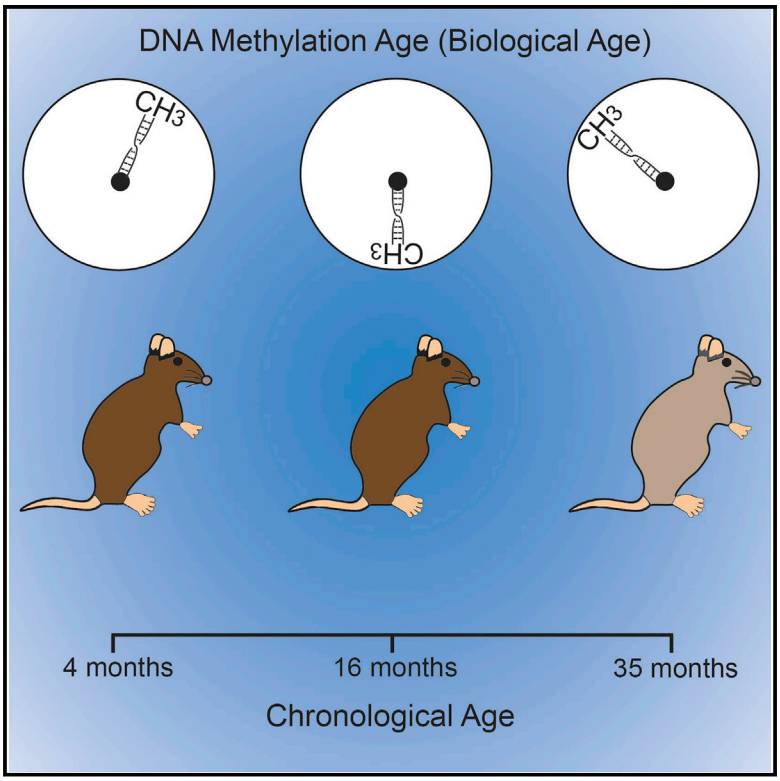
Short Article

Cell Metabolism

Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions

Graphical Abstract



Authors

Daniel A. Petkovich, Dmitriy I. Podolskiy, Alexei V. Lobanov, Sang-Goo Lee, Richard A. Miller, Vadim N. Gladyshev

Correspondence

vgladyshev@rics.bwh.harvard.edu

In Brief

Petkovich et al. developed a precise biomarker of mouse aging based on blood DNA methylation patterns. Their epigenetic clock can estimate the biological age of various mouse models and can be applied to evaluate the effects of genetic, pharmacological, and dietary interventions on average lifespan.

Highlights

- Mouse blood DNA methylation profiles are used to create an age-predicting clock
- CpG sites contributing to the DNA methylation clock are distinct in human and mouse
- The clock predicts modulation of lifespan by genetic and dietary interventions
- DNA methylomes are characterized by CpGs representing an organism's biological age

Petkovich et al., 2017, Cell Metabolism *25*, 954–960
April 4, 2017
http://dx.doi.org/10.1016/j.cmet.2017.03.016

CellPress

CellPress

Cell Metabolism
Short Article

Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions

Daniel A. Petkovich,[1,3] Dmitriy I. Podolskiy,[1,3] Alexei V. Lobanov,[1] Sang-Goo Lee,[1] Richard A. Miller,[2] and Vadim N. Gladyshev[1,4,*]

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA
[2]Department of Pathology and Geriatrics Center, University of Michigan, Ann Arbor, MI 48109, USA
[3]Co-first author
[4]Lead Contact
*Correspondence: vgladyshev@rics.bwh.harvard.edu
http://dx.doi.org/10.1016/j.cmet.2017.03.016

SUMMARY

The DNA methylation levels of certain CpG sites are thought to reflect the pace of human aging. Here, we developed a robust predictor of mouse biological age based on 90 CpG sites derived from partial blood DNA methylation profiles. The resulting clock correctly determines the age of mouse cohorts, detects the longevity effects of calorie restriction and gene knockouts, and reports rejuvenation of fibroblast-derived iPSCs. The data show that mammalian DNA methylomes are characterized by CpG sites that may represent the organism's biological age. They are scattered across the genome, they are distinct in human and mouse, and their methylation gradually changes with age. The clock derived from these sites represents a biomarker of aging and can be used to determine the biological age of organisms and evaluate interventions that alter the rate of aging.

INTRODUCTION

Organisms age at different rates, which are influenced by genotype, environment, and stochastic processes. Dietary, pharmacological, and genetic interventions offer an opportunity to adjust these rates and to examine, model, and ultimately regulate the process of aging in laboratory animals—and, potentially, in humans. However, identifying such interventions is currently both time consuming and cost prohibitive. An accurate estimator of the biological age of organisms subject to an intervention, as compared to their chronological age, can resolve this problem and find many applications in biomedical science. Although no reliable molecular methods are currently available that could determine the biological age of model organisms, estimates of aging rates have recently been developed for humans based on DNA methylation (DNAm) profiles (Hannum et al., 2013; Horvath, 2013; Weidner et al., 2014).

DNAm has pivotal roles in regulation of transcriptional programs and systematically varies as a function of age (Day et al., 2013; Jones, 2012; Jones et al., 2015; Maegawa et al., 2010; Schultz et al., 2015; Smith and Meissner, 2013; Zampieri et al., 2015). These changes start during embryogenesis and continue throughout the lifespan, affecting chromatin states, lineage specialization, gene expression, genome stability, and self-renewal of stem cells (Beerman and Rossi, 2014; Benayoun et al., 2015; Roadmap Epigenomics Consortium et al., 2015; Guo et al., 2014; Stelzer et al., 2015; Sun et al., 2014; Taiwo et al., 2013). The human DNAm clock model can predict certain health outcomes, such as increased future mortality (Chen et al., 2016; Christiansen et al., 2016; Horvath et al., 2015a; Marioni et al., 2015). In addition, accelerated DNAm aging was observed in patients with HIV infection and Down syndrome, and slower DNAm changes were reported for cerebellum aging (Horvath and Levine, 2015; Horvath et al., 2015b, 2015c). In this work, we sought to develop and employ a DNAm clock in the house mouse (*Mus musculus*), the main biological model of human aging and disease. The clock based on mouse blood DNA methylome was prepared and applied to test longevity interventions.

RESULTS AND DISCUSSION

To assess global changes in the DNA methylome, we subjected whole blood DNA of 141 male C57BL/6 mice (16 age groups, ranging from 3- to 35-month-old animals) to a modified version of reduced-representation bisulfite sequencing (RRBS) (Figure 1A; Table 1). Among 3.9–13.0 million CpG sites detected in individual samples (Tables S1 and S2, related to Table 1), more than 1.9 million sites were present in all samples (Figure S1A, related to Figure 1A and Table 1). These sites represented more than 8% of the mouse DNA methylome and mapped to 99% of UCSC RefGenes, 88% annotated CpG islands, 55% CpG island shores, 41% promoters, and 10% predicted enhancers (Figure S1B). Overall, this RRBS procedure resulted in significant coverage of the mouse blood DNA methylome (Figures S1A–S1C; Boyle et al., 2012) and allowed us to assess its changes as a function of age.

To characterize the resulting DNA methylome, we extracted the leading age-dependent DNAm signature using non-negative matrix factorization and estimated the behavior of average DNAm with age. This revealed a global pattern of hypomethylation slowly developing with age, akin to that observed in humans (Hannum et al., 2013; Horvath, 2013). Following development, the leading DNAm signature gradually decreased ($R^2 = 0.545$, bisquare robust linear fit) with age changing by ~5% ($p = 1.1 \times 10^{-4}$, two-sample t test) between 4 and 35 months

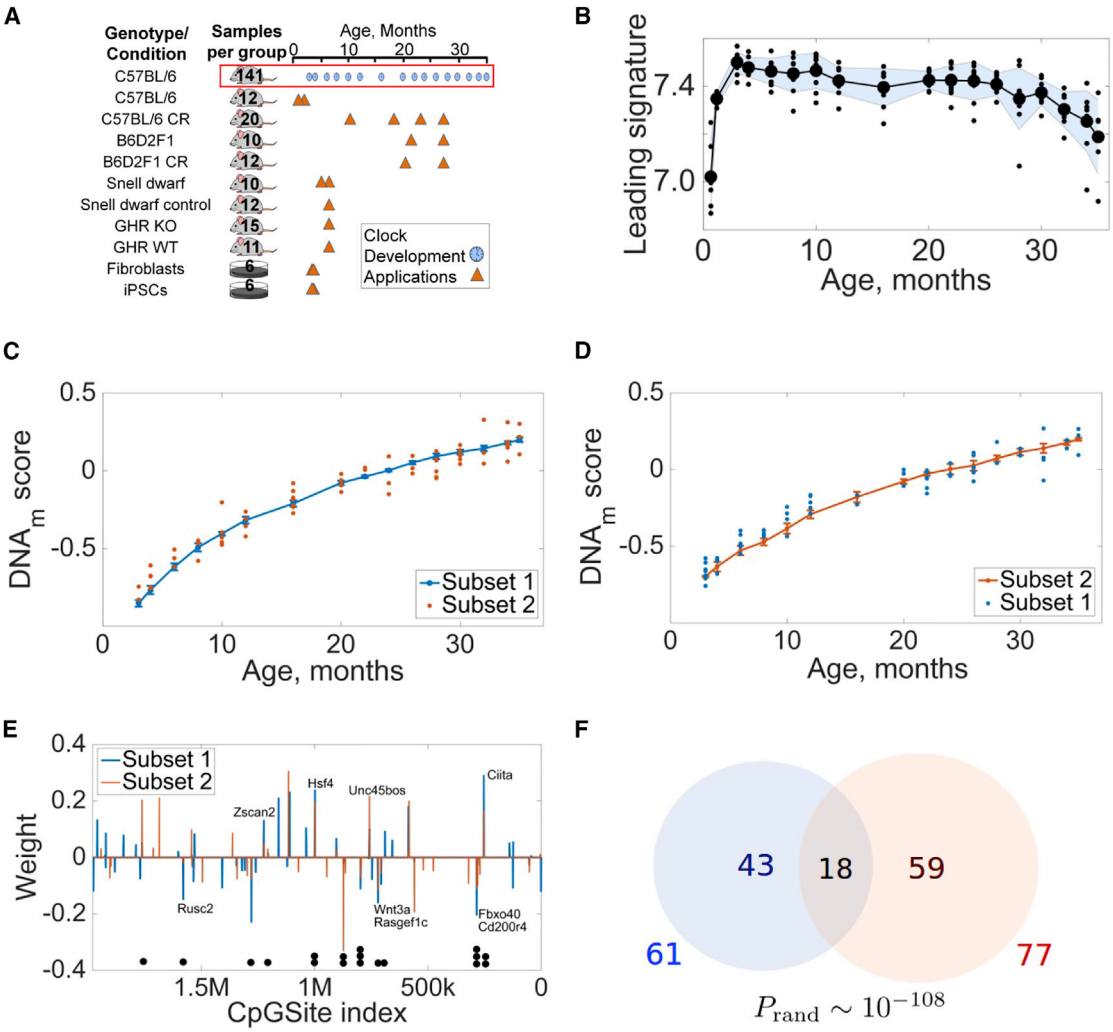


Figure 1. DNA Methylation Patterns Predict the Age of Mice

(A) Overview of mouse models used in the study. Blue circles show samples used to build the clock, and triangles show models analyzed with the clock. Numbers indicate the number of animals or cell cultures for each genotype and cohort; see also Tables 1 and S1.

(B) Behavior of the leading age-dependent DNAm signature for 141 C57BL/6 males with age.

(C) Behavior of the subset 1 clock (blue). Orange dots correspond to samples from subset 2. Goodness of fit $R^2 = 0.959$, $p = 1.16 \times 10^{-49}$.

(D) Behavior of the subset 2 clock (orange). Blue dots represent samples from subset 1. Goodness of fit $R^2 = 0.959$, $p = 6.26 \times 10^{-49}$.

(E) Weights and genome distribution of CpG sites contributing to subset 1 (blue) and subset 2 (orange) clocks. Black dots below the graph point to 18 CpG sites common to both clocks.

(F) Number of CpG sites contributing to subset 1 and 2 clocks. The probability of finding 18 common sites in two random sets derived from 1.9 million sites is $\sim 10^{-108}$.

of age (Figure 1B). The weights of different CpG sites in the leading signature suggested that the pattern of hypomethylation is determined by the dynamics of methylation levels on the whole DNA methylome.

To examine feasibility of the resulting partial DNA methylomes for describing the age of mice, we randomly divided 141 C57BL/6 mice into two subsets of 70 and 71 animals (subsets 1 and 2). DNAm clocks were then constructed separately for subsets 1 and 2 using elastic net regression with 10-fold cross-validation. The clock built using subset 1 was validated on the DNA methylation data from subset 2 (Figure 1C) and vice versa (Figure 1D); this revealed high correspondence between the methylation scores calculated using both subsets. Precision of the clocks built using subsets 1 or 2 and tested on subsets 2 or 1 generally decreased with age (Figure S2A, related to Figures 1C and 1D): in early life (<10 months) the chronological age corresponded to the DNA methylation age with precision of 1–3 months, whereas the precision dropped to 5–7 months for the oldest mice. However, the relative precision of the clock, defined as the ratio of the clock error to the chronological age, somewhat increased (Figure S2B, related to Figures 1C and 1D).

The subset 1 clock was determined by the weighted average methylation on 61 different CpG sites, and the subset 2 clock by the methylation on 77 CpG sites (Figure 1E). Interestingly, 18 sites were common to both clocks (Figure 1F), and the values of their contributing weights were close. A relatively small (but

Table 1. Characteristics of Samples Used in the Study

Strain (Treatment)	Source	Age (months)	Sex	Number of Samples
C57BL/6	NIA and BWH	0.67	Male	6
C57BL/6	NIA and BWH	1.17	Male	6
C57BL/6	NIA	3	Male	9
C57BL/6	NIA	4	Male	9
C57BL/6	NIA	6	Male	9
C57BL/6	NIA	8	Male	9
C57BL/6	NIA	10	Male	9
C57BL/6	NIA	12	Male	10
C57BL/6	NIA	16	Male	9
C57BL/6	NIA	20	Male	9
C57BL/6	NIA	22	Male	9
C57BL/6	NIA	24	Male	9
C57BL/6	NIA	26	Male	10
C57BL/6	NIA	28	Male	9
C57BL/6	NIA	30	Male	9
C57BL/6	NIA	32	Male	7
C57BL/6	NIA	34	Male	9
C57BL/6	NIA	35	Male	6
C57BL/6 (Calorie Restriction)	NIA	10	Male	5
C57BL/6 (Calorie Restriction)	NIA	18	Male	5
C57BL/6 (Calorie Restriction)	NIA	23	Male	5
C57BL/6 (Calorie Restriction)	NIA	27	Male	5
B6D2F1	NIA	20	Male	5
B6D2F1	NIA	27	Male	5
B6D2F1 (Calorie Restriction)	NIA	21	Male	7
B6D2F1 (Calorie Restriction)	NIA	27	Male	5
Snell Dwarf	U of Michigan	5	Male	2
Snell Dwarf	U of Michigan	6	Male	4
Snell Dwarf	U of Michigan	6	Female	4
Snell Dwarf Control	U of Michigan	5	Male	2
Snell Dwarf Control	U of Michigan	6	Male	5
Snell Dwarf Control	U of Michigan	6	Female	5
GHRKO	U of Michigan	6	Male	9
GHRKO	U of Michigan	6	Female	6
GHR WT	U of Michigan	6	Male	8
GHR WT	U of Michigan	6	Female	3

List of the strains, genotypes, sources, ages, and sexes of mice used in this study. NIH, National Institute on Aging; BWH, Brigham and Women's Hospital; U of Michigan, University of Michigan.

highly significant) overlap between the clocks built using Subsets 1 and 2 cannot be explained by a poor similarity of samples (Figure S1D, related to Figures 1C and 1D). Instead, we suggest that this is due to the fact that DNA methylation levels of many CpG sites are correlated with each other, and elastic net regularization picks relatively small subsets of these correlated sites (Figure S1E, related to Figures 1C and 1D). We also examined the previously built human DNA methylation clocks (Hannum et al., 2013; Horvath, 2013) and observed several common CpG sites (three sites for the 89-site clock and two for the 71-site clock), even though these clocks were built using different tissues. The data thus show that mouse and likely other mammalian genomes are characterized by certain CpG sites that may represent the age of organisms.

To develop a more precise mouse DNA methylation clock, we constructed a class of elastic net regression clocks based on the whole dataset representing 141 C57BL/6 mice. Among the clocks corresponding to different values of regularization parameter, we chose the model (further designated as mDNAm clock) that minimized both cross-validation error and the chance of overfitting (Figure 2A). We further recalculated the DNAm score into the methylation age, Age_{Met} (Figure 2B). On the training set, the estimated Age_{Met} of C57BL/6 mice coincided with the chronological age of the samples ($R^2 = 0.9971$, $p = 10^{-21}$) with

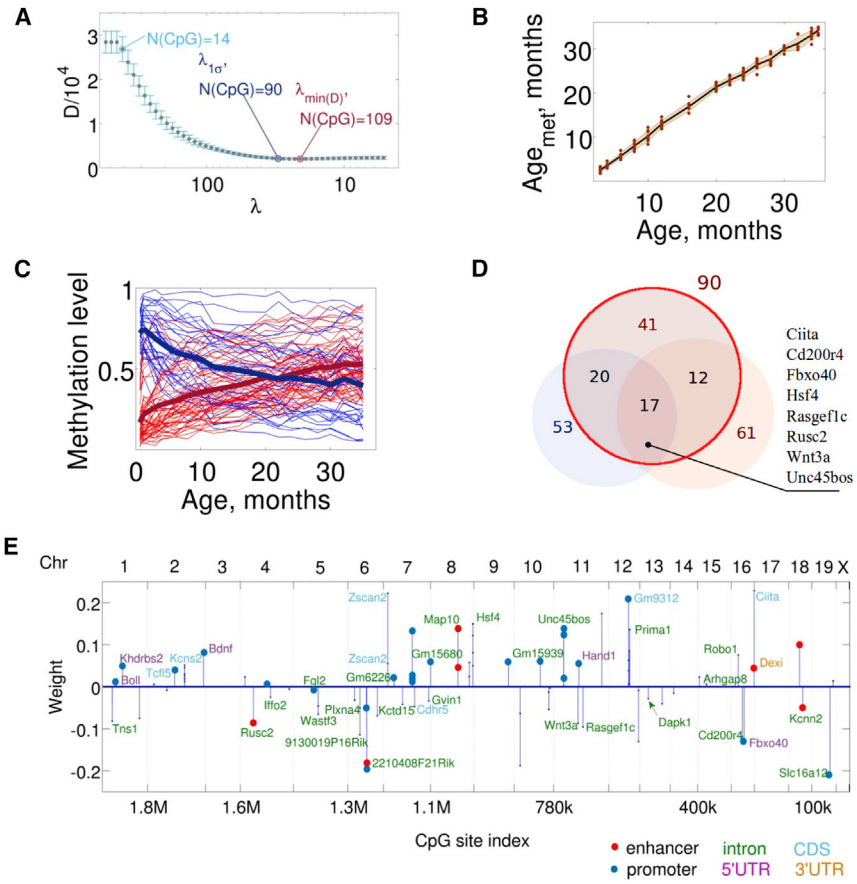


Figure 2. Development of the mDNAm Clock

(A) Selection of the optimally robust mDNAm clock. The clock corresponding to the minimum cross-validation deviation error is a weighted average of the DNAm levels of 109 CpG sites. The optimally robust clock is a weighted average of the DNAm levels of 90 CpG sites.

(B) Estimation of the mDNAm age of C57BL/6 control males.

(C) Trajectories of methylation levels of the 90 CpG sites that form the clock. Age-related increases in DNAm are shown in red, and decreases are shown in blue. Solid dark blue and red lines correspond to the signal averaged over the CpG sites, with methylation levels decreasing or increasing with age, respectively.

(D) The overlap between the CpG sites contributing to the 90-site mDNAm clock (red circle), subset 1 clock (lower left), and subset 2 clock (lower right). Gene list on the right shows genes that include 17 CpG sites common to all clocks.

(E) Distribution and weights of 90 CpG sites defining the clock along the genome. Positions of the contributing CpG sites within the genome are shown. CpGs were located within the bodies of particular genes, introns, and untranslated regions as well as in intergentic regions. Most mouse chromosomes (except 3, 17, X, and Y) contain at least one CpG site contributing to the clock. The color scheme shows the indicated sequence or functional elements within which the sites reside.

the mean square difference between the mDNAm age and the chronological age of 16 days. To determine an out-of-sample lower-bound estimate on the precision of the clock, we again used subsets 1 and 2 (Figures 1C and 1D). This estimation resulted in the lower out-of-sample bound on the precision of the clock $R^2 = 0.9008$. High accuracy of the model may be attributed to the use of a single mouse strain and identical environmental conditions, although these factors may also introduce a source of bias. There is also a potential for confounding as a result of changes of blood cell composition as a function of age (Marioni et al., 2015).

The mDNAm clock was defined by a weighted average DNA methylation over 90 CpG sites (Figures 2C and S1F, related to Figure 2C). The methylation state of these sites changed gradually, decelerating with age, with both hypomethylated and hypermethylated sites approaching more intermediate DNA methylation values (Figures 2C and S1F). The sites contributed unequally to the mDNAm clock (Table S3, related to Figure 2C) and formed several distinct clusters (~30) associated with genes *Hsf4*, *Kcns2*, *Map10*, *Tns2*, *Wnt3a*, and *Zscan2*. We found that 17 out of 18 CpG sites common to subset 1 and 2 clocks were also present among 90 CpG sites of the mDNAm clock (Figure 2D). Most of these 17 CpG sites were located within introns of *Ciita*, *Cd200r4*, *Rasgef1c*, *Wnt3a*, and *Zscan2*, and several were clustered. CpG sites in introns of particular genes often play a role of their secondary enhancers (Blattler et al., 2014), and we note that several identified genes are involved in development, differentiation, and tissue morphogenesis, consistent with a program-like behavior (Gladyshev, 2013). The CpG sites contributing to the mDNAm clock were distributed across the chromosomes (Figure 2E) and did not match the sites observed in human clocks (Hannum et al., 2013; Horvath, 2013). In addition, the pace of human and mouse clocks differed considerably, reflecting the difference in the rates of aging in humans and mice.

We further tested whether the mDNAm clock can detect the effects of interventions that are known to increase lifespan and delay age-dependent phenotypic changes and thus plausibly serve as an index of the biological age of mice rather than their chronological age. Estimation of Age_{met} for calorie-restricted C57BL/6 males (four age groups, with dietary intervention started at 14 weeks for all mice; Table 1) revealed on average 20% lower Age_{Met} than their chronological age (Figure 3A), consistent with the effect of calorie restriction on lifespan in this strain (Turturro et al., 1999). The lower biological age of calorie-restricted mice was highly significant in the tested age groups (Figure 3A; one-sided, two-sample t test for ΔAge_{Met} in the combined age groups, $p \approx 7.17 \times 10^{-12}$); however, the effect was less pronounced in younger animals (10-month-old mice) and was not observed in a separate experiment where 4-month-old mice were subjected to calorie restriction for 2 months (Figure S3A, related to Figure 3A; $p = 0.601$ for the t test of the control and calorie-restricted group DNA methylation ages belonging to different distributions). This observation is congruent with the finding that severe weight loss in humans

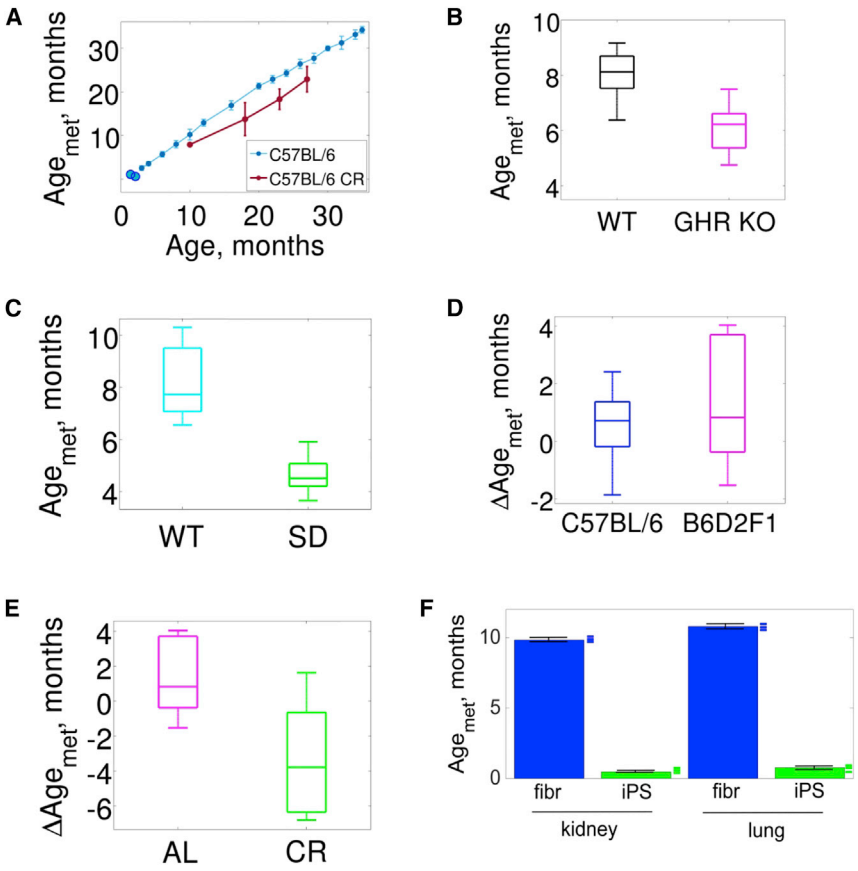


Figure 3. Applications of the mDNAm Clock

(A) Application of the mDNAm clock (light blue) to calorie-restricted (CR) C57BL/6 males (red). Light blue blobs not connected to the clock correspond to 20- and 35-day-old samples, with each cohort including six mice (not used to construct the clock).

(B) Whole-body growth-hormone-receptor knockout (GHR-KO) and WT ([C57BL/6J x BALB/cByJ]/F2) mice. The chronological age of GHR-KO mice was 5.9 ± 0.3 months, and the chronological age of WT mice was 5.9 ± 0.4 months.

(C) Snell dwarf (SD) and control (WT; [DW/J x C3H/HEJ]/F2) mice. The chronological age of both SD and WT mice was 5.9 ± 0.4 months.

(D) Comparing mDNAm ages of B6D2F1 and C57BL/6 mice. The differences ΔAge_{met} between chronological and mDNAm ages were calculated for cohorts of B6D2F1 and C57BL/6 mice.

(E) Comparing mDNAm ages of B6D2F1 control (AL) strain and the same calorie-restricted (CR) strain. The differences ΔAge_{met} between chronological and mDNAm ages were calculated for AL and CR cohorts.

(F) Mouse lung and kidney fibroblasts and fibroblast-derived iPSCs. Fibroblasts were collected from C57BL/6 mice and grown in culture; iPSCs were then generated from them. Blue and green marks on the right of the bars represent individual samples.

following bariatric surgery was not accompanied by a reduction of DNAm age (Horvath et al., 2014). Thus, the pattern of DNA methylation indicative of the difference between biological and chronological age appears to develop gradually, suggesting a cumulative function.

We also examined genetic interventions of longevity. Full-body growth hormone receptor knockout (GHR-KO) (Coschigano et al., 2003) showed a robust reduction in Age_{Met} as compared to control (one-sided, two-sample t test, $p = 5.6 \times 10^{-5}$; Figure 3B). Application of the mDNAm clock to Snell dwarf mice and their respective wild-type control ([DW/J × C3H/HEJ]/F2) showed, on average, 50% lower age (one-sided, two-sample t test, $p = 0.0243$) of Snell dwarf mice (Figure 3C), consistent with 42% mean survival extension in this longevity model (Flurkey et al., 2001). Although the clock was developed using samples of male mice, we found that it also correctly predicted the age of female mice (Figures S3B–S3E, related to Figure 3).

We further analyzed additional genetic backgrounds. The difference between Age_{Met} of B6D2F1 mice and their chronological age was not statistically significant (one-sided, two-sample t test, $p = 0.0510$; Figure 3D). The Age_{Met} of calorie-restricted B6D2F1 mice was 20% lower than that of the corresponding ad libitum mice (one-sided, two-sample t test, $p = 1.0 \times 10^{-5}$; Figure 3E). We also analyzed the Age_{Met} of 20- and 35-day-old C57BL/6 mice, which were the ages not covered by the clock, and the model correctly estimated the age of these mice (Figure 3A).

Finally, we collected kidney and lung fibroblasts from 10-week-old C57BL/6 mice, maintained and aged them in culture, and prepared six independent iPSC lines from them (Figure S4, related to STAR Methods sections "Primary Fibroblasts and Generation of Mouse iPSCs" and "Characterization of Mouse iPSCs"). The calculated Age_{Met} of kidney and lung fibroblasts was 295 ± 3 days and 323 ± 3 days, respectively, whereas all iPSCs were less than 1 month old (kidney, 14 ± 0.3 days; lung, 22 ± 1.2 days; Figure 3F). Although estimation of the exact age of fibroblasts and iPSCs may be influenced by the fact that the clock was developed based on mouse blood, the data show that the iPSC procedure modifies mDNAm to resemble that seen in younger blood. A similar effect was previously observed in human studies (Horvath, 2013).

We think that identification of a relatively small set of CpG sites forming the clock is possible because of a very large number of observables (CpG sites) produced by the RRBS procedure and then subjected to elastic net regression, which extracts a combination of sites responsible for a given age-dependent pattern of behavior. Thus, presumably, accurate molecular clocks measuring biological age of a sample may also be developed based on other high-throughput approaches, such as metabolite profiling and RNA-seq. The structure of the mDNAm clock may also change when the clock is developed across mouse tissues, as was previously observed for the DNA methylation clocks in humans (Hannum et al., 2013; Horvath, 2013). We also suggest that researchers utilizing the clock use experimental and control groups to determine the difference between chronological and

biological age, as the RRBS procedure is not fully standardized. For example, a discrepancy of ~2 months between the chronological and estimated biological age of the GHR-KO and Snell controls (Figures 3B and 3C) can, in principle, be attributed to a lower precision on the methylation clock out of sample and/or a difference in genetic background. It should be expected that whole-genome bisulfate sequencing data may further add to the power and resolution of the DNA methylation clocks due to a further increase in the number of age-dependent observables.

To conclude, we constructed a robust epigenetic clock based on the age-dependent behavior of blood DNA methylomes of C57BL/6 mice. The clock can estimate the age of cohorts of C57BL/6 mice, characterize the change in biological age in other mouse models, and evaluate the effects of genetic, pharmacological, and dietary interventions on average lifespan. More broadly, we found that there are certain CpG sites in mammals that are scattered across the genome, differ across species, and can form the biological clock. These sites can differentiate chronological and biological ages of organisms in a cumulative manner and represent a biomarker of aging. It may be expected that a related set of CpG sites may form a multitissue clock similar to that observed in humans (Horvath, 2013). We suggest that the mDNAm clock will find many applications in biomedical science, especially in areas where mouse models may be associated with changes in healthspan and lifespan.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Mouse Models
  - Genomic DNA Purification
  - RRBS
  - Primary Fibroblasts and Generation of Mouse iPSCs
  - Characterization of Mouse iPSCs
  - Generation of Embryoid Bodies and Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis of Sequence Reads
  - Estimating the Leading Methylation Signature and Its Age-Dependent Pattern
  - Controlling for Batch Effect
  - Constructing the Elastic Net Regression Clock and Estimating the Methylation Age of Samples
  - Statistical Analysis of Data Represented on Figures 1, 2, and 3
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cmet.2017.03.016.

AUTHOR CONTRIBUTIONS

V.N.G. supervised the study. D.A.P., D.I.P., and V.N.G. designed experiments and interpreted results. D.A.P. collected all mouse samples, developed the adjusted RRBS protocol, sequenced the samples, and contributed to bioinformatics analyses. D.I.P. developed the mDNAm clock and carried out other computational analyses. A.V.L. mapped sequencing reads, performed methylation calls, and carried out additional bioinformatics analyses. S.-G.L. prepared primary fibroblasts and iPSCs and contributed to mouse sample collection. R.A.M. provided growth hormone knockout and Snell dwarf models and the corresponding controls. D.I.P. and V.N.G. wrote the paper with significant contributions from D.A.P. and input from all authors.

ACKNOWLEDGMENTS

We thank NIA (Aged Rodent Colony) for providing the mice and Adeline Augereau, Andrei Avanesov, Maxim Gerashchenko, Alaattin Kaya, Marco Mariotti, Zalan Peterfi, Aleksei Mikhalchenko, Xuming Zhou, and Sun Hee Yim for discussion and help with sample preparation. This work was supported by NIH AG047745, AG047200, and AG019899.

Received: October 18, 2016
Revised: January 9, 2017
Accepted: March 21, 2017
Published: April 4, 2017

REFERENCES

Antoulas, A., and Sorensen, D. (2001). Approximation of large-scale dynamical systems: an overview. Int. J. Appl. Math. Comput. Sci. *5*, 1093–1121. http://pldml.icm.edu.pl/pldml/element/bwmeta1.element.bwnjournal-article-amcv11i5p1093bwm.

Beerman, I., and Rossi, D.J. (2014). Epigenetic regulation of hematopoietic stem cell aging. Exp. Cell Res. *329*, 192–199.

Benayoun, B.A., Pollina, E.A., and Brunet, A. (2015). Epigenetic regulation of ageing: linking environmental inputs to genomic stability. Nat. Rev. Mol. Cell Biol. *16*, 593–610.

Blattler, A., Yao, L., Witt, H., Guo, Y., Nicolet, C.M., Berman, B.P., and Farnham, P.J. (2014). Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. Genome Biol. *15*, 469.

Boyle, P., Clement, K., Gu, H., Smith, Z.D., Ziller, M., Fostel, J.L., Holmes, L., Meldrim, J., Kelley, F., Gnirke, A., and Meissner, A. (2012). Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. Genome Biol. *13*, R92.

Buganim, Y., Markoulaki, S., van Wietmarschen, N., Hoke, H., Wu, T., Ganz, K., Akhtar-Zaidi, B., He, Y., Abraham, B.J., Porubsky, D., et al. (2014). The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. Cell Stem Cell *15*, 295–309.

Carey, B.W., Markoulaki, S., Hanna, J., Saha, K., Gao, Q., Mitalipova, M., and Jaenisch, R. (2009). Reprogramming of murine and human somatic cells using a single polycistronic vector. Proc. Natl. Acad. Sci. USA *106*, 157–162.

Chen, B.H., Marioni, R.E., Colicino, E., Peters, M.J., Ward-Caviness, C.K., Tsai, P.-C., Roetker, N.S., Just, A.C., Demerath, E.W., Guan, W., et al. (2016). DNA methylation-based measures of biological age: meta-analysis predicting time to death. Aging (Albany, N.Y.) *8*, 1844–1865.

Christiansen, L., Lenart, A., Tan, Q., Vaupel, J.W., Aviv, A., McGue, M., and Christensen, K. (2016). DNA methylation age is associated with mortality in a longitudinal Danish twin study. Aging Cell *15*, 149–154.

Coschigano, K.T., Holland, A.N., Riders, M.E., List, E.O., Flyvbjerg, A., and Kopchick, J.J. (2003). 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.

Day, K., Waite, L.L., Thalacker-Mercer, A., West, A., Bamman, M.M., Brooks, J.D., Myers, R.M., and Absher, D. (2013). Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. Genome Biol. *14*, R102.

Flurkey, K., Papaconstantinou, J., Miller, R.A., and Harrison, D.E. (2001). Lifespan extension and delayed immune and collagen aging in mutant mice

with defects in growth hormone production. Proc. Natl. Acad. Sci. USA *98*, 6736–6741.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. *33*, 1–22.

Gladyshev, V.N. (2013). The origin of aging: imperfectness-driven non-random damage defines the aging process and control of lifespan. Trends Genet. *29*, 506–512.

Gu, H., Smith, Z.D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat. Protoc. *6*, 468–481.

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., et al. (2014). The DNA methylation landscape of human early embryos. Nature *511*, 606–610.

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.

Hockemeyer, D., Soldner, F., Cook, E.G., Gao, Q., Mitalipova, M., and Jaenisch, R. (2008). A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. Cell Stem Cell *3*, 346–353.

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

Horvath, S., and Levine, A.J. (2015). HIV-1 infection accelerates age according to the epigenetic clock. J. Infect. Dis. *212*, 1563–1573.

Horvath, S., Erhart, W., Brosch, M., Ammerpohl, O., von Schönfels, W., Ahrens, M., Heits, N., Bell, J.T., Tsai, P.-C., Spector, T.D., et al. (2014). Obesity accelerates epigenetic aging of human liver. Proc. Natl. Acad. Sci. USA *111*, 15538–15543.

Horvath, S., Pirazzini, C., Bacalini, M.G., Gentilini, D., Di Blasio, A.M., Delledonne, M., Mari, D., Arosio, B., Monti, D., Passarino, G., et al. (2015a). Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. Aging (Albany, N.Y.) *7*, 1159–1170.

Horvath, S., Garagnani, P., Bacalini, M.G., Pirazzini, C., Salvioli, S., Gentilini, D., Di Blasio, A.M., Giuliani, C., Tung, S., Vinters, H.V., and Franceschi, C. (2015b). Accelerated epigenetic aging in Down syndrome. Aging Cell *14*, 491–495.

Horvath, S., Mah, V., Lu, A.T., Woo, J.S., Choi, O.-W., Jasinska, A.J., Riancho, J.A., Tung, S., Coles, N.S., Braun, J., et al. (2015c). The cerebellum ages slowly according to the epigenetic clock. Aging (Albany, N.Y.) *7*, 294–306.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Jones, M.J., Goodman, S.J., and Kobor, M.S. (2015). DNA methylation and healthy human aging. Aging Cell *14*, 924–932.

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

Marioni, R.E., Shah, S., McRae, A.F., Chen, B.H., Colicino, E., Harris, S.E., Gibson, J., Henders, A.K., Redmond, P., Cox, S.R., et al. (2015). DNA methylation age of blood predicts all-cause mortality in later life. Genome Biol. *16*, 25.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N., Nery, J.R., Urich, M.A., Chen, H., et al. (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. Nature *523*, 212–216.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220.

Stelzer, Y., Shivalila, C.S., Soldner, F., Markoulaki, S., and Jaenisch, R. (2015). Tracing dynamic changes of DNA methylation at single-cell resolution. Cell *163*, 218–229.

Sun, D., Luo, M., Jeong, M., Rodriguez, B., Xia, Z., Hannah, R., Wang, H., Le, T., Faull, K.F., Chen, R., et al. (2014). Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. Cell Stem Cell *14*, 673–688.

Taiwo, O., Wilson, G.A., Emmett, W., Morris, T., Bonnet, D., Schuster, E., Adejumo, T., Beck, S., and Pearce, D.J. (2013). DNA methylation analysis of murine hematopoietic side population cells during aging. Epigenetics *8*, 1114–1122.

Turturro, A., Witt, W.W., Lewis, S., Hass, B.S., Lipman, R.D., and Hart, R.W. (1999). Growth curves and survival characteristics of the animals used in the Biomarkers of Aging Program. J. Gerontol. A Biol. Sci. Med. Sci. *54*, B492–B501.

Weidner, C.I., Lin, Q., Koch, C.M., Eisele, L., Beier, F., Ziegler, P., Bauerschlag, D.O., Jöckel, K.-H., Erbel, R., Mühleisen, T.W., et al. (2014). Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biol. *15*, R24.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al.; Mouse ENCODE Consortium (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355–364.

Zampieri, M., Ciccarone, F., Calabrese, R., Franceschi, C., Bürkle, A., and Caiafa, P. (2015). Reconfiguration of DNA methylation in aging. Mech. Ageing Dev. *151*, 60–70.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
OCT-4 antibody (C-10)	Santa Cruz Biotech	Cat#sc-5279; RRID:AB_628051
SOX2 antibody	Millipore	Cat#AB5603; RRID:AB_2286686
NANOG antibody	Millipore	Cat#AB5731; RRID:AB_2267042
SSEA-1 antibody	Santa Cruz Biotech	Cat#sc-21702; RRID:AB_626918
AFP antibody (C-19)	Santa Cruz Biotech	Cat#sc-8108; RRID:AB_633815
GATA4 antibody (C-20)	Santa Cruz Biotech	Cat#sc-1237; RRID:AB_2108747
Anti-α-Actinin (Sarcomeric) antibody	Sigma	Cat#A7732; RRID:AB_2221571
GFAP antibody	Dako	Cat#Z0334; RRID:AB_10013382
TUJ-1 antibody	Covance	Cat#MMS-435P; RRID:AB_2313773
Alexa Fluor® 488 AffiniPure Goat Anti-Rabbit IgG (H+L)	Jackson ImmunoResearch	Cat#111-545-144; RRID:AB_2338052
Alexa Fluor® 488 AffiniPure Goat Anti-Mouse IgG (H+L)	Jackson ImmunoResearch	Cat#115-545-003; RRID:AB_2338840
Alexa Fluor® 488 AffiniPure Mouse Anti-Goat IgG (H+L)	Jackson ImmunoResearch	Cat#205-545-108; RRID:AB_2339072
Chemicals, Peptides, and Recombinant Proteins		
MspI(20,000 units/ml)	New England Biolabs	Cat#R0106S
10x NEB buffer 2	New England Biolabs	Cat#B7002S
Klenow Fragment (3′→5′ exo-)	New England Biolabs	Cat#M0212L
T4 DNA Ligase (2,000,000 units/ml)	New England Biolabs	Cat#M0202M
AgencourtAMPure XP - 60 ml	Beckman Coulter	Cat#A63881
Pfu Turbo CxHotstart DNA Polymerase (500 units)	Agilent	Cat#600412
DNA/RNA Shield (2X Concentrate)	Zymo Research	Cat#R1200-25
Hoechst 33342	Life Technologies	Cat#H3570
Critical Commercial Assays		
Quick-gDNA Blood MiniPrep kit	Zymo Research	Cat#D3072
Quick-DNA Universal kit	Zymo Research	Cat#D4069
Genomic DNA Clean & Concentrator-10	ZymoReserach	Cat#D4011
Kapa Complete Universal Kit	KAPA Biosystems	Cat#KK4824
EZ DNA Methylation-Gold Kit	Zymo Research	Cat#D5006
Primers from TruSeq Nano LT Kits (Sets A and B)	Illumina	Cat#15041757 and 15041759
Alkaline Phosphatase Staining Kit	Stemgent	Cat#00-0055
Deposited Data		
Raw and analyzed data	This paper	GEO: GSE80672
Mouse reference genome NCBI build 38, GRCm38.p2	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/mouse/
Location of CpG islands for mouse reference genome	UCSC	http://hgdownload.cse.ucsc.edu/goldenPath/mm10/database/
Ensemble annotated mouse genome features, release 82	Ensembl	http://useast.ensembl.org/Mus_musculus/Info/Index
Location of predicted enhancers in mouse genome	Yue et al., 2014	N/A
Experimental Models: Cell Lines		
Primary fibroblasts from kidney of C57BL/6 male mice	This paper	N/A
Primary fibroblasts from lung of C57BL/6 male mice	This paper	N/A
Mouse iPSCs from kidney fibroblasts	This paper	N/A
Mouse iPSCs from lung fibroblasts	This paper	N/A
HEK293T cells	ATCC	CRL-3216

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
Mouse: C57BL/6	Aged Rodent Colony from National Institute on Aging	N/A
Mouse: B6D2F1	Aged Rodent Colony from National Institute on Aging	N/A
Mouse: GHR KO ([C57BL/6J x BALB/cByJ]/F2	Laboratory of Richard Miller (Coschigano et al., 2003)	N/A
Mouse: Snell dwarf ([DW/J x C3H/HEJ]/F2)	Laboratory of Richard Miller (Flurkey et al., 2001)	N/A
Oligonucleotides		
Primer: Illumina TruSeqprimer Forward: AATGATACGGC GACCACCGAGAT	This paper	N/A
Primer: Illumina TruSeqprimer Reverse: CAAGCAGAAG ACGGCATACGA	This paper	N/A
Recombinant DNA		
Plasmid: TetO-FUW-OSKM	Carey et al., 2009	Addgene Plasmid #20321
Plasmid: FUW-M2rtTA	Hockemeyer et al., 2008	Addgene Plasmid #20342
Software and Algorithms		
Testing quality of high throughput sequence libraries: FastQC v.0.10.1	N/A	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Adaptor removing and quality trimming: Trim Galore! v.0.4.0	N/A	http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Detection of methylation sites: Bismark v.0.14.5	N/A	http://www.bioinformatics.babraham.ac.uk/projects/bismark/
Dimensionality reduction by singular value decomposition: MATLAB	Antoulas and Sorensen (2001)	https://www.mathworks.com/help/matlab/ref/svds.html
Elastic net regression: MATLAB	Friedman et al. (2010)	https://www.mathworks.com/help/stats/lassoglm.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents should be directed to and will be fulfilled by the Lead Contact, Vadim Gladyshev (vgladyshev@rics.bwh.harvard.edu).

METHOD DETAILS

Mouse Models

All mice used in this study, with the exception of Snell dwarf, GHR-KO models and 3-5-week-old C57BL/6 mice, were obtained from the NIA Aged Rodent Colony (Table 1). The oldest mice obtained from NIA were 32 months; to obtain mice 34 and 35 months of age, 32-month old mice were aged at Brigham & Women's Hospital (BWH) for 2 and 3 months, respectively. Three of the six 35-month-old samples were from a different facility and denoted with an "R" (Table S1, related to Table 1). With the exception of mice at the 34 and 35 month time points, blood from all mice used for the clock was collected within 2 weeks of arrival from NIA. Snell dwarf and GHR-KO models and their controls were housed at the University of Michigan. The 3-5-week-old C57BL/6 mice were bred at BWH, Harvard Medical School, from parents obtained from NIA. Calorie restriction started at 14 weeks of age and continued until the time when the animals were sacrificed. For all other animals, food was provided ad libitum. Blood was isolated from the inferior vena cavae, with the exception of 3-5-week-old mice and Snell dwarf and GHR-KO mice (and their corresponding controls), in which blood was isolated directly from heart. The blood was immediately mixed with EDTA, and 100 μl was placed in 400 μl of lysis buffer (Zymo). In the case of Snell dwarf and GHR-KO mice, including the corresponding controls, 100 μl EDTA mixed blood was placed in 100 μl of DNA/RNA Shield (2X concentrate) (Zymo).

Genomic DNA Purification

DNA from samples in lysis buffer was purified by using Quick-gDNA Blood MiniPrep kit (Zymo) using manufacturer's instructions. DNA from samples in DNA/RNA 2X concentrate was purified by using Quick-DNA Universal kit (Zymo). DNA was eluted from columns

in 100 μl of 10 mM Tris-HCl buffer, pH 8.0. Then, 2 μl of RNase A (Life Technologies) was added to each sample. Samples were incubated at room temperature for 2 min, and DNA was prepared by using Genomic DNA Clean & Concentrator-10 (Zymo). DNA was eluted in 25 μl of TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) and quantified using a Qubit 2.0 (Life Technologies).

RRBS

Reduced Representation Bisulfite Sequencing (RRBS) was performed as previously reported with few modifications (Boyle et al., 2012; Gu et al., 2011). MspI digestions were set up by adding 100 ng of DNA sample to 1 μl of MspI (20 U/μl) (New England Biolabs), 3 μl of 10x NEB buffer 2 (New England Biolabs), and nuclease-free water to obtain a reaction volume of 30 μl. Reactions were incubated at 37°C for 17 hr. Then, 1 μl of Klenow fragment (3′-5′ exo-) (New England Biolabs) and 1 μl of dNTP mixture containing 10 mM dATP, 1 mM dCTP, and 1 mM dGTP were added to each sample for DNA end repair and A-tailing. The reactions were performed in a thermocycler without a heated lid. The program was set to 30°C for 20 min, 37°C for 20 min and 65°C for 20 min. Subsequently, ligation reactions were set up by adding 4.8 μl of T4 Ligation buffer, 1.6 μl of T4 DNA Ligase (New England Biolabs), and 3.2 of μl TruSeq Nano DNA LT adaptor from Illumina that was diluted 1:20, and 6.4 μl of nuclease-free water to the DNA samples. We used 6-7 different indexed adaptors per library. The reactions were incubated at 16°C for 16 hr.

After ligation, 1.2x SPRI bead clean-up was performed on the samples by adding 52 μl of water and 120 μl of Agencourt AMPure XP magnetic beads (Beckman Coulter). The mixture was pipetted 10 times and incubated at room temperature for 5 min. The samples were centrifuged and placed in 96-side magnet for 10 min. All liquid was removed without disturbing the bead pellets. Three washes were performed by adding 200 μl 80% ethanol to each sample still in the magnet, letting the samples sit at room temperature for 30 s, and removing ethanol without disturbing the beads. After the final wash, beads were dried at room temperature for 15-20 min. The samples were removed from the magnet and 40 μl of 10 mM Tris-HCl buffer was added to elute the DNA. Beads were then resuspended and samples were placed on the magnet for 10 min incubation at room temperature. Finally, the supernatant containing the eluted DNA was removed from the beads without disturbing them.

The purified DNA was quantified by qPCR using the Kapa Complete Universal Kit (KAPA Biosystems) following the manufacturer's instructions. Samples were pooled in libraries of 6 to 7. Subsequently, the libraries underwent bisulfite conversion using the EZ DNA Methylation-Gold Kit (Zymo) and eluted in 40 μl of buffer provided by the kit.

The bisulfite-converted samples were amplified by setting up 200 μl reactions consisting of 30 μl of library sample, 4 μl of Pfu Turbo Cx Hotstart DNA Polymerase, 2.5 U/μl (Agilent), 20 μl of 10X buffer (Agilent), 2 μl of dNTP mix (25 mM each, 100 mM total), 32 μl of Illumina TruSeq primer mix (2.5 μM each, 5.0 μM total), and 112 μl of nuclease-free water. The 200 μl reactions were then divided into aliquots of 50 μl and ran on a thermocycler as follows: Step 1: 95°C for 2 min; Step 2: for the least amount of cycles required for final amplification at 95°C for 30 s, 65°C for 30 s, and 72°C for 60 s; Step 3: 72° for 7 min; Step 4: hold at 4°C. To determine the least amount of cycles needed, 10 μl test reactions (1 μl of library sample, 0.2 μl of Pfu Turbo Cx Hotstart DNA Polymerase, 2.5 U/μl (Agilent), 1 μl of 10X buffer (Agilent), 0.1 μl of dNTP mix (25 mM each, 100 mM total), 1.6 μl of Illumina TruSeq primer mix (2.5 μM each, 5.0 μM total), and 6.1 of μl nuclease-free water) were performed using 9, 11, and 13 cycles. After the final PCR amplification, each individual sample was pooled into a 1.5 mL tube. A 1.2x SPRI bead clean-up was performed on the samples by adding 240 μl of Agencourt AMPure XP magnetic beads (Beckman Coulter, Inc.). The mixture was pipetted 10 times and incubated at room temperature for 5 min. Then, the samples were centrifuged and placed on a magnetic rack for 10 min. All liquid was removed without disturbing the bead pellets. Two washes were performed by adding 1 mL of 80% ethanol still in the magnet, letting the sample sit at room temperature for 30 s, and removing ethanol from the sample without disturbing the beads. After the final wash, beads were dried at room temperature for 15-20 min. The samples were removed from the magnet and 40 μl of 10 mM Tris-HCl buffer was added to each sample to elute the DNA. Beads were re-suspended and samples were placed on the magnet for 10 min at room temperature. Finally, the supernatant containing the eluted DNA was removed from the beads without disturbing them. Libraries were sequenced on the Illumina HiSeq 2500 platform using 75 paired-end sequencing with v4 reagents. Since RRBS libraries generate low complexity libraries, the samples were spiked with ~10% phiX.

Equipment required for the construction of RRBS libraries included a qPCR machine, thermal cycler, NanoDrop spectrophotometer or Qubit fluorimeter with associated reagents, a DynaMag-2 magnet (Life Technologies), Bioanalyzer with associated reagents or TapeStation with High Sensitivity D1000 ScreenTapes (Agilent Technologies), and a DyanMag 96-side magnet (Life Technologies).

Primary Fibroblasts and Generation of Mouse iPSCs

Primary fibroblasts were prepared from lungs and kidneys of 10-week-old C57BL/6 male mice. Cells were maintained in DMEM supplemented with 10% fetal bovine serum (FBS), 1% nonessential amino acids, 2 mM glutamine and antibiotics. Lentiviral vectors 4F2A and M2rtTA (Addgene plasmids #20321 and #20342) were used to reprogram fibroblasts (Carey et al., 2009). 48 hr after transfection of HEK293T cells, supernatants containing viral particles were harvested, filtered, and directly used for transduction of fibroblasts. Infected fibroblasts were seeded onto a feeder layer in ESC medium in the presence of 2 mg/ml doxycycline (DMEM supplemented with 15% FBS, 2 × 10^6 units leukemia inhibitory factor, 0.1 mM β-mercaptoethanol, 1% nonessential amino acids, 2 mM glutamine and antibiotics (Buganim et al., 2014). On day 10, iPSC colonies were picked, dissociated by trypsin digestion, and plated into new culture dishes. These iPSCs were subcultured every 2–3 days. All cells were maintained in a humidified incubator at 37°C and 5% CO_2. Testing for mycoplasma was routinely performed.

Characterization of Mouse iPSCs

Alkaline phosphatase staining was performed using the Alkaline Phosphatase Staining Kit (Stemgent) according to manufacturer's instructions. For fluorescence immunocytochemistry, iPSCs were cultured on mitotically inactivated MEFs for 2–3 days and fixed using PBS containing 4% PFA (Sigma) for 15 min at room temperature. After rinsing with PBS, fixed cells were blocked and permeabilised for 1 hr in PBS containing 10% (v/v) goat serum (Sigma) in 0.1% Triton X-100-PBS. Primary antibodies were as follows: OCT4 (Santa Cruz Biotech), SOX2 (Millipore), NANOG (Millipore), and SSEA-1 (Santa Cruz Biotech), which were diluted 1:100 in blocking solution. Fixed cells were incubated with the primary antibody solution overnight at 4°C. After washing out the primary antibody solution, fixed cells were incubated with secondary antibodies (labeled with Alexa 488, 1:200, Jackson ImmunoResearch) for 1 hr at room temperature. Nuclei were counterstained using 1 μg/ml Hoechst 33342 (Life Technologies). All fluorescence imaging was conducted using an LSM 700 confocal microscope (Zeiss).

Generation of Embryoid Bodies and Analysis

Colonies growing on MEFs were detached using 0.05% trypsin and grown as a suspension culture on low adherent plates using ESC medium without LIF. After 1 week of suspension growth, cells were transferred to 12- or 24-well plates coated with 0.1% gelatin and grown in DMEM supplemented with 20% FBS. Embryoid bodies were grown for 1–2 weeks prior to fixation and immunofluorescence staining. Cultures were fixed and stained as described above using the following antibodies: AFP (1:200, Santa Cruz Biotech), GATA4 (1:200, Santa Cruz Biotech), α-actinin (1:200, Sigma), GFAP (1:500, Dako), and TUJ-1(1:500, Covance).

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of Sequence Reads

Mouse genome assembly GRCm38.p2 and its annotation were downloaded from NCBI (ftp.ncbi.nih.gov). Locations of CpG islands for mm10 (Genome Reference Consortium Mouse Build 38) were obtained from UCSC (http://hgdownload.soe.ucsc.edu, released on 07-Mar-2012). CpG island shores were defined as regions 2 kp upstream and downstream of islands. Ensemble annotated features (release 82, ftp.ensemble.org) were utilized for the purpose of identification of promoters. Location of enhancers was based on a list of previously predicted enhancers (Yue et al., 2014). The quality of high throughput sequence libraries was verified using "FastQC v.0.10.1" package (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adaptor removal and quality trimming were performed using Trim Galore! v.0.4.0. The TrimGalore tool (www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used with settings optimized for RRBS. Methylation sites were detected using Bismark v.0.14.5 according to the program's manual. During postprocessing, the identified methylation sites present in samples were extracted using R procedures and custom Perl scripts and used for further analyses.

Estimating the Leading Methylation Signature and Its Age-Dependent Pattern

The leading methylation signature MSc_{total} was estimated by performing the dimensionality reduction (Antoulas and Sorensen, 2001) of the whole blood partial DNA methylomes of 141 C57BL/6 mice as follows:

All DNA methylomes/samples were combined into a single $(N\times M)$ dataset X, where N is the number of samples and M is the number of CpG sites, covered in each single observation ($\approx$ 1.9M), keeping the rows of the matrix X chronologically ordered, thee non-negative matrix factorization of the dataset X was performed:

$$X_{ij} = \sum_{\alpha} P_i^{(\alpha)} Q_j^{(a)}, \tag{1}$$

(Applying non-negative matrix factorization to DNA methylomes as observables of interest (methylation levels) are always manifestly non-negative, and the methylation signatures should encode this.)

The dimensionality reduction of the dataset X was performed, the result of the procedure approximating X by a rank 1 matrix

$$X_{ij}^{(\mathrm{red})} = P_i^{(1)} Q_j^{(1)}. \tag{2}$$

(The "low-rank" approximation above is valid in the sense of ratio of the Frobenius norms of the matrices X and $X^{(red)}$ being of the order 1. Namely, for the dataset of 141 DNA methylomes of C57BL/6 mice it was found that $\left(\left(||X - X^{(red)}|| \right) / \left(||X|| \right) \right) \approx 0.127$, where $||X||$ denotes the Frobenius norm of a matrix X.)

The age-dependent methylation signature MSc_{total} was defined as

$$MSc_{total} = Q^{(1)}. \tag{3}$$

The weights, which different CpG sites contribute to the leading age-dependent signature, were defined according to the prescription

$$Weight_i = \frac{P_i^{(1)}}{|P^{(1)}|}. \tag{4}$$

We found a very large number of sites contributing to the leading age-dependent signature represented in Figure 1B. While no sites had weights larger than $1.4 \cdot 10^{-3}$, ~194,000 CpG sites had weights larger than $1.3 \cdot 10^{-3}$, ~450,000 CpG sites weights larger than $1.2 \cdot 10^{-3}$, and ~570,000 CpG sites weights larger than $1.0 \cdot 10^{-3}$, confirming that the age-related hypomethylation pattern is a characteristics of the mouse DNA methylome as a whole. The non-negative matrix factorization was performed in MATLAB using nnmf command.

To avoid possible confusion, it has to be noted that the procedure described here is different from the principal component analysis (PCA) in an important respect: the first leading NNMF component of the data does not encode the variance of the data and instead roughly corresponds to the mean of the data. The mean of the data was not subtracted before performing non-negative matrix factorization of the data; the latter was performed to check whether a good low-rank approximation of the data matrix is possible and to find such an approximation.

Controlling for Batch Effect

In order to minimize the possible batch effect on the behavior of age-dependent methylation signatures, all samples used to construct the mDNAm clock (144 C57BL/6 males) were divided into 5 groups (sequenced by individual flow cells; Table S2, related to Table 1), with each group including samples from mice of different ages - from 3 to 35 months. Each flow cell had 8 sequencing lanes (columns in Table S2), and all sequencing lanes in any of the flow cells had samples representing different age groups. Other considered confounding variables were the adaptor numbers and library numbers (Table S2). The values of other confounding variables (such as a researcher preparing samples) were the same for all samples used to build the clock and could not contribute to a possible batch effect. The adaptor numbers were uncorrelated with the age group (correlation coefficient 0.0398, $p = 0.6394$) and accounting for the corresponding confounding variable did not contribute to the clock structure. By construction, library numbers had a high correlation with the flow cell number (correlation coefficient 0.9405, $p = 10^{-67}$) and thus together were reduced to a single independent confounding variable. Although samples of different ages were randomized, there was ~30% redundant anticorrelation between the age-group and the flow cell number, so the possibility of the batch effect due to this correlation was studied further, and its significance was estimated as follows.

First, the PCA analysis of the data matrix X_{ij} was performed. It was found that the leading PCA component (PCA_1) explains about 6% of the total variance in the data, the second one (PCA_2) contributes slightly more than 5% to the total variance, the 3rd - about 3%, etc. (Figure S2C, related to Figure 1B). The contribution of the first PCA component was then subtracted from the data, and the age-dependent behavior of the MSc_{total} score constructed using the remaining part of the data was found to be unchanged (Figure S2D, related to Figure 1B).

Second, the batch effect analysis analogous to the one described in (Johnson et al., 2007) was performed. Namely, the linear model establishing the correlation between DNA methylation data, the batch (flow cell) number and the age was constructed, assuming both additive batch contribution and the dependence of the error variance on the batch number. Again, it was found that the batch effect in the obtained data is negligible (Figure S2E; related to Figure 1B).

Finally, a linear mixed-effects model ($Age = \beta \cdot M + Z \cdot b + \epsilon$, where ϵ is the observation error vector) was constructed with methylation data as the fixed-effects variables M and flow cell numbers, adaptor numbers and library numbers as random-effects (confounding) variables Z. The vector of weights β was then compared with a similar vector β_{bare} constructed using a liner model $Age = \beta_{bare} \cdot M + \epsilon$ not taking into account the effect of confounding variables. It was found that the vectors β and β_{bare} correlated perfectly with each other (correlation coefficient 0.999, $\sim 10^{-97}$; Figure S2F).

Constructing the Elastic Net Regression Clock and Estimating the Methylation Age of Samples

The elastic net regression clock describing a systematic age-dependent change in whole blood partial DNA methylomes of C57BL/6 males was defined as a vector w of weights of CpG sites contributing to the clock. The clock DNAm score at the chronological age of the sample t was then given by

$$MSc(Age) = w_1 \cdot M_1(Age) + w_2 \cdot M_2(Age) + \ldots, \tag{5}$$

where $M_i(Age)$ is the DNAm level on a CpG site i contributing to the clock, estimated at the chronological age Age.

The vector w characterizing the clock was constructed as follows.

1. The $m = 141$ DNA methylome samples were combined into a single $N \times M$ matrix X, where the index M corresponds to the total number $(\approx 1.9\, M)$ of CpG sites of the overlap between different samples.

2. Numerical minimization of the target function

$$T = ||w_0 + w \cdot M^T - Age|| + \frac{\lambda}{2}\left(\frac{1}{2}(w \cdot w) + ||w||_1\right) =$$

$$= \frac{1}{2N}\sum_{i=1}^{N}(w \cdot M_i + w_0 - Age)^2 + \frac{1}{2}\sum_{j=1}^{n}\left(\frac{1}{2}w_j^2 + |w_j|\right),$$

(where N is the number of observations and n is the number of predictors) with different values of the Tikhonov regularization parameter λ was then performed with 20-fold cross-validation of the identified minimum. As usual, an k-fold cross-validation represented

a random separation of the original dataset into k subsets, with $k-1$ subsets used to train the clock and the remaining subset to validate it; the procedure was then repeated k times, the results of elastic net regression were averaged out over k repeats, and the deviation error D (cross-validation deviance) was estimated (Figure 2A).

3. The cross-validation deviance (as well as the mean square error in its estimation) was evaluated for every scanned value of λ, and the value of $\lambda = \lambda_{min(D)}$ corresponding to the minimum of cross-validation deviance was identified (Figure 2A).

4. To minimize the possibility of overfitting without much loss of predictability the value of $\lambda = \lambda_{1\sigma} = 33.1269$ corresponding to a 1σ deviation from $\lambda = \lambda_{min(D)}$ was chosen. The age estimation by the clock was performed as follows. Behavior of the DNAm score on the chronological age was approximated $(R^2 = 0.9958)$ by the power law

$$MSc = a \cdot Age^b + c, \tag{6}$$

where $a = 0.1666 \pm 0.1024$, $b = 0.4185 \pm 0.0767$ and $c = -1.712 \pm 0.3010$. The constants $a,\ b,\ c$ were found by performing the mean square regression of the DNAm scores of the actual samples calculated using the Equation (5) to the power law (6) (a nonlinear least-squares fit was performed using trust-region-reflective algorithm employed in MATLAB).

The Age_{Met} of the samples was then calculated by inverting the functional dependence (6) as

$$Age_{met} = \left(\frac{MSc - c}{a}\right)^{1/b} \pm \left(\frac{MSc - c}{a}\right)^{\frac{1}{b}-1} \frac{1}{b} \Delta(MSc), \tag{7}$$

where $\Delta(MSc)$ is the estimation error for the DNA methylation score.

It was found that the methylation clock corresponding to the minimum deviance is determined by methylation levels on 108 CpG sites (Figure 2A), while the more robust (subject to lesser overfitting) clock corresponding to the choice $\lambda = \lambda_{1\sigma}$ is determined by methylation levels on 90 CpG sites homogeneously distributed across the genome (Figure 2E). The elastic regression clock was constructed in MATLAB using the command lassoglm (MethylationLevels, ChronologicalAge, 'poisson', 'CV', 20, 'Alpha', 0.5), where the third argument implied a Poisson distribution of nonsystematic variation of responses, the fourth argument - 20-fold cross-validation and the 6th argument enforced the form of the target function described above. The choice of α, which we used, corresponds to the equal weights of lasso (pure L^1) and ridge (pure L^2) optimizations. The lassoglm command uses the coordinate descent algorithm (Friedman et al., 2010); in our case, the default value 10^{-4} for the convergence threshold was used.

Statistical Analysis of Data Represented on Figures 1, 2, and 3

Application of standard t tests for estimation of the effects of longevity interventions on the methylation age was validated by the distributions of methylation levels on individual CpG sites, which exhibited a distinct quasi-Gaussian peak. As the mDNAm clock is defined as a linear combination of methylation levels on $n \gg 1$ sites, the stochastic variable defining the value of mDNAm age, subject to the central limit theorem, is normally distributed with nonzero mean.

Figure 1: (B) Dots represent individual samples, solid dots connected by a line - mean for a cohort of particular age, and blue shaded region - s.d. (C) Error bars are s.d. of the DNAm score from the mean over the cohort of particular age.

Figure 2: (B) Dots correspond to individual samples, the dark red solid line - to mean mDNAm age for cohort with a particular chronological age, and shared region - to the behavior of s.d.

Figure 3: For all boxplots the center line is the median, the box boundaries are 25% and 75% percentile, and whiskers extend to the extreme data points, which are not considered as outliers. Outliers are plotted as individual points. (B) GHR-KO: 9 males and 6 females combined into a single dataset, WT: 8 males and 3 females combined into a single dataset. (C) SD: 6 males and 4 females combined into a single dataset, WT: 7 males and 5 females combined into a single dataset. (D) For B6D2F1 strain, the latter included 20- and 27-month-old male mice (5 in each cohort), for C57BL/6 strain – 20-, 22-, 26- and 28-month-old mice (9 males in each cohort except the one aged 26 months, which included 10 mice). (E) The CR cohorts included 20- and 27-month-old male mice (5 in each cohort), while the AL cohorts – 21- and 27-month-old mice (7 and 5 males in the corresponding cohort). (F) Fibroblasts were collected from 3 C57BL/6 males.

DATA AND SOFTWARE AVAILABILITY

The data obtained in this study were deposited to GEO under accession number GEO: GSE80672.

Cell Metabolism, Volume *25*

Supplemental Information

Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions

Daniel A. Petkovich, Dmitriy I. Podolskiy, Alexei V. Lobanov, Sang-Goo Lee, Richard A. Miller, and Vadim N. Gladyshev

Supplemental Information for

Using DNA methylation profiling to evaluate biological age and longevity interventions

Daniel A. Petkovich, Dmitriy I. Podolskiy, Alexei V. Lobanov, Sang-Goo Lee, Richard A. Miller & Vadim N. Gladyshev

Lead Contact: vgladyshev@rics.bwh.harvard.edu

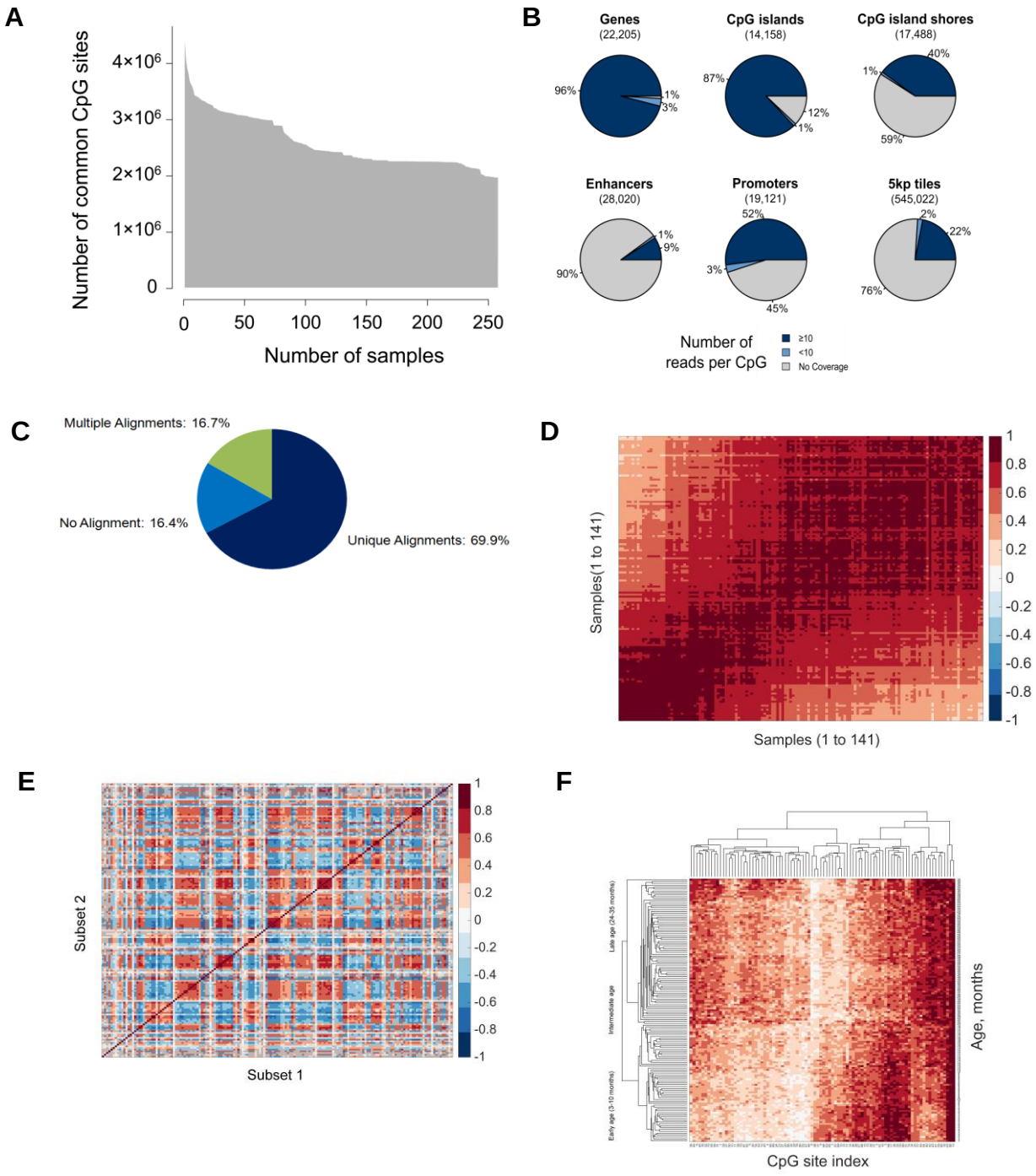


Figure S1. Details of the RRBS procedure and the identified clock sites. Related to Fig. 1A, 2C and Table 1. (A) The RRBS procedure targets similar regions of the mouse genome. This panel shows the number of CpG sites common among the samples as a function of the number of samples used. We observed a significant overlap in the sequences targeted by RRBS, with almost 2 million sites being common to all samples. (B) CpG coverage across genomic features. This panel shows sequence coverage for various genomic features based on CpG sites found in all samples. A feature was

considered covered if a CpG site seen in all samples was located in the respective feature. Numbers in parentheses show the number of features annotated in the mouse genome, which represent 100% in each category. The box shows the color scheme used for categories: ≥10-fold coverage, <10 fold coverage and no coverage. A list of predicted enhancers (see Star Methods) was used for annotation of CpG sites within enhancers. Promoters were defined as regions 1 kb downstream and upstream of the transcription start sites of Ensembl genes. 5 kb tiles were defined 5 kb non-overlapping consecutive tiles that divide the genome. (C) Alignment of RRBS reads to the mouse genome. Statistics for the paired-end sequence dataset (C57BL/6 male mice, 141 samples) is shown. Almost 70% of reads aligned uniquely, with 17% reads producing multiple alignments using Bismark. (D) Degree of similarity between different samples. Correlation between 141 samples used to construct the main clock. Only top 160 CpG sites contributing to different versions of mDNAm clocks built using Subsets 1 and 2 are used as observables. (E) Correlation between the pairs of top 160 CpG sites contributing to different versions of the clock built using Subsets 1 and 2. Most of the sites show strong correlation or anticorrelation between each other. (F) Clustered heat map of methylation levels on 90 CpG sites used to construct the mDNAm clock.

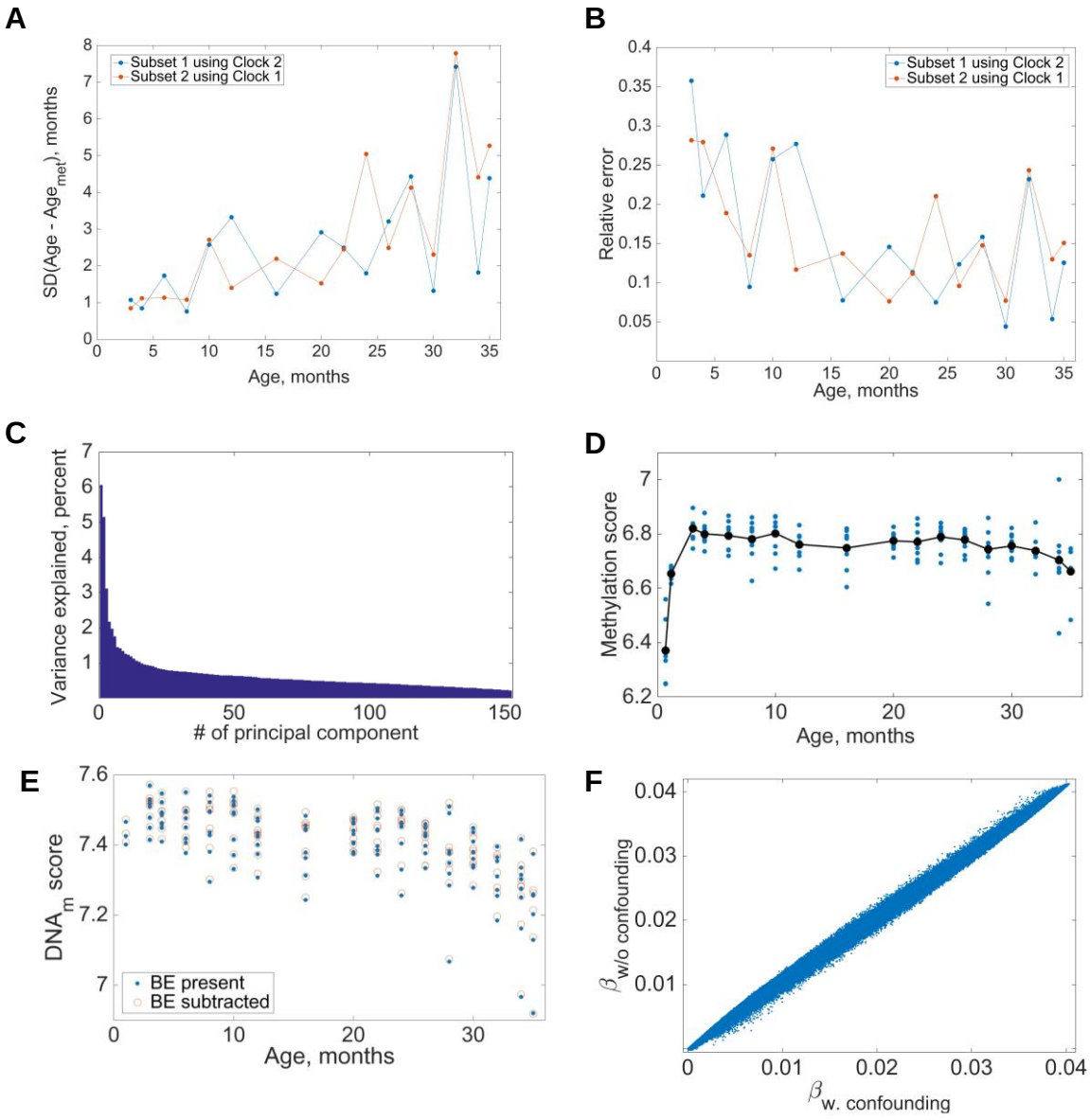


Figure S2. Testing for batch effect and error in estimating DNA methylation age. Related to Fig. 1C,D. (A) Behavior of the DNA methylation clock error with age. Error is defined as a square root of the mean squared difference between the detected DNA methylation age and the chronological age of a sample. Blue curve corresponds to the DNA methylation age of samples from Subset 1 as determined using the clock built from Subset 2 data. Orange curve corresponds to the DNA methylation age of samples from Subset 2 as determined using the clock built from Subset 1 data. (B) Behavior of relative DNA methylation clock error with age. Relative error is defined as a square root of the mean squared difference between the detected DNA methylation age and the chronological age of a sample, divided by the chronological age of the sample. Blue and red curves are defined as in (A). (C) Relative weights of contributions of different PCA components of the data matrix to the total variance of the data. The 1st PCA component of the data matrix $\boldsymbol{X_{ij}}$ explains about 6% of data variance, the 2nd PCA component -

about 5%, the 3rd PCA component - about 3%, etc. (D) Behavior of the MSc_{total} score with age after the contribution of the 1st principal component is subtracted from the data. (E) Behavior of the MSc_{total} score before and after the possible batch effect (BE subtracted) contribution is subtracted from the data. (F) Clock weights found by constructing linear models taking into account confounding variables ($\beta_{w.confounding}$) and ignoring them ($\beta_{w/o\ confounding}$). Weight vectors well correlate with each other (correlation coefficient 0.999, $p \sim \mathbf{10^{-97}}$).

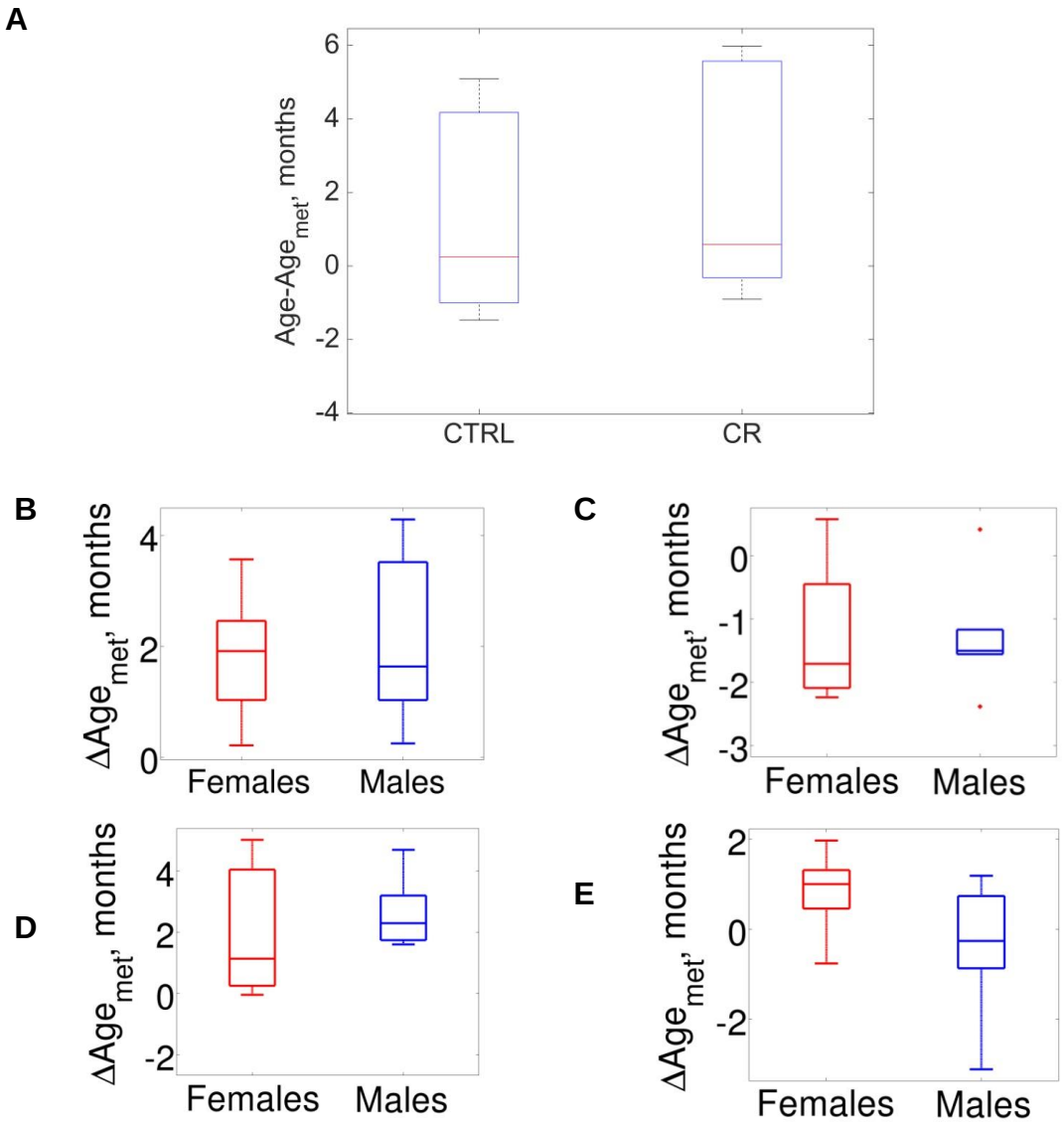


Figure S3. Additional testing of the DNA methylation clock using mouse longevity interventions. Related to Fig. 3A. (A) Short-term effect of calorie restriction. Comparison of DNA methylation ages between a control group and a calorie restricted group (6 mice each) of (DW/J x C3H/HEJ)/F2 mice. Mice were subjected to calorie restriction for 2 months starting from the age of 4 months. No significant difference in the DNA methylation age was detected ($\boldsymbol{p = 0.601}$). (B-E) The mDNAm clock developed based on male mice predicts the age of female mice. For all box plots the center line is the median, the box boundaries are 25% and 75% percentile, and whiskers extend to the extreme data points, which are not considered as outliers. Outliers are plotted as individual points. The panels show the difference $\boldsymbol{\Delta Age_{met}}$ between chronological and

mDNAm age for male and female mice. (B) Littermate control strain ((DW/J x C3H/HEJ)/F2) for the Snell dwarf mice (7 males, 5 females). (C) Snell dwarf mice (6 males, 4 females). (D) Littermate controls (genetically heterogeneous stock: 50% C57BL/6 and 50% BALB/cByJ) for the growth hormone receptor knockout mice (8 males, 3 females). (E) Growth hormone receptor knockout mice (9 males, 6 females).

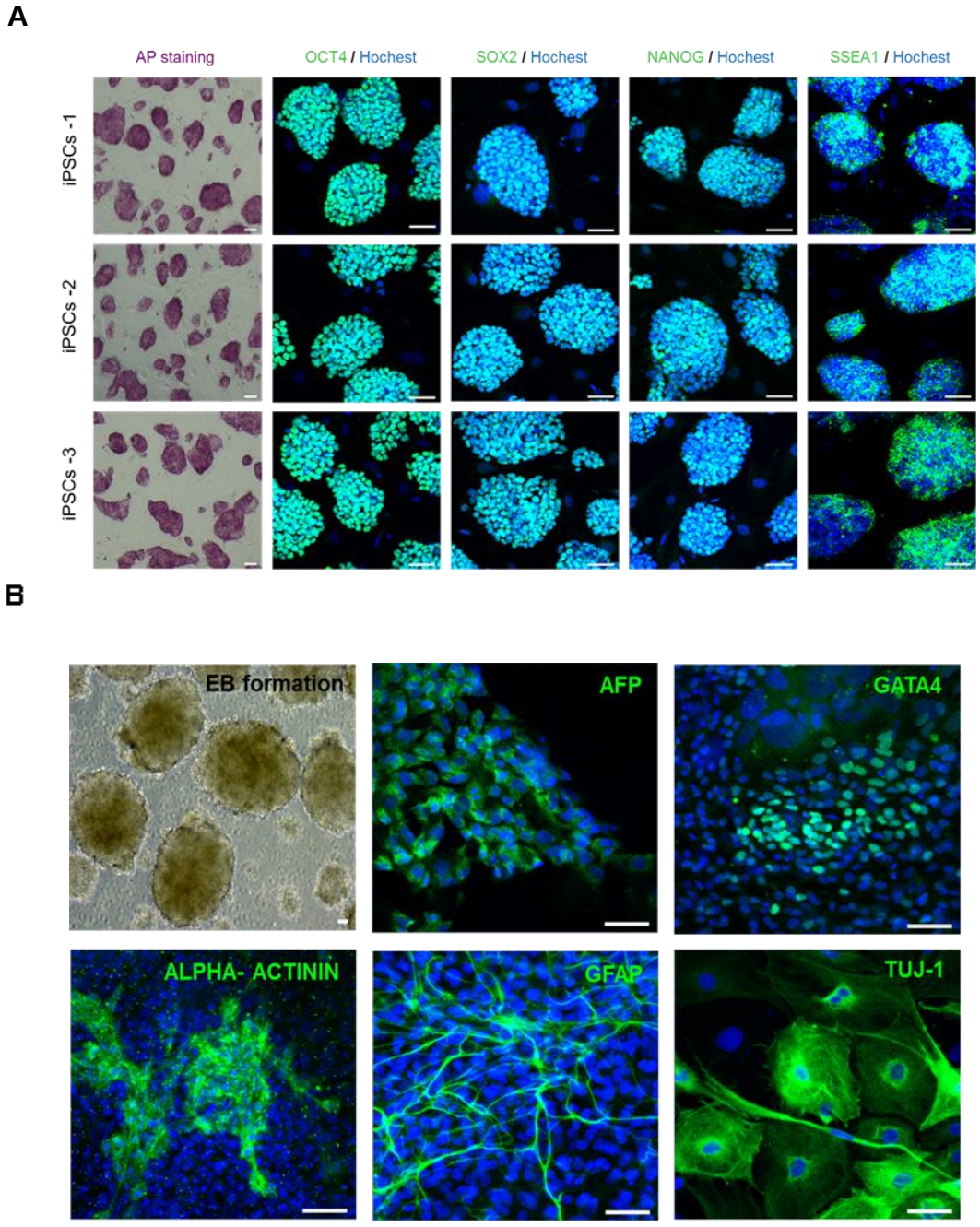


Figure S4. Generation and characterization of mouse iPSC lines. Related to STAR Methods, Sections "Primary fibroblasts and generation of mouse iPSCs" and "Characterization of mouse iPSCs", and Fig. 3F. (A) Alkaline phosphatase (AP) and immunofluorescence for the pluripotency markers OCT4, SOX2, NANOG, and SSEA-1 in mouse iPSCs. Nuclei were counterstained with Hochest33342. Scale bars, 50 um. (B) In vitro differentiation of mouse iPSCs. Embryoid body (EB) formation and immunostaining of differentiated EB with antibodies for the three germ layers as indicated. Scale bars, 50 um.