Reviewer's report

Title: A specific gene expression signature for visceral organ metastasis in breast cancer

Version: 0 Date: 07 Oct 2018

Reviewer: Mads Thomassen

Reviewer's report:

The manuscript by Savci-Heijink et al. describes the development of a 14 gene expression signature for visceral metastasis of breast cancer. The study is based on data from 157 primary tumors from patients that developed metastasis. The data has been published before where the authors showed association of molecular subtypes and site of metastasis. The results are significant also in independent testing samples but the sensitivity and specificity are not impressing showing the difficulties in predicting homing of metastasis. The problem of metastasis homing is both clinical and biological relevant. Several previous attempts has been performed with limited success.

The study is well performed and the data are well presented. The data set with 157 metastatis end points is large compared to previous published studies. The discussion of the classification results and the identified genes is balanced and interesting.

My main concern of the study is the statistical methods. Testing is performed in the training set which will lead to overfitting. Furthermore, the method for classification is a little primitive. This may partly explain why the performance of the classifier is not high.

Genes are selected by ANOVA and genes with p value less than 0.0001 were included in the classifier. Classification was performed with K-means clustering. This is a very simple approach. Why did the authors not optimize the number of features in the classifier? Did they try other classification methods? K-means clustering is a simple method compared to SVM, neural network, Random Forest etc.

The results are then validated in the test set and the training set. However, testing in training set is not optimal. Why did the authors not perform leave one out cross validation?

ComBat was used for batch reduction which is a good strategy. However, this method requires balanced subset of samples in each batch, i.e. different types of metastases must be balanced. Was this the case?

Minor comments:

There are several typing errors and the language could be improved.
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 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/). 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