Peer Review File

Article Information: https://dx.doi.org/10.21037/tlcr-21-729

Reviewer A

The manuscript presents interesting information about RNA profiles in plasma EV. Despite a small number of participants in the study, the authors were able to identify RNA signatures that discriminate samples of lung cancer patients and healthy controls. However, due to several weaknesses, the manuscript should not be published in the present form.

Comment 1: Analyzed group does not represent the “high risk of lung cancer” population. The majority of study participants have no smoking history (compared groups are unbalanced concerning smoking and no information on pack/year is provided). Moreover, the study participants are generally younger (ca. 55 years) than a high-risk group for screening study (usually inclusion criteria are 50-75 years, average ca. 65 years). Therefore, the discussion of presented data in the context of lung cancer screening is not justified. Hence, this aspect should be carefully considered in the manuscript. Moreover, validation of the proposed signature using a larger cohort that matches the screening criteria of being at high risk of lung cancer would markedly increase the impact of this finding in the future.

Reply 1:

Thank you for your helpful comment.

Our study focused on the diagnosis of early stage lung cancer. Restricted by conditions, the non-cancerous control samples in our study are from the healthy crowd in our hospital and the patients who underwent surgery in our department but proved to be benign nodule by pathological evaluation after surgery. What needs to be pointed out is that the epidemiologic studies reveal that the proportion of lung cancer in never smokers (LCINS) is higher in China, and approximately one third of all lung cancer patients are LCINS. (Zhou and Zhou, 2018). Huang et al. also revealed that Proportion of lung adenocarcinoma in female never-smokers has increased dramatically over the past 28 years (Huang et al., 2019). So, the study participants

are generally younger and the majority have no smoking history. We added the smoking history information for the healthy crowd in Table 1. This point was added to the *Discussion* section as a limitation of our study. Validation of the signature using a larger cohort that matches the screening criteria of being at high risk of lung cancer is needed in the future.

2) Figure 3 suggests that all samples were processed together. Hence, the information on precautions taken against information leakage between groups during data analyses should be presented.

Reply:

Thank you for your helpful comment.

Although the integration of data is carried out together, in fact, samples (especially plasma samples of tumor patients) are collected by different doctors at different time periods. So, Information about the three cohorts is relatively independent and not easy to leak. We added this point in the *Materials and Methods* section.

3) According to data presented in Table 2 even for the selected transcripts the difference between groups was not statistically significant if the FDR correction for multiple test correction was applied – please comment on the possible implications.

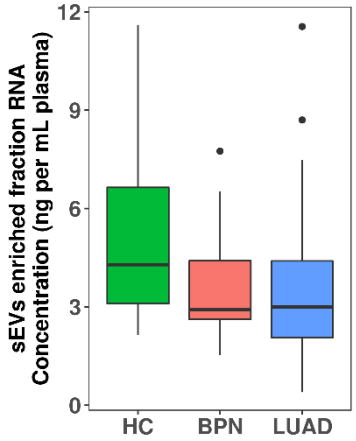
Reply: Thank you for your helpful comment.

The origin of exosomal RNA is complex. Compared with the RNA in the tissue, the differences in the RNA in the exosomes are diluted and the degree of differential expression is smaller. So, we didn't use the strict FDR standard. Instead, a more relaxed P-value criterion was used, and in subsequent marker screening, lasso was used to screen subsequent markers. There are also several similar studies that have also used P-value for marker screening (Min et al., 2019; Yuwen et al., 2019).

4) It would be important to know whether differences in the EV RNA profiles between cancer and control reflected differences in the populations of EVs between these groups. However, no information about hypothetical differences in number/amount and/or size of EV present in control and cancer samples is provided. Relevant data should be presented in the manuscript.

Reply: Thank you for your helpful comment.

A box plot comparing the RNA concentration (ng per mL plasma) of sEVs enriched fractions isolated from HC, BPN and LUAD groups was analyzed. Total RNA isolated from sEVs enriched fractions was analyzed and quantitated. No significant difference was identified in the RNA concentration of sEVs enriched fractions (ng/mL plasma) between HC, BPN patients and LUAD cases. We added it to supplementary Figure S2.



5) The introductory statement “However, the profile of EV long RNA (exLR) in LUAD and whether exLR could serve as a biomarker in early LUAD cases remain unknown” should be rephrased since several studies addressing the “exLR” in lung cancer were published, including ones quoted in the discussion (yet not limited to them).

Reply: Thank you for your helpful comment.

As you said, there are indeed related studies published before, such as reference 31 and 32(Xian et al., 2020; Zhang et al., 2017). We changed our statement “However, the profile of EV long RNA (exLR) in LUAD and whether exLR could serve as a biomarker in early LUAD cases remain unknown” to “However, there are few studies on the profile of EV long RNA (exLR) in LUAD and whether exLR could serve as a biomarker in early LUAD cases”.

6) The content of supplementary Table S1 does not match its description in the text (lines 239-242 page 7). Detailed information about differentially expressed RNA species is missed.

Reply: Thank you for your helpful comment.

In the previous version, Table S1 was omitted. We added Table S1 in this version.

7) Minor points:

- Figure 1H is not described in sufficient detail.

Reply: Thank you for your helpful comment. We have added a description of Figure 1H. The green part is mRNA, and the gray part is LncRNA.

- Font size is too small for the graphics.

Reply: Thank you for your helpful comment. We have redrawn the Figure 2D, Figure 2E, Figure 5A, and Figure 5C.

- Figure 5A&B is mislabeled in the text (lines 276-278 page 8).

Reply: Thank you for your helpful comment. Figure 5A&B is mislabeled in the text and we relabeled them.

Reviewer B

This manuscript entitled "Plasma extracellular vesicle long RNA profiling identifies a diagnostic signature for stage I lung adenocarcinoma." By Guo et. al. aims to identify potential diagnostic signatures for early LUAD. Specifically, the author used a case-control analysis to recruit 110 participants including LUAD, BPN, and HC for the study design. RNA-seq was performed for isolated sEVs and 8 exLR markers were identified to distinguish LUAD, NC, and HC with high AUC. Generally, the idea, results, and applications of this study are very important and interesting. However, in the mention of some methodological issues and data illustration consideration, the author should clarify several points carefully.

1. First of all, why the author selected sEVs as starting material for this study? Were these 8 signatures specifically presented in sEVs rather than other parts such as cell free nucleic acid. Therefore, I think the author should confirm whether these 8 signatures can be detected with

the same prediction power in a addition to those from sEVs.

Reply: Thank you for your helpful comment.

As tumor biomarkers, sEVs are more stable than cell free nucleic acid. The main reason is that sEVs have a double-layer membrane structure to prevent the contained RNA from being degraded. In the revised manuscript, qPCR was used to evaluate the performance and stability of these 8 exLRs. We found that there were 6 exLRs (NFKBIA, NDUFB10, ARPC5, SEPTIN9, H4C2, and lnc-PLA2G1B-2:3) with upregulated expression in LUAD patients compared to NC controls ($P < 0.05$). Moreover, we also verified that in sEVs enriched fraction NFKBIA, NDUFB10, SLC7A7, ARPC5, SEPTIN9, HMGN1, H4C2, and lnc-PLA2G1B-2:3 could not be degraded by the pretreatment of Proteinase K and RNase A. Those results suggested that these 8 exLRs were all well protected by sEV membrane in sEVs enriched fractions.

2. The author should mention that the criteria for selecting training set and internal validation set.

Reply: Thank you for your helpful comment.

Although the integration of data is carried out together, in fact, samples (especially plasma samples of tumor patients) are collected by different doctors at different time periods (training set, internal validation set and external validation set). Information about the three cohorts is relatively independent and not easy to leak. We added this point in the *Materials and Methods* section.

3. All RNA-seq data should be deposited in the public database such as Gene Expression Omnibus with an accession number.

Reply: Thank you for your helpful comment. All raw RNA-sequencing data have been deposited in the National Genomics Data Center (NGDC) under the accession code HRA002156 (https://bigd.big.ac.cn/gsa-human/browse/HRA002156). The deposited and publicly available data are compliant with the regulations of the Ministry of Science and Technology of the People's Republic of China. The raw sequencing data contain information unique to individuals and are available under controlled access.

4. The data quality of Western blot should be improved. Two issues need to be revised. Firstly, why is the size of each marker different in control cell lysate and in sEVs enriched fraction? Secondly, the image of Calnexin is over cropped. It is better to show more wide range of membrane.

Reply: Thank you for your helpful comment. The different positions of the same protein bands in cells and exosomes may be due to the different glycosylation and phosphorylation of the protein in exosomes and cells, or the distribution of different isoforms formed by alternative splicing in exosomes and cells. This is the main reason for low quality of Western blot. In the revised manuscript, we replaced Figure 1C.

5. According to the result, the author claimed a median read count for each sample is 35.3 million mapped reads (line 211). This may be different with Figure 1D and the author should check it.

Reply: Thank you for your helpful comment. In this study, we generated a median read count of 35.3 million mapped reads for each sample. However, Figure 1D depicts the number of genes we detected, not the median read count.

6. What does the color represent in Figure 1H?

Reply: Thank you for your helpful comment. We have added a description of Figure 1H. The green part is mRNA, and the gray part is LncRNA.

7. Figure 1 is the overview for QC of experiments such as sEVs isolation and RNA-seq. Is it possible to more confirm that RNA for sequencing is purely from sEVs without cell free nucleic acid contamination? The author may confirm by treating sEVs fraction with/without RNase or detergent.

Reply: Thank you for your helpful comment. In the revised manuscript, we verified that in sEVs enriched fraction NFKBIA, NDUFB10, SLC7A7, ARPC5, SEPTIN9, HMGN1, H4C2, and lnc-PLA2G1B-2:3 could not be degraded by the pretreatment of Proteinase K and RNase A. Those results suggested that these 8 exLRs were all well protected by sEV membrane in sEVs enriched fractions.

8. The author claimed “a total of 117 exLRs were identified as being differentially expressed in LUAD compared to noncancerous control cases (Table S1)” (line 239-240), however, it is difficult to identify these 117 exLRs in Table S1.

Reply: Thank you for your helpful comment.

In the previous version, Table S1 was omitted. We updated Table S1 in this version.

9. Table S2 is missing.

Reply: Thank you for your helpful comment.

We added Table S2 in this version.

10. The image quality of Figure 2D is poor. The size of words is too small to read. In addition, I am curious why only LUAD and HC in Figure 2D while HC, BPN, and LUAD in Figure 2A-C?

Reply: Thank you for your helpful comment. We have redrawn the Figure 2D. Non-cancerous cotrol (NC) is equal to the sum of healthy control (HC) and benign pulmonary nodule (BPN). As for Figure 2D, if we combine HC and BPN, the differential genes will be more clearly distinguished.

11. I am confused why the value of FDR is much greater than our consensus (Table 2)? In addition, what is the criteria of fold change for selecting these 8 genes? The author should clarify it.

Reply: Thank you for your helpful comment.

The origin of exosomal RNA is complex. Compared with the RNA in the tissue, the differences in the RNA in the exosomes are diluted and the degree of differential expression is smaller. So, we didn't use the strict FDR standard. Instead, a more relaxed P-value criterion was used, and in subsequent marker screening, lasso was used to screen subsequent markers. There are also several similar studies that have also used P-value for marker screening (Min et al., 2019; Yuwen et al., 2019).

12. In the mention of Figure 4, it is very important to detailly disclosure the model of "d-signature". How to draw ROC curve based on these 8 markers and what is the algorism to predict LUAD? How does the logistic regression work? This is a criteria issue that should be addressed.

Reply: Thank you for your helpful comment.

Detailed parameters of the model and the cutoff value:

Z = (-15.439212538296 + 0.276596940044276*NFKBIA + 0.116677939851743*NDUFB10 + 0.287717146886391*SLC7A7 + 0.10254802626883*ARPC5 + 0.152102286464288*SEPTIN9 + 0.43111045527139*HMGN1 + -0.00489333268680212*H4C2 + -0.226760347812566*lnc.PLA2G1B-2:3)

Probability=e^Z/(1+e^Z)

Cutoff value: 0.21

Probability is taken as the predicted value of the final model, if Probability > Cutoff, the sample is cancer patient.

13. There are several missing labels in the result such as Figure 6A should be Figure 5A (line 276), Figure 6B should be Figure 5B (line 27).

Reply: Thank you for your helpful comment. Figure 5A&B is mislabeled in the text and we relabeled them.

14. The author should perform some validations to confirm the result from RNA-seq by using independent platforms such as QPCR.

Reply: Thank you for your helpful comment. In the revised manuscript, qPCR was used to validate the performance and stability of these 8 exLRs. We found that there were 6 exLRs (NFKBIA, NDUFB10, ARPC5, SEPTIN9, H4C2, and lnc-PLA2G1B-2:3) with upregulated expression in LUAD patients compared to NC controls ($P < 0.05$).

15. The strength of discussion is weak. It should be discussed deeper. It is believed that numerous literatures mentioned about how these 8 signatures get involved in tumorigenesis. The author should more discuss why these 8 signatures prefer to be presented in sEVs or these

8 signatures were presented not only in sEVs. This is why the author selects sEVs as a target to identify potential biomarkers for early LUAD diagnostics.

Reply: Thank you for your helpful comment. In the *Discussion* section, we added some studies on the relationship between these 8 exLRs and cancer development. However, there are relatively few studies on the relationship between these 8 genes and EVs.

References:

Huang, C., Qu, X., and Du, J. (2019). Proportion of lung adenocarcinoma in female never-smokers has increased dramatically over the past 28 years. Journal of thoracic disease 11, 2685.

Min, L., Zhu, S., Chen, L., Liu, X., Wei, R., Zhao, L., Yang, Y., Zhang, Z., Kong, G., and Li, P. (2019). Evaluation of circulating small extracellular vesicles derived miRNAs as biomarkers of early colon cancer: a comparison with plasma total miRNAs. Journal of extracellular vesicles 8, 1643670.

Xian, J., Su, W., Liu, L., Rao, B., Lin, M., Feng, Y., Qiu, F., Chen, J., Zhou, Q., and Zhao, Z. (2020). Identification of Three Circular RNA Cargoes in Serum Exosomes as Diagnostic Biomarkers of Non–Small-Cell Lung Cancer in the Chinese Population. The Journal of Molecular Diagnostics 22, 1096-1108.

Yuwen, D., Ma, Y., Wang, D., Gao, J., Li, X., Xue, W., Fan, M., Xu, Q., Shen, Y., and Shu, Y. (2019). Prognostic role of circulating exosomal miR-425-3p for the response of NSCLC to platinum-based chemotherapy. Cancer Epidemiology and Prevention Biomarkers 28, 163-173.

Zhang, R., Xia, Y., Wang, Z., Zheng, J., Chen, Y., Li, X., Wang, Y., and Ming, H. (2017). Serum long non coding RNA MALAT-1 protected by exosomes is up-regulated and promotes cell proliferation and migration in non-small cell lung cancer. Biochemical and biophysical research communications 490, 406-414.

Zhou, F., and Zhou, C. (2018). Lung cancer in never smokers—the East Asian experience. Translational lung cancer research 7, 450.