Received: 17 May 2023 | Revised: 29 December 2023 | Accepted: 3 January 2024

DOI: 10.1111/1755-0998.13928

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

A cost-effective blood DNA methylation-based age estimation method in domestic cats, Tsushima leopard cats (*Prionailurus bengalensis euptilurus*) and *Panthera* species, using targeted bisulphite sequencing and machine learning models

Huiyuan Qi[1] | Qi Luan Lim[1] | Kodzue Kinoshita[1] | Nobuyoshi Nakajima[2] | Miho Inoue-Murayama[1]

[1]Wildlife Research Center, Kyoto University, Kyoto, Japan

[2]Biodiversity Division, National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan

Correspondence
Miho Inoue-Murayama, Wildlife Research Center, Kyoto University, Kyoto 606-8203, Japan.
Email: murayama.miho.5n@kyoto-u.ac.jp

Present address
Qi Luan Lim, Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Japan
Kodzue Kinoshita, Graduate School of Asian and African Area Studies, Kyoto University, Kyoto, Japan

Funding information
Environmental Restoration and Conservation Agency, Grant/Award Number: JPMEERF20214001; Japan Society for the Promotion of Science, Grant/Award Number: 20H00420 and 20H03008

Handling Editor: Maren Wellenreuther

Abstract

Individual age can be used to design more efficient and suitable management plans in both in situ and ex situ conservation programmes for targeted wildlife species. DNA methylation is a promising marker of epigenetic ageing that can accurately estimate age from small amounts of biological material, which can be collected in a minimally invasive manner. In this study, we sequenced five targeted genetic regions and used 8–23 selected CpG sites to build age estimation models using machine learning methods at only about $3–7 per sample. Blood samples of seven Felidae species were used, ranging from small to big, and domestic to endangered species: domestic cats (*Felis catus*, 139 samples), Tsushima leopard cats (*Prionailurus bengalensis euptilurus*, 84 samples) and five *Panthera* species (96 samples). The models achieved satisfactory accuracy, with the mean absolute error of the most accurate models recorded at 1.966, 1.348 and 1.552 years in domestic cats, Tsushima leopard cats and *Panthera* spp. respectively. We developed the models in domestic cats and Tsushima leopard cats, which were applicable to individuals regardless of health conditions; therefore, these models are applicable to samples collected from individuals with diverse characteristics, which is often the case in conservation. We also showed the possibility of developing universal age estimation models for the five *Panthera* spp. using only two of the five genetic regions. We do not recommend building a common age estimation model for all the target species using our markers, because of the degraded performance of models that included all species.

KEYWORDS
age estimation, DNA methylation, domestic cat, endangered Felidae species, *Panthera*, Tsushima leopard cat

1 | INTRODUCTION

Age information is important for both ex situ and in situ wildlife conservation, as it is relevant to individual behaviour, health, reproductive capacity and mortality (Blomqvist & Sten, 1982; Kirkwood & Austad, 2000; Nussey et al., 2013; Youn et al., 2022; Zhao et al., 2019). Age estimation can help determine demographic characteristics and predict current and future extinction risks for wildlife populations (Lacy, 2019; Oli & Dobson, 2003). Age estimation in injured or dead wild individuals also helps examine the relationships between age and the causes of injury or mortality (Conroy et al., 2019; Thorel et al., 2020); when certain age groups are found to have particularly high mortality rates, the causes of mortality can be deduced and preventive measures taken. Knowing the age of rescued wild individuals would help inform appropriate health care and enrichment activities that maximize the welfare of individuals of different age classes and thereby improve the efficiency of a breeding programme (Caselli et al., 2022; Eskelinen et al., 2015; Hecht, 2021).

Age estimation methods have been conducted based on morphological observations or measurements and mark–recapture, such as longtime tracking and direct observation in primates (Mori, 1979), mark–recapture in bats (Wilkinson & Brunet-Rossinni, 2009) and scar/speckle-counting in cetaceans (Hartman et al., 2016; Yagi et al., 2023). However, these methods are difficult to implement for species that are hard to observe or recapture or do not show prominent age-related changes in appearance. Age estimation of mammals can also be done via measuring the development and eruption of teeth and bones (Chevallier et al., 2017; White et al., 2016), but this requires either carcasses or long-term restraint of captured live animals.

Molecular ageing markers have been highlighted as new, less invasive age estimation tools that can determine individuals' ages by sampling and analysing small amounts of biological materials (Bocklandt et al., 2011; Gruber et al., 2021; Petkovich et al., 2017; Xia et al., 2017). DNA methylation is one of the most accurate age markers (Horvath, 2013; Li et al., 2018; Paoli-Iseppi et al., 2017; Stubbs et al., 2017). It is an epigenetic process in which 5-methylcytosine is formed via transfer of a methyl group, usually onto the C5 position of cytosine in the cytosine–guanine dinucleotide (CpG) sites in mammals (Bogdanović & Veenstra, 2009). Recently, epigenetic clocks have been developed based on the mammalian DNA methylation array HorvathMammalMethylChip40 (Arneson et al., 2022), which provides more than 37,000 highly conserved CpGs with high coverage; they have been used to accurately estimate age in mammalian species such as plains zebras (*Equus quagga*) (Larison et al., 2021), roe deer (*Capreolus capreolus*) (Lemaître et al., 2022), beluga whales (*Delphinapterus leucas*) (Bors et al., 2021) and naked mole-rats (*Heterocephalus glaber*) (Horvath et al., 2022). However, high costs (approximately $160/sample), requirement of large quantities of DNA (ideally more than 250 ng/sample) and relatively complex data processing still limit its wide application to conservation projects, especially those with limited budgets and few good-quality samples. Such projects will benefit more from a study design that uses only a few target genes selected from previous studies of related species.

Felids have received much conservation attention, however, few studies have focused on improving their age assessment for conservation implications. Recent publications for tigers (*Panthera tigris*) and lions (*P. leo*) were still based on teeth measurement (Sharma et al., 2022; White et al., 2016). DNA methylation-based age estimation that can be conveniently implemented for live felid individuals is still lacking, except for two studies focusing primarily on domestic cats (*Felis catus*) (Qi et al., 2021; Raj et al., 2021). Previously (Qi et al., 2021), we estimated domestic cat age with mean absolute error (MAE) at 3.83 years using a cost-effective RT-PCR-based method based on two gene regions. We also tried to build a pilot age estimation model for snow leopards ($n = 11$), and the MAE was 2.10 years. Raj et al. (2021) successfully developed high-accuracy models for domestic cats using the HorvathMammalMethylChip40, achieving MAE = 0.79 years, the most accurate value achieved by an epigenetic clock for cats, with 34 CpGs from about 10 gene regions (the gene regions used in the clock were not specified). The domestic cat clock developed by Raj et al. (2021) was also tested on some samples of other felid species: MAE = 1.65, 1.41 and 3.01 years for cheetahs (*Acinonyx jubatus*, $n = 14$), lions ($n = 7$) and tigers ($n = 8$), respectively. Although the HorvathMammalMethylChip40 provided accurate values, developing low-cost epigenetic clocks can facilitate wider application in conservation. Additionally, a larger sample size of other felid species will not only help develop more convincing epigenetic clocks for each species but also allow us to study whether the same age estimation models could be employed across multiple felid species. Although the body size and living environments of felids are varied, the genetic distance among cat families is relatively small; felid genomes show strong collinearity and recent genetic divergence (Cho et al., 2013; Davis et al., 2009; Wurster-Hill & Centerwall, 1982). The occurrences of interspecific hybrids in big and small cats in captivity (Gray, 1972) also supported this genome collinearity. These include many *Panthera* hybrids: ligers (male lions and female tigers), tigons (male tigers and female lions), jaglions (male jaguars and female lions), tipards (male tigers and female leopards) and leopon (male leopards and female lions); and domestic cat breeds that are interspecific hybrids, such as the Bengal and Savannah.

In the present study, we aimed to construct easily applicable age estimation models for conservation applications that yield acceptable accuracies using only a few gene regions and lower-cost targeted bisulphite sequencing. We compared the differences in DNA methylation profiles of different felid species and discussed the possibility of constructing common age estimation models across multiple felid species.

Our target species comprised small and big cat species. The domestic cat (*F. catus*) and Tsushima leopard cat (*Prionailurus bengalensis euptilurus*) were the target small cat species. The domestic cat is a beloved companion and model animal for other small felines. The age estimation model developed for the domestic cat can be used as a reference for other endangered small felines for which adequate sample sizes are difficult to obtain. Alternatively, the Tsushima leopard cat is a

critically endangered species that requires age estimation for improved conversation strategies. The Tsushima leopard cat is an isolated population of the Amur leopard cat (*Pr. bengalensis euptilurus*) found only on Tsushima Island in Japan, with about 100 wild and 30 captive individuals (Inoue, 2021). Rescue and breeding programmes have begun since 1999, and over time, many samples have been collected. We also included five *Panthera* spp. that require urgent conservation—jaguars (*Panthera onca*), Amur leopards (*P. pardus orientalis*), African lions (*P. leo leo*), snow leopards (*P. uncia*) and Amur tigers (*P. tigris altaica*).

In the models, we also assessed the influence of sex and health conditions, which may contribute to epigenetic ageing (Bors et al., 2021; Qi et al., 2021). Age estimation models developed in previous studies focused only on healthy individuals, necessitating revalidation before the models can be applied to diseased animals, which is common in wildlife conservation. In the present study, we overcame this shortcoming by including numerous samples from diseased individuals during model building. High applicability of models is important in the conservation context because rescued individuals have varied conditions.

2 | MATERIALS AND METHODS

2.1 | Ethics statement

All methods were carried out in accordance with relevant guidelines and regulations. The study was performed in compliance with the ARRIVE guidelines. All experimental protocols were approved by the ethical committee of Wildlife Research Center of Kyoto University, and all sample collection and experiments were conducted with permission from the ethical committee (approval numbers: WRC-2019&2020-012A, WRC-2021&2022-013A and WRC-2023-010A). All domestic cat samples were obtained with the consent of the cat owners. Other Felidae species samples were collected with the approval of each zoo/conservation centre. Approval from the Ministry of the Environment Japan was also obtained for the Tsushima leopard cat samples.

2.2 | Sample collection

2.2.1 | Domestic cat samples

A total of 139 residual blood samples were obtained from clinical health check-ups of 105 domestic cats from July to September 2020 from the Kyoto Medical Center, Daktari Animal Hospital and Anicom Specialty Medical Institute, Inc. The information recorded by the veterinarians on age, breed, neuter status, sex and health condition was provided by the institutions. All domestic cat samples were stored at −80°C for less than 1 month before DNA extraction. The ages of the cats ranged between 0.41 and 21.04 years, and the female-to-male sex ratio (F:M) was 50:55 = 10:11. Most domestic cat samples were obtained from mixed-breed individuals ($n = 65$). The remaining individuals ($n = 40$) were purebred cats. As few samples were available for each cat breed ($n \leq 8$), the influence of breed was not considered. Most, except six of the 112 adult individuals over 3.5 years old were neutered, while only one out of 15 individuals under 2 years old were neutered. Because of the extreme disparity in age distributions between neutered and un-neutered individuals, similar-age individuals from the two groups could not be found. Therefore, we did not investigate the effect of neutering on the accuracy of age estimation. Health conditions and other information are provided in Table 1 and Appendix S1.

2.2.2 | Tsushima leopard cat samples

Tsushima leopard cat blood samples were collected during health checks in Tsushima Wildlife Conservation Center and zoos from 2006 to 2021. These samples were stored at −20°C for 0–1 year or −80°C for 0–15 years. We included 84 samples of known age from 19 captive-born individuals (age ranged 0.50–15.32 years; F:M = 9:10; Table 1) and 15 samples from four rescued wild-born individuals of unknown age (F:M = 3:1; Table 1). Health conditions and other detailed information can be found in Appendix S2.

2.2.3 | *Panthera* spp. samples

A total of 96 blood samples from 35 individuals (age: 0.26–23.74 years, F:M = 16:19) of *Panthera* spp. including jaguars, Amur leopards, African lions, snow leopards and Amur tigers, were collected during routine health checks in Japanese zoos from 2001 to 2022 and stored at −80°C for 0–21 years. The sample size for each species is summarized in Table 1. Health conditions and other detailed information can be found in Appendix S3. Considering the small sample sizes of each species, small genetic distances (Li et al., 2016) and similar lifespans (e.g. maximum age = 25 years in Japanese zoos) (Animal Lifespan, 2017), we merged all *Panthera* spp. samples into one dataset for all subsequent analyses.

2.3 | DNA extraction and bisulphite conversion

Genomic DNA of all samples was extracted using the DNeasy Blood & Tissue Kit (QIAGEN GmbH, Hilden, Germany), followed by bisulphite conversion using the EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's protocol.

2.4 | Gene regions, primer design and PCR conditions

We targeted five DNA regions adjacent to five genes, namely, *TCF21* (Transcription factor 21), *PRMT8* (Protein Arginine Methyltransferase 8), *DLX5* (Distal-less homeobox 5), *RALYL* (RALY RNA binding protein like) and *ELOVL2* (ELOVL fatty acid elongase 2). Previously, we found

TABLE 1 Characteristics of samples from seven felid species used to build age estimation models.

Species	Scientific name	*n* (sample)	*n* (healthy sample)	Individual	Female	Min. age (years)	Max. age (years)	Ave. age (years)
Domestic cat	*Felis catus*	139	34	105	50	0.41	21.04	10.06
Tsushima leopard cat (age-known, captive)	*Prionailurus bengalensis euptilurus*	84	47	19	9	0.50	15.32	6.53
Tsushima leopard cat (age-unknown, wild-born)		15	0	4	3	-	-	-
Panthera		96	83	35	16	0.26	23.74	9.41
Jaguar	*Panthera onca*	4	4	3	1	10.79	19.73	16.33
Amur leopard	*Panthera pardus orientalis*	34	34	8	3	0.92	17.94	13.30
African Lion	*Panthera leo leo*	10	8	5	3	0.26	23.74	8.11
Snow leopard	*Panthera uncia*	33	27	15	6	0.93	17.66	7.77
Amur tiger	*Panthera tigris altaica*	15	10	4	3	3.58	18.13	13.41

Note: *n* (sample): number of samples. *n* (healthy sample): number of samples from healthy individuals. Individual: number of individuals. Female: number of female individuals.

that *RALYL* and *ELOVL2* were significantly correlated with chronological age in domestic cats and snow leopards (Qi et al., 2021); however, the accuracy of the two-marker age estimation model can be improved (domestic cat: MAE = 3.83 years, snow leopard: MAE = 2.10 years). In the present study, we included three more candidate genes: *TCF21*, *PRMT8* and *DLX5*, which showed age-related methylation rate changes in dogs, which like felids, fall under the order Carnivora (Lowe et al., 2018). These five genes are involved in many housekeeping/essential pathways, and their abnormal gene expression levels are detected in several types of cancer tissues (Safran et al., 2021; Stelzer et al., 2016). Homogeneous gene regions were searched against the reference genomes of each target species using BLAST+ 2.11.0 (Altschul et al., 1990; Camacho et al., 2009) provided by the National Centre for Biotechnology Information (NCBI). The reference genomes that were used are as follows: domestic cat—F.catus_Fca126_mat1.0 (GCF_018350175.1), Tsushima leopard cat—Fcat_Pben_1.1_paternal_pri (GCF_016509475.1), jaguar—PanOnc_v1_BIUU (GCA_004023805.1), leopard—Panpar1.0 (GCF_001857705.1), lion – P.leo_Ple1_pat1.1 (GCF_018350215.1), snow leopard—Puncia_PCG_1.0 (GCF_023721935.1) and tiger—P.tigris_Pti1_mat1.1 (GCF_018350195.1). The target regions are located in the gene bodies of the target genes, except for *ELOVL2*, which is located in the promoter/enhancer region. Bisulphited DNA was subjected to PCR amplification with TaKaRa EpiTaq™ HS (Takara Bio, Shiga, Japan), using the primers and PCR conditions provided in Table 2 (the sequence information on NCBI for each target region and each species is listed in Table S1). The singleplex PCR reactions were prepared separately for each primer pair at a final volume of 20 μL, containing 2 μL of bisulphited DNA (5–10 ng), 0.5 U TaKara EpiTaq HS, 1× EpiTaq PCR Buffer (Mg^{2+} free), 2.5 mM $MgCl_2$, 0.3 mM dNTP mixture and 500 nM each of forward and reverse primers.

2.5 | Next-generation sequencing

NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1 and Set 2, for sample indices) and NEBNext Ultra II DNA Library Prep Kit for Illumina (for adaptor ligation and PCR enrichment) (New England Biolabs, Ipswich, MA, USA) were used to generate MiSeq-compatible barcoded DNA sequencing libraries for each gene amplicon and sample following the manufacturer's protocol. Barcoded libraries were pooled together in equimolar amounts to make a single pooled library after concentration measurement using Tapestation D1000 (Agilent, Santa Clara, CA, USA). Next-generation sequencing was performed using Illumina MiSeq with the MiSeq Reagent Kit v2 (500 cycles) (Illumina, San Diego, CA, USA).

2.6 | Methylation data organizing

Quality check and trimming of sequence data were performed using fastp v0.20.1 with base quality threshold -q set to 20. The quality-filtered sequences were aligned to the bisulphite-converted

TABLE 2 Primers and PCR conditions.

Proximal gene (target region)	Primers	Size (bp)	PCR conditions	NCBI sequence ID: position (domestic cat)	Reference
TCF21 (gene body)	F: GTTTAGGAGAGGGGTTGTG R: TAATTAAATCTTACACTCCTTTTCCC	382	95°C (3 min) (95°C [30 s], 59°C [30 s], 72°C [30 s])×40 cycles 72°C (5 min)	NC_058372.1: 120538405–120538786	Lowe et al. (2018)
PRMT8 (gene body)	F: GTTTTGGTTTTGAAGGTGGT R: ACTCAAATCAATCCCCCCC	224	95°C (3 min) (95°C [30 s], 59°C [30 s], 72°C [30 s])×40 cycles 72°C (5 min)	NC_058374.1: 38848878–38849102	Lowe et al. (2018)
DLX5 (gene body)	F: AATTAGATTTTTATTTGGGTGG R: CTCTAAAACCATTATTTTAAAAACC	384	95°C (3 min) (95°C [30 s], 55°C [30 s], 72°C [30 s])×40 cycles 72°C (5 min)	NC_058369.1: 100169877–100170260	Lowe et al. (2018)
RALYL (gene body)	F: AGTTGGAGGAGTATGGTTTT R: ACAAAAATCTTATCAAAAAAAAAA	261	95°C (3 min) (95°C [30 s], 55°C [30 s], 72°C [30 s])×40 cycles 72°C (5 min)	NC_058385.1: 32317113–32317373	Lowe et al. (2018); Qi et al. (2021)
ELOVL2 (promoter/enhancer)	F: TGTYGTYGYGGYGTTTTTTGT R: CCAAAAACRAACRACRAATCC	118	95°C (3 min) (95°C [30 s], 57°C [30 s], 72°C [30 s])×45 cycles	NC_058372.1: 17621162–17621279	Bekaert et al. (2015); Qi et al. (2021)

Note: Sequence information on the National Center for Biotechnology Information (NCBI) database for other species are summarized in Table S1.

reference genome of respective species using these parameters: --score_min L, 0, -0.4, --non_directional in Bismark v0.22.3. The bisulphite-converted reference genomes were generated with bismark_genome_preparation. The count of methylated CpG sites and total coverage was output in cov.gz files with bismark_methylation_extractor (--comprehensive) and processed using the bsseq 1.26.0. package (Hansen et al., 2012) in R 4.0.5 (R Core Team, 2021). CpG sites with coverage lower than 100 for *DLX5* and 1000 for other regions were excluded from further analysis. *DLX5* had lower coverage (ranging from 42 to 1660, mean = 406); therefore, we set a lower threshold. The most often commonly used threshold in targeted bisulphite sequencing is 1000× (Leitão et al., 2018), but 100× could also be found in some studies (Chen et al., 2017). Samples with extremely lower coverage than others (0–2 samples per species) were not counted for sample collection (as detailed in Section 2.2) and were also excluded from the analysis.

2.7 | Age estimation model

2.7.1 | Data splitting

We designated four groups of samples to be included in our analysis (Figure 1). Domestic cats, Tsushima leopard cats and *Panthera* spp. each constituted a group, and the fourth group contained samples from all the species (ALL dataset). For the ALL dataset, unlike in the other three groups, relative age was used instead of age. Relative age is individual age relative to the maximum lifespan of its species and ranges between 0 and 1 (Appendix S4) (Lu et al., 2023; Raj et al., 2021). According to the global animal age database and records of Japanese zoos, the maximum lifespans are 30, 20 and 25 years for domestic cats, Tsushima leopard cats and *Panthera* spp., respectively (AnAge: Animal Lifespan, 2017; de Magalhães et al., 2007).

The workflow of our analysis is summarized in Figure 1. The data were first split into training and test datasets. To select CpGs that were stably selected across different datasets and also evaluate model performance more comprehensively, we prepared five sets of training and test data through one-time data splitting with stratified *k*-fold ($k = 5$, similar age and species distribution across folds) using the Python package scikit-learn 1.2.0 (Pedregosa et al., 2011) with StratifiedKFold and MultilabelStratifiedKFold functions in Python 3.8.8 (Van Rossum & Drake, 2009). In rotation, each fold was used as the test data and the remaining as training data for the following procedures of model building. The methods of feature/CpG selections are described in Section 2.7.3. Finally, we evaluated model performance on an ensemble of predictions conducted across the five test datasets.

2.7.2 | Data preprocessing

CpGs with zero or near-zero variance and those that were highly correlated (Pearson correlation ≥.8) were excluded from the training

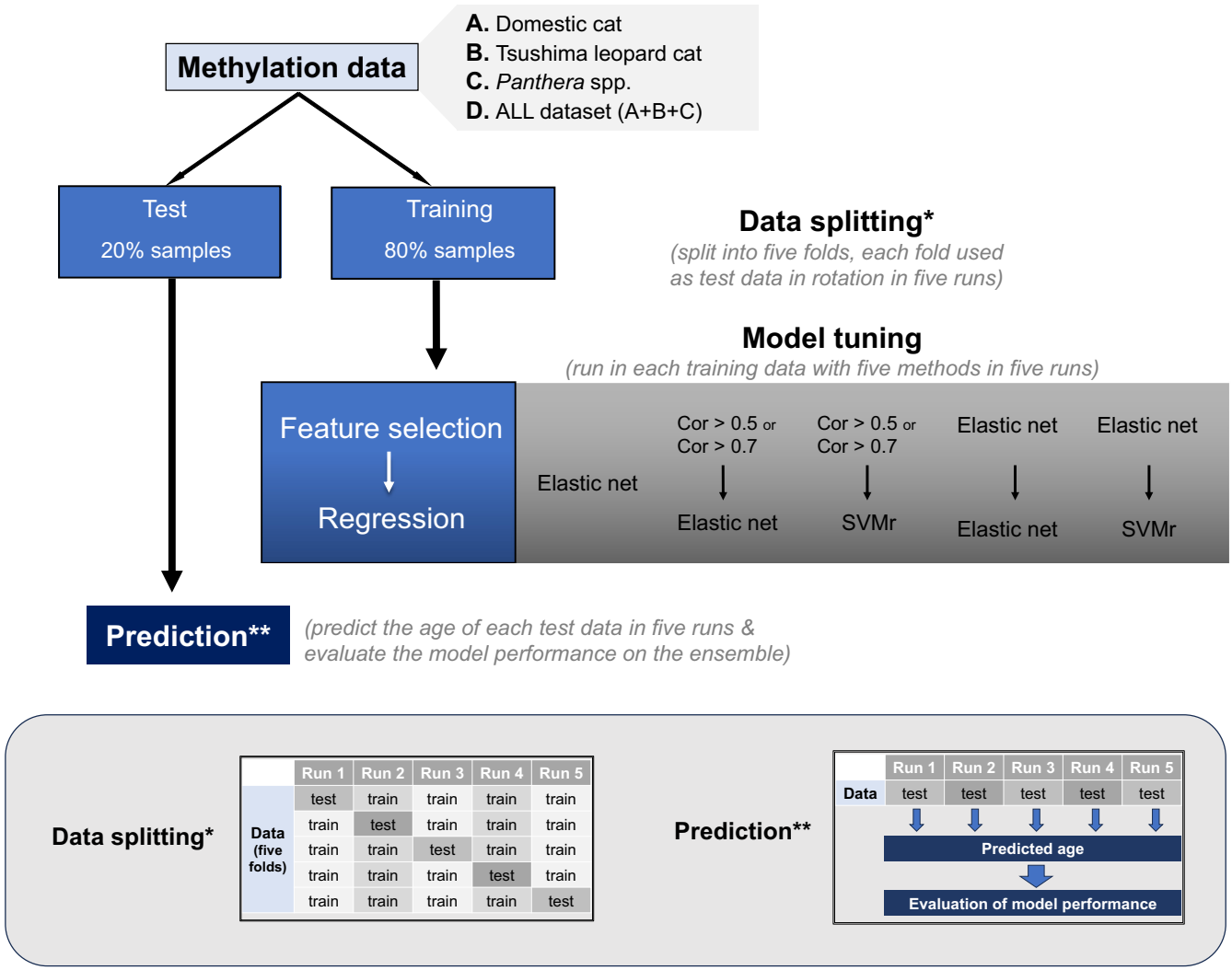

FIGURE 1 Workflow for model building. Data were first split into training (20%) and test sets (80%) in five runs, followed by model tuning using five different methods in the order of feature selection and regression. Cross-validations were conducted to find the best model in each model-tuning method in training data (leave-one-individual-out cross-validation [LOIOCV] for each study group, additional leave-one-species-out cross-validation [LOSOCV] for *Panthera* spp. and ALL dataset). Finally, predictions and model performance on test data were evaluated. SVMr, support vector machine radial.

data. Before model tuning, methylation rates of the remaining CpGs in training and test were standardized. The dataset was further processed using the R package caret 6.0–94 (Kuhn, 2008) and dplyr 1.1.3 (Hadley et al., 2023) in R 4.3.1 (R Core Team, 2021), following the method described by Anastasiadi and Piferrer (2023).

2.7.3 | Model tuning

We first used elastic net regression to tune our models, as the method is a mix of ridge and lasso regression which performs automatic feature shrinkage (i.e. feature selection) together with regression. Elastic net regression is widely adopted in many DNA methylation-based age estimation studies (Bors et al., 2021; Lu et al., 2023; Nakamura et al., 2023; Raj et al., 2021; Thompson et al., 2017; Vidaki et al., 2021).

Additionally, we conducted stepwise model tunings, that is, selecting features first and then conducting regression, to check whether model performances could be improved further. We performed feature selection via two methods. First, we adopted the most frequently used feature selection method; selecting CpGs based on correlations between their methylation and chronological age (the mean correlation of training data over five runs), with Pearson correlation coefficients over .5 or .7 as the thresholds (Figure 1). Second, we used the elastic net regression to select features before later regression, which is not frequently seen in age estimation studies but has been done in other machine learning studies (Topuz et al., 2018). We considered CpGs that were selected in the elastic net-based feature selection in over four of all five training sets (over 80%) as explanatory variables in later regression models. As shown in Figure 1, after feature selection, we created regression models with elastic net regression and support

vector machine radial (SVMr) in the training data. SVMr, similar to elastic net regression, has been found to produce high estimation accuracy (Krivonosov et al., 2022; Nakamura et al., 2023; Qi et al., 2021; Xu et al., 2015).

Model evaluation and parameter tuning based on leave-one-individual-out cross-validations (LOIOCV) were conducted in feature selection (elastic net-based) and regression. In LOIOCV, data from one individual are excluded for validation at each iteration of analysis. Additionally, leave-one-species-out cross-validations (LOSOCV) were performed for *Panthera* spp. and the ALL dataset to check species influence.

We performed elastic net regression and its hyperparameter tuning with cv.glmnet in the R package glmnet 4.1-8 (Friedman et al., 2010). Optimized alpha and lambda were determined under cross-validations based on the smallest mean absolute difference between predicted age and chronological age (MAE). For SVMr, we used the Python package scikit-learn 1.2.0 (Pedregosa et al., 2011) with GridsearchCV and SVR functions to find the best gamma in the range of 2^{-15} to 2^{6} and the best cost in the range of 2^{-5} to 2^{16}, under cross-validations based on MAE, as with elastic net regression.

2.7.4 | Age prediction and evaluation of model performance on test data

We predicted age from test data and evaluated model precision with MAE, median absolute error (MedAE), root mean square error (RMSE) and squared correlation between predicted age and chronological age (R^2). Model accuracy was evaluated using the Pearson correlation between predicted age and chronological age (r). MAE, MedAE and RMSE in ALL dataset models are measured based on relative age (percentages) and are not directly comparable with those of other study groups based on age (years).

2.8 | Factors affecting the deviation of predicted age from chronological age

To investigate what variables influenced ΔAge (difference between predicted age and chronological age) and |ΔAge| (absolute difference between predicted age and chronological age) in each group (i.e. domestic cats, Tsushima leopard cats, *Panthera* spp. and ALL dataset), we applied linear mixed models with individual ID as a random effect using the R package lmerTest 3.1.3 (Kuznetsova et al., 2017). Chronological age, sex (female and male) and health condition (healthy and diseased) were used as explanatory variables for all study groups. For groups comprising more than one species, the *Panthera* spp. and the ALL dataset, 'species' was used as an additional explanatory variable. In *Panthera* spp., the snow leopard was used as the standard in the variable 'species', because of the large sample size ($n = 33$) and relatively wide and even age distribution (minimum age = 0.93, maximum age = 17.66, average age = 7.77, Table 1). In the ALL dataset, domestic cats were used as the standard. The Akaike's information criterion (AIC) of models was used to determine whether interactions among each factor pair must be included—if including interactions made AIC smaller, then interactions were added as additional explanatory variables. Domestic cats had the largest sample size; for the variable 'health condition', we not only compared healthy ($n = 34$) versus diseased conditions ($n = 105$), but also used detailed disease categories as the explanatory variables: healthy ($n = 34$), cancer ($n = 13$), chronic kidney disease ($n = 48$), diabetes ($n = 23$), digestive disease ($n = 9$) and other diseases ($n = 33$). Note that several individuals had multiple diseases; therefore, the sum of the health categories was larger than the total sample size. The small sample sizes for each disease category in Tsushima leopard cats and *Panthera* spp. made separate analyses of disease groups statistically less plausible; therefore, we simply categorized samples into healthy and diseased—Tsushima leopard cats comprised 47 healthy and 37 diseased individuals, and *Panthera* spp. eighty-three healthy and 13 diseased individuals.

3 | RESULTS

3.1 | Correlation between methylation rates and chronological age

Over 100 CpGs were detected in domestic cats ($n = 106$), Tsushima leopard cats ($n = 108$) and *Panthera* spp. ($n = 105$). CpGs found in at least one group were listed with their NCBI position of each species in Appendix S5, of which 80.8% (97 out of 120) were found in all species. Pearson correlation coefficients between CpG methylation rate and age in the training data ranged from −0.12 to 0.77 (mean = 0.29) for domestic cats, −0.18 to 0.77 (mean = 0.23) for Tsushima leopard cats, −0.14 to 0.92 (mean = 0.43) for *Panthera* spp. and −0.04 to 0.68 (mean = 0.24) for ALL dataset.

3.2 | Age estimation model

The CpGs included in the best model determined for each study group are shown in Figure 2 (see Appendix S5 for detailed results). Overall, for all study groups, the workflow of elastic net feature selection followed by SVMr regression produced the best models. Table 3 shows the model performances evaluated by five indexes (i.e. MAE, MedAE, RMSE, R^2 and r). Model performances of all feature selection–regression methods are summarized in Table S2. The correlations between chronological age and the methylation rate of the CpGs selected by elastic net feature selection were not limited to high values (>.5, Figure 2). Selected CpGs of *Panthera* spp.—eight CpGs under LOIOCV and four CpGs under LOSOCV—were only located in two regions, *DLX5* and *ELOVL2*, which were different from domestic cats and Tsushima leopard cats (Figure 2). For domestic cats, all five targeted gene regions contributed to the best model (composed of 23 CpGs); for Tsushima leopard cats, 14 CpGs in four regions, excluding *TCF21*, yielded the best models. For the ALL

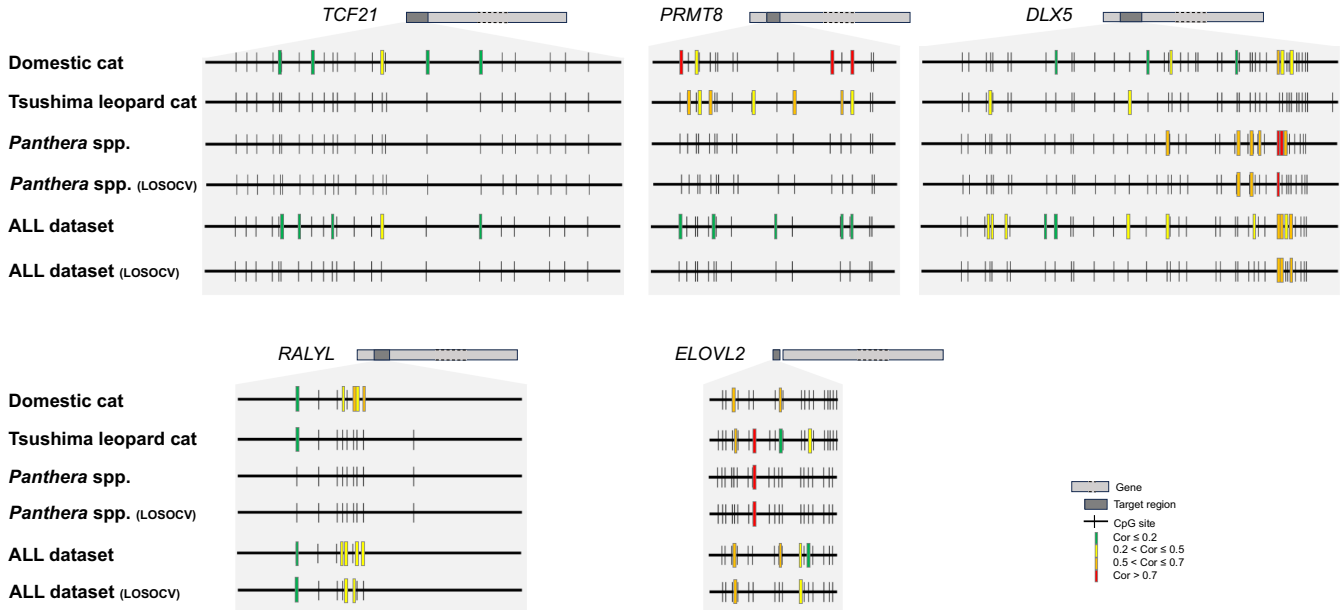

FIGURE 2 Selected CpGs of five targeted gene regions in the best model for each study group. Coloured bars represent CpG sites that were selected and show correlation coefficients <0.2 between methylation rates and chronological ages (green), 0.2–0.5 (yellow), 0.5–0.7 (orange) and >0.7 (red).

TABLE 3 Accuracy and precision of the best models.

	MAE	MedAE	RMSE	R^2	*r*
Domestic cat	1.966	1.595	2.514	.808	.899
Tsushima leopard cat	1.348	0.984	1.902	.805	.897
Panthera spp.	1.552	1.279	1.997	.873	.934
Panthera spp. (LOSOCV)	1.582	1.222	1.983	.875	.936
ALL dataset	0.086	0.071	0.110	.737	.859
ALL dataset (LOSOCV)	0.086	0.074	0.110	.735	.857

Note: MAE, MedAE and RMSE in ALL dataset models are measured based on relative age (percentages) and are not directly comparable with those of other study groups based on age (years).

Abbreviations: LOSOCV, leave-one-species-out cross-validation; MAE, mean absolute error; MedAE, median absolute error; *r*, Pearson correlation between predicted age and chronological age (*p*-values of *r* were all less than .001); R^2, squared correlation between predicted age and chronological age; RMSE, root mean square error.

dataset, the best LOIOCV model is composed of many more CpGs (31 CpGs) in all five targeted gene regions, while the best LOSOCV model only contained eight CpGs from *DLX5*, *RALYL* and *ELOVL2*.

Excluding the ALL dataset, Pearson correlation coefficients between predicted and chronological age (*r*) under the best models were higher than .890, and R^2 values were larger than .80. MAE was 1.966 for domestic cats; 1.348 for Tsushima leopard cats; 1.552 (under LOIOCV) and 1.582 (under LOSOCV) for *Panthera* spp. and .086 (under LOIOCV) and .086 (under LOSOCV) for the ALL dataset. Larger prediction errors (i.e. MAE, MedAE and RMSE) and lower R^2 and age correlations were obtained when adopting the best ALL dataset model in other data groups (i.e. domestic cats, Tsushima leopard cats and *Panthera* spp.) (Table S3). Plots of the best age estimation models are shown in Figure 3. Changes in the predicted age of individuals with multiple samples are shown in Figures S1–S3 for all study groups.

As a large sample size of diseased domestic cats was available ($n = 105$), we also attempted age prediction for the healthy domestic cats by using the diseased samples as training data and the healthy samples ($n = 34$) as test data. Very similar results (diseased [training]: MAE = 1.521, $r = .907$; healthy [test]: MAE = 1.971, $r = .902$, Table S4) were obtained compared to previous domestic cat models that included all samples (Table 4). This indicates that health conditions did not influence model performance.

3.3 | Factors affecting the deviation of predicted age from chronological age

The results of mixed linear regression for investigating the factors influencing ΔAge are summarized in Table 4 and Table S5. Older samples had a younger predicted age compared to chronological age

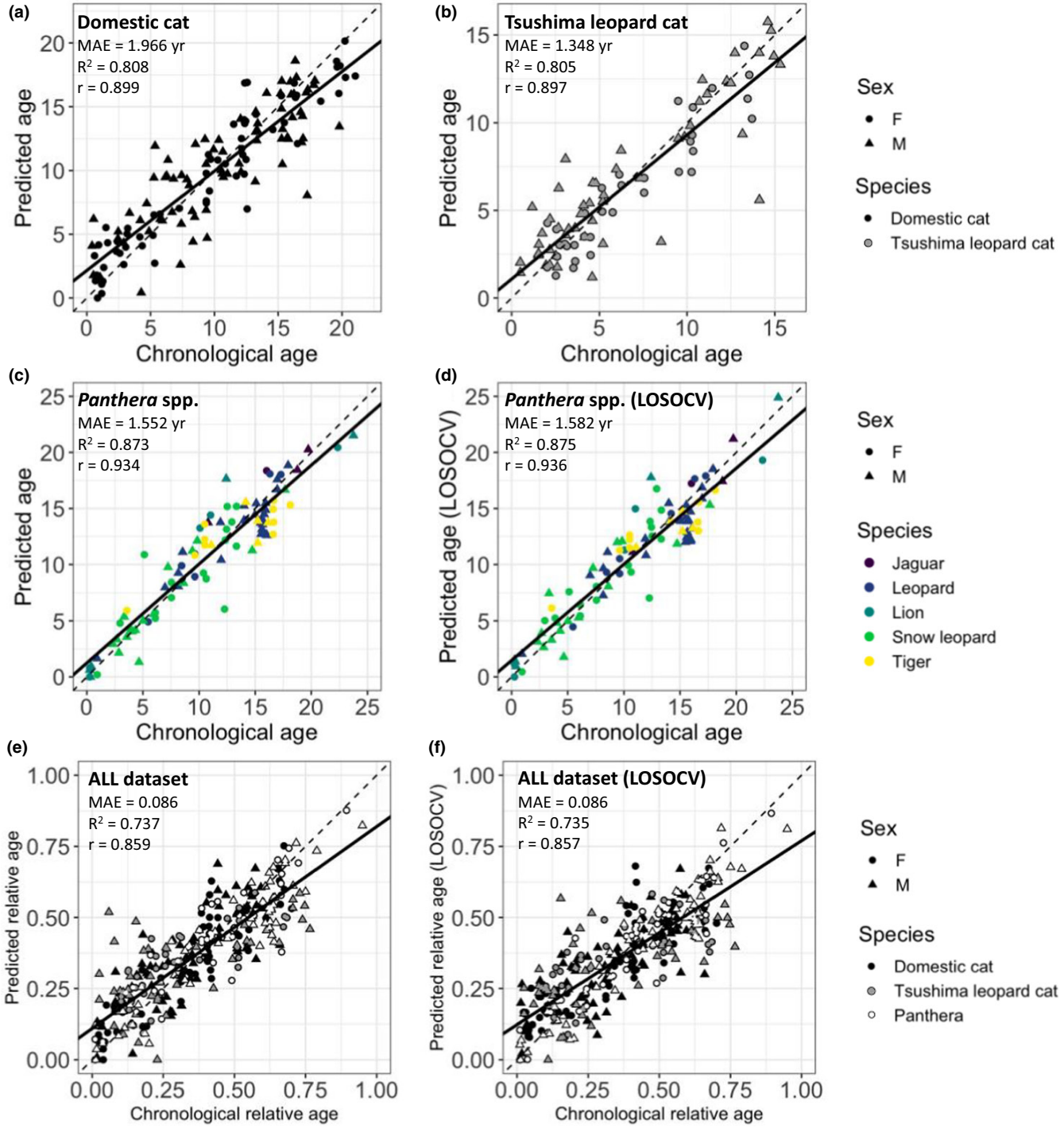

FIGURE 3 Best age estimation models for (a) domestic cats, (b) Tsushima leopard cats, (c) *Panthera* spp., (d) *Panthera* spp. (LOSOCV), (e) ALL dataset and (f) ALL dataset (LOSOCV). Detailed model accuracy and precision are summarized in Table 3. In the ALL dataset, age is converted to relative age (individual age relative to the maximum lifespan of its species). The solid lines and dashed lines represent the regression and identity lines ($y = x$) respectively. MAE in ALL dataset models are measured based on relative age (percentages) and are not directly comparable with those of other study groups based on age (years). MAE, mean absolute error; R^2, squared correlation between predicted age and chronological age; *r*, Pearson correlation between predicted age and chronological age.

in all groups ($p < .001$). Sex did not contribute to ΔAge in any group. Health conditions did not affect ΔAge in domestic cats, Tsushima leopard cats and the ALL dataset (Table 4; Table S5). However, the diseased samples showed smaller ΔAge than healthy samples in *Panthera* spp. ($p = .005$, Table 4).

Older Tsushima leopard cats tended to have smaller ΔAge in the best ALL-dataset models regardless of the cross-validation methods used (i.e. LOIOCV or LOSOCV, $p < .001$), while older *Panthera* spp. (especially lion, snow leopard and tiger, Table S5) tended to have larger ΔAge under LOSOCV ($p < .001$, Table 4). For *Panthera* spp.

TABLE 4 Coefficients and *p*-values for the mixed linear regression of ΔAge in the best age estimation models of *Panthera* spp. and the ALL dataset.

		Estimate	*p*-value
Panthera spp. Marginal R^2 = .311 Conditional R^2 = .532	(Intercept)	1.736	.009**
	Chronological age	−0.147	.0011**
	Sex (Male)	−0.336	.512
	Health condition (Diseased)	−2.294	<.001***
	Species (Jaguar)	2.390	.032*
	Species (Leopard)	0.100	.877
	Species (Lion)	0.738	.343
	Species (Tiger)	0.576	.474
Panthera spp. (LOSOCV) Marginal R^2 = .274 Conditional R^2 = .633	(Intercept)	1.778	.013*
	Chronological age	−0.161	<.001***
	Sex (Male)	0.028	.961
	Health condition (Diseased)	−1.635	.005**
	Species (Jaguar)	1.229	.292
	Species (Leopard)	−0.204	.778
	Species (Lion)	0.837	.335
	Species (Tiger)	0.286	.753
ALL dataset Marginal R^2 = .391 Conditional R^2 = .416	(Intercept)	0.085	<.001***
	Chronological relative age	−0.224	<.001***
	Sex (Male)	−0.002	.864
	Health condition (Diseased)	0.007	.576
	Species (Tsushima leopard cat)	0.098	<.001***
	Species (*Panthera* spp.)	−0.026	.303
	Age × Species (Tsushima leopard cat)	−0.301	<.001***
	Age × Species (*Panthera* spp.)	0.048	.413
ALL dataset (LOSOCV) Marginal R^2 = .477 Conditional R^2 = .568	(Intercept)	0.128	<.001***
	Chronological relative age	−0.367	<.001***
	Sex (Male)	−0.002	.888
	Health condition (Diseased)	0.000	.994
	Species (Tsushima leopard cat)	0.067	.007**
	Species (*Panthera* spp.)	−0.092	<.001***
	Age × Species (Tsushima leopard cat)	−0.215	.0011**
	Age × Species (*Panthera* spp.)	0.220	<.001***

Note: Results of the remaining study groups are presented in Table S5.
*p < .05. **p < .01. ***p < .001.

models, a larger ΔAge was found in jaguars than the other spp. under LOIOCV (p = .032, Table 4).

The results of mixed linear regression for investigating the factors influencing |ΔAge| are summarized in Table S6. Male domestic cats had larger |ΔAge| in the best domestic cat model (p = .028). Older *Panthera* had slightly larger |ΔAge| in the best *Panthera* spp. models, regardless of whether under LOIOCV (p = .0497) or LOSOCV (p = .037). Samples of male individuals (p = .007) and Tsushima leopard cats (p < .001) had larger |ΔAge| in the best ALL dataset model under LOIOCV. Older samples (p = .035) and samples of Tsushima leopard cats (p = .0013) had larger |ΔAge| in the best ALL dataset model under LOSOCV.

3.4 | Age prediction for wild samples of Tsushima leopard cats

The predicted ages of wild-born Tsushima leopard cats of unknown age are summarized in Table 5. Excluding the two samples from Leocat_w3, the estimated epigenetic ages of others were consistent with the ages estimated from morphological observation. The estimated ages of the samples taken at short intervals, such as within 1 year, also showed variation to some extent (e.g. individual Leocat_w1, mean = 10.438 years, SD = 0.556 years; individual Leocat_w4, mean = 10.144 years, SD = 1.269 years).

4 | DISCUSSION

From the performance of the age estimation models developed for domestic cats, Tsushima leopard cats and *Panthera* spp., age estimation in these species appeared to be successful with high accuracy using 8–23 CpGs from only 2–5 gene regions, cross-validated by LOIOCV (Pearson correlation coefficient between predicted age and chronological age [r] > .890, MAE ranging 1.348–1.966 years; Table 3). The cost for next-generation sequencing analysis per sample was approximately $7 based on five markers for a total of 334 samples, which was less than one-tenth of HorvathMammalMethylChip40 ($160/sample). Therefore, the method presented in this study is cost-effective and practical for conservation applications. To the best of our knowledge, this study is also the first to analyse several Felidae species, from small to big cat species (*Panthera* spp.), with appreciable sample sizes. We also acknowledge that model performance could be further improved by applying more age-correlated markers. For example, *SLC12A5*, as demonstrated by Raj et al. (2021), was highly related to the age of domestic cats and is a strong candidate.

DLX5 and *ELOVL2* were the two gene regions selected across all the targeted species that indicate possible usefulness as age estimation markers for other unstudied Felidae species. As was mentioned in Section 2.6, the coverage of *DLX5* was lower than other regions. We suspect that the skewed distribution of the number of reads for target regions is because during library preparation, all target regions were pooled per sample and tagged using a PCR-based approach; although the starting mixture was adjusted to be equimolar, the longest regions (*DLX5*) were the least amplified. Although the less coverage may have resulted in a larger deviation of *DLX5* methylation rates (Roeh et al., 2018), resulting in less correlation with age, the region is still found to be important for all species (Figure 2). *RALYL* showed considerable age-related methylation changes of its CpGs

TABLE 5 Predicted age and sample information of wild-born Tsushima leopard cats.

Individual ID	Sex	Rescued date	Sampling date	Predicted age/age stage at the time of sampling (morphology)	Predicted age (years) (DNA methylation)
Leocat_w1	F	2005/3/4	2013/5/23	11 years	10.306
Leocat_w1	F	2005/3/4	2013/6/21	11 years	11.048
Leocat_w1	F	2005/3/4	2013/7/28	11 years	9.961
Leocat_w2	M	2010/8/6	2015/3/12	6 years	6.044
Leocat_w2	M	2010/8/6	2021/1/11	12 years	13.138
Leocat_w3	F	2015/12/26	2018/2/13	Old	10.437
Leocat_w3	F	2015/12/26	2020/6/9	Old	6.166
Leocat_w4	F	2020/5/13	2020/6/28	Adult/Old	10.139
Leocat_w4	F	2020/5/13	2020/10/15	Adult/Old	8.636
Leocat_w4	F	2020/5/13	2020/11/19	Adult/Old	10.013
Leocat_w4	F	2020/5/13	2020/12/6	Adult/Old	9.057
Leocat_w4	F	2020/5/13	2020/12/16	Adult/Old	9.807
Leocat_w4	F	2020/5/13	2021/1/14	Adult/Old	9.448
Leocat_w4	F	2020/5/13	2021/6/22	Adult/Old	12.257
Leocat_w4	F	2020/5/13	2021/10/23	Adult/Old	11.796

for *Panthera* spp. (i.e. most of the correlation coefficients were larger than .5, Appendix S5), which was consistent with the observations made in Qi et al. (2021). However, *RALYL* was not the first choice in age estimation model building for *Panthera* spp. *PRMT8* appears to be an important gene for domestic cats and Tsushima leopard cats, because the methylation changes of its CpGs were found to be highly correlated with age, and about one-third of the sites were selected in the best models of these two small feline species.

We also built common models across all targeted felid species. However, the model performance was not as good as the species group-specific models (Table S3). This is especially so for Tsushima leopard cats, as the MAE increased drastically from 1.348 to 2.388 (under LOIOCV) and 2.468 (under LOSOCV). Younger Tsushima leopard cats tended to have an older predicted age, while older samples tended to have a younger predicted age, compared to their chronological age (Figure 3e,f). Although the selected CpGs in the ALL dataset LOIOCV model covered most of the CpGs in the Tsushima leopard cat-specific model (Figure 2), the contribution of CpGs was assumed to be largely different (SVMr is a non-linear model, so the actual feature contribution was not extracted). Additionally, the increased complexity of the ALL dataset LOIOCV model (i.e. larger sets of CpGs, that is, 31 CpGs were used) and significant species difference observed in both ΔAge and |ΔAge| of the best LOSOCV model (Table 4; Tables S5 and S6) implies that, at least for the five genomic regions explored in this study, combining all samples to construct a common age estimation model for felids seems not advisable.

A common age estimation model for *Panthera* spp. seems plausible, which is supported by the high prediction accuracy (MAE = 1.582 years) maintained in the *Panthera* LOSOCV model (Table 3; Figure 3f). The closer genetic distance among *Panthera* spp. could be the reason for the relatively high estimation accuracy (Li et al., 2016). Because only two regions (*DLX5* and *ELOVL2*) were used to construct the age estimation model for *Panthera* spp., the cost could be reduced to half—only about $3 per DNA sample. Nevertheless, species differences were observed to some extent in the LOIOCV model. The difference between the predicted and chronological age of jaguars tended to be larger than the other species in the LOIOCV model (Table 4); however, a definite conclusion cannot be made as only four samples of jaguars were available. The performance of the developed models with regard to some *Panthera* spp., especially jaguars, African lions and Amur tigers, cannot be ascertained due to the low sample size. The samples of *Panthera* spp. were obtained from captive individuals in Japanese zoos, which inevitably included some individuals with close or distant blood relations and a relatively uniform living environment, which may also be one of the reasons for the small model deviation. In the future, this should be addressed using a larger sample size from more diverse parentage and environment for constructing a more reliable common model for *Panthera* spp. and separate models for each species, to further investigate whether a common model or separate models is best suited for the *Panthera* genus.

Stepwise elastic net feature selection-SVMr regression yielded the best models in our study. In these models, some CpGs with a low correlation (Figure 2) between the methylation rate and chronological age were selected. This suggests that CpGs must not be selected based solely on the correlation coefficients. Sources of variation for the correlation between CpGs and age are limited when few genes are used. Furthermore, in some cases, CpGs that are highly correlated with age are likely to cluster at a limited number of gene locations, focusing only on the magnitude of the correlation coefficients with bias towards selecting only CpGs with a similar correlation with age. This results in a low diversity of explanatory variables in age estimation models and consequently, low accuracy in the age estimates.

Unlike most previous studies, we included both healthy samples and samples with a variety of diseases. In the context of wildlife conservation, rescued individuals of unknown age may be healthy or diseased; therefore, an age estimation model that could be applied with sufficient accuracy on individuals with varied health conditions would be required. Our models for domestic cats and Tsushima leopard cats did not find significant estimation differences between the healthy and unhealthy samples (Tables S5 and S6). For domestic cats, further evidence was provided by the very similar performance in the model trained on all samples and that trained on the diseased samples to predict the age of healthy samples. Consequently, it suggests that our models for the two small cat species are robust enough to estimate ages in samples whose health conditions are varied or unknown.

We noticed a better estimation accuracy in the diseased samples of *Panthera* spp. compared to that of the healthy samples (Table 4). Most of these diseased samples were from individuals between the ages of 10 and 20 years (Appendix S3), which may still be relatively healthy considering their relative ages to the species' maximum life spans, and thus expressed a young epigenetic age. Moreover, current disease diagnosis methods for *Panthera* spp. are not as extensive as those used for domestic cats. Some individuals who were unhealthy and displayed shifted epigenetic age, may have been misdiagnosed as healthy individuals, or vice versa. Nevertheless, only a few diseased samples ($n = 13$) were included compared to the healthy samples ($n = 83$) in the *Panthera* spp. group (Table 1). Therefore, the effect of disease on age estimation cannot be accurately estimated based on the samples included in this study.

Similar to the results of many previous studies (El Khoury et al., 2019; Prado et al., 2021; Raj et al., 2021), the developed models tended to underestimate the age of older individuals. As suggested by El Khoury et al. (2019), one explanation could be the saturation of methylation rates of targeted CpGs, that is, CpGs already reach either full methylation or complete de-methylation before the individuals age further. Another assumption is that long-lived individuals are biologically younger and have a younger predicted epigenetic age than their chronological age.

We used LOIOCV for the age estimation models built on captive Tsushima leopard cats to reduce the influence of different sample sizes in different individuals; however, future sampling should be improved further. We only had a small number of samples except for the age classes 2–6 years old, and most samples over 6 years old came from only two individuals (Leocat_3 and Leocat_5; Figure S2). We also predicted the ages of several wild-born Tsushima leopard cats of unknown age (Table 5), including both newly rescued (Leocat_w4) and those who lived long in a captive environment (Leocat_w1, Leocat_w2, Leocat_w3). The estimation results were satisfactory, which implied that this model, which was developed using captive individuals, can potentially also be applied to wild individuals.

Importantly, the estimation variation existed even among samples that were collected within 1 year from the same individual (see within-individual age change plot, Figures S1–S3). Therefore, for all species, we recommend multiple sampling within a 1-year window or less to obtain an average predicted age as the final predicted age. Environmental effects are also associated with DNA methylation change and age acceleration. To investigate the accuracy of age estimation on wild populations, a larger sample of wild-born individuals with known age, especially subadult and young adult individuals, is required.

In conclusion, we successfully developed epigenetic clocks using 8–23 CpGs from 2 to 5 gene regions with satisfactory accuracy for domestic cats (MAE = 1.966 years), Tsushima leopard cats (MAE = 1.348 years) and *Panthera* spp. (MAE = 1.552 years), using cost-effective next-generation sequencing and multiple machine-learning algorithms. Our models for domestic cats and Tsushima leopard cats are applicable to individuals of varying healthy conditions. We do not recommend building a common age estimation model for all the target species using our markers. Alternatively, we showed the possibility of developing a common model for five *Panthera* spp. The changes in predicted age for the same individual implied that multiple sampling within a 1-year window to obtain the mean of predicted age as the final predicted age is advisable for future applications.

AUTHOR CONTRIBUTIONS

H.Q. designed and conducted the study and wrote the first draft of the manuscript. Q.L.L. assisted in the laboratory work and the revision of the manuscript and analysis. K.K. and M.I-M. designed the study, communicated with the institutes for samples and revised the manuscript. N.N. led the Miseq-related work, including technical part and data analysis. All authors approved the final manuscript for publication.

ACKNOWLEDGEMENTS

We would like to express our sincere thanks to the Kyoto Medical Center, Daktari Animal Hospital and Dr. Minori Arahori of Anicom Specialty Medical Institute, Inc. for providing domestic cat samples and sample information. We greatly appreciate the Tsushima Wildlife Conservation Center, Nagoya Higashiyama Zoo and Botanical Gardens and Yokohama Zoological Gardens for providing Tsushima leopard cat samples and sample information. We are grateful to the Kobe Oji Zoo, Adventure World, Akita Omoriyama Zoo, Asahiyama Zoological Park and Wildlife Conservation Center, Hamamatsu Zoological Gardens, Nagoya Higashiyama Zoo and Botanical Gardens, Omuta City Zoo, Tama Zoological Park for providing samples and sample information of *Panthera* spp. We also thank the Japanese Association of Zoos and Aquariums (JAZA) and the Ministry of the Environment, Japan for their permission to sample and for providing sample information. This study was financially supported by JSPS (KAKENHI Grant Number 20H00420 to M. I.-M., 20H03008 to K. K.) and the Environment Research and Technology Development Fund (JPMEERF20214001) of the Environmental Restoration and Conservation Agency of Japan (M. I.-M.). We would like to thank Editage (www.editage.jp) for English language editing.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

The Appendices and other raw data, and the R and Python script of this study can be accessed at https://doi.org/10.5061/dryad.3r2280gn4.

ORCID

Huiyuan Qi https://orcid.org/0000-0003-0658-5970

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.

Anastasiadi, D., & Piferrer, F. (2023). Bioinformatic analysis for age prediction using epigenetic clocks: Application to fisheries management and conservation biology. *Frontiers in Marine Science*, *10*, 1096909. https://doi.org/10.3389/fmars.2023.1096909

Animal Lifespan. (2017). *The lifespan of animals*. August 29. https://www.zooing.net

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

Bekaert, B., Kamalandua, A., Zapico, S. C., Van de Voorde, W., & Decorte, R. (2015). Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics*, *10*(10), 922–930. https://doi.org/10.1080/15592294.2015.1080413

Blomqvist, L., & Sten, I. (1982). Reproductive biology of the snow leopard, *Panthera uncia*. *International Pedigree Book for Snow Leopards*, *3*, 71–79.

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*(6), e14821.

Bogdanović, O., & Veenstra, G. J. C. (2009). DNA methylation and methyl-CpG binding proteins: Developmental requirements and function. *Chromosoma*, *118*(5), 549–565. https://doi.org/10.1007/s00412-009-0221-9

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

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Caselli, M., Messeri, P., Dessì-Fulgheri, F., & Bandoli, F. (2022). Enriching zoo-housed ring-tailed lemurs (*Lemur catta*): Assessing the influence of three types of environmental enrichment on behavior. *Animals*, *12*(20), 2836.

Chen, G. G., Gross, J. A., Lutz, P.-E., Vaillancourt, K., Maussion, G., Bramoulle, A., Théroux, J.-F., Gardini, E. S., Ehlert, U., Bourret, G., Masurel, A., Lepage, P., Mechawar, N., Turecki, G., & Ernst, C. (2017). Medium throughput bisulfite sequencing for accurate detection of 5-methylcytosine and 5-hydroxymethylcytosine. *BMC Genomics*, *18*(1), 96. https://doi.org/10.1186/s12864-017-3489-9

Chevallier, C., Gauthier, G., & Berteaux, D. (2017). Age estimation of live arctic foxes *Vulpes lagopus* based on teeth condition. *Wildlife Biology*, *2017*(4), 1–6.

Cho, Y. S., Hu, L., Hou, H., Lee, H., Xu, J., Kwon, S., Oh, S., Kim, H.-M., Jho, S., Kim, S., Shin, Y.-A., Kim, B. C., Kim, H., Kim, C., Luo, S.-J., Johnson, W. E., Koepfli, K.-P., Schmidt-Küntzel, A., Turner, J. A., … Bhak, J. (2013). The tiger genome and comparative analysis with lion and snow leopard genomes. *Nature Communications*, *4*(1), 2433. https://doi.org/10.1038/ncomms3433

Conroy, M., O'Neill, D., Boag, A., Church, D., & Brodbelt, D. (2019). Epidemiology of road traffic accidents in cats attending emergency-care practices in the UK. *Journal of Small Animal Practice*, *60*(3), 146–152.

Davis, B. W., Raudsepp, T., Pearks Wilkerson, A. J., Agarwala, R., Schäffer, A. A., Houck, M., Chowdhary, B. P., & Murphy, W. J. (2009). A high-resolution cat radiation hybrid and integrated FISH mapping resource for phylogenomic studies across Felidae. *Genomics*, *93*(4), 299–304. https://doi.org/10.1016/j.ygeno.2008.09.010

de Magalhães, J. P., Costa, J., & Church, G. M. (2007). An analysis of the relationship between metabolism, developmental schedules, and longevity using phylogenetic independent contrasts. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *62*(2), 149–160.

El Khoury, L. Y., Gorrie-Stone, T., Smart, M., Hughes, A., Bao, Y., Andrayas, A., Burrage, J., Hannon, E., Kumari, M., Mill, J., & Schalkwyk, L. C. (2019). Systematic underestimation of the epigenetic clock and age acceleration in older subjects. *Genome Biology*, *20*(1), 283. https://doi.org/10.1186/s13059-019-1810-4

Eskelinen, H. C., Winship, K. A., & Borger-Turner, J. L. (2015). Sex, age, and individual differences in bottlenose dolphins (*Tursiops truncatus*) in response to environmental enrichment. *Animal Behavior and Cognition*, *2*(3), 241–253.

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

Gray, A. P. (1972). *Mammalian hybrids, a check-list and bibliography*. Revised edition. Commonwealth Agricultural Bureaux.

Gruber, H.-J., Semeraro, M. D., Renner, W., & Herrmann, M. (2021). Telomeres and age-related diseases. *Biomedicines*, *9*(10), 1335.

Hadley, W., Romain, F., Lionel, H., Kirill, M., & Davis, V. (2023). *Dplyr: A grammar of data manipulation*. https://dplyr.tidyverse.org

Hansen, K. D., Langmead, B., & Irizarry, R. A. (2012). BSmooth: From whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biology*, *13*(10), R83. https://doi.org/10.1186/gb-2012-13-10-r83

Hartman, K. L., Wittich, A., Cai, J. J., van der Meulen, F. H., & Azevedo, J. M. N. (2016). Estimating the age of Risso's dolphins (*Grampus griseus*) based on skin appearance. *Journal of Mammalogy*, *97*(2), 490–502. https://doi.org/10.1093/jmammal/gyv193

Hecht, L. (2021). The importance of considering age when quantifying wild animals' welfare. *Biological Reviews*, *96*(6), 2602–2616. https://doi.org/10.1111/brv.12769

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), 1–20.

Horvath, S., Haghani, A., Macoretta, N., Ablaeva, J., Zoller, J. A., Li, C. Z., Zhang, J., Takasugi, M., Zhao, Y., & Rydkina, E. (2022). DNA methylation clocks tick in naked mole rats but queens age more slowly than nonbreeders. *Nature Aging*, *2*(1), 46–59.

Inoue, N. (2021). *Japanese studbook of Tsushima leopard cat (Prionailurus bengalensis euptilurus)*. Japanese Association of Zoos and Aquariums.

Kirkwood, T. B., & Austad, S. N. (2000). Why do we age? *Nature*, *408*(6809), 233–238.

Krivonosov, M. I., Kondakova, E. V., Bulanov, N. A., Polevaya, S. A., Franceschi, C., Ivanchenko, M. V., & Vedunova, M. V. (2022). A new cognitive clock matching phenotypic and epigenetic ages. *Translational Psychiatry*, *12*(1), 1–9.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*, 1–26. https://doi.org/10.18637/jss.v028.i05

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lacy, R. C. (2019). Lessons from 30 years of population viability analysis of wildlife populations. *Zoo Biology*, *38*(1), 67–77. https://doi.org/10.1002/zoo.21468

Larison, B., Pinho, G. M., Haghani, A., Zoller, J. A., Li, C. Z., Finno, C. J., Farrell, C., Kaelin, C. B., Barsh, G. S., Wooding, B., Robeck, T. R., Maddox, D., Pellegrini, M., & Horvath, S. (2021). Epigenetic models developed for plains zebras predict age in domestic horses and endangered equids. *Communications Biology*, *4*(1), 1412. https://doi.org/10.1038/s42003-021-02935-z

Leitão, E., Beygo, J., Zeschnigk, M., Klein-Hitpass, L., Bargull, M., Rahmann, S., & Horsthemke, B. (2018). Locus-specific DNA methylation analysis by targeted deep bisulfite sequencing. *Methods in Molecular Biology*, *1767*, 351–366.

Lemaître, J.-F., Rey, B., Gaillard, J.-M., Régis, C., Gilot-Fromont, E., Débias, F., Duhayer, J., Pardonnet, S., Pellerin, M., Haghani, A., Zoller, J. A., Li, C. Z., & Horvath, S. (2022). DNA methylation as a tool to explore ageing in wild roe deer populations. *Molecular Ecology Resources*, *22*(3), 1002–1015. https://doi.org/10.1111/1755-0998.13533

Li, G., Davis, B. W., Eizirik, E., & Murphy, W. J. (2016). Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Research*, *26*(1), 1–11. https://doi.org/10.1101/gr.186668.114

Li, X., Li, W., & Xu, Y. (2018). Human age prediction based on DNA methylation using a gradient boosting regressor. *Genes*, *9*(9), 424. https://doi.org/10.3390/genes9090424

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*(1), 22. https://doi.org/10.1186/s13059-018-1397-1

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*, *3*, 1144–1166. https://doi.org/10.1038/s43587-023-00462-6

Mori, A. (1979). Analysis of population changes by measurement of body weight in the Koshima troop of Japanese monkeys. *Primates*, *20*(3), 371–397.

Nakamura, S., Yamazaki, J., Matsumoto, N., Inoue-Murayama, M., Qi, H., Yamanaka, M., Nakanishi, M., Yanagawa, Y., Sashika, M., Tsubota, T., Ito, H., & Shimozuru, M. (2023). Age estimation based on blood DNA methylation levels in brown bears. *Molecular Ecology Resources*, *23*(6), 1211–1225. https://doi.org/10.1111/1755-0998.13788

Nussey, D. H., Froy, H., Lemaitre, J.-F., Gaillard, J.-M., & Austad, S. N. (2013). Senescence in natural populations of animals: Widespread evidence and its implications for bio-gerontology. *Ageing Research Reviews*, *12*(1), 214–225. https://doi.org/10.1016/j.arr.2012.07.004

Oli, M. K., & Dobson, F. S. (2003). The relative importance of life-history variables to population growth rate in mammals: Cole's prediction revisited. *The American Naturalist*, *161*(3), 422–440.

Paoli-Iseppi, D., 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.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

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

Prado, N. A., Brown, J. L., Zoller, J. A., Haghani, A., Yao, M., Bagryanova, L. R., Campana, M. G., Maldonado, E., Raj, K., Schmitt, D., Robeck, T. R., & Horvath, S. (2021). Epigenetic clock and methylation studies in elephants. *Aging Cell*, *20*(7), e13414. https://doi.org/10.1111/acel.13414

Qi, H., Kinoshita, K., Mori, T., Matsumoto, K., Matsui, Y., & Inoue-Murayama, M. (2021). Age estimation using methylation-sensitive high-resolution melting (MS-HRM) in both healthy felines and those with chronic kidney disease. *Scientific Reports*, *11*(1), 1–10.

R Core Team. (2021). *R: A language and environment for statistical computing*. 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

Roeh, S., Wiechmann, T., Sauer, S., Ködel, M., Binder, E. B., & Provençal, N. (2018). HAM-TBS: High-accuracy methylation measurements via targeted bisulfite sequencing. *Epigenetics & Chromatin*, *11*(1), 1–10.

Safran, M., Rosen, N., Twik, M., BarShir, R., Stein, T. I., Dahary, D., Fishilevich, S., & Lancet, D. (2021). The GeneCards suite. In I. Abugessaisa & T. Kasukawa (Eds.), *Practical guide to life science databases* (pp. 27–56). Springer Nature. https://doi.org/10.1007/978-981-16-5812-9_2

Sharma, V., Sharma, C. P., Sharma, V., Goyal, S. P., Stevens, H., & Gupta, S. K. (2022). Age estimation of tiger *Panthera tigris* (Linnaeus, 1758) and lion *Panthera leo* (Linnaeus, 1758) (Mammalia: Carnivora: Felidae): Applicability of cementum annuli analysis method. *Journal of Threatened Taxa*, *14*(9), 810. https://doi.org/10.11609/jott.7727.14.9.21805-21810

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T. I., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., & Lancet, D. (2016). The GeneCards suite: From gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics*, *54*(1), 1.30.1–1.30.33. https://doi.org/10.1002/cpbi.5

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), 1–14.

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

Thorel, M., Pignon, C., Arne, P., Donnelly, T. M., & Rivière, J. (2020). Clouded leopard (*Neofelis nebulosa*) morbidity and mortality in captive-bred populations: A comprehensive retrospective study of medical data from 271 individuals in European, Asian, and Australian zoos. *Journal of Zoo and Wildlife Medicine*, *51*(1), 150–158.

Topuz, K., Uner, H., Oztekin, A., & Yildirim, M. B. (2018). Predicting pediatric clinic no-shows: A decision analytic framework using elastic net and Bayesian belief network. *Annals of Operations Research*, *263*(1), 479–499. https://doi.org/10.1007/s10479-017-2489-0

Van Rossum, G., & Drake, F. L., Jr. (2009). *Python 3 reference manual*. CreateSpace.

Vidaki, A., González, D. M., Jiménez, B. P., & Kayser, M. (2021). Male-specific age estimation based on Y-chromosomal DNA methylation. *Aging (Albany NY)*, *13*(5), 6442–6458. https://doi.org/10.18632/aging.202775

White, P. A., Ikanda, D., Ferrante, L., Chardonnet, P., Mesochina, P., & Cameriere, R. (2016). Age estimation of African lions *Panthera leo* by ratio of tooth areas. *PLoS One*, *11*(4), e0153648.

Wilkinson, G. S., & Brunet-Rossinni, A. K. (2009). Methods for age estimation and the study of senescence in bats. In T. H. Kunz & S. Parson (Eds.), *Ecological and behavioral methods for the study of bats* (pp. 315–325). The Johns Hopkins University Press.

Wurster-Hill, D. H., & Centerwall, W. R. (1982). The interrelationships of chromosome banding patterns in canids, mustelids, hyena, and felids. *Cytogenetic and Genome Research*, *34*(1–2), 178–192.

Xia, X., Chen, W., McDermott, J., & Han, J.-D. J. (2017). Molecular and phenotypic biomarkers of aging. *F1000Research*, *6*, 860. https://doi.org/10.12688/f1000research.10692.1

Xu, C., Qu, H., Wang, G., Xie, B., Shi, Y., Yang, Y., Zhao, Z., Hu, L., Fang, X., Yan, J., & Feng, L. (2015). A novel strategy for forensic age prediction by DNA methylation and support vector regression model. *Scientific Reports*, *5*(1), 17788. https://doi.org/10.1038/srep17788

Yagi, G., Kogi, K., & Sakai, M. (2023). Noninvasive age estimation for wild Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) using speckle appearance based on quantification-theory model I analysis. *Marine Mammal Science*, *1*(9), 662–670. https://doi.org/10.1111/mms.12999

Youn, S.-H., Efladl, A. K., Chung, M.-J., Jung, E., Shin, K.-Y., Shin, H.-J., Yang, J.-Y., Heo, K.-S., Chung, D.-H., & Yim, J.-H. (2022). Symmetric dimethylarginine is a promising biomarker for the early detection of age-related kidney dysfunction in zoo felids. *Acta Veterinaria*, *72*(3), 408–418.

Zhao, M., Klaassen, C. A. J., Lisovski, S., & Klaassen, M. (2019). The adequacy of aging techniques in vertebrates for rapid estimation of population mortality rates from age distributions. *Ecology and Evolution*, *9*(3), 1394–1402. https://doi.org/10.1002/ece3.4854

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Qi, H., Lim, Q. L., Kinoshita, K., Nakajima, N., & Inoue-Murayama, M. (2024). A cost-effective blood DNA methylation-based age estimation method in domestic cats, Tsushima leopard cats (*Prionailurus bengalensis euptilurus*) and *Panthera* species, using targeted bisulphite sequencing and machine learning models. *Molecular Ecology Resources*, *24*, e13928. https://doi.org/10.1111/1755-0998.13928