Author’s response to reviews

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

Authors:

Marta Lucchetta (malu@cancer.dk)
Isabelle da Piedade (isap@cancer.dk)
Mohamed Mounir (mohm@cancer.dk)
Marina Vabistsevits (marina.vabi@hotmail.com)
Thilde Terkelsen (thilde@cancer.dk)
Elena Papaleo (elenap@cancer.dk)

Version: 2 Date: 21 Jun 2019

Author’s response to reviews:

Dear Editor,

we attached the revised version of the manuscript where we have addressed in full all the remaining comments of reviewer2. We would like to sincerely thank you and the reviewer for the careful attention to our work and for the useful comments to improve further the data presentation.

Here our point by point response to the reviewer for sake of clarity:

1. 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.)

We have now added the information within Figure 1 in the main text
2. 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)

We have revised the sentence accordingly

3. 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.

The text has been shortened and redundancy removed

4. 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.

We have revised the results to include and discuss the new 16 DEA analyses correcting for the tumor microenvironment very early in the text. Table 1 includes now the pathway enrichment results that are robust independently on the correction for the microenvironment and only the common genes within these pathways. We have also revised the Table with candidate genes since three of them are no longer significant (RND3, SIPA1L2, and SLC1A3) when also the DEA with the correction for microenvironment are taken into account. The rest of the analyses
was not affected. Fig2S legend has been also revised to mention the full set of GO-enrichment for each DEA comparison (in Github).

5. 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.

We don’t think it will be useful to have it in the SI material due to the long list of predicted genes by Moonlight, so we prefer to leave them in the Github repository if readers interested in the details want to explore the list. The most important information of us was to use this analysis for the annotations of the candidate genes in Table 2.

6. 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.

This section is actually not related to differential expression analysis since it is the study of the coexpressed modules (so a different analysis not related to DEA) but we have revised the statements when we were linking these results to the DEA results.

7. Section 3.7:

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

We have revised these sections and Table (current Table 3), see our comments above (only 3 candidate genes no longer significant). We have also revised the survival analysis in section 3.8 and the validation on independent datasets of the candidate genes.

8. 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"

We have revised the survival analysis with the new list of candidate genes and added the hazard ratio column in the Table S4 in order to quantify the survival rate.
9. 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?

We apologize if the solution that we conceived at the previous revision sounded not appropriate to the reviewer but we now introduced (as explained above) all the DEA with corrections from microenvironment very early in the text, compare them to the consensus DEA without these corrections, revised the pathway enrichment, GO enrichment, candidate gene selection, and survival analyses that depends on this comparison and revised the discussion revising our statements where they didn’t hold robust after this comparison.

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.

The reviewer is right and we have now revised the text since it was not our intent to claim a link, it was some leftovers of a early version of the draft that we missed. We only aimed at suggesting that these three pathways could be interesting to explore further in LUSC in light of the recent review article that we cited and the importance of a more compromised immune response in LUSC than LUAD in our analyses.

10. If improvements to the English language within your manuscript have been requested, you should have your manuscript reviewed by someone who is fluent in English. If you would like professional help in revising this manuscript, you can use any reputable English language editing service. We can recommend our affiliates Nature Research Editing Service (http://bit.ly/NRES_BS) and American Journal Experts (http://bit.ly/AJE_BS) for help with English usage. Please note that use of an editing service is neither a requirement nor a guarantee of publication. Free assistance is available from our English language tutorial (https://www.springer.com/gb/authors-editors/authorandreviewertutorials/writinginenglish) and our Writing resources (BMC_WRITING_RESOURCES_URL http://bit.ly/NRES_BS) and American Journal Experts (http://bit.ly/AJE_BS) for help with English usage. Please note that use of an editing service is neither a requirement nor a guarantee of publication. Free assistance is available from our English language tutorial (https://www.springer.com/gb/authors-editors/authorandreviewertutorials/writinginenglish) and our Writing resources (BMC_WRITING_RESOURCES_URL
http://www.biomedcentral.com/getpublished/writing-resources). These cover common mistakes that occur when writing in English.

We used the resources suggested above to revise the English (along with the Grammarly Premium) and internal service available at our Institute for professional revision of scientific English. We also had one of our coauthors who is proficient in English revising the text carefully.