BIOINFORMATICS

Vol. 19 no. 6 2003, pages 694–703
DOI: 10.1093/bioinformatics/btg068

Statistical tests for identifying differentially expressed genes in time-course microarray experiments

Taesung Park [1,], Sung-Gon Yi [1], Seungmook Lee [1], Seung Yeoun Lee [2], Dong-Hyun Yoo [3], Jun-Ik Ahn [3] and Yong-Sung Lee [3]*

[1]Department of Statistics, Seoul National University, Seoul, Korea, [2]Department of Applied Mathematics, Sejong University, Seoul, Korea and [3]Department of Biochemistry, Hanyang University College of Medicine, Seoul, Korea

Received on July 17, 2002; revised on September 7, 2002; accepted on December 11, 2002

ABSTRACT

Motivation: Microarray technology allows the monitoring of expression levels for thousands of genes simultaneously. In time-course experiments in which gene expression is monitored over time, we are interested in testing gene expression profiles for different experimental groups. However, no sophisticated analytic methods have yet been proposed to handle time-course experiment data.

Results: We propose a statistical test procedure based on the ANOVA model to identify genes that have different gene expression profiles among experimental groups in time-course experiments. Especially, we propose a permutation test which does not require the normality assumption. For this test, we use residuals from the ANOVA model only with time-effects. Using this test, we detect genes that have different gene expression profiles among experimental groups. The proposed model is illustrated using cDNA microarrays of 3840 genes obtained in an experiment to search for changes in gene expression profiles during neuronal differentiation of cortical stem cells.

Availability: A set of programs written by *R* will be electronically sent upon request.

Contact: tspark@stats.snu.ac.kr

1 INTRODUCTION

Biological processes depend on complex interactions between many genes and gene products. To understand the role of a single gene or gene product in this network, many different types of information, such as genome-wide knowledge of gene expression, will be needed. Microarray technology is a useful tool to understand gene regulation and interactions. For example, cDNA microarray technology allows the monitoring of expression levels for thousands of genes simultaneously. cDNA microarrays consist of thousands of individual DNA sequences printed in a high density array on a glass slide. After being reversetranscribed into cDNA and labelled using red (Cy5) and green (Cy3) fluorescent dyes, two target mRNA samples are hybridized with the arrayed DNA sequences or probes. Then, the relative abundance of these spotted DNA sequences can be measured. For each gene the data consists of two fluorescence intensity measurements (R, G), showing the expression level of the gene in the red and green labelled mRNA samples. The ratio of the fluorescence intensity for each spot represents the relative abundance of the corresponding

DNA sequence. cDNA microarray technology has important applications in pharmaceutical and clinical research. By comparing gene expression in normal and tumor tissues, for example, microarrays may be used to identify tumor-related genes and targets for therapeutic drugs (Alizadeh *et al.*, 2000).

In microarray experiments, the identification of differentially expressed genes is an important issue (Friddle *et al.*, 2000; Galitski *et al.*, 1999; Golub *et al.*, 1999; Spellman *et al.*, 1998). To identify groups of genes with similar or correlated expression profiles (Alizadeh *et al.*, 2000), many clustering techniques have been applied. However, clustering methods are rather primitive and exploratory. Furthermore, as the number of genes becomes large, the clustering methods may not provide clear group patterns.

On the other hand, statistical test procedures can be useful tools for identifying differentially expressed genes especially in multiple-slide experiments. For a single-slide experiment, Chen *et al.* (1997) proposed a method for choosing cut-offs to identify differentially expressed genes. Recently, (Newton *et al.*, 2001) considered a hier-

*To whom correspondence should be addressed.

archical Bayesian model in order to identify differentially expressed genes based on the posterior odds of change. Nowadays, however, the importance of replication in microarray experiments has been pointed out by many researchers, mainly for increasing the precision of estimated quantities and to provide information about the uncertainty of estimates (Kerr *et al.*, 2001; Lee *et al.*, 2000).

A time-course experiment is a special case of a multiple-slide experiment, in which transcript abundance is monitored over time. Recently, a number of methods have been suggested for the identi- fication of differentially expressed genes in multipleslide cDNA microarray experiments based on statistical models such as the analysis of variance (ANOVA) model and the mixed effects model (Kerr *et al.*, 2000; Wolfinger *et al.*, 2001).

In this paper, we propose a statistical procedure to identify genes that have different gene expression profiles in time-course cDNA microarray data. We propose tests based on the ANOVA model to differentiate genes that have high variability from ones that do not. We propose two types of tests. One is the usual F-test which requires the normality assumption. The other is the permutation test which does not need the normality assumption. For the permutation test, we use residuals from the model with only time-effects. Using these tests, we detect genes that have different gene expression profiles among experimental groups.

The proposed procedure is illustrated using cDNA microarrays of 3840 genes obtained in a cortical stem cells experiment. From a developing fetal rat brain, 3840 genes were immobilized on a glass chip, and fluorescence-labelled target cDNA from cortical stem cells were hybridized. In this experiment, there are 3840 genes in each slide, two experimental groups for comparison and six different time points. Also, all experiments were replicated three times. The main objective of analysis is to identify genes with significant changes between the two experimental groups after adjusting for time effects.

The paper is organized as follows. The proposed ANOVA models and test procedures are presented in Section 2. Several test statistics and the calculation of adjusted p-values are also discussed. Section 3 describes normalization issues and presents the analysis results. Finally, Section 4 summarizes the concluding remarks.

2 METHODS

Suppose there are I experimental groups denoted by $i(= 1, \ldots, I)$, K time points denoted by $k(= 1, \ldots, K)$, and L replications denoted by $l(= 1, \ldots, L)$. Assume that there are N genes in one slide. We consider four typical types of time-course experiments.

(1) Two experimental groups
   (a) Without replication: $I = 2$, $L = 1$
   (b) With L replications: $I = 2$, $L \geq 2$

(2) $I\,(> 2)$ experimental groups
   (a) Without replication: $I > 2$, $L = 1$
   (b) With L replications: $I > 2$, $L \geq 2$

2.1 Two experimental groups: $I = 2$

Let y_{ikln} be the logarithm of the ratios of red and green background-corrected intensities from group i, time k, replication l, and gene n. Consider the following two models:

$$M_1 : y_{ikln} = \mu_n + \alpha_{in} + \beta_{kn} + (\alpha\beta)_{ikn} + \epsilon_{ikln},$$
$$M_2 : y_{ikln} = \mu_n + \alpha_{in} + \beta_{kn} + \epsilon_{ikln},$$

where $i = 1, 2$, $k = 1, \ldots, K$, $l = 1, \ldots, L$, and $n = 1, \ldots, N$. The gene effects μ_n capture the overall mean intensity in fluorescent signals for genes across the arrays, groups, and time points. The α_{in} terms account for gene specific group effects representing overall differences between two groups. The β_{ikn} account for time effects that capture differences in the overall concentration of mRNA in the samples from the different time points. The terms $(\alpha\beta)_{ikn}$ account for the interaction effect between group and time representing the signal contribution due to the combination of group and time. Note that the interaction terms $(\alpha\beta)_{ikn}$ cannot be estimated when $L = 1$. ϵ_{ikln} represent error terms. For the F-test they are assumed to follow a normal distribution. For the permutation test, however, the normality assumption is not required.

The above model is a two-way ANOVA model with group and time as two main factors. ANOVA models are commonly used to compare treatment means for some responses. In our ANOVA models, we are interested in comparing the mean values of two experiment groups as well as six time sequences. In our experiment, 2×6 experimental conditions are generated by group and time sequences.

In model M_1, the effects of interest are the interactions between group and time points, $(\alpha\beta)_{ikn}$. For gene n these terms capture differences from overall averages that are attributable to the specific combination of a time point k and group i. If these interaction terms are not significant, the effects of interest are the group effect α_{in} in model M_2. For gene n these terms capture differences from overall group averages. Thus, the hypotheses of interests are as follows:

$$H_{01} : (\alpha\beta)_{ikn} = 0 \text{ for } M_1,$$
$$H_{02} : \alpha_{in} = 0 \text{ for } M_2.$$

Testing significance of these effects involves the calculation of F-statistics for each gene.

Table 1. ANOVA table of model M_1

Source	Sum of squares	Degrees of freedom	Mean square error	F-statistic
Group	SS_G	$I-1$	MS_G	F_G
Time	SS_T	$K-1$	MS_T	F_T
Group × time	$SS_{G\times T}$	$(I-1)(K-1)$	$MS_{G\times T}$	$F_{G\times T}$
Error	SS_E	$IK(L-1)$	MS_E	
Total	SST	$IKL-1$		

For model M_1 Table 1 gives the typical form of the ANOVA table for each gene, if there are no missing observations. Sum of squares are given by

$$SS_G = KL\sum_i(\overline{y}_{i\cdot\cdot n} - \overline{y}_{\cdots n})^2,$$
$$SS_T = IL\sum_k(\overline{y}_{\cdot k\cdot n} - \overline{y}_{\cdots n})^2$$
$$SS_{G\times T} = L\sum_i\sum_k(\overline{y}_{ik\cdot n} - \overline{y}_{i\cdot\cdot n} - \overline{y}_{\cdot k\cdot n} + \overline{y}_{\cdots n})^2$$
$$SS_E = \sum_i\sum_k\sum_l(y_{ikln} - \overline{y}_{ik\cdot n})^2,$$
$$SST = \sum_i\sum_k\sum_l(y_{ikln} - \overline{y}_{\cdots n})^2.$$

The mean square errors are given by dividing sum of squares by degrees of freedom. For example, $MS_G = SS_G/(I-1)$, $MS_T = SS_T/(K-1)$ and so forth.

We need to fit N models of M_1 and to compute F-statistics for $(\alpha\beta)_{ikn}$, $F_{G\times T} = MS_{G\times T}/MS_E$. If each log ratio of intensities y has a normal distribution, then the F-statistics follow an F-distribution with $(K-1, 2K(L-1))$ degrees of freedom. One other issue for testing concerns p-values. Since we focus on N tests simultaneously, we need to adjust p-values caused by multiple comparison. Although there have been alternative methods including controlling false discovery rates, we use the method of adjusting p-values (Storey and Tibshirani, 2001).

When the interaction terms $(\alpha\beta)_{ikn}$ in M_1 are not significant, we need to consider reduced model M_2. The ANOVA table for this model is similar to that of M_1 without the group and interaction term. We also need to fit N models of M_2 and compute F-statistics for α_{in}, defined by $F_G = MS_G/MS_E$. If each log ratio of intensities y has a normal distribution, then the F-statistics follow an F-distribution with $(K-1, \phi_E)$ degrees of freedom, where $\phi_E = 2KL - K - 1$.

What if the normality assumption does not hold? Following the approach of Dudoit *et al.* (2000) we apply the permutation test that does not require any distributional assumption. The main idea of the permutation test is to derive the distribution of F-statistics from all possible permutations of the given observations and then compute the p-value of the F-statistic for the observed data.

If the microarray experiment has two factors of interest, say, A and B, which are not time-dependent, then the permutation test can be performed by permuting all levels of A and B simultaneously. In our experiment, however, we have two factors, *Group* and *Time*. In order to test the group effect, the data can be permuted over all levels of *Group*. However, if we permute *Time* and *Group* together, it might be difficult to extract information about the group difference. If we fix time, then the test is only for group effect for a specific time point. Furthermore, there are only few observations available to permute, which does not provide a sufficient sample size to determine the significance of group effect.

The proposed approach to handling this problem is to use two-stage models. At the first stage, remove the time effect and then at the second stage focus on the group effect. This can be done by fitting the following model with time effects only and then using the residuals:

$$M_3^1 : y_{ikln} = \mu_n + \beta_{ikn} + \epsilon_{ikln}, \quad (1)$$
$$M_3^2 : y_{ikln} = \mu_n + \beta_{kn} + \epsilon_{ikln}, \quad (2)$$

When there appears to be a strong interaction effect between group and time, it would be better to fit model M_3^1 in which β_{ikn} allows a different time effect for experimental groups. Otherwise, fit model M_3^2 which assumes the same time effect for each group, given by β_{kn}. Based on the test results of F-statistics, the decision for whether to fit M_3^1 or M_3^2 can also be made.

Let τ_{ikln} be the corresponding residual defined by

$$\tau_{ikln} = y_{ikln} - \hat{y}_{ikln} = \begin{cases} y_{ikln} - (\hat{\mu}_n + \hat{\beta}_{ikln}) & \text{for } M_3^1 \\ y_{ikln} - (\hat{\mu}_n + \hat{\beta}_{kn}) & \text{for } M_3^2 \end{cases} \quad (3)$$

The residuals do not have any information about time effect. They only have information about group effect. Thus, the permutation test can be performed using these residuals for testing group effects. More specifically, for gene n the KL residuals τ_{1kln} are from Group 1 and the other KL residuals τ_{2kln} are from Group 2, if there are not missing observations. For all $2KL$ residuals a two-sample permutation test can be performed using the following two-sample t-statistic,

$$T_g = \frac{\overline{\tau}_{1\cdot\cdot n} - \overline{\tau}_{2\cdot\cdot n}}{\sqrt{\frac{s_{1n}^2}{n_{1n}} + \frac{s_{2n}^2}{n_{2n}}}} \quad (4)$$

where n_{in} is the number of observations in Group i, and

$$\overline{\tau}_{i\cdot\cdot n} = \sum_{k=1}^{K}\sum_{l=1}^{L}\tau_{ikln}/n_{in},$$

$$s_{in}^2 = \frac{1}{n_{in}-1}\sum_{k=1}^{K}\sum_{l=1}^{L}(\tau_{ikln}-\overline{\tau}_{i\cdot\cdot n})^2$$

for $i = 1, 2$, $n = 1, \ldots, N$. By permuting all $2KL$ observations with KL observations in each group, the p-values for testing group effects can be obtained. In this case, there are N tests available. We need to adjust p-value by multiple comparison proposed by Westfall and Young (1993).

2.2 $I(>2)$ experimental groups

When I is larger than 2, the model development is the same as the case when $I = 2$. That is, we first need to fit models M_1 and M_2. The F-tests based on the normal assumption are exactly the same with those when $I = 2$, while the degrees of freedom differ. More specifically, we can use the same F-statistics given in the ANOVA table of Table 1. After fitting N ANOVA models, adjusted p-values can be obtained from the F-statistics in a similar manner.

For the permutation test, on the other hand, F-statistics based on the residuals in Equation (3) need to be used instead of two-sample t-test statistics in equation (4). For gene n, the KL residuals τ_{ikln} are from Group i, if there are no missing observations. For all IKL residuals a permutation test based on the F-statistics needs to be performed. That is, consider the following ANOVA model for residuals:

$$M_4 : \tau_{ikln} = \mu_n^R + \alpha_{in}^R + \epsilon_{ikln}, \qquad (5)$$

where $i = 1, \ldots, I$, $k = 1, \ldots, K$, $l = 1, \ldots, L$ and $n = 1, \ldots, N$. The superscript R is used to denote a model for residuals. The gene effects μ_n^R capture the overall mean intensity and α_{in}^R captures the ith group effect on residuals after removing the time effects. Thus, the hypothesis for testing group effect after removing time effects is given by

$$H_{03} : \alpha_{in}^R = 0 \quad \text{for all } i. \qquad (6)$$

For model M_4 Table 2 gives the typical form of the analysis of variance for each gene, if there are no missing observations. The F-statistic for α_{in}, defined by $F_G^R = MS_G^n/MS_E^R$, can be used for the permutation test. By permuting all IKL observations with KL observations in each group, the p-values for testing H_{03} can be approximately obtained by random permutation. We need to adjust p-values by multiple comparison proposed by Westfall and Young (1993).

Table 2. ANOVA table of model M_4 for the residuals

Source	Sum of squares	Degrees of freedom	Mean square error	F-statistic
Group	SS_G^R	$I-1$	MS_G^R	F_G^R
Error	SS_E^R	$I(KL-1)$	MS_E^R	
Total	SST	$IKL-1$		

3 RESULTS

3.1 Data

The data studied here are from a study of cortical stem rat cells. The goal of the experiment is to identify genes that are associated with neuronal differentiation of cortical stem cells. Although there have been many reports on the genes with changing their expression rates during neuronal differentiation, the thorough underlying mechanism of neuronal differentiation is not clear yet. Microarray slides were prepared using 3840 cDNA clones which were isolated from rat brain. Cortical neuronal stem cells were isolated from an E15 rat fetus and expanded under the presence of basic fibroblast growth factor (bFGF). After expansion, differentiation was induced by removing bFGF, and the cells were maintained for 12 h, 1, 2, 3, 4 and 5 days with or without ciliary neurotrophic factor (CNTF, 10 μg/ml). After extraction of total RNA at indicated time, reverse transcription was carried out using Cy 5-dUTP for fluorescence labelling and the expression patterns were compared to that of undifferentiated, expanded cortical stem cells, as a common reference, which was labelled with Cy 3. To get more reliable data, all the hybridization analyses were carried out three times against same RNA, and the scanned images were analyzed using an edge detection mode proposed by Kim *et al.* (2001).

In this experiment, there are 3840 genes in each slide, two experimental groups for comparison (No CNTF, CNTF), and six different time sequences (12 h, 1, 2, 3, 4 and 5 day). Since all experiments were replicated three times, all 36 slides were available for analysis. In this experiment, the reference design was used. That is, every sample of interest is hybridized to the same extraneous reference sample. The main objective of analysis is to identify genes with significant changes between two experimental groups after adjusting for time effects.

3.2 Normalization

Before applying the proposed models, we first describe normalization issues for handling spatial and intensity dependent effects on the measured expression levels. As pointed out by Yang *et al.* (2000), the purpose of normalization is to remove systematic variation in a

microarray experiment which affects the measured gene expression levels.

At a first step, we need to decide which set of genes to use for normalization. Yang *et al.* suggested three types of approaches: all genes on the array, constantly expressed genes, and controls. Recently, Tseng *et al.* (2001) suggested using the rank invariant genes. We tried rank invariant genes and then used all genes in the slide. However, the two approaches did not provide much different results. Thus, we decided to use all genes in the array.

Following the approaches of Yang *et al.* (2000), we applied global normalization using global median of log intensity ratios and intensity dependent non-linear normalization using a LOWESS curve. After applying these normalization methods to 36 slides, we found that the non-linear normalization method provided the most reasonable results.

3.3 Tests

Using the notation in the previous section, let y_{ikln} be the normalized log intensity ratios from group $i(=1,2)$, time $k(=1,\ldots,6)$, replication $l(=1,2,3)$, and gene $n(=1,\ldots,3840)$. Since $I=2$, we follow the test procedures for two experimental groups.

We first fit 3840 M_1 models and then perform the F-test for testing H_{01}. None of the genes were significant at the 5% significance levels using either F-tests or permutation tests. Next, we fit another 3840 M_2 after removing $(\alpha\beta)_{ikn}$ in M_1. The F-statistics for α_{in}, given by $F_G = MS_G/MS_E$ were obtained with (1,25) degrees of freedom. Under the normality assumption, the adjusted p-values were computed using the Bonferroni method (Holm, 1979; Shaffer, 1986, 1995). At the 5% significance level we found 53 genes.

Next, we perform the permutation test using the residuals τ_{ikln} given in Equation (3). Since the interaction terms $(\alpha\beta)_{ikn}$ are not significant, we use the second type of residuals. For each gene, 36 residuals are available. The possible number of permuted samples are $\binom{36}{18}$, which is too large to handle. Thus, for each gene we randomly generated 100 000 permutated samples and computed the two-sample t-statistics given in Equation (4). From these 100 000 t-statistics, the adjusted p-value was computed by Westfall and Young's method. At the 5% significance level, we identified 90 genes. We also computed the numbers of genes differentially expressed genes at different significance levels. For example, at the 1% significance level, the permutation test yielded 59, while the F-test yielded 37. At the 10% significance level, the permutation test yielded 106 genes, while the F-test yielded 64 genes.

We compare the list of 53 genes of F-tests and 90 genes of permutation tests that were selected at the 5% significance level. Note that the order of genes do not

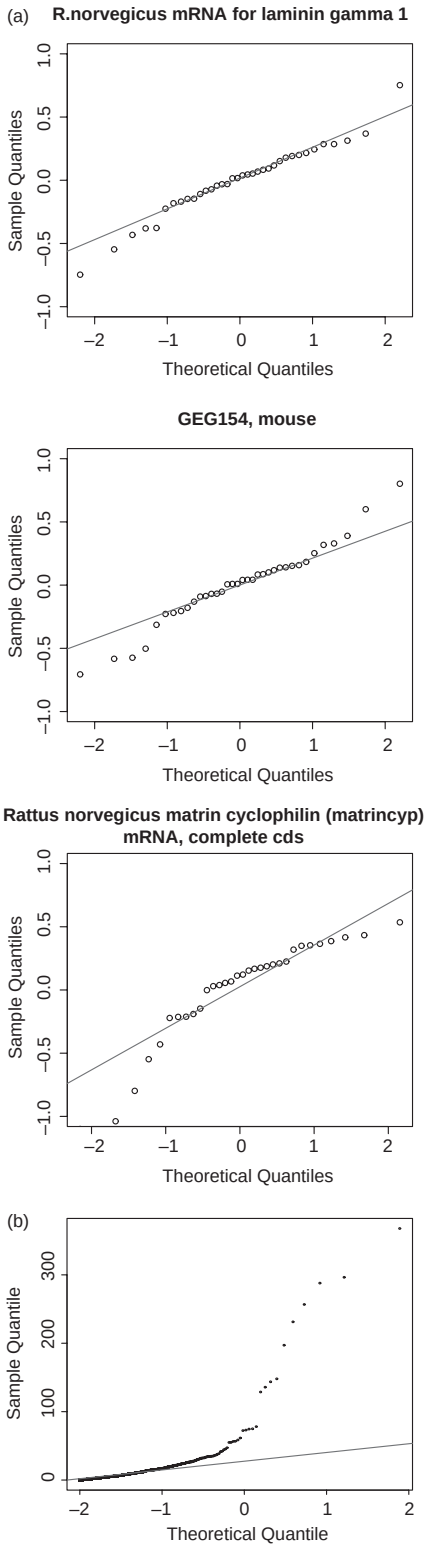

Fig. 1. The quantile–quantile plots. (a) The quantile–quantile plots of residuals for checking the normality assumption. The residuals are from the *1st, 6th, and 27th genes in Table 3*; **(b)** The quantile–quantile plot of F-statistic assuming all F-statistic follows $F(1,29)$ distribution.

differ much. The reason why the order of gene list is similar is that the order of genes in the F-test is preserved in computing the adjusted p-values of permutation test. In addition, all 53 genes are included in 90 genes. Table 3 shows the full list of all genes that are identified by F-tests and/or permutation tests. The first column shows the rank by permutation tests and the last column does the rank by F-tests. The blanks in F-tests show the same ranks as permutation tests. The difference between F-tests and permutation tests might be caused by the fact that the log-ratios of intensities do not follow normal distributions. For illustrative purpose, we select three genes in Table 3 and draw the quantile–quantile plots of residuals for checking the normality assumption. As shown in Figure 1a, the distributions are far from the normal distribution for some genes. Also, Figure 1b shows the quantile–quantile plots of F-statistics. For simplicity, we assume the same degrees of freedom, though they are slightly different due to missing data. If the F-statistics follow an F-distribution, they would scatter around the $Y = X$ line.

For further analysis, we focus on the these 53 genes that are selected by both F-tests and permutation tests. For the selected 53 genes, we performed a clustering analysis to confirm our findings. Using the Euclidian distance measures, we performed the K-means clustering analysis on the log-ratios. We tried different number of clusters. It appears that three clusters provide most reasonable grouping. The genes in the first cluster have green colors in both CNTF and No CNTF groups. The genes in the second cluster have green colors for the CNTF group but red colors for the No CNTF group. Finally, the genes in the third cluster have red colors for the CNTF group but green colors for the No CNTF group. Figure 2 shows the results of K-means clustering. We first tried a clustering analysis for all 3840 genes. However, we could not get any clear pattern of clusters. After selecting significant genes based on the statistical tests, we obtained clear cluster patterns.

For these three clusters, Figure 3 shows the profiles of log-ratio intensities over time. The left graphs are for CNTF groups and the right ones for No CNTF groups.

Table 3. List of significant genes. The following gene are significant for F-tests and/or permutation tests that are differentialy expressed between two groups

Rank	Gene name	Rank of F-test
1	*R.norvegicus* mRNA for *laminin gamma 1*	*
2	Unknown-B0484	3
3	Rat membrane guanylate cyclase mRNA, complete cds	2
4	*R.norvegicus* mRNA for NTR2 receptor	*
5	*Rattus norvegicus* (clone nclk) cdc2-related protein kinase mRNA, complete cds	*
6	GEG-154, mouse	*
7	Unknown-A1427	*
8	Poly(A) binding protein, mouse	*
9	Unknown-D0964a	*
10	Rat transcriptional repressor of myelin-specific genes (SCIP) mRNA, complete cds	*
11	Ribosomal p s6, rat	*
12	Rat retinol-binding protein (RBP) gene	*
13	Ini1 mRNA?,human	*
14	unknown-A0267	*
15	*Rattus norvegicus* low voltage-activated, T-type calcium channel alpha subunit (CACNA1G) mRNA, complete cds	*
16	Rat connexin 43 mRNA, complete cds	*
17	Rat alpha-prothymosin mRNA, complete cds	*
18	Rat mRNA for rhodanese	*
19	Unknown-B0388b	*
20	folate-binding protein, mouse	*
21	*Rattus norvegicus* phospholpase C delta-4 mRNA, complete cds	*
22	Unknown-D2622g	*
23	*Rattus norvegicus* thrombin mRNA, $3'$ end	*
24	*Rattus norvegicus* lactate dehydrogenase-B (LDH-B) mRNA, complete cds	*
25	Rat PRRHIS8 mRNA for ribosomal protein S8	*
26	Rat zinc finger protein (kid-1) mRNA, complete cds	*
27	*Rattus norvegicus* matrin cyclophilin (matrin-cyp) mRNA, complete cds	36
28	Rat mRNA for Distal-less 3 (Dlx-3) homeobox protein	27
29	Rat isoprenylated 67 kDa protein mRNA, complete cds	28
30	Rat mRNA for Ash-m, complete cds	29
31	microsatellite	30
32	Obiquinone oxidoreductase	31
33	Rat mRNA for T-cell marker CD2 antigen	32

Table 3. *Continued.*

Rank	Gene name	Rank of *F*-test
34	*Rattus norvegicus* interleukin-1 beta converting, enzyme (ILIBCE) mRNA, complete cds	33
35	*Rattus norvegicus* mRNA for proteasome p45/SUG, complete cds	34
36	Sum-transplantation antigen P198, mouse	35
37	*Rattus norvegicus* adenylyl cyclase type V mRNA, complete cds	*
38	*Rattus norvegicus* CAP1 gene	*
39	Unknown-C1330d	*
40	Insulin-like growth factor, mouse	*
41	Rat brain glyceraldehyde-3-phosphate dehydrogenase mRNA, 3′ end	43
42	Unknown-C1393b	44
43	*Rattus norvegicus* Sprague–Dawley protein tyrosine phosphatase mRNA, complete cds	41
44	Unknown-D1439b	42
45	*Rattus norvegicus* prostaglandin F2a receptor regulatory protein precursor, mRNA, complete cds	*
46	Rat mRNA for V-1 protein, complete cds	*
47	*Rattus norvegicus* GSK-3beta interacting protein rAxin mRNA, complete cds	*
48	Cytochrome oxidase, rat	*
49	Rat liver interleukin 6 receptor ligand binding chain mRNA, complete cds	*
50	*Rattus norvegicus* cysteine sulfinic acid decarboxylase mRNA, complete cds	*
51	rRNA 18S	*
52	Unknown-A1141	54
53	Rat mRNA for novel protein kinase PKN, complete cds	62
54	Unknown-B1571c	52
55	Rat liver mRNA for proteasomal ATPase (S4), complete cds	57
56	Acidic ribosomal phosphoprotein P1, rat	53
57	Splicing factor Sip1, human	59
58	Ribosoinal p. 40 kDa, rat	55
59	Unknown-B0430a	56
60	*Rattus norvegicus* mRNA encoding 45 kDa protein which binds to heymann nephritis antigen gp330	58
61	*Rattus norvegicus* polo like kinase (plk) mRNA, complete cds	60
62	Rat prohibitin (phb) mRNA, complete cds	64
63	*Rattus norvegicus* mRNA for 14-3-3 protein gamma-subtype, complete cds	61
64	Unknown-A0239	63
65	Rat brain glyceraldehyde-3 phosphate dehydrogenase mRNA, 3′ end	*
66	Unknown-D0626b	*
67	*R.norvegicus* mRNA for mammalian fusca protein	70
68	Rat mRNA for 230 kDa phosphatidylinositol 4-kinase, complete cds	67
69	*Rattus norvegicus* Lyn B tyrosine kinase (LynB) mRNA, complete cds	68
70	rVlaR (*R.norvegicus* mRNA for V1a arginine vasopressin receptor)	69
71	Rat alpha-crystallin B chain mRNA, complete cds	*
72	*Rattus norvegicus* CD59 protein precursor, mRNA, complete cds	*
73	Chromosome X region human	*
74	Unknown-A0222b	*
75	Unknown-D0180d	*
76	Unknown-C1635b	*
77	Unknown-D1441a	*
78	*R.norvegicus* P2X mRNA	*
79	*R.norvegicus* mRNA for cathepsin B	*
80	Unknown-D0037b	*
81	*Rattus norvegicus* cysteine sulfinic acid decarboxylase mRNA, complete cds	*
82	*Rattus norvegicus* endoplasmic reticulum protein ERp29 precursor, mRNA, complete cds	*
83	*Rattus norvegicus* p41-Arc mRNA, complete cds	*
84	Tyrosine phosphatase-like p. (brain), rat	*
85	*Rattus norvegicus* type 1 astrocyte and olfactory-limbic associated protein AT1-46 mRNA, complete cds	*
86	Rat calmodulin (RCM3) mRNA, complete cds	*
87	Unknown-D0284a	*
88	*Rattus norvegicus* mRNA for serine protease, complete cds	*
89	*R.norvegicus* MYR1 mRNA for myosin I heavy chain	94
90	*Rattus norvegicus* 190 kDa ankyrin isoform mRNA, complete cds	93

* Represents the same rank.

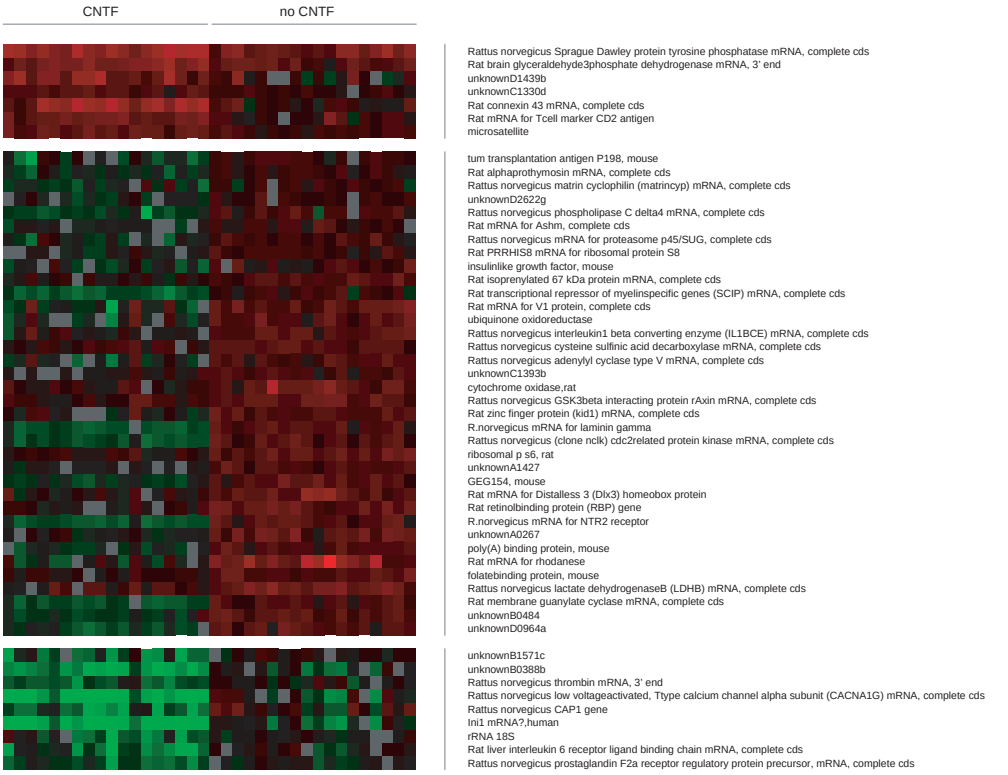

Fig. 2. The results of K-means clustering with three clusters.

Each line represents one gene. The solid lines with circles represent mean values. The genes in the first cluster tend to have lower log-ratios for No CNTF groups, while the genes in the second and third clusters do have higher log-ratios. Also, Figure 3 shows that there are higher variabilities for the CNTF group than No CNTF group. The solid lines do not change much over time, which confirms the finding from ANOVA models that $(\alpha\beta)$s are not significant.

4 DISCUSSION

In this paper, we proposed a statistical procedure to identify genes that have different gene expression profiles among experimental groups in time-course experiments. The proposed model is based on the usual ANOVA model and can detect genes that have different gene expression profiles among I experimental groups.

The proposed approach is an extension of the two-sample t-test proposed by Dudoit *et al.* (2000). They considered two group problem with replicated experiments. When there are more than two groups, the F-test based on the ANOVA model is a natural extension. However, when one factor is *Time* as in the time-course experiment, its extension is not straightforward. The key idea of the proposed model is using residuals after removing the time effect.

For the time sequence data, the main idea of the proposed model can also be applied to the ANOVA model of Kerr *et al.* (2000) as well as the mixed model of Wolfinger *et al.* (2001) by adding a time effect in the model. However, it needs some further considerations on how to fit the two-stage models, how to use residuals, and how to test differentially expressed genes.

Note that the proposed model is different from the ANOVA model of Kerr *et al.* (2000). First, the previous work modelled the log intensity while our approach models the log ratios, though there is a connection between these two models. The ANOVA model of Kerr *et al.* (2000) is more suitable for microarray data obtained from the loop design which has balanced dye effects. The example in this paper used the reference design in which every sample is hybridized to the same extraneous reference sample. Second, Kerr *et al.* (2000) fit one big model to microarray data for all genes simultaneously, while our method fits a model separately for each gene. They employ a bootstrap analysis of residuals and use bootstrap confidence intervals to detect differentially expressed genes. Instead of bootstrap analysis, we use permutation tests. Finally, the key difference here is that we consider two stage models: one for removing the effect of time and the second model is for detecting differentially expressed genes. Wolfinger *et al.* (2001)

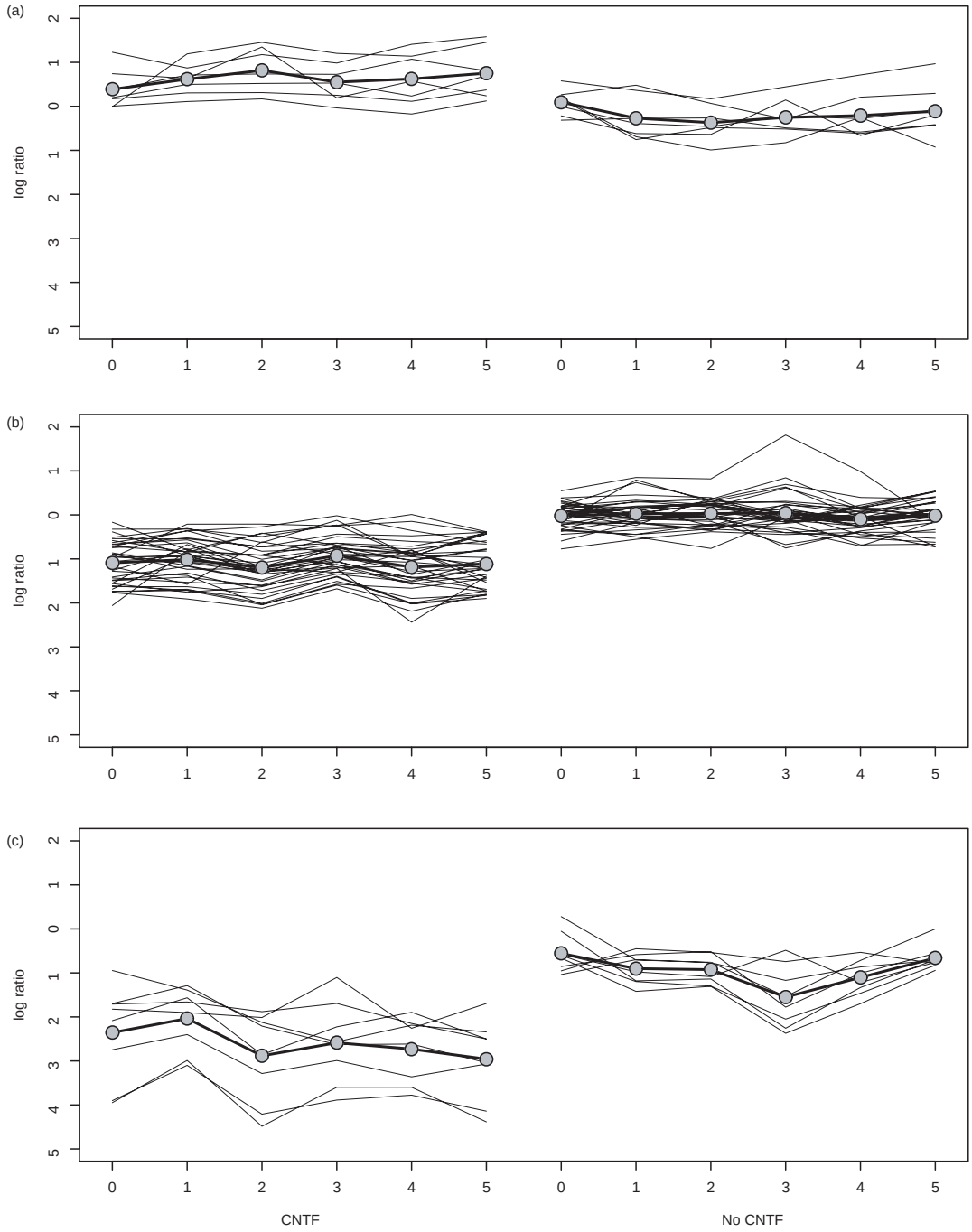

Fig. 3. Profiles of log-ratio intensities for clustered genes in Figure 2. (**a**) cluster 1, (**b**) cluster 2, (**c**) cluster 3.

used similar two-stage mixed models. However, the first stage model is mainly for normalization.

One advantage of the proposed permutation test is that it improves the discreteness of p-values. Note that if we permute data within each time sequence, there are $\binom{6}{3}^6 = 6.4 \times 10^7$ possible permuted samples. On the other hand, our approach yields $\binom{36}{18} = 9.075 \times 10^9$ possible permuted samples. Permutation tests usually suffer from the discreteness of p-values. Since the proposed method allows more permuted samples, the discreteness of p-value can be improved. However, it would be interesting to access the validity of the permutation test p-values and

to perform a power analysis. Furthermore, it is interesting to observe that the order of genes selected by F-tests does not differ from that of permutation tests. It is because the order of genes in the F-test is preserved in computing the adjusted p-values of permutation test. The idea of this p-value adjustment came from Westfall and Young, and later were used by Dudoit *et al.* in the analysis of microarray data. However, we think that the idea of this adjustment also needs a further evaluation. We are planning to conduct simulation studies for investigating the validity of permutation in a more general settings including the t-test of Dudoit *et al.* (2000). The results will be reported in a separate paper.

The proposed model is flexible and easy to extend. Since it is based on the ANOVA model, for example, it can be extended to the cases when there are more than two factors. For example, suppose that we have three factors of interest such as A and B as well as *Time*. Then, the ANOVA model M_1 and M_2 can be extended to the following models:

$$\begin{aligned} M_1^* &: y_{ijkln} = \mu_n + \alpha_{in}^A + \alpha_{jn}^B + \beta_{kn} \\ &\quad + (\alpha\beta)_{ikn}^A + (\alpha\beta)_{jkn}^B + \epsilon_{ijkln}, \\ M_2^* &: y_{ijkln} = \mu_n + \alpha_{in}^A + \alpha_{jn}^B + \beta_{kn} + \epsilon_{ijkln}, \end{aligned} \quad (7)$$

where $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, $l = 1, \ldots, L$, and $n = 1, \ldots, G$. In these models, α_{in}^A and α_{jn}^B represent the effects of A and B, respectively. We can also consider some models with only one interaction terms between M_1^* and M_2^*. The other test procedures described previously can be similarly extended.

ACKNOWLEDGEMENTS

The authors thank anonymous referees and the associate editor for many helpful comments The work for the first author was supported by a grant from the KOSEF through the Statistical Research Center for Complex

Systems at Seoul National University. The work for the last author was supported by grant number FG-5-01 of 21C Frontier Functional Human Genome Project From Ministry of Science Technology of Korea

REFERENCES

Alizadeh,A.A., Eisen,M.B., Davis,R.E., Ma,C., Lossos,I.S., Rosenwald,A., Boldrick,J.C., Sabet,H., Tran,T. Yu,X. *et al.* (2000) Different type of diffuse large b-cell lymphoma identified by gene expression profiling. *Nature*, **403**, 503–511.

Chen,Y., Dougherty,E.R. and Bittner,M.L. (1997) Ratio-based deisions and quantitative analysis of cDNA microarray images. *J. Biomed. Optics*, **2**, 364–374.

Dudoit,S., Yang,Y.H., Callow,M.J. and Speed,T.P. (2000) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray expriments. Submitted.

Friddle,C.J., Koga,T., Rubin,E.M. and Bristow,J. (2000) Expression profiling reveals distinct sets of genes altered during induction and resgression of cardiac hypertrophy. *Proc. Natl Acad. Sci. USA*, **97**, 6745–6750.

Galitski,T., Saldanha,A.J., Styles,C.A., Lander,E.S. and Mesirov,J.P. (1999) Ploidy regulation of gene expression. *Science*, **285**, 251–254.

Golub,T.R., Slonim,D.K., Tamayo,P., Huard,C., Gaasenbeek,M., Mesirov,J.P., Coller,H., Loh,M.L., Downing,J.R. Caligiuri,M.A. *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531–537.

Holm,S. (1979) A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, **6**, 65–70.

Kerr,M.K., Martin,M. and Churchill,G.A. (2000) Analysis of variance for gene expression microarray data. *J. Comput. Biol.*, **7**, 819–837.

Kerr,M.K., Afshari,C.A., Bennett,L., Bushel,P., Martinez,J., Walker,N. and Churchill,G.A. (2001) Statistical analysis of a gene expression microarray experiment with replication. *Statistica Sinica*, in press.

Kim,J.H., Kim,H.Y. and Lee,Y.S. (2001) A novel method using edge detection for signal extraction from cDNA microarray image analysis. *Experimental and Molecular Medicine*, **33**, 83–88.

Lee,M.T., Kuo,F.C., Whitmore,G.A. and Sklar,J. (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc. Natl Acad. Sci. USA*, **97**, 9834–9839.

Newton,M.A., Kendziorski,C.M., Richmond,C.S., Blattner,F.R, and Tsui,K.W. (2001) On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. *J. Comput. Biol.*, **8**, 37–52.

Shaffer,J.P. (1986) Modified sequentially rejective multiple test procedures. *JASA*, **81**, 826–831.

Shaffer,J.P. (1995) Multiple hypothesis testing. *Annu. Rev. Psychol.*, **46**, 561–584.

Spellman,P.T., Sherlock,G., Zhang,M.Q. Iyer,V.R. *et al.* (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.

Storey,J.D. and Tibshirani,R. (2001) Estimating false discovery rates under dependence, with applications to DNA microarrays. *Technical Report 2001-28*. Department of Statistics Stanford University.

Tseng,G.C., Oh,M.K., Rohlin,L., Liao,J.C. and Wong,W.H. (2001) Issues in cDNA microarray analysis: qualit filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res.*, **29**, 2549–2557.

Westfall,P.H. and Young,S.S. (1993) *Resampling-based multiple testing: examples and methods for p-value adjustment*, Wiley series in probability and mathematical statistics.

Wolfinger,R.D., Gibson,G., Wolfinger,E.D., Bennett,L., Hamadeh,H., Bushel,P., Afshari,C. and Paules,R.S. (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *J. Comput. Biol.*, **8**, 625–637.

Yang,Y.H,, Dudoit,S., Luu,P. and Speed,T.P. (2000) Normalization for cDNA microarray data. *Technical Report*.