Floating Prioritized Subset Analysis: a Powerful Method to Detect Differentially Expressed Genes

Wan-Yu Lin [1], Wen-Chung Lee [1,2 §]

[1] Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan.

[2] Research Center for Genes, Environment and Human Health, National Taiwan University, Taipei, Taiwan.

[§]Corresponding author

Correspondence & reprint requests: Prof. Wen-Chung Lee,

Rm. 536, No. 17, Xuzhou Rd., Taipei 100, Taiwan.

(FAX: 886-2-23511955; e-mail:wenchung@ntu.edu.tw)

Abstract

Controlling the false discovery rate (FDR) is a powerful approach to deal with a large number of hypothesis tests, such as in gene expression data analyses and genome-wide association studies. To further boost power, here we propose a floating prioritized subset analysis (floating PSA) that can more effectively use prior knowledge and detect more genes that are differentially expressed. Genes are first allocated into two subsets: a prioritized subset and a non-prioritized subset, according to investigators' prior biological knowledge. We allow the FDRs of the two subsets to vary freely (to float) but aim to control the overall FDR at a desired level. An algorithm for the floating PSA is developed to detect the largest number of true positives. Theoretical justifications of the algorithm are given, and computer simulation studies show that the method has good statistical properties. We apply this method to detect genes that are differentially expressed between acute lymphoblastic leukemia and acute myeloid leukemia patients. The result shows that our floating PSA identifies 32 more genes (permutation-based FDR = 0.0427) than the conventional (fixed) FDR control. Another example is a colon cancer study, and our floating PSA identifies 43 more genes (permutation-based FDR = 0.0502). The floating PSA method is to be recommended for the detection of differentially expressed genes, in light of its power, robustness, and ease of implementation.

Keywords: false discovery rate; gene expression; microarray; multiple comparisons; multiple hypothesis testing; simultaneous inference.

1. Introduction

Adjustment for multiple comparisons is a thorny issue in large-scale genomic studies, such as gene expression data analyses and genome-wide association studies. To deal with a large number of hypothesis tests, controlling the false discovery rate (FDR) provides a more practical balance between the numbers of true positives and false positives (Benjamini and Hochberg, 1995; Greenwood et al., 2007; Storey, 2002; Storey and Tibshirani, 2003). Compared with controlling the family-wise error rate, controlling the FDR is a more liberal and powerful approach.

A number of FDR related approaches allow incorporating prior knowledge to boost power, such as the weighted FDR control (Benjamini and Hochberg, 1997; Genovese et al., 2006; Roeder et al., 2006), the stratified FDR control (Sun et al., 2006), and the prioritized subset analysis (PSA) (Li et al., 2008). Among these methods, the PSA shows the most significant benefit gaining from using prior knowledge. To perform a PSA, a researcher, based on his/her prior biological knowledge, first picks out from among all the genes under study, a certain number of genes that he/she thinks are likely to be the true positives. He/she then places those selected genes in a 'prioritized subset' (henceforth referred to as the *P* subset) and those that are left behind in a 'non-prioritized' subset (the *N* subset). The FDR control is then applied to these two subsets separately, and the significant results are tallied and totaled. Li *et al.* (2008) showed that this simple prioritization scheme could detect more true positive genes, as compared to an 'aggregate analysis' (AA) where genes are not to be prioritized but pooled together for a collective FDR control.

Recently, Lin and Lee (2010) showed the prominent benefit of the PSA with a real genome-wide association study. However, for gene expression data, the above PSA that applies the same FDR criteria to the two subsets of the prioritized and the non-prioritized may not be an optimal approach. And worse, as will be demonstrated in this paper, the PSA can sometimes lead to fewer true positive genes being detected as compared to the AA approach. Such results shatter everyone's expectation since we all know that the PSA is equipped with extra prior knowledge (and therefore should perform better) while the AA isn't (and should perform less well). The cause for this occasional aberrant performance lies in PSA's fixing the FDR at the same value for the two subsets. Intuitively, the threshold for declaring a gene significant can be lower in the *P* subset, because genes in that subset may have a higher probability of being true positive. But genes in the *P* subset may also be more of those 'stronger' genes that have larger effect sizes. Should this be the case, the threshold for declaring a gene significant in that subset can be raised higher (instead of lower), because stronger genes can withstand more rigorous testing. However, we see that the PSA, with its fixed FDR control to the *P* and the *N* subsets, does not have the flexibility to regulate all these. It is then no wonder that it can sometimes ruin the precious biological knowledge that is put in and obtain a result even worse than when nothing is given.

To more effectively use the prior knowledge, here we propose a floating prioritized subset analysis (henceforth referred to as the 'floating PSA'). Instead of fixing the FDRs of the two subsets, the floating PSA allows them to vary freely but aim to control the overall FDR at a desired level. We develop an algorithm for the floating PSA that can detect the largest number of true positive genes. Theoretical justifications of the

algorithm are given and computer simulations are performed to study its performances. We also apply our method to an acute leukemia data set and a colon cancer data set.

2. Methods

2.1. The Algorithm for the Floating PSA

Suppose that a researcher has already allocated the study genes into a *P* subset (subscript *P*, a total of m_P genes) and an *N* subset (subscript *N*, a total of m_N genes) according to his/her prior knowledge. The desired level of the overall FDR is q^*. The following is our algorithm for the floating PSA.

1. Calculate the *p* values for the m_P genes in the *P* subset and the m_N genes in the *N* subset.

2. Apply the conventional FDR method (e.g., (Storey, 2002; Storey and Tibshirani, 2003)) to transform the *p* values to the *q* values, separately, for the m_P genes and for the m_N genes.

3. For all possible values of FDRs of the two subsets, q_P and q_N, both between 0 and 1, do,

3.1. In the *P* subset, count the number of tests with *q* value less than q_P. Denote this as S_P. When these S_P genes are declared significant, the expected number of false positives is $F_P = S_P q_P$, and the expected number of true positives, $T_P = S_P - F_P$.

3.2. Do the same to the *N* subset, but count the number of tests with *q* value less than q_N. Denote this as S_N. When these S_N genes are declared

significant, the expected numbers of false positives and true positives are $F_N = S_N q_N$ and $T_N = S_N - F_N$, respectively.

4. Among all pairs of q_P and q_N that satisfy $(F_P + F_N)/(S_P + S_N) \le q^*$, find the pair that maximizes the expected total number of true positives: $T_P + T_N$. Such q_P and q_N are the optimal values of our floating PSA for the *P* subset and the *N* subset, respectively.

An R code to perform the floating PSA algorithm is available upon request.

2.2. Theoretical Justifications

We can express the expected number of genes declared significant, S, the expected number of true positives, T, and the expected number of false positives, F ($S = T + F$), as functions of the level of significance, α. Under the same desired level of the overall FDR (q^*), the three methods of the floating PSA, the fixed PSA, and the AA amount to the following three different constraints:

$$\frac{F_P(\alpha_P) + F_N(\alpha_N)}{S_P(\alpha_P) + S_N(\alpha_N)} \le q^*, \text{ for floating PSA;} \quad (1)$$

$$\frac{F_P(\alpha_P)}{S_P(\alpha_P)} \le q^* \text{ and } \frac{F_N(\alpha_N)}{S_N(\alpha_N)} \le q^*, \text{ for fixed PSA;} \quad (2)$$

and

$$\frac{F(\alpha)}{S(\alpha)} \le q^*, \text{ for AA.} \quad (3)$$

The objective function of the floating PSA and the fixed PSA is to maximize $T(\alpha_P, \alpha_N) = T_P(\alpha_P) + T_N(\alpha_N) = \left[S_P(\alpha_P) + S_N(\alpha_N)\right] - \left[F_P(\alpha_P) + F_N(\alpha_N)\right]$, and that of the AA, to maximize $T(\alpha) = S(\alpha) - F(\alpha)$.

It is straightforward to see that $F_P(\alpha_P)/S_P(\alpha_P) \le q^*$ and $F_N(\alpha_N)/S_N(\alpha_N) \le q^*$ implies $\left[F_P(\alpha_P) + F_N(\alpha_N)\right]/\left[S_P(\alpha_P) + S_N(\alpha_N)\right] \le q^*$. Thus, any pair of (α_P , α_N) that satisfies (2) also satisfies (1). Next, we see that (3) can be expressed as $\left[F_P(\alpha) + F_N(\alpha)\right]/\left[S_P(\alpha) + S_N(\alpha)\right] \le q^*$. Thus, any value of α that satisfies (3) also satisfies (1). Let the α's that satisfy (1), (2), and (3) form the sets of $\Omega_{floating}$, Ω_{fixed}, and Ω_{AA}, respectively. We then have $\Omega_{fixed} \subseteq \Omega_{floating}$ and $\Omega_{AA} \subseteq \Omega_{floating}$. Therefore, $\max_{\Omega_{floating}}(T) \ge \max_{\Omega_{fixed}}(T)$, and $\max_{\Omega_{floating}}(T) \ge \max_{\Omega_{AA}}(T)$. And we see that the floating PSA will detect no fewer true positive genes than the fixed PSA and the AA.

Note that the subsets *P* and *N* play absolute symmetric role here. If the researcher's prior information is totally jumbled up, such that those genes likely to be true positives are placed (wrongly) in the *N* subset, and those unlikely to be true positives, in the *P* subset, we are still guaranteed to have $\max_{\Omega_{floating}}(T) \ge \max_{\Omega_{fixed}}(T)$ and $\max_{\Omega_{floating}}(T) \ge \max_{\Omega_{AA}}(T)$.

On the other hand, if the researcher prioritizes genes in a purely random manner (the prioritization probability is the same for each and every gene), we will have $\max_{\Omega_{floating}}(T) = \max_{\Omega_{fixed}}(T) = \max_{\Omega_{AA}}(T)$. To show this, first we note that under such random prioritization, $S_P(\alpha) = m_P s(\alpha)$ and $S_N(\alpha) = m_N s(\alpha)$, where $s(\alpha)$ is the probability that a randomly chosen gene will be declared significant under α. And $F_P(\alpha) = m_P f(\alpha)$ and $F_N(\alpha) = m_N f(\alpha)$, where $f(\alpha)$ is the probability that a randomly chosen gene will

become a false discovery under α (by false discovery, we mean that the gene is a true negative *and* the gene is declared significant). Let π be the probability that a randomly chosen gene is a true positive, and $\left(1-\bar{\beta}\right)$, the average power. Then,

$s(\alpha)=(1-\pi)\alpha+\pi\left(1-\bar{\beta}\right)$ and $f(\alpha)=(1-\pi)\alpha$. Express f as a function of s, $f=g(s)$. Because $\frac{d\left(1-\bar{\beta}\right)}{d\alpha}>0$ and $\frac{d^2\left(1-\bar{\beta}\right)}{d\alpha^2}<0$, we have

$$\frac{d^2f}{ds^2}=\left(\frac{d^2f}{d\alpha^2}\times\frac{ds}{d\alpha}-\frac{df}{d\alpha}\times\frac{d^2s}{d\alpha^2}\right)\times\left(\frac{ds}{d\alpha}\right)^{-3}$$

$$=-\pi(1-\pi)\times\frac{d^2\left(1-\bar{\beta}\right)}{d\alpha^2}\times\left[(1-\pi)+\pi\times\frac{d\left(1-\bar{\beta}\right)}{d\alpha}\right]^{-3}>0$$

. Thus, $g(.)$ is a convex function.

Assuming $s_P\neq s_N$ we have $\frac{m_Pg(s_P)+m_Ng(s_N)}{m_P+m_N}>g\left(\frac{m_Ps_P+m_Ns_N}{m_P+m_N}\right)$, and after re-arrangement of terms, $\frac{m_Pg(s_P)+m_Ng(s_N)}{m_Ps_P+m_Ns_N}>\frac{m_Pg(s^*)+m_Ng(s^*)}{m_Ps^*+m_Ns^*}$, where $s^*=\frac{m_Ps_P+m_Ns_N}{m_P+m_N}$. Therefore we see that the overall FDR will be higher given the same total number of genes declared significant ($m_Ps_P+m_Ns_N=m_Ps^*+m_Ns^*$), when the two subsets adopt different α's (and hence $s_P\neq s_N$) than when they adopt the same α (and hence $s_P=s_N=s^*$). This implies that for this random prioritization scheme, the 'floating' PSA should adopt a 'fixed' α level (and hence a 'fixed' FDR control) in the two subsets to maximize the total number of true discoveries.

3. Results

3.1. Simulation Studies

We assume that there are a total of m independent genes. Let the hypothesis indicator variables be H_i, where $H_i = 1$ (or 0) if the *i*th gene is a true positive (or a true negative), and $H_i \sim Bernoulli(\pi)$, $i = 1, \cdots m$. Following (Jung, 2005), we assume the effect size (δ) of half of the true positive genes to be 1.0, and that of the remaining half, to be 0.5. The effect sizes of all the true negative genes are fixed at exactly 0. Suppose that about $pri \times 100$ % of genes are to be prioritized, and the remaining $(1 - pri) \times 100$ % of genes, non-prioritized. Let the prioritization indicator variables be U_i, where $U_i = 1$ (or 0) if the *i*th gene is prioritized (or non-prioritized), and $U_i \sim Bernoulli(pri)$, $i = 1, \cdots m$.

Following (Genovese et al., 2006), we define two measures of the informativity of prioritization: $\eta_1 = \Pr(U = 1 | \delta = 0.5) / \Pr(U = 1 | \delta = 0)$ and $\eta_2 = \Pr(U = 1 | \delta = 1.0) / \Pr(U = 1 | \delta = 0)$. $\eta_1 = \eta_2 = 1$ implies that the prior knowledge is totally useless. The prior is informative if $\eta_1 > 1$ or $\eta_2 > 1$ (or both). When $\eta_1 = \eta_2 > 1$, a researcher's prior knowledge is impartial to the effect size of a true positive gene. And the larger the difference between η_1 and η_2 is, the more partial the prior knowledge is.

In our simulation, we set m=10,000, $\pi = 0.1$, and $pri = 0.1$. We first specify a value for η_2 (>1), and let $\eta_1 = \eta_2^f$ with $f = 1$ (no partiality), $1/2$ (moderate partiality), and 0 (strong partiality), respectively. We assume that the total sample size (n) is 30, 60, or 100 ($n/2$ cases and $n/2$ controls). For a true negative gene, its *p* value is generated from the uniform distribution over the interval [0, 1]; for the true positive gene with effect size of δ, from the following cumulative distribution function (Hung et al., 1997),

$$F(p)=\Phi\left(z_{p/2}+\frac{\delta}{\sqrt{\frac{2}{n}+\frac{2}{n}}}\right)+\Phi\left(z_{p/2}-\frac{\delta}{\sqrt{\frac{2}{n}+\frac{2}{n}}}\right),\ 0<p<1,$$

where z_ν and $\Phi(\cdot)$ are the ν th percentile and the cumulative distribution function of the standard normal distribution, respectively.

The three methods of AA, fixed PSA and floating PSA are then applied. We use Storey and Tibshirani's (2003) smoothing spline approach provided by the package 'fdrtool' (Strimmer, 2008a, 2008b) to estimate the proportions of true negative genes, and then transform the *p* values to *q* values using Storey and Tibshirani's algorithm (2003) (See Remark B in their Appendix). The FDRs (or the overall FDR for the floating PSA) are to be controlled at 0.05.

[Figure 1 is about here]

Figure 1 presents the powers (the proportions of true positive genes being detected among all true positive genes) of the three methods, averaged over 1,000 repetitions. We found that the floating PSA is uniformly more powerful than the fixed PSA and the AA for all the scenarios studied. The fixed PSA is more powerful than the AA when there is no partiality ($\eta_1=\eta_2>1$) or when the sample size is smaller. But the situation is quite different when the prior knowledge has strong partiality toward genes with larger effect size. From figure 1, we see that the power of the fixed PSA (where prior knowledge has been utilized) is even lower than that of the AA (where prior knowledge has not been utilized, or cannot be utilized due to lack of it at all). What is more, this power decrease gets even worse as the prior knowledge becomes more informative (larger η_2). By contrast, our proposed floating PSA does not show such paradoxical effects; it becomes

more powerful as the prior knowledge becomes more informative as it should be for all levels of partiality.

[Figure 2 is about here]

As pointed out by Korn et al. (Korn et al., 2007; Korn et al., 2004; Oura et al., 2009), the actual number of false discoveries may not be well controlled through FDR procedures. Hence we study the 'false discovery proportions' (defined as the proportions of true negative genes among all the genes declared significant) of the three methods. Figure 2 is the boxplot of the false discovery proportions in 1,000 repetitions for the three methods, when $\eta_2 = 10$. It can be seen that there is no apparent difference between the distributions of actual false discovery proportions for all the three methods.

[Figure 3 is about here]

[Figure 4 is about here]

Figure 3 presents the FDR levels (q_P and q_N), and figure 4, the significance levels (α_P and α_N), in the floating PSA method. We see that when the overall FDR is to be controlled at 0.05, the FDR level in the *P* subset (q_P) is floating toward a value lower than 0.05, and the FDR level in the *N* subset (q_N), a value higher than 0.05. And this floating away from the overall controlled value is more striking as the prior knowledge becomes more informative and more partial. By contrast, we see that the significance level in the *P* subset (α_P) is higher than the significance level in the *N* subset (α_N). And the difference is more pronounced with more informativity but less partiality. Thus we see that an optimal strategy is one that in terms of FDR control, the *P* subset be made more stringent than the *N* subset, but in terms of significance level, the *P* subset be made more liberal than the *N* subset. (The fixed PSA is less than optimal, because in terms of

FDR, the two subsets are forced to be equal, and in terms of significant level, the *P* subset is too liberal, and the *N* subset, too stringent.)

We present the simulation results when $\pi = 0.2$ and $pri = 0.2$ in the Supplementary Materials. The results are similar to those in Figures 1 ~ 4 when $\pi = 0.1$ and $pri = 0.1$. From figure 1 and the Supplementary Materials, we can see that when the sample size is not large ($n = 30$), the fixed PSA is always better than the AA, under all levels of partiality. When the sample size is larger ($n = 60$ or 100), the power of the fixed PSA is lower than that of the AA, under a strong partiality. This is because when the partiality is strong, most true positive genes in the *P* subset have strong effects, and a liberal selection criterion given by the fixed PSA is a waste (the strong-effect genes can be detected even without this favour), especially when the sample size is also large (this further boosts power of detecting those strong-effect genes). This liberalization on the *P* subset, however wasteful it may be, still needs to be offset by stringency on the *N* subset. And the net result no wonder is the decrease in the total number of true positives being detected. By contrast, our floating PSA avoids an over-liberal selection criterion given to the *P* subset, and trades it with a less stringent selection criterion to the *N* subset.

In our simulations, η_2 is specified to be no less than 1. When η_2 is less than 1, the merits of the floating PSA remain to be seen (see Supplementary Materials), because of the symmetric role for the *P* and *N* subsets. Essentially, the floating PSA works equally well on ‘authentic information’ (true positive proportion in the *P* subset is raised higher because of the information) or on ‘jumble up information’ (true positive proportion in the *P* subset on the contrary is being lowered down because of the information).

3.2. Application to a Leukemia Data Set

As an example, the three FDR methods were applied to a leukemia data set. Golub *et al.* (1999) reported 38 bone marrow samples obtained from acute leukemia patients, including 27 patients with acute lymphoblastic leukemia (ALL) and 11 patients with acute myeloid leukemia (AML). For each sample, the levels of a total of 7,129 'gene expressions' (henceforth referred to simply as 'genes') were monitored by DNA microarrays produced by Affymetrix. The purpose of the study is to detect genes that are differentially expressed between ALL and AML patients. The *p* values were calculated using the Mann-Whitney test for the 7,129 study genes. If we control the family-wise error rate at 0.05 (the level of significance for each gene being set at $0.05/7129 = 7.01\times10^{-6}$ with the Bonferroni correction), a total of 30 significant genes can be identified. By contrast, if we apply the AA method to control the FDR at 0.05 (the level of significance for each gene is 0.0091), a total of 826 significant genes can be identified.

Both the fixed PSA and the floating PSA need prior knowledge to prioritize genes. We acquired this information from the Leukemia Gene Database of the Bioinformatics Organization (www.bioinformatics.org/legend/leuk_db.htm) as well as other related literature (Antras et al., 1991; Blaser, 2002; Chertov et al., 1996; Dickstein et al., 2001; Domer et al., 1993; Guzman et al., 2001; Handen and Rosenberg, 1997; Hayakawa et al., 1998; Janssen and Marynen, 2006; Klein et al., 1996; Kohka et al., 1998; Levine et al., 2005; Melhem et al., 1997; Melki et al., 1999; Mullighan, 2009; Pigazzi et al., 2008; Rhee et al., 1995; Schafer et al., 1996; Shimada et al., 2002; Silver et al., 1999; Towatari et al., 1997; Wakioka et al., 1999; Wu et al., 1992; Yasumoto and Shibahara, 1997). Any of the 7,129 study genes is to be prioritized if it appears in the Leukemia Gene Database

or be mentioned in any of the above literature. We are not fastidious about the prioritization. For example, knowing that KIAA1438 (Dickstein et al., 2001), KIAA0202 (Blaser, 2002), and KIAA1509 (Levine et al., 2005) had been reported to be related to leukemia, we went straight ahead to prioritize the entire KIAA family, a total of 256 genes in this data set, although many of them may have little relevance to leukemia. In addition, we searched among this leukemia data set for gene names with 'leuk-', 'lymph-', 'myel-', 'hemo-', 'platelet', and 'monocyte' to prioritize haematologically-related genes. In the end, a total of 819 genes were prioritized. We then transformed *p* values into *q* values using Storey and Tibshirani's smoothing spline approach (Storey and Tibshirani, 2003) (with the package 'fdrtool' (Strimmer, 2008a, 2008b), and the proportions of true negative genes were estimated as $\hat{\pi}_0 = 0.63$, $\hat{\pi}_{0,P} = 0.46$, and $\hat{\pi}_{0,N} = 0.65$, respectively).

[Table 1 is about here]

Table 1 presents our analysis result of this leukemia data set. The FDRs (or the overall FDR for the floating PSA) are to be controlled at 0.05. It can be seen that although the fixed PSA identified many more genes ($+72$) in the *P* subset than the AA, it missed an even larger portion of genes (-120) in the *N* subset as compared to AA. As a result, the total number of genes identified by the fixed PSA in the two subsets combined is 48 genes less than the AA. As for the floating PSA, it identified 42 more genes in the *P* subset and 10 less genes in the *N* subset than the AA. And the net result is that it identified 32 more genes than the AA (and 80 more genes than the fixed PSA).

Table 1 also presents the FDR values in the two subsets for the floating PSA method. As expected, the FDR is being lowered (0.0314) in the *P* subset and being raised higher (0.0568) in the *N* subset, as compared to its overall controlled value of 0.05. In terms of

significance levels, we see that the fixed PSA is very liberal in the *P* subset ($+0.0245$ as compared to AA), but too stringent on genes in the *N* subset (-0.0029 as compared to AA). By comparison, the floating PSA strikes a middle ground. It is neither overly too liberal ($+0.0097$ as compared to AA) in the *P* subset, nor overly too stringent (-0.0004 as compared to AA) in the *N* subset.

To evaluate how well the FDR is controlled within the two subsets, we further estimated the permutation-based FDR (Xie et al., 2005). We randomly permuted the data and calculated the null *P* values, $p_i^{(b)}$, for the *i*th gene in the *b*th permutation (*i*=1,…, 7129). Through *B* permutations, the numbers of false positives (*FP*) of the two subsets are estimated as $\hat{FP}(\alpha_P) = \sum_{b=1}^{B} \#\left\{i : p_i^{(b)} \le \alpha_P\right\} / B$ (for those *i's* belonging to the prioritized subset) and $\hat{FP}(\alpha_N) = \sum_{b=1}^{B} \#\left\{i : p_i^{(b)} \le \alpha_N\right\} / B$ (for those *i's* belonging to the non-prioritized subset), where $\alpha_P = 0.0188$ and $\alpha_N = 0.0087$ (for the floating PSA), or $\alpha_P = 0.0336$ and $\alpha_N = 0.0062$ (for the fixed PSA). For AA, the number of false positives (*FP*) for the aggregate set is estimated as $\hat{FP}(\alpha) = \sum_{b=1}^{B} \#\left\{i : p_i^{(b)} \le \alpha\right\} / B$ (for all *i*), where $\alpha = 0.0091$. With *B*=100,000, the permutation-based FDR of AA is 0.0417, and those of the fixed PSA and the floating PSA are listed in Table 1. They are all less than our FDR control levels, suggesting satisfactory FDR controls for all the three methods.

In the Supplementary Materials, we detailed the results for the 7,129 genes.

3.3. Application to a Colon Cancer Data Set

Another example is a colon cancer study (Alon et al., 1999). The data set contains 2,000 gene expressions in 40 tumor and 22 normal colon tissue samples. The *p* values

were calculated using the Mann-Whitney test for the 2,000 study genes. Controlling the family-wise error rate at 0.05, a total of 16 significant genes can be identified with the Bonferroni correction (the level of significance for each gene is 2.5×10^{-5}). However, if we apply the AA method to control the FDR at 0.05 (the level of significance for each gene is 0.0037), a total of 128 significant genes can be identified.

We further obtained prior knowledge (to prioritize genes) from a number of papers related to colon cancer (Barnard et al., 1995; Bolmont et al., 1990; Burdick and Konstantopoulos, 2004; Easwaran et al., 1999; Kowalski and Denhardt, 1989; Kuo et al., 1995; Lee et al., 2004; Lee et al., 2006; Lee et al., 2008; Mori et al., 1993; Narisawa et al., 1994; Rothman et al., 1996; Sinicrope et al., 1995; Song et al., 2005; Tozzi et al., 1991; Yoo et al., 2004). Finally, a total of 156 genes were prioritized. We then transformed *p* values into *q* values using Storey and Tibshirani's smoothing spline approach (Storey and Tibshirani, 2003) (with the package 'fdrtool' (Strimmer, 2008a, 2008b), and the proportions of true negative genes were estimated as $\hat{\pi}_0 = 0.86$, $\hat{\pi}_{0,P} = 0.35$, and $\hat{\pi}_{0,N} = 0.91$, respectively). Table 2 presents the result of this example. The FDRs (or the overall FDR for the floating PSA) are to be controlled at 0.05. In this data set, our floating PSA identified a total of 171 differentially expressed genes, while the fixed PSA, a total of 149 genes.

[Table 2 is about here]

In this colon cancer data set, we see that the fixed PSA isn't so counter-productive as in the previous leukemia data set—the fixed PSA now identifies 21 more genes than the AA. But we see that it is again the floating PSA that identifies the largest number of genes (43 more genes than the AA and 22 more genes than the fixed PSA). We also list

the permutation-based FDRs of the fixed PSA and the floating PSA in Table 2 (that of AA is 0.0505), which are less than or very close to our FDR control levels, suggesting satisfactory FDR controls for all the three methods.

We detailed the results for the 2,000 genes of this example in the Supplementary Materials.

4. Discussion

When testing simultaneously for a very large number of genes, there is no guarantee that those true positive genes have the most extreme *p* values and stand out among all the noises—there are simply too many genes to compete with. The PSA works by creating two subsets of genes with different genetic make ups. (By genetic make up, we mean the proportion of true positives and the average effect size of a gene subset.) The true positive genes now need only to compete for standing out with the other genes in the same subset, but they don't need to compete with all those genes in the other subset. This may increase the chance of true positive genes' standing out (i.e., being detected). In this paper, we perfect the PSA technique by charging it with floating FDR control. We show that the floating PSA is uniformly more powerful than the fixed PSA and the AA.

The floating PSA algorithm is completely data driven. It is robust to any prior knowledge used for prioritization, be it informative or non-informative, partial or impartial. In essence, it can automatically 'sense' the difference in the genetic make ups between the two subsets of genes and adjust the FDR control values accordingly to maximize the overall power. Even if a researcher's prior knowledge is 'anti-informative' (the prioritized has a smaller proportion of true positives than the non-prioritized) and 'anti-partial' (the prioritized has smaller effect sizes than the non-prioritized), the floating

PSA can still make use of this jumbled up information to boost the power equally well. The floating PSA is also a foolproof method. The worst scenario one can conceive is when the prior knowledge amounts to nothing but tossing a coin. In that case, the floating PSA still holds up; it maintains the same power as the AA.

The floating PSA does have its limitations. First, similar with the stratified FDR control (Sun et al., 2006) and the fixed PSA (Li et al., 2008), the number of genes in both subsets should not be too small, if the within-subset and/or the overall FDR are to be adequately controlled. Second, as pointed out by Qiu and Yakovlev (2006), FDR estimates may become unstable because of the complicated correlation structure among gene expressions. In our analysis of the leukemia and the colon cancer datasets, the smallest subset we made contains 156 genes. Also, we performed permutation-based FDR (Xie et al., 2005) to evaluate how well the FDRs are actually controlled. And the results are quite satisfactory.

The floating PSA method can be compared to the *p*-value-weighting FDR approach (Genovese et al., 2006) and the Bayesian FDR method (Whittemore, 2007). To use the *p*-value-weighting FDR approach, one must assign a weight-of-evidence (of being a true positive) for each and every gene. The method then incorporates the information into the FDR control. To use Bayesian FDR control, one must supply the prior odds (of being a true positive) for each and every gene under study. Combining these with the data (the observed test statistics), the method then generates a posterior odds (of being a true positive) for each and every gene. The prior knowledge (weight-of-evidence or prior odds) in these two methods is in the form of a continuous variable, which must come from previous studies or from best available biological knowledge to date. By comparison, the

floating PSA method, with its binary categorization of genes, may seem to be a very crude use of prior knowledge. However, this crudeness turns out also to be its virtue. First, we see that dichotomizing the genes into a *P* and an *N* subset is much easier to do in practice than assigning painstakingly a continuous-scale value to each and every gene. Second, as mentioned earlier, the floating PSA is remarkably robust—to the point that it can feed on jumbled up information equally well, whereas the *p*-value-weighting method and the Bayesian FDR method will suffer from power loss if the prior knowledge is flawed.

A possible extension of the floating PSA method would be to construct more than two subsets. This would be applicable to the situations when the study genes by themselves are making several subsets, i.e., they come from several groups/families, each having distinct biological functions/pathways. The floating PSA method may also be suitable for detecting gene-gene interactions. In a gene-gene interaction study, one can allocate the gene pairs (or gene clusters) to the *P* subset if any one gene in the gene pairs/gene clusters were previously found to be associated with the study outcome, and to the *N* subset if otherwise. Finally, the floating PSA method can do away with pre-selecting a value (say, 0.05) for the overall FDR (the q^*). To achieve this, Cheng et al's method of adaptive threshold criteria (Cheng et al., 2004) can be extended to two-subset situation, to balance the false positive and false negative errors jointly in the *P* and the *N* subsets. All these topics are worthy of further studies.

In conclusion, the floating PSA method is to be recommended for the detection of differentially expressed genes, in light of its power, robustness, and ease of implementation.

Acknowledgements

We thank the anonymous reviewers for their constructive comments. We thank Golub *et al.* (Golub et al., 1999) and Alon et al. (Alon et al., 1999) for making their data available to the scientific community. This study was supported by National Science Councils, Taiwan.

References

Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D., Levine, A.J., 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci U S A 96, 6745-6750.

Antras, J., Lasnier, F., Pairault, J., 1991. Adipsin gene expression in 3T3-F442A adipocytes is posttranscriptionally down-regulated by retinoic acid. J Biol Chem 266, 1157-1161.

Barnard, G.F., Mori, M., Staniunas, R.J., Begum, N.A., Bao, S., Puder, M., Cobb, J., Redman, K.L., Steele, G.D., Jr., Chen, L.B., 1995. Ubiquitin fusion proteins are overexpressed in colon cancer but not in gastric cancer. Biochim Biophys Acta 1272, 147-153.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. B. 57, 289-300.

Benjamini, Y., Hochberg, Y., 1997. Multiple hypotheses testing with weights. Scand. J. Stat. 24, 407-418.

Blaser, S., Jersch, K., Hainmann, I., Wunderle, D., Zgaga-Griesz, A., Busse, A., Zieger, B., 2002. Human septin-septin interaction: CDCrel-1 partners with KIAA0202. FEBS Lett. 519, 169-172.

Bolmont, C., Lilienbaum, A., Paulin, D., Grimaud, J.A., 1990. Expression of desmin gene in skeletal and smooth muscle by in situ hybridization using a human desmin gene probe. J Submicrosc Cytol Pathol 22, 117-122.

Burdick, M.M., Konstantopoulos, K., 2004. Platelet-induced enhancement of LS174T colon carcinoma and THP-1 monocytoid cell adhesion to vascular endothelium under flow. Am J Physiol Cell Physiol 287, C539-547.

Cheng, C., Pounds, S.B., Boyett, J.M., Pei, D., Kuo, M.L., Roussel, M.F., 2004. Statistical significance threshold criteria for analysis of microarray gene expression data. Stat Appl Genet Mol Biol 3, Article36.

Chertov, O., Michiel, D.F., Xu, L., Wang, J.M., Tani, K., Murphy, W.J., Longo, D.L., Taub, D.D., Oppenheim, J.J., 1996. Identification of defensin-1, defensin-2, and

CAP37/azurocidin as T-cell chemoattractant proteins released from interleukin-8-stimulated neutrophils. J Biol Chem 271, 2935-2940.
Dickstein, J.I., Davis, E.M., Roulston, D., 2001. Localization of the chromosome 22 breakpoints in two cases of acute megakaryoblastic leukemia with t(1;22)(p13;q13). Cancer Genet Cytogenet 129, 150-154.
Domer, P.H., Fakharzadeh, S.S., Chen, C.S., Jockel, J., Johansen, L., Silverman, G.A., Kersey, J.H., Korsmeyer, S.J., 1993. Acute mixed-lineage leukemia t(4;11)(q21;q23) generates an MLL-AF4 fusion product. Proc Natl Acad Sci U S A 90, 7884-7888.
Easwaran, V., Song, V., Polakis, P., Byers, S., 1999. The ubiquitin-proteasome pathway and serine kinase activity modulate adenomatous polyposis coli protein-mediated regulation of beta-catenin-lymphocyte enhancer-binding factor signaling. J Biol Chem 274, 16641-16645.
Genovese, C., Roeder, K., Wasserman, L., 2006. False discovery control with P-value weighting. Biometrika 93, 509-524.
Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S., 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286, 531-537.
Greenwood, C.M., Rangrej, J., Sun, L., 2007. Optimal selection of markers for validation or replication from genome-wide association studies. Genet Epidemiol 31, 396-407.
Guzman, M.L., Neering, S.J., Upchurch, D., Grimes, B., Howard, D.S., Rizzieri, D.A., Luger, S.M., Jordan, C.T., 2001. Nuclear factor-kappaB is constitutively activated in primitive human acute myelogenous leukemia cells. Blood 98, 2301-2307.
Handen, J.S., Rosenberg, H.F., 1997. Intronic enhancer activity of the eosinophil-derived neurotoxin (RNS2) and eosinophil cationic protein (RNS3) genes is mediated by an NFAT-1 consensus binding sequence. J Biol Chem 272, 1665-1669.
Hayakawa, F., Towatari, M., Iida, H., Wakao, H., Kiyoi, H., Naoe, T., Saito, H., 1998. Differential constitutive activation between STAT-related proteins and MAP kinase in primary acute myelogenous leukaemia. Br J Haematol 101, 521-528.
Hung, H.M., O'Neill, R.T., Bauer, P., Kohne, K., 1997. The behavior of the P-value when the alternative hypothesis is true. Biometrics 53, 11-22.
Janssen, H., Marynen, P., 2006. Interaction partners for human ZNF384/CIZ/NMP4--zyxin as a mediator for p130CAS signaling? Exp Cell Res 312, 1194-1204.
Jung, S.H., 2005. Sample size for FDR-control in microarray data analysis. Bioinformatics 21, 3097-3104.
Klein, S.C., Kube, D., Abts, H., Diehl, V., Tesch, H., 1996. Promotion of IL8, IL10, TNF alpha and TNF beta production by EBV infection. Leuk Res 20, 633-636.
Kohka, H., Yoshino, T., Iwagaki, H., Sakuma, I., Tanimoto, T., Matsuo, Y., Kurimoto, M., Orita, K., Akagi, T., Tanaka, N., 1998. Interleukin-18/interferon-gamma-inducing factor, a novel cytokine, up-regulates ICAM-1 (CD54) expression in KG-1 cells. J Leukoc Biol 64, 519-527.
Korn, E.L., Li, M.C., McShane, L.M., Simon, R., 2007. An investigation of two multivariate permutation methods for controlling the false discovery proportion. Stat Med 26, 4428-4440.

Korn, E.L., Troendle, J.F., McShane, L.M., Simon, R., 2004. Controlling the number of false discoveries: application to high-dimensional genomic data. Journal of Statistical Planning and Inference 124, 379-398.

Kowalski, J., Denhardt, D.T., 1989. Regulation of the mRNA for monocyte-derived neutrophil-activating peptide in differentiating HL60 promyelocytes. Mol Cell Biol 9, 1946-1957.

Kuo, M.L., Huang, T.S., Lin, J.K., 1995. Preferential requirement for protein tyrosine phosphatase activity in the 12-O-tetradecanoylphorbol-13-acetate-induced differentiation of human colon cancer cells. Biochem Pharmacol 50, 1217-1222.

Lee, D.H., Anderson, K.E., Harnack, L.J., Folsom, A.R., Jacobs, D.R., Jr., 2004. Heme iron, zinc, alcohol consumption, and colon cancer: Iowa Women's Health Study. J Natl Cancer Inst 96, 403-407.

Lee, R.A., Kim, H.A., Kang, B.Y., Kim, K.H., 2006. Hemoglobin induces colon cancer cell proliferation by release of reactive oxygen species. World J Gastroenterol 12, 5644-5650.

Lee, W.S., Seo, G., Shin, H.J., Yun, S.H., Yun, H., Choi, N., Lee, J., Son, D., Cho, J., Kim, J., Cho, Y.B., Chun, H.K., Lee, W.Y., 2008. Identification of differentially expressed genes in microsatellite stable HNPCC and sporadic colon cancer. J Surg Res 144, 29-35.

Levine, R.L., Wadleigh, M., Sternberg, D.W., Wlodarska, I., Galinsky, I., Stone, R.M., DeAngelo, D.J., Gilliland, D.G., Cools, J., 2005. KIAA1509 is a novel PDGFRB fusion partner in imatinib-responsive myeloproliferative disease associated with a t(5;14)(q33;q32). Leukemia 19, 27-30.

Li, C., Li, M., Lange, E.M., Watanabe, R.M., 2008. Prioritized subset analysis: improving power in genome-wide association studies. Hum Hered 65, 129-141.

Lin, W.-Y., Lee, W.-C., 2010. Incorporating prior knowledge to facilitate discoveries in a genome-wide association study on age-related macular degeneration. BMC Research Notes 3, 26.

Melhem, R., Hailat, N., Kuick, R., Hanash, S.M., 1997. Quantitative analysis of Op18 phosphorylation in childhood acute leukemia. Leukemia 11, 1690-1695.

Melki, J.R., Vincent, P.C., Clark, S.J., 1999. Concurrent DNA hypermethylation of multiple genes in acute myeloid leukemia. Cancer Res 59, 3730-3740.

Mori, M., Barnard, G.F., Staniunas, R.J., Jessup, J.M., Steele, G.D., Jr., Chen, L.B., 1993. Prothymosin-alpha mRNA expression correlates with that of c-myc in human colon cancer. Oncogene 8, 2821-2826.

Mullighan, C.G., Downing,J.R., 2009. Genome-wide profiling of genetic alterations in acute lymphoblastic leukemia: recent insights and future directions. Leukemia in press.

Narisawa, T., Fukaura, Y., Terada, K., Umezawa, A., Tanida, N., Yazawa, K., Ishikawa, C., 1994. Prevention of 1,2-dimethylhydrazine-induced colon tumorigenesis by HMG-CoA reductase inhibitors, pravastatin and simvastatin, in ICR mice. Carcinogenesis 15, 2045-2048.

Oura, T., Matsui, S., Kawakami, K., 2009. Sample size calculations for controlling the distribution of false discovery proportion in microarray experiments. Biostatistics 10, 694-705.

Pigazzi, M., Manara, E., Baron, E., Basso, G., 2008. ICER expression inhibits leukemia phenotype and controls tumor progression. Leukemia 22, 2217-2225.
Qiu, X., Yakovlev, A., 2006. Some comments on instability of false discovery rate estimation. Journal of Bioinformatics and Computational Biology 4, 1057-1068.
Rhee, K., Bresnahan, W., Hirai, A., Hirai, M., Thompson, E.A., 1995. c-Myc and cyclin D3 (CcnD3) genes are independent targets for glucocorticoid inhibition of lymphoid cell proliferation. Cancer Res 55, 4188-4195.
Roeder, K., Bacanu, S.A., Wasserman, L., Devlin, B., 2006. Using linkage genome scans to improve power of association in genome scans. Am J Hum Genet 78, 243-252.
Rothman, T.P., Chen, J., Howard, M.J., Costantini, F., Schuchardt, A., Pachnis, V., Gershon, M.D., 1996. Increased expression of laminin-1 and collagen (IV) subunits in the aganglionic bowel of ls/ls, but not c-ret -/- mice. Dev Biol 178, 498-513.
Schafer, H., Trauzold, A., Siegel, E.G., Folsch, U.R., Schmidt, W.E., 1996. PRG1: a novel early-response gene transcriptionally induced by pituitary adenylate cyclase activating polypeptide in a pancreatic carcinoma cell line. Cancer Res 56, 2641-2648.
Shimada, H., Ichikawa, H., Ohki, M., 2002. Potential involvement of the AML1-MTG8 fusion protein in the granulocytic maturation characteristic of the t(8;21) acute myelogenous leukemia revealed by microarray analysis. Leukemia 16, 874-885.
Silver, A., Moody, J., Dunford, R., Clark, D., Ganz, S., Bulman, R., Bouffler, S., Finnon, P., Meijne, E., Huiskamp, R., Cox, R., 1999. Molecular mapping of chromosome 2 deletions in murine radiation-induced AML localizes a putative tumor suppressor gene to a 1.0 cM region homologous to human chromosome segment 11p11-12. Genes Chromosomes Cancer 24, 95-104.
Sinicrope, F.A., Hart, J., Michelassi, F., Lee, J.J., 1995. Prognostic value of bcl-2 oncoprotein expression in stage II colon carcinoma. Clin Cancer Res 1, 1103-1110.
Song, S., Byrd, J.C., Mazurek, N., Liu, K., Koo, J.S., Bresalier, R.S., 2005. Galectin-3 modulates MUC2 mucin expression in human colon cancer cells at the level of transcription via AP-1 activation. Gastroenterology 129, 1581-1591.
Storey, J.D., 2002. A direct approach to false discovery rates. J. R. Stat. Soc. B. 64, 479-498.
Storey, J.D., Tibshirani, R., 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100, 9440-9445.
Strimmer, K., 2008a. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics 24, 1461-1462.
Strimmer, K., 2008b. A unified approach to false discovery rate estimation. BMC Bioinformatics 9, 303.
Sun, L., Craiu, R.V., Paterson, A.D., Bull, S.B., 2006. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. Genet Epidemiol 30, 519-530.
Towatari, M., Iida, H., Tanimoto, M., Iwata, H., Hamaguchi, M., Saito, H., 1997. Constitutive activation of mitogen-activated protein kinase pathway in acute leukemia cells. Leukemia 11, 479-484.
Tozzi, M.G., Camici, M., Pesi, R., Allegrini, S., Sgarrella, F., Ipata, P.L., 1991. Nucleoside phosphotransferase activity of human colon carcinoma cytosolic 5'-nucleotidase. Arch Biochem Biophys 291, 212-217.

Wakioka, T., Sasaki, A., Mitsui, K., Yokouchi, M., Inoue, A., Komiya, S., Yoshimura, A., 1999. APS, an adaptor protein containing Pleckstrin homology (PH) and Src homology-2 (SH2) domains inhibits the JAK-STAT pathway in collaboration with c-Cbl. Leukemia 13, 760-767.

Whittemore, A., 2007. A Bayesian false discovery rate for multiple testing. Journal of Applied Statistics 34, 1-9.

Wu, B.Y., Hanley, E.W., Turka, L.A., Nabel, G.J., 1992. Isolation of a cDNA clone encoding a zinc finger protein highly expressed in T-leukemia lines. Blood 80, 2571-2576.

Xie, Y., Pan, W., Khodursky, A.B., 2005. A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. Bioinformatics 21, 4280-4288.

Yasumoto, K., Shibahara, S., 1997. Molecular cloning of cDNA encoding a human TFEC isoform, a newly identified transcriptional regulator. Biochim Biophys Acta 1353, 23-31.

Yoo, B.C., Jeon, E., Hong, S.H., Shin, Y.K., Chang, H.J., Park, J.G., 2004. Metabotropic glutamate receptor 4-mediated 5-Fluorouracil resistance in a human colon cancer cell line. Clin Cancer Res 10, 4176-4184.

Figures

Figure 1. Power of the three methods

Power of the AA (solid lines), fixed PSA (- △ -), and floating PSA (- + -). The total sample size is set at 30, 60, and 100 (from top to bottom), respectively.

Figure 2. Boxplot of false discovery proportions in 1,000 repetitions for the three methods

The total sample size is set at 30, 60, and 100 (from top to bottom), respectively. The value of η_2 is fixed at 10 for all the scenarios.

Figure 3. FDR levels in the floating PSA method

The FDR levels (q_P for the prioritized subset and q_N for the non-prioritized subset) in the floating PSA method (solid lines, $\eta_1 = \eta_2$; broken lines, $\eta_1 = \sqrt{\eta_2}$; dotted lines, $\eta_1 = 1$). The total sample size is set at 30, 60, and 100 (from top to bottom), respectively.

Figure 4. Significance levels in the floating PSA method

The significance levels (α_P for the prioritized subset and α_N for the non-prioritized subset) in the floating PSA method (solid lines, $\eta_1 = \eta_2$; broken lines, $\eta_1 = \sqrt{\eta_2}$; dotted lines, $\eta_1 = 1$). The total sample size is set at 30, 60, and 100 (from top to bottom), respectively.

Tables

Table 1. Data analysis for the leukemia data set. The FDRs (or the overall FDR for the floating PSA) are to be controlled at 0.05.

	Prioritized Subset (Number of genes = 819)	Non-Prioritized Subset (Number of genes = 6,310)	Total (Number of genes = 7,129)
Number of Significant Genes			
fixed PSA (difference between fixed PSA and AA)	258 (+ 72)	520 (– 120)	778 (– 48)
floating PSA (difference between floating PSA and AA)	228 (+ 42)	630 (– 10)	858 (+ 32)
False Discovery Rate			
fixed PSA	to be controlled at 0.05	to be controlled at 0.05	
[permutation-based FDR)	[0.0458]	[0.0408]	[0.0425]
floating PSA	0.0314	0.0568	to be controlled at 0.05
[permutation-based FDR]	[0.0285]	[0.0479]	[0.0427]
Level of Significance			
fixed PSA (difference between fixed PSA and AA)	0.0336 (+ 0.0245)	0.0062 (– 0.0029)	–
floating PSA (difference between floating PSA and AA)	0.0188 (+ 0.0097)	0.0087 (– 0.0004)	–

Table 2. Data analysis for the colon cancer data set. The FDRs (or the overall FDR for the floating PSA) are to be controlled at 0.05.

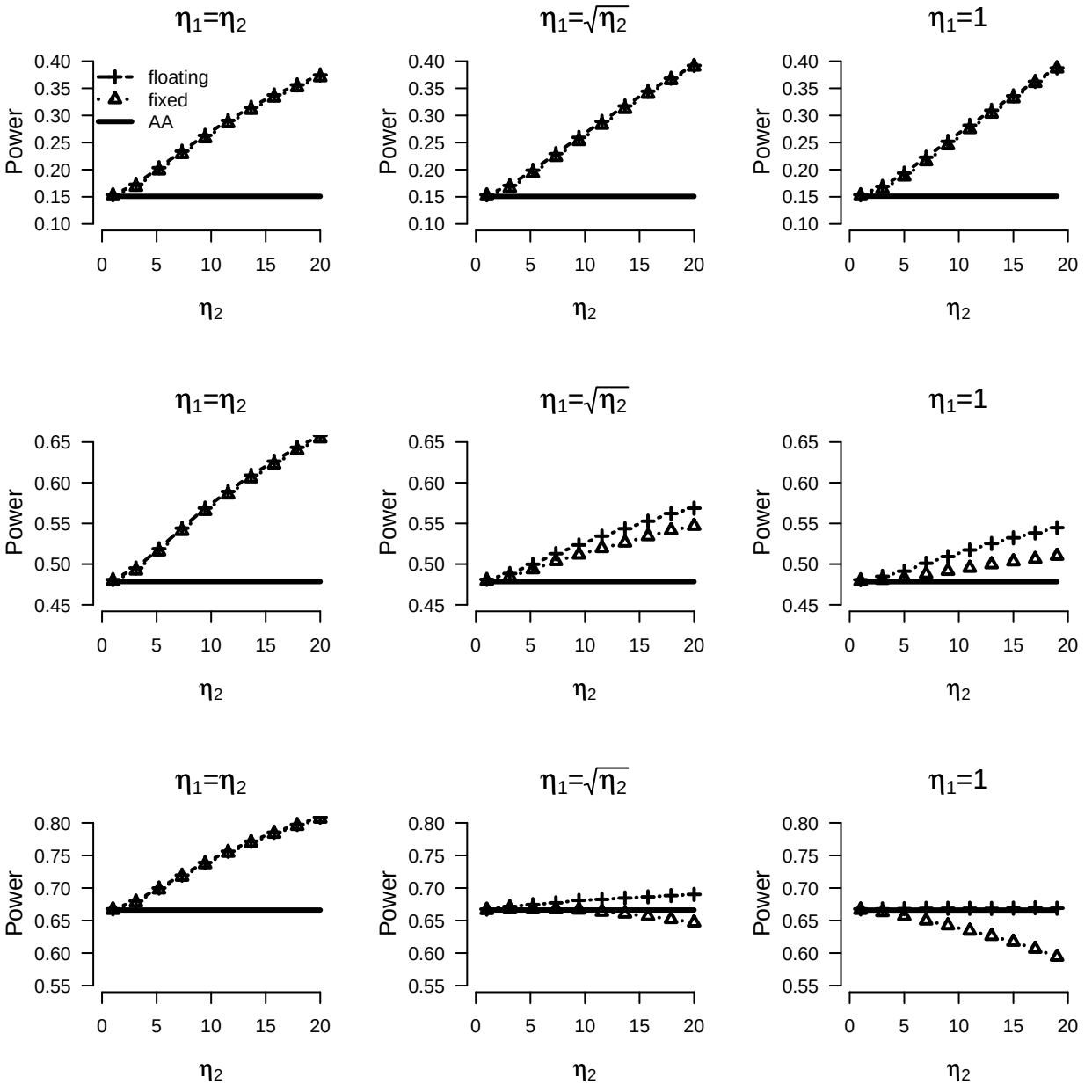
	Prioritized Subset (Number of genes = 156)	Non-Prioritized Subset (Number of genes = 1,844)	Total (Number of genes = 2,000)
Number of Significant Genes			
fixed PSA (difference between fixed PSA and AA)	67 (+43)	82 (−22)	149 (+21)
floating PSA (difference between floating PSA and AA)	58 (+34)	113 (+ 9)	171 (+43)
False Discovery Rate			
fixed PSA	to be Controlled at 0.05	to be controlled at 0.05	
[permutation-based FDR]	[0.0470]	[0.0511]	[0.0493]
floating PSA	0.0238	0.0635	to be controlled at 0.05
[permutation-based FDR]	[0.0233]	[0.0640]	[0.0502]
Level of Significance			
fixed PSA (difference between fixed PSA and AA)	0.0583 (+0.0560)	0.0025 (−0.0012)	—
floating PSA (difference between floating PSA and AA)	0.0250 (+0.0227)	0.0043 (+0.0006)	—

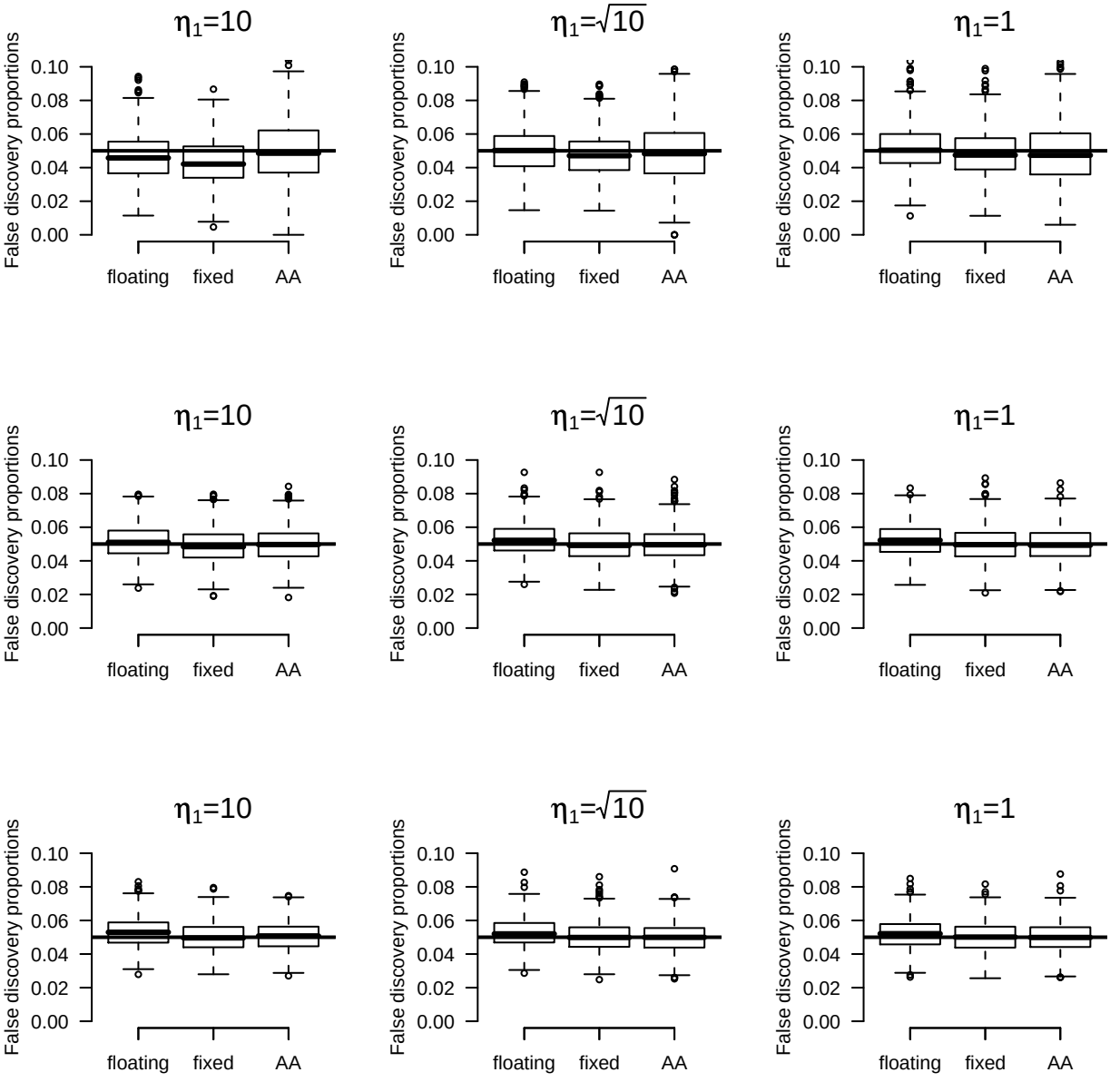
Supplementary Materials

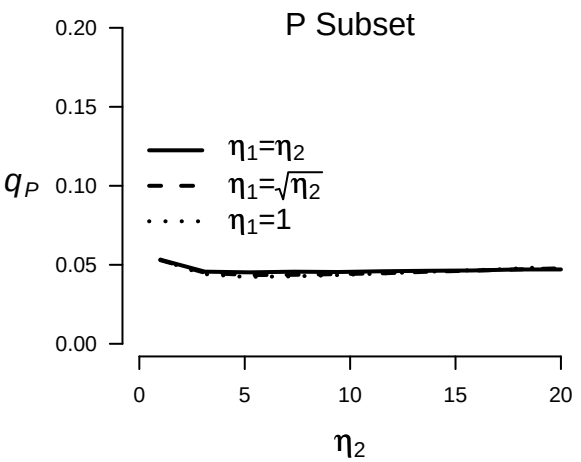
file 1 – Simulation results of other scenarios

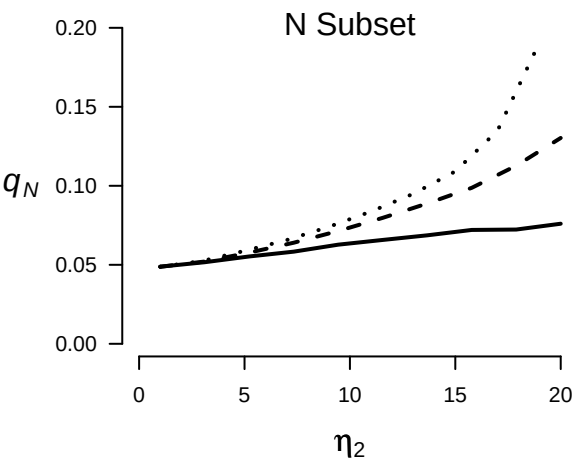
file 2 – The detailed results of the leukemia data analysis

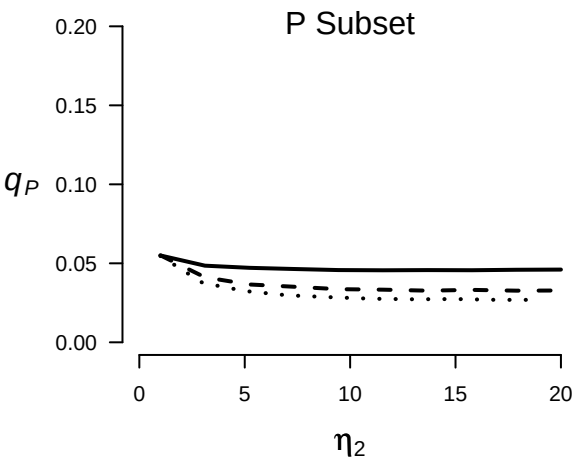
file 3 – The detailed results of the colon cancer data analysis

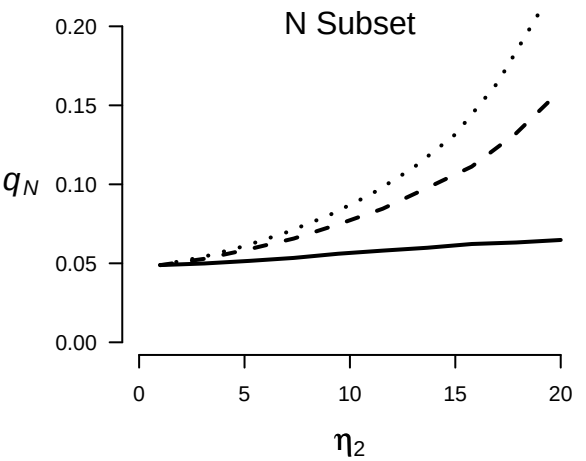

η1=η2
η1=√η2
η1=1
floating
fixed
AA
Power
η2
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.65
0.60
0.55
0.50
0.45
0.80
0.75
0.70
0.65
0.60
0.55
0
5
10
15
20

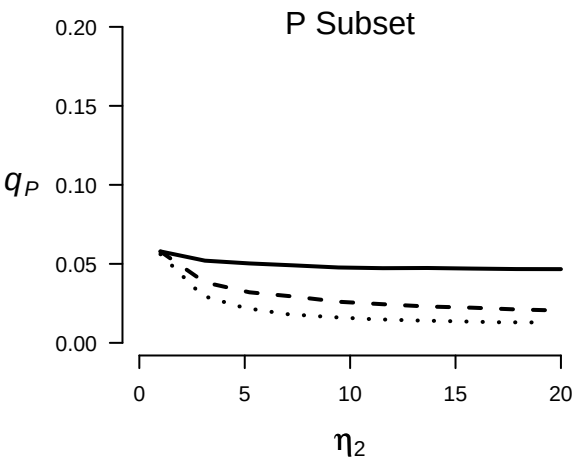

η1=10
η1=√10
η1=1
False discovery proportions
0.00
0.02
0.04
0.06
0.08
0.10
floating
fixed
AA

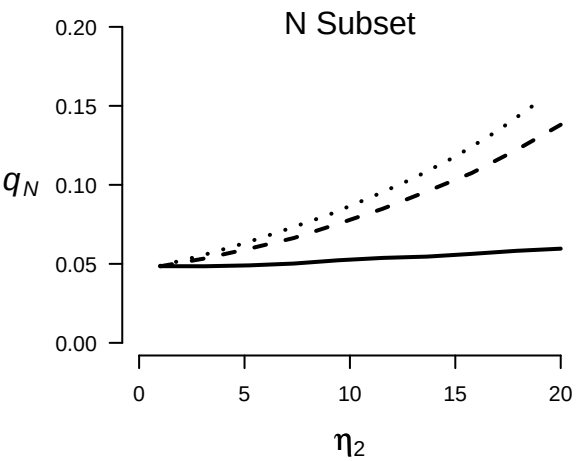

P Subset
q_P
0.20
0.15
0.10
0.05
0.00
$\eta_1=\eta_2$
$\eta_1=\sqrt{\eta_2}$
$\eta_1=1$
0
5
10
15
20
η_2

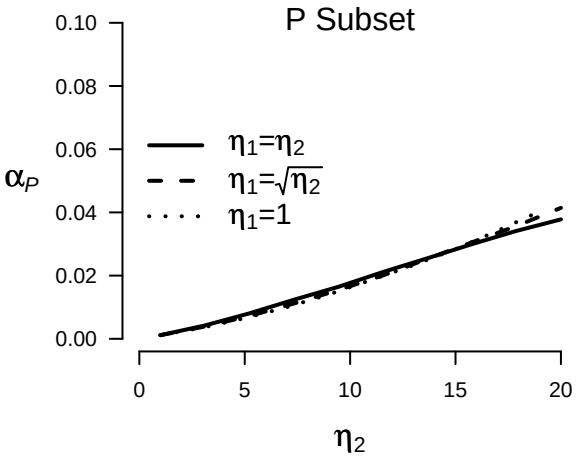

N Subset
q_N
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
η_2

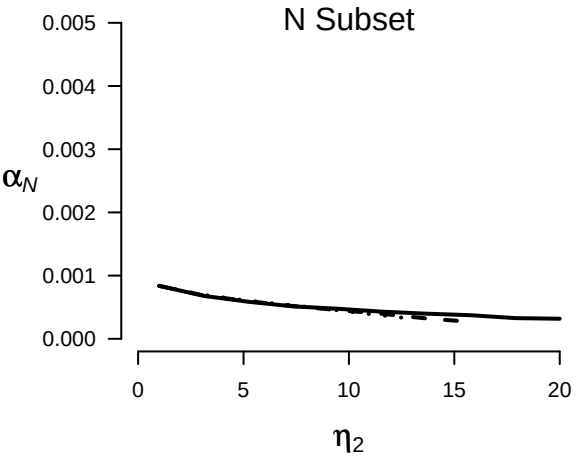

P Subset
q_P
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
η_2

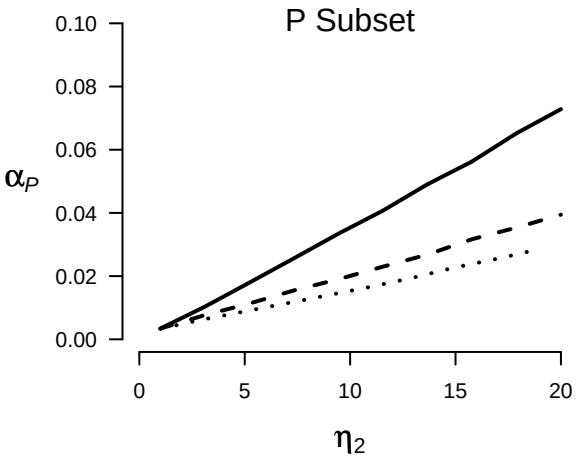

N Subset
q_N
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
η_2

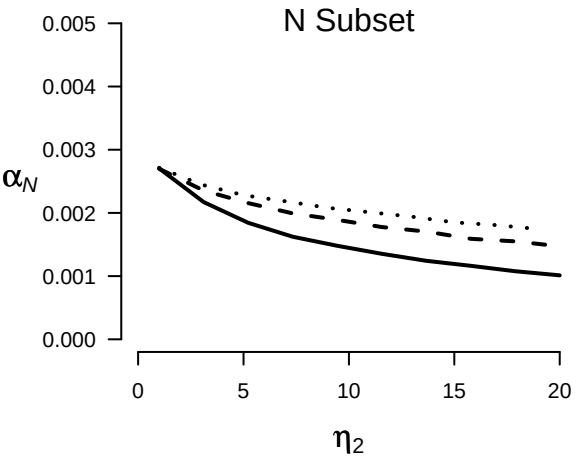

P Subset
q_P
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
η_2

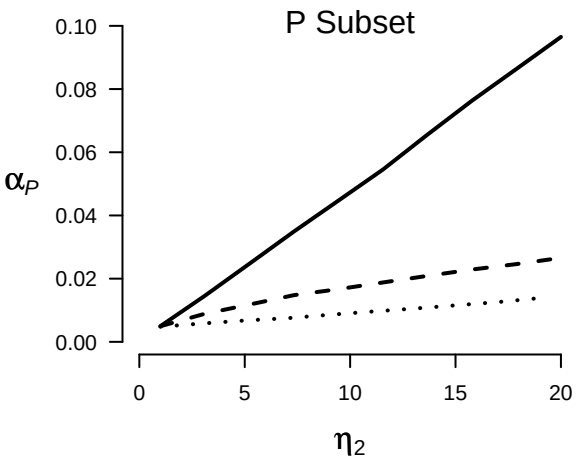

N Subset
q_N
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
η_2

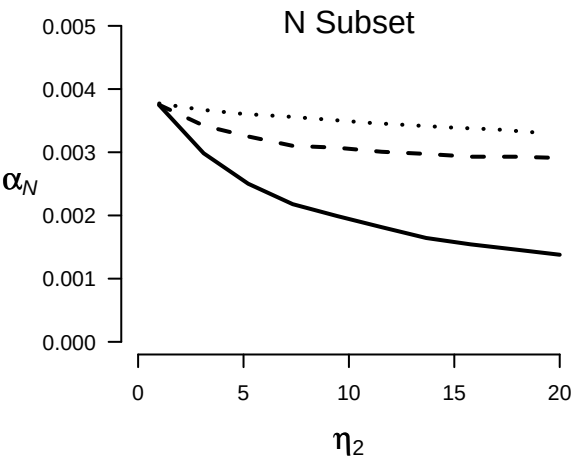

P Subset
α_P
$\eta_1=\eta_2$
$\eta_1=\sqrt{\eta_2}$
$\eta_1=1$
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
η_2


N Subset
α_N
0.005
0.004
0.003
0.002
0.001
0.000
0
5
10
15
20
η_2


P Subset
α_P
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
η_2


N Subset
α_N
0.005
0.004
0.003
0.002
0.001
0.000
0
5
10
15
20
η_2


P Subset
α_P
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
η_2


N Subset
α_N
0.005
0.004
0.003
0.002
0.001
0.000
0
5
10
15
20
η_2

Supplementary Materials

file 1. Simulation results of other scenarios

The simulation results when $\pi = 0.2$ and $pri = 0.2$:

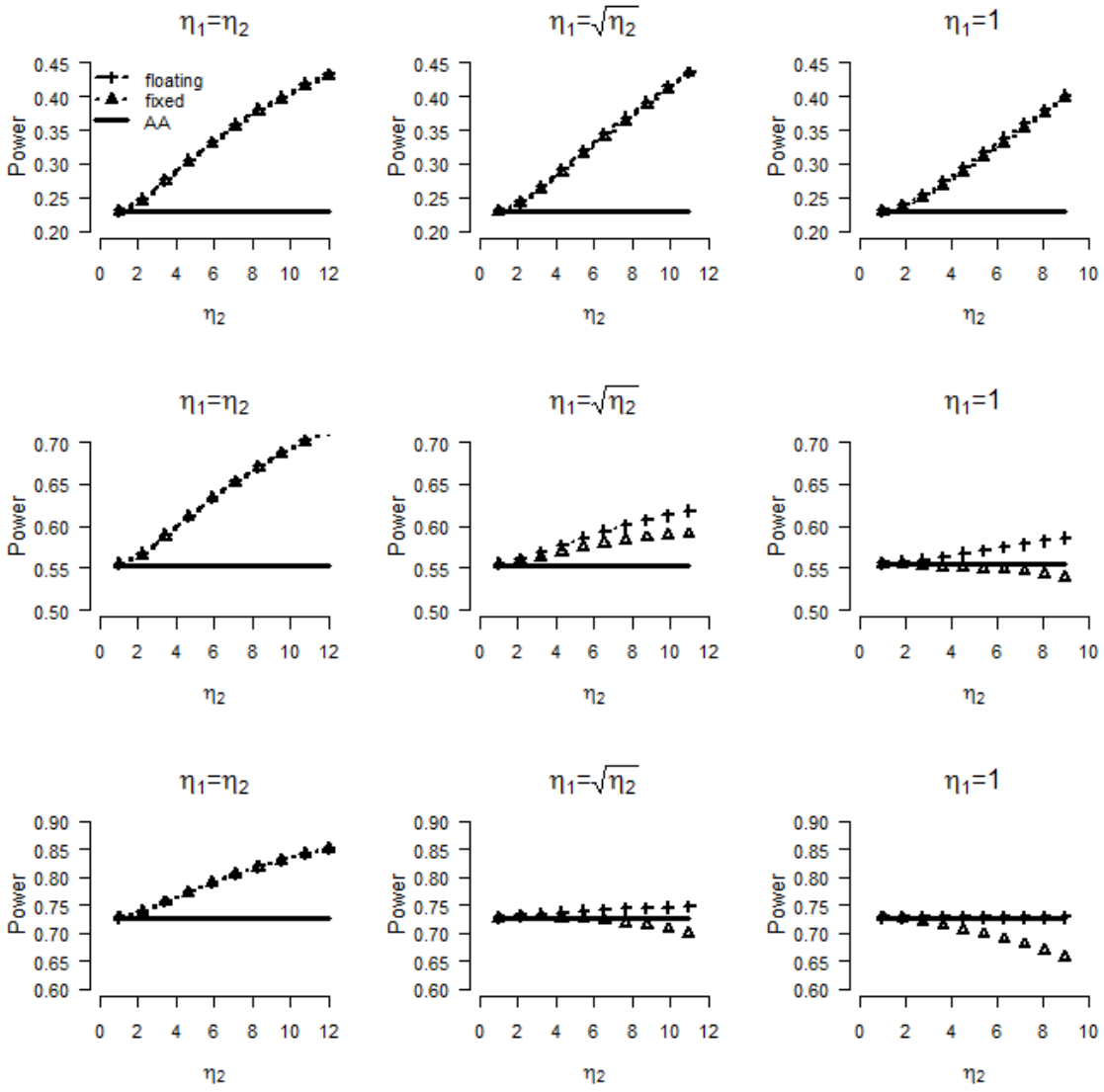


Figure 1. Power of the three methods

Power of the AA (solid lines), fixed PSA (- △ -), and floating PSA (- + -). The total sample size is set at 30, 60, and 100 (from top to bottom), respectively.

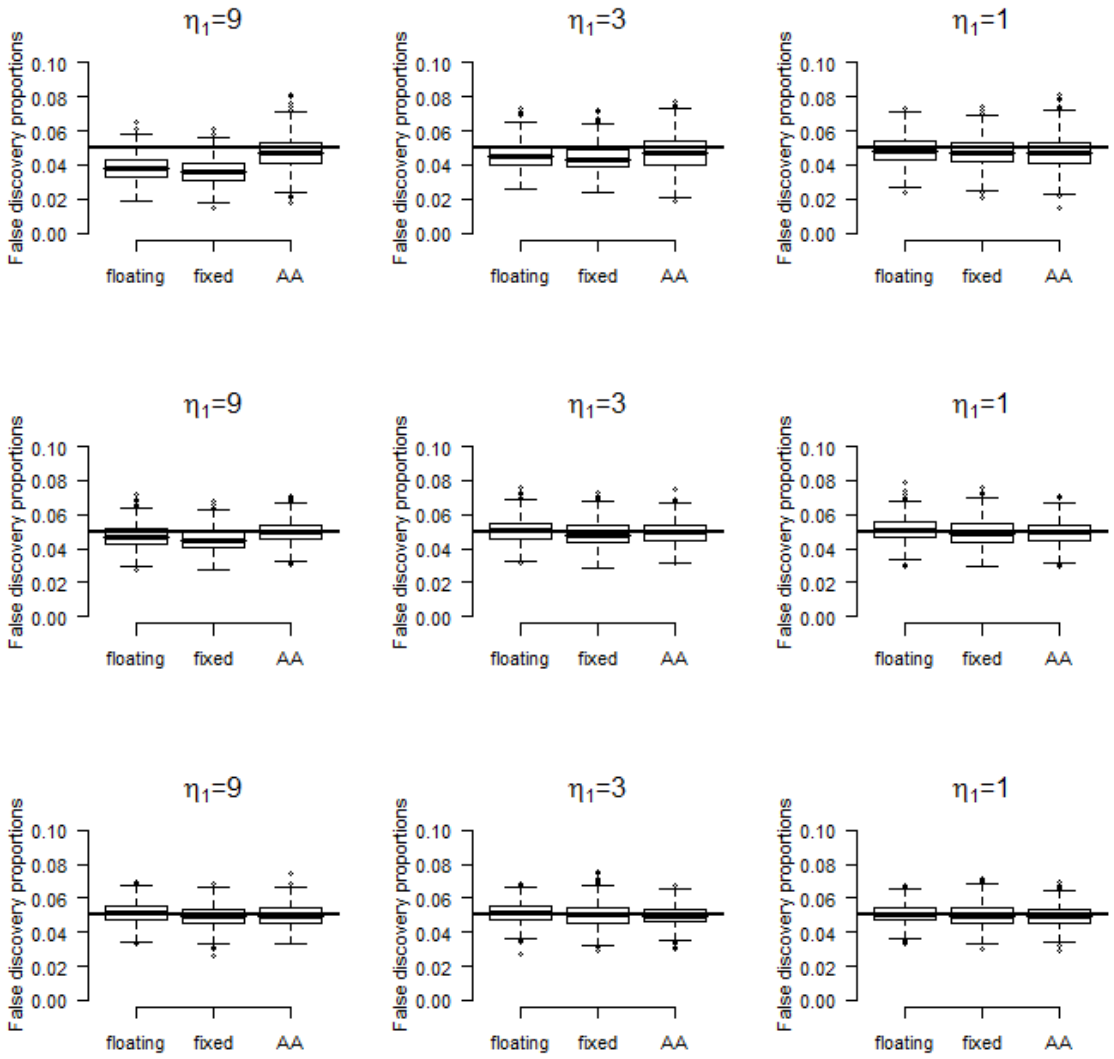


Figure 2. Boxplot of false discovery proportions in 1,000 repetitions for the three methods

The total sample size is set at 30, 60, and 100 (from top to bottom), respectively. The value of η_2 is fixed at 9 for all the scenarios.

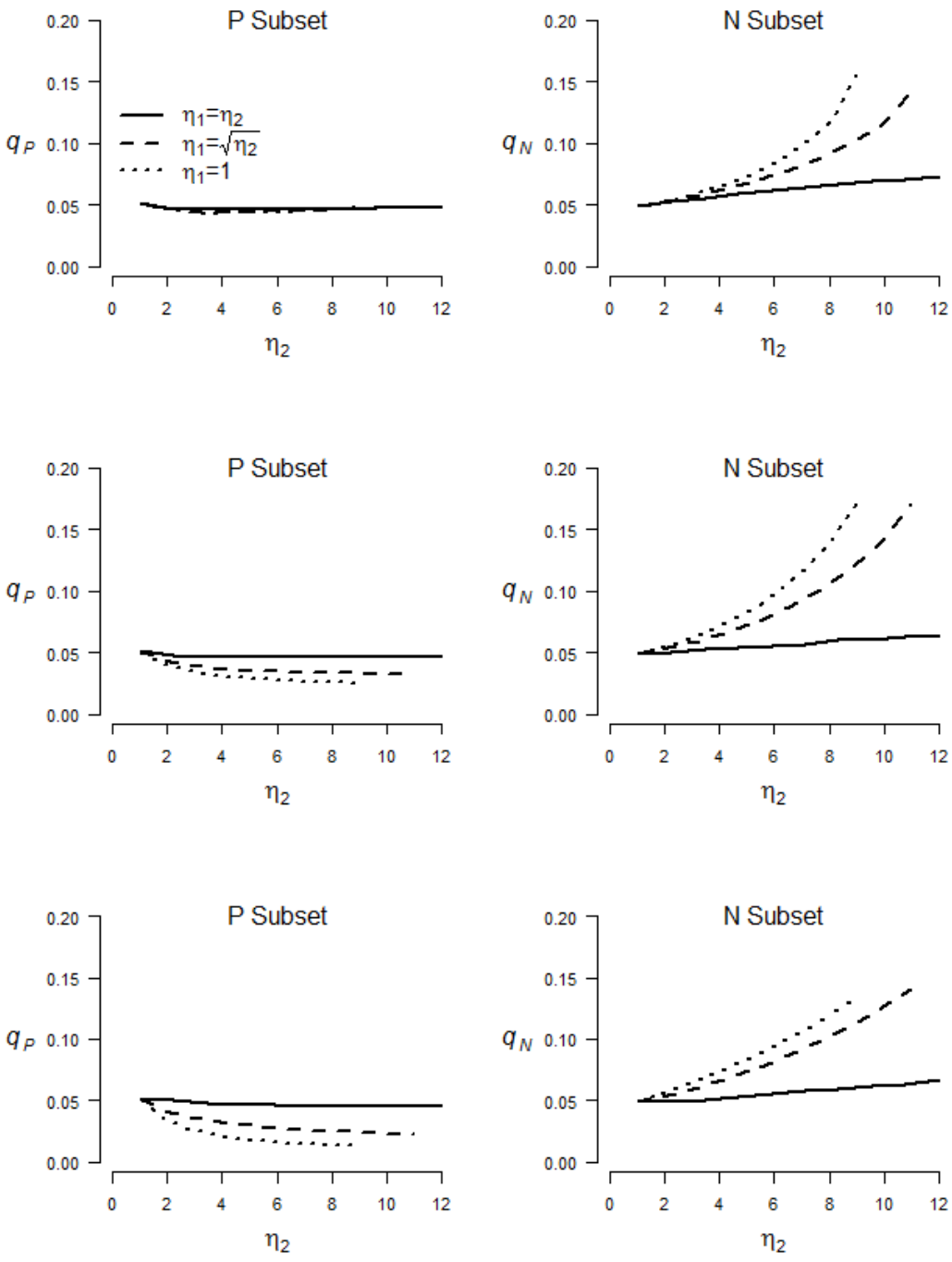


Figure 3. FDR levels in the floating PSA method

The FDR levels (q_P for the prioritized subset and q_N for the non-prioritized subset) in the floating PSA method (solid lines, $\eta_1 = \eta_2$; broken lines, $\eta_1 = \sqrt{\eta_2}$; dotted lines, $\eta_1 = 1$). The total sample size is set at 30, 60, and 100 (from top to bottom), respectively.

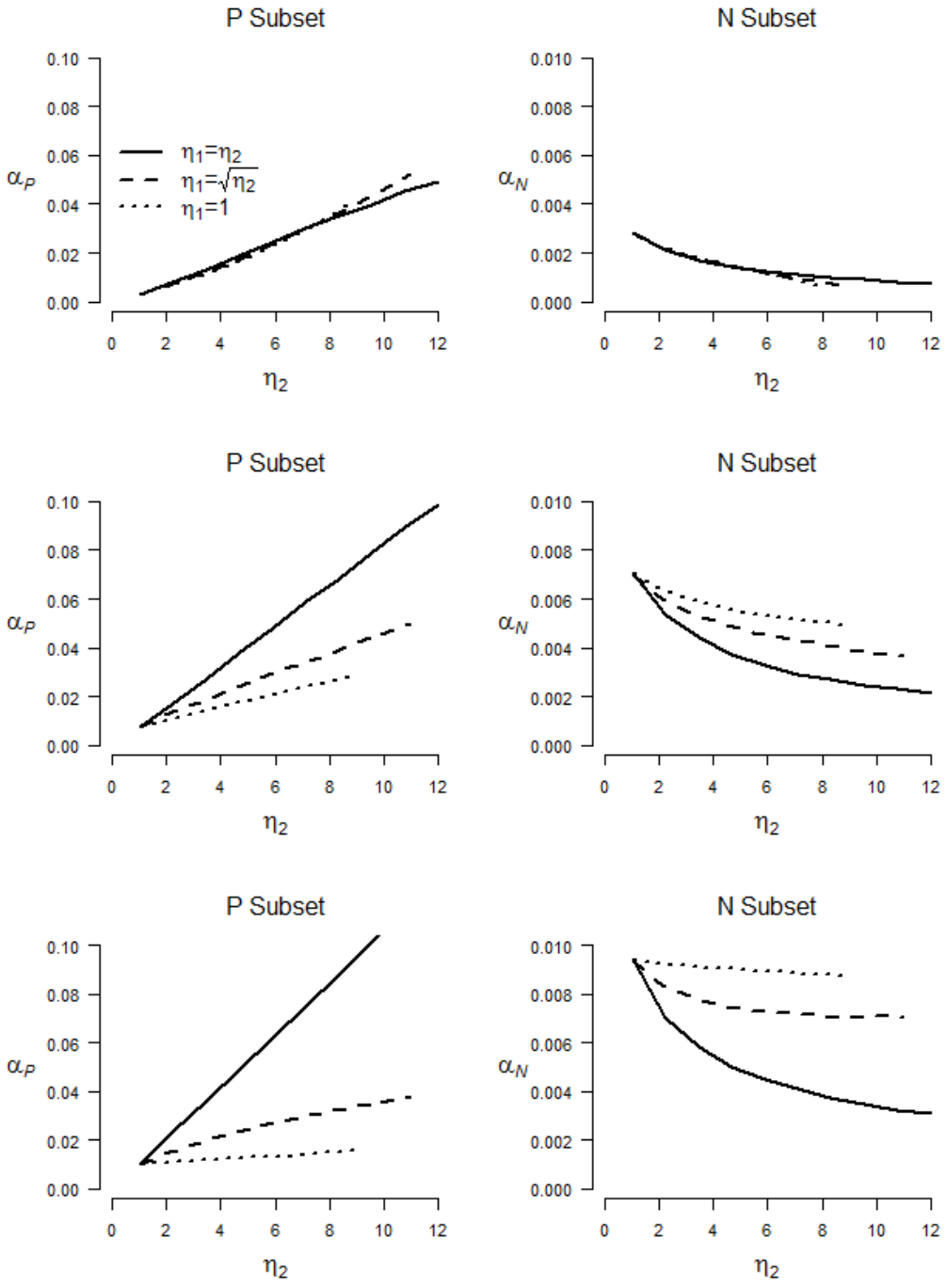


Figure 4. Significance levels in the floating PSA method

The significance levels (α_P for the prioritized subset and α_N for the non-prioritized subset) in the floating PSA method (solid lines, $\eta_1 = \eta_2$; broken lines, $\eta_1 = \sqrt{\eta_2}$; dotted lines, $\eta_1 = 1$). The total sample size is set at 30, 60, and 100 (from top to bottom), respectively.

When $\eta_2 \le 1$, $\pi = 0.1$, and $pri = 0.1$:

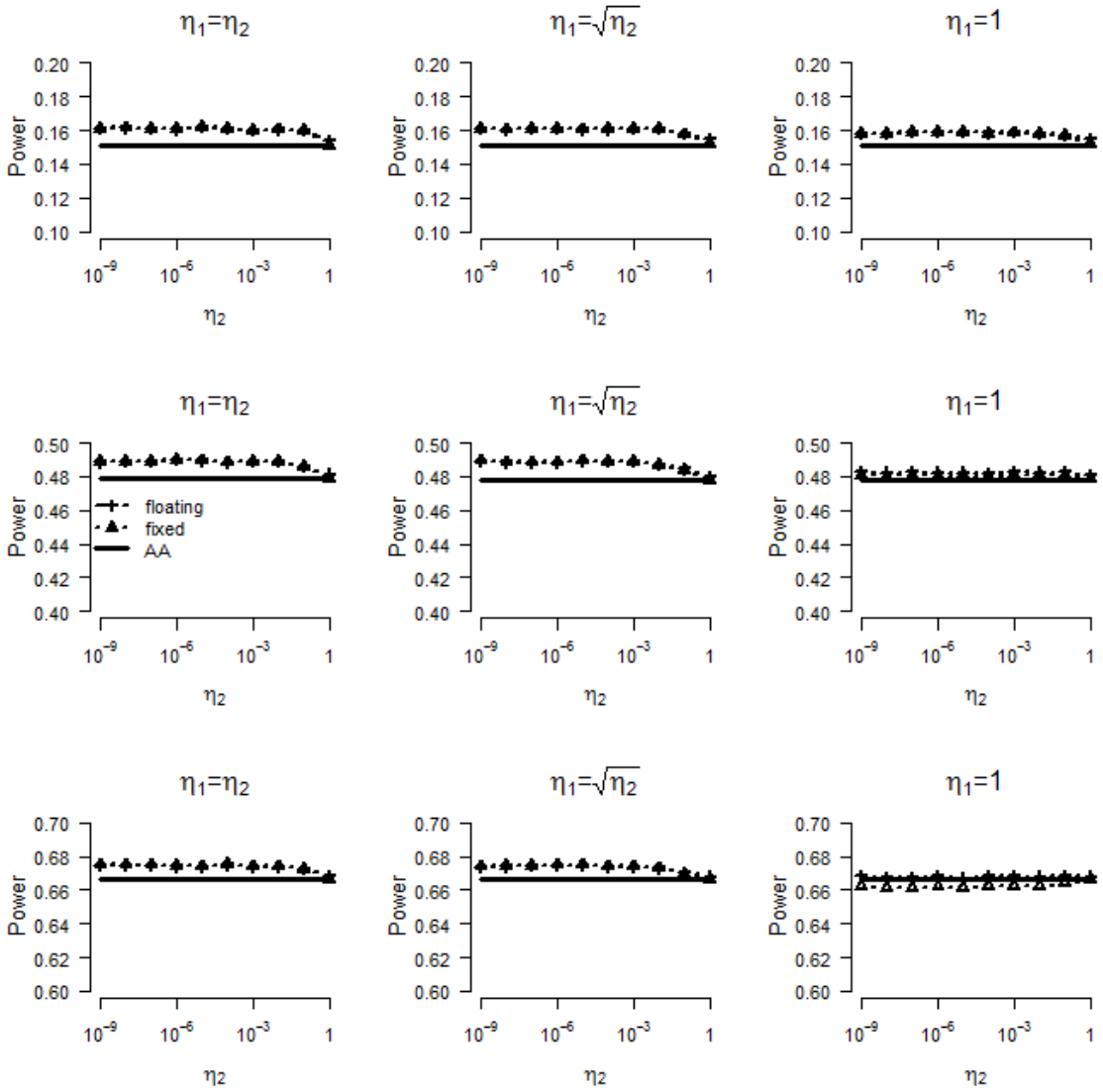


Figure 5. Power of the three methods

Power of the AA (solid lines), fixed PSA (- △ -), and floating PSA (- + -). The total sample size is set at 30, 60, and 100 (from top to bottom), respectively.

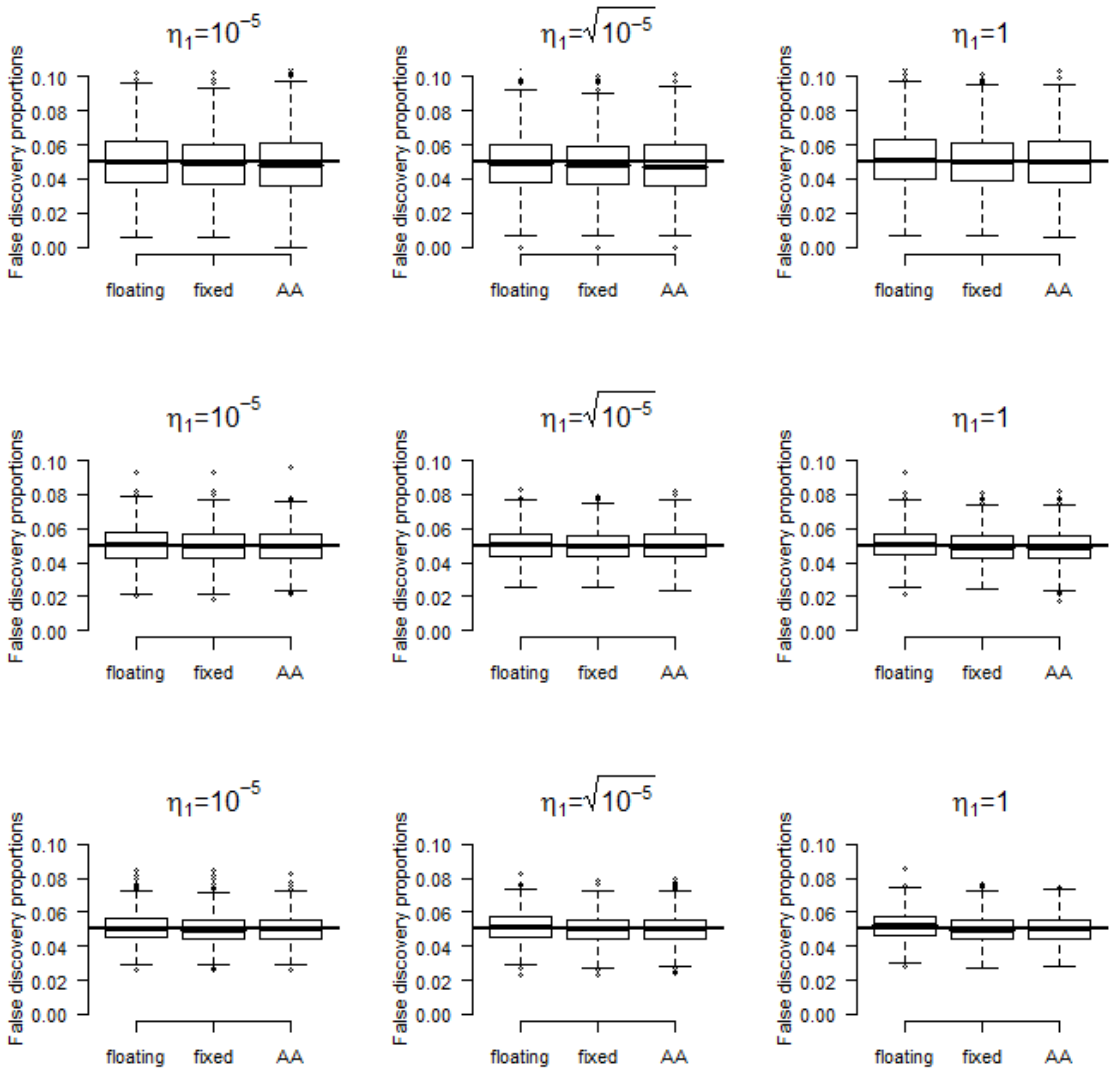


Figure 6. Boxplot of false discovery proportions in 1,000 repetitions for the three methods

The total sample size is set at 30, 60, and 100 (from top to bottom), respectively. The value of η_2 is fixed at 10^{-5} for all the scenarios.

When $\eta_2 \le 1$, $\pi = 0.2$, and $pri = 0.2$:

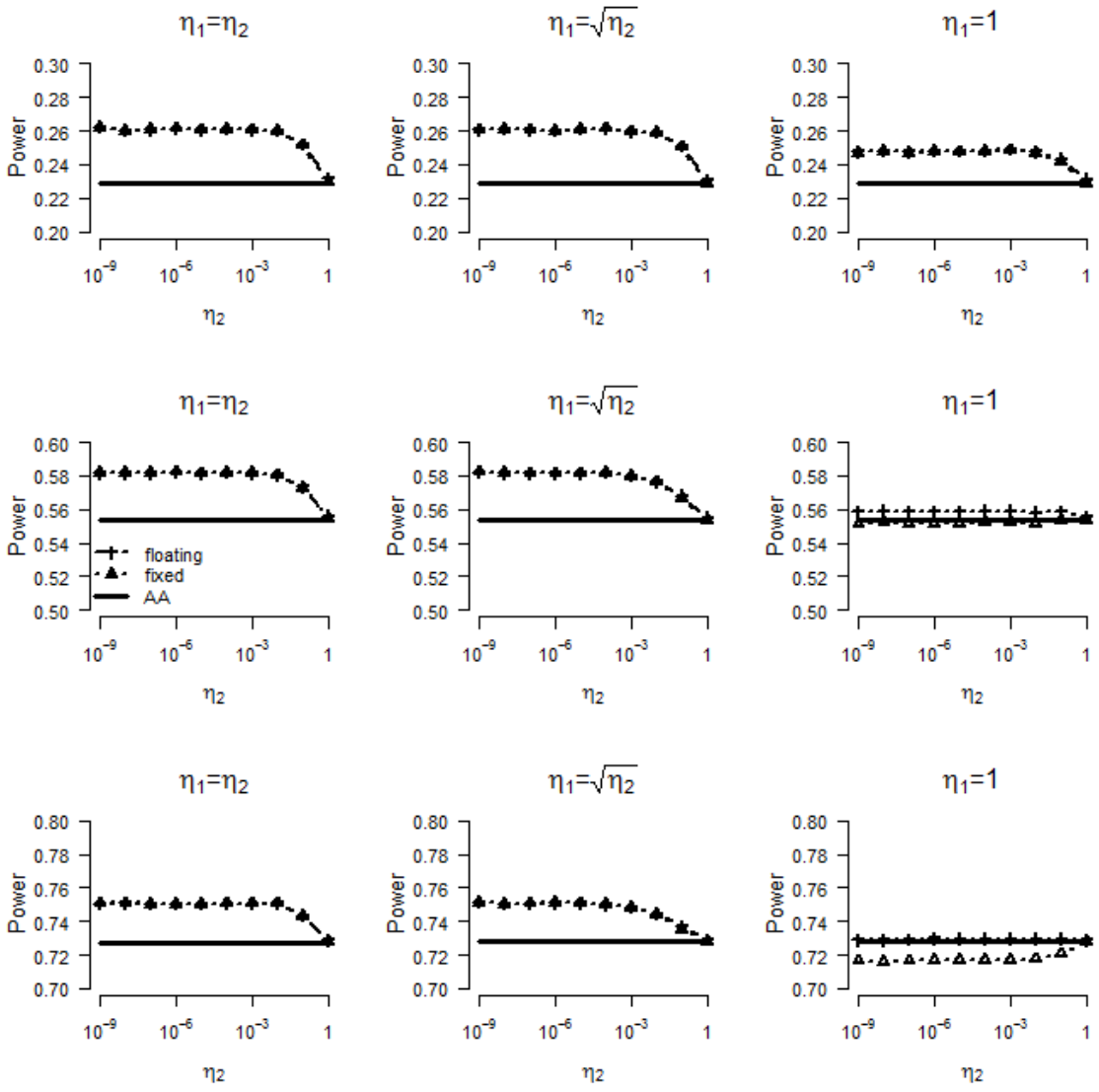


Figure 7. Power of the three methods

Power of the AA (solid lines), fixed PSA (- △ -), and floating PSA (- ＋ -). The total sample size is set at 30, 60, and 100 (from top to bottom), respectively.

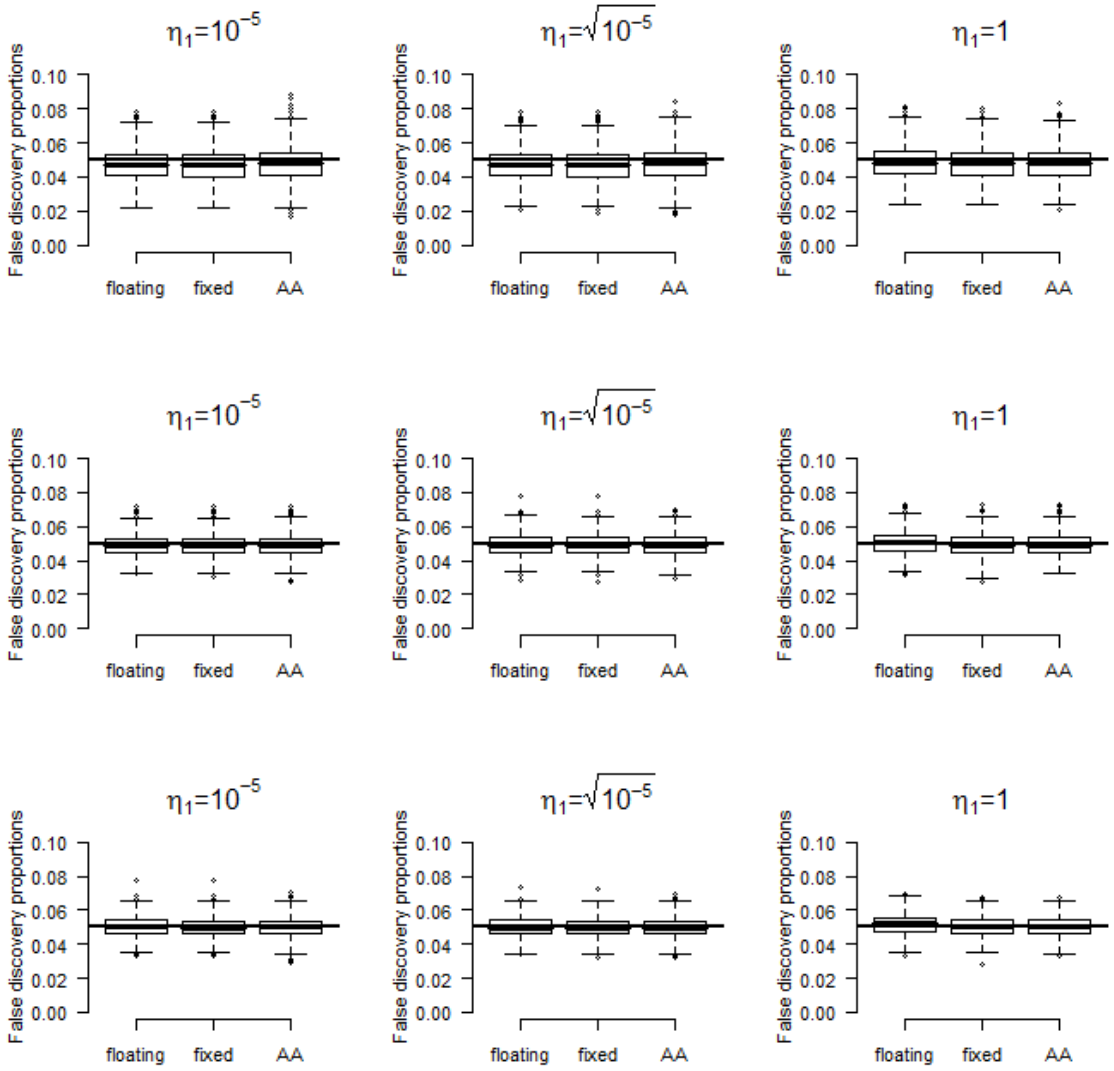


Figure 8. Boxplot of false discovery proportions in 1,000 repetitions for the three methods

The total sample size is set at 30, 60, and 100 (from top to bottom), respectively. The value of η_2 is fixed at 10^{-5} for all the scenarios.