Title: Designing epigenetic clocks for wildlife research

Running title: Wildlife epigenetic clocks

Levi Newediuk[1*], Evan Richardson[2], Alyssa Bohart[3], Amélie Roberto-Charron[3], Colin J Garroway[1£*], Meaghan J Jones[4£*]

[1] Department of Biological Sciences, University of Manitoba, Canada
[2] Environment and Climate Change Canada, Winnipeg, Manitoba, Canada
[3] Department of Environment, Government of Nunavut
[4] Department of Biochemistry and Medical Genetics, University of Manitoba, and Children's Hospital Research Institute of Manitoba

£ Denotes co-senior authors

* Corresponding authors: Levi Newediuk (Levi.Newediuk@umanitoba.ca), Colin J Garroway (Colin.Garroway@umanitoba.ca), Meaghan J Jones (Meaghan.Jones@umanitoba.ca)

Abstract

The potential applications of epigenetic clocks are expanding in wildlife conservation and management. The pace at which they are being adopted highlights the need for field-specific design best practices. Epigenetic clocks were originally developed for human studies, presenting challenges for their adoption in wildlife research. Most notably, the estimated ages of sampled wildlife can be unreliable, and sampling restrictions limit the number and variety of available samples, which can reduce the accuracy of epigenetic clocks for wildlife. In this article, we present a detailed workflow for designing, validating, and applying wildlife epigenetic clocks in a way that accounts for sampling constraints. We provide recommendations for two main applications of wildlife epigenetic clocks: estimating unknown ages and assessing cumulative biological aging. Our simulations and analyses, applied to an extensive polar bear dataset from across the Canadian Arctic, demonstrate that accurate epigenetic clocks for wildlife can be constructed and validated with limited samples, accommodating projects with small budgets and sampling constraints. With our workflow and examples, we hope to make epigenetic clock use more accessible and widespread in wildlife conservation and management.

Keywords: epigenetic clock, DNA methylation, biomarker, wildlife monitoring, biodiversity conservation, age estimation

Introduction

Over the past decade, epigenetic clocks—models that predict age based on DNA methylation patterns—have transformed human biomedicine by revealing how stressful life experiences accelerate biological aging, leading to disease (Lu et al., 2019) and early mortality (Chen et al., 2016; Marioni et al., 2015). Now, epigenetic clocks are poised for similar impacts in wildlife management and conservation biology, as biological age provides a novel and non-lethal means to estimate key conservation and management metrics like age structure and the cumulative lifetime stress underlying individual survival and population declines. Although other biological aging methods have provided some of these insights, the superior precision and accuracy of epigenetic clocks set them apart as a uniquely promising tool (Le Clercq et al., 2023).

While a universal clock was recently published for all mammals (Lu et al., 2023), the most accurate epigenetic clocks are species-specific. These custom clocks present a new design challenge. Relative to human and model organism samples collected under controlled settings, wildlife sampling is logistically challenging, often underfunded, and time-intensive, making it difficult to sample at the required intensity to account for the biological variation that leads to biases in tissue types, sexes, and ages of samples known to experience distinct DNA methylation patterns (McEwen et al., 2020; Simpkin et al., 2016; Yusipov et al., 2020). Despite several species-specific epigenetic clocks already developed (Bors et al., 2021; Czajka et al., 2024; Newediuk et al., 2024; Parsons et al., 2023), there has been limited discussion on best practices for sampling wildlife DNA and designing epigenetic clocks to deal with these biases. Moreover, few species-specific clocks have been independently validated for accuracy across different populations, making it difficult to detect when biases are present.

This paper is intended to provide practical recommendations for designing species-specific epigenetic clocks for wildlife, focusing on minimizing the impact of sampling bias on their accuracy. We begin with an overview of epigenetic clock models, covering what they measure and where they could be applied in wildlife conservation and management. Then, we discuss the key design considerations important for minimizing bias in wildlife epigenetic clocks: representative sampling, feature selection, and validation methods sensitive to small sample sizes. We frame our discussion around comparisons of epigenetic clock design approaches using simulations and an extensive DNA methylation dataset from several wild polar bear (*Ursus maritimus*) populations (Box 1). Accompanying the discussion, we provide a comprehensive workflow that guides the reader through each major step and decision in developing a species-specific epigenetic clock (Figure 1).

Overview of epigenetic clocks

Epigenetic clocks are regression models that estimate age based on predictable changes to DNA methylation that occur over lifetimes. DNA methylation (DNAm) is an important regulator of gene expression and cellular identity that consists of a methyl group on a cytosine-guanine sequence (CpG) in DNA (Bestor et al., 2015). Though biologically vital, DNAm is not static and its maintenance declines with age, resulting in increasing variability and higher rates of errors over the lifespan. The majority of mammalian genomes are highly methylated, and so age-related variability and error results in a global loss of DNAm with age (Jung & Pfeifer, 2015). However, specific CpG sites—conserved across mammals—undergo highly predictable changes with chronological age (Horvath, 2013; Lu et al., 2023). Epigenetic clocks leverage these predictable,

age-associated DNA methylation changes to estimate chronological age (Hannum et al., 2013; Horvath, 2013; Lu et al., 2023).

Most epigenetic clocks are constructed using elastic net regression. This penalized regression method identifies a small subset of thousands of CpG sites—sometimes as few as a dozen—that most accurately predict age across a set of DNA samples. The specific CpG sites selected often vary each time the elastic net regression model is fit to the same set of samples. This means caution should be used in causal interpretations of DNA methylation related to gene function at specific sites (Moqri et al., 2023). Still, the resulting age predictions on new samples are generally stable and accurate (Haftorn et al., 2023; Hannum et al., 2013; Horvath, 2013).

Epigenetic clock accuracy depends primarily on matching the sample size and the composition of the samples used to construct the model to the population it will be used on. Larger samples improve accuracy (Mayne et al., 2021), and theoretically, perfectly accurate age estimates are possible with sufficiently large samples (Q. Zhang et al., 2019). However, factors that influence DNAm patterns in mammals including sex, tissue type, and genetic ancestry, can also be important confounding factors that reduce the accuracy of epigenetic clocks if they are mismatched between the training and test samples. While clocks constructed using samples from one sex, tissue type, or genetic ancestry will likely produce stable and accurate age predictions for those groups, they may be less accurate when applied to new samples with different compositions (Carja et al., 2017; McEwen et al., 2020; Simpkin et al., 2016; Yusipov et al., 2020). Moreover, the rate of DNA methylation change is inconsistent across the lifespan, with changes occurring faster in early life (Horvath & Raj, 2018; McEwen et al., 2020), making it important to model the process using samples spanning a wide range of ages to ensure accuracy. We explore how these biases impact wildlife epigenetic clock accuracy in section A.

Wildlife applications of epigenetic clocks

Reasons for estimating wildlife epigenetic age

There are two main reasons to estimate epigenetic age in wildlife: accurately estimating the unknown ages of animals to improve information about population age structure and age-specific vital rates and assessing biological age acceleration. While age acceleration is the primary focus of biomedical epigenetic clock research due to the human health implications of epigenetic aging, both applications are potentially valuable for wildlife conservation and management.

Existing methods for estimating the ages of wildlife are often either limited in precision or require invasive sampling (Calvert & Ramsay, 1998; Y. Zhang et al., 2024). Morphological biomarkers, such as tooth cementum annuli or aspartic acid racemization in eye lenses, measure age-related changes but typically require post-mortem samples. Additionally, the accuracy of some of these methods varies with age, often providing imprecise age estimates for younger individuals (Garde et al., 2018). Telomere length is another age-associated molecular marker and a less invasive alternative to morphological approaches, requiring only samples like blood and skin that can be collected from live animals. However, this method is less accurate than epigenetic aging (Le Clercq et al., 2023). Epigenetic clocks, therefore, could potentially provide the best and least invasive information about age structure outside of long-term population tracking.

Epigenetic clocks assess biological age acceleration as the residual difference between an animal's known chronological age and its epigenetic age as predicted by the clock. Positive age acceleration has been linked to environmental stress (Zannas, 2019), disease (Lu et al., 2019; Perna et al., 2016), and early death (Marioni et al., 2015) in humans. As in humans, epigenetic clocks for wildlife can only estimate age acceleration when applied to known-age samples. Challenges in accurately aging wildlife using morphological or other methods, as discussed above, will limit the studies that can create or apply epigenetic clocks in this way. However, where known-age samples are available, new research has identified connections between ecologically relevant environmental stressors and epigenetic age acceleration (Anderson et al., 2021; Newediuk et al., 2024).

Applying epigenetic age estimates in wildlife conservation and management

The potential applications of accurate age predictions for wildlife management and conservation are two-fold. First, accurate age estimates could help track changes in wildlife population growth and survival rates, which often coincide with shifts in population age structures (Jackson et al., 2020). Age-structured population models, which track these dynamics (Holmes et al., 2007; Hostetter et al., 2021), rely on accurate age data from individuals. For samples from populations already aged using other accurate methods, epigenetic age acceleration could provide a standard means to assess the severity of environmental stressors and their consequences for population health. Unlike traditional wildlife stress biomarkers, such as glucocorticoid hormone levels—which are highly variable and lack a clear reference point for an "unstressed" animal (Romero & Beattie, 2022)—epigenetic age acceleration is relatively stable and has been consistently associated with stress and health (Lu et al., 2019; Perna et al., 2016; Zannas, 2019).

Most importantly, epigenetic aging has the potential to detect populations experiencing environmental stressors before declines occur, which could facilitate timely conservation and management interventions. Current metrics for assessing the impacts of stress on populations, such as population dynamics and genetic diversity, are lagging indicators that often reflect the cumulative effects of stress following several generations of poor survival and reproductive success. In contrast, epigenetic age accelerates in response to stress experienced within the lifespan of individual animals, positioning it as a leading indicator to identify populations at risk of future declines.

Recommended workflow

Building an epigenetic clock starts with collecting tissue samples (Figure 1A) and extracting the DNA. DNA extraction is often followed by bisulfite treatment to convert non-methylated cytosines to uracil, which enables their differentiation from methylated nucleotides. Methylation levels at target CpG sites in the bisulfite-converted DNA are then measured.

The Horvath Mammalian Array is the most common platform for measuring DNAm. Adapted from earlier microarrays designed for human DNA, the array includes just over 37,000 50-bp target sites, including conserved CpG sites and their flanking sequences (Arneson et al., 2022). The sites were selected from an alignment of 62 mammal species with the human genome. While not all sites are expected to align to the genome of every mammal species, most genomes tested align to at least half of the sites on the array, and DNAm at a subset of those is expected to change predictably with age (See Section C; Arneson et al., 2022).

In the following workflow, we assume readers measured methylation in bisulfite-converted DNA using the Horvath Mammalian Array. However, alternative workflows are also possible, such as quantifying DNAm in bisulfite-converted DNA through targeted or whole-genome next-generation sequencing (Kurdyukov & Bullock, 2016). Regardless of how DNAm is quantified, the considerations we discuss regarding sampling and clock design are broadly applicable to wildlife epigenetic clocks and remain relevant across different workflows.

Once the raw DNAm data are collected, they must be processed into a format suitable for elastic net regression. This process involves several optional decisions about quality control (Figure 1B) and pre-processing (Figure 1C). Finally, the clock's accuracy must be validated using one of several possible approaches (Figure 1D). Accuracy is typically assessed by the median absolute error (MAE) of the absolute differences between observed and predicted ages and either the coefficient of determination (R-squared) of the linear relationship between epigenetic age and chronological age or Pearson's correlation coefficient (the "age correlation" — Horvath & Raj, 2018). A low MAE indicates the clock estimates epigenetic age with high precision, and a high R-squared or age correlation indicates a stronger linear relationship between epigenetic age and chronological age. Together, these are characteristics of an accurate clock (Figure 2).

The intended application of a wildlife epigenetic clock must be considered before its construction, as key decisions made at this stage will influence its value for accurately estimating ages versus assessing epigenetic age acceleration. Because accuracy improves with sample size (Q. Zhang et al., 2019), constructing clocks with as many samples as possible is ideal for estimating chronological age in unknown-age samples. Variation in DNAm rates due to tissue, sex, and genetic ancestry can be dealt with during quality control, pre-processing, and validation (see sections B, C and D). However, perfectly accurate clocks cannot, by definition, detect residual differences between chronological and epigenetic age, making them unsuitable for estimating epigenetic age acceleration (Q. Zhang et al., 2019). Moreover, even after pre-processing and quality control, clocks designed for narrow applications—such as a single population—might perform poorly when applied to new sample types.

In the following sections, we explore these considerations in more detail, outlining epigenetic clock design decisions related to sample collection, data quality control checks, pre-processing, and clock validation. For each application of wildlife epigenetic clocks—estimating unknown ages and assessing epigenetic age acceleration—we provide tailored recommendations.

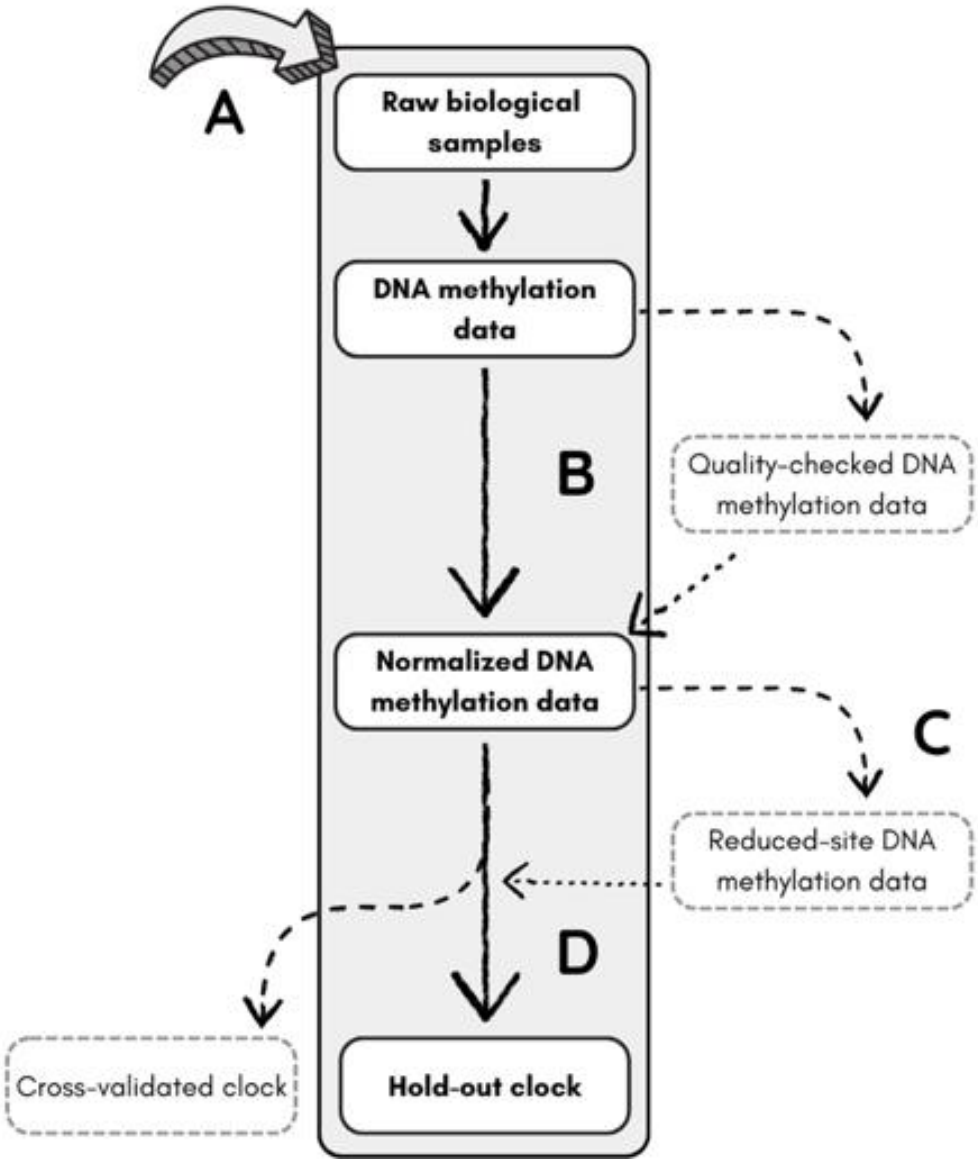


Figure 1 Our recommended workflow (solid lines) for developing epigenetic clocks includes (A) deciding on the sample size and characteristics required to train an accurate clock, extracting the DNA, and quantifying DNA methylation; (B) performing optional quality-control tests and normalizing the DNA methylation data; (C) performing pre-processing steps to limit the number of features used to fit the clock; and (D) validating the clock. Dotted lines indicate optional or alternative steps.

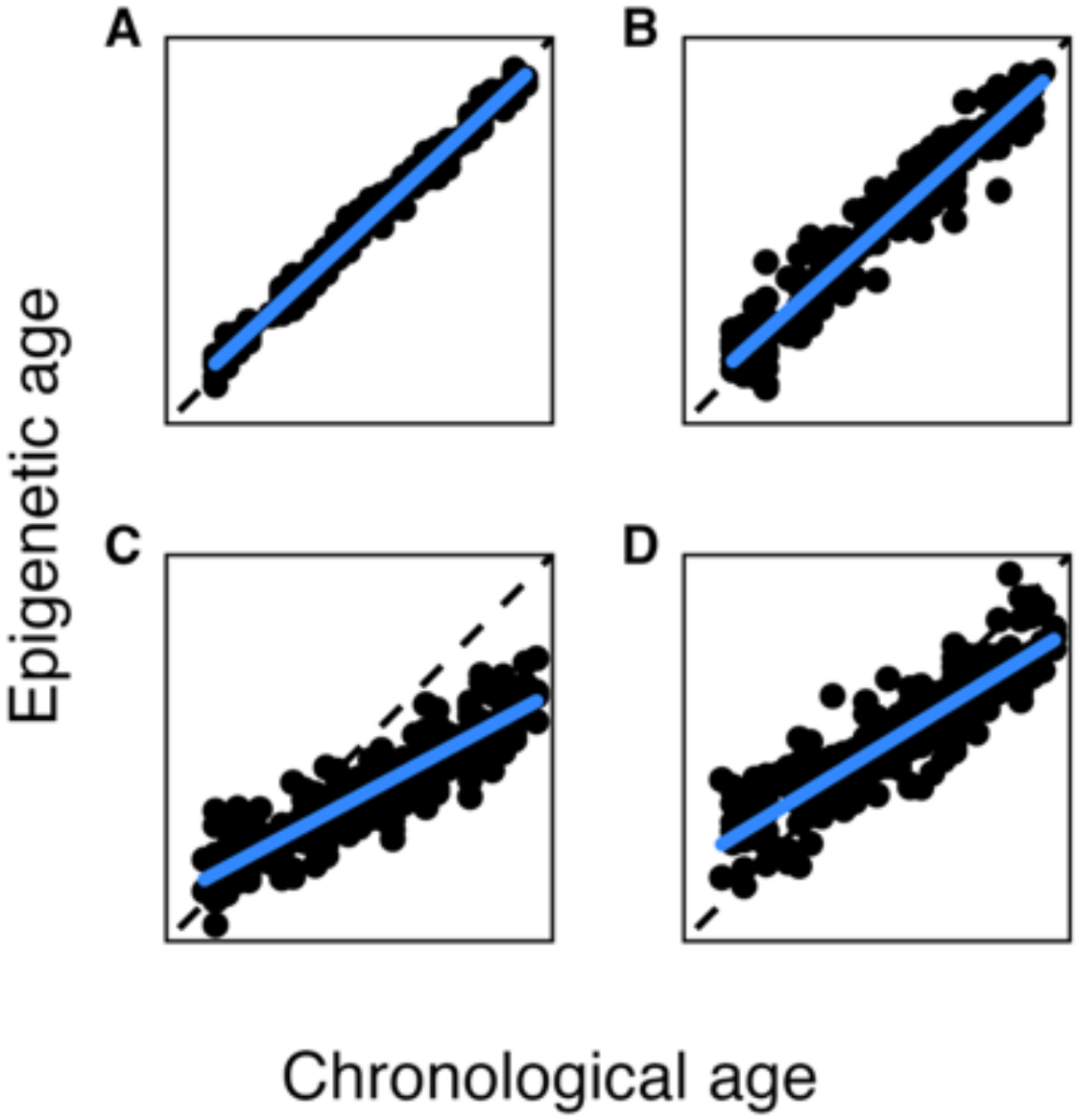


Figure 2 Simulated examples of epigenetic clocks with varying accuracy. The accuracy of epigenetic age estimates can be checked by comparing them to known chronological ages. Black points are observed chronological and predicted epigenetic ages, blue lines are regression lines through the points, and the dotted lines are guides for a 1:1 relationship between chronological and epigenetic age. (A) illustrates a clock with high accuracy. The regression line closely follows the 1:1 line, resulting in low median absolute error (MAE) and high R^2 and correlation between epigenetic and chronological age. Clock (B) is less accurate, with a higher MAE. Clocks (C) and (D) have a similar correlation and R^2, but (D) has a lower MAE, as it better tracks a 1:1 relationship between epigenetic and chronological age.

Section A—Design considerations: sample selection and bias

Sampling challenges in wildlife epigenetic clocks

This section addresses considerations for sample selection when training wildlife epigenetic clocks. In human studies, epigenetic clocks can vary in accuracy when the training set is biased toward one or a few classes of age, sex, tissue, or other factors that influence DNAm (Hannum et al., 2013; McEwen et al., 2020). To mitigate these *class biases* (Box 2 – *Class bias simulation*), human epigenetic clocks are typically trained on large samples that fully represent the classes to which the clocks will later be applied.

In contrast, wildlife sampling is often opportunistic or limited to specific age or sex groups, with genetic relationships frequently unknown and potential constraints on which tissues can be collected. Wildlife studies must also contend with *age biases*, which arise when sampling is restricted to one or a few age classes, followed by the clock's application to age ranges not represented in the training data (Box 3 – *Age bias simulation*). This problem is exacerbated by nonlinear changes in DNAm with age (Horvath & Raj, 2018). Additionally, inaccuracies in wildlife clocks could stem from having to estimate the chronological ages of samples (e.g., (Mayne et al., 2023; Thompson et al., 2017; Box 3 – *Age error simulation*), a challenge less important in human studies where precise ages are usually known.

Therefore, wildlife studies must recognize the potential limitations of epigenetic clocks trained on class-biased samples, avoid critical biases, and anticipate the future applications of clocks to guide sampling. In the following subsections, we discuss the evidence for and potential causes of reduced clock accuracy due to class and age biases, accompanied by clocks fit with our polar bear dataset, to demonstrate how these biases might affect clock accuracy in wildlife.

Class biases—Genetic population differences in aging

One of the major class-specific differences in DNAm with age occurs between populations. Population differences arise due to a combination of environmental factors, which account for some between-human population variation in the relationship between DNAm and age, and genetic differences, which play a more significant role (Carja et al., 2017; Fraser et al., 2012). For example, studies on human twins have shown that genetic differences between individuals can explain up to half of the variation in their epigenetic aging rates (Jylhävä et al., 2019).

Whether population differences represent an important class bias for wildlife epigenetic clocks is uncertain. The Horvath mammalian methylation array, used to measure DNAm in all mammals, differs from the human array by including only sequences conserved across a large number of mammalian species. This design should minimize bias caused by genetic variation among populations of the same species (Arneson et al., 2022). However, substantial genomic alignment differences to the mammalian methylation array still exist between species (Lu et al., 2023; Zoller & Horvath, 2024), suggesting that genetic variation at some sites on the array could also subtly affect clock accuracy between populations.

Testing the influence of genetic population structure on epigenetic clock performance in polar bears, we found minimal impact (Figure 3A), suggesting age-specific DNAm patterns in this species are largely unaffected by genetic differences. Despite this, genetic influences on DNAm might vary across species, and many recent species-specific clocks trained using samples from single populations have not yet been tested on other populations. This testing is particularly

important because genetic structure often correlates with spatial variation, a key factor when examining relationships between epigenetic aging rates and environmental variation.

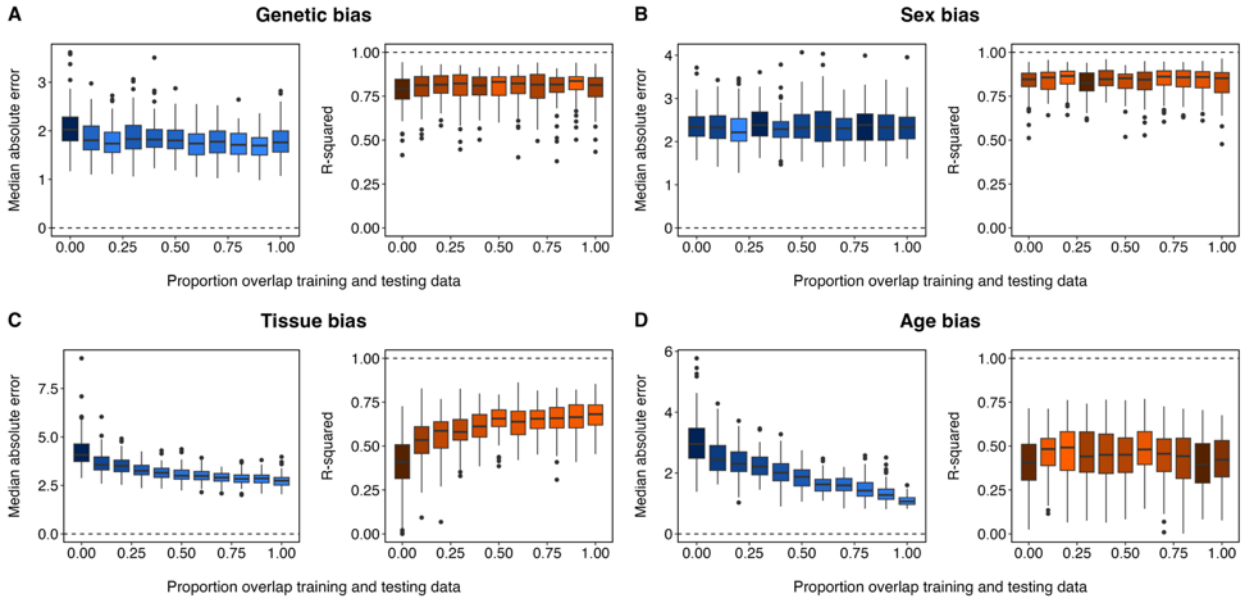


Figure 3 Predictive accuracy of polar bear epigenetic clocks trained with varying levels of class overlap with the testing data, measured by median absolute error (MAE, blue) of epigenetic age relative to chronological age and the R-squared (r, orange) of the linear relationship between epigenetic and chronological age. Brighter orange and blue boxes indicate more accurate clocks and darker-shaded boxes are less accurate. For each overlap proportion, we fit 100 clocks with new training and testing samples, and the resulting accuracy metrics are displayed as boxplots showing the median, interquartile range, and outliers. (A) predicts epigenetic age in 30 samples from two western-Arctic subpopulations (Southern and Northern Beaufort) using 75 samples from the same populations and a genetically distinct central-Arctic subpopulation (Western Hudson Bay), with overlap proportions ranging from genetically identical (0) to entirely distinct (1). (B) predicts epigenetic age in 30 male samples using 60 samples ranging from entirely female (overlap = 0) to entirely male (overlap = 1), with equal numbers from each subpopulation. (C) predicts epigenetic age in 75 muscle samples from seven subpopulations across the Canadian Arctic, using 100 samples ranging from only muscle (overlap = 1) to blood and skin (overlap = 0). (D) predicts epigenetic age in 30 mature bears (> 5 years) using 45 samples ranging from entirely mature (overlap = 1) to entirely immature (< 5 years), with equal representation from each subpopulation. The plots indicate that clock performance is most affected by biased tissue types and age groups in the training data and that these biases have a greater impact on the deviation of epigenetic age from chronological age than on the linear relationship between epigenetic and chronological age.

Class biases—Sex-specific DNA methylation

Female humans and other mammals live longer than males (Lemaître et al., 2020), which raised early concerns about possible sex-based differences in epigenetic aging that could affect the accuracy of epigenetic clocks. The majority of sex-specific DNAm patterns occur on the sex chromosomes, though some autosomes also show sex-specific effects (Gatev et al., 2021; McCartney et al., 2020). Some evidence for these intrinsic sex-associated aging mechanisms

comes from comparing aging rates and DNA methylation between sterilized and unsterilized animals. In these studies, androgen-sensitive CpG sites in sterilized animals show lower DNA methylation (Sugrue et al., 2021) and sterilized individuals also age faster epigenetically (Stubbs et al., 2017). In human epigenetic clocks, these biases are well-documented; differences in aging-related phenotypes between males and females align with distinct DNA methylation patterns (Grant et al., 2022). To prevent these differences from impacting accuracy, clocks designed for humans and model organisms often exclude markers present on the sex chromosomes (Hannum et al., 2013; Stubbs et al., 2017).

Assessing the impact of sex ratio in the training dataset on epigenetic clock accuracy using our polar bear dataset, we found no impact (Figure 3B). This suggests that in polar bears, few age-predictive sites selected by the clock model are likely sex-specific. However, sex-based differences in DNAm could be important in other species, as they have been documented elsewhere (e.g., Czajka et al., 2024; Prado et al., 2021; Robeck et al., 2021).

Class biases—Tissue-specific DNA methylation

Tissue-specific aging rates present a known challenge for human epigenetic clocks (Horvath & Raj, 2018; Porter et al., 2021). Research suggests that epigenetic clocks trained on specific tissues tend to be highly accurate for that tissue but less effective for predicting age in other tissues (Porter et al., 2021), as different tissues capture slightly different aspects of aging (Gibson et al., 2019; McEwen et al., 2020). For example, the human PedBE clock, trained using buccal epithelial cells from children and adolescents aged 0–20, remains one of the most accurate human clocks even when applied to older age groups (McEwen et al., 2020). However, its accuracy falls drastically when used to age non-epithelial tissues (Ibid). In contrast, multi-tissue clocks can be less accurate but more versatile across different tissues, as they tend to select sites that are not tissue-specific (Horvath, 2013; Porter et al., 2021).

We found tissue bias in the training set of our polar bear epigenetic clocks greatly impacted clock accuracy (Figure 3C). As for others (Robeck et al., 2021; Stubbs et al., 2017), tissue type is an important consideration for this species. At least in some species, the elastic net regression algorithm seems to favour DNAm patterns exclusive to the dominant tissue type in the sample (Robeck et al., 2021).

Age bias and age estimation bias

Many human clocks are less accurate for young individuals because DNAm changes occur several-fold faster in early than later life (Alisch et al., 2012). Rapid changes in DNAm during early life and adolescence are linked to genes related to growth and development that are less active in adulthood (McEwen et al., 2020). Additionally, changes in cell composition of tissues with age can also influence DNAm, as DNAm differs across cell types (Chen et al., 2016; Shireby et al., 2020). When trained on samples with a narrow age range, particularly those from older individuals, clock accuracy declines (Simpkin et al., 2016). For example, the Hannum clock, one of the earliest human clocks, was trained on samples from adults 19 years and older, making it less accurate for adolescents compared to the Horvath clock, which was trained on samples from newborns to older adults (Simpkin et al., 2016). Accounting for non-linear changes

in DNA methylation with age is thought to improve the accuracy of epigenetic clocks (Bernabeu et al., 2023; Haftorn et al., 2023).

Age bias is a critical consideration in designing wildlife epigenetic clocks, where sampling methods often favour some age classes over others (Bisi et al., 2011; Camacho et al., 2017; Smith et al., 1995). Thus, some ages are bound to be underrepresented or absent in many wildlife epigenetic clocks. Using our polar bear data, we found that the median absolute error between actual and predicted age was lowest when we trained clocks with more samples from mature individuals and used them to predict sample ages from other mature individuals (Figure 3D), suggesting epigenetic aging rates are likely faster in young polar bears as they are in humans (Alisch et al., 2012). However, when we simulated these non-linear epigenetic aging patterns, we found that training clocks with samples from older individuals and using them to predict the ages of younger samples resulted in an even more dramatic loss of accuracy (Box 3).

Moreover, unlike in human studies where chronological ages are typically known, wildlife researchers must often estimate the ages of their samples (e.g., Thompson et al., 2017), introducing further error (Mayne et al., 2023). Traditional methods for estimating wildlife age rely on body size or changes in the chemical and structural composition of teeth, eyes, baleen, ear plugs, and other features as animals age (reviewed in Morris, 1972). However, these methods can be inaccurate, leading to either over- or underestimation of epigenetic age (Box 3). For example, the accumulation of abnormal proteins in eye lenses is a standard aging method for bowhead whales (*Balaena mysticetus*). This method's low accuracy, in addition to the long lifespan of this species, may explain the poor accuracy of the pan-mammalian clock in this species (Lu et al., 2023).

Sampling recommendations for wildlife epigenetic clocks

Based on our simulations, analyses, and review of existing epigenetic clocks, their accuracy and reliability will be maximized by addressing key sources of bias and sampling either broadly or narrowly depending on the clock's intended use. We recommend the following approaches to sampling:

Minimize tissue and age biases. To ensure accuracy, we recommend even sampling across ages—particularly "prime" ages—and either focusing on a single tissue type for clocks designed for single tissues or sampling evenly across multiple tissues for broader applications. Our polar bear analysis found tissue and age biases most influence clock performance (Figure 3), consistent with human studies (Porter et al., 2021). The most accuracy is lost when training samples are skewed toward individuals older than the clock's target population (Box 3). However, accuracy improves greatly when younger samples are included in the training sample, even if the youngest and oldest individuals are not included (Box 3).

Despite not being important for polar bears, other class biases, such as population structure and sex differences in DNA methylation, can also influence clock performance (Fraser et al., 2012; Grant et al., 2022). If unavoidable, some of these factors can be mitigated using the quality control and pre-processing methods discussed in the following sections B and C. Differences in epigenetic aging rates due to genetic ancestry are particularly relevant for clocks

designed to assess epigenetic age acceleration across environments, which could be confounded with genetic variation across environments.

Tailor sampling to intended clock applications. Clocks trained on a single class—i.e., a single tissue type, sex, or age range—are likely to be most accurate, as they can identify both class-specific DNAm patterns and those generally related to aging. Our analyses indicate that the most accurate clocks are trained on samples closely matching the class characteristics of the test samples (Figure 3; Boxes 2 & 3). Even simply avoiding highly skewed distributions of sample ages in the training data was effective; our simulations showed that clocks trained on "prime-aged" individuals still generalized well to old and young individuals (Box 3). We recommend sampling from narrow age and class ranges matched to the test population for estimating unknown ages, where a high degree of accuracy is critical.

Conversely, sampling breadth is important for assessing epigenetic age acceleration, where class differences in aging rates—particularly between populations—could be mistaken for the effects of environmental stressors on epigenetic aging rates. For example, a clock trained with samples from a single population might predict faster aging in a different population either because of genetic differences in age-associated sites or exposure to distinct stressors. Drawing training samples from both populations should control for the genetic differences. Our simulations suggest that even a small proportion of samples from each class represented in the training sample can improve the clock's predictions across classes (Box 2).

Anticipate population dynamics and sampling constraints. If future samples will consistently come from the same tissues, age ranges, populations, and sexes, we recommend training the clock with the samples limited to those classes for maximum accuracy. All our analyses indicated these narrowly focused clocks were the most accurate, and other wildlife clock studies have made similar observations (Robeck et al., 2023). However, training on a broader sample range will better capture general age-related changes in DNA methylation and mitigate future biases from age- and class-specific sites if population demography or sampling methods could change. Using data from long-term research projects to examine past population dynamics will help anticipate these changes to ensure clocks remain robust to future demographic and genetic shifts.

Section B—Quality checks and data organization

In the microarray approach, DNA is extracted from tissue samples, and the bisulfite-converted DNA is hybridized to the array, stained, and imaged. The resulting probe images are processed to generate individual *beta values*, quantifying the proportion of methylation at each CpG site. These beta values are then normalized to correct for background fluorescence, a component of the technical variation (Triche et al., 2013). The normalized beta values are used as input for the elastic net regression model that constitutes the epigenetic clock. R packages—most popularly *SeSAMe* (Zhou et al., 2018) and *minfi* (Aryee et al., 2014)—provide functions for converting the images to raw DNA methylation data and normalizing them into beta values. Newediuk et al. (2024) is linked to a well-structured GitHub project with detailed R code covering the entire epigenetic clock workflow (Figure 1). A tailored R package, *MammalMethylClock*, also provides detailed workflows and functions for processing data from the Horvath Mammal Array into normalized betas (Zoller & Horvath, 2024).

Several quality-control checks are recommended to minimize technical variation and potential age or sample class biases. First, randomizing the positions of DNA samples on the array before hybridization can avoid misinterpreting positional effects as biologically meaningful differences in beta values. This randomization aims to physically distribute samples from the same classes across the array. After imaging, quality-control metrics can be assessed on raw probe data using standard pipelines in R (see *minfi* and *SeSAMe*). Hierarchical clustering on normalized beta values can also help visualize potential biases related to sample classes. Finally, if multiple arrays are used within a single study, possible batch effects can be assessed and corrected with the *sva* package (Leek et al., 2012) in R.

Section C—Design considerations: Data pre-processing methods

Overview of pre-processing methods and wildlife clock considerations

After preparing the raw data with normalization, optional batch correction, and other quality control steps, clock performance can still be improved with additional pre-processing steps before building the epigenetic clock. In this section, we discuss three pre-processing methods which improve performance by reducing the dimensionality of the data used to train the clock: genomic alignment, supervised feature selection, and unsupervised feature selection. The examples we provide are specific to beta values obtained from the Horvath Mammal Array, but the same principles of dimensionality reduction are germane to any high-dimensional DNAm data.

Pre-processing improves accuracy because DNAm and other high-throughput data are high-dimensional. This means that the number of features—in this instance, the proportion of methylation at CpG sites—is generally larger than the number of individuals from which the features are sampled. DNAm changes at many CpG sites are uninformative concerning age, potentially biasing clocks by introducing unnecessary complexity. The regularization algorithms used to fit epigenetic clocks reduce some of this complexity by imposing penalties on the number and importance of predictors—CpG sites—in the model (Kuhn & Johnson, 2013). However, feature selection, where the number of features is streamlined before model-fitting, is still common in machine-learning applications that use regularization. It reduces overfitting, improving predictive accuracy in new data (Theng & Bhoyar, 2024).

Feature selection is likely most beneficial for wildlife epigenetic clocks with small sample sizes. Unlike human epigenetic clocks, often designed using hundreds or even thousands of samples (Fransquet et al., 2019), wildlife clocks often rely on datasets with no more than a few dozen samples (e.g., Czajka et al., 2024; Thompson et al., 2017). This creates an inflated feature-to-sample ratio, making dimensionality reduction even more critical.

However, a key consideration when incorporating feature selection into epigenetic clock workflows is balancing model simplification with preserving predictive information. While feature selection helps to reduce overfitting, it also decreases the number of CpG sites available for epigenetic clock development, potentially excluding important predictive sites if the feature selection is too strict. The importance of retaining predictive sites is evident from studies showing that epigenetic clocks trained with progressively fewer CpG sites can still predict age but with substantially lower accuracy compared to clocks using dozens or hundreds of sites (Haftorn et al., 2023; Li et al., 2022). To find the best balance between minimizing the exclusion

of important predictive sites and reducing bias from uninformative ones, we applied two pre-processing approaches—genomic alignment and supervised feature selection—to the polar bear data and assessed their impact on epigenetic clock performance.

Pre-processing methods—Genomic alignment

An initial approach to reducing the number of features is to align the genome of the study species to the mammalian array before fitting a clock. The Horvath Mammal Array was designed for all eutherian mammals, and while at least half of the sites included on the array are conserved among 115 species on which it was tested (Arneson et al., 2022), differences in alignments are possible. Genomic alignment, regarded as a best practice for applying the universal clock for mammals and developing species-specific epigenetic clocks (Zoller & Horvath, 2024), reduces feature complexity by retaining only CpG sites that align to the species of interest's genome. Alignment has become a routine part of creating species-specific epigenetic clocks (e.g., Parsons et al., 2023; Raj et al., 2021; Thompson et al., 2017; Wilkinson et al., 2021).

We found a very modest improvement in accuracy when we removed 3,818 CpG sites that did not align to the polar bear genome (Figure 4). While genome alignment might be more beneficial for some species with fewer sites aligned to the array—resulting in the removal of more non-aligned sites— this approach is also limited because it is only possible for species with genome assemblies available.

Pre-processing methods—Supervised and unsupervised feature selection

Feature selection methods reduce complexity based on relationships among CpG sites. First, supervised or semi-supervised filtering methods select features according to their relationships with explicitly selected variables (Kuhn & Johnson, 2013). In epigenetic clocks, the target variable is often age; CpG sites are retained for significant relationships with age (e.g., Li et al., 2022; Zhuang et al., 2012). It is also possible to select features using other target variables. For example, CpG sites can also be excluded for class biases in their DNA methylation-age relationships (e.g., sex—Newediuk et al., 2024). Class bias can be detected with linear models predicting DNA methylation using age and common class-biased variables like sex and tissue type (Box 4).

In unsupervised filtering methods, the outcome variable—e.g., age—is ignored while features are retained or excluded according to unrelated patterns in the data (Kuhn & Johnson, 2013). In variance-filtering, for example, CpG sites with the most variation in methylation are retained because sites with low variation are less likely to discriminate among ages (Higgins-Chen et al., 2022; Sarac et al., 2017; Zhuang et al., 2012). Sites with signals that cluster with other sites are also targets for unsupervised filtering. These sites tend to be more reliable predictors, and retaining them results in accurate and stable clocks (Higgins-Chen et al., 2022). Related sites can be identified and retained with approaches like k-means clustering (Sarac et al., 2017), or clocks can be trained directly on the principal components of multicollinear CpG sites identified with principal components analysis (Higgins-Chen et al., 2022).

We found removing biased CpG sites from our polar bear clocks through supervised feature selection improved accuracy relative to genome alignment only (Figure 4). This is likely because combining the alignment with feature selection eliminated substantially more sites. However, eliminating too many sites compromised accuracy. For example, removing 35,387

sex-specific sites and those without a strong relationship with age left only 2,105 sites to create the clock, reducing the median absolute error (MAE) compared to clocks without feature selection (Figure 4B). However, removing tissue-biased sites eliminated 37,448 of the initial 37,492 CpG sites on the array, causing a sharp decline in R-squared (Figure 4A). These findings are consistent with our feature selection simulations, which showed that removing class-biased sites lowered clock MAE and maintained a high R-squared until the number of removed sites reached a threshold, beyond which accuracy declined sharply (Box 4). Building accurate wildlife clocks will require balancing feature selection while retaining enough sites to model age-related changes in DNA methylation.

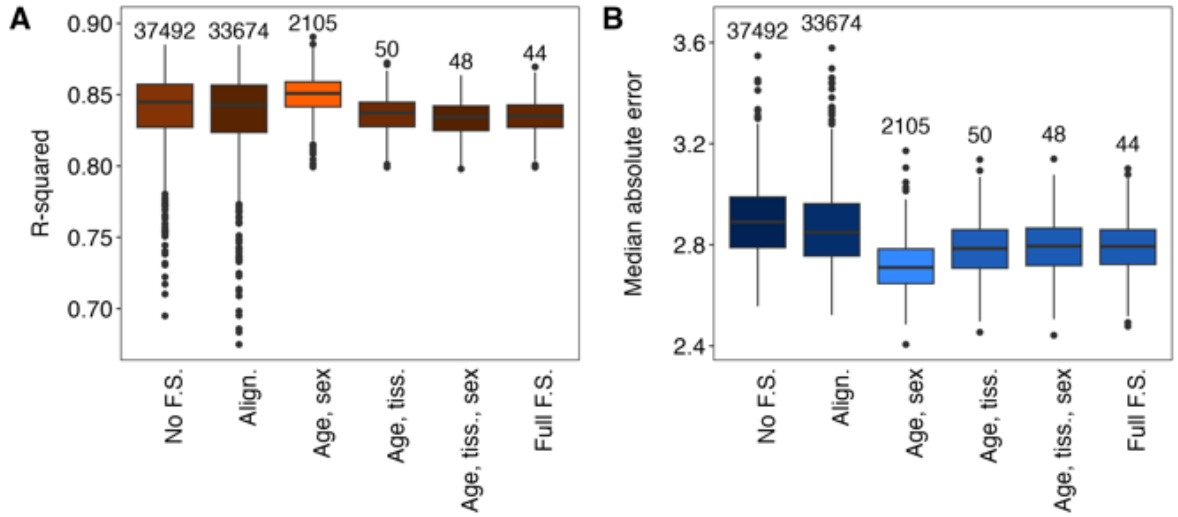


Figure 4. Accuracy, evaluated by R-squared (A) and median absolute error (B), compared between clocks fit with different feature pre-selection approaches using polar bear methylation data. Each box and whisker represents a different feature selection approach. From left to right, these approaches include no feature selection (No F.S.), sites removed if they did not align to the polar bear genome (Align.), sites removed if they lacked a significant relationship with age in both sexes (Age, sex), sites removed if they lacked a relationship with age in blood, skin, and muscle tissues (Age, tiss.), sites removed if they lacked a significant relationship with age in all tissues and both sexes (Age, tiss., sex), and sites removed if they lacked a relationship with age in all tissues and both sexes and did not align to the polar bear genome (Full F.S.). In each approach, we fit 500 clocks by selecting 319 individuals for training sampled evenly across subpopulations, ages 0–30, sexes, and all tissue types, then applied to predict the ages of the remaining 250 individuals. Numbers above the boxes and whiskers denote the number of initial sites retained for fitting the clock.

Pre-processing recommendations for wildlife epigenetic clocks

Our simulations and analyses highlight a fundamental consideration in building epigenetic clocks: reducing feature complexity improves accuracy, but enough features must be retained to ensure the clock is accurate. This balance is important for wildlife epigenetic clocks, where small sample sizes result in a high feature-to-sample ratio. The loss of accuracy from removing too

many sites is especially important for clocks designed for age estimation, where accuracy is critical. Based on our findings, we recommend the following pre-processing steps to enhance the accuracy of wildlife epigenetic clocks:

Align the species genome to the Horvath Mammal Array. While this approach yielded only modest improvements in accuracy for polar bears, it also did not remove enough CpG sites to reduce clock accuracy. Genomic alignment may be more impactful for other species, particularly when it eliminates more misaligned CpG sites than we could exclude with our polar bear alignment. Genomic alignment could exclude as many as 20,000 sites in some species (Arneson et al., 2022), making it an effective method for reducing dimensionality. The use of this approach depends on the availability of a reference genome for the species of interest. Fortunately, genome alignment has become a standard approach for designing species-specific epigenetic clocks (Zoller & Horvath, 2024), and many species reference genome alignments are available on the Mammalian Methylation Consortium's GitHub page at https://github.com/shorvath/MammalianMethylationConsortium/.

Remove only the minimum number of sites necessary to improve accuracy while retaining important predictive sites. To optimize accuracy—particularly for epigenetic clocks used for age estimation—we recommend identifying and closely approaching the threshold where further site removal starts to impact accuracy. This threshold is likely species- and dataset-specific. In our simulations, accuracy first declined when we removed approximately 75% of CpG sites. For polar bear clocks, the threshold occurred somewhere between the removal of 35,387 and 37,442 CpG sites—representing 94.4% and 99.9% of sites on the Horvath Mammal Array.

We recommend determining this threshold through iterative feature selection, targeting different combinations of potentially biased classes to find the combination that yields the highest accuracy. For example, if the samples include four tissue types and two sexes, eight linear regression models could be fit, one for each combination of sex and tissue type, to identify the CpG sites significantly related to age in each group. The next step will be constructing a series of clocks, starting with the sites shared across all groups and working gradually up to larger sets of sites shared by subsets of groups (e.g., sites common to only two tissue types and both sexes). In our polar bear analysis, clock performance improved most when we retained only sites strongly related to age for both sexes. In contrast, too few sites were associated with age in all groups (both tissue types and both sexes), which resulted in the removal of too many sites and reduced clock accuracy (Figure 4).

Section D—Design considerations: Validation approaches

This section discusses approaches for assessing clock accuracy, a challenge in wildlife studies where sample sizes are often small. Clock accuracy is evaluated using a validation dataset, the gold standard being a distinct hold-out set independent of the training data used to fit the model. This method, widely used in human epigenetic clock studies with typically large sample sizes (e.g., Hannum et al., 2013; Horvath, 2013; McEwen et al., 2020), leaves enough samples to create an accurate clock while avoiding inflated accuracy estimates caused by overfitting the training dataset (Hastie et al., 2009). However, in wildlife studies, small samples make it difficult to reserve a substantial hold-out set for validation without compromising the data available for

clock training. Validation strategies for wildlife studies must, therefore, maximize true accuracy while avoiding its overestimation. This section discusses the benefits of different validation approaches, which we compare using our polar bear data.

There are three primary approaches for selecting a validation set to estimate the accuracy of predictive models, including epigenetic clocks. In addition to setting aside a distinct hold-out set and using the same dataset for both training and validation, validation can be performed on a series of smaller subsets of the training data, with errors averaged across subsets sampled from the training data—a method known as cross-validation.

Cross-validation approaches differ by the size of equally sized subsets or folds, *k*. In *k-fold cross-validation*, a fold of size *k* is used for testing, while the remaining *k-1* folds are used for testing. *Leave-one-out (LOO) cross-validation* is a special case of *k-fold cross-validation*, where each fold contains only a single observation. In the context of epigenetic clock models, the single observation can also be a single grouping of individuals. For example, the universal clock for mammals was validated using leave-one-species-out cross-validation, where the clock, trained on all but a single species, was tested on each excluded species in turn (Lu et al., 2023). In species-specific clocks, the group could be population, sex, or tissue, with the remaining groups used for testing.

The small sample sizes typical of wildlife studies often make cross-validation the only practical option for epigenetic clock validation. Indeed, most wildlife clocks published since the release of the Horvath40Mammal array—including the universal clock for mammals—were validated using LOO cross-validation (e.g., Parsons et al., 2023; Prado et al., 2021; Raj et al., 2021; Robeck et al., 2021). LOO cross-validation best estimates true test error because it uses nearly all the data (n – 1) for training while iterating systematically through the testing data (James et al., 2013).

Using our polar bear dataset, we evaluated the accuracy of epigenetic clocks estimated through LOO cross-validation and compared it to evaluation on an independent hold-out set. From 661 polar bear samples, we randomly selected a subset of 500. Within this subset, we sampled 250 for training the clock, retaining the remaining samples as a hold-out set and also validating the clock by performing LOO cross-validation on the training samples. We repeated this process 100 times.

We found no difference in either the R-squared (Figure 5A) or median absolute error (Figure 5B) between the two approaches when evaluating accuracy (Figure 5B), indicating that LOO cross-validation reflects true accuracy. Importantly, our findings suggest that developers of wildlife clocks may be justified in using their full set of available samples to maximize true clock accuracy while reliably assessing the clock's predictive performance when applied to new samples.

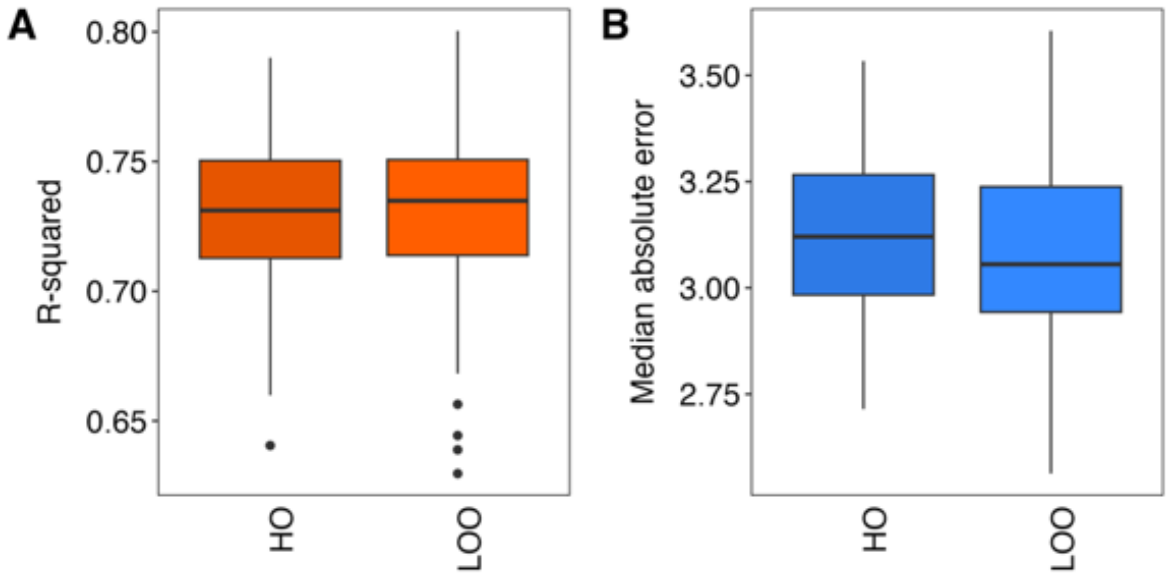


Figure 5 Accuracy, evaluated by R-squared (A) and median absolute error (B), compared between polar bear clocks validated using leave-one-out cross-validation (LOO) versus an independent hold-out set (HO). Bright orange and blue boxes represent higher accuracy, while darker colours represent lower accuracy.

Validation recommendations for wildlife epigenetic clocks

Use all samples available for training wildlife clocks. To improve accuracy, we recommend using all available samples when training species-specific epigenetic clocks, followed by leave-one-out cross-validation. Leave-one-out cross-validation is generally thought to approximate true model accuracy in machine learning applications (Hastie et al., 2009), and our results suggest this holds for epigenetic clocks. Moreover, retaining more samples for training should lead to more accurate cross-validated clocks, providing more reliable age estimates and measures of cumulative lifetime stress.

Conduct independent validations of existing clocks. Although leave-one-out cross-validation shows promise for approximating true accuracy, our analysis was limited to a single species. Many other species-specific clocks have not been tested on populations beyond those used to train them. To build confidence in the reliability of leave-one-out cross-validation as a validation approach, we recommend that follow-up studies independently validate existing clocks.

Conclusions

Epigenetic clocks have great potential to fill critical data gaps in wildlife conservation and management. However, challenges associated with collecting wildlife DNA samples, which limit sample sizes and can impact the accuracy of epigenetic clocks, have been largely unaddressed. The absence of a standardized workflow for developing wildlife epigenetic clocks also hampers their widespread use. To address these issues and encourage their development, we provided a detailed workflow for developing epigenetic clocks geared toward wildlife research (Figure 1), including sample selection, quality control, feature pre-selection, and validation. We demonstrated our recommended workflow using simulations and data from polar bears across the

Canadian Arctic, equipping practitioners with the tools and knowledge needed to design and develop accurate epigenetic clocks.

Through our polar bear analyses and simulations, we showed that thoughtful sampling, feature selection, and validation can ensure accurate epigenetic clocks for wildlife, even with small sample sizes. Among our most important recommendations is to plan ahead of clock development. Identifying target populations, tissues, age ranges, and sexes in advance will enable the design of wildlife clocks tailored to specific applications. Narrowly focused clocks are often the most accurate, except when applied to a broader range of samples than those on which they were trained. Clock accuracy can also be enhanced by reducing the initial number of sites used for training and using all available samples for training. While maximizing accuracy is particularly important for clocks used to estimate unknown ages, perfectly accurate clocks cannot measure epigenetic aging rates, emphasizing the need to clarify the clock's intended purpose from the outset. With planning, epigenetic clocks can provide highly accurate age data for wildlife conservation and management.

Acknowledgments

Financial and logistic support of field research was provided by Churchill Northern Studies Centre, Environment and Climate Change Canada, Isdell Family Foundation, Parks Canada, Polar Bears International, Wildlife Media, World Wildlife Fund, and an NSERC Discovery Grant awarded to CJG. We thank Steven Baryluk and Faye d'Eon Eggertson for contributing data from polar bear programs in the Northwest Territories and Ian Stirling, Dennis Andriashek and David McGeachy for their invaluable contributions to the western Hudson Bay polar bear program. We also thank Elizabeth Patterson for helpful feedback on the manuscript.

Box 1—Polar bear data from across the Canadian Arctic

We compiled an extensive DNA methylation dataset from polar bears across the Canadian Arctic to assess whether sampling biases, data pre-processing, and validation influence wildlife clock performance. Our dataset includes DNA from 10 distinct subpopulations, each with different proportions of blood, skin and muscle tissue, from male and female bears, and representing ages across the typical lifespan of a wild polar bear from age 0 to 30 (Table B1).

Table B1 Overview of polar bear DNA methylation samples from 10 genetically distinct subpopulations across the Canadian Arctic. DNA was extracted from three tissue types: blood (B), skin (S), and muscle, and male (M) and female (F) bears.

Subpopulation	Number of samples	Location	Age range	Tissue proportions	Sex proportions
Southern Beaufort	76	Western Arctic	0–20	B: 0.20; S: 0.80	F: 0.54; M: 0.46
Northern Beaufort	62	Western Arctic	0–24	B: 0.11; S: 0.89	F: 0.55; M: 0.45
Gulf of Boothia	36	Western Arctic	0–20	M: 1.0	F: 0.53; M: 0.47
Lancaster Sound	41	Western Arctic	0–21	M: 1.0	F: 0.46; M: 0.54
Mc'Clintock Channel	35	Western Arctic	0–17	M: 1.0	F: 0.66; M: 0.34
Foxe Basin	40	Central Arctic	0–21	M: 1.0	F: 0.50; M: 0.50
Western Hudson Bay	235	Central Arctic	0–30	B: 0.43; S: 0.57	F: 0.60; M: 0.50
Southern Hudson Bay	47	Central Arctic	0–22	M: 1.0	F: 0.51; M: 0.49
Davis Strait	41	Eastern Arctic	0–20	M: 1.0	F: 0.46; M: 0.54
Baffin Bay	40	Eastern Arctic	0–23	M: 1.0	F: 0.50; M: 0.50

We used the age, tissue, sex, and population structure of the data to assess the impacts of class bias, age bias, and feature selection on clock performance. We trained clocks using varying degrees of overlap (0-100%) between the age ranges, tissues, sexes, and populations in the training and testing sets. We fit these clocks using elastic net regression with the *glmnet* package (Friedman et al., 2010) in R v4.3.1 (R Core Team, 2023) and evaluated their performance based on median age error and R-squared.

For feature selection, we identified which classes—of age, tissue, sex, and population—showed significant differences in DNA methylation patterns. We fit multivariate linear models with the DNA methylation matrix as the response variable and tissue, sex, and population as predictors using the *limma* package (Ritchie et al., 2015). We excluded any CpG sites if DNA methylation differed significantly by class with $p < 0.001$.

Box 2—Class bias

Class biases—where certain categories, such as age, sex, or tissue type, are overrepresented in the data used to train an epigenetic clock—can lower its performance.

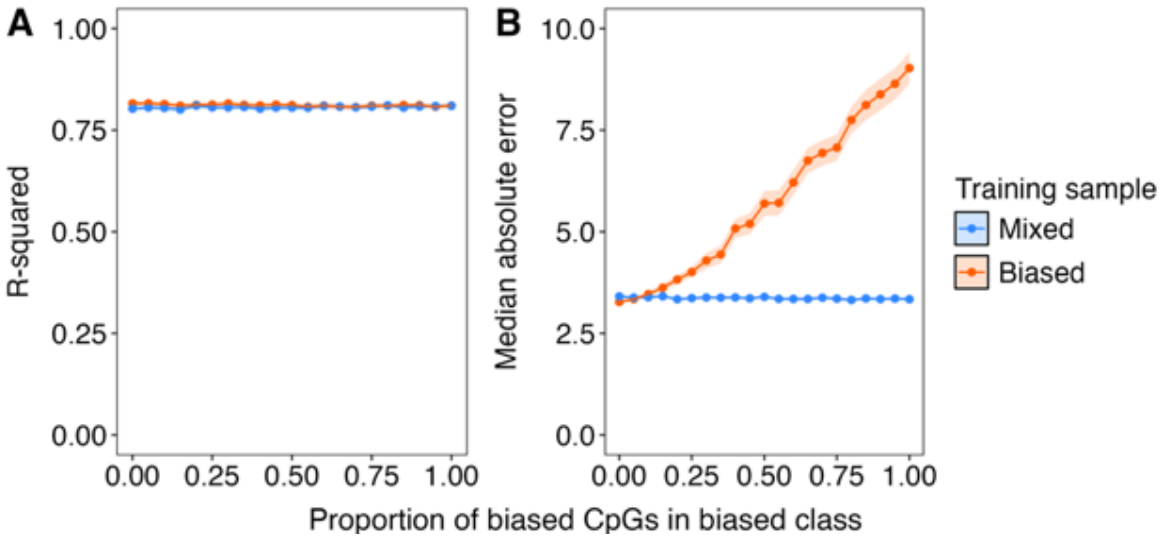
Simulation

To test the importance of class bias for epigenetic clock performance, we simulated DNA methylation data with a class bias. First, we simulated 500 ß values representing 500 age-associated CpG sites, where $\beta_i = xy_i + \varepsilon$. In our simulated data, y_i is a vector of chronological ages from 0 to 30, β_i represents the proportion of methylation at CpG site i, x is the slope of the relationship between y_i and β_i, and ε is normally distributed error (mean = 0, standard deviation = 0.5). We simulated x values for each β_i from a uniform distribution ranging from -0.1 to 0.1.

We then assigned the simulated samples to one of two classes: biased and unbiased. We simulated a weaker association between age and DNA methylation in the biased class by introducing additional error into ε in 5% up to 100% of the CpG sites for the biased class. We trained two clocks: one using a random sample of 150 each from the biased and unbiased class and another using only samples from the biased class. The second clock represents the case where a sampling bias might result in a clock designed for one class being applied to predict age in another. We compared the performance of the two clocks using an independent test set of 150 samples from the unbiased class. We fit the clocks using elastic net regression with the *glmnet* package (Friedman et al., 2010) in R v4.3.1 (R Core Team, 2023).

Conclusion

To ensure accuracy, epigenetic clocks should be trained with all classes of interest. Our results show that class bias does not affect the linear relationship between chronological and epigenetic age (Figure B1.1 A), but it increases the median absolute error (Figure B1.1 B), which grows as the proportion of biased CpG sites increases, suggesting chronological age is either over- or underestimated (Figure 2). The median absolute error is minimized when the training set includes samples from both the biased and unbiased classes, as the procedure can select enough age-related sites to predict age accurately.



690

691 **Figure B1.1** Accuracy of epigenetic clocks trained with two simulated sample types: one without
692 class bias and the other including varying proportions of biased CpG sites (ranging from 0 to 1).
693 The mixed training sample (blue) includes an equal number of samples from both the biased and
694 unbiased classes, while the biased training sample (red) contains only samples from the biased
695 class. The points and ribbons indicate each accuracy metric's mean and 95% confidence intervals
696 in 100 bootstrapped samples of CpG sites at each proportion.

697

Box 3—Age bias and aging error

Training a clock on a narrow age range introduces a form of age bias that limits the clock's performance when applied outside of that range (Simpkin et al., 2016). The problem is thought to stem from more rapid changes in DNA methylation in some periods of life (Alisch et al., 2012), and can be corrected by accounting for the nonlinear relationship between DNA methylation and age (Bernabeu et al., 2023; Haftorn et al., 2023).

Simulation

We simulated non-linear relationships between DNA methylation and age to test the influence of sampling bias on epigenetic clock performance. We simulated 500 ß values, where $\beta_i = {y_i}^x + \varepsilon$. In our simulated data, y_i is a vector of chronological ages from 0 to 30, β_i is the proportion of methylation at CpG site i, x is sampled from a normal distribution N(2, 0.35), and ε is normally distributed error N(0, 0.8).

Using our simulated data, we trained three clocks using 150 age-biased samples and tested them on different age groups. First, we trained a clock on 150 individuals aged 0–15 and tested it on 150 samples aged 16–30 to assess how well clocks trained on younger samples performed on older test sets. We then reversed this by training a clock on individuals aged 16–30 and testing it on younger samples aged 0–15. Finally, we trained a clock on samples aged 5–20 and tested it on a broader range of ages (0–30), simulating a common scenario in wildlife research where "prime age" individuals are oversampled (Camacho et al., 2017; Smith et al., 1995).

Another form of age bias arises when the true ages of samples are unknown. In a second simulation, we incrementally introduced aging error to simulate this form of age bias by adjusting the ages of the simulated samples with an error drawn from a random normal distribution with a mean of 0 and a standard deviation ranging from 1 to 5 years. We fit the clocks using elastic net regression with the *glmnet* package (Friedman et al., 2010) in R v4.3.1 (R Core Team, 2023).

Inaccurate aging also reduces clock accuracy. As we introduced error into sample ages, the median absolute error increased steadily, and the R-squared decreased (Figure B2.2).

Conclusion

To ensure epigenetic clock accuracy, training samples should be aged as accurately as possible while avoiding a bias toward exclusively older or younger individuals. We found that any form of age bias reduced clock accuracy. Median absolute error increased when clocks trained on samples of either older or younger individuals were applied to the opposite age class. Both biased clocks also showed a lower R-squared between chronological and epigenetic ages. Interestingly, the clock trained on prime-age individuals performed similarly to the unbiased clock, with a slightly lower median absolute error but worse R-squared. This suggests that accurate clocks can be made even if sampling regimes cannot capture very old and very young individuals.

We recommend avoiding under-sampling young individuals. Training clocks with samples from older individuals yielded far worse predictions for young individuals than the reverse, with almost triple the median absolute error of the unbiased clock and an R-squared

lower than 0.5 (Figure B2.1). This pattern is strikingly similar to findings from many human clocks (Simpkin et al., 2016), suggesting wildlife studies should be particularly cautious of training epigenetic clocks with samples skewed toward older age classes.

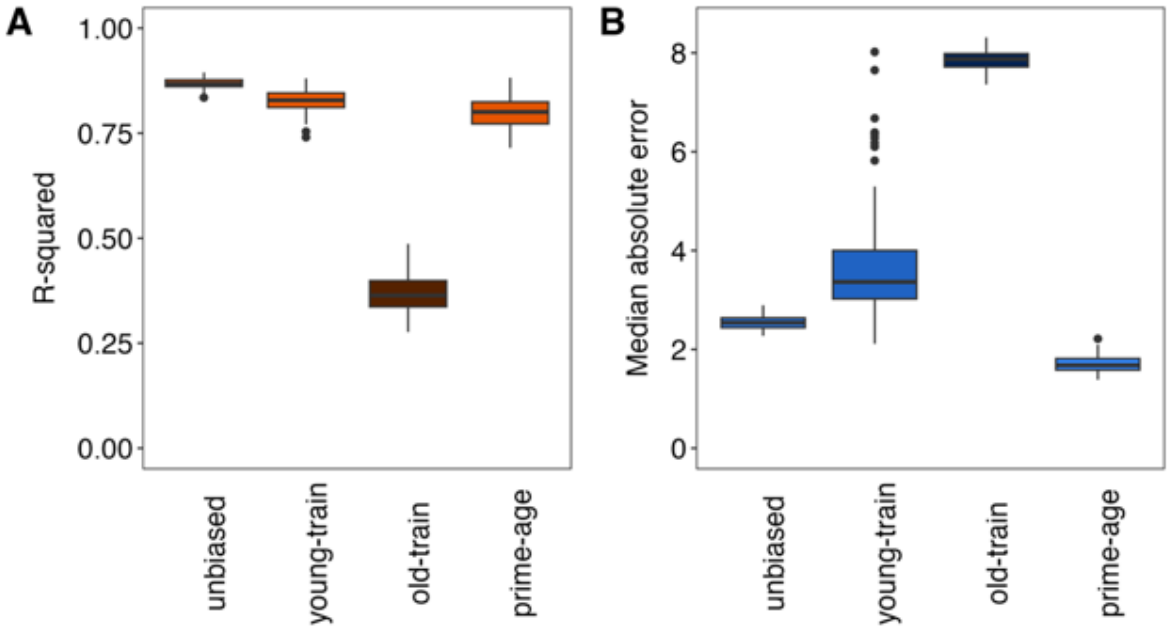


Figure B2.1 Accuracy of epigenetic clocks, evaluated by R-squared (A) and median absolute error (B), trained on simulated age-biased samples and tested on different age groups. From left to right, the clocks are unbiased: trained with the same ages it predicts; young-biased: trained on samples aged 15 years or younger and tested on individuals aged 16–30; old-biased: trained on samples aged 16–30 and tested on individuals under 15 years; and prime-aged: trained on samples aged 5–20 and tested on individuals aged 0–30. The colour gradient indicates accuracy. Brighter orange and blue boxes indicate more accurate clocks, and darker-shaded boxes are less accurate.

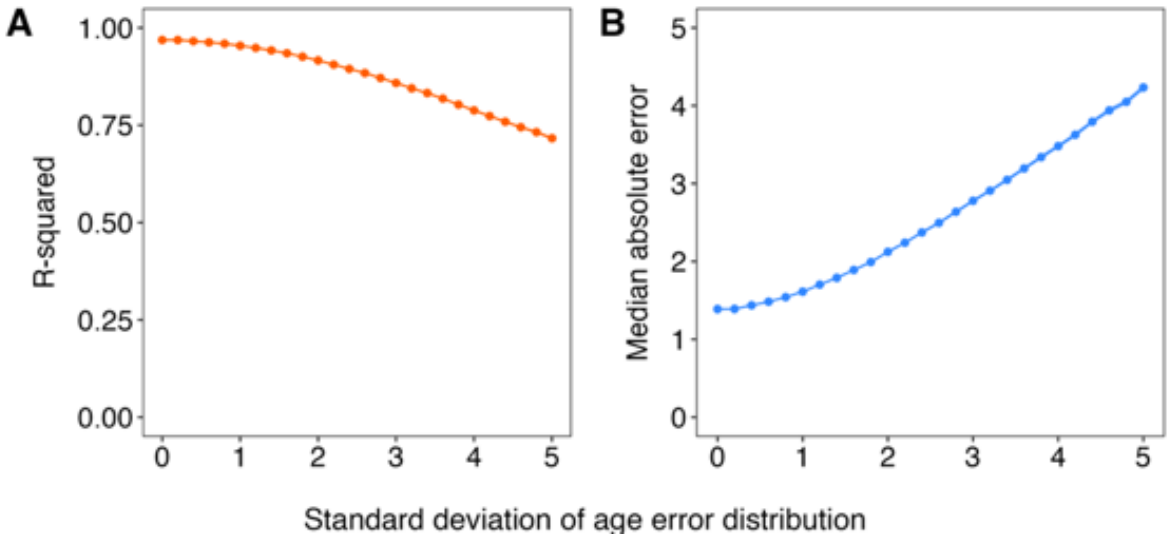


Figure B2.2 Accuracy of age error-biased epigenetic clocks in predicting chronological age, evaluated by R-squared (A) and median absolute error (B). The clocks were trained on simulated data with progressively increasing error (standard deviation) in the training sample ages relative to their true ages. The points and ribbons indicate each accuracy metric's mean and 95% confidence intervals in 100 bootstrapped samples of CpG sites at each proportion.

Box 4—Feature selection

Feature selection enhances predictive model performance by removing features that lack strong associations with the response variable (Theng & Bhoyar, 2024). For epigenetic clocks, supervised feature selection improves accuracy by removing CpG sites with class-specific relationships between DNA methylation and age. However, excessively reducing the initial pool of CpG sites also limits the features available to model relationships with age, which can also reduce model performance (Li et al., 2022).

Simulation

We used our simulated class-biased DNA methylation data, described in Box 2, to test the trade-off between feature selection and retaining biased features. We assessed the performance impact of retaining versus excluding CpG sites with class-specific relationships. In the feature selection scenario, we simulated supervised feature selection by sequentially removing the class-specific CpG sites—from 5% to 95% of the total CpG sites—before fitting the clock. We compared the accuracy of these clocks with those trained using the full set of class-biased CpG sites.

Conclusion

Our simulation demonstrates that feature selection for accurate epigenetic clocks requires balancing the removal of sites lacking any relationship with age and retaining sites important for predicting the relationship. Excluding the class-biased CpG sites with feature selection kept the median absolute error consistently low relative to clocks where the class-biased sites were retained. However, as we removed more CpG sites, the R-squared declined, and the median absolute error increased, indicating that excessively shrinking the initial CpG pool could compromise some aspects of accuracy while improving overall performance (Figure 2). In contrast, while class bias slightly reduced the R^2, the removal of class-biased CpG sites caused an even sharper decline, suggesting excessive feature selection might negatively impact epigenetic clock performance (Figure B3.1).

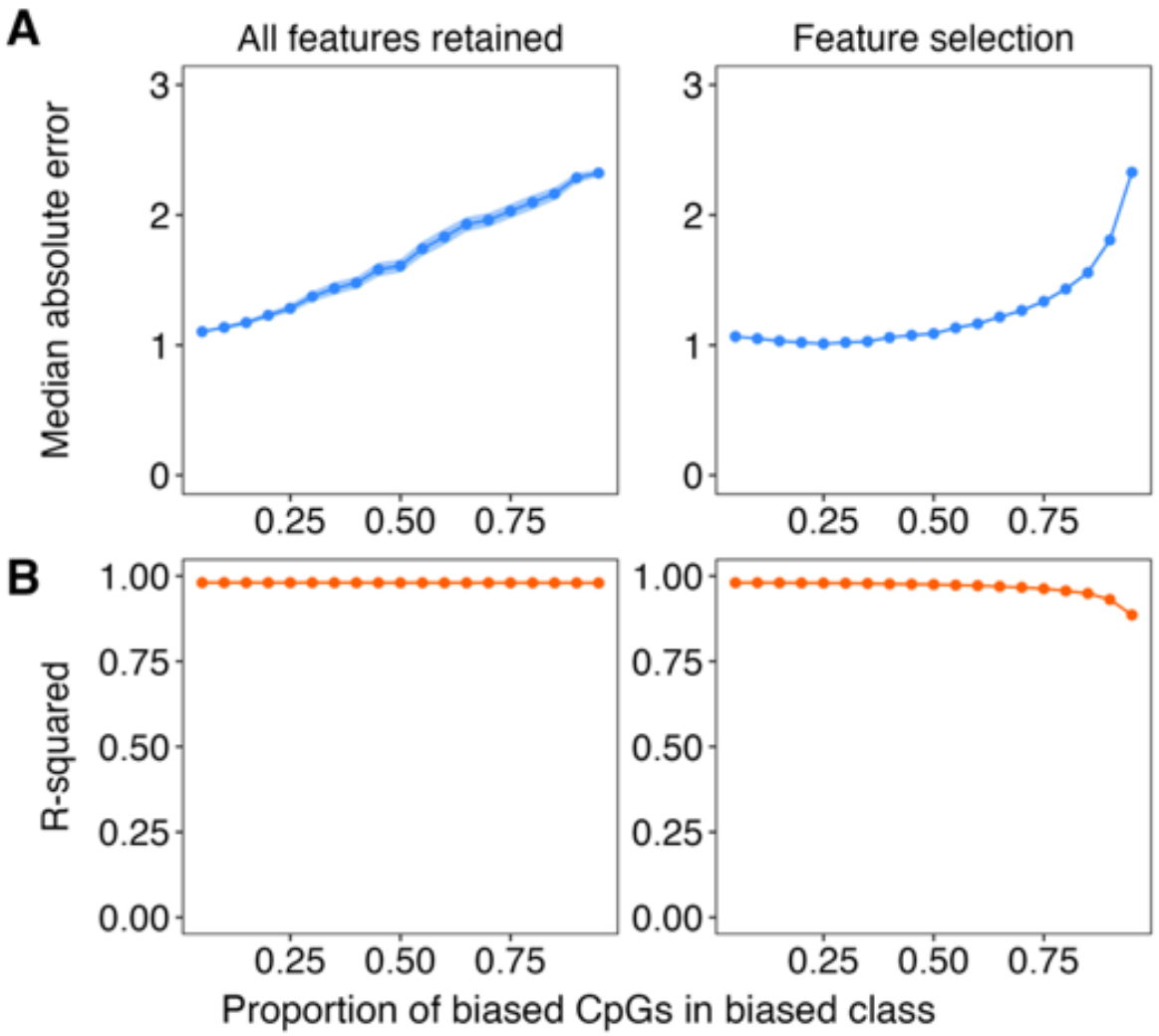


Figure B3.1 The accuracy of clocks fit using simulated data as the proportion of biased CpG sites increases in a set of 500 CpG sites. For each proportion, we fit a clock where we retained the biased CpG sites for training and another where we performed feature selection, removing all biased CpG sites before training. The points and ribbons indicate each accuracy metric's mean and 95% confidence intervals in 100 bootstrapped samples of CpG sites at each proportion.

References

Alisch, R. S., Barwick, B. G., Chopra, P., Myrick, L. K., Satten, G. A., Conneely, K. N., & Warren, S. T. (2012). Age-associated DNA methylation in pediatric populations. *Genome Research*, *22*(4), 623–632. https://doi.org/10.1101/gr.125187.111

Anderson, J. A., Johnston, R. A., Lea, A. J., Campos, F. A., Voyles, T. N., Akinyi, M. Y., Alberts, S. C., Archie, E. A., & Tung, J. (2021). High social status males experience accelerated epigenetic aging in wild baboons. *eLife*, *10*, e66128. https://doi.org/10.7554/eLife.66128

Arneson, A., Haghani, A., Thompson, M. J., Pellegrini, M., Kwon, S. B., Vu, H., Maciejewski, E., Yao, M., Li, C. Z., Lu, A. T., Morselli, M., Rubbi, L., Barnes, B., Hansen, K. D., Zhou, W., Breeze, C. E., Ernst, J., & Horvath, S. (2022). A mammalian methylation array for profiling methylation levels at conserved sequences. *Nature Communications*, *13*(1), 783. https://doi.org/10.1038/s41467-022-28355-z

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, *30*(10), 1363–1369. https://doi.org/10.1093/bioinformatics/btu049

Bernabeu, E., McCartney, D. L., Gadd, D. A., Hillary, R. F., Lu, A. T., Murphy, L., Wrobel, N., Campbell, A., Harris, S. E., Liewald, D., Hayward, C., Sudlow, C., Cox, S. R., Evans, K. L., Horvath, S., McIntosh, A. M., Robinson, M. R., Vallejos, C. A., & Marioni, R. E. (2023). Refining epigenetic prediction of chronological and biological age. *Genome Medicine*, *15*(1), 12. https://doi.org/10.1186/s13073-023-01161-y

Bestor, T. H., Edwards, J. R., & Boulard, M. (2015). Notes on the role of dynamic DNA methylation in mammalian development. *Proceedings of the National Academy of Sciences*, *112*(22), 6796–6799. https://doi.org/10.1073/pnas.1415301111

Bisi, F., Newey, S., Nodari, M., Wauters, L. A., Harrison, A., Thirgood, S., & Martinoli, A. (2011). The strong and the hungry: Bias in capture methods for mountain hares Lepus timidus. *Wildlife Biology*, *17*(3), 311–316. https://doi.org/10.2981/10-133

Bors, E. K., Baker, C. S., Wade, P. R., O'Neill, K. B., Shelden, K. E. W., Thompson, M. J., Fei, Z., Jarman, S., & Horvath, S. (2021). An epigenetic clock to estimate the age of living beluga whales. *Evolutionary Applications*, *14*(5), 1263–1273. https://doi.org/10.1111/eva.13195

Calvert, W., & Ramsay, M. A. (1998). Evaluation of Age Determination of Polar Bears by Counts of Cementum Growth Layer Groups. *Ursus*, *10*, 449–453.

Camacho, C., Canal, D., & Potti, J. (2017). Lifelong effects of trapping experience lead to age-biased sampling: Lessons from a wild bird population. *Animal Behaviour*, *130*, 133–139. https://doi.org/10.1016/j.anbehav.2017.06.018

Carja, O., MacIsaac, J. L., Mah, S. M., Henn, B. M., Kobor, M. S., Feldman, M. W., & Fraser, H. B. (2017). Worldwide patterns of human epigenetic variation. *Nature Ecology & Evolution*, *1*(10), 1577–1583. https://doi.org/10.1038/s41559-017-0299-z

Chen, B. H., Marioni, R. E., Colicino, E., Peters, M. J., Ward-Caviness, C. K., Tsai, P.-C., Roetker, N. S., Just, A. C., Demerath, E. W., Guan, W., Bressler, J., Fornage, M., Studenski, S., Vandiver, A. R., Moore, A. Z., Tanaka, T., Kiel, D. P., Liang, L., Vokonas, P., … Horvath, S. (2016). DNA methylation-based measures of biological age: Meta-

analysis predicting time to death. *Aging*, *8*(9), 1844–1865. https://doi.org/10.18632/aging.101020

Czajka, N., Northrup, J. M., Jones, M. J., & Shafer, A. B. A. (2024). Epigenetic clocks, sex markers and age-class diagnostics in three harvested large mammals. *Molecular Ecology Resources*, *24*(5), e13956. https://doi.org/10.1111/1755-0998.13956

Fransquet, P. D., Wrigglesworth, J., Woods, R. L., Ernst, M. E., & Ryan, J. (2019). The epigenetic clock as a predictor of disease and mortality risk: A systematic review and meta-analysis. *Clinical Epigenetics*, *11*(1), 62. https://doi.org/10.1186/s13148-019-0656-7

Fraser, H. B., Lam, L. L., Neumann, S. M., & Kobor, M. S. (2012). Population-specificity of human DNA methylation. *Genome Biology*, *13*(2), R8. https://doi.org/10.1186/gb-2012-13-2-r8

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1). https://doi.org/10.18637/jss.v033.i01

Garde, E., Bertelsen, M. F., Ditlevsen, S., Heide-Jørgensen, M. P., Nielsen, N. H., Frie, A. K., Olafsdottir, D., Siebert, U., & Hansen, S. H. (2018). Accuracy of the Aspartic Acid Racemization Technique in Age Estimation of Mammals and the Influence of Body Temperature. *NAMMCO Scientific Publications*, *10*. https://doi.org/10.7557/3.4400

Gatev, E., Inkster, A. M., Negri, G. L., Konwar, C., Lussier, A. A., Skakkabaek, A., Sokolowski, M. B., Gravholt, C. H., Dunn, E. C., Kobor, M. S., & Aristizabal, M. J. (2021). Autosomal sex-associated co-methylated regions predict biological sex from DNA methylation. *Nucleic Acids Research*, *49*(16), 9097–9116. https://doi.org/10.1093/nar/gkab682

Gibson, J., Russ, T. C., Clarke, T.-K., Howard, D. M., Hillary, R. F., Evans, K. L., Walker, R. M., Bermingham, M. L., Morris, S. W., Campbell, A., Hayward, C., Murray, A. D., Porteous, D. J., Horvath, S., Lu, A. T., McIntosh, A. M., Whalley, H. C., & Marioni, R. E. (2019). A meta-analysis of genome-wide association studies of epigenetic age acceleration. *PLOS Genetics*, *15*(11), e1008104. https://doi.org/10.1371/journal.pgen.1008104

Grant, O. A., Wang, Y., Kumari, M., Zabet, N. R., & Schalkwyk, L. (2022). Characterising sex differences of autosomal DNA methylation in whole blood using the Illumina EPIC array. *Clinical Epigenetics*, *14*(1), 62. https://doi.org/10.1186/s13148-022-01279-7

Haftorn, K. L., Romanowska, J., Lee, Y., Page, C. M., Magnus, P. M., Håberg, S. E., Bohlin, J., Jugessur, A., & Denault, W. R. P. (2023). Stability selection enhances feature selection and enables accurate prediction of gestational age using only five DNA methylation sites. *Clinical Epigenetics*, *15*(1), 114. https://doi.org/10.1186/s13148-023-01528-3

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.-B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., & Zhang, K. (2013). Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Molecular Cell*, *49*(2), 359–367. https://doi.org/10.1016/j.molcel.2012.10.016

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Higgins-Chen, A. T., Thrush, K. L., Wang, Y., Minteer, C. J., Kuo, P.-L., Wang, M., Niimi, P., Sturm, G., Lin, J., Moore, A. Z., Bandinelli, S., Vinkers, C. H., Vermetten, E., Rutten, B.

P. F., Geuze, E., Okhuijsen-Pfeifer, C., Van Der Horst, M. Z., Schreiter, S., Gutwinski, S., … Levine, M. E. (2022). A computational solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and longitudinal tracking. *Nature Aging*, *2*(7), 644–661. https://doi.org/10.1038/s43587-022-00248-2

Holmes, E. E., Fritz, L. W., York, A. E., & Sweeney, K. (2007). AGE-STRUCTURED MODELING REVEALS LONG-TERM DECLINES IN THE NATALITY OF WESTERN STELLER SEA LIONS. *Ecological Applications*, *17*(8), 2214–2232. https://doi.org/10.1890/07-0508.1

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

Horvath, S., & Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, *19*(6), 371–384. https://doi.org/10.1038/s41576-018-0004-3

Hostetter, N. J., Lunn, N. J., Richardson, E. S., Regehr, E. V., & Converse, S. J. (2021). Age-structured Jolly-Seber model expands inference and improves parameter estimation from capture-recapture data. *PLOS ONE*, *16*(6), e0252748. https://doi.org/10.1371/journal.pone.0252748

Jackson, J., Mar, K. U., Htut, W., Childs, D. Z., & Lummaa, V. (2020). Changes in age-structure over four decades were a key determinant of population growth rate in a long-lived mammal. *Journal of Animal Ecology*, *89*(10), 2268–2278. https://doi.org/10.1111/1365-2656.13290

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Jung, M., & Pfeifer, G. P. (2015). Aging and DNA methylation. *BMC Biology*, *13*(1), 7. https://doi.org/10.1186/s12915-015-0118-4

Jylhävä, J., Hjelmborg, J., Soerensen, M., Munoz, E., Tan, Q., Kuja-Halkola, R., Mengel-From, J., Christensen, K., Christiansen, L., Hägg, S., Pedersen, N. L., & Reynolds, C. A. (2019). Longitudinal changes in the genetic and environmental influences on the epigenetic clocks across old age: Evidence from two twin cohorts. *EBioMedicine*, *40*, 710–716. https://doi.org/10.1016/j.ebiom.2019.01.040

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer New York. https://doi.org/10.1007/978-1-4614-6849-3

Kurdyukov, S., & Bullock, M. (2016). DNA Methylation Analysis: Choosing the Right Method. *Biology*, *5*(1), 3. https://doi.org/10.3390/biology5010003

Le Clercq, L., Kotzé, A., Grobler, J. P., & Dalton, D. L. (2023). Biological clocks as age estimation markers in animals: A systematic review and meta-analysis. *Biological Reviews*, brv.12992. https://doi.org/10.1111/brv.12992

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, *28*(6), 882–883. https://doi.org/10.1093/bioinformatics/bts034

Lemaître, J.-F., Ronget, V., Tidière, M., Allainé, D., Berger, V., Cohas, A., Colchero, F., Conde, D. A., Garratt, M., Liker, A., Marais, G. A. B., Scheuerlein, A., Székely, T., & Gaillard, J.-M. (2020). Sex differences in adult lifespan and aging rates of mortality across wild mammals. *Proceedings of the National Academy of Sciences*, *117*(15), 8546–8553. https://doi.org/10.1073/pnas.1911999117

Li, A., Mueller, A., English, B., Arena, A., Vera, D., Kane, A. E., & Sinclair, D. A. (2022). Novel feature selection methods for construction of accurate epigenetic clocks. *PLOS Computational Biology*, *18*(8), e1009938. https://doi.org/10.1371/journal.pcbi.1009938

Lu, A. T., Fei, Z., Haghani, A., Robeck, T. R., Zoller, J. A., Li, C. Z., Lowe, R., Yan, Q., Zhang, J., Vu, H., Ablaeva, J., Acosta-Rodriguez, V. A., Adams, D. M., Almunia, J., Aloysius, A., Ardehali, R., Arneson, A., Baker, C. S., Banks, G., … Horvath, S. (2023). Universal DNA methylation age across mammalian tissues. *Nature Aging*. https://doi.org/10.1038/s43587-023-00462-6

Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D., Whitsel, E. A., Assimes, T. L., Ferrucci, L., & Horvath, S. (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging*, *11*(2), 303–327. https://doi.org/10.18632/aging.101684

Marioni, R. E., Shah, S., McRae, A. F., Chen, B. H., Colicino, E., Harris, S. E., Gibson, J., Henders, A. K., Redmond, P., Cox, S. R., Pattie, A., Corley, J., Murphy, L., Martin, N. G., Montgomery, G. W., Feinberg, A. P., Fallin, M. D., Multhaup, M. L., Jaffe, A. E., … Deary, I. J. (2015). DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biology*, *16*(1), 25. https://doi.org/10.1186/s13059-015-0584-6

Mayne, B., Berry, O., & Jarman, S. (2021). Optimal sample size for calibrating DNA methylation age estimators. *Molecular Ecology Resources*, *21*(7), 2316–2323. https://doi.org/10.1111/1755-0998.13437

Mayne, B., Berry, O., & Jarman, S. (2023). Calibrating epigenetic clocks with training data error. *Evolutionary Applications*, *16*(8), 1496–1502. https://doi.org/10.1111/eva.13582

McCartney, D. L., Zhang, F., Hillary, R. F., Zhang, Q., Stevenson, A. J., Walker, R. M., Bermingham, M. L., Boutin, T., Morris, S. W., Campbell, A., Murray, A. D., Whalley, H. C., Porteous, D. J., Hayward, C., Evans, K. L., Chandra, T., Deary, I. J., McIntosh, A. M., Yang, J., … Marioni, R. E. (2020). An epigenome-wide association study of sex-specific chronological ageing. *Genome Medicine*, *12*(1), 1. https://doi.org/10.1186/s13073-019-0693-z

McEwen, L. M., O'Donnell, K. J., McGill, M. G., Edgar, R. D., Jones, M. J., MacIsaac, J. L., Lin, D. T. S., Ramadori, K., Morin, A., Gladish, N., Garg, E., Unternaehrer, E., Pokhvisneva, I., Karnani, N., Kee, M. Z. L., Klengel, T., Adler, N. E., Barr, R. G., Letourneau, N., … Kobor, M. S. (2020). The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. *Proceedings of the National Academy of Sciences*, *117*(38), 23329–23335. https://doi.org/10.1073/pnas.1820843116

Moqri, M., Herzog, C., Poganik, J. R., Justice, J., Belsky, D. W., Higgins-Chen, A., Moskalev, A., Fuellen, G., Cohen, A. A., Bautmans, I., Widschwendter, M., Ding, J., Fleming, A., Mannick, J., Han, J.-D. J., Zhavoronkov, A., Barzilai, N., Kaeberlein, M., Cummings, S., … Gladyshev, V. N. (2023). Biomarkers of aging for the identification and evaluation of longevity interventions. *Cell*, *186*(18), 3758–3775. https://doi.org/10.1016/j.cell.2023.08.003

Morris, P. (1972). A review of mammalian age determination methods. *Mammal Review*, *2*(3), 69–104. https://doi.org/10.1111/j.1365-2907.1972.tb00160.x

Newediuk, L., Richardson, E. S., Biddlecombe, B. A., Kassim, H., Kathan, L., Lunn, N., Ruth Rivkin, L., Salama, O. E., Schmidt, C., Jones, M. J., & Garroway, C. J. (2024). *Climate change, age acceleration, and the erosion of fitness in polar bears*. https://doi.org/10.1101/2024.01.05.574416

Parsons, K. M., Haghani, A., Zoller, J. A., Lu, A. T., Fei, Z., Ferguson, S. H., Garde, E., Hanson, M. B., Emmons, C. K., Matkin, C. O., Young, B. G., Koski, W. R., & Horvath, S. (2023). DNA methylation-based biomarkers for ageing long-lived cetaceans. *Molecular Ecology Resources*, 1755-0998.13791. https://doi.org/10.1111/1755-0998.13791

Perna, L., Zhang, Y., Mons, U., Holleczek, B., Saum, K.-U., & Brenner, H. (2016). Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clinical Epigenetics*, *8*(1), 64. https://doi.org/10.1186/s13148-016-0228-z

Porter, H. L., Brown, C. A., Roopnarinesingh, X., Giles, C. B., Georgescu, C., Freeman, W. M., & Wren, J. D. (2021). Many chronological aging clocks can be found throughout the epigenome: Implications for quantifying biological aging. *Aging Cell*, *20*(11), e13492. https://doi.org/10.1111/acel.13492

Prado, N. A., Brown, J. L., Zoller, J. A., Haghani, A., Yao, M., Bagryanova, L. R., Campana, M. G., Maldonado, J., Raj, K., Schmitt, D., Robeck, T. R., & Horvath, S. (2021). Epigenetic clock and methylation studies in elephants. *Aging Cell*, *20*(7). https://doi.org/10.1111/acel.13414

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.3.1) [R]. R Foundation for Statistical Computing. https://www.R-project.org/

Raj, K., Szladovits, B., Haghani, A., Zoller, J. A., Li, C. Z., Black, P., Maddox, D., Robeck, T. R., & Horvath, S. (2021). Epigenetic clock and methylation studies in cats. *GeroScience*, *43*(5), 2363–2378. https://doi.org/10.1007/s11357-021-00445-8

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/nar/gkv007

Robeck, T. R., Fei, Z., Lu, A. T., Haghani, A., Jourdain, E., Zoller, J. A., Li, C. Z., Steinman, K. J., DiRocco, S., Schmitt, T., Osborn, S., Van Bonn, B., Katsumata, E., Mergl, J., Almunia, J., Rodriguez, M., Haulena, M., Dold, C., & Horvath, S. (2021). Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Communications Biology*, *4*(1), 642. https://doi.org/10.1038/s42003-021-02179-x

Robeck, T. R., Haghani, A., Fei, Z., Lindemann, D. M., Russell, J., Herrick, K. E. S., Montano, G., Steinman, K. J., Katsumata, E., Zoller, J. A., & Horvath, S. (2023). Multi-tissue DNA methylation aging clocks for sea lions, walruses and seals. *Communications Biology*, *6*(1), 359. https://doi.org/10.1038/s42003-023-04734-0

Romero, L. M., & Beattie, U. K. (2022). Common myths of glucocorticoid function in ecology and conservation. *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology*, *337*(1), 7–14. https://doi.org/10.1002/jez.2459

Sarac, F., Seker, H., & Bouridane, A. (2017). Exploration of unsupervised feature selection methods to predict chronological age of individuals by utilising CpG dinucleotics from whole blood. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3652–3655. https://doi.org/10.1109/EMBC.2017.8037649

Shireby, G. L., Davies, J. P., Francis, P. T., Burrage, J., Walker, E. M., Neilson, G. W. A., Dahir, A., Thomas, A. J., Love, S., Smith, R. G., Lunnon, K., Kumari, M., Schalkwyk, L. C., Morgan, K., Brookes, K., Hannon, E., & Mill, J. (2020). Recalibrating the epigenetic clock: Implications for assessing biological age in the human cortex. *Brain*, *143*(12), 3763–3775. https://doi.org/10.1093/brain/awaa334

Simpkin, A. J., Hemani, G., Suderman, M., Gaunt, T. R., Lyttleton, O., Mcardle, W. L., Ring, S. M., Sharp, G. C., Tilling, K., Horvath, S., Kunze, S., Peters, A., Waldenberger, M., Ward-Caviness, C., Nohr, E. A., Sørensen, T. I. A., Relton, C. L., & Smith, G. D. (2016). Prenatal and early life influences on epigenetic age in children: A study of mother–offspring pairs from two cohort studies. *Human Molecular Genetics*, *25*(1), 191–201. https://doi.org/10.1093/hmg/ddv456

Smith, G. C., Pugh, B., & Trout, R. C. (1995). Age and sex bias in samples of wild rabbits, *Oryctolagus cuniculus*, from wild populations in southern England. *New Zealand Journal of Zoology*, *22*(2), 115–121. https://doi.org/10.1080/03014223.1995.9518028

Stubbs, T. M., Bonder, M. J., Stark, A.-K., Krueger, F., Von Meyenn, F., Stegle, O., & Reik, W. (2017). Multi-tissue DNA methylation age predictor in mouse. *Genome Biology*, *18*(1), 68. https://doi.org/10.1186/s13059-017-1203-5

Sugrue, V. J., Zoller, J. A., Narayan, P., Lu, A. T., Ortega-Recalde, O. J., Grant, M. J., Bawden, C. S., Rudiger, S. R., Haghani, A., Bond, D. M., Hore, R. R., Garratt, M., Sears, K. E., Wang, N., Yang, X. W., Snell, R. G., Hore, T. A., & Horvath, S. (2021). Castration delays epigenetic aging and feminizes DNA methylation at androgen-regulated loci. *eLife*, *10*, e64932. https://doi.org/10.7554/eLife.64932

Theng, D., & Bhoyar, K. K. (2024). Feature selection techniques for machine learning: A survey of more than two decades of research. *Knowledge and Information Systems*, *66*(3), 1575–1637. https://doi.org/10.1007/s10115-023-02010-5

Thompson, M. J., vonHoldt, B., Horvath, S., & Pellegrini, M. (2017). An epigenetic aging clock for dogs and wolves. *Aging*, *9*(3), 1055–1068. https://doi.org/10.18632/aging.101211

Triche, T. J., Weisenberger, D. J., Van Den Berg, D., Laird, P. W., & Siegmund, K. D. (2013). Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Research*, *41*(7), e90–e90. https://doi.org/10.1093/nar/gkt090

Wilkinson, G. S., Adams, D. M., Haghani, A., Lu, A. T., Zoller, J., Breeze, C. E., Arnold, B. D., Ball, H. C., Carter, G. G., Cooper, L. N., Dechmann, D. K. N., Devanna, P., Fasel, N. J., Galazyuk, A. V., Günther, L., Hurme, E., Jones, G., Knörnschild, M., Lattenkamp, E. Z., … Horvath, S. (2021). DNA methylation predicts age and provides insight into exceptional longevity of bats. *Nature Communications*, *12*(1), 1615. https://doi.org/10.1038/s41467-021-21900-2

Yusipov, I., Bacalini, M. G., Kalyakulina, A., Krivonosov, M., Gensous, N., Ravaioli, F., Milazzo, M., Giuliani, C., Fiorito, G., Gagliardi, A., Polidoro, S., Ivanchenko, M., & Franceschi, C. (2020). Age-related DNA methylation changes are sex-specific: A comprehensive assessment. *Aging*, *12*(23), 24057–24080.

Zannas, A. S. (2019). Epigenetics as a key link between psychosocial stress and aging: Concepts, evidence, mechanisms. *Dialogues in Clinical Neuroscience*, *21*(4), 389–396. https://doi.org/10.31887/DCNS.2019.21.4/azannas

Zhang, Q., Vallerga, C. L., Walker, R. M., Lin, T., Henders, A. K., Montgomery, G. W., He, J., Fan, D., Fowdar, J., Kennedy, M., Pitcher, T., Pearson, J., Halliday, G., Kwok, J. B., Hickie, I., Lewis, S., Anderson, T., Silburn, P. A., Mellick, G. D., … Visscher, P. M. (2019). Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Medicine*, *11*(1), 54. https://doi.org/10.1186/s13073-019-0667-1

Zhang, Y., Bi, J., Ning, Y., & Feng, J. (2024). Methodology Advances in Vertebrate Age Estimation. *Animals*, *14*(2), 343. https://doi.org/10.3390/ani14020343

Zhou, W., Triche, T. J., Laird, P. W., & Shen, H. (2018). SeSAMe: Reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky691

Zhuang, J., Widschwendter, M., & Teschendorff, A. E. (2012). A comparison of feature selection and classification methods in DNA methylation studies using the Illumina Infinium platform. *BMC Bioinformatics*, *13*(1), 59. https://doi.org/10.1186/1471-2105-13-59

Zoller, J., & Horvath, S. (2024). MammalMethylClock R package: Software for DNA methylation-based epigenetic clocks in mammals. *Bioinformatics*, *40*(5), btae280. https://doi.org/10.1093/bioinformatics/btae280

Data Accessibility Statement: All data and code are available from https://github.com/ljnewediuk/how_to_clocks.git. The data will also be made available on Dryad upon acceptance.

Benefit-sharing Statement: All collaborators contributing community data for this study were offered authorship. The results of this work are shared with the local communities through these contributors.

Author Contributions: LN, MJJ, CJG, and ER conceived of the study. AB and ARC contributed data. MJJ contributed lab resources. LN analyzed the data and wrote the manuscript. All authors provided feedback on the manuscript.