Science
AAAS

Supplementary Materials for

DNA methylation networks underlying mammalian traits

Amin Haghani *et al*.

Corresponding authors: Amin Haghani, ahaghani@altoslabs.com; Steve Horvath, shorvath@mednet.ucla.edu

Science **381**, eabq5693 (2023)
DOI: 10.1126/science.abq5693

The PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S32
References

Other Supplementary Material for this manuscript includes the following:

MDAR Reproducibility Checklist
Tables S1 to S24
Data S1 to S19

Materials and Methods

Mammalian Data

The data included 15,456 samples derived from 70 tissue types of 348 different mammalian species (331 eutherians, 15 marsupials, 2 monotremes). We profiled 25 out of 26 mammalian taxonomic orders. These samples were collected from different age ranges of most of the species. Many of the species have been described in individual papers (*8-31*). The species-level characteristics such as maximum lifespan, average weight, and age at sexual maturity were chosen from anAge database (*47*). This data also includes DNA methylation from different dog breeds and mouse data from experimental lifespan intervention. Additional data sets included 57 horse transcriptome data generated from 29 different horse tissues (*27*). All DNA samples were analyzed by the custom-designed mammalian methylation array (*5*). Following data collection, the SeSaMe normalization method was used to define beta values for each probe (*48*). During network development, the data were divided into discovery (11,099 samples from 174 species) and validation (4,357 samples from 240 species) sets. The split was based on time of data generation. Skin biopsies, ear punches, and wing punches were all considered as skin in our analysis.

Selection of conserved CpGs in eutherians, marsupials, and monotremes

The high conserved CpGs were selected based on alignment of probes to 12 mammalian species from different taxonomic orders. These species included human (hg19), mouse (mm10), vervet monkey (ChlSab1.1.100), rhesus macaque (Mmul_10.100), cattle (ARS-UCD1.2), cat (Felis_catus_9.0.100), dog (CanFam3.1), African elephant (loxAfr3.100), bat (Rhinolophus_ferrumequinum.HLrhiFer5), killer whale (GCF_000331955.2_Oorc_1.1), opossum (Monodelphis_domestica.ASM229v1.100), and platypus (mOrnAna1.p.v1.100). These species were selected based on a large sample size in our data, a relatively high genome quality, and representation from different taxonomic orders.

Our phyloepigenetic tree analysis and our WGCNA analysis involved three sets of CpGs probes that were selected based on alignments to mammalian genomes. The first set consisted of probes (CpGs) that mapped (aligned uniquely) to the above mentioned eutherian species (human, mouse,vervet monkey, macaque, cat, dog, cattle, elephant, bat, killer whale), the second set was comprised of CpGs that map to opossum as a marsupial representative, and the third set consisted of CpGs that also map to platypus. These three sets were additionally filtered using calibration data generated from the array's performance on human, mouse, and rat synthetic DNA at different methylation levels (from 0–100% methylated) (*5*). Only the probes with a linear correlation ≥ 0.8 in all three species calibration data were kept as a mappable probe in mammalian species. The final number of remaining probes for the analysis were 14,705 in eutherians, 7,956 including marsupials, and 5,656 including monotremes.

Our EWAS of maximum lifespan involved a separate set of 28,318 CpGs that apply to mice and humans as detailed elsewhere.

Hierarchical clustering

DNA methylation data from tissues with more than 50 species were used for hierarchical clustering and comparison with the phylogenetic tree.

To define phyloepigenetic trees, we used hierarchical clustering of tissue samples (as opposed to CpGs) with average linkage. As input, the hierarchical tree method required a dissimilarity measure which was defined as 1 minus the Pearson correlation (**fig. S3** shows comparison of

different linkage methods). We aggregated all the animals from one species by forming mean values per tissue. The distances in the hierarchical trees (i.e., the height values) were directly compared with the evolutionary distances (based on estimated time) in a publicly available phylogenetic tree (*49*).

To construct phyloepigenetic trees, we used a correlation-based dissimilarity matrix (Diss=1 - rho where rho is the pairwise correlation coefficient between species). As pointed out by a reviewer, this approach can be justified on the following theoretical grounds. In a phylogeny of continuous traits, the variance for a trait in a given species is the sum of the branch lengths from the species to the root, and the covariance between any two species is the sum of the ancestral branch lengths common to the species (*50*). Thus, the tree defines the covariance matrix for a given trait across all species. After normalizing the tree to have height one, the covariances become correlations. In such a tree, the distances between any two species is 2 * (1 - rho) where rho is the pairwise correlation. The number 2 is a common scaling factor in the dissimilarity matrix and can be dropped. In other words, by estimating the correlation matrix and using the hierarchical clustering method on 1 - rho, one recovers the tree that generated the matrix (*50*).

We used picante::phylosignal() function (*51*) to calculate the phylogenetic signal of individual CpGs to traditional phylogenetic trees. Phylogenetic signal quantifies the extent to which a species' phylogeny predicts its ecological similarity. The K statistic measures phylogenetic signal by comparing the observed trait signal to the signal expected under a Brownian motion model of trait evolution on a phylogeny (*32*). A K value of 1 represents a Brownian motion process, suggesting some level of phylogenetic signal or conservatism. K values near zero indicate a random or convergent evolutionary pattern, while K values above 1 signify strong phylogenetic signal and trait conservatism. The statistical significance of phylogenetic signal can be assessed by comparing observed variance patterns of independent trait contrasts to a null model that shuffles taxa labels across the phylogeny's tips.

Unsupervised signed WGCNA

First, we formed two signed WGCNA networks based on the two sets of probes in our data. The first network was generated from 14,705 conserved CpGs in 10,927 samples of 167 eutherian species. The preservation of this network was evaluated in an independent dataset comprising 3,692 samples from 29 tissues of 228 mammalian species (164 new species, 64 overlapped with the training set). The independent data set was not used to update the weighted correlation networks.

The second network was a subset of 7,956 probes in 11,105 samples from 167 eutherian and nine marsupial species. Traditionally, WGCNA is applied for transcriptome data and uses an unsupervised clustering method to assign the co-expressed genes into modules (*33*). In this study, we used the signed WGCNA method to define the modules of co-methylated CpGs in mammalian samples. First, the adjacency matrix (defined as a power of 0.5+0.5*correlations between CpGs) was converted into a scale-free network using the soft threshold power (default power = 12) of the signed matrix. In this signed WGCNA, the network connection weight between two CpGs reflects a positive Pearson correlation. This means that negatively correlated CpGs have a negligible connection and will typically lie in separate modules.

The result was converted into a topological overlap matrix (TOM), and 1-TOM distance measure (dissimilarity), which was used for hierarchical clustering of the data. The trees were trimmed using a dynamic tree-cut algorithm to assign the modules containing at least 30 CpGs. Each module eigengenes (MEs) was defined as the singular vector (corresponding to the highest

singular value) from the singular value decomposition of the scaled CpGs that underlie the respective module. The eigengenes in the eutherian network (Net1) explained a range of 24–63% (average = 43%) of the variance in the methylation data in the training set, test set, and all data in each module (**table S3**). For a given module, we defined the measure of module membership, kME, as the Pearson correlation between the module eigengene and the CpGs. The association of module eigengenes were examined for different traits using multivariate linear regression models. The module colors in both networks were matched using the matchLabels() function in the WGCNA package.

Module preservation analysis

We used the training data to build several networks. As the study progressed, we generated additional data. We used these additional data as the test set to validate the existence of the modules. Toward this end, we considered two module preservation statistics: the cor.KME module preservation statistics, and the proportion of variance explained by the eigengenes.

Module preservation for each network was estimated using the "modulePreservation" R function in the WGCNA R package using primates as the reference for comparison (*34*).

Consensus Networks

All networks, including consensus networks, were developed in the discovery data. A total of seven consensus co-methylation networks were developed to effectively remove the confounding effects of species and/or tissue type. The constructed consensus networks are as follows: cNet3, 57 species-tissue strata (network where tissue/species effects were removed); cNet4, 35 species but ignore tissue strata (only species effects were removed); cNet5, 15 tissue types but ignore species (only tissue effects were removed); cNet6, 27 species blood (species effects were removed in blood); cNet7, seven species brain (species effects were removed in brain); cNet8, 10 species liver (species effects were removed in liver); and cNet9, 30 species skin data (species effects were removed in skin). Blood, skin, liver, and brain were the tissue types that had data from >7 mammalian species each.

Consensus WGCNA assumes that the DNA methylation network is conserved between multiple data strata. This network was generated following methods previously described (*33, 43*). Briefly, the adjacency matrices (correlation) were constructed using DNA methylation beta values in each data set. The matrices were converted into scale-free networks using a tuned soft threshold power for each dataset. Results were converted into TOM, merged, and then used to form a consensus tree network using a hierarchical clustering of dissimilarity matrix (1-TOM). Similarly, the colors were matched to network 1 colors.

Phylogenetic independent contrast method

We used the independent contrast method (*52*) to relate contrasts in module eigengenes (y-axis) to log transformed maximum lifespan using R function (pic from the R package ape, https://rdrr.io/cran/ape/man/pic.html). The independent contrast method combines closely related tree nodes to account for their high correlation. While the axes still correspond to the original variables, the original variables have been transformed to a different scale.

Dog methylation data

This data included n=756 blood samples (dogs) from 93 breeds (*29*). These samples were also used in the WGCNA analysis. Samples were provided by researchers at the National Human Genome Research Institute (NHGRI) and collection was approved by the Animal Care and Use Committee of the Intramural Program of NHGRI at the National Institutes of Health (Protocol 8329254).

Mortality analysis in human cohorts

To investigate if our modules are associated with mortality risks in humans, we analyzed Illumina 450k methylation arrays generated from blood samples of 4,651 individuals from the FHS offspring cohort (n=2,544 Caucasians, 54% females) and WHI cohort (n=2107 women). Each module eigengenes were calculated based on the overlap of CpGs between the 450k and Mammalian array. In FHS, the mean (SD) chronological age at the time of the blood draw was 66.3 (8.9) years old. During follow-up, 330 individuals died. The mean (SD) follow-up time (used for assessing time-to-death due to all-cause mortality) was 7.8 (1.7) years. The WHI cohort is a national study that enrolled postmenopausal women aged 50-79 years. Our WHI data consists of three ethnic/racial groups: 47% European ancestry (Caucasians), 32% African Americans, and 20% Hispanic ancestry. All the three ethnic groups have marginally the same age distribution, with mean (SD) of 65.4 (7.1) years old. The follow-up time has mean (SD) of 16.9 (4.6) years. During the follow-up, 765 women died. For each cohort and each module, we applied Cox regression analysis for time-to-death to assess the association with all-cause mortality. Our analysis was adjusted for age at blood draw and adjusted for gender as needed. To avoid confounding, we stratified the WHI cohort by ethnic/racial groups. Finally, for each module, we performed fixed effect models weighted by inverse variance to combine the 4 results across FHS and WHI sets into a single estimate. The meta-analysis was performed in R *metafor* function. We reported the meta-analysis P-value for the hazard ratio of mortality and a P-value for a test of heterogeneity Cochran Q test.

Growth hormone receptor KO studies

We analyzed liver samples from growth hormone receptor knock out (GHRKO) mice (n=11; 5 female, 6 male) and age matched C57BL/6 control mice (n=18; 9 male, 9 female). The age range of these animals was 6–8 months.

Light pollution experiment in mice

We analyzed murine liver samples. Mice in the control group, n=8, were exposed to standard light/dark cycles of 12 hours light (100 lux) followed by 12 hours dark. Mice in the circadian rhythm disrupted group, n=8, were exposed to 12 hours light (100), followed by 12 hours dim light (20 lux). Male C57BL/6 cohorts were exposed to these conditions from three months of age, for a period of 12 months, and sacrificed at age 1.17 years.

Caloric restriction and high fat diet in mice

Caloric restriction (CR) experiment included 56 liver samples from treated and 36 control C57BL/6 male mice. The mice were 18 months old at the time of sacrifice. High fat diet data (*12*) included liver samples: high fat diet n=133 (125 females, 8 males); control (*ad libitum*) n=212 (202 females, 10 males). Age range: 3–32 months.

Transient reprogramming in mice

The partial reprogramming experiment (*39*) was done by a systemic Yamanaka factors expression (Oct4, Sox2, Klf4, Myc) which was periodically induced by adding doxycycline to drinking water for 2 days per week in 4F het; R26::rtTAhet mice. The partial programming was done at three different durations: 1m 4F, n=3; 7m 4F, n=5; 10m 4F, n=3 samples per tissue type except for skin where only n=2 samples were available.The control group was C57BL/6 mice+dox, n=7.

EWAS of log maximum lifespan

We restricted the EWAS of life history traits to 28,318 CpGs that were shown to work in two species of great importance in biomedical research: mice and humans. Toward this end, we used calibration/titration data (correlation with calibration exceeds 0.8) and mappability information as described in (*5*).

Since the distribution of maximum lifespan and other life history traits was highly skewed, we imposed a log-transformation on these phenotypes before conducting EWAS. We carried out four types of analyses that differ by how they deal with two potential confounders: adult weight and phylogeny. Our "generic" EWAS corresponds to a marginal correlation analysis where the average methylation level of a given CpG per species was regressed on the (log-transformed) maximum lifespan using ordinary least squares regression. The second EWAS approach removed the confounding effect of average adult weight. To adjust for adult weight, we first regressed log maximum lifespan on log weight and formed residuals. Next the residuals become the dependent variables in the regression models. The third EWAS approach replaced ordinary least squares regression by phylogenetic regression, the variance-covariance matrix of which modeled evolutionary distances using branch lengths from the TimeTree project (*45, 53*). The fourth EWAS approach adjusted for both average weight and for phylogenetic relationships. Since phylogenetic regression takes into account sample covariance, it is more appropriate to report a phylogenetic independent contrast (PIC) as opposed to a simple scatter plot. Instead of using paired tip values from the tree, contrasts are calculated based on each node. The phylogenetic contrast model assumes that trait divergences occur independently at each node (*52*).

We carried out EWAS analyses in the following tissues/organs for which enough species (N>25 species) was available: skin (N = 137), blood (N = 133), liver (N = 147), skeletal muscle (N = 38), and brain (N = 26).

rankPvalue approach for CpGs Linked to Lifespan in Various Taxonomic Orders and Tissues

Using the non-parametric rankPvalue method and R function (*33*), we combined 25 EWAS of log transformed maximum lifespan results from various taxonomic order or tissue type strata, calculating the significance of a CpG's consistently high (or low) rank based on the 25 EWAS of log maximum lifespan (Data S13).

Here the function rankPvalue calculated the p-value for observing that CpG has a consistently high ranking (or low ranking) according to the EWAS of log maximum lifespan (corresponding to 25 columns of the input EWAS results).

The function uses the central limit theorem to calculate asymptotic p-values for test statistics that measure consistently high or low ordinal values (EWAS results). We used the percentile rank method which replaces each EWAS result by the ranked version. Low ranking and high rankings allow one to find consistently small values or consistently large values of the 25 EWAS results. Since the EWAS results are not independent, the rank p values should be interpreted as descriptive measures as opposed to inferential measures.

Significance thresholds

One of the main advantages of WGCNA is its focus on modules (e.g., 55 module eigengenes) as opposed to individual CpGs (e.g., 14,705 eutherian CpGs) which circumvents the multiple comparison problem (p= 0.001 suggested by the Bonferroni correction 0.05/55)

For our various EWAS of log transformed maximum lifespan, we adopted a nominal significance threshold of 1.8×10^{-6} (=0.01/28,318) based on a conservative Bonferroni adjustment. To list more top CpGs corresponding to relaxed thresholds, **table S12–S15** report the FDR adjustment associated with the P-values in different EWAS analysis. For generic EWAS, both the top 500 selection and the thresholding give these foreground inputs for all of our enrichment analyses: Lifespan negative/positive, 500/500 CpGs; Lifespan (AdjWeight) negative/positive, 500/500; Lifespan (AdjPhylo) negative/positive, 439/396; Lifespan (AdjPhyloWeight) hypo/hyper, 500/360.

Gene sets enrichment analysis

We performed several enrichment analyses for mammalian co-methylation modules. First, enrichment was performed at the level of genomic region using GREAT analysis (*54*) using all 14705 eutherian probes as the background. The analysis used human hg19 annotations, a 50kb window for extending the gene regulatory domain, and default settings for the other options. For each module, the input included up to 500 CpGs with highest module membership measures. The biological processes were subsumed under parent ontology terms using the "rrvgo" package (*55*). Second, the intra-module hub CpGs were also statistically tested for overlap with human GWAS results, in which gene p values were calculated by MAGENTA algorithm (*56*). Third, the genes next to the 500 CpGs were analyzed by Ingenuity Pathway Analysis (*40*) to identify potential upstream regulators.

Chromatin state analysis

We conducted chromatin state enrichment analyses using a universal chromatin state annotation of the human genome annotation that is not specific to one cell or tissue type. The annotation was based on a stacked ChromHMM model recently generated by integrating over 1000 epigenomic data sets from diverse human cell and tissue types (*41*). For each EWAS enrichment mentioned above, we utilized a hyper-geometric test to assess significant overlap between chromatin states and the two sets of CpGs that are significant in either positive or negative correlations with maximum lifespan. The background set for these hyper-geometric enrichment tests were the 28,318 CpGs that mapped to both human and mouse and were validated using calibration data (*5*).

Data from the Mammalian Methylation Consortium and description of Human Cohorts

Description of species

Primates (*57*)

Ethics

This research complied with all relevant ethical regulations overseen by seven ethics review boards. The human skin samples were acquired with informed consent prior to collection of human skin samples approved by the Oxford Research Ethics Committee in the UK; reference 10/H0605/1. Participants were not compensated. The secondary use of the other de-identified/coded human tissue samples (blood, postmortem tissues) is not interpreted as human subjects' research under U.S. Department of Health & Human Services 45 CFR 46. Therefore, the need to obtain written, informed consent from human study participants was waived (secondary use of de-identified tissues). Human samples were covered by University of California Los Angeles IRB#18-000315. All procedures related to non-human primates were approved by different committees: baboons (UTHSCSA Animal Care and Use Committee), strepsirrhini (Duke Institutional Animal Care and Use Committee and the DLC Research Committee), rhesus macaques (Animal Care and Use Committee of the NIA Intramural Program) (*57*), vervet monkey (UCLA and VA Institutional Animal Care and Use Committees) (*19*), marmosets (IACUC of UTHSA) (*18*).

Baboon care and maintenance

All animals were given a full veterinary examination prior to recruitment to the study and no obvious cause of ill health or pathology was observed. The animals were housed in group cages at the Southwest National Primate Research Center, at Texas Biomedical Research Institute (TBRI), in San Antonio, Texas in mixed sex groups of up to 16. The remaining 4 females were housed in individual cages at the UT Health Sciences Center San Antonio (UTHSCSA).

Twenty-eight females and the ten males were fed ad libitum Purina Monkey Diet 5038 (12% energy from fat, 0.29% from glucose and 0.32% from fructose and metabolizable energy content of 3.07 kcal/g protein; Purina LabDiets, St Louis, MO, USA) (CTR). Water was continuously available to all animals. Animal health status was recorded daily.

Necropsy

None of the animals were euthanized for this project. Rather, we used left-over frozen tissue samples that had previously been collected as part of other projects. Necropsies were performed by either a qualified, experienced veterinarian or M.D investigator. At TRBI, baboons were pre-medicated with ketamine hydrochloride (10 mg/kg IM) and anesthetized using isoflurane (2%) resulting in general anesthesia as previously described (*58*). Baboons were exsanguinated while under general anesthesia as approved by the American Veterinary Medical Association. At UTHSCSA four animals were euthanized using Pentobarbital at 390 mg/ml (Fatal- Plus Solution, Vortech, Dearborn, MI, USA) . Following cardiac asystole, respiratory failure and a lack of reflexive response to both skin pinch and eye touch stimulation, tissues (adipose, cerebellum, cerebral cortex, muscle, heart, liver) were rapidly dissected and immediately frozen in liquid nitrogen.

For the studies in which fetal tissue was obtained, all animals were housed in 20 foot × 20 foot × 15 foot metal and concrete group cages at the Texas Biomedical Research Institute. Experimental animals were obtained from appropriate groups of 16 healthy female baboons of similar pre-study body weights (10–15 kg) and morphometric features. The potential day of conception was determined based on the day of ovulation and changes in sex skin color and

pregnancy was confirmed at 30 days post ovulation by using ultrasonography. Details of housing, feeding, and environmental enrichment have been published elsewhere. All procedures were approved by the University of Texas Health Science Center and Texas Biomedical Research Institute internal animal care and use committees and performed in the Association for Assessment and Accreditation of Laboratory Animal Care–approved facilities.

Prior to Cesarean section, baboons were premedicated with ketamine hydrochloride (10 mg/kg, IM). Following tracheal intubation, isoflurane (2%, 2L/min, by inhalation) was used to maintain an appropriate plane of anesthesia throughout the surgery. A cesarean section was performed at gestational day 165 (0.9 of gestation) using standard sterile techniques. Following hysterotomy, the umbilical cord was identified and used for fetal exsanguination with both maternal and fetal baboons under general anesthesia as approved by the American Veterinary Medical Association Panel on Euthanasia. Postoperatively, mothers were placed in individual cages and watched until they were upright under their own power. Maternal analgesia was administered for 3 days (buprenorphine hydrochloride injection; Hospira, Inc., Lake Forest, IL, USA; 0.015 mg/kg/day) post-operatively or longer if indicated. They were returned to their group cage two weeks postoperatively.

Animals were individually fed to enable precise regulation of intake either between 7:00 am and 9:00 am or 11:00 am and 1:00 pm. Water was continuously available in each feeding cage (Lixit, Napa, California), and the animals were fed Purina Monkey Diet 5038 (Purina, St Louis, Missouri). For this study, we selected samples representing the entire primate lifespan, from neonate to old age.

Strepsirrhine primates

Strepsirrhini is a suborder of primates that includes the lemuriform primates, which consist of the lemurs of Madagascar, pottos and galagos from Africa, and the lorises from Southeast Asia. Lemuroids and lorisoids together form the more ancestral sister clade to all other living primates. As such, they lend unparalleled power to any comparative study within the primate clade.

For this study, we selected a total of 91 samples from individuals representing 26 strepsirrhine species, in most cases, the entire lifespan, from immature (infant or juvenile) to senile stages: 68 samples from peripheral blood, 23 samples from skin. The strepsirrhine primates (suborders Lemuriformes and Lorisiformes) used in this study were from the Duke Lemur Center (DLC) in Durham, NC (USA). The Duke Lemur Center is certified by both the Association for Assessment and Accreditation of Laboratory Animal Care and the American Zoological Association. The animal handling and sample collection procedures in this study were performed by a veterinarian after review and approval by the Duke Institutional Animal Care and Use Committee and the DLC Research Committee. Both housing and sample collection met or exceeded all standards of the Public Health Service's “Policy on the Humane Care and Use of Laboratory Animals”. The lemurs are housed in comparable social and housing conditions, habituated to human presence, and individually identifiable. The DLC also maintains a large collection of banked tissues, deriving from routine veterinary procedures and necropsies, amassed over the Center’s 55-year history. Detailed records of life and medical history, reproduction, and social-group membership are digitally maintained.

Peripheral blood was collected through venipuncture with standard procedures, either during a routine veterinary procedure or at time of necropsy. Skin tissues were collected during necropsies. Whole blood was preserved in either EDTA or Lithium Heparin, and stored at -80oC. Skin tissues were either frozen directly at -80oC or were first flash frozen and then stored at -80oC.

We profiled the following species: *Cheirogaleus medius* (Fat-tailed dwarf lemur), *Daubentonia madagascariensis* (Aye-aye), *Eulemur albifrons* (White-headed lemur), *Eulemur collaris* (Collared brown lemur), *Eulemur coronatus* (Crowned lemur), *Eulemur flavifrons* (Blue-eyed black lemur), *Eulemur fulvus* (Brown lemur), *Eulemur macaco* (Black lemur), *Eulemur mongoz* (Mongoose lemur), *Eulemur rubriventer* (Red-bellied lemur), *Eulemur rufus* (Red-fronted lemur), *Eulemur sanfordi* (Sanford's brown lemur), *Galago moholi* (South African galago), *Hapalemur griseus* (Bamboo lemur), *Lemur catta* (Ring-tailed lemur), *Loris tardigradus* (Slender loris), *Microcebus murinus* (Gray mouse lemur), *Mirza zaza* (Northern giant mouse lemur), *Nycticebus coucang* (Slow loris), *Otolemur crassicaudatus* (Greater galago), *Perodicticus potto* (Potto), *Propithecus diadema* (Diademed sifaka), *Propithecus tattersalli* (Golden-crowned sifaka), *Varecia rubra* (Red ruffed lemur).

Primates from Busch Gardens

The blood samples from chimpanzees (Pan troglodytes, n=2), gorillas (Gorilla, n=3)

Orangutan (n=1, Pongo pygmaeus), red ruffed lemur (n=1, Varecia variegata), and White-fronted marmoset (n=1, Callithrix geoffroyi) were opportunistically collected and banked during routine health exams from these zoo-based animals located at Busch Gardens Tampa (Tampa, Florida).

Human tissue samples

We analyzed previously generated methylation data from n=1352 human tissue samples (adipose, blood, bone marrow, dermis, epidermis, heart, keratinocytes, fibroblasts, kidney, liver, lung, lymph node, muscle, pituitary, skin, spleen) from individuals whose ages ranged from 0 to 101 years. Out of the 1352 tissues, n=655 came from women.

The tissue samples came from four sources: tissue and organ samples from the National NeuroAIDS Tissue Consortium (*59, 60*), Blood samples from the Cape Town Adolescent Antiretroviral Cohort study (*61*) and the PEG study (*62*), skin and other primary cells provided by Ken Raj (*63*). Ethics approval (IRB#18-000315). Human postmortem samples are from previously generated methylation data (adipose, blood, bone marrow, dermis, epidermis, heart, keratinocytes, fibroblasts, kidney, liver, lung, lymph node, muscle, pituitary, skin, spleen) from individuals whose ages ranged from 0 to 93. The tissue samples came from three sources. Tissue and organ samples from the National NeuroAIDS Tissue Consortium (*59, 60*). Blood samples from the Cape Town Adolescent Antiretroviral Cohort study (*61*). Additional blood samples from the PEG study. Skin and other primary cells provided by Kenneth Raj (*63*). Ethics approval (IRB#15-001454, IRB#16-000471, IRB#18-000315, IRB#16-002028).

Vervet monkeys (*19*)

All animals used in this study were Caribbean-origin vervet monkeys (*Chlorocebus sabaeus*) from the VRC at Wake Forest School of Medicine. The VRC colony is an extended multigenerational pedigree established from 57 founders imported from the islands of St. Kitts and Nevis in the West Indies. The introduction of new animals to the pedigree ended in the mid-1980s. The colony members are socially reared in extended family groups mimicking the natural social composition of vervet monkey troops in the wild. Group sizes range from 11 to 23 animals, with one or two intact adult males included in each group. Unfamiliar males are rotated into each group every 3–5 years. The pedigree structure is genetically confirmed. All colony-born vervets have known chronological age accurate to 1 day.

Beyond applications in aging studies, animals from VRC are used in a wide range of research in areas such as the efficacy and enhancement of vaccines for infectious diseases, e.g., influenza and dengue, investigations of diabetes, metabolic disease and obesity; and the development of novel non-invasive biomedical imaging methodologies.

Ethics statement: The Wake Forest School of Medicine facilities are certified by the Association for Assessment and Accreditation of Laboratory Animal Care. The animal handling and sample collection procedures in this study were performed by a veterinarian after review and approval by the UCLA and VA Institutional Animal Care and Use Committees. Both housing and sample collection were in compliance with the US National Research Council Committee's Guidelines for Care and Use of Laboratory Animals and met or exceeded all standards of the Public Health Service's "Policy on the Humane Care and Use of Laboratory Animals.

Vervet tissue samples

For this study, we selected a total of 240 samples representing the entire vervet lifespan, from neonatal to senile stages: 144 samples from the peripheral blood, 48 samples from the liver, and 48 samples from the cortical brain area BA10. The brains were perfused to remove blood prior to dissection. The targeted brain area BA10 was very small, and brain samples were dissected as bulk tissues, collecting, to the extent feasible without the benefit of microscopy, the full thickness of the cortex while avoiding the underlying white matter. One outlier blood sample (202943350003_R03C01 from animal 1992020) was excluded from analysis on the basis of the DNAm profile. The remaining 143 blood samples included 14 pairs of biological replicates collected from 14 individuals at two different time points 3.9–10.93 years apart. Peripheral blood was collected through venipuncture with standard procedures. Liver and brain cortical tissues were collected during necropsies.

Genomic DNA was isolated from blood and liver samples primarily through Puregene chemistry (Qiagen). DNA from the liver was extracted manually and that from the blood was extracted with an automated Autopure LS system (Qiagen). DNA was extracted from old liver tissues and clotted blood samples manually with a QIAamp DNA Blood Midi Kit and DNeasy Tissue Kit according to the manufacturer's protocol (Qiagen, Valencia, CA). DNA from BA10 was extracted on an automated nucleic acid extraction platform AnaPrep (Biochain) with a magnetic bead based extraction method and Tissue DNA Extraction Kit (AnaPrep).

Common marmosets (*18*)

Animal care and maintenance: All marmosets used in this research were housed at the Barshop Institute for Longevity and Aging Studies at UT Health San Antonio (UTHSA). The Institutional Animal Care and Use Committee (IACUC) of UTHSA is responsible for monitoring housing and animal conditions regularly to ensure all guidelines are met for the safety and health of the animals. This research was reviewed and approved by the UTHSA IACUC and experiments were conducted in compliance with the US Public Health Service's Policy on Humane Care and Use of Laboratory Animals and the Guide for the Care and Use of Laboratory Animals and adhered to the American Society of Primatologists (ASP) principles for the ethical treatment of non-human primates. Animals used in this study were based on age as well as a record of relatively good health as assessed by veterinary examination. Animals in this study were born either at UTHSCSA or transferred from the Southwest National Primate Research Center (SNPRC) in San Antonio TX. At UTHSA, animals were maintained using a modified specific pathogen-free barrier facility.

Each animal received three diet choices daily provided ad libitum: Harlan Teklad purified marmoset diet (TD99468), Mazuri Callitrichid gel diet (5MI5) and ZuPreem.

Blood draws and clinical blood counts/chemistry: All blood draws were taken during the morning hours of 08:00-11:00 from fed, non-anesthetized animals restrained in a custom assembly. Femoral vein blood collection (1.0-2.0 mL) was performed on each animal and blood was placed into PAXgene blood collection tubes (Qiagen). Whole blood was frozen and stored at -80˚ C for shipment to UCLA for methylation analysis.

Study samples: For these studies we selected samples representing the entire primate lifespan, from neonate to old age.

Genomic DNA was isolated from tissue samples mostly using Puregene chemistry (Qiagen). DNA from the liver was extracted manually and from blood using an automated Autopure LS system (Qiagen). From old liver tissues and clotted blood samples DNA was extracted manually using QIAamp DNA Blood Midi Kit and the DNeasy Tissue Kit according to manufacturer's protocol (Qiagen, Valencia, CA). DNA from BA10 was extracted on an automated nucleic acid extraction platform Anaprep (Biochain) using a magnetic bead-based extraction method and Tissue DNA Extraction Kit (AnaPrep).

Rhesus macaque (*17*)

In total, we analyzed N=281 rhesus macaque tissue samples from 8 different sources of DNA. The rhesus monkeys have been housed continuously at the NIH Animal Center, Poolesville, MD. The animal center is fully accredited by the American Association for Accreditation of Laboratory Animal Care, and all procedures were approved by the Animal Care and Use Committee of the NIA Intramural Program. Monkeys were of a heterogeneous genetic background, both Chinese and Indian origin.

Monkeys were housed individually in standard nonhuman primate caging on a 12h light/12h dark cycle, room temperature 78+/-2 degrees humidity at 60+/-20%. All monkeys had extensive visual, auditory, and olfactory but limited tactile contact with monkeys housed in the same room. Monkeys received 2 meals per day at estimated ad libitum levels throughout the study. Water was always available ad libitum. Monkeys were monitored minimally 3 times daily by trained animal care staff.

Sample Collection: Monkeys were fasted overnight, approximately 16-18 hours. Monkeys were anesthetized with either Ketamine, 7-10 mg/kg, IM or Telazol, 3-5 mg/kg, IM. Blood samples were obtained by venipuncture of the femoral vein using a vacutainer and EDTA tubes. Samples were immediately placed on dry ice and stored at -80 degrees. Skin samples were collected at the same time from an alcohol-wiped area of the back between the shoulder blades. Omental fat, kidney, liver, lung, skeletal muscle, and brain cortex were collected during necropsies scheduled for other study purposes. At that time, tissues were flash frozen in liquid nitrogen following collection and stored at -80 degrees. These tissues were selected for use based on having matching blood samples. None of the monkeys were sacrificed for this study.

Prairie voles (*64*)

Prairie vole colony: Male and female prairie voles (*Microtus ochrogaster*) were produced from laboratory-bred colonies at Cornell University, from breeding pairs that were offspring of wild caught animals captured in Champagne County, Illinois, USA. Voles were weaned and housed with littermates on postnatal day (PND) 21, and then housed with same-sex littermates after PND42-45. All animals received rodent chow (Laboratory Rodent Diet 5001, LabDiet, St.

Louis, MO, USA) and water *ad libitum* and were maintained under standard laboratory conditions (14L:10D cycle, lights on at 08:00, 20 ± 2 °C) in transparent polycarbonate cages (29 x 18 x 13 cm) lined with Sani-chip bedding and provided nesting material. All experimental procedures were conducted and approved by the Institutional Animal Care and Use Committee (IACUC) of Cornell University (2013-0102) and were in accordance with the guidelines set forth by the National Institutes of Health.

Prairie vole tissue sample collection: Ear, liver, and brain samples from the Cornell University prairie vole colony were collected from 48 male and female prairie voles at various life stages: neonatal (<1 month old), sub-adult (2-4 months old), mature adult (4-10 months old), and middle aged/old adult (>10 months old). The pair bonded male and female prairie voles used in our study cohabitated with their partners for several months and produced at least three generations of litters. Animals were euthanized via rapid decapitation, their tissues rapidly extracted and frozen on dry ice before being stored at -80C until further processing for genomic DNA extraction. Brains were coronally sectioned and brain regions from the pair bonding circuit (PBC) were micro-dissected and pooled for each animal. The PBC brain regions included the prefrontal cortex, nucleus accumbens, lateral septum, ventral pallidum, and medial amygdala, and ventral tegmental area. Genomic DNA was isolated and purified using the phenol-chloroform extraction and ethanol precipitation method. A total of 144 tissue samples were collected and processed for DNA methylation analysis. One animal was removed from the study due to a mismatch with the reported sex and our DNA methylation-based sex estimator.

Peromyscus (*15*)

Deer mice are maintained as outbred, genetically diverse closed colonies in the Peromyscus Genetic Stock center of the University of South Carolina. The study was approved by the Institutional Animal Care and Use Committee (IACUC) of the UofSC (protocol #: 2356-101506-042720) and were in accordance with the guidelines set forth by the National Institutes of Health. DNA was isolated from live animals by tail snips, or upon sacrifice from livers and brains by using the DNeasy DNA isolation kit (Qiagen).

Horses (*27*)

Ethics: This collection protocol was approved by the UC Davis Institutional Animal Care and Use Committee (Protocol#19037). All collection protocols were approved by the UC Davis Institutional Animal Care and Use Committee (Protocols #20751 and 21455, respectively).

We generated DNA methylation data from n=42 different horse tissues collected at necropsy. The tissue atlas was generated from two Thoroughbred mares as part of the FAANG initiative (*65*), with the following tissues profiled: adipose (gluteal), adrenal cortex, blood (PBMCs; only n=1 mare), cartilage (only n=1 mare), cecum, cerebellum (2 samples each from lateral hemisphere and vermis), frontal cortex, duodenum, fibroblast, heart (2 samples each from the right atrium, left atrium, right ventricle, left ventricle), hypothalamus, ileum, jejunum, keratinocyte, kidney (kidney cortex and medulla), lamina, larynx (i.e. cricoarytenoideus dorsalis muscle), liver, lung, mammary gland, mitral valve of the heart, skeletal muscle (gluteal muscle and longissimus muscle), occipital cortex, ovary, parietal cortex, pituitary, sacrocaudalis dorsalis muscle, skin, spinal cord (C1 and T8), spleen, suspensory ligament, temporal cortex, tendon (deep digital flexor tendon and superficial digital flexor tendon), uterus (*65*). These tissues were also used for RNAseq analysis.

Blood samples were collected via venipuncture into EDTA tubes from across 24 different horse breeds (buffy coat). Most of the samples were from the Thoroughbred (TB) (n=79) and

American Quarter Horse breeds (QH, n=62). For the following breeds, we had between one and six blood samples: Andalusian, Appaloosa, Arabian, Dutch Warmblood, Hanoverian, Holsteiner, Irish Sport Horse, Lipizzaner, Lusitano, mixed breed, Oldenburg, Paint or Paint cross, Percheron, Shire, Standardbred, Warmblood and Welsh Pony. The n=49 liver samples originated from necropsy collections of horses across 19 different breeds, with most of the liver samples from QHs (n=20).

Naked mole-rat (*24*)

The NMR tissue samples were provided by two different labs: 1) Vera Gorbunova and Andrei Seluanov from the University of Rochester and 2) Chris Faulkes from Queen Mary University of London.

Animals from the University of Rochester: All animal experiments were approved and performed in accordance with guidelines set up by the University of Rochester Committee on Animal Resources with protocol number 2009-054 (naked mole-rat). Naked mole-rats were from the University of Rochester colonies. All animals in the colonies are microchipped and their ages are recorded. All tissues except for skin biopsies and blood were obtained from frozen tissue collection at the University of Rochester from healthy animals that were euthanized for other studies. Skin biopsies (2 mm punch) were collected from the backs of the animals under local anesthesia. Blood samples were collected from the tails. The n=3 induced pluripotent stem cells from NMR were generated as described in (*66*). As a control set for the iPS study, we used n=3 fibroblasts samples from animals aged 1 and 2.

Genomic DNA was extracted using Qiagen DNeasy Blood and Tissue kit and quantified using Nanodrop and Qubit.als

Study Animals from Queen Mary University of London: Naked mole-rats were maintained in the Biological Services Unit at Queen Mary University of London in accordance with UK Government Animal Testing and Research Guidance. The tissues used in this study were obtained from post-mortem specimens from animals free from disease in compliance with national (Home Office) and institutional procedures and guidelines. Because sample collection was from post-mortem material, additional local ethical approval was not required for this study. Tissue samples were snap frozen in liquid nitrogen following dissection and transferred for storage at -80ºC.

Sheep (*13*)

Sheep DNA samples for this study were derived from two distinct tissues from two strains: ear tissue from New Zealand Merino, and blood from South Australian Merino.

Sheep ear: Ear tissue was obtained from females and both intact and castrated male Merino sheep during routine on-farm ear tagging procedures in Central Otago, New Zealand. As a small piece of tissue is removed during the ear tagging process that is usually discarded by the farmer, we were able to source tissue and record the year of birth without altering animal experience, in accordance with the New Zealand Animal Welfare Act (1999) and the National Animal Ethics Advisory Committee (NAEAC) Occasional Paper No 2. The exact date of birth for each sheep is unknown, however, this was estimated to be the 18th of October each year, according to the date at which rams were put out with ewes (May 10th of each year), a predicted mean latency until mating of 1Tab2 days, and the mean gestation period from a range of sheep breeds (149 days). Castration was performed by the farmer using the rubber ring method within approximately 5-50 days from birth as per conventional farming practice. Mass of yearlings was recorded by the farmer for both castrated and intact male sheep at 6.5 months of age, as a part of routine growth

assessment. In total, ear tissue from 138 female sheep aged 1 month to 9.1 years and 126 male sheep (63 intact, 63 castrates) aged 6 months to 5.8 years were collected and subjected to DNA extraction.

DNA was extracted from ear punch tissue using a Bio-On-Magnetic-Beads (BOMB) protocol which isolates DNA molecules using solid-phase reversible immobilisation (SPRI) beads. Approximately 3 mm punches of ear tissue were lysed in 200 µL TNES buffer (100 mM Tris, 25 mM NaCl, 10 mM EDTA, 10% w/v SDS), supplemented with 5 µL 20 mg/mL Proteinase K and 2 µL RNAse A and incubated overnight at 55 °C as per BOMB protocols. The remainder of the protocol was appropriately scaled to maximise DNA output while maintaining the necessary 2:3:4 ratio of beads:lysate:isopropanol. As such, 40 µL cell lysate, 80 µL 1.5X GITC (guanidinium thiocyanate), 40 µL TE-diluted Sera-Mag Magnetic SpeedBeads (GE Healthcare, GEHE45152105050250) and 80 µL isopropanol were combined. After allowing DNA to bind the SPRI beads, tubes were placed on a neodymium magnetic rack for ~5 minutes until the solution clarified and supernatant was removed. Beads were washed 1x with isopropanol and 2x with 70% ethanol, and then left to air dry on the magnetic rack. 25 µL of MilliQ H_2O was added to resuspend beads, and tubes were removed from the rack to allow DNA elution. Tubes were once again set onto the magnets, and the clarified solution (containing DNA) was collected.

DNA was quantified using the Quant-iT PicoGreen dsDNA assay kit (ThermoFisher Scientific, cat # P11496). 1 µL DNA sample was added to 14 µL TE diluted PicoGreen in MicroAmp optical 96-well plates (ThermoFisher Scientific, cat #N8010560) as per manufacturer directions, sealed, and placed into a QuantStudio qPCR machine for analysis. Samples with DNA content greater than the target quantity of 25 ng/µL were diluted with MilliQ.

Sheep blood: DNA methylation was analyzed in DNA extracted from the blood of 153 South Australian Merino sheep samples (80 transgenic Huntington's disease model sheep (OVT73 line) and 73 age-matched controls) aged from 2.9 to 7.0 years. All protocols involving OVT73 sheep were approved by the Primary Industries and Regions South Australia (PIRSA, Approval number 19/02) Animal Ethics Committee with oversight from the University of Auckland Animal Ethics Committee. The epigenetic age of the transgenic sheep carrying the *HTT* gene was not significantly different from controls (p=0.30, Mann-Whitney U test), therefore the data derived from these animals was subsequently treated as one dataset.

300 µL thawed blood samples were treated with 2 rounds of red cell lysis buffer (300 mM Sucrose, 5 mM $MgCl_2$, 10 mM Tris pH8, 1% Triton X-100) for 10 minutes on ice, 10 minute centrifugation at 1,800 RCF, and supernatant removed between each buffer treatment. The resulting cell pellet was incubated in a cell digestion buffer (2.4 mM EDTA, 75 mM NaCl, 0.5 % SDS) and Proteinase K (500 µg/ml) at 50 °C for two hours. Phenol:Chloroform:Isoamyl alcohol (PCI, 25:24:1; pH8) was added at equal volumes, mixed by inversion, and placed in the centrifuge for 5 minutes at 14,000 RPM at room temperature (repeated if necessary). The supernatant was collected and combined with 100% ethanol at 2x volume, allowing precipitation of DNA. Ethanol was removed and evaporated, and 50 µL TE buffer (pH8) was added to resuspend genomic DNA. DNA sample concentration was initially quantified using a nanodrop, followed by Qubit.

Pig (*21*)

Porcine samples: All animal procedures were approved by the University of Illinois and University of Wisconsin Institutional Animal Care and Use Committee, and all animals received humane care according to the criteria outlined in the Guide for the Care and Use of Laboratory Animals. Porcine whole blood samples (n=146) were collected from female Large White X

Landrace crossbred domestic pigs (n=84, age range 11 – 2,285 days) and Wisconsin Miniature Swine™ (n=60, age range 8 – 1,880 days) at the University of Wisconsin-Madison. Whole blood (n=16) and tissue samples (bladder, frontal cortex, kidney, liver, lung; n=16/tissue type) were collected from 16 Large White X Minnesota minipig crossbred pigs (n=9 female, n=8 male, age range 29 – 1,447 days) at the University of Illinois at Urbana-Champaign. All blood samples were collected in EDTA tubes, aliquoted, and flash frozen in liquid nitrogen within 10 minutes of collection. Tissue samples were collected and flash frozen within 10 minutes of euthanasia. All samples were stored at -80 until processing. Samples were shipped to the University of California, Los Angeles Technology Center for Genomics & Bioinformatics for DNA extraction and generation of DNA methylation data.

Odontocete species (*10*, *14*, *67*)

Ethics approval: The study was authorized by the management of each institution and was reviewed by their respective zoo research and animal use committees.

Study Animals: For model development, our study population included 293 animals from nine species of odontocetes, of which the majority were from four species including beluga (n = 66), Pacific white-sided dolphins (n = 17), killer whales (n = 37) and bottlenose dolphins (n = 137), housed at nine Association of Zoos and Aquarium (AZA), Alliance for Marine Mammal Parks and Aquariums or Japanese AZA accredited zoological institutions. Known (77.8%) or estimated (based on length at capture or rescue for stranded animals) birth dates were provided by each housing institution. In addition to zoo-based animals, we included 19 skin samples from free-ranging Norwegian killer whales to the training set because the ages of these animals could be estimated within sufficient accuracy (expected error less than 8%) on the basis of several lines of evidence including GLG counts (n = 3), whereby 1 GLG was assigned per year of age, length at necropsy (n = 1), juvenile at first identification (n = 2) or minimum age estimation based on length and maturity at first sighting (n = 13). The remaining 26 animals' ages could not be estimated within our accuracy parameters and skin samples from these animals were used to demonstrate model application for determining age and sex in wild animals.

Sample collection: Blood samples (0.5 ml min) were collected either voluntarily from the peripheral periarterial venous rete on the ventral tail fluke using an 18 to 22 gauge winged blood collection set or attached to a vacutainer collection system. Blood was collected by either the veterinary technician or veterinarian on staff and into BD Vacutainers (Becton Dickinson, Franklin Lakes, NJ) containing EDTA. Samples were inverted in the Vacutainer a minimum of 10 times and then frozen at -80°C until further testing.

Skin samples (~0.5 gm) were collected either under stimulus control or manual restraint using a sterile disposable dermal curette (Miltex, Integra Life Sciences Corp.,York, PA) from a location just posterolateral of the dorsal fin overlying the epaxial muscle. Prior to collection, a cold pack was placed on the site for several minutes prior to sampling to numb the sample site. Skin samples were placed into sterile cryovials (Nunc® Cryotubes, MilliporeSigma Corp., St. Louis, MO) and stored at -80^{o}C until shipment on dry ice. Skin samples from non-living animals were obtained from frozen (-80°C) specimens that had been previously collected and stored during standard necropsy procedures.

Skin samples were sectioned from previously collected killer whale biopsy samples of 45 unique individuals (photo-identified) collected in August and November 2017 and from April through July 2018 in northern Norway (Jourdain, 2017). The killer whales were biopsied using an ARTS darting system (Restech, Bodø, Norway) and 25 × 9 mm or 40 × 9 mm stainless steel tips

in 2017, and with an injection gun (Pneu-Dart Inc., Williamsport, PA) and 25 × 7 mm tips in 2018 as previously described (Jourdain, 2017). In addition, tissue samples were collected from six dead, stranded killer whales, and one other individual by-caught in a herring purse-seine, in northern Norway between 2015 and 2017 (E. Jourdain, unpublished data). Skin samples were collected from the region directly posterior to the dorsal fin and stored at -20°C until analysis.

Beluga whales (*67*)

Skin tissue samples were collected from carcasses of beluga whales that were beach-cast, stranded dead, or taken during subsistence hunting between 1992 and 2015 in Cook Inlet, Alaska, USA (NMFS Research Permit 932-1905-00/MA-009526 through the Marine Mammal Health and Stranding Response Program). Skin samples were preserved in a salt and dimethyl sulfoxide (DMSO) solution and archived at NOAA's Southwest Fisheries Science Center in La Jolla, California, USA. A total of 69 individuals were selected for the clock calibration dataset, and their chronological ages were estimated by counting tooth growth layer groups. The final calibration dataset included 67 individuals due to inconsistent molecular sex data. Teeth were analyzed by at least two readers using methods validated in Lockyer et al. (2007), and a consensus age provided by NOAA was used in this study. When individuals were represented by multiple teeth in the dataset, the oldest age estimate was used to mitigate error from tooth wear (*e.g.*, the count from the tooth with the greatest number of growth layer groups).

Samples of skin tissue from living CI beluga whales were collected with a biopsy dart in 2016, 2017, and 2018 (NMFS ESA/MMPA Permit #20465; McGuire et al., 2017). Biopsy samples were frozen in the field in liquid nitrogen and later subsampled at the NOAA Alaska Fisheries Science Center in Seattle, Washington, USA. Genomic DNA was extracted from tissue samples using a standard phenol-chloroform protocol. Extracted DNA was treated with RNAse A (1 μL of 1 mg/mL added to samples of 100 μL for 30 minutes at room temperature) and then purified and concentrated using a DNA Clean and Concentrator-5 Kit (Zymo Research Corp., USA). The concentration of genomic DNA was measured on a QUBIT 4 fluorometer (ThermoFisher Scientific, USA).

Killer whales and bowhead whales (*6*)

Animal Use and Ethics: For bowhead (*Balaena mysticetus*) subsistence hunts, indigenous hunters had the authorization to conduct hunts and collected samples on behalf of Fisheries and Oceans Canada. Bowhead whale biopsy samples were collected in 2019 under Fisheries and Oceans Canada (DFO) license to Fish for Scientific Purposes (LFSP) S-19/20-1007-NU and Animal Care approval (AUP) FWI-ACC-2019-14. Skin samples from eastern North Pacific killer whales (*Orcinus orca*) were collected as previously described (Ford et al. 2018b) under NMFS General Authorization No. 781–1725, and scientific research permits 781-1824-01, 16163, 532-1822-00, 532– 1822, 10045, 18786-03, 545-1488, 545-1761, and 15616.

Killer whale biopsy samples & DNA extraction: Killer whale populations in the eastern North Pacific are among the most intensively studied cetacean populations globally. The so-called 'resident' killer whale populations inhabiting the coastal waters from California to Alaska comprise individually identified whales that have been studied for over 40 years. These longitudinal studies and reliable identification of individual whales within populations through annual photographic census data has provided unique insight into population dynamics and demographics. The resolution provided by annual documentation of births, ages at physical

maturity (males) and age at first parturition (females) provides an unparalleled opportunity to validate epigenetic models of age from whale skin.

The killer whale validation dataset is rare in both the representation across age classes and the number of known-age samples from a wild population based on direct observations (*6*), providing an ideal training set for validating an epigenetic clock. Killer whale longevity is estimated to be 80 or 90 years and samples in the current dataset represent individual whales ranging from age 0 (neonate) to 79 yrs. Reflecting killer whale age-related mortality patterns, the number of individuals representing older age classes diminishes as expected with n = 11 whales estimated be > 50 yrs old based on size, physical and reproductive maturity at the time of first observation.

Ages of individual killer whales from the Southern Resident killer whale and Alaska Resident killer whale populations were determined based on the sex and size of the animal during the year that it was first documented. Whales born during the study (post-1974) were aged in reference to the year in which they were born. Ages for whales that were juveniles or adults when field observations began in the early 1970s were aged based on the year they reached physical maturity or, for females, the year they gave birth to their first viable offspring. Confidence estimates (0% - 100%) were assigned to each individual sampled whale included in the dataset reflecting the certainty around age estimates, frequency of encounters with the individual and age or state of physical maturity at the time of first identification. Genetic samples from 142 killer whales (supplementary info) were included in the dataset representing 118 different killer whales of age 0 yrs (neonate or fetus; 100% certainty) to age 75+ yrs (50% certainty).

Epidermal samples were collected from live killer whales using remote dart biopsy methods, and from dead stranded animals during routine post-mortem necropsy protocols. Identities of individual whales were recorded photographically whenever possible. Total genomic DNA was extracted from skin biopsies either using a silica-membrane kit following manufacturer's protocols (DNeasy Blood and Tissue kit, Qiagen, Valencia, CA), or following a standard proteinase K phenol/chloroform/isoamyl alcohol extraction protocol.

Bowhead whale biopsy samples & age estimates: Bowhead whales are very slow-growing and extremely long-lived baleen species that undergo periods with rapid growth (as a fetus), pauses in growth (ages 1 to about 6-7) and slow growth (age 8+) which gradually slows even more as they age. For this reason, accurate age estimates of individual bowhead whales are typically limited to early life stages when multiple sources of information are available.

Estimating age for individual bowhead whales requires a combination of several different types of data including body length, length of the longest baleen plate, body condition measurements, other morphological measurements that include frequency of scars, colour of the peduncle and chin region, and aspartic acid racemization (AAR) analyses from eye lenses. Despite multiple data sources, bowhead age can only be approximated for whales more than approximately 7 years old because the variation in size at age is many times larger than the annual growth rate of individual whales. Whales 2 - 7 years old cannot be accurately aged without information on the longest baleen plate, which is only available for harvested whales, although body condition provides information to assist with age estimation.

Bowhead whale skin samples were collected from remote biopsy using a crossbow during photographic studies conducted in Cumberland Sound, and samples collected during subsistence harvest. Data available for age estimates were limited to body length, body condition, frequency of scars, and colour of the peduncle region for biopsy samples. Skin samples collected during subsistence harvest included age estimates based on body length, length of the longest baleen plate,

notes on scars or peduncle colour and AAR ages. Where multiple sources of data were available, ages obtained from AAR or age-at-length estimates were adjusted to take account of the additional data as below.

Age at length was estimated using data in Figure 7 of Lubetkin et al. (2012) (*68*) with estimation for gradually reduced growth rates for whales older than 60 years old. Yearlings were confirmed among the smallest whales by body condition using criteria in Koski et al. (2010) (*69*). Body condition was also used to adjust the age of whales 2 - 7 years as determined by length because the youngest of these whales (i.e., 2 years old) appeared to have the poorest body condition and body condition appeared to improve as young whales age (see Figure 3 in (*69*)).

Age based on aspartic acid racemization was determined for 11 harvested bowhead whales (4 females; 7 males). Both eyeballs were stored at –20°C immediately after dissection from harvested whales, and one eyeball per whale was subsequently used for age estimation. Dissection of eye lenses and age estimation by the aspartic acid racemization (AAR) technique was performed using methods described in (*70*). Estimates of individual D/L ratios were converted to age estimates as described in (*71*).

Humpback whales

Ethics: Skin samples were collected by the Center for Coastal Studies under research permits issued by the U.S., National Marine Fisheries Service (21485, 16325, 20465, 14245, 633-1483, 633-1778, 932-1905), the Canadian Department of Fisheries and Oceans and IACUC #NWAK-18-02.

Humpback sample collection: Skin samples were collected from live North Atlantic humpback whales by biopsy sampling techniques (Palsbøll, 1991). Sampling was performed in the Gulf of Maine between 2003 and 2020 under the authorization of nationally-issued research permits. Samples were refrigerated or frozen in the field and then archived at -80C without chemical preservative.

Year of birth can be determined precisely for humpback whales first encountered as calves during an obligatory period of maternal dependency (Clapham, 1992;Baraff). Otherwise, there are no reliable outward indicators of chronological age in this species. We therefore selected known-age samples for this study based on data from a long-term study of individual whales. The identity of each sampled individual was confirmed through photo-identification techniques (Katona, 1981) using a reference catalog of the Gulf of Maine population curated by the Center for Coastal Studies (Provincetown, MA). For individuals first seen as calves, age at sampling was the number of years between sampling and birth. However, the exact date of birth in that year was not known for any individual, and sampling was performed over a wide window (April through November) outside of the winter breeding period. We therefore used an estimated day of birth at the peak of the winter breeding season for all individuals (February 15) to refine age at the time of sampling.

Photo-identification research on this population did not begin until the 1970s and individuals cataloged since time were not all seen first as calves. Thus, chronological age is unknown for many individuals, including the oldest whales in the population. We therefore selected samples from individuals with long sighting spans (as long or longer than the upper 20% of the available known-age data) to clarify epigenetic age patterns at and beyond the top of the validated age range. In these cases, we calculated a minimum age at sampling based on the fact the individual could not have been born later than the year before their first sighting. However, such whales could have been born in any earlier year and so this provided only a minimum bound on their chronological age.

The sex of sampled individuals was known independently from molecular genetic analysis (Bérubé, 1996), in some cases supplemented by observations of the genital slit (Glockner, 1983) or calving history.

Cats (*20*)

Ethics. Sample collection was approved by the Clinical Research Ethical Review Board of the RVC (URN: 2019 1947-2).

Feline and other animal blood samples: The DNA archive of the Royal Veterinary College (RVC) was searched for feline ethylenediaminetetraacetic acid (EDTA) blood samples that were residuals from previous routine hematology testing. Cats were selected to represent the widest age range possible based on the available samples with a uniform distribution across the entire range, available breeds and neutering status. As the samples originated from cats that were presented for veterinary investigation, cats were selected to have no or minimal abnormalities on available laboratory data (hematology, serum biochemistry, endocrinology), reviewed by a board certified veterinary clinical pathologists (BSz). The DNA samples were maintained frozen at -80 °C for various amounts of time (0-11 years). Samples from guinea pigs, rabbits, ferrets, and alpacas were also residual samples from routine patients presented for veterinary care. Genomic DNA from cat blood was extracted using the Zymo DNA extraction kit according to the manufacturer's instructions. DNA was eluted in water and quantified with a picogreen kit according to the instructions provided.

Non-domestic cat species: Blood samples from cheetah (Latin name Acinonyx jubatus), lion (Panthera leo nubica), and tiger (Panthera tigris) were opportunistically collected and banked during routine health exams from these zoo-based animals located at Busch Gardens and White Oak Conservation.

Elephants (*9*)

Ethics. This study was authorized by the management of each participating zoo and, where applicable, was reviewed and approved by zoo research committees. In addition, the study received IACUC approval (#18-29) at the NZP; and endorsement from the elephant Taxon Advisory Group and Species Survival Plan.

Study Animals: Our study population included 140 elephants (57 African and 83 Asian) housed in 27 AZA-accredited zoos in North America (including Canada). Known or estimated birthdates were gleaned from each species' studbooks.

Uncertain surrounding age information is encoded in our variable "ConfidenceInAgeEstimate". A value of 90% indicates that the chronological age could be off by around 5 percent. Our analysis omitted animals for whom the confidence in the age estimate was less than 90 percent. For all elephants, the birthdates used in our analyses were taken from those recorded in the regional studbooks maintained through the Association of Zoos and Aquariums Taxon Advisory Group and Species Survival Plan for each species. Elephants had known birthdates if there were captive-born (i.e., they were recorded as having been born in a captive facility in the U.S. or in a range country), or were recorded as having an estimated birth date if they were imported (i.e., capture location and date of capture was recorded in the studbook). How age was estimated for imported individuals is not described in the studbooks but was most likely done by the broker at the time of importation based on individual morphometric measurements. The majority (>70%) of the North American population is imported as captive breeding as a population management tool was not fully implemented until the late 1990s. In our sample set we

had about an equal proportion of captive vs imported individuals, although we did have more imported animals reflecting the management history of the population. Our samples are made up of 65 captive-born/known birthdates (21 Africans and 26 Asians), and 77 imported/estimated birthdates (36 African and 27 Asian).

Whole blood samples from either an ear or leg vein directly into an EDTA tube were collected between 1998 and 2019 during regular veterinary examinations and shipped frozen to the genetics lab at the Smithsonian Conservation Biology Institute Center for Conservation Genomics (SCBI-CCG). The samples were stored in an ultralow freezer (-80ºC) until DNA extraction.

Yellow-bellied marmots (*28*)

Ethics. Data and samples were collected under the UCLA Institutional Animal Care and Use protocol (2001-191-01, renewed annually) and with permission from the Colorado Parks and Wildlife (TR917, renewed annually).

All samples were collected as part of a long-term study of a free-living population of yellow-bellied marmots in the Gunnison National Forest, Colorado (USA), where marmots were captured and blood samples collected biweekly during their active season (May to August).

Individuals were monitored throughout their lives, and chronological age was calculated based on the date at which juveniles first emerged from their natal burrows. We only used female samples because precise age for most adult males is unavailable since males are typically immigrants born elsewhere. We selected 160 whole blood samples from 78 females with varying ages. From these, DNA methylation (DNAm) profiling worked well for 149 samples from 73 females with ages varying from 0.01 to 12.04 years.

Roe deer (*23*)

Study populations: We sampled roe deer living in two enclosed forests with markedly different environmental contexts: Trois-Fontaines (TF) and Chizé (CH). The Trois-Fontaines forest (1,360 ha) is located in north-eastern France (48°43'N, 4°55'E) and is characterized by a continental climate, moderately severe winters and warm and rainy summers. This site has rich soils and provides high quality habitat for roe deer (Pettorelli et al. 2006). In contrast, the Chizé forest (2,614 ha) is located in western France (46°50'N, 0°25'W) and is characterized by temperate oceanic climate with Mediterranean influences. This site has a low productivity due to poor quality soils and frequent summer droughts (Pettorelli et al. 2006), and thereby provides a quite poor habitat for roe deer in most years. Individuals from these two populations have been intensively monitored using a long-term Capture-Mark-Recapture program since 1975 and 1977 (for Trois-Fontaines and Chizé, respectively). In each site, 10-12 days of capture using drive-netting are organized every year between December and March (see Gaillard et al., 1993 for details on capture sessions), which allows capturing and measuring about half the population every year. Once a roe deer is captured, its sex and body mass (to the nearest 50g) are recorded and a basic clinical examination is performed. All individuals included in our analyses were of known age because they were either caught as newborn in spring (see Delorme et al. 1988 for further details) or as c.a. 8 months old during winter captures, when they still have their milk teeth (most often incisors and always premolars, Flerov 1952).

Roe deer blood samples and DNA extraction: In 2016 and 2017, we collected blood samples (up to 1mL per kg of body mass) from the jugular vein. Within 30 min of sampling, the blood was centrifuged at 3000 g for 10 min and the plasma layer was removed before washing the cells with an equivalent volume of 0.9% w/v NaCl solution. After a second centrifugation, the intermediate

buffy coat layer, comprising mainly leukocytes, was collected in a 1.5-mL Eppendorf tube and immediately frozen at −80 °C in a freezer (Telstar SF 8025) until further use.

We extracted genomic DNA from leukocytes using the Macherey-Nagel NucleoSpin® Blood QuickPure kit. DNA purity was assessed using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington DE, USA). For all samples, the purity absorption range was 1.7 - 2.0 for the 260/280 nm ratio and >1.8 for the 260/230 nm ratio. We selected 96 samples by balancing the numbers of individuals among ages, and between populations and sexes. DNA concentration was determined spectrophotometrically using the Qubit assay kit. DNA samples were then diluted in ultrapure water to reach a concentration of ~70 ng.µl^{-1} and displayed in a microplate to complete the DNA methylation protocol (see below). For 6 samples, the concentrations obtained after dilution were too low compared to the expected concentrations of 70 ng/µl and were excluded from the dataset. The 90 roe deer samples analyzed in this study correspond to 79 individuals (i.e. 11 individuals were sampled both in 2016 and 2017) aged from 8 months to 13.5 years of age. This age range encompasses most of the roe deer lifespan as individuals older than 15 years of age are rarely observed in the wild (the oldest age ever recorded for a roe deer monitored in the wild being 17.5 years old, Gaillard et al. 1998).

Zebras (*11*)

Ethics: Plains zebra samples were collected under a protocol approved by the Research Safety and Animal Welfare Administration, University of California Los Angeles: ARC # 2009-090-31, originally approved in 2009.

Samples: Both whole blood (96) and remote biopsy (24) samples were obtained from a captive population of zebras maintained in a semi-wild state by the Quagga Project (*72*) in the Western Cape of South Africa. The population was founded in 1989 with 19 wild individuals (9 from Etosha National Park in Namibia, 10 from the Kwazulu-Natal in South Africa). Since its inception, the population has undergone strong selection in an effort to reproduce the phenotype of the extinct quagga subspecies: no stripes on legs and hindquarters, and thinner and paler stripes in the head and barrel region. During sampling, individuals were uniquely identified by their stripe patterns and their ages were derived from studbook information about dates of birth, which were typically known within one month. One exception to this is a tissue sample from a founder that was sampled as a young mare and would have been at least 25 years old at sampling, but may have been slightly older. Remote biopsies were taken using an air-powered rifle affixed with a 1 mm wide by 20-25 mm deep biopsy dart and preserved in RNAlater (Qiagen). Blood samples were collected opportunistically during veterinarian visits, and preserved in EDTA tubes. Most samples were collected from different individuals, with the exception of two individuals that each were sampled twice some years apart. All samples were stored at -20 °C. After eliminating samples with low confidence for individual identity and age, we retained 76 blood samples and 20 biopsy samples, totaling 96 zebra samples.

Three zebra data sets were analyzed in our epigenetic models: (1) only blood samples, (2) only biopsy samples, and (3) blood and biopsy samples combined. The Grevy's zebra (n=5) and Somali wild ass (n=7), are samples from zoo-based animals that were opportunistically collected and banked during routine health exams and the DNA methylation profiles from these samples have been reported previously (*27*).

Rat (*73*)

The rat tissues came from 4 different labs across three countries:(i) India: Nugenics Research in collaboration with School of Pharmacy SVKM's NMIMS University (K. Singh), (ii) United States: University of Tennessee Health Science Center (H. Chen) and Medical College of Wisconsin (L.C. Solberg Woods), and (iii) Argentina: University of La Plata (R. Goya).

Rats from Tennessee and Wisconsin: Blood samples (n=48): Male and female heterogeneous stock rats were bred at the Medical College of Wisconsin (Solberg Woods Lab) or University of Tennessee Health Science Center (Hao Chen Lab). Heterogeneous Stock (HS) populations were originally developed by breeding together eight inbred strains, followed by maintaining the colony in a manner that minimizes inbreeding, allowing fine-resolution genetic mapping of a variety of complex traits. Rats were euthanized at different ages by an overdose of isoflurane (> 5%). Trunk blood was collected immediately and stored at −80 °C until processing. Blood samples were treated with streptokinase (60-80 IU/200 µl blood, overnight incubation at 37 °C) and DNA was extracted using the QiaAmp Blood Mini Kit (Qiagen Cat No./ID: 51304) following manufacturer's instructions. All procedures were approved by the Institutional Animal Care and Use Committee of the University of Tennessee Health Science Center or the Medical College of Wisconsin and followed the NIH Guide for the Care and Use of Laboratory Animals. Genomic DNA was isolated from tissue samples mostly using Puregene chemistry (Qiagen). DNA from the liver was extracted manually and from blood using an automated Autopure LS system (Qiagen). From tissues and clotted blood samples DNA was extracted manually using QiaAmp DNA Blood Midi Kit and the DNeasy Tissue Kit according to manufacturer's protocol (Qiagen, Valencia, CA). DNA from BA10 was extracted on an automated nucleic acid extraction platform Anaprep (Biochain) using a magnetic bead-based extraction method and Tissue DNA Extraction Kit (AnaPrep).

Rats from the University of La Plata (R. Goya lab): Multiple tissues/cell types (adipose, blood, cerebellum, hippocampus, hypothalamus, liver, neocortex, ovaries, pituitary, skin, substantia nigra): Young (3.7 mo., n=11), Late Adults (LA, 8.0 mo., n=9), Middle-Aged (M-A, 15.7 mo., n=6) and Old (25.5 mo., n=14) female Sprague-Dawley (SD) rats, raised in our Institute, were used. Animals were housed in a temperature-controlled room (22 ± 2ºC) on a 12:12 h light/dark cycle. Food and water were available *ad libitum*. All experiments with animals were performed in accordance to the Animal Welfare Guidelines of NIH (INIBIOLP's Animal Welfare Assurance No A5647-01) and approved by our Institutional IACUC (Protocol # P05-02-2017).

Tissue sample collection: Before sacrifice by decapitation, rats were weighed, blood was withdrawn from the tail veins with the animals under isoflurane anesthesia and collected in tubes containing 10µl EDTA 0.342 mol/l for 500µl blood. The brain was removed carefully, severing the optic and trigeminal nerves and the pituitary stalk (not to tear the pituitary gland), weighed and placed on a cold plate. All brain regions were dissected by a single experimenter (see below). The skull was handed over to a second experimenter in charge of dissecting and weighing the adenohypophysis. The rest of the body was handed to other 2 or 3 experimenters who dissected and collected whole ovaries, a sample of liver tissue, adipose tissue and skin tissue from the distal portion of tails.

Brain region dissection: Prefrontal cortex, hippocampus, hypothalamus, substantia nigra and cerebellum were rapidly dissected on a cold platform to avoid tissue degradation. After dissection, each tissue sample was immediately placed in a 1.5ml tube and momentarily immersed in liquid nitrogen. The brain dissection protocol was as follows. First a frontal coronal cut was made to discard the olfactory bulb, then the cerebellum was detached from the brain and from the medulla oblongata using forceps. To isolate the medial basal hypothalamus (MBH), brains were placed ventral side up and a second coronal cut was made at the center of the median eminence (−3,6 mm

referred to bregma). Part of the MBH was taken from the anterior block of the brain and the other part from the posterior block in both cases employing forceps. The hippocampus was dissected from the cortex in both hemispheres using forceps. This procedure was also performed on the anterior and posterior blocks, alternatively placing the brain caudal side up and rostral side up. To dissect the substantia nigra, in each hemisphere a 1-mm thick section of tissue was removed from the posterior part of the brain (−4,6 mm referred to bregma) using forceps. Finally, the anterior block was placed dorsal side up, to separate the prefrontal cortex. With a sharp scalpel, a cut was made 2 mm from the longitudinal fissure, and another cut was made 5 mm from it. Additionally, two perpendicular cuts were made, 3 mm and 6 mm from the most rostral point, obtaining a 9 mm^2 block of prefrontal cortex. This procedure was performed in both hemispheres and the two prefrontal regions collected in a code-labeled tube.

Anterior pituitary: Using forceps the dura mater that covers gland was removed leaving the organ free on the sella turcica. The neural lobe was carefully separated from the anterior pituitary (AP) which was then carefully lifted with fine curved tip forceps pointing upwards. It was rapidly weighed, then put in a tube and placed momentarily in liquid nitrogen.

Ovaries: The genital apparatus was dissected by cutting the mesentery to isolate the uterine horns, the tubular oviduct, the ovaries and the junction between the anus/rectum and the vulva/vagina, leaving the unit of the sexual organs and the urinary bladder isolated. The ovaries were carefully separated from the oviducts; the fat around the ovaries was also removed. Both gonads were placed in a single eppendorf tube and momentarily placed in liquid nitrogen.

Liver: Liver tissue extraction was made by cutting a piece of the median lobe (0.5 cm x 0.5 cm). Tissue was placed in a tube and momentarily stored immersed in liquid nitrogen.

Adipose tissue: Adipose tissue samples were obtained from the fatty tissue of the small intestine.

Tail skin: For skin tissue, 5 cm of a distal tail portion were cut with scissors. Skin was separated and hair removed using scalpel. Tissue was placed in a tube and stored as described for other tissues.

DNA was extracted from blood on an automated nucleic acid extraction platform called QiaSymphony (Qiagen) with a magnetic bead-based extraction kit, QIAsymphony DNA Midi Kit (Qiagen). DNA was extracted from tissue on an automated nucleic acid extraction platform called Anaprep (Biochain) with a magnetic bead-based extraction kit, Tissue DNA Extraction Kit (Biochain). DNA from brain regions was extracted using an automated nucleic acid extraction platform called QIAcube HT (Qiagen) with a column-based extraction kit, QIAamp 96 DNA QIAcube HT Kit (Qiagen).

Rats from Nugenics Research Group: Sprague Dawley rats of both sexes were used, from which blood, whole brain, heart and liver were harvested. Two batches or samples were prepared: the first batch was intended for training the epigenetic clock: n=42 blood samples, n=18 whole brain, n=18 heart, n=18 liver samples. The second batch of test set involved n=76 tissue samples (n=22 blood, n=18 liver, n=18 heart, n=18 hypothalamus). The test data were used to evaluate the effect of the treatment in 3 conditions: young (30-week old) and treated old samples (109 weeks old), and untreated old samples (again 109 weeks old). We evaluated 4 sources of DNA: blood (n=18), liver (n=18), heart (n=18) and hypothalamus (n=18). Ethics committee approval number - CPCSEA/IAEC/P-6/2018.

Male Sprague Dawley rats of 8 weeks (200–250 g) and 20 months (400-450g) were procured from the National Institute of Bioscience, Pune, India. Animals were housed in the animal house facility of School of Pharmacy, SVKM’s NMIMS University, Mumbai during the study under

standard conditions (12:12 h light: dark cycles, 55-70% of relative humidity) at 22±2°C temperature with free access to water and standard pellet feed (Nutrimix Std-1020, Nutrivet Life Sciences, India). The animals were acclimatized to the laboratory environment for seven days before initiation of the study. The experimental protocol was approved by the Institutional Animal Ethics Committee. The approval number is CPCSEA/IAEC/P-75/2018. Rats were euthanized at different ages by an overdose of isoflurane (> 5%). Trunk blood was collected immediately and stored at −80 °C until processing. 100 µl blood sample was treated with 20 µl Proteinase K and then the volume was adjusted to 220 µl with Phosphate Buffer Saline (PBS) in 1.5 ml or 2 ml microcentrifuge tube. 200 µl buffer AL was mixed thoroughly to this mixture by vortexing, and incubated at 56°C for 10 min. Then 200 µl ethanol (96–100%) was added to the sample and mixed thoroughly by vortexing and DNA was extracted using the Qiagen DNeasy blood and tissue kit, Qiagen Cat No./ID: 69504 following manufacturer's instructions. The study protocol was approved through the Institutional Animal Ethics Committee (approval no. CPCSEA/IAEC/P-6/2018) which was formed in accordance with the norms of the Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA), Government of India and complied with standard guidelines on handling of experimental animals.

According to the manufacturer's instructions, 20 µl Proteinase K was pipetted into a 1.5 ml or 2 ml microcentrifuge tube, to this 50–100 µl anticoagulated blood was added and the volume was adjusted to 220 µl with PBS. 200 µl Buffer AL was added and mixed thoroughly by vortexing, and incubated at 56°C for 10 min. Finally 200 µl ethanol (96–100%) was added to the sample, and mixed thoroughly by vortexing. The mixture was added to the DNeasy Mini spin column placed in a 2 ml collection tube and centrifuged at 6000 x g for 1 min. Further the DNeasy Mini spin column was placed in a new 2 ml collection tube (previous flow through and collection tube was discarded), 500 µl Buffer AW1 was added and centrifuged for 1 min at 6000 x g. Again flow-through and collection tube was discarded. The DNeasy Mini spin column was placed in a new 2 ml collection tube, 500 µl Buffer AW2 was added, and centrifuged for 3 min at 20,000 x g to dry the DNeasy membrane. Flow-through and collection tube was discarded. DNeasy Mini spin column was placed in a clean 1.5 ml or 2 ml microcentrifuge tube and 200 µl Buffer AE was pipetted directly onto the DNeasy membrane. Sample was incubated at room temperature for 1 min, and then centrifuged for 1 min at 6000 x g to elute.

Dog (*29*)

DNA samples from n = 742 dog blood samples from 93 breeds were provided by researchers at the National Human Genome Research Institute (NHGRI). The weight of individual dogs was unknown. Collection was approved by the Animal Care and Use Committee of the Intramural Program of NHGRI at the National Institutes of Health (Protocol #8329254).

Dog lifespan and breed characteristics: Standard breed weight (SBW), height (SBH) and lifespan were aggregated from several sources. SBW and SBH were taken from previously reported values (*74*, *75*), which were updated if AKC values differed (*76*). If the AKC did not specify SBW or SBH, we used data from Atlas of Dog Breeds of the World (*77*). Lifespan estimates were calculated as the average of the standard breed across sexes, compiled from numerous publications consisting primarily of multi-breed surveys of age and cause of death from veterinary clinics and large-scale breed-specific surveys, which are often conducted by purebred dog associations. Sources for lifespan are reported in Supplementary Note 1. When available, data was combined across surveys for number of dogs, minimum, maximum, mean, and median age at

death. The minimums, maximums, and medians were averaged across studies to produce a representative lifespan expectation for each breed. For three breeds (American Hairless Terrier, Sloughi, and Ibizan Hound), no published survey data was available. For these breeds, the maximum age expectation was obtained from the American Kennel Club website.

Bats (*26*)

Wing tissue samples. Wing punches were taken from 778 individually marked animals that were either kept in captivity (15 species) or recaptured as part of long-term field studies (11 species). We excluded 42 samples because we did not have independent evidence to confirm minimum age estimates. For 630 samples the individual was marked shortly after birth, so age estimates were exact. For the remainder, age represented a minimum estimate because the individual was not initially banded as a juvenile. We used minimum age estimates when other evidence, such as tooth wear or time since initial capture, indicated that the minimum age estimate was likely to be close to the real age. The study was approved by the University of Maryland Institutional Animal Care and Use Committee (FR-APR-18-16).

Here, we provide additional information on when and where samples were taken from either captive or free- ranging animals for each of the 26 species of bats used in this study.

Pallid bats, *Antrozous pallidus,* were captured between 2005 and 2008 at six sites in central Oregon (44.94° N, 120.38° W) using mist nets over a water source or outside a night roost or with a handnet on an extension pole outside a day-roosting crevice. Each bat was weighed, measured and marked with a numbered band. Adults were distinguished from juveniles by closed epiphyseal gaps. Tissue samples were obtained from wing membranes using 3 mm biopsy punches and stored in 95% ethanol until DNA was extracted using a Qiagen DNeasy Tissue Kit. DNA extracts were stored frozen at -80°C. Live animal procedures conformed to the American Society of Mammalogists guidelines and were approved by the University of Maryland Institutional Animal Care and Use Committee (protocols R-05-26 and R-08-39). Bat capture and sampling was conducted with permission of the Pine Creek Conservation Area, the Oregon Department of Fish and Wildlife (permit 081-95), and the John Day Fossil Beds National Monument, National Park Service (permit JODA-2005-SCI-0003).

Wing tissue from Artibeus jamaicensis, Cynopterus brachyotis, Eidolon helvum, Pteropus giganteus, P. hypomelanus, P. poliocephalus, P. pumilus, P. rodricensis, P. vampyrus, and Rousettus aegyptiacus was taken between 2006 and 2017 from bats kept at the Lubee Bat Conservancy, an AZA (Associated Zoos and Aquariums [https://www.aza.org/current-cert]) certified facility, in Gainesville, Florida. The bats are group-housed in twelve 1068 sq. ft. enclosures with indoor temperature-controlled roosting areas and outdoor flight rooms and are fed a diet of fruit, vegetables and nutritional supplements. Wing tissue biopsies are periodically taken from individually marked animals and kept at -20°C in 95% ethanol. The majority of 243 samples from these species were taken from animals that were born in captivity. DNA was extracted with a Zymo miniprep plus kit.

Wing tissue samples were taken in 2018 from captive Carollia perspicillata housed in a tropical zoo (Papiliorama, Kerzers FR, Switzerland). Approximately 400 bats roost in an artificial cave kept on a reversed light cycle and are fed twice a night with a fruit-based diet. Since 2011, the population has been monitored by capturing individuals using a harp-trap placed at the entrance to the cave. Forearm length, body weight, reproductive status and tooth-wear are recorded from every captured individual. At first capture, individuals are marked on the forearms with a unique combination of three colored plastic rings (A.C. Hughes, UK, size XB). All captures and markings

were authorized by the cantonal veterinary service (permits nb:2011_42_FR, 2013_10E_FR, 2014_59_FR). Between July and November 2018, 3mm biopsies were punched on the patagium and hermetically stored in silica gel. Based on the date of first capture and tooth-wear score, the age of each individual sampled was estimated3. DNA was extracted with a Zymo miniprep plus kit.

Biopsy punches (2 or 3 mm) were taken from the wing of captive common vampire bats, *Desmodus rotundus*, between 2010 and 2014. Bats were housed and fed blood4 in flight cages (3 x 2 x 1.5 m) as a captive group at the Cranbrook Institute of Science (24-39 bats, Bloomfield Hills, MI, USA) or at the University of Maryland (7 bats, University of Maryland Institutional Animal Care and Use Committee protocol R-10–63). Age was determined based on zoo birth records. Individuals were born at the Houston Zoo, Cincinnati Zoo, Chicago Brookfield Zoo, or the Cranbrook Institute of Science. Tissue samples were stored in 95% ethanol prior to DNA extraction using a Qiagen DNeasy kit. DNA extracts were frozen for long-term storage at -80°C. Live animal procedures conformed to the American Society of Mammalogists guidelines and were approved by the University of Maryland Institutional Animal Care and Use Committee (protocol R-10-63).

Wing tissues of big brown bats (*Eptesicus fuscus*) were sampled with a 3 mm biopsy punch from the wing of known age captive bats at Northeast Ohio Medical University (NEOMED; Rootstown, Ohio) in 2018 and 2019. These animals came from a colony previously maintained by Dr. Ellen Covey at the University of Washington, which was started in 2005 with bats caught in North Carolina that were banded according to year of capture or birth. These bats underwent natural hibernation and were exclusively fed an *ab libitum* diet of fresh water and mealworms (*Tenebrio molitor*). In 2014 some of the bats were transported to NEOMED and are now housed indoors on a 12 h light/dark cycle and fed the same *ab libitum* fresh water and mealworm diet. Wing punches were taken from a second colony of *E. fuscus* by Dr. Paul Faure and Lucas Greville in February or August, 2020. This colony is kept at McMaster University in Hamilton, Ontario, Canada in an indoor/outdoor enclosure (2.5 x 8.3 x 2.7m)7 in which the temperature fluctuates with ambient conditions, but is kept above freezing by a heater on a thermastat. Known-age animals from this colony were born in captivity from females captured in Ontario. Tissue samples from both colonies were stored in DNA Shield and kept frozen at -20°C prior to DNA extraction with a Zymo miniprep plus kit. Animal use protocols were approved by the NEOMED Institutional Animal Care and Use Committee or the Animal Research Ethics Board of McMaster University (AUP# 20-05-20).

During July 2019, wing membrane samples were obtained from sub-adult or adult female and subadult male lesser long-nosed bats, *Leptonycteris yerbabuenae*, with a 4 mm biopsy punch at the entrance of the Pinacate Cave in the Reserva de la Biosfera el Pinacate y Gran Desierto de Altar (31°38'51.6" N, 113°28'53.5" W), Sonora, Mexico. Bats were captured using mist nets (Avinet models: TB02, TB06, TB012; Portland, Maine, USA) set outside caves just prior to when bats emerged to forage. Individuals were sexed, weighed and the forearm measured. To discriminate subadults from adults, age was determined by the degree of fusion of the epiphyses at the metacarpal–phalangeal joint. Tissue samples were stored in a DNA/RNA Shield buffer (Zymo Scientific, Irvine, CA 92614, U.S.A.). DNA was extracted with a Zymo miniprep plus kit. Bat tissue samples were collected under permit SGPA/DGVS/06361/17 issued to R. A. Medellín by The Ministry of Environment and Natural Resources.

Samples of velvet free-tailed bats, *Molossus molossus*, come from a long-term study in Gamboa, Panama (09°07’ N 79°41’ W), where the bats roost in crevices in houses. We captured

social groups with mist nets (Ecotone, Gydnia, Poland) at the entrance of roosts during evening emergence and individually marked all bats with a subcutaneous passive integrated transponder (Trovan ID-100, Euro ID, Weilerswist, Germany) at first capture. Wing tissue samples were taken with a 3 mm biopsy punch and stored in 96% ethanol until DNA extraction using a Zymo miniprep kit. Capture and handling of animals were carried out under permits SE/A-112-13, SE/A-73-14, SE/A-95-15, and SE/A-32-17 from the Autoridad Nacional del Ambiente in Panama with approval from the Institutional Animal Care and Use Committee of the Smithsonian Tropical Research Institute (2012-0505-2015).

Little brown bats, *Myotis lucifugus*, were captured as they departed from an attic maternity colony in Chestertown, Maryland (39°12'N, 76°04'W), in September 1996. Captured bats were weighed, measured and banded with individually marked bands. Young of the year were identified by their weight and absence of tooth wear. Wing membrane biopsies were taken and stored in a 5M NaCl with 20% dimethyl sulfoxide solution and kept frozen at -80°C. DNA was extracted with a Zymo miniprep plus kit. Bat capture and handling was approved by the Maryland Department of Natural Resources (permit SCO-30403).

Wing tissue samples were taken from greater mouse-eared bats, *Myotis myotis*, between 2013 and 2018 as part of a long-term mark-recapture study conducted by Bretagne Vivante in Brittany, France (47°35'N, 2°14'W"). Bats were caught using modified harp traps as they left one of five different roosts. Individuals at first capture are fitted with PIT tags to facilitate identification on subsequent recaptures. Measurements taken from each individual include sex, forearm length, weight and transponder number. Age class (juvenile or adult) is determined by examining the degree of the epiphyseal closure of the metacarpal-phalangeal joints. Wing biopsies were taken with a 3 mm biopsy punch, flash frozen and stored in liquid nitrogen prior to extraction. All procedures were conducted with full ethical approval and permission (AREC-13- 38-Teeling) awarded by the University College Dublin ethics committee and in accordance with permits issued by 'Arrêté' by the Préfet du Morbihan. DNA was extracted from wing biopsies using a Promega Wizard SV DNA extraction kit (catalog no. A2371) or the Qiagen DNeasy Blood and Tissue kit (Qiagen). Extractions carried out with the Promega kit were partially automated using a Hamilton STAR Deck liquid handling robot.

Adult and juvenile Mexican fishing bats, *Myotis vivesi*, were captured by gloved hand from roosts in talus slopes on Isla Partida Norte in the Gulf of California, Mexico (29°03'N, 113°00'W) during the day between 2015 and 2018. Individuals were measured and banded with numbered metal bands on their forearms for identification upon recapture. In 2018 wing tissue was taken with a 3 mm biopsy punch and preserved in Zymo DNA shield. Bat capture and handling were conducted under permits #7668–15, 2492–17 and #5409-18 from Dirección General de Vida Silvestre, and permits #17–16, 21–17 and 20-18 from Secretaría de Gobernación, and the University of Maryland Institutional Animal Care and Use Committee protocols FR-15-10 and FR-18-20.

Common noctules, *Nyctalus noctula*, were captured as part of a long-term study16,17,18 at the Seeburgpark in Kreuzlingen, Switzerland (47.649928° N, 9.186123° E) where bats regularly roost in boxes. Each bat was marked with a subcutaneous pit-tag (ID100; Euro ID, Weilerswist, Germany) injected under the dorsal skin. Wing tissue samples were taken with a 3 mm biopsy punch and stored in 96% ethanol until DNA extraction using a Zymo miniprep kit. All handling and sampling of the bats in Switzerland was approved by the Veterinäramt Thurgau (permit FIBL1/12).

Wing tissue samples were taken in September 2018 from lesser spear-nosed bats, *Phyllostomus discolor*, kept in a breeding colony in the Department Biology II of the Ludwig-Maximilians-University in Munich. In this colony animals were kept under semi-natural conditions (12 h day/night cycle, 65 to 70 % relative humidity, 28°C) with free access to food and water. The license to keep and breed *P. discolor* was issued by the German Regierung von Oberbayern. Under German Law on Animal Protection a special ethical approval is not needed for wing tissue collection. Wing tissue was stored in RNAlater until DNA was extracted using QIAamp® MinElute columns following the manufacturer's instructions. The samples were eluted in 50µl of molecular grade water and concentrated to reduce their volume by approximately 50% using a Speedvac, with the following settings: duration 20 minutes, temperature in the chamber 30°C, H2O (water) mode. Greater spear-nosed bats, *Phyllostomus hastatus,* were captured and sampled between 1990 and 2018 in Trinidad, Lesser Antilles19,20,21,22. Most often, harem groups, which include one adult male plus 15-20 lactating females with pups, were captured during the day from within a solution depression in the ceiling of either Tamana (10.4711°N, 61.1958°W), Caura (10.7019°N, 61.3614°W), or Guanapo cave (10.6942°N, 61.2654°W) using a bucket trap. Captured bats were sexed, measured for size, weight, and tooth wear, and individually marked with stainless steel numbered bands. Age was determined exactly for adults that were recaptured after being banded as pups. Wing biopsy punches (4 mm) were stored frozen at -80°C in either a 5M NaCl with 20% dimethyl sulfoxide solution or Zymo DNA Shield prior to DNA extraction using a Qiagen Puregene or Zymo miniprep plus kit. Frozen samples were selected to maximize the number of known-age individuals with approximately equal numbers at all ages. Animal handling methods follow guidelines by the American Society of Mammalogists and were approved by the University of Maryland Institutional Animal Care and Use Committee (protocols R-91-33, R-93-22, R-94-25, R-01-07, R-11-21, R-13-77, FR-APR-18-16) under licenses from the Forestry Division of the Ministry of Agriculture, Land and Fisheries, Trinidad and Tobago.

Wing tissue samples were taken from greater horseshoe bats, *Rhinolophus ferrumequinum*, by using 3 mm biopsy punches between 2016 and 2018 from wild female bats as part of a long-term study at a maternity colony in Gloucestershire, UK (51.7107°N, 2.2777°W). Bats were captured at the roost with hand nets, and all individuals were weighed, ringed with aluminum alloy rings and morphometric data such as forearm length recorded under licenses (Natural England Project Licenses 2015-9918-SCI-SCI; 2016-23583- SCI-SCI; 2017-30137-SCI-SCI) issued to Roger Ransome. All bats studied were first marked as infants, so we could be certain of their age. Bats were aged between 1-21 years, with the 40 individuals selected in a fairly even manner across this age span. Sampling procedures were conducted under licenses (Natural England 2015-11974-SCI-SCI; 2016-25216-SCI-SCI; 2017- 31148-SCI-SCI) issued to Gareth Jones, with tissue biopsy additionally licensed under Home Office Project Licenses (PPL 30/3025 prior to 2018; P307F1428 from 2018 onwards) and Home

Tissue samples were stored in silica gel beads and then transferred to a -20C freezer for long-term storage. DNA was extracted from wing biopsies using a Promega

Wizard SV DNA extraction kit or a Qiagen DNeasy Blood and Tissue kit. Extractions carried out with the Promega kit were partially automated using a Hamilton STAR Deck liquid handling robot but otherwise followed the manufacturer's instructions.

Samples of proboscis bats, *Rhynchonycteris naso,* came from a long-term study between 2005 and 2016 at La Selva Biological Station in Costa Rica (10° 25′ N, 84° 00′W)26,27,28. Bats were mist-netted in the vicinity of their roosts. Wing tissue was sampled with a 4 mm biopsy punch, individuals were marked with colored plastic bands, sexed, measured and age class determined

(juvenile: 0-4 months, subadult: 5-10 months, or adult>10 months)26. Age was determined exactly for individuals that were banded as pups and recaptured as adults. Ethanol (80%) was used to preserve tissue samples, and a salt–chloroform procedure or Qiagen BioSprint 96 DNA Blood Kit was used for DNA isolation26,28. Research permits were granted by the MINAE (Ministerio del Ambiente y Energia) and the ACC (Área de Conservación Central). Animal treatment followed the Guide for Care and Use of Laboratory Animals of the National Institutes of Health. Animal handling complied with current Costa Rican laws and was approved by the Animal Care Review Committee of the SINAC (Sistema Nacional de Áreas de Conservación) and MINAE (permits 022-2005-OFAU, 108-2006-SINAC, 147-2007-SINAC,

Office personal licences.

183-2008-SINAC, 187-2009-SINAC, 130-2010-SINAC and 068-2011-SINAC, 115–2012-SINAC, 033–2013-SINAC, SINAC-SE-GASP-PI-R-121–2013, R-006–2015-OT-CONAGEBIO, SINAC-SE-CUS-PI-R-088–2016).

Samples of greater sac-winged bats, *Saccopteryx bilineata* came from long-term studies in Costa Rica (n = 21 from La Selva Biological Station, 10° 25′ N, 84° 00′W and n = 6 from Santa Rosa National Park, 10° 53′ N, 85° 46′ W) and Panama (n = 4 from the Biological Station Barro Colorado Island (BCI) of the Smithsonian Tropical Research Institute, 9° 9′ N/79° 51′ W) between 1994 and 2016. Bats were captured with mist nets when entering or leaving their day roosts, individually banded with two coloured plastic bands on their forearms, and a wing tissue biopsy sample (4mm) preserved in 80% ethanol was taken. Age was determined exactly for individuals that were banded as pups and recaptured as adults. DNA was extracted with a salt-chloroform procedure or with the Qiagen BioSprint 96 DNA Blood Kit. The process of acquiring data and protocols for capturing and handling bats complied with the current laws of Panama and were conducted in accordance with the relevant guidelines and regulations. Our study in Panama was approved by the Smithsonian Tropical Research Institute and its Animal Care and Use Committee (ACUC, permits: IACUC 100316-0910-12, ACUC 2013-1015-2016). For research in Costa Rica, permits were granted by the MINAE (Ministerio del Ambiente y Energia), the ACC (Área de Conservación Central) and the ACG (Área de Conservación Guanacaste). Animal treatment followed the Guide for Care and Use of Laboratory Animals of the National Institutes of Health. Animal handling complied with current Costa Rican laws and was approved by the Animal Care Review Committee of the SINAC (Sistema Nacional de Áreas de Conservación) and MINAE (permits 272-2003-OFAU, 135-2004-OFAU, 022-2005-OFAU, 108-2006-SINAC, 147-2007-SINAC, 183-2008-SINAC, 187-2009-SINAC, 130-2010-SINAC and 068-2011-SINAC, 115–2012-SINAC, 033–2013-SINAC, SINAC-SE-GASP-PI-R-121– 2013, R-006–2015-OT-CONAGEBIO, SINAC-SE-CUS-PI-R-088–2016).

Mexican free-tailed bats, *Tadarida brasiliensis,* are housed at Bat World Sanctuary, a licensed non-profit bat rehabilitation facility and accredited by the Global Federation of Animal Sanctuaries (https://www.sanctuaryfederation.org/sanctuaries/bat-world/) in Weatherford, Texas. Most individuals sampled were rescued as pups, although some were rescued as adults, making their exact age unknown. Individuals are group-housed in large indoor enclosures. Wing membrane biopsies (4 mm) were collected by Amanda Lollar in August 2019 and stored in Zymo DNA Shield until DNA was extracted using a Zymo miniprep plus kit.

Cattle (*8*)

Ethical authorization and animals: All animal procedures were carried out in accordance with the relevant guidelines at each institution. Specifically, procedures related to sample collection in Poland followed the EU Directive of the European Parliament and the Council on the protection of animals used for scientific purposes (22 September 2010; No 2010/63/EU), Polish Parliament Act on Animal Protection (21 August 1997, Dz.U. 1997 nr 111 poz. 724) with further novelization - Polish Parliament Act on the protection of animals used for scientific or educational purposes (15 January 2015, Dz.U. 2015 poz. 266). Blood and oocyte collection were approved by the Local Ethics Committee for Experiments on Animals, University of Warmia and Mazury in Olsztyn, Poland (Agreement No. LKE.065.27.2019). For animal procedures in the USA, approval from the University of Nebraska Institutional Animal Care and Use Committee was obtained (approval number is 1560). Samples were obtained from 357 female cattle (Bos Taurus). The animals were housed on the Dairy Farm of the Institute of Animal Reproduction and Food Research of Polish Academy of Sciences, (Wielki Las, Poland) and at the Eastern Nebraska Research and Extension Center at the University of Nebraska-Lincoln (Nebraska, USA). In the present study, samples were collected from Polish Red cattle and from the herd in Eastern Nebraska. In this herd, black Angus or composites of varying percentages of Simmental x Angus (black) or Red Angus were used as blood donors. Animals were free of Bovine Herpesvirus Type 1, Bovine Viral Diarrhea/Mucosal Disease, tuberculosis and Enzootic bovine leucosis.

Blood collection and further processing: The blood samples from both herds were collected during routine animal management activities, during the routine blood collection for disease prevention. Blood samples were taken only from cows in the luteal phase of the estrous cycle. Blood was collected into 8ml PAXgene Blood DNA Tubes (Quiagen, Cat No. 761115) and stored at -80°C until the shipment (USA Veterinary Permission Nr 138809) to the UCLA Technology Center for Genomics & Bioinformatics (Los Angeles, USA) for further analyses.

Oocyte Collection and further processing: In total 80 Bovine ovaries were collected immediately *post mortem* from 40 cows which were selected for routine culling due to management reasons. All cows were in the luteal phase of the estrous cycle. Before the isolation of both ovaries from each cow, blood samples were also collected into 8ml PAXgene Blood DNA Tubes (Quiagen, Cat No. 761115) to generate both sample types from one donor. Isolated ovaries were kept on ice and immediately transported to the laboratory. Afterwards, immature bovine cumulus–oocyte complexes (COCs) were recovered by aspirating ovarian follicles in the diameter of 2–8 mm. Cumulus cells were removed from COCs by pipetting them for 5min in a Petri dish containing 500 µl of Phosphate Buffered Saline (PBS) with 0.1% hyaluronidase (Sigma-Aldrich, St. Louis, MO, USA). Denuded oocytes from every donor were pooled (10-15 immature oocytes/ovary) and then processed for genomic DNA isolation.

Mouse data (*12*)

The mouse data come from several sources listed below.

UCLA Lab. Animal breeding and husbandry: All mice were maintained and bred under standard conditions consistent with National Institutes of Health guidelines and approved by the University of California, Los Angeles Institutional Animal Care and Use Committees. The cages were maintained on a 12:12 light/dark cycle, with food and water ad lib. Tissues were harvested from discervical mice and fresh frozen on dry ice.

The Tet3 knockout mice were generated at the Jackson Laboratory using CRISPR technology with deletion of exon 4. Tet3 heterozygous knockout mice were crossed with C57BL/6J (JAX 000664) to set up the colony. All mice were maintained and bred under standard conditions

consistent with National Institutes of Health guidelines and approved by the University of California, Los Angeles Institutional Animal Care and Use Committees. The cages were maintained on a 12:12 light/dark cycle, with food and water ad lib. Mice were euthanized by cervical dislocation, cortical and striatal tissues were then quickly dissected out and fresh frozen on dry ice until used for DNA extraction.

Bcl11b heterozygous mouse was obtained from MMRRC at University of California, Davis (Catalog # 046780-UCD) and crossed with C57BL/6J (JAX 000664) to set up the colony. All mice were maintained and bred under standard conditions consistent with National Institutes of Health guidelines and approved by the University of California, Los Angeles Institutional Animal Care and Use Committees. The cages were maintained on a 12:12 light/dark cycle, with food and water ad lib. Mice were euthanized by cervical dislocation, cortical and tissues were then quickly dissected out and fresh frozen on dry ice until used for DNA extraction.

BXD mice (*12*) University of Tennessee Health Science Center: Samples for this study were selected from a larger colony of BXD mice that were housed in a specific pathogen-free (SPF) facility at the University of Tennessee Health Science Center (UTHSC). All animal procedures were in accordance with protocol approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Tennessee Health Science Center. Detailed description of housing conditions and diet can be found in (Roy, 2020;Williams, 2020). Mice were given ad libitum access to water, and either standard laboratory chow (Harlan Teklad; 2018, 18.6% protein, 6.2% fat, 75.2% carbohydrates), or high-fat chow (Harlan Teklad 06414; 18.4% protein, 60.3% fat, 21.3% carbohydrates). Animals were first weighed within the first few days of assignment to either diets, and this was mostly but not always prior to introduction to HFD. Following this, animals were weighed periodically, and a final time (BWF) when animals were humanely euthanized (anesthetized with avertin at 0.02 ml per g of weight, followed by perfusion with phosphate-buffered saline) at specific ages for tissue collection. The present work utilizes the biobanked liver specimens that were pulverized and stored in -80 °C, and overlaps samples described in (Williams, 2020).

Growth hormone receptor knockout from the University of Michigan: Growth hormone receptor knockout mice (GHR -/- dwarfs) from the University of Michigan (Richard A. Miller). Full-body growth hormone receptor knockout (GHR-KO) (*78*).

Calorie restricted mice from the University of Texas Southwestern Medical Center (*79*)

Ethics: The Institutional Animal Care and Use Committee (IACUC) of the University of Texas Southwestern Medical Center approved the animal protocol (APN 2015-100925), which has been subsequently renewed every 3 years (2018 and 2021).

Animals

C57BL/6J male mice (6 week-old, n = 72) were obtained from the Mouse Breeding Core, Wakeland lab, UT Southwestern Medical Center, Dallas, TX, USA. After 2 weeks of acclimation, mice were individually housed in standard polycarbonate mouse cages (Fischer Scientific, Cat. Nos. 01-288-1B and 01-288-21) with stainless steel running wheel, inside isolation cabinets containing 12 cages each. Temperature and humidity levels were monitored, and the mice were housed under light/dark cycles of LD12:12 h (green LEDs, ~100 lux at the level of the cage floor). Water was provided ad libitum throughout the study.

Mice were fed with round pellets of 300 mg each containing 3.60 Kcal/g (Dustless Precision Pelletsâ, Rodent, Purified, F0075, BioServ, Flemington, NJ, USA). The food composition is similar to regular mouse chow (18.7% protein, 5.6% fat, 59.1% carbohydrates and 4.7% fiber). Food access was controlled by an automated feeder system designed in our lab with Phenome

Technologies Inc., Skokie, IL, USA that precisely controls how much, when, and how often the food is dispensed (Acosta-Rodríguez et al., Cell Metabolism 2017). After six weeks of recording under ad libitum food access, 12 mice/group were randomly assigned to one of 6 feeding conditions (Acosta-Rodriguez et al., under revision): 24h access ad libitum; 70% of baseline ad libitum levels fed at the beginning of the dark (CR-night) or light (CR-day) phase which they consumed in less than 2h (classic CR protocols); or fed under CR evenly spread over 12 hours during the dark (CR-night-12h) or light (CR-day-12h) by releasing one 300 mg pellet every 90 min; or evenly spread over 24 hours (CR-spread) by releasing one 300 mg pellet every 160 min. Mice were weighed during cage change every 21 days, and bedding was checked for any evidence of food spillage. ClockLab Chamber Control Software v3.401 (Actimetrics Inc., Wilmette, IL, USA) was used to schedule and record feeding events, and ClockLab Data Acquisition System v3.209 (Actimetrics Inc., Wilmette, IL, USA) was used to record wheel-running behavior.

Tissue collection and DNA extraction: At 19 months of age, mice were released into constant darkness and samples collected every 4 hours for 48 hours (n=12 mice/condition). Tissues were snap frozen and stored at -80°C until DNA extraction.

Liver tissue was digested using 1mL of digestion buffer (100mM NaCl, 50mM Tris-HCl pH 8, 100mM EDTA pH 8 and 1% SDS) plus 20uL of protein K (20mg/mL) incubated ON at 60ºC. After a RNAse (Ambion, AM2286, 2uL) treatment for 1h at 37ºC, DNA extraction was performed using UltraPure™ Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v, Thermo Fisher, #15593049) and phase lock gel (VWR, # 10847-802) to avoid contamination with interphase, according to manufacturer instructions.

South African species

Ethics: All biological samples from South Africa were either historical samples (collected prior to the formation of the Animal Use and Care Committee of the University of Pretoria), archived post-mortem samples or collected fresh from animals prior to this study in origin. All samples were stored at -80ºC.

The Animal Use and Care Committee of the University of Pretoria evaluated and approved the experimental protocol and collection of all samples (ethics clearance number: NAS022/2021, EC028-07, NAS209/2021, NAS021/2020, EC015-08, EC008-17, NAS018/2020, EC004-11, EC060-12), with DAFF section 20 approval (SDAH-Epi-21051907211, SDAH-Epi-12/11/1/1/8 (2002 LH), SDAH-Epi-20072707050). In addition, permission to capture the various species was obtained from all landowners, and a collecting permit was obtained from the relevant nature conservation authorities (Permit number: CPF6 Western Cape- CN44-87-13780, CN44-31-2285, Northern Cape – FAUNA 0715/2020, Eastern Cape- CRO 54/12CR and CRO 55/12CR, Gauteng- CPF6-0124, Kwa-Zulu Natal- OP1545/2021, Limpopo- CPM-333-00002). Necessary TOPS permits were also acquired for threatened species (Permit number: 68103).

Study animals: Except for *Heterocephalus glaber (derived from wild-caught animals from Kenya), Fukomys damarensis, Atelerix albiventris* and *Echinops telfairi*, which came from captive-raised populations, all other small mammal species were wild-caught in South Africa. All terrestrial small mammal species, *Crocidura cyanea, Elephantulus myurus, E. edwardii, Macroscelides proboscideus, Aethomys namaquensis, Mus minutoides, Rhabdomys pumilio* and *Suncus varilla* were caught using Sherman traps baited with oats and peanut butter. While all small subterranean mammals, *Georychus capensis, Bathyergus suillus, B. janetta, Cryptomys hottentotus hottentotus, C. h. pretoriae, C. h. mahali, C. h. natalenesis* and *Amblysomus hottentotus* were captured using Hickman live traps, baited with a small piece of sweet potato. All

traps were monitored for captures every 2-3 hours over the course of the day and left overnight, being checked first thing in the morning *Smutsia temminckii* that had become entangled and electrocuted on the farm's border electrified fence were collected and stored at -20 degrees until sampled.

Apodemus mice

Apodemus mice were culled during routine (yearly) population control, via cervical dislocation confirmed by exsanguination. All animal work was conducted in accordance with the UK Home Office in compliance with the Animals (Scientific Procedures) Act 1986, was approved by the University of Edinburgh Ethical Review Committee and was carried out under the approved UK Home Office Project License PP4913586.

(*80*)

Prof Thomas Little maintains a formerly-wild, but now lab-reared wood mouse colony in standard laboratory conditions at the University of Edinburgh. The colony has been in captivity for many generations, but the wood mice are purposely outbred to maintain genetic diversity. All mice are housed individually in ventilated cages (Techniplast, 1285L) with food and water ad libitum. DNA was extracted using a phenol-chloroform method from ear punches taken from individuals of known age from the colony. We obtained DNA from 48 mice of both sexes, spanning an age range of 88 to 496 days old. This slightly unusual choice of ages arises because we utilised mice that were part of other experiments. Lifespan in wild mice is not known for certain, and many will die from extrinsic mortality (e.g., predation), but in our own field work we recapture 10-20% of tagged mice the following year (Pedersen, unpublished data), meaning wild mice may can live for hundreds of days.

Spiny mouse

Following anesthesia with 4% (v/v) vaporized isoflurane (Henry Schein Animal Health, Dublin, OH), full thickness ear tissue biopsies were collected from spiny mice (*Acomys cahirinus*) housed in breeding colonies at the University of Kentucky, Lexington, KY or Monash University, Melbourne, Australia. Biopsies were used for ear tissue samples and to produce ear pinna fibroblast cultures. Animal protocols were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Kentucky (2019-3254). Spiny mouse fetal fibroblasts were derived from embryonic skin following enzymatic digestion with trypsin. Connective tissue fibroblasts from adult spiny mice were derived from ear biopsies and grown at 37 °C in 3% oxygen and 5% CO2 as previously described (Saxena et al., 2019).

Nova Scotia masked shrews (Sorex cinereus) (*22*)

Sampling, sexing, and aging shrews

Masked shrew samples were collected from five locations: Peterborough County, Ontario and Sandy Cove, North Mountain, Long Island and Bon Portage Island (BPI) in Nova Scotia, Canada. All shrews were trapped within six weeks of each other to limit seasonal morphological changes, such as skull and body size. DNA was extracted from liver, tail, and fetus using the DNeasy Blood & Tissue Kit from QIAGEN. Shrews were left in dermestid beetle tanks for a week to remove all remaining tissue from the skulls. A Leica EZ4 microscope was used to assess age class for each shrew based on teeth wear according to Pruitt's method. Age in months was estimated based on the trapping date, the age class, and the known reproductive season of masked shrews assuming a

maximum lifespan of ~17 months. Shrews were sexed by PCR amplification of the SRY gene. Gel electrophoresis was used to identify successful amplification and thus confirm the presence of the Y chromosome (i.e. male). Age class designations were confirmed based on morphological features; sex assignments were subsequently validated with the methylation assay.

Shrews and small animals from Museum of Biological Diversity at The Ohio State University.

Ethics statement

All wild-caught animals were collected and sacrificed in accordance with protocols approved by The Ohio State University IACUC (Institutional Animal Care and Use Committee) under protocol number 2017A00000036. All wild-caught animals were collected with scientific collecting permits issued from Ohio and Washington and according to guidelines established by the American Society of Mammalogy for the use of wild animals in research (Sikes & Animal Care and Use Committee of the American Society of Mammalogists 2016).

Tissue samples were taken from archived specimens that are in the process of being deposited in the Museum of Biological Diversity at The Ohio State University.

Tissue samples were acquired from wild animals using multiple collecting procedures, including (1) trapping live animals using Sherman live-traps or pitfalls and followed by euthanasia using isoflurane overdose, (2) sampling of roadkill animals, and (3) sampling of deceased animals received from wildlife rehabilitation centers. Abdominal and pelvic organs were removed from animals and archived in a -80C freezer.

Marsupials (*16*)

Opossum and mouse tissue samples

Opossum samples come from a pedigreed, breeding colony of gray short-tailed opossums (*Monodelphis domestica*) that was established by founder individuals purchased from the Southwest Foundation for Biomedical Research. Mouse samples come from a breeding colony of C57BL/6J mice originally purchased from the Jackson Laboratory. Both colonies are maintained by the Sears Lab at UCLA. Opossums were euthanized by CO2 inhalation to effect followed by bilateral thoracotomy. Mice were euthanized by CO2 inhalation to effect followed by cervical dislocation. These procedures are in accordance with the AVMA Guidelines for the Euthanasia of Animals 2013: https://www.avma.org/KB/Policies/Documents/euthanasia.pdf, and all animal procedures were approved by the UCLA IACUC. All tissue samples, e.g., liver, blood, tail, were taken from euthanized opossums and mice and stored at -20 C until use. DNA from samples was extracted and purified using a DNA Miniprep Plus Kit (Zymo), following manufacturer's protocols. PicoGreen fluorescent dsDNA was used to assess the concentration of resulting DNA and concentrations adjusted to 50 to 250 ng/ul. DNA samples were submitted to the Technology Center for Genomics & Bioinformatics at UCLA for generation of DNA methylation data and further analyses.

Tasmanian devil tissue samples (*16*)

Ear samples from Tasmanian devil (Sarcophilus harrisii) were collected from individuals in the Tasmanian devil insurance metapopulation, either living as an introduced population on Maria Island (N=29), or in the zoo-based population (N=17). Samples were collected by the Save the Tasmanian Devil Program (STDP), or the respective zoos, under the STDP's Standard Operating Procedure: Trapping and handling wild Tasmanian devils and shared with the University of

Sydney. The samples were collected for standard management practice, which received ethics approval over 15 years ago. DNA from these samples were made available for use in this study, no individuals were specifically sampled for the purposes of this research.

Tasmanian devils within the zoo-based insurance population are housed in a range of scenarios from intensive housing as individuals, or as pairs in breeding season; or in large group housed enclosures, ranging from 4-10 males and 4-10 females per enclosure. Tasmanian devils are not known to be a social species and housing within the Tasmanian devil insurance population is under the management of the Zoo and Aquarium Association Australasia. Devils are housed and fed a variety of diet items and managed according to the ZAA Husbandry Guidelines for the species. Devils living on Maria Island are wild devils and so there is no animal care or maintenance for these individuals other than twice yearly monitoring trapping trips.

Samples were selected for this study to include a range of known age individuals and sexes. DNA was extracted using either a modified phenol-chloroform protocol or the MagAttract HMW DNA kit (Qiagen, Germany; cat: 67563). DNA concentration and quality were assessed using a Nanodrop 2000 Spectrophotometer (ThermoFisher Scientific) and 0.8% agarose gel electrophoresis for 30 minutes at 90V.

Kangaroo and wallaby tissue samples (*16*)

The blood samples for kangaroo and wallaby were opportunistically collected from zoo-based animals during routine health exams. We analyzed blood from the following species: Macropus rufus (red kangaroo), Macropus giganteus (Eastern grey kangaroo), Macropus fuliginosus (Western grey kangaroo), and Macropus rufogriseus (red-necked wallaby).

Mammalian liver samples (Diego Villar Lozano and Duncan Odom)

Ethics

The use of all animals in this study was approved by the Animal Welfare and Ethics Review Board, under reference number NRWF-DO-02vs, and followed the Cancer Research UK Cambridge Institute guidelines for the use of animals in experimental studies. Tissue samples from humans were obtained from the Addenbrooke's Hospital at the University of Cambridge, under license number 08-H0308-117 ''Liver specific transcriptional regulation''.

Methods

We obtained mammalian tissue samples from routine euthanasia procedures (e.g., macaque, marmoset, rabbit, cat, dog, horse, and opossum), commercial providers or abattoirs (e.g., cow, pig, ferret, and guinea pig), and research tissue banks from specialty conservation programmes (e.g., cetaceans, Damaraland mole-rat and tissue samples from zoo post-mortem examinations). Tissue sources of these samples have been reported in previous publications (*2, 81, 82*). In most cases, tissues were obtained immediately post-mortem (typically within an hour) to maximize experimental quality. Tissues were kept on ice until processed to minimize potential loss of integrity during post-mortem time, and flash-frozen in dry ice. 20-50 mg of tissue was used for extraction of genomic DNA with standard commercial kits (Qiagen), and genomic DNA samples were plated in 96-well plates using a randomized layout obtained with the R package OSAT (*83*).

Mammalian liver samples from the University of Rochester

All experiments were performed according to procedures approved by the University of Rochester Committee on Animal Resources (UCAR). Animal protocol # 101939 / UCAR-2017-033

The tissues were from the Gorbunova and Seluanov tissue bank at the University of Rochester (*84*).

Human Cohorts

Framingham Heart Study Cohort (FHS)

The FHS cohort(*85*) is a large-scale longitudinal study started in 1948, initially investigating the common factors of characteristics that contribute to cardiovascular disease (CVD), https://www.framinghamheartstudy.org/index.php. The study at first enrolled participants living in the town of Framingham, Massachusetts, who were free of overt symptoms of CVD, heart attack or stroke at enrollment. In 1971, the study started FHS Offspring Cohort to enroll a second generation of the original participants' adult children and their spouses (n= 5124) for conducting similar (*86*). Participants from the FHS Offspring Cohort were eligible for our study if they attended eighth examination cycles and consented to having their molecular data used for study. We used the 2,544 participants with available DNA methylation profiles (measured at exam 8), from the group of Health/Medical/Biomedical (IRB, MDS) consent. The FHS data are available in dbGaP (accession number: phs000363.v16.p10 and phs000724.v2.p9). Deaths among the FHS participants that occurred prior to January 1, 2013 were ascertained using multiple strategies, including routine contact with participants for health history updates, surveillance at the local hospital and in obituaries of the local newspaper, and queries to the National Death Index. Death certificates, hospital and nursing home records prior to death, and autopsy reports were requested. When cause of death was undeterminable, the next of kin were interviewed. The date and cause of death were reviewed by an endpoint panel of 3 investigators.

DNA methylation quantification Peripheral blood samples were collected at the 8th examination. Genomic DNA was extracted from the buffy coat using the Gentra Puregene DNA extraction kit (Qiagen) and bisulfite converted using EZ DNA Methylation kit (Zymo Research Corporation). DNA methylation quantification was conducted in two laboratory batches using the Illumina Infinium HumanMethylation450 array (Illumina). Methylation beta values were generated using the Bioconductor *minfi* package with Noob background (*87*).

Women's Health Initiative (WHI)

The Women's Health Initiative is a national landmark study that enrolled postmenopausal women aged 50-79 years into the clinical trials (CT) or observational study (OS) cohorts between 1993 and 1998(*88, 89*). We included 2,017 WHI participants from "*Broad Agency Award 23*" (WHI BA23) with available phenotype and DNA methylation array data: 2,107 women from "*Broad Agency Award 23*" (WHI BA23). WHI BA23 focuses on identifying miRNA and genomic biomarkers of coronary heart disease (CHD), integrating the biomarkers into diagnostic and prognostic predictors of CHD and other related phenotypes. The study spans three WHI sub-cohorts including GARNET, WHIMS and SHARe.

DNA methylation quantification for BA23: In brief, bisulfite conversion using the Zymo EZ DNA Methylation Kit (Zymo Research, Orange, CA, USA) as well as subsequent hybridization of the HumanMethylation450k Bead Chip (Illumina, San Diego, CA), and scanning (iScan, Illumina) were performed according to the manufacturers protocols by applying standard settings. DNA methylation levels (β values) were determined by calculating the ratio of intensities between methylated (signal A) and un-methylated (signal B) sites. Specifically, the β value was calculated from the intensity of the methylated (M corresponding to signal A) and un-methylated (U corresponding to signal B) sites, as the ratio of fluorescent signals β =

Max(M,0)/[Max(M,0)+Max(U,0)+100]. Thus, β values range from 0 (completely un-methylated) to 1 (completely methylated).

Supplementary Text

Variability in sequence conservation does not explain phyloepigenetic relationships

To verify whether the ability to accurately ascertain phylogenetic relationships between species based on methylation levels of conserved CpGs is indeed due to differences in methylation levels as opposed to confounding by sequence conservation (*5*), we performed three sensitivity analyses in our dataset. First, we grouped the 14,705 eutherian probes by their conservation across species, and ascertained if this would change the phyloepigenetic-phylogenetic tree correlations (congruence). Out of the set of probes whose sequences are present in (mapped to) at least n=150 mammalian species, we selected a random set of 100 CpGs from which we constructed a phyloepigenetic tree, and then determined the Congruence(150), i.e. the Pearson correlation between distances of the two trees. To examine the effect of reduced sequence conservation, we repeated this analysis with phyloepigenetic trees that were constructed from 100 random CpG probes mappable to lower numbers of species (ranging from n=50 to n=150 mammalian species). Strikingly, n correlated strongly with Congruence(n) (r=0.77, p=0.0008, **fig. S2C**). This shows that methylation levels from highly conserved CpGs lead to a higher congruence between the trees than when using less conserved CpGs.

In a second sensitivity analysis, we binarized the CpGs based on alignment of probes to different mammalian genomes, and generated phyloepigenetic trees that are solely based on the presence/absence of a measurable CpG in the genome (**fig. S2d**). These mappability-based phyloepigenetic trees exhibited a moderate Congruence=0.59 when all 37,492 CpGs on the array were used, but a weaker correlation (Congruence=0.2–0.3) when only the conserved eutherian or marsupial probes or monotreme probes were considered (fig. S2d, Methods). A direct comparison of the mappability trees with the blood phyloepigenetic trees show a weak correlation (Congruence=0.32) even after all 37,492 CpGs are used to create the mappability tree (**fig. S2E**).

In the third sensitivity analysis, we used the detection pvalues of the CpGs and selected 180 CpGs that had detection pvalue<0.01 in all 384 species. The mappability tree from these 180 CpGs show no congruence with traditional phylogenetic (**fig. S2F**, congruence=0.08), and blood phyloepigenetic trees (**fig. S2F,G**, congruence= - 0.02). However, phyloepigenetic trees from these 180 CpGs still show a high congruence with traditional phylogenetic in all tissues (congruence = 0.41 – 0.71 depending on tissue, **fig. S2F, H**)

Overall, all sensitivity analyses demonstrated that methylation level of specific cytosines is the main determinant of the inter-species relationships that constitute the phyloepigenetic tree and illustrates that the high Congruence=0.90 in the blood phylogenetic tree is not directly driven by sequence features (**Fig. 1**). The close relationship between phylogenetic and phyloepigenetic trees reflects an intertwined evolution of the genome and epigenome that mediates the biological characteristics and traits of different mammalian species.

CpGs with phylogenetic signal

To identify CpGs exhibiting a strong phylogenetic signal between methylation and phylogenetic trees, we employed the K statistic as described by Blomberg et al. (2003) (*32*). We used the phylosignal() function is as a part of the picante package in R to calculate the K statistic of phylogenetic signal as well as P-value based on variance of phylogenetically independent contrasts relative to tip shuffling randomization.

The K statistic, introduced by Blomberg et al. (2003), is a measure used to quantify the phylogenetic signal of a continuous trait in a group of related species. The phylogenetic signal represents the degree to which closely related species resemble each other more than expected by chance alone. The K statistic compares the observed variance in the trait among species to the variance expected under a Brownian motion model of trait evolution (a random walk model) across the phylogenetic tree.

A K value of 1 indicates that the observed trait variation is consistent with the Brownian motion model, suggesting a strong phylogenetic signal. If the K value is greater than 1, it indicates that closely related species are even more similar than expected under the Brownian motion model, whereas a K value less than 1 implies that related species are less similar than expected, suggesting a weak phylogenetic signal. In summary, the K statistic provides a quantitative measure to assess the influence of phylogenetic relationships on the distribution of continuous traits among species.

The top 500 CpGs with significant phylogenetic signals (nominal Blomberg $p < 0.001$, additional selection by variance z score) were found to be enriched in upstream intergenic regions in analyzed mammalian tissues (**Fig. 1D, fig. S4**). Conversely, we observed a depletion of phylogenetic signals in promoters, particularly in liver samples (**fig. S4**). We further splitted the data into groups of 10 CpGs relative to the transcriptional start site (TSS) to find the regions with the strongest phylogenetic signal. This analysis also confirmed that regions with strong phylogenetic signals are located in intergenic regions, gene bodies (OR>3, fisher exact $p<0.05$), but not the promoter regions (**fig. S4D**). As a sensitivity test, we constructed phyloepigenetic trees using only the CpGs with a significant Blomberg K statistic ($p < 0.001$, variance z-score < -2.5), which resulted in a high degree of congruence in phyloepigenetic trees (Congruence = 0.97, **fig. S4**).

<u>In depth characterization of modules</u>

DNAm modules are correlated with mRNA across horse tissues

We employed canonical correlation analysis to detect potential correlations between co-methylation modules and mRNA levels for 29 horse tissues for which we have data (*27*). We found that 23 modules had a significant ($p<0.05$) canonical correlation with mRNA levels (**fig. S10A**). The strongest correlation was exhibited by the plum2 module ($r=0.84$, $p<2x10^{-16}$), which contains genes such as *FAM27B*, *YWHASE*, *FRG1BP* and *KLF7*. Detailed results from the canonical correlation analysis are provided in **table S3** and **table S6.**

<u>Modules that relate to taxonomic orders</u>

Modules that relate to specific taxonomic orders, or even species, are expected to be valuable starting points for experimental interrogation of mammalian evolution. For example, two modules were strongly associated with primates in positive and negative directions: Primate+ (plum2, increased methylation in primates) and Primate- (white, decreased methylation in primates) (**fig. S11A**). The Primate+ (plum2) module could even be considered as a human+ module because the eigengene takes on higher values in human tissues than in non-human primate tissues (**fig. S11B, table S3**).

A functional enrichment study revealed that the Primate- (white) module is involved in ERK/MAPK signaling (**Fig. 2C, table S3**).

Unknown modules

Some of the modules with no obvious relationship to specific traits were nevertheless found by functional enrichment studies to be related to ubiquitous cellular processes. As a case in point, the yellow module with 895 CpGs is the largest consensus module that is preserved in all mammalian tissues, but it did not relate to any of the primary traits examined (**Fig. 2A**).

Clues to its biological relevance could be found in a murine study of sleep disruption. Only the yellow module eigengene exhibited a significant positive association (p=0.02, **fig. S7B**) with sleep disruption in liver samples from mice that had been subjected to 12 months of light pollution at night (Methods). Interestingly, the yellow module has a CpG adjacent to the CLOCK gene (**table S2**).

Modules that relate to tissue type

Our mammalian methylation dataset is in effect an atlas of 70 tissue-types, cell-types, and tissue regions from a wide age range of mammalian species. This allowed us to create a tissue module atlas (Fig4c). DNA methylation could distinguish among tissues such as blood, skin, liver, muscle, cerebellum, cerebral cortex, and brain. The tissue modules were largely preserved in individual taxonomic orders, including marsupials and monotremes (**Fig. 4c**, **table S3, fig. S6B**). For example, the top blood modules in different taxonomic orders were Blood(+) (magenta) and Blood(-) (tan) modules.

Identification of consensus modules across mammalian tissues

To identify the most conserved modules across species and tissue types, we developed an additional seven consensus networks (cNet3-cNet9) (see Methods). Each consensus network can be interpreted as a meta-analysis of methylation networks (**fig. S4**). cNet3 is the consensus (conserved) network between 57 WGCNA networks developed in different species+tissue combinations (e.g., human-blood, mouse-blood) with at least 50 samples and hence represents the most conserved modules in all mammalian tissues. cNet4 is based on 35 species but ignores tissues. By contrast, cNet5 ignored species and formed consensus networks across 15 tissue types for which at least 50 samples were available. cNet6-cNet9 are tissue-specific consensus networks for different species. Although these supervised consensus networks cannot be used to identify modules that relate to differences between species and tissues, they are informative for studies of sex and chronological age, as elaborated below. These additional networks could identify Net1 modules that can be considered mammalian consensus modules due to high preservation in all mammalian tissues (**Fig. 2A**). Some of these consensus modules included the purple, lightsteelblue1, and mediumpurple3 modules, which will be discussed in the following sections.

Consensus modules that relate to sex

Sex was strongly associated with two modules: Female(+) and Female(-) (lightsteelblue1 and mediumpurple3 modules in Net1 and matching colors in the consensus networks, respectively) (**Fig. 4A**). Unsurprisingly, these modules mostly consist of CpGs that are located on human and mouse X-chromosomes (**Fig. 4B**). The only autosomal CpG in lightsteelblue1 modules is located on the POU3F2 exon in human chromosome 6 and mouse chromosome 4. POU3F2 is reported to contribute to maternal behavior differences in mammals (*90*). This gene was among the 11 autosomal genes that was identified by EWAS of sex in 30 primates (*57*).

Stratified analysis of lifespan and weight

Although the correlation between adult weight and maximum lifespan across the 348 species in our data was strong (Pearson R=0.73, p<2.2e-16, **fig. S12**), it was less so in some taxonomic

orders such as Rodentia (R=0.55, p=7e-9) and Chiroptera (r=0.38, p=0.01), while particularly strong in others e.g., Carnivora (R=0.82, p=2e-7) and Eulipotyphla (R=0.83, p=2e-7, **fig. S12**). A true lifespan-related module should exhibit a significant correlation with maximum lifespan even when the analysis is restricted to species of a given taxonomic order. Ideally, the significant correlation between the module eigengene and lifespan can be observed in several taxonomic orders even in orders with a weak correlation between lifespan and adult weight. The magenta module correlates positively with maximum lifespan in two taxonomic orders (r=0.4, p=5.9e-5 in Rodentia, and r=0.37, p=1.2e-04 in liver samples from Eulipothyphla**, table S3**). By contrast, the magenta module exhibits weaker or even negative correlation with adult weight in different taxonomic orders (non-significant correlations with weight in Rodentia r=0.30, p=0.002 and Eulipotyphla r=-0.38, p=0.053, **table S3**). There was an unexpected negative correlation between the magenta module and maximum lifespan in Artiodactyla (r=-0.60, p=2.2e-6), which mirrors the negative correlation between lifespan and average weight of different dog breeds as detailed in the following.

Life expectancy in pure dog breeds

The relationship between weight and lifespan is strong and yet paradoxical. Across species, a positive correlation between average weight and maximum lifespan is observed, but within species, it is typically the opposite (*91, 92*). In dogs, breed weight is negatively correlated with lifespan (*91, 92*), as is also observed across the 93 dog breeds in our study (r= -0.58) (*29*). We used 756 blood samples from 93 different dog breeds to correlate the module eigengenes with breed lifespan and breed weight. One of the top lifespan-related modules, the magenta module, exhibited the expected relationship within samples from this species: r= -0.30, p=0.003 for breed lifespan and r=0.28, p=0.006 for breed weight (**Fig. 3C**). The opposite associations of the magenta module with lifespan and weight across species and within species mirrors the above-mentioned paradoxical association between weight and lifespan. The dog methylation data allowed us to find additional modules that correlate with breed lifespan and breed weight. The skyblue3 module eigengene exhibited a significant correlation with dog breed lifespan (skyblue3, r= -0.39, $p=1x10^{-4}$) and a positive correlation with breed weight (skyblue3, r= 0.32, p=0.001) (**Fig. 3C**). The skyblue3 module is noteworthy since it also exhibits a consistent negative correlation with maximum lifespan across species (skyblue3, r=-0.26, $p=1.04x10^{-6}$, **table S3**).

Mortality in human cohorts

In total, a subset of 7521 and 5558 mammalian probes overlaps with the human Illumina EPIC and Illumina 450k beadchip array platforms, respectively (*5*). We could use these overlapping probes to calculate the eigengenes of the mammalian modules in the two human epidemiological cohort studies: the Framingham Heart Study (FHS) Offspring Cohort and the Women's Health Initiative (WHI). A Cox regression analysis allowed us to calculate hazard ratios for time to death due to all-cause mortality for each of the module eigengenes. The magenta ($Lifespan^{(+)}Blood^{(+)}$) module exhibited the strongest correlation with human mortality risk. One standard deviation of the magenta module eigengene was associated with a hazard ratio HR= 0.91 (MetaP=0.0016, **Fig. 3D**), which indicates that high methylation values in the magenta module are associated with lower mortality risk in humans.

In depth description of EWAS of maximum lifespan

EWAS of maximum lifespan

We carried out epigenome-wide association studies (EWAS) to relate the methylation levels of individual CpGs to the various life history traits. To reduce biases resulting from different levels of sequence conservation, our EWAS of life history traits focused on n = 333 eutherian species, excluding Marsupial species. We performed four types of EWAS analyses adjusting for different confounders: (1) Lifespan; a direct regression analysis of lifespan (generic EWAS). (2) Weight-adjusted lifespan (AdjWeight); a regression analysis of maximum lifespan after adjustment for adult weight, which identifies lifespan-related CpGs that are independent of the body mass of the species. (3) Phylogenetic-adjusted lifespan (AdjPhylo); a phylogenetic regression model (23) of lifespan, which adjusts for evolutionary relationships between species. (4) Phylogeny and Weight-adjusted lifespan (AdjPhyloWeight); a phylogenetic regression of lifespan after adjustment for average adult species weight. The results of these four categories of EWAS can be found in **table S12-S15**. For brevity, we will focus on categories 1 and 2 since categories 3 and 4 led to qualitatively similar conclusions (**table S12-S15**).

Each analysis category is further subdivided by tissue type. The all-tissue analysis (denoted "All", **fig. S16**) ignored tissue type. Within species, mean methylation levels are highly correlated across tissue types (R > 0.95), but the all-tissue analysis may miss longevity mechanisms that are specific to tissues and organs. Therefore, we also present EWAS for five tissues for which there were a sufficiently large number of samples: blood (n = 139 species), skin (n = 132), liver (n = 154), muscle (n = 46) and brain (n = 34). We observed positive pairwise correlations between the all-tissue EWAS results and those of tissue specific EWAS (**fig. S16**): such as blood (Pearson correlation R = 0.76), skin (R = 0.69), liver (R = 0.69), muscle (R = 0.49), and brain (R = 0.38). All tissue, Blood and skin lifespan EWAS are summarized in **fig. S17 - S19**. To assess the robustness of maximum lifespan EWAS, we observed high agreements, in most tissues, between our generic EWAS (category 1) and a separate maximum lifespan EWAS using only samples obtained from animals that were younger than their species' average age of sexual maturity and younger than 5 (**fig. S21**).

We identified the genes that are proximal to CpGs that are statistically most correlated with maximum lifespan. These are as follows: lifespan was positively-correlated with a CpG in the distal intergenic region neighboring TLE4 (Pearson R = 0.68, P = $5.9x10^{-46}$, **fig. S17a**) and two CpGs near the promoter region of HOXA4 (R = 0.67), and negatively-correlated with a CpG in an intron of GATA3 (R = -0.65, P = $8x10^{-42}$), exon in ZBTB7B (R = -6.1, P = $6x10^{-35}$), and the promoter region of CELF2 (R = -0.6, P = $1x10^{-34}$).

Many of these significant CpGs remain so after phylogenetic adjustment, such as the CpGs neighboring TLE4, HOXA4, PKNOX2, LMX1B, C15orf41, and ZEB2 (P = $2x10^{-5}$, P = $2x10^{-3}$, P = $3x10^{-4}$, P =$x10^{-4}$, P = $9x10^{-4}$, P = $1x10^{-}$6, respectively, **fig. S17A** and **table S14**). Phylogenetic EWAS (category 3 analysis) top CpGs are reported in the form of phylogenetic independent contrast (**fig. S27**), and generic EWAS and phylogenetic EWAS Z statistics agreements are summarized in **fig. S29**.

All the top-ranking CpGs mentioned above from the category 1 analysis remain in the top 500, in both directions, of weight adjusted EWAS (category 2 analysis) (**fig. S17d**), which indicates that these CpGs do not reflect confounding by body mass. But adjustment for adult weight (category 2) leads to a different set of top ranking CpGs: the top positively lifespan-related CpGs are in a promoter of PKNOX2 (R = 5.4, P = $5.2x10^{-27}$) and an intron of LMX1B (R = 0.51, P = $4.0x10^{-24}$) and the top CpGs negatively related to lifespan are in an intron of C15orf41 (R = -0.55, P = $8.0x10^{-27}$) and an intron of ZEB2 (R = -0.5, P = $5.1x10^{-23}$).

<u>Gene set enrichment analysis of maximum lifespan</u>

To uncover biological processes potentially linked to lifespan-related CpGs, we identified functional annotations associated with genes proximal to lifespan-related CpGs using the Genomic Regions Enrichment of Annotations Tool (GREAT) (*54*). GREAT automatically adjusts for biases arising from the array platform and biases of uneven coverage of genes.

The number of significant lifespan related CpGs per tissue type depends on the underlying sample size (number of species). We imposed an upper limit of 500 on the number of significant CpGs and referred to the top 500 CpGs with a positive and negative correlation with lifespan as lifespan.pos set and lifespan.neg set, respectively. These CpGs are further subject to a Bonferroni corrected significance threshold (P < 1.8e-6) before the enrichment analysis. Detailed results can be found in (**table S16**, and **fig. S22-S24**).

CpGs that have a positive correlation with maximum lifespan implicate genes that play a critical role in development including the HOXL gene group (GREAT P = 1.2x10-5, **fig. S22**) based on the following genes EVX1, HOXA2, HOXA3, HOXA4, HOXA5, HOXB1, HOXB2, HOXB3, HOXB4, HOXB7, HOXB8, HOXB9, HOXC4, HOXD10, HOXD8, HOXD9) (**table S16**). More significant enrichment for HOXL genes were obtained after adjusting the analysis for adult weight (reporting False Discovery Rate p-values as FDR) (GREAT FDR = 1.3x10-15). The EWAS of lifespan implicated embryonic organ morphogenesis with (category 2) or without (category 1) adjusting for adult weight (generic EWAS, GREAT FDR = 3.4x10-4, weight-adjusted EWAS, GREAT FDR = 2.5x10-7) and multicellular organism development (generic EWAS, GREAT FDR = 9x10-4, weight-adjusted EWAS, GREAT FDR = 4.4x10-5). Developmental pathways are even more enriched in skin samples, such as embryonic organ morphogenesis in generic EWAS (GREAT FDR = 2.7x10-30) and embryonic organ development in weight-adjusted EWAS (GREAT FDR = 1.4x10-15).

CpGs that are positively related to weight-adjusted maximum lifespan are enriched with genes involved in mouse phenotypes such as abnormal survival (GREAT P = 4.1x10-4) and mortality/aging (GREAT P = 7.2x10-5) (**fig. S22**, **table S16**).

The GREAT enrichment analysis revealed that CpGs negatively related to lifespan are located next to genes that play a role in abnormal eye morphology according to mouse knockout studies (GREAT FDR = 2.3x10-4), regulation of gene expression (GREAT FDR = 2.1x10-5) and DNA-templated regulation of transcription (GREAT P=4.0x10-5).

Both negatively and positively lifespan related CpGs are located near genes that play different roles in mRNA processing (**table S18**) and splicing including (CELF1; CELF2; CELF6; DAZAP1; FAM172A; HNRNPA1; HNRNPK; HNRNPU; JMJD6; MBNL1; MBNL2; NOVA2; QKI; RBFOX1; RBM15; RBM39; SF1; SON; SRPK1; SRPK2; SRSF12; TRA2A; TRA2B; YTHDC1.

Our transcription factor analysis based on GREAT demonstrates that CpGs positively related to weight adjusted maximum lifespan are located near binding sites of transcription factors HOXA4 (GREAT FDR = 1.8x10-10, **fig. S22**, **table S16**), GATA6 (GREAT FDR = 4.2x10-11), EVI1 (GREAT FDR = 0.001) while CpGs negatively related to lifespan are located near binding sites of transcription factor ER (GERAT FDR = 1.1x10-5 estrogen receptor) and IK3 (GREAT FDR = 2.7x10-4).

Overlap with human GWAS results

Genes next to CpGs with a significant positive correlation with weight adjusted mammalian maximum lifespan (category 2 analysis) exhibited significant overlap with gene sets identified by human genome-wide association studies (GWAS) of human birth length (FDR = 4.1x10-10). Genes next to CpGs with significant positive correlation with maximum lifespan (category 1

analysis) overlapped with gene sets implicated by GWAS of the human waist circumference in all ancestries (FDR = 8.3x10-4, **fig. S22**). Details of the enrichment terms are reported in **table S20**.

Exons and CpG islands

We observed that lifespan related CpGs could be found in genic and intergenic regions with methylation changes in both positive and negative directions.

We find that exons, which are supposed to exhibit high methylation to protect against cryptic transcription, exhibit high methylation levels in long lived species ($P < 10\text{-}4$, **fig. S18B**). Similarly, CpG islands, which are supposed to exhibit low methylation to maintain high expression levels of housekeeping genes (*93*), exhibit low methylation levels in long lived species (**fig. S18B-C**, Odds ratio = 3.70, $P < 0.0001$). Overall, these results suggest that the cells of long living mammalian species exhibit higher fidelity in methylation landscapes. Similar observations have been made in mouse strains with different lifespans (*12*).

Upstream regulator analysis using Ingenuity pathway analysis

We used the set of conserved CpGs on the mammalian array as background in all our enrichment analyses including the upstream regulator analysis. Thus, our results control for general sequence conservation. This can also be seen from our control analysis where we analyzed the upstream regulators of the 500 most highly conserved CpGs. The resulting gene list was not enriched for pluripotency factors (**fig. S25**). In other words, the enrichment for SOX2-OCT4-NANOG is specific to lifespan related CpGs.

Chromatin context of CpGs that are negatively associated with lifespan

To elucidate the genomic context of maximum lifespan-related cytosines, we related them to a universal chromatin state annotation that is based on chromatin marks from over 100 human cell and tissue types (*41*). The annotation is shared across cell and tissue types, and consists of 100 chromatin states that correspond to constitutive and cell-type-specific activities, with detailed descriptions about these states are available in **table S21**. The mean methylation levels of CpGs inside these chromatin states is remarkably conserved across all taxonomic orders (**fig. S30**).

CpGs with a significant negative correlation with maximum lifespan (top 500 per direction with $P < 1.8x10^{-6}$ Bonferroni corrected p-value threshold) are located near regions annotated by chromatin states associated with transcriptional start sites (TSS1, $P = 2.4x10^{-12}$, **Fig. 5D**, **table S22**) and flanking promoter regions (PromF4, $P = 5x10^{-}10$ and PromF5, $P = 1.9x10^{-9}$). These three chromatin states are associated with constitutively active promoter activities across most cell types, and they are lowly methylated in most mammalian species (**fig. S30**).

We observed significant enrichments of CpGs that are inversely related with lifespan in regions bound by both polycomb repressive complex 1 and complex 2 (PRC1, $P = 6x10^{-}11$ and PRC2, $P = 2.4x10^{-6}$). Unlike PRC1 target sites, PRC2 target sites harbor CpGs that correlate (positively) with the relative age associated module in mammalian species (**Fig. 7**).

The chromatin states cluster together under a hierarchical clustering dendrogram (**fig. S31**) derived from the emission probability profile of chromatin states' hidden Markov model (HMM) (*41*). We observe that top ranked CpGs from generic lifespan EWAS in the states are enriched for CpGs inversely associated with log maximum lifespan (**fig. S31** leftmost columns). This shows that CpGs that are proximal to transcription start sites and their immediate flanks share a strong inverse association with lifespan.

Additionally, two active enhancer states (EnhA3, and EnhA9) also exhibit low methylation in long lived species. EnhA3 corresponds to strong enhancers in most cells while EnhA9 is particularly pronounced in blood cells (*41*). These findings suggest that the cells of long-lived species exhibit low methylation in select enhancers that maintain the identity of blood cells.

An enrichment study of individual histone marks indicates that CpGs inversely related to lifespan are marked by H3K27ac ($P = 2.9x10^{-42}$), H3K4me2 ($P = 2.6x10^{-34}$), H3K4me3 ($P = 5.6x10^{-30}$), H3K9ac ($P = 7x10^{-27}$), H2A.Z ($P = 6.4x10^{-20}$) based on ENCODE ChIPseq data of different human tissues (**table S23**). These results indicate that promoters exhibit low methylation in long lived species.

Chromatin context of CpGs that are positively associated with lifespan

CpGs that are positively correlated with lifespan are found to be enriched in a different set of chromatin states (**Fig. 7, table S22**). These CpG are enriched in active enhancer chromatin states EnhA2 (active in select cell types, **table S21**) ($P = 2.8x10^{-4}$), and and EnhA17 ($P = 3.7x10^{-4}$), which is active specifically in embryonic stem cells (ESC) and induced pluripotent stem cells (iPSC) (**Fig. 7**). Other states that are also enriched with lifespan-positively-correlated CpGs correspond to bivalent regulatory regions specifically in a subset of cell and tissue types, including blood and ESC, BivProm3 and BivProm4 ($P = 4.4x10^{-4}$, $P = 1.6x10^{-}5$) (**Fig. 7**) (*41*). Among the top 10 CpG sites in the lifespan generic EWAS, all four CpG sites proximal to the HOX gene family (three CpGs neighboring HOXA4 and one CpG neighboring HOXB3) are in BivProm4 state (**table S12**). BivProm3-4 did not show enrichment for the relative age associated module. CpGs that are positively correlated with lifespan are also enriched for transcription associated states TxEnh8 ($P = 4x10^{-5}$), TxEx1 ($P = 3x10^{-4}$), and common highly methylated domains ($P = 1x10^{-4}$) (**Fig. 7**).

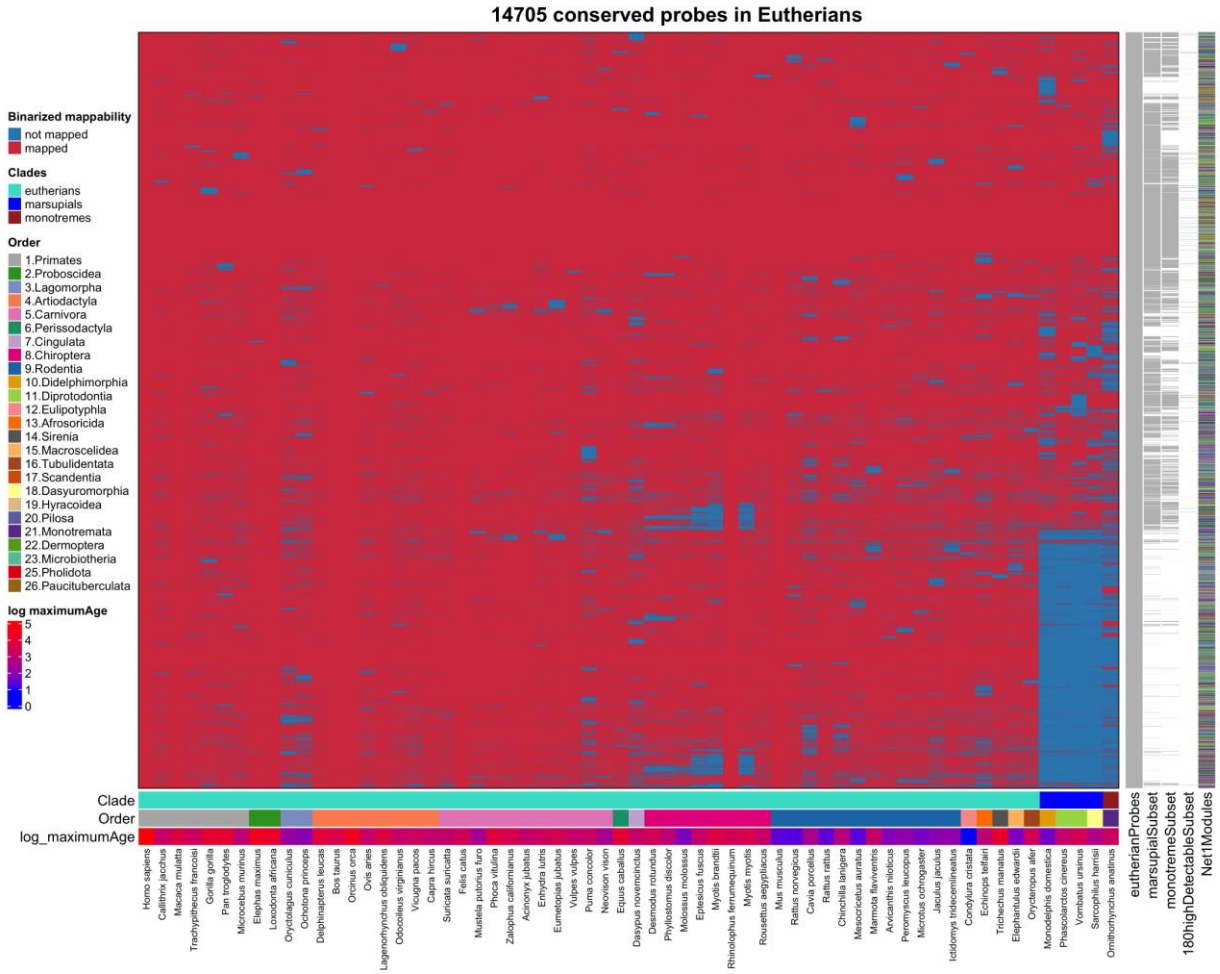


Fig. S1. Conserved CpGs in eutherians and mammals. The heatmap shows the a subset of highly conserved CpGs (14,705 out of 37,492 CpGs on mammalian array 40k) that were selected based on alignment of probes to 12 mammalian species from different taxonomic orders for the current study. These species included human (hg19), mouse (mm10), vervet monkey (ChlSab1.1.100), rhesus macaque (Mmul_10.100), cattle (ARS-UCD1.2), cat (Felis_catus_9.0.100), dog (CanFam3.1), African elephant (loxAfr3.100), bat (Rhinolophus_ferrumequinum.HLrhiFer5), killer whale (GCF_000331955.2_Oorc_1.1), opossum (Monodelphis_domestica.ASM229v1.100), and platypus (mOrnAna1.p.v1.100). The heatmap is ordered by hierarchical clustering (1-cor, average linkage) of rows, and mammalian clades for columns. The rows are also annotated with mammalian co-methylation colors to show that the mappability clustering (in rows) does not relate to co-methylation modules.

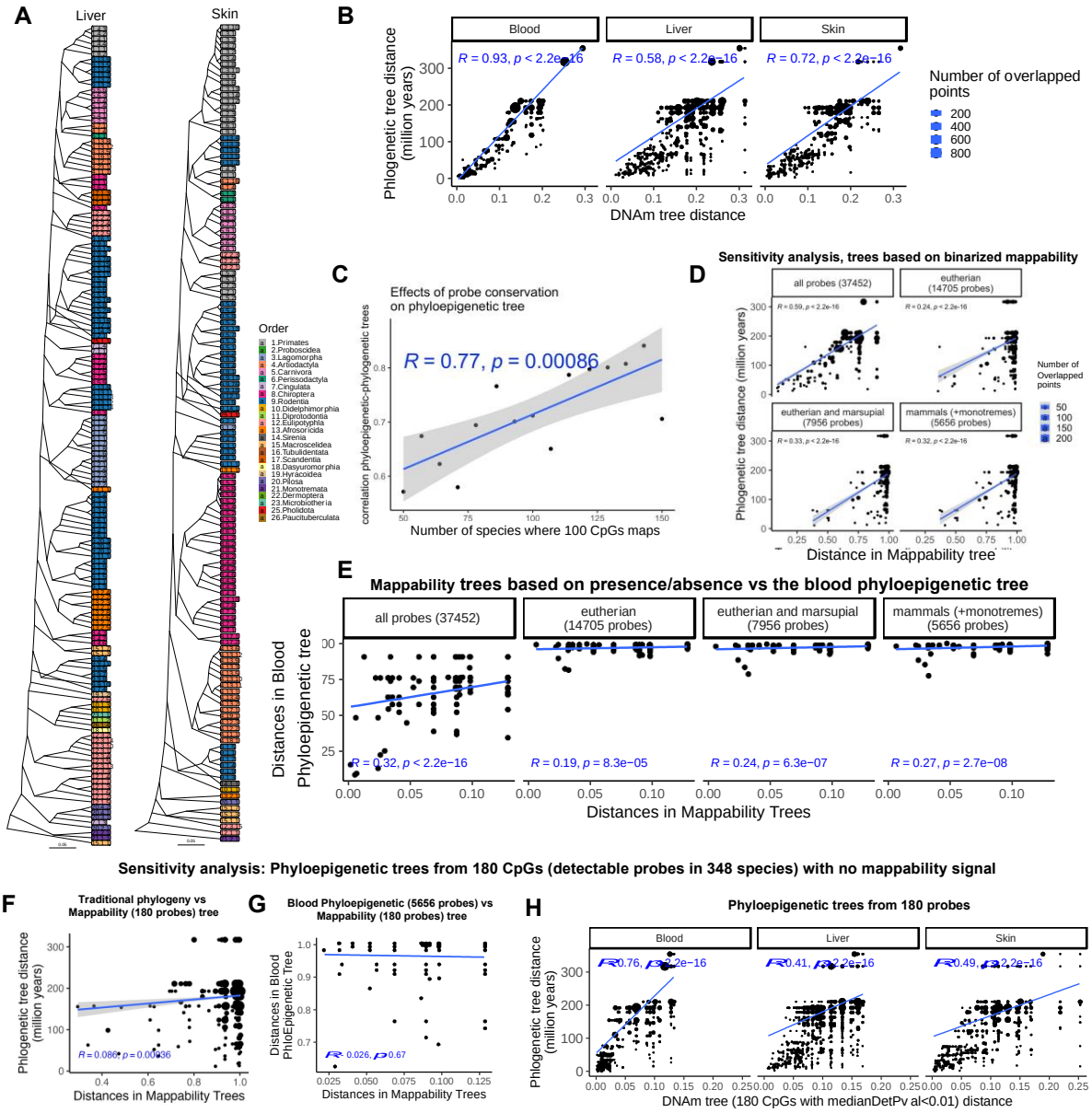


Fig. S2. DNA methylation is a good indicator of mammalian evolution. (**A)** Phyloepigenetic trees in blood (Fig. 1c), liver and skin of mammalian species. Distances: 1-cor. (**B)** The distances of phyloepigenetic (1-cor) and evolutionary trees are highly associated. The size of the dots indicates the number of overlapping points in the plot. (**C)** Sensitivity analysis of phyloepigenetic-phylogenetic trees relationship based on probe sequence conservation. One hundred random probes were selected 15 times based on reduction in conservation and alignment to mammalian genomes. The phyloepigenetic distances (1-cor) were calculated from hierarchical clustering of blood DNA methylation data. (**D)** trees that are only based on mammalian array mappability (sequence conservation) weakly track phylogenetic distances than DNAm based analysis (panel b). We created three trees based on different subsets of probes: 1) 37,452 CpGs presented in mammalian array, 2) 7,956 conserved CpGs in eutherians and marsupials, 3) 14,705 conserved CpGs in eutherians. 4) 5,656 conserved CpGs in all mammals. **(E)** Correlation of

Distances between blood phyloepigenetic and mappability trees. There is a weak relationship between phyloepigetic and mappability trees. **(F-H)** A sensitivity analysis to develop phyloepigenetic trees from 180 CpGs with no mappability signal. High detectable CpGs are a subset of monotreme CpGs with detection pvalue <0.01 in all 348 species. **(F)** There is no congruence between a conventional Phylogenetic tree and a mappability tree from highly detectable CpGs. **(G)** There is no congruence between the blood phyloepigenetic tree and a mappability tree from highly detectable CpGs. (H) Phyloepigenetic trees from 180 highly detectable CpGs show high congruence with the conventional phylogenetic tree. Distances: 1-cor. The phylogenetic distances are based on the TimeTree database.

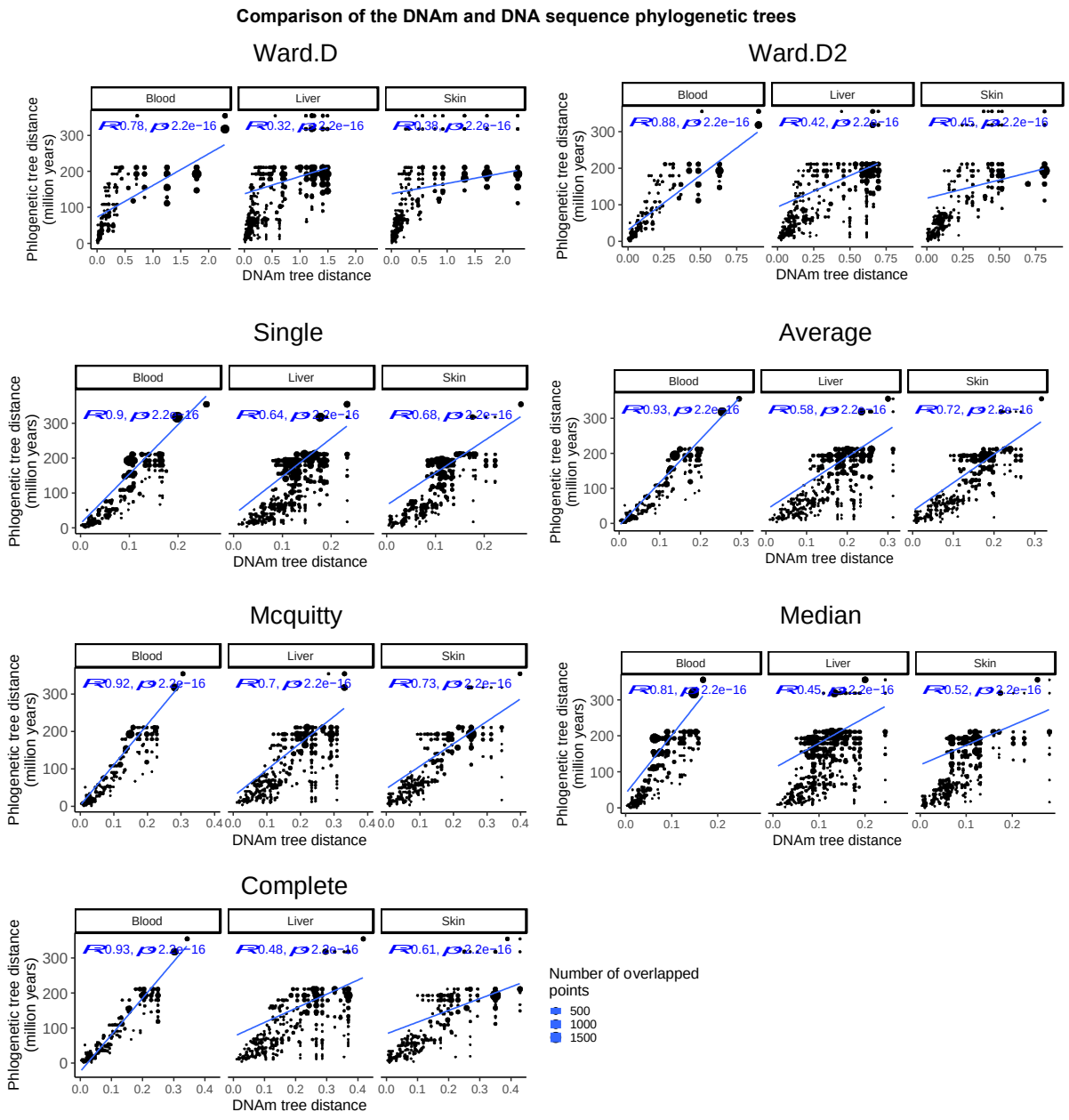
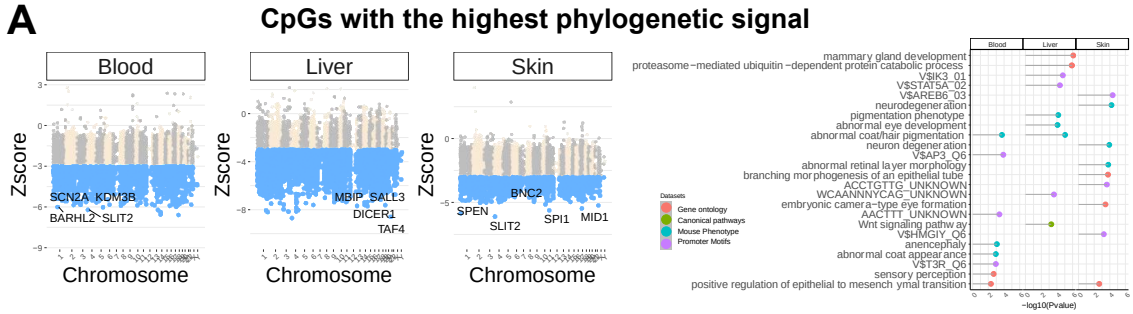
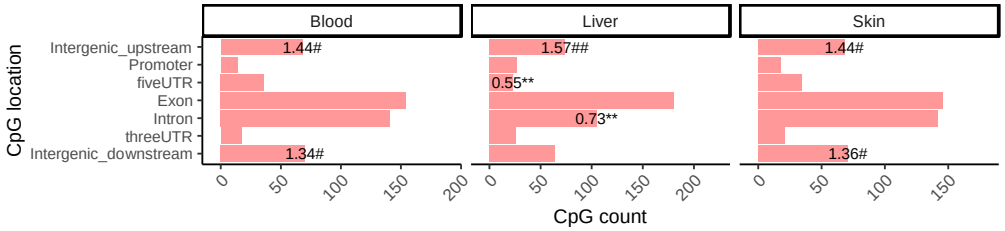
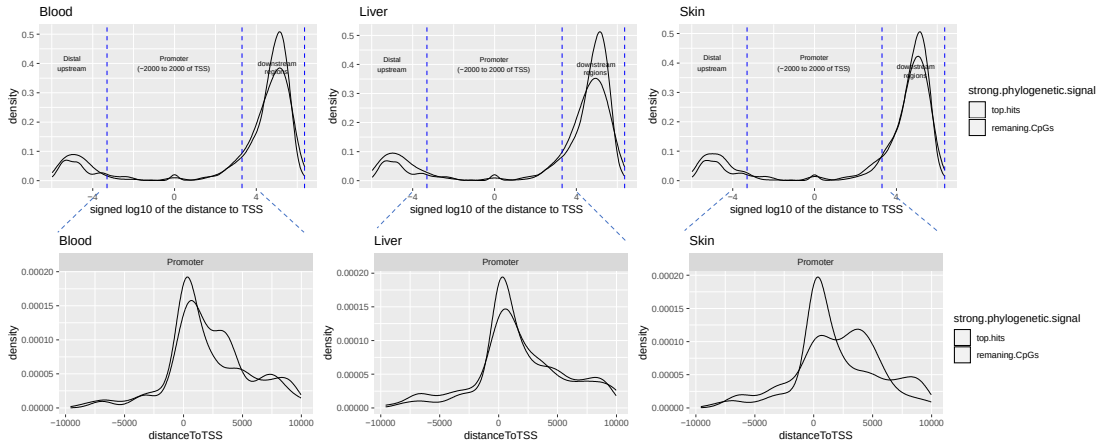
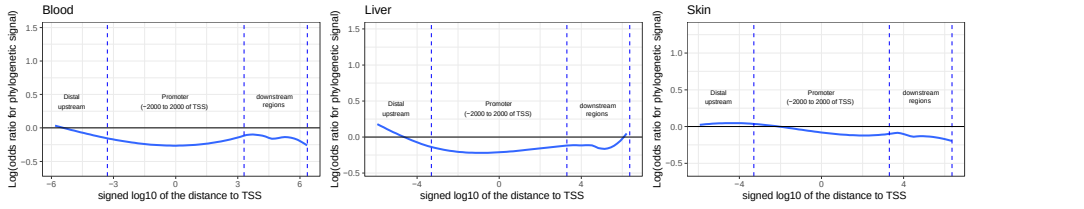


Fig. S3. Comparison of the effects of different linkage methods on phyloepigenetic-phylogenetic tree relationships. The panels compare different linkage methods to cluster DNAm trees in mammalian data. All distances are based on 1-cor in DNAm data between mammalian species. The phylogenetic distances are based on the TimeTree database. The average linkage method led to the highest agreement between phyloepigenetic-phylogenetic evolutionary trees.


A
CpGs with the highest phylogenetic signal
Blood
Liver
Skin
Zscore
Chromosome
SCN2A
KDM3B
BARHL2
SLIT2
MBIP
SALL3
DICER1
TAF4
BNC2
SPEN
SLIT2
SPI1
MID1
Datasets
Gene ontology
Canonical pathways
Mouse Phenotype
Promoter Motifs
mammary gland development
proteasome−mediated ubiquitin−dependent protein catabolic process
V$IK3_01
V$STAT5A_02
V$AREB6_03
neurodegeneration
pigmentation phenotype
abnormal eye development
abnormal coat/hair pigmentation
neuron degeneration
V$AP3_Q6
abnormal retinal layer morphology
branching morphogenesis of an epithelial tube
ACCTGTTG_UNKNOWN
WCAANNNYCAG_UNKNOWN
embryonic camera−type eye formation
AACTTT_UNKNOWN
Wnt signaling pathway
V$HMGIY_Q6
anencephaly
abnormal coat appearance
V$T3R_Q6
sensory perception
positive regulation of epithelial to mesench ymal transition
−log10(Pvalue)


B
Enrichment of the gene regions for top 500 CpGs with phylogenetic relationship
Blood
Liver
Skin
CpG location
Intergenic_upstream
Promoter
fiveUTR
Exon
Intron
threeUTR
Intergenic_downstream
1.44#
1.57##
0.55**
0.73**
1.34#
1.36#
CpG count


C
Blood
Liver
Skin
density
Distal upstream
Promoter (−2000 to 2000 of TSS)
downstream regions
signed log10 of the distance to TSS
distanceToTSS
strong.phylogenetic.signal
top.hits
remaning.CpGs


D
Blood
Liver
Skin
Log(odds ratio for phylogenetic signal)
signed log10 of the distance to TSS

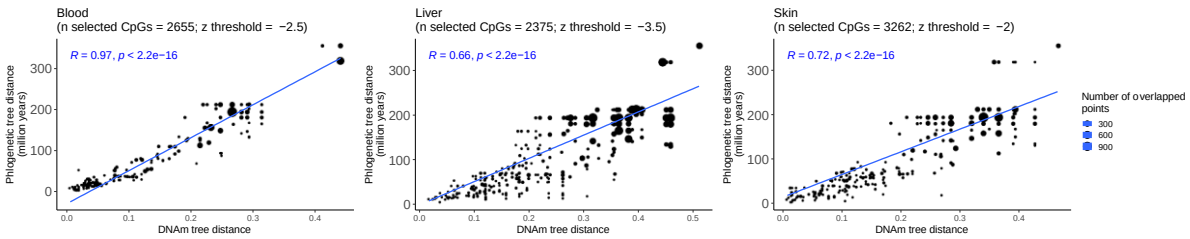

E
Phyloepigenetic trees based on CpGs with phylogenetic signal
Blood
(n selected CpGs = 2655; z threshold = −2.5)
Liver
(n selected CpGs = 2375; z threshold = −3.5)
Skin
(n selected CpGs = 3262; z threshold = −2)
R = 0.97, p < 2.2e−16
R = 0.66, p < 2.2e−16
R = 0.72, p < 2.2e−16
Phlogenetic tree distance (million years)
DNAm tree distance
Number of overlapped points
300
600
900

Fig. S4. Characterizing CpGs that contain the phylogenetic signal. We used phylosignal() R function is a part of the picante package (*51*) in R, which is specifically designed for the analysis of phylogenetic and ecological community data. The phylosignal() function measures the phylogenetic signal in a given set of traits (here CpGs) for a group of species (here mammalian species). The K statistic from Blomber was used to assess the degree to which closely related species resemble each other more than they resemble species chosen at random (*32*). **(A)** Manhattan plots of phylogenetic signals of individual CpGs in different mammalian tissues of our study. The y-axis is the variance-z score of the phylogenetic signal K statistics. The top CpGs with phylogenetic signals are labeled by the adjacent gene in the human Hg19 genome. The right panel shows the GREAT enrichment analysis of top 500 CpGs (p=0.001, selection by variance zscore) per tissue. We only report enrichment terms that are significant at pvalue<0.005 and contain at least five significant genes. The top three significant terms per column (EWAS) and enrichment database are shown in the panel. Several of these enrichment terms do not pass the multiple test correction using FDR. **(B)** Location of top 500 CpGs (p=0.001, selection by variance zscore) in each tissue relative to the closest transcriptional start site. The odds ratio of the proportion changes than the background are reported in for each bar. Fisher exact p values: * (or #) p<0.05, **(or ##) p<0.01, ***(or ###) p<0.001, ****(or ####)p<0.0001. # odds ratio>1, * odds ratio<1. **(C)** Density plots of the distribution of top 500 CpGs with phylogenetic signal (red color) vs the background CpGs (blue color) relative to the nearest transcription start site (TSS, x-axis). The lower panel zooms in on the promoter regions (-10000 to 10000 of the TSS). **(D)** Scatter plots displaying the log-odds ratios of regions exhibiting significant phylogenetic signals relative to the Transcription Start Site (TSS) are presented. In this analysis, CpGs were grouped into categories using sliding windows relative to the TSS, ensuring a minimum count of 10 CpGs per group. To assess enrichment, the Fisher exact overlap test was employed, focusing on the top 500 CpGs displaying phylogenetic signals within each region. The results indicate notable enrichment (OR>3) in certain intergenic and genic regions, but not in promoters. **(E)** Scatter plot of the distances in blood phyloepigenetic (1-cor) based on CpGs with a high phylogenetic signal vs the traditional evolutionary tree. The number of used CpGs and the zscore threshold for each phyloepigenetic tree is reported in the corresponding panel.

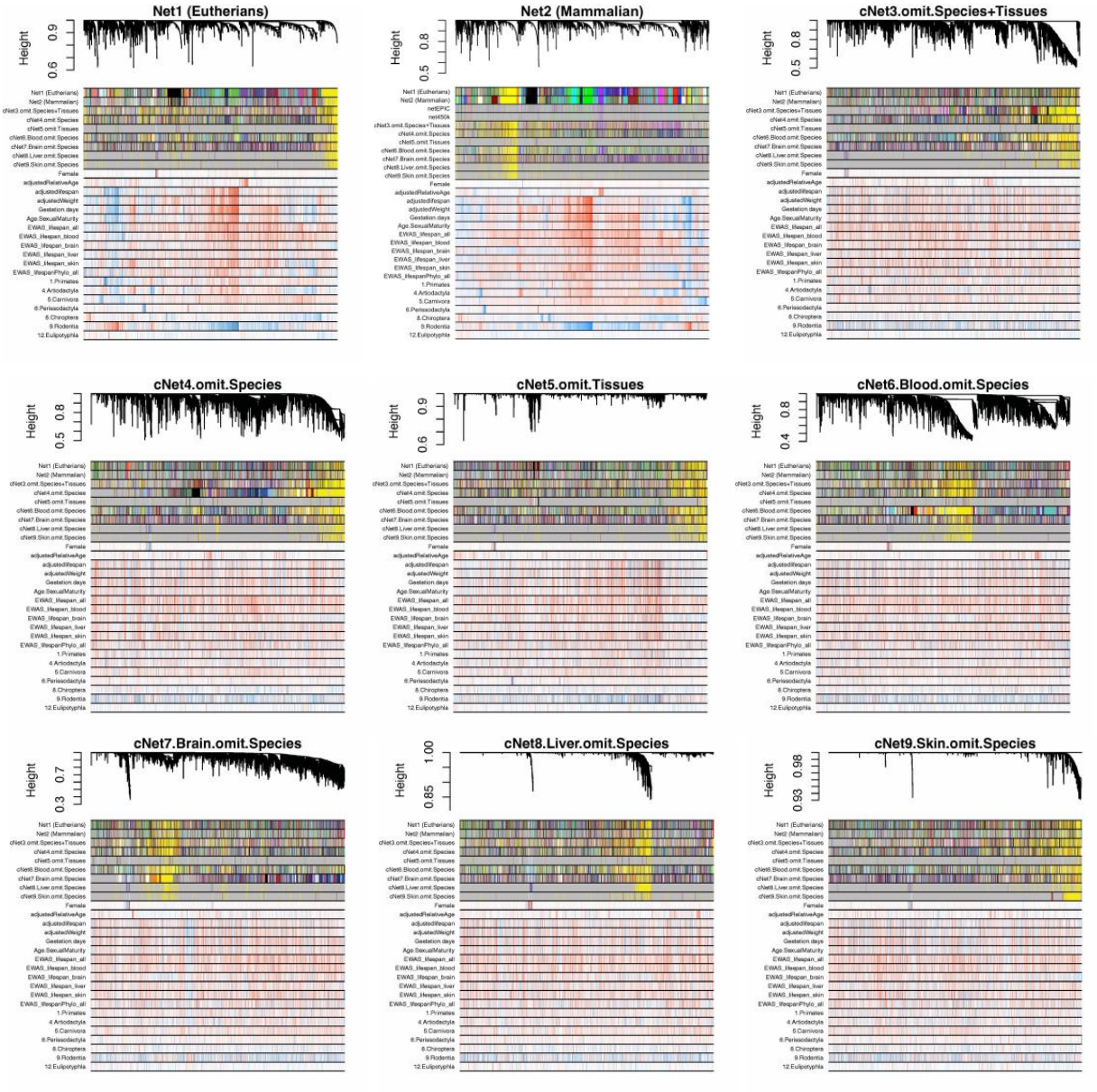


Fig. S5. Constructed WGCNA networks from mammalian DNA methylation data. A total of nine WGCNA networks were constructed and compared for module association with species and individual sample characteristics. Network 1, an unsupervised network of 14,705 conserved CpGs in 167 eutherian species, showed the strongest module-trait associations. An additional unsupervised network was developed based on subsets of probes for future study applications: Network 2, 7,925 probes that map to both eutherians and marsupials. Additionally, seven consensus networks of 14,705 eutherian probes by different tissue and species combinations were formed to identify the most conserved modules for studying individual sample characteristics. All module colors were matched to network 1 for comparison.

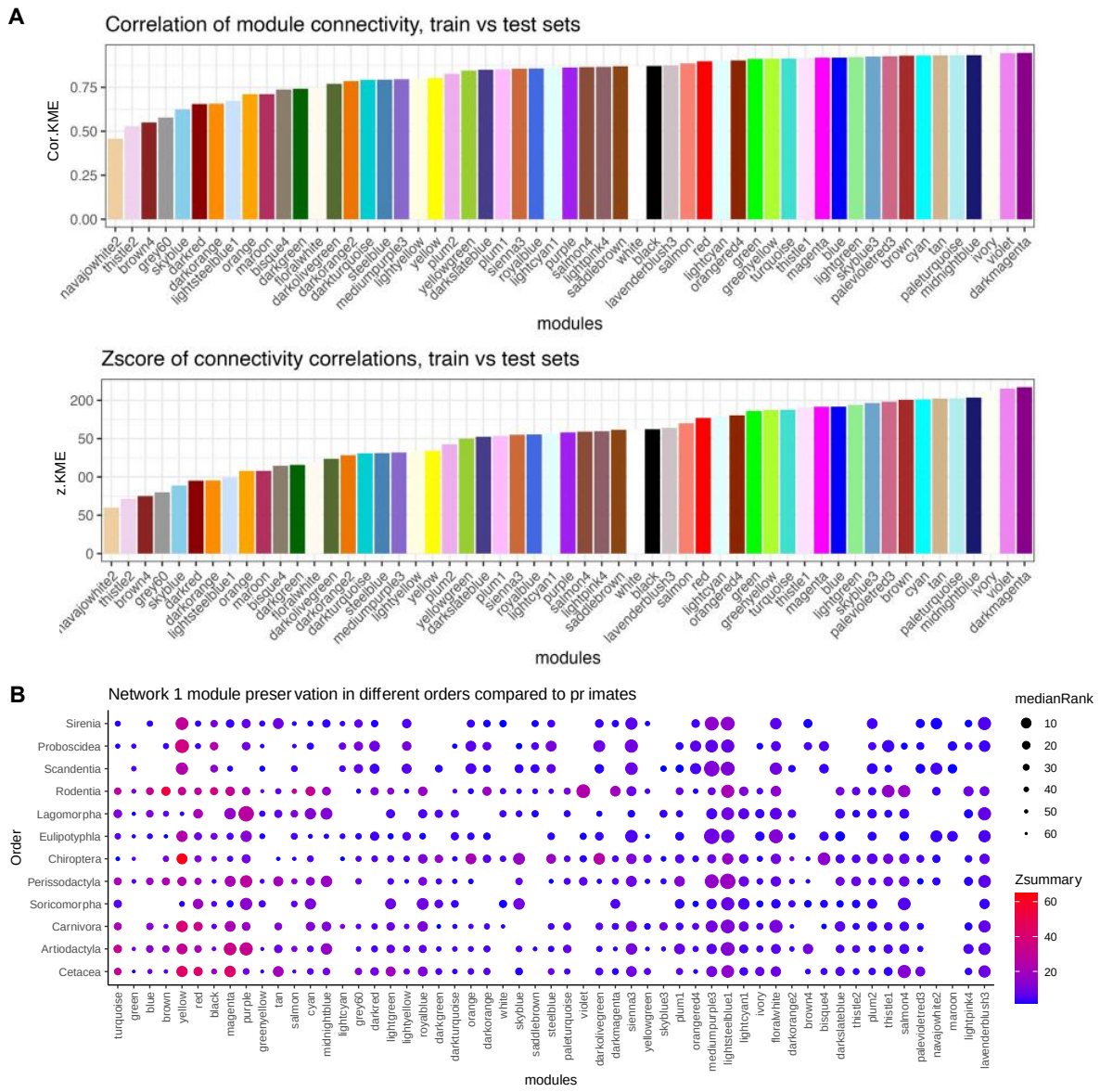


Fig. S6. The modules in the mammalian methylation network are highly conserved in an independent dataset and between phylogenetic orders. (A) mammalian methylation DNAm network was trained on 10,927 samples from 175 eutherian species. We examined the conservation of these modules in an independent dataset which contained 3,692 samples from 29 tissues of 228 mammalian species (164 new species, 64 shared with the training set). The module preservation was assessed by a correlation of module connectivity (KME) between these two independent (training and test) datasets. All modules had a significant (p<1e-315) and positive correlation. **(B)** The DNA methylation modules are highly conserved in different phylogenetic orders. Module preservation is estimated with permutation of networks in each order and comparison with primates. The figures show the summary z score of preservation, and median rank of preservation for each module.

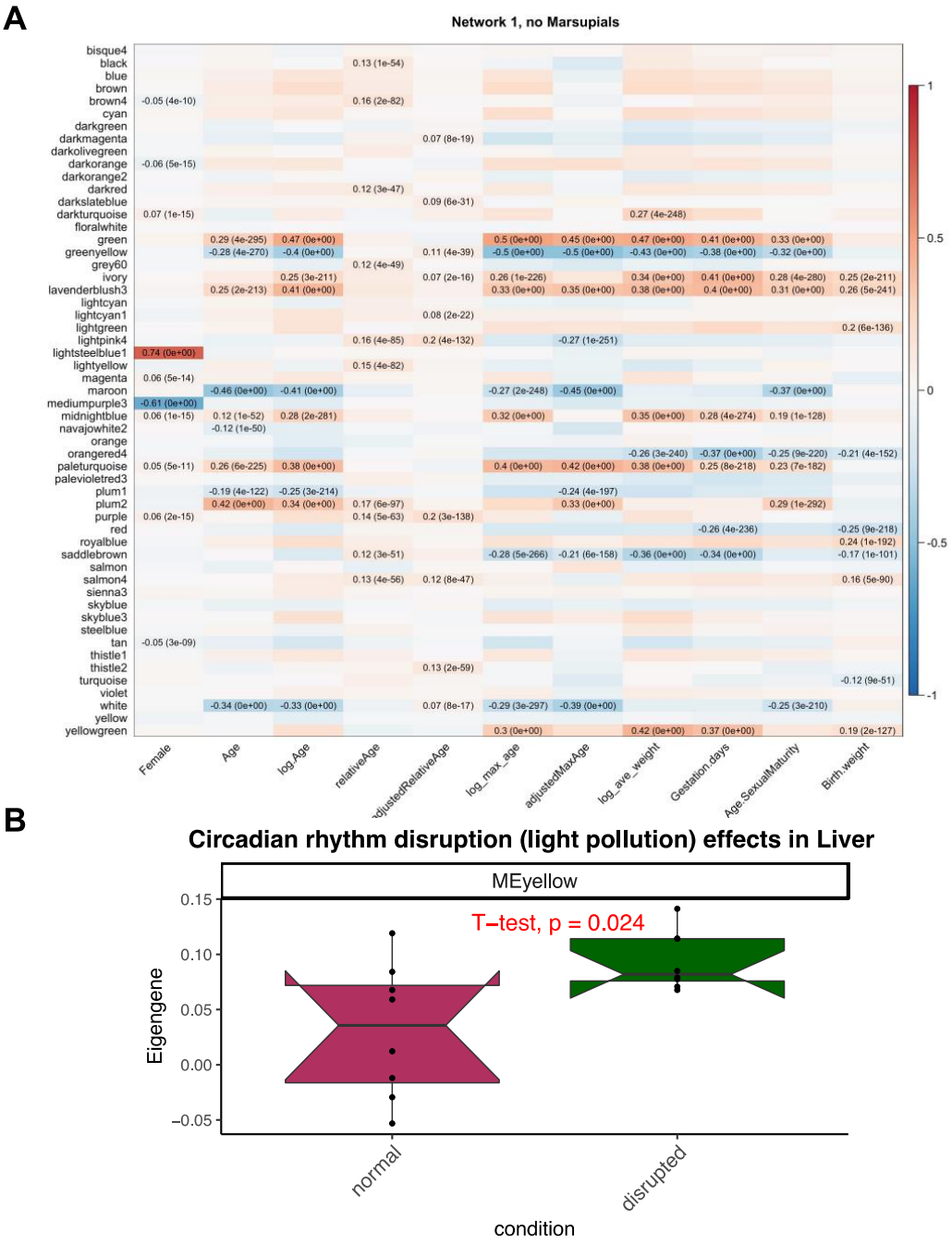


Fig. S7. DNA methylation Module-trait association in mammals. (**A)** Association of different individual and species-level traits with network 1 modules. Network 1 consists of 14,705 conserved CpGs in 324 eutherian species. The data include around 70 tissue types, from all age ranges of most of the species. (**B)** Long-term exposure to light pollution increases the yellow module eigengene. Control group, n=8, standard light/dark cycles of 12 hours light (100 lux) followed by 12 hours dark. Circadian rhythm disrupted group, n=8, the light cycle included 12 hours light (100), followed by 12 hours dim light (20 lux). Cohorts were exposed to these conditions from three months of age, for a period of 12 months.

Fig. S8. Relationship of mammalian co-methylation modules with EWAS of age and mammalian maximum lifespan. Box plots represent the z score of association for the top seven modules in each EWAS model. The dashed lines are z scores equal to 2 or -2, which is the equivalent to $p<0.05$. EWAS_age for each tissue is the Stouffer meta-analysis of DNAm-age association in each species. EWAS_lifespan is based on the association of mean DNAm levels to maximum reported lifespan for each species. EWAS_lifespanPhylo is the phylogenetic regression analysis of mean DNAm levels to maximum reported lifespan for each species. Detailed EWAS methods and results for age and maximum lifespan are reported in articles from the Mammalian Methylation Consortium (*1*, *2*).

Fig. S9. Gene ontology enrichment of the hub CpGs in each module. The gene-level enrichment was done using GREAT analysis and human hg19 assembly background. The background probes were limited to 14,705 conserved probes in eutherians. The biological processes were reduced to parent ontology terms, and the top 10 parent terms with nominal $p<10^{-3}$ for each module were reported in the heatmap. The larger ontology categories were defined manually and used to arrange the ontology terms in the heatmap. * Indicates the top ontology category for each module. The input includes up to 500 hub CpGs for each module.

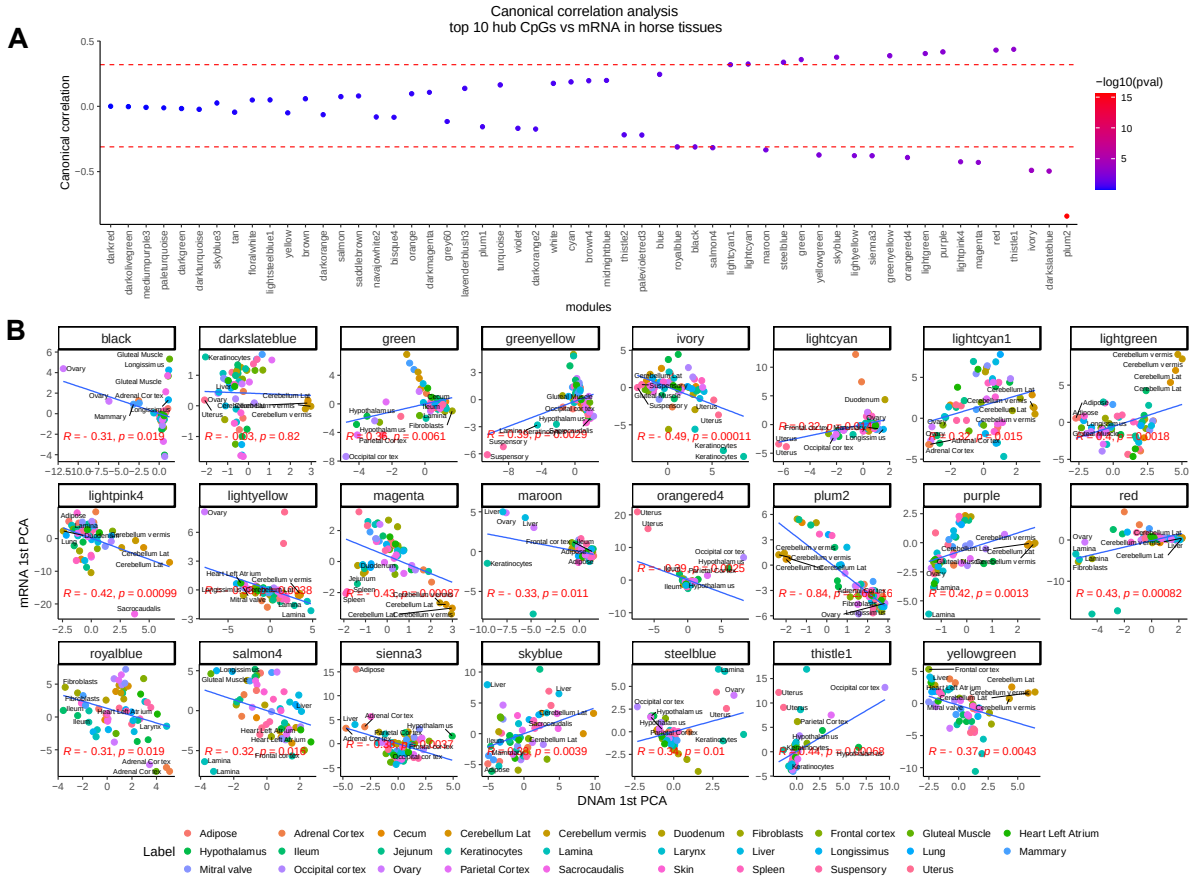


Fig. S10. DNA methylation modules relate to mRNA changes and are preserved in different phylogenetic orders. (A) DNA methylation modules and mRNA networks are canonically correlated in horse tissues. The scatter plots report the canonical correlation of 10 hub CpGs of each module to mRNA levels in horse tissues. Canonical correlation is done for the hub CpGs of each module and mRNA changes of their neighboring genes in the horse genome (EquCab3.0.100). The red lines indicate p<0.05. **(B)** Scatter plots of 1st principal component of 10 hub CpGs and mRNA levels of the neighboring gene in 57 horse tissue samples. The mRNA and DNA methylation data originated from 57 samples from 29 different horse (*27*).

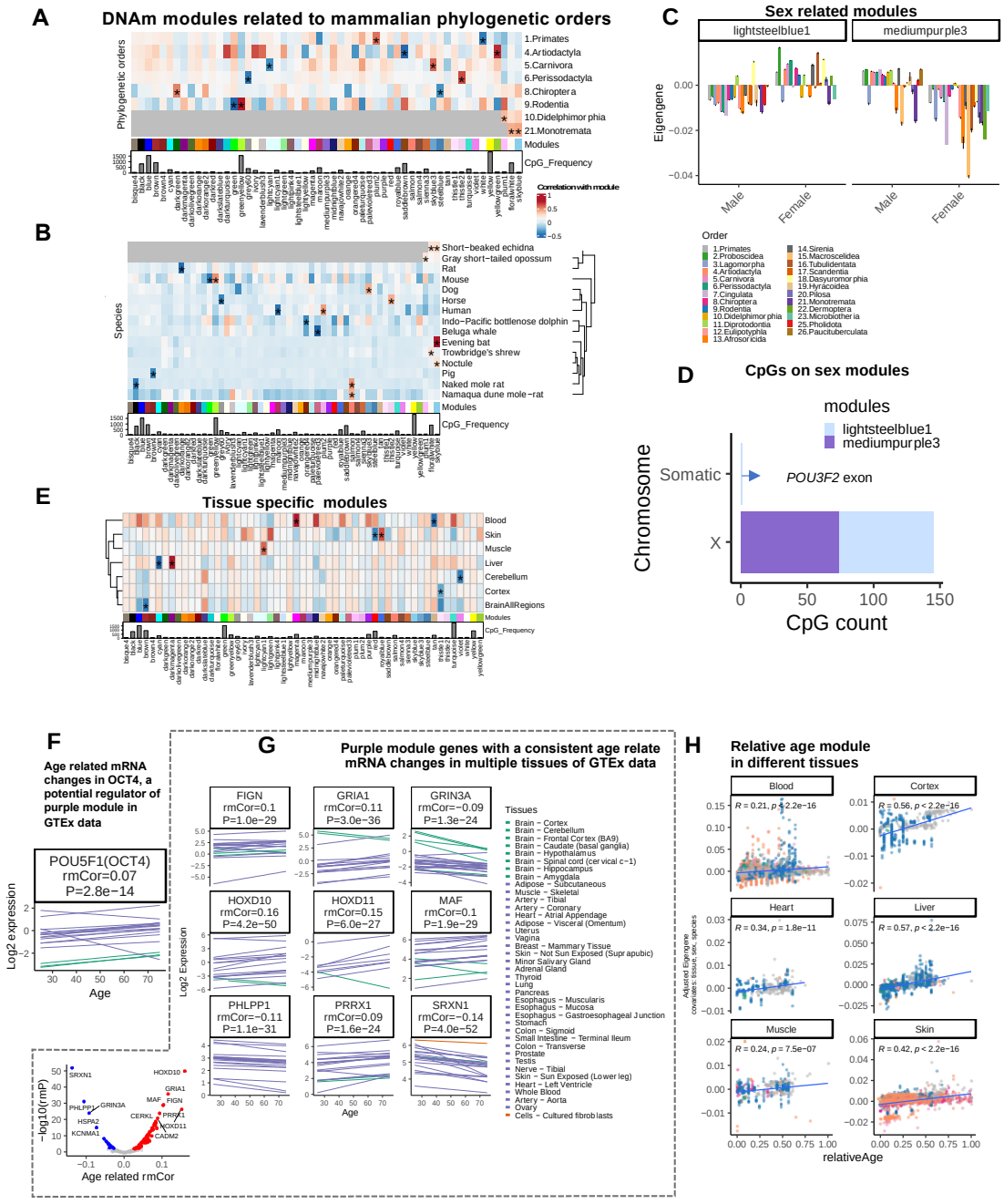


Fig. S11. Co-methylation modules that relate to specific species and taxonomic orders. **(A)** Heatmap of taxonomic order specific modules. * indicated the top two modules related to each taxonomic order with minimum absolute correlation of 0.4. The marsupial and monotreme analysis is done by subsetting the modules to the probes that also map to these clades. (**B)** DNA methylation modules associated with individual species. The top two modules for each species are labeled by *. The rows are clustered based on hierarchical clustering of Euclidean distances and complete method. The species with weak module association (r<0.2) were not shown in the heatmap. (**C)** Sex-specific modules in different taxonomic orders. (**D)** Distribution of sex module CpGs on sex

chromosomes. Only a CpG in POU3F2 exon is in the somatic chromosomes: humans chr6 and mice chr4. **(E)** Heatmap of tissue-specific modules. * Indicates the top two modules related to each tissue, cell type, or tissue region with minimum absolute correlation of 0.38 and p<1e-315. The tissues with no modules that passed these criteria were excluded from the heatmap. **(F)** Age-related mRNA changes in Yamanaka factors, which are the potential upstream regulators of the purple module based on IPA analysis. rmCor: repeated measure correlation analysis (*94*) of all tissues in the GTEx database. (**G)** Scatter plot of mRNA changes purple module genes with the highest rmCor in human GTEx data. The plots only show the regression lines of the tissues with significant (p<0.05) mRNA-age correlations. The color of the lines indicates different tissue types. The lower left plot is the volcano plot of repeated measures correlation (rmcorr R package) of all purple module's genes in GTEx data.

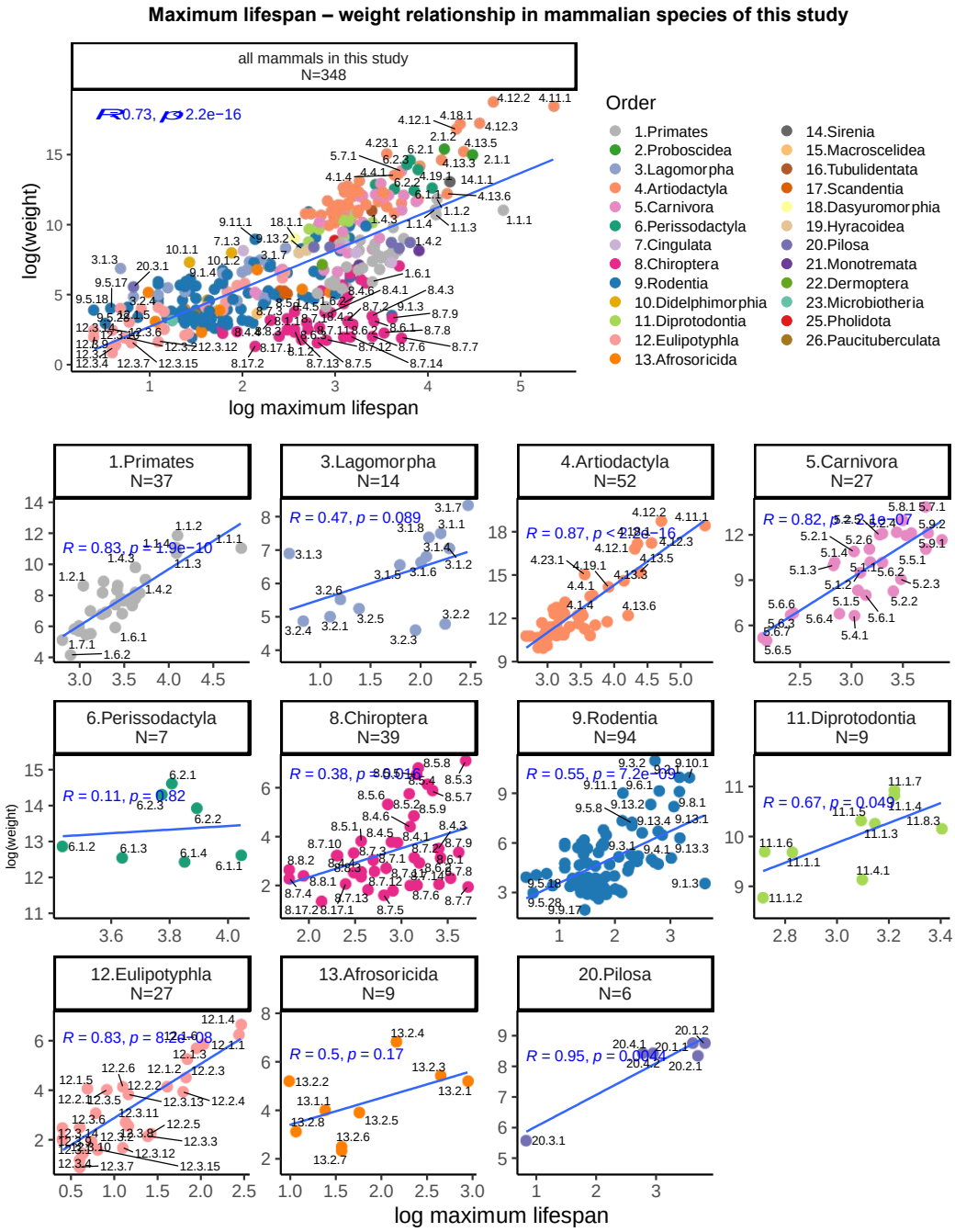


Fig. S12. Maximum lifespan versus average adult weight in mammalian species of this study. Each panel reports the relationship between log transformed (base e) values of maximum lifespan (x-axis) and average adult weight (y-axis) in different taxonomic orders. Points are colored by taxonomic order. R, Pearson correlation coefficient; P, two-sided T-test p-value.

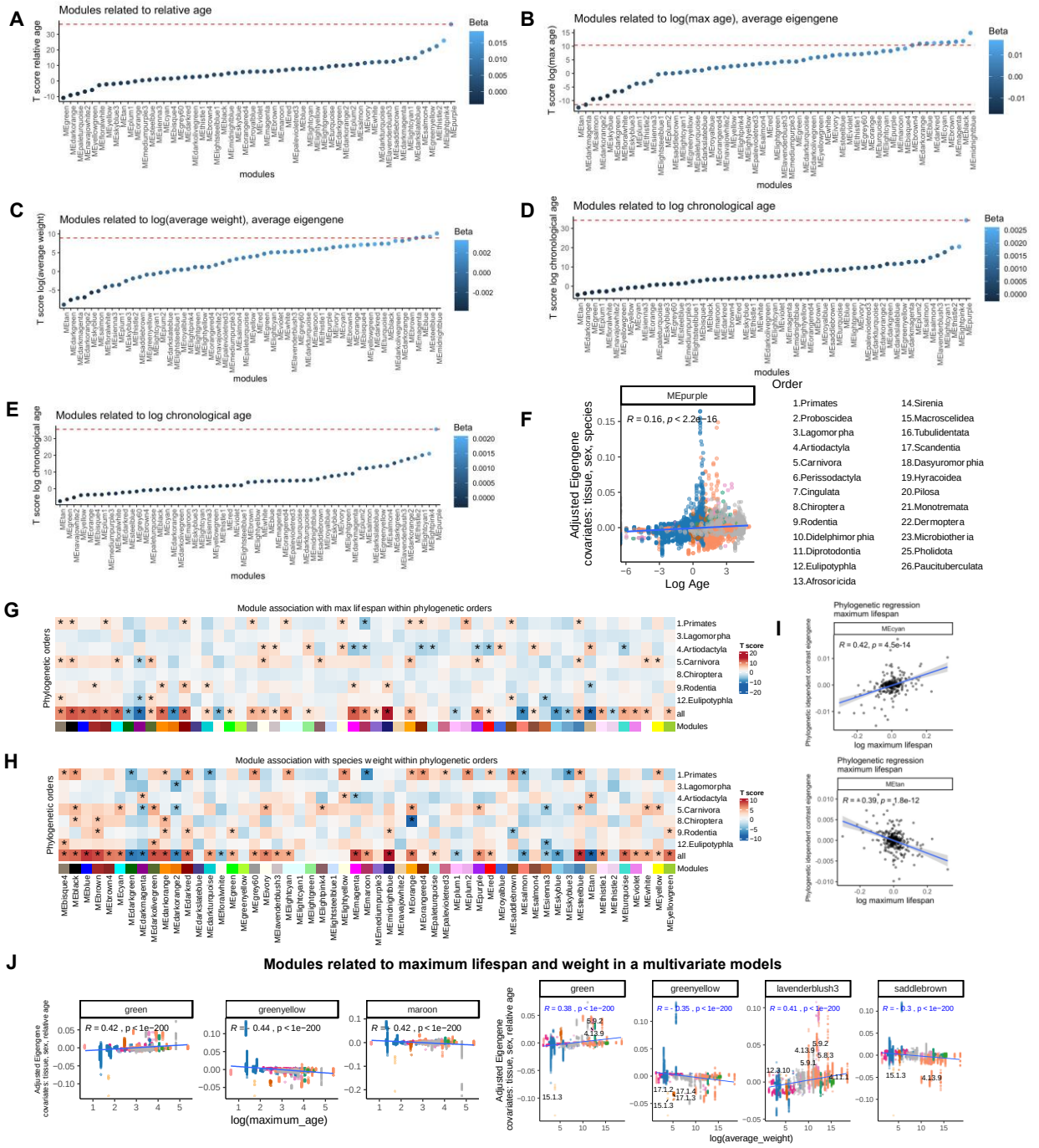


Fig. S13. DNA methylation modules' association with age, max age and weight in mammals. The association of modules' eigengenes with relative age (**A**), log species maximum age (**B**), log species average adult weight (**C**), chronological age (**D**), and log chronological age (**E**). Red line: RelativeAge, p<1e-300; chronological age, p<1e-100; log chronological age, p<1e-200; maximum age, p<1e-33; weight, p<1e-18. Covariates: tissue, sex and species differences for relative age, chronological age and log chronological age; relative age, tissue, and sex for log(max age) and log(average weight). (**F)** Purple module also relates to log chronological age in mammals.

Module association with log maximum age **(G)** and log species adult weight **(H)** within each phylogenetic order. * indicates the p<1e-3 for max age, and weight modules for each order. (**I)** Phylogenetic regression analysis of the modules with mammalian maximum lifespan. We used the independent contrast method (*52*) to relate contrasts in module eigengenes (y-axis) to log transformed maximum lifespan. The independent contrast method combines closely related tree nodes to account for their high correlation. While the axes still correspond to the original variables, the original variables have been transformed to a different scale. The evolutionary tree for phylogenetic regression was downloaded from the TimeTree database. Only the top two modules are presented in the scatter plot. Covariates: relative age, tissue, and sex. (**J)** The modules related to (log transformed) maximum age and average adult weight using multivariate regression analysis. Covariates: tissue, sex, relative age.

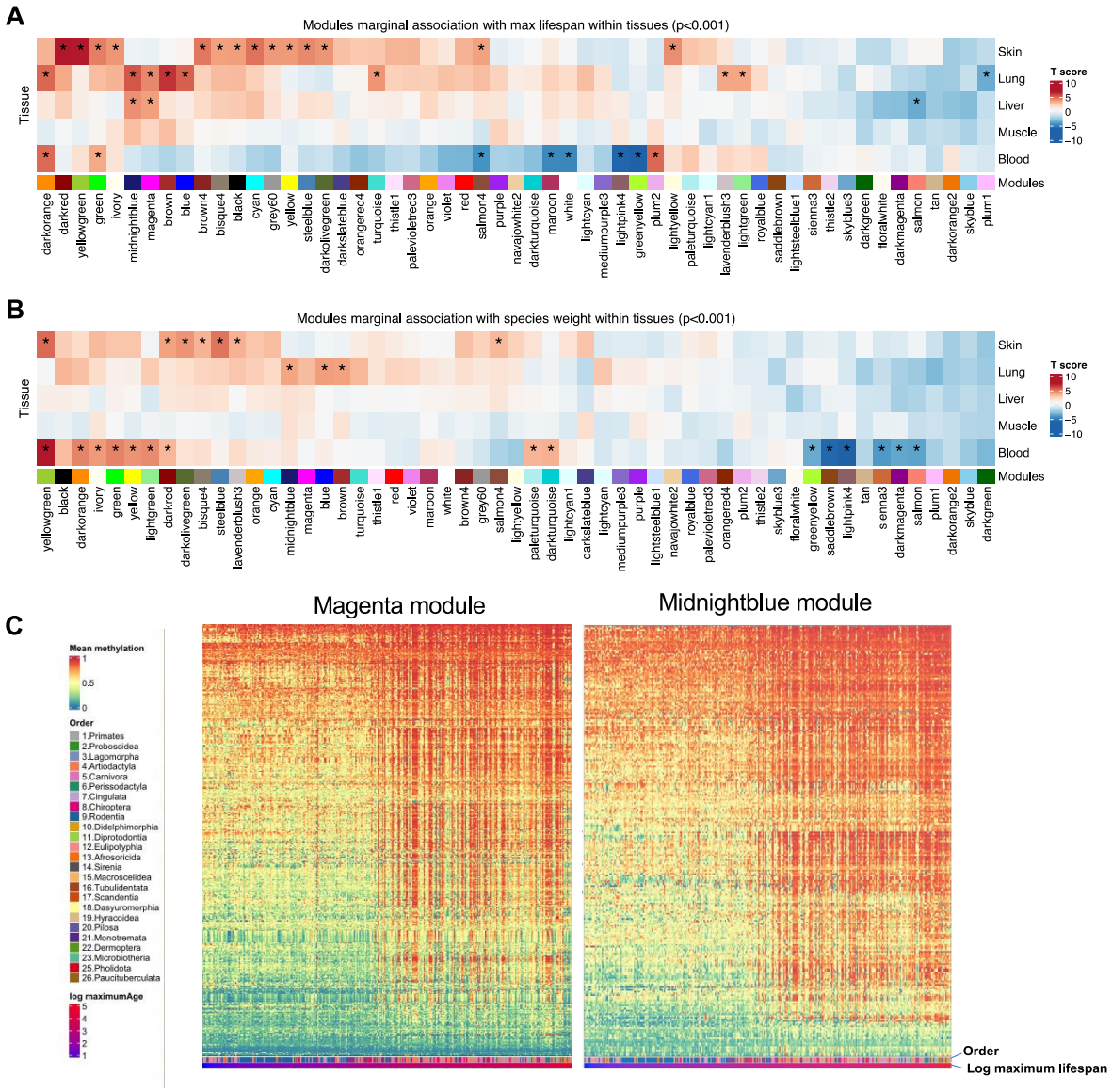


Fig. S14. DNA methylation modules' association with max age and weight in different tissues. The associations are evaluated by a marginal regression analysis of the log maximum lifespan (A), and log weight (B) of the mammals to the mean eigengene values of each module per species per tissue. * Indicates p<0.001. (C) Mean methylation of the CpGs per mammalian species in the magenta and midnightblue modules. The species (columns) are sorted based on their maximum lifespan.

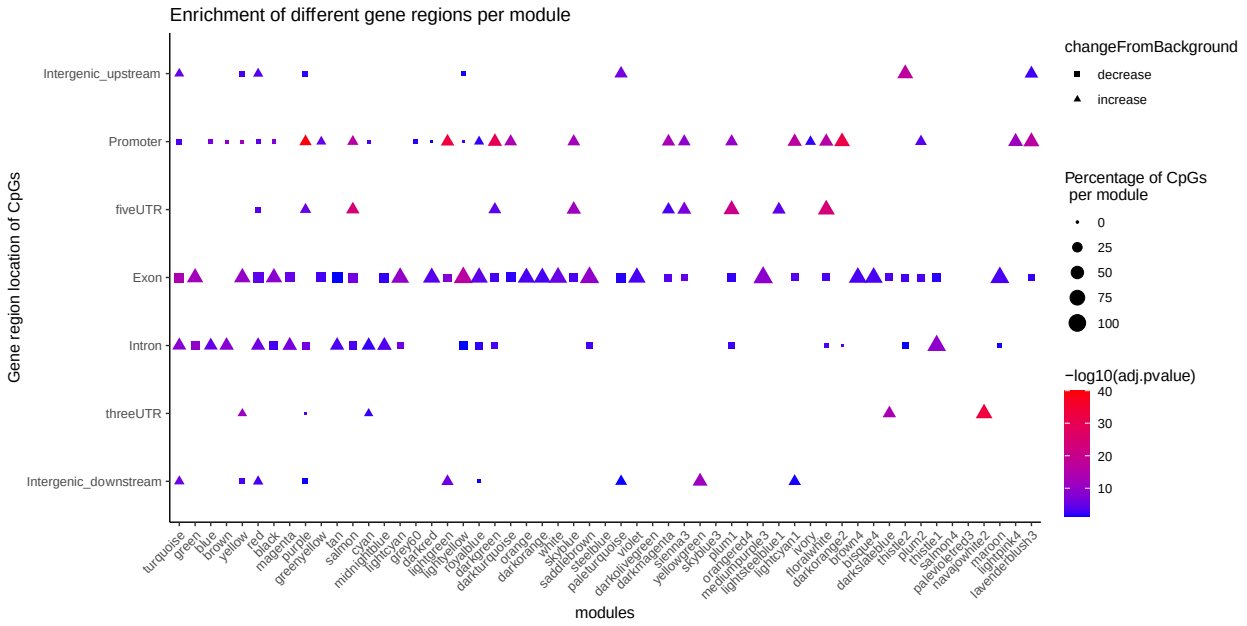


Fig. S15. Enrichment of gene regions for CpGs in each module. The CpGs are annotated relative to the nearest transcriptional start site in the human hg19 genome (*5*). Only genomic regions with a significant difference (5% false discovery rate) to the background (proportion chi-square test) are reported in the dot-plot.

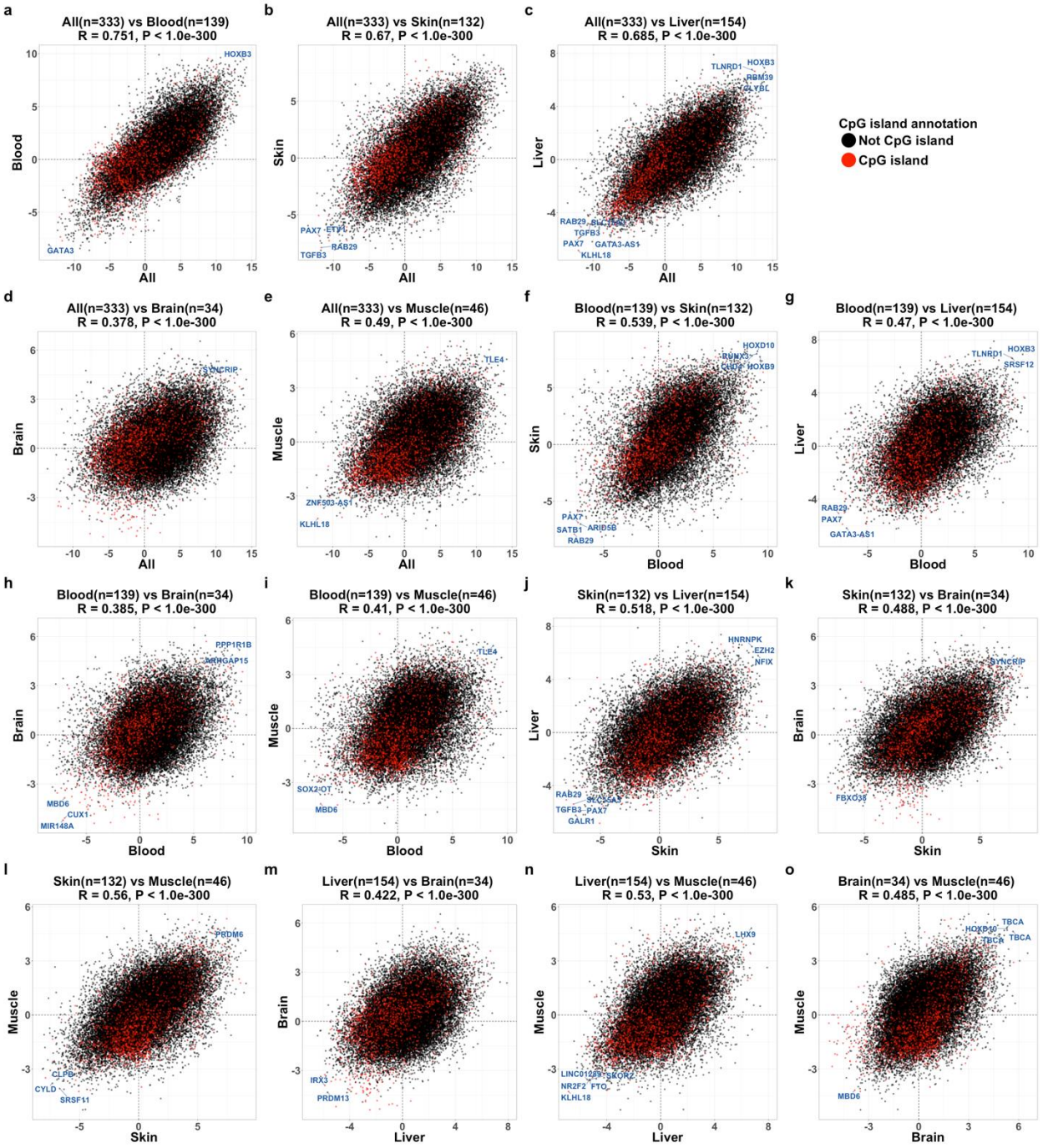


Fig. S16. Generic Lifespan EWAS in different tissues from Eutherian species. Scatter plot of CpG Z statistics agreements between tissues, color-coded by human CpG island annotations (not island: black, island: red). Both x- and y-axes are CpG Z statistics for the set of EWAS background CpG probes (28,318) consistent with the methods section (mappable to humans and mice and correlation with calibration exceeds 0.8). Panels show agreements between (**A)** blood vs. all, (**B)** skin vs. all, **(C)** liver vs. all, **(D)** brain vs. all, **(E)** muscle vs. all, **(F)** skin vs. blood, **(G)** liver vs. blood, **(H)** brain vs. blood, **(I)** muscle vs. blood, **(J)** liver vs. skin, **(K)** brain vs. skin, **(L)** muscle vs. skin, **(M)** brain vs. liver, **(N)** muscle vs. liver, **(O)** muscle vs. brain. Panel titles report r and p as Pearson's correlation and p-values, respectively.

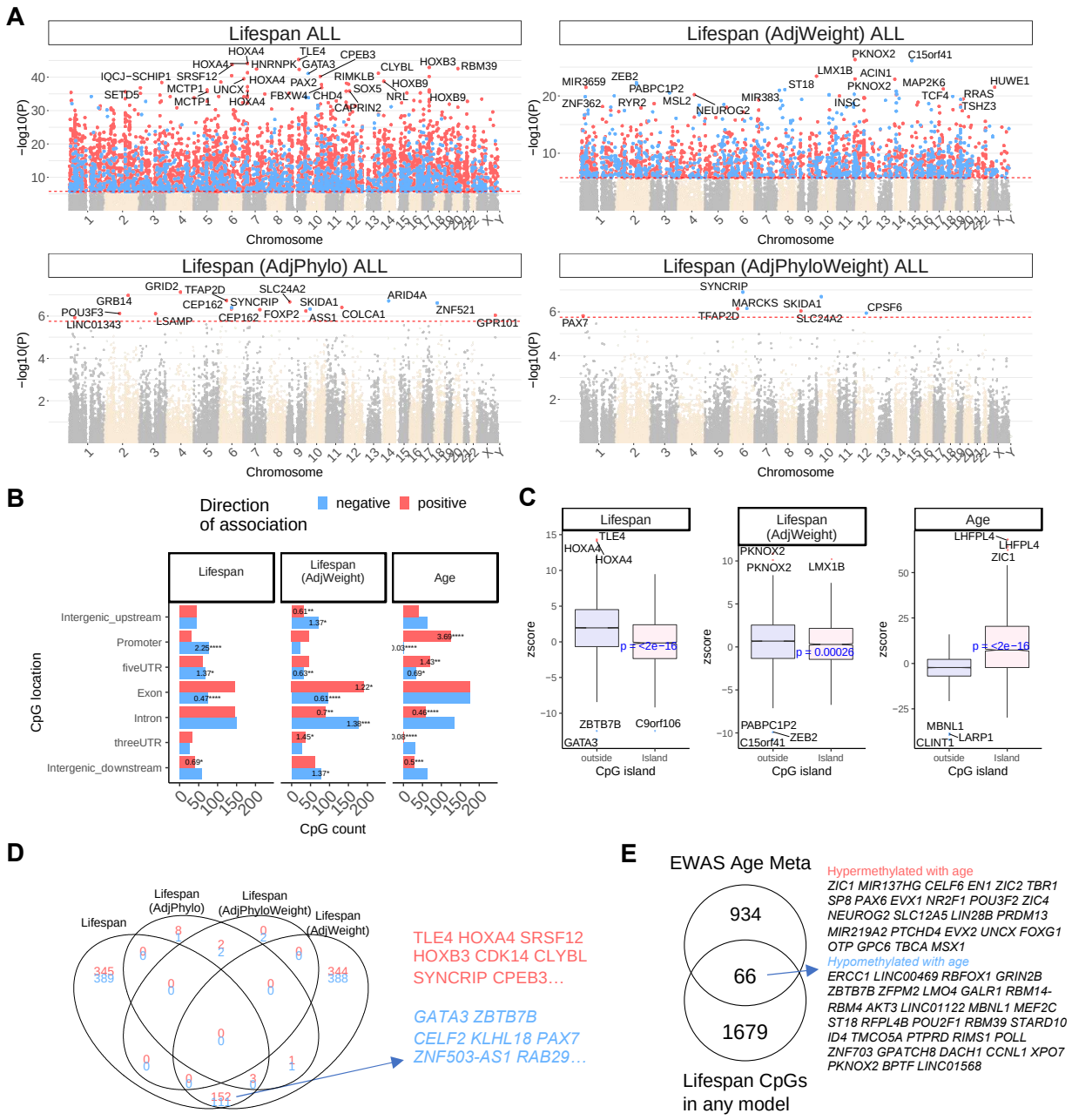


Fig. S17. EWAS of significant CpGs related to mammalian maximum lifespan, adjusted by weight and phylogeny. (A) are Manhattan plots reporting Manhattan plots of lifespan, lifespan EWAS adjusted by weight (AdjWeight), lifespan EWAS adjusted by phylogeny (AdjPhylo), and lifespan adjusted by both weight and phylogeny (AdjPhyloWeight). The background probes were limited to the set of EWAS background CpG probes (28,318) consistent with the methods section (mappable to humans and mice and correlation with calibration exceeds 0.8). (**B**) Location of the top CpGs in each tissue relative to the closest transcriptional start site. A panel for the top 1000 age related CpGs was added to the figure for comparison (*7*). The changes in gene regions were tested by a hypergeometric test in proportion to the background. The odd ratios and p-values (* <0.05, **<0.01, ***<0.001, ****<0.0001) of changes are reported for each bar. **(C)** Boxplot of

association with mammalian maximum lifespan by human CpG island status. The mean difference was tested by Student T-test. **(D)** Venn diagram of the overlap in the top 1000 (500 per direction) significant CpGs for different models of EWAS of lifespan from panel (A) The overlap hits were labeled by neighboring genes. **(E)** Overlap of CpGs associated with mammalian lifespan and the top 1000 CpGs that relate to chronological age in mammals (*7*).

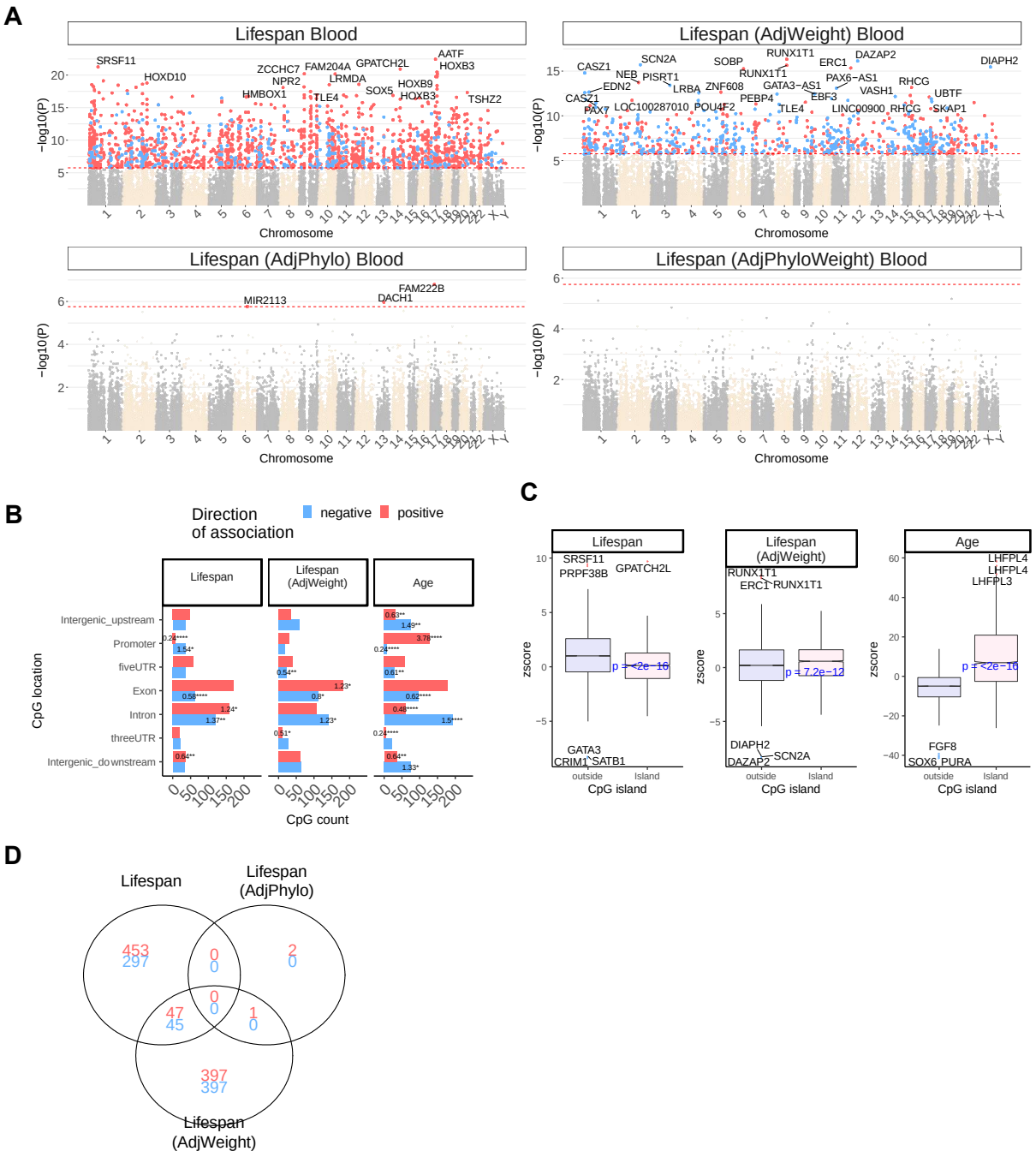


Fig. S18 | EWAS of mammalian maximum lifespan in blood. The associations were examined with four different models: 1) lifespan: each species as a datapoint in the model regardless of evolutionary distance. 2) lifespan adjusted for average species weight. 3) lifespan adjusted for evolutionary distance by phylogenetic regression. The evolutionary tree was acquired from TimeTree database (*45*). 4) lifespan adjusted for both average adult species weight and evolutionary distance. **(A)** Manhattan plots of EWAS of maximum lifespan in the set of EWAS background CpG probes (28,318) consistent with the methods section (mappable to humans and mice and correlation with calibration exceeds 0.8). The coordinates are based on the alignment to the human hg19 genome. The direction of associations with $p < 0.001$ (red dotted line) is

highlighted by red (hypermethylated) and blue (hypomethylated) colors. Some top CpGs were labeled by the neighboring genes, **(B)** Location of top CpGs relative to the closest transcriptional start site. A panel for the top 500 age-related CpGs in each direction was added to the figure for comparison (*7*). The changes in each gene region were tested by Fisher's exact test based on the same background. The odds ratios and p-values (* <0.05, **<0.01, ***<0.001, ****<0.0001) of changes are reported for each bar. **(C)** Boxplot of association with mammalian maximum lifespan by human CpG island status. The mean difference was tested by a Student's T test. A panel for the top 1000 age-related CpGs was added to the figure for comparison, **(D)** Venn diagram of the overlap in the top 1000 (500 per direction) significant CpGs for different models of EWAS of lifespan. The Venn diagram does not show AdjPhyloWeight because it contains zero CpG probe past the significance threshold.

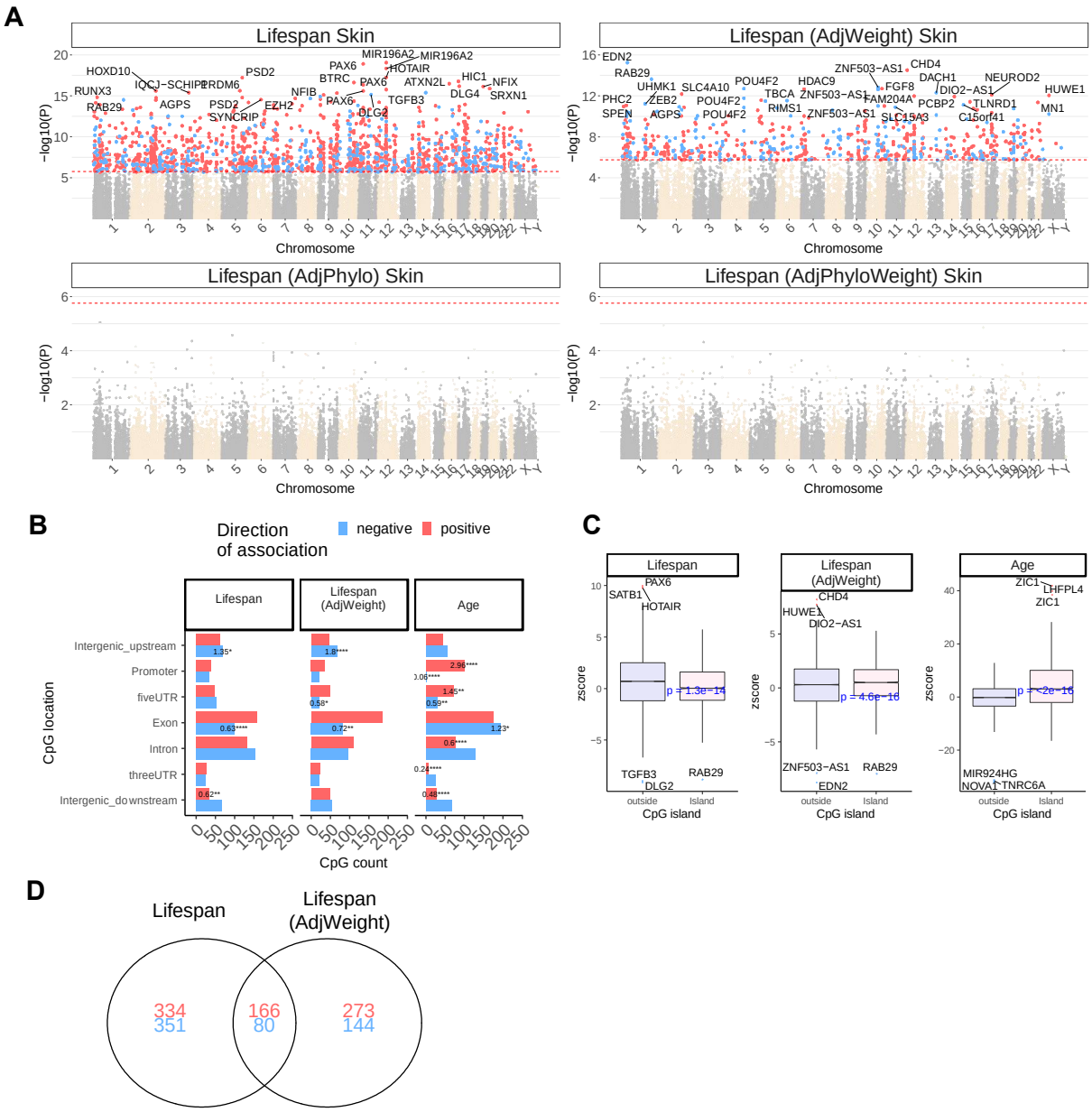


Fig. S19. EWAS of mammalian maximum lifespan in skin. The associations were examined with four different models: 1) lifespan: each species as a datapoint in the model regardless of evolutionary distance. 2) lifespan adjusted for average species weight. 3) lifespan adjusted for evolutionary distance by phylogenetic regression. The evolutionary tree was acquired from TimeTree database (*45*). 4) lifespan adjusted for both average adult species weight and evolutionary distance. **(A)** Manhattan plots of EWAS of maximum lifespan in the set of EWAS background CpG probes (28,318) consistent with the methods section (mappable to humans and mice and correlation with calibration exceeds 0.8). The coordinates are based on the alignment to the Human hg19 genome. The direction of associations with $p < 0.001$ (red dotted line) is highlighted by red (hypermethylated) and blue (hypomethylated) colors. The top few CpGs were labeled by the neighboring genes, **(B)** Location of top CpGs in each tissue relative to the closest transcriptional start site. A panel for the top 1000 age-related CpGs was added to the figure for

comparison. The changes in each gene region were tested by Fisher's exact test based on the same background. The odds ratios and p-values (* <0.05, **<0.01, ***<0.001, ****<0.0001) of changes are reported for each bar. **(C)** Boxplot of association with mammalian maximum lifespan by human CpG island status. The mean difference was tested by a student's T test. A panel for the top 1000 age-related CpGs was added to the figure for comparison, **(D)** Venn diagram of the overlap in the top 1000 (500 per direction) significant CpGs for different models of EWAS of lifespan.

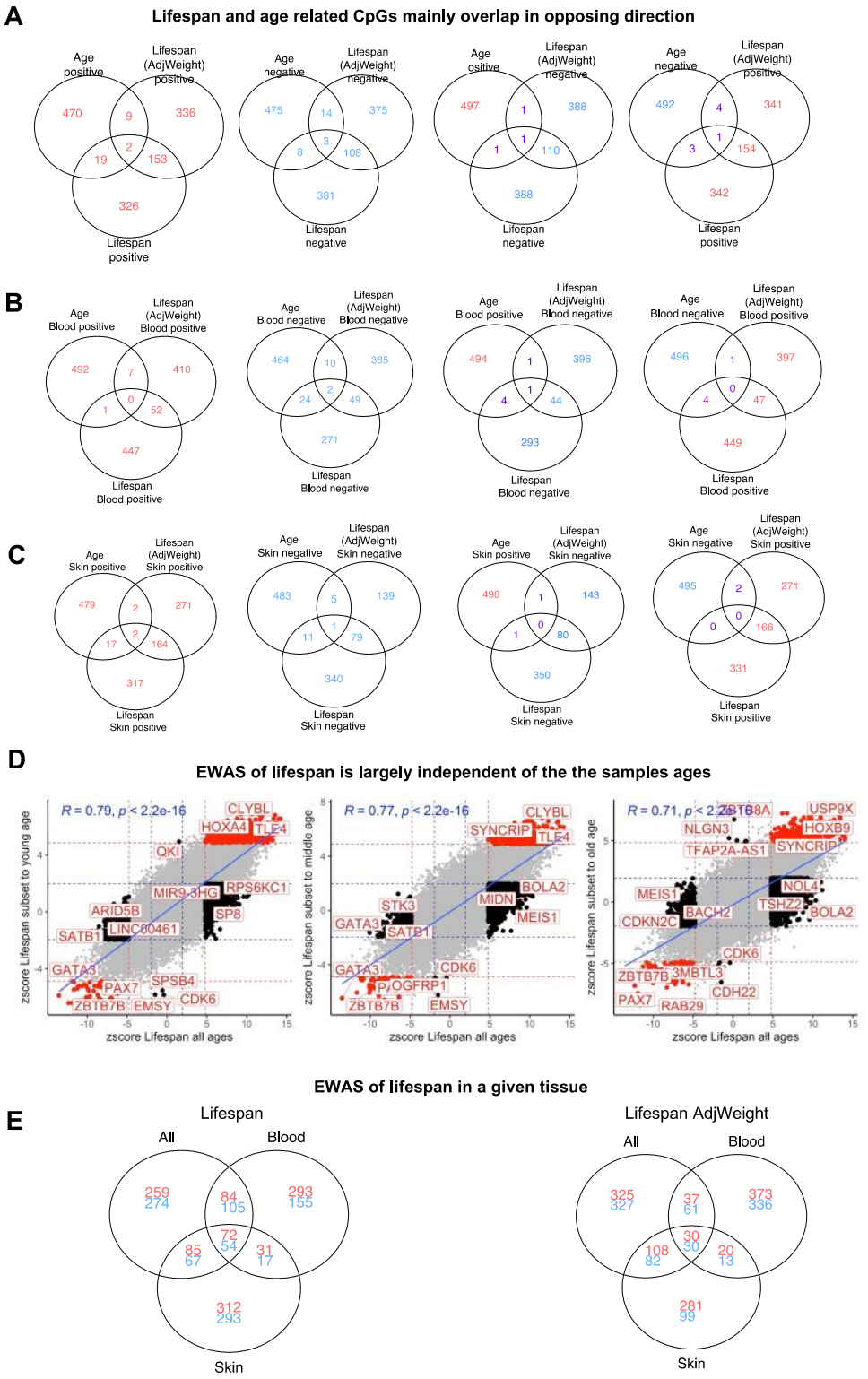


Fig. S20. The relationship of EWAS of lifespan with mammalian age and tissue. Venn diagrams showing the overlap of top 1000 (500 per direction) significant CpGs in EWAS of maximum lifespan, and top age related CpGs based on the meta-analysis of tissues **(A)**, blood **(B)**,

and skin **(C)**. **(D)** Scatter plots of the marginal EWAS of maximum lifespan based on all samples (y axis), and the samples limited to different age groups (x axis). **(E)** Venn diagram of the overlap of top 1000 (500 per direction) significant CpGs in EWAS of maximum lifespan with mean methylation in all tissues, blood, and skin of each species.

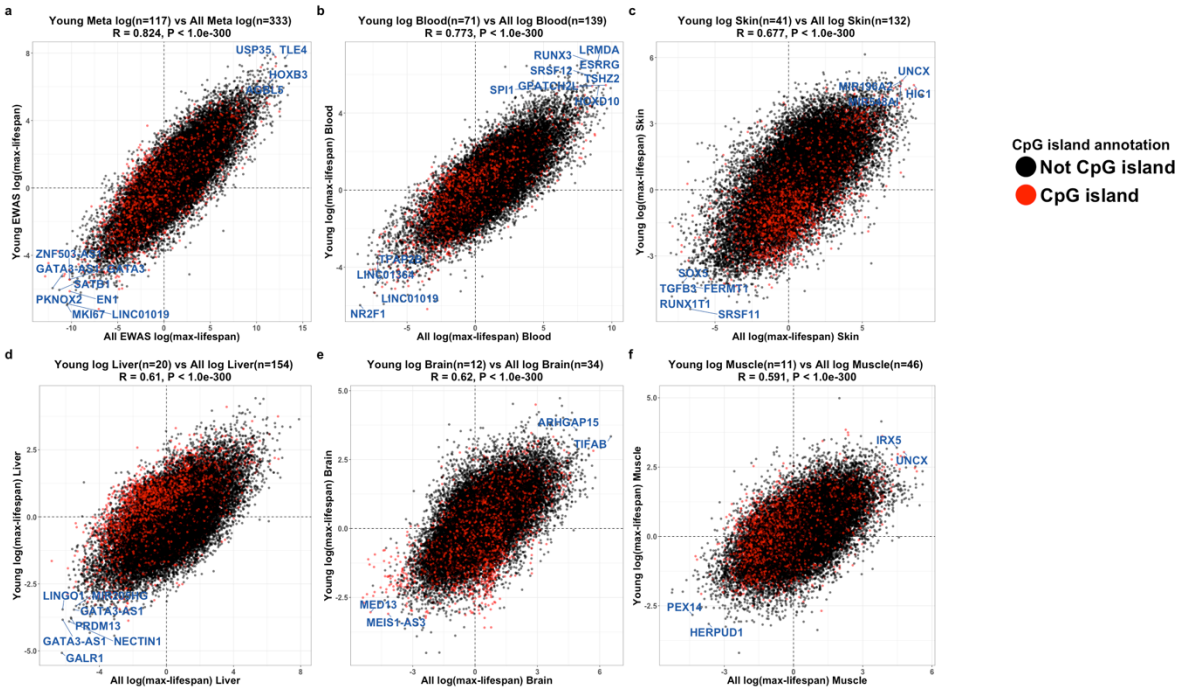


Fig. S21. Generic EWAS agreements between all samples and young samples. Agreements between EWAS based on young samples and EWAS based on all available samples. Young samples are defined as samples younger than five years of age and before the age of sexual maturity. Panels show agreements between, **(A)** all tissue all vs. young generic EWAS, **(B)** all vs. young generic EWAS in blood, **(C)** all vs. young generic EWAS in skin, **(D)** all vs. young generic EWAS in liver, **(E)** all vs. young generic EWAS in brain, **(F)** all vs. young generic EWAS in muscle. Panel titles report r and p as Pearson's correlation and p-values, respectively.

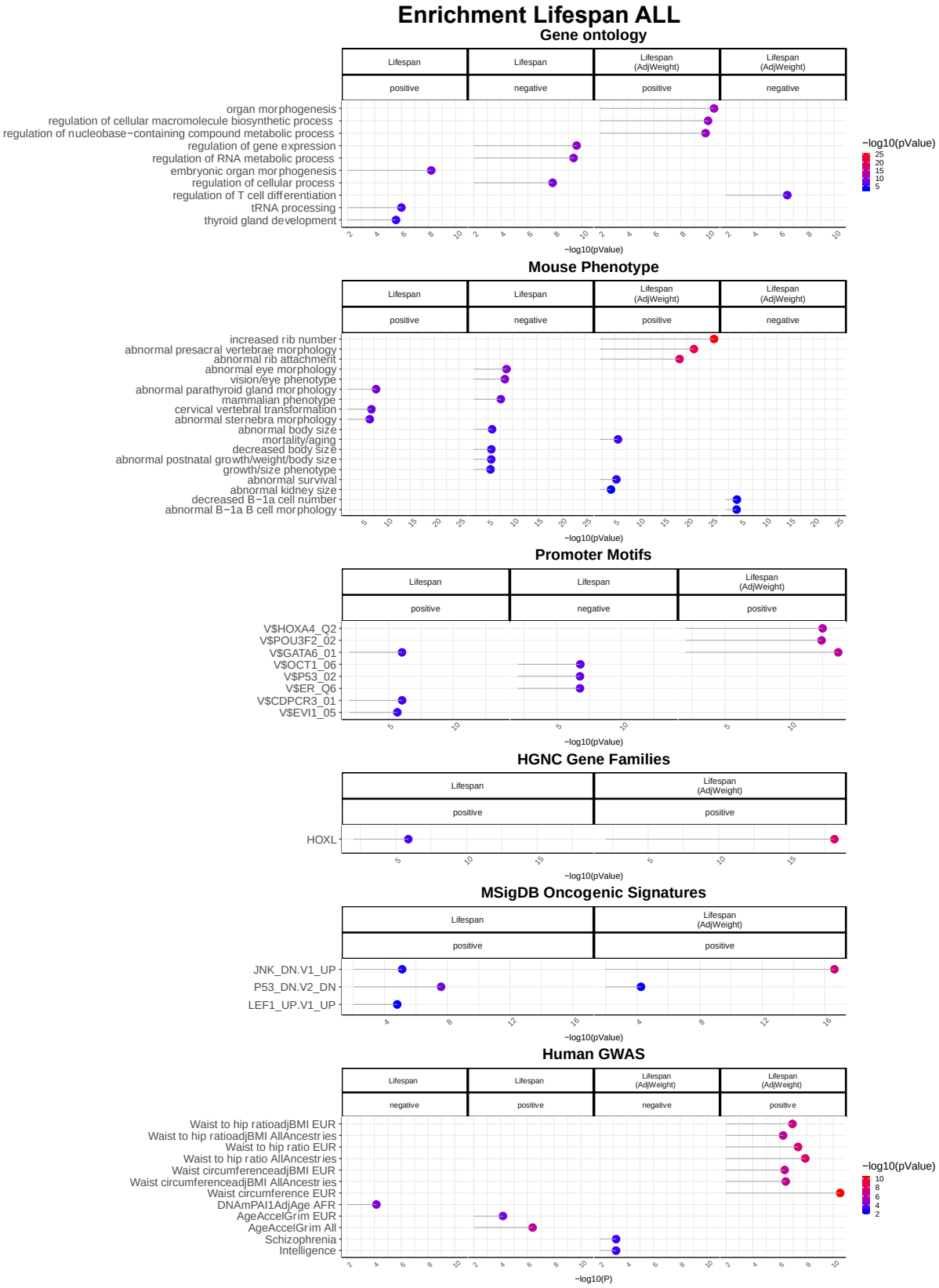

Enrichment Lifespan ALL
Gene ontology
Lifespan
Lifespan
Lifespan (AdjWeight)
Lifespan (AdjWeight)
positive
negative
positive
negative
organ morphogenesis
regulation of cellular macromolecule biosynthetic process
regulation of nucleobase−containing compound metabolic process
regulation of gene expression
regulation of RNA metabolic process
embryonic organ morphogenesis
regulation of cellular process
regulation of T cell differentiation
tRNA processing
thyroid gland development
−log10(pValue)
−log10(pValue)
Mouse Phenotype
increased rib number
abnormal presacral vertebrae morphology
abnormal rib attachment
abnormal eye morphology
vision/eye phenotype
abnormal parathyroid gland morphology
mammalian phenotype
cervical vertebral transformation
abnormal sternebra morphology
abnormal body size
mortality/aging
decreased body size
abnormal postnatal growth/weight/body size
growth/size phenotype
abnormal survival
abnormal kidney size
decreased B−1a cell number
abnormal B−1a B cell morphology
−log10(pValue)
Promoter Motifs
V$HOXA4_Q2
V$POU3F2_02
V$GATA6_01
V$OCT1_06
V$P53_02
V$ER_Q6
V$CDPCR3_01
V$EVI1_05
−log10(pValue)
HGNC Gene Families
HOXL
−log10(pValue)
MSigDB Oncogenic Signatures
JNK_DN.V1_UP
P53_DN.V2_DN
LEF1_UP.V1_UP
−log10(pValue)
Human GWAS
Waist to hip ratioadjBMI EUR
Waist to hip ratioadjBMI AllAncestries
Waist to hip ratio EUR
Waist to hip ratio AllAncestries
Waist circumferenceadjBMI EUR
Waist circumferenceadjBMI AllAncestries
Waist circumference EUR
DNAmPAI1AdjAge AFR
AgeAccelGrim EUR
AgeAccelGrim All
Schizophrenia
Intelligence
−log10(P)
−log10(pValue)

Fig. S22. Gene set enrichment analysis of significant CpGs related to mammalian maximum lifespan. The gene-level enrichment was done using GREAT analysis using human background. Foreground selection is consistent with the description in the methods section. The background probes were limited to the set of EWAS background CpG probes (28,318) consistent with the methods section (mappable to humans and mice and correlation with calibration exceeds 0.8). Human GWAS enrichment was calculated by a hypergeometric test of the top 2.5% genes involved in GWAS of complex traits-associated genes with the top lifespan-related gene regions in our analysis. The biological processes were reduced to parent ontology terms using the "rrvgo" package (Method). Input: Lifespan negative/positive, 500/500 CpGs; Lifespan (AdjWeight) negative/positive, 500/500. In each panel, the columns with no significant terms were removed to simplify the figure. Panels only show entries below a p-value threshold of $p<1x10-4$.

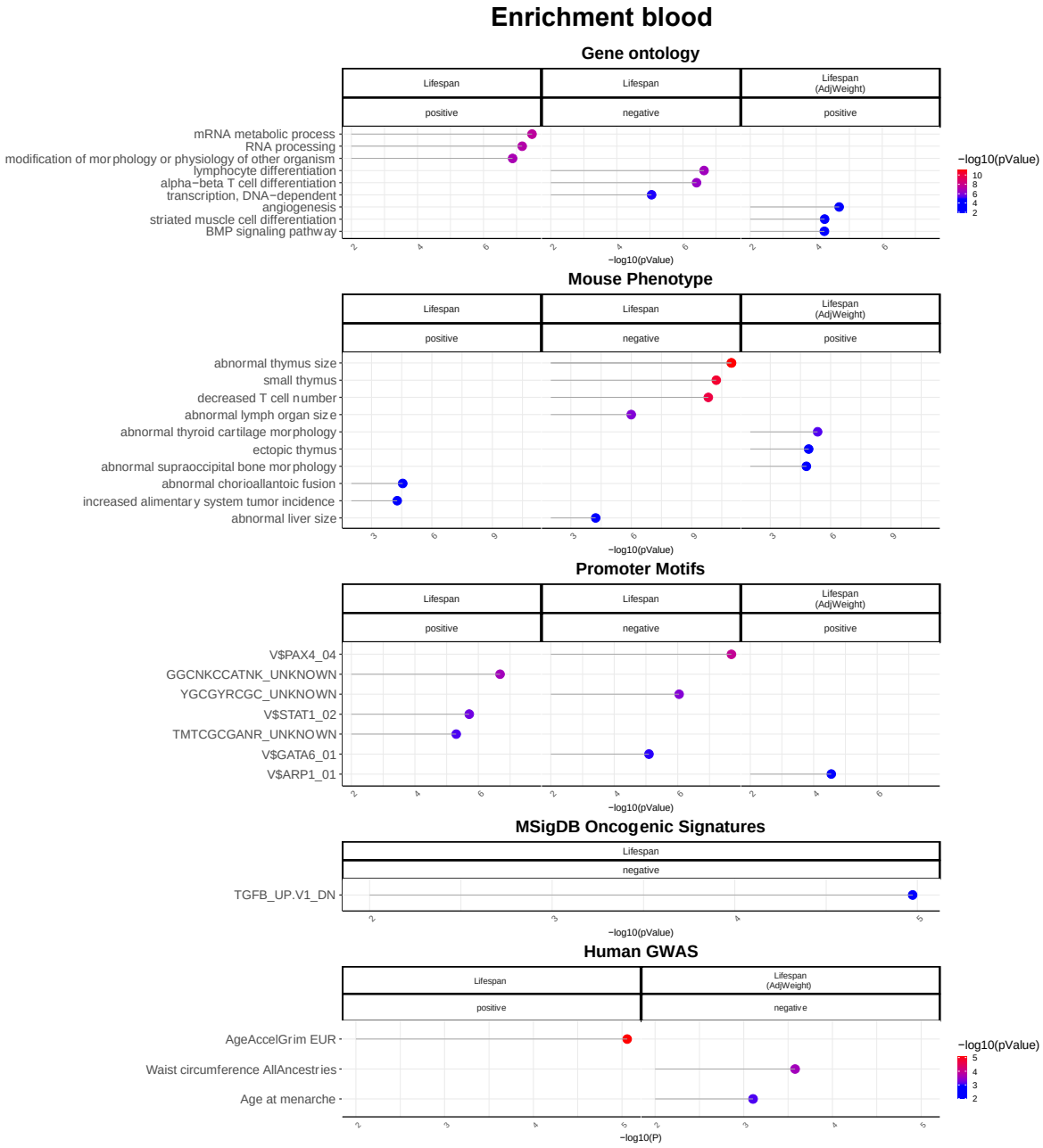


Fig. S23. Gene set enrichment analysis of significant CpGs related to mammalian maximum lifespan in blood. The gene level enrichment was done using GREAT analysis using human background. The background probes were limited to the set of EWAS background CpG probes (28,318) consistent with the methods section (mappable to humans and mice and correlation with calibration exceeds 0.8). Human GWAS enrichment was calculated by a hypergeometric test of the top 5% genomic regions involved in GWAS of complex traits-associated genes with the top lifespan-related gene regions in our analysis. The biological processes were reduced to parent ontology terms using the "rrvgo" package. Input: Lifespan hypo/hyper, 500/500 CpGs; Lifespan (AdjWeight) hypo/hyper, 500/500. In each panel, the columns with no significant terms were removed to simplify the figure. Panels only show entries below a p-value threshold of $p<1x10^{-4}$.

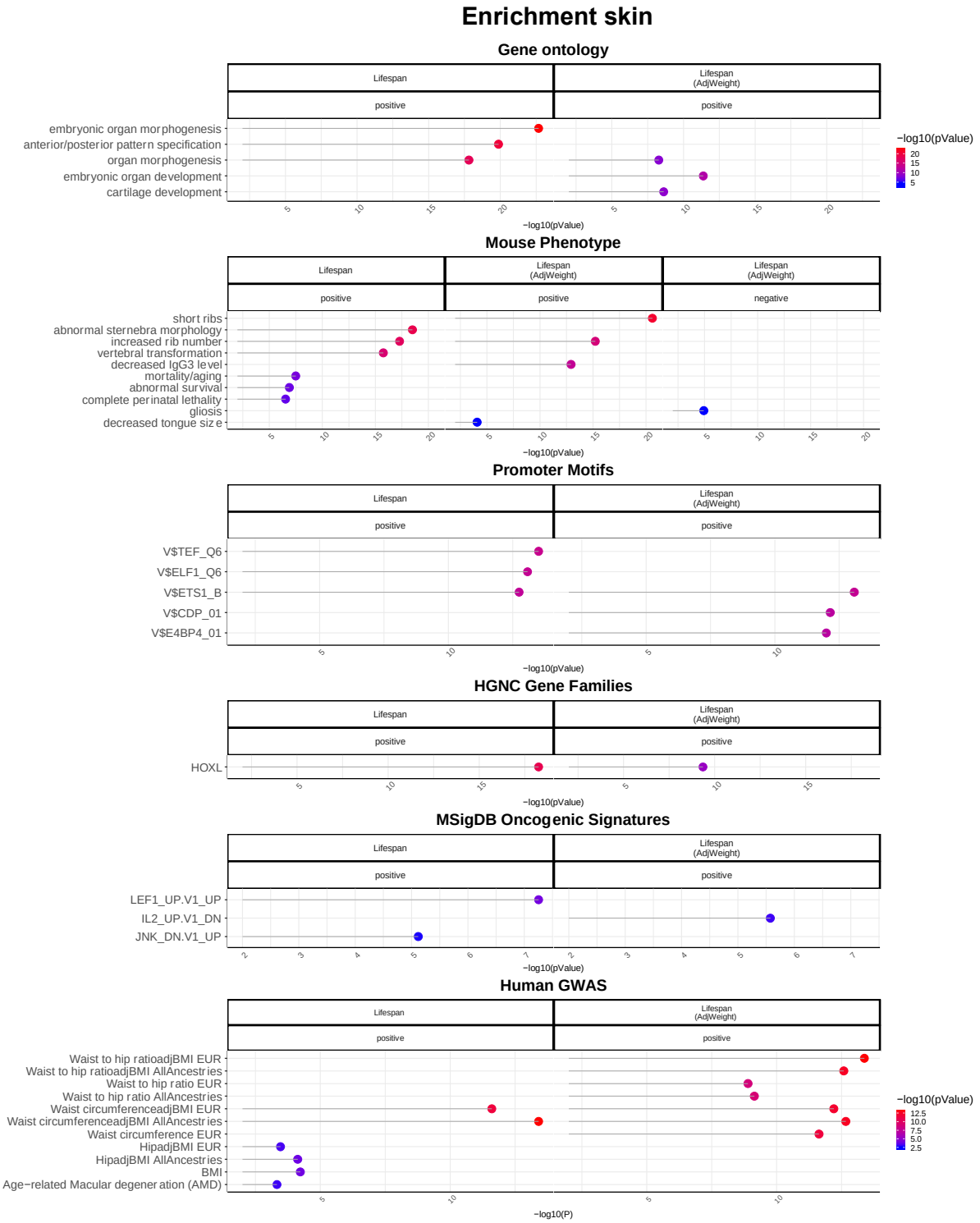


Fig. S24. Gene set enrichment analysis of significant CpGs related to mammalian maximum lifespan in skin. The gene level enrichment was done using GREAT analysis using human background. The background probes were limited to the set of EWAS background CpG probes (28,318) consistent with the methods section (mappable to humans and mice and correlation with calibration exceeds 0.8). Human GWAS enrichment was calculated by a hypergeometric test of the top 2.5% genomic regions involved in GWAS of complex traits-

associated genes with the top lifespan-related gene regions in our analysis. The biological processes were reduced to parent ontology terms using the "rrvgo" package. Input: Lifespan hypo/hyper, 500/500 CpGs; Lifespan (AdjWeight) hypo/hyper, 500/500; Lifespan (AdjPhylo) hypo/hyper, 12/22; Lifespan (AdjPhyloWeight) hypo/hyper, 38/13. In each panel, the columns with no significant terms were removed to simplify the figure. Panels only show entries below a p-value threshold of $p<1x10^{-4}$.

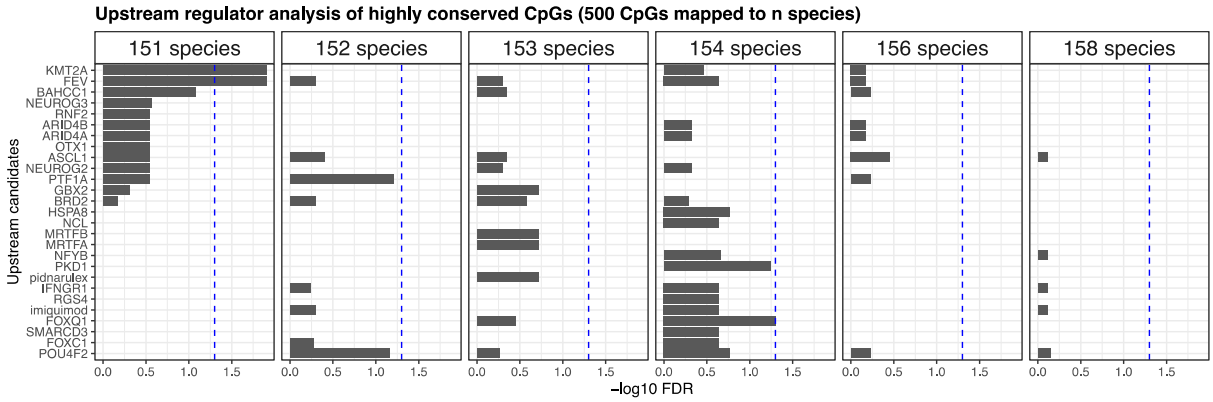


Fig. S25. A sensitivity analysis for potential confounding effects of sequence conservation on the upstream regulator enrichment results. The analysis is done using Ingenuity Pathway Analysis. Different sets of 500 random CpGs are selected based on alignment to 151-158 mammalian species. The background included 28,318 CpGs used for EWAS of maximum lifespan. These probes represent the most conserved CpGs that could be uniquely aligned to most analyzed mammalian genomes. Almost none of the enrichment terms passed the significant threshold (FDR<0.05), confirming that the enrichment analysis is not confounded by sequence conservation.

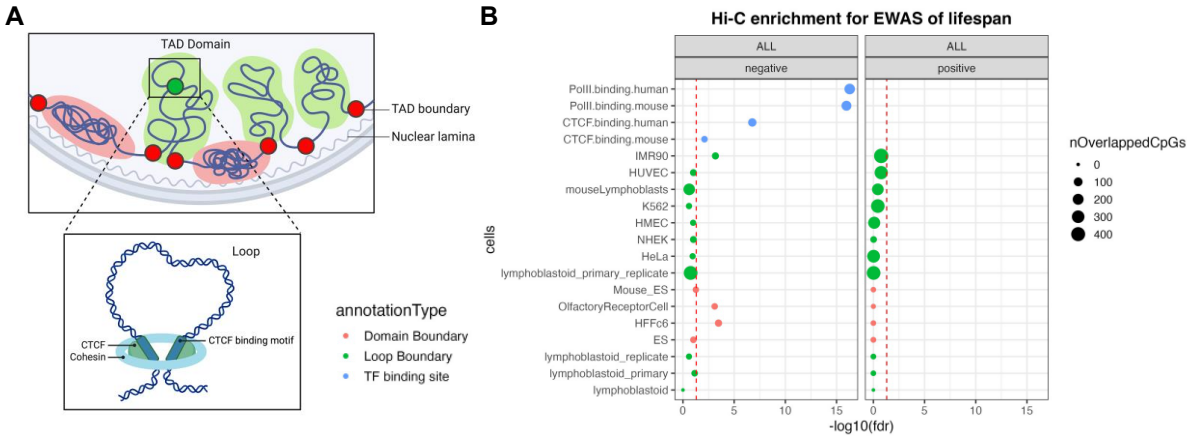


Fig. S26. Enrichment analysis of the Topologically associating domains (TAD) and Loop boundaries for CpGs related to mammalian maximum lifespan. (A) Schematic figure of TADs and loop organizations. **(B)** The curated datasets include summarized Hi-C data generated from 13 human and 2 mouse cell lines. TAD boundary data is downloaded from 4DNucleome portal (Accession IDs: 4DNESU4BQU4G, 4DNESKRDKZ3P, 4DNESWST3UBH, 4DNESFSCP5L8). Loop boundary data is from GSE63525 (*95*). The binding data is downloaded from ENCODE. The false discovery rate (FDR) for each hypergeometric test was calculated per database.

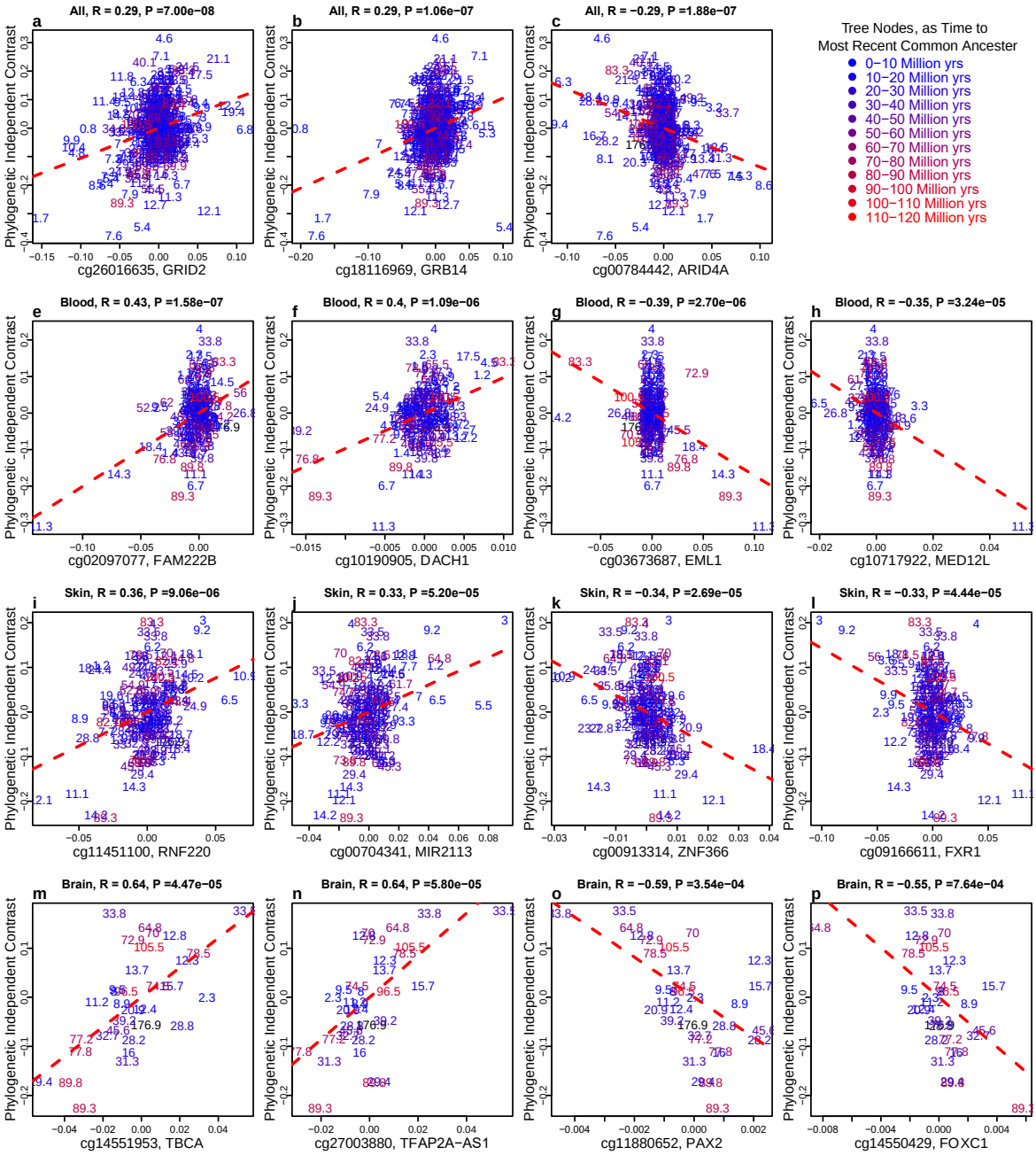


Fig. S27. Top Significant CpG sites in a phylogenetic independent contrast plot, Eutherians. Scatter plot of CpG methylation and maximum lifespan, transformed and scaled to phylogenetic independent contrasts, based on all available samples. To properly visualize sample correlations, phylogenetic independent contrast plots select parent nodes that are of relatively similar distances to each other (*52*). We color-coded these common ancestor nodes as time to present, in millions of years. Panels show scatter plots of top three CpGs from**(A)–(C),** all tissues, **(B)–(G),** top four CpG from blood tissues, **(H)–(K),** top four CpGs from skin tissues, **(L)–(O),** top four CpGs from brain tissues. P-values reported are based on phylogenetic generalized least squared (GLS) regression. Panel titles report r and p as Pearson's correlation and p-values, respectively.

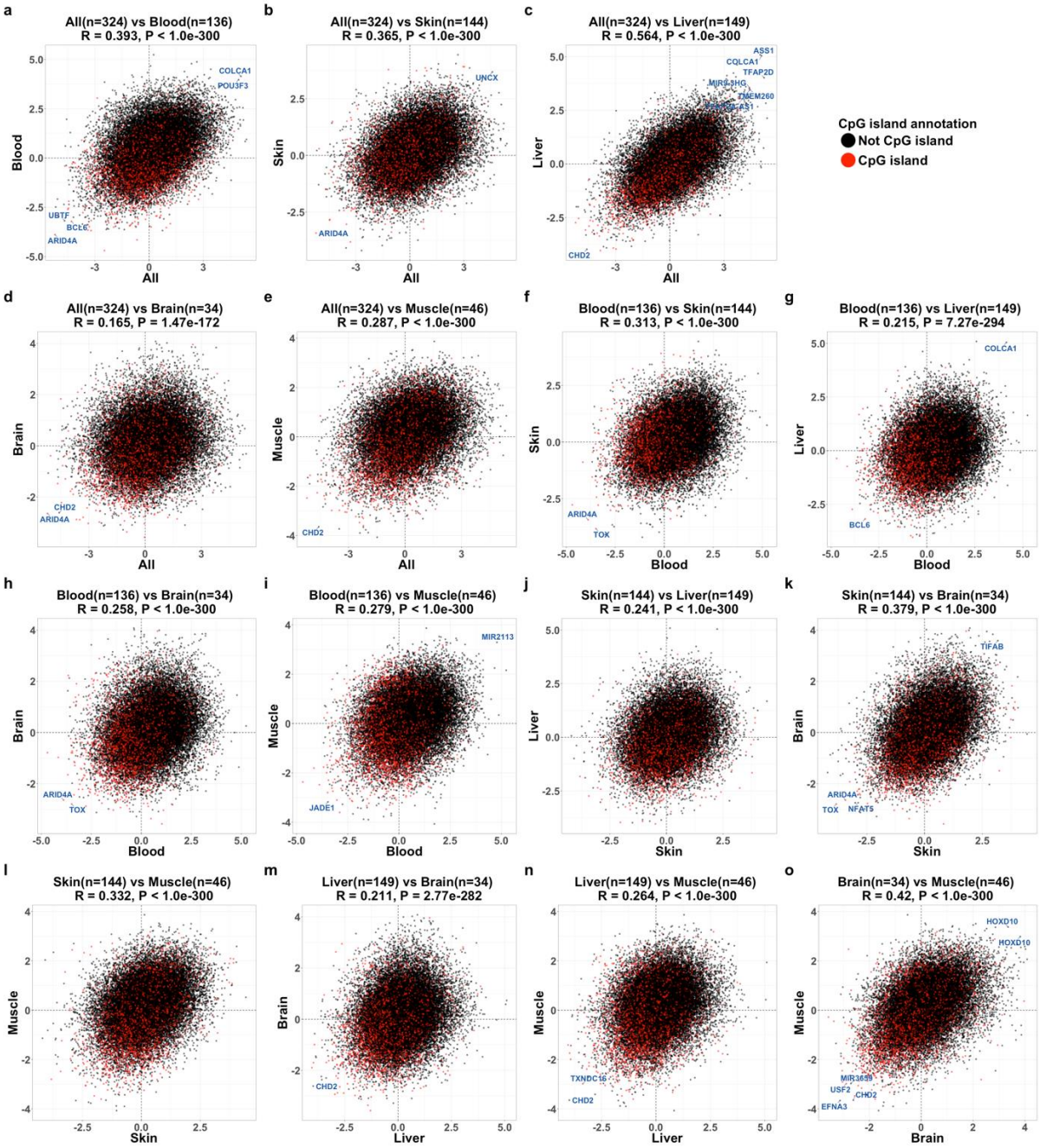


Fig. S28. Phylogenetic EWAS agreement in various tissues, Eutherians. Scatter plot of CpG Z statistics between tissues, color-coded by human CpG island annotations (not island: black, island: red). Both x- and y-axes are CpG Z statistics for the set of EWAS background CpG probes (28,318) consistent with the methods section (mappable to humans and mice and correlation with calibration exceeds 0.8). Agreements are defined identically to **Fig. S3.** Panels show agreements between **(A)** blood vs. all, **(B)** skin vs. all, **(C)** liver vs. all, **(D)** brain vs. all, **(E)** muscle vs. all, **(F)** skin vs. blood, **(G)** liver vs. blood, **(H)** brain vs. blood, **(I)** muscle vs. blood, **(J)** liver vs. skin, **(K)** brain vs. skin, **(I)** muscle vs. skin, **(M)** brain vs. liver, **(N)** muscle vs. liver, **(O)** muscle vs. brain. Panel titles report r and p as Pearson's correlation and p-values, respectively.

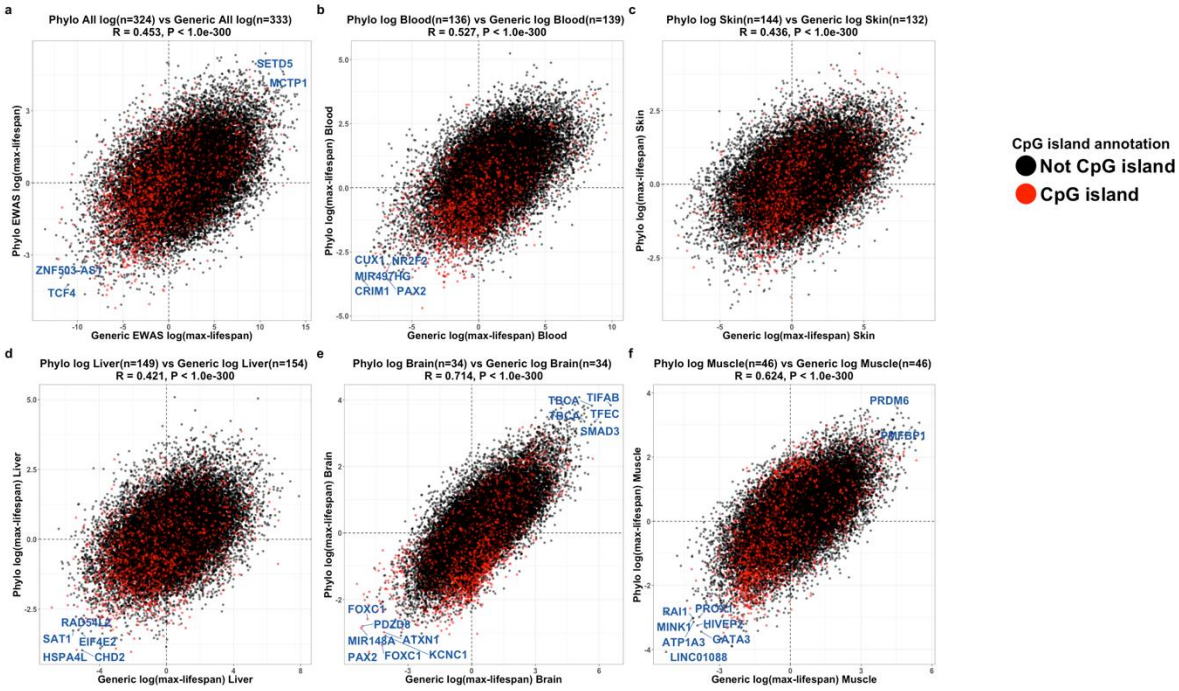


Fig. S29. Simple linear regression (generic) and phylogenetic regression EWAS agreement. Scatter plot of CpG Z statistics across phylogenetic Generic EWAS vs. Phylogenetic EWAS. Panel titles and axes labels report agreements between EWAS analyses. Panels show agreements between, **(A)** all tissue phylogenetic vs. generic EWAS, **(B)** phylogenetic vs. generic EWAS in blood, **(C)** phylogenetic vs. generic EWAS in skin, **(D)** phylogenetic vs. generic EWAS in liver, **(E)** phylogenetic vs. generic EWAS in brain, **(F)** phylogenetic vs. generic EWAS in muscle. Panel titles report r and p as Pearson's correlation and p-values, respectively.

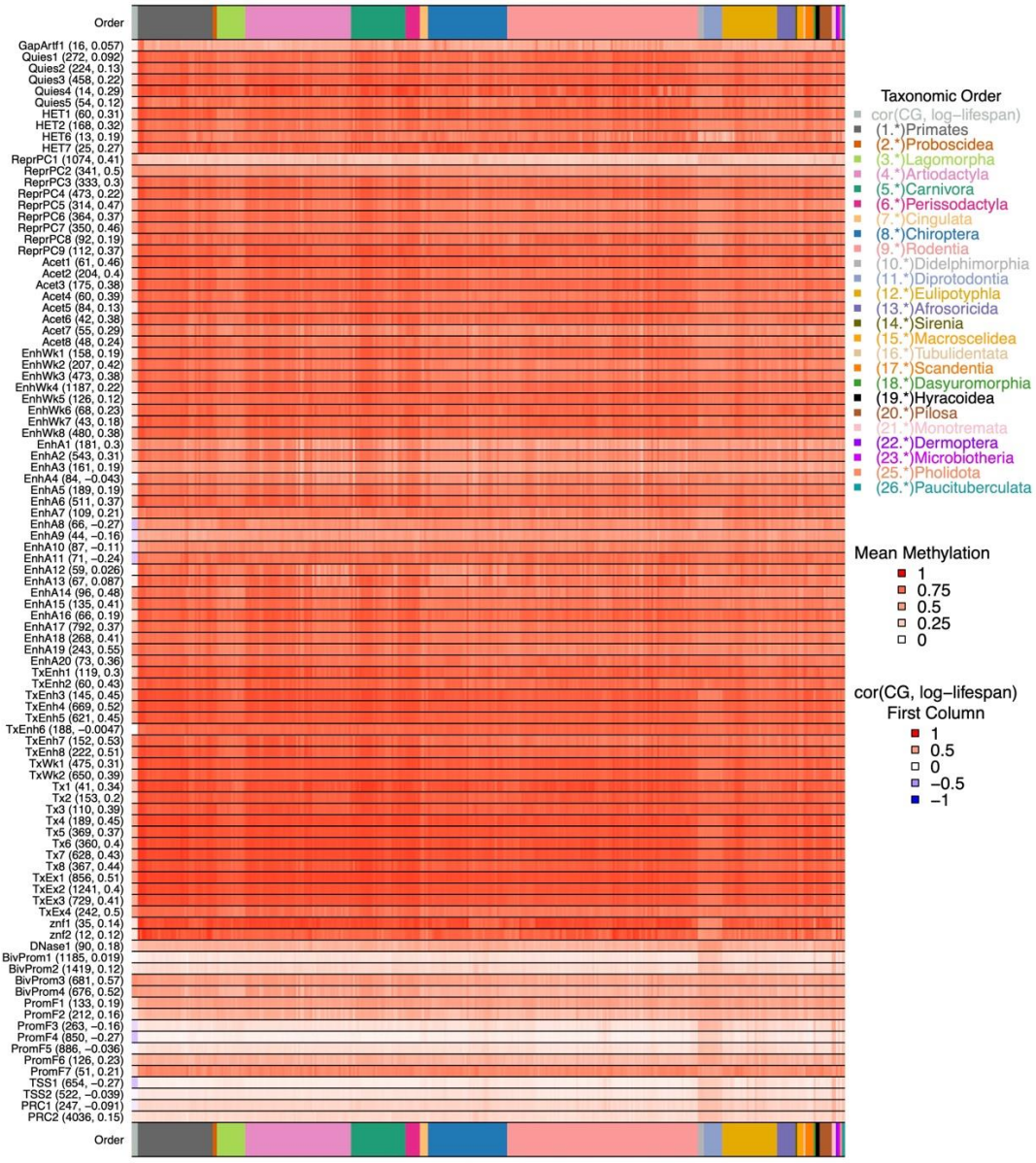


Fig. S30. Chromatin State Mean Methylation Level and Correlations with Lifespan

Heatmap of mean methylation levels of CpGs belonging to each chromatin state from the universal chromatin state annotation (*41*), stratified by species. The rows are chromatin states with more than 10 mammalian CpG sites. Each column is a mammalian species, grouped and color-coded by taxonomic order, shown in color bands on the top and bottom of the heatmap. Each cell (except cells in the first column), ranges from 0 to 1, representing the mean methylation level of each chromatin state's CpGs (filtered by probes considered in our EWAS, all probes mappable to human and mouse, and correlation exceeding 0.8 in calibration/titration data) in the given species. The first column on the left shows correlations between mean CpG methylation in each chromatin state and log-transformed maximum lifespan, which ranges from -1 to 1. Values in each row's parentheses are in the format of (number of CpGs in the chromatin state, Pearson's correlation with lifespan).

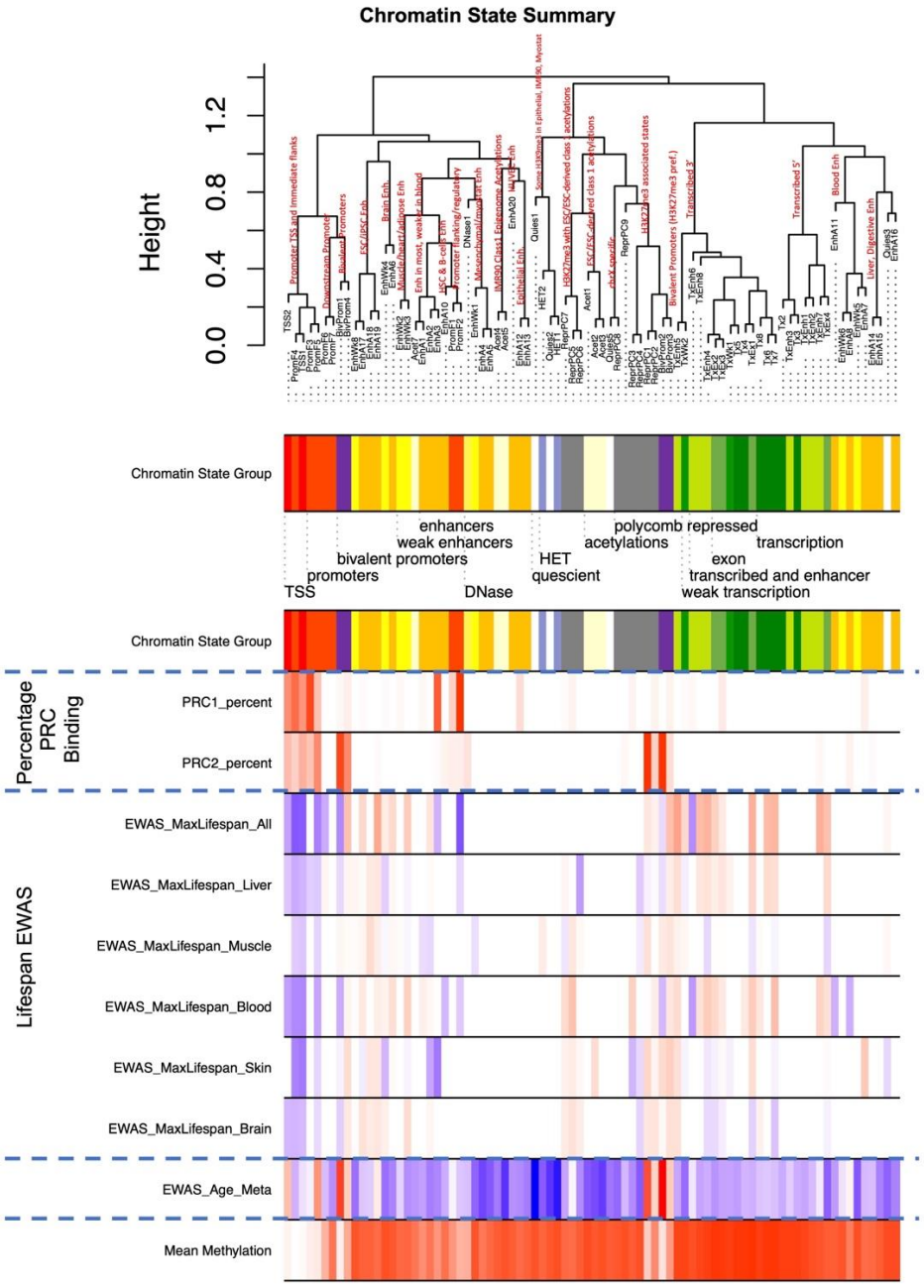


Fig. S31. Emission profile of different chromatin states and relationship with EWAS of lifespan. Figure summarizes EWAS Z statistics from different analyses (rows) by chromatin states. Chromatin states (columns) are clustered by their emission profiles. Descriptions of clusters of states are shown on the dendrogram. Like the chromatin enrichment analysis, only the top 500 positively (red shaded color in rows) and negatively (blue shaded color in rows) associated CpGs in each chromatin state are aggregated to form mean Z statistics for this figure. The EWAS results are from the marginal analysis of log lifespan with mean methylation levels in all tissues combined (EWAS_MaxLifespan_All) or different tissues.

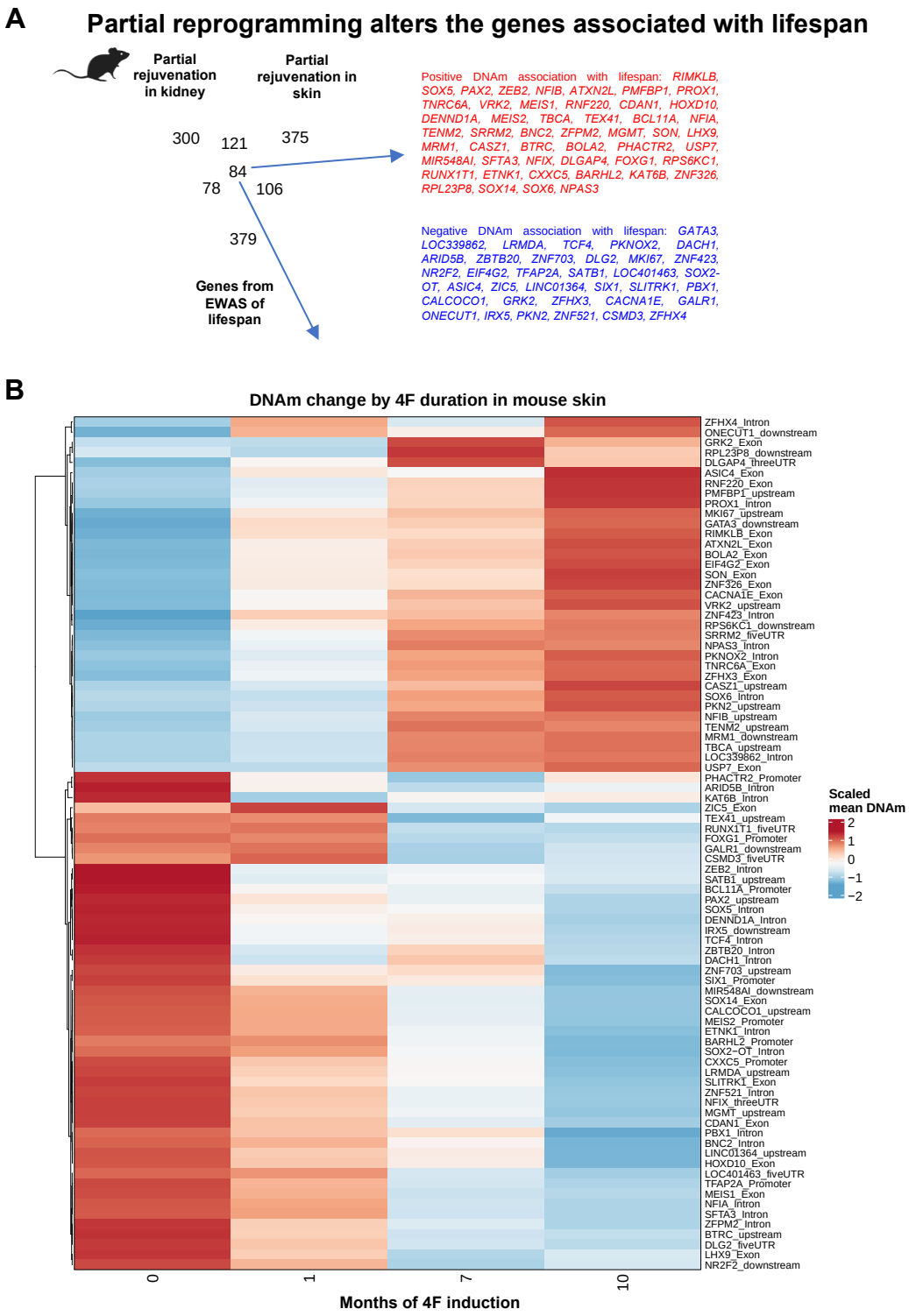


Fig. S32. Maximum lifespan genes are partially regulated by Yamanaka factors. (A) Venn diagram of 3 gene lists. First, the top 647 genes adjacent to 1000 lifespan related CpGs (500 positive and 500 negative). Gene lists 2 and 3 are based on CpGs that are differentially methylated (nominal Wald test p<0.005, up to 500 positive and 500 negatively related CpGs) after OSKM overexpression in murine kidney (583 genes) and skin (686 genes) (*39*). We observe significant

overlap between the gene lists (nominal Fisher exact p= $9.9x10^{-30}$ for skin and lifespan; p= $4.5x10^{-25}$ for kidney and lifespan) (**Fig. 4**) (*39*). **(B)** Heatmap of the scaled mean methylation changes in the skin of mice with different duration of 4F treatment. The row names indicate the gene region of the top CpG adjacent to each gene. The methylation value of each CpG is adjusted for chronological age differences between the four groups of this study.

Captions for Supplementary tables

Table S1.

Summary of samples in this study. The table reports minimum age, maximum age, median age. Whenever possible, our Mammalian Methylation Consortium profiled tissues from animals of the entire age range (from birth to old age). For some species, the age information was not available.

Table S2.

CpGs in Mammalian co-methylation networks. The table includes network colors from net1 (eutherian network), net2 (eutherian+marsupial network), and 7 consensus networks (cNet). The cNet are based on different mammalian species or tissues combinations as indicated in the column names. The CpG coordinates for human Hg19 and mouse MM10 genomes are also resporeted in the table.

Table S3.

Summary of module-trait association analyses. The table summarizes several characteristics of network 1 (eutherian network) modules and eigengene association analysis with different mammalian traits. The column names summarize the analyzed trait (e.g., Age, Phylogenetic Order) and model used for the analysis. In addition, the table shows different enrichment analysis for the CpGs within the module. The hub genes are defined based on module connectivity (kME).

Table S4.

Enrichment analysis for CpGs in different modules. The gene level enrichment was done using GREAT analysis using human background. The background probes were limited to the set of eutherian probes used for network development.

Table S5.

Module-CpG connectivity (kME) in network 1.

Table S6.

DNAm-mRNA association in horse tissues. The z scores are calculated from Fisher Z transformation of Pearson correlation coefficients.

Table S7.

EWAS-GWAS enrichment in network 1 modules. The intra-module hub CpGs were also statistically tested for overlap with human GWAS results, in which gene p values were calculated by MAGENTA algorithm.

Table S8.

Summary of dog breed samples.

Table S9.

Canonical pathway enrichment analysis of the network 1 mammalian modules using IPA software.

Table S10.

Upstream regulator analysis of the network 1 mammalian modules using IPA software.

Table S11.

EWAS-TWAS enrichment analysis for age-related TWAS datasets in GEO.

Table S12.

Selected Significant CpG sites, Generic log Maximum Lifespan EWAS, Eutherians. Sample size (n) indicates the number of species included in species-wise EWAS; Coordinates are start end coordinates for the nearest gene on human hg19 genome; Z statistics are rounded to the second significant decimal place, while p values are rounded to the first significant decimal place, as the power term is more relevant in comparing significance in EWAS. Z statistics is calculated by sign(t) * abs(qnorm(p/2, lower.tail = TRUE)) formula.

Table S13.

Selected Significant CpG sites, Generic log Species Maximum Lifespan EWAS, adjusted by Adult Body Weight, Eutherians. Sample size (n) indicates the number of species included in species-wise EWAS; Coordinates are start end coordinates for the nearest gene on human hg19 genome; Z statistics are rounded to the second significant decimal place, while p values are rounded to the first significant decimal place, as the power term is more relevant in comparing significance in EWAS. Z statistics is calculated by sign(t) * abs(qnorm(p/2, lower.tail = TRUE)) formula.

Table S14.

Selected Significant CpG sites, Phylogenetic log Species Maximum Lifespan EWAS, Eutherians. Sample size (n) indicates the number of species included in species-wise EWAS; Coordinates are start end coordinates for the nearest gene on human hg19 genome; Z statistics are rounded to the second significant decimal place, while p values are rounded to the first significant decimal place, as the power term is more relevant in comparing significance in EWAS. Z statistics is calculated by sign(t) * abs(qnorm(p/2, lower.tail = TRUE)) formula.

Table S15.

Selected Significant CpG sites, Phylogenetic log Species Maximum Lifespan EWAS, adjusted by adult body weight, Eutherians. Sample size (n) indicates the number of species included in species-wise EWAS; Coordinates are start end coordinates for the nearest gene on human hg19 genome; Z statistics are rounded to the second significant decimal place, while p values are rounded to the first significant decimal place, as the power term is more relevant in comparing significance in EWAS. Z statistics is calculated by sign(t) * abs(qnorm(p/2, lower.tail = TRUE)) formula.

Table S16.

Gene set enrichment analysis of different EWAS of lifespan. Class indicates the EWAS module: lifespan=Generic EWAS of log maximum lifespan; lifespanAdjW= Generic EWAS of log maximum lifespan adjusted for average weight of species. Group indicates the positive or negative relationship of CpGs in each EWAS. Only top 500 CpGs per direction in each EWAS were selected for the enrichment analysis. The gene level enrichment was done using GREAT analysis using human background. The background probes were limited to the set of human/mouse probes.

Table S17.

Reduced GO Terms GREAT Enrichment, log Species Maximum Lifespan EWAS, Eutherians. The biological processes were reduced to parent ontology terms using the "rrvgo" package.

Table S18.

Significant RNA processing Genes in Generic Lifespan EWAS.

Table S19.

Ingenuity Pathway Analysis (IPA) - Upstream Regulators of Generic Lifespan EWAS. The input is the top 1000 CpGs (500 negative and 500 positive) from EWAS of log lifespan in eutherians.

Table S20.

EWAS-GWAS enrichment analysis of the top hits from EWAS of lifespan. "EWAS" column indicates the EWAS module: lifespan=Generic EWAS of log maximum lifespan; lifespanAdjW= Generic EWAS of log maximum lifespan adjusted for average weight of species. "Class" column indicates the positive or negative relationship of CpGs in each EWAS. Only top 500 CpGs per direction in each EWAS were selected for the enrichment analysis.

Table S21.

Summary table of the universal chromatin state annotation.

Table S22.

Chromatin state enrichment analysis of modules and EWAS results.
Column Module/EWAS, reports the input list of CpGs. For example, the CpGs of a given module were selected based on the module eigengene based membership measure (kME). The EWAS results were selected based on a Wald test Z statistic resulting from a marginal model. All input lists were restricted to less than 500 CpGs.
Column state reports the universal chromatin state (*41*).
Column pval reports the nominal/uncorrected two-sided hypergeometric test p-value.
Column foldEnrichment reports the ratio of observed count divided by expected count.
Column nCommon reports the number of shared CpGs between the chromatin state and input CpGs.
Column totalCGwithState reports the total number of background CpGs (e.g., eutherian probes for modules or 28k mouse/human probes for EWAS probes with the respective chromatin state.

Table S23.

Histone Enrichment Analysis of the top hits from EWAS of lifespan. The histone annotations are defined based on ChIPseq data from ENCODE.

Table S24.

The top 1000 CpGs (500 per direction) based on Rank-pvalue meta-analysis of EWAS of lifespan in 25 eutherian tissue-order strata. The full list and additional statistical output are reported in Data S19.

Captions for Supplementary data

Data S1.

Mean methylation values per species in Blood.

Data S2.

Mean methylation values per species in Liver.

Data S3.

Mean methylation values per species in Skin.

Data S4.

Pearson correlation matrices for mammalian DNA methylation in Blood. The correlation matrix is calculated from the mean methylation values per species in blood.

Data S5.

Pearson correlation matrices for mammals in Liver. The correlation matrix is calculated from the mean methylation values per species in liver.

Data S6.

Pearson correlation matrices for mammals in Skin. The correlation matrix is calculated from the mean methylation values per species in skin.

Data S7.

Spearman correlation matrices for mammals in Blood. The correlation matrix is calculated from the mean methylation values per species in blood.

Data S8.

Spearman correlation matrices for mammals in Liver. The correlation matrix is calculated from the mean methylation values per species in liver.

Data S9.

Spearman correlation matrices for mammals in Skin. The correlation matrix is calculated from the mean methylation values per species in skin.

Data S10.

Bicor correlation matrices for mammals in Blood. The correlation matrix is calculated from the mean methylation values per species in blood.

Data S11.

Bicor correlation matrices for mammals in Liver. The correlation matrix is calculated from the mean methylation values per species in liver.

Data S12.

Bicor correlation matrices for mammals in Skin. The correlation matrix is calculated from the mean methylation values per species in skin.

Data S13.

Eutherians EWAS, in which the cells are Pearson's correlation between each mean methylation levels of each CpG per species and log(maximum lifespan), formula: cor(log-maximum lifespan, CpG); the column name indicate the tissue samples, the phylogenetic order, and number of species that was used for each EWAS. For example, "Generic.EWAS.Log.maxLifespan.OrderALL.TissueALL.N333" mean a generic correlation analysis of mean methylation levels of all tissues to log maximum lifespan of 333 eutherian species from all phylogenetic orders.

Data S14.

P-value results of the EWAS of mammalian lifespan in different phylogenetic orders and tissues of Eutherians. The column name follows the pattern as Data S13.

Data S15.

Pearson correlation results of the EWAS of mammalian lifespan (Adjusted for average weight) in different phylogenetic orders and tissues of Eutherians. The column name follows the pattern as Data S13.

Data S16.

P-value results of the EWAS of mammalian lifespan (Adjusted for average weight) in different phylogenetic orders and tissues of Eutherians**.** The column name follows the pattern as Data S13.

Data S17.

Z-score results of the EWAS of mammalian lifespan (phylogenetic regression) in different phylogenetic orders and tissues of Eutherians. In column name, PGLS indicates phylogenetic regression. The column name follows the pattern as Data S13.

Data S18.

P-value results of the EWAS of mammalian lifespan (phylogenetic regression) in different phylogenetic orders and tissues of Eutherians. In column name, PGLS indicates phylogenetic regression. The column name follows the pattern as Data S13.

Data S19.

Rank-pvalue meta-analysis of EWAS of lifespan in 25 eutherian tissue-order strata (Data S13).

References and Notes

1. S. Xiao, D. Xie, X. Cao, P. Yu, X. Xing, C.-C. Chen, M. Musselman, M. Xie, F. D. West, H. A. Lewin, T. Wang, S. Zhong, Comparative epigenomic annotation of regulatory DNA. *Cell* **149**, 1381–1392 (2012). doi:10.1016/j.cell.2012.04.029 Medline

2. D. Villar, C. Berthelot, S. Aldridge, T. F. Rayner, M. Lukk, M. Pignatelli, T. J. Park, R. Deaville, J. T. Erichsen, A. J. Jasinska, J. M. A. Turner, M. F. Bertelsen, E. P. Murchison, P. Flicek, D. T. Odom, Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015). doi:10.1016/j.cell.2015.01.006 Medline

3. J. Qu, E. Hodges, A. Molaro, P. Gagneux, M. D. Dean, G. J. Hannon, A. D. Smith, Evolutionary expansion of DNA hypomethylation in the mammalian germline genome. *Genome Res.* **28**, 145–158 (2018). doi:10.1101/gr.225896.117 Medline

4. J. Klughammer, D. Romanovskaia, A. Nemc, A. Posautz, C. A. Seid, L. C. Schuster, M. C. Keinath, J. S. Lugo Ramos, L. Kosack, A. Evankow, D. Printz, S. Kirchberger, B. Ergüner, P. Datlinger, N. Fortelny, C. Schmidl, M. Farlik, K. Skjærven, A. Bergthaler, M. Liedvogel, D. Thaller, P. A. Burger, M. Hermann, M. Distel, D. L. Distel, A. Kübber-Heiss, C. Bock, Comparative analysis of genome-scale, base-resolution DNA methylation profiles across 580 animal species. *Nat. Commun.* **14**, 232 (2023). doi:10.1038/s41467-022-34828-y Medline

5. A. Arneson, A. Haghani, M. J. Thompson, M. Pellegrini, S. B. Kwon, H. Vu, E. Maciejewski, M. Yao, C. Z. Li, A. T. Lu, M. Morselli, L. Rubbi, B. Barnes, K. D. Hansen, W. Zhou, C. E. Breeze, J. Ernst, S. Horvath, A mammalian methylation array for profiling methylation levels at conserved sequences. *Nat. Commun.* **13**, 783 (2022). doi:10.1038/s41467-022-28355-z Medline

6. K. M. Parsons, A. Haghani, J. A. Zoller, A. T. Lu, Z. Fei, S. H. Ferguson, E. Garde, M. B. Hanson, C. K. Emmons, C. O. Matkin, B. G. Young, W. R. Koski, S. Horvath, DNA methylation-based biomarkers for ageing long-lived cetaceans. *Mol. Ecol. Resour.* **23**, 1241–1256 (2023). doi:10.1111/1755-0998.13791 Medline

7. A. T. Lu, Z. Fei, A. Haghani, T. R. Robeck, J. A. Zoller, C. Z. Li, R. Lowe, Q. Yan, J. Zhang, H. Vu, J. Ablaeva, V. A. Acosta-Rodriguez, D. M. Adams, J. Almunia, A. Aloysius, R. Ardehali, A. Arneson, C. S. Baker, G. Banks, K. Belov, N. C. Bennett, P. Black, D. T. Blumstein, E. K. Bors, C. E. Breeze, R. T. Brooke, J. L. Brown, G. Carter, A. Caulton, J. M. Cavin, L. Chakrabarti, I. Chatzistamou, H. Chen, K. Cheng, P. Chiavellini, O. W. Choi, S. Clarke, L. N. Cooper, M. L. Cossette, J. Day, J. DeYoung, S. DiRocco, C. Dold, E. E. Ehmke, C. K. Emmons, S. Emmrich, E. Erbay, C. Erlacher-Reid, C. G. Faulkes, S. H. Ferguson, C. J. Finno, J. E. Flower, J. M. Gaillard, E. Garde, L. Gerber, V. N. Gladyshev, V. Gorbunova, R. G. Goya, M. J. Grant, C. B. Green, E. N. Hales, M. B. Hanson, D. W. Hart, M. Haulena, K. Herrick, A. N. Hogan, C. J. Hogg, T. A. Hore, T. Huang, J. C. Izpisua Belmonte, A. J. Jasinska, G. Jones, E. Jourdain, O. Kashpur, H. Katcher, E. Katsumata, V. Kaza, H. Kiaris, M. S. Kobor, P. Kordowitzki, W. R. Koski, M. Kruetzen, S. B. Kwon, B. Larison, S. G. Lee, M. Lehmann, J. F. Lemaitre, A. J. Levine, C. Li, X. Li, A. R. Lim, D. T. S. Lin, D. M. Lindemann, T. J. Little, N. Macoretta, D. Maddox, C. O. Matkin, J. A. Mattison, M. McClure, J. Mergl, J. J. Meudt, G. A. Montano, K. Mozhui, J. Munshi-South, A. Naderi, M. Nagy, P. Narayan, P. W. Nathanielsz, N. B. Nguyen, C. Niehrs, J. K. O'Brien, P. O'Tierney Ginn, D. T. Odom, A.

G. Ophir, S. Osborn, E. A. Ostrander, K. M. Parsons, K. C. Paul, M. Pellegrini, K. J. Peters, A. B. Pedersen, J. L. Petersen, D. W. Pietersen, G. M. Pinho, J. Plassais, J. R. Poganik, N. A. Prado, P. Reddy, B. Rey, B. R. Ritz, J. Robbins, M. Rodriguez, J. Russell, E. Rydkina, L. L. Sailer, A. B. Salmon, A. Sanghavi, K. M. Schachtschneider, D. Schmitt, T. Schmitt, L. Schomacher, L. B. Schook, K. E. Sears, A. W. Seifert, A. Seluanov, A. B. A. Shafer, D. Shanmuganayagam, A. V. Shindyapina, M. Simmons, K. Singh, I. Sinha, J. Slone, R. G. Snell, E. Soltanmaohammadi, M. L. Spangler, M. C. Spriggs, L. Staggs, N. Stedman, K. J. Steinman, D. T. Stewart, V. J. Sugrue, B. Szladovits, J. S. Takahashi, M. Takasugi, E. C. Teeling, M. J. Thompson, B. Van Bonn, S. C. Vernes, D. Villar, H. V. Vinters, M. C. Wallingford, N. Wang, R. K. Wayne, G. S. Wilkinson, C. K. Williams, R. W. Williams, X. W. Yang, M. Yao, B. G. Young, B. Zhang, Z. Zhang, P. Zhao, Y. Zhao, W. Zhou, J. Zimmermann, J. Ernst, K. Raj, S. Horvath, Universal DNA methylation age across mammalian tissues. bioRxiv 426733 [Preprint] (2021); https://doi.org/10.1101/2021.01.18.426733.

8. P. Kordowitzki, A. Haghani, J. A. Zoller, C. Z. Li, K. Raj, M. L. Spangler, S. Horvath, Epigenetic clock and methylation study of oocytes from a bovine model of reproductive aging. *Aging Cell* **20**, e13349 (2021). doi:10.1111/acel.13349 Medline

9. N. A. Prado, J. L. Brown, J. A. Zoller, A. Haghani, M. Yao, L. R. Bagryanova, M. G. Campana, J. E Maldonado, K. Raj, D. Schmitt, T. R. Robeck, S. Horvath, Epigenetic clock and methylation studies in elephants. *Aging Cell* **20**, e13414 (2021). doi:10.1111/acel.13414 Medline

10. T. R. Robeck, Z. Fei, A. T. Lu, A. Haghani, E. Jourdain, J. A. Zoller, C. Z. Li, K. J. Steinman, S. DiRocco, T. Schmitt, S. Osborn, B. Van Bonn, E. Katsumata, J. Mergl, J. Almunia, M. Rodriguez, M. Haulena, C. Dold, S. Horvath, Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Commun. Biol.* **4**, 642 (2021). doi:10.1038/s42003-021-02179-x Medline

11. B. Larison, G. M. Pinho, A. Haghani, J. A. Zoller, C. Z. Li, C. J. Finno, C. Farrell, C. B. Kaelin, G. S. Barsh, B. Wooding, T. R. Robeck, D. Maddox, M. Pellegrini, S. Horvath, Epigenetic models developed for plains zebras predict age in domestic horses and endangered equids. *Commun. Biol.* **4**, 1412 (2021). doi:10.1038/s42003-021-02935-z Medline

12. K. Mozhui, A. T. Lu, C. Z. Li, A. Haghani, J. V. Sandoval-Sierra, Y. Wu, R. W. Williams, S. Horvath, Genetic loci and metabolic states associated with murine epigenetic aging. *eLife* **11**, e75244 (2022). doi:10.7554/eLife.75244 Medline

13. V. J. Sugrue, J. A. Zoller, P. Narayan, A. T. Lu, O. J. Ortega-Recalde, M. J. Grant, C. S. Bawden, S. R. Rudiger, A. Haghani, D. M. Bond, R. R. Hore, M. Garratt, K. E. Sears, N. Wang, X. W. Yang, R. G. Snell, T. A. Hore, S. Horvath, Castration delays epigenetic aging and feminizes DNA methylation at androgen-regulated loci. *eLife* **10**, e64932 (2021). doi:10.7554/eLife.64932 Medline

14. T. R. Robeck, Z. Fei, A. Haghani, J. A. Zoller, C. Z. Li, K. J. Steinman, S. Dirocco, L. Staggs, T. Schmitt, S. Osborn, G. Montano, M. Rodriguez, S. Horvath, Multi-tissue methylation clocks for age and sex estimation in the common bottlenose dolphin. *Front. Mar. Sci.* **8**, 713373 (2021). doi:10.3389/fmars.2021.713373

15. S. Horvath, A. Haghani, J. A. Zoller, A. Naderi, E. Soltanmohammadi, E. Farmaki, V. Kaza, I. Chatzistamou, H. Kiaris, Methylation studies in *Peromyscus*: Aging, altitude adaptation, and monogamy. *Geroscience* **44**, 447–461 (2022). doi:10.1007/s11357-021-00472-5 Medline

16. S. Horvath, A. Haghani, J. A. Zoller, K. Raj, I. Sinha, T. R. Robeck, P. Black, A. Couzens, C. Lau, M. Manoyan, Y. A. Ruiz, A. Talbott, K. Belov, C. J. Hogg, K. E. Sears, Epigenetic clock and methylation studies in marsupials: Opossums, Tasmanian devils, kangaroos, and wallabies. *Geroscience* **44**, 1825–1845 (2022). doi:10.1007/s11357-022-00569-5 Medline

17. S. Horvath, J. A. Zoller, A. Haghani, A. J. Jasinska, K. Raj, C. E. Breeze, J. Ernst, K. L. Vaughan, J. A. Mattison, Epigenetic clock and methylation studies in the rhesus macaque. *Geroscience* **43**, 2441–2453 (2021). doi:10.1007/s11357-021-00429-8 Medline

18. S. Horvath, J. A. Zoller, A. Haghani, A. T. Lu, K. Raj, A. J. Jasinska, J. A. Mattison, A. B. Salmon, DNA methylation age analysis of rapamycin in common marmosets. *Geroscience* **43**, 2413–2425 (2021). doi:10.1007/s11357-021-00438-7 Medline

19. A. J. Jasinska, A. Haghani, J. A. Zoller, C. Z. Li, A. Arneson, J. Ernst, K. Kavanagh, M. J. Jorgensen, J. A. Mattison, K. Wojta, O.-W. Choi, J. DeYoung, X. Li, A. W. Rao, G. Coppola, N. B. Freimer, R. P. Woods, S. Horvath, Epigenetic clock and methylation studies in vervet monkeys. *Geroscience* **44**, 699–717 (2021). doi:10.1007/s11357-021-00466-3 Medline

20. K. Raj, B. Szladovits, A. Haghani, J. A. Zoller, C. Z. Li, P. Black, D. Maddox, T. R. Robeck, S. Horvath, Epigenetic clock and methylation studies in cats. *Geroscience* **43**, 2363–2378 (2021). doi:10.1007/s11357-021-00445-8 Medline

21. K. M. Schachtschneider, L. B. Schook, J. J. Meudt, D. Shanmuganayagam, J. A. Zoller, A. Haghani, C. Z. Li, J. Zhang, A. Yang, K. Raj, S. Horvath, Epigenetic clock and DNA methylation analysis of porcine models of aging and obesity. *Geroscience* **43**, 2467–2483 (2021). doi:10.1007/s11357-021-00439-6 Medline

22. M. L. Cossette, D. T. Stewart, A. Haghani, J. A. Zoller, A. B. A. Shafer, S. Horvath, Epigenetics and island-mainland divergence in an insectivorous small mammal. *Mol. Ecol.* **32**, 152–166 (2023). doi:10.1111/mec.16735 Medline

23. J. F. Lemaître, B. Rey, J.-M. Gaillard, C. Régis, E. Gilot-Fromont, F. Débias, J. Duhayer, S. Pardonnet, M. Pellerin, A. Haghani, J. A. Zoller, C. Z. Li, S. Horvath, DNA methylation as a tool to explore ageing in wild roe deer populations. *Mol. Ecol. Resour.* **22**, 1002–1015 (2022). doi:10.1111/1755-0998.13533 Medline

24. S. Horvath, A. Haghani, N. Macoretta, J. Ablaeva, J. A. Zoller, C. Z. Li, J. Zhang, M. Takasugi, Y. Zhao, E. Rydkina, Z. Zhang, S. Emmrich, K. Raj, A. Seluanov, C. G. Faulkes, V. Gorbunova, DNA methylation clocks tick in naked mole rats but queens age more slowly than nonbreeders. *Nat. Aging* **2**, 46–59 (2022). doi:10.1038/s43587-021-00152-1 Medline

25. 25. S. Horvath, A. Haghani, J. A. Zoller, A. T. Lu, J. Ernst, M. Pellegrini, A. J. Jasinska, J. A. Mattison, A. B. Salmon, K. Raj, M. Horvath, K. C. Paul, B. R. Ritz, T. R. Robeck, M. Spriggs, E. E. Ehmke, S. Jenkins, C. Li, P. W. Nathanielsz, Pan-primate studies of age and sex. *Geroscienc e* (2023). doi:10.1007/s11357-023-00878-3 Medline

26. G. S. Wilkinson, D. M. Adams, A. Haghani, A. T. Lu, J. Zoller, C. E. Breeze, B. D. Arnold, H. C. Ball, G. G. Carter, L. N. Cooper, D. K. N. Dechmann, P. Devanna, N. J. Fasel, A. V. Galazyuk, L. Günther, E. Hurme, G. Jones, M. Knörnschild, E. Z. Lattenkamp, C. Z. Li, F. Mayer, J. A. Reinhardt, R. A. Medellin, M. Nagy, B. Pope, M. L. Power, R. D. Ransome, E. C. Teeling, S. C. Vernes, D. Zamora-Mejías, J. Zhang, P. A. Faure, L. J. Greville, L. G. Herrera M, J. J. Flores-Martínez, S. Horvath, DNA methylation predicts age and provides insight into exceptional longevity of bats. *Nat. Commun.* **12**, 1615 (2021). doi:10.1038/s41467-021-21900-2 Medline

27. S. Horvath, A. Haghani, S. Peng, E. N. Hales, J. A. Zoller, K. Raj, B. Larison, T. R. Robeck, J. L. Petersen, R. R. Bellone, C. J. Finno, DNA methylation aging and transcriptomic studies in horses. *Nat. Commun.* **13**, 40 (2022). doi:10.1038/s41467-021-27754-y Medline

28. G. M. Pinho, J. G. A. Martin, C. Farrell, A. Haghani, J. A. Zoller, J. Zhang, S. Snir, M. Pellegrini, R. K. Wayne, D. T. Blumstein, S. Horvath, Hibernation slows epigenetic ageing in yellow-bellied marmots. *Nat. Ecol. Evol.* **6**, 418–426 (2022). doi:10.1038/s41559-022-01679-1 Medline

29. S. Horvath, A. T. Lu, A. Haghani, J. A. Zoller, C. Z. Li, A. R. Lim, R. T. Brooke, K. Raj, A. Serres-Armero, D. L. Dreger, A. N. Hogan, J. Plassais, E. A. Ostrander, DNA methylation clocks for dogs and humans. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2120887119 (2022). doi:10.1073/pnas.2120887119 Medline

30. K. J. Peters, L. Gerber, L. Scheu, R. Cicciarella, J. A. Zoller, Z. Fei, S. Horvath, S. J. Allen, S. L. King, R. C. Connor, L. A. Rollins, M. Krützen, An epigenetic DNA methylation clock for age estimates in Indo-Pacific bottlenose dolphins (*Tursiops aduncus*). *Evol. Appl.* **16**, 126–133 (2022). Medline

31. P. Chiavellini, M. Lehmann, M. Canatelli Mallat, J. A. Zoller, C. B. Herenu, G. R. Morel, S. Horvath, R. G. Goya, Hippocampal DNA methylation, epigenetic age and spatial memory performance in young and old rats. *J. Gerontol. A Biol. Sci. Med. Sci.* **77**, 2387–2394 (2022). doi:10.1093/gerona/glac153 Medline

32. S. P. Blomberg, T. Garland Jr., A. R. Ives, Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* **57**, 717–745 (2003). Medline

33. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). doi:10.1186/1471-2105-9-559 Medline

34. P. Langfelder, R. Luo, M. C. Oldham, S. Horvath, Is my network module preserved and reproducible? *PLOS Comput. Biol.* **7**, e1001057 (2011). doi:10.1371/journal.pcbi.1001057 Medline

35. D. Szklarczyk, A. L. Gable, K. C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, L. J. Jensen, C. von Mering, Correction to 'The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets'. *Nucleic Acids Res.* **49**, 10800 (2021). doi:10.1093/nar/gkab835 Medline

36. A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 3747–3752 (2004). doi:10.1073/pnas.0400087101 Medline

37. H. Pilcher, Money for old mice. *Nature* 10.1038/news030915-13 (2003). doi:10.1038/news030915-13

38. V. Acosta-Rodríguez, F. Rijo-Ferreira, M. Izumo, P. Xu, M. Wight-Carter, C. B. Green, J. S. Takahashi, Circadian alignment of early onset caloric restriction promotes longevity in male C57BL/6J mice. *Science* **376**, 1192–1202 (2022). doi:10.1126/science.abk0297 Medline

39. K. C. Browder, P. Reddy, M. Yamamoto, A. Haghani, I. G. Guillen, S. Sahu, C. Wang, Y. Luque, J. Prieto, L. Shi, K. Shojima, T. Hishida, Z. Lai, Q. Li, F. K. Choudhury, W. R. Wong, Y. Liang, D. Sangaraju, W. Sandoval, C. R. Esteban, E. N. Delicado, P. G. Garcia, M. Pawlak, J. A. Vander Heiden, S. Horvath, H. Jasper, J. C. Izpisua Belmonte, In vivo partial reprogramming alters age-associated molecular changes during physiological aging in mice. *Nat. Aging* **2**, 243–253 (2022). doi:10.1038/s43587-022-00183-2 Medline

40. A. Krämer, J. Green, J. Pollard Jr., S. Tugendreich, Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530 (2014). doi:10.1093/bioinformatics/btt703 Medline

41. H. Vu, J. Ernst, Universal annotation of the human genome through integration of over a thousand epigenomic datasets. *Genome Biol.* **23**, 9 (2022). doi:10.1186/s13059-021-02572-z Medline

42. W. Zhou, H. Q. Dinh, Z. Ramjan, D. J. Weisenberger, C. M. Nicolet, H. Shen, P. W. Laird, B. P. Berman, DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* **50**, 591–602 (2018). doi:10.1038/s41588-018-0073-4 Medline

43. P. Langfelder, S. Horvath, "Tutorials for the WGCNA package" (2014); https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/Tutorials.

44. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017). doi:10.1093/molbev/msx116 Medline

45. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi:10.1038/nature11247 Medline

46. A. Haghani, A. T. Lu, S. B. Kwon, A. Arneson, J. Ernst, S. Horvath, Data for: DNA methylation networks underlying mammalian traits, Zenodo (2023); https://doi.org/10.5281/zenodo.8180547.

47. J. P. de Magalhães, J. Costa, G. M. Church, An analysis of the relationship between metabolism, developmental schedules, and longevity using phylogenetic independent contrasts. *J. Gerontol. A Biol. Sci. Med. Sci.* **62**, 149–160 (2007). doi:10.1093/gerona/62.2.149 Medline

48. W. Zhou, T. J. Triche Jr., P. W. Laird, H. Shen, SeSAMe: Reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Res.* **46**, e123 (2018). doi:10.1093/nar/gky691 Medline

49. S. B. Hedges, J. Marin, M. Suleski, M. Paymer, S. Kumar, Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.* **32**, 835–845 (2015). doi:10.1093/molbev/msv037 Medline

50. J. Felsenstein, Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* **25**, 471–492 (1973). Medline

51. S. W. Kembel, P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, C. O. Webb, Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010). doi:10.1093/bioinformatics/btq166 Medline

52. J. Felsenstein, Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15 (1985). doi:10.1086/284325

53. A. Grafen, The phylogenetic regression. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **326**, 119–157 (1989). doi:10.1098/rstb.1989.0106 Medline

54. C. Y. McLean, D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010). doi:10.1038/nbt.1630 Medline

55. S. Sayols, rrvgo: A Bioconductor package to reduce and visualize Gene Ontology terms. *Aust. Dent. J.* (2020).

56. A. V. Segrè, L. Groop, V. K. Mootha, M. J. Daly, D. Altshuler; DIAGRAM Consortium; MAGIC investigators, Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLOS Genet.* **6**, e1001058 (2010). doi:10.1371/journal.pgen.1001058 Medline

57. S. Horvath, A. Haghani, J. A. Zoller, A. T. Lu, J. Ernst, M. Pellegrini, A. J. Jasinska, J. A. Mattison, A. B. Salmon, K. Raj, M. Horvath, K. C. Paul, B. R. Ritz, T. R. Robeck, M. Spriggs, E. E. Ehmke, S. Jenkins, C. Li, P. W. Nathanielsz, Pan-primate DNA methylation clocks. bioRxiv 402891 [Preprint] (2020); https://doi.org/10.1101/2020.11.29.402891.

58. N. E. Schlabritz-Loutsevitch, C. J. Dudley, J. J. Gomez, C. H. Nevill, B. K. Smith, S. L. Jenkins, T. J. McDonald, T. Q. Bartlett, P. W. Nathanielsz, M. J. Nijland, Metabolic adjustments to moderate maternal nutrient restriction. *Br. J. Nutr.* **98**, 276–284 (2007). doi:10.1017/S0007114507700727 Medline

59. S. Morgello, B. B. Gelman, P. B. Kozlowski, H. V. Vinters, E. Masliah, M. Cornford, W. Cavert, C. Marra, I. Grant, E. J. Singer; The National NeuroAIDS Tissue Consortium, A new paradigm in brain banking with an emphasis on infectious disease. *Neuropathol. Appl. Neurobiol.* **27**, 326–335 (2001). doi:10.1046/j.0305-1846.2001.00334.x Medline

60. S. Horvath, D. T. S. Lin, M. S. Kobor, J. A. Zoller, J. W. Said, S. Morgello, E. Singer, W. H. Yong, B. D. Jamieson, A. J. Levine, HIV, pathology and epigenetic age acceleration in different human tissues. *Geroscience* **44**, 1609–1620 (2022). doi:10.1007/s11357-022-00560-0 Medline

61. S. Horvath, D. J. Stein, N. Phillips, S. J. Heany, M. S. Kobor, D. T. S. Lin, L. Myer, H. J. Zar, A. J. Levine, J. Hoare, Perinatally acquired HIV infection accelerates epigenetic aging in South African adolescents. *AIDS* **32**, 1465–1474 (2018). doi:10.1097/QAD.0000000000001854 Medline

62. S. Horvath, B. R. Ritz, Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging* **7**, 1130–1142 (2015). doi:10.18632/aging.100859 Medline

63. S. Kabacik, S. Horvath, H. Cohen, K. Raj, Epigenetic ageing is distinct from senescence-mediated ageing and is not prevented by telomerase expression. *Aging (Albany NY)* **10**, 2800–2815 (2018). doi:10.18632/aging.101588 Medline

64. L. L. Sailer, A. Haghani, J. A. Zoller, C. Z. Li, A. G. Ophir, S. Horvath, Pair bonding slows epigenetic aging and alters methylation in brains of prairie voles. bioRxiv 313775 [Preprint] (2020); https://doi.org/10.1101/2020.09.25.313775.

65. E. N. Burns, M. H. Bordbari, M. J. Mienaltowski, V. K. Affolter, M. V. Barro, F. Gianino, G. Gianino, E. Giulotto, T. S. Kalbfleisch, S. A. Katzman, M. Lassaline, T. Leeb, M. Mack, E. J. Müller, J. N. MacLeod, B. Ming-Whitfield, C. R. Alanis, T. Raudsepp, E. Scott, S. Vig, H. Zhou, J. L. Petersen, R. R. Bellone, C. J. Finno, Generation of an equine biobank to be used for Functional Annotation of Animal Genomes project. *Anim. Genet.* **49**, 564–570 (2018). doi:10.1111/age.12717 Medline

66. L. Tan, Z. Ke, G. Tombline, N. Macoretta, K. Hayes, X. Tian, R. Lv, J. Ablaeva, M. Gilbert, N. V. Bhanu, Z.-F. Yuan, B. A. Garcia, Y. G. Shi, Y. Shi, A. Seluanov, V. Gorbunova, Naked mole rat cells have a stable epigenome that resists iPSC reprogramming. *Stem Cell Reports* **9**, 1721–1734 (2017). doi:10.1016/j.stemcr.2017.10.001 Medline

67. E. K. Bors, C. S. Baker, P. R. Wade, K. B. O'Neill, K. E. W. Shelden, M. J. Thompson, Z. Fei, S. Jarman, S. Horvath, An epigenetic clock to estimate the age of living beluga whales. *Evol. Appl.* **14**, 1263–1273 (2021). doi:10.1111/eva.13195 Medline

68. S. Lubetkin, J. Zeh, J. George, Statistical modeling of baleen and body length at age in bowhead whales (*Balaena mysticetus*). *Can. J. Zool.* **90**, 915–931 (2012). doi:10.1139/z2012-057

69. W. Koski, J. George, J. Zeh, J. Brandon, "Preliminary analyses on identifying yearling bowhead whales (*Balaena mysticetus*) in aerial photographs" (International Whaling Consortium, 2010), Report to the IWC Scientific Committee.

70. E. Garde, M. P. Heide-Jørgensen, S. H. Hansen, G. Nachman, M. C. Forchhammer, Age-specific growth and remarkable longevity in narwhals (*Monodon monoceros*) from West Greenland as estimated by aspartic acid racemization. *J. Mammal.* **88**, 49–58 (2007). doi:10.1644/06-MAMM-A-056R.1

71. C. Rosa, J. Zeh, J. Craig George, O. Botta, M. Zauscher, J. Bada, T. M. O'Hara, Age estimates based on aspartic acid racemization for bowhead whales (*Balaena mysticetus*) harvested in 1998–2000 and the relationship between racemization rate and body temperature. *Mar. Mamm. Sci.* **29**, 424–445 (2013). doi:10.1111/j.1748-7692.2012.00593.x

72. E. H. Harley, M. H. Knight, C. Lardner, B. Wooding, M. Gregor, The Quagga project: Progress over 20 years of selective breeding. *S. Afr. J. Wildl. Res* **39**, 155–163 (2009). doi:10.3957/056.039.0206

73. S. Horvath, K. Singh, K. Raj, S. Khairnar, A. Sanghavi, A. Shrivastava, J. A. Zoller, C. Z. Li, C. B. Herenu, M. Canatelli-Mallat, M. Lehmann, L. C. Solberg Woods, A. Garcia Martinez, T. Wang, P. Chiavellini, A. J. Levine, H. Chen, R. G. Goya, H. L. Katcher,

Reversing age: dual species measurement of epigenetic age with a single clock. bioRxiv 082917 [Preprint] (2020); https://doi.org/10.1101/2020.05.07.082917.

74. J. Plassais, J. Kim, B. W. Davis, D. M. Karyadi, A. N. Hogan, A. C. Harris, B. Decker, H. G. Parker, E. A. Ostrander, Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* **10**, 1489 (2019). doi:10.1038/s41467-019-09373-w Medline

75. J. Plassais, M. Rimbault, F. J. Williams, B. W. Davis, J. J. Schoenebeck, E. A. Ostrander, Analysis of large versus small dogs reveals three genes on the canine X chromosome associated with body weight, muscling and back fat thickness. *PLOS Genet.* **13**, e1006661 (2017). doi:10.1371/journal.pgen.1006661 Medline

76. American Kennel Club, *The Complete Dog Book: 20th Edition* (Ballantine Books, 2006).

77. B. Wilcox, C. Walkowicz, *The Atlas of Dog Breeds of the World*. (T.F.H. Publications, 1995).

78. K. T. Coschigano, A. N. Holland, M. E. Riders, E. O. List, A. Flyvbjerg, J. J. Kopchick, Deletion, but not antagonism, of the mouse growth hormone receptor results in severely decreased body weights, insulin, and insulin-like growth factor I levels and increased life span. *Endocrinology* **144**, 3799–3810 (2003). doi:10.1210/en.2003-0374 Medline

79. V. A. Acosta-Rodríguez, F. Rijo-Ferreira, C. B. Green, J. S. Takahashi, Importance of circadian timing for aging and longevity. *Nat. Commun.* **12**, 2862 (2021). doi:10.1038/s41467-021-22922-6 Medline

80. T. J. Little, A. N. O'Toole, A. Rambaut, T. Chandra, R. Marioni, A. B. Pederson, Methylation-based age estimation in a wild mouse. bioRxiv 203687 [Preprint] (2020); https://doi.org/10.1101/2020.07.16.203687.

81. C. Berthelot, D. Villar, J. E. Horvath, D. T. Odom, P. Flicek, Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat. Ecol. Evol.* **2**, 152–163 (2018). doi:10.1038/s41559-017-0377-2 Medline

82. M. Roller, E. Stamper, D. Villar, O. Izuogu, F. Martin, A. M. Redmond, R. Ramachanderan, L. Harewood, D. T. Odom, P. Flicek, LINE retrotransposons characterize mammalian tissue-specific and evolutionarily dynamic regulatory regions. *Genome Biol.* **22**, 62 (2021). doi:10.1186/s13059-021-02260-y Medline

83. L. Yan, C. Ma, D. Wang, Q. Hu, M. Qin, J. M. Conroy, L. E. Sucheston, C. B. Ambrosone, C. S. Johnson, J. Wang, S. Liu, OSAT: A tool for sample-to-batch allocations in genomics experiments. *BMC Genomics* **13**, 689 (2012). doi:10.1186/1471-2164-13-689 Medline

84. A. Seluanov, C. Hine, J. Azpurua, M. Feigenson, M. Bozzella, Z. Mao, K. C. Catania, V. Gorbunova, Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19352–19357 (2009). doi:10.1073/pnas.0905252106 Medline

85. T. R. Dawber, G. F. Meadors, F. E. Moore Jr., Epidemiological approaches to heart disease: The Framingham Study. *Am. J. Public Health Nations Health* **41**, 279–281 (1951). doi:10.2105/AJPH.41.3.279 Medline

86. W. B. Kannel, M. Feinleib, P. M. McNamara, R. J. Garrison, W. P. Castelli, An investigation of coronary heart disease in families. The Framingham offspring study. *Am. J. Epidemiol.* **110**, 281–290 (1979). doi:10.1093/oxfordjournals.aje.a112813 Medline

87. M. J. Aryee, A. E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A. P. Feinberg, K. D. Hansen, R. A. Irizarry, Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014). doi:10.1093/bioinformatics/btu049 Medline

88. The Women's Health Initiative Study Group, Design of the Women's Health Initiative clinical trial and observational study. *Control. Clin. Trials* **19**, 61–109 (1998). doi:10.1016/S0197-2456(97)00078-0 Medline

89. G. L. Anderson, J. Manson, R. Wallace, B. Lund, D. Hall, S. Davis, S. Shumaker, C.-Y. Wang, E. Stein, R. L. Prentice, Implementation of the women's health initiative study design. *Ann. Epidemiol.* **13**, S5–S17 (2003). doi:10.1016/S1047-2797(03)00043-7 Medline

90. M. Nasu, S. Yada, A. Igarashi, D. Sutoo, K. Akiyama, M. Ito, N. Yoshida, S. Ueda, Mammalian-specific sequences in pou3f2 contribute to maternal behavior. *Genome Biol. Evol.* **6**, 1145–1156 (2014). doi:10.1093/gbe/evu072 Medline

91. P. Dammann, R. Šumbera, C. Massmann, A. Scherag, H. Burda, Extended longevity of reproductives appears to be common in Fukomys mole-rats (Rodentia, Bathyergidae). *PLOS ONE* **6**, e18757 (2011). doi:10.1371/journal.pone.0018757 Medline

92. S. R. Urfer, M. Kaeberlein, D. E. L. Promislow, K. E. Creevy, Lifespan of companion dogs seen in three independent primary care veterinary clinics in the United States. *Canine Med. Genet.* **7**, 7 (2020). doi:10.1186/s40575-020-00086-8 Medline

93. L. D. Moore, T. Le, G. Fan, DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38 (2013). doi:10.1038/npp.2012.112 Medline

94. J. Z. Bakdash, L. R. Marusich, Repeated measures correlation. *Front. Psychol.* **8**, 456 (2017). doi:10.3389/fpsyg.2017.00456 Medline

95. S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014). doi:10.1016/j.cell.2014.11.021 Medline