Reviewer's report

Title: Modeling gene-by-environment interaction in comorbid depression with alcohol use disorders via an integrated bioinformatics approach

Version: 1 Date: 1 April 2008

Reviewer: Marylyn Ritchie

Reviewer's report:

“Modeling gene-by-environment interaction in comorbid depression with alcohol use disorders via an integrated bioinformatics approach” is a nice paper describing an important area of research in the study of complex disease. It is likely that complex diseases are due to the combination of genes and the environment and the detection of disease susceptibility risk factors is a major challenge. In this manuscript, the authors describe an approach whereby multiple public databases and several algorithms are integrated to discover genes important for comorbid depression and alcohol use disorders. This work is innovative and timely as there are many people in the field struggling with how to integrate data from multiple sources.

While this reviewer has much enthusiasm for the manuscript, there are some areas that need further clarification to make the methodologies and results more apparent.

Major Compulsory Revisions

- On page 6, where the 263 citations resulted from the PubMed query are discussed and are consistent with a statistically significant association between depression and AUD, did the authors manually curate or validate these 263 manuscripts to ensure that the context in which they share depression and AUD are consistent with the association you are looking for? Can you be certain that some of the papers are not reporting negative associations between the two disorders? Similarly can you be sure that the papers are looking at the association of these two disorders and it is not simply the report of different aspects of some other studies? It was not clear if these manuscripts were followed up after the query and this information is essential to comprehend what the query results really suggest.

- On page 8, the authors need to define their use of the word “interaction”. They discuss that the combinations of over-represented keywords provides evidence of interaction across a locus pair. Is this meant to be a statistical interactions, biological interaction, or simply co-occurrence? The word interaction has different meanings to different readers so it is important to define.

- On page 10, the authors indicate that the 21 keywords provide statistically significant evidence of interaction between TNF and MTHFR. Did the authors test this approach on data where we know there is a true interaction? I think that this
needs to be explained further. I am not sure that the co-occurrence of these keywords (even with permutation testing to demonstrate they occur more often than you would expect by chance) indicate “interaction”. Perhaps the word “co-occurrence” would be more appropriate. Again, this goes to my other point about the word interaction. For me, interaction implies a statistical interaction. I do not think that the co-occurrence of keywords allows one to infer a statistical interaction. Perhaps if these results were curated and a meta-analysis was performed, then one could make assessments of evidence for statistical interaction.

- Were the results on page 12 shown in Table 2 validated or curated?
- I don’t know much about the genes described here. Is the result of this manuscript a novel discovery and a new hypothesis describing how this pathway functions to lead to susceptibility for depression and AUD? Or is this a demonstration of a novel bioinformatics approach where it is shown to discover a pathway that we know a lot about? When I was reading page 13, I starting thinking this and was not sure. It would be nice to show that this approach works on something we know and then discover new things once we know the approach works. Perhaps this has already been done.

Minor Essential Revisions
- In the methods section, “Queries against NCBI Databases” section, the authors comment that Entrez gene surveys of the whole genome are complementary to whole genome association. It would be nice to have another sentence or two following this one to explain what is meant by this statement before describing the advantages and disadvantages.
- Also in this section, it is not clear what you are querying Entrez gene on. This is described in more detail later in the manuscript, but something should be said in this paragraph to provide examples for how these queries are done.
- On page 9, were no other genes returned from the Entrez gene query other than the three described? It was not clear if these were the best ones or the only ones.
- Table 3 is discussed in the manuscript after Table 4. Perhaps these table numbers should be switched.

Level of interest: An article of importance in its field

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