Reviewer’s report

Title: Distinct signatures of lung cancer types: aberrant mucin O-glycosylation and compromised immune response

Version: 1 Date: 23 Apr 2019

Reviewer: Aleksandar Dakic

Reviewer’s report:

Both Reviewer 1 and I suggested that the manuscript would benefit from some clarity when it comes to explaining the composition of three datasets ("all", "paired", "unpaired") and the workflow of the analyses. In revision, authors did some good work with the flowchart of the analyses in Figure 1, but the inset called "Datasets" does not help in explaining the relations between datasets. Can authors please replace the inset with the table attached to this review or have the table in the main/result section of the paper as Table 1. - so that a reader can easily see the numbers and understand how datasets and analyses relate to each other. (It is not helpful to put this in the supplementary text or table.)

In the Section 3.2, when describing the numbers of samples (tumor / normal, all / paired), the order of samples is used incorrectly in the context of "respectively" and the sentences do not read very well. I suggest something like "In our LUAD analyses, we used PAIRED dataset with 32 tumor and 27 normal samples, as well as ALL dataset with 324 tumor and 59 normal samples. For LUSC, PAIRED dataset contained 35 tumor and 35 normal samples, and ALL dataset 356 tumor and 51 normal samples. In both cases the PAIRED datasets were smaller subsets of the corresponding ALL datasets" (alternatively replace ALL with FULL throughout)

Entire Section 3.2 can be summed up with a good paragraph stating that the results from limma and edgeR pipelines considerably differs from those of edgeR-TCGAb pipeline (because the latter does not correct for patient-specific and batch effects). In a space of just one page authors repeat twice that this effect is more pronounced in PAIRED analysis and why.

Section 3.3:

The storyline about the involvement of the complement and innate immune system pathways is completely unchanged in both text, Table 1, and Fig 2S. It seems that authors report the same set of tumor-specific up- and down-regulated genes in this analysis as in the first draft, yet they claim in the Methods section that deconvolution was performed in the meantime. The immune cell composition estimates from this analysis and how it affected DE and pathway enrichment analysis were not mentioned anywhere. Only in the Conclusion section authors claim that, upon deconvolution, most of the genes in the complement and innate immune pathway were "validated" - except 9 genes in the innate immune and 2 genes in the complement pathway (although authors present it as only one complement gene! - C2 and C4BPA are contained in both complement and innate immune pathways). But the reported DE and pathway enrichment
results were not updated (text, Table 1, and Fig 2S) - the genes were not even removed from the Table 1. Of course just removing them from the table and leaving the FDR the unchanged would be misleading. If the authors want to claim that deconvolution has been performed, the entire DE and pathway enrichment analysis needs to be re-run and reported in the Results section. This would likely change the results in all the other sections - regardless whether they refer to immune involvement or not.

Section 3.5:

I do not understand why the described lists of potential oncogenes and tumor suppressors for LUAD and LUSC are on GitHub and not in the supplementary document.

Section 3.6:

This section also refers to the immune involvement, so the result would likely change if the analysis has been reported after the deconvolution.

Section 3.7:

It is not known if the final list of candidate genes in Table 2 would change after the deconvolution

Section 3.8:

In my previous revision, it was not my desire to see "group_low" coefficient explicitly mentioned in the text (this will mean little to a reader), but rather the effect size of the coefficient succinctly interpreted - quantified in terms of survival, rather than just called "significant" or "has a good prognosis"

Discussion and Conclusion:

The sections that refer to the immune involvement would likely change if the analyses has been reported after the deconvolution. It is not transparent enough to mention in the Discussion that quite a few immune genes reported in Table 1 actually should not be there according to the deconvolution result, but regardless "other" immune genes were "validated upon the deconvolution" so the immune involvement seems "genuine" - where is the result showing this?

Furthermore, in Discussion, Conclusion and the summary figure 7 authors make a claim that their work "links" LUSC with the immuno-evading oncogenic pathways activated by p53, cMyc and beta-catenin. I understand that there is a review where those regulators were connected with immune-evasion in cancer, but do not see such a claim substantiated anywhere in this work. The
fact that the LUSC-downregulated genes were enriched for the innate immune system pathway genes (before the deconvolution) does not directly "links" them to p53, cMyc and beta-catenin in particular.

**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.

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/). 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