Manuscript: *MGST2 WNT2* are candidate genes for comitant strabismus susceptibility in Japanese patients

Authors: Zhang J, Matsuo T

Summary: The authors perform association and linkage analyses of Comitant strabismus, a multi-factorial disorder, in families of Japanese ancestry. The statistical designed considered are case-control, TDT-based designs (trios), and linkage analysis in larger pedigrees. It is rare to read a manuscript where the authors state up front the challenges they face when dealing with quality assurance issues such as genotype misclassification errors. I think the reviewers are to be commended for this approach.

I list a number of questions observed in the manuscript, and based on those questions, suggest some modifications the authors may wish to consider.

Questions/Comments

1. Based on text lines 43-56, it appears that Strabismus is not a single phenotype, but a collection of phenotypes. If so, have the authors or others considered stratified analyses with "more genetic" subtypes?
2. Case-control association study (line 107, lines 111-112). Given all the other statistically analyses performed, the fact that the cases and controls were not independent, and the sample sizes were small, I recommend that this design be removed from the manuscript. It is most likely under-powered, and creates a greater multiple testing problem.
3. Lines 115-143, 184-208. In general, the TDTae analysis is used as a confirmatory analysis (although it can be used as a primary statistic). I would recommend not cleaning any of the data, running PLINK on the raw data (it will automatically filter out Mendelian inconsistencies) and then running TDTae with the DSB model. One can combine the P-values of the two methods by means of Fisher's combined p-value test. If the signal is real, one will see significant results in both the TDTae and TDT statistics. If the result is an artifact of genotype misclassification, the TDTae statistic will not be significant, and Fisher's combined p-value will most likely be non-significant. That is, the regions of interest for follow-up will be those where the TDT and TDTae are simultaneously significant for a given marker/set of markers. This point may also be reinforced through a modification of Figure 3.
4. Lines 209-227. It is worth noting that genotyping errors cause inflation in the recombination fraction between disease and marker loci (see, e.g., Lincoln and Lander 1992, Buetow, K. Am J Hum Genet 1991).
5. In general, there is no discussion of multiple test correction; the authors have performed multiple tests, and based on my read, have not applied any sort of correction like False-Discovery-Rate.

Suggestions:

1. As a suggestion, I recommend the authors performing some form of prinicipal components analysis, to obtain a single phenotype from the multiple phenotypes that have been considered. I noticed that Dr. Jurg Ott performed a number of the statistical analyses. Also, I know that he has worked in this methodological area for some time, so it would be auspicious to approach him about this topic.
2. As I mentioned above, remove the case-control design and results. I don't think it adds anything with the sample sizes given and the fact that subjects are not independent.
3. I would combine the TDT and TDTae results as mentioned above. In this way, with the Fisher's p-value, you get a single p-value per SNP rather than multiple ones. Technically speaking, Fisher's test is applied to independent statistics, but I think one can argue that the data sets will be sufficiently different, given that TDTae works on all the data, and the TDT on only the Mendelian consistent data.
4. I would add these references and a line acknowledging that recombination fractions may seem larger than they truly are.
5. The authors can use any of a number of multiple test correction methods for all p-value results from TDT/TDTae and linkage studies. In this way, they can state whether any SNPs are significantly associated after correction for multiple testing.