Journal of Data Science **3**(2005), 331-352

A Unified Computational Framework to Compare Direct and Sequential False Discovery Rate Algorithms for Exploratory DNA Microarray Studies

Danh V. Nguyen
University of California, Davis

Abstract: The problem of detecting differential gene expression with microarray data has led to further innovative approaches to controlling false positives in multiple testing. False discovery rate (FDR) has been widely used as a measure of error in this multiple testing context. Direct estimation of FDR was recently proposed by Storey (2002, *Journal of the Royal Statistical Society, Series B* **64**, 479-498) as a substantially more powerful alternative to the traditional sequential FDR controlling procedure, pioneered by Benjamini and Hochberg (1995, *Journal of the Royal Statistical Society, Series B* **57**, 289-300). Direct estimation to FDR requires fixing a rejection region of interest and then conservatively estimating the associated FDR. On the other hand, sequential FDR procedure requires fixing a FDR control level and then estimating the rejection region. Thus, sequential and direct approaches to FDR control appear very different. In this paper, we introduce a unified computational framework for sequential FDR methods and propose a class of more powerful sequential FDR algorithms, that link the direct and sequential approaches. Under the proposed unified compuational framework, both approaches simply approximate the least conservative (optimal) *sequential* FDR procedure. We illustrate the FDR algorithms and concepts with some numerical studies (simulations) and with two real exploratory DNA microarray studies, one on the detection of molecular signatures in *BRCA*-mutation breast cancer patients and another on the detection of genetic signatures during colon cancer initiation and progression in the rat.

Key words: BRCA-mutation, breast cancer, colon cancer, DNA microarray, false discovery rate, fatty acids, gene expression, multiple hypothesis testing, p-value.

1. Introduction and Background

Since the introduction of DNA microarray technology, microarray applications in research have flourished, especiallly in biomedical research. See Nguyen *et al.* (2002) for a more thorough description of DNA microarrays and their applications. One common application in microarray studies involves identifying differentially expressed genes between two or more biological conditions. An example, which we will consider in this paper, involves identifying differentially expressed genes in hereditary breast cancer patients with mutation in the *BRCA1* gene relative to patients with *BRCA2*-mutation. In this context, the null hypothesis of no differential expression is tested for each gene. Because the number of genes (or gene probes) are in the thousands, controlling for errors in this multiple testing situation is very important. Thus, methods that are able to identify truly alternative hypotheses (i.e., differentially expressed genes under a defined experimental condition) with few false positives are highly desirable. The application of such methods can result in a substantial reduction in research costs and efforts during the post-analysis and follow-up validation phase of microarray experiments (Chuaqui *et al.*, 2002).

Table 1: Notations for the possible outcomes of testing m hypotheses (Benjamini and Hochberg 1995). The proportion of true null hypotheses is $\pi_0 \equiv m_0/m$ and $\text{FDR} = E(\frac{V}{R} I_{\{R>0\}})$.

	Accept	Reject	Total
Null true	U	V	m_0
Alternative true	T	S	m_1
Total	W	R	m

A particularly promising measure of error in multiple testing is the false discovery rate (FDR), the expected proportion of false discoveries among R discoveries or rejections, introduced by Benjamini and Hochberg (1995). More precisely, FDR is defined as

$$\text{FDR} = E\left(\frac{V}{R} I_{\{R>0\}}\right) = E\left(\frac{V}{R} \,\middle|\, R > 0\right) \Pr(R > 0), \tag{1.1}$$

where V is the number of erroneous rejections (Type I errors), R is the total number of rejections and $I_{\{A\}}$ denotes the indicator function for event A. Please refer to Table 1, where the outcomes for testing m hypotheses are summarized. Note in Table 1 that only R, W, and m are observable and all other quantities in the table are not observable. Also note that the well-known familywise error rate (FWER), the probability of rejecting any null hypothesis erroneously, is

$\Pr(V > 0)$. Thus, FDR provides a much less strict criterion for control than the FWER in multiple testing. Hence, an obvious substantial gain in power is expected when controlling FDR compared to controlling FWER (Benjamini and Hochberg, 1995). The FWER is inappropriately strict for exploratory microarray studies, where m is in the thousands.

The traditional sequential approach to FDR, introduced by Benjamini and Hochberg (1995) (herein BH), requires fixing a FDR level of control, say α ($0 < \alpha < 1$). Denote the ordered observed p-values as $p_{(1)}, \ldots, p_{(m)}$. The Benjamini and Hochberg FDR (BH-FDR) controlling procedure is to find

$$\hat{k}_{\mathrm{BH}} = \max\left\{ j : p_{(j)} \leq \frac{j}{m}\alpha \right\} \tag{1.2}$$

and reject $p_{(1)}, \ldots, p_{(\hat{k}_{\mathrm{BH}})}$, where $\alpha \in (0, 1)$ is the pre-specified target control level. BH proved that procedure (1.2) results in FDR $\leq \pi_0\alpha$ for $0 \leq m_0 \leq m$, where $\pi_0 = m_0/m$ is the proportion of true null hypotheses. (It was later shown by Finner and Roters (2001) that the procedure (1.2) gives precisely FDR $= \pi_0\alpha$.) Since $0 \leq \pi_0 \leq 1$, it follows that FDR is controlled at level α for all configuration of m_0. However, note that the level of control is actually $\pi_0\alpha$, which is less than or equal to α. Thus, the BH-FDR controlling procedure (1.2) is increasingly conservative as π_0 approaches zero. This leads to a loss in power to detect true alternative hypotheses. Therefore, incorporating a less conservative, hence, more precise estimate of π_0 into the FDR controlling procedure (1.2) can improve power, as was done in Benjamini, Krieger, and Yekutieli (2001) and Benjamini and Hochberg (2000).

Storey (2002; 2003) and Storey and Tibshirani (2003) also recognized that estimation of π_0 is critical in the context of DNA microarray applications; however, they proposed a non-sequential, direct approach to estimate the FDR for a fixed rejection region. Storey (2002) proposed a direct estimate of FDR for a fixed rejection region $[0, \gamma]$. More precisely, the proposed estimator of FDR is

$$\widehat{\mathrm{FDR}}_\lambda(\gamma) = \hat{\pi}_0(\lambda)\frac{\gamma}{\widehat{\Pr}(P \leq \gamma)}, \tag{1.3}$$

where $\widehat{\Pr}(P \leq \gamma) = \#\{p_j \leq \gamma\}/m$ and $\hat{\pi}_0(\lambda)$ is a conservatively biased estimator of π_0 and P is a random p-value from any test. Under independence, the main case considered in this paper, P comes from the null distribution (i.e. $P \sim$ Uniform$[0, 1]$) with probability π_0. P is from the alternative distribution with probability $1 - \pi_0$. We will elaborate on the estimator $\hat{\pi}_0(\lambda)$ in Section 2.2. The estimator (1.3) is conservatively designed in the sense that $E[\widehat{\mathrm{FDR}}_\lambda(\gamma)] \leq \mathrm{FDR}(\gamma)$ for all γ and π_0 (Storey, 2002; Theorem 2).

We note that the benefit of estimating π_0 (or equivalently m_0) in multiple testing has previously been recognized, and dates back to at least Schweder and Spjøvtoll (1982). More theoretical studies of the sequential FDR procedure (1.2) and direct FDR estimation (1.3) can be found in Genovese and Wasserman (2002) and Storey, Taylor, and Siegmund (2004) respectively.

In this paper we propose a unified computational framework for numerical studies of sequential FDR methods for independent p-values based on estimation of π_0. We also introduce a more powerful class of sequential FDR algorithms. We show that the BH-FDR procedure along with other sequential procedures, such as the two stage FDR procedure (Benjamini, Krieger, and Yekutieli, 2001), fall within this class. In addition, we illustrate that when using the same estimate of π_0, the power of sequential FDR methods are very similar to the new, direct estimates of FDR proposed by Storey (2002).

The organization of the paper is as follows. In Section 2 we introduce a unified computational framework for comparing the FDR procedures. Under this framework, all sequential FDR methods approximate the least conservative FDR procedure through estimation of π_0. This provides a unified framework for numerical comparisons of sequential and direct methods. We introduce a new, more powerful, family of sequential FDR algorithms in Section 2.2. Next, we illustrate these sequential FDR algorithms using two DNA microarray gene expression data sets in Section 3. In section 4 we describe a simulation framework for exploratory DNA microarray studies. Using this simulation, a comparison of the power and FDR control for the proposed sequential FDR algorithms to other sequential FDR methods is also given in Section 4. In Section 5 we compare the power of the proposed sequential FDR algorithms to the direct estimation of FDR and illustrate that the sequential and direct estimation procedures are essentially "equivalent" (in terms of power).

2. A Unified Computational Framework for FDR Procedures

For the original sequential FDR controlling procedure (1.2) proposed by BH, namely BH-FDR, we have FDR $= \pi_0\alpha$. Setting π_0 to its upper bound of one gives the desired level of FDR control, α. Thus, for the BH-FDR procedure, we can define $\hat{\pi}_0(\text{BH}) \equiv 1$ because no information about π_0 was actually utilized from the distribution of observed p-values, $\{p_j\}_{j=1}^m$. Thus, the choice of $\hat{\pi}_0(\text{BH}) = 1$ represents the *most* conservatice choice. The extreme opposite to this most conservative choice would be to use π_0 itself in the BH-FDR procedure. Clearly, this is the optimal (benchmark) choice and it is the *least* conservative FDR procedure. Of course, π_0 is unknown in practice. However, for numerical (or computational) studies it can be used as a benchmark. More precisely, for numerical studies of the performance (in terms of power) of the various FDR

procedures, the choice of $\hat{\pi}_0(\text{BH}) = 1$ and π_0 itself provide the lower and upper bound on the performance of any given FDR procedure. Thus, in this section, we introduce a unified computational view of sequential FDR procedures based on incorporation of more precise information on π_0 from the data (Section 2.1). Next, the direct estimation approach to FDR is casted into this framework via estimation of π_0 (Section 2.2). This is possible since both sequential and direct approaches to FDR approximate the optimal FDR procedure.

2.1 Approximating the least conservative FDR controlling procedure

Since the original BH-FDR provides $\text{FDR} = \pi_0\alpha \leq \alpha$, it is conservative by a factor of $\pi_0 = m_0/m$. If π_0 (m_0) is known, then the conservativeness can be corrected by applying the BH-FDR procedure at level $\alpha' = \alpha/\pi_0$, instead of α. This correction provides FDR control at level α, since $\text{FDR} = \pi_0\alpha' = \alpha$. Thus, we define the optimal or *least conservative* FDR (LC-FDR) procedure, assuming that α is known, as finding

$$\begin{aligned} \hat{k}_{\text{LC}} &= \max\left\{j : p_{(j)} \leq \frac{j}{m}\left(\frac{\alpha}{\pi_0}\right)\right\} \\ &= \max\left\{j : p_{(j)} \leq \frac{j}{m_0}\alpha\right\} \end{aligned} \tag{2.1}$$

and then rejecting the hypotheses corresponding to the p-values $p_{(1)}, \ldots, p_{(\hat{k}_{\text{LC}})}$. The LC-FDR procedure, although useless in practice because π_0 is unknown, provides the benchmark for studying the precision of estimating FDR and the power to detect true positives (or true alternative hypotheses).

Thus, we can define a class of sequential FDR controlling procedures that approximate the LC-FDR procedure as finding

$$\hat{k} = \max\left\{j : p_{(j)} \leq \frac{j}{m}\left(\frac{\alpha^*}{\hat{\pi}_0}\right)\right\} \tag{2.2}$$

and then rejecting $p_{(1)}, \ldots, p_{(\hat{k})}$ such that $\text{FDR} \leq \alpha$, where $\hat{\pi}_0$ is a conservative estimate of π_0, $\alpha^* \in (0, 1)$, and $\alpha \in (0, 1)$ is the target FDR level of control. Under this framework, the least conservative data-based estimate of $\hat{\pi}_0$ such that $\text{FDR} \leq \alpha$ is desirable.

For example, the original sequential FDR controlling procedure, namely BH-FDR, trivially falls into class (2.2) with $\hat{\pi}_0$ given by $\hat{\pi}_0(\text{BH}) \equiv 1$ and $\alpha^* = \alpha$:

$$\hat{k}_{\text{BH}} = \max\left\{j : p_{(j)} \leq \frac{j}{m}\left(\frac{\alpha}{\hat{\pi}_0(\text{BH})}\right)\right\}.$$

As pointed out earlier, it is the most conservative FDR controlling procedure according to (2.2).

Recognizing that a gain in power to detect true positives would result from estimating π_0 less conservatively than $\hat{\pi}_0(\text{BH}) \equiv 1$, Benjamini, Krieger, and Yekutieli (2001) (herein BKY) proposed a novel two-stage FDR (2S-FDR) controlling procedure, where π_0 is estimated from stage 1. (See also related *adaptive* FDR procedures (Benjamini and Hochberg, 2000)). The procedure is as follows.

1. Let r_1 be the number of rejections from applying the BH-FDR procedure at level $\alpha' = \alpha/(1+\alpha)$. BKY proposed estimating π_0 by $\hat{\pi}_0(\text{BKY}) = (m - r_1)/m$.

2. Next, apply the BH-FDR procedure again (using the same data as in part 1), but at level $\alpha'/\hat{\pi}_0(\text{BKY})$:

$$\begin{aligned} \hat{k}_{\text{BKY}} &= \max\left\{ j : p_{(j)} \leq \frac{j}{m}\left(\frac{\alpha'}{\hat{\pi}_0(\text{BKY})}\right)\right\} \\ &= \max\left\{ j : p_{(j)} \leq \frac{j}{m - r_1}\alpha'\right\} \end{aligned} \tag{2.3}$$

and reject $p_{(1)}, \ldots, p_{(\hat{k}_{\text{BKY}})}$. Note that if $r_1 = 0$ no hypothesis is rejected, and if $r_1 = m$ then all m hypotheses are rejected. In both cases, the procedure terminates at stage 1. Note that the 2S-FDR procedure falls into class (2.2) with $\alpha^* = \alpha' = \alpha/(1+\alpha)$ and $\hat{\pi}_0 = \hat{\pi}_0(\text{BKY})$. It has been proven that for the BH-FDR and 2S-FDR procedure, FDR $\leq \alpha$ for all $0 \leq \pi_0 \leq 1$ (Benjamini and Hochberg, 1995; Benjamini, Krieger, and Yekutieli, 2001).

Since the choice of $\alpha^* = \alpha'$ is more strict than $\alpha^* = \alpha$ (i.e. $\alpha' < \alpha$), BKY also proposed the following two-stage *modified* FDR (2SM-FDR) procedure

$$\hat{k}_{\text{BKY-M}} = \max\left\{ j : p_{(j)} \leq \frac{j}{m}\left(\frac{\alpha'}{\hat{\pi}_0(\text{BKY-M})}\right)\right\},$$

where $\hat{\pi}_0(\text{BKY-M})$ is the first stage estimate of π_0 at level α instead of α'.

Note that under this unified view, given by (2.2), FDR methods, such as the BH-FDR, 2S-FDR, and 2SM-FDR, essentially "mimic" (or approximate) the least conservative FDR process. More precisely, they use estimates of the form $\alpha^*/\hat{\pi}_0(\cdot)$ as a plug-in for α/π_0, where α^* is some pre-chosen level. Thus, we can view these FDR controlling procedures as approximations to the LC-FDR procedure. In this setting, the least conservative data-based estimate of π_0 such that FDR $\leq \alpha$ is desirable. Next, we describe a new family of sequential FDR algorithms that provide a better approximation of the LC-FDR procedure. This procedure uses the estimator of π_0 proposed by Storey (2002) in his direct estimation approach to FDR.

2.2 A "new" family of sequential FDR algorithms: Linking direct FDR estimation to sequential FDR procedures

A gain in power to detect true alternative hypotheses will require a less conservative estimate of π_0. This gain in power was first shown by Storey (2002) for independent p-values and illustrated with extensive simulation by Nguyen (2004), including cases involving violation of key assumptions of the FDR framework. In this section we propose a family of sequential "FDR algorithms" by utilizing the estimator of π_0 proposed by Storey (2002) and further implemented by Storey and Tibshirani (2003) in their direct estimation approach to FDR.

The estimator of π_0 proposed by Storey (2002) is as follows. Since it is much more likely that very large p-values correspond to true null hypotheses, consider the set of large p-values falling into the upper interval $(\lambda, 1]$ to estimate π_0 (for some chosen $0 < \lambda < 1$). Furthermore, note that if no genes are differentially expressed, then the null p-values are uniformly distributed, denoted $P \sim \mathrm{U}(0,1)$. Hence, $\Pr\{P \in (\lambda, 1]\} = \Pr(P > \lambda) = 1 - \lambda$. It follows that the expected number of null p-values that would fall into the interval $(\lambda, 1]$ is $(1-\lambda)m_0$. In addition, if we know the number of null p-values in $(\lambda, 1]$, $\#\{\text{Null } p_j > \lambda\}$, then an unbiased estimate of π_0 is

$$\hat{\pi}_0(\mathrm{UB}) = \frac{\#\{\text{Null } p_j > \lambda\}}{m(1-\lambda)}, \tag{2.4}$$

since $E[\hat{\pi}_0(\mathrm{UB})] = m_0/m = \pi_0$.

The numerator of (2.4), $\#\{\text{Null } p_j > \lambda\}$, is not observable in practice. However, replacing the numerator with $\#\{p_j > \lambda\}$, an observable quantity, leads to a conservatively biased estimate of π_0:

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_j > \lambda\}}{m(1-\lambda)}. \tag{2.5}$$

It can be seen that the estimate, $\hat{\pi}_0(\lambda)$, is conservatively biased from the following simple inequality

$$\#\{p_j > \lambda\} = \#\{\text{Null } p_j > \lambda\} + \#\{\text{Alt. } p_j > \lambda\} \geq \#\{\text{Null } p_j > \lambda\}.$$

Thus, $E[\hat{\pi}_0(\lambda)] \geq E[\hat{\pi}_0(\mathrm{UB})] = \pi_0$. Note that the index parameter, λ, is actually a tuning parameter which balances bias and variance. More precisely, as λ approaches 1, $\#\{p_j > \lambda\}$ consists mostly of truly *null* p-values; therefore, the bias decreases. However, the interval used to estimate $\hat{\pi}_0(\lambda)$, specifically $(\lambda, 1]$, shrinks to zero as $\lambda \to 1$; hence, the variance increases. Thus, Storey and Tibshirani (2003) proposed an automatic algorithm for choosing the optimal λ to minimize the mean squared error of $\hat{\pi}_0(\lambda)$. The optimal estimator of π_0 is denoted as $\hat{\pi}_0(\mathrm{OPT})$ and is reproduced in the Appendix for convenience.

Using the estimator $\hat{\pi}_0(\lambda)$, we propose the following family of sequential FDR algorithms (indexed by λ) to better approximate the LC-FDR controlling procedure (2.1)

$$\hat{k}_\lambda = \max\left\{j : p_{(j)} \leq \frac{j}{m}\left(\frac{\alpha}{\hat{\pi}_0(\lambda)}\right)\right\}. \tag{2.6}$$

The proposed FDR algorithm simply replaces $\hat{\pi}_0(\text{BH}) \equiv 1$ in the ordinal BH-FDR procedure with $\hat{\pi}_0(\lambda)$, a less conservative estimate of π_0. We emphasize that the FDR algorithm given by (2.6) is sequential and it utilizes the exact information (estimator) of π_0 as with the direct estimation approach. Therefore, one might expect that the power of the two approaches would be equivalent. We will elaborate further on this in Section 5, using numerical studies.

We illustrate in the next section the proposed sequential FDR algorithms with two exploratory DNA microarray studies involving: (1) the detection of molecular signatures in *BRCA*-mutation breast cancer patients and (2) the detection of genomic signatures during colon cancer initiation and progression in the rat.

3. Examples

3.1 Detection of molecular signatures in *BRCA*-mutation breast cancer patients

One common application of DNA microarray technologies in biomedical research is the detection of differentially expressed genes between two or more biological conditions (groups). For example, Hendenfalk *et al.* (2001) applied cDNA microarray to the study of hereditary breast cancer. In particular, one goal of the microarray study was to identify genes, among $m = 3,226$ genes, that are differentially expressed in breast cancer patients with mutations in the *BRCA1* gene relative to patients with *BRCA2*-mutations. There were $n_1 = 7$ patients with *BRCA1*-mutation and $n_2 = 8$ patients with *BRCA2*-mutation. For each gene, we computed the two sample *t*-statistic and the corresponding *p*-value for testing the null hypothesis of no differential gene expression. Figure 3A displays the density histogram of the observed *p*-values, $\{p_j\}_{j=1}^m$. For this distribution of *p*-values, it is estimated that the proportion of true null hypotheses, π_0, is $\hat{\pi}_0(\lambda) = 0.678$ using $\lambda = 1/2$. The optimal estimate, $\hat{\pi}_0(\text{OPT})$, gave a similar estimate of 0.668. We applied sequential FDR methods, namely BH-FDR, 2S-FDR, 2SM-FDR, and the proposed FDR algorithm (2.6) using $\hat{\pi}_0(\lambda)$. For illustration, we applied the methods to control false discoveries among the m tests at 10% ($\alpha = .10$).

Note that all of the FDR sequential methods described earlier are of the following form: (1) find $\hat{k} = \max\{j : (j/m)(\alpha^*/\hat{\pi}_0)\}$ and (2) reject $p_{(1)}, \ldots, p_{(\hat{k})}$. This is equivalent to plotting the $p_{(j)}$ versus j/m, and finding the first time in the (reversed) sequence of ordered *p*-values, $\{p_{(m)}, p_{(m-1)}, \ldots, p_{(1)}\}$, that crosses

the line with slope $\alpha^*/\hat{\pi}_0$ (Genovese *et al.* 2002). All p-values below this point are rejected. For example, Figure 1 plots the ordered observed p-values (y-axis) versus j/m (x-axis). Also plotted in Figure 1 are three straight lines with slopes $\alpha/\hat{\pi}_0(\text{BH})$, $\alpha'/\hat{\pi}_0(\text{BKY})$, and $\alpha/\hat{\pi}_0(\lambda)$ corresponding to the BH-FDR, 2S-FDR, and the proposed FDR method using $\hat{\pi}_0(\lambda)$ respectively. For the proposed FDR method using $\hat{\pi}_0(\lambda)$, 281 genes were identified as differentially expressed, while the BH-FDR identified only 162 differentially expressed genes. The two-stage procedures, 2S-FDR and 2SM-FDR, identified the same number of differentially expressed genes as the BH-FDR procedure in this application.

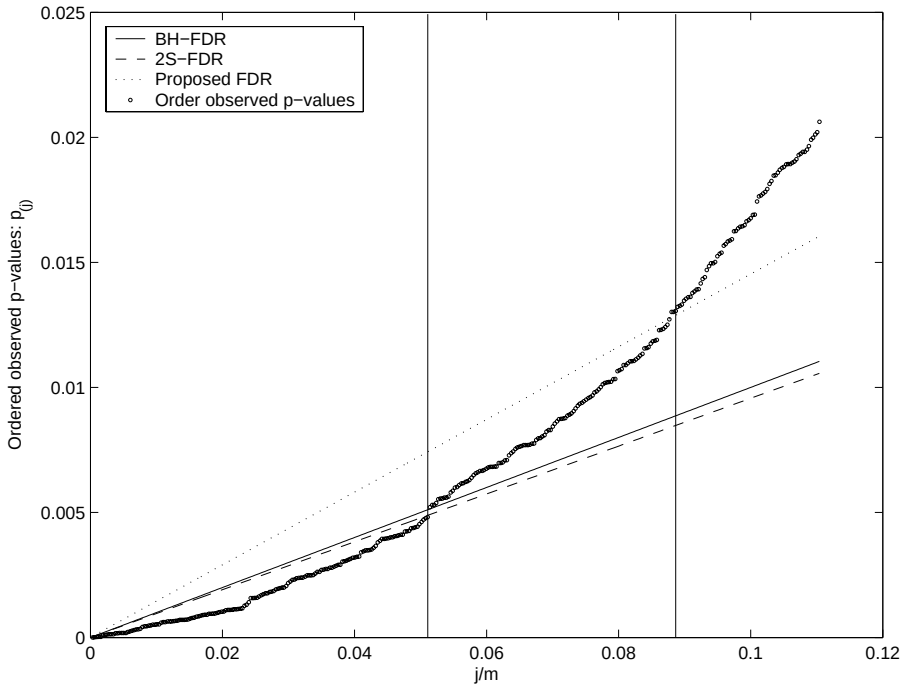

Figure 1: Sequential FDR methods applied to the *BRCA* data. The circles plotted are j/m versus $p_{(j)}$. A sequential FDR method is equivalent to finding the first time in the (reversed) sequence of ordered p-values, $\{p_{(m)}, p_{(m-1)}, \ldots, p_{(1)}\}$, that crosses the line with slope $\alpha^*/\hat{\pi}_0$. The right solid vertical line indicates this crossing, at $j = 281$, for the proposed FDR algorithm (dotted line). The left vertical line indicate the crossing for both the BH-FDR and the 2S-FDR procedure, which occurred at the same location ($j = 162$). It is also the same for 2SM-FDR (not shown).

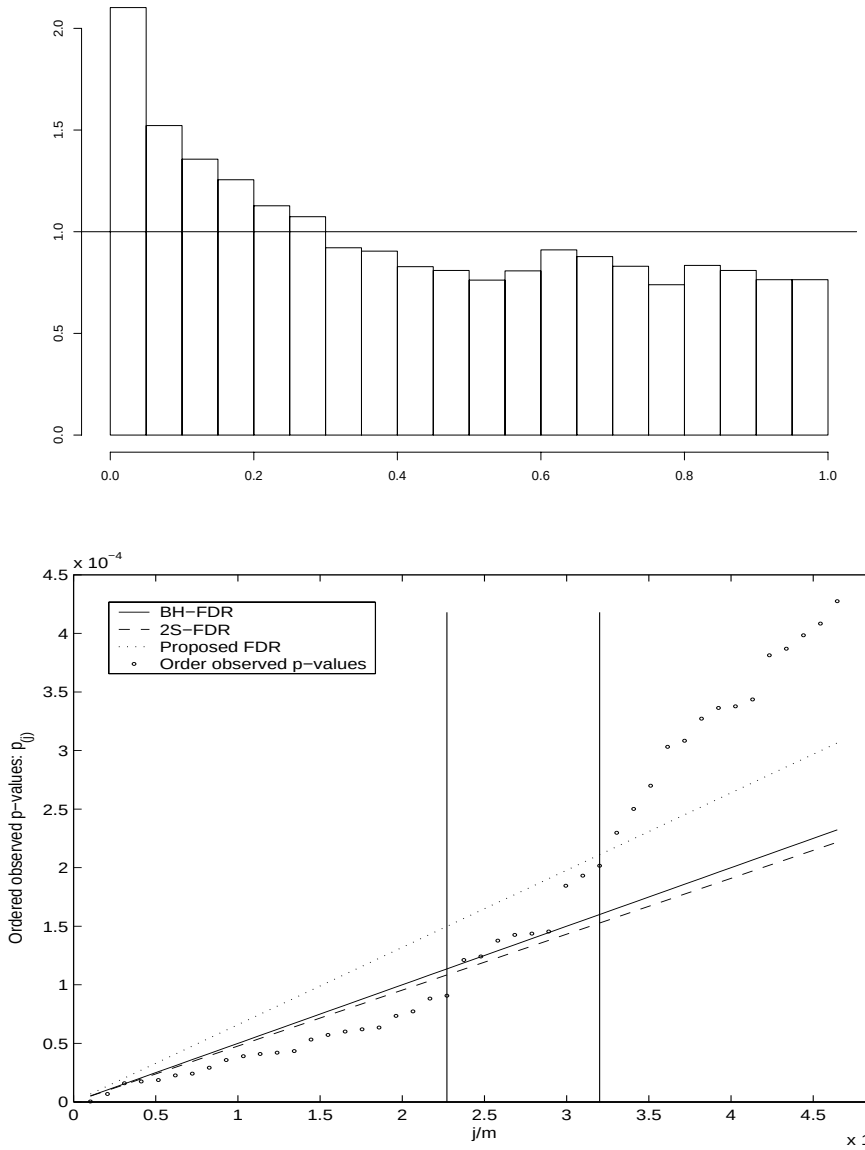

Figure 2: **Sequential FDR methods applied to the colon cancer data**. **(A, top)** Displayed is the density histogram of the 9,685 observed p-values comparing n-3 PUFA and n-6 PUFA enriched diets. **(B, bottom)** Displayed is a graphical summary of the sequential FDR analyses. Estimate of π_0, the proportion of truly non-differentially expressed genes, is $\hat{\pi_0}((\text{OPT})) = 0.758$.

3.2 Detection of genomic signatures during colon cancer initiation and progression in the rat

In a study of the chemopreventive n-3 polyunsaturated fatty acids (PUFA) as it relates to colon cancer initiation and progression in the rat, Davidson, Nguyen *et al.* (2004) used CodeLink^{TM} oligonucleotide microarrays (Ramakrishnan et al., 2002) to monitor gene expression profiles from 93 rat tissue samples. Each microarray contains 9,685 gene probes. The study utilized a $3 \times 2 \times 2$ factorial design with three types of dietary fat (rich in n-6 PUFA, n-3 PUFA or n-9 monounsaturated fatty acid (MUFA)), two treatments (injection of carcinogen azoxymethane (AOM) or saline) and two time points (12 hours (initiation) and 10 weeks (progression) post first injection). 93 RNA samples from 59 rats (including 34 replicates) were randomly assigned into the three-factor combination. It is of interest in the study to examine the molecular mechanisms by which dietary fat composition exerts tumor enhancing or inhibiting effects. More specifically, it is of interest to identify gene expressions that are significantly different between n-3 PUFA and n-6 PUFA enriched diets. For illustration, Figure 2A displays the distribution (density histogram) of the 9,685 observed p-values contrasting n-3 PUFA and n-6 PUFA enriched diets from a linear mixed model for repeated measures. Applying Storey and Tibshirani (2003) estimator for π_0, the the proportion of truly non-differentially expressed genes between n-3 PUFA and n-6 PUFA enriched diets, gives $\hat{\pi}_0(\text{OPT}) = 0.758$. Using this estimate, the proposed FDR method results in 31 differetially expressed genes at FDR level $\alpha = 0.05$. The original BH-FDR and the two-stage FDR procedures both yield 22 genes declared differetially expressed at the same FDR level.

4. Simulation Framework for Exploratory DNA Studies: Some Computational Experiments and Results

In this section we describe the simulation framework to compare FDR algorithms for exploratory DNA microarray studies. Using this framework, we compare the power of the proposed sequential FDR algorithm (2.6) to detect true alternative hypotheses. We illustrate via simulation the gain in power from using the proposed sequential FDR algorithm relative to the BH-FDR, 2S-FDR, 2SM-FDR, and, more importantly, to the least conservative FDR (LC-FDR) procedure. Recall that the LC-FDR procedure has optimal power and FDR control. We also examine the FDR control for the proposed FDR algorithm (2.6).

4.1 Simulation design considerations

The simulation was designed in the context of a two-sample comparison, anal-

ogous to the *BRCA*-mutation breast cancer example in Section 3. More specifically, we generated an $m \times n$ gene expression data matrix,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(01)}_{m_0 \times n_1} & \mathbf{X}^{(02)}_{m_0 \times n_2} \\ \mathbf{X}^{(11)}_{m_1 \times n_1} & \mathbf{X}^{(12)}_{m_1 \times n_2} \end{bmatrix}_{m \times n}.$$

Of the total m genes, m_0 are truly null and the remaining $m_1 = m - m_0$ are truly alternative.

Under the null setting, we generated the expression value for gene j in both groups $i = 1$ and 2, independently from a $\mathrm{N}(\mu_0, \sigma_0^2)$ distribution: $x_{ji} \sim \mathrm{N}(\mu_0, \sigma_0^2)$ for $j = 1, \ldots, m_0$ and $i = 1, 2$. This is the first m_0 rows of the data matrix $\mathbf{X}$ above. Under the null hypothesis, there is no difference in gene expression between groups 1 and 2 for all m_0 genes.

However, there is a difference in gene expression between groups 1 and 2 under the alternative setting. Thus, expression values for gene j in group 2, x_{j2} $(j = m_0 + 1, \ldots, m)$, were independently generated from a $\mathrm{N}(\mu_1, \sigma_1^2)$ distribution. For group 1, $x_{j1} \sim \mathrm{N}(\mu_0, \sigma_0^2)$. Rather than setting μ_1 to a fixed value, the true alternative mean expression for group 2 was allowed to vary above the null mean value of μ_0. Also, in order to more reflect the heterogeneity of variance often encountered with microarray data in practice, we similarly allowed the alternative variance parameter, σ_1^2 to vary. Table 2 summarizes the simulation model and parameters.

In this simulation framework, our intent is to test the null hypothesis that there is no differential expression for each gene between groups 1 and 2. Thus, after the data generation we computed the two sample *t*-statistic and the corresponding *p*-value for each gene, as was done in the *BRCA*-mutation breast cancer example. Because the data was generated with heterogeneity of variance, we examined both sets of *p*-values obtained from (1) erroneously assuming a t distribution for the null distribution and (2) using a permutation method to approximate the null distribution. For (1) the *p*-value for gene j was computed as $p_j = \Pr(|t(n-2)| > t_j)$, where t_j is the observed *t*-statistics and $t(n-2)$ denotes the t distribution with $n - 2$ degrees of freedom. For (2), we obtained B permutations of the n sample labels. For the bth permutation, denote the re-calculated *t*-statistics based on the permuted data by $\{t_j^{0b}\}_{j=1}^m$ $(b = 1, \ldots, B)$. The permutation-based *p*-values can be computed as $p_j = \sum_{b=1}^B \#\{k : |t_k^{0b}| \geq |t_j|, k = 1, \ldots, m\}/(mB)$, for $j = 1, \ldots, m$ (Storey and Tibshirani, 2003).

For example, Figure 3B displays the density histogram of the observed *p*-values for a simulated data set with $m = 1,000$ genes, group sample size $n_1 = n_2 = 8$, and $\pi_0 = m_0/m = 0.70$. Note that the simulated data resembles closely the *BRCA*-mutation hereditary breast cancer data described earlier in Section

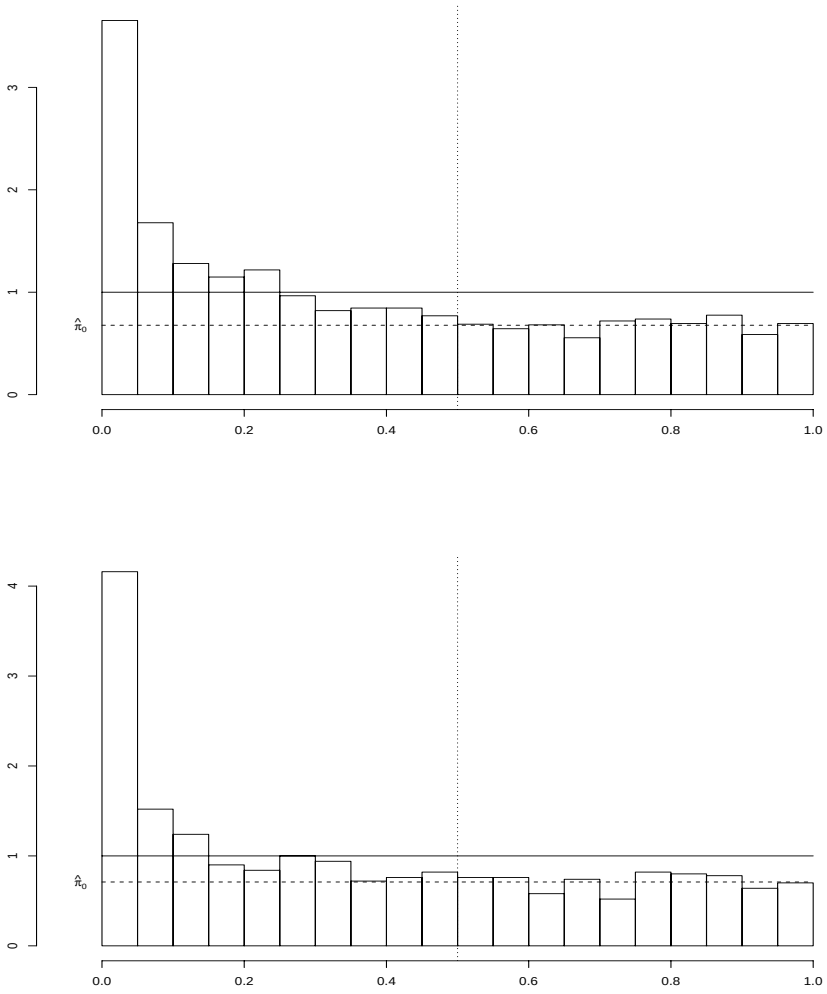

Figure 3: **(A, top) *BRCA* data**. Density histogram of the observed p-values corresponding to each gene in the *BRCA*-mutation breast cancer data. The solid horizontal line at one represents the $\mathrm{U}(0,1)$ distribution when all genes are not differentially expressed. The estimate of π_0 for $\lambda = 1/2$ (dotted vertical line) is $\hat{\pi}_0(\lambda) = \#\{p_i > \lambda\}/(m(1-\lambda)) = 0.678$ (dashed horizontal line). **(B, bottom) Simulated data**. Density histogram of the observed p-values corresponding to each gene for a simulated data with $\pi_0 = 0.70$ and $m = 1,000$ genes. The p-values correspond to two-sample t-statistics with group sample size $n_1 = n_2 = 8$. For this simulated data, $\hat{\pi}_0(\lambda = 1/2) = 0.710$ and $\hat{\pi}_0(\mathrm{OPT}) = 0.699$.

3. The estimate of the proportion of true null hypotheses is $\hat{\pi}_0(\lambda) = 0.710$ for the choice of $\lambda = 1/2$.

We summarized below some basic points of the simulation model which we believe more accurately reflect real microarray data.

- The number of genes, m, is large.
- The range of the sample size is small to moderate and reflects the range afforded in real microarray studies.
- The model encompasses heterogeneity of variance, often encountered with real microarrray data. This applies to both the null and alternative settings.
- The mean expression for genes differentially expressed is allowed to vary. Again, this applies to both the null and alternative settings.
- The above two items (alone and/or together) imply that the sampling distribution of the test statistic is not exact. Therefore, the p-values generated are only *approximations*.
- Biological theory and assumptions imply dependence in gene expression, so violation of independence should be considered in the simulation model. This was done, for example, in Nguyen (2004) and we refer the reader there for details.
- Measurement error are well recognized with gene expression measurements and can be incorporated into the simulation model. (For example, see Zien *et al.*, 2002). Although measurement error is an important issue with microarray data, we will not address this point further and will address it in future work. However, the framework described in Section 2 can be applied to any simulation model, including models with measurement errors and dependent data.

Under these more realistic simulation conditions, it will be interesting to examine the power and FDR control. For illustration, we will describe the results for the model given in Table 2 with $\mu_0 = 0$, $\sigma_0^2 = 1$, $\mu_1 \in \{\mu_{1d}\}_{d=1}^{D} = \{1, 2, 3, 4\}$ and $\sigma_1^2 \in \{\sigma_{1l}\}_{l=1}^{L} = \{1, 2, 4, 6\}$. Both μ_1 and σ_1^2 are randomly selected for each iteration (data set simulated). Also, we only describe the case where $n_i = 8$ per group, similar to the *BRCA*-mutation data set. Details and the patterns of the other cases are similar. However, some of the above cases, including violation of independence, non-normal models for gene expression, and different configuration of sample sizes can be found in Nguyen (2004), although not under the unified computational framework presented here.

Table 2: Simulation model and parameters. Data were generated such that there is no differential expression in m_0 genes from a total of m genes.

	Group 1 (n_1)	Group 2 (n_2)
True Null (m_0)	$\mathbf{X}^{(01)} : x \sim \mathrm{N}(\mu_0, \sigma_0^2)$	$\mathbf{X}^{(02)} : x \sim \mathrm{N}(\mu_0, \sigma_0^2)$
True Alt. (m_1)	$\mathbf{X}^{(11)} : x \sim \mathrm{N}(\mu_0, \sigma_0^2)$	$\mathbf{X}^{(12)} : x \sim \mathrm{N}(\mu_1, \sigma_1^2)$
Parameters	$\mu_0 \in \{\mu_{0d}\}_{d=1}^{D_1}$ $\sigma_0^2 \in \{\sigma_{0l}^2\}_{l=1}^{L_1}$	$\mu_1 \in \{\mu_{1d}\}_{d=1}^{D_2}$ $\sigma_1^2 \in \{\sigma_{1l}^2\}_{l=1}^{L_2}$

4.2 Power

Figure 4A displays the true proportion of null hypotheses, π_0, and its various conservative estimates, namely $\hat{\pi}_0$(BH), $\hat{\pi}_0$(BKY), $\hat{\pi}_0$(BKY-M), and $\hat{\pi}_0(\lambda)$ based on 10,000 simulations for each $\pi_0 \in \{0.1, 0.2, \ldots, 0.9\}$. Thus, a total of 90,000 data sets were generated. Also, for this simulation experiment we used a sample size of $n_i = 8$ per group, similar to the *BRCA* microarray data. It is clear from Figure 4A that $\hat{\pi}_0(\lambda)$ is the least conservative estimate of π_0; hence, much closer to the target, π_0. The resulting power, for the same simulation experiment, is given in Figure 4B. It is evident that the proposed FDR algorithm using $\hat{\pi}_0(\lambda)$ is substantially more powerful than the other sequential FDR methods. In addition, it is much closer to the optimal power, given by the LC-FDR procedure.

We note that there is little difference between the two stage procedure, 2S-FDR, and the modified version, 2SM-FDR.

4.3 FDR control

For the original BH-FDR controlling procedure, FDR $= \pi_0\alpha$; hence, it is conservative because it actually controls FDR at a lower level of $\pi_0\alpha$, rather at the target level of α. It follows that the actual level of FDR control for the BH-FDR procedure is linearly decreasing from α, as $\pi_0 \to 0$. On the other extreme, the least conservative or optimal procedure, namely LC-FDR, completely corrects for this conservativeness by utilizing π_0 itself. Thus, as expected and confirmed by the results in Figure 5, the LC-FDR control is at the target level of α and the BH-FDR control is lower, at precisely $\pi_0\alpha$. (For illustration, the results in Figure 5 is for a target FDR control level of $\alpha = 0.05$.) Also, the two-stage procedures control FDR below the target α; however, they are nearly as conservative as the BH-FDR control, the most conservative case.

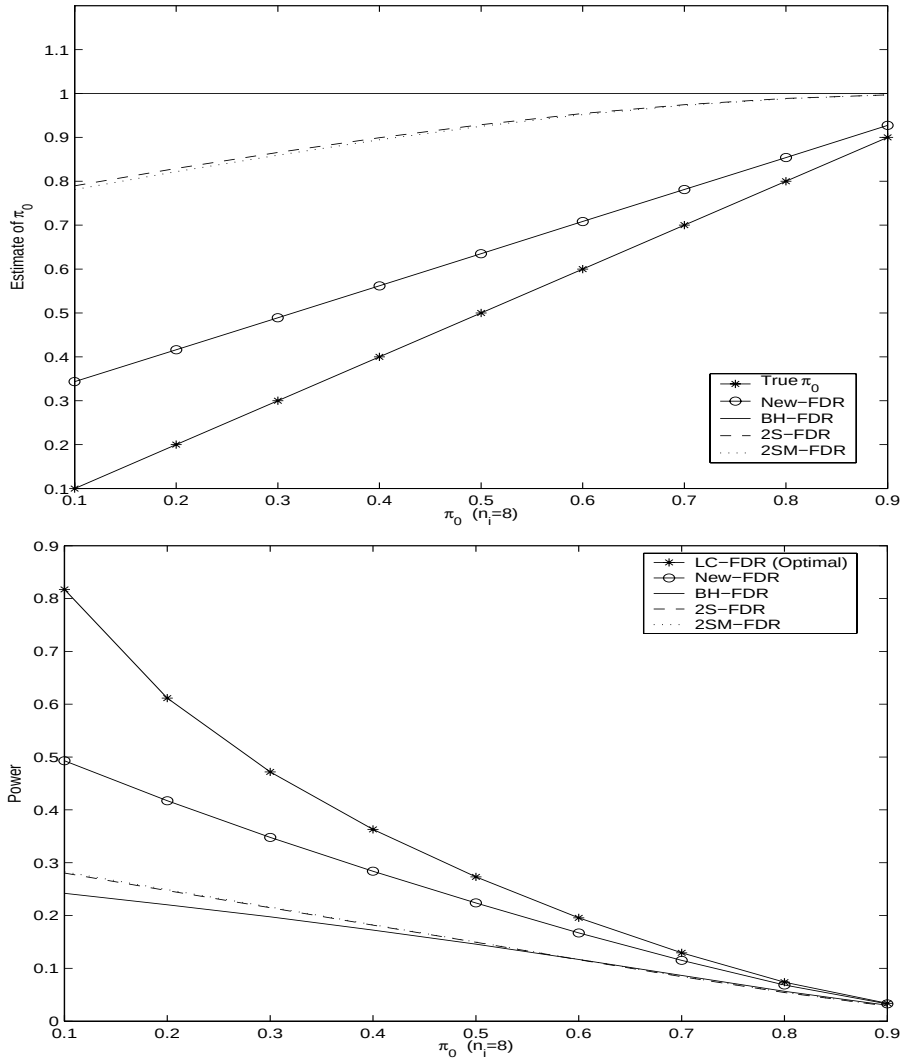

Figure 4: **(A, top) Conservativeness in estimating** π_0. Displayed are π_0 values (x-axis) versus their conservative estimates $\hat{\pi}_0$, averaged over 10,000 simulation runs, for the $\hat{\pi}_0(\text{BH}) \equiv 1$ (BH-FDR), $\hat{\pi}_0(\text{BKY})$ (2S-FDR), $\hat{\pi}_0(\text{BKY-M})$ (2SM-FDR), and $\hat{\pi}_0(\lambda)$. Also displayed for comparison are the true π_0 values $(-*-*-)$. **(B, top) Power curves**. Displayed are the corresponding power curves for the same 10,000 simulation runs, given in part (A). The proposed FDR algorithm, which uses the less conservative estimate $\hat{\pi}_0(\lambda)$, has power $(-\circ-\circ-)$ closest to the optimal power given by the least conservative FDR (LC-FDR) procedure $(-*-*-)$. Power is the proportion of true alternative hypotheses correctly rejected and was averaged over the 10,000 simulations for each π_0.

As can be seen from Figure 5, our proposed FDR algorithm (2.6), which uses Story's less conservative estimate, $\hat{\pi}_0(\lambda)$, also controls FDR at the target level of $\alpha = 0.05$. In other words, FDR $\leq \alpha$ for the proposed FDR algorithm (2.6). However, the FDR control is substantially less conservative than the other procedures, thus affording a large gain in power, as was described earlier.

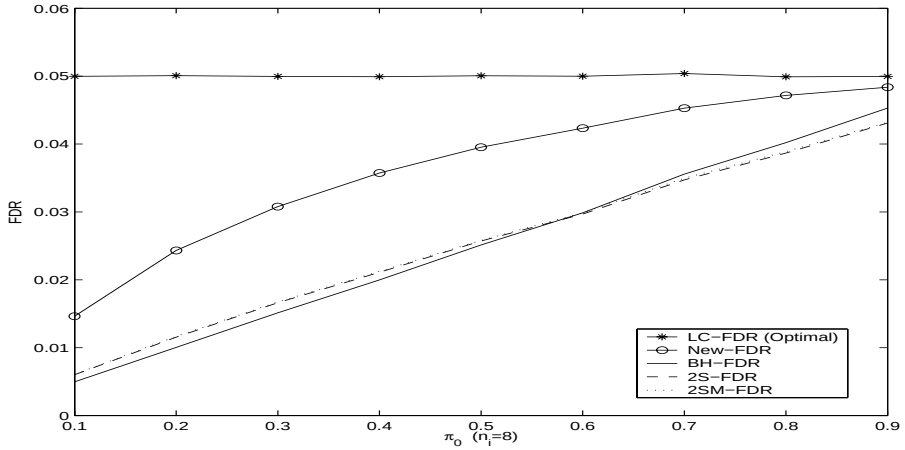

Figure 5: **FDR control for sequential methods**. Displayed are the observed levels of FDR control, averaged over the 10,000 simulation runs ($\alpha = 0.05$). Note that he BH-FDR control is lower, at level $\pi_0\alpha$. The proposed FDR algorithm using $\hat{\pi}_0(\lambda)$ is less conservative and, hence, closest to the target control level.

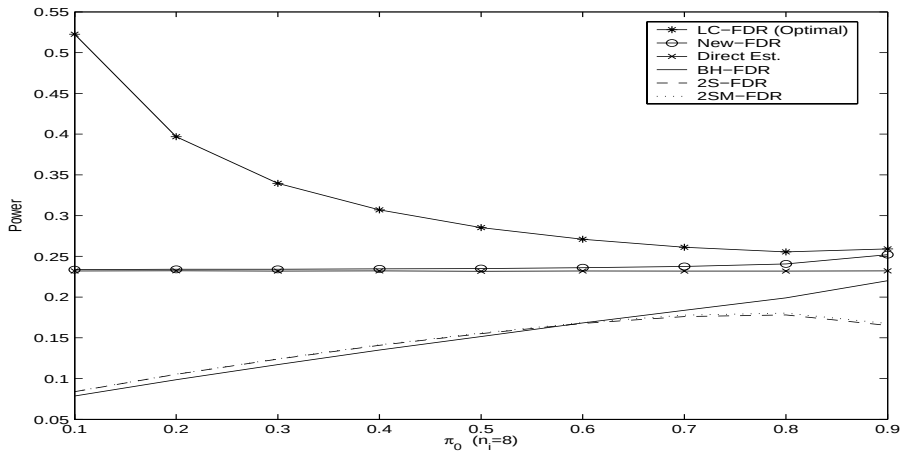

Figure 6: **FDR control for sequential methods**. Displayed are the observed levels of FDR control, averaged over the 10,000 simulation runs ($\alpha = 0.05$). Note that he BH-FDR control is lower, at level $\pi_0\alpha$. The proposed FDR algorithm using $\hat{\pi}_0(\lambda)$ is less conservative and, hence, closest to the target control level.

5. A "Reconciliation" Between Sequential and Direct FDR Approaches: Diminishing the Power Gap

Multiple testing in microarray applications often involves conducting thousands of tests simultaneously to screen for differentially expressed genes, as illustrated by the *BRCA* breast cancer data in Section 3. The widely popular sequential method, namely BH-FDR, completely ignores important information regarding π_0. Hence, it critically lacks power to detect truly alternative hypotheses, especially in microarray applications where m is extremely large. In fact, increasing information regarding π_0 becomes available when $m \to \infty$, and this is precisely the case in micrarray applications. However, the landmark paper introducing FDR by Benjamini and Hochberg (1995) did not envision applications with this scale. This is clearly apparent from their numerical study with $m \leq 64$. When m is extremely large, the investigator often must accept a high level of false discovery rate (α) in order to obtain some rejections or significant genes for follow-up studies under the BH-FDR procedure.

Storey (2002) recognizes some of the shortcomings of the BH-FDR procedure in microarray applications. He proposed estimating π_0 and FDR directly for a fixed rejection region. For example, if the investigator decides, *a priori*, to reject all genes with p-values less or equal to $\gamma = 0.005$, what is her/his expected FDR? The direct approach provides a conservatively biased estimator, $\widehat{\text{FDR}}(\gamma)$. Thus, the investigator fixes, before hand, the rejection; in this example, the rejection region is $[0, \gamma]$. The direct estimation approach to FDR (Storey, 2002) is an important and fundamental departure from the tradition sequential FDR approach in the following sense. Essentially, traditional sequential methods conservatively estimate the rejection region for a *fixed* FDR level, and the direct estimation approach conservatively estimate FDR directly for a *fixed* rejection region.

Thus, the interpretation and implementation of the two approaches are different. Differing views, advocating the practice of one approach over the other remains (Storey, 2002). However, based on numerical studies, there is a substantial gain in power from using direct estimation relative to using the sequential BH-FDR procedure. (See Figure 1 and Table 2 of Storey (2002).) This is due to the fact that the direct estimate of FDR uses a better estimate of π_0, and not to whether one fixes the rejection region (direct method) or the FDR level (sequential method). As we will illustrate, the key lies in utilizing a less conservative, hence more precise, estimate of π_0. More precisely, if we use *exactly* the same information (estimate of π_0) in both the sequential and direct FDR methods, will the differences in power of the two approaches diminish?

Storey (2002) proposed a direct estimate of FDR for a fixed rejection region $[0, \gamma]$, which uses $\hat{\pi}_0(\lambda)$. As described in the Introduction section, the direct

estimator of FDR is

$$\widehat{\mathrm{FDR}}_\lambda(\gamma) = \hat{\pi}_0(\lambda)\frac{\gamma}{\widehat{\Pr}(P \le \gamma)},$$

where $\widehat{\Pr}(P \le \gamma) = \#\{p_j \le \gamma\}/m$.

We compare the power curve of $\widehat{\mathrm{FDR}}_\lambda(\gamma)$ to the proposed sequential FDR algorithm (2.6), which uses the same estimate of π_0, namely $\hat{\pi}_0(\lambda)$. We note that sequential FDR methods cannot be directly compared with the direct estimation approach (Storey, 2002). This is because, as pointed out earlier, the former estimates the rejection region, whereas the later estimates FDR directly. However, this can be circumvented by using the sequential methods to control FDR at level $\alpha = \widehat{\mathrm{FDR}}_\lambda(\gamma)$ for each iteration (simulation run). The results, based on 10,000 simulation runs and using the same setup as described in Section 4, are given in Figure 6. Sequential methods using poor estimates of π_0, namely BH-FDR, 2S-FDR, and 2SM-FDR, have very low power; hence their power curves fall far below the optimal power curve (LC-FDR). The power for the proposed sequential FDR method (2.6) is close to the power for the direct estimator, $\widehat{\mathrm{FDR}}_\lambda(\gamma)$, because they both use the same estimate for π_0. This result is not surprising in light of the connection between the BH-FDR procedure and $\widehat{\mathrm{FDR}}_\lambda(\gamma)$ shown by Storey (2002) for independent p-values; BH-FDR is a special case of $\widehat{\mathrm{FDR}}_\lambda(\gamma)$ when $\hat{\pi}_0(\lambda) = 1$. Similarly, it is not difficult to show analytically that the proposed sequential FDR algorithm is equivalent to $\widehat{\mathrm{FDR}}_\lambda(\gamma)$ for independent p-values.

6. Discussion

We have provided a unified computational framework for comparing FDR procedures in numerical studies, encompassing traditional sequential methods and the new direct approach to FDR. We have demonstrated that when using the same information regarding π_0 in both sequential FDR methods and in direct estimation of FDR, the power curves are equivalent. In particular, we introduced a class of sequential FDR of the form $\hat{k} = \{j : p_{(j)} \le (j/m)(\alpha^*/\hat{\pi}_0)\}$ and showed that various sequential FDR methods fall within this class. In addition, when a less conservative estimate of π_0 is used, specifically $\hat{\pi}_0 = \hat{\pi}_0(\lambda)$, the power is equivalent to the direct estimator of FDR.

These conclusions also hold for other simulation configurations, in addition to those described earlier. For example, note that in the simulation we used a sample size of $n_i = 8$ per group, similar to the *BRCA* data. In the context of two sample comparison in microarray experiments, this is a moderate sample size. However, the results hold for even smaller sample sizes. Of course, as n_i grows larger, the power for all methods converge to the optimal power of the LC-FDR procedure.

We emphasize that the family of sequential FDR algorithms given by (2.6) was motivated and derived at by examining the estimation of π_0 in the *sequential FDR procedures* of Benjamimi and colleages. Thus, along this same line, utilizing the more precise estimator $\hat{\pi_0}(\lambda)$ of Storey's to improve power is sensible. However, it is important to note that the later algorithms do not guarantee the FDR control, in the sense of Bejamini and Hochberg (1955) for the BH-FDR procedure. Thus, one needs to take care to examine the FDR control in numerical studies (on average) as described in Section 4.3. This is particularly true also for numerical studies involving violation of the assumptions in the FDR framework. In these cases, even the procedures with proven FDR control no longer hold because of the violation of assumptions. Furthermore, it is important to note here that Storey (2002), more or less, already recognized the link between his direct estimation approach to the sequential approach early on. (See also Storey *et al.* (2004).) However, their numerical comparisons aims to demonstrate the improvement in power within the assumptions of the FDR framework. As described in Section 4 the statistical assumptions do not hold under the minimal conditions of real microarray studies, including inhomogenous mean and variance, small sample sizes, dependence in gene expression, and so on. Thus, it is important to examine the performance of both FDR approaches under these more realistic conditions and also gauge their performance relative the the (optimal) least conservative FDR procedure. The unified computation framework proposed here, base on estimation of π_0 and approximating the least conservative FDR procedure, is one way to make such comparisons.

Finally, we note that the real data analyses use the optimal choice of λ for direct estimation, namely $\pi_0(\text{OPT})$. (See also the appendix section.) However, to reduce the computational cost in the large simulation we used the approximation $\hat{\pi}_0(\lambda = 1/2)$. A prelimary simulation suggests that the power of the direct estimation method using $\pi_0(\text{OPT})$ has slightly higher power than the approximation $\hat{\pi}_0(\lambda = 1/2)$, as expected. Therefore, when comparing the power of direct estimation to sequential FDR methods, the results reported here remain valid.

Acknowledgements

We are grateful to a referee and the Editor for helpful comments. The research reported here was partially supported by an American Cancer Society IRG program grant, through the UC Davis Cancer Center.

Appendix: Optimal Selection of λ to Minimize Mean Squared Error

Storey and Tibshirani (2003) proposed the following automatic algorithm for choosing the optimal λ to minimize the mean squared error of $\hat{\pi}_0(\lambda)$.

1. For each $\lambda_k \in \mathcal{R} = \{0, 0.01, 0.02, \ldots, 0.95\}$ compute $\hat{\pi}_0(\lambda_k)$.
2. Fit a natural cubic spline with 3 degrees of freedom, $\hat{f}$, through the data points $\{\lambda_k, \hat{\pi}_0(\lambda_k)\}_{k=1}^{96}$. The data points are weighted by $1 - \lambda_k$.
3. Estimate π_0 by $\hat{\pi}_0(\mathrm{OPT}) = \hat{f}(1)$.

References

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57**, 289-300.

Benjamini, Y., and Hochberg, Y. (2000). On the adaptive control of false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Statist.* **25**, 60-83.

Benjamini, Y., Krieger, A., and Yekutieli, D. (2001). Two staged linear step up FDR controlling procedure. Technical report, Department of Statistics and Operation Research, Tel-Aviv University, and Department of Statistics, Wharton School, University of Pennsylvania.

Chuaqui, R. F., Bonner, R. F., Best, C. J,, Gillespie, J. W., Flaig, M. J., Hewitt, S. M., Phillips, J. L., Krizman, D. B., Tangrea, M. A., Ahram, M., Linehan, W. M., Knezevic, V., Emmert-Buck, M. R. (2002). Post-analysis follow-up and validation of microarray experiments. *Nature Genetics* **32** Supplement, 509-514.

Davidson, L. A., Nguyen, D. V., Hokanson, R. M., Callaway, E. S., Isett, R. B., Turner, N. D., Dougherty, E. R., Lupton, J. R., Carroll, R. J., Chapkin, R. S. (2004) Chemopreventive n-3 polyunsaturated fatty acids reprogram genetic signatures during colon cancer initiation and progression in the rat. *Cancer Research*, to appear.

Finner, H., and Rotter, M. (2000). On the false discovery rate and expected type I errors. *Biometrical Journal* **43**, 985-1005.

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* **15**, 870-878.

Genovese, C. R. and Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statististical Society Series B*, **64** 499-518.

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Raffeld, M., Yakhini, Z., Ben-Dor, A., Dougherty, E. R., Kononen, J., Bubendorf, L., Fehrle, W., Pittaluga, S., Gruvberger, S., Loman, N., Johannsson, O., Olsson, H., Wilfond, B., Sauter, G., Kallioniemi, O.-P., Borg, A., and Trent J. (2001). Gene-Expression Profiles in Hereditary Breast Cancer. *The New England Journal of Medicine* **344**, 539-548.

Nguyen, D. V. (2004). On estimating the proportion of true null hypotheses for false discovery rate controlling procedures in exploratory DNA microarray studies. *Computational Statistics and Data Analysis* **46**, 407-425.

Nguyen, D. V., Arpat, A. B., Wang, N., and Carroll, R. J. (2002). DNA Microarray experiments: biological and technological aspects. *Biometrics* **58**, 701-717.

Ramakrishnan, R., Dorris, D., Lublinsky, A., Nguyen, A., Domanus, M., Prokhorova, A., Gieser, L., Touma, E., Lockner, R., Tata, M., Zhu, X., Patterson, M., Shippy, R., Sendera, T. J., and Mazumder, A. (2002) An assessment of Motorola CodeLink™ microarray performance for gene expression profiling applications. *Nucleic Acids Research* **30**, e30.

Schweder, T., and Spjøvtoll, E. (1982). Plot of p-values to evaluate many tests simultaneously. *Biometrika* **69**, 493-502.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B* **64**, 479-498.

Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics* **31**, 2013-2031.

Storey J. D., Taylor, J. E., and Siegmund, D. (2004). Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society Series B* **66**, 187-205.

Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences USA* **100**, 9440-9445.

Zien, A., Fluck, J., Zimmer, R., Lengauer, T. (2003) Microarrays: how many do you need? *Journal of Computational Biology* **10**, 653-667.

Received June 29, 2004; accepted September 27, 2004.

Danh V. Nguyen
Division of Biostatistics
Department of Public Health Sciences
School of Medicine
University of California
Davis, CA, 95616 USA
Email: ucdnguyen@ucdavis.edu