UC Riverside

UC Riverside Previously Published Works

Title
An epigenetic clock to estimate the age of living beluga whales.

Authors
Bors, Eleanor
Baker, C
Wade, Paul
et al.

Peer reviewed

ORIGINAL ARTICLE

An epigenetic clock to estimate the age of living beluga whales

Eleanor K. Bors[1] [iD] | C. Scott Baker[1] [iD] | Paul R. Wade[2] [iD] | Kaimyn B. O'Neill[1] |
Kim E. W. Shelden[2] [iD] | Michael J. Thompson[3] | Zhe Fei[4] | Simon Jarman[5] [iD] |
Steve Horvath[4,6] [iD]

[1]Marine Mammal Institute, Oregon State University, Newport, OR, USA

[2]Marine Mammal Laboratory, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, Seattle, WA, USA

[3]Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA, USA

[4]Department of Biostatistics, School of Public Health, University of California-Los Angeles, Los Angeles, CA, USA

[5]School of Biological Sciences, University of Western Australia, Perth, WA, Australia

[6]Department of Human Genetics, Gonda Research Center, David Geffen School of Medicine, Los Angeles, CA, USA

Correspondence
Eleanor K. Bors, Cooperative Institute for Climate, Oceans, and Ecosystem Studies, University of Washington, Seattle, WA, USA.
Email: ekbors@gmail.com

C. Scott Baker, Marine Mammal Institute, Hatfield Marine Science Center, Oregon State University, Newport, OR, USA.
Email: scott.baker@oregonstate.edu

Steve Horvath, Department of Biostatistics, School of Public Health, University of California Los Angeles, Los Angeles, CA, USA.
Email: shorvath@mednet.ucla.edu

Funding information
North Pacific Research Board, Grant/Award Number: 1723; Cooperative Institute for Marine Resource Studies, Grant/Award Number: NB293T; Allen Fronteirs Group

Abstract

DNA methylation data facilitate the development of accurate molecular estimators of chronological age or "epigenetic clocks." We present a robust epigenetic clock for the beluga whale, *Delphinapterus leucas*, developed for an endangered population in Cook Inlet, Alaska, USA. We used a custom methylation array to measure methylation levels at 37,491 cytosine–guanine sites (CpGs) from skin samples of dead whales ($n = 67$) whose chronological ages were estimated based on tooth growth layer groups. Using these calibration data, a penalized regression model selected 23 CpGs, providing an $R^2 = 0.92$ for the training data; and an $R^2 = 0.74$ and median absolute age error = 2.9 years for the leave one out cross-validation. We applied the epigenetic clock to an independent dataset of 38 skin samples collected with a biopsy dart from living whales between 2016 and 2018. Age estimates ranged from 11 to 27 years. We also report sex correlations in CpG data and describe an approach of identifying the sex of an animal using DNA methylation. The epigenetic estimators of age and sex presented here have broad applications for conservation and management of Cook Inlet beluga whales and potentially other cetaceans.

KEYWORDS
aging, Alaska, cetaceans, conservation biology, Cook Inlet, DNA methylation, endangered species, wildlife management

1 | INTRODUCTION

Age is a fundamental life-history parameter in organismal biology, population dynamics, and ecology. The age of an animal is important for understanding characteristics such as age of reproductive maturity, fecundity rates, and survival rates. These characteristics can vary between healthy and compromised populations. Moreover, knowing the age of animals in a population can improve the study of population dynamics. For example, age-specific estimates of fecundity and survival can be used to predict population growth rate (Brault & Caswell, 1993), and the age structure of a population can imply past trajectory of the population (Venuto et al., 2020). Additionally, age can enhance the interpretation of genetic analyses in some cases (e.g., kinship analysis). Therefore, the ability to determine the age of animals is an important tool in wildlife studies. In cetaceans, age is critically important but often unknown due to the difficulty of determining age in long-lived, mobile species.

The development of molecular aging biomarkers (MABs) for mammals in particular has been of interest for decades (Jarman et al., 2015), with numerous lines of inquiry into the role of molecular mechanisms in aging. Molecular aging studies in cetaceans initially focused on the relationship between telomere length and age, but that line of inquiry proved unfruitful (Dunshea et al., 2011; Jarman et al., 2015; Olsen et al., 2012). Attention has turned to other MABs such as epigenetic markers, specifically DNA methylation, which have received considerable attention in recent years (De Paoli-Iseppi et al., 2017; Horvath & Raj, 2018; Jarman et al., 2015; Jylhävä et al., 2017). Epigenetics is broadly understood as the study of any gene-regulating activity that does not involve changes to a DNA sequence and can be, but is not necessarily, heritable (Pennisi, 2001). The term encompasses myriad molecular processes ranging from chromatin state to direct chemical modification of DNA (e.g., methylation). At specific cytosine–guanine dinucleotides (CpGs), the cytosine nucleotide can be methylated to generate 5-methylcytosine, a chemical modification that affects gene expression (Field et al., 2018; Razin & Cedar, 1991). Methylation levels at some of these sites have been shown to correlate with age.

Numerous DNA methylation-based age predictors, often referred to as "epigenetic clocks," have been developed for humans (Bocklandt et al., 2011; Garagnani et al., 2012; Hannum et al., 2013; Horvath, 2013; Levine et al., 2018; Lin et al., 2016; Lowe et al., 2018). Because methylation patterns are known to be tissue-specific, some epigenetic clocks use a single tissue such as blood (Hannum et al., 2013), while pan-tissue clocks appear to apply to all sources of DNA except sperm (Horvath, 2013). In humans, clocks have been designed using varying numbers of CpG sites, from one site (e.g., an age predictor based on a CpG in the *ELOVL2* gene; Garagnani et al., 2012) to several hundred sites (e.g., 353 sites; Horvath, 2013). Epigenetic clocks have also been developed for other species including the mouse (Meer et al., 2018; Petkovich et al., 2017; Stubbs et al., 2017; Thompson et al., 2018), chimpanzee (Ito et al., 2018), bat (Wright et al., 2018), canid (Ito et al., 2017; Thompson et al., 2017), humpback whale (Polanowski et al., 2014), minke whale (Tanabe et al., 2020),

and bottlenose dolphin (Beal et al., 2019). Accurate age estimates can be valuable for conservation efforts and species management. For example, the use of age structure data for harbor porpoise allowed for estimation of the maximum rate of increase for the population, leading to the conclusion that bycatch mortality in commercial fisheries had led to population decline (Moore & Read, 2008).

Our focus here is the beluga whale, *Delphinapterus leucas* (Pallas, 1776). Beluga whales inhabit the circumpolar north with southernmost populations occurring in the Saint Lawrence Estuary in Eastern Canada, the Sea of Okhotsk in Eastern Russia, and Cook Inlet in Alaska, USA. The Cook Inlet (CI) beluga whale population does not migrate, is geographically and genetically isolated (O'Corry-Crowe et al., 1997), and is the focus of conservation and management efforts (NOAA, 2016). Estimates of abundance for this population numbered over 1000 whales in the late 1970s and early 1990s (Shelden et al., 2015). From 1994 to 1998, abundance declined steeply from 653 to 347 whales, in part, due to unregulated hunting (Hobbs et al., 2000; Mahoney & Shelden, 2000). In 2000, hunting regulations were implemented; the CI beluga population was designated as a distinct population segment (DPS), recognizing that CI beluga whales constitute a population that is "discrete from other populations and significant in relation to the entire taxon" (65 FR 38788 22 June 2000); and CI beluga whales were listed as depleted under the U.S. Marine Mammal Protection Act (65 FR 34590, 21 May 2000). Eight years later, the DPS was listed as endangered under the U.S. Endangered Species Act (ESA) (73 FR 62919, 22 October 2008). Today, there are an estimated 279 individuals in the population, and the number is declining (Wade et al., 2019). The U.S. National Oceanic and Atmospheric Administration (NOAA) released the *Cook Inlet Beluga Whale Recovery Plan* in 2016 (NMFS, 2016) pursuant to the requirements of the ESA. This plan highlights the importance of determining the population age structure of CI beluga whales in order to understand growth, reproduction, and survival rates.

Age determination of beluga whales to date has relied on data derived from tooth growth layer groups (Lockyer et al., 2007; Waugh et al., 2018), a method that is also applicable to some other toothed whale species (Hamilton & Evans, 2018; Perrin & Myrick, 1980). The tooth samples required for aging studies are typically acquired from dead animals. Efforts to develop methods that estimate the age of living animals have led to the development of length–age curves for adult belugas (Vos et al., 2019) as well as fetuses and neonates (Robeck et al., 2015). Methods that use length to estimate age encounter complications due to asymptotic growth curves in adults, meaning that after a certain age only a minimum age estimate is possible.

Here, we present an epigenetic clock and a sex-predictive logistic model for beluga whales based on DNA methylation data. This study leverages long-term sampling of the CI beluga population, recent development of beluga genomic resources, advances in methylation array technology, and machine learning to develop a novel method to age living beluga whales based on DNA from skin samples. The epigenetic age estimator (epigenetic clock) presented here will aid in the management and conservation of this endangered cetacean

population, as demonstrated by our application of the beluga epigenetic clock to estimate the age of living beluga whales sampled with a biopsy dart.

2 | METHODS

2.1 | Sample collection, chronological age estimation, and DNA extraction

Skin tissue samples were collected from carcasses of beluga whales that were beach-cast, stranded dead, or taken during subsistence hunting between 1992 and 2015 in Cook Inlet, Alaska, USA (NMFS Research Permit 932-1905-00/MA-009526 through the Marine Mammal Health and Stranding Response Program). Skin samples were preserved in a salt and dimethyl sulfoxide (DMSO) solution and archived at NOAA's Southwest Fisheries Science Center in La Jolla, California, USA. A total of 69 individuals were selected for the clock calibration dataset (Table S1), and their chronological ages were estimated by counting tooth growth layer groups (Vos et al., 2019). The final calibration dataset included 67 individuals due to inconsistent molecular sex data (see below). Teeth were analyzed by at least two readers using methods validated in Lockyer et al. (2007), and a consensus age provided by NOAA was used in this study. When individuals were represented by multiple teeth in the dataset, the oldest age estimate was used to mitigate error from tooth wear (e.g., the count from the tooth with the greatest number of growth layer groups; Vos et al., 2019).

Samples of skin tissue from living CI beluga whales were collected with a biopsy dart in 2016, 2017, and 2018 (NMFS ESA/MMPA Permit #20465; McGuire, Michaud, et al., 2017). Biopsy samples were frozen in the field in liquid nitrogen and later subsampled at the NOAA Alaska Fisheries Science Center in Seattle, Washington, USA. Genomic DNA was extracted from tissue samples using a standard phenol–chloroform protocol modified for small skin samples by Baker et al. (1994). Extracted DNA was treated with RNAse A (1 µl of 1 mg/ml added to samples of 100 µl for 30 min at room temperature) and then purified and concentrated using a DNA Clean and Concentrator-5 Kit (Zymo Research Corp.). The concentration of genomic DNA was measured on a QUBIT 4 fluorometer (Thermo Fisher Scientific).

2.2 | Molecular sex identification

A multiplexed polymerase chain reaction (PCR) was used to sex individual whales in both the calibration and biopsy datasets. The PCR primers and reaction protocol followed those in Olavarría et al. (2007), which is based on Gilson et al. (1998). This assay amplifies fragments of the male-specific SRY gene and the ZFY/ZFX genes of males and females as a control band. Sex-specific bands were visualized by agarose gel electrophoresis. In three cases, molecular sex identified tissue samples that did not correlate with sex metadata in

the original records (z35345, z143907, z144309). These cases have been noted and amended in the records presented here (Table S1). Two of these individuals were removed from the dataset because the molecular sex could not be reconciled with information from necropsies. One was retained because it did not conflict with known information. Therefore, the final calibration dataset included 67 individuals.

2.3 | DNA methylation measurements

Genomic DNA aliquots were sent to the UCLA Neurosciences Genomics Core facility where they were quantified and bisulfite converted using the Zymo EZ-96 DNA Methylation-Gold Kit (Zymo, Inc.; Cat# D5007). When possible, a total of 250 ng of genomic DNA were used for each individual (in a few cases, lower quantities were used when DNA concentration was too low to achieve 250 ng with a maximum volume loadable of 20 µl). A custom mammalian methylation array (HorvathMammalMethylChip40) assembled with 37,491 oligonucleotide probes, each 50 nucleotides long terminating in a C-G dinucleotide, was used to determine methylation state of CpGs (Arneson et al., 2021). Genomic regions of interest were located in conserved regions across mammalian genomes, and probes were designed using human, mouse, and other mammal sequences with the expectation that any given probe would likely work in a certain subset of mammalian species (Arneson et al., 2021). The particular subset of species for each probe is provided in the chip manifest file can be found at Gene Expression Omnibus (GEO) at NCBI as platform GPL28271.

The clock training/calibration dataset (historic samples) was evaluated in one round of array assays (one sample per array, 24 arrays per chip), and the biopsy dataset (recent skin tissue samples) was evaluated in another. Fluorescence at the terminal nucleotide of each probe was read by an Illumina iScan machine, and raw data were provided in iDAT files. Raw data were normalized using the SeSAMe pipeline (Zhou et al., 2018) resulting in a methylation estimate (beta value) corresponding to each array probe for every individual in the dataset and a detection p-value corresponding to the confidence in the normalized beta value. Beta values are derived from the ratio of the fluorescence intensity of a methylated probe for a specific CpG to the total overall probe intensity (the sum of signal from both the methylated and unmethylated probes plus a constant) (Du et al., 2010). Beta values range from zero to one with a value of zero indicating that no copies of the gene were methylated.

To identify technical outliers after SeSAMe normalization, we used unsupervised hierarchical clustering analysis based on the interarray correlation. As a consequence, data for one sample were removed from the dataset and replaced with data generated for another DNA extraction of the same tissue sample that did not exhibit anomalous clustering. Data were filtered by detection p-value as calculated in SeSAMe (Zhou et al., 2018). To test the effect of p-value filtering on downstream analyses, we evaluated analyses on data with different thresholds for the number of individuals with

a passable detection *p*-value (e.g., a detection *p*-value of <0.05 in more than one individual, in over 10 individuals, in over 20). Ultimately, CpG sites that had a detection *p*-value <0.05 for 10 or more individuals were considered in further clock-building analyses, resulting in the use of data from 28,875 CpG probes from the array (Table S2).

2.4 | Sex-correlated CpGs and methylation-based sex prediction

The correlation between CpG methylation levels and whale sex was evaluated using Pearson's correlation using *NymPy* in a *Python* environment. To determine how many of the beluga sex-correlated CpGs are likely on X or Y chromosomes, and how many represent sexual dimorphism in autosomal methylation levels, we used information about the location of each probe in the human genome from the methylation array manifest. A logistic model for sex prediction was built using LASSO regression in *cv.glmnet()* ($\alpha = 1$) for binomial parameters (numerical coding was 1 = female; 0 = male).

2.5 | Age correlation of CpG sites and epigenetic clock development through elastic net regularization

Pearson's correlations between beta values for individual CpGs and chronological age were calculated using *NumPy* (Oliphant, 2006) and *Pandas* (McKinney, 2010). The absolute values of individual CpG correlations were ranked using a custom *Python* script.

The *glmnet* package (Friedman et al., 2010) was implemented in *R* (R core Team, 2013) to fit penalized regression models. The two main parameters used in this machine learning regularization method are lambda, which is known as the regularization parameter that sets the stringency of the penalty during regularization (high lambdas lead to stronger penalization); and alpha, which is the elastic net mixing parameter that is used to determine the blend between a ridge regression ($\alpha = 0$) and a least absolute shrinkage and selection (LASSO) regression ($\alpha = 1$). For all runs, the lambda used was the *lambda.min* parameter calculated by *cv.glmnet* using a 10-fold cross-validation. In elastic net regularization, the alpha parameter will determine the number of sites used in the clock: ridge regression will retain all the sites and LASSO regression will retain fewer sites.

Alpha values of 0.1 through 0.9 with a 0.1 interval were evaluated through multiple runs of *cv.glmnet()* (note that an alpha value of 0.0 would have yielded a model using all 28,875 CpGs). The resulting models were then used to calculate the age of each individual in the calibration dataset. The relationship between model ages and chronological ages based on tooth growth layer groups was evaluated using linear regression in *NymPy* and *Pandas* within a *Python* environment. Age error was calculated for each individual, which was defined as the difference between the estimated chronological age from tooth growth layer groups and the age prediction resulting

from the multivariate linear regression model. Regression coefficients, mean absolute age error, and median absolute age error were calculated for the dataset as a whole and for each sex independently.

To evaluate the likely accuracy of each model for estimating the age of future experimental samples, leave one out cross-validation (LOOCV) was run by executing the *cv.glmnet()* program on $n - 1$ samples, looping through each of the 67 samples. The predicted age of the omitted sample was calculated with a model developed with the remaining 66 samples. For each LOOCV iteration, a different model was generated, but the same alpha value was used in each iteration of *cv.glmnet()* (only the samples used changed). LOOCV elastic net models were run with a lambda value of *lambda.min* calculated within *cv.glmnet()* program by running a 10-fold internal cross-validation. LOOCV were executed at numerous alpha values to better understand the effect of alpha on model performance. LOOCV models were assessed in the same manner as the full elastic net regression models, using linear regression as well as age error calculations for the full dataset and for each sex independently.

2.6 | Genomic location and identity of clock CpGs

The location of each clock CpG probe in the human or mouse genomes was known from methylation array design. The genomic locations of each clock CpG and flanking sequences (200 bp in both directions) were extracted from the human or mouse genomes through the NCBI genome data viewer, leveraging the RefSeq database (O'Leary et al., 2016). The extracted mouse or human sequence was then located in the beluga genome (GenBank genome scaffold accession number: ASM228892v3) with NCBI BLAST (Altschul, 1990; Johnson et al., 2008). Annotations at each CpG site, or for the closest gene, were recorded. The relative locations of each CpG in the final epigenetic clock were assessed to identify any potentially linked sites. The same methods were also used to determine the annotation of the single CpG used in the sex prediction model.

2.7 | Age determination of living whales using skin biopsy samples

We applied the beluga epigenetic clock model to the dataset of living beluga whales sampled by biopsy dart. Beta values generated by the custom mammalian methylation array for each of the clock CpG sites were used in the beluga epigenetic clock to calculate the epigenetic ages of sampled whales. The absolute median age error from LOOCV of the calibration dataset was used as a course approximation for the potential range of the epigenetic age estimation generated by the epigenetic clock. When appropriate and possible, age estimates generated with the beluga epigenetic clock were compared to photographs of each individual and compared to subjective color-classes used to assess age in the field (McGuire et al., 2018; McGuire, Stephens, et al., 2017; unpublished data, P. Wade; Figure S1).

3 | RESULTS

3.1 | Chronological ages and molecular sex of clock calibration samples

The estimates of chronological ages derived from tooth growth layer groups for the 67 clock calibration samples ranged from −1 (fetus) to 49 years with a median age of 21 years (Figure 1). Previous records of sex were confirmed by PCR amplification of sex-specific primers. After correction, the ratio of males to females in our calibration dataset was 36 to 31. The median age of males in the dataset was 20 while that of females was 22. However, the three oldest samples in the dataset are males based on molecular sex data (Figure 1).

3.2 | Sex-correlated CpG sites and methylation-based sex prediction model

In addition to correlations between methylation and age, the relationship between methylation and sex was also investigated. Methylation levels at 165 CpG sites had Pearson's correlation values of 0.9 or greater with sex (Table S3). Of these 165 sites, 160 were located on the X chromosome in humans (Table S3), one was located on the Y chromosome in humans, two were located on autosomal chromosomes in humans, and two did not have known coordinates in the human genome. The two autosomal sites included probe cg26452915 (Chr20:58911021, annotated as GNAS) and cg25449272 (Chr15:56654033, annotated as ZNF280D). A logistic model ($p_{female} = 1/1 + e^{-(0.6717 - 1.1579*\beta value)}$) generated using a LASSO logistic regression machine learning method implemented in *cv.glmnet()* selected a single CpG: probe cg15451847, with a Pearson's correlation of $r = -0.999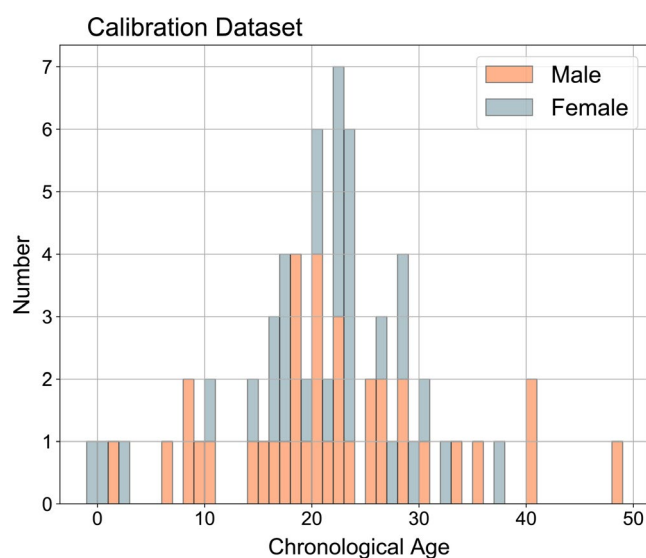$. The model predicted sex in the calibration samples (after thresholding the predicted probability >0.5 indicating a female and outputs <0.5 indicating a male). Probe cg15451847 corresponds to a site on the Y chromosome in humans (ChrY:19715996, annotated as KDM5D), indicating that its utility in predicting sex does not come from sexually dimorphic methylation patterns, but rather from the detection of a Y chromosome.

3.3 | Age-correlated CpG sites and the beluga epigenetic clock

The majority of CpG sites showed correlations with age of $r < 0.5$ ($n = 28,232$), and only 1.9% of CpG sites had a correlation of $r > 0.5$ ($n = 551$) (Table 1). No single site had a correlation coefficient larger than 0.9 but 21 sites exhibited correlations of between 0.8 and 0.9. The majority of CpGs assayed on the methylation array have negative correlations with age, and nearly 100% of the strongest correlations (>0.7) were negative (Table 1).

Using elastic net regularization with values of alpha between 0.1 and 1 at 0.1 intervals, *cv.glmnet()* yielded models of varying sizes using between 20 and 134 CpGs with R^2 values ranging from 0.923 to 0.942 (Table S4). The final model selected as the beluga epigenetic clock ($\alpha = 0.9$) uses 23 CpG sites to generate age predictions, meaning the clock model has 24 terms, including the *y*-intercept (Table 2). The information in Table 2 comprises the multiple linear regression model and is all the information needed to calculate age for new samples using data from the custom methylation array. We selected this specific model to optimize for median age error, R^2, and the *y*-intercept to reduce the tendency to overestimate the age of young whales. A linear regression of ages calculated with this model versus the tooth ages for each calibration sample

FIGURE 1 The distribution of chronological age by sex estimated from tooth growth layer groups for the calibration dataset ($n = 67$). Note that each bin represents one year and negative ages are fetal samples

TABLE 1 The frequency of absolute values of Pearson's correlation coefficients (r) for the relationship between methylation (beta values) at 28,875 CpG sites and chronological age based on teeth growth layers of the 67 calibration samples, with age in 0.1 bins (each range is inclusive of the lower bound and exclusive of the upper bound). The right column shows the percent of the CpGs in each bin that have a negative correlation

Absolute value Pearson's *r*	Number of CpG sites	% with negative correlation
0.0–0.1	14,133	56.5
0.1–0.2	8663	59.9
0.2–0.3	3887	58.1
0.3–0.4	1200	67.1
0.4–0.5	427	78.9
0.5–0.6	253	87.7
0.6–0.7	173	96.0
0.7–0.8	115	99.1
0.8–0.9	24	100
0.9–1.0	0	NA
Total	28,875	59.2

TABLE 2 The CpG sites selected for the beluga epigenetic clock with associated model coefficients (to be multiplied by the CpG beta value), including the y-intercept. The CpG sites are referenced to the array probe names (Table S5). Pearson's correlation coefficient (r) for the methylation ratio with the tooth growth layer count is shown for each individual CpG site (third column)

Probe ID	Model coefficient	CpG correlation
(Intercept)	77.9708623	NA
cg00952468	−4.525104457	−0.7578
cg02534193	−9.85801758	−0.6457
cg02714609	−14.10074358	−0.8206
cg07279255	14.33543764	0.6956
cg07493173	−0.175897809	−0.7854
cg09622321	−8.372865623	−0.8686
cg12584622	−6.025047791	−0.8685
cg14043264	−5.78974944	−0.7817
cg14671961	−5.341584389	−0.6566
cg15809488	−0.469547261	−0.7487
cg15992086	−0.379907718	−0.6122
cg16678811	−2.632026911	−0.7887
cg17856858	−1.525713524	−0.8161
cg18629679	−2.129559849	−0.8387
cg21419180	30.73879442	0.4590
cg21420547	5.593128066	0.5444
cg22069272	−5.967431136	−0.8290
cg22416332	−9.317426596	−0.7402
cg25579908	−19.26589983	−0.7404
cg26286303	−0.357946396	−0.7184
cg26313355	−1.762140707	−0.7743
cg26899365	−0.44830474	−0.7246
cg27600712	−0.104727313	−0.8134

3.4 | Genomic location and identity of clock CpG sites

The locations of all 23 CpG sites in the beluga genome were identified using BLAST and the NCBI genome viewer (Table S5). Annotations and gene ontology information for each CpG site revealed a wide range of gene identities and putative functions (Table S5). Of interest because of the role of epigenetics in gene regulation, four of the CpG sites are located in genes that have GO terms related to nuclear chromatin (cg00952468, cg15809488, cg21419180), promoter-specific chromatin binding (cg21419180), and chromatin remodeling (cg26899365). Fifteen of the 23 CpG sites were located within known genes that are annotated in the beluga genome; 18 of the sites are annotated within a gene in the human genome. Each annotation is unique, indicating that the CpGs are not linked within the same gene.

3.5 | Application of the beluga epigenetic clock to living beluga whales in the Cook Inlet

The methylation states of all 37,491 array CpGs were measured for genomic DNA from 38 skin tissue samples taken from living beluga whales. Data for the 23 clock CpG sites were used as input into the clock model yielding ages from approximately 11 to 27 years old, with potential range of ±2.9 years, using the LOOCV median age error of the calibration dataset (Figure 3; Table S6). The lower end of the estimated age distribution is consistent with the field practice of only sampling whales that are large juveniles or older. Additionally, epigenetic ages are in agreement with broad color categories that can be used to determine age classes of younger whales before they are entirely white (Wade, unpublished data; Figure S1).

4 | DISCUSSION

This study reports a robust epigenetic clock for beluga whales, enabling age estimation of living whales with just a small piece of skin tissue. The beluga epigenetic clock is based on 23 CpG sites that were selected from 37,491 CpG probes on a custom mammalian methylation array. Age estimation based on the multivariate age estimation model greatly outperforms age estimation based on a single CpG, which is consistent with what has been observed in other mammalian species including humans (Bocklandt et al., 2011; Hannum et al., 2013; Horvath, 2013; Thompson et al., 2018). The leave one out cross-validation (LOOCV) analysis suggests that this beluga epigenetic clock estimates age with a median absolute error of 2.9 years for samples of unknown age. Future independent test data are needed to fully validate the applicability of the clock to new datasets. Beluga whale longevity has been estimated to be 60 or 70 years (Burns & Seaman, 1986; Suydam, 2009), which would mean the clock approximates age within ±5% of the beluga lifespan.

resulted in a training set estimate of $R^2 = 0.92$ (Figure 2a; other stats in Table 3). However, this training set estimate of the predictive accuracy is biased. To arrive at an estimate of the accuracy that is less biased by the nature of the training data, we employed leave one out cross-validation (LOOCV). The LOOCV run for an alpha of 0.9, which is intended to approximate the model's performance on unknown data, had an R^2 value of 0.74, a mean absolute age error of 3.65 years, and a median age error of 2.87 years (Figure 2b). The absolute age error for each sample in the LOOCV (the difference between the LOOCV predicted age and the estimated chronological age) showed a trend of underestimating the age of old whales while overestimating the age of young whales (regression slope = 0.65; regression y-intercept = 7.22).

The R^2 values for regressions between model predicted age and chronological age were consistently, but only slightly, higher for females than males in both the full clock model and LOOCV. None of the CpGs used in the final clock showed a sex correlation of 0.5 or greater.
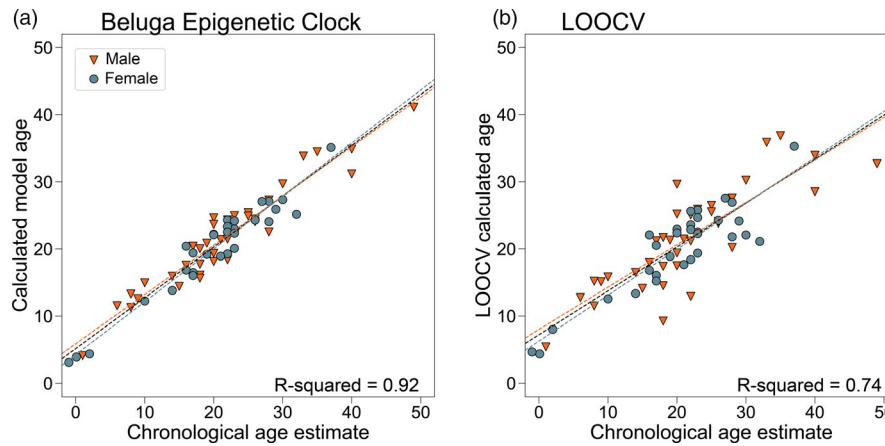
FIGURE 2 (a) Epigenetic ages calculated using the beluga epigenetic clock regressed against estimated chronological ages (based on GLG) for the calibration dataset. Data for males are represented by orange triangles and those for females are represented by gray-blue circles. Sex-specific regression lines as well as the overall regression line are shown (orange dashed line for males, gray/blue dashed line for females, black dashed line for overall regression). The training data showed an overall $R^2 = 0.92$ ($p = 3.50e^{-38}$). See Table 3 for all other statistics and sex-specific values. (b) Leave one out cross-validation (LOOCV) of the *cv.glmnet()* model parameters ($\alpha = 0.7$, *lambada.min*) with the same color scheme for males and females as panel A. Overall LOOCV $R^2 = 0.74$ ($p = 1.14e^{-20}$). See Table 3 for all other statistics and sex-specific values

TABLE 3 Statistics for the beluga epigenetic clock model and leave one out cross-validation for $\alpha = 0.9$: mean age error (mae), median age error (medae), *r*-squared for the regression, *p*-value for the regression, regression slope, and *y*-intercept for the regression

Sex	mae	medae	R^2	p-Value	Regression slope	y-Intercept
Beluga epigenetic clock						
All	2.34	1.97	0.92	3.50E−38	0.76	5.17
m	2.64	1.96	0.92	4.31E−20	0.73	5.89
f	1.99	2.06	0.94	6.33E−19	0.79	4.25
Leave one out cross-validation						
All	3.65	2.87	0.74	1.14E−20	0.65	7.22
m	4.16	2.98	0.70	2.35E−10	0.63	8.01
f	3.06	2.56	0.81	7.58E−12	0.68	6.27

The LOOCV showed a pattern of age underestimation for older whales and overestimation of younger whales, a pattern that is present but less pronounced in the training dataset. This pattern could partially be driven by data scarcity of older and younger whales in our calibration dataset, but it is also observed in other epigenetic clocks (Beal et al., 2019; Polanowski et al., 2014). Future research is needed to clarify the clock's accuracy at the two ends of the age distribution. The young samples in our calibration dataset—one fetus and three calves—may have unique epigenetic changes occurring due to developmental processes or stress. Special considerations may be needed for fetuses and neonates from stranded mothers or still births.

It can be advantageous to carry out a nonlinear transformation of age (e.g., a log transformation) to account for faster epigenetic changes occurring during development (Hannum et al., 2013; Horvath, 2013). Here, we did not carry out a nonlinear age transformation because we found no evidence that it would improve the model fit in our data. This approach is consistent with other studies that directly regressed age on the CpGs (Polanowski et al., 2014; Thompson et al., 2017, 2018; Wright et al., 2018).

Future work that combines epigenetic methods presented here with research describing the relationship between morphometric features of beluga whale calves and age could offer a line of research that would improve clock performance for very young whales and fetuses (Robeck et al., 2015; Shelden et al., 2019). The development of alternative clock models could better describe variation in the rate of aging with life phase.

We found it important to prefilter the normalized CpG data based on detection *p*-value. Without any detection *p*-value filtering, we developed a clock with 59 CpGs ($\alpha = 0.5$). While this clock led to an excellent predictive accuracy for chronological age as measured by R^2 values (Table S2), we found that several of the underlying probes did not align to a CpG in the beluga genome. To alleviate concerns about overfitting, we built the final beluga clock (based on 23 CpGs) using only CpGs in a filtered dataset that required a significant detection *p*-value in at least 10 individuals.

Methylation data were also used to impute the sex of beluga whales. Many sex-associated sites are located on the X chromosome in humans, but those that were not could be the focus of future study on sexual dimorphism in methylation patterns. Sex-based
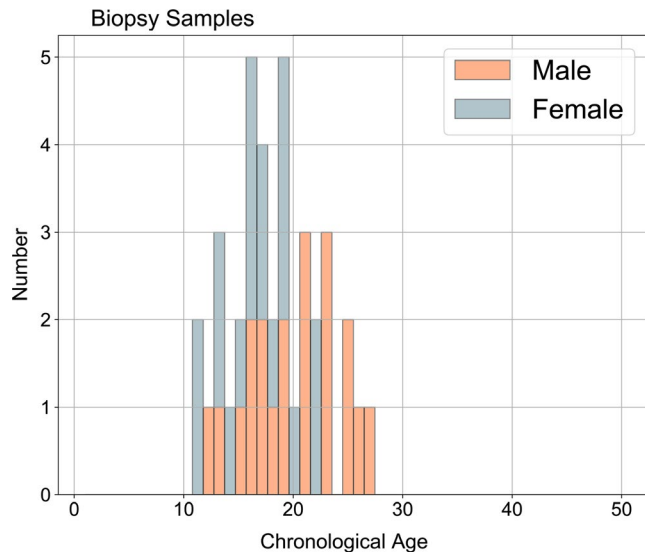
FIGURE 3 The estimated epigenetic ages for biopsy samples collected from living beluga whales between 2016 and 2018. Ages ranged from approximately 11 to 27. Note, this is not a representative age distribution of the populations due to bias in individuals biopsied during field seasons (e.g., younger whales not sampled with biopsy dart)

correlation analyses allowed us to assess the possibility that some CpGs were included in the clock due to sex-based patterns instead of age alone. This was not the case: None of the 23 CpGs in the clock had a Pearson correlation of more than 0.5 with sex. Furthermore, because there was no substantial sex-based difference in clock performance, our results support the use of this epigenetic clock for both sexes. While sexual dimorphism in morphology and behavior can be observed in beluga whales (Hauser et al., 2017), we found no indication that sex-specific clocks are required for estimation of chronological age. Furthermore, using one set of CpGs for both sexes will enable the development of an accessible age and sex assay in the laboratory by sequencing just 24 genomic regions (the 23 clock CpGs and the one sex-predictor CpG).

The sequencing of the beluga whale genome (Jones et al., 2017) increased the capacity for molecular research on this nonmodel species, enabling us to identify and map the 23 CpG sites in the beluga epigenetic clock. The 23 sites in the beluga epigenetic clock are found in genes related to critical biological activity like transcription, metabolism, and cell membranes. The function and mechanistic relationship of these genes with age is an open question. Clock development using an array, instead of targeted candidate genes, allows for comparison of important clock sites across mammal species without predisposing researchers to use just a handful of genes that have shown some relationship in the past.

The beluga epigenetic clock was successfully applied to skin tissue samples collected with a biopsy dart from living whales in Cook Inlet, Alaska. We present data from 38 skin tissue samples, but photo ID evaluation after the field season has indicated that three of these samples may be from the same individual (MML_RA180909_B01, MML_RA180910_B04, and MML_RA180912_B02). The estimated

epigenetic ages for those three potential repeat samples are 26, 28, and 26; and the samples are all male, so the results perhaps support the photo identification results that this may be a recaptured whale. Genotyping is the best method to ultimately confirm. Age from biopsy samples will be useful in contributing to many different studies related to the conservation and management of beluga whales. With further development, it may be possible to partition CpG sites that correlate with chronological age from those that reflect biological age. Whereas "chronological age" is important for demographics parameters, "biological age" could be used to investigate the numerous physiological changes associated with aging (De Paoli-Iseppi et al., 2017; Horvath & Raj, 2018). In some populations or individuals, this biological aging is accelerated due to stress and exposure to environmental contaminants (a concept known as accelerated epigenetic aging). Future studies should explore whether epigenetic age acceleration can be observed in different whale populations, potentially reflecting genetic differences or various stress factors.

Research that compares epigenetic aging of Cook Inlet beluga whales with other populations will inform the applicability of this epigenetic clock to circumpolar populations of beluga whales. Data from other populations of beluga whales could also improve the accuracy of the beluga epigenetic clock by increasing sample size (the most accurate human clocks were trained on thousands of samples, e.g., Horvath, 2013) and help to mitigate error from chronological age estimates based on tooth growth layer groups. Tooth aging is subjected to unknown error: beluga teeth wear with age at a rate that has not been quantified and is possibly individual-specific (Vos et al., 2019). Our analysis critically relies on the assumption that GLG patterns are well calibrated when it comes to estimating the chronological age of beluga whales. Beyond other beluga populations, this research also facilitates phylogenic comparisons of epigenetic clock CpGs with other cetacean species. Using a methylation array and machine learning to develop clocks will enable interspecific comparisons of age-relevant methylation patterns, potentially improving our understanding of the evolutionary function of age-correlated methylation.

Center (Kelly Robertson), Northwest Fisheries Science Center (Kim Parsons), Group for Research and Education on Marine Mammals (Robert Michaud, Michel Moisan), Alaska SeaLife Center, Alaska Veterinary Pathology Services (Kathy Burek-Huntington), Joint Base Elmendorf Richardson (Christopher Garner), and the Cook Inlet Beluga Whale Photo-ID Project (Tamara McGuire) provided personnel, expertise, and/or data. The views implied or expressed here are those of the authors and do not necessarily represent the views of the National Marine Fisheries Service, NOAA. Reference in this document to trade names does not imply endorsement by the National Marine Fisheries Service, NOAA.

CONFLICT OF INTEREST

SH is a founder of the non-profit Epigenetic Clock Development Foundation which plans to license several of his patents from his employer UC Regents. The other authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Analysis scripts can be accessed on the author's GitHub (https://github.com/ekbors/belugas). Raw methylation data are available in the NCBI GEO database, GSE164465, "Genome Methylation in Wild Beluga Whales." The manifest for the Methylation Array is available at Gene Expression Omnibus (GPL28271: Illumina HorvathMammalianMethylChip40 BeadChip). Processed data (CSV files) that have been normalized have also been archived in DataONE under NPRB project #1723.

ORCID

Eleanor K. Bors https://orcid.org/0000-0003-1751-1467
C. Scott Baker https://orcid.org/0000-0003-2183-2036
Paul R. Wade https://orcid.org/0000-0003-2428-9323
Kim E. W. Shelden https://orcid.org/0000-0002-9831-0933
Simon Jarman https://orcid.org/0000-0002-0792-9686
Steve Horvath https://orcid.org/0000-0002-4110-3589

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Arneson, A., Haghani, A., Thompson, M. J., Pellegrini, M., Kwon, S. B., Vu, H. T., Li, C. Z., Lu, A. K., Barnes, B., Hansen, K. D., Zhou, W., Breeze, C., Ernst, J., & Horvath, S. (2021). A mammalian methylation array for profiling methylation levels at conserved sequences. *bioRxiv* 2021.01.07.425637. https://doi.org/10.1101/2021.01.07.425637

Baker, C. S., Slade, R. W., Bannister, J. L., Abernethy, R. B., Weinrich, M. T., Lien, J., Urban, J., Corkeron, P., Calambokidis, J., Vasquez, O., & Palumbi, S. R. (1994). Hierarchical structure of mitochondrial DNA gene flow among humpback whales *Megaptera novaeangliae*, worldwide. *Molecular Ecology*, 3, 313–327. https://doi.org/10.1111/j.1365-294X.1994.tb00071.x

Beal, A. P., Kiszka, J. J., Wells, R. S., & Eirin-Lopez, J. M. (2019). The Bottlenose Dolphin Epigenetic Aging Tool (BEAT): A molecular age estimation tool for small cetaceans. *Frontiers in Marine Science*, 6, 561. https://doi.org/10.3389/fmars.2019.00561

Bocklandt, S., Lin, W., Sehl, M. E., Sánchez, F. J., Sinsheimer, J. S., Horvath, S., & Vilain, E. (2011). Epigenetic predictor of age. *PLoS One*, 6, e14821. https://doi.org/10.1371/journal.pone.0014821

Brault, S., & Caswell, H. (1993). Pod-specific demography of killer whales (*Orcinus Orca*). *Ecology*, 74, 1444–1454.

Burns, J. J., & Seaman, G. A. (1986). Investigations of belukha whales in coastal waters of western and northern Alaska. II. Biology and ecology. *U.S. Department of Commerce, NOAA, OCSEAP Final Report*, 56(1988), 221–357.

De Paoli-Iseppi, R., Deagle, B. E., McMahon, C. R., Hindell, M. A., Dickinson, J. L., & Jarman, S. N. (2017). Measuring animal age with DNA methylation: From humans to wild animals. *Frontiers in Genetics*, 8, 106. https://doi.org/10.3389/fgene.2017.00106

Du, P., Zhang, X., Huang, C. C., Jafari, N., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, 11, 587. https://doi.org/10.1186/1471-2105-11-587

Dunshea, G., Duffield, D., Gales, N., Hindell, M., Wells, R. S., & Jarman, S. N. (2011). Telomeres as age markers in vertebrate molecular ecology. *Molecular Ecology Resources*, 11, 225–235. https://doi.org/10.1111/j.1755-0998.2010.02976.x

Field, A. E., Robertson, N. A., Wang, T., Havas, A., Ideker, T., & Adams, P. D. (2018). DNA methylation clocks in aging: Categories, causes, and consequences. *Molecular Cell*, 71, 882–895. https://doi.org/10.1016/j.molcel.2018.08.008

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularized paths for generalized linear models via coordinate descent (Technical Report). *Journal of Statistical Software*, 33(1), 1–22.

Garagnani, P., Bacalini, M. G., Pirazzini, C., Gori, D., Giuliani, C., Mari, D., Di Blasio, A. M., Gentilini, D., Vitale, G., Collino, S., Rezzi, S., Castellani, G., Capri, M., Salvioli, S., & Franceschi, C. (2012). Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*, 11, 1132–1134. https://doi.org/10.1111/acel.12005

Gilson, A., Syvanen, M., Levine, K., & Banks, J. (1998). Deer gender determination by polymerase chain reaction: Validation study and application to tissues, bloodstains, and hair forensic samples from California. *California Fish and Game*, 84, 159–169.

Hamilton, V., & Evans, K. (2018). Establishing growth chronologies from marine mammal teeth: A method applicable across species. *Journal of Experimental Marine Ecology*, 505, 24–34. https://doi.org/10.1016/j.jembe.2018.04.006

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S. V., 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, 359–367.

Hauser, D. D. W., Laidre, K. L., Stern, H. L., Moore, S. E., Suydam, R. S., & Richard, P. R. (2017). Habitat selection by two beluga whale populations in the Chukchi and Beaufort seas. *PLoS One*, 12, e0172755. https://doi.org/10.1371/journal.pone.0172755

Hobbs, R. C., Rugh, D. J., & DeMaster, D. P. (2000). Abundance of beluga whales, Delphinapterus leucas, in Cook Inlet, Alaska, 1994–2000. *Marine Fisheries Review*, 62(3), 37–45.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, 14, 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, 371–384. https://doi.org/10.1038/s41576-018-0004-3

Ito, G., Yoshimura, K., & Momoi, Y. (2017). Analysis of DNA methylation of potential age-related methylation sites in canine peripheral blood leukocytes. *Journal of Veterinary Medical Science*, 79, 745–750. https://doi.org/10.1292/jvms.16-0341

Ito, H., Udono, T., Hirata, S., & Inoue-Murayama, M. (2018). Estimation of chimpanzee age based on DNA methylation. *Scientific Reports*, *8*, 9998. https://doi.org/10.1038/s41598-018-28318-9

Jarman, S. N., Polanowski, A. M., Faux, C. E., Robbins, J., De Paoli-Iseppi, R., Bravington, M., & Deagle, B. E. (2015). Molecular biomarkers for chronological age in animal ecology. *Molecular Ecology*, *24*, 4826–4847. https://doi.org/10.1111/mec.13357

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: A better web interface. *Nucleic Acids Research*, *36*, 5–9.

Jones, S. J. M., Taylor, G. A., Chan, S., Warren, R. L., Hammond, S. A., Bilobram, S., Mordecai, G., Suttle, C. A., Miller, K. M., Schulze, A., Chan, A. M., Jones, S. J., Tse, K., Li, I., Cheung, D., Mungall, K. L., Choo, C., Ally, A., Dhalla, N., … Haulena, M. (2017). The genome of the beluga whale (*Delphinapterus leucas*). *Genes*, *8*, 378. https://doi.org/10.3390/genes8120378

Jylhävä, J., Pedersen, N. L., & Hägg, S. (2017). Biological age predictors. *EBioMedicine*, *21*, 29–36. https://doi.org/10.1016/j.ebiom.2017.03.046

Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Bandinelli, S., Hou, L., Baccarelli, A. A., Stewart, J. D., Li, Y., Whitsel, E. A., Wilson, J. G., Reiner, A. P., Aviv, A., Lohman, K., Liu, Y., Ferrucci, L., & Horvath, S. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging*, *10*, 573–591.

Lin, Q., Weidner, C. I., Costa, I. G., Marioni, R. E., Ferreira, M. R. P., Deary, I. J., & Wagner, W. (2016). DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy. *Aging*, *8*, 394–401.

Lockyer, C., Hohn, A. A., Doidge, W. D., Heide-Jørgensen, M. P., & Suydam, R. (2007). Age determination in belugas *Delphinapterus leucas*: A quest for validation of dentinal layering. *Aquatic Mammals*, *33*(3), 293–304. https://doi.org/10.1578/am.33.3.2007.293

Lowe, R., Barton, C., Jenkins, C. A., Ernst, C., Forman, O., Fernandez-Twinn, D. S., Bock, C., Rossiter, S. J., Faulkes, C. G., Ozanne, S. E., Walter, L., Odom, D. T., Mellersh, C., & Rakyan, V. K. (2018). Ageing-associated DNA methylation dynamics are a molecular readout of lifespan variation among mammalian species. *Genome Biology*, *19*, 22. https://doi.org/10.1186/s13059-018-1397-1

Mahoney, B. A., & Shelden, K. E. W. (2000). Harvest history of belugas, *Delphinapterus leucas*, in Cook Inlet, Alaska. *Marine Fisheries Review*, *62*(3), 124–133.

McGuire, T., Michaud, R., Moisan, M., & Garner, C. (2017). *Cook Inlet beluga whale biopsy: Field report for 2016 feasibility study*. Prepared by LGL Alaska Research Associates, Inc., GREMM, and JBER for NMFS (67 pp.). Retrieved from https://www.cookinletbelugas.com/reports-and-publications

McGuire, T., Stephens, A., & McClung, J. (2018). Photo-identification of beluga whales in Cook Inlet, Alaska: Inclusion of biopsy and hexa-copter photographs from 2017. Report prepared by The Cook Inlet Beluga Whale Photo-ID Project, Anchorage, AK, for National Marine Fisheries Service, Marine Mammal Laboratory (91 pp.). Retrieved from https://www.cookinletbelugas.com/reports-and-publications

McGuire, T., Stephens, A., Michaud, R., Moisan, M., & Garner, C. (2017). *Cook Inlet beluga whale biopsy: Photo-identification of biopsied whales during the 2016 feasibility study*. Report prepared by LGL Alaska Research Associates, Inc., GREMM, and JBER for NMFS (33 pp.). Retrieved from https://www.cookinletbelugas.com/reports-and-publications

McKinney, W. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th Python Science Conference*. https://doi.org/10.3828/ajfs.41.3.62

Meer, M. V., Podolskiy, D. I., Tyshkovskiy, A., & Gladyshev, V. N. (2018). A whole lifespan mouse multi-tissue DNA methylation clock. *eLife*, *7*, e40675. https://doi.org/10.7554/eLife.40675

Moore, J. E., & Read, A. J. (2008). A Bayesian uncertainty analysis of cetacean demography and bycatch mortality using age-at-death data. *Ecological Applications*, *18*, 1914–1931. https://doi.org/10.1890/07-0862.1

National Marine Fisheries Service (2016). *Recovery plan for the Cook Inlet beluga whale (*Delphinapterus leucas*)*. National Marine Fisheries Service, Alaska Region, Protected Resources Division, Juneau, AK.

O'Corry-Crowe, G. M., Suydam, R. S., Rosenberg, A., Frost, K. J., & Dizon, A. E. (1997). Phylogeography, population structure and dispersal patterns of the beluga whale *Delphinapterus leucas* in the western Nearctic revealed by mitochondrial DNA. *Molecular Ecology*, *6*(10), 955–970. https://doi.org/10.1046/j.1365-294X.1997.00267.x

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*, D733–D745. https://doi.org/10.1093/nar/gkv1189

Olavarría, C., Baker, C. S., Garrigue, C., Poole, M., Hauser, N., Caballero, S., Flórez-González, L., Brasseur, M., Bannister, J., Capella, J., Clapham, P., Dodemont, R., Donoghue, M., Jenner, C., Jenner, M. N., Moro, D., Oremus, M., Paton, D., Rosenbaum, H., & Russell, K. (2007). Population structure of South Pacific humpback whales and the origin of the eastern Polynesian breeding grounds. *Marine Ecology Progress Series*, *330*, 257–268.

Oliphant, T. E. (2006). *A guide to NumPy*. Trelgol Publishing.

Olsen, M. T., Bérubé, M., Robbins, J., & Palsbøll, P. J. (2012). Empirical evaluation of humpback whale telomere length estimates; quality control and factors causing variability in the singleplex and multiplex qPCR methods. *BMC Genetics*, *13*, 77. https://doi.org/10.1186/1471-2156-13-77

Pennisi, E. (2001). Behind the Scenes of Gene Expression. *Science*, *293*(5532), 1064–1067. https://doi.org/10.1126/science.293.5532.1064

Perrin, W. F., & Myrick, A. C. (1980). *Age determination of toothed whales and sirenians*. International Whaling Commission.

Petkovich, D. A., Podolskiy, D. I., Lobanov, A. V., Lee, S. G., Miller, R. A., & Gladyshev, V. M. (2017). Using DNA methylation profiling to evaluate biological age and longevity interventions. *Cell Metabolism*, *25*(4), 954–960.

Polanowski, A. M., Robbins, J., Chandler, D., & Jarman, S. N. (2014). Epigenetic estimation of age in humpback whales. *Molecular Ecology Resources*, *14*, 976–987. https://doi.org/10.1111/1755-0998.12247

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Razin, A., & Cedar, H. (1991). DNA methylation and gene expression. *Microbiology Reviews*, *55*(3), 451–458.

Robeck, T. R., Schmitt, T. L., & Osborn, S. (2015). Development of predictive models for determining fetal age-at-length in belugas (*Delphinapterus leucas*) and their application toward in situ and ex situ population management. *Marine Mammal Science*, *31*, 591–611.

Shelden, K. E. W., Goetz, K. T., Rugh, D. J., Calkins, D. G., Mahoney, B. A., & Hobbs, R. C. (2015). Spatio-temporal changes in beluga whale, *Delphinapterus leucas*, distribution: Results from aerial surveys (1977–2014), opportunistic sightings (1975–2014), and satellite tagging (1999–2003) in Cook Inlet, Alaska. *Marine Fisheries Review*, *77*, 1–31. https://doi.org/10.7755/MFR.77.2.1

Shelden, K. E. W., Robeck, T. R., Goertz, C. E. C., McGuire, T. L., Burek-Huntington, K. A., Vos, D. J., & Mahoney, B. A. (2019). Breeding and calving seasonality in the endangered Cook Inlet beluga whale population: Application of captive fetal growth curves to fetuses and newborns in the wild. *Marine Mammal Science*, *36*(2), 700–708.

Stubbs, T. M., Bonder, M. J., Stark, A. K., Krueger, F., von Meyenn, F., Stegle, O., Reik, W., Bolland, D., Butcher, G., Chandra, T., Clark, S. J., Corcoran, A., Eckersley-Maslinc, M., Field, L., Frising, U. C., Gilbert, C., Guedes, J., Hernando-Herraez, I., Houseley, J., … Veldhoenb, M. (2017). Multi-tissue DNA methylation age predictor in mouse. *Genome Biology*, *18*, 68. https://doi.org/10.1186/s13059-017-1203-5

Suydam, R. S. (2009). *Age, growth, reproduction, and movements of beluga whales (*Delphinapterus leucas*) from the eastern Chukchi Sea* (170 pp.). Ph.D. dissertation, University of Washington.

Tanabe, A., Shimizu, R., Osawa, Y., Suzuki, M., Ito, S., Goto, M., Pastene, L. A., Fujise, Y., & Sahara, H. (2020). Age estimation by DNA methylation in the Antarctic minke whale. *Fisheries Science*, *86*, 35–41.

Thompson, M. J., Chwia, K., Rubbi, L., Lusis, A. J., Richard, C., Srivastava, A., Korstanje, R., Churchill, G. A., Horvath, S., & Pellegrini, M. (2018). A multi-tissue full lifespan epigenetic clock for mice. *Aging*, *10*, 2832–2854.

Thompson, M. J., von Holdt, B., Horvath, S., & Pellegrini, M. (2017). An epigenetic aging clock for dogs and wolves. *Aging*, *9*, 1055–1068.

Venuto, R., Botta, S., Barreto, A. S., Secchi, E. R., & Fruet, P. F. (2020). Age structure of strandings and growth of Lahille's bottlenose dolphin (*Tursiops truncatus gephyreus*). *Marine Mammal Science*, *36*(3), 813–827. https://doi.org/10.1111/mms.12683

Vos, D. J., Shelden, K. E. W., Friday, N. A., & Mahoney, B. A. (2019). Age and growth analyses for the endangered belugas in Cook Inlet, Alaska. *Marine Mammal Science*, *36*(1), 293–304.

Wade, P. R., Boyd, C., Shelden, K. E. W., & Sims, C. L. (2019). Group size estimates and revised abundance estimates and trend for the Cook Inlet beluga population. In K. E. W. Shelden, & P. R. Wade (Eds.), *Aerial surveys, distribution, abundance, and trend of belugas (*Delphinapterus leucas*) in Cook Inlet, Alaska, June 2018* (pp. 53–89) AFSC Processed Rep. 2019-09. NOAA.

Waugh, D. A., Suydam, R. S., Ortiz, J. D., & Thewissen, J. G. M. (2018). Validation of Growth Layer Group (GLG) depositional rate using daily incremental growth lines in the dentin of beluga (*Delphinapterus leucas* (Pallas, 1776)). *PLoS One*, *13*(1), e0190498. https://doi.org/10.1371/journal.pone.0190498

Wright, P. G. R., Mathews, F., Schofield, H., Morris, C., Burrage, J., Smith, A., Dempster, E. L., & Hamilton, P. B. (2018). Application of a novel molecular method to age free-living wild Bechstein's bats. *Molecular Ecology Resources*, *18*, 1374–1380. https://doi.org/10.1111/1755-0998.12925

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*, *46*, e123. https://doi.org/10.1093/nar/gky691

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.