Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-22-909

Reviewer A

Comment A-1:

Please define NSCLC, PUMCH, LUAD, LUSC in the Abstract, and define PUMCH in line 101 in the Introduction.

Reply A-1:

We thank the reviewer for providing this suggestion to make these abbreviations clearer. We provided full names for these abbreviations when they were first mentioned.

Changes in the text A-1:

1. Page 2, line 24

We replaced "NSCLC" with "non-small cell lung cancer (NSCLC)".

2. Page 2, line 32

We added the following description for the patients:

"who underwent surgery in Peking Union Medical College Hospital (PUMCH)"

3. Page 2, line 38

We replaced "LUAD" with "lung adenocarcinoma (LUAD)".

4. Page 2, line 39

We replaced "LUSC" with "lung squamous cell carcinoma (LUSC)".

5. Page 6, line 114

We replaced "PUMCH" with "Peking Union Medical College Hospital (PUMCH)".

Comment A-2:

The resolution of figures 3 to 6 is not good making them very hard/impossible to read.

Reply A-2:

We are sorry for this inconvenience, there might be something wrong when our uploaded files merged into one pdf. file. This time we will carefully check the resolution and make sure the figures are clear to read.

Changes in the text A-2:

No edits have been made under this comment.

Reviewer B

Comment B-1:

This study aims to develop a target sequencing panel suitable for Chinese NSCLC. So, it may be better to submit this MS to a Chinese domestic journal. It is true that the fact is important that there is a difference between Chinese and Western populations, is this MS useful to other readers than Chinese? If so, please describe the reason. If this MS is useful only for the Chinese population, please make this point clear in the text.

Reply B-1:

This is an important point, we appreciate the reviewer raising it and here is our clarification. We think our research is interesting not only for Chinese readers, but also for worldwide readers who study the driver mutations of NSCLC, mainly due to the following two reasons.

First, although we performed the target panel sequencing in a Chinese NSCLC cohort, the results can reflect the mutational landscape of the East Asian NSCLC population due to the genetic similarity among East Asian countries. Given the fact that the total number of lung cancer patients in East Asia accounts for 40% - 50% of the world's lung cancer patients, our study could make great contributions to the global lung cancer research field.

Second, our study was not limited to using this panel to detect driver mutations in Chinese NSCLC patients, we also conducted more comprehensive analysis through inter-subtype and inter-cohort comparison, especially the comparison between our in-house cohort with OrigiMed cohort and TCGA cohort further revealed the differences of driver mutations and downstream pathway alterations, which could provide useful insights for the understanding of NSCLC tumor heterogeneity.

Overall, we believe that our study will be of interest to worldwide readers and suitable for international journals like Journal of Thoracic Disease. We also thank the reviewer for questioning that, and we have made the following modifications in Discussion section to make it clearer.

Changes in the text B-1:

1. Page 18, line 381

We added the following sentence to emphasize the impact of our study to the whole NSCLC research field:

“Moreover, the comparison between our in-house cohort with OrigiMed cohort and TCGA cohort further revealed the inter-cohort similarity and difference of driver mutations and downstream pathway alterations, which could provide useful insights for the understanding of NSCLC tumor heterogeneity.”

Comment B-2:

Many undefined abbreviations are used in Abstract, which prevents readers from quick understanding the content by reading Abstract.

Reply B-2:

We thank the reviewer for providing this suggestion to make these abbreviations clearer. We provided full names for these abbreviations when they were first mentioned.

Changes in the text B-2:

1. Page 2, line 24

We replaced “NSCLC” with “non-small cell lung cancer (NSCLC)”.

2. Page 2, line 32

We added the following description for the patients:

“who underwent surgery in Peking Union Medical College Hospital (PUMCH)”

3. Page 2, line 38

We replaced “LUAD” with “lung adenocarcinoma (LUAD)”.

4. Page 2, line 39

We replaced “LUSC” with “lung squamous cell carcinoma (LUSC)”.

Reviewer C

Comment C-1:

The authors argue strongly about the need for a unique NGS targeted panel for Chinese patients. But this is not borne out by the evidence in the literature or by their conclusions. The paper references (ref 17,18) show that the difference between Chinese and caucasion population is highlighted by whole genome sequencing, not by targeted panel. In addition, the identified difference was attributed to mutational signatures in non-coding regions in Chinese patients that suggest a role of infiltrating inflammatory cells. The mutated genes in coding regions tend to be the identical between the two populations. The difference between the two populations in the coding genes turns out to be the mutation distribution or the mutational frequency of these genes i.e. EGFR mutations are higher in Chinese but the genes mutated tend to be very similar. This is borne out by the fact that the genes in their targeted panel are not very unique to what is in the literature.

Reply C-1:

The reviewer pointed out a very important fact that the mutational spectrum differences between Chinese and Caucasian NSCLC patients are mainly reflected by the different mutation frequency of genes, not the mutated genes themselves. We totally agree with this point, and it is also the motivation for us to design our targeted panel due to following reasons.

First, the cost of targeted panel sequencing (TPS) still remains a heavy burden for both the NSCLC patients and the whole medical system, especially for developing countries like China, whole exome sequencing or large gene panels including hundreds of genes may not be applicable to most NSCLC patients. It is necessary to design a clinically cost-effective gene panel, which contains most important genes closely related to cancer treatment and prognosis but excludes redundant genes.

Second, the gene mutational landscape especially the mutation frequency is different between Chinese and Caucasian NSCLC patients, although some targeted panels have already been designed mainly for Caucasian NSCLC patients, it may not be reasonable to directly apply the Caucasian NSCLC patients-based panels to Chinese patients. In order to design a small-size but effective sequencing panel suitable for Chinese patients, the genes should be selected and adjusted according to Chinese NSCLC mutational characteristics. Our study selected 21 driver genes with both high prevalence and clear oncogenic role, designed a sequencing panel for the 21 genes, and validated the robustness of the panel to identify driver mutations and guide targeted therapies for Chinese NSCLC patients.

We thank the reviewer for the comment, to make our statement clear and accurate in the article, in Introduction section, we further highlighted the need for this panel was based on the clinical cost-effectiveness of applying TPS for Chinese NSCLC patients.

Changes in the text C-1:

1. Page 5, line 93

We added the following sentence to emphasize the rationale of our study:

“However, the cost of TPS still remains a heavy burden for both the NSCLC patients and the whole medical system [13, 14], especially for developing countries like China, WES and large gene panels including hundreds of genes may not be applicable to most NSCLC patients. It is important to design a clinically cost-effective

gene panel, which contains most important genes closely related to cancer treatment and prognosis but excludes redundant genes [15]."

2. Page 6, line 106

We added the following sentence to emphasize the rationale of our study:

"In order to design a small-size but effective sequencing panel suitable for Chinese patients, the genes should be selected and adjusted according to Chinese NSCLC mutational characteristics."

Comment C-2:

Authors seem to identify EGFR as a frequently occurring mutation in lung squamous cell carcinoma, having both passenger and driver mutations. This is still a controversial issue, whether EGFR mutations occur in LUSC or whether the specimen tested were actually adenosquamous carcinoma. Though evidence is mounting for its occurrence in LUSC, authors need to address this question and demonstrate that the diagnosis of LUSC in their samples were rigorously confirmed.

Reply C-2:

This is an excellent question and here we would address the reviewer's concern from two aspects.

On the one hand, the diagnosis of LUSC in our samples were rigorously confirmed. First, our samples were all collected from surgically resected gross specimen, not percutaneous or bronchoscopic biopsy, thus the samples used for pathological diagnosis could basically represent the whole tumor. Second, for the enrollment conditions, we only selected tumors with pathological diagnosis of LUAD and LUSC for sequencing and subsequent analysis, and tumors with diagnosis of other rare NSCLC subtypes like adenosquamous carcinoma or larger cell lung carcinoma were excluded. Third, all of our pathological diagnoses were made by a team of professional and qualified pathologists with strict quality control. Therefore, the possibility that the adenosquamous cell carcinoma samples were mixed in the squamous cell carcinoma samples was very small.

On the other hand, we have detected many genetic variants on the *EGFR* gene in LUSC, but most of them were non-protein affecting variants or passenger mutations (Fig. 4B), actually the frequency of important driver mutations was 9%, significantly lower than 27% of that in LUAD (Fig. 5C). LUSC was previously reported to harbor lower *EGFR* mutation frequency than LUAD, about 3%-18% (PMID: 34195081), the mutation frequency in our LUSC cohort was relatively high but within the reasonable range, possibly due to normal bias of random sampling or genuine tumor heterogeneity among LUSC cohorts.

We thank the reviewer for arising this question, to improve the rigor of the manuscript, in Methods section we further emphasized the criteria that only pathologically diagnosed LUAD and LUSC samples were selected for sequencing and downstream analysis, while other rare NSCLC subtypes were excluded from this study.

Changes in the text C-2:

1. Page 6, line 124

We replaced "patients" with "sample collection".

2. Page 7, line 132

We added the following sentence to describe the selection criteria:

"Only samples with definite diagnosis of adenocarcinoma or squamous cell carcinoma were sent for DNA sequencing, while other rare NSCLC subtypes like adenosquamous carcinoma or large cell lung carcinoma were excluded".

Comment C-3:

Authors compare their 21 gene panel with 4 panels previously published. However, it must be noted that these lung panels are in no way comprehensive or representative of currently available lung panels. Therefore, the conclusion from this comparison that their panel is unique is fraught and not consistent with the numerous lung panels in the literature. e.g. commercial targeted panels exist with 50-400 genes that contain all the genes in their 21 gene panel.

Reply C-3:

We agree with the reviewer that these 21 genes themselves were not the novelty of our panel, because they can be contained by a larger panel composed hundreds of genes. As we replied to the reviewer comment C-1, we would like to highlight that the significance and novelty of our panel is that it could be the most suitable panel for many Chinese NSCLC patients when taking into account both clinical benefits and costs.

We are sorry for confusion caused by the statement of novelty when comparing with other 4 panels. Here the rationale of the comparison was because these panels were all small panels composed of 20-50 target genes, with similar size to ours, and the comparison results indicated that we not only included classic driver genes like *TP53*, *EGFR*, but also added some newly reported driver genes like *ROS1*, *RET*.

For this comparison, we removed the description of "novelty" and emphasized that the inter-panel comparison was made under the condition of similar size.

Changes in the text C-3:

1. Page 9, line 183

We added the following description the panels used for comparison:

“which have similar size to ours”.

2. Page 9, line 187

We deleted the description “and novelty”.

Comment C-4:

Authors correlate TP53 mutations with smoking in LUSC. However, in LUAD, TP53 is associated with age and poor differentiation. Nothing is said about its relationship to smoking. Authors need to clarify whether this is actually the case or give some insight of their thinking on this.

Reply C-4:

We thank the reviewer for the comment, and to address the reviewer’s concern, we carefully checked our data and confirmed the accuracy of the results.

For the inconsistent clinical association results of *TP53* mutation between LUAD and LUSC, we speculate that it might be due to the inter-subtype differences of mutation characteristics between LUAD and LUSC so that one significant mutation-clinical factor association in one subtype might be insignificant in the other subtype. On the other hand, the epidemiological analysis about clinical association of driver mutations in LUAD/LUSC requires a large amount of patient data and comparison of results obtained by different research groups. our results were obtained from 134 LUAD patients and 126 patients in a single center, and it might be challenging to make interpretations for all NSCLC patients only based on our own data.

And to make our data more transparent for further verification, we also provided the raw data as Supplementary Table. S5 including the contingency tables and p-values for each driver mutation and clinical factor, optimized the figures Fig. 3E and 4E by making the color reflect the –log2 p-value of each significance test for clinical association.

Changes in the text C-4:

1. We added Supplementary Table.S5.

2. We modified main figures Fig.3E and 4E to show the –log2 p-value of significance test for clinical association.

Comment C-5:

Overall, this is an interesting well written study that attempts to address questions about mutations in lung carcinoma and their correlates with populations, clinical features, signaling pathways.

Reply C-5:

We thank the reviewer for appreciating the significance of our work, and we hope our responses above can address the reviewer's concerns.

Changes in the text C-5:
No edits have been made under this comment.