Identifying temporally differentially expressed genes through functional principal components analysis

XUELI LIU*
Division of Biostatistics, City of Hope, Duarte, CA 91010-3000, USA
xuliu@coh.org

MARK C. K. YANG
Department of Statistics, University of Florida, Gainesville, Florida 32611-8545, USA

SUMMARY
Time course gene microarray is an important tool to identify genes with differential expressions over time. Traditional analysis of variance (ANOVA) type of longitudinal investigation may not be applicable because of irregular time intervals and possible missingness due to contamination in microarray experiments. Functional principal components analysis is proposed to test hypotheses in the change of the mean curves. A permutation test under a mild assumption is used to make the method more robust. The proposed method outperforms the recently developed extraction of differential gene expression and a 2-way mixed effects ANOVA under reasonable gene expression models in simulation. Real data on transcriptional profiles of blood cells microarray from treated and untreated individuals were used to illustrate this method.

Keywords: False discovery rate; Functional hypothesis testing; Functional principal components analysis; Permutation test; Time course gene expression profiles.

1. INTRODUCTION
Statistical methods to identify differentially expressed genes from a single time point were developed in the early 2000s (e.g. Kerr and others, 2000; Efron and others, 2001; Storey and Tibshirani, 2003). At the same time, clustering and classification methods were also developed for temporally observed gene expression under 1 condition (e.g. Eisen and others, 1998; Bar-Joseph and others, 2003; Liu and Müller, 2003). As the costs of microarray experiments decrease rapidly, researchers can begin to investigate temporally differentially expressed genes across different experimental conditions.

Many time course gene expression profiles assume an independent sampling scheme in the sense that these microarray experiments sacrifice different animals at each time point. However, in this paper, our focus is to propose an approach to detecting time differentially expressed genes with longitudinal correlation structure data, which often arise from time course microarray experiments where repeated
measurements are taken from the same subject. A motivating example is provided by a time course 2-color complementary DNA microarray study relating to transcriptional profiles of blood cells from multiple sclerosis (MS) patients—2 individuals treated with interferon-β (IFN-β) and 4 untreated control. A total of 130 samples were analyzed. One key feature of this study is that repeated peripheral blood samples were taken from the same subject irregularly (for details, see Section 4.1), across approximately 7 days, leading to within-subject correlation among measurements from the same individual in the data set. In addition, missing values were common in the 2-color spotted microarray, resulting in unbalanced longitudinal data.

Mixed effects models are often invoked to analyze such data (Laird and Ware, 1982; Diggle and others, 1994; Wang and Kim, 2003). Although the covariance matrix defined in most statistical software can be very general, these procedures have not incorporated the within-subject correlation into consideration. Recently, several other models have been proposed (Storey and others, 2005; Hong and Li, 2006; Yuan and Kendziorski, 2006). However, the random components are not generally modeled. For example, the pioneer work of extraction of differential gene expression (EDGE) by Storey and others (2005) assumed a subject-dependent but time-independent random mean shift.

In the meanwhile, several statistical methods have been developed along the line of decomposing the covariance function to better reflect sources of random variation in the time dimension through functional data analysis (Rice and Wu, 2000; James and others, 2001; James and Sugar, 2003; and many others). More recently, Yao and others (2005) proposed a nonparametric method to perform functional principal analysis through conditional expectation (PACE) for sparse/irregular functional data. One specific goal of these methods is to better estimate the mean curve and the covariance function and predict individual trajectories under a single experiment condition.

In this paper, our goal was to integrate these functional covariance methods into testing for changes in temporal gene expression profiles under different experimental conditions. Specifically, we carry out a functional principal component analysis (FPCA) to predict each replicate’s expression trajectory, respectively, under the null and the alternative hypotheses. An $F$-type statistic is subsequently constructed, and a permutation test is conducted to obtain a $p$ value for each gene. One distinguishing feature of our approach is that, by estimating each individual replicate expression profile more accurately, it is more sensitive in detecting temporally differentially expressed genes. The performance of the proposed method in both simulations and real data demonstrates that in many situations, it is more powerful than the EDGE and a 2-way mixed effects analysis of variance (ANOVA).

2. Significance analysis of time course gene expressions

2.1 The PACE model and its estimates

Since time course gene expression data should change smoothly with respect to time, we adapt functional data analysis tools PACE to analyze such data because it is conceptually simple and straightforward to implement.

Consider a square integrable stochastic process $X(t)$ with an unknown mean function $\mathbb{E}(X(t)) = \mu(t)$, $t \in [0, T]$, $T > 0$, and an unknown covariance function $\text{cov}(X(s), X(t)) = G(s, t)$, $(s, t) \in [0, T] \times [0, T]$. Let $X_i(t)$, $i = 1, 2, \ldots, m$, be $m$ realizations of $X(t)$. Further, let the $j$th observation $X_{ij}(t)$ of the $i$th subject observed at a time $T_{ij}$ be $Y_{ij}$ for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, N_i$. To reflect the fact that data are sparse, the PACE model incorporated uncorrelated additive measurement errors into model as follows:

$$Y_{ij} = X_i(T_{ij}) + \epsilon_{ij} = \mu(T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(T_{ij}) + \epsilon_{ij}, \quad T_{ij} \in [0, T],$$

(2.1)
Identifying differentially expressed genes through FPCA

where \( \epsilon_{ij} \) are i.i.d. measurement errors independent of \( X(t) \) and \( \phi_k(t) \) and \( \lambda_k \), respectively, are the eigenfunctions and eigenvalues of \( G(s, t) \) through Karhunen–Loève expansion (for details, see chapter 6 of Ramsay and Silverman, 1997). This model, in some sense, can be cast as Laird and Ware (1982) random effects model or the Diggle and Al Wasel (1997) random effects model in spectral analysis, except that all relevant unknown quantities are data driven and estimated nonparametrically.

Model (2.1) has a simple biological interpretation. It is conceivable that a gene’s function is governed by many factors (\( \phi_k(t) \)), such as the hormone level or the circadian rhythm, but their influence on the gene expression level varies by individuals, that is the strength of the \( k \)th influential factor (\( \phi_k(t) \)) on individual \( i \) is a random variable (\( \hat{\xi}_{ik} \)), usually called the functional principal component score. Note that each \( \hat{\xi}_{ik} \) is a fixed value for any given individual. Since the overall mean has been described in \( \mu(t) \), \( \hat{\xi}_{ik} \) should have mean 0. How many curves (the bound of \( K \)) are necessary depends on the time complexity. Usually, the process can be well approximated by the first few, say \( K \), eigenfunctions with the first \( K \) largest eigenvalues. The number \( K \) can be chosen by Bayesian information criterion or the fraction of total variance explained.

Quantities to be estimated in (2.1) include the unknown mean function \( \mu(t) \), the eigenfunctions and eigenvalues of the covariance function \( G \), and the functional principal component scores \( \hat{\xi}_{ik} \). Denote by \( Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^T \), \( \mu_i = (\mu(T_{i1}), \mu(T_{i2}), \ldots, \mu(T_{iN_i}))^T \), and \( \phi_{ik} = (\phi_k(T_{i1}), \phi_k(T_{i2}), \ldots, \phi_k(T_{iN_i}))^T \), the estimates of functional principal component scores are then given by

\[
\hat{\xi}_{ik} = \hat{E}[\xi_{ik}|Y_i] = \hat{\lambda}_k \hat{\phi}_{ik}^T \hat{\Sigma}_{Y_i}^{-1}(Y_i - \hat{\mu}_i),
\]

where \( \hat{\mu}_i \), \( \hat{\Sigma}_{Y_i} \), and \( \hat{\Sigma}_{Y_i} = \text{cov}(Y_i, Y_i) \) (the \( N_i \times N_i \) matrix of subject \( i \)). For technical details, see Yao and others (2005). The predicted trajectory for subject \( i \) for a prechosen \( K \) is then given by

\[
\hat{X}_i(t) = \hat{\mu}(t) + \sum_{k=1}^{K} \hat{\xi}_{ik} \hat{\phi}_k(t).
\]

In the context of our analysis, the dynamic activity of each gene is viewed as a stochastic process and each replicate is thus a realization of this process.

### 2.2 Functional hypothesis test

Now that we have constructed a functional principal components framework for modeling time course gene expression data, we proceed to formulate functional hypothesis testing to identify temporally differentially expressed genes. To simplify our presentation, we will focus on searching for temporally differentially expressed genes under 2 experimental conditions, for example a treatment versus control design. Unless otherwise noted, we denote the treatment group as Trt and the control group as Ctr. We assume there are a total number of \( G \) genes on each array slide and let there be \( n_l \) replicates in the \( l \)th group, \( l = \{\text{Trt}, \text{Ctr}\} \). In this paper, we only consider separate analyses of individual genes and drop the gene index accordingly.

For 1 gene, let \( Y_{ilij} \) be the expression level of the \( i \)th replicate in the \( l \)th group observed at time point \( T_{ilij} \), where \( l = \{\text{Trt}, \text{Ctr}\} \), \( i = 1, 2, \ldots, n_l \), \( j = 1, 2, \ldots, N_i \). Following the PACE model, the approximate trajectory using the first \( K \) eigenfunctions for the \( i \)th replicate in the \( l \)th group can be predicted as:

\[
\hat{X}_{i,l}(t) = \hat{\mu}_l(t) + \sum_{k=1}^{K} \hat{\xi}_{i,ik} \hat{\phi}_{l,k}(t).
\]
To identify differentially expressed genes over time, the functional hypothesis testing can be formulated in terms of the mean expression curves as follows:

\[ H_0: \mu_{\text{Ctr}}(t) = \mu_{\text{Trt}}(t) = \mu(t), \quad t \in [0, T] \quad \text{versus} \quad H_1: \mu_{\text{Ctr}}(t) \neq \mu_{\text{Trt}}(t), \quad \text{for some} \quad t \in [0, T]. \]

Under \( H_0 \) of no differential expression, all subjects are assumed to have come from the same stochastic process and thus they share the same mean curve \( \mu(t) \) and a common dependence structure. Therefore, all subjects from the 2 groups are pooled together to estimate the mean curve \( \hat{\mu}(t) \) and the covariance surface function \( G(s, t) \). The individual trajectory is then predicted based on these estimations. We denote the estimated overall mean curve by \( \hat{\mu}(t) \) and the predicted trajectory for the \( i \)th subject by \( \hat{X}_i^0(t) \) (as in (2.3)). Under \( H_1 \) in the presence of differential expression, we conduct the PACE analysis separately for the 2 groups. In this context, we denote the predicted mean gene expression curve for the \( l \)th group by \( \hat{\mu}_l(t) \), \( l = \{\text{Trt, Ctr}\} \). The \( i \)th replicate in the corresponding group can be predicted by \( \hat{X}_i^1(t) \), as in (2.4).

### 2.3 Model inference

Based on the predicted individual trajectories, we define the residual to be the difference between the observed values and the fitted values. It is well known that the residual sum of squares (RSS) is a reasonable measure of the goodness of fit for a given model. Therefore, the RSS under the null and the alternative models can be represented by \( \text{RSS}^0 = \sum_i \int (Y_i(t) - \hat{X}_i^0(t))^2 dt \) and \( \text{RSS}^1 = \sum_i \int (Y_i(t) - \hat{X}_i^1(t))^2 dt \), respectively. In practice, the integral is replaced by summation since observed data are on some discrete points. The test statistic is defined to be an analogue of the \( F \)-statistic (following the idea of Storey and others, 2005) as follows:

\[
F = \frac{\text{RSS}^0 - \text{RSS}^1}{\text{RSS}^1},
\]

where the statistic \( F \) measures the relative variance reduced by the alternative model.

There is a sizable literature on its asymptotic distribution for (2.5), for example Shen and Faraway (2004) and references therein. Such methods are readily applicable for classical functional data where the number of subjects are moderate or large and the time points are dense. However, their applicability is not easy to assess when the sample size (the number of subjects) is small, which is often the case in time course microarray experiments.

For small sample size, we propose a modified fitting procedure under an additional assumption. We assume that the principal component curves \( \phi_k(t) \) are not affected by the experimental conditions. Under this assumption, we can estimate the covariance structure of the stochastic process by pooling data from both groups. This practice is not uncommon in modeling second-order stochastic processes representing 2 groups of functional data, for instance, Müller and others (2006) and references therein. Under this assumption, the subjects labels are exchangeable under the null hypothesis. Therefore, we can use the permutation test for formal statistical decisions. The \( p \) value for gene \( i \) is then calculated in the usual way by letting

\[
p_i = \frac{\sum_{b=1}^B I\{F_b \geq F_i\}}{B},
\]

where \( B \) is the maximum number of permuted samples, \( I\{\cdot\} \) is an indicator function, \( F_b \) is the null statistic from the \( b \)th permutation, and \( F_i \) is the observed test statistic constructed from the original data for gene \( i \).
In contrast, Storey and others (2005) defined the $p$ value for gene $i$ by considering the null statistics from all genes:

$$p_i = \frac{\sum_{b=1}^{B} \# \{ F_{jb} \geq F_i, j = 1, \ldots, G \}}{G \times B},$$

where $G$ is the number of genes. This definition is rather liberal (see details in supporting information: supporting appendix of Storey and others, 2005, p. 17) because the distribution of the pooled $p$ values often depends on the magnitudes of the observed statistics. Our simulation results implied this fact too (see Table 1 for details).

When we apply the proposed procedure to obtain a significance measure for each gene, it is necessary to consider $p$ value adjustment because thousands of hypotheses are tested simultaneously. A commonly used strategy is to control the false discovery rate (FDR), as illustrated in Benjamini and Hochberg (1995), Storey and Tibshirani (2003), and Storey and others (2005). We adopt the one proposed in Benjamini and Hochberg (1995) since the concept is very intuitive and simple to compute.

3. Simulation studies

To assess the performance of the proposed method and to compare it with results from the EDGE and a 2-way mixed effects ANOVA model, we carried out analyses on simulated data sets under 2 scenarios: One falls in the proposed FPCA model framework, while the other a mixed effects ANOVA model.

3.1 Simulation based on the FPCA model

The time course expression profiles were generated according to (2.1). To simplify presentation, we took $\mu(t) = \alpha + \beta t$. We denote by $\alpha_1, \beta_1, \alpha_2, \beta_2$ the corresponding $\alpha, \beta$ in groups 1 and 2, respectively. Table 1 summarizes the 15 parameter settings (with total 45 cases when combined with 3 different eigenfunctions) used to generate data, their hypothesis types, and the percentages of correct conclusions from ANOVA, EDGE, and FPCA.

The eigenfunctions $\phi_k(t)$ were assumed to be the same for the 2 groups to mimic possible rhythms in a biological process. We used only 1 eigenfunction $\phi_1(t)$, which can take 1 of 3 forms, $-\sqrt{\frac{2}{T}} \cos\left(\frac{2\pi t}{T}\right)$ (single cycle), $-\sqrt{\frac{2}{T}} \cos\left(\frac{4\pi t}{T}\right)$ (double cycle), or $\frac{1}{\sqrt{T}}$ (corresponding to the case that an individual deviation from the mean curve is a random shift, as in the EDGE model), where $t \in [0, T]$, $T = 10$. There were 50 genes in each case. Thus, we had a total of 2250 genes of which 600 were under $H_0$ and 1650 were under $H_1$. In each group, there were 10 subjects. The number of time points $N_i$ at which individual $i$ were observed were generated uniformly on $\{4, 5, \ldots, 10\}$, and the points $T_{ij}, j = 1, \ldots, N_i$ were generated uniformly on $[0, T]$. Further, the coefficient (principal score) $\xi_{i1}$ of $\phi_1(t)$ for subject $i$ was modeled as normal random variable with mean 0 and standard deviation 0.5. The additive errors $\epsilon_{ij}$ were generated from $N(0, \sigma_0^2)$ with $\sigma_0 = 0.1$.

We applied the proposed procedure FPCA as described in Section 2.3, EDGE, using the settings for analyzing time course microarray data, with 4 polynomial basis functions and a 2-way mixed effects ANOVA model with subject-specific random effect to the simulated data. In the ANOVA model, the $p$ value of the time by group interaction term is the indicator of whether the gene is temporally differentially expressed. The claim of $H_1$ was set at $p$ value less than or equal to 0.01 for all 3 methods.

Table 2 summarizes the results in terms of false positives, false negatives, and FDRs. Both the proposed FPCA and EDGE performed better than the ANOVA model, but the proposed FPCA method is the most powerful method in correctly identifying the significant genes. Possible reasons for the performance
<table>
<thead>
<tr>
<th>Case/H0 or H1</th>
<th>((\alpha_1, \alpha_2))</th>
<th>((\beta_1, \beta_2))</th>
<th>ANOVA</th>
<th>EDGE</th>
<th>FPCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1s/H0</td>
<td>(0, 0)</td>
<td>(0, 0)</td>
<td>100</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td>1d/H0</td>
<td>(0, 0)</td>
<td>(0, 0)</td>
<td>100</td>
<td>88</td>
<td>98</td>
</tr>
<tr>
<td>1f/H0</td>
<td>(0, 0)</td>
<td>(0, 0)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2s/H0</td>
<td>(0, 0)</td>
<td>(0.1, 0.1)</td>
<td>100</td>
<td>94</td>
<td>100</td>
</tr>
<tr>
<td>2d/H0</td>
<td>(0, 0)</td>
<td>(0.1, 0.1)</td>
<td>100</td>
<td>92</td>
<td>100</td>
</tr>
<tr>
<td>2f/H0</td>
<td>(0, 0)</td>
<td>(0.1, 0.1)</td>
<td>98</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>3s/H0</td>
<td>(0, 0)</td>
<td>(0.2, 0.2)</td>
<td>100</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>3d/H0</td>
<td>(0, 0)</td>
<td>(0.2, 0.2)</td>
<td>100</td>
<td>88</td>
<td>98</td>
</tr>
<tr>
<td>3f/H0</td>
<td>(0, 0)</td>
<td>(0.2, 0.2)</td>
<td>98</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>4s/H0</td>
<td>(0, 0)</td>
<td>(0.3, 0.3)</td>
<td>98</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td>4d/H0</td>
<td>(0, 0)</td>
<td>(0.3, 0.3)</td>
<td>100</td>
<td>76</td>
<td>100</td>
</tr>
<tr>
<td>4f/H0</td>
<td>(0, 0)</td>
<td>(0.3, 0.3)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>5s/H1</td>
<td>(0, 0.1)</td>
<td>(0, 0)</td>
<td>0</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>5d/H1</td>
<td>(0, 0.1)</td>
<td>(0, 0)</td>
<td>0</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>5f/H1</td>
<td>(0, 0.1)</td>
<td>(0, 0)</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6s/H1</td>
<td>(0, 0.2)</td>
<td>(0, 0)</td>
<td>2</td>
<td>12</td>
<td>86</td>
</tr>
<tr>
<td>6d/H1</td>
<td>(0, 0.2)</td>
<td>(0, 0)</td>
<td>0</td>
<td>10</td>
<td>62</td>
</tr>
<tr>
<td>6f/H1</td>
<td>(0, 0.2)</td>
<td>(0, 0)</td>
<td>2</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>7s/H1</td>
<td>(0, 0.5)</td>
<td>(0, 0)</td>
<td>0</td>
<td>18</td>
<td>100</td>
</tr>
<tr>
<td>7d/H1</td>
<td>(0, 0.5)</td>
<td>(0, 0)</td>
<td>0</td>
<td>10</td>
<td>94</td>
</tr>
<tr>
<td>7f/H1</td>
<td>(0, 0.5)</td>
<td>(0, 0)</td>
<td>0</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>8s/H1</td>
<td>(0, 0)</td>
<td>(0.1, 0)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>8d/H1</td>
<td>(0, 0)</td>
<td>(0.1, 0)</td>
<td>100</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>8f/H1</td>
<td>(0, 0)</td>
<td>(0.1, 0)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>9s/H1</td>
<td>(0, 0)</td>
<td>(0.2)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>9d/H1</td>
<td>(0, 0)</td>
<td>(0.2)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>9f/H1</td>
<td>(0, 0)</td>
<td>(0.2)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10s/H1</td>
<td>(0, 0)</td>
<td>(0.1, 0.1)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10d/H1</td>
<td>(0, 0)</td>
<td>(0.1, 0.1)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10f/H1</td>
<td>(0, 0)</td>
<td>(0.1, 0.1)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>11s/H1</td>
<td>(0, 0)</td>
<td>(0.1, 0.3)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>11d/H1</td>
<td>(0, 0)</td>
<td>(0.1, 0.3)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>11f/H1</td>
<td>(0, 0)</td>
<td>(0.1, 0.3)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>12s/H1</td>
<td>(0, 0.1)</td>
<td>(0.1, 0)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>12d/H1</td>
<td>(0, 0.1)</td>
<td>(0.1, 0)</td>
<td>100</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>12f/H1</td>
<td>(0, 0.1)</td>
<td>(0.1, 0)</td>
<td>100</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>13s/H1</td>
<td>(0, 0.1)</td>
<td>(0.1, 0.1)</td>
<td>0</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>13d/H1</td>
<td>(0, 0.1)</td>
<td>(0.1, 0.1)</td>
<td>0</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>13f/H1</td>
<td>(0, 0.1)</td>
<td>(0.1, 0.1)</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>14s/H1</td>
<td>(0, 0.2)</td>
<td>(0.2, 0)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>14d/H1</td>
<td>(0, 0.2)</td>
<td>(0.2, 0)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>14f/H1</td>
<td>(0, 0.2)</td>
<td>(0.2, 0)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>15s/H1</td>
<td>(0, 0.1)</td>
<td>(0.2, 0.2)</td>
<td>2</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>15d/H1</td>
<td>(0, 0.1)</td>
<td>(0.2, 0.2)</td>
<td>2</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>15f/H1</td>
<td>(0, 0.1)</td>
<td>(0.2, 0.2)</td>
<td>0</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>16s/H1</td>
<td>(0, 0.1)</td>
<td>(0.2, 0.2)</td>
<td>0</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

s, single-cycle sinusoidal; d, double-cycle sinusoidal; f, flat; ANOVA, % detected correctly by ANOVA; EDGE, % detected correctly by EDGE; FPCA, % detected correctly by FPCA.
Table 2. Comparison of ANOVA, EDGE, and FPCA using simulated data

<table>
<thead>
<tr>
<th></th>
<th>ANOVA</th>
<th>EDGE</th>
<th>FPCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total $H_0$ claimed</td>
<td>1343</td>
<td>1232</td>
<td>1117</td>
</tr>
<tr>
<td>Total $H_1$ claimed</td>
<td>907</td>
<td>1018</td>
<td>1133</td>
</tr>
<tr>
<td>$H_1$ claimed as $H_0$ (false negatives)</td>
<td>746</td>
<td>687</td>
<td>520</td>
</tr>
<tr>
<td>$H_0$ claimed as $H_1$ (false positives)</td>
<td>3</td>
<td>55</td>
<td>3</td>
</tr>
<tr>
<td>Sensitivity (proportion of $H_1$ claimed as $H_1$) (%)</td>
<td>54.8</td>
<td>58.4</td>
<td>68.5</td>
</tr>
<tr>
<td>FDR</td>
<td>0.025</td>
<td>0.022</td>
<td>0.019</td>
</tr>
</tbody>
</table>

Fig. 1. Simulated expression levels of group 1 data (circles), group 2 data (x-marks), estimated mean expression curves (dotted curves for group 1 and dashed curves for group 2), and the first eigenfunctions (solid curves). The $x$-axis is time and the $y$-axis is expression intensity. The bandwidths used to obtain the smoothed mean curves and covariance surfaces are 2.0 and 2.5, respectively. These 4 subjects (starting at the top left panel then moving to the right) were generated based on parameters given in rows 6s, 7d, 13s, and 7f, respectively, in Table 1. It is clear that the eigenfunctions used to generate them were recovered. Note that the real magnitudes of the eigenfunctions in the observation are modified by a scalar factor $\xi_{ik}$ in (2.3).

In this study, there were 167 genes that were correctly identified as differentially expressed genes by our method, but not by EDGE. Figure 1 depicts 4 such examples. These 4 genes correspond to cases generated by rows 6s, 7d, 13s, and 7f, respectively, in Table 1. It is immediately clear that the proposed algorithm is able to recover both the mean and the eigenfunction very well. It was very surprising to see that the case corresponding to the gene in the lower right panel was missed by EDGE 100% of the time. We believe the reason might be due to the ways they treat the random shift $\gamma$ for each subject (see details in supporting information: supporting appendix of Storey and others, 2005, p. 8).
surprising that the new method is more powerful because the functional principal scores have been proved to be the best linear consistent predictors (see discussion in Yao and others, 2005, pp. 9 and 17). We also examine the percentages that were claimed correctly for the \( H_0 \) cases (the first 12 rows in Table 1). It is immediately seen that both ANOVA and FPCA are able to correctly identify almost all these \( H_0 \)s, agreeing well with the significance level \( \alpha = 0.01 \). For EDGE, it can only correctly identify those cases with the flat eigenfunction (the cases that EDGE model is aimed for). There are considerably higher percentages of misidentifications in other cases, with the worst case of 24% in 4d. In all, 110 genes of the 600 null (about 18.3%) were declared to be nonnull at the significance level of 0.01. The results confirm that EDGE is on the liberal side.

### 3.2 Simulation based on a mixed effects ANOVA model

We also conducted simulation based on a mixed effects ANOVA model where temporal gene expression data were generated based on a subset of the real data from the transcriptional profiles of blood cells presented in Section 4.

First, we chose 2500 “homogeneous” genes from the real data where homogeneous is in the sense that none of them is differentially expressed according to both FPCA and EDGE. The group mean curve for each gene, \( \mu_0(t) \), was nonparametrically estimated based on all 18 replicates profiles (6 from the treatment group and 12 from the control group). For each replicate, the gene expression profile was generated at time points 0, 3.5, 6.25, 7.25, 9.5, 10, 25, 11.5, 13, 15.5, 16.5, 25, 33, 41, and 49 h based on a linear mixed effects model. Specifically, for the \( i \)th replicate measured at the \( j \)th point \( t_{ij} \), the simulated gene expression profile can be expressed as

\[
Y_{ij} = \mu_0(t_{ij}) + \alpha + \beta t_{ij} + a + b t_{ij} + \epsilon_{ij},
\]

where \( a, b \) are the random intercept and slope distributed as uniform random variables on \([-0.2, 0.2] \); \( \alpha, \beta \) are, respectively, the intercept and slope for the fixed linear effects, although \( \alpha \) would not affect the results because the \( \mu_0(t) \) did not have an explicit parametric form; and \( \epsilon_{ij} \) are additive normal random measurement errors with mean 0 and variance \( \sigma^2 \). To vary the degree of differential gene expressions, 2 values of the error variance are considered: \( \sigma^2 = 0.1 \) and \( \sigma^2 = 0.5 \).

Again, we denote, respectively, by \( \alpha_1, \beta_1, \alpha_2, \beta_2 \) the corresponding groups 1 and 2 parameter values of \( \alpha, \beta \). We simulated 10%, that is 250, of the 2500 genes to be time differentially expressed where the intercepts and slopes have 5 settings with 50 genes in each setting. For the remaining 2250 nondifferentially expressed genes, the 2 groups have the same intercepts and slopes ascribed also to 5 settings with 450 genes in each setting. Table 3 presents the 10 parameter settings used to generate data and their hypothesis types.

<table>
<thead>
<tr>
<th>Row ID/H0 or H1</th>
<th>( (a_1, a_2) )</th>
<th>( (\beta_1, \beta_2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/H1</td>
<td>(0, 0.04)</td>
<td>(0, 0)</td>
</tr>
<tr>
<td>2/H1</td>
<td>(0, 0.10)</td>
<td>(0, 0)</td>
</tr>
<tr>
<td>3/H1</td>
<td>(0, 0)</td>
<td>(0.02, 0.06)</td>
</tr>
<tr>
<td>4/H1</td>
<td>(0, 0.04)</td>
<td>(0.04, 0)</td>
</tr>
<tr>
<td>5/H1</td>
<td>(0, 0.02)</td>
<td>(0.02, 0)</td>
</tr>
<tr>
<td>6/H0</td>
<td>(0, 0)</td>
<td>(0, 0)</td>
</tr>
<tr>
<td>7/H0</td>
<td>(0, 0)</td>
<td>(0.02, 0.02)</td>
</tr>
<tr>
<td>8/H0</td>
<td>(0, 0)</td>
<td>(0.04, 0.04)</td>
</tr>
<tr>
<td>9/H0</td>
<td>(0, 0)</td>
<td>(0.06, 0.06)</td>
</tr>
<tr>
<td>10/H0</td>
<td>(0, 0)</td>
<td>(0.1, 0.1)</td>
</tr>
</tbody>
</table>
When the 3 methods were applied to the simulated data, the claim of $H_1$ was set at $p$ value less than or equal to 0.01 or 0.002. Tables 4 and 5 summarize the results of the comparison. Our findings show that none of the 3 methods is clearly superior to the others. Due to the simple structure of the covariance function, we expect that FPCA has no advantage, although one would expect that the ANOVA mixed effects model should have performed better.

4. TRANSCRIPTOMICAL PROFILES OF BLOOD CELLS DATA AND RESULTS

4.1 Data description

MS is an inflammatory, chronic, and demyelinating disease that affects the central nervous system. IFN-β, a member from a family of cytokines that mediate a range of diverse functions including antiviral, antiproliferative, antitumor, and immunomodulatory activities (Stark and others, 1998), is currently the most prescribed treatment regimen for controlling exacerbations in MS patients (Fernald and others, 2007). However, IFN-β’s mechanism of action at the genomic level in humans has not been fully delineated.

Fernald and others (2007) used DNA microarrays to analyze the transcriptional profile of blood cells through the application of time course microarray data. In addition to its time course nature, the study was a treatment versus control design. Specifically, repeated peripheral blood draws were conducted in 2 IFN-β-treated MS patients, whereas samples from 4 healthy controls (untreated) matched for age and gender were also collected accordingly. For 1 treated individual, triplicate samples were obtained at 0, 3.5, 6.25, 9.5, 11.5, 16.5, 25, 49, and 156 h. Since IFN-β stimulated gene expression levels remain steady at later time points (25 h postinfection), the main interest from a biological point of view, therefore, lies in the higher detail in the early part of the curves. Because of this, it would not hurt to treat the last 3 time points as 25, 41, and 49 h, which led to well-defined shapes of the gene expression profiles. For the other treated...
individual, triplicate samples were collected at 7.25, 10.25, 13, 15.5, and 33 h. The collection times for
the 4 healthy controls followed those of the 2 treated individuals, with 2 controls following 1 time series
each. Triplicate samples were also obtained at each time point for 4 controls (except that for control 3 at
time point 13, there was only 1 sample). A total of 130 samples were analyzed. The RNA purification,
fluorescent labeling, microarray hybridization, and other technical details were described in Fernald and
others (2007). The Lowess-normalized $\log_2$ ratios data are publicly available at Gene Expression Omnibus

Each microarray contained over 22 000 spotted 70-mer oligonucleotide probes. Because the study was
a 2-color spotted microarray experiment, missing values were not uncommon due to dust or scratches on
the slide, spotting problems, or hybridization problems. We here used a less rigorous approach to deal
with missing values by simply throwing away probes with 10% or more missings. We then further filtered
the data by removing genes with variance in the lowest 10% across all samples because they showed little
variation over time and treatment. This resulted in a subset of 8206 genes for each microarray slide. These
8206 genes were used to illustrate our method.

4.2 Results of differentially expressed genes

In our analysis, we choose $K$, the number of eigenfunctions, to be the smallest integer that explains at least
85% of total variation. We applied model fitting and inference procedure in Section 2.3 and identified a
total number of 4177 genes that are IFN-β stimulated with a FDR of 0.02.

It is instructive to have a look at some examples of temporally differentially expressed genes. Figure 2
shows 3 such cases. In these plots, replicates from each group display different expression patterns over
the time period examined. The first 2 eigenfunctions show how each replicate can deviate from the mean
structure over time, with the coefficient $\xi_{ij}$ indicating the individual strength along the direction of the
$j$th eigenfunction. Furthermore, the similarity of the second eigenfunction shows that our algorithm is not
an artificial extra level of complexity in modeling sources of random variation. Our results at the mean
level demonstrate that the IFN-β treatment induces time-dependent changes in individual immunomodu-
latory gene expression profiles. Plots in Figure 2 may provide a deeper understanding of the underlying
regulatory networks at play.

4.3 Comparisons with EDGE and ANOVA

Of interest is a comparison of the proposed FPCA model with the EDGE method and with the 2-way
mixed effects ANOVA model similar to that we did in the simulation study. In summary, applying the
EDGE analysis to the data results in a total number of 1881 genes that are IFN-β stimulated with a FDR
of 0.044. The difference in the FDR and the identified genes seem to indicate that FPCA is more efficient.
For example, none of the 3 differentially expressed genes displayed in Figure 2 has been detected by
EDGE. Small sample size may also be a factor contributing to the big difference between the 2 results.

Using the 2-way mixed effects ANOVA model to the same data set only identified 358 genes that
changed expressions in the time course with a FDR of 0.23. Such performance of ANOVA in this application
is not too surprising. One possible reason may be that the 2-way ANOVA with repeated measurements
uses every available degree of freedom and, consequently, the time structure is ignored. As pointed out
in Storey and others (2005), “As long as a model using less degrees of freedom appropriately captures
the signal cross time, then standard statistical theory says that this model will be more powerful.” This
explains why both EDGE and FPCA are more powerful in revealing more number of differentially genes
at much lower FDRs. The findings in this case study further confirm the effectiveness of the proposed
FPCA model.
Identifying differentially expressed genes through FPCA

Fig. 2. Top panels: Smooth estimates of the first 2 eigenfunctions (solid the first and dashed the second) for 3 selected temporally differentially expressed genes. Bottom panels: Observed treated individuals expression levels (circles), untreated individuals expression levels (x marks), and estimated mean expression curves (dotted curves for treatment group and dashed curves for control group). The x-axis is time in hours and the y-axis is the log-transformed gene expression intensity. The symbols from the annotation platform GPL4228 for these 3 genes are SSBP1, LSM4, and COX5AP1 (moving from the left to the right).

5. DISCUSSION

Technological progress in genome-wide measurements results in large amounts of time course gene expression data necessary to discover the dynamic activities of cells. We have proposed a method for detecting temporally differentially expressed genes by integrating a functional principal components method for sparse longitudinal data (Yao and others, 2005) into a hypothesis test model framework. The method is illustrated by simulation study and an illustration to transcriptional profiles of blood cells microarray data. Both findings showed that, in many situations, our method is more powerful than the EDGE model (Storey and others, 2005) and a 2-way mixed effects ANOVA model.

We did not investigate extensively the issue of bandwidth selection for smoothing the covariance surfaces and the mean curves. Data-driven methods, such as cross-validation and generalized cross-validation, can be used to select the bandwidth. However, numerical and application results were robust over a range of bandwidths, that is overall shapes of the mean and eigenfunctions are almost invariant for a range of bandwidth values (Liu and Müller, 2003).

In our analyses, we have assumed that genes are independent, although this assumption is often not valid since it is well known that genes frequently interact. As a future research direction, we will consider gene interactions and how the interactions among genes affect the gene expression trajectory over time.
to improve the understanding of genome-wide dependence structure. A comparison of the $\phi_k(t)$ between genes may lead to discoveries in gene interactions.

As a final comment, this is a generic method in the sense that it may be applied to other types of time course data where time dependence occurs, such as the mass spectrometry data in cancer research. We expect to see more applications of the proposed method in the future.

ACKNOWLEDGMENTS

We would like to thank the Editor, the co-Editor, and the referee for many valuable and informative comments, which have significantly improved the paper. Conflict of Interest: None declared.

FUNDING

Taiwan National Science Council (NSC 98-2811-M-005-001) to M.Y.

REFERENCES


[Received June 6, 2008; revised May 27, 2009; accepted for publication June 22, 2009]