Author's response to reviews

Title: Tobacco use induces anti-apoptotic, proliferative patterns of gene expression in circulating leukocytes of Caucasian males

Authors:

Peter C Charles (pcharles@med.unc.edu)
Brian D Alder (bdalder@gmail.com)
Eleanor G Hilliard (eleanorh@med.unc.edu)
Jonathan C Schisler (jonathan_schisler@med.unc.edu)
Robert E Lineberger (rob_lineberger@med.unc.edu)
Sabeen Mapara (sabeen.mapara@gmail.com)
Joel S Parker (jparker@expressionanalysis.com)
Samuel S Wu (ssw@med.unc.edu)
Andrea Portbury (andrea_portbury@med.unc.edu)
Cam Patterson (cpatters@med.unc.edu)
George A Stouffer (rick_stouffer@med.unc.edu)

Version: 2 Date: 1 May 2008

Author's response to reviews: see over
Thursday, May 1st, 2008

Melissa Norton, M.D.
Editor-in-Chief
BioMed Central – Medical Genomics
BioMed Central Ltd
Middlesex House
34-42 Cleveland Street
London, W1T 4LB, UK

Dear Dr. Norton:

On behalf of all co-authors, I would like to submit to BioMed Central Medical Genomics our revised manuscript, "Tobacco use induces anti-apoptotic, proliferative patterns of gene expression in circulating leukocytes of Caucasian males", for consideration as a research article.

We are very grateful to the reviewers for their insightful comments, all of which we have incorporated into our revised manuscript. We believe these changes have significantly improved this article, which we hope is now in acceptable format for publication. A comprehensive discussion of all the suggestions made by the reviewers follows this letter.

In addition to the revisions that were made at the request of the reviewers, we have made the following changes that we believe strengthen our manuscript considerably:

- Figures 2 and 3 have been transposed and data has been added to Figure 3 to demonstrate both the technical and biological verification of our results.
- A new table, Table III, has been added to illustrate the top three gene sets identified by GSA.
- A supplemental table has been added as Additional file 2 and lists the 109 genes identified by SAM and GSA.

All authors have read and approved submission of the manuscript, and material in the manuscript has not been published elsewhere and is not being considered for publication. Permissions have been obtained for all citations, acknowledgments, and communications.

My corresponding address is:

Cam Patterson, M.D.
Carolina Cardiovascular Biology Center
8200 Medical Biomolecular Research Bldg.
103 Mason Farm Road
Chapel Hill, NC 27599-7126
Phone: (919) 843-6477
Fax: (919) 843-4585
E-mail: cpatters@med.unc.edu
Thank you for your time and consideration.

Sincerely yours,

Cam Patterson, MD
Enclosure
Response to Reviewers

Reviewer #1, Travis Dunkley:

Major Revision item #1: “It is unclear if the 40 samples used for the qRT-PCR validation shown in figure 2 were included among the samples that were expression profiled or if these represent completely independent samples”

Thank you for pointing out the importance of biological replication – we agree with your point. Initially, all of the samples used in the qRT-PCR validation study were taken from the same pool of subjects used for expression profile analysis simply to validate the microarray results. However, we have now addressed your concern for verifying our finding in two ways. First, we have expanded the number of genes analyzed by qRT-PCR from the original set of samples (Caucasian males used in the microarray analysis) to demonstrate the robustness of our microarray technique and refer to this analysis as technical verification. Furthermore, we selected another cohort of biological replicates (Caucasian females) to confirm the cotinine-dependent effects on two of these genes and refer to this as biological verification. These changes are reflected in Figure 3 and well as in the results section. We hope the Reviewer finds that this approach adds scientific merit to the study.

Reviewer #2, Joost van Deft:

Major Revision item #1: “In light of these discrepancies, the authors must exclude these patients [patients with high cotinine levels who report to be non-smokers and patients with low cotinine levels who report to be smokers] from their analyses…Thus the authors MUST restrict their groups of patients to only smokers and only non-smokers”

The Reviewer raises an excellent point that inclusion of the discordant subjects in our analyses could confound our results and conclusions. After considering this issue in depth, we realized that we needed to clarify in our manuscript by what exact criteria we were defining our subject groups. The 2 major subject groups in our study are defined as “high cotinine” and “low cotinine” based on whether or not their plasma cotinine levels are above 100 ng/ml or below 50 ng/ml respectively. Although we also asked the subjects to define themselves as either “smokers” or “non-smokers”, their answer was not used in determining to which subject group they would be assigned. Therefore, even though the 7 subjects who had cotinine levels over 100 ng/ml reported themselves as non-smokers, they were included in the “high cotinine” group. Likewise, the 3 subjects who had cotinine levels below 50 ng/ml but who reported themselves as “smokers” were included in the “low cotinine” group. We have amended the manuscript (page 10, paragraph 1) to reflect this and hope that our clarification satisfies the reviewer as to why these subjects were retained in the overall analysis.

Although we hope that the Reviewer will agree that defining our subject groups by plasma cotinine levels does remove the risk of confounding our results by including discordant subjects, we wanted to stress that we thought this issue was significant and
worthy of further investigation. Therefore, we performed a simple correlation between the results obtained with and without inclusion of these subjects in the sample pool (Reviewer’s Figure 1). Log ratios were calculated for low versus high cotinine using all samples (including those whose smoking history did not match plasma cotinine levels), and for the subset that only contained samples in which smoking history and plasma cotinine levels matched. Genes with log ratios near zero were more variable, but there was still good agreement as the log ratios deviate from zero. The correlation coefficient between the data set with discordant subjects and the data set without discordant subjects is 0.86.

Reviewer Figure 1. Scatter plot demonstrating the relationship between the log ratio estimates from overlapping cohorts. The agreement is quantified using Pearson’s correlation ($r = 0.86$).
Minor Revision item # 1: “The abstract mentions prominently an effect of smoking on cell cycling, but this is hardly discussed in the paper”.

We thank the Reviewer for this observation. We agree that cell cycling is not a prominent factor in our manuscript and so we have removed mention of it from the Abstract (page 2, paragraph 3).

Minor Revision item # 2: “Relevant studies on smoking and gene expression in humans are missing”.

Thank you for bringing these studies to our attention. They are indeed relevant and were missing thus we have included these references in our manuscript (page 4, paragraph 1; page 15, paragraph 2).

Minor Revision item # 3: In the paper the authors hardly compare their data to those of others (microarray studies or rt-PCR based studies). They should at least mention similarities and differences”.

Thank you for making this observation. We have added in a discussion comparing our findings with those published in a study by van Leeuwen et al., 2007 (page 15, paragraph 2). We have also amended the manuscript by adding more references to related studies in the introduction (page 4, paragraph 1).

Minor Revision # 4: “Especially the absence of CYP1B1 [in genes retrieved in this study] is very surprising, as this is the prototypical gene to be affected by cigarette smoke. The authors must clarify this.”

The Reviewer makes a very important observation that we have now addressed in the conclusions section of our manuscript (page 15, paragraph 2). We agree with the Reviewer that some historically “smoking-related” genes appear to be absent from our genetic profile, and adding this discussion of CYP1B1 and SOD2 to our study will enhance the readers confidence in our results.

Reviewer #3, Avrum Spira:

Major Revision # 1a:…”[RNA and DNA isolated from whole blood in which RBC had been lysed] would imply that they are studying gene expression in all leukocytes, not simply the peripheral mononuclear cell portion as suggested by the title of their manuscript”.

This is an excellent point and we appreciate the Reviewer mentioning it. We have revised the title of our manuscript to reflect the nature of the cells actually included in this study. In addition, throughout the manuscript we have removed the phrase “peripheral mononuclear cell” and replaced it with “circulating leukocyte”.
Major Revision # 1b: “Were complete blood counts with WBC differentials performed on these subjects to address this issue?”

Thank you for making this point. We have now included this data in Table I and mention it in the Materials and Methods section (page 5, paragraph 1). In comparing the relative numbers of various white blood cell subtypes, we found no significant differences between high cotinine and low-cotinine subjects. Therefore, we believe that the genetic profile differences we are reporting in this study are not due to a change in WBC proportions between the 2 subject groups. We feel inclusion of the WBC counts in our demographics chart (Table 1 in the manuscript) adds to the confidence our study results and thank the reviewer for this suggestion.

Major Revision # 2a: “..is it not likely that this interferon gene set simply reflects a change in the proportion of various WBC cell types in setting of pulmonary or systemic disease?”

We agree with the Reviewer that this is indeed a possible explanation for our results. Therefore, we compared the WBC counts in our subject demographics chart and found no significant difference in WBC subtypes (Table 1). This is consistent with our conclusions that the genetic profile differences we report between high cotinine versus low cotinine subjects are not due to differences in WBC subtype numbers or other co-morbidities. We further explore other co-morbidities relating to this issue below.

Major Revision # 2b: “What about other co-morbidities that might relate to smoking including COPD and lung cancer?”

Thank you for asking this highly relevant question. We agree with the Reviewer that taking into account co-morbidities is an important issue in the interpretation of the data we are presenting in this manuscript. In order to address this issue, we performed a principal components analysis (PCA) of the genes identified in this paper using the combined significant gene list and visualized in the context of COPD, diabetes, CAD class and cotinine levels. We have added this data as a Supplementary figure (Additional file 1) and have made reference to it in the manuscript (page 10, paragraph 1). As expected, the top component of variation appears associated with smoking status. Additionally, it does not appear associated with the remaining variables. To formally test this hypothesis, the PC1 loadings were tested for association with each of the four clinical variables. Smoking status was found to be significantly associated with PC1 (p < 0.001). However, none of the remaining clinical variables were associated with the top component of variation (COPD p = 0.91; CAD p = 0.15; Diabetes p = 0.55) indicating that the primary axis of variation for these genes is not confounded with COPD and other disease states. Other co-morbidities, and cancer treatment, were cause for exclusion from the study (page 5, paragraph 1).

Major Revision # 3: “These findings [Figure 3] could be strengthened if the authors could build a gene expression classifier using a training set of samples and then correctly
predict the class of test samples….Additionally, it would be informative to see where subjects who call between cotinine levels of 50 – 100 ng/ml classify as an independent test set. Furthermore, this type of analysis would allow the authors to assess whether the 7 subjects who were self-reported as nonsmokers but had high cotinine levels classify with their appropriate group”.

Thank you for pointing out that building and testing a training set would add to the strength of our manuscript; in fact, we gave this suggestion very lengthy and careful consideration. The objective of this paper was to utilize hyperclustering to gain insight into the mechanism by which groups of genes function coordinately within a cellular or pathophysiological process that are tobacco-dependent, not to identify biomarkers or gene signatures, per se, although given your suggestion we realize the potential in this approach. We are currently working on testing ‘functional’ training sets that employ the hyperclustering method, but are still developing this methodology. Hopefully you find our other changes to our manuscript sufficient to recommend for publication. Regarding the 7 discordant patients, we have addressed this issue in Reviewer Figure 1 and the accompanying text.

Major Revision # 4: “Temporal replicates on a small number of subjects (e.g. collection of blood in early morning on 2 consecutive days) would provide a measure of the reproducibility of gene expression in peripheral WBC within an individual”

Although we agree with the Reviewer that this would be a valuable method of measuring the reproducibility of our findings within subjects, the design of our study was such that blood was collected from subjects as they were being treated for a day procedure at a hospital clinic. Therefore, the ability to gather samples on consecutive days from each subject was not within the bounds of our study parameters.

Major Revision # 5: “Given the identification of an additional 69 genes on GSA, validation of select genes of biological interest from that list (e.g. the interferon responsive genes that they focus on in their conclusion) would strengthen the biological conclusions of the authors”.

This is an excellent point and, as suggested, we have now included one of the GSA-identified genes as part of our qRT-PCR analysis for both technical and biological verification (see Figure 3 and page 14, paragraph 1). The change in direction of all genes detected by qRT-PCR, irrespective of their identification (SAM or GSA) was consistent with the microarray analysis.