Ultra-cheap and scalable epigenetic age predictions with TIME-Seq

Patrick T Griffin[1], Alice E Kane[1], Alexandre Trapp[2], Jien Li[1], Maeve S McNamara[1], Margarita V Meer[2,3], Michael R MacArthur[4], Sarah J Mitchell[4], Amber L Mueller[1], Colleen Carmody[1], Daniel L Vera[1], Csaba Kerepesi[2], Nicole Noren Hooten[5], James R Mitchell[4#], Michele K Evans[5], Vadim N Gladyshev[2], David A Sinclair[1*]

Affiliations

[1]Blavatnik Institute, Dept. of Genetics, Paul F. Glenn Center for Biology of Aging Research at Harvard Medical School, Boston, MA 02115 USA

[2]Brigham and Women's Hospital, Division of Genetics, Department of Medicine, Harvard Medical School, Boston, MA, 02115 USA

[3]Yale University School of Medicine, Department of Pathology, New Haven, CT

[4] Department of Health Sciences and Technology, ETH Zurich, Zurich 8005 Switzerland.

[5]Laboratory of Epidemiology and Population Science, National Institute on Aging, National Institutes of Health, Baltimore, MD

[#]Deceased

*Corresponding author: david_sinclair@hms.harvard.edu

ABSTRACT

Epigenetic "clocks" based on DNA methylation (DNAme) are the most robust and widely employed aging biomarker. They have been built for numerous species and reflect gold-standard interventions that extend lifespan. However, conventional methods for measuring epigenetic clocks are expensive and low-throughput. Here, we describe Tagmentation-based Indexing for Methylation Sequencing (TIME-Seq) for ultra-cheap and scalable targeted methylation sequencing of epigenetic clocks and other DNAme biomarkers. Using TIME-Seq, we built and validated inexpensive epigenetic clocks based on genomic and ribosomal DNAme in hundreds of mice and human samples. We also discover it is possible to accurately predict age from extremely low-cost shallow sequencing (e.g., 10,000 reads) of TIME-Seq libraries using *scAge,* a probabilistic age-prediction algorithm originally applied to single cells. Together, these methods reduce the cost of DNAme biomarker analysis by more than two orders of magnitude, thereby expanding and democratizing their use in aging research, clinical trials, and disease diagnosis.

INTRODUCTION

Aging "clocks" are highly accurate, machine learning-derived biomarkers trained to predict age or age-associated phenotypes. Both physiological measurements[4,5] and biomolecules[6-8] have been used to develop clocks, and they are becoming increasingly common in basic research and clinical studies. However, methods to measure clocks that are inexpensive and scalable are lacking, limiting their use in large cohort studies such as clinical trials of age-related disease.

The most robust and widely used aging clocks are based on DNA cytosine methylation (DNAme) and are interchangeably referred to as DNAme clocks or epigenetic clocks. These clocks comprise sets of CpGs (usually dozens to hundreds) and corresponding algorithms that use methylation levels to predict age. Epigenetic clocks have been built for humans[9-11], mice[1,3,12], and a multitude of other mammals[13-16], and they have been shown to reflect interventions that are canonically associated with longevity[17], accelerated aging[18], and even cellular rejuvenation[19,20]. While much focus has been put on developing more accurate clocks, clocks in different organisms, or clocks trained to predict aging-adjacent phenotypes[11,21], very little work has been done to make the use of epigenetic clocks more experimentally tractable.

Epigenetic clocks have been predominantly built and assayed using methylation microarray (e.g. Illumina MethylationEPIC BeadChip[22]) or Reduced Representation Bisulfite Sequencing[23] (RRBS). While these approaches measure hundreds of thousands to millions of CpGs, making them useful for biomarker discovery, they are expensive and low throughput, costing hundreds of dollars per sample and requiring laborious and expensive DNA preparation.

To address this limitation, we developed Tagmentation-based Indexing for Methylation Sequencing (TIME-Seq) for scalable and inexpensive targeted methylation sequencing of DNAme biomarkers. We use TIME-Seq to cheaply and rapidly build epigenetic clocks from hundreds of mice and validate these clocks on independent cohorts of longitudinally tracked mice. Using a recently reported algorithm for epigenetic age analysis in single cells[24], we discover it is possible to predict age with similar accuracy as traditional clocks from shallow-sequenced TIME-Seq libraries, dramatically reducing the sequencing depth needed to predict age. Finally, we applied TIME-Seq to study ribosomal DNA methylation variation in over 600 humans and built an epigenetic clock from libraries enriched for human clock loci. By decreasing cost more than two orders of magnitude and expediting large-scale experiments, TIME-Seq enables more widespread use of DNAme clocks and other epigenetic biomarkers.

RESULTS

We designed TIME-Seq, a novel targeted sequencing method to build and measure DNAme-based biomarkers (e.g., epigenetic clocks) for low-cost in hundreds to thousands of samples. TIME-Seq leverages barcoded and bisulfite (BS)-resistant Tn5-transposomes to rapidly index sample DNA for a pooled library preparation (Fig. 1a), which streamlines large-scale experiments and minimizes the cost of consumables (Table S1). After tagmentation and pooling, methylated end-repair (5-methyl-dCTP replaces dCTP) is performed, and pools are prepared for in-solution hybridization enrichment using biotinylated-RNA baits (Fig. S1a-e). Unlike BS-compatible single-cell indexing approaches[25], we designed barcoded TIME-Seq adaptors to be short (38-nt) for optimal enrichment efficiency since longer adaptors are more likely to daisy-chain with off-target DNA[26] (Fig. S1f-h). Baits are produced in-house from single-stranded oligonucleotide libraries (Fig. S1a and Table S2), providing inexpensive enrichments from a regenerable source. After BS conversion of captured DNA and indexed-PCR amplification of each pool, Illumina short-read sequencing is performed (Fig. S2) and sample reads are demultiplexed based on pool and Tn5-adaptor indexes. From mapped reads, a matrix of methylation values for CpGs in each sample is used to train or readout a DNAme biomarker.

To assess TIME-Seq, we performed a pilot experiment using 100 nanograms (ng) of blood DNA from 12 mice across 4 age groups (Fig. 1b). Pools were enriched using baits targeting a previously described epigenetic clock in ribosomal DNA (rDNA) repeats[27]. Samples efficiently demultiplexed from each pool (Fig. 1c) and DNA methylation was accurately measured (Fig. 1d) with high correlation (R=0.9) between replicate CpG levels (Fig. 1e and Fig. S3a) and with deep coverage at targeted epigenetic clock loci from less than 600,000 reads (Fig. 1f). In comparison with

RRBS libraries of the same samples, TIME-Seq libraries had substantially higher overlap with target clock CpGs (Fig. 1g). Age prediction using an existing RRBS-based rDNA clock, however, showed only moderate correlation with age in our pilot (R=0.53; Fig. S3b), possibly due to the differences in CpG coverage between TIME-Seq and RRBS at several clock loci (Fig. S3c).

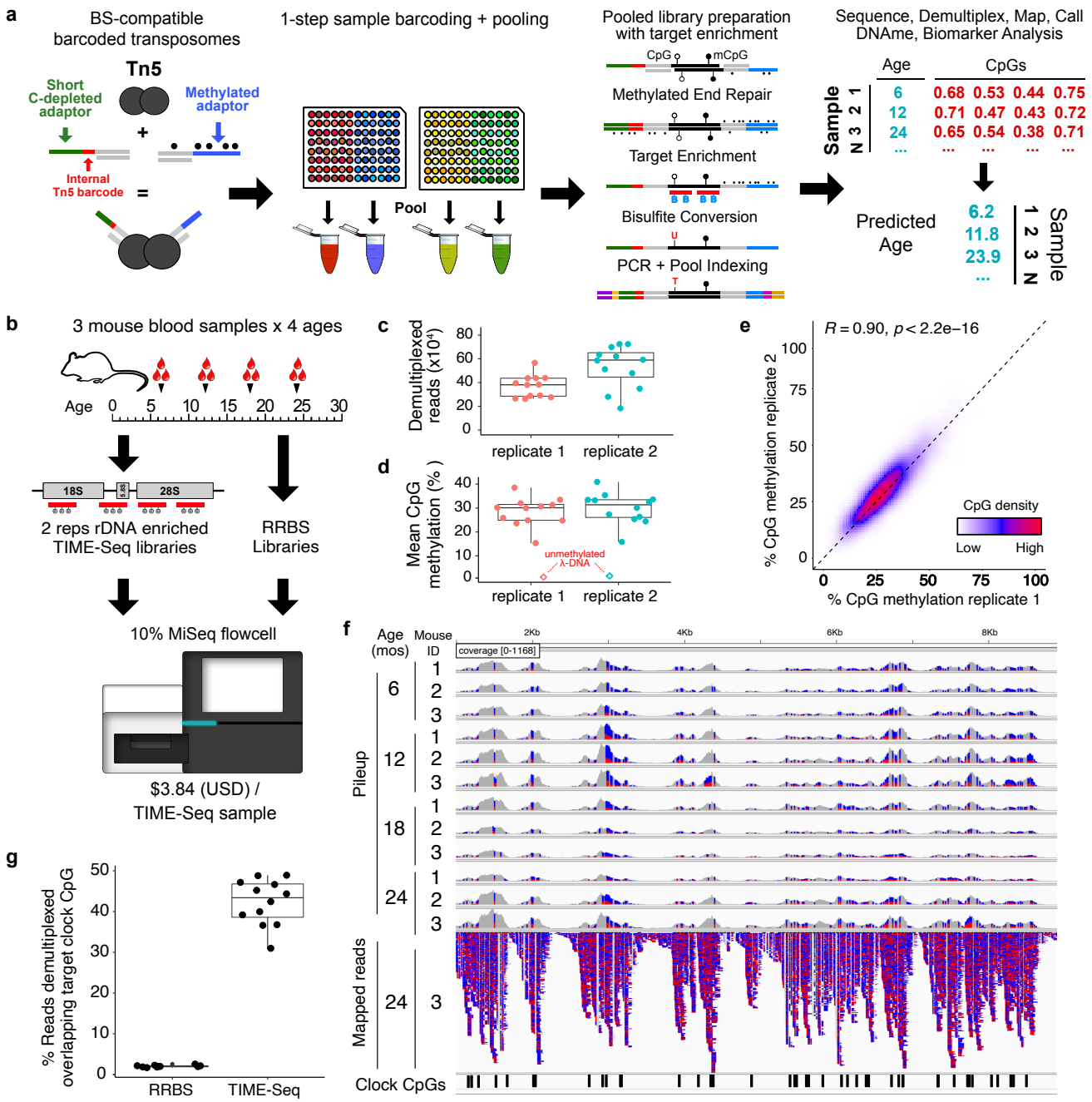

Figure 1. TIME-Seq enables highly multiplexed targeted methylation sequencing for biomarker discovery and measurement. ***a****, Schematic of the TIME-Seq library preparation for highly multiplexed targeted methylation sequencing to build and measure DNA methylation (DNAme)-based biomarkers.* ***b****, TIME-Seq pilot experimental design using mouse blood DNA from 4 age groups and preparing 2 replicates of each sample with rDNA baits (version 1) as well as RRBS libraries to be sequenced as a fraction of a Illumina MiSeq sequencing run.* ***c****, Demultiplexed reads from the 12-sample replicate TIME-Seq pools.* ***d****, Mean CpG methylation from reads mapped to the mouse ribosomal DNA meta-locus in TIME-Seq pools. Unmethylated lambda phage DNA control is represented as a diamond.* ***e****, Percent methylation from reads mapped to ribosomal DNA meta-locus in replicate 1 and replicate 2 in CpGs with coverage at least 125. Correlation from*

coverage at least 50-250 ranged from R=0.86-0.93 (see Fig. S3a). ***f****, Screenshot of IGV browser showing pileups of mapped reads in samples from a TIME-Seq pool (replicate 1) as well as mapped reads from sample 24-3. Reads are colored by mismatch, which for cytosines is blue for T (unmethylated) and red for C (methylated). Target Clock CpG coordinates are illustrated on the bottom track by black rectangles.* ***g****, Percent of reads directly overlapping target clock CpGs from TIME-Seq libraries (N=12; mean from 2 replicates) and shallow-sequenced RRBS libraries (N=10).*

To build a more accurate rDNA clock compatible with TIME-Seq, enrichment baits tiling the entire rRNA promoter and coding regions were designed and used to enrich TIME-Seq libraries from 191 mouse blood DNA samples (ages 2-35 months, 182 passed quality filters) in pools of 47-48 (Fig. 2a). Pools were combined and sequenced on an Illumina MiSeq for a per-sample cost of less than five US dollars (USD) (see Table S3 for sequencing costs). The majority of demultiplexed reads (Fig. 2b) from each sample (57-80%) mapped to the rDNA repeat meta-locus (Fig. 2c), resulting in high coverage for each sample at rDNA CpGs (Fig. 2d). Mean methylation levels from each sample (Fig. 2e) significantly increased with age (Fig. S4a) and were the main source of variance between samples based on principal component analysis (PCA) (Fig. S4b-c). Age was noticeably associated with variance in the second principal component (Fig. 2f), consistent with high age correlation for individual CpGs (Fig. 2g). To train the age prediction model, samples were split approximately 80:20 into training and testing sets (Fig. 2h) and a mostly-ridge (α=0.07) elastic net regression was applied (Fig. S4d). Age predictions using the resulting 187 CpG TIME-Seq rDNA (TS-rDNA) clock showed high correlation with age (training, R=0.97; testing, R=0.94) and a median absolute error (MedAE) of only 1.86 months in the testing samples (Fig. 2i). To build a clock that could be applied to both TIME-Seq and RRBS samples, we trained a model from TIME-Seq data using only CpGs with high coverage in RRBS data[1] (Fig. 2j). This clock showed high age correlation (R=0.87) in RRBS data and reflected the longevity benefit of caloric restriction. These data serve as a proof-of-concept for TIME-Seq as a robust, low-cost, and high throughput epigenetic age prediction method.

Most aging clocks are built from non-repetitive loci such as CpG islands in gene promoters or other gene regulatory elements[28]. We sought to build a clock from non-repetitive loci and designed baits to target CpGs that were previously reported to have high age-correlation in mouse blood[1] and multi-tissue clocks[3,12] (Fig. 2k and Fig. S4e). These baits—termed mouse discovery (MD) baits—were used to enrich five TIME-Seq pools made from 201 mouse blood DNA samples (186 passed quality filters). Target loci were enriched with 71% median mapped reads on-target (Fig. S4f). As expected, methylation levels of many of the target CpGs in the dataset had high correlation—both positive and negative—with age (Fig. 2l). The same training strategy (Fig. 2m and Fig. S4g) was applied to build a robust epigenetic clock, with high age-correlation (training, R=0.98; testing, R=0.91) and a MedAE of only 1.68 months (Fig. 2n). Like RRBS- and Illumina microarray-based clocks, the 173 CpG TIME-Seq blood (TS-blood) clock has CpGs with both positive and negative age trajectories (Fig. 2o). This experiment further validates TIME-Seq for cheap, scalable, and accurate age prediction and demonstrates the versatility of the approach.

We hypothesized that the cost of age prediction could be even further reduced by applying *scAge*[24]—a recently described method for epigenetic age prediction in single cells—to shallow-sequenced TIME-seq libraries, which might resemble sparse and binary single-cell DNAme data. To test this, we shallow-sequenced 3 MD-enriched TIME-Seq pools containing 121 mouse-blood DNA samples (119 passed quality filters; Fig. 3a). Sample reads ranged from 3,610 to 32,588 (median 11,560; Fig. 3b) with a per-sample cost of just $1.85 (USD). While average methylation levels were approximately the same as deep-sequenced data (Fig. S5a), the vast majority of CpGs were only covered by one or a few reads (Fig. 3c). To predict age from such low-pass sequencing, *scAge* leverages existing deep-sequenced methylation data to construct linear models for maximum-likelihood age prediction (Fig. 3d). We first applied *scAge* to shallow-TIME-Seq data using published RRBS data[1,3] as reference. Since TIME-Seq libraries are enriched for age-correlated loci in RRBS datasets, there was high intersection (median 49.1%) between model CpGs and shallow-TIME-Seq data compared to a more random distribution of sparse data, such as whole-genome single-cell methylation data[2] (Fig. 3e).

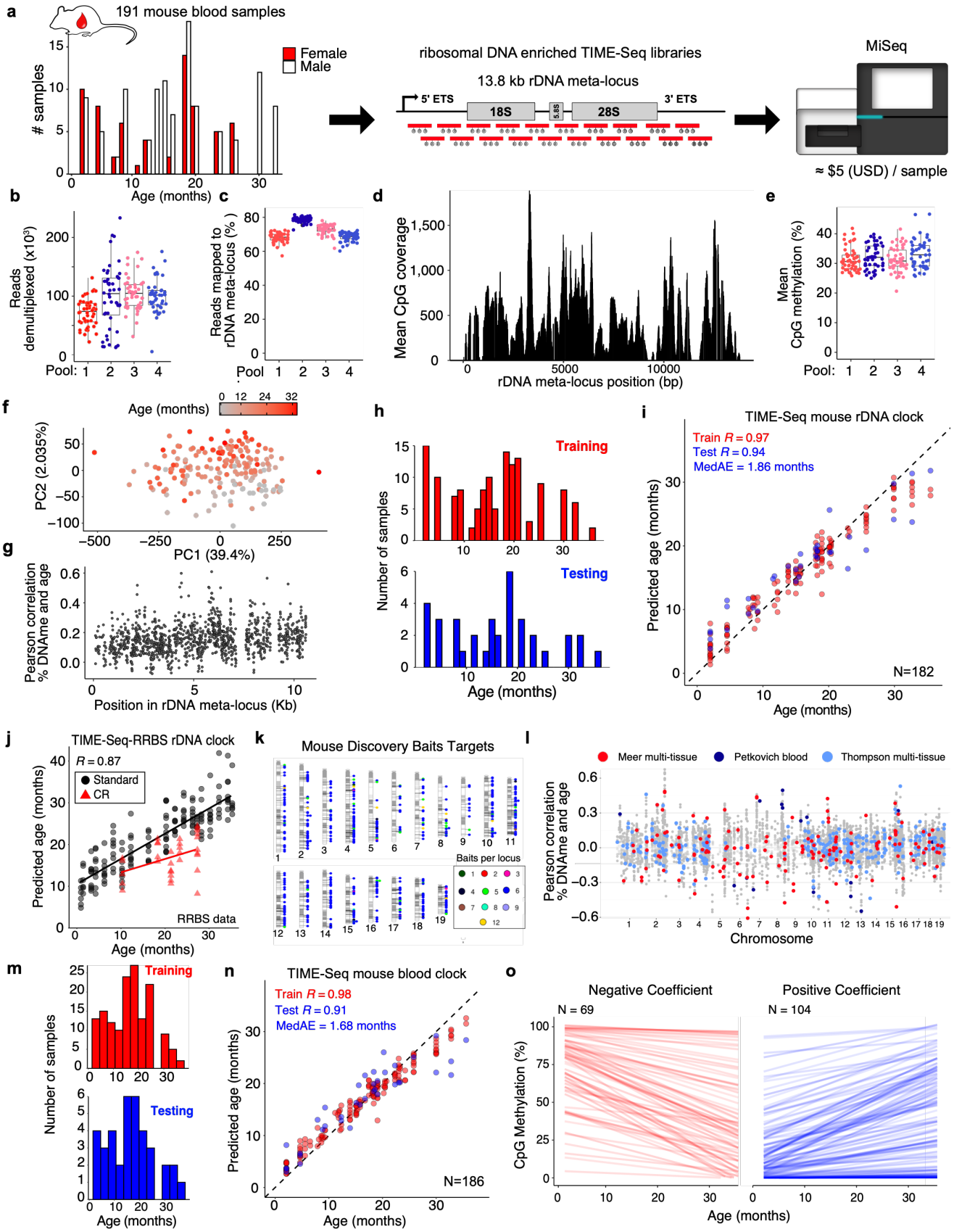

Figure 2. Inexpensive and accurate blood epigenetic clocks trained on a large cohort of mice using TIME-Seq. a, *Schematic of 191 mouse blood DNA samples (histogram), library preparation, and sequencing for TIME-Seq rDNA*

clock training and testing. ***b****, Reads demultiplexed from each rDNA clock pool of 47-48 samples.* ***c****, Percent of demultiplexed reads from each sample that mapped to the rDNA meta-locus.* ***d****, Mean coverage at rDNA meta-locus CpGs in rDNA-enriched TIME-Seq libraries.* ***e****, Mean CpG methylation from each sample in the four pools.* ***f****, Principal component (PC) 1 plotted against PC2 values from PCA of rDNA-enriched TIME-Seq libraries. Samples are colored by age with grey (young) to red (old).* ***g****, Pearson correlation between methylation ratio (0-1) and age at individual CpGs across the rDNA meta-locus.* ***h****, Histogram of training (N=145; red) and testing (N=37; blue) samples used to develop the TIME-Seq rDNA clock.* ***i****, TIME-Seq mouse rDNA clock showing age prediction for training (red) and testing (blue). Pearson correlation and median absolute error (MedAE) on test are shown in the top left corner.* ***j****, TIME-Seq-based clock developed using only CpGs with at least 50 coverage in RRBS data used to develop the original mouse rDNA clock[1]. Caloric restricted (CR) mice are represented as red triangles.* ***k****, Illustration of target loci (circles) from across the mouse genome represented in a chromosome ideogram (PhenoGram). Colors represent differing numbers of probes at each target locus (legend, bottom right).* ***l****, Pearson correlation between individual CpG ratios across chromosomes represented in a Manhattan plot style. Clock loci from three previously described mouse clocks are shown colored.* ***m****, Histograms of training (N=149; red) and testing (N=37; blue) samples for the TIME-Seq mouse blood clock.* ***n****, TIME-Seq mouse blood clock showing age prediction for training (red) and testing (blue). Pearson correlation and MedAE are shown in the top left corner.* ***o****, Linear regression models fit to CpG levels across age for each clock CpG, split by clock coefficient sign. Transparency is dependent on the absolute value of the coefficient (i.e., CpGs that carry more weight are more opaque).*

Remarkably, we found that age predictions were extremely accurate (Fig. 3f-g and Fig S5b-c) with a correlation of R=0.86 (Females, R=0.94; Males, R=0.85) and MedAE of just 2.23 months. Next, we used the deep-sequenced TIME-Seq data as the reference for *scAge*, which further improved age correlation (R=0.91; MedAE=2.87; Fig. 3h-i) to a similar level as deep sequenced elastic-net-based clocks. To test if this approach is generalizable to other mouse tissues, we prepared and shallow-sequenced TIME-Seq libraries from 104 mouse liver samples (ages 3-29 months), finding age predictions remarkably accurate using both deep-sequenced TIME-Seq liver data (Fig. 3j and Fig. S5d) and RRBS liver data (Fig. S5e-f) as reference. This combination of shallow-TIME-Seq and *scAge* decreases the cost of age prediction by more than 2 orders of magnitude compared to traditional sequencing and epigenetic clock approaches.

To validate TIME-Seq-based age prediction approaches, we prepared both rDNA- and MD-enriched TIME-Seq libraries in two separate experiments using blood DNA from an independent cohort of mice (N=75-93; ages 1.9-32.9 months). A subset of these mice were tracked longitudinally, assessed using the mouse frailty-index[29]—a composite biomarker of age-related decline—and had blood composition parameters measured (Fig. 4a). The TS-rDNA clock (R=0.83; Fig. 4b) and TS-blood clock (R=0.93; Fig. 4c), as well as the combination of *scAge* and shallow-TIME-Seq using both RRBS (R=0.84; Fig. S6a) and the original TIME-Seq data for reference CpGs (R=0.87; Fig. 4d) gave accurate age predictions that reflected longitudinal age change. While the TS-rDNA clock had the least accurate predictions, it maintained rank-order prediction in many mice from the longitudinal cohort (Fig. S6b), demonstrating an ability to measure consistent aging trajectories. Further, mice from The Jackson Labs (JAX)—the same colony as the original clock training set—had metrics comparable to the TS-rDNA testing set (JAX mice, R=0.93; Fig. S6c), suggesting that inter-colony differences might decrease TS-rDNA clock accuracy even in mice of the same lineage (C57BL/6). Accuracy of the rDNA clock also improved when filtering for samples with higher coverage (Fig. S6d), which will help inform the target reads-per-sample in future applications. TS-blood clock and *scAge* predictions were highly correlated with each other (R=0.91) and both were less well correlated with rDNA age (Fig. 4e).

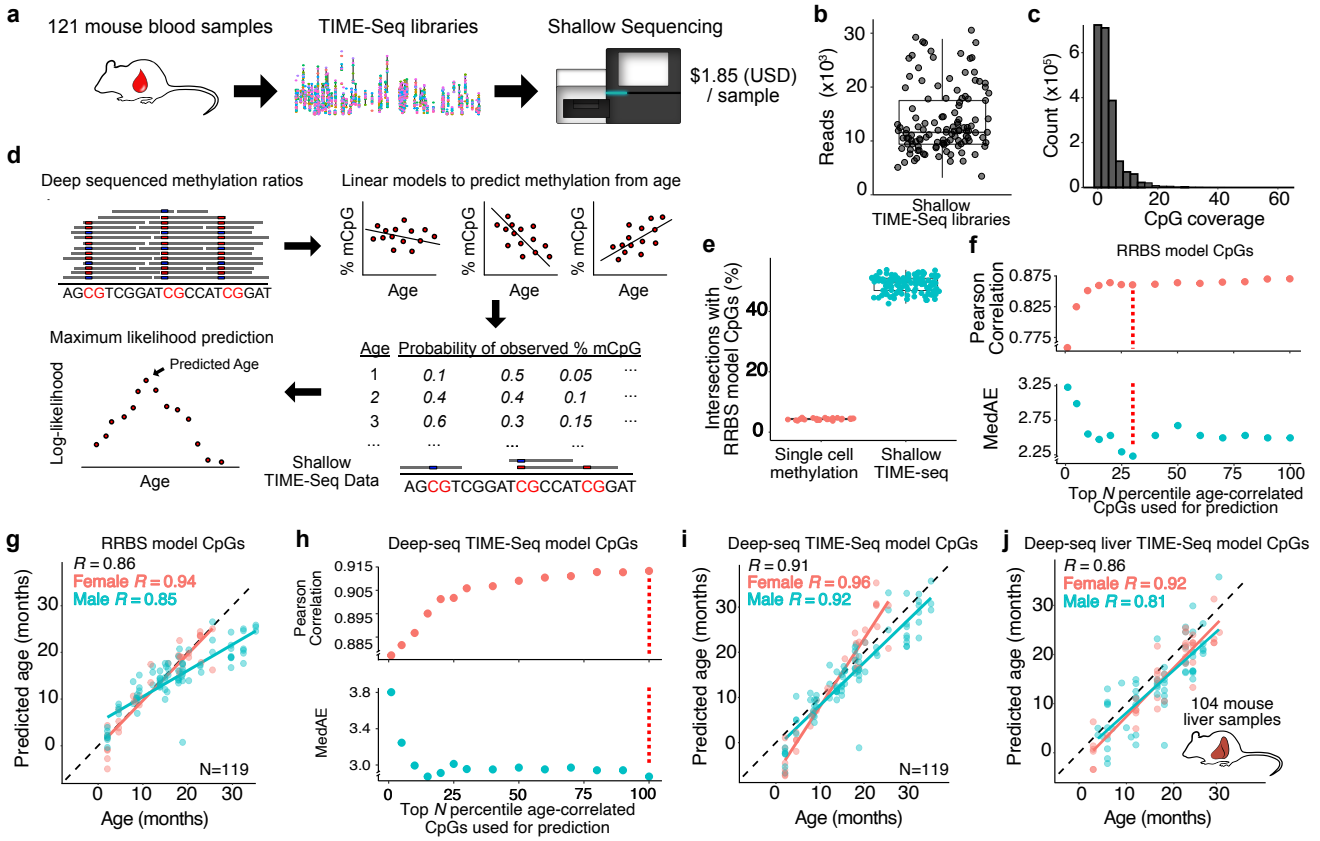

Figure 3. Ultra-cheap age prediction with scAge from shallow sequencing of TIME-Seq libraries. a*, Schematic of shallow sequencing of 121 MD-enriched TIME-Seq libraries.* ***b****, Demultiplexed read number from each shallow sequenced sample.* ***c****, Histogram of CpG coverages in shallow sequenced samples.* ***d****, Schematic of the scAge framework for maximum likelihood age prediction from shallow TIME-Seq methylation data.* ***e****, Percent of scAge model CpGs covered in each shallow sequenced TIME-Seq library and a previously reported single-cell methylation dataset*[2]*.* ***f****, Pearson correlation and MedAE from scAge-based age prediction using reference RRBS data (Thompson et al., 2018). CpGs in each sample were ranked by the absolute value of their correlation with age (based on deeply sequenced data), and only the top N percentile was used for maximum likelihood prediction at each point. The red line indicates the percentile presented in (g).* ***g****, Age predictions from scAge in shallow-sequenced TIME-Seq libraries (N=119) using RRBS reference data (top 30% age-associated CpGs). Pearson correlations are shown in the top left corner.* ***h****, Pearson correlation and MedAE from scAge-based age prediction of the top N percentiles of CpGs using deep-sequenced, MD-enriched TIME-Seq data as reference (data used for Fig. 2k-o). The red line indicates the percentile presented in (i).* ***i****, scAge predictions from shallow-TIME-Seq data using deep-sequenced TIME-Seq data (100% of CpGs) as models. Pearson correlations are shown in the top left corner.* ***j****, scAge predictions from 104 shallow-sequenced liver samples using deep-sequenced TIME-Seq liver data as model CpGs (100% of CpGs). Pearson correlations are shown in the top left corner.*

The difference between predicted age from epigenetic clocks and chronological age (ΔAge) has been shown to correlate with a wide variety of age-associated phenotypes[7]. To test if TIME-Seq predictions related to other measures of health or aging, we compared ΔAges from each approach to mouse frailty index and blood composition measurements (Fig. 4f-g). To control for raw age correlation of each variable (top Fig. 4f and Fig. S6e), measurements from each mouse were subtracted by the median value of that variable in similar aged animals—abbreviated ΔMed^{age}(blood) and ΔMed^{age}(FI). ΔMed^{age}(blood) values were not correlated with ΔAges from the deep-sequenced clocks, suggesting blood cell composition is not driving predictive variance. However, ΔMed^{age}(blood) values from the parameters with significant age correlations, mean platelet volume (MPV), neutrophil percent (NE %), and lymphocyte percent (LY %) were also significantly correlated with *scAge* ΔAges in the same direction.

Since frailty index is also highly correlated with age and indicative of age-related decline, we hypothesized that ΔAges might be positively correlated with ΔMed^{age}(FI) values (i.e., mice that are more frail for their age are also predicted older). However, when comparing ΔMed^{age}(FI) to ΔAges (Fig. 4g), we found no correlation with any TIME-Seq prediction method. While further investigation will be needed to delineate the association of blood composition and frailty with epigenetic age predictions, this analysis demonstrates the usefulness of TIME-Seq in assessing large cohorts with paired data.

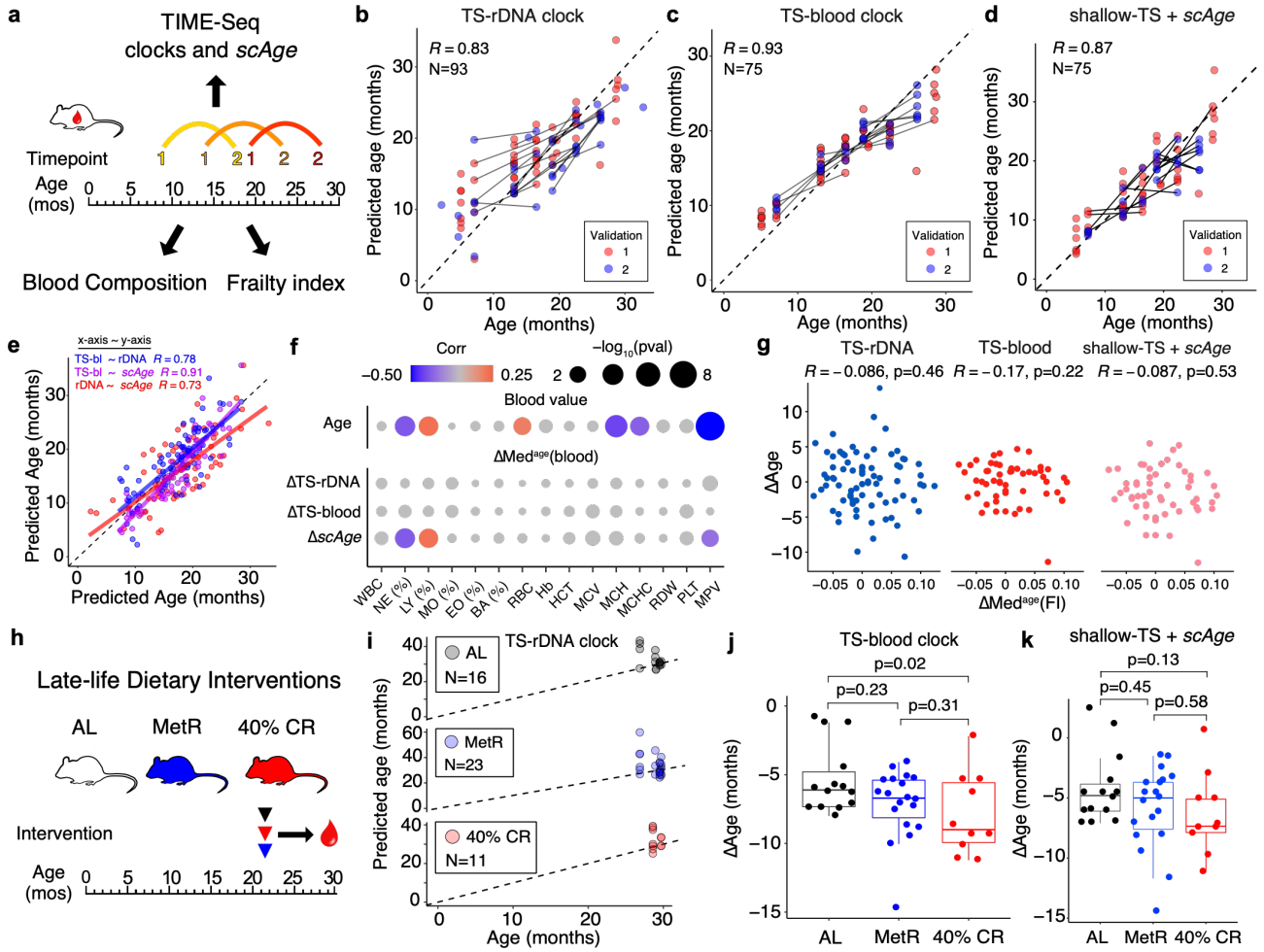

Figure 4. Validation of TIME-Seq age predictions in longitudinally tracked mice and late-life dietary interventions. a*, Experimental schematic for validation of TIME-Seq age prediction methods in an independent cohort of mice with longitudinal timepoints, paired frailty index, and blood composition data.* ***b-d,*** *TIME-Seq age predictions in two independent validation library preparations using (****b****) TIME-Seq mouse rDNA (TS-rDNA) clock, (****c****) TIME-Seq mouse blood (TS-blood) clock, and (****d****) scAge predictions from shallow-TIME-Seq data (shallow-TS + scAge) using deep-TIME-Seq data (original MD-enriched libraries shown in Fig. 2) as reference. Lines connect the same mouse at two different ages. Pearson correlations and n values are shown in the top left corner.* ***e****, Correlation between age predictions in the validation sets from all three approaches.* ***f****, Pearson correlation and significance matrix between ΔAge from each approach and ΔMed^{age}(blood), i.e., the difference in median value from similar aged mice for each blood measurement. Color and size of each circle represent the Pearson correlation and p-value significance, respectively. WBC = white blood cell count (1000 cells/μl), NE (%) = percent of neutrophils, LY (%) = percent of lymphocytes, MO (%) = percent of monocytes, EO (%) = percent of eosinophils, BA (%) = percent of basophils, RBC = red blood cell count (10^6/μl), Hb = hemoglobin (g/dL), HCT = hematocrit, MCV = mean corpuscular volume (fL), MCH = mean corpuscular hemoglobin (pg), MCHC = mean corpuscular hemoglobin concentration (g/dL), RDW = red blood cell distribution width, PLT = platelets (1000 cells/μl), MPV = mean platelet volume (fL)* ***g****, ΔMed^{age}(FI), i.e., the difference in frailty index for each mouse from the median frailty index from similar aged mice, plotted against ΔAge from each approach. Pearson correlations and p-values are shown above each graph.* ***h****, Schematic of late-life dietary interventions and blood collection. AL = ad libitum (N=14, black); MetR = methionine restriction (N=15, blue); 40% CR = 40% caloric restriction (N=10,*

red). Colors carry on to panels (i-j) ***i****, TS-rDNA clock predictions in intervention mice. TS-blood clock* ***(j)*** *and shallow-TS + scAge* ***(k)*** *predictions represented as a boxplot. Analysis of variance (ANOVA) was used to assess difference between groups. ANOVA p-values for group comparisons are shown above data.*

To test if TIME-Seq predictions could detect interventions that moderately extend lifespan in late-life, we measured epigenetic age in the blood of mice that were 40% caloric restricted (CR) or methionine restricted (MetR) for six months starting at 22-24 months of age (Fig. 4h). While TS-rDNA clock predictions from samples in these cohorts were largely accurate (Fig. 4i), the treated mice ΔAges were not significantly different from mice fed *ad libitum* (AL) (Fig. S6f). This contrasts with the negative ΔAges of lifelong CR mice using the RRBS-TS-rDNA clock (see Fig. 2j) and suggests a more drastic intervention might be needed to detect shifting age trajectories with the TS-rDNA clock in late-life mice. MD-enriched TIME-Seq libraries were also prepared and both deep and shallow sequenced. CR mice were predicted significantly younger than AL mice using the TS-blood clock (Fig 4j; ANOVA, p=0.02), though there was no difference between the MetR and AL groups, similar to frailty index analysis (Fig. S6g). While the pattern of age prediction using *scAge* resembled TS-blood clock predictions, none of the groups were statistically significant in comparison (Fig. 4k). These data demonstrate that TIME-Seq-based age predictions can detect even short-term longevity interventions started in late life.

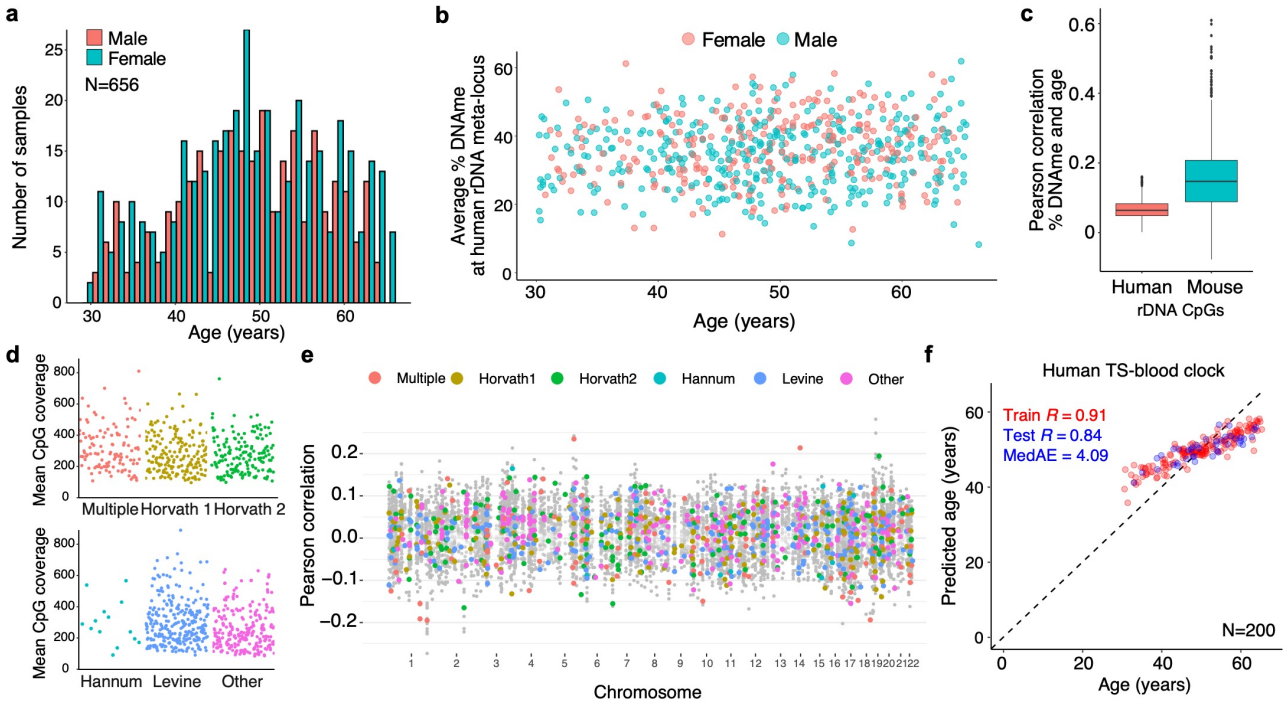

Figure 5. TIME-Seq accurately predicts age in hundreds of human samples. a*, Histogram of age colored by sex in the human DNA samples (N=656) used for TIME-Seq library preparations.* ***b****, Average methylation at human ribosomal DNA meta-locus. Points are colored by sex. Males (N=285), blue; Females (N=369), red.* ***c****, Boxplot of Pearson correlations between age and CpG methylation ratios in human (red) and mouse (blue) ribosomal repeat meta-locus from TIME-Seq data.* ***d****, Mean coverage of enriched CpGs from target human clock loci in HD-enriched TIME-Seq libraries.* ***e****, Pearson correlation between age and DNA methylation at CpGs across the human genome presented in Manhattan-plot style. CpGs that are colored are from each of the 11 target clocks.* ***f****, Human TS-blood clock built using the top 200 samples, filtered by number of reads on-target. Pearson correlation and median absolute error (MedAE) on test are shown in the top left corner.*

Next, we applied TIME-Seq to predict age in blood DNA samples from a large cohort of humans[30,31] (N=656; Fig. 5a) using hybridization baits that targeted human rDNA repeats or human discovery (HD) baits targeting 11 human clocks[28]. rDNA-enriched TIME-Seq libraries showed considerable variation in average methylation across reads mapped to a human rDNA meta-locus (8.5%-61.7%; Fig. 5b), which exceeds the variation for mice and did not correlate with

age and sex (Fig. 5b). Since there is high copy number variation of rDNA in humans[32], the methylation variation might result from hypermethylated repeats used to suppress rRNA dosage in individuals with more rDNA copies. Individual human rDNA CpGs had lower age-correlation than the mouse rDNA-enriched TIME-Seq data (Fig. 5c), and we could not produce an accurate epigenetic clock (defined as R>0.8)[7] from human rDNA (Fig. S7a). HD-enriched TIME-Seq libraries had high enrichment of target CpGs (Fig. 5d) with a noticeable age-correlation in the second and third principal components (Fig. S7b) and age correlated CpGs across the genome (Fig. 5e). Like rDNA, this dataset had lower age-correlation than the analogous mouse dataset, perhaps due to the more restricted age range (majority of samples were aged 40-60 years). While age prediction models from the entire dataset produced only moderate results (Fig. S7c), we built a clock from 200 samples with the highest on-target read number (testing, R=0.84; Fig. 2f). These data exemplify the scale that TIME-Seq enables, and refinement of human clocks to improve accuracy from broader age-ranges is the subject of ongoing research.

DISCUSSION

DNA methylation-based biomarkers are increasingly ubiquitous tools for both clinical and basic research as well as disease diagnosis, but they have been expensive and laborious, precluding the analysis of large cohorts. TIME-Seq is a flexible and scalable targeted sequencing approach that decreases costs for methylation biomarkers by more than two orders of magnitude. Using TIME-Seq, we built and validated epigenetic clocks to predict age from hundreds of mice and human samples in over 2000 unique libraries. This scale was enabled by immediate sample barcoding that facilitates low-cost, pooled library preparation compatible with efficient hybridization-based enrichment and bisulfite conversion. Compared to traditional RRBS library preparations, which can take anywhere from 4-9 days[33] and cost upwards of $30-$50 per sample, TIME-Seq libraries cost only $0.65 per sample (see Table S1) and take only 1-1.5 days to complete. Further, input DNA for TIME-Seq (100 ng) is the same as standard RRBS libraries and 3-5 times less than DNAme microarray, enabling longitudinal measurement of epigenetic age from low-yield DNA extractions such as a mouse cheek bleed.

Using a combination of shallow-TIME-Seq and *scAge*, we discovered that it is possible to predict age in bulk samples from ten thousand sequencing reads or less. We applied *scAge* with both deep-sequenced RRBS and TIME-Seq data as reference, finding both capable of highly accurate age prediction in mouse blood and liver. Using *scAge* in combination with TIME-Seq library preparation and shallow sequencing reduces the cost of accurate age prediction more than 200-fold. While targeted approaches have been described for clock analysis[34-36], such as pyrosequencing or digital PCR, these methods are more expensive and less scalable than TIME-Seq predictions—both elastic-net-based clocks and using *scAge*. Further, their reliance on low-CpG clocks (e.g., 3-15 CpGs) limits their biological relevance to the small set of measured loci. TIME-Seq is flexible and capable of high enrichment of thousands to tens-of-thousands of CpGs.

In the current study, we rationally designed baits to target loci that had high correlation with the target phenotype (i.e., age). However, when such data is not available, more genomic area might be enriched to identify CpGs with high-correlation relative to the phenotype of interest (e.g., metabolic syndrome, cancer, or frailty). These libraries would be more costly to sequence initially, however, the baits could be downscaled once model CpGs were discovered, requiring fewer reads per sample to attain a readout to the biomarker. This separation of more expensive biomarker discovery from low-cost measurement will be key to the widespread adoption and routine use of DNAme biomarkers as well as those based on other biomolecules[8,37].

Being designed for maximal cost and labor efficiency, TIME-Seq has the same trade-offs as other Tn5-based library preparations, and DNA samples that are improperly quantified tend to drop-out due to either over- or under-tagmentation. This limitation might be addressed using on-bead tagmentation[38] that controls for variation in starting DNA amount. Ultimately, strategies for high-throughput DNA extraction and normalization, combined with an automated TIME-Seq library preparation could facilitate thousands of samples being processed by a scientist in a few days and enable low-cost population-scale age predictions or DNAme biomarker survey.

METHODS

Animal assessments

All mouse experiments were approved by the Institutional Animal Care and Use Committee of the Harvard Medical Area Standing Committee on Animals. Male and female C57BL/6Nia mice were obtained from the National Institute on Aging (NIA, Bethesda, MD), and group housed (3-4 mice per cage) at Harvard Medical School in ventilated microisolator cages with a 12:12 hour light cycle, at 71°F with 45-50% humidity. Mouse blood samples (150-300 µl) were collected in anesthetized mice (3% isoflurane) from the submandibular vein into tubes containing approximately 10% by-volume of 0.5M EDTA. Blood was spun at 1500 RCF for 10 minutes and plasma removed. Blood cell pellets were stored frozen at -80ºC. For validation experiments, a sub-sample of whole blood was processed with the Hemavet 950 (Drew Scientific) to give 20 whole blood count parameters. Frailty was assessed using the mouse clinical frailty index[29], a non-invasive assessment of 31 health deficits in mice. For each item, mice were scored 0 for no deficit, 0.5 for a mild deficit, and 1 for a severe deficit. Scores were added and divided by 31 to give a frailty index between 0 and 1, where a higher score indicates a greater degree of frailty. For more details see http://frailtyclocks.sinclairlab.org. 200 mouse (C57BL/6N) ocular-vein blood samples were collected by researchers at The Jackson Laboratory's Nathan Shock Center (Bar Harbor, Maine) according to methods described previously[39]. These samples were used for TIME-Seq clock training and testing, as well as shallow-sequencing analysis.

Late-life dietary interventions

Male and female C57BL/6Nia mice were obtained from the NIA at 19 months of age and housed at the Harvard T.H. Chan School of Public Health (Boston, MA). Mice were group housed 3-4 per cage for the duration of the study in static isolator cages at 71°F with 45–50% humidity, on a 12:12 hour light-dark (07:00am–07:00 pm) cycle. After arrival, mice were fed a control diet (Research Diets, New Brunswick, NJ) until the start of the study. Mice were then randomized to one of three groups: *ad libitum* diet, methionine restriction (0.1% methionine) or 40% CR. CR was started in a stepwise fashion decreasing food intake by 10% per week until they reached 40% CR at week 4. CR intake was based upon *ad libitum* intake. Mice were monitored weekly for bodyweights and food intakes. Fasting blood samples (4-6h) were taken at sacrifice (after 6 months on the diet) by cardiac puncture. Approximately 200 µl of whole blood in 1 µl of 0.5 mM EDTA was collected. The tube was spun, and the plasma removed. The remaining blood pellet was frozen at -80ºC until further analysis. Custom mouse diets were formulated at Research Diets (New Brunswick, NJ; catalog #'s A17101101 and A19022001).

DNA extraction and quantification

100-300 µl of pelleted mouse whole blood (plasma removed) was resuspended in 1mL of red blood cell (RBC) lysis buffer (155 mM NH4Cl, 12 mM NaHCO3, 0.1 mM EDTA, pH 7.3), incubated for 10 minutes on ice, and centrifuged at 2000 RCF for 5 minutes. Pelleted cells were resuspended in 1.5mLs of RBC lysis buffer and spun twice before being lysed in 800 µl TER (50 mM Tris-HCl pH8, 10 mM EDTA, 40 µg/mL RNase A) with 50 µl of 10% SDS added. Lysates were incubated at 37ºC to allow for RNA degradation and then 50 µl of 20 mg/mL proteinase K was added and samples were incubated overnight at 65ºC. To purify DNA, 500 µl of 1:5 diluted (dilution buffer: 20% PEG 8000, 2.5 M NaCl, 10 mM Tris-HCl pH8, 1 mM EDTA, 0.05% Tween-20) SPRI DNA binding beads were added to each sample, and they were incubated with rotation for 30 minutes at room temperature. Tubes were then placed on magnetic racks to capture SPRI beads, and the beads were washed with 1 mL of ice-cold 80% ethanol twice. DNA was eluted in 75 µl of 10mM Tris-HCl (pH 8). Purified DNA was quantified using the qubit double-stranded DNA broad range kit (Catalog No. Q32850, ThermoFisher) and diluted to 10 ng/µl for TIME-Seq reactions. To check for contaminants, a subset of samples from each extraction were assessed by nanodrop.

Human blood DNA samples

Human blood DNA came from a cohort of individuals participating in the HANDLS study (Healthy Aging in Neighborhoods of Diversity across the Life Span), a longitudinal fixed cohort study of 3,720 community-dwelling

African American and white adults aged 30-64 in Baltimore, MD, USA[31]. Methods for selection criteria, sample collection, extraction of DNA from blood, and more are described in previous publications[30,31,40]. Purified aliquots of DNA from 675 individuals were received and used for this study. HANDLS is approved by the National Institutes of Health Institutional Review Board (IRB; 09AG248). All participants provided written informed consent. The work done with these samples including data storage, analysis, and handling are approved by the IRB of the Harvard Faculty of Medicine (IRB21-0693).

Tn5 transposase purification

Tn5 transposase was purified according to a protocol originally described in Nguyen Ba et al. (pre-print)[41], which is based on published protocols[42] with several adjustments to increase yield. Briefly, 1 mL of overnight culture containing pTXB1-Tn5 (Addgene plasmid #60240) was used to inoculate 1 L of ZYM-505 growth media containing 100 µg/mL ampicillin and 0.001% polypropylene glycol (L14699-AE, Alfa Aesar). After the culture was grown for 4 hours at 37ºC, IPTG (0487-10G, VWR) was added to 0.25 mM, and the culture was grown for an additional 4 hours at 18ºC. The culture was centrifuged at 25,000 RPM for 25 minutes, and the pelleted culture was flash-frozen in liquid nitrogen before being stored overnight at -80ºC. The pellet was thawed on ice, resuspended in 10 mL of HEGX buffer (20mM HEPES-KOH pH 7.2, 0.8 M NaCl, 1mM EDTA, 10% glycerol, 0.2% Triton X-100) with a Roche protease inhibitor (SKU11697498001, Millipore Sigma), and 1% w/v of pre-dissolved (50% w/v) octyl-thioglucoside (O-130-5 Gold Bio) was added to help lysis. After a 10-minute incubation on ice, 100 mL of HEG-X was added, and the lysate was transferred to a glass beaker for sonication on the 550 Sonic Dismembrator (ThermoFisher) using 15 cycles (15 seconds on, 30 seconds off) on 70% duty and power 7. The sonicated lysate was pelleted at 30,000 RPM for 30 minutes at 4ºC. The supernatant was transferred to a clean beaker with a stir bar, placed on a magnetic stir plate, and 10% PEI was added dropwise while stirring to remove excess bacterial DNA. After 15 minutes, PEI and precipitated DNA were removed by spinning the mixture at 30,000 RPM for another 30 minutes at 4ºC. The lysate was added to 2 chromatography columns (7321010, BioRad) packed with 25 mL each of chitin resin (S6651S, NEB) and equilibrated with 100 mL of HEG-X each. The supernatant was added to the columns in equal proportion and allowed to flow through, before the column was washed with 30 column volumes of HEG-X. To elute the purified Tn5, 25 mLs of HEG-X with 100 mM DTT was added to each column and 10 mLs was allowed to flow through before sealing the column and letting it stand for 44 hours. 27 mL of elution was collected, and a Bradford assay (23200, ThermoFisher) was used to quantify protein. The elution was then concentrated to 20 mL using Amplicon Ultra 30 kDA filters (UFC900308, Millipore Sigma) and dialyzed in a Slide-A-Lyzer (66212, ThermoFisher) cassette with 1 L dialysis buffer (50 mM HEPES-KOH, pH 7.2 0.8M NaCl, 0.2 mM EDTA, 2 mM DTT, 20% glycerol). After two rounds of dialysis totaling 24 hours, the eluted protein was removed, aliquoted into 1 mL microcentrifuge tubes, and flash frozen in liquid nitrogen before being stored at -80ºC. The final concentration of purified protein was 1.5 mg/mL, and we estimate that 1L of culture produced enough Tn5 for approximately 16,000 TIME-Seq reactions with 100 ng of DNA per sample.

Activation of TIME-Seq Tn5

Oligonucleotides (oligos) were ordered from IDT and HPLC purified except for TIME-Seq indexed adaptors and hybridization blocking oligos. To anneal TIME-Seq adaptors, 100 µM TIME-Seq adaptor B containing a 5-bp internal barcode and (separately) 100 µM methylated adaptor A were combined in equal volume with the 100 µM Tn5 reverse ME oligo. Enough methylated A adaptor was annealed to be added in equal proportion with each indexed adaptor B. Oligos were denatured at 95ºC for 2 minutes and then ramped down to 25ºC at 0.1ºC/s. The annealed oligos were then diluted with 85% glycerol to 20 µM, and the methylated A adaptor, indexed TIME-Seq B adaptor, and 50% glycerol were combined in a ratio of 1:1:2. The resulting 10µM adaptors were combined in equal volume with purified Tn5 (1.5 mg/mL), mixed thoroughly by pipetting 20 times, and incubated at room temperature for 50 minutes. Activated transposomes were stored at -20ºC and no loss of activity has been observed up to 8 months.

To test activity of TIME-Seq transposomes, 100 ng of human genomic DNA (11691112001, Roche) was tagmented in 25 µl reactions by adding 12.5 µl of 2X tagmentation buffer (20mM Tris-HCl pH 7.8, 10 mM dimethylformamide, 10 mM $MgCl_2$) using 1.5 µl of each barcoded transposome. After reactions were incubated at 55ºC for 15 minutes, STOP buffer (100mM MES pH5, 4.125M guanidine thiocyanate 25% isopropanol, 10mM EDTA) was added to denature Tn5 and release DNA. To assess tagmentation, 90% of each reaction was run in separate lanes of a 1% agarose gel at 90 volts (V) for 1 hour. The gel was stained with 1x SYBR gold (S11494, ThermoFisher) in Tris-Acetate-EDTA (TAE) buffer, and DNA fragment size was determined using a ChemiDoc (Bio-Rad) for gel imaging. The remaining DNA was pooled, cleaned up using a DNA Clean & Concentrator-5 (D4013, Zymo) kit, and the DNA was amplified with barcoded TIME-Seq PCR primers. Amplified pools were spiked into sequencing runs to assess relative barcode activity when new transposase was activated.

Biotin-RNA bait design and production

Mouse and human discovery baits were designed to enrich for previously described epigenetic clock CpGs[1,3,12,28] using mm10 and hg19 genomic coordinates. Using bedtools[43] (v2.28.0), regions around each target CpG were first expanded 125 base pairs (bps) up- and downstream (*bedtools slop*) and overlapping loci were merged (*bedtools merge*). Next, 110 bp bait windows were defined every 20 bps in the region (*bedtools window*) and the baits were intersected (*bedtools intersect*) with a file containing RepeatMasker (http://www.repeatmasker.org) annotated regions to identify baits overlapping repetitive DNA regions. Next, the FASTA sequence of each bait was gathered (*bedtools getfasta*), and blat (version 35) was used to get the copy number for each bait with options, -fastMap -maxIntron=50 -stepSize=5 -repMatch=2253 -minScore=40 -minIdentity=0. Using custom R (v. 4.0.2) scripts, information was gathered from the output files from each probe including the percent of each nucleotide, the overlap with repeats, and the bait copy number as determined by blat. Baits were automatically filtered that had overlap with repeats, however, each locus that had none or very few (< 4) baits after filtering were inspected manually and, if the blat copy number was low (< 10), baits were added back, either to the exact locus or shifted slightly to avoid the annotated repeat. In preliminary biotin-RNA hybridization experiments (data not shown), we noticed that baits with high or low percentage T (less than 8% or greater than 30%) had low coverage, possibly due to stalling of the RNA polymerase while incorporating biotin-UTP. Therefore, when more than half the baits at a target locus had a percent T greater than 30% or lower than 8%, the reverse complement strand was captured instead for all baits at the locus.

To design enrichment baits for mouse and human rDNA, FASTA sequences were prepared from GeneBank accessions BK000964.3 (mouse) and U13369.1 (human) according to previously described methods[27] by moving the last 500 bps of each sequence (rDNA promoter) to be in front of the 5' external transcribed spacer (ETS). From the region comprising the promoter to the 3' ETS (mouse, 13,850bps; human, 13,814bps), 110-nt baits were designed using the *bedtools window* function to create baits every 20bps. Version 1 rDNA baits used in the pilot and targeting the original rDNA clock were designed to specifically enrich rDNA clock CpGs[27] using the same approach described for non-repetitive clock CpGs (i.e., 250bp windows were merged and 110-nt baits were designed to tile the regions).

The sequence of each bait set was appended with a promoter (Sp6 or T7) for *in vitro* transcription (IVT), as well as a promoter-specific reverse priming sequence that contained a BsrDI restriction enzyme motif. Bait sets containing Sp6 and T7 promoters were ordered together in a single-stranded DNA oligo pool (Twist), and pools were resuspended to 10 ng/µl. Bait sets were amplified in reactions containing 12.5 µl of 2X KAPA HiFi HotStart Polymerase Mix (7958927001, Roche), 0.75 µl of each 10 µM primer, and 0.5 µl of the bait pool using the following thermocycler program: initial denaturation, 95ºC for 3 min; 10 cycles amplification, 98ºC for 20 seconds, 61ºC (Sp6) or 58ºC (T7) for 15 seconds, 72ºC for 15 seconds; a final elongation for 1 minute at 72ºC. Amplified DNA was cleaned up using a Clean & Concentrator-5 kit (D4013, Zymo) and then digested with 1 µl BsrDI (R0574S, NEB) at 65ºC for 30 minutes. This reaction was again purified with a Clean & Concentrator-5 kit, and IVT reactions were set up according to the HiScribe™ T7 (E2040S, NEB) or Sp6 (E2070S, NEB) High Yield kits using half of the cleaned DNA template for each reaction and storing the rest at -80ºC. All ribonucleotides were added to a final concentration of 5mM, including a 1:4 ratio of

biotin-16-UTP (BU6105H, Lucigen) to UTP at 5 mM. After the reactions were incubated for 16 hours overnight at 37ºC, 25 µl of nuclease free water and 2 µl of DnaseI (M0303S, NEB) were added to degrade the DNA template. RNA was purified using the 50 µg Monarch® RNA Cleanup Kit (T2040S, NEB). The concentration of RNA was measured using Qubit RNA BR Assay Kit (Q10210, ThermoFisher) and the size of RNA was measured using an RNA ScreenTape (5067-5576, Agilent) on an Agilent Tapestation.

While the yield from each DNA amplification and IVT reaction varies depending on the size and composition of the bait sets, we estimate that 1,000-10,000 hybridization reactions-worth of bait could be produced from just 1 single-stranded oligo pool. For TIME-Seq libraries of 48-64 samples, this could enrich tens to hundreds of thousands of samples. Another advantage of this approach is that baits can be easily shared with other researchers, either the ssDNA template, amplified DNA, or prepared biotin-RNA.

TIME-Seq library preparation

For TIME-Seq library preparations, samples were organized into relatively even pools, and 10 µl of DNA (10 ng/µl, 100 ng total) from each sample was distributed into separate wells of strip tubes (or 96-well plates) for tagmentation. 100 ng of unmethylated lambda phage DNA (D1521, Promega) was tagmented with each pool. Lambda DNA that came through at a low percentage of demultiplexed reads served to estimate bisulfite conversion efficiency. To tagment samples, 12.5 µl of 2X tagmentation buffer (20mM Tris-HCl pH 7.8, 10mM dimethylformamide, 10mM $MgCl_2$) was added to each sample. Next, 2.5 µl of uniquely indexed TIME-Seq transposome was added, and the reaction was immediately mixed by pipetting 20 times. Once transposome was added to each sample in a pool, the samples were placed at 55ºC for 15 minutes. After incubation, 7 µl of STOP buffer (100mM MES pH5, 4.125M guanidine thiocyanate 25% isopropanol, 10mM EDTA) was added, pools were vortexed and pulse spun in a centrifuge, and the reaction was incubated at 55ºC for an additional 5 minutes.

After stopping the reactions, samples from each pool were combined into a single tube, typically a 5 mL Lo-bind tube (0030122348, Eppendorf) or 15 mL falcon tube (229410, Celltreat), and 118 µl per sample of DNA Binding Buffer (D4004-1-L, Zymo) was added. Pools were then applied to Clean & Concentrate-25 (D4033, Zymo) columns. If the volume of the pool exceeded 5 mL, each pool was passed in equal volume through 2 separate columns. After 2 washes, pools were eluted in 41 µl (typically yielding 39 µl after elution) and 1 µl was removed to assess tagmentation fragment size and yield by D5000 ScreenTape (Catalog No. 5067- 5588, Agilent) on an Agilent Tapestation.

For methylated end-repair, eluted pools were combined with 5 µl of NEB Buffer 2, 5 µl of 5mM dNTPs containing 5-methyl-dCTP (N0356S, NEB) instead of dCTP, and 2 µl of Klenow Fragment (3'→5' exo-) (M0212L, NEB). The reactions were incubated at 37ºC for 30 minutes and then cleaned up with a Clean and Concentrate-5 column (Zymo). To elute pools, 6 µl of heated elution buffer was applied to the column and incubated for 1 minute before being spun. Eluted DNA was then passed through the column a second time, yielding approximately 5 µl for the hybridization enrichment reaction. If the pool exceeded 50 samples, the pool would be cleaned with a Clean and Concentrate-25 column, eluted in 30 µl, and concentrated to 5 µl with a SpeedVac Concentrator (Eppendorf).

For each pool, DNA, RNA, and hybridization mixtures were prepared in separate strip tubes (1 per pool). On ice, DNA mixtures were prepared by adding 5 µl of concentrated tagmented DNA from each pool, 3.4 µl of 1µg/µl mouse cot-1 (18440016, ThermoFisher) or human cot-1 (15279011, ThermoFisher), and 0.6 µl of 100 µM TIME-Seq hybridization blocking primers (IDT). RNA mixtures were prepared on ice by combining 4.25 µl of nuclease free H2O with 1 µl of Superase•In RNase inhibitor (AM2696, ThermoFisher), mixing, and then adding 0.75 µl (750 ng total) of the biotin-RNA baits. Hybridization mixtures were kept at room temperature and comprised 25 µl 20X SSPE (AM9767, ThermoFisher), 1 µl 0.5 M EDTA, 10 µl 50X Denhardt's Buffer (1% w/v Ficoll 400, 1% w/v polyvinylpyrrolidone, 1% w/v bovine serum albumin), and 13 µl of 1% SDS. Once the mixtures were prepared for each pool, the DNA mixtures were placed in a thermocycler and incubated for 5 minutes at 95ºC. Next, the thermocycler cooled to 65ºC, and the hybridization mix was added to the thermocycler. After 3 minutes at 65ºC, the RNA mix was added to the thermocycler

and incubated for 2 minutes at 65ºC. Next, the thermocycler lid was opened and, keeping all tubes in the thermocycler well, 13 µl of heated hybridization buffer was transferred to the RNA baits mixture, followed by 9 µl of the denatured TIME-Seq pooled DNA. This step was done quickly to limit temperature change during transfer, typically with a multichannel pipette for multiple pools. The combined mixtures were pipetted to mix 3-5 times, capped, and the thermocycler lid was closed. The hybridization reaction was then incubated at 65ºC for 4 hours.

To capture biotin-RNA:DNA hybrids, 125 µl of streptavidin magnetic beads were washed three times in 200 µl of binding buffer (1M NaCl, 10mM Tris-HCl pH 7.5, 1 mM EDTA) and resuspended in 200 µl of binding buffer. With the reaction still in the thermocycler, the streptavidin beads were added to the reactions and then quickly removed to room temperature. The reactions were rotated at 40 RPM for 30 minutes to allow for biotin-streptavidin binding and then placed on a magnetic separation rack (20-400, Sigma-Aldrich) until the solution was clear. Next, the beads were resuspended in 500 µl of hybridization wash buffer 1 (1X SSC, 0.1% SDS) and incubated at room temperature for 15 minutes. The beads were separated again on the magnetic separation rack and quickly resuspended in 500 µl of pre-heated 65ºC wash buffer 2 (0.1X SSC, 0.1% SDS), then incubated for 10 minutes at 65ºC. This step was repeated for a total of 3 heated washes. On the final wash, beads were magnetically separated, resuspended in 22 µl of 0.1N NaOH, and immediately removed to a MAXYMum Recovery PCR tube (PCR-02D-L-C, Axygen Scientific). After 10 minutes, beads were separated and 21 µl of the eluted DNA from each pool was moved to a new MAXYMum Recovery PCR tube.

Bisulfite conversion was done using the EpiTect Fast Bisulfite Conversion Kit (59824, Qiagen). The volume of the eluted DNA was adjusted to 40 µl using nuclease free water, 85 µl of Bisulfite Solution was added, followed by 15 µl of the DNA Protect Buffer, and the solution was mixed thoroughly. Bisulfite conversion and clean-up proceeded according to standard kit instruction. DNA was eluted in 23 µl of kit elution buffer. The initial elution was passed through the column a second time.

PCR amplification was done in a 50 µl reaction containing 23 µl of the eluted DNA, 1 µl of 25 µM P7 indexed primer, 1 µl of 25µM P5 indexed primer, and 25 µl NEB Q5U 2X Master Mix. Reactions were amplified with the following program: initial denaturation at 98ºC for 30 second; 19 cycles of 98ºC for 30 seconds, 65ºC for 30 seconds, and 72ºC for 1 minute; a final elongation at 72ºC for 3 minutes. After PCR reactions were finished, they were cleaned using 1.8X CleanNGS SPRI Beads (CNGS005, Bulldog-Bio). Library fragment size and yield was assessed using a D1000 (5067-5582, Agilent) or High Sensitivity D1000 ScreenTape (5067-5584, Agilent) on an Agilent Tapestation. Pools were combined for sequencing.

Sequencing

TIME-Seq library sequencing requires two custom sequencing primers for read 2 (Tn5 index) and index read 1 (i7 index), which were spiked into standard primers for all sequencing runs. Most deep-sequenced TIME-Seq libraries were sequenced on an Illumina MiSeq using a 150 cycle MiSeq v3 kit (MS-102-3001, Illumina) or on a NextSeq 500 using a 150 cycle NextSeq High (20024907, Illumina) or Mid (20024904, Illumina) Output v2.5 kit. The following read protocol was used for 150 cycle kits: read 1, 145-153 cycles; i7 index read, 8 cycles; i5 index read (if needed), 8 cycles; read 2, 5 cycles. This custom read protocol helps maximize sequencing efficiency and decrease cost, since the fragment size of amplified TIME-Seq libraries were typically 80-200bps and sequencing with larger kits results in a large portion of overlapping (unused) cycles. Optimization experiments, smaller deep-sequenced pools, and shallow sequencing data were typically generated by sequencing on a MiSeq using a MiSeq Reagent Micro v2 kit (MS-103-1002, Illumina) using standard paired-end and dual indexed read cycle numbers. The human HD-enriched data was generated by sequencing on a NovaSeq 6000 using a 200 cycles NovaSeq 6000 SP Reagent Kit v1.5. High GC genome from *Deinococcus radiodurans* was spiked into sequencing runs in different proportions (1-3% on MiSeq, 10% on NovaSeq, and 15-20% on NextSeq) to increase base diversity and improve sequencing quality[44]. For specific information on sequencing for each experiment, including cost and the number of samples, see Table S3.

Sequenced read demultiplexing and processing

TIME-Seq pools were demultiplexed using sabre (https://github.com/najoshi/sabre) to identify the internal Tn5 barcode with no allowed mismatches and separate reads into unique FASTQ files for each sample. Cutadapt (version 2.5) was used to trim adaptors (PE: -G AGATGTGTATAAGAGANAG -a CTNTCTCTTATACACATCT -A CTNTCTCTTATACACATCT; SE option: -a CTNTCTCTTATACACATCT). Reads were mapped using Bismark[45] (version v0.22.3; options -N 1 --non_directional) to bisulfite converted genomes (*bismark_genome_prepararation*) mm10, hg19 (iGenomes, Ilumina) or custom rDNA loci (see bait design methods), and reads were subsequently filtered using the bismark function *filter_non_conversion* (option --threshold 11). This step removes reads that are artificially fully methylated during the methylated end-repair (0.5%-3%) of the reads[44]. Typically, samples with greater than 3-5% non-conversion were excluded from analysis since this was indicative of poor tagmentation. Next, the bismark function *bismark_methylation_extractor* was used to call methylation for each sample with options to avoid overlapping reads (--no-overlap) in PE sequencing and to ignore the first 10 bps of each read (if SE: --ignore 10 and --ignore_3prime 10; if PE: --ignore_r2 10 --ignore_3prime_r2 10 as well), which precludes bias from methylated cytosines added in the Tn5 insertion gap during end-repair. Bisulfite conversion efficiency was assessed from unmethylated lambda DNA mapped to the BS-converted *Enterobacterphage lambda* genome (iGenomes, Ilumina) and was generally ≥99%.

Epigenetic clock training, testing, and analysis in validation cohort data

R (version 4.0.2) was used for all data analysis, including data organization, clock training and testing, and applying clocks to validation data. For clock training, methylation and coverage data were taken from bismark.cov files (output of *bismark_methylation_extractor*) for each sample and arranged into a data matrix with samples (rows) and CpGs (columns). For TS-rDNA clock training, the methylation matrix was filtered to comprise only CpGs with high coverage (≥ 200) in greater than 90% of samples at each CpG in the coverage matrix. For the mouse and human discovery datasets, CpGs were filtered to have at least coverage 10 in 90% of the samples.

To build epigenetic clocks from deep-sequenced TIME-Seq data, samples were randomly taken from discrete age groups (e.g., 25-55 weeks for mice) in approximately 80:20 training to testing ratio. Penalized regression models were built to predict age from methylation matrixes with the R package glmnet[46] with alpha set to 0.05 (elastic net, mostly ridge). Clocks were trained to predict age in units of weeks for mice and years for humans. After models were trained, they were applied to the testing set using the *predict* function in R. The model from a split with high Pearson correlation and low median error in the testing set was selected for application to the independent validation cohorts. This same process was applied to build the RRBS-TS-rDNA clock, filtering for CpGs in TIME-Seq data that had minimum coverage of 50 in the RRBS mouse blood clock dataset[1]. Mouse clocks were applied to validation and intervention data by organizing bismark.cov files into a methylation matrix comprising only the clock CpGs, summing the weighted methylation (e.g., methylation ratios [0-1] x clock CpG coefficient) for each sample, and adding the model's intercept. When a clock locus was not covered in a sample that passed quality filters, the missing value was replaced by the average DNAme percent for all other samples at that CpG.

***scAge*-based epigenetic age predictions from shallow TIME-Seq**

To obtain accurate epigenetic age predictions from shallow TIME-seq data, we employed the use of the recently introduced *scAge* framework[24]. Unlike to the original *scAge* application in single-cell data, all methylation values from shallow sequencing were retained unmodified (i.e., without introducing a forced binarization step for fractional methylation values). We used bulk blood- and liver-specific methylation matrices to compute linear regression equations and Pearson correlations between methylation level and age for each CpG within a particular training set. These equations were in the form:

$$f_{CpG}(age) = Met = m_{CpG} * age + b_{CpG}$$

where *age* is treated as the independent variable predicting methylation, and *m* and *b* are the slope and intercept of the CpG-specific regression line, respectively. This enabled the creation of reference linear association metrics between methylation level and age for each CpG covered in the training datasets. Five separate training datasets were used to

compute epigenetic ages in shallow sequencing data: (1) Bulk blood RRBS methylation profiles from 50 normally-fed C57BL/6J mice from the Thompson et al. (2018) study across 1,202,751 CpG sites; (2) Bulk liver RRBS methylation profiles from 29 normally-fed C57BL/6J mice from the Thompson et al. (2018) study across 1,042,996 CpG sites; (3) Bulk blood RRBS methylation profiles from 153 normally-fed C57BL/6J mice from the Petkovich et al. (2017) study across 1,918,766 CpG sites; (4) Bulk blood TIME-seq methylation profiles from 120 normally-fed C57BL/6J mice from the present study across 6,884 CpG sites; and (5) Bulk liver TIME-seq methylation profiles from 103 normally-fed C57BL/6J mice from the present study across 12,480 CpG sites. Methylation values from the Thompson et al. (2018) study were concatenated to the positive strand, as described in the original study.

We intersected individual methylation profiles of shallow TIME-seq data with the reference datasets, producing a set of n common CpGs shared across both datasets. For each shallow sample, we filtered these n CpGs based on the absolute value of their correlation with age ($|r|$) in the reference data, selecting the most age associated CpGs in every sample. We opted to use a percentile-based approach to select informative CpGs for predictions, which takes into account differential coverage across samples. Furthermore, this enabled more consistent correlation distributions among shallow TIME-Seq profiles of different coverage.

For each selected CpG per sample, we iterated through age in steps of 0.1 months from a minimum age to a maximum age value. We chose -20 months and 60 months, respectively, to encompass a wide range of possible predictions without artificially bounding outputted predicted values. We observed in our testing that all predictions generated by *scAge* for these data were within this generous range and not near the extremes. Using the linear regression formula calculated per individual CpG in a reference set, we computed the age-dependent predicted methylation, $f_{CpG}(age)$, which by the nature of the data normally lies between 0 or 1. If this predicted value was outside of the range (0, 1), it was instead replaced by 0.01 or 0.99 depending on the proximity to either value. This ensured that predicted bulk methylation values were bounded in the unit interval, corresponding to a biologically meaningful range between fully unmethylated and fully methylated.

Next, we computed the probability of observing a particular methylation state at every age in the given range based on the reference linear model estimate. For this, we calculated the absolute value of the distance between the observed methylation fraction in the shallow sample and the estimated methylation value from the reference linear model. Next, we subtracted this absolute distance from 1; hence, the closer the observed value is from that predicted by the linear model, the higher the probability of observing this state at a particular age. This provided an age-dependent probability for every common CpG retained in the algorithm.

Lastly, assuming that all CpGs in a particular sample are independent from each other, the product of each of these CpG-specific probabilities will be the overall probability of the observed methylation pattern: $P(age) = \prod_{k=1}^{n} Pr_k(age)$, where k represents individual CpGs. We then found the maximum of that product among different ages (i.e., to find the most probable age for observing that particular methylation pattern). To do this, we compute the sum across CpGs of the natural logarithm of the individual age-dependent probabilities, preventing underflow errors when many CpGs are considered. This gave us $\sum_{k=1}^{n} ln(P_k(age))$ for each age step. By harnessing the relationship of methylation level and age at many CpGs, these logarithmic sums provide a single likelihood metric for every age step within the defined bounds. We picked the age of maximum likelihood as our predictor of epigenetic age for a particular shallow TIME-Seq sample.

DATA AVAILABILITY

Raw FASTQ sequencing data will be deposited to NCBI Sequence Read Archive (SRA) and organized in a BioProject.

CODE AVAILABILITY

Code for demultiplexing and read processing as well as analysis of clocks and *scAge* prediction will be provided on Github.

ACKNOWLEDGMENTS

We would like to thank Dr. Alex Nguyen Ba and Dr. Sarah Boswell for providing the Tn5 purification and reaction protocols, as well as Dr. Chiara Ricci-Tam for providing hybridization enrichment protocols. We would like to thank the Bauer Core Facility at Harvard for sequencing and DNA extraction help as well as the Harvard Microbiology Department for allowing us access to their Illumina MiSeq. We would also like to thank the Seidman Lab at Harvard Medical School for allowing us to use their Agilent Tapestation. This work was supported by the National Institutes of Health grant to The Jackson Laboratory Nathan Shock Center of Excellence in the Basic Biology of Aging (AG038070). P.T.G was supported by the National Science Foundation Graduate Research Fellowship (DGE1745303), and A.E.K was supported by a Diamond/AFAR Postdoctoral Transition Award in Aging (DIAMOND19036). N.N.H. and M.K.E. are supported by the Intramural Research Program (IRP) of the National Institute on Aging (NIA), National Institutes of Health. HANDLS is funded by the NIA-IRP (Project Number: AG000513). This work was also funded by NIH/NIA (R01AG019719 to D.A.S; P01AG055369 to S.J.M.; AG065403 and AG047200 to V.N.G) and the Glenn Foundation for Medical Research (to D.A.S.).

AUTHOR INFORMATION

P.T.G conceived the project, designed TIME-Seq library preparations, and carried out most preliminary experiments and optimization of the TIME-Seq protocol. A.E.K organized mouse cohorts and performed frailty index and blood parameter measurement. P.T.G., A.E.K., M.S.M., A.L.M., J.L., C.C., and D.L.V. collected blood samples and extracted DNA from samples. P.T.G., A.E.K., and J.L. performed library preparations. P.T.G. developed the bioinformatic pipeline for read processing and performed most of the bioinformatic analyses. C.K. provided the coverage-filtered RRBS methylation matrix used for TS-RRBS rDNA clock training. A.T. developed an application of *scAge* to shallow-sequenced-TIME-Seq data, with input from C.K., and performed s*cAge* analyses. A.T., M.V.M., and V.N.G. provided JAX mice blood samples used for clock training. N.N.H. and M.K.E. provided human DNA samples. M.R.M., J.R.M., and S.J.M. provided dietary intervention samples and corresponding frailty data. D.A.S. provided feedback on project goals and directions. V.N.G. supervised the development and application of scAge. P.T.G. wrote the manuscript with input from D.A.S. and A.E.K. All authors edited and approved the manuscript.

ETHICS DECLARATIONS

P.T.G. and D.A.S. are named inventors on a patent application related to TIME-Seq methods filed by Harvard Medical School and licensed to Longevity Sciences. P.T.G. is an equity owner of Longevity Sciences. D.A.S is a consultant to, inventor of patents licensed to, and in some cases board member and investor in MetroBiotech, Cohbar, Life Biosciences and affiliates, Zymo, EdenRoc Sciences and affiliates, Alterity, InsideTracker, Immetas, Segterra, and Galilei Biosciences. He is also an inventor on patent applications licensed to Bayer Crops, Merck KGaA, and Elysium Health. Additional info at sinclair.hms.harvard.edu/david-sinclairs-affiliations. A.T., C.K, V.N.G. are named inventors on a patent application related to *scAge* filed by Brigham and Women's Hospital. S.J.M, A.E.K, M.R.M and J.R.M. have nothing to disclose.

REFERENCES

1 Petkovich, D. A. *et al.* Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell metabolism* **25**, 954-960.e956, doi:10.1016/j.cmet.2017.03.016 (2017).

2 Gravina, S., Dong, X., Yu, B. & Vijg, J. Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome biology* **17**, 150, doi:10.1186/s13059-016-1011-3 (2016).

3 Thompson, M. J. *et al.* A multi-tissue full lifespan epigenetic clock for mice. *Aging* **10**, 2832-2854, doi:10.18632/aging.101590 (2018).

4 Schultz, M. B. *et al.* Age and life expectancy clocks based on machine learning analysis of mouse frailty. *Nature communications* **11**, 4618, doi:10.1038/s41467-020-18446-0 (2020).

5 Bell, C. G. *et al.* DNA methylation aging clocks: challenges and recommendations. *Genome biology* **20**, 249, doi:10.1186/s13059-019-1824-y (2019).

6 Lehallier, B. *et al.* Undulating changes in human plasma proteome profiles across the lifespan. *Nat Med* **25**, 1843-1850, doi:10.1038/s41591-019-0673-2 (2019).

7 Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature reviews. Genetics* **19**, 371-384, doi:10.1038/s41576-018-0004-3 (2018).

8 Fleischer, J. G. *et al.* Predicting age from the transcriptome of human dermal fibroblasts. *Genome biology* **19**, 221, doi:10.1186/s13059-018-1599-6 (2018).

9 Horvath, S. DNA methylation age of human tissues and cell types. *Genome biology* **14**, R115, doi:10.1186/gb-2013-14-10-r115 (2013).

10 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* **49**, 359-367, doi:10.1016/j.molcel.2012.10.016 (2013).

11 Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573-591, doi:10.18632/aging.101414 (2018).

12 Meer, M. V., Podolskiy, D. I., Tyshkovskiy, A. & Gladyshev, V. N. A whole lifespan mouse multi-tissue DNA methylation clock. *eLife* **7**, doi:10.7554/eLife.40675 (2018).

13 Wang, T. *et al.* Quantitative Translation of Dog-to-Human Aging by Conserved Remodeling of the DNA Methylome. *Cell Syst* **11**, 176-185.e176, doi:10.1016/j.cels.2020.06.006 (2020).

14 Levine, M. *et al.* A rat epigenetic clock recapitulates phenotypic aging and co-localizes with heterochromatin. *eLife* **9**, doi:10.7554/eLife.59201 (2020).

15 Robeck, T. R. *et al.* Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Commun Biol* **4**, 642, doi:10.1038/s42003-021-02179-x (2021).

16 Wilkinson, G. S. *et al.* DNA methylation predicts age and provides insight into exceptional longevity of bats. *Nature communications* **12**, 1615, doi:10.1038/s41467-021-21900-2 (2021).

17 Wang, T. *et al.* Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome biology* **18**, 57, doi:10.1186/s13059-017-1186-2 (2017).

18 Horvath, S. *et al.* Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging* **10**, 1758-1775, doi:10.18632/aging.101508 (2018).

19 Lu, Y. *et al.* Reprogramming to recover youthful epigenetic information and restore vision. *Nature* **588**, 124-129, doi:10.1038/s41586-020-2975-4 (2020).

20 Kerepesi, C., Zhang, B., Lee, S. G., Trapp, A. & Gladyshev, V. N. Epigenetic clocks reveal a rejuvenation event during embryogenesis followed by aging. *Sci Adv* **7**, doi:10.1126/sciadv.abg6082 (2021).

21 Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303-327, doi:10.18632/aging.101684 (2019).

22 Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome biology* **17**, 208, doi:10.1186/s13059-016-1066-1 (2016).

23 Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic acids research* **33**, 5868-5877, doi:10.1093/nar/gki901 (2005).

24 Trapp, A., Kerepesi, C. & Gladyshev, V. N. Profiling epigenetic age in single cells. *bioRxiv*, 2021.2003.2013.435247, doi:10.1101/2021.03.13.435247 (2021).

25 Mulqueen, R. M. *et al.* Highly scalable generation of DNA methylation profiles in single cells. *Nature biotechnology* **36**, 428-431, doi:10.1038/nbt.4112 (2018).

26 Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome research* **22**, 939-946, doi:10.1101/gr.128124.111 (2012).

27 Wang, M. & Lemos, B. Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. *Genome research*, doi:10.1101/gr.241745.118 (2019).
28 Liu, Z. *et al.* Underlying features of epigenetic aging clocks in vivo and in vitro. *Aging cell*, e13229, doi:10.1111/acel.13229 (2020).
29 Whitehead, J. C. *et al.* A clinical frailty index in aging mice: comparisons with frailty index data in humans. *J Gerontol A Biol Sci Med Sci* **69**, 621-632, doi:10.1093/gerona/glt136 (2014).
30 Chitrala, K. N. *et al.* Race-specific alterations in DNA methylation among middle-aged African Americans and Whites with metabolic syndrome. *Epigenetics* **15**, 462-482, doi:10.1080/15592294.2019.1695340 (2020).
31 Evans, M. K. *et al.* Healthy aging in neighborhoods of diversity across the life span (HANDLS): overcoming barriers to implementing a longitudinal, epidemiologic, urban study of health, race, and socioeconomic status. *Ethn Dis* **20**, 267-275 (2010).
32 Malinovskaya, E. M. *et al.* Copy Number of Human Ribosomal Genes With Aging: Unchanged Mean, but Narrowed Range and Decreased Variance in Elderly Group. *Front Genet* **9**, 306, doi:10.3389/fgene.2018.00306 (2018).
33 Gu, H. *et al.* Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature protocols* **6**, 468-481, doi:10.1038/nprot.2010.190 (2011).
34 Han, Y. *et al.* Epigenetic age-predictor for mice based on three CpG sites. *eLife* **7**, doi:10.7554/eLife.37462 (2018).
35 Han, Y. *et al.* New Targeted Approaches for Epigenetic Age Predictions. *bioRxiv*, 799031, doi:10.1101/799031 (2019).
36 Han, Y. *et al.* Targeted methods for epigenetic age predictions in mice. *Scientific reports* **10**, 22439, doi:10.1038/s41598-020-79509-2 (2020).
37 Lehallier, B., Shokhirev, M. N., Wyss-Coray, T. & Johnson, A. A. Data mining of human plasma proteins generates a multitude of highly predictive aging clocks that reflect different aspects of aging. *Aging cell*, e13256, doi:10.1111/acel.13256 (2020).
38 Bruinsma, S. *et al.* Bead-linked transposomes enable a normalization-free workflow for NGS library preparation. *BMC genomics* **19**, 722, doi:10.1186/s12864-018-5096-9 (2018).
39 Ackert-Bicknell, C. L. *et al.* Aging Research Using Mouse Models. *Curr Protoc Mouse Biol* **5**, 95-133, doi:10.1002/9780470942390.mo140195 (2015).
40 Tajuddin, S. M. *et al.* Novel age-associated DNA methylation changes and epigenetic age acceleration in middle-aged African Americans and whites. *Clin Epigenetics* **11**, 119, doi:10.1186/s13148-019-0722-1 (2019).
41 Nguyen Ba, A. N. *et al.* Barcoded Bulk QTL mapping reveals highly polygenic and epistatic architecture of complex traits in yeast. *bioRxiv*, 2021.2009.2008.459513, doi:10.1101/2021.09.08.459513 (2021).
42 Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome research* **24**, 2033-2040, doi:10.1101/gr.177881.114 (2014).
43 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
44 Suzuki, M. *et al.* Whole-genome bisulfite sequencing with improved accuracy and cost. *Genome research* **28**, 1364-1371, doi:10.1101/gr.232587.117 (2018).
45 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572, doi:10.1093/bioinformatics/btr167 (2011).
46 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).

SUPPLEMENTAL FIGURES

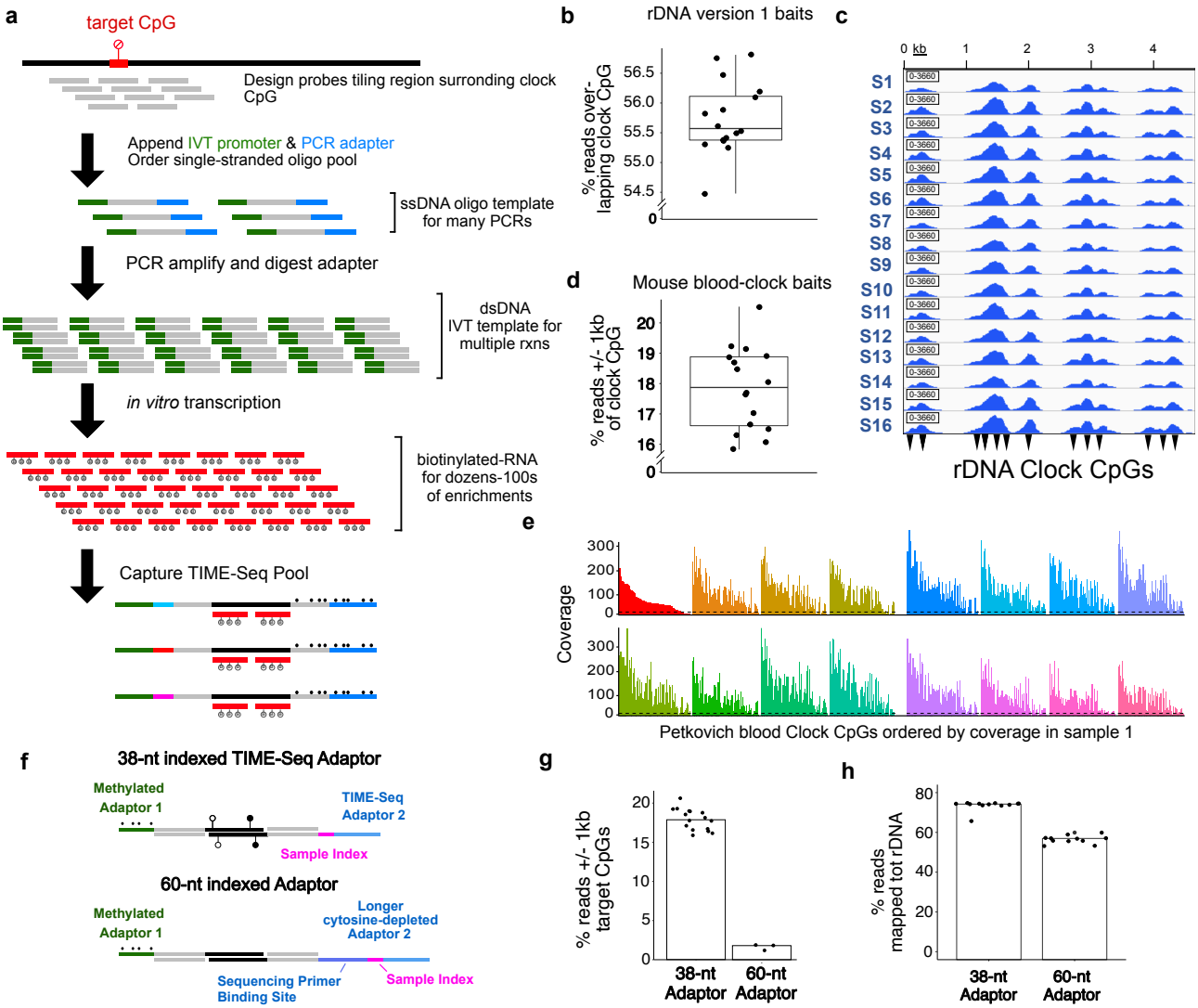

Figure S1. Biotinylated-RNA bait production and initial hybridization enrichment testing. a, Schematic of steps involved in production of biotin-RNA baits from single-stranded oligo pools for target enrichment in TIME-Seq libraries. The percent of reads overlapping target RRBS mouse rDNA clock CpGs (**b**) and an IGV browser screenshot of mapped-read pileups (**c**) using version 1 rDNA baits for enrichment of a TIME-Seq pool. Reads on-target (**d**) and mouse RRBS blood clock (Petkovich et al., 2017) CpG coverage (**e**) using mouse-blood specific baits in a pilot experiment targeting non-repetitive clock loci. Dotted line represents coverage cut-off of 10. Pools in both rDNA and blood clock pilot enrichments were sequenced with approximately 1 million paired end (PE) reads each in pools of 16 samples. **f**, Adaptor design schematic for comparison of TIME-Seq adaptors with longer barcoded adaptors. Comparison of on-target reads in short TIME-Seq and long cytosine-depleted adaptor designs for both mouse blood clock (**g**) and (**h**) rDNA (version 1) baits enrichments

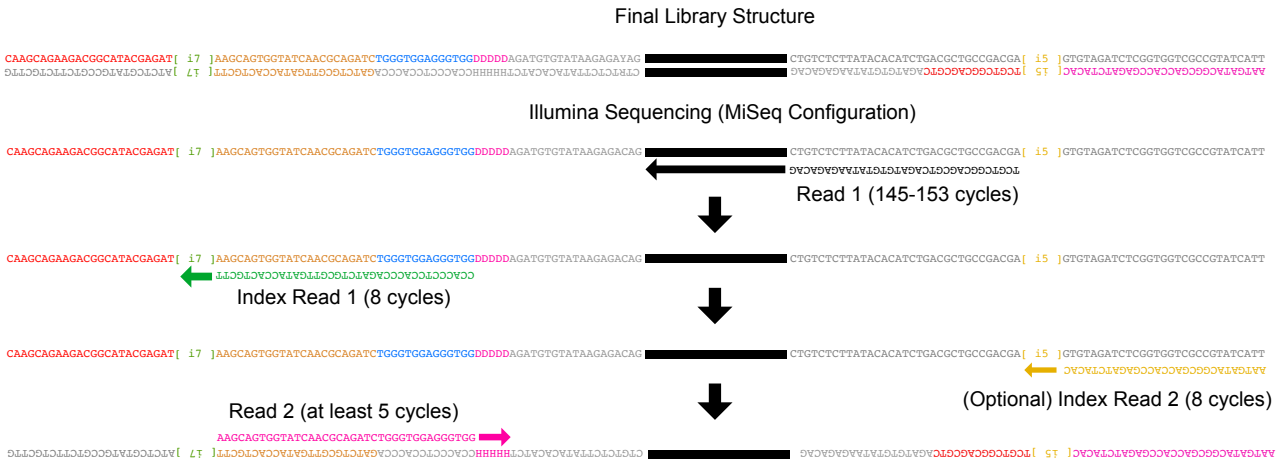

Figure S2. Final TIME-Seq library and sequencing schematic. Schematic representation of final library structure (top) and Illumina sequencing (bottom) steps required to sequence TIME-Seq libraries. Index read 1 and Read 2 primers are custom primers.

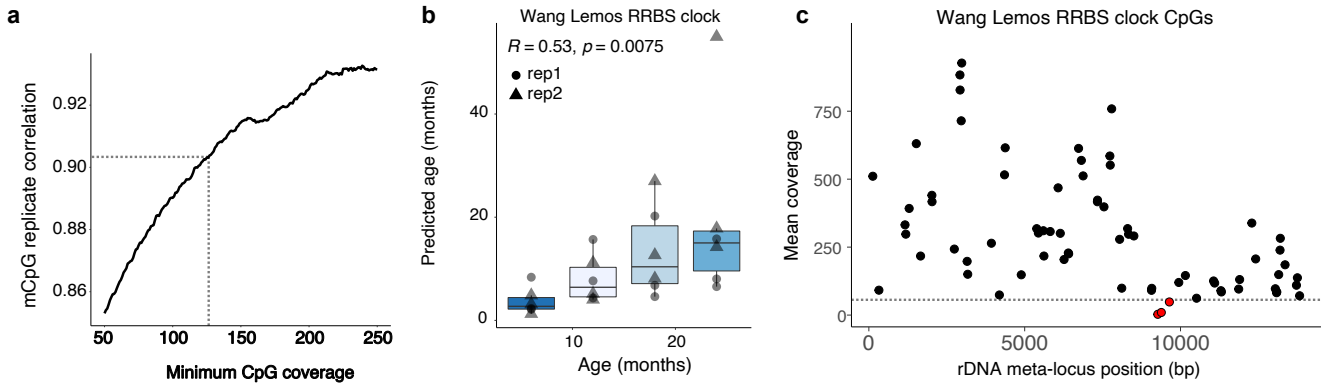

Figure S3. Pilot-experiment correlation data and age predictions using the Wang and Lemos RRBS rDNA clock. a, Replicate correlation from different coverage cutoffs in the rDNA (version 1) pilot experiment. CpG coverage of 125 is presented in Figure 1. **b**, Wang and Lemos (2019) RRBS clock predictions using TIME-Seq data enriched for clock loci. Pearson correlation and associated p-value are shown in the top left corner. **c,** Coverage of each clock locus in the original RRBS rDNA clock. CpGs shown in red have mean coverage less than 50.

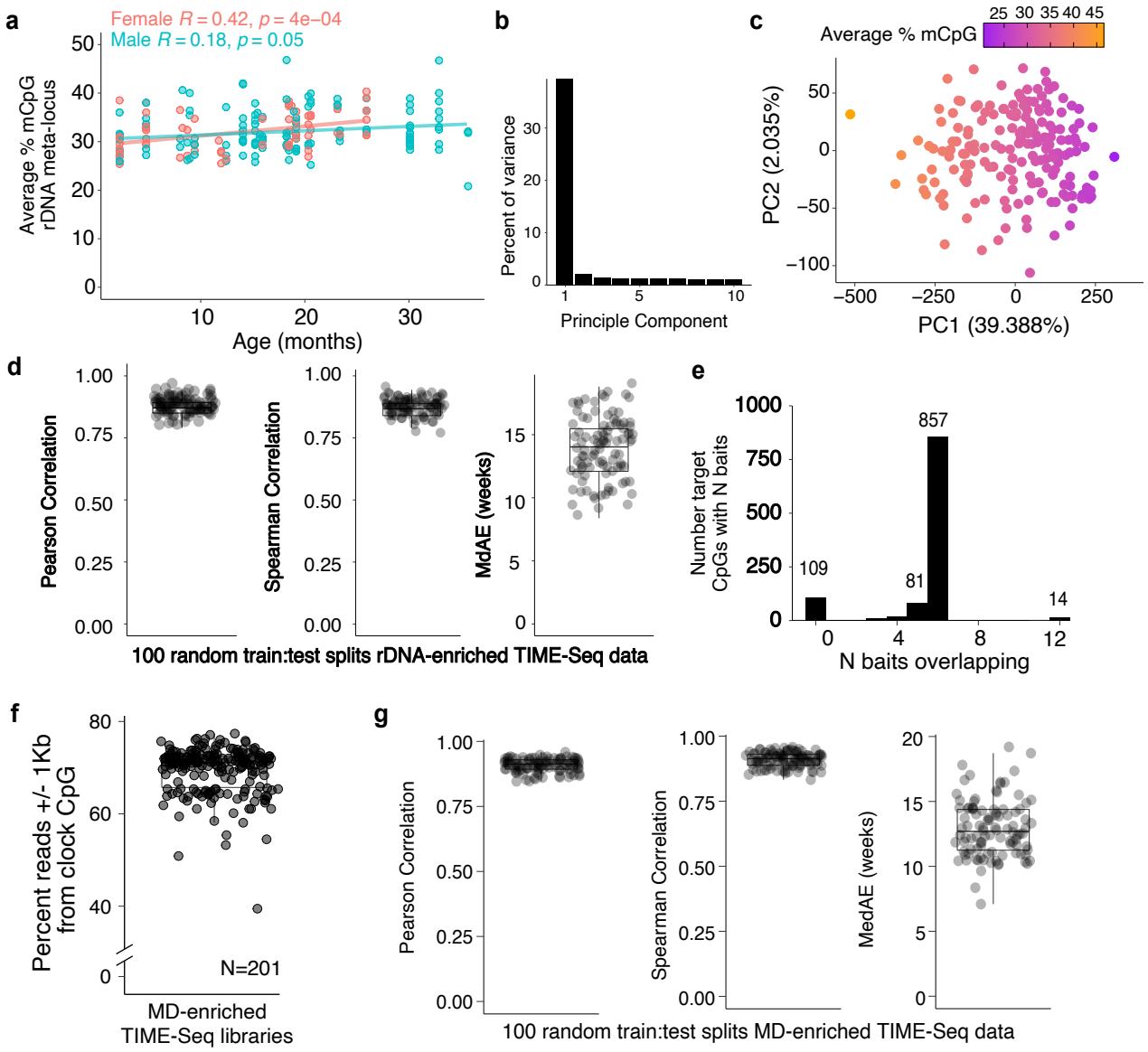

Figure S4. TIME-Seq clock training and testing using rDNA and mouse discovery baits to enrich libraries from a large cohort of mice. a, Average percent methylation at rDNA for male (N=123; blue) and female (N=68; pink) plotted against age. Pearson correlations and associated p-values are shown in the upper left corner. **b**, Eigenvalues of the first 10 principal components from TIME-Seq rDNA data. **c**, Principal component analysis showing each sample colored by average methylation. **d**, Metrics (Pearson correlation, left; Spearman correlation, middle; MedAE, right) for predicted ages from age prediction models from the testing set in 100 random splits (80 test:20 train) of rDNA data. **e**, Histogram of the number of baits overlapping each targeted clock CpG in the mouse discovery bait set used to make the TIME-Seq mouse blood clock. **f**, Percent of mapped reads within 1 Kb (+/-) of target loci from 201 mouse blood samples in TIME-Seq libraries enriched with mouse discovery baits. **g**, Metrics (Pearson correlation, left; Spearman correlation, middle, MedAE, right) for age predictions in the testing set from 100 random sample splits of 186 samples that passed quality filters when developing the TIME-Seq mouse blood clock.

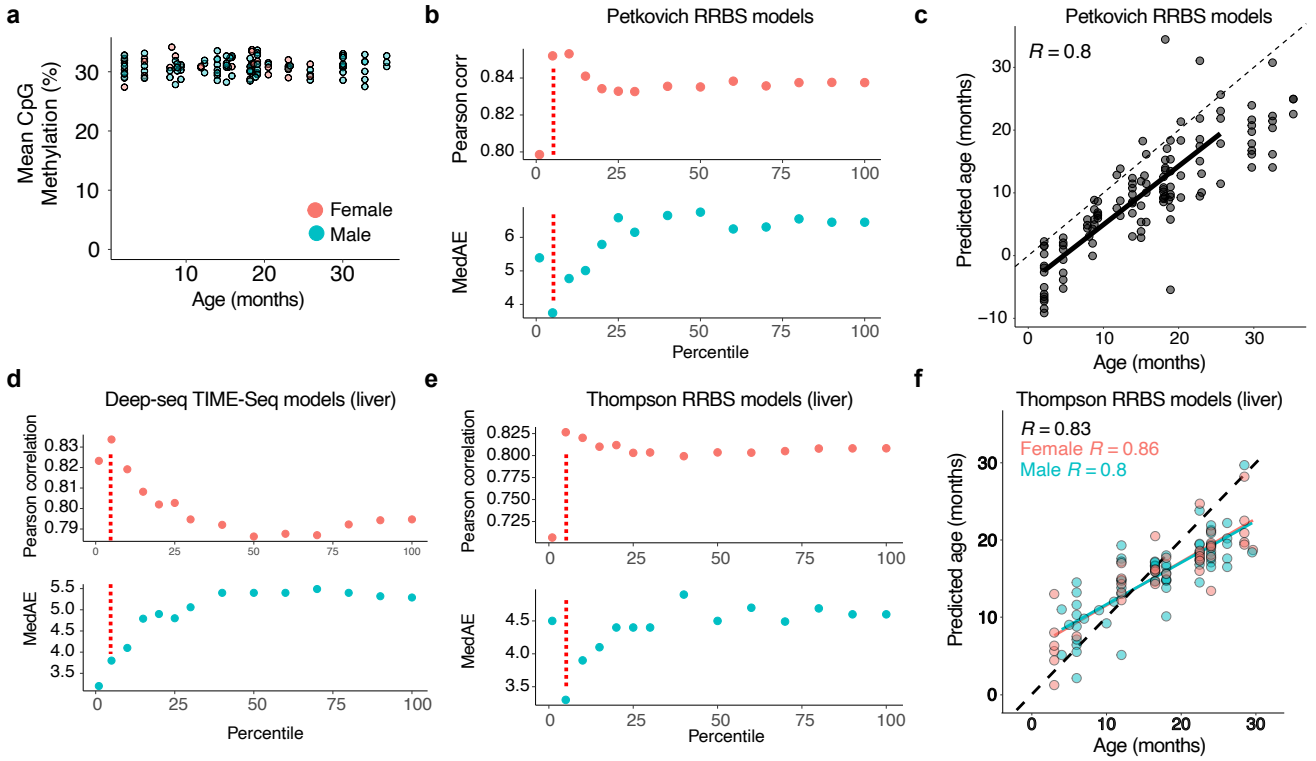

Figure S5. Data related to *scAge*-based shallow-sequencing predictions in TIME-Seq samples. a, Mean CpG methylation in shallow sequenced samples plotted against ages and colored by sex. **b**, *scAge* prediction statistics using Petkovich et al. (2017)[1] data as model CpGs. The red line indicates the percentile chosen to be represented. **c**, Predictions in TIME-Seq samples (N=119) using the top 5% of intersecting CpGs Petkovich et al. (2017) data as reference. Pearson correlation is shown in the top left corner. **d-e**, Prediction statistics in mouse liver samples using deep-sequenced TIME-Seq liver libraries (**d**) or RRBS liver data from Thompson et al. (2018)[3] (**e**) as reference data. The red line indicates the percentile chosen to be represented. **f**, Age predictions in TIME-Seq liver libraries using the top 5% of CpGs from RRBS liver data as reference. Overall and sex-specific Pearson correlation coefficients are shown in the top left.

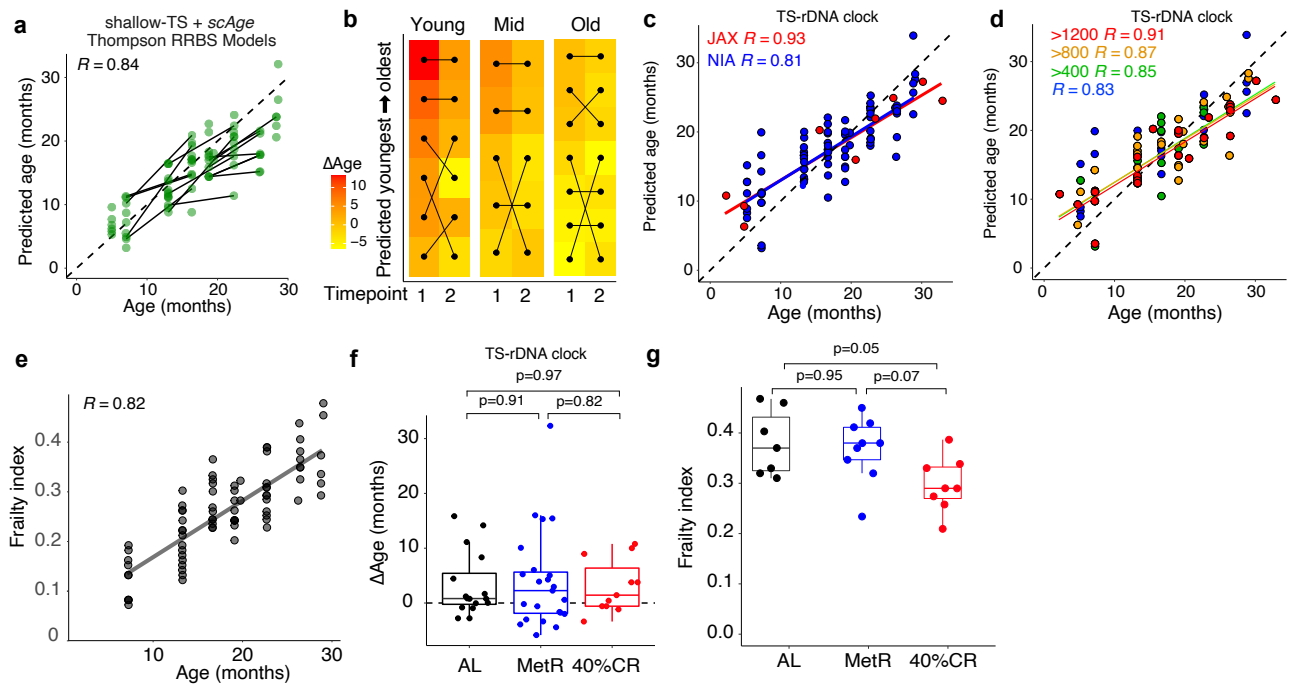

Figure S6. Validation of TIME-Seq-based age prediction approaches. a, Predictions from shallow TIME-Seq data using *scAge* with the top 20% of intersecting age-associated CpGs in RRBS data as reference models. Lines connect the same mouse at different ages. Pearson correlation is shown in the top left corner. **b**, Rank order of age prediction using TS-rDNA clock for longitudinally tracked mice at the first and second timepoints. Squares are colored by ΔAge (scale shown on the left). **c-d,** TS-rDNA clock predictions colored by the originating colony (**c**) or average coverage cutoff at clock CpGs (**d**). Pearson correlations are shown in the top left corner. **e**, Frailty index for mice used in the validation sets plotted against age. Pearson correlation is shown in the top left corner. **f**, Delta age **(**ΔAge) for TS-rDNA clock predictions in intervention samples. **g**, Frailty index for samples in intervention treatment group. Not all intervention mice were assessed for frailty. ANOVA p-values for group comparisons are shown above data in panels f and j.

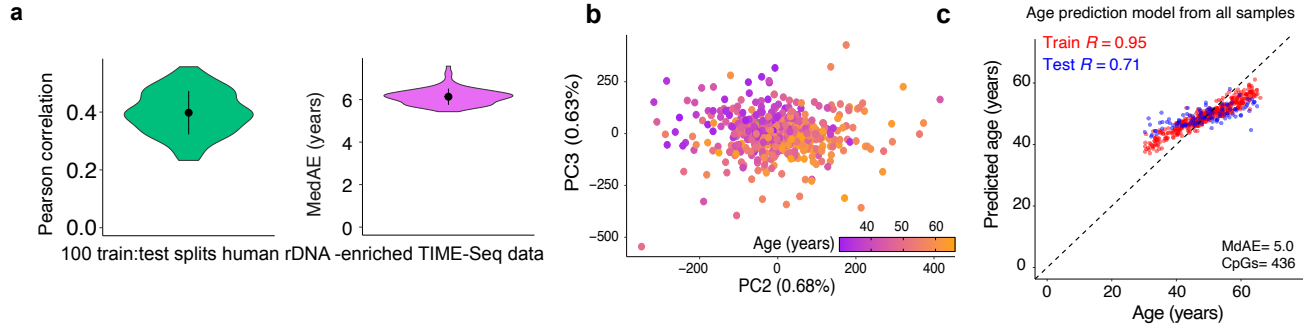

Figure S7. TIME-Seq age predictions from rDNA and genomic DNA in hundreds of human samples. **a**, Prediction metrics (MedAE, median absolute error) from the testing set of 100 random 80:20 training to testing splits of the human rDNA-enriched TIME-Seq libraries. **b**, Samples plotted using the second and third principal components from human TIME-Seq libraries enriched with human discovery baits. Samples are colored by age. **c**, Predicted age from a regression model (not a clock, defined as R>0.8 in the testing set) built to predict age using all samples that passed quality filters in the human dataset. Pearson correlations are shown in the top left corner, and MedAE (years) shown in the bottom right.

SUPPLEMENTARY TABLES

Step	Component	Subcomponents	Catalog #	Cost (USD)	Reagent Size / Conc	Volume per reaction	Samples per reaction	Cost per sample (USD)
Tagmentation	Unmethylated lambda phage DNA	-	Promega D1521	81	250µg (1µg/µL)	0.1µL	Pool (48-64 samples)	-
	Barcoded TIME-Seq transposase	Purified Tn5 transposase (1.5mg/mL)	*Homemade*	300	16mL (1.5mg/mL)	2.5µL	1 sample	$ 0.02
		TIME-Seq adaptor B (48-64)	IDT	5.38 / barcode	300µL / barcode (100µM)			
		Methylated Adaptor A		342	0.785 (100µM)			
		Tn5 reverse Oligo		47	1.4mL (100µM)			
	2X Tagmentation Buffer	*See Recipes*	*Homemade*	-	NA	12.5µL		-
	STOP Buffer	*See Recipes*	*Homemade*	-	NA	7µL		-
	Clean and Concentrator-25	-	Zymo D4034	75	50 preps	NA	Pool (48-64 samples)	$ 0.03
	Tapestation D500	D5000 Reagents	5067-5589	84.87	105 samples	1 unit	Pool (48-64 samples)	$ 0.06
		D5000 Screen Tapes	2067-5588	198	105 samples	1 unit		
End repair	NEB Klenow Fragment (3'→5' exo-) + NEB Buffer 2	-	NEB M0212L	196	0.25 (5000U/mL)	2	Pool (48-64 samples)	$ 0.03
	Methylated End Repair dNTPs Mix (5mM)	5-methyl-dCTP *(See Recipes)*	N0356S	73	100µL (10mM)	5µL		$ 0.04
		dATP (See Recipes)	NEB N0446S	176	250µL each (100mM)			
		dTTP (See Recipes)						
		dCTP (See Recipes)						
	Clean and Concentrator-5	Zymo	D4013	75	50 preps	NA	Pool (48-64 samples)	
Hybridization	Human Cot-1 DNA	-	Thermo 15279011	269	500µL (1µg/µL)	3.4µL	Pool (48-64 samples)	$ 0.01
	Blocking Primer Mix	Adaptor A Blocking Primer	IDT	4.82	300µL (100µM)	0.6µL		-
		Adaptor B Blocking Primer		5.38	300µL (100µM)			
	20xSSPE	-	Thermo 15591043	76.25	NA	25µL		-
	0.5M EDTA	-	*Homemade*	-	NA	1µL		-
	50x Denhardt's	*See Recipes*	*Homemade*	-	NA	10µL		-
	1% SDS	-	*Homemade*	-	NA	13µL		-
	SUPERase•In™ RNase Inhibitor	-	Thermo AM2696	164	125µL	1µL		$ 0.03
	Biotin-RNA Bait Pool	-	*Homemade* (Twist IVT Template)	Variable	*See methods*	1µL		$ 0.01
Biotin-RNA:DNA Pull-down	Streptavidin Magnetic Beads	-	NEB S1420S	334	5mL (4mg/mL)	125µL	Pool (48-64 samples)	$ 0.17
	Bead Binding Buffer	*See Recipes*	*Homemade*	-	NA	600µL		-
	Wash Buffer 1	*See Recipes*	*Homemade*	-	NA	500µL		-
	Wash Buffer 2	*See Recipes*	*Homemade*	-	NA	1500µL		-
	Elution Buffer (0.1N NaOH)	-	*Homemade*	-	NA	25µL		-
Bisulfite Conversion	EpiTect Fast Bisulfite Conversion Kit	-	Qiagen 59824	236	50 Preps	NA	Pool (48-64 samples)	$ 0.10
PCR Amplification and Clean-up	NEBNext® Q5U® Master Mix		NEB M0597L	125	50 reactions	25µL	Pool (48-64 samples)	$ 0.05
	20uM P7 Index [i7] PCR Primer (HPLC)	-	IDT	61	100µL (100µM)	1µL		-
	20uM P5 Index [i5] PCR Primer (HPLC)	-		56	100µL (100µM)	1µL		-
	CleanNGS SPRI DNA & RNA SPRI Bead Purification Kit		Bulldog Bio CNGS005	125	5mL	90µL	Pool (48-64 samples)	$ 0.05
	Tapestation D1000	D1000 Reagents	5067-5589	85.15	105 samples	1 unit	Pool (48-64 samples)	$ 0.06
		D1000 Screen Tapes	2067-5582	198	105 samples	1 unit		
						TOTAL	*Per sample*	*$ 0.65*
							Per pool (48 samples)	*$ 31.33*

Table S1: Estimated cost of reagents for TIME-Seq library preparation. Reagents that are estimated to be used in quantities that cost less than $0.01 (USD) per sample were excluded from the total cost per sample calculation.

Probes	Target	# unique targets	Kb targeted	IVT Promoter
Mouse ribosomal DNA v.1 baits	Wang and Lemos rDNA blood clock	21	9.5*	Sp6
Mouse blood-clock baits	Petkovich et al., mouse blood clock	59	15.7	T7
Mouse ribosomal DNA v.2 baits	Tiling mouse rDNA meta-locus	1 repetitive	13.8*	Sp6
Mouse discovery baits	Thompson et al., multi-tissue mouse clock. Meer et al., multi-tissue mouse clock. Petkovich et al., mouse blood clock	854	215	Sp6
Human discovery baits	11 previously described human clocks (listed in Liu et al.)	1289	324	T7
Human ribosomal DNA baits	Tiling human rDNA meta-locus	1 repetitive	13.8*	T7

Table S2: Information on Biotinylated-RNA bait pools used for targeted enrichment of TIME-Seq libraries. * = Kb of targeted equals the target area is multiplied by copy number. Mean ≈1400 per haploid (C57BL/6 mice); mean ≈400 per haploid (Human)

Experiment	# samples	Sequencing kit	Cost of kit (USD)	% of pool	Cost per sample (USD)
rDNA (v.1 baits) pilot (Fig. 1)	24	MiSeq v3 150 cycle	$ 922.21	10%	$ 3.84
TS-rDNA clock train / test (Fig. 2)	191	MiSeq v3 150 cycle	$ 922.21	100%	$ 4.75
TS-blood clock train / test (Fig. 2)	198	NextSeq High v2.5 150 cycle	$ 2,866.75	100%	$ 14.48
shallow TIME-Seq + ScAge blood (Fig. 3)	121	MiSeq v2 Micro 300 cycle	$ 447.41	50%	$ 1.85
shallow TIME-Seq + ScAge liver (Fig. 3)	104	MiSeq v2 Micro 300 cycle	$ 447.41	94%	$ 4.04
TS-rDNA validation set 1 (Fig. 4)	43	MiSeq v3 150 cycle	$ 922.21	34%	$ 7.29
TS-rDNA validation set 2 (Fig. 4)	53	MiSeq v2 Micro 300 cycle	$ 447.41	73%	$ 6.16
TS-blood validation set 1 + 2 (Fig. 4)	81	NextSeq High v2.5 150 cycle	$ 2,866.75	26%	$ 9.20
TS-shallow + ScAge validation set 1 and 2 (Fig. 4)	81	MiSeq v2 Micro 300 cycle	$ 922.21	27%	$ 3.07
TS-rDNA intervention (Fig. 4)	67	MiSeq v2 Micro 300 cycle	$ 447.41	82%	$ 5.48
TS-blood intervention (Fig. 4)	57	NextSeq High v2.5 150 cycle	$ 2,866.75	18%	$ 9.05
TS-shallow + ScAge intervention (Fig. 4)	57	MiSeq v3 150 cycle	$ 922.21	12%	$ 2.01
Human rDNA-enriched libraries (Fig. 5)	654	NextSeq Mid v2.5 150 cycle	$ 1,110.00	100%	$ 1.67
Human HD-enriched libraries (Fig. 5)	608	NovaSeq SP v1.5 200 cycle	$ 2,745.00	100%	$ 4.51
Total unique libraries	**2080**				

Table S3: Sample number, sequencing kit, and estimated cost of sequencing for each experiment.