Author's response to reviews

Title: SNP-SNP Interactions in Breast Cancer Susceptibility

Authors:

Venus U Onay (onay@mshri.on.ca)
Laurent Briollais (laurent@mshri.on.ca)
Julia A Knight (knight@mshri.on.ca)
Ellen Shi (Ellen.Shi@cancercare.on.ca)
Yuanyuan Wang (y32wang@uwaterloo.ca)
Sean Wells (sean@mshri.on.ca)
Hong Li (li@mshri.on.ca)
Isaac Rajendram (idr1978@hotmail.com)
Irene L Andrulis (andrulis@mshri.on.ca)
Hilmi Ozcelik (ozcelik@mshri.on.ca)

Version: 4 Date: 2 March 2006

Author's response to reviews: see over
March 2, 2006

BioMed Central Editorial Team

Re: MS1523829166824828
SNP-SNP Interactions in Breast Cancer Susceptibility

Dear Editor,

Please find in this letter the revisions of our manuscript MS1523829166824828 and our responses to the reviewers’ comments. First of all, we would like to express our appreciation for the input provided by the two reviewers, which helped us to considerably improve our manuscript. We feel strongly that we have tried to address all the issues raised, hopefully, to the satisfaction of the reviewers. All the changes are described in the letter and highlighted in the manuscript. We have also made substantial language corrections following the editor and reviewer’s recommendations.

Sincerely Yours,

Hilmi Ozcelik, Ph.D.
Mount Sinai Hospital,
600 University Ave., Rm 992A,
Toronto, Ontario M5G 1X5
Tel: 416-586-4996, Fax: 416-586-8869
e-mail: ozcelik@mshri.on.ca.
Reviewer: David Goldgar

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. There should be an additional table describing in more detail the significant interactions, in terms of the effect size for particular combinations of SNP 2-locus genotypes.

We have added the suggested table (Table 5, page 37) describing the effect size of each genotype combination for the four significant interactions in more detail. We feel that this table now better explains the nature of the four significant SNP-SNP interactions.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. It would be useful to add a column to table 1 or 2 describing any previous evidence for association of the SNPs with Breast Cancer in other (and perhaps larger) studies.

We have added a column to Table 1 (page 31), which describes the results from previous studies on the polymorphisms included in our study. Association studies with larger sample size were indicated with references and comments where applicable. Relevant footnotes were also added as “§ The studies that showed statistically significant overall SNP-disease risk associations were considered (N_{cases}>250 and N_{controls}>250). †Cases are women diagnosed with breast cancer before age 40”

2. Also the column in table 1 labeled 'Interacting SNPs' is not clear as to what is meant by interacting.

We have added a footnote to Table 1 (page 31) to better explain the term “interacting SNPs”, as “*No and Yes indicates whether the SNP has been shown to be “interacting” or “not interacting” with other SNPs in this study.”

3. On page 10, what sample size was used for the bootstrap samples?

We have indicated the sample size used for bootstrap analysis as “(each random sample comprising 398 cases and 372 controls)” on page 10 (line 20).

Discretionary Revisions (which the author can choose to ignore)

1. Table 4 might be better as the first table since it more or less describes the study population.
Former Table 4 has been expanded and presented earlier (Table 2 on page 32) as suggested.

2. It might be interesting to use the family history data to increase power for the main effects analysis using e.g., the method outlined in Thompson D, Witte J, Slattery M, Goldgar D. (2004) Increased power for case-control studies of single nucleotide polymorphisms through Incorporation of family history and genetic constraints. GeneticEpidemiology27:215-24.

This is an interesting method that we are planning to use in our future analyses. Instead of stratifying by family history, some of our analyses were also adjusted for family history. Our results remained unchanged after this adjustment.
Reviewer: Thomas A Sellers

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. The article has a significant number of statements that need to be supported with citations. Probably half of the statements in the background are undocumented, yet are stated with conviction of fact rather than opinion. Each sentence needs to be reviewed throughout the manuscript.

We have revised the entire background section as suggested by the reviewer, and added the relevant references to support our statements. We have done this by reviewing each sentence systematically and by modifying the text according to the reviewer’s suggestion (see pages 4-6).

2. The thesis presented would be much more compelling if the introduction chose to focus more on the genes and SNPs selected for investigation than the lengthy discourse on polygenic models. Indeed, interpretation of the results is difficult without such a background.

As suggested by the reviewer, the background section (from page 4 last paragraph to page 5) now focuses more on the genes and SNPs, and the strategy used to select them. To improve this, we have also added a column in Table 1 (page 31) to better describe the biological processes the genes are involved in.

3. The sampling strategy, subselection of participants from the larger study, and the control sampling is difficult to follow. It would be helpful to have Table 4 presented earlier, and with more information on the study subjects (demographics, clinical characteristics, more non-genetic risk factor data).

As also raised by the other reviewer, Table 4 has been expanded and presented earlier in the manuscript (Table 2 on page 32). To further address this issue, we have added the following section to page 7 (line 5) in Methods, Subject Population section, as “Therefore, the cases should better represent all cases without enrichment for genetic risk criteria such as family history. In Table 2, 21.6% of cases in the present study had a first-degree family history of breast cancer, which is consistent with the 17 to 22% frequency reported in cases in a number of large case-control studies [44-46].”

4. Table 3 results are impossible to interpret as presented. The significance tests do not permit interpretation of the nature of the interactions. Two rows include age interactions, which is not consistent with the title of the table. I don’t understand the two ER SNP interactions. Are they in linkage disequilibrium? A haplotype approach should be applied.
Our former Table 3 has been revised and is now presented as Table 4 on page 36. The age-SNP interactions were removed. We have also added information about the two ESR1 SNPs: linkage disequilibrium and haplotype analysis. The following explanations have been incorporated in the text:

i-Methods, page 10 (line 10) “We have also estimated the amount of linkage disequilibrium (LD) between the two ESR1 SNPs separated by about 140 kb on chromosome 6 and investigated their haplotype effect on breast cancer using the software “Unphased” from Dudbridge [53].”

ii-Results, page 14 (line 3) “The amount of LD between the two ESR1 SNPs was relatively small, with a D’ [59] of 0.07 in cases and 0.15 in controls, and none of the four haplotypes was significantly associated with breast cancer. Therefore, only the interaction effect between the two SNPs is presented.”

---

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. pg 4, line 1: Change “Genetic epidemiological” to “Genetic association” (traditional genetic epidemiology is gene mapping without consideration of function or relevance of the marker).

We thank to reviewer for this clarification. Due to changes in our introduction, the section including this expression has been removed.

2. A careful distinction between a mutation and a polymorphism must be made. Geneticists use differences in frequency, but the authors sometimes confuse that with functional effects of the DNA alteration.

We have described the term polymorphism in the background section, starting in the second paragraph on page 4 (line 9) as “SNPs have been historically classified as commonly occurring (>1%) genetic variation in the general population, whereas the rare variants with obvious functional consequences on the protein have been classified as mutations. Compared to mutations, SNPs have been perceived as functionally insignificant, however, current evidence emphasizes that a considerable fraction affects the intrinsic properties and the function of the proteins to a variable degree [7-9].”

3. pg 4, line 14. I would have expected this statement to be supported with references to BRCA variants. I don’t understand the point as intended.

As requested, references supporting contribution of BRCA mutations to breast cancer development have been provided in Background, page 4, line 2 as “The existence of dominant predisposition alleles/mutations, conferring a high breast cancer risk, has been confirmed with the discovery of BRCA1 and BRCA2 [1,2].”
4. pg 4, line 8. Do we really know this to be true?

Due to changes in the introduction, this sentence has been removed.

5. pg 5, line 2 from bottom. A brief description of the genetic risk criteria is necessary for the reader to interpret the results.

We have added a brief description of the genetic risk criteria (page 6, line 15) in the “Methods - Subject Population” section as “All respondents who met a defined set of genetic risk criteria (i.e., Ashkenazi Jewish; diagnosed before age 36 years; previous ovarian or breast diagnosis; one or more first- or two or more second-degree relatives with breast or ovarian cancer; one or more second- or third-degree relatives with either breast cancer diagnosed before age 36 years, ovarian cancer diagnosed before age 61 years, multiple breast or breast and ovarian primaries, or male breast cancer; three or more first-degree relatives with any combination of breast, ovarian, colon, prostate, or pancreatic cancer or sarcoma, with at least one diagnosis before age 51 years) were included in the study [43].”

6. pg 8. More detail is needed on the “control” SNPs. This is an unusual approach.

As suggested, we have revised the background significantly and have better explained our SNP selection strategy. The SNPs with no functional evidence was defined as “control” SNPs in the text. We have added the following explanation on page 5 (line 2) as “SNPs were initially selected from the best evidence from published studies in the beginning of the project, in year 2000, and subsequently classified under three categories (high-, medium- and low- rank), representing SNPs with a wide range of functional evidence. High-rank SNPs were supported by studies, which demonstrated the effect of the SNP on the regulation of expression or protein function. The medium-rank category is more likely to include functionally relevant SNPs, as the substitutions are predicted to significantly affect function, although this was not confirmed experimentally. This category also includes SNPs, which were associated with breast cancer risk factors. The low ranking category, on the other hand, contained SNPs with no functional information. Among the SNPs studied, XPD-[Lys751Gln], MTHFR-[Ala222Val], COMT-[Met108/158Val], GSTP1-[Ile105Val] and CCND1-[Pro241Pro], have been shown to alter the function or post-translational modification of their encoded protein [15-27]. MMP1-[1G(-1607)2G] and IL10-[G(-1082)A] have been shown to alter the transcription and expression of these genes [28-32]. IL13-[Arg130Gln] has been suggested to have functional consequences, while GSTM3-[4595 (3bp ins/del)] was predicted to create a YY1 transcription factor binding site [33,34]. The TNFA-[G(-308)A] forms a haplotype with some nearby SNPs and some studies observed increased haplotype dependent transcriptional activity change while some others do not [35-38]. CYP17-[C518T], and IL13-[Arg130Gln] were found to be associated with other cancer related variables, such as serum estrogen and IgE levels, respectively [39-41]. BARD1-[Pro24Ser] changes a structurally important non-polar proline residue to a positively charged serine. There were no functional speculations for ESR1-[Ser10Ser], ESR1-[Pro325Pro], PTEN-
[(IVS4+109)ins/delACTAA], IL1A-[Ala114Ser], G-CSF-[Leu185Leu] and GADD45- [C(IVS3+168)T. Thus, the 19 SNPs studied represent SNPs with a wide range of functional knowledge and evidence.”.

7. pg 11. SNPs don’t show evidence for HWE. It’s their distribution that is being tested.

We have added a sentence in the “Results” section, page 12 (line 2) as “None of the SNP distributions showed deviation from Hardy-Weinberg equilibrium in this sample.”

8. Table 1. The column heading “Interacting SNPs” is unclear. Does this relate to the author’s reference to “control” SNPs (the “no’s” in the column)?

We have added a footnote to Table 1 (page 31), describing the term “interacting SNPs” as “No and Yes indicates whether the SNP has been shown to be “interacting” or “not interacting” with other SNPs in this study.”

9. Table 2. Epidemiologic risk factors for breast cancer should be explored, and included when significant in the model. This is critical given the uncertainty as to the appropriateness of the various selection and matching algorithms that were applied.

Epidemiological risk factors have been explored and presented in Table 2 (page 33) and the following text was added:

i-Page 9 (line 2) “Several epidemiological risk factors were also assessed for association with breast cancer including age, BMI, education status, smoking status, family history, menopausal status, age at menarche, age at menopause, parity and age at first live birth (Table 2). Some of our analyses were also carried out adjusting the SNP main effect for the statistically significant epidemiological risk factors.”

ii- Page 10 (line 7) “All these analyses were also performed adjusting the interaction effects for the risk factors found to be associated with breast cancer risk at a significance level of 5% (BMI and family history).”

iii-Page 12 (line 3) “The distribution of selected epidemiologic risk factors in cases and controls is shown in Table 2. Cases and controls were similar with respect to the distribution of smoking status, menopausal status, age at menarche, age at menopause, parity, and age at first birth. Controls tended to have a higher BMI (p=0.05) and level of education (p=0.06) than cases. Cases were also more likely to have a positive family history of breast cancer than controls, and this difference was highly significant (p=<10^-5).”
Page 12 (line 14) “Our results remained unchanged when the models were also adjusted for BMI and family history.”

Page 13 (last paragraph) “The results of our multivariate analyses adjusted for age, BMI and family history confirmed the role of the most important interactions. The bootstrap $P$-values associated with XPD-[Lys751Gln] and IL10-[G(-1082)A], COMT-[Met108/158Val] and CCND1-[Pro241Pro], GSTP1-[Ile105Val] and COMT-[Met108/158Val], and BARD1-[Pro24Ser] and XPD-[Lys751Gln] were all significant (respectively, $P=0.014$, $P=0.020$, $P=0.022$ and $P=0.020$), however the significance of the tests decreased due to the high proportion of individuals missing BMI information. Therefore, only the analyses adjusted for age are presented in Table 4.”

Discretionary Revisions (which the author can choose to ignore)

1. pg 3, line 3 from bottom: “Although the majority of SNPS are ^silent and thought to be^ harmless, a considerable…” (unless a citation is included otherwise).

Due to the changes in the introduction, the mentioned sentence has been removed.

2. pg 7, line 4. This sentence should be broken into separate sentences.

This sentence was under the “Methods-SNP Selection Strategy” section. This section has now been moved to the “Background”. The mentioned sentence has also been changed due to the revision of this section.

3. pg 8, line 3 from bottom. Delete (it is redundant with first sentence in the section).

This sentence has been deleted.

4. pg 8, last sentence: “The ^reliability of the^ results were ^determined^…” (this isn’t really a validation).

We have revised the sentence as “The reliability of the results was determined by re-genotyping a randomly selected 10% portion of the total study population.”, as suggested by the reviewer (page 8, line 15).

5. Table 2. One could just give the additive model results for the three genotypes. The $p$-value isn’t important since the 95% CI’s are presented, but it would be nice to have the number of cases and controls in the table.

Table 3 in page 34 (previous Table 2) has been revised and the number of cases and controls was included. Only the results from the co-dominant model are now presented in this table, as suggested.
6. pg 12. Rather than talk about modes of transmission (which is appropriate to segregation of a trait), consider genetic risk model.

“Modes of transmission” in the text has been replaced by “genetic risk models” in pages 10 (line 5) and 12 (line 16).

7. pg 16, last paragraph. The point of this was lost upon this reader.

We have revised this point (page 17, line 7) as “SNP-SNP interactions in breast cancer development have been also reported in other studies, which targeted the SNPs of the carcinogen metabolism genes, including GSTM1, GSTT1, GSTP1, GSTM3 and CYPs [71-75].”