Reviewer's report

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

Version: 1 Date: 25 November 2010

Reviewer: Carlos Flores

Reviewer's report:

Dear Editor,

The manuscript entitled “NOVEL VARIANTS IN THE PRDX6 GENE AND THE RISK OF ACUTE LUNG INJURY FOLLOWING MAJOR TRAUMA” by Doran and coworkers describes an association study of selected SNPs in PRDX6 gene and acute lung injury (ALI) risk following major trauma. After re-sequencing the entire gene region in 48 samples, no association was found for any of the SNPs and haplotypes tested in a multiethnic sample of ICU patients enrolled in a cohort study. In addition, they present novel variants and the in silico study of their predicted effects. This is a well conducted study using an appropriate cohort sample size, and with a reasonable hypothesis given the importance of redox regulation in ALI. As such, the findings by Doran et al. are meaningful and add knowledge to this growing field (Flores et al. Crit Care 2008, 12:R130). Thus, the manuscript should be considered for publication in BMC Medical Genetics with the following suggestions, mostly to adhere it to STREGA standards for reporting genetic association studies:

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.

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.

3) Methods, Statistical analysis of ALI associations. 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)?

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.

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.

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".

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.

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

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.

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.

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

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).

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.

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

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

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.

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?

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

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).
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.

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.

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.

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……”.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests