DNA methylation rates scale with maximum lifespan across mammals

Samuel. J. C. Crofts, Eric Latorre-Crespo*, Tamir Chandra*

*These authors contributed equally

Abstract

DNA methylation rates have previously been found to broadly correlate with maximum lifespan in mammals, yet no precise relationship has been observed. We compared methylation rates at conserved age-related sites across mammals in both skin and blood and found that methylation rates scale with maximum lifespan to the power of negative one. The emergence of an explicit scaling law suggests that methylation rate is either a fundamental limiting factor in maximum lifespan across species or is linked to an underlying universal constraint.

Main

Organisms display enormous variation as the result of evolution, spanning many orders of magnitude in characteristics such as size, energy requirements, and lifespan. Despite this remarkable diversity, it has been observed that life is often bound by shared underlying mechanisms and constraints[1]. These fundamental connections between organisms can be reflected in scaling laws, which mathematically describe an association between two physical quantities over several orders of magnitude.

A notable example of a scaling law in the field of biology is Max Kleiber's observation that an animal's metabolic rate is proportional to its mass to the power of three-quarters[2]. This observation was later shown to hold across not just whole organisms, but also cells, mitochondria, and enzymes, spanning a total of 27 orders of magnitude in mass[3]. It has been proposed that this relationship arises from the transport of materials through branching fractal-like networks and that evolution tends to minimise the energy required to supply these materials[4]. Such an explanation demonstrates the power of scaling laws to reveal fundamental processes that govern biological systems.

DNA methylation is an epigenetic modification in which a methyl group is added to a cytosine base followed by a guanine (CpG). Methylation status at a given CpG can vary between cells, meaning a methylation proportion can be calculated for each CpG across a population of cells. Methylation proportions of some CpGs change in a predictable way with age. This observation led to the development of the first "epigenetic clocks" in the early 2010s[5–7] which used methylation proportions of selected CpGs to predict chronological age in humans. Since then, epigenetic clocks have been extended to numerous other organisms, including the development of clocks which measure age across mammalian species[8].

Recently, DNA methylation rates have been shown to generally correlate with a species' maximum lifespan, although no consistent scaling law has been observed and the biological mechanisms behind the correlation remain unclear. Lowe et al. (2018)[9] looked at age-related CpGs across six mammals and found a negative trend between methylation rates and maximum lifespan. Similarly, Wilkinson et al. (2021)[10] looked at age-related CpGs in 26 species of bat and again found a negative correlation between methylation rate and longevity. More generally, methylation dynamics have even recently been used to develop epigenetic predictors of life-history traits[11] and to attempt to identify specific CpGs and associated genes involved in both ageing and longevity[12].

We compared the blood methylation rates of conserved age-related CpGs across seven species of mammal, representing four taxonomic orders and covering approximately two orders of magnitude in maximum lifespan. In contrast to previous studies, we restricted our analyses to CpGs which were age-related in each mammal being compared. We found that methylation rates scale with maximum lifespan to the power of negative one. We corroborated this finding by repeating our analysis on skin samples from 23 species of bat and 4 other mammals. We observed approximately the same trend in this dataset, indicating that this scaling law is independent of tissue type. The emergence of an explicit scaling law suggests that epigenetic mechanisms are linked to fundamental constraints on lifespan that are shared across species.

For each species, we conducted univariate linear regressions for each CpG of the form methylation ~ age. A CpG was considered age-associated if the adjusted R-squared value from this regression was greater than 0.1. The methylation rate of each CpG was defined as the slope from these regressions. In the case of datasets with multiple measurements for individuals, linear mixed-effects models were used with random intercepts for individuals and the marginal R^2 value was used.

We then selected a baseline species (see *Methods* for details) and compared each other species to this baseline species in a pairwise manner. For each comparison we:
1) Selected the set of conserved and age-related CpGs common to both species.
2) For each CpG in this set, calculated the absolute methylation rate for each species
3) Calculated the methylation rate ratio of the comparison species compared to the baseline species for each CpG. That is, for each CpG, *methylation rate ratio = absolute rate of comparison species/absolute rate of baseline species*. For example, a ratio of 2 would mean that the methylation rate of the species in question was twice as fast as the methylation rate of the baseline species in a particular CpG.
4) Calculated the median and interquartile range (IQR) of the resulting values across all CpGs included in the analysis.

We plotted the median methylation rate ratios in blood against maximum lifespan **(Fig. 1a)**. This revealed a relationship in which methylation rates decay to an asymptote as lifespan increases. Taking the logarithm of both axes (see *Methods* for details) resulted in a consistent linear association with the slope equal to approximately -1 (slope = -1.04, 95% CI -1.23 to -0.87) **(Fig. 1b)**. This result supports a simple relationship in which methylation rates, M, scale proportional to lifespan, L, to the power of negative 1:

$$M \propto L^{-1} \quad \textbf{(Equation 1)}$$

The error bars in **Fig. 1b** represent the IQR of rate ratios, demonstrating that this scaling is observed not just on average across CpGs, but is generally observed at the level of individual CpGs. Additionally, while the sample sizes were moderate in terms of individuals, the number of CpGs used in each comparison was large - generally in the order of a few hundred to a few thousand (**Supp. Tables 1** and **2**).

Similar results were seen in skin samples (**Fig. 1c** and **1d)** (slope = -0.87, 95% CI -1.16 to -0.59), indicating that the scaling is a general property of methylation rates across different tissue types. However, the association was less tight than seen in blood, potentially in part because the vast majority of samples were from bats, for which maximum longevity is known with less certainty than better-studied mammals such as mice and chimpanzees. Additionally, skin is an external tissue which may be more impacted by environmental adaptations, such as those acquired for living in water (e.g. beluga whale) or underground (e.g. naked mole rat).

Fig. 1e shows an individual CpG that exhibits the trend we observe on average, with decreasing methylation rates as maximum lifespan increases.

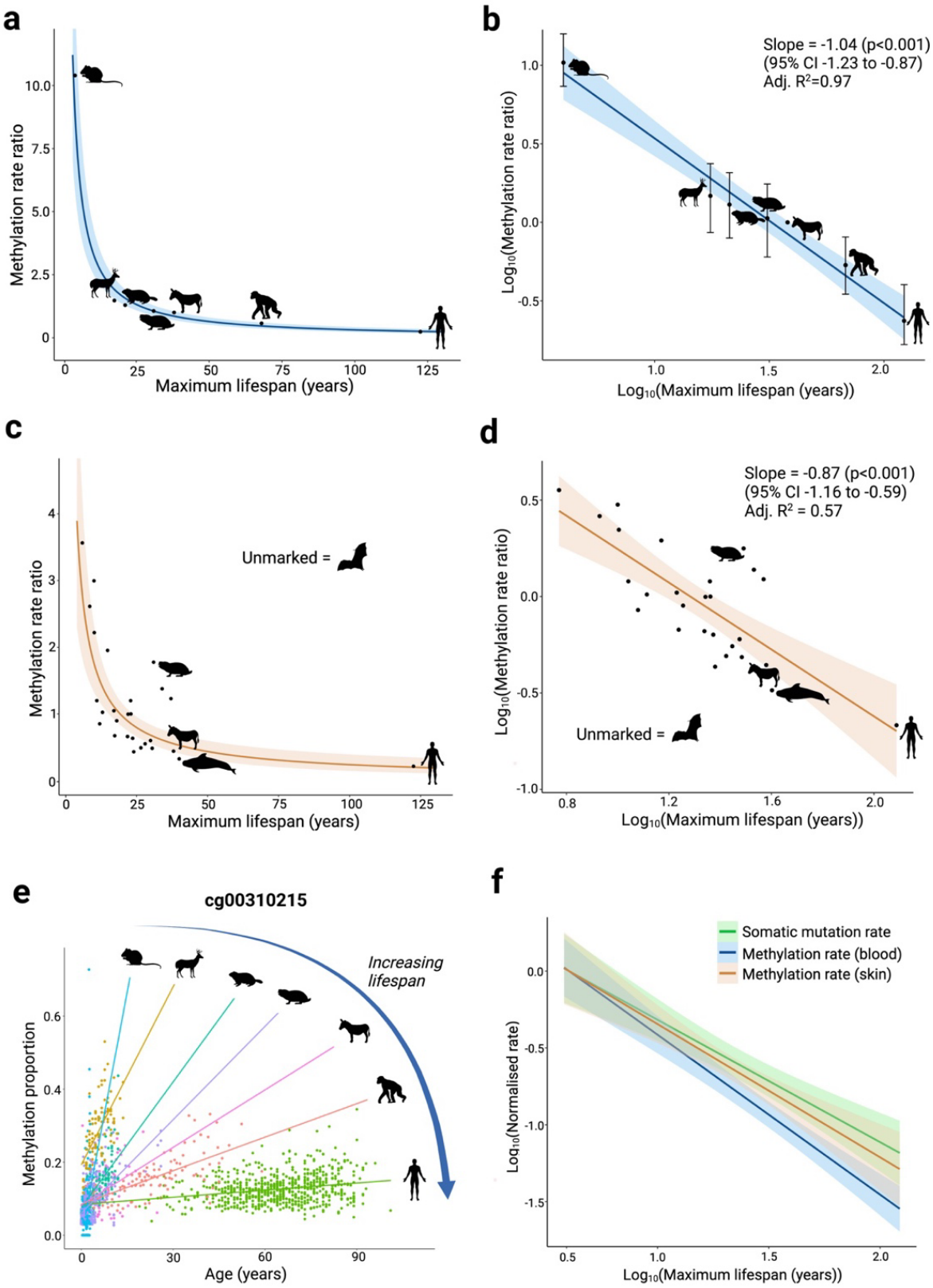


Figure 1: **a,** Methylation rate ratio versus maximum lifespan in blood samples, using rates in the zebra as a baseline. Each point is the median value. Regression line plotted from the power-law association shown in b. Shaded region represents the 95% confidence interval. From left to right, species are: house mouse (*Mus musculus*) (n=113), roe deer (*Capreolus capreolus*) (n=94), yellow-bellied marmot (*Marmota flaviventris*) (n=149 samples from 73 individuals), naked mole rat (*Heterocephalus glaber*)

(n=92), plains zebra (*Equus quagga burchellii*) (n=96), chimpanzee (*Pan troglodytes*) (n=113 samples from 83 individuals), human (*Homo sapiens*) (n=656). **b,** Same data as in a), but with axes log-transformed. Error bars show the interquartile range of values. Regression line from a simple linear regression of the form y~x. **c,** and **d,** show an equivalent analysis as a) and b), but in skin samples using the big brown bat (*Eptesicus fuscus*) as baseline. Error bars omitted for clarity. Icons represent the same species as in a) and b), but with different sample sizes (human (n=41), mole rat (n=63), zebra (n=22)) and with the addition of the beluga whale (*Delphinapterus leucas*) (n=69). Unmarked points are various species of bat (n>=15) (see Methods). **e,** Example of an individual conserved CpG in blood samples in which the methylation rates (i.e. slopes resulting from the linear regressions) decrease as lifespan increases. **f,** Log-log plots of blood and skin methylation rates versus maximum lifespan, presented alongside somatic mutation rate (per Mb) versus maximum lifespan. Y-axis represents normalised rates. Regression lines shown with points omitted, with shaded regions representing 95% confidence intervals. Intercept of regression lines adjusted so that they start at zero on the plot. Mutation rate data taken from Cagan et al. (2022)[13]. Created with BioRender.com.

Our analysis of DNA methylation data in mammals reveals a scaling law between maximum lifespan and DNA methylation rate over approximately two orders of magnitude and in two distinct tissue types. Many organism characteristics, including lifespan, scale with body mass[4,14]. However, the relationship we observe is independent of body mass, with no clear trend seen when regressing against mass instead of lifespan (**Ext. Data Fig. 1**). Notably, the naked mole rat, which is a common outlier in such cross-species comparisons due to its remarkably long life for its body mass, fits well on our trend line in blood samples.

Specifically, we find that methylation rate is approximately proportional to lifespan to the power of negative 1. This relationship means, for example, that we would expect the shortest-lived mammal (Northern short-tailed shrew) to have a methylation rate approximately 100 times faster than the longest-lived (bowhead whale), given that its maximum lifespan is approximately 100 times shorter[15]. An interesting application of such a scaling relationship is that it allows estimations of maximum lifespan for newly discovered species through longitudinal sampling, even in the absence of any knowledge of true ages.

The fact that a specific and quantitative relationship exists between methylation rate and maximum lifespan suggests that there is a fundamental biological constraint acting across diverse mammalian lineages. To further make sense of this, we can describe methylation rate, M, at a particular CpG as the product of two underlying quantities: R, the rate of stem cell division, and p, the probability that a cell division results in a change in methylation state.

$$M \propto pR \propto L^{-1} \quad \textbf{(Equation 2)}$$

This reveals that the quantity pR must be inversely proportional to lifespan. As for which these two factors may be responsible for the scaling we observe, we can gain additional insight by linking our results with the recent finding that somatic mutation rate also scales approximately with lifespan to the power of negative one[13] **(Fig. 1f)**. The fact that these two phenomena display a similar relationship with lifespan suggests two main scenarios that would explain the data, which we discuss below.

Firstly, if methylation rates and somatic mutation rates are independent, then our results may indicate that aberrant methylation levels are themselves a constraint on maximum lifespan. In other words, age-related changes to DNA methylation are deleterious and so mechanisms to

mitigate this are selected for in longer-lived organisms (i.e. to decrease p in **Eq. 2**). If true, the fact that we see similar scaling as observed in somatic mutation rates is intriguing, as it might imply a similar level of functional relevance for both phenomena, despite epimutations being widely considered to have a weaker direct impact on organism health. This scenario would support an instructive role of DNA methylation in the associations observed between epigenetic changes and physiological outcomes, in rejuvenating interventions and ageing[16–19].

Secondly, it is possible that both phenomena are driven by a constraint on a common underlying mechanism. As for what the common underlying mechanism may be, **Eq. 2** raises two possibilities. Given cell division plays a role in both processes, one possibility is that the scaling of both methylation and mutation rates is being driven by stem cell replication rates. Indeed, it has previously been suggested that the rate of haematopoietic stem cell divisions in mammals decreases with lifespan such that the total number of divisions per stem cell is approximately constant, regardless of lifespan[20]. However, as Cagan et al. note, there is recent evidence that replication rates may not be a major driver of mutation rates[21,22], potentially making this explanation less likely. Alternatively, both mutation and methylation rates may share a common genomic maintenance mechanism that increases in effectiveness with lifespan. However, given that the molecular processes behind methylation changes and mutations are quite distinct, it is not readily apparent what this common mechanism may be.

Although our study contains a relatively small number of datasets, the trend is consistent. This is especially striking considering that the datasets are from entirely distinct studies, comprise both array-based and reduced representation bisulfite sequencing (RRBS) methylation data, and that age and maximum lifespan estimates in species other than humans are very imprecise. When more data become publicly available in the future (e.g. the upcoming Mammalian Methylation Consortium data release), our findings will be able to be explored further.

Methods

Analysis

In contrast to previous studies[9,10], we restricted our analysis to CpGs which were age-related in each mammal being compared. We did this because even a conserved CpG site may behave markedly differently between species. For example, the ELOVL2 CpG (cg16867657) is the strongest age-related CpG in humans and chimpanzees (R^2=0.74 and 0.90, respectively), but shows no age-association at all in the other species analysed (**Ext. Data Fig. 2**). As such, using this CpG to calculate methylation rates across species may not be appropriate.

For each species, we conducted univariate linear regressions for each CpG of the form methylation ~ age. A CpG was considered age-associated if the adjusted R-squared value from this regression was greater than 0.1. The methylation rate of each CpG was defined as the slope from these regressions. In the case of datasets with multiple measurements for individuals, linear mixed-effects models were used with random intercepts for individuals and the marginal R^2 value was used.

Our initial approach was to find age-related CpGs common to all species and compare the average slope. However, very few CpGs satisfied this criterion, resulting in unstable estimates. Additionally, this method would extend poorly to additional animals as the number of common CpGs would decrease with each addition.

To overcome this issue, we instead used one species as a baseline. Zebra (*Equus quagga burchellii*) was chosen as the baseline species in blood and the big brown bat (*Eptesicus fuscus*) for skin samples as they a) had a relatively large sample size (n=96 and n=115, respectively), b) had an intermediate maximum lifespan, and c) methylation levels were assessed using the Mammalian Methylation Array[23], as were most of the other species, allowing a large crossover of common CpGs.

We then compared each species to these baseline species in a pairwise manner. For each comparison we:
1) Selected the set of conserved and age-related CpGs common to both species. Age-related methylation changes had to have the same direction in both species.
2) For each CpG in this set, calculated the absolute methylation rate for each species
3) Calculated the methylation rate ratio of the comparison species compared to the baseline species for each CpG. That is, for each CpG, *methylation rate ratio = absolute rate of comparison species/absolute rate of baseline species*. For example, a ratio of 2 would mean that the methylation rate of the species in question is twice as fast as the methylation rate of the baseline species in a particular CpG.
4) Calculated the median and interquartile range (IQR) of the resulting values across all CpGs included in the analysis. Use of the median was chosen over the mean as the mean was severely affected by large outliers.

This method ensured that the number of CpGs analysed in each comparison was large (usually hundreds to a few thousand) and could in theory be extended to any number of animals.

Scaling and power laws

Mathematically, scaling can generally be described as a power law relationship of the form:

$$y = ax^b$$

in which x and y are the variables of interest, and a and b are constants.

By taking the logarithm of both the x and y axes, power law relationships can be represented as a linear relationship in which the slope is equal to the original power:

$$\rightarrow log(y) = log(ax^b)$$
$$\rightarrow log(y) = log(a) + b * log(x)$$
$$\rightarrow Y = C + bX$$

in which $Y = log(y), C = log(a)$ and $X = log(x)$.

Data availability

All datasets used are publicly available. Pre-processed datasets were used without further adjustment, except in the case of the mouse (*Mus musculus*) dataset, which was processed as outlined in Simpson et al. 2023[24]. The yellow-bellied marmot (*Marmota flaviventris*) dataset is available on the Gene Expression Omnibus at GSE174544[25]; human (*Homo sapiens*) blood dataset at GSE40279[6]; mole rat (*Heterocephalus glaber*) dataset at GSE174777[26]; deer (Capreolus capreolus) dataset at GSE184216[27]; zebra (*Equus quagga burchellii*) dataset at GSE184223[28]; chimpanzee (*Pan troglodytes*) dataset at GSE136296[29], mouse (*Mus musculus*) dataset at GSE80672[30], bat (various species) dataset at GSE164127[10], beluga whale (*Delphinapterus leucas*) dataset at GSE164465[31], and human (*Homo sapiens*) skin dataset from the EWAS datahub Download page ("tissue_methylation_v1.zip" file)[32].

Data on maximum lifespan and mass were taken from the AnAge database[15] where possible, or taken from other estimates used in the original paper in the case of the bat dataset[10].

In the case of the bat dataset[10], only species with more than 15 samples were included in the analysis.

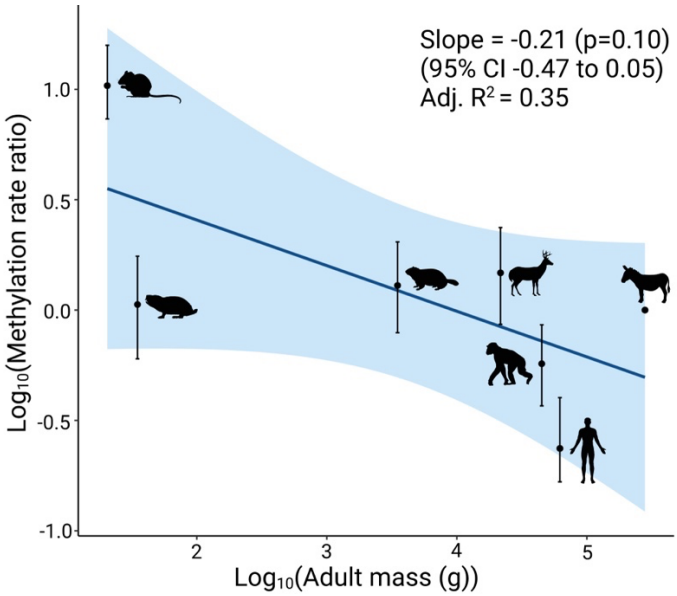
Code availability

Code used for the analysis (conducted in R version 4.2.1) is available at https://github.com/samuel-crofts/methylation_scaling.

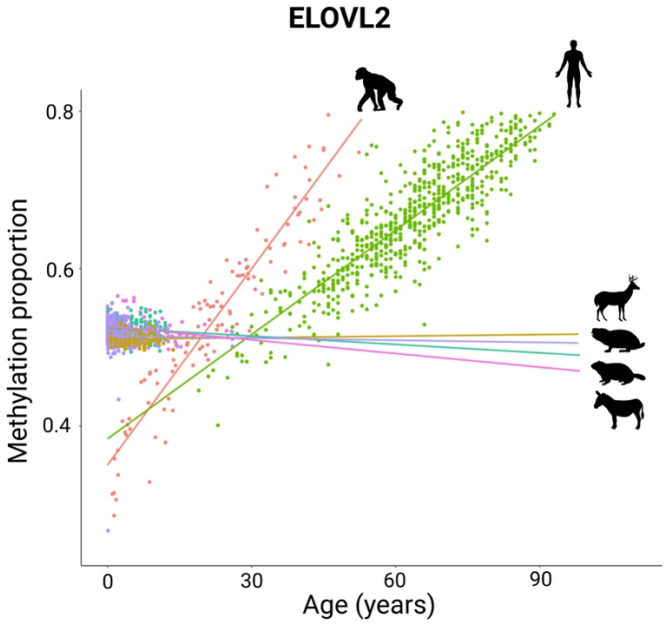
Acknowledgements

We thank Daniel Simpson for previously processing the mouse dataset. We thank Thomas Little for critical reading of the manuscript. We thank all members of the Chandra lab for their input. S.J.C.C is supported by the Wellcome Trust Hosts, Pathogens & Global Health Programme (grant number: grant.226831/Z/22/Z). E.L.C. is a cross-disciplinary postdoctoral fellow supported by funding from the University of Edinburgh and the Medical Research Council (MC_UU_00009/2). T.C. is supported through a Chancellor's Fellow at the University of Edinburgh and the MRC Human Genetics Unit.

Extended data



Extended Data Figure 1: Log(Methylation rate ratio) vs log(adult mass) in blood samples, using rates in the zebra as a baseline. Each point is the median value. Regression line from simple linear regressions of the form y~x. Shaded region represents the 95% confidence interval. Error bars show the interquartile range of values. From left to right, species are: house mouse (*Mus musculus*), naked mole rat (*Heterocephalus glaber*), yellow-bellied marmot (*Marmota flaviventris*), roe deer (*Capreolus capreolus*), chimpanzee (*Pan troglodytes*), human (*Homo sapiens*), plains zebra (*Equus quagga burchellii*).



Extended Data Figure 2: Example of a CpG (ELOVL2) that is strongly age-related in some species but age-invariant in others. Regression lines from simple linear regressions of the form methylation ~ age.

Correspondence

Eric Latorre-Crespo: eric.latorrecrespo@ed.ac.uk

Tamir Chandra: tamir.chandra@igmm.ed.ac.uk

Affiliations

MRC Human Genetics Unit, University of Edinburgh, Edinburgh, UK

Samuel. J. C. Crofts, Eric Latorre-Crespo & Tamir Chandra

School of Biological Sciences, University of Edinburgh, Edinburgh, UK

Samuel. J. C. Crofts

School of Informatics, University of Edinburgh, Edinburgh, UK

Eric Latorre-Crespo

Author Contributions

E.L.C. and T.C conceived and supervised the study. S.J.C.C, E.L.C and T.C. wrote the manuscript. S.J.C.C., E.L.C and T.C. conducted data analysis.

Competing Interest Statement

None applicable.

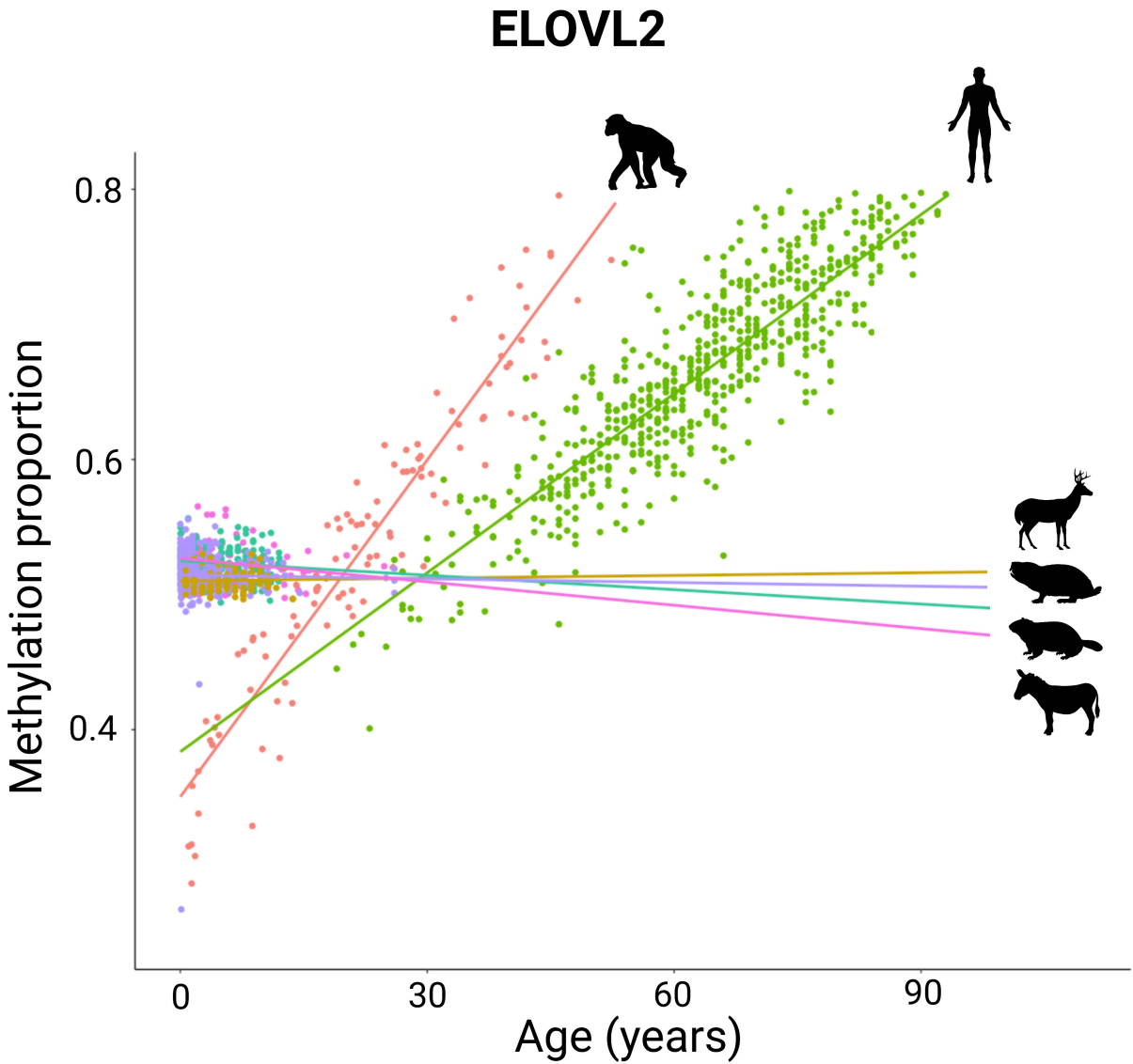
References

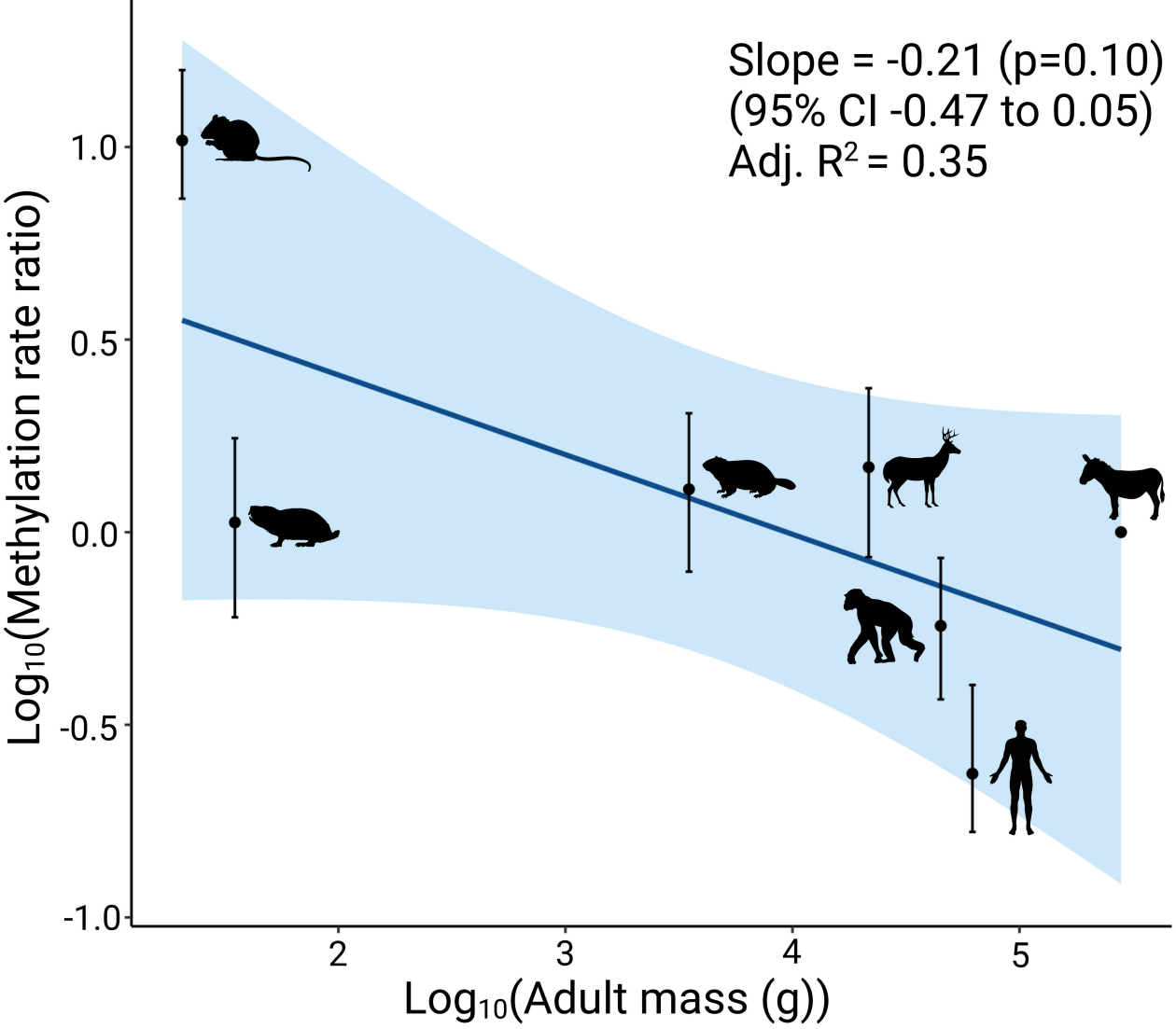
1. West, G. B. & Brown, J. H. The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *J. Exp. Biol.* **208**, 1575–1592 (2005).
2. Kleiber, M. Body size and metabolism. *Hilg.* **6**, 315–353 (1932).
3. West, G. B., Woodruff, W. H. & Brown, J. H. Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proc Natl Acad Sci USA* **99 Suppl 1**, 2473–2478 (2002).
4. West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).

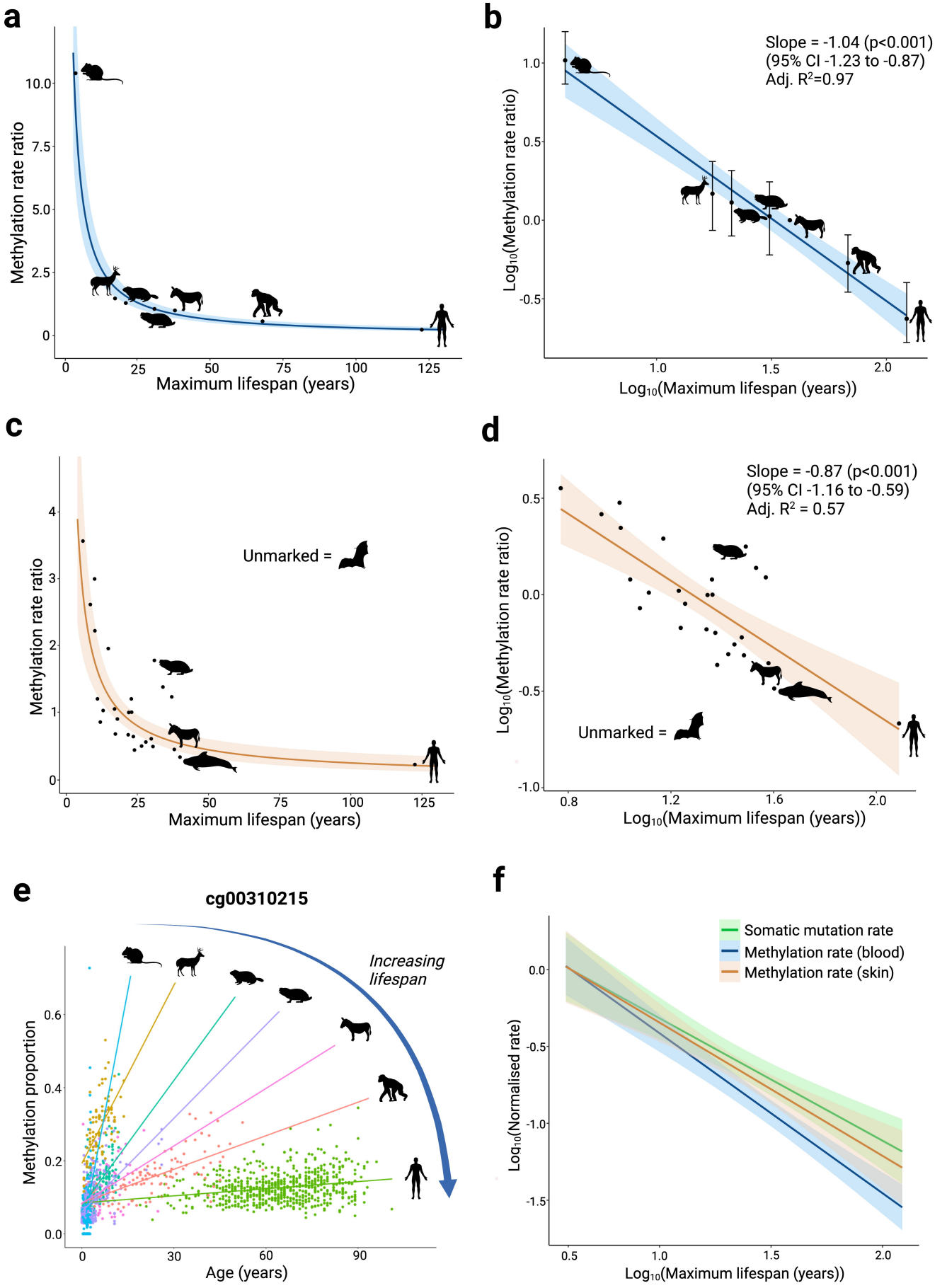
5. Bocklandt, S. *et al.* Epigenetic predictor of age. *PLoS ONE* **6**, e14821 (2011).
6. Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
7. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
8. Fei, Z., Raj, K., Horvath, S. & Lu, A. Universal DNA methylation age across mammalian tissues. *Innov. Aging* **5**, 412–412 (2021).
9. Lowe, R. *et al.* Ageing-associated DNA methylation dynamics are a molecular readout of lifespan variation among mammalian species. *Genome Biol.* **19**, 22 (2018).
10. Wilkinson, G. S. *et al.* DNA methylation predicts age and provides insight into exceptional longevity of bats. *Nat. Commun.* **12**, 1615 (2021).
11. Li, C. Z. *et al.* Epigenetic predictors of maximum lifespan and other life history traits in mammals. *BioRxiv* (2021) doi:10.1101/2021.05.16.444078.
12. Haghani, A. *et al.* Divergent age-related methylation patterns in long and short-lived mammals. *BioRxiv* (2022) doi:10.1101/2022.01.16.476530.
13. Cagan, A. *et al.* Somatic mutation rates scale with lifespan across mammals. *Nature* **604**, 517–524 (2022).
14. Seluanov, A. *et al.* Telomerase activity coevolves with body mass not lifespan. *Aging Cell* **6**, 45–52 (2007).
15. Tacutu, R. *et al.* Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res.* **46**, D1083–D1090 (2018).
16. Olova, N., Simpson, D. J., Marioni, R. E. & Chandra, T. Partial reprogramming induces a steady decline in epigenetic age before loss of somatic identity. *Aging Cell* **18**, e12877 (2019).
17. Ocampo, A. *et al.* In Vivo Amelioration of Age-Associated Hallmarks by Partial Reprogramming. *Cell* **167**, 1719-1733.e12 (2016).
18. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).

19. Dabrowski, Jan. K. *et al.* Probabilistic inference of epigenetic age acceleration from cellular dynamics. *BioRxiv* (2023) doi:10.1101/2023.03.01.530570.
20. Dingli, D., Traulsen, A. & Pacheco, J. M. Dynamics of haemopoiesis across mammals. *Proc. Biol. Sci.* **275**, 2389–2392 (2008).
21. Abascal, F. *et al.* Somatic mutation landscapes at single-molecule resolution. *Nature* **593**, 405–410 (2021).
22. Blokzijl, F. *et al.* Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260–264 (2016).
23. Arneson, A. *et al.* A mammalian methylation array for profiling methylation levels at conserved sequences. *Nat. Commun.* **13**, 783 (2022).
24. Simpson, D. J. *et al.* Region-Based Epigenetic Clock Design Improves RRBS-Based Age Prediction. *BioRxiv* (2023) doi:10.1101/2023.01.13.524017.
25. Pinho, G. M. *et al.* Hibernation slows epigenetic ageing in yellow-bellied marmots. *Nat. Ecol. Evol.* **6**, 418–426 (2022).
26. Horvath, S. *et al.* DNA methylation clocks tick in naked mole rats but queens age more slowly than nonbreeders. *Nat. Aging* **2**, 46–59 (2022).
27. Lemaître, J.-F. *et al.* DNA methylation as a tool to explore ageing in wild roe deer populations. *Mol. Ecol. Resour.* **22**, 1002–1015 (2022).
28. Horvath, S. *et al.* DNA methylation aging and transcriptomic studies in horses. *Nat. Commun.* **13**, 40 (2022).
29. Guevara, E. E. *et al.* Age-associated epigenetic change in chimpanzees and humans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190616 (2020).
30. Sziráki, A., Tyshkovskiy, A. & Gladyshev, V. N. Global remodeling of the mouse DNA methylome during aging and in response to calorie restriction. *Aging Cell* **17**, e12738 (2018).
31. Bors, E. K. *et al.* An epigenetic clock to estimate the age of living beluga whales. *Evol. Appl.* **14**, 1263–1273 (2021).
32. Xiong, Z. *et al.* EWAS Open Platform: integrated data, knowledge and toolkit for

epigenome-wide association study. *Nucleic Acids Res.* **50**, D1004–D1009 (2022).


ELOVL2
Methylation proportion
0.8
0.6
0.4
0
30
60
90
Age (years)


Slope = -0.21 (p=0.10)
(95% CI -0.47 to 0.05)
Adj. R^2 = 0.35
1.0
0.5
0.0
-0.5
-1.0
2
3
4
5
Log_{10}(Methylation rate ratio)
Log_{10}(Adult mass (g))


a
Methylation rate ratio
Maximum lifespan (years)
b
Slope = -1.04 (p<0.001)
(95% CI -1.23 to -0.87)
Adj. R²=0.97
Log10(Methylation rate ratio)
Log10(Maximum lifespan (years))
c
Unmarked =
Methylation rate ratio
Maximum lifespan (years)
d
Slope = -0.87 (p<0.001)
(95% CI -1.16 to -0.59)
Adj. R² = 0.57
Unmarked =
Log10(Methylation rate ratio)
Log10(Maximum lifespan (years))
e
cg00310215
Increasing lifespan
Methylation proportion
Age (years)
f
Somatic mutation rate
Methylation rate (blood)
Methylation rate (skin)
Log10(Normalised rate)
Log10(Maximum lifespan (years))