**Author's response to reviews**

**Title:** Association between Paraoxonase gene and stroke in the Han Chinese Population

**Authors:**

Guojun Zhang (zguoj1018@yahoo.com.cn)
Wenjin Li (wenjinliwj@gmail.com)
Zhiqiang Li (lizqsiju@gmail.com)
Hong Lv (lvhonghong@yahoo.com.cn)
Yonghong Ren (yhren@capitalbio.com)
Ruimin Ma (ma.ruimin@yahoo.com.cn)
Xiaohong Li (xhli@capitalbio.com)
Xixiong Kang (kangxx@VIP.sina.com)
Yongyong Shi (shiyongyong@gmail.com)
Yimin Sun (ymsun@capitalbio.com)

**Version:** 6  **Date:** 12 October 2012

**Author's response to reviews:** see over
Dear Editor,

Thank you for your hard working and for the reviewers’ comments on our manuscript entitled “Association between Paraoxonase gene and stroke in the Han Chinese Population”. Those comments are very valuable and helpful for revising and improving our paper, as well as the important guiding significance to our researches. We have studied comments carefully and made correction which we hope will meet the criteria. Revised tests are marked in red in the manuscript.

Sincerely yours,

Wenjin Li, wenjinliwjl@gmail.com

Corresponding author:

Name: Yimin Sun, E-mail: ymsun@capitalbio.com;
Name: Yongyong Shi, E-mail: shiyongyong@gmail.com.

The following is a point-to-point response to the three reviewers’ comments.

Reviewer #1:

Thank you the reviewer for the comments.

Reviewer #2:

MAJOR COMPULSORY REVISIONS

Q1: Specifically, please revisit my previous comment 3.A, on your choice to perform
adjusted analyses. In general, we should be very clear in why we are adjusting for some factors, and what the adjusted estimates mean. According to standard epidemiological thinking, we should not be adjusting indiscriminately. For example, it is not a good strategy to adjust for any factor that was significantly different in the two groups, without some additional rationale and clarification (and this the strategy you state in the methods). In your reply to my comment, you give a very different justification for your adjustments, namely that several factors are associated with stroke. This is getting closer to a better explanation, but it is not sufficient. To address my question, you have to describe how you think that the major factors are causally related. For example, in your reply to my comment in your cover letter, you indicate that you believe that the following directed acyclic graph (DAG) is correct: GENE ----> LDL ----> Stroke. Then you should *not* adjust for LDL - you do not adjust for intermediates in the path. If you believe that the GENEs affect Stroke not only through intermediaries but also directly (there is an arrow from GENE to Stroke), again to get the total effect of the GENE you would *not* adjust for LDL (the intermediate factor). If you adjust for the intermediate factor in the second case, you are attempting to do a mediation analysis, and you should do a lot more explaining, and perhaps use more advanced methods that what you did here. So, if you opt to keep the adjusted analyses, please show the DAG you operate under, and explain what the adjusted results mean.

**Answer:** Thank you very much for the comment. Your suggestion is very insightful.

In our study, age, HDL and the incidence of hypertension were included in the
logistic regression model for adjustment. The main reason why we adjust for these three covariates is that we want to reduce the false-positive rate. If age, HDL, and the incidence of hypertension which have significant difference between the subjects, to some degree, could make the sample stratified into cases and controls with bias, then genes that have correlation with these factors but do not have true association with the disease will be falsely recognized as being associated with the disease. Accordingly, when the correlation between these factors and the disease are not clear to us or when we cannot get a clear DAG like path, adjusting for these factors may be a more rigorous way to assess the association results. Moreover, we keep both the adjusted and the unadjusted results, which both show that there exists a significant association with the disease. Additionally, we have found that multiple logistic regression is conducted with adjustment for factors such as age, sex, history of hypertension, diabetes as covariates in some similar researches on stroke [1-3]. Therefore, we do believe your view is quite right when we could get a clear DAG path, but in order to reduce the effects of these confounding factors, we opt to choose a stricter way and to keep those adjusted analysis.

Reference:


Q2: I meant that you could just discuss the findings in the context of the previous meta-analyses, but you went a step further. That is fine, and I commend you for this extra work. I would think that you should add in the methods section that you used 2 previous meta-analyses and added your study to them to contextualize your findings. Acknowledge that this is not a real meta-analysis update (you have not searched for additional studies that may have appeared in the meanwhile, nor performed real analyses, not even analyzed other genetic models). I would use only the random effects model.

Answer: Thank you for the comment. We have used only the random effects model as suggested and have re-written this part according to the suggestion as shown in the revised method and meta-analysis sections.

Q3: Note that in the haplotype analyses, the OR for each haplotype refers to a reference haplotype. Usually, this is the most common haplotype. Please mention the reference haplotype in the table (or its footnote). As you know, the choice of the
reference haplotype is arbitrary. For example, I could ask you to use any haplotype as a reference. All the ORs would change, but the omnibus test will remain exactly the same. Therefore, the interpretation of the ORs is very much unique to the reference haplotype. The way that you describe the haplotypes' ORs as protective or predisposing in the Discussion is not clear -- you have to say with respect to the reference haplotype XYZ.

**Answer:** Thank you for the comment. In the haplotype analysis, the OR in one block for each haplotype was calculated by using all the other haplotypes in the same block as the reference haplotype. We have added this footnote in the revised Table 5. We think the previous contents concerning the OR in the discussion part were not clear enough, therefore, we omitted this part.

**MINOR:**

Q4: Remove the term cohort from the Abstract.

**Answer:** Thank you for the comment. We have removed the term cohort in our revised manuscript as suggested.

Q5: Drop the fixed effects meta-analysis graph.

**Answer:** Thank you for the comment. We have dropped the fixed effects meta-analysis graph and opted to use only the random effects model as suggested.

Q6: Some attention to language is needed. The edits have introduced some syntax and grammar errors.
**Answer:** Thank you for the comment. Efforts have been made to improve the English of the manuscript.

Special thanks to you for your good comments.

**Reviewer #3:**

Thank you the reviewer for the comments.