Reviewer’s report

Title: Molecular discrimination between follicular thyroid carcinoma and adenoma based on gene expression profiling in formalin-fixed paraffin-embedded samples

Version: 1 Date: 25 July 2013

Reviewer: Adetayo Kasim

Reviewer’s report:

The manuscript proposed a highly predictive 5-gene classifier to discriminate between malignant follicular thyroid cancer (FTC) and benign follicular thyroid adenoma (FTA). The manuscript is well written and a good piece to read.

Major Compulsory Revisions:

1. A few genes are desirable, but Figures A and B SUPPL1 do not support your choice of 8 genes. 12 genes seem more appropriate based on the misclassification errors. The authors should justify why the number of genes selected is not arbitrary.

2. I commend the authors for comparing their 5-gene classifier with the 76-gene classifier by Borup et al. However, there are other studies such as Weber et al with 3-gene and Hinsich et al with 5-gene classifier. It would greatly improve the manuscript if the authors could compare their 5-gene classifier with other classifiers with a similar number of genes. The authors should also discuss whether their 5 genes overlap with other studies (see Table 4 of Borup et al). I understand the discussion about pre- or post-operative gene signatures, but it would still be good to compare the different sets of genes.

3. The manuscript lacks any information about the uncertainty of identifying true gene signatures. Specifically, the 8 genes identified by the authors are not necessarily “the genes” but a set of genes to discriminate between FTC and FTA. It would be helpful if the authors could provide information about how these 8 genes are ranked in all the datasets in terms of fold changes and p-values. For example, are they consistently ranked in the top 8 across all the datasets?

Minor Essential Revisions

1. The authors should use either “p” (page 3: p < 0.05) or “p value” (page 8: pvalue < 0.0005) and not both.

2. Page 8: p value < 0.0005 (i.e. 0.09 false positive expected). The 0.09 false positives should be clarified because the significant level of 0.0005 implies 0.05% chance of finding a false positive.

3. Check “!” in “0.11-0.69![1]” on page 12.

Discretionary Revisions
1. How does the K-folds cross-validation results compare with LOOCV for dataset C and D?

**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