Reviewers report

Title: Association of PALB2 sequence variants with the risk of familial and early-onset breast cancer in a South-American population.

Version: 3 Date: 13 October 2014

Reviewer: Dirce Carraro

Reviewers report:

The manuscript is about the identification of genetic variants in PALB2 gene in Chilean breast cancer patients. PALB2 (partner and localizer of BRCA2) is a new gene recently characterized as associated to familial Breast cancer. The protein coded by this gene interacts with BRCA2 and BRCA1 proteins and plays crucial role for key BRCA1/2 genome caretaker functions. Germline loss-of-function mutations in PALB2 gene lead to Fanconis anemia, at biallelic status, while monoallelic status mutations are associated with higher risk of breast cancer (BC). The frequency of loss-of-function mutations varies among different populations. Based on that, the argument of this manuscript is the lack of data about the real contribution of PALB2 mutations to familial breast cancer in the South American population, including Chile.

In an attempt to contribute with this issue, in this study Leyton and collaborators screened the coding regions and exon-intron boundaries of the PALB2 gene in 100 Chilean BRCA1/2-negative women with familial BC. They did not find any loss-of-function mutation, but found three missense variants c.1676A>C (exon 4), c.1861C>A (exon 5), and c.2993C>T (exon 9). The authors investigated the variants in 436 breast cancer samples and 809 healthy Chilean used as Control group. They evaluated the distribution of allele frequency in both groups for estimating risk for breast cancer.

The topic of the manuscript is actual and important and the question posed by author is very clear and authentic. However, some weak points were identified and have to be clarified.

Major Compulsory Revisions

Screening of coding sequence and intron/exon boundaries of PALB2 gene:

The authors used a conformational sensitive gel electrophoresis method for detecting DNA sequence variation (single-base changes or small insertions and deletions) in the 100 patients and found only three variants. This is a customized method based on heteroduplex that should be minimally described. The complete sequence of the sample used as reference in the assay should be informed as well as the transcript reference sequence used for variants descriptions. Additionally as the sensitivity of the method depends on fragment length besides other factors, authors should mention the number and size of amplicons. I also would like to see the list of primers used for PALB2 complete screening and the authors should present it as supplementary material or
mention in the manuscript that the list would be available under requesting. Anyway, minimal information about the primer designing should be stated in the manuscript. The author, without given any reason, presented two primer sequences designed for exon 5. Additionally the reference presented by authors that was supposed to give details about the method seems to be incorrect.

References:
Surprisingly the largest study regarding the risk of developing breast cancer for women who are PALB2 loss of function mutation carriers (Antoniou et al., N Engl J Med. 2014 Aug 7;371(6):497-506) was not referred in this manuscript. It is important not only to refer it but also discuss considering the current result of their manuscript.

In studies of genetic association using case-control design is important to match in both groups the frequency of ethnic and geographic origin of individuals, besides age and socioeconomic strata as described in manuscript, in order to avoid confounding effects that may lead to false positive association. In this sense, it is important to present information that certifies that these aspects were controlled or the author should comment the limitation of the study design that can implicate in the positive association identified. Additionally, conclusion in abstract and in the end of the manuscript should be revised taking this issue into consideration.

Bioinformatic analysis of PALB2 variants
I would recommend changing “bioinformatic analysis” (line 259, 331 and..) for “Prediction of functional effects of PALB2 variants” or something similar. Additionally, terms such as “PolyPhen software indicated” should be avoided and replaced by “PolyPhen software classified … or predicted…”. All analysis using these softwares are not definitive and should be used with caution. In fact, with exception of the variant 1676A>G, which is found in high frequency in dbSNP (around 14%), both additional variants c.1861C>A and 2993C>T could be classified as variant of unknown significance, since no robust evidence was provided yet to prove that they are pathogenic or not. Additionally, the fact that the variant c.1861C>A was detected only in BC group does not exclude the possibility to be only a rare variant (as stated in lines 263 and 264). Taking this into consideration it is recommended that all paragraph (lines 260 to 276) be rewritten.

Minor Essential Revisions
To prevent confusion regarding the intrinsic meaning of terms such as polymorphism (a non disease-causing alteration) and mutations (a disease-causing alteration), I recommend using the word “genetic variant or allelic variant” instead “polymorphism” as cited in lines 205, 206, 317, 324, etc.

Line 200: … “variation database of PALB2 gene” instead “database of PALB2 polymorphism and mutation”

Table 4: To facilitate the comprehension of the composite Genotype, I suggest to
add in the first column the number of variants in dbSNP (rs152451-rs455551636)

**Level of interest:** An article whose findings are important to those with closely related research interests

**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