Author’s response to reviews

Title: The BRCA2 variant c.68-7T>A is associated with breast cancer

Authors:

Pål Møller (moller.pal@gmail.com)
Eivind Hovig (ehovig@ifi.uio.no)

Version: 1 Date: 04 Aug 2017

Author’s response to reviews:

General comments to referee comments:

We thank the referees for to-the-point comments that our data indicate an association between the variant and breast cancer, and that the sum of available data is conflicting with respect to whether or not the variant is pathogenic. This is indeed why we wrote the report, i.e. to make our observations known. The referee comments focus the underlying problem that current classification systems is not considering degree of pathogenicity (penetrance), but only the probability for any variant to have high penetrance, which we do discuss in the report. Going deeper reading all the underlying statistical papers to our references 14 and 15, one will find that they are based on the assumption that any genetic variant is either pathogenic with high penetrance for cancer, or normal without increased risk for cancer, and that there is nothing in-between. Moreover, the algorithm is based on Bayesian logic and tolerates neither zero nor NULL (unknown) as an argument. A very low a priori value will be a probability trap not to get out of, for which reason no argument may be very low, zero or NULL, and any low penetrance variant will be expected to give conflicting result, because it will not meet any of the two underlying assumptions for the categorization. The mathematical system for classification of pathogenicity theoretically does not demonstrate low and intermediate penetrant variants, confer to that for MMR genes, there is an obvious shortage of disease-causing demonstrated PMS2 variants. As discussed at the Variant interpretation committee at the InSiGHT meeting in Florence in this July, the current classification system (reference 15 in the paper) does not identify most disease-associated PMS2 variants as pathogenic, and this is partly so also for MSH6. If it were to be true that the variant in question has low penetrance, it would explain the conflicting reports. Also, the internationally recognized prevalence cut-off point for wild-type versus mutant alleles is 0.01 (0.02 heterozygote prevalence), and this has been used as cut-off point for pathogenic variants in the references mentioned. Any low-penetrant allele not strongly subjected to Darwinistic elimination may have much higher prevalence. The variant in question has a world-wide prevalence of 0.002, and also the higher Finnish prevalence is compatible with high penetrance. The population prevalence is no argument against neither high nor low penetrance, but it makes high penetrance less probable. Compliant with these argument, we have
not concluded the variant to be ‘pathogenic’ which to us infers high penetrance, but rather we describe an association. We decided when writing the report not to dig deep in these problems (to separate our presentation of results from a deeper theoretical discussion needing more arguments than our report of one variant), but in response to the referees’ comments we now include a few more lines on these.

Reviewer #1: This manuscript is interesting for familial breast cancer where authors describe that a splice variant of BRCA2 (c.68-7T>A) is associated with breast cancer. However, I am not convinced with pathogenic nature of this variant. Reasons are several fold as

a) CADD score of this variant is 3.7 and conservation scores are very low (phastCons 0.514, GERP++ 0.355 and phyloP 0.456).

b) Additionally there is no evidence of impact by SNPeff.

c) Vlinvar database is also predicting it to be begin

Clinvar: phenotype not specified likely benign ClinVar
RCV000168529.2
Breast-ovarian cancer, familial 2 conflicting interpretations of pathogenicity ClinVar
RCV000077384.7
Familial cancer of breast benign ClinVar
RCV000074550.4
Hereditary breast and ovarian cancer syndrome conflicting interpretations of pathogenicity

d) It also appears to be common variant for European populations like allele frequencies are European (Finnish) - 0.006566 and European (Non-Finnish) - 0.002366. I assume a familial variant to be rare in the population like allele frequencies below 0.001. If not requires proper explanation.

As you see completely reverse results are existing databases and by bioinformatics tools for pathogenic nature of this variant.
Response: See above. Existing data are to us not completely reverse to our finding, but did prompt us to examine all our information and write the report. None of the findings mentioned by the referee are, however, conclusive arguments against the variant being associated with breast cancer. To the opposite, all arguments mentioned are to us in keeping with a variant with intermediate penetrance and/or co-operating with variants in other genes to cause increased risk for breast cancer as concluded.

I recommend some experiments to confirm the totally opposite results.

Response:

In silico analyses suggested variant splicing, and this has been confirmed in vitro. What more experiments to undertake is unclear to us, and we have no such at hand.

I will like to the improvements as I am supporter of familial cancer genetics, but given facts improvements are required.

Reviewer #2: This paper examines the potential pathogenicity of a specific intronic BRCA2 variant previously identified in HBOC families and more recently at a frequency in population controls that raises concerns over the previous classification as pathogenic. The authors report a high frequency of the variant in their clinic based series of HBOC families and go on to examine the segregation of the variant with phenotype in 3 families as well as noting the incidence of 2 further breast cancers in prospective follow up of 24 carriers. Overall the data presented are supportive of a causative role - particularly the very high frequency of the variant compared to relevant population controls - but the limited numbers of carriers included in the analyses as well as the relatively simplified approach to the analysis significantly limits the strength of the conclusions. I agree that the overall conclusion, that the data provides evidence of an association with breast cancer, of unknown strength at this time, is correct but have questions about the analyses presented:

Segregation analysis - why are only 3 of 18 pedigrees included in the analysis?

Response:

All 18 families were included, and the results were as described. We expanded the first family as described and concluded the variant was pathogenic. The later families are as they are from our clinical activity where we passively await which family members on their own initiative come for testing without being directly invited, and lack of informative meiosis in this context is as expected. Going back to these families to extend them for segregation analyses up to 20 years
after the clinical service given will, if at all possible when former generations and cancer cases in recent years are dead, require time and money and is not doable within reasonable time. Our report may be read as a request to anyone with suitable families to do so.

Acknowledging that it is the author's own described method, why use a significantly simplified segregation method that requires pruning of most of the pedigree data when other robust options such as the full likelihood method, or the co-segregation likelihood method are capable of managing this and are now openly available and widely supported by the literature?

Response:

See introduction above. Any form of such more sophisticated segregation analysis would be assumptions instead of testing to get results based on retrospective data non-randomly selected, and would to us not give any definite proof. Specifically, segregation analyses without testing all family members for additional disease-causing variants and doing so in highly selected breast cancer kindreds only, may give misleading results. We have many kindreds with more than one variant causing breast cancer. Also, and theoretically, segregation analysis is not to us a valid method to determine penetrance when most breast cancers in the population are phenocopies, and this will merge with the probability of in-mating in small genetic isolates as confounders in highly selected families. Segregation analysis gives proof when a variant has high penetrance and there are few phenocopies in the population, but this is not the case in the present study. In short, segregation analyses do indeed indicate disease-causing variants, but should not be considered confirmative evidence in families where phenocopies in the population overwhelmingly outnumber those genetically caused, which will be the case in families that are not fully tested. Thus, we noted in the report that segregation analysis supported our conclusion, but did not prove it.

Prospective follow-up / annual incidence: again this is a very simplified analysis on small numbers. At least a standardized incidence ratio is needed to account for the age distribution of the carriers in follow-up. Even so the width of the confidence interval around the annual incidence figure means that the strictly correct interpretation of this figure is that it provides no significant additional support for pathogenicity at this stage.

Response:

We are familiar with numbers needed to prove penetrance by follow-up (https://www.ncbi.nlm.nih.gov/pubmed/28754778, https://www.ncbi.nlm.nih.gov/pubmed/22320316) and numbers allowed nothing but collapsing all observations irrespective of age for the reports. We gave the point estimate of events divided
by observation years as annual incidence rate with the corresponding 95% confidence interval, and compared it in the text with references. To us, numbers did not allow more sophisticated statics. Our report may be read as an invitation to others with follow-up data to add them ours to arrive at firm conclusions.

Other queries: what was the overall prevalence of pathogenic BRCA1 and BRCA2 mutations in the 714 families included in this analysis? At 2.5% was the c.68-7 variant the most frequent variant detected?

Response:

The report describes the prevalence of the variant in families not having another pathogenic variant for breast cancer. We apologize that this information by mistake was deleted in the manuscript submitted. Overview of the most prevalent Norwegian pathogenic variants is available at https://www.ncbi.nlm.nih.gov/pubmed/17574839. The large number of families detected to have pathogenic BRCA1/2 variants were demonstrated by direct testing for these variants and not sequenced. Prevalence of the variant in question in these families is not obtainable from our series.

With only 18 index cases/families found to be carriers it would be useful to include a table summarising their clinical features, particularly in comparison to the cohort as a whole i.e. age of cancer dx, history of ovarian cancer, bilateral breast cancer etc.

Response:

All families were subjected to segregation analysis and the results were as described. A ‘family’ is an ill-defined object, because it has no delineation (how distant relatives to include, and which lineages?), and has to us limited value if not starting with a method on how to describe all families the same way. We report results of clinical activity having no such underlying method, and being dependent of family members opting for testing. Describing the families as is would to us not meet scientific standards.

It should be noted in the text that the population frequency figures provided come from 1000Genomes and the ExAC database respectively as the properties of these datasets are well known.

Response:
We have now made ExAC explicit in the text, referencing Gnomad, which is the incarnation of Exac used. The 1000Genomes reference is to a 1000Genomes of Norwegian genomes, which is listed as a reference.

Did the ExAC figures quoted have the data from TCGA removed?

Response: Yes. Gnomad has these removed, and contains higher numbers of genomes.

Please review for language and sentence structure e.g. para 4: electronical (is not a real word)

Response: Done.