Author's response to reviews

Title: Effect of GSTM2-5 polymorphisms in relation to tobacco smoke exposures on lung function growth: a birth cohort study.

Authors:

Melannie Alexander (alexn49@email.sc.edu)
Wilfried Karmaus (karmaus1@memphis.edu)
John W Holloway (J.W.Holloway@soton.ac.uk)
Hongmei Zhang (hzhang@mailbox.sc.edu)
Graham Roberts (G.C.Roberts@soton.ac.uk)
Ramesh J Kurukulaaratchy (Ramesh@jk7771.fsnet.co.uk)
Syed H Arshad (S.H.Arshad@soton.ac.uk)
Susan Ewart (ewart@cvm.msu.edu)

Version: 2 Date: 8 May 2013

Author's response to reviews: see over
Effect of \textit{GSTM2-5} polymorphisms in relation to tobacco smoke exposures on lung function growth: a birth cohort study.

\textbf{Point-by-point response to the reviewers’ critique}

We thank the reviewers and the editor for their insightful suggestions and revisions. We have incorporated their suggestions and thus improved our manuscript, which is now suitable for publication in the BMC Pulmonary Medicine. In the following pages, we have addressed each critique point by point:

Comments from the Reviewer

Reviewer # 1

- Major Compulsory Revisions

1. It remains unclear from the methods section and figure 1, whether all CpG loci in the Illumina 450K array were interrogated with respect to \textit{GSTM2-5} genetic variation or only select loci. If only a selection, then how were these chosen? If all were used, what is the rationale? Given the hypotheses of the paper, it would seem the authors would be most interested in evaluating only the CpG methylation within the \textit{GSTM2-5} genes. Please provide clarification, as power to detect associations will be affected by which approach was employed.

\textbf{Response:}

We agree with this need for clarification and changed the following descriptions:

- In the methods section (“Statistical analysis” subsection of the “Methods” section), the fourth paragraph incorporates more specific information regarding selection of CpG sites:
  o “Methylation quantitative trait loci (methQTL) analyses were performed via Wilcoxon-Mann-Whitney tests to select CpG sites housed in the \textit{GSTM2-5} cluster (k=52) that varied by diplotype among a subset of participants with methylation information (n=245).”

- Figure 1 has been updated to reflect that only CpG sites within the \textit{GSTM2-5} cluster were interrogated. The four steps in Part B of the figure have been updated as follows:
  o “Are \textit{GSTM2-5} diplotypes influencing methylation of CpG sites in \textit{GSTM2-5} cluster?”
  o “Are methylation levels in \textit{GSTM2-5} CpG sites associated with \textit{GSTM2-5} diplotype-smoke interaction?”
“Are methylation levels in GSTM2-5 CpG sites associated with lung function at age 18?”

“Are GSTM2-5 CpG site methylation levels modifying the effect of GSTM2-5 diplotypes on lung function at age 18?”

2. What quality control measures were applied to the Illumina 450K array data? For instance, did the authors remove CpG loci on X and Y chromosomes, or did they remove CpG loci known to be affected by a SNP at the exact locus, etc? (This filtering is largely irrelevant if they only evaluated methylation within GSTM2-5 genes). Nevertheless, if they assessed >200 subjects some commentary on evaluation of batch or plate effects is warranted.

Response:
We agree that the quality control measures must be stated. In addition to explicitly stating that only CpG sites within the GSTM2-5 cluster were interrogated (specified in changes in previous response), we added the following to the “Laboratory analysis” subsection, which is under “Methods”: “Arrays were processed using a standard protocol as described elsewhere [21], with multiple identical control samples assigned to each bisulphite conversion batch to assess assay variability and samples randomly distributed on microarrays to control against batch effects. This analysis is based on CpG sites determined in a single batch.”

3. Please clarify that none of your observed significant methQTLs are in fact a result of a base change that removes or creates a CpG site directly?

Response:
We agree that base changes may alter CpG site methylation; therefore we cross-checked all SNPs involved in the present study against our list of SNPs that alter GSTM2-5 CpG site methylation. We have also examined all CpG sites whose methylation may have been affected by probe SNPs and removed them from further analysis. The following statement has been added to the “Laboratory analysis” subheading within the “Methods” section, first paragraph: “All GSTM2-5 SNPs of interest were checked against a database containing SNPs that may artificially alter GSTM2-5 CpG site methylation. Also, CpG sites were excluded from further analyses if nearby probe SNPs affected their methylation.”

- Minor Essential Revisions

4. Not all references in 2nd paragraph of introduction are formatted properly.
**Response:**

Sorry for this. The revision reads as follows in the second paragraph of the introduction:

“Regarding growth, only one study found an association between *GSTM1* gene function and lung function growth, where children with the *GSTM1* deletion had slower lung function growth compared to those with the normal genotype [6].”

5. Please specify in the methods section what type of blood sample was used for methylation analyses and at what age i.e. cord blood, whole blood, separated cells, etc?

**Response:**

Thanks for pointing at this oversight in the description. The following revision has been made in the Methods section, Laboratory analysis subsection, first subsection:

“Details on genomic DNA extraction from whole blood samples collected at age 18 and genotyping can be found in the online supplement (Additional file 1).”

- Discretionary Revisions

6. One additional reason that the authors only observed a significant interaction between SHS at age 18 and diplotype on methylation of GSTM2 may be the timing of methylation assay. As noted by the authors, methylation was assayed at age 18, and most directly correlates in time. Because methylation is not likely to be static over time, evaluation of methylation only at age 18 may not represent early methylation levels enough to correlate with in utero or early childhood exposures.

**Response:**

We agree with this as an additional explanation; thus, the following has been added to the discussion section as a part of our limitations (Discussion section, paragraph 5). The revision reads as follows:

“In our study, the joint effect of maternal smoking and *GSTM2*-5 diplotype did not influence methylation levels of CpG sites within the *GSTM2*-5 cluster; however, these methylation levels were captured at age 18 and therefore we cannot exclude whether epigenetic changes took place due to smoking at an earlier age nor can we determine that methylation levels at age 18 are representative of early-life methylation profiles.”

7. It would be interesting if the authors could conduct a mediation analysis to test whether altered DNA methylation of GSTM 2 and 5 is on the causal pathway between diplotype and lung function.
Response:

This is an interesting suggestion. Details on implementation of path analysis have been added as the last paragraph of the Methods section, path analysis subsection in Additional File 1. It reads as follows:

“To investigate whether CpG sites that were modified by the joint effect of diplotype and tobacco smoke exposure acted as an intervening variable between the diplotype and the lung function outcome at age 18, a path analytical model was constructed using the CALIS procedure in SAS. A two-stage estimation was performed. First, unweighted least-squares estimates of the model parameters and their residuals were computed. Estimates were then used as initial values for the optimization process to compute maximum likelihood parameter estimates. The path coefficients represent the partial correlation between the dependent and independent variables adjusted for other covariates. These path coefficients are also known as the direct effects (the effect of a risk factor on an outcome that is not moderated by other variables), indirect effects (the effect of the risk factor on an outcome variable with an intervening variable), and total effects (the sum of direct and indirect effects of the path). Figure E1 illustrates the theoretical model from GSTM2-5 diplotype to GSTM2-5 CpG site methylation to lung function levels at age 18 (Additional File 1: Figure E1).”

Additionally, findings from the path analyses were explained in the last paragraph of the Results section. It reads as follows:

“Results of the path analyses showed that although several direct effects of diplotype on CpG site methylation were seen, no indirect effects of diplotype on lung function at age 18 were detected (Additional File 3: Table E14).”

In addition, results from the path analyses were discussed in the first paragraph of the Discussion section. It reads as follows:

“Path analyses show that methylation at cg06970744 did not lie on the pathway between GSTM5 diplotypes and lung function, further providing evidence that methylation at certain sites can modify the effect of genetic variation on an outcome. Also, methylation at this site appears to be unaffected by nearby SNPs (unpublished observations).”

Reviewer #2

Reviewer’s report:

Minor Essential Revisions
1. Smoking has been shown to affect epigenetic factors such as DNA methylation. In my opinion the manuscript will benefit from stratification of smokers versus non-smokers. This analysis is especially important for the effects on methylation status of the GSTM2-5 loci. The authors should also discuss this in the Discussion.

Thanks for this suggestion. Use of interaction terms in general linear models assesses the interaction of an additive scale and is identical with stratification. The use of interaction terms within a modeling approach, has the advantage that is does not require multiple strata, which would reduce the statistical power of the analysis. This was the reason that we did not apply stratification.

2. The biological importance of the findings should be discussed.

For the following reasons, we respectfully disagree with this suggestion:

- Only one study has reported an interaction of maternal smoking with \( \text{GSTM2-5} \) genes. It was published in a prestigious journal, however without a replication study. Hence, there was a need of a replication study to determine if we can rely on the prior study.
- Based on results of our study, it is clear that certain genes within the \( \text{GSTM2-5} \) loci explain some differences in lung function in late adolescence; however, in contrast with the prior study on smoking and \( \text{GSTM2-5} \) genes, joint effects were not detected. In our study, findings were validated through examination of methylation levels.

We believe that future studies are necessary to explain conflicting results before we can start to discuss the biological implication.