Reviewer's report

Title: Gene signatures ESC, MYC and ERG-fusion are early markers of a potentially dangerous subtype of prostate cancer

Version: 2
Date: 10 April 2014
Reviewer: Qianben Wang

Reviewer's report:

This article aims to address the critical need for accurate prognostic biomarkers capable of predicting indolent vs. malignant prostate cancer to inform treatment practices. To accomplish this, they have analyzed expression levels across 15 gene signatures identified in a previous study, which were demonstrated as markers of good vs. poor prognosis. The samples utilized in this study, 116 cancer and 40 normal fresh frozen tissues from 41 patients, may be considered a validation cohort in some respects, and indeed it is the authors’ purpose to provide support for the use of these gene signatures in the clinical setting. The authors report success in validating the 15 gene set signature’s ability to distinguish good vs. poor prognosis based primarily on the finding that normal samples were assigned exclusively to the good prognosis subtype. Beyond this it is difficult to gauge the success as there is no long-term follow up data to confirm subtype assignment in this cohort and as only 47 of the 116 cancer samples were assignable to a subtype at a p-value < 0.05. Next, the authors demonstrate the significant impact that sample composition has on subtype assignment, reporting that unassigned samples were composed of nearly equal parts cancer and stroma while samples of higher cancer to normal ratios were more often assignable. This is somewhat troubling, as 31/40 normal samples were assignable, yet heterogeneous samples, whose overall gene signature is highly influenced by the normal expression signature, were paradoxically obscured from assignment. They attempt to address the issue of sample heterogeneity with a method that should be of conceptual value to the field, subtraction of a “normal signature” from cancer samples, and find that this approach improves their ability to divide the cohort into the two subtypes (in reality it has merely improved the ability to assign, but we cannot know if it has improved the appropriate or accurate assignment rate without follow up data). An important finding, and one that seems to be reiterated in the field is that gene set signatures related to poor and good prognosis show little relationship to Gleason score. The robustness of this concept is demonstrated by the fact that distinct samples from the same patient but with different Gleason scores show high gene signature correlation (though we are not told if these distinct samples are assigned to the same subtype) as opposed to comparison of samples with the same Gleason score, which revealed low gene signature correlation. Further refinement of this approach may then be able to improve upon the current prognostic value of the Gleason score. The authors elaborate on this idea by showing that three key gene sets (ESC, MYB, and ERG-fusion), whose overexpression was associated
with poor prognosis in the MSKCC cohort, exhibited high expression levels in samples with a low Gleason score from this validation cohort. This potentially identifies an important gene signature that predicts poor prognosis in men who might choose a less aggressive treatment option in light of the low Gleason score of their disease. Again, further refinement of this concept and validation in a cohort that has follow up data is essential before any conclusions can be definitively drawn.

I believe this manuscript is premature, and the conclusions drawn from the work are limited by the lack of recurrence and survival data for this validation cohort. The authors should seek an appropriate dataset with requisite clinical follow up data in which to demonstrate the validity of the gene set signature as a prognostic tool, the methods of improving subtyping by subtraction of normal signatures, and the concept of Gleason score-independent subtyping. In addition, I have several comments concerning various aspects of the methods and interpretation of the results that should be addressed before resubmission.

1) How were the original 21 gene sets organized in the final 15 gene sets?
2) Within the gene sets you combined to form the final 15, is it possible that certain sets could have a stronger influence on the composite gene set score? Could this obscure the ability to assign samples to a subtype when assignment is possible if the 21 gene sets were utilized as originally proposed?
3) It seems to me that the approach to correlating sample signatures to subtype signatures might result in artificial subtype assignment. Assigning a sample to a subtype based on the correlation with that subtype being “significantly better” than the correlation to other subtypes does not seem rigorous. Did you explore an independent correlation strategy?
4) I would also like to see the authors explore a normal signature subtraction method using an “individual normal signature” in cases where a matched normal sample is available for a particular cancer sample. The method utilized seems problematic, as Figure 1C (Top – Normal) displays a dramatic left skew with outliers that could negatively impact the assembly of an average normal signature. Perhaps characterizing the gene set signature differences in normal samples from patients with different Gleason scores would lead to a more refined method.
5) Throughout the results section, it would be appropriate for the authors to describe how many of the samples they are describing are from a single patient. It is critical in allowing the reader to judge the robustness of the approach, and readers should be able to identify redundancies within the data.
6) Regarding Comment 5, the impact of this work is severely limited by the small size of the cohort, which is in reality only 41 patients. I would imagine that several of your significant results would be lost if you performed these analyses based on a single, composite gene set signature (normal and cancer) for each patient. For example, clusters 1, 2, and 3 in Figure 2A are composed largely of multiple samples from the same patient. You would very likely lose cluster 3 and perhaps cluster 2 if you considered patients rather than samples.
The authors are very unlikely to find a validation cohort of appropriate size for which clinical follow up data and thorough pathologic analysis (such as that required for their subtraction method) is available. Thus I suggest a major revision that focuses either on a true validation of the gene set signature subtyping approach in a new cohort or on the valuable technical advance offered by their subtraction method in their current cohort.

**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.