Reviewer's report

Title: Meta analysis of papillary thyroid carcinoma microarray data and independent validation: New insights from old data

Version: 1 Date: 17 December 2010

Reviewer: Kyle Furge

Reviewer's report:

This article by Vierlinger et al. describes the meta-analysis of several gene expression microarray dataset to identify papillary thyroid cancer-specific genes and pathways. The paper is reasonably well written however the conclusions that are drawn lack support from the analytical approaches. In addition, the problem question posed by that authors are not well defined "...that there is yet undiscovered information in the wealth of data deposited in public microarray databases". The idea of using meta-analysis to reveal additional information in gene expression data is sound however the authors do not convincingly demonstrate that they reveal information that could not have been gained by analysis of the individual gene expression datasets. The identification of the SERPINA1 as a gene that can discriminate between papillary thyroid cancer and other subtypes of papillary cancer and benign lesions has been previously reported.

Major Compulsory Revisions:

1) The distance weighted discrimination (DWD) method to combine datasets could lead to more robust results, but application of a dataset specific transformation prior to meta-analysis is not that unusual. The authors should show why other integration models were "...discarded do to poor performance"

2) The requirement for distance weighted discrimination of the datasets before classification is not clear. The authors show that without DWD correction, when multiple datasets are combined, they can still accurately classify the remaining dataset. If this is the case, why do they perform this procedure?

3) To identify individual genes that discriminate between papillary thyroid cancer, it is not clear if the same results could not have been obtained using multifactor ANOVA in which study site would be one of the factors in the analysis. The authors should demonstrate that their approach identified genes that could not be identified using a multifactor model. If point 2 and point 3 cannot resolved, the paper should be re-written to exclude Figure 1 and the DWD correction method.

4) The gene set enrichment analysis applied to the combined dataset if a good first step but how those pathways are associated with thyroid cancer are largely unclear. Perhaps the authors should explore further analysis to extend these preliminary studies into novel insights into papillary thyroid cancer.
Minor essential revisions:

1) The first paragraph of the results section is not written well and should be re-worked.

2) The qRT-PCR results of SERPINA1 are shown as ratios between a housekeeping gene and SERPINA1. This approach is seems atypical. The authors should provide references for this approach or consider using the more typical delta-deltaCT for reported qRT-PCR values.

3) The Figure 1B could be misleading. The authors claim that DWD preserves that biological information in these samples and infer this effect by showing the PTC and NG labels along the bottom of the clustering diagram. However, the co-clustering of the PTC and NG labels in the DWD data seems to be due to the way the dendrogram is drawn as it is difficult to find a cutpoint in the dendrogram that cleanly separates the NG and the PTG samples. The authors should show where cutting the tree on the DWD treated data leads to a significantly different PTC/NG classification versus cutting the tree on the non DWD-treated data.

Minor discretionary revisions:

2) The article should be edited for grammatical consistency

Level of interest: An article of limited interest

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

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

Declaration of competing interests:

I declare that I have no competing interests