Reviewer's report

**Title:** A Towards-Multidimensional Screening Approach to Predict Candidate Genes of Rheumatoid Arthritis based on SNP, Structural and Functional Annotations

**Version:** 1  **Date:** 25 February 2010

**Reviewer:** Nora L Nock

**Reviewer's report:**

Major Compulsory Revisions (which the author must respond to before a decision on publication can be reached)

1) Hypothesis & Conclusions: The specific research question posed by the authors is not well defined. The Abstract, in particular, does not adequately describe the authors’ approach, how this approach is novel and how the approach improves upon current findings. The authors state in the Abstract that “our method could return more ‘precise’ results”; however, no results comparing their method to a “traditional” genome-wide association analysis approach are provided in the Abstract. It appears the authors did perform a “traditional” genome-wide association (GWA) analysis of the Genetic Analysis Workshop (GAW) 16 Rheumatoid Arthritis (RA) data but the results of this are not mentioned in the Abstract (and the methodological details provided on p.17 are scant, see “Minor Revisions” below. Furthermore, although the authors discuss the “traditional” GWA results on p.10 and compare the methods using a Venn Diagram (Figure 5), the authors should also provide a list of the specific candidate genes (& specific variant rs# from which the genes was identified) for each method (Note: This information could be placed in Supplemental Materials if there is not room in the manuscript). From Figure 5, it appears only 22 candidate genes overlap between the two approaches; however, on p.10, the authors state their significance level on the “traditional” GWA was set to 0.01 (not the more widely accepted level of 1 X 10-7 for a 500K panel). Therefore, the authors should discuss how changing the significance criteria would affect the results of this comparative analysis. Furthermore, the authors need to justify how they concluded that there are “34 known disease genes of RA” – is this based on individual candidate gene papers?

2) Acknowledgement of Prior Work (Introduction & Discussion): The authors do not adequately address and reference prior methodological work. For example, on page 4, the authors’ state: “A limited number of studies have used genome-wide association studies, function clustering algorithms or pattern recognition methods based on structural genomics knowledgebase of disease and non-disease genes to identify candidate genes, respectively. However, these studies to date do not identify candidate genes from multidimensional genomic annotations or knowledgebase.” However, these statements are not substantiated with supporting details or references.
In addition, in the Results and Discussion section on p.6, the authors’ state: “Consistent with all the previous associated studies of rheumatoid arthritis, genes in the strongest gene set were members of the immunoglobulin protein family, protein kinase domain family, SH3 domain family and ligand-binding domain of nuclear hormone receptor family and so on, and they included several high disease risk and hotly-reported genes such as CD4, FGFR1 and KDR.”; however, they provide no supporting references.

3) Databases: The authors need to provide the accession dates and versions/blocks for the various databases utilized since the information in these databases is continually being updated and could affect their results and conclusions.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1) General: There are various typos and spacing issues through the document – too numerous to list here. The authors should also define ALL acronyms at first use. The transitions between sentences are very weak making it difficult to determine critical points the authors are trying to convey. The authors should carefully proof their manuscript and revise accordingly.

2) Abstract: The Abstract is, perhaps, the most important part of the manuscript. The authors should revise the Abstract to make it clearer and more informative. For example, the authors state that “modest risk genes are likely to be missed after adjustment for multiple testing” (presumably using “traditional” GWA analysis approaches); however, the authors never provide specific results comparing the standard “GWA” analyses approaches to their method nor do they clarify how their method improves upon the multiple testing issue.

3) Conclusion: It appears the authors’ method could be utilized in other diseases; however, the authors never comment about this. It would be helpful for the authors to discuss this, if true, in the Discussion section (and Abstract).

4) Methods: On p. 13, Table 1 is not “below”; and, it would be helpful to reference the reader to Table 1 again on p.14 where the variables enter into equations #2 and #5 (since the variables are not defined in the text).

On p.17, it is unclear if the carrot (“^”) refers to an “and” or an “or” command. That is, does the gene need to be identified in just one or all three of the databases? It appears from Figure 1 that the “^” indeed refers to an “and” command; but, in the text on p.17, it says that “…we use f(gi)...by retaining genes that share one or more similar function annotations”. Thus, this issue needs clarified. Furthermore, the authors should comment on how the results might change if the algorithm was modified to include genes in only 1 (“or”), 2 or all 3 (“and”) databases. On p.17, additional details regarding the “traditional” GWA should be provided.
5) Figures: The figures need renumbered to avoid confusion for the reader (e.g., Figure 2 is on page 7 but Figure 1 is on p. 17). Figure 3A and 3B should be linked together.

**Level of interest:** An article of importance in its field

**Quality of written English:** Not suitable for publication unless extensively edited

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

**Declaration of competing interests:**

I declare that I have no competing interests