Reviewer’s report

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

Version: 1 Date: 27 March 2014

Reviewer: Jing Wang

Reviewer’s report:

- Major Compulsory Revisions

1. In this paper, the authors performed several analysis including predicting samples from a new dataset into the previously defined subtypes of PCa, analyzing how the sample tissue composition affects the subtype prediction and how the sample source (from same patient or not) and Gleason score affect the subtype assignments. But the main question to be answered in this paper is not well defined.

According to the title “Gene signatures ESC, MYC and ERG-fusion are early markers of a potentially dangerous subtype of prostate cancer”, it seems that the main finding of this paper is the three signatures. But in the original paper (Markert et al., 2011), which defined the PCa subtypes used in this paper already indicated that patient groups with poor survival are characterized by ESC and ERG-fusion.

In the abstract, the authors indicated “Our study validates a previous molecular subtyping of PCa in a new patient cohort, and identifies a subgroup of PCa samples highly interesting for detecting high risk PCa at an early stage”. But the authors also indicated “Our data did not distinguish between four different subtypes of PCa as previously published, but rather highlighted two groups of samples”. The authors should clarify the confusion of these statements. What’s the purpose of this analysis (validating four subtypes or just assigning samples into groups with poor or good outcomes)? Are the survival data available for the new dataset to support the subtype assignment (poor outcome group and good outcome group)? For subtype assignment, p < 0.05 is used. Is the multiple testing correction performed?

2. For “Gene Set Enrichment Analysis” in Methods: (1) dsig(j) should be defined; (2) in the original paper (Subramanian et al., 2005), the Enrichment Score is defined as the maximum deviation from zero of Phit - Pmiss. What’s the relationships between dGS(j) = square_root((N-S)/s) and Phit, dGS(j) = -square_root(s/(N-S)) and Pmiss?

- Minor Essential Revisions

1. “Prognoses” should be “prognosis”.

2. The authors should indicate which kind of correlation (Spearman correlation or Pearson correlation) was used in this paper.

3. Legend of table 1: Table 1: Number of samples assigned exclusively and significantly (p>0.05 and p>0.25, dependent correlations by Steiger [36]). Should it be p < 0.05?

**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.