

A rat epigenetic clock recapitulates phenotypic aging and co-localizes with heterochromatin-associated histone modifications

Morgan Levine^{1,2*}, Ross McDevitt³, Margarita Meer¹, Kathy Perdue³, Andrea Di Francesco³, Theresa Meade³, Colin Farrell⁴, Kyra Thrush², Meng Wang², Christopher Dunn³, Matteo Pellegrini⁴, Rafael de Cabo^{3*†}, Luigi Ferrucci^{3*†}

¹*Department of Pathology, Yale University School of Medicine, New Haven, CT, USA*

²*Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA*

³*Translational Gerontology Branch, Biomedical Research Center, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA*

⁴*Molecular Biology Institute and Departments of Energy Laboratory of Structural Biology and Molecular Medicine, and Chemistry and Biochemistry, University of California, Box 951570, Los Angeles, CA 90095-1570*

*Joint Corresponding Authors

†Joint senior authors

ABSTRACT

Aging has been shown to be a strong driver of DNA methylation changes, leading to the development of robust biomarkers in humans and more recently, in mice. This study aimed to generate a novel epigenetic clock in rats—a model with unique physical, physiological, and biochemical advantages for studying mammalian aging. Additionally, we incorporated behavioral data, unsupervised machine learning, and network analysis to identify epigenetic signals that not only track with age, but also relate to phenotypic aging and reflect higher-order molecular aging changes. We used DNAm data from reduced representation bisulfite sequencing (RRBS) to train an epigenetic age (DNAmAge) measure in Fischer 344 CDF (F344) rats. In an independent sample of n=32 F344 rats, we found that this measure correlated with age at ($r=0.93$), and related to physical functioning (5.9e-3), after adjusting for age and differential cell counts. DNAmAge was also found to correlate with age in C57BL/6 mice ($r=0.79$), and was decreased in response to caloric restriction (CR), such that the longer the animal was on a CR diet, the greater the decrease in DNAm. We also observed resetting of DNAm when kidney and lung fibroblasts were converted to induced pluripotent stem cells (iPSCs). Using weighted gene correlation network analysis (WGCNA) we identified two modules that appeared to drive our DNAmAge measure. These two modules contained CpGs in intergenic regions that showed substantial overlap with histone marks H3K9me3, H3K27me3, and E2F1 transcriptional factor binding. In moving forward, our ability to unravel the complex signals linking DNA methylation changes to functional aging would require experimental studies in model systems in which longitudinal epigenetic changes can be related to other molecular and physiological hallmarks of aging.

INTRODUCTION

Since first appearing in the literature in 2011¹, ‘epigenetic clocks’ have emerged as some of the most promising potential biomarkers of aging². Epigenetic clocks are traditionally measured by combining information on DNA methylation (DNAm) levels at hundreds of cytosine-guanine dinucleotides (CpGs) to produce age predictions. These estimates have been shown to strongly correlate with observed age (often $r > 0.7$)³⁻⁷. More importantly, the divergence in the predicted age relative to the observed age has been shown to reflect differential mortality and/or disease risk/prevalence⁸⁻¹⁵—suggesting it captures differences in biological rather than chronological aging. Furthermore, in mice, evidence is beginning to emerge to show that genetic or behavioral interventions known to influence aging produce differences in the epigenetic age of birth cohorts^{5-7,16}—leading to the growing enthusiasm surrounding the application of epigenetic clocks as powerful biomarkers in aging research.

The quest to identify measures that can assess biological age is currently a major goal in Geroscience research¹⁷⁻¹⁹. With the growing list of potential therapeutics to target aging, there is an immediate need to establish gold-standard measures for efficacy testing. Accurate biomarkers of aging will also enable better prognostic evaluation; they may inform treatment, study inclusion, and personal decisions; and perhaps most importantly, they will help uncover the underlying mechanisms driving biological aging. However, in order for a biomarker to be valuable in these endeavors, we must distinguish between changes that simply track with chronological time from those that relate to biological function (both entropic damage and/or compensatory mechanisms).

While up to this point, the vast majority of epigenetic clock studies have been conducted using human cohorts, our ability to elucidate the underlying biology of epigenetic aging will require mechanistic studies using mammalian animal models. Epigenetic clocks based on blood and multiple tissues have been developed for mouse models^{5-7,20}; however, because of their small size, drawing sufficient blood volumes for epigenetic analysis require terminal bleeds. Thus, the use of mice presents a problem when it comes to tracking epigenetic aging longitudinally or using it as a prognostic indicator. Conversely, rats share many of the same advantages as mice, yet also are approximately 10 times larger in mass, enabling the safe collection of substantially more blood at a given point in time without undue harm to the animal²¹. The ability to safely collect a larger quantity of blood will allow researchers to track animals longitudinally and relate molecular changes to changes in phenotypic characteristics—including *in vivo* brain imaging or individual differences in age-related cognitive decline, for which rats may be better models than mice²¹. Allowing for multiple assays from the same samples will enable researchers to draw links between various hallmarks of aging.

The ability to link epigenetics to other phenotypic data is of critical importance in developing biomarkers of aging. Traditionally, correlations with chronological age have been used to assess validity of aging biomarkers^{2,4-7,22,23}. However, evidence is mounting that this may not be the best approach. Indeed, the epigenetic clocks that appear most informative for predicting future health and wellness are not necessarily the strongest age predictors^{2,11}. This is due to the fact that a number of molecular and physiological traits change over the lifespan, but the degree of change does not necessarily reflect their importance in the biological aging processes manifesting as death and disease. Therefore, the only way to distinguish changes that track chronological time versus those that track biological aging is to validate potential biomarkers using variables other than age.

The aim of this study was to measure DNAm using blood samples from a large, multi-age cohort of Fischer 344 CDF (F344) rats ($n=134$) and associate DNAm patterns with chronological age along with behavioral and cellular changes. In doing so, we identified key epigenetic age changes and developed a novel rat epigenetic clock, which we show tracks age and functioning in rats. We also demonstrate that this clock can be used to track aging in mice and is responsive to interventions such as caloric restriction and cellular reprogramming.

RESULTS

Age Differences in DNAm

Male F344 rats were acquired from the NIA Aged Rodent Colony. The samples were evenly distributed over a large age range, spanning 1 to 27 months, with six rats in each age group ($n=162$). Within three weeks of arrival, rats were weighed, blood was drawn for FACS and methylation sequencing, and behavioral testing was performed. Our analytic sample included 134 rats. Excluded samples represent animals that died or were euthanized prior to blood draw, those that did not pass quality control for bisulfite sequencing, and/or those without complete FACS, rotarod, or open field data.

Whole blood from retro-orbital sampling was used to perform reduced representation bisulfite sequencing (RRBS) and imputation was conducted via a k-nearest neighbor (kNN) sliding window, yielding information on DNAm levels at 5,505,909 CpGs across the genome for all animals. To train an age predictor, samples were split into training ($n=102$) and test sets ($n=32$) based on age to ensure equal age distributions (*see Methods*). By applying elastic net regression to the training set, we built an age predictor that selected 68 of the 5.5 million CpGs. The age predictor based on the 68 CpGs, which we call DNAMAge, exhibited a correlation of $r=0.90$ with observed age in the validation sample (**Fig. 1**). As with many of the other human and rodent clocks, the relationship between DNAMAge and age exhibits nonlinear properties, suggesting a sigmoidal relationship between the two variables.

We hypothesize that the aging trends exhibited across the methylome are highly redundant, and as a result, dimensionality can be greatly reduced without compromising the underlying signal being captured. To test this, we performed principal component analysis (PCA) using the training sample to identify orthogonal signals and test for associations with age. Results suggest that PC1 only captures 6.7% of the variance in our data, and that jointly, the first ten PCs capture 28% of the variance. Nevertheless, PC1 explains 87% of the variance in age ($r=0.93$) in our independent test sample (**Fig. 1**). Furthermore, when we applied these PCs (in contrast to CpGs) to train an age predictor, based on 10-fold cross-validation in the training sample, our model suggested that age prediction was not improved by including PCs beyond PC1 alone.

DNAm Associations with FACS and Phenotypic Variables

To explore the association between DNAm and aging-related functional variables, we tested the chronological age-adjusted associations between the DNAMAge measures and performance in rotarod and open field behavioral tests of locomotor function. Data was available for eight summarized open field test variables and two rotarod variables (max time and mean time). Additionally, we also tested for confounding variables heterogeneity in blood cell types using Fluorescence Activated Cell Sorter (FACS) comprising information on 39 variables related to white blood cell composition. As with the DNAm data, given the redundancy in these measures, we ran two PCAs using the training sample—one to reduce the dimensionality of the

FACS data and one to reduce dimensionality of the behavioral phenotype data (open field and rotarod). When examining the age correlations of the first three PCs, we found that PC1 for both types of variables are strongly associated with age in the validation sample (**Fig. S1**). PC1 based on FACS data shows a correlation of $r=0.91$ with age, and the PC1 based on phenotype data (open field and rotarod) correlates with age at $r=0.63$. As shown in **Table S1**, PC1 for phenotype data is most strongly related to the open field variables, except for time spent in the center zone (an anxiety-sensitive measure) and duration of stereotypic movements, including grooming and sniffing.

Using multivariate linear regression models, we tested whether age-adjusted DNAmAge was associated with phenotypic variables and/or confounded by FACS (Table 1). In model 1, which included DNAmAge regressed on chronological age alone, as expected we observed that DNAmAge is strongly related to chronological age—similar to the correlation results. Results for Model 2, which included the addition of the phenotypic PC1, showed a significant association between DNAmAge and the physical functioning after controlling for age in the model. This suggests that among rats of the same chronological age, those with higher epigenetic age had worse locomotor function. Moreover, the addition of the phenotypic variable somewhat attenuated the chronological age effect, suggesting that part of the age-associated increase in DNAmAge is explained by declines in physical functioning (16%). In Model 3, we included PC1 from the FACS data as an additional covariate. In doing so, we observed further attenuation of the age association and a small attenuation of the phenotypic association, suggesting that cell composition may account for a small proportion of the associations between DNAmAge and either age and/or physical functioning. Nevertheless, the associations with phenotypic PC1 remained significant, suggesting that among animals of the same chronological age, age-related functional decline is related to epigenetic aging independent of cell composition. This was further supported by an expanded model in which the association between phenotypic PC1 and DNAmAge remains ($\beta=1.7$, $p=6.3e-3$), even after adjustment for the first 10 PCs from the FACS PCA (results not shown).

Table 1

	Beta Coefficient (P-value)		
	Model 1	Model 2	Model 3
DNAmAge (PC1)			
Age	0.75 (6.0e-15)	0.63 (1.27e-11)	0.42 (4.4e-4)
Pheno PC1	--	1.46 (3.4e-3)	1.29 (5.9e-3)
FACS PC1		--	1.91 (2.7e-2)

Validation in C57BL/6 Mice

To add additional out-of-sample validation, we used publicly available RRBS data from mouse models. However, given the sparsity in RRBS coverage, after mapping genomic coordinates from rat (rn6) to mouse (mm10), only 3,625 CpGs were available across all samples. Given this limitation, we re-calculated PC1 using only CpGs in common between the two species. When re-evaluating the performance in the rat data, the new DNAmAge measure from the overlapped CpGs was able to equivalently recapture the age signal. This is despite the fact that we started with a much smaller number of parameters (~3,600 compared to 5.5 million). For instance, as shown in **Fig. 2**, the new model based on PC1 explained 85% of the variance in age in the rat validation sample ($r=0.92$ correlation). Furthermore, the association between epigenetic

aging and phenotypic aging in rats remained, and if anything, was slightly more robust (**Table S2**).

To estimate DNAmAge in mice, we applied the PCA model that was calculated in the rat data using the 3,625 overlapped CpGs to the overlapping mouse RRBS data. We found that DNAmAge explains 62% of the variance in age in the mouse validation sample ($r=0.79$ correlation). Interestingly, the mouse data showed high intercept and lower slope for DNAmAge as a function of age, suggesting that this DNAmAge measure is initially over-estimated in mouse, but that the rate of increase with age is slower than for rats. However, again, we found a non-linear sigmoidal association between DNAmAge and chronological age in both the rat and the mouse validation data.

In addition to the age association, we also used the mouse data to test the effect of dietary intervention on the rat DNAmAge. Of the $n=177$ mice, 20 underwent caloric restriction (CR). Using an age-adjusted model, we found that DNAmAge was decreased in CR versus ad libitum fed animals (**Table S3**), such that CR was associated with an average decrease of 1.20 months in DNAmAge ($s.e.=0.31$, $p=1.6e-4$). Although mice were subjected to CR starting at 14 weeks, the time they remained on the diet varied from about 6.5-23 months, at which point DNAm was assessed. Therefore, we tested whether prolonged CR amplified the deceleration of DNAmAge using a model with an interaction between age and CR. Results showed that DNAmAge was significantly associated with the interaction between chronological age and CR ($\beta=-0.12$, $p=9.1e-3$). This suggests a negative correlation between CR duration and DNAmAge. To test this directly, we calculated the linear model for DNAmAge regressed on age, using only control animals. We then applied the equation to CR animals to calculate the age residual of DNAm, which we correlated with duration of CR. As shown in **Fig. S2**, duration of CR was correlated with DNAmAge (relative to the expected value based on age) at $r=-0.55$, $p=0.012$.

RRBS data generated from mouse kidney, kidney-derived induced pluripotent stem cells (iPSCs), lung, and lung-derived iPSCs was used to test whether cellular reprogramming altered DNAmAge. Results (**Fig. 3**) showed that the DNAmAge was significantly decreased in both kidney ($p=1.13e-4$) and lung ($p=6.57e-3$), by a little over 3 months. Given that the increase in DNAmAge as a function of chronological age is compressed in mice, when refitting so that the slope is equal to 1, the difference between tissue and tissue-derived iPSCs actually translates to ~20 months.

Deconstruction of Epigenetic Aging Measures

We hypothesize that the DNAmAge measure captures a composite of diverse epigenetic aging phenomena. Thus, in order to deconstruct this aging signal into distinct components from which we can better study mechanisms, we clustered CpGs based on co-methylation patterns across samples. For instance, using the 3,755 CpGs shared across the rat and mouse RRBS data, we performed weighted gene correlation network analysis (WGCNA), with the goal of identifying co-methylation modules. Co-methylation modules represent clusters of tightly related CpGs, whose DNAm values are highly correlated across samples in both the rat and the mouse data. Using a signed network with a power of 1, we identified four co-methylation modules (**Fig. S3**)—blue (64 CpGs), pink (47 CpGs), purple (44 CpGs), and green (38 CpGs). The majority of CpGs ($n=3,432$) were not assigned to a module (denoted as the grey module). Next, we tested whether certain modules contributed more to the DNAmAge measure. To do this, we compared the weights of CpGs as a function of which module they were assigned to. We found that the green, purple, and blue modules contained CpGs with higher loadings compared to the grey

module. Interestingly, the green module contained CpGs with both the highest positive and lowest negative loadings (CpGs that were strongly hypermethylated and hypomethylated with age, respectively), whereas the purple and blue modules were exclusively made up of CpGs that positively loaded on DNAmAge (CpGs that trended towards hypermethylation with age). Conversely, the pink module was made up of CpGs with loadings at or near 0, suggesting that these CpGs did not contribute much to the overall DNAmAge score (**Fig. 4**).

Given the high loadings in the green, blue and purple modules—suggesting that CpGs in these modules may be driving the original DNAmAge measure—we generate DNAmAge measures based on CpGs within each module (module-specific DNAmAge measures). This was done using the same PCA method in which we constructed the original score (based on all 3,625 CpGs), but in this case, it was carried out one-by-one for each module, restricting to CpGs in that module. As seen in **Fig S3**, high age correlations were observed for the green module ($r_{rat}=0.86$, $r_{mouse}=0.58$), the purple module ($r_{rat}=0.88$, $r_{mouse}=0.68$), and the grey module ($r_{rat}=0.89$, $r_{mouse}=0.84$), while moderate age correlations were observed for the blue module ($r_{rat}=0.33$, $r_{mouse}=0.39$), and non-significant age correlations were found for the pink module ($r_{rat}=-0.05$, $r_{mouse}=0.11$).

We next evaluated the associations between module-specific DNAmAge (excluding grey) and the age-related physical functioning in rats, CR in mice, and reprogramming in lung and kidney fibroblasts. Results using a fully adjusted linear model (controlling for age and FACS PC1-5, **Table S2**) suggest that the green module DNAmAge ($\beta=0.51$, $p=4.9e-3$) and the blue module DNAmAge ($\beta=0.34$, $p=3.1e-3$) were positively associated with age-related physical functioning, such that higher DNAmAge in these two modules (accounting for chronological age and cell composition) was associated with a more functionally impaired phenotype in rats (**Table S2**). When comparing module-specific DNAmAge measures on the basis of CR, we found that all four modules were significantly reduced in CR versus control mice ($p_{blue}=2.5e-8$, $p_{pink}=2.9e-10$, $p_{green}=5.2e-4$, $p_{purple}=2.2e-3$). Finally, only the blue and green DNAmAges showed significant decline after reprogramming in both kidney ($p_{blue}=5.2e-4$, $p_{green}=2.3e-4$) and lung ($p_{blue}=7.6e-3$, $p_{green}=1.8e-2$).

Genomic Features of Methylation Modules

In order to identify underlying epigenetic mechanisms associated with the four modules, we used the Cistrome Project database (<http://cistrome.org>) to assess enrichment for binding overlap of transcription factors (TFs), chromatic regulators, histone marks, and variants. The most striking results were for CpGs in the blue module, which exhibited substantial overlap with the repressive histone modification marks H3K9me3 and H3K27me3 and E2F1 TF binding. The green module also exhibited high enrichment for H3k9me3 and H3k27me3, but to a somewhat lesser degree. Although other modules showed slight enrichments for TFs, chromatic regulators, and histone marks, they were substantially lower than what was observed in the blue module (Supplemental Material).

Next, we tested for enrichments based on UCSC Genome Browser features. We found that as much as 98% of the CpGs in the blue module and 55% of the CpGs in the green module were located far from TSS (10^3 to 10^6 downstream (**Fig 5A**)). Conversely, only about 10% of CpGs in the overall data ($n=3,625$) were located within this proximity to TSS. Similarly, both the blue and green modules were substantially enriched for CpGs in intergenic regions—nearly every CpG in the blue module is in an intergenic region, as are approximately 60% of CpGs in the green module (**Fig 5B**). This represented enrichments of 7.4-fold and 4.5-fold for the blue

and green modules, respectively, relative to all CpGs in the data (only ~13% were found to be located in intergenic regions).

Trends in CpG density was another striking feature that differentiated modules. For instance, results showed that the blue module was highly skewed towards CpGs in CG dense regions (**Fig 5C-D**). The green module showed bimodal distribution, such that about half the CpGs it contains were in highly dense regions and the other half were in very sparse regions. When examining the behaviors of all the CpGs in the models, we found that the CpGs from the green module, in the sparse regions, were those that became hypomethylated with age yet contributed substantially to the overall DNAmAge score—as indicated by their strong negative loadings (**Fig 5D**). In contrast, the green module CpGs in dense regions had high positive loadings, suggesting that hypermethylation of these CpGs was strongly reflective of accelerated aging in the DNAmAge variable. Similarly, the blue and purple modules were made up of CpGs for which hypermethylation was associated with aging, but the difference was that the blue module contained CpGs in dense regions, whereas the purple module contained CpGs in sparser regions. Finally, the pink module contained CpGs that did not appear to have a strong aging signal, nor have defining characteristics in regards to genomic location.

DISCUSSION

This study presents the first epigenetic aging clock developed for rats. While it shares many of the same experimental advantages with its close relative, the mouse, the rat represents an exciting rodent model for studying mammalian aging. Its larger size better enables longitudinal tracking and simultaneous assessment of multiple assays and it is also better suited to *in vivo* brain scanning and performing assessments to study heterogeneity in cognitive aging²¹. In moving forward, our ability to track how molecular features change over time in accordance with functional phenotypes will be critical for disentangling the causal pathways through which aging hallmarks directly contribute to morbidity and mortality. It will also better enable assessments of geroprotective therapeutics given that samples can be taken pre- and post-treatment without causing undue harm to the animal.

Our novel epigenetic clock was shown to strongly correlate with chronological age and relate to physical performance, independent of cell composition and/or chronological age. Furthermore, we show that it is also applicable to mice. For instance, we estimated our epigenetic clock in whole blood samples from C57BL/6 and found that it correlates with age at $r=0.79$. Moreover, we found that it is decreased in animals that have undergone CR. This is consistent with previous studies showing the CR prevent DNA methylation drift and epigenetic age in mice^{24,25}. We further observed, that the longer the animals remained on a CR diet, the lower their DNAmAge was relative to their chronological age, suggesting that the longer CR is maintained, the greater the decrease in epigenetic age.

Nevertheless, there were also interesting observations that differentiated the performance of our DNAmAge measure in mice versus rats. Most notably, the functional form of DNAmAge regressed on age. While DNAmAge exhibits a sigmoidal fit with age in rats, this is more dramatic in the mouse data. Furthermore, the measure appears compressed at both younger and older ages, such that the range in mouse chronological age is 0.66-34.4 months (mean=17.0, $sd=10.8$), yet the range in DNAmAge is 11.1-28.8 months (mean=16.9, $sd=2.8$). The non-linear fit of DNAmAge as a function of age has been repeatedly reported in both human and rodent samples². One possibility for this observation is that the changes in DNAm reflect other non-linear cellular changes, such as cellular division or mutation accumulation, which increase

exponentially in early life and then decelerate upon maturation²⁶. Relatedly, the proportion of senescent and/or precancerous cells may accumulate in a non-linear manner with age, such that these changes are slow in early life and then accelerate towards the end of the lifespan²⁷. The combination of these non-linear patterns in composite DNAAge measure may therefore produce the sigmoidal age trajectories that have been observed. When considering the compression of DNAAge across age in mice versus rats, one explanation is that some of the signals captured in epigenetic clocks are conserved, while others are not. If not all age-related DNA changes in rats are conserved in mice this could offset, distend, or constrict the DNAAge measure when applied to mice, even if the overall score still exhibits a significant age correlation.

To understand the “types” of age-related DNA changes that comprise the epigenetic clock, we performed weighted network analysis, which enabled us to identify co-methylation modules²⁸. These modules represent groups of CpGs with highly correlated DNA levels across samples that are thought to capture specific epigenetic aging phenomena. Five modules were identified, four of which exhibited significant age correlations in both rats and mice. However, our ability to relate the modules to phenotypic data—rather than age alone—allowed us to identify the most promising aging signals. For instance, the two modules with the strongest age correlations (grey and purple) were not related to physical functioning in rats. Conversely, the green and blue modules, which had more conservative age correlations, were related to physical functioning in rats, independent of age. Interestingly, they exhibited deceleration in response to CR and were reset upon reprogramming. The majority of epigenetic clock studies evaluate measures on the basis of their age correlations. Here, we were able to start dissecting signals that are purely age-derived from those that are related to phenotypic changes and, therefore, are more relevant for biological aging. Had we relied exclusively on the robustness of the age predictions, it would have been impossible to pinpoint these two modules as being the most informative.

When examining their characteristics, we found that CpGs in the blue and green modules tended to lie in intergenic regions that were more than 10k bp downstream of TSS. Thus, these signatures are unlikely to represent promoter hypermethylation that directly repress gene transcription. Moreover, enrichment analyses showed that CpGs in these two modules tended to co-locate with H3K9me3, H3K27me3, and E2F1 TF, which suggests they are likely located in constitutive heterochromatin and/or facultative heterochromatin domains²⁹. Interestingly, these patterns are reminiscent of an epigenetic landscape that has been associated with cellular senescence—another major hallmark of aging. In the paper by Narita et al.³⁰, the authors described a distinct heterochromatic profile of senescent cells, which is characterized by heterochromatin protein recruitment (including H3K9me), and stable repression of E2F target genes, which have essential roles in cell cycle control.

Overall, we were able to develop a composite DNAAge measure in rats that was associated with age and with physical functioning (independent of age). It was also found to be conserved in mice and exhibited robust responses to aging interventions—CR and cellular reprogramming. Unlike previous epigenetic clocks, our novel clock was built using PCA, rather than elastic net or other supervised machine learning methods. Interestingly, while we considered the additive effects of multiple PCs in generating the DNAAge measure, PC1 alone was sufficient to produce a highly age correlated variable in both the training and validation datasets. One explanation is that the F344 rats used in this study were highly homogenous—with the same sex, genetic background, and environment. Therefore, the main variance to be explained was age, which was subsequently captured in PC1. Because the rats may be aging in the same manner

without differencing aging patterns/phenotypes, additional PCs were not picking-up age-related changes. That being said, we hypothesize that PCs beyond PC1 will add additional information when assessing aging in heterogenous populations—genetic diverse groups (e.g. humans), environmentally heterogeneous populations, diverse tissues, etc.

In moving forward, our ability to continue to unravel the complex signal linking DNAm changes to functional aging outcomes will require experiments in diverse model systems in which epigenetic changes can be both tracked longitudinally and related to other informative molecular and physiological hallmarks of aging.

METHODS

Animals

All experimental procedures were conducted in accordance with the Guide for the Care and Use of Laboratory Animals and approved by the NIA Animal Care and Use Committee. Male F344 rats were obtained from the NIA Aged Rodent Colony housed at the Charles River Laboratories (Frederick, MD). After receipt into NIA intramural housing facility (Baltimore, MD), animals were housed with Nylabone supplementation and ad libitum access to food (NIH-31 diet) and water. Rats younger than 3 months were housed in groups of three; all other rats were singly housed. Rats were maintained on a 12/12 lighting schedule, with all procedures carried out during the animals' light cycle. Rats were habituated to the facility for at least 3 days before 500 μ l of whole blood was collected via retro-orbital bleedings for DNA and FACS analysis. Blood for DNA was collected in heparinized tubes, spun, and the plasma removed; buffy coat and red blood cells were frozen at -80°C until DNA extraction. Blood for FACS analysis was collected in EDTA-treated tubes, chilled on ice, and tested immediately.

The RRBS data from mouse blood and iPSC samples were acquired from Gene Expression Omnibus, under the accession number GSE80672.

Behavior

Rats were recovered from bleeds for at least eight days prior to behavioral testing. For open field test, rats were placed for 30 min in a 42x42x33cm Plexiglas chamber. Activity was monitored through infrared beams detected with AccuScan Fusion software (Omnitech Electronics; Columbus, OH). On a separate day, rats were tested in a rotarod apparatus (MED-Associates; St Albans, VT) consisting of a 7-cm rotating drum placed 60 cm above a horizontal surface. Rats were placed on the drum rotating at a constant 8 RPM for 5 min; rats that fell off during that time were placed back on. Rats were then tested in three trials with accelerating drum rotation (4-40 RPM) and tested for latency to fall up to 5-min maximum trial duration. Rats recovered in their homecage for 15 min between trials. Nails of hindlimbs were trimmed in old rats (18+ mo-old) one day prior to testing.

FACS

One hundred μ L of whole rat blood from chilled EDTA-treated tubes was stained and then processed using a Beckman Coulter TQ-prep and the Beckman Coulter immunoprep reagent system. Immunophenotyping data was acquired on a BD FACSCanto II and analyzed using BD FACSDiva. Antibodies used for fluorescence analysis are as follows: FITC-conjugated anti-rat CD3 (clone 1F4), PE-conjugated anti-rat CD25 (clone OX-39), PerCP-conjugated anti-rat CD8a (clone OX-8), PE-Cy7 conjugated anti-rat CD11b/c (clone OX-42), APC-Cy7 conjugated anti-rat CD4 (clone W3/25) from Biolegend (City, state), and AF647-conjugated anti-rat RT1B (clone

OX-6), BV421-conjugated anti-rat CD45RA (clone OX-33), and BV605-conjugated anti-rat CD45 (clone OX-1) from BD Biosciences (City, state).

DNA Analysis, RRBS Libraries, Sequencing Alignment, and Methylation Matrix Assembly

Following proteinase K and RNase A treatments, DNA was isolated from rat blood cells using QIAamp DNA Mini Kit (Qiagen, City, state) following manufacturer's instructions using a QIAcube automated device. DNA was eluted from columns in 200 μ L of AE buffer, concentrated in 20 μ L 10 mM Tris-HCl, pH 8.5, 0.1 mM EDTA using Genomic DNA Clean & Concentrator-10 (Zymo, City, state), and quantified using a Qubit 2.0 (Life Technologies, City, state). RRBS libraries were generated as described previously^{31,32}. Briefly, 100 ng of isolated DNA was digested with MspI restriction enzyme (NEB, Ipswich, MA), carried out end-repair/adenylation (NEB) and ligation with TruSeq barcoded adapters (Illumina, San Diego, CA). DNA fragment length between 200-300bp were selected with SPRI magnetic beads (ABM, Richmond, BC, Canada), followed by bisulfite treatment (Millipore, Billerica, MA), and PCR amplification (Bioline, Taunton, MA). The libraries were sequenced by multiplexing 8 libraries per lane on the Illumina HiSeq2500 sequencer, with 100bp single-end reads.

Sequencing reads were trimmed using CutAdapt³³ to remove adapter sequences. Trimmed reads were mapped and methylation beta values called from the alignment file using BS-Seeker2³⁴. Aligned data were filtered to keep only CpG sites with greater than 10x coverage in at least 75% of the samples. Missing values were imputed using a kNN sliding window; missing methylation values were assigned the average value of the 5 nearest neighbors by Euclidean distance within a 3Mb window.

Statistical Analysis

Elastic net penalized regression was used to generate a DNAm predictor of chronological age in rats based on RRBS data containing approximately 5.5 million CpGs. Prior to training, rats were grouped into training and testing sample based on age. For instance, rats ages 2m, 6m, 10m, 14m, 18m, 22m, and 26m were selected for the test set (n=32), while all others were included in the training set (n=104). Next PCA was run using the rats in the training set and elastic net was used to train and predictor of age based on PCs rather than individual CpGs. Thus, each PC contained the information from all CpGs (weighted according to contribution). Both the PC based and the CpG based DNAmAge measures were evaluated in the test sample based on biweight midcorrelations with age and multivariate linear regression analysis to assess the association between them and PC1 from a PCA of phenotypic variables (i.e. rotarod and open field), after adjusting for age and cell composition (PC1-5 from PCA of FACS data).

LiftOver chain file from rn6 to mm10 genome assembly was downloaded from UCSC genome browser (<https://genome.ucsc.edu/index.html>) to provide the alignment from rat to mouse genome. rtracklayer liftOver function (<https://www.bioconductor.org/help/workflows/liftOver/>) was used to load the chain file and map the rat CpG to corresponding mouse genome coordinates. We then assessed CpG overlap between the mouse and rat data and found that 3,625 of the 5.5 million CpGs were available in all datasets. As a result, we reran PCA in the training rat sample and fit a new DNAmAge measure based on the PCs from the 3,625 CpGs. We also re-evaluated age correlations and associations with phenotypic variables in the rat test data. We then fit this DNAmAge predictor in the mouse data and used biweight midcorrelations to assess age associations, and multivariate

linear regression analysis to assess the effect of CR on DNAmAge, relative to chronological age, as well as the effect of iPSC reprogramming in mouse lung and kidney fibroblasts.

To identify co-methylation modules, we applied consensus WGCNA to the mouse and rat RRBS data. For this analysis, we input data as a “signed” topological overlap matrix, with a power=1. We also used deepSplit=1, minModuleSize=25, and minKMEtoStay=0.4. After grouping CpGs in modules, we generated module-specific DNAmAge measures based on PCA. We then tested for age association and phenotype associations in a manner consistent with models we have previously run.

Finally, LOLA (<http://databio.org/regiondb>) and Cistrome (<http://cistrome.org>) were used to assess enrichment for binding overlap of transcription factors, chromatic regulators, histone marks, variants, and genomic locations. For enrichment we used a background comprising all 3,625 CpGs from the overlapped mouse and rat RRBS data. Enrichment was conducted independently for each of the four modules (excluding grey). Mouse (mm10) region sets were used for all enrichment analysis, given that Cistrome and LOLA only contain databases for human and mouse.

REFERENCES

- 1 Bocklandt, S. *et al.* Epigenetic predictor of age. *PLoS One.* **6**, doi:10.1371/journal.pone.0014821 (2011).
- 2 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).
- 3 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115, doi:10.1186/gb-2013-14-10-r115 (2013).
- 4 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* **49**, doi:10.1016/j.molcel.2012.10.016 (2013).
- 5 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).
- 6 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.
- 7 Thompson, M. J. *et al.* A multi-tissue full lifespan epigenetic clock for mice. *Aging (Albany NY)* **10**, 2832-2854, doi:10.18632/aging.101590 (2018).
- 8 Levine, M., Lu, A., Bennett, D. & Horvath, S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging (Albany NY)* **Dec** (2015).
- 9 Levine, M. E. *et al.* DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany NY)* **7**, 690-700, doi:10.18632/aging.100809 (2015).
- 10 Levine, M. E. *et al.* Menopause accelerates biological aging. *Proc Natl Acad Sci U S A* **113**, 9327-9332, doi:10.1073/pnas.1604558113 (2016).
- 11 Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* **10**, 573-591, doi:10.18632/aging.101414 (2018).

12 Ambatipudi, S. *et al.* DNA methylome analysis identifies accelerated epigenetic ageing associated with postmenopausal breast cancer susceptibility. *Eur J Cancer* **75**, 299-307, doi:10.1016/j.ejca.2017.01.014 (2017).

13 Horvath, S. *et al.* Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci U S A* **111**, 15538-15543, doi:10.1073/pnas.1412759111 (2014).

14 Horvath, S. *et al.* Accelerated Epigenetic Aging in Down Syndrome. *Aging Cell* **14**, doi:DOI: 10.1111/acel.12325 (2015).

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

16 Sziraki, A., Tyshkovskiy, A. & Gladyshev, V. N. Global remodeling of the mouse DNA methylome during aging and in response to calorie restriction. *Aging Cell* **17**, e12738, doi:10.1111/acel.12738 (2018).

17 Justice, J. N. *et al.* A framework for selection of blood-based biomarkers for geroscience-guided clinical trials: report from the TAME Biomarkers Workgroup. *Geroscience* **40**, 419-436, doi:10.1007/s11357-018-0042-y (2018).

18 Kennedy, B. K. *et al.* Geroscience: linking aging to chronic disease. *Cell* **159**, 709-713 (2014).

19 Sierra, F. The Emergence of Geroscience as an Interdisciplinary Approach to the Enhancement of Health Span and Life Span. *Cold Spring Harb Perspect Med* **6**, a025163, doi:10.1101/cshperspect.a025163 (2016).

20 Stubbs, T. M. *et al.* Multi-tissue DNA methylation age predictor in mouse. *Genome Biology* **18**, 68, doi:10.1186/s13059-017-1203-5 (2017).

21 Ellenbroek, B. & Youn, J. Rodent models in neuroscience research: is it a rat race? *Dis Model Mech* **9**, 1079-1087, doi:10.1242/dmm.026120 (2016).

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

23 Stubbs, T. M. *et al.* Multi-tissue DNA methylation age predictor in mouse. *Genome Biol* **18**, 68, doi:10.1186/s13059-017-1203-5 (2017).

24 Maegawa, S. *et al.* Caloric restriction delays age-related methylation drift. *Nat Commun* **8**, 539, doi:10.1038/s41467-017-00607-3 (2017).

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

26 Rozhok, A. I. & DeGregori, J. The evolution of lifespan and age-dependent cancer risk. *Trends Cancer* **2**, 552-560, doi:10.1016/j.trecan.2016.09.004 (2016).

27 Zietkiewicz, E., Wojda, A. & Witt, M. Cytogenetic perspective of ageing and longevity in men and women. *J Appl Genet* **50**, 261-273, doi:10.1007/BF03195682 (2009).

28 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

29 Saksouk, N., Simboeck, E. & Dejardin, J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* **8**, 3, doi:10.1186/1756-8935-8-3 (2015).

30 Narita, M. *et al.* Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell* **113**, 703-716, doi:10.1016/s0092-8674(03)00401-x (2003).

31 Orozco, L. D. *et al.* Epigenome-wide association of liver methylation patterns and complex metabolic traits in mice. *Cell Metab* **21**, 905-917, doi:10.1016/j.cmet.2015.04.025 (2015).

32 Smith, Z. D., Gu, H., Bock, C., Gnirke, A. & Meissner, A. High-throughput bisulfite sequencing in mammalian genomes. *Methods* **48**, 226-232, doi:10.1016/j.ymeth.2009.05.003 (2009).

33 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3, doi:10.14806/ej.17.1.200 (2011).

34 Guo, W. *et al.* BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14**, 774, doi:10.1186/1471-2164-14-774 (2013).

FIGURE LEGENDS

Figure 1: Principal Component Analysis (PCA) Based Construction of DNAmAge in rats

A) Screeplot of variance explained for PCs1-60. A large amount of the variance is captured by PCs 1 and 2, with an elbow forming around PC7, suggesting that the amount of variance explained drops-off significantly after the first six PCs.

B) Supervised approach: A sample of male F344 rats, ages 1 to 27 months, were split into training (n=102) and test sets (n=32). Using DNAm levels measured at 5,505,909 CpGs across the genome as input, we applied elastic net penalized regression to train a predictor of age. Based on training data, 68 CpGs were selected for the DNAmAge measure. The plot shows the correlation between age (x-axis) and DNAmAge (y-axis) in the test sample.

C) Unsupervised Approach: We converted PC1 (estimated based on PCA in the training sample) to units of years by regressing age on PC1 and then multiplying the coefficient by PC1 and adding the constant in the test sample. We then tested the correlation between PC1 DNAmAge and chronological age.

Figure 2: Age correlation in rats and mice using the restricted overlapping CpGs

A) PC1 DNAmAge was re-estimated using only the CpGs that were overlapped between rat and mouse (n=3,625). Although this variable contained only 0.066% of the CpGs in the original measure, we were able to observe an equivalent age correlation.

B) This variable was then applied to data from C57BL/6 mice and was found to be strongly correlated with age. We also observed the calorie restricted mice (magenta) trended towards lower DNAmAge than ad libitum fed animals (green).

Figure 3: Rejuvenation of fibroblast-derived iPSCs

We applied our DNAmAge measure to data from kidney and lung fibroblast controls and derived iPSCs (GSE80672). We find that a significant reduction in DNAmAge for lung- and kidney-derived iPSCs versus controls.

Fig 4: PC1 Loadings (for Overlapped CpGs from mouse data) by module.

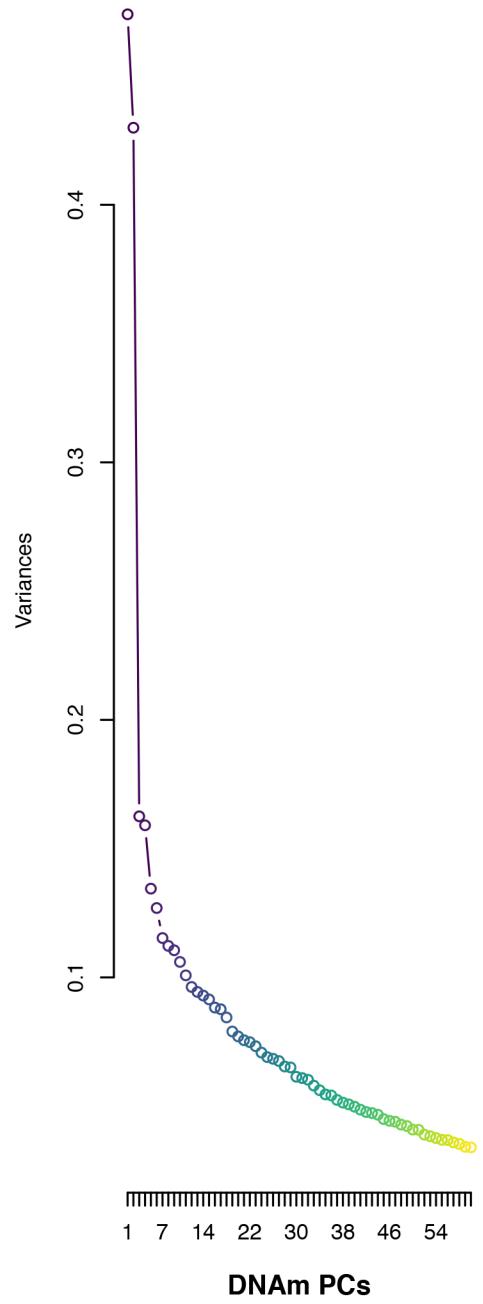
Each of the 3,625 CpGs was assigned to a module (i.e. green, purple, pink, blue, grey). The y-axis shows the loadings from PC1 that was used to signify the DNAmAge measure. The Purple and Blue modules tend to have CpGs with positive loadings, signifying that DNAm levels for CpGs in these modules were more strongly related to higher DNAmAge. The green module had CpGs with both high positive and high negative loadings, suggesting that half the CpGs in this

module are hypomethylated in accordance with higher DNAmAge whereas the other half are hypermethylated with higher DNAmAge.

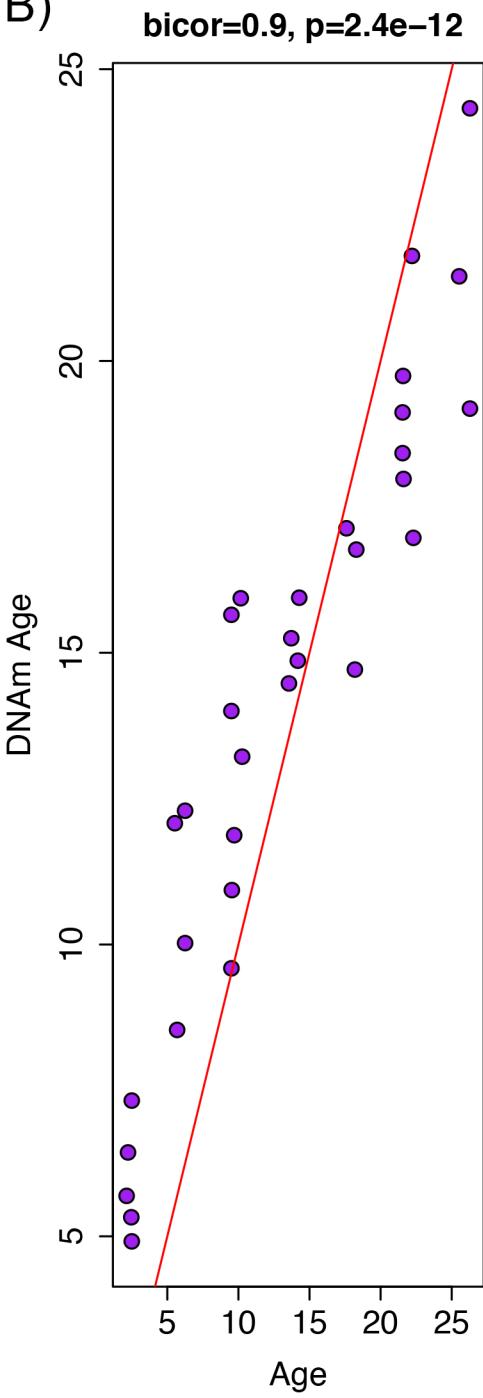
Fig 5: Module characteristics.

- A) Base-pair proximities of CpGs to Transcription Start Sites (TSS), according to module assignment (denoted by color). We find a high proportion of CpGs in the blue module (and to some extent the green module) are located 10,000 to 1,000,000 downstream of TSS.
- B) Proportion of CpGs in various genomic regions as a function of module (denoted by color). Results suggest that 98% of CpGs in the blue module and 60% of CpGs in the green module are in intergenic regions.
- C) Distribution of surrounding CpGs densities according to module (denoted by color). CpG density was (x-axis) was calculated as the number of CpGs within a 100-bp window (50 bp on either side of the CpG of interest). We observed that the blue module tended to be comprised of CpGs located in regions of higher CpG density (island), while the green module was bimodal, with half the CpGs in it being located in high density regions, and the other half in low density regions.
- D) We then plotted CpG density as a function of the CpG loading for PC1 that was used to estimate DNAmAge from overlapped CpGs. We find that for the green module, CpGs with strong negative loadings are in low density regions, while those with high loadings are in both low and high density regions.

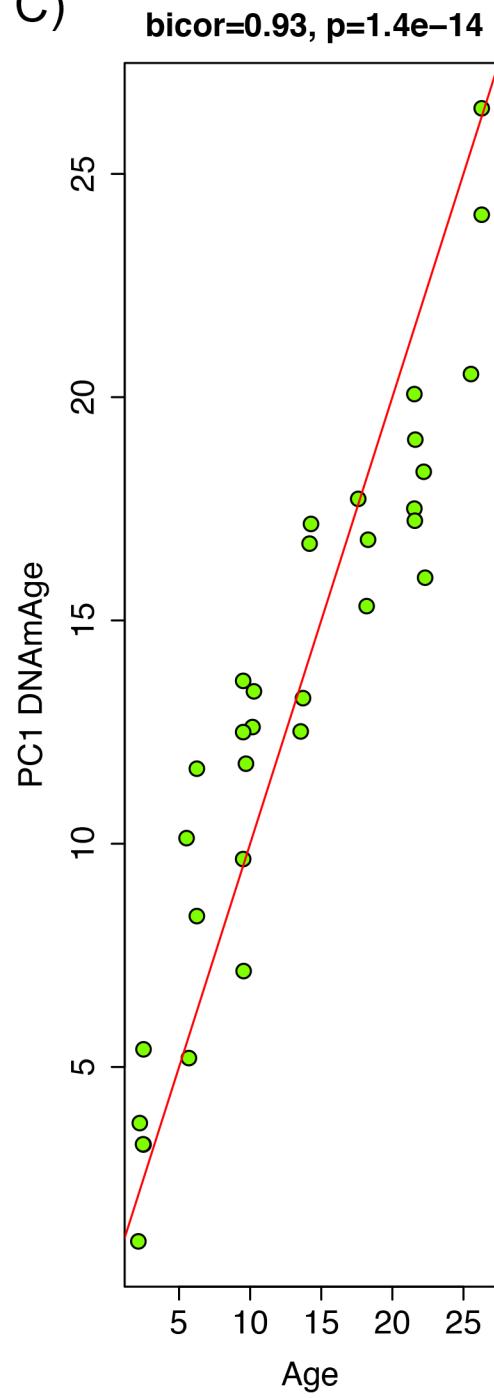
A)



B)

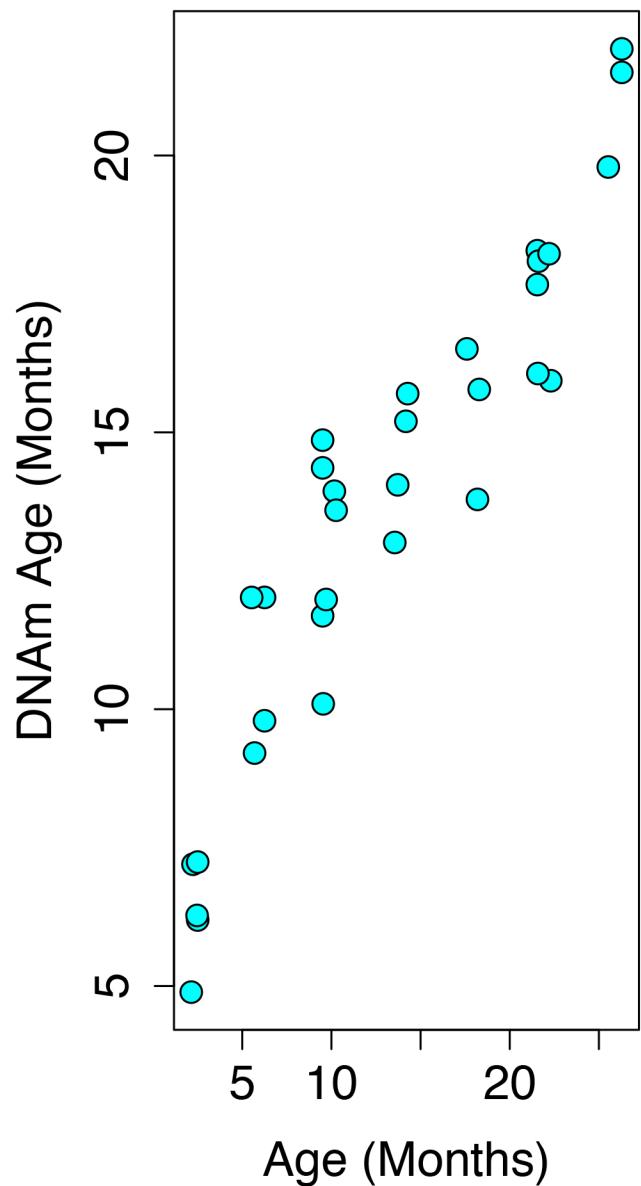


C)



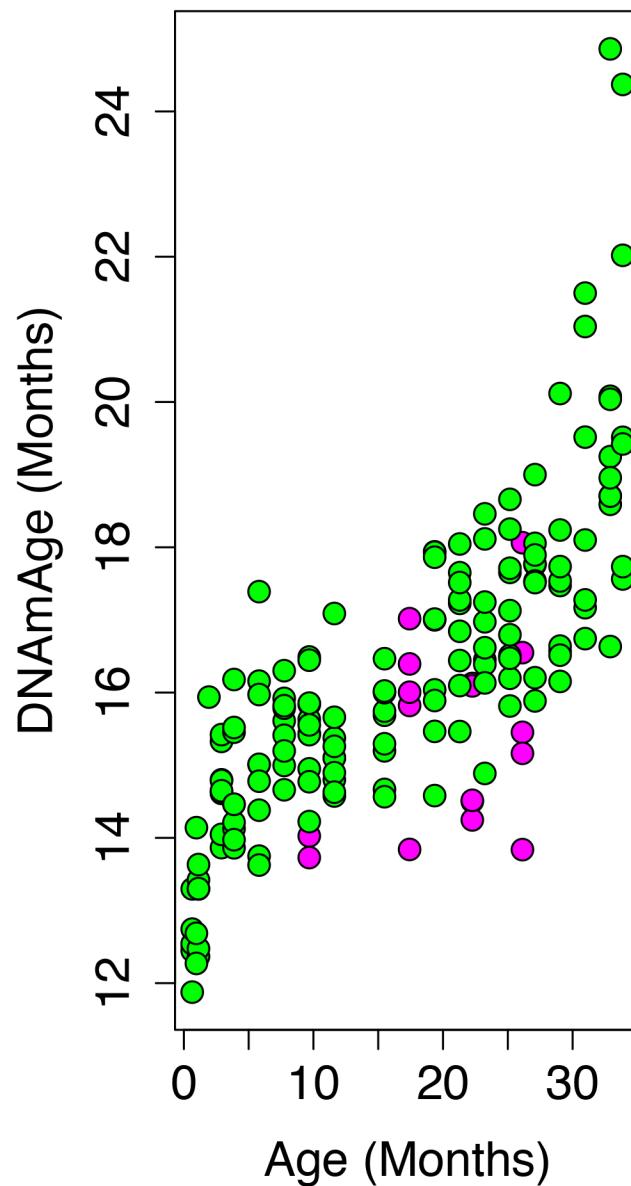
A)

Rat bicor=0.92, p=9.8e-14

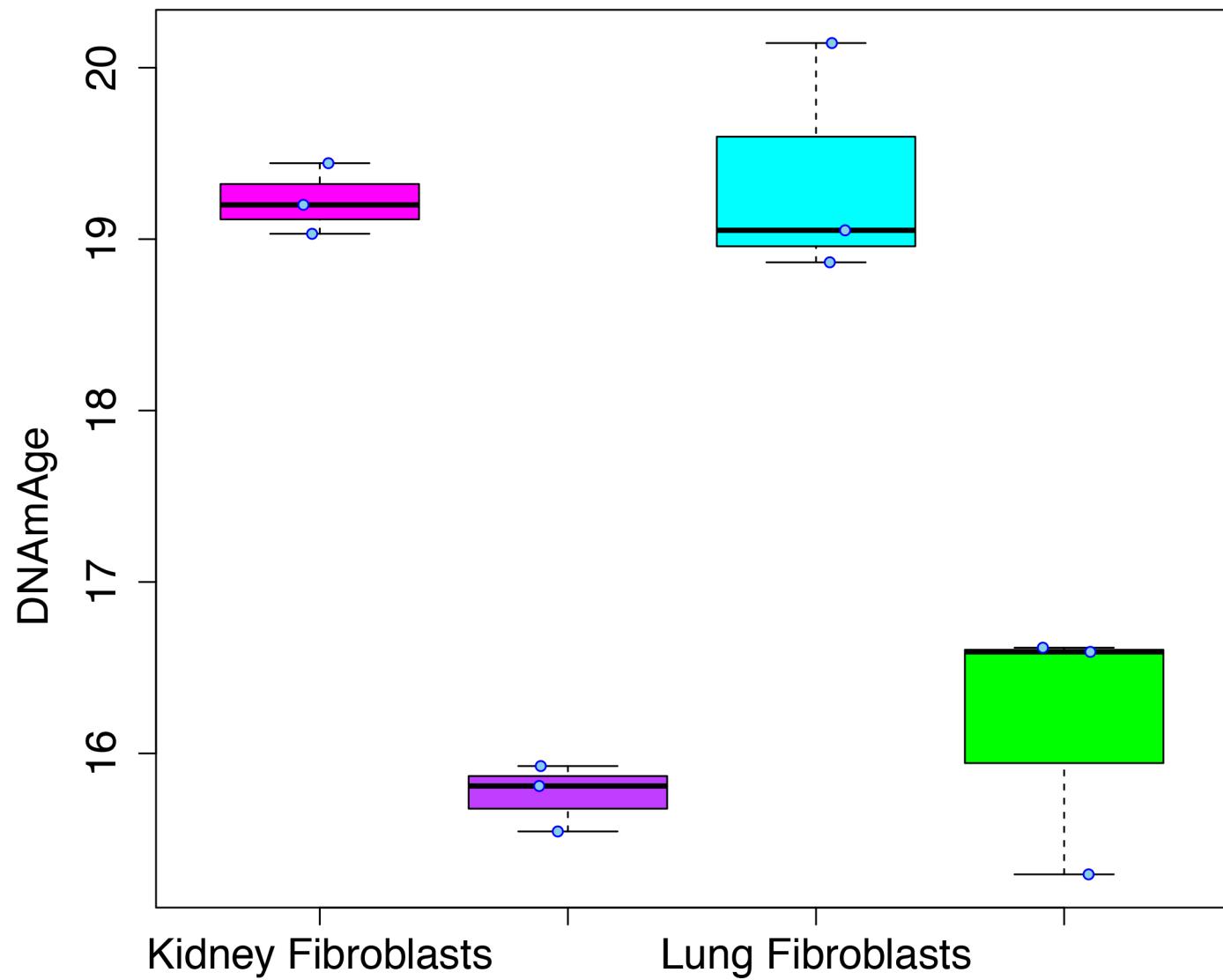


B)

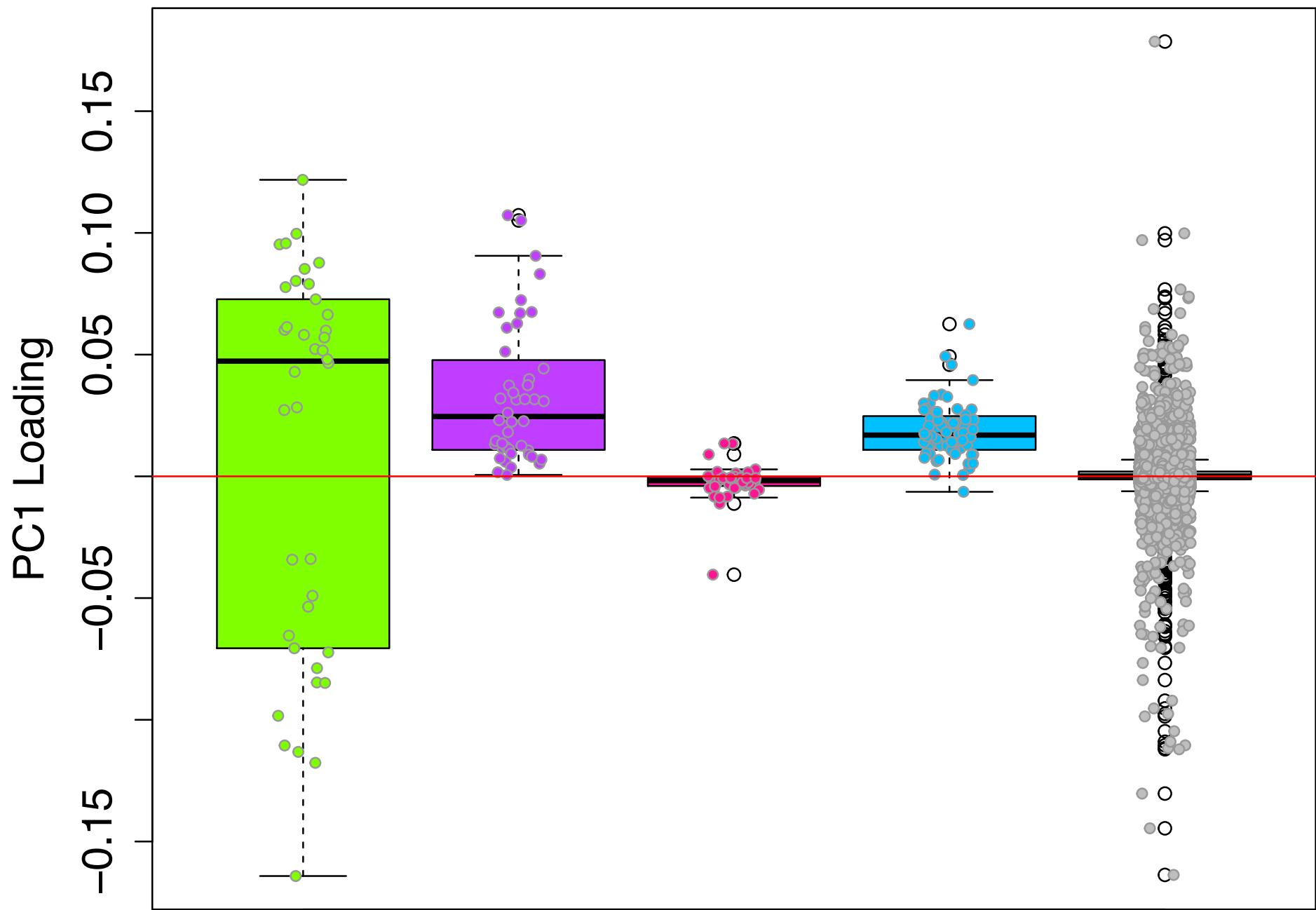
Mouse bicor=0.79, p=5e-39

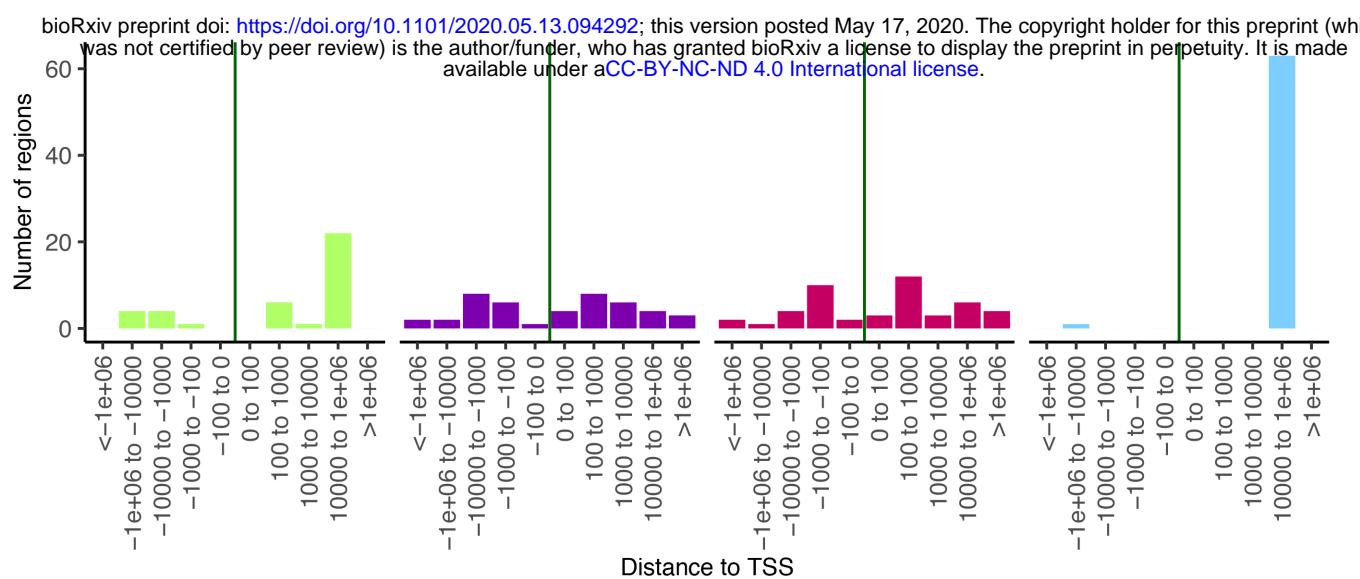
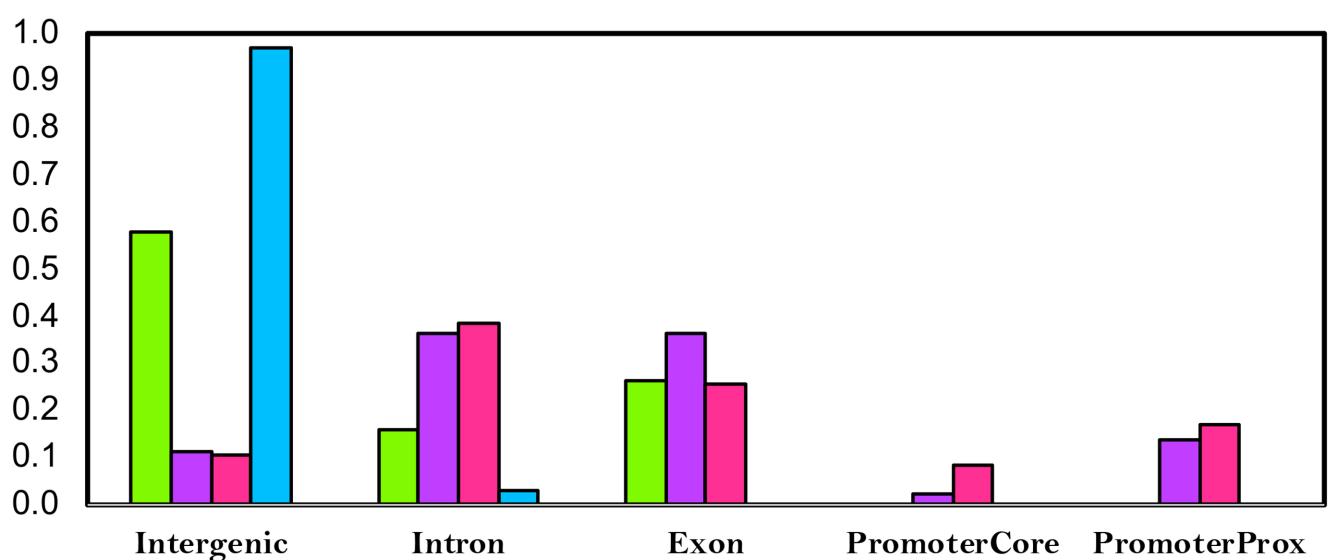
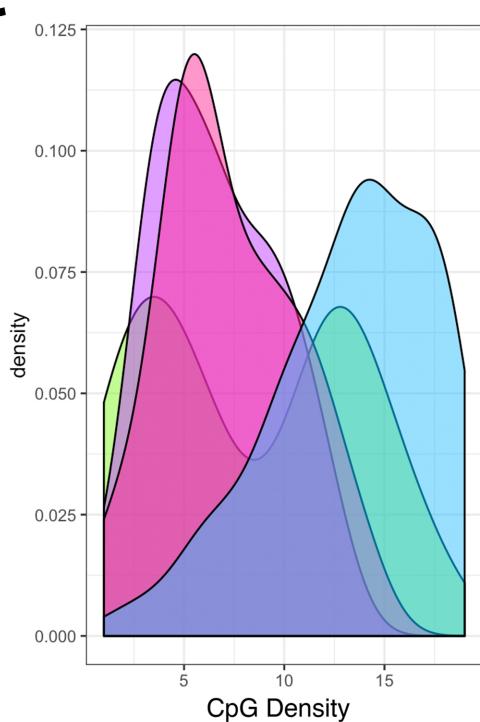


Mouse Fibroblasts $p = 0.038$



$p = 1.1e-52$



A**B****C****D**