Author's response to reviews

Title: Polymorphisms in IL-1beta, Vitamin D Receptor Fok1, and Toll-like Receptor 2 Are Associated with Extrapulmonary Tuberculosis

Authors:

Alison A Motsinger-Reif (alison.motsinger@gmail.com)
Paulo RZ Antas (pzupuim@ioc.fiocruz.br)
Noffisat Oki (onoki@ncsu.edu)
Shawn Levy (shawn.levy@vanderbilt.edu)
Steven M Holland (smh@nih.gov)
Timothy R Sterling (timothy.sterling@vanderbilt.edu)

Version: 2 Date: 2 November 2009

Author's response to reviews: see over
Enclosed is a revised version of the manuscript “Polymorphisms in IL-1β, Vitamin D Receptor Fok1, and Toll-like Receptor 2 Are Associated with Extrapulmonary Tuberculosis” by Motsinger-Reif and Antas et al. Our revisions are based on the thoughtful suggestions of the three reviewers, and we feel that the manuscript is much improved as a result of these changes. We have responded to each of their concerns individually below, with particular attention on the comments of referee 3.

We hope that our revisions have made this manuscript suitable for publication in *BMC Medical Genetics*. We appreciate your review of our manuscript.

Sincerely,

Alison Motsinger-Reif, Ph.D.
Timothy R. Sterling, M.D.

**Referee 3: Guy Brock**

**Reviewer's report:**
The authors present a genetic association analysis of 24 extrapulmonary tuberculosis cases, 24 pulmonary tuberculosis controls and 57 PPD+ controls in two stages. The first stage consisted of 22 SNPs and 3 microsatellites which were previously associated with tuberculosis risk, and the second consisted of 613 SNPs in 26 candidate genes involved in tuberculosis pathogenesis. One potential concern with this study is the small sample size. MDR has been previously documented to have adequate power in sample sizes as small as 200 cases and 200 controls (Ritchie, *Genet Epidemiol.* 2003 Feb;24(2):150-7]. The sample sizes in this study, however, are much smaller, especially when considering just the EP cases (24 patients) vs the PPD+ controls (57 patients), and even MDR will have low power except for markers with large effect sizes. The authors do address this limitation in the Discussion and provide a table with detectable effect sizes for main effects. However, no statement is provided concerning the power to detect gene-gene interactions, which is typically smaller given the number of comparisons involved. Therefore, it is likely not surprising that MDR found no associations with EP tuberculosis in the 2nd stage of
analysis. Further, sparse cell counts will be a concern for MDR when assessing 3-way and higher order interactions with these sample sizes (27 cells and only 24 cases). The authors should add an additional statement that addresses these specific limitations.

We agree with the reviewer’s points, and have added these issues to the Discussion (pages 10-11 of the manuscript).

Another concern is in Table 5 – the average prediction error for this SNP is 52%, which is worse than a coin flip. It is a bit surprising that a SNP with this prediction error has a significant p-value.

In Table 5, both an “average balanced accuracy” and “average prediction error” were presented. Because of the class imbalance (a deviation of the ratio of cases to controls from 1.0) in the dataset, the expected prediction error is not 50%. An error of 52% is better than chance for this level of class imbalance. To address concerns regarding class imbalance, balanced accuracy (the arithmetic mean of sensitivity and specificity) was implemented in the MDR method. A detailed discussion (with references) of these issues has been added to a new Appendix of the manuscript.

The reviewer’s comment (and the comments of referee 2, addressed below) demonstrate that reporting both balanced accuracy and prediction error can cause confusion in interpretation. To improve the interpretation of the models, we have removed the prediction error column from Table 5. The balanced accuracy is a better measure for such data.

**Minor Essential Revisions**

1. **The number of black patients for the sub-analysis should be mentioned in text of the abstract and results, especially since black patients are actually a majority in this study.**

   This is an excellent suggestion, and we have added these numbers to the Results section of the abstract, as well as to the Results section of the manuscript (page 6).

2. **There is no information on the % missing for each marker, or how missing genotypes were handled / imputed.**

   We have clarified our description of missing data patterns and how they were handled. We have also added a description of the level of missingness for the markers used in the current study, and how missingness was addressed in the analysis. This information has been added to the Results (pages 6-7), and the analysis section (page 16).

3. **Table 4: Are the single-locus tests based on the 2 df chi-squared test (for SNPs), the trend test, or allele-test? This should be clarified in the methods.**

   The single locus tests were performed using 2 d.f. chi-squared tests. We apologize for
this omission, and have added this information to the Methods (page 17).

Also, results for significant SNPs should include more information such as
number and percent in each case/control group within each SNP genotype, and
number and percent in each case/control group within each SNP genotype, and odds
ratios with 95% CIs relative to a baseline genotype group.

Due to the “winner’s curse” phenomenon, based on the small sample size in this analysis,
we did not feel that it would be appropriate to report effect sizes for the association
results, given the limitations of the small sample size and the potential to mislead the
interpretation of the results. It has been shown that accurate effect sizes cannot be
estimated from small sample sizes, even though associations can be detected. We have
added a discussion of this to the manuscript, in the Discussion (page 10).

4. In Table 5, the authors should specify whether the ‘Average balanced
accuracy’ and ‘average prediction error’ is based on the CV training or testing data.
Also, the CV consistency should be reported as a fraction, eg 4/10 and 3/10. Why
were the MAF and HWE p-value included in the second portion of Table 5 and not
the first?

In Table 5 we have clarified that the average balanced accuracy is for the testing set from
cross-validation, and have presented the cross-validation results as fractions.
Additionally, we have added the MAF to the first part of the table.

5. The footnote to Table 5 mentions a post-hoc logistic regression analysis of the 2
significant SNPs found by MDR, but gives no details other than the model ‘appears’
to be recessive for both SNPs. It would be more informative for readers to see this
model to understand exactly how these two SNPs are interacting with each other.

With this statement, we were referring to a post hoc regression analysis, where recessive
model encoding of each SNP (as opposed to a dominant, additive encoding) maximized
model fit, leading to our interpretation that the model “appears” recessive. We agree with
the reviewer’s comment that this was unclear as previously worded in the manuscript,
and we have added a more detailed explanation in the footnote to Table 5.

6. The lower detectable odds ratio in Table 6 is redundant, and should be
removed to avoid the value ‘0.00’ which is misleading.

We have removed the lower detectable odds ratio information from Table 6.

Referee 1: Fumio Kishi
Reviewer's report:
In this study, the authors investigated several polymorphisms among three
groups, extrapulmonary tuberculosis cases, pulmonary tuberculosis controls and
PPD+ controls. They concluded that genetic variants in IL1-beta, VDF-FokI and
TLR2 were associated with the risk of extrapulmonary tuberculosis. The authors
chose Multifactor Dimensionality Reduction (MDR) analysis because the number of
samples was small. Even though MDR analysis is powerful method for small size
samples, the sample number is not enough in this study. It is necessary the authors collect much more number of samples, such as more than 50 or more.

While we agree that the small sample size is a limiting factor in our study, unfortunately no additional samples are available. We mention this limitation in the Discussion (page 10). While we are limited in interpreting negative associations, this did not eliminate our ability to identify and interpret positive associations. An increased sample size would improve power to detect associations with smaller effect size, but it does not change the positive associations that were identified.

They described gene variants of VDF-FokI and TLR2 are important risk especially for blacks. As compared to blacks, the number of whites or Asians analyzed is too small. Whether the gene variants are specific to blacks could not be judged properly.

We agree with the reviewer’s point, and have added specific discussion points addressing this issue in the Discussion section (page 8).

The style of references is not consistent through the ref number 1-50. The authors should confirm the style appeared in ref.

We believe that the reference style was consistent before, but we have re-checked the references and confirmed their correct order.

The tables, page 21-27, is not arranged in the order of explanations appeared in the section of results. For example, Table 3 is first in the text. Need better arrange.

We have re-ordered the tables according to their appearance in the text, and renumbered them accordingly. We apologize for this oversight.

Referee 2: Thomas Hawn
Reviewer's report:
Motsinger-Rief et al examine the role of candidate gene polymorphisms and susceptibility to TB. The paper is well written and presents data with a statistical technique that is not common and may offer some advantages over existing techniques. Although interesting, the data is somewhat difficult to interpret in its current format. In addition, the small sample size is a serious limitation for any negative conclusions of the study.

Major Comments.
1. Study Design: The small sample size is a major limitation of this study. As illustrated in Table 6, the power of this study is extremely limited to detect anything except major associations. Given that most of the selected candidate genes have previously been shown to have smaller effects than those that can be detected in this cohort, the value of the studies in the current manuscript is severely compromised and cannot be properly be considered a validation sample. Although the authors acknowledge this limitation, the fact remains that this severely limits any negative
findings from this study.

As discussed in the previous responses, we agree with the limitations of the small sample size, and have added additional discussion addressing these points in the Discussion (pages 8 and 10).

2. Statistical analysis: Multifactor dimensionality reduction (MDR) is not a standard technique used in genetic association studies. Given its statistical complexity, a full review by a genetic statistician may be helpful. Although the technique appears to present some advantages, its use in such a small sample set may not take advantage of its strengths. For example, the authors state that this technique enables analysis of gene-gene interactions. Such an analysis is not practically possible with such a small cohort, which is small even for assessing single SNP associations. The MDR data is difficult to analyze and compare to more traditional ways of presenting genetic data. For example, no odds ratios or genotype frequencies are presented in Tables 4 and 5. At a minimum, a more traditional display of data (with odds ratios, allele, and genotype frequencies) alongside the MDR analysis would be helpful for this reviewer to assess the strength of the association.

As discussed in responses to previous reviewers, the prediction error estimate is the measure of effect size used with the MDR method. We have provided an Appendix with additional explanation and discussion of the MDR method and the use of prediction error, to aid in the interpretation of the results.

Specific points:
1. For the single locus association tests in Table 4, what model is used for analysis? Comparisons of alleles or genotypes? Is an allele trend test used or a dominant or recessive model? Why doesn’t the MDR analysis also find associations with IL-1beta393 and VDR Fok1.

A chi-square test of association with 2 degrees of freedom (i.e., a genotypic encoding) was used for the association tests. This description has been added to the Methods (page 17).

In regards to discussing the differences in the results of the MDR analyses and the single locus tests, it is crucial to keep in mind that while the single locus tests report the results of EACH potential association, MDR performs variable selection that reports and tests the BEST MDR model. These two approaches ask two slightly different questions, and present the results in two slightly different ways. A discussion of this difference has been added to the Appendix (page 3).

2. Table 5: what does average balanced accuracy mean? Or average prediction error? Or cross-validation consistency? Given the lack of familiarity that most readers have with MDR, a more detailed explanation of terms and how the data compares to a conventional analysis would be helpful.
Based on the reviewer’s helpful comment, the description of the MDR method and its associated terminology, as well as the figure demonstrating it, have been rewritten and expanded. This expanded explanation is now in the Appendix.

3. Extrapulmonary phenotype: The extrapulmonary cases are composed of small numbers of many different types. The pathogenesis may be very different among these types—particularly lymphatic vs miliary vs meningeal vs all other. Such heterogeneity among the extrapulmonary cases may limit any conclusions.

The extent to which tuberculosis pathogenesis (and therefore immunogenetic risk factors) differs for extrapulmonary disease at different anatomic sites is unclear. We agree that this is an important point, and one worth pursuing in a much larger study population. One would need a large number of persons with extrapulmonary disease from each of several distinct anatomic sites: lymphatic, pleural, miliary, meningeal, bone/joint, pericardial, peritoneal, etc.