Reviewer's report

Title: Genetic pathways influencing breast size are shared with breast cancer

Version: 1 Date: 17 January 2012

Reviewer: Ryan Abo

Reviewer's report:

Genetic pathways influencing breast size are shared with breast cancer

The authors present a study investigating the genetic factors underlying variation in breast size in a European ancestry cohort of 11,423 women using a genome wide association study. From their study, they identified four single-nucleotide polymorphisms (SNPs) near four genes that were significantly associated with breast size.

The authors present interesting associations that appear to add knowledge regarding the genetic factors underlying breast size and possibly breast cancer risk, although, there are a number of major and minor points to address.

Major Compulsory Revisions:

1. The introduction is not very thorough or salient in presenting the background literature regarding the significance of breast size as a phenotype for studying and how breast size may be linked to breast cancer.
   a. There is an intuitive indirect association between breast cup size and breast cancer risk, which is that women with larger breast cup size are most likely obese and at a higher risk for breast cancer. The introduction mentions a correlation between obesity and breast size and that obesity has been shown to play a role in breast cancer risk, but does expand upon what other studies have found or hypothesized based on their findings.
   b. Specific details from the cited works are lacking
      i. What is the heritability estimate of breast size and how much is estimated to be driven by obesity from Wade et al. 2010.
      ii. Have there been any other genetic studies for breast size or have there been any breast cancer genetic studies with breast size as a measured covariate?
      iii. What is the relationship between lean women and breast cancer from Kusano et al. 2006 and Egan et al. 1999? Do these relationships contribute to any of the hypothesized meanings from the discussed findings?

2. Study design is limited with no mention of relevant covariates that could possibly influence the results. What other covariates were measured that are relevant for this study? These should be listed with summary statistics in a table or supplemental data. What is the correlation between BMI and breast size, age and breast size, breast cancer status and breast size, and menopausal status and breast size. A lot of potentially important covariates are not even mentioned in
the method section. If breast cancer status was collected for these subjects, why not determine and report the association and effect between these top findings for breast cancer risk in your data set? This seems to be logical, especially if your conclusion is that these variants have a role in breast cancer risk.

3. Given the three different genotyping platforms for the subjects, why not use imputation methods to obtain a common set of SNPs to perform your analysis? Even with a single platform imputation would be a standard procedure to determine any other variants of interest within the regions of significant association.

4. Figure 1 was not present.

5. The title of the manuscript is misleading. There are four significantly associated SNPs presented near genes, and to abstract these findings to genetic pathways seems to be an overstatement. The SNP near ESR1 is an interesting finding and has a clear association with breast cancer due to its role in encoding an estrogen receptor. The SNP “near MDM2” is actually quite far from MDM2 with a number of genes between the SNP and MDM2. The SNP 140kb from INHBB is also quite far from the gene, and the link between rs7816345 and ZNF703 is tenuous. Linking these SNPs to the genes is not very well done other than by pure location, and the role of these genes in specific “pathways” is not well presented.

Minor essential revisions:
1. Discussion of methods for controlling for covariates for the top findings and presentation of the figures in the Discussion may be better placed in the Methods and Results sections, respectively.
2. The following sentence was confusing: “The average breast size among the 1.8% of our cohort carrying zero of the “larger size” alleles is about…”.
3. Were any of the findings below statistical significance in or around any known breast cancer genes?
4. Did you conduct any pathway or gene set analyses?

Discretionary revisions:
1. More stringent SNP quality control thresholds that are used as standards: SNP call rate < 95%, MAF < 5%, HWE p-value < 10E-3
2. Report all findings above a p-value cutoff.

**Level of interest:** An article of importance in its field

**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.