Reviewer’s report

Title: Integrative analysis of genetic and epigenetic profiling of lung squamous cell carcinoma (LSCC) patients to identify smoking level relevant biomarkers

Version: 0 Date: 18 Jun 2019

Reviewer: Federica Lo Sardo

Reviewer's report:

The work is interesting because it is aimed to characterize the molecular profile of LSCC patients, representing a big fraction (25%) of lung cancers. Up to date, several mutations have been characterized as oncogenic drivers in lung cancer and proposed as potential therapeutic targets. The present work in particular analyzes the genes that are differentially expressed in high smoking patients LSCC respect to low smoking patients in order to provide the rationale for new potential therapeutic targets, in particular in patients with a known smoking history. The work is potentially interesting. However, it is very correlative and poor of biological readout. The authors should address some points prior to publication.

1. Add some recent review about correlation of smoke, air pollution and molecular profiles of the patients with the risk of lung cancer i.e "The health and social implications of household air pollution and respiratory diseases". Simkovich et al., 2019, "Tobacco biomarkers and genetic/epigenetic analysis to investigate ethnic/racial differences in lung cancer risk among smokers". Musphy et al., 2018

2. In the results sections, explain what all the acronimous stand for i.e : DEGs=Differentially expressed genes?

3. Remove the GOTERMS from the text to make the results section more easy to be read.

4. It is difficult to understand the difference between figure 1 and figure 2. Do both of them show upregulated genes in high smoking patients? Explain better in the legend. In general, use bigger fonts in the figures or add tables to see in details the lists of genes.

5. Page 10, line 17, while describing the genes belonging to the Hippo pathway, add the recent review about the importance of YAP and TAZ in lung carcinogenesis "YAP and TAZ in lung cancer: oncogenic role and clinical targeting" Lo Sardo et al., 2018. This review supports also the finding showed in figure 5A, (the strong overexpression of TAZ in high smokers vs low smoking patients). In the discussion section TAZ is defined as tafazzin. Are the authors sure that it is tafazzin and not the hippo pathway transcriptional coactivator (TAZ/WWTR?)
6. In page 12, it is not clear why AIRE, SLC6A3 and PENK genes were selected. Authors should explain clearly on their basis of what rationale they were selected. In the discussion section, add some references about the biological role of these genes in lung cancer.

7. Finally, in order to give stronger biological readout of this in descriptive analyses based on published datasets, authors should show at least one experiment that validates some of the genes that are differentially regulated in high and low smoking patients. For example, they can analyze the expression of the proteins encoded by those genes in patient. They can also use already deposited data of immunoistochemistry, if present. More appreciated would be some functional expriment in lung cancer cell lines where the expression of a pair (ore one) of those interesting genes is manipulated in order to see the phenotipic effect on cell proliferation, colony formation, cell cycle profile, or some oncogenic mechanism. Otherwise the work appears a list of differentially regulated genes that are not placed in their biological context.

8. English editing is required

**Level of interest**
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

**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 have no conflict of interest

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

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

Yes