nature portfolio

Corresponding author(s): Tsankov, Alexander M.

Last updated by author(s): 06/21/2026

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Published single-cell RNA-sequecned data was used in this study.<br>10x Genoimics Xenium platform for spatial transcriptomics data. |
| Data analysis | R V.4.1.1<br>CRAN<br>https://www.r-project.org/<br>Python V.3.10.14<br>Anaconda<br>https://docs.anaconda.com/anaconda/install/<br>Seurat V.4.0.5<br>Stuart et al24<br>https://satijalab.org/seurat/<br>ggplot2 V3.4.3<br>CRAN<br>https://CRAN.R-project.org/package=ggplot2<br>pheatmap V.1.0.12<br>CRAN<br>https://CRAN.R-project.org/package=pheatmap<br>harmony V.0.1.0<br>Korsunsky et al84<br>https://github.com/immunogenomics/harmony |

cNMF V1.3.4
Kotliar et al35
https://github.com/dylkot/cNMF
Muscat V.1.8.2
Corwell et al85
https://github.com/HelenaLC/muscat
Monocle3 V 1.0.0
Trapnell et al39
https://cole-trapnell-lab.github.io/monocle3/
CellPhoneDB V4.0.0
Garcia-Alonso et al86
https://www.cellphonedb.org/
scRepertoire V2.0.0
Borcherding et al87
https://github.com/ncborcherding/scRepertoire
igraph V1.5.1
CRAN
https://CRAN.R-project.org/package=igraph
XGBoostV2.1.0
Chen & Guestrin88
https://github.com/dmlc/xgboost
scikit-learn V1.5.1
Pedregosa et al89
https://scikit-learn.org/dev/whats_new/v1.5.html
Optuna V3.6.1
Akiba et al90
https://github.com/optuna/optuna

survival R package (https://rdrr.io/cran/survival/)
CellPhoneDB (v2.1.7) (https://github.com/Teichlab/cellphonedb)
pySCENIC (v0.11.2) (https://www.10xgenomics.com/products/loupe-browser)
scanpy (v1.10.3) (https://github.com/scverse/scanpy)
squidpy (v1.1.0) (https://github.com/scverse/squidpy)
scvi-tools (v0.20.3) (https://github.com/scverse/scvi-tools)
stLearn (v0.4.12) (https://github.com/BiomedicalMachineLearning/stLearn)
SOAPy (1.0.1) (https://github.com/LiHongCSBLab/SOAPy)
cell2location (v0.1) (https://github.com/BayraktarLab/cell2location)
pythologist (https://github.com/jason-weirather/pythologist)
Custom analysis code, which is available on Github (https://github.com/TsankovLab/sc_Aging_clock)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The Xenium spatial transcriptomics data generated in this study is accessible through the Cellular Senescence Network (SenNet) consortium online data portal at https://doi.org/10.60586/SNT797.JGHF.784 and at Gene Expression Omnibus (GEO) under accession GSE335761. The processed Xenium object is accessible at 10.5281/zenodo.20144719. The HLCA scRNA-seq data is accessible at https://cellxgene.cziscience.com/collections/6f6d381a-7701-4781-935c-db10d30de293. PBMC data is available at https://www.synapse.org/Synapse:syn49637038/files/ and GTEx data is accessible at https://gtexportal.org/home/downloads/adult-gtex/. All data supporting the findings of this study are available within the paper and its supplementary information files. Source data are provided with this paper.

## Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

| | |
|---|---|
| Reporting on sex and gender | Sex was determined by self-report, where 41% (52/184) of patients in our single-cell cohort , 45.5% (10/22) in spatial cohort, and 31.7% (183/578) of patients in bulk RNA-seq validation cohort were female. |
| Reporting on race, ethnicity, or other socially relevant groupings | Ethnicity was determined by self-report, in single-cell cohort 85.3% (157/184) of patients on the data are white, 7.1% (13/184) are black, 1.6% (3/184) are asian, 1.1% (2/184) are latino, and 1.1% (2/184) are mixed.<br>In spatial transcriptomics cohort 13.6% (3/22) of patients are white, 31.8% (7/22) are black, 31.8% (7/22) are other, and ethnicity for 22.7% (5/22) was unknown. |
| Population characteristics | Figure 1a, 5a, Supplementary Figure 1a, Supplementary Data 1, Supplementary Data 3. |

| | |
|---|---|
| Recruitment | Normal lung autopsy samples and age information was deidentified and obtained through the Mount Sinai biorepository. |
| Ethics oversight | Human lung autopsy samples used for Xenium spatial transcriptomics profiling were obtained via the Biorepository and Pathology CoRE from de-identified deceased individuals, which was exempted by the Institutional Review Board at the Icahn School of Medicine at Mount Sinai. Written informed consent was obtained from the next of kin or legally authorized representatives in accordance with institutional policies, CARE guidelines, and the principles of the Declaration of Helsinki. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | 184 lung parenchyma scRNA-seq samples from normal donor lung tissue and 70 normal lung samples from 22 deceased individuals. We conducted power analysis in Supp Fig. 1d. Sample sizes were determined primarily by the publicly available datasets, and for Xenium, the sample availability and data quality. |
| Data exclusions | For compositional analysis, all samples with < 1,000 total cells were excluded from this analysis, due to greater uncertainty of cell proportions in samples with fewer cells. For XGBoost model generation, modules derived from cell types with less than 400 cells (Goblet, Mesothelium, pDCs, HSCs, Ionocytes, PNECs, Tuft, Hillock-like) were excluded from the feature space for training the model. |
| Replication | Our findings were highly reproducible across three independent cohorts and technologies (HLCA scRNA-seq, GTEx bulk RNA-seq, and Xenium spatial transcriptomics) when using rigorous statistic testing, including correction for multiple hypothesis testing. |
| Randomization | Randomization was not applicable to our study design. We assigned samples to age groups accepted in the literature but in parallel performed rank-based association with age using all samples, both yielding highly reproducible results. |
| Blinding | Pathologist was blinded to age-related hypothesis during the selection of regions of interest in lung tissue samples collected from the Mount Sinai biorepository. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |
| ☒ | ☐ Plants |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | anti-KRT8(DSHB) Clone ID:TROMA-I 1:100 dilution, Lot no. 4/25/24;<br>Donkey anti-Rat IgG Secondary Antibody(Invitrogen) Catalog #A48269 1:1000 dilution, Lot no. YC364986;<br>DAPI(Fisher Scientific) Catlog #D1306 10ug/mL Lot no. 3092003 |
| Validation | Rat monoclonal anti-KRT8 (clone TROMA-I; Developmental Studies Hybridoma Bank, Cat# TROMA-I; RRID: AB_531826) recognizes keratin 8 and is validated by the supplier and independent studies (DeLaForest A, Quryshi AF, Frolkis TS, Franklin OD, Battle MA. GATA4 Is Required for Budding Morphogenesis of Posterior Foregut Endoderm in a Model of Human Stomach Development. Front Med (Lausanne). 2020 Feb 19;7:44. doi: 10.3389/fmed.2020.00044. PMID: 32140468; PMCID: PMC7042400.). The donkey anti-rat IgG (H+L) secondary antibody (Invitrogen/Thermo Fisher, Alexa Fluor™ Plus 488; Cat# A48269; RRID: AB_2893137) is validated by the |

manufacturer for ICC/IF and IHC with minimal cross-reactivity. DAPI (Thermo Fisher, Cat# D1306) is a widely used DNA-binding nuclear counterstain and was used according to the manufacturer's instructions solely to label nuclei.

## Plants

Seed stocks

*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes

*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

Authentication

*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*