Author's response to reviews

Title: Gene-expression patterns in peripheral blood classify familial breast cancer susceptibility

Authors:

Stephen R Piccolo (stephen_piccolo@byu.edu)
Irene L Andrulis (ANDRULIS@lunenfeld.ca)
Adam L Cohen (adam.cohen@hci.utah.edu)
Thomas Conner (tom.conner@hci.utah.edu)
Philip J Moos (Philip.Moos@pharm.utah.edu)
Avrum E Spira (aspira@bu.edu)
Saundra S Buys (saundra.buys@hci.utah.edu)
W E Johnson (wej@bu.edu)
Andrea H Bild (andreab@genetics.utah.edu)

Version: 3
Date: 29 September 2015

Author's response to reviews: see over
September 29, 2015

Dear Dr. Sands:

We thank you and the reviewers for thoroughly reviewing our manuscript entitled "Gene-expression patterns in peripheral blood classify familial breast cancer susceptibility" (#1944373553172151). The comments have been extremely helpful as we have revised the manuscript. We have addressed each of these comments carefully.

Below we provide a point-by-point reply to the reviewers’ comments. In addition, we have updated the article’s format to comply with this journal’s requirements. We are confident that you will now find it suitable for publication in BMC Medical Genomics.

Thank you for taking the time to review our work. Please contact us any time with any questions you may have.

Warmest regards,

Andrea H. Bild, PhD
Director, Genome Sciences Program
Department of Pharmacology & Toxicology
University of Utah
Email: andreab@genetics.utah.edu
Phone: 801-581-6353

W. Evan Johnson, PhD
Assistant Professor
Division of Computational Biomedicine
Boston University School of Medicine
Email: wej@bu.edu
Below we provide in-line responses to the reviewers. The reviewer’s comments are shown in gray. Our comments are shown in black.

Reviewer 1

Reviewer: Logan C Walker
Reviewer's report:
This study presents a new tool (biomarker) which has potential for estimating individual risk of breast cancer. Although important, further studies would be required to test efficacy.

We thank the reviewer for providing a careful review of our manuscript. We agree that further studies will be necessary to test the efficacy of our approach before it will be suitable for clinical use. In the Discussion, we now emphasize this point more strongly.

Major Compulsory Revisions:

1. Tables S1 and S2 - The markers appeared to be unable to predict 1) BRCA1/2 mutation status in patients with family history, and 2) patients with (or without) family history, which limits the clinical relevance of the study in relation to improving genetic screening.

Tables S1 and S2 correspond to the probabilistic predictions shown in Figure 1. Many women who carry BRCA1/2 mutations do not develop breast cancer. Accordingly, the goal of this study was not to predict BRCA1/2 mutation status but to evaluate how well we could differentiate between women who developed familial breast cancer and those who did not, whether or not those individuals carried a mutation in BRCA1 or BRCA2. As illustrated in Figures 1 and 3, our analysis resulted in accuracy (AUC) levels of 0.76 and 0.73 for the Utah and Ontario cohorts, respectively. To evaluate these predictions further, we used an ANOVA statistical test to compare the predictions among the various subgroups in our cohorts. As can be seen in Figure 1A, the predictions were significantly higher for BRCA1/2 and BRCAX women who developed cancer compared to the remaining subgroups. In several cases, differences between individual subgroups were statistically significant. For example, the predictions were significantly higher for BRCA1/2 women who developed breast cancer than for BRCA1/2 women who did not develop breast cancer in the Utah cohort. In contrast, there was not a significant difference between BRCA1/2 women who developed breast cancer and BRCAX women who developed breast cancer; this observation was in line with our expectation that we would see consistent gene-expression patterns among women who developed familial breast cancer, irrespective of BRCA1/2 status. In addition, we did not see significant differences between any of the control subgroups.

2. The mutation status is noted but not the type of mutations. Which patients had BRCA1 mutations and which had BRCA2 mutations?

We thank the reviewer for this question. We have added this information to Additional data file 4 for the Utah cohort.

3. Batch effects are significant when comparing expression profile in blood between individuals. The authors noted that the arrays had been processed at different facilities and had used a tool (ComBat) to minimise these effects. It would be beneficial to see the before and after effect of ComBat to the data. This could be presented as supplementary information.

We thank the reviewer for this suggestion. We have added Figure S1, which uses principal component analysis to show that strong batch effects existed before ComBat adjustment and that ComBat effectively corrected for these effects. Below are Figures S1A and S1B.
4. The relevance of the pathway analysis is unclear. Genes/pathways that are characteristic of the familial breast cancer group are being assessed in non-breast tissue. Are the authors implying that the 250 genes used for the biomarker are also characteristic of asymptomatic breast tissue and/or are playing an important role in breast cancer susceptibility?

Our biomarker analysis indicates a potential to use peripheral-blood gene-expression levels to distinguish between individuals who develop familial breast cancer and individuals who do not. We used a pathway-analysis to gain biological insight into the genes that enabled these predictions. Although expression patterns in peripheral blood surely differ from patterns that would be observed in asymptomatic breast tissue, we believe the peripheral-blood expression patterns provide insight into the biological processes that may alter familial breast cancer risk in individuals who develop familial breast cancer relative to controls. Previous studies have associated cell-adhesion perturbations and dysregulation of ECM components with breast tumorigenesis [1–10] (see references below); however, these pathways have not previously been linked to breast cancer susceptibility. We now emphasize this point more directly in the Results section. We are conducting an extensive follow-up study to examine this hypothesis using experimental and functional approaches in breast cells.


Minor Essential Revisions:

1. Abstract Methods - The training set should be referenced as the Utah cohort.

   **We thank the reviewer for noticing that this was unclear. We have updated the abstract to clarify that the training set was from Utah.**

2. Page 6 Modify the following into 2-3 sentences for clarity “Although such aberrant expression may not manifest itself phenotypically ...carry these mutations[14-26]”

   **We have clarified this text and separated it into 3 sentences.**

3. Suppl Data Files 1 and 2 need a key to define column headers

   To define these column headers, we have added an Additional data file (#3) that defines each of the clinical/demographic/treatment variables that are shown in Additional data files 1 and 2.

4. There are Supplementary Data Files 1 and 2, and Table S1 and S2, is this normal for BMC Med Genomics formatting.

   **We have changed the naming of these files to “Additional data files” to be consistent with the BMC formatting standards and to avoid confusion with the supplementary table names.**

Level of interest: An article of importance in its field
Quality of written English: Acceptable
Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.
Declaration of competing interests:
I declare that I have no competing interests

Again we thank the reviewer for these comments.
Reviewer: Lisa Hines
Reviewer's report:

This work explores the ability to use gene-expression levels in peripheral blood mononuclear cells as a means of breast cancer risk prediction among individuals who have a family history of breast cancer (FBC). In this study, they utilized two study populations: a training set of 124 women (63 controls and 61 cases) from Utah and a validation set of 73 women (37 controls and 36 cases) from Ontario, Canada. These data were additionally stratified according to BRCA1 and BRCA2 mutation status.

We thank the reviewer for providing a careful review of our manuscript. We note that the training (Utah) set included 39 cases (individuals who developed familial breast cancer) and 85 controls (individuals who had a family history of breast cancer but had not developed a tumor and individuals who did not have a family history of breast cancer). For the Ontario cohort, our study included 28 cases and 45 controls. These numbers can be obtained from Table 1.

Major:
1. One of the biggest limitations with this study is that it is retrospective blood samples were obtained after diagnosis and treatment. While these breast cancer patients may have completed their course of treatment, it is plausible that the differences that are being observed are still a consequence of treatment regimen. While the inclusion of women with sporadic breast cancer does provide some control of this confounding variable, tumor pathology and/or treatment regimens for women with a family history could have been more aggressive than sporadic cases. Were these groups comparable with respect to clinical/pathological characteristics and treatment regimen? The manuscript should provide summary tables that reflect risk factors and prognostic factors among the groups being compared for both the test and validation sets. The current Tables 1 and 2 provide no information regarding these datasets, other than sample size and median age. Supplemental Tables 1 and 2 do not provide summary data based according the groups being compared. There is no data for Ontario population. Was the Ontario population similar with respect to these factors when compared to the Utah population?

We thank the reviewer for this observation and agree that these are important issues to consider. As the reviewer mentions, the women with sporadic breast cancer helped to control for confounding effects due to treatment regimens and/or pathology. Due to the retrospective nature of this study, we do not have details about tumor pathology and treatment regimens for these patients. Thus it is possible that these factors may have differed between familial breast cancer cases and sporadic cases. However, we do provide extensive clinical, demographic, and health data for the subset of Utah samples for whom we were able to collect health-survey data. We identified genes whose expression correlated with these variables (see Gene-expression data filtering subsection). Thus, we performed extensive analysis and corrections for potential biases. These data are provided in raw and summarized forms in Additional data files 1 and 2. As the reviewer states, future studies will help to address these issues further.

With respect to Table 2, it appears that these comparison groups were not age-matched? This is another limitation. It appears that, overall, there were differences among these groups. The groups without breast cancer tended to be younger, so we cannot assume that they will not develop breast cancer by the time they reach the average age for the cases.

The reviewer asks whether we matched the patients by age. We did match patients by age. The previous version of the manuscript did not state this clearly enough, so we have added clarifying text in the Materials and methods section. Although the median ages for the non-cancer groups were younger in some cancers than the median ages for the cancer groups, we performed an analysis-of-variance test to ensure that the overall differences in age across all six groups was not statistically significant. In addition, because the goal of our paper was to demonstrate an ability to differentiate between individuals who developed familial breast cancer and individuals who did not, we also verified that there was not a significant difference in ages between these two groups (t-test p-value = 0.28). We have added these details to the manuscript.
2. Clarification and corrections are needed in the Results section, as well as Tables and Figures. There are many inconsistencies with respect to Figures and Tables (including Supplementary Tables) that are being referenced in the Results section, and Figure legends are not consistent with what is being illustrated in the Figure. Additional clarification and interpretation should be made with respect to what is being depicted in the Figure, i.e. define genomic model score, do the lines depict 95% confidence intervals, are their differences according to BRCA status (Figures 2C and 2D), etc.

We apologize that we inverted the figure numbers for Figures 2 and 3. For this reason, the figure legends were inconsistent with what was shown in the figures. In our revised submission, we have fixed the figure numbers. In addition, we have updated the captions for all of the tables and figures to improve clarity. We now define the genomic model scores and provided additional explanations on how to interpret the graphs. We also clarify what the dotted lines represent in Figures 3C and 3D.

The Results Section could provide more clarity when interpreting data results. For example, it states that similar levels of accuracy were obtained for women with and without BRCA, but Figures 2C and 2D reflect differences that are not discussed. Was any comparison done to see if the gene-expression profile was any better at predicting risk when compared with existing models for estimating risk based solely on personal health history and demographics? This would be an important point to address.

We agree that it would be interesting to compare our model against existing models that use personal health history and demographics. Due to the retrospective nature of our study and limitations to our IRB protocol, we are unable to address this question for the present data set. However, now that various risk models have been reported in the literature (for example, Wacholder, et al. 2010, Xu, et al. 2013, Sawyer, et al. 2012), we feel that it would be useful to perform a systematic comparison across these models and to compare them directly against the widely used Gail model (Gail, et al. 1989).

3. The Discussion section should address the limitations of this study. The sample size is actually quite small once subdivided into the different groups, with 23 being the largest subgroup. The Discussion states that the sample size was large enough to obtain statistically meaningful results, yet the data illustrated in Figure 1 appear to reflect very large confidence intervals (assuming these do in fact depict 95% CIs). This small retrospective study provides interesting suggestive data, but prospective studies are imperative to assess potential as a risk prediction tool.

The boxes in Figure 1 represent quartile boundaries for the probabilistic predictions that our model produced. When the Discussion section mentions confidence intervals, it refers to the bootstrap-derived confidence intervals that are described in the Results section. These accuracy levels suggest that our approach has potential to predict familial breast cancer status at statistically meaningful accuracy levels. However, we agree that prospective studies will be imperative for evaluating whether our model can be applied in clinical settings. We now state this more strongly in the Discussion section of manuscript.

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Acceptable
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests: I declare that I have no competing interests

Again we thank the reviewer for these comments and for a thorough review.