Author's response to reviews

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

Authors:

  Joanna K Ledwoń (joanna.ledwon@gmail.com)
  Ewa E Hennig (hennige@coi.waw.pl)
  Natalia Maryan (natalia@maryan.pl)
  Krzysztof Goryca (kgoryca@gmail.com)
  Dorota Nowakowska (donow@coi.waw.pl)
  Anna Niwińska (annaniwinska@gmail.com)
  Jerzy Ostrowski (jostrowski@warman.com.pl)

Version: 2 Date: 2 October 2013

Author's response to reviews: see over
Dear Editor,

We are submitting a revised manuscript entitled “Common low-penetrance risk variants associated with breast cancer in Polish women” (Ms: 2788241397930877). We were very pleased that the Reviewers found our manuscript of some interest. They made several helpful suggestions to improve presentation of our work which we have carefully considered and followed in revising this manuscript. Below, please find the specific responses to the Reviewers’ comments and the changes that were made in the manuscript. All significant corrections are indicated in blue all along the revised manuscript.

We believe that revising the manuscript along the lines suggested by the Reviewers strengthen this paper. We hope that you agree and now find the paper acceptable for publication in the BMC Cancer.

Yours sincerely,
Jerzy Ostrowski, MD, PhD,
Professor of Medicine

Note for Academic Editor:

As was recommended by one of the Reviewers, several statistical analyses were corrected or added in the revised manuscript. Such analyses required an involvement of one new person in the study, which is now added as a new co-author of the manuscript (underlined): Joanna K. Ledwoń, Ewa E. Hennig, Natalia Maryan, Krzysztof Goryca, Dorota Nowakowska, Anna Niwińska, Jerzy Ostrowski.

Responses to Reviewer 1: Sajjad Rafiq

Major revisions:
1. In the methods section the authors should provide sample size calculations.

As recommended, information about sample size calculation is added in the Methods part of the revised manuscript. Additionally, results of sample size calculation are provided in Additional Table 2 of the revised manuscript.

2. In the results section of the abstract, the authors need to provide the p-value which would suggest a potential replication based on the Bonferroni criteria. Also the p-values of SNPs which are specifically mentioned as to be associated with familial and sporadic cancer should be presented.

The relevant Bonferroni corrected $p$-value, $p_{corr}$ have been provided all along the Abstract, as was suggested by Reviewer.

Minor essential:
3. In the methods section the last sentence has an error, as the Bonferroni corrected p-value for 16 association tests cannot be 0.05. It has to be 0.05 divided by $16 = 3.1 \times 10^{-3}$. 


We agree with the Reviewer that for the Bonferroni corrected $p$-value, for 16 association tests, “raw” $p$-values (before correction) need to be 16 times lower. In our manuscript, the Bonferroni correction was implemented by multiplying “raw” $p$-values by 16 (number of tested SNPs) instead of lowering significance threshold. Consider columns “$p$-value_cor” in Table 4 of the manuscript. Only such $p$-value$_{cor}$ of less than 0.05 were consider as significant.

4. **In the results section the second paragraph the SNP selection can be moved to the methods section.** In an earlier section of the manuscript the authors mention that 35 SNPs are known to be associated with susceptibility of breast cancer. Are all of these 35 SNPs known to be associated from GWAs or have you included candidate gene study SNPs as well. Why were only 8 GWAs SNPs tested. Were these 8 the most strongly associated SNPs in European populations? Were the 5 SNPs chosen from previous candidate gene studies associated with susceptibility in any of the genome wide association studies. Or were any SNPs in linkage disequilibrium with these 5 SNPs associated with susceptibility at GWAs level of significance?

As suggested, all relevant information of SNP selection, mostly included in the second paragraph of the Result section, are presented in the Methods part of the revised manuscript.

According to database “A Catalog of Published Genome-Wide Association Studies” [www.genome.gov/gwastudies] provided by Hindorff L. et al, which includes only GWAS results, there were data on 36 SNPs associated with breast cancer, at the moment when we have started our manuscript preparation. Thus, such information was included in Background section of the manuscript.

When we began our study, the data on only about 20 GWAS SNPs was available, however, the most consistent associations were indicated for 15 SNPs representing 7 loci. We have chosen 8 tagging SNPs for confirmation (two SNPs represented $FGFR2$ loci). Additional 3 SNPs were selected based on well documented candidate gene study. Final 5 SNPs were relatively frequent among patients coming forward to Genetic Counseling of Cancer Center-Institute of Oncology in Warsaw; association of 2 of these SNPs were suggested in other previously published studies and remaining 3 were not documented so far.

None of 5 SNPs chosen from previous candidate studies associated with susceptibility in any of GWAS and also none of GWAS associated SNPs was in linkage disequilibrium with these 5 SNPs.

5. **In the results section in the second paragraph line 4; Is the sentence supposed to read “Additional five SNPs were selected based on the data”**?

Yes, the misspelling was corrected.

6. **In paragraph 3 line 1 of the results section the authors should mention exactly what they refer to as significant.**

As recommended, the “raw” $p$-value significance of less than 0.05 was mentioned in this paragraph of the revised manuscript. Further, the exact $p$-value$_{cor}$ was added in parallel with associated SNP.

7. **In paragraph 4 of the results section the authors mention that rs1219648 and rs1799944 were excluded from the dataset and then represented by tagging SNPs. Please advice how strongly (r$^2$ value) did these tagging SNPs tag rs1219648 and rs1799944.**
The exact $r^2$ values for tagging SNPs rs1219648 and rs17999444 were included in the Results part of the manuscript (paragraph 4). In the revised manuscript, these statements are included in paragraph 3 of the Results: “SNPs rs2981582 and rs766173 are in the same linkage disequilibrium (LD) blocks with rs1219648 ($r^2 = 0.967$) and rs1799944 ($r^2 = 1$), respectively”.

**Responses to Reviewer 2: Nick Orr**

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.*

As suggested, the sentence has been corrected.

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

The phrase was replaced.

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.*

We agree that all estimations on SNPs association with a disease have rather population than personal value. And, according to this and next recommendation of the Reviewer, relevant statement was removed from the revised manuscript, in order to make the Introduction part more concise.

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.*

We have carefully revised and shortened the Introduction part, in order to make it more concise. In the agreement with Reviewer suggestion, the discussion on prediction and prevention was removed from the revised manuscript.

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

Yes, the phrase has been changed.

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.*

In our study the blood samples were collected at the Department of Breast Cancer and Reconstructive Surgery or Genetic Counseling Unit from women with newly-diagnosed breast cancer, predominately exhibiting early age of onset. Especially among the Genetic Counseling Unit patients, those with positive family history were more frequent. According to our previous results, over 25% of BRCA1 mutation carriers with hereditary cancers were of less than 50 years at diagnosis [Gaj P et al, *Fam. Cancer* 2012, 11:623]. It was shown that
such enrichment of early onset and familial cancer cases, might additionally improve the power of the study [McCarthy MI et al, Nat. Rev. Genet. 2008, 9:356].

Our previous studies [Gaj P et al, Fam. Cancer 2012, 11:623; Brozek I et al, J Appl Genet 2011, 52:325] and also others [Malone KE et al, Cancer Res. 2006, 66:8297] indicated that the frequency of high risk mutations is higher in early onset and familial breast cancers than in older and with negative family history. However, in Poland, also in sporadic cancers, $BRCA1$ mutation was more frequently indicated before 50 years of age at diagnosis than after.

Although, almost equal number of familial and sporadic cancer cases was recruited to our study, substantial portion was under 50. Therefore, the high frequency of $BRCA1$ mutation in group of patients was not unexpected.

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.

Please consider the relevant explanation on 11 SNPs selection which is included in the respond to Reviewer 1, point 4 of this letter.

Actually, the TOX3 variant rs3803662 was one of selected SNPs (consider Table 3, position 14). However, in our study, it shows significant association ($p$-value < 0.05) only without Bonferroni correction (Table 4 of the revised manuscript).

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?

National Colorectal Cancer Screening Program offers screening colonoscopy for all Polish citizens aged 50 – 65 years and younger with family history of colorectal cancer. Control group was recruited predominantly from people who took part in the Screening Program in Warsaw and Szczecin in years 2003-2010. Women with normal results of mammography and colonoscopy, without personal history of cancer and family history of breast/ovarian cancer were chosen to our study.

9. **The QC criteria for sample and assay completion at < 10% are a bit too relaxed; 5% (or less) seems more appropriate.**

As recommended, we have changed QC criteria to more stringent, and took 5% in the analyses of data in the revised paper.

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).

By additive gene action model we mean the binominal logistic regression model (with each susceptibility allele modifying baseline risk by the same factor), fitted for single loci.

According to Reviewer suggestion, additive gene action model was replaced by the Cochran-Armitage test for trend in the revised manuscript.
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.

We agree that applied method of correction is conservative and may lead to some false negative results. Notwithstanding, we believe that correction for multiple testing should be applied regardless the status of tested SNPs. As genetic association studies in the context of breast cancer are conducted independently by many research teams, the 5% significance threshold may be too relaxed and may lead to propagating false positive results. The danger of false positive is especially high in environment where negative results are less likely to be published.

However, in Table 4 with the results of our analyses, all associations with at least 5% significance for uncorrected p-values are provided for concerned readers. Also, full results for all 16 SNPs in both allelic and CA trend tests are included in Additional Table 3.

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).

We certainly agree that three, not four, separate loci were validated in this study. In order to stress this finding, the appropriate statements were added into Abstract and Conclusions of the revised manuscript. Tagged sets of variants were used as kind of control of correctness of analyses.

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!

The presentation of SNPs in the results Table 4 was changed in relation to chromosome location, as recommended by the Reviewer. The Additional Table 3 replaced Additional Tables 2 and 3 in the revised manuscript. The results of recessive and dominant logistic regression analyses were removed. Instead, all results of allelic and CA trend tests are presented. All values greater than 1 were corrected.

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

All recommendations of the Reviewer on statistics were full filed. Mainly, QC criteria were changed to more stringent and additive test analysis was replaced by Cochran-Armitage test for trend.

15. Needs some language corrections before being published

We have carefully check the manuscript after revision in regard to correct the language.