Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-23-1890

Reviewer A

The manuscript called 'Identification of prognostic biomarkers of smoking-related lung cancer' tries to demonstrate the association of BPIFA1, SLPI, and SCGB3A1 with lung cancer in smokers, with SCGB3A1 revealing a notable correlation with patient prognosis. The manuscript's topic could be interesting and reach the interest of the Journal of Thoracic Disease's readers. However, I identified essential flags that make it impossible to accept the manuscript in its present form.

1. The profile expression in cancer is unstable depending on the tumor stage; however, in this study, it is unclear which tumor stage is used.

Reply 1: The profile expression on the tumor stage is based on the online database. We have added the tumor stage in the Supplementary table 1 which including the clinical information.(see page 6, lines 171)

2. Concerning bioinformatic tools, please describe the versión package R, version R, and enrichments functional analysis. For example, in line 161, the use of p-value adjustment is established on the statistical significance. However, the author uses a p-value, which is not correct.

Reply 2: Thanks for the reviewer's comments. This p-value used is |log2(fold change)|, we have recorrected. The detailed *P*-values are in the Table 1 and Table 2. (see page 7, lines 212-219)

3. The authors have used a dataset published in 2008 (GSE12428); they need to show the complete list of the genes differential analysis and supplementary data. These data are essential because it could be interesting to compare these results against the data presented in 'Smoking and cancer-related gene expression in bronchial epithelium and non-small-cell lung cancers' (https://doi.org/10.1002/path.2039). In this sense, the authors could obtain more interesting results by applying WGCNA analysis to this data.

Reply 3: Thanks for the reviewer's important comments very much. We have added the GSE12428 details in the Supplementary table 2 and Supplementary table 3 which classified by tumor and non-tumor/ smoking and non-smoking.

4. Has enrichment functional analysis been done using ORA or GSEA needs to be clarified. ORA methods differ from GSEA because they only consider the query gene set of interest and

require a strict cutoff to classify genes as up- and down-regulated.

Reply 4: Thanks for the reviewer's comments. In this study, we don't use ORA or GSEA to analysis the enrichment functional. In the future, we will conduct more in-depth analysis with more samples.

5. Regarding cohorts TCGA, the dataset is not classified based on Smoking. I don't know if repression genes are suitable prognostic biomarkers.

Reply 5: Thanks for the reviewer's comments. The TCGA dataset is not currently disaggregated by smoking status. We will conduct more in-depth analysis to verification the biomarker s function.

6. Last, the authors present poor quality on the figures; I recommended better it.

Reply 6: We have increased the resolution of the figures so that it is clear.

Reviewer B

1) First, I suggest the authors to indicate the bioinformatics analysis and the datasets used for the current analysis.

   Reply 1): Thanks for the reviewer's suggestion. We used the dataset published in GSE12428 and online TCGA database.

2) Second, the abstract needs some revisions. The background did not explain the clinical needs for this research focus and what the potential clinical significance is. The methods need to describe the clinical samples in the datasets, the definition of smoking, the prognosis outcomes, and how the prognostic role of the identified biomarkers was ascertained. The results need to briefly describe the clinical samples used and quantify the findings by providing expression levels and accurate P values, as well as the HR values. The conclusion needs comments for the limitations of this study.

   Reply 2): Thanks for the reviewer's comments. We have revised the part of abstract which has been marked. (see page 1-2, lines 34-67)

3) Third, in the introduction of the main text, the authors need to clearly define smoking-related lung cancer, what has been known on its prognosis and prognostic biomarkers, and analyze why its biomarkers deserve to be studied. The authors need to have comments on the limitations and knowledge gaps of prior studies.

   Reply 3): Thanks for the reviewer's comments. The definition of smoking-related lung cancer has been described in the Methods. We have changed the limitations and

knowledge gaps of prior studies in the introduction.(see page 4-5, lines 135-140; see page 5, lines 157-159)

4) Fourth, in the methodology of the main text, please provide the definition of smoking-related lung cancer and explain why ex-smokers were included. The authors need to describe the clinical samples, clinical variables, and prognosis outcomes. In statistics, please describe the test of the independent prognostic role of the identified biomarkers.

Reply 4): Thanks for the reviewer's comments. We have provided the definition of smoking-related lung cancer in the Methods. Including the ex-smokers is for identifying genetic differences between smokers and ex-smokers, as well as for grouping. We also supplemented the sample with clinical information and described the test of the independent prognostic role of the identified biomarkers. (see page 5, lines 157-159; see page 6, lines 170)

5) Finally, please cite several related papers: 1. Hanash S. Lung cancer susceptibility beyond smoking history: opportunities and challenges. Transl Lung Cancer Res 2022;11(7):1230-1232. doi: 10.21037/tlcr-22-477. 2. Zhang X, Guo X, Gao Q, Zhang J, Zheng J, Zhao G, Okuda K, Tartarone A, Jiang M. Association between cigarette smoking history, metabolic phenotypes, and EGFR mutation status in patients with non-small cell lung cancer. J Thorac Dis 2023;15(10):5689-5699. doi: 10.21037/jtd-23-1371.

Reply 5): We expanded the Discussion section.(see page 9, lines 296-29, and page 10, lines 306-310).

"Although the relationship between smoking and lung cancer is well known, much work remains to fully elucidate the risk factors associated with lung cancer among smokers and non-smokers. （PMID: 35958329）"

"Zhang et al. found that compared with NSCLC patients who smoked, non-smoking patients were more sensitive to EGFR tyrosine kinase inhibitors and had better prognosis. In addition, it was found that non-smoking patients had a higher maximum standardized uptake value of primary tumors and a lower incidence of EGFR mutations. (PMID: 37969305)"

Reviewer C

The paper titled "Identification of prognostic biomarkers of smoking-related lung cancer" is interesting. Three genes, BPIFA1, SLPI, and SCGB3A1, were identified as being associated with smokers with lung cancer, with SCGB3A1 showing a close correlation with patient prognosis. These findings provide potential new targets for the treatment of lung cancer. However, there are several minor issues that if addressed would significantly improve the

manuscript.

1) It is recommended to increase the evaluation of the correlation between SCGB3A1 expression and prognosis and clinicopathological factors in patients with smoking-related lung cancer.

Reply 1): Thanks for the reviewer's comments. We have added a description of this section in discussion part.(see page 11, lines 356-365)

2) This study is only the result of bioinformatics analysis and requires experimental validation with a larger sample size.

Reply 2): Thanks for the reviewer's comments. We will conduct more in-depth analysis with more samples in the following study.

3) It is recommended to add in vivo and in vitro experiments to study the biological function of SCGB3A1.

Reply 3): Thanks for the reviewer's suggestion very much. Due to limited experimental conditions, we will study the biological function of SCGB3A1 in vivo and in vitro in the following study.

4) How to judge the prognostic characteristics of smoking-related lung cancer based on the results of this study? How to provide candidate targets for the treatment of smoking-related lung cancer? It is recommended to include relevant descriptions in the discussion.

Reply 4): Thanks for the reviewer's comments. It may be predicted by judging the expression level of SCGB3A1 to prognostic characteristics of smoking-related lung cancer. The key molecules that can be used as therapeutic targets for smoking-related lung cancer can be identified through gene and molecular target research, immunotherapy, clinical trials and drug development. We have added this describe in the discussion. (see page 11, lines 356-367).

5) The introduction part of this paper is not comprehensive enough, and the similar papers have not been cited, such as "Fusion gene recurrence in non-small cell lung cancers and its association with cigarette smoke exposure, Transl Lung Cancer Res, PMID: 36386463". It is recommended to quote this article.

Reply 5): We expanded the introduction section.(see page 4, lines 117-120).

"Vellichirammal et al. reported a positive correlation between smoking and fusion frequency in lung adenocarcinoma and found that as a fusion gene associated with cigarette smoke exposure, downregulation of the P53 pathway resulted in higher gene fusion formation in lung adenocarcinoma.（PMID: 36386463）"

6) What is the relationship between SCGB3A1 and tumor-infiltrating immune cells? What role does SCGB3A1 play in prognosis in tumor? It is recommended to add relevant content.

Reply 6): Thanks for the reviewer's comments. Through literature research, we found that SCGB3A1 may serve as a potential prognostic biomarker and immune-related therapeutic target for LUAD. We have added this describe in the discussion part and has been cited. (see page 11, lines 355-363)