Title: Association of common variants identified by recent genome-wide association studies with obesity in Chinese children: a case control study

Version: 1  Date: 25 November 2015

Reviewer: Robert Sladek

Reviewer’s report:

Hai-Jun Wang et al., Association of common variants identified by recent genome-wide association studies with obesity in Chinese children: a case control study.

Overall, this is a conservative, thorough and clearly-written study of the impact of genetic variation on obesity in Chinese children.

I have suggested some modifications to the manuscript (points 4 and 7) that the authors could make at their discretion; and have identified some typographical and grammatical errors that should be corrected (point 10).

Detailed comments follow:

1. Is the question posed by the authors well defined?

Yes. The authors wished to determine whether SNPs associated with BMI or obesity in large GWAS of Caucasians and Asians were individually associated with, or had collective value for predicting obesity in Chinese children.

2. Are the methods appropriate and well described?

Yes. The authors have identified 40 obesity related loci, based on a review of genome-wide association studies that were conducted mainly using Caucasian adults. In six cases, genetically linked proxy SNPs were chosen to allow a multiplexed assay design. Association was tested in 2030 unrelated Chinese children, who were previously recruited in two cross-sectional studies performed in Beijing. Obese or overweight participants were identified using age- and sex-specific BMI percentiles. Logistic and linear regression was used to identify SNPs associated with the weight classification and BMI standard deviation score, respectively. Of the 32 SNPs that passed quality control, 6 showed nominal association (P<0.05) with childhood obesity, of which 2 showed association at a more stringent threshold reflecting multiple testing (P<0.00156).

My major concern with the methods used in this study surrounds the SNP selection, which is based on studies published in 2010 and 2012. Since then, the GIANT consortium has gone on to identify 56 additional loci associated with obesity (Locke, Nature. 2015 Feb 12;518(7538):197-206). Since this more than doubles the number of obesity-related loci, the paper would be stronger if these
loci had been tested.

3. Are the data sound?

Yes. The methods used for data cleaning are clearly described as are statistical approaches used to detect confounders (including study group), heterogeneity and bias. The rate of genotyping failure seems appropriate for the genotyping technology (Sequenom MassARRAY) and data cleaning procedure.

4. Do the figures appear to be genuine, i.e. without evidence of manipulation?

Yes. The figures, tables and supplemental tables are clear and provide a full description of the study results (including results for the SNPs excluded from the analysis for technical reasons). As minor points, the authors should consider redrawing Figure 1 to better separate the confidence intervals seen in the initial and present study (either by using colors that are more easily distinguishable or displacing the overlapping CI whiskers and centroids). It's also hard to identify the SNPs that were used to assess directional consistency of effects. Finally, it would also help to add summary statistics to the body or legend of Figure 2.

5. Does the manuscript adhere to the relevant standards for reporting and data deposition?

Yes. The summary level data is fully reported.

6. Are the discussion and conclusions well balanced and adequately supported by the data?

Yes. The authors have taken a conservative approach in their discussion and conclusions.

7. Are limitations of the work clearly stated?

The main limitation of this work is the sample size used for validation. As the authors point out, only 6 of the 42 previously associated SNPs showed association in their population sample of 2030 individuals. Differences in ethnic origin could account for the discrepancy; and, in fact, 8 of the SNPs were monomorphic or low-frequency (MAF<1%) in this Chinese population and 3 SNPs showed differences in minor allele frequency compared to Europeans (FST>0.10). However, none of the four SNPs that were initially associated with in East Asians achieved significance in this study; and 23 of the remaining 26 SNPs showed the same direction of effect on obesity as reported in the original study (as well as a dosing effect on obesity), suggesting that the failure to individually validate these SNPs could be better explained by either the younger age of the study sample or lower statistical power. The authors discuss this well in the paper; but I think that the discussion should be expanded to note that none of the SNPs previously associated with obesity in East Asians, showed association in this study (this information can be confirmed by digging around in the supplemental tables, but should be mentioned in the text). I would also add a
sentence to explain the implications of the heterogeneity of effect sizes that was seen between the current and the discovery studies (Page 16, line 9). It's probably also worth noting that none of the proxy SNPs showed significant association.

8. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
Yes.

9. Do the title and abstract accurately convey what has been found?
Yes. The abstract clearly defines the source of the candidate SNPs, the study's sample population, the major results of the paper and their implications.

10. Is the writing acceptable?
Yes. The paper is well-organized and clearly written.

There are a few typographical errors:
page 5, line 12 "identified 18 new loci associated with BMI at a ..."
page 7, line 12 "were approved by the ethics ..."
page 9, line 3 "Multiplex SNP assays designs failed for 6 out of 40 SNPs (rs10968576, rs4771122, rs10150332, rs12444979, rs29941 and rs261967); these were replaced by ..."
page 9, line 9 "(CEU) for the other 5 SNPs)."
page 9, line 14 "group, 31 oof the 32 SNPs showed ..."
page 9, line 21 "representing the effect of population ..."
page 10, line 12 "weight the risk alleles ..."
page 10, line 16 "that weighting the risk alleles ..."
page 15, line 22: "our study would be underpowered to detect ..."
page 16, line 18 "consequently, reduced statistical power."
page 16, line 18 "by the number of loci tested, which is growing as new genome-wide meta-analyses are conducted ...

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

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**
I declare that I have no competing interests.