Reviewer's report

Title: Polymorphisms in folate-metabolizing genes, chromosome damage, and risk of Down syndrome: identification of key factors using artificial neural networks

Version: 1 Date: 19 May 2010

Reviewer: Yian Chen

Reviewer's report:

• Major Compulsory Revisions

  1. The authors emphasized that “traditional methods” is not good for detection of complex relationships several times in the paper, and mentioned in the result section that “linear correlation index between input variables and the target variable was generally very low, with the exception of BNMN”. Linear correlation is not an appropriate traditional statistical test for this research question under this context to predict a binary outcome (MDS vs. control mothers). It is also mentioned that “these networks determine a noteworthy improvement, compared with traditional methods of analysis” (page 4). To support these claims, the authors should at least use appropriate traditional methods of analyses, such as logistic regression.

  2. How were the 32 MDS mothers and 29 controls mothers selected? Are they age matched, ethnicity matched or…? The sample size is very small. If MDS and control mothers are from different ethnic groups, the results on polymorphism could largely reflect the ethnic differences.

  3. How was the genetic data coded for the analysis? For instance, is MTHFR677C>T coded as 0, 1, and 2 for CC, CT and TT, respectively? Or is it coded as binary (0 and 1)? If it is coded as 0 and 1, which model is used? (CC, and CT as 1 and TT as 0 or CT and TT as one group…)? The genetic coding is not mentioned in the manuscript. Is dominant, recessive, or log-additive model considered? This will have an impact on the numeric results and also on the interpretation of the results regardless of which analysis method (ANN or logistic regression) is used.

• Minor Essential Revisions

  1. Although we will get results using ANN for a small dataset like this, it does not mean the results necessarily would be generalizable to a larger population. An independent dataset in the future to validation the current finding is essential.

  2. Currently, 5x2 cross-validation was used so that multiple models were developed and validated during this process. In the future, when an independent dataset (with the same independent variables) is available, which model will be used for the validation?

  3. What are the ethnic backgrounds of the MDS and control mothers? The
haplotypes and LD structures between the SNPs could potentially be used in the model development. LD information is currently only discussed in the discussion while the relationship between SNPs could be also dealt with before putting them into the models.

**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.