Reviewer’s report

Title: A 25-gene classifier predicts overall survival in resectable pancreatic cancer

Version: 0 Date: 06 Jul 2017

Reviewer: Francisco Real

Reviewer's report:

In this ms. Birnbaum et al. perform an analysis of a large dataset of published mRNA expression studies to identify a poor prognosis signature. They perform an extreme case-case comparison, select 25 best performers among 1400 differentially expressed genes and apply this signature to the whole dataset, the individual datasets and analyze associations with clinical/pathological factors and with survival. They compare the performance of the signature with that of other previously published classifiers and the data support the notion that this signature outperforms the published ones in the multivariable analyses. The work will be of interest to the community and should foster independent prospective validation.

1. Non-ductal tumors represent 10% of the whole series. They should be analyzed separately and a precise description of the tumor types should be provided.

2. Figures 2a and 2b are very difficult to read and they should be drawn in a different manner.

3. The training set is remarkably small compared with the test set; this may reduce the ability of the used strategy to capture the best genes for the classifier.

4. Although it is said that a comparison of the performance of the 25 gene-classifier in comparison with random signatures was analyzed, I cannot find that detailed information in the manuscript. Only a short statement is provided in the Results section.

5. Why use a two-tiered classifier rather than developing a 25-gene based score for continuous classification?

6. A pathway analysis of the 1400 genes should be provided in the main text as a Table.

7. Table 3: do the results presented correspond to associations or to correlations, as stated in the title? I believe that these are associations.

8. The authors may not have access to the full Australian ICGC dataset but indeed this would be the largest single dataset in which to test the classifier. It is not correct to say that the largest study corresponds to 102 cases.
9. A statement should be included in the Discussion indicating that, despite the very high p value, the HR for the signature in both the univariable and multivariable analyses is relatively low, around 2, and therefore of uncertain clinical value.

10. The discussion of specific genes in the signature is not useful, in my view: most likely it does not reflect causal effects but associations. The pathway discussion may be more relevant.

11. I am not convinced about the statement in the conclusion regarding the possible usefulness of the signature in determining whether patients should undergo immediate surgery vs. neoadjuvant chemotherapy since PDAC is most likely a systemic disease in the vast majority of cases.

**Are the methods appropriate and well described?**

If not, please specify what is required in your comments to the authors.

Yes

**Does the work include the necessary controls?**

If not, please specify which controls are required in your comments to the authors.

Yes

**Are the conclusions drawn adequately supported by the data shown?**

If not, please explain in your comments to the authors.

Yes

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**

If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I recommend additional statistical review

**Quality of written English**

Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**

Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

No competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal