iScience, Volume *26*

Supplemental information

Age or lifestyle-induced accumulation of genotoxicity is associated with a length-dependent decrease in gene expression

Olga Ibañez-Solé, Irantzu Barrio, and Ander Izeta

Supplemental figures and legends

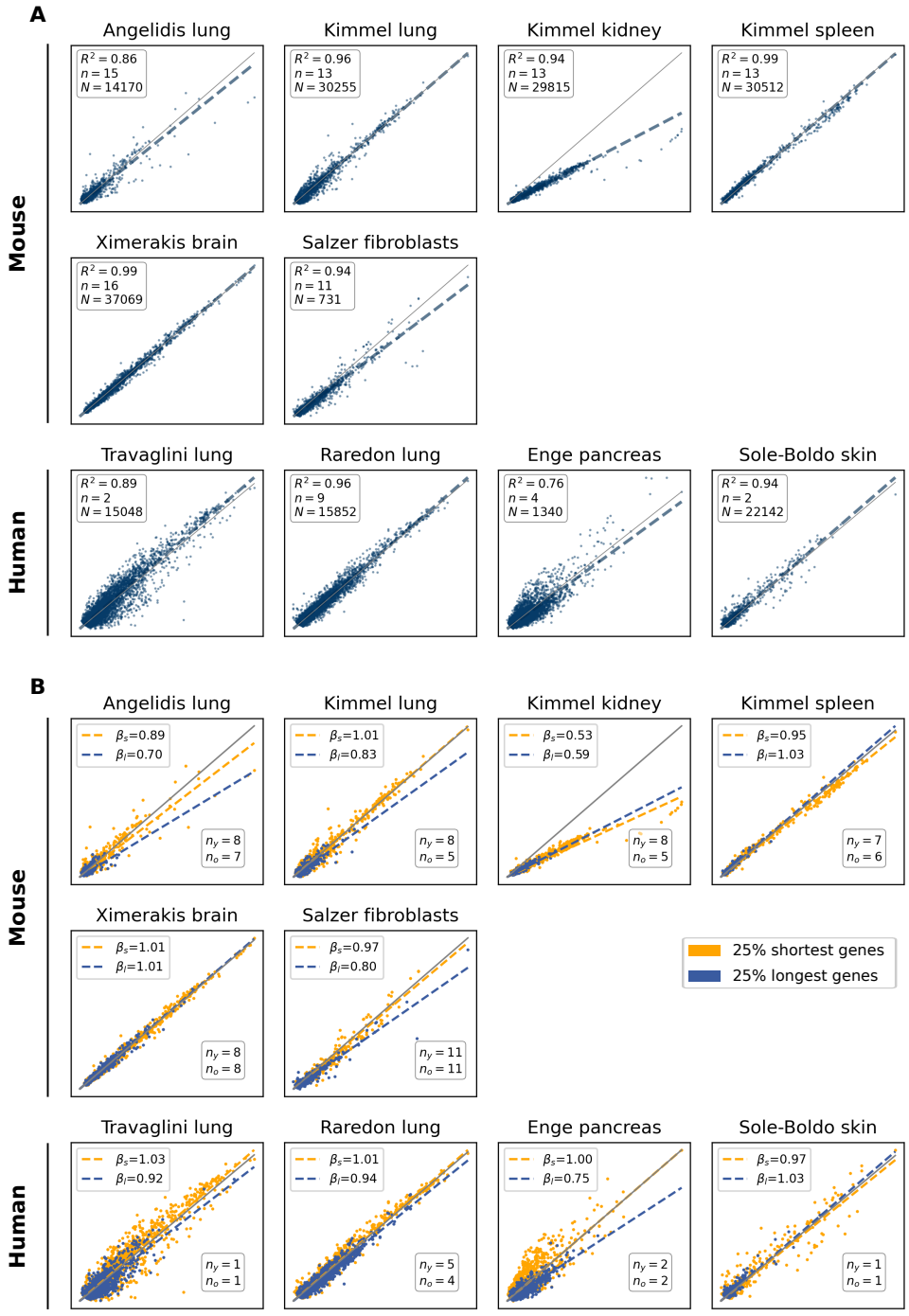

Figure S1. Downregulation of long genes with aging is replicated in several datasets of different species, related to Figure 1. (**A**) Gene expression is conserved with aging in several datasets of different species. Average gene expression in old against young cells in six mouse and four human datasets of several tissues. R^2: coefficient of determination; N: total number of cells; n: number of biological replicates. (**B**) Age-associated shutdown of transcription is found to be gene length-dependent in several datasets of different species. β_s and β_l correspond to the slopes of the multiple linear regression models with interaction fitted on the 1st and 4th quartiles (top 25% shortest and top 25% shortest genes). Number of biological replicates in each age category: young (n_y) and old (n_o).

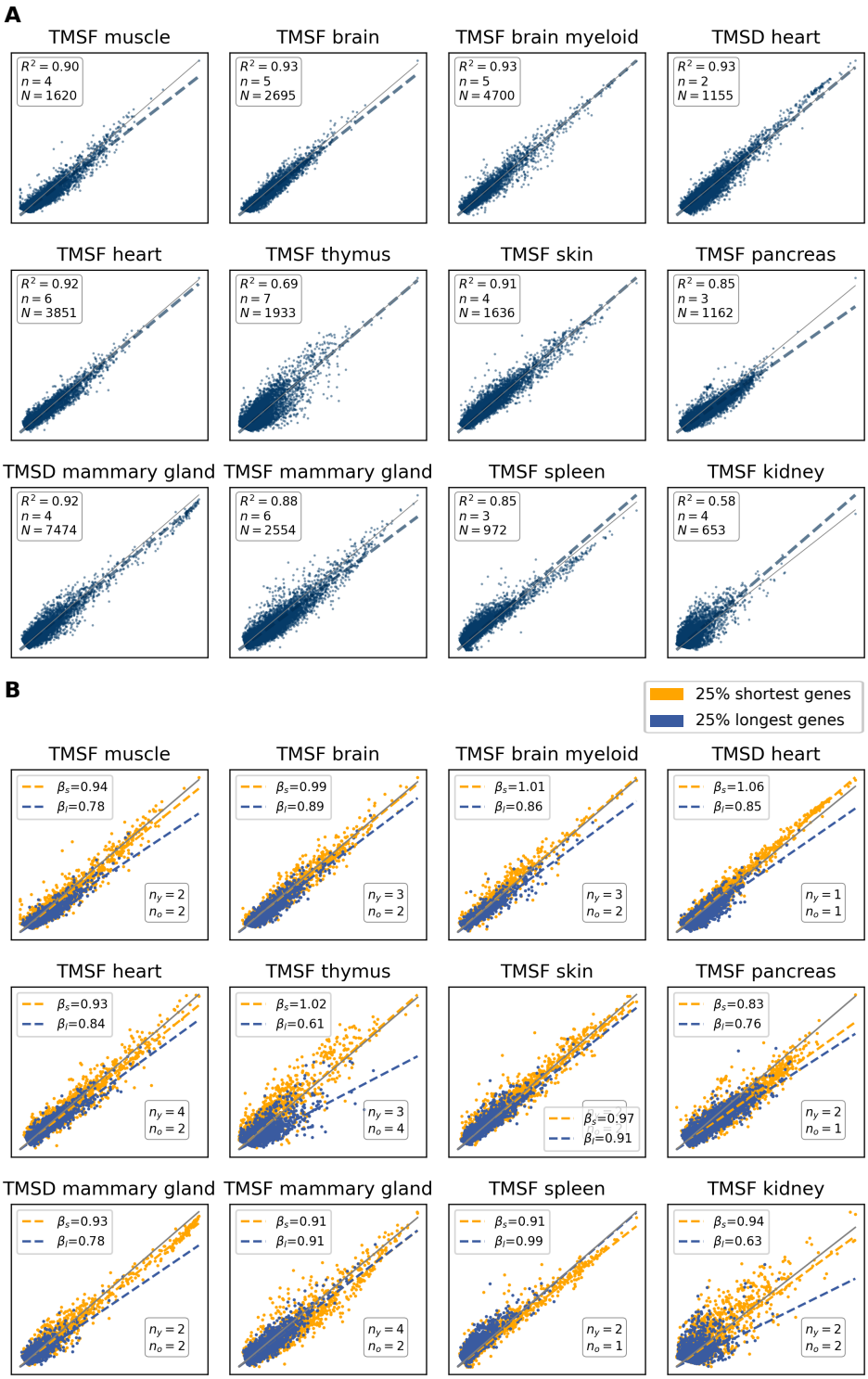

Figure S2. Age-associated shutdown of transcription is also detected in 18-month-old females, related to Figure 1. (**A**) Gene expression is highly conserved but shows a detectable decay with aging in 18-month-old female mice as well. Scatter plots showing the average gene expression in 18-month-old female mice against average gene expression in 3-month-old female mice in 11 tissues (12 comparisons) from the TMS FACS and the TMS droplet datasets[20] . Each dot represents a gene. N: number of single cells; n: number of biological replicates. R^2: coefficient of determination. (**B**) Age-associated shutdown of transcription preferentially affects long genes. The scatter plots show the average gene expression in 18-month-old versus 3-month-old female mice. The top 25% and bottom 25% of the total genes according to their gene length are shown in blue and yellow, respectively. β_s and β_l correspond to the slopes of the multiple linear regression models with interaction fitted on the 1st and 4th quartiles (top 25% shortest and top 25% shortest genes). Number of young (n_y) and old (n_o) biological replicates are shown.

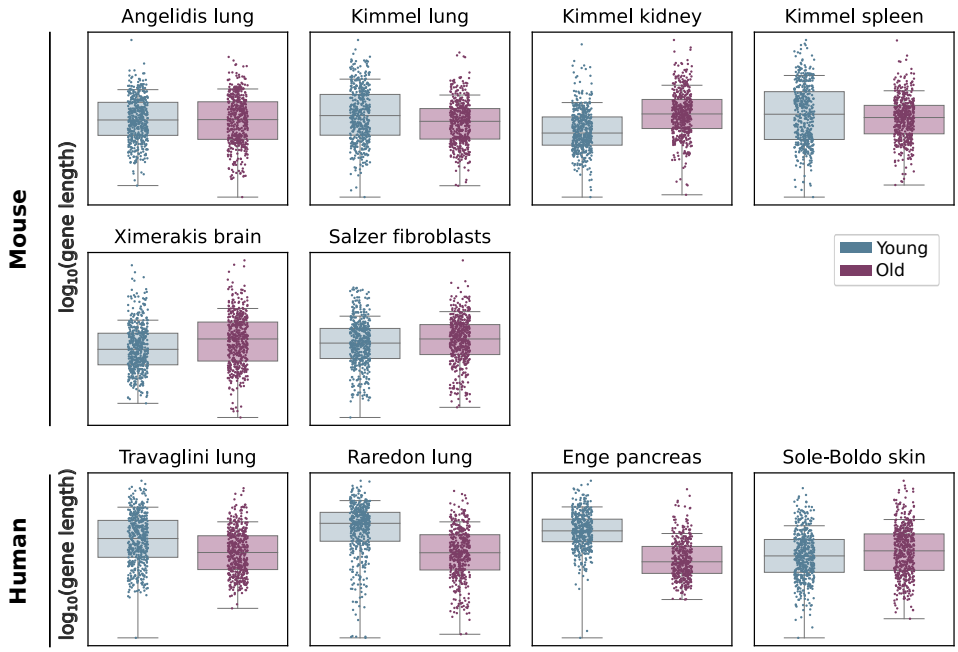

Figure S3. Downregulation of long genes is found in several datasets of different species, related to Figure 2. Top 300 DEGs between young and old cells in 10 independent aging datasets from mouse and human. The 300 differentially expressed genes between young and old individuals were obtained using the Wilcoxon method.

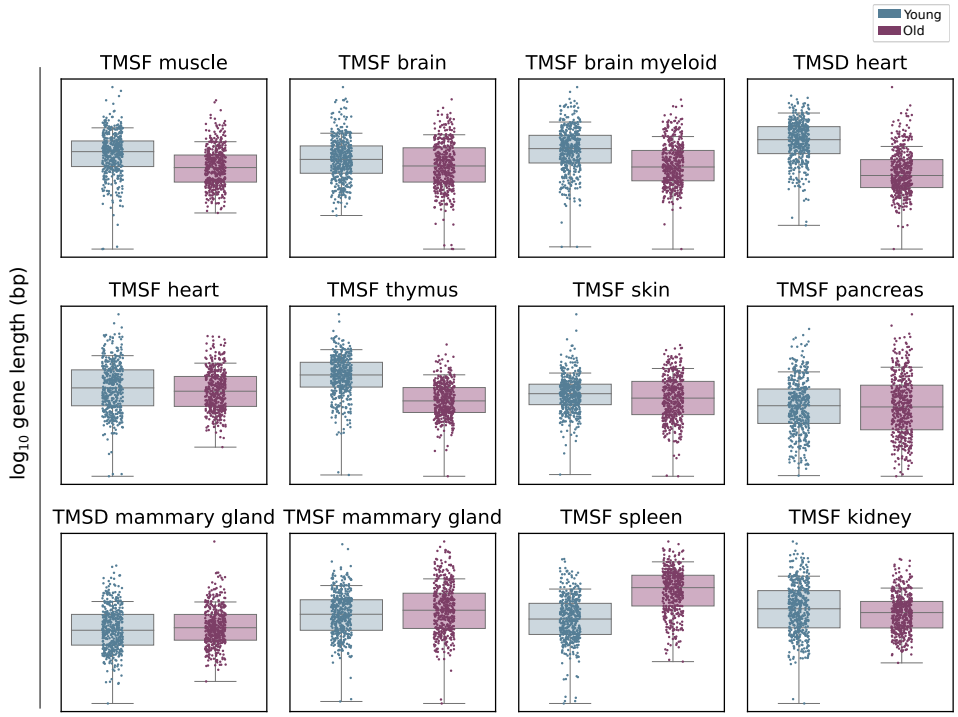

Figure S4. Downregulation of long genes is also detected in 18-month-old female mice, related to Figure 2. Top 300 DEGs between young and old cells in 12 aging datasets from the Tabula Muris Senis[20]. The 300 differentially expressed genes between 3-months-old and 18-months-old female mice were obtained using the Wilcoxon method.

Supplemental tables

	25% shortest (Q1)			25% longest (Q4)		
	short_min	short_median	short_max	long_min	long_median	long_max
Bladder	65	5,373	10,218	58,647	104,278	2,270,723
Brain	63	5,402	10,143	52,003	93,157	1,211,426
Brain myeloid	69	5,521	10,418	57,559	103,285	2,257,271
Heart	69	5,196	9,713	52,057	95,055	2,270,723
Kidney	63	5,402	9,905	49,274	88,337	1,503,513
Liver	108	5,494	10,367	56,554	101,606	2,960,898
Lung	63	5,574	10,615	59,384	108,898	2,960,898
Muscle	64	5,897	11,150	63,517	118,298	2,960,898
Pancreas	67	5,526	10,471	55,760	100,580	2,960,898
Skin	63	5,411	10,151	56,744	101,808	2,960,898
Spleen	63	5,379	9,987	50,379	89,933	1,503,513
Thymus	63	5,538	10,287	54,303	99,058	2,960,898

Table S1. Length of the Q1 (25% shortest) and Q4 (25% longest) genes used in the analysis of Figure 1. The minimum (min), median and maximum (max) gene lengths (bp) are shown for the two gene categories (short, long).

	Q1		Q1-Q2		Q1-Q3		Q1-Q4	
	Est. (SE)	p-val	Est. (SE)	p-val	Est. (SE)	p-val	Est. (SE)	p-val
Bladder	1.02 (<0.01)	0	-0.03 (0.01)	<0.001	-0.05 (0.01)	<0.001	-0.07 (0.01)	<0.001
Brain	0.86 (<0.01)	0	-0.25 (0.01)	<0.001	-0.35 (0.01)	<0.001	-0.43 (0.01)	<0.001
Brain myeloid	1.03 (<0.01)	0	-0.06 (0.01)	<0.001	-0.11 (0.01)	<0.001	-0.29 (0.01)	<0.001
Heart	1.02 (<0.01)	0	-0.09 (0.01)	<0.001	-0.17 (0.01)	<0.001	-0.28 (0.01)	<0.001
Kidney	0.93 (<0.01)	0	-0.03 (0.01)	3.85E-02	-0.08 (0.02)	<0.001	-0.19 (0.02)	<0.001
Liver	0.86 (<0.01)	0	-0.06 (0.01)	<0.001	-0.08 (0.01)	<0.001	-0.20 (0.02)	<0.001
Lung	1.16 (0.01)	0	-0.24 (0.02)	<0.001	-0.39 (0.02)	<0.001	-0.50 (0.02)	<0.001
Muscle	1.26 (0.01)	0	-0.36 (0.02)	<0.001	-0.50 (0.02)	<0.001	-0.62 (0.02)	<0.001
Pancreas	0.85 (<0.01)	0	-0.03 (0.01)	2.83E-02	-0.06 (0.01)	<0.001	-0.14 (0.01)	<0.001
Skin	1.09 (<0.01)	0	-0.13 (0.01)	<0.001	-0.21 (0.01)	<0.001	-0.29 (0.01)	<0.001
Spleen	0.99 (<0.01)	0	0.05 (0.01)	<0.001	-0.03 (0.01)	3.80E-02	-0.19 (0.01)	<0.001
Thymus	1.02 (<0.01)	0	-0.13 (0.02)	<0.001	-0.25 (0.02)	<0.001	-0.41 (0.02)	<0.001

Table S2. Linear models fit on short and long genes are significantly different in 12 murine aging mouse datasets, related to Figure 1. We test for the difference between the slope that best fits the old vs young average gene expression using the Q1 genes (25% shortest) and the slope that corresponds to each of the other three quartiles (Q2, Q3, Q4). Q1-Q2, Q1-Q3 and Q1-Q4 represent the differences between the slopes fitted on Q1 and each of the quartiles. Est. (estimate), SE (standard error), p-val (*p*-value).

	U statistic	*p*-value
Bladder	28948.5	5.23e-10
Brain	27401.0	2.52e-10
Brain myeloid	13075.0	1.45e-43
Heart	12005.5	5.54e-45
Kidney	22024.0	9.91e-21
Liver	31636.0	0.000227
Lung	10844.0	7.16e-52
Muscle	8774.5	7.31e-59
Pancreas	25380.0	6.00e-12
Skin	12953.5	1.35e-44
Spleen	19386.0	5.45e-25
Thymus	10888.5	1.97e-50

Table S3. Mann-Whitney test comparing lengths of DEG between young and old cells U statistic and *p*-value associated with each comparison, related to Figure 2. The test compares the mean log_{10} gene length (bp) of the top 300 DEGs between young and old cells in 12 murine tissues.

	Q1		Q1-Q2		Q1-Q3		Q1-Q4	
	Est. (SE)	p-val	Est. (SE)	p-val	Est. (SE)	p-val	Est. (SE)	p-val
TMSD F (3-18)	1.06 (0.00)	<0.001	-0.07 (0.01)	<0.001	-0.14 (0.01)	<0.001	-0.21 (0.01)	<0.001
TMSD F (3-21)	1.06 (0.00)	<0.001	-0.04 (0.01)	<0.001	-0.12 (0.01)	<0.001	-0.22 (0.01)	<0.001
TMSD M (1-18)	0.97 (0.00)	<0.001	-0.04 (0.01)	<0.001	-0.05 (0.01)	<0.001	-0.09 (0.01)	<0.001
TMSD M (1-24)	0.98 (0.00)	<0.001	-0.03 (0.01)	<0.001	-0.04 (0.01)	<0.001	-0.07 (0.01)	<0.001
TMSF F (3-18)	0.93 (0.00)	<0.001	-0.01 (0.01)	7.84E-2	-0.05 (0.01)	<0.001	-0.09 (0.01)	<0.001
TMSF M (3-24)	1.02 (0.01)	<0.001	-0.09 (0.01)	<0.001	-0.17 (0.01)	<0.001	-0.28 (0.01)	<0.001

Table S4. Output of the statistical analysis comparing the effects of the different gene length groups based on a linear model with interaction, related to Figure 3. We test for the difference between the slope that best fits the old vs young average gene expression using the Q1 genes (25% shortest) and the slope that corresponds to each of the other three quartiles (Q2, Q3, Q4). Q1-Q2, Q1-Q3 and Q1-Q4 represent the differences between the slopes fitted on Q1 and each of the quartiles. Datasets: murine heart and aorta TMSD (Tabula muris senis droplet), TMSF (Tabula muris senis FACS), age of the cohorts are shown in parentheses (months).

		Q1		Q1-Q2		Q1-Q3		Q1-Q4	
Data	Comparison	Est. (SE)	p-val	Est. (SE)	p-val	Est. (SE)	p-val	Est. (SE)	p-val
Lin	H vs UV	0.91 (<0.01)	0	-0.07 (<0.01)	<0.001	-0.10 (0.01)	<0.001	-0.11 (0.01)	<0.001
	VD vs UV	0.91 (<0.01)	0	-0.08 (<0.01)	<0.001	-0.10 (0.01)	<0.001	-0.13 (0.01)	<0.001
	H vs VD	0.98 (<0.01)	0	0.01 (<0.01)	0.392	-0.01 (0.01)	0.201	0.00 (0.01)	0.903
Gold.	H vs Smoker	1.01 (<0.01)	0	-0.03 (<0.01)	<0.001	-0.06 (0.01)	<0.001	-0.12 (0.01)	<0.001
Muld.	WT vs ADH5KO	1.00 (<0.01)	0	0.02 (<0.01)	<0.001	0.01 (0.01)	0.171	0.01 (0.01)	0.0969
	WT vs CSBKO	1.00 (<0.01)	0	0.03 (<0.01)	<0.001	0.03 (<0.01)	<0.001	0.04 (0.01)	<0.001
	WT vs DKO	0.99 (<0.01)	0	-0.02 (<0.01)	4.68E-03	-0.05 (0.01)	<0.001	-0.06 (0.01)	<0.001
Wang	GC: ct vs UV	0.94 (<0.01)	0	0.00 (<0.01)	0.691	0.01 (0.01)	9.30E-03	0.00 (0.01)	0.893
	mut: ct vs UV	0.94 (<0.01)	0	-0.04 (<0.01)	<0.001	-0.07 (0.01)	<0.001	-0.10 (0.01)	<0.001
	ct: GC vs mut	0.97 (<0.01)	0	0.02 (<0.01)	<0.001	0.02 (<0.01)	<0.001	0.00 (<0.01)	0.958
	UV: GC vs mut	0.99 (<0.01)	0	-0.03 (<0.01)	<0.001	-0.08 (<0.01)	<0.001	-0.12 (0.01)	<0.001

Table S5. Statistical significance of the analyses done on premature aging datasets, related to Figures 4-7. We test for the difference between the slope that best fits *y* (condition 2) vs *x* (condition 1) average gene expression using the Q1 genes (25% shortest) and the slope that corresponds to each of the other three quartiles (Q2, Q3, Q4). Q1-Q2, Q1-Q3 and Q1-Q4 represent the differences between the slopes fitted on Q1 and each of the quartiles. Datasets: Lin[33], Gold. (Goldfarbmuren)[35], Muld. (Mulderring)[38] and Wang[31]. Comparisons between conditions: H (healthy), UV (UV-radiated), VD (UV-radiated upon vitamin D treatment), ADH5KO ($Adh5^{-/-}$), CSBKO ($Csb^{m/m}$), DKO ($Adh5^{-/-}$ $Csb^{m/m}$ double knock out), GC (gene-corrected), ct (control), mut (mutant). Est. (estimate), SE (standard error), p-val (*p*-value).