Reviewer's report

Title: Interrogating differences in expression of targeted gene sets to predict breast cancer outcome

Version: 1 Date: 9 January 2013

Reviewer: Victor Jin

Reviewer's report:

Major Compulsory Revisions

The study presented by Andres et al used a multivariable gene expression model for predicting each outcome and from 32 candidate genes analyzed from their studies and public results, they found that molecular signatures consisting of five genes (PGR, GABRP, TBC1D9, SLC39A6 and LRBA) for disease mortality and of six genes (PGR, ESR1, GABRP, TBC1D9, SLC39A6 and LRBA) for disease recurrence. The work is very interesting, the presentation of the work is very clear. I have a few minor comments.

1. Can authors give out the pathological characteristics for each patient in a Suppl. Table?

2. In Table 1, it looks like that summary statistic for “Mortality” and “Recurrence” are same. Why?

3. Author should randomly select 32 genes for a control test. See how bad (or maybe as good as the selected set) for a random set. The random set of 32 genes may be selected as similar gene expression values. Also repeat the random set a few times.

4. There are many patient cohorts in a public database. The author may test this 32-gene signatures on these cohorts.

5. What is the possibility of next for the project, patent or clinical trial?

Level of interest: An article of importance in its field

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.