DNA methylation networks underlying mammalian traits

Title: DNA Methylation Networks Underlying Mammalian Traits

Authors: Amin Haghani[1, 2 †*]; Caesar Z. Li[3, 4 †]; Todd R. Robeck[5]; Joshua Zhang[1]; Ake T. Lu[1, 2]; Julia Ablaeva[6]; Victoria A. Acosta-Rodríguez[7]; Danielle M. Adams[8]; Abdulaziz N. Alagaili[9, 10]; Javier Almunia[11]; Ajoy Aloysius[12]; Nabil M.S. Amor[13]; Reza Ardehali[14]; Adriana Arneson[15, 16]; C. Scott Baker[17]; Gareth Banks[18]; Katherine Belov[19]; Nigel C. Bennett[20]; Peter Black[21]; Daniel T. Blumstein[22, 23]; Eleanor K. Bors[17]; Charles E. Breeze[24]; Robert T. Brooke[25]; Janine L. Brown[26]; Gerald Carter[27]; Alex Caulton[28, 29]; Julie M. Cavin[30]; Lisa Chakrabarti[31]; Ioulia Chatzistamou[32]; Andreas S. Chavez[27, 33]; Hao Chen[34]; Kaiyang Cheng[35]; Priscila Chiavellini[36]; Oi-Wa Choi[37, 38]; Shannon Clarke[28]; Joseph A. Cook[39]; Lisa N. Cooper[40]; Marie-Laurence Cossette[41]; Joanna Day[42]; Joseph DeYoung[37, 38]; Stacy Dirocco[43]; Christopher Dold[44]; Jonathan L. Dunnum[39]; Erin E. Ehmke[45]; Candice K. Emmons[46]; Stephan Emmrich[6]; Ebru Erbay[47, 48, 49]; Claire Erlacher-Reid[43]; Chris G. Faulkes[50, 51]; Zhe Fei[3, 52]; Steven H. Ferguson[53, 54]; Carrie J. Finno[55]; Jennifer E. Flower[56]; Jean-Michel Gaillard[57]; Eva Garde[58]; Livia Gerber[59, 60]; Vadim N. Gladyshev[61]; Rodolfo G. Goya[36]; Matthew J Grant[62]; Carla B. Green[7]; M. Bradley Hanson[46]; Daniel W. Hart[20]; Martin Haulena[63]; Kelsey Herrick[64]; Andrew N. Hogan[65]; Carolyn J. Hogg[19]; Timothy A. Hore[66]; Taosheng Huang[67]; Juan Carlos Izpisua Belmonte[2]; Anna J. Jasinska[37, 68, 69]; Gareth Jones[70]; Eve Jourdain[71]; Olga Kashpur[72]; Harold Katcher[73]; Etsuko Katsumata[74]; Vimala Kaza[75]; Hippokratis Kiaris[76]; Michael S. Kobor[77]; Pawel Kordowitzki[78]; William R. Koski[79]; Michael Krützen[60]; Soo Bin Kwon[16, 15]; Brenda Larison[22, 80]; Sang-Goo Lee[61]; Marianne Lehmann[36]; Jean-François Lemaître[57]; Andrew J. Levine[81]; Xinmin Li[82]; Cun Li[83, 84]; Andrea R. Lim[1]; David T.S. Lin[85]; Dana M. Lindemann[43]; Schuyler W. Liphardt[86]; Thomas J. Little[87]; Nicholas Macoretta[6]; Dewey Maddox[88]; Craig O. Matkin[89]; Julie A. Mattison[90]; Matthew McClure[91]; June Mergl[92]; Jennifer J. Meudt[93]; Gisele A. Montano[5]; Khyobeni Mozhui[94]; Jason Munshi-South[95]; William J. Murphy[96, 97]; Asieh Naderi[76]; Martina Nagy[98]; Pritika Narayan[62]; Peter W. Nathanielsz[83, 84]; Ngoc B. Nguyen[14]; Christof Niehrs[99, 100]; Batsaikhan Nyamsuren[101]; Justine K. O'Brien[42]; Perrie O'Tierney Ginn[72]; Duncan T Odom[102, 103]; Alexander G. Ophir[104]; Steve Osborn[105]; Elaine A. Ostrander[65]; Kim M. Parsons[46]; Kimberly C. Paul[81]; Amy B. Pedersen[87]; Matteo Pellegrini[106]; Katharina J. Peters[60, 107]; Jessica L. Petersen[108]; Darren W. Pietersen[109]; Gabriela M. Pinho[22]; Jocelyn Plassais[65]; Jesse R. Poganik[61]; Natalia A. Prado[110, 26]; Pradeep Reddy[111, 2]; Benjamin Rey[57]; Beate R. Ritz[112, 113, 81]; Jooke Robbins[114]; Magdalena Rodriguez[115]; Jennifer Russell[105]; Elena Rydkina[6]; Lindsay L. Sailer[104]; Adam B. Salmon[116]; Akshay Sanghavi[73]; Kyle M. Schachtschneider[117, 118, 119]; Dennis Schmitt[120]; Todd Schmitt[64]; Lars Schomacher[99]; Lawrence B. Schook[117, 121]; Karen E. Sears[22]; Ashley W. Seifert[12]; Aaron B.A. Shafer[122]; Anastasia V. Shindyapina[61]; Melanie Simmons[45]; Kavita Singh[123]; Ishani Sinha[22]; Jesse Slone[67]; Russel G. Snell[62]; Elham Soltanmohammadi[76]; Matthew L. Spangler[108]; Maria Spriggs[21]; Lydia Staggs[43]; Nancy Stedman[21]; Karen J. Steinman[124]; Donald T Stewart[125]; Victoria J. Sugrue[66]; Balazs Szladovits[126]; Joseph S. Takahashi[7, 127]; Masaki Takasugi[6]; Emma C. Teeling[128]; Michael J. Thompson[106]; Bill Van Bonn[129]; Sonja C. Vernes[130, 131]; Diego Villar[132]; Harry V. Vinters[133]; Ha Vu[15, 16]; Mary C. Wallingford[72]; Nan Wang[37, 38]; Gerald S. Wilkinson[8]; Robert W. Williams[134]; Qi Yan[3, 2]; Mingjia Yao[3]; Brent G. Young[54]; Bohan Zhang[61]; Zhihui Zhang[6]; Yang Zhao[6]; Peng Zhao[14, 135]; Wanding Zhou[136, 137]; Joseph A. Zoller[3]; Jason Ernst[15, 16]; Andrei Seluanov[138]; Vera Gorbunova[138]; X. William Yang[37, 38]; Ken Raj[139]; Steve Horvath[1, 2 *]

Affiliations:

1 Dept. of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
2 Altos labs, San Diego, CA, USA
3 Dept. of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, USA
4 Janssen Research & Development, Spring House, PA, USA
5 Zoological Operations, SeaWorld Parks and Entertainment, Orlando, Florida, USA
6 Dept. of Biology, University of Rochester, Rochester, NY, USA
7 Dept. of Neuroscience, Peter O'Donnell Jr. Brain Institute, University of Texas Southwestern Medical Center, Dallas, TX, US
8 Dept. of Biology, University of Maryland, College Park, USA
9 Dept. of Zoology, College of Science, King Saud University, P.O. Box 2455, Riyadh, Saudi Arabia
10 DSFP, King Saud University, Riyadh, Saudi Arabia
11 Loro Parque Fundacion, Avenida Loro Parque, Puerto de la Cruz, Tenerife, Spain
12 Dept. of Biology, University of Kentucky, Lexington, KY, USA
13 Laboratory of Biodiversity, Parasitology, and Ecology, University of Tunis El Manar, Tunisia
14 Division of Cardiology, Dept. of Internal Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA
15 Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA, USA
16 Dept. of Biological Chemistry, University of California, Los Angeles, Los Angeles, California, USA
17 Marine Mammal Institute, Oregon State University, Newport, OR, USA
18 Mammalian Genetics Unit, MRC Harwell Institute, Harwell Science and Innovation Campus, Oxfordshire, UK
19 School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia
20 Dept. of Zoology and Entomology, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa
21 Busch Gardens Tampa, Tampa, Florida, USA
22 Dept. of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA, USA
23 The Rocky Mountain Biological Laboratory, Crested Butte, Colorado, USA
24 Altius Institute for Biomedical Sciences, Seattle, WA, USA
25 Epigenetic Clock Development Foundation, Los Angeles, CA, USA
26 Center for Species Survival, Smithsonian Conservation Biology Institute, Front Royal, VA, USA
27 Dept. of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH, USA
28 AgResearch, Invermay Agricultural Centre, Mosgiel, Otago, New Zealand
29 Dept. of Biochemistry, University of Otago, Dunedin, Otago, New Zealand
30 Gulf World Marine Park - Dolphin Company, Panama City Beach, FL, USA
31 School of Veterinary Medicine and Science, University of Nottingham, UK.
32 Dept. of Pathology, Microbiology & Immunology, School of Medicine, University of South Carolina, SC, USA
33 Translational Data Analytics Institute, The Ohio State University, Columbus, OH, USA
34 Dept. of Pharmacology, Addiction Science and Toxicology, The University of Tennessee Health Science Center, Memphis, TN, USA
35 Medical Informatics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
36 Biochemistry Research Institute of La Plata, Histology and Pathology, School of Medicine, University of La Plata, La Plata, Argentina
37 Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, Dept. of Psychiatry and Biobehavioral Sciences, University of California Los Angeles, Los Angeles, California, USA

[38] Dept. of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA
[39] University of New Mexico, Dept. of Biology and Museum of Southwestern Biology, Albuquerque, New Mexico, United States of America
[40] Dept. of Anatomy and Neurobiology, Northeast Ohio Medical University, Rootstown, Ohio, USA
[41] Dept. of Environmental & Life Sciences, Trent University, Peterborough Ontario, Canada K9J 7B8
[42] Taronga Institute of Science and Learning, Taronga Conservation Society Australia, New South Wales, Australia
[43] SeaWorld Orlando, 7007 SeaWorld Drive, Orlando, FL USA
[44] Zoological Operations, SeaWorld Parks and Entertainment, 7007 SeaWorld Drive, Orlando, Florida, USA
[45] Duke Lemur Center, Durham, North Carolina, USA
[46] Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, Washington, USA
[47] Altos Labs, Bay Area Institute of Science, Redwood City, USA
[48] Dept. of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[49] David Geffen School of Medicine, University of California, Los Angels, CA, USA
[50] School of Biological and Chemical Sciences, Queen Mary University of London, London, UK
[51] School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK
[52] Dept. of Statistics, University of California Riverside, CA, USA
[53] Dept. of Biological Sciences, University of Manitoba, Winnipeg, Canada
[54] Fisheries and Oceans Canada, University Crescent, Winnipeg, Canada
[55] Dept. of Population Health and Reproduction, University of California, Davis School of Veterinary Medicine, Davis, CA, USA
[56] Mystic Aquarium, Mystic, Connecticut, USA
[57] University of Lyon, CNRS, Laboratoire de Biométrie et Biologie Evolutive, UMR5558, Villeurbanne, France
[58] Greenland Institute of Natural Resources, Nuuk, Greenland
[59] School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia
[60] Evolutionary Genetics Group, Dept. of Evolutionary Anthropology, University of Zurich, Zurich, Switzerland
[61] Division of Genetics, Dept. of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[62] Applied Translational Genetics Group, School of Biological Sciences, Centre for Brain Research, The University of Auckland, Auckland, New Zealand
[63] Vancouver Aquarium, Vancouver, Canada
[64] SeaWorld San Diego, San Diego, CA, USA
[65] Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
[66] Dept. of Anatomy, University of Otago, Dunedin, New Zealand
[67] Division of Human Genetics, Dept. of Pediatrics, University at Buffalo, New York, USA
[68] Division of Infectious Diseases, Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States
[69] Department of Molecular Genetics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland
[70] School of Biological Sciences, University of Bristol, Bristol, UK
[71] Norwegian Orca Survey, Andenes, Norway
[72] Mother Infant Research Institute, Tufts Medical Center, Boston, MA, USA
[73] Yuvan Research, Mountain View, CA, USA
[74] Kamogawa Sea World, Kamogawa, Chiba, Japan
[75] Peromyscus Genetic Stock Center, University of South Carolina, SC, USA
[76] Dept. of Drug Discovery and Biomedical Sciences, College of Pharmacy, University of South Carolina, SC, USA
[77] Edwin S.H. Leong Healthy Aging Program, Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, Canada
[78] Institute of Animal Reproduction and Food Research of Polish Academy of Sciences, Olsztyn, Poland
[79] LGL Limited, King City, ON, Canada

[80] Center for Tropical Research, Institute for the Environment and Sustainability, UCLA, Los Angeles, CA, USA
[81] Dept. of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
[82] Technology Center for Genomics and Bioinformatics, Dept. of Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, CA, USA
[83] Texas Pregnancy and Life-course Health Center, Southwest National Primate Research Center, San Antonio, Texas, USA
[84] Dept. of Animal Science, College of Agriculture and Natural Resources, Laramie, Wyoming, USA
[85] Centre for Molecular Medicine and Therapeutics, BC Children's Hospital Research Institute, University of British Columbia, Vancouver, Canada
[86] Biology Department, University of New Mexico, Albuquerque, NM, USA
[87] Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK
[88] White Oak Conservation Center, Yulee FL, USA
[89] North Gulf Oceanic Society, Homer, Alaska,USA
[90] Translational Gerontology Branch, National Institute on Aging Intramural Research Program, National Institutes of Health, Baltimore, MD, USA
[91] ABS Global, DeForest, WI, USA
[92] Marineland of Canada, Niagara Falls, Ontario, Canada
[93] Biomedical and Genomic Research Group, Dept. of Animal and Dairy Sciences, University of Wisconsin Madison, Madison, Wisconsin, US
[94] Dept. of Preventive Medicine, University of Tennessee Health Science Center, College of Medicine, Memphis, TN, USA
[95] Louis Calder Center - Biological Field Station, Dept. of Biological Sciences, Fordham University, Armonk, NY USA.
[96] Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas 77843, USA
[97] Interdisciplinary Program in Genetics and Genomics, Texas A&M University, College Station, Texas 77843, USA
[98] Museum fur Naturkunde, Leibniz-Institute for Evolution and Biodiversity Science, Berlin, Germany
[99] Institute of Molecular Biology (IMB), Mainz, Germany
[100] Division of Molecular Embryology, DKFZ-ZMBH Alliance, Heidelberg, Germany
[101] Mammalogical Society of Mongolia, Ulaanbaatar, Mongolia
[102] Cancer Research UK Cambridge Institute, University of Cambridge, Robinson Way, Cambridge, CB2 0RE, UK
[103] Deutsches Krebsforschungszentrum, Division of Regulatory Genomics and Cancer Evolution - B270, Im Neuenheimer Feld 280, Heidelberg, Germany
[104] Dept. of Psychology, Cornell University, Ithaca, NY, USA
[105] SeaWorld San Antonio, San Antonio, Texas, USA
[106] Dept. Molecular Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA, USA
[107] School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, Australia
[108] Dept. of Animal Science, University of Nebraska, Lincoln, USA
[109] Mammal Research Institute, Dept. of Zoology and Entomology, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa
[110] Dept. of Biology, College of Arts and Science, Adelphi University. One South Avenue, Garden City, NY, USA
[111] Salk Institute for Biological Studies, La Jolla, CA USA
[112] Dept. of Epidemiology, UCLA Fielding School of Public Health, Los Angeles, CA, USA
[113] Dept. of Environmental Health Sciences, UCLA Fielding School of Public Health, Los Angeles, CA, USA
[114] Center for Coastal Studies, Provincetown, MA, USA
[115] Miami Seaquarium, Miami, Florida, USA
[116] The Sam and Ann Barshop Institute for Longevity and Aging Studies, and Dept. of Molecular Medicine, UT Health San Antonio, and the Geriatric Research Education and Clinical Center, South Texas Veterans Healthcare System, San Antonio TX, USA
[117] Dept. of Radiology, University of Illinois at Chicago, Chicago, IL, USA
[118] Dept. of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, IL, USA

[119] National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[120] College of Agriculture, Missouri State University, Springfield MO, USA
[121] Dept. of Animal Sciences, University of Illinois at Urbana-Champaign, USA
[122] Dept. of Forensic Science, Environmental & Life Sciences, Trent University, Peterborough Ontario, Canada K9J 7B8
[123] Shobhaben Pratapbhai Patel School of Pharmacy & Technology Management, SVKM'S NMIMS University, Mumbai, India
[124] Species Preservation Laboratory, SeaWorld San Diego, California, USA
[125] Biology Department, Acadia University, Wolfville, Nova Scotia, Canada B4P 2R6
[126] Dept. of Pathobiology and Population Sciences, Royal Veterinary College, Hatfield, UK
[127] Howard Hughes Medical Institute, Dept. of Neuroscience, University of Texas Southwestern Medical Center, Dallas, Texas, USA
[128] School of Biology and Environmental Science, University College Dublin, Belfield, Dublin, Ireland
[129] Animal Care and Science Division, John G. Shedd Aquarium, Chicago, Illinois, USA
[130] School of Biology, The University of St Andrews, Fife, UK
[131] Neurogenetics of Vocal Communication Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[132] Blizard Institute, Faculty of Medicine and Dentistry, Queen Mary University of London, London, UK
[133] Dept. of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, USA
[134] Dept. of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, College of Medicine, Memphis, TN, USA
[135] Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, USA
[136] Center for Computational and Genomic Medicine, Children's Hospital of Philadelphia, Philadelphia, USA
[137] Dept. of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, USA
[138] Depts. of Biology and Medicne, University of Rochester, Rochester, NY, USA
[139] Altos Labs, Cambridge, UK

† Co-first authors

* Corresponding authors: shorvath@mednet.ucla.edu ahaghani@altoslabs.com

Address: 5510 Morehouse Dr San Diego, CA 92121 United States

Abstract: Using DNA methylation profiles (n=15,456) from 348 mammalian species, we report the construction of *phyloepigenetic* trees that bear remarkable similarities to traditional phylogenetic ones. Using unsupervised clustering across all samples, we identified 55 distinct cytosine modules, of which 30 are related to traits such as maximum lifespan, adult weight, age, sex, and human mortality risk. Maximum lifespan is found to be associated with methylation levels in *HOX* genes, with developmental processes, and potential regulation by pluripotency transcription factors. The methylation state of some modules responds to perturbations such as caloric restriction, ablation of growth hormone receptors, consumption of high-fat diets, and expression of Yamanaka factors. This study reveals an intertwined evolution of the genome and epigenome that mediates the biological characteristics and traits of different mammalian species.

One-Sentence Summary: Phyloepigenetic trees, derived from DNA methylation profiles, mirror mammalian evolution and are related to mammalian lifespan and other species characteristics.

Main Text:

Comparative epigenomics is a burgeoning field that integrates epigenetic signatures with phylogenetic relationships to decipher gene-to-trait functions (*1-3*). Prior research has investigated the capacity of DNA methylation patterns in regulatory sequences to reflect evolutionary relationships among species (*3, 4*). A recent study compared methylation data across multiple animal species at orthologous gene promoters using a sequencing-based assay that did not specifically target conserved CpGs (*4*). Previous investigations faced limitations regarding the measurement platform, particularly the low sequencing depth at conserved CpGs and the sample size per species.

Our study overcomes these constraints in several ways. First, we utilized a measurement platform ensuring high sequencing depth at conserved CpGs, allowing for a more precise analysis of DNA methylation patterns in highly conserved DNA regions. Second, we significantly increased the sample size per species, aiming for around 30 samples. We profiled 348 species from 25 of the 26 mammalian taxonomic orders. This comprehensive dataset enables a rigorous examination of phylogenetic relationships, co-methylation relationships between cytosines, and their associations with maximum lifespan and other species characteristics.

We profiled 15,456 samples (**Fig. 1A; table S1**) using a methylation array platform that provides extremely effective sequencing depth at highly conserved CpGs across mammalian species (*5*). This vast dataset is the product of the multi-national Mammalian Methylation Consortium, which consists of 191 collaborators from diverse areas of scientific expertise. In previous studies, we applied supervised machine learning methods to generate DNA methylation-based predictors of age, called epigenetic clocks, for numerous species (*6-31*).

Here, we perform a large-scale cross-species unsupervised analysis of the entire dataset to reveal the relationship of DNA methylation (DNAm) with mammalian phylogeny. We show that we can construct *phyloepigenetic* trees that parallel traditional phylogenetic ones. We then proceed to interrogate the extent to which DNA methylation underpins specific biological traits by employing unsupervised weighted correlation network analysis (WGCNA) to minimize the influence of bias on our observations. This approach readily identifies clusters of co-methylated CpGs (co-methylation modules) that are associated with species characteristics, including taxonomy, tissue type, sex, lifespan, and aging.

Results:

Evolution and DNA methylation

We generated a dataset consisting of DNA methylation profiles of 15,456 DNA samples derived from 70 tissue types, from 348 mammalian species using the mammalian methylation array (*5*). We evaluated whether methylation levels of cytosines (CpGs) in DNA sequences that are conserved across species would allow us to construct what could be termed a phyloepigenetic tree. To avoid potential confounding by different tissue types, we generated tissue-specific phyloepigenetic trees (**Fig. 1B; fig. S2; fig. S3**). We defined the '*Congruence'* between traditional phylogenetic trees and phyloepigenetic trees as the Pearson correlation coefficient between distances (branch length) based on phyloepigenetic trees and evolutionary distances in traditional phylogenetic trees. We observe high congruence (*Congruence*=0.93, **Fig. 1C; fig. S2**) for the blood-based phyloepigenetic tree (124 species), and lower congruence values for non-blood tissues (*Congruence*=0.58 for liver and *Congruence*=0.72 for skin, **fig. S2**). The lower congruence in liver

(158 species) and skin (133 species) may be due to potential variability in sampling between species. The varying congruence across tissue types shows that the CpG probes do not serve as genotyping proxies. The tissue dependence of congruence indicates that phyloepigenetic trees are derived based on differences in methylation levels and not sequence conservation. This point is also corroborated by three sensitivity analyses, which confirmed that the high congruence was indeed due to differences in methylation levels (supplementary text). In particular, the phyloepigenetic trees based on the 180 CpGs with the most significant detection p values across all 348 species still are congruent with traditional trees (**fig. S2F-G**).
In order to identify CpGs that exhibit a pronounced phylogenetic signal in relation to methylation and phylogenetic trees, we utilized the K statistic method described by Blomberg et al. (2003) (*32*). Among the top 500 CpGs showing significant phylogenetic signals (nominal Blomberg $p < 0.001$, additionally selected by variance z-score), we observed an enrichment in upstream intergenic regions (odds ratio OR = 1.4, Fisher exact $p < 0.05$, **fig. S4B**). To further investigate regions with the strongest phylogenetic signal, we divided the data into groups of 10 CpGs relative to the transcriptional start site (TSS). This analysis also confirmed that intergenic regions exhibit robust phylogenetic signals (OR > 3, Fisher exact $p < 0.05$), while the promoter regions did not show such signals (**Fig. 1D**).

DNA methylation networks relate to individual and species traits

We used signed weighted correlation network analysis (WGCNA, an unsupervised analysis) (*33*) to cluster CpGs with similar methylation dynamics across samples into co-methylation modules. We then summarized their methylation profiles as "module eigengenes". The respective eigengenes of these modules were used to identify their potential correlations with various traits within and across mammalian species.

Our data analysis proceeded in two sequential phases. First, we developed several co-methylation networks using data from 11,099 DNA samples from 174 species (discovery dataset, finalized March 2021). A eutherian network (Net1) was formed from 14,705 conserved CpGs using this dataset (**Fig. 2A**). Later (March 2022), we generated a second data set of 4,357 samples from 30 tissues of 240 mammalian species (174 new species, and 66 that are represented in the discovery set), which were not used to define modules and were used as an independent validation set. All of the eutherian modules were present in the independent validation dataset according to module preservation statistics (*corKME*) (*34*); validating the presence of these modules (*corKME*>0.43, $p<10^{-22}$; median corKME=0.84) (**fig. S5**). These modules were designated with colors according to the WGCNA convention (**Fig. 2A**). The smallest module (lavenderblush3) consisted of 33 CpGs, while the largest (turquoise) had 1,864 CpGs.

To characterize the 55 modules with respect to species characteristics (e.g., maximum lifespan and average adult weight), module eigengenes were calculated in all samples (discovery and replication set combined, 331 eutherian species). As information on taxonomic order, tissues, maximum lifespan, age, sex and adult weight of each species were available, we were able to assess whether any of the module eigengenes correlated with these traits. Of the 55 modules, 30 were found to be correlated with at least one trait (**Fig. 2B; fig. S7; table S3**). Specifically, 15 modules were related to taxonomic orders such as primates, rodents, or carnivores (**Fig. 2B;** see also **fig. S11**). Ten modules related to tissue type (**fig. S11**), while two were related to sex (**fig. S11**), one to age, seven modules to maximum lifespan, and four to average adult species weight. Some modules were related to multiple characteristics. In the following sections, we mainly focus on the modules that relate to mammalian maximum lifespan, adult weight, and age. Other modules

(related to taxonomic order, tissue type, or sex) are described in the supplement (**fig. S11**). We performed two analyses to ascertain whether these eutherian modules are also applicable to marsupials and monotremes. First, we trained a network (Net2) in both eutherians and marsupials based on only 7,956 probes that are mappable to both. The color bands underneath the hierarchical tree reveal that all the Net1 modules were also preserved in Net2 (**Fig. 2A**). Second, we selected CpGs in Net1 modules that are also mapped to marsupials or monotremes and confirmed that their eigengene relationships to primary traits were retained in these mammalian clades (**table S3**). For example, the magenta module, which is related to blood in eutherians, was also found to be so in monotremes (**table S3**), which confirms that the Net1 modules can indeed be applied to other mammalian clades, by selecting probes that are also mapped to those clades.

Relationship with protein-protein interactions

A functional enrichment study, which adjusted for the methylation array background, showed that the 500 most connected CpGs per module were adjacent to genes associated with a wide range of biological processes including development, immune function, metabolism, reproduction, stem cell biology, stress responses, aging, and several signaling pathways (**Fig. 2C, fig. S9**).

We examined whether the proteins encoded by cognate genes (closest to respective CpGs) within modules are known to mutually interact or predicted to do so by STRING protein-protein interaction networks, which integrate known and predicted protein associations from over 14,000 organisms (*35*). A permutation test analysis evaluating the global cluster coefficient (*36*) of each module showed that 14 modules are significantly ($p<0.001$) enriched for genes encoding mutually-interacting proteins (**Fig 2D**). Overall, these results suggest that co-methylation relationships can be reflected at the protein level for a subset of modules.

Modules related to maximum lifespan

To adjust for potential confounders, we used four regression modeling approaches to identify modules that are associated with log transformed maximum lifespan (dependent variable): 1) a univariate regression model whose covariate was the module eigengene (averaged per species), 2) a phylogenetic regression model whose covariate was again the module eigengene (averaged per species), 3) a multivariate linear regression model that included the module eigengene, sex, tissue, and relative age as covariates, 4) model approach 1 applied to specific tissue types.

The marginal analysis identified four modules (magenta, black, midnightblue, and tan) that related significantly to maximum lifespan (the absolute value of the Pearson correlation exceeded $r=0.6$, Student T test $p<1\times10^{-33}$). The CpGs underlying the implicated modules exhibit the sample patterns as can be seen from corresponding heatmaps (**fig. S14C**). Phylogenetic regression also identified associations of the same modules (**table S3**). Our fourth modeling approach, i.e. the tissue-stratified marginal analysis, indicates that the relationship of modules to maximum lifespan is often tissue-specific. For example, the magenta and midnightblue modules relate to maximum lifespan in lung and liver (**fig. S14A**). In contrast, the black module relates to maximum lifespan only in skin, while the tan module exhibited a weak relationship to lifespan in the tissue-specific analysis.

For ease of comprehension, modules were labeled with the trait and direction of relationship by superscript +/- signs (e.g. magenta = $Lifespan^{(+)}Weight^{(+)}Blood^{(+)}$ module). The two modules (magenta with 480 CpGs, and midnightblue with 249 CpGs) that correlated with lifespan in lung and liver also correlated significantly with average adult weight across all eutherian species

(r=0.47 to 0.55, $p<1x10^{-18}$, **Fig. 3**). The magenta module (Lifespan$^{(+)}$Weight$^{(+)}$Blood$^{(+)}$) is enriched with developmental genes such as *HOXA5* and *VEGFA*, *SOX2*, and *WNT11* **(table S4)**. The midnightblue (Lifespan$^{(+)}$Weight$^{(+)}$) module implicates genes involved in tRNA metabolism ($p=2x10^{-6}$, e.g. *URM1*), lipopolysaccharides ($p=5x10^{-6}$, e.g. *CERCAM*), development ($p=10^{-4}$, HOXL gene family), and fatty acids ($p=2x10^{-3}$, e.g. *ACADVL*). The magenta module also relates to lifespan and average weight of dog breeds(**Fig. 3C**, r= -0.30, p=0.003). Furthermore, it is related to the hazard of human death (hazard ratio HR= 0.91, MetaP=0.0016, **Fig. 3D**) in epidemiological cohort studies.

After adjustment for phylogeny, the cyan module relates to mammalian lifespan phylogenetic contrast (r=0.42, $p=4x10^{-14}$, **fig. S13I**). The Lifespan$^{(+)}$Liver$^{(-)}$ (cyan) module consists of genes that play a role in adaptive immunity ($p=2x10^{-6}$), histone and protein demethylation (p=0.0001), and metabolism (p=0.0004) (**table S4**).

The multivariate model analysis included sex, tissue type, and relative age as covariates to reveal the modules that relate to lifespan in different tissues. The regression analysis found two modules with opposing correlations with maximum lifespan: green module (lifespan r=0.42, average weight r=0.38, $p<10^{-300}$) and the greenyellow module (lifespan r= -0.44, average weight r= -0.35, $p<10^{-300}$, **fig. S13J**). The CpGs of the Lifespan$^{(-)}$Weight$^{(-)}$Rodentia$^{(+)}$ (greenyellow) are located near genes that play a role in development ($p=5x10^{-13}$**, table S4**) and in RNA metabolism ($p=6x10^{-12}$).

Age-related consensus module in mammals

The purple module (denoted subsequently as RelativeAge$^{(+)}$ module) exhibited the strongest positive correlation with relative age (Relative age r=0.35, $p<10^{-300}$, **Fig. 3E; fig. S13**).

To remove the confounding effects of species and/or tissue type, we also constructed seven consensus networks (denoted cNet3,...,cNet9, Description in supplement and methods). The purple module was preserved in 3 different consensus networks (cNet3, cNet4, and cNet6, **Fig. 2A**), suggesting conservation in different species and tissues (scatter plot in **fig. S11H**). The RelativeAge$^{(+)}$ module is positively enriched for CpGs in regulatory regions (e.g. promoters and 5'UTR) and depleted in intron regions (**fig. S15**). Functional enrichment of this module highlighted embryonic stem cell regulation, axonal fasciculation, angiogenesis, and diabetes-related pathways (**table S3**). The CpGs in this module are adjacent to Polycomb repressor complex 2 (PRC2, EED) targets which are marked by H3K27me3 (**table S3**).

Ingenuity pathway analysis implicates POU5F1 (alias OCT4), SHH, ASCL1, SOX2, and NEUROG2 proteins as putative upstream regulators of the RelativeAge$^{(+)}$ module. We used GTEx data to examine if the mRNA levels of any of these upstream regulators are altered with age in human tissues. *OCT4* (repeated measures correlation, rmCor=0.07, $p=2x10^{-14}$), which is among the four known Yamanaka factors for cellular dedifferentiation, showed a positive increase with age in several but not all human tissues (**fig. S11F**). Nine other genes (e.g. *HOXD10*, rmCor=0.16, $p=4x10^{-50}$; *SRXN1*, rmCor=-0.14, $p=4x10^{-52}$) from the RelativeAge$^{(+)}$ module also had a nominally significant rmCorr (p<0.005) in GTEx data (**Fig. 3F; fig. S11G**), although opposite aging patterns could be found in select tissues. These observations highlight the relevance of genes in the RelativeAge$^{(+)}$ module to stem cell biology and aging in human tissues.

Interventional studies in mice

We related our methylation modules to interventions that are known to modulate the lifespan of mice (**Fig. 4A–C**). This included growth hormone receptor knockout (i.e. dwarf mice) (*37*) and

caloric restriction (*38*), which extended life, and high-fat diet, which elicited the opposite effect (*12*). Six modules, including the purple module (RelativeAge$^{(+)}$) showed a significant ($p<0.05$) decrease of the module eigengene in dwarf mice and after caloric restriction, and conversely a modest increase after a high-fat diet. Although magenta, black, midnightblue, tan, and greenyellow modules are related to maximum lifespan, these modules were not significantly ($p>0.05$) associated with interventions that affect murine lifespan (GHRKO, CR and HF diet). Instead the purple, ivory, lavenderblush3, royalblue, salmon4 and skyblue modules, none of which are related to maximum lifespan, were the ones that are significantly associated with these interventions. In other words, the lifespan modules and lifespan-affecting interventions modules are mutually exclusive.

Transient expression of Yamanaka factors

We examined if a transient expression of the Yamanaka factors in the 4-factor (4F) mouse affects the module eigengenes. The experimental design is shown in **Fig. 4D**, with additional details reported in the original article (*39*). Five out of six of the above mentioned murine intervention modules showed a nominally significant dose-dependent rejuvenation in murine skin ($p<0.06$) and 2 modules showed the same in kidney (dose refers to the duration of 4F treatment: 0-, 1-, 7-, and 10-months intermittent expression of 4F factors) (**Fig. 4E–F**). The purple, ivory, and lavenderblush3 modules were particularly sensitive to the 4F treatment (Pearson correlation $\leqslant$-0.72 in skin). In addition, the purple RelativeAge$^{(+)}$ module's response to the 4F treatment is consistent with bioinformatic findings that OCT4 is an upstream regulator of this module.

Epigenome-wide association analysis of maximum lifespan

We carried out epigenome-wide association studies (EWAS) to identify individual CpGs with methylation levels that correlate with maximum lifespan. To reduce bias resulting from different levels of sequence conservation, our EWAS of maximum lifespan focused on n = 333 eutherian species, excluding marsupial and monotreme species. We restricted the analysis to 28,318 high quality probes that are conserved between humans and mice.

When relating individual CpGs to log-transformed maximum lifespan, we used several modeling approaches (detailed in the Supplementary text). Briefly, our first approach, generic modeling, applied regression analysis ignoring tissue type and age. Second, we repeated the regression analysis after focusing on a given tissue type. Third, we focused on specific non-overlapping age groups: young animals (defined as age younger than 1.5 times the age at sexual maturity), middle-aged, and old (defined as Age>3.5 times the age at sexual maturity), see fig S19. Some of these regression models were further adjusted for average species weight (denoted lifespanAdjWeight).

For brevity, we will focus on linear regression models since phylogenetic regression models led to qualitatively similar conclusions (**tables S13–S14**). The most significant lifespan-related CpGs are located in the distal intergenic region neighboring *TLE4* (Pearson $r = 0.68$, $p = 5.8\times10^{-46}$, **Fig. 5A, table S11**) and two CpGs near the promoter region of *HOXA4* ($R = 0.66$, $p = 7.5\times10^{-45}$, midnight blue module, **Fig. 5A**), and negatively-correlated with a CpG in an intron of *GATA3* ($R = -0.65$, $p = 8.8\times10^{-42}$, **Fig. 5A**). Many of these significant CpGs remain so after phylogenetic adjustment, such as the CpGs neighboring *TLE4*, *HOXA4* ($p = 4.2\times10^{-5}$, $p = 4.8\times10^{-3}$ respectively, **fig. S17** and **table S11-S12**). The top 1,000 lifespan-related CpGs (comprising 500 positively and 500 negatively lifespan related CpGs) significantly overlapped (Fisher exact $p = 5.5\times10^{-134}$) with those found in our weight-adjusted analysis (lifespanAdjWeight).

In general, methylation of lifespan-related CpGs does not change with age in mammalian tissues (**Fig. 5B, fig. S20**). The same can be seen from EWAS of lifespan restricted to animals of a given age group (e.g., only very young animals, **fig. S20D**). The EWAS of lifespan in all animals (irrespective of age) is highly correlated (r>0.7) with the analogous EWAS restricted to animals that are young, middle-aged, or old, animals respectively.

EWAS of lifespan showed good consistency with the eigengene-based analysis in the mammalian co-methylation network. As expected, the previously discussed lifespan-related modules were enriched with CpGs implicated by our EWAS of lifespan: midnightblue (hypergeometric test P = $2.2x10^{-47}$; 67/249 overlapped CpGs), greenyellow (hypergeometric P = $2.1x10^{-36}$; 70/398 overlapped CpGs), tan (hypergeometric P = $6.7x10^{-23}$; 52/365 overlapped CpGs), and green (hypergeometric P = $5.0x10^{-18}$; 104/1542 overlapped CpGs) module.

In total, 1006 genes had a differential methylation association with lifespan (union of cognate genes resulting from the marginal model analysis for lifespan and lifespanAdjWeight). The gene expression levels of 17 of these genes exhibited a highly significant repeated measures correlation with chronological age (repeated measures Cor p value < 10^{-50}) in different human tissues (**Fig. 5C**). Two of these genes, *PTCHD4* and *ZBTB7B,* were also implicated by EWAS of weight-adjusted lifespan (lifespanAdjWeight). The cognate genes next to the top 500 positively lifespan-related CpGs play a critical role in animal organ morphogenesis (marginal model lifespan GREAT enrichment false discovery rate, FDR = 3x10-4 and LifespanAdjWeight FDR=$3.3x10^{-7}$, Fig. 5D), increased rib number in mice (FDR=$1x10^{-21}$, **Fig. 5D**), and implicates the HOXL gene group (FDR = 0.004 and weight adjusted LifespanAdjWeight FDR=$1.3x10^{-15}$), and abnormal survival in mice (FDR <$4x10^{-4}$).

Upstream regulators of maximum lifespan

We employed Ingenuity Pathway analysis (*40*) to identify potential upstream regulators of the genes cognate to the top 500 positively and top 500 negatively lifespan-related CpGs. The top-ranked candidate regulators of both gene lists included SOX2-OCT4-NANOG pluripotency factors (FDR = $5.7x10^{-4}$ lifespan negative, FDR = $5.7x10^{-4}$ lifespan positive), which play critical roles in cellular reprogramming. We performed a control analysis that ruled out potential confounding by sequence conservation (**fig. S25**). Upstream regulators also included several candidates related to development: sonic hedgehog (SHH), lifespan negative FDR = $1.3x10^{-4}$; POU4F2, lifespan negative FDR = $3.3x10^{-7}$ and ASCL1, lifespan negative FDR = $1.6x10^{-3}$ (**Fig. 5E**). These findings suggest that expression of lifespan-related genes might be regulated to some extent by pluripotency factors. This prompted us to investigate whether expression of any of the lifespan-related genes identified above are altered by transient expression of pluripotency inducing factors (Yamanaka factors OSKM) in a mouse model (*39*). Indeed, this analysis revealed that transient expression of OSKM altered the expression of 195 out of 646 lifespan-related genes in skin and 166 lifespan-related genes in the kidney (nominal Fisher exact p=$3.9x10^{-52}$ for skin and lifespan; p=$1.4x10^{-42}$ for kidney and lifespan, **Fig. 5F, fig. S32**). Genomic positions that are known to be bound by pluripotency factors (in at least one human/murine cell type according to ChIP-seq data from Encode) are located near CpGs that are associated with maximum species lifespans: NANOG binding sites are enriched for CpGs that are positively correlated with lifespan (FDR=0.002) and to CpGs underlying the midnightblue module (FDR=0.0006), which has high methylation levels in long-lived species (**Fig. 5G**). OCT4 (POU5F1) (FDR=0.02) and cMYC (FDR=0.003) binding sites are enriched with CpGs in the greenyellow module, which has low methylation levels in long-lived species (**Fig. 5G**). The ChIP-seq binding location analysis also

implicates other noteworthy factors such as POLII, CTCF, RAD21, YY1, and TAF1, which show the strongest enrichment for negatively lifespan-related CpGs (**Fig. 5G**).

Given the role of CTCF in regulating the 3D organization of the genome, we conducted an enrichment analysis of Topologically Associating Domain (TAD) boundaries and loop boundaries identified in both human and mouse cell lines (**fig. S26**). We found that both TAD and loop boundaries demonstrated significant enrichment of negatively lifespan-related CpGs (FDR=$3x10^{-4}$ for TAD boundaries and FDR=$6.7x10^{-4}$ for loop boundaries in various cell lines, such as olfactory receptor cells, as well as human fibroblasts IMR90 and HFFc6; **fig. S26**). This finding aligns with the significant enrichment observed for CTCF (FDR=10^{-7}).

CpGs Linked to Lifespan in Various Taxonomic Orders and Tissues

To pinpoint CpGs associated with log maximum lifespan independent of phylogenetic order or tissue type, we conducted a meta-analysis of EWAS findings from 25 distinct strata, comprising phylogenetic order and tissue type. Using a non-parametric meta-analysis approach (rankPvalue), we assessed the EWAS of lifespan (meta lifespan) in these strata to identify CpGs unconfounded by tissue type or phylogenetic order (table S24). Our meta.lifespan results demonstrated significant overlap with the previously mentioned EWAS of lifespan in all eutherian species (hypergeometric P = $1x10^{-175}$, **Fig. 6A**). In contrast, none of the meta.lifespan CpGs overlapped with EWAS of age, which further support the statement that methylation of lifespan-related CpGs does not change with age in mammalian tissues. The top 4 CpGs from the meta.lifespan analysis are depicted in **Fig. 6B**, showing significant positive correlations for CpGs near LOXL1 and ZSCAN29 (exons), and negative correlations for those near RAB29 (exon) and GATA3 (downstream) with log maximum lifespan across various taxonomic orders and tissue types. Similar to our above mentioned results, CpGs implicated by our meta lifespan analysis (FDR<0.05) overlap significantly (FDR<0.01) with genes involved in organ morphogenesis, RNA biosynthesis, increased rib number in mice, Wnt signaling (**Fig. 6C**), and genes altered by transient expression of pluripotency-inducing factors in mouse models (nominal Fisher exact p<10^{-5} for skin and lifespan meta; p<10^{-11} for kidney and lifespan meta, **Fig. 6D**).

Chromatin state analysis

Our large-scale mammalian DNAm data confirms that CpGs located in promoter regions (-2000 to 2000 bp of TSS regions) have low methylation levels (**Fig. 7A**, mean=15%). In contrast, those in gene bodies and distal regions are highly methylated (**Fig. 7A,** mean value ~70%). CpGs having a high/low mean methylation level tend to have positive/negative Z statistics for lifespan, respectively (**Fig. 7B-C**). We find that CpGs with low methylation levels in long-lived species are located close to the transcriptional start site of genes and near binding sites of polycomb repressive complex 1 (PRC1, p=$6.4x10^{-11}$, **Fig. 7D**) and polycomb repressive complex 2 (PRC2, p=$2x10^{-6}$). To test the hypothesis that long-lived species exhibit high/low methylation levels in chromosomal regions that are expected to have high/low methylation patterns, we used chromatin states that were identified and annotated based on over 1000 epigenetic data sets encompassing a diverse range of human cell and tissue types (*41*).

The lifespan related CpGs are enriched with transcriptional start site chromatin state (TSS1, p=$2.5x10^{-12}$), and flanking promoter states (PromF4, p=$5.6x10^{-10}$; PromF5, p=$2.0x10^{-9}$; PromF2, p=$3.0x10^{-4}$, **Fig. 7D**).

The CpGs with high methylation levels in blood samples of long-lived species are enriched in gene body associated states (notably transcribed state TxEx1, $p=7.5x10^{-8}$ and highly transcribed state TxEx4 $p=1.7x10^{-6}$, **Fig. 7E**). Detailed description of the chromatin state enrichment for EWAS of maximum lifespan is in the supplementary text.

A bi-clustering analysis between chromatin states and co-methylation modules based on fold enrichments (**Fig. 8; table S21; table S22**) revealed that the 55 mammalian co-methylation modules fall into three large groupings (referred to as meta modules). The bar plot to the left of **Fig. 8** shows different mean methylation levels of the CpGs underlying the 3 meta modules: mean methylation=0.23, 0.66, and 0.77 for meta modules 1, 2, and 3, respectively.

Meta module 1 contains several chromatin states that are associated with polycomb repression, including bivalent regulatory regions (BivProm1, 2) and ReprPC1. Further, meta module 1 contains chromatin states related to transcriptional start sites (TSS1, TSS2), and several flanking promoters (PromF2,3,4,5). TSS1, PromF2, and PromF4-5 (associated with negatively lifespan-related CpGs) were previously associated as the universal chromatin states with the strongest enrichments for CpG islands (54-101 fold) (*41*). The color band underneath **Fig. 8** reveals that six modules underlying meta-module 1 are sensitive to murine lifespan interventions. Meta module 1 is enriched with CpGs that have low methylation levels in long lived species (significant overlap with EWAS of lifespan, tan/greenyellow modules, **Fig. 8**).

Meta-module 2 can be considered as a partially methylated module (mean methylation 0.66) and is enriched with several enhancer states, late replicating domains (partially methylated domains, common PMD (*42*)), and solo CpGs (WCGW,(*42*)). Meta-module 2 also contains the module most significantly related with lifespan (midnightblue) and the human mortality risk module (magenta). These two modules overlap with the CpGs that are positively related to lifespan. Three out of four average weight-related modules are also located in meta-module 2.

Discussion:

In this study, we present an analysis of the most extensive cross-species DNA methylation dataset to date, obtained from a mammalian array platform. This platform specifically focuses on highly conserved regions of DNA, making it a valuable resource for studying methylation patterns across mammalian species (*5*). The successful construction of mammalian phyloepigenetic trees suggests that the divergence of DNA methylation profiles is closely aligned with genetic changes throughout evolution. Numerous sensitivity assessments reveal that the observed phyloepigenetic associations are not due to technical issues associated with our measurement platform. Instead, the phyloepigenetic signal may stem from sources like upstream regulators, transcription factors, or DNA sequence variations in distant regions.

The conserved CpGs exhibiting the strongest phylogenetic signals are situated in intergenic regions, while promoter regions do not display such signals. Previous studies report a rapid evolutionary rate of enhancers as a shared feature among mammalian genomes, while promoters demonstrate either full or partial conservation across species (*2*).

We found that 30 of the resulting 55 modules identified from an unsupervised machine learning method were readily associated with species traits (taxonomic order, maximum lifespan, average adult weight) or individual traits (chronological age, tissue, sex). We expect that many of the remaining 25 modules will be associated with biological characteristics about which we currently have no information. As a case in point, although the yellow module was not associated with any of our primary tested traits, it did show association with response to a murine circadian rhythm

disruption study (light pollution during the night, **fig. S7B**). The upstream regulator analysis of the EWAS of lifespan identified the pluripotency transcription factors (OCT4, SOX2, and NANOG). We show that the transient overexpression of OSKM in murine tissues affects the methylation levels of CpGs near genes implicated by our EWAS of maximum lifespan (Fig. 5E). We speculate that the enhanced activity of the pluripotency network in long-lived species results in more efficient tissue repair and maintenance, ensuring a longer lifespan.

Both the EWAS and eigengene-based analysis identified methylation signatures of maximum lifespan, and most of these were independent of aging, and presumably set at birth, and could be stable predictors of lifespan at any point in life. Several CpGs that are more highly methylated in long-lived species are located near *HOXL* genes and other genes that play a role in morphogenesis and development. Some of these lifespan-related CpGs are located next to genes that are also implicated in our analysis of upstream regulators (e.g., *ASCL1* and *SMAD6*). CpGs with methylation levels that are inversely related to lifespan are enriched in transcriptional start site (TSS1) and promoter flanking (PromF4, PromF5) associated chromatin states. Genes located in chromatin state TSS1 are constitutively active and enriched for nucleic acid metabolic processes (*41*). This suggests that long-living species evolved mechanisms that maintain low methylation levels in chromatin states that would favor higher expression levels of selected genes that are potentially essential for an organism's survival.

References and Notes

1. S. Xiao *et al.*, Comparative epigenomic annotation of regulatory DNA. *Cell* **149**, 1381-1392 (2012).
2. D. Villar *et al.*, Enhancer evolution across 20 mammalian species. *Cell* **160**, 554-566 (2015).
3. J. Qu *et al.*, Evolutionary expansion of DNA hypomethylation in the mammalian germline genome. *Genome research* **28**, 145-158 (2018).
4. J. Klughammer *et al.*, Comparative analysis of genome-scale, base-resolution DNA methylation profiles across 580 animal species. *Nature Communications* **14**, 232 (2023).
5. A. Arneson *et al.*, A mammalian methylation array for profiling methylation levels at conserved sequences. *Nat Commun* **13**, 783 (2022).
6. K. M. Parsons *et al.*, DNA methylation-based biomarkers for ageing long-lived cetaceans. *Molecular ecology resources*, (2023).
7. A. T. Lu *et al.*, Universal DNA methylation age across mammalian tissues. *bioRxiv*, 2021.2001.2018.426733 (2021).
8. P. Kordowitzki *et al.*, Epigenetic clock and methylation study of oocytes from a bovine model of reproductive aging. *Aging Cell* **20**, e13349 (2021).
9. N. A. Prado *et al.*, Epigenetic clock and methylation studies in elephants. *Aging Cell* **20**, e13414 (2021).
10. T. R. Robeck *et al.*, Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Commun Biol* **4**, 642 (2021).
11. B. Larison *et al.*, Epigenetic models developed for plains zebras predict age in domestic horses and endangered equids. *Communications Biology* **4**, 1412 (2021).
12. K. Mozhui *et al.*, Genetic loci and metabolic states associated with murine epigenetic aging. *Elife* **11**, (2022).
13. V. J. Sugrue *et al.*, Castration delays epigenetic aging and feminizes DNA methylation at androgen-regulated loci. *Elife* **10**, (2021).
14. T. R. Robeck *et al.*, Multi-Tissue Methylation Clocks for Age and Sex Estimation in the Common Bottlenose Dolphin. *Frontiers in Marine Science* **8**, (2021).
15. S. Horvath *et al.*, Methylation studies in Peromyscus: aging, altitude adaptation, and monogamy. *Geroscience* **44**, 447-461 (2022).
16. S. Horvath *et al.*, Epigenetic clock and methylation studies in marsupials: opossums, Tasmanian devils, kangaroos, and wallabies. *Geroscience* **44**, 1825-1845 (2022).

17. S. Horvath *et al.*, Epigenetic clock and methylation studies in the rhesus macaque. *Geroscience* **43**, 2441-2453 (2021).
18. S. Horvath *et al.*, DNA methylation age analysis of rapamycin in common marmosets. *Geroscience* **43**, 2413-2425 (2021).
19. A. J. Jasinska *et al.*, Epigenetic clock and methylation studies in vervet monkeys. *Geroscience*, (2021).
20. K. Raj *et al.*, Epigenetic clock and methylation studies in cats. *Geroscience* **43**, 2363-2378 (2021).
21. K. M. Schachtschneider *et al.*, Epigenetic clock and DNA methylation analysis of porcine models of aging and obesity. *Geroscience* **43**, 2467-2483 (2021).
22. M. L. Cossette *et al.*, Epigenetics and island-mainland divergence in an insectivorous small mammal. *Molecular ecology* **32**, 152-166 (2023).
23. J. F. Lemaitre *et al.*, DNA methylation as a tool to explore ageing in wild roe deer populations. *Molecular ecology resources* **22**, 1002-1015 (2022).
24. S. Horvath *et al.*, DNA methylation clocks tick in naked mole rats but queens age more slowly than nonbreeders. *Nat Aging* **2**, 46-59 (2022).
25. G. S. Wilkinson *et al.*, Author Correction: DNA methylation predicts age and provides insight into exceptional longevity of bats. *Nat Commun* **12**, 2652 (2021).
26. G. S. Wilkinson *et al.*, DNA methylation predicts age and provides insight into exceptional longevity of bats. *Nat Commun* **12**, 1615 (2021).
27. S. Horvath *et al.*, DNA methylation aging and transcriptomic studies in horses. *Nature Communications* **13**, 40 (2022).
28. G. M. Pinho *et al.*, Hibernation slows epigenetic ageing in yellow-bellied marmots. *Nature Ecology & Evolution*, (2022).
29. S. Horvath *et al.*, DNA methylation clocks for dogs and humans. *Proceedings of the National Academy of Sciences of the United States of America* **119**, e2120887119 (2022).
30. K. J. Peters *et al.*, An epigenetic DNA methylation clock for age estimates in Indo-Pacific bottlenose dolphins (Tursiops aduncus). *Evolutionary Applications* **n/a**, (2022).
31. P. Chiavellini *et al.*, Hippocampal DNA Methylation, Epigenetic Age and Spatial Memory Performance in Young And Old Rats. *J Gerontol A Biol Sci Med Sci*, (2022).
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).
33. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
34. P. Langfelder, R. Luo, M. C. Oldham, S. Horvath, Is my network module preserved and reproducible? *PLoS Comput Biol* **7**, e1001057 (2011).
35. D. Szklarczyk *et al.*, 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).
36. A. Barrat, M. Barthelemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3747-3752 (2004).
37. H. Pilcher. (Nature Publishing Group, 2003).
38. V. Acosta-Rodriguez *et al.*, Circadian alignment of early onset caloric restriction promotes longevity in male C57BL/6J mice. *Science* **376**, 1192-1202 (2022).
39. K. C. Browder *et al.*, In vivo partial reprogramming alters age-associated molecular changes during physiological aging in mice. *Nature Aging* **2**, 243-253 (2022).
40. A. Kramer, J. Green, J. Pollard, Jr., S. Tugendreich, Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-530 (2014).
41. H. Vu, J. Ernst, Universal annotation of the human genome through integration of over a thousand epigenomic datasets. *Genome Biol* **23**, 9 (2022).
42. W. Zhou *et al.*, DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet* **50**, 591-602 (2018).
43. P. Langfelder, S. Horvath. (UCLA, 2014).
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).
45. E. P. Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

46. J. P. de Magalhaes, 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).
47. 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 Research* **46**, e123-e123 (2018).
48. 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).
49. J. Felsenstein, Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am J Hum Genet* **25**, 471-492 (1973).
50. S. W. Kembel *et al.*, Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463-1464 (2010).
51. J. Felsenstein, Phylogenies and the Comparative Method. *The American Naturalist* **125**, 1-15 (1985).
52. A. Grafen, The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **326**, 119-157 (1989).
53. C. Y. McLean *et al.*, GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501 (2010).
54. S. Sayols, rrvgo: a Bioconductor package to reduce and visualize Gene Ontology terms. *Aust Dent J*, (2020).
55. A. V. Segre, L. Groop, V. K. Mootha, M. J. Daly, D. Altshuler, Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* **6**, (2010).
56. S. Horvath *et al.*, Pan-primate DNA methylation clocks. *bioRxiv*, 2020.2011.2029.402891 (2020).
57. N. E. Schlabritz-Loutsevitch *et al.*, Metabolic adjustments to moderate maternal nutrient restriction. *British journal of nutrition* **98**, 276-284 (2007).
58. S. Morgello *et al.*, The National NeuroAIDS Tissue Consortium: a new paradigm in brain banking with an emphasis on infectious disease. *Neuropathol Appl Neurobiol* **27**, 326-335. (2001).
59. S. Horvath *et al.*, HIV, pathology and epigenetic age acceleration in different human tissues. *Geroscience*, (2022).
60. S. Horvath *et al.*, Perinatally acquired HIV infection accelerates epigenetic aging in South African adolescents. *AIDS* **32**, 1465-1474 (2018).
61. S. Horvath, B. R. Ritz, Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany NY)* **7**, 1130-1142 (2015).
62. 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).
63. L. L. Sailer *et al.*, Pair bonding slows epigenetic aging and alters methylation in brains of prairie voles. *bioRxiv*, 2020.2009.2025.313775 (2020).
64. E. N. Burns *et al.*, Generation of an equine biobank to be used for Functional Annotation of Animal Genomes project. *Animal genetics* **49**, 564-570 (2018).
65. L. Tan *et al.*, Naked Mole Rat Cells Have a Stable Epigenome that Resists iPSC Reprogramming. *Stem cell reports* **9**, 1721-1734 (2017).
66. E. K. Bors *et al.*, An epigenetic clock to estimate the age of living beluga whales. *Evolutionary Applications*, (2020).
67. S. Lubetkin, J. Zeh, J. George, Statistical modeling of baleen and body length at age in bowhead whales (*Balaena mysticetus*). *Canadian Journal of Zoology* **90**, 915 - 931 (2012).
68. W. Koski, J. George, J. Zeh, J. Brandon, Preliminary analyses on identifying yearling bowhead whales (Balaena mysticetus) in aerial photographs. *Unpublished paper to IWC. SC/62/BRG329*, (2010).
69. 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).
70. C. Rosa *et al.*, 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. *Marine Mammal Science* **29**, 424-445 (2013).
71. E. H. Harley, M. H. Knight, C. Lardner, B. Wooding, M. Gregor, The Quagga project: progress over 20 years of selective breeding. *African Journal of Wildlife Research* **39**, 155-163 (2009).
72. S. Horvath *et al.*, Reversing age: dual species measurement of epigenetic age with a single clock. *bioRxiv*, 2020.2005.2007.082917 (2020).
73. J. Plassais *et al.*, Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat Commun* **10**, 1489 (2019).

74. J. Plassais *et al.*, Analysis of large versus small dogs reveals three genes on the canine X chromosome associated with body weight, muscling and back fat thickness. *PLOS Genetics* **13**, e1006661 (2017).
75. AmericanKennelClub, *The Complete Dog Book: 20th Edition*. (Ballantine Books, 2006).
76. B. Wilcox, C. Walkowicz, *The Atlas of Dog Breeds of the World*. (T.F.H. Publications, 1995).
77. K. Coschigano *et al.*, 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).
78. V. A. Acosta-Rodríguez, F. Rijo-Ferreira, C. B. Green, J. S. Takahashi, Importance of circadian timing for aging and longevity. *Nature Communications* **12**, 2862 (2021).
79. T. J. Little *et al.*, Methylation-Based Age Estimation in a Wild Mouse. *bioRxiv*, 2020.2007.2016.203687 (2020).
80. 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).
81. M. Roller *et al.*, LINE retrotransposons characterize mammalian tissue-specific and evolutionarily dynamic regulatory regions. *Genome Biol* **22**, 62 (2021).
82. L. Yan *et al.*, OSAT: a tool for sample-to-batch allocations in genomics experiments. *BMC Genomics* **13**, 689 (2012).
83. A. Seluanov *et al.*, Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. *Proceedings of the National Academy of Sciences* **106**, 19352-19357 (2009).
84. 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).
85. 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).
86. M. J. Aryee *et al.*, Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-1369 (2014).
87. A. Anonymous, Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials* **19**, 61-109 (1998).
88. G. Anderson *et al.*, Implementation of the women's health initiative study design. *Ann Epidemiol* **13**, S5-17 (2003).
89. M. Nasu *et al.*, Mammalian-specific sequences in pou3f2 contribute to maternal behavior. *Genome Biol Evol* **6**, 1145-1156 (2014).
90. P. Dammann, R. Sumbera, 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).
91. 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).
92. L. D. Moore, T. Le, G. Fan, DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23-38 (2013).
93. J. Z. Bakdash, L. R. Marusich, Repeated Measures Correlation. *Front Psychol* **8**, 456 (2017).
94. S. S. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).

Acknowledgments:

Steve Horvath, Adriana Arneson, and Jason Ernst are inventors on patent/patent application (publication number WO2020150705) held/submitted by University of California Los Angeles that covers the mammalian methylation array technology. Nicolas Cermakian, Steve Brown and Stuart Peirson were involved in the mouse study of light pollution. Images of animals are from Phylopic ("http://phylopic.org") or Wikimedia, which are under public domains or CC BY 3.0 license ("https://creativecommons.org/licenses/by/3.0/"). Fibroblasts from multiple species were accumulated using resources of the Longevity Consortium (Richard A. Miller). DNA samples from several species were provided by Museum of Vertebrate Zoology (MVZ), Museum of Southwestern Biology (MSB), University of New Mexico, and Chris J. Conroy from the

University of California, Berkeley, which are linked by the Arctos Museum database (https://arctosdb.org/).

Sonja C Vernes was supported by a Max Planck Research Group Award from the Max Planck Gesellschaft and a UKRI Future Leaders Fellowship (MR/T021985/1). Nigel Bennett was funded by a DST-NRF SARChI chair of Mammalian Behavioural Ecology and Physiology (GUN 64756). Collection of plains zebra samples was supported by the National Geographic Society to Brenda Larison (8941-11).

Funding:

The majority of the funding was contributed by the Paul G. Allen Frontiers Group (Horvath) and Open Philanthropy (Horvath). Additional financial support for specific aspects of data generation was obtained from the following sources:

National Geographic Society grant 8941-11 (BL),

British Heart Foundation (FS/18/39/33684) (DV),

Wellcome (WT202878/Z/16/Z) (DTO),

European Research Council (788937) (DTO),

Cancer Research UK (20412) (DTO),

Max Planck Research Group Award from the Max Planck Gesellschaft (SCV)

UKRI Future Leaders Fellowship (MR/T021985/1) (SCV)

DST-NRF SARChI chair of Mammalian Behavioural Ecology and Physiology (GUN 64756) (NCB)

Science Foundation Ireland Future Frontiers 19/FFP/6790 (ECT)

National Institute on Aging Intramural Research Program, NIH. (JAM), U01 AG060908 (Horvath), AG055841 (Mozhui), and AG043930/AG/NIA (Williams)

National Institute of Health AG065403 (Gladyshev), AG047200 (Gladyshev), AG067782 (Gladyshev) and AG076607 (Gladyshev), AG047200 (Gorbunova).

A.H. Schultz Foundation, Department of Evolutionary Anthropology, University of Zurich (Michael Krützen),

Leverhulme Trust RPG-2019-404 (Amy Pederson),

UCLA Jonsson Comprehensive Cancer Center and Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research Ablon Scholars Program (Jason Ernst)

Funds provided by the College of Computer, Mathematical and Natural Sciences at the University of Maryland, College Park (Gerald Wilkinson)

Taronga Conservation Society Australia (Jusine K. O'Brien)

Author contributions:

Conceptualization: SH, AH

Formal Analysis: AH, CZL, ATL, QY, JZH

Investigation: All authors

Visualization: AH, CZL

Funding acquisition: SH, BL, DV, DTO, SCV, NCB, ECT, JAM

Project administration: SH

Supervision: SH

Writing – original draft: AH, CZL

Writing – review & editing: AH, SH, KR, JE and all other co-authors

Competing interests:

Steve Horvath, Adriana Arneson, and Jason Ernst are inventors on patent/patent application

(publication number WO2020150705) held/submitted by University of California Los Angeles that covers the mammalian methylation array technology.

S.H. and Robert Brooke are founders of the non-profit Epigenetic Clock Development Foundation, which has licensed several patents from UC Regents, and distributes the mammalian methylation array.

Data and materials availability:

All data from the Mammalian Methylation Consortium is posted on Gene Expression Omnibus (Complete dataset: GSE223748). Subsets of the data sets can also be downloaded from accession numbers, GSE174758, GSE184211, GSE184213, GSE184215, GSE184216, GSE184218, GSE184220, GSE184221, GSE184224, GSE190660, GSE190661, GSE190662, GSE190663, GSE190664, GSE174544, GSE190665, GSE174767, GSE184222, GSE184223, GSE174777, GSE174778, GSE173330, GSE164127, GSE147002,GSE147003, GSE147004). The mammalian array platform is distributed by the non-profit Epigenetic Clock Development Foundation (https://clockfoundation.org/).

The mammalian data can also be downloaded from the Clock Foundation webpage: https://clockfoundation.org/MammalianMethylationConsortium.

The manifest file of the mammalian array, genome annotations of the CpGs, and codes can be found on Github (doi: 10.5281/zenodo.7574747).

Method summary

The Mammalian Methylation Consortium generated cytosine methylation data from n=15,456 DNA samples derived from 70 tissue types of 348 mammalian species (331 eutherians, 15 marsupials, 2 monotremes) using a custom-designed mammalian methylation array that targets CpGs at conserved loci in mammals (*5*). DNA methylation data were used for phyloepigenetic tree development using 1-cor dissimilarity applied to mean methylation values per species. The choice of the correlation-based dissimilarity matrix is justified in Supplementary Methods.

For unsupervised analysis, we formed WGCNA networks based on two sets of CpG 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 second network was a subset of 7,956 conserved CpGs in 11,105 samples from 167 eutherian and nine marsupial species. In addition, we developed seven consensus co-methylation networks to remove the confounding effects of species and tissue type. Consensus WGCNA can be interpreted as a meta-analysis across networks in different species and tissue types (*33, 43*).

For the eutherian network (Net1), module eigengenes (MEs) were defined as singular vectors (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, replication 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 was examined for different traits using individual regression models.

EWAS of lifespan was done in 28,318 CpGs that apply to mice and humans according to 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 were highly skewed, we imposed a log-transformation on these phenotypes before conducting EWAS. Our tissue type specific EWAS was conducted in tissues with enough species (N>25 species) available. 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 the conservative Bonferroni adjustment. We report a false discovery rate in our enrichment studies to adjust for multiple comparisons.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S32

Tables S1 to S24

Data S1 to S19

References (1–94)

Fig. 1. Phyloepigenetic trees parallel the mammalian evolutionary tree. (A) the traditional phylogenetic tree from the TimeTree database (*44*) based on 321 (out of 348) species in our study. A full description of the species in our study is reported in **table S1**. (**B)** Blood-based phyloepigenetic tree created from hierarchical clustering of DNA methylation data in this study (additional analysis in fig. S3A,B). We formed the mean value per cytosine across samples for each species. The clustering used 1 minus the Pearson correlation (1-cor) as a pairwise dissimilarity measure and the average linkage method as intergroup dissimilarity. Phyloepigenetic trees for skin and liver can be found in fig. S2. Additional analyses, e.g., involving different choices of CpGs or intergroup dissimilarity measures, are reported in the supplement (**fig. S2**). The colored bars reflect the branch height. **(C)** Scatter plot of the distances in blood phyloepigenetic (1-cor) vs

the traditional evolutionary tree. **(D)** Scatter plots displaying the log-odds ratios of regions exhibiting significant phylogenetic signals relative to the Transcription Start Site (TSS) are presented. The phylogenetic signal is determined using Blomberg's K statistic (*32*). 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. Additional analysis in **fig. S4**.

Fig. 2. DNA methylation network relates to species and individual characteristics in mammalian species. (**A)** the WGCNA network of 14,705 conserved CpGs in eutherian species (Network 1). The identified modules related to species, or individual sample characteristics. Network 1 modules were compared to eight additional networks (fig. S5). The modules with strong associations with species and sample characteristics were labeled below the dendrogram. Grey color codes CpGs that are outside of modules. **(B)** summary of the modules that showed strong associations with species and individual sample characteristics. The +/- labels are the direction of association with each trait. **(C)** Top defined functional biological processes related to network 1 modules (details in **fig. S9**, **table S4**). **(D)** mammalian co-methylation modules form clusters of proteins in the STRING protein-protein interaction (PPI) network. For the sake of visualization, the analysis was limited to the top 50 CpGs with the highest module membership value per module. colors: mammalian network 1. The lollipop plot shows the global cluster coefficient (*36*) of the proteins within a module (up to 500 top CpGs) in a PPI network. Our permutation analysis matched the distribution of the original module sizes. We evaluated 1100 random permutations, i.e. 20 for each of the 55 modules. The boxplot reports the global clustering coefficient per module (y-axis) versus permutation status: module resulting from a random selection of proteins (left) versus original module resulting from WGCNA (right). The modules with cluster coefficients larger than the maximum permutation cluster coefficient were considered as significant at p=0.001. The dashed vertical line corresponds to the maximum global clustering coefficient observed in the 1100 random permutations.

Fig. 3. Co-methylation modules related to mammalian maximum lifespan, weight, human mortality, and age. Modules associated with log maximum lifespan ($p<10^{-20}$) **(A)** or log average species weight ($p<10^{-17}$) (**B)** in marginal association: correlation test with the mean module eigengene of the species. The module eigengene is defined as the 1st principal component of the scaled CpGs underlying a module. The species are randomly labeled by their animal number (**table S1**). (**C)** The top modules associated with median life expectancy, upper limit life expectancy, or average adult weight of 93 dog breeds, model: marginal correlation test of the mean module eigengene with target variables (detailed breed characteristics are in **table S8**). R, Pearson correlation coefficient; p, correlation test p-value. (**D)** Forest plots of the top modules associated with mortality risk in the Framingham Heart Study Offspring Cohort (FHS), and Women's Health Initiative (WHI) study, totaling 4651 individuals (1095, 24% death). The right panel indicates the number of deaths/total number of individuals in each study. We report the meta-analysis p-value in the title of the forest plot. **(E)** Module that correlates significantly ($p<1x10^{-300}$) with relative age (defined as ratio of age/maximum lifespan) across mammalian species using a multivariate regression model. Covariates: tissue, sex, and species differences. Each dot corresponds to a

eutherian tissue sample (n=14,542). Dots are colored by taxonomic order. (**F)** The volcano plot of the rmCorrelation of all purple module genes in GTEx data (Additional analysis in fig. S11).

Fig. 4. The effects of different pro-aging and anti-aging interventions on selected DNAm modules. Six DNA methylation modules are sensitive to lifespan-related intervention experiments and relate to the life expectancy of the mouse models. (**A)** Changes in the intervention modules in the liver parallel smaller size and longer life expectancy of growth hormone receptor mouse models (GHRKO). Sample size: GHRKO, 11 (5 female, 6 male); Wt, 18 (9 male, 9 female). The age range: 6–8 months. (**B)** Caloric restriction (CR) DNA methylation module signature predicts longer lifespan in this treated group. Age=18 months; Sex=Male; N=CR, 59; control, 36. (**C)** High-fat diet accelerates aging in the age module. N=high-fat diet, 133 (125 females, 8 males); control (*ad libitum*), 212 (202 females, 10 males). Age range: 3–32 months. (**D), (E), (F)** Examining the effects of *in vivo* partial reprogramming on intervention modules. (**D)** a schematic view of the partial programming experiment in 4F mice (*39*). A systemic Yamanaka factors expression (Oct4, Sox2, Klf4, Myc) was periodically induced by adding doxycycline to drinking water for two days per week. The partial programming was done at three different durations. Sample size: control (C57BL/6+dox), n=7; 1m 4F, n=3; 7m 4F, n=5; 10m 4F, n=3 (all tissues except skin, n=2 for skin). (**E), (F)** scatter plots of the linear changes of the intervention modules in the skin (**E**) and kidney (**F**) of mice treated with different durations (dosages) of Yamanaka factors. R, Pearson correlation coefficient; p, correlation test p-value. The intervention modules indicate a dose-dependent rejuvenation of skin and kidney by this partial programming regimen.

Fig. 5. Epigenome-wide association study (EWAS) of mammalian log-transformed maximum lifespan. (A) The figure represents the CpG-specific association with maximum lifespan across n=333 eutherian species. For EWAS, the mean methylation values of each CpG (per species) were regressed on log maximum lifespan. The right portion of the panel reports EWAS results after adjustment for average adult weight. Genome annotation: human hg19. Red dotted line: Bonferroni corrected two-sided p-value < $1.8x10^{-6}$. The point colors indicate the corresponding modules. The bar plot indicates the top enriched (hypergeometric test, eutherian probes as background) modules for the top 1000 (500 negative CpGs nominal p<$1.1x10^{-11}$, FDR=$1x10^{-10}$; 500 negative CpGs positive CpGs nominal p<$1.5x10^{-21}$, FDR=$7.5x10^{-20}$) significant CpGs for different EWAS. **(B)** Venn diagram of the overlaps between top hits from EWAS of maximum lifespan and meta-analysis of age (meta-analysis results from (*7*), additional analysis in fig. S20). **(C)** Venn diagram of the overlaps between the genes adjacent to the EWAS results and top age-related mRNA changes in human tissues (p<1e-50). (**D)** Gene set enrichment analysis of the genes proximal to CpGs associated with mammalian maximum lifespan. We only report enrichment terms that are significant after adjustment for multiple comparisons (hypergeometric FDR <0.01) and contain at least five significant genes. The top three significant terms per column (EWAS) and enrichment database are shown in the panel. (**E)** Ingenuity potential upstream regulator analysis (*40*) of the differentially methylated genes related to mammalian maximum lifespan. (**F)** Venn diagram of 3 gene lists. First, the top 646 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 (601 genes) and skin (695 genes) (*39*). We observe significant overlap between the gene lists (nominal Fisher exact p=$3.9x10^{-52}$ for skin and lifespan; p=$1.4x10^{-42}$ for kidney and lifespan) (**G)** transcriptional factor motif enrichment analysis of lifespan modules and lifespan related CpGs. The enrichment results for

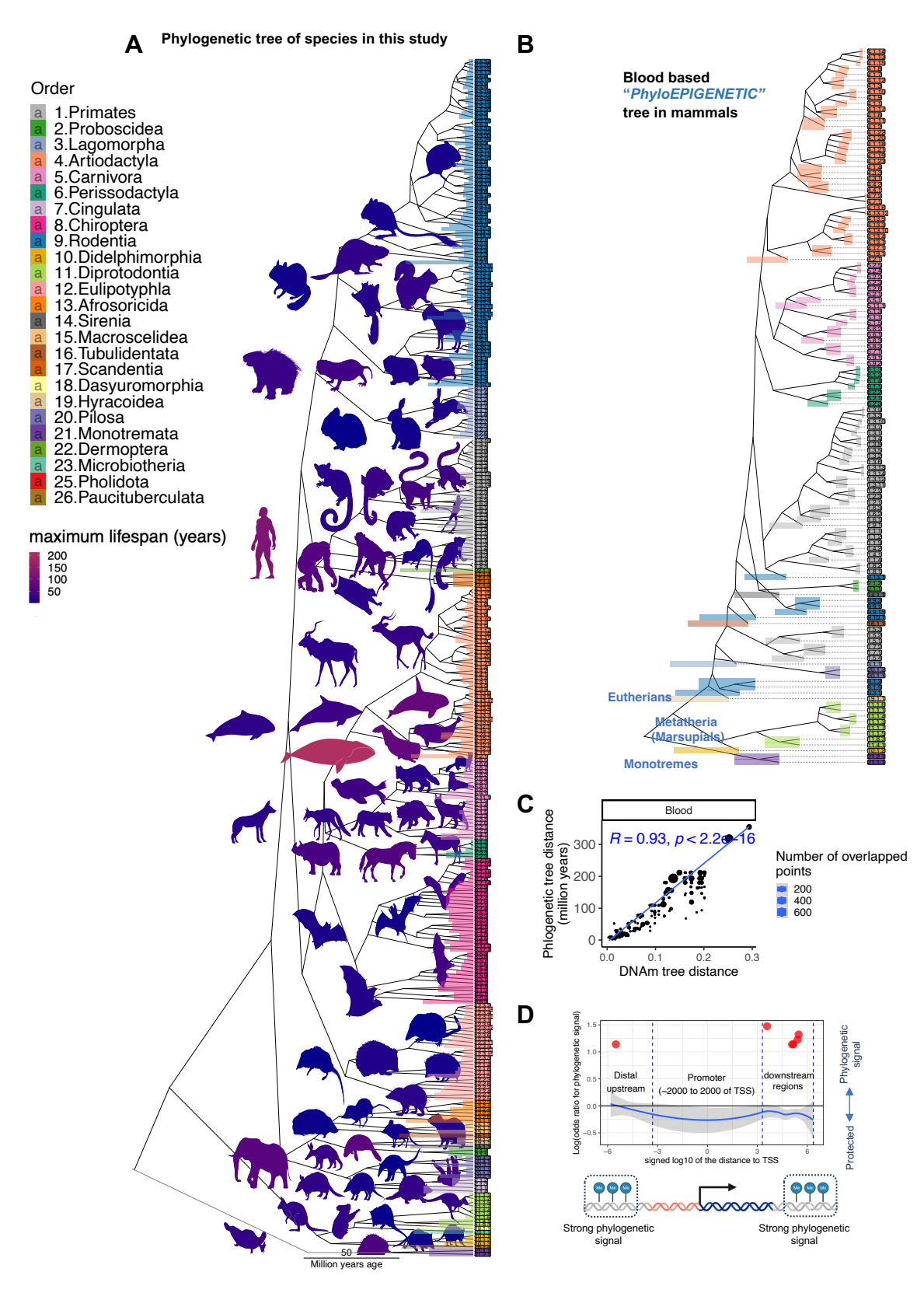
LifespanAdjWeight.negative were not significant. The overlap is assessed by a hypergeometric test for the CpGs within the motifs based on the human hg19 genome.

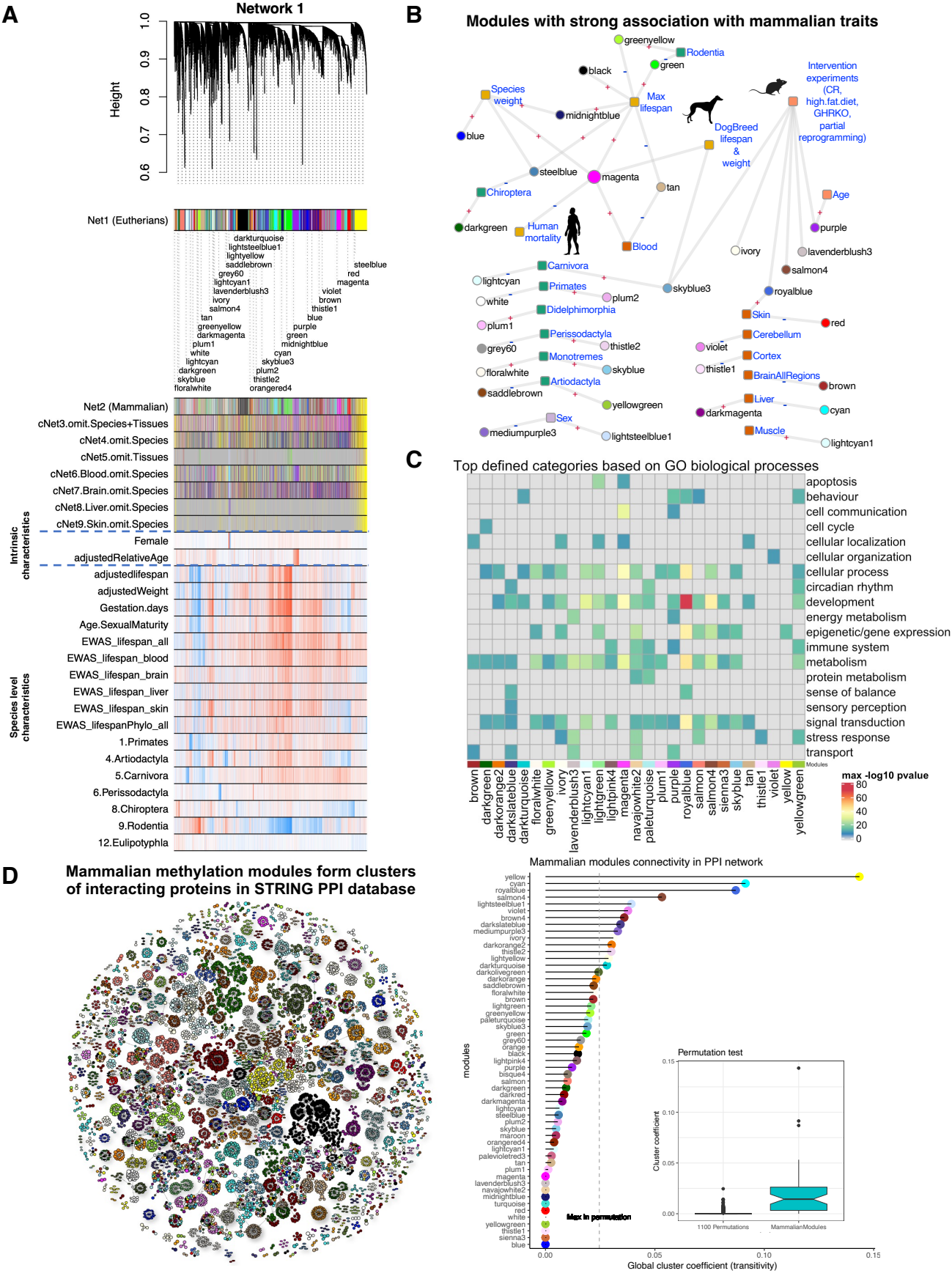
Fig. 6. CpGs Linked to Lifespan in Various Taxonomic Orders and Tissues. Using the non-parametric rankPvalue method (*33*), we combined 25 EWAS of lifespan results from various taxonomic order or tissue type strata, calculating the significance of a CpG's consistently high (or low) rank based on the 25 EWAS of log maximum lifespan (meta lifespan, underlying EWAS results can be found in table S24, Data S19). (A) The overlap of top 1000 (500 per direction) meta-lifespan CpGs with EWAS of lifespan in all eutherians (nominal Fisher exact $p=1x10^{-175}$). (B) Scatter plots illustrating the top meta-lifespan CpGs categorized into different tissue-phylogenetic order strata are presented. Each panel displays only the strata that exhibit significant relationships. The first row represents the phylogenetic order strata combining all tissues. (C) Gene set enrichment analysis of the genes proximal to CpGs associated with mammalian maximum lifespan. We only report enrichment terms that are significant after adjustment for multiple comparisons (hypergeometric FDR <0.01) and contain at least five significant genes. The top three significant terms per column (EWAS) and enrichment database are shown in the panel. (D) Venn diagram of 3 gene lists. First (the bottom circle), the top 407 genes adjacent to 1000 meta-lifespan CpGs (500 positive and 500 negative). Gene lists 2 and 3 (the top circles) 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 (601 genes) and skin (695 genes) (*39*).

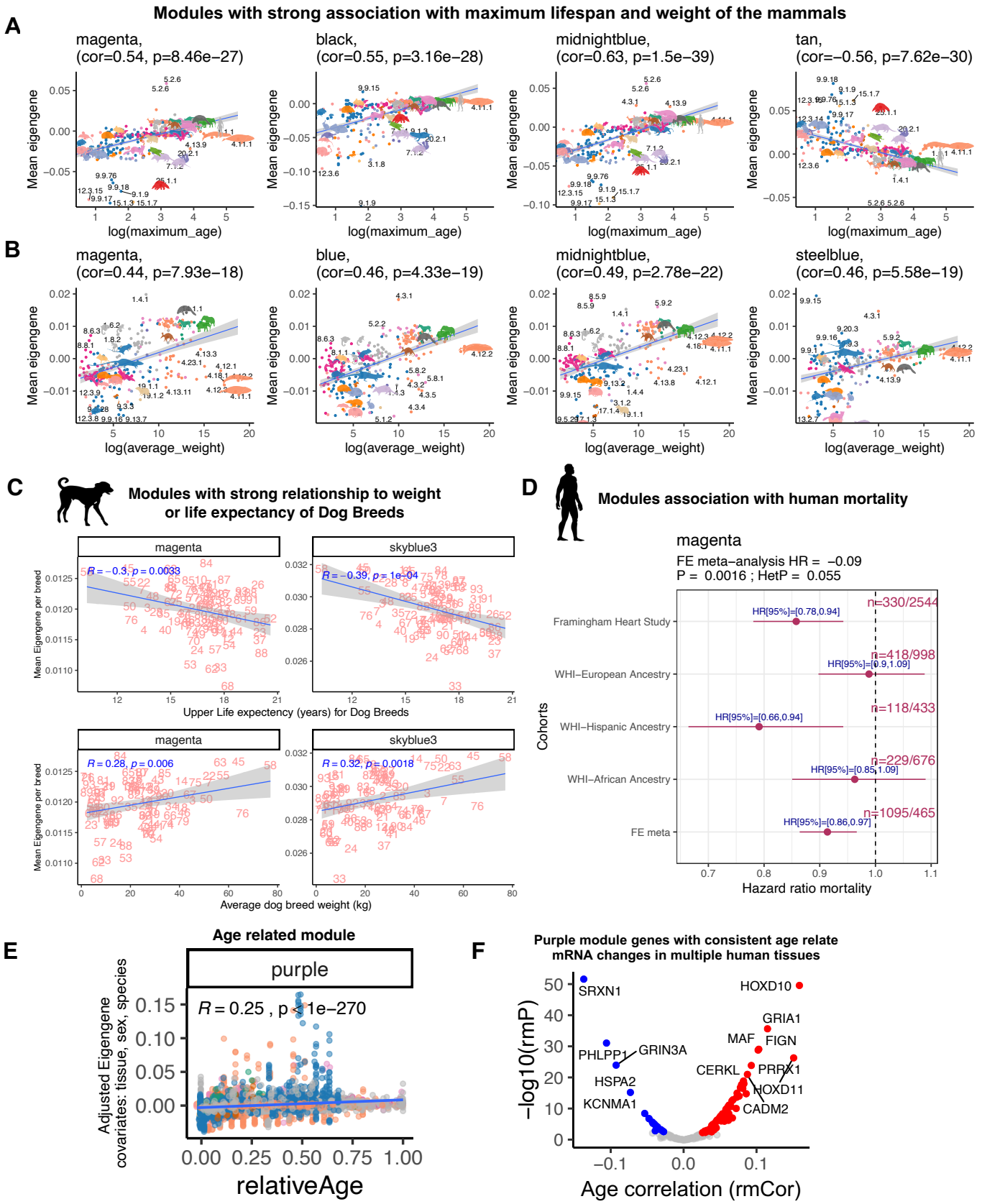
Fig. 7. Chromatin state analysis and distance to the transcriptional start site for the lifespan-related CpGs. (A) scatter plot showing, for each CpG (each datapoint), mean methylation across species (y-axis) vs. distances to the nearest transcription start site (x-axis). The color and shape of each datapoint corresponds to the chromatin state and gene region annotation for each CpG site, as in legend below (B). (**B)** For each CpG, the epigenome-wide association study (EWAS) Z-statistics for log maximum lifespan. the distance to the nearest transcriptional start site. **(C)** scatter plots showing, for each CpG (each datapoint), mean methylation in eutherians and EWAS Z statistics for log maximum lifespan in different genomic regions (intergenic, promoter, gene body). The CpGs are colored based on the overlapping chromatin state in the human genome, and shaped based on the annotated gene region, as in legend on right. Additional EWAS results after adjustment for phylogenetic relationships can be found in **fig. S17-20** and corresponding enrichment results can be found in **fig. S22-S24**. Pearson correlation coefficients and p-values are reported in different panels. Chromatin state enrichment analysis of (**D)** the top 500 negatively lifespan related CpGs, and (**E)** top 500 positively lifespan related CpGs . The columns in each panel correspond to EWAS results for log transformed maximum lifespan across i) all tissues combined (Lifespan.All), ii) blood samples only (Lifespan.Blood), iii) skin samples only (Lifespan.Skin), meta-lifespan and the corresponding results after adjustment for average adult weight, Lifespan(AdjWeight). The last column reports enrichment with respect to the relativeAge$^{(+)}$ module (purple). We use the same significance thresholds as in Figure 5. The cells' shadings correspond to fold enrichment between co-methylation modules andeach chromatin state . The cells' numeric values correspond to the p-value of such enrichments based on the hypergeometric test, and only cells' values with significant p-value<0.001 (equivalent to FDR<0.02) are shown. The chromatin states are learned based on epigenetic datasets profiling chromatin mark signals in different human cell and tissue types, resulting in a genome annotation

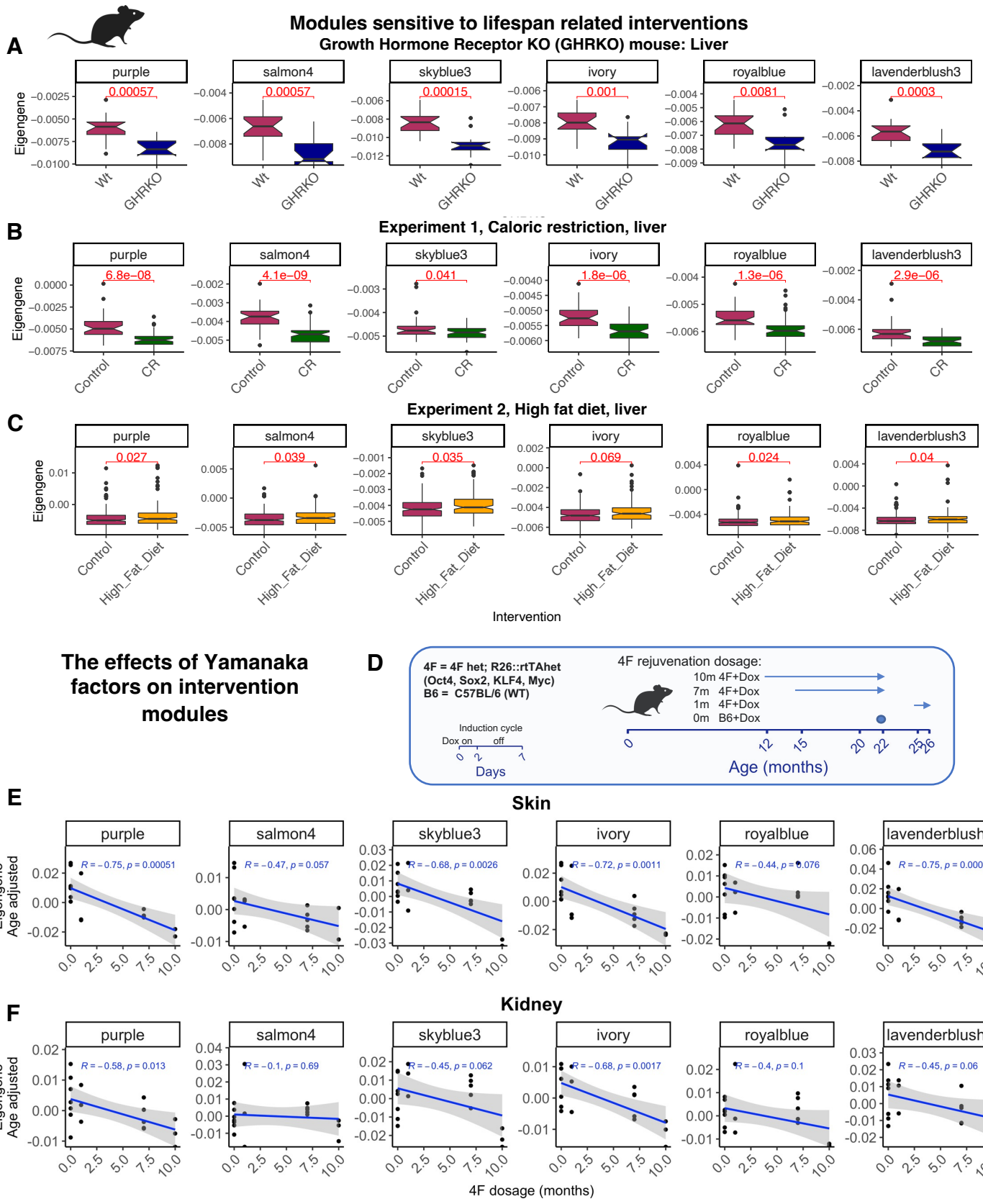
shared across cell types (*41*). The common partially methylated domains (commonPMD), solo CpGs (WCGW), and highly methylated domain (HMD) annotations are from (*42*). Polycomb repressor complexes (PRC) 1 and PRC 2 binding sites are obtained from the ChIP-seq datasets of PCR 1 and 2 from ENCODE (ENCODE Project Consortium, 2012)(*45*).

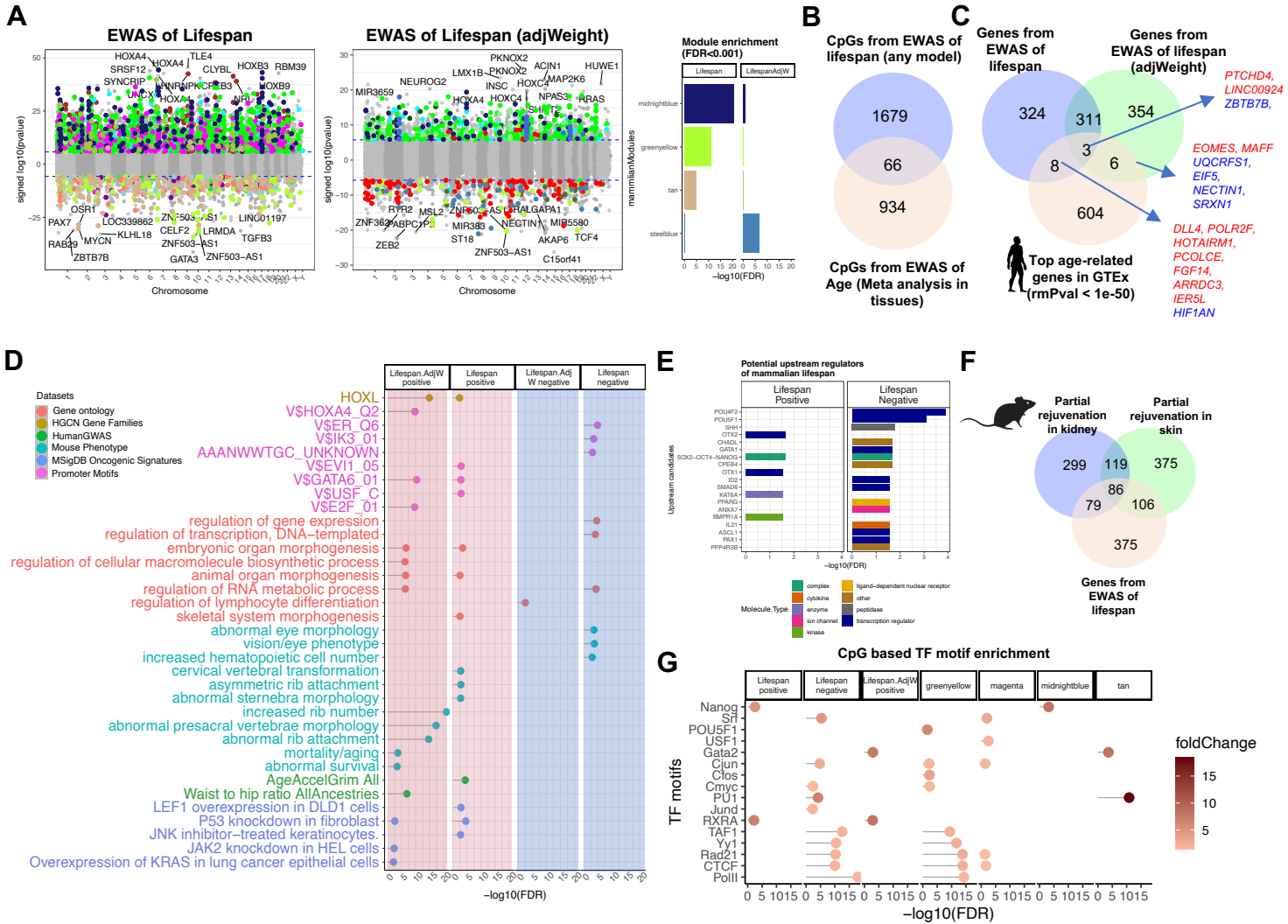
Fig. 8. Mammalian methylation meta modules based on the chromatin states and external genome annotations. The heatmap shows the enrichments between (1) mammalian co-methylation modules and significant lifespan-related EWAS CpG groups (x-axis) EWAS and (2) chromatin states or other genomic annotation (y-axis). The cells' shadings correspond to log transformed fold-enrichment values (observed count divided by expected count). Hypergeometric tests were used to evaluate the enrichment significance in each cell, and * indicates a nominal p-value<0.001 (FDR<0.10). Only chromatin states and external genome annotations with at least one significant enrichment (FDR<0.10) are shown. The chromatin states are based on a human based universal chromatin annotation of human cell and tissue types (*41*). Other genomic annotations include the common partially methylated domains (commonPMD), solo CpGs (WCGW), and highly methylated domain (HMD) annotations, which are from (*42*). In addition, Polycomb repressor complexes (PRC) 1 and PRC 2 binding sites are defined from the ChIP-seq data of PRC 1 and 2 from ENCODE (ENCODE Project Consortium, 2012)(*45*). The row and column hierarchical clustering trees (average linkage) are based on a dissimilarity (1 minus the pairwise Pearson correlation between log transformed fold enrichment values). The left side barplot indicates the mean methylation levels of the CpGs in each state for all eutherian samples in our data. We used the 14,705 eutherian CpGs as the background for enrichment of the co-methylation modules. In contrast, 28,318 CpGs (high quality probes in humans and mice) were used as a background for enrichment of significant lifespan-related EWAS CpG groups with chromatin states and genome annotations. Each EWAS CpG group includes up to 500 most significant CpGs per direction (positively/negatively related with lifespan) as detailed in the caption of **Fig. 5**.

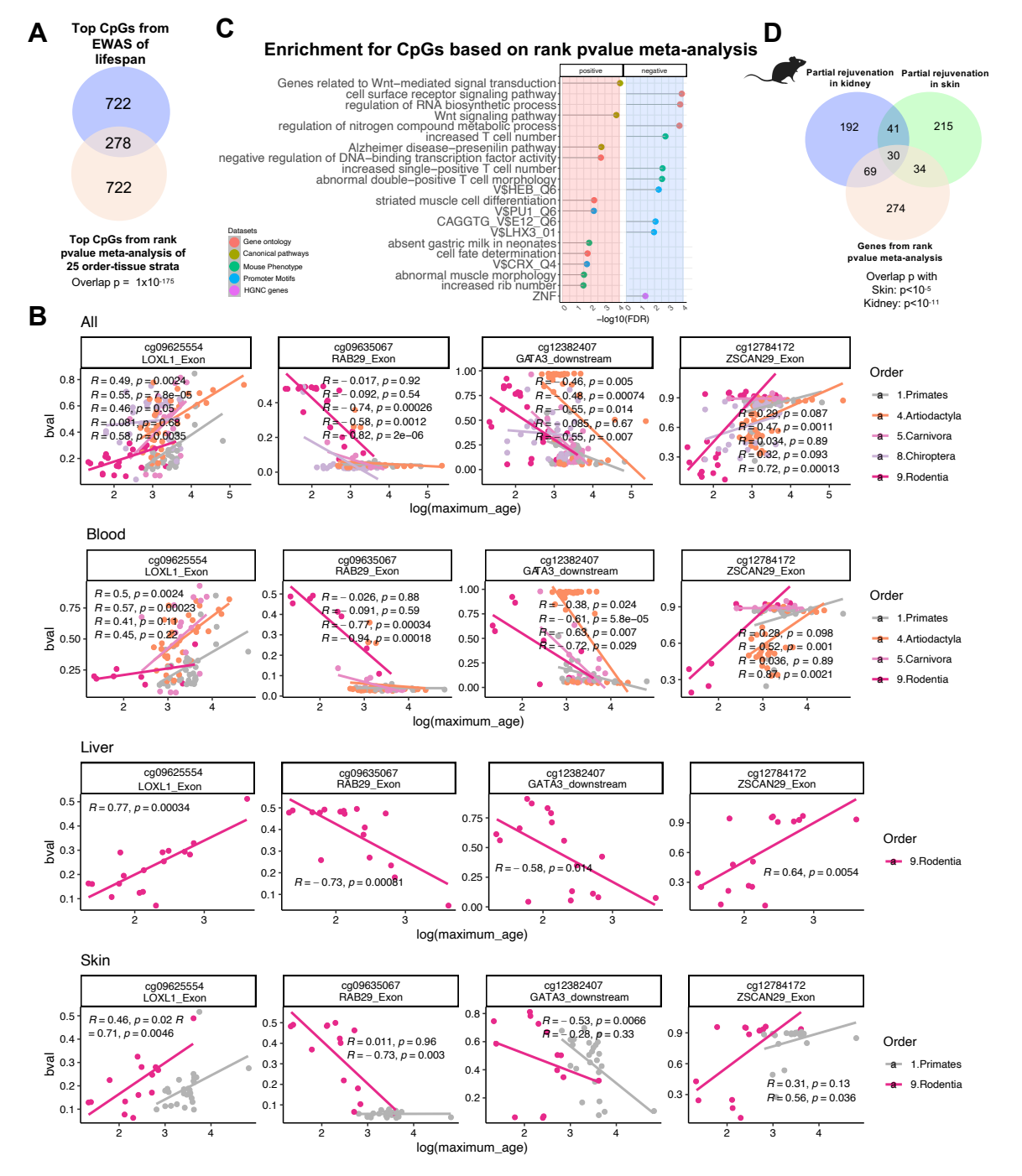

A
Phylogenetic tree of species in this study
Order
1.Primates
2.Proboscidea
3.Lagomorpha
4.Artiodactyla
5.Carnivora
6.Perissodactyla
7.Cingulata
8.Chiroptera
9.Rodentia
10.Didelphimorphia
11.Diprotodontia
12.Eulipotyphla
13.Afrosoricida
14.Sirenia
15.Macroscelidea
16.Tubulidentata
17.Scandentia
18.Dasyuromorphia
19.Hyracoidea
20.Pilosa
21.Monotremata
22.Dermoptera
23.Microbiotheria
25.Pholidota
26.Paucituberculata
maximum lifespan (years)
200
150
100
50
50
Million years age
B
Blood based
"PhyloEPIGENETIC"
tree in mammals
Eutherians
Metatheria
(Marsupials)
Monotremes
C
Blood
R = 0.93, p < 2.2e−16
Phlogenetic tree distance
(million years)
DNAm tree distance
Number of overlapped
points
200
400
600
D
Log(odds ratio for phylogenetic signal)
Distal
upstream
Promoter
(−2000 to 2000 of TSS)
downstream
regions
signed log10 of the distance to TSS
Phylogenetic
signal
Protected
Strong phylogenetic
signal
Strong phylogenetic
signal

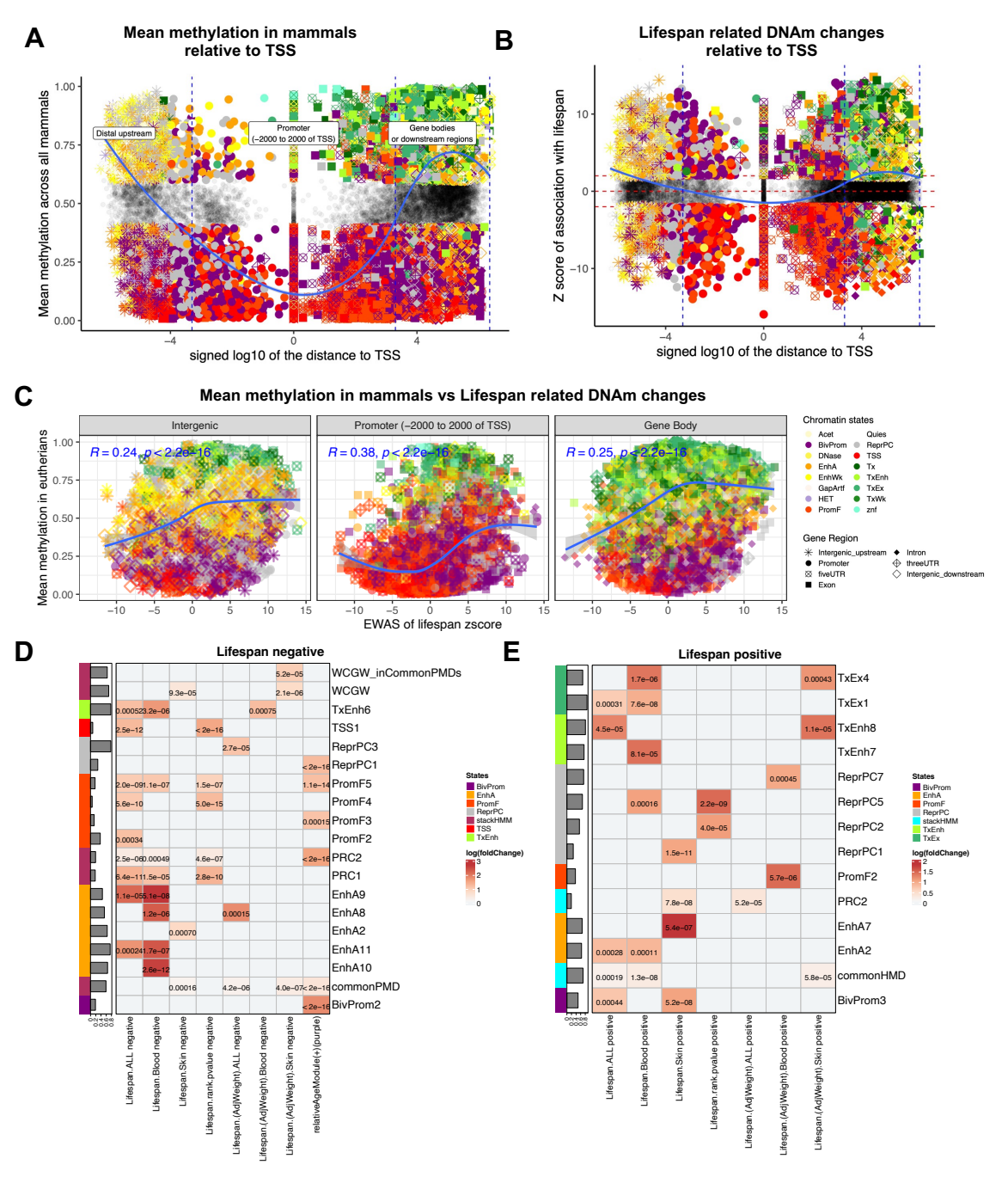

A
Network 1
Height
1.0
0.9
0.8
0.7
0.6
Net1 (Eutherians)
darkturquoise
lightsteelblue1
lightyellow
saddlebrown
grey60
lightcyan1
lavenderblush3
ivory
salmon4
tan
greenyellow
darkmagenta
plum1
white
lightcyan
darkgreen
skyblue
floralwhite
steelblue
red
magenta
violet
brown
thistle1
blue
purple
green
midnightblue
cyan
skyblue3
plum2
thistle2
orangered4
Net2 (Mammalian)
cNet3.omit.Species+Tissues
cNet4.omit.Species
cNet5.omit.Tissues
cNet6.Blood.omit.Species
cNet7.Brain.omit.Species
cNet8.Liver.omit.Species
cNet9.Skin.omit.Species
Intrinsic characteristics
Female
adjustedRelativeAge
adjustedlifespan
adjustedWeight
Gestation.days
Age.SexualMaturity
EWAS_lifespan_all
EWAS_lifespan_blood
EWAS_lifespan_brain
EWAS_lifespan_liver
EWAS_lifespan_skin
EWAS_lifespanPhylo_all
Species level characteristics
1.Primates
4.Artiodactyla
5.Carnivora
6.Perissodactyla
8.Chiroptera
9.Rodentia
12.Eulipotyphla
B
Modules with strong association with mammalian traits
greenyellow
Rodentia
green
black
Species weight
Max lifespan
midnightblue
blue
DogBreed lifespan & weight
Intervention experiments (CR, high.fat.diet, GHRKO, partial reprogramming)
steelblue
magenta
tan
Chiroptera
Age
darkgreen
Human mortality
purple
Blood
ivory
lavenderblush3
Carnivora
salmon4
lightcyan
Primates
skyblue3
royalblue
white
plum2
Didelphimorphia
Skin
red
plum1
Perissodactyla
Cerebellum
grey60
thistle2
violet
Monotremes
Cortex
floralwhite
skyblue
thistle1
Artiodactyla
BrainAllRegions
brown
saddlebrown
yellowgreen
Liver
Sex
cyan
darkmagenta
mediumpurple3
lightsteelblue1
Muscle
lightcyan1
C
Top defined categories based on GO biological processes
apoptosis
behaviour
cell communication
cell cycle
cellular localization
cellular organization
cellular process
circadian rhythm
development
energy metabolism
epigenetic/gene expression
immune system
metabolism
protein metabolism
sense of balance
sensory perception
signal transduction
stress response
transport
Modules
brown
darkgreen
darkorange2
darkslateblue
darkturquoise
floralwhite
greenyellow
ivory
lavenderblush3
lightcyan1
lightgreen
lightpink4
magenta
navajowhite2
paleturquoise
plum1
purple
royalblue
salmon
salmon4
sienna3
skyblue
tan
thistle1
violet
yellow
yellowgreen
max -log10 pvalue
80
60
40
20
0
Mammalian modules connectivity in PPI network
modules
yellow
cyan
royalblue
salmon4
lightsteelblue1
violet
brown4
darkslateblue
mediumpurple3
ivory
darkorange2
thistle2
lightyellow
darkturquoise
darkolivegreen
darkorange
saddlebrown
floralwhite
brown
lightgreen
greenyellow
paleturquoise
skyblue3
green
grey60
orange
black
lightpink4
purple
bisque4
salmon
darkgreen
darkred
darkmagenta
lightcyan
steelblue
plum2
skyblue
maroon
orangered4
lightcyan1
palevioletred3
tan
plum1
magenta
lavenderblush3
navajowhite2
midnightblue
turquoise
red
white
yellowgreen
thistle1
sienna3
blue
Max in permutation
0.00
0.05
0.10
0.15
Global cluster coefficient (transitivity)
Permutation test
Cluster coefficient
0.15
0.10
0.05
0.00
1100 Permutations
MammalianModules
D
Mammalian methylation modules form clusters of interacting proteins in STRING PPI database


A
Modules with strong association with maximum lifespan and weight of the mammals
magenta, (cor=0.54, p=8.46e−27)
black, (cor=0.55, p=3.16e−28)
midnightblue, (cor=0.63, p=1.5e−39)
tan, (cor=−0.56, p=7.62e−30)
Mean eigengene
log(maximum_age)
B
magenta, (cor=0.44, p=7.93e−18)
blue, (cor=0.46, p=4.33e−19)
midnightblue, (cor=0.49, p=2.78e−22)
steelblue, (cor=0.46, p=5.58e−19)
log(average_weight)
C
Modules with strong relationship to weight or life expectancy of Dog Breeds
magenta
skyblue3
R = −0.3, p = 0.0033
R = −0.39, p = 1e−04
R = 0.28, p = 0.006
R = 0.32, p = 0.0018
Mean Eigengene per breed
Upper Life expectency (years) for Dog Breeds
Average dog breed weight (kg)
D
Modules association with human mortality
magenta
FE meta−analysis HR = −0.09
P = 0.0016 ; HetP = 0.055
Framingham Heart Study
WHI−European Ancestry
WHI−Hispanic Ancestry
WHI−African Ancestry
FE meta
Cohorts
HR[95%]=[0.78,0.94]
HR[95%]=[0.9,1.09]
HR[95%]=[0.66,0.94]
HR[95%]=[0.85,1.09]
HR[95%]=[0.86,0.97]
n=330/2544
n=418/998
n=118/433
n=229/676
n=1095/465
Hazard ratio mortality
E
Age related module
purple
R = 0.25 , p < 1e−270
Adjusted Eigengene
covariates: tissue, sex, species
relativeAge
F
Purple module genes with consistent age relate mRNA changes in multiple human tissues
SRXN1
PHLPP1
GRIN3A
HSPA2
KCNMA1
HOXD10
GRIA1
MAF
FIGN
CERKL
PRRX1
HOXD11
CADM2
−log10(rmP)
Age correlation (rmCor)


Modules sensitive to lifespan related interventions
A
Growth Hormone Receptor KO (GHRKO) mouse: Liver
purple
salmon4
skyblue3
ivory
royalblue
lavenderblush3
Eigengene
0.00057
0.00057
0.00015
0.001
0.0081
0.0003
Wt
GHRKO
B
Experiment 1, Caloric restriction, liver
6.8e−08
4.1e−09
0.041
1.8e−06
1.3e−06
2.9e−06
Control
CR
C
Experiment 2, High fat diet, liver
0.027
0.039
0.035
0.069
0.024
0.04
High_Fat_Diet
Intervention
The effects of Yamanaka factors on intervention modules
D
4F = 4F het; R26::rtTAhet (Oct4, Sox2, KLF4, Myc)
B6 = C57BL/6 (WT)
Induction cycle
Dox on
off
Days
4F rejuvenation dosage:
10m 4F+Dox
7m 4F+Dox
1m 4F+Dox
0m B6+Dox
Age (months)
E
Skin
Eigengene Age adjusted
R = −0.75, p = 0.00051
R = −0.47, p = 0.057
R = −0.68, p = 0.0026
R = −0.72, p = 0.0011
R = −0.44, p = 0.076
R = −0.75, p = 0.00056
F
Kidney
R = −0.58, p = 0.013
R = −0.1, p = 0.69
R = −0.45, p = 0.062
R = −0.68, p = 0.0017
R = −0.4, p = 0.1
R = −0.45, p = 0.06
4F dosage (months)


A
EWAS of Lifespan
EWAS of Lifespan (adjWeight)
signed log10(pvalue)
Chromosome
Module enrichment
(FDR<0.001)
Lifespan
LifespanAdjW
midnightblue
greenyellow
tan
steelblue
mammlianModules
−log10(FDR)
B
CpGs from EWAS of lifespan (any model)
1679
66
934
CpGs from EWAS of Age (Meta analysis in tissues)
C
Genes from EWAS of lifespan
Genes from EWAS of lifespan (adjWeight)
324
311
354
3
6
8
604
PTCHD4, LINC00924 ZBTB7B,
EOMES, MAFF UQCRFS1, EIF5, NECTIN1, SRXN1
DLL4, POLR2F, HOTAIRM1, PCOLCE, FGF14, ARRDC3, IER5L HIF1AN
Top age-related genes in GTEx (rmPval < 1e-50)
D
Datasets
Gene ontology
HGCN Gene Families
HumanGWAS
Mouse Phenotype
MSigDB Oncogenic Signatures
Promoter Motifs
Lifespan.AdjW positive
Lifespan positive
Lifespan.Adj W negative
Lifespan negative
HOXL
V$HOXA4_Q2
V$ER_Q6
V$IK3_01
AAANWWTGC_UNKNOWN
V$EVI1_05
V$GATA6_01
V$USF_C
V$E2F_01
regulation of gene expression
regulation of transcription, DNA−templated
embryonic organ morphogenesis
regulation of cellular macromolecule biosynthetic process
animal organ morphogenesis
regulation of RNA metabolic process
regulation of lymphocyte differentiation
skeletal system morphogenesis
abnormal eye morphology
vision/eye phenotype
increased hematopoietic cell number
cervical vertebral transformation
asymmetric rib attachment
abnormal sternebra morphology
increased rib number
abnormal presacral vertebrae morphology
abnormal rib attachment
mortality/aging
abnormal survival
AgeAccelGrim All
Waist to hip ratio AllAncestries
LEF1 overexpression in DLD1 cells
P53 knockdown in fibroblast
JNK inhibitor−treated keratinocytes.
JAK2 knockdown in HEL cells
Overexpression of KRAS in lung cancer epithelial cells
−log10(FDR)
E
Potential upstream regulators of mammalian lifespan
Lifespan Positive
Lifespan Negative
Upstream candidates
−log10(FDR)
Molecule.Type
complex
cytokine
enzyme
ion channel
kinase
ligand−dependent nuclear receptor
other
peptidase
transcription regulator
F
Partial rejuvenation in kidney
Partial rejuvenation in skin
299
119
375
86
79
106
375
Genes from EWAS of lifespan
G
CpG based TF motif enrichment
Lifespan positive
Lifespan negative
Lifespan.AdjW positive
greenyellow
magenta
midnightblue
tan
TF motifs
Nanog
Sri
POU5F1
USF1
Gata2
Cjun
Cfos
Cmyc
PU1
Jund
RXRA
TAF1
Yy1
Rad21
CTCF
PolII
−log10(FDR)
foldChange
15
10
5


A
Top CpGs from EWAS of lifespan
722
278
722
Top CpGs from rank pvalue meta-analysis of 25 order-tissue strata
Overlap p = 1x10-175
C
Enrichment for CpGs based on rank pvalue meta-analysis
positive
negative
Genes related to Wnt−mediated signal transduction
cell surface receptor signaling pathway
regulation of RNA biosynthetic process
Wnt signaling pathway
regulation of nitrogen compound metabolic process
increased T cell number
Alzheimer disease−presenilin pathway
negative regulation of DNA−binding transcription factor activity
increased single−positive T cell number
abnormal double−positive T cell morphology
V$HEB_Q6
striated muscle cell differentiation
V$PU1_Q6
CAGGTG_V$E12_Q6
V$LHX3_01
absent gastric milk in neonates
cell fate determination
V$CRX_Q4
abnormal muscle morphology
increased rib number
ZNF
−log10(FDR)
Datasets
Gene ontology
Canonical pathways
Mouse Phenotype
Promoter Motifs
HGNC genes
D
Partial rejuvenation in kidney
Partial rejuvenation in skin
192
41
215
30
69
34
274
Genes from rank pvalue meta-analysis
Overlap p with
Skin: p<10-5
Kidney: p<10-11
B
All
Blood
Liver
Skin
cg09625554
LOXL1_Exon
cg09635067
RAB29_Exon
cg12382407
GATA3_downstream
cg12784172
ZSCAN29_Exon
bval
log(maximum_age)
Order
1.Primates
4.Artiodactyla
5.Carnivora
8.Chiroptera
9.Rodentia


A
Mean methylation in mammals relative to TSS
Mean methylation across all mammals
Distal upstream
Promoter (−2000 to 2000 of TSS)
Gene bodies or downstream regions
signed log10 of the distance to TSS
B
Lifespan related DNAm changes relative to TSS
Z score of association with lifespan
signed log10 of the distance to TSS
C
Mean methylation in mammals vs Lifespan related DNAm changes
Intergenic
Promoter (−2000 to 2000 of TSS)
Gene Body
R = 0.24, p < 2.2e−16
R = 0.38, p < 2.2e−16
R = 0.25, p < 2.2e−16
Mean methylation in eutherians
EWAS of lifespan zscore
Chromatin states
Gene Region
D
Lifespan negative
WCGW_inCommonPMDs
WCGW
TxEnh6
TSS1
ReprPC3
ReprPC1
PromF5
PromF4
PromF3
PromF2
PRC2
PRC1
EnhA9
EnhA8
EnhA2
EnhA11
EnhA10
commonPMD
BivProm2
Lifespan.ALL negative
Lifespan.Blood negative
Lifespan.Skin negative
Lifespan.rank.pvalue negative
Lifespan.(Adj)Weight).ALL negative
Lifespan.(Adj)Weight).Blood negative
Lifespan.(Adj)Weight).Skin negative
relativeAgeModule(+)(purple)
States
log(foldChange)
E
Lifespan positive
TxEx4
TxEx1
TxEnh8
TxEnh7
ReprPC7
ReprPC5
ReprPC2
ReprPC1
PromF2
PRC2
EnhA7
EnhA2
commonHMD
BivProm3
Lifespan.ALL positive
Lifespan.Blood positive
Lifespan.Skin positive
Lifespan.rank.pvalue positive
Lifespan.(Adj)Weight).ALL positive
Lifespan.(Adj)Weight).Blood positive
Lifespan.(Adj)Weight).Skin positive
States
log(foldChange)

Enrichment of the modules for different chromatin and genomic states (* p<1e−3)

Meta module 1 | Meta module 2 | Meta module 3

Cluster 1 mainly TSS,PRC sites

PromF2
GapArtf1
neither
ReprPC1
PRC2
BivProm2
PromF6
PromF5
BivProm1
PRC1
TSS2
PromF3
TSS1
PromF4

Cluster 2 mainly Enhancer states

TxEnh6
PromF1
EnhA3
EnhA7
EnhWk5
PromF7
EnhA9
EnhA10
EnhA8
EnhA11
EnhA5
EnhA13
EnhA12
Tx3
EnhWk6
EnhA4
EnhA15
commonPMD
BivProm4
BivProm3
TxEnh1
EnhA6
TxEnh7
EnhA19
TxEx4
EnhWk2
EnhA1
EnhA2
EnhA17

Cluster 3 high mean methylation state

Quies5
ReprPC8
EnhA16
ReprPC6
EnhWk3
ReprPC3
ReprPC4
WCGW_inCommonPMDs
EnhWk4
commonHMD
WCGW
znf1
Quies2
Tx1
Quies3
Quies1
TxEnh5
TxEx1
TxEnh4
Tx5
TxWk1
Tx7
Tx6
TxWk2
TxEx2
TxEx3
Tx2
Tx4
Acet3
ReprPC2
ReprPC5
HET1
Acet2
HET6
Acet1
TxEnh8
TxEnh3
EnhA14

Mean DNAm in mammals (0, 0.2, 0.4, 0.6, 0.8)

States
Acet
BivProm
EnhA
EnhWk
GapArtf
HET
PMD annotation
PRC TFs
PromF
Quies
ReprPC
TSS
Tx
TxEnh
TxEx
TxWk
znf

log(fold enrichment)
6
4
2
0
−2
−4

Modules
ageModules
InterventionModules
maxLifespanModules
weightModules
humanMortalityModules
dogLifespanModules
orderModules
sexModules
tissueModules

tan
greenyellow
Lifespan negative
salmon4
purple
ivory
royalblue
lightgreen
lightcyan1
darkslateblue
yellowgreen
plum2
lightsteelblue1
darkmagenta
skyblue3
darkorange2
darkgreen
salmon
sienna3
floralwhite
plum1
skyblue
lightpink4
lavenderblush3
darkturquoise
thistle2
Lifespan.AdjW positive
magenta
midnightblue
Lifespan positive
brown
blue
turquoise
red
Lifespan.AdjW negative
green
paleturquoise
navajowhite2
cyan
white
thistle1
orangered4
palevioletred3
lightcyan
black
violet
darkolivegreen
grey60
darkorange
darkred
brown4
steelblue
yellow
lightyellow
mediumpurple3
bisque4
maroon
orange
saddlebrown