Reviewer’s report

Title: Molecular Risk Assessment of BIG 1-98 Participants by Expression Profiling using RNA from Archival Tissue

Version: 1 Date: 19 October 2009

Reviewer: Virginia Espina

Reviewer’s report:

Two major topics are addressed in the manuscript “Molecular risk assessment of BIG 1-98 participants by expression profiling using RNA from archival tissues”. The first topic is compatibility of archived FFPE tissue with qRT-PCR based molecular scores, using fresh frozen tissue as a comparison. The second topic is the development of genomic molecular profiles as prognostic indicators for post-menopausal women with ER+ breast cancer. The paper has an acceptable level of writing and ease of reading. For patients with positive lymph nodes, the authors show a molecular score capable of indicating likelihood of poor disease free survival (figure 3E).

Major Compulsory Revisions

1. Although Supplemental Table 1 lists gene identifiers, it was difficult to determine the biological rationale for selecting the genes, rationale for determining the parameters of the algorithm and the rationale for selection of the algorithm itself.

2. The authors have not adequately discussed the sources of bias in their study. Specifically, tumor size is known to relate to aggressiveness. Moreover, as shown in figure 2 the subdivisions of the histological grading, which are determined by the subjective scores by the pathologists, have a better resolution than the PRO_10 score. This raises the general question as to what value the molecular score has over the standard histological grading.

3. The samples were not microdissected and were biased against tumors with a low percentage of cells because tumors with less than 30% tumor were removed from analysis (page 5). In routine pathologic settings, typical core needle biopsies contain less than 30% tumor. This means that an entire class of tumors was not available for scoring. The percentage of cells in the biopsy is entirely dependent on the biopsy sampling technique and not reflective of the tumor size. A biopsy sample could be obtained at the tumor perimeter, or center, resulting in samples with vastly different tumor cell percentages.

4. The authors do not show that their gene score reveals any additional information that is not demonstrable through histopathological grading (Figure 2). The PRO_10 score only changes one unit across the spectrum of histopathologic grading shown in Figure 2. The mean values fall within a very narrow range (14.0 – 15.0 units).

5. The most serious deficiency in the analysis is the lack of a test set to develop
the molecular score and an independent sample set to validate the score. The models were not validated with an independent, blinded validation set. Bootstrapping will not overcome biases built into the tissue collection methods and cellular heterogeneity.

Minor Essential Revisions

1. The table of the gene identifiers (Supplemental Table 1) should be included in the main body of the paper. This is essential information for the reader to understand which genes were used to develop the molecular scores.

2. The paper suffers from a lack of focus. Initially the paper appears to focus on compatibility of archival FFPE tissue for generating molecular scores but the remainder of the paper discusses generation of a molecular score for predicting DFS. The authors may wish to clearly delineate the two studies.

3. There is a discrepancy on page 5 regarding the number of gene controls utilized (3 or 5). Please refer to the following sentences: “RNAs were tested by quantitative reverse transcription PCR (qRT-PCR) with five control genes leading to the further exclusion of 35 samples (8%)...”. “For the remaining 342 RNAs... the expression of 34 genes (3 control genes and 31 genes for computing scores...”. Which 5 control genes were used to exclude samples and why were 5 genes used to exclude data, yet 3 genes were used to develop the molecular score?

4. Figure 2 legend does not indicate what is shown in the box and whiskers plot, i.e. mean ±2SD or some other statistical data.

5. Figure 3 is difficult to interpret and would benefit from additional labeling for each Kaplan-Meier plot as well as additional explanations of each plot in the figure legend. Specifically, to each figure please add the molecular score category and the patient category or group used in each plot.

6. Figure 4B – the reader needs to understand why the PGR_5 score has an inverse relation to the number of events, compared to the positive correlation for the PRO_10 and RISK_25 scores. Additional text describing the biology of the score associations would be very beneficial.

7. In general the authors did not provide adequate rationale regarding choice of genes, algorithms, or relevance to breast cancer biology. The reader would greatly benefit from enhanced discussion of these topics.

8. Additional discussion regarding effects of tissue heterogeneity should be discussed because these tumor samples were not microdissected. We can not know which cell population (tumor or stoma) contributed which percentage of signal to the molecular score.

9. Page 11 – the authors compare their molecular score to OncotypeDX recurrence score and claim that their RISK_25 score can predict DFS. This statement is unwarranted unless the authors use their score on a blinded validation sample set, or on the original samples used to generate the OncotypeDX recurrence score.

Discretionary Revisions
1. Additional information regarding the percentage of tumor cells would be appreciated. For example, a list of the number of tumors with >90% tumor, number with 75% tumor, number with 50% tumor, etc.

2. The conclusions are very general and are not supported with specific data or confidence intervals or p values. These values are listed in the results section but the conclusions would appear stronger if they were supported statistically.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare I have no competing interests.