Reviewer’s report

**Title:** Genetic Variation in SPAG16 Regions Encoding the WD40 Repeats is not Associated with Reduced Sperm Motility and Axonemal Defects in a Population of Infertile Males

**Version:** 2 **Date:** 26 June 2012

**Reviewer:** Jim A Mossman

**Reviewer’s report:**

Genetic Variation in SPAG16 Regions Encoding the WD40 Repeats is not Associated with Reduced Sperm Motility and Axonemal Defects in a Population of Infertile Males

Nagarkatti-Gude et al.

This manuscript describes an investigation of Spag16 gene polymorphisms and sperm parameters and ultrastructural morphology in a sample of 60 infertile men of Italian origin. The authors found no previously undescribed mutations in this gene when comparing their sample sequences to the HapMap cohort. Furthermore, there were no associations between genetic variants and any measure of sperm parameters or morphology in this sample.

I think this manuscript is well written, concise, and deals with an interesting area of research which will be of general interest to the readers of BMC Urology.

Minor essential revisions:

1) In Figure 1, I did not understand the description of the shading and boxing. The manuscript (Results, para1) “Alignments are shaded and boxed to indicate identical and conserved amino acids, respectively, relative to the human sequence”. In my opinion, the shading does not correspond to identical amino acid sequence relative to human (e.g. position 12, human = “V”, mouse and bull = “L” all with the same shading). Perhaps this needs a better description of shading or what is considered identical or conserved.

2) (Results, para 2) “…and primates (Fig.3).” Should this be Fig. 2?

3) (Results, para 3) The authors state the probe set was sensitive to the rs10167688 polymorphism since it was used in an independent study using West African DNA samples. Is there a reference for this statement?

4) (Results, para 5) linkage disequilibrium, (Table 3 title): Dysequilibrium is misspelled.

5) The two SNPs in LD are adjacent polymorphisms (according to their IDs), with one in an exon and the other very close in the adjacent intron. If so, this could be briefly mentioned in the discussion because it is perhaps not unexpected. Since
they are adjacent polymorphisms, they should perhaps be next to each other in the Supplementary Table 2.

6) In Supplementary Table 2, two of the three significant associations are in bold type yet one (rs16851495 vs normal acc. fibers, p=0.041) is not. Is this a typo or is there a reason for this difference?

7) (Results, para 8) “..3D structure model shown in Figure 2” Should this be Figure 3?

8) (Discussion, para 1) Given the limitations of the chosen sample from males with very poor semen characteristics, it is perhaps a pity that a larger or more representative sample was not investigated. My main concern of this investigation is that the inclusion criteria for the sample has selected a population of men that had very narrow variation for the semen and morphological traits of interest because (Methods, para 1) “..Inclusion criteria for patients were…absence of..ultrastructural sperm defects of possible genetic origin.” My interpretation of the sentence is that men with pathological sperm morphology were removed from the sample before genetic effects were investigated, which is exactly what the aim of the study was! The lack of sensitivity to detect associations when the variation of the sample is so narrow (e.g. WHO A+B #10%) make the null results look prime for further investigation anyway, which the authors discuss, but which could possibly be elaborated. I would also mention future inclusion of control samples from men with high values of semen characteristics, or patients with semen characteristics conferring known fertility to contrast the subfertile cohort, which as the authors rightly acknowledge is a main weakness of the study. While the minor allele frequencies were not different to those of the HapMap cohort, the latter is not an appropriate control for individuals with good semen characteristics. Nevertheless, this is a null result but an important one. I would suggest including a justification for the patient inclusion criteria, especially given the reduced sensitivity. Was a power calculation performed at all?

9) In Supplementary Table 3, the Haplotype coded 22, for total motility is reported as a coefficient=0, SE=0, P=1.0. Is this correct that there was no variance at all for 20 patients? In addition, Supplementary Table 4 is rather confusing because the coding of the SNPs is not clear. What do 0,1, and 2 represent? It is not in the Table description.

10) (Discussion, para 2) The binominal nomenclature for species in the alignment are not correct. Genus name is capitalized, species name is lower case. In the Figure 2 Maximum likelihood tree, some species are named, and the vernacular is used for others. The names (of species) should be used and be consistent throughout.

11) In Table 2, only 58 individuals were typed for the r22042792 SNP. Is this correct and there were some missing data? If so, this could be noted in the Table legend. (See my point #9) The SNP code in Supplementary Table 3 does not show missing genotypes, or describe the SNP data. Again, perhaps a nucleotide letter or letters would make more sense for a reader to interpret the haplotype data themselves.
12) In the table with phenotyping data (supplementary Table 1), the Fertility index (FI) values are extremely variable and the manuscript suggests those >100 were scored as 100. Is this correct? And if so, is this justifiable?

In my opinion this study warrants publication after minor changes, although I think the lack of fertile or high motility controls is a major weakness, especially when the experiment was, perhaps, unlikely to detect genetic associations with small effect sizes throughout.

**IMPORTANT NOTE**

The version of the manuscript I reviewed did not contain either Table 3 or Figure 3. I have not seen these and cannot comment on their agreement with the manuscript, or their content.

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

**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