Peer Review File

Article Information: Available at http://dx.doi.org/10.21037/tlcr-20-612

Comments to reviewer A

Comments: This manuscript discusses prognostic significance of cribriform adenocarcinoma of the lung based on manual evaluation of H&Es and molecular data, clinicopathologic and demographic characteristics of a L-ADC cohort comprising over 1000 cases. This study is a continuation of the authors recent work on L-ADC – with the past work focused on prognostic significance of the micropapillary pattern of L-ADC.
The current study is well designed and adequately powered. Results are valid and sound, and the conclusions reflect the main points of the study. What I think is missing, is the evaluation of concordance in histopathological reading of slides between the two pathologists. In the Discussion the authors noted that the inter-observer agreement was a limitation, but they did not provide any quantitative data to show the extent it exists in this study. The two pathologists should eventually evaluate all H&E slides, and the Cohen's kappa be calculated (or other suitable metric) to measure the concordance between them. Papers showing that the inter and intra rater disagreement in reading L-ADC in H&E slides exists should also be cited. To put this problem in a perspective, in Discussion, the authors should mention that artificial intelligence solutions that are being developed can alleviate this problem and eventually read the H&E slides automatically. Here are some papers to consider:
https://www.nature.com/articles/s41598-019-40041-7
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6365499/

Response to the reviewer:

Thank you for pointing out this important issue. In this study, we determined the subtype with a discussion between two pathologists. At the beginning of the study, we did not reach an agreement to recognize the "cribriform pattern"; thus, this difficulty was found to be the first limitation. This difficulty was due to the fact that the image of the "cribriform pattern" was different among the pathologists. Thus, we went back to previous papers at that time and carefully referred to the pictures of the "cribriform pattern" (training step). After that, the interobserver agreement increased, but we have no data regarding the concordance at that time. As the reviewer addressed, we also agree that it is important to determine the concordance between the pathologists. Therefore, we randomly selected 100 cases in this cohort and calculated the Cohen's kappa value. As a result, the interobserver agreement was fair (kappa =0.621), indicating that the interobserver concordance was not very low between the pathologists. We think this is because we work in the same institute and discuss issues routinely. In addition, unfortunately, we could not find papers dealing with interobserver agreement. Therefore, we would like to conclude that the "training step" to recognize the pattern was the most important one. Further, we completely agree that deep learning techniques are promising to recognize or classify such growth patterns, as the reviewer mentioned. Taken together, we would like to change the paragraph in the discussion section as below;

Discussion section, Page 19, line 1:

There were some limitations in this study, the most notable of which was the interobserver agreement. Although we subclassified Cribri-p from acinar-pattern tumors after careful discussion, we noted some difficulty in distinguishing simple acinar glands from fused-glands because most were continuously present in the same regions. However, we consider that we can overcome interobserver disagreement by training to recognize the pattern carefully because our concordance rate was not very low when analyzing the selected 100 cases (data not shown). In contrast, if the Cribri-ADC is recognized as the sixth histological subtype of L-ADC, it will potentially be more difficult to reach interobserver agreement than it would be with the current scheme. Deep learning techniques may be promising to overcome this difficulty. (Ref. 18, 19) In addition, we also considered the classification by tumor grading to be more optimal than determining the tumor subtype for predicting prognosis (Ref, 20).

18. Wei JW, Tafe LJ, Linnik YA, et al. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. Sci Rep. 2019;9(1):3358

19. Gertych A, Swiderska-Chadaj Z, Ma Z, et al. Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. Sci Rep. 2019;9(1):1483.

Comments to reviewer B

Comment 1: Definition of cribriform growth: In some studies the fuzed gland pattern is interpreted under complex glandular growth. This point should be addressed.

Reply 1: Thank you for your comment. To clarify this point, we added the following sentence to the patients and methods paragraph.

Page 8, Line 1: In some previous studies, the fused-gland pattern was interpreted under a complex glandular pattern. [5, 7]

Comment 2: If I understood correctly, the mutation analyses were used according to the actual clinical practice and no additional analyses were performed for this study?

Reply 2: I apologize for the lack of clarity. In this study, no additional genetic alterations were analyzed. To evaluate the association between Cribri-p-exhibiting tumors and gene alterations, we used the data that we had obtained in our previous studies (Ref. 10-15). In previous studies, some genetic analyses were performed as clinical practice, while the others were performed for the studies. We have added an explanation about these concerns as follows.

Page 8, Line 12:
To evaluate the association between Cribri-p-exhibiting tumors and gene alterations, such as *EGFR*, *KRAS*, *HER2*, *BRAF*, *ALK*, and *ROS1*, we used the data that we had performed in our previous studies.

Comment 3: Large proportion of stage I patients (78.4%) stage II (10.3%), (Stage III 9.3%) and never-smokers (47.1%) Is this distribution reflective national clinical situation or related to this specific center? For example In this European study the distribution was much more even between stages I-III (stage I 39,2%, stage II 21,5%, stage III 36,1%), and the proportion of never-smokers was 9.2%. (Ref. 6)

Reply 3: As the reviewer mentioned, the proportion of Stage I patients in our cohort was much higher and the proportion of smokers in our cohort was lower than that in the European cohort. However, the proportions were comparable to the other data from Japan or the Asian country (Ref. 8, 9). Thus, we consider that this distribution is representative of the national (or Asian) clinical situation. We added this discussion to the Results section.

Page 10, Line 10: The proportion of Stage I patients in our cohort was much higher and the proportion of smokers in our cohort was lower than that in the European cohort.[Ref. 6] However, the proportions were comparable to the other data from Japan or Asian countries.[Ref. 8, 9]

Comment 4: "Our data may be the first to indicate that TKIs should be included in postoperative adjuvant therapy regimens for patients with Cribri-ADC." a statement this strong is unjust.

Reply 4: I understand the reviewer's remark. We have deleted this statement from the paper.

Page 17, Line 8: We deleted the following sentence; "Our data may be the first to indicate that TKIs should be included in postoperative adjuvant therapy regimens for patients with Cribri-ADC."