Biometrics 000, 000–000 DOI: 000
000 0000

Joint high-dimensional Bayesian variable and covariance selection with an application to eQTL analysis

Anindya Bhadra*

Department of Statistics, Purdue University, West Lafayette, IN 47907-2066.

**email:* bhadra@purdue.edu

and

Bani K. Mallick

Department of Statistics, Texas A&M University, College Station, TX 77843-3143.

Summary: We describe a Bayesian technique to (a) perform a sparse joint selection of significant predictor variables and significant inverse covariance matrix elements of the response variables in a high-dimensional linear Gaussian sparse seemingly unrelated regression (SSUR) setting and (b) perform an association analysis between the high-dimensional sets of predictors and responses in such a setting. To search the high-dimensional model space, where both the number of predictors and the number of possibly correlated responses can be larger than the sample size, we demonstrate that a marginalization-based collapsed Gibbs sampler, in combination with spike and slab type of priors, offers a computationally feasible and efficient solution. As an example, we apply our method to an expression quantitative trait loci (eQTL) analysis on publicly available single nucleotide polymorphism (SNP) and gene expression data for humans where the primary interest lies in finding the significant associations between the sets of SNPs and possibly correlated genetic transcripts. Our method also allows for inference on the sparse interaction network of the transcripts (response variables) after accounting for the effect of the SNPs (predictor variables). We exploit properties of Gaussian graphical models to make statements concerning conditional independence of the responses. Our method compares favorably to existing Bayesian approaches developed for this purpose.

Key words: eQTL Analysis; Gaussian Graphical Model; Hyper-inverse Wishart distribution; Joint Variable and Covariance Selection; Sparse Seemingly Unrelated Regression.

1. Introduction

The study of expression quantitative trait loci (eQTL) associates genetic variation in populations (typically, single nucleotide polymorphisms or SNPs, modeled as categorical predictors) with variations in gene expression levels (typically, continuous response variables) in order to identify polymorphic genomic regions affecting gene expression (Kendziorski et al., 2006; Sillanpää and Noykova, 2008; Chun and Keles, 2009). Using DNA microarray or deep sequencing it is now possible to measure both genetic variation and gene expression simultaneously. Hence, theoretically, eQTL analysis essentially allows for the study of the association of all regions in a genome with the expression of all genes. In that respect, an eQTL study can be viewed as a generalization of a traditional quantitative trait loci (QTL) analysis, but with a large number of phenotypes (gene expressions).

The basic statistical modeling framework to analyze eQTL data for understanding the genetic basis of gene regulation is to (i) treat the high-dimensional set of gene expression as multiple, possibly correlated, responses and (ii) identify their assoication with the genetic markers by treating them as covariates. Joint analyses of these high-dimensional data have been performed using multivariate regression models where the responses are quantitative measures of gene expression abundances and the predictors encode DNA sequence variation at a number of loci. Traditional multivariate regression models fail in these situations due to the high dimensionality of these problems which frequently exceeds the sample size. Existing Bayesian approaches (Yi and Shriner, 2008) are mainly based on the concept of covariate selection (Jia and Xu, 2007; Monni and Tadesse, 2009; Richardson et al., 2010; Petretto et al., 2010) to identify the markers with evidence of enhanced linkage (hot spots).

Gaussian graphical models have been applied to infer the interaction relationship among genes at the transcriptional level (Schäfer and Strimmer, 2005; Segal et al., 2005; Peng et al., 2009). Conditional independence arising out of a Gaussian graphical model (i.e., pattern of

expression for a given gene can be predicted by only a small subset of other genes) induces sparisity in the inverse covariance matrix which is crucial to analyze such high-dimensional data. However, gene expression data alone is unable to fully capture the gene activities. We note that eQTL data also contains essential information about gene regulation and have been used to infer the interaction network among genes (Zhu et al., 2004; Chen et al., 2007). Incorporating available covariates such as information on genetic variants from an eQTL study results in improved estimation of the graphical model (Yin and Li, 2011).

Sparse variable selection has recently received a lot of attention in statistics literature, especially with the emergence of high-dimensional data in fields like genetics or finance and several popular approaches, both frequentist, e.g., LASSO of Tibshirani (1996), elastic net of Zou and Hastie (2005), and Bayesian, e.g., Geweke (1996), stochastic search variable selection (SSVS) of George and McCulloch (1993) have been developed. Covariance matrix selection, indicating the selection of the sparse subset of non-zero inverse covariance matrix elements, for the case of a Gaussian graphical model, also has a long history, going back to Dempster (1972) and recent work in this field include Meinhausen and Bühlmann (2006) and Carvalho and Scott (2009). While the two problems (covariate and covariance selections) separately have a long history in the statistics literature, a joint analysis has, however, been lacking. Therefore, in this paper, we propose a joint sparse modeling approach for the responses (gene expressions) as well as covariates (genetic markers). This joint model simultaneously performs a Bayesian selection of (a) significant covariates and (b) the significant entries in the adjacency matrix of the correlated responses. Furthermore, our method is well equipped to work in a high-dimesional framework. In frequentist setting, similar joint modeling has been recently attempted by Yin and Li (2011) and Cai et al. (2011).

We adopt seemingly unrelated regression (SUR) (Zellner, 1962) framework where multiple predictors affect multiple responses, just as in the case of a multivariate regression, with the

additional complication that the response variables exhibit a correlation structure (which is typically assumed to be the same for all the samples). Bayesian variable selection techniques have been developed in the SUR context by Holmes et al. (2002) and Smith and Kohn (2000). In both the papers, a relatively modest dimension of the responses has been considered, ignoring any sparsity or graphical structure in the error inverse covariance matrix. In our problem frequently the number of predictors (p) and the number of correlated responses (q) can be much larger than the sample size (n). Also, we consider SUR models in which both regression coefficients and the error inverse covariance matrix have many zeroes. Zeroes in regression coefficients arise when each gene expression only depends on a very small set of genetic markers. Similarly, zeroes in inverse covariance matrix arise due to sparse conditional independence structure of the Gaussian gene interaction network. In this setting, it is necessary to induce sparsity in both the regression coefficients and the elements of the inverse covariance matrix while performing regression (Rothman et al., 2010; Wang, 2010). We achieve this via a combination of variable selection and Gaussian graphical modeling techniques.

Joint variable selection and inverse covariance estimation methods are challenging to implement, specifically in high dimensions. In a recent example Yin and Li (2011) apply independent L_1 regularized selection of the significant SNPs and the significant entries of the inverse covariance matrix of the responses, following the approach described by Rothman et al. (2010). In this approach one estimates jointly the vector of regression coefficients and the inverse covariance matrix, imposing sparsity in both. However, numerical instability is often encountered in high-dimensional maximization problems that are inherent in classical L_1 regularization-based techniques. Also, often the interest lies in identifying the non-zero elements in the vector of regression coefficients and in the inverse covariance matrix, rather than knowing their exact value. Keeping this in mind, we propose a Bayesian hierarchical

formulation that allows us to analytically integrate out the nuisance parameters of the actual regression coefficient and the covariance matrix, leaving in the model the parameters of interest - a vector of indicators denoting whether a particular predictor (SNP) is important and an adjacency matrix (equivalently, a graph) of the responses indicating the zeroes in the inverse covariance matrix of the responses. This marginalization based collapsed Gibbs sampling plays a key role in the computational efficiency of our method. The quantities that are integrated out from the model can always be regenerated in the posterior when an association analysis is desired.

The rest of the manuscript is organized as follows: In section 2, we describe our Bayesian hierarchical model using the notation of the matrix-variate normal distribution of Dawid (1981) and outline the MCMC search strategy. In section 3, we use our technique on two simulated data sets - one where we simulate directly from our proposed hierarchical model of section 2 and another where we use the related simulated data set of Richardson et al. (2010), which studies a similar problem but where our model assumptions are violated slightly. We show that we can successfully infer the significant predictors and the graph, as well as performing the required association analysis for eQTL mapping. In section 4, we apply the technique on a real dataset where the SNP information and gene expression data are publicly available on the International HapMap project and the Sanger Institute websites respectively. Finally, we conclude by proposing some extensions of the current work in section 5.

2. The model

We consider the following model

$$\mathbf{Y} = \mathbf{X}_{\boldsymbol{\gamma}}\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{Y}$ is an $n \times q$ matrix of standarized gene expression data for n individuals, for the same q genes; $\mathbf{X}_{\gamma}$ is an $n \times p_{\boldsymbol{\gamma}}$ is a matrix of predictors encoding of $p_{\boldsymbol{\gamma}}$ genetic markers; $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ is a $p_{\boldsymbol{\gamma}} \times q$ matrix of regression coefficients. Let p be the total number of predictors, of which

only a sparse subset of length $p_{\boldsymbol{\gamma}}$ is present in the model. Thus we need a vector of indicators $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)$. Here γ_i is 1 if the ith predictor is present in the model and 0 otherwise. Therefore, $p_{\boldsymbol{\gamma}} = \sum_{i=1}^{p} \boldsymbol{\gamma}_i$.

We further assume that $\boldsymbol{\epsilon}$ follows a zero mean matrix-variate normal distribution (Dawid, 1981), denoted as $\mathrm{MN}_{n\times q}(\mathbf{0}, \mathbf{I}_n, \boldsymbol{\Sigma}_{\mathbf{G}})$, where $\mathbf{0}$ is an $n \times q$ matrix of zeroes, $\boldsymbol{\Sigma}_{\mathbf{G}}$ is the $q \times q$ covariance matrix of q possibly correlated responses. Here and in the rest of the manuscripts $\mathbf{I}_k$ denotes an identity matrix of size k.

This model is similar to the SUR model, where the goal is to estimate the regression parameters $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ by borrowing gene expression information among different genes. In eQTL studies, each row of $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ is assumed to be sparse since early studies in the genetics of gene expression indicate that the transcripts are expected to have only a few genetic regulators (Schadt et al., 2003; Brem and Kruglyak, 2005). The inverse covariance matrix $\boldsymbol{\Sigma}_{\mathbf{G}}^{-1}$ is also expected to be sparse, since typical genetic networks have limited links (Leclerc, 2008). We allow $\mathbf{G}$ to be a symmetric, undirected, decomposable graph and presence of an off-diagonal edge in $\mathbf{G}$ indicate conditional dependence in the corresponding two response variables given the rest. Maximum possible number of off-diagonal edges in $\mathbf{G}$ is $q(q-1)/2$. The goal is now a joint sparse estimation of both $\boldsymbol{\gamma}$ and $\mathbf{G}$ - which would respectively tell us the predictors that are marginally significant (considering all the responses) and the gene interaction network, after accounting for the effect of the genetic variability (the predictors). In order to perform an eQTL analysis, conditional on $\boldsymbol{\gamma}$ and $\mathbf{G}$, one can then sample the matrix $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$, the entries of which, if away from zero, indicate a non-zero association between the respective SNPs and the genetic transcripts.

2.1 *Baysian Gaussian graphical models*

In a Bayesian setting, Gaussian graphical modeling is based on hierarchical specifications for the covariance matrix (or inverse covariance matrix) using global conjugate priors in the

space of positive-definite matrices, such as inverse Wishart priors or its equivalents. Dawid and Lauritzen (1993) introduced an equivalent form as the hyper-inverse Wishart (HIW) distribution. The construction for decomposable graph enjoys many advantages, such as computational efficiency due to its conjugate formulation and exact calculation of marginal likelihoods (Carvalho and Scott, 2009; Jones et al., 2004). Below, we describe some notations and previously known results which have been used in the rest of the paper.

We specify an undirected graph $\mathbf{G} = (V, E)$ where V is a set of vertices and $E = (i, j)$ is a set of edges for $i, j \in V$. A graph (or, a subgraph) is termed as complete if all its vertices are connected. Given this set of complete subgraphs, a clique is defined as a complete subgraph which is not completely a part of another subgraph (Carvalho et al., 2007). A decomposable graphs is one which can be split into a set of cliques $C_1, \ldots, C_k$ (Lauritzen, 1996). A clique is thus a complete maximal subset of a graph. Define $H_{j-1} = C_1 \cup \ldots \cup C_{j-1}$ and $S_j = H_{j-1} \cap C_j$. Then S_j s are called the separators.

Now, following the notation and Proposition 5.2 of Lauritzen (1996), we have the following fact about Gaussian graphical models: If $Y = (Y_\nu)_{\nu \in V}$ is a random vector in $\mathbb{R}^{|V|}$ that follows a multivariate normal distribution with mean vector $\boldsymbol{\xi}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{G}}$, then

$$Y_\nu \perp Y_\mu | Y_{V \setminus \{\nu,\mu\}} \iff \omega_{\nu\mu} = 0,$$

where $\boldsymbol{\Sigma}_{\mathbf{G}}^{-1} = \{\omega_{\nu\mu}\}_{\nu,\mu \in V}$ is the inverse of the covariance matrix. Thus, the elements of the adjacency matrix of the graph $\mathbf{G}$ have a very specific interpretation, in the sense that they model conditional independence among the components of the multivariate normal. Presence of an off-diagonal edge in the graph indicates non-zero partial correlation.

The advantage of having a decomposable graph allows us to split the density of the multivariate normal into products and ratios over the set of cliques and separators. The clique finding problem is polynomial time for decomposable graphs but is known to be NP-complete for a general graph and thus we restrict ourselves to decomposable graphs for this

article. If the data are denoted by y, then from section 5.3 of Lauritzen (1996),

$$f(y) \;=\; \frac{\prod_{j=1}^{k} f(y_{C_j})}{\prod_{j=2}^{k} f(y_{S_j})}, \tag{2}$$

where $f(\cdot)$ denotes a generic density. Given this graph $\mathbf{G}$, Dawid and Lauritzen (1993) derived a conjugate prior distribution for the covariance matrix $\mathbf{\Sigma_G}$, which they termed the hyper-inverse Wishart (HIW) distribution. Specifically, if q dimensional i.i.d random variables $Y_i \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma_G})$ for $i = 1, \ldots, n$ and $\mathbf{\Sigma_G} \sim \mathrm{HIW_G}(\delta, \mathbf{\Phi})$ is the prior law for a positive integer δ and a positive definite $q \times q$ matrix $\mathbf{\Phi}$, then the posterior is $\mathbf{\Sigma_G}|\mathrm{Y} \sim \mathrm{HIW_G}(\delta + n, \mathbf{\Phi} + \mathbf{Y'Y})$, where $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ is an $n \times q$ matrix. This particular distribution plays a key role in our model in sections 2.2 and 2.5.

2.2 *Hierarchical model and MCMC computation*

Note that the model in equation (1) is equivalent to

$$\mathrm{Vec}(\mathbf{Y} - \mathbf{X_\gamma B_{\gamma,G}})' \;\sim\; \mathrm{N}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Sigma_G}),$$

where $\mathrm{N}(\boldsymbol{\mu}, \mathbf{\Psi})$ denotes a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Psi}$. Here $\otimes$ denotes the Kronecker product and $\mathbf{A'}$ denotes the transpose of a matrix $\mathbf{A}$. The operator $\mathrm{Vec}(\cdot)$ forms a column vector by stacking the columns of a matrix on one another. Note that in this formulation the individuals (n) are uncorrelated in our model, which is reasonable if the data are not familial, but through the $q \times q$ covariance matrix $\mathbf{\Sigma_G}$, we allow the responses to be correlated within each individual, which is reasonable in an eQTL analysis where the phenotypic traits (responses) can be correlated. The complete hierarchical model is now specified as:

$$(\mathbf{Y} - \mathbf{X_\gamma B_{\gamma,G}})|\mathbf{B_{\gamma,G}}, \mathbf{\Sigma_G} \;\sim\; \mathrm{MN}_{n\times q}(\mathbf{0}, \mathbf{I}_n, \mathbf{\Sigma_G}), \tag{3}$$

$$\mathbf{B_{\gamma,G}}|\boldsymbol{\gamma}, \mathbf{\Sigma_G} \;\sim\; \mathrm{MN}_{p_\gamma \times q}(\mathbf{0}, c\mathbf{I}_{p_\gamma}, \mathbf{\Sigma_G}), \tag{4}$$

$$\mathbf{\Sigma_G}|\mathbf{G} \;\sim\; \mathrm{HIW_G}(b, d\mathbf{I}_q), \tag{5}$$

$$\gamma_i \overset{\text{i.i.d}}{\sim} \text{Bernoulli}(w_\gamma) \quad \text{for} \quad i = 1, \ldots, p, \tag{6}$$

$$\text{G}_k \overset{\text{i.i.d}}{\sim} \text{Bernoulli}(w_G) \quad \text{for} \quad k = 1, \ldots, q(q-1)/2, \tag{7}$$

$$w_\gamma, w_G \sim \text{Uniform}(0, 1), \tag{8}$$

where b, c, d are fixed, positive hyper-parameters and w_γ and w_G are prior weights that control the sparsity in $\boldsymbol{\gamma}$ and $\mathbf{G}$ respectively. We denote by γ_i and G_k the indicators for the ith element for the vector $\boldsymbol{\gamma}$ and the kth off-diagonal edge in the lower triangular part of the adjacency matrix of the graph $\mathbf{G}$. The diagonal elements in the adjacency matrix of $\mathbf{G}$ are always restricted to be 1, since a zero on the diagonal violates the positive definiteness of $\boldsymbol{\Sigma}_\mathbf{G}$.

Equations (4) and (5) specify the prior for the mean and covariance respectively. Priors of the type of equation (4) is also known as a "spike and slab" prior (George and McCulloch, 1993). With p predictors there is total of 2^p possible models, corresponding to the presence or absence of each individual predictor. The spike and slab type prior allows for the use of a set of p_γ predictors out of a possible p, reducing the number of "active" covariates during a given MCMC iteration. Note in equation (4) that the matrix $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ is assumed a priori independent in the direction of the SNPs but not in the direction of the responses. An alternative will be to use Zellner's g-prior (Zellner, 1986) with the formulation $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}} \sim \text{MN}_{p_\gamma \times q}(\mathbf{0}, c(\mathbf{X}_{\boldsymbol{\gamma}}\mathbf{X}'_{\boldsymbol{\gamma}})^{-1}, \boldsymbol{\Sigma}_{\boldsymbol{G}})$. Now, from equation (4) we have

$$\mathbf{X}_{\boldsymbol{\gamma}}\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}|\boldsymbol{\gamma}, \boldsymbol{\Sigma}_\mathbf{G} \sim \text{MN}_{n\times q}(\mathbf{0}, c(\mathbf{X}_{\boldsymbol{\gamma}}\mathbf{X}'_{\boldsymbol{\gamma}}), \boldsymbol{\Sigma}_\mathbf{G}).$$

Using this in equation (3) we get

$$\mathbf{Y}|\boldsymbol{\gamma}, \boldsymbol{\Sigma}_\mathbf{G} \sim \text{MN}_{n\times q}(\mathbf{0}, \mathbf{I}_n + c(\mathbf{X}_{\boldsymbol{\gamma}}\mathbf{X}'_{\boldsymbol{\gamma}}), \boldsymbol{\Sigma}_\mathbf{G}),$$

which has the effect of integrating out $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$. Now, let $\mathbf{A}$ be the Cholesky decomposition of

the matrix $\{\mathbf{I}_n + c(\mathbf{X}_{\boldsymbol{\gamma}}\mathbf{X}'_{\boldsymbol{\gamma}})\}^{-1}$, i.e.,

$$\mathbf{A}\mathbf{A}' \;=\; \{\mathbf{I}_n + c(\mathbf{X}_{\boldsymbol{\gamma}}\mathbf{X}'_{\boldsymbol{\gamma}})\}^{-1}.$$

The Cholesky decomposition exists for positive values of c. Then, if we define $\mathbf{T} = \mathbf{A}\mathbf{Y}$, we have

$$\mathbf{T}|\boldsymbol{\gamma}, \boldsymbol{\Sigma}_{\mathbf{G}} \;\sim\; \mathrm{MN}_{n\times q}(\mathbf{0}, \mathbf{I}_n, \boldsymbol{\Sigma}_{\mathbf{G}}). \tag{9}$$

Noting the definition of the prior for $\boldsymbol{\Sigma}_{\mathbf{G}}$ from equation (5), integrating out $\boldsymbol{\Sigma}_{\mathbf{G}}$ from equations (9) and (5) gives rise to the hyper-matrix t distribution of Dawid and Lauritzen (1993), and we get

$$\mathbf{T}|\boldsymbol{\gamma}, \mathbf{G} \;\sim\; \mathrm{HMT}_{n\times q}(b, \mathbf{I}_n, d\mathbf{I}_q), \tag{10}$$

where $\mathrm{HMT}(\cdot)$ denotes the hyper-matrix t distribution. This is a special type of t distribution which, given the graph, splits into products and ratios over the cliques and separators as in equation (2). More specifically, given n observations and the graph $\mathbf{G}$, we know the sequence of cliques $C_1, \ldots, C_k$ and separators $S_2, \ldots S_k$. For any $A \subseteq C_i$, we choose the nodes in A and write T^n_A as the corresponding $(n \times |A|)$ matrix where $|A|$ denotes the cardinality of the set A. Then the hyper-matrix t density on a given clique C_j can be written as

$$\begin{aligned} f(\mathbf{t}^n_{C_j}) \;=\; & \pi^{-n/2}\frac{\Gamma_{|C_j|}((b+n+|C_j|-1)/2)}{\Gamma_{|C_j|}((b+|C_j|-1)/2)}\{\det(d\mathbf{I}_{|C_j|})\}^{-n/2} \\ & \times[\det\{\mathbf{I}_n + (\mathbf{t}^n_{C_j})(d\mathbf{I}_{|C_j|})^{-1}(\mathbf{t}^n_{C_j})'\}]^{-(b+n+|C_j|-1)/2}, \end{aligned}$$

by equation (45) of Dawid and Lauritzen (1993). The densities on the separators can be written similarly. The overall density is now given by an application of equation (2) as

$$f(\mathbf{t}^n) \;=\; \frac{\prod_{j=1}^k f(\mathbf{t}^n_{C_j})}{\prod_{j=2}^k f(\mathbf{t}^n_{S_j})}. \tag{11}$$

Note that now we have effectively integrated out both $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ and $\boldsymbol{\Sigma}_{\mathbf{G}}$. Thus the joint search

for the sparse set of predictors and sparse set of inverse covariance matrix elements can just cycle between $\boldsymbol{\gamma}$ and $\mathbf{G}$, provided we can evaluate the HMT density in a computationally efficient manner. We accomplish this via MCMC using random addition/deletion of variables (or edges, in the case of the graph) for both $\boldsymbol{\gamma}$ and $\mathbf{G}$. This collapsed Gibbs sampler works efficiently in high-dimensional problems.

2.3 *MCMC for* $\boldsymbol{\gamma}$ *given* $\mathbf{G}$ *and* $\mathbf{T}$

We choose to work with a simple addition/deletion of variables for $\boldsymbol{\gamma}$. We have $p(\boldsymbol{\gamma}|w_\gamma) \propto \prod_{i=1}^{p} w_\gamma^{\gamma_i}(1-w_\gamma)^{(1-\gamma_i)}$, where γ_i is the ith element of the vector $\boldsymbol{\gamma}$ and from equation (8) we note $w_\gamma \sim \text{Uniform}(0,1)$. This gives $p(\boldsymbol{\gamma}) \propto \{(p+1)\binom{p}{p_\gamma}\}^{-1}$ as the prior law after integrating out w_γ where $p_\gamma = \sum_{i=1}^{p} \gamma_i$.

(1) Given the current $\boldsymbol{\gamma}$, propose $\boldsymbol{\gamma}^*$ by either (a) changing a non-zero entry in $\boldsymbol{\gamma}$ to zero with probability $(1-\alpha_\gamma)$ and set $q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^*)/q(\boldsymbol{\gamma}^*|\boldsymbol{\gamma}) = \alpha_\gamma/(1-\alpha_\gamma)$, or (b) changing a zero entry in $\boldsymbol{\gamma}$ to one, with probability α_γ and set $q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^*)/q(\boldsymbol{\gamma}^*|\boldsymbol{\gamma}) = (1-\alpha_\gamma)/\alpha_\gamma$.

(2) Calculate $f(\mathbf{t}|\boldsymbol{\gamma}^*, \mathbf{G})$ and $f(\mathbf{t}|\boldsymbol{\gamma}, \mathbf{G})$ where f denotes the HMT density of equation (11).

(3) Jump from $\boldsymbol{\gamma}$ to $\boldsymbol{\gamma}^*$ with probability

$$r(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*) \;=\; \min\left\{1, \frac{f(\mathbf{t}|\boldsymbol{\gamma}^*, \mathbf{G})p(\boldsymbol{\gamma}^*)q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^*)}{f(\mathbf{t}|\boldsymbol{\gamma}, \mathbf{G})p(\boldsymbol{\gamma})q(\boldsymbol{\gamma}^*|\boldsymbol{\gamma})}\right\}.$$

2.4 *MCMC for* $\mathbf{G}$ *given* $\boldsymbol{\gamma}$ *and* $\mathbf{T}$

Similar to $\boldsymbol{\gamma}$, we add or delete off-diagonal edges at random to $\mathbf{G}$ to perform the MCMC. We have $p(\mathbf{G}|w) \propto \prod_{k=1}^{q(q-1)/2} w_G^{G_k}(1-w_G)^{(1-G_k)}$, where G_k is the kth off-diagonal element in lower triangular part of the (symmetric) adjacency matrix of $\mathbf{G}$. The diagonal elements are always 1 to ensure non-zero variance. We again note from equation (8) that $w_G \sim \text{Uniform}(0,1)$ and therefore, similar, to section 2.3, integrating out w_G gives $p(\mathbf{G}) \propto \left[\{q(q+1)/2\}\binom{q(q+1)/2}{r_G}\right]^{-1}$, where $r_G = \sum_{k=1}^{q(q-1)/2} G_k$.

(1) Given the current decomposable graph $\mathbf{G}$, propose a decomposable graph $\mathbf{G}^*$ by either

(a) changing a non-zero off-diagonal entry in the lower triangular part of $\mathbf{G}$ to zero with probability $(1-\alpha_G)$ and set $q(\mathbf{G}|\mathbf{G}^*)/q(\mathbf{G}^*|\mathbf{G}) = \alpha_G/(1-\alpha_G)$, or (b) changing a zero off-diagonal entry in the lower triangular part of $\mathbf{G}$ to one, with probability α_G and set $q(\mathbf{G}|\mathbf{G}^*)/q(\mathbf{G}^*|\mathbf{G}) = (1-\alpha_G)/\alpha_G$. It is understood that the corresponding upper triangular elements are also changed to maintain the symmetry of the adjacency matrix.

(2) Calculate $f(\mathbf{t}|\mathbf{G}^*, \boldsymbol{\gamma})$ and $f(\mathbf{t}|\mathbf{G}, \boldsymbol{\gamma})$ where f denotes the HMT density of equation (11).

(3) Jump from $\mathbf{G}$ to $\mathbf{G}^*$ with probability

$$r(\mathbf{G}, \mathbf{G}^*) \;=\; \min\left\{1, \frac{f(\mathbf{t}|\mathbf{G}^*, \boldsymbol{\gamma})p(\mathbf{G}^*)q(\mathbf{G}|\mathbf{G}^*)}{f(\mathbf{t}|\mathbf{G}, \boldsymbol{\gamma})p(\mathbf{G})q(\mathbf{G}^*|\mathbf{G})}\right\}.$$

2.5 *Regeneration of* $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ *in the posterior for association analysis*

Since the $p_\gamma \times q$ matrix of regression coefficients is integrated out during the analysis, the recovered predictors with high posterior probabilities are the globally significant ones. However, this does not provide information on which of the q responses each of these significant predictors is associated with. In order to perform an association study between the p predictors and the q responses, we can simulate $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ directly conditional on $\boldsymbol{\gamma}$ and $\boldsymbol{G}$ due to its conjugate structure:

- Generate $\boldsymbol{\Sigma}_{\boldsymbol{G}}|\mathbf{Y}, \mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}, \boldsymbol{\gamma}, \boldsymbol{G}$ from $\text{HIW}_{\boldsymbol{G}}\{b+n, d\mathbf{I}_q + (\mathbf{Y}-\mathbf{X}_{\boldsymbol{\gamma}}\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}})'(\mathbf{Y}-\mathbf{X}_{\boldsymbol{\gamma}}\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}})\}$.
- Generate $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}|\mathbf{Y}, \boldsymbol{\Sigma}_{\boldsymbol{G}}, \boldsymbol{\gamma}, \boldsymbol{G}$ from $\text{MN}_{p_\gamma \times q}\{(\mathbf{X}'_{\boldsymbol{\gamma}}\mathbf{X}_{\boldsymbol{\gamma}}+c^{-1}\mathbf{I}_{p_\gamma})^{-1}\mathbf{X}'_{\boldsymbol{\gamma}}\mathbf{Y}, (\mathbf{X}'_{\boldsymbol{\gamma}}\mathbf{X}_{\boldsymbol{\gamma}}+c^{-1}\mathbf{I}_{p_\gamma})^{-1}, \boldsymbol{\Sigma}_{\boldsymbol{G}}\}$.

where the starting values for $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{G}}$ can be chosen from the respective priors.

We may sample $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{G}}$ this way in each iteration of the MCMC for association analysis. However, it should be noted that while sampling $\boldsymbol{\gamma}$ in the posterior, the marginalized sampler in section 2.3 would perform better and conditioning on $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ is to be avoided due to the hierarchical structure of the model and dependence of $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ on $\boldsymbol{\gamma}$. An alternative and fast way to sample $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ is to generate the quantity conditional on $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{G}}$, where "hat" denotes some form of posterior summary statistic. Specifically, for the purpose of this article,

we determine the entries of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{G}}$ by controlling the Bayesian false discovery rate (FDR) on the posterior mean at a certain level, say 5%, using the method of Müller et al. (2007).

2.6 *Specification of hyperparameters*

The hyperparameters we need to specify in order to perform the MCMC are b, c, d, α_γ and α_G. A crucial parameter is c, which essentially is a global shrinkage parameter akin to ridge penalty in a frequentist analysis. This parameter needs to be chosen in a way so that the variability in the two series ($\mathbf{X}$ and $\mathbf{Y}$) are matched. We suggest taking for c the value $\text{Var}(\mathbf{Y})/\text{Var}(\mathbf{X})$, where variance is computed over all the $n \times q$ elements for $\mathbf{Y}$ and $n \times p$ elements for $\mathbf{X}$. Web Appendix A presents simulation results documenting the stability of the estimated graph $\mathbf{G}$ over a range of choice for c, justifying our guideline.

3. Simulation study

We study two simulation examples, one where we directly simulate from the model described in section 2 and another simulation example from Richardson et al. (2010), where our model assumptions are violated slightly, in order to verify the robustness of our method.

3.1 *Simulation example 1*

Here, we simulate directly from the model in section 2. We choose $p = 498$, $q = 300$ and $n = 120$. $\mathbf{X}$ is a 120×498 matrix of ones and zeroes, coding a binary SNP, taken from Richardson et al. (2010). Clearly this is a *large p, large q, small n* setting. The true predictors are {30, 40, 57, 62, 161, 239, 269, 322, 335, 399, 457} and thus true $p_\gamma = 11$. Figure 1 shows the true adjacency matrix for the graph $\mathbf{G}$. We choose $b = 10$, $c = 0.3^2$, $d = 1$.

[Figure 1 about here.]

We perform 100,000 MCMC iterations after a burn-in period of 50,000 and monitor the likelihood of the model to ascertain that around 100,000 MCMC iterations are enough for a reasonable search in the high-dimensional model space. In order to ensure good exploration, we perform independent parallel MCMC searches with initial values chosen at random. The

search is extremely fast, with about 2000 MCMC iterations/minute (implemented in the MATLAB programming language and running on an Intel Xeon 2.5 GHz processor) because (a) we successfully avoid having to sample the high-dimensional $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ and $\boldsymbol{\Sigma}_{\mathbf{G}}$ in the posterior in our collapsed Gibbs sampler, from a multivariate normal and a hyper inverse Wishart respectively, having integrated these parameters out of the model and (b) since $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ and $\boldsymbol{\Sigma}_{\mathbf{G}}$ depend on $\boldsymbol{\gamma}$ and $\mathbf{G}$, a full Gibbs sampler will suffer from poor mixing - if at all it is possible to design such a sampler for a high-dimensional problem like the one at hand. The only computationally intensive step in our search procedure is the evaluation of an HMT density in the MCMC steps 2.3.2 and 2.4.2. We, however, note that our MCMC procedure for $\boldsymbol{\gamma}$ and $\mathbf{G}$ in sections 2.3 and 2.4 are extremely simple in principle - in the sense that they search the model space via simple addition or deletion of variables (or edges) in the model, i.e, at any given MCMC iteration we only change one variable at a time via a local move of the Markov chain. We anticipate that, in principle, ideas similar to parallel tempering (Geyer, 1991) or stochastic approximation Monte Carlo (Liang et al., 2007), where the Markov chain is allowed both local and global moves, will result in an even more efficient search. However, we will not focus on these more advanced MCMC techniques further in this article, noting that our simple MCMC appears to work well.

For an element of $\boldsymbol{\gamma}$ and $\mathbf{G}$, given a threshold on the posterior probability for inclusion in the model, we first determine the true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) in this simulation example. Define the true positive rate (TPR) and the false positive rate (FPR) as,

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}}.$$

[Figure 2 about here.]

A plot of TPR and FPR as we vary the decision threshold on the posterior probability of inclusion is called the receiver operating characteristic (ROC). In Figures 2(a) and (b)

we present the ROC curves corresponding to $\boldsymbol{\gamma}$ and $\mathbf{G}$ respectively. As we noted in the introduction of this paper, most techniques developed so far for high-dimensional analysis treat the variable and the covariance selections in a model separately. Therefore, an important question to address is whether the "joint" selection of variables and covariance structures offers any advantage over the case where these two were to be estimated separately. In Figure 2(a), the solid blue line corresponds to the ROC for $\boldsymbol{\gamma}$ where the joint selection is performed and the broken line corresponds to the ROC where the responses are assumed to be independent and hence the adjacency matrix $\mathbf{G}$ is a fixed diagonal matrix and is not updated. With a diagonal matrix for the response graph, the model is identical to section 3.1 of Richardson et al. (2010). Comparing the ROC curves, we note that the joint selection performs better. Similarly, in Figure 2(b), the solid blue line corresponds to the ROC curve for $\mathbf{G}$ where both $\boldsymbol{\gamma}$ and $\mathbf{G}$ are updated and the broken blue line is the ROC curve for a zero mean model (i.e. there are no predictors). Covariance selection for a similar, zero mean model has been considered by Carvalho and Scott (2009). We again see the joint selection is advantageous in this setting.

The plot of posterior marginal probabilities of inclusion for the predictors (Figure 2(c)) correctly identifies most of the eleven true covariates as having high posterior probabilities (circled in red). One must note from Figure 1 that the number of true non-zero elements in the inverse covariance matrix is much higher than the available sample size. This is different from the true number of predictors, which is much smaller than the sample size. Therefore, indeed it is a strength of our method that it can recover the true adjacency structure in this very high-dimensional setting. Figure 2(d) shows the plot of of the posterior probabilities for $\mathbf{G}$ on a gray scale, and the true structure is clearly identifiable as a region of higher posterior probability at the bottom right corner of the plot.

3.2 *Simulation example 2*

We study the simulation examples of Richardson et al. (2010), section 6. The setting for their first simulation experiment is as follows. There are $q = 100$ response variables or genetic transcripts, $p = 498$ SNPs and $n = 120$ samples. This is again a *large p, large q, small n* setting. Table 1 shows the responses that are affected by a given SNP in the simulation example entitled "Sim 1" in Richardson et al. (2010). We first select a set of SNPs $\tilde{p} \in \{1, \ldots, p\}$ and a set of responses $\tilde{q}_p \in \{1, \ldots, q\}$ corresponding to each element of the vector $\tilde{p}$, in order to let a single SNP affect multiple responses in Table 1.

[Table 1 about here.]

Thus, we see there are groups of transcripts that are affected by one or more of the SNPs, e.g., transcripts 17-20 are affected by SNP 30, 161 and 239 simultaneously. The responses are now simulated with $\beta_{\tilde{p},\tilde{q}_p} \sim N(0, 0.3^2)$ for $\tilde{p}$ and $\tilde{q}_p$ listed in Table 1 and noise terms are simulated from a $N(0, 0.1^2)$. The rest of the simulated responses are just $N(0, 0.1^2)$ noise terms. The predictor matrix $\mathbf{X}$ is the same as section 3.1. A simultaneous analysis of significant variables and adjacency matrix elements should now identify the groups of correlated transcripts (those which are afftected by the same SNP) as well as the set of significant predictors.

Figure 3(a) plots the marginal posterior probabilities for the elements of $\boldsymbol{\gamma}$, with SNPs 30, 161 and 239 circled in red. Thus we see the true SNPs again have high marginal posterior probabilities of inclusion in the model. ROC curve for $\boldsymbol{\gamma}$ looked very similar to Simulation example 1. Note that here we do not know the true adjacency matrix since the simulation setting differs from our model assumptions, but, we will still expect the transcripts that are affected by the same SNP to be exhibit correlation among themselves. Figure 3(b) shows the recovered adjacency matrix where the cutoff on the posterior probability for inclusion was set to 0.4. Comparing Fig 3(b) with Table 1, we see that in this case we are able to recover the correlation structure among the transcripts as well.

[Figure 3 about here.]

Similar to section 3.1, our search procedure for both $\boldsymbol{\gamma}$ and $\mathbf{G}$ again proceeds via simply adding and deleting edges at random. This simple procedure is at a contrast with the more involved Hierarchical Evolutionary Stochastic Search (HESS) of Richardson et al. (2010), where an idea similar to simulated tempering is followed. There are multiple Markov chains with different temperatures (i.e., different state transition probabilities) and there are swapping between these chains. However, unlike Richardson et al. (2010), our method has the added benefit of borrowing the conditional dependence information among the responses through the graph $\mathbf{G}$ while searching the model space for $\boldsymbol{\gamma}$ and vice versa.

For this example, we are now well suited to study if we can recover the $p_\gamma \times q$ matrix of regression coefficients $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ to reflect the association between the predictors and the responses, since we know these associations (between the SNPs and genetic transcripts) from Table 1. We take the approach described in section 2.5 and simulate $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ in the posterior conditional on $\hat{\boldsymbol{\gamma}}$ and $\hat{\mathbf{G}}$ chosen to restrict the FDR on their respective posterior means below 5% (Figures 3(a) and (b)). Basically, the inferred vector $\boldsymbol{\gamma}$ tells us the predictors that are globally significant (for one or more responses), however we can make a statement concerning the association between p SNPs and the q transcripts only through this recovered matrix $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$.

As an example, Figures 3(c) and (d) show the plots of $\mathbf{B}_{161,\cdot}$ and $\mathbf{B}_{239,\cdot}$ respectively, i.e., it shows the association of SNP 161 and 239 with all the 100 transcripts. We see that $\mathbf{B}_{161,17-20}$ in Figure 3(c) and $\mathbf{B}_{239,1-20}$ and $\mathbf{B}_{239,71-80}$ Figure 3(d) to be away from zero, compared to the rest. Formally, we can compute the mean and standard deviations for the elements of $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ over posterior samples and then an association is declared significant via a t-test accounting for multiple testing involved. The dashed horizontal lines in Figures 3(c) and (d) are the mean and the Bonferroni corrected 95% confidence intervals, giving us upper and lower bounds for declaring an association significant.

4. eQTL analysis on publicly available human data

We focus on a human eQTL analysis to demonstrate the effectiveness of the methodology proposed in this article. We consider 60 unrelated individuals of Northern and Western European ancestry from Utah (CEU), whose genotypes are available from the International Hapmap project and publicly downloadable via the hapmart interface (`http://hapmart.hapmap.org`). The genotype is coded as 0,1 and 2 for homozygous rare, heterozygous and homozygous common alleles. We focus on the SNPs found in the 5' UTR (untranslated region) of mRNA (messenger RNA) with a minor allele frequency of 0.1 or more. The UTR possibly has an important role in the regulation of gene expression and has previously been subject to investigation by Chen et al. (2008). In total, there are 3125 SNPs. The gene expression data of these individuals were analyzed by Stranger et al. (2007). There were four replicates for each individual. The raw data were background corrected and then quantile normalized across replicates of a single individual and then median normalized across all individuals. The gene expression data are again publicly available from the Sanger Institute website (`ftp://ftp.sanger.ac.uk/pub/genevar`). Out of the 47293 total available probes corresponding to different Illumina TargetID, we select the 100 most variable probes - each corresponding to a different transcript. Supplementary Table 1 provides a list of the corresponding TargetID for these 100 probes. Thus, here, $n = 60$, $p = 3125$ and $q = 100$.

Controlling for FDR at 5% level on the posterior probabilities of $\boldsymbol{\gamma}$ yields 8 globally significant SNPs. These are `rs2285635 (1421), rs12026825 (2735), rs757124 (3057), rs 2205912 (3058), rs 929762 (2788), rs 3810711 (2543), rs 708463 (2921), rs 7770751 (2653)`. These are listed in the order of decreasing posterior probability and the numbers in the parentheses denote the serial number of the SNP among all the 3125 SNPs listed in Supplementary Table 2.

Conditional on the inferred $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{G}}$ we simulate $\mathbf{B}_{\boldsymbol{\gamma},\mathbf{G}}$ in the posterior and determine the

significant associations via a t-test. Table 2 provides a list of significant associations of the 8 SNPs identified above with the corresponding transcripts.

[Table 2 about here.]

[Figure 4 about here.]

Thus, in Table 2 there is a total of 43 associations detected between the SNPs and the transcripts. Chen et al. (2008) detected a slightly higher number of associations using a frequentist p-value thresholding approach and considering both the 3' and 5' UTRs simultaneously. The inferred graph has a total of 55 edges and is shown in Figure 4. The TargetIDs correspond to unique gene locations and can be queried in standard bioinformatics databases.

5. Discussion

This paper discusses an approach of joint Bayesian selection of the significant predictor variables and inverse covariance matrix elements of the response variables in a computationally feasible way. The two problems are separately known in literature as variable selection and covariance selection. The marginalization-based collapsed Gibbs sampler is shown to be effective in high dimensions. Association analysis, where the dimensions of both the predictors and responses are high compared to the sample size, is also discussed in the context of eQTL analysis. An immediate extension of the current work is to deploy more advanced MCMC schemes in sections 2.3 and 2.4 to include possibilities of global move of the Markov chain. This will result in a more thorough search of the model space. Apart from eQTL analysis, there are numerous problems that arise in finance, econometrics and biological sciences where sparse SUR can be a useful approach to modeling and therefore, we expect our inference procedure to be effective. We have only concerned ourselves with a linear Gaussian model in this article, in order to take advantage of conjugate structures of the priors. As long as marginalization is feasible, relaxing the linearity assumption is relatively straightforward. As an example, analytically integrating out the regression coefficients should still be possible if the covariates enter the model through a spline function and thus the

current work can be extended to include flexible nonparametric techniques. One should note that the problem dimension for inferring the covariance graph for a given sample size n scales as $O(q^2)$ where q is the number of correlated traits. Therefore, our approach is useful if one is interested in inferring the interaction among a modest number of traits. However, the raw number of traits our method can handle is inherently smaller compared to methods that assume the traits to be independent (Kendziorski et al., 2006; Bottolo et al., 2011) or assume a dependence structure through a covariance matrix $\mathbf{\Sigma}$ (but no underlying graph structure) but then integrate out $\mathbf{\Sigma}$ (Petretto et al., 2010), thus not inferring $\mathbf{G}$. How well the method performs is also a factor of how sparse the true covariance structure is and it is natural to expect that for a given sample size n, performance degrades for a denser graph.

6. Supplementary Materials

Web Appendix A: Additional simulation results referenced in section 2.6; Web Table 1: The 100×60 matrix of gene expression data for 100 transcripts (with unique Illumina TargetID) for each of the 60 CEU individuals & Web Table 2: the 3125 5' UTR SNPs considered for the 60 CEU individuals listed by their rsID, referenced in section 4, are available with this paper at the Biometrics website on Wiley Online Library.

Acknowledgements

We thank Biometrics co-editor Thomas A. Louis, the associate editor and two anonymous referees for valuable suggestions. This publication is based on work supported by Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST).

References

Bottolo, L., Petretto, E., Blankenberg, S., Cambien, F., Cook, S. A., Tiret, L., and Richardson, S. (2011). Bayesian detection of expression quantitative trait loci hot spots. *Genetics*

189, 1449–1459.

Brem, R. B. and Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 1572–1577.

Cai, T., Li, H., Liu, W., and Xie, J. (2011). Covariate adjusted precision matrix estimation with an application in genetical genomics. *Technical Report, The Wharton School Statistics Department* http://www-stat.wharton.upenn.edu/∼tcai/paper/CAPME.pdf,.

Carvalho, C. M., Massam, H., and West, M. (2007). Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika* **94,** 647–659.

Carvalho, C. M. and Scott, J. G. (2009). Objective Bayesian model selection in Gaussian graphical models. *Biometrika* **96,** 497–512.

Chen, L., Emmert-Streib, F., and Storey, J. (2007). Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biology* **8,** R219.

Chen, L., Tong, T., and Zhao, H. (2008). Considering dependence among genes and markers for false discovery control in eqtl mapping. *Bioinformatics* **24,** 2015–2022.

Chun, H. and Keles, S. (2009). Expression quantitative trait loci mapping with multivariate sparse partial least squares regression. *Genetics* **182,** 79–90.

Dawid, A. P. (1981). Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika* **68,** 265–274.

Dawid, A. P. and Lauritzen, S. L. (1993). Hyper markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics* **21,** 1272–1317.

Dempster, A. P. (1972). Covariance selection. *Biometrics* **28,** 157–175.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88,** 881–889.

Geweke, J. (1996). Variable selection and model comparison in regression. In Bernardo,

J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 5*, pages 609–620. Oxford University Press.

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics. Proceedings of the 23rd Symposium on the Interface*, pages 156–163. Interface Foundation of North America.

Holmes, C. C., Denison, D. G. T., and Mallick, B. K. (2002). Accounting for model uncertainty in seemingly unrelated regressions. *Journal of Computational and Graphical Statistics* **11,** 533–551.

Jia, Z. and Xu, S. (2007). Mapping quantitative trait loci for expression abundance. *Genetics* **176,** 611–623.

Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C., and West, M. (2004). Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science* **20,** 388–400.

Kendziorski, C. M., Chen, M., Yuan, M., Lan, H., and Attie, A. D. (2006). Statistical methods for expression quantitative trait loci (eqtl) mapping. *Biometrics* **62,** 19–27.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press.

Leclerc, R. D. (2008). Survival of the sparsest: robust gene networks are parsimonious. *Molecular Systems Biology* **4,** 213.

Liang, F., Liu, C., and Carroll, R. J. (2007). Stochastic approximation in Monte Carlo computation. *Journal of the American Statistical Association* **102,** 305–320.

Meinhausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Annals of Statistics* **34,** 1436–1462.

Monni, S. and Tadesse, M. G. (2009). A stochastic partitioning method to associate high-dimensional responses and covariates (with discussion). *Bayesian Analysis* **4,** 413–436.

Müller, P., Parmigiani, G., and Rice, K. (2007). FDR and Bayesian multiple comparisons

rules. In *Proceedings of the Valencia/ISBA 8th World Meeting on Bayesian Statistics*. Oxford University Press.

Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association* **104,** 735–746.

Petretto, E., Bottolo, L., Langley, S. R., Heinig, M., McDermott-Roe, C., Sarwar, R., Pravenec, M., Hbner, N., Aitman, T. J., Cook, S. A., and Richardson, S. (2010). New insights into the genetic control of gene expression using a Bayesian multi-tissue approach. *PLoS Computational Biology* **6,** e1000737.

Richardson, S., Bottolo, L., and Rosenthal, J. (2010). Bayesian models for sparse regression analysis of high dimensional data. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., editors, *Bayesian Statistics 9*, pages 539–569. Oxford University Press.

Rothman, A. J., Levina, E., and Zhu, J. (2010). Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics* **19,** 947–962.

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., Ruff, T. G., Milligan, S. B., Lamb, J. R., Cavet, G., Linsley, P. S., Mao, M., Stoughton, R. B., and Friend, S. H. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422,** 297–302.

Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology* **4,** Article 32.

Segal, E., Friedman, N., Kaminski, N., Regev, A., and Koller, D. (2005). From signatures to models: understanding cancer using microarrays. *Nature Genetics* **37 Suppl,** S38–S45.

Sillanpää, M. J. and Noykova, N. (2008). Hierarchical modeling of clinical and expression quantitative trait loci. *Heredity* **101,** 271–284.

Smith, M. and Kohn, R. (2000). Nonparametric seemingly unrelated regression. *Journal of Econometrics* **98,** 257–281.

Stranger, B., Nica, A., Forrest, M., Dimas, A., Bird, C., Beazley, C., Ingle, C., Dunning, M., Flicek, P., Montgomery, S., Tavare, S., Deloukas, P., and Dermitzakis, E. (2007). Population genomics of human gene expression. *Nature Genetics* **39,** 1217–1224.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **58,** 267–288.

Wang, H. (2010). Sparse seemingly unrelated regression modelling: Applications in finance and econometrics. *Computational Statistics and Data Analysis* **54,** 2866–2877.

Yi, N. and Shriner, D. (2008). Advances in Bayesian multiple quantitative trait loci mapping in experimental crosses. *Heredity* **100,** 240–252.

Yin, J. and Li, H. (2011). A sparse conditional Gaussian graphical model for analysis of genetical genomics data. *Annals of Applied Statistics* **5,** 2630–2650.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* **57,** 348–368.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In Goel, P. K. and Zellner, A., editors, *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pages 233–243. Elsevier, North-Holland.

Zhu, J., Lum, P. Y., Lamb, J., GuhaThakurta, D., Edwards, S. W., Thieringer, R., Berger, J. P., Wu, M. S., Thompson, J., Sachs, A. B., and Schadt, E. E. (2004). An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenetic and Genome Research* **105,** 363–374.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67,** 301–320.

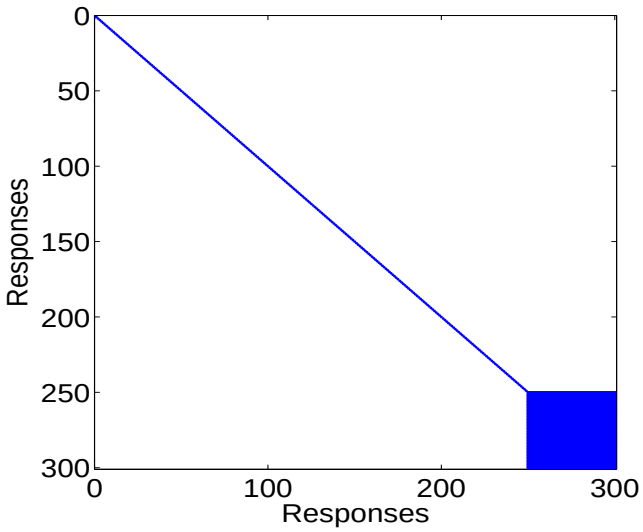
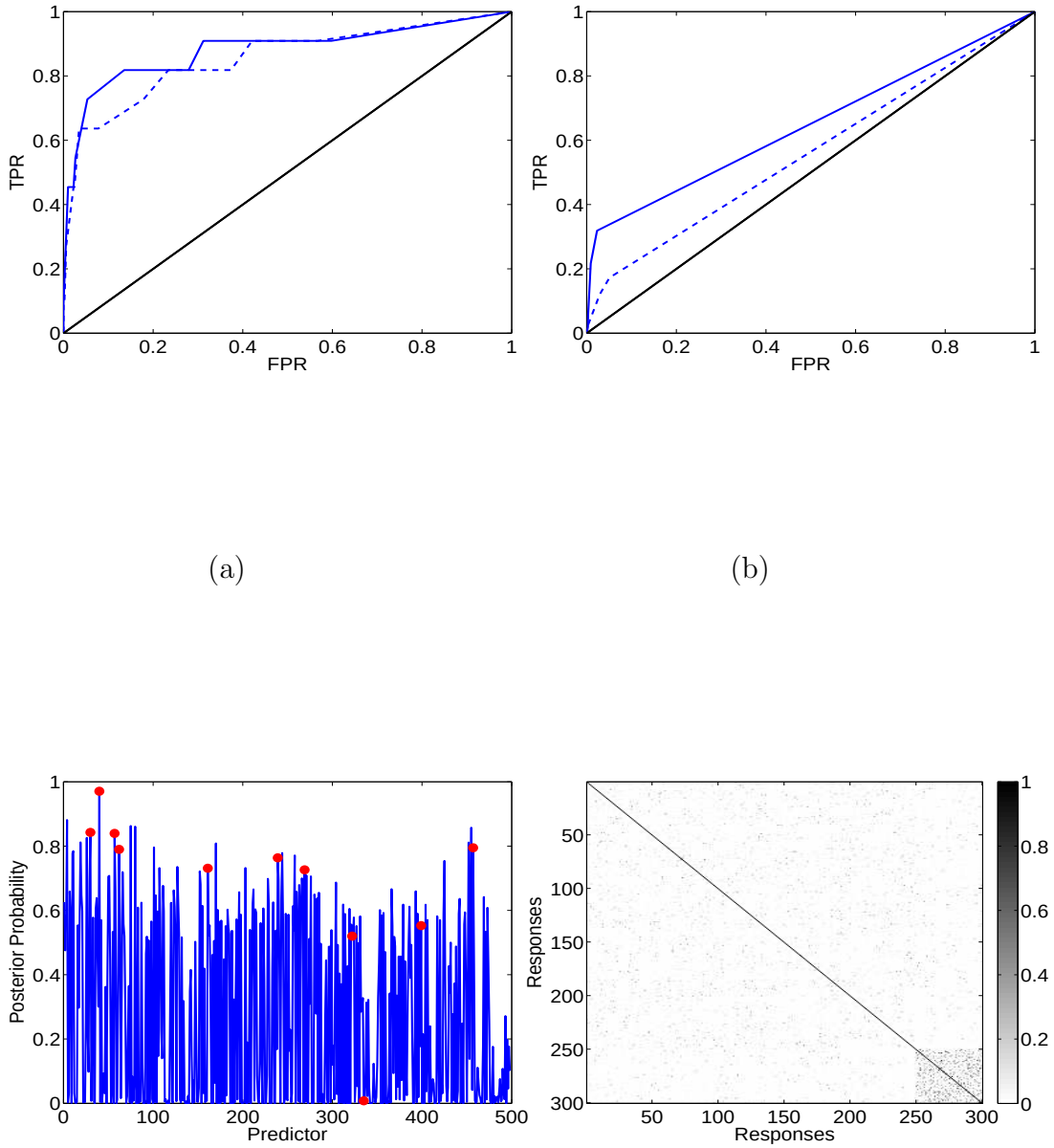


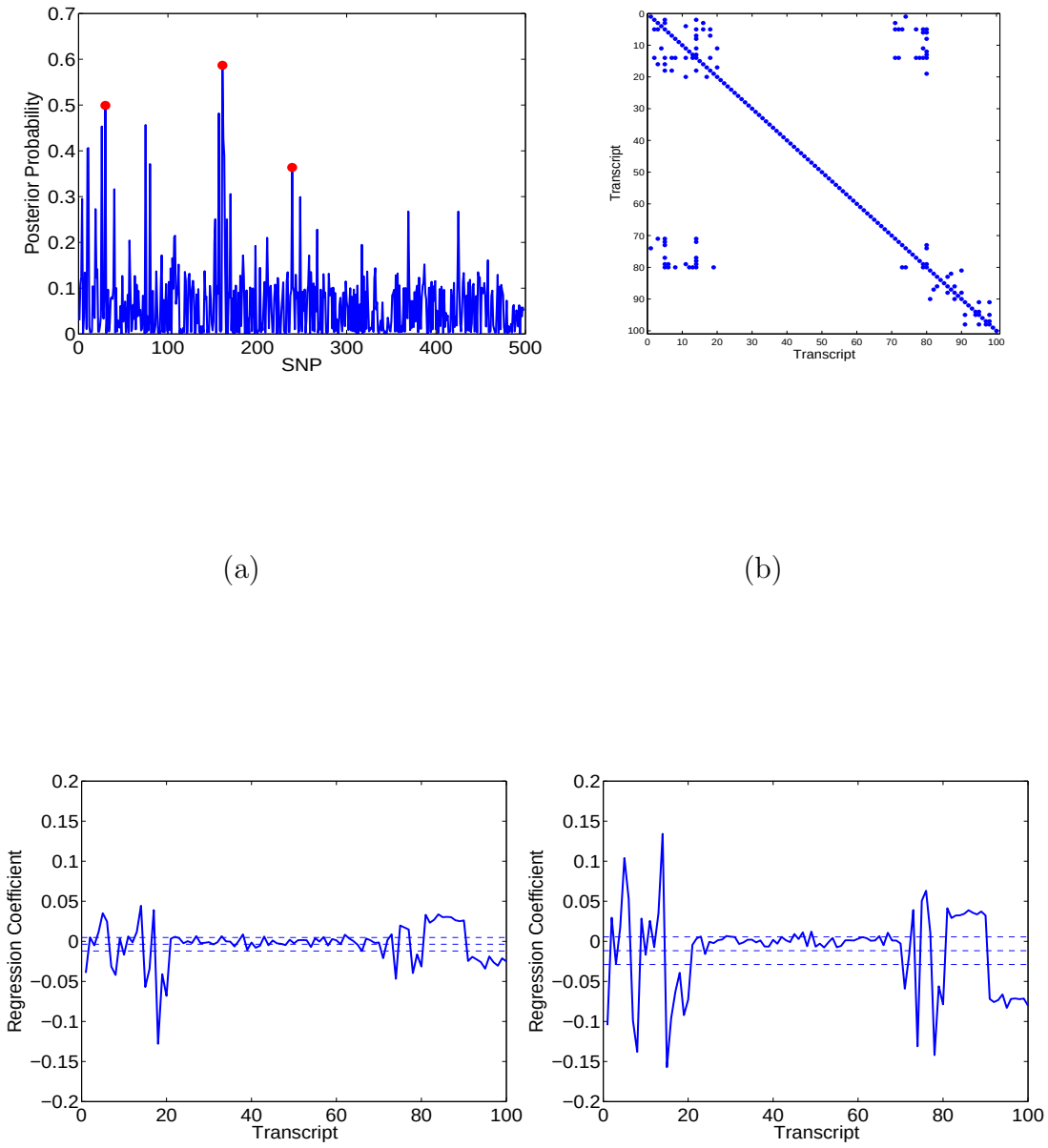
Figure 1. Setting for simulation example 1: The true adjacency matrix of the inverse covariance graph for the responses. Shaded area indicates the presence of an edge and otherwise an edge is absent. This figure appears in color in the electronic version of this article.



(a) (b)

(c) (d)

Figure 2. Result for simulation example 1: Plot of TPR v. FPR as the threshold on the posterior probabilities of inclusion in the model is varied (i.e., ROC curves) for (a) $\boldsymbol{\gamma}$, joint estimation (solid curve) and fixed, diagonal graph (broken curve) and (b) $\mathbf{G}$, joint estimation (solid curve) and zero mean model (broken curve). The diagonals correspond to the ROC curves for random guess. (c) The marginal posterior probabilities for $\boldsymbol{\gamma}$ with the true variables circled. (d) The marginal posterior probabilities for $\mathbf{G}$ on a gray scale. This figure appears in color in the electronic version of this article.



(a) (b)

(c) (d)

Figure 3. Result for simulation example 2: (a) the marginal posterior probabilities for $\boldsymbol{\gamma}$ with the true variables circled (b) the recovered adjacency matrix with a cutoff on the posterior probabilities of edge inclusion set to 0.4 (c) association of SNP 161 with all the 100 transcripts, showing enhanced association for transcripts 17-20 and (d) association of SNP 239 with all the 100 transcripts, showing enhanced association for transcripts 1-20 and 71-80. The dashed lines in the last two plots are the mean and Bonferroni corrected 95% intervals. This figure appears in color in the electronic version of this article.

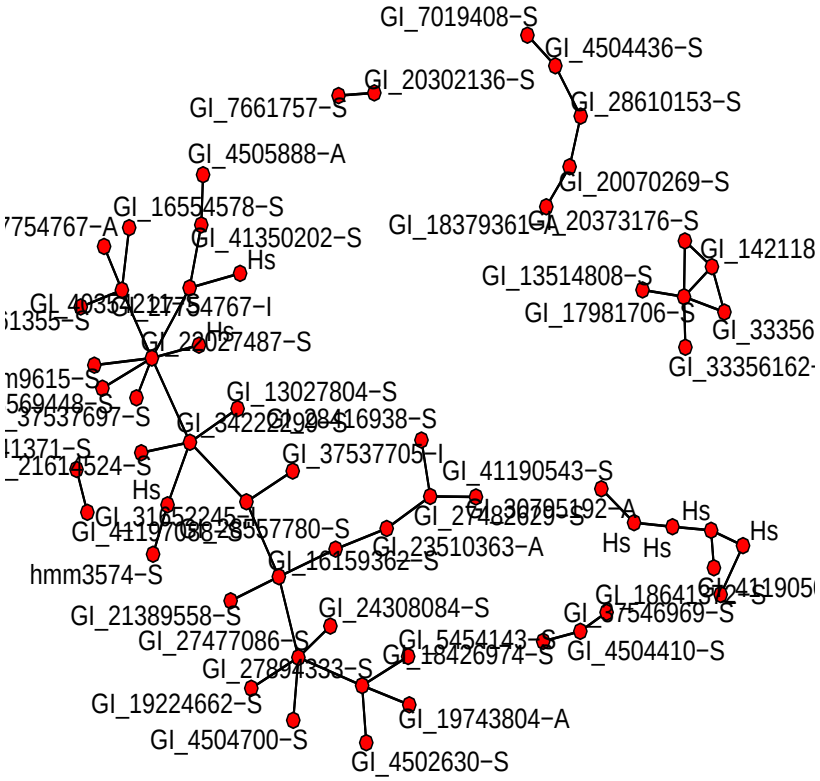


Figure 4. The inferred graph for the CEU data, showing 55 significant interactions among the 100 traits considered. Traits that are conditionally independent of all the other traits are not shown. This figure appears in color in the electronic version of this article.

Table 1
Simulation setting for the association between significant SNPs and corresponding transcripts in simulation example 2.

SNP ($\tilde{p}$)	Transcript ($\tilde{q}_p$)
30	1-20, 71-80
161	17-20
239	1-20, 71-80

Table 2

Association between significant SNPs and corresponding transcripts in the eQTL study identified by a t-test as significant at 5% level after Bonferroni correction.

SNP	Transcript (Illumina Probe ID)
rs2285635	GI_17981706-S, GI_33356162-S, Hs.449602-S, hmm10289-S, GI_20373176-S
rs12026825	GI_41197088-S, GI_41190507-S, GI_18641371-S, Hs.449602-S, GI_23065546-A, GI_24797066-S, GI_18641372-S
rs757124	hmm3574-S, Hs.449602-S, hmm3577-S, GI_21389378-S, GI_24797066-S
rs2205912	GI_18426974-S, GI_37546026-S, Hs.406489-S, hmm3574-S, Hs.449602-S
rs929762	GI_41190507-S, Hs.449605-S, Hs.512137-S, Hs.406489-S, hmm3574-S
rs3810711	GI_18426974-S, GI_41197088-S, Hs.406489-S, hmm3574-S, Hs.449572-S
rs708463	GI_37546026-S,GI_37546969-S,Hs.449609-S, GI_42662536-S, GI_18641372-S
rs7770751	GI_41190507-S, GI_37546026-S, Hs.512137-S, hmm3577-S, hmm10289-S, GI_28416938-S