Reviewer’s report

Title: Associations between male infertility and ancestry in South Americans: A case control study

Version: 0 Date: 30 Dec 2016

Reviewer: Wolfgang Forstmeier

Reviewer’s report:

This study examines genetic causes of male infertility by associating genetic information (mitochondrial haplotype, Y-chromosome haplotype) with (1) fertility problems in a case-control comparison, and (2) sperm characteristics among male patients with fertility problems. The first of the two approaches does not detect any differences in haplotype frequencies between infertile men and controls. The second approach yields an apparent association between the Y-chromosome haplotype F(xK) and sperm morphology, however, the test that is implemented (logistic regression) lacks statistical power because it fails to make use of the full information content of the data that appears to be available. While sperm characteristics of patients were measured on a continuous scale (sperm count, motility, morphology), patients were grouped according to whether the fell below or above a certain cut-off value, and then haplotype frequencies were tested as a predictor of this grouping. I do not understand the reasons for choosing such an analysis strategy where most of the information content is discarded before testing. I strongly recommend re-analysing the data with a more powerful test, namely an ANOVA, where haplotype is used as a predictor of the sperm trait that is measured on a continuous scale (sperm counts should probably be ln-transformed to approach normality). This test should reveal with much greater statistical power whether the claimed association is likely to be real or likely to represent a false-positive finding. The latter are very frequent in exploratory studies of genotype-phenotype association studies like the present one. In my opinion, publication of such exploratory studies should not be conditional on reaching statistical significance (since unreplicated positive findings in association studies are predominantly false-positives anyway), but it is essential that statistical tests make full use of the available data.

Furthermore, the study should present more details about the sperm characteristics that were measured. Currently, the sperm traits and their quantification are not described in the methods and the obtained values are hardly visible from the results section and they are not discussed in comparison to reference values from control subjects that do not suffer from fertility problems.

Finally, the writing and use of grammar should be improved.

Specific comments:

1. The Abstract would benefit from a substantial revision. I recommend mentioning "mitochondrial haplotype" and "Y-chromosome haplotype" already in line 34. Then I would cut the end of the sentence in lines 35-36 (from "looking" to "failure"). Line 39 should be changed to "compared among haplogroups by ANOVA". Line 41: I suggest "The genotyping confirmed the
known admixture...". In line 42, "both groups": specify that these are infertile and controls. Line 44: specify that this is a Y-chromosome haplotype. Line 46: consider replacing "maternal ancestry" with "mitochondrial haplogroups". The current "Conclusion" section is mostly redundant and should be replaced with a statement that the possible association between the Y-chromosome haplotype F(xK) and sperm morphology needs further confirmatory testing.

2. Line 85: I suggest replacing "applied in medicine from the perspective" with "used to study"

3. Lines 116-118: Why is this finding not used to make a specific confirmatory (rather than exploratory) test? Wouldn't this study predict that men of haplotype H have higher sperm motility? This could be tested with the present data.

4. Line 140: should be changed to "where compared among haplogroups"

5. Line 150: I do not understand the logic of this procedure. If all included patients were already tested negative for Y-chromosome microdeletions, why were they tested here again (lines 192-197), and does this double-testing explain why there was only one single positive case (line 214)?


7. Lines 208-210: I had great difficulties making sense of these numbers provided. The term "azoospermia" should be defined for the non-specialist reader (is this sharp zero sperm or <50,000 sperm per ml?). Then, the number provided seem to say that 18% had 0 sperm, 60% ranged between 50,000 and 5 Mio sperm, and 22% ranged between 5 and 13 Mio sperm, with an average concentration of 3.5 Mio among the latter 82%. However, then in Table 5 it looks like 40 patients actually had sperm counts higher than 15 Mio sperm per ml. These results should be presented in a way that allows for no misunderstandings.

8. Lines 211-212: 38% and 65% out of how many (120 or 98)?

9. Line 212: "below 4% of normal sperm morphology" This is not understandable to the non-specialist reader and no reference to normal values is given.

10. Line 274: explain TSPY

11. Line 282 and 326: Note that, for the non-recombining part of the Y-chromosome, proximity does not predict association like it does for autosomes.

12. Line 288: Note that this should be true for any Y-chromosome association with infertility, because any mutation that leads to unconditional infertility gets immediately removed from the gene-pool.
13. Lines 316-317: unclear. Maybe it is worth noting that in such cases of population admixture, a relatively lower Y-chromosome diversity has been maintained compared to the mitochondrial diversity (where both European and South American genotypes were maintained)

14. Line 618: "see text": no such text is provided anywhere.

15. Tables 5 and 6: Since numbers do not add up to 120 patients, an additional column with "other haplogroups" or "not genotyped" should be added.

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

No

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Yes

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

No

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

**Quality of written English**
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal