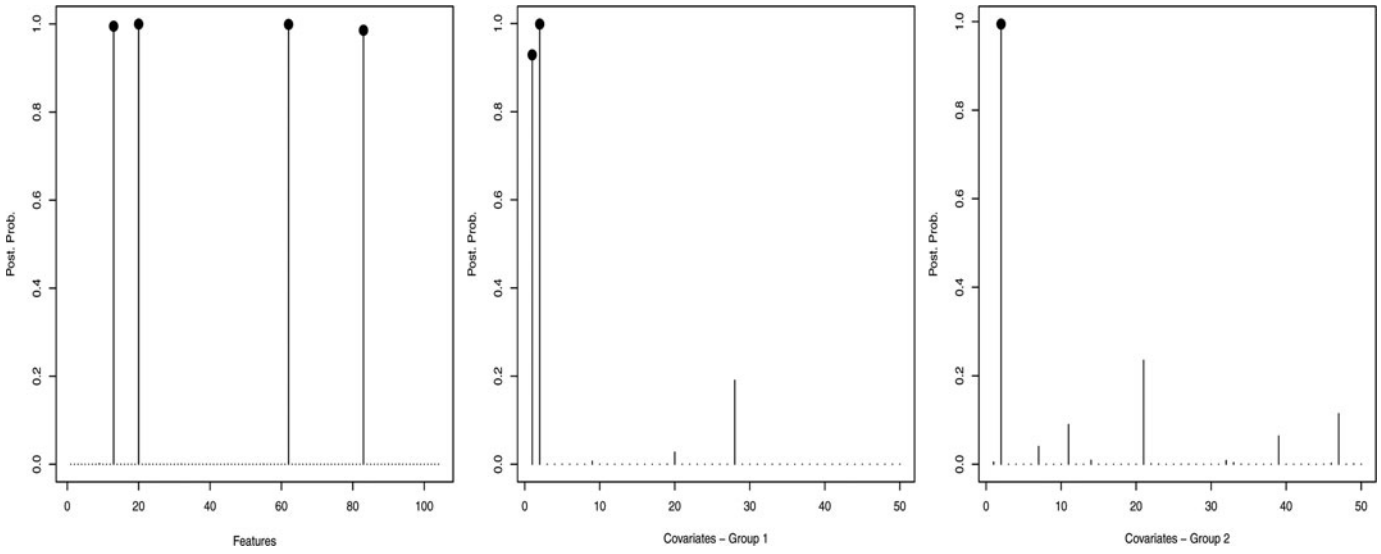

Figure 3. Simulation study: marginal posterior probabilities of inclusion for features (left) and covariates (middle and right).

we considered three different pairs of discriminatory ROIs and significant SNPs. More specifically, we assumed different starting counts of included ROIs and SNPs, that is, 2 ROIs and 2 SNPs, 10 ROIs and 5 SNPs, and 15 ROIs and 10 SNPs, respectively. The actual ROIs and SNPs initially included were randomly selected from the set of available features and covariates. To assess the agreement of the results between the two chains, we looked at the correlation between the marginal posterior probabilities for ROI selection, $p(\gamma_j|\mathbf{Z}, \mathbf{X})$, and SNP selection, $p(\delta_{lk}|\mathbf{Z}, \mathbf{X})$ for $k = 1, 2$. These indicated good concordance between the three MCMC chains, with pairwise correlation coefficients ranging from 0.999 to 1 for ROIs and from 0.997 to 0.998 ($k = 1$) and from 0.993 to 0.999 ($k = 2$) for SNPs. Concordance among the marginal posterior probabilities was confirmed by looking at scatterplots of the marginal probabilities across each pair of MCMC chains (figures not shown). We also used the Gelman and Rubin's convergence diagnostics (Gelman and Rubin 1992) to assess the convergence of the parameters μ_{0kj}'s to their posterior distributions. Those statistics were all below 1.1, ranging from 1.0051 to 1.0212, clearly indicating that the MCMC chains were run for a satisfactory number of iterations.

We then computed marginal posterior probabilities for feature and covariate selection, $p(\gamma_j = 1|\mathbf{Z}, \mathbf{X})$ and $p(\delta_{rk} = 1|\mathbf{Z}, \mathbf{X})$, respectively. These are displayed in Figure 3, for all 104 features and the 50 covariates. All four relevant features were correctly identified by our model, with high posterior probability (> 0.98), while noisy features showed posterior probabilities smaller than 1%. In addition, the two relevant covariates were also selected for the first group, with $p(\delta_{r1} = 1|\mathbf{Z}, \mathbf{X}) > 0.92$, and one relevant covariate was selected for the second group, with $p(\delta_{r2} = 1|\mathbf{Z}, \mathbf{X}) > 0.99$. Sensitivity analyses with different choices of e and w_{rk} indicated that the posterior inference is quite robust. Specifically, our selection results were practically invariant when we let e vary between -3.5 and -2, which is equivalent to an a priori expected proportion of included features between 3% and roughly 10%, and w_{rk} vary between 3% and 10%. Regarding the other hyperparameters of the model, we noticed that smaller values of h would encourage the inclusion of smaller effects and that setting this parameter to 1 would lead to the inclusion of a few false positive covariates.

Posterior distributions of the intercept parameters $\boldsymbol{\mu}_{01}$ and $\boldsymbol{\mu}_{02}$, given the selected features, indicated that parameters characterizing the two groups are clearly different. For example, Figure 4 shows the posterior distributions of the elements μ_{011} and μ_{021}, corresponding to the first selected feature. These distributions were obtained from the MCMC output conditioning upon the inclusion of the four relevant features. Similar plots were obtained for other selected features (not shown).

We estimated the dependence structure among selected features by looking at the posterior correlations among the corresponding intercepts, calculated based on the MCMC samples. These were

$$\operatorname{Corr}\boldsymbol{\mu}_{01} = \begin{pmatrix} 1.0000 & 0.5480 & 0.5916 & 0.4426 \\ 0.5480 & 1.0000 & 0.6075 & 0.4848 \\ 0.5916 & 0.6075 & 1.0000 & 0.4930 \\ 0.4426 & 0.4848 & 0.4930 & 1.0000 \end{pmatrix}$$

and

$$\operatorname{Corr}\boldsymbol{\mu}_{02} = \begin{pmatrix} 1.0000 & 0.5322 & 0.5740 & 0.5379 \\ 0.5322 & 1.0000 & 0.5075 & 0.4535 \\ 0.5740 & 0.5075 & 1.0000 & 0.4924 \\ 0.5379 & 0.4535 & 0.4924 & 1.0000 \end{pmatrix}$$

for Groups 1 and 2, respectively. These estimates show that our model is indeed able to capture correlation structure among selected features. Furthermore, we summarized the distance of the estimated correlation matrices from the true ones using the RV-coefficient, a measure of similarity between positive semidefinite matrices (Robert and Escoufier 1976; Smilde et al. 2009). The RV coefficient $R_V \in (0, 1)$ may be considered as a generalization of the Pearson's coefficient of determination. For both Groups 1 and 2, we got values very close to 1, that is, $R_V = 0.994$ for $\operatorname{Corr}\boldsymbol{\mu}_{01}$ and $R_V = 0.997$ for $\operatorname{Corr}\boldsymbol{\mu}_{02}$, suggesting good agreement with the true correlation structures.

Finally, we assessed predictive performance. We used the four selected features and the three selected covariates, 2 for the first

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

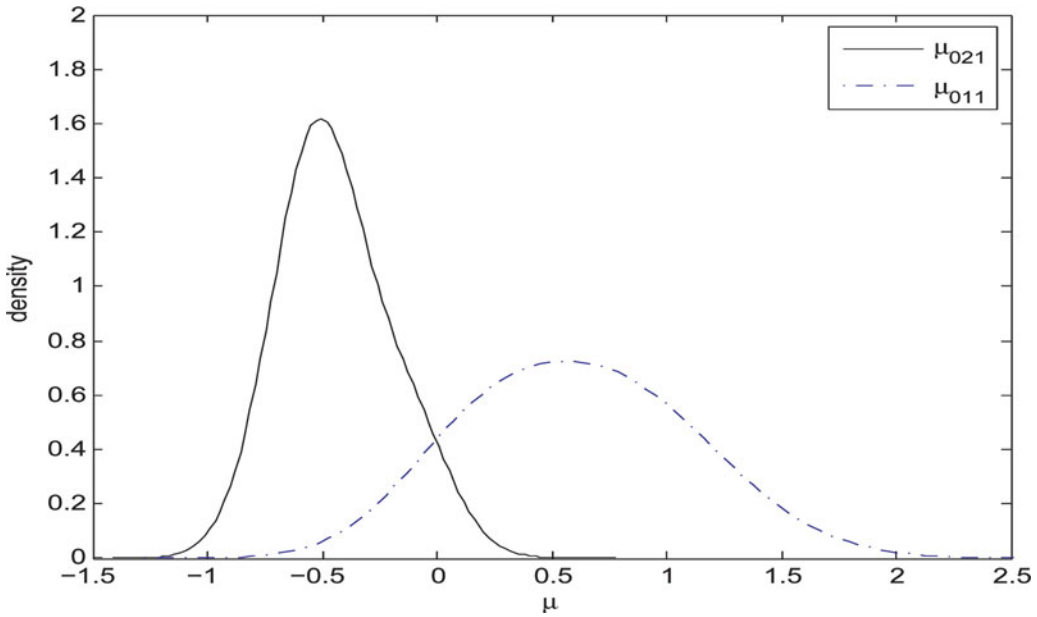

Figure 4. Simulation study: density Kernel estimate of the posterior distribution of μ_{011} and μ_{021}. The online version of this figure is in color.

group and 1 for the second one, and classified the observations in the validation test using formula (9). Our model led to a misclassification of 26% of the subjects. Figure 5 shows the posterior probabilities of membership in Group 1 for the 100 observations in the validation set, with the numbers 1 and 2 indicating the actual group membership. Units are sorted, from the highest to the lowest, within each group. Units within the boundaries of the dashed line are misclassified and provide a visual representation of the misclassification rate, whereas units above and below the solid line are correctly classified in Group 1 and Group 2, respectively. To assess the variability of our prediction result, we repeated our procedure over 100 splits of the data into training and validation sets. We obtained an average misclassification error of 25.2% of the units, with a 95% CI of (17,36)%.

For comparison, we looked into approaches that use a two-step approach, by (1) first classifying subjects based on the feature (ROIs) data only, and (2) then applying variable

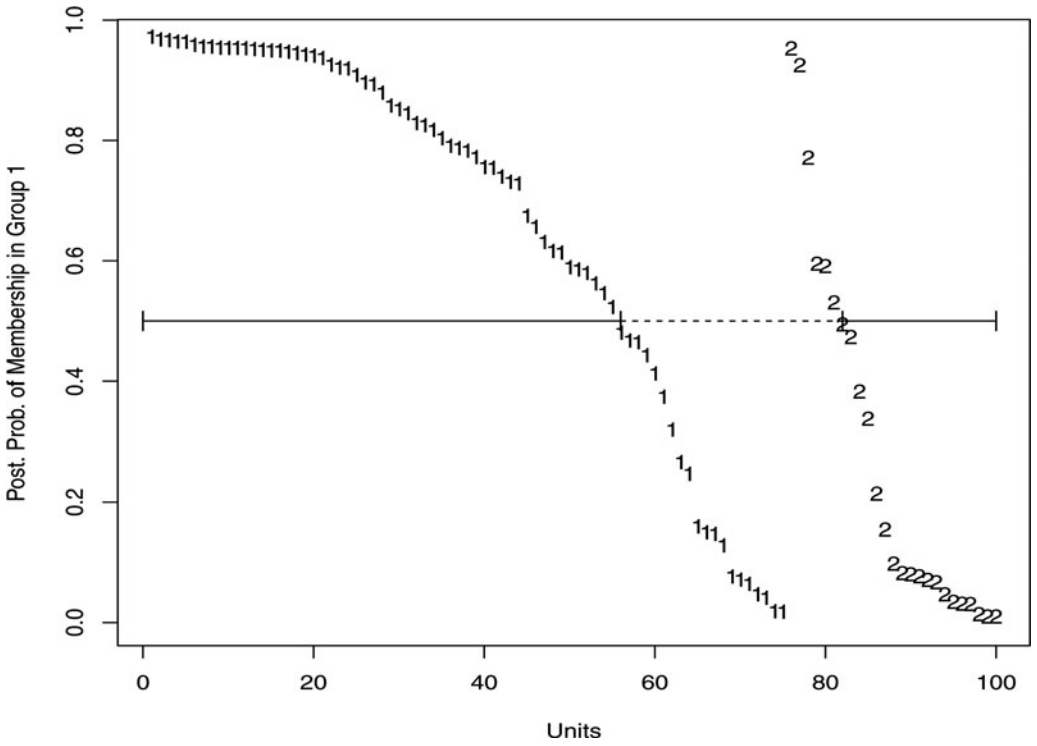

Figure 5. Simulation study: posterior probabilities of membership in Group 1 for the 100 observations in the validation set. The numbers 1 and 2 indicate the actual group membership. Units within the boundaries of the dashed line are misclassified; units above and below the solid line are correctly classified in Groups 1 and 2, respectively.

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

selection in linear models that regress the discriminatory features on the genetic covariates. In Step 1, we looked at the prediction performances of the method for Bayesian variable selection in probit models by Sha et al. (2004), that use the $\mathbf{x}$ data and a binary response indicating the group membership of the samples. Averaging over 100 splits we obtained a misclassification error of 24.8%, with a 95% CI of (19,32)%. Also, an alternative classification method, based on support vector machine (SVM; Cristianini and Shawe-Taylor 2000), resulted in a misclassification error of 27.8%, with a 95% CI of (21,36)%. In Step 2, to avoid selection biases, we used the truly discriminatory features as response variables. We applied the method by Guan and Stephens (2011), separately for each group of subjects and each feature, and compared the results with our joint estimation method. Recall that our method identified three true and no false significant SNPs. In general, the method by Guan and Stephens (2011) led to more false positives (FP). For example, a threshold of 0.5 on the marginal posterior probability of inclusion led to two true positives (TP) and two FP for feature 13 in the control group and feature 20 in the schizophrenia group, and two TP and one FP for feature 83 in the control group and feature 13 in the schizophrenia group. These results suggest that our joint selection scheme performs comparably to single-step methods in retrieving true discriminatory features but leads to a more reliable identification of the true significant SNPs.

4.2 Case Study on Schizophrenia

Schizophrenia is a severe psychiatric disorder that disrupts normal thinking, speech, and behavior. It has a strong genetic component with heritability estimated up to 80% based on family and twin studies (Cardno and Gottesman 2000). Recently, there have been increasing efforts to use fMRI and examine genetic variation to study potential schizophrenia biomarkers, to better understand the pathology of schizophrenia (see e.g., Colantuoni et al. 2008; Liu et al. 2009; Meda et al. 2010, among others). Here we have available imaging data on activation profiles captured with fMRI during a sensorimotor task for $n = 210$ subjects, of which $n_1 = 118$ are healthy controls and $n_2 = 92$ are subjects affected by schizophrenia. Participant recruitment and data collection were conducted by the MCIC, a collaborative effort of four research teams from Boston, Iowa, Minnesota, and New Mexico. Prior to inclusion in the study, all healthy participants were screened to ensure that they were free of any medical, neurological, or psychiatric illnesses, including any history of substance abuse. Additional information, including participants demographics, can be found in the article by Chen et al. (2012).

The original fMRI data were collected from all participants at four MCIC sites during a sensorimotor task, a block-design motor response to auditory stimulation. The data were then preprocessed in SPM5 (*http://www.fil.ion.ucl.ac.uk/spm*). To correct for the subject movements during the experiment, images were realigned using the INRIAlign algorithm by Freire, Roche, and Mangin (2002). Data were then spatially normalized into the standard Montreal Neurological Institute space (Friston et al. 1995a) and resliced to $3 \times 3 \times 3$ mm, resulting in $53 \times 63 \times 46$ voxels. Spatial smoothing was further applied with a 10 mm Gaussian kernel. Following the article by Chen et al. (2012), a multiple regression model was fit to the data from each participant, incorporating regressors of the stimulus and its temporal derivative plus an intercept term. The resulting regression coefficient values were used to create stimulus-on versus stimulus-off contrast images, which are sometimes also referred to as statistical parametric maps (Friston et al. 1995b). All the voxels with missing measurements were excluded. In this work, we additionally segmented the individual contrast images into $p = 116$ ROIs according to the MNI space AAL atlas by Tzourio-Mazoyer et al. (2002). Finally, we summarized activations in each region by computing the median of the statistical parametric map values for that region. The individual region-based summary statistics were also centered by subtracting the overall mean within the region. In addition to the imaging data, we have available measurements on $R = 81$ genetic covariates (SNP) for each participant in the study. The SNPs were selected by accessing the schizophrenia research forum (*http://www.schizophreniaforum.org/*) and identifying SNPs that had been previously implicated in schizophrenia.

The goal of the MCIC study was to identify a subset of ROIs with discriminating activation patterns and a set of SNPs that are relevant to explain the ROI's activations. Chen et al. (2012) applied a two-step procedure involving PCA and ICA and found a significant correlation between fMRI data and the genetic covariates in both cases and controls. Here we use our unified modeling framework to relate the patterns of observed brain activity in subjects with different conditions to the individuals' specific genome. Our model jointly infers discriminating activation patterns and identifies genetic features related to those patterns.

We split the observations into a training set and a validation set. Given the high complexity of the scientific problem we want to study, we decided to allocate a larger number of the observations to the training set. Specifically, we randomly assigned 174 participants to the training set and 36 participants to the validation set. To obtain a training set with enough information on both groups, we followed a balanced allocation scheme and randomly assigned to this set 87 subjects from the control group and the same number of subjects affected by schizophrenia.

For this data analysis, we constructed a spatial network among ROIs based on the physical location of the regions in the brain. The spatial coordinates, in a three-dimensional space, of the centroids of the individual regions allowed us to calculate a distance matrix among ROIs, based on the Euclidean distance. Given such matrix we constructed a spatial network by including an edge in the network if the distance between the corresponding pair of ROIs was less than a given threshold. We selected the threshold value such that the average number of neighbors was approximately five, therefore reproducing an overall connectivity close to the one of a three-dimensional lattice structure. We used the resulting network for the specification of the MRF prior (6) on $\boldsymbol{\gamma}$. The result is a spatially based selection process of the ROIs that enables sparsity in the resulting network structure.

Results we report below were obtained by combining the outputs from three MCMC chains. For each chain, we randomly selected different discriminatory ROIs and significant SNPs as starting points. More specifically, we initially assumed 2 ROIs and 2 SNPs, 10 ROIs and 5 SNPs, and 15 ROIs and 10 SNPs as

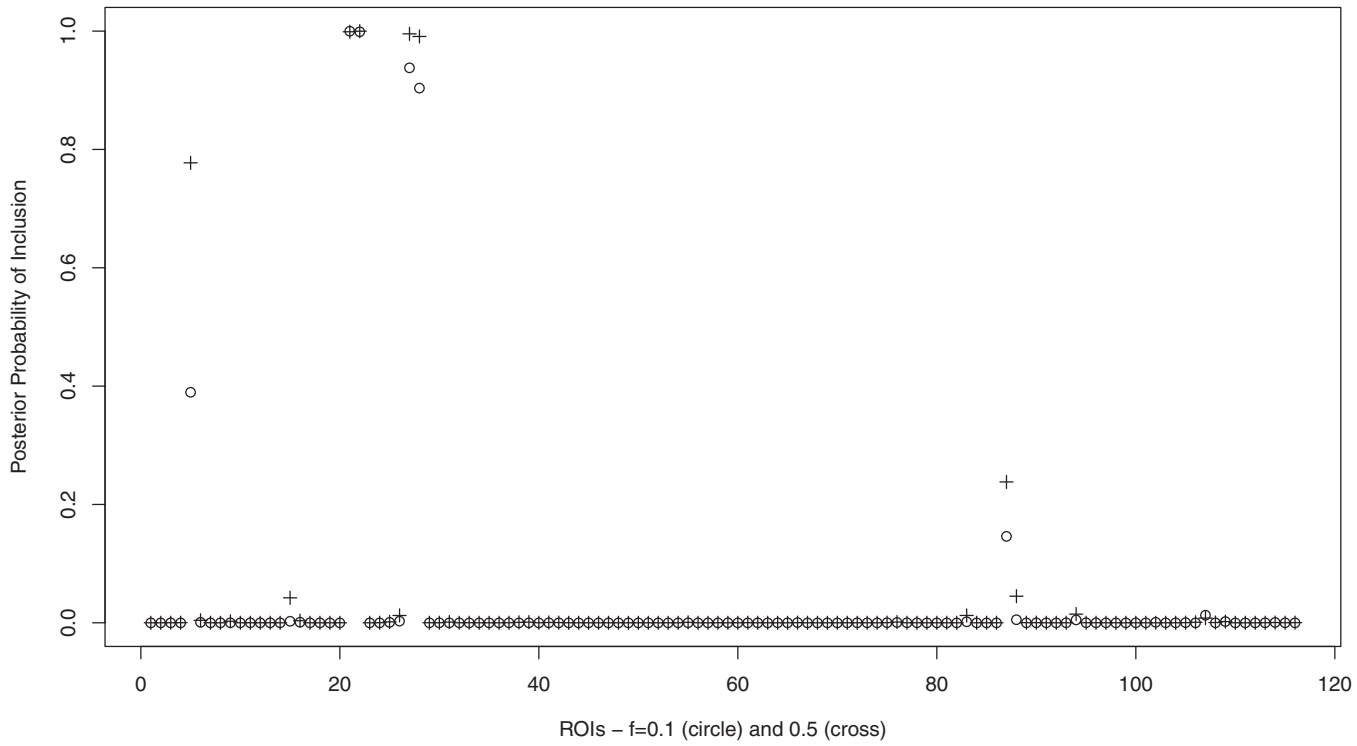

Figure 6. Schizophrenia dataset: marginal posterior probabilities of inclusion for ROIs for $f = 0.1$ (o) and $f = 0.5$ (+).

initial counts of selected features and SNPs. We used 200,000 iterations with a burn-in of 1000 iterations. We set the hyperparameter of the MRF that controls the sparsity at the ROI level to $e = -4.5$, which corresponds to setting the proportion of ROIs expected a priori to be included in the model to at least 1.1% of the total number of ROIs. We considered two values for f, setting this parameter to either 0.1 or 0.5, representing a very small or a moderate effect of the prior information, respectively. For example, for $f = 0.1$ the prior probability of inclusion of an ROI with five of its neighbors already selected is 1.8%, that is, 1.6 times the prior probability of an ROI that does not have any of its neighbors selected, while for $f = 0.5$ the same probability is 11.9%, that is, 10.8 times the probability of an ROI with no neighbors selected. These small-to-moderate effects allow the data to mostly drive the selection. For the SNP selection, we set $w_r = 0.1$, indicating that a priori approximately 10% of the SNPs are expected to be selected. In addition, although several susceptibility genes have been identified from linkage and association studies, the associations between a diagnosis of schizophrenia and individual polymorphisms have been often found weak (Harrison and Owen 2003; Duan, Sanders, and Gejman 2010). We therefore set $h_1^* = 1$, favoring the selection of small SNP effects. For the other hyperparameters, we used the same settings as described in Section 4.1.

The trace plots for the number of included ROIs and SNPs showed good mixing (figures not shown). The MCMC samplers mostly visited models with 10–15 ROIs and 1–3 SNPs, indicating that the data mostly determine the amount of sparsity in the model, since overall the numbers of selected ROIs and SNPs were not close to the corresponding expected numbers a priori. We assessed the agreement of the results between any pair of chains by looking at the correlation coefficients between marginal posterior probabilities for ROI selection, $p(\gamma_j|\mathbf{Z}, \mathbf{X})$, and SNP selection, $p(\delta_{lk}|\mathbf{Z}, \mathbf{X})$ for $k = 1, 2$. These indicated good concordance between the three MCMC chains, with pairwise correlation coefficients ranging from 0.998 to 0.999 for ROIs and from 0.998 to 0.999 ($k = 1$) and from 0.994 to 0.998 ($k = 2$) for SNPs, in the case $f = 0.5$; and from 0.996 to 0.998, from 0.997 to 0.999, and from 0.998 to 0.999, respectively, for $f = 0.1$. Also, the Gelman and Rubin's statistics, calculated on the sample values of the μ_{0kj}'s, ranged from 1.0001 to 1.0043, indicating that our chains converged to the posterior distribution.

Posterior probabilities for all ROIs and SNPs are summarized in Figures 6 and 8, respectively. These probabilities can be used to prioritize the relevant ROIs and SNPs for further experimental work. Brain regions with high posterior probability are listed in Table 1, both for $f = 0.5$ and $f = 0.1$. As it is evident from the table, the lists of selected ROIs under the two settings largely overlapped. A threshold of 0.5 on the marginal posterior probabilities selected a subset of five ROIs in the case $f = 0.5$ and four ROIs for $f = 0.1$. This threshold corresponds to an expected false discovery rate (Bayesian FDR) of 4.8% for $f = 0.5$ and 3.1% for $f = 0.1$, which we calculated according to the formulation suggested by Newton et al. (2004). The increase in the posterior probability of ROI 5 in the first row of Table 1 is an effect of the MRF prior, since ROI 5 is connected to ROIs 21, 27, and 28. The selected ROIs, shown in Figure 7, were all mid-line frontal regions, and mostly connected in the MRF prior. The frontal superior orbital, olfactory, and rectus regions have all reported schizophrenia deficits (see Nakamura et al. 2008; Diaz et al. 2011; Langdon et al. 2011, among many others). The left frontal region included the Brodmann area 10, identified by the cross-hair in Figure 7, which is often implicated in schizophrenia

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

Table 1. Schizophrenia dataset: list of selected ROIs and SNPs and corresponding posterior probabilities

ROI	Name	$p(\gamma_j \mid \mathbf{Z}, X)$ for $f = 0.1$	$p(\gamma_j \mid \mathbf{Z}, X)$ for $f = 0.5$
ROI 5	Frontal Sup Orb L	0.37	0.77
ROI 21	Olfactory L	1.00	1.00
ROI 22	Olfactory R	1.00	1.00
ROI 27	Rectus L	0.95	1.00
ROI 28	Rectus R	0.93	0.99
		Schizophrenia	
SNP	Name	$p(\delta_{2l} \mid \mathbf{Z}, X)$ for $f = 0.1$	$p(\delta_{2l} \mid \mathbf{Z}, X)$ for $f = 0.5$
SNP 25	rs1934909	0.52	0.48
SNP 31	rs875462	0.93	0.84
SNP 44	rs17101921	0.86	0.86
		Control	
SNP	Name	$p(\delta_{1l} \mid \mathbf{Z}, X)$ for $f = 0.1$	$p(\delta_{1l} \mid \mathbf{Z}, X)$ for $f = 0.5$
SNP 16	rs6794467	0.98	0.99
SNP 50	rs2421954	0.99	0.99
SNP 70	rs2270641	0.99	0.99

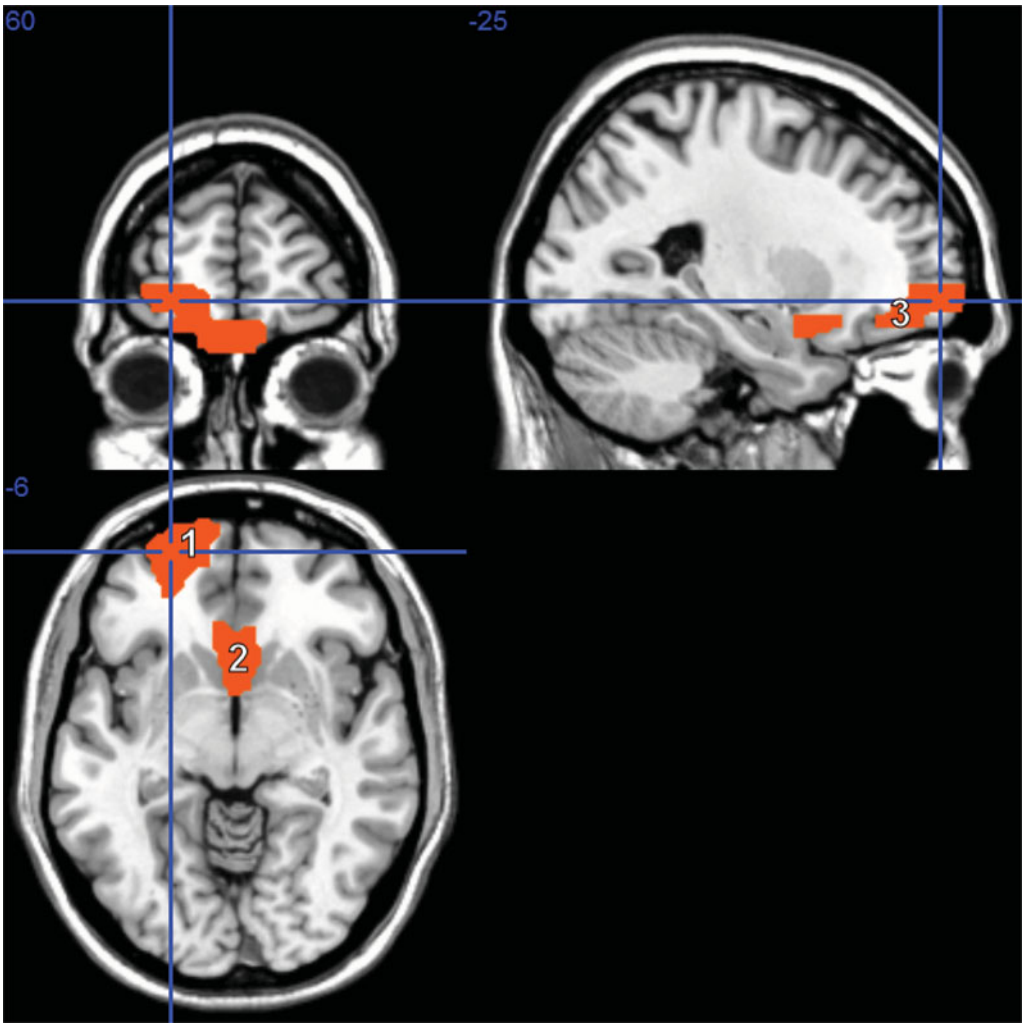

Figure 7. Schizophrenia dataset: map of brain network for the identified ROIs. The highlighted regions are the orbital part of the superior frontal gyrus (ROI 5, coded as "1," spanning superior frontal gyrus, middle frontal gyrus, inferior frontal gyrus), the olfactory cortex (ROIs 21 and 22, coded as "2," spanning subcallosal gyrus and anterior cingulate), and the gyrus rectus (ROIs 27 and 8, coded as "3," spanning medial frontal gyrus, rectal gyrus, and superior frontal gyrus). The online version of this figure is in color.

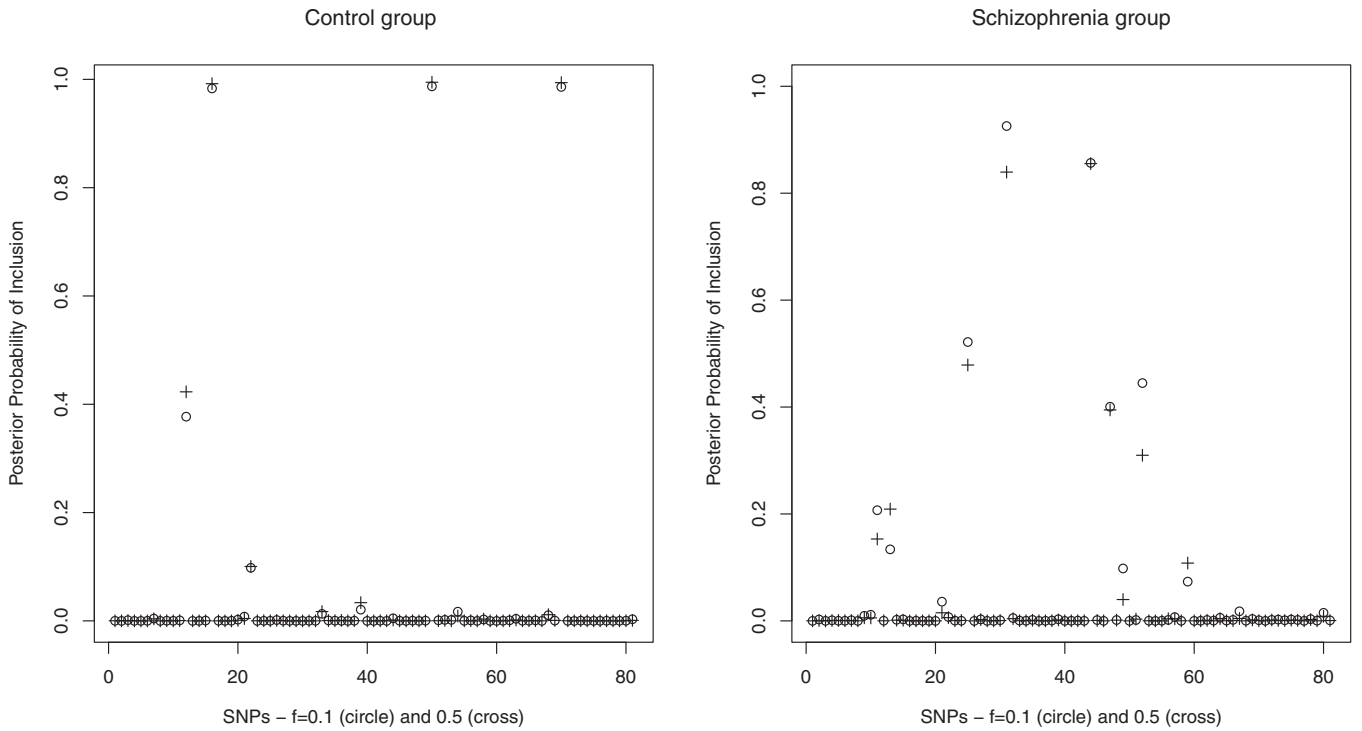

Figure 8. Schizophrenia dataset: marginal posterior probabilities of inclusion for SNPs for $f = 0.1$ (o) and $f = 0.5$ (+).

(Vogeley et al. 2003; Camchong et al. 2008; Schneiderman et al. 2011).

Posterior inference on the intercept parameters $\boldsymbol{\mu}_{01}$ and $\boldsymbol{\mu}_{02}$, given the selected features, and estimates of their correlations can give us indication on characteristic features of the two groups of participants and on dependence structures among selected ROIs suggesting new activation patterns. Estimated correlations for the case $f = 0.5$ were

$$\text{Corr}_{\boldsymbol{\mu}_{01}} = \begin{pmatrix} 1.0000 & 0.0149 & 0.0267 & 0.0295 & 0.0328 \\ 0.0149 & 1.0000 & 0.0246 & 0.0293 & 0.0235 \\ 0.0267 & 0.0246 & 1.0000 & 0.0373 & 0.0506 \\ 0.0295 & 0.0293 & 0.0373 & 1.0000 & 0.0539 \\ 0.0328 & 0.0235 & 0.0506 & 0.0539 & 1.0000 \end{pmatrix}$$

and

$$\text{Corr}_{\boldsymbol{\mu}_{02}} = \begin{pmatrix} 1.0000 & 0.3532 & 0.3403 & 0.3310 & 0.3562 \\ 0.3532 & 1.0000 & 0.4509 & 0.4193 & 0.4227 \\ 0.3403 & 0.4509 & 1.0000 & 0.3617 & 0.4024 \\ 0.3310 & 0.4193 & 0.3617 & 1.0000 & 0.3818 \\ 0.3562 & 0.4227 & 0.4024 & 0.3818 & 1.0000 \end{pmatrix}$$

for control and schizophrenic groups, respectively, showing higher correlations among ROIs in the schizophrenic group. This finding is consistent with previous work in fMRI, which has shown less unique brain activity in cases versus controls, and supports the presence of a generalized cognitive deficit (Calhoun et al. 2006). Similar correlations were obtained for $f = 0.1$.

Table 2 reports estimates and 95% CIs of the regression coefficients β_{jk} corresponding to the selected SNPs, for the selected ROIs (case $f = 0.5$), in the two groups of participants. Again, our findings are supported by other studies in the literature. Among those SNPs identified in the schizophrenia group, SNP 25 is in gene DISC1 on chromosome 1, a gene that is disrupted in schizophrenia (Kang et al. 2011; Kim et al. 2012). Also, SNP 31 in gene DTNBP1 on chromosome 6 has been found associated with schizophrenia (Duan et al. 2007). Interestingly, Colantuoni et al. (2008) reported age-related changes in the expression of putative schizophrenia genes, including DISC1 and DTNBP1, in the human prefrontal cortex, including Brodmann area 10. SNP 44 is on chromosome 10 and it has also been associated to schizophrenia (O'Donovan et al. 2009). This locus was identified via a genome-wide association study (GWA) analysis and has not yet been assigned to a specific gene. As for the control group, SNP 16 is in gene CLDN11 on chromosome 3 and SNP 70 in gene SLC18A1 on chromosome 8. Both genes are implicated in the functioning of the central nervous system (CNS; Lohoff et al. 2008a, b; Lal-Nag and Morin 2009). Together with the peripheral nervous system, the CNS has a fundamental role in the control of behavior.

Estimates of the regression coefficients reported in Table 2 inform us on the effects of the selected SNPs on the activations of the selected ROIs. The corresponding 95% CIs, also reported in the same table, allow us to assess uncertainty on those parameters. As emphasized at the end of Section 2.2, a meaningful interpretation of these regression coefficients needs to take into account the nature of the dependent variable in the regression models, which is the set of *discriminatory* ROIs in the two groups. An interesting feature of our results is that, while effects are all significant across selected ROIs in the control group, differential effects are clearly indicated in the schizophrenia group. For example, SNP 25 (rs1934909 in gene DISC1) has a significant effect on the Rectus L ROI only and SNP 31 (rs875462 in gene DTNBP1) on the Olfactory ROIs only. A close look at the alleles' frequencies of the selected SNPs in the

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

Table 2. Schizophrenia dataset: posterior mean (and 95% CI) of regression coefficients β_{jk} for the selected SNPs on the selected ROIs ($f = 0.5$)

ROI	Name	Schizophrenia group	Control group
		SNP 25	SNP 16
ROI 5	Frontal Sup Orb L	0.0714 (−0.0794, 0.2221)	−0.1854 (−0.3175, −0.0534)
ROI 21	Olfactory L	0.0659 (−0.1029, 0.2347)	−0.2819 (−0.4444, −0.1193)
ROI 22	Olfactory R	0.0664 (−0.1041, 0.2368)	−0.2779 (−0.4172, −0.1386)
ROI 27	Rectus L	0.2341 (0.0443, 0.4238)	−0.2712 (−0.4394, −0.1031)
ROI 28	Rectus R	0.1662 (−0.0202, 0.3527)	−0.2915 (−0.4347, −0.1483)
		SNP 31	SNP 50
ROI 5	Frontal Sup Orb L	0.0192 (−0.0631, 0.1016)	0.2054 (0.0541, 0.3568)
ROI 21	Olfactory L	0.1416 (0.0494, 0.2338)	0.3275 (0.1412, 0.5137)
ROI 22	Olfactory R	0.1392 (0.0461, 0.2324)	0.2471 (0.0874, 0.4068)
ROI 27	Rectus L	0.1022 (−0.0015, 0.2059)	0.2245 (0.0318, 0.4172)
ROI 28	Rectus R	0.0753 (−0.0266, 0.1771)	0.2449 (0.0808, 0.4090)
		SNP 44	SNP 70
ROI 5	Frontal Sup Orb L	0.1013 (−0.0987, 0.3014)	0.1824 (0.0356, 0.3292)
ROI 21	Olfactory L	0.2002 (−0.0239, 0.4243)	0.2487 (0.0680, 0.4293)
ROI 22	Olfactory R	0.2293 (0.0031, 0.4556)	0.2797 (0.1249, 0.4346)
ROI 27	Rectus L	0.2398 (−0.0120, 0.4917)	0.2913 (0.1044, 0.4781)
ROI 28	Rectus R	0.2585 (0.0111, 0.5060)	0.2663 (0.1071, 0.4254)

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

schizophrenia and control groups revealed that the selection is not driven by differences in genotype variation within groups. These findings call for further experimental validation.

Our model also showed good predictive performance. For $f = 0.5$, using all five selected ROIs and the selected SNP, the model misclassified 12 of the 36 samples, that is, 33% of the validation set. Similarly for the case $f = 0.1$. As with the simulation study, we compared our results with methods that use a two-step approach, first classifying subjects based on the imaging data only, and then applying variable selection in regression models that regress the discriminatory ROIs on the genetic covariates. In the first step, the method by Sha et al. (2004) misclassified 14 out of the 36 samples, that is, 39%, and the SVM method 11 out of the 31 samples, that is, 69%. In the second step, we used the five ROIs selected by our method as response variables and applied the method by Guan and Stephens (2011) separately for each group of subjects and each ROI. Using a threshold of 0.5 on the marginal posterior probability of inclusion, none of the SNPs were identified in the control group, while in the schizophrenia group the method selected SNP9 for ROI5 ($p = 0.53$), SNP47 for ROI21 ($p = 0.77$) and SNP21 for ROI22 ($p = 0.62$).

5. CONCLUDING REMARKS

We have presented a hierarchical modeling framework for the analysis of data that arise in imaging genetics. Focusing on the relationships between genetic factors and brain functional patterns, we have analyzed data from a study where the interest lies in identifying brain regions (ROIs) with discriminating activation patterns between schizophrenic patients and healthy controls, and in relating the observed activation patterns with available genetic information on the subjects (SNPs). As patterns of brain connectivity in fMRI scans are known to be related to the subjects' genome, the ability to model the link between the imaging and genetic components could indeed lead to improved diagnostics and therapeutic interventions. Overall, our results have confirmed the complex nature of genetic effects on the functional brain abnormality in schizophrenia. Particularly, the identified SNPs have implicated the DISC1 and DTNBP1 genes, associated to schizophrenia, as well as genes associated with the central nervous system, and appear to be related mainly to the frontal regions including the Brodmann area 10.

The hierarchical model we have developed has several innovative characteristics: It is integrative, in that it combines activation patterns with genetic information on the subjects; it achieves the simultaneous selection of a set of discriminatory regions and relevant SNPs; it employs spatially based selection process priors that capture available knowledge on connectivity among regions of the brain, so that regions having the same activation patterns are more likely to be selected together. An additional feature of our model is that it takes into account correlation among selected ROIs, as captured by the covariance matrices of the distribution of the intercept parameters.

Our integrative model aims at selecting SNPs that are implicated in the differential activation patterns observed in patients and healthy controls. In our analysis, we have considered a relatively small set of SNPs that were already known to be of interest based on previous studies. Typically, the initial list of potentially interesting SNPs can be the result of GWA studies, where the outcome, that is, the known phenotype (disease/not disease), is related to the genetic covariates (SNPs) in a logistic model. Our model can then be used to determine if the selected SNPs are additionally significantly associated to an alteration of the normal brain functions. Clearly, discovery of new associations between SNPs and the disease under study is also possible with our model. In such situations, analyses with large number of SNPs can be tackled by considering gene-based

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

summaries of SNP data. Indeed, more and more GWA studies confirm the involvement of aggregates of common SNPs that collectively account for a substantial proportion of variation in risk to the disorder (see Purcell et al. 2009, for a schizophrenia study).

Several extensions of our model are worth investigating. First, we can learn about the fixed hyperparameters. For example, we can assume prior distributions for the parameters of the Ising model, as in the articles by Liang (2010) and Stingo et al. (2011), as well as the scale parameters of the spike and slab priors (4), as sometimes done in Bayesian variable selection (Chipman et al. 2001; Scott and Berger 2010). Second, extensions to mixture models for clustering the subjects are possible. In particular, infinite mixture models can be fit based on the Dirichlet process, as in the article by Kim, Tadesse, and Vannucci (2006), or based on the probit stick-breaking process proposed by Chung and Dunson (2009) for a mixture of univariate regression models, which also incorporates selection of the covariates. Furthermore, even though our method already suggests connections among distant features that can form new activation patterns, a more formal approach could assume $\boldsymbol{\mu}_{0k(\gamma)}$ to be a realization of a network of connected components (nodes) that describes general relationships among the selected features as

$$\boldsymbol{\mu}_{0k(\gamma)}|G_{k(\gamma)} \sim N_{p_\gamma}(\boldsymbol{\nu}_{k(\gamma)}, \boldsymbol{\Gamma}_{0k(\gamma)}), \quad k = 1, \ldots, K, \quad (10)$$

with $G_{k(\gamma)}$ the graph encoding the relationships (see Dobra, Lenkoski, and Rodriguez 2011). This would allow us to estimate component-specific networks among selected ROIs. Finally, another interesting avenue is to extend our model to handle data observed at different time points. Hidden Markov models can be employed. For example, in an application to genomic data, Gupta, Qu, and Ibrahim (2007) considered a hierarchical Bayesian hidden Markov regression model for determining groups of genes being influenced by separate sets of covariates over time.

APPENDIX: MCMC ALGORITHM

We start by writing the expression of the marginal posterior distribution of $(\boldsymbol{\gamma}, \boldsymbol{\delta}_k, \boldsymbol{\mu}_{0k(\gamma)})$ explicitly. The dimension of $\boldsymbol{\mu}_{0k(\gamma)}$ varies depending on the number of nonzero components in $\boldsymbol{\gamma}$. This type of issues has been traditionally addressed in the context of transdimensional MCMC algorithms. In this article, we follow the hybrid Gibbs/Metropolis approach discussed by Dellaportas, Forster, and Ntzoufras (2002), based on the article by Carlin and Chib (1995), and assume a pseudo-prior for the parameter $\boldsymbol{\mu}_{0k(\gamma^c)}$ of the nonselected features, that is, $\boldsymbol{\mu}_{0k(\gamma^c)} \sim N(0, \boldsymbol{\Gamma}_{0k(\gamma^c)})$, where $\boldsymbol{\Gamma}_{0k(\gamma^c)} = \text{diag}(\sigma^2_{p_\gamma+1}, \ldots, \sigma^2_p)$, with $\sigma^2_j \sim \text{IG}(\tilde{a}_j, \tilde{b}_j)$, $j = p_\gamma + 1, \ldots, p$. Alternatively, methods based on mixtures of singular distributions could be usefully employed in this setting (Petris and Tardella 2003; Gottardo and Raftery 2008). For a review of available transdimensional MCMC algorithms, we refer, among others, to Frühwirth-Schnatter 2006, chap. 5, and the discussion in the literature by Fan and Sisson (2011).

The unnormalized full conditionals can be derived from the conditional independencies of our model (see Figure 1). Let $\boldsymbol{\mu}_{0k} = (\boldsymbol{\mu}_{0k(\gamma)}, \boldsymbol{\mu}_{0k(\gamma^c)})$. Then, the joint marginal posterior distribution of $(\boldsymbol{\gamma}, \boldsymbol{\delta}_k, \boldsymbol{\mu}_{0k})$ is

$$\begin{aligned} &p(\boldsymbol{\gamma}, \boldsymbol{\delta}_k, \boldsymbol{\mu}_{0k}|\mathbf{X}, \mathbf{Z}) \\ &\quad \propto \prod_{j:\gamma_j=0} p(\mathbf{x}_{j(\gamma^c)}|\boldsymbol{\gamma}) \prod_{j:\gamma_j=1} \prod_{k=1}^{K} p\left(\mathbf{x}_{jk(\gamma)}|\mathbf{Z}_{k(\delta_k)}, \mu_{0jk}, \boldsymbol{\delta}, \boldsymbol{\gamma}\right) \\ &\qquad \times \prod_{k=1}^{K} \left[p(\boldsymbol{\mu}_{0k(\gamma)}|\boldsymbol{\gamma}) \prod_{j:\gamma_j=0} p(\mu_{0jk(\gamma^c)}|\boldsymbol{\gamma}) \right] p(\boldsymbol{\gamma}) p(\boldsymbol{\delta}) \end{aligned}$$

with

$$\begin{aligned} &p(\mathbf{x}_{j(\gamma^c)}|\boldsymbol{\gamma}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{\Gamma(a_0')}{\Gamma(a_0)} \frac{b_0^{a_0}}{b_0'^{a_0'}}, \\ &p\left(\mathbf{x}_{jk(\gamma)}|\mathbf{Z}_{\mathbf{k}(\delta_\mathbf{k})}, \mu_{\mathbf{0jk}}, \boldsymbol{\delta}, \boldsymbol{\gamma}\right) \\ &\quad = \left(\frac{1}{\sqrt{2\pi}}\right)^{n_k} \frac{\Gamma(a_k')}{\Gamma(a_k)} \frac{b_k^{a_k}}{b_k'^{a_k'}} \left|\mathbf{Z}_{k(\delta_k)}\mathbf{Z}^T_{k(\delta_k)} + h^{-1}\mathbf{I}_{n_k}\right|, \\ &p(\boldsymbol{\mu}_{0k(\gamma)}|\boldsymbol{\gamma}) = \pi^{-\frac{p_\gamma}{2}} h_1^{*\frac{-p_\gamma}{2}} \\ &\quad \times \frac{|\mathbf{Q}|^{(d_k+p_\gamma+1)/2}}{|\mathbf{Q} + (\boldsymbol{\mu}_{0k} - \mathbf{m}_{0k})(\boldsymbol{\mu}_{0k} - \mathbf{m}_{0k})^T|^{(d_k+p_\gamma)/2}} \frac{\Gamma((d_k + p_\gamma)/2)}{\Gamma((d_k + p_\gamma + 1)/2)}, \\ &p(\mu_{0jk(\gamma^c)}|\boldsymbol{\gamma}) = \left(\frac{1}{\sqrt{2\pi}}\right) \frac{\tilde{b}_j^{\tilde{a}_j}}{\tilde{b}_{jk}'^{\tilde{a}_j'}} \frac{\Gamma(\tilde{a}_j')}{\Gamma(\tilde{a}_j)}, \end{aligned}$$

and where $h_1^* = 2h_1$, $\mathbf{Z}_{k(\delta)}$ is the $n_k \times |\boldsymbol{\delta}_k|$ data matrix of the selected covariates, n_k is the number of subjects allocated to the kth component, and $|\boldsymbol{\delta}_k| = \sum_{r=1}^{R} \delta_{rk}$, and with

$$\begin{aligned} b_0' &= b_0 + \frac{\mathbf{x}_j^T \mathbf{x}_j}{2}, \\ a_0' &= a_0 + n/2, \\ b_k' &= b_k + \left(\mathbf{x}_{jk} - \mathbf{1}_{n_k}\mu_{0jk}\right)^T \left(h^{-1}\mathbf{I}_{n_k} + \mathbf{Z}_{k(\delta_k)}\mathbf{Z}^T_{k(\delta_k)}\right)^{-1} \\ &\qquad \times \left(\mathbf{x}_{jk} - \mathbf{1}_{n_k}\mu_{0jk}\right)/2, \\ a_k' &= a_k + n_k/2, \\ \tilde{a}_j' &= \tilde{a}_j + 1/2, \\ \tilde{b}_{jk}' &= \tilde{b}_j + \frac{(\mu_{0jk} - m_{0jk})^2}{2}, \end{aligned}$$

for $k = 1, \ldots, K$ and $j = p_\gamma + 1, \ldots, p$, with $\mathbf{x}_j$ the $n \times 1$ vector of the observed values for feature j and $\mathbf{x}_{jk}$ the $n_k \times 1$ vector of the observed values for feature j and group k.

Our MCMC algorithm comprises the following three steps:

1. A Metropolis-Hastings step on $\boldsymbol{\gamma}$: It consists of randomly choosing between changing the value of a single γ_j (either 0 to 1 or 1 to 0) and swapping two γ_j's (with opposite values). The transition from the current value $\boldsymbol{\gamma}^O$ to the proposed value $\boldsymbol{\gamma}^N$ is accepted with probability

$$\begin{aligned} &\min\left[\left(\prod_{j:\gamma_j=0} p(\mathbf{x}_{j(\gamma^c)}|\boldsymbol{\gamma}^N) \prod_{j:\gamma_j=1} \prod_{k=1}^{K} p(\mathbf{x}_{jk(\gamma)}|\mu_{0jk}, \boldsymbol{\delta}, \boldsymbol{\gamma}^N) \right.\right. \\ &\quad \left. \times \prod_{k=1}^{K} \left[p(\boldsymbol{\mu}_{0k(\gamma)}|\boldsymbol{\gamma}^N) \prod_{j:\gamma_j=0} p(\mu_{0jk(\gamma^c)}|\boldsymbol{\gamma}^N) \right] p(\boldsymbol{\gamma}^N) \right) \\ &\quad \Big/ \left(\prod_{j:\gamma_j=0} p(\mathbf{x}_{j(\gamma^c)}|\boldsymbol{\gamma}^O) \prod_{j:\gamma_j=1} \prod_{k=1}^{K} p(\mathbf{x}_{jk(\gamma)}|\mu_{0jk}, \boldsymbol{\delta}, \boldsymbol{\gamma}^O) \right. \\ &\quad \left.\left. \times \prod_{k=1}^{K} \left[p(\boldsymbol{\mu}_{0k(\gamma)}|\boldsymbol{\gamma}^O) \prod_{j:\gamma_j=0} p(\mu_{0jk(\gamma^c)}|\boldsymbol{\gamma}^O) \right] p(\boldsymbol{\gamma}^O) \right), 1 \right]. \end{aligned}$$

2. A Metropolis-Hastings step for $\boldsymbol{\delta}_k$: Similarly to the previous step, for each of the K components we randomly choose between changing the value of a single δ_{rk} and swapping two δ_{rk}'s with opposite values. The transition from the current value $\boldsymbol{\delta}_k^O$ to the proposed value $\boldsymbol{\delta}_k^N$ is accepted with probability

$$\min\left[\frac{\prod_{j:\gamma_j=1} p\left(\mathbf{x}_{jk(\gamma)}\middle|\mu_{0jk}, \boldsymbol{\delta}_k^N, \boldsymbol{\gamma}\right) p\left(\boldsymbol{\delta}_k^N\right)}{\prod_{j:\gamma_j=1} p\left(\mathbf{x}_{jk(\gamma)}\middle|\mu_{0jk}, \boldsymbol{\delta}_k^O, \boldsymbol{\gamma}\right) p\left(\boldsymbol{\delta}_k^O\right)}, 1\right].$$

3. A random walk Metropolis-Hastings step on the $\boldsymbol{\mu}_{0k(\gamma)}$'s: We update $\boldsymbol{\mu}_{0k(\gamma)}$ by proposing new values for all μ_{0kj} included in the model at that particular iteration, according to $\mu_{0kj}^N = \mu_{0kj}^O + \epsilon$, with $\epsilon \sim N(0, v^2)$ for $j = 1, \ldots, p_\gamma$ and $k = 1, \ldots K$, for some fixed choice of the proposal parameter v. The transition from μ_{0kj}^O to μ_{0kj}^N is accepted with probability

$$\min\left[\frac{p\left(\mathbf{x}_{jk(\gamma)}\middle|\mu_{0jk}^N, \boldsymbol{\delta}, \boldsymbol{\gamma}\right) p\left(\boldsymbol{\mu}_{0k(\gamma)}^N\middle|\boldsymbol{\gamma}\right)}{p\left(\mathbf{x}_{jk(\gamma)}\middle|\mu_{0jk}^O, \boldsymbol{\delta}, \boldsymbol{\gamma}\right) p\left(\boldsymbol{\mu}_{0k(\gamma)}^O\middle|\boldsymbol{\gamma}\right)}, 1\right].$$

Note that in all three steps the proposal distribution is symmetric and therefore does not appear in the Metropolis-Hastings acceptance ratio.

[Received May 2012. Revised May 2013.]

REFERENCES

Ashby, F. G. (2011), *Statistical Analysis of fMRI Data*, Cambridge, MA: MIT Press. [876]

Banerjee, S., Gelfand, A., and Carlin, B. (2003), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman & Hall/CRC. [879,880]

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society,* Series B, 36, 192–236. [879]

Bowman, F., Caffo, B., Bassett, S., and Kilts, C. (2008), "Bayesian Hierarchical Framework for Spatial Modeling of fMRI Data," *NeuroImage*, 39, 146–156. [876]

Brown, P., Vannucci, M., and Fearn, T. (1998), "Multivariate Bayesian Variable Selection and Prediction," *Journal of the Royal Statistical Society,* Series B, 60, 627–641. [879]

Calhoun, V., Adali, T., Kiehl, K., Astur, R., Pekar, J., and Pearlson, G. (2006), "A Method for Multi-Task fMRI Data Fusion Applied to Schizophrenia," *Human Brain Mapping*, 27, 598–610. [887]

Camchong, J., Dyckman, K., Austin, B., Clementz, B., and McDowell, J. (2008), "Common Neural Circuitry Supporting Volitional Saccades and Its Disruption in Schizophrenia Patients and Relatives," *Biological Psychiatry*, 64, 1042–1050. [887]

Cardno, A., and Gottesman, I. (2000), "Twin Studies of Schizophrenia: From Bow-and-Arrow Concordances to Star Wars Mx and Functional Genomics," *American Journal of Medical Genetics*, 97, 12–17. [884]

Carlin, B., and Chib, S. (1995), "Bayesian Model Choice via Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society,* Series B, 157, 473–484. [889]

Chen, J., Calhoun, V., Pearlson, G., Ehrlich, S., Turner, J., Ho, B., Wassink, T., Michael, A., and Liu, J. (2012), "Multifaceted Genomic Risk for Brain Function in Schizophrenia," *NeuroImage*, 61, 866–875. [876,877,884]

Chipman, H., George, E. I., and McCulloch, R. E. (2001), "The Practical Implementation of Bayesian Model Selection," in *IMS Lecture Notes*, 38, 65–134. [878,889]

Chung, Y., and Dunson, D. (2009), "Nonparametric Bayes Conditional Distribution Modeling With Variable Selection," *Journal of the American Statistical Association*, 104, 1646–1660. [889]

Colantuoni, C., Hyde, T., Mitkus, S., Joseph, A., Sartorius, L., Aguirre, C., Creswell, J., Johnson, E., Deep-Soboslay, A., Herman, M., Lipska, B. K., Weinberger, D. R., and Kleinman, J. (2008), "Age-Related Changes in the Expression of Schizophrenia Susceptibility Genes in the Human Prefrontal Cortex," *Brain Structure & Function*, 213, 255–271. [877,884,887]

Cristianini, N., and Shawe-Taylor, J. (2000), *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge: Cambridge University Press. [884]

Dellaportas, P., Forster, J., and Ntzoufras, I. (2002), "On Bayesian Model and Variable Selection Using MCMC," *Statistics and Computing*, 12, 27–36. [889]

Diaz, M., He, G., Gadde, S., Bellion, C., Belger, A., Voyvodic, J., and McCarthy, G. (2011), "The Influence of Emotional Distraction on Verbal Working Memory: An fMRI Investigation Comparing Individuals With Schizophrenia and Healthy Adults," *Journal of Psychiatric Research*, 45, 1184–1193. [885]

Dobra, A., Lenkoski, A., and Rodriguez, A. (2011), "Bayesian Inference for General Gaussian Graphical Models With Application to Multivariate Lattice Data," *Journal of the American Statistical Association*, 106, 1418–1449. [889]

Duan, J., Martinez, M., Sanders, A., Hou, C., Burrell, G., Krasner, A., Schwartz, D., and Gejman, P. (2007), "DTNBP1 (Dystrobrevin Binding Protein 1) and Schizophrenia: Association Evidence in the 3' End of the Gene," *Human Heredity*, 64, 97–106. [887]

Duan, J., Sanders, A., and Gejman, P. (2010), "Genome-Wide Approaches to Schizophrenia," *Behavioural Brain Research*, 83, 93–102. [885]

Etzel, J. A., Gazzola, V., and Keysers, C. (2009), "An Introduction to Anatomical ROI-Based fMRI Classification Analysis," *Brain Research*, 1282, 114–125. [877]

Fan, Y., and Sisson, S. (2011), "Reversible Jump MCMC," in *Handbook of Markov Chain Monte Carlo*, eds. Brooks S., Gelman, A., Jones, G. L. and Meng, X.-L., Boca Raton, FL: Chapman & Hall/CRC, pp. 67–92. [889]

Fearn, T., Brown, P., and Besbeas, P. (2002), "A Bayesian Decision Theory Approach to Variable Selection for Discrimination," *Statistics and Computing*, 12, 253–260. [881]

Freire, L., Roche, A., and Mangin, J.-F. (2002), "What is the Best Similarity Measure for Motion Correction in fMRI Time Series?" *IEEE Transactions on Medical Imaging*, 21, 470–484. [884]

Friston, K. (1994), "Functional and Effective Connectivity in Neuroimaging: A Synthesis," *Human Brain Mapping*, 2, 56–78. [876,879]

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D., and Frackowiak, R. S. J. (1995a), "Spatial Registration and Normalization of Images," *Human Brain Mapping*, 3, 165–189. [884]

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. J. (1995b), "Statistical Parametric Maps in Functional Imaging: A General Linear Approach," *Human Brain Mapping*, 2, 189–210. [884]

Frühwirth-Schnatter, S. (2006), *Finite Mixture and Markov Switching Models*, New York: Springer. [889]

Gelman, A., and Hill, J. (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, New York: Cambridge University Press. [879]

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–472. [882]

George, E., and McCulloch, R. (1993), "Variable Selection via Gibbs Sampler," *Journal of the American Statistical Association*, 88, 881–889. [879]

——— (1997), "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7, 339–373. [879]

Gottardo, R., and Raftery, A. E. (2008), "Markov Chain Monte Carlo With Mixtures of Mutually Singular Distributions," *Journal of Computational and Graphical Statistics*, 17, 949–975. [889]

Guan, Y., and Stephens, M. (2011), "Bayesian Variable Selection Regression for Genome-Wide Association Studies and Other Large-Scale Problems," *Annals of Applied Statistics*, 5, 1780–1815. [884,888]

Guo, Y., Bowman, F., and Kilts, C. (2008), "Predicting the Brain Response to Treatment Using a Bayesian Hierarchical Model With Application to a Study of Schizophrenia," *Human Brain Mapping*, 29, 1092–1109. [876]

Gupta, M., Qu, P., and Ibrahim, J. (2007), "A Temporal Hidden Markov Regression Model for the Analysis of Gene Regulatory Networks," *Biostatistics*, 8, 805–820. [889]

Harrison, P., and Owen, M. (2003), "Genes for Schizophrenia? Recent Findings and Their Pathophysiological Implications," *Lancet*, 361, 417–419. [885]

Hoff, P. (2006), "Model-Based Subspace Clustering," *Bayesian Analysis*, 1, 321–344. [878]

Jbabdi, S., Woolrich, M., and Behrens, T. (2009), "Multiple-Subjects Connectivity-Based Parcellation Using Hierarchical Dirichlet Process Mixture Models," *NeuroImage*, 44, 373–384. [876]

Kang, E., Burdick, K., Kim, J., Duan, X., Guo, J., Sailor, K., Jung, D., Ganesan, S., Choi, S., Pradhan, D., Lu, B., Avramopoulos, D., Christian, K., Malhotra, A., Song, H., and Ming, G. (2011), "Interaction Between FEZ1 and DISC1 in Regulation of Neuronal Development and Risk for Schizophrenia," *Neuron*, 72, 559–571. [887]

Kim, J., Liu, C., Zhang, F., Duan, X., Wen, Z., Song, J., Feighery, E., Lu, B., Rujescu, D., St Clair, D., Christian, K., Callicott, J., Weinberger, D., Song, H., and Ming, G. (2012), "Interplay Between DISC1 and GABA Signaling

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

Downloaded by [Fondren Library, Rice University] at 06:43 30 September 2013

Regulates Neurogenesis in Mice and Risk for Schizophrenia," *Cell*, 148, 1051–1064. [887]

Kim, S., Smyth, P., Stern, H., and Turner, J. (2005), "Parametric Response Surface Models for Analysis of Multi-Site fMRI Data," *Lecture Notes in Computer Science*, 3749, 352–359. [877]

Kim, S., Tadesse, M., and Vannucci, M. (2006), "Variable Selection in Clustering via Dirichlet Process Mixture Models," *Biometrika*, 93, 877–893. [878,889]

Lal-Nag, M., and Morin, P. (2009), "The Claudins," *Genome Biology*, 10, 235. [887]

Langdon, R., McGuire, J., Stevenson, R., and Catts, S. (2011), "Clinical Correlates of Olfactory Hallucinations in Schizophrenia," *British Journal of Clinical Psychology*, 50, 145–163. [885]

Lazar, N. (2008), *The Statistical Analysis of fMRI Data*, New York: Springer-Verlag. [876]

Li, F., and Zhang, N. (2010), "Bayesian Variable Selection in Structured High-Dimensional Covariate Space With Application in Genomics," *Journal of American Statistical Association*, 105, 1202–1214. [879]

Liang, F. (2010), "A Double Metropolis-Hastings Sampler for Spatial Models With Intractable Normalizing Constants," *Journal of Statistical Computing and Simulation*, 80, 1007–1022. [889]

Lindquist, M. (2008), "The Statistical Analysis of fMRI Data," *Statistical Science*, 23, 439–464. [876]

Liu, J., Pearlson, G., Windemuth, A., Ruano, G., Perrone-Bizzozero, N., and Calhoun, V. (2009), "Combining fMRI and SNP Data to Investigate Connections Between Brain Function and Genetics Using Parallel ICA," *Human Brain Mapping*, 30, 241–255. [877,884]

Lohoff, F., Lautenschlager, M., Mohr, J., Ferraro, T., Sander, T., and Gallinat, J. (2008a), "Association Between Variation in the Vesicular Monoamine Transporter 1 Gene on Chromosome 8p and Anxiety-Related Personality Traits," *Neuroscience Letters*, 434, 41–45. [887]

Lohoff, F., Weller, A., Bloch, P., Buono, R., Doyle, G., Ferraro, T., and Berrettini, W. (2008b), "Association Between Polymorphisms in the Vesicular Monoamine Transporter 1 Gene (VMAT1/SLC18A1) on Chromosome 8p and Schizophrenia," *Neuropsychobiology*, 57, 55–60. [887]

Meda, S., Jagannathan, K., Gelernter, J., Calhoun, V., Liu, J., Stevens, M., and Pearlson, G. (2010), "A Pilot Multivariate Parallel ICA Study to Investigate Differential Linkage Between Neural Networks and Genetic Profiles in Schizophrenia," *NeuroImage*, 53, 1007–1015. [884]

Nakamura, M., Nestor, P., Levitt, J., Cohen, A., Kawashima, T., Shenton, M., and McCarley, R. (2008), "Orbitofrontal Volume Deficit in Schizophrenia and Thought Disorder," *Brain*, 131, 180–195. [885]

Newton, M., Noueiry, A., Sarkar, D., and Ahlquist, P. (2004), "Detecting Differential Gene Expression With a Semiparametric Hierarchical Mixture Method," *Biostatistics*, 5, 155–176. [885]

O'Donovan, M., Norton, N., Williams, H., Peirce, T., Moskvina, V., Nikolov, I., Hamshere, M., Carroll, L., Georgieva, L., Dwyer, S., Holmans, P., Marchini, J., Spencer, C., Howie, B., Leung, H., Giegling, I., Hartmann, A., Moller, H., Morris, D., Shi, Y., Feng, G., Hoffmann, P., Propping, P., Vasilescu, C., Maier, W., Rietschel, M., Zammit, S., Schumacher, J., Quinn, E., Schulze, T., Iwata, N., Ikeda, M., Darvasi, A., Shifman, S., He, L., Duan, J., Sanders, A., Levinson, D., Adolfsson, R., Osby, U., Terenius, L., Jonsson, E., Cichon, S., Nothen, M., Gill, M., Corvin, A., Rujescu, D., Gejman, P., Kirov, G., Craddock, N., Williams, N., and Owen, M. (2009), "Analysis of 10 Independent Samples Provides Evidence for Association Between Schizophrenia and a SNP Flanking Fibroblast Growth Factor Receptor 2," *Molecular Psychiatry*, 14, 30–36. [887]

O'Hara, R., and Sillanpaa, M. J. (2009), "A Review of Bayesian Variable Selection Methods: What, How and Which," *Bayesian Analysis*, 4, 85–118. [878]

Penny, W., Trujillo-Barreto, N., and Friston, K. (2005), "Bayesian fMRI Time Series Analysis With Spatial Priors," *NeuroImage*, 24, 350–362. [876,879]

Petris, G., and Tardella, L. (2003), "A Geometrical Approach to Transdimensional MCMC," *The Canadian Journal of Statistics*, 31, 469–482. [889]

Propp, J., and Wilson, D. (1996), "Exact Sampling With Coupled Markov Chains and Applications to Statistical Mechanics," *Random Structures and Algorithms*, 9, 223–252. [879]

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., Sklar, P., International Schizophrenia Consortium. (2009), "Common Polygenic Variation Contributes to Risk of Schizophrenia and Bipolar Disorder," *Nature*, 460, 748–752. [889]

Raftery, A., and Dean, N. (2006), "Variable Selection for Model-Based Clustering," *Journal of the American Statistical Association*, 101, 168–178. [877]

Robert, P., and Escoufier, Y. (1976), "A Unifying Tool for Linear Multivariate Statistical Methods: The RV- Coefficient," *Journal of the Royal Statistical Society,* Series C, 25, 257–265. [882]

Schneiderman, J., Hazlett, E., Chu, K., Zhang, J., Goodman, C., Newmark, R., Torosjan, Y., Canfield, E., Entis, J., Mitropoulou, V., Tang, C., Friedman, J., and Buchsbaum, M. (2011), "Brodmann Area Analysis of White Matter Anisotropy and Age in Schizophrenia," *Schizophrenia Research*, 130, 57–67. [887]

Scott, J., and Berger, J. O. (2010), "Bayes and Empirical-Bayes Multiplicity Adjustment in the Variable Selection Problem," *The Annals of Statistics*, 38, 2587–2619. [889]

Servin, B., and Stephens, M. (2007), "Imputation-Based Analysis of Association Studies: Candidate Regions and Quantitative Traits," *PLoS Genetics*, 3, e114. [881]

Sha, N., Vannucci, M., Tadesse, M., Brown, P., Dragoni, I., Davies, N., Roberts, T., Contestabile, A., Salmon, N., Buckley, C., and Falciani, F. (2004), "Bayesian Variable Selection in Multinomial Probit Models to Identify Molecular Signatures of Disease Stage," *Biometrics*, 60, 812–819. [884,888]

Smilde, A. K., Kiers, H. A. L., Bijlsma, S., Rubingh, C. M., and van Erk, M. J. (2009), "Matrix Correlations for High-Dimensional Data: The Modified RV-Coefficient," *Bioinformatics*, 25, 401–405. [882]

Smith, M., and Kohn, R. (1996), "Nonparametric Regression Using Bayesian Variable Selection," *Journal of Econometrics*, 75, 317–344. [878]

Smith, M., Putz, B., Auer, D., and Fahrmeir, L. (2003), "Assessing Brain Connectivity Through Spatial Bayesian Variable Selection," *NeuroImage*, 20, 802–815. [876,879]

Stingo, F., Chen, Y., Tadesse, M., and Vannucci, M. (2011), "Incorporating Biological Information Into Linear Models: A Bayesian Approach to the Selection of Pathways and Genes," *Annals of Applied Statistics*, 5, 1978–2002. [879,889]

Stingo, F., Vannucci, M., and Downey, G. (2012), "Bayesian Wavelet-Based Curve Classification via Discriminant Analysis With Markov Random Tree Priors," *Statistica Sinica*, 22, 465–488. [878,879]

Tadesse, M., Sha, N., and Vannucci, M. (2005), "Bayesian Variable Selection in Clustering High Dimensional Data," *Journal of American Statistical Association*, 100, 602–617. [877,878,879]

Telesca, D., Muller, P., Parmigiani, G., and Freedman, R. (2012), "Modeling Dependent Gene Expression," *Annals of Applied Statistics*, 6, 542–560. [879]

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002), "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain," *NeuroImage*, 15, 273–289. [876,884]

Vogeley, K., Tepest, R., Schneider-Axmann, T., Hütte, H., Zilles, K., Honer, W., and Falkai, P. (2003), "Automated Image Analysis of Disturbed Cytoarchitecture in Brodmann Area 10 in Schizophrenia," *Schizophrenia Research*, 62, 133–140. [887]

Woolrich, M., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., and Smith, S. (2009), "Bayesian Analysis of Neuroimaging Data in FSL," *NeuroImage*, 45, S173–S186. [876]

Woolrich, M., Jenkinson, M., Brady, J., and Smith, S. (2004), "Fully Bayesian Spatio-Temporal Modeling of fMRI Data," *IEEE Transactions on Medical Imaging*, 23, 213–231. [876]