Author's response to reviews

Title: Genetic variants associated with 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: 3 Date: 8 May 2012

Author's response to reviews:

Thank you for your suggestions. We have made the requested changes; details are below.

Associate Editor:

> This is a significantly improved paper and the authors have provided good answers. However, there are still several points that should be addressed. The title suggests causative relationships between these signals, breast cancer risk and breast size, however genetic association does not necessarily mean 'influence'. Please substitute 'influencing' by 'associated with'.

Done.

> This is a giant work and a possible example for future papers. Thus, the authors have to be extra careful and transparent with what they do and how they present the results and interpretations. For example, the cross-platform imputation is a serious topic by itself and I trust your technical capabilities of doing so but some quality metric should be provided or referred to if it was reported by you before.

We have added some statistics on the cross-platform imputation to the methods and the tables: "Imputation quality was slightly higher overall for the batches using the denser platform (average $r^2$ of 0.91 versus 0.87). "
In the tables, we now report a pair of $r^2$ values for the imputation quality on the less dense genotyping platform and for the more dense platform, denoted as $r^2_{V2}$ and $r^2_{V3}$ respectively.

While there are a few SNPs that are imputed to higher quality using the dense platform (for example rs17625845 near INHBB has $r^2_{V2}$ of 0.72 $r^2_{V3}$ of 0.99), for all SNPs reported, both $r^2$ measures are at least 0.73.

The discussion has to include more on limitations of this study. Specifically, it has to be mentioned that most of the results reported here are based on imputed variants. There is known genetic diversity between individuals of European ancestry. Even though the imputation quality seems to be good for these markers, it’s strongly dependent on LD pattern in 1000 genomes (CEU), which includes a very small sample set compared to the set you infer the genotypes for. Usually, it is taken care of by genotyping of all relevant imputed variants but in your case it's understandably impossible. Nevertheless, this has to be discussed, as the temptation to publish results based on sole imputation is very high.

We have added some discussion of limitations:

"We note that the estimation of breast volume via self-reported bra size is likely to be far from perfect. Thus, it would be interesting to see what effects the SNPs found here would have in a more exactly phenotyped population.

Likewise, many of the SNPs reported here were only imputed and not directly typed. While the estimated $r^2$ values are generally quite high, indicating good imputation quality, ideally these SNPs would be directly typed in a replication cohort."

Please include HWE p-values for all markers you report in the tables. These values for individual markers got to be good, no matter your genome-wide HWE threshold.

We have added HWE p-values calculated from rounded dosages. All p-values are at least 0.01 for reported SNPs.
> There is still mention of 'shared pathways' in the abstract.

Removed and replaced with: "These results provide insight into the genetic factors underlying normal breast development and show that some of these factors are shared with breast cancer."

Reviewer 2:

> In this revised manuscript they do not outline study limitations, in particular, the weakness of using self-reported information which may not be as accurate as other pieces of information. The authors should include a sentence or two on study limitations and how this may affect their results.

We have added a discussion of limitations (pasted above).