Reviewer's report

Title: Estrogen and progesterone-related gene variants and colorectal cancer risk in women

Version: 1 Date: 4 April 2011

Reviewer: Deborah Thompson

Reviewer's report:

This is a competent study which has been well designed and analysed. The presentation is clear, thorough and logical. This type of candidate gene study has to a large extent been superseded by the advent of the GWAS era, and the authors acknowledge that none of the GWAS colorectal studies to date has found evidence of risk alleles in hormone-related genes. They therefore take a more detailed approach, looking at haplotypes and at the combined effects of alleles within a gene.

Minor Essential Revisions

1/ The OR quote in the abstract claims to be for carriers of at least one A allele, but seems to be the same (apart from the lower 95% CI) as the OR in table 3, which looks to be for the per-allele model. This needs to be clarified.

2/ Statistical Analysis, paragraph 1. Presumably the exclusion criteria was for SNPs with a MAF<5% ‘or’ which deviated from HWE, not ‘and’?

3/ Statistical Analysis, paragraph 4. The description of the set-based tests is unclear.

4/ Statistical Analysis, paragraph 6. It is not clear why it was necessary to perform permutation tests rather than the conventional asymptotic tests.

5/ Statistical Analysis, paragraph 7. A RR of 1.7 would be quite a large effect for a SNP – what would the power be for a more realistic effect size?

6/ Results, paragraph 1. The number of SNPs is given as 242, but 241 on the next page and in Table 1.

7/ Results, penultimate paragraph. The authors talk about the joint effect of rs1772453 and rs10883782. It would be very useful to know what the LD between these SNPs was and how many women were carriers for both SNPs.

8/ Discussion, paragraph 6, last line. This suggests that mutation breaks down LD, but the general model is that mutation creates LD, which is in turn destroyed by recombination. LD could be broken down if one or both SNPs was a recurrent mutation, although this would be very unusual.

9/ Table 3. A column containing the MAF of each SNP and an overall n could
usefully replace the numbers of each genotype. A footnote should clarify that the OR is for the per-allele model.

10/Table 4. The results for blocks 1-4 could optionally be moved to a supplementary table, as they add little to the main manuscript. The interesting results for block 5 are hard to interpret without knowing which SNPs they refer to e.g. it is clear that haplotype 4 is completely tagged by the first SNP, but it is only in the Discussion that we find out which SNP this is.

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.