Inter-tissue convergence of gene expression and loss of cellular identity during ageing

Hamit Izgi[1], DingDing Han[2,+], Ulas Isildak[1], Shuyun Huang[2], Ece Kocabiyik[1], Philipp Khaitovich[3*], Mehmet Somel[1*], Handan Melike Dönertaş[4*]

[1] Department of Biological Sciences, Middle East Technical University, Ankara, Turkey
[2] CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China
[3] Center for Neurobiology and Brain Restoration, Skolkovo Institute of Science and Technology, Moscow, Russia
[4] European Molecular Biology Laboratory, European Bioinformatics Institute EMBL-EBI, Wellcome Trust Genome Campus, Cambridge, UK
[+] present address: Department of Clinical Laboratory, Shanghai Children's Hospital, Shanghai Jiaotong University, Shanghai, China

*Correspondence: msomel@metu.edu.tr (M.S.), donertas.melike@gmail.com (H.M.D), P.Khaitovich@skoltech.ru (P.K)

Abstract (124/150 words)

Gene expression tends to reverse its developmental trajectory direction in ageing, a phenomenon previously linked to cellular identity loss. Our analysis of cortex, lung, liver and muscle transcriptomes of 16 mice, covering development and ageing intervals, revealed widespread but tissue-specific ageing-associated expression reversals. Cumulatively, these reversals create a unique phenomenon: mammalian tissue transcriptomes diverge from each other during postnatal development, but during ageing, they tend to converge towards similar expression levels, a process we term divergence-convergence, or DiCo. We confirmed the DiCo pattern using data from independent mouse and human datasets. Further, using publicly available single-cell transcriptome data, we showed that DiCo could be driven both by alterations in tissue cell type composition and also by cell-autonomous expression changes within particular cell types.

Keywords
Ageing, development, transcriptome, mouse model

Introduction

Development and ageing in multicellular organisms are highly intertwined processes. On the one hand, certain ageing-related phenotypes, such as presbyopia, have been suggested to be caused by the continuation of developmental processes into adulthood (Blagosklonny 2013; de Magalhães and Church 2005). Such “runaway development” phenomena may be due to insufficient natural selection pressure to optimise regulation after sexual reproduction starts. Conversely, molecular studies have also reported a reversal of the ageing transcriptome towards pre-adult levels in various contexts, including primate brain regions (Somel et al. 2010; Dönertaş et al. 2017; Colantuoni et al. 2011), and mouse liver and kidney (Anisimova et al. 2020). Studying the functional consequences of this reversal pattern in the ageing human brain, we previously interpreted it as an indication of loss of cellular identity in neurons, possibly exacerbated by a relative reduction in neuronal proportions (Dönertaş et al. 2017). Such changes, in turn, could be caused by the accumulation of stochastic damage at the genetic, epigenetic, and proteomic levels over an adult lifetime, causing deregulation of gene expression networks.

Several major questions remain. First, the prevalence of reversal phenotypes across tissues is unclear, as most research has been conducted in the brain (Somel et al. 2010; Dönertaş et al. 2017). A second question pertains to the similarity of reversal-exhibiting genes and pathways across tissues. Ageing-related expression changes are partly shared among organs (Zahn et al. 2007), and reversal trends are also shared across different regions of the primate brain (Dönertaş et al. 2017). Distinct tissues might hence show parallel reversal patterns. Alternatively, as mammalian tissues diverge from each other during development in their transcriptome profiles (Cardoso-Moreira et al. 2019), one may hypothesise that during ageing, tissues converge back toward similar transcriptome profiles. Such a putative late-age convergence phenomenon would be consistent with the notion of ageing-related cellular identity loss (Yang et al. 2019; Dönertaş et al. 2017). A final question concerns the mechanism behind the observed reversal trends at the bulk tissue level. Specifically, the contribution of cell type composition and cell-autonomous changes to the reversals at the tissue level needs to be explored.

Documenting the reversal phenomenon is critical to better understand the proximate mechanisms of mammalian ageing, and its ultimate mechanisms, such as the stochastic disruption versus hyperfunction of developmental genes. However, such work has been limited by the scarcity of studies that comprise both development and ageing periods of the same organism and across different tissues. This work presents an age-series analysis of bulk transcriptome profiles of mice, including samples of four tissues across postnatal development and ageing periods covering the whole postnatal lifespan. Using this dataset, we study the prevalence, mechanisms, and functional consequences of the reversal phenomenon in different mouse tissues. We further test the related hypothesis of tissue convergence during ageing and investigate the contribution of cell type composition and cell-autonomous changes.

Results

We generated bulk RNA-seq data from 63 samples covering the cerebral cortex (which we refer as cortex), liver, lung, and skeletal muscle (which we refer as muscle) of 16 male C57BL/6J mice, aged between 2 to 904 days of postnatal age (see Methods). As mice reach sexual maturity by around two months (Tacutu et al. 2018), we treated samples from individuals aged between 2 and 61 days (n=7) as the development series, and those aged between 93 and 904 days (which roughly correspond to 80-year-old humans (Flurkey, M. Currer, and Harrison 2007)) (n=9) as the ageing series **(Figure S1)**. The final dataset contained n=15,063 protein-coding genes expressed in at least 25% of the 63 samples (one 904 days old mouse lacked cortex data).

Tissues diverge during postnatal development. Consistent with earlier work (Brawand et al. 2011; Cardoso-Moreira et al. 2019), we found that variation in gene expression is largely explained by tissue differences, such that the first three **p**rincipal **c**omponents (PCs) separate samples according to tissue (ANOVA $p<10^{-16}$ for PC1-3), with the cortex most distant from the others **(Figure 1a)**. Meanwhile, PC4, which explains 8% of the total variance, displayed a shared age-effect across tissues in development (Spearman correlation coefficient ρ=[0.88, 0.99], nominal p<0.01 for each test; **Figure 1b**, **Table S1**). Also, after the tissue effect was removed by standardisation, the PCA showed a strong influence of age on the first two PCs, which explain 31% of the variance in total **(Figure S2, Table S1)**.

In the PC3-PC4 space of Figure 1a, we further observed a trend of higher similarity among juvenile samples than among adult samples across tissues (change in mean Euclidean distance among tissues with age during development ρ_{dev}=0.9, p_{dev}=0.006; during ageing ρ_{age}=-0.3, p_{age}=0.429), which resonates with previous reports of inter-tissue transcriptome divergence during development (Cardoso-Moreira et al. 2019). This divergence pattern was also observed when **p**rincipal **c**omponents **a**nalysis (PCA) was performed with developmental samples only but was not significant (days 2 to 61; **Figure S3**).

Tissues involve common gene expression changes with age. We next characterised the age-related changes in gene expression shared across tissues by studying i) the trends at the transcriptome level and ii) the significant changes at the gene level. First, we studied similarities in overall trends in age-related gene expression changes using the Spearman correlation coefficient (ρ) between expression levels and age, for each gene, in each tissue, separately for the developmental and ageing periods (see Methods, tissue-specific age-related gene expression changes and functional enrichment test results are available as **Table S2**). We then examined similarities across tissues during development and ageing by comparing the gene-wise expression-age correlation coefficients **(Figure 1c)**. Considering the whole transcriptome without a significance cutoff, we found a weak correlation of age-related expression changes in tissue pairs, both during development (ρ=[0.17, 0.39], permutation test p<0.05 for all the pairs, **Table S3**), and ageing (ρ= [0.23, 0.33], permutation test p<0.05 in 4/6 pairs, **Table S3**). The difference between inter-tissue similarities within the development and those within ageing was not significant (Wilcoxon signed-rank test, p=0.31). Moreover, the number of genes with the same direction of change across four tissues was consistently more than expected by chance (p<0.05), except for genes upregulated in ageing (**Figures 1e, S4**). This attests to overall similarities across tissues both within the development and within ageing datasets, albeit of modest magnitudes.

In the second approach, we focused on genes with a significant age-related expression change, identified separately during development and ageing (using Spearman correlation and **f**alse-**d**iscovery **r**ate (FDR)<0.1, **Figure 1d**). We found that the developmental period was accompanied by a large number of significant changes (n=[1,941, 6,151], 13-41% across tissues), with the most manifest changes detected in the cortex. The genes displaying significant developmental change across all four tissues also showed significant overlap (**Figure S5, S6**; permutation test: p_{shared_up}=0.027, p_{shared_down}<0.001, false-positive rate < 20%). Using the **G**ene **O**ntology (GO), we found no functional enrichment for these shared developmentally up-regulated genes (**B**enjamini **Y**ekutieli (BY)-corrected p<0.1). In contrast, shared developmentally down-regulated genes were enriched in functions such as cell cycle and cell division (BY-corrected p<0.1; **Table S4**). Contrary to widespread expression change during development (13-41%), the proportion of genes undergoing significant expression change during ageing was between 0.0001-15% **(Figure 1d)**. This contrast between postnatal development and ageing was also observed in previous work on the primate brain (Somel et al. 2010; Işıldak et al. 2020). In terms of the number of genes with the significant ageing-related change, the most substantial effect we found was in the lung (n=2,319), while close to no genes showed a statistically significant change in the muscle (n=2), a tissue previously noted for displaying a weak ageing

transcriptome signature across multiple datasets (Turan et al. 2019). Not unexpectedly, we found no common significant ageing-related genes across tissues **(Figure S5a)**. Considering the similarity between the ageing and development datasets **(Figure 1c)** and the similar samples sizes in development (n=7) and ageing periods (n=9), the lack of overlap in significant genes in ageing might be due to low signal-to-noise ratios in the ageing transcriptome, as ageing-related changes are subtler compared to those in development **(Figure S5b)**.

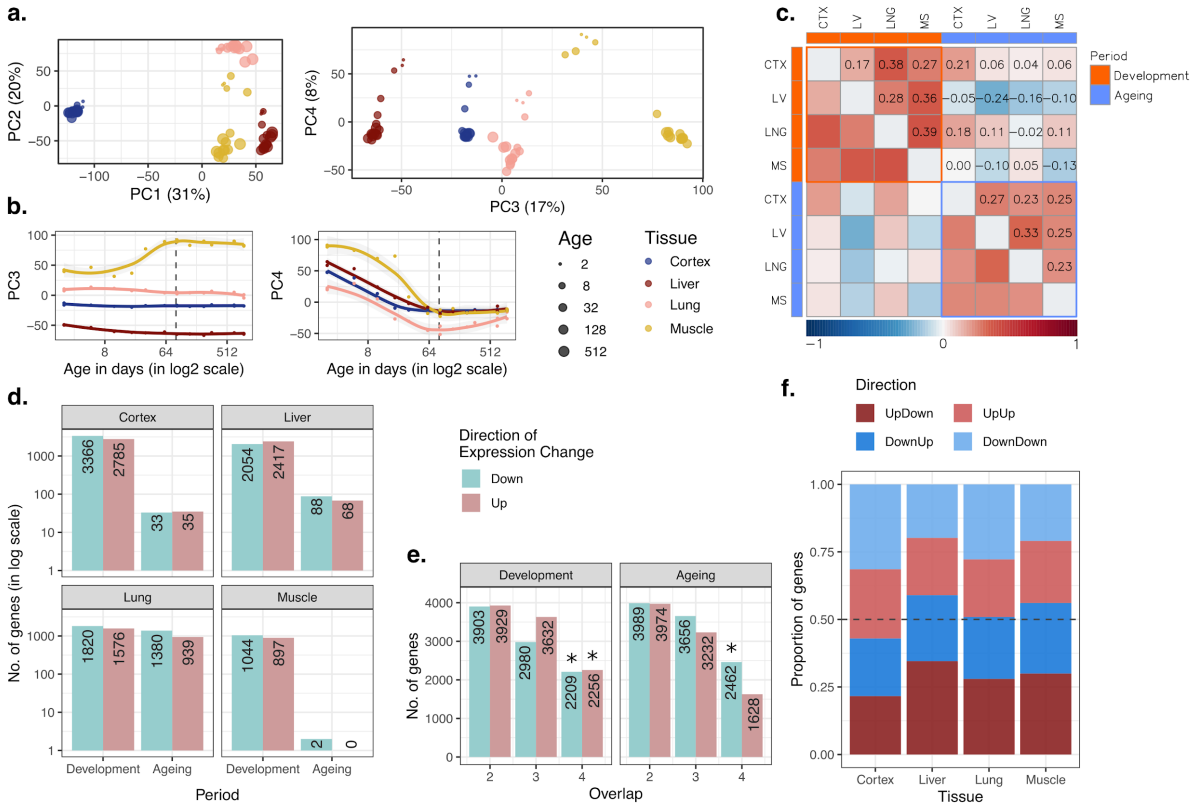

Figure 1: *a) Principal components analysis (PCA) of expression levels of 15,063 protein-coding genes across four tissues of 16 mice. Values in the parentheses show the variation explained by each component. b) Age trajectories of PC3 (left) and PC4 (right). Spearman correlation coefficients between PC4 and age in each tissue in development range between 0.88 and 0.99 (See* ***Table S1*** *for all tests). The dashed vertical line indicates 90 days of age, separating development and ageing periods. c) Similarity between the age-related gene expression changes (Spearman's correlation coefficient between expression and age) across tissues in development and ageing. Similarities were calculated using Spearman correlations coefficient between expression-age correlations across tissues. CTX: cortex, LV: liver, LNG: lung, MS: muscle. d) The number of significant age-related genes in each tissue (FDR<0.1). e) Shared age-related genes among tissues without using a significance cutoff. The x-axis shows the number of tissues among which age-related genes are shared. Significant overlaps are indicated with asterisk (*)* ***(Figure S4)****. f) The proportion of age-related expression change trends in each tissue across the lifetime. No significance cutoff was used. UpDown: up-regulation in development and down-regulation in the ageing; DownUp: down-regulation in development and up-regulation in the ageing; UpUp: up-regulation in development and up-regulation in the ageing; DownDown: down-regulation in development and down-regulation in ageing.*

Gene expression reversal is a common phenomenon in multiple tissues. We then turned to investigate the prevalence of the reversal phenomenon (*i.e.* an opposite direction of change during development and ageing) across the four tissues. We first compared age-related expression change between development and ageing periods in the same tissue **(Figure 1c)**. This revealed weak negative correlation trends in liver and muscle (though not in the lung and cortex), *i.e.* genes up- or down-regulated during development tended to be down- or up-regulated during ageing, respectively. These reversal trends were comparable when the analysis was repeated with the genes showing

relatively high levels of age-related expression change (|ρ|>0.6; **Figure S7**). We further studied the reversal phenomenon by classifying each gene expressed per tissue into those showing up- or down-regulation (irrespective of statistical significance) during development and during ageing, thus summarising trends of continuous change versus reversal in each tissue. In line with the above results, as well as earlier observations in the brain, kidney, and liver (Dönertaş et al. 2017; Anisimova et al. 2020), we found that ~50% (43-59%) of expressed genes showed reversal trends (**Figure 1f**), although these proportions were not significantly more than randomly expected in permutation tests (**Figure S8**, see Methods). Overall, we conclude that the expression trajectories of the genes do not necessarily continue linearly into the ageing period.

Pathways related to development, metabolism and inflammation are associated with the reversal pattern. We then asked whether genes displaying reversal patterns in each tissue may be enriched in common functional categories. Our earlier study focusing on different brain regions had revealed that up-down genes, *i.e.* genes showing developmental up-regulation followed by down-regulation during ageing, were enriched in tissue-specific pathways, such as neuronal functions (Dönertaş et al. 2017). Analysing up-down genes in the present study, we also found significant enrichment (BY corrected p-value<0.1) in functions such as “synaptic signaling” in the cortex, as well as “tube development” and “tissue morphogenesis” in the lung, “protein catabolic process” in the liver and “cellular respiration” pathways in the muscle **(Table S5)**. Meanwhile, down-up genes showed significant enrichment in functions such as “wound healing”, and “peptide metabolic process” in the cortex, “translation” and “nucleotide metabolic process” in the lung, and “inflammatory response” in the liver **(Table S5)**.

Genes showing reversal pattern are not shared among tissues. As tissues displayed modest correlations in development- and ageing-related expression change trends **(Figures 1c, S7)**, and as we had previously observed that distinct brain regions show similarities in their reversal patterns (*i.e.* the same genes showing the same reversal type), different tissues might also be expected to show similarities in their reversal patterns. Interestingly, we found no overlap between gene sets with the reversal pattern (up-down or down-up genes) across tissues, relative to random expectation ($p_{up\text{-}down}$>0.07, $p_{down\text{-}up}$=0.53, **Figure S9**). Such a lack of overlap might be explained if genes showing reversal patterns in each tissue tend to be tissue-specific. It would also be consistent with the notion that reversals involve loss of cellular identities gained in development, during which tissue transcriptomes appear to diverge from each other (**Figures 1a, S3**) (Cardoso-Moreira et al. 2019). This result led us to ask whether, in accordance with the reversal phenomenon, inter-tissue transcriptome divergence may be followed by increasing inter-tissue similarity, or convergence, during ageing.

Inter-tissue divergence during development and convergence during ageing. We studied the inter-tissue divergence/convergence question using two methods. In the first, we studied how expression variation among tissues changes with age. To do this, for each individual, we calculated the **c**oefficient **o**f **v**ariation (CoV) across the four tissues for each commonly expressed gene (n=15,063), which represents a measure of expression variation among tissues. Then, we assessed how such inter-tissue variation changes over the lifetime, by calculating the Spearman correlation between CoV and age, separately for development and ageing periods **(Table S6)**.

Using the CoV values calculated across all 15,063 genes (excluding one 904 days-old individual for which we lacked the cortex data), we observed a significant mean CoV increase in development (Spearman correlation ρ=0.77, p=0.041), confirming that tissues diverge as development progresses **(Figure 2a)**. Interestingly, during ageing, we observed a decrease in mean CoV with age, albeit not significant (ρ=-0.50, p=0.2, **Figure 2a**), suggesting that tissues may tend to converge during ageing. Repeating this analysis without cortex data (which lacks one individual) and thus including all 16 individuals revealed a stronger and significant decrease of mean CoV during ageing (ρ=-0.67,

p=0.047; **Figure S10**), suggesting the lack of significant result with all tissues might be due to decreased power. We obtained the same divergence-convergence pattern by calculating the median CoV values for each individual instead of the mean (**Figures S10-11**). **Figure 2b** exemplifies this pattern of increasing and decreasing CoV for the gene displaying the strongest such signal. We next studied the transition points between divergence and convergence by clustering genes showing divergence-convergence pattern (without using a significance cutoff, n=4,802) based on their CoV change with age **(Figure S24)**. Notably, the cluster with slightly delayed divergence (Cluster 1) was associated with metabolic and respiration-related processes (BY-corrected p<0.1), and the cluster with a relatively delayed convergence (Cluster 5) was enriched in categories related to vascular development (BY-corrected p<0.1) **(Table S7)**. To assess the contribution of tissues to divergence-convergence pattern, we clustered the genes (n=4,802) based on their expression change with age **(Figure S25)**. Not surprisingly, the clusters with relatively higher expression levels of a tissue (*e.g.* muscle in Cluster 9) were enriched in categories (BY-corrected p<0.1) related to that tissue (*e.g.* muscle cell development) **(Table S8)**.

We then studied this pattern at the single-gene level. We tested each gene for significant CoV change in their expression levels (*i.e.* divergence or convergence) in development and ageing (Spearman correlation test with FDR<0.1). We found that the ratio of divergent and convergent genes differed significantly during development (70% divergence among 2,581 significant genes) and ageing (68% convergence among 62 significant genes) **(Figure 2d-e)**. The same pattern was also observed without using significance cutoff **(Figure S12)**.

To our knowledge, inter-tissue convergence during ageing is a novel phenomenon. In order to verify this trend, we used a second approach to test it in our mouse dataset. For each individual, we calculated correlations between pairs of tissues across their gene expression profiles. Under the divergence-convergence pattern, we would expect pairwise correlations to decrease during development and increase during ageing. Among all pairwise comparisons, we observed a strong negative correlation during development (ρ=[-0.61, -0.9], nominal p<0.05 in 5 out of 6 tests), while during ageing, 4 out of 6 comparisons showed a moderate positive correlation (ρ=[0.16, 0.69], nominal p<0.05 in 1 out of 6 comparisons, **Figure S13**). Calculating the mean of pairwise correlations among tissues for each individual, we observed a significant divergence-convergence pattern **(Figure S14)**.

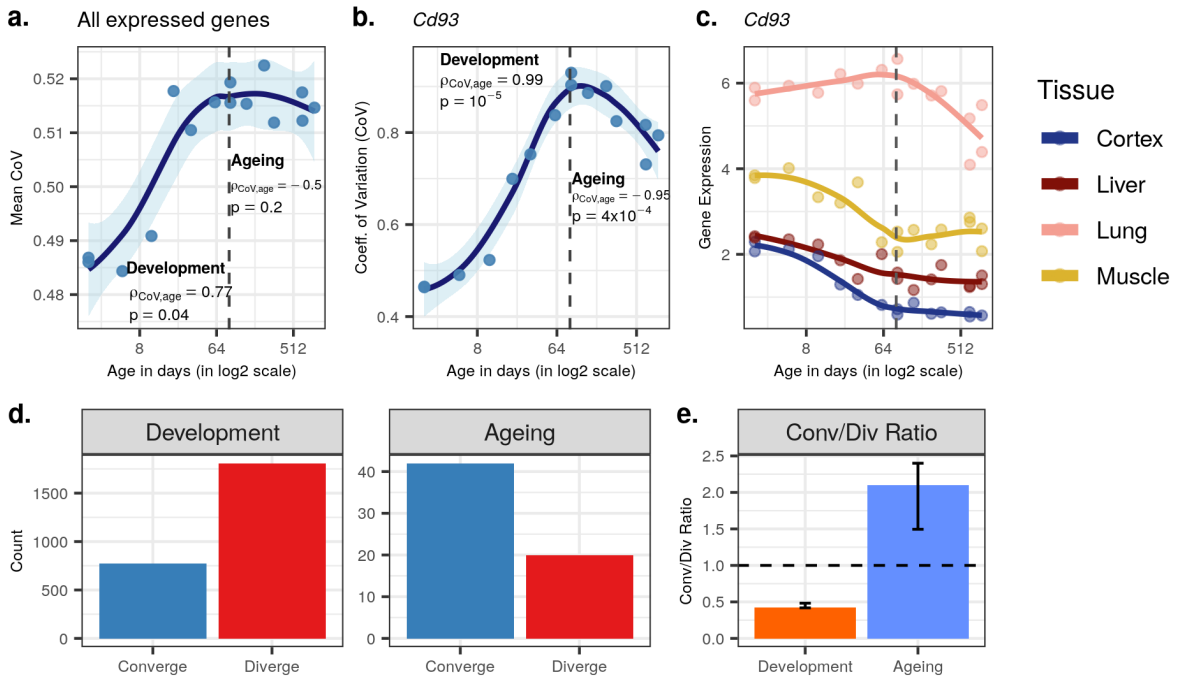

Figure 2. *The divergence-convergence (DiCo) pattern among tissues. a) Transcriptome-wide mean CoV trajectory with age. Each point represents the mean CoV value of all protein-coding genes (15,063) for each mouse except the one that lacks expression data in the cortex. b) Age effect on CoV value of the Cd93 gene which has the highest rank for the DiCo pattern, in four tissues (Methods). CoV increases during development and decreases during ageing, indicating expression levels show DiCo patterns among tissues. c) Expression trajectories of the gene Cd93. d) The number of significant CoV changes with age (FDR<0.1) during development (left, $n_{conv.}$=772, $n_{div.}$=1,809) and ageing (right, $n_{conv.}$=42, $n_{div.}$=20). Converge: genes showing a negative correlation (ρ) between CoV and age; Diverge: genes showing a positive correlation between CoV and age. e) The ratio of convergent/divergent genes in development and in ageing. The graph represents only genes showing significant CoV changes (at FDR<0.1, given in panel d). Error bars represent 95% confidence intervals calculated with the jackknife procedure.*

Inter-tissue convergence during ageing is a reproducible trend in another mouse dataset and in humans. We investigated inter-tissue convergence during ageing in another mouse ageing transcriptome dataset that comprises 5 tissue samples (brain, lung, liver, kidney, spleen) collected from the same mice (n=18) (Jonker et al. 2013). We confirmed the decrease in mean CoV during ageing (ρ = -0.58, p = 0.01) as well as the increased pairwise tissue correlations in 7/10 tissue pairs (**Figure S15**).

We next tested the inter-tissue convergence in GTEx dataset, which includes transcriptome data from humans covering a large set of tissues during the ageing period (GTEx Consortium et al. 2017). We first confirmed the results using 47 individuals with samples in the same four tissues and observed an age-related increased similarity between tissue samples of the same individual using PCA (ρ = -0.27, p = 2.2×10^{-15}, **Figure S17**). Mean and median CoV values ($\rho_{meanCoV}$ = -0.12, $p_{meanCoV}$ = 0.41; ρ_{medCoV} = -0.18, p_{medCoV} = 0.22) and pairwise inter-tissue correlations (5/6 comparisons positive, none was significant) also followed a trend compatible with inter-tissue convergence during ageing, albeit not significant **(Figure S18)**. We further tested the inter-tissue convergence using 10 tissues in GTEx, covering 35 samples from adipose, tibial artery, cerebellum, lung, skeletal muscle, tibial nerve, pituitary, sun-exposed skin, thyroid, and whole blood (see Methods). Similar to our previous results, inter-tissue convergence during ageing was observed using through PCA (ρ = -0.26, p = 6.7×10^{-7}, **Figure S20**), mean and median CoV change ($\rho_{meanCoV}$ = -0.13, $p_{meanCoV}$ = 0.47; ρ_{medCoV} = -0.29, p_{medCoV} = 0.089, **Figure S21**), and pairwise tissue correlations (29 positive, 16 negative correlations, none is significant after multiple testing (BY) correction). Notably, 8 of 16 negative correlations involved the

skin tissue (**Figure S21**).

These results suggest that inter-tissue convergence in ageing is a weak but widespread phenomenon, observed in multiple tissues in mice and humans. Overall, while mouse and human tissues display divergence in development (**Figures 1a, 2a**, (Cardoso-Moreira et al. 2019)) this is followed by convergence in ageing **(Figures 2a, S10-S21)**. This we refer to as the divergence-convergence (DiCo) pattern of postnatal lifetime.

The divergence-convergence (DiCo) pattern indicates loss of tissue-specificity during ageing. Potential explanations of DiCo pattern involve two scenarios consistent with the age-related loss of cellular identity: i) decreased expression of tissue-specific genes in their native tissues, or ii) non-specific expression of tissue-specific genes in other tissues. To test these predictions, we first identified tissue-specific gene sets based on relatively high expression of that gene in a particular tissue (cortex: 1,175, lung: 839, liver: 986, muscle: 766 genes). We noted that tissue-specific genes show clear up-down reversal patterns, being mostly up-regulated during development, and down-regulated during ageing (**Figure 3**, OR = [1.66, 4.41], $p<10^{-6}$, for each tissue, **Table S9**). Tissue-specific genes were also enriched among divergent-convergent genes (**Table S10**, OR=1.15, Fisher's exact test $p<0.001$).

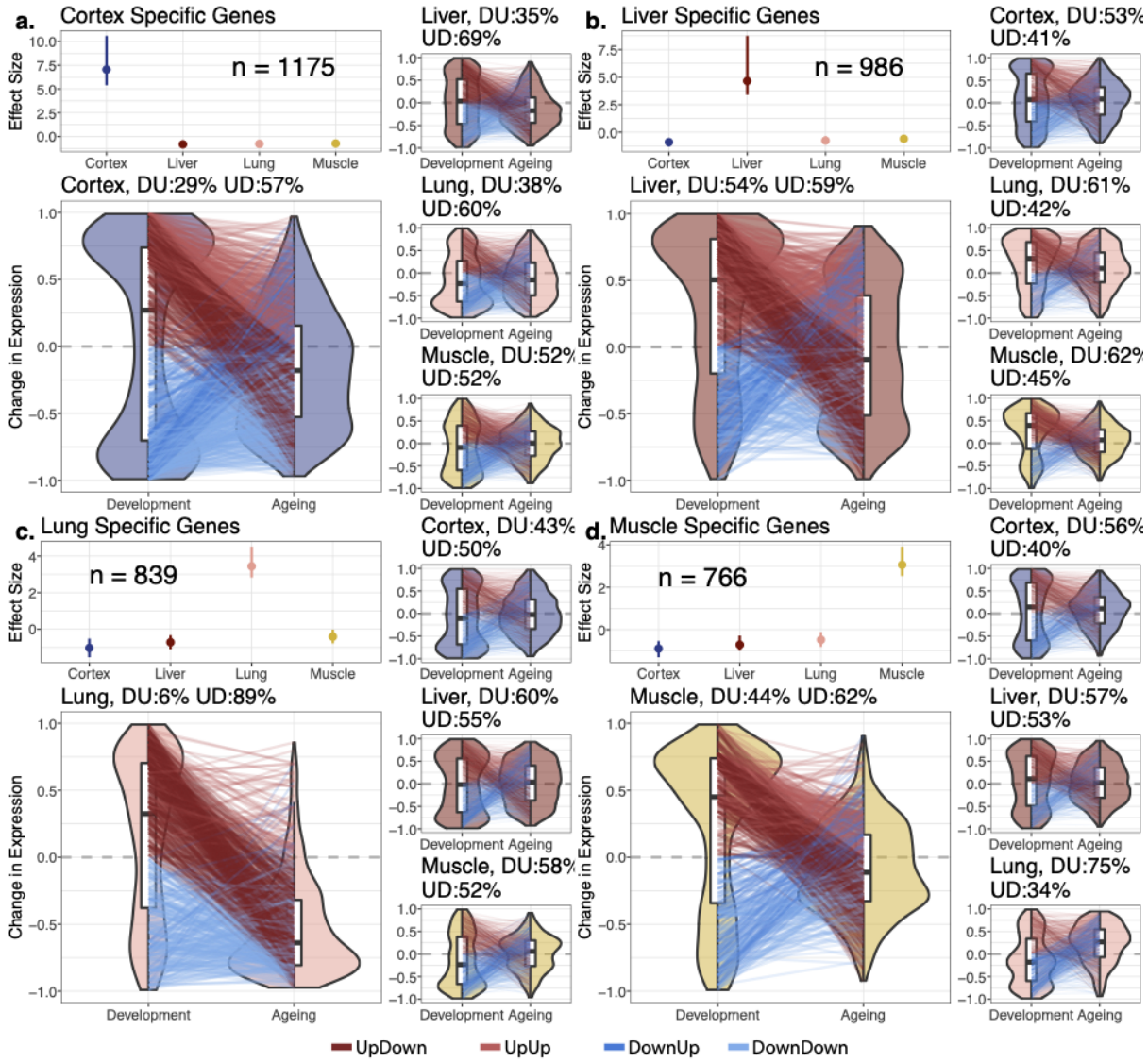

Figure 3: *Age-related expression changes of the tissue-specific genes. In each panel a-d, the upper left*

subpanels show effect size (ES) calculated with the Cohen's D formula, using expression levels of each gene among tissues (Methods). The IQR (linerange) and median (point) effect size for each tissue is shown. The number of tissue-specific genes is indicated inside each subpanel. The lower left subpanels show violin plots of the distribution of age-related expression change values (See Method) among tissue-specific genes, in development and in ageing. Each quadrant represents the plots for each tissue-specific gene group. The red and blue lines connect gene expression changes for the same genes in development and ageing. DU: percentage of down-up reversal genes among down-regulated, tissue-specific genes in development. UD: percentage of up-down reversal genes among up-regulated, tissue-specific genes in development. Tissue specific genes are enriched among UD reversal genes (Fisher"s Exact Test; OR_{cortex}*=1.66,* OR_{lung}*=4.41,* OR_{liver}*=1.47,* OR_{muscle}*=2.12,* $p<10^{-6}$ for each test).

We then tested our initial prediction that the DiCo pattern is related to tissue-specific genes losing their expression in their native tissue during ageing, or gaining expression in non-native tissues. We first tested the loss of tissue-specific expression during ageing by considering all tissue-specific genes and found a positive odds ratio (OR = 5.50, Fisher's exact test $p=2.1x10^{-129}$, **Figure 4a**). The same analysis with only the DiCo genes yielded a much stronger association (OR=74.81, Fisher's exact test $p=5.9x10^{-203}$, **Figure 4b**). This suggests that loss of tissue-specific expression is observed across the transcriptome, with a particularly strong association among DiCo genes.

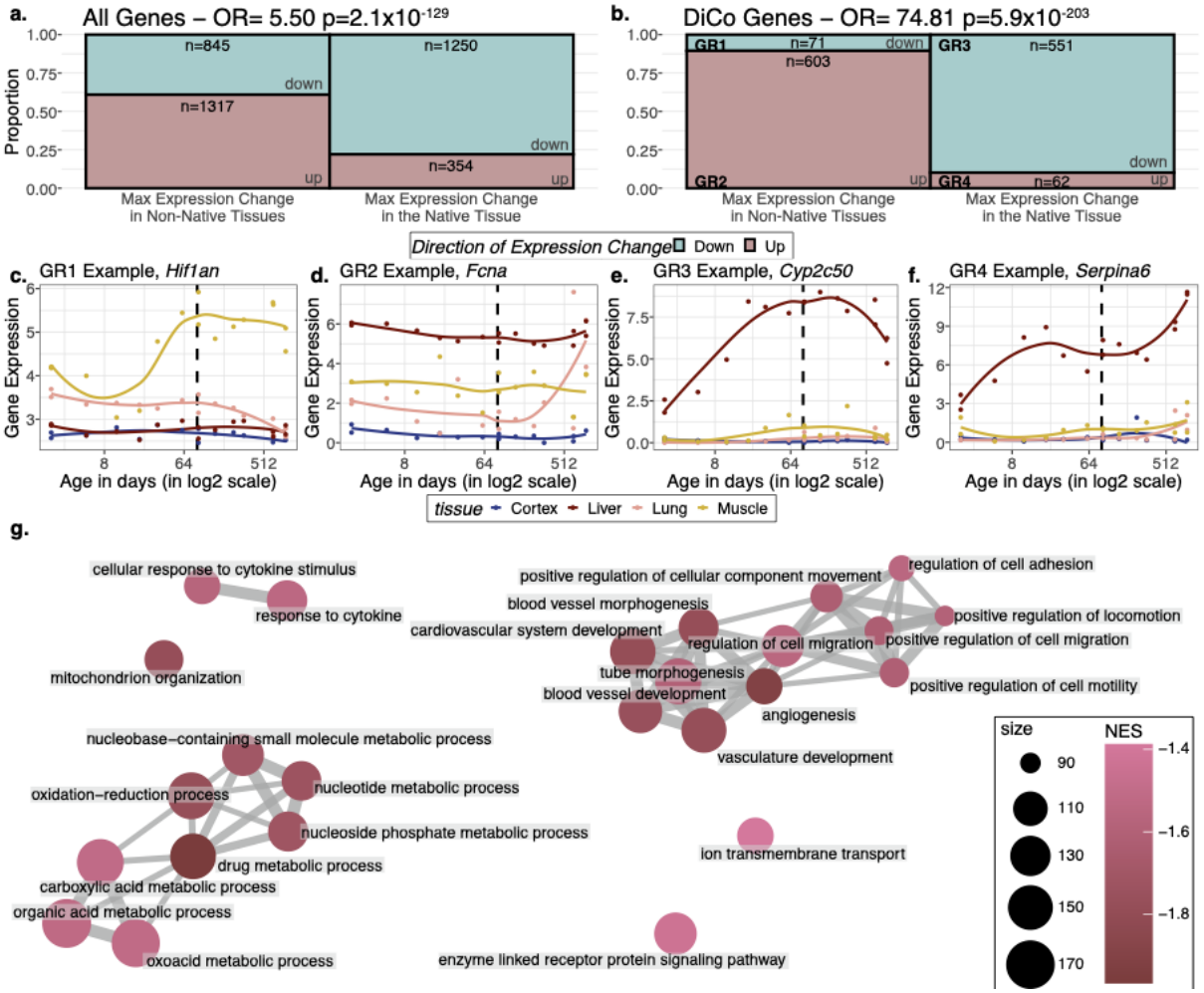

Figure 4. *a) Mosaic plot showing the association between maximal expression change in native vs. non-native tissues (x-axis) vs. down (cyan) or up (pink) regulation across all tissue-specific genes (n=3,766). The highly significant odds ratio indicates that genes native to a tissue tend to be down-regulated during ageing in that native tissue, if they show maximal expression change during ageing in that tissue. Conversely, if they show maximal expression change during ageing in non-native tissue, those genes are up-regulated during ageing. Consequently, tissue-specific expression patterns will tend to be lost. b) The same as (a) but using only the*

tissue-specific genes that show the DiCo pattern (n=1,287). The 4 groups are annotated as GR1-4 and gene expression changes for each group is exemplified in c-f. g) Functional enrichment results for the genes following DiCo pattern (n=4,802) among all divergent genes (n=9,058) during development. The nodes are GO Biological Processes (BP) and the edges show the similarity between the nodes. The shorter and thicker edges represent stronger similarity among categories measured with the Jaccard index. The size of each node shows the number of genes in categories. The more negative NES (normalized enrichment score) values, the stronger the enrichment for DiCo pattern. Only the 25 GO BPs chosen as representatives by the 'emapplot' function in enrichplot R package are shown. The full list is available in ***Table S11****.*

We then asked whether genes displaying the DiCo pattern may be related to specific functional pathways. Using GO, we searched for functional enrichment among convergent genes during ageing, using developmentally divergent genes as the background (Methods). We found enrichment for 237 GO **B**iological **P**rocess (BP) categories for the DiCo pattern (**K**olmogorov-**S**mirnov (KS) Test, BY-corrected p<0.1, **Table S11**). In addition to mitochondria, energy metabolism and other metabolic pathways, we also found categories related to tissue development and differentiation **(Figure 4g)**, consistent with the notion of ageing-related loss of cell identity.

Cellular composition changes and cell-autonomous expression changes can both explain the divergence-convergence pattern. Ageing-related transcriptome changes observed using bulk tissue samples may be explained by temporal changes in cell type proportions within tissues, by cell-autonomous expression changes, or both. To explore whether the observed inter-tissue DiCo patterns may be attributed to changes in cell type proportions, we used published data from a mouse single-cell RNA-sequencing experiment (Tabula Muris Consortium 2020). For each of the four tissues in our original experiment, we collected cell type-specific expression profiles from 3-month-old young adult mice in the Tabula Muris Senis dataset. We deconvoluted bulk tissue expression profiles in our mouse dataset using the corresponding tissue's cell type-specific expression profiles by regression analysis (Methods), and studied the relative contributions of each cell type to tissue transcriptomes and how these change with age. The analysis was performed with three gene sets; all genes (n=[12,492, 12,849]), DiCo (n=[4,007, 4,106]) and non-DiCo genes (n=[8,485, 8,743]). Studying these deconvolution patterns, we observed a weak but consistent trend involving the most common cell types in different tissues. For instance, analysing DiCo genes in the liver and lung, we found that the most common cell type's contribution (hepatocyte in the liver and bronchial smooth muscle cell in the lung) tends to increase during development (Spearman correlation ρ_{liver}=0.96, ρ_{lung}=0.81, nominal p<0.05). This contribution then decreases during ageing (ρ_{liver}=-0.77, ρ_{lung}=-0.86, nominal p<0.05) (**Figure 5a, S22**). This pattern was also observed in muscle and cortex, albeit not significantly (**Figure 5a, S22, Table S12**). These changes most likely reflect shifts in cellular composition demonstrated directly in mice using *in situ* RNA staining (Tabula Muris Consortium 2020). Repeating the analysis with non-DiCo genes resulted in highly similar patterns considering the most common cell types in tissues except in muscle ageing in which the age-related decrease was significantly higher with DiCo genes than the non-DiCo genes (permutation test with re-sampling all genes, $p_{skeletal\text{-}muscle\text{-}satllite\text{-}cell}$=0.04) **(Figure 5a, S22, S29)**. Hence, these results indicate that the observed cellular composition changes partly explain but not exclusive to DiCo pattern.

Next, we investigated the possible role of cell-autonomous changes in the DiCo pattern. Cell-autonomous changes could contribute to inter-tissue convergence during ageing in two ways. First, similar or corresponding cell types present in different tissues, such as immune cells, might become more similar with age in their expression profiles. Another possible scenario, consistent with the notion of age-related cellular identity loss, is that the expression profiles of unrelated cell types, such as tissue-specific cell types in different tissues, become more similar, *i.e.* converge, with age. To test these scenarios, we first ordered the pairwise correlations between cell types in different tissues to determine the most similar and dissimilar cell types across tissues (Methods). Then, we studied how these similarities (*i.e.* pairwise correlations) change with age **(Figure 5b)**. Intriguingly, we found that similar cell types (*i.e.* those with the highest correlations) among tissues become less similar with

age (36/54 of pairwise comparisons, **Table S13**). On the contrary, the most distinct cell types (*i.e.* those with the lowest correlations) among tissues become more similar with age (45/54, **Table S14**). Repeating the analysis considering the DiCo genes only yielded a similar trend (30/54 decrease in correlation among the most similar cell types, permutation test, p>0.1; and 47/54 increase in correlation among the most distinct cell types, permutation test, p>0.1). These results suggest that cell-autonomous changes also contribute to inter-tissue convergence during ageing and provide evidence for age-related cellular identity loss.

Finally, we tested the possibility of intra-tissue convergence of cell types in the Tabula Muris Senis dataset, by calculating expression variation among cell types using the CoV measure for each individual. However, we did not observe a consistent trend of increasing similarity among cell types within tissue from 3m- to 24m-old mice **(Figure S23)**.

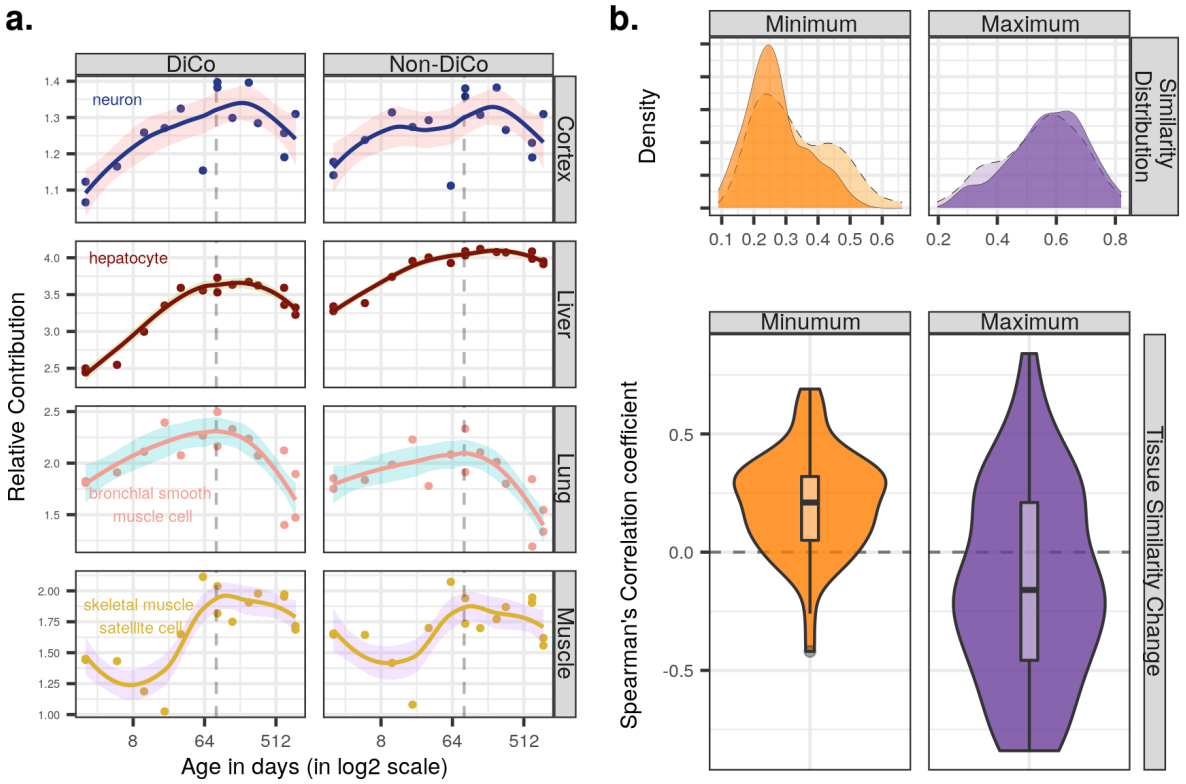

Figure 5. *a) Deconvolution analysis of our mouse dataset with the 3-month-old scRNA-seq data (Tabula Muris Senis) using DiCo* (n=[4,007, 4,106]) *and non-DiCo* (n=[8,485, 8,743]) *genes. Only the cell types with the highest relative contributions to each tissue bulk transcriptome are shown (cell type names are given within each plot). Contributions of all cell types to bulk tissue transcriptomes are shown in* ***Figure S22****. b) Distribution of correlations for minimally (left) and maximally (right) correlated cell type pairs among tissues (n=54 pairs). For each cell type of a given tissue, one minimally (or maximally) correlated cell type is chosen from other tissues among the 3-month age group of the Tabula Muris Senis dataset (density plots with solid line edges). Dashed lines show the correlation distributions in 24-months age of minimally or maximally correlated cell type pairs identified in the 3-months age group. Bottom panel shows age-related expression similarity (ρ) changes of minimally (left) and maximally (right) correlated cell type pairs. The correlation between age and tissue similarity (expression correlations) were calculated for each pair of cell types identified in the 3-months age group.*

Discussion

Our findings confirm a number of ageing-associated phenomena identified earlier, while also revealing new patterns. First, we report parallel age-related expression changes among the four tissues studied, during development, as well as in ageing. The inter-tissue correlation distributions were modest and

also comparable between development and ageing (**Figure 1c**). This last point may appear surprising given the stochastic nature of ageing relative to development (Bahar et al. 2006; Martinez-Jimenez et al. 2017; Angelidis et al. 2019; Somel et al. 2006), and also given earlier observations that developmental expression changes tend to be evolutionarily conserved, while ageing-related changes much less so (Zahn et al. 2007; Somel et al. 2010). Parallel expression changes among tissues during ageing may be related to common responses to damage and inflammation and processes such as reduced energy metabolism, as observed earlier (Zahn et al. 2007).

Second, we verify the generality of the reversal pattern, *i.e*. up-down or down-up expression change patterns across the lifetime, among distinct mouse tissues that include both highly mitotic (lung and liver) and less mitotic ones (skeletal muscle and cortex). Consistent with earlier observations in fewer tissues (Anisimova et al. 2020; Dönertaş et al. 2017), we find that about half the expressed genes display reversal in all cases studied. We thus propose that reversal is a general phenomenon in mammalian ageing.

Two observations are notable. One is that reversal-displaying genes, especially those displaying the up-down pattern in each tissue, can be associated with tissue-specialization-related pathways (*e.g.* morphogenesis) and tissue-specific functions (*e.g.* synaptic activity). The second is the lack of significant overlap among reversal genes among tissues. These imply that the parallel changes among tissues observed during development and ageing, described earlier, do not include reversals. Therefore, we hypothesised that reversals might instead be linked to inter-tissue divergence and tissue specialisation during development, and possibly convergence during ageing.

We indeed observed that the tissue-specific genes show an up-down reversal pattern in their native tissue. Studying inter-tissue similarity across mouse lifespan, we further found that the four tissues' transcriptomes diverged during postnatal development, and we further detected a trend, albeit weak, towards inter-tissue convergence during ageing. Studying two additional datasets from mice and humans, we observed that convergence during ageing is common, although possibly not ubiquitous. Importantly, these results reliably show that developmental inter-tissue divergence does not continue into ageing. We predict that future work on ageing with datasets, including multiple distinct tissues and larger sample sizes may establish the prevalence of the inter-tissue divergence-convergence phenomenon.

Finally, we report three interesting observations related to inter-tissue divergence-convergence. We determine that tissue-specific genes tend to be down-regulated in the tissues that they belong to during ageing, while non-tissue-specific genes are up-regulated. Second, using deconvolution, we infer that cell types most common in a tissue (*e.g.* hepatocytes in the liver) tend to increase in frequency during development, but then decrease in frequency during ageing, as also shown recently using immunohistochemistry in a number of mouse tissues (Tabula Muris Consortium 2020). Accordingly, the DiCo phenomenon can at least partly be explained by shifts in cellular composition. This is intriguing as both highly mitotic and low mitotic tissues share this trend, indicating that an explanation based on stem cell exhaustion may not be applicable here. Third, during ageing, we find growing expression similarity between normally distinct cell types in different tissues, but decreased similarity between normally similar cell types. Cell-autonomous expression changes, therefore, may also contribute to the divergence-convergence phenomenon.

These results support a model where expression reversals and the newly discovered divergence-convergence phenomenon are associated with loss of specialisation, at the tissue level and possibly also at the cellular level.

Materials and Methods

Sample Collection

We collected time-series 16 male C57BL/6J mice samples representing the whole lifespan of *Mus musculus,* comprising both postnatal development and the ageing period from four different tissues; cerebral cortex, liver, lung and skeletal muscle. One 904 days-old mouse had no cortex tissue sample, and thus excluded from the analysis. As a result, we have generated 63 RNA-seq libraries in total.

Separation of development and ageing periods:

In order to compare the gene expression changes during postnatal development and ageing we studied the samples before sexual maturation (covering 2 to 61 days of age , n=6) as postnatal development period, and samples covering 93 to 904 days (n=8 in all except in cortex n=7) as the ageing period.

RNA-Seq Library Preparation

RNA sequencing was performed as previously described (Liu et al. 2016) with slight modifications. Briefly, total RNA was extracted using Trizol reagent (Invitrogen). For sequencing library construction, all samples were randomized first and prepared with TruSeq RNA Sample Preparation Kit (Illumina) according to the manufacturer's instruction. Libraries were then sequenced on the Illumina Hi-seq 4000 system in three lanes within one flow-cell, using the 150-bp paired-end module.

RNA-Seq Data Preprocessing

The quality assessment of the raw RNA-seq data was done using FastQC (Andrews 2010) and the adapters were removed using Trimmomatic (Bolger, Lohse, and Usadel 2014). The low-quality reads were filtered using the parameters: PE ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:1:0:8:true, SLIDINGWINDOW:4:15, MINLEN:25. Remaining high-quality reads were aligned to the reference genome GRCm38 using STAR-2.5.3 (Dobin et al. 2013) with parameters: --sjdbOverhang 99 --outSAMattrIHstart 0 --outSAMstrandfield intronMotif --sjdbGTFfile GRCm38.gtf. Percentage of uniquely mapped reads in libraries ranged from 80 to 93%. We used cufflinks (Trapnell et al. 2010) to generate read counts for uniquely aligned reads (samtools -q 255 filter) and calculated expression levels as fragment per kilobase million (FPKM). There were 50 duplicated genes, and the sum of their FPKM values were used. In total we had 51,608 genes. We restricted the whole analysis to only protein-coding genes obtained by the 'biotype' feature of the biomaRt library (Durinck et al. 2009). We also excluded genes which were not detected (zero FPKM) in 25% or more of the samples (at least 15 of 63), resulting in 15,063 protein-coding genes in total. Using all the samples together, FPKM values were log2 transformed (after adding 1) and quantile normalized with 'normalize.quantiles' function from 'preprocessCore' library (Bolstad 2020) to be used in downstream analysis.

Principal component analysis:

We estimated the source of variation in the data using Principal component analysis (PCA) on the scaled expression matrix with 'prcomp' function in the R base. The first four components, from PC1 to PC4 in order, explained 31%, 20%, 17% and 8% of the total variance with a clear separation of tissues in PC1 and PC2 and a strong age effect in PC4. We performed ANOVA between each PC1-3 and tissue labels to confirm tissue differences. The magnitude of age effect on PCA analysis was measured with Spearman correlation between individual age and each principal component. We further calculated euclidean distance in pairwise manner among tissues of each individual in PC3-4 space constructed using only the developmental samples. Then, we tested the effect of age on mean euclidean distance among tissues using Spearman correlation.

Age-related gene expression change

To identify genes showing age-related expression change in each tissue, we used Spearman correlation between individual age and expression level, separately for development and ageing

periods. To capture potential non-linear but monotonic changes in expression, we chose the non-parametric Spearman correlation test for both periods. Significance of age-related genes was assessed with FDR<0.1 cutoff, calculated with the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg 1995) using the 'p.adjust' function in the R base library.

Functional associations:
We tested the functional associations of age-related gene expression change in separate tissues for each period separately, following gene set over-representation analysis (GORA) procedure with Gene Ontology (GO) (Ashburner et al. 2000) Biological Process (BP) categories using the 'topGO' package (Alexa and Rahnenfuhrer 2019). We applied the classical algorithm and performed Fisher's Exact Test on categories that satisfy the criteria of a minimum 10 and maximum 500 number of genes. We used the whole set of expressed genes (n=15,063) as the background. P-values were corrected for multiple testing using Benjamini-Yekutieli (BY) (Benjamini and Yekutieli 2001) procedure, which does not assume independent tests and thus is suitable for multiple testing correction for GO enrichment with potentially overlapping gene sets. Categories with BY corrected p-value<0.1 were considered as significant.

Correlation between age-related gene expression changes in different tissues
We calculated Spearman correlation coefficients between age-related gene expression change values in each tissue pair (ρ from the correlation between gene expression levels and age) (**Figure 1c**). In order to test the statistical significance of the correlations we used a permutation scheme as the expression levels across tissues are not independent but of the same mice. In order to account for the dependence, the individual ages were permuted but the permuted values kept the same across tissues (similar to permutation tests applied in (Dönertaş et al. 2017; Işıldak et al. 2020; Dönertaş et al. 2018)). Specifically, we randomized the individual ages for 1000 times, in development and ageing separately, using the 'sample' function in R, while keeping the permuted age labels the same for individuals across tissues. We calculated the age-related gene expression changes with permuted ages, thus simulating the null distribution with no age effect.

We then calculated the Spearman correlation between the age-related gene expression levels from the permutations across tissues and we assigned the p value by calculating the proportion of permuted calculations with a more extreme correlation. FPR or false positive rate was calculated as the median value of expectations divided by the observed value (**Table S3**).

Shared gene expression changes across tissues:
We summarized the number of shared age-related genes among tissues in both development and ageing and for up- and down-regulated genes separately, using FDR<0.1 (**Figure S5**). For each gene, we counted the number of tissues with the same direction of expression change with age. We also calculated the same overlap among tissues with all genes without using any significance cutoff (**Figure 1e**).

Permutation test:
We again used a permutation scheme to assess the significance of shared age-related genes to account for the dependence between tissues. We tested the significance of shared up- and down-regulated genes, selected with or without FDR cutoff, and development and ageing periods separately. We used the permuted age-related expression changes that are described before for the test of correlations. To test the significance of the overlap of significantly up- or down- regulated genes (FDR<0.1), we ranked the permuted age-related expression changes (ρ) for each tissue in each period separately, and chose the highest N (to test the up-regulation), or lowest N (to test the down-regulation) number of genes, where N is the number of significantly up- or down- regulated genes in a given tissue (FDR<0.1). We then calculated the number of overlaps across tissues and thus generated a distribution of shared expression changes when there is no age effect. We calculated the p value as the number of the permuted values with a higher number of overlap than the

observed value divided by the number of permutations (n=1,000). FPR was calculated as the median number of shared expression changes in permutations divided by the observed value.

Functional Associations:
We tested the functional associations of shared expression change trends among tissues in each period, separately, following the GORA procedure using the same criteria and algorithms explained in the previous section. To test shared up-regulated (n=45) or down-regulated genes (n=138) in development, we chose all significant age-related genes across tissues (n=10,305) in the development period. Since we could not identify any shared ageing-related genes across tissues (**Figure S5**), we did not perform a functional test for the ageing period.

Analysis of gene expression reversals
We compared the direction of gene expression change during development and ageing to identify reversal genes in each tissue separately. Genes showing up-regulation (positive correlation with age) in development and down-regulation (negative correlation with age) in ageing were assigned as up-down (UD) reversal genes, while the genes with the opposite trend (down-regulation in development and up-regulation in ageing) were assigned as down-up (DU) reversal genes. Without using any significance level for expression-age correlation values, we calculated the proportion of genes showing reversal by keeping the expression change direction in development the same, *i.e.* UD%=UD/(UU+UD) and DU%=DU/(DD+DU).

Permutation test:
To test the significance of reversal proportions, we kept the developmental changes constant and randomly permuted the individual ages only in the ageing period (as described earlier). Among developmental up-regulated genes, we calculated an UD% from each permutation, simulating a null distribution for UD reversal. We applied the same principle for the DU genes. Thus, we created a null distribution with the expected reversal ratios and tested the significance of observed values for each tissue separately.

Functional associations:
We used the GORA procedure as described earlier but kept the developmental changes constant in the background. More specifically, we tested the functional enrichment of UD reversal genes against UU genes, and DU genes against DD genes to keep the developmental changes fixed and test the reversal pattern not developmental associations.

Overlap of reversal genes - permutation test
We tested the significance of overlap using the same permutation scheme described above. Specifically, among developmental up- (or down-) regulated genes shared among tissues (n_{up}=2,255, n_{down}=2,209), we constructed the null distribution by calculating the ratio of UD vs UD+UU (or DU vs DU+DD) genes shared among tissues in ageing permutations (**Figure S8**).

Tissue convergence and divergence calculations using coefficient of variation (CoV)
For each individual mouse and gene (n=15,063), we calculated a coefficient of variation (CoV) with normalized expression levels from the four tissues by dividing the standard deviation by the mean. We studied the (inter-tissue) expression-variation change with age in development and ageing periods separately, using two approaches.

Mean/median CoV across genes:
We assessed the global variation among the tissues of each individual mouse by calculating the mean (or median) CoV of genes and then using Spearman correlation test between mean-CoV (or median-CoV) and individual age.

CoV at the gene level
In the second approach, we tested each commonly expressed gene using Spearman correlation test between the CoV value of a gene and age. P-values were corrected for multiple testing, using the 'BH' procedure and the significance was assigned to corrected p-value<0.1. The genes showing positive correlation between CoV and age were called divergent, and the ones showing negative correlation between CoV and age were called convergent **(Figure 2d)**. Genes that show divergent pattern during development and convergent pattern in ageing (without using a significance level) were called divergent-convergent (DiCo) genes (n=4,802).

Permutation Test
To test the significance of DiCo genes (n=4,802), we kept the developmental divergent genes constant (n=9,058, without a significance cutoff) and randomly permuted the individual ages only in the ageing period (as described earlier). Among developmental divergent genes, we calculated DiCo% for each permutation, simulating a null distribution for DiCo pattern (**Figure S16**).

DiCo Genes Clustering
We used the k-means algorithm to cluster DiCo genes according to their CoV or expression changes with age, separately. To find the optimum number of clusters for both procedures, we applied gap statistics using the 'clusGap' function in the 'cluster' package with 500 simulations (Tibshirani, Walther, and Hastie 2001). We used the 'kmeans' function in base R with 'iter.max=20' and 'nstart=50' parameters to cluster CoV values or expression levels which were standardized to mean=1 and sd=0 across genes.

Functional association:
To test the functional associations of the genes showing divergent-convergent pattern among tissues, we performed GSEA using GO BPs. We retrieved developmental divergent genes (CoV-age ρ>0, n=9,058) and multiplied these correlation values with the ones calculated in the ageing period. Therefore, the genes with a negative value represent a divergent-convergent pattern quantitatively, while the ones with a positive product represent a divergent pattern throughout lifetime. We then ranked the genes according to the calculated product values and sought enrichment for the upper and lower tail of the distribution using the Kolmogorov-Smirnov (KS) test implemented in the 'clusterProfiler' package (Yu et al. 2012). 'gseGO' function was used with parameters: nPerm=1000, minGSSize=10, maxGSSize=500 and pValueCutoff=1. Therefore, the enriched categories for the genes in the lower tail of the distribution would represent DiCo enrichment.

We further sought enrichment among DiCo genes that were clustered with the k-means algorithm in the previous step for both CoV and expression clusters, separately. Genes in each cluster were tested among all DiCo genes using the same GORA procedure as described before.

Jackknife to test the Di/Co ratio between dev and ageing
We tested the significance of divergent/convergent gene ratios using a jackknife resampling procedure in development and ageing periods, separately. Leaving out an individual in each iteration, we re-calculated the number of significant divergent and convergent genes and their ratios, denoted as s_i. Then, we constructed a 95% confidence interval using the following formula:

CI(95%) = $\bar{s} \pm 2 * \sqrt{\frac{var}{n}}$ where,

var is the variance of divergent/convergent ratios (s_i) in leave-one-out resamples and $\bar{s}$ is the mean of s_i values.

Pairwise tissue divergence-convergence test
In order to further verify the inter-tissue divergent-convergent pattern that we observed between

development and ageing periods, we used another approach based on expression correlations among tissues. We calculated pairwise Spearman correlation coefficients among tissues of the same individual mouse using all commonly expressed genes. For each tissue pair, we tested the effect of age on expression-correlations using Spearman correlation test in development and ageing periods, separately. We then calculated the mean (or median) of all six pairwise tissue correlations for each individual mouse (**Figure S13**). The same procedure was applied to other imported datasets (**Figure S15, S18**).

Determination of tissue-specific genes

To identify which tissue(s) contribute to the reversal pattern, we assigned each gene to a tissue to identify tissue-specific expression patterns. First, we calculated an effect size (ES) between the expression of a gene in a tissue versus other three tissues using the development samples only, and repeated this procedure for all tissues. Hence, we obtained ES for each commonly expressed gene in each tissue. ES was calculated using the 'Cohen's d' formula defined as the difference between the two means divided by the pooled standard deviation. We then assigned each gene to a tissue in which the gene has the highest ES. Finally, we retrieved only the fourth quartile (>Q3) of genes assigned to a tissue to define tissue-specific expression. Using this approach, we found 3,766 tissue-specific genes in total (cortex: 1,175, lung: 839, liver: 986, muscle: 766 genes).

Enrichment test with the direction of age-related change

We tested the association between tissue-specificity and the age-related change in expression during ageing using Fisher's exact test. Specifically, we constructed the contingency table with two groups; the first group defines the direction of maximum expression change of each tissue specific gene (being either positive or negative), the second group defines whether this maximum expression change a tissue specific gene occurs in its native tissue or not (either yes or no). Hence, a test result where a positive OR suggest either the expression of genes decrease the most in their native tissue or increase the most in non-native tissues.

Enrichment in DiCo pattern

We tested the association between tissue-specificity and DiCo pattern using Fisher's exact test, calculating the enrichment of tissue-specific genes within DiCo genes.

Additional publicly available bulk tissue transcriptome datasets

Jonker

We downloaded the raw data from the GEO database with GSE34378 accession number (Jonker et al. 2013) and followed the same analysis pipeline described above using all the samples from 5 tissues ("Brain - Cortex", "Lung", "Liver", "Kidney", "Spleen") of 18 mouse comprising 90 samples in total. This dataset represents ageing period of the mouse ranging from 90 to 900 days. Using the oligo package (Carvalho and Irizarry 2010), we retrieved the expression matrices and removed the probesets that are annotated to more than one gene. We confined the analysis to only the protein-coding genes expressed in at least 25% of all samples. Resulting 16,647 genes were log2 transformed (after adding 1) and quantile normalized using the preprocessCore library (Bolstad 2020) across all samples. Downstream analysis was the same as described above.

GTEx

We downloaded GTEx v8 data (GTEx Consortium et al. 2017) from the data portal and repeated the analysis in human tissues. We first confirmed our results in the same 4 tissues ("Brain - Cortex", "Lung", "Liver", "Muscle - Skeletal") and then expanded the analysis to 10 tissues ("Adipose - Subcutaneous", "Artery - Tibial", "Brain - Cerebellum", "Lung", "Muscle - Skeletal", "Nerve - Tibial", "Pituitary", "Skin - Sun Exposed (Lower leg)", "Thyroid", "Whole Blood"). In order to choose which tissues to analyse, we first choose the minor tissues with the highest number of samples for each major tissue, which prevents the representation of the same tissue multiple times. We then performed

hierarchical clustering of tissues based on the presence of samples from the same individuals (**Figure S19**) and cut the tree into 3 clusters based on visual inspection. We selected the cluster with the highest number of overlapping individuals to analyse. The same procedure was followed for both 4- and 10-tissue analyses. In particular, we restricted the analysis to the individuals with samples in all tissues analysed and with a death circumstance of 1 (violent and fast deaths due to an accident) and 2 (fast death of natural causes) on the Hardy Scale (n = 47 for 4 tissue, n = 35 for 10 tissue). We removed duplicated genes from the analysis. Similar to our analysis with the mice data, we used only the protein-coding genes that are expressed in at least 25% of all samples, totalling 16,211 for 4 tissues and 16,305 for 10 tissues. The TPM values obtained from the GTEx data portal are log2 transformed (after adding 1), and quantile normalised using the preprocessCore library (Bolstad 2020) in R. Downstream analysis was the same as other datasets.

Single-cell RNAseq

Preprocessing

We used Tabula Muris Senis dataset (Tabula Muris Consortium 2020) for scRNA-seq analysis as it is the only dataset to our knowledge that includes time-series samples covering old age, and the tissues present in our dataset. Seurat processed FACS data of the tissues: lung, liver, skeletal muscle and non-myeloid brain, were downloaded from figshare database (Pisco 2020). Seurat package (Stuart et al. 2019) was used to retrieve the expression matrix of the cells that are annotated to cell types in the original article. Each tissue contains samples from three time points; 90 (3m), 540 (18m) and 720 (24m) days old mice, totaling 14 samples each in lung, liver and brain and 9 samples in liver. We excluded cell types with less then 15 cells among all samples, and excluded genes if the expression level is 0 for all cells at a given age. This resulted in a median number of 99-382 cells assigned to cell types, 6-24 cell types and 16,951-22,122 genes across tissues. Using 3 month old mice, we calculated cell type specific expressions in each tissue such that we first calculated the mean expression levels among cells of an individual mouse for each cell type and then calculated the mean among individuals to obtain an average expression value for each cell type. Uniprot gene symbols were converted to Ensembl gene ids using biomaRt R package (Durinck et al. 2009).

Deconvolution

We used cell type specific expression profiles of 3-month-old mice to estimate relative contributions of cell types to the transcriptome profiles of tissues in our mouse dataset. We used a linear-regression based deconvolution method for each tissue using three genesets; all genes (n=[12,492, 12,849]), DiCo genes (n=[4,007, 4,106]) and non-DiCo genes (n=[8,485, 8,743]). Regression coefficients were used as relative contribution of cell types. We then tested the effect of age on cell type contributions using Spearman correlation test in development and ageing.

Cell type similarity & their change with age

To investigate the contribution of cell autonomous changes to inter-tissue convergence in ageing, we calculated pairwise cell type expression correlations among tissues and studied how these correlations change with age. Based on pairwise correlations in the 3-months age group, we identified the maximally and minimally correlated cell type pairs among tissues. Specifically, for each cell type in a given tissue (*e.g.* 10 cell types *in* the liver), we chose the minimally correlated cell types from the other three tissues (*e.g.* choose the minimally correlated cell type among 15 cell types in the cortex, choose the minimally correlated one among 24 cell types in the lung, and the same procedure for 6 cell types in the muscle*)* and repeated this procedure for the other three tissues, resulting in 54 cell type pairs. Then, we calculated Spearman correlation coefficients between age and minimally correlated cell type pairs identified in the 3-months age group. Likewise, we repeated the same analysis for the maximally correlated cell type pairs among tissues.

Permutation Tests:

To test whether DiCo genes are significantly more associated with cell type proportion changes than

non-DiCo genes, we performed a permutation test based on re-sampling procedure. For each tissue, we took random samples among all genes (n=[12,492, 12,849]) with size N (N being the number of DiCo genes in that tissue), and repeated deconvolution as explained above. By calculating cell type proportion changes with age for each random sample repeated 1000 times, we created the null distribution for each cell type. Then, we calculated the p values as the number of random samples having the same or higher cell type proportion change values divided by the observed value (cell type proportion changes with DiCo genes).

We applied a similar permutation scheme explained above to test cell type similarity change differences between DiCo and non-DiCo genes. For each random sample of non-DiCo genes with size N, we calculated the pairwise correlations among cell types of tissues and identified maximally and minimally correlated cell types in the 3-months age group. Then, we calculated age-related changes of those correlations using Spearman correlation to construct null distribution.

Analysis of within tissue convergence of cell types

Analogous to inter-tissue convergence analysis, we also studied intra-tissue convergence of cell types in scRNA seq data by calculating CoV among cell types within a tissue for each individual in 3m, 18m and 24m separately. We filtered the data to obtain cell types present in at least 2 individual mice in every time point for each tissue which yielded 4, 7, 20 and 6 cell types in brain, liver, lung and muscle, respectively. We then tested the mean CoV (or CoV per gene) change with age using Spearman correlation test.

Ethics Statement

Post-mortem samples were obtained from 16 C57BL/6J mice aged between 2 days and 904 days, housed in groups and fed under standard conditions in the MPI-EVA Animal Facility (Leipzig, Germany). The mice were sacrificed for reasons independent of this study, their tissues were harvested and frozen immediately, and stored at -80°C.

Competing Interests

The authors report no competing interests.

Data Availability

Raw and processed RNA-seq data have been deposited in GEO with the accession number GSE167665. All summary statistics and analysis output are also provided as supplementary tables.

Code Availability

All the code used to perform analyses is available in GitHub: https://github.com/hmtzg/geneexp_mouse

Author Contributions

H.M.D, M.S. and P.K. conceived and conceptualized the study. H.I, H.M.D, and M.S. design the models to analyze the data. H.M.D validated the results. H.I. analysed the data with contributions by U.I., H.M.D, and E.K. D.D.H conducted the experiments and produced the raw data. S.H. helped D.D.H. with experimental procedures. H.I., U,I., H.M.D. curated and visualised the data. H.I., H.M.D., and M.S. wrote the manuscript. M.S and P.K. acquired funding for the experiments. H.M.D., M.S., and P.K. supervised and administered the study. All authors read, revised and approved the final version of this manuscript.

Acknowledgement

We thank Dr Nurcan Tuncbag, Dr Nihal Terzi Çizmecioğlu, and the whole METU CompEvo team for helpful comments and fruitful discussions, and Gozde Turan and Melih Yıldız for the critical reading of the manuscript and their suggestions.

Funding Statement

This work was supported by EMBL (H.M.D), the Scientific and Technological Research Council of Turkey (TÜBİTAK 2232, M.S.), the Science Academy (Turkey) BAGEP Award (M.S.), by METU internal grant (BAP).

References

Alexa, Adrian, and Jorg Rahnenfuhrer. 2019. “topGO: Enrichment Analysis for Gene Ontology.”

Andrews, Simon. 2010. “FastQC: A Quality Control Tool for High Throughput Sequence Data.” 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Angelidis, Ilias, Lukas M. Simon, Isis E. Fernandez, Maximilian Strunz, Christoph H. Mayr, Flavia R. Greiffo, George Tsitsiridis, et al. 2019. “An Atlas of the Aging Lung Mapped by Single Cell Transcriptomics and Deep Tissue Proteomics.” *Nature Communications* 10 (1): 963. https://doi.org/10.1038/s41467-019-08831-9.

Anisimova, Aleksandra S., Mark B. Meerson, Maxim V. Gerashchenko, Ivan V. Kulakovskiy, Sergey E. Dmitriev, and Vadim N. Gladyshev. 2020. “Multifaceted Deregulation of Gene Expression and Protein Synthesis with Age.” *Proceedings of the National Academy of Sciences of the United States of America*, June. https://doi.org/10.1073/pnas.2001788117.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, et al. 2000. “Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium.” *Nature Genetics* 25 (1): 25–29. https://doi.org/10.1038/75556.

Bahar, Rumana, Claudia H. Hartmann, Karl A. Rodriguez, Ashley D. Denny, Rita A. Busuttil, Martijn E. T. Dollé, R. Brent Calder, et al. 2006. “Increased Cell-to-Cell Variation in Gene Expression in Ageing Mouse Heart.” *Nature* 441 (7096): 1011–14. https://doi.org/10.1038/nature04844.

Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 57 (1): 289–300. http://www.jstor.org/stable/2346101.

Benjamini, Yoav, and Daniel Yekutieli. 2001. “The Control of the False Discovery Rate in Multiple Testing under Dependency.” *Annals of Statistics* 29 (4): 1165–88. https://doi.org/10.1214/aos/1013699998.

Blagosklonny, Mikhail V. 2013. “Aging Is Not Programmed: Genetic Pseudo-Program Is a Shadow of Developmental Growth.” *Cell Cycle* 12 (24): 3736–42. https://doi.org/10.4161/cc.27188.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Bolstad, Ben. 2020. “preprocessCore: A Collection of Pre-Processing Functions.” https://github.com/bmbolstad/preprocessCore.

Brawand, David, Magali Soumillon, Anamaria Necsulea, Philippe Julien, Gábor Csárdi, Patrick Harrigan, Manuela Weier, et al. 2011. “The Evolution of Gene Expression Levels in Mammalian Organs.” *Nature* 478 (7369): 343–48. https://doi.org/10.1038/nature10532.

Cardoso-Moreira, Margarida, Jean Halbert, Delphine Valloton, Britta Velten, Chunyan Chen, Yi Shao, Angélica Liechti, et al. 2019. “Gene Expression across Mammalian Organ Development.” *Nature* 571 (7766): 505–9. https://doi.org/10.1038/s41586-019-1338-5.

Carvalho, Benilton S., and Rafael A. Irizarry. 2010. “A Framework for Oligonucleotide Microarray Preprocessing.” *Bioinformatics* 26 (19): 2363–67. https://doi.org/10.1093/bioinformatics/btq431.

Colantuoni, Carlo, Barbara K. Lipska, Tianzhang Ye, Thomas M. Hyde, Ran Tao, Jeffrey T.

Leek, Elizabeth A. Colantuoni, et al. 2011. “Temporal Dynamics and Genetic Control of Transcription in the Human Prefrontal Cortex.” *Nature* 478 (7370): 519–23. https://doi.org/10.1038/nature10524.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. “STAR: Ultrafast Universal RNA-Seq Aligner.” *Bioinformatics* 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dönertaş, Handan Melike, Matías Fuentealba Valenzuela, Linda Partridge, and Janet M. Thornton. 2018. “Gene Expression-Based Drug Repurposing to Target Aging.” *Aging Cell* 17 (5): e12819. https://doi.org/10.1111/acel.12819.

Dönertaş, Handan Melike, Hamit İzgi, Altuğ Kamacıoğlu, Zhisong He, Philipp Khaitovich, and Mehmet Somel. 2017. “Gene Expression Reversal toward Pre-Adult Levels in the Aging Human Brain and Age-Related Loss of Cellular Identity.” *Scientific Reports* 7 (1): 5894. https://doi.org/10.1038/s41598-017-05927-4.

Durinck, Steffen, Paul T. Spellman, Ewan Birney, and Wolfgang Huber. 2009. “Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor Package biomaRt.” *Nature Protocols* 4 (8): 1184–91. https://doi.org/10.1038/nprot.2009.97.

Flurkey, Kevin, Joanne M. Currer, and D. E. Harrison. 2007. “Chapter 20 - Mouse Models in Aging Research.” In *The Mouse in Biomedical Research (Second Edition)*, edited by James G. Fox, Muriel T. Davisson, Fred W. Quimby, Stephen W. Barthold, Christian E. Newcomer, and Abigail L. Smith, 637–72. Burlington: Academic Press. https://doi.org/10.1016/B978-012369454-6/50074-1.

GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, et al. 2017. “Genetic Effects on Gene Expression across Human Tissues.” *Nature* 550 (7675): 204–13. https://doi.org/10.1038/nature24277.

Işıldak, Ulaş, Mehmet Somel, Janet M. Thornton, and Handan Melike Dönertaş. 2020. “Temporal Changes in the Gene Expression Heterogeneity during Brain Development and Aging.” *Scientific Reports* 10 (1): 4080. https://doi.org/10.1038/s41598-020-60998-0.

Jonker, Martijs J., Joost Pm Melis, Raoul V. Kuiper, Tessa V. van der Hoeven, P. F. K. Wackers, Joke Robinson, Gijsbertus Tj van der Horst, et al. 2013. “Life Spanning Murine Gene Expression Profiles in Relation to Chronological and Pathological Aging in Multiple Organs.” *Aging Cell* 12 (5): 901–9. https://doi.org/10.1111/acel.12118.

Liu, Xiling, Dingding Han, Mehmet Somel, Xi Jiang, Haiyang Hu, Patricia Guijarro, Ning Zhang, et al. 2016. “Disruption of an Evolutionarily Novel Synaptic Expression Pattern in Autism.” *PLoS Biology* 14 (9): e1002558. https://doi.org/10.1371/journal.pbio.1002558.

Magalhães, João Pedro de, and George M. Church. 2005. “Genomes Optimize Reproduction: Aging as a Consequence of the Developmental Program.” *Physiology* 20 (August): 252–59. https://doi.org/10.1152/physiol.00010.2005.

Martinez-Jimenez, Celia Pilar, Nils Eling, Hung-Chang Chen, Catalina A. Vallejos, Aleksandra A. Kolodziejczyk, Frances Connor, Lovorka Stojic, et al. 2017. “Aging Increases Cell-to-Cell Transcriptional Variability upon Immune Stimulation.” *Science* 355 (6332): 1433–36. https://doi.org/10.1126/science.aah4115.

Pisco, Angela. 2020. “Tabula Muris Senis Data Objects.” https://doi.org/10.6084/m9.figshare.12654728.v1.

Somel, Mehmet, Song Guo, Ning Fu, Zheng Yan, Hai Yang Hu, Ying Xu, Yuan Yuan, et al. 2010. “MicroRNA, mRNA, and Protein Expression Link Development and Aging in Human and Macaque Brain.” *Genome Research* 20 (9): 1207–18. https://doi.org/10.1101/gr.106849.110.

Somel, Mehmet, Philipp Khaitovich, Sabine Bahn, Svante Pääbo, and Michael Lachmann. 2006. “Gene Expression Becomes Heterogeneous with Age.” *Current Biology: CB* 16 (10): R359–60. https://doi.org/10.1016/j.cub.2006.04.024.

Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William

M. Mauck 3rd, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. 2019. “Comprehensive Integration of Single-Cell Data.” *Cell* 177 (7): 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

Tabula Muris Consortium. 2020. “A Single-Cell Transcriptomic Atlas Characterizes Ageing Tissues in the Mouse.” *Nature* 583 (7817): 590–95. https://doi.org/10.1038/s41586-020-2496-1.

Tacutu, Robi, Daniel Thornton, Emily Johnson, Arie Budovsky, Diogo Barardo, Thomas Craig, Eugene Diana, et al. 2018. “Human Ageing Genomic Resources: New and Updated Databases.” *Nucleic Acids Research* 46 (D1): D1083–90. https://doi.org/10.1093/nar/gkx1042.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie. 2001. “Estimating the Number of Clusters in a Data Set via the Gap Statistic.” *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 63 (2): 411–23. https://doi.org/10.1111/1467-9868.00293.

Trapnell, Cole, Brian A. Williams, Geo Pertea, Ali Mortazavi, Gordon Kwan, Marijke J. van Baren, Steven L. Salzberg, Barbara J. Wold, and Lior Pachter. 2010. “Transcript Assembly and Quantification by RNA-Seq Reveals Unannotated Transcripts and Isoform Switching during Cell Differentiation.” *Nature Biotechnology* 28 (5): 511–15. https://doi.org/10.1038/nbt.1621.

Turan, Zeliha Gözde, Poorya Parvizi, Handan Melike Dönertaş, Jenny Tung, Philipp Khaitovich, and Mehmet Somel. 2019. “Molecular Footprint of Medawar’s Mutation Accumulation Process in Mammalian Aging.” *Aging Cell* 348 (May): e12965. https://doi.org/10.1111/acel.12965.

Yang, Jae-Hyun, Patrick T. Griffin, Daniel L. Vera, John K. Apostolides, Motoshi Hayano, Margarita V. Meer, Elias L. Salfati, et al. 2019. “Erosion of the Epigenetic Landscape and Loss of Cellular Identity as a Cause of Aging in Mammals.” *Cold Spring Harbor Laboratory*. https://doi.org/10.1101/808642.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. “clusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters.” *Omics: A Journal of Integrative Biology* 16 (5): 284–87. https://doi.org/10.1089/omi.2011.0118.

Zahn, Jacob M., Suresh Poosala, Art B. Owen, Donald K. Ingram, Ana Lustig, Arnell Carter, Ashani T. Weeraratna, et al. 2007. “AGEMAP: A Gene Expression Database for Aging in Mice.” *PLoS Genetics* 3 (11): e201. https://doi.org/10.1371/journal.pgen.0030201.