Author's response to reviews

Title: Gene-environment Interaction Effects on Lung Function- A Genome-Wide Association Study within the Framingham Heart Study

Authors:

    Shu-Yi Liao (shuyi.liao@mail.harvard.edu)
    Xihong Lin (xlin@hsph.harvard.edu)
    David C Christiani (dchris@hsph.harvard.edu)

Version: 2 Date: 22 October 2013

Author's response to reviews: see over
Response Cover Letter

The authors would like to thank the reviewers for their valuable comments and time. We have addressed the comments and provide responses point-by-point below. We have also re-formatted our manuscript according to the editorial team's suggestions. The revisions have significantly improved the manuscript, and we hope to receive a favorable reply. Thank you for your consideration.

Response to reviewers’ comments

Reviewer: Dr. Lisa Maier

Major Compulsory Revisions:
1. Re: they do not characterize the lung function abnormalities into pathologic changes, such as those associated with COPD, which can be caused by the environment. Would it be beneficial to pick spirometry measurements from exam dates of the Offspring cohort that would match in age to the Third Generation Cohort?

Thank you for this suggestion; however, the FHS data used in this study do not provide enough detail to stratify by pathologic processes. In addition, we only have occupational information from Exam 8 in the Offspring cohort (at exam 8 the average age is about 60 years old).

2. Re: should note somewhere that the Framingham study only included Caucasians, as this may impact genetic results.

We have added this information in the Methods: Study Population section, as follows (p. 6, paragraph 3):

Our study population derives from the Framingham Heart Study [25], which includes only Caucasians.

3. Re: they should also report information about health, medications, and if available reported diagnosis of lung disease, like COPD in the demographic table. It would also be helpful to compare the attributes of the two generations as they appear to have differences from a demographic standpoint- lung function, smoking, that could impact the ability to find similar results in these groups.

We agree. Unfortunately, we do not have information on reported diagnosis of COPD. However, the ratio of FEV1/FVC is available. The most likely attribute is age. We have added the corresponding information in Table 2 and in the Results, as follows (p. 9, paragraph 3):

...However, the proportion of COPD cases (defined by FEV1/FVC less than 70%) was higher in the group with highly likely dust exposure. Lung function was higher and the
The proportion of COPD cases was lower in the Third Generation as compared to the Offspring Generation, most likely attributable to the younger age of participants in the Third Generation.

4. Re: the authors mention that SNP analysis was performed on 9,321 subjects from the three generations of participants. However, the analyses focus on just 2 generations of participants. This should be clarified.

We have clarified in Methods: Genotyping and Quality Control, as follows (p. 7, paragraph 2):

...We used 4,785 subjects from two generations in our study.

5.1. Re: Is there more information as to how the JEM was constructed? It would be helpful if there were references or an explanation why certain jobs were considered dusty or not dusty. Who classified the jobs- is the classification from the Framingham study?

We have added this information to the Methods: Occupational Exposure, as follows (p. 7, paragraph 3, continued to p. 8):

...The questionnaire asked, "Using the occupation coding sheet choose the code that best describes your occupation". There are 29 job categories (exclude retired) on the FHS occupation coding sheet, and we classified 4 of them as highly likely dust exposure including factory/assembly/mechanic, skilled labor, general labor, and heavy labor, as modified from UCSF COPD Job Exposure Matrix (January, 2009 revision) [26].

5.2. Re: If only the occupation reported in Exam 8 was used then possible confounding effects should be addressed or at least mentioned in the discussion.

Thank you for this suggestion. We now address the possible limitation in the Discussion, as follows (p. 13, paragraph 3, continued to p. 14):

...Finally, our occupational exposure status was classified based on a JEM having less detailed occupational information than the UCSF COPD JEM; thus, our estimates may not reflect the true frequency. The occupational information in our study derives from a cross-sectional questionnaire, which may not reflect the longest-held job (e.g., participants may have worked in a dusty job for 10 years but had switched to a non-dusty job by the time of the survey). These measurement errors might result in an underestimation of the gene by occupational exposure interaction association.

6. Re: The use of network analysis for the GWAS study was novel and intriguing. It would benefit from greater explanation as to why the authors chose the gene list from the Huang study. Would other studies be appropriate to be used in addition? It would also be helpful if the authors provided the p values of the networks studied.
There is no p-value in our network analysis. The way to build the network by this method is to combine the two potential susceptibility gene lists based on previously reported gene interactions. Huang's study is the only study, to the best of our knowledge, which investigated gene expression in human airway cells after dust exposure or PM exposure. We have added a statement to this effect in Methods: Network Analysis, as follows (p. 8, paragraph 3):

...To the best of our knowledge, no other studies focus on the effects of dust exposure or PM on gene expression in human airway cells.

7. Re: The discussion is quite long but and It would help if it was more connected and included some integration between findings of networks, function and or similar findings from other studies of exposure or genetics.

We have rewritten the Discussion accordingly.

8. Re: The paper needs major editing focusing on grammar, tense and typographical errors.

We apologize for these errors and have revised the text accordingly.

Discretionary Revisions:

1. Re: An overall description of the study design would enhance the methods section and readability of the paper.

We have added a reference about the study design of the Framingham Heart Study in the Methods: Study Population, as follows (p. 6, paragraph 3):

Our study population derives from the Framingham Heart Study [25], which includes only Caucasians. This study has recruited participants since 1948; there have been three generations of participants: the Original Cohort, their Offspring, and the Third Generation.

2. Re: With significant data available to indicate changes in spirometry over time, the authors would have had an advantage to evaluate this measurement and determine if genetics and exposure are predictors in decline or stability of lung function.

Thank you for the recommendation. We did consider analyzing the longitudinal lung function. However, we tried to develop a better method to do this kind of analysis since this is a "longitudinal", "family" data. This would be a consideration in our next publication when the new methods were developed.

3. Re: The authors found higher lung function in those with dusty exposures. This is counterintuitive. It would be helpful if the discussion addressed this issue and how it might have impacted the authors’ findings.
We have added the information in the Discussion as follows (p. 13, paragraph 3):

….In addition, the mean FEV\textsubscript{1} is higher in the group with highly likely dust exposure. This may be explained by gender distribution, since the proportion of males was higher in the group with high likelihood of dust exposure. Another potential explanation is the healthy worker effect confounding bias [55-57]. Participants in the highly likely dust exposure group were likely those in better health since most dusty jobs (e.g. heavy labor) require workers to be in better health. This bias may result in underestimation of the occupational exposure effect and reduce the effect size of our findings.

4. Re: It would be helpful in the discussion if some additional discussion of the function of some of the genes was provided, such as ZNF804A and OPRM1.

Thank you for the recommendation. Unfortunately, the functions of these genes are not well-known. We have added the phrase, “Although their functions remain to be uncovered …” to the Discussion (p. 13, paragraph 2).
Major Compulsory Revisions:

1. Re: Given the lack of replication data and the low power to detect interaction, the authors should put these results in perspective with the published GWAS of lung function, both in the FHS sample and in meta-analyses. For example, what is the result for the HHIP, HTR4, ADAM19, AGER, etc. gene regions in these analyses? Further, the gene SMOC2 has previously been highlighted in FHS studies of pulmonary function, with a specific gene-smoking effect modification, and this gene is observed among the top results in Table 4. How do these results add value to prior findings?

We have added relevant information to the Discussion, as follows (p. 11, paragraph 2):

...For other top SNPs in our study, SNP rs6941466 in SMOC2 (p=3.21 x 10^{-5}) was associated with FEV1/FVC through gene by occupational exposure interaction. SMOC2 has been previously reported in both a gene main effect study [29] and a gene by smoking interaction study [24]. Another top SNP, rs1289714 (p=5.25 x 10^{-4}), in HHIP had a gene by occupational exposure effect modification on FEV1/FVC. The HHIP region was previously highlighted in the GWAS of FEV1/FVC focus on gene main effect using FHS The recent gene by smoking interaction on lung function study [24] discovered three gene regions, DNER, HLA-DQB1/HLA-DQA2, and KCNJ2/SOX9. However, the p-values for the SNPs in these genes were not statistically significant in our study.

2.1 Re: The comment in the Background of the Abstract "previous studies focused only on main effect of occupational exposure or genetics on lung function" is misleading in light of the 2012 GWAS meta-analysis paper examining gene-smoking interactions.

We apologize for the misleading statement and have rewritten the sentence in the Abstract: Background (p. 2), and added the study to our references in Background (p. 5, paragraph 2), as follows:

Previous studies in occupational exposure and lung function have focused only on the main effect of occupational exposure or genetics on lung function.

...The majority of the gene-environment interaction studies focused on gene by smoking interaction [21-24].

2.2 Re: The authors should consider the importance and relevance of those findings to their results (DNER, SOX9). Are there overlapping pathways? HDAC4 was identified in the gene-smoking interaction study as well as the main effects GWAS.

We were unable to identify an overlapping pathway. We have added the following statements to the Discussion (p. 11, paragraph 2 and p. 12, 1st paragraph):

...
...The recent gene by smoking interaction of lung function study [24] discovered three
gene regions, DNER, HLA-DQB1/HLA-DQA2, and KCNJ2/SOX9, with significant
SNPs. However, the p-values of the SNPs for these genes were not significant in our
study.

...Our results are consistent with a recent large-scale GWAS that identified an HDAC4
association with lung function [35], which was also found in the gene by smoking
interaction study [24].

3. Re: Are the analyses really not adjusted for age?

Thank you for pointing out this oversight; we did adjust for age but forgot to mention it
in the text. This point is now included in the Methods: SNP-level Analysis section, as
follows (p. 8, paragraph 2):

...For each SNP, we included age, gender, height (inch), pack-years, smoking status,
occupational exposure status, and tested for the main SNP effect.

4. Re: For the network analysis, how did you "map" a given SNP result to a gene? Is
this simply the nearest gene? Was LD considered?

We mapped the SNP to the gene according to the gene prediction track "RefSeqGenes" in
the UCSC browser (http://genome.ucsc.edu). Therefore, not all of the SNPs can be
mapped. This information is included in the Methods: Network Analysis, as follows (p. 9,
1st paragraph):

...Gene annotation was performed using the gene prediction track "RefSeqGenes" in the
UCSC browser (http://genome.ucsc.edu).

Minor Essential Revisions:
5. Re: Why have the authors focused solely on genotyped variants instead of using
imputed genotypes?

Thank you for the question. In this paper, we focused on actual genotyped SNPs since the
imputed genotypes may not reflect true genotype. However, we take this suggestion and
will consider using the imputed genotypes in subsequent work.

6. Re: It is not clear how the final SNP count became 300,709. How many SNPs were
MAF<.05?

We have amended the Methods: Genotyping and Quality Control, as follows (p. 7,
paragraph 2):

...We also excluded 146,203 SNPs with minor allele frequency lower than 5% in our
study subjects.
7. Re: On page 8, the text refers to Table 2 and states that FEV1 was higher in the group with highly likely dust exposure, but this group has 2x as many men. The higher FEV1 is likely related to the sex distribution.

We now address this in the Discussion, as follows (p. 13, paragraph 3):

...In addition, the mean FEV1 was higher in the group with highly likely dust exposure. This may be explained by gender distribution, since the proportion of males was higher in the group with high likelihood of dust exposure. Another potential explanation is the healthy worker effect confounding bias [55-57]. Participants in that group were likely those in better health since most dusty jobs require workers to be in better health. This bias may result in underestimation of the occupational exposure effect and reduce the effect size of our findings.

8. Re: Several key references are missing including the original FHS GWAS of these traits and the recent GWAS meta-analysis of gene*smoking interaction

We have added these key publications in our references.