Reviewer's report

Title: Common low-penetrance risk variants associated with breast cancer in Polish women

Version: 1 Date: 29 August 2013

Reviewer: Nick Orr

Reviewer's report:

1) In the background section you state that, “the majority of breast cancer cases are sporadic, but a substantial portion result from highly penetrant mutations”. These are conflicting statements – only a small proportion arise from highly penetrant mutations.

2) In the fourth sentence of the background section I think you should replace “disease development” with “disease susceptibility” or something similar.

3) The statement that carrying 14 susceptibility alleles increases the risk of breast cancer six-fold compared to carriers of the protective alleles is a bit misleading in this context since the absolute numbers are so small as to be irrelevant. The point Pharoah was making in his paper was that susceptibility alleles were of no value for personalised prediction, but could have value for population based screening.

4) Following on from my previous comment, I think you should make the introduction a lot more concise. The discussion on prediction and prevention has recently been covered in great depth (and better clarity) elsewhere.

5) In the first sentence, third paragraph of the background section I think you mean risk of cancer, rather than prevalence.

6) Could you clarify why your study is so enriched for early breast cancer? Similarly, were you surprised by the high frequency of BRCA1 mutations in your cases? Neither is particularly reflective of breast cancer in the general population.

7) SNP selection is puzzling. Even prior to the recent COGS papers there were more than 20 well validated breast cancer susceptibility loci. Why did you only analyse 11 of these. Also, why was the TOX3 variant rs3803662 not included? It confers one of the largest increased relative risks of breast cancer of the known loci.

8) Could you describe the control population in more detail please as I'm not familiar with the screening program to which you refer? How representative are the controls to the general population?

9) The QC criteria for sample and assay completion at < 10% are a bit too relaxed; 5% (or less) seems more appropriate.
10) Could you confirm what you mean by additive gene action model? I’m assuming you mean test for trend? It isn’t clear though. It feels as though you’ve thrown the kitchen sink at your dataset in terms of the models analysed. A CA trend test or similar would be sufficient (it’s robust to deviations from Hardy Weinberg proportions, unlike the allelic test presented here).

11) Correction for multiple testing doesn’t seem necessary (particularly if you lose some of the genetic models reported) given that these SNPs are mostly already confirmed susceptibility loci; p<0.05 is probably ok.

12) Why did you choose to type two sets of tightly correlated variants (in FGFR2 and BRCA2)? As you say, they don’t add much to your overall story (and you should acknowledge this in the paper – really you’ve just validated three loci, not four).

13) Presentation of the results tables could be improved dramatically, e.g. please order SNP by chromosome and location. Also, in supplemental table 3, there are multiple p-values greater than 1!

**Level of interest:** An article of limited interest

**Quality of written English:** Needs some language corrections before being published

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

**Declaration of competing interests:**

I declare that I have no competing interests