Author’s response to reviews

Title: A Kallikrein 15 (KLK15) single nucleotide polymorphism located close to a novel exon is associated with poor ovarian cancer survival

Authors:

Jyotsna Batra (Jyotsna.batra@qut.edu.au)
Christina M Nagle (christina.nagle@qimr.edu.au)
Tracy OMara (t.omara@qut.edu.au)
Melanie Higgins (mhiggins@uvic.ca)
Ying Dong (y.dong@qut.edu.au)
Olivia L Tan (livandkhai@yahoo.co.nz)
Felicity Lose (felicity.lose@qimr.edu.au)
Lene Marie Skeie (lene@nettnyheter.net)
Srilakshmi Srinivasan (s1.srinivasan@qut.edu.au)
Kelly L Bolton (boltonk@mail.nih.gov)
Honglin Song (honglin@srl.cam.ac.uk)
Susan J Ramus (s.ramus@ucl.ac.uk)
Simon A Gayther (s.gayther@ucl.ac.uk)
Paul D P Pharoah (paul1@srl.cam.ac.uk)
Mary-Anne Kedda (MaryAnne.Kedda@pmintl.com)
Amanda B Spurdle (Amanda.Spurdle@qimr.edu.au)
Judith A Clements (j.clements@qut.edu.au)

Version: 3 Date: 23 December 2010

Author’s response to reviews:

Dear Editor,

Thank you for the opportunity to revise the manuscript “A Kallikrein 15 (KLK15) single nucleotide polymorphism located close to a novel exon shows evidence of association with poor ovarian cancer survival” by Batra et al. Please find the point-wise answers to the reviewer’s comments below, which helped us to improve our submission. We would also like to add that informed consent was taken from all the participating individuals and this statement is included in the revised manuscript. We sincerely hope that our m/s will now be acceptable in BMC cancer.

Reviewer: kathryn terry

Reviewer’s report:

This is a study of polymorphisms in the KLK15 gene identified through a variety of in silico tools to identify likely functional polymorphisms. In this effort the authors identified a novel exon likely involved in alternative splicing. 9 SNPs were selected to tag 22 likely functional SNPs and genotyped in 319 ovarian cancer
cases from a population based case control study and Royal Brisbane Hospital in Australia. The most promising SNP (rs266851) in the Australian data was genotyped in 1815 cases from multiple UK studies as well as 413 cases from the TCGA pilot project. The SNP was not significantly associated with ovarian cancer survival in these validation data sets. Strengths of this study includes the authors exhaustive attention to in silico modeling to identify the most promising SNPs in the KLK15 gene, control for age, stage, histologic subtype, and grade in the analysis, and genotyping in 3 independent data sets. However, there are several limitations of the study that need to be addressed.

Major Compulsory Revisions
1. Although the initial findings in the Australian data set are interesting for rs266851 the genotyping in the UK and TCGA data sets show no significant associations. Table 1 shows hazard ratio estimates from the UK GWAS (HR=1.07, 95CI% 0.94-1.24) and TCGA (HR=1.24, 95% CI=0.90-1.61) consistent with a null association and have confidence intervals that span 1. Granted these estimates go in the same direction as the original finding but could not be considered a “validation” of the original finding. With that said, the authors may want to point out that there could be differences in these populations that may have lead to different results including differences in histologic grade/subtype distributions, longer time to case ascertainment leading to a survival bias (that is, cases may be too sick to participate or have died before enrollment if the time between diagnosis and enrollment in the UK and TCGA data sets is longer than the Australian data set then a survival advantage for this SNP could be missed entirely in the UK and TCGA data sets). However, with the data presented here that is impossible to assess. Given these differences are not explored, and the UK and TCGA data sets are much larger than the original (and therefore better powered to detect a true association), the current data leads to a conclusion of no association. Consequently, the main conclusion of the paper and the title of the manuscript need to be revised.

Answer: The title and the result section have been revised to understate the findings as per the reviewer’s suggestion. We do not believe that the current data leads to a conclusion of no association, but we do openly acknowledge that findings need to be confirmed with additional replication sets with similar clinical characteristics. This has been stated in the conclusions. As per the reviewer’s suggestions, possible differences in the datasets has been explored and
described in the revised manuscript to argue the difference in magnitude of risk estimates observed in the UK and TCGA datasets. Please see the revised discussion section on page 16-17.

2. More detailed information about case ascertainment, including time between diagnosis and enrollment, should be added to the methods. Also, how were the 207 cases from the population based study selected? Presumably the original study includes far more than 207 cases. Similarly, how were the hospital cases identified... were these consecutive incident cases? If not, how were they selected?

Answer: Additional details on case ascertainment have been provided in the revised manuscript. Please see the revised ‘Study Participants and Genotyping’ section on page 5.

Minor Essential Revisions

3. The authors should include a discussion of the limitations of the study including selection of the cases, lack of validation, and possibility of survival bias.

Answer: The Discussion section has been revised as per the reviewer’s suggestions. Please see page 16-17 of the revised manuscript.

4. Residual disease is an important predictor of survival. If this data is available it should be added to the statistical analysis, if not this needs to be noted in the limitations.

Answer: Data on residual disease is not available and has been noted in the Study Participants and Genotyping section on page 6.

Discretionary Revisions

5. The statistical analysis description indicates that survival analyses were censored at Sept 1, 2004. Presumably, this date coincides with the end of follow up. It would be useful to update this analysis with more recent survival data if possible.

Answer: We did not follow up the patients after the censored date, so unfortunately, we don’t have the more recent censored data.

6. In the section describing the sequencing of cancer cell lines and aggressive cancer patients, were the authors looking at germline or somatic mutations in the aggressive cases?

Answer: We were looking at the germline sequence variations.

Reviewer: Majid Masso
Reviewer’s report:
Minor Essential Revisions:
1. Abstract: Identify HR as hazard ratio upon first use.
   Answer: HR has been defined as suggested.

2. Background: “…structural similarity with KLK3 (PSA) and is…” – no italics for KLK3 here, refers to the protein not the gene.
   Answer: Italics has been removed.

3. Several websites addresses for the tools in Supplementary Table 2 have changed, here are the new addresses for those that I could find:
   - mfold http://mfold.bioinfo.rpi.edu/
   - Dragon ERE finder ver. 2 (web address no longer exists, cannot locate new address)
   - Promoter Scan Version 1.7 http://www-bimas.cit.nih.gov/molbio/proscan/
   - CpG Islands http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html
   - NHR Scan http://www.cisreg.ca/cgi-bin/NHR-scan/nhr_scan.cgi
   - ESEfinder http://rulai.cshl.edu/tools/ESE/
   - miRanda
   Answer: Thank you very much for the information. We have updated the Supplementary table 2 accordingly.

4. Typo in Fig. 2A: below and to the right of Exon B, replace 56038167 with 56028167.
   Answer: It has been replaced. Thank you for correcting.

5. Results, at the beginning of the “in silico promoter analysis …” section: replace Cluster with Cister.
   Answer: We have corrected this error.

6. In 3 places reference is made to predictions obtained by using one or more software packages, but the actual names of those packages are not mentioned and should be included (web addresses are not necessary since I assume they are already included in Supplementary Table 2):
   a. Results, section “SNP information and prediction of functional effects”, subsection “Gene”: “We also predicted miRNA binding sites using three different software programs.” – either end sentence with a colon and list the programs, or if miRanda is the only one of the three in Supplementary Table 2, then end sentence with a comma followed by “one of which was miRanda.”
b. Same section as in a., subsection “Promoter”: second sentence, ending with “…using two different software (Fig. 2B).” – provide software package names.

c. Caption to Fig. 2B: “AREs (boxes), … as predicted by a single software.” – not only is the software name not provided, but on the surface this contradicts what is mentioned in the text (i.e., two different software packages were used from b.); however, closer inspection reveals that the AREs, EREs, and NHRBs were predicted with a single package (I have to assume Cister, but how can I know for sure), while the effect of SNPs on creating/destroying ARE/ERE sites were predicted using two unnamed packages. But this presentation is confusing at best and should be reworked for greater clarity, and it would help to include the names of the software packages in each case as the text is rewritten.

Answer: Software package names have been provided and clarified in the text.

7. I suggest moving Supplementary Table 6 (SNP selection for the survival studies) to the main body of the paper - it provides a good summary of the final list of SNPs obtained from the extensive bioinformatics work, and it serves as a starting point for the next part of the study. Within the main manuscript, two changes need to be made to Table 3: the heading “n dead” should be changed to “n censored”, and CT & TT under “Australian” should be changed to CT/TT for consistency.

Answer: Supplementary Table 6 (SNP selection for the survival studies) has been moved to the main body of the paper. Table 3 (Now Table 4) has been modified as per the suggestions.