Reviewer's report

Title: Identification of a candidate prognostic gene signature by transcriptome analysis of matched pre- and post-treatment prostatic biopsies from patients with advanced prostate cancer

Version: 2
Date: 8 October 2014

Reviewer: Nuno L Barbosa-Morais

Reviewer's report:

MAJOR COMPULSORY REVISIONS:

1. The selection of differentially expressed genes for downstream knowledge-based examination is insufficiently explained (both in the main text and in the figure legends) and appears to involve some level of arbitrariness. The authors mention that differential expression was tested with a paired linear model using edgeR (lines 172-173, page 8). One therefore assumes that the up- and down-regulated genes mentioned in lines 214-215 (page 10) are differentially expressed between the group of 4 post-docetaxel samples and the group of 4 pre-docetaxel samples. However, in lines 223-225 (page 10) a selection of genes with consistent expression changes in at least 3 patients is mentioned. How was differential expression for each individual patient tested? The authors mention selecting the top 10 differentially-regulated genes but then state they identified 6 differentially expressed genes. How were the genes ranked for differential expression (LFC or FDR?)? What happened to the other 4? Is it that those 10 were significantly differentially expressed but only 6 of them showed consistent expression changes in at least 3 patients? Or is it that no statistical significance was used in ranking the genes and only 6 of the top 10 were significant? Whatever the case, it is not what is conveyed by the way it is phrased. It is also not clear why only 10 (6) genes were chosen, given that cBioPortal can handle longer user-defined lists of genes. This immediately limits the number of biomarker candidates to be found. Again, what was the criterion for deeming FAM72B and ADAM7 particularly impactful in terms of survival (line 236, page 11)? Were all the combinations involving any of the 6 genes tested and that proved to be most significantly associated with survival? Similarly, how were the 5 genes in Figure 3B selected from the 11 in Figure 3A? The authors mention that those 5 are the ones “which were significantly altered” (line 272, page 12) but so should the other 6, given that they were deemed to be up- or down-regulated (line 248, page 11).

2. In lines 207-212 (pages 9-10), the authors state they have observed 3 novel inter-chromosomal gene fusions. Apart from the inherent usefulness of identifying new disease biomarkers, what could be the relevance of these findings? Can something be said about the functionality of the resulting products? Moreover, given that the seminal reference on gene fusions in prostate cancer dates from the distant 2008, wouldn’t it be straight forward, for instance, to run the gene
fusion identification pipeline on a large number of samples from the TCGA dataset?

MINOR ESSENTIAL REVISIONS:

3. The second paragraph of the Background (line 96, page 5) omits which treatment confers a survival advantage, although one guesses it is docetaxel based on the first paragraph. This ends up slowing down the reading with somewhat misconveyed information that is already included in the first paragraph. I recommend rephrasing accordingly.

4. Lines 217-218 (page 10) and Figure 1B: the statistical significance of the correlation (and its actual value) between KLK3 expression and PSA levels should be stated.

DISCRETIONARY REVISIONS:

5. Figure 1B: are there zero values or wouldn’t a log-log scale show the trend more emphatically?

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests.