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.Shii@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: 3 Date: 17 February 2006

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

Please find the comments and revisions for the manuscript, MS1523829166824828 Version 2, in response to reviewer’s comments. First of all, we would like to express our appreciation for the input provided by the reviewers, which helped us to considerably improve the message of the manuscript. We feel strongly that we have tried to address all the issues raised, hopefully, to the satisfaction of the reviewers. We finished our corrections by the given deadline (i.e. February 16). All the changes are done are described in the letter and highlighted in the text.

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 an additional table (Table 5, page 37) describing the more detailed effect size of the significant interactions. We feel that the findings in this table have strengthened our message of 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, in page 31, to describe the previous studies reporting on the polymorphisms covered in our manuscript. Association studies with larger sample size were indicated with references and comments where applicable.
Relevant footnotes are 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 explanation to Table 1 (page 31), describing the term “interacting SNPs” as “* The SNPs that were found to have statistically significant interactions in this study. The majority of interacting SNPs were found in higher rank group.”

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 with 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.

As suggested, this table has been expanded and presented earlier as Table 2 on page 32.
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 significantly, as suggested by the reviewer, and added the relevant references to support the statements. We have done this by reviewing each sentence, modifying the text, as suggested by the reviewer (please 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, in the background section (from page 4 last paragraph to page 5), we have focused more on the introductory information on genes and SNPs, and their selection strategy. To improve this, we have also added a column in Table 1 (on page 31) to further describe the biological processes of the genes studied.

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 as Table 2 on page 32. To further describe this we have added the following sections to page 6 line 14 in 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].”

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.**

Former Table 3 is revised and presented as Table 4 on page 36. The age-SNP interactions were removed since they were not consistent with the title of the table. We have studied the two ESR1 SNPs for linkage disequilibrium and haplotype analysis. The following explanations have been incorporated in different sections of the manuscript.

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].”

Results, page 14, line 1” 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 clarification of the terminology. Due to the text change in the introduction (as suggested) the section including this terminology 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. They are frequently found in the genome, and a considerable fraction has been shown to affect 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 has been provided in Background in 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 the text change in the introduction (as suggested) the mentioned 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 on page 6, line 14 in Methods, under Subject Population 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]”. This description was also supported by reference [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, better explaining our SNP selection strategy. The description of the SNPs with no functional evidence was termed as “control” SNPs previously. We have added the following text in page 5 line 1 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 CyclinD1-[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 results, page 12, line 2 as “None of the SNP distributions showed evidence for 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 explanation to Table 1 in page 31, describing the term “interacting SNPs” as “* The SNPs that were found to have statistically significant interactions in this study. The majority of interacting SNPs were found in higher rank group;” The “no” and the “yes” indicates whether the SNP has been shown to be “interacting” or “not interacting” with others in the 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 as Table 2 on page 33, and the relevant text were added as following

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.”

Page 10, line 7 “All these analyses were also performed adjusting the interaction effects for the risk factors found significantly associated with breast cancer risk (BMI and family history).

Page 12, line 3 “The distribution of selected epidemiologic risk factors in cases and controls is shown in Table 2. Cases and controls were relatively similar in smoking status, menopausal status, age at menarche, age at menopause, parity, and age at first birth. Controls tended to have a higher BMI and also to be more educated than cases.”
The two groups differed significantly only on first-degree family history of breast cancer, 
\((p=<10^{-5})\), which was more likely to occur in cases as expected.”

Page 12, line 14” Our results remain unchanged when the models were also adjusted for 
BMI and family history.”

Page 13, last paragraph as “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 CyclinD1-[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.”

------------------------------------------------------------------------------------------------------------------------
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 text change in the introduction (as suggested) the mentioned sentence has been removed.

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

This section (SNP selection strategy) has been moved to introduction (as suggested) and 
due to the revision the structure of the mentioned sentence has been changed.

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

This sentence is 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 results were confirmed by re-genotyping a 
randomly selected 10% portion of the total study population.” on 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 (frequency) of cases and controls was included. Only the additive model results are given in the 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 the point on page 17 (line 4) 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 [72-76].”