nature communications

Article

https://doi.org/10.1038/s41467-026-75427-5

Cellular hallmarks and aging clock of the human lung parenchyma

Received: 23 February 2026

Accepted: 1 July 2026

Published online: 15 July 2026

Check for updates

Ke Xu[1,2,3,4,10], Grace S. Kim[1,2,3,4,10], Atharva Bhagwat[1,2,3,4,10], Sepideh E. Meimand[1,2,3,4], Vignesh V. Venkat[1,2,3,4], Giang Pham[2,3,4], Jing Zhang[2,3,4], Jamshid Abdul-Ghafar[5], Mohamad D. Bairakdar[1,6], Komal Dolasia[1,2,3,4], Isha R. Sahasrabudhe[7], Alter Klausner[8], Yijia Chen[1,9], Thinh T. Nguyen[1,2,3,4], Bruno Giotti[1,2,3,4], Rachel Brody[5], So-Jin Kim[8], Jaymin J. Kathiriya[7,8], Daniel J. Puleston[2,3,4], Patty J. Lee[8] & Alexander M. Tsankov[1,2,3,4]

Aging affects lung function, predisposing older adults to respiratory diseases; however, the cellular and molecular mechanisms of lung aging are not fully understood. Leveraging single-cell and spatial transcriptomics data from 184 and 70 lung parenchyma samples, respectively, we present an analytical platform to dissect the cell composition, gene expression modules, and regulatory changes linked to multiple hallmarks of lung aging. Our findings show cell type-specific age-association of senescence markers and a decline in alveolar cell proliferation, autocrine WNT signaling, and stemness indicators with advancing age. Analysis of myeloid cells reveals a global reduction in macrophage subsets and a surge in mitochondrial dysfunction and inflammatory signaling. In contrast, lung parenchyma T cells expand with age and exhibit heightened interferon gamma expression, cytotoxic activity, and exhaustion in older lung and blood samples, indicative of age-related immune dysfunction. Cell interaction and spatial analysis demonstrate aberrant myeloid-T cell cross-talk, leading to an increase in T cell chemotaxis and activation. Lastly, we use machine learning to predict lung biological age and identify putative biomarkers of lung aging and disease risk.

Aging is a complex and multifaceted process that affects all tissues and organs, leading to a decline in physiological function and an increased susceptibility to both acute and chronic diseases[1]. Underlying decline in function are several hallmarks of aging, including stem cell exhaustion, altered intercellular communication, cellular senescence, mitochondrial dysfunction, loss of proteostasis, genomic instability, and chronic inflammation[2,3]. The lung plays a fundamental role in gas exchange while interacting with the outside environment, and age-related cellular, molecular, and spatial changes have significant impact on lung compliance, respiratory function, host immune interactions, and overall health[4–6]. Hence, understanding the intricate mechanisms underlying lung aging is of paramount importance, as it can shed light on the pathophysiology of age-related respiratory diseases, such as chronic obstructive pulmonary disease, interstitial lung disease (ILD),

[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai (ISMMS), New York, NY, USA. [2]Department of Immunology and Immunotherapy, ISMMS, New York, NY, USA. [3]Lipschultz Precision Immunology Institute, ISMMS, New York, NY, USA. [4]The Tisch Cancer Institute, ISMMS, New York, NY, USA. [5]Department of Pathology, Molecular and Cell-Based Medicine, ISMMS, New York, NY, USA. [6]Presently at University of Washington, Seattle, WA, USA. [7]Departments of Medicine and Cell, Developmental, and Regenerative Biology, ISMMS, New York, NY, USA. [8]Division of Pulmonary, Critical Care and Sleep Medicine, ISMMS, New York, NY, USA. [9]Presently at University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [10]These authors contributed equally: Ke Xu, Grace S. Kim, Atharva Bhagwat. e-mail: alexander.tsankov@mssm.edu

pneumonia, and COVID-19[7]. Better understanding of the cellular and molecular drivers of aging can in turn lead to the development of targeted interventions and therapeutic strategies that can mitigate the decline in respiratory function in older adults[8]. However, the mechanisms of lung aging remain elusive and our understanding as to why age increases the risk of respiratory infections, obstructive and fibrotic lung diseases is lacking[5,9–11]. This gap persists in part because of the organ's intricate cellular composition and the limited resolution of traditional bulk genomic analyses, which obscure cell type–specific changes critical to understanding the aging process.

High throughput single-cell RNA-sequencing (scRNA-seq) and analysis have greatly refined our understanding of the repertoire of >80 cellular identities and functions in lung homeostasis, including the discoveries of the pulmonary ionocyte, tuft-ionocyte progenitor (TIP) cells, deuterosomal cells in airway epithelium, and bipotent alveolar AT0 progenitors capable of differentiating into alveolar or secretory cell lineages[12–19]. Several groups have leveraged the single-cell resolution of these and other data sets to identify a number of cellular changes associated with aging in murine[20–22] and human lungs[23,24], including increased cellular transcriptional noise, granzyme K (*GZMK*) expressing CD8 T cell abundance, and disrupted alveolar macrophage and alveolar type II (AT2) cell regeneration. Despite providing important mechanistic insight, murine studies do not capture many aspects of human lung biology due to vast species differences while scRNA-seq human lung aging studies still remain underpowered to systematically dissect in vivo aging dynamics at a population level. Moreover, prior studies lack high resolution spatial information needed to understand how cellular interactions are altered with age in the context of lung tissue. Efforts by the Human Lung Cell Atlas (HLCA) network built an integrated scRNA-seq atlas of the human lung in health and disease across hundreds of individuals with consensus cellular annotations and harmonized clinical information (e.g., age, sex, ethnicity, smoking history)[25]. We reasoned that similarly processed scRNA-seq lung parenchyma samples within the HLCA cohort would provide improved cellular resolution and statistical power lacking in prior human studies to examine the cellular composition, gene expression modules, regenerative lineages, and ligand-receptor interactions associated with lung aging[23,24,26].

In this work, we present a comprehensive computational framework seeking to characterize the human lung aging process at single-cell resolution by leveraging HLCA scRNA-seq data from 184 lung parenchyma samples as a discovery cohort (Fig. 1a). To validate our findings, our study also provides single-cell resolution spatial transcriptomics data from 70 lung parenchyma samples and 332,063 cells using the Xenium platform and integrates publicly available data from 578 normal lung parenchyma bulk RNA-seq samples from the Genotype-Tissue Expression Program (GTEx) and 317 scRNA-seq and single-cell T cell receptor sequencing (scTCR-seq) samples from peripheral blood to provide additional support and context[27,28]. Drawing from the HLCA and independent validation cohorts, our analysis reveals reproducible, age-associated alterations in cellular composition and gene expression. Consistent with prior observations[5,20,21,26], we observe reductions in alveolar epithelial and macrophage subsets with advancing age, coupled with increase in T, NK, and stromal cell populations—changes suggestive of declining regenerative capacity and chronic immune activation in the aged lung. Moreover, we report an increase in mitochondrial dysfunction in macrophages with age and a *CXCL9*+ myeloid spatial niche that likely contributes to increase in T cell chemotaxis and activation in older lungs. Finally, we leverage machine learning to capture the non-linear relationship between the de novo discovered age-associated cell type expression modules and age, building a single-cell resolution clock of the lung parenchyma.

Results

Single-cell view of age-related changes in lung cell composition

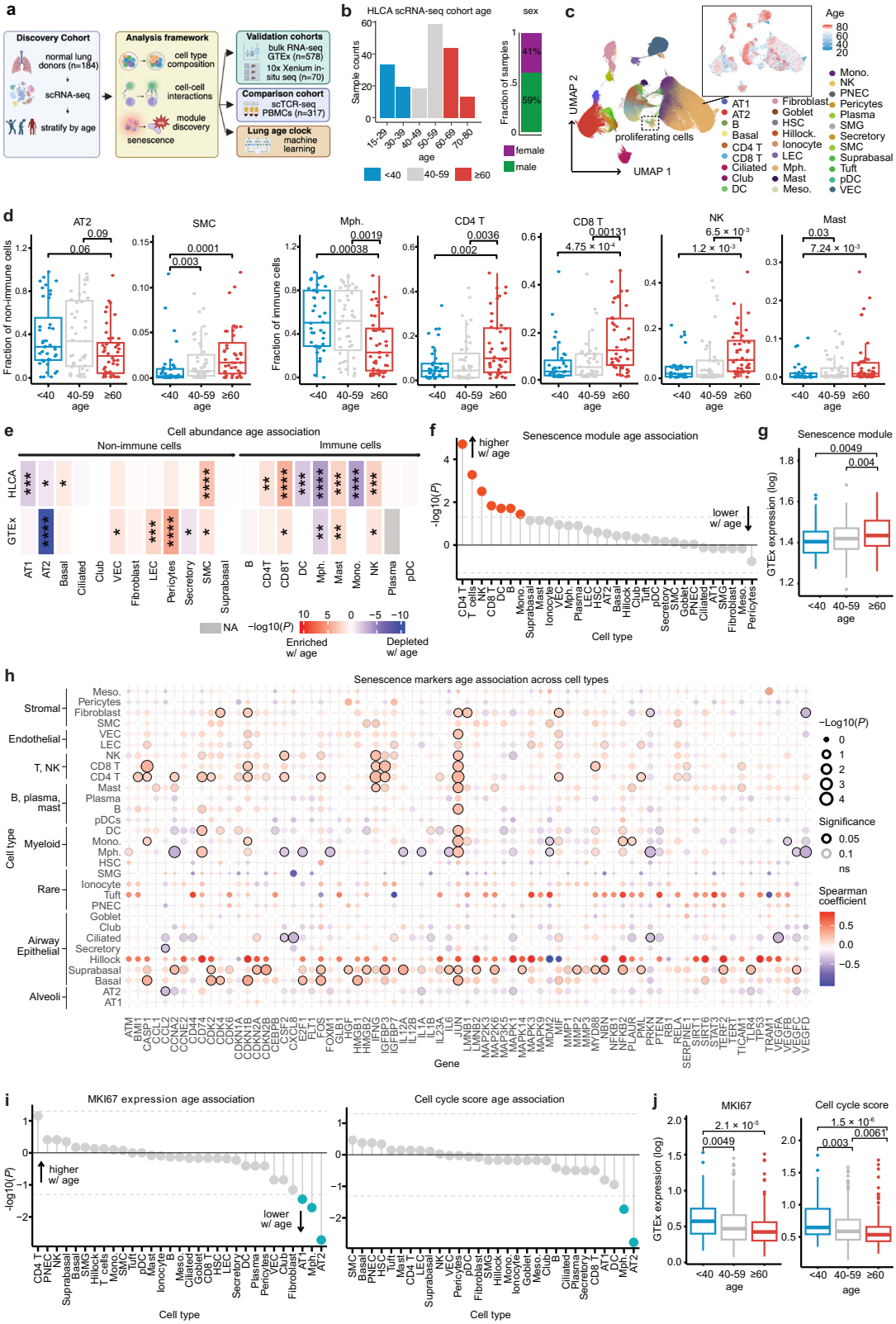
To comprehensively study lung aging, we integrated similarly processed scRNA-seq data from 184 HLCA normal lung parenchyma donor samples with no reported lung disease comorbidities, spanning 15-80 years of age (Supplementary Data 1; Fig. 1b, Supplementary Fig. 1a; Methods). Next, we filled in missing cell annotation information in our HLCA discovery cohort (Fig. 1c; 464,785 total cells) using the HLCA core dataset consensus annotations as a reference[29] (Methods). To investigate age-related changes in lung parenchyma cell compositions, we employed two complementary analytical strategies that both corrected for multiple hypothesis testing using the Benjamini-Hochberg method (Methods)[30]. First, we grouped lung donors into three discrete age groups (<40, 40–59, ≥60 years old; Fig. 1b), in line with metabolomic and proteomic data demonstrating two inflection points around 40 and 60 years of age[31–33], and then computed differences in cell proportions in a pairwise manner between all age groups. Second, we used a rank-based Spearman correlation method that did not rely on age group classifications and rather leveraged all scRNA-seq samples with sufficient cellular sampling (≥1000 cells) in the analysis. Using both approaches and after removing dependencies between cell type fractions to sum to one (Methods), we found pronounced decreases in macrophages (Mph.) and AT2 cells and elevated levels of stromal cells (e.g., smooth muscle (SMC), vascular endothelial cells (VEC), and pericytes) in old lung donors (Fig. 1d, e), where the rank-based analysis consistently showed higher levels of significance (Supplementary Fig. 1b). These results are in agreement with prior knowledge based on murine models and bulk deconvolution analysis of human cohorts[5,20,21,26,34]. While decline in alveolar macrophage fractions has been well established in aged mice[34,35], to our knowledge this has not been reported at a single-cell resolution in human cohorts. Furthermore, we identified highly specific cell type markers (Supplementary Data 2) from the scRNA-seq data and performed bulk deconvolution analysis across the GTEx validation cohort, which showed highly consistent results (Fig. 1e, bottom). To ascertain if the observed age-associated changes in cell proportions were not due to the sex, ethnicity, or smoking history of patients, we regressed out the effects of these covariates using a generalized linear model[36] and observed highly concordant trends (Supplementary Fig. 1c). Finally, we quantified the number of scRNA-seq donor samples required to robustly detect age-associated cell composition changes in the HLCA cohort and found that 100–120 samples (with ≥1000 cells) provided sufficient statistical power across 100 random subsampling runs (Supplementary Fig. 1d).

Examining all cell fractions, we also found elevated levels in CD8 T, natural killer (NK), and mast cells in both our scRNA-seq and bulk validation cohorts (Fig. 1d, e). Prior work noted age-related increases of lymphocytes in human lavage samples[37]; however, to our knowledge changes in immune cell fractions with age have not been comprehensively characterized in human lungs at single-cell resolution. Furthermore, the scRNA-seq data allowed us to explore age-associated changes at higher, cell subset resolution, which uncovered a decline in AT2 and alveolar macrophage proliferating populations and elevated levels of aerocytes, myofibroblasts, and alveolar fibroblasts in older lung donors (Supplementary Fig. 1e). The decline in proliferative capacity of AT2 cells has been well characterized previously[20,24] and is likely a reflection of dysregulated stem cell niches in the lung, a hallmark of aging. In sum, we establish an integrated scRNA-seq lung cohort and analysis framework that can serve as a platform for systematic cellular and molecular dissection of lung parenchyma aging.

Senescence and proliferation markers' age-association is highly cell-type specific

As with changes in cell composition, cellular senescence represents another important hallmark of aging, where Senescence-Associated

https://doi.org/10.1038/s41467-026-75427-5



Secretory Phenotype (SASP) genes have been identified and extensively characterized in prior studies[38,39]. However, the cell type specificity of SASP and other canonical senescence markers and their age-association has not been quantified in the human lung. Towards this goal, we manually curated 75 SASP/senescence marker genes (Supplementary Data 3) and leveraged the HLCA cohort and analysis pipelines to establish the age association for each lung cell type (Fig. 1f). We observed that our senescence signature was highly correlated with age, which was significant for multiple immune cell subsets after multiple hypothesis correction, including CD4 T cells, CD8 T cells, NK cells, dendritic cells (DC), monocytes, and B cells (Fig. 1f). The link between our senescence signature and age was further

Fig. 1 | Single-cell view of age-related changes in lung cell composition and senescence. a Schematic illustration of the analytical strategy and workflow of this study, created in BioRender. Tsankov, A. (2026) https://BioRender.com/upv29gj. **b** Distribution of age (left) and sex (right) for the HLCA normal lung parenchyma scRNA-seq samples used in this study. **c** Uniform Manifold Approximation and Projection (UMAP) visualization of healthy lung parenchyma scRNA-seq dataset (464,785 total cells) colored by cell type annotation and age (top right inset, red color indicating older samples and blue indicating younger samples). Proliferating cells are highlighted by dashed boxes. **d** Box plot of selected cell type fractions in HLCA cohort stratified by age groups (left-to-right: $n = 40, 43, 44$ samples). FDR-corrected two-sided Mann-Whitney-Wilcoxon tests on the CLR-transformed cell-type-abundances were used to calculate changes between age groups. **e** Heatmap of correlation between age and CLR-transformed cell type abundances as measured using HLCA scRNA-seq (top) and deconvoluted GTEx bulk RNA-seq (bottom) data cohorts. scRNA-seq cell type proportions were normalized relative to immune and non-immune cell compartments. Rare cells were excluded from the cell proportion analysis due to small sample size variation. FDR-adjusted, two-sided Spearman correlation -log10(*P* value) are displayed signed by correlation direction, where $^{*}P \le 0.1$, $^{**}P \le 0.05$, $^{***}P \le 0.01$, $^{****}P \le 0.001$, and $P > 0.1$ not shown. **f** Plot showing the correlation between age and the average expression of a curated senescence signature ranked across all HLCA cell types. Y axis reports -log10 (FDR-corrected, two-sided Spearman correlation *P* values) signed by correlation direction. **g** Boxplot of curated senescence signature expression in GTEx dataset. FDR-corrected, two-sided Mann-Whitney-Wilcoxon test *P* values were calculated between age groups (left-to-right: $n = 73, 293, 212$ samples). **h** Dot plot showing the correlation with age for each curated senescence marker (columns) gene expression across all HLCA cell types (rows). Dot color represents the Spearman correlations and dot size represents the two-sided, FDR-corrected *P* values. **i** Plot showing correlation between age and expression of MKI67 (left) or a cell cycle signature (right) across all HLCA cell types. Y axis reports -log10 (FDR-corrected, two-sided Spearman correlation *P* values) signed by correlation direction. **j** Boxplot showing the expression of MKI67 (left) or a cell cycle signature (right) in the GTEx cohort split by age groups (left-to-right: $n = 73, 293, 212$ samples). FDR-corrected, two-sided Mann-Whitney-Wilcoxon test *P* values were calculated between age groups. Meso. = Mesothelium. Mono. = Monocytes. Mph. = Macrophages. HSC = Hematopoietic stem cells. Hillock = Hillock-like cells. LEC = Lymphatic endothelial cells. PNEC = Pulmonary neuroendocrine cells. SMG = Submucosal gland cells. SMC = Smooth muscle cells. pDC = Plasmacytoid dendritic cells. VEC = Vessel endothelial cells. Box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually. *P* values > 0.05 are not shown.

corroborated in the GTEx data (Fig. 1g). To test the generalizability of these findings, we also evaluated the cell type age association using 11 additional senescence signatures from other studies, including Senmayo[38], SenSig[40], SASP[41], CSGene[42], Quest[43], Casella et al.[44], Inflammatory Network[45], CellAge[46], GeneAge[47], DeepScence[48], and the Senescence Network (SenNet) consortium's cross-tissue gene set[49]. Different senescence signature scores increased with age across most cell types and largely in a similar fashion, with the greatest enrichment observed amongst immune cell subsets, in agreement with our manually curated gene set (Supplementary Fig. 2a). We also examined the age association of a previously reported lung epithelial senescence signature[50], and observed a significant age association across several epithelial cell types–club, basal, goblet, ciliated, and ionocytes (Supplementary Fig. 2b).

Next, we examined the age-association of each senescence gene across all lung cell types (Fig. 1h). NK cells, CD8 T cells, and CD4 T cells showed the most consistent upregulation of SASP genes in older lung donors, especially for interferon gamma (*IFNγ*), *IGFBP3*, *CDKN1B*, and *JUN*. In contrast, monocytes and macrophages exhibited both down-regulation (e.g., *VEGFD*, *IL12A*) and upregulation (e.g., Macrophage Migration Inhibitory Factor (*MIF*), *PLAUR*) of SASP genes with age, where *MIF* was also more highly expressed in epithelial and stromal cells of old lung donors. Overall, while senescence markers' expression predominantly increased with age, we observe highly cell type-specific trends for most SASP genes.

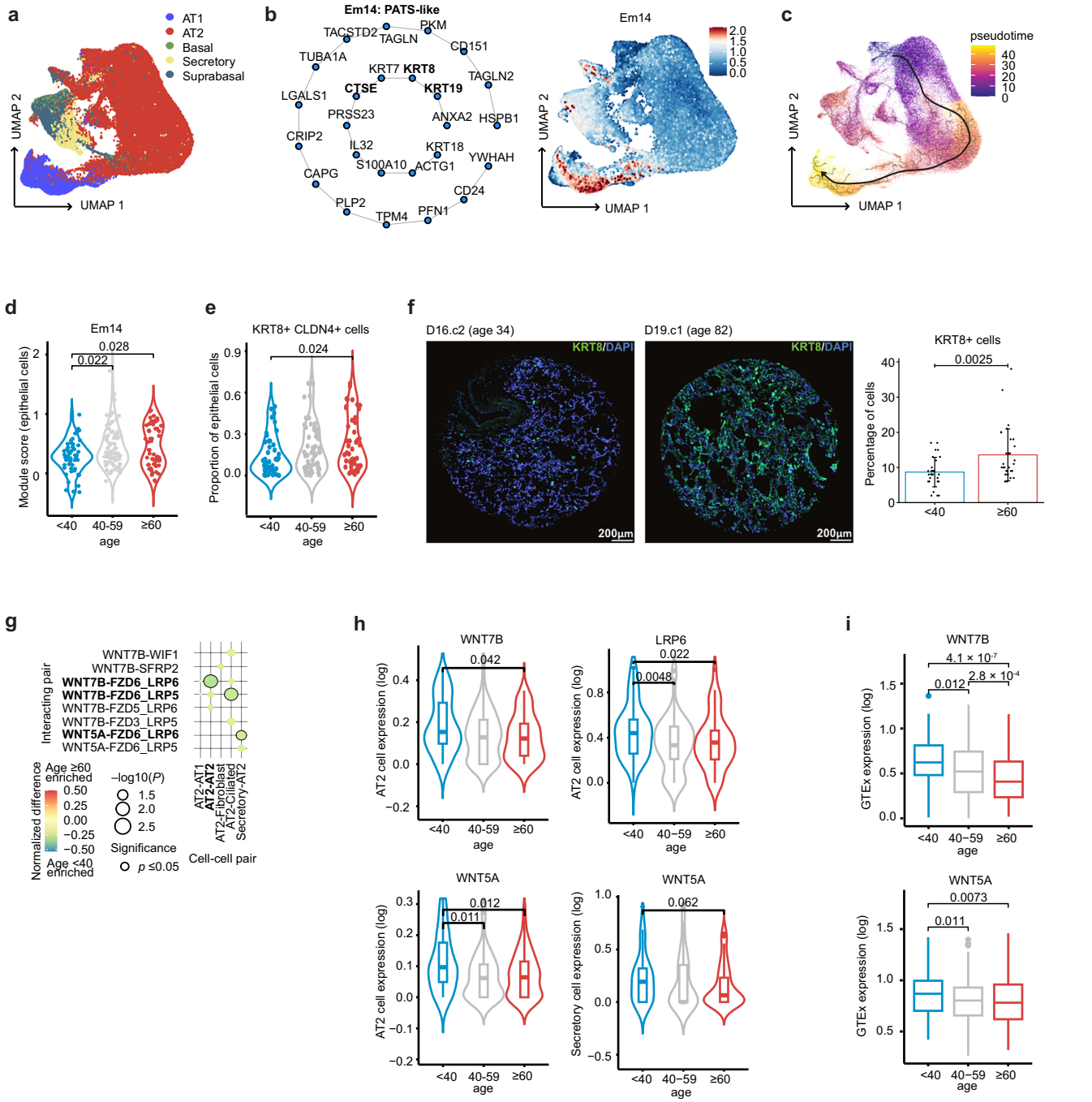
Given the overall increase in cell type-specific expression of senescence markers with age, we next investigated the link between cell proliferation and age across all cell types in our scRNA-seq cohort. Our analysis demonstrated cell type specific age association for the expression of *MKI67*, a cell cycle signature (Supplementary Data 2), and multiple other cell cycle markers (Fig. 1i, Supplementary Fig. 2c), where proliferation decreased amongst most cell types and most prominently in macrophage, AT1, and AT2 cell populations (Fig. 1i), in agreement with our cell subset proportion analysis (Supplementary Fig. 1e). Bulk expression of *MKI67* and our cell cycle signature (Fig. 1j) in the GTEx cohort showed an overall negative correlation with age, in agreement with our higher-resolution findings and prior work[26]. Taken together, we find global increase in senescence and decrease in cell proliferation with age; however, these patterns are cell type and gene specific underscoring the utility of our single-cell resolution platform for de novo discovery of cellular and molecular changes associated with age in the lung parenchyma.

Dysregulation of alveolar regeneration mechanisms in older lung donors

Our ability to detect global aging trends in cell composition, senescence, and cell proliferation prompted us to delve deeper into the molecular pathways associated with aging in each cell compartment, starting with epithelial cells. AT2 cells' stem-like properties and ability to differentiate into AT1 cells are of vital importance in maintaining alveolar homeostasis, supporting gas exchange, and facilitating alveolar regeneration and repair after lung injury[51,52]. Our cell composition analysis above highlighted a reduction of AT2 cell abundance and proliferation in older lung donors (Fig. 1d, e, Supplementary Fig. 1e), suggesting a decline in AT2 regenerative capacity with age. To investigate this further, we used consensus non-negative matrix factorization (cNMF)[53] to de novo discover functional expression modules across lung parenchyma epithelial cells (Fig. 2a). We identified 27 distinct epithelial modules (Ems; Supplementary Data 4) of which 13 demonstrated age-associated expression (Supplementary Fig. 3a). Among these, Em18 exhibited the highest increase in activity with age (Supplementary Fig. 3a, b), expressing basal markers *TP63*, *KRT15* as well as the *STING1* gene, which is a key component of the cGAS-STING pathway and known to sense DNA from intracellular pathogens and regulate type I interferon genes[54] that were also present in this module (e.g., *IFITM1/2/3*, *IFI16*).

Interestingly, Em14 co-expressed *KRT8*, *KRT18*, and *CTSE* genes (Fig. 2b) are reminiscent of pre-alveolar type 1 transitional cell state (PATS), which was shown to be induced during lung fibrosis[55] and COVID-19 autopsy samples[56]. Indeed, Em14 was highly correlated (R = 0.7; Spearman correlation $P < 2.2e\text{-}16$; Supplementary Fig. 3c) with a PATS signature previously defined in ref.[55]. Moreover, pseudotime trajectory analysis using Monocle3[57] confirmed the enrichment of Em14 expression during cell state transition from AT2 to AT1 cells (Fig. 2b, c), further suggesting that Em14 represents an intermediate state during alveolar cell differentiation. Importantly, Em14 expression gradually increases with age in our cohort (Fig. 2d). As a complementary analysis, we quantified the proportion of lung parenchyma epithelial cells expressing both *KRT8* and *CLDN4*, previously identified as the two core markers of PATS[55], and found an elevated fraction in older lung donors (Fig. 2e). To assess KRT8 expression across age groups at the protein level, we performed immunofluorescence staining for KRT8 on tissue microarray (TMA) sections of 59 lung parenchyma tissue cores from 19 autopsy cases of different ages and lacking any lung disease comorbidities at time of death. Confocal

https://doi.org/10.1038/s41467-026-75427-5



imaging of TMA cores showed visibly increased KRT8 staining in samples from older individuals compared to younger individuals (Fig. 2f). Quantitative analysis across all TMA cores (Fig. 2f, right) demonstrated a significantly higher percentage of KRT8$^+$ cells among DAPI$^+$ nuclei in the old group (age ≥ 60 years; $n = 31$ cores) relative to the young group (age <40 years; $n = 28$ cores). While these trends are supportive, we acknowledge that some of the KRT8+ cells may come from airway epithelial cells. In our scRNA-seq cohort, we also observed a reduction in *TP53* expression in AT2 cells with age (Supplementary Fig. 3d), which aligns with prior work directly linking Trp53 to regulation of PATS[55] and showing an accumulation of PATS-like cells in Trp53-null mice[58]. While the TP53 regulatory mechanisms in alveolar cells are highly complex[55,58–60], our aging atlas suggests that reduced *TP53* expression may contribute to the impaired differentiation of AT2 cells in older lung donors, further emphasizing the impact of aging on lung alveolar cell repair and function. We also quantified the age association of four recently defined pathological epithelial progenitors (PEP) signatures[61], including basaloid[15], DATP[62], KRT8 + ADI[63] and PATS-like[55] signatures. We observed that both PATS-like and basaloid signatures were significantly associated with age in our scRNA-seq cohort (Supplementary Fig. 3e), consistent with observed accumulation of these cells in response to chronic and acute lung injury[15,55]. The PATS-like module Em14 was also correlated (R = 0.47; Spearman correlation $P < 2.2e\text{-}16$; Supplementary Fig. 3f) with the basaloid signature, highlighting the transcriptional coupling between these 2 expression programs. In sum, we identified several epithelial cell functional modules with elevated expression in older lung donors, including a KRT8-expressing PATS-like signature that we validated at the protein level.

In lung homeostasis, WNT signaling likely plays an important role in maintaining the AT2 cell stem niche, capacity to self-renew, and ability to differentiate into AT1 cells, which are essential for efficient

Fig. 2 | Dysregulation of alveolar regeneration mechanisms in older lung donors. a UMAP visualization of epithelial cells ($n = 91,479$) colored by cell type annotation. Ciliated, goblet, and rare epithelial cells were excluded from the visualization. **b** Wheel plot of the top 25 genes representing PATS-like module Em14 (left) and UMAP visualization of Em14 module score (right). **c** UMAP visualization of epithelial cell pseudotime analysis. Arrow indicates inferred trajectory direction. **d** Violin plot of Em14 module expression in epithelial cells across HLCA samples stratified by age. FDR-corrected, two-sided Mann-Whitney-Wilcoxon test *P* values were calculated between age groups (left-to-right: $n = 54, 77, 56$ samples). MDM=Monocyte-derived Macrophages. **e** Violin plot of KRT8, CLDN4 double positive cell fractions relative to all epithelial cells across HLCA samples stratified by age. Two-sided Mann-Whitey-Wilcoxon test *P* values were calculated between age groups (left-to-right: $n = 45, 63, 50$ samples). **f** Left: Representative confocal images from two TMA cores illustrating KRT8 (green) and DAPI (blue) staining. Right: Quantification of the proportion of KRT8+ cells among DAPI nuclei across 28 cores from a young group (age <40 years) and 31 cores from the old group (age ≥ 60 years). Bars represent the mean ± standard deviation. Significance between the old and young groups was calculated using a two-sided Mann-Whitney-Wilcoxon test. **g** Dot plot of significantly age-associated ligand-receptor (rows) interactions between AT2 cells and other cell types. Dot color represents direction of age-association and dot size represents the -log10(Fisher's exact test *P* values). **h** Violin plots of log normalized WNT7B (top left), WNT5A (bottom left), and LRP6 (top right) gene expression in AT2 cells from HLCA samples stratified by age (left-to-right: $n = 52, 70, 48$ samples) and WNT5A (bottom right) gene expression in secretory cells between HLCA age groups (left-to-right: $n = 45, 60, 52$ samples). FDR-corrected, two-sided, Mann-Whitey-Wilcoxon test *P* values were calculated between age groups. For visualization purposes, the minimum and maximum range of the y axis was set to *Q1-1.5IQR* and *Q3 + 1.5IQR*, respectively. **i** Box plots of log-normalized WNT7B (top) and WNT5A (bottom) gene expression in the GTEx bulk RNA-seq cohort split by age groups (left-to-right: $n = 73, 293, 212$ samples). FDR-corrected, two-sided Mann-Whitney-Wilcoxon test *P* values were calculated between age groups. For visualization purposes, the minimum and maximum range of the y axis was set to *Q1–1.5IQR* and *Q3 + 1.5IQR*, respectively. Box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually. *P* values > 0.05 are not shown.

gas exchange[52,64,65]. Given the observed reduction in AT2 cells in aging populations, we sought to examine the relationship between WNT signaling and age. Using ligand-receptor interaction inference[66] on the scRNA-seq cohort, we demonstrated a substantial decrease in *WNT7B* and *WNT5A* interactions involving AT2 cells with age (Fig. 2g), not observed for other regenerative pathways (e.g., FGF, NOTCH). The decrease in interactions is likely due to downregulated *WNT7B*, *WNT5A*, and *LRP6* expression in AT2 cells and decreased *WNT5A* production in secretory cells of older lung donors (Fig. 2h). We observed a similar decline in *WNT7B* and *WNT5A* bulk expression in the GTEx cohort (Fig. 2i). While *WNT7B* predominantly activates canonical WNT/β-catenin signaling in the context of lung development[67], mesenchymal proliferation[68], and lung injury repair[69,70], *WNT5A* generally acts through non-canonical pathways. However, its reduction with aging reflects a broader deterioration of the AT2 autocrine WNT niche. This decline in WNT ligand availability, coupled with reduced *LRP6* expression, may contribute to impaired canonical WNT tone required for AT2 homeostasis. The decreased AT2 cell autocrine WNT interactions and proliferation with age are reminiscent of prior work showing induction of autocrine WNT signaling and progenitor expansion during murine lung injury[52].

In summary, we report decreased AT2 cell abundance, proliferative capacity, and autocrine WNT signaling interactions, and increased PATS-like cells in older lung donors, emblematic of a decline in alveolar proliferation, regeneration, and overall lung function with age.

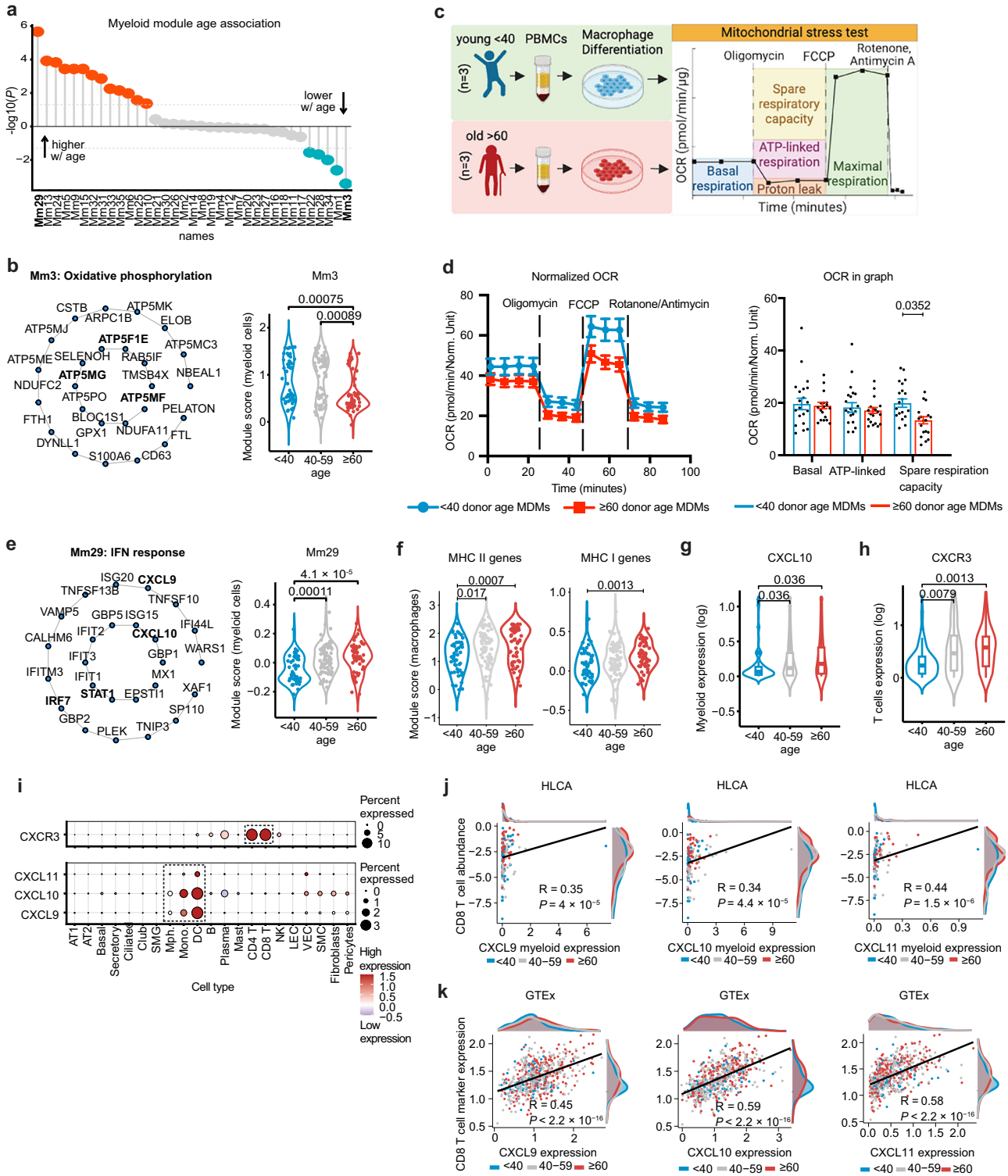
Myeloid expression of *CXCL9/10/11* and T cell abundance correlate with age

Myeloid cells are essential for maintaining an effective innate immune response against pathogens that frequently enter the lung airways and their function can deteriorate with age[71]. It has been previously reported that alveolar macrophages decline in number in aging mice[20,22,35,72], and our cell composition analysis extends this observation to several macrophage cell subsets in the human lung parenchyma, including alveolar, monocyte-derived, and proliferating macrophages (Supplementary Fig. 1e). To investigate age-associated mechanisms that may underlie the decline of macrophage abundance and function, we again applied cNMF to all myeloid cells in our scRNA-seq cohort and identified 18 age-related, gene co-expression, myeloid modules (Mms; Fig. 3a, Supplementary Data 4). Amongst these, Mm3 showed the greatest decline in activity with age (Fig. 3a, b) that was also evident in decreased expression of nearly all constituent genes within the module and across different myeloid cell subsets, including DCs, macrophages, and monocytes (Supplementary Fig. 4a, b). Mm3 was enriched for genes associated with oxidative phosphorylation, respiration, and ATP production, including several ATP synthase genes (e.g., *ATP5F1E*, *ATP5MF*; Fig. 3b). ATP synthase protein complexes have been shown to separate from dimers to monomers with age, leading to a loss of the characteristic mitochondrial membrane folds and reduced cellular energy production capacity[73]. To test this functionally, we evaluated oxygen consumption rate (OCR) using a mitochondrial stress test of monocyte-derived macrophages (MDMs) differentiated from human PBMC donors of different ages (Fig. 3c, Supplementary Data 5; Methods). Consistent with our hypothesis, the stress test data showed impaired OCR in old (age ≥60) versus young (age <40) donor MDMs, as indicated by the reduced spare respiratory capacity after induction of FCCP (Fig. 3d), a compound that disrupts the mitochondrial membrane proton gradient and drives respiration to work at maximal capacity. In sum, our computational and experimental data argue that reduced expression of ATP synthase and oxidative phosphorylation genes with age in myeloid cells likely lead to mitochondrial dysfunction, one of the original hallmarks of aging[39].

On the other end of the spectrum, Mm29 was the most upregulated module in myeloid cells of older lung donors (Fig. 3a) and was marked by expression of genes (e.g., *ISG20*, *IFIT1/2/3*) associated with response to interferon (IFN) signaling (Fig. 3e). This module also co-expresses transcription factors (TFs) *IRF7* and *STAT1* (Fig. 3e), known to regulate IFN signaling[74,75], and displayed consistent upregulation with age across module genes and myeloid cell subsets (Supplementary Fig. 4a, b). The established connection between IFN signaling and expression of the major histocompatibility complex (MHC)[76,77] prompted us to investigate the age-association of MHC class I and II gene in myeloid cells. Indeed, we found higher expression of MHC class I in macrophages and MHC class II in both macrophages and DCs in older lung donors (Fig. 3f, Supplementary Fig. 4c). This suggests an age-associated increase in antigen-presentation activity by DCs and macrophages in the lung parenchyma, which we will explore further in the next section.

We also noticed co-expression of C-X-C motif chemokine ligands 9 and 10 (*CXCL9*, *CXCL10*) in Mm29, which are known to be regulated downstream of *STAT1* and IFN signaling[78] and to attract activated T cells to sites of inflammation via CXCR3 receptor[79,80]. We observed significant upregulation of *CXCL10* in myeloid cells of older lung donors (Fig. 3g) and a more modest increase in expression of *CXCL9* and *CXCL11* (Supplementary Fig. 4d). These chemokines are all specifically expressed by myeloid cells in the lung parenchyma, while their cognate receptor *CXCR3* is primarily expressed in lung T cells and also upregulated with age (Fig. 3h, i). Given that we previously noted higher T cell abundance in older lung donors (Fig. 1d, e), we next investigated

https://doi.org/10.1038/s41467-026-75427-5



the potential age-associated role of *CXCL9/10/11* in T cell chemotaxis. We found a significant correlation between CD8 or total T cell abundance and *CXCL9*, *CXCL10*, and *CXCL11* expression in myeloid cells across our scRNA-seq cohort (Fig. 3j, Supplementary Fig. 4e; Spearman correlation P value ≤ 0.0001). This trend was corroborated by the GTEx cohort when correlating *CXCL9*, *CXCL10*, and *CXCL11* expression and deconvolved CD8 and total T cell abundance (Fig. 3k, Supplementary Fig. 4f; Spearman correlation P value < $2.2*10^{-16}$). Taken together, we de novo discovered 18 age-associated myeloid expression modules, including an oxidative phosphorylation and IFN response module; the IFN module co-expresses TFs *STAT1*, *IRF7* and chemokines *CXCL9*, *CXCL10*, thus likely contributing to induction of MHCII genes in macrophages/DCs and to a higher T cell abundance observed in older lung donors.

Elevated *IFNγ* expression and T cell activation in older lung donors

Aging is associated with decreased potential of the adaptive immune system to generate memory T cells, as the pool of naive T cells and the diversity of TCR clonotypes decreases with age[81]. This likely

Fig. 3 | Myeloid expression of CXCL9/10/11 and T cell abundance correlate with age. a Plot showing the correlation between age and myeloid module scores (averaged across myeloid cells) of HLCA samples, ranked by FDR-corrected, Spearman correlation *P* value. Y axis reports -log10 (*P* value) signed by correlation direction. **b** Wheel plot of the top 25 genes in Mm3 (left) and violin plot of Mm3 module average expression in myeloid cells (right) for all HLCA samples stratified by age. FDR-corrected, two-sided Mann-Whitney-Wilcoxon test *P* values were calculated between age groups (left-to-right: $n = 50, 75, 56$ samples). **c** Schematic illustration of the mitochondrial stress test experimental workflow, including donor PBMC acquisition, macrophage differentiation, and OCR stress analysis, created in BioRender. Tsankov, A. (2026) https://BioRender.com/upv29gj. **d** Metabolic stress plot and bar plot of mitochondria stress test measured by Seahorse XF analyzer for human PBMC monocyte-derived macrophages (MDMs) stratified by age. Two-way ANOVA with Tukey's correction was used to identify significant differences between age groups ($n = 19$ and $n = 18$ replicates for young and old groups, respectively). Data are presented as mean values +/- standard error of the mean. **e** Wheel plot of the top 25 genes representing Mm29 (left) and violin plot of Mm29 module average expression in myeloid cells (right) for all HLCA samples stratified by age. FDR-corrected, two-sided Mann-Whitney-Wilcoxon test *P* values were calculated between age groups (left-to-right: $n = 50, 75, 56$ samples). **f** Violin plot of MHC class I and class II module scores averaged across macrophages for all HLCA samples stratified by age. FDR-corrected, two-sided Mann-Whitney-Wilcoxon test *P* values were calculated between age groups (left-to-right: $n = 50, 74, 56$ samples). **g** Violin plot of log-normalized CXCL10 expression in myeloid cells for all HLCA samples stratified by age. FDR-corrected, two-sided Mann-Whitey-Wilcoxon test *P* values were calculated between age groups (left-to-right: $n = 50, 75, 56$ samples). For visualization purposes, the minimum and maximum range of the y axis was set to *Q1-1.5IQR* and *Q3 + 1.5IQR*, respectively. **h** Violin plot of log-normalized CXCR3 expression in T cells for all HLCA samples stratified by age. FDR-corrected, two-sided Mann-Whitey-Wilcoxon test *P* values were calculated between age groups (left-to-right: $n = 48, 66, 55$ samples). For visualization purposes, the minimum and maximum range of the y axis was set to *Q1-1.5IQR* and *Q3 + 1.5IQR*, respectively. **i** Dot plot of scaled log-normalized expression for CXCR3, CXCL11, CXCL10, CXCL9, across all HLCA cell types with at least 400 cells. **j** Spearman correlation and two-sided *P* values were calculated between log normalized CXCL9, CXCL10, and CXCL11 expression in myeloid cells and CD8 T cell abundance relative to immune cells across HLCA scRNA-seq samples, where samples with no chemokine expression or zero CD8T cell abundance were excluded. **k** Spearman correlation and two-sided *P* values were calculated between log normalized CXCL9, CXCL10, and CXCL11 expression and CD8 T cell marker expression (proxy for T cell abundance) across GTEx bulk RNA-seq samples. Box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually. *P* values > 0.05 are not shown.

contributes to diminished protection against pathogens[82] and higher incidence of respiratory diseases in older individuals, including influenza[83] and COVID-19[7]. To gain further insight into the functional changes in lung T cells during aging, we identified 12 age-associated T and NK cell modules (Tms; Fig. 4a, Supplementary Data 4) using cNMF[53]. Among these, two environmental stress response related modules were upregulated with age, including Tm4 associated with tumor necrosis factor alpha (TNFα) pathway genes (e.g., *NFKBIA*, *FOS*; Supplementary Fig. 5a) and Tm12 expressing heat shock protein (HSP; Fig. 4b) genes, which play a critical role in both heat shock and general stress response[84]. Both Tm4 and Tm12 module scores and their constituent genes were consistently upregulated across CD4 T, CD8 T, and NK cells in older lung donors (Supplementary Fig. 5b, c). We also found increased activity of IFNγ signaling module Tm7 expressing *IFNγ*, chemokines *CCL3/4*, and TF *IRF8* in T cells of older lung donors (Fig. 4c). Given the importance of IFNγ signaling in enhancing anti-viral response[85], we next asked if its expression in lung T/NK cells increased with age; indeed we observed significantly higher expression of *IFNγ* in older lung donors across all T cells (Fig. 4d), as well as across NK cells, CD4 and CD8 T cells (Supplementary Fig. 5d). T and NK cells also exhibit the highest expression of *IFNγ* amongst all lung cell types (Supplementary Fig. 5e), arguing that increased *IFNγ* expression in T and NK cells of older lung donors likely contributes to the elevated activity of IFN response gene module Mm29 in myeloid cells (Fig. 3e), and suggesting a feedforward loop of increased Mm29 *CXCL9/10/11* myeloid expression, T cell recruitment, and *IFNγ* expression in the lung parenchyma of older individuals. Finally, we report upregulation of module Tm2 expressing genes related to cytotoxicity (e.g., *PRF1*, *GZMA*) in CD8 T and NK cells of older lung donors (Supplementary Fig. 5c,f), which is likely related to higher IFNγ signaling in these samples.

We also observed an age-associated upregulation of module Tm8 expressing T cell activation (e.g., *CD3D/E*, *CD8A/B*) and effector (e.g., *GZMA*, *GZMK*) markers (Fig. 4e), where the majority of Tm8 genes were consistently upregulated with age across T cell subsets (Supplementary Fig. 5g). To investigate the mechanisms that may underlie increased expression of T cell activation genes, we performed differential ligand-receptor interaction analysis involving T cells and all other cell types. This analysis uncovered a number of increased ligand-receptor interactions known to play a role in T cell activation (Fig. 4f). For example, older lung donors demonstrated increased interactions of *CD40LG* and *CD40* between CD4 T and myeloid cells; this interaction is known to act as a secondary, costimulatory signal for helper T cell activation by antigen presenting cells (APCs)[86]. We also found a significant enrichment of interactions involving CD4 and CD8 T cells expressing integrins, including integrin *ITGAL* that is highly expressed on T and NK cells, with myeloid cells expressing *ICAM1* in old lung donors (Fig. 4f, Supplementary Fig. 5h). *ICAM1*–integrin/*ITGAL* cell–cell adhesion functions to stabilize the interaction between APCs and T cells, promoting T cell activation[87,88]. Moreover, we also show age-associated increase in interactions between *ANXA1* expressing T cells and formyl peptide receptor (FPR; e.g., *FPR1/3*) expressing myeloid cells (Fig. 4f). ANXA1 is a known molecular tuner of helper T cell activation and differentiation functioning via the FPR family[89], where *FPR1* and *FPR3* are most highly expressed in lung myeloid cells capable of antigen presentation (Supplementary Fig. 5h). In support of these findings, older lung donors demonstrated elevated expression of *CD40LG*, *ITGAL* and *ANXA1* in the relevant T cell subsets and *ICAM1* and *CD40* in DCs and macrophages (Fig. 4g, Supplementary Fig. 5i). Taken together, older lung donors exhibit elevated expression of *IFNγ*, TNFα pathway, stress response, and T cell activation gene modules, where the latter is likely, in part, due to increased costimulatory interactions involving *ICAM*, *CD40LG*, and *ANXA1*.

Spatially-resolved mechanisms of aging in the lung parenchyma

To further explore the age-associated cellular crosstalk in lung parenchyma spatial tissue context, we performed highly-multiplexed in-situ spatial transcriptomics on 70 lung parenchyma samples from 22 deceased individuals spanning different ages, sexes, and ethnicities but lacking any lung disease comorbidities (Fig. 5a, Supplementary Data 6). The 70 sample cores (1.5 mm in diameter) were selected by a board-certified pulmonary pathologist on our team to include bronchial, alveolar, and vascular regions of interest from each case, and placed on two tissue microarrays (TMAs; Fig. 5b). High-resolution spatial transcriptomics profiling of the TMAs was performed across 389 gene probes using the Xenium platform, including 100 custom selected markers of senescence and aging (Supplementary Data 6). The spatial data provided true single-cell resolution as cells clustered by cell type but not donor identity (Fig. 5c, d, Supplementary Fig. 6a), and higher sample size and quality compared to two recently published Xenium datasets from human lung tissue[90,91] (Supplementary Fig. 6b). Importantly, when correlating cell composition with sample age, the Xenium cohort showed highly consistent trends as the two other independent cohorts from HLCA and GTEx already examined

https://doi.org/10.1038/s41467-026-75427-5

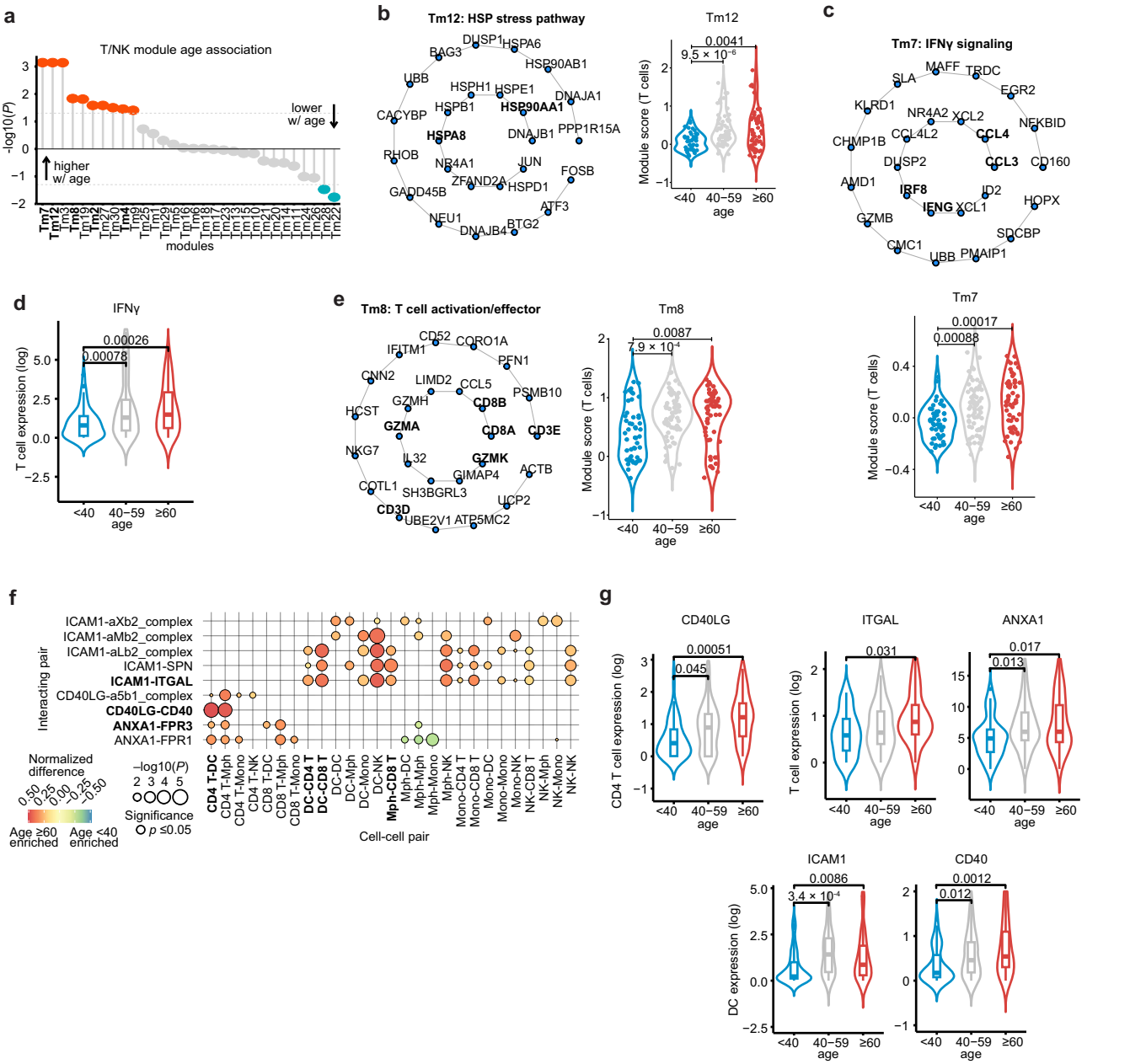
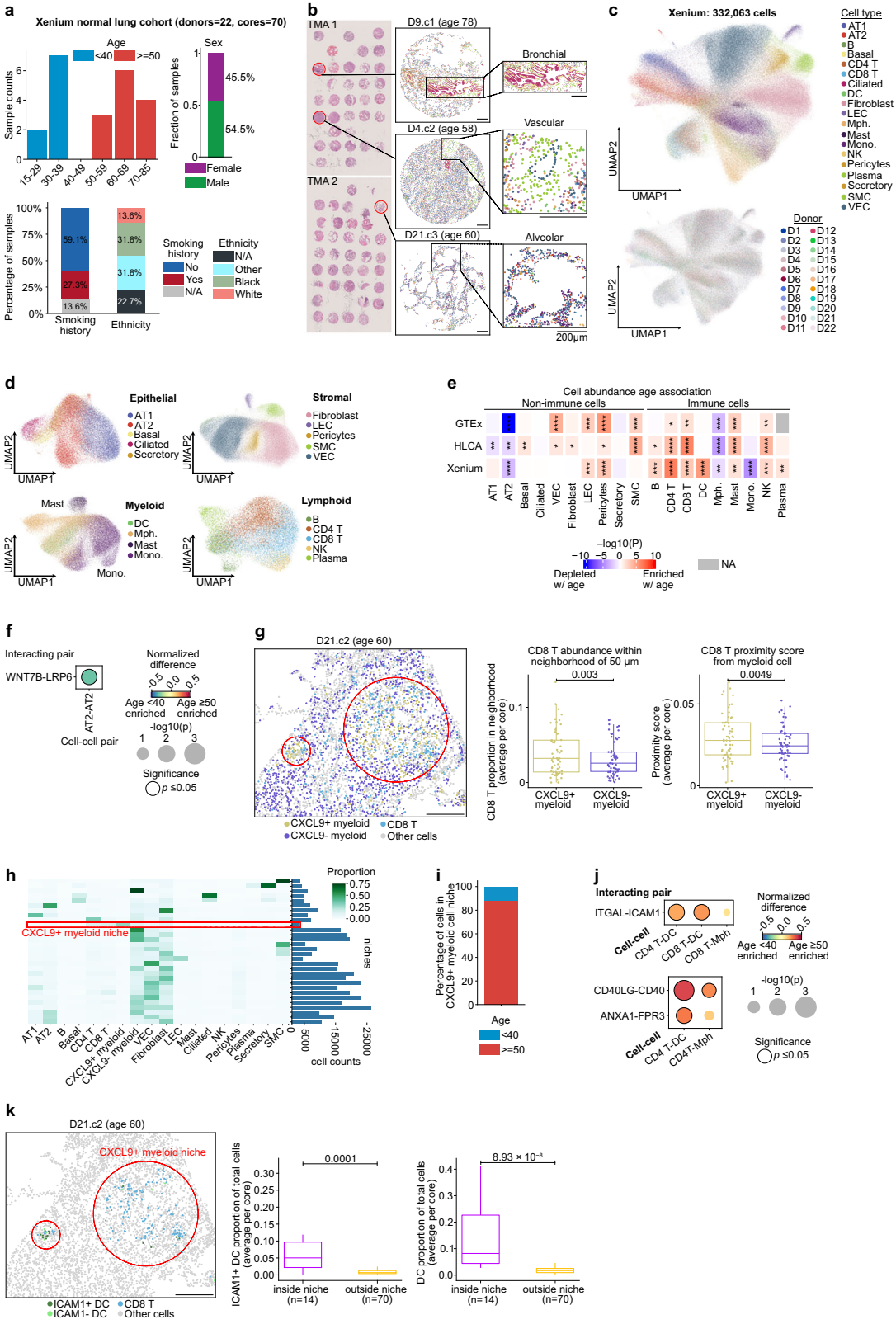


Fig. 4 | Elevated *IFNγ* expression and T cell activation in old lung donors. a Plot showing the correlation between age and T/NK module scores (averaged across T and NK cells) of HLCA samples, ranked by FDR-corrected, Spearman correlation *P* value. Y axis reports -log10 (*P* value) signed by correlation direction. **b** Wheel plot of the top 25 genes included in Tm12 (left) and violin plot of Tm 12 module average expression in T cells (right) for all HLCA samples stratified by age. FDR-corrected, two-sided Mann-Whitney-Wilcoxon test *P* values were calculated between age groups (left-to-right: $n = 48, 66, 54$ samples). **c** Wheel plot of the top 25 genes included in Tm7 (top) and violin plot of Tm7 module average expression in T cells (bottom) for all HLCA samples stratified by age. FDR-corrected, two-sided Mann-Whitney-Wilcoxon test *P* values were calculated between age groups (left-to-right: $n = 48, 66, 54$ samples). **d** Violin plot of log-normalized IFNγ expression in all T cells for all HLCA samples stratified by age. FDR-corrected, two-sided Mann-Whitey-Wilcoxon test *P* values were calculated between age groups (left-to-right: $n = 48, 66, 54$ samples). For visualization purposes, the minimum and maximum range of the y axis was set to *Q1-1.5IQR* and *Q3 + 1.5IQR*, respectively. **e** Wheel plot of the top 25 genes included in Tm8 (left) and violin plot of Tm8 module average expression in T cells (right) for all HLCA samples stratified by age. FDR-corrected, two-sided Mann-Whitney-Wilcoxon test *P* values were calculated between age groups (left-to-right $n = 48, 66, 54$ samples). **f** Dot plot of significantly age-associated ligand-receptor (rows) interactions involved in T cell activation between CD4 T, CD8 T, NK, DC, Monocytes and Macrophages (columns). Dot color represents direction of age-association and dot size represents the -log10 (Fisher's exact test *P* values). DC = dendritic cells. Mono.=Monocytes. Mph.=Macrophages. **g** Violin plot of log-normalized gene expression of selected ligand and receptor genes from the cell-cell interaction analysis in (**f**) for all HLCA samples stratified by age. FDR-corrected, two-sided Mann-Whitey-Wilcoxon test *P* values were calculated between age groups (left-to-right: $n = 45, 65, 53$ samples for CD40LG expression; $n = 48, 66, 54$ samples for ANXA1 and ITGAL expression; $n = 48, 72, 55$ samples for ICAM1 and CD40 expression). For visualization purposes, the minimum and maximum range of the y axis was set to *Q1-1.5IQR* and *Q3 + 1.5IQR*, respectively. Box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually. *P* values > 0.05 are not shown.

https://doi.org/10.1038/s41467-026-75427-5



(Fig. 5e, Supplementary Fig. 6c), including decrease in AT2 and macrophage abundance and elevated numbers of immune (T, mast, and NK) and stromal (pericytes) cell types in older lung samples. These trends were largely preserved after removing the constraint for cell type fractions to sum to one (Supplementary Fig. 6d).

Beyond cell fractions, the Xenium data also provides spatial information that can be used to investigate how cellular crosstalk and ligand-receptor interactions alter with age. Consistent with our observations in the HLCA cohort, the spatial data demonstrated a depletion of interactions between *WNT7B* and *LRP6* (*FZD6* and *LRP5* were not present on our Xenium gene panel) at neighboring AT2 cells (Fig. 5f). The spatial data also supported our prediction that *CXCL9/10* expressing myeloid cells play a role in T cell chemotaxis, especially in older lung samples, where we observed increased abundance and

Fig. 5 | Investigating the age-associated cellular and spatial crosstalk using Xenium spatial data from normal lung autopsy samples. a Distribution of age, sex, smoking history, and ethnicity across our normal lung cohort spatially profiled using Xenium. **b** Left: H&E staining of two TMAs consisting of 70 cores from 22 autopsy cases (4 cores per case sampling distinct anatomical regions–bronchial ($n = 7$), vascular ($n = 22$), alveolar ($n = 40$), and undetermined ($n = 1$); core diameter 1.5 mm), profiled in a single Xenium run to eliminate batch effects. Middle: Spatial maps of representative sample cores, colored by cell annotations. Right: Zoomed in bronchial, vascular, and alveolar regions as annotated by a pulmonary pathologist. All scale bars represent a distance of 200 μm. **c** Uniform Manifold Approximation and Projection (UMAP) visualization of xenium cohort (332,063 total cells) colored by cell type annotation (top) and colored by donor IDs (bottom) after batch correction using Harmony on donor IDs. **d** UMAP plot of clustered Xenium spatial transcriptomics data split into cell type compartments—Epithelial (top-left), Stromal, Myeloid, and Lymphoid (bottom-right)—colored by cell type annotation after batch correction using Harmony. **e** Heatmap of Spearman correlation between age and cell proportions measured using deconvolved GTEx RNA-seq (top), HLCA scRNA-seq (middle), and Xenium (bottom) data cohorts, where cell proportions were normalized relative to immune and non-immune cell types. Cell types with ≤10 average number of cells per sample in HLCA scRNA-seq are not shown. FDR-adjusted, two-sided Spearman correlation -log10(*P* values) are displayed and signed by correlation direction, where $^{*}P \le 0.1$, $^{**}P \le 0.05$, $^{***}P \le 0.01$, $^{****}P \le 0.001$, $P > 0.1$ not shown. **f** Dot plot of WNT7B-LRP6 ligand-receptor spatial interaction between AT2 cell types. Dot color represents direction of age-association and dot size represents the -log10 (Fisher's exact test *P* value). **g** Left: Spatial map of D21.c2 core, where cells (dots) are colored by annotations of interest. Red circles indicate CXCL9+ myeloid niche. Scale bar represents a distance of 200 μm. Right: Comparison of CD8 T cell abundance (middle) and proximity (right) between CXCL9+ myeloid and CXCL9- myeloid centered neighborhoods of radius 50 μm. Only the cores with a minimum of 5 CXCL9+ myeloid and CXCL9- myeloid cells were considered ($n = 61$). Two-sided Wilcoxon signed-rank paired test *P* value was computed. **h** Heatmap proportion of cells belonging to 30 spatial niches identified using SOAPy. **i** Stacked bar plot representing percentage of cells in CXCL9+ myeloid cell niche colored by age group. **j** Dot plot of T cell activation interactions (rows) between T and antigen presenting cells (columns). Dot color represents direction of age-association and dot size represents the -log10 (Fisher's exact test *P* value). **k** Left: Spatial map of D21.c2 core, where cells (dots) are colored by annotations of interest. Red circles indicate CXCL9+ myeloid niche. Scale bar represents a distance of 200 μm. Right: Box plots comparing the proportion of ICAM1+ DCs (left) and total DCs (right) per core either within or outside the CXCL9+ myeloid niche. Significance was assessed using a two-sided Mann-Whitney-Wilcoxon test. Box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually. *P* values > 0.05 are not shown.

proximity of CD8 and total T cells near *CXCL9*+ and *CXCL10*+ versus *CXCL9*- and *CXCL10*- myeloid cells, respectively (Fig. 5g, Supplementary Fig. 6e,f). Interestingly, in a handful of samples (e.g., D21.c2 in Fig. 5g) we observed large immune aggregates composed of *CXCL9/10* expressing myeloid and CD8 T cells. To further characterize these spatial structures, we performed de novo cellular neighborhood detection analysis and identified 30 cellular niches, including a niche enriched for *CXCL9*+ myeloid cells (Fig. 5h, highlighted in red). Cells within this niche originated predominantly from older lung samples, including sample D21.c2 (Fig. 5i). Given the higher abundance of T cells in these niches, we next utilized the spatial data to investigate interactions related to T cell activation we previously observed to be enriched in HLCA older lung donors (Fig. 4f). Indeed, we observed higher spatial co-localization of ligand-receptor interacting gene pairs *ITGAL-ICAM1*, *ANXA1-FPR3*, and *CD40LG-CD40* expressed on T cells and APCs, respectively, in older lung spatial samples (Fig. 5j, Supplementary Fig. 6g-i). Finally, we also observed higher abundance of both *ICAM1*+ and total DCs in the *CXCL9*+ myeloid niches compared to cells found outside of these niches (Fig. 5k). Taken together, we find that *CXCL9*+ myeloid niches predominantly found in older lung samples have a higher abundance of both T cells and DCs, which may contribute to an enrichment of T cells and T cell activation interactions in older lung donors that we initially observed in the HLCA scRNA-seq discovery cohort.

Age-associated T cell modules are coupled to TCR clonotype size

Our findings of heightened T cell abundance, activation, and *IFNγ* expression in the aging lung prompted us to examine whether these age-related changes are consistent across tissues, where cross-tissue analysis[92–94] has the potential to provide a broader understanding of T cell function during aging. Towards this goal, we used our computational framework to reanalyze scRNA-seq and scTCR-seq data from a recent study characterizing 317 peripheral blood mononuclear cell (PBMC) samples from healthy individuals[28]. Comparing immune cell proportion association with age between lung and blood (Fig. 6a, Supplementary Fig. 7a), we found a similar increase in CD4 T and NK cell proportion in older donors. We also observed notable differences, where monocyte, DC, and pDC abundance increased with age in PBMCs, which contrasts with our observations in the lung (Fig. 6a). Notably, we found a divergence in CD8 T cell abundance trends, where their proportion increases in the lung but decreases in the blood with age. Prior studies have indicated differences in T cell compositions between tissues and blood[93,94], prompting us to explore how the CD8 T cell subset fraction changes with age. To do this, we classified lung CD8 T cells subsets using label transfer[29] of PBMC CD8 T cell annotations as a reference. Interestingly, CD8 T cell subsets in both the lung and blood showed largely similar aging trends (except the rare *KLRC2* + , *GZMB*- subset that made up ~0.5% and ~3.5% of CD8 T cells in blood and lung, respectively), where naive CD8 T cell abundance decreased, while *GZMB*+ effector memory T cell subsets increased with age (Fig. 6b, Supplementary Fig. 7b). Examining the composition of T cell subsets in the lung and blood (Fig. 6c), we observed that the naive subset represents the largest fraction of CD8 T cells in the blood (34.9%) but <0.4% of CD8 T cells in the lung. In contrast, *GZMB*+ and *GZMK* + T effector memory (Tem) cells make up 58% of lung but only 32% of blood CD8 T cells, which may explain why total CD8 T cell proportions increase with age in lung but not in PBMCs dominated by naive CD8 T cells. Homing mechanisms via receptor CXCR3 and its ligands CXCL9/10/11, which we showed to be upregulated in older lung donors (Fig. 3g,h, Supplementary Fig. 4d), may also contribute to the divergent age trends in lung versus blood for CD8 T cell abundance, where prior studies have demonstrated that these chemokines efficiently recruit activated effector CD4 and CD8 T cells into the lung[95,96]. Accordingly, we also observed increased *CXCR3* expression in older blood donors (Supplementary Fig. 7c), indicating that aging is associated with an expanded circulating pool of *CXCR3* + T cells capable of responding to elevated CXCL9/10/11 chemokine gradients in the lung.

To gain further insight into T and NK cell functional changes observed in old lung donors, we examined the age-association of *IFNγ* expression and T/NK cell modules presented in the previous section across both lung and PBMC donors. Interestingly, Tm2, Tm7, Tm8, and *IFNγ* CD8 T cell expression correlated with age in older blood donors, while Tm4 and Tm12 expression declined with age in PBMCs, unlike the lung (Fig. 6d). These trends were also highly similar in NK and CD4 T cells from lung and blood donors (Supplementary Fig. 7d, e). Since chronic activation and cytotoxicity of T cells upon stimulation can lead to exhaustion[97,98], we also examined if markers of exhaustion (*TIGIT*, *HAVCR2*, *PDCD1*, and *LAG3*) demonstrate heightened expression with age. Indeed, we found increased T cell expression of exhaustion markers with age in blood and lung samples (Fig. 6d, Supplementary Fig. 7d, e). Taken together, we see that signatures associated with hyperactivity of T cells (*IFNγ*, exhaustion, activation, cytotoxicity) are elevated with age in both lung and blood, while TNF*α* pathway and HSP stress response modules have opposing age trends in the two tissues.

https://doi.org/10.1038/s41467-026-75427-5

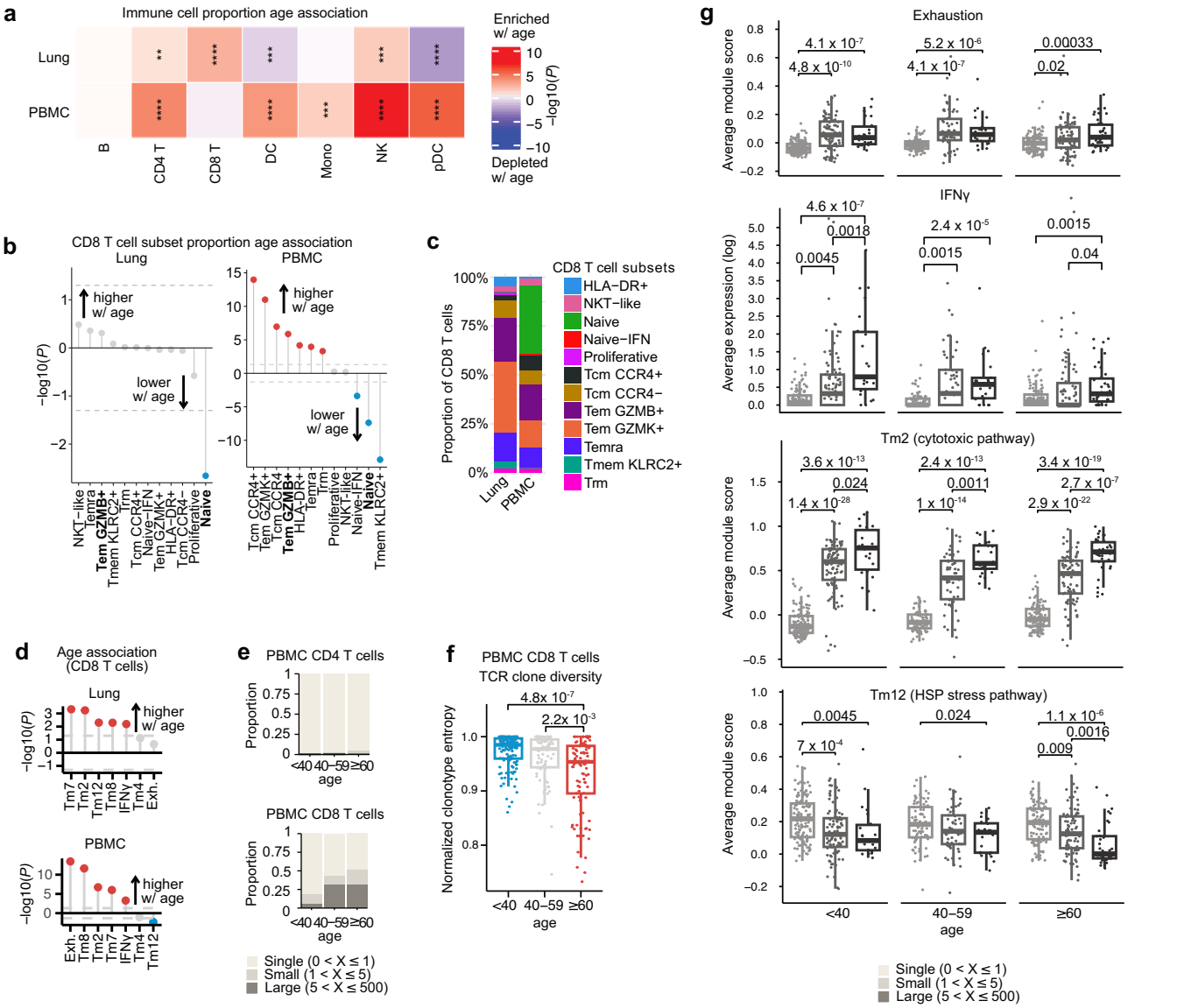


Fig. 6 | Investigating T cell subset abundance and functional module aging patterns in human lung and blood. a Heatmap of CLR-transformed cell abundance correlation (Spearman) with age across HLCA lung (top) and human PBMC (bottom) samples for cell types found in both data sets. DC = dendritic cells. Mono.=monocytes. pDCs=plasmacytoid DCs. Two-sided Spearman *P* values were computed. **P* ≤ 0.05, ***P* ≤ 0.01, ****P* ≤ 0.001, *****P* ≤ 0.0001, and *P* > 0.05 not shown. **b** Plot showing the correlation between CD8 T cell subset proportions and age across HLCA lung and PBMC samples, ranked by FDR-corrected, two-sided Spearman correlation *P* value. Y axis reports -log10 (*P* value) signed by correlation direction. **c** Stacked boxplot showing lung and PBMC CD8 T cell subset relative proportions when aggregated across all samples. **d** Plot showing the correlation between lung T/NK modules scores and age across HLCA lung (top) and PBMC (bottom) in CD8 T samples, ranked by FDR-corrected, two-sided Spearman correlation *P* value. Y axis reports -log10 (*P* value) signed by correlation direction. **e** Stacked bar plots of PBMC CD4 (top) and CD8 (bottom) T cells' clone size proportion across age groups (CD4, left-to-right, single to large clonotypes: *n* = 129, 74, 6, 80, 45, 1, 108, 80, 16 samples; CD8, left-to-right, single to large clonotypes: *n* = 128, 91, 22, 80, 54, 24, 108, 85, 38 samples). **f** Boxplot of PBMC CD8 T cell normalized entropy/diversity of clonotypes across age groups (left-to-right: *n* = 128, 80, 108 samples). FDR-corrected, two-sided Mann-Whitney-Wilcoxon test was used to compute *P* values. **g** Box plots of Exhaustion (top left), Tm2 (bottom left), Tm12 (bottom right) average module scores, and IFNγ (top right) average expression across CD8 T cells separated by age groups and clonotype size (left-to-right: *n* = 128, 91, 22, 80, 54, 24, 108, 85, 38 samples). FDR-corrected, two-sided Mann-Whitney-Wilcoxon test *P* values were computed between clonotype size groups. Box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually. *P* values > 0.05 are not shown.

The availability of single-cell TCR clonotype information in the PBMC cohort enabled us to study clonal expansion during aging, where a prior longitudinal human blood study using bulk RNA-seq study showed reduction of TCR repertoire richness and increased clonal expansion with age[99]. Our single-cell resolution analysis corroborated these findings, where expanded TCR clonotypes made up a larger fraction of both CD4 and CD8 T cells in older donors (Fig. 6e) while TCR entropy (richness) declined with age (Fig. 6f). We next asked if *IFNγ*, exhaustion markers, and CD8 T module expression that changes with age are associated with clonotype expansion. Indeed, we saw that *IFNγ*, exhaustion, Tm2, Tm7, and Tm8 expression was significantly elevated in CD8 T cells with expanded clonotypes amongst all age groups, while Tm4 and Tm12 showed the opposite trend (Fig. 6g, Supplementary Fig. 7f). To ensure that these biological findings were not skewed by a few outlier expanded TCR clonotypes, we removed all PBMC samples where the largest clonotype accounted for

more than 5% of the CD8 T cells in a given sample and observed highly similar results (Supplementary Fig. 7g, h).

In sum, comparing blood and lung enabled us to gain a deeper understanding of aging in CD8 T cells, where (a) we see opposing aging trends in total CD8 T cells but similar cross-tissue trends in CD8 T cell subset abundance, (b) we see opposing age-association in stress-related T/NK cell modules expression (Tm4 and Tm12) but similar cross-tissue increase in modules related to CD8 T cell hyperactivity (*IFNγ*, exhaustion, activation, cytotoxicity), and (c) we find that T cell modules with age-associated changes in expression also exhibit similar changes in expression in expanded clonotype T cells, which could explain alterations in their function with age. These findings highlight both tissue-specific and cross-tissue aspects of immune aging, revealing conserved and divergent patterns that contribute to our understanding of age-related changes in the immune system.

Multi-cellular aging mechanisms in the lung parenchyma

To map coordinated changes in the lung cellular ecosystem that underlie aging, we performed multi-cellular community analysis to link cell subsets with similar cell composition changes across samples, which were shown previously to be perturbed in the aging human brain and Alzheimer's disease[100]. AT2 and macrophage subsets showed similar patterns in cellular decline with age across samples and were strongly anti-correlated with stromal (fibroblast, endothelial, smooth muscle) and immune (T, NK, mast) cell subsets that demonstrated consistently higher abundance in older lung donors (Fig. 7a). In agreement with a decline of alveolar regenerative capacity, both AT2 proliferative cells and transient bipotent progenitor AT0 cells[18] also showed coordinated decrease in abundance with age across lung donors.

To identify aging mechanisms that are shared across cell types, we also examined the similarity between cell type modules using Jaccard index (Fig. 7b). We observed several distinct clusters of age-associated modules with similar signatures across cell types, including TNFα pathway, HSP/stress, ATP/mitochondrial dysfunction, IFN response related modules that were highly similar to the previously discussed Tm4, Tm12, Mm3, and Mm29 modules, respectively. Worth noting, similar modules within clusters also exhibited similar age-association across cell types (Fig. 7b; color bars). Interestingly, TNFα pathway module that increased in expression with age co-expressed several activator protein 1 (AP-1) transcription factors (JUN, JUND, FOS, and FOSB), which form dimers and have been shown to drive the enhancer landscape and transcriptional program of senescence[101]; hence, our analysis argues this age-related mechanism is likely shared between cell types. Overall, these findings provide valuable insights into the complex interplay between different cell types in the aging lung, highlighting both cross-cell type and cell type-specific aging mechanisms.

Single-cell resolution predictor of lung parenchyma age

Vast progress has been made in using chronological age plus other measurable features to predict biological age[1,102,103], which represents the most important risk factor in many diseases, including respiratory diseases such as COVID-19[7]. Prior work to quantify lung biological age used bulk genomic data and primarily focused on classification of patients into young versus old categories[104–106]. We reasoned that our systematic, single-cell resolution study of cellular aging can more accurately model the biological age of lung donors and also uncover putative biomarkers of aging. To achieve this, we trained an XGBoost (eXtreme Gradient Boosting) machine learning (ML) model (Fig. 7c; Methods) to predict the biological age of lung samples using cNMF module scores as features (Supplementary Data 4). The model starts with all cellular module scores and chronological age as input, is trained using 5-fold cross-validation to predict biological age, and iteratively reduces the set of features through backward feature selection, keeping the 200 most informative features at first and thereafter removing the 5 least informative features per training cycle (Methods). To compare the predictive power of single-cell versus bulk RNA-seq data, we used the same feature space (scRNA-seq derived cNMF modules) to train an analogous model using module scores calculated on bulk RNA-seq data and age information from GTEx. Examining the models' R^2 performance when using scRNA-seq versus bulk data across 50 independent runs showed that the single-cell HLCA cohort significantly outperformed the bulk GTEx cohort regardless of the number of features used, demonstrating the higher predictive power of single-cell data (Fig. 7d, Supplementary Fig. 8a). The improved model performance of the scRNA-seq cohort was evident despite >3X fewer samples than the GTEx cohort, which might also explain the higher variance in performance observed for the HLCA cohort. We also evaluated the performance of our model on two unseen scRNA-seq validation cohorts[107,108] and found similar or superior prediction accuracy as the HLCA scRNA-seq cohort, confirming that our model robustly captures age-related transcriptomic changes in different single-cell datasets.

Our ML framework also enables ranking the predictive importance of cellular features, which represent potential biomarkers of lung aging. When examining the 20 features most frequently appearing within the 30 features most predictive of biological age, which exhibited similar median performance as models with the 200, 100, and 50 most informative features (Fig. 7d, Supplementary Fig. 8a), we identified several modules that closely align with our previous analyses, such as the Macrophage.25 IFN response and the Macrophage.4 ATP module that correspond to Mm29 and Mm3, respectively (Fig. 7e). Worth noting, 10 of the 20 most informative features were highly correlated with age, which is significantly higher than expected by chance (Supplementary Fig. 8b, c; permutation test P value = 0.0072). Hence, our ML approach not only serves as a future diagnostic of lung disease risk but can also rank the importance of lung cellular biomarkers in predicting age-related decline.

In summary, we built a single-cell resolution clock of the lung parenchyma that shows superior predictive power over similarly trained models based on bulk RNA-seq data. The model performs well on validation cohorts not used in training and also aids in identifying the most predictive cellular biomarkers of lung aging, which consist of both cell type-specific and cross-cell type gene modules.

Discussion

In this study, we devised a data-driven analytical strategy to systematically dissect human lung aging at cellular resolution by leveraging 184 scRNA-seq lung parenchyma samples from the HLCA[25] and two independent validation cohorts–Xenium spatial transcriptomics data generated here and bulk RNA-seq data from GTEx. Our computational approach uncovered consistent cell composition alterations in older lung donors across all three cohorts, including a decrease in alveolar and macrophage cell subsets and elevated abundance in T, NK, mast and stromal cells populations. These findings largely confirmed prior work[5,20,21,26] and extended it to cell subset resolution in humans, including demonstrating age-associated decline in AT0 progenitors, proliferating macrophages and AT2 cells that likely impact the alveolar and myeloid regenerative capacity in the lung.

The ability of our computational platform to consistently replicate known age-associated cell composition changes in three independent cohorts encouraged us to define the cellular and molecular aging mechanism using de novo module discovery across our scRNA-seq cohort. In total, we identified 142 age-associated gene expression modules (Supplementary Data 4) related to several established hallmarks of aging:

Stem cell exhaustion-We identified a PATS-like epithelial cell module Em14 that was upregulated with age, which likely represents an aberrant alveolar cell state that is also enriched in respiratory diseases,

https://doi.org/10.1038/s41467-026-75427-5

such as ILD and COVID-19[55,56]. Coupled with decreased AT2 cell proliferation and autocrine WNT signaling interactions, our data points to dysfunctional AT2 stem cell niche and alveolar regeneration in older lung donors.

Mitochondrial dysfunction-Age-associated decline in activity of module Mm3 co-expressing several ATP synthase and oxidative phosphorylation genes are indicative of decreased mitochondrial energy production capacity, which we functionally validated in human monocyte-derived macrophages. Similar modules were also found to significantly decrease with age in basal, suprabasal, AT1, NK, mast, vascular, and lymphatic endothelial cells (Fig. 7b), suggesting that mitochondrial dysfunction is shared across multiple cell types.

Chronic inflammation-Modules related to inflammation and TNFα signaling (e.g., Tm4) were also found to be significantly upregulated with age across multiple cell types, including macrophages, secretory, T, plasma, mast, and NK cells (Fig. 7b). Interestingly, prior work showing that reduced oxidative phosphorylation acts to keep macrophages in a pro-inflammatory state, suggests a self-reinforcing relationship between mitochondrial dysfunction and inflammation in macrophages[109].

Loss of Proteostasis-The most widely upregulated module across cell types with age consisted of several HSPs (e.g., Tm12), known to act as chaperones and mediate proper protein folding[110]. Modules significantly upregulating HSPs with age were found in DCs, monocytes,

Fig. 7 | Single-cell resolution predictor of lung parenchyma age. a Community analysis network plot across HLCA samples using fine-grained cell type annotations. Edges represent significant Spearman correlation (|R| ≥ 0.3) between cell type abundances, size of nodes represent Spearman correlation significance (-log10 *P* value) between cell type abundances and age, and color of nodes/edges represents correlation sign (red = positive and blue = negative correlation). **b** Heatmap of Jaccard similarity indexes between top 25 genes for each pair of cNMF modules, discovered in cell types with at least 400 cells. Only modules with significant age-association (FDR-adjusted Spearman correlation $P \le 0.05$) and having a Jaccard similarity ≥ 0.3 with another module were included to facilitate interpretation and visualization. **c** Schematic of the ML model trained to predict lung age based on cNMF modules discovered in this study, created in BioRender. Tsankov, A. (2026) https://BioRender.com/upv29gj. **d** R^2 performance of XGBoost ML model for the top 200, 100, 50, and 30 features (after backward feature selection) across 50 training runs using the GTEx bulk (gray), HLCA scRNA-seq (blue), and two validation scRNA-seq (green) datasets (left-to-right: $n = 578$, 181, 8, and 10 scRNA-seq samples). Two-sided Mann-Whitney-Wilcoxon test *P* values were calculated to compare performance between datasets across 50 training runs for different numbers of features used. **e** Bar plot of the top 20 features ranked by the fraction of times a feature appeared in the 30 most important features following backward features across 50 model runs. We chose 30 most important features for evaluation, since it represented an inflection point in performance exhibiting the highest median R^2 among runs with <50 features (Supplementary Fig. 8a). Module activities with significant age-association (FDR-adjusted, two-sided Spearman correlation *P* value ≤ 0.05) are colored in blue or red. Box plot horizontal lines show the 25th, 50th (median) and 75th percentiles, with vertical whiskers extending to a maximum distance of 1.5× the interquartile range from the hinge. Data beyond the whisker ends are plotted individually. *P* values > 0.05 are not shown.

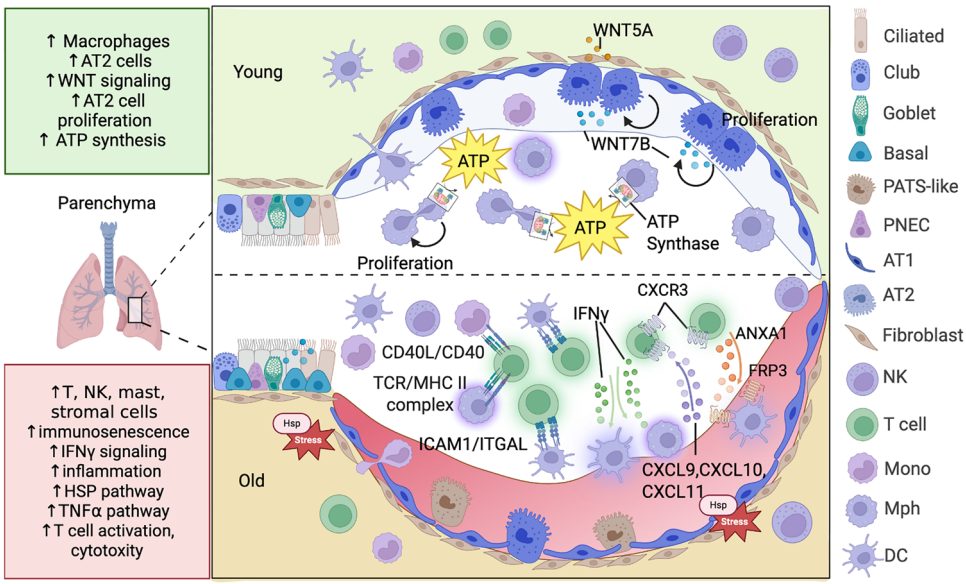


Fig. 8 | Cellular hallmarks of lung parenchyma aging revealed by single-cell and spatial transcriptomics. Schematic illustrating the main findings from this study, created in BioRender. Tsankov, A. (2026) https://BioRender.com/upv29gj.

basal, suprabasal, T, B, NK and mast cells (Fig. 7b), and may be a compensatory mechanism to counteract the decreased stability and functionality of proteins in older lung donor samples.

Genomic instability-We identified basal cell module Em18 expressing *STING1* and other genes related to the cGAS-STING pathway, which is known to be induced by aberrant host DNA that accumulates in the cytosol following DNA damage[111].

Senescence-Using 12 established gene signatures[38,40–47,49,50], we showed an overall increase in senescence with age that was highly cell type specific. While *CDKN2A* and *CDKN1A* orthologs have been shown to be effective markers of senescence in mice, in the human scRNA-seq data we observed highly cell type specific age-association in expression of different senescence signatures and markers. Notably, we observed higher expression of MIF in old lung donors, which our team identified to be a critical modulator of age- and smoke-related chronic lung disease, thereby revealing MIF as a future lung-targeted senotherapeutic[112].

Integration of the scRNA-seq and spatial data also enabled us to take a system-level approach to studying another hallmark of aging, altered intercellular communication (Fig. 8). Most prominently, we found increased expression of *IFNγ* in T and NK cells of older lung donors, which likely led to higher responsiveness to this molecule observed in a number of cell types, including macrophages (Mm29), monocytes, and DCs (Fig. 7b). Mm29 in turn expresses chemokines *CXCL9/10* that are known to recruit T cells, likely leading to a higher T cell abundance observed in both HLCA and Xenium older lung samples (Fig. 8). We also observed increased ligand-receptor gene interactions between T and APCs that are associated with T cell activation, including *CD40LG-CD40*, *ITGAL-ICAM1*, and *ANXA1-FPR3*. In support of these observations, de novo spatial cell neighborhood detection identified a *CXCL9*+ myeloid niche predominantly found in older lung samples with elevated proportions of T cells and DCs, which likely contributes to increased T cell activation and abundance. The aberrant T-myeloid cell cross-talk in older lung donors likely establishes a hyperactive adaptive host immune response with higher effector, cytotoxic, and exhausted module activities in older lung and blood donors, especially in expanded TCR clonotypes. This T cell overstimulation may be a compensatory mechanism due to the decline in diversity of memory T cells and effectiveness of immune surveillance in older individuals[82].

The comprehensive catalog of age-associated cellular modules also enabled us to build a single-cell resolution predictor of lung parenchyma biological age. Our ML model trained on the HLCA scRNA-seq data showed similar performance on two unseen scRNA-seq validation cohorts and outperformed a similar model trained on bulk RNA-seq data. Future studies that collect other types of data modalities (e.g., DNA methylation, metabolomic, epigenomic) along with scRNA-seq from the same lung donors will likely further improve on the performance of the method presented here. Nevertheless, our framework helped identify the most predictive gene modules of lung

biological age, including IFN responsive and oxidative phosphorylation myeloid modules Mm29 and Mm3, that can form the basis of improved cellular biomarkers of risk for respiratory diseases[7].

Despite providing a holistic, system-level view of lung cellular aging, we acknowledge several limitations to our study. First, while our goal was to study normal human aging using donor lungs with no reported lung disease comorbidities, these lung samples are not fully healthy and lack information about several important covariates/confounders, including exposure to pollution, history of mechanical ventilation, viral infections, subclinical fibrosis, and functional lung decline. 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. Fourth, the unpaired nature of our lung and blood cohort do not permit direct tracking of TCR clonotypes or cell migration trends and other functional assays will be necessary to ascertain the role of chemokines CXCL9/10/11 in lung T cell chemotaxis. Finally, epigenetic and proteomic data from a large cohort of lung samples can enable a direct comparison of our single-cell lung age predictor and other proteomic and epigenetic clocks in the future.

Overall, our data-driven approach revealed age-related shifts in cell composition, signaling dysfunction, and intercellular communication in the aging lung, underscoring the multifaceted impact of aging on cellular mechanisms, such as impairing lung regeneration. Our work will serve as a rich data and analytical resource (see Data and Code availability) for future studies focused on lung cellular aging and will form the foundation for putative biomarkers of lung biological age and disease risks.

Methods

Ethical regulations

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. All autopsy tissue samples were elected to have a post-mortem interval (PMI) ≤ 18 h from donors with no lung disease comorbidities, and were confirmed to be clinically and histologically normal.

Lung single-cell RNA-seq data acquisition and processing

Human Lung Cell Atlas scRNA-seq data from declined donor lungs was downloaded from (https://cellxgene.cziscience.com/collections/6f6d381a-7701-4781-935c-db10d30de293) and loaded into a Seurat object using the Seurat V.4.0.5 R package. Only scRNA-seq normal lung parenchyma samples from the same technology (Chromium platform, 10X Genomics) were subsetted using the metadata information, excluding samples that lacked age information, all cellular compartments (e.g., CD45+ sorted samples), with known lung disease comorbidities (e.g., lung cancer) and samples from brushings, scrapings, biopsies, resections that are known to affect cell type compositions (e.g., brushings, scrapings enrich for epithelial cells). Ambient RNA was removed from the HLCA core data sets as described in ref.[25], while extended HLCA data sets underwent ambient RNA correction using FastCAR[113]. The final set of samples used and clinical information is displayed in Supplementary Data 1. Data normalization, variable feature selection, scaling, and principal component analysis (PCA) dimensionality reduction were performed with Seurat functions NormalizeData, FindVariableFeature, ScaleData, and RunPCA, respectively. The data was integrated using the RunHarmony function from harmony V.0.1.0, correcting for both "dataset" and "sample". The Seurat object was then subsetted into epithelial, myeloid, T-NK, B-plasma-mast, mesenchymal, and endothelial compartments based on known cell class markers. Each compartment was renormalized and harmonized using the same functions as described above. For studying alveolar regeneration in our epithelial compartment analysis (Fig. 2, Supplementary Fig. 2), we focused on cells directly related to the alveolar lineage, including AT1, AT2, secretory, suprabasal and basal cells. Cells with missing annotation information were annotated within each cell compartment using label transfer[29], where cells with annotation information formed the reference population. Transferred annotations were independently validated by canonical marker gene expression, with high concordance between predicted labels and known lineage markers. Furthermore, the increased statistical significance observed with higher scRNA-seq sample sizes in our power analysis (Supplementary Fig. 1d) argues that the label transferred annotations were biologically meaningful and did not suffer from technical (non-biological) batch effects.

Tissue microarray (TMA) construction

To maximize sample size per Xenium run and minimize batch effects between samples, multiple normal lung cores were placed on a TMA of similar size as a Xenium slide (10.45 mm×22.45 mm). Alveolar, bronchial, and vascular lung tissue regions were annotated on a hematoxylin and eosin (H&E) stain by our team's pulmonary pathologist and a circular punch of 1.5 μm diameter was used to acquire individual cores from each formalin-fixed paraffin-embedded (FFPE) block. Two TMA blocks were created with 44 cores on each TMA block (5×9) and sectioned for downstream immunofluorescence staining and spatial profiling via the Xenium platform (10X Genomics). Several cores detached during the sectioning and attachment process; hence, a total of 70 cores remained intact for Xenium spatial transcriptomics profiling.

Xenium data generation and preprocessing

Two TMA blocks were loaded into a single Xenium run, with a gene panel of 389 genes (Supplementary Data 6) consisting of the standard lung Xenium gene panel and 100 custom genes related to lung aging. Cell segmentation of the Xenium 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.

Raw data were loaded as an anndata V0.9.1 object using scanpy V1.10.3. Quality metrics were calculated using 'scanpy.pp.calculate_qc_metrics' and cells were filtered out using 'scanpy.pp.filter_cells' with parameters min_genes=20, min_counts=30, and max_counts=1000. Raw counts were normalized using 'scanpy.pp.normalize_total' with parameter target_sum=1e4 and log normalized with 'scanpy.pp.log10p'. Principal component analysis (PCA) dimensionality reduction was performed using 'sc.pp.pca'. The data was

integrated using Harmony with 'scanpy.external.pp.harmony_integrate' on donor id.

Xenium cell type annotation, composition, cell-cell interaction, and niche analyses

HLCA single-cell RNA seq data was used as reference. SCVI model from scvi-tools V0.20.3 was used to integrate the single-cell and xenium data, with parameters: n_layers=3, n_latent=32, n_epochs=20. SCANVI model was used to assign cell type labels to cells in the xenium cohort, n_epochs=20. Transferred annotations were independently validated by canonical marker gene expression, with high concordance between predicted labels and known lineage markers.Composition of each cell type per sample was calculated using pandas V2.0.3 method value_counts. Association between age and cell proportions was determined using Spearman correlation with the age values.

For ligand receptor analysis 'st.tl.cci.run' from stLearn V0.4.12 with parameters n_pais=10000, radius=50 for ligand-receptor pairs *ICAM1-ITGAL* and *CD40LG-CD40* and radius=100 for *WNT7B-LRP6* and *ANXA1-FPR3*. 'st.tl.cci.run_cci' was used for getting significant cell type-cell type interactions with parameters min_spots=3 and n_perms=1000. SOAPy V1.0.1 library was used to detect de-novo niches. Neighborhood network was generated using 'sp.pp.make_network' with parameters method='radius' and cutoff=50. Niches were detected using 'sp.tl.get_c_niches' with parameters k_max=30.

Age-associated cell type composition and senescence analysis

Composition of each cell type per sample was calculated by R function prop.table. Association between age and cell proportions was determined using Mann-Whitney-Wilcoxon test between any two age groups, or Spearman correlation with the actual age values, which treated age as a continuous variable and did not rely on stratifying samples into groups. All samples with <1,000 total cells were excluded from this analysis, due to greater uncertainty of cell proportions in samples with fewer cells. We performed correction for multiple hypothesis testing (multiple cell types) using the Benjamini-Hochberg method[30] and considered a False Discovery Rate (FDR) of 0.05 as significant.

Centered log ratio (CLR) transform for removing dependencies between cell fractions

First, cell type proportions were normalized to one separately within the immune and non-immune cell compartments to alleviate differential enrichments of immune and non-immune cell types during lung tissue dissociation prior to scRNA-seq profiling. For CLR transformation, we let there be D cell types per sample and $x_i \in \{x_1, \ldots, x_D\}$ where x_i represented the observed proportion for each cell type i in a given sample. A small pseudocount $\epsilon = 10^{-6}$ was added to all x_i to avoid log(0) operations. The CLR transformation was then computed as:

$$clr(x_i) = \log(x_i + \varepsilon) - \frac{1}{D}\sum_{j=1}^{D} \log(x_j + \epsilon). \quad (1)$$

This transformation centers the log-transformed proportions within each sample, yielding values with zero mean across cell types, thus removing the unit-sum constraint for cell fractions within a sample. The resulting CLR-transformed proportions were used as continuous response variables in our generalizable linear model analyses and rank-based Spearman correlation analyses. Although the CLR transformation introduces a linear dependency (i.e. $\sum_{i=1}^{D} clr(x_i) = 0$) among components, this does not affect statistical inference in our framework because each cell type was modeled separately rather than jointly.

Gene expression and visualization

The entire lung parenchyma Seurat object was subsetted to cell type level objects and log-normalized. For each cell type, average gene expression across each sample was calculated using the AverageExpression function from the Seurat package. Age groups gene expression comparison was performed using the Mann-Whitney-Wilcoxon test. Gene expression was visualized by dot plots using Seurat function Dotplot. For senescence analysis dot plots, all cells were included. Cell cycle scores were calculated as the average log-normalized expression of five established cell proliferation marker genes (Supplementary Data 2). For other analysis dot plots, cell type that have fewer than 400 cells (Goblet, Ionocyte, Mesothelium, Pulmonary neuroendocrine cells - PNECs, Hematopoietic stem cells - HSCs, Tuft, Hillock-like cells, Plasmacytoid dendritic cells - pDCs) were excluded to facilitate visualization.

cNMF module discovery across cell compartments

Consensus non-negative matrix factorization (cNMF) was implemented in the Python package cNMF V.1.3.4 to identify gene modules in each of the following cellular compartments: epithelial, myeloid, and T/NK. For each compartment, we tested from 5 to 35 K with 100 replicates and filtered outlier components with Euclidean distance >0.3 from their nearest neighbors. We observed a stable regime in which gene programs were reproducibly identified and enriched for coherent biological pathways. Lower K values captured broad, conserved programs, whereas higher K values refined these into more specific cell state or condition-associated programs. Given our focus on lung aging, we opted for selecting higher K values for different compartments to capture cell type specific age-associated responses. Then, based on the trade-off between reconstruction error and factorization stability as done in ref.[114], we selected the most appropriate Ks, where we used K of 27, 35, and 30 for epithelial, myeloid, and T/NK compartments, respectively. We found that different K values could more robustly capture age-associated programs in different cellular compartments due to lower/higher diversity of cellular identity and cell activity transcriptional programs. For example, the myeloid compartment ($K = 35$) contained higher diversity of cell type identities/activities than the T/NK cell compartment ($K = 30$).

Next, we computed cNMF module scores by first taking the top 25 genes ranked by spectra scores for each cNMF module and then using Seurat function AddModuleScore. To find associations between cNMF module scores and age, we first computed the mean score for each cNMF module across cells of a given compartment and then ran two different analyses: 1) Mann-Whitney-Wilcoxon test between age ≥60 and age <40 samples and 2) Spearman correlations with FDR correction between mean cNMF scores and age across samples. For each cell compartment, different numbers of age-associated cNMF modules were identified (Supplementary Data 4).

Differential expression heatmap for cNMF module genes

Muscat V.1.8.2 package was used to compute differential gene expression between different age groups across relevant cell types using the DEseq2 method. The top 25 genes for each age-associated module were visualized using pheatmap() from pheatmap V.1.0.12 package.

Pseudotime trajectory analysis

Monocle3 V.1.0.0 was used to convert the epithelial seurat object into CellDataSet (cds) object and UMAP size factors estimation and graphic learning was achieved by estimate_size_factors and learn_graph functions. The earliest node was defined by the closest node from "AT0" cell types and pseudotime was visualized by plot_cells.

CellphoneDB ligand-receptor analysis

CellPhoneDB V.4.0.0 was used to predict significant ligand-receptor interactions using normalized raw counts and annotation of broad cell

classes as input. The method was run with default parameters separately for each sample ID to avoid batch effects between samples. Interactions with a *P* value below 0.05 were classified as significant. The significant means output was converted into a binary matrix based on presence or absence of a significant interaction. Normalized enrichment difference (*ED*) of a specific interaction was calculated as done previously[59,115]:

$$ED = \frac{nA}{NA} - \frac{nB}{NB} \quad (2)$$

where *nA* and *nB* represent the number of samples with the interaction in groups A and B, respectively, while *NA* and *NB* represent the total number of samples in these groups. Significance of enrichment for an interaction was calculated using Fisher's exact test. Negative log10 *P* values and normalized enrichment differences for each ligand-receptor interaction and cell pair combination were plotted using the ggplot2 V.3.4.3 package in R.

Bulk RNA-Seq datasets acquisition and analysis

Reads Per Kilobase per Million mapped reads (RPKM) matrix including 578 bulk RNA-seq lung samples as part of the Genotype-Tissue Expression (GTEx) database was downloaded from the website https://gtexportal.org/home/downloads/adult-gtex/bulk_tissue_expression along with metadata information. Ensembl gene IDs were converted to gene names and duplicated mappings were removed. To identify age association with cell type abundances, we binned samples into three age groups (<40, 40–59 and ≥60 years of age) and estimated cell type abundances using the average expression of a curated list of normal lung cell type markers (Supplementary Data 2). Differences between any two age groups were assessed for significance using the Mann-Whitney-Wilcoxon test. As a complementary rank-based analysis, we also performed Spearman correlation between age and estimated lung cell proportions across GTEx samples.

Immunofluorescence

Immunofluorescence staining was performed on serial sections of the TMAs used for Xenium spatial transcriptomics. Sections were deparaffinized, rehydrated, and subjected to antigen retrieval using RNAscope Target Retrieval Reagent (10X) (Catalog #322001, Advanced Cell Diagnostics). Following antigen retrieval, sections were blocked for 2 hours in a solution containing 5% horse serum (Catalog #16050122, Gibco), 10% BSA (Catalog #BP1600-100, Fisher Scientific), and 0.5% Tween-20 (Catalog #P2287, Millipore Sigma) in 1× PBS. Slides were then incubated overnight at 4 °C with an anti-KRT8 primary antibody (TROMA-I; DSHB, Catalog #AB_531826; 1:100 dilution). The following day, sections were incubated with a fluorescently conjugated secondary antibody (Invitrogen, Catalog #A48269; 1:1000 dilution) for 1 h at room temperature. Nuclei were counterstained with DAPI (Catalog #D1306, Fisher Scientific, 10ug/mL) for 5 min. Z-stack and tile images were acquired using a Zeiss Axio microscope under identical acquisition settings across all samples. Image analysis and quantification were performed using Zeiss ZEN software. For each TMA core, the number of $KRT8^+$ cells was quantified and expressed as a percentage of total $DAPI^+$ nuclei/cells. A total of 28 cores from the young group (age ≤ 40) and 31 cores from the old group (age ≥ 60) were analyzed. The remaining cores either detached during TMA construction, resulted in poor quality images, or fell outside of the two age ranges. Data are presented as the percentage of $KRT8^+$ cells per core, with bars representing group means ± standard deviation (SD).

Macrophage differentiation procedure

Peripheral blood mononuclear cells (PBMCs) were isolated from healthy donors using Lymphoprep™ (STEMCELL Technologies) density gradient centrifugation. Human monocytes were then enriched using the MojoSort™ Human Pan Monocyte Isolation Kit (BioLegend) according to the manufacturer's instructions. Isolated monocytes were counted and plated onto non-tissue culture-treated plates. Cells were cultured in complete RPMI medium supplemented with 10% fetal bovine serum (FBS), 1% penicillin-streptomycin, 1× non-essential amino acids, 2 mM L-glutamine, and 1 mM sodium pyruvate. To promote macrophage differentiation, 50 ng/mL of macrophage colony-stimulating factor (M-CSF) was added. Media were replaced every 3 days. On day 8, differentiated macrophages were detached using 10 mM EDTA in PBS.

Mitochondrial stress Seahorse assay

Oxygen consumption (OCR) was determined using a Seahorse XF-96 Extracellular Flux Analyzer (Seahorse Bioscience). Macrophages are plated at 70,000 cells/well in 96-well plates overnight. Cells are washed and incubated with XF RPMI stock media adding 10 mM Glucose, 1 mM pyruvate and 2 mM glutamine at 37 °C in a non-CO2 incubator for 60 min. The measurements were performed at 5 min intervals (1 min mixing, 1 min recovery, 3 min measuring) for 1.5 h. Inhibitors were used at the following concentrations: 1 µM oligomycin, 1.7 µM FCCP, 1 µM rotenone, and 1 µM antimycin A (all from Seahorse Bioscience).

PBMC data acquisition and analysis

Single cell RNA-seq data and TCR contigs from 317 peripheral blood mononuclear cell (PBMC) samples from Synapse were retrieved from accession code syn49637038. One TCR sample was omitted from the Synapse submission; hence, all subsequent analysis was performed with 316 samples. Donors with inflammatory conditions, cancer, certain lifestyle variables like smoking, abnormal clinical blood panel tests, and cold or flu symptoms in the prior month were excluded. Seurat objects created for each cell type were created following code provided at https://github.com/teresho4/scRNA-seq_atlas_Hs_PBMC_aging. PBMC composition and boxplot analysis followed the same analysis pipeline as for the lung.

Blood scTCR-seq analysis

Using scRepertoire V.2.0.0 package, contigs were combined using combineTCR function, removing any cell barcode with an NA value in at least one of the chains, and any cell barcode with more than 2 immune receptor chains. Multiple reads were combined into clone calls by using the VDJC gene sequence. Clonotypes were labeled as single ($n = 1$), small ($n > 1$ and $n \leq 5$), and large ($n > 5$). To assess association with clonotype size and/or age, the *IFNγ* gene expression per sample was computed using the Seurat AverageExpression across CD8 T cells, and all other module scores were assessed by using the Seurat AddModuleScore function and then averaged across CD8 T cell subsets per sample and per clonotype.

Clonotype entropy/diversity analysis

Clonotype entropy, or diversity, was calculated using the clonalDiversity function from scRepertoire package with the norm.entropy metric. Clones were called using VDJC genes and grouped by sample. Mann-Whitney-Wilcoxon test was used with FDR correction on *P* value between age groups[100].

Community analysis

Community network construction and visualization were generated using the igraph V.1.5.1 package, by following the methods described before[100]. Node size scales with the age-association of each cell subset composition, while edges were created by Spearman correlation between different cell subset compositions. Only edges with absolute Spearman correlation coefficient > 0.3 were visualized in the network. Node size represents Spearman correlation significance (-log10 *P* value) between cell type abundances and age, and color of nodes/

edges represents correlation sign (red = positive and blue = negative correlation).

Jaccard similarity matrix

A Jaccard matrix was created by computing the Jaccard similarity between the top 25 genes of each pair of cell type cNMF modules. The Jaccard similarity was calculated as follows:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

where A represents genes for cell type cNMF module A, B represents genes for cell type cNMF module B, and $J(A,B)$ represents the Jaccard similarity in genes between a pair of modules. $J(A,B)$ is always between [0,1]. To facilitate interpretation and visualization, cNMF modules of cell types with at less than 400 cells were removed (Goblet, Mesothelium, pDCs, HSCs, Ionocytes, PNECs, Tuft, Hillock-like) and modules representing expression of two cell types likely arising from doublets (a total of 39 modules) were removed. Furthermore, only modules with significant age-association (FDR-adjusted Spearman correlation $P < 0.05$) and that have Jaccard similarity ≥0.3 with another module were displayed in Fig. 7b.

XGBoost validation cohorts and data acquisition

To evaluate the prediction accuracy of our ML model on unseen data, we used two independent single-cell RNA-sequencing cohorts. The first validation cohort of 8 parenchyma samples was derived from the LungMAP project as published by Guo et al.[107] and ranged from ages 21 to 63, where data can be downloaded under Human CellCards Multi-Study CellRef 1.0 Atlas in the Lungmap atlas. The second validation dataset was obtained from Natri et al.[108] where data can be downloaded using GEO accession number GSE227136. From this dataset, we removed samples already in HCLA and only used the following 10 samples with an age range of 17 to 60: THD0002, THD0016, THD0021, THD0022, THD0023, VUHD112, VUHD113, VUHD115, VUHD117, and VUHD65. Cell type annotations were label transferred using the Seurat function FindTransferAnchors, where the HCLA scRNA-seq data served as a reference.

XGBoost regression model training for age prediction

For predicting lung donor biological age, we trained XGBoost regression models using XGBRegressor from the Python XGBoost package V.2.1.0 expanding on our previous work[116]. Input features were created by running cNMF to discover 30 gene expression modules per cell type (29 cell types total) (Supplementary Data 4), and taking the average module score of the 25 top genes for each module across cells in the corresponding HLCA cell type per sample using Seurat addModuleScore function. 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, due to low and uneven cell sampling across donors. Modules of cell subsets that were absent in fewer than 50% of samples were excluded from analysis, resulting in the removal of 150 modules. To ensure comparability across datasets, the module scores were subsequently quantile normalized to achieve consistent distribution across samples. Bulk RNA-seq samples under 20 years or over 79 years were removed to match the age range of the HLCA cohort. Hence, the total number of features used summed up to 480, while the number of samples included 578 bulk RNA-seq samples and 181 scRNA samples. Since GTEx age information was provided in terms of decades (e.g., 20 s, 30 s, etc.), for training/testing the model we converted both HLCA and GTEx ages to the midpoint of the age decade (e.g., donors with ages between 20-29 were all mapped to age 25).

We randomly split the data into 5 train/test folds with a 80/20% split of training and validation respectively for cross validation. For each of the 5 runs of cross validation and model testing per fold, 10 different seeds (seeds 1 to 10) were used for seeding the XGBoost model building process. Since we used each of the 5 folds as a test set 10 times (with a different seed each time), we in effect obtained 50 estimates of model performance on an unbiased test set.

After training the model using the full feature space, we selected the top 200 features according to our feature importance score (see next paragraph). Using this feature space, we then performed iterative backward feature selection with a step size of 5 until 5 features remained. To be more specific, after reducing the feature space to 200 features, we trained a model using this reduced feature space, picked the best model according to mean performance across validation folds when performing cross-validation, ranked the 200 features based on their feature importance score for the chosen model, and eliminated the bottom 5 features. This was then repeated iteratively until 5 features remained.

Feature importance was calculated using the "permutation_importance" function from scikit-learn V.1.5.1, setting the "n_repeats" parameter to 10, after picking the best model according to cross-validation. This function implements a permutation importance mechanism: for each feature in the feature space, it randomly permutes its values and keeps all other features constant, then measures the effect of this permutation on a model's chosen performance metric by comparing the change in performance to the baseline performance when no permutation is performed. The larger the negative effect of this process on the chosen performance metric, the more important a feature is deemed to be. Since this permutation is a random process, it is useful to perform this process multiple times, and measure the mean effect of permutation on performance. We chose to repeat the process 50 times. We note that since we used cross-validation, feature importance was averaged across any given feature.

We assessed model performance by computing the R^2 score to explain variance of our model. This is computed as:

$$R^2 = 1 - \frac{RSS}{TSS}, \quad (4)$$

where RSS is the residual sum of squares, and TSS is the total sum of squares. A negative value for R^2 may occur when the RSS is greater than TSS, which means that the model's predictions are on average further from the actual values than the mean of the data.

Hyperparameter optimization was performed using the automated hyperparameter search framework Optuna[117] V.3.6.1. We used the default hyperparameter optimization strategy, Tree-structured Parzen Estimator (TPE) using the same seed as the corresponding model training run (seeds 1-10). In contrast to the naive but commonly employed strategy of randomly choosing and testing model hyperparameters settings such as grid search or random search,which is an inefficient and non-optimal strategy, we took advantage of Bayesian optimization strategies to use the history of model performance under different hyperparameter settings to explore the hyperparameter search space. Each training run in Optuna is called a "study." Each study consisted of multiple "trials," each corresponding to a specific model hyperparameter setting. At the end of a study, the best model hyperparameters were chosen based on the trial that performed best according to the mean variance explained across validation folds. The number of trials per study was set to 50 ('n_trials' in 'study.optimize' function was set to 50), to test 50 hyperparameter settings per run. During backward feature selection, this process was repeated from scratch for each feature space. The best models chosen for different feature spaces very likely differ in their hyperparameter settings. The full description of each hyperparameter can be found at https://xgboost.readthedocs.io/en/release_1.7.0/python/python_api.html, under xgboost.XGBRegressor.

Permutation test for fraction of age-associated features observed

The observed test statistic was calculated using the proportion of features with FDR-adjusted P value ≤ 0.05 among the top 20 most predictive features out of 480 total features that the model was trained on. To compute an empirical P value, we conducted a permutation test by random sampling with replacement subsets of 20 features 10,000 times and calculating the proportion of significant features within each subset of 20 randomly selected features to create a null distribution (Supplementary Fig. 8c), against which the observed test statistic was compared.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

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 https://doi.org/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.

Code availability

All source code for reproducing our results and for enabling future lung aging exploration has been deposited at our lab's GitHub space (https://github.com/TsankovLab/sc_Aging_clock).

References

1. Ferrucci, L. et al. Measuring biological aging in humans: a quest. *Aging Cell* **19**, e13080 (2020).
2. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
3. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: an expanding universe. *Cell* **186**, 243–278 (2023).
4. Bowdish, D. M. E. The Aging Lung: Is Lung Health Good Health for Older Adults? *Chest* **155**, 391–400 (2019).
5. Schneider, J. L. et al. The aging lung: Physiology, disease, and immunity. *Cell* **184**, 1990–2019 (2021).
6. Budinger, G. R. S. et al. The intersection of aging biology and the pathobiology of lung diseases: A joint NHLBI/NIA workshop. *J. Gerontol. A Biol. Sci. Med. Sci.* **72**, 1492–1500 (2017).
7. Zhang, X. et al. Viral and host factors related to the clinical outcome of COVID-19. *Nature* **583**, 437–440 (2020).
8. de Cabo, R. & Mattson, M. P. Effects of intermittent fasting on health, aging, and disease. *N. Engl. J. Med.* **381**, 2541–2551 (2019).
9. Cho, S. J. & Stout-Delgado, H. W. Aging and lung disease. *Annu Rev. Physiol.* **82**, 433–459 (2020).
10. Venosa, A. Senescence in pulmonary fibrosis: between aging and exposure. *Front. Med.* **7**, 606462 (2020).
11. Gulati, S. & Thannickal, V. J. The aging lung and idiopathic pulmonary fibrosis. *Am. J. Med. Sci.* **357**, 384–389 (2019).
12. Montoro, D. T. et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018).
13. Travaglini, K. J. et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).
14. Reyfman, P. A. et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536 (2019).
15. Adams, T. S. et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* **6**, (2020).
16. Habermann, A. C. et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.* **6**, eaba1972 (2020).
17. Vieira Braga, F. A. et al. A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* **25**, 1153–1163 (2019).
18. Kadur Lakshminarasimha Murthy, P. et al. Human distal lung maps and lineage hierarchies reveal a bipotent progenitor. *Nature* **604**, 111–119 (2022).
19. Shah, V. S. et al. Single cell profiling of human airway identifies tuft-ionocyte progenitor cells displaying cytokine-dependent differentiation bias in vitro. *Nat. Commun.* **16**, 5180 (2025).
20. McQuattie-Pimentel, A. C. et al. The lung microenvironment shapes a dysfunctional response of alveolar macrophages in aging. *J. Clin. Investig.* **131**, (2021).
21. Angelidis, I. et al. An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nat. Commun.* **10**, 963 (2019).
22. Mogilenko, D. A. et al. Comprehensive profiling of an aging immune system reveals clonal GZMK CD8 T cells as conserved hallmark of inflammaging. *Immunity* **54**, 99–115.e12 (2021).
23. De Man, R. et al. Single-cell atlas of human lung aging identifies cell type dyssynchrony and increased transcriptional entropy. *Nat. Commun.* **17**, (2026).
24. Liu, X. et al. Transcriptomics analysis identifies the decline in the alveolar type II stem cell niche in aged human lungs. *Am. J. Respir. Cell Mol. Biol.* https://doi.org/10.1165/rcmb.2023-0363OC (2024).
25. Sikkema, L. et al. An integrated cell atlas of the lung in health and disease. *Nat. Med.* **29**, 1563–1577 (2023).
26. Chow, R. D., Majety, M. & Chen, S. The aging transcriptome and cellular landscape of the human lung in relation to SARS-CoV-2. *Nat. Commun.* **12**, 4 (2021).
27. GTEx Consortium Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
28. Terekhova, M. et al. Single-cell atlas of healthy human blood unveils age-related loss of NKG2CGZMBCD8 memory T cells and accumulation of type 2 memory T cells. *Immunity* **56**, 2836–2854.e9 (2023).
29. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
30. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289–300 (2018).
31. Shen, X. et al. Nonlinear dynamics of multi-omics profiles during human aging. *Nat. Aging* **1**, 16 (2024).
32. Ma, L.-Z. et al. Plasma proteomics identify novel biomarkers and dynamic patterns of biological aging. *J. Adv. Res.* https://doi.org/10.1016/j.jare.2025.05.004 (2025).
33. Lehallier, B. et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
34. Lee, S. et al. Molecular programs of fibrotic change in aging human lung. *Nat. Commun.* **12**, 6309 (2021).
35. Wong, C. K. et al. Aging impairs alveolar macrophage phagocytosis and increases influenza-induced mortality in mice. *J. Immunol.* **199**, 1060–1068 (2017).
36. Nelder, J. A. & Wedderburn, R. W. M. Generalized linear models. *J. R. Stat. Soc.* https://doi.org/10.2307/2344614 (1972).

37. Meyer, K. C. & Soergel, P. Variation of bronchoalveolar lymphocyte phenotypes with age in the physiologically normal human lung. *Thorax* **54**, 697–700 (1999).
38. Saul, D. et al. A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. *Nat. Commun*. **13**, (2022).
39. Coppé, J.-P., Desprez, P.-Y., Krtolica, A. & Campisi, J. The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu. Rev. Pathol.* **5**, 99 (2010).
40. Cherry, C. et al. Transfer learning in a biomaterial fibrosis model identifies in vivo senescence heterogeneity and contributions to vascularization and matrix production across species and diverse pathologies. *GeroScience* **45**, 2559–2587 (2023).
41. De Cecco, M. et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73–78 (2019).
42. Zhao, M., Chen, L. & Qu, H. CSGene: a literature-based database for cell senescence genes and its application to identify critical cell aging pathways and associated diseases. *Cell Death Dis.* **7**, e2053 (2016).
43. Gorgoulis, V. et al. Cellular senescence: defining a path forward. *Cell* **179**, 813–827 (2019).
44. Casella, G. et al. Transcriptome signature of cellular senescence. *Nucleic Acids Res.* **47**, 7294–7305 (2019).
45. Freund, A., Orjalo, A. V., Desprez, P.-Y. & Campisi, J. Inflammatory networks during cellular senescence: causes and consequences. *Trends Mol. Med.* **16**, 238–246 (2010).
46. Tacutu, R. et al. Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res.* **46**, D1083–D1090 (2018).
47. de Magalhães, J. P., Curado, J. & Church, G. M. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* **25**, 875–881 (2009).
48. Qu, Y. et al. Single-cell and spatial detection of senescent cells using DeepScence. *Cell Genom.* **5**, 101035 (2025).
49. Suryadevara, V. et al. SenNet recommendations for detecting senescent cells in different tissues. *Nat. Rev. Mol. Cell Biol.* **25**, 1001–1023 (2024).
50. DePianto, D. J. et al. Molecular mapping of interstitial lung disease reveals a phenotypically distinct senescent basal epithelial cell population. *JCI Insight* **6**, (2021).
51. Han, S., Scott Budinger, G. R. & Gottardi, C. J. Alveolar epithelial regeneration in the aging lung. *J. Clin. Investig.* **133**, (2023).
52. Nabhan, A. N., Brownfield, D. G., Harbury, P. B., Krasnow, M. A. & Desai, T. J. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* **359**, 1118–1123 (2018).
53. Kotliar, D. et al. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. https://doi.org/10.7554/eLife.43803 (2019).
54. Ishikawa, H. & Barber, G. N. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature* **455**, 674–678 (2008).
55. Kobayashi, Y. et al. Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat. Cell Biol.* **22**, (2020).
56. Delorey, T. M. et al. COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* **595**, (2021).
57. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
58. Kaiser, A. M. et al. p53 governs an AT1 differentiation programme in lung cancer suppression. *Nature* **619**, 851–859 (2023).
59. Zhao, W. et al. A cellular and spatial atlas of TP53-associated tissue remodeling defines a multicellular tumor ecosystem in lung adenocarcinoma. *Nat. Cancer* 1–23, (2025).
60. Yao, C. et al. Senescence of alveolar type 2 cells drives progressive pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **203**, 707–717 (2021).
61. Xie, T. et al. Abnormal respiratory progenitors in fibrotic lung injury. *Stem Cell Res. Ther.* **13**, 1–13 (2022).
62. Choi, J. et al. Inflammatory signals induce AT2 cell-derived damage-associated transient progenitors that mediate alveolar regeneration. *Cell Stem Cell* **27**, 366–382.e7 (2020).
63. Strunz, M. et al. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat. Commun.* **11**, 3559 (2020).
64. Zacharias, W. J. et al. Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature* **555**, 251–255 (2018).
65. Abdelwahab, E. M. M. et al. Wnt signaling regulates trans-differentiation of stem cell like type 2 alveolar epithelial cells to type 1 epithelial cells. *Respir. Res.* **20**, 204 (2019).
66. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, (2020).
67. Rajagopal, J. et al. Wnt7b stimulates embryonic lung growth by coordinately increasing the replication of epithelium and mesenchyme. *Development* **135**, 1625–1634 (2008).
68. Shu, W., Jiang, Y. Q., Lu, M. M. & Morrisey, E. E. Wnt7b regulates mesenchymal proliferation and vascular development in the lung. *Development* **129**, 4831–4842 (2002).
69. Wang, S., Dang, H., Xu, F., Deng, J. & Zheng, X. The Wnt7b/β-catenin signaling pathway is involved in the protective action of calcitonin gene-related peptide on hyperoxia-induced lung injury in premature rats. *Cell Mol. Biol. Lett.* **23**, 4 (2018).
70. Aros, C. J., Pantoja, C. J. & Gomperts, B. N. Wnt signaling in lung development, regeneration, and disease progression. *Commun. Biol.* **4**, 601 (2021).
71. Linehan, E. & Fitzgerald, D. C. Ageing and the immune system: focus on macrophages. *Eur. J. Microbiol. Immunol.* **5**, 14–24 (2015).
72. Park, M. D. et al. Hematopoietic aging promotes cancer by fueling IL-1α-driven emergency myelopoiesis. *Science* **386**, eadn0327 (2024).
73. Daum, B., Walter, A., Horst, A., Osiewacz, H. D. & Kühlbrandt, W. Age-dependent dissociation of ATP synthase dimers and loss of inner-membrane cristae in mitochondria. *Proc. Natl. Acad. Sci. USA* **110**, 15301–15306 (2013).
74. Darnell, J. E. Jr, Kerr, I. M. & Stark, G. R. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* **264**, 1415–1421 (1994).
75. Platanias, L. C. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat. Rev. Immunol.* **5**, 375–386 (2005).
76. Zhou, F. Molecular Mechanisms of IFN-γ to Up-Regulate MHC Class I Antigen Processing and Presentation. *Int. Rev. Immunol.* https://doi.org/10.1080/08830180902978120 (2009).
77. Siebeler, R., de Winther, M. P. J. & Hoeksema, M. A. The regulatory landscape of macrophage interferon signaling in inflammation. *J. Allergy Clin. Immunol.* **152**, 326–337 (2023).
78. Luster, A. D. & Ravetch, J. V. Biochemical characterization of a gamma interferon-inducible cytokine (IP-10). *J. Exp. Med.* **166**, 1084–1097 (1987).
79. Dufour, J. H. et al. IFN-gamma-inducible protein 10 (IP-10; CXCL10)-deficient mice reveal a role for IP-10 in effector T cell generation and trafficking. *J. Immunol.* **168**, 3195–3204 (2002).
80. Loetscher, M. et al. Chemokine receptor specific for IP10 and mig: structure, function, and expression in activated T-lymphocytes. *J. Exp. Med.* **184**, 963–969 (1996).

81. Griffith, A. V., Fallahi, M., Venables, T. & Petrie, H. T. Persistent degenerative changes in thymic organ function revealed by an inducible model of organ regrowth. *Aging Cell* **11**, 169–177 (2012).
82. Messaoudi, I., Lemaoult, J., Guevara-Patino, J. A., Metzner, B. M. & Nikolich-Zugich, J. Age-related CD8 T cell clonal expansions constrict CD8 T cell repertoire and have the potential to impair immune defense. *J. Exp. Med.* **200**, 1347–1358 (2004).
83. Acosta, E. et al. Determinants of influenza mortality trends: Age-period-cohort analysis of influenza mortality in the United States, 1959-2016. *Demography* **56**, 1723–1746 (2019).
84. Santoro, M. G. Heat shock factors and the control of the stress response. *Biochem. Pharmacol.* **59**, 55–63 (2000).
85. Samuel, C. E. Antiviral actions of interferons. *Clin. Microbiol. Rev.* **14**, 778–809, table of contents (2001).
86. Bretscher, P. A. A two-step, two-signal model for the primary activation of precursor helper T cells. *Proc. Natl. Acad. Sci. USA.* **96**, 185–190 (1999).
87. Goldstein, J. S., Chen, T., Gubina, E., Pastor, R. W. & Kozlowski, S. ICAM-1 enhances MHC-peptide activation of CD8(+) T cells without an organized immunological synapse. *Eur. J. Immunol.* **30**, 3266–3270 (2000).
88. Real, E. et al. Immature dendritic cells (DCs) use chemokines and intercellular adhesion molecule (ICAM)-1, but not DC-specific ICAM-3-grabbing nonintegrin, to stimulate CD4+ T cells in the absence of exogenous antigen. *J. Immunol.* **173**, 50–60 (2004).
89. D'Acquisto, F. et al. Annexin-1 modulates T-cell activation and differentiation. *Blood* **109**, 1095–1102 (2007).
90. Vannan, A. et al. Spatial transcriptomics identifies molecular niche dysregulation associated with distal lung remodeling in pulmonary fibrosis. *Nat. Genet.* **57**, 647–658 (2025).
91. Quach, H. et al. Early human fetal lung atlas reveals the temporal dynamics of epithelial cell plasticity. *Nat. Commun.* **15**, 1–24 (2024).
92. Domínguez Conde, C. et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022).
93. Szabo, P. A. et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 1–16 (2019).
94. Thome, J. J. C. et al. Spatial map of human T cell compartmentalization and maintenance over decades of life. *Cell* **159**, 814–828 (2014).
95. Campanella, G. S. V., Medoff, B. D., Manice, L. A., Colvin, R. A. & Luster, A. D. Development of a novel chemokine-mediated in vivo T cell recruitment assay. *J. Immunol. Methods* **331**, 127–139 (2008).
96. Brownlie, D. et al. Comparison of lung-homing receptor expression and activation profiles on NK Cell and T cell subsets in COVID-19 and Influenza. *Front Immunol.* **13**, 834862 (2022).
97. Blank, C. U. et al. Defining 'T cell exhaustion. *Nat. Rev. Immunol.* **19**, 665–674 (2019).
98. McKinney, E. F., Lee, J. C., Jayne, D. R. W., Lyons, P. A. & Smith, K. G. C. T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. *Nature* **523**, 612–616 (2015).
99. Sun, X. et al. Longitudinal analysis reveals age-related changes in the T cell receptor repertoire of human T cell subsets. *J. Clin. Investig.* **132**, (2022).
100. Cain, A. et al. Multicellular communities are perturbed in the aging human brain and Alzheimer's disease. *Nat. Neurosci.* **26**, 1267–1280 (2023).
101. Martínez-Zamudio, R. I. et al. AP-1 imprints a reversible transcriptional programme of senescent cells. *Nat. Cell Biol.* **22**, 842–855 (2020).
102. Comfort, A. Test-battery to measure ageing-rate in man. *Lancet* **294**, 1411–1415 (1969).
103. Ingram, D. K. Toward the behavioral assessment of biological aging in the laboratory mouse: concepts, terminology, and objectives. *Exp. Aging Res.* **9**, 225–238 (1983).
104. Yu, D. et al. CellBiAge: Improved single-cell age classification using data binarization. *Cell Rep.* **42**, (2023).
105. Ren, X. & Kuan, P. F. RNAAgeCalc: a multi-tissue transcriptional age calculator. *PLoS One* **15**, e0237006 (2020).
106. Jia, M. et al. Transcriptional changes of the aging lung. *Aging Cell* **22**, e13969 (2023).
107. Guo, M. et al. Guided construction of single cell reference for human and mouse lung. *Nat. Commun.* **14**, 1–20 (2023).
108. Natri, H. M. et al. Cell-type-specific and disease-associated expression quantitative trait loci in the human lung. *Nat. Genet.* **56**, 595–604 (2024).
109. Van den Bossche, J. et al. Mitochondrial dysfunction prevents repolarization of inflammatory macrophages. *Cell Rep.* **17**, 684–696 (2016).
110. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332 (2011).
111. Härtlova, A. et al. DNA damage primes the type I interferon system via the cytosolic DNA sensor STING to promote anti-microbial innate immunity. *Immunity* **42**, 332–343 (2015).
112. Sauler, M., Bucala, R. & Lee, P. J. Role of macrophage migration inhibitory factor in age-related lung disease. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **309**, L1–L10 (2015).
113. Berg, M. et al. FastCAR: fast correction for ambient RNA to facilitate differential gene expression analysis in single-cell RNA-sequencing datasets. *BMC Genomics* **24**, 722 (2023).
114. Giotti, B. et al. Single-Cell View of Tumor Microenvironment Gradients in Pleural Mesothelioma. *Cancer Discov.* **14**, 2262–2278 (2024).
115. Soni, N. et al. Single-cell dissection of the genotype-immunophenotype relationship in glioblastoma. *Brain* https://doi.org/10.1093/brain/awaf129 (2025).
116. Bairakdar, M. D. et al. Learning the cellular origins across cancers using single-cell chromatin landscapes. *Nat. Commun.* **16**, 1–20 (2025).
117. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 2623–2631 (Association for Computing Machinery, 2019).

Acknowledgements

We thank all members of the Tsankov lab as well as members of the SenNet consortium and publication working group for valuable insights and feedback. This research was also conducted with support from the Biorepository and Pathology Core at Mount Sinai. Figures 1a, 3c, 7c, and 8, were created using Biorender.com (Tsankov, A. (2026) https://BioRender.com/upv29gj).

Author contributions

A.M.T. conceived and designed all aspects of the study with help from K.X. K.X., G.S.K. and A.B. carried out the analysis on lung scRNA-seq, PBMC, and spatial data, respectively. S.E.M. and V.V.V. performed scRNA-seq analysis for the revised manuscript. G.P. and J. Z. performed the mitochondrial stress test experiment, designed by D.P. J.A-G. procured lung autopsy samples annotated regions for spatial analysis, with input from R.B., K.D., and A.M.T. M.B., G.S.K. and K.X. performed lung age model training and predictions. K.D., I.S., and J.K. performed the immunofluorescence experiments, with supervision from A.M.T and J.K. S.K., A.K., and P.J.L. provided senescence/aging expertise, resources, and performed the spatial profiling. Y.C., B.G., and T.T.N. contributed with code and feedback. K.X., G.S.K., A.B., and A.M.T wrote the manuscript with input from all authors.

Funding
This research was funded by the NIH Common Fund, through the Office of Strategic Coordination/Office of the NIH Director under awards U54AG075936-01 to P.J.L. and by NIH grants R01AG089078-01A1 and R21AG095627-01 to A.M.T. This work was supported in part through the Minerva computational and data resources and staff expertise provided by Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai and supported by the Clinical and Translational Science Awards (CTSA) grant UL1TR004419 from the National Center for Advancing Translational Sciences. Research reported in this publication was also supported by the Office of Research Infrastructure of the National Institutes of Health under award number S10OD026880 and S10OD030463. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-75427-5.

Correspondence and requests for materials should be addressed to Alexander M. Tsankov.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.