

Inferring metabolic networks using the Bayesian adaptive graphical lasso with informative priors

CHRISTINE PETERSON*, MARINA VANNUCCI, CEMAL KARAKAS,
WILLIAM CHOI, LIHUA MA AND MIRJANA MALETIĆ-SAVATIĆ

Metabolic processes are essential for cellular function and survival. We are interested in inferring a metabolic network in activated microglia, a major neuroimmune cell in the brain responsible for the neuroinflammation associated with neurological diseases, based on a set of quantified metabolites. To achieve this, we apply the Bayesian adaptive graphical lasso with informative priors that incorporate known relationships between covariates. To encourage sparsity, the Bayesian graphical lasso places double exponential priors on the off-diagonal entries of the precision matrix. The Bayesian adaptive graphical lasso allows each double exponential prior to have a unique shrinkage parameter. These shrinkage parameters share a common gamma hyperprior. We extend this model to create an informative prior structure by formulating tailored hyperpriors on the shrinkage parameters. By choosing parameter values for each hyperprior that shift probability mass toward zero for nodes that are close together in a reference network, we encourage edges between covariates with known relationships. This approach can improve the reliability of network inference when the sample size is small relative to the number of parameters to be estimated. When applied to the data on activated microglia, the inferred network includes both known relationships and associations of potential interest for further investigation.

KEYWORDS AND PHRASES: Graphical models, Bayesian adaptive graphical lasso, Informative prior, Metabolic network, Neuroinflammation.

1. INTRODUCTION

In the graphical modeling framework, a graph structure $G = (V, E)$ is used to represent the conditional dependence relationships among a set of variables [28]. Each vertex $i \in V = \{1, \dots, p\}$ corresponds to a random variable, and there is no edge in $E = V \times V$ between variables i and j if and only if these two variables are conditionally independent given the remaining variables. Each edge $(i, j) \in E$ therefore represents a conditional dependence relationship. Since these relationships are assumed to be symmetric, $(i, j) \in E$ if and only if $(j, i) \in E$.

*Corresponding author.

When the data $\mathbf{Y} = (y_1, \dots, y_n)'$ are multivariate normal, Gaussian graphical models enforce the conditional independence relationships expressed in a graph G through constraints on the precision matrix $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$. Specifically, if $(i, j) \notin E$, then the corresponding entry ω_{ij} in the precision matrix is constrained to be zero. The nonzero entries ω_{ij} can be used to estimate partial correlations $\rho_{ij} = -\omega_{ij} / \sqrt{\omega_{ii}\omega_{jj}}$ that reflect the strength of the relationship between variables i and j after conditioning on all remaining variables. Inference on Gaussian graphical models requires both learning the network structure G and estimating the precision matrix $\mathbf{\Omega}$. Since the zeros in $\mathbf{\Omega}$ correspond to the graph G , the goal of inference can be framed as estimation of a sparse version of $\mathbf{\Omega}$.

Bayesian graphical lasso estimation techniques encourage sparsity by placing double exponential priors on the off-diagonal entries of the precision matrix. The Bayesian adaptive graphical lasso allows each double exponential prior to have a unique shrinkage parameter. These shrinkage parameters share a common gamma hyperprior. In Section 2 we first review graphical lasso methods and then extend the Bayesian adaptive models to create informative prior structures by formulating tailored hyperpriors on the shrinkage parameters. We include a brief simulation study comparing the performance of adaptive graphical lasso methods for inference of network structures. In Section 3 we illustrate our method with an application to inference of the cellular metabolic network in activated microglia, briefly described below. There we choose parameter values for each hyperprior that shift probability mass toward zero for nodes that are close together in a reference network, therefore encouraging edges between covariates (metabolites) with known relationships. Our approach improves the reliability of network inference when the sample size is small relative to the number of parameters to be estimated. The network we infer includes both known relationships and associations of potential interest for further investigation. Section 4 concludes the paper with a discussion.

1.1 Cellular metabolic network of activated microglia

We infer the cellular metabolic network from the metabolite concentration measurements from activated microglia. Microglia are innate immune cells which become activated

as a response to infection or injury of the brain [25]. The chronic activation of microglia characterizes neuroinflammation and is a hallmark of many neurodegenerative diseases, including Alzheimer’s disease and Huntington’s disease [11, 21]. Despite the significance of neuroinflammation, the mechanisms by which it leads to disruptions in cellular function are yet to be revealed. Understanding how neuroinflammation perturbs the cellular metabolic network could improve our knowledge of the mechanisms behind neurodegenerative disease, help identify biomarkers, and provide targets for potential therapeutic interventions.

The mainstay of our approach is that the changes that occur due to microglial activation are associated not only with individual metabolite perturbations, but also with changes to metabolic pathways in which metabolite relationships are altered. These relationships make up the connections in the metabolic network. To infer the structure of this network, we consider a Gaussian graphical model where each metabolite corresponds to a node in the graph. In such a network, if the concentrations of two metabolites are dependent given the concentrations of all other metabolites, then the two corresponding nodes will be connected by an edge. We interpret an edge in the network as signifying that the two connected metabolites are related through cellular reactions, either directly or indirectly through reactions involving intermediate molecules.

The literature on statistical approaches to infer metabolic networks is quite sparse. Early attempts focus on linear associations between metabolites as captured by Pearson correlation coefficients [6, 44] or on dependency measures based on the Kullback-Leibler information divergence [4]. More recently, Krumsiek et al. [27] construct networks as indirect graphs where conditional dependencies between variables are captured by partial correlation coefficients. In our approach, we propose adaptive Bayesian graphical models with prior parameters that encourage edges between metabolites with known relationships. By interrogating public databases, we can map the metabolites of interest to biological pathways. Of these databases, the Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.genome.ad.jp/kegg/) is the most complete. The KEGG database provides information not only on metabolic interactions but also on the compounds, enzymes and genes involved. We use the established metabolite relationships in the construction of our prior.

2. METHODS

We start by briefly reviewing lasso and graphical lasso regularization methods and their Bayesian counterparts. We then focus on adaptive methods and on our proposed informative priors for adaptive Bayesian inference.

2.1 Graphical lasso

In recent years, many statistical approaches have been developed that rely on regularization to reduce model complexity and prevent overfitting. In regression models, the

lasso method, one of the most popular regularization techniques, adds an L_1 penalty on the absolute value of the regression coefficients to the least squares criterion [45]. This penalty not only achieves shrinkage in the nonzero coefficient estimates but also performs variable selection since some coefficients are forced to be exactly zero. The elastic net, another regularized regression approach, includes both an L_1 and an L_2 penalty on the regression coefficients [53]. Whereas the lasso will select only one from a group of highly correlated predictors, the elastic net encourages a grouping effect in which strongly correlated predictors tend to come in or out of the model together.

In graphical models, regularized methods enable inference of sparse graphs. For example, the graphical lasso achieves sparsity in the estimation of the precision matrix $\mathbf{\Omega}$ by imposing a penalty on its L_1 norm [17, 34, 51]. Let $\mathbf{S} = \mathbf{Y}'\mathbf{Y}/n$ represent the sample covariance based on the column-centered data $\mathbf{Y}_{n \times p}$. The estimation procedure entails maximization of the penalized multivariate normal log-likelihood

$$(1) \quad \log(\det \mathbf{\Omega}) - \text{tr}(\mathbf{S}\mathbf{\Omega}) - \rho \|\mathbf{\Omega}\|_1,$$

given the constraint that $\mathbf{\Omega}$ must be symmetric and positive definite. To solve this optimization problem, an efficient algorithm using coordinate descent has been proposed. Although this approach is computationally fast, the use of a single fixed penalty parameter causes large values to be over-penalized. To avoid this bias, Fan et al. [16] have proposed the adaptive graphical lasso that uses a penalty term in which the absolute values of the entries in the precision matrix ω_{ij} are adaptively weighted. The criterion to be maximized is then

$$(2) \quad \log(\det \mathbf{\Omega}) - \text{tr}(\mathbf{S}\mathbf{\Omega}) - \lambda \sum_{i=1}^p \sum_{j=1}^p w_{ij} |\omega_{ij}|,$$

where the adaptive weights w_{ij} are defined as $w_{ij} = 1/|\tilde{\omega}_{ij}|^\gamma$ for some $\gamma > 0$ and any consistent initial estimate of the precision matrix $\tilde{\mathbf{\Omega}} = (\tilde{\omega}_{ij})$. Such penalty structure results in better prediction and improves specificity with respect to the standard lasso.

2.2 Bayesian graphical lasso

Bayesian regularization methods achieve shrinkage through the choice of a prior that favors values close to zero. In the original proposal of the lasso, Tibshirani [45] notes that the lasso coefficient estimates match the maximum *a posteriori* (MAP) estimates in the Bayesian framework when independent double exponential priors are placed on the regression coefficients. Park and Casella [39] explore this setting, demonstrating that as the shrinkage parameter λ is increased, the Bayesian lasso coefficient estimates tend to zero more slowly than under the original version of the lasso, but that for appropriately chosen penalty parameters the posterior median estimates are very close to those from the original lasso. The Bayesian version of the elastic net uses a combination of double exponential and normal

priors to capture the L_1 and L_2 penalty terms of the original elastic net [5, 30]. The Bayesian adaptive lasso avoids over-penalization of large effects through a prior formulation which allows the scale parameter to vary across coefficients [19]. One advantage of the Bayesian lasso methods over the original versions lies in the selection of the penalty parameters λ . In the classical framework, these are fixed and chosen via cross-validation techniques, which have been shown to be unstable. In the Bayesian lasso models, uncertainty in the selection of the penalty parameters is accounted for by imposing gamma hyperpriors and including these parameters in the Markov Chain Monte Carlo (MCMC) sampling of the posterior. See [39] and [30] for Monte Carlo EM algorithms.

In graphical models, the Bayesian graphical lasso shrinks the off-diagonal entries of $\mathbf{\Omega}$ toward zero using a double exponential prior [47]. The sharpness of the double exponential distribution is controlled by the shrinkage parameter λ . The resulting parameter estimates can be linked to the frequentist results in a similar way as for the Bayesian lasso: given the choice of $\lambda = \rho/n$, the posterior mode of $\mathbf{\Omega}$ under the Bayesian graphical lasso is the frequentist graphical lasso estimate. And again, the shrinkage parameter λ does not need to be fixed: instead, uncertainty over λ can be expressed through a hyperprior and it can be included in posterior sampling. Wang [47] demonstrates that the Bayesian graphical lasso has reduced standard errors versus the original graphical lasso, due in part to the fact that the final estimates of $\mathbf{\Omega}$ are averaged over the sampled values of λ .

2.3 Informative priors for adaptive estimation

The Bayesian adaptive graphical lasso, proposed as an analogue to the adaptive graphical lasso, allows different shrinkage parameters λ_{ij} for different entries in $\mathbf{\Omega}$ [47]. These shrinkage parameters share a common $\text{Gamma}(r, s)$ hyperprior where r and s are fixed hyperparameters. Given these parameters, the posterior conditional mean of λ_{ij} given ω_{ij} will be small for large $|\omega_{ij}|$, and vice-versa. This means that the λ_{ij} can be inferred in a way that retains the advantage of the adaptive lasso in reducing the bias incurred by a single penalty.

Although the prior formulation of the adaptive Bayesian graphical lasso does express the belief that the overall network structure is sparse, it does not use specific prior knowledge on likely interactions. We extend the adaptive Bayesian graphical lasso to allow an informative prior by specifying unique values of s_{ij} for each off-diagonal λ_{ij} based on prior reference information. We thereby take advantage of one of the major strengths of the Bayesian approach, which is the ability to incorporate valuable information from previous research through the choice of an appropriate prior.

Our model formulation assumes that the data $\mathbf{Y}_{n \times p}$ follow a multivariate normal likelihood

$$(3) \quad p(\mathbf{y}_i | \mathbf{\Omega}) = \mathcal{N}(0, \mathbf{\Omega}^{-1}), \quad i = 1, \dots, n,$$

where the precision matrix $\mathbf{\Omega}$ is the inverse of the covariance matrix $\mathbf{\Sigma}$. Following the formulation in [47], the prior on the precision matrix $\mathbf{\Omega}$ is

$$(4) \quad p(\mathbf{\Omega} | \{\lambda_{ij}\}_{i < j}) = \frac{1}{C} \prod_{i < j} \left[\underbrace{\frac{\lambda_{ij}}{2} \exp\{-\lambda_{ij}|\omega_{ij}|\}}_{\text{Double exponential prior on } \omega_{ij}} \right] \prod_{i=1}^p \left[\underbrace{\frac{\lambda_{ii}}{2} \exp\{-\frac{\lambda_{ii}}{2}\omega_{ii}\}}_{\text{Exponential prior on } \omega_{ii} > 0} \right] 1_{\mathbf{\Omega} \in M^+},$$

where C is the normalizing constant and M^+ is the cone of positive-definite matrices of dimension p . A hyperprior on the shrinkage parameters for the off-diagonal entries of $\mathbf{\Omega}$ can be formulated as

$$(5) \quad p(\{\lambda_{ij}\}_{i < j} | \{\lambda_{ii}\}_{i=1}^p) \propto C \prod_{i < j} \underbrace{\frac{s_{ij}^r}{\Gamma(r)} \lambda_{ij}^{r-1} \exp\{-\lambda_{ij}s_{ij}\}}_{\text{Gamma}(r, s_{ij}) \text{ prior on } \lambda_{ij}},$$

where $\Gamma(\cdot)$ represents the gamma function. Wang [47] suggests a noninformative prior formulation where s is chosen small relative to ω_{ij} , in order for the inference of λ_{ij} to be truly adaptive. He specifically chooses the parameter setting $s = 10^{-6}$ and $r = 10^{-2}$.

When relevant prior network information is available, this can be integrated into the model specification through the choice of the hyperparameters s_{ij} . Since smaller values of λ_{ij} imply that an edge between i and j is more likely, we would like to shift the prior density of λ_{ij} toward zero when i and j are more closely linked according to a reference network G^* that summarizes known relationships between the covariates. Let d_{ij} be the length of the shortest undirected path between nodes i and j in G^* . Several approaches exist to calculate such distances within a graph structure. Here we use the implementation found in the R package `igraph` that relies on breadth-first search of the graph [13]. If i and j are not mutually reachable, we assume that d_{ij} is infinite. We then set

$$(6) \quad s_{ij} = \begin{cases} d_{ij}^{-1} \cdot 10^{-6+c} & \text{for finite } d_{ij} \\ 10^{-6} & \text{if } d_{ij} = \infty, \end{cases}$$

where $0 < c < 6$ is a positive constant, with $c = 2$ being a reasonable choice. Our prior setting encourages smaller shrinkage parameters λ_{ij} for those entries in $\mathbf{\Omega}$ that reflect closer connections in the prior network. If instead no prior connection exists, then we set s_{ij} to the noninformative setting which allows the posterior distribution of λ_{ij} to be primarily determined by ω_{ij} . Our choice of hyperparameters makes edges between nodes that are closer in the prior network more likely, while still allowing edges between nodes that are further apart or not mutually reachable in the prior network. The inverse distance is a sensible measure to use in the prior since we assume *a priori* that we are less likely

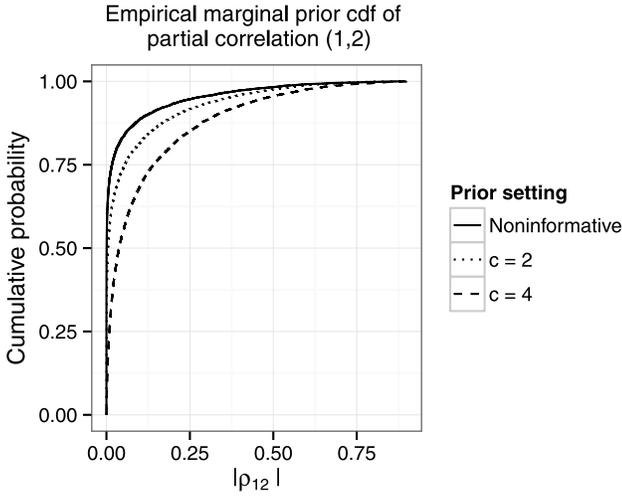


Figure 1. Empirical marginal prior cdf for the magnitude of the partial correlation ρ_{12} for noninformative prior vs. two settings of informative prior favoring the edge (1, 2).

to be able to observe connections which require many intermediate steps.

To better understand the implications of the choice of s_{ij} in the informative prior, we can look at the implied prior distribution on the partial correlation ρ_{ij} . The marginal prior on ρ_{ij} is difficult to assess analytically due to the positive definite constraint on Ω . However, we can obtain empirical estimates of the marginal prior density through simulation. Specifically, we ran an MCMC sampler following the approach given in section 2.4 of [47] to generate a sample of the precision matrix Ω using 5,000 iterations burn-in and 10,000 iterations after the burn-in under each of three prior settings. The first setting is the noninformative prior with $s_{ij} = 10^{-6}$ for all (i, j) . In the two informative prior settings, we favor edge (i, j) by assuming a prior reference network with $d_{12} = 1$ and $d_{ij} = \infty$ for all other (i, j) . We compare the effect of this assumption under the parameter settings $c = 2$ and $c = 4$. The empirical marginal cumulative distribution function (cdf) for the absolute value of ρ_{12} under these three conditions is given in Figure 1. As expected, larger values of s_{12} shift the marginal prior density away from zero.

2.4 Posterior inference

To find the conditional posterior distribution of λ_{ij} for $i < j$ we consider the terms in the posterior distribution that include λ_{ij} :

$$\begin{aligned}
 (7) \quad p(\lambda_{ij} | \Omega) &\propto p(\Omega | \{\lambda_{ij}\}_{i \leq j}) \cdot p(\lambda_{ij}) \\
 &\propto \frac{\lambda_{ij}}{2} \exp\left\{-\lambda_{ij}|\omega_{ij}|\right\} \cdot \frac{s_{ij}^r}{\Gamma(r)} \lambda_{ij}^{r-1} \exp\{-\lambda_{ij}s_{ij}\} \\
 &\propto \lambda_{ij}^r \exp\left\{-\lambda_{ij}(|\omega_{ij}| + s_{ij})\right\}.
 \end{aligned}$$

We recognize this as the kernel of the Gamma($1 + r, |\omega_{ij}| + s_{ij}$) distribution. This means that posterior sampling can be performed using a straightforward modification of the data-augmented block Gibbs sampler proposed in [47]. Specifically, we extend the algorithm to allow a full matrix of values s_{ij} rather than the fixed scalar s . At each iteration, we sample a new value for each λ_{ij} from the Gamma($1 + r, |\omega_{ij}| + s_{ij}$) distribution.

Since the sampled matrices do not include exact zeros, the inference of the graph structure is not straightforward. Wang [47] suggests that entry ω_{ij} is nonzero if and only if the ratio of the posterior sample mean estimate of ρ_{ij} under the graphical lasso prior to the posterior expected value of ρ_{ij} using the standard conjugate Wishart prior $W(3, \mathbf{I}_p)$ is greater than 0.5. Another approach to edge selection in the Bayesian framework is to include edge (i, j) in the final model if the 95% posterior credible interval (CI) for ω_{ij} does not include 0. As discussed in Section 3.5, we found that for our case study the former approach yielded graphs that were not sufficiently sparse, while the latter was too stringent. Instead, we chose to determine the graph structure by selecting edges corresponding to partial correlations with absolute value greater than 0.1. Since the partial correlations reflect the strength of the conditional association between two nodes, this selection strategy is sensible from the practical perspective that the strongest connections are of greatest interest. Liu et al. [32] use a similar approach of selecting edges to summarize the dependence among variables in a regression by thresholding the correlations among regression coefficients.

As our final inference is based on the partial correlations, the sampling scheme could be modified to estimate the partial correlation matrix directly. The parametrization in terms of the precision matrix Ω is more straightforward, however, since Ω is the natural parameter for the multivariate normal likelihood. In addition, it is more computationally tractable to sample the precision matrix since the constraint that the partial correlations must have magnitude less than 1 renders a block Gibbs update of the off-diagonal elements of the partial correlation matrix infeasible.

2.5 Simulation study

To assess both the impact of the informative prior and the edge selection method, we include a simulation study comparing the performance of adaptive graphical lasso methods in learning the structure of an AR(2) model and of a scale-free network. We chose to include an AR(2) model since this type of network is commonly used as a test case for graph structure learning, including in [47]. For the AR(2) model, the precision matrix Ω is chosen to be the symmetric matrix

$$\Omega = \begin{pmatrix} 1 & 0.5 & 0.25 & 0 & \cdots & 0 \\ & 1 & 0.5 & 0.25 & \cdots & 0 \\ & & 1 & 0.5 & \cdots & 0 \\ & & & 1 & \cdots & 0 \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{pmatrix}.$$

Table 1. Simulation study comparing specificity (SP), sensitivity (SE), and Matthews correlation coefficient (MCC) for adaptive graphical lasso methods and edge selection procedures

	AR(2) graph with $p = 15$ and $n = 20$			Scale-free network with $p = 30$ and $n = 45$			
	SP	SE	MCC	SP	SE	MCC	
	Adaptive GLasso	0.74	0.59	0.31	0.79	0.75	0.39
Bayesian adaptive GLasso with noninformative/ informative prior	Ratio selection	0.90/ 0.87	0.17/ 0.23	0.10/ 0.13	0.88/ 0.86	0.37/ 0.40	0.22/ 0.22
	CI selection	1.00/ 1.00	0.03/ 0.04	0.10/ 0.15	1.00/ 1.00	0.07/ 0.08	0.23/ 0.24
	ρ threshold	0.95/ 0.94	0.28/ 0.36	0.33/ 0.39	0.95/ 0.95	0.39/ 0.42	0.38/ 0.39

For our second simulation, we chose a graph structure with properties more closely reflecting those of biological networks. Specifically, we used a scale-free network since Jeong et al. [23] found that metabolic network topologies are well-described by scale-free networks. Scale-free networks have the property that the degree distribution $P(k)$ follows a power-law distribution of the form $k^{-\gamma}$ where k is the node degree and γ is a positive constant [3]. Scale-free networks typically have a small percentage of hub nodes which have very high degree and many peripheral nodes that have few connections. We constructed a scale-free network on 30 nodes with power-law exponent $\gamma = 2$ using the function `barabasi.game()` from the `igraph` R package [13]. The resulting graph, which has 47 edges, includes a highly-connected central node. We created the precision matrix Ω for this graph by setting the entries $\omega_{ij} = \omega_{ji} = 0.2$ for all edges (i, j) .

In each iteration of the simulations, a sample \mathbf{Y} of size n is generated from the distribution $\mathcal{N}(\mathbf{0}, \Omega^{-1})$, and the sample covariance is computed as $\mathbf{S} = \mathbf{Y}'\mathbf{Y}/n$. To demonstrate the utility of our method in settings similar to that of the case study, we used comparable n/p ratios. We also experimented with other n/p settings and found that as expected, the added benefit from the informative prior declines with increasing n .

For each simulation setting, we apply the frequentist version of the adaptive graphical lasso [16] using 10-fold cross-validation for parameter selection. As in [16], the criteria for edge selection is that the estimated value of ω_{ij} is at least 0.001. We compare this approach to the Bayesian adaptive graphical lasso using both a noninformative and an informative prior. For the noninformative prior, we set $s = 10^{-6}$ and $r = 10^{-2}$. For the informative prior, we set $c = 2$ and use the true network to calculate the pairwise prior distances.

For the Bayesian methods, we compare three edge selection procedures: the approach from [47] which includes an edge if and only if the ratio of the posterior estimate of ρ_{ij} to the posterior expected value of ρ_{ij} using the Wishart prior $W(3, \mathbf{I}_p)$ is greater than 0.5, selection using 95% credible intervals, and selection using the criteria that the posterior estimate of ρ_{ij} has magnitude greater than a threshold of 0.1.

The selection performance is described in terms of specificity, sensitivity, and the Matthews correlation coefficient

(MCC), which represents an overall summary of classification success. Specificity is defined as

$$\text{Specificity} = \frac{TN}{TN + FP},$$

where TN represents the number of true negatives and FP is the number of false positives. Sensitivity is defined as

$$\text{Sensitivity} = \frac{TP}{TP + FN},$$

where TP is the number of true positives and FN is the number of false negatives. Finally, the MCC is defined as

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}.$$

The results from 50 iterations are given in Table 1. The frequentist version of the adaptive graphical lasso has better sensitivity than the Bayesian methods, but results in more false positive selections. For the Bayesian adaptive graphical lasso, the informative prior improves sensitivity for all three selection methods. Although there is a slight trade-off in terms of the specificity, use of the informative prior improves the overall performance as summarized by the MCC. Among the edge selection procedures, thresholding the posterior estimate of ρ results in more true positive selections than the alternative approaches. Selection using 95% credible intervals eliminates false positive edge selection, but at the expense of very low selection of true edges.

3. CASE STUDY

To infer the metabolic network of activated microglia under inflammatory conditions, we apply our method to a data set which consists of the estimated concentrations of 17 critical metabolites in 24 samples of cultured activated microglia. These metabolites, which include amino acids and other small molecules, represent both inputs and byproducts of the reactions taking place within the cell.

3.1 Metabolomics screening of activated microglia

Metabolomics refers to the study of global metabolite content in cells, tissues, and organisms. It provides the infor-

mation on “the metabolome” or the collection of all metabolites (small molecules, such as amino acids, fatty acids, lipids and carbohydrates) in a biological system. The level of each metabolite within the metabolome depends on the specific physiological, developmental, and pathological state of a cell or tissue. Therefore, the metabolome reflects the phenotype of a cell or tissue resulting from different genetic or environmental influences [10, 12, 36]. Over the past several years, the use of metabolomics for biomarker discoveries has been explored in cancer cell biology, drug development, toxicology and other medical disciplines, offering the prospect of being increasingly important in medicine and biotechnology (see for example [8, 22, 38] and [9]).

Here we utilize nuclear magnetic resonance (NMR) spectroscopy-based metabolomics to investigate the metabolic status of inflammation in the central nervous system. We model this state by activating the primary innate immune cells of the brain parenchyma, i.e. microglia, with a gram-negative bacterial endotoxin, lipopolysaccharide (LPS). Microglia represent 10–20% of the brain glial cell population and are the first responder to perturbation of the brain parenchyma [2]. In steady-state condition of healthy brains, microglia function as sentinels by continuously sampling the environment with their processes for any aberrations to homeostasis. During insult as seen in pathological conditions such as neurodegenerative diseases, stroke, microbial infections, and tumor metastasis, microglia become activated. At this time, microglia secrete cytokines, which act as molecular signals to alarm other cellular players in the niche of the impending danger, resulting in inflammation [20, 26]. Several neurological disorders, such as multiple sclerosis, Parkinson’s disease, Alzheimer’s disease, schizophrenia, and even autism, are associated with chronic activation of microglia, [29, 37, 41]. Most neurological disorders related to chronic inflammation, and by extension activation of microglia, are not detected until the occurrence of the symptoms or a brain lesion is noticeable. Therefore, biomarkers that can detect the onset of a disease early in the time course can lead to better disease management.

3.2 Data collection and processing

Our data consist of estimated concentrations of 17 critical metabolites in 24 samples of cultured activated microglia. The metabolites were identified and their concentrations were estimated based on proton nuclear magnetic resonance (^1H NMR) spectra. NMR spectroscopy allows the acquisition of highly reproducible and resolved spectra which consist of latent metabolites resonating at various chemical shifts measured in parts per million (ppm). Such spectra contain thousands of resonances that may belong to hundreds of metabolites [36]. The ^1H NMR data were collected on Bruker Avance 800 MHz NMR spectrometers equipped with a CryoProbe. The temperature of the data acquisition was 25°C. The chemical shift and the concentrations of the

metabolites are referenced to 0.05 millimolar trimethylsilyl propanoic acid (TSP). The spectra were processed and analyzed on a Linux workstation running TopSpin 2.1.

The statistical analysis of NMR metabolic profiles typically involves a number of different stages [14, 15]. Many programs are available to perform standard data pre-processing techniques and obtain phased, baseline-corrected, chemical shift referenced and normalized spectra. We use the commercial software Chenomx (Chenomx Inc., Edmonton, Canada) to perform metabolite identification and quantification [48]. In NMR spectroscopy, identification of metabolites is a crucial step of any data analysis. The chemical structure of a molecule uniquely determines the number of peaks it generates in an NMR spectrum, together with their location and ratio of heights. Given an observed spectrum, the area under the peaks from a specific metabolite is directly related to its abundance. This makes it possible to quantify the concentrations of selected metabolites by matching the sample spectra to the reference NMR spectra of pure compounds that can be downloaded from public databases. We used the Chenomx 800 MHz reference library to identify and quantify the abundant metabolites from the processed NMR spectra.

Some authors, including [27], transform the metabolite concentrations to the log scale to improve normality, but we found that our data were reasonably normal on the original scale and the log transform did not offer significant improvement. All analyses were therefore done on the untransformed concentration estimates. As desired, the effect of the prior is not strong, but it clarifies the selection decision for values close to the cutoff.

3.3 Prior formulation

Although we have sufficient sample size to ensure that the sample covariance matrix \mathbf{S} is nonsingular, we still need to estimate more parameters than we have data points since there are 136 possible edges in a graph on 17 nodes. This makes the use of regularized estimation and relevant prior information of particular importance.

We obtained reference information on metabolic pathways from the KEGG database, which provides reaction networks based on manually collated results from published research [24]. The KEGG network can be exported as KGML, an XML-based representation, and imported into R using the package `KEGGgraph` [52]. To obtain a prior network relating the metabolites in our data set, we merged the KEGG global metabolic pathways network with additional lipid metabolism pathways (glycerophospholipid metabolism and ether lipid metabolism) that include sn-glycero-3-Phosphocholine. We then calculated pairwise distances d_{ij} as previously described. For metabolites that were mutually reachable in the reference network, distances ranged from direct connections with distance 1 to more distant connections with distances up to 20. Pairwise distances greater than 16 are possible since the reference network is

Table 2. Length of shortest undirected path between metabolites in reference network

	Acetate	Alanine	Alloisoleucine	Choline	Creatine	Creatine phosphate	Glutamate	Glutamine	Glutathione	Glycine	Isoleucine	Lactate	Leucine	O-Phosphocholine	Pyruvate	Valine	sn-g-3-P
Acetate		3	Inf	6	6	7	4	5	5	4	15	3	7	6	2	8	5
Alanine			Inf	7	5	6	2	3	4	3	13	2	8	8	1	7	6
Alloisoleucine				Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf
Choline					10	11	8	9	9	8	19	7	12	1	6	11	1
Creatine						1	5	5	3	2	15	5	11	11	4	10	9
Creatine phosphate							6	6	4	3	16	6	12	12	5	11	10
Glutamate								1	2	3	13	4	9	9	3	7	7
Glutamine									3	4	14	5	10	10	4	8	8
Glutathione										1	15	4	10	10	3	9	8
Glycine											14	3	9	9	2	10	7
Isoleucine												14	19	20	13	18	18
Lactate													8	8	1	9	6
Leucine														12	7	12	11
O-Phosphocholine															7	12	2
Pyruvate																8	5
Valine																	10
sn-g-3-P ¹																	

¹sn-glycero-3-Phosphocholine

Trace plots of MCMC sampling

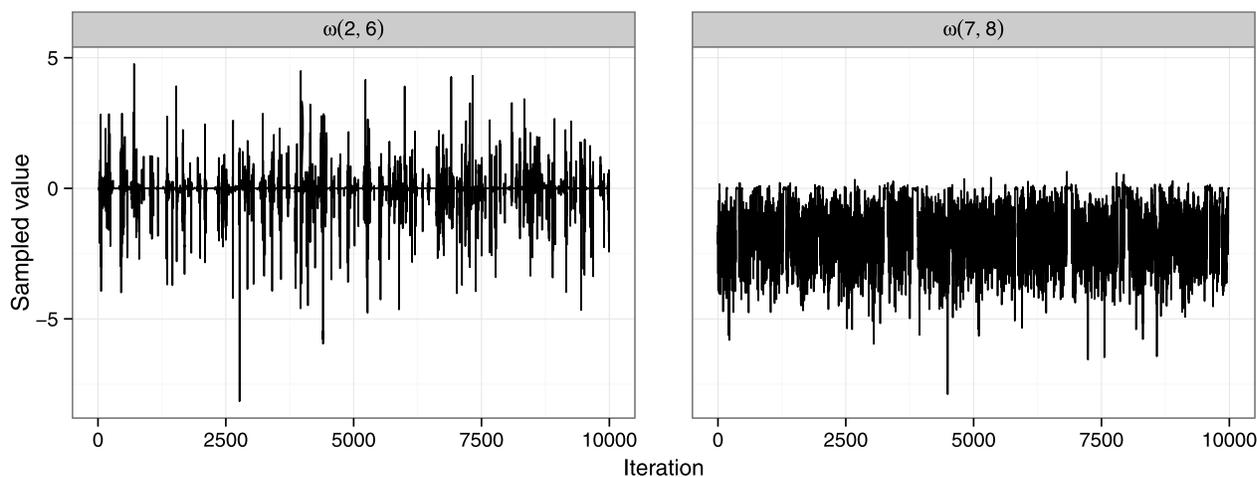


Figure 2. Example trace plots for elements of Ω . The plot at left, where the values are centered around 0, shows the MCMC sample of $\omega(2,6)$. This relationship, which is between alanine and creatine, was not included as an edge in the final graph. The plot at right corresponds to the nonzero entry $\omega(7,8)$ which was included in the final graph as the edge between glutamate and glutamine. As desired, these plots show good mixing and no overall trends.

defined on a larger set of nodes. Since there was no information in the KEGG database for alloisoleucine, its distance to other metabolites is considered to be infinite, resulting in a noninformative setting for s_{ij} . See Table 2 for a full summary.

3.4 Results

Posterior inference was based on 10,000 MCMC iterations after 5,000 iterations of burn-in using the mildly informative setting $c = 2$. This choice of prior has only a slight

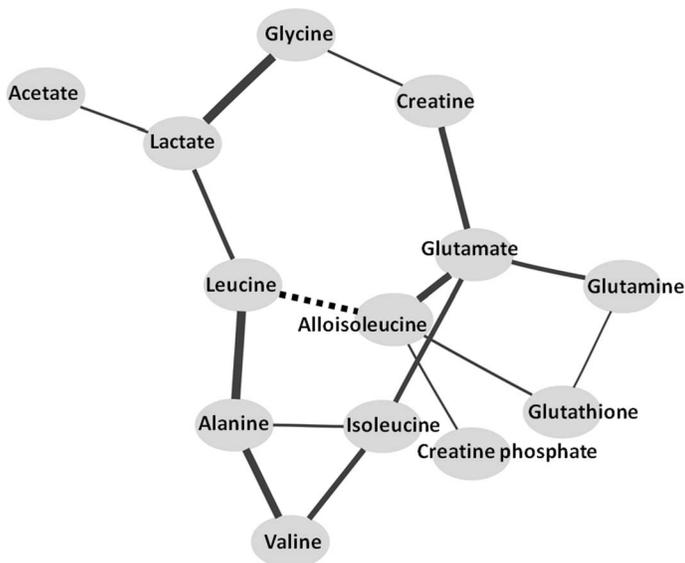


Figure 3. Inferred metabolic network with solid edges for negative partial correlations and dashed edges for positive partial correlations. Edges are included for partial correlations greater than 0.1 in magnitude, with thicker line widths for larger values.

effect on the network structure versus the noninformative prior, but does result in stronger estimated partial correlations for relationships favored by the prior reference network information (see also Section 3.5 below). Trace plots for the elements of Ω showed good convergence (see Figure 2), and results were stable across multiple runs of the sampler.

A network based on the estimated partial correlations is shown in Figure 3. Each node corresponds to a metabolite, and each edge represents a partial correlation between two metabolites with absolute value greater than 0.1 (see Table 3 for the full set of estimated partial correlations). The partial correlation summarizes the strength of the dependency between the concentrations of two metabolites given the concentrations of all other metabolites in the sample data. To depict the strength of these relationships in the network diagram, thicker edges are used for partial correlations of greater magnitude. Since this network is based on a limited set of metabolites, some edges may reflect indirect associations rather than direct reactions. An indirect association may occur when two metabolites are connected through reactions involving intermediate molecules not measured in the study. All estimated partial correlations in the network diagram are negative except for the partial correlation between the amino acids leucine and alloisoleucine. A negative partial correlation between two metabolites indicates that given the concentrations of all other metabolites, higher levels of one are associated with lower levels of the other.

3.5 Sensitivity

To assess the impact of our proposed prior, we ran the data analysis under a range of settings for the s_{ij} param-

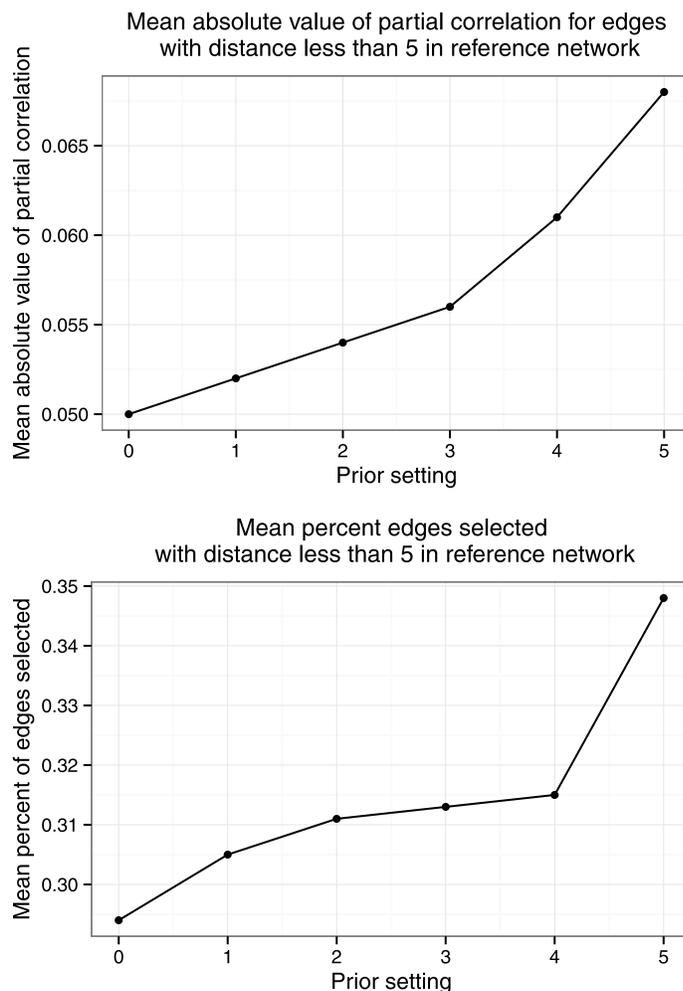


Figure 4. Mean of absolute values for estimated partial correlations for edges with distance less than 5 in reference network (top) and percentage of selected edges with distance less than 5 in reference network (bottom). Summary includes results under noninformative prior (labeled as prior setting 0) and under informative prior with c ranging from 1 to 5 (labeled as prior settings 1 to 5).

ters. The noninformative prior uses $s_{ij} = 10^{-6}$ for all pairs $i < j$. To test the impact of modifying s_{ij} , we studied additional settings where s_{ij} was set to $d_{ij}^{-1} \cdot 10^{-6+c}$, with $c = 1, 2, \dots, 5$. For cases resulting in s_{ij} smaller than 10^{-6} we simply retained the value from the noninformative prior, that is $s_{ij} = 10^{-6}$. For each setting, we performed 25 runs of the MCMC sampler, each with 5,000 iterations for burn-in and 10,000 iterations as the basis for inference.

As shown at the top of Figure 4, increasing c clearly reduces the shrinkage of the partial correlation estimates for edges favored under the prior across the range of values tested. The average estimated partial correlations for edges with distance less than five in the prior network increased steadily from 0.050 under the noninformative prior, to 0.054

Table 3. Estimated partial correlations

	Acetate	Alanine	Alloisoleucine	Choline	Creatine	Creatine phosphate	Glutamate	Glutamine	Glutathione	Glycine	Isoleucine	Lactate	Leucine	O-Phosphocholine	Pyruvate	Valine	sn-g-3-P
Acetate	-0.02	-0.01	-0.07	-0.03	-0.01	-0.07	-0.01	0.00	-0.05	-0.02	-0.15	0.01	-0.04	-0.01	-0.01	-0.01	-0.02
Alanine		-0.01	0.00	-0.01	0.00	-0.02	0.01	0.00	-0.03	-0.17	-0.02	-0.52	-0.01	-0.02	-0.44	-0.02	
Alloisoleucine			-0.02	-0.02	-0.14	-0.41	-0.01	-0.16	-0.04	-0.01	0.02	0.39	-0.02	0.00	-0.06	-0.03	
Choline				-0.03	-0.02	-0.05	-0.02	-0.02	-0.07	-0.01	-0.02	0.00	-0.03	-0.01	-0.01	-0.01	
Creatine					-0.03	-0.37	0.00	-0.02	-0.15	-0.07	0.00	0.00	-0.07	-0.01	-0.03	-0.04	
Creatine phosphate						-0.04	-0.01	-0.02	-0.01	-0.01	0.00	0.01	-0.01	0.00	-0.01	-0.01	
Glutamate							-0.26	-0.05	-0.04	-0.29	0.01	0.01	-0.05	0.00	-0.03	-0.07	
Glutamine								-0.12	-0.01	0.01	0.00	0.02	-0.01	0.00	0.01	0.00	
Glutathione									-0.04	-0.01	0.01	0.00	-0.02	-0.03	-0.01	-0.04	
Glycine										-0.03	-0.51	-0.02	-0.05	-0.04	-0.02	-0.03	
Isoleucine												0.00	-0.01	-0.03	-0.01	-0.33	-0.03
Lactate													-0.27	-0.01	-0.02	0.00	0.00
Leucine														0.00	-0.01	-0.01	-0.01
O-Phosphocholine															-0.01	-0.01	-0.02
Pyruvate																-0.02	-0.05
Valine																	-0.05
sn-g-3-P ²																	

²sn-glycero-3-Phosphocholine

with the mildly informative setting of $c = 2$, to 0.068 with the strongest prior setting of $c = 5$.

The three edge selection approaches compared in Section 2.5 resulted in very different levels of sparsity for the inferred graph. Under the noninformative prior, the ratio selection of Wang [47] resulted in an average of 58.5 selected edges, while the CI method resulted in an average of 5.5 selected edges. The partial correlation threshold offered a reasonable compromise with 15.6 edges selected on average. Although the number of selections varied widely, the trend is consistent with the results of the simulation study which showed that the CI method has a high specificity but very low sensitivity, while the ratio selection method of Wang [47] has lower specificity than the other two methods.

As shown at the bottom of Figure 4, the impact of the informative prior on edge selection using the partial correlation threshold criteria was moderate, with 29.4% of edges selected under the noninformative prior having distance less than five in the prior network, 31.1% of edges selected under the mildly informative prior with $c = 2$ having distance less than five in the prior network, and 34.8% of edges selected under the strongest prior with $c = 5$ having distance less than five in the reference network. The impact of the informative prior under the other two edge selection approaches was similar in that the percentage of edges selected reflecting close connections in the reference network increased for larger values of c .

Figure 5 shows the partial correlations under the non-informative and informative priors. The informative prior

causes one additional edge to be selected corresponding to the edge between glutamine and glutathione, which are separated by a path of length 3 in the reference network. The informative prior reduces shrinkage on the partial correlations which correspond to closer relationships in the reference network, but does not substantially change the structure of the selected network.

3.6 Findings

While many of the edges in the graph shown in Figure 3 correspond to known relationships, other connections potentially relevant to neuroinflammation are novel. The interpretation of the inferred connections requires an understanding of cellular metabolism in general, and of microglia specifically.

One of the key metabolites in the network is glutamate, which functions as an excitatory neurotransmitter in neurons. High levels of glutamate are associated with excitotoxicity and may account for some of the damage that occurs in persons suffering from neurodegenerative disease. Under normal conditions, levels of the glutamate transporters GLT-1 and GLAST are minimal. When microglia are activated by treatment with LPS, GLT-1 and GLAST are expressed, increasing the uptake of glutamate into the cell [25, 40]. The discovered connections involving glutamate are therefore of particular relevance. The negative partial correlation between glutamate and glutamine reflects the glutamate-glutamine cycle, an important metabolic pathway in the brain in which glutamate is converted to glu-

Magnitude of partial correlations under noninformative and informative priors

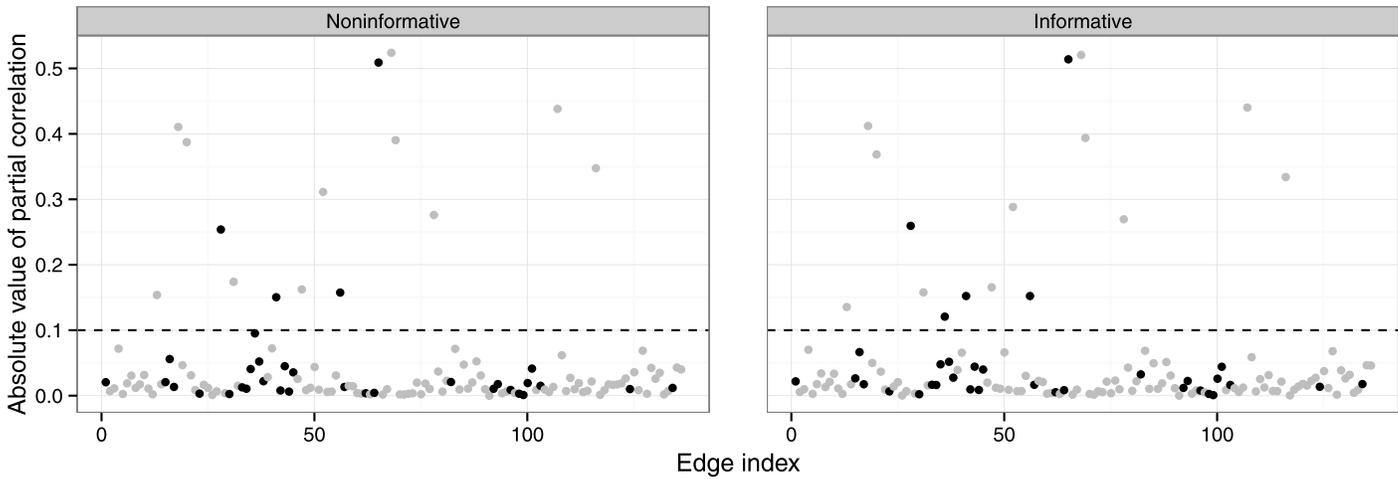


Figure 5. Magnitudes of partial correlation for each pairwise relationship under noninformative and informative priors. Darker points indicate a distance in the reference network less than 5. The selection cutoff of 0.1 is marked with a dashed line.

tamine, and glutamine in turn is converted back to glutamate [7]. This cycle is one mechanism for the homeostatic regulation of glutamate. The strong negative association of creatine with glutamate, which have an estimated partial correlation of -0.37 , is noteworthy because of its relevance to treatment of neuroinflammation. This relationship is consistent with research demonstrating that the supplementation of creatine lowers brain glutamate levels in patients with neurodegenerative disease [18].

The connection between glutamine and glutathione, which was added due to the informative prior, is both biologically plausible and of potential significance in the treatment of neuroinflammation. The inferred negative partial correlation between these two metabolites reflects glutamine’s known role as a precursor of glutathione. This connection is of importance since the LPS-treated microglia are in a state of increased oxidative stress, a condition also found under inflammation and neurological disorders. Glutathione is an antioxidant which can modulate the oxidative stress [1].

Other relevant connections include the negative partial correlation of glycine and creatine, which can be explained by the fact that glycine is an input into the production of creatine [49]. Glycine is in turn essential for the production of nitric oxide (NO) by microglia [50]. Increased production of NO by activated microglia is a possible cause of pathology in neurodegenerative disease [42].

An interesting indirect association uncovered by this analysis is the negative association of lactate with acetate. This relationship is biologically plausible since lactate is a byproduct of energy production when oxygen levels are low [33], whereas acetate is used as an input to Krebs cycle, which generates energy when oxygen levels are higher. Another set of associations of interest highlight the connection between the metabolic network of inflamed microglia

and NF- κ B activation. The transcription factor NF- κ B is responsible for the production of many inflammatory signals and has been shown to play a role in the activation of microglia [35]. Inhibition of NF- κ B has been shown to reduce glutamate levels and thereby reduce neurotoxicity [54]. Several metabolites in our inferred network are involved in NF- κ B regulation: glutamate has been shown to activate NF- κ B, whereas glutathione may inhibit its activation [31]. These relationships highlight the connection between the metabolic and gene regulatory networks, and offer a potential target for therapeutic intervention.

The connections involving metabolites related to the activation of the mTOR signaling pathway, which regulates cell growth, proliferation, and survival, are potentially noteworthy as well. Leucine has been shown to activate mTOR [46]. Russo et al. [43] found that mTOR activation was associated with the activation of microglia treated with LPS, and that the inhibition of mTOR reduced the response of microglia to pro-inflammatory cytokines. Thus, the inferred metabolic relationships related to mTOR signaling could be explored for potential avenues for treatment of neuroinflammation.

4. DISCUSSION

We have demonstrated that the Bayesian adaptive graphical lasso with informative priors is a useful tool in the inference of cellular metabolic networks. Our informative prior formulation encourages stronger shrinkage on connections that are not supported by prior information, but still allows novel relationships to be inferred. This method could also be used to infer other biological networks of interest such as protein-protein interaction networks or gene regulatory networks where relevant prior information is available from public databases, such as KEGG.

ACKNOWLEDGEMENTS

Christine Peterson's research is funded by a training fellowship from the Keck Center of the Gulf Coast Consortia, on the NLM Training Program in Biomedical Informatics, National Library of Medicine (NLM) T15LM007093. Marina Vannucci's research is partially funded by NIH/NHLBI P01-HL082798 and NSF/DMS 1007871. Mirjana Maletić-Savatić's research is funded by the NIH Intellectual and Developmental Disabilities Research Grant (P30HD024064), the Dana Foundation, and the McKnight Endowment Fund for Neuroscience.

Received 29 October 2012

REFERENCES

- [1] AMORES-SÁNCHEZ, M. and ÁNGEL-MEDINA, M. (1999). Glutamine, as a precursor of glutathione, and oxidative stress. *Mol. Genet. Metab.* **67** 100–105.
- [2] BANATI, R. (2003). Neuropathological imaging: *in vivo* detection of glial activation as a measure of disease and adaptive change in the brain. *Br. Med. Bull.* **65**(1) 121–131.
- [3] BARABÁSI, A. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. [MR2091634](#)
- [4] BANG, J., CROCKFORD, D., HOLMES, E., PAZOS, F., STERNBERG, M., MUGGLETON, S. and NICHOLSON, J. (2008). Integrative top-down system metabolic modeling in experimental disease states via data-driven Bayesian methods. *J. Proteome Res.* **7** 497–503.
- [5] BORNN, L., GOTTARDO, R. and DOUCET, A. (2010). *Grouping Priors and the Bayesian Elastic Net*. Technical report, Department of Statistics, University of British Columbia.
- [6] CAMACHO, D., DE LA FUENTA, A. and MENDES, P. (2005). The origin of correlations in metabolomics data. *Metabolomics* **1**(1) 53–63.
- [7] COOPER, A. (2012). The role of glutamine synthetase and glutamate dehydrogenase in cerebral ammonia homeostasis. *Neurochem. Res.* <http://dx.doi.org/10.1007/s11064-012-0803-4>.
- [8] COEN, M., HOLMES, E., LINDON, J. and NICHOLSON, J. (2008). NMR-based metabolic profiling and metabolomic approaches to problems in molecular toxicology. *Chem. Res. Toxicol.* **21**(1) 9–27.
- [9] DAVIS, V., BATHE, O., SCHILLER, D., SLUPSKY, C. and SAWYER, M. (2011). Metabolomics and surgical oncology: Potential role for small molecule biomarkers. *J. Surg. Onc.* **103**(5) 451–459.
- [10] DUNN, W., BAILEY, N. and JOHNSON, H. (2005). Measuring the metabolome: Current analytical technologies. *Analyst* **130**(5) 606–625.
- [11] DORING, A. and YONG, V. (2011). The good, the bad and the ugly. Macrophages/microglia with a focus on myelin repair. *Front. Biosci. (Schol. Ed.)* **3** 846–856.
- [12] FIEHN, O. (2002). Metabolomics – the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**(1) 155–171.
- [13] CSARDI, G. and NEPUSZ, T. (2006). The **igraph** software package for complex network research. *InterJournal, Complex Systems*, 1695.
- [14] EBBELS, T. and CAVILL, R. (2007). Bioinformatic methods in NMR-based metabolic profiling. *Prog. Nucl. Mag. Res. Sp.* **55**(4) 361–374.
- [15] EBBELS, T., LINDON, J. and COEN, M. (2011). Processing and modeling of nuclear magnetic resonance (NMR) metabolic profiles. *Methods Mol. Biol.* **708** 365–388.
- [16] FAN, J., FENG, Y. and WU, Y. (2009). Network exploration via the adaptive lasso and SCAD penalties. *Ann. Appl. Stat.* **3**(2) 521–541. [MR2750671](#)
- [17] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**(3) 432–441.
- [18] GENIUS, J., GEIGER, J., BENDER, A., MÖLLER, H., KLOPSTOCK, T. and RUJESCU, D. (2012). Creatine protects against excitotoxicity in an *in vitro* model of neurodegeneration. *PLoS One* **7**(2) e30554.
- [19] GRIFFIN, J. and BROWN, P. (2007). Bayesian adaptive Lasso with non-convex penalization. Technical report, University of Kent.
- [20] HANISCH, U. (2002). Microglia as a source and target of cytokines. *Glia* **40**(2) 140–155.
- [21] HOARAU, J., KREJBICH-TROTOT, P., JAFFAR-BANDJEE, M., DAS, T., THON-HON, G., KUMAR, S., NEAL, J. and GASQUE, P. (2011). Activation and control of CNS innate immune responses in health and diseases: A balancing act finely tuned by neuroimmune regulators (NIReg). *CNS Neurol. Disord. Drug Targets* **10**(1) 25–43.
- [22] HOLMES, E., WILSON, I. and NICHOLSON, J. (2008). Metabolic phenotyping in health and disease. *Cell* **134**(5) 714–717.
- [23] JEONG, H., TOMBOR, B., ALBERT, R., OLTVAI, Z. and BARABÁSI, A. (2000). The large-scale organization of metabolic networks. *Nature* **407**(6804) 651–654.
- [24] KANEHISA, M., GOTO, S., SATO, Y., FURUMICHI, M. and TANABE, M. (2012). KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Res.* **40** D109–D114.
- [25] KETTENMANN, H., HANISCH, U., NODA, M., and VERKHRATSKY, A. (2011). Physiology of microglia. *Physio. Rev.* **91** 461–553.
- [26] KREUTZBERG, G. (1996). Microglia: A sensor for pathological events in the CNS. *Trends Neurosci.* **19**(8) 312–318.
- [27] KRUMSIEK, J., SUHRE, K., ILLIG, T., ADAMSKI, J. and THEIS, F. (2011). Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst. Biol.* **5**(21) 1–16.
- [28] LAURITZEN, S. (1996). *Graphical Models*. Clarendon Press, Oxford. [MR1419991](#)
- [29] LEE, J., TRAN, T. and TANSEY, M. (2009) Neuroinflammation in Parkinson's disease. *J. Neuroimmune Pharmacol.* **4**(4) 419–429.
- [30] LI, Q., and LIN, N. (2010). The Bayesian elastic net. *Bayesian Anal.* **5**(1) 151–170. [MR2596439](#)
- [31] LI, X., SONG, L. and JOPE, R. (1998). Glutathione depletion exacerbates impairment by oxidative stress of phosphoinositide hydrolysis, AP-1, and NF- κ B activation by cholinergic stimulation. *Brain Res. Mol. Brain Res.* **53** 196–205.
- [32] LIU, F., CHAKRABORTY, S., FAN, L. and LIU, Y. (2012). Bayesian regularization via the graph Laplacian prior. 1–31. http://www.stat.duke.edu/~fl35/APT/top_ten_papers/gl_JASA11_submit.pdf.
- [33] LUNT, S. and VANDER HEIDEN, M. (2011). Aerobic glycolysis: Meeting the metabolic requirements of cell proliferation. *Annu. Rev. Cell Dev. Bi.* **27** 441–464.
- [34] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the Lasso. *Ann. Statist.* **34**(3) 1436–1462. [MR2278363](#)
- [35] MÉMET, S. (2006). NF- κ B functions in the nervous system: From development to disease. *Biochem. Pharmacol.* **72** 1180–1195.
- [36] METZ, T. (2011). *Metabolic Profiling: Methods and Protocols* Springer, New York.
- [37] MEYER, U., FELDON, J. and DAMMANN, O. (2011). Schizophrenia and autism: both shared and disorder-specific pathogenesis via perinatal inflammation? *Pediatr. Res.* **69**(5) 26R–33R.
- [38] NICHOLSON, J., CONNELLY, J., LINDON, E. and HOLMES, E. (2002). Metabolomics: A platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* **1** 153–161.
- [39] PARK, T., and CASELLA, G. (2008). The Bayesian lasso. *J. Am. Stat. Assoc.* **103**(482) 681–686. [MR2524001](#)
- [40] PERSSON, M., BRANTEFJORD, M., HANSSON, E. and RÖNNBÄCK, L. (2005). Lipopolysaccharide increases microglial GLT-1 expression and glutamate uptake capacity *in vitro* by a mechanism dependent on TNF- α . *Glia* **51** 111–120.

- [41] PETERSON, L. and FUJINAMI, R. (2007). Inflammation, demyelination, neurodegeneration and neuroprotection in the pathogenesis of multiple sclerosis. *J. Neuroimmunol.* **184**(1-2) 37–44.
- [42] POSSEL, H., NOACK, H., PUTZKE, J., WOLF, G. and SIES, H. (2000). Selective upregulation of inducible nitric oxide synthase (iNOS) by lipopolysaccharide (LPS) and cytokines in microglia: in vitro and in vivo studies. *Glia* **32** 51–59.
- [43] RUSSO, C., LISI, L., TRINGALI, G. and NAVARRA, P. (2009). Involvement of mTOR kinase in cytokine-dependent microglial activation and cell proliferation. *Biochem. Pharmacol.* **78** 1242–1251.
- [44] STEUER, R., KURTHS, J., FIEHN, O. and WECKWERTH, W. (2003). Observing and interpreting correlations in metabolic networks. *Bioinformatics* **19**(8) 1019–1026.
- [45] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the Lasso. *J. R. Statist. Soc. B* **58**(1) 267–288. [MR1379242](#)
- [46] TOKUNGA, C., YOSHINO, K. and YONEZAWA, K. (2004). mTOR integrates amino acid- and energy-sensing pathways. *Biochem. Biophys. Res. Co.* **313** 443–446.
- [47] WANG, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis* **7**(2) 771–790.
- [48] WELJIE, A., NEWTON, J., MERCIER, P., CARLSON, E. and SLUPSKY, C. (2006). Targeted profiling: Quantitative analysis of ¹H-NMR metabolomics data. *Anal. Chem.* **78** 4430–4442.
- [49] WYSS, M., and KADDURAH-DAOUK, R. (2000). Creatine and creatinine metabolism. *Physio. Rev.* **83**(3) 1107–1213.
- [50] YANG, L., TANAKA, J., ZHANG, B., SAKANAKA, M. and MAEDA, N. (1998). Astrocytes modulate nitric oxide production by microglial cells through secretion of serine and glycine. *Biochem. Biophys. Res. Co.* **251** 277–282.
- [51] YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94**(1) 19–35. [MR2367824](#)
- [52] ZHANG, J. and WIEMANN, S. (2009). KEGGgraph: A graph approach to KEGG PATHWAY in R and Bioconductor. *Bioinformatics* **25**(11) 1470–1471.
- [53] ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B* **67**(Part 2) 301–320. [MR2137327](#)
- [54] ZOU, J. and CREWS, F. (2005). TNF α potentiates glutamate neurotoxicity by inhibiting glutamate uptake in organotypic brain slice cultures: Neuroprotection by NF- κ B inhibition. *Brain Res.* **1034** 11–24.

Christine Peterson
 Department of Statistics, MS 138
 Rice University
 6100 Main St.
 Houston, TX 77251-1892
 USA
 E-mail address: cbp2@rice.edu

Marina Vannucci
 Department of Statistics, MS 138
 Rice University
 6100 Main St.
 Houston, TX 77251-1892
 USA
 E-mail address: marina@rice.edu

Cemal Karakas
 Departments of Pediatrics, Neuroscience, and Program in
 Developmental Biology
 Baylor College of Medicine
 Jan and Dan Duncan Neurological Research Institute at
 Texas Children's Hospital
 1250 Moursund St., Suite N1250.16
 Houston, TX 77030
 USA
 E-mail address: karakas@bcm.edu

William Choi
 Departments of Pediatrics, Neuroscience, and Program in
 Developmental Biology
 Baylor College of Medicine
 Jan and Dan Duncan Neurological Research Institute at
 Texas Children's Hospital
 1250 Moursund St., Suite N1250.16
 Houston, TX 77030
 USA
 E-mail address: wchoi@bcm.edu

Lihua Ma
 Departments of Pediatrics, Neuroscience, and Program in
 Developmental Biology
 Baylor College of Medicine
 Jan and Dan Duncan Neurological Research Institute at
 Texas Children's Hospital
 1250 Moursund St., Suite N1250.16
 Houston, TX 77030
 USA
 E-mail address: lihuam@bcm.edu

Mirjana Maletić-Savatić
 Departments of Pediatrics, Neuroscience, and Program in
 Developmental Biology
 Baylor College of Medicine
 Jan and Dan Duncan Neurological Research Institute at
 Texas Children's Hospital
 1250 Moursund St., Suite N1250.16
 Houston, TX 77030
 USA
 E-mail address: maletics@bcm.edu