Author's response to reviews

Title: Molecular prediction for atherogenic risks across different cell types of leukocytes

Authors:

Feng Cheng (fc8b@virginia.edu)
Ellen C Keeley (eck6v@virginia.EDU)
Jae K Lee (jaeklee@virginia.edu)

Version: 2  Date: 19 November 2011

Author's response to reviews: see over
November 17, 2011

Dear Editor of BMC Medical Genomics:

We greatly thank you for your consideration of our manuscript “Molecular prediction for atherogenic risks across different cell types of leukocytes” by Cheng et al. (MS#: 1733709600594895). We also appreciate reviewers’ insightful comments on our study which greatly helped to improve our manuscript. We carefully revised it fully reflecting all reviewers’ comments as described in our point-by-point responses below. We hope the revision is well suited for the publication in your high standard journal.

Thank you.

Sincerely,

Jae K. Lee, Ph.D.
Departments of Public Health Sciences
University of Virginia School of Medicine
1335 Hospital Drive, Rm. 3181, P.O.Box 800717
Charlottesville, VA 22908-0717
To Reviewer #1

This paper displayed impressive results. However, I am concerned about the following three issues:

1) When COXEN was applied, my understanding is that FH1 and FH2 were used as training data to identify genes that preserved highly concordant gene expression patterns between FH1 and FH2. Subsequently, these biomarkers were used back to analyze FH2. Please explain the validity of this method.

Response: Thanks much for raising an excellent point. It is correct that FH1 and FH2 were from the identical set of familial hyperglycemia patients and healthy controls and possible that our significant prediction on FH2 is correlated with the use of the identical patient set. However, we cautiously believe that our significant prediction on FH2 is not mainly due to the use of the same patient set for the following reasons. First, the molecular data of the two sets were from completely different immune cells—FH1 from monocytes and FH2 from circulating T cells. Since our biomarker discovery and predictive modeling were performed strictly based on monocyte cells of the FH1 set, FH2 is independent of FH1 for its molecular characteristics and data. Second, we observed that our identical prediction model performed considerably better for white blood samples on FH3, a completely independent set of FH patients and controls from the FH1 set. We think this was due to the fact that monocytes are partially included in white blood cells so our monocyte-based predictor presented a better predictability for that set. Therefore, we believe the common molecular information is more important than the use of a specific patient set for our training and prediction. We added these points in the Discussion of our revision.

2) FH2 and FH1 share the same sample. Are they still independent?

Response: Please refer to the response above.

3) In Figure 3, the authors are trying to show the clusters. But it seems that the results are not good in terms of the sample level. The cluster of sample is messy. Please revise it or explain it.

Response: In fact, in our original clustering heatmap analysis (Figure 3), several FH patients and healthy controls were clustered with opposite groups. Note that this kind of clustering analysis merely provides exploratory information and may not provide any
conclusive statistical results so we presented this heatmap analysis simply to show a classification potential of our biomarkers. Nevertheless, in this revision we attempted to use a slightly different clustering algorithm, the so-called McQuitty’s or WPGMA method, which is known to be robust with highly varying sizes of clusters by using the average distance between clusters weighted by uneven cluster sizes. This algorithm resulted in a much cleaner clustering heatmap showing all but two FH patients clustered into their respective groups now in our revised heatmap. Also note that we used a rigorous multivariate statistical classification modeling such as linear discriminant analysis to obtain our high-performing prediction model which could consistently provide significant stratification performance on all our independent test sets.

To Reviewer #2

The diagnosis of atherosclerosis is a very difficult problem in clinic because there is no obvious symptom during the development of atherosclerosis. There is an urgent need to address this issue. In the manuscript “Molecular prediction for atherogenic risks across different cell types of leukocytes” by Cheng et al., the authors proposed a new atherosclerosis diagnosis tool based on gene information.

1. The microarray platform of the test set 3 ATHERO1 is Affymetrix HG-U133A GeneChips which is different from the training set and other 2 sets (Affymetrix HG-U133 plus 2.0). How did authors deal with these different microarray platforms?

Response: We very much appreciate the positive and encouraging comments by the reviewer. Affymetrix HG-U133A is a part of Affymetrix HG-U133 plus 2.0 so we chose the common probe sets for our analysis in this paper. The identical probes in the two array platforms generally show consistent expression patterns and also our multi-gene predictor has an effect to average out the differences, mostly random noises. We added this point in “Material and Methods” of our revision.

2. This is a pioneering work in atherosclerosis diagnose using gene profiling. The authors should discuss more about the potential implication and problem of this approach in diagnosis of atherosclerosis.

Response: We completely agree with the reviewer’s suggestion. We are also currently collecting more molecular data from patients with atherosclerosis at our institution for our continuous study. We further emphasized this critical need in the Discussion section in this revision.