Author's response to reviews

Title: Identifying Significant Genetic Regulatory Networks in the Prostate Cancer from Microarray Data Based on Transcription Factor Analysis and Conditional Independence

Authors:

Hsiang-Yuan Yeh (d926708@oz.nthu.edu.tw)
Shih-Wu Cheng (tknl45@hotmail.com)
Yu-Chun Lin (g9565505@alumni.nthu.edu.tw)
Cheng-Yu Yeh (g9665535@oz.nthu.edu.tw)
Shih-Fang Lin (linfion07@hotmail.com)
Von-Wun Soo (soo@cs.nthu.edu.tw)

Version: 11 Date: 13 October 2009

Author's response to reviews: see over
Summary for Revision

Identifying Significant Genetic Regulatory Networks in the Prostate Cancer from Microarray Data Based on Transcription Factor Analysis and Conditional Independency

We appreciate the comments from the associate editor and all reviewers in this paper. We have made necessary experiments and modifications to cope with all the comments.

The basic notations for different fonts are:

- **Bold face** fonts are from reviewers’ original comments.
- *Italic face* fonts are the modified/added texts in the paper.
- Plain fonts are our answers to reviewers’ comments.

Answers to Reviewer #1

**Q1: Minor Essential Revisions.** The authors have carefully responded to my comments and make modification correspondingly. However, authors should also clearly express their point in the manuscript that Bonferroni correction is not necessary for initial independent tests in page 10. The example in page 10 is kind of misleading as it only includes one pair of genes and of course it's not necessary to perform multiple test correction. This is important as to understanding why they did not perform Bonferroni correction at the beginning of this study while they did in the conditional independency tests later, according to the newest version of this manuscript.

A1: Thanks for your comments for the detail information of Bonferroni correction. In page 10, it is one hypothesis testing for a transcript regulatory gene and the other gene in the microarray data. We use a pair of genes as an individual independent testing and we do not perform Bonferroni correction to reconstruct the networks. According to the large amount of genes and lots of permutations, the appropriate p-value calculated by Bonferroni correction is too small and conservative and there would be little power without massive sample sizes. For example, there are more than ten thousand of genes in the microarray and the number of the permutations may be over ten millions. After reconstructing the initial network, we use biological knowledge to filter out the results calculated by independent tests.

After we get the pair of interactions from transcription regulatory genes and their
co-expressed genes, we start to check the indirect edges should be needed and it may included several variables to be tested. For a given structure of the concept of d-separation see in Figure 4, we extract the minimum d-separate genes between start and end genes and we determine whether an edge between two nodes should be removed. Comparing with the whole nodes in the networks, it is a small set of nodes should be tested in conditional independent testing with Bonferroni correction.

Figure 6 Example of find_ d-separation set procedure

Take an example in Figure 6, we want to verify the direct link between node X and Y should be deleted or not in the sub-network. We use the procedure in Figure 5 to extract the minimum d-separate genes, node T and U, to help us determine whether an edge between two nodes should be removed. According to the small size of the d-separating genes, we do the tests in each of the predicted conditional independence relations from 2 tests of CI(X,Y|T) and CI(X,Y|U) and reject the null that both tests are independent with the p-value less than 0.05/2 = 0.025.

On P.11 line 1-4
We use a pair of genes as an individual independent testing and we do not perform Bonferroni correction to reconstruct the networks. Because the large amount of genes and lots of permutations, the appropriate p-value calculated by Bonferroni correction is too small and conservative.

On P.16 line 7-12
With the small size of the minimum d-separating sets, we can do the permutation comparisons by applying Bonferroni correction for multiple testing to renew the significant threshold for each of the n individual tests to maintain an experiment-wise error rate. Comparing with the whole nodes in the network, it is a small set of nodes should be tested in conditional independent testing with Bonferroni correction.
Take an example in Figure 6, we want to verify the direct link between node X and Y should be deleted or not in the sub-network. We use the procedure in Figure 5 to extract the minimum d-separate genes, node T and U, to help us determine whether an edge between two nodes should be removed. According to the small size of the d-separating genes, we do the tests in each of the two predicted conditional independence relations as $CI(X,Y|T)$ and $CI(X,Y|U)$ and reject the null hypothesis that both tests are independent with the p-value less than $0.05/2 = 0.025$. For different d-separated genes, we can get different significant p-value and so on.