A comparative review of estimates of FDR in small microarray experiments

Meaza Demissie
Department of Statistics, University of Örebro, Sweden

Abstract

The focus of this paper is to illustrate and compare three recently suggested methods for the identification of differentially expressed genes and estimation of the false discovery rate (FDR) and to inversigate whether these FDR estimation methods are biased when we have a small sample size of 3 subjects per group. The methods are estimation of FDR based on averaging of fdr1d (FDR.avg), FDR estimation based on averaging of fdr2d (FDR.avg.fdr2d) and direct estimation of FDR using p-values (FDR.p). The properties are studied in a simulation experiment and a real data example. The comparison results shows that FDR.avg.fdr2d performs better than FDR.p and FDR.avg. The results further show the limitation of performance of the methods in very small sample size such as three subjects per group indicating the importance of developing a new FDR estimation procedure which works well under a situation of small sample size and unequal group variances.

Keywords: Differentially expressed genes, False discovery rate, local false discovery rate, multidimentional false discovery rate.

1 Introduction

In recent years, large scale simultaneous hypothesis testing problem has become an important issue in many scientific, medical and social study settings, where there is a need to consider hundreds or thousands of variables simultaneously with a limited number of observations. Among these areas is microarray technology, whose advent has made it possible to study the variation of expression of several thousands of genes simultaneously via multiple hypotheses testing (Pawitan *et al.*, 2005a).

Microarrays have been useful in addressing a wide range of problems in applied medical and biological research. However, the main focus of this work is with the most commonly posed question of identifying genes that are differentially expressed across two or more kinds of tissue samples (e.g. tissues like normal and cancerous cervical) or samples obtained under different experimental conditions (Dudoit *et al.*, 2003). Testing statistical significance of differential expression in microarray data analysis requires testing of each gene or inferring on many null hypothesis simultaneously. When many hypotheses are tested, the probability

that a type I error is committed increases sharply with the number of hypotheses and this leads to the multiple testing problem.

Some authors have proposed different techniques to address the problem of multiple testing. The most common method has been to control the family wise error rate(FWER), the probability of committing at least one type I error (Westfall and Young, 1993). Dudiot *et al.* (2002) has later applied this method to microarray data to systematically assess and measure the significance of any observed changes in gene expression levels. However, most of the traditional approaches are conservative and uses the standard p-value for every experiment which leads to the risk of accumulating a large rate of false positive results (Benjamini and Yekutieli, 2001; Sinha *et al.*, 2004; Pawitan *et al.*, 2005a). However, the false discovery rate (FDR), which has been proposed by Benjamini and Hochberg (1995) is a better and widely used method as an error measure for multiple testing problems.

These authors propose procedures to control the FDR, which is the proportion of false positives among all of the null hypotheses rejected (Benjamini and Hockberg, 1995, 2000; Keselman *et al*, 2002; Pawitan *et al*, 2005a). The false discovery rate, has further been developed for several purposes by some authors (e.g. Storey 2002, Storey and Tibishirani, 2003; Pawitan *et al.*, 2005b; Ploner *et al.*, 2006)

In addition to multiple testing problems, the relatively high cost of research in microarray experiments forces researchers in the area to use very small sample size (e.g. Carolina *et al.*, 2001 ; Gadbury *et al.*, 2003; Kayo *et al.*, 2001; Zhao and Pan, 2003). In cases when the sample size is small, the robustness of parametric tests such as the t-test is questionable due to instability in the estimation of gene specific variances (Tusher *et al.*, 2001; Smith *et al.*, 2003; Yang and Churchill, 2006). The small sample problem together with the common feature of many genes with small effects in microarray experiments intensifies the problem of estimation of FDR (Pawitan *et al.*, 2005b). Three recently proposed methods for estimation of FDR are the direct estimate of FDR based on p-values(FDR.p) as described in Pawitan *et al.* (2005), estimation of FDR based on averaging of local false discovery rate(FDR.avg) and estimation of FDR based on averaging of multidimensional false discovery rate (FDR.avg.fdr2d). The authors who propose the methods have studied the properties of the FDR estimates using different sample sizes. Nevertheless, none of the methods introduced are checked for validity for very small sample sizes such as 3 samples per group. Hence, the purposes of this work are (1) to illustrate and compare the performances of the three FDR estimation methods ; (2) to investigate whether these FDR estimation methods are relevant (or are not biased) when three samples per group are used. Both simulated and real data examples are used to investigate and illustrate the methods using different sample sizes and different sets of proportion of none differentially expressed genes(non-DE).

2 Materials and Methods

In testing many hypotheses, the probability of making a type I error (false positive) increases rapidly with the number of hypotheses. The common approach of multiple testing is to construct a procedure that controls the FWER. Control of the FWER leads to conservative

procedures which may be useful for situations where the aim is to identify a small number of hypotheses that are truly rejected such as in a study of clinical trials to approve a drug. However, in the case of data analyses with several thousands of variables as in microarray data, approaches of controlling the FWER are too conservative. In such cases another measure for the erroneous rejection of a number of true null hypotheses, the FDR which is introduced by Benjamini and Hochberg (1995) is prefered. The false discovery rate (FDR) is the expected proportion of wrongly rejected null hypotheses among the rejected ones.

2.1 Test statistics

A test statistic, Z has been used to test whether a gene is differentially expressed (DE) or equally expressed (EE) between two groups. Following Efron *et al.* (2001), Pawitan *et al.* (2005b) and Ploner *et al.* (2006), the distribution F of the observed statistics has the form of the mixture model

$$F(z) = \pi_0 F_0(z) + (1 - \pi_0) F_1(z) \tag{1}$$

where π_0 is the proportion of truly non-DE genes, F_0 is the distribution of the statistics of non-DE genes and F_1 is the distribution of the statistics of DE genes.

If the null hypothesis, H_0 of non-DE genes is rejected for $|Z| > c$ for a given critical value $c > 0$ then

$$FDR(c) = \frac{\pi_0 Pr_0(|Z| > c|H_0)}{Pr(|Z| > c)} \tag{2}$$

where Pr refers to probability.

2.2 Direct FDR estimation using p-values

For a fixed cut-off value c of a given t-statistics Z with rejection region $|Z| > c$ and m number of genes, the following empirical FDR can be obtained as

$$\widehat{FDR}(c) = \frac{\hat{\pi_0}\{1 + \hat{F}_0(-c) - \hat{F}_0(c)\}}{1 + \hat{F}(-c) - \hat{F}(c)}$$

where $\hat{\pi_0}$ is the estimator of π_0, $\hat{F}_0$ is the estimator of F_0 and $\hat{F}$ is the estimator of F. As described in Pawitan *et al.*, 2005b, $\hat{F}$ is obtained from the empirical distribution of the observed statistics, $\hat{F}_0$ is obtained directly from F_0 which is generated by permutation argument and $\hat{\pi_0}$ is estimated as (Efron et al., 2001)

$$\pi_0 \leq \hat{\pi_0} = min_c\{\widehat{f(c)}/f_0(c)\} \tag{3}$$

In terms of the distribution of p-values equation (2) can be expressed as (Pawitan *et al.*, 2005b)

$$\widehat{FDR}(\alpha) = \frac{\hat{\pi_0}\alpha}{\hat{D}(\alpha)}$$

where $\hat{D}(\alpha)$ is the proportion of p-values less than or equal to α. When α is the i^{th} order of p values ($p_{(i)}$) then $\hat{D}(\alpha) = i/m$. Equation (3) can then be written as

$$\widehat{FDR}(p_{(i)}) = \frac{\hat{\pi}_0 m p_{(i)}}{i} \leq \frac{m p_{(i)}}{i} \tag{4}$$

the upper bound is the unconstrained version of the Benjamini and Hochberg (1995) adjusted p-values. Monotonocity is secured by taking a cumulative minimum over larger p-values (Pawitan *et al.*, 2005b).

2.3 FDR estimation using average of local false discovery rate

For a fixed value z of the statistics Z the mixture framework of equation (1) could similarly be expressed as

$$f(z) = \pi_0 f_0(z) + (1 - \pi_0) f_1(z) \tag{5}$$

where π_0 is proportion of truly non-DE genes, f_0 is the density of the statistics of non-DE genes and f_1 is the density of the statistics of DE genes.

The local false discovery rate (fdr) in this setting is defined as the posterior probability of a gene being non-DE, given that $Z{=}z$ (Efron and Tibishirani, 2002; Ploner *et al.*, 2006). Using Bayes rule it can be expressed as

$$fdr(z) = \frac{\pi_0 f_0(z)}{f(z)} \tag{6}$$

The empirical fdr is obtained from the estimates of π_0, $f_0(z)$ and $f(z)$. The estimate π_0 is obtained as in equation 3, the estimate of $f(z)$ is obtained as described in Ploner *et al.*, 2006, using non-parametric density estimation of the observed Z and the estimate of $f_0(z)$is obtained using similar smoothing technique as for $f(z)$ but this time using the statistics obtained by permutation of the null distribution.

The global FDR can be obtained by taking the average of the local fdr expressed in equation (6). Global false discovery rate is expressed as the posterior probability of a case being non-DE given that its test statistics value (Z) is less than some value (z) as

$$FDR(z) = prob\{non - DE | Z \leq z\} = \frac{\pi_0 F_0(z)}{F(z)} \tag{7}$$

where $F_0(z)$ and $F(z)$ are the cdfs of $f_0(z)$ and $f(z)$respectively. Equation (7) can then be expressed as

$$FDR(z) = \frac{\pi_0 \int_{-\infty}^{z} f_0(Z)(y)dy}{\int_{-\infty}^{z} f(Z)(y)dy} \tag{8}$$

substituting $f_0(Z)$ in (8) by the expression $\frac{fdr(Z)f(Z)}{\pi_0}$ we get

$$FDR(z) = E\{fdr(Z) \mid Z \leq z\}$$

Similarly the empirical FDR is obtained as

$$\widehat{FDR(z)} = E\{\widehat{fdr(Z)} \mid Z \leq z\}$$

where $\widehat{fdr(Z)}$ is the estimator of fdr.

2.4 FDR estimation using average of multidimensional false discovery rate

Ploner *et al.* (2006) define the two-dimensional false discovery rate (fdr2d) as follows :

$$fdr2d(c_1, c_2) = \frac{\pi_0 f_0(c_1, c_2)}{f(c_1, c_2)}$$

where c_1 is some fixed value of the t-statistics C1 and c_2 is some fixed value of the standard error of the statistic C2.

The empirical fdr2d is obtained as (Ploner et al 2006)

$$\widehat{fdr2d(c)} = \hat{\pi_0} \frac{r(c)}{p\{1 - r(c)\}}$$

where r(c) is the proportion of success (all statistics generated under permutation) as a function of c and is given by $\frac{pf_0(c)}{f(c)+pf_0(c)}$. A non-parametric density estimation is performed based on success-failure probability as a function of c where the probability of failure is assigned to the observed statistics.

The global FDR can also be found by taking average of the ordered fdr2d values as follows:

$$FDR(S) = E\{fdr2d(C_1, C_2) \mid S\}$$

where S represents the rejection region of the two dimensional statistics $Z = (C_1, C_2)$ such that $Z \in S$ are called DE .

2.5 Simulation set up

Three different simulation designs have been used to demonstrate the performance of the different FDR estimation methods in the comparison of two independent groups using t-statistics.

In the first simulation design(set-up 1), m=10000 genes of two-sample microarray data has been drawn from a normal distribution with constant standard deviation (sigma=1) across genes. In the simulation a fixed effect size of D=1 for differentially expressed genes, in units of gene-specific standard deviation has been used. Different sample sizes ranging from 5 to 20 per group and the proportions of truly non-DE genes $\pi_0 = 0.95$, 0.9 and 0.8 have been used.

In the second simulation design (set-up 2), a DNA microarray experiment with gene expression data of biopsy samples of diffuse large B-cell lymphoma of 240 patients as described

in Rosenwald *et al.* (2002) has been used to simulate a new data set with sample size n=3 subjects per group. The samples in each group have been simulated using sampling with replacement. The mixture estimate used has been $\hat{\pi_0}$=0.59 (Pawitan *et al.*, 2005). Among the 240 patients or cases 102 of them were found to survive after average clinical follow up of 4.4 years and so the two groups compared were 102 survivors and 138 non-survivors. Following Ploner *et al.* (2006), the censoring information for this data set has been ignored.

In the third simulation design (set-up 3) a biologically simulated data (Spike-in), which is a subset of the data used in Tan *et al.*, (2006) containing 3 versus 3 samples from control and spike-in groups with proportion of non-DE genes π_0=0.9 has been used. The total number of genes used has been m=14010.

3 Results

3.1 Simulation study

Figure 1 compares the performance of the FDR estimation methods based on direct computation of p-values (FDR.p) and averaging of fdr1d (FDR.avg) with the theoretical FDR (FDR.true) as described in the simulation set-up 1. Figure 1a. shows the curve of FDR as a function of t-statistics for n=10 and π_0=0.9, Figure 1b shows same curves for π_0=0.95, Figure 1c shows the curves for π_0=0.8 and Figure 1d shows the curves for n=5 and π_0=0.9. Visual inspection of these four plots indicate the accuracy of the FDR estimation methods for reasonably large n such as n=10 per group up to small n such as n=5 when using π_0=0.9. However, as shown in Figure 2, as sample size decreases down to n=3 the FDR estimation methods are no more accurate. This will also be shown later. When the percentage of DE genes is decreased down to 0.05 the two methods under-estimate the true FDR (see Figure 1b). Likewise when percentage of DE genes is increased up to 0.2 both methods over-estimate the true FDR (see Figure 1c). This result indicates how a change of small π_0 influences FDR estimation.

For n=10 samples per group, the 5 percent level two sided critical value for the t-statistics and associated FDR when π_0=0.9 is approximately 2.06 and 0.46 respectively. Whereas, for n=5 samples per group as reported by Pawitan et al. (2005), the 5 percent level two sided critical value for the t-statistic is 2.31 and the associated FDR at this point is >0.6. At same value of π_0 when sample size is reduced down to 3 samples per group the 5 percent two-sided critical value for the t-statistic is approximately 2.73 and one expects to get FDR of 74 percent. This result idicates that FDR increases sharply and unacceptably as sample size decreases. This is true even if critical values are set large.

For an overall comparison, the global FDRs associated with the first 10 percent top ranking genes have been computed and the resulting FDR curves as a function of the proportion of genes declared DE have been drawn in the consecutive figures. Figure 3 shows the FDR plots for simulation set up 1 described in section 2.5 for sample sizes n=20 to n=5 and π_0=0.9 for comparison of FDR.p, FDR.avg and FDR.avg.fdr2d. False discovery rates have been computed by ranking genes according to their true FDR values which are computed

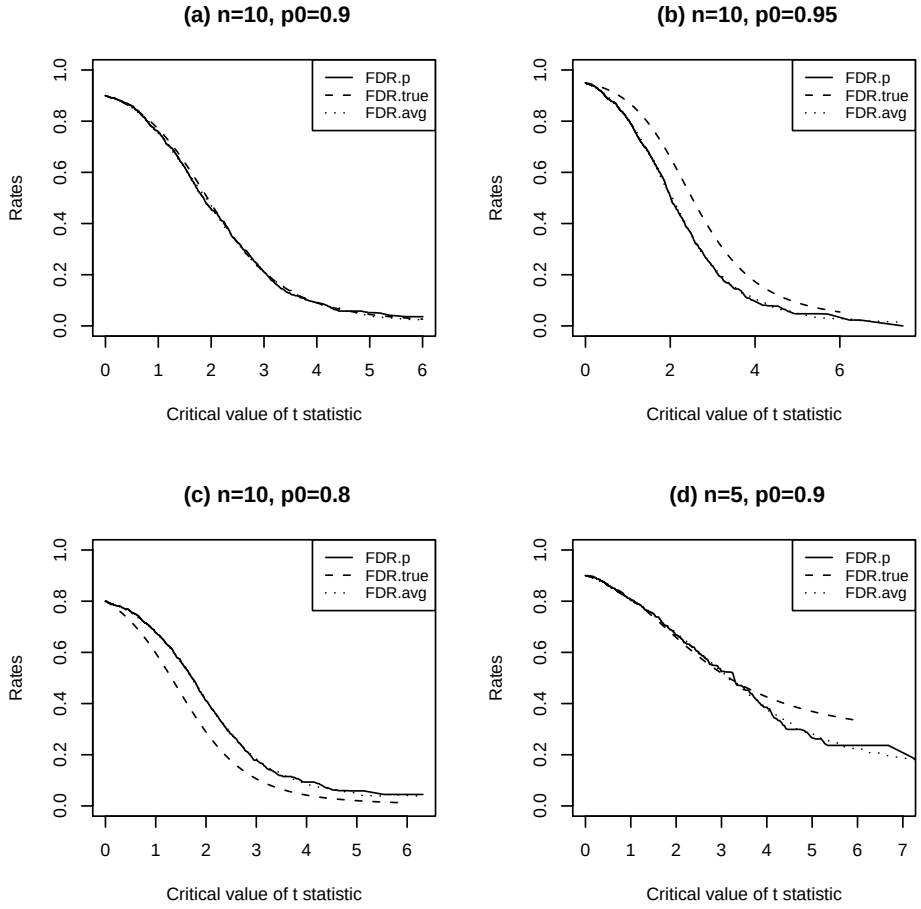

Figure 1: *FDR curves as a function of critical value of t statistics. (a) Comparison of FDR p value based and FDR based on averaging of local false discovery rate with the theoretical FDR under simulation set up 1 for the sample size n=10 subjects per group and* π_0*=0.9. (b) Same comparison as in (a) with* π_0*=0.95. (c) Same comparison as in (a) and (b) using* π_0*=0.8. (d) Same comparison as (a) using sample size n=5 subjects per group*

from the known DE status of each simulated gene. One can see from Figure 3 that FDR.p is less smooth than FDR.avg. This indicates the departure of the distribution of the empirical p-values from the theoretical continuous uniform (0,1) distribution. The two main reasons for the departure are:

1. The p-values which are obtained from random permutation of the permutation test under the null hypothesis are correlated.

2. The problem of identifying the real DE genes because of the common feature of many genes with small effects which is common in Microarray data in a situation when one has only a small number of samples (Pawitan *et al.*, 2005b).

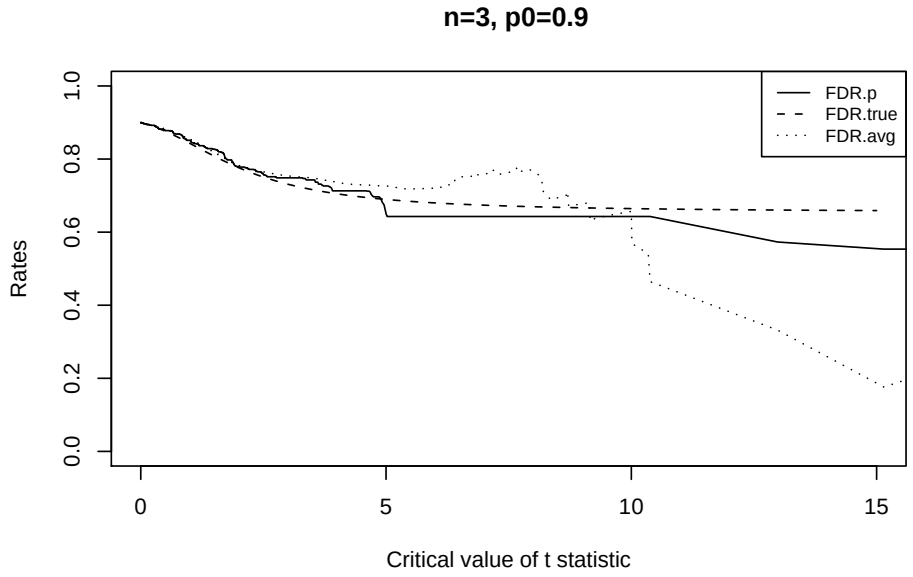

Figure 2: *FDR curves as a function of critical value of t statistics for n=3 and* π_0*=0.9.*

However, the FDR estimate which has been obtained by averaging of fdr is smooth as the computation is based on smoothing operation to estimate the densities in equation (6).

Figure 3a-Figure 3d show the performance of FDR.avg.fdr2d over the other two estimates and FDR.avg as well as FDR.p for large sample size and for sample size as small as 5 subjects per group. Same performance of FDR.avg.fdr2d over FDR.p and FDR.avg has been observed when increasing and decreasing proportion of DE genes to 0.2 and 0.05 respectively. The increase of FDR can also be seen from the plots as sample size decreases.

The analysis of Lymphoma data which has been described in setup 2 of section 2.5 and shown in Figure 4, indicates that when the sample size in the simulation is reduced down to 3 samples per group, one can only get 20 possible permutations. Therefore, FDR.p is severely affected by the very small number of possible test statistics and hence of p-values. The comparison of FDR.p and FDR.avg.fdr2d to the 3 by 3 sample data sub sampled from the full data has been made in Figure 4a and Figure 4b respectively. Both plots show the very large difference in the performance of whole sample estimate versus the small sample (3 versus 3) estimate. According to results when the sample size is decreased down to 3 subjects per group the tests loose sensitivity, giving larger FDR compared to the whole sample estimates. This is true for both FDR.p and FDR.avg.fdr2d as shown in Figure 4a and Figure 4b respectively.

Figure 5 presents the analysis of the Spike-in data described under simulation set-up 3 of section 2.5. It compares FDR.p and FDR.avg.fdr2d to the true false discovery proportion (FDP) for the sample size 3 subjects per group. It shows that neither FDR.p nor FDR.avg.fdr2d gives close FDR estimate to the true FDP. This result indicates the problem of FDR estimation in very small samples for the two methods. It is basically expected that

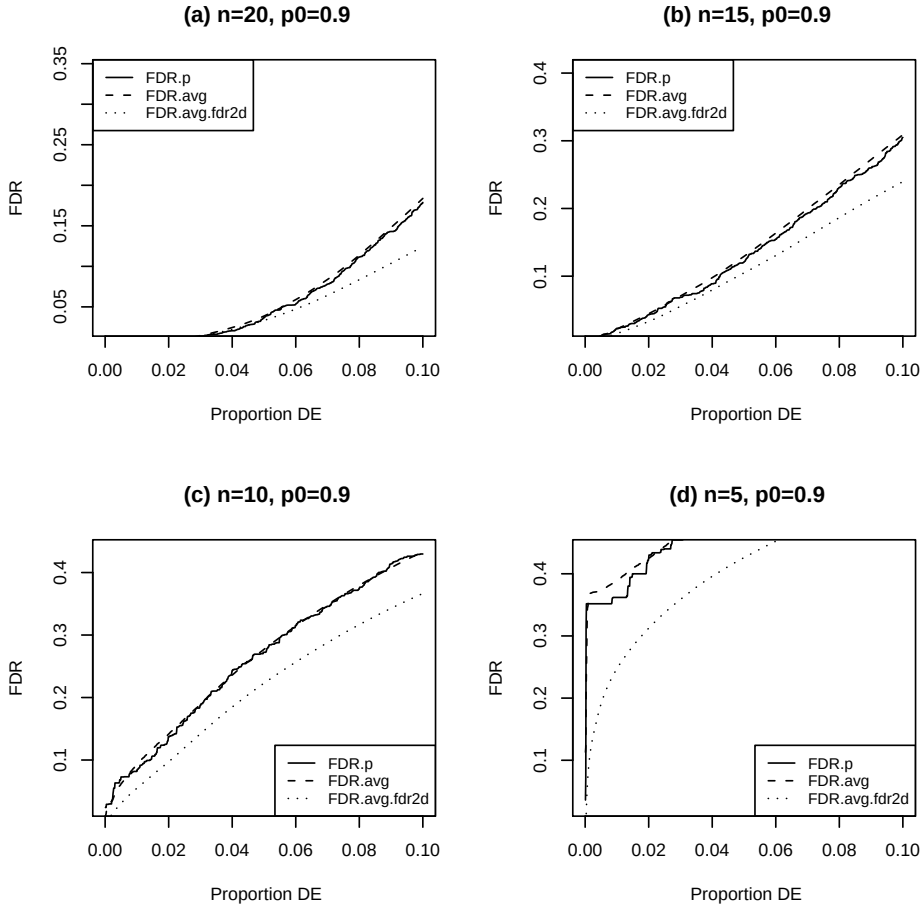

Figure 3: *Plot showing FDR of the first 10 percent top regulated genes (a) Comparison of FDR p-value based, FDR based on averaging of fdr1d and FDR based on averaging of fdr2d under simulation set up 1 with* π_0*for the sample size n=20 subjects per group (b)Same comparison as in (a) using the sample size n=15 subjects per group (c)Same comparison as in (a) and (b) using the sample size n=10 subjects per group. (d) Same comparison as the three plots using the sample size n=5 subjects per group.*

the two FDR estimation procedures would not work well in a situation where one has a very small sample size. This is due to the fact that the null distribution of both procedures is unknown and there is a requirement of large enough samples to use permutation tests.

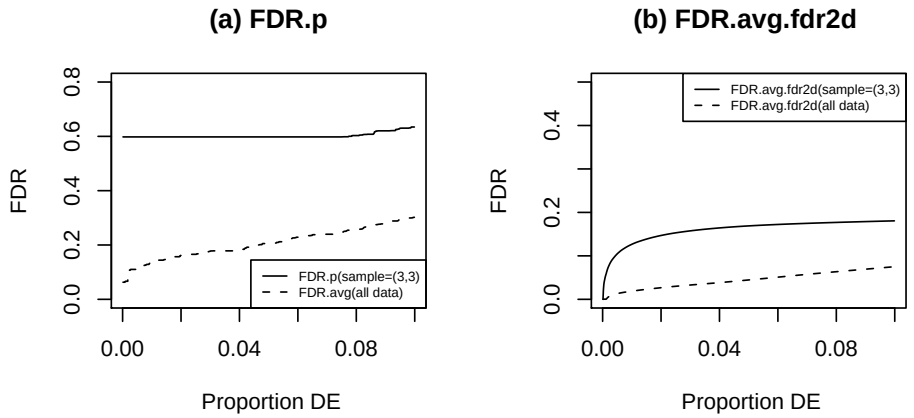

Figure 4: *Plot showing FDR of the first 10 percent top regulated genes (a) Comparison of direct FDR estimate using p-values for the 3 versus 3 sample simulated from Lymphoma data described in section 2.5 to whole sample based estimate. (b) Same comparison as in (a) but this time FDR based on averaging of fdr2d.*

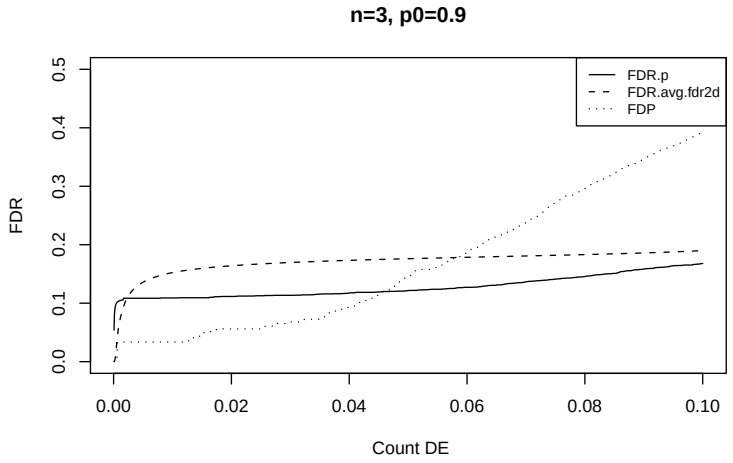

Figure 5: *Plot showing FDR of the first 10 percent top regulated genes for Spike in data as described in section 2.5 to compare FDR p-value based and based on averaging of fdr2d to the true false discovery proportion.*

3.2 Application to real data

The breast cancer data, which has been described in Hedenfalk *et al.* (2000) to investigate how tumours resulting from two different mutations (BRCA1 and BRCA2) differ in their genetic activity has been used. The sample sizes of the first and second mutation groups; BRCA1 and BRCA2 are 7 and 8 respectively; more details can be found in Hedenfalk *et al.* (2001). Following Ploner *et al.*, (2006), only 370 genes have been analysed using the mixture

estimate $\hat{\pi_0}$=0.7.

The resulting analysis is shown in Figure 6. It shows the performance of FDR based on averaging of fdr2d over FDR.p and FDR.avg. FDR based on averaging of fdr1d performs similar to FDR.p. This result agrees with the results obtained from simulation study.

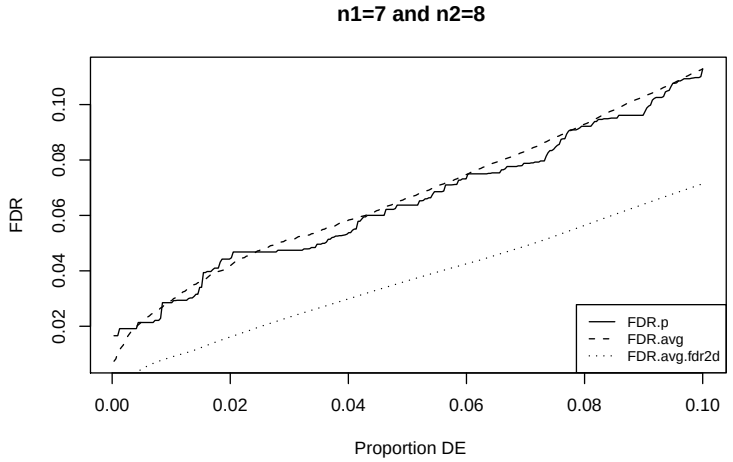

Figure 6: *Plot showing FDR of the first 10 percent top regulated genes for Hedenfalk data as described in section 3.2 to compare FDR p-value based, FDR based on averaging of fdr1d and FDR based on averaging of fdr2d.*

Table 1 displays the numbers of selected differentially expressed genes at FDR cut-offs(values) 0.0001, 0.001, 0.01, or 0.05. These results again show the performance of FDR.avg.fdr2d over FDR.avg and FDR.p. Notice that the estimated FDR of FDR.avg.fdr2d shows a sharp increase in number of significant genes over the range of FDR rates. This table shows that threshholding genes with a cutoff value of 0.0001, 0.001 and 0.01 for FDR.avg.fd2d yields 4 , 7 and 37 DE genes respectively. This implies that no genes are expected to be false positives at these threshholds. Whereas thresholding FDR.avg.fdr2d at 0.05 yields 226 DE which is very large number of genes as compared to the total. It is expected that 11 genes out of the 226 are false positives. On the other hand, thresholding FDR.p at 0.05 yields 131 DE and around 7 genes are expected to be false positives. In general, while controlling about the same FDR, FDR.avg.fdr2d identifies more differentially expressed genes than FDR.p and FDR.avg. Hence the results of the analysis of this data indicates that the method based on FDR.avg.fdr2d is more powerful than FDR.p and FDR.avg. In light of these results, it is recommended that the method based on FDR.avg.fdr2d towards finding differentially expressed genes in microarray data can provide useful information beyond that of other methods. Pound (2006) also recommends choosing an appropriate FDR method for specific applications. It is however recommended that the cutoff values be used as exploratory guide for the researcher to investigate and further know which features are false positives before making any decisions. It is also good to make an association study of misclassification with false discovery rate as there might be genes which are most differentially expressed but which

may not necessarily be of most biological relevance.

Table 1. Count of differentially expressed genes at different FDR thresholds of the three methods

FDR cutoff	FDR.p	FDR.avg	FDR.avg.fdr2d	total number of genes
0.0001	1		4	370
0.001	1		7	370
0.01	5	5	37	370
0.05	131	115	226	370

4 Discussion

In this work, the comparative study of performances of three FDR estimation procedures for identification of differentially expressed genes and investigation of the methods for their validity when using very small sample size; three subjects per group have been presented. These are FDR p-value based (FDR.p), FDR based on averaging of fdr1d (FDR.avg) and FDR based on averaging of fdr2d (FDR.avg.fdr2d). Both simulation and real data examples show that FDR.avg.fdr2d performs better than FDR.p and FDR.avg for large sample sizes and for sample sizes as small as 5 subjects per group. FDR based on averaging of fdr1d is observed to perform as well as FDR.p. However, as predicted by the theoretical analysis where the three procedures require large enough samples to run permutation tests (Pawitan *et al.*, 2005; Ploner *et al.*, 2006), the simulation study has confirmed that the problem of FDR estimation still persists in a situation when very small sample sizes such as 3 subjects per group is used. Smyth (2004) has recently developed a procedure which works well in very small samples but not in a case when the two groups under comparison do not have equal variance. In conclusion, this work indicates that there is a need to develop a new FDR estimation procedure which is suited for very small sample sizes such as three observations per group when the group variances are unequal.

References

[1] Benjamini,Y. and Yekutieli,D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29 (4): 1165-1188.

[2] Benjamini,Y. and Hochberg,Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser.* B, 57, 289-300.

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

[4] Corominola,H. *et al.* (2001). Identification of novel genes differentially expressed in omental fat of obese subjects and obese Type 2 diabetic Patients. *Diabets*, 50, 2822-2830.

[5] Dudoit,S. *et al.* (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Stat. Sin.*, 12, 111-139.

[6] Dudoit,S. *et al.* (2003). Multiple hypothesis testing in microarray experiments. *Stat. Sin.*, 18, 71-103.

[7] Efron,B. and Tibshirani,R. (2002) Empirical Bayes methods and false discovery rates for microarrays. *Genet. Epidemiol.*, 23, 70-86.

[8] Efron,B. *et al.* (2001). Empirical Bayes analysis of a microarray experiment. *J. Am. Stat. Assoc.*, 96, 1151-1160.

[9] Gadbury,G.L. *et al.* (2003). Randomization tests for small samples: an application for genetic expression data. *Applied Statistics* 52: 365-376

[10] Hedenfalk,D. *et al.* (2001) Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.*, 344, 539548.

[11] Keselman,H.J. *et al.* (2002). Controlling the rate of type I error over a large set of statistical tests. *Br. J. Math. Stat. Psychol.*, 55, 27-39.

[12] Kayo, T., *et al.* (2001). Influences of aging and caloric restriction on the transcriptional profile of skeletal muscle from rhesus monkeys. *Proc. Natl Acad. Sci. USA* 98, 5093-5098.

[13] Pawitan,Y. *et al.* (2005a). False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics*, 21, 3017-3024.

[14] Pawitan,Y. *et al.* (2005b). Bias in the estimation of false discovery rate in microarray studies. *Bioinformatics*, 21, 3865-3872.

[15] Ploner, A. *et al.* (2006) Multidimensional local false discovery rate for micorarray studies. *Bioinformatics*, 22: 556-65.

[16] Rosenwald,A. *et al.* (2002) The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.*, 346, 19371947.

[17] Sinha, R. *et al.* (2005). Local false discovery rate and minimum total error rate approaches to identifying interesting chromosomal regions. *BMC Genetics* 6 S23

[18] Smyth, G.K. (2004). Smyth,G.K. (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol, 3(1):Article 3

[19] Storey,J.D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B*, 64, 479-498.

[20] Storey,J.D. and Tibshirani,R. (2003). Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA*, 100, 9440-9445.

[21] Tan *et al.* (2006). Finding regions of significance in SELDI measurements for identifying protein biomarkers. *Bioinformatics* 22(12):1515-1523

[22] Tusher, V.G. *et al.* (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA*, 98, 5116- 5121.

[23] Yang, H. and Churchill, G. (2006). Estimating p-values in small microarray experiments. *Bioinformatics*, 23, 38-43.

[24] Westfall,P.H. and Young,S.S. (1993). Resampling Based Multiple Testing. *Wiley, NY.*

[25] Zhao, Y. and Pan W. (2003) Modified nonparametric approaches to detecting differentially expressed genes in replicated microarray experiments. *Bioinformatics* 19 (9): 1046-1054 .