Author's response to reviews

Title: Identification of SERPINA1 as single marker for papillary thyroid carcinoma through microarray meta analysis and quantification of its discriminatory power in independent validation

Authors:

Klemens M Vierlinger (klemens.vierlinger@ait.ac.at)
Markus H Mansfeld (Markus.mansfeld@ait.ac.at)
Oskar Koperek (oskar.koperek@meduniwien.ac.at)
Christa Noehammer (christa.noehammer@ait.ac.at)
Klaus Kaserer (klaus.kaserer@meduniwien.ac.at)
Friedrich Leisch (friedrich.leisch@stat.uni-muenchen.de)

Version: 2 Date: 27 January 2011

Author's response to reviews: see over
To the Editor

BMC Medical Genomics

**Subject:** Response to the reviewers, MS: 147150156475552 - Meta analysis of papillary thyroid carcinoma microarray data and independent validation: New insights from old data

Dear Editor, Dear Reviewers

First of all, we want to thank all reviewers very much for critically reading the manuscript and for their comments which are highly helpful, interesting and constructive. According to their suggestions we have improved the manuscript.

**Reviewer 4: Amit Aggarwal**

**Major Revisions:** None

**Minor Revisions:** None

**Discretionary Revisions:**

The sensitivity/specificity of SERPINA1 assay in FNAB’s is likely a critical for its usability in clinical setting. See prior published report in use of sham-FNA data such as Durand et al, Evaluation of Gene Expression Profiles in Thyroid Nodule Biopsy Material to Diagnose Thyroid Cancer, Journal of Clinical Endocrinology & Metabolism, 93(4) 1195-1202, 2008.

> Assessment of FNABs will be a very critical next step in the evaluation of SERPINA1 as a single mRNA marker for PTC but is without the scope of the present study.

I personally feel that the claim on “new insights” is bit weak as SERPINA1 and several other genes have previously been described by other researchers as a marker of PTC. The ‘bioinformatically’ driven functional analysis of top differentially expressed genes has not been developed or addressed in sufficient depth or detail.

> We have toned down the ‘new insights’ claim in the revised submission and expanded the function analysis of the significant genes.

The qPCR Ct data can be published as a supplemental data section.

> Link is included in the revised submission

**Reviewer 1: Kyle Furge**
Major Revisions:

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"

We have included a description of how we assessed the performance of other methods but did not include any figures. We feel that a thorough comparison of data integration methods would be beyond the scope of this paper.

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?

There are several ways to conduct a microarray meta-analysis, also without data integration (see bootstrap approach by Fujarewicz et.al. as stated in the discussion). However one needs to be careful when combining datasets without removing the study-bias (Ezlinger et.al., 2006). In the light of the results of principal component analysis it would have been likely to introduce a strong dataset bias in feature selection when omitting data integration. We thank you for pointing out that the paragraph on study-crossvalidation is misleading in that respect. We have therefore deleted it, as it is a mere description of technicalities and adds more confusion than value to the overall claims in the paper.

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.

Yes you are right, ANVOA based methods like the one propsed by Choi et.al. (Bioinformatics 2003) would have been a good option for identifying differentially expressed genes. However, we set out not only to identify differentially expressed genes, but also to test these for their discriminatory power. ANOVA would have been a very good option to achieve the first goal, but it would be difficult to assess the performance of any classification rule if the data is not on the same scale. This might have been feasible to do in a case where the classification rule consists of only one gene (as is the case here, but one doesn’t know this in advance), but is surely much more complicated for more complex classification tasks. As stated in the paper, there are other methods to achieve this, but we believe to have shown the advantages of using DWD to generate a homogenous dataset. We clarified this in the revised manuscript.

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.

Results and Conclusions on the GSEA part have been largely rewritten.

Minor Revisions:
The first paragraph of the results section is not written well and should be re-worked.

Rewritten in the revised submission.

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

This was an embarrassing mistake on my part, for which I apologise. It is corrected in the revised submission.

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.

Absolutely, the way the dendrogram is drawn is arbitrary; in fact we did not influence this at all (R default settings being used). The claim in the text is that the mixing of the branches shows removal of the dataset bias. All claims about biological information were based on the combination of the dendrogram and the PCA plots, where it is quite clear that the biological information is being preserved. We rephrased the figure caption to avoid this misunderstanding.

Reviewer 3: Radha Krishna Murthy Karuturi

Major Revisions:

Main finding/focus of the paper is not clear other than providing some support for DWD.

Thanks for pointing this out. It seems we need to emphasise our point more clearly: While there are other microarray meta-analysis approaches published on papillary thyroid carcinoma (each having their own advantages and disadvantages), this is the only one which was able to identify SERPINA1 as potent single mRNA marker, a fact that has been confirmed by independent validation. Up to now the most comprehensive meta-analysis relies on summary statistics and hence cannot assign a measure of confidence to their findings. Another meta analysis achieves a similar classification accuracy but needs a twenty-gene classification rule, which limits its potential clinical applicability. Both studies lack independent validation. While SERPINA1 showed up as being somewhat associated with PTC in some studies before, this is the first study to quantify its discriminative power on a large sample cohort (total of 181 samples). We changed some sections in the manuscript to further emphasise these important findings.

The meta-analysis is not something new for the authors to keep claiming "our approach"

You are right; we have toned down these claims in the revised submission.
As noted by the authors, SERPINA1 has already been shown to be a good marker even for PTC. I haven’t understood the importance of their experiments on SERPINA1.

Previous studies showed SERPINA1, amongst many other genes, to be somewhat associated to PTC in single DNA-array or IHC studies of mostly very small to small sample sizes. As stated above, this is the first study to quantify the discriminative power of SERPINA1 - mRNA on a rather large sample cohort. We emphasised this more rigorously in the revised manuscript.

Why was the GSEA analysis limited to top 20 at adj.pval<5e-25 which is way too stringent?

*Thank you for pointing out this misunderstanding. The methods section states that GSEA was calculated using the student’s t-statistic but fails to mention that no cut-off was used (see reference). Table 3 is a mere mentioning of the top genes in the analysis and has nothing to do with GSEA. The relevant statement was changed.*

How would the GSEA result be different if it were done on the signatures obtained from the individual datasets?

*Individual studies identify the same pathways. This is included in the revised submission of the manuscript.*

What other methods had been tried as alternative to DWD and how worse is their performance?

*Discretisation methods as well as Bayesian probabilistic methods have been tried. Their performance was evaluated by principal component analysis. This is included in the revised submission of the manuscript.*

Why DAD1 was chosen as normalizing gene?

*Of the potential housekeeping genes from Eisenberg et.al. ([http://www.compugen.co.il/supp_info/Housekeeping_genes.html](http://www.compugen.co.il/supp_info/Housekeeping_genes.html)) we picked three genes which showed stable expression in the meta analysis data. As stated in the Material section of the manuscript, the method of Vandesompele et.al. picked DAD1 out of these three genes as the gene with the most stable expression. Changing the housekeeping gene does not alter the overall results of the analysis. No changes to the manuscript were made.*

Minor Revisions:
- Grammatical errors were corrected
- Size of the array platforms were included in Table1

Reviewer 2: Ganesan Kumaresan

Major Revisions:

Since the authors found i) 9 gene set and, ii) 11 gene sets as equally good classifier of PTC/NG with 99% accuracy, the results should be included as major figure adjacent to the top left figure in Figure 2. Further, RT-PCR could be moved as next figur.
We included a heatmap to visualise their discriminative power of these genes.

The annotations in figure 1 must be modified to appropriate font size to increase the clarity and for the easy understanding of readers.

There are 99 leaves in the dendrogram, so these will not be readable in any case. Therefore we decided to colour-code them. The key for the colour code is in the legend. We changed the Figure caption to clarify this point.

In the results, in the last paragraph, the speculations on transcription factor motif must be deleted.

Done

Accordingly, With reference to Fig. 3, since the authors didn’t observe any significant results for cytoband enrichment and Transcription Factor Target site enrichment, the right top and right bottom figures in Fig.3 must be deleted. Similarly, the annotation in the Fig.3 must be modified to enhance the readability. (Fig 3 also could be deleted totally).

In order to focus more clearly on our main claims, we deleted Figure 3 and stated the results of Gene Set Enrichment Analysis in the text.

Discussion must be written again by excluding the unnecessary portions to minimize the content and to discuss exclusively on the most relevant aspects of the results.

Done

Minor Revisions:

In the abstract fifth line, authors may avoid the word “own”. In the Table 1, legends, one by email is not appropriate and needs to be changed.

Done

Sincerely,

Klemens Vierlinger