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

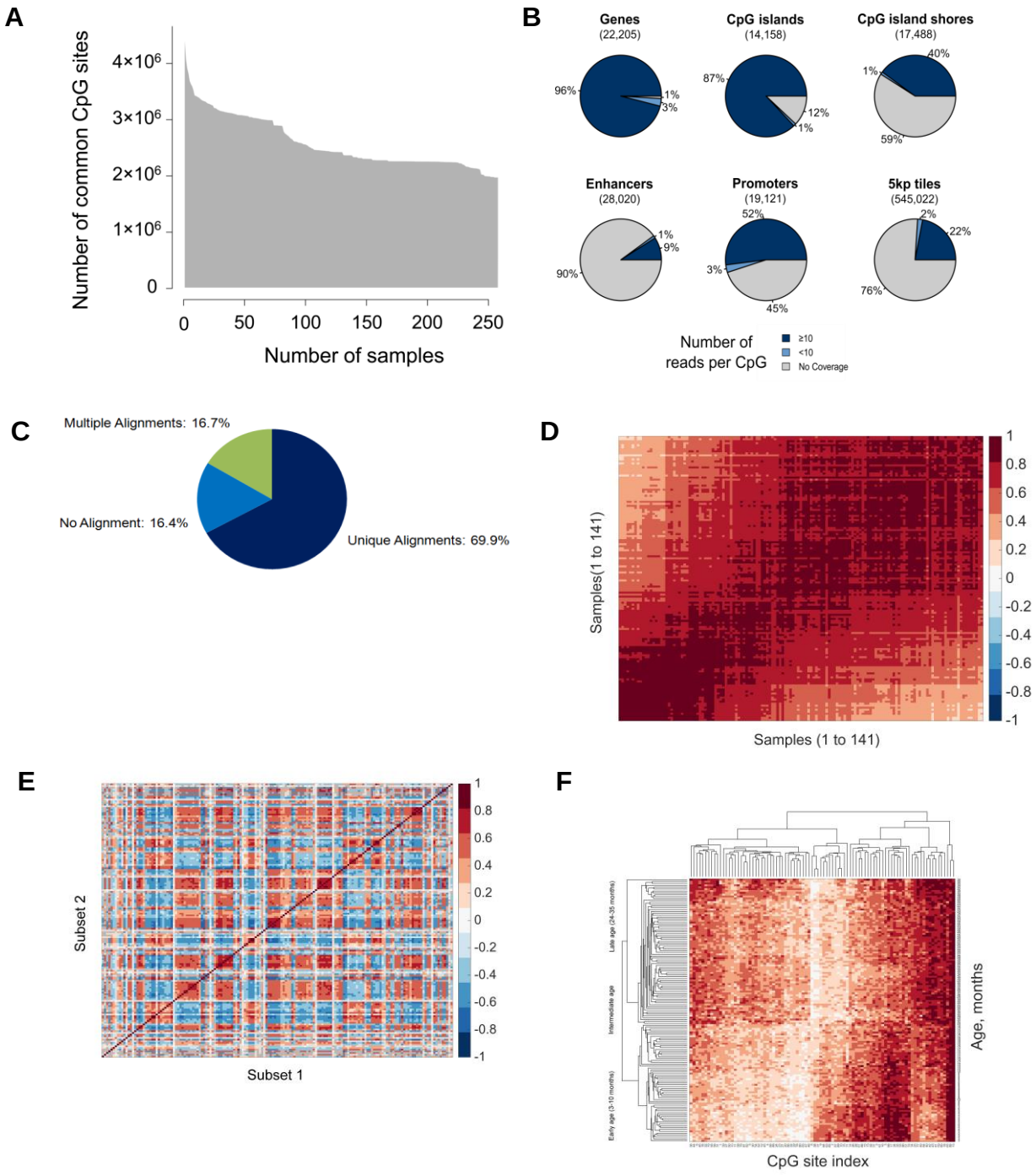

Figure S1. Details of the RRBS procedure and the identified clock sites. Related to Fig. 1A, 2C and Table 1. (A) The RRBS procedure targets similar regions of the mouse genome. This panel shows the number of CpG sites common among the samples as a function of the number of samples used. We observed a 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.

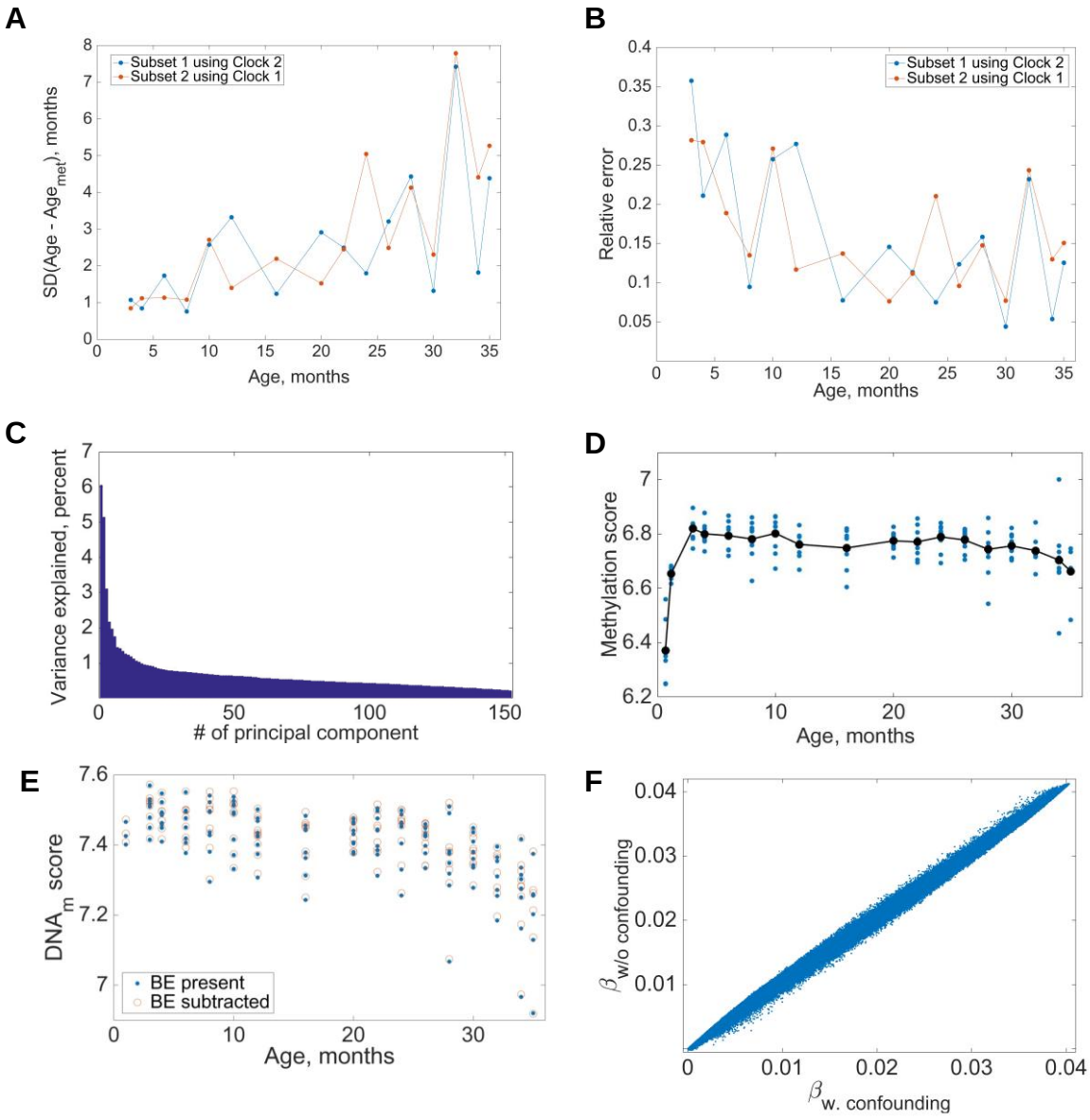

Figure S2. Testing for batch effect and error in estimating DNA methylation age. Related to Fig. 1C,D. (A) Behavior of the DNA methylation clock error with age. Error is defined as a square root of the mean squared difference between the detected DNA methylation age and the chronological age of a sample. 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, $p \sim \mathbf{10^{-97}}$).

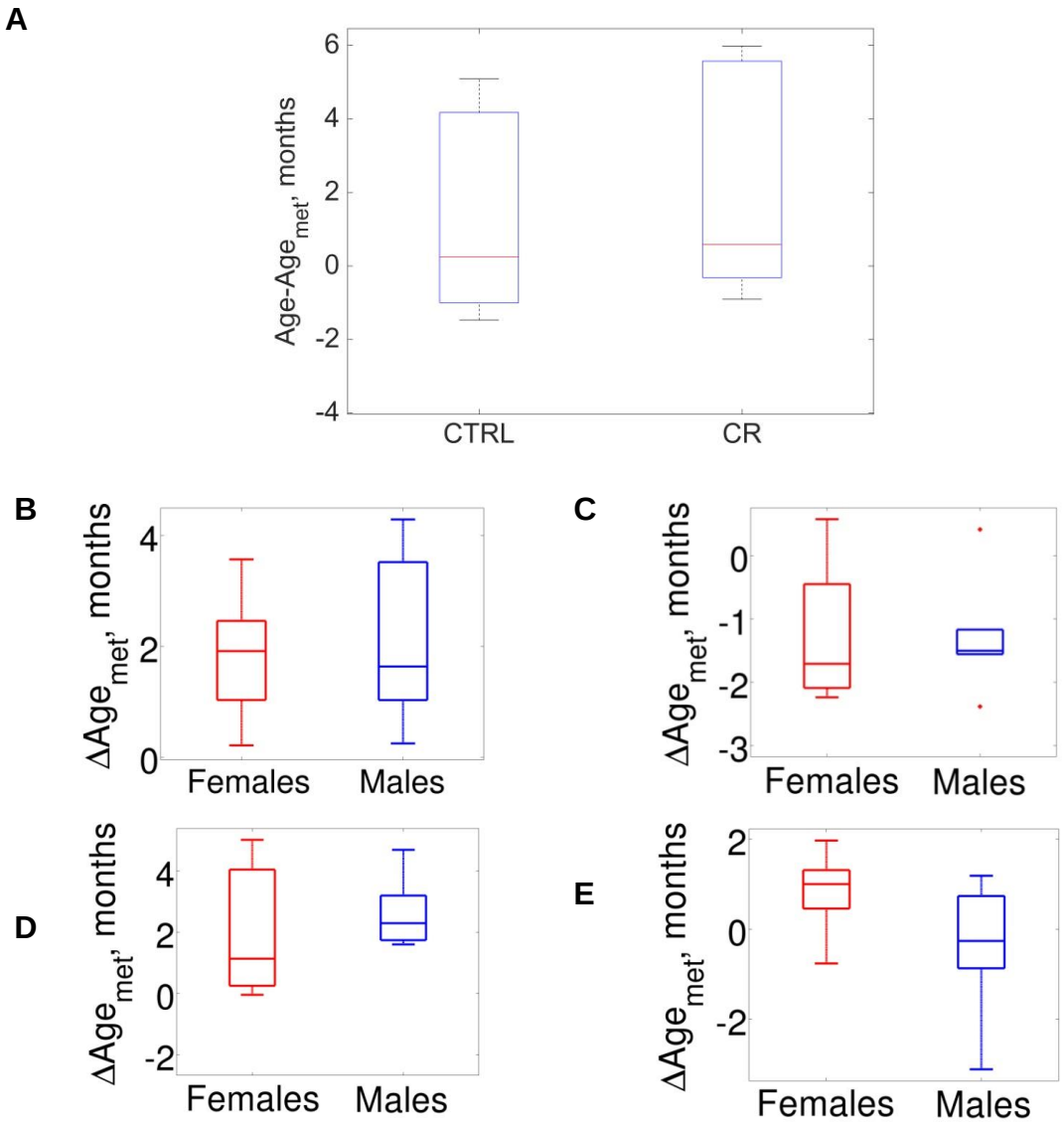

Figure S3. Additional testing of the DNA methylation clock using mouse longevity interventions. Related to Fig. 3A. (A) Short-term effect of calorie restriction. Comparison of DNA methylation ages between a control group and a calorie restricted group (6 mice each) of (DW/J x C3H/HEJ)/F2 mice. Mice were 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).

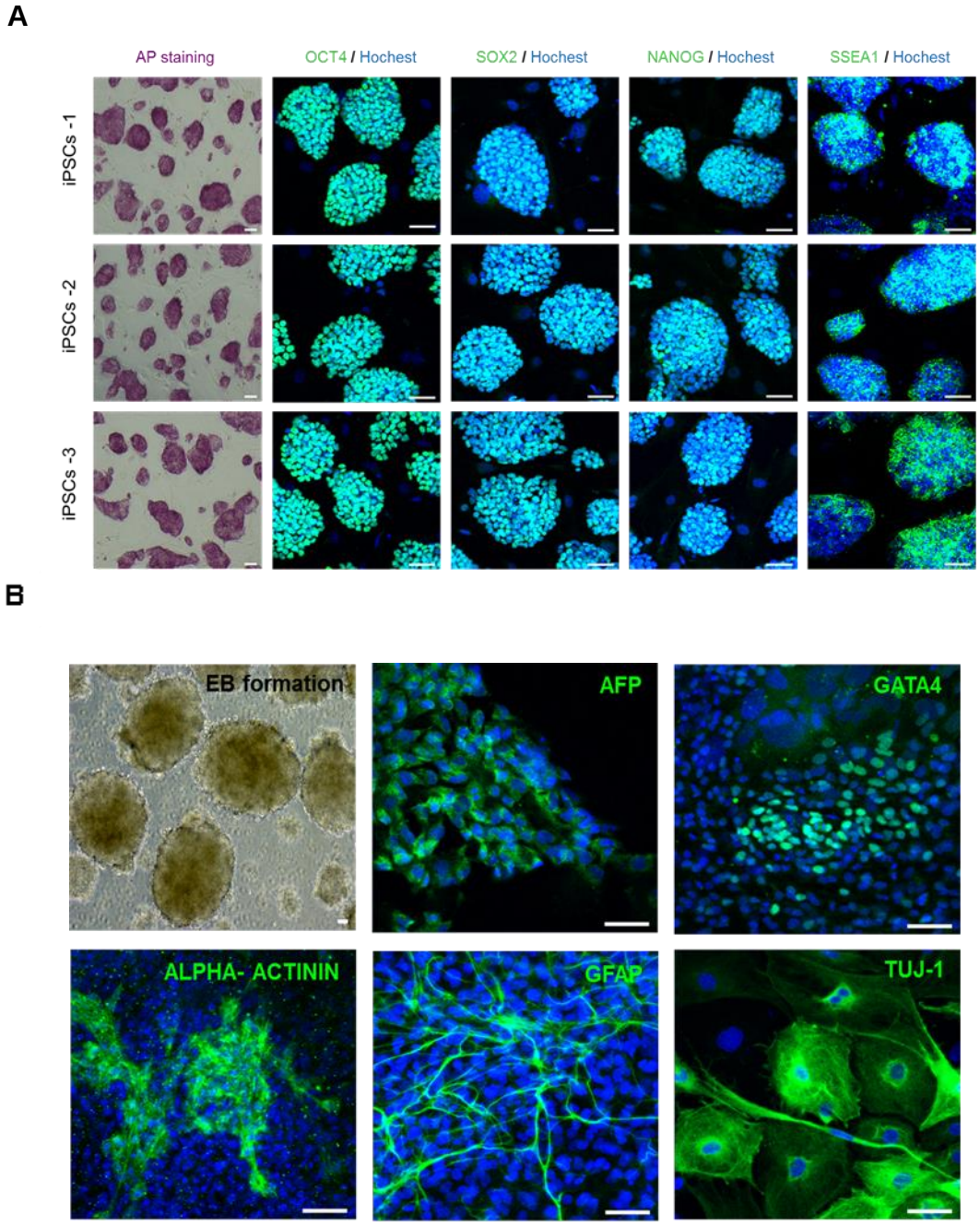

Figure S4. Generation and characterization of mouse iPSC lines. Related to STAR Methods, Sections "Primary fibroblasts and generation of mouse iPSCs" and "Characterization of mouse iPSCs", and Fig. 3F. (A) Alkaline phosphatase (AP) and immunofluorescence for the pluripotency markers OCT4, SOX2, NANOG, 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.