Reviewer's report

Title: Psychosis candidate genes in the prefrontal cortex: Meta-analysis of gene expression microarray studies

Version: 1 Date: 26 June 2008

Reviewer: Paul Pavlidis

Reviewer's report:

Choi et al present an analysis across seven gene expression datasets obtained from the Stanley Genomics database, to identify expression changes between psychotic and non-psychotic patients. They identify several genes including metallothionein genes and some neuropeptides.

Major Compulsory revisions

In an earlier publication by the authors (Elashoff, et a., J Mol Neurosci. (2007)), a similar analysis is carried out on BPD and control data, using some of the same data stratified in a different way. In that study the authors discuss the potential impact of the samples having subjects in common, and the use of a degree of freedom correction factor to correct for study interdependencies (this correction is not described in any detail in the earlier paper). It is not clear why a similar correction was not deemed necessary in the current study. The authors should address this lack of independence head-on and thoroughly describe any correction they do decide to make. As it stands, the pvalues they computed cannot be taken seriously and the results are of questionable statistical significance.

I'm concerned that the authors are dredging some of the same data they used in Elashoff et al. There they studied the bipolar disease effect. Now they are restratifying the same samples based on psychosis (and adding additional samples), which is obviously not independent of BPD. It seems important to discuss the relationship between these two analyses of the same data. They even got one of the same genes coming up (MT1X). It's rather confusing that these two analyses of the same data (in part) are giving the same result when analyzed different ways, especially since the current study corrected for bipolar disease! (page 9) Surely this is something that requires an explanation.

Also pertaining to the shared subjects across datasets, Table 1 should be revised. As it currently reads, the reader can misinterpret ‘Subject Samples’ to mean the number of subjects under study, in which case the numbers would be incorrect. The row should be re-named ‘# of Samples’ and below it another row should be inserted to display the ‘# of Subjects’ (i.e. 82 and 81 for the Control and Psychosis groups respectively) for which the demographic data actually applies to. It would also be worth re-wording in the abstract ‘…over 400 individual microarray samples…’ to ‘….over 400 microarray samples comprised of 162...’
individual subjects…’.

The Methods section is a bit unclear. Did the authors use the pre-processed data, or did they use the raw data (.cel files) and run the same normalization methods across all datasets? In the Individual study analysis once the genes identified as significantly correlated with a demographic variable, it is not clear how this information was integrated with the psychosis analysis to obtain FC values. (p9. ‘….analysis of psychosis was performed…..adjusting for the demographic terms as listed above…’)? Please elaborate.

Minor essential revisions

In the discussion (pg 21) the authors state "Meta-analysis of schizophrenia shows that neuropeptide genes such as SST, NPY and TAC1 are not significantly regulated although these genes show a similar pattern of down-regulation as compared to the meta-analysis of psychosis". What meta-analysis of schizophrenia is being referred to? Furthermore, what does it mean for these genes to be not significantly regulated, but also showing down-regulation? This makes no sense to me.

Finally, the term meta-analysis seems somewhat inappropriate and even misleading for this study. The appeal of a meta-analysis is that it in effect combines findings from independent studies into one large study with many participants. In this study although there are multiple studies being combined, they are not independent, as they are all using samples from the same subject pool. If the authors use the term "combined analysis" or "cross-study analysis", there would be no problem.

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

Quality of written English: Acceptable

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

Declaration of competing interests:

I declare that I have no competing interests