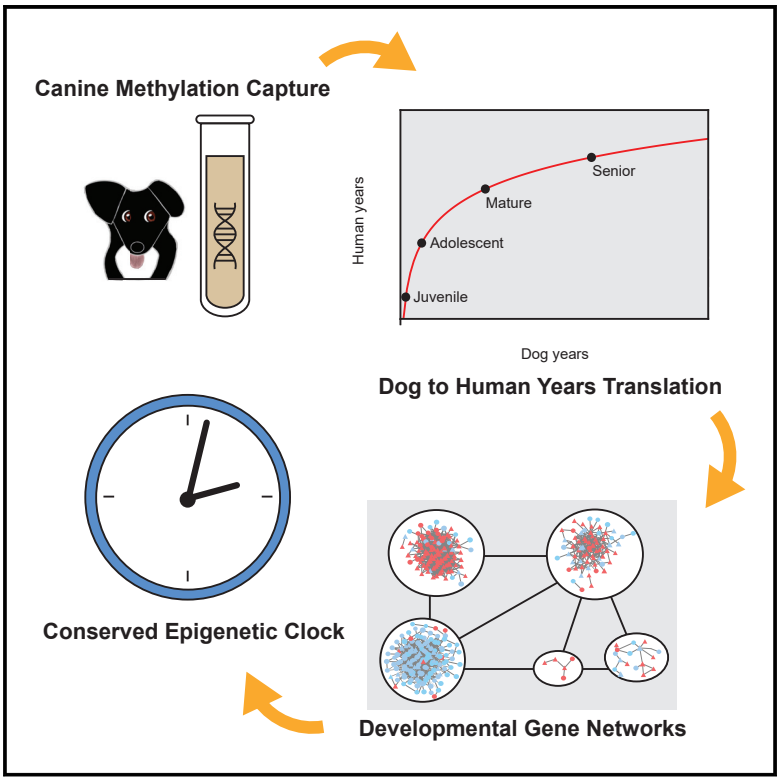
Cell Systems

Report

Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome

Graphical Abstract



Highlights

- Oligo-capture sequencing of methylomes from 104 Labradors, 0–16 years old
- Methylome similarity translates dog years to human years logarithmically
- Conserved age-related changes predominately impact developmental gene networks
- Formulation of a conserved epigenetic clock transferable across mammals

Authors

Tina Wang, Jianzhu Ma, Andrew N. Hogan, ..., Danika L. Bannasch, Elaine A. Ostrander, Trey Ideker

Correspondence

tideker@ucsd.edu

In Brief

Wang et al. create an oligo-capture system to characterize the canine DNA methylome, targeting syntenic regions of the genome conserved across all mammals. Cross-species comparisons reveal a nonlinear epigenetic signature that aligns the progression of life events in dogs, humans, and mice. This conserved signature occurs primarily in modules of developmental genes, leading the team to create a conserved epigenetic clock model of aging that can be trained and operated across different species.

Wang et al., 2020, Cell Systems *11*, 1–10
August 26, 2020
https://doi.org/10.1016/j.cels.2020.06.006

CellPress

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

Cell Systems

CellPress

Report

Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome

Tina Wang,[1] Jianzhu Ma,[1] Andrew N. Hogan,[2] Samson Fong,[3] Katherine Licon,[1] Brian Tsui,[1] Jason F. Kreisberg,[1] Peter D. Adams,[4] Anne-Ruxandra Carvunis,[5] Danika L. Bannasch,[6] Elaine A. Ostrander,[2] and Trey Ideker[1,3,7,*]

[1]Division of Genetics, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA
[2]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA
[3]Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA
[4]Sanford Burnham Prebys Medical Discovery Institute, San Diego, La Jolla, CA 92093, USA
[5]Department of Computational and Systems Biology, Pittsburgh Center for Evolutionary Biology and Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA
[6]Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA
[7]Lead Contact
*Correspondence: tideker@ucsd.edu
https://doi.org/10.1016/j.cels.2020.06.006

SUMMARY

All mammals progress through similar physiological stages throughout life, from early development to puberty, aging, and death. Yet, the extent to which this conserved physiology reflects underlying genomic events is unclear. Here, we map the common methylation changes experienced by mammalian genomes as they age, focusing on comparison of humans with dogs, an emerging model of aging. Using oligo-capture sequencing, we characterize methylomes of 104 Labrador retrievers spanning a 16-year age range, achieving >150× coverage within mammalian syntenic blocks. Comparison with human methylomes reveals a nonlinear relationship that translates dog-to-human years and aligns the timing of major physiological milestones between the two species, with extension to mice. Conserved changes center on developmental gene networks, which are sufficient to translate age and the effects of anti-aging interventions across multiple mammals. These results establish methylation not only as a diagnostic age readout but also as a cross-species translator of physiological aging milestones.

INTRODUCTION

The wisdom that every year in a dog's life equates to seven human years reflects our deep intuition that development and aging are conserved processes that occur at different rates in different species. All mammals, whether dog, human, or other creature, pass through similar life stages of embryogenesis, birth, infancy, youth, adolescence, maturity, and senescence (Withers, 1992). Although embryonic developmental programs have been relatively well studied, many of the molecular events governing postnatal life stages, including those tied to aging, are still unresolved (Khan et al., 2017). Over the past decade, it has become clear that a prominent molecular alteration during aging is remodeling of the DNA methylome, the pattern of epigenetic modifications whereby methyl groups are present at some cytosine-guanine dinucleotides (methyl-CpGs) but absent from others (Field et al., 2018). The methylation states of tens of thousands of CpGs have been found to change predictably over time, enabling the construction of mathematical models, known as "epigenetic clocks," that use these shifting patterns to accurately measure the age of an individual (Hannum et al., 2013; Horvath, 2013; Petkovich et al., 2017; Stubbs et al., 2017; Thompson et al., 2017; Wang et al., 2017).

Most questions regarding the relationship between DNA methylation and mammalian life stage remain unanswered (Figure 1). While the rate of methylation change appears to depend on maximal lifespan (Maegawa et al., 2017; Lowe et al., 2018), whether lifespan is the sole factor in determining how the methylome progresses with age, or if CpG states are aligned to specific intermediate milestones in development and aging, remains unknown. It is also unclear if the major epigenetic changes that occur with age involve the same or different (or random) CpG sites in different species. While DNA encoding a highly conserved ribosomal RNA family shows increasing methylation over time at the same conserved CpG sites in mice and humans (Wang and Lemos, 2019), epigenetic clocks trained in humans have not been strong predictors of age in other mammals (Petkovich et al., 2017; Stubbs et al., 2017). Perhaps this is not surprising, as epigenetic clocks are formulated using regression techniques that select only a small subset of the available CpG

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

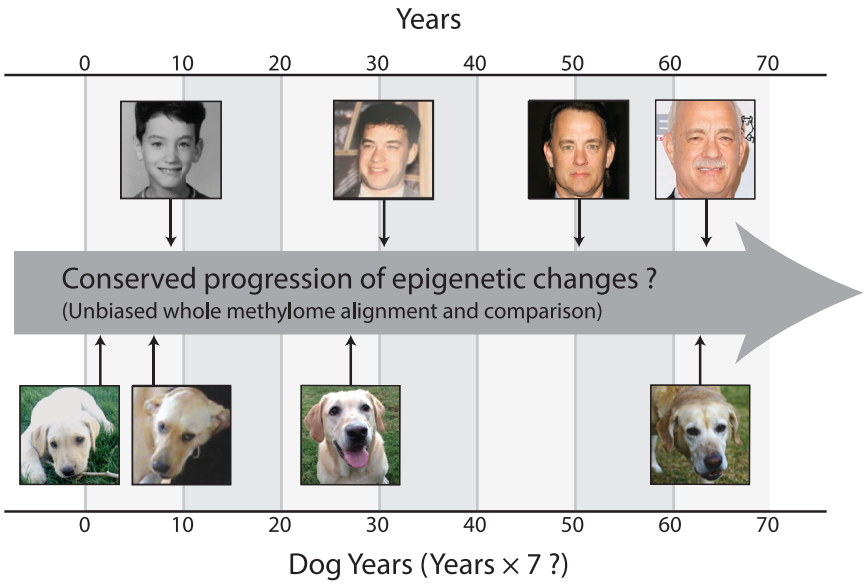


Figure 1. Physiological versus Epigenetic Change during Development and Aging

Aging yields similar physiological changes in humans and dogs, yet these changes occur along different time scales. Are these different timescales reflected in the progression of epigenetic changes observed during aging? If so, is this progression consistent with the adage "one dog year equals seven human years," or does it suggest a different cross-species alignment of time?

sites (<500) out of millions of potential subsets that can be used to measure age equivalently (Field et al., 2018).

Thus far, a major impediment to understanding the relationship between methylation and mammalian life stages has been our inability to characterize conserved epigenetic features in different mammals. Most studies of the human methylome use Illumina methylation arrays, which include oligonucleotide probes designed to explicitly measure the methylation states of >450,000 CpG sites (http://www.illumina.com/). Since a similar array platform has not yet been developed for other species, techniques such as whole-genome bisulfite sequencing (WGBS) have been used, which measure many CpGs without the need for a specific capture system but at relatively low coverage per sequencing run (Chatterjee et al., 2017). Other efforts, such as reduced representation bisulfite sequencing (RRBS), increase coverage for some CpGs but do not guarantee consistent measurements of the same sites across samples (Chatterjee et al., 2017). Therefore, what is needed is (1) to select a suitable model species, and (2) within this species, to develop a CpG capture system that can be more directly aligned with human methylation arrays.

Domestic dogs provide a unique opportunity to address these challenges (Gilmore and Greer, 2015; Kaeberlein et al., 2016). Dogs have been selectively bred by humans for occupation and esthetics (Ostrander et al., 2017), generating over 450 distinct breeds whose members share morphologic and behavioral traits. Most breeds derive from small numbers or popular sires, and dogs have been domesticated for only about 15,000–30,000 years (Vonholdt et al., 2010), leading to strong phenotypic and genetic homogeneity within breeds (Dreger et al., 2016). Dogs share nearly all aspects of their environment with humans, including factors associated with aging such as diet and chemical exposure. They also experience similar levels of health observation and health care intervention as humans (Gilmore and Greer, 2015; Kaeberlein et al., 2016). While the average lifespan differs dramatically across breeds, there is also considerable variability within some breeds, such as the Labrador retriever. Despite extensive phenotypic differences, however, all domestic breeds are members of the same species with a similar developmental, physiological, and pathological trajectory as humans (Gilmore and Greer, 2015; Kaeberlein et al., 2016). Importantly for aging studies, dogs accomplish this progression in many fewer years than humans, generally fewer than 20. Finally, epigenetic clocks have been demonstrated in dogs (Thompson et al., 2017), establishing them as a system for studies of age-related epigenetic remodeling.

Here, we develop the dog as a model system for epigenetic aging, using a specifically designed CpG oligo-capture system to generate high quality methylation data that align directly to human methylomes. Comparison of each dog methylome to its nearest human counterparts reveals a conserved but nonlinear progression of epigenetic changes, with rapid remodeling in puppies, relative to children, which slows markedly in canine adulthood. We show that most of the conserved epigenetic changes occur within specific developmental gene networks, such that a set of 439 conserved CpG sites is sufficient to build a pan-species epigenetic clock of aging. Unlike previous clocks, which are predictive in a single species only, the conserved developmental clock translates to multiple species with relative accuracy.

RESULTS

Characterization of Dog Methylomes with Syntenic Bisulfite Sequencing (SyBS)

Commonly used techniques such as WGBS and RRBS measure large numbers of random CpGs, but such CpGs are often not present in regions conserved across species. To enable high quality evolutionary comparisons of dog methylomes with other mammals, we performed targeted-bisulfite sequencing to systematically characterize CpGs in regions of the dog genome that are syntenic with those measured by Illumina human methylome arrays. Since Illumina arrays have been used to characterize epigenetic aging in many human studies (Alisch et al., 2012; Hannum et al., 2013; Horvath, 2013), our goal was to create a high quality panel of dog methylomes with substantial coverage of CpGs noted in prior human datasets (Figure 2A). Our strategy, henceforth called synteny bisulfite sequencing (SyBS), was designed to capture approximately 90,000 CpGs of the approximately 232,000 conserved CpGs on the Illumina array (STAR Methods).

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

CellPress

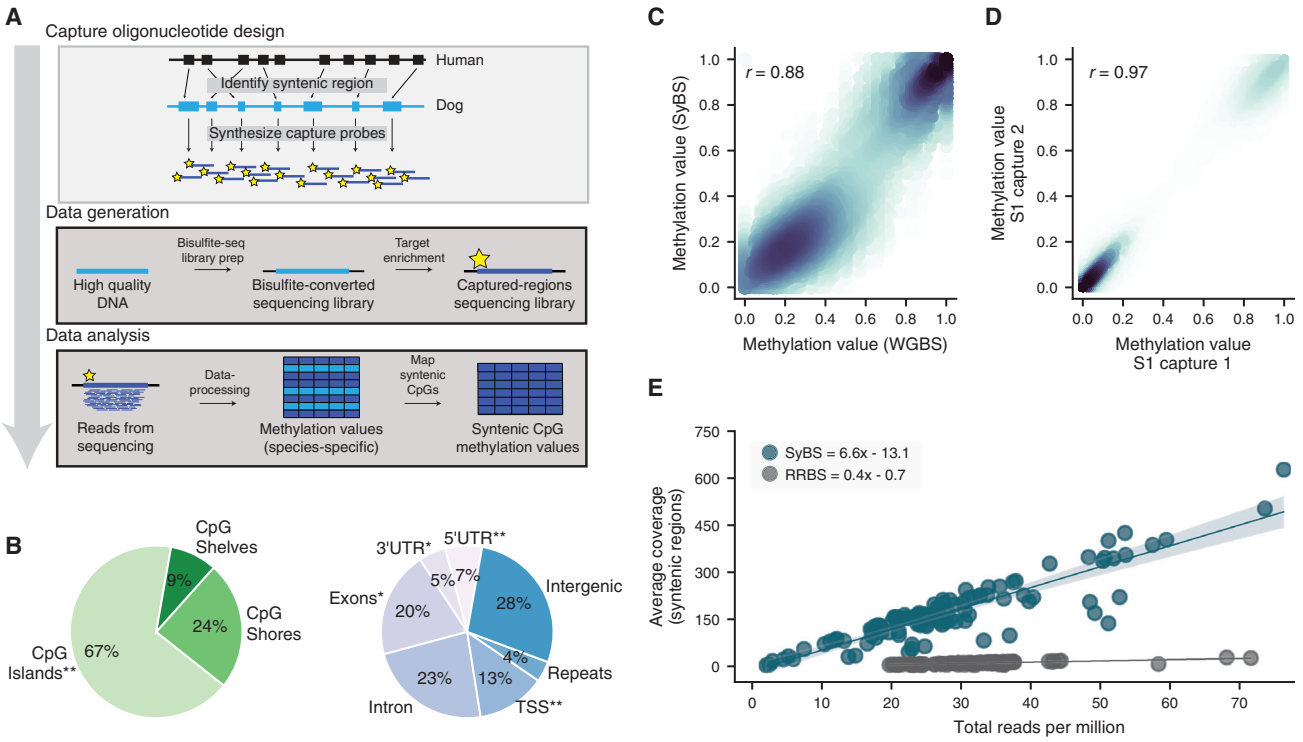


Figure 2. Interrogating Mammalian Methylomes by Syntenic Bisulfite Sequencing (SyBS)

(A) Strategy used to profile and compare CpG methylation states within blocks of synteny in the mammalian genome. Capture oligonucleotide design: regions of DNA (blue blocks) characterized by the Illumina 450K methylation array in humans are mapped to their syntenic region in dogs using whole-genome alignments between the two species. These regions are used to design oligonucleotides (yellow stars) for capture and enrichment of DNA in the second species. Data generation: A sequencing library is constructed from high quality DNA and bisulfite converted, analogously to WGBS. Syntenic sequences are captured, sequenced, and aligned to the mammalian genome under study. CpG methylation values are called and then filtered to select those conserved with humans for further analysis. For more details see STAR Methods.

(B) Pie charts showing representation of targeted genomic regions. Regions exhibiting significant enrichment ($p < 10^{-10}$) are indicated using asterisks with * indicating odds ratio > 2.5 and ** odds ratio > 4. UTR, untranslated region; TSS, transcription start site.

(C) Ten dog methylomes were sequenced twice, either with enrichment for syntenic regions (SyBS hybridization) or without enrichment (WGBS). Methylation values (per CpG site per animal) are shown for SyBS (y axis) versus WGBS (x axis). Sites were considered if they were covered by >5 reads with both SyBS and WGBS.

(D) Concordance of SyBS values for one canine DNA sample (S1), for which two independent captures were performed. In (C) and (D) the color captures the density of observations at each point (darker colors represent higher densities), and the r value is the Pearson correlation.

(E) Average coverage of syntenic segments versus total reads in millions, contrasting SyBS with RRBS.

We applied SyBS to characterize the methylomes of 104 dogs, primarily consisting of Labrador retrievers and representing the entire lifespan, from 0.1 to 16 years at the time of blood draw (Figure S1A; Table S1). Libraries were sequenced to an average depth of 163×, with nine dogs removed due to lack of coverage. Captured CpG sites spanned the genome and were enriched for genomic regions that included exons, transcription start sites, and CpG islands (odds ratio > 2.5 and $p < 10^{-10}$ by Fisher's exact test, Figure 2B). The methylation values associated with captured CpGs were similar to those obtained using WGBS (Pearson correlation r > 0.85, Figure 2C) and showed excellent replication across independent captures from the same samples (r > 0.95, Figures 2D and S1B–S1F). As expected, SyBS achieved substantially higher coverage of syntenic regions compared with non-targeted reduced representation bisulfite methods (~13-fold increase, Figure 2E). For comparison, we obtained previously published methylation profiles from the blood of 320 human individuals aged from 1 to 103 years at the time of sample isolation (Alisch et al., 2012; Hannum et al., 2013). Based on these data, we identified 54,469 well profiled CpGs in both species, thus enabling systematic evolutionary studies of epigenetic changes during life (STAR Methods).

A Concordant but Nonlinear Relationship between Dog and Human Age

We observed the highest methylome similarities (Pearson correlation, STAR Methods) when pairing young dogs with young humans, or aged dogs with older humans. In contrast, the lowest similarities were obtained when pairing young dogs with old humans or vice versa (Figure 3A). The relationship between methylome similarity and age was lost upon permutation (FDR < 0.01; Figure S2A), which indicated that a conserved set of CpG sites are affected during aging in the two mammalian species. Notably, this signal was sufficiently strong to arise in an unsupervised methylome-wide analysis without sub-selection of markers. This result suggested that the conserved methylation

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

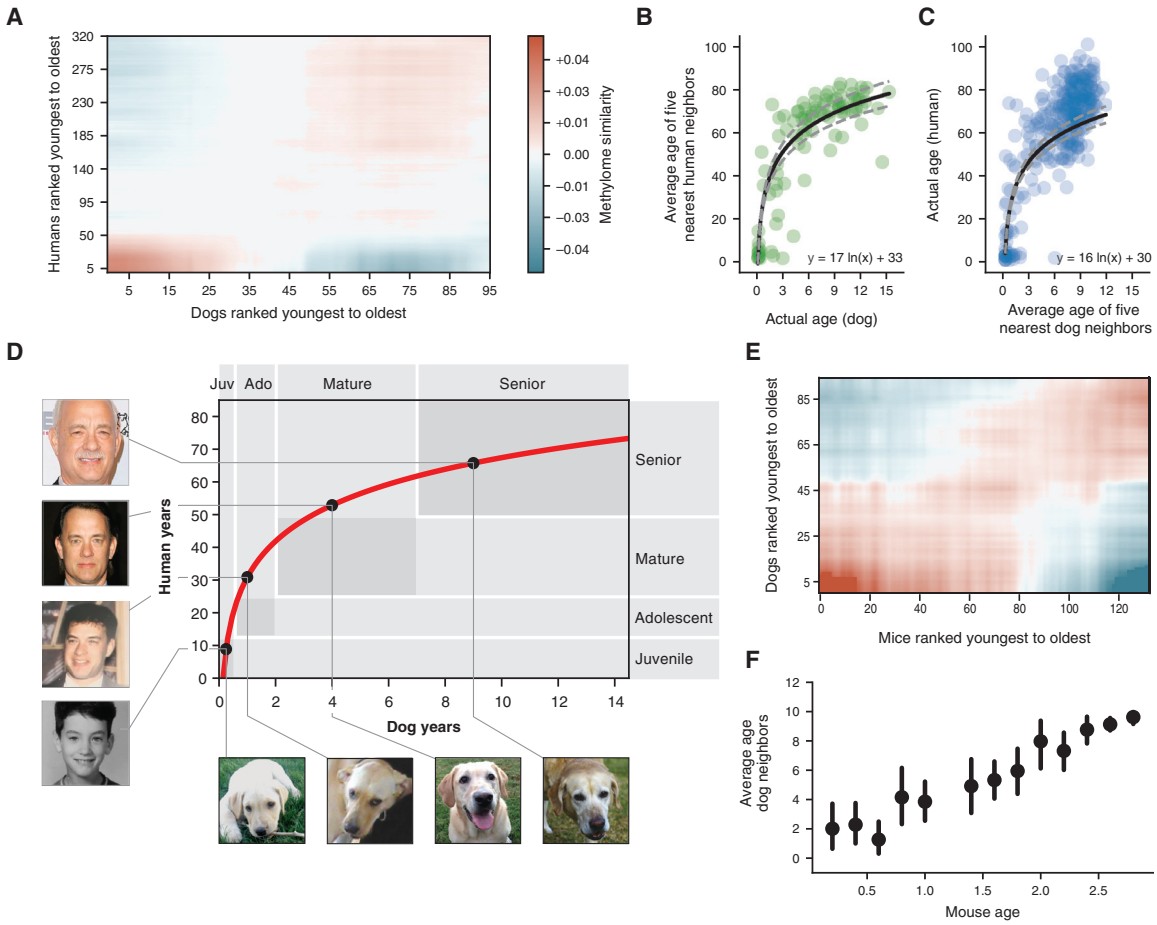


Figure 3. A Nonlinear Transformation from Dog-to-Human Age

(A) Dog-human methylome similarities (Pearson correlation, blue-red color range) are shown with dogs and humans ranked from youngest to oldest. Data are lightly smoothed in both dimensions using Gaussian interpolation in matplotlib.

(B) The age of each dog methylome (x axis) is plotted against the average age of the five nearest human methylomes (y axis), 95 dogs are depicted.

(C) Reciprocal plot in which the age of each human methylome (y axis) is plotted against the average age of the five nearest dog methylomes (x axis), 320 humans are depicted.

(D) Logarithmic function for epigenetic translation from dog age (x axis) to human age (y axis). Outlined boxes indicate the approximate age ranges of major life stages as documented qualitatively based on common aging physiology. Juvenile refers to the period after infancy and before puberty, 2–6 months in dogs, 1–12 years in humans; adolescent refers to the period from puberty to completion of growth, 6 months to 2 years in dogs, approximately 12–25 years in humans; Mature refers to the period from 2–7 years in dogs and 25–50 years in humans; Senior refers to the subsequent period until life expectancy, 12 years in dogs, 70 years in humans. Dog life stages are based on veterinary guides and mortality data for dogs (Fleming et al., 2011; Bartges et al., 2012; Inoue et al., 2015). Human life stages are based on literature summarizing life cycle and lifetime expectancy (Bogin and Smith, 1996; CIA, 2013; Arias et al., 2017). Black dots on the curve connect to images of the same yellow Labrador taken at four different ages (courtesy of Sabrina and Michael Mojica, with permission) and to images of a representative human at the equivalent life stages in human years (photos of Tom Hanks drawn from a public machine-learning image repository, Chen et al., 2015).

(E) Mouse-dog methylome similarities shown as in (A).

(F) Data from (E) are summarized by sorting mice according to 0.2-year bins (x axis) and, for each mouse, plotting the average age of the 5 nearest dogs by methylome similarity (y axis). Points illustrate the mean of each bin and bars represent the 95% confidence interval obtained from bootstrapping.

changes with age in ribosomal RNA genes, noted previously (Wang and Lemos, 2019), extend more generally to the greater mammalian methylome. It contrasts with previous observations using epigenetic clocks, which did not find strong conservation across species, likely because these clocks are restricted to 80–300 CpGs, selected for optimal age prediction in humans independent of other species (Stubbs et al., 2017).

We next investigated whether the conserved methylation changes in dogs and humans follow a constant rate of change with age, with the rate constant depending on the lifespan of each species as suggested previously (Maegawa et al., 2017; Lowe et al., 2018), or whether there was evidence for a more complex trajectory. For this purpose, we assigned the age of each dog to the average age of its nearest humans by methylome-wide similarity (STAR Methods). This analysis revealed a monotonic, time-resolved, nonlinear relationship between dog and human age (Figure 3B considering the k = 5 nearest humans, Figures S2B–S2G considering other k values). Similar results were obtained in a reciprocal analysis assigning each human to its nearest dogs (Figure 3C), as evidenced by the similarity in fitted functions:

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

human_age = 17 *ln(dog_age)* + 33 (Figure 3B) and human_age = 16 *ln(dog_age)* + 30 (Figure 3C). Therefore, we combined the reciprocal analyses to generate the single function: human_age = 16 *ln*(*dog_age*) + 31 (Figure 3D).

We found that this function showed strong agreement between the approximate times at which dogs and humans experience common physiological milestones during both development and lifetime aging, i.e., infant, juvenile, adolescent, mature, and senior (Lebeau, 1953; Bogin and Smith, 1996; Bartges et al., 2012) (Figure 3D). The observed agreement between epigenetics and physiology was particularly close for infant/juvenile and senior stages. For instance, the epigenome translated approximately 8 weeks in dogs (0.15 years) to approximately 9 months in humans (0.78 years), corresponding to the infant stage when deciduous teeth develop in both puppies and babies (Bogin and Smith, 1996; Bartges et al., 2012). In seniors, the expected lifespan of Labrador retrievers, 12 years, correctly translated to the worldwide lifetime expectancy of humans, 70 years (Fleming et al., 2011; CIA, 2013). For adolescent and mature stages, the correspondence was more approximate, with the epigenome showing faster changes for dogs, relative to humans, than expected by physiological tables (Inoue et al., 2015; Arias et al., 2017) (Figure 2D). Thus, the canine epigenome progresses through a series of conserved biological states that align with major physiological changes in humans, occurring in the same sequence but at different chronological timepoints during each species' lifespan.

A conserved nonlinear and weaker epigenetic progression was also observed by comparing the dog methylomes to those measured previously from 133 mice (Petkovich et al., 2017) (Figures 3E and 3F). This weaker effect may be due to the limited number of mice sampled during the developmental period (Table S3). Nevertheless, the ability to translate age among these three diverse mammals indicates that shared physiology may yield conserved molecular transitions in epigenome remodeling with age.

Identification of Genes Exhibiting Conserved Methylation Dynamics with Age

To determine whether the conserved changes were concentrated within particular genes or gene functions, we examined CpG methylation states near 7,942 genes for which orthologs were present in all three species (dogs, humans, and mice, STAR Methods). This analysis identified 394 genes for which methylation values showed conserved time-dependent behavior across species (empirical $p < 0.05$, Figure S3; Table S2). To understand the underlying gene functions, we mapped them onto the parsimonious composite network (PCNet), a database of approximately 2 10^6 molecular interactions capturing physical and functional relationships among genes and gene products, in which each interaction has support from multiple sources (Huang et al., 2018). The genes clustered into five highly interconnected network modules (Figure 4), nearly all of which were enriched for developmental functions. Methylation changes at these developmental genes were concentrated near transcription start sites (Figure S4A) and in CpG islands and shores for all three species (Figure S4B). Four modules predominantly increased in methylation with age (FDR < 0.05) and included modules associated with synapse assembly (18 genes) and neuroepithelial cell differentiation (5 genes) and two modules associated with anatomical patterning (117 and 69 genes). These four modules were enriched for polycomb repressor targets, which are predominantly silenced in adult tissues (Xie et al., 2013). A fifth module was enriched in leukocyte differentiation and nucleic-acid metabolism (144 genes) and demonstrated decreasing methylation with age. We also noted that orthologs from all five modules were among the most highly conserved in DNA sequence in the mammalian genome, even accounting for high sequence conservation of developmental genes in general (Figure S5).

Further indication of the importance of developmental gene modules was observed when calculating dog-human methylome similarity using CpGs at developmental genes only versus a comparison using all CpGs except those at developmental genes (Figure S6; STAR Methods). This comparison showed that CpGs near developmental genes are both necessary and sufficient to recapitulate the cross-species alignments of age observed earlier (Figure 3). Enrichment of age-related methylation increases in developmental genes has been previously observed in humans (Rakyan et al., 2010) and mice (Maegawa et al., 2010). Our findings extend such observations by highlighting developmental genes as predominant drivers of the ability to align mammalian methylomes and in specifying more precisely where in the developmental gene modules such changes occur.

Translating Age and Aging Effects Using a Conserved Epigenetic Clock

Epigenetic clocks have garnered recent interest due to their ability to translate an individual's methylome to an accurate prediction of age (Hannum et al., 2013; Horvath, 2013; Petkovich et al., 2017; Stubbs et al., 2017; Thompson et al., 2017; Wang et al., 2017). However, they have typically been applied to study methylome data in one species only, with less success in cross-species application (Stubbs et al., 2017). Therefore, in our final analysis, we examined whether the conserved developmental gene modules could be used to construct a conserved epigenetic clock capable of predicting age in multiple mammalian species.

In particular, we formulated a conserved development clock based on methylation values of the 394 CpGs within the conserved developmental gene modules we had identified earlier (Figure 4; STAR Methods). By training in dogs or alternatively mice, this clock could be used to score the age of a dog or mouse using its methylation profile. Implicit in this analysis was the translation of a dog methylome to its equivalent mouse age or vice versa (Figure 5A). For baseline comparison, we also constructed single species methylome-wide clocks for dogs and mice as per the usual procedure (STAR Methods). In this case, the model was allowed to select the most optimal CpGs for age prediction from the entire methylome of each species (Figure 5A).

When training and predicting on individuals within a single species, the ages measured by the single species methylome-wide clocks were very accurate (dog *rho* = 0.99, Figures 5B and 5D; mouse *rho* = 0.86, Figures 5C and 5D) relative to the conserved development clocks (dog *rho* = 0.81; mouse *rho* = 0.78; Figure 5D). However, when applying a clock trained in one species to make predictions in the other, the ages predicted by the conserved development clock (dogs-to-mice *rho* = 0.73;

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

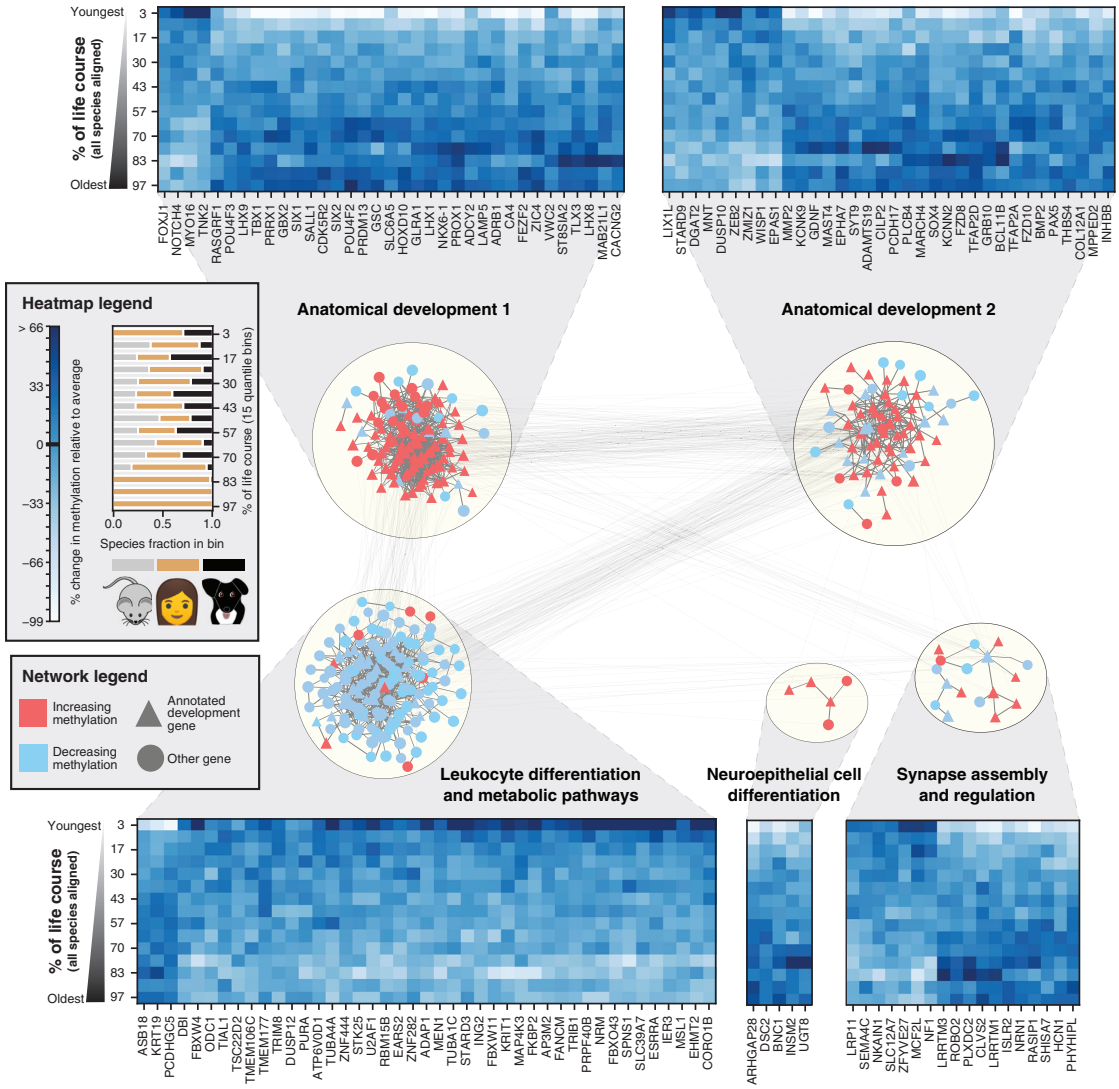


Figure 4. Conserved Lifetime Methylation Changes Aggregate in Developmental Networks

Genes exhibiting conserved age-related methylation behavior were mapped onto a composite molecular interaction network which was subsequently clustered to reveal five major modules, labeled according to enriched Gene Ontology functions (STAR Methods). Colors represent the conserved direction of change with age, with red representing genes that increase in methylation with age and blue representing genes that decrease in methylation with age. Heatmaps show the conserved methylation patterns of a random subset of genes in each module. Columns represent distinct orthologs, while rows represent the average values of all species ranked according to their age in human years and divided into 15 age bins (quantiles). Values are normalized according to the mean and standard deviation of methylation for each ortholog. The fractional species composition of each bin is visualized in the legend.

mice-to-dogs *rho* = 0.71) were substantially more accurate than those of the single species methylome-wide clocks (dogs-to-mice *rho* = 0.22; mice-to-dogs *rho* = 0.32; Figure 5D). In addition, clocks based on the developmental gene modules were more accurate than clocks based on methylation of ribosomal RNA sequences (*rho* = 0.6 dogs-to-mice and 0.5 mice-to-dogs) (Wang and Lemos, 2019). When applying the conserved development clock to mice treated with lifespan-extending interventions (STAR Methods), the epigenetic ages were 30% less, on average, than those of control mice ($p < 10^{-6}$, Figure 5E). These same results were observed when using the development clock trained in dogs to predict mouse age (Figure 5F). Together, these results demonstrate that the methylation states of developmental gene modules can be used to construct a conserved model of age that is transferable between mammalian species. In contrast, previous models trained in a single species can appear to achieve extremely high accuracy when testing in that same species, but these models do not readily transfer outside of the original species context.

DISCUSSION

By using targeted oligonucleotide capture (SyBS), we have produced methylomes of high quality for comparative studies among dogs and other mammals. Analysis of these data shows that multiple mammalian species experience conserved

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

CellPress

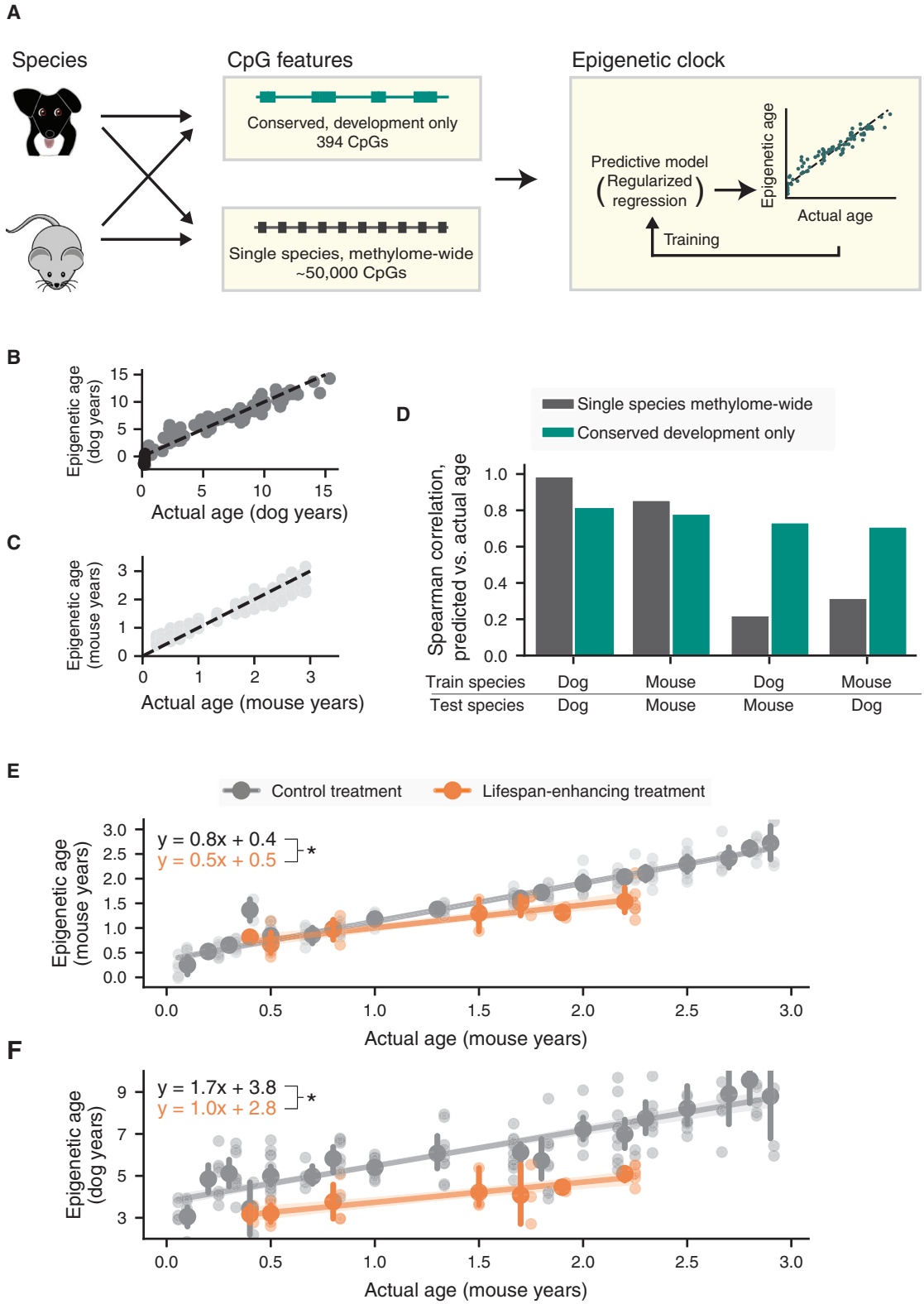


Figure 5. A Conserved Development Clock Measuring Age and Physiological Aging

(A) Construction of epigenetic clocks. Four clocks are constructed, depending on whether the training data are from dogs or mice and whether the input features are from all methylome-wide CpGs or from CpGs in conserved developmental modules only. All four cases yield a regression model for predicting age from CpG markers (STAR Methods).

(legend continued on next page)

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

methylation changes during aging, and that the scope of these changes is methylome-wide. Second, the trajectory of changes followed by one species as it ages is not necessarily the same as that followed by another. In particular, dog methylomes remodel very rapidly in early life compared with the methylomes of their human counterparts. Our further analysis in this regard (Figure 3D) suggests that the rate of remodeling is not only determined by the lifespan of a species but also by the timing of key physiological milestones. Previously, CpG methylation states in humans were proposed to exhibit a nonlinear trajectory over time, with non-constant rates of change (Horvath, 2013). The analysis here demonstrates a different point, that methylation changes in one species can be nonlinear with respect to another.

We observe that epigenetic changes during aging center on highly conserved modules of developmental genes, in which methylation generally increases with age. Although the enrichment of developmental pathways has been generally observed in mammals previously (Ciccarone et al., 2018; Field et al., 2018; Horvath and Raj, 2018), our findings show that specific changes at these loci translate the effects of age and anti-aging interventions across species. While the biology of aging has historically been considered as separate from that of development (Miller and Nadon, 2000; Kowald and Kirkwood, 2016), their strong association, demonstrated here, supports a model in which at least some aspects of aging are a continuation of development rather than a distinct process.

Limitations and extensions of our findings are as follows. First, we have used DNA isolated from whole blood, for which age-dependent shifts in leukocyte populations have been described (Jaffe and Irizarry, 2014). In particular, previous studies have found that $CD4^+$ T cells, $CD8^+$ T cells, and B cells decline with age. Although it is possible that such conserved shifts may influence our findings, such decline occurs in both dogs (Greeley et al., 2001) and humans (Jaffe and Irizarry, 2014). Second, our study has focused exclusively on Labrador retrievers, a popular and heterogeneous breed for which we could collect large numbers of unrelated dogs in order to control for population structure. Distinct breeds exhibiting widely varying lifespans (Gilmore and Greer, 2015; Kaeberlein et al., 2016) could yield different age-translation functions.

Further efforts to characterize epigenetic changes across breeds and species may help to address these and broader questions. For example, does the timing of epigenetic changes early in life influence the overall lifespan of a species, or of an individual within that species? Does modulating the timing of developmental events affect lifespan? Again, comparisons of species or sub-species that experience developmental milestones at similar times but with different lifespans (such as distinct dog breeds) may help address these questions, providing critical and complementary data to inform the ongoing cross-species aging studies (Kaeberlein et al., 2016), including clinical trials of aging interventions (Urfer et al., 2017).

Finally, our study has demonstrated that the methylome can be used to quantitatively translate the age-related physiology experienced by one organism (i.e., a model species like dog) to the age at which physiology in a second organism is most similar (i.e., a second model or humans; Figures 3 and 5). These results enable the methylome to act not only as a diagnostic readout of age in a single species, as per the usual epigenetic clock studies, but also for cross-species translation of age and physiological state of aging. Such translation may provide a compelling tool in the quest to understand aging and identify interventions for maximizing healthy lifespan.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- METHOD DETAILS
  - Annotations
  - Public Datasets
  - Canine Samples
  - SyBS Target Selection
  - SyBS Library Preparation and Sequencing
  - SyBS Data Processing
  - Public RRBS Data Processing
  - Human Methylation Array Data Processing
  - Determining Orthologous CpGs
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Dog-Human Pairwise Methylome Similarity
  - k-nearest Neighbors Analysis
  - Fitting the Epigenetic Age Transfer Function
  - Validation Using Mouse Methylomes
  - Conserved Methylation Changes in Orthologs
  - Network Analysis
  - Developmental Genes Analysis
  - Conserved Development Clock Analysis

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cels.2020.06.006.

(B) Scatterplot of predicted versus actual ages for the dog methylome-wide model.
(C) Scatterplot of predicted versus actual ages for the mouse methylome-wide model.
(D) Performance of single species methylome-wide clocks (gray) or conserved developmental clocks (turquoise) as measured by the Spearman correlation between predicted and actual ages within species or across species.
(E) The conserved development clock distinguishes the effects of lifespan-enhancing treatments (orange) from control treatments (gray). For each treatment, mouse epigenetic ages are measured (y axis; conserved development clock trained in mice) and plotted against actual mouse ages binned in 0.1-year bins (x axis). Mean ± 95% confidence intervals shown for each bin and each observation.
(F) As for (E) but training the conserved development clock using data for dogs. For each treatment (orange lifespan-enhancing; gray control), epigenetic ages of each mouse are measured and plotted against actual mouse ages binned in 0.1-year bins (x axis). * denotes $p < 0.05$ in all panels.

ACKNOWLEDGMENTS

This work is supported by the following: the California Institute for Regenerative Medicine (CIRM Epigenome Center to T.I.), the National Institute for Environmental Health Sciences (T.I., ES014811), the National Institute of General Medical Sciences (A.R.C., GM108865 and T.W., GM008666), the National Institute on Aging (P.D.A., AG031862), the National Institute of Dental and Craniofacial Research (D.L.B., DE022532), the Maxine Adler Endowed Chair Funds (D.L.B.) and the Intramural Program of the National Human Genome Research Institute (A.H. and E.A.O.). We also gratefully acknowledge Sabrina and Michael Mojica for providing images that document the lifespan of their beloved Labrador, Cosmo

AUTHOR CONTRIBUTIONS

T.W. and T.I. initiated and conceptualized the study; T.W. carried out all experiments and implemented the main analyses; A.H., E.A.O., and D.L.B. provided canine samples; J.M., S.F., B.T., J.F.K., D.L.B., and A.R.C. assisted with miscellaneous analyses and provided feedback; P.D.A. provided key input on human and animal aging; T.W., T.I., P.D.A., and E.A.O. interpreted results and wrote the manuscript.

DECLARATION OF INTERESTS

T.I. is co-founder of Data4Cure Inc, is on the Scientific Advisory Board, and has an equity interest. T.I. is on the Scientific Advisory Board of Ideaya Biosciences, has an equity interest, and receives sponsored research funding. The terms of these arrangements have been reviewed and approved by the University of California San Diego in accordance with its conflict of interest policies. T.I. and T.W. hold a patent entitled "Methylome profiling of animals and uses thereof," international patent application #PCT/US18/49103, US application #16/638,454.

Received: November 2, 2019
Revised: March 16, 2020
Accepted: June 9, 2020
Published: July 2, 2020

REFERENCES

Alisch, R.S., Barwick, B.G., Chopra, P., Myrick, L.K., Satten, G.A., Conneely, K.N., and Warren, S.T. (2012). Age-associated DNA methylation in pediatric populations. Genome Res. *22*, 623–632.

Andrews, S., et al. (2010). FastQC: a quality control tool for high throughput sequence data.

Arias, E., Heron, M., and Xu, J. (2017). United States life tables, 2013. Natl Vital Stat. Rep. *66*, 1–64.

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

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat. Genet. *25*, 25–29.

Bartges, J., Boynton, B., Vogt, A.H., Krauter, E., Lambrecht, K., Svec, R., and Thompson, S. (2012). AAHA canine life stage guidelines. J. Am. Anim. Hosp. Assoc. *48*, 1–11.

Bell, G.I., Karam, J.H., and Rutter, W.J. (1981). Polymorphic DNA region adjacent to the 5′ end of the human insulin gene. Proc. Natl. Acad. Sci. USA *78*, 5759–5763.

Bogin, B., and Smith, B.H. (1996). Evolution of the human life cycle. Am. J. Hum. Biol. *8*, 703–716.

Bolstad, B.M. (2013). 'preprocessCore: a collection of pre-processing functions', R package version, 1. https://github.com/bmbolstad/preprocessCore.

Chatterjee, A., Rodger, E.J., Morison, I.M., Eccles, M.R., and Stockwell, P.A. (2017). Tools and strategies for analysis of genome-wide and gene-specific DNA methylation patterns. Methods Mol. Biol. *1537*, 249–277.

Chen, B., Chen, C., and Hsu, W.H. (2015). Face recognition and retrieval using cross-age reference coding With cross-age celebrity dataset. IEEE Trans. Multimedia *17*, 804–815.

CIA (2013). The World Factbook 2013–14 (Central Intelligence Agency).

Ciccarone, F., Tagliatesta, S., Caiafa, P., and Zampieri, M. (2018). DNA methylation dynamics in aging: how far are we from understanding the mechanisms? Mech. Ageing Dev. *174*, 3–17.

Dreger, D.L., Rimbault, M., Davis, B.W., Bhatnagar, A., Parker, H.G., and Ostrander, E.A. (2016). Whole-genome sequence, SNP chips and pedigree structure: building demographic profiles in domestic dog breeds to optimize genetic-trait mapping. Dis. Models Mech. *9*, 1445–1460.

Field, A.E., Robertson, N.A., Wang, T., Havas, A., Ideker, T., and Adams, P.D. (2018). DNA methylation clocks in aging: categories, causes, and consequences. Mol. Cell *71*, 882–895.

Fleming, J.M., Creevy, K.E., and Promislow, D.E.L. (2011). Mortality in North American dogs from 1984 to 2004: an investigation into age-, size-, and breed-related causes of death. J. Vet. Intern. Med. *25*, 187–198.

Gilmore, K.M., and Greer, K.A. (2015). Why is the dog an ideal model for aging research? Exp. Gerontol. *71*, 14–20.

Greeley, E.H., Ballam, J.M., Harrison, J.M., Kealy, R.D., Lawler, D.F., and Segre, M. (2001). The influence of age and gender on the immune system: a longitudinal study in Labrador Retriever dogs. Vet. Immunol. Immunopathol. *82*, 57–71.

Gross, A.M., Jaeger, P.A., Kreisberg, J.F., Licon, K., Jepsen, K.L., Khosroheidari, M., Morsey, B.M., Swindells, S., Shen, H., Ng, C.T., et al. (2016). Methylome-wide analysis of chronic HIV infection reveals five-year increase in biological age and epigenetic targeting of HLA. Mol. Cell *62*, 157–168.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.B., Gao, Y., et al. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol. Cell *49*, 359–367.

Hastie, T., Tibshirani, R., Narasimhan, B., and Chu, G. (2020). impute: impute: Imputation for microarray data. R package version 1.62.0. https://www.bioconductor.org/packages/release/bioc/html/impute.html.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biol. *14*, R115.

Horvath, S., and Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. *19*, 371–384.

Huang, J.K., Carlin, D.E., Yu, M.K., Zhang, W., Kreisberg, J.F., Tamayo, P., and Ideker, T. (2018). Systematic evaluation of molecular networks for discovery of disease genes. Cell Syst. *6*, 484–495.e5.

Inoue, M., Hasegawa, A., Hosoi, Y., and Sugiura, K. (2015). A current life table and causes of death for insured dogs in Japan. Prev. Vet. Med. *120*, 210–218.

Jaffe, A.E., and Irizarry, R.A. (2014). Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. *15*, R31.

Kaeberlein, M., Creevy, K.E., and Promislow, D.E.L. (2016). The dog aging project: translational geroscience in companion animals. Mamm. Genome *27*, 279–288.

Khan, S.S., Singer, B.D., and Vaughan, D.E. (2017). Molecular and physiological manifestations and measurement of aging in humans. Aging Cell *16*, 624–633.

Kowald, A., and Kirkwood, T.B.L. (2016). Can aging be programmed? A critical literature review. Aging Cell *15*, 986–998.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al.; SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nat. Methods *17*, 261–272.

Krueger, F. (2015). https://github.com/FelixKrueger/TrimGalore.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lebeau, A. (1953). L'âge du chien et celui de l'homme Essai de statistique sur la mortalité canine. Bul. de l'Ac. Vét. de France *26*, 229–232.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lowe, R., Barton, C., Jenkins, C.A., Ernst, C., Forman, O., Fernandez-Twinn, D.S., Bock, C., Rossiter, S.J., Faulkes, C.G., Ozanne, S.E., et al. (2018). Ageing-associated DNA methylation dynamics are a molecular readout of lifespan variation among mammalian species. Genome Biol. *19*, 22.

Madar, A., Greenfield, A., Vanden-Eijnden, E., and Bonneau, R. (2010). DREAM3: network inference using dynamic context likelihood of relatedness and the inferelator. PLoS One *5*, e9803.

Maegawa, S., Hinkal, G., Kim, H.S., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L.A., and Issa, J.P. (2010). Widespread and tissue specific age-related DNA methylation changes in mice. Genome Res. *20*, 332–340.

Maegawa, S., Lu, Y., Tahara, T., Lee, J.T., Madzo, J., Liang, S., Jelinek, J., Colman, R.J., and Issa, J.J. (2017). Caloric restriction delays age-related methylation drift. Nat. Commun. *8*, 539.

Merico, D., Isserlin, R., and Bader, G.D. (2011). Visualizing gene-set enrichment results using the cytoscape plug-in enrichment map. Methods Mol. Biol. *781*, 257–277.

Miller, R.A., and Nadon, N.L. (2000). Principles of animal use for gerontological research. J. Gerontol. A Biol. Sci. Med. Sci. *55*, B117–B123.

Ostrander, E.A., Wayne, R.K., Freedman, A.H., and Davis, B.W. (2017). Demographic history, selection and functional diversity of the canine genome. Nat. Rev. Genet. *18*, 705–720.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J.T., et al. (2011). Scikit-learn: machine learning in Python. Journal of machine learning research *12*, 2825–2830.

Petkovich, D.A., Podolskiy, D.I., Lobanov, A.V., Lee, S.G., Miller, R.A., and Gladyshev, V.N. (2017). Using DNA methylation profiling to evaluate biological age and longevity interventions. Cell Metab. *25*, 954–960.e6.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

R Core Team (2018). R: a language and environment for statistical computing (R Foundation for Statistical Computing). https://www.R-project.org.

Rakyan, V.K., Down, T.A., Maslau, S., Andrew, T., Yang, T.P., Beyan, H., Whittaker, P., McCann, O.T., Finer, S., Valdes, A.M., et al. (2010). Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res. *20*, 434–439.

Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. *43*, D670–D681.

Ryan, D. (2017). MethylDackel: a (mostly) universal methylation extractor for BS-seq experiments, 2017. https://github.com/dpryan79/MethylDackel.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Sheffield, N.C., and Bock, C. (2016). LOLA: enrichment analysis for genomic region sets and regulatory elements in R and bioconductor. Bioinformatics *32*, 587–589.

Siepel, A., Pollard, K.S., and Haussler, D. (2006). New methods for detecting lineage-specific selection. In Lecture Notes in Computer Science (Springer), pp. 190–205.

Stubbs, T.M., Bonder, M.J., Stark, A.K., Krueger, F., BI Ageing Clock Team, von Meyenn, F., Stegle, O., and Reik, W. (2017). Multi-tissue DNA methylation age predictor in mouse. Genome Biol. *18*, 68.

Teschendorff, A.E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., and Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics *29*, 189–196.

Thompson, M.J., vonHoldt, B., Horvath, S., and Pellegrini, M. (2017). An epigenetic aging clock for dogs and wolves. Aging *9*, 1055–1068.

Picard Toolkit (2019). Broad Institute, GitHub Repository. http://broadinstitute.github.io/picard/.

Urfer, S.R., Kaeberlein, T.L., Mailheau, S., Bergman, P.J., Creevy, K.E., Promislow, D.E.L., and Kaeberlein, M. (2017). A randomized controlled trial to establish effects of short-term rapamycin treatment in 24 middle-aged companion dogs. GeroScience *39*, 117–127.

Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. *19*, 327–335.

Vonholdt, B.M., Pollinger, J.P., Lohmueller, K.E., Han, E., Parker, H.G., Quignon, P., Degenhardt, J.D., Boyko, A.R., Earl, D.A., Auton, A., et al. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature *464*, 898–902.

Wang, M., and Lemos, B. (2019). Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. Genome Res. *29*, 325–333.

Wang, T., Tsui, B., Kreisberg, J.F., Robertson, N.A., Gross, A.M., Yu, M.K., Carter, H., Brown-Borg, H.M., Adams, P.D., and Ideker, T. (2017). Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. Genome Biol. *18*, 57.

Withers, P.C. (1992). Comparative Animal Physiology (Cole Publishing Company). https://books.google.com/books/about/Comparative_Animal_Physiology.html?hl=&id=Ba_wAAAAMAAJ.

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell *153*, 1134–1148.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. Nucleic Acids Res. *44*, D710–D716.

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Labrador retrievers whole blood samples	UC Davis	N/A
Labrador retrievers whole blood samples	NHGRI	N/A
Critical Commercial Assays		
Kapa HTP Library Prep Illumina	Roche	KK8234
SeqCap Adapter kit A	Roche	7141530001
SeqCap Adapter kit B	Roche	7141548001
SeqCap Epi Developer S	Roche	7139071001
SeqCap Epi Developer Reagent	Roche	6684335001
SeqCap EZ HE-Oligo Kit A	Roche	6777287001
SeqCap EZ HE-Oligo Kit B	Roche	6777317001
VK SeqCap Epi Reagent Kit plus	Roche	7185723001
Seqcap EZ Pure Capture Bead Kit	Roche	6977952001
Zymo EZ DNA methylation	Zymo	11-334
Kapa Library qPCR quantification	Roche	KK4824
Bioanalyzer (DNA 1000)	Agilent	5067-1504
Bionalyzer machine	Agilent	2938C
Kapa HiFi HotStart Uracil Polymerase	Roche	KK2801
Covaris microTUBES	Covaris	520166
Ampure beads	Beckman Coulter	A63881
HiSeq 4000 Reagent kit	Illumina	PE-410-1001
Deposited Data		
Raw and analyzed data	This paper	GEO: GSE146920
Code and associated files	This paper	https://doi.org/10.5281/zenodo.3864683
Dog genome (CanFam 3.1)	Yates et al., 2016	http://dec2017.archive.ensembl.org/
Mouse genome (mm10)	Yates et al., 2016	http://dec2017.archive.ensembl.org/
Human genome (hg19)	Yates et al., 2016	http://dec2017.archive.ensembl.org/
CpG Islands & repeat regions	Rosenbloom et al., 2015	https://hgdownload.soe.ucsc.edu/downloads.html
Gene & ortholog annotations	Vilella et al., 2009	http://dec2017.archive.ensembl.org/
Mouse RRBS data	Petkovich et al., 2017	GEO: GSE80672
Human (children) 450K array	Alisch et al., 2012	GEO: GSE36054
Human (adults) 450K array.	Hannum et al., 2013	GEO: GSE40279
Dog and wolf RRBS	Thompson et al., 2017	SRA: SRP065319
Gene Ontology	Ashburner et al., 2000	http://geneontology.org/docs/download-ontology/
Software and Algorithms		
LiftOver	Rosenbloom et al., 2015	http://hgdownload.soe.ucsc.edu/downloads.html#utilities_downloads
FASTQC	N/A	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
TrimGalore (v0.4.5)	N/A	http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Bismark (v0.14.3)	Krueger and Andrews, 2011	http://www.bioinformatics.babraham.ac.uk/projects/bismark/

(*Continued on next page*)

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MethylDackel	N/A	https://github.com/dpryan79/MethylDackel
Picard Tools (v1.141)	N/A	https://broadinstitute.github.io/picard/
SAMtools	Li et al., 2009	http://www.htslib.org/
BEDtools (v2.25.0)	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/
fancyimpute	N/A	https://pypi.org/project/fancyimpute/
Minfi (R)	Aryee et al., 2014	https://bioconductor.org/packages/release/bioc/html/minfi.html
Impute (R)	Hastie et al., 2020	https://www.bioconductor.org/packages/release/bioc/html/impute.html
Modified beta-mixture quantile dilation (BMIQ, R)	Horvath, 2013; Teschendorff et al., 2013	PMID: 24138928; Additional File 24
LOLA (R)	Sheffield and Bock, 2016	https://bioconductor.org/packages/release/bioc/html/LOLA.html
preprocessCore (R)	Bolstad, 2013	http://bioconductor.org/packages/release/bioc/html/preprocessCore.html
statsmodel (v0.8.0 Python)	Perktold et al.	http://www.statsmodels.org/stable/index.html
SciPy (v1.1.1 Python)	Virtanen et al., 2020	https://www.scipy.org/
scikit-learn (v0.19.2 Python)	Pedregosa et al., 2011	https://scikit-learn.org/stable/
Cytoscape (v3.7)	Shannon et al., 2003	https://cytoscape.org/download.html
clusterMaker2	Shannon et al., 2003	http://apps.cytoscape.org/apps/clustermaker2
Enrichment Map	Merico et al., 2011	http://apps.cytoscape.org/apps/enrichmentmap

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Trey Ideker (tideker@ucsd.edu)

Materials Availability

This study did not generate new materials.

Data and Code Availability

The sequencing reads and processed files generated during this study are available at GEO (GSE146920). All original code is freely available for download at 10.5281/zenodo.3864683.

METHOD DETAILS

Annotations

Reference genomes were downloaded from Ensembl for dog (CanFam3.1), mouse (mm10) and human (hg19). Ensembl Biomart version 91 was used for gene, 3′UTR and 5′UTR annotations (Yates et al., 2016). CpG islands, repeat annotations, and chain files were downloaded from the UCSC Genome Browser (Rosenbloom et al., 2015). CpG shores were designated as regions 2 kilobases (kb) outside each CpG island, and CpG shelves were designated as regions 2kb outside of CpG shores. Promoters were designated as regions 2kb upstream and 100 basepairs (bp) downstream of the transcription start sites (TSS) based on gene annotations from Ensembl (Yates et al., 2016). Whole genes were divided into exonic and intronic sequences. Intergenic regions were then defined as the remaining regions of the genome after subtracting all other annotated regions. Definitions of one-to-one orthologs were downloaded from Ensembl Compara (Vilella et al., 2009) for dogs, humans and mice.

Public Datasets

The following datasets were obtained from Gene Expression Omnibus (GEO) or Sequence Read Archives (SRA) [number of individuals included in study in brackets]:

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

- GSE80672 (Petkovich et al., 2017): Methylomes from postnatal mice. Blood, Reduced Representation Bisulfite Sequencing (RRBS) method. [133]
- GSE36054 (Alisch et al., 2012): Methylomes from human children. Blood, Infinium 450K array. [35]
- GSE40279 (Hannum et al., 2013): Methylomes from human adults. Blood, Infinium 450K array. [285]
- SRP065319 (Thompson et al., 2017): Methylomes from dogs and wolves. Blood, RRBS method [92].

Canine Samples

Information on each dog sample used, including age, breed, and source, is given in Table S1, with the age distribution also provided in Figure S1A. For samples sourced from NHGRI, domestic dogs were collected with owners' signed consents in accordance with standard protocols approved by the NHGRI IACUC committee. Samples were collected at canine-centric events such as dog shows. Alternatively, owners were supplied with a mail-in kit which included instructions, tubes for blood draws and a general information sheet requesting the AKC number (when available), pedigree and date of birth. Blood draws were performed by licensed veterinarians or veterinary technicians. For samples sourced from UC Davis, blood was collected from privately owned dogs through the William R. Pritchard Veterinary Medical Teaching Hospital. Owners specified the breed of each dog. Standard collection protocols were reviewed and approved by the UC Davis IACUC. DNA was extracted either using the Puregene kit (Qiagen) or using the cell lysis protocol described by (Bell et al., 1981), followed by phenol/chloroform extraction with phase separation in 15-mL phase-lock tubes (5-Prime, Inc. Gaithersburg, MD, USA).

SyBS Target Selection

The strategy for syntenic bisulfite sequencing was to base our Illumina Human 450K probe locations were extended 50bp with respect to the strand of each probe. The resulting locations were mapped to the dog genome using liftOver (Rosenbloom et al., 2015) using default parameters. After excluding regions that mapped to sex, mitochondrial and unplaced contigs in the dog genome, we identified approximately 230,000 probes that were syntenic between human and dogs. Hybridization probes were generated to target these regions using the Roche SeqCap-Epi platform. This process produced an 18.8 megabase sequencing library in dogs, containing approximately 90,000 CpGs that were also profiled by the Illumina 450K array in humans.

SyBS Library Preparation and Sequencing

We followed the protocol specified by the Roche SeqCap-Epi platform. Briefly, approximately 500ng of lambda phage DNA (bisulfite-conversion control) was added to 1ug of dog DNA, then sheared to an average of 175bp (Covaris). Sheared DNA was end-repaired, A-tailed and ligated to barcoded adapters. Adapter-ligated libraries were subjected to bisulfite treatment (Zymo EZ DNA methylation lightning kit) following manufacturer's instructions. Bisulfite-treated libraries were cleaned and amplified using 25 cycles of PCR with a uracil-tolerant enzyme (Kapa). Libraries were pooled equimolarly into 4-plex or 6-plex hybridization capture reactions to a total of 1ug per reaction. Captured product was PCR amplified (10 cycles). Hybridizations were pooled before sequencing and split among 10 lanes on an Illumina HiSeq 4000 in 2x100bp cycles.

SyBS Data Processing

Reads obtained from sequencing were demultiplexed and their quality was verified using FastQC (Andrews et al., 2010). Reads were trimmed using TrimGalore (Krueger, 2015) (4bp) then aligned to a bisulfite-converted dog genome (CanFam3.1) using Bismark (v0.14.3) (Krueger and Andrews, 2011), which produced alignments with Bowtie2 (v2.1.0) (Langmead et al., 2009) with parameters "-score_min L,0,-0.2". Methylation values for CpG sites were determined using MethylDackel (v0.2.1) (Ryan, 2017). Custom Python scripts using BEDtools (v2.25.0) (Quinlan and Hall, 2010) were used to determine on-target reads. Optical PCR duplicates were determined using Picard tools (v1.141) (Picard Toolkit, 2019) and removed using Samtools (v0.1.18) (Li et al., 2009). Coverage of syntenic regions was determined using the number of unique on-target reads that were orthologous to humans, divided by the expected sequencing space. Only CpG sites that were on-target, covered by at least five reads and present across 90% of samples were selected for further analysis. Samples missing more than 30% of CpGs were removed from further analyses resulting in the removal of nine dogs. Missing data for selected CpGs were imputed by performing k-nearest neighbors (k = 10) using fancyimpute in Python. To assess the concordance of methylation values obtained using SyBS with conventional approaches, we also sequenced 10 of the same dogs using whole-genome bisulfite sequencing (libraries prior to enrichment with SyBS probes). Reads were processed and aligned with the canine genome as described above. We saw an average Pearson correlation of r = 0.85 among these 10 samples (range 0.75 - 0.97) (Figure 2C). We also performed independent replicate hybridizations for 6 samples. We saw an average $r = 0.97$ (range: 0.96 - 0.98) for these technical replicates (Figures 2D and S1B–S1F). We verified that lambda phage DNA exhibited complete conversion (>99.5%). We tested the significance of the enrichment of our captured sequences and genome region annotation using the LOLA package (Sheffield and Bock, 2016) in R (version 3.5.1) (R Core Team, 2018). Enrichment tests are performed using Fisher's exact tests, with the possible 'universe' defined by restriction digestion fragmentation of autosomes in the canine reference genome.

Public RRBS Data Processing

For data generated using Reduced Representation Bisulfite Sequencing (RRBS), methods for alignment and CpG selection were identical to those described above. Since RRBS fragments are generated using restriction enzymes with specific recognition sites,

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

optical PCR duplicates could not be removed and on-target CpGs were not determined. For evolutionary comparative analysis, we included 133 control mice aged between 3 months to 2.5 years (Petkovich et al., 2017). To compare the coverage of syntenic regions between SyBS and non-targeted bisulfite technology, we used a RRBS study in dogs and wolves (Thompson et al., 2017) (Figure 2E).

Human Methylation Array Data Processing

Illumina Infinium 450K methylome array data were quantile normalized using Minfi (Aryee et al., 2014) and missing values were imputed using the Impute package (Hastie et al., 2020) in R. These values were adjusted for cell counts as previously described (Gross et al., 2016). To enable comparisons across different methylation array studies, we implemented beta-mixture quantile dilation (BMIQ) (Horvath, 2013; Teschendorff et al., 2013) and used the median of the (Hannum et al., 2013) dataset as the gold standard. To mitigate residual batch effects, we selected human samples that clustered closely in the first two principal components using scikit-learn v0.19.2 (Pedregosa et al., 2011) and verified that such filtering had little effect on the distribution of ages. We also removed samples for which more than 10% of probes were not adequately detected. This procedure resulted in methylome profiles for 320 humans that could be compared to the SyBS-generated dog methylomes.

Determining Orthologous CpGs

Human Illumina 450K methylation array CpGs were extended by 50bp with respect to the strand using BEDtools and mapped to the target genome (mouse or dog) using liftOver with "-minMatch=0.5". We verified that the coordinate alignment obtained using 50bp was identical to that obtained using the exact coordinate (1bp) at "-minMatch=0.95". This procedure allowed us to determine an exact orthologous region for each human CpG and each dog CpG. When multiple dog CpGs were assigned to one human CpG probe region, we took the average methylation value of the aligned CpGs in dogs. This procedure resulted in 54,469 dog-human orthologous CpGs for further analysis. To mitigate batch effects specific to sequencing and/or array platforms, we normalized the sequencing methylation values using BMIQ and performed quantile normalization using the preprocessCore package in R (normalize.quantiles.use.target function) (Bolstad, 2013).

For dog-to-mouse comparisons, CpGs that were separated by 1bp were merged into one region using BEDtools. Each region was then extended by 50bp. The resulting region files were aligned to the target genome using liftOver "-minMatch=0.5". Only regions that were concordant between the two alignments (i.e., dog to mouse or mouse to dog) were selected for further analysis. CpGs that were assigned to the same aligned regions were averaged to generate 9,404 bins, consisting of 87,915 CpGs from dogs.

QUANTIFICATION AND STATISTICAL ANALYSIS

Dog-Human Pairwise Methylome Similarity

Methylation values of orthologous CpGs were normalized by subtracting the mean and dividing by the standard deviation over individuals (i.e., z-transformed, separately for each species). The resulting z-values represent the tendency to decrease or increase relative to the mean of each CpG within a species. Using these values we calculated the pairwise Pearson correlation between the methylomes of each dog-human pair. Correlation was computed across all orthologous CpG values using the SciPy Python package (Virtanen et al., 2020), forming a 95 x 320 (dog x human) methylome similarity matrix (Figure 3). We also created a coarsened version of this matrix, in which the pairwise similarities were averaged over two-year age windows in both species, forming an 8 x 51 (dog x human) methylome similarity matrix (*MSA*, Figure S2A).

Given this matrix, we evaluated the significance of association between age and methylome similarity using permutations. Specifically, we generated the following two-by-two contingency table:

	Ages More Different $AD(i,j)>\underline{AD}$	Ages More Similar $AD(i,j)\leq\underline{AD}$
Methylomes more different $MSA(i,j)\leq\underline{MS}$	$Count_1$	$Count_3$
Methylomes more similar $MSA(i,j)>\underline{MS}$	$Count_2$	$Count_4$

where *MSA* is the methylome similarity matrix, $AD(i,j)$ is the age difference computed as $\left| Age\ bin_{dog} - Age\ bin_{human} \right|$ and *Count* is the number of occurrences (cells within the *MS* similarity matrix) for which the table row and column conditions are met. Using these counts, we calculated the p-value using the one tailed Fisher's exact test and compared this p-value to that obtained when permuting the membership of dogs and humans in two-year age bins across 1000 permutations (Figure S2A).

k-nearest Neighbors Analysis

To achieve a robust assignment of reciprocal nearest neighbors, we used a strategy inspired by Context Likelihood of Relatedness (Madar et al., 2010). Specifically, we z-normalized the *MS* methylome similarity matrix to form *MSZ*, as follows:

$$MSZ_{row}(i,\ j) = max\left(0,\ \frac{MS_{ij} - \underline{MS_{i*}}}{\sigma_{i*}}\right)$$

$$MSZ_{column}(i,j) = max\left(0, \frac{MS_{i,j} - \overline{MS_{*j}}}{\sigma_{*j}}\right)$$

$$MSZ(i,j) = mean\ [MSZ_{row}(i,j),\ MSZ_{column}(i,j)]$$

k-nearest neighbors were assigned to each dog or to each human with respect to *MSZ* values. This process was implemented in Python using scikit-learn.

Fitting the Epigenetic Age Transfer Function

The nearest neighbor analysis was fit using non-linear regression with the SciPy package in Python. The model fit was specified using the following formula:

$$Dog\ Age = A * ln(Human\ age) + B$$

Here, "Dog age" was represented by the chronological ages of dogs, and "Human age" was the average age of the nearest human neighbors with respect to methylome similarity. The converse was performed as well, i.e., dog age was represented as the average age of the nearest dog neighbors and human age was the chronological age in humans. For the final age transfer function, the coefficients (*A*,*B*) were estimated by bootstrapping an equal number of both dogs and humans. The standard error was estimated using 1000 bootstraps.

Validation Using Mouse Methylomes

Dog-mouse methylome similarity was calculated identically as for dog-human comparisons. A *k*-nearest neighbors analysis (as described for dogs and humans above) was repeated using the orthologous CpGs for pairwise comparisons involving mice. The mouse methylome data had a highly canalized age distribution which was different from that of the dogs or humans in our study. That is, mice had been sampled at discrete ages, we therefore visualized these data according to 0.2 year bins (Figure 3F).

Conserved Methylation Changes in Orthologs

We considered 14,652 one-to-one orthologs in dogs, humans and mice that were within 2.5kb of orthologous CpGs. Among these, we identified 7,934 orthologous genes for which methylation values were available. Methylation values were then logit-transformed; multiple CpGs assigned to one gene were represented by the average methylation value. We assigned to each ortholog a 'methylation conservation score' using the following procedure. First, the age of each dog or mouse individual was translated to the equivalent human age using the epigenetic age translation functions built using the *k*-nearest neighbors analysis. We ranked all individuals according to their age in human years and divided this ranking into 15 quantile bins. Logit-transformed methylation values were averaged within each bin and species. For each gene and species we calculated the Spearman correlation between the gene's methylation values and age. Genes were then ranked by $sign(correlation) * -log_{10}(correlation_{p\ value})$ within each of the three species. We computed the Euclidean norm of the three ranks and sought genes with very low norms (for which methylation was consistently among the most increasing with age across species) or with very high norms (for which methylation was consistently among the most decreasing with age across species). Significance was determined using a two-sided empirical p-value < 0.05, yielding 394 genes.

We examined methylation within gene bodies by calculating the distance of each CpG relative to the transcription start site and normalized these distances by the gene length. We then grouped CpGs into 10 bins and calculated the average methylation using a rolling window (window = 3) among genes grouped according to their conservation status: conserved increasing or decreasing methylation with age and not conserved. We then calculated the difference between the oldest 20% and youngest 20% for each species. We repeated this analysis using all one-to-one orthologs or grouped according to their developmental gene status (Figure S4A). To test enrichment of genomic regions for conserved developmental genes, for each developmental gene, we annotated its underlying CpGs according to their genomic region (TSS, intergenic, 5′UTR, repeats, exonic, 3′UTR, intronic, CpG shelves, CpG shores, CpG islands). We then tested for enrichment using the Fisher's exact test between conserved and not conserved developmental genes using a p-value cutoff of < 0.005 for each species (Figure S4B).

Network Analysis

We downloaded the PCNet parsimonious composite human functional interaction network from (Huang et al., 2018) and subselected gene orthologs with significantly conserved methylation trajectories (see above) resulting in a subnetwork with 355 nodes and 2003 edges. We visualized the network using Cytoscape (Shannon et al., 2003) (version 3.7) and performed community detection using clusterMaker2 (Shannon et al., 2003). To annotate modules, we performed functional enrichment using a hypergeometric test for each term within the Biological Process branch of the human Gene Ontology (GO) (Ashburner et al., 2000) and adjusted for false-discovery rate using a very strict Benjamini-Hochberg procedure (FDR < 0.001) implemented using statsmodel in python. Significant GO terms were clustered according to gene-set similarity using Enrichment Map (Merico et al., 2011), and gene modules were clustered according to their Jaccard overlap, revealing high-level functional categories (Figure 4).

Please cite this article in press as: Wang et al., Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome, Cell Systems (2020), https://doi.org/10.1016/j.cels.2020.06.006

Developmental Genes Analysis

Genes were ranked according to their methylation conservation score (see above) and subdivided into 25 evenly spaced bins, separating genes with significantly conserved decreases or increases in methylation for a total of 27 bins. We then obtained PhyloP (Siepel et al., 2006) sequence conservation scores according to the orthologous CpGs assigned to each gene. Finally, we averaged the PhyloP scores in each methylation conservation score bin, estimating the 95% confidence interval by bootstrapping (Figure S5). We assessed the significance of the interaction between methylation conservation score and developmental gene status using ANOVA.

We restricted to orthologous CpGs profiled across dogs, humans and mice (6,906 CpGs) that were within 2.5kb of the gene bodies of all orthologous genes ('all CpGs'). From this set, we identified CpGs near development genes ('devCpGs'); we also controlled for the number of CpGs with 100 randomly-sampled subsets of CpGs that were equal in size from those not near developmental genes ('not devCpGs'). We calculated the methylome similarity (as described above) based on these CpG subsets for pairwise comparisons of species (dog and human, dog and mouse). For each pairwise comparison (Species 1, Species 2), we identified the 5-nearest neighbors in Species 2 for each individual of Species 1, then binned the actual age of Species 1 into five discrete bins and calculated the average neighbor age for each bin with the 95% confidence interval estimated by bootstrapping (Figure S6).

Conserved Development Clock Analysis

Dog and mouse epigenetic clocks were built with Elastic net (scikit-learn in Python) using either methylome-wide CpG values (~90000 CpGs across both species) or 394 CpG values associated with developmental gene modules (Figure 5). We refer to the ages predicted from this model as ''epigenetic ages''. Hyperparameters were tuned using five-fold cross validation in the dog data. Since some other clocks in the literature use ten-fold cross validation, we also tested the parameters selected using a ten-fold cross-validation procedure. We found an increase in median absolute error (MAE) in this case, thus we proceeded with five-fold. Performance of the final model was assessed by Spearman correlation of actual versus epigenetic age (output of the Elastic net model) for 11 dogs which had not been used for training, and for the control mice described above. As controls in Figures 5E and 5F, we built 100 control clocks using 100 randomly sampled sets of 394 CpGs that had been profiled in dogs and mouse but were not in developmental gene modules. For analysis involving lifespan-enhancing intervention mice, we obtained DNA methylation data profiled from whole blood from (Petkovich et al., 2017), processed as described above. We removed GHRKO from further analysis, as principal component analysis using the 394 conserved CpGs revealed clustering due to treatment. All remaining mice used in this analysis are described in Table S3. We applied the epigenetic clock, trained in mice or trained in dogs, and evaluated the effect of longevity-enhancing interventions using a log-likelihood ratio test.

Cell Systems, Volume *11*

Supplemental Information

Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome

Tina Wang, Jianzhu Ma, Andrew N. Hogan, Samson Fong, Katherine Licon, Brian Tsui, Jason F. Kreisberg, Peter D. Adams, Anne-Ruxandra Carvunis, Danika L. Bannasch, Elaine A. Ostrander, and Trey Ideker

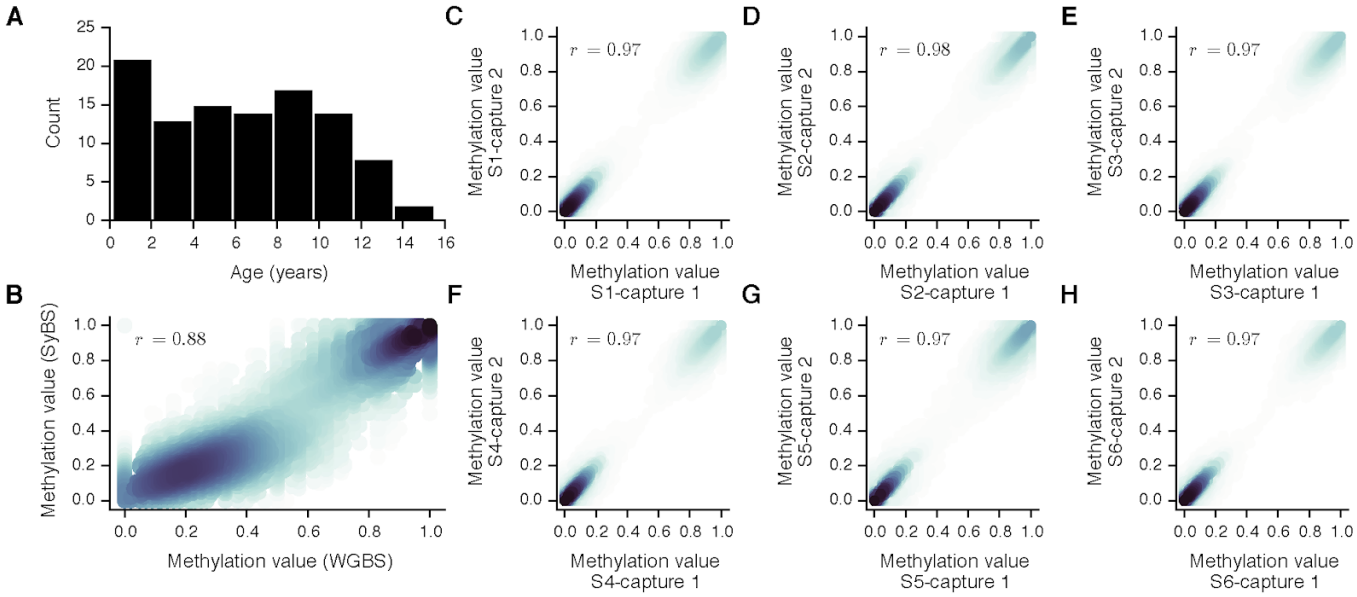


Figure S1: Concordance of SyBS with non-targeted sequencing. Related to Figure 1. (A) The distribution of ages of 104 dogs used in this study, depicted as a histogram. (**B-F**) Concordance of SyBS values for six canine DNA samples (S1-S6), for which two independent captures were performed. The x-axis represents values obtained for the first capture, and the y-axis represents values obtained for the second capture. The Pearson correlation of the two captures is shown. Colors represent density of observations.

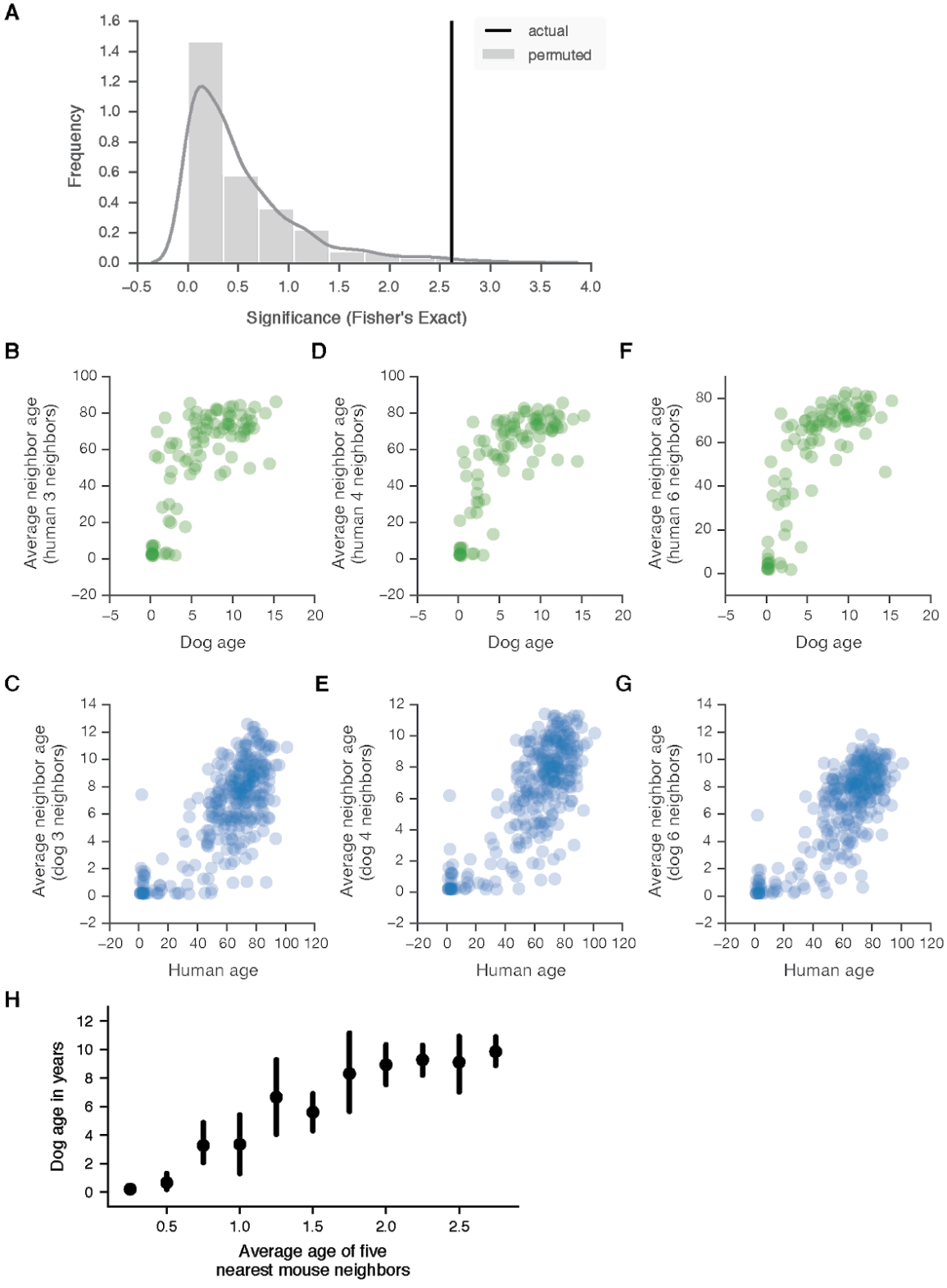


Figure S2: Evaluating methylome similarities observed for dogs and humans. Related to Figure 2. (**A**) Black line: the observed p-value of association between methylome similarity for each pair of species and age. P-value computed by Fisher's exact test. The 'significance' ($-\log_{10}$(p-value)) is compared to those obtained from 1000 randomizations in which dog and humans were permuted (gray bars). (**B-G**) Varying the number of nearest neighbors *k* in dog-to-human age alignments. The actual ages of dogs (in years, x-axis, **B,D,F**) or humans (x-axis, **C,E,G**) versus the average age of *k* human or dog neighbors (in years, y-axis), respectively, when varying *k*: (**B,C**) *k* = 3, (**D,E**) *k* = 4 and (**F,G**) *k* = 6. Results for *k* = 5 are shown in the main text (**Figure 2**). (**H**) The average age of five mouse neighbors (x-axis) and actual dog age (y-axis), where the observations are binned according to 0.2-year bins along the x-axis, such that the visualizations are identical to (**Fig. 2F**). Points illustrate the mean of each bin and bars represent the 95% confidence interval obtained from bootstrapping.

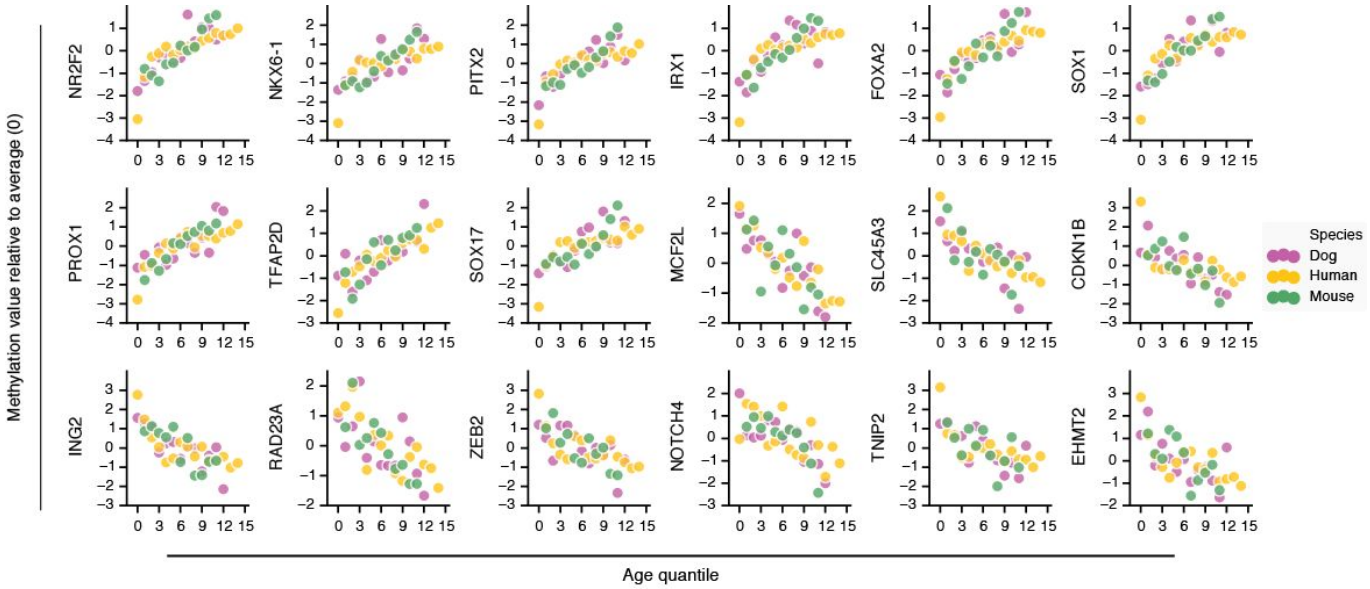


Figure S3: Representative genes exhibiting conserved methylation changes with age. Related to Figure 3. All dog, human and mouse individuals are ranked according to their ages translated to dog years (see text), then sorted into 15 age bins (quantiles, x-axis). Eighteen out of 394 orthologs for which CpG methylation values (y-axis) exhibit conserved direction of change with age are shown. Methylation values are normalized by the mean and standard deviation within each species. Colors indicate species.

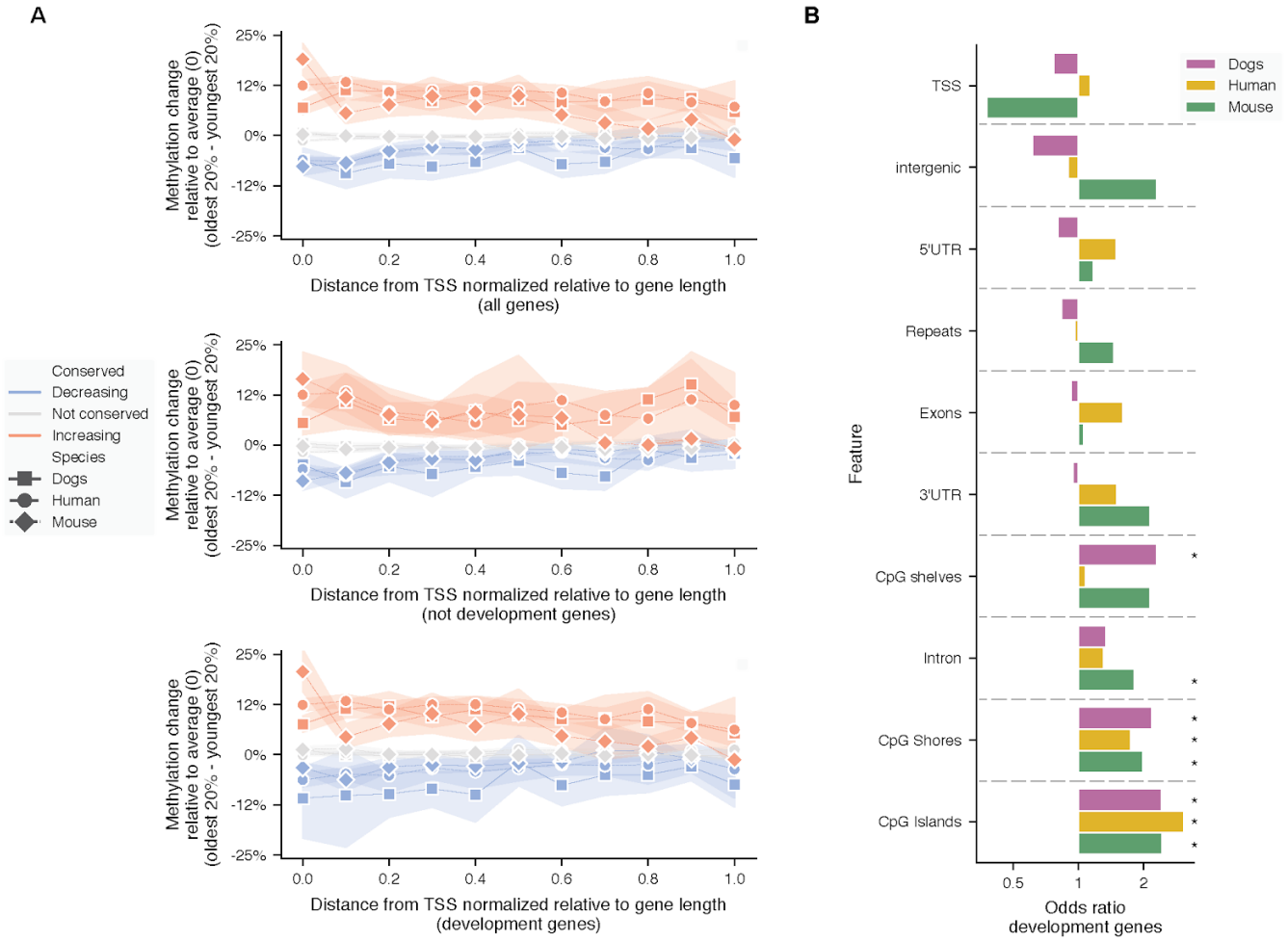


Figure S4: Conserved methylation changes with age in orthologous regions. Related to Figure 3. (**A**) CpGs were binned into 10 quantiles according to the distance from the transcription start site (TSS) normalized by the length of the gene for the 398 genes exhibiting conserved changes with age. The median methylation difference between the oldest 20% and youngest 20% of each species is shown according to a 3 CpG rolling window. In red are those genes that increase methylation with age, in not conserved genes and in blue are genes that decrease methylation with age. Each species is differentiated by markers as shown in the legend. Each point reflects the average difference with bands representing the 95% confidence interval, estimated from 1000 random samplings. The three plots show trends according to their developmental status: all genes (top), not development genes (middle) and developmental genes (bottom). (**B**) Odds ratio for the enrichment of genomic regions within conserved developmental genes when considering all development genes with colors representing different species. Asterisks (*) represent regions with exhibiting $p < 0.005$ by Fisher's exact test.

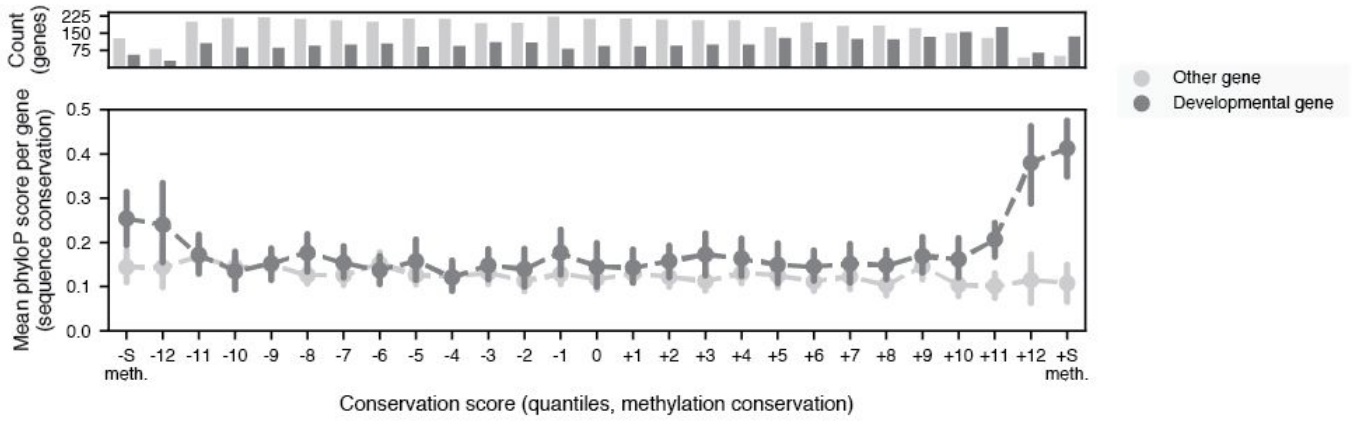


Figure S5: Relationship between sequence constraints and conserved methylation changes with age for developmental gene modules. Related to Figure 3. We analyzed 7,942 genes with 1:1 orthologs across humans, dogs and mice and for which methylation data were available in all species. For each ortholog we computed a methylation conservation score (x-axis, binned into quantiles) and a sequence conservation score (y-axis, PhyloP). The orthologs with most significant conserved increases or decreases in methylation with age (198 and 196 genes respectively) are indicated by −S and +S at the extreme ranges of x. Vertical error bars indicate the 95% confidence interval estimated by bootstrapping. Orthologs are treated separately according to their developmental status (dark points, developmental gene; light points, non-developmental gene). The plot reveals a strong association between methylation conservation and sequence conservation, in a manner that is specific to developmental genes. The top barplot depicts the number of genes within each quantile bin. Further details on computation of conservation scores are given in **STAR Methods**.

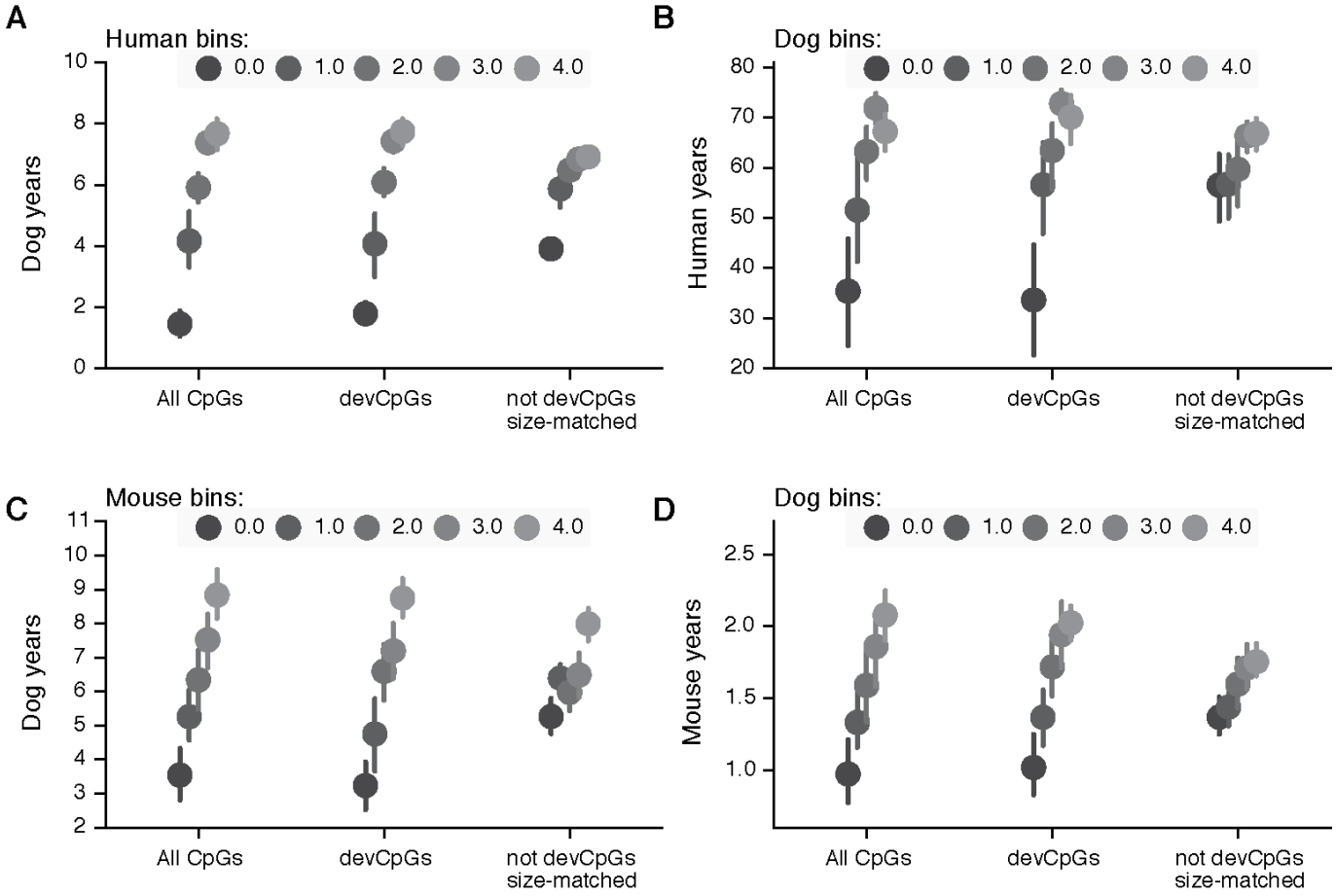


Figure S6: Dependence of the conserved epigenetic progression on CpGs within developmental gene modules. Related to Figure 3. The plots show alignments of age across species based on methylome similarity: (**A**) dogs versus humans, (**B**) humans versus dogs, **(C)** dogs versus mice, **(D)** mice versus dogs. Within each species comparison, methylome similarity is calculated using All CpGs, CpGs near development genes (devCpGs), or randomly sampled CpG sets of equal size but not near development genes (not devCpGs size-matched). In each comparison, the x-axis represents increasing age bins for one species (quintiles), and the y-axis represents the average ages of the 5 nearest neighbors in the opposite species according to methylome similarity. Vertical bars show the 95% confidence intervals of the averages obtained from 1000 bootstrapped samplings. Note that in all comparisons (A-D), the strong trends observed for All CpGs are preserved for devCpGs but are substantially weakened for the size-matched controls.

Table S1: Dog sample description. Related to Figure 2.†

Sample	Age (years)	Breed	Center		Sample	Age (years)	Breed	Center
2650	0.4	Lab	Davis		2623	15.5	Lab	Davis
7935	0.5	Lab	Davis		7936	0.5	Lab	Davis
3352	0.6	Lab	Davis		3018	1.8	Lab	Davis
3591	1.0	Lab	Davis		2519	2.5	Lab	Davis
5866	1.7	Lab	Davis		2538	5.7	Lab	Davis
3033	2.2	Lab	Davis		2967	6.3	Lab	Davis
2661	2.2	Lab	Davis		5856	6.6	Lab	Davis
7008	2.3	Lab	Davis		2667	7.6	Lab	Davis
7295	2.4	Lab	Davis		2935	8.3	Lab	Davis
2640	2.5	Lab	Davis		3592	9.3	Lab	Davis
2564	2.5	Lab	Davis		6537	10.0	Lab	Davis
7007	2.6	Lab	Davis		2624	10.6	Lab	Davis
2569	3.2	Lab	Davis		2888	10.9	Lab	Davis
3654	4.3	Lab	Davis		3274	10.9	Lab	Davis
3539	4.7	Lab	Davis		2714	11.0	Lab	Davis
3209	4.8	Lab	Davis		3696	11.3	Lab	Davis
2616	4.8	Lab	Davis		2638	11.4	Lab	Davis
2665	5.4	Lab	Davis		2664	11.4	Lab	Davis
2928	5.5	Lab	Davis		2542	11.5	Lab	Davis
2554	5.5	Lab	Davis		2266	11.9	Lab	Davis
2597	5.6	Lab	Davis		2477	12.1	Lab	Davis
2583	5.6	Lab	Davis		2489	12.3	Lab	Davis
2613	5.6	Lab	Davis		2270	12.7	Lab	Davis
2557	5.6	Lab	Davis		2849	14.5	Lab	Davis
2810	6.1	Lab	Davis		3592	9.3	Lab	Davis
3589	6.3	Lab	Davis		23855	0.1	Lab	NHGRI
2625	6.4	Lab	Davis		22139	0.2	Lab	NHGRI
2649	6.4	Lab	Davis		23848	0.2	Lab	NHGRI
7335	6.6	Lab	Davis		23852	0.2	Lab	NHGRI
7153	6.7	Lab	Davis		23840	0.2	Lab	NHGRI
6750	6.8	Lab	Davis		23826	0.2	Lab	NHGRI
3054	6.9	Lab	Davis		23820	0.2	Lab	NHGRI
2619	7.3	Lab	Davis		23824	0.2	Lab	NHGRI
7046	7.8	Lab	Davis		23792	0.2	Lab	NHGRI
3073	7.8	Lab	Davis		23800	0.2	Lab	NHGRI
3436	8.0	Lab	Davis		23796	0.2	Lab	NHGRI
3255	8.0	Lab	Davis		25293	1.8	Lab	NHGRI
2500	8.2	Lab	Davis		9137	3.0	Lab	NHGRI
3220	8.3	Lab	Davis		31279	3.1	Lab	NHGRI
7328	8.3	Lab	Davis		25415	3.8	Lab	NHGRI
2604	8.4	Lab	Davis		9127	5.3	Lab	NHGRI
2668	8.5	Lab	Davis		22211	9.5	Lab	NHGRI
2776	8.8	Lab	Davis		31117	9.5	Lab	NHGRI
2927	9.1	Lab	Davis		30379	11.7	Lab	NHGRI
2755	9.2	Lab	Davis		7725	12.6	Lab	NHGRI
3328	9.5	Lab	Davis		13295	13.3	Lab	NHGRI
2862	9.6	Lab	Davis		D1	1.5	Mix	CV
2903	9.7	Lab	Davis		D6	1.9	Mix	CV
2916	9.8	Lab	Davis		D4	4.2	Mix	CV
3151	9.9	Lab	Davis		D5	4.6	Mix	CV
6529	9.9	Lab	Davis		D2	6.8	Pug	CV
2747	10.0	Lab	Davis					

† The table lists the 104 dogs used in this study: the sample ID, age in years and center where the sample was obtained are shown. Abbreviations: Labrador retriever (Lab), Mixed breed (Mix), UC Davis (Davis), National Human Genome Research Institute (NHGRI), Consented Volunteer (CV).