Supplementary Material for

Cellular age explains variation in age-related cell-to-cell transcriptome variability

Ming Yang[1], Benjamin R. Harrison[1], Daniel E.L. Promislow[1,2,*]

[1]Department of Laboratory Medicine and Pathology, University of Washington School of Medicine, Seattle, WA 98195 USA

[2]Department of Biology, University of Washington, Seattle, WA 98195 USA

*To whom correspondence should be addressed. Email: promislo@uw.edu

This PDF file includes:

Supplemental Methods

Supplemental Figures S1 to S14

Supplemental Methods

Modeling cell population dynamics

To estimate the age distribution of specific cells, we assume a constant (age-independent) survival rate for a given cell type. We first transformed cellular lifespan of each cell type in days into survival rate per day. Given a finite survival rate per day as p, the instantaneous mortality rate $\mu = -ln\ (p)$ (Caswell 1972). Assuming an exponentially distributed survival function, the expected mean lifespan is $T = \frac{-1}{ln\ (p)}$. Inversely, given cellular lifespan, we can calculate finite survival rate as

$$p = exp\ \left(-\frac{1}{T}\right) \qquad (1)$$

which ranges from 0 to 1.

Using estimates of p, we constructed a demographic model describing cell population dynamics, adapted from a standard class-structured life-history model (Caswell 2000), to derive cell type-specific turnover rates, and then working back from cell turnover rates, to determine the cellular age distribution at any organismal age. Considering the simplest case which only contains a progenitor compartment and the downstream fully-differentiated compartment holding cells of different age classes, we aim to derive the distribution of cellular ages in the downstream compartment provided a continuous influx of cells from the progenitor compartment. A life-cycle graph for this model is shown in **Supplemental Figure S1**. For the purposes of this model, we make the following simplifying assumptions. Initially, there are C cells in the cell population at the onset of maturation, and all cellular ages are at age class A=0. At each time unit, three events happen. The progenitor compartment produces F new cells which flow into age class A=0. Existing cells advance from one age class to the next age class with a survival probability of p irrespective of cellular age, or die with a probability of $1\text{-}p$. Cell turnover starts at the onset of maturation and is defined as $b = F/C$.

Assuming a discrete time model, the number of cells in age class i at organismal age t is $N(i, t)$, where, $N(i, t)$ can be formulated as

$$N(i,t) = p^i * F, \qquad for\ 0 \le i \le t-1$$

$$N(i,t) = p^i * C,\ for\ i = t \qquad (2)$$

Thus, at organismal age t, the total number of cells in a cell population, $S(t)$, is

$$S(t) = \sum_{i=0}^{t} N(i,t)\ = F + pF + p^2F + \cdots + p^{t-1}F\ +\ p^tC$$

$$= \sum_{i=0}^{t-1} p^{t-1} * F + p^t * C$$

$$= \frac{F(1-p^{t-1})}{1-p} + p^t * C \quad (3)$$

We assume that the cell population size C remains constant, and so, dividing both side of equation (3) by *C*, then we have

$$\frac{S(t)}{C} = \frac{F(1-p^{t-1})}{C*(1-p)} + p^t \Leftrightarrow 1 = \frac{F(1-p^{t-1})}{C*(1-p)} + p^t \Leftrightarrow$$

$$1 - p^t = \frac{F(1-p^{t-1})}{C*(1-p)} \Leftrightarrow 1 = \frac{F}{C*(1-p)} \Leftrightarrow$$

$$F = C * (1-p) \quad (4)$$

Thus, under the assumption of constant cell population size, the three parameters must satisfy the above relationship. This makes intuitive sense. In order for the cell population to stay constant, the number of cells lost from the fully differentiated compartment per time unit, $C * (1-p)$, should be equal to the replenishment (F) from the progenitor compartment. As a result, the cell turnover rate is

$$b = F/C = (1-p) \quad (5)$$

Next, to get the number of cells in age class *i* at organismal age *t*, we replace F with $C * (1-p)$ in the equation (2) of *N*(*i, t*),

$$N(i,t) = p^i * C * (1-p), \ for\ 0 \le i \le t-1$$

$$N(i,t) = p^i * C, \ for\ i = t \quad (6)$$

The proportion of cells in age class *i* at organismal age *t* as $f(k,t)$ is given by dividing equation (6) by *C* on both sides:

$$f(i,t) = (1-p) * p^i, when\ i = 0, \ldots, t-1$$

$$f(i,t) = p^i, when\ i = t \quad (7)$$

This closed-form analytical expression allows the derivation the mean of the cellular age distribution at organismal age *t*,

$$E(A) = \sum_{i=0}^{t} i * f(i,t) = t * p^t - \frac{t*p^{t+1} - p^{t+1} - t*p^t + p}{p-1} \quad (8)$$

As shown above, E(*A*) is fully determined by cell survival rate, *p*, and the organismal age, *t*. As *p* approaches 1, E(*A*) approaches *t* (i.e., $\lim_{p \to 1} E(A) = t$).

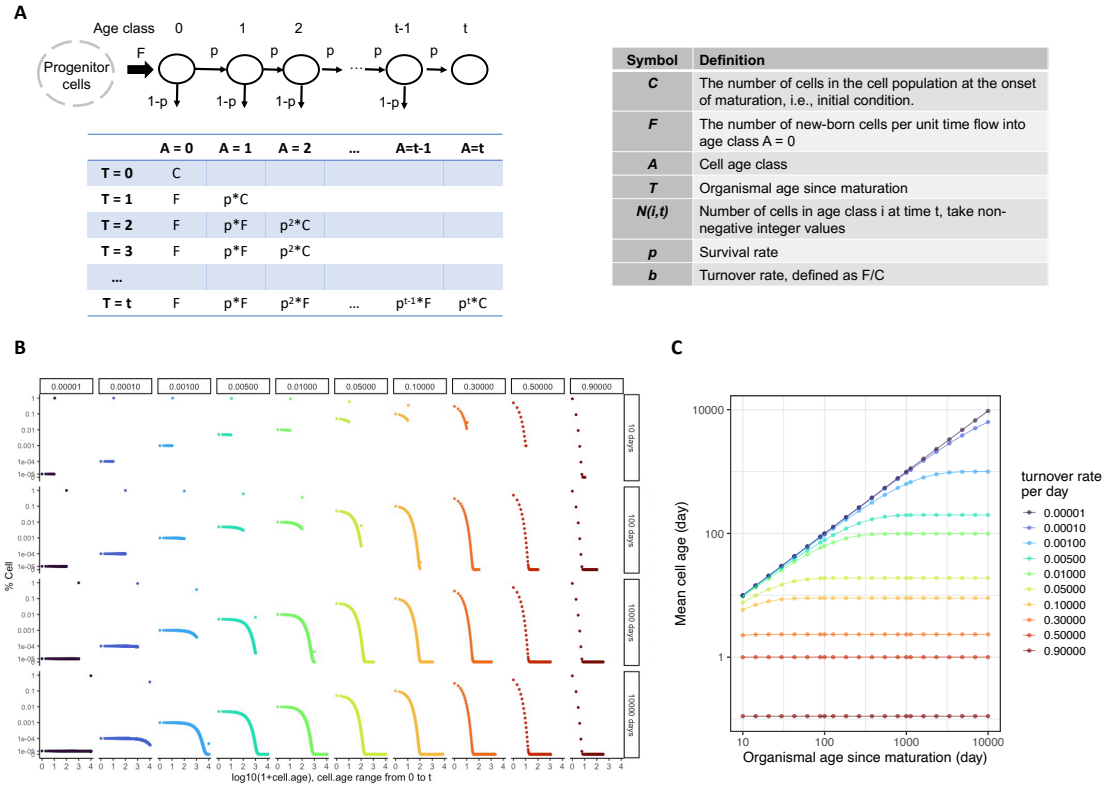
From equation (5), the cell turnover rate is $b = 1 - p = 1 - exp\left(-\frac{1}{T}\right)$. Assuming the onset of maturation in mice at 2 months old (Brust et al. 2015) and using equation (8), we

computed the mean cellular age over organismal age for each cell type (**Supplemental Fig. S2**).

At a fixed cell turnover rate b, $E(A)$ changes with organismal age t. Denoting the mean cell age at 3 months old as E($A \mid t$ = 3 months) and that at 24 months old as E(A | t = 24 months) , Δ_{age}, or the change in the mean cell age of each cell type within 3-month-old and 24 month-old mice, is calculated as,

$$\Delta_{age} = E(A \mid t = 24 \text{ months}) - E(A \mid t = 3 \text{ months}) \quad (9)$$

Supplemental Figures



	A = 0	A = 1	A = 2	...	A=t-1	A=t
T = 0	C					
T = 1	F	p*C				
T = 2	F	p*F	p^2*C			
T = 3	F	p*F	p^2*C			
...						
T = t	F	p*F	p^2*F	...	p^{t-1}*F	p^t*C

Symbol	Definition
C	The number of cells in the cell population at the onset of maturation, i.e., initial condition.
F	The number of new-born cells per unit time flow into age class A = 0
A	Cell age class
T	Organismal age since maturation
N(i,t)	Number of cells in age class i at time t, take non-negative integer values
p	Survival rate
b	Turnover rate, defined as F/C

Figure S1. A population dynamic model of cell age distribution. (**A**) An age-structured life-history model of cell population dynamics. Each black circle represents one age class. Cells in class *i* might die with probability *1-p* or advance into the next age class *i*+1 with probability *p* per time unit. We assume a constant cell population size, with the progenitor compartment producing *F* new cells flowing into age class A = 0 each time unit. The matrix shows the distribution of cells in each age class over organismal ages. (**B**) The cell age distributions at different organismal ages with varying turnover rates, following Equation (7) in the Supplemental Methods. Each column represents one specification of cell turnover rate and each row represents one given organismal age. (**C**) The mean of cellular ages of cell populations at different organismal ages with varying turnover rates. Colors denote different cell turnover rates.

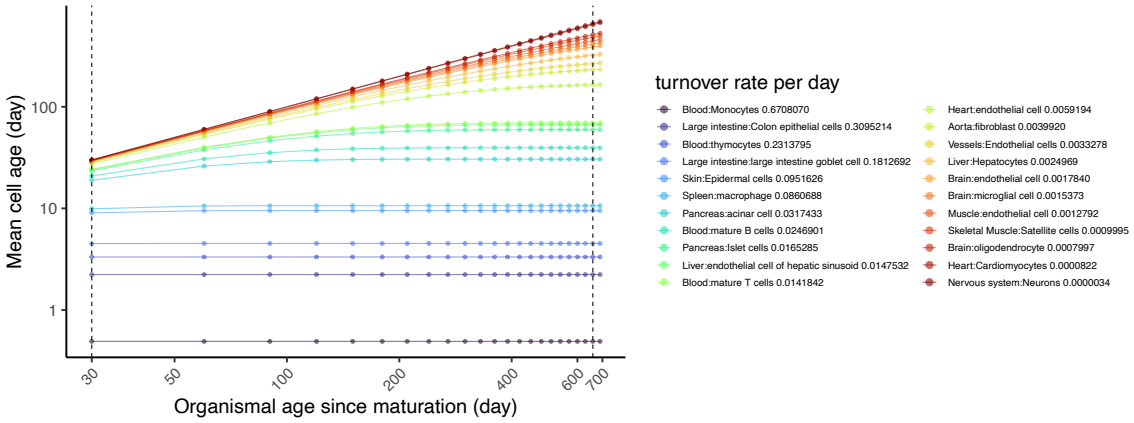


Figure S2. The effect of organismal age on cell age. Based on the model presented in the Supplemental Material, assuming the onset of maturation in mice at 2 months old (Brust et al. 2015), this figure shows the relationship between cell type-specific turnover rates and organismal age for specific cell types. Rodent cell-type specific lifespan data are from Sender and Milo (2021) (Sender and Milo 2021). Colors denote different cell types each with its own turnover rate.

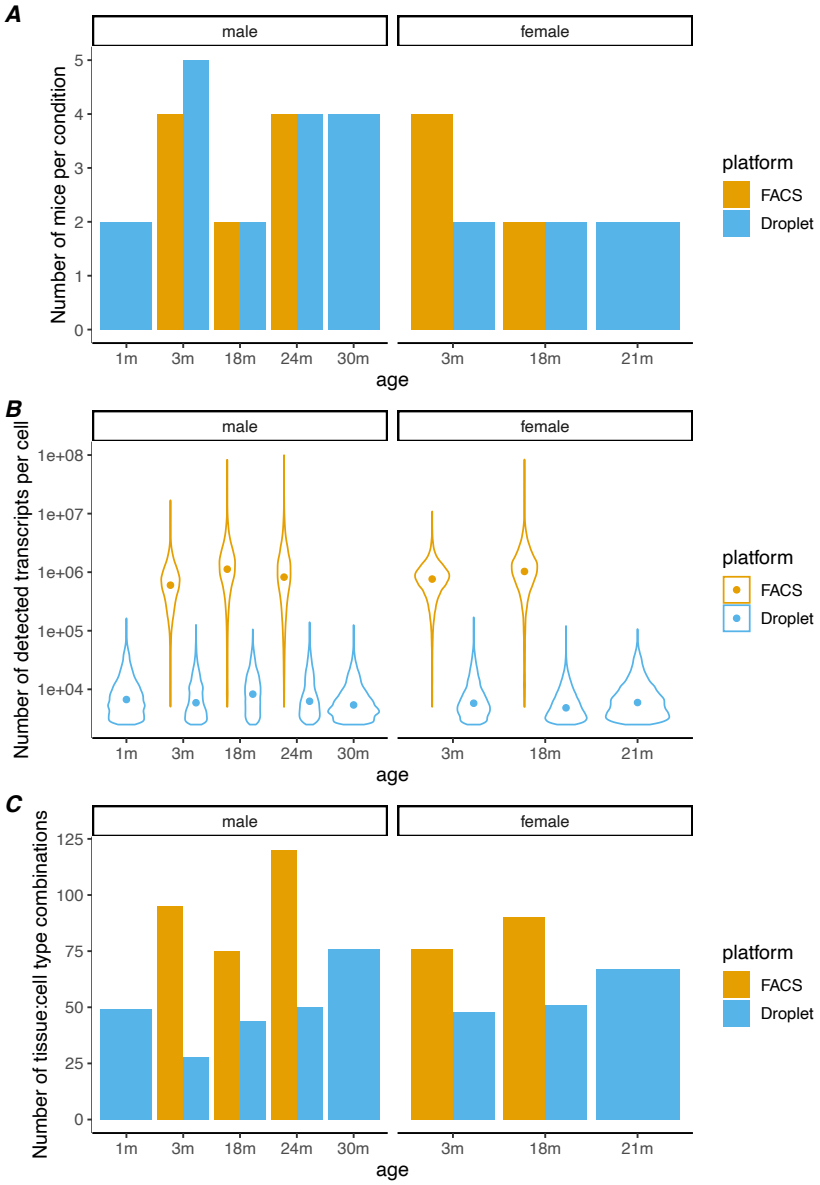


Figure S3. *Tabula muris senis* (TMS) dataset overview. (A) The number of profiled mice per condition. The TMS FACS data was collected from 10 male mice with ages ranging from 3 months to 24 months, and from 6 female mice ranging from 3 months to 18 months. The TMS droplet data was collected from 17 male mice with ages ranging from 1 months to 30 months, and from 6 female mice ranging from 3 months to 21 months. **(B)** The distribution of detected transcripts per cell across conditions. **(C)** The number of unique tissue:cell type combinations captured per condition. Within each sex and age group, cells with at least 500 expressed genes and 5000 transcripts were kept. From these cells, the tissue:cell type combinations with at least 40 cells were kept. We focused on the TMS FACS data as it has a much higher sequencing depth and more comprehensive tissue:cell type coverage.

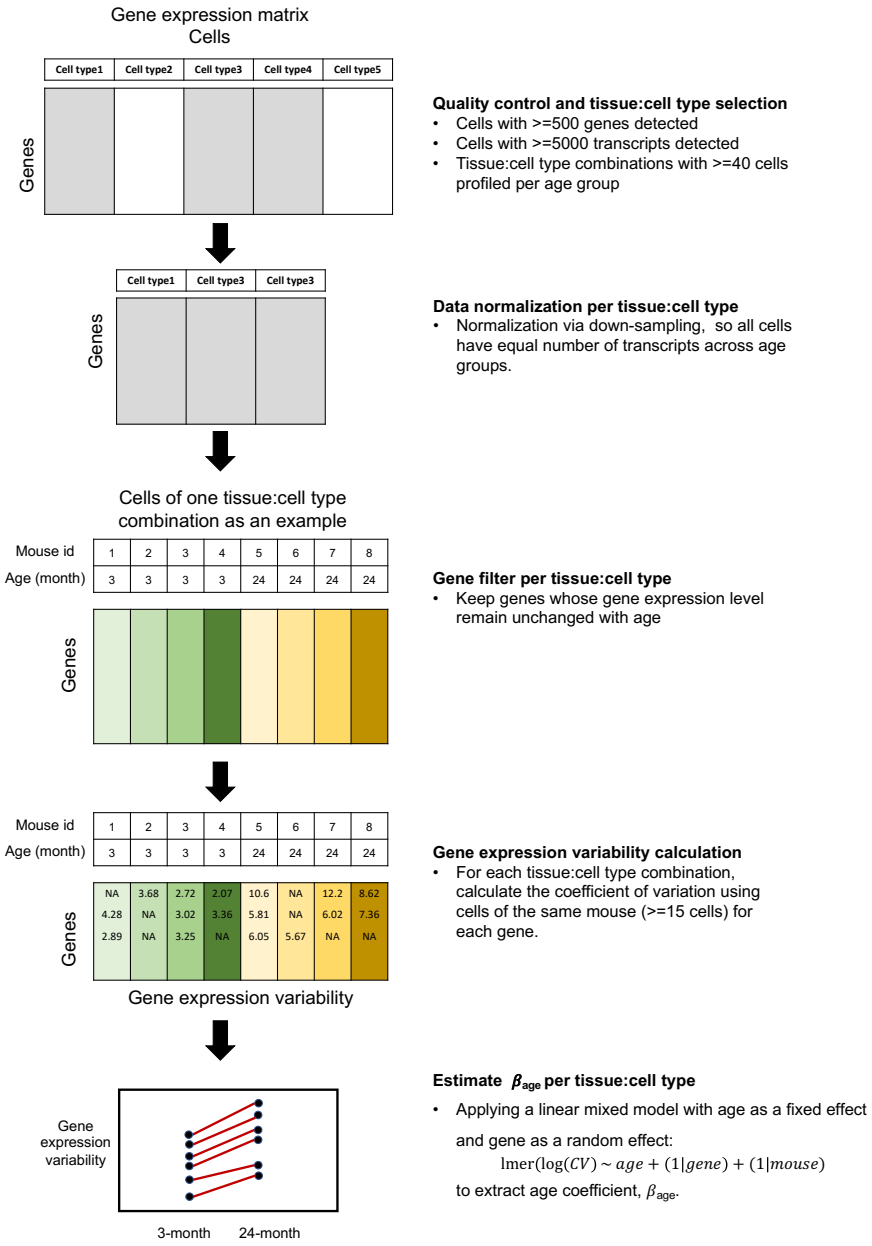


Figure S4. Overview of transcriptome variation analysis pipeline.


Aorta:aortic endothelial cell
beta_age=0.00939
FDR < 0.05
BAT:B cell
beta_age=0.01198
FDR < 0.05
BAT:endothelial cell
beta_age=0.00514
FDR = 0.175
Brain_Myeloid:microglial cell
beta_age=0.00875
FDR < 0.05
Brain_Non-Myeloid:endothelial cell
beta_age=0.01474
FDR < 0.05
Brain_Non-Myeloid:neuron
beta_age=0.02981
FDR < 0.05
Brain_Non-Myeloid:oligodendrocyte
beta_age=0.01279
FDR < 0.05
Diaphragm:endothelial cell
beta_age=0.00907
FDR = 0.0929
Diaphragm:skeletal muscle satellite cell
beta_age=0.01332
FDR < 0.05
GAT:endothelial cell
beta_age=0.01176
FDR < 0.05
Heart:atrial myocyte
beta_age=0.00606
FDR = 0.354
Heart:endothelial cell of coronary artery
beta_age=0.00983
FDR < 0.05
Heart:monocyte
beta_age=0.00785
FDR < 0.05
Large_Intestine:epithelial cell of large intestine
beta_age=0.00659
FDR = 0.0979
Large_Intestine:large intestine goblet cell
beta_age=0.00443
FDR = 0.0659
Limb_Muscle:endothelial cell
beta_age=0.01484
FDR < 0.05
Limb_Muscle:skeletal muscle satellite cell
beta_age=0.01812
FDR < 0.05
Liver:endothelial cell of hepatic sinusoid
beta_age=0.00371
FDR = 0.161
Liver:hepatocyte
beta_age=0.0102
FDR = 0.19
Marrow:macrophage
beta_age=0.00486
FDR = 0.242
Marrow:mature alpha-beta T cell
beta_age=0.01278
FDR < 0.05
Marrow:NK cell
beta_age=0.01713
FDR < 0.05
MAT:B cell
beta_age=0.00492
FDR < 0.05
MAT:endothelial cell
beta_age=0.00431
FDR = 0.0929
Pancreas:endothelial cell
beta_age=0.00559
FDR = 0.255
Pancreas:pancreatic A cell
beta_age=0.00495
FDR < 0.05
Pancreas:pancreatic acinar cell
beta_age=0.00258
FDR = 0.155
Pancreas:pancreatic B cell
beta_age=0.01061
FDR < 0.05
SCAT:B cell
beta_age=-2e-05
FDR = 0.996
SCAT:endothelial cell
beta_age=0.01234
FDR < 0.05
Skin:epidermal cell
beta_age=0.00708
FDR < 0.05
Spleen:B cell
beta_age=-0.00162
FDR = 0.674
Spleen:CD4-positive, alpha-beta T cell
beta_age=0.00244
FDR = 0.34
Spleen:CD8-positive, alpha-beta T cell
beta_age=0.00117
FDR = 0.451
Thymus:DN4 thymocyte
beta_age=0.00011
FDR = 0.996
Thymus:thymocyte
beta_age=-0.00971
FDR < 0.05
Trachea:endothelial cell
beta_age=0.01451
FDR < 0.05
Trachea:fibroblast
beta_age=0.01428
FDR < 0.05
Trachea:macrophage
beta_age=0.00632
FDR < 0.05
cv2 (log10)
age
3m
24m

Figure S5. Effect of age on transcriptome variability. Each panel represents one tissue:cell type combination. In each panel, the squared coefficient of variation (CV^2) is plotted for each gene at each age. To avoid an effect driven by a mean-variance correlation, this analysis was limited to genes that did not show age-related changes in mean expression levels (P value > 0.1, Methods in the main text). In each panel, lines connect the same gene measured at the two ages. The number of genes per tissue:cell type ranges from 286 to 3943. Age-related changes in gene expression variability are represented by the slope of these lines. For each tissue:cell type, the slope (β_{age}) describing the change in CV^2 with age was estimated using a linear mixed model with age as a fixed predictor and both gene and mouse as random effects ($\mathrm{lmer}(\log(CV^2) \sim \mathrm{age} + (1|\mathrm{gene}) + (1|\mathrm{mouse}))$, via the lme4 package in R (Bates et al. 2015).

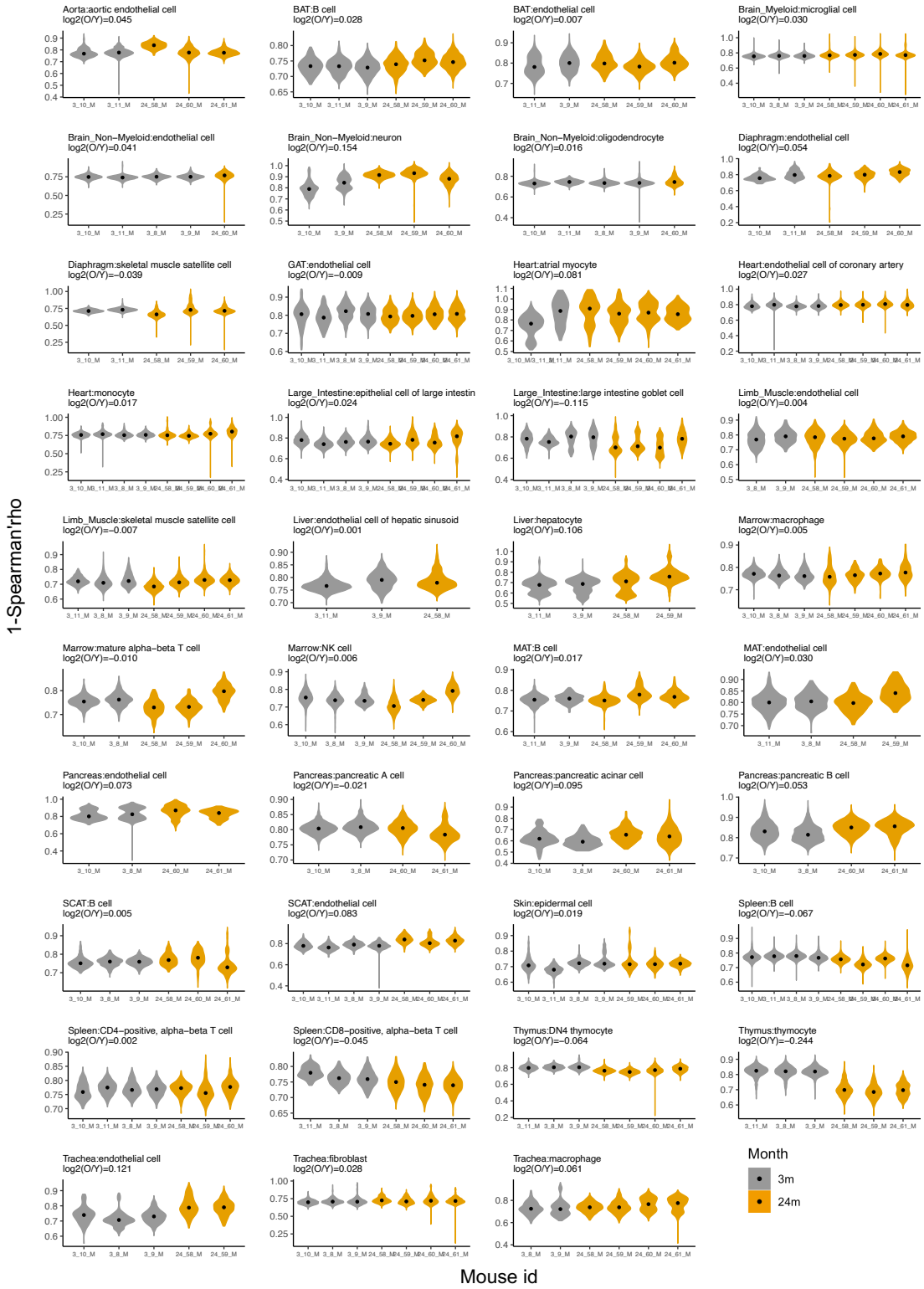


Figure S6. Effect of age on transcriptome variability, as measured by cell-cell correlations. Cell-to-cell correlation analyses were performed among cells from each

mouse individual ($\geqslant$15 cells) per tissue:cell type combination, using Spearman's rank correlation coefficients (ρ) as the measure of similarity between cells. Each ρ value corresponds to a pair of cells. The number of cells per tissue:cell type per age group ranges from 15 to 1215. The x-axis shows mouse individual id in the TMS. The y-axis shows 1-ρ and black dots denote the median value. We took the median of 1–ρ values within each mouse individual, and then used the mean across mice of the same age group to represent cell type-level transcriptome variability measurement. The age effect is measured as log2(O/Y), where the transcriptome variability in the young sample is Y and that of the old is O. Larger values of O compared with Y indicate larger cell-to-cell distances.

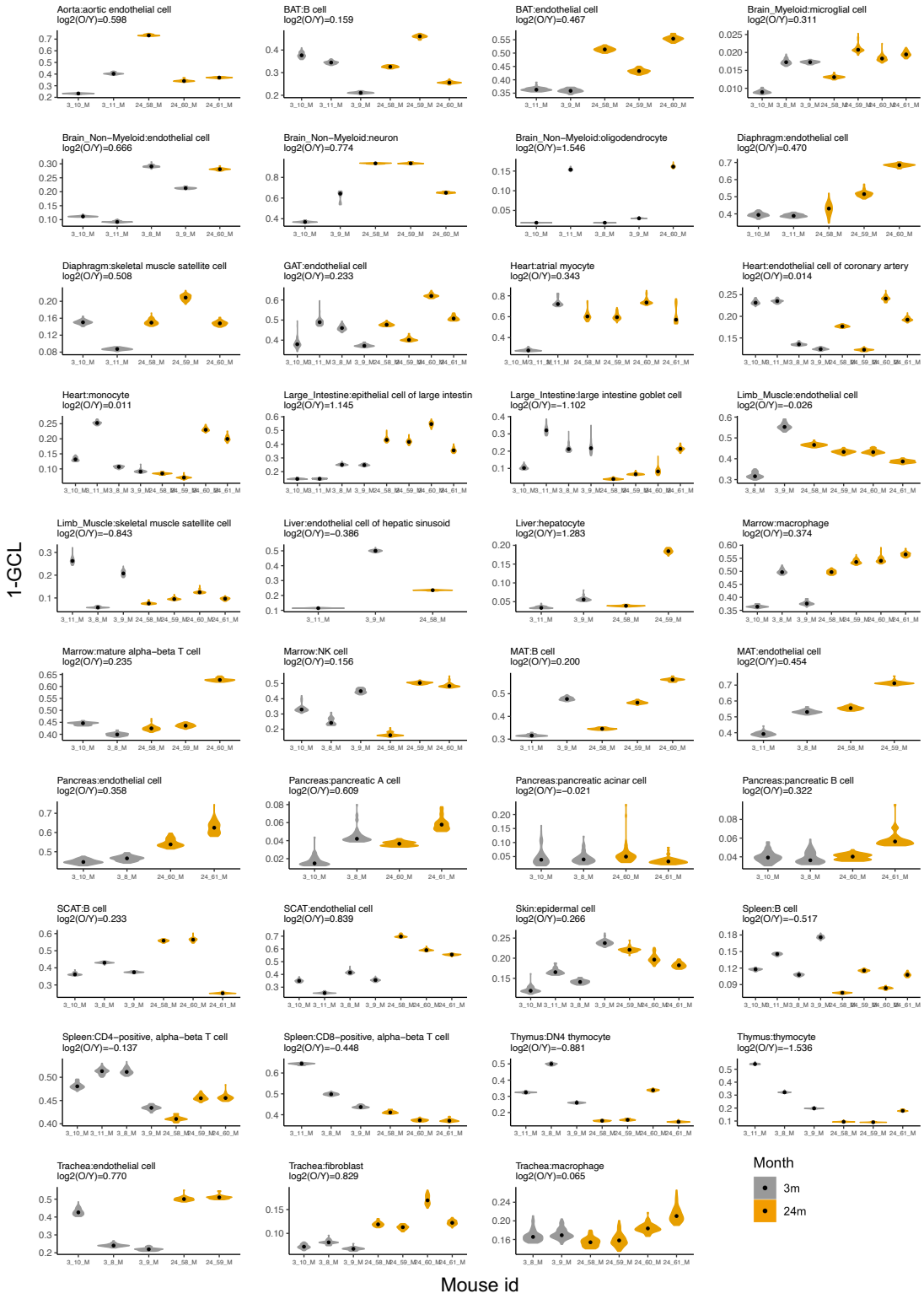


Figure S7. Effect of age on transcriptome variability as measured by GCL. The Global Coordination Level (GCL) metric, in which the similarity among cells is assessed, based

on their distance in gene expression space within subsets (samples) of genes, has been used to measure similarity among cells (Levy et al. 2020). The x-axis shows mouse individual id in the TMS. The y-axis shows the 1-GCL value, which increases with an increase in gene expression divergence. Each dot represents the distance between cells, measured by the similarity of expression matrices, made from cells of each mouse individual per tissue:cell type combination, when divided in half using random halves of genes expressed (Levy et al. 2020). There are 100 such random samples for each mouse individual per tissue:cell type combination. The black dot inside each violin plot denotes the median value which represents the mouse individual-specific discoordination value. We used the mean of mouse individual-specific discoordination values of the same age group to represent cell type-level transcriptome variability measurement. We express the change in transcriptome variability captured by this measure as log2(O/Y), where the transcriptome variability in the young sample is Y and that of the old is O.

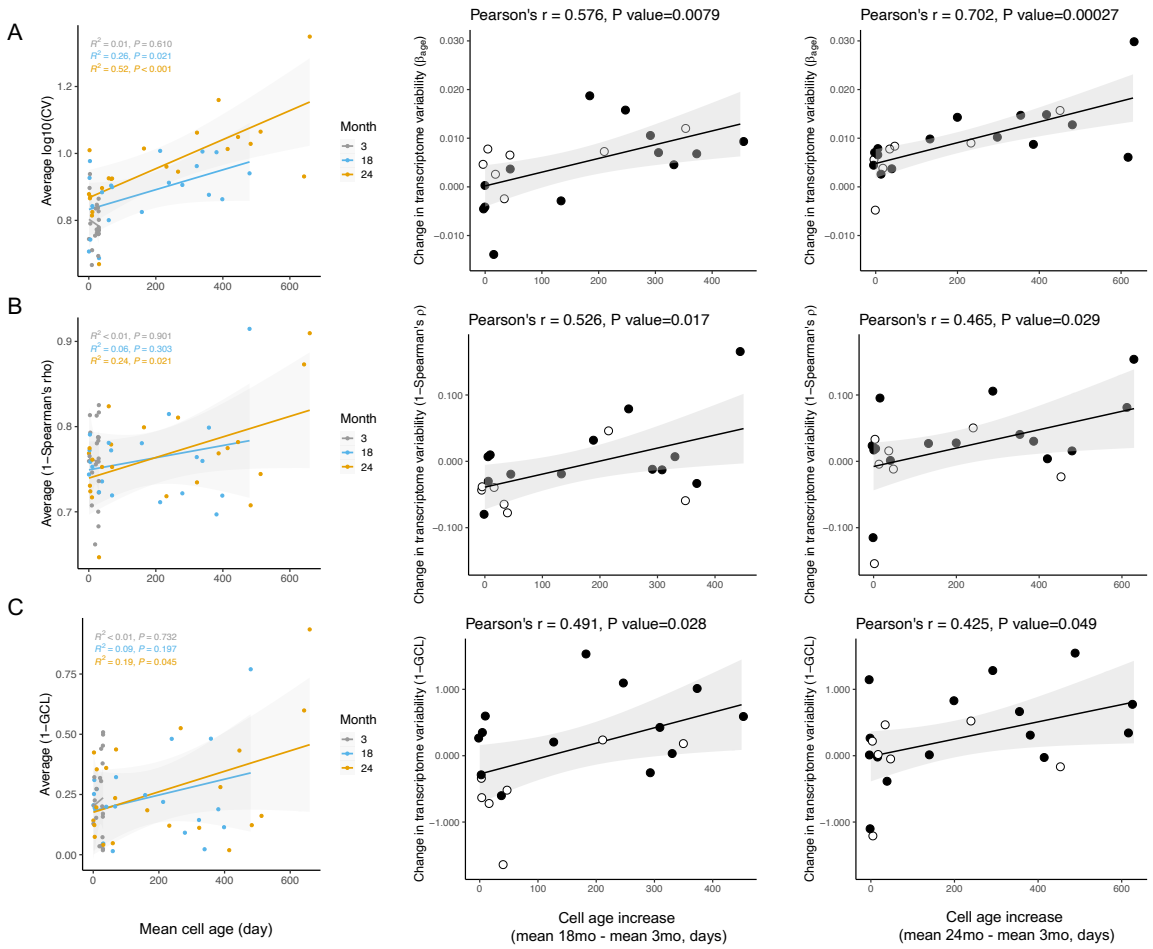


Figure S8. Different metrics for the effect of cell age on transcriptome variability. Each row of figures illustrates the relationship between age-related increases in transcriptome variability and cell age, using gene expression variability as shown in main Figure 2 (**A**), or two alternative approaches, Spearman rank correlations (**B**), and GCL (**C**) (see **Methods**). In each row, the leftmost panel shows the transcriptome variability of each cell type of each of three organismal ages (Month) against their respective mean cell ages (in days). The middle panels plot mean change in transcriptome variability observed in cells over 15-month period against the change in estimated mean cell age across cell types. The rightmost panels plot mean change in transcriptome variability observed in cells over 21-month period against the change in estimated mean cell age across cell types. In all panels, the lines show ordinary least squares regressions, with shading indicating 95% confidence intervals. Filled circles indicate the 15 cell types with tissue-specific lifespan data, and open circles the 7 cell types without tissue-specific lifespan estimates, and so their β_{age} is the mean from all tissues in which they are detected (see **Methods**).

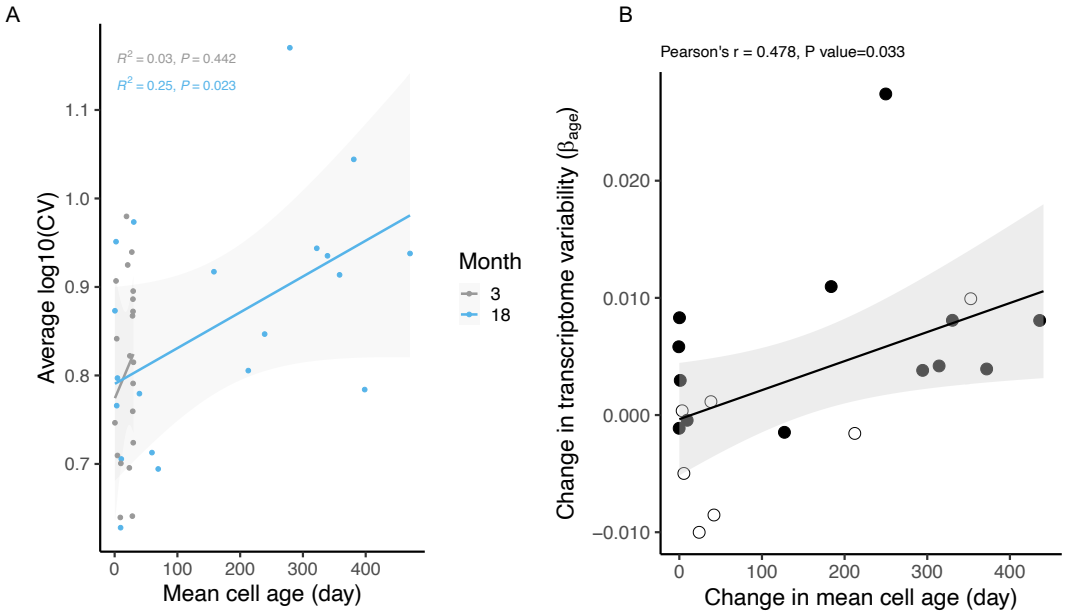


Figure S9. Cellular age dynamics predict the increase in transcriptome variability in female mice from 3-month to 18-month old. (**A**) The average log10 coefficient of variation (CV) of each cell type at each of three organismal ages (Month) is plotted over their respective mean cell ages (in days). The colored lines show ordinary least squares regressions, with shading indicating 95% confidence intervals, each corresponding to one age group. (**B**) The mean change in transcriptome variability (βage, Methods) observed in cells over a 15-month period, from 18-month-old mice compared to cells from 3-month-old mice, is correlated with the change in estimated mean cell age for 20 cell types. In both panels, the lines show ordinary least squares regressions, with shading indicating 95% confidence intervals. Filled circles indicate the 13 cell types with cell- and tissue-specific lifespan data, and open circles the 7 cell types without tissue-specific lifespan estimates, and so their βage is the mean from all tissues in which they are detected (see **Methods**).

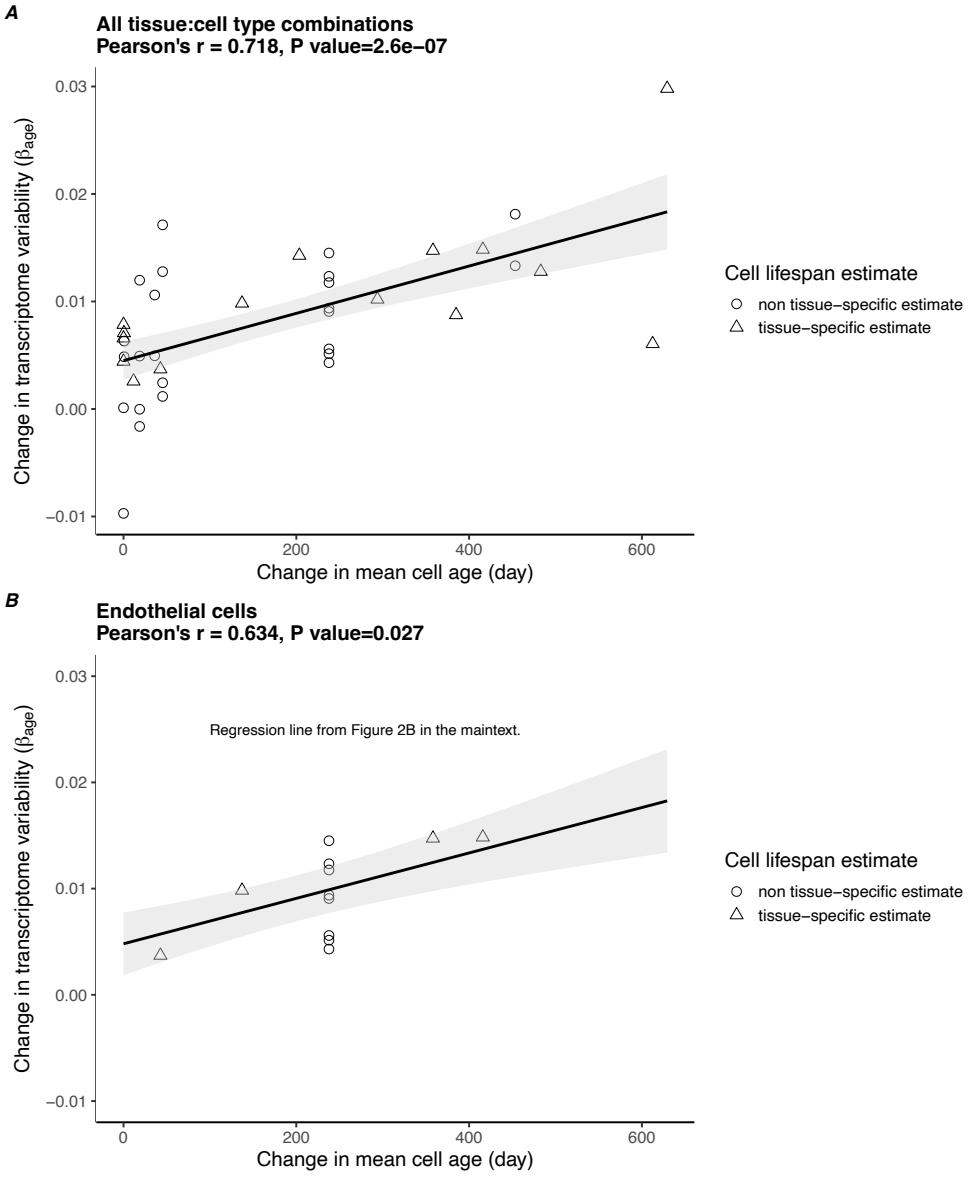


Figure S10. Within cell type, transcriptome variability across tissues may be explained by tissue-specific effects on cell turnover. (A) The change in transcriptome variability with age (β_{age}) across tissue:cell types, where point shapes denote cell types with (triangles) or without (circles) tissue-specific lifespan estimates. The black line shows an ordinary least squares regression, with shading indicating 95% confidence intervals. **(B)** Among the 12 tissues that had transcriptome data for endothelial cells, there are four with tissue-specific estimates of cell lifespan (triangles), from which we derived four separate cellular age distributions to get the change in cellular age over a 21-month period, from 24-month-old mice compared to cells from 3-month-old mice. These tissue-specific β_{age} and change in mean cell age values are shown in comparison to the regression of all data (black line taken from **Fig. 2B** in the main text).

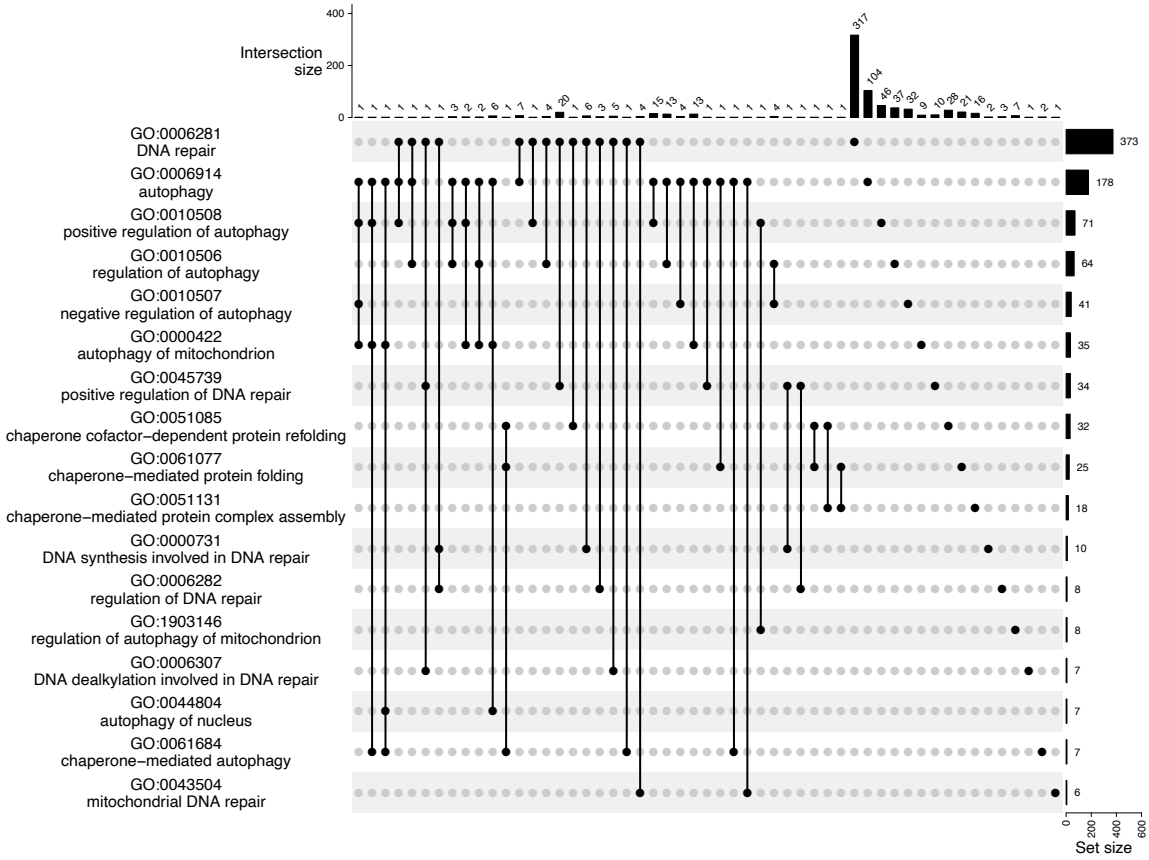


Figure S11. Intersection between the Gene Ontology (GO) biological process terms related to 'chaperone', 'DNA repair', and 'autophagy'. Each GO term contains at least six gene members shown as the horizontal bars (Set size). The top margin shows the number of genes that intersect between the GO terms. We used gene member expression abundance to calculate GO term expression and limited this analysis to non-redundant terms; for the 17 GO terms, there were at most 20 intersecting genes shared between terms, and 11 of the terms shared fewer than eight genes.

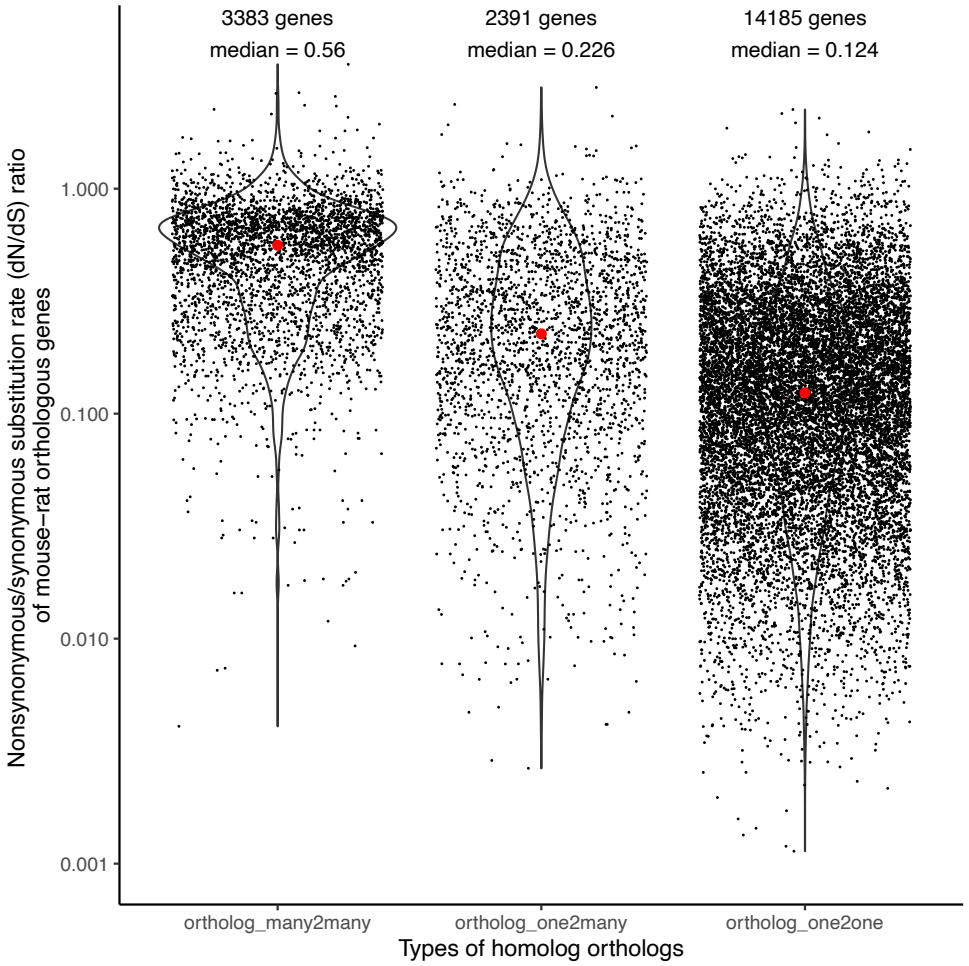


Figure S12. The distribution of d_N/d_S values (Ensembl Biomart (v.99) (Yates et al. 2020)**) among different types of orthologs between mouse and rat.**

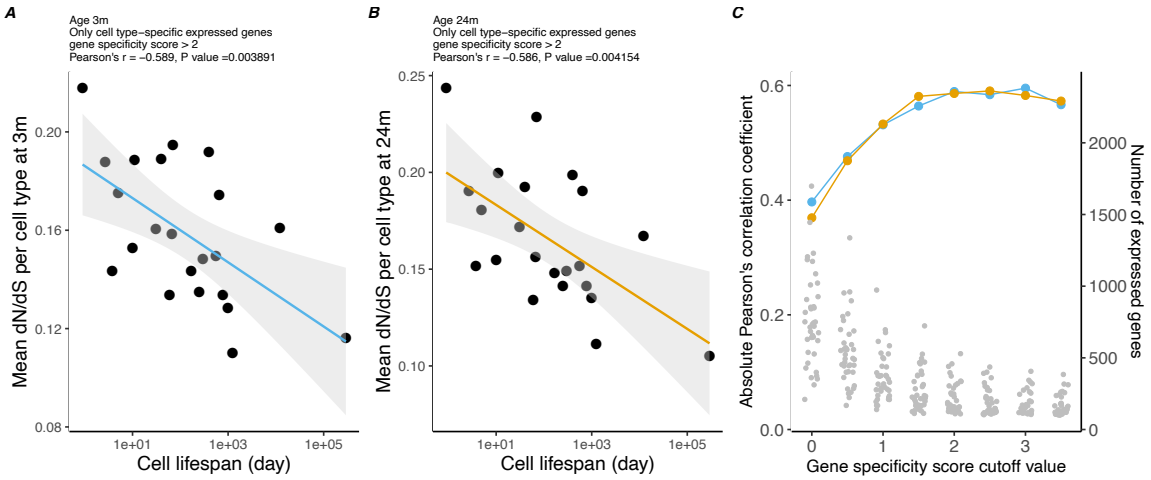


Figure S13. After removing genes under positive selection, the transcriptome of short-lived cells still shows relaxed evolutionary constraint. The relaxation of evolutionary constraint (d_N/d_S) of genes expressed in 22 cell types in young (panel **A**) and old (panel **B**) mice, is negatively correlated with cell lifespan. Smaller d_N/d_S values connote greater evolutionary constraint. The black line shows the ordinary least squares regression with shading of 95% confidence intervals. The analysis was performed with 14,109 genes after excluding 76 genes whose $d_N/d_S > 1$. **(C)** Correlation between d_N/d_S and cell lifespan for transcriptomes of increasing cell-type specificity. The blue (age 3m) and yellow (age 24 m) dots correspond to the left-side axis, and show the correlation strength. The grey dots correspond to the right-side axis and show the number of expressed genes per cell type at each level of specificity.

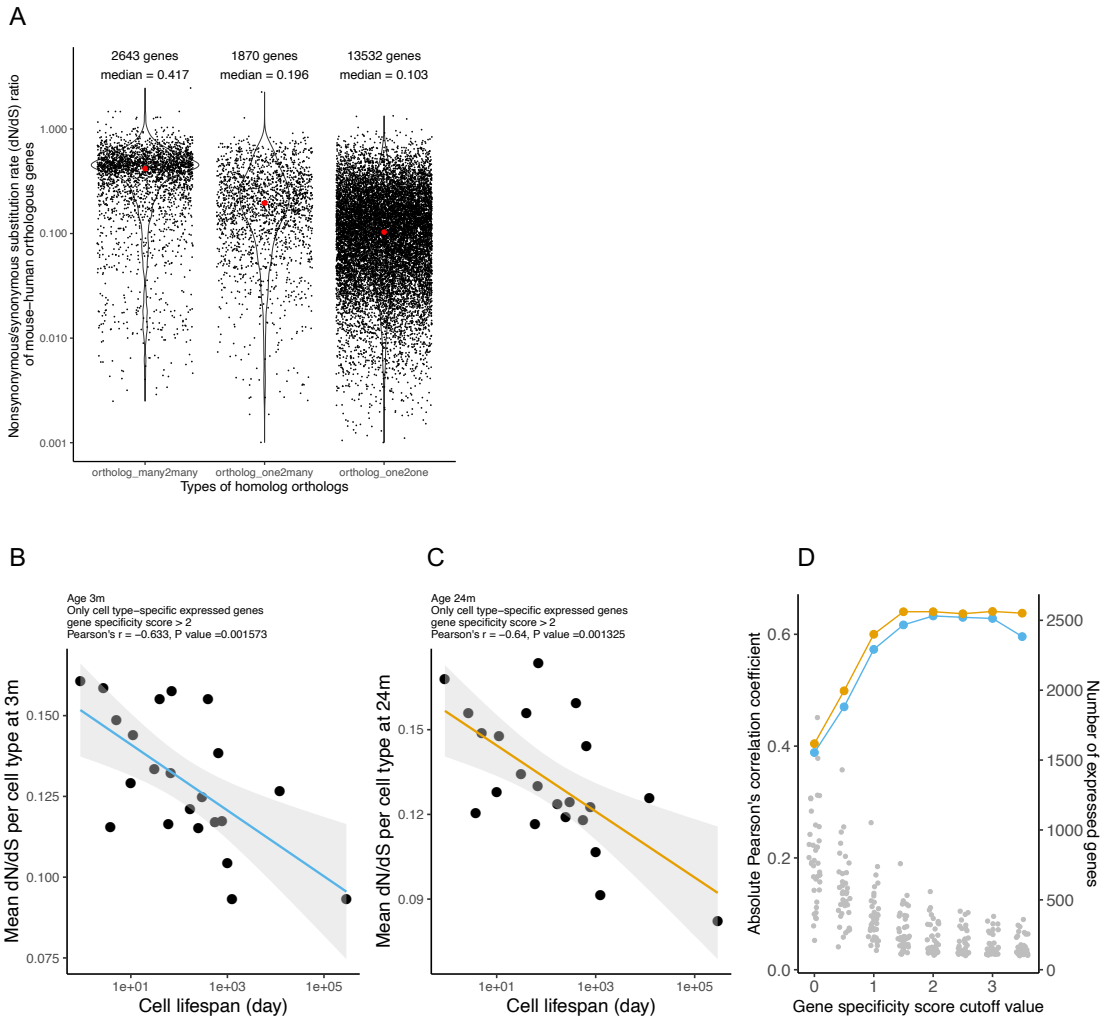


Figure S14. Cell lifespan still predicts evolutionary constraint among expressed genes with mouse-human orthologous genes. (A) The distribution of dN/dS values among different types of homolog orthologs between mouse and human. We retrieved dN/dS ratios of mouse-human orthologous genes from the Ensembl Biomart (v.99) (Yates et al. 2020). Only one-to-one orthologs estimated by Ensembl were included in the study, which gave 13,532 gene pairs. The relaxation of evolutionary constraint (d_N/d_S) of genes expressed in 22 cell types in young (panel **B**) and old (panel **C**) mice is negatively correlated with cell lifespan. Smaller d_N/d_S values connote greater evolutionary constraint. The black line shows the ordinary least squares regression with shading of 95% confidence intervals. The analysis was performed with 13,527 genes after excluding 5 genes whose $d_N/d_S > 1$. **(D)** Correlation between d_N/d_S and cell lifespan for transcriptomes of increasing cell-type specificity. The blue (age 3m) and yellow (age 24 m) dots correspond to the left-side axis, and show the correlation strength. The grey dots correspond to the right-side axis and show the number of expressed genes per cell type at each level of specificity.

References

Almanzar N, Antony J, Baghel AS, Bakerman I, Bansal I, Barres BA, Beachy PA, Berdnik D, Bilen B, Brownfield D, et al. 2020. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**: 590–595.

Bates D, Mächler M, Bolker BM, Walker SC. 2015. Fitting linear mixed-effects models using lme4. *J Stat Softw* **67**: 1–48.

Brust V, Schindler PM, Lewejohann L. 2015. Lifetime development of behavioural phenotype in the house mouse (Mus musculus). *Front Zool* **12**: 1–14.

Caswell H. 2000. *Matrix population models*. Sinauer Sunderland, MA.

Caswell H. 1972. On instantaneous and finite birth rates. *Limnol Oceanogr* **17**: 787–791.

Levy O, Amit G, Vaknin D, Snir T, Efroni S, Castaldi P, Liu YY, Cohen HY, Bashan A. 2020. Age-related loss of gene-to-gene transcriptional coordination among single cells. *Nat Metab* **2**: 1305–1315.

Sender R, Milo R. 2021. The distribution of cellular turnover in the human body. *Nat Med* **27**: 45–48.

Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, et al. 2020. Ensembl 2020. *Nucleic Acids Res* **48**: 682–688.

Zhang MJ, Pisco AO, Darmanis S, Zou J. 2021. Mouse aging cell atlas analysis reveals global and cell type-specific aging signatures. *Elife* **1**: e62293.