Bayesian Hierarchical Latent Model for Gene Set Analysis

Yi Chao

Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science

in

Statistics

Inyoung Kim, Chair

Jeffrey B. Birch

Scotland Leman

April 29, 2009

Blacksburg, Virginia

Keywords: Bayesian hierarchical model, Generalized linear mixed model, Latent variable, Pathway based analysis, Point-mass density, Probit regression model

Bayesian Hierarchical Latent Model for Gene Set Analysis

Yi Chao

ABSTRACT

Pathway is a set of genes which are predefined and serve a particular celluar or physiological function. Ranking pathways relevant to a particular phenotype can help researchers focus on a few sets of genes in pathways. In this thesis, a Bayesian hierarchical latent model was proposed using generalized linear random effects model. The advantage of the approach was that it can easily incorporate prior knowledge when the sample size was small and the number of genes was large. For the covariance matrix of a set of random variables, two Gaussian random processes were considered to construct the dependencies among genes in a pathway. One was based on the polynomial kernel and the other was based on the Gaussian kernel. Then these two kernels were compared with constant covariance matrix of the random effect by using the ratio, which was based on the joint posterior distribution with respect to each model. For mixture models, log-likelihood values were computed at different values of the mixture proportion, compared among mixtures of selected kernels and point-mass density (or constant covariance matrix). The approach was applied to a data set (Mootha *et al.*, 2003) containing the expression profiles of type II diabetes where the motivation was to identify pathways that can discriminate between normal patients and patients with type II diabetes.

ACKNOWLEDGMENTS

I would like to thank my advisor, Inyoung Kim, for guiding me through the Bayesian area with application to bioinformatics over the years, and for inspiring me when I was depressed in the darkness prior to the brightness.

I would like to thank my other two committee members who have been very supportive. Jeffrey B. Birch has long been a director to me in both academics and routine life in U.S. His classes have proved to be one of my favorite learning experiences at VT. He has also supported me during the previous semesters so I could concentrate on research. Scotland Leman proved to be who first lead me into the world of Bayesian. His door is always open and always welcomes free talk on academics.

I would like to thank Leanna House for her teaching in hierarchical model, providing a powerful tool to explore the truth behind the complexness and a wide space to start my research. I would like to thank Marion Reynolds for his teaching in statistical inference. I have been delighted talking about classical statistics with him. I would thank to C. Michael Box for his installing new packages of R software in the lab for me. I would like to thank to all the faculty and staff here, where I feel like living in a big warm family.

I would like to thank my friends and colleagues at VT. These include Huaiye Zhang, Ciro Velasco-Cruz, Dipayan Maiti, Chongrui Yu, Sai Wang, Chao Han, Xiaobo Guo, Abdel-salam, Nels Johnson, David Peng, Laura Freeman, Jonathan Duggins and many others.

Finally, I would like to thank my family for their greatest and purest love. Long distance could not separate our beliefs and love. We are one, forever.

Contents

List of Tables

List of Figures

1 Introduction

High-throughput microarray has become one of the most important tools which have been widely used for functional genomics studies. Numerous statistical methods have been developed for use with these methods. However, most of the methods are single-gene based analyses which have not considered the dependencies among genes. In recent years, researchers have started looking at sets of genes instead of one gene at a time. This set of genes is predefined and is called a pathway. Ranking pathways relevant to a particular phenotype can help researchers focus on a few sets of genes. The advantage of the pathway based analysis is that it can detect subtle changes in gene expression levels which may not be possible with the single-gene based analysis (Mootha *et al.*, 2003; Hosack *et al.*, 2003; Rajagopalan and Agarwal, 2005).

A number of methods have been proposed to identify pathways relevant to a particular disease. Several papers have described the advantages of performing pathway based analysis. Goeman *et al.* (2004) proposed a global test based on the generalized linear random effects model. Random forest based analysis was proposed by Pang *et al.* (2006). The global test and the random forest approach are applicable to both continuous and binary outcomes. The global test is a model based analysis, while random forest is a tree based analysis. These two methods, and many other existing methods, are frequentist approaches.

In this thesis, a model based analysis is proposed using a generalized linear random effects model for binary events only, assuming there are only two categories. The model is based on a Bayesian hierarchical latent model. The advantage of the Bayesian approach is that it can not only clearly express a complex statistical model, but also easily incorporate prior knowledge when the sample size is small and the number of genes is large. The Gaussian random process was developed to construct the dependencies among genes in a pathway. For the covariance matrix of this process, two kernels were implemented: one was the polynomial

kernel and the other was the Gaussian kernel. Five models, corresponding to five different covariance structures, were considered as follows: Model 1 with the polynomial kernel, Model 2 with the Gaussian kernel, Model 3 with constant covariance matrix of the random effect, Model 4 with the mixture kernel of the polynomial kernel (or Gaussian) and point-mass density, Model 5 with the mixture kernel of the polynomial kernel (or Gaussian) and constant covariance matrix of the random effect.

The interesting question was whether τ^{-1} was zero in the covariance structure $\tau^{-1}\mathbf{K}(\mathbf{X})$ for each pathway. If τ^{-1} was zero, then it meant that the gene expression profile did not help to distinguish between the two binary groups.

The pathway where the Bayesian credible interval of τ^{-1} was far away from zero was selected. The top 50 pathways ranked by τ^{-1} were selected from 277 pathways using the different covariance structures of the Gaussian random process: the polynomial kernel based covariance matrix, the Gaussian kernel based covariance matrix. Pathways that overlapped were selected for analysis.

The constant covariance matrix of the random effect was proposed for the case that the genes expressions of the pathway did not help to distinguish between the two binary groups. The point-mass density was proposed for the case that the random effect $\boldsymbol{\gamma}(\mathbf{X})$ did not exist, implying that the genes expressions of the pathway did not help to distinguish between the two binary groups. The pathway was considered as not significant if the model had a larger likelihood value with constant covariance matrix of the random effect (or the point-mass density) than the polynomial kernel (or Gaussian). Models 1 and 2 were compared with Model 3 (which had a constant covariance matrix of the random effect) by using the ratio of the joint posterior distribution with respect to each model. The predictive power of a new observation was computed using a leave-one-out cross-validation approach.

For mixture models, log-likelihood values were computed at different values of the mixture

proportion, π, compared among the mixtures of selected kernels and point-mass density (or constant covariance matrix). When the polynomial or the Gaussian kernel had a higher log-likelihood value to have a large value of the mixture proportion π, the pathway was selected as significant.

This thesis is organized as follows. In Section 2, the Bayesian hierarchical latent model for the generalized linear mixed model is presented. The polynomial and the Gaussian kernels are introduced for constructing the covariance matrix for a set of random variables. In Section 3, the Bayesian approach is described and the full conditional distributions for each structure are derived. In Section 4, a Bayesian inference approach is suggested for this study. First of all, overlapping pathways between the top 50 pathways are selected with the polynomial and the Gaussian kernels ranked by τ^{-1}. Based on the overlapping pathways, Models 1 and 2 are compared with Model 3 by using the ratio of the joint posterior distribution with respect to each model and the predictive classification is obtained using leave-one-out cross-validation. Log-likelihood values for Models 4 and 5 are computed using different values of the mixture proportion π, to maximize the log-likelihood value of mixture models. In Section 5, the Bayesian approach is applied to type II diabetes data (Mootha *et al.*, 2003) which contains a microarray expression profile of 277 pathways and the result of data analysis is summarized. Section 6 contains concluding remarks.

2 Bayesian Hierarchical Model

Let Y_i be the binary response variable, $i = 1, \ldots, n$. The $Y_i = 1$ denotes that the sample i is diseased or one type of cancer and $Y_i = 0$ denotes that the sample i is normal. Let X_{ij} denotes the gene expression level of the jth gene in a pathway for ith sample, where $j = 1, ..., p$. Then the data matrix $\mathbf{X}$ can be expressed as

$$\mathbf{X}_{n\times p} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ . \\ . \\ . \\ \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & . & . & . & X_{1p} \\ X_{21} & X_{22} & . & . & . & X_{2p} \\ . & . & & . & & . \\ . & . & & . & & . \\ . & . & & . & & . \\ X_{n1} & X_{n2} & . & . & . & X_{np} \end{pmatrix}.$$

The Bayesian hierarchical model with a binary outcome $\mathbf{Y}_{n\times 1}$, a clinical covariate $\mathbf{C}_{n\times 2}$ where $\mathbf{C}_{n\times 2}$=($\mathbf{1}$, $\mathbf{c}$) and $\mathbf{c}$ is the $n \times 1$ vector of age, and a gene expression matrix $\mathbf{X}_{n\times p}$ is

$$\begin{aligned} Pr\{\mathbf{Y} = \mathbf{1}|\mathbf{C}, \boldsymbol{\gamma}(\mathbf{X})\} &= \Phi\{\mathbf{C}\boldsymbol{\alpha}' + \boldsymbol{\gamma}(\mathbf{X})\}, \\ \boldsymbol{\gamma}(\mathbf{X}) &\sim MN\{\mathbf{0}, \tau^{-1}\mathbf{K}(\mathbf{X})\}, \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal cumulative density function, $\boldsymbol{\alpha}$=(α_0,α_1) is the regression coefficient vector, the random effect matrix $\boldsymbol{\gamma}(\mathbf{X}) = \{\gamma(\mathbf{x}_1), \ldots, \gamma(\mathbf{x}_n)\}'$, and $\gamma(\cdot)$ follows the Gaussian random process with mean 0 and covariance $\text{cov}\{\gamma(\mathbf{x}_i), \gamma(\mathbf{x}_j)\} = \tau^{-1}K(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{K}$ is $n \times n$ matrix with ijth component $K(\mathbf{x}_i, \mathbf{x}_j)$.

The $\Phi(\cdot)$ links the linear function $\mathbf{C}\boldsymbol{\alpha}' + \boldsymbol{\gamma}(\mathbf{X})$ to the conditional probability of $\mathbf{Y} = \mathbf{1}$, which transforms a continuous model space to probability space $(0, 1)$. This is known as the probit regression model. Based on Albert and Chib (1993), latent variables $\mathbf{Z} = (Z_1, Z_2, ..., Z_n)'$ were defined with $\mathbf{Z} \sim MN\{\mathbf{C}\boldsymbol{\alpha}' + \boldsymbol{\gamma}(\mathbf{X}), \mathbf{I}\}$ such that

$$Y_i = \begin{cases} 1 & \text{if } Z_i \geq 0 \\ 0 & \text{if } Z_i < 0 \end{cases},$$

where $Pr\{Z_i \geq 0|\mathbf{C}_i, \gamma(\mathbf{x}_i)\} = Pr\{Y_i = 1|\mathbf{C}_i, \gamma(\mathbf{x}_i)\}$ and $Pr\{Z_i \geq 0|\mathbf{C}_i, \gamma(\mathbf{x}_i)\} = 1 - \Phi\{-(\mathbf{C}_i\boldsymbol{\alpha}' + \gamma(\mathbf{x}_i))\}$.

For the covariance $\text{cov}\{\gamma(\mathbf{x}_i), \gamma(\mathbf{x}_j)\}$, two kernels were considered: one was the polynomial kernel and the other was the Gaussian kernel as follows.

- Polynomial kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \mathbf{x}_j'$,
- Gaussian kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{||\mathbf{x}_i - \mathbf{x}_j||}{2})$.

The polynomial kernel made the dependence of the pathway effect stronger by increasing the absolute correlation between genes expressions of two samples if they had the same sign. The dependence structure based on the Gaussian kernel depended on the Euclidean distance between two genes expressions of two samples. The smaller the Euclidean distance, the stronger the dependence.

Five models were proposed corresponding to five covariance structure of $\boldsymbol{\gamma}(\mathbf{X})$:

- Model 1: $\boldsymbol{\gamma}(\mathbf{X}) \sim MN\{\mathbf{0}, \tau_p^{-1}\mathbf{K}(\mathbf{X})\}$, $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \mathbf{x}_j'$ (Polynomial kernel),
- Model 2: $\boldsymbol{\gamma}(\mathbf{X}) \sim MN\{\mathbf{0}, \tau_g^{-1}\mathbf{K}(\mathbf{X})\}$, $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{||\mathbf{x}_i - \mathbf{x}_j||}{2})$ (Gaussian kernel),
- Model 3: $\boldsymbol{\gamma}(\mathbf{X}) \sim MN\{\mathbf{0}, \tau_c^{-1}\mathbf{I}\}$ (Constant covariance matrix of the random effect),
- Model 4: $\boldsymbol{\gamma}(\mathbf{X}) \sim \pi MN\{\mathbf{0}, \tau^{-1}\mathbf{K}(\mathbf{X})\} + (1-\pi)\delta_0(\boldsymbol{\gamma})$ (The mixture kernel of the polynomial kernel (or Gaussian) and point-mass density),
- Model 5: $\boldsymbol{\gamma}(\mathbf{X}) \sim \pi MN\{\mathbf{0}, \tau^{-1}\mathbf{K}(\mathbf{X})\} + (1-\pi)MN\{\mathbf{0}, \tau_c^{-1}\mathbf{I}\}$ (The mixture kernel of the polynomial kernel (or Gaussian) and constant covariance matrix of the random effect).

The top 50 pathways ranked by τ^{-1} were selected from candidate pathways using the different covariance structures of the Gaussian random process: the polynomial kernel based covariance matrix, the Gaussian kernel based covariance matrix. The pathways that overlapped were selected for analysis.

The constant covariance matrix of the random effect was proposed for the case that the genes expressions of the pathway were useless in the random effect. The point-mass

density was proposed for the case that the random effect $\boldsymbol{\gamma}(\mathbf{X})$ did not exist. The pathway was considered as not significant if the model had a larger likelihood value with constant covariance matrix of the random effect (or point-mass density) than the polynomial kernel (or Gaussian).

Models 1 and 2, with the polynomial and Gaussian kernels, were compared with Model 3 which had constant covariance matrix of the random effect by using the ratio of the joint posterior distribution with respect to each model. The predictive power of a new observation was computed using a leave-one-out cross-validation approach.

For mixture models, log-likelihood values were computed at different values of the mixture proportion, π, compared among the mixtures of selected kernels and point-mass density (or constant covariance matrix). When the polynomial or the Gaussian kernel had a larger likelihood value with a large value of the mixture proportion π in the mixture model, the pathway was selected as significant.

3 Bayesian Approach

In this section, full conditional distributions of parameters were derived based on each model.

3.1 The prior and full conditional distributions

It was assumed that $\boldsymbol{\gamma}(\mathbf{X}) \sim MN\{\mathbf{0}, \tau^{-1}\mathbf{K}(\mathbf{X})\}$. In model 1, the polynomial kernel was used, $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x_i}\mathbf{x_j'}$, while the Gaussian kernel $K(\mathbf{x_i}, \mathbf{x_j}) = \exp(-||\mathbf{x_i} - \mathbf{x_j}||/\mathbf{2})$ was used in model 2.

Using the prior distributions $\boldsymbol{\alpha} \sim MN(\mathbf{0}, \boldsymbol{\phi}^{-1})$ and $\tau \sim \text{Gamma}(a, b)$, the joint posterior

distribution was derived as

$$\begin{aligned}[\mathbf{Z}, \alpha_0, \alpha_1, \boldsymbol{\gamma}, \tau | \mathbf{y}, \mathbf{C}, \mathbf{X}] &\propto \prod_{i=1}^{n} (I_{Z_i \geq 0} I_{y_i=1} + I_{Z_i<0} I_{y_i=0}) \times \prod_{i=1}^{n} N(Z_i; \alpha_0 + c_i\alpha_1 + \gamma_i, 1) \\ &\times MN\{\boldsymbol{\gamma}; \mathbf{0}, \tau^{-1}\mathbf{K}(\mathbf{X})\} \times \text{Gamma}(\tau; a, b) \\ &\times N(\alpha_0; 0, \phi^{-1}) \times N(\alpha_1; 0, \phi^{-1}),\end{aligned}$$

where $I_{(event)}$ was an indicator function, which equaled 1 if the event was true and otherwise equaled 0.

Since the model was a probit regression model, it allowed to have closed forms of the full conditional distributions of Z_i. The full conditional distributions for Z_i were

$$\begin{aligned}[Z_i | y_i = 0, \alpha_0, \alpha_1, \gamma_i, \tau] &\propto I_{Z_i<0}\ I_{y_i=0}\ N(Z_i; \alpha_0 + c_i\alpha_1 + \gamma_i, 1), \\ [Z_i | y_i = 1, \alpha_0, \alpha_1, \gamma_i, \tau] &\propto I_{Z_i\geq 0}\ I_{y_i=1}\ N(Z_i; \alpha_0 + c_i\alpha_1 + \gamma_i, 1),\end{aligned}$$

which were truncated normal distributions.

The full conditional distributions for $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, τ were proportional to

$$\begin{aligned}[\alpha_0 | Z_i, \alpha_1, \boldsymbol{\gamma}, \tau] &\propto N\{\frac{\sum_{i=1}^{n}(Z_i - c_i\alpha_1 - \gamma_i)}{n + \phi}, (n + \phi)^{-1}\}, \\ [\alpha_1 | Z_i, \alpha_0, \boldsymbol{\gamma}, \tau] &\propto N\{\frac{\sum_{i=1}^{n} c_i(Z_i - \alpha_0 - \gamma_i)}{\sum_{i=1}^{n} c_i^2 + \phi}, (\sum_{i=1}^{n} c_i^2 + \phi)^{-1}\}, \\ [\boldsymbol{\gamma} | Z_i, \alpha_0, \alpha_1, \tau] &\propto MN[\{\mathbf{I} + (\tau^{-1}\mathbf{K}(\mathbf{X}))^{-1}\}^{-1}(\mathbf{Z} - \mathbf{C}\boldsymbol{\alpha}'), \{\mathbf{I} + (\tau^{-1}\mathbf{K}(\mathbf{X}))^{-1}\}^{-1}], \\ [\tau | Z_i, \alpha_0, \alpha_1, \boldsymbol{\gamma}] &\propto \text{Gamma}\{a + \frac{1}{2}, b + \frac{1}{2}\boldsymbol{\gamma}'(\mathbf{K}(\mathbf{X}))^{-1}\boldsymbol{\gamma}\}.\end{aligned}$$

Since the full conditional distributions for all parameters had closed forms, the Gibbs sampling algorithm was used to generate a sequence of variables from the distribution of each variable in order, which was conditional on the previously obtained variables.

For model 3, it was assumed that $\boldsymbol{\gamma} \sim MN(0, \tau_c^{-1}I)$ where the kernel $\mathbf{K}(\mathbf{X})$ was replaced with the identity matrix, implying that the gene expression $\mathbf{X}$ in the pathway provided no information. The full conditional distribution for $\boldsymbol{\gamma}$ was derived as

$$[\boldsymbol{\gamma} | Z_i, \alpha_0, \alpha_1, \tau_c] \propto MN\{(\mathbf{I} + \tau_c\mathbf{I})^{-1}(\mathbf{Z} - \mathbf{C}\boldsymbol{\alpha}'), (\mathbf{I} + \tau_c\mathbf{I})^{-1}\}.$$

For other parameters, the full conditional distributions were derived in a similar way.

3.2 The prior and full conditional distributions based on mixture models

In model 4, it was assumed that $\boldsymbol{\gamma}(\mathbf{X}) \sim \pi MN\{\mathbf{0}, \tau^{-1}\mathbf{K}(\mathbf{X})\} + (1-\pi)\delta_0(\boldsymbol{\gamma})$. The random variable was defined as $L \sim Ber(\pi)$. Using the prior distributions $\boldsymbol{\alpha} \sim MN(\mathbf{0}, \boldsymbol{\phi}^{-1})$ and $\tau \sim \text{Gamma}(a, b)$, the joint posterior distribution was derived as

$$\begin{aligned}
[\mathbf{Z}, L, \alpha_0, \alpha_1, \boldsymbol{\gamma}, \tau | \mathbf{y}, \mathbf{C}, \mathbf{X}] &\propto \prod_{i=1}^{n}(I_{Z_i \geq 0} I_{y_i=1} + I_{Z_i<0} I_{y_i=0}) \times \prod_{i=1}^{n} N(Z_i; \alpha_0 + c_i\alpha_1 + \gamma_i, 1) \\
&\times [I_{L=1} MN\{\boldsymbol{\gamma}; \mathbf{0}, \tau^{-1}K(\mathbf{X})\} + I_{L=0}\delta(0)] \\
&\times \text{Gamma}(\tau; a, b) \times N(\alpha_0; 0, \phi^{-1}) \times N(\alpha_1; 0, \phi^{-1}).
\end{aligned}$$

Then the full conditional distributions were proportional to

$$\begin{aligned}
[Z_i | y_i = 0, L, \alpha_0, \alpha_1, \boldsymbol{\gamma}, \tau] &\propto I_{Z_i<0}\ I_{y_i=0}\ N(Z_i; \alpha_0 + c_i\alpha_1 + \gamma_i, 1), \\
[Z_i | y_i = 1, L, \alpha_0, \alpha_1, \boldsymbol{\gamma}, \tau] &\propto I_{Z_i \geq 0}\ I_{y_i=1}\ N(Z_i; \alpha_0 + c_i\alpha_1 + \gamma_i, 1), \\
[\alpha_0 | Z_i, L, \alpha_1, \boldsymbol{\gamma}, \tau] &\propto N\{\frac{\sum_{i=1}^{n}(Z_i - c_i\alpha_1 - \gamma_i)}{n+\phi}, (n+\phi)^{-1}\}, \\
[\alpha_1 | Z_i, L, \alpha_0, \boldsymbol{\gamma}, \tau] &\propto N[\frac{\sum_{i=1}^{n} c_i(Z_i - \alpha_0 - \gamma_i)}{\sum_{i=1}^{n} c_i^2 + \phi}, \{\sum_{i=1}^{n}(c_i^2 + \phi)\}^{-1}], \\
[\boldsymbol{\gamma} | L = 1, Z_i, \alpha_0, \alpha_1, \tau] &\propto MN[\{\mathbf{I} + (\tau^{-1}K(\mathbf{X}))^{-1}\}^{-1}(\mathbf{Z} - \mathbf{C}\boldsymbol{\alpha}'), \{\mathbf{I} + (\tau^{-1}K(\mathbf{X}))^{-1}\}^{-1}], \\
[\boldsymbol{\gamma} | L = 0, Z_i, \alpha_0, \alpha_1, \tau] &\propto MN\{(\mathbf{Z} - \mathbf{C}\boldsymbol{\alpha}'), \mathbf{I}\}, \\
[\tau | L = 1, Z_i, \alpha_0, \alpha_1, \boldsymbol{\gamma}] &\propto \text{Gamma}[a + \frac{1}{2}, b + \frac{1}{2}\boldsymbol{\gamma}'\{K(\mathbf{X})\}^{-1}\boldsymbol{\gamma}], \\
[\tau | L = 0, Z_i, \alpha_0, \alpha_1, \boldsymbol{\gamma}] &\propto \text{Gamma}(a, b).
\end{aligned}$$

In model 5, it was assumed that $\boldsymbol{\gamma}(\mathbf{X}) \sim \pi MN\{\mathbf{0}, \tau^{-1}\mathbf{K}(\mathbf{X})\} + (1-\pi)MN\{\mathbf{0}, \tau_c^{-1}\mathbf{I}\}$, the joint posterior distribution was

$$\begin{aligned}
[\mathbf{Z}, L, \alpha_0, \alpha_1, \boldsymbol{\gamma}, \tau | y] &\propto \prod_{i=1}^{n}(I_{Z_i \geq 0} I_{y_i=1} + I_{Z_i<0} I_{y_i=0}) \times \prod_{i=1}^{n} N(Z_i; \alpha_0 + c_i\alpha_1 + \gamma_i, 1) \\
&\times [I_{L=1} MN\{\boldsymbol{\gamma}; \mathbf{0}, \tau^{-1}K(\mathbf{X})\} + I_{L=0} MN(\boldsymbol{\gamma}; \mathbf{0}, \tau_c^{-1}\mathbf{I})]
\end{aligned}$$

$$\times \quad \text{Gamma}(\tau; a, b) \times N(\alpha_0; 0, \phi^{-1}) \times N(\alpha_1; 0, \phi^{-1}),$$

and the full conditional distributions for $\boldsymbol{\gamma}$ were derived as

$$\begin{aligned}
[\boldsymbol{\gamma}|L=1, Z_i, \alpha_0, \alpha_1, \tau] &\propto MN[\{\mathbf{I} + (\tau^{-1}K(\mathbf{X}))^{-1}\}^{-1}(\mathbf{Z} - \mathbf{C}\boldsymbol{\alpha}'), \{\mathbf{I} + (\tau^{-1}K(\mathbf{X}))^{-1}\}^{-1}], \\
[\boldsymbol{\gamma}|L=0, Z_i, \alpha_0, \alpha_1, \tau_c] &\propto MN\{(\mathbf{I} + \tau_c\mathbf{I})^{-1}(\mathbf{Z} - \mathbf{C}\boldsymbol{\alpha}'), (\mathbf{I} + \tau_c\mathbf{I})^{-1}\}.
\end{aligned}$$

For other parameters, the full conditional distributions were the same as the previous mixture case.

3.3 The Algorithm

Since the full conditional distributions have closed forms for all models, the Gibbs sampling is applicable. The Gibbs sampling algorithm is following:

- Step 1: Initialize $[\mathbf{Z}^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\gamma}^{(0)}, \tau^{(0)}]$.
- Step 2: At the t-th iteration,
  - (i) Draw $\mathbf{Z}^{(t)}$ from $[\mathbf{Z}^{(t)}|\mathbf{Z}^{(t-1)}, \boldsymbol{\alpha}^{(t-1)}, \boldsymbol{\gamma}^{(t-1)}]$.
  - (ii) Draw $\boldsymbol{\alpha}^{(t)}$ from $[\boldsymbol{\alpha}^{(t)}|\mathbf{Z}^{(t)}, \boldsymbol{\alpha}^{(t-1)}, \boldsymbol{\gamma}^{(t-1)}]$.
  - (iii) Draw $\boldsymbol{\gamma}^{(t)}$ from $[\boldsymbol{\gamma}^{(t)}|\mathbf{Z}^{(t)}, \boldsymbol{\alpha}^{(t)}, \tau^{(t-1)}]$.
  - (iv) Draw $\tau^{(t)}$ from $[\tau^{(t)}|\boldsymbol{\gamma}^{(t)}]$.
  - (v) Draw $L^{(t)}$ from $[L^{(t)}|\pi]$ (The draw for L is considered only for the case of the mixture model 4 or 5).
- Step 3: Increase t until the required the number of iterations, $M = 10,000$.
- Stop

Typically in Bayesian computing, $M = 10,000$ was preferred. For the parameters of priors, it was proposed that $a = 50$, $b = 0.5$ and $\phi^{-1} = 0.5$ here. In addition, different parameters would be chosen in the future work for sensitivity analysis.

4 Bayesian Inference

Using the Gibbs sampling algorithm, the MCMC samples at the t th iteration, $\{\mathbf{Z}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\gamma}^{(t)}, \tau^{(t)}, t = 1, \ldots, M\}$, were obtained after burn-in period. These samples were used for posterior inference and prediction. A 95% Bayesian credible interval was calculated. The median and mean were obtained for the point estimator. MCMC trace plots and histograms were used to examine whether the posterior distributions had converged. The interesting question was whether τ^{-1} was zero in the covariance structure $\tau^{-1}\mathbf{K}(\mathbf{X})$ for each pathway since τ^{-1} was zero meant that the gene expression profile did not help to distinguish between the two binary groups. The pathway where the Bayesian credible interval of τ^{-1} was far away from zero was selected.

Another interesting question was whether $\tau^{-1}\mathbf{K}(\mathbf{X})$ was the same as $\tau^{-1}\mathbf{I}$ which also implied that the gene expression profile did not help to distinguish between the two binary groups. The ratio, using the joint posterior distribution with respect to each model, was used to compare the three models. Using the posterior samples, it was counted that how many times the ratio was larger than 1 between model j and model 3, $j = 1, 2$

$$\begin{aligned} Ratio_{j3}^{t} &= \frac{\prod_{i=1}^{n} N(Z_{i,j}^{t}; \alpha_{0,j}^{t} + \alpha_{1,j}^{t} C_i + \gamma_{i,j}^{t}, 1) \times MN\{\boldsymbol{\gamma}_j{}^{t}; \mathbf{0}, \tau_j^{-1,t}(\mathbf{K}(\mathbf{X}))\}}{\prod_{i=1}^{n} N(Z_{i,3}^{t}; \alpha_{0,3}^{t} + \alpha_{1,3}^{t} C_i + \gamma_{i,3}^{t}, 1) \times MN(\boldsymbol{\gamma}_3{}^{t}; \mathbf{0}, \tau_3^{-1,t}\mathbf{I})} \\ &\times \frac{\text{Gamma}(\tau_j^{t}; a, b) \times N(\alpha_{0,j}^{t}; 0, \phi^{-1}) \times N(\alpha_{1,j}^{t}; 0, \phi^{-1})}{\text{Gamma}(\tau_3^{t}; a, b) \times N(\alpha_{0,3}^{t}; 0, \phi^{-1}) \times N(\alpha_{1,3}^{t}; 0, \phi^{-1})}. \end{aligned}$$

The predictive classification of a new observation $Y_{i,new}$ was obtained using the leave-one-out cross-validation, conditioning on the expression levels in each pathway as

$$Pr(Y_{i,new} = 1|\mathbf{X}, \mathbf{C}) = \frac{1}{M} \sum_{t=1}^{M} P(Y_{i,new} = 1|\mathbf{X}, \mathbf{C}, \mathbf{Z}^{(t)}, \tau^{(t)}, \boldsymbol{\gamma}^{(t)}),$$

where $P\{Y_{i,new} = 1|\mathbf{X}, \mathbf{C}, \mathbf{Z}^{(t)}, \tau^{(t)}, \boldsymbol{\gamma}^{(t)}\} = \Phi\{\mathbf{C}_i \boldsymbol{\alpha}'^{(t)}_{-i} + \boldsymbol{\gamma}_i^{(t)}\}$ and $\boldsymbol{\gamma}_i^{(t)} \sim MN\{\mathbf{0}, \tau_{-i}^{-1,(t)}\mathbf{K}(\mathbf{x}_i)\}$.

For mixture models, log-likelihood values were computed at different values of the mixture proportion, π, where $\pi = (0.1, 0.3, 0.5, 0.7, 0.9)$, compared among the mixtures of selected

kernels and point-mass density (or constant covariance matrix). When the polynomial or the Gaussian kernel had a higher log-likelihood value to have a large value of the mixture proportion π, the pathway was selected as a significant.

Likelihood for Model 4, the mixture kernel of the polynomial kernel (or Gaussian) and point-mass density, where $\boldsymbol{\theta}$=($\mathbf{Z}$,$\boldsymbol{\gamma}$),

$$\int \prod_{i=1}^{n} (I_{Z_i \geq 0} I_{y_i=1} + I_{Z_i<0} I_{y_i=0}) \times \prod_{i=1}^{n} N(Z_i; \alpha_0 + c_i\alpha_1 + \gamma_i, 1) \times [\pi MN\{\boldsymbol{\gamma}; \mathbf{0}, \tau^{-1}K(\mathbf{X})\} + (1-\pi)\delta(0)] d\boldsymbol{\theta}.$$

Likelihood for Model 5, the mixture kernel of the polynomial kernel (or Gaussian) and constant covariance matrix of the random effect, where $\boldsymbol{\theta}$=($\mathbf{Z}$,$\boldsymbol{\gamma}$),

$$\int \prod_{i=1}^{n} (I_{Z_i \geq 0} I_{y_i=1} + I_{Z_i<0} I_{y_i=0}) \times \prod_{i=1}^{n} N(Z_i; \alpha_0 + c_i\alpha_1 + \gamma_i, 1) \times [\pi MN\{\boldsymbol{\gamma}; \mathbf{0}, \tau^{-1}K(\mathbf{X})\} + (1-\pi) MN(\boldsymbol{\gamma}; \mathbf{0}, \tau_c^{-1}\mathbf{I})] d\boldsymbol{\theta}.$$

5 Example

The Bayesian approach was applied to the microarray gene expression data with $22,283$ genes on type II diabetes (Mootha *et al.*, 2003). In the data, there were 17 samples with normal glucose tolerance and 18 samples with type II diabetes mellitus. 277 pathways were considered including 128 KEGG pathways and 149 curated pathways. The KEGG pathway database(http://www.genome.jp/kegg/pathway.html) was a collection of curated pathways representing current knowledge on the molecular interaction and reaction Networks for metabolism, genetic information processing, environmental information process, cellular processes, and human disease. The 149 curated pathways were constructed from known biological experiments by Mootha and colleagues. The example data was given in table 1.

In the analysis, Y was set as the binary clinical outcomes corresponding to either normal or type II diabetes, a clinical covariate $\mathbf{C}_{n\times 2}$ where $\mathbf{C}_{n\times 2}$=($\mathbf{1}$, $\mathbf{c}$) and $\mathbf{c}$ was the $n \times 1$ vector of age for each subject, and $\mathbf{X}$ was the $n \times p$ gene expression, where $n = 35$, p was the number of genes in a specific pathway, where p varied from 2 to 200 across these pathways. The goal was to identify pathways to distinguish between two groups (normal vs. type II diabetes). To identify significant pathways, the Bayesian hierarchical model was used and parameters estimated. The top 50 pathways were selected using the polynomial and the Gaussian kernels based on the rank of τ^{-1}. Using 23 pathways overlapped these top 50 pathways, it was determined as follows. One of overlapping pathways is pathway 229, "Oxidative phosphorylation", known to be related to diabetes (Misu *et al.*, 2007; Mootha *et al.*, 2003; Mootha *et l.*, 2004). This is a process of cellular respiration in humans (or in general eukaryotes). The pathway contains coregulated genes across different tissues which are related to insulin/glucose disposal. It consists of ATP synthesis, a pathway involved in energy transfer. Another pathway is pathway 36, c17 U133 probes, which is also selected as significant (Kim *et al*, 2009).

6 Conclusion and Discussion

In this thesis, a Bayesian method was developed for pathway based analysis. The approach was a model based analysis for a generalized linear random effects model. The Bayesian probit regression model was used because it can derive closed forms of full conditional distribution of parameters so that the Gibbs sampling algorithm can be applied. The Gaussian random process was considered to construct the dependencies among genes in a pathway. For covariance matrix of this process, two kernels were implemented: one was the polynomial kernel and the other was the Gaussian kernel. The polynomial kernel was used because it made that the dependence of the pathway effect strong, between genes expres-

sions of two samples, if they had the same sign. The dependence structure based on the Gaussian kernel depended on the Euclidean distance between genes expressions of the two samples. Five models, corresponding to five different covariance structures, were considered as follows: Model 1 with the polynomial kernel, Model 2 with the Gaussian kernel, Model 3 with constant covariance matrix of the random effect, Model 4 with the mixture kernel of the polynomial kernel (or Gaussian) and point-mass density, Model 5 with the mixture kernel of the polynomial kernel (or Gaussian) and constant covariance matrix of the random effect.

The interesting question was whether τ^{-1} was zero in the covariance structure $\tau^{-1}\mathbf{K}(\mathbf{X})$ for each pathway since a zero value of τ^{-1} meant that the gene expression profile did not help to distinguish between the two binary groups.

The pathway where the Bayesian credible interval of τ^{-1} was far away from zero was selected. The approach was applied to a data set from (Mootha *et al.*, 2003) which was the gene expression profiles of type II diabetes. The top 50 pathways ranked by τ^{-1} were selected from 277 candidate pathways using the different covariance structures of the Gaussian random process: the polynomial kernel based covariance matrix, the Gaussian kernel based covariance matrix. The overlapped pathways were selected for analysis.

The constant covariance matrix of the random effect was proposed for the case that the genes expressions of the pathway did not help to distinguish between the two binary groups. The point-mass density was proposed for the case that the random effect $\boldsymbol{\gamma}(\mathbf{X})$ did not exist, implying that the genes expressions of the pathway did not help to distinguish between the two binary groups. The pathway was considered as not significant if the model had a larger likelihood value with constant covariance matrix of the random effect (or the point-mass density) than the polynomial kernel (or Gaussian). Models 1 and 2 were compared with Model 3 (which had a constant covariance matrix of the random effect) by using the ratio of

the integration of the joint posterior distribution with respect to parameters. The predictive power of a new observation was computed using a leave-one-out cross-validation approach.

For mixture models, log-likelihood values were computed at different values of the mixture proportion, π, compared among the mixtures of selected kernels and point-mass density (or constant covariance matrix). When the polynomial or the Gaussian kernel had a higher log-likelihood value to have a large value of the mixture proportion π, the pathway was selected as significant.

The example data was given in table 1. The results based on the polynomial kernel and the Gaussian kernel were given in Table 2-3 and Table 4-5, respectively. The proportion of $Ratio_{i3} < 1$ in $10,000$ iterations which was slightly larger than 50%, meant that Model 3 was slightly better. But it could not help guarantee the pathway was significant or not for they were almost too close to 50%. Then the log-likelihood comparison based on mixture model was proposed given different mixture proportions. The result based on the mixture of selected kernels and point-mass density was shown in Table 8. The mixture of selected kernels and constant covariance structure was summarized in Table 9. The results suggested that when the mixing proportion was 0.9, the largest log-likelihood values were obtained, meaning that the selected pathways were highly significant. The 23 overlapping pathways were given in Table 6, including pathway 229 and 36. Pathway 229, "Oxidative phosphorylation", was known to be related to diabetes (Misu *et al.*, 2007; Mootha *et al.*, 2003; Mootha *et l.*, 2004). Pathway 36, c17 U133 probes, was also selected as significant (Kim *et al*, 2009). Using leave-one-out classification, the predictive classification of a new observation $Y_{i,new}$ was obtained, conditioning on the expression levels in each overlapped pathway. These results were given in Table 7. Five number summary statistics values were summarized. The predictive probability of pathway 229 was about 0.53 for the diabetes group and was about 0.47 for the normal group. For pathway 36, the predictive probability was about 0.53 for

the diabetes group and was about 0.47 for the normal group. Similar results were obtained using both Kernels.

The MCMC traces plots of pathway 229 and pathway 36 based on the Gaussian kernel were shown in figure 1 and 2, respectively. For the mixture of the Gaussian kernel and constant covariance matrix, the MCMC traces plots of them were shown in figure 3 and 4.

For the parameters of priors, it was proposed that $a = 50$, $b = 0.5$ and $\phi^{-1} = 0.5$ here. In addition, different priors values would be chosen in the future work for sensitivity analysis. Simply using kernels may not be enough to characterize the dependencies among genes for all pathways. The multivariate adaptive regression splines (Friedman, 1991) approach may be one possible way to model more flexible dependence than the kernel approach. But this approach may require a large sample size.

Pathway	PID	Gene	NGT_1	...	NGT_{17}	DM_1	...	DM_{18}
1	200862_at	DHCR24	4.9912		5.0194	5.2512		4.8611
1	207708_at	ALOXE3	4.9813		5.3201	5.4803		5.1668
2	207386_at	CYP7B1	4.3190		4.0180	3.9538		3.6331
2	207708_at	ALOXE3	4.9813		5.3201	5.4803		5.1668
2	218760_at	COQ6	4.0682		4.7695	4.8249		4.1371
3	$200844_{sa}t$	PRDX6	6.9080		6.9473	6.7995		6.8190
3	$200845_{sa}t$	PRDX6	6.4951		6.1733	6.4355		6.5557
3	207708_at	ALOXE3	4.9813		5.3201	5.4803		5.1668
4	200027_at	NARS	5.9270		5.3933	6.0312		5.7391
4	200708_at	GOT2	6.8930		6.4758	7.2996		6.8327
4	201000_at	AARS	5.5510		5.6695	5.7400		5.8711
4	$201623_{sa}t$	DARS	6.2898		5.8505	6.1421		6.1796
4	201624_at	DARS	4.4580		3.9546	3.8449		4.1070
4	$202144_{sa}t$	ADSL	6.3618		6.8327	6.6805		6.8396
4	$204476_{sa}t$	PC	4.6400		2.5805	3.8519		4.5906
4	$205843_{xa}t$	CRAT	4.0182		4.3577	4.8755		5.1928
4	206030_at	ASPA	4.7185		4.2734	4.4220		3.8625
4	206527_at	ABAT	4.3968		4.8115	4.5499		4.5225
4	206780_at	GAD2	2.8073		3.3191	2.7466		3.1017
4	$207076_{sa}t$	ASS	5.3591		4.9363	5.3544		5.3197
4	208813_at	GOT1	6.8694		6.2766	6.4117		5.9442
4	$209522_{sa}t$	CRAT	5.5242		5.1917	5.8957		5.8748
4	$210250_{xa}t$	ADSL	6.5176		6.7020	6.5267		6.5326
4	210326_at	AGXT	3.0206		4.0713	3.2321		2.8512
4	$216651_{sa}t$	GAD2	2.9466		3.6256	4.0045		4.0190
4	221761_at	ADSS	4.0723		4.2449	4.6208		4.3391
.				...			...	
277	$209610_{sa}t$	PSAT1	5.5405		4.7737	4.6406		5.1657
.				...			...	
277	$220892_{sa}t$	PSAT1	4.0806		3.4867	4.4169		4.4060

Table 1: Part of the microarray gene expression data on type II diabetes including $22,283$ genes within 277 pathways, NGT_i=sample i normal and DM_i=sample i with type II diabetes.

Rank	Pathway	$\widehat{(\tau^{-1})}$	$\widehat{\sigma(\tau^{-1})}$	2.5% *Bayesian C.I.*	97.5% *Bayesian C.I.*	$(Ratio_{13} < 1)\%$
1	*27*	0.0157	0.0028	0.0120	0.0175	50.66%
2	*229*	0.0157	0.0029	0.0121	0.0174	49.89%
3	*60*	0.0157	0.0028	0.0120	0.0175	50.23%
4	*177*	0.0157	0.0029	0.0120	0.0175	50.35%
5	*52*	0.0157	0.0029	0.0120	0.0175	50.76%
6	*43*	0.0157	0.0029	0.0120	0.0175	50.13%
7	*269*	0.0157	0.0029	0.0121	0.0174	50.33%
8	*34*	0.0157	0.0028	0.0120	0.0174	51.16%
9	*50*	0.0157	0.0029	0.0120	0.0175	50.12%
10	*36*	0.0157	0.0029	0.0120	0.0175	50.49%
11	*232*	0.0157	0.0028	0.0120	0.0174	49.90%
12	*32*	0.0157	0.0029	0.0120	0.0175	50.04%
13	*126*	0.0157	0.0028	0.0120	0.0174	50.13%
14	*109*	0.0157	0.0028	0.0120	0.0174	49.66%
15	*70*	0.0157	0.0028	0.0121	0.0174	50.43%
16	*248*	0.0157	0.0028	0.0120	0.0174	49.44%
17	*35*	0.0157	0.0028	0.0120	0.0175	50.19%
18	*67*	0.0157	0.0028	0.0120	0.0174	49.69%
19	*30*	0.0156	0.0028	0.0120	0.0174	49.80%
20	*249*	0.0156	0.0028	0.0120	0.0174	49.82%
21	*213*	0.0156	0.0028	0.0120	0.0174	49.61%
22	*102*	0.0156	0.0028	0.0121	0.0174	50.31%
23	*66*	0.0156	0.0028	0.0120	0.0174	49.89%
24	*259*	0.0156	0.0028	0.0120	0.0174	50.00%
25	*104*	0.0156	0.0028	0.0120	0.0174	49.72%

Table 2: Top 1-25 pathways selected by Bayesian approach using the polynomial kernel and ranked by τ^{-1}, $\widehat{(\tau^{-1})}$ =the estimation of τ^{-1}, $\widehat{\sigma(\tau^{-1})}$ =the standard deviation of τ^{-1}, *Bayesian C.I.*=the Bayesian credible interval of τ^{-1} and $(Ratio_{13} < 1)\%$=the proportion of $Ratio_{13} < 1$ in $10,000$ iterations.

Rank	Pathway	$\widehat{(\tau^{-1})}$	$\widehat{\sigma(\tau^{-1})}$	2.5% *Bayesian C.I.*	97.5% *Bayesian C.I.*	$(Ratio_{13} < 1)\%$
26	*44*	0.0156	0.0028	0.0120	0.0174	49.45%
27	*115*	0.0156	0.0028	0.0120	0.0174	49.10%
28	*59*	0.0156	0.0028	0.0120	0.0174	50.76%
29	*186*	0.0156	0.0028	0.0120	0.0174	48.79%
30	*63*	0.0156	0.0028	0.0120	0.0173	49.50%
31	*56*	0.0156	0.0028	0.0120	0.0174	50.13%
32	*137*	0.0156	0.0028	0.0120	0.0174	50.11%
33	*242*	0.0156	0.0028	0.0120	0.0174	49.88%
34	*278*	0.0156	0.0028	0.0120	0.0173	50.15%
35	*107*	0.0156	0.0028	0.0120	0.0174	49.93%
36	*98*	0.0156	0.0028	0.0120	0.0174	50.24%
37	*149*	0.0156	0.0028	0.0120	0.0174	50.22%
38	*221*	0.0156	0.0028	0.0121	0.0173	50.12%
39	*154*	0.0156	0.0028	0.0120	0.0174	50.55%
40	*222*	0.0156	0.0028	0.0120	0.0174	50.50%
41	*91*	0.0156	0.0028	0.0120	0.0173	50.45%
42	*38*	0.0156	0.0028	0.0120	0.0174	49.20%
43	*49*	0.0156	0.0028	0.0120	0.0174	50.89%
44	*39*	0.0156	0.0027	0.0121	0.0173	50.27%
45	*253*	0.0156	0.0028	0.0120	0.0174	50.16%
46	*53*	0.0156	0.0028	0.0120	0.0174	50.29%
47	*41*	0.0156	0.0028	0.0120	0.0174	50.98%
48	*86*	0.0156	0.0028	0.0120	0.0174	49.78%
49	*29*	0.0156	0.0028	0.0120	0.0174	49.91%
50	*238*	0.0156	0.0028	0.0120	0.0174	49.77%

Table 3: Top 26-50 pathways selected by Bayesian approach using the polynomial kernel and ranked by τ^{-1}, $\widehat{(\tau^{-1})}$ =the estimation of τ^{-1}, $\widehat{\sigma(\tau^{-1})}$ =the standard deviation of τ^{-1}, *Bayesian C.I.*=the Bayesian credible interval of τ^{-1} and $(Ratio_{13} < 1)\%$=the proportion of $Ratio_{13} < 1$ in $10,000$ iterations.

Rank	Pathway	$\widehat{(\tau^{-1})}$	$\widehat{\sigma(\tau^{-1})}$	2.5% *Bayesian C.I.*	97.5% *Bayesian C.I.*	$(Ratio_{23} < 1)\%$
1	*249*	0.0158	0.0029	0.0121	0.0176	49.79%
2	*27*	0.0157	0.0028	0.0120	0.0175	50.38%
3	*253*	0.0157	0.0028	0.0121	0.0174	50.47%
4	*213*	0.0157	0.0029	0.0120	0.0175	50.47%
5	*229*	0.0157	0.0028	0.0121	0.0175	50.53%
6	*257*	0.0157	0.0028	0.0121	0.0174	50.92%
7	*37*	0.0157	0.0028	0.0121	0.0175	49.92%
8	*238*	0.0157	0.0028	0.0121	0.0174	50.65%
9	*50*	0.0157	0.0028	0.0121	0.0175	50.19%
10	*97*	0.0157	0.0028	0.0120	0.0174	50.77%
11	*270*	0.0157	0.0029	0.0120	0.0174	50.75%
12	*138*	0.0157	0.0028	0.0120	0.0175	51.11%
13	*92*	0.0157	0.0029	0.0120	0.0174	50.28%
14	*62*	0.0157	0.0028	0.0121	0.0174	49.93%
15	*177*	0.0157	0.0028	0.0120	0.0175	49.96%
16	*93*	0.0157	0.0028	0.0120	0.0174	50.06%
17	*28*	0.0157	0.0028	0.0121	0.0174	50.32%
18	*230*	0.0157	0.0028	0.0120	0.0174	50.47%
19	*135*	0.0157	0.0028	0.0120	0.0173	51.01%
20	*256*	0.0157	0.0028	0.0120	0.0174	50.00%
21	*43*	0.0157	0.0028	0.0121	0.0174	50.68%
22	*56*	0.0156	0.0028	0.0121	0.0174	49.77%
23	*44*	0.0156	0.0028	0.0120	0.0175	50.40%
24	*79*	0.0156	0.0028	0.0120	0.0174	50.11%
25	*35*	0.0156	0.0028	0.0120	0.0174	50.11%

Table 4: Top 1-25 pathways selected by Bayesian approach using the Gaussian kernel and ranked by τ^{-1}, $\widehat{(\tau^{-1})}$ =the estimation of τ^{-1}, $\widehat{\sigma(\tau^{-1})}$ =the standard deviation of τ^{-1}, *Bayesian C.I.*=the Bayesian credible interval of τ^{-1} and $(Ratio_{23} < 1)\%$=the proportion of $Ratio_{23} < 1$ in $10,000$ iterations.

Rank	Pathway	$\widehat{(\tau^{-1})}$	$\widehat{\sigma(\tau^{-1})}$	2.5% *Bayesian C.I.*	97.5% *Bayesian C.I.*	$(Ratio_{23} < 1)\%$
26	*110*	0.0156	0.0028	0.0120	0.0174	49.85%
27	*47*	0.0156	0.0028	0.0121	0.0173	50.06%
28	*246*	0.0156	0.0028	0.0121	0.0174	50.34%
29	*218*	0.0156	0.0028	0.0120	0.0174	50.25%
30	*54*	0.0156	0.0028	0.0120	0.0174	49.65%
31	*102*	0.0156	0.0029	0.0120	0.0174	49.68%
32	*221*	0.0156	0.0028	0.0120	0.0174	50.01%
33	*104*	0.0156	0.0028	0.0120	0.0174	50.11%
34	*60*	0.0156	0.0028	0.0120	0.0174	49.73%
35	*36*	0.0156	0.0028	0.0120	0.0174	50.20%
36	*105*	0.0156	0.0028	0.0120	0.0174	50.20%
37	*63*	0.0156	0.0028	0.0120	0.0174	49.85%
38	*248*	0.0156	0.0028	0.0120	0.0174	50.11%
39	*116*	0.0156	0.0028	0.0120	0.0174	49.94%
40	*243*	0.0156	0.0028	0.0120	0.0174	50.00%
41	*31*	0.0156	0.0028	0.0121	0.0174	50.00%
42	*153*	0.0156	0.0028	0.0120	0.0173	50.24%
43	*39*	0.0156	0.0028	0.0120	0.0174	50.32%
44	*94*	0.0156	0.0028	0.0120	0.0174	49.91%
45	*57*	0.0156	0.0028	0.0120	0.0174	49.94%
46	*268*	0.0156	0.0028	0.0119	0.0174	50.40%
47	*107*	0.0156	0.0028	0.0120	0.0173	49.92%
48	*232*	0.0156	0.0028	0.0119	0.0173	50.77%
49	*32*	0.0156	0.0028	0.0120	0.0174	50.87%
50	*95*	0.0156	0.0028	0.0120	0.0174	50.25%

Table 5: Top 26-50 pathways selected by Bayesian approach using the Gaussian kernel and ranked by τ^{-1}, $\widehat{(\tau^{-1})}$ =the estimation of τ^{-1}, $\widehat{\sigma(\tau^{-1})}$ =the standard deviation of τ^{-1}, *Bayesian C.I.*=the Bayesian credible interval of τ^{-1} and $(Ratio_{23} < 1)\%$=the proportion of $Ratio_{23} < 1$ in $10,000$ iterations.

	Polynomial Kernel				Gaussian Kernel			
Pathway	$\widehat{(\tau^{-1})}$	$\widehat{\sigma(\tau^{-1})}$	2.5% *C.I.*	97.5% *C.I.*	$\widehat{(\tau^{-1})}$	$\widehat{\sigma(\tau^{-1})}$	2.5% *C.I.*	97.5% *C.I.*
27	0.0157	0.0028	0.0120	0.0175	0.0157	0.0028	0.0120	0.0175
32	0.0157	0.0029	0.0120	0.0175	0.0156	0.0028	0.0120	0.0174
35	0.0157	0.0028	0.0120	0.0175	0.0156	0.0028	0.0120	0.0174
36	0.0157	0.0029	0.0120	0.0175	0.0156	0.0028	0.0120	0.0174
39	0.0156	0.0027	0.0121	0.0173	0.0156	0.0028	0.0120	0.0174
43	0.0157	0.0029	0.0120	0.0175	0.0157	0.0028	0.0121	0.0174
44	0.0156	0.0028	0.0120	0.0174	0.0156	0.0028	0.0120	0.0175
50	0.0157	0.0029	0.0120	0.0175	0.0157	0.0028	0.0121	0.0175
56	0.0156	0.0028	0.0120	0.0174	0.0156	0.0028	0.0121	0.0174
60	0.0157	0.0028	0.0120	0.0175	0.0156	0.0028	0.0120	0.0174
63	0.0156	0.0028	0.0120	0.0173	0.0156	0.0028	0.0120	0.0174
102	0.0156	0.0028	0.0121	0.0174	0.0156	0.0029	0.0120	0.0174
104	0.0156	0.0028	0.0120	0.0174	0.0156	0.0028	0.0120	0.0174
107	0.0156	0.0028	0.0120	0.0174	0.0156	0.0028	0.0120	0.0173
177	0.0157	0.0029	0.0120	0.0175	0.0157	0.0028	0.0120	0.0175
213	0.0156	0.0028	0.0120	0.0174	0.0157	0.0029	0.0120	0.0175
221	0.0156	0.0028	0.0121	0.0173	0.0156	0.0028	0.0120	0.0174
229	0.0157	0.0029	0.0121	0.0174	0.0157	0.0028	0.0121	0.0175
232	0.0157	0.0028	0.0120	0.0174	0.0156	0.0028	0.0119	0.0173
238	0.0156	0.0028	0.0120	0.0174	0.0157	0.0028	0.0121	0.0174
248	0.0157	0.0028	0.0120	0.0174	0.0156	0.0028	0.0120	0.0174
249	0.0156	0.0028	0.0120	0.0174	0.0158	0.0029	0.0121	0.0176
253	0.0156	0.0028	0.0120	0.0174	0.0157	0.0028	0.0121	0.0174

Table 6: Overlapping pathways between top 50 pathways selected by Bayesian approach with the polynomial and the Gaussian kernels and ranked by τ^{-1}, $\widehat{(\tau^{-1})}$ =the estimation of τ^{-1}, $\widehat{\sigma(\tau^{-1})}$ =the standard deviation of τ^{-1}, $C.I.$=the Bayesian credible interval of τ^{-1}.

		Polynomial Kernel				Gaussian Kernel			
		Normal	Normal	Diabetes	Diabetes	Normal	Normal	Diabetes	Diabetes
Prediction	Pathway	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Min	*27*	0.2739	0.2784	0.2992	0.2947	0.2477	0.2428	0.3024	0.3053
*Q*1		0.4093	0.4078	0.4676	0.4694	0.4028	0.4010	0.4647	0.4701
*Q*2		0.4702	0.4705	0.5284	0.5275	0.4700	0.4692	0.5291	0.5319
*Q*3		0.5320	0.5329	0.5858	0.5840	0.5347	0.5342	0.5928	0.5950
Max		0.6737	0.6755	0.7383	0.7279	0.7035	0.7047	0.7467	0.7445
Min	*32*	0.2549	0.2693	0.3054	0.3069	0.2490	0.2517	0.2929	0.3103
*Q*1		0.4098	0.4133	0.4669	0.4695	0.4068	0.4072	0.4617	0.4621
*Q*2		0.4678	0.4673	0.5309	0.5328	0.4693	0.4677	0.5249	0.5233
*Q*3		0.5310	0.5303	0.5918	0.5912	0.5333	0.5335	0.5916	0.5882
Max		0.6891	0.7004	0.7425	0.7411	0.6976	0.6987	0.7465	0.7485
Min	*35*	0.2639	0.2653	0.3305	0.3400	0.2502	0.2482	0.3006	0.3020
*Q*1		0.4091	0.4118	0.4694	0.4709	0.4040	0.4009	0.4653	0.4677
*Q*2		0.4706	0.4733	0.5279	0.5281	0.4626	0.4614	0.5259	0.5247
*Q*3		0.5337	0.5310	0.5881	0.5865	0.5281	0.5313	0.5908	0.5947
Max		0.6901	0.6838	0.7388	0.7369	0.7040	0.7025	0.7527	0.7403
Min	*36*	0.2658	0.2755	0.3205	0.3199	0.2531	0.2657	0.2944	0.2991
*Q*1		0.4074	0.4085	0.4722	0.4735	0.4018	0.3993	0.4653	0.4658
*Q*2		0.4701	0.4719	0.5301	0.5324	0.4709	0.4734	0.5306	0.5318
*Q*3		0.5303	0.5292	0.5888	0.5892	0.5345	0.5345	0.5924	0.5925
Max		0.6942	0.6943	0.7248	0.7291	0.7215	0.7153	0.7495	0.7444
Min	*39*	0.2636	0.2589	0.3137	0.3091	0.2580	0.2707	0.2955	0.3021
*Q*1		0.4099	0.4099	0.4704	0.4703	0.4055	0.4049	0.4640	0.4637
*Q*2		0.4672	0.4676	0.5278	0.5291	0.4712	0.4732	0.5265	0.5251
*Q*3		0.5272	0.5300	0.5860	0.5844	0.5412	0.5468	0.5934	0.5957
Max		0.7049	0.7154	0.7550	0.7591	0.7070	0.7002	0.7455	0.7352
Min	*43*	0.2803	0.2865	0.3010	0.3227	0.2564	0.2654	0.3066	0.3081
*Q*1		0.4177	0.4168	0.4708	0.4726	0.4068	0.4079	0.4635	0.4657
*Q*2		0.4746	0.4760	0.5244	0.5254	0.4736	0.4706	0.5270	0.5293
*Q*3		0.5283	0.5289	0.5828	0.5804	0.5363	0.5333	0.5921	0.5922
Max		0.6845	0.6900	0.7217	0.7220	0.7151	0.7087	0.7599	0.7590
Min	*44*	0.2586	0.2589	0.3290	0.3385	0.2555	0.2512	0.3034	0.3000
*Q*1		0.4114	0.4126	0.4695	0.4661	0.4070	0.4087	0.4684	0.4686
*Q*2		0.4708	0.4696	0.5312	0.5301	0.4673	0.4700	0.5313	0.5326
*Q*3		0.5294	0.5298	0.5883	0.5862	0.5371	0.5421	0.5943	0.5905
Max		0.6936	0.6922	0.7237	0.7264	0.7008	0.7095	0.7536	0.7473
Min	*50*	0.2660	0.2699	0.3016	0.2955	0.2516	0.2491	0.2932	0.2875
*Q*1		0.4121	0.4154	0.4674	0.4692	0.4070	0.4075	0.4655	0.4626
*Q*2		0.4711	0.4725	0.5273	0.5283	0.4707	0.4725	0.5291	0.5292
*Q*3		0.5329	0.5322	0.5847	0.5859	0.5355	0.5373	0.5892	0.5891
Max		0.6944	0.7040	0.7404	0.7372	0.6920	0.6908	0.7453	0.7318
Min	*56*	0.2725	0.2710	0.3028	0.3012	0.2581	0.2567	0.2988	0.2953
*Q*1		0.4072	0.4064	0.4685	0.4695	0.4061	0.4046	0.4651	0.4677
*Q*2		0.4677	0.4672	0.5321	0.5320	0.4681	0.4712	0.5287	0.5277
*Q*3		0.5292	0.5318	0.5937	0.5932	0.5351	0.5383	0.5917	0.5920
Max		0.7103	0.7017	0.7541	0.7540	0.6967	0.6905	0.7496	0.7359
Min	*60*	0.2722	0.2613	0.3123	0.3235	0.2449	0.2497	0.3037	0.3063
*Q*1		0.4134	0.4134	0.4711	0.4703	0.4022	0.4058	0.4650	0.4642
*Q*2		0.4727	0.4720	0.5307	0.5291	0.4709	0.4768	0.5284	0.5264
*Q*3		0.5334	0.5348	0.5886	0.5874	0.5324	0.5333	0.5891	0.5846
Max		0.6974	0.6940	0.7341	0.7298	0.6966	0.7030	0.7509	0.7475

Table 7: Prediction probability of overlapping pathways between top 50 pathways selected by Bayesian approach with the polynomial and the Gaussian kernels.

			Polynomial	Kernel			Gaussian	Kernel	
		Normal	Normal	Diabetes	Diabetes	Normal	Normal	Diabetes	Diabetes
Prediction	Pathway	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Min	*63*	0.2543	0.2620	0.3021	0.3153	0.2459	0.2456	0.2924	0.2988
$Q1$		0.4145	0.4138	0.4686	0.4696	0.4070	0.4091	0.4622	0.4634
$Q2$		0.4770	0.4772	0.5294	0.5317	0.4702	0.4685	0.5284	0.5242
$Q3$		0.5339	0.5292	0.5901	0.5865	0.5310	0.5367	0.5909	0.5861
Max		0.6798	0.6806	0.7370	0.7407	0.6818	0.6765	0.7494	0.7516
Min	*102*	0.2457	0.2529	0.3221	0.3148	0.2404	0.2441	0.2978	0.3001
$Q1$		0.4131	0.4125	0.4688	0.4666	0.4057	0.4042	0.4652	0.4677
$Q2$		0.4730	0.4688	0.5294	0.5325	0.4713	0.4708	0.5277	0.5276
$Q3$		0.5352	0.5390	0.5878	0.5902	0.5358	0.5357	0.5908	0.5912
Max		0.6987	0.6923	0.7350	0.7304	0.7003	0.6866	0.7512	0.7612
Min	*104*	0.2363	0.2339	0.3018	0.2983	0.2511	0.2560	0.3118	0.3206
$Q1$		0.4054	0.4044	0.4603	0.4568	0.4050	0.4053	0.4667	0.4681
$Q2$		0.4744	0.4732	0.5288	0.5302	0.4734	0.4709	0.5276	0.5297
$Q3$		0.5400	0.5366	0.5951	0.5964	0.5326	0.5332	0.5839	0.5814
Max		0.7270	0.7073	0.7810	0.7741	0.7107	0.7012	0.7466	0.7461
Min	*107*	0.2778	0.2830	0.3212	0.3160	0.2587	0.2640	0.2900	0.2930
$Q1$		0.4122	0.4147	0.4715	0.4730	0.4046	0.4074	0.4704	0.4694
$Q2$		0.4686	0.4669	0.5312	0.5330	0.4668	0.4658	0.5317	0.5315
$Q3$		0.5285	0.5281	0.5814	0.5802	0.5336	0.5356	0.5918	0.5912
Max		0.6789	0.6767	0.7231	0.7184	0.7031	0.6991	0.7574	0.7437
Min	*177*	0.2574	0.2565	0.3227	0.3215	0.2419	0.2359	0.2962	0.2983
$Q1$		0.4120	0.4122	0.4758	0.4769	0.4019	0.4011	0.4642	0.4621
$Q2$		0.4694	0.4705	0.5293	0.5304	0.4695	0.4683	0.5262	0.5262
$Q3$		0.5275	0.5279	0.5863	0.5878	0.5325	0.5339	0.5878	0.5896
Max		0.6838	0.6883	0.7339	0.7230	0.7158	0.7150	0.7433	0.7373
Min	*213*	0.2546	0.2575	0.2934	0.2923	0.2496	0.2475	0.2995	0.3027
$Q1$		0.4065	0.4076	0.4682	0.4656	0.4071	0.4067	0.4640	0.4625
$Q2$		0.4734	0.4728	0.5294	0.5305	0.4722	0.4721	0.5302	0.5318
$Q3$		0.5392	0.5381	0.5920	0.5897	0.5346	0.5361	0.5949	0.5954
Max		0.7074	0.7134	0.7452	0.7486	0.6900	0.6807	0.7465	0.7604
Min	*221*	0.2683	0.2668	0.3180	0.3232	0.2405	0.2443	0.3029	0.3089
$Q1$		0.4127	0.4103	0.4734	0.4742	0.4097	0.4101	0.4638	0.4660
$Q2$		0.4697	0.4706	0.5293	0.5304	0.4716	0.4699	0.5275	0.5280
$Q3$		0.5237	0.5267	0.5872	0.5879	0.5329	0.5344	0.5927	0.5916
Max		0.6874	0.6892	0.7261	0.7205	0.6844	0.6764	0.7521	0.7535
Min	*229*	0.2776	0.2799	0.3087	0.3207	0.2420	0.2498	0.3063	0.3100
$Q1$		0.4115	0.4117	0.4712	0.4735	0.4098	0.4099	0.4640	0.4664
$Q2$		0.4691	0.4693	0.5336	0.5326	0.4740	0.4704	0.5250	0.5266
$Q3$		0.5275	0.5307	0.5929	0.5919	0.5415	0.5392	0.5933	0.5881
Max		0.6908	0.6906	0.7156	0.7131	0.7092	0.6996	0.7428	0.7440
Min	*232*	0.2646	0.2583	0.3120	0.3167	0.2455	0.2460	0.2992	0.3054
$Q1$		0.4115	0.4128	0.4691	0.4715	0.4081	0.4076	0.4653	0.4654
$Q2$		0.4645	0.4647	0.5266	0.5270	0.4706	0.4687	0.5271	0.5287
$Q3$		0.5261	0.5286	0.5825	0.5795	0.5359	0.5362	0.5931	0.5896
Max		0.6710	0.6632	0.7303	0.7246	0.7111	0.7056	0.7419	0.7413
Min	*238*	0.2705	0.2723	0.3166	0.3211	0.2583	0.2599	0.2915	0.2935
$Q1$		0.4083	0.4102	0.4716	0.4720	0.4115	0.4117	0.4598	0.4624
$Q2$		0.4719	0.4699	0.5299	0.5323	0.4728	0.4766	0.5265	0.5296
$Q3$		0.5318	0.5342	0.5877	0.5878	0.5368	0.5364	0.5919	0.5927
Max		0.6885	0.6902	0.7264	0.7248	0.7194	0.7173	0.7729	0.7675

			Polynomial	Kernel			Gaussian	Kernel	
		Normal	Normal	Diabetes	Diabetes	Normal	Normal	Diabetes	Diabetes
Prediction	Pathway	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Min	*248*	0.2678	0.2709	0.3110	0.3183	0.2387	0.2342	0.3001	0.3029
$Q1$		0.4137	0.4147	0.4719	0.4736	0.4066	0.4062	0.4679	0.4677
$Q2$		0.4716	0.4719	0.5274	0.5267	0.4693	0.4695	0.5293	0.5301
$Q3$		0.5270	0.5258	0.5826	0.5856	0.5367	0.5372	0.5915	0.5919
Max		0.6723	0.6776	0.7240	0.7214	0.6987	0.6890	0.7553	0.7532
Min	*249*	0.2673	0.2645	0.3199	0.3265	0.2503	0.2508	0.2966	0.3003
$Q1$		0.4106	0.4084	0.4730	0.4744	0.4031	0.4031	0.4652	0.4642
$Q2$		0.4742	0.4768	0.5298	0.5286	0.4705	0.4715	0.5304	0.5294
$Q3$		0.5287	0.5293	0.5870	0.5872	0.5344	0.5338	0.5960	0.5937
Max		0.6844	0.6832	0.7319	0.7349	0.7099	0.6982	0.7471	0.7446
Min	*253*	0.2412	0.2417	0.3039	0.3098	0.2500	0.2541	0.2943	0.3015
$Q1$		0.4059	0.4045	0.4734	0.4747	0.4038	0.4051	0.4693	0.4695
$Q2$		0.4729	0.4689	0.5324	0.5321	0.4732	0.4723	0.5290	0.5283
$Q3$		0.5315	0.5325	0.5905	0.5884	0.5366	0.5364	0.5918	0.5903
Max		0.6971	0.6898	0.7310	0.7309	0.7042	0.7112	0.7471	0.7440

	Polynomial	and	Point-mass			Gaussian	and	Point-mass	
Pathway	$\pi = 0.1$	0.3	0.5	0.7	0.9	$\pi = 0.1$	0.3	0.5	0.7	0.9
27	-89.6097	-84.3250	-93.6767	-100.8029	-27.4078	-92.2183	-83.9393	-98.1210	-101.3295	-36.9661
32	-74.5988	-91.1763	-89.7667	-94.2849	-45.4785	-82.3963	-85.7547	-93.7509	-101.2650	-40.1362
35	-82.1249	-79.4311	-91.2623	-89.6418	-57.7610	-75.0304	-82.6301	-93.3203	-96.7034	-51.7525
36	-76.7462	-92.6822	-91.2217	-96.7731	-29.3416	-85.9767	-82.1339	-90.6975	-101.9851	-29.4243
39	-78.9130	-86.8862	-96.6282	-87.1791	-36.0094	-68.1786	-85.5657	-91.6920	-104.4558	-37.1595
43	-77.6200	-75.4491	-95.2109	-87.3411	-42.6807	-92.6393	-91.3018	-83.4547	-102.4193	-44.6771
44	-79.2289	-78.6459	-99.0842	-105.0286	-36.8892	-75.2051	-85.0127	-94.5287	-92.7939	-43.0213
50	-77.0942	-91.6952	-106.4643	-93.4494	-24.2755	-74.6403	-83.8038	-104.2355	-96.9197	-29.8003
56	-73.0555	-86.2489	-95.9689	-103.5378	-33.4283	-75.6710	-81.0434	-104.0168	-102.5679	-33.6712
60	-76.8654	-84.4777	-86.9530	-96.6614	-35.5391	-80.7878	-92.6847	-98.7081	-101.9840	-30.2318
63	-78.5407	-82.2566	-94.9332	-102.3404	-37.4382	-85.8454	-86.7251	-89.2277	-98.9844	-56.9666
102	-73.9452	-77.4506	-94.1905	-100.3804	-16.6375	-70.0733	-90.2126	-101.1307	-101.6901	-24.0634
104	-71.0426	-85.9550	-84.1409	-101.0087	-37.4511	-72.1228	-85.0801	-99.8887	-95.0613	-45.4940
107	-81.3486	-87.9075	-95.7056	-99.9949	-50.3171	-92.3075	-77.3575	-98.2703	-96.4968	-32.6405
177	-83.1158	-98.0737	-89.1788	-100.2707	-41.4139	-77.7612	-85.8690	-92.3877	-101.9098	-37.0035
213	-81.4581	-88.4611	-104.8559	-82.7978	-46.4415	-85.0286	-87.4152	-86.7290	-106.0232	-46.9100
221	-75.6350	-82.1727	-94.2206	-112.5444	-53.3895	-81.3517	-78.9361	-101.0304	-109.0720	-49.8237
229	-79.4675	-81.9152	-101.5734	-107.7748	-38.1227	-76.3921	-80.6297	-90.5961	-101.4590	-28.5732
232	-77.1862	-87.7314	-91.2930	-107.0081	-74.2807	-82.7225	-97.2582	-95.3039	-100.5103	-36.5818
238	-72.2161	-81.3961	-94.9845	-86.6786	-36.0718	-71.3444	-92.1969	-97.6423	-99.1545	-46.4910
248	-75.1838	-85.1340	-96.0070	-98.6222	-37.4782	-79.4949	-78.2865	-102.4643	-106.0986	-31.9894
249	-80.2707	-77.6329	-99.2291	-100.2209	-47.7806	-77.4326	-93.4519	-101.1277	-108.7311	-56.7668
253	-82.3137	-84.6110	-96.3320	-102.3442	-29.9983	-74.6661	-89.4633	-89.2047	-92.9129	-39.9102

Table 8: Log-likelihood value based on the mixture of selected kernels and point-mass density, π=the mixture proportion.

	Polynomial and Constant					Gaussian and Constant				
Pathway	$\pi = 0.1$	0.3	0.5	0.7	0.9	$\pi = 0.1$	0.3	0.5	0.7	0.9
27	-77.3526	-76.4807	-67.6022	-67.2045	-41.0632	-77.3813	-78.6254	-75.0449	-61.8915	-35.7982
32	-75.8610	-77.6761	-73.0162	-63.8918	-28.4798	-76.3342	-75.4908	-75.4959	-60.0886	-31.7715
35	-76.2100	-78.2027	-73.8664	-66.1623	-35.9609	-81.6746	-77.9857	-75.0873	-65.9640	-26.8334
36	-78.4893	-78.9714	-77.5009	-74.7068	-28.2839	-77.3073	-78.0263	-76.3573	-72.4890	-28.2570
39	-77.2743	-75.9493	-76.9360	-62.4703	-35.1099	-78.7436	-76.8563	-76.3170	-71.7533	-23.3679
43	-77.6872	-78.9708	-74.3449	-70.4164	-13.6850	-78.5912	-80.4497	-76.1053	-59.6081	-24.5087
44	-78.2807	-80.7319	-84.4580	-72.3985	-15.5506	-78.7976	-79.5194	-78.7301	-70.7466	-18.1691
50	-77.9895	-77.1855	-78.2188	-70.6400	-18.1024	-78.4341	-75.9071	-75.9879	-69.1629	-26.9938
56	-77.9876	-74.4020	-71.0717	-69.4398	-30.9465	-78.2774	-75.7292	-73.6655	-69.6070	-36.1730
60	-77.9563	-74.3952	-78.3353	-71.9692	-30.2357	-78.4729	-76.5551	-75.5927	-68.5790	-20.2411
63	-78.0331	-76.4274	-73.0871	-62.5960	-31.0437	-77.1595	-76.0394	-79.0706	-69.2809	-34.6454
102	-77.0453	-77.6759	-72.1555	-70.9117	-26.4245	-78.4785	-75.4058	-68.1873	-66.5453	-25.2462
104	-76.6318	-77.2742	-73.1585	-64.3935	-31.8190	-76.8476	-76.0488	-77.7017	-63.2991	-36.2175
107	-81.2087	-81.5950	-82.1734	-69.5193	-29.3148	-78.1376	-78.4586	-76.9339	-77.3053	-18.4683
177	-79.4237	-81.1129	-82.9973	-68.5479	-31.4826	-79.5161	-82.2463	-85.9490	-82.0351	-30.5218
213	-78.0967	-76.8449	-74.1528	-70.1644	-41.3122	-78.4815	-75.4670	-74.0991	-68.2019	-37.7294
221	-79.8148	-86.7610	-91.7771	-91.6506	-35.2403	-79.4049	-87.2871	-89.5863	-91.6451	-32.1177
229	-77.8063	-76.5115	-73.6989	-66.9360	-24.0074	-75.7090	-78.5973	-76.4080	-64.6714	-19.0580
232	-78.0321	-82.9603	-90.7997	-73.8113	-20.9800	-80.8000	-82.0010	-83.5842	-84.3040	-27.5046
238	-76.9709	-75.1968	-79.9104	-66.4742	-40.5238	-76.3860	-80.5384	-78.6250	-67.6159	-28.4909
248	-78.1180	-81.4038	-84.1633	-68.4664	-35.6159	-77.6021	-80.0115	-77.7066	-70.4652	-29.9161
249	-78.4498	-78.6412	-82.6710	-78.0350	-17.1158	-78.2846	-76.3153	-80.6119	-85.7281	-14.1024
253	-79.1677	-78.0527	-73.6698	-70.8110	-32.3504	-77.2881	-75.2101	-73.4178	-66.7466	-33.1872

Table 9: Log-likelihood value based on the mixture of selected kernels and constant covariance structure, π=the mixture proportion.

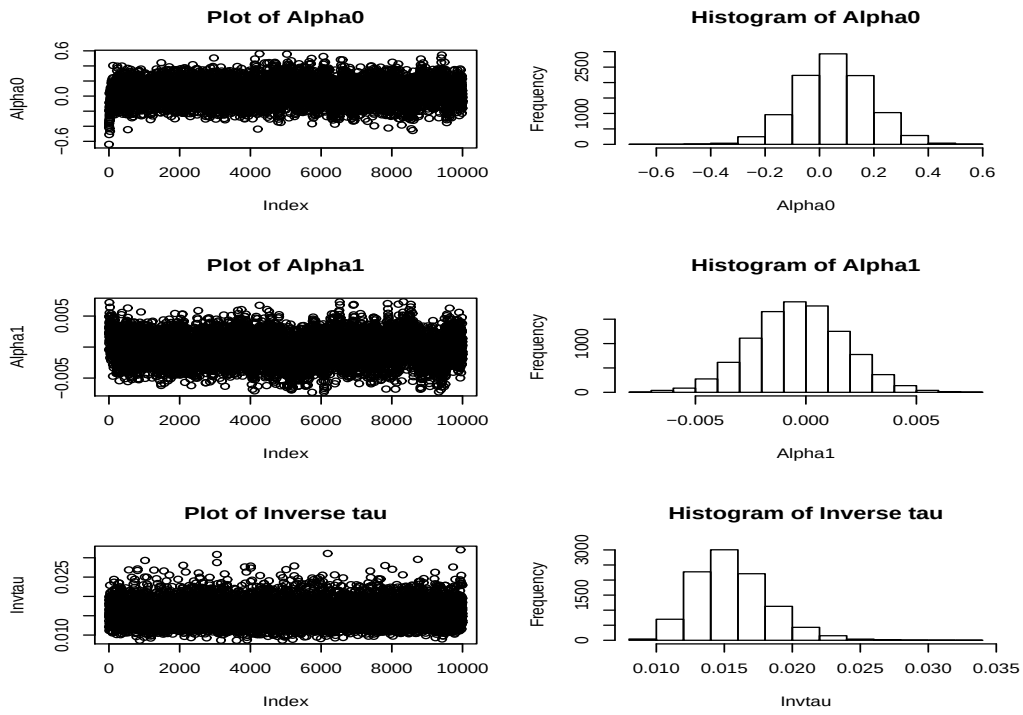

Figure 1: MCMC trace plots and histograms based on the Gaussian kernel using pathway 229

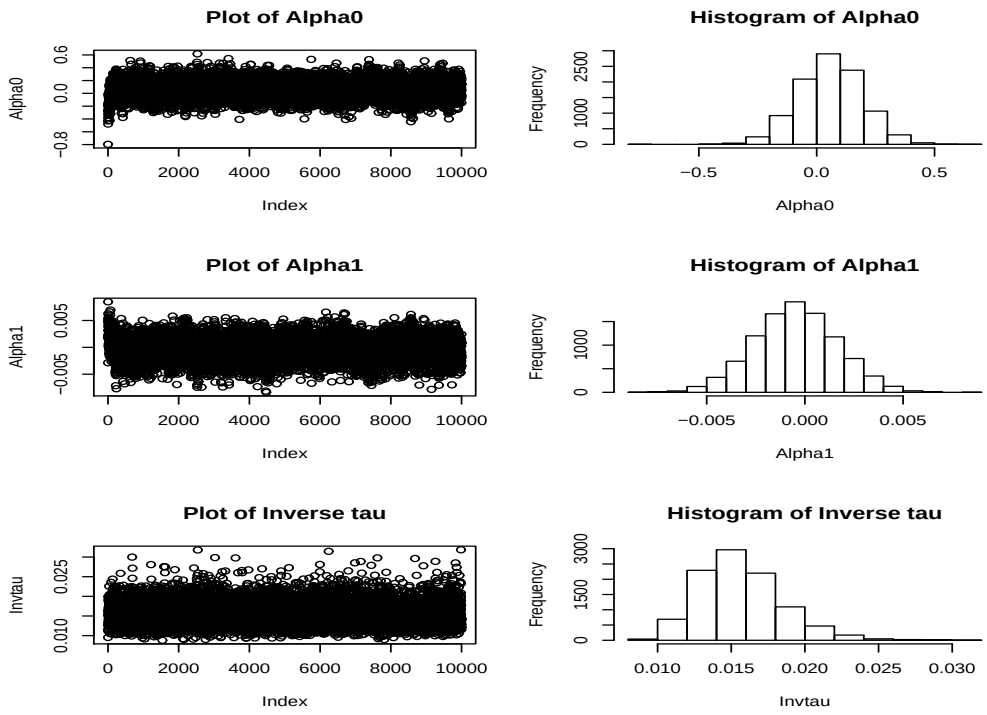

Figure 2: MCMC trace plots and histograms based on the Gaussian kernel using pathway 36

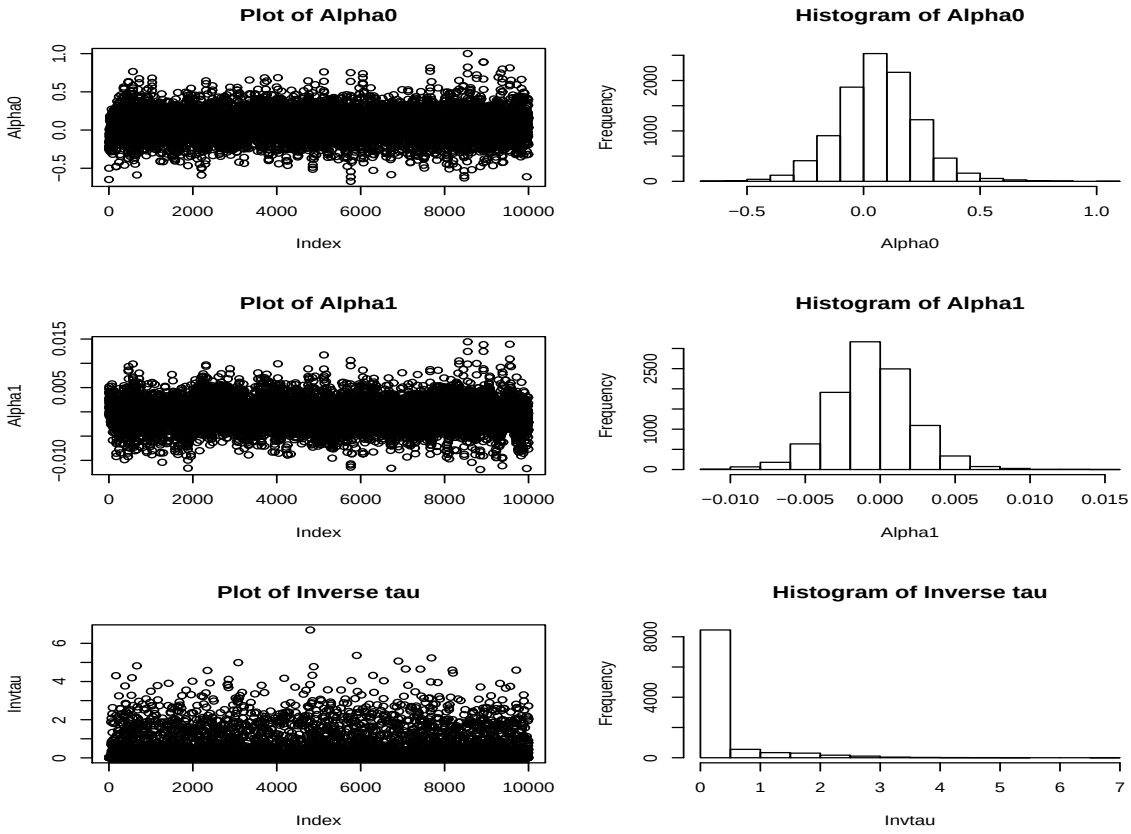

Figure 3: MCMC trace plots and histograms based on the mixture of the Gaussian kernel and constant covariance matrix using pathway 229

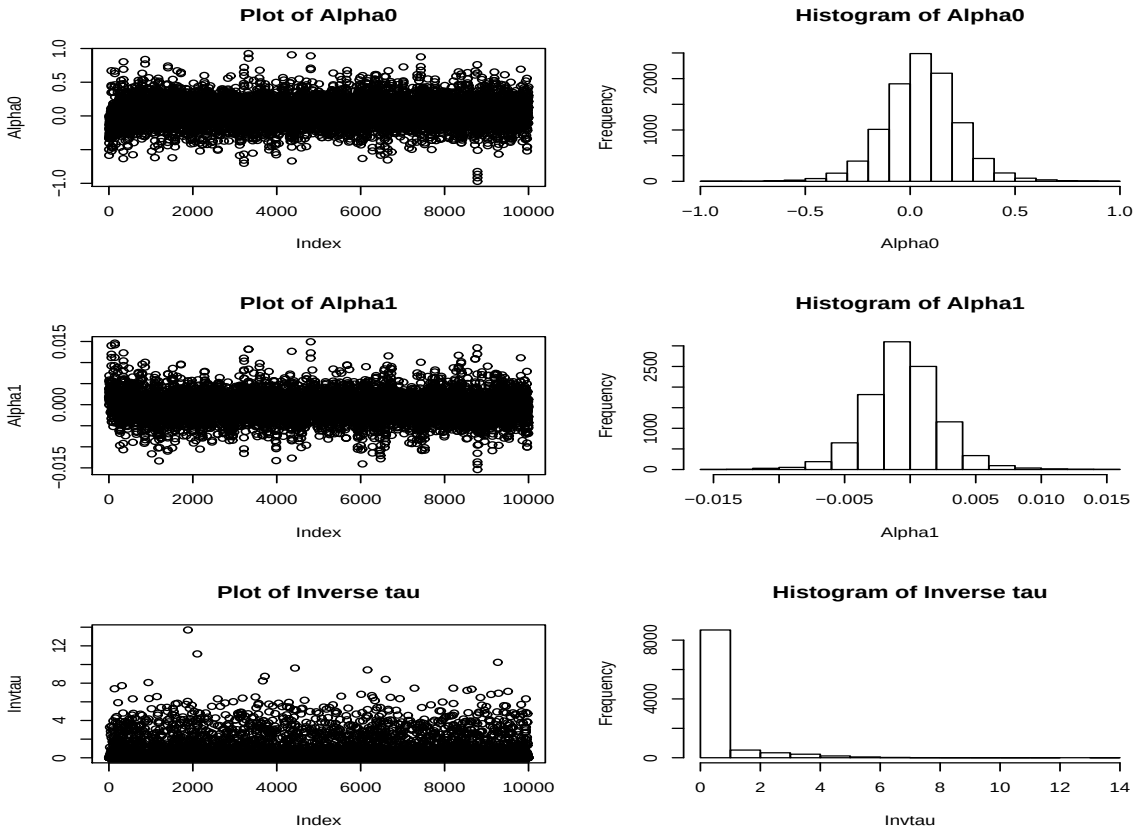

Figure 4: MCMC trace plots and histograms based on the mixture of the Gaussian kernel and constant covariance matrix using pathway 36

References

Albert, J. and Chib, S. (1993) Bayesian analysis of binary and polychotomous response data. *Journal of American Statistical Association*, 88, 669-679.

Friedman, J.H. (1991) Multivariate adaptive regression splines (Disc: P67-141). *The Annals of Statistics*, 19, 1-67.

Goeman, J., van de Geer, S.A., de Kort, F., van Houwelingen, H.C. (2004) A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics*, 20, 93-99.

Hosack, D.A., Dennis, G., Jr., Sherman, B.T., Lane, H.C., Lempicki, R.A. (2003) Identifying biological themes within lists of genes with ease. *Genome Biology*, 4, P4.

Kim, I., Pang, H., and Zhao, H. (2009). Semiparametric methods for evaluating pathway effects on clinical outcomes using gene expression data. (Submitted)

Misu, H., Takamura, T., Matsuzawa, N., Shimizu, A., Ota, T., Sakurai, M., Ando, H., Arai, K., Yamashita, T., Honda, M., Yamashita, T., and Kaneko, S. (2007). Genes involved in oxidative phosphorylation are coordinately upregulated with fasting hyperglycaemia in livers of patients with type 2 diabetes. *Diabetologia*, 50, 268-277.

Mootha, V.K., et al. (2003). Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proceedings of the National Academy of Sciences*, 100, 605-610.

Mootha, V. K., Handschin, C., Arlow, D., Xie, X., Pierre, J. S., Sihag, S., Yang, W., Altshuler, D., Puigserver, P., Patterson, N., Willy, P. J., Schulman, I. G., Heyman, R. A., Lander, E. S., and Spiegelman, B. M. (2004). Errα and Gabpa/b specify PGC-1α-dependent oxidative phosphorylation gene expression that is altered in diabetic muscle. *Proceedings of the National Academy of Sciences*, 101, 6570-6575.

Pang H, Lin A, Holford M, Enerson BE, Lu B, Lawton MP, Floyd E, Zhao H. (2006) Pathway analysis using random forests classification and regression. *Bioinformatics*, 22, 2028-2036.

Rajagopalan, D.A. and Agarwal, P. (2005) Inferring pathways from gene lists using a literature-derived network of biological relationships. *Bioinformatics*, 21, 788-793.