Reviewer’s report

Title: An 8-gene qRT-PCR-based gene expression score that has prognostic value in early breast cancer.

Version: 3 Date: 21 January 2010

Reviewer: Mads Thomassen

Reviewer’s report:

The authors analyze the expression of 83 genes from the 70-gene profile; the recurrence score and a 2-gene score in 153 paraffin-embedded breast tumor samples. The data for the 3 profiles has previously been compared and published by the authors. In this manuscript they develop a new 8-gene profile and compare it to the 3 other profiles. Although several prognostic and predictive profiles are published it is important to do more analyses to identify the best combination of genes in the profiles. Furthermore, it is an advantage to make tests with few genes that can be performed in local labs.

Discretionary Revisions

1) The filtering from 83 to 53 genes and use of fresh frozen samples should be explained in more detail.

2) The feature selection method using p-values and correlation should also be explained in more detail

Major Compulsory Revisions

1) All samples in the data set are used to select the 8 genes. This introduces an information leakage in the leave-one-out cross validation (LOUCV). The bias is considerably smaller in this 83-gene (53 after filtering) data set compared to genome-wide analyses, but the manuscript would be improved by selection of genes in each LOUCV cycle, so that no information from the test set (the left out sample) was used in training. To build a final overall profile for validation in external data sets, the entire data set could be used.

2) The principal component method results in the "8-gene score". The cut-off point in this score is set to include 60 % of patients in the low risk group. Why was the cut-off set to 60 %? An alternative would be to select the optimal cut-off point in the single rounds of LOUCV.

3) It would be helpful for the reader to have sensitivity and specificity for the 8-gene profile and the other 3 profiles in this data set. This would also help to compare the performance of 8-gene profile and the 70-gene profile in the NKI data set.

4) The group of patients in the study is very diverse, resulting in a profile that is partially prognostic and partially predictive for a mixture of treatments. An sub-analysis of patients only receiving endocrine therapy would show the relevance of the 8-gene profile for decision of chemotherapy.
**Level of interest:** An article whose findings are important to those with closely related research interests

**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