Author's response to reviews

Title: Novel Variants in the PRDX6 Gene and the Risk of Acute Lung Injury Following Major Trauma

Authors:

Melanie Rushefski (meldoran@mail.med.upenn.edu)
Richard Aplenc (raplenc@mail.med.upenn.edu)
Nuala Meyer (Nuala.Meyer@uphs.upenn.edu)
Mingyao Li (mingyao@mail.med.upenn.edu)
Rui Feng (ruifeng@mail.med.upenn.edu)
Paul N. Lanken (Paul.Lanken@uphs.upenn.edu)
Robert Gallop (rgallop@wcupa.edu)
Scarlett Bellamy (bellamys@mail.med.upenn.edu)
A. Russell Localio (rilocalio@mail.med.upenn.edu)
Sheldon I. Feinstein (sif@mail.med.upenn.edu)
Aron B. Fisher (abf@exchange.upenn.edu)
Steven M. Albelda (albelda@mail.med.upenn.edu)
Jason D. Christie (jchristi@mail.med.upenn.edu)

Version: 2 Date: 10 March 2011

Author’s response to reviews:

Author Response:

We thank the editors and reviewers their careful reviews of this manuscript. We have responded to all reviewer comments and changed the manuscript accordingly. We believe the resulting manuscript is much improved. In response to reviewer comments, we have adjusted the details of the statistical analytical methods for association testing and haplotype analysis; and extensively clarified the data presented in the results section. A point by point response to the reviewers’ concerns is included below.

Reviewer 1:

Major Compulsory Revisions.

1) “The author’s considerations about recombination rates and LD in both the abstract and results sections are confusing. To me, LD does not seem to be as strong in this gene in none of the two populations as the authors suggest in the abstract. In fact the authors suggest the opposite somewhere in results. Is this relevant information for the article? My suggestion is to ether clarify this issue or remove any reference to it from the manuscript.”
R1.1) We apologize for any inconsistencies between the abstract and the data. In the last sentence of the Results Haplotype Analysis section, the phrase “with high haplotype diversity” has been removed as suggested.

2) “The authors have to emphasize the fact that sample size (if split by ancestry) is suboptimal. Thus, this gene might be relevant for ALI even though associations were not detected in this study. This is one of the main messages of the study and it should be stated starting from the abstract. Providing details (as an xls file supplement) of their re-sequencing results as multilocus genotypes split by ancestry and disease status is a must for future studies with larger sample sizes with ALI or other diseases where this gene has a key role.”

R1.2) We agree. In response, a statement of statistical power is now included in both the abstract and methods. Furthermore, we now present resequencing results as a supplemental file, as suggested.

3) “The authors tested the individual SNP associations using chi squared tests. However, the inheritance model used for their analysis is not specified and the degrees of freedom of the chi squared test cannot be deduced. This needs to be clarified. In case the authors utilized a co-dominant test (my guess), I would strongly suggest re-testing them either using a trend test (additive) or an allelic test, hoping that they may improve their findings. At least for those showing association (if any), the authors need to show the effects in terms of ORs and 95%CIs. The remaining of this section is confusing because either methods were not clearly described or simply because results are not shown. For example, was association tested for the two populations by separate? Why using 500 subjects for power calculations when association tests were done by population and these divide to about 250 each and less than a half were ALI cases? Where are the results of the interaction? Why testing interactions? Where are results from multivariate analysis? Why using such a stringent significance level when there are correlated tests (note that there’s a typo with the level 0.05/37 is 0.0013, not 0.00013) and the final number of tests was not 37 in any of the two samples (I counted 21 tests for Caucasians and 36 for African Americans)?”

R 1.3) We now specify the inheritance models as suggested. Specifically, we considered an additive (trend) model in primary analyses. In addition, we now present a supplemental file with dominant and recessive models for our data. Regarding the power calculations, we agree that our power calculations should be revised for race specific analyses, and have now corrected this in the revised manuscript. In addition, we have corrected the typo in the level of significance, readjusted the level of significance based on the actual number of SNPs tested (0.0024 for EA and 0.0014 for AA). Multivariable analyses were indeed performed on our SNPs of interest, adjusting for age and ISS. These results are now presented in a new table 7. There were no associations with ALI in
multivariable analyses as now presented in the revised results section. Finally, we did not specifically test for interaction with clinical variables. This sentence in the methods has been rewritten accordingly.

4) Results. There’s a risk in testing SNPs with poor completion rate for association. The large missing information for many SNPs in Table 6 may indicate suboptimal genotyping assay designs among other problems which might be biasing the results. I suggest dropping these SNPs from the analysis, not only from the haplotype analysis (say by including an initial cutoff at 95% completion rate to be considered for further analysis). This shouldn’t affect dramatically the tagging coverage of the final SNP set used for association given that the authors used many other tagged SNPs.

R1.4 We agree and have now excluded SNPs with low completion rate as suggested. We have added a sentence to the methods indicating our completion threshold of 95%. Results have been rewritten accordingly as well.

5) Results. The haplotype analysis is not clear and results are not adding much more information to the manuscript. Haplotype blocks are strongly dependent on the SNPs being tested in the region. Thus, limiting the analysis to them might bias the results. In any case, the results reported for this part are negative, patchy, and inconsistent with methods: they are referred by the authors as they were from sliding windows (not shown), not showing which SNPs were included (see table 7) nor the frequencies of the haplotypes in ALI and non-ALI. This needs to be clarified. In addition, the authors need to switch to alternative haplotype tests, e.g. comprising the whole gene or in sliding/LD-accommodated windows, use haplotype information to infer untyped variants, etc…so that a less biased view of the associations is offered to the reader.

R1.5 We have excluded sliding window results and now present traditional haplotype analyses using methods of Gabriel (Science 2002). This is now clarified in the methods section.

Minor Essential Revisions

1) The title is too broad and should be more informative of the results. I suggest modifying it to something like "Polymorphisms in the PRDX6 gene are not associated with acute lung injury secondary to major trauma".

R1.1) We have respectfully decided to keep the original title because we do not want to state that our polymorphisms are not associated with ALI since it is possible that they are associated, but we were unable to detect it due to our low sample size.
2) In the abstract and at the end of the background section, please, emphasize that this is the first time this gene is analyzed for association with ALI and that re-sequencing was performed to uncover “common” variants. Also, Peroxiredoxin 6 should be in normal font.

R1.2) We have included this detail in the abstract and background and made changes to the text as suggested.

3) Apportion sample size by ethnicity for the samples used for SNP discovery and for association from the abstract.

R1.3) The sample size is now reported by ethnicity for both the SNP discovery set and the cohort used for association.

4) ICU as the source for trauma samples should be stated in the abstract. What is the source of the other 24 samples of European and African descent used for re-sequencing? Are they a subset of those used for association? Please include this information in the abstract and describe in methods.

R1.4) This has now been clarified that the source for all trauma samples in this study is the ICU.

5) Starting from the abstract, some terminology is confusing. The authors refer to tagged SNPs when they should be referring to tagging SNPs. Please, modify it throughout the manuscript.

R1.5) We agree and have modified it throughout the text.

6) In the background, please define ALI as “an inflammatory syndrome characterized by acute respiratory failure due to non-cardiogenic pulmonary edema and hypoxemia”.

R1.6) The definition of ALI has been changed as suggested.

7) In the background, please complete the evidence available for the potential role of peroxiredoxins in inflammation including their relation with cytokine levels and signal transduction (Rhee et al. Free Radic Biol Med 2005, 38:1543).

R1.7) We agree the details in this paper are important to the background section of our manuscript. We have included a statement that Prdx6 has been shown to play a role in both inflammation and cell signaling cascades to the introduction as suggested.
8) The authors analyzed a well-characterized sample of a severe phenotype, which, in part, explains the limited sample size. However, it has been improving with time as compared to previous studies by the group. As part of Methods, authors should mention that patients had major trauma from any cause and report as well all demographic and clinical variables that were collected. This is only shown in a Table (which should be Table 1 instead of 5). Additionally, when authors refer to ALI they are actually referring to ALI + ARDS. If so, please, state that with a sentence in Methods. Additionally, it would be interesting to know how many of the ALI patients developed ARDS. If available, please provide details in the patient’s characteristics table.

R1.8 The demographic and clinical variables collected have been added to the methods section. Also, we agree that describing the patient population as having ALI only is unclear. We have now made it clear in the methods that our patient population consists of both patients with ALI and ARDS, and report the proportion with ARDS.

9- Methods, PRDX6 resequencing. The authors should provide a supplement with primers and PCR conditions used for re-sequencing.

R1.9 An xls file with the primers and the initial cycling conditions has been provided as a supplement. We did not include the exact cycling condition for each primer since they were all within the temperature gradient described in the supplement.

10- Methods, PRDX6 resequencing. The program used for tagging SNP selection and the algorithm utilized should be provided.

R1.10 We used Tagger for tagging SNP selection. We have now added a sentence to Methods, under SNP genotyping, stating that the pairwise Tagger algorithm was used and have provided a reference for the algorithm.

11- Methods, SNP genotyping. Why did the authors adopted a strategy of genotyping both novel SNPs and tagging SNPs? This decision is not explained in the manuscript. In addition, in Table 6, the authors show 9 more SNPs that are absent from Tables 1 and 2 and with a new nomenclature. Are these coming from a different source? Do they have rs numbers as well? Please, try to be consistent.

R1.11 We genotyped novel and tagging SNPs to test for association with ALI, as well as to validate the novel SNPs discovered via sequencing and gain more knowledge about the haplotype structure of PRDX6 using the tagging SNPs. We have now clarified this strategy in the SNP genotyping section of the revised
methods section. The nine additional SNPs in Table 6 are tagging SNPs selected for mid throughout genotyping of association with ALI. Not all of these were observed while sequencing, therefore are not present in Tables 1 and 2. We agree the nomenclature should be consistent and have updated the table where applicable, since some of the novel SNPs do not have rs numbers. We also noted which SNPs were selected from the sequencing project and which were tagging SNPs.

12- Methods, SNP genotyping. To comply with existing guidelines, please give more details regarding genotyping quality controls, namely: Were genotyping calls performed simultaneously for the entire study or were they performed in subsamples (e.g. plates)? Was genotyping blind to disease status? Was all genotyped performed in the same lab (state where)? Did the authors duplicate a fraction of the samples to monitor genotyping concordance? If so, what was the rate of genotyping discordance? Were genotype calls performed manually or automatically? If automatically, which algorithm was used?

R1.12 We have now added a paragraph to the SNP genotyping methods section to clarify the quality control measures used during genotyping and data analysis.

13- Methods, Statistical analysis of ALI associations. Please, state that the LD was calculated in terms of r2 values.

R1.13 We agree has and have now corrected the text in the methods section.

14- Methods, Statistical analysis of ALI associations. Hardy-Weinberg equilibrium should be tested, at least for controls, as an extra level of genotyping quality control. These values can be included in a supplement table along with the completion rate (or the missing genotype rate as referred in table 6).

R1.14 We did test for Hardy-Weinberg equilibrium, but failed to report it in the initial paper. We now include a supplement with the Hardy-Weinberg equilibrium test results and summarized these results in the results.

15- Results. Resequencing results should be summarized in a single table including both novel and previously described SNPs (put together tables 1 and 2) with MAF (not observations) split by population. The author should add another column indicating, at least, 10 bases of each flanking sequence for novel SNPs.

R1.15 Tables 1 and 2 are now combined into one table. We now report the MAF rather than the number of observations. We also include approximately 10-15 bp of flanking sequence for novel SNPs. We agree that this makes the table more informative.
16- Results. Table 6 needs homogeneity for the decimals; say 2 for MAF and 3 for p-values. Additionally, it would be informative to provide MAF for ALI and non-ALI by separate.

R1.16 We adjusted the decimal places as suggested and also now provide the MAF for ALI and non-ALI.

17- Discussion. No empirical assessment for the presence of population stratification or a correction for it was considered. This is a limitation of the study that needs to be clearly identified and discussed. Authors should discuss also whether the study design utilized any step to limit confounding by population stratification, sample size limitations, limitations of the study to detect rare variants and their effects, among others. As a conclusion, the authors need to report the negative association found for ALI in the context of study limitations.

R1.17 We now state in the discussion that population stratification was not formally assessed. We also comment on the limitation of potential confounding, as suggested.

Discretionary Revisions.

1. Methods, PRDX6 resequencing. The second and third sentences of this section should be combined into a single sentence like “……selected for sequencing of PCR fragments, providing a power of 99% to detect…. “.

R1.1 We agree that the original text was repetitive and have modified it as suggested.

Reviewer 2:

Major Compulsory Revisions

1. The study is underpowered to detect associations with relative risks roughly below 1.7 at common allele frequencies. Lessons from numerous GWAS have highlighted that the overwhelming majority of identified variants have relative risks well below 1.7. Since this article is reporting a negative result, it should be very clear in the discussion to what spectrum of allele frequencies and effect sizes the negative result applies. An additional figure where the power of the study in relationship to allele frequency and effect size would further aid the reader.

R2.1 We agree that the power of this study was not clearly stated in the initial submission. We now include the sample size along with the power for common
minor allele frequencies in the methods, include a figure as suggested, and have added consideration of the findings in this context in the revised discussion section.

2. The 1000 genomes project www.1000genomes.org was designed to capture all human genetic variation at minor allele frequencies greater than 1%. In principle, gene resequencing to harvest new variants becomes unnecessary. The 1% maf is around what the re-sequencing experiment in this paper was set up to achieve. It would be very helpful to the readers to gain information about the need to still perform these re-sequencing efforts. Therefore, a section comparing the variants identified with those from the 1000 genomes project would be highly desirable.

R2.2 We compared our results with the 1000 Genomes project as suggested and found that 28 of our SNPs were not registered in the database. We now include a table detailing which of our novel sequencing SNPs are in the 1000 genomes database and which are not. We interpret this as being due to the fact that we performed sequencing on subjects with an extreme phenotype, and have added text to the discussion section accordingly.

Minor Essential Revisions

1. The sub-section “Statistical Analysis of ALI association” in the “Methods” needs to be more descriptive. Is the p-value <0.01 later corrected by the 2x37 tests? There is a reference to multivariable analyses using logistic regression but no mention of this is made in the Results. How was the sliding window haplotype analysis implemented?

R2.1 We apologize for the error in reporting an improper p-value cutoff for significance. We have corrected this in the revised methods section. In addition, multivariable models are now presented. We would also like to apologize for the description of the haplotype analysis as sliding window. Haplotype analyses were performed using standard methods as now clarified in the revised methods section.

2. Several variants were identified in only one individual. It would be very useful to include in Tables 1 and 2 the confidence scores for the genotype calls.

R2.2 Confidence scores for variants identified in only one individual are now presented in Table 2 as requested, reported as the percentage of overlap between heterozygote peaks.

3. Caption of figure 1: The figure has arrows below the gene structure and I assume these are the regions for which amplicons could be designed. This
should be specified in the caption.

R2.3 The reasoning for the arrows below the gene schematic has been added to the caption as well as the reasoning for the arrows and lines in the schematic. We agree this could be confusing to the potential reader.

Discretionary Revisions

1. The sections on transcription factor binding sites and microRNA sites are not useful. In my experience, particularly for transcription factors, the findings are largely dependent on the method used. Given the small length of the binding motifs, the chances of false positive findings are enormous. Instead, if the authors would like to keep this section, I would advice to draw from ENCODE and other large consortia that are charting the epigenome. This way the localization of SNPs within regulatory elements and binding sites for transcription factors can be based on experimental evidence.

R2.1 We compared the transcription factors returned from the TESS search with the ENCODE data and found that only 3 of our transcription factors overlapped with data in the UCSC genome browser ENCODE tables. We filtered the ENCODE data for the promoter of PRDX6 and found that these 3 transcription factors have not been experimentally shown to bind the target sequence in that region. These findings are now reported in the revised methods and results sections, and commented upon in the revised discussion section.

Reviewer 3:

Major Compulsory Revisions.

1. Methods, SNP genotyping. Please describe whether quality control approaches were implemented in genotyping, if so, how (e.g., percentage of random samples repeated). Given the high genotype data missing rate (up to 66% for 7 SNPs) mentioned in the discussion section, both the genotyping completion rate and data concordance rate should be reported.

R3.1 Please see our response for R1.12. We have now added text to the SNP genotyping methods section stating the quality control measures taken. Also, we have now noted in this section that the missing genotype frequency is reported in revised Table 6.

2. Methods, Statistical analysis of ALI association. It's not clear whether the association testing was performed separately for the European and African Americans. Also authors need provide more details on potential confounding factors adjusted in the multivariable genetic association analyses.
R3.2 Testing was performed separately for European and African Americans. This is now clearly stated in the revised methods section. Confounding variables in the multivariable analyses are now clarified.

3. Results, Identification of novel polymorphisms in PRDX6, line 5. Table 2 listed 37 SNPs which are known. Providing allele frequency information in addition to number of subjects carrying the SNP (last column in Table 2), and a comparison of allele frequency information available in the public databases (e.g., dbSNP, HapMap or the 1000 Genomes Database), will be more informative.

R3.3 Please see our reply for R1.15. We are now presenting tables 1 and 2 as one table (Table 2) and removed the number of observations and replaced them with the MAF, as well as included 10-15 bp of flanking sequence for all novel SNPs. We have also added MAF information from dbSNP were applicable. We agree that it is more informative to report the MAF from a public database as a comparison to our data.

4. Results, Identification of novel polymorphisms in PRDX6, line 6. Authors mentioned that 25 of the novel SNPs uncovered were submitted to NCBI, are dbSNP ss# available for those? If so, please list them in Table 1.

R3.4 dbSNP ss# numbers are available, and now presented in Table 2.

5. Results, In Silico function of novel SNPs in PRDX6. Given that no association was found between PRDX6 SNPs studied and ALI and no variants in the coding regions was identified, the major strength of this paper is the discovery of potential functional variants in either gene promoter that could regulate gene transcriptional activity or in 3'UTR that might be altering mRNA stability. Authors described the findings of TESS database search in Table 3, however, more details are needed, i.e., 1) how many SNP sites within PRDX6 gene either create or abolish a transcription factor binding site? 2) Of them, how many are novel?

R3.5 We have added the details as suggested to the revised results section. We agree this information better describes our findings.

6. Results, Association of PRDX6 with ALI. Since “Blunt Mechanism (%))” only appeared in Table 5, and was not described in Methods/Patient population, please provide the description in Table 5 for “Blunt Mechanism (%))”.

R3.6 The description for “Blunt Mechanism (%))” is now included in revised Table 1 (formerly Table 5).
7. Results. Association of PRDX6 with ALI. Is there a reason why the authors chose 3 and 10 consecutive SNPs for the sliding window haplotype analysis?

R3.7 Please see response for R2.1. We apologize for the error of reporting the haplotype analysis as being from a sliding window. We have corrected this throughout the manuscript.

8. Was the association testing performed for severity (APACHE) and outcome variables?

R3.8 We did not test for association with outcome variables since there is not enough mortality (low power) in this cohort. We did not associate with APACHE as it a composite score driven in part by chronic health conditions.

9. Discussion. Given that the major finding of this study is the identification of novel SNPs within the PRDX6 gene and its 5’ and 3’ flanking regions via direct sequencing, the subscript is to justify future studies in identifying PRDX6 gene as genetic risk factors in other diseases. Therefore, I believe that more discussion directly on the involvement of Prdx6 in other diseases and conditions (e.g., lung cancer) needs to be provided in the text in order that the reader can evaluate more precisely the significance of the findings of the study. Similarly, authors briefly mentioned the important interaction between Prdx6 and GSTpi in the background section, but didn’t further explore that as possible future direction testing gene x gene interactions in the Discussion, especially given that genetic association between glutathione s-transferase (GST) gene variants and a variety of human diseases have been established.

R3.9 We have now added a paragraph to the discussion describing future directions for PRDX6 research including involvement in lung cancer and the importance of gene-gene interactions with GSTpi.

10. In summary, this is a potentially interesting piece of work given the protective role of PRDX6 against oxidant injury. However, discussing the biochemical and mechanistic implications of the resequencing results (rather than the genetics) would enhance the value of the manuscript to the general reader.

R3.10 We have added a section to the discussion describing potential biochemical and mechanistic changes based on the regions of SNP discovery, focusing on the promoter and 3’UTR.

Minor Essential Revisions.

1. Background, 2nd paragraph. Please spell out class pi
glutathione-S-transferase (GSTpi) when it first appeared in the text.

R3.1 The text has been modified as suggested.

2. Background, 3rd paragraph. Please provide URLs for SNPbrowser and HapMap.

R3.2 We now provide the URL for HapMap, but have removed SNPbrowser from the paper as it is software provided by Applied Biosystems that is used for ordering assays more so than obtaining allele frequency information.

3. Methods, PRDX6 resequencing. Authors referenced published literature (Kruglyak L, et al., Nature Genetics) for the power calculation method used in this study, please provide direct information on the software or program and parameters used for the power calculation.

R3.3 The Nature Genetics reference was for determining the number chosen to sequence, we apologize for not including the additional reference for our power calculations. PGA was used for power calculations, and is now referenced accordingly.

4. Methods, Statistical analysis of ALI association. Line 7, please spell out LD when it first appeared in the text.

R3.4 The text has been modified as suggested.

5. Results, Identification of novel polymorphisms in PRDX6, line 4. Please provide NCBI dbSNP build number for the SNP matching comparison and the URL for Genewindow.

R3.5 The information requested has been added to the revised results section. We agree this needed to be clarified.

6. Results, Haplotype Analysis. Authors mentioned that a total of 37 SNPs were genotyped and 7 were excluded from the haplotype analysis due to a high genotyping failure rate. However, in Figures 2 and 3 legends, authors showed the haplotype structure generated from 31 SNPs (should be 30).

R3.6 Please see reply to R1.4. We readjusted our haplotype analysis to include a 95% genotype completion rate. We apologize for the error and the typo in figures 2 and 3 – these have now been corrected.
Discretionary Revisions.

1. In the abstract, authors may want to clearly state the study design: trauma- ALI cases were compared to ICU sick controls.

R3.1 We have now clarified this in the abstract as suggested.

2. Results, Identification of novel polymorphisms in PRDX6, line 10, Figure 1. For the readership, I would suggest label all (if possible) or some prioritized novel SNPs with their corresponding IDs. And more descriptions are need for Figure 1 legend to explain PRDX6 gene structure and symbols/abbreviations used for illustration.

R3.2 Please see reply for R2.3. We now present a more descriptive legend for Figure 1.

3. Results, In Silico function of novel SNPs in PRDX6. Since haplotype analyses did not strengthen the single marker association signals, I would suggest move Table 7 to the supplementary section.

R3.3 We agree this does not add much to the analysis and now present the haplotype data as a supplement.

Other minor issues:

1. Methods, In silico modeling of putative function in SNP sites, first paragraph.

Italic font should be used for “in silico”.

R3.1 The text has been modified as suggested.

2. Methods, Statistical analysis of ALI association. line 6. Please delete extra period after “ALI =0.30”.

R3.2 The extra period has been deleted. Thank you for bringing it to our attention.

3. Discussion, 4th paragraph. Should “mrR-942” be replaced by “miR-942”?

R3.3 Yes. We have corrected the typo.