Author's response to reviews

Title: Weighted Multiple Testing Procedures for Genomic Studies

Authors:

Jiang Gui (jiang.gui@dartmouth.edu)
Tor D Tosteson (Tor.Devin.Tosteson@dartmouth.edu)
Mark Borsuk (mark.borsuk@dartmouth.edu)

Version: 2 Date: 10 April 2012

Author's response to reviews: see over
March 26th, 2012

Dear Editor,

I would like to submit a revised manuscript by Gui et al. to be considered for publication. The title of the paper is “Weighted Multiple Testing Procedures for Genomic Studies”. We greatly appreciate the thoughtful reviews and have made changes to the paper where appropriate. As a result, we think the clarity of the paper has been greatly improved. Please find below our detailed responses to the reviewers. Thank you.

Sincerely,

Jiang Gui, Ph.D.
Assistant Professor of Community and Family Medicine
883 Rubin Building, HB 7927
One Medical Center Dr.
Dartmouth-Hitchcock Medical Center
Lebanon, NH 03756

Phone: 603-653-6083
Fax: 603-653-9093
Email: jiang.gui@dartmouth.edu
Responses to the Reviewers  
Manuscript 1483557767668926, Bio-Data Mining

Reviewer 1

This paper is articulate and well written, but as a review of multiplicity control for genetic studies this paper is incomplete.

Major revisions:
There are several methods that control the FWER by deriving the effective number of independent tests from correlated SNP data, and then using this quantity in either a Bonferroni or Sidak equation to discover the appropriate point-wise threshold for significance. These methods vary in formulation, and generally fall into the basic categories of PCA-based (Gao et al, genetic epidemiol, 2011; Nyholt et al, am j hum genet, 2004; Li and Ji, heredity, 2005) numerical integration-based (Conneely and Boehnke, Am J hum genet, 2007) and (Han et al, PloS Genet, 2009), and permutation testing. Also see (Dudbridge and Gusnanto, Genet Epidemiol, 2008), and (Hoggart et al., Genet Epidemiol, 2008), for discussions of this topic. The FDR-based methods described in this paper do not handle the correlated tests issue, and this is a major concern for most genetic epidemiologists.

This is very good point. We added a paragraph to the Discussion section to discuss the methods for adjusting multiple comparisons under dependency and potential application for weighted hypothesis testing.

Minor revisions
Figure I is not very useful, and could be dropped

We deleted Figure 1.

The plural of stratum is strata
“Equation (2) can be further simplified” on page 4, but I think it should say Equation (1)

We changed the sentence.

Reviewer 2

The enclosed article is a review of an important topic in genomic research – how to control for the concerns related to false positive findings when analyzing high-throughput data with a large number of statistical tests. The authors review several commonly used approaches, and do an excellent job describing the methods. This review would be an excellent resource for researchers that are new to the topic.

While this is an interesting review, I think there are a few topics that could be covered that would improve the impact and breadth of the paper.

First, they do a good job of introducing the methods, but do not include a detailed discussion of the consequences of issues related to multiple testing (from both a study design and biological point of view as well as a statistical one). Additionally, as the characteristics of different genomic experiments are very different in the expectation of the types/amount of associations that come through, the authors
should discuss these issues (expectations for microarrays vs. SNP-chips for example). This discussion would provide important guidance for readers in thinking about which methods to use for their data.

We intend for this paper to provide an introduction to the available methods for researchers who are already motivated to address issues of multiple testing. As the sources and consequences of these issues are highly dependent on the nature of the study design, we do not attempt to provide a comprehensive review of relevant application areas. However, we do agree with the point that the application may guide the choice of methods.

We have added a paragraph discussing the methods for adjusting multiple comparisons under dependency. This is an issue arising with both microarray and SNP-chip data.

Additionally, more discussion on how to choose a method as well as how to choose a cut-off (the consequences of false positive vs. false negative decisions, etc.) would be very helpful.

We have added a paragraph to the Discussion section, giving a more specific recommendation regarding choice of method. Typically, the False Discovery Rate (FDR) is fixed and then power is maximized subject to the FDR. For the purposes of consistency with the literature, we adopt this assumption here, rather than couch the problem in a broader decision theoretic framework.

Also, while the methods they choose to discuss are important ones, references to additional methods would make the paper more comprehensive.

We added additional references, in particular in response to Reviewer 1’s comments regarding dependence.

Finally, on a more minor note, I dont feel like the figure is very informative. How do they quantify these overlaps and bodies of work? What does the figure add to just a description? Maybe something like a timeline of development with overlap might be more informative. This would allow for the addition of references to less commonly used approaches to also be added.

We deleted Figure 1.

Reviewer 3

Major Compulsory Revisions:
This paper reviewed recent developments of weighting procedures in multiple testing content.

My main concern is that how this article can be useful for the general audience of BioDataMining. It did neither show the importance of using weighted multiple testing, nor give clear guidance on which method / software to use.

As mentioned in response to Reviewer 1, we intend for this paper to provide an introduction to the available methods for readers who are already motivated to address issues of multiple testing rather than attempting to convince the reader that multiple testing is important.

We added links to software for settings I and III. We did not find any existing software for setting II.
To improve this manuscript, I encourage the authors to emphasize the importance of weighting with more reference on their successful applications. It is still hard for the reader to implement those procedures without reading the original papers by Wasserman and Roeder (2006), Roeder et al. (2007) for Setting I; Roquain and van de Wiel (2009) for Setting II; Sun et al. (2006) for Setting III. Why do not you skip the detailed algorithms that have been already in those articles, and only give intuitive explanations and pros/cons for each Setting.

We added a discussion on pros/cons for each setting. In fact, successful previous applications are still rare given the relatively recent development of these methods. We believe that the main contribution is to provide a summary of existing methods for WFDR and grouping them into three distinct but closely related settings. The details are essential to explain these methods to general audience.

It will be also useful to address when weighting should be considered in different scenarios, e.g. different sample sizes, or number of tests, or even the violation of the Independence/normality assumption.

At the suggestion of Reviewer 1, we added a paragraph to discuss the methods for adjusting multiple comparisons under dependency. Recommendations for other scenarios are not easy to generalize.

Minor Essential Revisions:
\( r_k \) is not defined in Setting I

We define \( r_k \) in setting I now.