SenePy: Unveiling the Cell-Type Specific Landscape of Cellular Senescence through Single-Cell Analysis in Living Organisms

Mark A. Sanborn[1,2], Xinge Wang[1,2,3], Shang Gao[1,2,3], Yang Dai[2,3], Jalees Rehman[1,2,3,4]

Affiliations:

[1] Department of Biochemistry and Molecular Genetics, University of Illinois College of Medicine, Chicago, Illinois, USA

[2] Center for Bioinformatics and Quantitative Biology, University of Illinois at Chicago, Chicago, Illinois, USA

[3] Department of Biomedical Engineering, University of Illinois College of Medicine, Chicago, Illinois, USA

[4] University of Illinois Cancer Center, Chicago, Illinois, USA

Please address correspondence to:

Jalees Rehman, MD (jalees@uic.edu)

Department of Biochemistry and Molecular Genetics

The University of Illinois, College of Medicine,

900 S. Ashland Ave (M/C 669), Chicago, IL, 60607

Phone: (312) 996-7670, Fax: (312) 413-0353

ABSTRACT

Senescent cells accumulate in tissues with organismal age and contribute causally to multiple chronic diseases. *In vivo* senescent cell phenotypes are heterogeneous because cellular context and stressors vary by cell type and tissue. Due to the variability of senescence programs, there is no universal method to identify senescent cells and even widely used markers, such as *CDKN2A*, are not ubiquitous. Therefore, we interrogated the *Tabula Muris Senis* mouse single-cell aging atlas and an array of single-cell datasets from human donors that spanned many ages to find cell-specific signatures of cellular senescence. We derived 75 mouse and 65 human senescence signatures from individual cell populations. *CDKN2A* and other markers of senescence were overrepresented in these signatures but there were many novel senescence genes present at higher rates. Within individual cell populations, we observed multiple programs of senescence with distinct temporal and transcriptional characteristics. We packaged the signatures along with a single-cell scoring method into an open-source package: *SenePy*. *SenePy* signatures better recapitulate cellular senescence than available methods when tested on multiple *in vivo* RNA-seq datasets and a $p16^{ink4a}$ reporter single-cell dataset. We used *SenePy* to map the kinetics of senescent cell accumulation across 97 cell types from humans and mice. *SenePy* also generalizes to disease-associate senescence and we used it to identify an increased burden of senescent cells in COVID-19 and myocardial infarction. This work provides a significant advancement towards our ability to identify and characterize *in vivo* cellular senescence.

INTRODUCTION

Aging is a key risk factor for many chronic diseases[1]. One biological manifestation of organismal aging is cellular senescence (CS), a phenomenon characterized by permanent cell cycle arrest, impaired homeostatic cellular function, and activation of the senescence-associated secretory phenotype (SASP) which involves the release of pro-inflammatory proteins, proteases, and other bioactive paracrine factors[2]. Senescent cells accumulate in tissues with increasing organismal age, but senescent cells are found even in young organisms and can accrue prematurely due to exogenous stressors[3, 4]. Accumulated senescent cells contribute to sterile inflammation, tissue remodeling, and local dysfunction which ultimately drives various pathologies[5]. CS contributes to a wide array of chronic diseases, including cardiovascular disease, neurodegeneration and diabetes[6-8]. The senescent cell burden in aged organisms also contributes to runaway inflammation and poor outcomes in acute diseases, such as those from coronavirus infection[9]. Targeted clearance of senescent cells with senolytics can mitigate disease severity and increase healthspan[7-10], but in some contexts their elimination may exacerbate disease[11]. Despite the growing understanding of the role of cellular senescence in aging and various diseases, *in vivo* cellular senescence remains poorly phenotypically and mechanistically characterized[2, 12]. The majority of CS markers have been identified in cultured cells subjected to experimental conditions that may not accurately represent a living system. More comprehensive sets of markers are required to robustly study cellular senescence in living systems.

One of the biggest challenges in studying *in vivo* cellular senescence is the high degree of heterogeneity in which it presents[2]. Senescence has been observed in numerous cell types across all major organs. Senescent cells partially lose their pre-senescence identities and phenotypes but this suggests that the mechanistic paths to the senescent states varies between cell types[13, 14]. Cells are also subject to various cell-intrinsic and extrinsic stressors which drive senescence. Telomere attrition is a well-known senescence trigger, but telomere-independent DNA damage, oxidative stress, and oncogenic signaling can also induce cellular senescence[5]. Paracrine factors released in the SASP state and cell-surface signaling from senescent cells can induce secondary senescence among otherwise healthy cells

in close proximity[15, 16]. Drivers of senescence shift the transcriptional landscape of senescent cells[13, 14], but this has been primarily studied in cultured cells. The tissue context or cell identity-specific transcriptional landscapes of cellular senescence have not been fully defined. For example, it is not clear whether the senescence transcriptional programs in skin fibroblasts exposed to UV light differ from the senescence programs of fibroblasts in internal organs protected from light. The heterogeneity arising from different stressors, tissues, and cell types makes it difficult to broadly apply transcriptional signatures of senescence derived from *in vitro* cultured cells which have been removed from their *in situ* tissue environment. There is no universal signature or marker of CS exists[2]. Even the cell cycle arrest inducer $p16^{ink4a}$, which is widely accepted as one of the most specific markers of cellular senescence, is not always required for senescence induction and its use as a sole marker in transcriptomics data is confounded by the fact that the corresponding *CDKN2A* locus encodes for multiple genes with overlapping sequence identity[17-19]. Other markers of senescence may be constitutively expressed in some cell types or upregulated in general with organismal age or inflammation. Recent work has utilized literature screening and transcriptomics to find a gene set that is broadly differentially abundant in senescent cells[20], but this does not account for tissue-, cell-, or stress-specific heterogeneity and may not capture all programs of cellular senescence. There remains a need to identify and characterize tissue- and cell-specific cellular senescence programs[21].

In this study, we aggregate and interrogate large-scale single-cell RNA-sequencing datasets across tissues and ages, both in mice and humans to define *in vivo* cellular senescence heterogeneity. We developed a novel algorithmic approach to identify cell-type-specific senescence signatures. We used a $p16^{ink4a}$ reporter mouse model dataset along with other transcriptomics datasets to validate our approach. We have generated the open-source Python package *SenePy* (https://github.com/jaleesr/senepy). *SenePy* allowed us to map the kinetics of cellular senescence in many tissues and cell types with respect to organismal age and in the context of disease. Using *SenePy* we were able to identify senescent cells across several tissues and examine similarities as well as tissue-specific and cell-type-specific signatures of cellular senescence.

RESULTS

Known senescence markers are cell-type-specific and poorly characterize *in vivo* cellular senescence. We examined the expression of established CS markers in comprehensive mouse and human single-cell atlases to determine their dynamics with age and their cell-type-specificity. For mouse data, we utilized the *Tabula Muris Senis*[22] resource which is comprised of 328K cells from 19 tissues collected between 1 and 30 months of age (**Sup Fig 1A**). The human data utilized in this study are from 7 studies[23-29] that span 37 tissues from individuals ranging in age from 1 to 92 years old, altogether comprising 1.6M cells (**Sup Fig 1B**). We took the union of *SenMayo*[20] (n = 125), which is a recently published list of senescence-associated genes, and a novel curated senescence gene set (n = 110) to establish a panel of 181 experimentally validated CS maker genes (∩= 52, hypergeometric $p = 6.4x10^{-94}$). This panel of known or validated senescence markers served as a starting point for our downstream analyses.

To assess how this set of senescence markers overlapped with cell-specific and universal organismal aging genes in mice, we compared them to a study that identified 76 cell-specific signatures and a universal aging signature (n = 330 genes) from *Tabula Muris Senis*[30] (**Fig 1A**). Only the CS markers *Cd9, Ctnnb1, and Jun* were present in the universal organismal aging signature (defined by genes upregulated with age in half of the cell types), which can be explained by random chance (hypergeometric p = 0.58). *CDKN2A* ($p16^{ink4a}$ encoding gene), which is widely accepted as one of the most universal and specific markers of CS, was not present in any of the cell-specific organismal aging signatures. Furthermore, only 15 of the 76 cell-specific aging signatures were enriched for CS markers (**Sup Fig2A,B**). This observed cell-type-specific CS enrichment is not explainable by known CS dynamics and was negatively correlated to population-specific proliferation, as determined by the proportion of *Mki67*+ cells (Pearson's R = -0.24, p = 0.04) (**Sup Fig 2C**). These findings suggest that a universal CS signature may be obfuscated by cellular heterogeneity and that differential expression between subpopulations of cells is not suitable to extract cell-type specific CS markers. Instead, we show that the proportions of cells

positive for *Cdkn2a* and other CS markers increase significantly with age in *Tabula Muris Senis* (**Fig 1B**), thus indicating that changes in cell positivity is a more useful metric for identifying dynamic CS genes.

We next analyzed the aging dynamics of CS markers across all tissues and cell types in the mouse and human datasets with respect to the proportion of positive cells. Some of the most widely used markers of CS, such as *Cdkn2a* and *Cxcl13*, showed an overall increase in the proportion of cells expressing these genes with age but had tropism for specific tissue and cell types (**Fig 1C,D**, **Sup Fig 3A**). Human skin from the face, for example, had appreciable levels of *CDKN2A*+ cells in young individuals and a large increase with age (**Sup Fig 3B**). However, other important markers of CS, such as *CDKN1A* ($p21^{cip1}$ encoding gene), were more constitutively expressed in both young and old mice and humans (**Sup Fig 3C**).

We examined the dynamics of cellular senescence in 60 mouse and 50 human cell types. Overall, the landscape of all CS markers was highly heterogeneous when stratified by tissue and cell type in both species (**Fig 1E,F**). Only 16 of 1,770 pairwise combinations of cell types showed significant senescence marker gene overlap (Hypergeometric, FDR $p < 0.05$), thus highlighting how senescence program transcription profiles differ widely between tissues and cell types. Cells from the same tissue were most likely to have significant marker overlap (Chi-square, $p = 3.2 \times 10^{-10}$). While fibroblasts from different tissues showed some overlap in senescent cell transcriptional profiles, this was not statistically enriched (Chi-square, $p = 0.2$). The marker found to increase dynamically with age in the largest number of human cell types was *CCL4*, but this marker was only dynamic in 21 of 50 (42%) cell populations. *Ccl5* and *Ccl8* were the most universally dynamic in mice, but only in 33% of cell types. *CDKN2A* was one of the most enriched markers of senescence in both humans and mice, but it was only dynamic in 26% of human and 32% of mouse cell types. These data indicate that there is no universal senescence marker gene set for all tissues and cell types, and that each cell type within each tissue takes distinct transcriptional paths to the senescence state. Instead, our data maps the suitability of known senescence markers in different organisms, tissues, and cells.

147 ***De novo* cell-type specific signatures derived algorithmically from single-cell transcriptomes.** We
148 used an unbiased computational method to identify putative senescence signatures in distinct mouse
149 and human cell populations (**Fig 2A**). These signatures were derived from 47 mouse and 50 human cell
150 types that spanned wide organismal ages. We leveraged the prior biological knowledge that senescent
151 cells accumulate with increasing organismal age but remain a minority population and this allowed us to
152 identify genes that were dynamic with increasing organismal age within each cell population. Using cell-
153 by-gene positivity matrixes with the selected genes and cells, we constructed networks with hubs of
154 genes likely to be expressed in the same cells. Genes not associated with hubs were removed as they
155 likely represented noise or stochastic processes which increase with organismal age but are not directly
156 related to cellular senescence. Some cell types were comprised of multiple distinct signatures which
157 separated during network analysis (**Fig 2B**). In total, we derived 75 mouse and 65 human putative CS
158 cell-type signatures. The signatures contained both genes as well as values which indicate the degree of
159 correlation they have to other genes within the same signature.

160 Cell-type signatures were highly heterogenous but several known markers of CS were enriched
161 in selected cell types, although not in a consistent manner (**Fig 2C,D**). Importantly, similar gene
162 expression signatures (Hypergeometric FDR < 0.05) were more likely to be found between cells from the
163 same tissue (Chi-square, $p = 2.6 \times 10^{-9}$) than between cells of the same cell type (Chi-square, $p = 0.001$).
164 However, there were some exceptions in which signatures from cell types found in multiple tissues shared
165 high similarity. For example, senescent fibroblasts found in mouse lungs were most similar to senescent
166 fibroblasts from mouse tracheas and happened to also share the overall highest similarity between any
167 two mouse senescent cell type signatures. However, fibroblast signatures taken from all of the tissues
168 were not more likely to similar to each other (Chi-square, $p = 0.3$). Overall, the signatures did not form
169 distinct clades when clustered based on similarity, emphasizing heterogeneity and suggesting tissue-
170 specific influences on CS signatures may be more influential than universal drivers of CS (**Sup Fig 4**).

171 Many cell-type signatures contained overlapping genes despite the high degree of overall
172 signature heterogeneity (**Fig 2E**). There were 387 of 1081 (36%) mouse cell-type-signature pairs that

shared significant overlap compared to 16 out of 1770 (0.9%) when using established markers. Based on their genetic profile, cell-type signatures clustered distinctly with each other but not with organismal aging signatures. Importantly, the previously defined set of unspecific senescence markers was more closely associated with the newly derived signatures and fell within the same neighborhood cluster. These observations indicate that the signatures we derived share some underlying genetic characteristics despite being highly distinct and that these signatures likely represent bonified *in vivo* CS programs. Therefore, we developed the open-source *SenePy* Python software package to score single-cells based on their expression of these CS signature genes. *SenePy* rapidly processes thousands of cells and provides relative senescent scores for every cell within a given population.

Distinct modes and phenotypes of senescence exist within the same cell populations. Cell populations are exposed to multiple stressors and therefore may harbor multiple independent modes of senescence. We observed that multiple cell-type signatures consisted of correlated gene hubs that clustered during network analysis (**Fig 2B, 3A**). Of the 46 mouse cell populations with derived cell-type signatures, 24 contained multiple gene hubs. From 47 human cell populations, 14 signatures were comprised of multiple hubs. Multiple signature hubs likely represent modes of CS with distinct kinetics and gene expression patterns. For example, aging mouse tongue keratinocytes consisted of two hubs, both enriched in established CS marker genes (**Fig 3A**). Gene enrichment analysis indicated different phenotypes and functional roles between cells expressing these separate hubs (**Fig 3B,C**). One of the tongue keratinocyte hubs was more proinflammatory in nature and primarily enriched for immune cell chemotaxis, cytokine, and TNF signaling pathways, typical of the senescence-associated secretory phenotype. We have termed this “type-A” senescence. The other, “type-B”, was enriched for innate immune processes and pathways that typically respond to pathogenic stimuli, for example, the NOD-like receptor signaling pathway (Fisher’s Exact FDR p = 10-8). These observations suggest that these different modes of senescence drive different inflammatory pathways and may contribute differentially to sterile inflammation. Both type-A/B tongue keratinocyte gene hubs were most similar to gene signatures

from tongue basal cells, further emphasizing the importance of tissue context on the modes of senescence. However, they were similar to other hubs from multiple tissues and cell types (**Fig 3D**), either presenting as type-A or type-B senescence hubs but not both, suggesting this multimodal pattern is not restricted to tongue keratinocytes. Next, we used *SenePy* to score tongue keratinocytes based on these type-A and type-B signatures. The number of cells that were outliers ($> 3\sigma + \mu$) in their signature-specific distributions increased significantly as a function of organismal age (**Fig 3E,F**), indicating a higher number of senescent keratinocytes in old mouse tongues compared to young. The increase in senescent cells with age followed a similar pattern in both CS programs but the type-B senescent cells appeared before type-A senescent cells temporally and also reached higher percentages.

We next examined senescent fibroblasts from three different tissues to determine the CS characteristics of similar cell types in different contexts. The mouse fibroblasts from hearts, lungs, and tracheas each had two distinct CS signature hubs. Most fibroblast hub signatures shared little genetic similarity and high cosine distance (**Fig 3G**). The senescent cell gene hubs with the highest pairwise similarities were present in the cells of the lungs and trachea, possibly indicative of the spatial proximity and function in the respiratory system may have resulted in similar senescence phenotypes. Functionally, these similar hubs in lungs and trachea shared common biological processes, such as inflammatory response, cytokine signaling, and immune cell chemotaxis (**Fig 3H**). However, the trachea hub was uniquely and highly enriched for genes involved in B-cell signaling, neutrophil activation, and lymphocyte proliferation. When the fibroblast populations were scored with *SenePy,* they showed distinct temporal kinetics (**Fig 3I**). In all cell populations, there was a small proportion of senescent cells in young mice and a drastic increase in old mice. The biggest increase in the proportion of senescent cells occurred between the ages of 18 and 24 months. Interestingly, cells identified using the most similar trachea and lung hubs had comparable temporal kinetics and nearly identical high proportions of senescent cells in 24-month-old mice. Senescent fibroblast populations in the heart and lungs also followed parallel kinetics despite greater gene and ontological distance. Together, these results suggest multiple modes of senescence even within the same populations and that these modes are temporally and phenotypically distinct.

Cell-specific signatures are unique but share common stress response and inflammatory pathways. We explored the senescence markers we derived for gene and ontological commonality to determine whether there are universal senescence signatures. No gene was present in every signature or hub, but selected genes were statistically overrepresented (random permutations) in the signatures (**Fig 4A**). There were 29 genes with greater than 25% prevalence (n >= 12, p <= 0.0014) in the 47 mouse cell types tested. The most common gene, and the only one above 50% prevalence, was *Hba-a1,* which was present in 28 (60%, $p = 5 \times 10^{-5}$) of the 47 signatures. For comparison, *Cdkn2a* (gene that encodes p16) was also overrepresented in the data but only present in 12 signatures (25.5%, p = 0.0014). Among other known senescence markers, multiple Bcl2 family genes, which are common senolytic targets, were overrepresented in these data. Interestingly, along with *Hba-a1*, the hemoglobin subunits *Hba-a2*, *Hbb-b1*, and *Hbb-b2* were among these most representative genes, despite the absence of erythrocytes in the upstream analysis. Genes present in 8 or more signatures (n = 141, p < 0.01 by random permutation) were enriched for multiple biological processes involved in inflammation, immunity, cytokine signaling, and chemotaxis (**Fig 4B**). The NF-kappa B signaling pathway, a known driver of cell senescence, was among the most commonly enriched pathways (KEGG, BH-corrected p = 0.0004), emphasizing that NF-kappa B plays an important role in some programs of *in vivo* senescence. However, only 9 of the 47 signatures were individually enriched for NF-kappa B signaling, indicating that it is far from universal (**Fib 4C**). Therefore, to test for transcription factors that might be active, we tested our signatures for TF binding enrichment in their gene promoters (**Fig 4D**). The most universally enriched binding motif was RREB1, which was enriched in 38 of 47 signatures. Additionally, we found 46 other transcription factors enriched in over 50% of the mouse signatures.

In human cells, only two genes were present in >25% of the human senescence signatures (**Fig 4E**): Myosin light chain 9 (*MYL9*) and matrix metallopeptidase 9 (*MMP9*). *MMP9* is a known SASP component and is present in our curated set of senescence markers. In comparison, *CDKN2A* was only present in 8 of 51 signatures but still higher than what would be expected by random chance (p = 0.01).

There were an additional 222 genes present in 8 or more of the cell types. Genes present in more signatures than expected by chance ($p < 0.05$) were enriched for DNA replication and repair, chemokine signaling, cell cycle inhibition, NF-kappa B signaling, and other biological pathways (**Fig 4F**). When signatures were tested individually, the most commonly enriched pathways included neutrophil-mediated immunity, platelet degranulation, cytokine signaling pathways, and other inflammation and immune pathways (**Fig 4G**), consistent with SASP.

The most universal genes and pathways active in the cell-type signatures from both species shared some characteristics. There were 46 common genes that were enriched in both the mouse and human cell-type signatures. This was only a marginal overrepresentation compared to random chance (Hypergeometric, $p = 0.09$), suggesting low gene-wise concordance between species. *CDKN2A, CXCR2*, and *CCL3* were the only common genes that were previously established senescence markers. However, the pathway concordance between species was high and both sets of common genes were enriched for 32 common pathways, such as NF-kappa B signaling, AGE-RAGE signaling, and chemokine signaling. Likewise, eight of the top 20 most commonly enriched transcription factors from mouse signatures were also commonly enriched in human signatures (**Fig 4H**). This suggests that core senescence pathways between species are conserved but the individual genes that are enriched in senescent cells show a high degree of genetic variation. Together, these data indicate that our *de novo* cell-type signatures are enriched for known senescence phenotypes and share some commonality between cell types and species despite their high degree of heterogeneity.

The cell-specific kinetics of senescent cell accumulation with organismal age. We next used *SenePy* to determine the proportion of senescent cells in distinct populations from young and old organisms. The numbers and proportions of cells identified as senescent increased drastically with age in both mice (**Fig 5A**) and humans (**Fig 5B**) overall but showed distinct cellular tropism and kinetics. Within the tested mouse cells, kidney epithelial cells had the highest proportion of senescent cells at old age, followed by tracheal fibroblasts (**Fig 5C**). Out of the 47 tested cell populations in humans with

appreciable senescence signatures, type II pneumocytes had the highest percentage of senescent cells reaching a level of greater than 11% in patients aged 68-77. This was followed by several other populations of skin and lung cells (**Fig 5D**). However, in the human cells we observed temporal heterogeneity between patients. For example, cardiomyocytes from a patient in the 58-year age bin were 7.6% senescent while a patient in the 68-year bin had only 0.4% senescent cardiomyocytes. This disparate pattern of fewer senescent heart cells in the older patient persisted in every other heart cell type except adipocytes, suggesting patient-specific differences are important drivers for whole-tissue senescence. In younger individuals (ages 40-60), the heart cell type with the highest proportion of senescent cells was invariably lymphoid cells (**Fig 5E**). Conversely, there was a large increase in the proportion of senescent cells within solid tissue cell types taken from most 60-year+ donors. One 68-year-old patient in particular had a surprisingly high proportion of cardiomyocytes and fibroblasts with high senescence scores. But in the oldest heart tested, there was a relatively low proportion of senescent cells in all cell types except for lymphocytes. These data could be indicative of how circulating cells may be able to promote tissue senescence after transmigrating into tissues or have reduced efficacy in clearing senescent cells.

The proportions of cells within populations predicted to be senescent by *SenePy* were not correlated to the replicative potential in their respective populations (**Sup Fig 5A**,**B**). Surprisingly, replicative populations, such as large intestine enterocytes, had minimal increases in the proportion of senescent cells with age. This is corroborated by an undetectable change in the number of *Cdkn2a*+ enterocytes with age and the general lack of correlation between *Ckdn2a*+ cells in replicative populations. Likewise, we did not observe a negative correlation between the population-level expression of telomerase and the calculated senescence burden (**Sup Fig 5C**,**D**). Thus, indicating that the well-known driver of *in vitro* senescence, telomere attrition, is not the primary driver of senescence within organisms.

***SenePy* identifies ground-truth *in vivo* senescence more robustly than established markers of senescence.** To test *SenePy*'s ability to robustly detect senescent cells in additional datasets, we utilized

single-cell RNA-seq data from p16-Cre^{ERE2}-tdTomato reporter mice in which p16+ cells become red fluorescent[31]. We first performed differential expression analysis between the tdTomato+ and tdTomato- kidney cells (**Fig 6A**). The mRNAs that were differentially abundant in tdTomato+ cells were enriched for kidney-specific signatures derived in this study. Our kidney-specific signatures were more enriched than previously defined gene sets or a senescence gene panel derived from the AI model GPT-4 that was tasked to identify senescence signatures (n = 100 genes, "senGPT") (**Fig 6B**). A notable shift was observed in the distribution of cells scored with *SenePy* using kidney-specific signatures but not in a heart endothelial cell hub (**Fig 6C**). Likewise, the mRNAs more abundant in tdTomato+ liver cells were more enriched for *SenePy's* liver Kupffer cell signature than any other gene sets (**Fig 6D**). These data indicate that *SenePy*'s senescence signatures and derived senescence scores identify cells that become p16+.

We also examined the ability of *SenePy* signatures to identify transcriptional changes due to senolytic treatment as well as those seen in experimental conditions that induce senescence in multiple *in vivo* and *in vitro* models. The genes downregulated in mouse lungs following therapeutic senolysis were enriched for multiple *SenePy* lung- or airway-specific signatures (**Fig 6E**). *SenePy* signatures were more enriched than cell-type agnostic gene sets. *SenePy* does not contain a specific skeletal muscle senescence signature due to data availability, yet mRNA less abundant in mouse muscle tissue following senolytic treatment was enriched for multiple *SenePy* signatures, including one from myocytes (**Sup Fig 6A**). Multiple *SenePy* endothelial signatures were enriched in mRNA more abundant after radiation-induced senescence of human endothelial cells but the advantage of *SenePy* over other gene sets was diminished in this *in vitro* context (**Fig 6F, Sup Fig 6B**). The discrepancy between the *in vitro* and *in vivo* efficacy of *SenePy* was even more apparent in models of senescent fibroblasts in culture (F**ig 6G,H**). In these *in vitro* contexts, *SenePy* was outperformed. These data suggest that *SenePy* recapitulates *in vivo* cellular senescence and that gene sets derived primarily through previous *in vitro* experiments do not.

We next tested the marker suitability of genes that encode for p16^{ink4a} and p21^{cip1} in the p16-Cre^{ERE2}-tdTomato liver cells (**Fig 6I**). Only 8% of p16^{high} cells, as indicated by tdTomato, had detectible levels of *Cdkn2a* RNA (p16 encoding gene) (**Fig 6J**). This indicates that either *Cdkn2a* expression was

not detectible due to single-cell dropout or the expression of *Cdkn2a* is transitory in the majority of these senescent cells. Another important marker of cellular senescence, *Cdkn1a* (p21 encoding gene), was found in the majority of both tdTomato+ and tdTomato- cells, making its binary expression an inadequate metric because it's more universally expressed in non-senescent cells (**Fig 6K, L**). These data represent a striking example of why sole reliance on known cellular senescence genes like p16 and p21 is not sufficient, especially in single-cell transcriptomics because of low single-cell resolution, dropout, and marker-independent senescence programs.

***SenePy* predicts elevated senescence burden in severe disease.** We posited that *SenePy* could be applied to datasets to quantify the burden of senescence in disease. We first used *SenePy* to analyze a single-cell RNAseq lung dataset of 19 individuals who died from COVID-19 and 7 age-matched controls[32] (**Fig 7A**). There were senescent lung cells in both control and COVID-19 patients but there was an observable increase in the senescence of lung epithelial cells among COVID-19 patients (**Fig 7B**). COVID-19 mortality was significantly associated with an increased proportion of senescent lung epithelial cells and were present in AT1, AT2, and general airway epithelial cell populations (Mann-Whitney, $p = 0.004$) (**Fig 7C**). The proportions of senescence in non-epithelial cells, such as immune cells, fibroblasts, and endothelial cells, were not significantly associated with COVID-19 mortality (**Fig 7D**). Some of the deceased patients had relatively elevated senescence burdens in multiple cell types (**Fig 7E**).

Next, we used *SenePy* to score spatiotemporally resolved mouse transcriptomics data following myocardial infarction[33]. We used the hub signatures we previously derived from mouse hearts to score the spatially resolved spots (**Fig 7F**). Senescent loci were found even in the control heart, corroborating earlier observations that even young organisms have baseline levels of cellular senescence. However, since the spatially resolved spots consist of multiple single cells, the number of single senescent cells in these data are unknowable. The proportion of spots with high senescence burden was highest at day 7 but the change was not statistically significant, likely due to the small sample size (ANOVA, $n <= 3$). We observed a strong spatial association between spots with high senescence burden and heart fibrosis

after MI (**Fig 7G**). Senescent-like spots strongly colocalized in regions of the hearts expressing fibrotic markers such as *Col1a1*. This correlation becomes readily apparent by 7 days post-MI but was not observed in the control heart or hearts shortly after MI. The 8 hubs used to score the spots contributed to the overall senescence burden with distinct temporal patterns (**Fig 7H**). The endothelial cell hub had the highest contribution at day one but continued to decrease up to day 14. The score from a myocyte hub jumped from baseline on day 7 then dropped back down. Other senescence gene programs remained otherwise unperturbed by MI.

Additionally, we observed strong spatial autocorrelation between the spatially-resolved spots with high senescence burden (**Fig 7F, I**). Only one of 9 samples did not have highly significant spatial clustering of senescent-like spots. Unsurprisingly, this same day 14 sample has low senescent spot association to *Col1a1* as well as the lowest overall senescence burden. To further investigate this finding of senescent cell clustering and to see if this phenomenon is apparent in other tissues, we utilized spatial transcriptomics data from mouse brains before and after inflammatory insult (**Fig 7J**). The control brains had a small number of senescent-like spots which prevented significant autocorrelation; however, the brains from LPS-treated animals had amplified senescence signatures and highly significant clustering of senescent-like spots. These data indicate that senescent cells are more likely to be found in close proximity across multiple *in vivo* systems.

DISCUSSION

There is a paucity of tissue- and cell-specific markers for senescent cells due to the heterogeneity of cells that undergo cellular senescence (CS). This is especially challenging in single-cell transcriptomics because the high rate of dropout and limited sequencing depth of the technology poses a challenge for using classical senescence markers such as *Cdkn2a* and *Cdkn1a* as sole indicators of cell senescence. Furthermore, the cell-specific heterogeneity of cellular senescence represents a major challenge for the development of a universal senescence signature gene panel. Therefore, in this study, we took an unbiased large-data approach to identify cell-specific programs of cellular senescence and created

SenePy as an open-source platform (https://github.com/jaleesr/SenePy) to identify senescent cells in single-cell transcriptomic data. We validated the *SenePy* approach using single-cell RNA-seq data in p16 reporter mice and applied *SenePy* to determine the kinetics and heterogeneity of senescence across several human and mouse cell types in aging and disease.

Previous studies have generated transcriptomic signatures of cellular senescence based on the generation of senescence in controlled *in vitro* environments [13, 20]. While these signatures have helped advance the mechanistic understanding of cellular senescence, it is challenging to use such signatures for highly variable *in vivo* contexts. We show that cells express many of these genes at higher rates with age at the organism or tissue level, but none of the genes obtained from such senescence panels were applicable in the majority of cell types tested. Even the widely used marker *CDKN2A* was identified in senescent cells in less than a third of mouse and human cell types. The lack of universality of such senescence marker genes is likely multifactorial, arising from true biological variation and compounded by technical limitations. For example, many SASP genes are also genes that increase with inflammation and organismal aging. A transcript more abundant with age in all tested cells of a population would not be a suitable marker for a phenomenon that is present in only a minority of cells. Furthermore, our analysis of the p16-Cre^{ERE2}-tdTomato mouse cells highlights the limits of using a small number of markers such as p16^{ink4a} or *Cdkn2a* in transcriptomics data. While the tdTomato+ cells likely represented bona fide senescent cells[31], only a small proportion of individual cells had detectible levels of *Cdkn2a* at the time of tissue harvest and sequencing. This likely arose from the transient expression of p16 earlier on in the senescence program in combination with gene dropout inherent in single-cell sequencing. Nevertheless, many highly visible and impactful studies are forced to rely on a small unspecific set of markers because better alternatives did not exist. In addition to finding novel cell-specific signature genes, we provide a cell type suitability map for well-known senescence markers.

Our approach focused on markers specific to individual populations that would not be confounded by transcriptional differences with age. We were able to extract sub-population level programs of CS by setting kinetic thresholds based on the prior knowledge that senescent cells increase with age but are

present in the minority of cells and by examining single-cell co-expression as opposed to differential expression. Unsurprisingly, many of our computationally-derived signatures contained and were statistically enriched for pre-established CS markers. However, these comprised the minority and most genes have not been previously thought of as senescence markers. While we did find well-known senescence genes like *CDKN2A*, *BCL2*-family genes, and various SASP factors to be statistically common within our signatures, the most universal markers were novel ones. For example, in mice, the most common signature gene we identified was the alpha hemoglobin subunit. Hemoglobin has been previously reported as a response gene to oxidative stress in non-erythrocytes[34, 35], but has yet to be reported in CS. Interestingly, in human signatures there was no significant overrepresentation of globin genes. We observed this and other important differences between mouse and human signatures, suggesting organism-specific CS marker panels to be more specific. Additionally, we used our signatures to map the kinetics of senescent cell accumulation in many different tissues and cell types. Recent work has mapped the abundance of senescent marker mRNA in 13 different tissues as a function of mouse age and in a progeria model[36]. This work, however, was agnostic of cell type and relied on a small set of senescence markers. To our knowledge, this study is the first that comprehensively maps the increase in senescent cells in many different mouse and human tissues with respect to cell type.

The CS marker genes identified in this study may contain novel and cell-specific targets for senolytics. Senolytics have been effective in treating age-associated disease and at extending healthspan in model organisms[7-10]. Many senolytics inhibit anti-apoptotic factors such as BCL2-family proteins[37], and we indeed found numerous BCL2-family protein genes as markers in many of our cell-specific signatures. Therefore, this indicates that our methodology was able to extract actual senolytic targets. Our data will also inform which tissues and cell types BCL2-family inhibitors and other senolytics are likely more effective. Recent work has used a computational approach to identify a novel surface marker on senescent cells to direct targeted CAR-T senolysis[38], expanding the potential field of senolytic targets and emphasizing the importance of the novel markers found in our study. Furthermore, clearance of senescent cells is not beneficial in every context and cell-specific targeting of senescent cells may be

warranted in some scenarios. For example, liver regeneration is dependent on hepatic stellate cell senescence but is hindered by senescence of the liver parenchyma[39]. Our work uncovers potential cell-specific targets for specific senolytic therapy.

Senescent cells contribute to cardiovascular pathology[6, 40], but their role in disease has never been spatiotemporally characterized. We show that senescent cells localize at sites of heart fibrosis. The proportion of senescent cells in each heart throughout the time series did not significantly change and were present even in the control heart. We also observed highly significant spatial clustering of senescent foci in the hearts and brains of mice which may be supporting evidence for an *in vivo* bystander effect[15]. The spatial distributions of senescent cells have been previously examined from spatially resolved transcriptomics data in aged mouse brains [41]. Their results show that *Cdkn2a*+ spots are adjacent to activated microglia but no spatial clustering of *Cdkn2a*+ spots. Though, their methodology relies on a narrow definition of CS which may not translate to actual $p16^{ink4a}$ and does not account for Cdkn2a dropout or p16-independent forms of senescence. These and other data would greatly benefit from a reexamination with more comprehensive gene sets, such as those proposed herein.

Here we report that senescence burden in human lung endothelium is associated with COVID-19 mortality. Senolytic treatment prior to infection with betacoronavirus has been shown to reduce mortality in mice suggesting that pre-existing senescent cells increase disease risk[9]. In humans, deceased patients infected with SARS-CoV-2 had an increased number of mucosal cells expressing senescence markers and infection itself can induce senescence[4]. Their transcriptomics data also showed CS marker abundance to be highest in epithelial cells relative to other mucosal cell types. Taken together, these studies support our findings that senescent epithelial cells contribute to COVID-19 mortality.

By design, our methodology removes genes that are constitutively expressed at baseline or in aged cells to maintain a distinction from organismal aging. Inherently, this discounts genes that may be part of the senescence program which overlap with the transcriptional shift with age. We also do not account for genes that are downregulated in senescent cells. Negative markers of senescence would add extra information to better identify senescent cells, but to find negative correlations in all pairwise

459 combinations of genes with this methodology was computationally limiting. The comprehensiveness of
460 our signature panel is also limited by the data available at the time of study and the exact set of tissues
461 and cells tested influences any conclusions of universality or comparison between species. We do not
462 expect our signatures to negate the need for large scale future efforts such as SenNet[21]. We instead
463 expect our work to be complementary and assist these efforts.

464 This work comprehensively identified gene expression programs and signatures of senescent
465 cells that are stratified by species, tissue, and cell type and used them to broadly characterize senescent
466 cells in mice and humans. We created *SenePy*: a computational platform that assigns a senescence
467 score to individual cells in single-cell transcriptomic data, which can serve as a resource to uncover cell-
468 type and tissue-specific mechanisms of cellular senescence *in vivo*.

469
470
471

472 **METHODS**

473 **Data collection**

474 Single-cell RNA mouse data were collected from the *Tabula Muris Senis* atlas (ref). *Tabula Muris Senis*
475 consists of single-cells from 30 mice from 1 to 30 months of age taken from 19 tissues. Human single-
476 cell data were collected from 7 studies. The liver data were obtained from five donors ranging from 21 to
477 65 years old[29]. Single skin cells were obtained from 6 patients ranging in age from 18 to 48 years old[25].
478 Lung data were collected from 17 donors ranging from 21 to 72 years old[28]. Human heart cells were taken
479 from 14 patients ranging from 40 to 75 years old[26]. Human hippocampal cells were collected from 37
480 patients ranging from newborn to 92 years old[23]. These tissue-specific datasets were given priority for
481 downstream analysis in their respective tissues, but we used additional multi-tissue atlases. Cells from
482 the *Human Cell Landscape* came from 51 donors ranging from 21 to 66 years old and from 25 different
483 tissues[27]. Cells from *Tabula Sapiens* came from 15 patients ranging from 22 to 74 years of age[24].

484

485 **Data annotation**

The data were available in a range of formats from fastq to processed and annotated count data. Data from *Tabula Muris Senis*, *Tabula Sapiens,* and the human heart study were provided with cell type annotations. The Lung and hippocampal studies provided unannotated counts. Fastq data were processed through 10x CellRanger or the Dropseq protocol (https://mccarrolllab.org/dropseq/) depending on the technology used to prepare the libraries. All fastqs were aligned to GRCh38. Processed single-cell counts were handled with Scanpy[42]. Cells were filtered out if they had a relatively low or high number of detected genes or high relative proportion of mitochondrial reads (thresholds varied based on dataset distribution). We used a variety of methods to annotate cell types. Since the *Human Cell Landscape* contained many tissues and cell types, we transferred the annotations from *Tabula Sapiens* using scANVI[43] after processing the raw data with scVI-tools[44]. If a cluster, which was called by the Leiden algorithm on the scVI embeddings, had lower than 85% cell-type agreement, those cells were not used in downstream analysis. Cell types from the liver, skin, and lung studies were annotated similarly but clusters with poor label transfer were instead manually annotated using known cell-type markers[45]. For lack of a reference dataset, the hippocampal data were annotated exclusively using known markers. Annotations were harmonized across datasets (e.g., “kidney endothelial cell” changed to “endothelial cell”) and mapped back on to the raw counts. Cells lacking annotations because they failed QC or label transfer were discarded. For total dataset visualizations, the species-specific raw data were integrated using scVI and the embeddings were projected via UMAP.

Mouse cell-type specific age-dynamic genes

Mouse data came from mice aged 1, 3, 18, 21, 24, 30 months (m) but age availability varied by tissue. Cells were stratified by tissue, age, and cell type. The starting baseline was chosen as 3m if there were at least 200 3m cells, if not the starting baseline was aggregated with 1m cells. Likewise, 30m was prioritized for old cells if at least 200 were present, otherwise the old baseline fell back to 24m. The proportion of cells expressing one or more UMI copies of a gene was determined in each population (**Equation 1**). Zero values at 3m or 1m were imputed with the inverse of the cell count ($p_{age} =$

$n_{total\ cells}{}^{-1}$). Young cells were used as baselines and the ratios in old cells were determined relative to them (**Equation 2**). Genes were only considered dynamic if they were found in less than 5% of young cells ($p_{3m|1m} < 0.05$), found in less than 20% of old cells ($p_{24m|30m} < 0.2$), were in greater than 1% of old cells ($p_{24m|30m} > 0.01$), increased at least 0.5% ($p_{24m|30m} - p_{3m|1m} > 0.005$), and the old to young ratio was greater than 2.5 or the gain was greater than 5% ($r_{24m|30m} > 2.5 \mid p_{24m|30m} - p_{3m|1m} > 0.05$). These thresholds assume that the proportions of cells expressing senescence marker increase with age but remain in the minority. Dynamic genes are cell-specific markers and do not account for small changes to baseline levels of constitutively expressed genes, which may be senescence-associated genes but are not specific markers.

$$p_{age} = n_{gene^{+}cells}/n_{total\ cells}$$

(**Equation 1**) p_{age}: the proportion of cells positive for a gene at a given age. Where $p_{age} = n_{gene^{+}cells}$ is the number of cells positive for a given gene and $n_{total\ cells}$ is the number of total cells in the same population.

$$r_{30m|24m} = p_{30m|24m}/p_{3m|1m}$$

(**Equation 2**) $r_{30m|24m}$: ratio of old cells positive for a gene relative to young cells. Where $p_{30m|24m}$ represents the proportion of cells positive for a gene in cells from 30- or 24-month-old mice and $p_{3m|1m}$ are the proportion of cells positive for the gene from 3- or 1-month mice.

Human cell-type specific age-dynamic genes

Human ages were binned into 10-year bins to account for the continuous range of human ages. Bins ranged from 8 (8-17 years old) to 88 (88-97 years old). To be considered for further analysis a cell-type population must 1) have three unique bins with at least 100 cells in each bin or a bin <=28 and a bin >=58 with at least 100 cells and 2) have a bin >=48 with at least 100 cells. These criteria were required in

individual datasets because downstream analysis avoids confounding effect from multiple studies. Cells were stratified by dataset, tissue, age bin, and cell type. The proportions of cells expressing genes were calculated for each age bin (**Equation 1**). The young starting populations were selected from the 8, 18, or 28 10-year bins if one was present, else the starting proportion was calculated by regressing the age and known proportion values and solving for 18 years (**Equation 3**). The old ending populations were selected from the oldest age bin (p_{old}). Genes were only considered dynamic if their age by proportion slope was positive (**Equation 4**), their maximum positivity was less than 25% ($p_{max} < 0.25$), their starting proportion was less than 5% ($p_{young} < 0.05$), their gain was greater than 1.5% ($p_{old} - p_{young} > 0.015$), the maximum proportion was not two times higher than the ending proportion ($2p_{max} < p_{old}$), and the maximum proportion had to be in the 48 year bin or higher ($\max(p_{48+}) = p_{\max}$).

$$\mathrm{cov}(\mathrm{age}, p) = \frac{\sum(\mathrm{age}_i - \overline{\mathrm{age}})(p_i - \overline{p})}{n}$$

$$\mathrm{var}(\mathrm{age}) = \frac{(\mathrm{age}_i - \overline{\mathrm{age}})^2}{n}$$

$$p_{18} = \frac{18 \cdot \mathrm{cov}(\mathrm{age}, p)}{\mathrm{var}(age)} + \bar{p} - \frac{\overline{age} \cdot \mathrm{cov}(\mathrm{age}, p)}{\mathrm{var}(age)}$$

(**Equation 3**) p_{18}: extrapolated proportion of cells positive in the 18-year bin. Where $\mathrm{cov}(age, p)$ represents the covariance between age and proportion p, $\mathrm{var}(age)$ is the variance of age, and n is the sample size.

$$m = \frac{cov(age, p)}{var(age)}$$

(**Equation 4**) m: slope of the linear regression line for the proportions of a given gene with age. See equation 3.

Identifying novel senescence signatures from mice

Age-dynamic genes for each tissue-cell-type were found as described earlier. Count data were subset by these genes and further subset to only include cells from mice >21m. Subsets with fewer than 100 cells were not tested further. The count matrixes were binarized to represent cell by gene positivity. Every pairwise combination of genes was tested for Pearson's correlation (**Equation 5**). To test the statistical significance, each pairwise comparison was randomly permutated 500 times. Pairwise correlations were kept if they had a positive r value and if their r value was at least 0.05 higher than the respective q99 r value from the random permutations (**Equation 6**). The filtered correlations were used to construct networks with NetworkX (https://networkx.org/). The Louvain algorithm was used to group genes into clusters. Network clusters with fewer than 5 genes or genes with no correlations were removed. Genes loosely connected to clusters were removed if they had fewer than $\log(n_{cluster\ genes})$ (Where $n_{cluster\ genes}$ is the number of genes in a Louvain cluster) connections to other genes in the network. The cleaned clusters are hereby referred to as hubs and the aggregated hubs for each cell type as a cell-specific signatures.

$$r_{i,j} = \frac{\sum_{k=1}^{n}\left(x_{k,i} - \bar{x_i}\right)\left(x_{k,j} - \bar{x_j}\right)}{\sqrt{\sum_{k=1}^{n}\left(x_{k,i} - \bar{x_i}\right)^2}\sqrt{\sum_{k=1}^{n}\left(x_{k,j} - \bar{x_j}\right)^2}}$$

(**Equation 5**) $r_{i,j}$: Pearson's correlation coefficent for dynamic genes i and j. Where $x_{k,i}$ represents the binary expression value of gene i in cell k; $x_{k,j}$ is the binary expression value of gene j in cell k; and n is the total number of cells in the population.

$$t = q_{99}\left(r_{\text{perm}(i,j)}\right) + 0.05$$

$$r_{i,j} > t \; and \; r_{i,j} > 0$$

(**Equation 6**) t: significance threshold. Where $r_{\text{perm}(i,j)}$ represents the distribution of correlation coefficients for 500 random permutations of gene i and k. q_{99} represents the 99th percentile value of this distribution. The inequality depicts one criteria of gene selection based on t.

583

584 **Identifying novel senescence signatures from humans**

585 Age-dynamic genes for each dataset-tissue-cell-type were found as described earlier. Count data were
586 then subset by these genes and further subset to only include cells from patients 48 years of age or older.
587 Significant correlations, networks, hubs, and signatures were generated similarly to those from mice.

588

589 **Novel signature comparison**

590 Each signature or hub has a set of genes and corresponding weights for how many connections a gene
591 shares with other genes. Pairwise cosine similarity was calculated by comparing the union of each gene
592 list and imputing 0s (**Equation 7**).For pairwise hypergeometric similarity between two signatures, the
593 cumulative distribution function for two lists of genes was determined using the genes present in the
594 original species aggregated counts as the background list (**Equation 8**). For signature network analysis,
595 all pairwise hypergeometric sf values (i.e., p-values) were corrected with the Bonferroni method,
596 converted to $-\log_{10} sf_{corrected}$, and used as similarity scores between signatures if they were significant.

597 To find genes represented in the signatures more than expected by chance, we used a random
598 permutation method. A set of hubs with random genes identical in size to the original signatures were
599 generated 1000 times from the background set of expressed genes in the dataset. A distribution was
600 created representing the number of times a gene was found in each of the 1000 permutations. The actual
601 number of signatures a gene was found in was compared to this distribution to determine significance.

602

603 $$cosine(L_i, L_j) = \frac{\sum_{k=1}^{n} a_{i,k} a_{j,k}}{\sqrt{\sum_{k=1}^{n} a_{i,k}^2} \sqrt{\sum_{k=1}^{n} a_{j,k}^2}}, where\ a_{i,k} = \begin{cases} w_i k\ \ if\ g_k \in L_i \\ 0, otherwise \end{cases}$$

604 (**Equation 7**) $cosine(L_i, L_j)$: cosine distance between two signatures L_i and L_j. Where $a_{i,k}$ represent the
605 weight of gene k in gene list i, and is equal to the corresponding connection value $w_i k$ if gene g_k is in

gene list i, or 0 otherwise. Where n represents the total number of genes in the union of signatures L_i and L_j.

$$sf = 1 - P\left(\left|L_i \cap L_j\right| - 1, |N|, |L_i|, \left|L_j\right|\right) = 1 - \sum_{k=0}^{x-1} \frac{\binom{L_j}{k}\binom{N-L_j}{L_i-k}}{\binom{N}{L_i}}$$

(**Equation 8**) sf: survival function of the hypergeometric distribution $P(x, N, I, j)$. Where L_i and L_j represent two gene lists and N represents the background gene list. x is equal to the cardinality of the L_i and L_j intersection minus 1.

Gene set enrichment – GO, KEGG, transcription factor binding

We used the Enrichr python API gseapy[46] for gene set enrichment against the GO and KEGG databases (refs). The background set of genes used came from all expressed genes from their respective datasets. Only FDR-corrected p-values below 0.05 were considered significant. A custom "senescence" gene set was added which was comprised of the union between all literature-based senescence markers collected for this study and senMayo[20].

For transcription factor binding analysis, the regions 1000 bp upstream and 500 bp downstream of the transcription start sites were extracted for each gene in a gene list. JASPAR 2020 core vertebrate non-redundant position frequency matrices were us as the input motifs (ref). The extracted regions were examined for relative motif enrichment using the MEME-suite simple enrichment analysis.

Only Benjamini-Hochberg-corrected p-values below 0.05 were considered significant.

Scoring cells using SenePy

Gene signatures are comprised of genes and their respective number of edges in their network (termed importance value). We developed SenePy, a lightweight and fast scoring algorithm specific for our gene sets that borrows from Seurat's AddModuleScore() and Scanpy's tl.score_genes(). SenePY is built in python and integrates well with scanpy and anndata. SenePy has four core functions: load_hubs(),

translator(), score_hub(), and score_all_cells(). The load_hubs() function initializes the hub object which includes the hubs themselves along with additional metadata, such as each hub's enrichment for known senescence genes. Depending on the input data and its respective reference, the optional translator() function can be used to harmonize gene symbols based on known gene aliases. The score_hub() function takes one input hub and anndata and returns a list of scores for each cell. The score_all_cells() takes one input hub and anndata and stratifies the data based on input categories, for example, to score individual cell types separately to avoid confounding the score.

The scoring happens in multiple steps. First, the mean is calculated for each gene in the dataset across all cells (**Equation 9**). All genes are ranked by their mean and split into n_{bins} (default: 25) expression bins (**Equation 10**). Next, n_{ctrl_size} (default: 50) background genes are selected for each input signature gene from its corresponding expression bin (**Equation 11**). The counts data are then optionally binarized (default: True) to represent the binary senescence cellular state and the gene-cell positivity from which the underlying networks were derived . Next, the counts are optionally amplified (Default: True) by their corresponding importance value from the input signature (e.g., if [*Cdkn2a*, 2] is in the signature all *Cdkn2a* values would be multiplied by 2) (**Equation 12**). Then the cell-by-signature-gene matrix is averaged across the cell axis and subtracted from the mean of the cell-by-background matrix also averaged across the cell axis (**Equation 13**).

$$m_j = (1/n)\sum_{i=1}^{n} X_{ij}$$

(**Equation 9**) Where X is a matrix which contains the expression level of gene j in cell i and m_j is the average expression of gene j across all cells.

$$\mathrm{B}_1,\ \mathrm{B}_2,\ \ldots,\ \mathrm{B}_{\mathrm{n\_bins}}$$

(**Equation 10**) Where B_{n_bins} is the number of bins used to categorize every gene based on their mean expression.

$$s \in B_k, \quad \text{for } s \in S$$

$$BG_s = \{g \mid g \in B_k^{(n_{ctrl\,size})}\}, g \neq s$$

$$BG = \bigcup_{s \in S} BG_s$$

(**Equation 11**) BG: background gene set. Where S is the gene signature and s is a gene within S. B_k is a subset of genes that fall within the k-th expression bin based on their mean expression. Where g is a background gene selected from expression bin B_k and $n_{ctrl\,size}$ is the number of background genes selected for each signature gene s from the corresponding expression bin. Where BG_s is a set of background genes randomly selected from the same expression bin B_k as the signature gene s. BG is the union of all background genes selected for each signature gene s.

$$Y_{ij} = \{1\ if\ X_{ij} > 0, 0\ otherwise\}$$

$$Z_{ij} = Y_{ij} * I_j$$

(**Equation 12**) Z_{ij}: modified expression matrix. Where Y_{ij} is the optionally binarized expression matrix X_{ij} and I_j represents the optional importance values for gene j.

$$\overline{Y_{i,S}} = \frac{1}{|S|} \sum_{s \in S} Z_{i,s}, \quad \text{for } i \in \text{all cells}$$

$$\overline{Y_{i,BG}} = \frac{1}{|BG|} \sum_{g \in BG} Z_{i,g}, \quad \text{for } i \in \text{all cells}$$

$$\text{Score}_i = \overline{Y_{i,S}} - \overline{Y_{i,BG}}, \quad \text{for } i \in \text{all cells}$$

(**Equation 13**) Score$_i$: *SenePy* score for cell i. Where $|S|$ and $|BG|$ are the cardinality (number of elements in the set) of gene signatures S and BG. Where $Z_{i,s}$ and $Z_{i,g}$ represent the optionally amplified expression values of the genes in the gene signature S and background gene set BG, respectively.

Senescence burden in spatially resolved transcriptomics

Data was preprocessed in Scanpy and pots with fewer than 1000 detected genes were removed. Cells were normalized to 10,000 counts and log converted. The 8 heart-specific mouse hub signatures were used to score the spatially resolved mouse hearts independently using senepy.score_hub() with a translator() and with binarize and importance set to False because Visium data has higher gene counts than single-cell data. Outlier spots were identified in each sample if they fell 3 standard deviations outside the mean for their respective sample distribution in addition to a combined sample distribution ($Outlier > \mu + 3\sigma$). The outliers from each signature were merged to determine if any given spot was an outlier. Relative senescence burden is presented as the proportion of outlier spots. For the mouse brains, we used the top 150 most common genes in all the signatures because we had no specific mouse brain signatures. To determine spatial autocorrelation, we used the ESDA python package (https://pysal.org/esda/). The weights of the autocorrelation were weighted by the inverse of the Euclidean distance between two spots with a value of 1 to denote an outlier and 0 for normal spots. Three is used as a maximum value for Euclidean distance and the weights for distances beyond three are set to 0 (**Equation 14**).

$$I = \frac{n \sum_i \sum_j \left(\frac{1}{d(i,j)}\right) \cdot \delta(d(i,j) \leq 3)(x_i - \bar{x})(x_j - \bar{x})}{\sum_i \sum_j \left(\frac{1}{d(i,j)}\right) \cdot \delta(d(i,j) \leq 3) \sum_i (x_i - \bar{x})^2}, \text{p-value} = 1 - \left(\Phi(I)\right)$$

(**Equation 14**) I: Moran's I. Where n is the number of spots; $d(i,j)$ is the Euclidean distance between spot i and spot j; $\delta(d(i,j) \leq 3)$ is 1 if the distance is greater or equal to 3 and otherwise 0; and x_i and x_j are the values at spot i and spot j. $\Phi(I)$ is the CDF of the standard normal distribution at the Moran's I value.

Senescence burden in COVID-19 mortality

Single-cell lung data from 20 patients that died from COVID-19 and 7 control patients were collected from an available atlas[32]. Doublets were removed from each individual sample using SOLO[47] in combination with SCVI-tools. Cells with low counts or high mitochondrial reads were removed. SCVI-tools was used

to integrate the 27 samples, using sample ID as a categorical covariate and mitochondrial read percent, ribosomal read percent, and total counts as continuous covariates. Cell types were manually annotated using known cell-type markers (PanglaoDB). Cell types were scored with respective cell type hubs from SenePY (e.g., epithelial cells were scored with ciliated epithelial, basal cell, club cell, and pneumocyte hubs) using the senepy.score_all_cells() function. Cells were divided and scored as individual subtypes (e.g., AT1, AT2, airway epithelium). Cell outliers were identified in each sample if they fell 3 standard deviations outside the mean within every respective cell-type distribution ($Outlier > \mu + 3\sigma$). Outliers were merged across hubs to identify all cells with potential senescence burden and output as a proportion of total cells.

REFERENCES

1. Niccoli T, Partridge L. Ageing as a risk factor for disease. Curr Biol. 2012;22(17):R741-52. doi: 10.1016/j.cub.2012.07.024. PubMed PMID: 22975005.
2. Hernandez-Segura A, Nehme J, Demaria M. Hallmarks of Cellular Senescence. Trends Cell Biol. 2018;28(6):436-53. Epub 20180221. doi: 10.1016/j.tcb.2018.02.001. PubMed PMID: 29477613.
3. Childs BG, Durik M, Baker DJ, van Deursen JM. Cellular senescence in aging and age-related disease: from mechanisms to therapy. Nat Med. 2015;21(12):1424-35. doi: 10.1038/nm.4000. PubMed PMID: 26646499; PMCID: PMC4748967.
4. Lee S, Yu Y, Trimpert J, Benthani F, Mairhofer M, Richter-Pechanska P, Wyler E, Belenki D, Kaltenbrunner S, Pammer M, Kausche L, Firsching TC, Dietert K, Schotsaert M, Martinez-Romero C, Singh G, Kunz S, Niemeyer D, Ghanem R, Salzer HJF, Paar C, Mulleder M, Uccellini M, Michaelis EG, Khan A, Lau A, Schonlein M, Habringer A, Tomasits J, Adler JM, Kimeswenger S, Gruber AD, Hoetzenecker W, Steinkellner H, Purfurst B, Motz R, Di Pierro F, Lamprecht B, Osterrieder N, Landthaler M, Drosten C, Garcia-Sastre A, Langer R, Ralser M, Eils R, Reimann M, Fan DNY, Schmitt CA. Virus-induced senescence is a driver and therapeutic target in COVID-19. Nature. 2021;599(7884):283-9. Epub 20210913. doi: 10.1038/s41586-021-03995-1. PubMed PMID: 34517409.
5. van Deursen JM. The role of senescent cells in ageing. Nature. 2014;509(7501):439-46. doi: 10.1038/nature13193. PubMed PMID: 24848057; PMCID: PMC4214092.
6. Chimenti C, Kajstura J, Torella D, Urbanek K, Heleniak H, Colussi C, Di Meglio F, Nadal-Ginard B, Frustaci A, Leri A, Maseri A, Anversa P. Senescence and death of primitive cells and myocytes lead to premature cardiac aging and heart failure. Circ Res. 2003;93(7):604-13. Epub 20030904. doi: 10.1161/01.RES.0000093985.76901.AF. PubMed PMID: 12958145.
7. Aguayo-Mazzucato C, Andle J, Lee TB, Jr., Midha A, Talemal L, Chipashvili V, Hollister-Lock J, van Deursen J, Weir G, Bonner-Weir S. Acceleration of beta Cell Aging Determines Diabetes and Senolysis

735 Improves Disease Outcomes. Cell Metab. 2019;30(1):129-42 e4. Epub 20190530. doi:
736 10.1016/j.cmet.2019.05.006. PubMed PMID: 31155496; PMCID: PMC6610720.
737 8. Zhang P, Kishimoto Y, Grammatikakis I, Gottimukkala K, Cutler RG, Zhang S, Abdelmohsen K,
738 Bohr VA, Misra Sen J, Gorospe M, Mattson MP. Senolytic therapy alleviates Abeta-associated
739 oligodendrocyte progenitor cell senescence and cognitive deficits in an Alzheimer's disease model. Nat
740 Neurosci. 2019;22(5):719-28. Epub 20190401. doi: 10.1038/s41593-019-0372-9. PubMed PMID:
741 30936558; PMCID: PMC6605052.
742 9. Camell CD, Yousefzadeh MJ, Zhu Y, Prata L, Huggins MA, Pierson M, Zhang L, O'Kelly RD,
743 Pirtskhalava T, Xun P, Ejima K, Xue A, Tripathi U, Espindola-Netto JM, Giorgadze N, Atkinson EJ, Inman
744 CL, Johnson KO, Cholensky SH, Carlson TW, LeBrasseur NK, Khosla S, O'Sullivan MG, Allison DB,
745 Jameson SC, Meves A, Li M, Prakash YS, Chiarella SE, Hamilton SE, Tchkonia T, Niedernhofer LJ, Kirkland
746 JL, Robbins PD. Senolytics reduce coronavirus-related mortality in old mice. Science. 2021;373(6552).
747 Epub 20210608. doi: 10.1126/science.abe4832. PubMed PMID: 34103349; PMCID: PMC8607935.
748 10. Baker DJ, Childs BG, Durik M, Wijers ME, Sieben CJ, Zhong J, Saltness RA, Jeganathan KB,
749 Verzosa GC, Pezeshki A, Khazaie K, Miller JD, van Deursen JM. Naturally occurring p16(Ink4a)-positive
750 cells shorten healthy lifespan. Nature. 2016;530(7589):184-9. Epub 20160203. doi:
751 10.1038/nature16932. PubMed PMID: 26840489; PMCID: PMC4845101.
752 11. Born E, Lipskaia L, Breau M, Houssaini A, Beaulieu D, Marcos E, Pierre R, Do Cruzeiro M, Lefevre
753 M, Derumeaux G, Bulavin DV, Delcroix M, Quarck R, Reen V, Gil J, Bernard D, Flaman JM, Adnot S, Abid
754 S. Eliminating Senescent Cells Can Promote Pulmonary Hypertension Development and Progression.
755 Circulation. 2023;147(8):650-66. Epub 20221214. doi: 10.1161/CIRCULATIONAHA.122.058794. PubMed
756 PMID: 36515093.
757 12. Gorgoulis V, Adams PD, Alimonti A, Bennett DC, Bischof O, Bishop C, Campisi J, Collado M,
758 Evangelou K, Ferbeyre G, Gil J, Hara E, Krizhanovsky V, Jurk D, Maier AB, Narita M, Niedernhofer L,
759 Passos JF, Robbins PD, Schmitt CA, Sedivy J, Vougas K, von Zglinicki T, Zhou D, Serrano M, Demaria M.
760 Cellular Senescence: Defining a Path Forward. Cell. 2019;179(4):813-27. doi:
761 10.1016/j.cell.2019.10.005. PubMed PMID: 31675495.
762 13. Hernandez-Segura A, de Jong TV, Melov S, Guryev V, Campisi J, Demaria M. Unmasking
763 Transcriptional Heterogeneity in Senescent Cells. Curr Biol. 2017;27(17):2652-60 e4. Epub 20170830.
764 doi: 10.1016/j.cub.2017.07.033. PubMed PMID: 28844647; PMCID: PMC5788810.
765 14. Casella G, Munk R, Kim KM, Piao Y, De S, Abdelmohsen K, Gorospe M. Transcriptome signature
766 of cellular senescence. Nucleic Acids Res. 2019;47(14):7294-305. doi: 10.1093/nar/gkz555. PubMed
767 PMID: 31251810; PMCID: PMC6698740.
768 15. Nelson G, Wordsworth J, Wang C, Jurk D, Lawless C, Martin-Ruiz C, von Zglinicki T. A senescent
769 cell bystander effect: senescence-induced senescence. Aging Cell. 2012;11(2):345-9. Epub 20120209.
770 doi: 10.1111/j.1474-9726.2012.00795.x. PubMed PMID: 22321662; PMCID: PMC3488292.
771 16. Teo YV, Rattanavirotkul N, Olova N, Salzano A, Quintanilla A, Tarrats N, Kiourtis C, Muller M,
772 Green AR, Adams PD, Acosta JC, Bird TG, Kirschner K, Neretti N, Chandra T. Notch Signaling Mediates
773 Secondary Senescence. Cell Rep. 2019;27(4):997-1007 e5. doi: 10.1016/j.celrep.2019.03.104. PubMed
774 PMID: 31018144; PMCID: PMC6486482.
775 17. Prieur A, Besnard E, Babled A, Lemaitre JM. p53 and p16(INK4A) independent induction of
776 senescence by chromatin-dependent alteration of S-phase progression. Nat Commun. 2011;2:473.
777 Epub 20110913. doi: 10.1038/ncomms1473. PubMed PMID: 21915115.

18. Li J, Poi MJ, Tsai MD. Regulatory mechanisms of tumor suppressor P16(INK4A) and their relevance to cancer. Biochemistry. 2011;50(25):5566-82. Epub 20110606. doi: 10.1021/bi200642e. PubMed PMID: 21619050; PMCID: PMC3127263.
19. Haferkamp S, Scurr LL, Becker TM, Frausto M, Kefford RF, Rizos H. Oncogene-induced senescence does not require the p16(INK4a) or p14ARF melanoma tumor suppressors. J Invest Dermatol. 2009;129(8):1983-91. Epub 20090212. doi: 10.1038/jid.2009.5. PubMed PMID: 19212341.
20. Saul D, Kosinsky RL, Atkinson EJ, Doolittle ML, Zhang X, LeBrasseur NK, Pignolo RJ, Robbins PD, Niedernhofer LJ, Ikeno Y, Jurk D, Passos JF, Hickson LJ, Xue A, Monroe DG, Tchkonia T, Kirkland JL, Farr JN, Khosla S. A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. Nat Commun. 2022;13(1):4827. Epub 20220816. doi: 10.1038/s41467-022-32552-1. PubMed PMID: 35974106; PMCID: PMC9381717.
21. SenNet C. NIH SenNet Consortium to map senescent cells throughout the human lifespan to understand physiological health. Nat Aging. 2022;2(12):1090-100. Epub 20221220. doi: 10.1038/s43587-022-00326-5. PubMed PMID: 36936385; PMCID: PMC10019484.
22. Tabula Muris C. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. Nature. 2020;583(7817):590-5. Epub 20200715. doi: 10.1038/s41586-020-2496-1. PubMed PMID: 32669714; PMCID: PMC8240505.
23. Zhou Y, Su Y, Li S, Kennedy BC, Zhang DY, Bond AM, Sun Y, Jacob F, Lu L, Hu P, Viaene AN, Helbig I, Kessler SK, Lucas T, Salinas RD, Gu X, Chen HI, Wu H, Kleinman JE, Hyde TM, Nauen DW, Weinberger DR, Ming GL, Song H. Molecular landscapes of human hippocampal immature neurons across lifespan. Nature. 2022;607(7919):527-33. Epub 20220706. doi: 10.1038/s41586-022-04912-w. PubMed PMID: 35794479; PMCID: PMC9316413.
24. Tabula Sapiens C, Jones RC, Karkanias J, Krasnow MA, Pisco AO, Quake SR, Salzman J, Yosef N, Bulthaup B, Brown P, Harper W, Hemenez M, Ponnusamy R, Salehi A, Sanagavarapu BA, Spallino E, Aaron KA, Concepcion W, Gardner JM, Kelly B, Neidlinger N, Wang Z, Crasta S, Kolluru S, Morri M, Pisco AO, Tan SY, Travaglini KJ, Xu C, Alcantara-Hernandez M, Almanzar N, Antony J, Beyersdorf B, Burhan D, Calcuttawala K, Carter MM, Chan CKF, Chang CA, Chang S, Colville A, Crasta S, Culver RN, Cvijovic I, D'Amato G, Ezran C, Galdos FX, Gillich A, Goodyer WR, Hang Y, Hayashi A, Houshdaran S, Huang X, Irwin JC, Jang S, Juanico JV, Kershner AM, Kim S, Kiss B, Kolluru S, Kong W, Kumar ME, Kuo AH, Leylek R, Li B, Loeb GB, Lu WJ, Mantri S, Markovic M, McAlpine PL, de Morree A, Morri M, Mrouj K, Mukherjee S, Muser T, Neuhofer P, Nguyen TD, Perez K, Phansalkar R, Pisco AO, Puluca N, Qi Z, Rao P, Raquer-McKay H, Schaum N, Scott B, Seddighzadeh B, Segal J, Sen S, Sikandar S, Spencer SP, Steffes LC, Subramaniam VR, Swarup A, Swift M, Travaglini KJ, Van Treuren W, Trimm E, Veizades S, Vijayakumar S, Vo KC, Vorperian SK, Wang W, Weinstein HNW, Winkler J, Wu TTH, Xie J, Yung AR, Zhang Y, Detweiler AM, Mekonen H, Neff NF, Sit RV, Tan M, Yan J, Bean GR, Charu V, Forgo E, Martin BA, Ozawa MG, Silva O, Tan SY, Toland A, Vemuri VNP, Afik S, Awayan K, Botvinnik OB, Byrne A, Chen M, Dehghannasiri R, Detweiler AM, Gayoso A, Granados AA, Li Q, Mahmoudabadi G, McGeever A, de Morree A, Olivieri JE, Park M, Pisco AO, Ravikumar N, Salzman J, Stanley G, Swift M, Tan M, Tan W, Tarashansky AJ, Vanheusden R, Vorperian SK, Wang P, Wang S, Xing G, Xu C, Yosef N, Alcantara-Hernandez M, Antony J, Chan CKF, Chang CA, Colville A, Crasta S, Culver R, Dethlefsen L, Ezran C, Gillich A, Hang Y, Ho PY, Irwin JC, Jang S, Kershner AM, Kong W, Kumar ME, Kuo AH, Leylek R, Liu S, Loeb GB, Lu WJ, Maltzman JS, Metzger RJ, de Morree A, Neuhofer P, Perez K, Phansalkar R, Qi Z, Rao P, Raquer-McKay H, Sasagawa K, Scott B, Sinha R, Song H, Spencer SP, Swarup A, Swift M, Travaglini KJ, Trimm E, Veizades S, Vijayakumar S, Wang B, Wang W, Winkler J, Xie J, Yung AR, Artandi SE, Beachy PA, Clarke MF, Giudice

LC, Huang FW, Huang KC, Idoyaga J, Kim SK, Krasnow M, Kuo CS, Nguyen P, Quake SR, Rando TA, Red-Horse K, Reiter J, Relman DA, Sonnenburg JL, Wang B, Wu A, Wu SM, Wyss-Coray T. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science. 2022;376(6594):eabl4896. Epub 20220513. doi: 10.1126/science.abl4896. PubMed PMID: 35549404; PMCID: PMC9812260.

25. Zou Z, Long X, Zhao Q, Zheng Y, Song M, Ma S, Jing Y, Wang S, He Y, Esteban CR, Yu N, Huang J, Chan P, Chen T, Izpisua Belmonte JC, Zhang W, Qu J, Liu GH. A Single-Cell Transcriptomic Atlas of Human Skin Aging. Dev Cell. 2021;56(3):383-97 e8. Epub 20201124. doi: 10.1016/j.devcel.2020.11.002. PubMed PMID: 33238152.

26. Litvinukova M, Talavera-Lopez C, Maatz H, Reichart D, Worth CL, Lindberg EL, Kanda M, Polanski K, Heinig M, Lee M, Nadelmann ER, Roberts K, Tuck L, Fasouli ES, DeLaughter DM, McDonough B, Wakimoto H, Gorham JM, Samari S, Mahbubani KT, Saeb-Parsy K, Patone G, Boyle JJ, Zhang H, Zhang H, Viveiros A, Oudit GY, Bayraktar OA, Seidman JG, Seidman CE, Noseda M, Hubner N, Teichmann SA. Cells of the adult human heart. Nature. 2020;588(7838):466-72. Epub 20200924. doi: 10.1038/s41586-020-2797-4. PubMed PMID: 32971526; PMCID: PMC7681775.

27. Han X, Zhou Z, Fei L, Sun H, Wang R, Chen Y, Chen H, Wang J, Tang H, Ge W, Zhou Y, Ye F, Jiang M, Wu J, Xiao Y, Jia X, Zhang T, Ma X, Zhang Q, Bai X, Lai S, Yu C, Zhu L, Lin R, Gao Y, Wang M, Wu Y, Zhang J, Zhan R, Zhu S, Hu H, Wang C, Chen M, Huang H, Liang T, Chen J, Wang W, Zhang D, Guo G. Construction of a human cell landscape at single-cell level. Nature. 2020;581(7808):303-9. Epub 20200325. doi: 10.1038/s41586-020-2157-4. PubMed PMID: 32214235.

28. Reyfman PA, Walter JM, Joshi N, Anekalla KR, McQuattie-Pimentel AC, Chiu S, Fernandez R, Akbarpour M, Chen CI, Ren Z, Verma R, Abdala-Valencia H, Nam K, Chi M, Han S, Gonzalez-Gonzalez FJ, Soberanes S, Watanabe S, Williams KJN, Flozak AS, Nicholson TT, Morgan VK, Winter DR, Hinchcliff M, Hrusch CL, Guzy RD, Bonham CA, Sperling AI, Bag R, Hamanaka RB, Mutlu GM, Yeldandi AV, Marshall SA, Shilatifard A, Amaral LAN, Perlman H, Sznajder JI, Argento AC, Gillespie CT, Dematte J, Jain M, Singer BD, Ridge KM, Lam AP, Bharat A, Bhorade SM, Gottardi CJ, Budinger GRS, Misharin AV. Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. Am J Respir Crit Care Med. 2019;199(12):1517-36. doi: 10.1164/rccm.201712-2410OC. PubMed PMID: 30554520; PMCID: PMC6580683.

29. MacParland SA, Liu JC, Ma XZ, Innes BT, Bartczak AM, Gage BK, Manuel J, Khuu N, Echeverri J, Linares I, Gupta R, Cheng ML, Liu LY, Camat D, Chung SW, Seliga RK, Shao Z, Lee E, Ogawa S, Ogawa M, Wilson MD, Fish JE, Selzner M, Ghanekar A, Grant D, Greig P, Sapisochin G, Selzner N, Winegarden N, Adeyi O, Keller G, Bader GD, McGilvray ID. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nat Commun. 2018;9(1):4383. Epub 20181022. doi: 10.1038/s41467-018-06318-7. PubMed PMID: 30348985; PMCID: PMC6197289.

30. Zhang MJ, Pisco AO, Darmanis S, Zou J. Mouse aging cell atlas analysis reveals global and cell type-specific aging signatures. Elife. 2021;10. Epub 20210413. doi: 10.7554/eLife.62293. PubMed PMID: 33847263; PMCID: PMC8046488.

31. Omori S, Wang TW, Johmura Y, Kanai T, Nakano Y, Kido T, Susaki EA, Nakajima T, Shichino S, Ueha S, Ozawa M, Yokote K, Kumamoto S, Nishiyama A, Sakamoto T, Yamaguchi K, Hatakeyama S, Shimizu E, Katayama K, Yamada Y, Yamazaki S, Iwasaki K, Miyoshi C, Funato H, Yanagisawa M, Ueno H, Imoto S, Furukawa Y, Yoshida N, Matsushima K, Ueda HR, Miyajima A, Nakanishi M. Generation of a p16 Reporter Mouse and Its Use to Characterize and Target p16(high) Cells In Vivo. Cell Metab. 2020;32(5):814-28 e6. Epub 20200918. doi: 10.1016/j.cmet.2020.09.006. PubMed PMID: 32949498.

32. Melms JC, Biermann J, Huang H, Wang Y, Nair A, Tagore S, Katsyv I, Rendeiro AF, Amin AD, Schapiro D, Frangieh CJ, Luoma AM, Filliol A, Fang Y, Ravichandran H, Clausi MG, Alba GA, Rogava M, Chen SW, Ho P, Montoro DT, Kornberg AE, Han AS, Bakhoum MF, Anandasabapathy N, Suarez-Farinas M, Bakhoum SF, Bram Y, Borczuk A, Guo XV, Lefkowitch JH, Marboe C, Lagana SM, Del Portillo A, Tsai EJ, Zorn E, Markowitz GS, Schwabe RF, Schwartz RE, Elemento O, Saqi A, Hibshoosh H, Que J, Izar B. A molecular single-cell lung atlas of lethal COVID-19. Nature. 2021;595(7865):114-9. Epub 20210429. doi: 10.1038/s41586-021-03569-1. PubMed PMID: 33915568; PMCID: PMC8814825.

33. Yamada S, Ko T, Hatsuse S, Nomura S, Zhang B, Dai Z, Inoue S, Kubota M, Sawami K, Yamada T, Sassa T, Katagiri M, Fujita K, Katoh M, Ito M, Harada M, Toko H, Takeda N, Morita H, Aburatani H, Komuro I. Spatiotemporal transcriptome analysis reveals critical roles for mechano-sensing genes at the border zone in remodeling after myocardial infarction. Nature Cardiovascular Research. 2022;1(11):1072-83. doi: 10.1038/s44161-022-00140-7.

34. Liu W, Baker SS, Baker RD, Nowak NJ, Zhu L. Upregulation of hemoglobin expression by oxidative stress in hepatocytes and its implication in nonalcoholic steatohepatitis. PLoS One. 2011;6(9):e24363. Epub 20110912. doi: 10.1371/journal.pone.0024363. PubMed PMID: 21931690; PMCID: PMC3171444.

35. Nishi H, Inagi R, Kato H, Tanemoto M, Kojima I, Son D, Fujita T, Nangaku M. Hemoglobin is expressed by mesangial cells and reduces oxidant stress. J Am Soc Nephrol. 2008;19(8):1500-8. Epub 20080430. doi: 10.1681/ASN.2007101085. PubMed PMID: 18448584; PMCID: PMC2488266.

36. Yousefzadeh MJ, Zhao J, Bukata C, Wade EA, McGowan SJ, Angelini LA, Bank MP, Gurkar AU, McGuckian CA, Calubag MF, Kato JI, Burd CE, Robbins PD, Niedernhofer LJ. Tissue specificity of senescent cell accumulation during physiologic and accelerated aging of mice. Aging Cell. 2020;19(3):e13094. Epub 20200125. doi: 10.1111/acel.13094. PubMed PMID: 31981461; PMCID: PMC7059165.

37. Zhu Y, Tchkonia T, Fuhrmann-Stroissnigg H, Dai HM, Ling YY, Stout MB, Pirtskhalava T, Giorgadze N, Johnson KO, Giles CB, Wren JD, Niedernhofer LJ, Robbins PD, Kirkland JL. Identification of a novel senolytic agent, navitoclax, targeting the Bcl-2 family of anti-apoptotic factors. Aging Cell. 2016;15(3):428-35. Epub 20160318. doi: 10.1111/acel.12445. PubMed PMID: 26711051; PMCID: PMC4854923.

38. Amor C, Feucht J, Leibold J, Ho YJ, Zhu C, Alonso-Curbelo D, Mansilla-Soto J, Boyer JA, Li X, Giavridis T, Kulick A, Houlihan S, Peerschke E, Friedman SL, Ponomarev V, Piersigilli A, Sadelain M, Lowe SW. Senolytic CAR T cells reverse senescence-associated pathologies. Nature. 2020;583(7814):127-32. Epub 20200617. doi: 10.1038/s41586-020-2403-9. PubMed PMID: 32555459; PMCID: PMC7583560.

39. Cheng N, Kim KH, Lau LF. Senescent hepatic stellate cells promote liver regeneration through IL-6 and ligands of CXCR2. JCI Insight. 2022;7(14). Epub 20220616. doi: 10.1172/jci.insight.158207. PubMed PMID: 35708907; PMCID: PMC9431681.

40. Roos CM, Zhang B, Palmer AK, Ogrodnik MB, Pirtskhalava T, Thalji NM, Hagler M, Jurk D, Smith LA, Casaclang-Verzosa G, Zhu Y, Schafer MJ, Tchkonia T, Kirkland JL, Miller JD. Chronic senolytic treatment alleviates established vasomotor dysfunction in aged or atherosclerotic mice. Aging Cell. 2016;15(5):973-7. Epub 20160805. doi: 10.1111/acel.12458. PubMed PMID: 26864908; PMCID: PMC5013022.

41. Kiss T, Nyul-Toth A, DelFavero J, Balasubramanian P, Tarantini S, Faakye J, Gulej R, Ahire C, Ungvari A, Yabluchanskiy A, Wiley G, Garman L, Ungvari Z, Csiszar A. Spatial transcriptomic analysis reveals inflammatory foci defined by senescent cells in the white matter, hippocampi and cortical grey

910 matter in the aged mouse brain. Geroscience. 2022;44(2):661-81. Epub 20220131. doi:
911 10.1007/s11357-022-00521-7. PubMed PMID: 35098444; PMCID: PMC9135953.
912 42. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis.
913 Genome Biol. 2018;19(1):15. Epub 20180206. doi: 10.1186/s13059-017-1382-0. PubMed PMID:
914 29409532; PMCID: PMC5802054.
915 43. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and
916 annotation of single-cell transcriptomics data with deep generative models. Mol Syst Biol.
917 2021;17(1):e9620. doi: 10.15252/msb.20209620. PubMed PMID: 33491336; PMCID: PMC7829634.
918 44. Gayoso A, Lopez R, Xing G, Boyeau P, Valiollah Pour Amiri V, Hong J, Wu K, Jayasuriya M,
919 Mehlman E, Langevin M, Liu Y, Samaran J, Misrachi G, Nazaret A, Clivio O, Xu C, Ashuach T, Gabitto M,
920 Lotfollahi M, Svensson V, da Veiga Beltrame E, Kleshchevnikov V, Talavera-Lopez C, Pachter L, Theis FJ,
921 Streets A, Jordan MI, Regier J, Yosef N. A Python library for probabilistic analysis of single-cell omics
922 data. Nat Biotechnol. 2022;40(2):163-6. doi: 10.1038/s41587-021-01206-w. PubMed PMID: 35132262.
923 45. Franzen O, Gan LM, Bjorkegren JLM. PanglaoDB: a web server for exploration of mouse and
924 human single-cell RNA sequencing data. Database (Oxford). 2019;2019. doi: 10.1093/database/baz046.
925 PubMed PMID: 30951143; PMCID: PMC6450036.
926 46. Fang Z, Liu X, Peltz G. GSEApy: a comprehensive package for performing gene set enrichment
927 analysis in Python. Bioinformatics. 2023;39(1). doi: 10.1093/bioinformatics/btac757. PubMed PMID:
928 36426870; PMCID: PMC9805564.
929 47. Bernstein NJ, Fong NL, Lam I, Roy MA, Hendrickson DG, Kelley DR. Solo: Doublet Identification
930 in Single-Cell RNA-Seq via Semi-Supervised Deep Learning. Cell Syst. 2020;11(1):95-101 e5. Epub
931 20200626. doi: 10.1016/j.cels.2020.05.010. PubMed PMID: 32592658.
932

Figure 1. Known senescence markers are cell-type-specific and poorly characterize *in vivo* cellular senescence. (**A**) There is an insignificant overlap between a universal organismal aging signature and previously reported senescence markers (p = 0.58, Hypergeometric). The universal signature are genes present in at least 50% of the cell-specific gene sets elucidated by differential expression between young and old cells previously. (**B**) A histogram depicting the 3-month to 24-month ratio of all cells expressing every gene in the *Tabula Muris Senis* dataset. The top 5% of the distribution is marked by the dashed line and known senescence markers above this are labeled. (**C, D**) UMAPs (left) representing *CDKN2A*+ ($p16^{ink4a}$ encoding gene) cells from the mouse and human datasets. Cells from 24-, and 30-month mice are denoted old while 1-, and 3-month mice are young. UMAPs (right) showing all cells in the datasets and are labeled by broad cell classifications. Bar graphs show the percentage of *CDKN2A*+ cells relative to all cells in the respective datasets. (**E, F**) Cell-specific maps of marker dynamics in mice and humans. Vertical dashed lines represent the start of a tissue and cell types from that tissue are classified and depicted by shape. Multiple cell types belonging to the same class are overplotted. Gain represents the percent increase of cells expressing the marker relative to young organisms. Bar plots depict the percentage of cell populations which the respective gene is a suitable marker and they are colored by senescence-associated function.

Figure 2. *De novo* cell-type specific signatures derived algorithmically from single-cell transcriptomes. (**A**) Overview of the algorithm used to define cell-specific signatures from mice and humans (see methods). Workflow ends with a cosine signature comparison used to determine signature similarity. (**B**) Example signatures derived from mouse cardiomyocytes and human hippocampal choroid plexus cells. Each node represents a gene and the connections represent positive cell co-positivity. Connections are weighted by Pearson's R. The colors represent distinct hub signatures within the overall cell signatures. (**C**, **D**) Representative diagram of all derived signatures from mouse and humans. Each dot represents a signature and is sized by its number of genes. The dot color is the respective enrichment for each signature compared to previously known senescence

markers (Hypergeometric). Each signature is connected to its most similar signature and the color of the connection is based on the cosine similarity. (**E**) Network similarity analysis of mouse cell-specific novel senescence signatures and organismal aging signatures. Each shape represents a signature and lines represent significant similarity between them. Similarity (strength of connections) is defined as -$\log_{10}$(BH-corrected Hypergeometric P). Network is clustered and colored by Louvain's algorithm.

.

Figure 3. Multiple modes of senescence exist within the same cell populations. (**A**) Signature derived from mouse tongue keratinocytes. Each node represents a gene and the connections represent positive cell co-positivity. Connections are weighted by Pearson's R. Nodes are colored by Louvain-based assignment to distinct hub signatures. (**B**) GO and (**C**) KEGG gene set enrichment of the two keratinocytes hub signatures. The "senescence" gene set is the pre-defined senescence markers used in this study. Horizonal dashed line represents FDR p = 0.05. (**D**) Pairwise enrichment of the two keratinocyte hubs against all other senescence signatures. (**E**) The strip plot depicts the score of each keratinocyte determined by *SenePy* using the forementioned hubs. Horizontal dashed lines represent three standard deviations above the mean. (**F**) Temporal kinetics of the proportion of cells scored three standard deviations above the mean by *SenePy* for the two keratinocyte hubs. (**G**) Hierarchical clustering of fibroblast hub signatures from mouse lungs, tracheas, and hearts based on cosine similarity. (**H**) Temporal kinetics of the proportion of lung, trachea, and heart fibroblast cells scored high by *SenePy* using their respective signatures. (**I**) GO gene set enrichment of the most similar trachea and lung fibroblast hubs. Gene sets specific to the trachea hub are colored. All enrichment plots use BH-corrected Fisher's Exact P values.

Figure 4. Cell-specific signatures are unique but share some genes and biological pathways. (**A**) Plot depicting the most commonly found genes from the novel mouse cell-specific signatures. Significance was determined from 1000 random sets of signatures equal in size from the background set of genes. (**B**) The 12 most enriched KEGG pathways from a "universal" signature comprised of

genes with a high mouse signature prevalence ($p < 0.01$). (**C**) The most commonly enriched KEGG and GO gene sets in every mouse signature. The bars note the percent of signatures enriched for the given pathway. (**D**) The most commonly enriched transcription factor motifs in the promotors of signature genes from mice. Bars represent the percent of signatures enriched for the given pathway. (**E**) Plots depicting the most commonly found genes from the novel human cell-specific signatures. (**F**) The 12 most enriched KEGG pathways from a "universal" signature comprised of genes with a high human signature prevalence ($p < 0.04$). (**G**) The most commonly enriched KEGG and GO gene sets in every human signature. The bars note the percent of signatures enriched for the given pathway. (**D**) The most commonly enriched transcription factor motifs in the promotors of signature genes from humans. Bars represent the percent of signatures enriched for the given pathway.

Figure 5. The cell-specific kinetics of senescent cell accumulation with organismal age. UMAPs of (**A**) mouse and (**B**) human cells depicting broad cell classification and overlayed with cells which were outliers determined by their *SenePy* score. (**C**) The proportional increase of *SenePy* outlier cells in old mice (24- or 30-month) relative to 3-month-old mice. (**D**) The proportion of *SenePy* human outlier cells in across age bins. Each row represents 0-16% and grey rectangles note that no data is available. (**E**) The fraction of *SenPy* outliers in individual cell types stratified by heart tissue donor. Age increases along the x-axis from left to right.

Figure 6. *SenePy identifies ground-truth in vivo senescence more robustly than established markers of senescence.* (**A**) UMAPs of single-cells from the kidney which were enriched for td-Tomato+ cells. (**B**) Enrichment analysis of differentially abundant genes in the td-Tomato+ kidney cells. Gene sets were derived in this study and we also include the SenMayo signature. (**C**) Density plots depicting *SenePy* score distributions calculated in kidney cells using kidney-specific *SenePy* signatures. (**D**) Enrichment analysis of differentially abundant genes in the td-Tomato+ liver cells. (**E**) Enrichment analysis of lung tissue genes downregulated after senolytic treatment in mice. Blue lables

indicate *SenePy* signatures. (**F**) Enrichment analysis of gene mRNAs more abundant in irradiated endothelial cells and (**G**) fibroblasts. (**H**) Enrichment analysis of gene mRNAs more abundant in replicative fibroblast senescence. (**I**) UMAPs of single-cells from the liver which were enriched for td-Tomato+ cells. (**J**) UMAP and representative bar graph of td-Tomato+ cells with *Cdkn2a*+ cells labeled. (**K**) UMAPs and representative bar graphs of td-Tomato+ and (**L**) td-Tomato- cells with *Cdkn1a*+ cells labeled.

Figure 7. *SenePy predicts elevated senescence burden in severe disease* (**A**) UMAP showing cells from uninfected control lungs (n = 7) and patients who died from COVID-19 (n = 22). (**B**) UMAP of control cells (left) and COVID-19 cells (right) with *SenePy* outlier cells labeled. (**C**) Proportion cells identified with *SenePy* from AT1, AT2, and airway epithelial cells (p = 0.004, Mann-Whitney). (**D**) Distributions of *SenePy* score from the major cell lung cell classes. (**E**) Heatmap with the relative senescence burden of each cell type in each patient. (**F**) Representative whole heart H&E staining overlayed with spatially resolved 10x Visium spots. Yellow spots are identified as senescence outliers from their *SenePy* score. Box plot (right) shows the proportion of identified spots at each time point (p = 0.30, one-way ANOVA) (**G**) The bottom images represent post-MI fibrosis via normalized *Col1a1* expression. Scale bars represent 100 μM. The correlation between senescence-like spots and *Col1a1* expression for each sample is shown by the right box plot ($-\log_{10}$[Pearson's R p-value]). (**H**) The relative contribution to the overall calculated senescence burden from the 8 hubs used to score the spots (**I**) Spatial clustering of the senescence-like spots ($-\log_{10}$[Moran's I p-value]). Horizontal line represents p = 0.05. (**J**) Representative H&E image of coronal sections spatially resolved by 10x Visium which were taken from mice 24h after exposure to LPS. Yellow spots are identified as senescence outliers from their *SenePy* score. Summary plot (right) depicts spatial autocorrelation of spots in LPS and saline treated mice ($-\log_{10}$[Moran's I p-value]).

Supplemental figure 1. (**A, B**) UMAPs of all mouse and human cells used that are colored by tissue of origin, age, and sex.

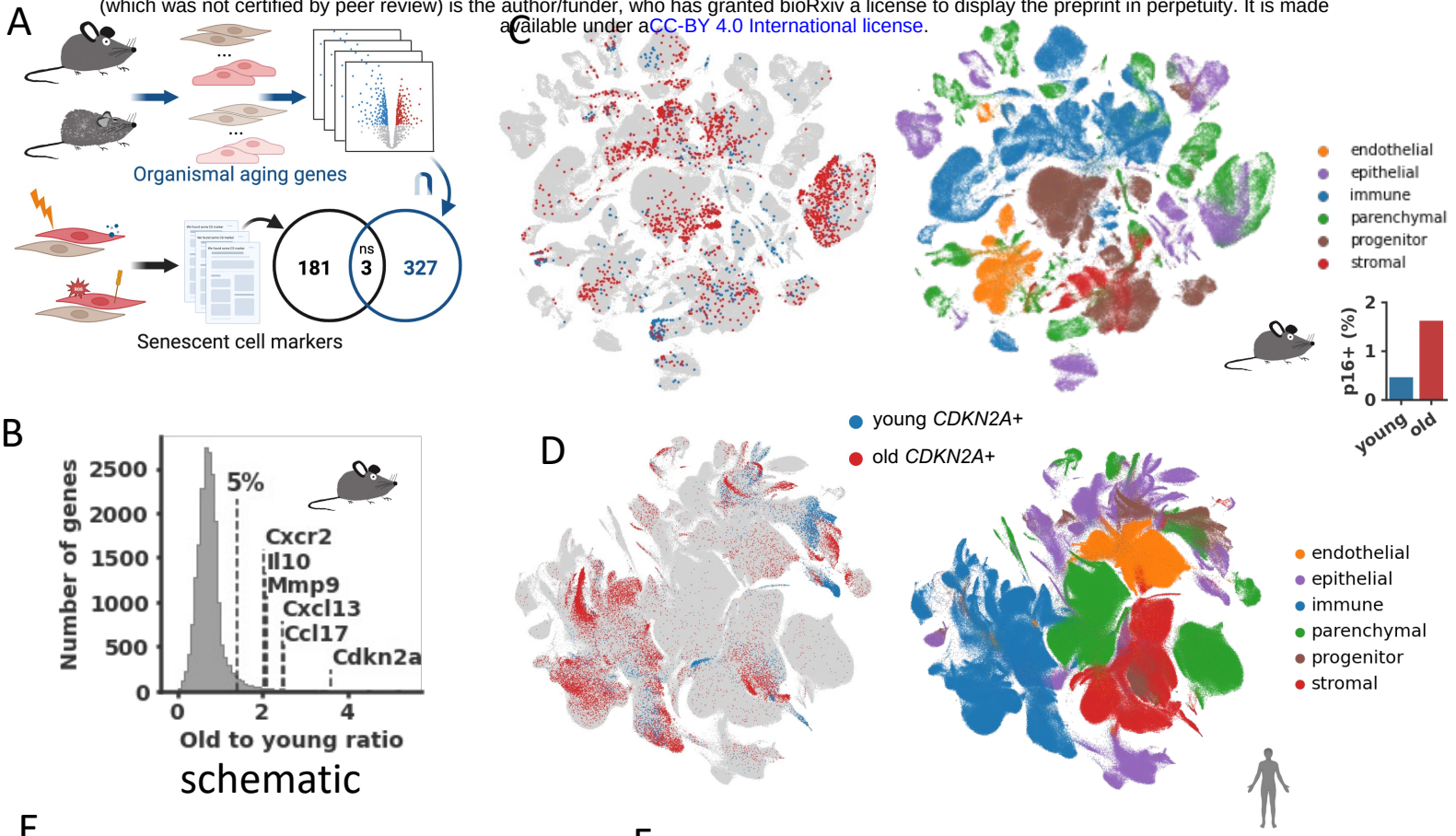
Supplemental figure 2. (**A**) Overview of the senescence marker gene set enrichment analysis. Cell-specific organismal aging signatures determined in a previous study were compared to a curated set of senescence markers taken from multiple studies. (**B**) The statistical enrichment of every organismal aging signature for senescence markers. Color represents the significance of enrichment (Hypergeometric, BH-corrected) and the size of the bar represents the percentage of senescence markers present in the aging signature. The aligned plot on the right shows the proportion of cells in each population that were expressing the proliferation marker Ki-67. (**C**) There is a negative correlation between senescence gene set enrichment and the proliferation state of a population ($p = 0.038$, Pearson's R).

Supplemental Figure 3. (**A**) UMAP representing *Cxcl13*+ cells in all mouse cells. (**B**) UMAP (top) of human skin cells colored by cell type. UMAP (bottom) of human skill cells positive for *CDKN2A* from young (age 18-34) and old (age 34-58) donors. Dot plot depicts the percentage of *CDKN2A*+ cells from old and young human skin cells. (**C**) UMAP representing *CDKN1A*+ (gene that encodes $p21^{cip1}$) cells in all mouse and human cells.

Supplemental Figure 4. Hierarchal clustering of signatures based on gene-set similarity. Similarity is defined as the inverse of $\log_{10}$(BH-corrected Hypergeometric P) for each pairwise comparison

Supplemental Figure 5. (**A**) The proportion of mouse cells expressing Tert and (**B**) Ki67. (**C,D**) The relationship between cell types expressing Tert and Ki67 with their gain in *SenePy* identified cells.

129 **Supplemental Figure 6.** (**A**) Enrichment analysis of gene mRNAs less abundant in mouse muscle
130 tissue after treatment with a senolytic. Blue lables indicate *SenePy* signatures. (**B**) Enrichment analysis
131 of gene mRNAs more abundant after irradiation induced senescence of cultured endothelial cells.



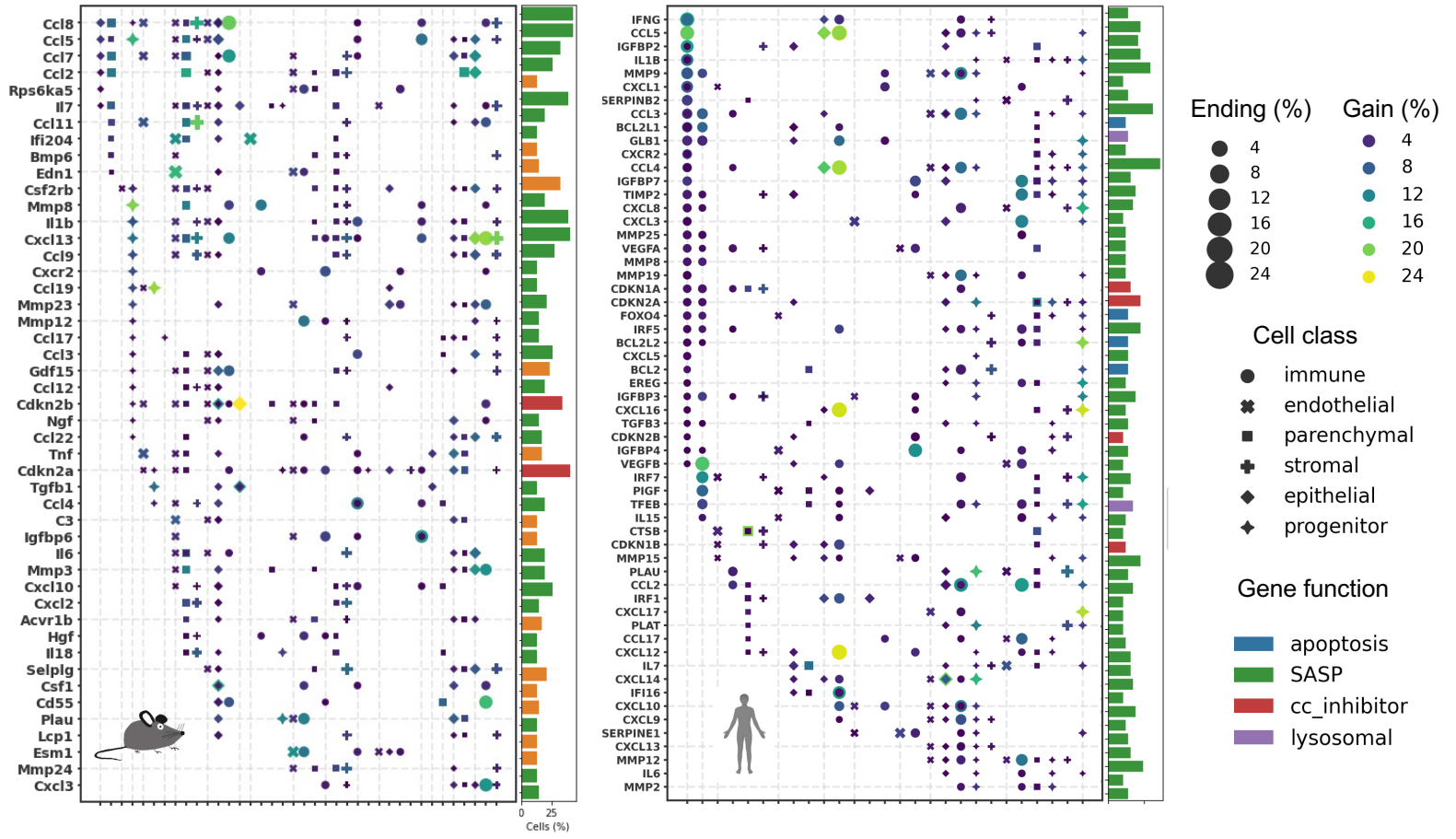


Fig 1 Known senescence markers are cell-type-specific and poorly characterize in vivo cellular senescence.

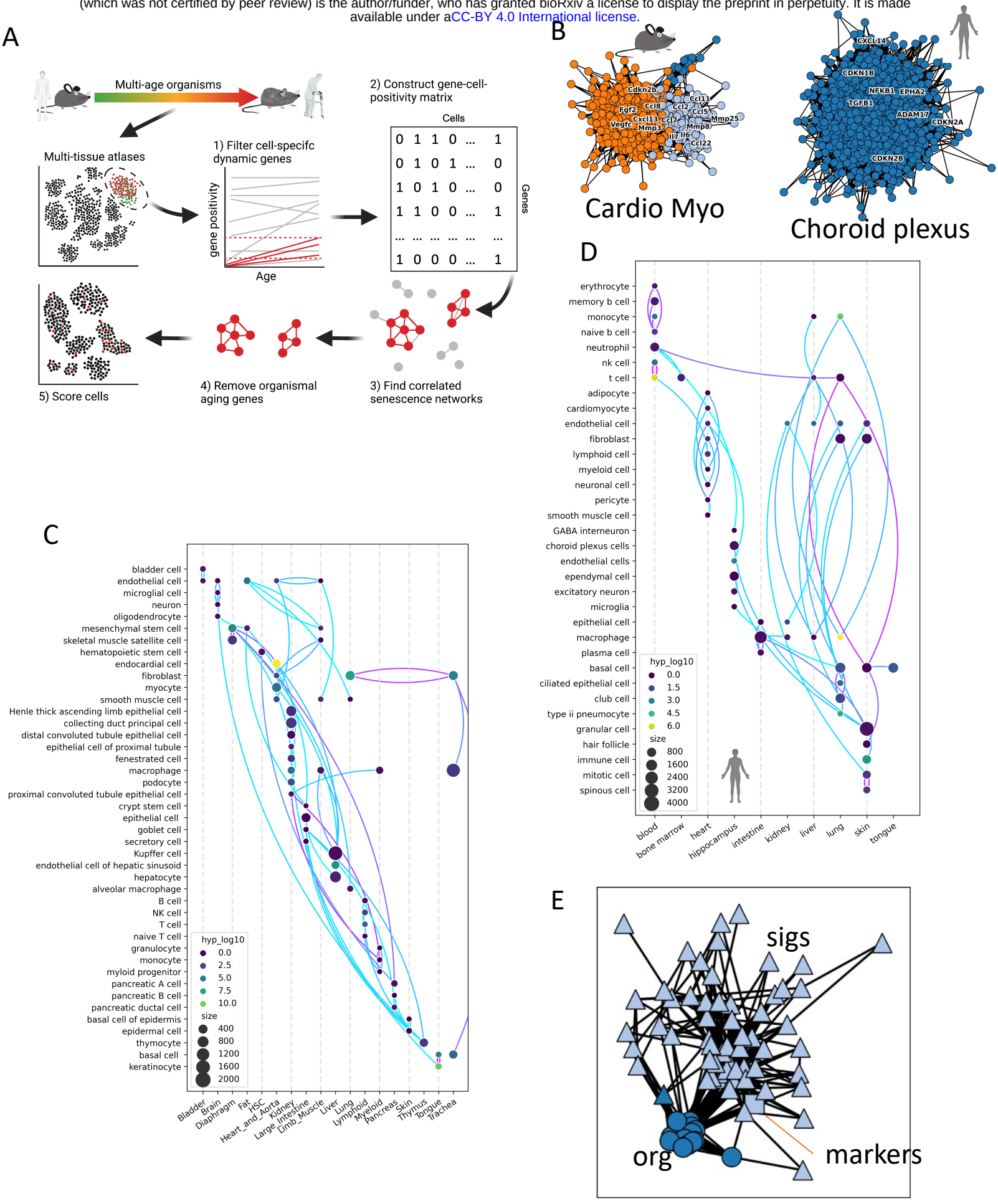


Fig 2 *De novo* cell-type specific signatures derived algorithmically from single-cell transcriptomes.

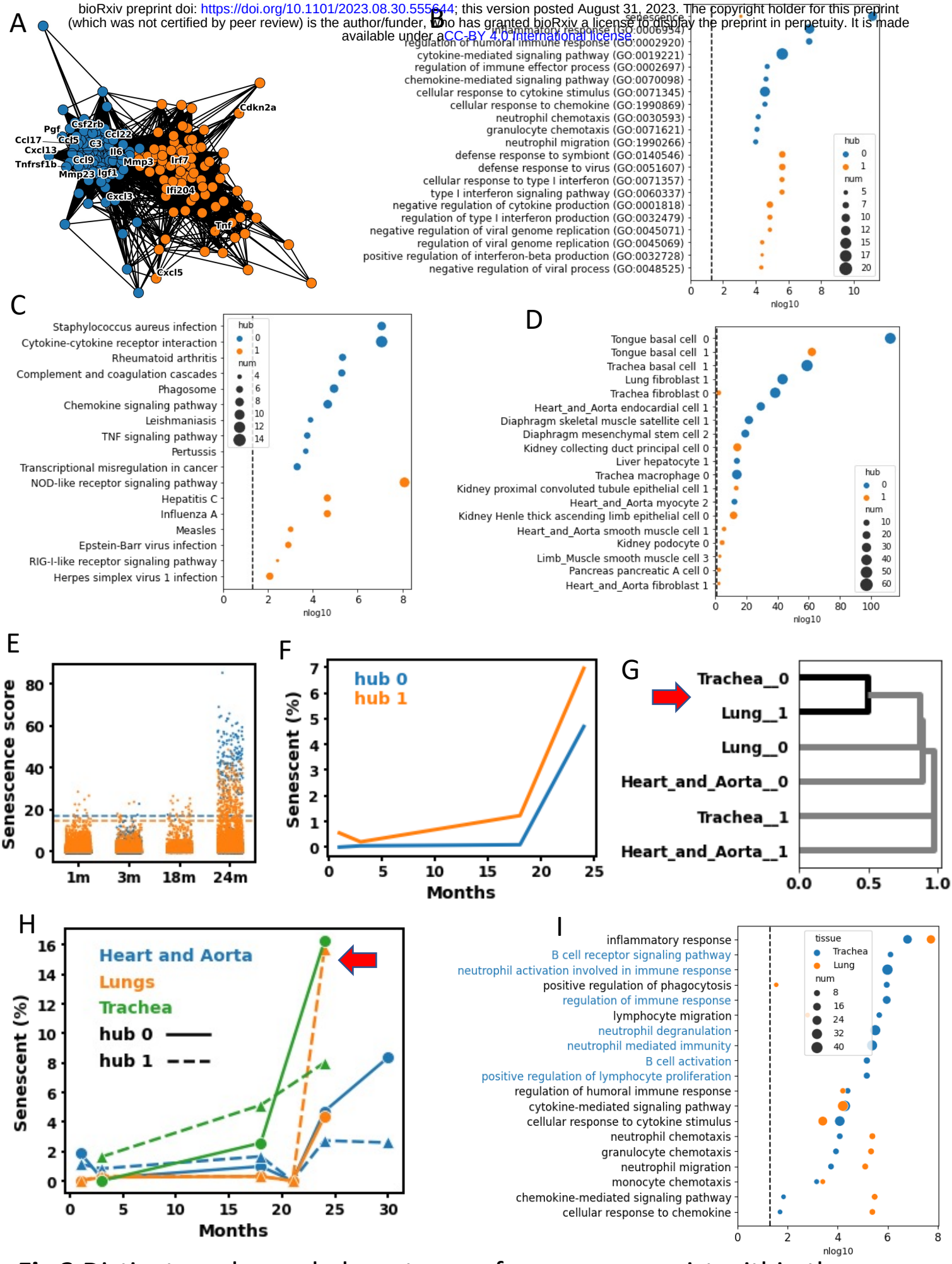


Fig 3 Distinct modes and phenotypes of senescence exist within the same cell populations

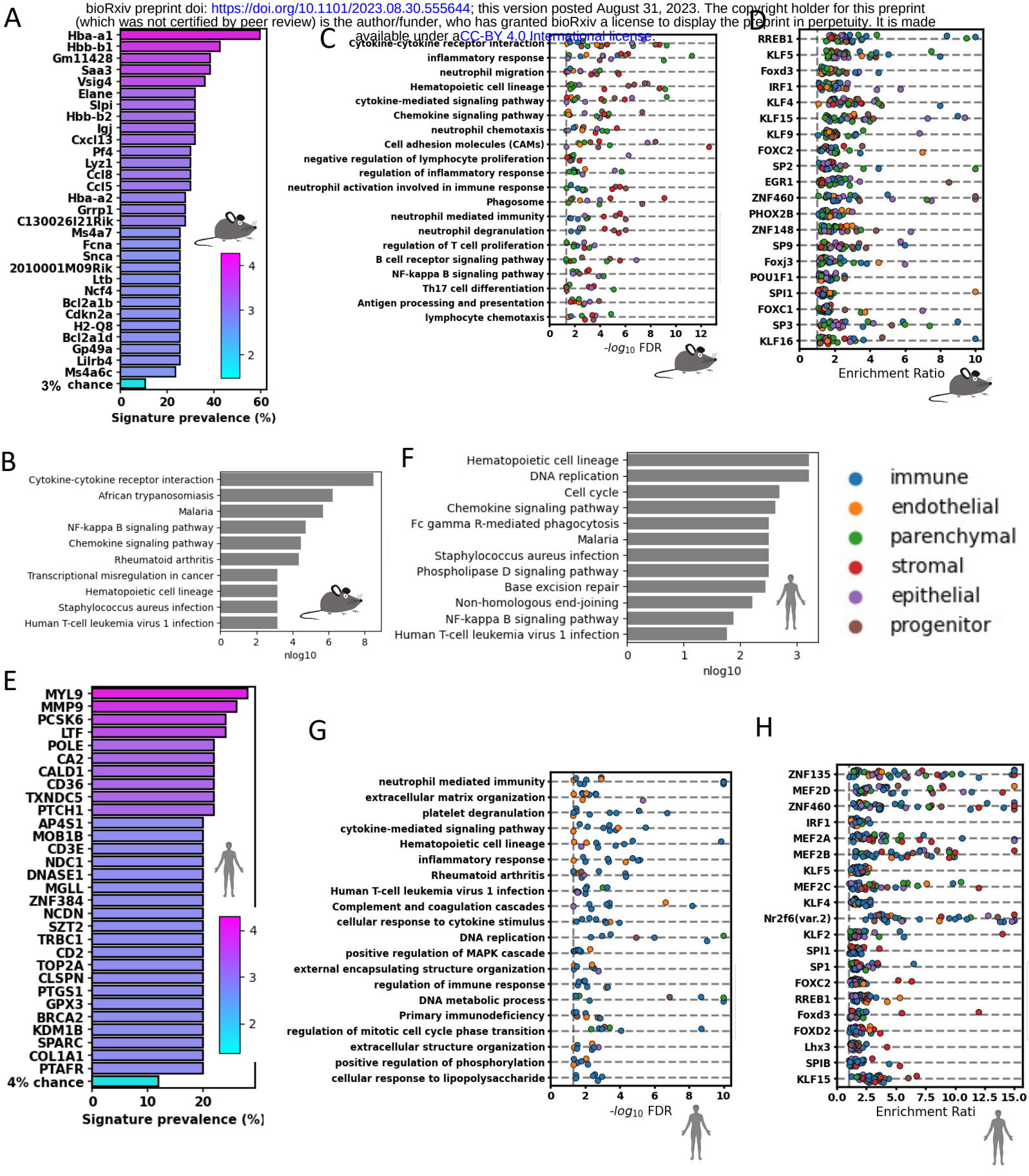


Fig 4 Cell-specific signatures are unique but share common stress response and inflammatory pathways

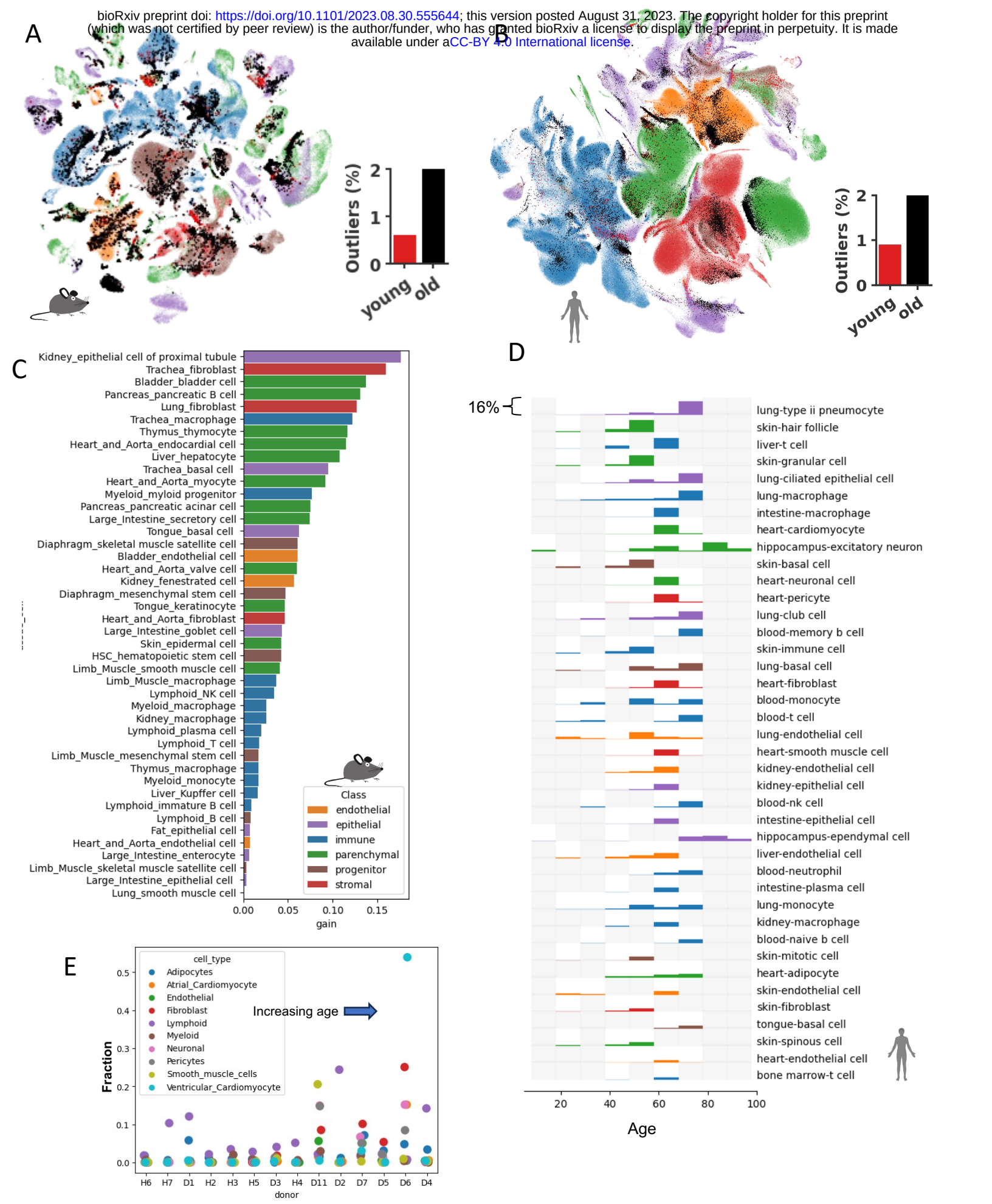


Fig 5 The cell-specific kinetics of senescent cell accumulation with organismal age

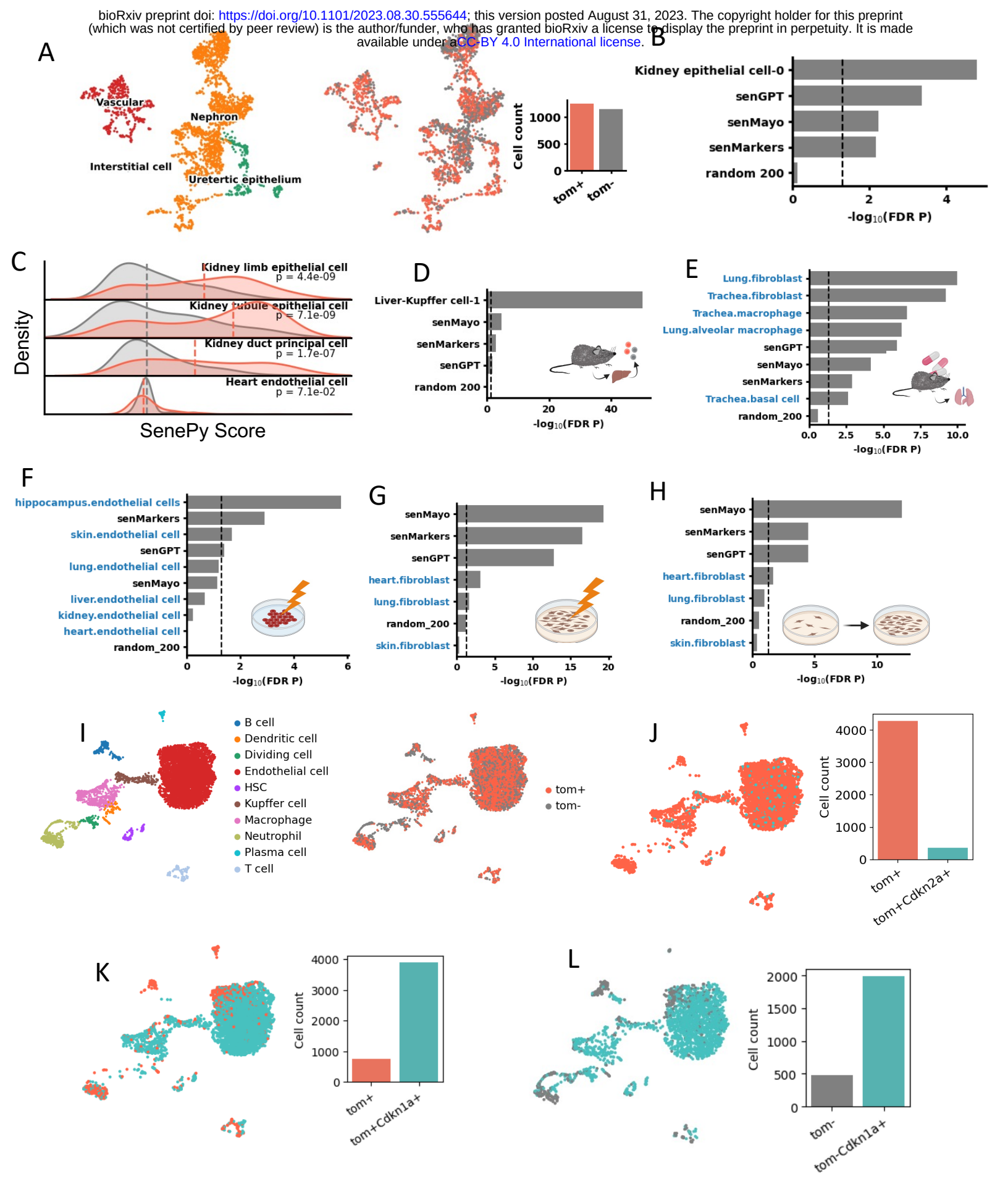


Fig 6 SenePy identifies ground-truth in vivo senescence more robustly than established markers of senescence.

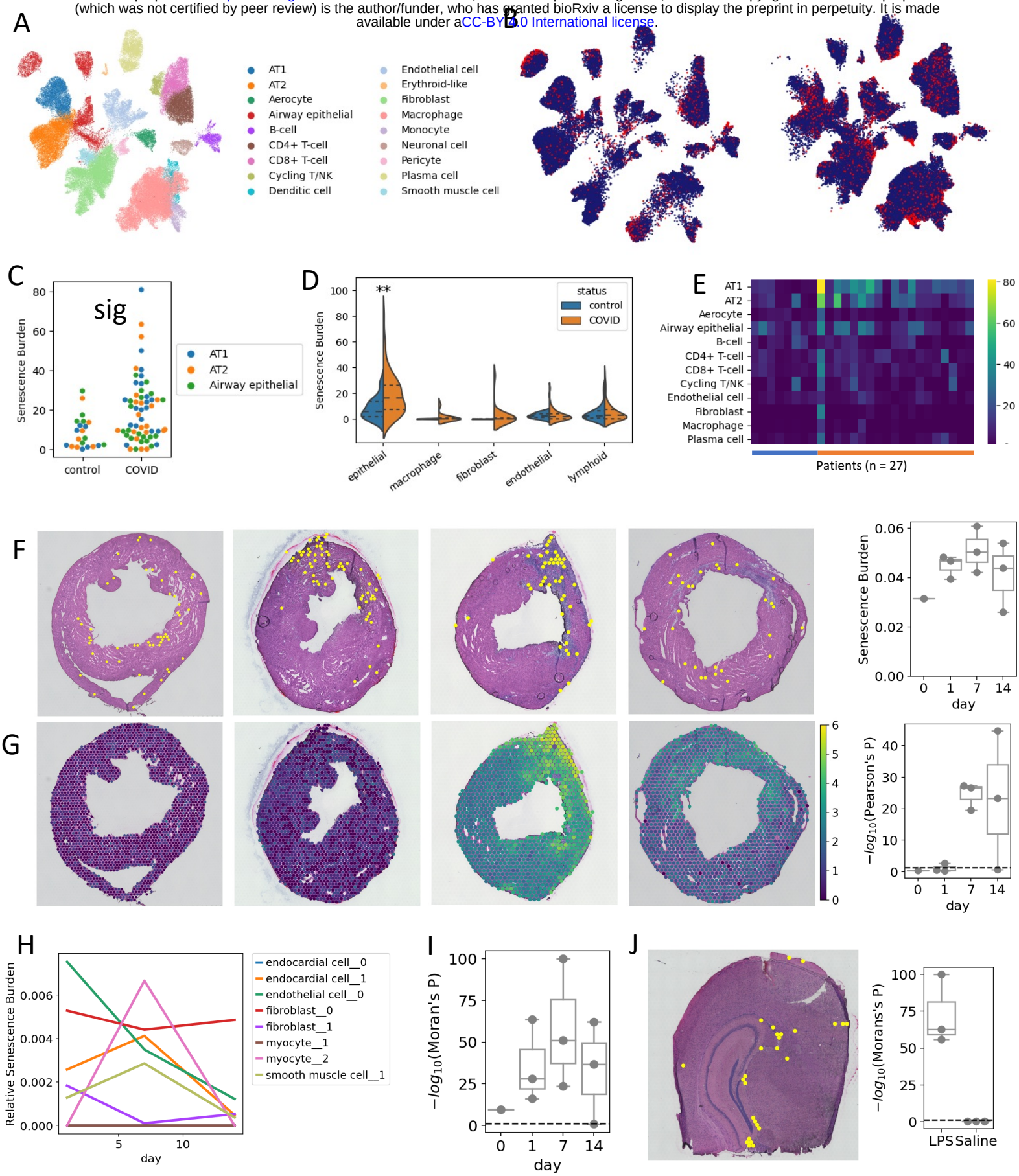
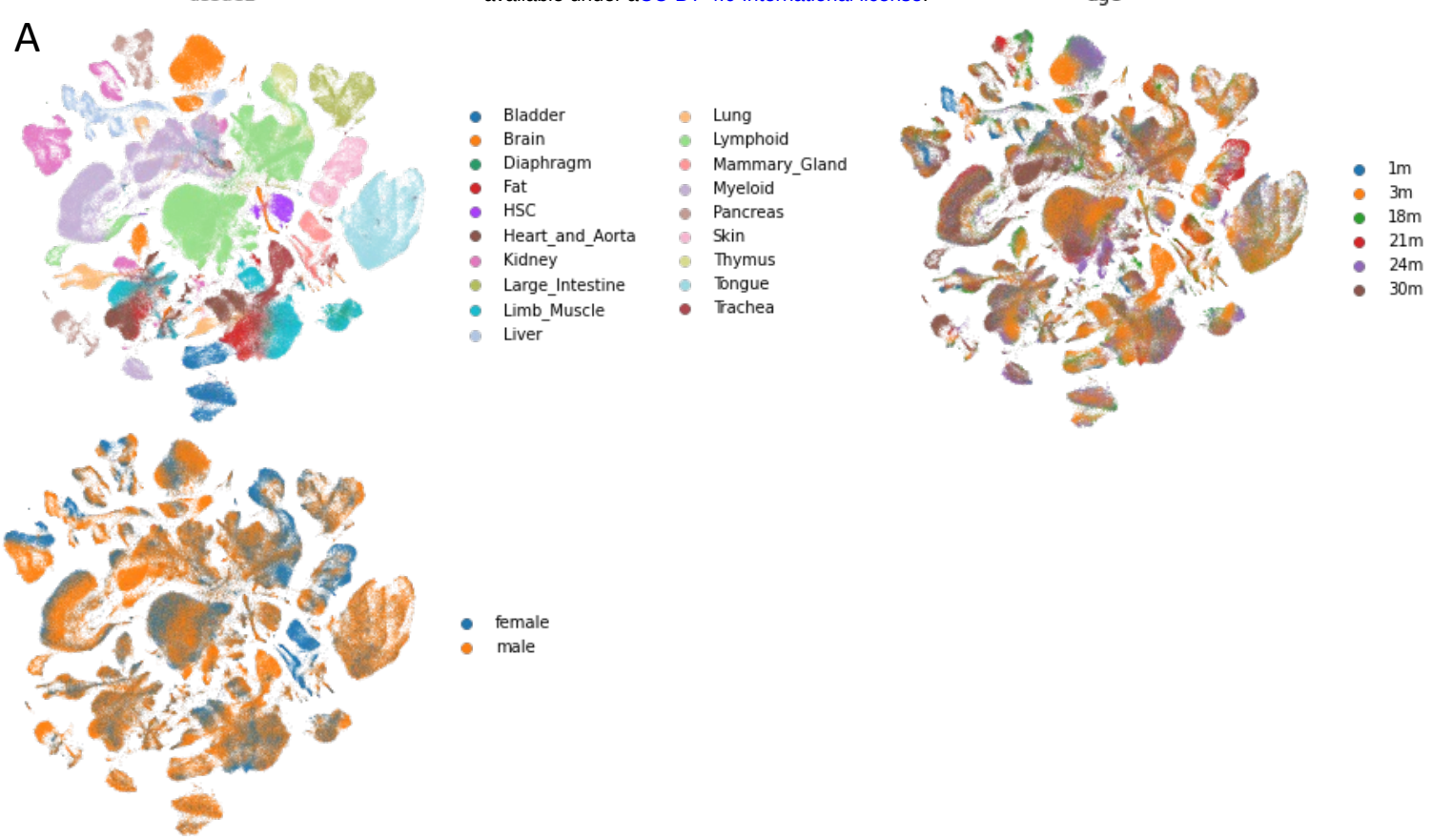
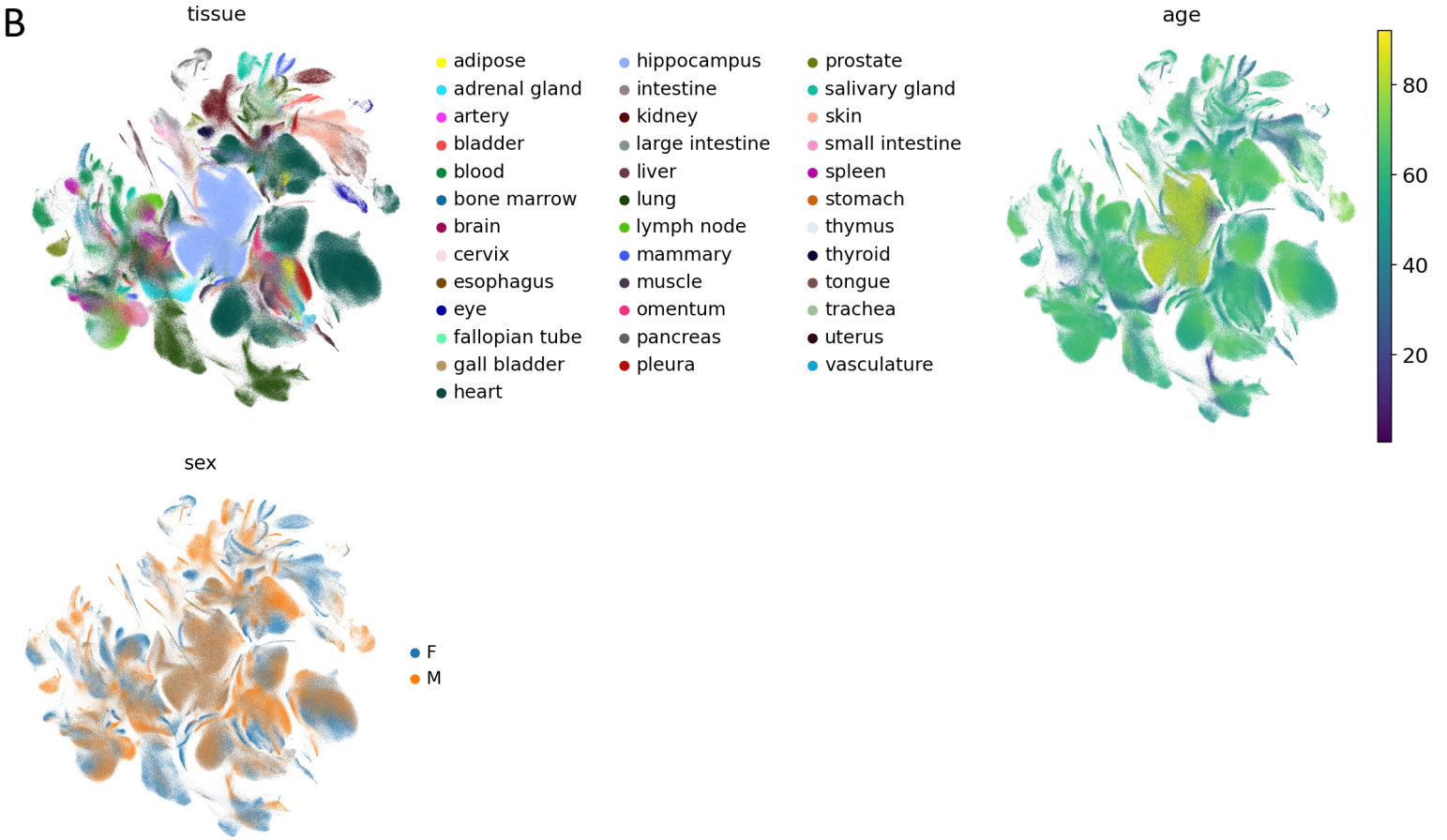
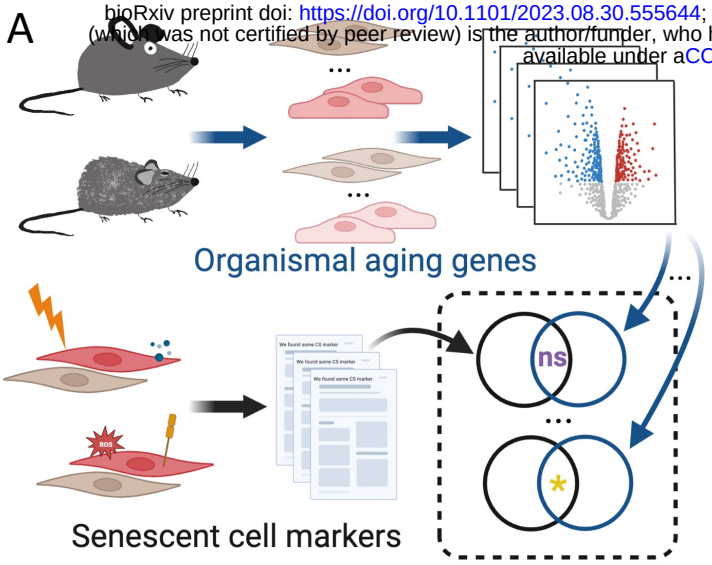


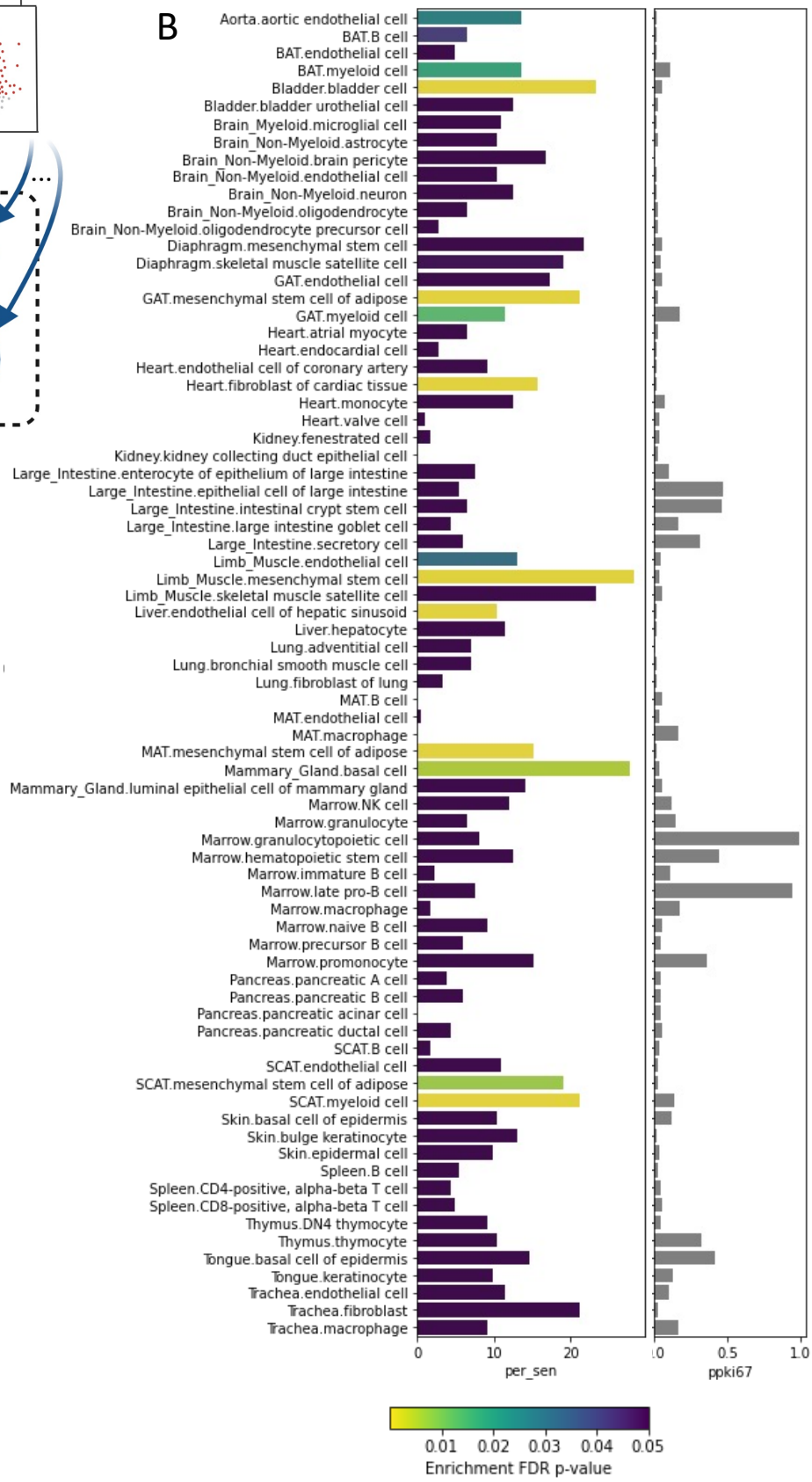
Fig 7 SenePy predicts elevated senescence burden in severe disease

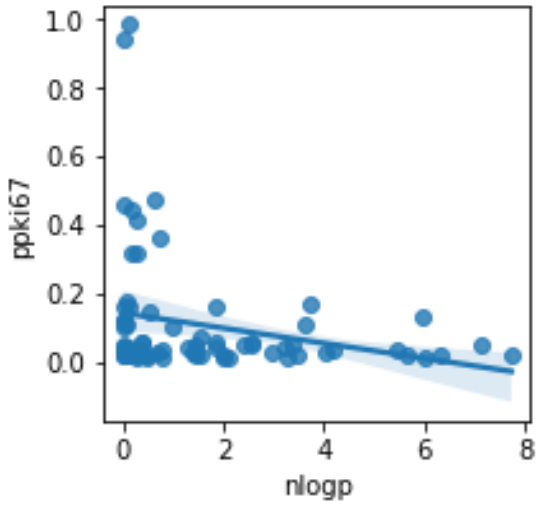




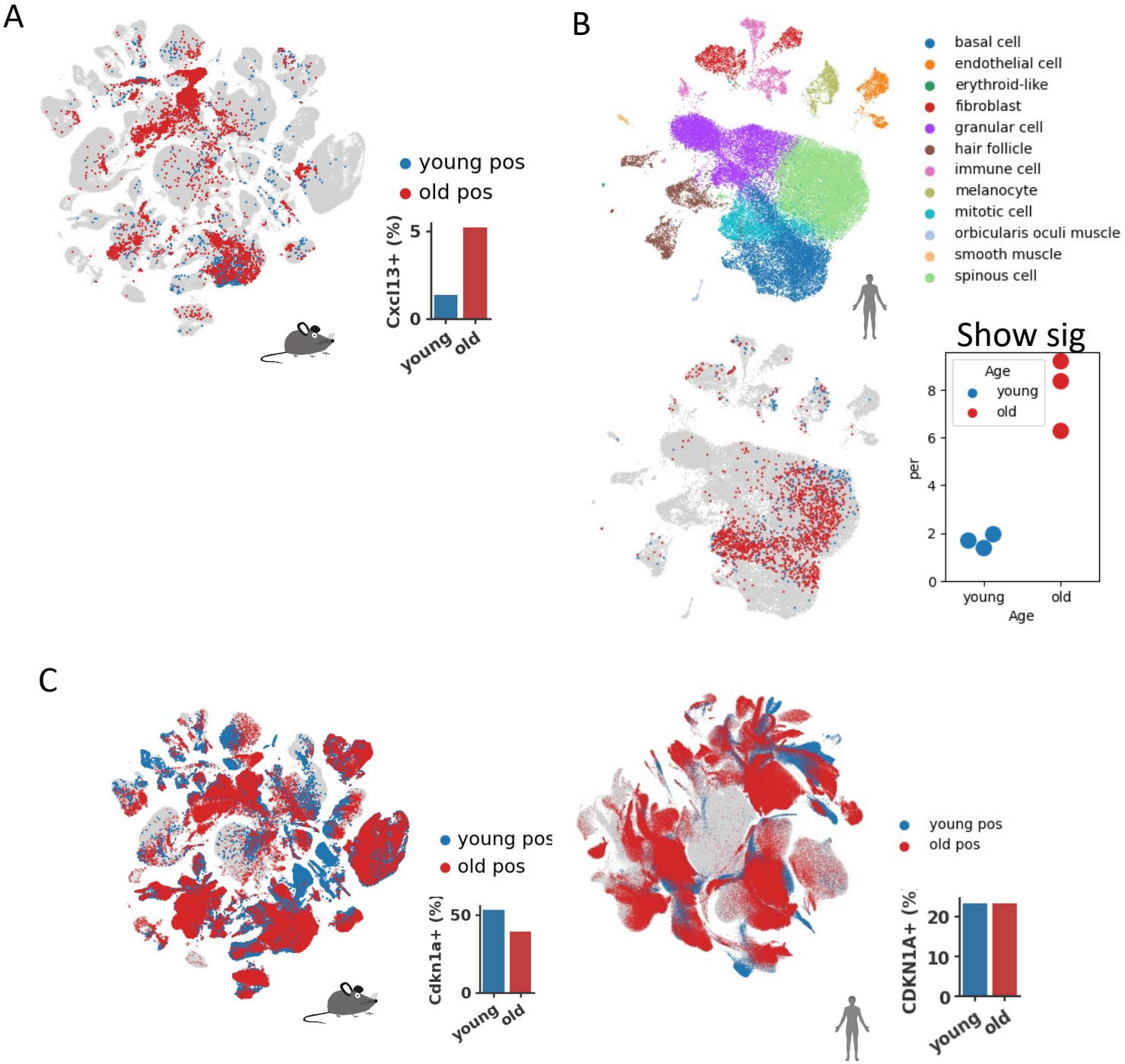
Supplemental Figure 1





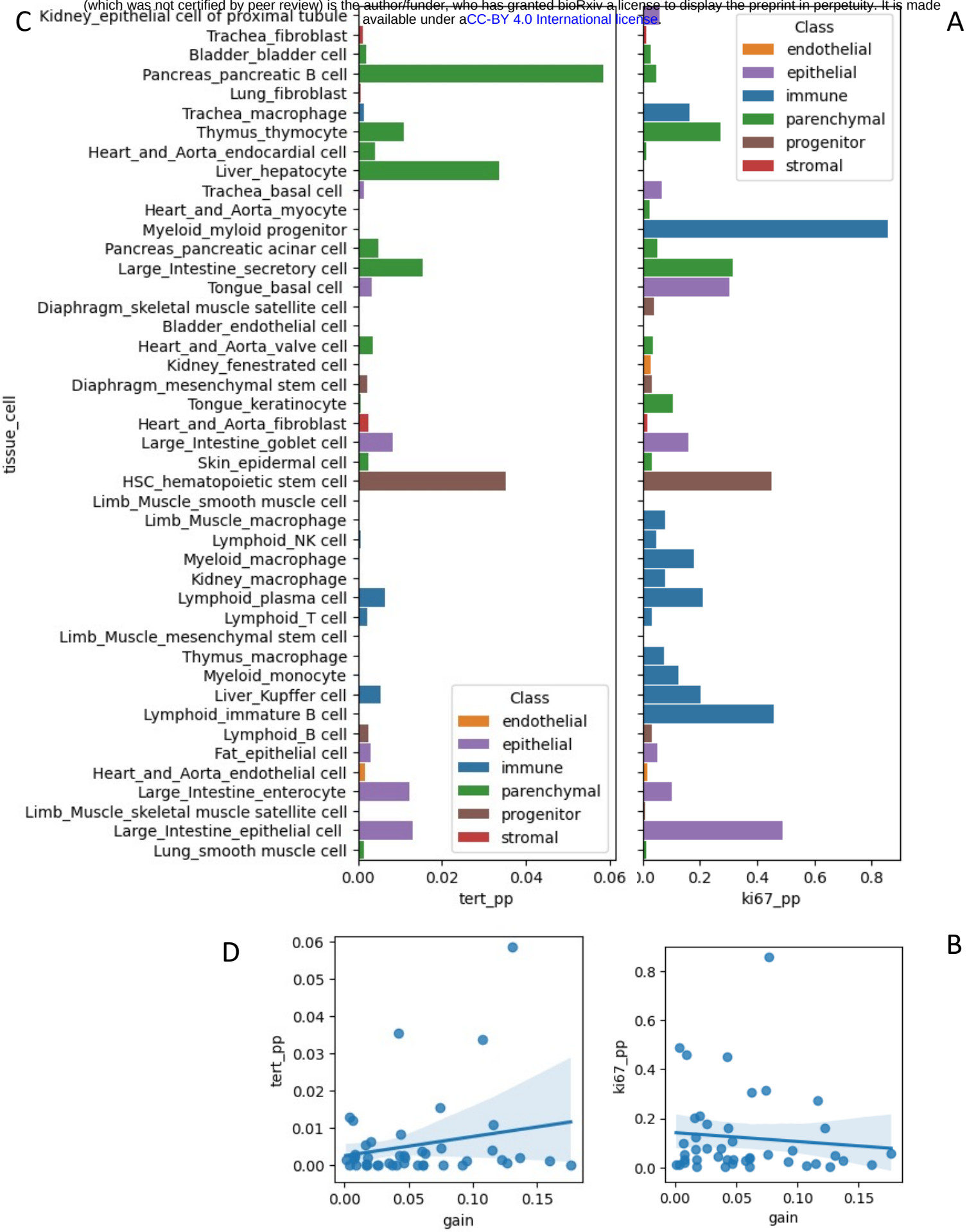


Supplemental Figure 2

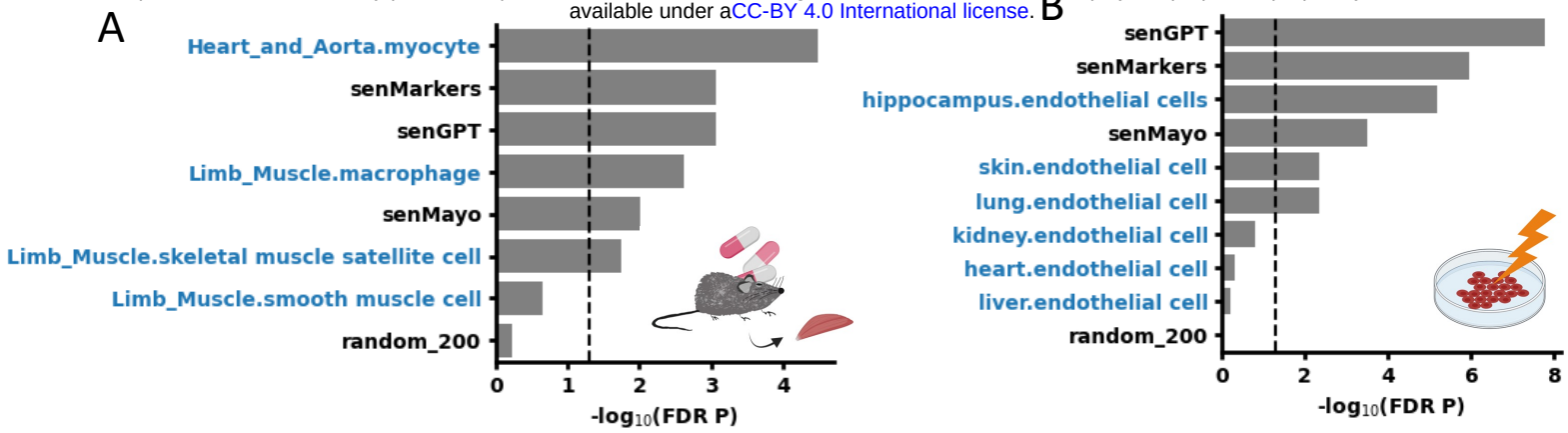


Supplemental figure 3

Supplemental Figure 4



Supplemental Figure 5



Supplemental Figure 6