Reviewer's report

Title: Significant Random Signatures Reveals New Biomarker for Breast Cancer

Version: 0 Date: 19 Jun 2019

Reviewer: Xiang Chen

Reviewer's report:

Based on the conclusion presented by Venet et al in 2012, the authors developed a new computational method to extract signals embedded in significant random signatures. Based on this approach, the authors identified a set of 840 significant genes from 47 published breast cancer-related signatures that were also assessed in two other datasets. Finally, the authors reported a new biomarker TAT for breast cancer.

Major comments:

1) The repeatability of the calculation method is important. The author should provide the script, the pipeline of script execution and the relevant test data (the datasets described in the manuscript) to facilitate the application and test of the method.

2) FDR correction should be applied in the calculation of empirical p-value in the part of methods, because the author has conducted 1000 hypothesis test.

3) What is "random permutation" in the section of "Identifying Significant Genes by Permutation Procedure"? How this is done needs to be clarified.

4) Figure 3 cannot show that most enriched pathways can significantly separate patients with good or poor prognosis. Is it necessary to exist? How to determine the good prognostic group and the poor prognostic group when calculating the score of phenotype N for each patient.

5) In the section of "Association of Top Twenty Genes with DMFS and RFS datasets", it is risky to use the t-test approach to calculate the significance p-value. Traditionally, the log-rank approach should be used to calculate the significance difference for the variable of survival time.

6) The test results of 7 breast cancer patients are meaningless to indicate the expression tendency of TAT, samples need to be added, and relevant test data need to be provided in the supporting material. Also, how is the significance assessment done here?

7) In the part of "Extracting significant genes embedded in empirically significant random signatures", the author needs to describe in detail the process of obtaining significant genes. For
example, how to produce the random signatures set and which genes to use; How many significant random signatures were obtained and their p-values respectively; Which gene-related protein interactions are used to compute significant genes.

Minor comments:

1) genes-&gt;gens, line 7, page 8

2) The author needs to clarify the relationship between signal and gene, which is confused in the manuscript.

3) Formula (1) is incorrect and incomprehensible.

4) The legend in Table 2 should specify the headers of the Table.

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

No

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Unable to assess

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

No

**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 am able to assess the statistics

**Quality of written English**
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published
**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.

I declare that I have 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/](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