Reviewer's report

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

**Version:** 2  **Date:** 27 December 2009

**Reviewer:** Virginia Espina

**Reviewer's report:**

The author’s have provided detailed responses to this, and other reviewer’s comments, but they have not sufficiently incorporated their statements into the revised manuscript. The author’s comments provide the additional level of detail that could transform this paper into an informative publication for the field of genomic diagnostics.

The authors should incorporate additional technical details. These details are necessary for this particular manuscript because the author’s stated purpose is to “report reliable molecular profiles using FFPE tissues” from participants in a multi-center clinical trial. The reader will greatly benefit from the additional details supporting this manuscript.

The paper remains weak in the following areas:

1. **Methods:** description of in silico selection of genes.
2. **Methods and/or Results section:** Comparison of frozen and FFPE samples, and sample descriptions.
3. **Discussion:** Implications of discarding approximately 10% of the FFPE samples due to low tumor content.

Suggestions to strengthen the manuscript are listed below.

1. **Methods**
   - Please add these statements from your reviewer comments to clarify the in silico selection of genes for the molecular score. “The genes used in this study were prospectively selected from publicly available microarray data and the scores were defined by giving equal weight to each gene in the four groups (proliferation, estrogen response, progesterone response, Her2 response). Thus, a training set was not used as the scores were based on in silico gene selection.”
   - Please list the microarray data sets that were used in the meta-analysis. This could be listed in supplemental information.

2. **Methods or Results section**
   - Please incorporate your statements regarding the percentage of tumor cells in the samples that were analyzed, such as “30% of the sections contained 30-50% tumor cells, and about 60% of tumors contained 50-100% tumor cells”.
   - Include quantitative information regarding the quality of the RNA from the frozen
and FFPE tissues.

- To avoid ambiguity regarding sample type, please include the following statements from your reviewer comments to definitively indicate the sample types: “All the patients from the BIG 1-98 were treated by mastectomy or breast conserving surgery” and “The blocks that were available were derived from representative tumor regions.”

3. Discussion

- The new sentence in the manuscript on page 12 of the Discussion is misleading and should be revised: “Tumor cells contain considerably more RNA than tumor-surrounding fat cells and therefore, molecular parameters are not or only marginally affected by contaminating fat cells.” Cells with high lipid content, such as breast adipose cells, may have lower RNA yield compared to tumor cells due to contaminating lipids. It is misleading to conclude that a lower RNA yield in breast adipose tissue compared to breast tumor indicates that the adipose cells are not affecting the molecular parameters, particularly if RNA amplification techniques such as PCR are used. It is well known that RNA can degrade at different rates in different tissues/cell types. It is highly unlikely in the BIG 1-98 study that all the samples were processed under identical conditions. Therefore, based on the variability of tumor content and lack of microdissection, it is unrealistic to conclude that the adipose cells are not contributing to the molecular profile in all the samples.

- The authors may wish to acknowledge that the samples were not microdissected and the possibility exists that other cell types in the sample may contribute to the molecular profile depending on the tumor content of the section.

- How is the molecular score going to be used clinically if samples with <30% tumor are not included in the score? This could cause significant bias in patients treated with neo-adjuvant therapy. How do the authors envision this score being used?

**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 that I have no competing interests.