On the gene ranking of replicated microarray time course data

Yu Chuan Tai[1] and Terence P. Speed[2,3]

[1] Institute for Human Genetics, University of California, San Francisco.

[2] Department of Statistics, University of California, Berkeley.

[3] Walter and Eliza Hall Institute of Medical Research, Australia.

August 9, 2007

Abstract

Consider the gene ranking problem of replicated microarray time course experiments where there are multiple biological conditions, and genes of interest are those whose temporal profiles are different across conditions. We derive the multi-sample multivariate empirical Bayes statistic for ranking genes in the order of differential expression, from both longitudinal and cross-sectional replicated developmental microarray time course data. Our longitudinal multi-sample model assumes that time course replicates are *i.i.d.* multivariate normal vectors. On the other hand, we construct our cross-sectional model using a normal regression framework with any appropriate basis for the design matrices. In both cases, we use natural conjugate priors in our empirical Bayes setting which guarantee closed form solutions

for the posterior odds. Our simulations and two case studies using published worm and mouse microarray time course datasets indicate that the proposed approaches work well.

keywords: longitudinal; cross-sectional; microarray time course; gene ranking; empirical Bayes.

1   Introduction

Many important biological events occur in a temporal fashion. Microarray time course experiments permit the monitoring of temporal profiles for thousands of genes simultaneously, hence are useful tools for investigating the biological processes of interest dynamically. Two major categories of time course experiments are those involving periodic and developmental phenomena, respectively. Periodic time courses typically involve natural biological processes (*e.g.* cell cycle (Cho et al., 1998; Spellman et al., 1998; Chu et al., 1998), circadian rhythms (Storch et al., 2002)) whose temporal profiles follow regular patterns, while in developmental time courses the temporal patterns are more arbitrary. The latter can be further divided into two subclasses: longitudinal and cross-sectional experiments. In longitudinal time course experiments, mRNA samples are extracted from the *same* experimental units, while in cross-sectional ones they are extracted from *different* units. The analysis of developmental time course experiments has been a great challenge, mainly due to the fact that the time series are very short (on average 3-12 time points), very few replicates (1-5 replications) per gene, and thousands of genes in an experiment. The time series are so short such that analysis techniques for standard time series data with hundreds of time points (*e.g.* Fourier transform and wavelets) can not be applied. Given very few replicates per gene, the replicate variances may be poorly estimated. For experiments with longitudinal design, the fact that gene expression measurements are correlated over time further complicates the analysis. We refer the reader to Tai and Speed (2005) for a detailed review on the background and analysis challenges of developmental microarray time course experiments.

Following a microarray experiment, typically only a limited number of genes can be further verified and studied. Thus, one immediate concern is how to identify genes of interest from such a large gene set. In Tai and Speed (2006), we constructed the multivariate empirical Bayes statistic (*MB*-statistic) to rank genes in order of interest from

longitudinal replicated developmental microarray time course experiments when there are one or two biological conditions. The former leads to what we called the one-sample problem, where genes of interest are those which change over time, perhaps in some specific pattern. The latter refers to the two-sample problem, where genes of interest are those whose temporal profiles are different between two biological conditions. We built our hierarchical model by assuming the independent time course vectors of longitudinal gene expression levels are *i.i.d.* from multivariate normal distribution, with gene-specific means and variance-covariance matrices. Natural conjugate priors were assigned to the gene-specific mean and variance-covariance matrix to ensure the existence of closed-form formula for the posterior odds against the null hypothesis that the gene stays constant (one-sample problem), or the temporal profiles are the same across two biological conditions (two-sample problem). These posterior odds were shown to be equivalent to their corresponding $\widetilde{T}^2$ statistics, variants of Hotelling T^2 statistics, when the sample size(s) are identical across genes. The proposed statistics were illustrated in a simulation study and on an *Arabidopsis* microarray time course dataset in Wildermuth et al. (2007).

In this paper, we extend our longitudinal model in Tai and Speed (2006) to the multi-sample problem, where genes of interest are those whose temporal profiles are different across two or more biological conditions. The corresponding models for cross-sectional design with any number of biological conditions are also developed. As noted in Tai and Speed (2005, 2006), only a few methods have been proposed to deal with this gene selection problem in the time course context, *e.g.* the B-spline approaches proposed in Bar-Joseph et al. (2003); Storey et al. (2005); Hong and Li (2006), and the Hidden Markov Model method in Yuan and Kendziorski (2006). We refer the reader to Tai and Speed (2005, 2006) for a review of these approaches.

This paper is organized as follows. In section 2, we present our multivariate empirical Bayes model for longitudinal replicated time course data with multiple biological

conditions. A simulation study is performed to assess our proposed statistic and hyperparameter estimation procedures. Our model for cross-sectional data is described in section 3. In section 4, we illustrate our cross-sectional method on two published microarray time course datasets and compare it to the moderated F statistic (Smyth, 2004) and the functional hierarchical method in Hong and Li (2006) or ANOVA. We summarize and discuss our approaches in section 5.

2 Longitudinal Design

We first introduce our notation for longitudinal experiments. Suppose that our mRNA samples represent D independent biological conditions. Suppose further that each gene g has n_{gd} $k \times 1$ independent time course biological replicates $\mathbf{X}_{gdi}$, $d = 1, ..., D$, $i = 1, ..., n_d$, where k is the number of time points. Our basic model for this multi-sample problem starts with the assumption that $\mathbf{X}_{gdi}$ are $i.i.d.$ multivariate normal with condition-specific mean $\boldsymbol{\mu}_{gd}$ and a common variance-covariance matrix $\boldsymbol{\Sigma}_g$, denoted by $N(\boldsymbol{\mu}_{gd}, \boldsymbol{\Sigma}_g)$. For simplicity, the subscript g will be dropped for the rest of this paper. The statistical models presented in the rest of this paper are for an arbitrary single gene g. As in Tai and Speed (2006), we denote our (composite) null and alternative hypotheses by H and K, respectively, where $H : \boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_D = \boldsymbol{\mu}, \boldsymbol{\Sigma} > 0$ and $K : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$, for some $i \neq j$, $\boldsymbol{\Sigma} > 0$.

2.1 The Moderated Wilks' lambda Statistic

A direct approach for testing the null H is the classical one-way multivariate analysis of variance (MANOVA). We now summarize the results in section 12.3 from Mardia et al.

(2000), first defining the within- and between- sums of squares and products

$$\mathbf{W} = \sum_{d=1}^{D} \sum_{i=1}^{n_d} (\mathbf{X}_{di} - \overline{\mathbf{X}}_d)(\mathbf{X}_{di} - \overline{\mathbf{X}}_d)',$$

$$\mathbf{B} = \sum_{d=1}^{D} n_d (\overline{\mathbf{X}}_d - \overline{\mathbf{X}})(\overline{\mathbf{X}}_d - \overline{\mathbf{X}})',$$

where $\overline{\mathbf{X}}_d = n_d^{-1} \sum_{i=1}^{n_d} \mathbf{X}_{di}$, $\overline{\mathbf{X}} = n^{-1} \sum_{d=1}^{D} \sum_{i=1}^{n_d} \mathbf{X}_{di}$, $n = \sum_{d=1}^{D} n_d$ are the condition-specific and overall average time course vectors, and the total sample size, respectively. The total sums of squares and products $\mathbf{T}$ equals to $\mathbf{W} + \mathbf{B}$. By Mardia et al. (2000), the likelihood-based Wilks' lambda is the ratio of two determinants

$$\mathbf{LR} = \frac{|\mathbf{W}|}{|\mathbf{T}|}.$$

A moderated Wilks' lambda can be defined by

$$\widetilde{\mathbf{LR}} = \frac{|\widetilde{\mathbf{W}}|}{|\widetilde{\mathbf{T}}|}, \tag{1}$$

where $\widetilde{\mathbf{T}} = \mathbf{T} + \nu \mathbf{\Lambda}$, $\widetilde{\mathbf{W}} = \mathbf{W} + \nu \mathbf{\Lambda}$, ν is a suitable positive number and $\mathbf{\Lambda}$ is a suitable positive definite matrix. This idea will be developed in what follows.

2.2   The Multivariate Empirical Bayes Model

2.2.1   Models and Priors

As in Tai and Speed (2006), we define a Bernoulli random variable I with success probability p to indicate the status of any single gene

$$I = \begin{cases} 1 & \text{if } K \text{ is true} \\ 0 & \text{if } H \text{ is true.} \end{cases}$$

The model for the data can then be written as follows:

$$\begin{cases} \mathbf{X}_{di} | \boldsymbol{\mu}_d, \boldsymbol{\Sigma}, I = 1 \sim N(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}) & d = 1, ..., D, \quad i = 1, ..., n_d; \\ \mathbf{X}_{di} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, I = 0 \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \end{cases} \tag{2}$$

We define the prior for the variance-covariance matrix $\boldsymbol{\Sigma}$ to be inverse-Wishart with degrees of freedom ν and scale matrix $\nu\boldsymbol{\Lambda}$ (Gelman et al., 2000):

$$\boldsymbol{\Sigma} \sim Inv\text{-}Wishart_\nu \left((\nu\boldsymbol{\Lambda})^{-1} \right). \tag{3}$$

In the case that the expected temporal patterns are different across biological conditions, the priors for $\boldsymbol{\mu}_d, d = 1, ..., D$ given $\boldsymbol{\Sigma}$ and I are assumed to be independent multivariate normal with condition-specific mean $\boldsymbol{\alpha}_d$ and scale parameter β_d, while when $I = 0$ ($\boldsymbol{\mu}_1 = ... = \boldsymbol{\mu}_D = \boldsymbol{\mu}$), all the $\boldsymbol{\mu}_d$ have a common prior:

$$\begin{cases} \boldsymbol{\mu}_d | \boldsymbol{\Sigma}, I = 1 \sim N \left(\boldsymbol{\alpha}_d, \beta_d^{-1}\boldsymbol{\Sigma} \right), & d = 1, ..., D; \\ \boldsymbol{\mu} | \boldsymbol{\Sigma}, I = 0 \sim N \left(\boldsymbol{\alpha}, \beta^{-1}\boldsymbol{\Sigma} \right), \end{cases} \tag{4}$$

where $\beta_d > 0$ and $\beta > 0$ are hyperparameters. For two-channel comparative microarray experiments where relative gene expression levels are measured, it is reasonable to assume $\boldsymbol{\alpha} = \boldsymbol{\alpha}_d = \mathbf{0}$, while in the single channel case these hyperparameters need to be estimated.

2.2.2 Posterior Odds

The posterior odds against the null that the expected time course profiles are the same can be easily derived using the above priors and models, and are

$$\mathbf{O} = \frac{p}{1-p} \frac{P(data|I=1)}{P(data|I=0)}$$

$$= \frac{p}{1-p} \left(\frac{n+\beta}{\beta} \right)^{\frac{k}{2}} \prod_{d=1}^{D} \left(\frac{\beta_d}{n_d + \beta_d} \right)^{\frac{k}{2}} \left(\frac{|\mathbf{T} + \mathbf{M}_H + \nu\boldsymbol{\Lambda}|}{|\mathbf{W} + \mathbf{M}_K + \nu\boldsymbol{\Lambda}|} \right)^{\frac{1}{2}(n+\nu)}, \tag{5}$$

where $\mathbf{M}_K = \sum_{d=1}^{D} (n_d^{-1} + \beta_d^{-1})^{-1} (\overline{\mathbf{X}}_d - \boldsymbol{\alpha}_d)(\overline{\mathbf{X}}_d - \boldsymbol{\alpha}_d)'$, and $\mathbf{M}_H = \left(n^{-1} + \beta^{-1} \right)^{-1} (\overline{\mathbf{X}} - \boldsymbol{\alpha})(\overline{\mathbf{X}} - \boldsymbol{\alpha})'$. When all genes have the same number of replicates n_d for the d-th biological condition, the posterior odds are equivalent to the last term in equation (5)

$$\frac{|\mathbf{T} + \mathbf{M}_H + \nu\boldsymbol{\Lambda}|}{|\mathbf{W} + \mathbf{M}_K + \nu\boldsymbol{\Lambda}|} = \frac{\left| \widetilde{\mathbf{T}} + \mathbf{M}_H \right|}{\left| \widetilde{\mathbf{W}} + \mathbf{M}_K \right|}. \tag{6}$$

Equation (6) is our EB analogue of Wilks' likelihood-based lambda from MANOVA with both its denominator and numerator moderated by (condition-specific) matrices involving prior means and variance-covariance matrices. When n_d is identical across genes, one can just use equation (6) for ranking instead of equation (5) since they give the same results.

Several special and limiting cases of the posterior odds are presented in Web Appendix A. We describe how we estimate the hyperparameters associated with the posterior odds in Web Appendix B.

2.3 Simulation Study

2.3.1 Method

In this section we report a simulation study to compare our fully moderated Wilks' lambda involving $\mathbf{M}_H$ and $\mathbf{M}_K$, the moderated F-statistic (Smyth, 2004), and the likelihood-based moderated Wilks' lambda without $\mathbf{M}_H$ and $\mathbf{M}_K$ to see if it helps to include these two extra terms involving the β_d and β. We further assess if hyperparameter estimation procedures we propose are satisfactory. We do this by plugging in their true values into the formula and comparing the number of false positives and false negatives with those obtained from our fully moderated Wilks' lambda with all hyperparameters estimated and likelihood-based moderated Wilks' lambda. We simulate 100 datasets, each with 10,000 genes. Again, genes are simulated independently as it makes sense for comparison purposes. We assign 200 genes ($p = 0.02$) to have different temporal profiles across biological conditions. The simulation has 3 biological conditions ($D = 3$), 2 replicates ($n_1 = n_2 = n_3 = 2$) within conditions, and 8 time points ($k = 8$). The hyperparameters are assigned based on a real dataset we have analyzed: $\nu = 13$, $\beta_1 = \beta_2 = \beta_3 = \beta = 1$, $\boldsymbol{\alpha}'_1 = (5, 5, 5, 5, 5, 5, 5, 5)$, $\boldsymbol{\alpha}'_2 = (6, 6, 6, 6, 6, 6, 6, 6)$, $\boldsymbol{\alpha}'_3 =$

$(7, 7, 7, 7, 7, 7, 7, 7), \boldsymbol{\alpha}' = (6, 6, 6, 6, 6, 6, 6, 6),$ and

$$
\boldsymbol{\Lambda} = \begin{pmatrix}
29.2 & 1 & 0.9 & 0.6 & 0.1 & 0.2 & 0.3 & 0.3 \\
1 & 29.2 & 1 & 1 & 0.4 & 0.5 & 0.7 & 0.6 \\
0.9 & 1 & 17.2 & 1 & 0.5 & 0.4 & 0.5 & 0.6 \\
0.6 & 1 & 1 & 18.3 & 1 & 0.5 & 0.5 & 0.4 \\
0.1 & 0.4 & 0.5 & 1 & 27.3 & 2 & 0.4 & 0.3 \\
0.2 & 0.5 & 0.4 & 0.5 & 2 & 30.3 & 1 & 0.2 \\
0.3 & 0.7 & 0.5 & 0.5 & 0.4 & 1 & 20.2 & 0.8 \\
0.3 & 0.6 & 0.6 & 0.4 & 0.3 & 0.2 & 0.8 & 19.2
\end{pmatrix} \times 10^{-2}.
$$

We compare our MB-statistic (equation 6), the likelihood-ratio moderated Wilks' lambda (equation 1), our MB-statistic (equation 6) with all the estimated hyperparameters replaced by their true values, and the moderated F-statistic under the linear model setting (Smyth, 2004) implemented in the Bioconductor package *limma* (Smyth, 2005), by looking at the numbers of false positives and negatives.

2.3.2 Simulation Results

Web Table 1 gives the mean and SD for the hyperparameter estimates. The mean for all $\boldsymbol{\alpha}_d$ and $\boldsymbol{\alpha}$ are identical to their true values. The hyperparameters β_1, β_2, β_3, ν are somewhat underestimated, while λ and β are a bit over estimated. To assess the effects of hyperparameter estimates on the results, we plug in the true values for the β only while keeping others fixed at their estimated values, and do the same for ν and $\boldsymbol{\Lambda}$, then calculate the MB-statistic. The rank correlations between the MB-statistics with the above procedure for the β, $(\nu, \boldsymbol{\Lambda})$ and all the estimated hyperparameters are 0.97 and 0.99, respectively. The correlation between the MBs with all estimated and all true hyperparameters is 0.95. Web Figure 1 compares the average numbers of false positives and negatives between the four statistics. The lines from left to right refer to fully mod-

erated Wilks' lambda with all true hyperparameters, fully moderated Wilks' lambda with all hyperparameters estimated, the likelihood-based moderated Wilks' lambda, and the moderated F-statistic. Our fully moderated Wilks' lambda with all hyperparameters estimated achieved both lower numbers of false positives and false negatives than the likelihood-ratio moderated Wilks' lambda. This suggests that it is worthwhile to estimate the β_d and β and to include the two extra terms $\mathbf{M}_H$, $\mathbf{M}_K$ in gene ranking, rather than just use likelihood-based moderated Wilks' lambda, which is equivalent to using flat priors on $\boldsymbol{\mu}_d$ and $\boldsymbol{\mu}$. The fact that the red line is closer to the green line than to the blue line suggests that although not entirely precise, our hyperparameter estimation procedures lead to better results than would be obtained by setting them to 0 as in the case of likelihood-ratio moderated Wilks' lambda. The moderated F-statistic induces more false positives and false negatives. The reason for this may be that it ignores the correlations between gene expression values at different time points.

3   Cross-sectional Design

The longitudinal model in section 2 treats the entire time course as a vector, and make use of the multivariate normality assumption. A similar scheme for cross-sectional data is presented in this section. There are three key differences between longitudinal and cross-sectional models. First, in cross-sectional experiments, there is no true biological correlations among gene expression values over time. It should therefore be reasonable to assume time course observations are independent across times. Thus, instead of a general covariance matrix $\boldsymbol{\Sigma}$, we assume a common variance σ^2 for any observation. Second, we are able to model the expected temporal profile(s) with both structured and unstructured means, each having closed-form solutions for the posterior odds. By contrast, a longitudinal model with structured means apparently requires computational techniques such as MCMC for calculating the posterior odds. Third, our cross-sectional

model no longer has the constraint of a fixed set of time points, as is the case in our longitudinal models, although extending the longitudinal models to allow for arbitrary times is a future research topic of interest here.

Now we introduce our notation. Suppose the d–th biological condition has K_d time points, denoted by $t_{d1}, ..., t_{dK_d}$. The number of time points maybe different across biological conditions, and the sampling times maybe arbitrary within and between biological conditions. Let $\mathbf{Y}_{d,t_{dj}}$ be the $n_{dj} \times 1$ random vector of all observations at time t_{dj} for condition d and $\mathbf{Y}_d$ be the $n_d \times 1$ random vector of observations for the d–th biological condition: $\mathbf{Y}'_d = (\mathbf{Y}_{d,t_{d1}}, ..., \mathbf{Y}_{d,t_{dK_d}})$. The total number of observations is $n = \sum_{d=1}^{D} n_d$. We denote the mean gene expression level for condition d at time t by $\mu_d(t)$.

3.1 Multi-sample Problem

We begin with the multi-sample problem where genes of interest are those whose temporal profiles are different among biological conditions. The model for the one-sample problem where genes of interest are those which change over time can be derived using a very similar argument.

3.1.1 Structured Means

The structured means model is particularly useful when mRNA samples are taken at *arbitrary* times within and between biological conditions, for example, from mice of differing ages. In such cases, we are no longer able to compare the temporal profiles using the averages and replicate variances at a fixed set of time points. Instead, we compare them by comparing the regression coefficients, assuming each mean temporal profile can be modeled by a suitable function of *time*. This approach automatically takes into account the time ordering into the analysis. Under the alternative, we treat each biological condition as a single population and fit a linear model to each biological

condition *separately*, while under the null the heterogeneity among biological conditions is ignored, *i.e.*, all the observations are assumed to come from the same population and a *common* linear model is fitted to all the data. In other words, under the null, we assume the mean temporal profiles for all the biological conditions lie on the same curve. In this way, we can derive the posterior odds that the D mean temporal profiles of a gene come from different populations (i.e. biological conditions) against their coming from a common population.

3.1.2 Models and Priors

For the structured means model, $\mu_d(t)$ is modelled as a linear combination of $m - 1$ functions in time. We write

$$\mu_d(t) = \sum_{q=0}^{m-1} \beta_{d,q} f_q(t), \quad d = 1, ..., D.$$

The functions $f_q(t), q = 0, ..., m - 1$ are m spline basis functions. The constant m associated with the degree of freedom is subject to the constraint $0 < m \leq min(K_d), d = 1, ..., D$, so that the coefficients can be compared across biological conditions. Using matrix notation, we write $\boldsymbol{\theta}'_d = (\beta_{d,0}, ..., \beta_{d,m-1})$. Under the null ($I = 0$) that the expected time courses are *identical* among biological conditions, all the $\boldsymbol{\theta}_d$s equal to a common vector, while under the alternative, they differ, i.e.

$$H : \boldsymbol{\theta}_1 = ... = \boldsymbol{\theta}_D = \boldsymbol{\theta}, \sigma^2 > 0$$

$$K : \boldsymbol{\theta}_{d_1} \neq \boldsymbol{\theta}_{d_2}, \sigma^2 > 0 \qquad \text{for some pair} \quad d_1 \neq d_2.$$

This model can be written

$$\begin{cases} \mathbf{Y}_d | \boldsymbol{\theta}_d, \sigma^2, I = 1 \sim N\left(\mathbf{X}_d \boldsymbol{\theta}_d, \sigma^2 \mathbf{I}\right), \quad d = 1, ..., D; \\ \\ \\ \mathbf{Y}_d | \boldsymbol{\theta}, \sigma^2, I = 0 \sim N\left(\mathbf{X}_d \boldsymbol{\theta}, \sigma^2 \mathbf{I}\right), \end{cases}$$

where $\mathbf{X}_d$ is a $n_d \times m$ design matrix of full rank with suitable spline bases.

We further define the overall time course observation vector $\mathbf{Y}$ and overall design matrix $\mathbf{X}$ by

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_D \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_D \end{pmatrix}.$$

The natural conjugate priors for our multivariate normal models can be constructed using inverse-gamma and multivariate normal distributions, denoted by

$$\sigma^2 \sim Inv\text{-}gamma\left(\frac{\nu}{2}, \frac{\nu\lambda^2}{2}\right)$$

$$\begin{cases} \boldsymbol{\theta}_d|\sigma^2, I = 1 \sim N\left(\boldsymbol{\alpha}_d, \sigma^2\boldsymbol{\Omega}_d^{-1}\right) \quad d = 1, ..., D \\ \\ \boldsymbol{\theta}|\sigma^2, I = 0 \sim N\left(\boldsymbol{\alpha}, \sigma^2\boldsymbol{\Omega}^{-1}\right), \end{cases}$$

where $\boldsymbol{\Omega}_d^{-1}, d = 1, ..., D$ and $\boldsymbol{\Omega}^{-1}$ are $m \times m$ diagonal matrices, with diagonal elements ω_{dj}^{-1} and $\omega_j^{-1}, j = 1, ..., m$, respectively. Using the priors and models decribed above, we obtain the posterior odds in the next section.

3.1.3 Posterior Odds

The posterior odds against the null that the expected temporal profiles among biological conditions can be easily derived, and are

$$\mathbf{O} = \frac{p}{1-p}\frac{P(data|I = 1)}{P(data|I = 0)}$$

$$= \frac{p}{1-p}\left(\frac{|\mathbf{I}_m + (\sum_{d=1}^D \mathbf{X}_d'\mathbf{X}_d)\boldsymbol{\Omega}^{-1}|}{\prod_{d=1}^D |\mathbf{I}_m + (\mathbf{X}_d'\mathbf{X}_d)\boldsymbol{\Omega}_d^{-1}|}\right)^{\frac{1}{2}}\left(\frac{TSS + m_H + \nu\lambda^2}{WSS + m_K + \nu\lambda^2}\right)^{\frac{1}{2}(n+\nu)},$$

(7)

where

$$\hat{\boldsymbol{\theta}}_d = (\mathbf{X}_d' \mathbf{X}_d)^{-1} \mathbf{X}_d' \mathbf{Y}_d,$$

$$WSS_d = (\mathbf{Y}_d - \mathbf{X}_d \hat{\boldsymbol{\theta}}_d)'(\mathbf{Y}_d - \mathbf{X}_d \hat{\boldsymbol{\theta}}_d), \ WSS = \sum_{d=1}^{D} WSS_d, \ \text{and}$$

$$m_d = (\hat{\boldsymbol{\theta}}_d - \boldsymbol{\alpha}_d)'((\mathbf{X}_d' \mathbf{X}_d)^{-1} + \boldsymbol{\Omega}_d^{-1})^{-1}(\hat{\boldsymbol{\theta}}_d - \boldsymbol{\alpha}_d), \ m_K = \sum_{d=1}^{D} m_d,$$

are least squares estimate for $\boldsymbol{\theta}_d$, within-condition residual sums of squares from the d-th regression fit, and quantities involving condition-specific prior means, respectively. In addition, the single population quantities (under H) are

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y},$$

$$TSS = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}}), \ \text{and}$$

$$m_H = (\hat{\boldsymbol{\theta}} - \boldsymbol{\alpha})'((\mathbf{X}' \mathbf{X})^{-1} + \boldsymbol{\Omega}^{-1})^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\alpha})$$

are least squares estimate for $\boldsymbol{\theta}$, the total residual sums of squares, and quantity involving prior means, respectively.

If the design matrices are orthonormalized, equation (7) can be further simplified by replacing $\mathbf{X}_d' \mathbf{X}_d$ by $\mathbf{I}$. When $\mathbf{X}_d$ is the same across genes, $\mathbf{O}$ is a monotonic increasing function of

$$\frac{TSS + m_H + \nu\lambda^2}{WSS + m_K + \nu\lambda^2}. \tag{8}$$

The above expression is our fully moderated F-statistic. Under the alternative when the expected temporal profiles come from different curves (populations), the total residual sums of squares (TSS) from fitting the same curve to all the data is larger than the within-condition residual sums of squares (WSS) from fitting a separate curve to each

biological condition. If $\omega_{dj} \to 0$ and $\omega_j \to 0$, m' and m vanish and the posterior odds are equivalent to

$$\frac{TSS + \nu\lambda^2}{WSS + \nu\lambda^2}.$$

In the limiting case that $\nu \to \infty$, the posterior odds are

$$\frac{p}{1-p} \left(\frac{|\mathbf{I}_m + (\sum_{d=1}^{D} \mathbf{X}'_d \mathbf{X}_d) \mathbf{\Omega}^{-1}|}{\prod_{d=1}^{D} |\mathbf{I}_m + (\mathbf{X}'_d \mathbf{X}_d) \mathbf{\Omega}_d^{-1}|} \right)^{\frac{1}{2}} exp\left\{ TSS + m - WSS - m' \right\}.$$

On the other hand, when $\nu \to 0$, the posterior odds are just simply equation (7) with ν replaced by 0.

3.2 Unstructured Means

The above structured means model can be used for aribitrary sampling times. Suppose gene expression levels are measured at a fixed set of k time points across biological conditions, i.e. $K_d = k, d = 1, ..., D$, and $t_{d_1,j} = t_{d_2,j} = t_j, d_1 \neq d_2, j = 1, ..., k$, we can also use the unstructured means model which estimates the expected time course using the (condition-specific) sample averages rather than the model-based fitted values. Now the vector of regression coefficients equal to the mean time course vector, i.e. $\boldsymbol{\theta}_d = \boldsymbol{\mu}_d$, $\boldsymbol{\theta} = \boldsymbol{\mu}, m = k$, and $\mathbf{X}_d$ and $\mathbf{X}$ are design matrices with 0s and 1s. Let n_j and n_{dj} be the total number of replicates at the j–th time point and the number of replicates for the d–th condition at the j–th time point, respectively. That is $n_j = \sum_{d=1}^{D} n_{dj}$. The posterior odds become

$$\frac{p}{1-p} \left(\frac{\prod_{j=1}^{k}(1 + n_j/\omega_j)}{\prod_{d=1}^{D} \prod_{j=1}^{k}(1 + n_{dj}/\omega_{dj})} \right)^{\frac{1}{2}} \left(\frac{TSS + m_H + \nu\lambda^2}{WSS + m_K + \nu\lambda^2} \right)^{\frac{1}{2}(n+\nu)}, \qquad (9)$$

where $\overline{Y}_{dj} = n_{dj}^{-1} \sum_{i=1}^{n_{dj}} Y_{dji}$, $\overline{Y}_j = n_j^{-1} \sum_{d=1}^{D} n_{dj} \overline{Y}_{dj}$, $TSS = \sum_{d=1}^{D} \sum_{j=1}^{k} \sum_{i=1}^{n_{dj}} (Y_{dji} - \overline{Y}_j)^2$, $WSS = \sum_{d=1}^{D} \sum_{j=1}^{k} \sum_{i=1}^{n_{dj}} (Y_{dji} - \overline{Y}_{dj})^2$, $m_H = \sum_{j=1}^{k} (n_j^{-1} + \omega_j^{-1})^{-1} (\overline{Y}_j - \alpha_j)^2$, $m_K = \sum_{d=1}^{D} \sum_{j=1}^{k} (n_{dj}^{-1} + \omega_{dj}^{-1})^{-1} (\overline{Y}_{dj} - \alpha_{dj})^2$ are the average gene expression value at the j–th

time point for condition d only and for all the conditions, the total and within sums of squares, and quantities involving (condition-specific) prior means, respectively. Under the alternative, the total sums of squares (TSS) is larger than within sums of squares (WSS) compared to the null.

If n_{dj} are the same across genes, and $\omega_j \to 0$, $\omega_{dj} \to 0$, and $\nu \to 0$, then the posterior odds are equivalent to the standard F-statistic from one-way ANOVA treating biological condition as the factor

$$\mathbf{O} \propto \frac{TSS}{WSS}.$$

3.3 Outliers

The presence of outlying gene expression measurements in the microarray context is not unusual. It is well known that least squares method is vulnerable to outliers (see *e.g.* Rousseeuw and Leroy (1987)). Although our structured means model is based on least squares estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_d$, it is straightforward to deal with outliers by replacing least squares estimates in equation (7) with robust estimators. Examples are the M-estimators (Huber, 1981), including L_1, Huber, and Tukey's bisquare estimators, see *e.g.* Rousseeuw and Leroy (1987). The latter two estimator are more commonly used, and their estimation procedures are usually conducted using iterative reweighted least squares (IRLS).

3.4 One-sample Problem

When there is only one biological condition ($D = 1$) and genes of interest are those which change over time, we use the same framework in section 3.1 with slightly different priors and models to derive the posterior odds $\mathbf{O}$.

3.5   Structured Means

3.5.1   Models and Priors

Using the same notation as section 3.1 with the subscript d dropped, the null hypothesis is $H : \beta_1 = ... = \beta_{m-1} = 0, \sigma^2 > 0$ and the alternative hypothesis is $K : \boldsymbol{\theta} \neq \mathbf{0}, \sigma^2 > 0$. The models become

$$\begin{cases} \mathbf{Y}|\boldsymbol{\theta}, \sigma^2, I = 1 \sim N\left(\mathbf{X}\boldsymbol{\theta}, \sigma^2\mathbf{I}\right) \\ \\ \mathbf{Y}|\beta_0, \sigma^2, I = 0 \sim N\left(\mathbf{1}\beta_0, \sigma^2\mathbf{I}\right). \end{cases}$$

The priors are

$$\sigma^2 \sim Inv\text{-}gamma\left(\frac{\nu}{2}, \frac{\nu\lambda^2}{2}\right)$$

$$\begin{cases} \boldsymbol{\theta}|\sigma^2, I = 1 \sim N\left(\boldsymbol{\alpha}, \sigma^2\boldsymbol{\Omega}^{-1}\right) \\ \\ \beta_0|\sigma^2, I = 0 \sim N\left(\alpha_0, \omega^{-1}\sigma^2\right). \end{cases}$$

The posterior odds are

$$\mathbf{O} = \frac{p}{1-p}\left(\frac{1 + n\omega^{-1}}{|\mathbf{I}_m + (\mathbf{X}'\mathbf{X})\boldsymbol{\Omega}^{-1}|}\right)^{\frac{1}{2}}\left(\frac{RSS_H + m_H + \nu\lambda^2}{RSS_K + m_K + \nu\lambda^2}\right)^{\frac{1}{2}(n+\nu)}, \tag{10}$$

where $RSS_K = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})$ and $RSS_H = (\mathbf{Y} - \mathbf{1}\hat{\beta}_0)'(\mathbf{Y} - \mathbf{1}\hat{\beta}_0)$ are the residual sums of squares under the alternative and null, respectively. The quantity $m_K = m$ in section 3.1.1, and

$$m_H = (n^{-1} + \omega^{-1})^{-1}(\hat{\beta}_0 - \alpha_0)^2.$$

In the case of a two-channel comparative experiment or a paired two-sample experiment, it is reasonable to have the null hypothesis as $H : \boldsymbol{\theta} = 0, \sigma^2 > 0$. In this case, the

17

posterior odds against the null hypothesis become

$$\mathbf{O} = \frac{p}{1-p} \left| \mathbf{I}_m + (\mathbf{X}'\mathbf{X})\mathbf{\Omega}^{-1} \right|^{-\frac{1}{2}} \left(\frac{RSS_H + m_H + \nu\lambda^2}{RSS_K + m_K + \nu\lambda^2} \right)^{\frac{1}{2}(n+\nu)}, \tag{11}$$

where RSS_K and RSS_H are the residual sums of squares under the alternative and null, respectively, and

$$m_H = n\hat{\beta}_0^2.$$

3.6   Unstructured Means

As in section 3.1, when the sampling times are fixed, it is possible to use the unstructured means model. Using the same notation as in section 3.2 with d dropped, we get the posterior odds

$$\mathbf{O} = \frac{p}{1-p} \left(\frac{1 + n\omega^{-1}}{\prod_{j=1}^{k}(1 + n_j\omega_j^{-1})} \right)^{\frac{1}{2}} \left(\frac{RSS_H + m_H + \nu\lambda^2}{RSS_K + m_K + \nu\lambda^2} \right)^{\frac{1}{2}(n+\nu)}, \tag{12}$$

where $\overline{\mathbf{Y}}_j = n_j^{-1} \sum_{i=1}^{n_j} \mathbf{Y}_i$, $\overline{\overline{\mathbf{Y}}} = n^{-1} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \mathbf{Y}_{ji}$, $RSS_K = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\mathbf{Y}_{ji} - \overline{\mathbf{Y}}_j)^2$ and $RSS_H = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\mathbf{Y}_{ji} - \overline{\overline{\mathbf{Y}}})^2$ are the average gene expression value at the j-th time point and across all time points, within time and total sums of squares, respectively, m_K is defined as before with $\hat{\boldsymbol{\theta}} = \overline{\mathbf{Y}}$, and

$$m_H = (n^{-1} + \omega^{-1})^{-1}(\overline{\overline{\mathbf{Y}}} - \alpha_0)^2.$$

In the case of $H : \boldsymbol{\theta} = \mathbf{0}, \sigma^2 > 0$, the posterior odds are

$$\mathbf{O} = \frac{p}{1-p} \left(\prod_{j=1}^{k} (1 + n_j\omega_j^{-1}) \right)^{-\frac{1}{2}} \left(\frac{RSS_H + m_H + \nu\lambda^2}{RSS_K + m_K + \nu\lambda^2} \right)^{\frac{1}{2}(n+\nu)}, \tag{13}$$

$$m_H = n\overline{\overline{\mathbf{Y}}}^2.$$

We describe how we estimate the hyperparameters associated with the posterior odds in Web Appendix C.

3.7 Simulation Study

3.7.1 Method

As in the longitudinal model, we perform a simulation study to assess the performance of our cross-sectional MB statistic. We simulate the data based on a real two-sample data example (D=2). Specifically, $1,000$ datasets are simulated. Each dataset contains 10,000 genes, and among them, 200 (p=0.02) are differentially expressed. The first biological condition has 14 samples ($n_1 = 14$) at ages (days) 58, 58, 79, 79, 96, 96, 102, 124, 124, 124, 149, 157, 157, 157. The second condition has 15 samples ($n_2 = 15$) at ages 58, 79, 79, 80, 96, 97, 97, 117, 124, 124, 124, 149, 149, 155, 166. The data are simulated based on a polynomial basis design matrix with quadratic terms (m=3). The hyperparameters are $\nu = 4$, $\lambda = 1$, $\alpha_1 = (17, 10, -5)$, $\alpha_2 = (5, 3, -2)$, $\alpha = (11, 2, -3)$, and

$$\boldsymbol{\Omega}_1^{-1} = \boldsymbol{\Omega}_2^{-1} = \begin{pmatrix} 12 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 1.6 \times 10^{-7} \end{pmatrix} \qquad \boldsymbol{\Omega}^{-1} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 1.7 \times 10^{-7} \end{pmatrix}.$$

We compare our MB-statistic (equation 7) with design matrices of B-spline basis of 3 and 6 degrees of freedom (3-df and 6-df MB), and the fully moderated likelihood ratio statistic (equation 7 with the diagonal elements of $\boldsymbol{\Omega}^{-1}$, $\boldsymbol{\Omega}_1^{-1}$ and $\boldsymbol{\Omega}_2^{-1}$ approach ∞), and the moderated F-statistic (Smyth, 2004) under the linear model setting.

3.7.2 Simulation Results

The estimated hyperparameters are all very close to their true values, so we omit the details here. This shows that our estimation procedures worked reasonably well. Web Figure 2 shows the ROC curves for these four statistics. Our cross-sectional 3-df MB performs slightly better than the 6-df MB when the number of false nagatives is less than 14, and became worse than the 6-df MB when the number of false nagatives is more than 14. They both achieve fewer numbers of false positives and false negatives than the fully moderated likelihood ratio statistic and the moderated F-statistic, except that the 3-df MB performed slightly wrose than the latter two statistics when the number of false negatives was larger than 28. The moderated F-statistic and the fully moderated likelihood ratio statistic perform similarly.

4 Applications

We illustrate the differences between our cross-sectional MB statistic and other published methods using two published microarray time course datasets of cDNA and Affymetrix platforms. Our goal is to find genes with large differences in their temporal patterns between conditions. We should be aware that there are many ways to define differential expression between such patterns, and investigators should choose the one most appropriate to the aims of their study.

4.1 Dauer exit time course study

Under unfavorable conditions (e.g. low amounts of food, high population density or temperature), *C. elegans* can develop into dauer larvae, which possess several properties (arrested, non-feeding, long-lived, stress-resistant) leading to enhanced survival. When a favorable environment returns, the dauer will continue growth. By contrast, L1 worms

do not have these properties as dauer, although they will also be arrested under a low food regime. In addition, L1 worms will resume their growth when more food is available.

Wang and Kim (2003) performed a cDNA microarray experiment for the feeding responses of dauer exit and L1 starvation processes over a 12-hour period (12 time points). Each time point had either 3 or 4 replicates. One of the aims of this study was to find dauer-recovery specific genes, which are the genes changing over time in the dauer exit process and with different temporal profiles from the L1 starvation process. To identify genes that change in expression during the dauer exit timecourse, Wang and Kim (2003) performed a standard one-way ANOVA. A total of 2,430 genes were identified, which were followed up for differential temporal patterns between the dauer exit and L1 starvation processes.

Hong and Li (2006) used their functional hierarchical method to these 2,430 genes to identify those with different temporal patterns between the two conditions. To compare the MB with their method, we calculate the cross-sectional MB for the same set of 2,430 genes. We use a B-spline basis with 5 degrees of freedom in our analysis. In addition, we also compare our MB results with the moderated F-statistic under the linear model setting, testing the differences in time effects between the two conditions to be 0.

4.1.1 Results

Figures 1-3 show the top 8 ranked genes from our 5-df MB statistic, the moderated F statistic and Hong and Li (2006), respectively. Seven out of the top 8 genes by MB are all in the top 8 genes by the moderated F. The MB and the moderated F perform very similarly. The top genes from our MB and the moderated F show large differences in temporal profiles between these two conditions. It is clear to see that our B-spline

basis fits the data well. Many genes with large temporal profile differences between conditions and ranked highly by our MB or moderated F are ranked differently by Hong and Li (2006). For example, the gene ranked 1 by MB and moderated F (Kim Lab ID=Y59A8C.D) is ranked 979 by Hong and Li (2006) even though this gene exhibits a large difference in temporal profiles between conditions. The top 8 genes by MB are ranked 1, 2, 3, 5, 4, 6, 9, and 7 by the moderated F model. In addition, these genes are ranked 979, 930, 748, 982, 301, 64, 150 and 660 by the functional hierarchical model (Hong and Li, 2006). On the other hand, the MB ranks for genes ranked from 1 to 8 by moderated F are 1, 2, 3, 5, 4, 6, 8, and 13. The MB ranks for genes ranked from 1 to 8 by Hong and Li (2006) are 1,541, 426, 1,110, 428, 654, 958, 161 and 270. The top 8 genes by MB and moderated F show large difference in temporal profiles, however, most of them are not ranked as the top 400 by Hong and Li (2006). The MB ranks for the top genes are very different from those obtained using the Hong and Li (2006) ranking. On average, the gene rankings by MB and moderated F seem more fitted to our goal.

4.2   Ames dwarf mice aging study

Ames dwarf mice have a mean life span at least 49% longer than their wildtype siblings. Homozygous for the *df* allele at the transcription factor *Prop1* responsible for proper embryonic development, Ames dwarf mice suffer from multiple hormone deficiency resulting in severe growth retardation. Amador-Noguez et al. (2004) studied the biochemical and metabolic pathways that slow aging in this mouse model using oligonucleotide arrays. Liver expression levels in 12 Ames dwarf mice and their wildtype strains were measured at ages 3, 6, 12, and 24 months using Affymetrix MOE430A microarrays. Three biological replicates were used for each time/strain, so there were 24 arrays in total. To identify alterations in gene expression that discriminate between

mutant and wildtype mice, Amador-Noguez et al. (2004) performed ANOVA and found 1,125 differentially expressed genes. We applied our cross-sectional MB to this dataset, using the structured mean model with a B-spline basis with 3 degrees of freedom.

4.2.1 Results

Figures 4 and 5 show the genes of different ranks by the MB, the moderated F and ANOVA by Amador-Noguez et al. (2004), respectively. By visual inspection on the temporal profiles of those top genes, the MB tends to favor genes with large differences in temporal patterns. The ANOVA approach tends to favor genes with small replicate variabilities regardless of the temporal differences. For example, 1452073_at is ranked 15 by ANOVA, however, it seems that there is barely any temporal difference between the two genotypes. The gene 1455640_a_at is ranked higher than 1421075_s_at by ANOVA even though the temporal difference of the former is smaller. There is a certain amount of similarity between MB and moderated F (70% overlapping for the top 100), and both methods select the same number 1 gene.

5   Discussion

In this paper, we revisit the gene ranking problem for microarray time course data. We derive a multivariate empirical Bayes (MB) statistics for both the longitudinal and cross-sectional multi-sample problem, where genes of interest are those having different temporal profiles across multiple biological conditions. As in Tai and Speed (2006), our focus here is on gene ranking, not on testing significance. Such a statistic is useful in practice, since typically only a limited number of genes can be followed up. The simulation studies suggest that our proposed statistics and estimation procedures give reasonable results.

Our longitudinal model assumes the same gene-specific $\boldsymbol{\Sigma}$ across biological conditions. Although this assumption may not be entirely satisfactory in many cases, it does not preclude us from getting sensible results. It is possible to relax this assumption at the cost of more intensive computation. Without this assumption, the likelihood-based method requires iterative procedures to compute the statistic, see section 5.4 in Mardia et al. (2000), and in our context, a MCMC soluation would be needed. Our cross-sectional model is also demonstrated using two published microarray time course studies. One limitation of our cross-sectional model is that it assumes a common variance σ^2 at all time points. Compared to the cost in computational burden when this assumption is relaxed, there is much more gain than loss with this limitation. As shown in our case studies, with this limitation it still gives good results.

Bayesian MANOVA or multivariate regression models were visited previously, see *e.g.* Raiffa and Schlaifer (1961); Press (1980, 1982); Press and Shigemasu (1985); Jelenkowska and Press (1997). These works used either noninformative, natural conjugate or generalized natural conjugate priors. In the multivariate normal setting, generalized natural conjugate prior comes after ordinary natural conjugate prior to remove certain constraints on the parameters from the prior density. The generalized natural conjugate prior differs from the ordinary one in that the former assumes that the priors on each parameters are independent (Press, 1982). However, both produce a Normal-Inverse Wishart density (Aitchison and Dunsmore, 1975) for the joint posterior distribution of the parameters (Press, 1980, 1982). We use conjugate priors to get closed-form solutions for the longitudinal posterior odds. To the best of our knowledge, this is the first *fully moderated* Wilks' lambda derived in the literature.

The B-spline fitting for our cross-sectional model works reasonably well for both long and short time courses as in the two case studies in section 4. Alternatively, for short time courses with fixed time points across conditions (*e.g.* the Ames dwarf mice aging study), the unstructured mean model could be used.

24

6   Acknowledgements

References

Aitchison, J. and Dunsmore, I. R. (1975). *Statistical Prediction Analysis.* Cambridge University Press.

Amador-Noguez, D., Yagi, K., Venable, S., and Darlington, G. (2004). Gene expression profile of long-lived Ames dwarf mice and Little mice. *Aging Cell*, 3(6):423–441.

Bar-Joseph, Z., Gerber, G., Simon, I., Gifford, D. K., and Jaakkola, T. S. (2003). Comparing the continuous representation of time-series expression profiles to identify differentially expressed genes. *Proceedings of the National Academy of Sciences USA*, 100(18):10146–10151.

Cho, R., Campbell, M., Winzeler, E., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T., Gabrielian, A., Landsman, D., Lockhart, D., and Davis, R. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell.*, 2(1):65–73.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O., and Herskowitz, I. (1998). The transcriptional program of sporulation in budding yeast. *Science*, 282(5389):699–705.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2000). *Bayesian data analysis.* Chapman & Hall/CRC.

Hong, F. and Li, H. (2006). Functional hierarchical models for identifying genes with different time-course expression profiles. *Biometrics*, 62(2):534–544.

Huber, P. J. (1981). *Robust Statistics.* John Wiley & Sons, Inc.

Jelenkowska, T. and Press, S. (1997). Bayesian inference in the multivariate mixed model MANOVA. *American Journal of Mathematical and Management Sciences*, 17:97–116.

Mardia, K., Kent, J., and Bibby, J. (2000). *Multivariate analysis.* Academic Press.

Press, S. (1980). Bayesian Inference in MANOVA. In Krishnaiah, P., editor, *Handbook of Statistics, Vol. I.* North-Holland Publishing Company. 117-132.

Press, S. (1982). *Applied multivariate analysis : using Bayesian and frequentist methods of inference.* Dover Publications.

Press, S. and Shigemasu, K. (1985). Bayesian MANOVA and MANOCOVA under exahangeability. *Communications in Statistics-Theory and Methods*, 14(5):1053–1078.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory.* Harvard Business School Publications.

Rousseeuw, R. J. and Leroy, A. (1987). *Robust regression and outlier detection.* John Wiley & Sons, Inc., New York.

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1):article 3.

Smyth, G. K. (2005). Limma: linear models for microarray data. In Gentleman, R., Carey, V., Dudoit, S., Irizarry, R., and Huber, W., editors, *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Springer, New York.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol. Biol. Cell*, 9(12):3273–3297.

Storch, K.-F., Lipan, O., Leykin, I., Viswanathan, N., Davis, F. C., Wong, W. H., and Weitz, C. J. (2002). Extensive and divergent circadian gene expression in liver and heart. *Nature*, 417:78–83.

Storey, J. D., Xiao, W., Leek, J. T., Tompkins, R. G., and Davis, R. W. (2005). Significance analysis of time course microarray experiments. *Proceedings of the National Academy of Sciences USA*, 102(36):12837–12842.

Tai, Y. C. (2005). *Multivariate empirical Bayes models for replicated microarray time course data*. Ph.D. dissertation, Division of Biostatistics, University of California, Berkeley.

Tai, Y. C. and Speed, T. P. (2005). Statistical analysis of microarray time course data. In Nuber, U., editor, *DNA Microarrays*. BIOS Scientific Publishers Limited, Taylor & Francis, 4 Park Square, Milton Park, Abingdon OX14 4RN. Chapter 20.

Tai, Y. C. and Speed, T. P. (2006). A multivariate empirical Bayes statistic for replicated microarray time course data . *Annals of Statistics*, 34(5):2387–2412.

Wang, J. and Kim, S. K. (2003). Global analysis of dauer gene expression in Caenorhabditis elegans. *Development*, 130(8):1621–1634.

Wildermuth, M. C., Tai, Y. C., Dewdney, J., Denoux, C., Hather, G., Speed, T. P., and Ausubel, F. M. (2007). Application of $\widetilde{T}^2$ statistic to temporal global arabidopsis expression data reveals known and novel salicylate-impacted processes. *In preparation.*

Yuan, M. and Kendziorski, C. (2006). Hidden markov models for microarray time course data under multiple biological conditions. *Journal of the American Statistical Association*, 101(476):1323–1340.
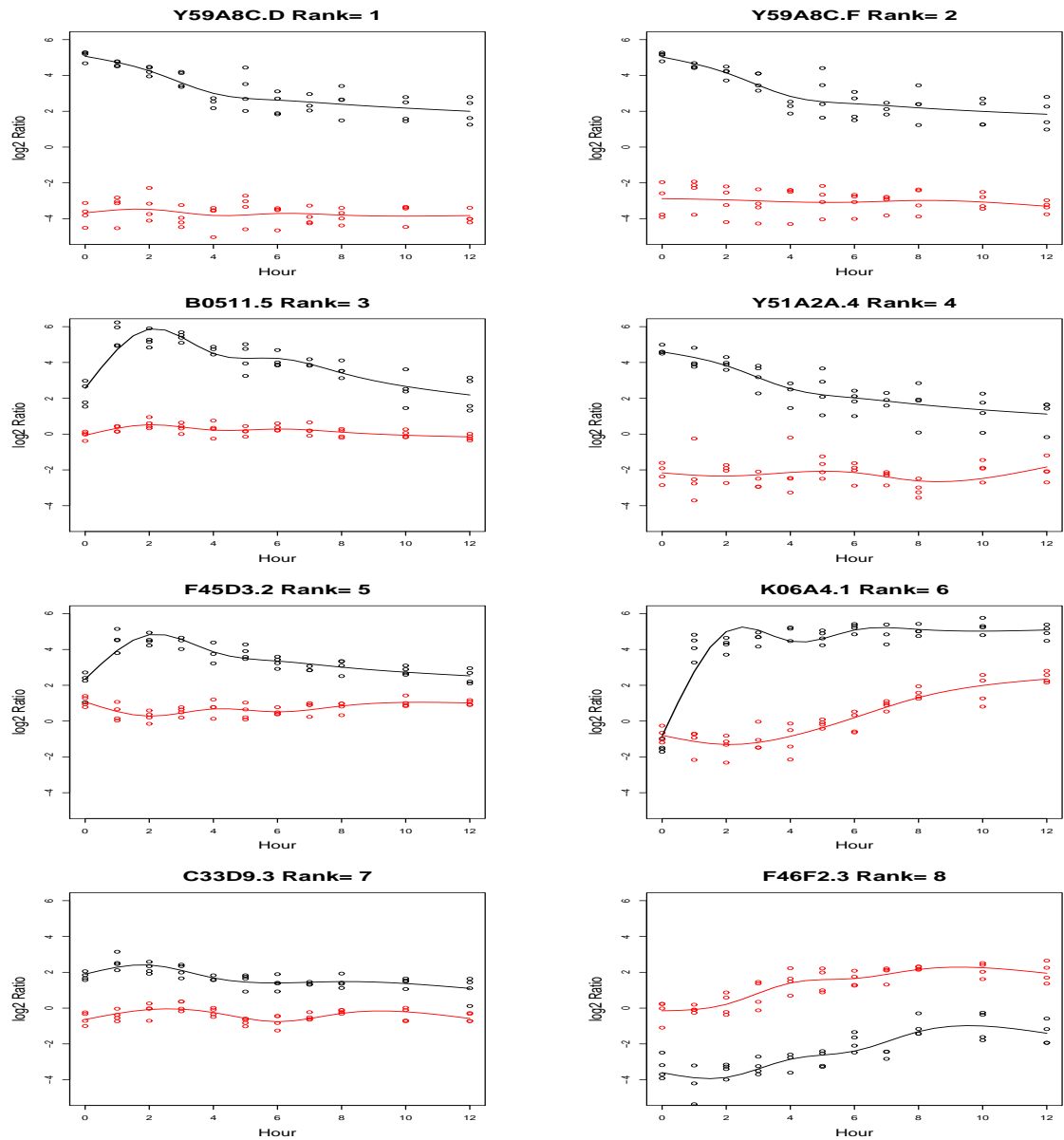
Figure 1: Top 8 genes with the highest MB statistic using B-spline basis design matrices. The solid curves are fitted values for the dauer exit timecourse (black) and the L1 starvation time course (red).
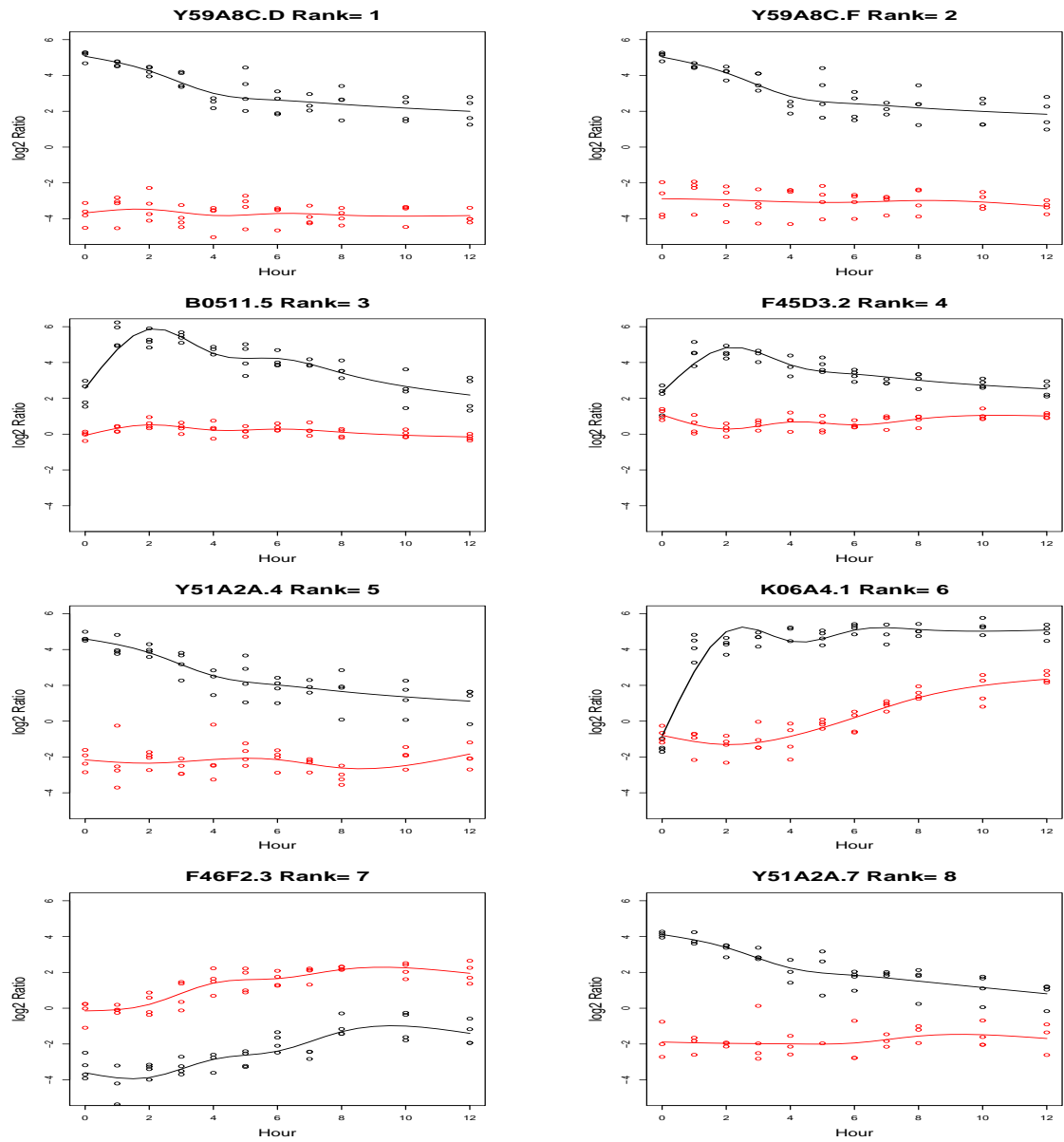
Figure 2: Top 8 genes from the moderated F-statistic. The solid curves are fitted values for the dauer exit timecourse (black) and the L1 starvation time course (red) from the MB statistic
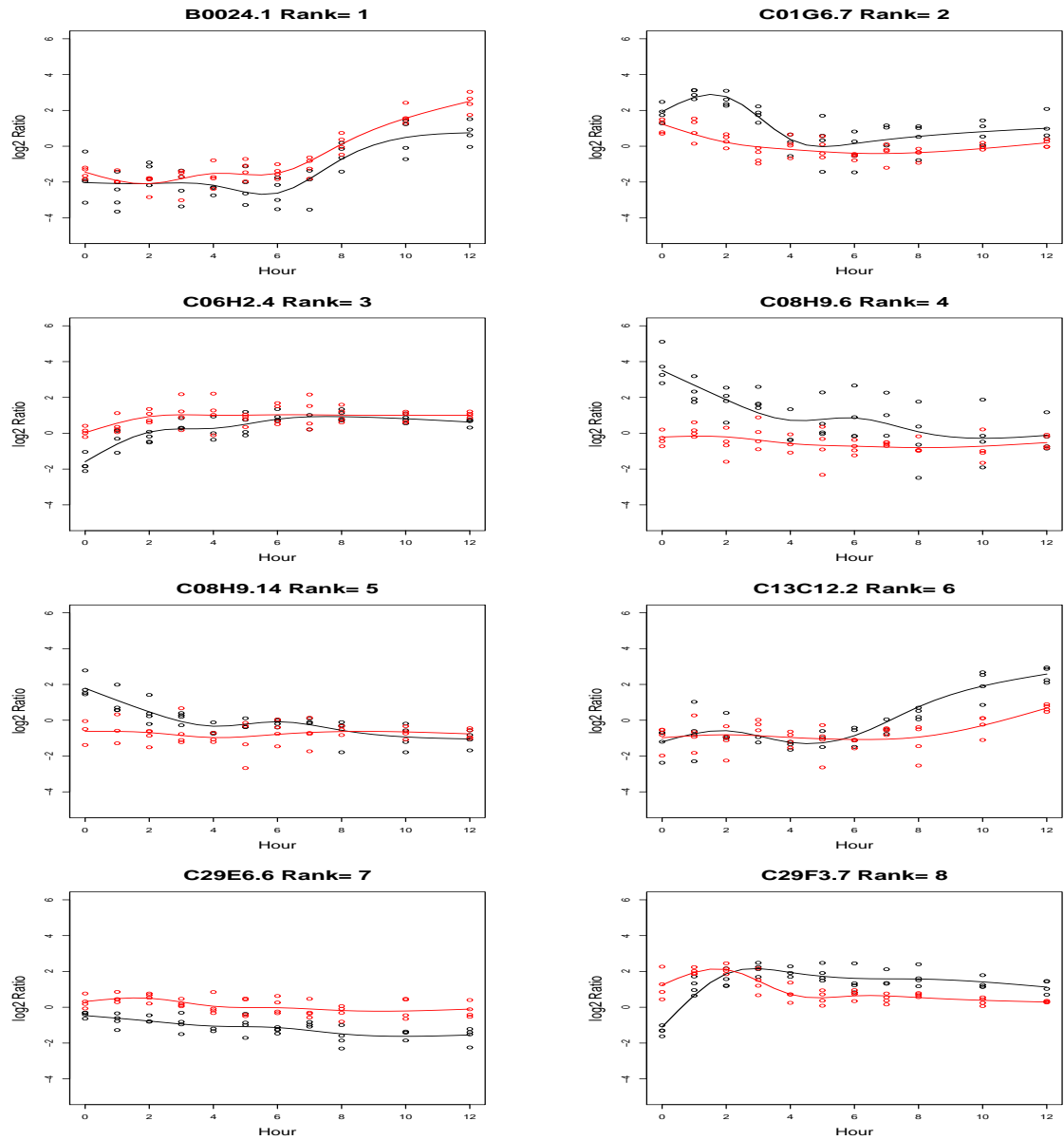
Figure 3: Top 8 genes from Hong and Li (2006). The solid curves are fitted values for the dauer exit timecourse (black) and the L1 starvation time course (red) from the MB statistic
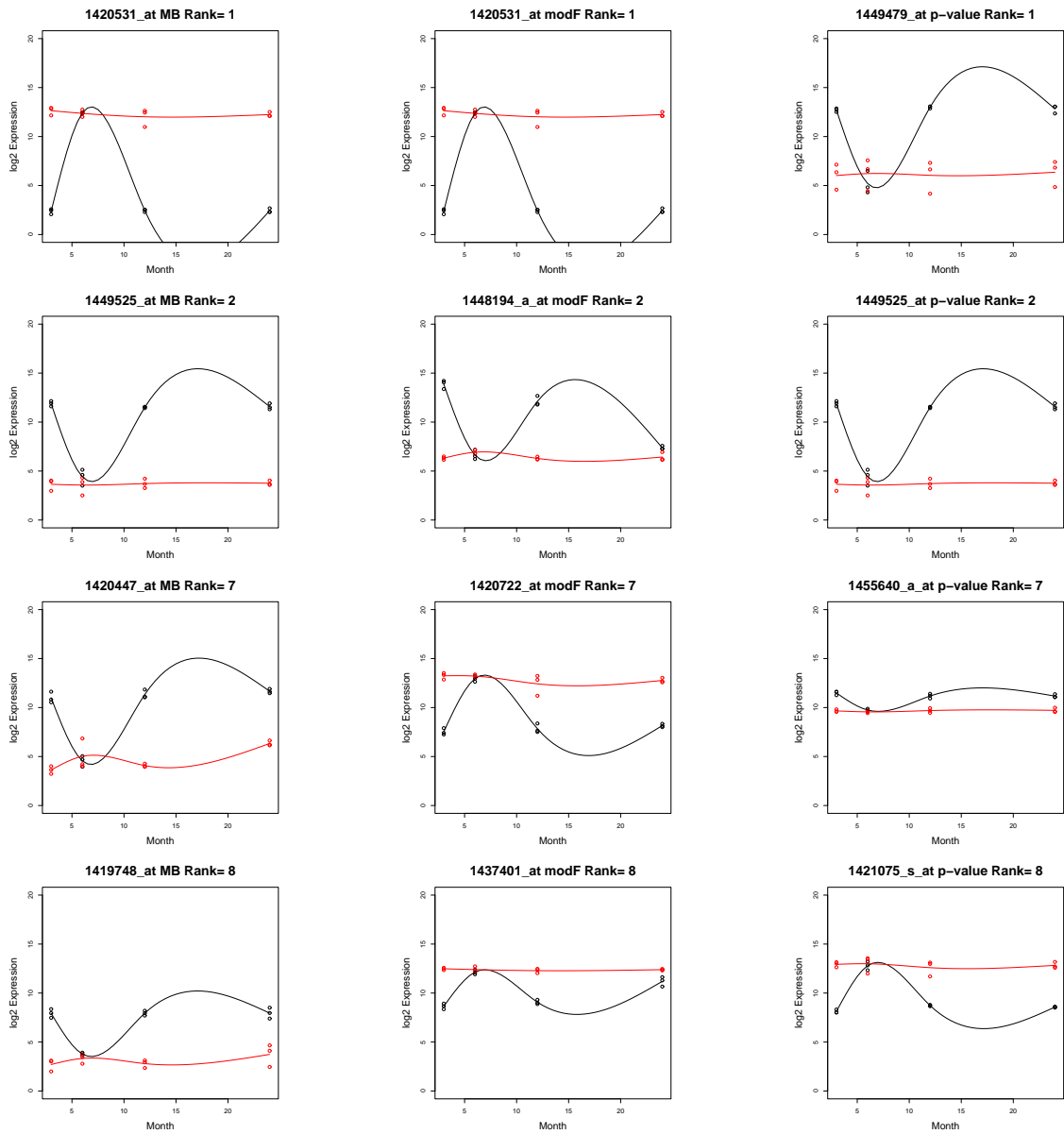
Figure 4: Genes of ranks 1, 2, 7, 8 by MB (left panel), moderated F (middle panel) and ANOVA (right panel). The solid curves are fitted values for the Ames dwarf mice (black) and the wildtype mice (red).
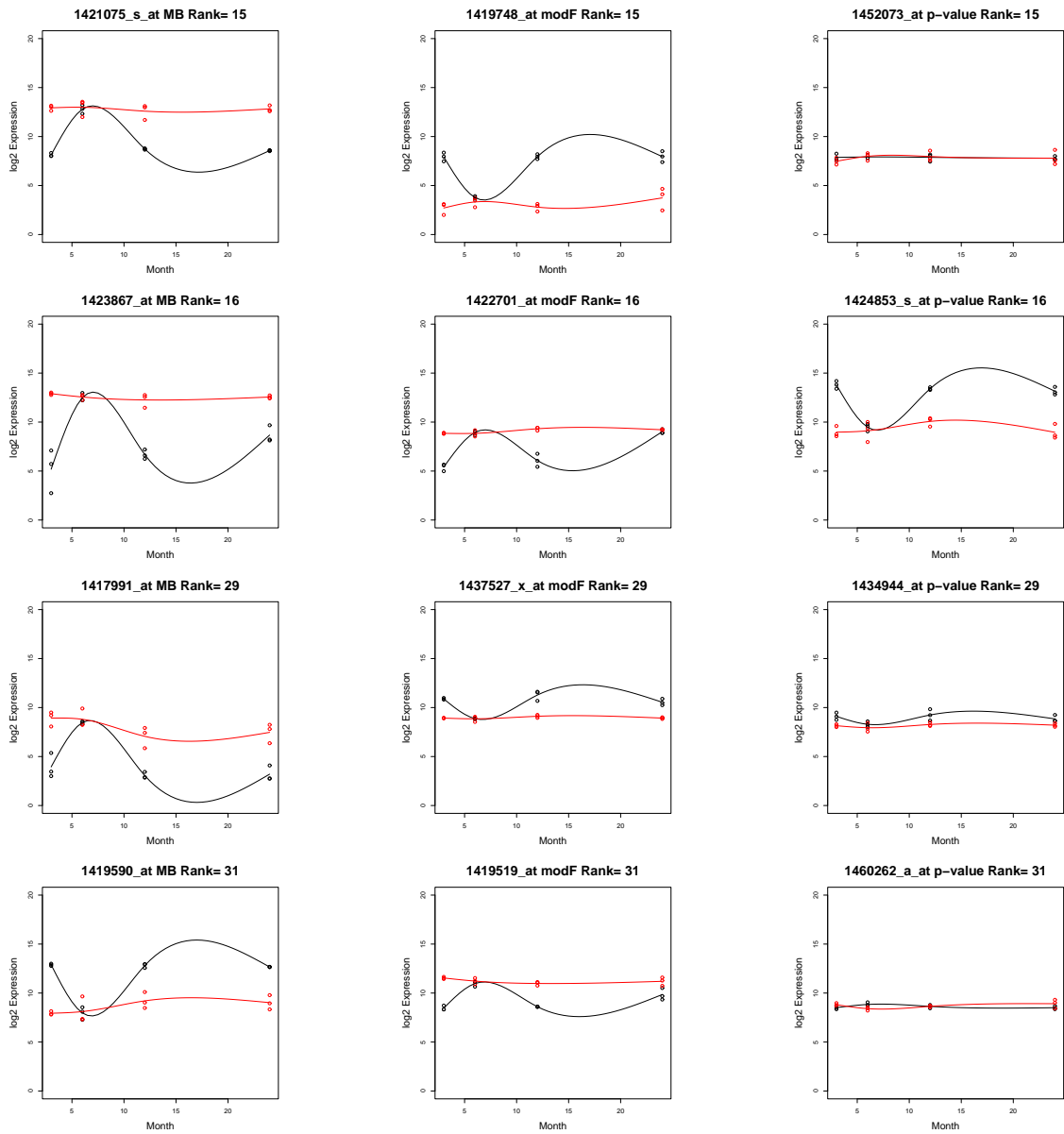
Figure 5: Genes of ranks 15, 16, 29, 31 by MB (left panel), moderated F (middle panel) and ANOVA (right panel). The solid curves are fitted values for the Ames dwarf mice (black) and the wildtype mice (red).