Mouse Aging Cell Atlas Analysis Reveals Global and Cell Type Specific Aging Signatures

Martin Jinye Zhang[1,2*], Angela Oliveira Pisco[3*], Spyros Darmanis[3], and James Zou[1,3,4*]

[1]Department of Electrical Engineering, Stanford University, Palo Alto, 94304 USA
[2]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, 02120 USA
[3]Chan-Zuckerberg Biohub, San Francisco, 94158 USA
[4]Department of Biomedical Data Science, Stanford University, Palo Alto, 94304 USA
[*]Correspondence should be addressed to martinjzhang@gmail.com, angela.pisco@czbiohub.org and jamesz@stanford.edu

ABSTRACT

Aging is associated with complex molecular and cellular processes that are poorly understood. Here we leveraged the Tabula Muris Senis single-cell RNA-seq dataset[35] to systematically characterize gene expression changes during aging across diverse cell types in mouse. We identified aging-dependent genes in 76 tissue-cell types from 23 tissues and characterized both shared and tissue-cell-specific aging behaviors. We found that the aging-related genes shared by multiple tissue-cell types change their expression congruently in the same direction during aging in most tissue-cell types, suggesting a coordinated global aging behavior at the organismal level. We integrated the aging-related genes to construct a cell-wise aging score that allowed us to investigate the aging status of different cell types from a transcriptomic perspective. Overall, our analysis provides one of the most comprehensive and systematic characterization of the molecular signatures of aging across diverse tissue-cell types in a mammalian system.

Introduction

Aging leads to the functional decline of major organs across the organism and is the main risk factor for many diseases, including cancer, cardiovascular disease, and neurodegeneration[21,26]. Past studies have highlighted different hallmarks of the aging process, including genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication[3,21,27,41]. However, the primary root of aging remains unclear, and the underlying molecular mechanisms are yet to be fully understood.

To gain a better insight into the mammalian aging process at the organismal level, the Tabula Muris Consortium, which we are members of, created a single cell transcriptomic dataset called Tabula Muris Senis (TMS)[35]. TMS is one of the largest expert-curated single-cell RNA sequencing (scRNA-seq) datasets, containing 529,823 cells from 23 tissues and organs of male and female mice (Mus musculus). The cells were collected from mice of diverse ages, making this data a tremendous opportunity to study the genetic basis of aging across different tissues and cell types. The TMS data is organized into scRNA-seq expression of different tissue-cell type combinations (e.g., B cells in spleen) via expert annotation and clustering.

The original TMS paper focused primarily on the cell-centric effects of aging, aiming to characterize changes in cell-type composition within different tissues. Here we provide a systematic gene-centric study of gene expression changes occurring during aging across different cell types. The cell-centric and gene-centric perspectives are complementary, as the gene expression can change within the same cell type during aging, even if the cell type composition in the tissue does not vary over time.

Our analysis focused on the TMS FACS data (acquired by cell sorting in microtiter well plates followed by Smart-seq2 library preparation[34]) because it has more comprehensive coverage of tissues and cell types (Supplementary Figures 1A-B) and is more sensitive at quantifying levels of gene expression. As shown in Figure 1A, the FACS data was collected from 16 C57BL/6JN mice (10 males, 6 females) with ages ranging from 3 months (20-year-old human equivalent) to 24 months (70-year-old human equivalent). It contains 120 cell types from 23 tissues (Supplementary Figures 1A-B), totalling 164,311 cells. We also used the TMS droplet data (derived from microfluidic droplets), for those tissues for which the data was available, to further validate our findings on an additional dataset generated by a different method.

We investigated the comprehensive expression signatures of aging across tissues and cell types in the mouse. We performed systematic differential gene expression (DGE) analysis to identify aging-related genes in 76 different tissue-cell type combinations across 23 tissues (Figure 1B). In addition, we characterized both shared and tissue-cell-specific aging signatures.

Our study identified global aging genes, namely genes whose expression varies substantially with age in most (>50%) of the tissue-cell types. Interestingly, the gene expression changes are highly concordant across tissue-cell types and exhibit strong bimodality, i.e., a gene tends to be either down-regulated during aging in most of the tissue-cell types or up-regulated across the board. We leveraged this coordinated dynamic to construct an aging score based on global aging genes. This aging score informs the biological age (as opposed to the chronological age) and is correlated with the cell turnover rate of different tissue-cell types from a transcriptomic perspective. The aging score reflects the aging-related genomic changes and captures different information from the commonly-used DNA methylation features[12,14]. In addition, our work investigates the biological age at the cell-type level, distinguishing itself from virtually all previous works which focus on predicting the individual-wise biological age[11,12,14,32,33]. Overall, our analysis highlights the power of scRNA-seq in studying aging and provides a comprehensive catalog of aging-related gene signatures across diverse tissue-cell types.

Results

Identification of aging-related genes

We performed differential gene expression analysis to identify aging-related genes for 76 tissue-cell types with a sufficient sample size; each tissue-cell type is required to have more than 100 cells in both young (3m) and old (18m, 24m) age groups. For each tissue-cell type, we tested if the expression of each gene was significantly related to aging using a linear model treating age as a numerical variable while controlling for sex. We applied an FDR threshold of 0.01 and an age coefficient threshold of 0.005 (corresponding to $\sim$10% fold change). For details, please refer to the differential gene expression analysis subsection in Methods.

As shown in Figure 1B, the number of significant age-dependent genes per tissue-cell type ranges from hundreds to thousands. Moreover, tissue-cell types with a higher number of cells (right panel in Figure 1B) tend to have more discoveries, likely due to their higher detection power. Interestingly, 62 out of 76 tissue-cell types have more down-regulated aging genes than up-regulated aging genes, indicating a general decrease in gene expression over aging. Cells from older mice (18m, 24m) were actually sequenced deeper, yet they still had much fewer expressed genes and lower expression in the expressed genes (Supplementary Figure 1C). This suggests that the decreasing expression phenomenon is genuine and unlikely to be confounded by sequencing depth or size factor normalization.

In contrast to the overall trend of decreasing expression, many immune cells have a higher number of up-regulated genes during aging, including B cells, T cells, as well as liver endothelial cells of the hepatic sinusoid, which are known to have the ability to become activated in response to diverse inflammatory stimuli[20]. This observation is particularly interesting given the strong link between the aging process and the immune system, as suggested by previous literature[27,35].

Tissue-cell level global aging markers

We visualized all discovered aging genes (significant in $\geq$1 tissue-cell type) in Figure 2A, where the color indicates the number of genes. The x-axis shows the proportion of tissue-cell types (out of 76 tissue-cell types) where the gene is significantly related to aging, while the y-axis shows the proportion of tissue-cell types where the gene is up-regulated. The visualization makes it clear that there are more down-regulated aging genes than up-regulated aging genes, consistent with the number of up/down-regulated discoveries, as shown in Figures 1B. Moreover, perhaps more strikingly, a bimodal pattern is apparent for aging-dependent genes. Genes tend to have a consistent direction of change during aging across different tissue-cell types—the expression either increases across most of the tissue-cell types or it decreases across the board. A similar bimodality was also observed recently in mouse brain cells[44].

We define *global aging genes* to be genes whose expression changes significantly with age in more than half of tissue-cell types. The bimodality pattern is especially striking in these genes. We identified 292 global aging genes in total (Supplementary Table 1), among which 93 are consistently up-regulated (in >80% of tissue-cell types) and 169 are consistently down-regulated (in >80% tissue-cell types). Only 30 global aging genes have an inconsistent direction of regulation in different tissue-cell types (up-regulated in 20%-80% tissue-cell types). We also used a heatmap to visualize all global aging genes in Figure 2B, where the pattern of the consistent direction of change becomes more apparent (see a larger version in Supplementary Figure 2).

We visualized the 20 top global aging genes in Figure 2C; these are genes that showed the most substantial change during aging. Many of these have been previously shown to be highly relevant to aging. For example, the down-regulation of *Lars2* has been shown to result in decreased mitochondrial activity and increase the lifespan for *C. elegans*[19]. On the other hand, *Jund* is a proto-oncogene known to protect cells against oxidative stress and its knockout may cause a shortened lifespan in mice[18]. Moreover, we also found *Rpl13a*, a key component of the GAIT (gamma interferon-activated inhibitor of translation) complex which mediates interferon-gamma-induced transcript-selective translation inhibition in inflammation processes[23], to be up-regulated. As a negative regulator of inflammatory proteins, *Rpl13a* contributes to the resolution phase of the inflammatory response, ensuring that the inflamed tissues are completely restored back to normal tissues, which also contributes to preventing cancerous growth of the injured cells caused by prolonged expression of inflammatory genes[45].

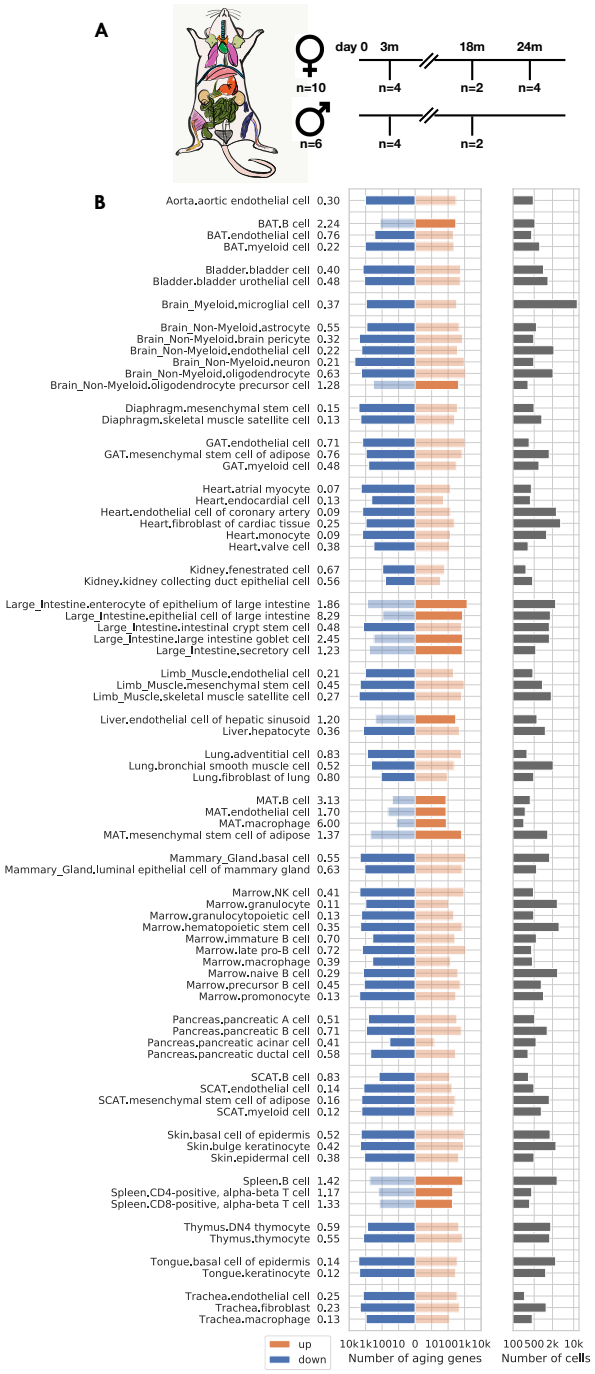

Figure 1. Analysis overview. A: Sample description. The data was collected from 16 C57BL/6JN mice (10 males, 6 females) with ages ranging from 3 months (20-year-old human equivalent) to 24 months (70-year-old human equivalent). B: Significant aging-dependent genes in all 76 tissue-cell types. The left panels show the number of aging genes (discoveries) for each tissue-cell type, broken down into the number of up-regulated genes (orange) and the number of down-regulated genes (blue), with the numbers on the right showing the ratio (up/down). The higher number of the two bars is shown in a solid color, and the lower one is shown in a transparent color. We can see most tissue-cell types (62/76) have more down-regulated aging genes. The right panels show the number of cells sequenced for each tissue-cell type.

Many of the top global aging genes were also identified in previous studies or correspond to important aging-related biological processes. For example, *Cat*, *Grn*, *Gpx2*, *Apoe* were identified as aging-related genes in mouse in a previous

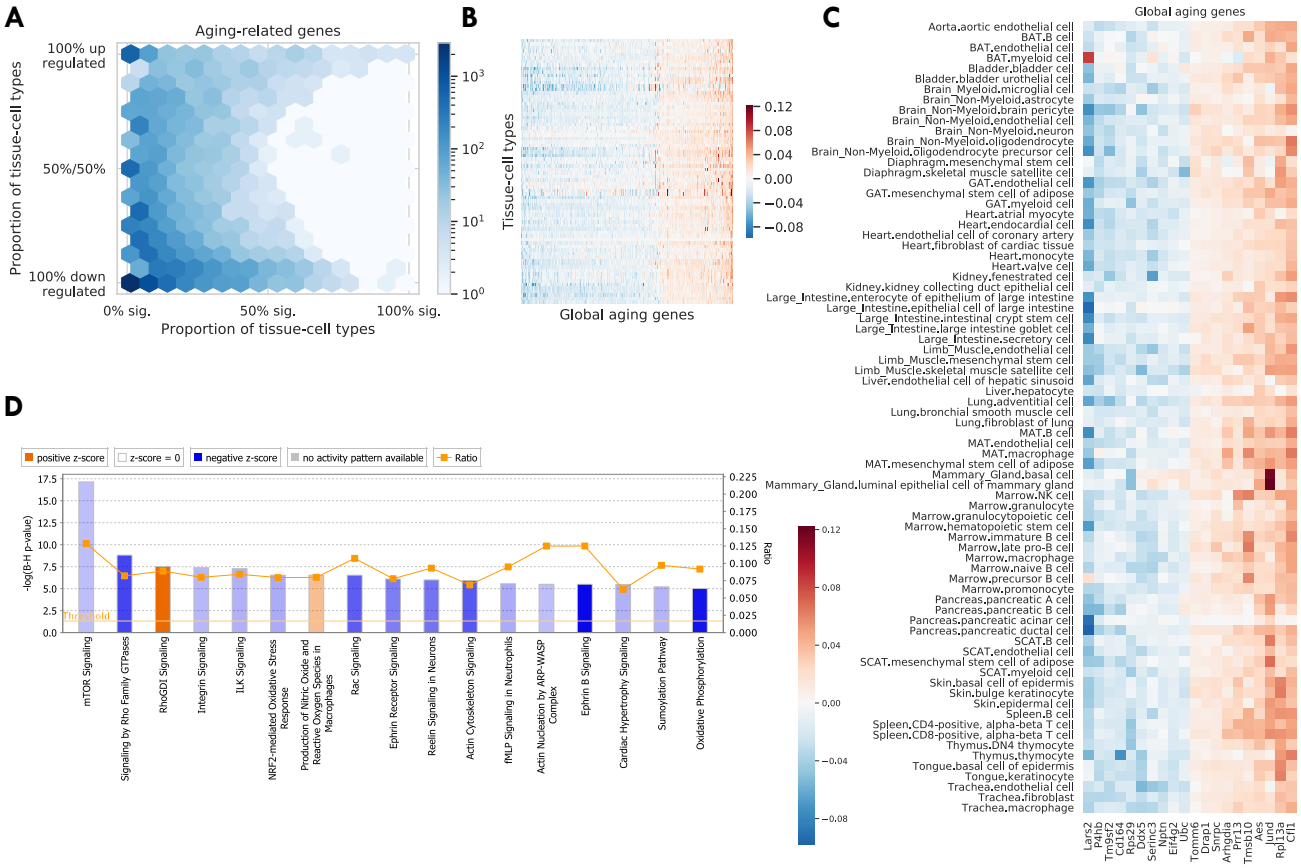

Figure 2. Tissue-cell level global aging genes. A: Tissue-cell level aging related genes. The color indicates the number of genes. The x-axis shows the proportion of tissue-cell types (out of all 76 tissue-cell types) where the gene is significantly related to aging, while the y-axis shows the proportion of tissue-cell types where the gene is up-regulated. B: Heatmap of age coefficient for all global aging genes. C: Heatmap of age coefficient for the top 20 global aging genes. For Panels B-C, blue/red represents down-/up- regulation. Also, 20 times the age coefficient roughly corresponds to the log fold change from young (3m) to old (24m). D: Top pathways associated with the global aging genes. The negative z-score (blue) means the pathway is predicted to be down-regulated and the positive z-score (red) means the opposite. The ratio represents the proportion of pathway genes that are also global aging genes.

study[39] (Supplementary Figure 3A). *App*, *Ctnnb1*, *Mapk1*, *Rac1*, *Arf1*, *Junb* are related to senescence, a hallmark of aging[3] (Supplementary Figure 3B). Many other global aging genes correspond to transcription factors (Supplementary Figure 4A), eukaryotic initiation factors (Eif genes, Supplementary Figure 4B), and ribosomal protein genes (Rpl/Rps genes, Supplementary Figure 5). Additional annotations are provided in Supplementary Figure 6.

Next, we performed a gene pathway analysis on the entire set of 292 global aging genes using the Ingenuity Pathway Analysis software (IPA)[16]. As shown in Figure 2D, the global aging genes are enriched for biological processes, including protein synthesis, apoptosis, cell cycle and tissue development and function. In general, we observed decreased activities of these pathways. Of note is the implication of the mTOR pathway, highlighted in Supplementary Figure 7, which is known to be related to aging[13,30,42]. mTOR and translation initiation go hand in hand[17] and it is therefore not surprising that we also observe a significant correlation of aging and translation. mTOR down-regulation has been shown to promote longevity[38] and it is interesting that when comparing old with young mice the global aging genes we identify point towards a collective down-regulation of this pathway.

Aging score based on global aging genes

Following the analysis of global aging genes, we next leveraged these marker genes to characterize the holistic aging status of individual cells. We use the global aging genes to define an aging score for each cell to explore how the aging process varies between different tissue-cell types in the same animal. More precisely, the aging score for a cell is defined to be the average

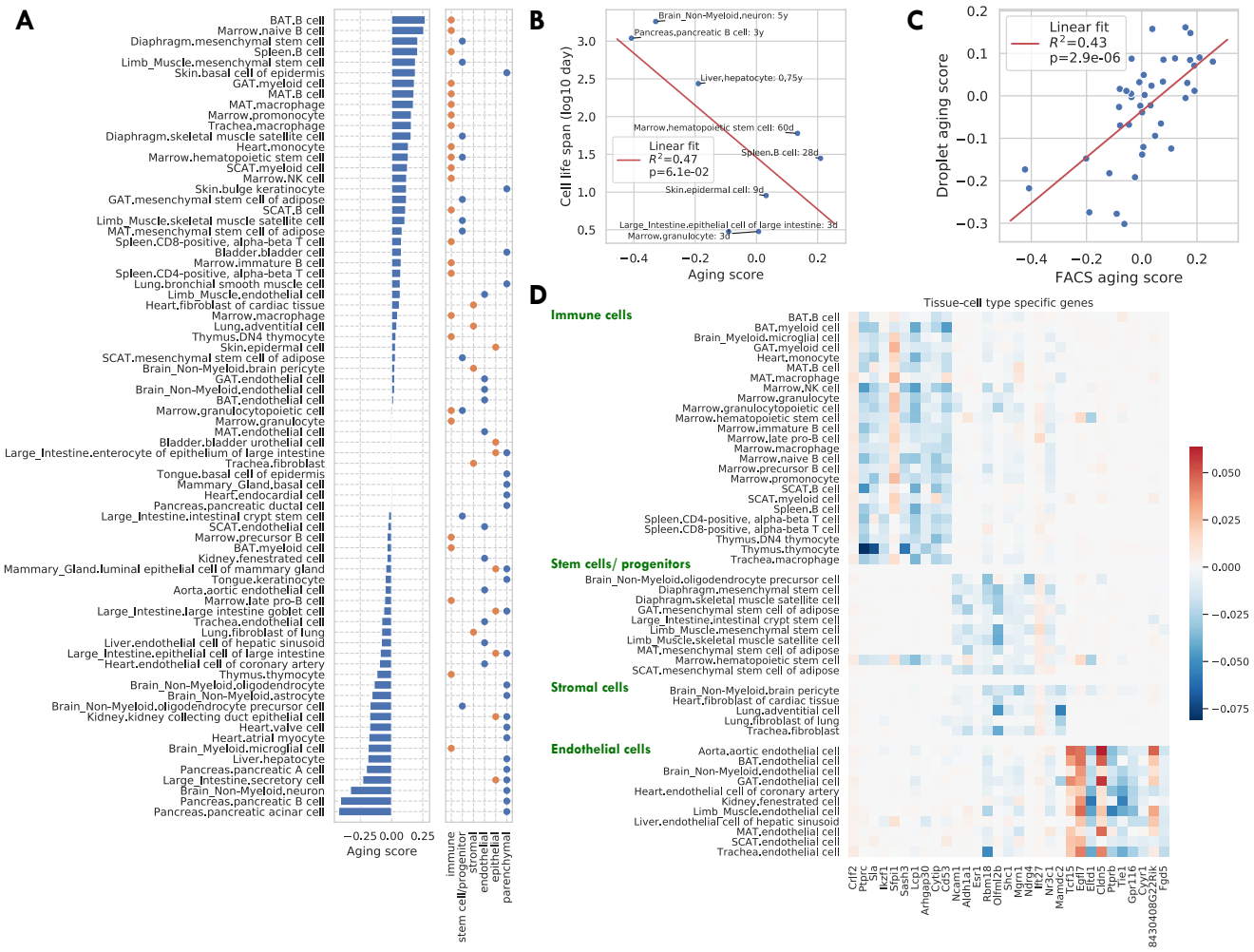

Figure 3. Tissue-cell-specific aging process. A: Aging score for different tissue-cell types. The left panel shows the aging score while the right panel shows the functional categories of the tissue-cell types. B: Correlation of the aging score with the average life span for a subset of cell types. C: Aging scores computed using both FACS and droplet data are consistent. D: age coefficients for aging genes that are enriched in different cell categories. Blue/red represents down-/up- regulation and 20 times the age coefficient roughly corresponds to the log fold change from young (3m) to old (24m). Epithelial cells and parenchymal cells are excluded because we did not find aging-dependent genes that are specific for these two categories.

expression of the up-regulated global aging genes (up-regulated in >80% tissue-cell types) minus the average expression of the down-regulated global aging genes (down-regulated in >80% tissue-cell types), while adjusting for the background gene expression level. For each tissue-cell type, we calculate the aging score by averaging over all cells from the tissue-cell type, and regressing out sex and the chronological age. See the aging score subsection in Methods for more details.

Intuitively, a larger aging score suggests the corresponding cells could be molecularly more sensitive to aging compared to other cells in the same animal. As shown in Figure 3A, immune cells and stem cells are amongst the cell types with higher aging scores, and they are also known to undergo substantial changes during aging. Aging of the immune system is commonly linked to the impaired capacity of elderly individuals to respond to new infections[25]. Also, adult stem cells are critical for tissue maintenance and regeneration, but the increased incidence of aging-related diseases has been associated with a decline in stem cell function[6] . On the other hand, parenchymal cells like pancreatic cells, neurons, and hepatocytes have lower aging scores. This could be due to the fact that their highly specialized functions change less during the aging process. Please also refer to Supplementary Figure 8 for the average aging score of all cell functional categories and Supplementary Table 3 for the cell functional annotations.

Interestingly, our aging score shows a strong negative correlation with the average life span of different cell types, as shown in Figure 3B. We curated the cell life span data from literature[24] for a subset of tissue-cell types where such data for mice or

rats is available (Supplementary Table 3); it is generally hard to measure the average life span for most cell types *in vivo*, so this comparison is limited to those cell types for which there is data. Among tissue-cell types with lower aging scores, neuron does not regenerate, pancreatic β-cell has a very low turnover rate of 0.07% per day[40], and liver hepatocyte has a life span of 0.5-1 year[22]. On the other hand, the life spans are shorter for tissue-cell types with a higher aging score, e.g., hematopoietic stem cell ($\sim$ 60 days[4]) and spleen B cell ($\sim$ 9 days[9]). Tissue-cell types known to have an extremely high turnover rate also have relatively higher aging scores, including skin epidermal cell ($\sim$9 days[15]), large intestine epithelial cell ($\sim$3 days[36]), and granulocyte ($\sim$3 days[10]). One possible explanation for the negative correlation is that the global aging genes tag biological processes related to cell proliferation, development, and death, which are more active in cell types with a higher turnover rate. This strong correlation is also consistent with the intuition that cells that have undergone more divisions could have molecular memories for being "older".

As validation, we repeated the analysis and computed the aging score using the droplet data; we observed a high concordance between the aging score computed using both data (p-value<1e-5, Figure 3C), indicating the robustness of the aging score. We also considered various potential confounding factors like the number of discoveries for each tissue-cell type, which are shown to not have a significant effect on the aging scores (Supplementary Figure 11). Among related works, Peters *et al.* identified 1497 aging-related genes using human whole blood samples and also provided a transcriptomic age predictor using all genes[32]. We found that these 1497 genes did not significantly overlap with our global aging genes (p-value 0.16), but they did significantly overlap with the immune-specific aging genes (p-value 8.3e-8, see below for details on the immune-specific aging genes). It is not surprising because Peters *et al.* identified those genes using the whole blood sample, which is most relevant to immune cells. This further highlights the need to systematically examine the aging signatures across different tissue-cell types as we did here. We also computed the transcriptomic age for each tissue-cell type using the predictor in Peters *et al.* and we found that it was strongly correlated with our aging score when we divided the tissue-cell types into adaptive immune cells (p-value 5.4e-4), innate immune cells (p-value 0.08) and other cells (p-value 1.8e-5) (one-sided Fisher's exact test, Supplementary Figure 9). Moreover, we compared our aging score with DNA mythelation age[12] and did not observe a significant correlation (Supplementary Figure 10), consistent with previous observations that DNA mythelation age and transcriptomic age capture different aspects of biology[14,32].

Tissue-cell-specific aging genes

As our analysis so far focused on global aging genes, we next consider tissue-cell specific aging genes. We first divided tissue-cell types into 6 categories based on their functionality, i.e., immune cells, stem cells/progenitors, stromal cells, parenchymal cells, endothelial cells, and epithelial cells (Supplementary Table 3). Then we identified genes that were over-represented in each cell category using Fisher's exact test with an FDR threshold of 0.25. We consider a gene to be over-represented in a cell category if such a cell category is enriched with tissue-cell types where the gene is significantly related to aging. We found many cell-category-specific genes for immune cells (204), stem cells (197), and endothelial cells (104), but we only found one gene specific to stromal cells (*Mamdc2*) and no genes for epithelial or parenchymal cells (Supplementary Table 2).

A subset of top cell-category-specific aging genes is shown in Figure 3D, where the tissue-cell types are ordered by their functional categories. Perhaps not surprisingly, marrow hematopoietic stem cell was also similar to immune cells. As the literature suggests, these genes are highly relevant to their corresponding cell functionalities[29]. For example, *Crlf2* encodes a member of the type I cytokine receptor family, which is a receptor for thymic stromal lymphopoietin (TSLP); it controls processes such as cell proliferation and development of the hematopoietic system[31]. *Ncam1* encodes a cell adhesion protein which is a member of the immunoglobulin superfamily and is involved in cell-to-cell interactions as well as cell-matrix interactions during development and differentiation, specifically in the nervous system. *Tcf15* encodes a protein that is found in the nucleus and may be involved in the early transcriptional regulation of patterning of the mesoderm.

We further used IPA[16] to predict the upstream regulators for the cell-category-specific aging genes for immune cells, stem cells, and endothelial cells, respectively. The results are shown in Supplementary Figure 12. For immune-specific genes, we found that *IL10*, *IL4* and *STAT6* were predicted to be down-regulated with age, in line with known trends towards pro-inflammatory phenotype acquisition with age[28]. For stem-cell-specific genes, we found that extracellular matrix (ECM) associated genes to be significantly changing with age. For example, matrix metalloproteases, which have a role in controlling the proinflammatory response and also have been implicated in neurological disease progression[8], were down-regulated. As for endothelial-specific genes, we found that *Vegf*/*VEGFA*, a major pro angiogenic factor, was predicted to be inhibited with age, along with *CD36* and *PECAM1*. *Vegf* inhibition has been previously linked with aging[1] and altogether this finding suggests that endothelial function is impaired[5].

Tissue level analysis and validation

We performed similar analyses at the tissue level in order to compare and assess the robustness of our tissue-cell level findings. We carried out a DGE analysis for every tissue by pooling all cells from different cell types in the tissue. The number of discoveries for each tissue is shown in Figure 4A. 20 out of 23 tissues had more down-regulated aging genes while there were

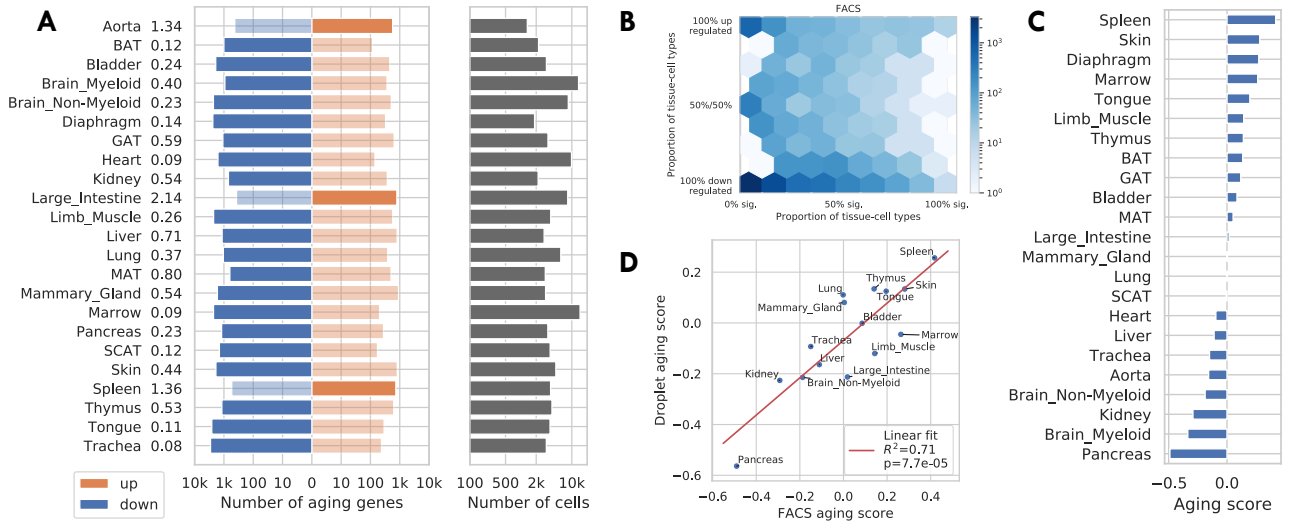

Figure 4. Tissue level analysis as validation. A: Number of discoveries for each tissue. The left panels show the number of discoveries for each tissue, broken down into the number of up-regulated genes (orange) and the number of down-regulated genes (blue), with the numbers on the right showing the ratio (up/down). The higher number of the two bars is shown in a solid color and the lower one is shown in a transparent color. We can see most tissues (20/23) have more down-regulated aging genes. The left panels show the number of cells for each tissue. B: Tissue level aging related genes. The color codes the number of genes. The x-axis shows the proportion of tissue types (out of 23 tissue types) where the gene is significantly related to aging, while the y-axis shows the proportion of tissues where the gene is up-regulated. C: FACS tissue aging score. D: Comparison between FACS tissue aging score and droplet tissue aging score.

more up-regulated aging genes in aorta, large intestine, and spleen, consistent with the tissue-cell level analysis in shown Figures 1B-C.

Figure 4B summarizes the DGE findings of all of the genes. The color indicates the number of genes; the x-axis shows the proportion of tissues (out of 23 tissues) where the gene is significantly related to aging, and the y-axis shows the proportion of tissues where the gene is up-regulated. Similar to the tissue-cell level analysis, we observed a bimodal pattern, and the direction of regulation becomes more consistent as the aging gene is shared by more tissues. We also considered the mouse tissue bulk data from a companion study to TMS[37] (see Supplementary Figure 1B for the data description) and observed a similar bimodal pattern (Supplementary Figure 13A). We used the threshold of 0.8 for selecting tissue global aging genes and found 131 such genes (Supplementary Figure 14, Supplementary Table 1), out of which 114 were also tissue-cell level global aging genes that we analyzed before, indicating a strong consistency (p-value 1.1e-199 via one-sided Fisher's exact test).

We next computed a tissue-level aging score for each tissue using the tissue-level global aging genes. As shown in Figure 4C, the tissue aging score is qualitatively similar to the tissue-cell level aging score (Figure 3A), with the immune tissues (spleen, marrow, thymus) and the epithelial tissues (skin, tongue) at the top, and brain myeloid and pancreas at the bottom. We also computed the tissue aging score using the droplet data and the bulk data, respectively. We observed a strong consistency between the aging scores computed using different data (Figure 4D for comparison with droplet data, Supplementary Figure 13B for comparison with bulk data), indicating the robustness of the aging score.

Discussion

This study provides a systematic and comprehensive analysis of aging-related transcriptomic signatures by analyzing 76 tissue-cell types in the TMS FACS data. Together with the analysis in the first publication of Tabula Muris Senis[35], this forms one of the largest analysis to date of the mammalian aging process at the single-cell resolution. Of particular interest are the 292 global aging genes identified in the study. These genes exhibit aging-dependent expression in a majority of the tissue-cell types in the mouse. Senescence genes, transcription factors, eukaryotic initiation factors, and ribosomal protein genes are enriched among the global aging genes. Interestingly, most of the global aging genes are strongly bimodal—their expression either decrease across almost all tissue-cell types or it increases across the board. This suggests that these genes have a uniform response to aging, which is robust to the specific tissue or cellular context. Moreover, we find a systematic decrease in expression for most genes, suggesting a turning off of transcription activity as the animal ages.

While we have validated our finding using the droplet data and bulk data, it is important to have further validations in future

studies. In particular, the bimodal expression pattern is less apparent in the droplet data, perhaps due to its limited tissue-cell types coverage and relatively shallower sequencing depth. The remarkable bimodal consistency of global aging genes makes them useful as biomarkers to characterize the aging status of individual cells. We proposed a new aging score based on the global aging genes. This aging score quantifies how sensitive each tissue-cell type is to aging and is positively correlated with the cell division rate. For example, immune cells tend to have a high aging score, which reflects the phenomenon that they undergo many cycles of cell division and also change substantially during the animal's lifespan. One hypothesis is that the aging score captures some aspects of the true biological age of the cells, which could be different from the birth age of the animal. This is partially supported by the fact that the aging score is concordant with the whole blood transcriptomic aging score of Peters *et al.*, which is predictive of the chronological age of different people[32]. An interesting direction of future work is to further investigate this model with functional experiments. In line with this, it would be important to study how some of the transcriptomic changes we quantify here, e.g., the down-regulation of mTOR, point towards healthy aging. Overall, our study provides a comprehensive characterization of aging genes across a wide range of tissue-cell types in the mouse. In addition to the biological insights, it also serves as a comprehensive reference for researchers working on related topics.

Methods

Data preprocessing

For the MACA FACS data and the MACA droplet data[35], we filtered out genes expressed in fewer than 5 cells, filtered out cells expressing fewer than 500 genes, and discarded cells with a total number of counts fewer than 5000 for the FACS data and 3000 for the droplet data. Then, we normalized the cells to have 10,000 reads per cell, followed by a log (natural base) transformation. Such a procedure is the same as in the original paper. The bulk data is from a companion study[37]. For the bulk data, we filtered out genes expressed in fewer than 5 samples, and filtered out samples expressing fewer than 500 genes. We normalized each sample to have 10,000 reads per sample, followed by a log transformation.

Differential gene expression analysis

For the differential gene expression analysis, as shown in Supplementary Figures 1A-B, we consider all 23 tissues and 76 tissue-cell types with more than 100 cells in both young (3m) and old (18m, 24m) age groups for the MACA FACS data; all 11 tissues and 24 tissue-cell types with more than 500 cells in both young (1m, 3m) and old (18m, 21m, 24m, 30m) age groups for the MACA droplet data; and all 17 tissues for the bulk data[37]. We required more cells for the droplet data than the FACS data because droplet data has a much shallower sequencing depth (6000 UMIs per cell, as compared to 0.85 million reads per cell for the FACS data). Also, we did not focus on the MACA droplet data for the downstream analysis due to its limited tissue and tissue-cell type coverage.

We identified genes significantly related to aging (DGE analysis) using a linear model treating age as a continuous variable while controlling for sex:

$$\text{gene expression} \sim \text{age} + \text{sex}. \tag{1}$$

We used the MAST package[7] (version 1.10.0) in R to perform the DGE analysis. For covariates, we did not control for the number of genes expressed in each cell (referred to as cellular detection rate in the MAST paper) because we found that old cells had fewer expressed genes per cell, which was not confounded by the sequencing depth (Supplementary Figure 1C). We used the Benjamini-Hochberg (FDR) procedure[2] to control for multiple comparison. We applied an FDR threshold of 0.01 and an age coefficient threshold of 0.005 (corresponding to $\sim$10% fold change) for detecting genes significantly related to aging.

Tissue-cell level global aging genes

For the tissue-level analysis, we selected a gene as a global aging age if it was significantly related to aging in more than 50% of tissue-cell types. For Figure 2C, we picked 10 up-regulated global aging genes and 10 down-regulated genes that are significant in most of the tissue-cell types.

Aging score

We computed the aging score using the *scanpy.tl.score_genes* function from the Scanpy python package[43] (version 1.4.3). Given a list of target genes and a list of background genes, the *scanpy.tl.score_genes* function computes the gene score for a cell as the average expression of the target genes, subtracting the average expression of the background genes. We computed the aging score for each cell as the gene score of the up-regulated global aging genes (up-regulated in >80% tissue-cell types) minus the gene score of the down-regulated global aging genes (down-regulated in >80% tissue-cell types). We used the set of all genes as the background genes for computing both gene scores. After that, we further regressed out sex, chronological age, and their intersection term for each cell. The procedure can be summarized as

1. Compute the raw aging score for each cell:

$$\text{raw aging score} = \textit{scanpy.tl.score_genes}(\text{list of up-regulated global aging genes}) - \tag{2}$$
$$\textit{scanpy.tl.score_genes}(\text{list of down-regulated global aging genes}). \tag{3}$$

2. Compute aging score as the residual of the regression

$$\text{raw aging score} \sim \text{age} + \text{sex} + \text{age} \times \text{sex}. \tag{4}$$

For comparison with Peters *et al.*[32], we first compared the set of aging-related genes. Peters *et al.* considered 11,908 human genes from the whole blood data, among which 1,497 were significantly related to aging. 10,093 out of all 11,908 genes, and 1,328 out of 1,497 aging-related genes had their mouse gene counterpart in our data. Therefore, we focused on these 10,093 genes that were present in both studies. Among these genes, 217 were global aging genes, and 168 were immune-specific genes. The intersection between the aging-related genes in Peters *et al.* and global aging genes in our study could be summarized by the contingency table below, yielding a p-value of 0.16 via one-sided Fisher's exact test.

	Not related to aging in Peters *et al.*	Related to aging in Peters *et al.*	Total
Not global aging gene	8,582	1,294	9,876
Global aging gene	183	34	217
Total	8,765	1,328	10,093

(5)

The intersection between the aging-related genes in Peters *et al.* and immune-specific genes in our study could be summarized by the contingency table below, yielding a p-value of 8.3e-8 via one-sided Fisher's exact test.

	Not related to aging in Peters *et al.*	Related to aging in Peters *et al.*	Total
Not immune-specific gene	8,645	1,280	9,925
Immune-specific gene	120	48	168
Total	8,765	1,328	10,093

(6)

We next computed the Peters' whole blood transcriptomic age for all tissue-cell types in the FACS data based on the predictor in Peters *et al.* (Supplementary Data 5B, PREDICTOR-for-NEW-COHORTS[32]). We used the coefficients of the 10,093 genes that are present in both studies to compute a raw Peters' whole blood transcriptomic age for each cell of the processed FACS data (size factor normalization and log transformation). We then regressed out sex, chronological age and their interaction term based on Equation (4) to have the final Peters' whole blood transcriptomic age for each cell. The Peters' whole blood transcriptomic age for each tissue-cell type is averaged over all cells from the tissue-cell type. We note that the procedures for computing our aging score and the Peters' whole blood transcriptomic age are only different by the genes used and their weights. The comparison are shown in Supplementary Figure 9.

Gene pathway analysis and upstream regulator prediction

Both analyses were done using the ingenuity Pathway Analysis software (IPA)[16] core analysis functionality. We generated Figure 2D for the pathway analysis of global aging genes with a FDR threshold 1e-5 and an absolute z-score threshold 0.7. For the upstream regulator prediction, top ten regulators were visualized for each category. The full list of predicted upstream regulators can be found in Supplementary Table 2.

Data availability

All data can be downloaded at https://figshare.com/articles/tms_gene_data/11413869

Code availability

The code for reproducing all results is at https://github.com/czbiohub/tabula-muris-senis/tree/master/2_aging_signature

Supplementary Information

1. Supplementary Figures
2. Supplementary Data: Tissue-cell level DGE results for the FACS data and the droplet data; tissue-level DGE results for the FACS data, the droplet data, and the bulk data.

3. Supplementary Table 1: Tissue-cell level and tissue-level global aging genes.
4. Supplementary Table 2: Tissue-cell specific aging genes for immune cells, stem cells, and endothelial cells.
5. Supplementary Table 3: Functional annotation for the 76 tissue-cell types in the FACS data, and the average life span for a subset of cell types, with reference.

References

1. A Ahluwalia, MK Jones, S Szabo, and AS Tarnawski. Aging impairs transcriptional regulation of vascular endothelial growth factor in human microvascular endothelial cells: implications for angiogenesis and cell survival. 2014.
2. Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.
3. Judith Campisi. Aging, cellular senescence, and cancer. *Annual review of physiology*, 75:685–705, 2013.
4. Samuel H Cheshier, Sean J Morrison, Xinsheng Liao, and Irving L Weissman. In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proceedings of the National Academy of Sciences*, 96(6):3120–3125, 1999.
5. Anthony J Donato, R Garrett Morgan, Ashley E Walker, and Lisa A Lesniewski. Cellular and molecular biology of aging endothelial cells. *Journal of molecular and cellular cardiology*, 89:122–135, 2015.
6. Maria Ermolaeva, Francesco Neri, Alessandro Ori, and K Lenhard Rudolph. Cellular and epigenetic drivers of stem cell ageing. *Nature Reviews Molecular Cell Biology*, 19(9):594, 2018.
7. Greg Finak, Andrew McDavid, Masanao Yajima, Jingyuan Deng, Vivian Gersuk, Alex K Shalek, Chloe K Slichter, Hannah W Miller, M Juliana McElrath, Martin Prlic, et al. Mast: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell rna sequencing data. *Genome biology*, 16(1):278, 2015.
8. Sandra Freitas-Rodriguez, Alicia R Folgueras, and Carlos Lopez-Otin. The role of matrix metalloproteinases in aging: Tissue remodeling and beyond, 2017.
9. DA Fulcher and A Basten. B cell life span: a review. *Immunology and cell biology*, 75(5):446–455, 1997.
10. Barbara Geering, Christina Stoeckle, Sébastien Conus, and Hans-Uwe Simon. Living and dying for inflammation: neutrophils, eosinophils, basophils. *Trends in immunology*, 34(8):398–409, 2013.
11. Alice C Holly, David Melzer, Luke C Pilling, William Henley, Dena G Hernandez, Andrew B Singleton, Stefania Bandinelli, Jack M Guralnik, Luigi Ferrucci, and Lorna W Harries. Towards a gene expression biomarker set for human biological age. *Aging cell*, 12(2):324–326, 2013.
12. Steve Horvath. Dna methylation age of human tissues and cell types. *Genome biology*, 14(10):3156, 2013.
13. Simon C Johnson, Peter S Rabinovitch, and Matt Kaeberlein. mtor is a key modulator of ageing and age-related disease. *Nature*, 493(7432):338–345, 2013.
14. Juulia Jylhävä, Nancy L Pedersen, and Sara Hägg. Biological age predictors. *EBioMedicine*, 21:29–36, 2017.
15. Maranke I Koster. Making an epidermis. *Annals of the New York Academy of Sciences*, 1170:7, 2009.
16. Andreas Krämer, Jeff Green, Jack Pollard Jr, and Stuart Tugendreich. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics*, 30(4):523–530, 2013.
17. Mathieu Laplante and David M Sabatini. mtor signaling in growth control and disease. *cell*, 149(2):274–293, 2012.
18. Gaëlle Laurent, Florence Solari, Bogdan Mateescu, Melis Karaca, Julien Castel, Brigitte Bourachot, Christophe Magnan, Marc Billaud, and Fatima Mechta-Grigoriou. Oxidative stress contributes to aging by enhancing pancreatic angiogenesis and insulin signaling. *Cell metabolism*, 7(2):113–124, 2008.
19. Siu Sylvia Lee, Raymond YN Lee, Andrew G Fraser, Ravi S Kamath, Julie Ahringer, and Gary Ruvkun. A systematic rnai screen identifies a critical role for mitochondria in c. elegans longevity. *Nature genetics*, 33(1):40–48, 2003.
20. Yinglin Liu, Carol R Gardner, Jeffrey D Laskin, and Debra L Laskin. Classical and alternative activation of rat hepatic sinusoidal endothelial cells by inflammatory stimuli. *Experimental and molecular pathology*, 94(1):160–167, 2013.
21. Carlos López-Otín, Maria A Blasco, Linda Partridge, Manuel Serrano, and Guido Kroemer. The hallmarks of aging. *Cell*, 153(6):1194–1217, 2013.

22. Yasushi Magami, Takeshi Azuma, Hideto Inokuchi, Shinichiro Kokuno, Fuminori Moriyasu, Keiichi Kawai, and Takanori Hattori. Cell proliferation and renewal of normal hepatocytes and bile duct cells in adult mouse liver. *Liver*, 22(5):419–425, 2002.

23. Barsanjit Mazumder, Prabha Sampath, Vasudevan Seshadri, Ratan K Maitra, Paul E DiCorleto, and Paul L Fox. Regulated release of l13a from the 60s ribosomal subunit as a mechanism of transcript-specific translational control. *Cell*, 115(2):187–198, 2003.

24. Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. Bionumbers—the database of key numbers in molecular and cell biology. *Nucleic acids research*, 38(suppl_1):D750–D753, 2009.

25. Encarnacion Montecino-Rodriguez, Beata Berent-Maoz, and Kenneth Dorshkind. Causes, consequences, and reversal of immune system aging. *The Journal of clinical investigation*, 123(3):958–965, 2013.

26. Teresa Niccoli and Linda Partridge. Ageing as a risk factor for disease. *Current biology*, 22(17):R741–R752, 2012.

27. Janko Nikolich-Žugich. The twilight of immunity: emerging concepts in aging of the immune system. *Nature immunology*, 19(1):10–19, 2018.

28. Yvonne Nolan, Frank O Maher, Darren S Martin, Rachael M Clarke, Miriam T Brady, Anthony E Bolton, Kingston HG Mills, and Marina A Lynch. Role of interleukin-4 in regulation of age-related inflammatory changes in the hippocampus. *Journal of Biological Chemistry*, 280(10):9354–9362, 2005.

29. Nuala A O'Leary, Mathew W Wright, J Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, Barbara Robbertse, Brian Smith-White, Danso Ako-Adjei, et al. Reference sequence (refseq) database at ncbi: current status, taxonomic expansion, and functional annotation. *Nucleic acids research*, 44(D1):D733–D745, 2015.

30. David Papadopoli, Karine Boulay, Lawrence Kazak, Michael Pollak, Frédérick Mallette, Ivan Topisirovic, and Laura Hulea. mtor as a central regulator of lifespan and aging. *F1000Research*, 8, 2019.

31. Linda S Park, Unja Martin, Kirsten Garka, Brian Gliniak, James P Di Santo, Werner Muller, David A Largaespada, Neal G Copeland, Nancy A Jenkins, Andrew G Farr, et al. Cloning of the murine thymic stromal lymphopoietin (tslp) receptor: Formation of a functional heteromeric complex requires interleukin 7 receptor. *Journal of Experimental Medicine*, 192(5):659–670, 2000.

32. Marjolein J Peters, Roby Joehanes, Luke C Pilling, Claudia Schurmann, Karen N Conneely, Joseph Powell, Eva Reinmaa, George L Sutphin, Alexandra Zhernakova, Katharina Schramm, et al. The transcriptional landscape of age in human peripheral blood. *Nature communications*, 6:8570, 2015.

33. Daniel A Petkovich, Dmitriy I Podolskiy, Alexei V Lobanov, Sang-Goo Lee, Richard A Miller, and Vadim N Gladyshev. Using dna methylation profiling to evaluate biological age and longevity interventions. *Cell metabolism*, 25(4):954–960, 2017.

34. Simone Picelli, Omid R Faridani, Åsa K Björklund, Gösta Winberg, Sven Sagasser, and Rickard Sandberg. Full-length rna-seq from single cells using smart-seq2. *Nature protocols*, 9(1):171, 2014.

35. Angela Oliveira Pisco, Nicholas Schaum, Aaron McGeever, Jim Karkanias, Norma F Neff, Spyros Darmanis, Tony Wyss-Coray, Stephen R Quake, et al. A single cell transcriptomic atlas characterizes aging tissues in the mouse. *bioRxiv*, page 661728, 2019.

36. Christopher S Potten. Stem cells in gastrointestinal epithelium: numbers, characteristics and death. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1370):821–830, 1998.

37. Nicholas Schaum, Benoit Lehallier, Oliver Hahn, Shayan Hosseinzadeh, Song E Lee, Rene Sit, Davis P Lee, Patricia Morán Losada, Macy E Zardeneta, Róbert Pálovics, et al. The murine transcriptome reveals global aging nodes with organ-specific phase and amplitude. *bioRxiv*, page 662254, 2019.

38. Giovanni Stallone, Barbara Infante, Concetta Prisciandaro, and Giuseppe Grandaliano. mtor and aging: An old fashioned dress. *International journal of molecular sciences*, 20(11):2774, 2019.

39. Robi Tacutu, Daniel Thornton, Emily Johnson, Arie Budovsky, Diogo Barardo, Thomas Craig, Eugene Diana, Gilad Lehmann, Dmitri Toren, Jingwei Wang, et al. Human ageing genomic resources: new and updated databases. *Nucleic acids research*, 46(D1):D1083–D1090, 2017.

40. Monica Teta, Simon Y Long, Lynn M Wartschow, Matthew M Rankin, and Jake A Kushner. Very slow turnover of β-cells in aged adult mice. *Diabetes*, 54(9):2557–2567, 2005.

41. Jan Vijg and Yousin Suh. Genome instability and aging. *Annual review of physiology*, 75:645–668, 2013.

42. Thomas Weichhart. mtor as regulator of lifespan, aging, and cellular senescence: a mini-review. *Gerontology*, 64(2):127–134, 2018.

43. F Alexander Wolf, Philipp Angerer, and Fabian J Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19(1):15, 2018.

44. Methodios Ximerakis, Scott L Lipnick, Sean K Simmons, Xian Adiconis, Brendan T Innes, Danielle Dionne, Lan Nguyen, Brittany A Mayweather, Ceren Ozek, Zachary Niziolek, et al. Single-cell transcriptomics of the aged mouse brain reveals convergent, divergent and unique aging signatures. *bioRxiv*, page 440032, 2018.

45. Xiang Zhou, Wen-Juan Liao, Jun-Ming Liao, Peng Liao, and Hua Lu. Ribosomal proteins: functions beyond the ribosome. *Journal of molecular cell biology*, 7(2):92–104, 2015.

Acknowledgements

We would like to thank S. Quake, R. Sinha, R. Sit, J. Cool, B. van de Geijn, H. Shi, X. Xu for feedback. M.J.Z. and J.Z. are supported by NSF CCF 1763191, NIH R21 MD012867-01, NIH P30AG059307, and grants from the Silicon Valley Foundation and the Chan-Zuckerberg Initiative.

Author contributions

M.J.Z., A.O.P., and S.D. analyzed the data. M.J.Z., A.O.P., and J.Z. wrote the manuscript. J.Z. supervised the research. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.