Author's response to reviews

Title: Gene expression profiling in whole blood identifies distinct biological pathways associated with obesity

Authors:

Sujoy Ghosh (sghosh@nccu.edu)
Robert Dent (bdent@ottawahospital.on.ca)
Mary-Ellen Harper (maryellen.harper@uottawa.ca)
Shelby A Gorman (shelby.a.gorman@gsk.com)
Joan S Stuart (joan.s.stuart@gsk.com)
Ruth McPherson (rmcpherson@ottawaheart.ca)

Version: 2 Date: 12 October 2010

Author's response to reviews: see over
October 7, 2010

Dear Dr. Demichelis,

We wish to re-submit manuscript # 1774792093408751 by Ghosh, et al., entitled ‘Gene expression profiling in whole blood identifies distinct biological pathways associated with obesity’ for reconsideration for publication in BMC Medical Genomics. We thank the reviewers for their insightful critique of our work and for many helpful suggestions. We have revised the manuscript based on the recommendations and have highlighted the areas of revision using magenta colored text in the body of the manuscript. Below are our point-by-point responses to the reviewers’ comments. We look forward to a favorable re-consideration of this work.

Sincerely,

Ruth McPherson, MD, PhD
Professor, Department of Medicine
Director, Atherogenomics Laboratory
University of Ottawa Heart Institute
40 Ruskin St- H4203
Ottawa, ON, Canada K1Y 4W7
rmcpherson@ottawaheart.ca

Responses to reviewer #1 (Camargo):
-Major Compulsory Revisions

1. Table 1 shows the characteristics of the subjects included in the current study (20 obese and 20 lean), but the data from 3 samples from each group were not used due to poor hybridization. Table should show the characteristics of the subjects included in the data analysis (17 obese and 17 lean), where the results come from.

We have made the recommended changes and the corrected Table 1 with 17 lean and 17 obese subjects is now provided. The changes have also been noted under “Phenotypic characterization of study subjects” on page 5.

2. The study was carried out in men and women. A new analysis showing significant changes for each sex would be worth and would provide an accurate analysis regarding gender dimorphism.

We have carried out gene-set enrichment analysis separately for males and females and the results have been added to Additional File 7. Comparison of the pathway ranks among ALL, FEMALES and MALES only analyses show that ‘apoptosis’ (ranked 7th, 8th and 3rd for ALL,
FEMALES and MALES respectively) and ‘oxidative phosphorylation’ (ranked 10th, 12th and 18th for ALL, FEMALES and MALES respectively) pathways remain in the top scoring pathways for all comparisons suggestive of concordant changes in both genders. The ‘ribosome’ pathway was the #1 ranked pathway when comparing obese to lean for ALL and FEMALES only analysis but was ranked #27th in the MALES, suggesting that this pathway may be preferentially altered only in females. Due to smaller sample numbers in the subgroups, the statistical measures (nominal p-values and false discovery rates) were uniformly poorer in the subgroup analyses compared to the full group (males and females combined). These results have been included on Page 10.

3. The authors have confirmed gene expression changes by RT-PCR. What is the rational for choosing genes? Are they representative of several levels of gene expression/pathways?

As correctly assumed by the reviewer, the genes chosen cover a range of approximately 7 logs (base 2) representing over 100-fold differences in the magnitude of gene expression. This information is now added on page 5 of the manuscript under Ascertainment of Data Quality.

4. Results are not clearly described. Results section is actually a Results and Discussion section where the results are discussed in a long and speculative way. Further, Discussion section (the second discussion) focuses in the advantages of the use of whole blood instead to others strategies such as isolation of PBMCs. It should be taking into consideration that, as the authors comment, “compared to whole blood, several cell types including neutrophils, basophils, eosinophils, platelets, reticulocytes and erythrocytes are depleted in PBMCs which lead to loss of important transcription information”, the use of a very heterogeneous cell population could introduce noise and could mask gene expression differences in specific cells. Different amount of each one of the types cell between obese and lean could be also responsible of the different gene expression.

We have further clarified the limitations of the PAXGENE system, as aptly pointed out by the reviewer, under “Discussion” on page 14.

5. The manuscript would improve if the main issues are clearly splitted along the text, the gene expression changes found between groups (obese and lean) and the advantages and desvantages of the use of whole blood versus cell types isolation.

Based on the reviewer’s suggestion, we have now clearly separated the Discussion and Results sections.

6. Although authors comment that “transcriptional differences are causal or caused”, sometimes along the manuscript is confused the relationship between gene expression and genetic variation. It should be clarified the fact that genetic variation could affect the expression of the genes, but gene expression can be modified because obesity independently of the genetic background.

Page 2. Abstract: Conclusion: “This study represents a novel approach for the elucidation of the genetic bases of obesity”

Page 4. “Second, as differences in gene expression are often driven by sequence variants in gene regulatory regions, our study provides a mechanism for the selection of obesity-associated candidate genes for the determination of possible regulatory sequence variants”.
To clarify the differences further, as suggested by the reviewer, we have now added lines elaborating the differences between genetic and environmental influences on gene expression under “Background” on page 4.

Page 10: “Accordingly, we examined whether biological pathways implicated from gene-set enrichment analysis of the current study could provide a set of mechanism-based gene predictors that would be capable of predicting obese and lean subjects with high accuracy”. What predicting value have differentially expressed genes because the obesity? Would be these genes differentially expressed in prone people to develop obesity?

Since the subjects examined were already obese, we do not know whether the differentially expressed genes are the cause or consequence of obesity. Our motivation for the predictive analysis reported in the paper was to generate proof-of-concept data to test if blood gene expression patterns are indeed useful for the classification of metabolic phenotypes (obese and lean in this case). Success in this endeavor would provide the preliminary evidence required for attempting to develop blood gene expression based classifiers for more relevant phenotypes such as predicting the weight-loss success of an individual prior to dietary intervention or bariatric surgery. We have now added this discussion to the manuscript under “Class prediction via blood gene expression” on page 10.

7. In order to clarify the differential gene expression in the 11 predictor genes associated with the phenotype, it would be worth look for sequence variants in these gene regulatory regions (e.g. promoter regions).

We have interrogated the NCBI dbSNP (Build 131) database to identify common sequence variants in a 2000 base region upstream of the start site for the 11 classifier genes. This approach can help us select candidate genes for discovering and screening regulatory sequence variants that control transcription in the obese and lean subjects and can lead to the identification of expression QTLs (eQTLs). This possibility has been added under Discussion on page 16.

- Minor Essential Revisions
1. In results section, it is not mentioned the number of transcripts detected in whole blood, and how many are up-regulated and down-regulated.

This information is now added under “Identification of differentially expressed genes” on page 6

2. Quality of the RNA samples was not shown, have the authors assessed the RNA quality before the microarrays performing? Which method did they use?

We used Agilent Bioanalyzer 2100 to assess RNA integrity prior to microarray processing. We have included this information now under “Sample preparation for transcriptome analysis” on page 17.
Responses to reviewer #2 (Koza):
Major compulsory Revisions:
1. The interpretation of the gene expression data is questionable because the contribution of RNA from the red blood cell component and remaining cellular components of whole blood appears to differ between obese and lean individuals. This is indicated by the author's own observation that the majority of genes that were upregulated in the obese subjects are most highly expressed in erythrocytes and reticulocytes. Several references to increased hemocrit in obese individuals are noted in the literature as well (i.e., Wysocki M et al Atherosclerosis 1991, 88(1):21-28). Although the reasons noted in the manuscript as to why PAXgene whole blood RNA was used instead of fractionated blood components are reasonable, is there a way to adjust gene expression data for hemocrit or for erythrocyte cell specific mRNA expression?

This is an excellent suggestion. We have carried out the analysis as reported below, but would like to make two comments here. First, since the same amount of cRNA was used from each sample for hybridization, the relative enrichment of cell types will only have a real effect on gene expression for genes that are differentially expressed among the cell types to begin with (e.g. hemoglobin transcripts which are expressed only in reticulocytes and not lymphocytes). For genes expressed at similar levels in all cell types, the differential cell type representation should not have an effect on expression – the only factor changing gene expression there is a true upregulation or downregulation of genes between the two populations (obese and lean), although the cellular origin(s) for the differential expression cannot be known. Scaling the gene expression data by the expression of a reticulocyte/erythrocyte specific gene will however not distinguish between the above two mechanisms of enhanced gene expression in the obese subjects and can potentially lead to incorrect conclusions. We scaled the gene expression data independently by the expression of 2 erythrocyte-specific transcripts, hemoglobin D (HBD) and erythrocyte membrane protein, band 2 (EMPB2) and subjected the scaled data to gene-set enrichment analysis. Of the original 3 pathways found to be differentially upregulated in the obese subjects, the “ribosome” pathway remained the top differentially expressed pathway (with the scaled data) whereas the “apoptosis” and “oxidative phosphorylation” pathways were no longer significantly enriched, with either of the scaled datasets. This suggests that the increase in erythrocyte/reticulocyte numbers in the obese is a possible explanatory mechanism for the observed increase in transcript levels for “apoptosis” and “oxidative phosphorylation” in the obese subjects. The results for the “ribosome” pathway suggest significant upregulation of the transcripts for the component genes of this pathway in the obese subjects.

These analyses using scaled data have been included as a new Additional File 13. The new results have also been incorporated in page 9 and page 14-15 of the manuscript. Additionally, we scaled the FTO expression data by expression values of two leukocyte specific genes, CCL5 and CCR7. In both cases, the expression levels of FTO were not statistically significantly different between the obese and lean subjects after scaling. Based on this finding, the interpretation of whole-blood FTO gene expression in obesity becomes complex and speculative. As such, we have removed the FTO analysis results from the revised manuscript.

2. Why is there a difference in the number of individuals used for phenotypic characterization of subjects (i.e 12 per group; results, page 5 line 8/9 and Table 1) different from the number of individuals analyzed via Affymetrix microarray (i.e 17 per group; page 6 line 1; Figure 3)? The number 12 refers to only the number of women; the number 17 refers to all subjects (women and men).
3. The data in Figure 1 are confusing. The X-axis indicates log2 ratio differences between the obese and lean cohorts; however, the bars are stacked and are no longer associated with the X-axis in a meaningful way. Would it be possible to re-design this figure to show a more direct correlation between the TaqMan and microarray analysis. The data are now plotted as side-by-side bars (instead of stacked bars) to compare the Affymetrix and Taqman results in a more meaningful way.