Reviewer's report

Title: Systematic assessment of prognostic gene signatures for breast cancer shows distinct influence of time and ER status

Version: 4  Date: 1 November 2013

Reviewer: Charles C Theillet

Reviewer's report:

This paper presents an extensive comparative analysis of 7 molecular signatures and 2 prognostic indicators in two large breast cancer expression profiling datasets. They conclude to important differences in prognostic power depending on the ER status, with most signatures, except the hypoxia signature, being informative in ER+, but not in ER- patients. This observation is not exactly new, having been pointed out in several studies before. What is interesting here is the extensive comparison of these signatures and the multivariate analysis with classical clinical parameters.

The second point the authors wish to make concerns the erosion of the prognostic power of the signatures with time. While it is an interesting finding, I was not entirely convinced by the strength of the conclusions.

Remarks and concerns

in the comparative analysis of the different signatures, the absence of the recent Guedj et al. (2012 Oncogene 1;31(9):1196) signature is regrettable. This signature has shown to segregate accurately ER+ from ER- and proliferative from non proliferative subtypes. In this sense it could perform just as well as the GGI.

The dependance of prognostic power to time could be undermined by multiple confounding factors. The test set is a compound dataset formed of 6 independent Affymetrix U133A datasets, with different aims, different treatment regimen and non unified clinical follow up procedures. Furthermore, it is now a known fact that while Basal-like tumors are at elevated risk of early recurrence, the number of metastatic events drops dramatically after 5 years. In ER+ breast cancer the situation is more complex, since two groups can be identified, one at risk of rapid recurrence, the second one showing late (between 6 and 10 years) or very late recurrence (>10 years). These late recurrences may, thus correspond to different biological settings, explaining why the signatures that are mainly fitted for aggressive breast cancer (high vs. low proliferation, high vs low hypoxia) perform poorly for late or very late recurrence.

Both these points should be discussed.

Minor points

Section "possible effects of cohort differences" the text states that cohort
differences do not contribute significantly to the tested models, but no data are shown to support this claim. Please correct.

The HER2 stratification came to no conclusion, which could be expected considering the work by Johan Staaf and colleagues in JCO (2010) 28, 1813, showing that HER2+ cancers do not form a homogeneous breast tumor subset and that major differences exist between HER2+/ER+ and HER2+/ER−. Hence, a double stratification could have been more informative.

In the discussion "we did not find any notable effect of cohort differences...." followed by a call on supplementary figure 1 and Box5. Supplementary figure 1 has nothing to do with dataset differences and I could not find Box5. please correct.

**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:**

no competing interests