Nature of epigenetic aging from a single-cell perspective

Andrei E. Tarkhov[1,*], Thomas Lindstrom-Vautrin[1], Sirui Zhang[1], Kejun Ying[1], Mahdi Moqri[1,2,3], Bohan Zhang[1], Vadim N. Gladyshev[1,#]

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA
[2]Department of Obstetrics & Gynecology, Stanford School of Medicine, Stanford University, Stanford, CA, USA
[3]Institute for Stem Cell Biology and Regenerative Medicine, Stanford School of Medicine, Stanford, CA, USA
Corresponding authors: *andrei_tarkhov@hms.harvard.edu, #vgladyshev@rics.bwh.harvard.edu

Abstract

Age-related changes in DNA methylation (DNAm) form the basis for the development of most robust predictors of age, epigenetic clocks, but a clear mechanistic basis for what exactly they quantify is lacking. Here, to clarify the nature of epigenetic aging, we analyzed the aging dynamics of bulk-tissue and single-cell DNAm, together with single-cell DNAm changes during early development. We show that aging DNAm changes are widespread, but are relatively slow and small in amplitude, with DNAm levels trending towards intermediate values and showing increased heterogeneity with age. By considering dominant types of DNAm changes, we find that aging manifests in the exponential decay-like loss or gain of methylation with a universal rate, independent of the initial level of DNAm. We further show that aging is dominated by the stochastic component, yet co-regulated changes are also present during both development and adulthood. We support the finding of stochastic epigenetic aging by direct single-cell DNAm analyses and modeling of aging DNAm trajectories with a stochastic process akin to radiocarbon decay. Finally, we describe a single-cell algorithm for the identification of co-regulated CpG clusters that may provide new opportunities for targeting aging and evaluating longevity interventions.

Introduction

Epigenetic clocks have been used for almost a decade to accurately predict chronological ages of cells, tissues and organisms[1–6]. More recently, new epigenetic clocks have been introduced that are trained on mortality and/or phenotypic, pathological and physiological readouts[7,8], as well as on the pace of aging[9]. In all these cases, the basic feature for age prediction or model construction is DNA methylation (DNAm) levels averaged over macroscopic tissue samples, i.e. bulk DNAm levels. The inherent challenge with such an approach is that the DNAm signal is averaged over a large number of different cells present in the tissue. Even though the role of various factors contributing to bulk DNAm changes has been extensively discussed[10–14], what biological processes drive the epigenetic aging clocks remains unknown. The observed internal age-related changes across all cells could be

confounded by multiple factors, such as changes in cell-type composition[15] and clonal expansion[16]. An important advance in this area is the development of a single-cell DNA methylation (scDNAm) clock known as scAge[17], relying on bulk DNAm data for calibration. Yet, it also does not provide a mechanistic basis for chronological age prediction.

To clarify the principles and biological mechanisms behind epigenetic aging, we turned our attention to single-cell DNAm aging and embryonic development data. Using mouse as a model, we examined inter-cell correlations as a marker of co-regulation. We were able to classify aging changes into two distinct categories: stochastic and co-regulated. Stochastic changes are not correlated across cells, organisms and CpG sites, thus representing the accumulation of molecular damage[22]. Co-regulated changes are coherent across different cells and animals, thereby changes at different CpG sites are correlated. Co-regulation of a genomic region in the current context means that there is a shared biological mechanism protecting cells in the same and other organisms of the same species from accumulating stochastic changes. However, without longitudinal scDNAm data we are unable to exclude the stochasticity of dynamics for co-regulated clusters — a cluster may be co-regulated in the genomic space, but, as a whole, change stochastically in time. Ideally, the only possibility of a CpG cluster to change its DNAm levels is to do it concordantly within the cell. The co-regulated clusters are thus good candidates for programmatic, regulatory epigenetic changes in the sense that they are irreducible to epigenetic damage accumulation. The embryonic development scDNAm data can be used as a control for co-regulated changes because of a tightly controlled genetic program that governs development.

By applying these approaches, we show that aging scDNAm changes are dominated by the stochastic component, in agreement with the concept of aging as the entropic loss of complexity[23,24], yet co-regulated changes are also present. Embryonic scDNAm data are dominated by a global wave of methylation from day E4.5 to E5.5, on top of which we managed to identify a fraction of co-regulated changes. Therefore, our analyses suggest that the nature of epigenetic aging is largely stochastic. The co-regulated CpG clusters, despite their sparsity, may show promise as better candidates for testing anti-aging interventions and aging quantification. For example, the current approaches for building epigenetic clocks are mostly based on penalized regression (Lasso, Ridge or ElasticNet), which penalize correlation across CpG sites in order to reduce the number of CpG sites used to build the clock. The predictions of age or other traits would be insensitive to the composition of the clock in the correlative analysis, thus the currently built clocks may be biased towards stochastic CpG sites. However, different parts of co-regulated clusters and stochastic sites may respond qualitatively differently in the case of interventions applied to an organism. A more mechanistic understanding of epigenetic aging based on co-regulation may improve performance of epigenetic clocks in evaluating anti-aging interventions.

Results

Bulk tissue epigenetic aging

Fig. 1.

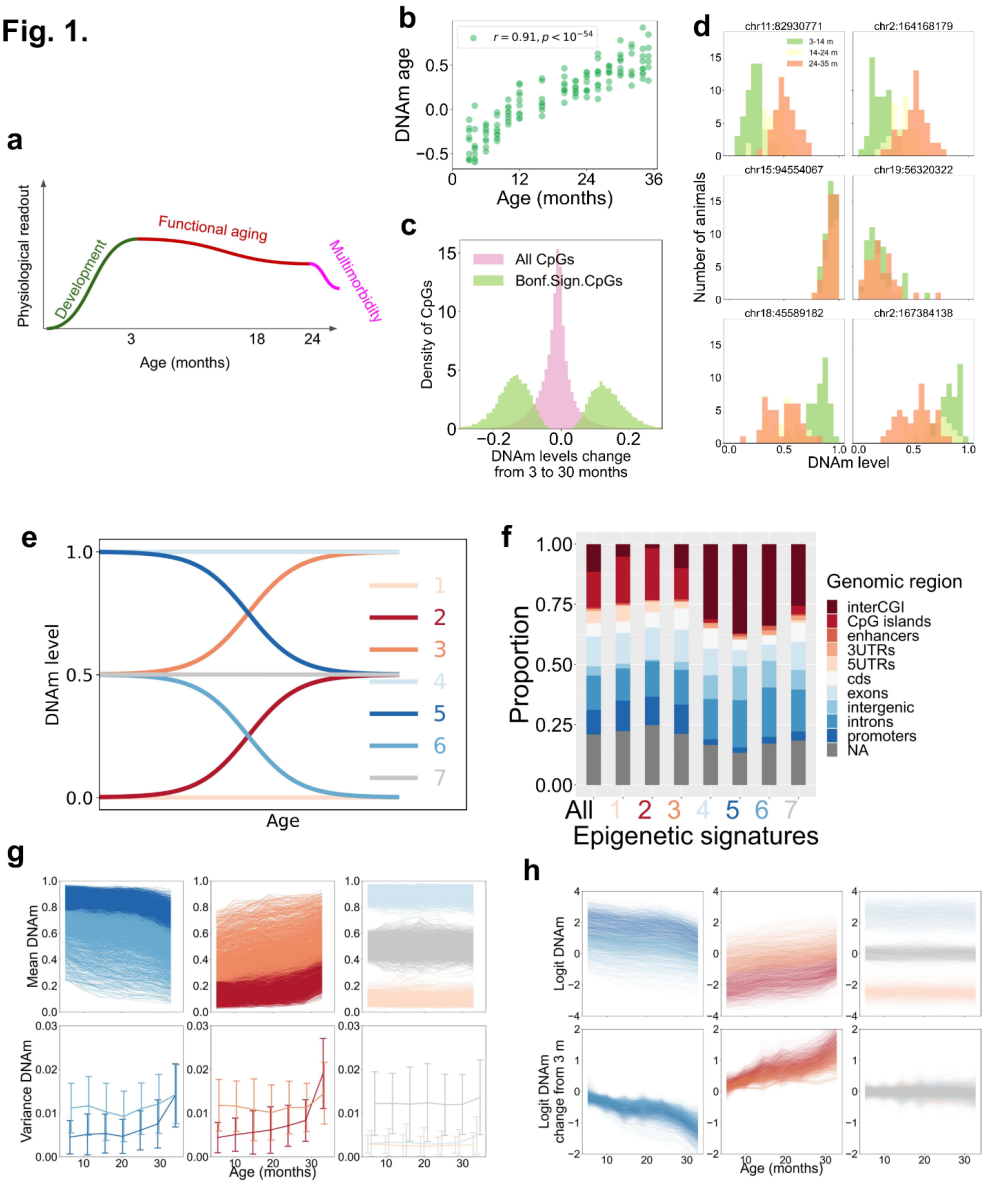


Fig. 1. Key features of bulk DNAm changes during aging. a. Three major life stages of mice represented by age-related changes in function — development, functional aging and multimorbidity[25,26]. **b.** Predicted Petkovich DNAm age of animals (16 age groups, from 3 till 35 months old) based on the Petkovich *et at.* dataset[4]. **c.** The range of DNAm changes at CpGs sites in the dataset. Histograms are shown for all CpG sites and for the CpGs significantly changing with age. **d.** Six representative CpG histograms in aging mouse cohorts of 3-14 (green), 14-24 (yellow) and 24-35 (red) months old for the following CpG sites: significantly hypermethylated with age (upper row), not changing with age (middle row), and significantly hypomethylated with age (lower row)[4]. **e.** Schematics of seven possible aging DNAm dynamics observed in experimental data. Other types of dynamics are rare but may also be present in the data. **f.** Genomic enrichment analysis for the CpG sites comprising the seven aging dynamics from **e**. Enrichment profiles are similar for the CpG sites sharing the direction of age-related changes (dynamics 2 and 3, and 5 and 6). Note that entropic changes do not share a common genomic profile (2 and 5), and the same applies to anti-entropic changes (3 and 6). **g.** Aging dynamics of mean DNAm levels (upper row) and the variance of DNAm levels (lower row) for the seven types of aging dynamics from **e**. Color scheme is the same as in **e**. **h.** Aging dynamics of logit of mean DNAm levels (upper row), and of the logit of mean DNAm levels relative to the logit of mean DNAm levels at 3 months of age. The logit was defined as $\log \frac{x}{1-x}$, where x is the DNAm level.

To assess the behavior of DNAm over the time course of mouse life, we divided it into three categories according to the typical aging dynamics of physiological readouts: development, functional aging and multimorbidity[25,26] (**Fig. 1a**). Similar life-course staging can be applied to humans[27–29]. Development, especially early development, is known to be tightly regulated, even if molecular aging already proceeds there[30]. Functional aging corresponds to the period of life between the end of development, accompanied by a gradual functional decline without major accumulation of comorbidities. Multimorbidity is the final stage of life, when the functional decline becomes incompatible with survival and normal homeostasis. We focused our analyses on the period of functional aging and made use of developmental data for uncovering dissimilarities in DNAm dynamics between these two life stages.

We first analyzed bulk blood DNA methylation (DNAm) changes during aging of male C57Bl/6 mice (16 age groups from 3 to 35 months, 8 animals per group) based on our previous reduced representation bisulfite sequencing dataset[4,31]. There were 268,044 CpG sites that significantly correlated with age (13.6% out of all 1,976,056 measured CpG sites). After the Bonferroni correction for multiple testing, the number of such CpG sites was 16,889 (0.85% of all sites). Given that a substantial fraction of CpGs change significantly with age, the prediction of chronological age based on the DNAm levels could be made (**Fig. 1b**), as was previously done in the case of the Petkovich *et al.* clock[4]. Epigenetic clocks typically comprise a few hundred of CpG sites out of hundreds of thousands that significantly change with age, hence the major challenge is how to select CpGs that contain the most biologically relevant information about aging and would produce valuable biomarkers of this process. Interestingly, the overall change of DNAm level for each of the numerous age-related CpG sites over functional aging was small. The largest change of methylation was about 30%, whereas for most of the sites it was less than 10% throughout lifespan (**Fig. 1c**). For illustration, we present six representative CpG histograms of aging mouse cohorts for pairs of CpGs significantly hypermethylated with age, not changing with age, and significantly hypomethylated with age (**Fig. 1d**).

We further classified the aging trajectories of DNAm into seven categories depending on the initial methylation value at the end of development (at 3 months), and the direction of subsequent changes (**Fig. 1e, g, h**). Dynamics 1, 4 and 7 represent CpGs, whose expression remains constant with age and differs only by the initial DNAm level. Dynamics 2 and 3 correspond to the growth of methylation with age, whereas dynamics 5 and 6 to the loss of methylation. Dynamics 2 and 5 trend towards the methylation level of 0.5 in accordance with the growth of entropy. At the same time, dynamics 3 and 6 correspond to an apparent loss of entropy, exhibiting “anti-entropic” behaviors. One may expect different biologies behind the entropic and “anti-entropic” behaviors of DNAm levels. To characterize these dynamics, we carried out genomic enrichment analyses of the epigenetic signatures corresponding to the seven types of dynamics (**Fig. 1f**). Surprisingly, dynamics 2 and 3 (increasing methylation) and dynamics 5 and 6 (losing methylation) displayed more similarity than entropic dynamics 2 and 5 or “anti-entropic” ones 3 and 6. Therefore, aging dynamics cluster according to the direction of DNAm level change over functional aging rather than according to the initial methylation level at the completion of development. Overall, the dynamics corresponding to the loss of methylation and higher initial methylation levels (4, 5, 6 and 7) compared to those gaining methylation or having low initial methylation levels (1, 2 and 3) are relatively enriched with inter-CpG island regions, or open sea, enhancers,

introns,intergenic regions and 3'-UTRs, i.e. generally non-functional regions. Likewise, they contain relatively fewer CpG islands, coding sequences, exons, 5'-UTRs and promoters.

Interestingly, the aging change was accompanied by the growth of heterogeneity of DNAm levels within cohorts of the same chronological age, which is indicated by the broadening of methylation level distribution with age (**Fig. 1d)**, and the growth of DNAm levels variance (**Fig. 1g**). Not only the mean DNAm level for CpGs corresponding to dynamics 2 and 5 showed a pronounced increase with age, but also the heterogeneity of DNAm levels in age cohorts grew with age (**Fig. 1g**). Additionally, they shared a similar rate of DNAm change notwithstanding the initial methylation level. We applied the logit transformation of DNAm levels, $\log\frac{x}{1-x}$ (x is the DNAm level), and found that after this transformation the changes were essentially linear (**Fig. 1h**), thus resembling exponential-decay-like changes in the average methylation levels within the defined dynamics of DNAm changes. This is consistent with the aging-related increase in entropy metrics based on various molecular and functional readouts[12,31–34].

The above analysis of bulk blood DNAm aging changes can be summarized by the following key points. The changes of DNAm during functional aging are omnipresent in the genome, and are relatively slow and small in the amplitude. DNAm levels, in general, trend towards the methylation level 0.5 and display increased heterogeneity with age. The aging DNAm dynamics can be clustered into seven dominant types. Aging manifests in the exponential-decay-like loss (or gain) of methylation with a universal rate, independent of the initial level of DNAm. These key points suggest that the DNAm aging changes demonstrate common features indicative of a stochastic process.

Stochastic single-cell model for simulating bulk tissue epigenetic aging

Fig. 2.

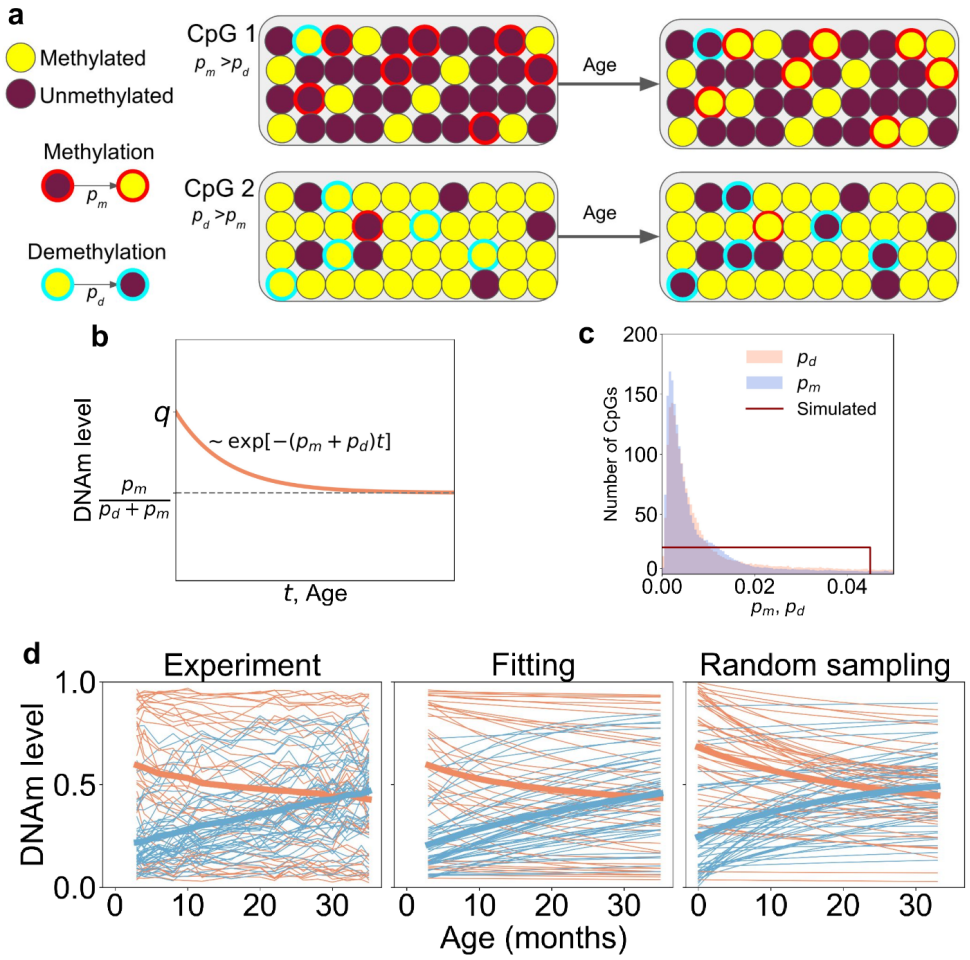


Fig. 2. Stochastic single-cell model for modeling bulk DNAm dynamics. a. Sketch for a three-parameter stochastic decay model: q — initial methylation level, p_d and p_m — rates of demethylation and methylation in a unit of time. The flips of methylation occur randomly with the corresponding rates. The behavior of two CpG sites in single cells comprising a tissue sample is illustrated. CpG 1 is initially hypomethylated and gets more methylated with age, whereas CpG 2 is initially hypermethylated and loses its methylation. **b.** The stochastic decay model produces exponential decay law for the dynamics of average methylation. The rate of decay equals the sum of p_m and p_d. The limiting value is defined by the dynamic equilibrium of methylation and demethylation. **c.** Distributions of p_m and p_d for the sites significantly changing with age[4], which were observed in the experiment. The random sampling was performed for a uniform distribution of q, p_m and p_d on the following intervals of parameters: $q \in [0, 1]$, $p_m, p_d \in [0, 0.045]$. **d.** Dynamics of CpG sites included in the Petkovich et al.[4] DNAm clock (left), results of fitting of those trajectories to the exponential decay defined in **c** (middle), and random sampling of aging trajectories **b** sampled from the distribution in **c** (right). Blue lines show trajectories losing methylation, and red lines — gaining methylation. Thick red and blue lines are the mean trajectories of red and blue lines.

In an attempt to explain a part of age-related changes of DNAm caused by stochastic damage accumulation, we tested the hypothesis that it is possible to reproduce the experimental DNAm aging trajectories with a stochastic single-cell model (see Methods). The stochastic model assumes that the DNAm dynamics for a CpG site in a cell are purely stochastic and controlled by three parameters: q — the initial methylation level, p_d and p_m — the rates of demethylation and methylation in a unit of time. We illustrate the stochastic model with an example of two CpG sites in a tissue sample of 40 cells — for CpG 1 the rate of methylation is higher than the rate of demethylation, whereas the initial level of methylation is low, and for CpG 2 the parameters are the opposite (**Fig. 2a**). As the time passes, some of the cells would flip their methylation state according to the predefined probabilities. However, the bulk tissue DNAm would average single-cell levels of methylation across all cells and produce a single number for the mean DNAm level, which is predicted by the model. For a sample of n cells, the CpG site is methylated in n_m cells. During functional aging, the CpG can randomly acquire or lose methylation according to the probability rates p_d and p_m defined above. We derived the aging trajectory equation for the average level of methylation across the whole sample $x(t) = n_m(t)/n$ to be an exponential decay curve (**Fig. 2b**):

$$x(t) = \frac{p_m}{p_m+p_d} + \left(q - \frac{p_m}{p_m+p_d}\right) \exp(-(p_m + p_d)t). \qquad (1)$$

To compare real aging trajectories of CpG sites with the model predictions, we analyzed the DNAm aging dynamics[4] and calculated the rates p_m, p_d for all sites changing significantly with age (**Fig. 2c**), and plotted the aging dynamics for 90 CpGs comprising the Petkovich *et al.* epigenetic clock (**Fig. 2d, left**). Next, we fitted the experimental DNAm levels dynamics to $x(t)$ for each clock CpG site (**Fig. 2d, middle**). We also plotted a subset of trajectories corresponding to the parameters q, p_d, p_m randomly drawn from the uniform distributions on the following intervals: $q \in [0, 1]$, $p_m, p_d \in [0, 0.045]$ (**Fig.2c**) for each CpG site (**Fig. 2d, right**).

The stochastic model predictions can fit the observed behavior of clock CpG aging trajectories (**Fig. 2d left** and **middle**). At the same time, similarity of experimental and randomly sampled trajectories (**Fig. 2d left** and **right)** suggests that the stochastic model is able to qualitatively reproduce the experimental behavior within a stochastic framework with the use of a single parameter for all CpG sites — the maximal rate of exponential decay among all sites $(p_m + p_d)_{max}$ defined explicitly via the upper limit of sampling intervals (**Fig.2c**). This parameter may be used to characterize instability of the aging epigenome and to predict a potential limit of lifespan due to epigenetic instability.

Bulk epigenetic clocks are agnostic of the single cell patterns of aging

Fig. 3.

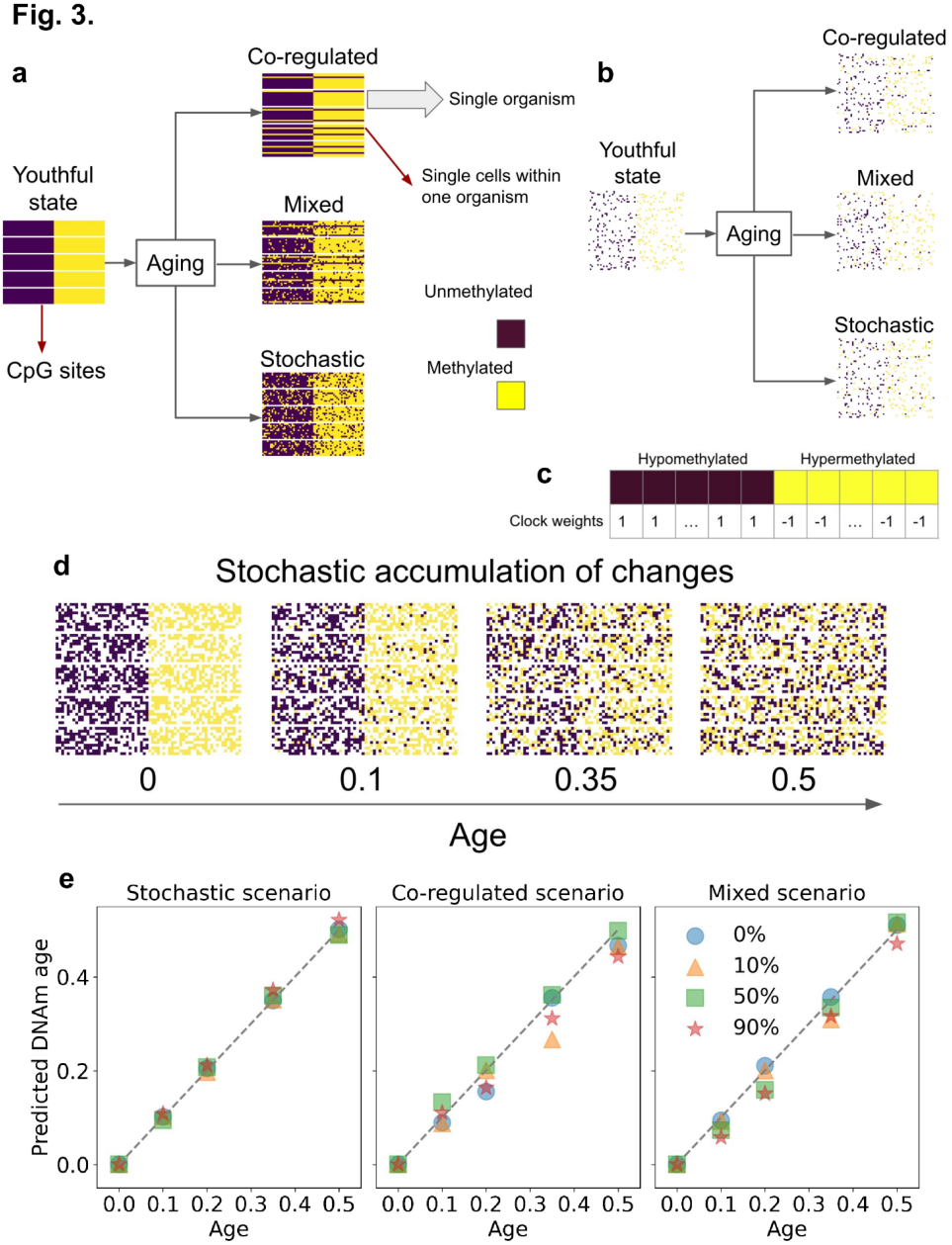


Fig. 3. Three single-cell scenarios of DNAm aging changes. a. An example of three single-cell scenarios used to produce the same 20% change of tissue-bulk DNAm levels with age: co-regulated, stochastic and mixed. The example is shown for 5 aging animals (horizontal rectangles) with DNAm levels measured for a set of cells (represented by rows in the rectangles). Each CpG is represented by a column, the left half of the CpGs are assumed to be hypomethylated in the youthful state, whereas the right half is initially hypermethylated. Co-regulated aging changes occur in a correlated manner within one cell (changes of different CpG sites are correlated), and in a coherent way in different cells and animals. Stochastic changes are uncorrelated among different cells and among different CpG sites. Mixed changes are a combination of the two. Bulk-tissue DNAm clocks are unable to distinguish these three scenarios since they all correspond to the same change in DNAm levels. **b.** Realistic coverage of NGS in single cells would produce a sparse subset of CpGs measured in only a handful of the cells. The sparsity in the plot is 90%. **c.** A hypothetical clock with the left half of all CpG sites hypomethylated in the youthful state, having weight +1, whereas the right half of CpG sites hypermethylated in the youthful state, having weight -1. For age prediction, the clock weights are normalized by the number of CpGs in the clock. **d.** Illustration of the dynamics of stochastic accumulation of epigenetic changes with age. Age corresponds to the fraction of average DNAm

levels (from 0 to 0.5). The sparsity level is 50%. **e.** Prediction of the hypothetical clock from **c** (normalized to have a unit length) on the simulated data for the three scenarios from **a**. The clock from **c** is able to accurately predict chronological age for any used single-cell scenario, and for any NGS coverage (see legend for the level of sparsity).

Different single-cell DNAm distributions can produce the same bulk tissue DNAm levels. Specifically, single-cell aging changes can be caused by drastically different biological processes without being detected by bulk tissue DNAm signals. To illustrate this point, we sketch three possible scenarios of single-cell DNAm changes that correspond to the same bulk DNAm pattern (**Fig. 3a**). The left half of CpGs were initially unmethylated in all cells and organisms, whereas the right half — methylated. Then, over the course of aging, each unmethylated CpG site acquires 20% of methylation, and each methylated CpG site loses 20% of methylation. Three single-cell scenarios are shown: stochastic, co-regulated and mixed. The stochastic one assumes that all changes are scattered across cells in an uncorrelated manner — the DNAm changes occur independently at CpGs in cells and organisms. This scenario corresponds to damage accumulation. The co-regulated scenario represents a model wherein there are two states for a cluster of CpGs, young and old, and during aging 20% of cells switch from the young state to the old state. This scenario corresponds, for instance, to the case of senescent cells, whose population grows with age. The third scenario is a mixture of the two described above — all cells accumulate stochastic changes of methylation, whereas some genomic regions are co-regulated. Bulk DNAm data are unable to distinguish these scenarios. In real scDNAm data, low sequence coverage produces a severely sparse signal (**Fig. 3b**), making the analysis of scDNAm challenging but not impossible in mitotic or clonally expanding cells. Below, for our analysis, we use single-cell muscle stem cells and embryonic cells to partially overcome this challenge.

We use a hypothetical example of a clock built for the simulated single-cell scenarios (**Fig. 3c**). Since bulk changes of DNAm are the same 20% for all CpGs, the clock has two constant parts: it has weights +1 for the right half of CpGs, and -1 for the left half (normalized to the total number of CpGs). We illustrate the dynamics of accumulation of aging changes in the stochastic scenario example (**Fig. 3d**). We show that the prediction of chronological age is possible no matter what single-cell scenario or the level of sparsity of the scDNAm signal were used to simulate the aging dynamics (**Fig. 3e**). At the same time, the clock is unable to distinguish the single-cell scenarios used to generate scDNAm levels. The mechanism behind such a clock is a gradual accumulation of stochastic damage, or the growth of the population of “old” cells. In both cases the predictive power of the clock is based on the inevitable change of average methylation levels with age. Further, we turn to real scDNAm data to clarify what scenario is realized in an experiment.

Single-cell epigenetic dynamics during functional aging and embryonic development

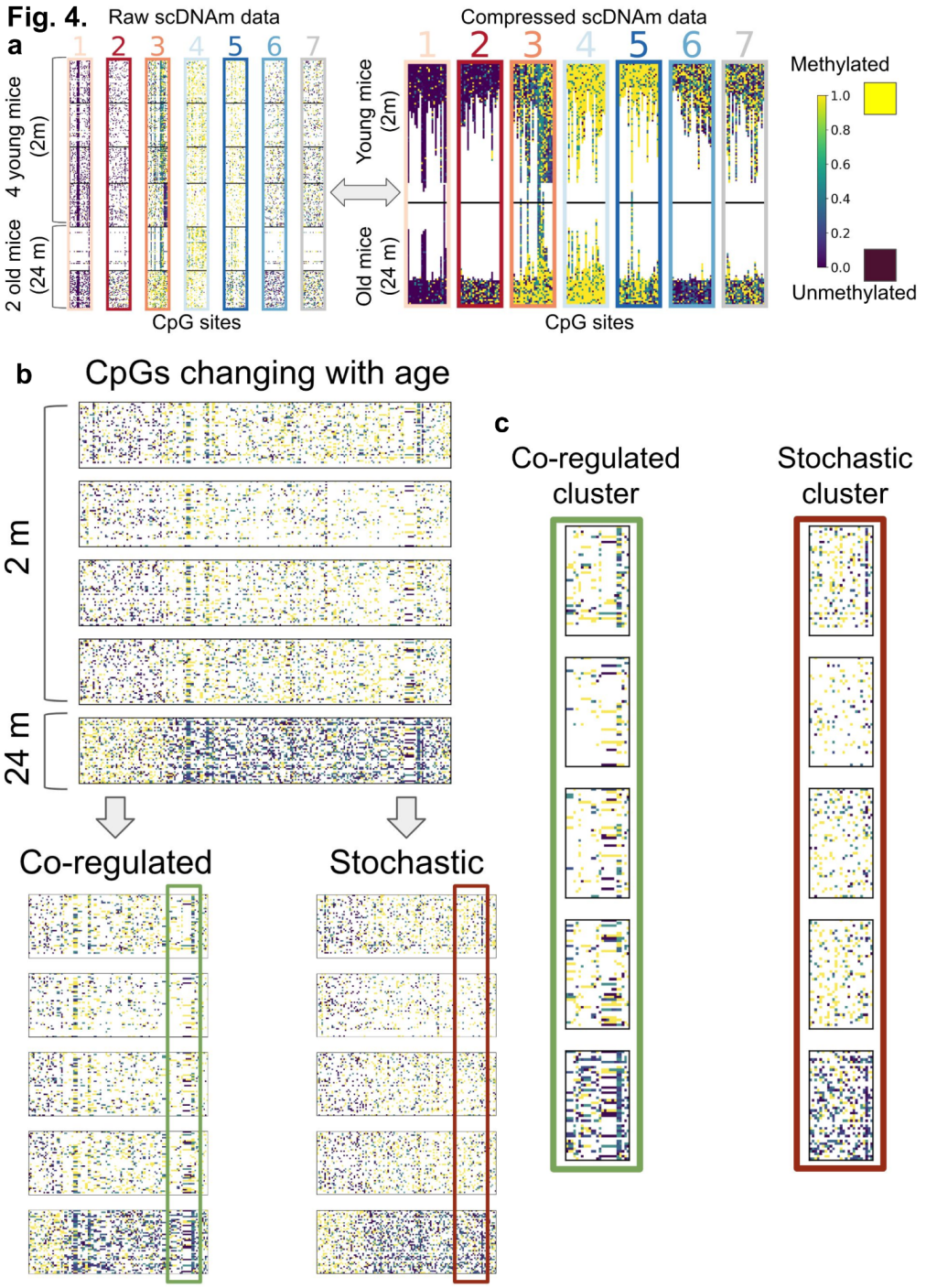


Fig. 4. Single-cell DNAm dynamics during functional aging. a. All seven types of aging trajectories are present in the muscle stem cells (muSC) scDNAm data[35] (**Fig. 1e**). Raw sparse scDNAm data are shown for a subset of 275 cells from 4 young mice (2 months old) and 2 old mice (24 months old), single cells from which represent horizontal lines (**left panel**). Black lines separate cells corresponding to different mice. For each CpG (vertical column), compression omits non-measured cells, and shows only the measured ones (**right panel**). All measured cells for the young mice are collapsed to the top of the figure, and for the old mice to the bottom. One old mouse showed a significantly lower coverage than other animals, and was censored from the following analyses. **b.** Dynamics of a subset of CpG sites that change significantly with age in scDNAm data (from 2 month-old to 24 month-old mice) (**upper panel**). All age-related CpGs were classified into co-regulated and stochastic clusters by their inter-cell correlation (**lower panel**). The co-regulated

cluster has characteristic stripes of CpGs whose DNAm levels change concordantly (**lower left panel**). The stochastic cluster does not show strong correlation among different cells (**lower right panel**). The largest co-regulated cluster (**green box**), and a stochastic cluster of the same size (**brown box**) were chosen for an additional close-up view. **c.** Close-up view of the co-regulated cluster (**left**) and a stochastic cluster (**right**) from **b**. The co-regulated cluster indicates the stretches of CpGs that show high correlation among themselves, and change their DNAm levels as a single unit. The stochastic cluster does not have similar clustering of CpG sites. Each CpG site in the stochastic cluster changes independently from others.

We analyzed two scDNAm datasets that provide high NGS coverage for each cell: mouse embryos prior and during gastrulation[36] and aging muscle stem cells (muSC)[37]. In the case of aging muSCs, we had 275 single cells derived from four 2-month-old mice and two 24-month-old mice (one old mouse was censored because of the low coverage of NGS in its cells). First, we filtered CpGs by coverage: each CpG must be measured in at least 15 cells of young mice, and 15 cells of old mice. Out of 35,584,147 CpGs measured in at least one cell, only 155,359 had sufficient coverage and passed the filter. Second, we identified CpGs changing significantly with age, thus leaving 502 CpG sites. Within those CpGs, we managed to observe all seven types of bulk DNAm aging dynamics present in **Fig. 1d** (**Fig. 4a**). For convenient comparison with the bulk DNAm dynamics, we present both raw and compressed scDNAm data, where for each CpG we omit non-measured cells and collapse all the cells for the young mice to the top, and for the old mice to the bottom of the figure (**Fig. 4a left** and **right**).

Out of 502 CpGs changing methylation with age, we identified 121 CpG sites increasing methylation and 381 CpG sites losing methylation with age (**Fig. 4b**). To identify the co-regulated clusters of CpGs, for each pair of CpGs, we calculated the inter-cell correlation coefficient separately for the young mice, for the old mice, and for all mice together. The values of correlation had to be higher than 0.4 in all three cases to include the pair of CpGs into the co-regulated cluster. Non-measured values of DNAm were imputed by zeros. The method of identifying co-regulated clusters is conceptually similar to the global coordination level (GCL) metric developed independently for scRNAseq data analysis[38]. The majority of age-related CpG sites, 76%, changed according to the stochastic scenario, whereas 24% of the CpG sites changed in a co-regulated manner. Such a definition of co-regulation is relatively strict, and would rather tend to mark a CpG site stochastic if it has low coverage. However, for a larger number of sequenced cells, the method would be able to identify more co-regulated clusters. Therefore, the ability to identify real co-regulated clusters may be strongly influenced by experimental limitations. With the advancements in sequencing technologies, it may be possible to improve the quality of prediction of co-regulation.

Commonly used for building epigenetic clocks, penalized regression methods have the advantage of selecting CpG sites that add the most new information to the regression model at the cost of reducing the number of collinear CpG sites. However, their major drawback in the context of scDNAm data is that they may be biased towards stochastically changing CpG sites (see **Extended Fig. 1, Lasso** and **ElasticNet clocks**). The reason for the bias is that the CpG sites comprising co-regulated clusters are strongly collinear to each other. For the purposes of mathematical regression, they do not add any additional information to the model, whereas the stochastic sites are poorly correlated with each other, and the addition of each new stochastic CpG site is beneficial for the mathematical algorithm. The biological

meaning of co-regulation implies that there is a well-defined biological process that controls each co-regulated cluster and does not allow the components of this cluster to accumulate epigenetic damage. Therefore, the mathematical collinearity is coherent with the biological meaning behind the cluster, which is ignored by penalized regression models.

Fig. 5.

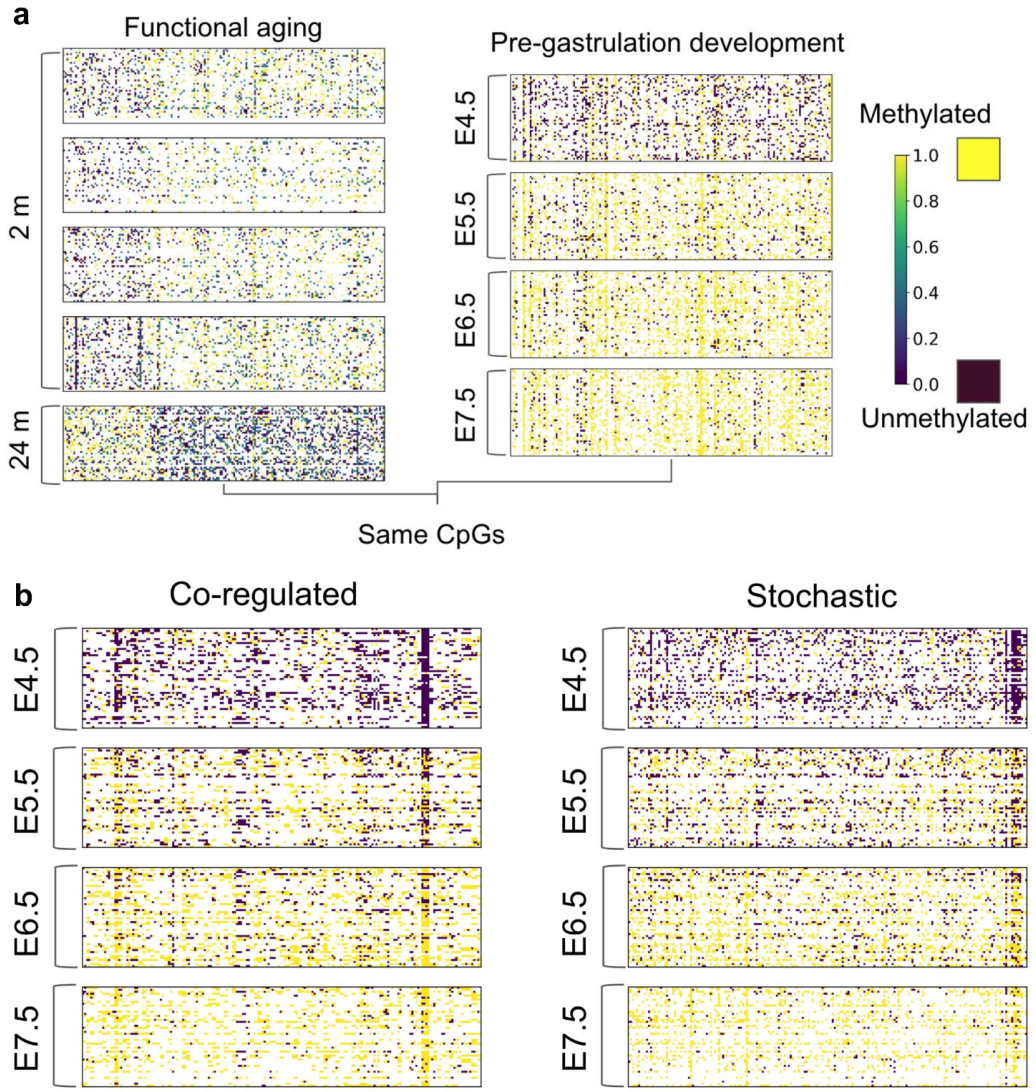


Fig. 5. Comparison of single-cell DNAm dynamics during functional aging and embryonic development. a. Comparison of aging dynamics of CpGs (**left**) to their dynamics in mouse embryos before and during gastrulation (**right**). Common CpG sites measured in both experiments were used for plotting. Both hyper- and hypo-methylated with age CpGs (**left**) were initially hypomethylated at embryonic day E4.5, and largely acquired global methylation by day E5.5 (**right**). Afterwards little qualitative change of methylation was detected up to day E7.5. The qualitative character of changes during functional aging comprises both co-regulated and stochastic clusters, whereas the pre-gastrulation development is characterized by the program-like global hypermethylation. **b**. Co-regulated and stochastic clusters identified for changes for days E4.5, E5.5, E6.5 and E7.5. The dominant process is gaining methylation at all co-regulated CpG sites.

To test the co-regulation scenario in the context of developmental genetic programs, we further analyzed embryonic scDNAm data[36]. There, we had 758 single-cell samples for

embryonic days E4.5, E5.5, E6.5 and E7.5. First, we chose the CpGs changing during functional aging (**Fig. 5a left**) that were also measured in the embryos (370 out of 502 CpGs). These CpGs changed in a coherent co-regulated and program-like manner: at E4.5 the global methylation was low, whereas by E5.5 it was largely set to a methylated state, and only slightly further increased during E6.5 and E7.5 (**Fig. 5a right**). Overall, the dynamics of scDNAm changes during functional aging was different from that during gastrulation. Some of CpGs losing methylation in old animals gained it during gastrulation. The global change of DNAm during gastrulation follows a trend opposite to the aging changes of DNAm age. At the same time, aging co-regulated clusters also changed during embryonic development, which may signify the developmental activation of the same biological mechanisms becoming prominent later in aging.

The global change of methylation during gastrulation represents a complication for our algorithm — in the case of global changes of methylation, the inter-cell correlation turns out to be less meaningful because of minor variation of methylation levels among cells and organisms at each developmental stage. In a hypothetical example of the methylation switching from 0 to 1 in all cells of all organisms, it is unclear how to identify individual clusters responsible for the global change of methylation. Therefore, we look for co-regulation in those regions that do not exactly follow the global wave of methylation. To identify such co-regulated clusters in the DNAm changes pre- and during gastrulation, we filtered CpGs by coverage: each CpG must be measured in at least 25 cells for samples corresponding to each embryonic day. Out of 20,073,742 CpGs measured in at least one cell only 44,711 had sufficient coverage and passed the filtering, whereas 6,000 CpGs changed significantly with age. An application of our algorithm produced a co-regulated cluster of 304 CpGs, and 5,796 stochastic CpGs (**Fig. 5b**). For the embryonic dataset, it turned out that there were only 5% co-regulated CpGs that we could identify in the background of the global methylation event.

Overall, our analysis showed that the major component in the aging scDNAm signal is stochastic. However, a quarter of aging CpG sites were identified as co-regulated with some of the clusters spanning extensive genomic regions. In the following section, we present a biological annotation of the identified clusters.

Biological annotation of co-regulated and stochastic clusters

Fig. 6.

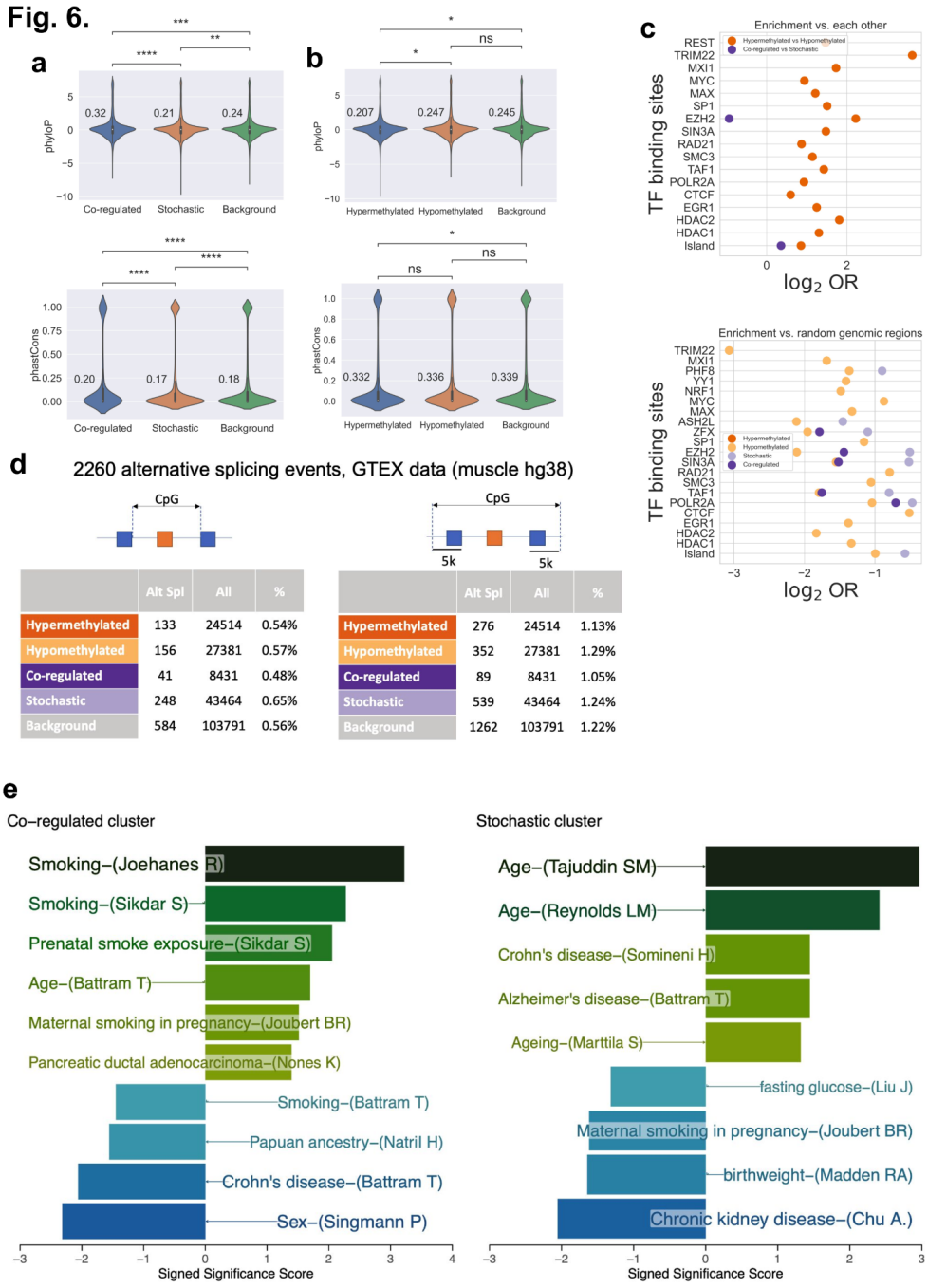


Fig. 6. Biological annotation of co-regulated and stochastic clusters. a. PhyloP (**upper panel**) and phastCons (**lower panel**) evolutionary conservation score distributions for CpGs comprising stochastic and co-regulated clusters compared to random regions of the genome. Co-regulated clusters show significantly higher evolutionary conservation than the random regions, whereas stochastic clusters are significantly less conserved than both the random regions and co-regulated clusters. **b**. PhyloP (**upper panel**) and phastCons (**lower panel**) evolutionary conservation score distributions for CpGs comprising hypermethylated and hypomethylated clusters compared to random regions of the genome. Hypermethylated clusters show significantly lower evolutionary conservation scores than both the random regions and hypomethylated regions. **c.** Enrichment with transcription-factor (TF) binding sites and CpG islands for co-regulated clusters vs. stochastic clusters and hypermethylated vs. hypomethylated regions (**upper panel**), and co-regulated and stochastic clusters, hypermethylated and hypomethylated regions vs. random regions of the genome (**lower panel**). **d.** Enrichment analysis of age-associated alternative splicing events (Alt Spl) for the CpGs

comprising co-regulated, stochastic, hypermethylated and hypomethylated clusters (**left**), and for the alternative splicing events within a 5 kb distance of the CpGs clusters in the genome (**right**). Co-regulated clusters showed a lower percentage of alternative splicing events than stochastic clusters, though the difference was not statistically significant (Fisher's exact test). **e**. Enrichment of co-regulated and stochastic clusters against EWAS hits. Each horizontal bar represents an enriched term. The X-axis shows the -log10(P-value), signed by log2 (Odds ratio). Only the EWAS trait with significant enrichment (P < 0.05) are included and annotated. *Independent t-test p-values for legend in* **a**: *ns:* $0.05 < p \leq 1$, **:* $0.01 < p \leq 0.05$, ***:* $10^{-3} < p \leq 0.01$, ****:* $10^{-4} < p \leq 10^{-3}$, *****:* $0.05 < p \leq 10^{-4}$.

To obtain more CpGs belonging to co-regulated and stochastic clusters for subsequent biological annotation, we modified our filters. First, we lowered the filter for CpGs based on coverage: instead of requiring 15 young and 15 old cells, we required 5 cells in each case. We identified 5,999,943 CpGs out of 35,584,147 CpGs measured in at least one cell, including 51,895 CpGs that changed significantly with age. In this case, the co-regulated and stochastic clusters comprised 8,431 and 43,464 CpGs, respectively. As expected from the algorithm's properties being biased towards a stricter criterion for co-regulated sites, the fraction of stochastic sites increased from 76% to 84%. In addition to 51,895 CpGs changing methylation with age, we selected 51,895 random CpGs from the genome for subsequent enrichment analyses.

We calculated evolutionary conservation scores phyloP[39,40] and phastCons[41–45] for co-regulated and stochastic clusters (**Fig. 6a**), as well as for genomic regions corresponding to dynamics 2 and 3, hypermethylated with age, and dynamics 5 and 6, hypomethylated with age (**Fig. 6b**). Co-regulated clusters showed a significantly higher evolutionary conservation than stochastic clusters and random regions, in agreement with the hypothesis of a tighter regulatory control and a higher biological importance of those regions (**Fig. 6a**). At the same time, stochastic regions showed a significantly lower evolutionary conservation than random genomic regions (**Fig. 6a**). The hypermethylated clusters also showed a significantly lower evolutionary conservation than hypomethylated clusters and random genomic regions (**Fig. 6b**).

We also examined the enrichment of co-regulated clusters vs stochastic clusters and hypermethylated vs hypomethylated clusters with transcription factor (TF) binding sites (**Fig. 6c upper panel**), and compared the trends with random genomic regions (**Fig. 6c lower panel**). The co-regulated regions contained significantly fewer EZH2 binding sites than the stochastic clusters, and fewer ZFX, EZH2, SIN3A, TAF1, POLR2A binding sites than the random regions. The stochastic sites contained fewer PHF8, ASH2L, ZFX, EZH2, SIN3A, TAF1, POLR2A binding sites than random regions. Overall, both co-regulated and stochastic clusters showed fewer binding sites than random genomic regions.

We further analyzed age-associated splicing events in the muscle tissue from the Genotype-Tissue Expression (GTEx) database[46]. For each CpG site from the co-regulated, stochastic clusters and random regions, we checked if it was located in the region from the beginning of the first exon to the end site of the last exon in an alternative splicing event (**Fig. 6d left**), or surrounding the splicing event within 5 kb (**Fig. 6d right)**. The stochastic

clusters were enriched with alternative splicing events in comparison with co-regulated clusters, however, the difference was not statistically significant.

Finally, we analyzed the enrichment of CpG sites from the co-regulated and stochastic clusters with hits from 6,993 epigenome-wide association studies (EWAS)[47]. The co-regulated clusters were enriched with the phenotypes related to smoking and age (**Fig. 6e**). The stochastic clusters were enriched with the phenotypes related to aging, Alzheimer's and Crohn's diseases.

Discussion

We analyzed the aging dynamics of bulk-tissue DNAm and single-cell DNAm, together with scDNAm changes during gastrulation, with the primary goal of separating stochastic accumulation of DNAm changes from co-regulated DNAm changes driven by a common biological mechanism.

By examining bulk DNAm changes with age and bulk DNAm aging clocks, we observed that epigenetic aging is omnipresent in the genome (for example, 13.6% of the measured CpGs changed significantly with age), and is a relatively slow process. It is also very small in amplitude (most sites change only by 10%), shares common temporal dynamics and shows increased heterogeneity with age. By simulating a null-hypothesis that DNAm aging changes occur randomly in single cells, we managed to reproduce the experimental aging clocks dynamics with the derived stochastic decay model. All these observations are consistent with the concept of epigenetic aging being a stochastic process characterized by increasing entropy.

In order to test whether other aging dynamics are present in DNAm data, we turned to aging scDNAm data and found that 76% of measured CpGs behaved in a stochastic manner during aging, whereas 24% changed in a co-regulated way. In contrast, during gastrulation scDNAm changes were dominated by a global methylation event, in the background of which we managed to identify only 5% of co-regulated CpGs. Even though the available scDNAm data are currently scant to provide more detailed insights into epigenetic aging, the methods developed in the present paper would be applicable to future single-cell data generated with advanced sequencing techniques. In particular, the algorithm we developed for the identification of co-regulated CpG clusters may allow improving the accuracy and interpretability of DNAm aging clocks.

We applied typical epigenetic clock-building routines (Lasso and ElasticNet penalized regressions) in order to build epigenetic aging clocks, and showed that they may be biased towards the stochastic CpG clusters. Without lowering accuracy of chronological age predictions, they may ignore most of the co-regulated CpG sites due to their high collinearity. In response to an intervention, the co-regulation of a cluster may be disrupted, which may be missed by such clocks. It is likely that the stochastic CpG sites bear less information regarding the biology involved because of their high tolerance to stochastic epigenetic changes. In other words, the clocks built to measure stochastic accumulation of epigenetic changes, might not perform well where one expects reversal of biological pathways and

processes, for example, in the case of rejuvenation therapies. On the other hand, clocks built on the co-regulated cluster of CpG sites may perform worse in the context of chronological age prediction but may be able to better capture the effects of longevity interventions.

Both bulk and single-cell DNAm analyses suggest that the high accuracy of epigenetic aging clocks may be predetermined by the stochastic decay of the epigenetic state set during early development. We describe a mechanism for epigenetic clocks which is strikingly similar to radiocarbon decay often used for dating in archeology: there is no need to have a biologically relevant mechanism as soon as the mean concentration of radioactive carbon or of the fraction of methylated DNA change monotonically with age. Radiocarbon dating works surprisingly well even though it is based on a purely stochastic process of radioactive decay according to the exponential decay law. In contrast to the radioactive decay of carbon-14, in the case of epigenetic clocks, we deal with two separate stochastic processes of gaining methylation in some genomic regions, and losing methylation in others, which are obviously actively driven by metabolism. Thus, the analogy with radiocarbon decay is rather mathematical and conceptual rather than biological. The two processes of loss and gain of methylation are controlled by two different kinds of biological machinery, but the measured mean methylation level changes are affected by both processes. Stochasticity implies that over time those machineries unavoidably make mistakes, which accumulate gradually with age and can be used as robust predictors of age.

It is important to note that our analyses are limited to the process of functional aging, and do not consider the effects of rejuvenation therapies on the epigenome[48–50]. Stochasticity of age-related epigenetic changes does not imply the impossibility of reversal, as is the case for epigenetic reprogramming protocols resetting the DNAm patterns. At the same time, stochasticity behind the process of accumulation of epigenetic changes with age does not preclude programmatic behavior, a quasi-program of aging, defined by the developmental biology predisposing species to follow a particular aging trajectory. Moreover, components of the stochastic part of epigenetic clocks would be predetermined by development and biological organization of the organism. The sites that were initialized in the hypo- or hyper-methylated states by the end of early development would tend to stochastically gain or lose methylation with age, hence they would make good candidates for epigenetic clock CpG sites. Thus, the developmental program initializes the epigenome into a state that later stochastically decays during aging. Therefore, the multispecies epigenetic clocks[51] may work well because closely related species, such as mammals, share the associated developmental biology setting them into similar initial states of the epigenome. Overall, the effects of early embryonic development on aging need to be further investigated[30,52].

The question of the biological meaning of existing epigenetic clocks deserves separate discussion. The causal relationship between the molecular changes during aging and functional decline resulting in mortality is also the subject of an ongoing debate. A highly desirable feature of aging clocks is the ability to predict mortality events and lifespan; however, the state-of-the-art epigenetic clocks continue ticking in immortalized cell cultures[18,19] and in naked-mole rats[20,21], where mortality exhibits minimal changes with age. These observations raise questions about the use of epigenetic clocks for the prediction of mortality. At the same time, the absence of a mechanistic explanation behind epigenetic clocks impedes their clinical use as aging biomarkers.

Stochasticity behind the clock does not imply that there is no biological value in the clock. The stochastic accumulation of damage may be influenced by lifestyle, diet, interventions and other factors, hence it is possible both to reset the stochastic changes to some other state (younger or older), and to change the rate of damage accumulation (in both directions, up and down). Therefore, stochasticity of epigenetic clocks might be a good indicator of cumulative deleteriousness of the environment in which an organism lives, or of cumulative non-specific damage. However, it is less clear how stochastic epigenetic clocks would capture the effects of target-specific therapies or some non-promiscuous aging changes. We anticipate that there might be two different kinds of clocks necessary for the quantification of aging: stochastic for estimation of cumulative damage, and co-regulated for estimation of programmatic effects of longevity interventions. The current approaches for building epigenetic clocks mix up these two qualitatively different components, and may have a limited predictive power for testing interventions.

Acknowledgements

The authors thank Didac Santesmasses, Alexander Tyshkovskiy, Jeyoung Bang, Wayne Mitchell, Anastasia Shindyapina for discussion. The work was supported by NIA grants to VNG.

Methods

Single-cell stochastic model of DNAm changes

For each CpG site, we assume the aging trajectory of methylation level is controlled by three parameters: q — the initial methylation level, p_d — the rate of demethylation in a unit of time, p_m — the rate of methylation in a unit of time. The model is assumed to be purely stochastic, which means that the state of a CpG site in a short time interval Δt has probability to get methylated $p_m \Delta t$ and probability to get demethylated $p_d \Delta t$.

Assuming that there are n cells in a sample, we can denote by n_d and n_m the numbers of cells that are demethylated or methylated for a given CpG site. The aging dynamics for n_m and n_d would be describable by the following rate equations:

$$n_m(t + \Delta t) = n_m(t)\,(1 - p_d \Delta t) + n_d(t) p_m \Delta t,$$

$$n_d(t + \Delta t) = n_d(t)\,(1 - p_m \Delta t) + n_m(t) p_d \Delta t.$$

The conservation of the total number of sites is satisfied, which is shown by the summation of the two equations above: $n_m(t + \Delta t) + n_d(t + \Delta t) = n_m(t) + n_d(t) = n$. To derive the differential equation for the average methylation level $x(t) = \frac{n_m(t)}{n}$, we use the fact that $n_d(t) = n - n_m(t)$,

$$\frac{n_m(t+\Delta t) - n_m(t)}{\Delta t} = - n_m(t) p_d + (n - n_m(t)) p_m,$$

and by dividing both sides of the equation by n, we obtain:

$\frac{dx(t)}{dt} \approx \frac{x(t+\Delta t)-x(t)}{\Delta t} = p_m - x(t)(p_d + p_m)$.

The exact solution of the above equation reads

$$x(t) = \frac{p_m}{p_m+p_d} + (q - \frac{p_m}{p_m+p_d}) \exp(-(p_m + p_d)t). \qquad \text{(S1)}$$

To understand the meaning of the three introduced parameters q, p_m, p_d, let us further analyze the above equation. The initial value of the average methylation level is defined by $x(0) = q$, whereas with time the methylation level tends to the asymptotic value $\lim_{t\to\infty} x(t) = \frac{p_m}{p_m+p_d}$. The rate of exponential decay is equal to the sum of the rates of methylation and demethylation $p_m + p_d$.

It is worth noting that even though the process generating the dynamics is purely stochastic, it doesn't necessarily lead to the saturation of the methylation level at the level 0.5. To the contrary, by varying the three parameters we may obtain an aging trajectory of a CpG site starting at any point from 0 to 1, and tending to any other methylation level from 0 to 1, whereas the rate of change would be controlled by the absolute values of the methylation and demethylation rates for each particular site. The above analysis considers a single CpG site and its aging trajectories. The values of the model parameters q, p_m, p_d may be characteristic of a genomic position, and may bear some biological meaning.

Fitting experimental bulk DNAm aging trajectories to the stochastic model prediction

In order to fit the experimental aging trajectories for CpG sites comprising the Petkovich *et al.* clock, we use the three-parametric stochastic aging trajectory derived above in Eq. (S1), and apply Python's fitting tool $\mathrm{scipy.optimize.curve_fit}$[53].

Simulating random subset of aging trajectories predicted by the stochastic model

To show how a subset of aging trajectories corresponding to randomly sampled parameters q, p_m, p_d, we use Eq. (S1) and randomly draw the parameters from the uniform distributions defined on the following intervals: $q \in [0, 1]$, $p_m, p_d \in [0, 0.0015]$. The number of sampled sets is equal to the number of CpG sites in the Petkovich *et al.* clock.

The bundle of random trajectories is thus fully defined by a single parameter — the upper bound for the rates p_m, p_d, which is set here to 0.0015. The maximal rate of exponential decay among all CpG sites $(p_m + p_d)_{max}$ hence represents the critical parameter of the stochastic model. It might be related to the typical level of deleteriousness of the environment for the organism. Therefore, $(p_m + p_d)_{max}$ might be a proxy to the identification of the maximal lifespan for a species.

Genomic enrichment analysis for epigenetic profiles

For genomic annotation of epigenetic profiles we used R package annotatr[54].

Single-cell DNAm data analysis

Due to high sparsity of single-cell DNAm data, we had to extensively use filtering by coverage. For illustrative purposes of identifying visually recognizable co-regulated clusters, we set a threshold of coverage for each CpG for it to be measured in 15 young cells and 15 old cells. In each of the cells, we had to keep all CpGs covered by at least 1 NGS read due to low coverage. For enrichment analysis, we lowered the filter to 5 young and 5 old cells. For muSCs, one old mouse had to be censored because of an extremely low coverage. For embryonic scDNAm, we set the threshold at 25 covered cells for each embryonic stage for each CpG.

For correlation analysis we developed custom correlative algorithms that were able to ignore omitted CpGs in some cells, and use only those that were measured. To avoid the confounding factor of different coverage in young and old cells for identifying CpGs whose methylation significantly correlated with age, we also produced a pseudo-bulk sample for each of the mice by calculating the average methylation level for each of the CpGs. A CpG was considered significantly correlated with age if it was associated with age in both single cell data, and in the pseudo-bulked data.

For inter-cell correlative analysis, we imputed absent methylation values with zeros. The threshold correlation value for a pair of CpGs to be considered co-regulated was chosen to be 0.4, which must be satisfied in three groups of mice: only young mice, only old mice and all mice. For embryonic data, there were four groups corresponding to each of the embryonic days E4.5-7.5. That further allowed lowering the limit for false positive identification of co-regulation in the case of low coverage.

Enrichment analysis in phenome-wide EWAS signals

We analyzed the enrichment of CpG sites from co-regulated and stochastic clusters in hits from 6,993 epigenome-wide association studies in humans, obtained from the EWAS catalog[47]. Mouse CpG sites from the mm10 reference mouse genome were mapped to the conserved human CpG sites in the hg19 human reference genome with the help of the UCSC Genome Browser liftover tool[55]. Then, the nearest human CpG sites in the Illumina EPIC array within 100 base-pairs were identified and used for the enrichment. For each EWAS hit, Fisher's exact test was performed to determine the enrichment of either co-regulated or stochastic clusters for a given trait.

Enrichment analysis in aging-associated splicing events

To detect age-associated splicing events, we downloaded the RNA-sequencing files of muscle tissue from Genotype-Tissue Expression (GTEx) database[46], quantified all alternative splicing events and modeled the association with age using linear regression. In total, 2260 aging-associated alternative splicing events were identified. For each CpG site, we checked if the CpG site fell in at least one aging-associated splicing event (from the start site of the first exon to the end site of the last exon in an alternative splicing event), or surrounding the splicing event within 5kb. We counted this type of CpG site in each CpG cluster and compared it with the random regions of the background genome.

Transcription factor binding regions

We used the ENCODE transcription factor binding database[56] for identification of genomic regions corresponding to TF bindings sites. The database is based on ChIP-seq combining chromatin immunoprecipitation with DNA sequencing to infer the possible binding sites of DNA-associated proteins. Prior to use mouse CpG sites from the mm10 reference mouse genome were mapped to the conserved human CpG sites in the hg38 human reference genome with the help of the UCSC Genome Browser liftover tool[55].

References

1. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
2. Hannum, G. *et al.* Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol. Cell* **49**, 359–367 (2013).
3. Johansson, Å., Enroth, S. & Gyllensten, U. Continuous Aging of the Human DNA Methylome Throughout the Human Lifespan. *PLoS ONE* **8**, e67378 (2013).
4. Petkovich, D. A. *et al.* Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell Metab.* **25**, 954-960.e6 (2017).
5. Thompson, M. J. *et al.* A multi-tissue full lifespan epigenetic clock for mice. *Aging* **10**, 2832–2854 (2018).
6. Meer, M. V., Podolskiy, D. I., Tyshkovskiy, A. & Gladyshev, V. N. A whole lifespan mouse multi-tissue DNA methylation clock. *eLife* **7**, e40675 (2018).
7. Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573–591 (2018).
8. Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303–327 (2019).
9. Belsky, D. W. *et al.* DunedinPACE, a DNA methylation biomarker of the pace of aging. *eLife* **11**, e73420 (2022).
10. Vershinina, O., Bacalini, M. G., Zaikin, A., Franceschi, C. & Ivanchenko, M. Disentangling age-dependent DNA methylation: deterministic, stochastic, and nonlinear. *Sci. Rep.* **11**, 9201 (2021).

11. Seale, K., Horvath, S., Teschendorff, A., Eynon, N. & Voisin, S. Making sense of the ageing methylome. *Nat. Rev. Genet.* (2022) doi:10.1038/s41576-022-00477-6.

12. Tarkhov, A. E., Denisov, K. A. & Fedichev, P. O. Aging clocks, entropy, and the limits of age-reversal. *bioRxiv* (2022) doi:10.1101/2022.02.06.479300.

13. Levine, M. E., Higgins-Chen, A., Thrush, K., Minteer, C. & Niimi, P. *Clock Work: Deconstructing the Epigenetic Clock Signals in Aging, Disease, and Reprogramming*. http://biorxiv.org/lookup/doi/10.1101/2022.02.13.480245 (2022) doi:10.1101/2022.02.13.480245.

14. Bell, C. G. *et al.* DNA methylation aging clocks: challenges and recommendations. *Genome Biol.* **20**, 249 (2019).

15. Marioni, R. E. *et al.* DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* **16**, 25 (2015).

16. Minteer, C. J. *et al. Revisiting the bad luck hypothesis: Cancer risk and aging are linked to replication-driven changes to the epigenome*. http://biorxiv.org/lookup/doi/10.1101/2022.09.14.507975 (2022) doi:10.1101/2022.09.14.507975.

17. Trapp, A., Kerepesi, C. & Gladyshev, V. N. Profiling epigenetic age in single cells. *Nat. Aging* **1**, 1189–1201 (2021).

18. Minteer, C. *et al.* Tick tock, tick tock: Mouse culture and tissue aging captured by an epigenetic clock. *Aging Cell* **21**, (2022).

19. Kabacik, S. *et al.* The relationship between epigenetic age and the hallmarks of ageing in human cells. *Nat. Aging* (2022) doi:10.1038/s43587-022-00220-0.

20. Kerepesi, C. *et al.* Epigenetic aging of the demographically non-aging naked mole-rat. *Nat. Commun.* **13**, 355 (2022).

21. Horvath, S. *et al.* DNA methylation clocks tick in naked mole rats but queens age more slowly than nonbreeders. *Nat. Aging* **2**, 46–59 (2022).

22. Gladyshev, V. N. *et al.* Molecular damage in aging. *Nat. Aging* **1**, 1096–1106 (2021).

23. Hayflick, L. Entropy Explains Aging, Genetic Determinism Explains Longevity, and

Undefined Terminology Explains Misunderstanding Both. *PLoS Genet.* **3**, e220 (2007).

24. Lipsitz, L. A. Loss of 'Complexity' and Aging: Potential Applications of Fractals and Chaos Theory to Senescence. *JAMA* **267**, 1806 (1992).

25. Yanai, S. & Endo, S. Functional Aging in Male C57BL/6J Mice Across the Life-Span: A Systematic Behavioral Analysis of Motor, Emotional, and Memory Function to Define an Aging Phenotype. *Front. Aging Neurosci.* **13**, 697621 (2021).

26. Petr, M. A. *et al.* A cross-sectional study of functional and metabolic changes during aging through the lifespan in male mice. *eLife* **10**, e62952 (2021).

27. Wang, S., Lai, X., Deng, Y. & Song, Y. Correlation between mouse age and human age in anti-tumor research: Significance and method establishment. *Life Sci.* **242**, 117242 (2020).

28. Thomas, J. *et al.* Running the full human developmental clock in interspecies chimeras using alternative human stem cells with expanded embryonic potential. *Npj Regen. Med.* **6**, 25 (2021).

29. Pyrkov, T. V. *et al.* Longitudinal analysis of blood markers reveals progressive loss of resilience and predicts human lifespan limit. *Nat. Commun.* **12**, 2765 (2021).

30. 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**, eabg6082 (2021).

31. Sziráki, 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 (2018).

32. Clemens, Z. *et al.* The biphasic and age-dependent impact of klotho on hallmarks of aging and skeletal muscle function. *eLife* **10**, e61138 (2021).

33. Menichetti, G., Bianconi, G., Castellani, G., Giampieri, E. & Remondini, D. Multiscale characterization of ageing and cancer progression by a novel network entropy measure. *Mol. Biosyst.* **11**, 1824–1831 (2015).

34. Sivakumar, S., LeFebre, R. W., Menichetti, G., Mugler, A. & Ambrosio, F. *The fidelity*

of genetic information transfer with aging segregates according to biological processes. http://biorxiv.org/lookup/doi/10.1101/2022.07.18.500243 (2022) doi:10.1101/2022.07.18.500243.

35. Angermueller, C. *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* **13**, 229–232 (2016).

36. Argelaguet, R. *et al.* Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487–491 (2019).

37. Hernando-Herraez, I. *et al.* Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. *Nat. Commun.* **10**, 4361 (2019).

38. Amit, G., Vaknin Ben Porath, D., Levy, O., Hamdi, O. & Bashan, A. Global coordination level in single-cell transcriptomic data. *Sci. Rep.* **12**, 7547 (2022).

39. Cooper, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901–913 (2005).

40. Siepel, A., Pollard, K. S. & Haussler, D. New Methods for Detecting Lineage-Specific Selection. in *Research in Computational Molecular Biology* (eds. Apostolico, A., Guerra, C., Istrail, S., Pevzner, P. A. & Waterman, M.) vol. 3909 190–205 (Springer Berlin Heidelberg, 2006).

41. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).

42. Thomas, J. W. *et al.* Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788–793 (2003).

43. Siepel, A. & Haussler, D. Computational identification of evolutionarily conserved exons. in *Proceedings of the eighth annual international conference on Computational molecular biology - RECOMB '04* 177–186 (ACM Press, 2004). doi:10.1145/974614.974638.

44. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).

45. Felsenstein, J. & Churchill, G. A. A Hidden Markov Model approach to variation

among sites in rate of evolution. *Mol. Biol. Evol.* **13**, 93–104 (1996).

46. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

47. Battram, T. *et al.* The EWAS Catalog: a database of epigenome-wide association studies. *Wellcome Open Res.* **7**, 41 (2022).

48. Zhang, B. *et al. Multi-omic rejuvenation and lifespan extension upon exposure to youthful circulation*. http://biorxiv.org/lookup/doi/10.1101/2021.11.11.468258 (2021) doi:10.1101/2021.11.11.468258.

49. Olova, N., Simpson, D. J., Marioni, R. E. & Chandra, T. Partial reprogramming induces a steady decline in epigenetic age before loss of somatic identity. *Aging Cell* **18**, e12877 (2019).

50. Gill, D. *et al.* Multi-omic rejuvenation of human cells by maturation phase transient reprogramming. *eLife* **11**, e71624 (2022).

51. MAMMALIAN METHYLATION CONSORTIUM *et al. Universal DNA methylation age across mammalian tissues*. http://biorxiv.org/lookup/doi/10.1101/2021.01.18.426733 (2021) doi:10.1101/2021.01.18.426733.

52. Zhang, B. *et al. Epigenetic profiling and incidence of disrupted development point to gastrulation as aging ground zero in* Xenopus laevis. http://biorxiv.org/lookup/doi/10.1101/2022.08.02.502559 (2022) doi:10.1101/2022.08.02.502559.

53. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

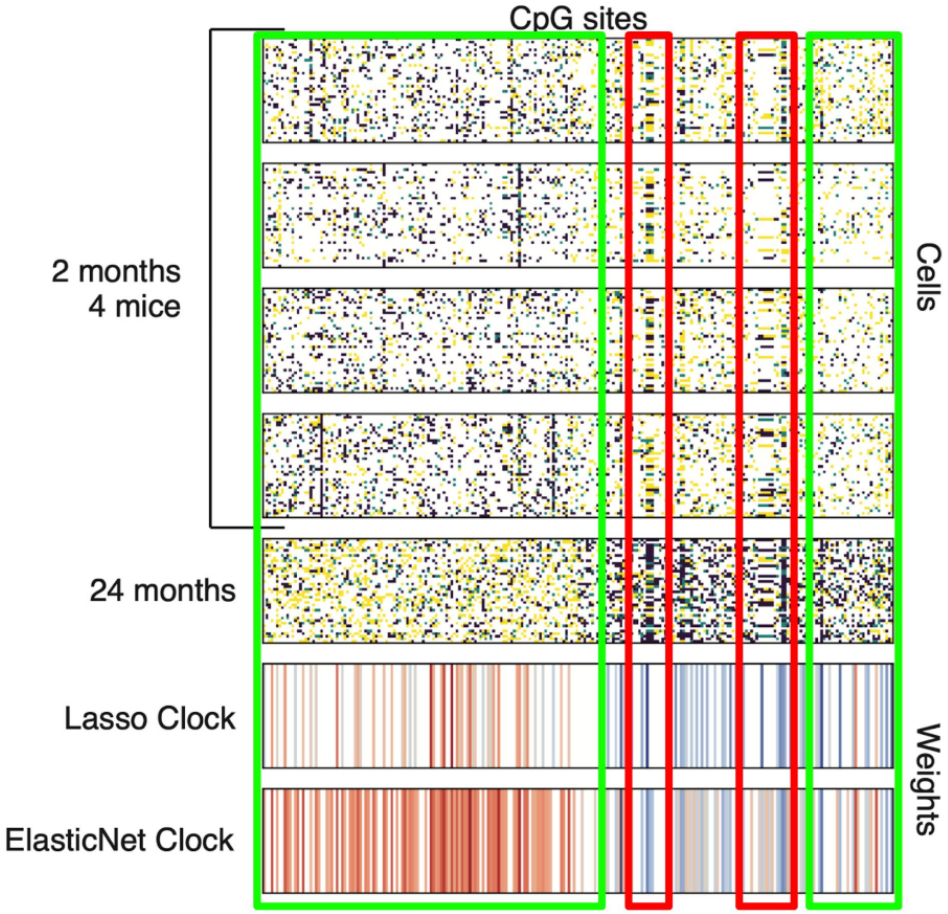
54. Cavalcante, R. G. & Sartor, M. A. annotatr: genomic regions in context. *Bioinformatics* **33**, 2381–2383 (2017).

55. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

56. Sloan, C. A. *et al.* ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44**, D726–D732 (2016).

Extended Data

Extended Data Fig. 1.



Extended Data Fig. 1. Dynamics of CpG sites significantly associated with age in scDNAm data[35]. The dominant cluster of CpG sites changes stochastically (green boxes), whereas the co-regulated clusters represent 24% of all CpG sites changing with age (red boxes). The penalized regression models are biased towards stochastic CpG sites, since they penalize any kind of correlation across the CpG sites used in the model (**Lasso** and **ElasticNet clocks**). The color represents the value of the regression weights. The co-regulated clusters contain fewer non-zero weights in the regression models. Therefore, the penalized regression models further lower the fraction of co-regulated sites used for building DNAm aging clocks.