The Road Less Traveled: Uncovering the Convergence Toward Specific Pleiotropic Phenotypes in Aging

Thomas Stoeger[1]

Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA
Correspondence: thomas.stoeger@northwestern.edu

Abstract

Aging is a complex process influenced by a wide range of environmental and molecular factors. Despite this complexity, individuals tend to age in highly similar ways, leading to the question of what drives this convergence. Recent research, including my own discoveries, suggests that the length of transcript molecules plays a crucial role in age-dependent changes to the transcriptome. Drawing inspiration from the road trip analogy of cellular transcription, I propose that a non-linear scaling law drives convergence towards specific pleiotropic phenotypes in biological aging. This scaling law is based on the notion that molecular changes observed during aging may reflect unspecific damage to cellular physiology. By validating this hypothesis, I can improve our understanding of biological aging and identify new candidate compounds for anti-aging interventions, as well as re-identify one known intervention. This work has actionable implications for improving human health and extending lifespans.

Background

As humans age, they exhibit a range of molecular, cellular, physiological, behavioral, and reproductive changes that are thought to result from damage to various components of cells, including telomeres, the genome, the epigenome, the proteome, and organelles[1]. While many theories have been proposed to explain the aging process[2], my hypothesis builds upon these frameworks to offer a new perspective that can explain a long-standing puzzle in our understanding of aging: Despite the many environmental and molecular factors that influence biological aging, individuals tend to experience similar age-related changes[3-6].

At the heart of my hypothesis is the transcriptome[7] - the complete set of transcript molecules produced by an organism's genes. Recent research, including my own, has revealed that as an organism ages, its transcriptome changes in a manner that is informed by the lengths of transcript molecules. This relationship between transcript length and aging has been first observed by the lab of Jan Hoeijmakers[8], and since then been described for *C. elegans*[9], fruit flies[10], killifish[11], mice[8,9,11-13], rats[11], and humans[8,11,14]. In most cases, the correlation between transcript length and age-dependent change is negative, although some tissues and cell types exhibit a positive correlation[11]. The strength of these correlations can be influenced by various factors, such as UV radiation exposure[9] or smoking[13] or caloric restriction[8].

It remains unclear whether these length correlations are caused by multiple factors or a single factor, such as DNA break-induced stalling of RNA polymerase[9], the enzyme responsible for

transcribing DNA into RNA. Additional molecular causes of length-correlated changes to the transcriptome have been identified in research on neuropathies, muscular dystrophies, and the heat shock response, including insufficient telescripting[15], loss of a splicing factor[16,17], loss of proteins involved in resolving DNA supercoiling[18] and loss of a transcriptional repressor that binds methylated DNA[19,20].

Recent research by Magdalena Maslon and colleagues provides the first in-vivo evidence of the physiological consequences of reduced transcriptional elongation rates in mice[21]. They engineered a point mutation in RNA polymerase, which resulted in altered splicing patterns, a decrease of long transcripts, a loss of neuronal stem cell renewal, and embryonic lethality. These findings suggest that the length-dependent changes in the transcriptome observed during aging may have profound consequences for cellular function and overall health.

Changes in the transcriptome during aging may affect all genes, but the most significant and reliable changes are observed in genes with the shortest and longest transcripts[8,11,13]. As I will demonstrate later in this manuscript, changes observed for the shortest transcripts may be an indirect consequence of changes in long transcripts. In humans, genes with the shortest transcripts are enriched for genes known to extend lifespan in model organisms, while genes with the longest transcripts enrich for genes involved in lifespan extension[11]. Additionally, the genes with the shortest transcripts are responsible for maintaining telomeres, chromatin, proteome formation, and mitochondrial function[11], all of which are associated with aging[1,22]. These findings suggest that changes in the transcriptome may play a critical role in aging and that targeting these changes may have therapeutic potential in age-related diseases.

Transcriptional Road Trip: Explaining Length-Dependent Changes in Aging

I propose that the road trip analogy of transcription can help explain why distinct environmental and molecular factors influence biological aging, and why different individuals experience aging in highly similar manners. In analogy to a road trip, RNA polymerase moves along chromatin to transcribe the information contained in the DNA sequence of a gene towards gene-specific transcript molecules (Figure 1A). However, the ability of RNA polymerase to complete its trip, and hence the production of a nascent transcript molecule, could be hindered by many factors, such as impaired function of RNA polymerase itself[21], DNA breaks[8,23], polymerase stalling[9], epigenetic consequences of methylated DNA[20], and DNA supercoiling[18,24].

Furthermore, RNA polymerase speed increases in aging[25,26], and cryptic polyadenylation sites no longer being masked by telescripting due to increased RNA polymerase speeds[15] could be viewed as corresponding to a car no longer being able to follow permitted speed limits and hence having an increased probability of being policed out of traffic (Figure 1A). The absence of any incidence that prevents the completion of a road trip, scales exponentially with the travelled distance, and a constant increase in the failure rate hence disproportionately affects long trips (Figure 1B,C).

The road trip analogy illustrates that increased inability to complete the trip may stem from various factors affecting the car's components or their interplay. Advancing age of the car would disproportionately affect long trips, irrespective of the specific cause for why the trip cannot be completed (Figure 1B,C). The probability of any component of the car to fail could depend on inherent material properties, past usage, accrued damage, maintenance, and extrinsic factors such as weather conditions or road quality.

Non-literal renditions of the road trip analogy are also thinkable as the number of introns is loosely correlated with transcript length[11,14], and intron retention occurs in aging[27,28] and can trigger the destruction of transcript molecules[29]. Longer transcripts may also provide more opportunities for RNA editing, and polymorphisms in RNA editing enzymes are associated with exceptional human longevity[30].

Minute Changes in Processivity Match Age-Dependent Changes in Transcriptome

In order to investigate whether the road trip analogy can explain transcriptomic changes observed in aging, I developed a simplified[31,32] biophysical model that approximates the scaling behavior of transcript abundance changes according to transcript length (Supplementary Figure 1). This model assumes that RNA polymerase has a certain probability of successfully transcribing each DNA nucleotide of a gene into an RNA nucleotide of its transcript. To calibrate the model, I first used a net rate of distance-dependent failure of transcript completion determined from artificial reporter constructs of different lengths[33]. However, this calibration led to the incorrect inference that most human genes are unable to form transcript molecules (Supplementary Figure 2). Instead, I plotted transcript lengths against previously observed copy numbers of transcript molecules[34] and inferred an empirical failure rate that matched the decrease in transcript copy numbers (see Methods, Supplementary Figure 3).

Next, I tested the plausibility of the road trip analogy by modulating the failure rate for completing a given nucleotide. Increasing the failure rate by 20% (which corresponds to an absolute drop in the success rate per nucleotide from 0.99999247 to 0.999990964) resulted in a 40-50% decrease in the formation of the longest transcripts (Figure 2A-C), matching or exceeding the magnitude of age-dependent decreases of transcripts observed empirically[8,11,35]. Therefore, I conclude that length-dependent changes of the transcriptome with aging are highly likely, unless it is assumed that aging cells maintain a nearly constant success rate to complete transcript molecules, which has been shown to be incorrect[9].

Transcriptomic measurements typically do not measure the absolute copy number of transcript molecules. Instead, they implicitly normalize the expression of one transcript to that of other transcripts[35,36]. To better reflect this approach, I incorporated the empirically observed distribution of transcript lengths (Figure 2A). By increasing the failure rate by 400%, I was able to match the magnitude of age-dependent increases of transcripts observed in previous studies[11,35] (Figure 2D). This finding suggests that short transcripts can increase with aging relative to other transcripts, but in reality, they decrease less in absolute copy numbers. These

results have important implications when interpreting reports of age-dependent increases of biomolecules.

Exploring the Link between Aging and Cellular Physiology: Many Perturbations Yield Length Correlations

In this section, I will explore the hypothesis that many distinct perturbations converge toward a length-correlated change of the transcriptome in aging. This postulate is conceptually similar to the hypothesis proposed by Vadim Gladyshev in 2013 on aging originating from imperfectness-driven non-random damage[37] and the 1962 hypothesis of V.J. Wulff and colleagues on changes in transcript formation during aging originating from the sites of RNA synthesis sustaining damage with their usage[38].

One of the main challenges in understanding the relationship between aging and damage is the lack of global molecular readouts of damage[39]. However, using the length-correlated changes of the transcriptome as a readout, I can distinguish between changes specific to the passage of time, loss of functionality of some components of the cell, or loss of functionality of any component of the cell.

Specifically, I re-examined recent data from Jonathan Weissman's lab[40], which conducted a perturb-seq study of cultured cells, where they impaired the function of human protein-coding genes by repressing their transcription through CRISPRi while monitoring transcriptomic changes. I found that the average effect of the loss of the function of a gene, and thus a cellular component, on the transcriptome is a slight, but significant, negative length correlation (Figure 3A; Spearman correlation between transcript lengths and relative changes of transcripts: -0.0359 +/- 0.0018 S.E.M., P=5.2e-79).

Furthermore, I found that perturbations of 41% of the tested human protein-coding genes lead to a negative correlation, and for 14%, it leads to a positive correlation (at P<0.01) (Figure 3B). This result aligns with the hypothesis that there are more ways for perturbations to make biological systems work less well than there are ways for perturbations to make cellular components work better[36].

I conclude that age-dependent changes of the transcriptome are not specific to biological aging per se, but correspond to generic impairments of cellular physiology (that can arise in the presence of CRISPRi). This finding provides further mechanistic insight into my hypothesis and confirms for the transcriptome the idea that "biological aging" refers to the consequences of "damage"[41].

However, the impairment of any component of the cell will not yield equally sized changes to the transcriptome. Perturbations of genes that positively regulate transcriptional elongation lead to the most negative length correlations (Supplementary Figure 4A), and the global mapping of gene-specific loss-of-function effects reported by the study of Jonathan Weissman and colleagues can be shown to reflect upon the inflicted length correlations in the

transcriptome (Spearman correlation of -0.41 between UMAP1 in Figure 2 of their study[40] and length correlations; Supplementary Figure 4B). In an accompanying preprint[42], I describe how I identified the specific regulator of transcriptional elongation whose impairment I believe to usually trigger the age-dependent length correlations observed in C57BL/6J mice kept under specific-pathogen-free conditions, and potentially in humans as well.

Additionally, these findings do not imply that damage per se is sufficient to inform on length dependent changes of the transcriptome. The yeast *S. cerevisiae*, for example, exhibits transcript formation that increases in aging[43] (due to a loss of nucleosome occupancy) whereas transcript formation primarily decreases in animal aging[38,44-49]. As expected by the same non-linear scaling behavior that leads to a disproportionate decrease of long transcripts when transcript completion is decreased (Supplementary Figure 1), long transcripts show a disproportionate increase with aging in *S. cerevisiae* (Supplementary Figure 5).

The Length Correlation and its Potential Role in Integrating Inputs and Outputs during Biological Aging

While the road trip analogy provides valuable insights into the transcriptomic changes that occur with age, the true value of my hypothesis lies in its potential to explain why hundreds of distinct perturbations can influence biological aging, and why aging affects individuals in similar ways despite various manifestations of advanced age[3-6].

The integration of inputs towards outputs during biological aging has been previously hypothesized to stem from various factors such as telomeres[50], the epigenome[51], signaling pathways[6,52], and network prinicples[53]. While these hypotheses are not mutually exclusive, no specific integratory mechanism has been revealed, except for the shortening of telomeres[50].

To determine whether the road trip analogy can fit into existing conceptual insights on the integration of distinct cues during aging, I turn to a recent review by Alan Cohen and colleagues[6]. They identified ten input cues and eleven output consequences of aging, which they postulated could be connected through a limited set of cellular signaling pathways. Interestingly, four of their ten input cues and nine of their eleven outputs have already been connected to the length correlation of the transcriptome[8,9,11,13,21] (Figure 4). The number of connected inputs increases to at least six when considering a meta-analysis of 4,796 studies of the EBI Gene Expression Atlas[54], which I included in the preprint version of a recent manuscript[55] and predates the review by Alan Cohen and colleagues (see Methods). These findings suggest that the length correlation may play a critical role in integrating many aspects associated with aging.

Moreover, length-correlated changes in the transcriptome have been observed in Alzheimer's disease[56,57], and the potential duality with length-correlated changes in aging was first proposed by Sourena Soheili-Nezhad in 2018[58]. The length correlation may therefore represent a promising avenue for understanding the age-dependency of age-related diseases.

Identifying Novel Anti-Aging Interventions

Antagonizing the length correlations observed during biological aging may pave the way for the development of a novel class of anti-aging interventions. This potential has been noted in a recent review[59], which appeared during the final stages of writing this manuscript.

It is also possible that some well-established anti-aging interventions have already been countering aging by counteracting age-dependent length correlations of the transcriptome. For example, caloric restriction predominantly affects genes encoding proteins involved in RNA splicing and subcellular transport of transcripts[60].

To identify novel candidate anti-aging interventions, one paradigm is to understand the primary causes of length correlations encountered during human aging and then design interventions based on these insights. Established behavioral interventions such as avoiding smoking and sources of excess DNA damage (such as UV irradiation) counter known causes of accelerated aging and length-correlated changes of the transcriptome[13].

Another paradigm is to act on the hypothesis that the net success rate of transcript formation defines the length correlation (Supplementary Figure 1). Potential anti-aging interventions do not necessarily need to counter rate-limiting or causal processes acting during the passage of time. Instead, they may modulate other aspects of cellular physiology (Figure 3).

To evaluate the second paradigm, I reanalyzed data from LINCS1000, which paired compounds with measurements of the transcripts of nearly 1,000 protein-coding genes in exposed cultured cells[61] (Figure 5A). Among 3,935 compounds, 111 always yielded a more positive length correlation than controls obtained by shuffling transcript lengths between different genes (10,000 of 10,000 randomizations). Many of these compounds (Figure 5B) appear bio-plausible as impacting negative regulators of chromatin accessibility. Notably, one well-documented anti-aging intervention, rapamycin[62], is among the 111 compounds, although it yields the weakest positive correction among all 111 compounds (Spearman correlation: +0.03). I will not disclose the full list of compounds for the sake of public health, as their efficacy and safety will depend on further considerations beyond LINCS1000 and their ability to antagonize the primary direction of the length correlation observed in human aging. More sophisticated analyses are thinkable and encouraged.

To develop viable anti-aging interventions, it is important to recall that decreased[21] and increased[15] speeds of RNA polymerase can both lead to a loss of processivity of RNA polymerase and thereby a decrease of long transcripts[15,21]. Therefore, anti-aging interventions that target the length correlation may require continuous monitoring and adjustments of the dosage regime. This emphasizes the need for a nuanced approach to developing and monitoring anti-aging interventions, to avoid unintended consequences and ensure optimal outcomes.

Discussion

Theories that explain biological aging often suggest that damage accumulation is a key factor[1,2]. There are many reviews on organismal[63,64] and molecular damage[39,65-69], but it is challenging to find a well-cited review on the various types of transcript damage that occur during biological aging. These damages extend beyond premature termination[9] and include intron retention[27,28], oxidative damage[70], and lost synchronization toward the circadian rhythm[71]. It is surprising how few reviews discuss transcript damages[22,39,72,73], given that research by Lorena Harries and colleagues[74] found that out of 1,065 considered biological and metabolic pathways, six of the seven pathways significantly associated with age in human peripheral blood leukocytes are related to RNA metabolism. The seventh pathway relates to chromatin assembly and disassembly.

In Box 1, I list the main aspects of my hypothesis, which is the first tangible hypothesis to explain why hundreds of distinct environmental and molecular factors influence biological aging, and why biological aging still affects different individuals in highly similar manners. Specifically, I offer a systems-level scaling principle that explains how the copy number of transcripts of thousands of genes changes during the passage of time.

During the preparation of this manuscript, Olga Ibáñez Solé and Ander Izeta independently arrived at the same starting hypothesis of the length-dependent changes of transcripts during aging relating to a generic loss of processivity of RNA polymerase and the non-linear scaling with transcript length[13].

I have challenged our hypothesis by using it to demonstrate that small increases in the failure rate of RNA polymerase suffice to match the changes observed in aging, and that transcriptomic changes observed during aging mirror the impact of unspecific loss of function of cellular components. Moreover, I demonstrated that the length correlation in the transcriptome connects many known input cues and output consequences of aging, and that our hypothesis adds sufficient explanatory value to identify in existing data one proven and many potential anti-aging interventions. Our hypothesis and the road trip analogy might even extend beyond transcripts toward polypeptides as ribosome pausing increases during *S. cerevisiae* and *C. elegans* aging and leads to a subsequent aggregation of nascent polypeptides[75].

However, the downstream consequences of the length correlation cannot be investigated entirely through experiments for the foreseeable future without altering the length correlation itself. This is because the length correlation is a genome-wide phenomenon, while most experimental approaches to modify gene transcription are focused on individual genes. Additionally, for most genes, the magnitude by which their transcripts change during aging is small compared to current experimental approaches[35], and recreating these comparably small magnitudes would require finely tuned experimental modulations of numerous individual genes that are not feasible at scale. As a result, it remains unclear if observed downstream effects

toward the transcript abundance of specific subsets of genes contribute to age-associated phenotypes by increasing or lessening them.

Given these limitations, I referred to my insights as a "hypothesis" and hope that its merits will be evaluated positively through its ability to spur the discovery and development of successful interventions to improve human health and extend lifespans.

Box 1: Main aspects

1. Passage of Time: Living organisms experience continuous perturbations as time passes.
2. Transcript Completion: The scaling behavior of transcript formation with transcript length informs on how transcripts change over time.
3. Damage: Impeded functionality of cellular components primarily leads to a negative correlation between transcript abundance and transcript length.
4. Convergence: Transcriptomes converge over time due to length-dependent non-linear scaling.
5. Pleiotropic Phenotypes: The molecular components and processes encoded by short and long transcripts are distinctively affected by the passage of time, leading to diverse phenotypes.

Methods

Artificial Intelligence-assisted Manuscript Preparation

The illustration of the car and its road trip was generated using DALL-E2 (https://openai.com/product/dall-e-2). After manuscript writing, the text was edited using ChatGPT (http://chat.openai.com) to enhance clarity and appeal for scientific journals.

Length Correlations

The primary transcript lengths for humans, mice, and *S. cerevisiae* were obtained from BioMart (http://useast.ensembl.org/biomart/martview), specifically from GRCh38.p13, GRCm39, and R64-1-1, respectively. For each gene, the median length across all primary transcripts was calculated, and genes marked as protein-coding by NCBI's reference file (https://ftp.ncbi.nlm.nih.gov/gene/DATA/GENE_INFO/All_Data.gene_info.gz) were selected. For mice and humans, only genes with an unambiguous 1:1 mapping between Entrez and Ensembl gene identifiers were retained. Length correlations were defined as the Spearman correlation between the lengths of primary transcripts and fold-changes of transcripts and computed using SciPy[76] 1.2.1.

Box and Regression Plots

Boxplots and regression plots, including bootstrapped estimates of the median and regression trends, were generated using seaborn[77] 0.11.2.

Estimation of Empirical Failure Rate

For the initial estimate, two measurements provided by Lee and Greenleaf[33] were used: an overall success rate of 0.65 for the first 867 nucleotides of a reporter construct and a success

rate of 0.35 for 1764 nucleotides between the 124th and 1888th nucleotide position of a reporter. The success rate per nucleotide was calculated by taking the 867th and 1764th root of these success rates, respectively. The per-nucleotide estimates were averaged and subtracted from 1 to obtain the per-nucleotide failure rate.

For the second and final estimate, transcript copy numbers were obtained from Schwanhäusser et al.[34] and filtered and aggregated for protein-coding genes as described in the length correlations section. The default empirical failure rate within the biophysical model was set such that for mice, the median transcript copy number of the 10% longest transcripts is only 45% of the median transcript copy number of the 10% shortest transcripts.

Reanalysis of Perturb-seq

The fold-changes and coordinates of Figure 2 of Replogle and Saunders et al.[40] were obtained from the author-provided files at https://doi.org/10.25452/figshare.plus.21632564.v1. For the analysis of Gene Ontology annotations[78], process annotation groups with at least 20 CRISPRi targeted genes in Replogle and Saunders et al.[40] were selected from https://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2go.gz. Targeted genes were excluded from the computation of length correlations.

Mapping of inputs and outputs

To identify the inputs connected to the length correlation, I first reviewed published studies and then looked at the top 50 length correlations, both positive and negative, reported in my preprint[55]. I conducted this analysis prior to the review by Cohen et al.[6] to avoid any retrospective fitting of my analysis. I had investigated 4,796 studies in the EBI Gene Expression Atlas[54] to identify length correlations. To improve the overall clarity of the resulting manuscript[11], I removed the corresponding analysis as suggested by a reviewer.

In support for a connection of the length correlation to "Nutrients", I found that 2-fucosyllactose supplementation resulted in the 5th most negative length correlation in mice, while a fast-food diet resulted in the 33rd most positive correlation. Similarly, in support of the connection to "Pollution", exposure to polluted groundwater resulted in the 9th most negative length correlation in mice.

Although the support for a connection to "Microbiota" is less direct, several microbes, such as Fusobacterium nucleatum, resulted in a negative length correlation, with Fusobacterium nucleatum resulting in the 10th most negative correlation in humans. Similarly, while the support for a connection to "Psychological stress" is less direct as this category is not directly contained in the EBI Gene Expression Atlas, sleep deprivation resulted in the 13th most negative length correlation in mice.

For outputs connected to the length correlation, I reviewed findings described in the main text and supplemental tables of prior publications on the length correlation of the transcriptome.

Reanalysis of LINCS1000

To conduct the reanalysis of LINCS1000, I used level 5 (fully normalized) data from the Gene Expression Omnibus record GSE92742, as provided by the authors. I only considered experimentally measured genes, and further I excluded from the analysis compounds of the Broad Institute, whose name would start with the prefix BRD-; As the readout of the length correlation, I considered for each compound the median length correlation across all conditions (e.g.: dosage regimes, cell types) in which a compound had been tested.

Acknowledgements

I am deeply grateful to my mentors, Elizabeth McNally, Luis Amaral, Richard Morimoto, and Scott Budinger, for their unwavering support and guidance throughout my journey as an independent scholar. I am equally grateful to my informal mentor Jacob Sznajder, who first proposed that I should follow our earlier study on length-correlated changes in the transcriptome with an article that embedded our findings in the broader context of aging research. I would also like to acknowledge Alexander Misharin, Douglas Vaughan and the research groups of Luis Amaral, Richard Morimoto, and Scott Budinger for their valuable contributions to my understanding of biological aging through thoughtful discussions. I am also thankful to the generous donors of Northwestern University, particularly Kimberly Querrey and Louis Simpson, and the Potocsnak family, for creating an environment that fosters critical thinking on health and aging. Additionally, this manuscript would not have been possible without the support from the NIH-National Institute on Aging. Their financial contribution (K99AG06854) and non-financial support were instrumental in encouraging the author's personal development and line of inquiry that ultimately led to the creation of this manuscript.

Competing Interests

I declare the absence of competing interests.

Materials & Correspondence

Should be addressed to Thomas Stoeger (thomas.stoeger@northwestern.edu).

Figures

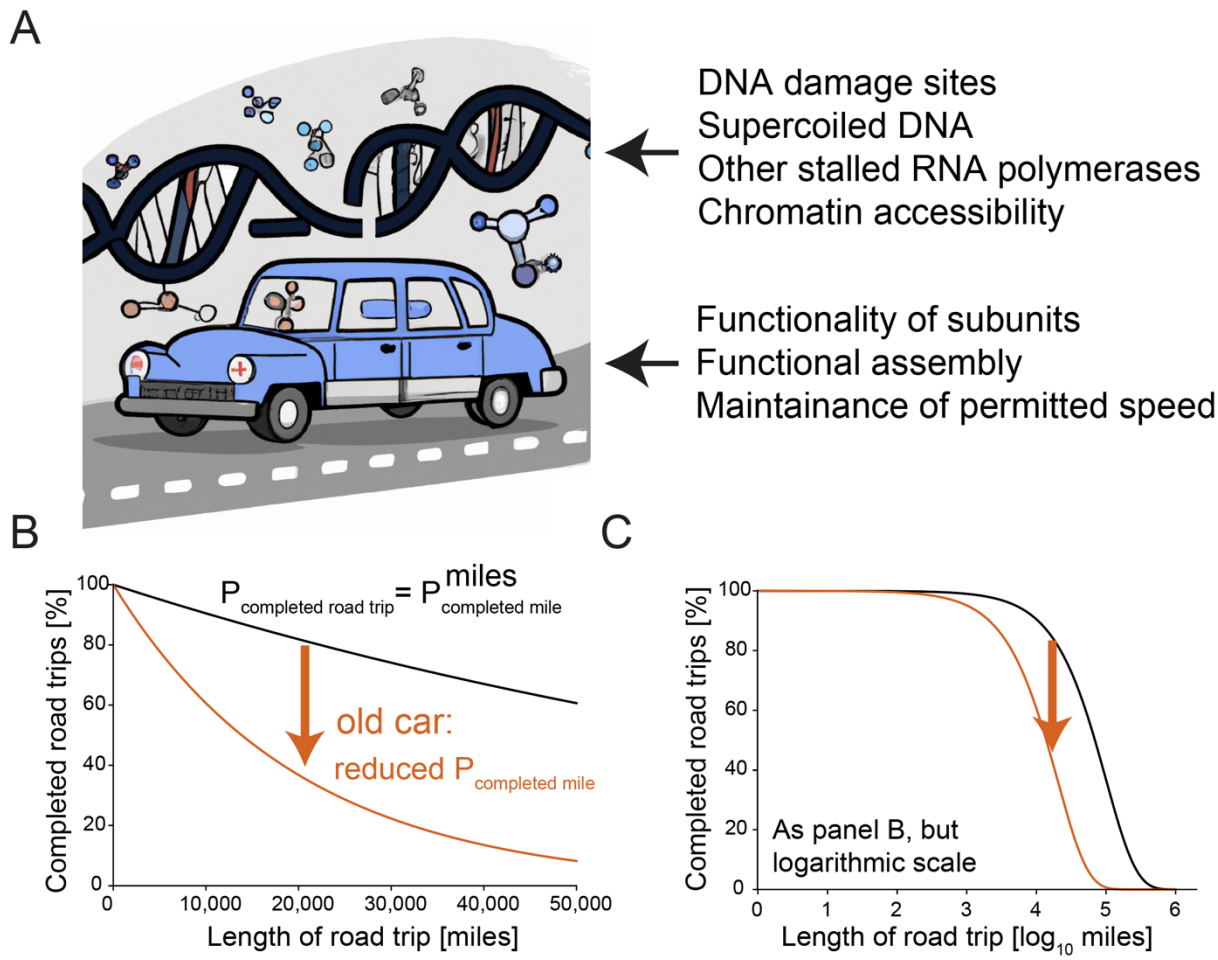

Figure 1: The road trip analogy of transcription proposes that an age-related decline in functionality affects long road trips more severely. A) The analogy between transcription and a road trip highlights the similarities in the challenges faced in both processes. B) The probability of completing a road trip ($P_{completed\ road\ trip}$) exponentially decreases with distance in miles. The orange line illustrates the probability of completing a trip when the probability of completing a mile ($P_{completed\ mile}$) is only one fifth. My rough estimate for $P_{completed\ road}$ trip is based on browsing internet forums, as actual failure rates of cars do not seem publicly reported, but the overall scaling behavior is independent of the specific numeric values. C) Logarithmic scale representation of the data shown in panel B enhances the visualization of the exponential decrease in $P_{completed\ road\ trip}$ with distance.

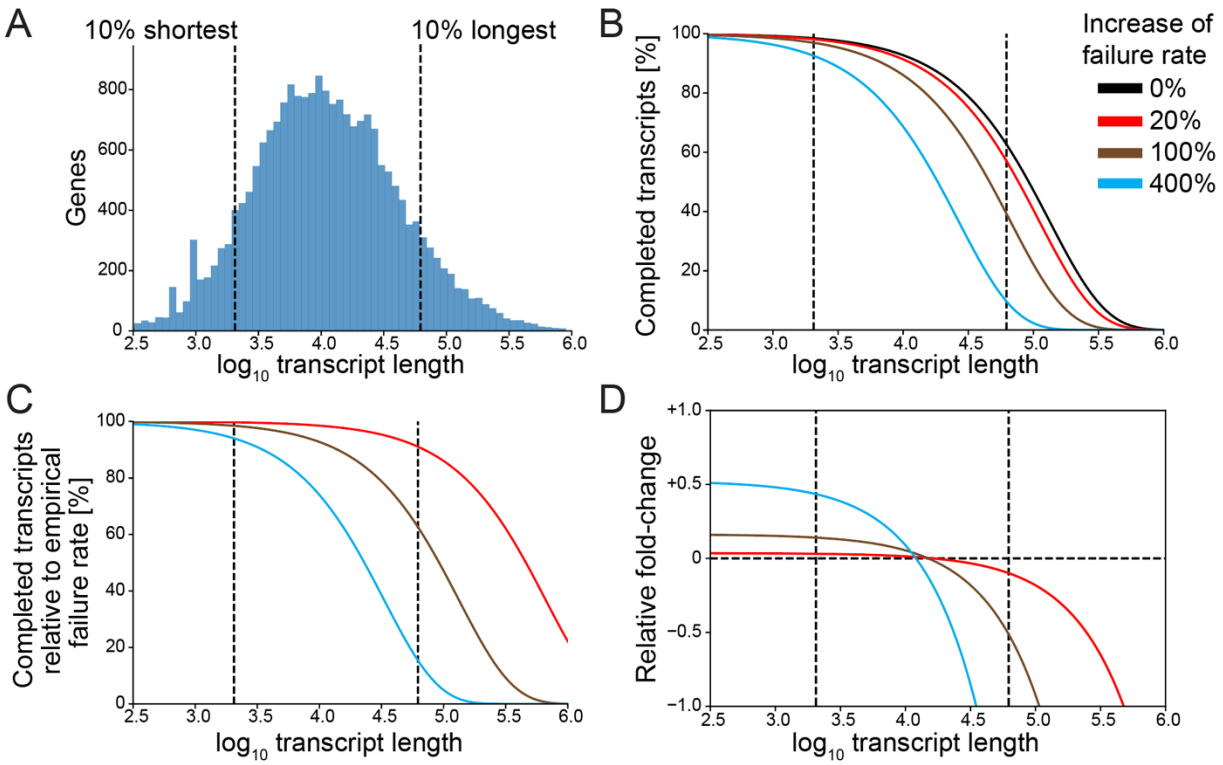

Figure 2: Small increases in the failure rate are sufficient to match changes observed in biological aging. A) Histogram of median transcript lengths of protein-coding genes in mice (for human, see Supplementary Figure 2A). Dashed lines indicate the 10% shortest and 10% longest transcript lengths. B) Increasing the empirically inferred failure rate (Supplementary Figure 3) by variable amounts (*c(a)* in Supplementary Figure 1) results in a reduction of long transcripts. C) Normalization of the data in panel B to the empirically inferred failure rate. Colors as in panel B. D) The relative fold-change is presented as a log2-ratio of simulated transcriptomic measurements in the presence of increased failure rate compared to the empirically inferred failure rate. Unlike panel C simulated transcriptomic measurements were normalized to total RNA abundance, assuming equal initiation frequency for each transcript in panel A. The horizontal dashed line indicates absence of a relative fold-change. Colors as in panel B.

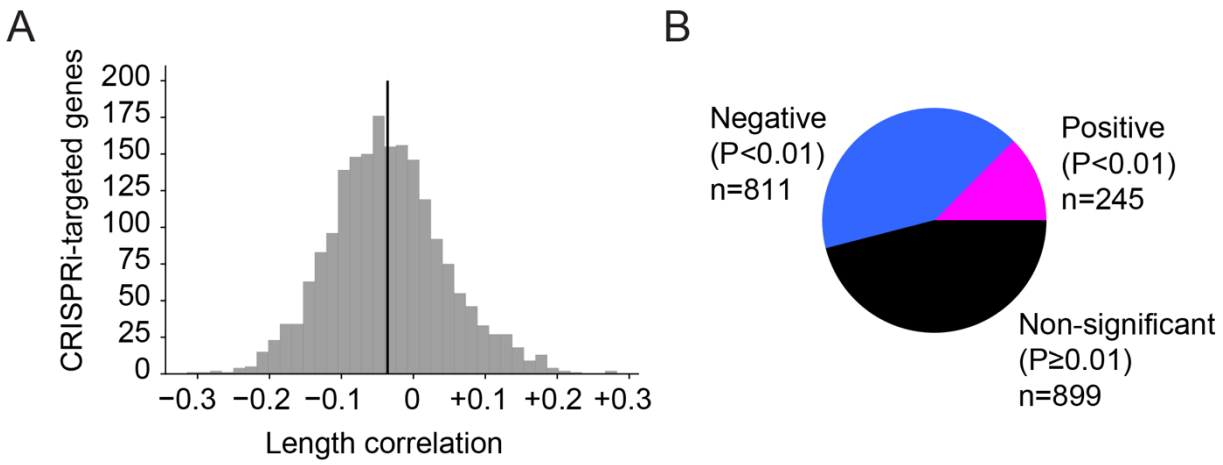

Figure 3: Loss of cellular components results in negative correlation between transcript length and relative fold-changes of transcripts. A) Histogram showing the length correlations observed in data by Replogle and Saunders et al. for 1,955 protein-coding genes targeted by CRISPRi[40]. B) Fraction of CRISPRi-targeted genes from panel A showing significant negative or positive correlation using Kendall and Stuart's method for P values of Spearman correlations[79].

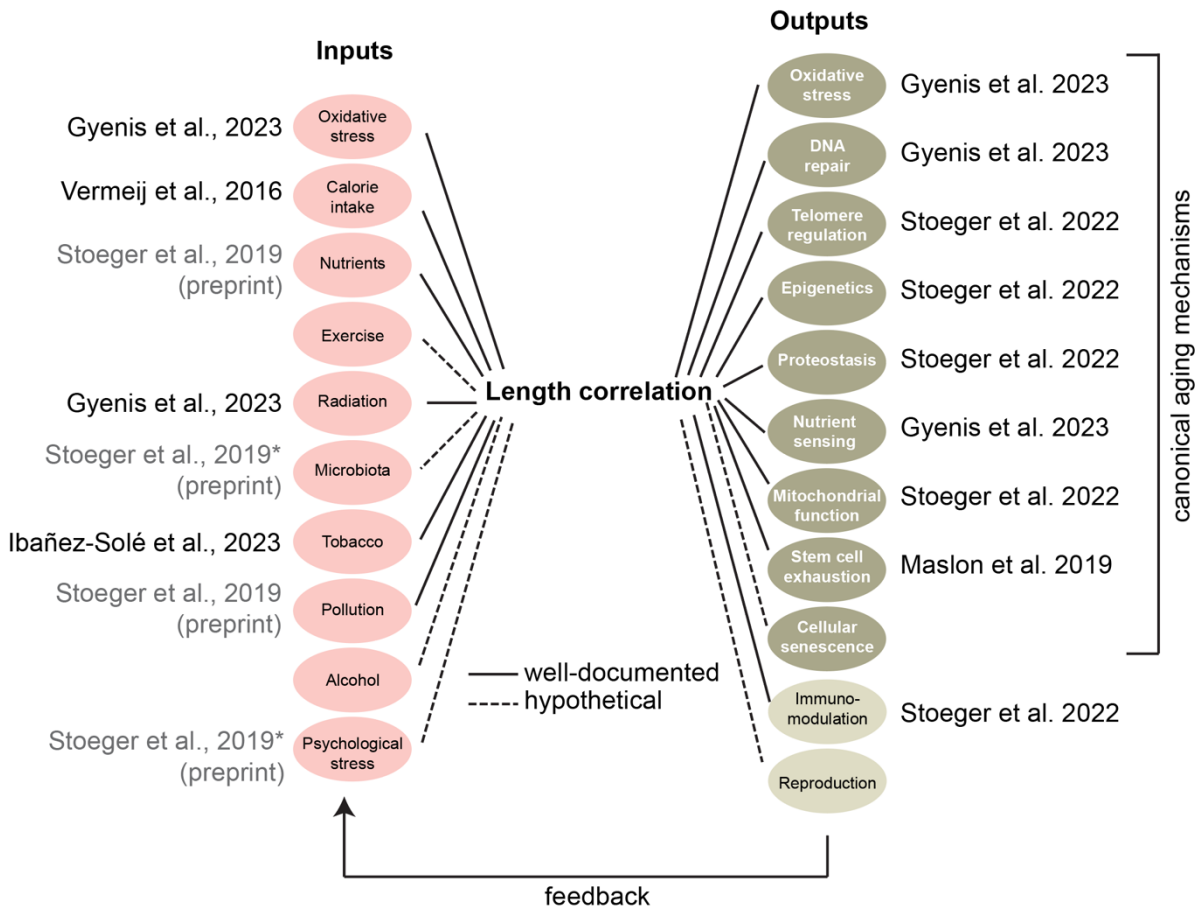

Figure 4: The length correlation connects inputs and outputs in aging processes. This figure, adapted from Cohen et al.[6] (licensed under CC BY 4.0), illustrates the inputs and outputs connected to length-correlated changes of the transcriptome. References for the specific inputs and outputs are provided[8,9,11,13,21,55]. Note that this figure includes references to the preprint version[55] of a recent manuscript of mine, as I had removed some material following reviewer feedback to enhance focus. Furthermore, the observed length correlation after sleep deprivation may be linked to psychological stress, a hypothetical input, while the correlation observed following exposure to microbes may be linked to microbiota, another hypothetical input.

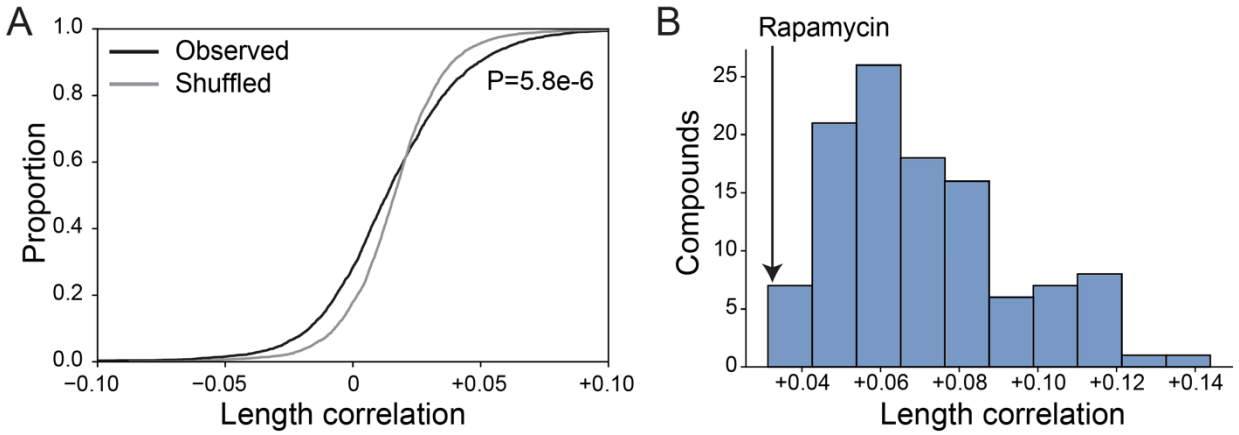

Figure 5: Length correlations in LINCS1000 Identify candidate anti-aging Interventions. A) Cumulative empirical density function for length correlations in LINCS1000[61] as observed with real transcript lengths (black) or when transcript lengths were shuffled among genes (grey). The P value is from the Kolmogorov-Smirnov test. It should be noted that both the observed and shuffled conditions yield a positive correlation, which I suspect to arise from the length correlation not being explicitly accounted for in the normalization scheme used and provided by LINCS1000. B) Histogram showing 111 compounds that consistently produce a more positive length correlation than 10,000 randomizations in which transcript lengths were shuffled. Among these compounds, rapamycin has the weakest positive length correlation.

References

1 Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186**, 243-278 (2023). https://doi.org:10.1016/j.cell.2022.11.001

2 Medvedev, Z. A. An attempt at a rational classification of theories of ageing. *Biol Rev Camb Philos Soc* **65**, 375-398 (1990). https://doi.org:10.1111/j.1469-185x.1990.tb01428.x

3 Spencer, R. P. & Clive, J. M. Evidence for senescent convergence during human aging. *Mech Ageing Dev* **91**, 185-194 (1996). https://doi.org:10.1016/s0047-6374(96)01786-1

4 Cohen, A. A. Complex systems dynamics in aging: new evidence, continuing questions. *Biogerontology* **17**, 205-220 (2016). https://doi.org:10.1007/s10522-015-9584-x

5 Freund, A. Untangling Aging Using Dynamic, Organism-Level Phenotypic Networks. *Cell Syst* **8**, 172-181 (2019). https://doi.org:10.1016/j.cels.2019.02.005

6 Cohen, A. A. *et al.* A complex systems approach to aging biology. *Nature Aging* **2**, 580-591 (2022). https://doi.org:10.1038/s43587-022-00252-6

7 Izgi, H. *et al.* Inter-tissue convergence of gene expression during ageing suggests age-related loss of tissue and cellular identity. *Elife* **11** (2022). https://doi.org:10.7554/eLife.68048

8 Vermeij, W. P. *et al.* Restricted diet delays accelerated ageing and genomic stress in DNA-repair-deficient mice. *Nature* **537**, 427-431 (2016). https://doi.org:10.1038/nature19329

9 Gyenis, A. *et al.* Genome-wide RNA polymerase stalling shapes the transcriptome during aging. *Nature Genetics* (2023). https://doi.org:10.1038/s41588-022-01279-6

10 Hall, H. *et al.* Transcriptome profiling of aging Drosophila photoreceptors reveals gene expression trends that correlate with visual senescence. *BMC Genomics* **18**, 894 (2017). https://doi.org:10.1186/s12864-017-4304-3

11 Stoeger, T. *et al.* Aging is associated with a systemic length-associated transcriptome imbalance. *Nature Aging* **2**, 1191-1206 (2022). https://doi.org:10.1038/s43587-022-00317-6

12 Vougioukalaki, M. *et al.* Different responses to DNA damage determine ageing differences between organs. *Aging Cell* **e13562** (2022). https://doi.org:https://doi.org/10.1111/acel.13562

13 Ibañez-Solé, O., Barrio, I. & Izeta, A. Age or lifestyle-induced accumulation of genotoxicity is associated with a length-dependent decrease in gene expression. *iScience*, 106368 (2023). https://doi.org:https://doi.org/10.1016/j.isci.2023.106368

14 Lopes, I., Altab, G., Raina, P. & de Magalhães, J. P. Gene Size Matters: An Analysis of Gene Length in the Human Genome. *Front Genet* **12** (2021).

15 Cugusi, S. *et al.* Heat shock induces premature transcript termination and reconfigures the human transcriptome. *Mol Cell* **82**, 1573-1588 e1510 (2022). https://doi.org:10.1016/j.molcel.2022.01.007

16 Hosokawa, M. *et al.* Loss of RNA-Binding Protein Sfpq Causes Long-Gene Transcriptopathy in Skeletal Muscle and Severe Muscle Mass Reduction with Metabolic Myopathy. *iScience* **13**, 229-242 (2019). https://doi.org:10.1016/j.isci.2019.02.023

17 Takeuchi, A. *et al.* Loss of Sfpq Causes Long-Gene Transcriptopathy in the Brain. *Cell Rep* **23**, 1326-1341 (2018). https://doi.org:10.1016/j.celrep.2018.03.141
18 King, I. F. *et al.* Topoisomerases facilitate transcription of long genes linked to autism. *Nature* **501**, 58-62 (2013). https://doi.org:10.1038/nature12504
19 Gabel, H. W. *et al.* Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89-93 (2015). https://doi.org:10.1038/nature14319
20 Kinde, B., Wu, D. Y., Greenberg, M. E. & Gabel, H. W. DNA methylation in the gene body influences MeCP2-mediated gene repression. *Proc Natl Acad Sci U S A* **113**, 15114-15119 (2016). https://doi.org:10.1073/pnas.1618737114
21 Maslon, M. M. *et al.* A slow transcription rate causes embryonic lethality and perturbs kinetic coupling of neuronal genes. *EMBO J* **38** (2019). https://doi.org:10.15252/embj.2018101244
22 Frenk, S. & Houseley, J. Gene expression hallmarks of cellular ageing. *Biogerontology* **19**, 547-566 (2018). https://doi.org:10.1007/s10522-018-9750-z
23 Yang, J. H. *et al.* Loss of epigenetic information as a cause of mammalian aging. *Cell* **186**, 305-326 e327 (2023). https://doi.org:10.1016/j.cell.2022.12.027
24 Baranello, L. *et al.* RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* **165**, 357-371 (2016). https://doi.org:10.1016/j.cell.2016.02.036
25 Bozukova, M. *et al.* Aging is associated with increased chromatin accessibility and reduced polymerase pausing in liver. *Mol Syst Biol* **18**, e11002 (2022). https://doi.org:10.15252/msb.202211002
26 Debès, C. *et al.* Aging-associated changes in transcriptional elongation influence metazoan longevity. *bioRxiv*, 719864 (2022). https://doi.org:10.1101/719864
27 Mazin, P. *et al.* Widespread splicing changes in human brain development and aging. *Mol Syst Biol* **9**, 633 (2013). https://doi.org:10.1038/msb.2012.67
28 Mariotti, M., Kerepesi, C., Oliveros, W., Mele, M. & Gladyshev, V. N. Deterioration of the human transcriptome with age due to increasing intron retention and spurious splicing. *bioRxiv*, 2022.2003.2014.484341 (2022). https://doi.org:10.1101/2022.03.14.484341
29 Braunschweig, U. *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* **24**, 1774-1786 (2014). https://doi.org:10.1101/gr.177790.114
30 Sebastiani, P. *et al.* RNA editing genes associated with extreme old age in humans and with lifespan in C. elegans. *PLoS One* **4**, e8210 (2009). https://doi.org:10.1371/journal.pone.0008210
31 Hamelin, C., Cousineau, L., Dion, M. & Yelle, J. Increased DNA topoisomerase I activity in aging human cell chromatin. *Biosci Rep* **4**, 861-868 (1984). https://doi.org:10.1007/bf01138168
32 Tubbs, A. & Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **168**, 644-656 (2017). https://doi.org:10.1016/j.cell.2017.01.002
33 Lee, J. M. & Greenleaf, A. L. Modulation of RNA polymerase II elongation efficiency by C-terminal heptapeptide repeat domain kinase I. *J Biol Chem* **272**, 10990-10993 (1997). https://doi.org:10.1074/jbc.272.17.10990

34 Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011). https://doi.org:10.1038/nature10098
35 Cellerino, A. & Ori, A. What have we learned on aging from omics studies? *Semin Cell Dev Biol* **70**, 177-189 (2017). https://doi.org:10.1016/j.semcdb.2017.06.012
36 Kaelin, W. G., Jr. Common pitfalls in preclinical cancer target validation. *Nat Rev Cancer* **17**, 425-440 (2017). https://doi.org:10.1038/nrc.2017.32
37 Gladyshev, V. N. The origin of aging: imperfectness-driven non-random damage defines the aging process and control of lifespan. *Trends Genet* **29**, 506-512 (2013). https://doi.org:10.1016/j.tig.2013.05.004
38 Wulff, V. J., Quastler, H. & Sherman, F. G. An hypothesis concerning RNA metabolism and aging. *Proc Natl Acad Sci U S A* **48**, 1373-1375 (1962). https://doi.org:10.1073/pnas.48.8.1373
39 Gladyshev, V. N. *et al.* Molecular damage in aging. *Nature Aging* **1**, 1096-1106 (2021). https://doi.org:10.1038/s43587-021-00150-3
40 Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559-2575 e2528 (2022). https://doi.org:10.1016/j.cell.2022.05.013
41 Kirkwood, T. B. A systematic look at an old problem. *Nature* **451**, 644-647 (2008). https://doi.org:10.1038/451644a
42 Richardson, R. A., Navarro, H. T., Amaral, L. A. N. & Stoeger, T. Data-driven design of a tool to promote the investigation of understudied genes. *bioRxiv*, 2023.2002.2028.530483 (2023). https://doi.org:10.1101/2023.02.28.530483
43 Hu, Z. *et al.* Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. *Genes Dev* **28**, 396-408 (2014). https://doi.org:10.1101/gad.233221.113
44 Wulff, V. J. & Freshman, M. Age-related reduction of the RNA content of rat cardiac muscle and cerebellum. *Arch Biochem Biophys* **95**, 181-182 (1961). https://doi.org:10.1016/0003-9861(61)90125-4
45 Yannarell, A., Schumm, D. E. & Webb, T. E. Age-dependence of nuclear RNA processing. *Mech Ageing Dev* **6**, 259-264 (1977). https://doi.org:10.1016/0047-6374(77)90026-4
46 Uemura, E. Age-related changes in neuronal RNA content in rhesus monkeys (Macaca mulatta). *Brain Res Bull* **5**, 117-119 (1980). https://doi.org:10.1016/0361-9230(80)90182-3
47 Richardson, A., Birchenall-Sparks, M. C., Staecker, J. L., Hardwick, J. P. & Liu, D. S. The transcription of various types of ribonucleic acid by hepatocytes isolated from rats of various ages. *J Gerontol* **37**, 666-672 (1982). https://doi.org:10.1093/geronj/37.6.666
48 Davie, K. *et al.* A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* **174**, 982-998 e920 (2018). https://doi.org:10.1016/j.cell.2018.05.057
49 Zhang, M. J., Pisco, A. O., Darmanis, S. & Zou, J. Mouse aging cell atlas analysis reveals global and cell type-specific aging signatures. *Elife* **10** (2021). https://doi.org:10.7554/eLife.62293
50 Shammas, M. A. Telomeres, lifestyle, cancer, and aging. *Curr Opin Clin Nutr Metab Care* **14**, 28-34 (2011). https://doi.org:10.1097/MCO.0b013e32834121b1

51 Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245-254 (2003). https://doi.org:10.1038/ng1089
52 Liu, G. Y. & Sabatini, D. M. mTOR at the nexus of nutrition, growth, ageing and disease. *Nat Rev Mol Cell Biol* **21**, 183-203 (2020). https://doi.org:10.1038/s41580-019-0199-y
53 Vural, D. C., Morrison, G. & Mahadevan, L. Aging in complex interdependency networks. *Phys Rev E Stat Nonlin Soft Matter Phys* **89**, 022811 (2014). https://doi.org:10.1103/PhysRevE.89.022811
54 Papatheodorou, I. *et al.* Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res* **46**, D246-D251 (2018). https://doi.org:10.1093/nar/gkx1158
55 Stoeger, T. *et al.* Aging is associated with a systemic length-driven transcriptome imbalance. *bioRxiv* (2019). https://doi.org:https://doi.org/10.1101/691154
56 Barbash, S. & Sakmar, T. P. Length-dependent gene misexpression is associated with Alzheimer's disease progression. *Sci Rep* **7**, 190 (2017). https://doi.org:10.1038/s41598-017-00250-4
57 Soheili-Nezhad, S., van der Linden, R. J., Olde Rikkert, M., Sprooten, E. & Poelmans, G. Long genes are more frequently affected by somatic mutations and show reduced expression in Alzheimer's disease: Implications for disease etiology. *Alzheimers Dement* **17**, 489-499 (2021). https://doi.org:10.1002/alz.12211
58 Soheili-Nezhad, S. Alzheimer's disease: the large gene instability hypothesis. *bioRxiv*, 189712 (2018). https://doi.org:10.1101/189712
59 Sims, A. A. & Gurkar, A. U. DNA damage-induced stalling of transcription drives aging through gene expression imbalance. *DNA Repair* **125**, 103483 (2023). https://doi.org:https://doi.org/10.1016/j.dnarep.2023.103483
60 Rhoads, T. W. *et al.* Caloric Restriction Engages Hepatic RNA Processing Mechanisms in Rhesus Monkeys. *Cell Metab* **27**, 677-688 e675 (2018). https://doi.org:10.1016/j.cmet.2018.01.014
61 Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e1417 (2017). https://doi.org:10.1016/j.cell.2017.10.049
62 Tsang, C. K., Qi, H., Liu, L. F. & Zheng, X. F. Targeting mammalian target of rapamycin (mTOR) for health and diseases. *Drug Discov Today* **12**, 112-124 (2007). https://doi.org:10.1016/j.drudis.2006.12.008
63 Clegg, A., Young, J., Iliffe, S., Rikkert, M. O. & Rockwood, K. Frailty in elderly people. *Lancet* **381**, 752-762 (2013). https://doi.org:10.1016/S0140-6736(12)62167-9
64 Franceschi, C. & Campisi, J. Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. *J Gerontol A Biol Sci Med Sci* **69 Suppl 1**, S4-9 (2014). https://doi.org:10.1093/gerona/glu057
65 Shigenaga, M. K., Hagen, T. M. & Ames, B. N. Oxidative damage and mitochondrial decay in aging. *Proc Natl Acad Sci U S A* **91**, 10771-10778 (1994). https://doi.org:10.1073/pnas.91.23.10771
66 Hoeijmakers, J. H. DNA damage, aging, and cancer. *N Engl J Med* **361**, 1475-1485 (2009). https://doi.org:10.1056/NEJMra0804615

67 Blackburn, E. H., Epel, E. S. & Lin, J. Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science* **350**, 1193-1198 (2015). https://doi.org:10.1126/science.aab3389
68 Labbadia, J. & Morimoto, R. I. The biology of proteostasis in aging and disease. *Annu Rev Biochem* **84**, 435-464 (2015). https://doi.org:10.1146/annurev-biochem-060614-033955
69 Liguori, I. *et al.* Oxidative stress, aging, and diseases. *Clin Interv Aging* **13**, 757-772 (2018). https://doi.org:10.2147/CIA.S158513
70 Nunomura, A. *et al.* Oxidative damage to RNA in aging and neurodegenerative disorders. *Neurotox Res* **22**, 231-248 (2012). https://doi.org:10.1007/s12640-012-9331-x
71 Talamanca, L., Gobet, C. & Naef, F. Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in humans. *Science* **379**, 478-483 (2023). https://doi.org:10.1126/science.add0846
72 Muller, W. E. *et al.* Poly(A) metabolism and aging: a current view. *Arch Gerontol Geriatr* **9**, 231-250 (1989). https://doi.org:10.1016/0167-4943(89)90043-5
73 Stegeman, R. & Weake, V. M. Transcriptional Signatures of Aging. *J Mol Biol* **429**, 2427-2437 (2017). https://doi.org:10.1016/j.jmb.2017.06.019
74 Harries, L. W. *et al.* Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Aging Cell* **10**, 868-878 (2011). https://doi.org:10.1111/j.1474-9726.2011.00726.x
75 Stein, K. C., Morales-Polanco, F., van der Lienden, J., Rainbolt, T. K. & Frydman, J. Ageing exacerbates ribosome pausing to disrupt cotranslational proteostasis. *Nature* **601**, 637-642 (2022). https://doi.org:10.1038/s41586-021-04295-4
76 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261-272 (2020). https://doi.org:10.1038/s41592-019-0686-2
77 Waskom, M. *et al.* mwaskom/seaborn: v0.9.0. *zenodo* https://doi.org:10.5281/zenodo.1313201
78 Gene Ontology, C. Gene Ontology Consortium: going forward. *Nucleic Acids Res* **43**, D1049-1056 (2015). https://doi.org:10.1093/nar/gku1179
79 Kendall, M. G. & Stuart, A. Inference and Relationship *The Advanced Theory of Statistics* **2** (1973).