nature portfolio

Peer Review File

Cellular Hallmarks and Aging Clock of the Human Lung Parenchyma

Corresponding Author: Professor Alexander Tsankov

This manuscript has been previously reviewed at another journal that is not operating a transparent peer review scheme. The manuscript was considered suitable for publication without further review at Nature Communications.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have made substantial revisions to the manuscript, including new and revised analyses and the addition of 70 spatial transcription data sets from 22 individuals. The central findings remain similar and largely align with prior studies; the cross-compartment comparisons, spatial data and aging clock model are interesting and informative advances for the field.

Major comments
1. The authors revised analysis and updates to the figures, including the addition of new validation data have largely addressed my prior concerns and comments. I have several additional comments regarding data that have been added at resubmission.
2. Figure 2f – the KRT8 immunofluorescence data are supportive, but the percentage of KRT8+ cells is far higher than plausible for PATS (10-15% of all cells as depicted) – some of the KRT8 is clearly in the airway epithelium. The limited specificity of this protein for PATS in human lung sections, and likelihood his is an overestimate, should be noted.
3. Xenium data – how was segmentation performed? What strategy was used to remove doublets? A dotplot or heatmap with marker genes should be included to support label transfer fidelity.
4. While likely limited by the probeset, it would be informative to assess how the cNMF aging model performs in the Xenium dataset if feasible.
5. The discussion section remains on the speculative side – while this is reasonable given many potential functional relationships that would need validation, couching these findings as transcriptional inference that will need functional testing is needed

(Remarks on code availability)
looks reasonable

Reviewer #2

(Remarks to the Author)
All my previously raised conccerns hahve been addressed properly.

(Remarks on code availability)
-

Point by point response to reviewers' comments

Reviewer #1 (Remarks to the Author):

The authors have made substantial revisions to the manuscript, including new and revised analyses and the addition of 70 spatial transcription data sets from 22 individuals. The central findings remain similar and largely align with prior studies; the cross-compartment comparisons, spatial data and aging clock model are interesting and informative advances for the field.

We appreciate the reviewer's supportive comments about how our study advances the field and for recognizing the substantial efforts we made to further improve this manuscript.

Major comments

1. The authors revised analysis and updates to the figures, including the addition of new validation data have largely addressed my prior concerns and comments. I have several additional comments regarding data that have been added at resubmission.

We would like to thank the reviewer for their detailed and thoughtful review of our original and revised manuscript.

2. Figure 2f – the KRT8 immunofluorescence data are supportive, but the percentage of KRT8+ cells is far higher than plausible for PATS (10-15% of all cells as depicted) – some of the KRT8 is clearly in the airway epithelium. The limited specificity of this protein for PATS in human lung sections, and likelihood his is an overestimate, should be noted.

We agree with the reviewer that while these results are supportive, some of the KRT8+ cells may in fact not be specific to PATS. We have acknowledged this in the results section of the revised manuscript (lines **234-235**).

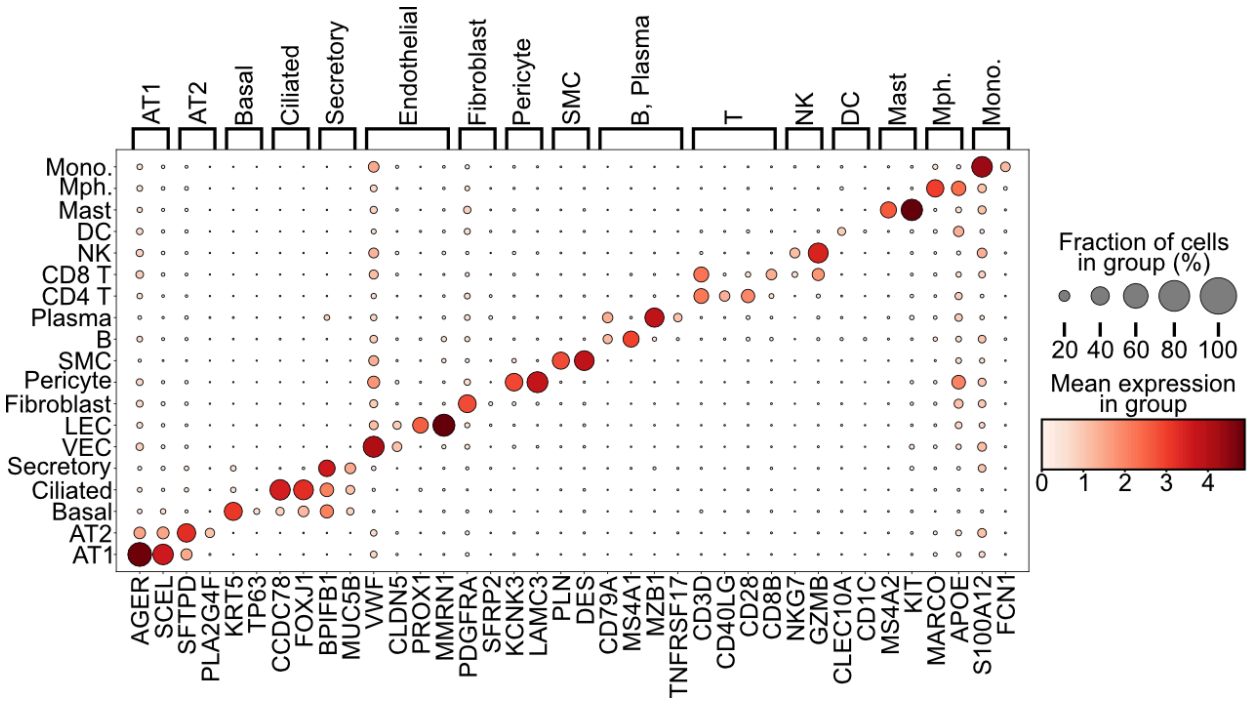
3. Xenium data – how was segmentation performed? What strategy was used to remove doublets? A dotplot or heatmap with marker genes should be included to support label transfer fidelity.

We thank the reviewer for raising these important points. Cell segmentation for our Xenium spatial transcriptomics (ST) data was performed using the 10x Genomics Xenium multimodal cell segmentation pipeline. This approach uses information from multiple imaging channels for segmentation, including DAPI for nuclear staining, membrane markers, and intracellular RNA staining information. The segmentation is prioritized hierarchically, where cells are first segmented using boundary staining when clear membrane information is available. If boundary information is less informative, boundaries are inferred by expansion from the nucleus to the interior RNA stain edge, and if neither is sufficient, a default isotropic nuclear expansion is applied. This multimodal strategy is designed to improve transcript-to-cell assignment relative to nuclear expansion alone. We have explained the cell segmentation strategy in more details in the Methods section of the revised manuscripts (**lines 725-733**)

Occurrence of doublets is a well-defined problem in single-cell RNA-sequencing (scRNA-seq) data, where multiple cells can be encapsulated in one droplet and sequenced together. In Xenium ST data, the main technical challenge is potential misassignment of transcripts arising from diffusion of RNA on the tissue section, imperfect cell segmentation, and other factors. How to correct for this source of noise remains an active area of methodological

development in ST data analysis[1,2]. Since these methods have not yet been benchmarked extensively and the effect on the underlying biological signal is still unknown, we did not apply any such corrections to the underlying Xenium ST data. We are however confident in the conclusions of our study, since we were able to integrate the respective advantages of scRNA-seq (better defined single-cell transcriptomes) and Xenium (spatial information) data to show highly concordant aging trends.

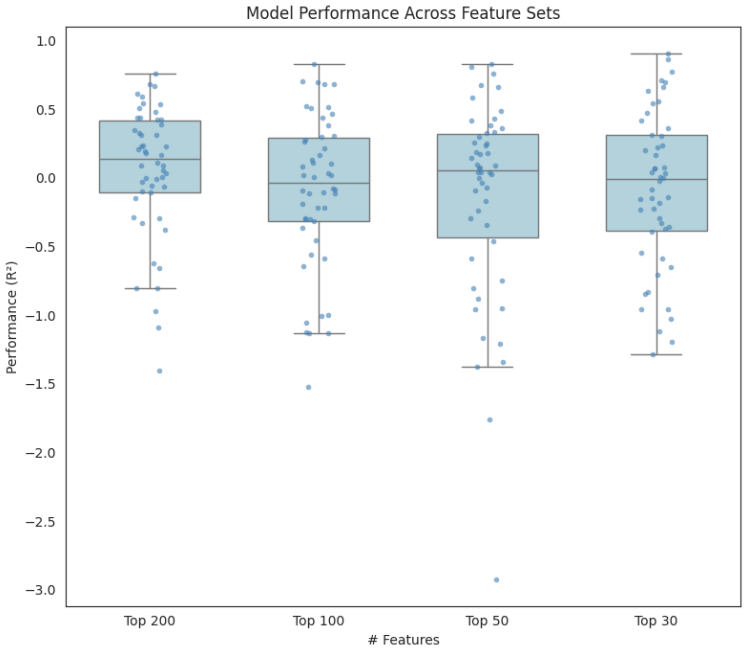
To address the reviewer's request and support the fidelity of our Xenium ST data cell annotation, we have now included a dot plot showing the expression of canonical cell type marker genes (new **Extended Data Fig. 6a**, also shown as **Reviewer Fig. 1**). The strong cell type specific expression patterns observed in this analysis provide additional support for the accuracy of the assigned labels.



Reviewer Figure 1 | Dot plot showing expression of representative markers across each annotated cell type in the Xenium data.

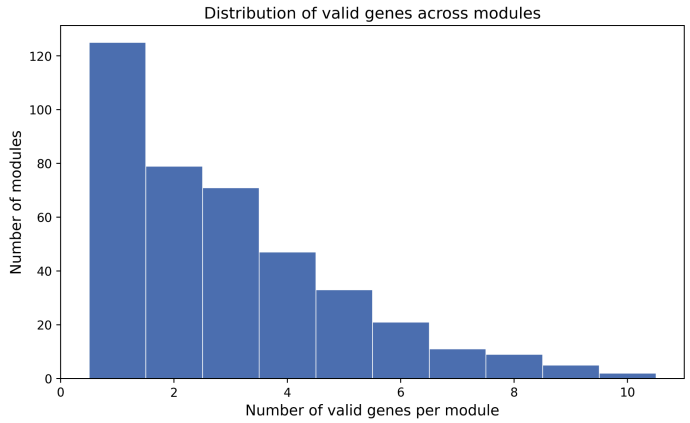
4. While likely limited by the probeset, it would be informative to assess how the cNMF aging model performs in the Xenium dataset if feasible.

We agree that applying the cNMF module aging model to the Xenium dataset is an interesting endeavor, despite the limitations imposed by the targeted probeset. To address this point, we trained our machine learning (ML) model to predict the age using the Xenium dataset, in the same manner as done for the GTEx RNA-seq and HLCA scRNA-seq data in the manuscript (**Figure 7c-d**), and evaluated the R^2 performance (**Reviewer Figure 2**). The performance of the model trained on the Xenium data was substantially lower compared to training on the scRNA-seq data. Specifically, the model achieved a median R^2 of 0.14 when using the top 200 features, and even worse performance for top 100 and top 30 features (**Reviewer Figure 2**).



Reviewer Figure 2 | R^2 performance of XGBoost ML model for the top 200, 100, 50, and 30 features across 50 training runs using the Xenium dataset.

As pointed out by the reviewer, it is possible that this poor performance is driven by limited gene coverage in the Xenium probeset. While each cNMF feature/module contains 25 genes, the overlap between these module genes and the Xenium panel is small. To quantify this, we found that 77 cNMF modules used as features for the scRNA-seq data model did not contain any genes on the Xenium panel. Additionally, the majority of remaining modules are represented by ≤3 genes on the Xenium dataset (**Reviewer Figure 3**), which compromises the ability to accurately compute module scores and parse out feature importance.



Reviewer Figure 3 | Histogram of modules included in model training (≥1 overlapping gene), stratified by the number of valid genes per module.

We note that spatial transcriptomics platforms with broader gene coverage, such as the 5,000-plex 10X Xenium or 10X Visium, would be better suited for evaluating the cNMF aging model, as these platforms would provide more complete representation of gene modules.

5. The discussion section remains on the speculative side – while this is reasonable given many potential functional relationships that would need validation, couching these findings as transcriptional inference that will need functional testing is needed

We agree with the reviewer and have added additional text to the discussion that further functional validation is needed of our transcriptomics findings (lines **661-666**). While we acknowledge this limitation, we also believe that our discussion section highlights the high resource value of our study, providing integrated single-cell and spatial data sets and age-related modules that can be used to generate a plethora of data-driven hypotheses in specific cellular contexts pertinent to lung parenchyma aging, which can be the focus of many future studies by other experimental laboratories.

Reviewer #1 (Remarks on code availability):

looks reasonable

Reviewer #2 (Remarks to the Author):

All my previously raised concerns have been addressed properly.

Thank you for taking the time to assess our revised manuscript.

Reviewer #2 (Remarks on code availability):

-

References Cited:

1. Mitchel, J., Gao, T., Petukhov, V., Cole, E. & Kharchenko, P. V. Impact and correction of segmentation errors in spatial transcriptomics. *Nature Genetics* **58**, 434–444 (2026).
2. Jones, D. C. *et al.* Cell simulation as cell segmentation. *Nature Methods* **22**, 1331–1342 (2025).

Point-by-point response to reviewers' comments

Reviewer #2 (Remarks to the Author):

With respect to your response to concerns raised by Reviewer #1, Reviewer #2 requested that you further expand the discussion on the limitations of relying on transcriptomic findings specifically by outlining a plan for future studies to achieve functional validation.

We thank the reviewer for taking the time to evaluate our point-by-point responses to Reviewer #1. We have now provided additional specific examples of the follow-up experiments to validate our transcriptomics findings in the Discussion section (**lines 676-685**) of the revised manuscript. We have included the edited text in the Discussion section below for convenience:

"Second, additional experiments will be needed to validate the cell type-specific transcriptomic findings reported in this study at the protein-level, including staining of key senescence markers (e.g., CDKN1A, CDKN2A, HMGB1), WNT signaling (e.g., WNT7B, LRP6), T cell chemotaxis (e.g., CXCL9, CXCL10, CXCR3), and T cell activation (e.g., ITGAL, ICAM1) ligands and receptors across age. Third, future experiments to establish causality (e.g., lineage tracing, lung injury, inhibition/rescue) in murine or other models will be required to validate our transcriptomic inference and discern the exact aging mechanisms of PATS, TP53, and WNT signaling in AT2 cells. For example, alveolar cell lineage tracing following lung injury and recovery in young versus old mice can help uncover possible mechanisms underlying accumulation of PATS with age."