

Qvalue Methods May not Always Control False Discovery Rate in Genomic Applications

Xiao Yang

Genomics Technology, Monsanto Company

St. Louis, MO 63167 USA

xiao.yang@monsanto.com

Abstract

The qvalue method by Storey (2002, 2003) has been proved to be theoretically sound for controlling false discovery rate in many high throughput genomic applications. However, empirical evidences suggest that this method can be more stringent than other methods, such as Bonferroni adjustment and the FDR method by Benjamini and Hochberg (1995). We compare these methods for detection of gene differential expression in microarray data analysis. For microarray experiment with the purpose of gene discovery, where many genes are expected to be differentially expressed across different experimental conditions, the qvalue method generally performs well. However, for experiments with only a few genes expected to be differentially expressed, the qvalue method performs much worse than other methods. Some insights are provided to examine this discrepancy. Adjustments to q-value method are recommended to accommodate many applications.

1. Introduction

Many high throughput genomic applications (e.g., microarray, genomewide scan for quantitative trait loci) have led to recent advancements in statistical multiple-test methods. Two types of adjustment methods have been used: methods for controlling familywise error rate (FWER), and methods for controlling false discovery rate (FDR). Methods controlling FWER include methods such as Bonferroni, step-up Bonferroni [1], step-down Bonferroni [2], and resampling based methods in [3]. Controlling FWER is useful for studies in which it is necessary to produce very few false positives, but they are not appealing for many genomics studies.

On the other hand, methods controlling FDR are intended to control the proportion of false positives among “significant” features, thus resulting in greater

chance of finding those truly significant features in a family of tests. Such methods include the FDR adjustment [4], and the qvalue method [5, 6]. Although both methods control FDR, they are quite different in the way FDR is controlled [6, 7]. Additionally, the qvalue method has been recommended over other methods.

In this paper, we compared the performance of qvalue method with other methods using microarray experiments. We discovered that for microarray experiments with many genes expected to be differentially regulated, the qvalue method typically worked well. However, for experiments with few genes significantly differentially regulated, the qvalue method did not perform well. We further examined the discrepancy between results from the qvalue method and those from other methods, and we recommended a modification of the qvalue method to accommodate many types of microarray experiments.

2. Comparing Qvalue Method with Other Methods

We compared results from 4 microarray experiments. The first two experiments (Figure 1, plot (1) and (2)) were intended to identify genes regulated by varying light conditions in plants, where many genes were expected to be significant. And the last two experiments (Figure 1, plot (3) and (4)) studied genes regulated by over-expressing two genes in plants, where only few genes were expected to be significant. For each family of tests, about 2,4000 tests were conducted individually, and density histograms were plotted in Figure 1, and different p-value adjustment methods were applied to each experiment. Table 1 summarizes the results from different methods for each experiment. For experiments (1) and (2), qvalue method identified as many significant as expected, while the qvalue method failed to pick up those

significant genes supposed to be regulated by the two transgenes for experiments (3) and (4).

Table 1. The table compares the number of significant features from 4 microarray experiments as shown in Figure 1. Abbreviations: Holm's for step-down Bonferroni, Hochberg's for step-up Bonferroni, and BH's FDR for the FDR method by Benjamini and Hochberg.

Methods	(1)	(2)	(3)	(4)
Qvalue	87	1219	0	0
BH's FDR	77	1183	2	13
Holm's	11	223	2	5
Hochberg's	11	223	2	5
Bonferroni	11	222	2	4

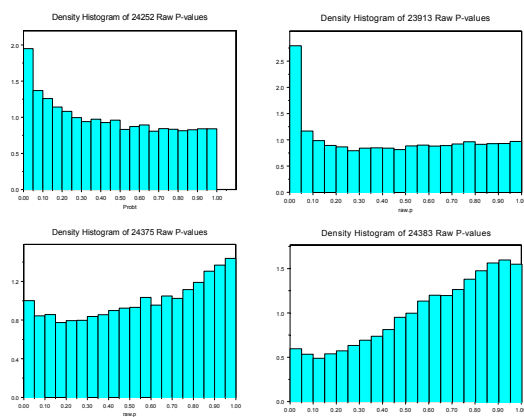


Figure 1. Density histograms of raw- p values from 4 microarray experiments. (1): top left, testing genes regulated by low light in maize leaf tissue; (2): top right, testing genes regulated by FarRed light in maize leaf tissue; (3): bottom left, testing genes regulated by a transgene; (4): bottom right, testing genes regulated by another transgene.

3. Discussions and Conclusions

What has caused the problem with the q value method? The q value method starts the adjustment with an estimation of expected proportion of null features among all tests, π_0 , which is based on the assumption that p values of null features are expected to be uniformly distributed in the range of $[0, 1]$. For microarray experiments with many significant features expected, this assumption seems valid from our experience, which can also be verified by the density histogram for the first two experiments. From this perspective, it is useful to use smoothing spline

technique [6] to estimate this parameter. However, for the last two transgenic microarray experiments (Figure 1), due to the irregular shape of the distribution of raw p values, smoothing spline estimation tends to over-estimate this parameter, thus resulting in no significant features for these experiments. Although not shown here, the case of under-estimation of π_0 can also happen.

To avoid over- or under- estimation of π_0 and to automate the process, we have implemented two alternative modifications to current q value method:

(1). Whenever the estimate of π_0 is greater than 1, it is automatically set to 1. As a result, adjusted p values agree with results from Hochberg's FDR method.

(2). Another modification would be to apply other robust methods to estimate π_0 without using smoothing spline technique. Instead, we compute a weighted average "proportion" from distribution of raw p values greater than a threshold (say 0.4), which gives a better control of estimation of π_0 , and the final estimate of π_0 is typically in the range of $(0, 1)$.

In summary, we have implemented these modifications in a SAS macro % p value, which has worked quite well for many microarray experiments.

4. References

- [1] Hochberg, Y., "A Sharper Bonferroni Procedure for Multiple Significance Testing", *Biometrika*, 75, 800-3, 1988.
- [2] Holm, S., "A Simple Sequentially Rejective Bonferroni Test Procedure", *Scandinavian Journal of Statistics*, 6, 65 – 70, 1979.
- [3] Westfall, P.H., S.S. Young, *Resampling-Based Multiple Testing: Examples and Methods for p -Value Adjustment*, John Wiley & Sons, New York, 1993.
- [4] Benjamini, Y., Y. Hochberg, "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing", *J. Royal Stat. Soc B*, 57, 289-300, 1995.
- [5] Storey, J.D., "A direct approach to false discovery rates", *J. Royal Stat. Soc. B* 64, 479-498, 2002.
- [6] Storey, J.D., "Statistical significance for genomewide studies", *Proc. Natl. Acad. Sci.*, 100: 9440-45, 2003.