POINT-BY-POINT RESPONSE TO REVIEWER 3

****

Independent Review Report, Reviewer 3

EVALUATION

This manuscript reported an integrated analysis of 1370 NMIBC transcripts from nine public datasets to combine clinicopathological features and recent molecular subtyping studies, aiming to help filter the patients with a high risk of repeated recurrence or muscle-invasive progression. The authors established a robust immuno-prognostic model dividing NMIBC patients into low-risk versus high-risk progression group, showing good prognostic stratification of progression-free survival (PFS), disease-free survival (DFS), and overall survival (OS) of NMIBC patients. Overall, the study is interesting and has a certain clinical application prospect for bladder cancer. The research scheme is reasonable and the data analysis is credible. This data is promising, but some results are confusing. Comments are as follows.:

Response: We appreciate the time and effort the Reviewer has dedicated to providing positive comments and valuable feedback on our manuscript. We have incorporated most of the reviewer's suggestions regarding ambiguity issues and highlighted the changes within the manuscript. However, for the unanswerable questions stemming from the lack of high-quality survival data or therapeutic response data, we further amended our discussion regarding the study's limitations. Here is a point-by-point response to the Reviewer's comments and concerns.

****

The authors established a robust immune-prognostic model for predicting the progression of NMIBC patients. How does it perform in predicting the response to immunotherapy, which was highly correlated with immune cells?

Response: We appreciate the Reviewer pointing this out. We couldn't agree more with this comment. When designing the study framework, we looked for transcriptomic profiles of bladder cancer patients who had undergone immune checkpoint inhibitor therapy, particularly those datasets that provided response data. We were only able to locate the IMvigor210 RNA-seq dataset (EGAD00001003977), which has been made available under control by Mariathasan et al. (Mariathasan et al., 2018). Unfortunately, the IMvigor210 study is a phase 2 trial investigating the clinical activity of PD-L1 blockade therapy using atezolizumab in patients with locally advanced and metastatic urothelial carcinoma (a/mUC), not non-muscle-invasive bladder cancer (NMIBC). Considering that previous studies have demonstrated the mutational heterogeneity between the NMIBC and MIBC cohorts (Glaser et al., 2017), we considered the IMvigor210 dataset unsuitable for our study. Having only a prognostic but not a predictive value is one of the weaknesses of our model due to the lack of high-quality response data for therapies used by patients with NMIBC. We have discussed this limitation in our work on Lines 378-381.

This study included 1370 NMIBC transcripts data from nine public datasets, was potential cross-dataset batch effect removed and investigated?

Response: We appreciate the Reviewer bringing this issue up here so that we may discuss it furthermore. There is no doubt that the cross-dataset batch effect is a priority problem that must be addressed before moving on to the subsequence analyses. In addition, it is one of the most frequently discussed topics in our interactions with Reviewer 1. We are pleased to conclude here with our trying results under the guidance of Reviewer 1 and what makes our manuscript the way it appears today.

Yes, both cross-platform and cross-dataset batch effects exist among the nine datasets. Our method of 1) analyzing differentially expressed genes (DEGs) based on each dataset independently and combining the results using the well-established p-value combination-based meta-analysis method; and 2) evaluating the tumor-infiltrated immune cells enrichment score using the ssGSEA method as it is a method that can be applied regardless of the platform or dataset sources, allowed us to bypass the aforementioned difficulty without sacrificing the actual prognosis-related signal. Due to the use of the enriched ssGSEA matrix in the modeling process, we no longer have to worry about cross-platform and cross-dataset differences in the original express matrices. The detailed evaluation results can be found in answer to question Major 1 in our first round interaction in Reviewer 1's review panel.

When expanding validation of the model in predicting clinical outcomes. How does it perform in other databases besides DFS in GSE32894 or OS GSE13507?

Response: We wish to thank the Reviewer for this question. Please kindly refer to supplementary table 1, where we summarize the available survival information across nine eligible datasets. In view of the fact that we validated the performance of our model primarily by the significant difference between two risk groups using Kaplan-Meier survival analysis, GSE32894 with available DFS data and GSE13507 with OS data were the only two datasets that could be analyzed. Although we were unable to provide a more comprehensive assessment of our model in this article, other researchers possessing survival time information are welcome to provide further validations using our code repositories on Github or Gitee. In addition, we have provided an explanation in the footnote of Figure 5 regarding why only GSE32894 and GSE13507 were used to validate the model (Lines 608-612). We hope that this result will be less confusing for our readers.

* In the nine eligible datasets, PFS status was assessed in E-MTAB-4321, GSE13507, and GSE32894, while only E-MTAB-4321 provided survival time. DFS status was assessed in GSE32894, GSE13507, and GSE48075, while only GSE32894 provided survival time. OS status was assessed in GSE13507 and E-MTAB-1940, while only GSE13507 provided survival time.

ssGSEA was used in building the immuno-prognostic model based on the score matrix of all nine candidate immune cells. What are the advantages of this model compared to other models based on several genes with other methods?

Response: We apologize for the possible ambiguity. As an initial matter, the ssGSEA was only used to obtain the enrichment matrix of nine candidate immune cells as the input characteristics. The ridge regression model was the method we chose to build the immuno-prognostic model. In order to avoid this misunderstanding, we added the modeling strategy

to the modeling portion of Figure 1. Regarding the question raised by the Reviewer, we have discussed the advantages of our model from two perspectives (Lines 310-325). First, we identified key DEGs for each candidate immune cell by combining results from nine datasets instead of relying on default lists from public tools. Consequently, we believe that this approach would enable us to overcome the limitations caused by the single source bias or non-specific selection of markers in previous studies. Second, we incorporated all nine candidate immune cells into our final model, believing that the combination of prior knowledge about innate and adaptive immune systems contributing to the prognostic differences would result in a more robust model, as it represented both biological and statistical insights. As indicated in our title, this is also the main conclusion of our study. Neither of these advantages would have been achieved without the integrated design of the study, which distinguished us from previous models based solely on genes or other single-dimensional data. A detailed comparison was discussed on lines 363-368.

Reference

Glaser, A. P., Fantini, D., Shilatifard, A., Schaeffer, E. M., and Meeks, J. J. (2017). The evolving genomic landscape of urothelial carcinoma. *Nat Rev Urol* 14, 215–229. doi: 10.1038/nrurol.2017.11.

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGF-β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 554, 544–548. doi: 10.1038/nature25501.