Reviewer's report

Title: A candidate pathway-based strategy to filter sets of genes for gene-environment interaction studies: The asthma example

Version: 1 Date: 17 April 2013

Reviewer: karine audouze

Reviewer's report:

The manuscript proposes a strategy to select gene-environment interactions. This is a very challenging and important step to increase knowledge of complex diseases. Asthma and tobacco smoke are used as a case study.

- Major Compulsory Revisions

The manuscript presents an interesting approach to select gene-environment interactions. However, there are significant shortcomings that the authors should address.

1. How do the authors select the GO term ‘response to oxidative stress’? The authors should state it.

2. Why do the authors use only 1 GO term? An integration of all significant GO terms related to asthma will increase the value of the study, as it will increase the specificity of the study for asthma.

3. The method describes in the manuscript need further details. It is not clear how the authors select the first gene list. Do they extract genes related to GO6979? If yes, why the authors use asthma as a disease reference? The link between asthma, GO 6979 and the gene list is not obvious.

Considering that obesity or Type II diabetes are also strongly connected to the GO term ‘response to oxidative stress’ and to tobacco smoke, the final 104 selected genes in the present study could have been selected for obesity as well. The present study is missing a direct and specific link to asthma. This should be clarified and/or corrected.

4. For the GO data extraction, a more complete description of the approach is needed. Regarding the gene selection from GO, which organism(s) was/were used? Which evidence(s) was/were considered? This is very important for such study as the use of all evidence might add noise and reduce the quality of the results.

5. Authors should add references used when increasing the initial GO gene list (237 genes to 265 genes).

6. Concerning the selection of the environmental factor. Why do the authors choose tobacco smoke only? It is well known that people are daily exposed to a
large range of environmental factors; therefore integration of more factors (e.g. dust, vehicle emission…) connected to asthma will increase the robustness of the approach.

7. A table with the top results from Step 1 will be useful to have (with p-values). A comparison between the results obtained using both pathways analysis will be interesting (overlap between both…).

8. Regarding Step 3, the authors should mention how they select the 12 pathways. They wrote ´…that are the most relevant to tobacco…´ on what do they base this statement?

9. The study allows selecting 104 genes. A comparison with the already well-know genes connected to asthma should be included to promote the significant gain of such study.

- Minor Essential Revisions

1. page 5. ´….we propose a filtering strategy for obtaining a large gene set…´ Is the focus is not to obtain a significant gene list rather than a large gene list?

2. Authors should include date of access to website references.

3. page 7. A table with the 12 selected pathways (and p-val) will be clearer to the reader than an enumeration in the text.

4. Table 1? Does the p-values correspond to the IPA analysis? It will be nice to mention it in the legend. A ranking by significance (p-value) would be helpful for the reader.

5. A table with the 104 selected genes is missing.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

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.