Reviewer's report

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

Version: 1 Date: 21 February 2011

Reviewer: Sarah West

Reviewer's report:

This is an interesting article on a promising area with good functional reasons for a potential effect on CRC risk. The small sample size of this exploratory study has limited its power to detect an effect and it would be valuable to see similar studies in much larger cohorts, perhaps from the GWAS community to further investigate whether variants in sex hormone genes influence CRC risk.

1. Introduction, paragraph 2.

The identified susceptibility loci for CRC were detected in 3 genome-wide association studies. References 13, 15, 17 and 18 were not separate studies, but all from the same GWA study just with additional samples and meta-analyses added to improve power to detect an association.

Sentence 3 in this section is a confusing and needs changing to reflect the above comment. I suggest change to:

“To date, at least 3 phase-design genome-wide association (GWA) studies of colorectal cancer have been undertaken, which identified several novel susceptibility loci mapping to....”

3. There were 14 novel susceptibility loci detected, not 5.

Please add chr1q41, chr12q13.13, chr3q26.2 and chr20q13.33 (published by Houlston et al. 2010 in Nature Genetics 42(11) 973-7), and chr20p12.3, chr19q13.1, chr14q22.2 and chr16q22.1 (published by Houlston et al. 2008 in Nature Genetics 40(12) 1426-35)

Do any of these regions harbour genes involved in sex hormone synthesis or actions?

4. Statistical analysis, paragraph 2 – adjustment for age and hysterectomy status – what properties are being adjusted for by correcting for age? Also if hysterectomy status is an issue then affected samples should ideally have been removed or analysed separately. A comment on this in the text might be useful.

5. In Results, paragraph 3, there is a mistake with a SNP name, I think it should read: “When multiple comparisons were accounted for, only the CYP17A1 rs17724534 variant remained...”

6. Results, paragraph 6. It is worth mentioning that the number of African-American cases and controls included in the study is very small (at 58
cases and 116 controls) and that the lack of a statistically significant association is not surprising. The results cannot really be convincingly interpreted as showing that these SNPs show no association in this group of samples.

Discretionary Revisions

1. Statistical analysis, paragraph 1. Were SNPs excluded if they failed to have genotypes for more than 5% (or a certain threshold) of samples? Also what was the average genotyping call rate of the samples? 20% missing data seems quite high.

2. Results, paragraph 2 – in the comparison of baseline characteristics with CRC – alcohol consumption, family history and E and E-P therapy did not show statistically significant differences from the control population, but this is not hinted at in the text. Also, the smoking status and history of polyps is mentioned as showing no difference, however, there is a difference it is just not statistically significant...

3. It would be useful to state which part of Europe the samples were from - north European ancestry or south or a mix of the two - as allele frequency can vary between Northern and Southern Europe, as seen in the lactase persistence variant, and this could affect the results. Or was this corrected for?

4. Supplementary table – the chromosome and positions of the SNPs would be helpful in the table to facilitate comparisons.

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