Reviewer’s report

Title: A gene expression profile for accurate tumour cell percentage scoring of breast tumour tissue: microarray diagnosis eligibility

Version: 1 Date: 23 February 2009

Reviewer: Lance David Miller

Reviewer’s report:

In this work, the authors describe the discovery and validation of a gene-expression based classifier capable of distinguishing low and high percentage of tumor cell tissue in frozen breast tumor samples. Given the subjective nature of pathological review for tumor cell percentage, and the requirement of tumor-section staining which is not easily adaptable for core and fine-needle biopsies, an accurate gene-expression based predictor of tumor cell percentage could hold merit as a rapid and quantitative alternative to conventional methods. Overall, this version of the manuscript is a bit short on details and interpretation of results, but the experimental design is mostly sound, and the findings are intriguing and worthy of further investigation.

Major Compulsory Revisions

1) The Methods section is scant on details. Where did the tumor samples come from? By what criteria were they selected (other than tumor cell percentage)? Was the use of these specimens approved by an institutional review board? How were the microarray data processed and normalized? What was the rationale for using a nearest-mean classifier?

2) The training and testing logic for constructing the classifier appears sound up until the final selection of 13 genes. The authors explain that of their initial set of 35 genes capable of optimal classification in the first training set, those with large variation in an additional set of 70 tumors with high (>50%) TCP were removed, resulting in 13 genes for setting the classification thresholds/weights (in the training set) and validating in the test set. How these 13 genes were selected is a bit ambiguous. What variation threshold was applied and why?

3) The authors do not provide a clear interpretation of how the TCP classifier should be used with respect to its predictive performance. In the title it is said to be for “accurate tumor cell percentage scoring…”, and it is suggested in the abstract that it essentially replaces histopathological analysis for initial tumor percentage scoring. However, the predictive performance of the classifier does not demonstrate accuracy in scoring tumor cell percentage, per se, but rather indicates an ability to distinguish low TCP from medium/high TCP better than chance. While it is clear by the overall accuracy and kappa score that conventional scoring and the TCP predictor are associated with one another, a more relevant question is what does it mean that 22% of the low TCP samples were incorrectly classified as medium TCP (30-49% TCP) or high TCP (>/>= 50%
TCP)? And the fraction of medium and high TCP incorrectly classified as low TCP is not disclosed or discussed. These questions are of practical concern (ie, if the utility of the classifier lies in its ability to distinguish low TCP from medium/high TCP tumors), and should be considered in the context of conventional histopathological review variation. Which is more reliable, an approximate 10% average variation in TCP scoring attributable to inter-pathologist disagreement, or the performance of the classifier given its classification error in distinguishing low from medium/high TCP? Some analysis addressing this perspective would provide much needed context for interpreting the utility of the classifier.

Minor Essential Revisions

1) In the last paragraph of the results section, it is stated that, “The difference between medium and high TCP samples was not significant…as the majority of the medium TCP samples are correctly classified as high TCP by the molecular profile.” Cross-classification (“medium” classified as “high”) is really “incorrect” classification, but I think the point is more at the fact that medium and high TCP samples, while viewed by the classifier in a similar way, still remain measurably different than the low TCP samples on the whole. Perhaps the authors can rewrite this sentence for greater clarity.

Discretionary Revisions

1) As prognostic gene expression predictors of outcome in breast cancer have been developed using the same (or similar?) microarray platform as was used by the authors, it might be interesting to attempt to evaluate how the TCP classifier can add value to the performance of such a prognostic signature. Presumably, tumors with suboptimal TCP will confound performance of a prognostic signature, as the signature would not be reliably assessed in those tumors. If the TCP classifier can identify tumors predicted to have low TCP, then might the exclusion of those tumors enhance the classification performance of the prognostic signature in question?

2) Does tumor histology (ductal, lobular, mucinous, tubular, medullary) make a difference in classifier performance?

3) It would be nice to demonstrate that the predictor’s performance is robust to issues related to choice of microarray normalization procedure and final gene selection method. For example, does 10, 15, 20 genes have essentially the same test-set performance as the 13 selected genes?

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

**Quality of written English:** Acceptable

**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.