Reviewer's report

Title: Genome wide association for substance dependence: convergent results from epidemiologic and research volunteer samples

Version: 1 Date: 1 October 2008

Reviewer: Ian Everall

Reviewer's report:

Genome wide association for substance dependence: convergent results from epidemiologic and research volunteer samples

Catherine Johnson, Tomas Drgon, Qing-Rong Liu, Ping-Wu Zhang, Donna Walther, Chuan-Yun Li, James Anthony, Yulan Ding, William Eaton and George R Uhl.

The authors have undertaken a genome wide association study for substance dependence in two cohorts: the Baltimore epidemiological catchment area study (ECA) and the Baltimore Molecular Neurobiology Branch (MNB) study. Both samples identified SNPs in 172 genes, which were unlikely to be due to chance, which differed between substance dependent versus control subjects. While the findings are novel and interesting the reviewer’s enthusiasm was dampened by the somewhat confusing narrative of the manuscript. The concerns include:

1. In describing the ECA sample in the introduction the authors state that the sample size is 3,481 from whom DNA was sampled in 1,071 subjects. However in the methods section the authors mention 662 European-American respondents whom 85%, which equates to a sample of 550 subject provided DNA. I am not sure what the sample included.

2. Similarly, the MNB sample is described in the background as recruiting 3,800 volunteers whereas in the methods it is described as 34 pools of 400 “abusers” and 280 “controls. So again there is confusion as to what constitutes this cohort. In addition abuse is different from dependence, what are the authors examining as the manuscript states dependence but the MNB is comprised of abusers.

3. In the methods it is stated that DNA is prepared from blood or cell lines but in the description of the cohorts it is stated that DNA samples were obtained. Were the DNA samples obtained from blood or other tissue specimens? In addition there is a complete lack of description of the cell lines that the authors had access to. Were these immortalized lymphoblastoid cells? If so this need to be explained. I could not find description in the references either.

4. In the methods it is stated that the authors utilized 100,000 Monte Carol simulation trials for SNPs of interest. However in the results it is stated that 25,000 or 10,000 Monte Carlo simulations were performed. Again the reviewer is confused by the lack of consistency.

5. The rationale for the comparison to brain libraries is inadequate. Also there is
no description of the brain libraries utilized. Were the authors examining particular brain regions or was this global brain cDNA libraries? Were the libraries for particular SNPs or was it to do with gene expression (see point below)?

6. In the results the authors state that they preferential brain and brain regional expression and refers to Table 1. However there is no information in this table on regional brain expression for the listed genes, again the reviewer is confused as to what the comparison exactly was with the brain cDNA libraries.

7. The statistical p values listed in the results are inadequately explained and do not tally with the table.

8. There was no attempt at discerning any potential biological meaning of the 172 genes of interest. Do they populate particular GO categories that may implicate certain functional processes in substance dependence?

**Level of interest:** An article of limited interest

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

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I no competing interests