Author's response to reviews

Title: Derivation of a Bronchial Genomic Classifier for Lung Cancer in a Prospective Study of Patients Undergoing Diagnostic Bronchoscopy

Authors:

Duncan H Whitney (duncan.whitney@veracyte.com)
Michael R Elashoff (elsahoff@gmail.com)
Kate Porta (kate@veracyte.com)
Adam C Gower (agower@bu.edu)
Anil Vachani (avachani@mail.med.upenn.edu)
J. Scott Ferguson (jsferguson@medicine.wisc.edu)
Gerard A Silvestri (silvestri@musc.edu)
Jerome S Brody (jbrody@bu.edu)
Marc E Lenburg (mlenburg@bu.edu)
Avrum Spira (aspira@bu.edu)

Version: 4
Date: 16 April 2015

Author's response to reviews: see over
Dear Dr. Sands,

Thank you for providing us with the opportunity to revise and resubmit our manuscript entitled "Derivation of a Bronchial Genomic Classifier for Lung Cancer in a Prospective Study of Patients Undergoing Diagnostic Bronchoscopy" (MS# 887984151601445). We have made major revisions in the manuscript based upon the very constructive recommendations of the reviewers. Below please find the reviewers comments in *italics* followed by our point-by-point responses.

**Reviewer 1 Major Compulsory Revisions**

*The authors did not give a convincing explanation on how the markers could be used to identify cancer patients. Indeed, the result on the test data-set shows a high sensitivity but a very low PPV (0.47) which mean that a large fraction of patients will be wrongly predicted to have cancer. I would suggest to discuss those results in more details and explain how the 23 markers and the associated classifier could be used in a clinical setting.*

We apologize that we did not make the intended clinical use of our classifier clearer. We have modified the conclusions section of the abstract and the first paragraph of the discussion to highlight the potential utility of a test with high sensitivity and high NPV to effectively rule out the presence of lung cancer and spare those patients with both negative bronchoscopy and a negative classifier from further invasive testing. In current clinical practice many patients with negative bronchoscopy are sent to more invasive testing and approximately 30% of surgical lung biopsies are performed in patients who are determined to have benign lesions [Detterbeck, et al., CHEST Journal 2013]. A classifier with high NPV has the potential to improve decision-making downstream of bronchoscopy by ruling out cancer and thus helping to avoid unnecessary invasive procedures in those subjects with benign lung lesions. These patients can then be followed noninvasively with repeat imaging of the chest. A test providing high PPV, on the other hand, might be useful for lung cancer screening, but our classifier was not developed for that clinical setting. There are two subsequent validation trials of our classifier (~ 700 patients evaluable patients in total) that are reported in a manuscript that is being revised for publication in the New England Journal of Medicine. The larger sample size in this study allows us to go into considerably more detail about the clinical scenarios in which the biomarker might have clinical utility.

*The test(s) used for the p-value computations should be described on the text.*

In the Methods section, we originally described that a Fisher's exact test was used to calculate statistical significance of all categorical variables and a t-test was used for continuous variables. Some of the p-values had not been originally included in Table 1 and have subsequently been added. The Methods section (p. 10) has also been updated to list the specific continuous and categorical variables for which p-values were calculated using t-tests and Fisher’s exact test, respectively.

*I understand that using genes which expression is correlated with clinical variables, like smoking status and pack years, could give more information to detect cancer patients. But I am not convinced that a gene associated with patient gender could help to identify
cancer patients. What is the effect of using the patient gender instead of the RPS4Y1 expression on the result of the classifier?

We thank the reviewer for making this astute comment and recognize that we could provide greater clarity to the gene selection process. We first identified clinical covariates in our training cohort in order to adjust for those covariates when selecting cancer associated genes. We found that the most highly correlated covariates in our training set were smoking status, smoking history (pack-years), sex, and age. We used gene expression to represent each covariate in order to rely on the residuals of those genes for selecting the cancer genes. Smoking status, smoking history, and sex could be accurately predicted with simple gene expression models. Differential expression of genes associated with these covariates and lung cancer has been shown in prior studies including genes that have been shown to differ between males and females [Spivack, et al., Clin Cancer Res. 2003; Mollerup, et al. Int J. Cancer 2006].

The expression of RPS4Y1 is so strongly associated with clinically ascertained gender that using either gene expression or the clinical indication of sex would yield an equivalent prediction of lung cancer risk. In clinical use, a classifier based on gene expression avoids the possibility of clinical data recording errors and can therefore be used to confirm critical data required to predict cancer likelihood. We have added a sentence to the discussion highlighting this latter point.

*The 3 sub-sections in the “CLASSIFIER DEVELOPMENT” section should be expended to provide more details about the methods used.*

We thank the reviewers for providing feedback that additional details would be helpful to further explain our novel approach of deriving an accurate gene expression classifier for lung cancer. We have added those details to the appropriate sections in the Methods.

*page 9: “First, the HG-U133A RMA expression values were adjusted by a gene-wise constant which shifted the mean of each gene’s expression levels in the test set to the mean observed in the training set”. How the constant have been computed?*

We apologize for the lack of clarity about how the test set data was adjusted prior to applying the classifier. For each gene, the expression level in the test set samples was adjusted by a constant such that the mean value for all test set samples was the same as the mean value in the training set samples. We updated the text in the Methods section (p. 9) to define how the gene-wise constants were calculated. Specifically, the constants were defined as the difference between the mean signal intensity of the test set samples compared to the mean of the training set samples which were calculated separately for each gene. This procedure resulted in a shift of the mean of each gene’s expression levels in the test set to be equivalent to the mean observed in the training set.

*page 9: “Second, for the classifier genes where a corresponding HG-U133A probeset was not available (LYPD2 and RNF150), the gene’s mean expression value in the training set was used for all of the test set samples.” As it is not possible to evaluate the contribution of the genes LYPD2 and RNF150 to identify cancer patient on the test data-set, I would suggest to remove them from the classifier. The result on the test data-set should be similar, and this will lead to a shorter list of markers.*
While we agree that the microarray platform used to generate our test set data forced us to validate a classifier lacking two of the genes (LYPD2 and RNF150), our training set cross validation studies found that inclusion of these genes improve classification performance. Since the classifier will be made available clinically as a microarray-based test using the Affymetrix Gene_ST array (which includes the two genes), we feel it is important to retain these genes in the final algorithm. Moreover, in the manuscript under revision at NEJM, we report the validation of the complete biomarker (including LYPD2 and RNF150) in two additional cohorts. However we thank the reviewer for the comment and have modified the text in our Discussion section (p. 14) accordingly.

**Reviewer 1 Minor Compulsory Revisions**

*page 5 line 95: “Additional file 10”

*The additional file number should be ordered in the order that they appear on the text.*

Additional file 10 was re-numbered as Additional file 2 after minor edits were made in the corresponding section of the Methods. Other Additional files were re-numbered accordingly.

*“page 11 line 217: What is T score ?”*  *“page 11 line 228: “entire training set..””* - > “entire training set.”

The T score is calculated using t statistics to represent the significance of the association of the 232 genes with cancer. The T score of 2.7 is equivalent to a two-tailed p-value of 0.007.*

*page 15 line 312: “However, the differences in the specific genes selected this may be due to differences in the feature selection process.” This sentence is not clear and should be rewritten.*

The section has been re-written as follows:

However, the differences in the specific genes selected in the classifier described here this may be due to differences in the feature selection process, specifically, the method of accounting for gene expression strongly associated with clinical covariates while selecting cancer genes.

*“page 19: the table 1 legend seems to be truncated.”

We apologize for this error which has been amended.

**Reviewer 2 Major Compulsory Revisions**

1. *Both Caucasians and African american’s were enrolled into the test set. Given that ethnic differences may exist. Could the authors comment if any gene expression ethnic differences were observed in the training set? Demographics of the test set should also*
We thank the reviewer for this important point and have performed additional analyses in response. The training set included 227 Caucasians and 60 African-Americans. We found that there was no difference in mean age (p=0.67) or cancer prevalence (p=0.616) but Caucasians smoked more (p=0.036) compared to African-Americans. We then found that there was no difference in the classifier performance by AUC (p=0.622). We have included this data as an additional file (see below) and amended the text in the Results section (p.12). Unfortunately, we do not have race data on previously published dataset that we are using as validation set and therefore are unable to add that to the demographics of our test set as requested.

**Additional File 10:** Analysis of clinical characteristics and classifier performance according to race in the training set

<table>
<thead>
<tr>
<th></th>
<th>Caucasian</th>
<th>African-Am</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>227</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>62.7</td>
<td>63.5</td>
<td>0.67</td>
</tr>
<tr>
<td>PY</td>
<td>46.1</td>
<td>37.6</td>
<td>0.04</td>
</tr>
<tr>
<td>Prevalence</td>
<td>74%</td>
<td>78%</td>
<td>0.62</td>
</tr>
<tr>
<td>AUC</td>
<td>0.81</td>
<td>0.77</td>
<td>0.62</td>
</tr>
</tbody>
</table>

2. In the selection of lung cancer genes, the authors used a model where they excluded genes related to clinical co-variates e.g. male/ female, current/ ex smokers and pack years. It would be useful if the differences in gene expression profiles e.g. through a PCA plot for these clinical co-variates be presented.

We wish to clarify that we did not exclude any genes based on association with clinical covariates as candidate cancer genes. Rather, cancer genes were selected after accounting for the effects of covariates by using the residuals of covariate genes, and therefore those same genes conceptually could still have been selected as cancer genes.

3. Although the sample size of this study is large, a drawback is the heterogeneous cohorts particularly with respect to multiple histologies in both the training and test sets. Could the authors again comment on the impact of the major histology groups (e.g. SCLC, adenocarcinoma and squamous) on their 17 gene expression classifier?

The reviewer raises an interesting point. In both this and our previous work we have been unable to identify either a significant effect of tumor subtype on classifier performance, or genes whose expression in mainstem airway brushings is significantly associated with tumor histology. Based on these data, we have speculated previously (Spira et al. Nature Medicine 2007) that the classifier detects something common to smoking-related lung cancers that is independent of subtype (such as damage from...
cigarette smoke, aberrant repair of cellular damage, tumor-dependent inflammation, etc) but did not feel that the results presented in the current MS justified a revisiting of this speculation. The sensitivity of the classifier in the test set stratified by tumor histology is shown in Table 4.

4. Further, this classifier was developed specifically in current or ex smokers. The conclusion is should be moderated to propose larger validation cohorts, and that this signature is potentially useful primarily in the context of smokers.

We agree with both important points made by the reviewer. We have modified the abstract and the discussion to highlight that the classifier is applicable to current and former smokers, and have modified the Abstract (p.3) and Conclusions (p.17) accordingly and to argue for the need for larger validation trials (p.15).