









## Epigenetic models developed for plains zebras predict age in domestic horses and endangered equids

Brenda Larison <sup>1,2,13</sup>✉, Gabriela M. Pinho <sup>1,13</sup>, Amin Haghani<sup>3</sup>, Joseph A. Zoller <sup>3</sup>, Caesar Z. Li <sup>3</sup>, Carrie J. Finno<sup>4</sup>, Colin Farrell<sup>5</sup>, Christopher B. Kaelin <sup>6,7</sup>, Gregory S. Barsh <sup>6,7</sup>, Bernard Wooding<sup>8</sup>, Todd R. Robeck <sup>9</sup>, Dewey Maddox<sup>10</sup>, Matteo Pellegrini<sup>5</sup> & Steve Horvath <sup>3,11,12</sup>✉

Effective conservation and management of threatened wildlife populations require an accurate assessment of age structure to estimate demographic trends and population viability. Epigenetic aging models are promising developments because they estimate individual age with high accuracy, accurately predict age in related species, and do not require invasive sampling or intensive long-term studies. Using blood and biopsy samples from known age plains zebras (*Equus quagga*), we model epigenetic aging using two approaches: the epigenetic clock (EC) and the epigenetic pacemaker (EPM). The plains zebra EC has the potential for broad application within the genus *Equus* given that five of the seven extant wild species of the genus are threatened. We test the EC's ability to predict age in sister taxa, including two endangered species and the more distantly related domestic horse, demonstrating high accuracy in all cases. By comparing chronological and estimated age in plains zebras, we investigate age acceleration as a proxy of health status. An interaction between chronological age and inbreeding is associated with age acceleration estimated by the EPM, suggesting a cumulative effect of inbreeding on biological aging throughout life.

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA. <sup>2</sup>Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095, USA. <sup>3</sup>Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA. <sup>4</sup>Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, CA 95616, USA. <sup>5</sup>Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA, USA. <sup>6</sup>HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA. <sup>7</sup>Department of Genetics, Stanford University, Stanford, CA 94305, USA. <sup>8</sup>Quagga Project, Elandsberg Farms, Hermon 7308, South Africa. <sup>9</sup>Zoological Operations, SeaWorld Parks and Entertainment, 7007 SeaWorld Drive, Orlando, FL, USA. <sup>10</sup>White Oak Conservation, 581705 White Oak Road, Yulee, FL 32097, USA. <sup>11</sup>Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, USA. <sup>12</sup>Altos Labs, San Diego, CA, USA. <sup>13</sup>These authors contributed equally: Brenda Larison and Gabriela M. Pinho. ✉email: [blarison@ucla.edu](mailto:blarison@ucla.edu); [shorvath@mednet.ucla.edu](mailto:shorvath@mednet.ucla.edu)

Effective management of threatened species relies on the ability to estimate demographic trends, which depend, in turn, on accurate information about age distributions within populations<sup>1</sup>. Growth rates shape age distributions, can reflect past and current environmental and anthropogenic perturbances<sup>2,3</sup>, and can also be used to predict future population growth<sup>4</sup>. However, age is challenging to quantify in wild animals. Age estimation typically requires either investment in long-term field studies or invasive approaches that may not be feasible in live animals<sup>4,5</sup>. Another problem is the limited accuracy of some methods, which may negatively impact conservation efforts<sup>2</sup>. The challenges and importance of obtaining accurate age information have motivated efforts to develop an accurate and non-invasive approach to aging wild animals<sup>4,6</sup>.

Epigenetic aging models, particularly epigenetic clocks (ECs), promise to improve the aging of wild animals and thereby make valuable contributions to wildlife conservation and population biology<sup>4,6</sup>. These highly accurate clocks use information from genomic methylation patterns and have been studied extensively in humans<sup>7–9</sup> and mice<sup>10–12</sup>. The development and use of epigenetic models in other species are still limited but are becoming increasingly common<sup>6</sup>. A critical limitation to developing epigenetic aging models for wildlife is that these models need to be trained on samples from individuals of known age. Therefore populations of non-model organisms with known-age individuals are of extreme importance<sup>13</sup>. Here, we develop epigenetic models for a wild equid (plains zebras, *Equus quagga*) using both blood and remote biopsy samples collected from known-age individuals in a captive-bred population.

Besides their high accuracy, four other features of epigenetic aging models should make them attractive for wildlife managers. First, they can be developed from many different tissue types<sup>6,14</sup>. Second, they can be created based on very few genomic sites. The most accurate clocks for humans involve only a few hundred CpG sites<sup>14,15</sup>, and far fewer CpGs have been used to build ECs in some wild vertebrates<sup>3,16,17</sup>. Third, an accurate clock can be developed using relatively few individuals of known age<sup>3,16,17</sup>. Finally, ECs developed for one species have been shown to predict age accurately in closely related taxa (e.g., humans and chimps<sup>14</sup>), and therefore can be developed for a less threatened species with the intent of using them in threatened sister species. Using this rationale, we aim to test the performance of the EC developed for plains zebras to predict age in domestic horses (*E. caballus*) and two threatened species: Grevy's zebras (*E. grevyi*) and Somali asses (*E. africanus somaliensis*).

Since individual chronological ages are known for our plains zebra population, it is possible to estimate how fast individuals are aging compared to others by the discrepancy between epigenetic age and chronological age, dubbed age acceleration. Positive age acceleration indicates that an individual is biologically older than expected based on its chronological age. Age acceleration is predictive of all-cause mortality in humans<sup>18–22</sup>, which suggests that epigenetic models can be a powerful approach to study the impact of different factors on biological aging. Age-acceleration has also been associated with stress and adversity<sup>23–25</sup>, elevated glucocorticoids<sup>26,27</sup>, and inbreeding<sup>28–31</sup>, all of which are relevant for managing wild populations. The plains zebra population sampled in this study has a complex pedigree due to semi-captive breeding, which creates the opportunity to test whether inbreeding is associated with accelerated epigenetic aging in this population. The strong correlation of ECs to age can sometimes negatively affect their ability to detect age acceleration associated with biological variation<sup>32,33</sup>. We therefore also develop a second model for the plains zebra, the epigenetic pacemaker model (EPM). The EPM has previously been found to be useful for investigating how environmental and life-history factors influence aging<sup>34,35</sup>.

Our main goals for this study are to (1) develop epigenetic aging models for plains zebras; (2) test the ability of an EC developed for plains zebras to predict age in three equid species; and (3) estimate the influence of inbreeding levels on plains zebra aging patterns by correlating it with age acceleration predicted by both an EC and EPM. Our main results include an EC that predicts age accurately in plains zebras and three congeners tested, including domestic horse and two endangered sister species. We further show that inbreeding associated age acceleration increases with age, suggesting that inbreeding may have a cumulative effect on age acceleration throughout life. The development of epigenetic aging models in a wild equid stands to have broad impact because the crown group of the genus *Equus* comprises a closely related group of six extant species, five of which range from near threatened to critically endangered<sup>36,37</sup>.

## Results

**Epigenetic aging models.** We developed epigenetic models using methylation data profiles from three plains zebra data sets: (1) 76 blood samples, (2) 20 biopsy samples, and (3) 96 blood and biopsy samples combined (Table 1). For each data set, we developed both an EC and an EPM. We evaluated the effectiveness of applying the blood-based zebra EC to predict age in other equids using known-age domestic horses (*E. caballus*,  $n = 188$ ), Grevy's zebras (*E. grevyi*,  $n = 5$ ), and Somali asses (*E. africanus somaliensis*,  $n = 7$ ).

To develop the ECs we fit a generalized linear model with elastic-net penalization using leave-one-out (LOO) cross-validation. To improve EC fit<sup>14</sup> we square root transformed chronological age prior to fitting the models. The blood EC (Pearson's  $r = 0.96$ , median absolute error (MAE) = 0.56 years, Fig. 1a) and the combined tissue EC ( $r = 0.89$ , MAE = 0.62, Fig. 1c) predicted age more accurately than the biopsy EC ( $r = 0.62$ , MAE = 1.79 years, Fig. 1b). The blood EC selected 70 CpG sites, the biopsy clock 31 CpGs, and the combined clock selected 99 CpGs. The biopsy EC had no CpG sites in common with the blood and combined ECs. The blood EC and combined tissue EC shared only two CpGs. We report coefficients, intercepts, and lambdas in Supplementary Data 1 and present the results from using untransformed ages in Supplementary Fig. 1.

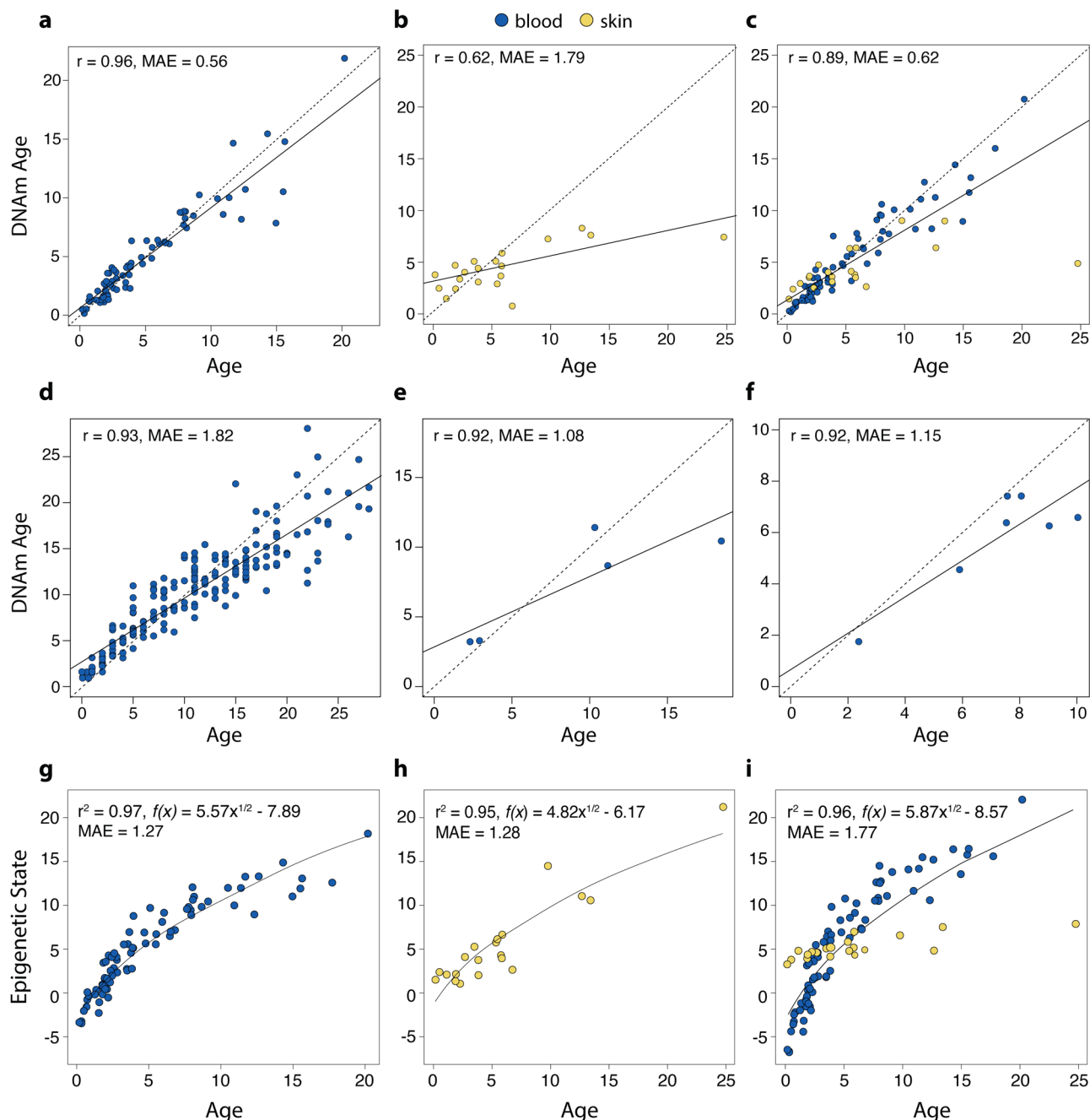
Cross-species predictive ability was high (Fig. 1d–f). The zebra-blood EC predicted horse age with high accuracy ( $r = 0.93$ , MAE = 1.82). The error when predicting the ages of horses younger than 15 years is lower (MAE 1.15) than when predicting the ages of older horses (MAE = 3.97). While prediction errors for Grevy's zebra and Somali wild ass ages were even lower (MAE of 1.08 and 1.15 respectively), this should be viewed with caution as we had limited sample sizes from Grevy's zebra ( $n = 5$ ) and Somali wild ass ( $n = 7$ ).

To construct plains zebra EPMs we used sites in which methylation levels were highly correlated with individual chronological age. Epigenetic states were estimated using a leave-one-out cross-validation. The blood, biopsy, and combined-tissue EPMs included 391, 242, and 248 CpG sites, respectively. We provide details of the CpGs selected by each model in

**Table 1** Description of the zebra data.

Tissue	N	No. female	Mean age	Min. age	Max. age
Blood	76	42	5.21	0.156	20.2
Biopsy	20	9	5.87	0.162	24.8

We restrict the description to animals whose ages could be estimated with high confidence (90% or higher). Tissue type,  $N$  = Total number of samples/arrays. Number of females. Age: mean, minimum and maximum.



**Fig. 1 Predictive ability of epigenetic aging models.** **a–c** Plains zebra epigenetic clocks (EC). We developed 3 ECs for plains zebras using square-root transformed ages: **a** blood samples ( $n = 76$ ), **b** biopsy samples ( $n = 20$ ), and **c** combined tissue types. Leave-one-sample-out (LOO) estimate of DNA methylation age are plotted against chronological age. Linear regressions of epigenetic age are indicated by a solid line while the diagonal dashed line depicts  $y = x$ . **d–f** Tests of the ability of the plains zebra blood clock to predict chronological age in other equids: **d** domestic horse  $n = 188$ , **e** Grevy's zebra  $n = 5$ , **f** Somali wild ass  $n = 7$ . **g–i** Epigenetic pacemaker (EPM) models for plains zebras. Epigenetic states (or epigenetic age) of plains zebras predicted from the EPM using **g** blood ( $n = 76$ ), **h** remote biopsy tissue ( $n = 20$ ), and **i** both sample types combined. Predictions are based on 76 blood samples and 20 biopsy samples. The equation of the fitted curve (solid line) is described for each plot. MAE are based on ages translated by the equation.

Supplementary Data 1. Epigenetic state was strongly correlated with chronological age in both tissue types: blood ( $r = 0.97$ , Fig. 1g) and biopsy ( $r = 0.95$ , Fig. 1h). EPMs based on the two tissues used largely distinct sets of CpG sites, sharing only 40 sites. Despite retaining a strong overall correlation in the combined EPM ( $r = 0.96$ ), the difference in the model's performance for the two sample types is apparent (Fig. 1i). The combined tissue pacemaker shared 138 sites with blood and only 34 with biopsies.

**Association of inbreeding with biological aging.** Inbreeding was estimated both as the inbreeding coefficient  $F$  and by the proportion of the genome in runs of homozygosity (FROH). We derived the genotypes used to estimate inbreeding from two sources: RAD sequencing data (42 samples) and genotypes imputed at the same set of loci from low-coverage whole-genome sequencing data using GLIMPSE<sup>38</sup> (28 additional samples). GLIMPSE produced high-quality imputations (mean dosage  $r^2$  of 80%; mean concordance between imputed and true genotype of

92%) as assessed by running leave-one-out imputations on 35 samples present in both the RADseq and low coverage data sets (Supplementary Fig. 2a, b). The Mendelian error rate across loci averaged 0.06 ( $\pm 0.06$ ) in the full data set. ROH were discoverable across 90% of the genome and ranged in size from 1.5 MB (minimum allowed size) up to >60 MB (Supplementary Fig. 2c). ROH over 10 MB are expected to reflect inbreeding loops occurring within the last five generations<sup>39,40</sup>. Individuals from more recent generations showed a pattern of excessive total ROH relative to the number of ROH segments as indicated by an upward shift in ROH size relative to number (Supplementary Fig. 2d)<sup>39</sup>.  $F_{\text{ROH}}$  ranged from 0 to 0.37, and  $F$  statistics ranged from  $-0.21$  to 0.32 (one outlier with an extreme negative  $F$  value was removed from the analysis). The Pearson correlation between  $F$  and  $F_{\text{ROH}}$  was 0.84.

We used multiple linear regressions to assess whether inbreeding is associated with age acceleration in the plains zebra population. Age acceleration was calculated by subtracting each individual's chronological age from its predicted age, and was calculated separately for the EPM and EC. Sex was added to the linear models as a covariate. EPM age acceleration was significantly associated with an interaction between chronological age and both  $F$  and  $F_{\text{ROH}}$  (Fig. 2 and Supplementary Table 1a, c), indicating that the impacts of inbreeding on biological age increase with chronological age. EC age acceleration was not associated with either measure of inbreeding. Sex was not associated with age acceleration in any model. A re-run of our analyses using only the RADseq samples gave similar results (Supplementary Table 1e–h).

**EWAS and functional analysis of plains zebra tissues.** We performed the EWAS analysis on the 31,836 probes that could be uniquely aligned to specific adjacent loci in the horse genome. Since the mammalian methylation array is based on stretches of DNA that are conserved in all mammals, the horse annotation can be applied to the zebra data<sup>41</sup>. At a nominal  $p < 10^{-4}$ , a total of 9757 and 331 probes were related to age in blood ( $n = 76$ , age-range 0.15–20.2 years) and biopsies ( $n = 20$ , age-range 0.16–24.8 years) respectively (Fig. 3a). The top age-related changes per tissue are as follows (Fig. 3a): blood, hypomethylation in *FANCL* upstream, *MAF* downstream, *ZNF608* upstream, and *PBX3* intron; biopsy, hypermethylation in *PLCB1*, *NEUROD1*, and *BARHL2* upstream. DNAm aging was distributed in both genic

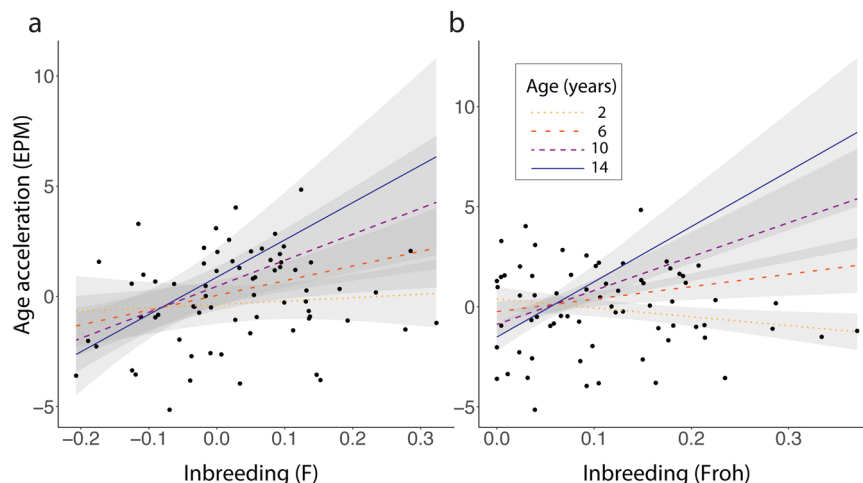
and intergenic regions relative to transcriptional start sites (Fig. 3b). Promoters and 5'UTR regions, which can be considered expression regulatory regions, mainly gained methylation with aging in both tissues. This observation paralleled a systematic positive correlation of CpG islands with age (Fig. 3c).

The association of CpG sites with chronological age in blood and biopsy was relatively similar, with a moderate positive correlation between the  $z$ -scores from the EWAS for each tissue type ( $r = 0.25$ , Fig. 3e). Of the CpGs with significant association with chronological age, only 81 overlapped between blood and biopsy samples (Fig. 3d). Some of these shared CpGs include hypermethylation in *PLCB1* exon, *RIMS1* exon, and hypomethylation in *NOVA2* intron and *NFIA* intron (Supplementary Fig. 3a, b). In contrast to results based on specific genes, functional enrichment analysis using GREAT<sup>42</sup> identified that age-related CpGs in both blood and biopsies were significantly related with regard to biological pathways, specifically development (e.g., nervous system) and survival, and were enriched with polycomb repressor complex 2 (e.g., *EED*, *SUZ12*, *PCR2*) target genes (Supplementary Fig. 4d).

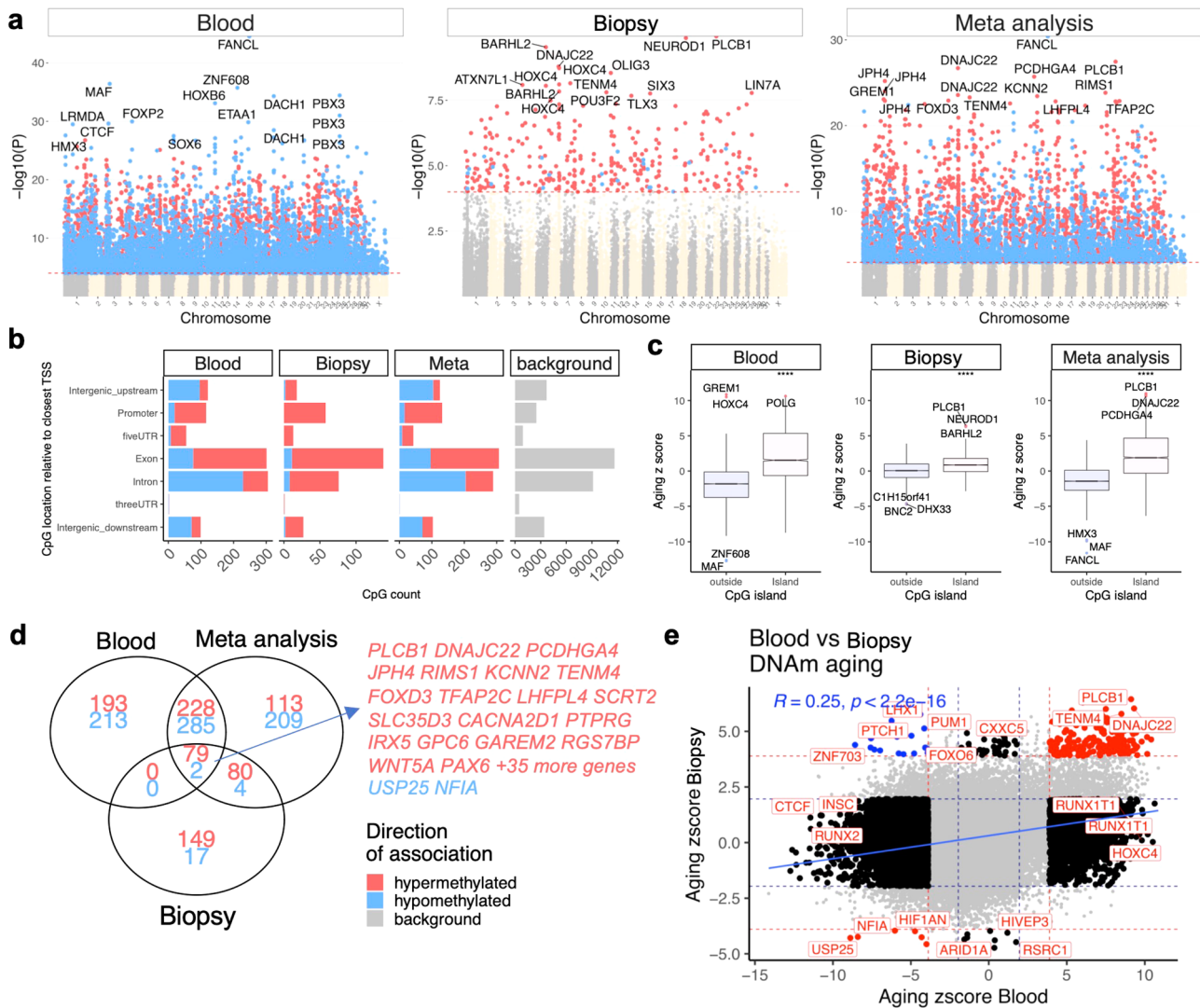
## Discussion

To the best of our knowledge, this is the first study to present DNA methylation-based age estimators for any wild equid. The high accuracy of the EC models reflects that we used a custom array that profiled 36 thousand probes that were highly conserved across numerous mammalian species. These robust data allowed us to construct highly accurate epigenetic aging models for plains zebras. The best model to predict chronological age was the EC developed from blood samples, which predicted individual age with a 6-month error.

Developing a highly accurate biopsy-based EC may be challenging due to the variability of tissue types within such a sample. Biopsy samples consist of three skin layers—the epidermis, the dermis, and the hypodermis—and may even contain deeper tissues. These individual layers can exhibit different methylation patterns<sup>43</sup> and are often present in different proportions across samples because they vary in thickness across the body and among individuals. A comparison between blood and skin-based (dermis and epidermis only) odontocete clocks also found the skin-based clocks to be less accurate<sup>44</sup>. Despite the inherent difficulties of using biopsy samples and the small sample size and skewed age



**Fig. 2 Relationship between epigenetic age acceleration calculated from the epigenetic pacemaker (EPM) model and inbreeding in plains zebras.** Lines represent the predicted age acceleration for individuals with different chronological ages and different levels of inbreeding. Gray areas show 95% confidence intervals. Black dots represent the individual plains zebra data. Inbreeding was calculated in PLINK as **a**  $F$  and **b**  $F_{\text{ROH}}$ .



**Fig. 3 Epigenome wide association study (EWAS) of chronological age in blood and skin of plains zebras.** **a** Manhattan plots of the EWAS of chronological age. Since a genome assembly was not available for zebra, the coordinates are estimated based on the alignment of Mammalian array probes to EquCab3.0.100 (domestic horse) genome. The direction of associations with  $p < 10^{-4}$  (red dotted line) is highlighted by red (hypermethylated) and blue (hypomethylated) colors. The top 15 CpGs were labeled by the neighboring genes. **b** Location of top age-related CpGs in each tissue relative to the closest transcriptional start site. Top CpGs were selected at  $p < 10^{-4}$  and further filtering based on z-score of association with chronological age for up to 500 in a positive and negative direction. The number of selected CpGs: blood, 1000; biopsy, 331; meta-analysis, 1000. The gray color represents the location of 3,1836 mammalian BeadChip array probes mapped to EquCab3.0.100 genome. **c** Boxplot of z-scores from a correlation of age with CpG location (within or outside CpG islands). The median Z statistics are significantly different ( $p < 10^{-4}$ ). **d** Venn diagram of the top age-related CpGs in blood and biopsy samples from plains zebras. **e** Sector plot of DNA methylation aging in plains zebra blood and biopsy tissues. Red dotted line:  $p < 10^{-4}$ ; blue dotted line:  $p > 0.05$ ; Red dots: shared CpGs; black dots: tissue-specific changes; blue dots: CpGs whose age correlation differs between blood and biopsy tissue.

distribution of the biopsy samples (Supplementary Fig. 5), our biopsy clock predicted age with an error of  $\pm 2$  years.

The plains zebra blood EC accurately predicted the chronological ages of horses, Grevy’s zebras, and Somali wild asses. This was expected since caballine and non-caballine equids are somewhat closely related (4–4.5 MYA)<sup>45,46</sup>. Non-caballine species of equids are more closely related to plains zebras (1.28–1.75 MYA) than domestic horses, which may explain the lower errors found for age estimation in Grevy’s zebras and Somali asses. In fact, chimpanzees and bonobos have a similar divergence time to those observed within the non-caballine equids<sup>47</sup> and align more closely to each other in DNAm age than either does with humans<sup>14</sup>.

The extraordinary accuracy of ECs stems from their utilization of sites that maximize a linear relationship between epigenetic age

and chronological age. This accuracy can sometimes negatively affect their ability to detect age acceleration associated with biological variation<sup>32,33</sup>. Methylation levels change in a non-linear fashion throughout individuals’ lifetimes in several species, with accelerated changes in early life and slower changes once individuals reach adulthood<sup>14,34</sup>. The EPM was developed to model these non-linear changes<sup>34,48,49</sup>. The EPM estimates epigenetic age by maximizing the similarity between estimated and observed methylation levels, and therefore does not make any assumptions about linearity but rather identifies the shape of the relationship between age and methylation directly from the data. In this sense, the EPM is potentially more associated with biological than chronological aging<sup>34</sup>, which may be particularly useful for investigating how environmental and life-history factors influence aging<sup>35</sup>.

The EPM models developed here reveal that epigenetic changes occur in a non-linear fashion throughout the lifespan of plains zebras. As has been found in humans and other species<sup>34,35</sup>, young zebras undergo faster epigenetic changes than adult zebras. We also observed that the variance in the estimates of epigenetic age is lower in young compared to old zebras. Increased variation in epigenetic age in adults is observed in humans and other species as well and may be a consequence of lifetime accumulation of environmental and physiological factors on the epigenome<sup>7,14</sup>. In agreement with the idea of cumulative effects on aging, the effects of inbreeding on epigenetic aging were more apparent in older individuals. We found a significant effect of the interaction between inbreeding measures and chronological age on age acceleration estimated from the blood EPM, wherein inbred individuals exhibited higher age acceleration at older chronological ages. The association between inbreeding and DNA methylation has been described in plants<sup>28,30</sup>, chickens<sup>31</sup>, and salmon<sup>29</sup>. Increases in inbreeding effects with age, such as those we describe, are predicted by theory<sup>50</sup> and have been shown empirically in an insect<sup>51</sup> and mammal<sup>52</sup> species.

Because the probes in the mammalian array were selected based on conservation in mammalian genomes, we expect our findings will have high translatability into humans and other mammals. The age-related gain of methylation in promoters (Fig. 3b) is consistent with observations in humans and many other species<sup>53,54</sup>. The low overlap of significant CpGs between tissues (Fig. 3d) may reflect the relatively low sample size ( $n = 20$  skin biopsy samples) or biological differences between the biopsy and whole blood samples. Some of the genes close to the most significant CpGs in the EWAS of blood and biopsy samples play key roles in the DNA damage pathway and maintenance of genomic integrity (*FANCL*)<sup>55</sup>, regulation of cellular and/or developmental processes (*MAF*<sup>56</sup>, *PBX3*<sup>57</sup>, *NEUROD1*<sup>58</sup>, *BARHL2*<sup>59,60</sup>, *NOVA2*<sup>61,62</sup>, *NFIA*<sup>63,64</sup>), and extra- and/or intra- cellular signaling (*PLCB1*<sup>65,66</sup>, *RIMS1*<sup>67,68</sup>).

Most of these genes are near to the top significant CpG sites from an EWAS that included multiple mammalian species and tissues: *MAF*, *NEUROD1*, *BARHL2*, *NFIA*, *PLCB1*, and *RIMS1*<sup>69</sup>. In mice, *MAF* promotes osteoblast differentiation and may be an important gene for age-related bone disease therapy<sup>70,71</sup>. *PBX3* is at the center of the most enriched network associated with aging based on differentially methylated regions in human blood<sup>72</sup>, and its expression levels in rat frontal cortex with aging may depend on the amount of stress experienced by the previous generations<sup>73</sup>. Many of these genes are potential targets for therapies to prevent and treat age-related cognitive decline and neurodegeneration. *BARHL2* is differentially expressed in the hippocampus of young and old rats<sup>74</sup>, and methylation levels in nearby CpGs are associated with aging across different tissues in naked mole rats<sup>75</sup>. *PLCB1*'s expression levels are associated with aging in the human prefrontal cortex<sup>76</sup>. SNPs near *ZNF608* have been associated with early stage of cognitive decline<sup>77</sup>, Alzheimer's disease risk<sup>78</sup>, and body mass index<sup>79,80</sup> in humans. *NEUROD1* is differentially expressed with aging in the mouse hippocampus and is a critical regulator of neurogenesis<sup>81</sup>. *RIMS1* is important for synaptic transmission at neuromuscular junctions in mammals<sup>82</sup>. The abundance of the *NOVA2* protein in the cytoplasm decreases with aging in human neurons<sup>83</sup>. *NFIA* modulates the plasticity of local circuits in the adult hippocampus and may be involved in the cortical atrophy associated with Alzheimer's disease<sup>84</sup>. In addition to the EWAS results, the GREAT enrichment analysis implicated developmental pathways, bivalent chromatin, and regions suppressed by polycomb repressor complex 2. These results are consistent with observations in many other mammalian species and corroborate

findings on DNAm aging in many other mammalian species<sup>3,5,14,17,32,35,69,75,85</sup>.

We expect that epigenetic aging models will be valuable for ecological studies and population management in wild species since age estimates can be used to assess reproductive potential and population viability. In known age populations, both ECs<sup>27,41,86,87</sup> and EPMS<sup>34,48,49</sup> have the potential to identify causes of individual accelerated aging. These models can be especially useful when combined with ecological or other genetic data<sup>4</sup>. Given the small number of CpG sites required, an aging project in a wild population could be done relatively inexpensively using bisulfite sequencing or pyrosequencing<sup>88–90</sup>, and the costs of the mammalian array are decreasing. While epigenetic models based on blood are more accurate and less invasive than many other options for aging mammals, there are drawbacks in that animals must be immobilized to obtain the samples. Because biopsy samples can be obtained with minimal disruption<sup>91</sup>, a highly reliable clock based on biopsy samples will be a worthwhile direction for future research. Since the methods for extracting genomic DNA from feces have improved<sup>92–94</sup>, it will be worthwhile to explore whether epigenetic aging models can be adapted to this non-invasive source of DNA.

## Methods

**Samples.** We obtained both whole blood (96) and remote biopsy (24) samples from a captive population of zebras maintained in a semi-wild state by the Quagga Project<sup>95</sup> in the Western Cape of South Africa. The population was founded in 1989 with 19 wild individuals (9 from Etosha National Park in Namibia and 10 from the Kwazulu-Natal in South Africa). Since its inception, the population has undergone artificial selection to reproduce the phenotype of the extinct quagga subspecies: no stripes on legs and hindquarters, and thinner and paler stripes in the head and barrel region. At sampling, we identified individuals by their unique stripe patterns and derived their chronological ages from studbook information, in which dates of birth are typically accurate to within one month. One exception is a biopsy sample from a founder that was captured for the project as a young mare and would have been at least 25 years old at sampling. We obtained remote biopsies using an air-powered rifle affixed with a 1 mm wide by 20–25 mm deep biopsy dart and preserved in RNAlater (Qiagen). Veterinarians collected blood opportunistically during activities of the Quagga Project and preserved them in EDTA tubes. All but four samples were collected from different individuals; two individuals were sampled twice at different ages (one and three years apart, respectively). We stored all samples at  $-20^{\circ}\text{C}$ . After eliminating samples (24 of 120) with <90% confidence for individual identity or age, we retained 76 blood samples and 20 biopsy samples, totaling 96 plains zebra samples (Table 1).

The collection of 188 whole-blood samples from domestic horses is described in detail in<sup>85</sup>. The Grevy's zebra ( $n = 5$ ) and Somali wild ass ( $n = 7$ ) are samples from zoo-based animals that were opportunistically collected and banked during routine health exams. The DNA methylation profiles from these samples have been reported previously<sup>69</sup>.

**Ethics approval.** We collected plains zebra samples under a protocol approved by the Research Safety and Animal Welfare Administration, University of California Los Angeles: ARC # 2009-090-31, approved initially in 2009.

**DNA methylation data.** We generated all DNA methylation data (plains zebra, horse, Somali wild ass, Grevy's zebra) using a custom Illumina methylation array (HorvathMammalMethylChip40)<sup>96</sup>. The array contains 36 thousand probes, 31,836 of which mapped uniquely to the horse genome<sup>97,98</sup>. We normalized methylation values from each species (plains zebra, horse, Somali wild ass, and Grevy's zebra) and tissue (blood and biopsy) using SeSAMe<sup>99</sup>. Unsupervised hierarchical clustering revealed that the plains zebra samples clustered by tissue (Supplementary Fig. 6).

**Epigenetic aging models.** We studied epigenetic aging in plains zebras using both EC<sup>14,100,101</sup> and EPM models<sup>34,49</sup>. For the ECs we fit generalized linear models in glmnet v.4.0-2 in R v.4.1.0<sup>102,103</sup>. We treated the methylation data from other equids (domestic horse, Grevy's zebra, Somali wild ass) independently and did not use them for zebra clock development. In addition to the LOO cross-validation presented here, we also conducted analyses using 10-fold cross-validation. The two forms of cross-validation did not produce appreciably different results.

We tested the ability of the blood-based EC to predict chronological age for other equids (domestic horse, Grevy's zebra, Somali wild ass) by inputting the DNA methylation profiles of these species into the plains zebra model. We used the

median absolute error of chronological age estimation and the Pearson correlation between predicted ages and known ages to assess accuracy for each species.

To construct plains zebra EPMs we used sites in which methylation levels were highly correlated with individual chronological age. The Pearson correlation ( $r$ ) thresholds for entry into the model were absolute values of 0.75 for blood and biopsy, and 0.6 for the combined EPM. The threshold used to select the sites for input into the combined EPM was lower because only one CpG site had  $r$  higher than 0.75. Epigenetic states were estimated using a leave-one-out cross-validation with EpigeneticPacemaker 0.0.3<sup>48</sup> in Python 3.7.4<sup>104</sup>. Supplementary Data 2 provides all Pearson coefficients for methylation levels against chronological age.

**Association of inbreeding with biological aging.** We prepared libraries and conducted  $2 \times 150$  bp paired-end RAD sequencing as described in<sup>105</sup>. To maximize the number of SNPs ascertained, we sequenced reads produced by two restriction enzymes, SbfI and PacI (74155 and 127429 cut sites in the horse genome, respectively). We aligned reads to the horse genome EquCab3<sup>97</sup> using BWA<sup>106</sup>. We called genotypes using the haplotype-based callers FreeBayes<sup>107</sup> and Sentieon<sup>108</sup>, retaining only the intersection of variants called by these callers. We retained only variants within the first read of each paired-end read (a total of ~60.5 Mb). We also removed indels, multiallelic and non-autosomal sites, loci genotyped in <10% of individuals, and individuals genotyped at <20% of loci. We followed GATK's basic guidelines for additional filtering (<https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants>), excluding SNPs with  $QD < 2$ ,  $FS > 60$ ,  $SOR > 3$ ,  $MQ < 40$ ,  $MQRankSum < -12.5$ , and  $ReadPosRankSum < -8$ . Finally, we removed SNPs with  $MAF < 0.01$ . Our filtering strategy resulted in 56 individuals genotyped at 322542 loci. The number of loci is consistent with levels of heterozygosity observed in inbred wild populations of plains zebras<sup>105</sup>.

The 56 samples with RADseq genotypes were used as a reference panel to impute these same 322542 SNPs in 89 individuals sequenced at low coverage. Low coverage libraries were constructed from 200 ng genomic DNA using Truseq Nano kit (Illumina), indexed with unique dual indices (Integrated DNA Technologies), and sequenced 24 libraries per lane on a HiSeqX platform (Illumina), generating  $6.2 \pm 1.4$  raw Gbp per sample. We aligned sequences to the horse genome EquCab3<sup>97</sup> using BWA<sup>106</sup>. Genotypes were then imputed using GLIMPSE<sup>38</sup>. As part of the GLIMPSE pipeline, genotype likelihoods were called from the low coverage bam using mpileup in bcftools<sup>109</sup>, ignoring indels and duplicate reads, and recalculating base alignment quality on the fly. Thirty-five of the 89 imputed individuals were also present in the RADseq data and were used to assess imputation quality using a leave-one-out approach.

Forty-two individuals from the RAD-seq data and 28 from the imputed data had associated epigenetic data. The combined sample of 70 individuals spans seven generations of the Quagga Project. With the exception of removing singletons and private doubletons, we did not further MAF prune or LD prune the combined RAD and imputed data, as such pruning can bias the detection of ROH in an inbred population<sup>110</sup>. We used 313,645 autosomal SNPs to estimate the inbreeding coefficient  $F$  and to detect runs of homozygosity (ROH).  $F$  was estimated in PLINK<sup>111</sup> using methods of moments. We used PLINK's default settings to detect runs of homozygosity (ROH) with the exception that we increased the stringency to require detection of runs in 150 rather than 100 bp windows, and final runs had to be at least 1.5 MB long. In addition, we allowed only two missing SNPs per homozygous window. We converted ROH to the inbreeding coefficient  $F_{ROH}$ <sup>112</sup> by dividing the total length of ROH for each individual by the length of the genome over which we screened for ROH<sup>110</sup>.

Linear regression models to assess the relationship between inbreeding and age acceleration were fitted with the lme4 v.0.9-38<sup>113</sup> package in R v.4.1.0<sup>103</sup>. We fitted four linear models in total, running separate analyses for the two estimates of inbreeding,  $F$  and  $F_{ROH}$ , and separate analyses for the two measures of age acceleration calculated based on EPM or EC. Age acceleration was the dependent variable and was calculated as the residuals of chronological age regressed on predicted age. In each analysis the independent variables were sex, chronological age, inbreeding, and the interaction between chronological age and inbreeding. We checked the residuals for normality and adjusted for heteroskedasticity via Huber-White with the package sandwich v.3.0-1<sup>114,115</sup>. We repeated our analyses using only the individuals genotyped directly with RAD-seq data to ensure imputed genotypes did not bias our results.

**EWAS and functional analysis of plains zebra tissues.** To identify genes potentially associated with aging, we performed EWAS in each tissue separately using the R function "standardScreeningNumericTrait" from the "WGCNA" R package<sup>116</sup>. The results were combined across tissues using Stouffer's meta-analysis method<sup>117</sup>. We estimated the distance of each significant CpG site to the closest transcriptional start site. We retained only the 500 CpGs with the most positive  $z$ -scores and the 500 with the most negative  $z$ -scores from each EWAS (blood, biopsy, and the combined meta-analysis). Restricting the number of analyzed CpGs did not have a drastic impact on the enriched pathway results. We used these CpGs as the input for GREAT analysis software<sup>42</sup>. The background was the human Hg19 genome, limited to 31,836 CpG sites mapping to the horse genome. The options in the analysis included "Basal plus extension" and a maximum of 50 kb flanking window for the CpGs coordinates.

**Statistics and reproducibility.** Statistical analyses were performed for the epigenetic models, inbreeding and age-acceleration, and EWAS and functional analysis. The analyses are described in the corresponding Methods sections, including all parameters used to allow reproducibility.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Methylation data for plains zebras can be downloaded from Gene Expression Omnibus GSE184223. RAD sequencing data is available as fastq files on SRA, BioProject ID: PRJNA670933. Data and scripts associated with model development and the creation of Figs. 1 and 2 are published on DRYAD (doi:10.5068/D1W39K).

## Code availability

Scripts used to run EWAS and functional analyses are at <https://github.com/shorvath/MammalianMethylationConsortium>.

Received: 27 August 2021; Accepted: 2 December 2021;

Published online: 17 December 2021

## References

- Beissinger, S. R. & Westphal, M. I. On the use of demographic models of population viability in endangered species management. *J. Wildl. Manag.* **62**, 821–841 (1998).
- Campana, S. Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *J. Fish. Biol.* **59**, 197–242 (2001).
- Polanowski, A. M., Robbins, J., Chandler, D. & Jarman, S. N. Epigenetic estimation of age in humpback whales. *Mol. Ecol. Resour.* **14**, 976–987 (2014).
- Jarman, S. N. et al. Molecular biomarkers for chronological age in animal ecology. *Mol. Ecol.* **24**, 4826–4847 (2015).
- Thompson, M. J., vonHoldt, B., Horvath, S. & Pellegrini, M. An epigenetic aging clock for dogs and wolves. *Aging* **9**, 1055–1068 (2017).
- De Paoli-Iseppi, R. et al. Measuring animal age with DNA methylation: from humans to wild animals. *Front. Genet.* **8**, 106 (2017).
- Bell, C. G. et al. DNA methylation aging clocks: challenges and recommendations. *Genome Biol.* **20**, 249 (2019).
- Field, A. E. et al. DNA methylation clocks in aging: categories, causes, and consequences. *Mol. Cell* **71**, 882–895 (2018).
- Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
- Petkovich, D. A. et al. Using DNA methylation profiling to evaluate biological age and longevity interventions. *Cell Metab.* **25**, 954–960 e956 (2017).
- Stubbs, T. M. et al. Multi-tissue DNA methylation age predictor in mouse. *Genome Biol.* **18**, 68 (2017).
- Wang, T. et al. Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction, and rapamycin treatment. *Genome Biol.* **18**, 57 (2017).
- Nussey, D. H., Froy, H., Lemaitre, J. F., Gaillard, J. M. & Austad, S. N. Senescence in natural populations of animals: widespread evidence and its implications for bio-gerontology. *Ageing Res. Rev.* **12**, 214–225 (2013).
- Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
- Voisin, S. et al. An epigenetic clock for human skeletal muscle. *J. Cachexia Sarcopenia Muscle* <https://doi.org/10.1002/jcsm.12556> (2020).
- De Paoli-Iseppi, R. et al. Age estimation in a long-lived seabird (*Ardenna tenuirostris*) using DNA methylation-based biomarkers. *Mol. Ecol. Resour.* **19**, 411–425 (2019).
- Ito, H., Udono, T., Hirata, S. & Inoue-Murayama, M. Estimation of chimpanzee age based on DNA methylation. *Sci. Rep.* **8**, 9998 (2018).
- Chen, B. H. et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging* **8**, 1844–1865 (2016).
- Christiansen, L. et al. DNA methylation age is associated with mortality in a longitudinal Danish twin study. *Aging Cell* **15**, 149–154 (2016).
- Horvath, S. et al. Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. *Aging* **7**, 1159–1170 (2018).
- Marioni, R. E. et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* **16**, 25 (2015).
- Perna, L. et al. Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin. Epigenetics* **8**, 64 (2016).
- Mitchell, C., Schepner, L. M. & Notterman, D. A. DNA methylation, early life environment, and health outcomes. *Pediatr. Res.* **79**, 212–219 (2016).

24. Pérez, R. F., Santamarina, P., Fernández, A. F., & Fraga, M. F. Epigenetics and Lifestyle: The Impact of Stress, Diet, and Social Habits on Tissue Homeostasis. In *Epigenetics and Regeneration* (ed. Palacios, D.) pp. 461–489 (Academic Press, 2019).
25. Szyf, M., Tang, Y. Y., Hill, K. G. & Musci, R. The dynamic epigenome and its implications for behavioral interventions: a role for epigenetics to inform disorder prevention and health promotion. *Transl. Behav. Med.* **6**, 55–62 (2016).
26. Lee, R. S. et al. Chronic corticosterone exposure increases expression and decreases deoxyribonucleic acid methylation of Fkbp5 in mice. *Endocrinology* **151**, 4332–4343 (2010).
27. Zannas, A. S. et al. Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. *Genome Biol.* **16**, 266 (2015).
28. Biemont, C. Inbreeding effects in the epigenetic era. *Nat. Rev. Genet.* **11**, 234 (2010).
29. Venney, C. J., Johansson, M. L. & Heath, D. D. Inbreeding effects on gene-specific DNA methylation among tissues of Chinook salmon. *Mol. Ecol.* **25**, 4521–4533 (2016).
30. Vergeer, P., Wagemaker, N. C. & Ouborg, N. J. Evidence for an epigenetic role in inbreeding depression. *Biol. Lett.* **8**, 798–801 (2012).
31. Han, W. et al. Genome-wide analysis of the role of DNA methylation in inbreeding depression of reproduction in Langshan chicken. *Genomics* **112**, 2677–2687 (2020).
32. Thompson, M. J. et al. A multi-tissue full lifespan epigenetic clock for mice. *Aging* **10**, 2832–2854 (2018).
33. Zhang, Q. et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med.* **11**, 54 (2019).
34. Snir, S., Farrell, C. & Pellegrini, M. Human epigenetic ageing is logarithmic with time across the entire lifespan. *Epigenetics* **14**, 912–926 (2019).
35. Pinho, G. M. et al. Hibernation slows epigenetic aging in yellow-bellied marmots. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.07.434299> (2021).
36. Moehlman, P. D. *Equids: Zebras, Asses, and Horses Status Survey and Conservation Action Plan* Vol. 37, 190 pp (IUCN/SSC Equid Specialist Group, 2002).
37. Moehlman, P. D. & King, S. R. B. IUCN SSC Equid Specialist Group 2020 Report. <https://www.iucn.org/commissions/ssc-groups/mammals/mammals-a-e/equid> (2020).
38. Rubinacci, S., Ribeiro, D. M., Hofmeister, R. & Delaneau, O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* **53**, 120–126 (2021).
39. Ceballos, F. C., Hazelhurst, S. & Ramsay, M. Runs of homozygosity in sub-Saharan African populations provide insights into complex demographic histories. *Hum. Genet.* **138**, 1123–1142 (2019).
40. Curik, I., Ferenčaković, M. & Sölkner, J. Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci.* **166**, 26–34 (2014).
41. Anderson, J. A. et al. The costs of competition: high social status males experience accelerated epigenetic aging in wild baboons. *eLife* **10**, e66128 (2020).
42. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* <https://doi.org/10.1038/nbt.1630> (2010).
43. Gronniger, E. et al. Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS Genet.* **6**, e1000971 (2010).
44. Robeck, T. R. et al. Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Commun. Biol.* <https://doi.org/10.1038/s42003-021-02179-x> (2021).
45. Jonsson, H. et al. Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc. Natl Acad. Sci. USA* **111**, 18655–18660 (2014).
46. Vilstrup, J. T. et al. Mitochondrial phylogenomics of modern and ancient equids. *PLoS One* **8**, e55950 (2013).
47. Jensen-Seaman, M. I. & Hooper-Boyd, K. A. in *Encyclopedia of Life Sciences (ELS)* (John Wiley & Sons, Ltd., 2008).
48. Farrell, C., Snir, S. & Pellegrini, M. The epigenetic pacemaker—modeling epigenetic states under an evolutionary framework. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btaa585> (2020).
49. Snir, S. & Pellegrini, M. An epigenetic pacemaker is detected via a fast conditional expectation maximization algorithm. *Epigenomics* **10**, 695–706 (2018).
50. Charlesworth, B. & Hughes, K. A. Age-specific inbreeding depression and components of genetic variance in relation to the evolution of senescence. *Proc. Natl Acad. Sci. USA* **93**, 6140–6145 (1996).
51. Fox, C. W. Inbreeding depression increases with maternal age. *Evolut. Ecol. Res.* **12**, 961–972 (2010).
52. Benton, C. H. et al. Inbreeding intensifies sex- and age-dependent disease in a wild mammal. *J. Anim. Ecol.* **87**, 1500–1511 (2018).
53. Mayne, B., Berry, O., Davies, C., Farley, J. & Jarman, S. A genomic predictor of lifespan in vertebrates. *Sci. Rep.* **9**, 17866 (2019).
54. McClain, A. T. & Faulk, C. The evolution of CpG density and lifespan in conserved primate and mammalian promoters. *Aging* **10**, 561–572 (2018).
55. Alpi, A. F., Pace, P. E., Babu, M. M. & Patel, K. J. Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. *Mol. Cell* **32**, 767–777 (2008).
56. Kannan, M. B., Solovieva, V. & Blank, V. The small MAF transcription factors MAFF, MAFG, and MAFK: current knowledge and perspectives. *Biochim. Biophys. Acta* **1823**, 1841–1846 (2012).
57. Li, Z. et al. PBX3 is an important cofactor of HOXA9 in leukemogenesis. *Blood* **121**, 1422–1431 (2013).
58. Malecki, M. T. et al. Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nat. Genet.* **23**, 323–328 (1999).
59. Ding, Q., Joshi, P. S., Xie, Z. H., Xiang, M. & Gan, L. BARHL2 transcription factor regulates the ipsilateral/contralateral subtype divergence in postmitotic dII neurons of the developing spinal cord. *Proc. Natl Acad. Sci. USA* **109**, 1566–1571 (2012).
60. Mo, Z., Li, S., Yang, X. & Xiang, M. Role of the Barhl2 homeobox gene in the specification of glycinergic amacrine cells. *Development* **131**, 1607–1618 (2004).
61. Giampietro, C. et al. The alternative splicing factor Nova2 regulates vascular development and lumen formation. *Nat. Commun.* **6**, 8479 (2015).
62. Yano, M., Hayakawa-Yano, Y., Mele, A. & Darnell, R. B. Nova2 regulates neuronal migration through an RNA switch in disabled-1 signaling. *Neuron* **66**, 848–858 (2010).
63. Deneen, B. et al. The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord. *Neuron* **52**, 953–968 (2006).
64. Hiraike, Y. et al. NFIA co-localizes with PPARgamma and transcriptionally controls the brown fat gene program. *Nat. Cell Biol.* **19**, 1081–1092 (2017).
65. Caricasole, A., Sala, C., Roncarati, R., Formenti, E. & Terstappen, G. C. Cloning and characterization of the human phosphoinositide-specific phospholipase C-beta 1 (PLCβ1). *Biochim. Biophys. Acta* **1517**, 63–72 (2000).
66. McOmish, C. E., Burrows, E. L., Howard, M. & Hannan, A. J. PLC-beta1 knockout mice as a model of disrupted cortical development and plasticity: behavioral endophenotypes and dysregulation of RGS4 gene expression. *Hippocampus* **18**, 824–834 (2008).
67. Mittelstaedt, T., Alvarez-Baron, E. & Schoch, S. RIM proteins and their role in synapse function. *Biol. Chem.* **391**, 599–606 (2010).
68. Schoch, S. et al. RIM1a forms a protein scaffold for regulating neurotransmitter release at the active zone. *Nature* **415**, 321–326 (2002).
69. Lu, A. T. et al. Universal DNA methylation age across mammalian tissues. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.01.18.426733> (2021).
70. Nishikawa, K. et al. Maf promotes osteoblast differentiation in mice by mediating the age-related switch in mesenchymal cell differentiation. *J. Clin. Invest.* **120**, 3455–3465 (2010).
71. Saidak, Z., Hay, E., Marty, C., Barbara, A. & Marie, P. J. Strontium ranelate rebalances bone marrow adipogenesis and osteoblastogenesis in senescent osteopenic mice through NFATc/Maf and Wnt signaling. *Aging Cell* **11**, 467–474 (2012).
72. McClay, J. L. et al. A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. *Hum. Mol. Genet.* **23**, 1175–1185 (2014).
73. Ambeskov, M. et al. Ancestral stress programs sex-specific biological aging trajectories and non-communicable disease risk. *Aging* **12**, 3828–3847 (2020).
74. Burger, C., Lopez, M. C., Baker, H. V., Mandel, R. J. & Muzyczka, N. Genome-wide analysis of aging and learning-related genes in the hippocampal dentate gyrus. *Neurobiol. Learn. Mem.* **89**, 379–396 (2008).
75. Horvath, S. et al. DNA methylation clocks show slower progression of aging in naked mole-rat queens. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.15.435536> (2021).
76. Rapoport, S. I., Primiani, C. T., Chen, C. T., Ahn, K. & Ryan, V. H. Coordinated expression of phosphoinositide metabolic genes during development and aging of human dorsolateral prefrontal cortex. *PLoS One* **10**, e0132675 (2015).
77. Dube, J. B. et al. Genetic determinants of “cognitive impairment, no dementia”. *J. Alzheimers Dis.* **33**, 831–840 (2013).
78. Hinney, A. et al. Genetic variation at the CELF1 (CUGBP, elav-like family member 1 gene) locus is genome-wide associated with Alzheimer’s disease and obesity. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **165B**, 283–293 (2014).
79. Ntalla, I. et al. Replication of established common genetic variants for adult BMI and childhood obesity in Greek adolescents: the TEENAGE study. *Ann. Hum. Genet.* **77**, 268–274 (2013).
80. Speliotes, E. K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
81. Gao, Z. et al. Neurod1 is essential for the survival and maturation of adult-born neurons. *Nat. Neurosci.* **12**, 1090–1092 (2009).
82. Badawi, Y. & Nishimune, H. Presynaptic active zones of mammalian neuromuscular junctions: Nanoarchitecture and selective impairments in aging. *Neurosci. Res.* **127**, 78–88 (2018).



83. Tollervey, J. R. et al. Analysis of alternative splicing associated with aging and neurodegeneration in the human brain. *Genome Res.* **21**, 1572–1582 (2011).
84. Kim, B. H., Nho, K. & Lee, J. M., Alzheimer's Disease Neuroimaging, I. Genome-wide association study identifies susceptibility loci of brain atrophy to NFIA and ST18 in Alzheimer's disease. *Neurobiol. Aging* **102**, 200 e201–200 e211 (2021).
85. Horvath, S. et al. DNA methylation aging and transcriptomic studies in horses. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.11.435032> (2021).
86. Benayoun, B. A., Pollina, E. A. & Brunet, A. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat. Rev. Mol. Cell Biol.* **16**, 593–610 (2015).
87. Quach, A. et al. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging* **9**, 419–446 (2017).
88. Cray-Dooley, F. K. et al. A comparison of existing global DNA methylation assays to low-coverage whole-genome bisulfite sequencing for epidemiological studies. *Epigenetics* **12**, 206–214 (2017).
89. Reed, K., Poulin, M. L., Yan, L. & Parisenti, A. M. Comparison of bisulfite sequencing PCR with pyrosequencing for measuring differences in DNA methylation. *Anal. Biochem.* **397**, 96–106 (2010).
90. Tost, J., Dunker, J. & Gut, I. G. Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques* **35**, 152–156 (2003).
91. Karesh, W. B. in *Zoo and Wild Animal Medicine: Current Therapy* (eds Fowler Murray, E. & Eric Miller, R.) 298–308 (Saunders Elsevier, 2008).
92. Chiou, K. L. & Bergey, C. M. Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. *Sci. Rep.* **8**, 1975 (2018).
93. Orkin, J. D. et al. The genomics of ecological flexibility, large brains, and long lives in capuchin monkeys revealed with fecalFACS. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.2010632118> (2021).
94. Snyder-Mackler, N. et al. Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples. *Genetics* **203**, 699–714 (2016).
95. Harley, E. H., Knight, M. H., Lardner, C., Wooding, B. & Gregor, M. The Quagga project: progress over 20 years of selective breeding. *South African J. Wildlife Res.* <https://doi.org/10.3957/056.039.0206> (2009).
96. Arneson, A. et al. A mammalian methylation array for profiling methylation levels at conserved sequences Preprint at *bioRxiv* <https://doi.org/10.1101/2021.01.07.425637> (2021).
97. Kalbfleisch, T. S. et al. Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun. Biol.* **1**, 197 (2018).
98. Wade, C. M. et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* <https://doi.org/10.1126/science.1178158> (2009).
99. Zhou, W., Triche, T. J. Jr., Laird, P. W. & Shen, H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Res.* **46**, e123 (2018).
100. Bocklandt, S. et al. Epigenetic predictor of age. *PLoS One* **6**, e14821 (2011).
101. Hannum, G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
102. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
103. R Core Team. R: A language and environment for statistical computing (R Foundation for Statistical Computing, 2020).
104. Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual* (CreateSpace, 2009).
105. Larison, B. et al. Population structure, inbreeding and stripe pattern abnormalities in plains zebras. *Mol. Ecol.* **30**, 379–390 (2021).
106. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
107. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at <https://arxiv.org/abs/1207.3907> (2012).
108. Freed, D., Aldana, R., Weber, J. A. & Edwards, J. S. The Sentione Genomics Tools—A fast and accurate solution to variant calling from next-generation sequence data. Preprint at *bioRxiv* <https://doi.org/10.1101/115717> (2017).
109. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
110. Meyermans, R., Gorssen, W., Buys, N. & Janssens, S. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics* **21**, 94 (2020).
111. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
112. McQuillan, R. et al. Runs of homozygosity in European populations. *Am. J. Hum. Genet.* **83**, 359–372 (2008).
113. Zeileis, A. & Hothorn, T. Diagnostic checking in regression relationships. *R News* **2**, 7–10 (2002).
114. Zeileis, A. Econometric computing with HC and HAC covariance matrix estimators. *J. Stat. Softw.* **11**, 1–17 (2004).
115. Zeileis, A., Köll, S. & Graham, N. Various versatile variances: an object-oriented implementation of clustered covariances in R. *J. Stat. Softw.* **95**, 1–36 (2020).
116. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
117. Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A. & Williams, R. M. *J. Adjustment During Army Life* (Princeton University Press, 1949).

## Acknowledgements

This work was supported by the Paul G. Allen Frontiers Group (SH). G.P. was supported by the Science Without Borders program of the National Counsel of Technological and Scientific Development of Brazil. Sample collection was supported by National Geographic Grant 8941-11 (BL). We thank the following people for their assistance with acquiring the plains zebra blood and tissue samples: Colleen O’Ryan, Stephen Mitchell, Mick D’Alton, Tom Turner, Boet le Roux, Evert Grobbelar, Ewald Groenewald, Basie and Coenraad Bezuidenhout, Linda Mason, Patricia Swanepoel, Fernando Rueda, Ross Cowlin, Melissa Stander, Hanna Lindstadt, Ansel Abels, J.P. Hugo, Cobus van Collier, Jannie du Plessis, and SANParks. White Oak Conservation kindly provided Grevy’s zebra and Somali wild ass samples.

## Author contributions

B.L., S.H. and G.P. conceived of the study. B.L. and G.P. analyzed data, and B.L., G.P., and S.H. co-wrote the article. The remaining authors helped with the statistical analysis (A.H., J.Z., C.L., C.F., and M.P.), or the data generation (B.W., C.J.F., C.K., G.B., T.R., and D.M.). All authors reviewed and edited the article.

## Competing interests

The authors declare the following competing interests: S.H. is a founder of the non-profit Epigenetic Clock Development Foundation which plans to license several patents from his employer UC Regents. These patents list S.H. as inventor. The other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02935-z>.

**Correspondence** and requests for materials should be addressed to Brenda Larison or Steve Horvath.

**Peer review information** *Communications Biology* thanks Aaron Shafer and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: Christopher Hine and Luke R. Grinham. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021