Methodology article

Sample size calculation for microarray experiments with blocked one-way design

Sin-Ho Jung*[1], Insuk Sohn[1], Stephen L George[1], Liping Feng[2] and Phyllis C Leppert[2]

Address: [1]Department of Biostatistics and Bioinformatics, Duke University Medical Center, North Carolina 27710, USA and [2]Center for Fibroid Biology and Therapy, Duke University Medical Center, North Carolina 27710, USA

Email: Sin-Ho Jung* - sinho.jung@duke.edu; Insuk Sohn - insuk.sohn@duke.edu; Stephen L George - stephen.george@duke.edu; Liping Feng - liping.feng@duke.edu; Phyllis C Leppert - phyllis.leppert@duke.edu

* Corresponding author

Abstract

Background: One of the main objectives of microarray analysis is to identify differentially expressed genes for different types of cells or treatments. Many statistical methods have been proposed to assess the treatment effects in microarray experiments.

Results: In this paper, we consider discovery of the genes that are differentially expressed among K (> 2) treatments when each set of K arrays consists of a block. In this case, the array data among K treatments tend to be correlated because of block effect. We propose to use the blocked one-way ANOVA F-statistic to test if each gene is differentially expressed among K treatments. The marginal p-values are calculated using a permutation method accounting for the block effect, adjusting for the multiplicity of the testing procedure by controlling the false discovery rate (FDR). We propose a sample size calculation method for microarray experiments with a blocked one-way design. With FDR level and effect sizes of genes specified, our formula provides a sample size for a given number of true discoveries.

Conclusion: The calculated sample size is shown via simulations to provide an accurate number of true discoveries while controlling the FDR at the desired level.

Background

Clinical and translational medicine have benefited from genome-wide expression profiling across two or more independent samples, such as various diseased tissues compared to normal tissue. DNA microarray is a high throughput biotechnology designed to measure simultaneously the expression level of tens of thousands of genes in cells. Microarray studies provide the means to understand the mechanisms of disease. However, various sources of error can influence microarray results [1].

Microarrays also present unique statistical problems because the data are high dimensional and are insufficiently replicated in many instances. Methods of adjustment for multiple testing therefore become extremely important. Multiple testing methods controlling the false discovery rate (FDR) [2] have been popularly used because they are easy to calculate and less strict in controlling the false positivity compared to the family-wise error rate (FWER) control method [3].

Numerous sample size calculation methods have been proposed for comparing *independent* groups while controlling the FDR in designing microarray studies. Lee and Whitmore [4] considered comparing multiple groups using ANOVA models and derived the relationship between the effect sizes and the FDR using a Bayesian approach. Their power analysis does not address the multiple testing issue. Muller *et al.* [5] chose a pair of testing errors, including FDR, and minimized one while controlling the other at a specified level using a Bayesian decision rule. Jung [6] proposed a closed form sample size formula for a specified number of true rejections while controlling the FDR at a desired level. Pounds and Cheng [7] and Liu and Hwang [8] proposed similar sample size formulas which can be used for comparison of K independent samples. These methods are for the FDR-control methods based on independence or a weak dependency assumption among test statistics. Recently, Shao and Tseng [9] introduced an approach for calculating sample sizes for multiple comparisons accounting for dependency among test statistics.

In some studies, specimens for K treatments are collected from the same subject and means are compared across treatment groups. In this case, the gene expression data for the K treatments may be dependent since they share the same physiological conditions. For example, Feng *et al.* [10] conducted a study to discover the genes differentially expressed between center (C) and edge (E) of the uterine fibroid and the matched adjacent myometrium (M). In this study, specimens are taken from the three sites for each patient. The patients are blocks and the three sites (K = 3), C, E and M, are treatments (or groups) to be compared.

Since a set of K specimens are collected from each patient, we require a much smaller number of patients than a regular unblocked design. Furthermore, the observations within each block tend to be positively correlated, so that a blocked design requires a smaller number of arrays than the corresponding unblocked design just as a paired two-sample design with a positive pairwise correlation requires a smaller number of observations than a two independent sample design. The more heterogeneous the blocks are, the greater the savings in number of arrays for the blocked design.

In this paper, we consider a non-parametric blocked F-test statistic to compare the gene expression level among K dependent groups. We adjust for multiple testing and control the FDR by employing a permutation method. We propose a sample size calculation method for a specified number of true rejections while controlling the FDR at a specified level. Through simulations, we show that the blocked F-test accurately controls the FDR using the permutation resampling method and the calculated sample

size provides an accurate number of true rejections while controlling the FDR at the desired level. For illustration, the proposed methods are applied to the fibroid study [10] mentioned above.

Methods
Non-parametric block **F**-*test statistic*
Suppose that we want to discover genes that are differentially expressed among K sites (treatments or groups). For each of n patients (blocks), a specimen is collected from each site for a microarray experiment on m genes. In this case, the gene expression data from the K sites tend to be correlated. Let Y_{ijk} denote the expression level of gene i (= 1,..., m) from treatment k (= 1,..., K) of block j (= 1,..., n). We consider the blocked one-way ANOVA model

$$Y_{ijk} = \mu_i + \delta_{ik} + \gamma_{ij} + \varepsilon_{ijk}, \tag{1}$$

where, for gene i, μ_i is the population mean, δ_{ik} is a fixed treatment effect and the primary interest, γ_{ij} is a random block effect, and ε_{ijk} is a random error term. We assume that $\sum_{k=1}^{K} \delta_{ik} = 0$, $\gamma_{i1},..., \gamma_{in}$ are independent and identically distributed (IID) with mean 0 and variance v_i, (ε_{ijk}, $1 \le j \le n$, $1 \le k \le K$) are IID with mean 0 and variance σ_i^2, and error terms and block effects are independent. The standard ANOVA theory using parametric F distributions to test the treatment effect assumes a normal distribution for ε_{ijk}. However, in this paper, we avoid the normality assumption by using a permutation resampling method in testing and a large-sample approximation in sample size calculation.

For gene i(= 1,..., m), the hypotheses for testing the treatment effect are described as

$$H_i : \delta_{i1} = \delta_{i2} = \dots = \delta_{iK} = 0$$

against

$$\bar{H}_i : \delta_{ik} \neq 0 \text{ for some } k = 1,\dots, K.$$

We reject H_i in favor of $\bar{H}_i$ for a large value of F-test statistic

$$F_i = \frac{n \sum\limits_{k=1}^{K} (\bar{Y}_{i.k} - \bar{Y}_{i..})^2 / K - 1)}{\sum\limits_{j=1}^{n} \sum\limits_{k=1}^{K} (Y_{ijk} - \bar{Y}_{ij.} - \bar{Y}_{i.k} + \bar{Y}_{i..})^2 / \{(K-1)(n-1)\}}, \tag{2}$$

where $\bar{Y}_{i.k} = n^{-1} \sum_{j=1}^{n} Y_{ijk}$, $\bar{Y}_{ij.} = K^{-1} \sum_{k=1}^{K} Y_{ijk}$, and
$\bar{Y}_{i..} = (Kn)^{-1} \sum_{j=1}^{n} \sum_{k=1}^{K} Y_{ijk}$. If the error terms are normally distributed, F_i marginally has the $F_{K-1, (K-1)(n-1)}$ distribution under H_i. The normality assumption can be relaxed if n is large.

Without the normality assumption, the joint null distribution of the statistics can be approximated using a block permutation method, where the array data sets for K treatments are randomly shuffled within each block: the permuted data may be represented as $\{(Y_{ij\tilde{1}}, \ldots Y_{ij\tilde{K}}), 1 \le i \le m, 1 \le j \le n\}$, where $(\tilde{1}, \ldots, \tilde{K})$ is a random permutation of $(1, \ldots, K)$. Note that there are $(K!)^n$ different permutations, among which $(K!)^{n-1}$ give different F-statistic values. The R language package multtest [11] can be used to implement the permutation-based multiple testing procedure for blocked microarray data. We consider adjusting for the multiplicity of the testing procedure by controlling the FDR [12,13].

Permutation-based multiple testing for FDR-control
(i) Compute the F-test statistics $(F_1, \ldots, F_m)$ from the original data, $(f_1, \ldots, f_m)$.

(ii) From the b-th permutation data ($b = 1, \ldots, B$), compute the F-test statistics $(F_1^{(b)}, \cdots, F_m^{(b)})$.

(iii) For gene i, estimate the marginal p-value by

$$p_i = \frac{1}{B} \sum_{b=1}^{B} I(F_i^{(b)} \ge f_i),$$

where $I(A)$ is an indicator function of event A.

(iv) For a chosen constant $\lambda \in (0, 1)$, estimate the q-value by

$$q_i = \frac{p_i \sum_{l=1}^{m} I(p_l > \lambda)}{(1-\lambda) \sum_{l=1}^{m} I(p_l \le p_i)}.$$

(v) For a specified FDR level q^*, discover gene i (or reject H_i) if $q_i < q^*$.

Sample size calculation
Let $_0$ and $_1$ denote the sets of indices of genes that are equally and differentially expressed, respectively, in K treatments, and $\{\bar{\delta}_{ik} = \delta_{ik}/\sigma_i, i \in {}_1, 1 \le k \le K\}$ denote the standardized effect sizes for the differentially expressed genes. Let m_0 and $m_1 = m - m_0$ denote the cardinalities of $_0$ and $_1$, respectively.

Suppose that we want to discover gene i (or reject H_i) if the marginal p-value p_i is smaller than $\alpha \in (0, 1)$. For large m and under the independence assumption or weak dependence among the F-test statistics, the FDR corresponding to the cutoff value α is approximated by

$$FDR(\alpha) = \frac{m_0\alpha}{m_0\alpha + \sum_{i \in \mathcal{M}_1} \beta_i(\alpha)}, \qquad (3)$$

where $\beta_i(\alpha) = P(p_i \le \alpha)$ is the marginal power of a single α-test applied to gene $i \in {}_1$ and $\sum_{i \in \mathcal{M}_1} \beta_i(\alpha)$ denotes the expected number of true rejections when we reject H_i for $p_i < \alpha$, see Jung [6].

Now, we derive $\beta_i(\alpha)$ for gene $i \in {}_1$. By the standard blocked one-way ANOVA theory under the normality assumption for ε_{ijk},

$$\sum_{k=1}^{K} (\bar{Y}_{i.k} - \bar{Y}_{i..})^2 \sim \frac{\sigma_i^2}{n} \chi_{K-1}^2 \left(n \sum_{k=1}^{K} \bar{\delta}_{ik}^2\right)$$

and

$$\frac{1}{(K-1)(n-1)} \sum_{j=1}^{n} \sum_{k=1}^{K} (Y_{ijk} - \bar{Y}_{ij.} - \bar{Y}_{i.k} + \bar{Y}_{i..})^2 \sim \sigma_i^2 \chi_{(K-1)(n-1)}^2$$

are independent, where $\chi_v^2(\eta)$ is the noncentral χ^2-distribution with v degrees of freedom and noncentrality parameter η, and $\chi_v^2 = \chi_v^2(0)$. Hence, for the F-test statistic (2), we have

$$F_i \sim F_{(K-1),(K-1)(n-1)}\left(n \sum_{k=1}^{K} \bar{\delta}_{ik}^2\right)$$

where $F_{v_1, v_2}(\eta)$ is the noncentral F-distribution with v_1 and v_2 degrees of freedom, and noncentrality parameter η. Note that, for $i \in {}_0$, $\bar{\delta}_{i1} = \cdots = \bar{\delta}_{iK} = 0$ and $F_i \sim F_{(K-1),(K-1)(n-1)}(0) = F_{(K-1),(K-1)(n-1)}$, the central F-distribution.

The marginal powers are expressed as

$$\beta_i(\alpha) = P\{F_{(K-1),(K-1)(n-1)}(n\sum_{k=1}^{K}\bar{\delta}_{ik}^2) \ge F_{(K-1),(K-1)(n-1),\alpha}\},$$

$$(4)$$

where $F_{v_1,v_2,\alpha}$ denotes the $100(1 - \alpha)$ percentile of F_{v_1,v_2} distribution. The marginal powers can be calculated using R, SAS or some other packages. Suppose we want r_1 true rejections while controlling the FDR at q^*. By combining this with (3) and (4), we obtain two equations

$$r_1 = \sum_{i \in \mathcal{M}_1} P\{F_{(K-1),(K-1)(n-1)}(n\sum_{k=1}^{K}\bar{\delta}_{ik}^2) \ge F_{(K-1),(K-1)(n-1),\alpha}\}$$

$$(5)$$

and

$$q^* = \frac{m_0\alpha}{m_0\alpha + r_1}. \qquad (6)$$

Note that r_1/m_1 denotes the probability of true rejection. At the design stage of a study, m is given by the number of genes included in the chips to be used for microarray experiment, m_1 and $\{\bar{\delta}_{ik}, i \in_1, 1 \le k \le K\}$ are projected based on biological knowledge or estimated from pilot data, and K, r_1 (or r_1/m_1) and q^* are prespecified. The only unknown variables in (5) and (6) are α and n. By solving (6) with respect to α, we obtain $\alpha^* = r_1 q^*/\{m_0 (1 - q^*)\}$ and, by plugging this in (5), we obtain an equation for r_1 depending only on n,

$$r_1 = \sum_{i \in \mathcal{M}_1} P\{F_{(K-1),(K-1)(n-1)}(n\sum_{k=1}^{K}\bar{\delta}_{ik}^2) \ge F_{(K-1),(K-1)(n-1),\alpha^*}\}.$$

$$(7)$$

The marginal power function (4) includes n in the degrees of freedom of the denominator as well as the noncentrality parameter of the F-distributions. The impact of the degrees of freedom of the denominator of the F-statistic on the marginal power is much weaker than that of the noncentrality parameter, so that $\beta_i(\alpha)$ is a monotone increasing function of n, and consequently equation (7) has a unique solution. Figure 1 demonstrates the relationship between n and $\beta_i(\alpha)$ with $\alpha = 0.05$; $\bar{\delta}_{ik} = \{k - (K + 1)/2\}/K$ for $1 \le k \le K$; $K = 3$, 4 or 5. This monotone relationship becomes clear for large n as shown by an approxi-
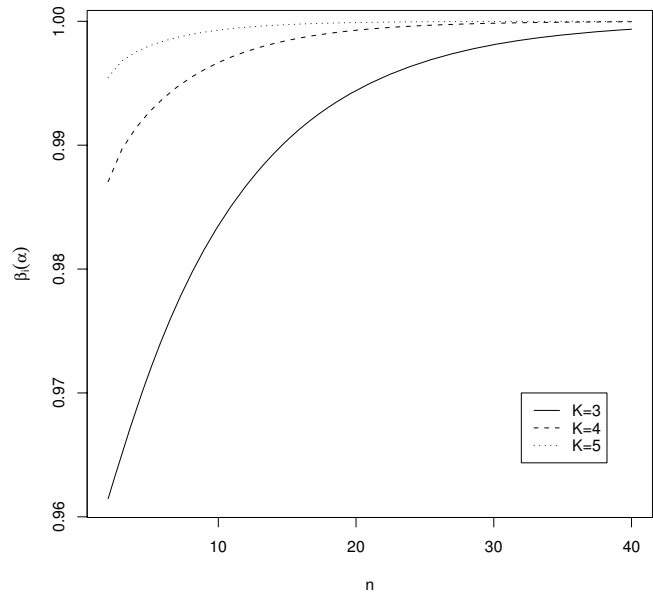

Figure 1 $\bar{\delta}_{ik}$
Monotone relationship between *n* and marginal power $\beta_i(\alpha)$ with α = 0.05 and $\bar{\delta}_{ik}$ = {k - (K + 2)/2}/K for 1 $\le$ k $\le$ K.

mate sample size formula given below. Note that the variance of block effect v_i has no impact on the sample size and power of the test statistic for treatment effect.

In summary, the sample size (i.e., number of blocks) n for $r_1 (\le m_1)$ true rejections is calculated as follows, assuming that the error terms in model (1) are normally distributed.

Sample size calculation based on the noncentral F-*distribution*
(i) Specify the input variables:

    - K = number of treatments;

    - m = total number of genes for testing;

    - m_1 = number of genes differentially expressed in K treatments (m_0 = m - m_1);

    - $\{\bar{\delta}_{ik}, i \in_1, 1 \le k \le K\}$ = standardized effect sizes for prognostic genes;

    - q^* = FDR level;

    - r_1 = number of true rejections

(ii) Using the bisection method, solve

$$r_1 = \sum_{i \in \mathcal{M}_1} P\{F_{(K-1),(K-1)(n-1)}(n \sum_{k=1}^{K} \bar{\delta}_{ik}^2) \geq F_{(K-1),(K-1)(n-1),\alpha^*}\}.$$

with respect to n, where $\alpha^* = r_1 q^* / \{m_0(1 - q^*)\}$.

(iii) The required sample size is n blocks, or nK array chips.

In the sample size formula based on the noncentral F-distribution, the relationship between n and the marginal power functions based on the F-distribution is complicated and a normal distribution assumption of the error terms is required. In the large sample case, we can loosen the normality assumption and simplify this relationship. If the error terms have a finite 4-th moment, then, for large n, the distribution of F_i is approximated by

$$\frac{1}{(K-1)} \chi_{K-1}^2 (n \sum_{k=1}^{K} \bar{\delta}_{ik}^2).$$

A proof is given in the Appendix. Similarly, for large n, the $F_{(K-1),(K-1)(n-1)}$ distribution can be approximated by $(K-1)^{-1} \chi_{K-1}^2$, so that $F_{(K-1),(K-1)(n-1),\alpha} \approx (K-1)^{-1} \chi_{K-1,\alpha}^2$, where $\chi_{\nu,\alpha}^2$ is the $100(1 - \alpha)$ percentile of the χ^2 distribution with ν degrees of freedom. Hence, the marginal power for F_i is approximated by

$$\beta_i(\alpha) = P\{\chi_{K-1}^2 (n \sum_{k=1}^{K} \bar{\delta}_{ik}^2) \geq \chi_{K-1,\alpha}^2\},$$

and a sample size based on the χ^2-distribution approximation is obtained by solving

$$r_1 = \sum_{i \in \mathcal{M}_1} P\{\chi_{K-1}^2 (n \sum_{k=1}^{K} \bar{\delta}_{ik}^2) \geq \chi_{K-1,\alpha^*}^2\} \qquad (8)$$

with respect to n, where $\alpha^* = r_1 q^* / \{m_0(1 - q^*)\}$. In this equation, n appears only in the noncentrality parameter of the χ^2 distributions.

Equation (8) is especially useful when we want to compare the powers between a blocked one-way design and an unblocked one-way design. Using similar approximations, it is easy to show that an approximate sample size $N = nK$ for a study with unblocked one-way design with a balanced allocation is obtained by solving

$$r_1 = \sum_{i \in \mathcal{M}_1} P\{\chi_{K-1}^2 (n \sum_{k=1}^{K} \bar{\delta}_{ik}^2) \geq \chi_{K-1,\alpha^*}^2\} \qquad (9)$$

with respect to n, where $\tilde{\bar{\delta}}_{ik} = \delta_{ik} / \sqrt{v_i + \sigma_i^2}$. The only difference between (8) and (9) is the standardized effect sizes, $\bar{\delta}_k = \delta_{ik}/\sigma_i$ and $\tilde{\bar{\delta}}_{ik} = \delta_{ik}/\sqrt{v_i + \sigma_i^2}$. The latter is always smaller than the former because of the variance among blocks, v_i. If v_i is large compared to the variance of experimental errors, σ_i^2, then a blocked one-way design requires much smaller number of arrays than an unblocked one-way design. Let n_u and n_b denote the sample sizes n calculated under an unblocked and a blocked design, respectively. If v_i / σ_i^2 are constant f among the prognostic genes, then from (8) and (9), we have $n_u = (1 + f)n_b$. As an example, consider the design of the fibrosis study as discussed in Background Section and suppose that the variance of the block effects is half of that of measurement errors for the prognostic genes, i.e. $f = 0.5$. In this case, if a blocked design requires $n_b = 100$ patients and $3n_b = 300$ array chips, then the corresponding unblocked design with a balanced allocation requires $n_u = 150$ patients per group or a total 450 patients. For an unblocked design, the number of array chips is identical to that of patients, and compared to the blocked design, the unblocked design requires 1.5 times more chips and 4.5 times more patients.

Results and discussion
Simulations

First, we investigate the accuracy of the FDR control based on blocked one-way ANOVA tests and the sample size formulas via simulations. For the simulations on FDR control, we consider blocked one-way designs with $K = 3$ treatments and $n = 10, 30$, or 50 blocks. For gene i (= 1,..., m) from treatment k (= 1,..., K) of block j (= 1,..., n), block effect γ_{ij} and error terms $_{ijk}$ are generated from $N(0, 0.5^2)$ and $N(0,1)$, respectively. For differentially expressed genes $i \in _1$, the standardized treatment effects are set at $(\bar{\delta}_{i1}, \bar{\delta}_{i2}, \bar{\delta}_{i3}) = (1, 0, -1)$ or $(1, -2, 1)$. We set the total number of genes $m = 4000$; the number of differentially expressed genes $m_1 = 40$ or 200; and the nominal FDR level $q^* = 0.05, 0.1, 0.2, 0.3, 0.4$, or 0.5. We conducted $N = 1000$ simulations under each setting, and the null distribution of the test statistics is approximated from $B = 1000$ permutations for each simulation sample. In simulation $l(= 1,..., N)$, the FDR-control multiple testing method is applied to the simulated data using tuning parameter $\lambda = 0.95$ [12] to count the numbers of total rejections $\hat{r}$ and

false rejections $\hat{r}_0$ and to estimate the FDR, $\hat{q}_l = \hat{r}_0 / \hat{r}$.
Then the empirical FDR is obtained as

$$\hat{q} = \frac{1}{N} \sum_{l=1}^{N} \hat{q}_l.$$

Table 1 reports the simulation results. The testing procedure controls the FDR accurately, i.e. $\hat{q} \approx q^* \approx q*$, when m_1 is large (m_1 = 200), but tends to be anti-conservative, i.e. $\hat{q} > q^* > q*$, when m_1 is small (m_i = 40). Jung and Jang [13] made similar observations for two-sample t-tests and Cox regression.

For the simulations on sample size calculation, we set m = 4000; m_1 = 40 or 200; number of treatment K = 3; treatment effects $(\bar{\delta}_{i1}, \bar{\delta}_{i2}, \bar{\delta}_{i3})$ = (1/4, 0, -1/4) or (1/4, -1/2, 1/4) for $i \in {}_1$; $\gamma_{ij} \sim N (0, 0.5^2)$ and ${}_{ijk} \sim N (0. 1)$. We want the number of true rejections r_1 to be 30%, 60% or 90% of m_1 while controlling the FDR level at $q*$ = 1%, 5% or 10%. For each design setting, we first calculate the sample size n based on the F-distribution or the chi-square approxima-

tion, and then generate N = 1000 samples of size n under the same setting. From each simulation sample, the number of true rejections are counted while controlling the FDR at the specified level using λ = 0.95. The first, second and third quartiles, Q_1, Q_2 and Q_3, of the observed true rejections, $\hat{r}_1$, are estimated from the 1000 simulation samples.

Table 2 summarizes the simulation results by the two methods. As expected, sample size increases in r_1 and decreases in m_1 and $q*$. Since the standardized effect sizes for the differentially expressed genes influence the sample size through their sum of squares, the combination of effect sizes (1/4, 0, -1/4) requires a larger sample size than (1/4, -1/2, 1/4). The sample size based on the chi-square approximation is always smaller than that based on the F-distribution. The median (Q_2) of the empirical true rejections $\hat{r}_1$ is smaller than the nominal r_1 for the sample size based on the chi-square approximation, especially with a small n, while the sample size based on the F-distribution is always accurately powered, i.e. $Q_2 \approx r_1$.

Table 1: Empirical FDR from N = 1000 simulations with *B* = 1000 permutations for each simulation data set

			n		
m_1	$(\delta_{i1}, \delta_{i2}, \delta_{i3})$	$q*$	10	30	50
40	(1, 0, -1)	0.05	0.1766	0.0921	0.0925
		0.1	0.1819	0.1647	0.1705
		0.2	0.2736	0.2462	0.2506
		0.3	0.3636	0.3478	0.3512
		0.4	0.4546	0.4449	0.4431
		0.5	0.5435	0.5389	0.5399
	(1, -2, 1)	0.05	0.0936	0.0899	0.0915
		0.1	0.1619	0.1663	0.1665
		0.2	0.2402	0.2498	0.2421
		0.3	0.3373	0.3469	0.3461
		0.4	0.4347	0.4481	0.4421
		0.5	0.5318	0.5446	0.5340
200	(1, 0, -1)	0.05	0.0653	0.0573	0.0603
		0.1	0.1120	0.1093	0.1130
		0.2	0.2076	0.2105	0.2146
		0.3	0.3079	0.3086	0.3176
		0.4	0.4070	0.4056	0.4171
		0.5	0.5051	0.5013	0.5162
	(1, -2, 1)	0.05	0.0567	0.0554	0.0591
		0.1	0.1108	0.1079	0.1111
		0.2	0.2142	0.2061	0.2116
		0.3	0.3120	0.3052	0.3113
		0.4	0.4124	0.4049	0.4148
		0.5	0.5141	0.5010	0.5162

Table 2: $Q_2 (Q_1, Q_3)/n$, where n is the sample size and Q_k (k = 1, 2, 3) are the k-th quartile of the empirical true rejections $\hat{r}_1$ from N = 1000 simulations

		Based on the chi-square approximation		
m_1	r_1	$q^* = 1\%$	5%	10%
$(\bar{\delta}_{i1}, \bar{\delta}_{i2}, \bar{\delta}_{i3})$ = (1/4, 0, -1/4)				
40	12	11 (9, 13)/123	10 (8, 13)/100	11 (8, 14)/90
	24	23 (20, 26)/166	23 (21, 26)/138	23 (21, 25)/125
	36	36 (34, 37)/242	36 (34, 37)/207	36 (35, 37)/191
200	60	56 (49, 61)/100	55 (47, 61)/77	55 (49, 61)/67
	120	115 (109, 120)/138	118 (112, 124)/110	117 (110, 122)/96
	180	179 (176, 182)/207	178 (176, 182)/171	179 (175, 182)/154
$(\bar{\delta}_{i1}, \bar{\delta}_{i2}, \bar{\delta}_{i3})$ = (1/4, -1/2, 1/4)				
40	12	8 (6, 10)/41	8 (5, 10)/34	7 (5, 10)/30
	24	21 (19, 23)/56	21 (18, 24)/46	21 (19, 24)/42
	36	35 (33, 37)/81	35 (34, 36)/70	36 (34, 37)/64
200	60	42 (36, 48)/34	41 (35, 47)/26	44 (36, 52)/23
	120	103 (98, 109)/46	108 (101, 114)/37	104 (98, 111)/32
	180	176 (173, 180)/70	177 (173, 180)/57	178 (174, 180)/52
		Based on the *F*-distribution		
m_1	r_1	$q^* = 1\%$	5%	10%
$(\bar{\delta}_{i1}, \bar{\delta}_{i2}, \bar{\delta}_{i3})$ = (1/4, 0, -1/4)				
40	12	12 (10, 15)/129	12 (10, 14)/104	12 (10, 15)/94
	24	24 (21, 27)/171	25 (23, 27)/142	24 (22, 26)/129
	36	36 (35, 37)/246	36 (35, 38)/211	36 (35, 38)/194
200	60	60 (55, 66)/104	61 (54, 66)/80	62 (54, 70)/70
	120	123 (117, 128)/142	122 (118, 128)/113	120 (114, 126)/99
	180	179 (177, 184)/211	180 (177, 183)/174	181 (178, 184)/157
$(\bar{\delta}_{i1}, \bar{\delta}_{i2}, \bar{\delta}_{i3})$ = (1/4, -1/2, 1/4)				
40	12	13 (10, 15)/47	13 (10, 15)/38	13 (10, 16)/34
	24	23 (21, 26)/60	25 (23, 27)/50	25 (23, 27)/46
	36	36 (35, 37)/86	35 (35, 38)/73	36 (34, 37)/67
200	60	61 (55, 67)/38	66 (60, 72)/30	66 (59, 71)/26
	120	121 (116, 127)/50	123 (116, 128)/40	121 (116, 126)/35
	180	180 (177, 183)/73	181 (177, 184)/60	182 (178, 185)/55

Example

We applied the permutation-based blocked one-way ANOVA and the sample size calculation method to the fibroid study discussed in the Background Section. From each patient, specimens are taken from two sites of fibroid tissue, center (C) and edge (E), and one normal myometrium (M). Five patients are accrued to the study. We regard the three sites as treatments (K = 3) and the patients as blocks (n = 5). mRNA was amplified and hybridized onto HG-U133 GeneChips according to the protocols recommended by Affymetrix (Santa Clara, CA), and m = 54675 probe sets on the array were analyzed. Expression values were calculated using the Robust Multichip Average (RMA) method [14]. RMA estimates are based upon a robust average of background corrected PM intensities. Normalization was done using quantile normalization [15]. We filtered out all "AFFX" genes and genes for which there were 4 or fewer present calls (based on Affymetrix's present/marginal/absent (PMA) calls using mismatch probe intensity, the ratio of PM to MM). That is, a gene is included only if there are at least 3 present calls among the 15 PMA calls. Filtering yielded 30711 genes to be used in the subsequent analyses.

In order to group the samples according to the degree of similarity present in the gene expression data, we first applied a hierarchical clustering analysis to the filtered 30711 gene expression data and generated a dendrogram
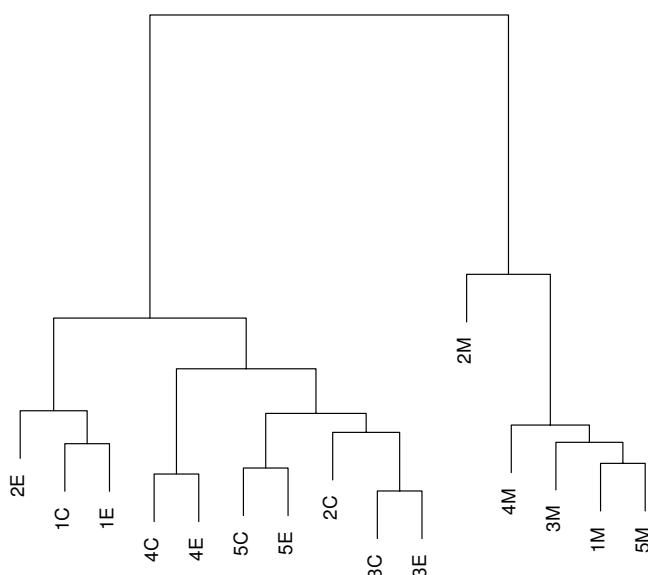
Figure 2
Hierarchical clustering dendrogram. *k*A means site A
(= E, C or M) for patient *k* (= 1,..., 5).

(Figure 2). We used the Complete Linkage method [16] and Pearson's correlation coefficient as a measure of similarity. In the dendrogram, the height of each branch point indicates the similarity level at which each cluster was generated. We obtained the same clustering using the L_2 norm as a measure of similarity. Except for patient 2, E and C are clustered together for each patient. In spite of the block effect, M is clustered separately from E and C regardless of patient assignment. We conclude that C and E have similar gene expression profiles, but M has a different gene expression profile from either C or E. While the clustering analysis investigates the genome wide expression profile, blocked one-way ANOVA helps us identify individual genes differentially expressed among the three sites. Using the blocked one-way ANOVA method, we selected the top 50 genes in terms of parametric p-values (Table 3). The expression patterns of six genes that are identified as differentially expressed are presented in Figure 3. The expression levels of each patients are connected among three sites. These genes are similarly expressed between C and E, but differentially expressed in M. Further, 220273_*at*, 210255_*at*, 229160_*at*, 204620_*s_at* and 217287_*s_at* are under-expressed in M while 1553194_*at* is over-expressed in M.

The results of our analysis of the two sites of fibroid tissue, center and edge, compared to the normal myometrium using a blocked one-way design suggest that reduced FDR provides an enhanced approach to clinical microarray studies. Our findings are consistent with previously reported genome-wide profiling studies [17,18]. We believe that these results support the hypothesis that uterine fibroids develop through altered wound healing signaling pathways leading to tissue fibrosis [19,20]. Using the method described in this paper, genes differentially over-expressed in the fibroid tissue compared to myometrium are related to extracellular matrix (ECM) and ECM regulation such as collagen IV, alpha 1, versican (chondroitin sulfated 2) and IL-17β [21]. IL-17β, a cell-cell signaling transducer has been reported to enhance MMP secretion and to rapidly induce phosphorylation of the extracellular signal-related kinases (ERK) 1/2 and p38MAPK in colonic myofibroblasts and has been shown to stimulate MMP-1 expression in cardiac fibroblasts through ERK 1/2 and p38 MAPK [22,23]. Thus IL-17β is important in remodelling of the extracellular matrix. According to our analysis, RAD51-like 1, a recombinational repair gene, is also over-expressed in fibroids, which is consistent with a report that RAD51B is the preferential translocation partner of high mobility group protein gene (HMGIC) in uterine leiomyomas [24]. HMGIC codes for a protein that is a non-histone DNA binding factor that is expressed during development in embryonic tissue and is an important regulator of cell growth, differentiation and transformation as well as apoptosis [25]. Arrest of apoptosis appears to be a hallmark of uterine fibroids, a finding that is characteristic of altered wound healing as well [19]. HMGIC appears to play a role in the development of uterine fibroids [19,26,27].

Suppose that we want to design a new fibroid study using the data analyzed above as pilot data. In the sample size calculation, we set $m = 30, 000$. We assume that the $m_1 = 50$ genes which were selected as the top 50 genes in terms of parametric p-value are differentially expressed in the three sites ($K = 3$). From the pilot data, we estimate the standardized treatment effect δ_{ik}. For illustration, the effect sizes of these $m_1 = 50$ genes are taken to be $\delta_{ik} = 0.1$ $\hat{\delta}_{ik}$. We need $n = 15$ patients (blocks) to discover 90% of the prognostic genes, i.e. $r_1 = [0.9 \times 50] = 45$, while controlling the FDR at $q^* = 5\%$ level. In a simulation study, we generated $N = 1000$ microarray data sets of size $n = 15$ under this design setting. With $q^* = 0.05$, we observed the quartiles $Q_2(Q_1, Q_3) = 46(45, 47)$ from the empirical distribution of the observed true rejections.
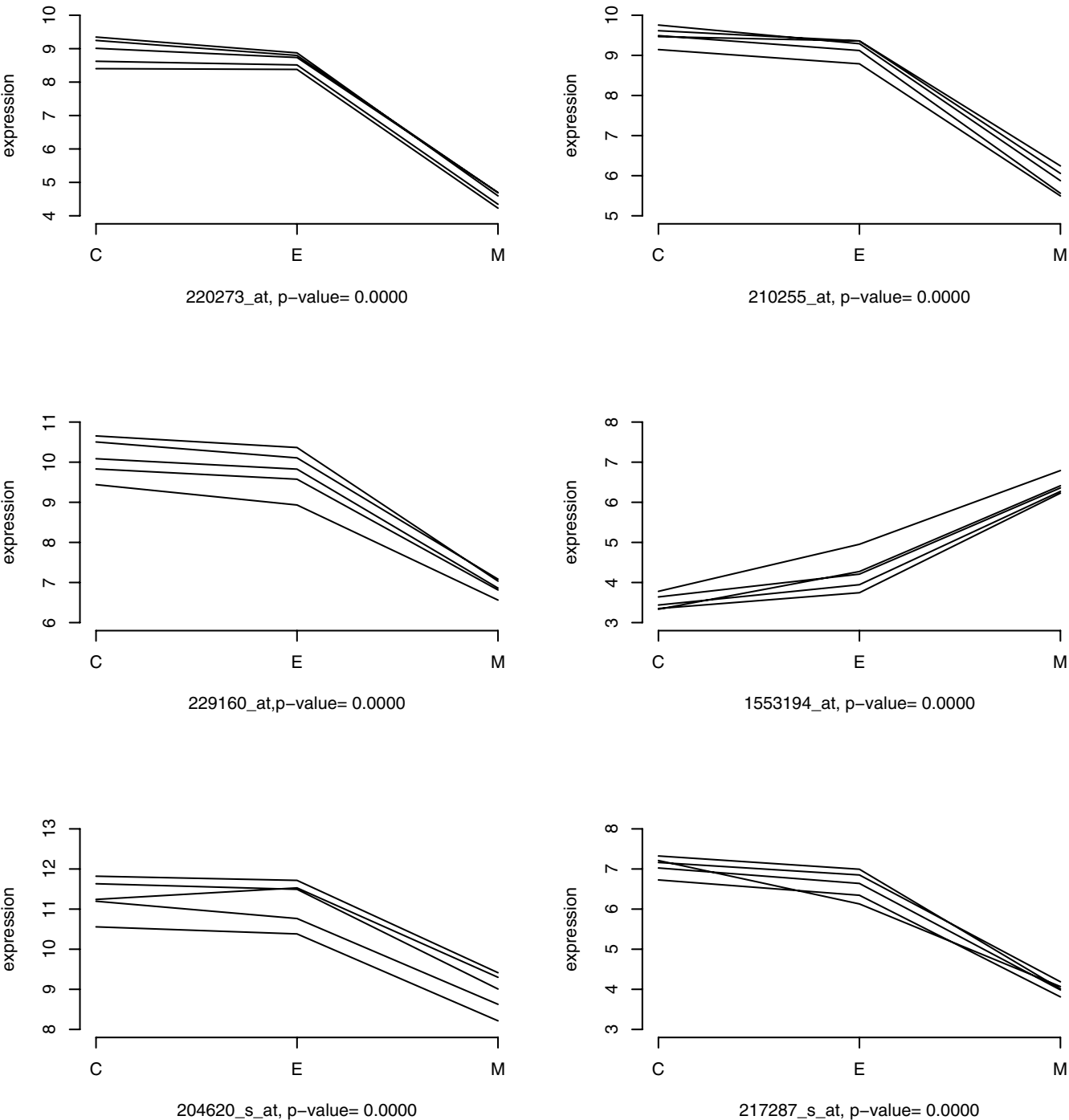
Figure 3
Expression patterns of six genes that are significantly differentially expressed in three sites.

Table 3: The result of unterine fibroid tissue and adjacent myometrium microarray experiment

probe_set_id	Gene_Descriptor	parametric		non-parametric	
		p-value	q-value	p-value	q-value
220273_at	interleukin 17B	0.0000	0.0000	0.0008	0.0131
213479_at	neuronal pentraxin II	0.0000	0.0000	0.0015	0.0144
210255_at	RAD51-like 1 (S. cerevisiae)	0.0000	0.0000	0.0008	0.0131
205833_s_at	prostate androgen-regulated transcript 1	0.0000	0.0000	0.0077	0.0219
229160_at	melanoma associated antigen (mutated) 1-like 1	0.0000	0.0000	0.0008	0.0131
1561122_a_at	RAD51-like 1 (S. cerevisiae)	0.0000	0.0000	0.0046	0.0189
210817_s_at	calcium binding and coiled-coil domain 2	0.0000	0.0000	0.0015	0.0144
1553194_at	neuronal growth regulator 1	0.0000	0.0000	0.0008	0.0131
202965_s_at	calpain 6	0.0000	0.0000	0.0108	0.0239
204620_s_at	chondroitin sulfate proteoglycan 2 (versican)	0.0000	0.0000	0.0054	0.0196
217287_s_at	transient receptor potential cation channel, subfamily C, member 6	0.0000	0.0000	0.0008	0.0131
227875_at	kelch-like 13 (Drosophila)	0.0000	0.0000	0.0023	0.0156
205286_at	transcription factor AP-2 gamma (activating enhancer binding protein 2 gamma)	0.0000	0.0000	0.0046	0.0189
242737_at	RAD51-like 1 (S. cerevisiae)	0.0000	0.0000	0.0062	0.0206
209965_s_at	RAD51-like 3 (S. cerevisiae)	0.0000	0.0000	0.0008	0.0131
202007_at	nidogen 1	0.0000	0.0000	0.0015	0.0144
221731_x_at	chondroitin sulfate proteoglycan 2 (versican)	0.0000	0.0000	0.0077	0.0219
244813_at	RAD51-like 1 (S. cerevisiae)	0.0000	0.0000	0.0015	0.0144
201310_s_at	chromosome 5 open reading frame 13	0.0000	0.0000	0.0008	0.0131
210258_at	regulator of G-protein signalling 13	0.0000	0.0000	0.0008	0.0131
202589_at	thymidylate synthetase	0.0000	0.0000	0.0054	0.0196
228766_at	gb:AW299226	0.0000	0.0000	0.0054	0.0196
218380_at	NLR family, pyrin domain containing 1	0.0000	0.0000	0.0008	0.0131
201417_at	SRY (sex determining region Y)-box 4	0.0000	0.0000	0.0015	0.0144
215972_at	Prostate androgen-regulated transcript 1	0.0000	0.0000	0.0093	0.0231
212942_s_at	KIAA1199	0.0000	0.0000	0.0046	0.0189
202966_at	calpain 6	0.0000	0.0000	0.0108	0.0239
205943_at	tryptophan 2,3-dioxygenase	0.0000	0.0000	0.0015	0.0144
213668_s_at	SRY (sex determining region Y)-box 4	0.0000	0.0000	0.0015	0.0144
219454_at	EGF-like-domain, multiple 6	0.0000	0.0000	0.0008	0.0131
235503_at	ankyrin repeat and SOCS box-containing 5	0.0000	0.0000	0.0069	0.0212
222834_s_at	guanine nucleotide binding protein (G protein), gamma 12	0.0000	0.0000	0.0008	0.0131
210198_s_at	proteolipid protein 1 (Pelizaeus-Merzbacher disease, spastic paraplegia 2, uncomplicated)	0.0000	0.0000	0.0015	0.0144
220565_at	chemokine (C-C motif) receptor 10	0.0000	0.0000	0.0008	0.0131
237671_at	RAD51-like 1 (S. cerevisiae)	0.0000	0.0000	0.0093	0.0231
201220_x_at	C-terminal binding protein 2	0.0000	0.0000	0.0039	0.0180
217771_at	golgi phosphoprotein 2	0.0000	0.0000	0.0015	0.0144
224002_s_at	FK506 binding protein 7	0.0000	0.0000	0.0008	0.0131
213170_at	glutathione peroxidase 7	0.0000	0.0000	0.0008	0.0131
211980_at	collagen, type IV, alpha 1	0.0000	0.0000	0.0031	0.0167
211981_at	collagen, type IV, alpha 1	0.0000	0.0000	0.0031	0.0167
212282_at	transmembrane protein 97	0.0000	0.0000	0.0008	0.0131
2013090_x_at	chromosome 5 open reading frame 13	0.0000	0.0000	0.0015	0.0144
211917_s_at	prolactin receptor///prolactin receptor	0.0000	0.0000	0.0008	0.0131
212281_s_at	transmembrane protein 97	0.0000	0.0001	0.0008	0.0131
231930_at	ELMO/CED-12 domain containing 1	0.0000	0.0001	0.0123	0.0248
205347_s_at	thymosin-like 8	0.0000	0.0001	0.0015	0.0144
223571_at	C1q and tumor necrosis factor related protein 6	0.0000	0.0001	0.0015	0.0144
204619_s_at	chondroitin sulfate proteoglycan 2 (versican)	0.0000	0.0001	0.0046	0.0189
231741_at	endothelial differentiation, sphingolipid G-protein-coupled receptor, 3	0.0000	0.0001	0.0054	0.0196

Conclusion

We have considered studies where microarray data for K treatment groups are collected from the same subjects (blocks). We discover the genes differentially expressed among K groups using non-parametric F-statistics for blocked one-way ANOVA while controlling the FDR. We employ a permutation method to generate the null distribution of the F-statistics without a normal distribution assumption for the gene expression data. The permutation-based multiple testing procedure can be easily modified for controlling the familywise error rate, see e.g. Westfall and Young [28] and Jung *et al*. [29].

We propose a simple sample size calculation method to estimate the required number of subjects (blocks) given the total number of genes m, number of differentially expressed genes m_1 and their standardized effect sizes ($\bar{\delta}_{ki}$, $1 \le i \le m_1$, $1 \le k \le K$) and the number of true rejections r_1 at a specified FDR level q^*. Through simulations and analysis of a real data set, we found that the permutation-based analysis method controls the FDR accurately and the sample size formula performs accurately. While we specify the individual effect sizes for the prognostic genes, some investigators [30,31] use a mixture model for the marginal p-values by specifying a distribution for the effect sizes among m genes.

Glueck *et al*. [32] propose an exact calculation of average power for the Benjamini-Hochberg [2] procedure for controlling the FDR. Their formula may is useful for deriving sample sizes when the test statistics are independent and the number of hypotheses m is small. However, it is not appropriate for designing a microarray study with a large number of dependent test statistics.

A sample size calculation program in R is available from http://www.duke.edu/~is29/BlockANOVA/.

Appendix

We want to prove that F_i converges to $(K-1)^{-1}\sigma_i^2 \chi_{K-1}^2(n\sum_{k=1}^{K}\bar{\delta}_{ik}^2)$ in distribution regardless of the normal distribution assumption on $_{ijk}$ and γ_{ij}. We only assume that $E(\varepsilon_{ijk}^4) < \infty$. The following is one of key lemmas used to derive the distribution of the F-statistics in the standard ANOVA theory, see e.g. Section 3b.4 of Rao [33].

<u>Lemma:</u> Suppose that, for $k = 1,..., K$, z_k are independent $N(\mu_k, 1)$ random variables and A is an idempotent $K \times K$

matrix with rank v. Let $z = (z_1,..., z_K)^T$ and $\mu = (\mu_1,..., \mu_K)^T$. Then, $z^T A z \sim \chi_v^2(\mu^T A \mu)$.

We have

$$\frac{1}{(K-1)(n-1)}\sum_{j=1}^{n}\sum_{k=1}^{K}(Y_{ijk} - \bar{Y}_{ij.} - \bar{Y}_{i.k} + \bar{Y}_{i..})^2 = \frac{1}{(K-1)(n-1)}\sum_{j=1}^{n}\sum_{k=1}^{K}(\varepsilon_{ijk} - \bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i.k} + \bar{\varepsilon}_{i..})^2$$

$$= \frac{1}{(K-1)(n-1)}(\sum_{j=1}^{n}\sum_{k=1}^{K}\varepsilon_{ijk}^2 - K\sum_{j=1}^{n}\bar{\varepsilon}_{ij.}^2 - n\sum_{k=1}^{K}\bar{\varepsilon}_{i.k}^2 + nK\bar{\varepsilon}_{i..}^2),$$

where $\bar{\varepsilon}_{ij.} = K^{-1}\sum_{k=1}^{K}\varepsilon_{ijk}$, $\bar{\varepsilon}_{i.k} = n^{-1}\sum_{j=1}^{n}\varepsilon_{ijk}$ and

$\bar{\varepsilon}_{i..} = (Kn)^{-1}\sum_{j=1}^{n}\sum_{k=1}^{K}\varepsilon_{ijk}$. By the strong law of large numbers, we have $n^{-1}\sum_{j=1}^{n}\varepsilon_{ijk}^2 \to \sigma_i^2$, $n^{-1}\sum_{j=1}^{n}\bar{\varepsilon}_{ij.}^2 \to \sigma_i^2/K$, $\bar{\varepsilon}_{i.k} \to 0$ and $\bar{\varepsilon}_{i..} \to 0$ almost surely (a.s.).

Hence,

$$\frac{1}{(K-1)(n-1)}\sum_{j=1}^{n}\sum_{k=1}^{K}(Y_{ijk} - \bar{Y}_{ij.} - \bar{Y}_{i.k} + \bar{Y}_{i..})^2 \to \sigma_i^2 \text{ a.s.}$$

Let $z_k = \sigma_i^{-1}\sqrt{n}(\delta_{ik} + \bar{\varepsilon}_{i.k})$ and $\bar{z} = K^{-1}\sum_{k=1}^{K}z_k = (K\sigma_i)^{-1}\sum_{k=1}^{K}\bar{\varepsilon}_{i.k}$. Then, $z_1,..., z_K$ are independent and, by the central limit theorem, z_k is approximately $N(\bar{\delta}_{ik}\sqrt{n}, 1)$. Let I be the $K \times K$ identity matrix, $\mathbf{1} = (1,..., 1)^T$ the $K \times 1$ vector with components 1, $z = (z_1,..., z_K)^T$ $A = I - K^{-1}\mathbf{11}^T$. Note that A is an idempotent matrix with rank $K - 1$ and $\mu^T A \mu = n\sum_{k=1}^{K}\bar{\delta}_{ik}^2$, where $\mu = \sqrt{n}(\bar{\delta}_{i1},..., \bar{\delta}_{ik})^T$. Then, $\sum_{k=1}^{K}(z_k - \bar{z})^2 = z^T A z$ is approximately distributed as $\chi_{K-1}^2(n\sum_{k=1}^{K}\bar{\delta}_{ik}^2)$ by the lemma. Since $\sum_{k=1}^{K}(\bar{Y}_{i.k} - \bar{Y}_{i..})^2 = \sigma_i^2\sum_{k=1}^{K}(z_k - \bar{z})^2$, $\sum_{k=1}^{K}(\bar{Y}_{i.k} - \bar{Y}_{i..})^2$ is approximately distributed as $\sigma_i^2\chi_{K-1}^2(n\sum_{k=1}^{K}\bar{\delta}_{ik}^2)$. By combining this result with (A.1) using the Slutsky's theorem, we complete the proof.

Authors' contributions

SJ proposed the research project and wrote the manuscript. IS performed statistical analysis. SLG supported the

research and participated in the writing of the manuscripted. PCL was responsible for the study design, conduct and oversight of the experiments and interpretation of results. She contributed to the preparation of the manuscript. LF was responsible for preparing the tissue samples for microarray analysis and interpretation of results and in manuscript preparation. The authors are solely responsible for the content of this study. All authors read and approved the final manuscript.

References

1. Catherino WH, Leppert PC, Segars JH: **The promise and perils of microarray analysis.** *Am J Obstet Gynecol* 2006, **195:**389-393.
2. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *JR Statist Soc B* 1995, **57:**289-300.
3. Westfall PH, Wolfinger RD: **Multiple tests with discrete distributions.** *American Statistician* 1997, **51:**3-8.
4. Lee MLT, Whitmore GA: **Power and sample size for DNA microarray studies.** *Stat Med* 2002, **22:**3543-3570.
5. Muller P, Parmigiani G, Robert C, Rousseau J: **Optimal sample size for multiple testing: the case of gene expression microarrays.** *J Am Stat Assoc* 2004, **99:**990-1001.
6. Jung SH: **Sample size for FDR-control in microarray data analysis.** *Bioinformatics* 2005, **21:**3097-3103.
7. Pounds S, Cheng C: **Sample size determination for the false discovery rate.** *Bioinformatics* 2005, **21:**4263-4271.
8. Liu P, Hwang JTG: **Quick calculation for sample size while controlling false discovery rate with application to microarray analysis.** *Bioinformatics* 2007, **23:**739-746.
9. Shao Y, Tseng CH: **Sample size calculation with dependence adjustment for FDR-control in microarray studies.** *Stat Med* 2007, **26:**4219-4237.
10. Feng L, Walls M, Sohn I, Behera M, Jung SH, Leppert PC: **Novel approach to the analysis of genome-wide expression profiling.** *Reproductive Sciences* 2008, **15:**298A.
11. Ge Y, Dudoit S, Speed TP: **Resampling-based multiple testing for microarray data hypothesis.** *Test* 2003, **12:**1-44.
12. Storey JD: **A direct approach to false discovery rates.** *J of Roy Stat Soc Ser B* 2002, **64:**479-498.
13. Jung SH, Jang W: **How accurately can we control the FDR in analyzing microarray data?** *Bioinformatics* 2006, **22:**1730-1736.
14. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data.** *Biostatistics* 2003, **4:**249-264.
15. Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance.** *Bioinformatics* 2003, **19:**185-193.
16. Hartigan JA: **Clustering Algorithms.** Wiley: New York; 1975.
17. Catherino WH, Prupas C, Tsibris JC, Leppert PC, Payson M, Nieman LK, Segars JH: **Strategy for elucidating differentially expressed genes in leiomyomata identified by microarray technology.** *Fert Steril* 2003, **80:**282-290.
18. Skubitz K, Skubitz AP: **Differential gene expression in uterine leiomyoma.** *Journal of Laboratory and Clinical Medicine* 2003, **141:**279-308.
19. Leppert PC, Catherino WH, Segars JH: **A new hypothesis about the origin of uterine fibroids based on gene expression profiling with microarrays.** *Am J Obstet Gynecol* 2006, **195:**415-420.
20. Rogers R, Norian J, Malki M, Christman G, Abu-Asab M, Chen F, Korecki C, Iatridis J, Catherino WH, Dhillon N, Leppert P, Segars JH: **Mechanical homeostatis is altered in uterine leiomyoma.** *Am J Obstet Gynecol* 2008, **198:**474.e1-11.
21. Malik M, Webb J, Catherino WH: **Retinoic acid treatment of human leiomyoma cells transformed the cell phenotype to one strongly resembling myometrial cells.** *Clin Endocrinol (Oxf)* 2008 in press.
22. Yagi Y, Andoh A, Inatomi O, Tsujikawa T, Fujiyama Y: **Inflammatory responses induced by interleukin-17 family members in human colonic subepithelial myofibroblasts.** *J Gastroenterol* 2007, **42:**746-753.
23. Cortez DM, Feldman MD, Mummidi S, Valente AJ, Steffenses B, Vincenti M, Barnes J, Chandrasekar B: **IL-17 stimulates MMP-1 expression in primary human cardiac fibroblasts via p38 MAPK- and ERK 1/2-dependent C/EBP-, NF-kB, and AP-1 activation.** *Am J Physiol Heart Circ Physiol* 2007, **293:**3356-3365.
24. Schoenmakers EFPM, Huysmanns C, Ven WJM Van de: **Allelic knockout of novel spice variants of human recombination repair gene RAD51B in t(12;14) uterine leiomyomas.** *Cancer Res* 1999, **59:**19-23.
25. Reeves R: **Molecular biology of HMGA proteins: hubs of nuclear function.** *Gene* 2001, **277:**63-81.
26. Gatas GJ, Quade BJ, Nowak RA, Morton CC: **HMGIC expression in human adult and fetal tissues and in uterine leimomyomata.** *Genes Chromosomes and cancer* 1999, **25:**316-322.
27. Peng Y, Laser J, Shi G, Mittal K, Melamed J, Lee P, Wei J-J: **Antiproliferative effects by Let-7 repression of high-mobility group A2 in uterine leiomyoma.** *Mol Cancer Res* 2008, **6:**663-673.
28. Westfall PH, Young SS: **Resampling-based Multiple Testing: Examples and Methods for P-value Adjustment.** Wiley: New York; 1993.
29. Jung SH, Bang H, Young S: **Sample size calculation for multiple testing in microarray data analysis.** *Biostatistics* 2005, **6:**157-169.
30. Hu J, Zou F, Wright FA: **Practical FDR-based sample size calculations in microarray experiments.** *Bioinformatics* 2005, **21:**3264-3272.
31. Jørstad TS, Midelfart H, Bones AM: **A mixture model approach to sample size estimation in two-sample comparative microarray experiments.** *BMC Bioinformatics* 2008, **9:**117.
32. Glueck DH, Mandel J, Karimpour-Fard A, Hunter L, Muller K: **Exact calculations of average power for the Benjamini-Hochberg procedure.** *International Journal of Biostatistics* 2008, **4:**Article 11.
33. Rao CR: **Linear Statistical Inference and Its Applications.** Wiley: New York; 1965.