Peer Review File
Article information: https://dx.doi.org/10.21037/tlcr-24-696

Reviewer A:

Comment:

The authors propose a few genes as prognostic factors in lung adenocarcinoma patients.

The major concern with this manuscript is that it solely relies on analysis of data from publicly available datasets, without any in-house validation in clinical samples.

It is also debatable whether it is possible to reproduce the results presented by the authors from what is provided in the material and methods section.

The only provided wet-lab experimental result is displayed in Figure 10. However, this figure lacks quantification of the signals and is therefore difficult to interpret.

In order for the proposed set of genes to potentially be of any value, authors would need to, at the very minimum, validate the uncovered set of 30 genes, presented in Figure 1, in an in-house cohort of clinical lung adenocarcinoma vs normal samples. Without such validation, it is impossible to assess the impact of the proposed findings.

Reply: Thank you very much for your valuable comments and constructive criticism. We acknowledge that our study primarily relies on the analysis of publicly available datasets, which indeed may introduce some limitations. Although we have conducted thorough analysis based on these data, we recognize that the inherent constraints of these databases might have influenced our findings. As for the experimental validation, we fully agree with the reviewer on the importance of validating the identified gene set through in-house experiments. The experimental data we presented, particularly in Figure 10, is limited and lacks signal quantification, making the results harder to interpret. We have reanalyzed and quantified the data in Figure 10 to provide clearer and more reliable insights. As for the experimental validation, we fully agree with the reviewer on the importance of validating the identified gene set through in-house experiments. The experimental data we presented, particularly in Figure 10, is limited and lacks signal quantification, making the results harder to interpret. We will reanalyze and quantify the data in Figure 10 to provide clearer and more reliable insights. We plan to perform further validation of the differential expression of these genes using clinical samples from our laboratory, specifically comparing lung adenocarcinoma tissues with normal tissues. This validation will help better assess the clinical relevance of our findings and address the current limitations. Once again, we are grateful for your insightful comments, which will significantly contribute to improving our manuscript.

Changes in the text: We have revised Figure 10. We have performed statistical analysis of PAICS and DARS2 protein expression differences between normal lung

tissues and LUAD specimens and results were shown in Figure 10. (see Page 9 line 269-271; Page 28 line 715-721) We have added the shortcomings of the research in the discussion and started the next research plan. (see Page 15 line 454-459) Although we have conducted an in-depth analysis based on the available data set, it is undeniable that the limitation of the database may affect the robustness of the results of this study. Therefore, we expect that more real-world studies or prospective clinical trials can improve the above data and conduct a more comprehensive analysis to further verify the results of this article.

Reviewer B:

Comment:

The authors reported "the expression of CENPL, DARS2, and PAICS would determine the prognosis of pulmonary adenocarcinoma".

At first, it is not appropriate to examine prognosis and PFS for eligible patients who have undergone a variety of treatments. Reliable prognostic factors cannot be identified in this study.

Please describe how analyze the drug sensitivity in detail.

Line 245-249 should be described at Methods part.

Evaluation method of immunohistochemistry should be needed.

Reply: Thank you very much for your thoughtful comments. We recognize the limitations of conducting a study based solely on database analysis. Although we have performed an extensive analysis using the available datasets, we acknowledge that database limitations may impact the robustness of our findings. We have addressed these limitations in the discussion section and outlined our future research plans to further validate and strengthen our conclusions. (see Page 15, lines 454-459) Regarding your concern about prognosis and PFS, we would like to clarify that the eligible patients included in this study did not receive any treatments prior to their enrollment. Additionally, we have provided further details on how drug sensitivity was analyzed, ensuring clarity in our methodology. We also appreciate your suggestion to include the information from lines 245-249 in the Methods section. This has been done in the revised manuscript. Lastly, we have now added a detailed description of the immunohistochemistry evaluation method, which can be found both in the revised Methods section and in Figure 10. Thank you again for your valuable feedback, which has significantly helped us improve the quality and clarity of our manuscript.

Changes in the text: We have added the shortcomings of the research in the discussion and started the next research plan. (see Page 15 line 454-459) Although we have performted an in-depth analysis based on the available data set, it is undeniable that the limitation of the database may affect the accuracy of the results. Therefore, we

expect that more real-world studies or prospective clinical trials could improve the above data and conduct a more comprehensive analysis to further verify the results of this study. We have added details on how to analyze the drug sensitivity. (see Page 8 line 221-235) Evaluation method of immunohistochemistry has been added and shown in Figure 10. (see Page 9 line 269-271; Page 28 line 715-721)

Reviewer C:

Comment:

Although authors aimed to describe possible new tumor biomarkers for lung adenocarcinoma, in the main text there was no mention of common oncogenic drivers in the study sample, as well as their frequency and co-mutations. It is highly recommended that not only baseline features such as age and gender are used to predict tumor responses and its prognosis. Are these data available? Did the authors adjust the multivariate Cox analysis to common driver mutations? If not, would the authors believe this might have affected the results?

Also, most of the drugs listed in the paper as positively correlated with the genes enlisted in the main text are not used to treat lung cancer. Did the authors managed to assess proper and currently recommend therapies for lung cancer treatment?

Reply： Thank you very much for your valuable comments. As this study is based solely on publicly available databases, our primary goal was to identify potential new prognostic biomarkers for lung adenocarcinoma using data from multiple sources. In future studies, we plan to use data from our center and real-world cohorts to further validate and improve the accuracy of our findings. Regarding the absence of information on common oncogenic drivers, such as their frequency and co-mutations, we acknowledge the importance of these factors in predicting tumor responses and prognosis. Unfortunately, due to the limitations of the databases used, we were unable to obtain the relevant data on common oncogenic drivers and co-mutations. Consequently, we did not adjust the multivariate Cox analysis for these factors in the current study. While we recognize that the inclusion of such data could potentially refine the results, we believe the main conclusions of our study remain robust. We will prioritize the inclusion of these variables in future analyses when more comprehensive datasets become available. As for the drug sensitivity analysis, we recognize that many of the drugs listed are not currently used in lung cancer treatment. However, several of these compounds from the GDSC and CTRP databases are small molecules that could represent novel therapeutic agents. Our intention was to explore potential treatments beyond the scope of established lung cancer therapies, as this could uncover new possibilities for treatment. We have now included additional details about the drug sensitivity analysis in the revised manuscript.Once again, we are grateful for your

insightful feedback, which has helped us clarify and improve the manuscript.

Changes in the test: We have added details on how to analyze the drug sensitivity. (see Page 8 line 221-235)

Reviewer D:

Comment:

The study aims to provide better biomarkers of LUAD to facilite the diagnosis and care of LUAD patients. It is well designed and well written. It nonetheless only relies on prediction based on mRNA expression data.

It is well appreciated that mRNA and protein expression are often not correlated in tumours. It would be thus interesting to see whether IHC experiments also support the model and the relationship with clinicopathological data. A database of proteomic data in LUAD is available for that : https://ualcan.path.uab.edu and may match and confirm mRNA expression relationships.

It would also be nice to test, at least in vitro, the relationship with compounds evaluated and highlighted in the manuscript. At least, 1 or 2 as a proof of concept that these predictions hold true.

Reply: Thank you very much for your insightful comments and suggestions. We appreciate your recognition of the importance of validating our findings beyond mRNA expression data. As you pointed out, mRNA and protein expression levels can indeed differ in tumors, and we fully agree that validating our results through immunohistochemistry (IHC) and proteomic data would strengthen our study. In response to your suggestion, we plan to further investigate the expression differences and molecular mechanisms of the identified genes between normal tissues and lung adenocarcinoma tissues. The LUAD proteomic database you mentioned (https://ualcan.path.uab.edu) will be an invaluable resource for us to confirm the relationship between mRNA and protein expression, and we intend to incorporate this analysis in our future work. Additionally, we agree that in vitro validation of drug-gene interactions is crucial. In our follow-up studies, we will evaluate the drug sensitivity of the identified genes, and we plan to begin with testing 1 or 2 key compounds as proof of concept, to confirm the predictive validity of our findings.

Changes in the test: We have performed statistical analysis of PAICS and DARS2 protein expression differences between normal lung tissues and LUAD specimens and results were shown in Figure 10. (see Page 9 line 269-271; Page 28 line 715-721)