Author's response to reviews

Title: Genetic variants influencing breast size also influence breast cancer risk

Authors:

Nicholas Eriksson (nick@23andme.com)
Geoffrey M Benton (gbenton@23andme.com)
Chuong B Do (cdo@23andme.com)
Amy K Kiefer (akiefer@23andme.com)
Joanna L Mountain (joanna@23andme.com)
David A Hinds (dhinds@23andme.com)
Uta Francke (ufrancke@23andme.com)
Joyce Y Tung (joyce@23andme.com)

Version: 2 Date: 23 March 2012

Author's response to reviews: see over
Dear Editor and Reviewers,

Thank you for your comments. In response, we have updated the manuscript significantly, as follows:

- Between submission and revision, we greatly expanded the size of the cohort. It now contains 16,175 women, as opposed to 11,423 before.

- We have imputed our genotype data against a recent 1000 genomes dataset.

- As covariates in this analysis, we have included age, 5 principal components, bra band size (in inches), and indicator variables for breast augmentation surgery, breast reduction surgery, mastectomy, ever pregnant, and currently pregnant or breastfeeding.

- As a consequence of these changes (almost entirely due to the increased sample size), the set of associated SNPs has changed. Now there are 6 regions with a genome-wide significant association, one of which has two independent significant SNPs. Of the four SNPs reported in the previous version, three remain genome-wide significant, the fourth (near MDM2) is no longer significant (p=4.6e-6).

- There are now three SNPs that were previously associated with breast cancer that show association with breast size in this study. We have included a table of 28 previously identified breast cancer SNPs and their p-values for breast size. We believe this result substantially strengthens our claim that breast size and breast cancer share genetic factors in common.

Detailed responses to your comments follow.

**Comments from editor**

- The title of the paper is misleading, because the identified variants are located quite far away from the reported breast cancer genes and no pathways were studied in this paper. Consider the following title - “A possible overlap between genetic regions associated with breast size and breast cancer”.

- We have changed the title to “Genetic variants influencing breast size also influence breast cancer risk.” We believe that the additional results (with 3 SNPs associated with breast cancer also associated with breast size) make the link between breast cancer SNPs and breast size SNPs direct. We have removed references to pathways throughout the paper.

- the genetic analysis is not sufficiently described. Why there was no imputation done to combine the data between these platforms?

- We have now performed imputation and detailed the genetic analysis.

- It is obvious that breast size is dependent on age and BMI. It was stated that these 4 SNPs explain 2% of breast size, while addition of age and BMI adds it up to 20%. It is clear that these factors have to be included into the model as they have stronger affects than SNPs? please provide updated results in table 1.
In this revised manuscript, we have included as covariates age, 5 principal components, bra band size (in inches), and indicator variables for breast augmentation surgery, breast reduction surgery, mastectomy, ever pregnant, and currently pregnant or breastfeeding. We used bra band size as a proxy for BMI due to good correlation (about 0.5) between the measures and much greater completion rate for band size (nearly 100% compared to about 50% for BMI). None of the SNPs we report in Table 1 are associated with BMI in other studies. We have included a short discussion of BMI-associated SNPs in the paper (next to last paragraph before Conclusions).

It looks like PCA analysis has been done but it should be better described. Were all the samples considered of European ancestry (based on PCA) and included into analysis?

Yes, all samples were of European ancestry and all samples had PCA results included. We have provided a discussion of this in the Methods as well as citations to earlier papers using our database that discuss the selection of European individuals [Eriksson et al., 2010, Tung et al., 2011] and the PCA calculation [Eriksson et al., 2010] in much more detail.

Provide number of samples for each of the association results, as it’s unclear how many samples had info of age, BMI and corresponding genotypes.

This is now included in the Methods

How is the HWE threshold of 10E20 justified? This is only possible in case of severe population stratification or technical issues; the completion rate of 90% is also quite low.

In our imputation procedure, we have updated the completion rate to 95%. The HWE threshold was chosen because these p-values were computed using our entire (European) database, thus they were based on about 100,000 people. The p-value scales quadratically with sample size, so this is equivalent to a cutoff of 1e-5 for 25,000 people, for example. We have added this point to the methods.

In a separate table for each of associated variants provide distance from the breast cancer genes (translation start site and location upstream or downstream) and LD measures (D’ and r2) and distances for the corresponding/reported breast cancer variants.

We have been much more explicit about the links between each variant and breast cancer variants in the text. We have not included D’ information as we didn’t feel it was that informative beyond r2 for our examples.

Realistically, it is understandable that this is not a specific well-designed study and it’s hard to significantly improve it. However, all limitations have to be clearly discussed in the paper?technical issues with 3 genotyping platforms and no imputation attempts; incomplete genotyping rate; possible population stratification issues and how they were taken care of in this analysis; factors affecting breast size (such as age, BMI, number of pregnancies and lactations, breast augmentation, etc.) and their inclusion and non-inclusion into the model.

We have expanded the methods section greatly to take into account these points.

The possible overlap with breast cancer regions is interesting and should be discussed but
with explanations that these are still just very wide gene regions and the functional variants responsible for one or both traits still have to be identified and validated by multiple methods.

We have tried to make this distinction more clear, e.g., “the others are near genes with links to breast cancer and development” in the Introduction and “The other associations we have found are near genes involved in other aspects of breast cancer and estrogen pathways.” in the Conclusion.

If the results hold after re-analysis with inclusion of age and BMI, it should be clear that these are just very preliminary results and no connection could be made between breast size and breast cancer risk without additional genetic fine mapping, epidemiological data and molecular studies.

We have added some clarification of this issue to the Abstract and Conclusion, e.g., “While these results do not directly support any possible epidemiological relationships between breast size and cancer, …” and “It should be noted that the shared relationships between breast size and breast cancer at these three regions are not strong enough to drive the possible epidemiological connection that has been reported elsewhere between breast size and breast cancer.”

**Reviewer 1**

Major Compulsory Revisions:

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.

- 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.

- Specific details from the cited works are lacking

  - What is the heritability estimate of breast size and how much is estimated to be driven by obesity from Wade et al. 2010.
  - 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?
  - 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?

The introduction has been greatly expanded in these directions. There is now a more detailed discussion about the relationship between breast size and cancer as well as a statement of the differences in risk between early and late obesity. Heritability numbers are
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 import 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.

As stated above, in response to this, we have included as covariates age, 5 principal components, bra band size (in inches), and indicator variables for breast augmentation surgery, breast reduction surgery, mastectomy, ever pregnant, and currently pregnant or breastfeeding. We used bra band size as a proxy for BMI due to good correlation between the measures (about 0.5) and much greater completion rate for band size (nearly 100% compared to about 50% for BMI). Breast cancer status was collected, however the number of women with breast cancer is small enough that there is not sufficient power to detect associations with breast cancer. Therefore we instead searched the literature for breast cancer associations to test against breast size, finding that three SNPs are associated with both. We have included this data in Table 2.

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.

Agreed: we have now imputed all subjects using 1000 genomes data.

Figure 1 was not present.

It should be present in this upload

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.

We have changed the title to “Genetic variants influencing breast size also influence breast cancer risk”, see further explanations above.

Minor essential revisions:

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.

We have added a discussion about covariates to the Methods. The figures are now called out in the “Results and Discussion” section.

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”.

This has been rewritten to be more clear and now reads: “We used these 7 SNPs to compute a genetic propensity score for breast size by counting the number of alleles associated with larger size that each participant carried. The average cup size among women in the top 5% of this score (e.g., women carrying 9 or more of the 14 possible ‘large’ alleles) was 0.83 cup sizes bigger (5.39 versus 4.56) than the average cup size among women in the bottom 5% of this score (women carrying 4 or fewer ‘large’ alleles).”

Were any of the findings below statistical significance in or around any known breast cancer genes?

There is now one association that is not quite significant that is near a breast cancer gene - CHEK2. We have added a table of breast cancer SNPs to further address this point.

Did you conduct any pathway or gene set analyses?

No, thus we have removed references to “pathways” in the title and the paper at large.

Discretionary revisions:

More stringent SNP quality control thresholds that are used as standards: SNP call rate <95%, MAF < 5%, HWE p-value < 10E-3

We have adopted the 95% call rate threshold in our imputation pipeline. We have left the MAF threshold to be 0.001, although all SNPs reported are over 5% MAF. The HWE threshold was chosen to be so small because the p-values are calculated in a very large group (nearly 100,000 people, see discussion above).

Report all findings above a p-value cutoff.

We have added a supplemental table with all SNPs under 1e-4.

Reviewer 2

Major compulsory revisions:

In addition to cup size, the study collected information on breast cancer status, age height, weight, and bra band size. The authors are missing some details on study inclusion and exclusion. There is an unstated assumption that none of the women included in this study were pregnant or breast-feeding or had breast augmentation or reductions. If this information was collected and individuals were excluded from study for any of these (or other) reasons this should be made clear. If this information was not available then the authors should mention these possible confounders to breast size.
We have included all women who reported breast size (and were European and unrelated). We have changed the methods so that we correct for these covariates (the full list of covariates is now age, 5 principal components, bra band size (in inches), and indicator variables for breast augmentation surgery, breast reduction surgery, mastectomy, ever pregnant, and currently pregnant or breastfeeding. There is now an extended discussion of this topic in the Methods.

There does not appear to be much of an additive effect of the four SNPs given a 0.93 cup size difference between ladies with none versus 5-8 of the “large” snps . The authors should comment on this.

The numbers have changed slightly, so that women with 9 or more “large” SNPs have 0.83 cup sizes bigger than those with 4 or less. The average effect size for the SNPs is about 0.15 cup sizes per allele, so the additive model would imply that a difference of 5+ alleles would correspond to a little over 0.75 cup sizes, which is consistent with what we see.

What was the result of the rs9397435 SNP and cup size in the people who were not typed for the rs9397436 SNP?

We have addressed this by imputation; now there are no missing genotypes.

Minor essential revisions:

The authors should include a description of what the * mean in Figure 1.

These figures have been revised to only use one symbol (the various symbols corresponded to different annotations of SNPs, e.g., conserved or non-synonymous).

Discretionary revisions:

The authors allude to previous studies about breast size and breast cancer risk. They should include information in their population if there was any evidence of breast cancer risk based on breast size and if this associates with any of the four most significant SNPs in this study.

We only have about 500 women with breast cancer in our database, so we are not well powered to detect associations with breast cancer. However, we have added a table of other breast cancer SNPs and their associations (or lack thereof) with breast size.

References
