Reviewer's report

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

Version: 1 Date: 12 March 2008

Reviewer: david reif

Reviewer's report:

Reviewer's report (David M. Reif)
---------------------------------

The authors describe a novel approach to combining bioinformatic tools/resources toward the exploration of genetic and environmental influences on comorbid depression and alcohol use disorders (AUD).

While there is necessarily a great deal of human input in the approach described, the authors do a reasonable job of laying out a systematic method. Because public databases are used, the soundness of the data is predicated on the assumption of quality within PubMed, Entrez Gene, GeneGO, and DAVID. The authors present a strong argument for the validity of this assumption in the ‘Methods’ section.

Reproducibility of these experiments is subject to changing data within the databases used; therefore, subsequent analyses could foreseeably give different results. However, the authors have provided details of query terms used and other information as supplementary material, which is all that can be reasonably expected (short of requiring future applications to use the specific database versions queried here).

It is not clear from the manuscript or attached PDG-ACE document whether the software is available. Otherwise, the reporting and data deposition are satisfactory.

Overall, the introductory material, methods description, results, and discussion are well-organized and well-written. There are specific comments given below on recommended additions.

- Minor Essential Revisions

1. Figure 1 (Analysis Flow) does not seem to match the analysis progression described in the text. In particular, the middle layer in Figure 1 presents three hypotheses in parallel that do not interact directly with each other, whereas in the text, there was an ordered progression of these steps. The figure should be refined to show information flow as written in ‘Methods’ and ‘Results’.

2. The initials given preceding the email addresses on the title page for Erika FH
Saunders are out of order.

3. In the first paragraph of ‘Methods’, the authors should mathematically define positive predictive value (PPV), since it is used frequently in the paper. I assume their definition is:

$$PPV = \frac{TP}{TP+FP} = \frac{TruePositives}{TruePositives+FalsePositives}$$

4. The last paragraph of ‘Conclusions’ should be extended to discuss HOW the increased volume of data in public repositories might affect this analysis approach and what issues exist with current data. For example: How publication bias toward positive results affect this approach? Do only commonly-studied genes accumulate enough evidence to be relevant? What additional data would have the greatest effect on the authors’ approach?

5. Since the title includes “gene-by-environment interaction”, more introductory and/or discussion text should be devoted the limitations of this integrated bioinformatics approach in characterizing the GxE interaction with respect to dose, exposure, heritability, and other common issues in those fields. It is beyond the scope of this paper to add a comprehensive review of these issues, but some discussion is necessary to put these results in context.

- Discretionary Revisions

1. In the introduction, the authors should address the feasibility of using a natural language processing (NLP) approach in obtaining indications of association in PubMed queries. It may not be appropriate for this application, but the authors should acknowledge NLP as an alternative approach or perhaps as a future direction.

2. The authors could discuss the impact of analysis order on subsequent steps in their integrated informatics approach. For example, justification of why APOE was not included in formal analysis is provided; however, what if APOE was included in the PDG-ACE step rather than only in subsequent PubMed and GeneGo steps?

3. The authors should discuss availability of PDG-ACE software or if there are plans to distribute the tool.

What next?: Accept after minor essential revisions

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.