Reviewer's report

Title: Identification of candidate breast cancer predisposing variants by performing whole exome sequencing on index patients from BRCA1 and BRCA2-negative breast cancer families

Version: 1  Date: 16 Dec 2018

Reviewer: Melissa Southey

Reviewer's report:

Shahi et al., report the findings of a small case-control study of genetic breast cancer predisposition that has applied whole exome sequencing. Sharing/publication of all data generated in this pursuit is valuable but I have a number of concerns about the way the authors have presented this data that reduces its value in its current form. The whole exome sequencing scope of the study is essentially lost to a more specific and hypothesis driven analysis of very much small group of genes.

Abstract

Please define "rare" in the abstract.

Please substantiate the claim that nonsense, frame-shift indels and splice-site variants observed in BC cases with the genes of the panel have the highest probability of predispose to BC?

How were these variants further validated?

What does the expression "especially functional" mean? There does not seem to be any functional characterisation of these variants presented?

BRCA1/2 should be written as BRCA1 and BRCA2 throughout the manuscript.

There is contradictory information about what genes are breast cancer predisposition genes e.g.- should RINT1 be included in this category in the abstract without further explanation?

What does "vaguely" mean in this context? - a more scientific description is required here.

Is it possible to investigate the frequency of these variants in appropriate publically available data?

Introduction.

Please justify the information provided about penetrance of pathogenic variants in PALB2.
Please provide further information about common genetic variants that are associated with very small increments in risk - eg which ones are associated with 2-fold increased risk (this seems to be an overstatement)?

Please update the figures presented about the proportion of the familial risk that the currently identified genetic risk factors for breast cancer account for - this figure is much closer to 50%.

Line 34 - what does "weakly or strongly" mean on Page 4.

Lines 46-49 - Please substantiate the claim that nonsense, frame-shift indels and splice-site variants observed in BC cases with the genes of the panel have the highest probability to predispose to BC?

Methods

Line 53 page 5 "including genes that are not included in the Easton et al list" should be deleted.

Line 11, page 6, please replace "vaguely" with a scientific expression.

Results

Line 15, page 10 please delete "markedly".

Line 20, page 15. After adjustment this result is not statistically significant. The focus on this result throughout the remainder of the paper appears unjustified (although the lack of significance is acknowledged).

Discussion.

Line 41, page 11, increasing the sample size may not enable this analysis to reach statistical significance. The current statement should be deleted or edited - "Although a larger samples size is required to reach statistical significance"…

Please explain how the following was calculated and how the statement can be justified "the likelihood of those nonsense variants found in the genes of the CAGP in BC cases to be implicated in the molecular mechanism modulating BC risk is more than 50%."
The discussion is extremely long and arduous to read. This information is better presented in tabulated form so the text can be substantially reduced to the essence of the findings. Care needs to be taken to present the data accurately in terms of evidence for cancer and breast cancer predisposition eg Line 60, page 15 - what does "low impact mean"? Care should also be taken to ensure that the reader understands if the data being cited in from germline or somatic analyses. Care should also be taken with the use of the word "polymorphism" in this context.

The conclusion should be re-written with attention to the above suggestions.

**Ae the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

No

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Yes

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

No

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I recommend additional statistical review

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable
Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal