A universal limit for mammalian lifespan revealed by epigenetic entropy

Juan José Alba-Linares[1,2,3,4], Juan Ramón Tejedor[1,2,3,4], Agustín F. Fernández[1,2,3,4], Raúl F. Pérez[1,2,3,4,5,6,*], Mario F. Fraga[1,2,3,4,7,*]

[1]Cancer Epigenetics and Nanomedicine Laboratory, Centro de Investigación en Nanomateriales y Nanotecnología-Consejo Superior de Investigaciones Científicas (CINN-CSIC), Universidad de Oviedo, 33011 Oviedo, Spain.

[2]Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), 33011 Oviedo, Spain.

[3]Instituto Universitario de Oncología del Principado de Asturias (IUOPA), Universidad de Oviedo, 33003 Oviedo, Spain.

[4]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III (ISCIII), 28029 Madrid, Spain.

[5]Fundació de Recerca Clínic Barcelona - Institut d'Investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), 08036 Barcelona, Spain.

[6]Departamento de Bioquímica y Biología Molecular, Facultad de Veterinaria, Universidad Complutense de Madrid, 28040 Madrid, Spain.

[7]Departamento de Biología de Organismos y Sistemas, Área de Fisiología Vegetal, Universidad de Oviedo, 33006 Oviedo, Spain.

*Correspondence to: Raúl F. Pérez (raufer14@ucm.es), Mario F. Fraga (mffraga@cinn.es)

ABSTRACT

Age-associated DNA methylation patterns have shown strong associations with species lifespan. However, it remains unclear whether epigenetic noise levels can account for the observed differences between mammalian species. In this study, we examined the rate of loss of epigenetic information with age by measuring entropy at mammalian conserved CpG sites across a diverse range of species. Longer-lived mammals tend to gain fewer noisy CpGs with age, irrespective of whether these originate from hyper- or hypomethylation processes. Importantly, we found that the rate of epigenetic entropy gain declines in a linear fashion with species maximum lifespan, pointing to the existence of a universal limit for mammalian lifespan in the vicinity of 220 years.

MAIN

The species that comprise the tree of life encompass a vast array of organisms that are distinguished by unique life traits, covering a dynamic range of biological features such as body mass, metabolic rate, and developmental and reproductive patterns. Nevertheless, despite the fact that maximum lifespan is a precise, organism-specific life trait [1], the molecular mechanisms that underpin differences across species remain unidentified.

In this regard, a plethora of theories have been proposed to elucidate the causal nature of aging and lifespan. However, classical models have failed to provide a comprehensive explanation of this universal phenomenon, as they tend to focus on specific facets of aging, rather than offering a holistic perspective [2]. Among the various documented molecular mechanisms that regulate lifespan [3], there are stochastic (enzymatic errors, reactive by-products) and deterministic (genetic programs, environmental conditions) components that ultimately contribute to the accumulation of molecular damage and disorder. In this sense, Cagan and colleagues have revealed that loss of genetic information, in terms of somatic mutation rate per year, scales with lifespan in mammals [4]. Similarly, recent research has posited that the accumulation of epigenetic noise may represent the underlying cause of aging. For instance, loss of cellular identity is a well-known hallmark of aged tissues and appears to be driven by a convergent transcriptional activation of tissue-specific genes lacking CpG islands as a consequence of heterochromatin erosion with age [5,6]. In this line, loss of epigenetic information due to the repair of double-strand breaks is sufficient to trigger aging *in vivo*, including cellular dedifferentiation and acceleration of DNA methylation clocks, but can also be reversed by cellular reprogramming using Yamanaka factors [7]. Epigenetic noise is typically measured through various metrics, of which entropy (see Methods) is one of the most commonly used.

The preceding evidence led us to hypothesize that epigenetic noise gain with age is a species-specific characteristic, and thus that a lower noise gain corresponds to a longer maximum lifespan. Previous research by Crofts and colleagues revealed that the rate of DNA methylation (compared to baseline species) follows an exponential decay law with maximum lifespan in mammals [8]. These interesting findings raise several questions that remain unanswered: (1) does this DNA methylation rate actually link to the loss of epigenetic information?; (2) can DNA methylation loss rate scale with species maximum lifespan through an epigenetic entropy gain?; (3) if so, can entropy patterns estimate the global limit of maximum lifespan that any mammalian species can reach? Accordingly, a recent study conducted by Bertucci-Richter and Parrot provided an initial approximation, demonstrating that the rate of regional methylation disorder in PRC2-target genes shared across four mammalian species exhibited a deceleration with longer lifespan [9]. Similarly, Horvath and

colleagues noted that DNA methylation rate scales with age primarily at PRC2-bound bivalent promoters [10], which tend to gain methylation with age [11]. Nevertheless, further investigation is still required to ascertain the extent to which global DNA methylation loss mechanisms also contribute to differences in epigenetic noise gain with age, while expanding the number of mammalian species analyzed. This will enable a definitive conclusion to be reached regarding the scaling laws between the rate of DNA methylation noise gain and the limits in species maximum lifespan.

To address the aforementioned issues, we conducted a comprehensive investigation into epigenetic entropy gain rates in conserved CpG sites from Horvath's mammalian array [12–14] across 18 species that met the necessary criteria for rigorous statistical analysis (see Methods). The recent discovery of tissue-to-tissue disparities in the prediction of species lifespan via novel epigenetic clocks has led to the identification of blood as the optimal tissue for consistently yielding the highest lifespan estimations across species [15]. Accordingly, our approach entailed the examination of blood samples from diverse taxonomic orders, encompassing the majority of maximum lifespan observations documented in mammals.

Initially, we aimed to investigate whether epigenetic entropy gain with age showed any correlation with maximum lifespan across species. Strikingly, we demonstrated that the number of conserved CpG sites that accumulate epigenetic noise with age (FDR<0.05, logFC>0) is inversely proportional to species maximum lifespan (Fig. 1A: $R = -0.5530$, $P < 0.05$). To circumvent any potential bias associated with the unequal numbers of samples per species, which modifies the statistical power to detect significant hits, we extracted the residuals from regressing the number of entropy-gaining CpG sites and species sample size, confirming the observed negative trend while adjusting for the confounding factor (Fig. S1, see Methods). Additionally, we further segregated entropy-gaining CpGs according to the direction of average DNA methylation change. In this way, we found that the accumulation of epigenetic noise mediated by both DNA methylation gains and losses is inversely associated with lifespan (Fig. 1B, hyper: $R = -0.7808$, $P < 0.001$; Fig. 1C, hypo: $R = -0.4183$, $P = 0.08$). It is our contention that the strength of the anticorrelation may be greater for DNA hypermethylation than for hypomethylation because it involves fewer CpGs in the majority of species (Fig. 1D).

The results presented herein demonstrate that the conserved CpG sites that gain epigenetic entropy with age, irrespective of the direction of methylation change, contain information able to trace maximum lifespan trajectories in mammals. Consequently, we explored pan-species patterns of entropy gain in Horvath's mammalian array using multivariate linear models that incorporated species and sex covariates. Thus, we defined DMEPs as those CpG sites that gain epigenetic

entropy as a function of age/lifespan ratio (FDR<0.05, logFC>0, Methods), an approach that allows for the simultaneous analysis of the 18 selected species in our generalized model. We subsequently discovered that the overwhelming majority of conserved CpG sites (21,902) were considered DMEPs, exceeding the sum of pan-species CpG sites that lost epigenetic entropy (7,469) or remained stable (8,117) with age/lifespan ratio (Fig. 2A). Of these DMEPs, it should be noted that a higher proportion (64.3%) originated from DNA hypomethylation processes.

Importantly, when we compared the rate of epigenetic entropy gain at these DMEPs across species, we found an inversely proportional, significant, and robust linear relationship with respect to mammals' maximum lifespan (Fig. 2B: $R = -0.6802$, $P < 0.01$). This negative correlation was also evident using specifically hypermethylated DMEPs (Fig. 2C: $R = -0.7523$, $P < 0.001$) and hypomethylated DMEPs (Fig. 2D: $R = -0.5078$, $P < 0.05$). This result is noteworthy for several reasons. First, it provides evidence that the rate of epigenetic entropy gain is a genuinely informative measure of maximum lifespan, operating independently of the deterministic direction of methylation changes. Second, it shows that the accumulation of noise can be a more direct predictor of species lifespan than other proposed epigenetic indicators such as the rate of DNA methylation [8], a finding which is more biologically consistent with the information theory of aging [16]. Third, the discovery of a negative linear scaling between noise and lifespan, which differs from the exponential law previously described [8], suggests the existence of epigenetically-based evolutionary constraints that limit the maximum lifespan of any putative mammalian species.

To explore this concept, we extrapolated our lineal model to the null entropy rate by taking the cut-off point (X-axis) of the regression line between the entropy gain rate and maximum lifespan (Fig. 2E). This yielded a value of 118 years, which aligns closely with the longest document lifespan in humans (122.5 years) [1] , the species with the maximum value for which blood data were available in the Mammalian Methylation Consortium dataset. Similarly, by extrapolating the error of the model, we determined that the range of maxima considered by our regression oscillates approximately between 118 and 220 years, with the latter representing the maximum estimated limit of mammalian lifespan. This prediction is also consistent with current knowledge, whereby the bowhead whale is the longest-living mammal, with a lifespan of up to 211 years [1]. Therefore, our evidence points to the existence of a potential pan-mammalian limit to maximum lifespan, which can be tracked by studying the species-specific rate of epigenetic noise accumulation over lifetime.

Finally, we also investigated whether the epigenetic entropy gain rate could recapitulate documented sex differences in lifespan. On average, female median lifespan is longer than that of their male counterparts, but previous studies have found no consistent differences in aging rates [17]. To explore this question, we defined epigenetic entropy rate acceleration as the difference

between entropy rate in males and females, divided by the rate in females (Fig. S2). Out of 17 species tested, 10 showed positive acceleration values, but only 5 species (human, vervet, rat, beluga whale, and Asian elephant) exhibited values higher than +40% threshold, selected to ensure reliable conclusions. Conversely, out of the 7 species with negative acceleration values, only two of them (mouse and killer whale) were below the limit of -40%. Consequently, it can be reasonably proposed that epigenetic entropy gain rates may offer a potential mechanism to explain the divergence in lifespan between sexes. However, further research is still required to provide conclusive predictors of this phenomenon.

Our study develops an independent statistical framework that analyzes Horvath's mammalian array in a global, unbiased manner, considering the same CpGs in all samples without performing any prior selection based on intersections or pairwise comparisons by species. Nevertheless, our comprehensive approach has a number of considerations to take into account. First, certain changes in DNA methylation with age, and thus in epigenetic entropy, are due to changes in blood cell-type composition during aging. The absence of deconvolution strategies that can be applied to all the species studied currently prevents its correction. Nevertheless, given that the observed trends encompass the majority of conserved CpG sites in the mammalian array, it is very unlikely that this limitation influenced the robust conclusions reached in our study. Second, further research will be needed to elucidate the role of stem cell turnover in determining mammalian lifespan, considering that these divisions are the main factor driving the increase of epigenetic entropy with age in murine tissues [18].

All in all, our results provide compelling evidence that the rate of epigenetic noise accumulation exhibits an inversely proportional relationship with maximum lifespan in mammals, irrespective of the direction of methylation change. Importantly, this pan-species trend reveals a global limit for mammalian lifespan, with an approximate maximum of 220 years. Therefore, our findings give insights into how loss of epigenetic information explains differences in lifespan across species and offer a rationale for the existence of a universal lifespan limit in mammals derived from DNA methylation entropy.

METHODS

Data selection and statistical considerations

In agreement with previous studies [15], we selected blood data from the Mammalian Methylation Consortium [12–14] as the optimal tissue for consistently yielding the highest estimates of lifespan per species. In order to perform pan-species analyses, we also normalized the age of samples by the

maximum lifespan of their species (age/lifespan ratio). Subsequently, we excluded blood samples if they (1) belonged to species with fewer than 30 samples, to ensure an adequate statistical power to detect robust epigenetic entropy rates; (2) belonged to species with fewer than 5 samples with an age/lifespan ratio lower than 0.16 (equivalent to 19.60 years in humans); (3) belonged to species with fewer than 5 samples with an age/lifespan ratio higher than 0.5 (61.25 years in humans). With these filters, we aimed to select methylation data with similar ranges of age/lifespan ratios, thus reducing potential biases. After this process, 3,107 blood samples from 18 different species were considered for downstream analyses, covering a wide range of documented mammalian lifespans (from 3.8 to 122.5 years).

Probe-level differential analyses of entropy and methylation by mammalian species

First, we aimed to identify those evolutionary-conserved CpG sites from Horvath's mammalian array (with 37,488 probes) that gain noise with age in each species. To measure loss of epigenetic information and noise accumulation, we calculated probe-level Shannon entropy levels from normalized beta values (β) according to the following formula [19,20]:

$$ShannonEntropy_{CpG} = \frac{1}{log(0.5)} \cdot \left(\beta_{CpG} \cdot log(\beta_{CpG}) + (1-\beta_{CpG}) \cdot log(1-\beta_{CpG})\right) \qquad \textbf{(1)}$$

We then employed linear models built with the *limma* package (v.3.44.3) [21] to fit epigenetic entropy values as a function of age in years. Noisy CpGs per species were defined as those with FDR<0.05 and logFC>0. Furthermore, we performed differential methylation analyses to categorize noisy CpGs based on their average methylation change, by regressing their corresponding beta values against age. Accordingly, we classified them as hypermethylated (hyper: FDR<0.05, logFC>0), hypomethylated (hypo: FDR<0.05, logFC<0) or without a clear direction of methylation change (NS: FDR≥0.05). In this way, the different collections of CpGs that gain noise with age were linearly correlated with species maximum lifespan. To further validate the observed negative trends and remove biases resulting from different numbers of samples per species, which cause differences in statistical power, we computed the residuals from regressing the number of noisy CpG sites against the number of blood samples per species, which are independent of the latter, and we finally tested the linear associations between residuals and species lifespan.

Pan-species differential analyses of entropy and methylation

In contrast to previous studies that restricted their analyses to gene CpGs shared among species [9] or common sites that change with age in pairwise comparisons [8], we were interested in identifying epigenome-wide longevity signatures by defining pan-mammalian CpGs that gain noise with age/lifespan ratio. First, using the same collection of CpGs across species allows to directly

compare epigenetic entropy rates without the need to calculate other relative transformations. Second, the intersection of CpGs that gain noise with age yields a smaller subset as the number of species increases, limiting its applicability.

Consequently, we used the *limma* suite (v.3.44.3) to design multivariate linear regressions that modeled entropy levels by each CpG as a function of age/lifespan ratio, while accounting for species and sex as covariates. Our prior selection of 18 species encompassing analogous ranges of age/lifespan ratios was necessary to enable the application of generalized linear models and the derivation of reliable pan-species information. Therefore, we defined differentially methylated entropic positions (DMEPs) as those conserved CpG sites that gain noise with age/lifespan ratio (FDR<0.05, logFC>0) across mammalian species. Similarly, by using differential methylation analyses that regress beta values against age/lifespan ratio (with species and sex as covariates), DMEPs were further classified as hypermethylated (hyper: FDR<0.05, logFC>0), hypomethylated (hypo: FDR<0.05, logFC>0) or without a clear direction of average methylation change (NS: FDR≥0.05).

Entropy gain rate calculations

First, we calculated the accumulation of epigenetic noise occurring in DMEPs for each sample according to the following formula, where *N* is the number of CpG sites:

$$ShannonEntropy_N = \frac{1}{N \cdot log(0.5)} \cdot \sum_{i=1}^{N} \left(\beta_i \cdot log(\beta_i) + (1-\beta_i) \cdot log(1-\beta_i)\right) \qquad \textbf{(2)}$$

Subsequently, we obtained the epigenetic entropy gain rate per species by the slope of the regression of cumulative entropy levels versus the age of samples. We then investigated the linear correlations between entropy gain rates and species maximum lifespan, revealing a robust negative association between the two variables. Consequently, to estimate the universal lifespan limit that any mammalian species can reach, we extrapolated the errors of the prior regression line at its intersection with the X-axis, thereby obtaining a range of maxima within which the longest known mammal (bowhead whale) [1] is found.

Finally, we also explored the differences in epigenetic entropy gain rates between sexes, by using the slope of the regression of cumulative entropy levels versus the age of male or female samples. The entropy gain rate acceleration was then defined as the difference between male and female entropy rates divided by the female entropy rate. Acceleration values exceeding ±40% threshold were considered representative changes.

Visualization of results

All graphs were generated in R using the *ggplot2* package (v.3.3.2) [22].

REFERENCES

1. De Magalhães JP, Costa J. A database of vertebrate longevity records and their relation to other life-history traits. Journal of Evolutionary Biology. 2009;22:1770–4.

2. Gladyshev VN. Aging: progressive decline in fitness due to the rising deleteriome adjusted by genetic, environmental, and stochastic processes. Aging Cell. 2016;15:594–602.

3. Tian X, Seluanov A, Gorbunova V. Molecular Mechanisms Determining Lifespan in Short- and Long-Lived Species. Trends in Endocrinology & Metabolism. 2017;28:722–34.

4. Cagan A, Baez-Ortega A, Brzozowska N, Abascal F, Coorens THH, Sanders MA, et al. Somatic mutation rates scale with lifespan across mammals. Nature. 2022;604:517–24.

5. Lee J-Y, Davis I, Youth EHH, Kim J, Churchill G, Godwin J, et al. Misexpression of genes lacking CpG islands drives degenerative changes during aging. Science Advances. 2021;7:eabj9111.

6. Izgi H, Han D, Isildak U, Huang S, Kocabiyik E, Khaitovich P, et al. Inter-tissue convergence of gene expression during ageing suggests age-related loss of tissue and cellular identity. Benayoun BA, Cheah KSE, editors. eLife. 2022;11:e68048.

7. Yang J-H, Hayano M, Griffin PT, Amorim JA, Bonkowski MS, Apostolides JK, et al. Loss of epigenetic information as a cause of mammalian aging. Cell. 2023;186:305-326.e27.

8. Crofts SJC, Latorre-Crespo E, Chandra T. DNA methylation rates scale with maximum lifespan across mammals. Nat Aging. 2024;4:27–32.

9. Bertucci-Richter EM, Parrott BB. The rate of epigenetic drift scales with maximum lifespan across mammals. Nat Commun. 2023;14:7731.

10. Horvath S, Zhang J, Haghani A, Lu AT, Fei Z. Fundamental equations linking methylation dynamics to maximum lifespan in mammals [Internet]. 2023 [cited 2024 Aug 18]. Available from: http://biorxiv.org/lookup/doi/10.1101/2023.05.21.541643

11. Rakyan VK, Down TA, Maslau S, Andrew T, Yang T-P, Beyan H, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res. 2010;20:434–9.

12. Arneson A, Haghani A, Thompson MJ, Pellegrini M, Kwon SB, Vu H, et al. A mammalian methylation array for profiling methylation levels at conserved sequences. Nat Commun. 2022;13:783.

13. Haghani A, Li CZ, Robeck TR, Zhang J, Lu AT, Ablaeva J, et al. DNA methylation networks underlying mammalian traits. Science. 2023;381:eabq5693.

14. Lu AT, Fei Z, Haghani A, Robeck TR, Zoller JA, Li CZ, et al. Universal DNA methylation age across mammalian tissues. Nat Aging. 2023;3:1144–66.

15. Li CZ, Haghani A, Yan Q, Lu AT, Zhang J, Fei Z, et al. Epigenetic predictors of species maximum life span and other life-history traits in mammals. Science Advances. 2024;10:eadm7273.

16. Lu YR, Tian X, Sinclair DA. The Information Theory of Aging. Nat Aging. 2023;3:1486–99.

17. Lemaître J-F, Ronget V, Tidière M, Allainé D, Berger V, Cohas A, et al. Sex differences in adult lifespan and aging rates of mortality across wild mammals. Proceedings of the National Academy of Sciences. 2020;117:8546–53.

18. Vaidya H, Jeong HS, Keith K, Maegawa S, Calendo G, Madzo J, et al. DNA methylation entropy as a measure of stem cell replication and aging. Genome Biology. 2023;24:27.

19. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. Molecular Cell. 2013;49:359–67.

20. Shannon CE, Weaver W. Mathematical Theory of Communication. Urbana; 1963.

21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.

22. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. New York, NY: Springer; 2009 [cited 2023 Apr 7]. Available from: https://link.springer.com/10.1007/978-0-387-98141-3

FIGURES

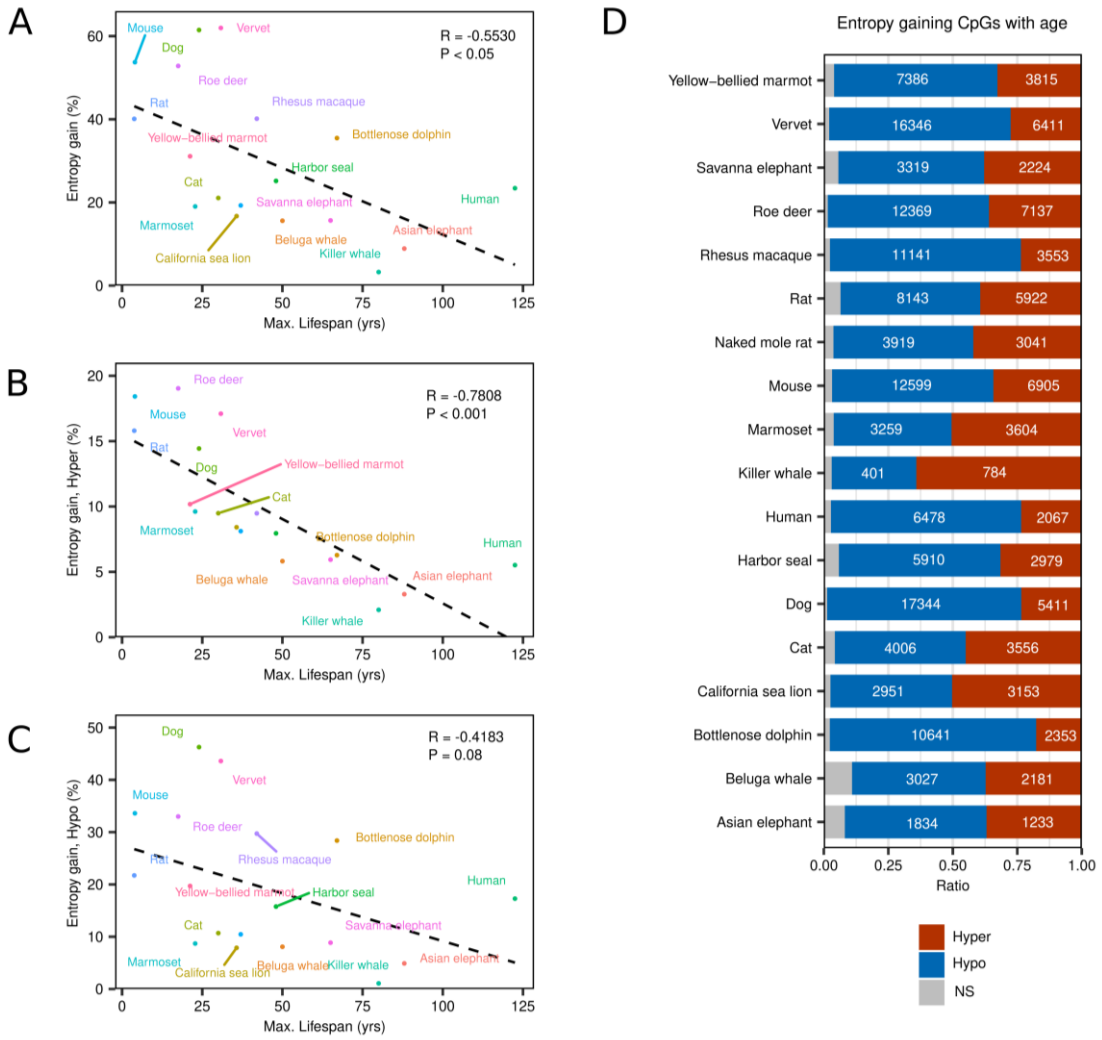

Figure 1. The number of age-associated noisy CpGs declines with species maximum lifespan. **(A-C)** Scatter plot showing the percentage of CpGs gaining epigenetic entropy with age (A), separated by hyper- (B) and hypomethylation (C) processes. **(D)** Barplots depicting the number and proportion of age-associated noisy CpGs by species, segregated by methylation gain and loss.

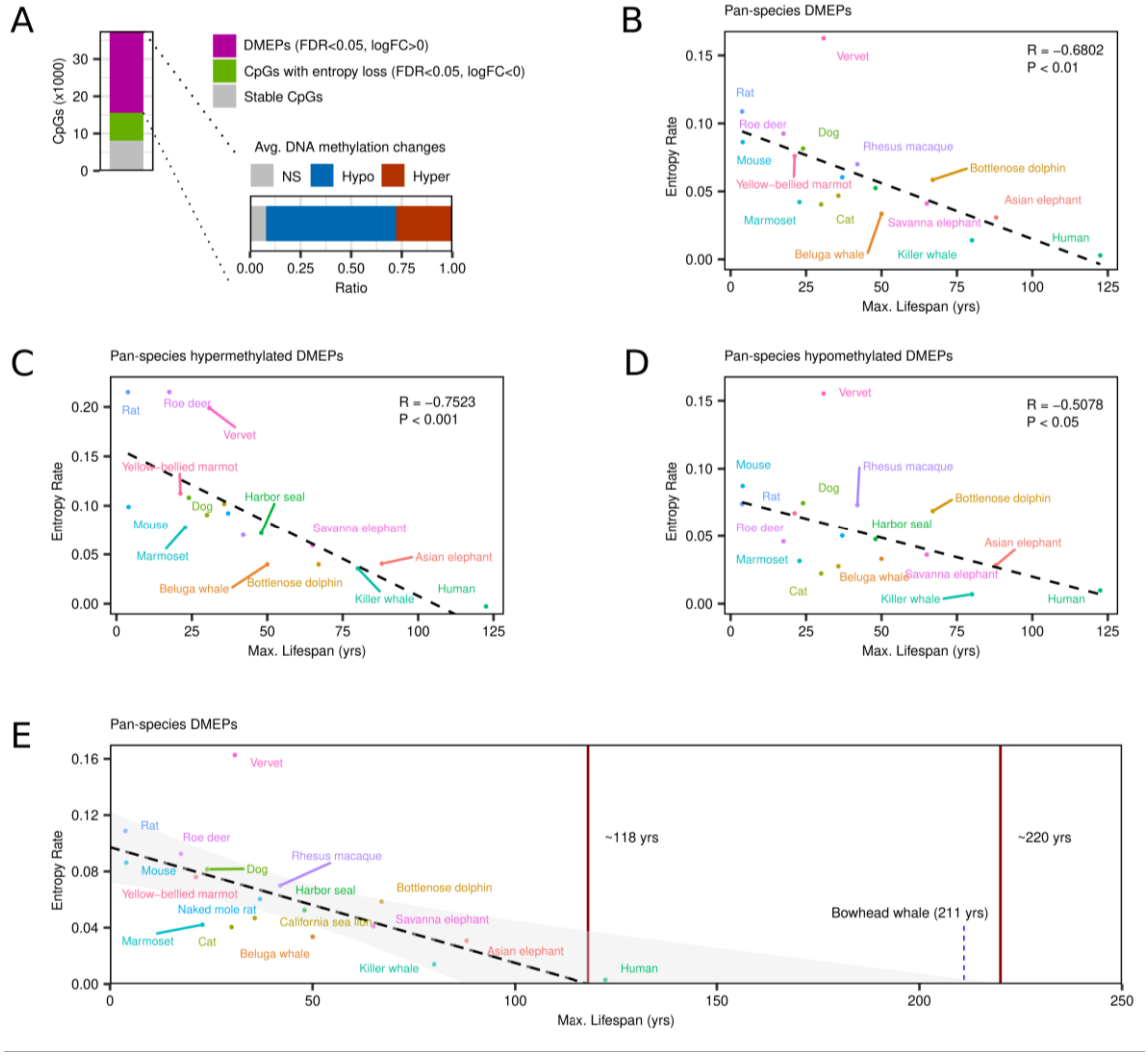

Figure 2. The rate of methylation entropy across conserved CpG sites tracks species maximum lifespan and evidences a universal limit for mammalian lifespan. **(A)** Barplot classifying mammalian array CpGs according to their contribution to entropy gain or loss across species. DMEPs are further grouped according to the direction of average methylation change. **(B-E)** Scatter plot showing the linear decrease in epigenetic entropy rate gain with age as a function of species maximum lifespan (B), distinguishing hyper- (C) and hypomethylation (D) processes. By including confidence intervals in regression (E), the epigenetic entropy rate reveals a range of limits on mammalian lifespan (118-220 years), which encompasses the longest-known mammal, the bowhead whale.

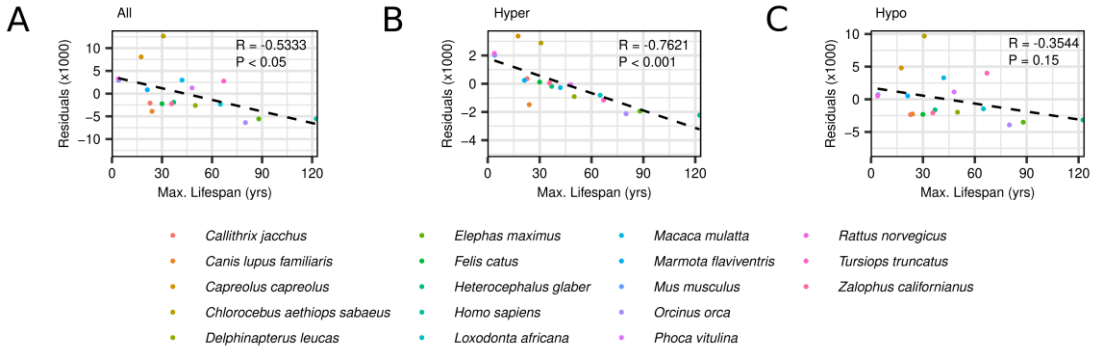

Figure S1. Scatter plots depicting the relationship between species maximum lifespan and residuals of the linear regression between the number of noisy CpGs with age and the number of samples by species (A), separating by hyper- (B) and hypomethylation (C).

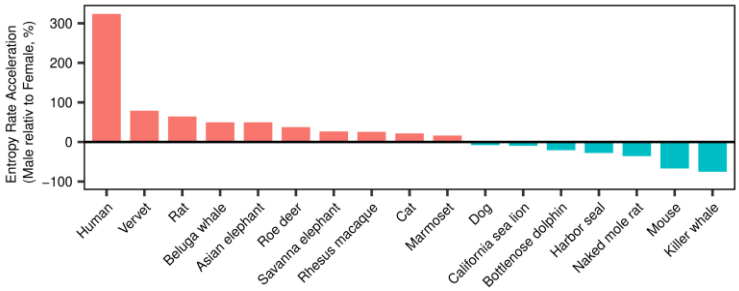

Figure S2. Barplots showing epigenetic entropy rate acceleration in males relative to females in different mammalian species.