A concerted increase in readthrough and intron retention drives transposon expression during aging and senescence

Authors: Kamil Pabis 123, Diogo Barardo 23, Kumar Selvarajoo 5678, Jan Gruber 14, Brian K. Kennedy 123

1 Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore
2 Healthy Longevity Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore
3 Centre for Healthy Longevity, National University Health System
4 Science Divisions, Yale-NUS College, Singapore, 138527, Singapore
5 Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore.
6 Singapore Institute of Food and Biotechnology Innovation, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore.
7 NUS Synthetic Biology for Clinical and Technological Innovation (SynCTI), National University of Singapore, Singapore, Singapore.
8 School of Biological Sciences, Nanyang Technological University, Singapore, Singapore.

Aging and senescence are characterized by pervasive transcriptional dysfunction, including increased expression of transposons and introns. Our aim was to elucidate mechanisms behind this increased expression. Most transposons are found within genes and introns, with a large minority being close to genes. This raises the possibility that transcriptional readthrough and intron retention are responsible for age-related changes in transposon expression rather than expression of autonomous transposons.

To test this, we compiled public RNA-seq datasets from aged human fibroblasts, replicative and drug-induced senescence in human cells and RNA-seq from aging mice and senescent mouse cells. Indeed, our reanalysis revealed a correlation between transposons expression, intron retention and transcriptional readthrough across samples and within samples. Both intron retention and readthrough increased with aging or cellular senescence and these transcriptional defects were more pronounced in human samples as compared to those of mice.

In support of a causal connection between readthrough and transposon expression, analysis of models showing induced transcriptional readthrough confirmed that they also show elevated transposon expression. Taken together, our data shows that elevated transposon reads during aging seen in various RNA-seq dataset are concomitant with multiple transcriptional defects. Intron retention and transcriptional readthrough are the most likely explanation for the expression of transposable elements that lack a functional promoter.

38

39 **INTRO**

40 Repetitive DNA makes up at least 50% of the human genome (**de Koning et al. 2011**)
41 and has been linked to genomic instability, cancer and – somewhat less consistently –
42 species longevity (**Pabis et al. 2021, Khristich and Mirkin 2020**). Transposons are one
43 particularly abundant class of repetitive sequences found in the nuclear genome of
44 eukaryotes and are strongly associated with cancer-causing structural variants like DNA
45 deletions, insertions or inversions (**Rodriguez-Martin et al. 2020**). Although transposons
46 are often considered parasitic, harmful or, at best, neutral recent evidence suggests
47 they may contribute to genomic diversification on evolutionary timescales and to the
48 cellular stress-response by providing transcription factor binding sites (**Villanueva-Cañas**
49 **et al. 2019**).

50 The major transposon families in the human genome are long interspersed nuclear
51 element (LINEs), short interspersed nuclear element (SINEs), long terminal repeat
52 transposons (LTRs) and DNA transposons comprising 21%, 11%, 8% and 3% of the
53 human genome, respectively (**Kazazian and Moran 2017**).

54 In recent years, transposon expression, particularly of elements belonging to the LINE-1
55 family, has been hypothesized to be a near universal marker of aging. Several lines of
56 evidence support a link between transposon reactivation and aging (**Gorbunova et al.**
57 **2021**).

58 Using RNA-seq, for example, to quantify reads mapping to transposable elements it was
59 shown that expression of these was higher in 10 vs 1 day-old nematodes (**LaRocca,**
60 **Cavalier, Wahl 2020**), 40 vs 10 day-old flies (**Wood et al. 2016**) and in muscle and liver
61 of aged mice (**De Cecco et al. 2013**). Transposon expression also increases during in vitro
62 senescence (**Colombo et al. 2018**) and in fibroblasts isolated from people between the
63 ages of 1 and 96 years (**LaRocca, Cavalier, Wahl 2020**), whereas lifespan extending
64 mutations and interventions in mice reduce expression of transposons (**Wahl et al. 2021**)

65 However, not all studies report such consistent age-related increases across all
66 transposon classes (**Ghanam et al. 2019**). Moreover, it would be a mistake to conflate
67 changes in RNA-seq-, RNA-, DNA- and protein-based measurements of transposon
68 expression as evidence for one and the same phenomenon.

69 Quantification of transposon expression using RNA-seq or PCR techniques remains
70 challenging given their repetitive nature and the fact that 98 to 99% of all transposons
71 are co-expressed with neighboring transcriptional units (**Deininger et al. 2017, Stow et**

al. 2021). In fact, only a small fraction of LINE-1 and SINE elements, the latter relying on co-transposition by LINE elements, are expressed from a functional promoter. It is believed that such "hot" LINE-1 loci drive most transposition events and genomic instability (**Deininger et al. 2017, Rodriguez-Martin et al. 2020**).

The large number of co-expressed transposons would likely mask any signal from autonomous elements in standard RNA-seq experiments. Therefore, alternative explanations for increased transposon expression in such datasets are needed and these could include age-related changes in transposon adjacent genes e.g. intron retention and transcriptional readthrough.

Most eukaryotic genes contain introns which need to be removed during a complex co-transcriptional process called splicing. Defects in splicing are the cause of many, often neurologic, hereditary conditions (**Scotti and Swanson 2016**). One such type of splicing defect is intron retention. Although basal levels of intron retention may be benign or even physiologic, high levels are considered harmful (**Zheng et al. 2020**). Aging is accompanied by many changes to splicing patterns (**Wang et al. 2018**), including intron retention, which is increased during cellular senescence (**Yao et al. 2020**) and in aging drosophila, mouse hippocampus and human prefrontal cortex. In addition, levels in human AD brain tissues are elevated even further than in the aged brain (**Adusumalli et al. 2019**).

Transcription is terminated by the cleavage and polyadenylation machinery when the polymerase transcribes through the polyadenylation signal leading to pre-mRNA cleavage, polymerase pausing, conformational change and eventually polymerase detachment (**Rosa-Mercado et al. 2021**). However, this termination may fail leading to so called readthrough transcription which increases during stress conditions (**Rosa-Mercado and Steitz 2022**) and during cellular senescence, at least for a subset of genes (**Muniz et al. 2017**).

Emerging evidence suggests that intron retention, readthrough and transposon expression are linked (**Rosa-Mercado et al. 2021, Hadar et al. 2022**), but this has not been studied in the context of aging. In this manuscript we show that intron retention and readthrough taken together can explain most of the apparent increase in transposon expression seen in RNA-seq datasets of aging.

RESULTS

1. Genomic location of transposons implicates them as a marker of multiple transcriptional defects

During aging and senescence expression of all four major transposon families (LTR, DNA, LINE, SINE) increases (**Table S1**). However, such a global increase in transposon expression cannot be explained by transposon biology alone since most transposons are ancient, and, having accumulated many mutations, lack a functional promoter (**Özgen et al. 2019).** Instead, we reasoned that transposon expression is associated with the expression of nearby genes.

Consistently, out of 4.5 million annotated transposons 60% were located within genes (most of them intronic), 10% within 5 kb of a gene and the remaining 30% were between genes, albeit usually still within 100 kb (**Table 1**). For transposons whose expression significantly changed with age or senescence, the number of intronic and downstream-of-gene transposons is even higher.

Based on these findings we studied intronic transposons as a potential marker for age-related intron retention and downstream-of-gene transposons as a marker of age-related readthrough transcription.

Dataset	Genic	Intronic	ds	us	ds	us	n
all transposons	0.604	0.493	0.045	0.049	0.187	0.174	4520928
Fleischer et al. (aging, significant)	0.857	0.564	0.082	0.025	0.105	0.037	7673
Colombo et al. (senescence, significant)	0.698	0.540	0.107	0.053	0.177	0.118	8730
	within genes		within 5kb		within 100kb		

Table 1.
Fraction of genic or intronic transposons and of transposons downstream (ds) or upstream (us) of genes in two different datasets.

2. Elevated transposon expression and intron retention with age

First, we confirmed that age-related transposon expression is indeed elevated as shown by **LaRocca et al. (2020).** This analysis is based on a dataset originally published by **Fleischer et al. (2018)**, who performed RNA-seq of 143 fibroblasts isolated from donors between the ages of 0 to 96. This dataset will be hereafter referred to as the “aging dataset”. We find that transposon expression increases slowly and non-linearly during aging, with a sudden and sharp increase around 80 years of age (**Fig. 1A, Fig. S1**). Next,

we showed that intron retention is also elevated with aging as shown by **Yao et al. (2020).** We find that intron retention increases with age (**Fig. 1B**) and that samples with high transposon expression also showed elevated intron retention (**Fig. 1C**).

In our analysis we corrected for gene co-expression by normalizing the expression levels of introns and transposons to the expression levels of the closest gene. However, another complementary approach would be to plot and compare the raw, unadjusted log-fold changes for genes, introns, transposons and other elements. In this analysis, we found that, on average, age-related upregulation of transposons and introns is more pronounced than upregulation of genes (**Fig. S2A**), consistent with **Fig. 1A** and **1B**.

Next, we re-analyzed a dataset originally published by **Colombo et al. (2018)** which will be referred to as the "senescence dataset". Here, we tested whether MDAH041 cells induced into senescence with H2O2, 5-azacytidine, adriamycin or serial passaging show elevated transposon expression and intron retention. As expected, all these treatments increased the expression of transposons (**Fig. 1D**, **Fig. S2B**), as documented before in this dataset (**Colombo et al. 2018**). The retention of introns was also increased (**Fig. 1E**) and expression levels of introns and transposons correlated with each other (**Fig. 1F**).

Having established that transposons reads and retained introns are elevated with aging and senescence, we tested whether this is conserved between mouse and human. Transposon expression ($p<0.0001$) was elevated, and intron retention showed a trend towards elevation ($p=0.14$), in the liver of 26 month-old mice, suggesting these age-related transcriptional defects might be conserved between species (**Fig. S3A, B**). In contrast, we found that increased expression of transposons and intron retention in models of cellular senescence is not conserved between species. To the contrary, early passage mouse fibroblasts showed elevated transposon expression and intron retention compared to senescent fibroblasts (**Fig. S3C, D**), which contrasts with the mouse liver and human senescence data.

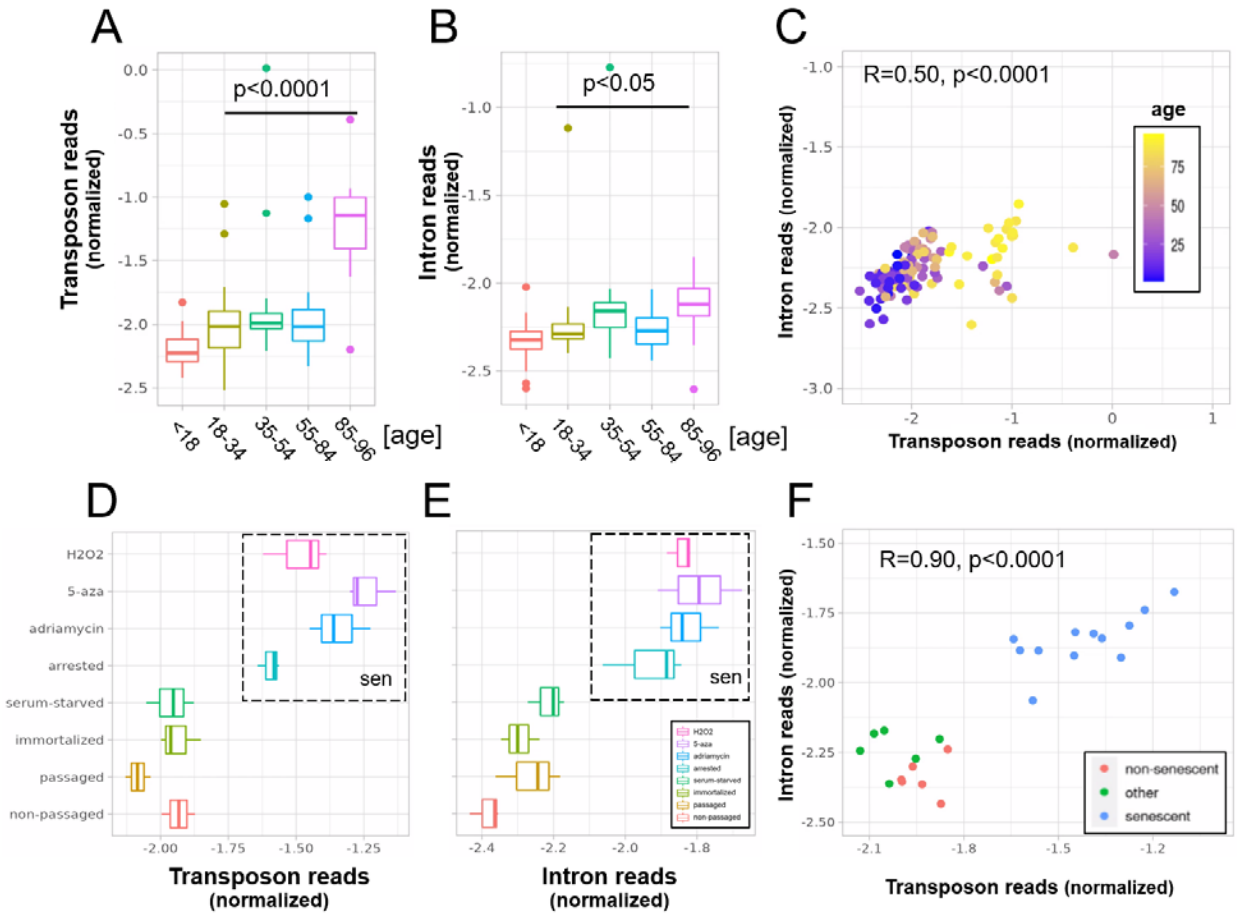


Fig. 1. A concerted increase in transposon and intron expression with aging and senescence.

Transposon (A) and intron (B) expression is increased in skin fibroblasts isolated from aged individuals. What is more, expression of both transposons and introns shows a significant correlation (C). Similarly, transposon (D) and intron (E) expression is increased in MDAH041 cells induced into senescence (sen) by treatment with drugs or via passaging. Normalized reads from the top 1000 differentially expressed genes, transposons and introns were used in this analysis.

A) Transposon reads normalized by the expression of adjacent genes plotted across five age groups (adolescent n=32, young n=31, middle-aged n=22, old n=37 and very old n=21).

B) Intron reads normalized by the expression of adjacent genes plotted across five age groups (adolescent n=32, young n=31, middle-aged n=22, old n=37 and very old n=21).

C) Scatterplot between transposon and intron expression (normalized as in A and B) for all 143 samples. Each sample is colored by age.

D) Transposon reads normalized by the expression of adjacent genes plotted across four senescent conditions (H2O2, 5-azacytidine, adriamycin, replicative senescence) and four

other conditions (serum-starved, immortalized, intermediate passage and early passage). N=3 per group.
E) Intron reads normalized by the expression of adjacent genes plotted across four senescent conditions (H2O2, 5-azacytidine, adriamycin, replicative senescence) and four other conditions (serum-starved, immortalized, intermediate passage and early passage). N=3 per group.
F) Scatterplot between transposon and intron expression (normalized as in E and F) for all 24 samples. Each sample is colored by senescence status.

3. Transposon expression as a marker of intron retention

So far these age-related events have been considered independently, but given the localization of transposons being mostly intragenic we reasoned that intron retention could bias the transposon findings. Indeed, we found that samples with high intron retention also show high transposon expression. This is true both in the aging dataset (**Fig. 1C**) and in the cellular senescence dataset (**Fig. 1F**), where intronic reads accounted for most of the observed variability in transposon expression (or vice versa). Looking at individual genomic loci, we also found a strong correlation between the log-fold change of a transposon and the log-fold change of the intron it is located in (**Fig. S4**).

As a more direct test of this hypothesis, we determined the correlation between the counts of retained introns and the counts of intronic transposons for each sample individually. Non-intronic transposons and randomized transposons served as a negative control. Consistent with the log-fold change data, we found a specific correlation between intronic transposon and intron counts in both datasets, with stronger correlations in the senescence dataset (**Fig. S5**).

Altogether this data suggests that intron retention contributes to the age-related increase in the expression of transposons (**Fig. 2**).

Fig. 2. A schematic model of transposon expression.
In our model, transcription can give rise to mRNAs and pre-mRNAs that contain retained introns when co-transcriptional splicing is impaired, as seen during aging and senescence, and these often contain transposon sequences. In addition, transcription can give rise to mRNAs and pre-mRNAs that contain transposon sequences towards the 3'-end of the mRNA when co-transcriptional termination at the polyadenylation signal (PAS) is impaired as seen with aging and senescence. Some of these RNAs may be successfully polyadenylated whereas others will be subject to nonsense mediated decay. Image created by Biorender.

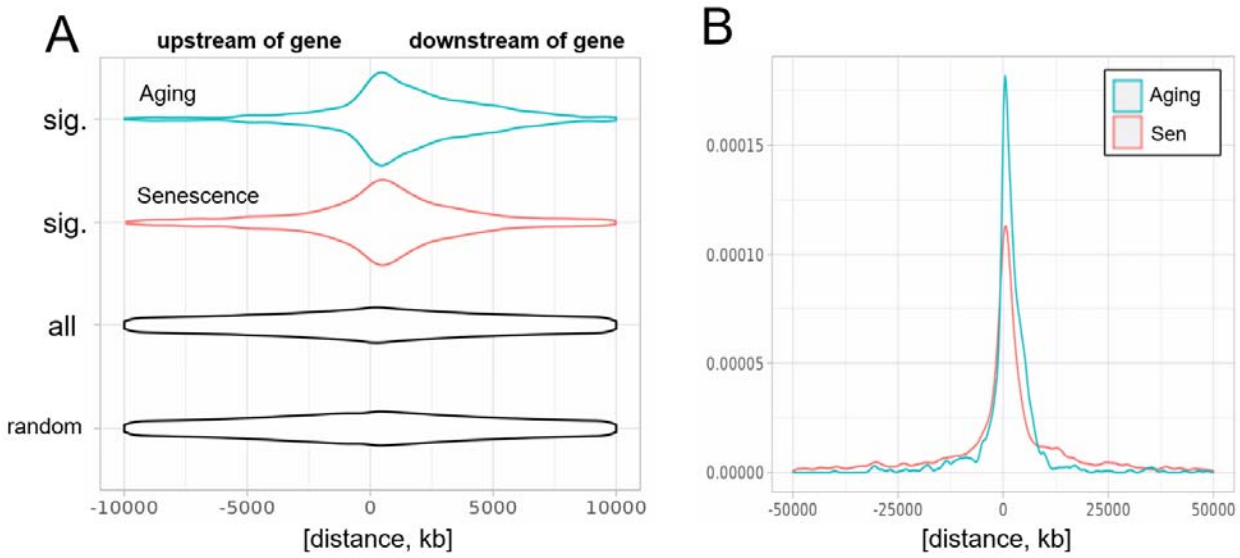
4. Expression of extragenic transposons as a marker of readthrough transcription

Most transposons showing increased expression with aging or cellular senescence mapped within genes or introns and their expression changes may be attributed to intron retention. How can we explain expression of the remaining 14 to 30% of transposons? (**Table 1**)

To answer this question, we analyzed the distribution of transposons mapping to extragenic regions. Out of all the transposons, showing significantly changed expression during aging or cellular senescence, most within 5 kb of genes and they were found predominantly downstream of genes (**Fig. 3A**). The distribution of transposons in the aging and senescence dataset was very similar except for a wider distribution of senescence-associated transposons within a 100 kb window around genes (**Fig. 3B**). This

223 contrasts with the almost completely even distribution of randomly permuted
224 transposons.

225



226

227 **Fig. 3. Biased distribution of extragenic transposons relative to genes.**
228 (A) Extragenic transposons with significantly (sig.) changed expression during aging or
229 cellular senescence show a biased distribution, when mapped back onto the genome,
230 with a preference towards a 5 kb region at the 3'-end of genes when compared to all
231 annotated transposons (all) or randomly permuted transposons (random). (B)
232 Extragenic transposons whose expression changes during cellular senescence are spread
233 out further from genes compared to aging-associated transposons.

234 Given the proximity of significantly changed transposons and genes, this suggested that
235 an age-related increase in transcriptional readthrough could explain elevated
236 transposon expression. To test this, we next quantified readthrough in a 10 to 20 kb
237 region downstream of genes. We chose a region not immediately adjacent to genes to
238 avoid biases through basal levels of readthrough, although the data was comparable for
239 other distances (**Fig. S7**).

240 Indeed, readthrough transcription was increased with aging (**Fig. 4A, B**), and in the
241 senescence dataset, levels of readthrough transcription increased with replicative or
242 chemically induced senescence (**Fig. 4D, E**). Transposon expression correlated with
243 readthrough levels on a per sample basis in both the aging and the senescence dataset
244 (**Fig. 4C, F**).

As shown before for intron retention, these findings may not translate to other species. We found that readthrough was unchanged in the liver of 26 month-old mice (p=0.22, **Fig. S6A**). In addition, readthrough decreased with replicative senescence of mouse fibroblasts, contrary to our expectations (**Fig. S6C, D**).

Next, we quantified readthrough across a 100 kb region downstream of genes. We found that both aging and senescence lead to readthrough across the whole region, although senescence was associated with more extensive readthrough further downstream of genes (**Fig. S7, Fig. 3B**). As another line of evidence we quantified transcription upstream of genes, which is called read-in. This kind of transcription is often attributed to widespread readthrough extending from the end of a gene to the start of another (**Roth et al. 2020**). Consistent with the extensive readthrough levels observed we found elevated read-in transcription during cellular senescence (**Fig. S8A**) but not aging (**Fig. S8B**).

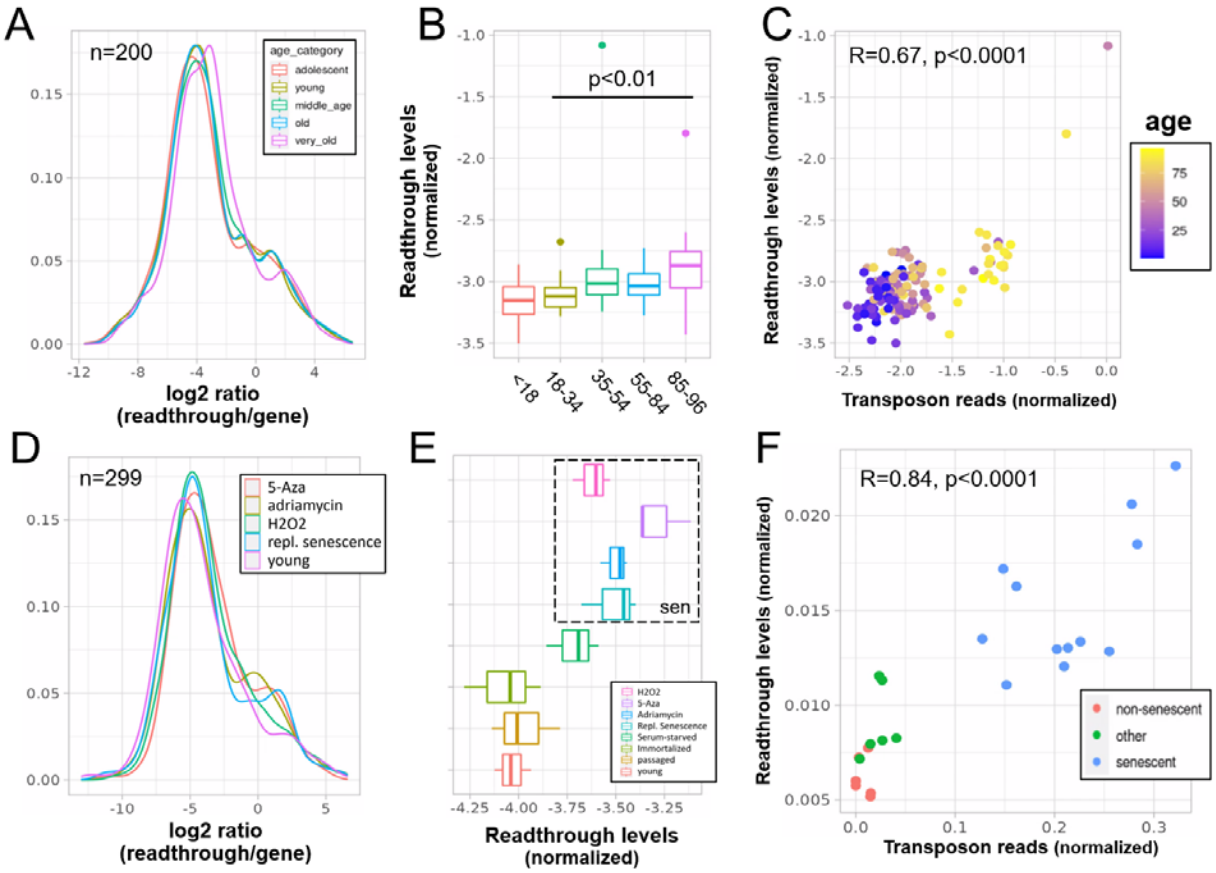


Fig. 4. Increased readthrough levels with aging and senescence.
Readthrough transcription is increased in fibroblasts isolated from the very old (A, B) and after induction of senescence in vitro (D, E). Readthrough transcription is correlated with transposon expression (C, F).
A) Readthrough was determined in a region 10 to 20 kb downstream of genes for a subset of genes that were at least 20 kb away from the nearest neighboring gene (n=206 regions). The log2 ratio of readthrough to gene expression is plotted across five age groups (adolescent n=32, young n=31, middle-aged n=22, old n=37 and very old n=21).
B) As in (A) but data is plotted on a per sample basis.
C) Scatterplot between transposon expression and readthrough levels (normalized as in A and Fig. 1) for all 143 samples. Each sample is colored by age.
D) Readthrough was determined in a region 10 to 20 kb downstream of genes for a subset of genes that were at least 20 kb away from the nearest neighboring gene (n=299 regions). The log2 ratio of readthrough to gene expression is plotted across four senescent conditions (H2O2, 5-Aza, Adriamycin, replicative senescence) and for early passage cells. N=3 per group.
E) As in (D) but data is plotted on a per sample basis and for additional control datasets (serum-starved, immortalized, intermediate passage and early passage). N=3 per group.
F) Scatterplot between transposon expression and readthrough levels (normalized as in A and Fig. 1) for all 24 samples. Each sample is colored by senescence status.

Having shown that samples with higher levels of readthrough also have higher transposon expression (per sample correlation), we asked whether transposons are more often than expected located in readthrough regions (per locus correlation). Indeed as expected, in both datasets, transposons significantly changed with age or senescence were more often found in significant readthrough regions compared to all transposons (**Table S2**). Moreover, there was a significant correlation between the log-fold changes of transposons downstream of genes with the log-fold changes of readthrough downstream of the same genes (**Fig. S9**). Transposons located elsewhere showed a weaker correlation with readthrough (**Table S3**).

To further test whether these correlations between log-fold changes are also reflected on the count level, we determined the correlation between readthrough counts and the counts of transposons downstream of genes for each sample individually. Transposons located upstream of genes and randomized transposons served as a negative control. Consistent with the above data, we found a specific correlation between transposon and readthrough counts in both datasets (**Fig. S10**).

296 Finally, we asked whether readthrough levels are predictive of transposon expression on
297 a per sample basis independently of intron retention. Using the aging dataset with 143
298 samples we compared a linear model just including intron retention with one also
299 including readthrough and found the second model to perform significantly better
300 ($p<0.0001$, **Table S4**). In addition, we found that the correlation between readthrough
301 levels and transposon expression remained significant even when correcting for intron
302 retention using partial correlation ($R=0.53$, $p<0.0001$). We did not test this in the
303 senescence dataset due to its limited size.
304

305 **5. Readthrough is a causal inducer of transposon expression**

306 Based on this correlative evidence (**Fig. 4, Fig. S9**), we hypothesized that specific
307 inducers of readthrough would also increase transposon expression. To test this, we re-
308 analyzed literature data from studies involving multiple treatments that promote
309 readthrough. These included viral infection, hyperosmotic stress (KCl treatment), heat
310 shock and oxidative stress.

311 We found that KCl treatment of HEK293 cells for 1 hour (**Rosa-Mercado et al. 2021**)
312 indeed strongly induced both readthrough (**Fig. 5A**) and transposon expression (**Fig. 5B**)
313 as determined by RNA-seq and TTLseq. Consistent with this, in our reanalysis of a study
314 using mNET-seq profiling of nascent mRNA (**Bauer et al. 2018**), KCl treatment of HEK293
315 cells also induced readthrough (**Fig. 5D**) and transposon expression (**Fig. 5E**), whereas
316 influenza infection caused an intermediate phenotype. Finally, we analyzed RNA-seq
317 data from NIH-3T3 mouse fibroblasts subjected to three stressors over two hours
318 (**Vilborg et al. 2017**). Heatshock caused extensive readthrough (**Fig. S11A**) and
319 transposon expression (**Fig. S11B**), whereas hydrogen peroxide treatment and KCl
320 caused moderate expression of these.

321 We also reasoned that readthrough should specifically increase the expression of
322 transposons that are close to genes. Consistent with this notion, induced readthrough
323 led to preferential expression of gene proximal transposons (i.e. those within 25 kb of
324 genes), when compared with senescence or aging (**Fig. S12A**). Moreover, the fraction of
325 significantly upregulated transposons found outside of genes was higher with induced
326 readthrough than with senescence or aging (**Fig. S12B**).

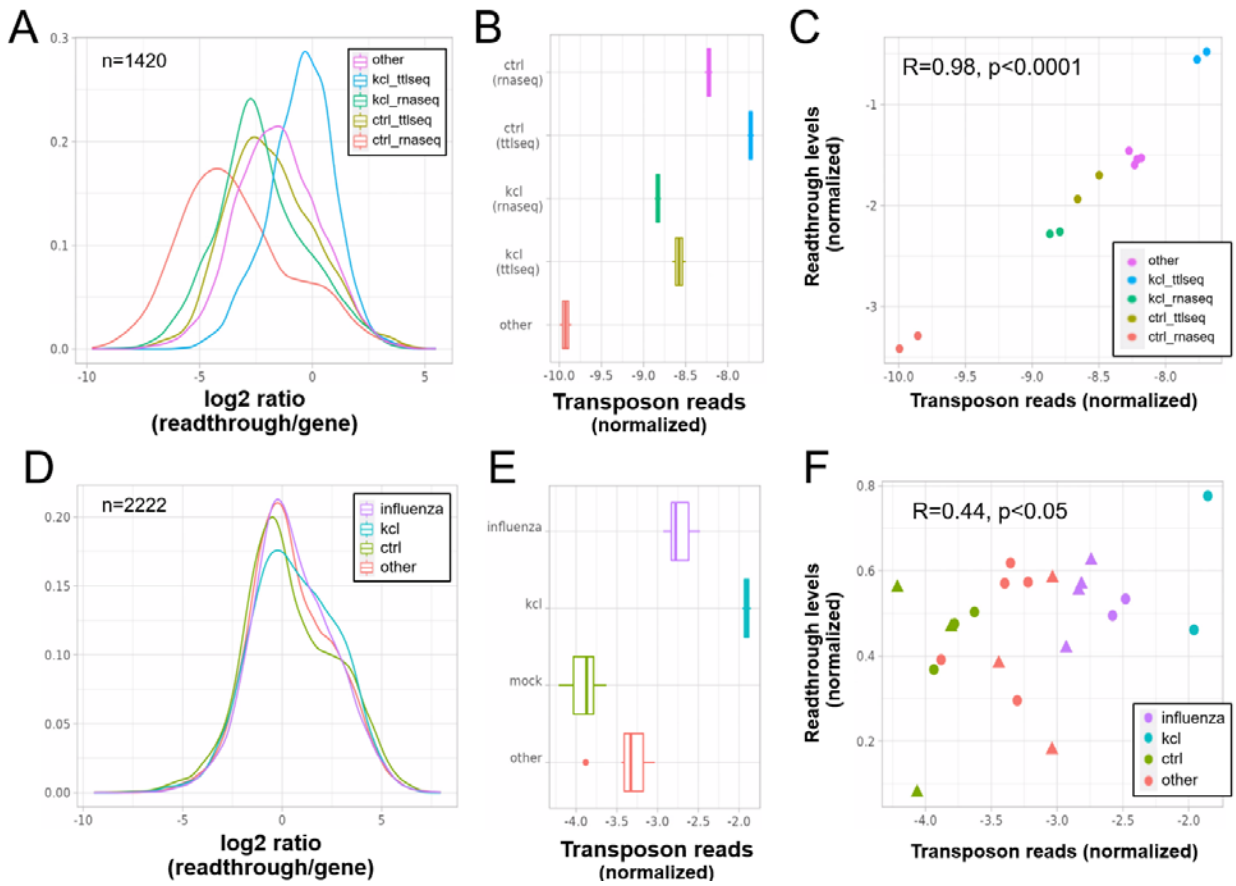


Fig. 5. Elevated transposon expression after induced readthrough.
Readthrough transcription is increased after KCl treatment of HEK293 cells (A, D) and to a lesser extent during influenza infection (D). These readthrough inducing treatments also promote increased transposon expression (B, E). Giving rise to a correlation between readthrough and transposon expression on a per sample basis (C, F). Data in (A-C) is from **Rosa-Mercado et al. (2021)**. In this study HEK293 cells were subjected to 1 hour of hyperosmotic stress (KCl) after which RNA-seq and TTLseq were performed. Data in (D-F) is from **Bauer et al. (2018)**. In this study HEK293 cells were subjected to 1 hour of hyperosmotic stress (KCl) after which mNETseq was performed.
A) Readthrough was determined in a region 10 to 20 kb downstream of genes for a subset of genes that was at least 20 kb away from the nearest neighbouring gene (n=1420 regions).
B) To normalize transposon expression counts for each transposon were corrected for the expression of the nearest gene. Normalized transposon counts for each sample are shown as box-whisker plot.
C) Normalized transposon counts (as in B) and readthrough levels are plotted for each sample (n=12).

D) Readthrough was determined in a region 10 to 20 kb downstream of genes for a subset of genes that was at least 20 kb away from the nearest neighbouring gene (n=2222 regions).
E) To normalize transposon expression counts for each transposon were corrected for the expression of the nearest gene. Normalized transposon levels for each sample are shown as box-whisker plot.
F) Normalized transposon counts (as in B) and readthrough counts are plotted for each sample (n=22). The shapes indicate the cell type (circles=HEK293, triangles=A549).

6. Increased expression of autonomous LINE-1 elements with senescence but not aging

Many expressed LINE-1 elements are found within genes (**Table S5**), which poses a challenge for the quantification of active LINE-1 elements. Given this proximity it is unsurprising that a substantial correlation between the expression of transposons and neighboring genes (**Fig. 6**) is seen in RNA-seq.

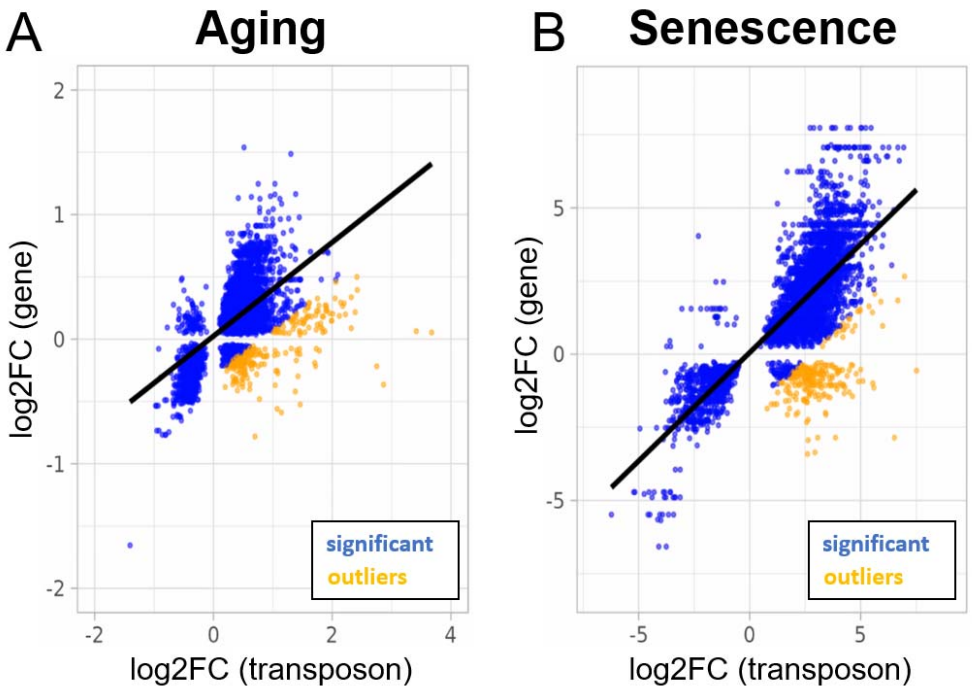
To mitigate this problem, we selected 250 transposons that showed higher than expected expression levels compared to the nearest gene. We reasoned that LINE-1 elements with a functional promoter ("active" LINE-1s) would be more likely autonomous and should be enriched among transposons that show a divergent expression profile from their neighboring genes.

We found no enrichment for active LINE-1 elements in the aging dataset (**Fig. 6A**). In contrast, transposons significantly upregulated with senescence and, among those, especially the ones with higher-than-expected expression, showed significant enrichment for active LINE-1 elements (**Fig. 6B**). Moreover, we found that almost 100% of active LINE-1 elements were increased with senescence which was higher than for all LINE-1 elements and other transposons (**Table S5**).

Finally, we redid the same analysis with a different definition of active LINE-1 elements as those potentially encoding functional ORF1p and ORF2p. In this analysis, we included all elements encoding an open reading frame > 900 nt with high sequence similarity to ORF1p and >3400 nt with high sequence similarity to ORF2p ($p<0.05$ for pairwise alignment). Out of 122626 annotated LINE-1 elements 585 encoded a putative ORF1p and 133 encoded an ORF2p.

ORF1p-encoding LINE-1 elements were significantly enriched among transposons with higher-than-expected expression and among transposons significantly changed with senescence or aging (**Table S7**). These findings were more robust for senescence than

for aging. ORF2p-encoding LINE-1 elements, in contrast, were too rare to be studied reliably as to their enrichment.



	active LINE1	LINE1	active%	p value
outliers	2	79	2.53	NS
significant	55	2820	1.95	NS
all expressed	549	12214	4.49	

	active LINE1	LINE1	active%	p value
outliers	21	91	25.61	<0.0001
significant	340	2506	13.57	<0.0001
all expressed	1538	27167	5.66	

Fig. 6. Increased expression of active LINE-1 elements with senescence but not aging. A scatterplot between the log2-fold change of every transposon ($p<0.05$) differentially expressed with age or senescence and the log2-fold change of the nearest gene. The percentage of active LINE-1 elements among transposons with higher-than-expected expression levels is unchanged in aging (A) and increased during senescence (B). For the tables, we compared the percentage of active LINE-1 elements among all LINE-1 elements for 250 transposons with higher-than-expected expression levels (in orange; labeled "outliers"), all significantly changed transposons (blue plus orange; labeled "significant") and all transposons passing read filtering (not plotted; labeled "all expressed"). The p-value is based on the comparison with "all expressed".

DISCUSSION

396 We show that transposons would serve as a useful biomarker of aging, even if most of
397 the transposon signal in RNA-seq was due to co-expression with nearby transcriptional
398 units, because they capture different kinds of transcriptional dysfunction.

399 **Introns**

400 Intron retention is one prominent example of dysfunctional splicing during aging, but
401 the association with age-related transposon expression has not been evaluated. A
402 recent study by **Gualandi et al. (2022)** using RNA-seq data from the 1000 Genomes
403 project found a strong per sample correlation between intron and transposon
404 expression in the general population. In line with these results, we showed that intron
405 retention may explain a large fraction of transposon reads in aging and senescence
406 datasets (**Fig. 1**).

407 **Readthrough, read-in and transposons**

408 In a study of oncogene-Induced senescence, **Muniz et al. (2017)** found increased
409 readthrough transcription at 91 so called START loci, which was associated with
410 suppression of neighboring genes through the production of antisense RNAs and several
411 of these antisense RNAs were necessary for the maintenance of the senescent state.
412 Our data is consistent with their model, although we find even more extensive
413 readthrough transcription. Analyzing hundreds of loci we show that senescence
414 promotes transcriptional readthrough leading to the overexpression of intergenic
415 transposons (>25 kb from the nearest gene) and read-in at neighboring genes (**Fig. 3B,**
416 **Fig. S8B**). In contrast, aging-related readthrough was qualitatively different, producing
417 shorter readthrough and no read-in (**Fig. 3B, Fig. S8A**).

418 Our correlative data suggested that readthrough transcription could account for the
419 expression of many transposons that are outside of genes and thus cannot be explained
420 by intron retention (**Fig. 4C, Fig. 4F**). To test this hypothesis, we reanalyzed data from
421 models of induced readthrough showing that readthrough is indeed a causal inducer for
422 the expression of gene proximal transposons after hyperosmotic stress and viral
423 infection (**Fig. 5**). Given the technical challenges of measuring readthrough and the
424 unreliability of intron retention detection tools (**David et al. 2022**) our results are likely
425 underestimating the true magnitude of correlation between readthrough and
426 transposon expression.

427 In a previous study it was observed that senescence leads to increased expression of
428 transposons and of genes with low abundances (**Zhang et al. 2021**). This “leakage” was
429 attributed to loss of H3K9me3 histone marks and increased heterochromatin
430 accessibility, but is also consistent with readthrough transcription, as we observed.

431 Indeed, it has been suggested that readthrough is the most likely cause for apparent
432 intergenic transcription (**Agostini et al. 2021**).

433 **Is there genuine expression of autonomous transposons during aging?**

434 The distribution of transposons showing apparent age-related expression within the
435 genome is an important question since intragenic transposons are more likely to be co-
436 expressed with genes. Using Sanger sequencing of long-range amplicons to study
437 replicative senescence, **De Cecco et al. (2019)** found a large fraction of transposons
438 mapping to intergenic (69%) regions and few to intragenic/gene-proximal regions (31%).
439 In contrast using RNA-seq data, we find that most transposons are located within genes
440 (≥70%; **Table 1**), although, there are more intergenic transposons significantly changed
441 with senescence than with aging (**Fig. 3**). One explanation for this discrepancy is the
442 mapping approach used. Since their analysis did not use an expectation-maximization
443 algorithm to map back transposons onto the genome, it may underestimate the number
444 of intragenic and gene-adjacent transposons in contrast to our analysis using TELocal.

445 Given that inactive LINE-1 loci greatly outnumber active loci (**Deininger et al. 2017, Stow**
446 **et al. 2021**), and we saw only limited evidence for active LINE-1 transcription in
447 fibroblasts isolated from donors of varying ages (**Fig. 6A**), our data appears to question
448 the role of transposons in organismal aging. An intriguing explanation is provided by **De**
449 **Cecco et al. (2019)**, who found that LINE-1 ORF1 protein-positive cells colocalize with a
450 small subset of senescent cells. Most likely, such changes would not be obvious from
451 bulk sequencing, perhaps explaining why we did not see elevated LINE-1 expression in
452 the aging dataset. In contrast, the signal would be readily detectable in cultures induced
453 into senescence where most cells do enter senescence, which is consistent with our
454 findings (**Fig. 6B**).

455 As for another explanation, we speculate that LINE-1 elements expressed through intron
456 retention and readthrough are compatible with translation via leaky scanning or re-
457 initiation (**Mouilleron et al. 2015**), accounting for increased LINE-1 ORF1 protein levels
458 detected in senescent cells, even if leaky scanning is orders of magnitude less efficient
459 than normal initiation.

460 **Summary**

461 Taken together, our work and existing literature suggest that RNA-seq, and by extension
462 qPCR, of transposable elements overestimates their true expression and their age-
463 related changes. Such data alone is insufficient as support for a "transposon hypothesis
464 of aging" that postulates an increase in autonomous expression of such elements.
465 Instead, a concerted increase in transposon expression, intron retention and

466 transcriptional readthrough points to a global increase in transcriptional dysfunction
467 during aging and senescence, at least in human cells.

468 Nevertheless, there are several other techniques such as immunoblotting for ORF1
469 protein, rapid amplification of cDNA ends (RACE), or genomic PCR that may (and do)
470 provide evidence in favor of age-related reactivation of potentially autonomous
471 transposons, specifically LINE-1 elements.

472 Although we have shown increased transcriptional dysfunction with aging and cellular
473 senescence, so far we lack a model to explain these findings. In this context it is
474 noteworthy that the elongation rate of Pol II is correlated with readthrough
475 transcription (**Fong et al. 2015**), intron processing, and may increase with aging (**Debès**
476 **et al. 2019**), potentially providing a unifying explanation.

477 **METHODS**

478 **Datasets, read preprocessing and alignment**

479 We searched the SRA for suitable datasets that used total RNA extraction because this
480 improves the detection of non-canonical transcripts that may be degraded before
481 successful polyadenylation. Publicly available human RNA-seq datasets of aging and
482 senescence, respectively, were obtained from Colombo et al. 2018 (GSE60340) and
483 Fleischer et al. 2018 (GSE113957). Human RNA-seq and TTLseq data was obtained from
484 Rosa-Mercado et al. 2021 (GSE152059), human mNET-seq data from Bauer et al. 2018
485 (PRJNA432639) and mouse RNA-seq data from Vilborg et al. 2017 (GSE98906).

486 Mouse liver RNA-seq data from ad libitum and dietary restricted mice was obtained
487 from Hahn et al. 2017 (GSE92486) and mouse replicative senescence data from Wang et
488 al. 2022 (GSE179880).

489 Reads were filtered, adaptor-trimmed and repaired using fastp (**Chen et al. 2018**) with a
490 min Phred score of 25. Afterwards reads passing these filters were aligned to the human
491 genome (GRCh38) or mouse genome (GRCm39) using STAR 2.7.8a (**Dobin et al. 2013**) in
492 single-pass mode. To optimize the quantification of transposons the STAR parameters --
493 outFilterMultimapNmax and --winAnchorMultimapNmax were set to 100. The full
494 pipeline with all the configuration files is available on github (pabisk/aging_transposons).

495 **Detection of readthrough and intron retention**

496 We employed a modified Automatic Readthrough Transcription Detection (ARTDeco)
497 pipeline, using their provided annotations (**Roth et al. 2020**) but with a different
498 quantification approach. Readthrough candidate regions in the ARTDeco pipeline were

499 defined based on the longest transcript to exclude the effects of differential 3'-UTR
500 usage. Genic and readthrough reads were counted using featureCounts from the
501 Rsubread package in 10 kb bins downstream of genes, in a window between 0 to 100 kb.
502 Readthrough regions with low coverage, in close proximity to, or overlapping, nearby
503 protein-coding genes were excluded from the analysis. Intron retention was detected
504 using the intron REtention Analysis and Detector (iREAD), which is a tool that detects
505 retention events based on both splice junction reads and intron expression levels (**Li et**
506 **al. 2020**). All samples were treated as unstranded.

507 **Locus specific detection of transposon reads and transcripts**

508 We used TELocal v1.1.1 (mhammell-laboratory/TElocal) to quantify transposon reads,
509 which is functionally similar to TETranscripts (**Jin et al. 2015**) except it provides locus-
510 specific expression levels. Both tools use an expectation maximization algorithm to
511 assign ambiguous reads. Annotation files were downloaded from the mhammell lab and
512 from genecode (v39 for human and M27 for mouse). TELocal was also used to quantify
513 transcriptomic counts. Again, all samples were treated as unstranded.

514 **Expression and normalization of genes, introns, transposons and readthrough**
515 **transcripts**

516 Transcripts and elements with low expression counts were filtered. After filtering, intron,
517 readthrough, read-in or transposon elements were combined with transcript counts for
518 the analysis in DESeq2. For the aging fibroblast dataset age was normalized and
519 centered and the design model considered age+gender+ethnicity. For the induced
520 senescence dataset all four inducers were included together and the remaining
521 treatment groups were considered non-senescent.

522 In order to study expression of these elements unbiased by gene coexpression, we
523 selected the top 1000 overexpressed elements and divided their read counts by the
524 read counts of the nearest gene. When not specified in the annotation the nearest
525 genes or elements were identified with the bedtools closest or intersect function, as
526 appropriate. Finally, to provide a single expression value for each sample we used the
527 mean log of the normalized expression values.

528 **Expression of potentially active and protein-coding LINE-1 elements**

529 We defined potentially active LINE-1 elements as those that have a functional promoter
530 (**McKerrow and Fenyö 2020**). Whereas to define potentially protein-coding LINE-1
531 elements we predicted open reading frames for each annotated element using the
532 predORF function from the systemPipeR package. Global pairwise alignment to ORF1p

533 (uniprot: Q9UN81) and ORF2p (uniprot: O00370) was performed and only elements with
534 a significant alignment were included for further analyses ($p<0.05$).

535

536 **ACKNOWLEDGEMENTS**

537 We would like to thank Juliane Liepe, Marco Malavolta and Chin-Tong Ong for
538 support and/or their help in improving this manuscript and VitaDAO for financial
539 support.

540

541 **REFERENCES**

542 Adusumalli, S., Ngian, Z.-K., Lin, W.-Q., Benoukraf, T., & Ong, C.-T. (2019). Increased intron retention is a
543 post-transcriptional signature associated with progressive aging and Alzheimer's disease. Aging Cell,
544 18(3), e12928. https://doi.org/10.1111/acel.12928

545 Agostini, F., Zagalak, J., Attig, J., Ule, J., & Luscombe, N. M. (2021). Intergenic RNA mainly derives from
546 nascent transcripts of known genes. Genome Biology, 22(1), 136. https://doi.org/10.1186/s13059-021-
547 02350-x

548 Bauer, D. L. V., Tellier, M., Martínez-Alonso, M., Nojima, T., Proudfoot, N. J., Murphy, S., & Fodor, E.
549 (2018). Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. Cell
550 Reports, 23(7), 2119-2129.e3. https://doi.org/10.1016/j.celrep.2018.04.047

551 Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor.
552 Bioinformatics, 34(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

553 Colombo, A. R., Elias, H. K., & Ramsingh, G. (2018). Senescence induction universally activates
554 transposable element expression. Cell Cycle, 17(14), 1846–1857.
555 https://doi.org/10.1080/15384101.2018.1502576

556 David, J. K., Maden, S. K., Wood, M. A., Thompson, R. F., & Nellore, A. (2022). Retained introns in long
557 RNA-seq reads are not reliably detected in sample-matched short reads (p. 2022.03.11.484016). bioRxiv.
558 https://doi.org/10.1101/2022.03.11.484016

559 De Cecco, M., Ito, T., Petrashen, A. P., Elias, A. E., Skvir, N. J., Criscione, S. W., Caligiana, A., Brocculi, G.,
560 Adney, E. M., Boeke, J. D., Le, O., Beauséjour, C., Ambati, J., Ambati, K., Simon, M., Seluanov, A.,
561 Gorbunova, V., Slagboom, P. E., Helfand, S. L., … Sedivy, J. M. (2019). L1 drives IFN in senescent cells and
562 promotes age-associated inflammation. Nature, 566(7742), Article 7742.
563 https://doi.org/10.1038/s41586-018-0784-9

564 Debès, C., Papadakis, A., Grönke, S., Karalay, Ö., Tain, L., Mizi, A., Nakamura, S., Hahn, O., Weigelt, C.,
565 Josipovic, N., Zirkel, A., Brusius, I., Sofiadis, K., Lamprousi, M., Lu, Y.-X., Huang, W., Esmaillie, R., Kubacki,
566 T., Späth, M. R., … Beyer, A. (2022). Aging-associated changes in transcriptional elongation influence
567 metazoan longevity (p. 719864). bioRxiv. https://doi.org/10.1101/719864

568 Deininger, P., Morales, M. E., White, T. B., Baddoo, M., Hedges, D. J., Servant, G., Srivastav, S., Smither,
569 M. E., Concha, M., DeHaro, D. L., Flemington, E. K., & Belancio, V. P. (2017). A comprehensive approach
570 to expression of L1 loci. Nucleic Acids Research, 45(5), e31. https://doi.org/10.1093/nar/gkw1067

571 Deniz, Ö., Frost, J. M., & Branco, M. R. (2019). Regulation of transposable elements by DNA
572 modifications. Nature Reviews Genetics, 20(7), Article 7. https://doi.org/10.1038/s41576-019-0106-6

573 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras,
574 T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21.
575 https://doi.org/10.1093/bioinformatics/bts635

576 Fleischer, J. G., Schulte, R., Tsai, H. H., Tyagi, S., Ibarra, A., Shokhirev, M. N., Huang, L., Hetzer, M. W., &
577 Navlakha, S. (2018). Predicting age from the transcriptome of human dermal fibroblasts. Genome
578 Biology, 19(1), 221. https://doi.org/10.1186/s13059-018-1599-6

579 Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M. A., Sheridan, R. M., Nguyen, T., Karp, S., &
580 Bentley, D. L. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA
581 Polymerase II Termination Suggest Widespread Kinetic Competition. Molecular Cell, 60(2), 256–267.
582 https://doi.org/10.1016/j.molcel.2015.09.026

583 Ghanam, A. R., Cao, J., Ouyang, X., & Song, X. (2019). New Insights into Chronological Mobility of
584 Retrotransposons In Vivo. Oxidative Medicine and Cellular Longevity, 2019, e2818415.
585 https://doi.org/10.1155/2019/2818415

586 Gorbunova, V., Seluanov, A., Mita, P., McKerrow, W., Fenyö, D., Boeke, J. D., Linker, S. B., Gage, F. H.,
587 Kreiling, J. A., Petrashen, A. P., Woodham, T. A., Taylor, J. R., Helfand, S. L., & Sedivy, J. M. (2021). The
588 role of retrotransposable elements in ageing and age-associated diseases. Nature, 596(7870), Article
589 7870. https://doi.org/10.1038/s41586-021-03542-y

590 Gualandi, N., Iperi, C., Esposito, M., Ansaloni, F., Gustincich, S., & Sanges, R. (2022). Meta-Analysis
591 Suggests That Intron Retention Can Affect Quantification of Transposable Elements from RNA-Seq Data.
592 Biology, 11(6), Article 6. https://doi.org/10.3390/biology11060826

593 Hadar, S., Meller, A., Saida, N., & Shalgi, R. (2022). Stress-induced transcriptional readthrough into
594 neighboring genes is linked to intron retention. IScience, 25(12), 105543.
595 https://doi.org/10.1016/j.isci.2022.105543

596 Hahn, O., Grönke, S., Stubbs, T. M., Ficz, G., Hendrich, O., Krueger, F., Andrews, S., Zhang, Q., Wakelam,
597 M. J., Beyer, A., Reik, W., & Partridge, L. (2017). Dietary restriction protects from age-associated DNA
598 methylation and induces epigenetic reprogramming of lipid metabolism. Genome Biology, 18(1), 56.
599 https://doi.org/10.1186/s13059-017-1187-1

600 Jin, Y., Tam, O. H., Paniagua, E., & Hammell, M. (2015). TEtranscripts: A package for including
601 transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics, 31(22),
602 3593–3599. https://doi.org/10.1093/bioinformatics/btv422

603 Kazazian, H. H., & Moran, J. V. (2017). Mobile DNA in Health and Disease. New England Journal of
604 Medicine, 377(4), 361–370. https://doi.org/10.1056/NEJMra1510092

605 Khristich, A. N., & Mirkin, S. M. (2020). On the wrong DNA track: Molecular mechanisms of repeat-
606 mediated genome instability. Journal of Biological Chemistry, 295(13), 4134–4170.
607 https://doi.org/10.1074/jbc.REV119.007678

608 Koning, A. P. J. de, Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. (2011). Repetitive Elements May
609 Comprise Over Two-Thirds of the Human Genome. PLOS Genetics, 7(12), e1002384.
610 https://doi.org/10.1371/journal.pgen.1002384

611 LaRocca, T. J., Cavalier, A. N., & Wahl, D. (2020). Repetitive elements as a transcriptomic marker of aging:
612 Evidence in multiple datasets and models. Aging Cell, 19(7), e13167. https://doi.org/10.1111/acel.13167

613 Li, H.-D., Funk, C. C., & Price, N. D. (2020). iREAD: A tool for intron retention detection from RNA-seq
614 data. BMC Genomics, 21(1), 128. https://doi.org/10.1186/s12864-020-6541-0

615 McKerrow, W., & Fenyö, D. (2020). L1EM: A tool for accurate locus specific LINE-1 RNA quantification.
616 Bioinformatics, 36(4), 1167–1173. https://doi.org/10.1093/bioinformatics/btz724

617 Mouilleron, H., Delcourt, V., & Roucou, X. (2016). Death of a dogma: Eukaryotic mRNAs can code for
618 more than one protein. Nucleic Acids Research, 44(1), 14–23. https://doi.org/10.1093/nar/gkv1218

619 Muniz, L., Deb, M. K., Aguirrebengoa, M., Lazorthes, S., Trouche, D., & Nicolas, E. (2017). Control of Gene
620 Expression in Senescence through Transcriptional Read-Through of Convergent Protein-Coding Genes.
621 Cell Reports, 21(9), 2433–2446. https://doi.org/10.1016/j.celrep.2017.11.006

622 Pabis, K. (2021). Triplex and other DNA motifs show motif-specific associations with mitochondrial DNA
623 deletions and species lifespan. Mechanisms of Ageing and Development, 194, 111429.
624 https://doi.org/10.1016/j.mad.2021.111429

625 Retained introns in long RNA-seq reads are not reliably detected in sample-matched short reads |
626 bioRxiv. (n.d.). Retrieved December 27, 2022, from
627 https://www.biorxiv.org/content/10.1101/2022.03.11.484016v1.abstract

628 Rodriguez-Martin, B., Alvarez, E. G., Baez-Ortega, A., Zamora, J., Supek, F., Demeulemeester, J.,
629 Santamarina, M., Ju, Y. S., Temes, J., Garcia-Souto, D., Detering, H., Li, Y., Rodriguez-Castro, J., Dueso-
630 Barroso, A., Bruzos, A. L., Dentro, S. C., Blanco, M. G., Contino, G., Ardeljan, D., … Tubio, J. M. C. (2020).
631 Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1
632 retrotransposition. Nature Genetics, 52(3), Article 3. https://doi.org/10.1038/s41588-019-0562-0

633 Rosa-Mercado, N. A., & Steitz, J. A. (2022). Who let the DoGs out? – Biogenesis of stress-induced
634 readthrough transcripts. Trends in Biochemical Sciences, 47(3), 206–217.
635 https://doi.org/10.1016/j.tibs.2021.08.003

636 Rosa-Mercado, N. A., Zimmer, J. T., Apostolidi, M., Rinehart, J., Simon, M. D., & Steitz, J. A. (2021).
637 Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription
638 despite widespread transcriptional repression. Molecular Cell, 81(3), 502-513.e4.
639 https://doi.org/10.1016/j.molcel.2020.12.002

640 Roth, S. J., Heinz, S., & Benner, C. (2020). ARTDeco: Automatic readthrough transcription detection. BMC
641 Bioinformatics, 21(1), 214. https://doi.org/10.1186/s12859-020-03551-0

642 Scotti, M. M., & Swanson, M. S. (2016). RNA mis-splicing in disease. Nature Reviews Genetics, 17(1),
643 Article 1. https://doi.org/10.1038/nrg.2015.3

644 Stow, E. C., Kaul, T., deHaro, D. L., Dem, M. R., Beletsky, A. G., Morales, M. E., Du, Q., LaRosa, A. J., Yang,
645 H., Smither, E., Baddoo, M., Ungerleider, N., Deininger, P., & Belancio, V. P. (2021). Organ-, sex- and age-
646 dependent patterns of endogenous L1 mRNA expression at a single locus resolution. Nucleic Acids
647 Research, 49(10), 5813–5831. https://doi.org/10.1093/nar/gkab369

648 Vilborg, A., Sabath, N., Wiesel, Y., Nathans, J., Levy-Adam, F., Yario, T. A., Steitz, J. A., & Shalgi, R. (2017).
649 Comparative analysis reveals genomic features of stress-induced transcriptional readthrough.
650 Proceedings of the National Academy of Sciences, 114(40), E8362–E8371.
651 https://doi.org/10.1073/pnas.1711120114

652 Villanueva-Cañas, J. L., Horvath, V., Aguilera, L., & González, J. (2019). Diverse families of transposable
653 elements affect the transcriptional regulation of stress-response genes in Drosophila melanogaster.
654 Nucleic Acids Research, 47(13), 6842–6857. https://doi.org/10.1093/nar/gkz490

655 Wahl, D., Cavalier, A. N., Smith, M., Seals, D. R., & LaRocca, T. J. (2021). Healthy Aging Interventions
656 Reduce Repetitive Element Transcripts. The Journals of Gerontology: Series A, 76(5), 805–810.
657 https://doi.org/10.1093/gerona/glaa302

658 Wang, K., Wu, D., Zhang, H., Das, A., Basu, M., Malin, J., Cao, K., & Hannenhalli, S. (2018).
659 Comprehensive map of age-associated splicing changes across human tissues and their contributions to
660 age-associated diseases. Scientific Reports, 8(1), Article 1. https://doi.org/10.1038/s41598-018-29086-2

661 Wang, Y., Liu, L., Song, Y., Yu, X., & Deng, H. (2022). Unveiling E2F4, TEAD1 and AP-1 as regulatory
662 transcription factors of the replicative senescence program by multi-omics analysis. Protein & Cell,
663 13(10), 742–759. https://doi.org/10.1007/s13238-021-00894-z

664 Wood, J. G., Jones, B. C., Jiang, N., Chang, C., Hosier, S., Wickremesinghe, P., Garcia, M., Hartnett, D. A.,
665 Burhenn, L., Neretti, N., & Helfand, S. L. (2016). Chromatin-modifying genetic interventions suppress
666 age-associated transposable element activation and extend life span in Drosophila. Proceedings of the
667 National Academy of Sciences, 113(40), 11277–11282. https://doi.org/10.1073/pnas.1604621113

668 Yao, J., Ding, D., Li, X., Shen, T., Fu, H., Zhong, H., Wei, G., & Ni, T. (2020). Prevalent intron retention fine-
669 tunes gene expression and contributes to cellular senescence. Aging Cell, 19(12), e13276.
670 https://doi.org/10.1111/acel.13276

671 Zhang, X., Liu, X., Du, Z., Wei, L., Fang, H., Dong, Q., Niu, J., Li, Y., Gao, J., Zhang, M. Q., Xie, W., & Wang,
672 X. (2021). The loss of heterochromatin is associated with multiscale three-dimensional genome
673 reorganization and aberrant transcription during cellular senescence. Genome Research, 31(7), 1121–
674 1135. https://doi.org/10.1101/gr.275235.121

675 Zheng, J.-T., Lin, C.-X., Fang, Z.-Y., & Li, H.-D. (2020). Intron Retention as a Mode for RNA-Seq Data
676 Analysis. Frontiers in Genetics, 11. https://www.frontiersin.org/articles/10.3389/fgene.2020.00586