Reviewer's report

Title: 118 SNPs of Folate-Related Genes and Risks of Selected Congenital Anomalies

Version: 1 Date: 10 February 2009

Reviewer: Anne Parle-McDermott

Reviewer's report:

This manuscript by Shaw et al., sought to test a number of polymorphisms for risk of spina bifida and conotruncal heart defects in a Californian population. Their candidate gene, case-control approach consisted of 13 genes involved in the metabolism or transport of folate. The list of candidate genes is not novel; many if not all have been examined previously in relation to spina bifida risk (some in relation to conotruncal heart defects) with often conflicting results. Although, some of the variants showed an association with spina bifida, the overall conclusions indicate lack of a strong association of any of the genes with risk of either malformation. However, the genetic risk associated with these malformations has not been addressed extensively and thus, appropriately designed association studies, yielding positive or negative results should be published. I have a number of issues with the manuscript as follows:

Major compulsory revisions

1. Population stratification: the breakdown of the ethnic groups as described in Table 2 is a major cause for concern. There is a significant difference in the percentage of individuals of White Hispanic versus non-Hispanic in the case group compared to the control group. Allele frequencies that simply vary between ethnic groups and are unrelated to these malformations may produce a false positive association in such a study design. Although the authors do indicate that they adjusted for this in their logistic regression analysis more detail in relation to how they adjusted for this should be provided.

2. Abstract & Main Text: Results- a number of SNPs from the same gene are listed as having significant associations: the level of linkage disequilibrium between markers from the same gene should be included i.e, D' and r^2 values. Are they separate risks from the same gene or are they simply acting as markers of each other? The positive results should be put into the context of LD patterns of the gene itself- this is most important.

3. The genomic DNA samples were extracted from dried blood spots on filter paper and subsequently amplified by whole genome amplification (WGA). The authors should provide some details into relation to the WGA procedure in terms of ensuring both copies of each gene are efficiently amplified.

4. Statistical correction: given the large number of comparisons performed some form of correction should be applied. The authors should address this or justify why they feel it unnecessary to correct the P-values. In addition, P-values are not
included anywhere in the manuscript even though significance can be inferred from the OR. P-values for all significant associations should be included.

5. Table 1: the information contained within Table 1 would be easier for the reader to access if presented in the form of a diagram with each gene represented separately (probably more appropriate in a supplementary document). Each diagram should include the location of each polymorphism from the 5' to 3' end of the gene. Also Table 1 is not actually referred to in the text.

6. Table 4: also very long. This information could be contained within a supplementary section or simply ‘data not shown’, except for significant associations. Also details of how exactly the haplotype associations were computed should be included. ‘Identified blocks were assessed with odds ratios’ is too vague. A permutation analysis using Phase 2.1.1. software is more commonly used to test for haplotype risk (Stephens & Donnelly, Am J Hum Genet, 73(5): 1162-1169, 2003.)

7. Discussion: acknowledgement that full gene coverage for each gene was not achieved with this study design i.e., enough markers genotyped across the entire gene to ensure all variants are captured.

Minor Essential Revisions

1. The manuscript would benefit from a review of the structure of some of the sentences and the overall ‘flow’ of the information. I’ve highlighted a couple of sentences:

   • Abstract background: should probably read ‘Folic acid taken in early pregnancy….’
   • ‘Eligible were live born infants only…….’ ??
   • ‘Few odds ratios (Ors) revealed sizable departures from 1.0.’

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

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

Yes, hold a joint patent in relation to a polymorphism within one of the folate genes mentioned in the manuscript.