Peer Review File

Article information: https://dx.doi.org/10.21037/tlcr-23-16

Reviewer A

The present work shows a very interesting proposal on a supplementary molecular classification in pN2 lung adenocarcinomas. Authors have performed a very complete and interest analysis and the molecular classification of tumors is and will be fundamental in the treatment of patients. However, I have a series of questions that I would like to comment on:

Comment 1: In the methods section, the number of samples used in some experiments, such as IHC, is missing.

Reply 1: Thank you for your suggestions. To verify our findings, we included samples from 45 patients with lung adenocarcinoma of pN2 in our center for immunohistochemical staining. According to your suggestion, we added the number of patients enrolled in the section of Methods (see Page 5, line 12).

Comment 2: Some of the bioinformatics methods should be explained in a more didactic way so that researchers who are not experts in the field can understand it.

Reply 2: Thank you for your professional advice. Bioinformatics analysis is widely used in various aspects of cancer research because it can provide rich and detailed data for cancer research. Based on this, our team members have undergone rigorous and standardized bioinformatics training, and all the methods we employ are evidence-based(1-3). Therefore, in order to enable our analysis to be repeated by subsequent interested researchers, the methodology used is approved, and all the methodological terms are technical terms.

Comment 3: In Table 1, the total data of types N2 A and B do not match.

Reply 3: We apologize for our negligence. We have checked all the data in Table 1 and represented the final data in Table 1 (Page 23).

Comment 4: The authors have made an exhaustive analysis with the available data, but I think it is necessary a comparison of the results with the histopathological characteristics of adenocarcinomas and with the driver mutations associated with each tumour. Both, two fundamental characteristics in the current diagnosis of lung AD.

Reply 4: Thank you for your suggestion. In this study, based on transcriptomic information of patients with pN2 lung adenocarcinoma, we divided these patients into low-risk subtypes of pN2-A and high-risk subtypes of pN2-B, and analyzed the correlation between molecular typing and clinicopathological features (see Table 1). Meanwhile, we analyzed the correlation between molecular typing and driver gene mutations (see Figure S1A,1B).

Reviewer B

Comment 1: pN2-A and pN2-B sub-groups from the abstract through the text lack definition.

Reply 1: According to your suggestion, we have defined the two types. Patients with pN2 LUAD were classified into two distinct molecular categories (pN2-A and pN2-B) based on transcriptome information, pN2-A and pN2-B represent low-risk and high-risk patients, respectively (see Page 2, line 18-20).

Comment 2: What are the molecular characteristics of each sub-group. Do they contain different driver mutations, or have differences in CD8 T cell infiltrations, differences in Treg enrichment?

Reply 2: Thank you for your advice. In the present study, based on transcriptomic information of patients with pN2 lung adenocarcinoma, we divided these patients into low-risk subtypes of pN2-A and high-risk subtypes of pN2-B, and analyzed the correlation between molecular typing and driver gene mutations (see Figure S1A,1B). To comprehensively characterize the cell composition and immune microenvironment of different types of pN2-LUAD, we collected samples from two patients with pN2-A (DFS >60 months) and pN2-B (DFS <12 months), respectively, for snRNA-seq. Notably, T/NK cells were enriched in pN2-A patients, while endothelial cells were enriched in pN2-B patients. To explore the role of different T/NK cell subtypes in pN2-LUAD molecular typing, we clustered T/NK cells into five subtypes, including 2 subtypes of Helper CD 4+ T cells (T/NK-C0 and T/NK-C4), 2 subtypes of cytotoxic immune cells (CD8+ T or NK) (T/NK-C1 and T/NK-C5), and 1 subtype of Tregs (T/NK-C2), as shown in Figure 4A. GESA confirmed that cytotoxic immune cells (T/NK-C1 and T/NK-C4) mainly activated the T-cell receptor and chemokine signaling pathways to exert an immune surveillance function (Figure 4B). The cell-cell interaction network also confirmed that cancer cells primarily communicate with cytotoxic immune cells rather than Treg cells (Figure 4C).

Comment 3: It is difficult to understand the clinical application of the method. Do you believe that it is a validated method to predict recurrence-free survival? Is it superior to detection of ctDNA?

Reply 3: The molecular classification of pN2 LUAD has good clinical guiding significance. First, we found that pN2-B patients had a markedly shorter DFS than both the pN0 ($P<0.001$) and pN1 ($P=0.019$) patients. Also, pN2-A and pN1 patients had a similar DFS ($P=0.523$). Compared with pN0 patients, pN2-A patients did not exhibit significantly different results ($P=0.242$), as shown in Figure 1F. Similar results were observed for OS (Figure 1G). Thus, for patients with pN2-A, the treatment pattern is consistent with patients with pN0, while

patients with pN2-B may require multidisciplinary treatment. In addition, for patients with different subtypes of LUAD, the characteristics of tumor microenvironment (TME) are different, suggesting that there are differences in the efficacy of immunotherapy. Because of this, this molecular typing has clinical guiding value.

Reviewer C

Comment 1: First, the title is unclear and inaccurate, which needs to indicate the development and validation of gene-based classification model.

Reply 1: According to your suggestions, we have optimized the title of the study as "Development and validation of a gene-based classification model for pN2 lung adenocarcinoma"(see Title page).

Comment 2: Second, the abstract needs some revisions. The background did not indicate the knowledge gap on the molecular subtype of LUAD and the clinical significance of this research focus. The methods need to indicate the clinical factors, genes, and prognosis outcomes in the databases, how the gene-based classification model was developed, and statistical methods for validating the model. The results need to quantify the findings by using detailed statistics and accurate P values, such as the HR value and the AUC values. The conclusion needs more detailed comments for the clinical implications of the findings.

Reply 2: According to your suggestions, we have detailed description of relevant methods in the section of Abstract, and quantitative presentation of the resulting data (see Page 2 line 10-25).

Comment 3: Third, the introduction of the main text needs to have a brief review on the methodology of the development of the molecular classification model in lung cancer and methods for assessing the validity of the classification model. Insights are needed for clarifying the clinical significance of this research focus.

Reply 3: As described in the section of Introduction, the staging of TNM to guide clinical practice is currently challenged, especially for pN2 stage LUAD, whose prognosis and treatment are highly heterogeneous. Previous studies have focused on the anatomical site of lymph node metastasis, but these patients still cannot be clearly categorized. Based on this, our study aims to establish molecular typing models by using transcriptome information to explore the internal molecular mechanism of pN2 stage lung adenocarcinoma and provide basis for individualized diagnosis and treatment of such patients.

Comment 4: Fourth, the methodology of the main text needs to indicate the research design, describe the clinical variables and the prognosis outcomes in the databases, and describe the

threshold AUC values, sensitivity and specificity for a good classification model. Details of how the classification model was developed should be described in detail.

Reply 4: Thank you for your professional advice. Since our data comes from TCGA-LUAD, GSE68465 and the database of our center, there is a lot of data information. In order to make the expression clear and efficient, we present many data objectively in the charts. According to your suggestion, we have described the relevant data of the molecular model in the manuscript(see Page 11 line 18-20).

Comment 5: Finally, please consider to cite the below relevant paper: Hao X, Li W, Li W, Gu M, Wang Z, Nakahashi K, Antonoff MB, Adachi H, Zhou S, Xu S. Re-evaluating the need for mediastinal lymph node dissection and exploring lncRNAs as biomarkers of N2 metastasis in T1 lung adenocarcinoma. Transl Lung Cancer Res 2022;11(6):1079-1088. doi: 10.21037/tlcr-22-207

Reply 5: According to your suggestions, we have made a comprehensive reading of the literature you mentioned(4) and cited it in the manuscript.

References:

1. Xiong Y, Lei J, Zhao J, et al. A gene-based survival score for lung adenocarcinoma by multiple transcriptional datasets analysis. *BMC CANCER* 2020; **20**:1046.
2. Xiong Y, Feng Y, Qiao T, Han Y. Identifying prognostic biomarkers of non-small cell lung cancer by transcriptome analysis. *CANCER BIOMARK* 2020; **27**:243-50.
3. Xiong Y, Lei J, Zhao J, et al. Gene expression-based clinical predictions in lung adenocarcinoma. *Aging (Albany NY)* 2020; **12**:15492-503.
4. Hao X, Li W, Li W, et al. Re-evaluating the need for mediastinal lymph node dissection and exploring lncRNAs as biomarkers of N2 metastasis in T1 lung adenocarcinoma. *Transl Lung Cancer Res* 2022; **11**:1079-88.