Title

An epigenetic clock for Xenopus tropicalis

Author

Bennett, Ronan

Publication Date

2025

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

An Epigenetic Clock for *Xenopus tropicalis*

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Bioinformatics

by

Ronan Robert Bennett

2025

ABSTRACT OF THE THESIS

An Epigenetic Clock for *Xenopus tropicalis*

by

Ronan Robert Bennett
Master of Science in Bioinformatics
University of California, Los Angeles, 2025
Professor Matteo Pellegrini, Chair

DNA methylation clocks have been widely used for accurate age prediction, but most studies have been carried out on mammals. Here we present an epigenetic clock for the aquatic frog *Xenopus tropicalis*, a widely used model organism in developmental biology and genomics. To construct the clock, we collected DNA methylation data from 192 frogs using targeted bisulfite sequencing at genomic regions containing CpG sites previously shown to have age-associated methylation in *Xenopus*. We found highly positively and negatively age-correlated CpGs are enriched in heterochromatic regions marked with H4K20me3 and H3K9me3. Positively age-correlated CpGs are enriched in bivalent chromatin and gene bodies with H3K36me3, and tend to be proximal to lowly expressed genes. These epigenetic features of aging are similar to those found in mammals, suggesting evolutionary conservation of epigenetic aging mechanisms. Our clock enables future aging biology experiments that leverage the unique properties of amphibians.

The thesis of Ronan Robert Bennett is approved.

Robert D. Damoiseaux

Shantanu H. Joshi

Matteo Pellegrini, Committee Chair

University of California, Los Angeles

2025

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

This thesis is adapted from Bennett, R., Morselli, M., Petrova, K., Peshkin, L., Pellegrini, M. An epigenetic clock for *Xenopus tropicalis*. npj Aging 11, 38 (2025). https://doi.org/10.1038/s41514-025-00236-x. The article is licensed under CC-BY 4.0 http://creativecommons.org/licenses/by/4.0/. Text and figures are repurposed here in compliance with the license. Document formatting and figure numbers were modified to fit the thesis submission guidelines.

Contributions: (Introduction) R.B: Initial draft, editing. M.M, L.P, M.P: Editing. (Results and Methods) R.B: Data analysis, initial draft, editing. K.P: Data analysis for "Gene expression proximal to highly age-associated CpGs". M.M, L.P, M.P: Data generation and project supervision, manuscript editing. (Discussion) R.B, L.P: Initial draft, editing. M.M, M.P: Editing.

I sincerely thank the co-authors above for their generous assistance and supervision throughout the project. I thank Silvia Nigro (University of Parma) for helping pre-process bisulfite sequencing data, and William Ratzan (Harvard Medical School) for preparation of *Xenopus* Hybloc DNA. This project used computational and storage services associated with the Hoffman2 Cluster, operated by the UCLA Office of Advanced Research Computing's Research Technology Group, and the High Performance Computing facility at the University of Parma (Parma, Italy). Frog samples were obtained from the National Xenopus Resource (RRID:SCR_01373). Supported by the National Institutes of Health R24 award OD031956 (L.P.).

CHAPTER 1

Introduction

DNA methylation clocks have been broadly used for age prediction in mammals [1]. A common approach for constructing these clocks is to use a regularized linear regression model to predict chronological age based on DNA methylation levels. Certain cytosine-phosphate-guanine dinucleotides (CpG sites) have methylation levels that reliably increase or decrease with age, so these linear clocks are able to accurately predict age based on these sites. These clocks allow the determination of the age of animals caught in the wild, which opens up new research opportunities in conservation biology.

Another common application of epigenetic clocks is to determine the rate of aging of an individual, by measuring whether they are aging faster or slower than other individuals with the same chronological age. This property is useful for studies that evaluate the impact of interventions on lifespan. For long-lived species, an experiment that directly measures the time until death can be prohibitively long, but an epigenetic clock can provide a more immediate endpoint to evaluate the effectiveness of an intervention, thus accelerating the pace of aging research. In humans, epigenetic age acceleration (the number of years that epigenetic predicted age exceeds chronological age) is correlated with all-cause mortality and a number of major diseases such as cancer and heart disease [2–4]. The predicted age from DNA methylation clocks, also known as epigenetic age, has been shown to slow down with interventions that extend lifespan in mice such as caloric restriction [5]. Epigenetic clocks evaluated in human fibroblasts before and after treatment with Yamanaka factors to obtain induced pluripotent stem cells showed epigenetic age decrease to near zero after

the treatment, similar to the epigenetic age of other stem cells [1,5]. Furthermore, partial reprogramming of somatic cells induced a steady decline in epigenetic age over 20 days of OSKM expression until it reached zero [6]. People with diseases of accelerated aging such as Werner syndrome and Hutchinson-Gilford progeria syndrome have epigenetic ages that are significantly higher than controls [7,8]. These observations suggest that DNA methylation clocks can measure cellular aspects of the biological aging process. However, the mechanisms underlying age-associated methylation changes are poorly understood. Age-correlated CpG sites have been preferentially found in loci that are targets of PRC2 [9,10], and near promoters of transcription factors involved in developmental processes in humans [9,11] and *Xenopus* [12].

While the majority of the previous studies of epigenetic aging have been carried out in mammals, clocks have also been developed for other vertebrates, including amphibians. Amphibians are invaluable models for biological research due to their unique life cycles, diverse ecological niches, and physiological adaptability. Their ability to live in both aquatic and terrestrial environments provides insights into developmental biology, environmental adaptation, and evolutionary processes. Amphibians, like frogs and salamanders, are particularly useful in studying regeneration, as many can regenerate limbs and organs, offering models for tissue repair and stem cell research. Additionally, their permeable skin and sensitivity to environmental changes make them excellent indicators for ecological and toxicological studies, helping researchers understand the impact of environmental stressors on living organisms. An amphibian epigenetic clock could advance aging research by providing a tool to measure biological age with precision in species that exhibit unique regenerative abilities and variable lifespans. By studying how epigenetic markers correlate with aging in amphibians, researchers may uncover mechanisms of longevity, regeneration, and cellular resilience, offering insights that could inform therapies for aging and age-related diseases in humans.

An example of an amphibian epigenetic clock is the study by Zoller et al. [12] which constructed DNA methylation clocks in *Xenopus tropicalis* and *Xenopus laevis*. This study used

the mammalian methylation array [13] to measure the methylation state of 4,635 CpG sites that have high sequence conservation between *Xenopus* and mammals. They demonstrated evolutionary conservation in DNA methylation changes between *Xenopus* frogs and humans by constructing joint human-frog clocks that are able to predict the age of either species. This supports the use of *Xenopus* as a model organism to study epigenetic aging. While this dataset is well-suited for studying similarities in DNA methylation levels between *Xenopus* and mammals, it is limited by the fact that the use of the mammalian methylation array excludes the vast majority of CpG sites where methylation changes are specific to *Xenopus*. Moreover, the study is also limited in the number and type of samples used. They reported samples from 6 tissues pooled from a number of individual frogs (35 X. laevis and 30 X. tropicalis tissue samples). However, they encountered technical challenges that resulted in the loss of over half of the samples [14]. Thus, while this study represents a promising first step, because of the above limitations it is unlikely that the resulting clock or the mammalian array will be broadly used for aging studies in frogs.

In the present work, we sought to construct a robust DNA methylation clock based on data from 192 frogs, focusing on a single easily sampled tissue, and measuring a comprehensive set of *Xenopus*-specific age-associated CpGs. We generated methylation data from *Xenopus tropicalis* frogs using a skin punch from the hindlimb webbing, followed by DNA methylation measurement using targeted bisulfite sequencing [15]. We targeted CpGs that had the most significantly age-correlated methylation across the entire *Xenopus tropicalis* genome based on whole-genome bisulfite sequencing data from a previous study of 9 *Xenopus tropicalis* frogs [16]. Our approach captures *Xenopus*-specific age-associated methylation signals and is not restricted to sites that are conserved between amphibians and mammals. We therefore believe that our clock is a valuable tool for future studies in *Xenopus* aging biology and that our approach can be easily extended to other amphibians.

CHAPTER 2

Results

2.1 Data collection and filtering

We collected DNA methylation data from 192 *Xenopus tropicalis* frogs with known ages from the National Xenopus Resource at the Marine Biological Laboratory (Woods Hole, MA, USA). The tissue samples used for DNA extraction were hindlimb webbing skin punches in the 187 adult frogs (Figure 2.1). For the 5 tadpoles, the entire organism was used for DNA

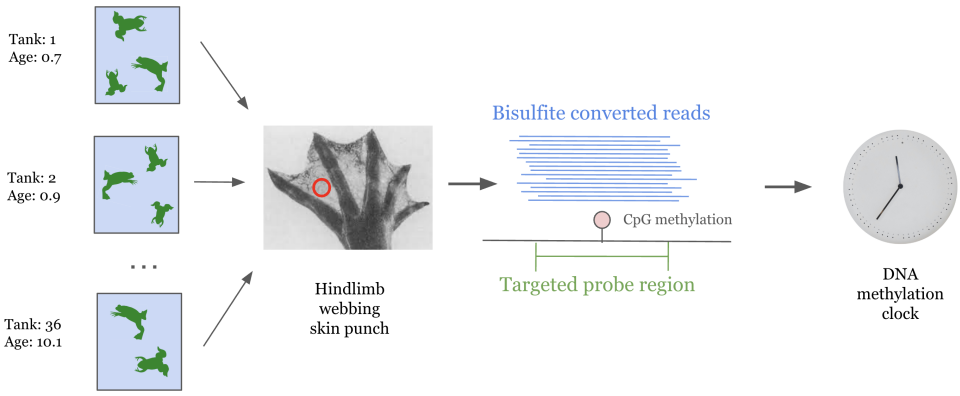

Figure 2.1: Overview of data collection workflow. Xenopus tropicalis frogs of different ages and strains were raised in tanks from birth, a skin punch from each frog's hindlimb webbing was collected for DNA extraction, and targeted bisulfite sequencing was performed to measure DNA methylation levels at a collection of predetermined genomic regions. These targeted regions were designed to capture CpG sites that were previously shown to be highly age-associated in a genome-wide study of Xenopus tropicalis.

extraction. Each frog was raised from birth in an aquatic tank with other frogs of the same age and strain. We collected samples from 26 distinct strains present across 34 physical tanks. The youngest animals in the dataset are 3-month-old tadpoles, and the oldest are 10.9-year-old adults. Supplementary Data 1 contains the age and strain information for each frog (See Appendix: Data and Code Availability).

We used targeted bisulfite sequencing (TBS) to profile DNA methylation [15]. We designed DNA probes to cover regions of the genome that were previously shown to contain CpG sites with methylation highly correlated with age in our whole genome bisulfite sequencing (WGBS) study of 9 *Xenopus tropicalis* frogs of three different ages [16] (see Methods section). The WGBS dataset was also derived from hindlimb webbing skin, and the ages ranged from 1 to 10 years old. The 3,443 targeted probe regions are available in Supplementary Data 2. Bisulfite-converted reads underwent quality control, trimming, and alignment to the v10.0 reference genome [17] using BiSulfite Bolt [18] (see Methods section). In total, we measured 6,717 CpGs with sequencing coverage of at least 100x in all 192 frogs. Among the CpG sites shared between the whole genome and targeted bisulfite sequencing datasets, the age correlations calculated under the WGBS and TBS datasets have a significant positive correlation (Pearson r=0.37, p=5.6×10^{-113}, n=3506, Figure 2.2). As expected, many sites in our TBS data display a strong positive or negative correlation with age (see two examples in Figure 2.3).

We noted that some CpG sites show a characteristic pattern with 3 separate levels of methylation across samples (Figure 2.3C). This pattern of methylation is likely caused by germline C to T mutations at these genomic positions across some of our strains [19]. We identified these sites as those with methylation values that had a better fit in a Gaussian Mixture Model (GMM) with 3 components than a GMM with 1 (see Methods section). After filtering out these sites, 6,183 CpG sites remained for further analysis. Principal component analysis of our TBS methylation data for each frog shows that age is highly correlated with the first principal component (r = 0.768, p=1.34×10^{-37}, n=187, Figure 2.4A). In the WGBS

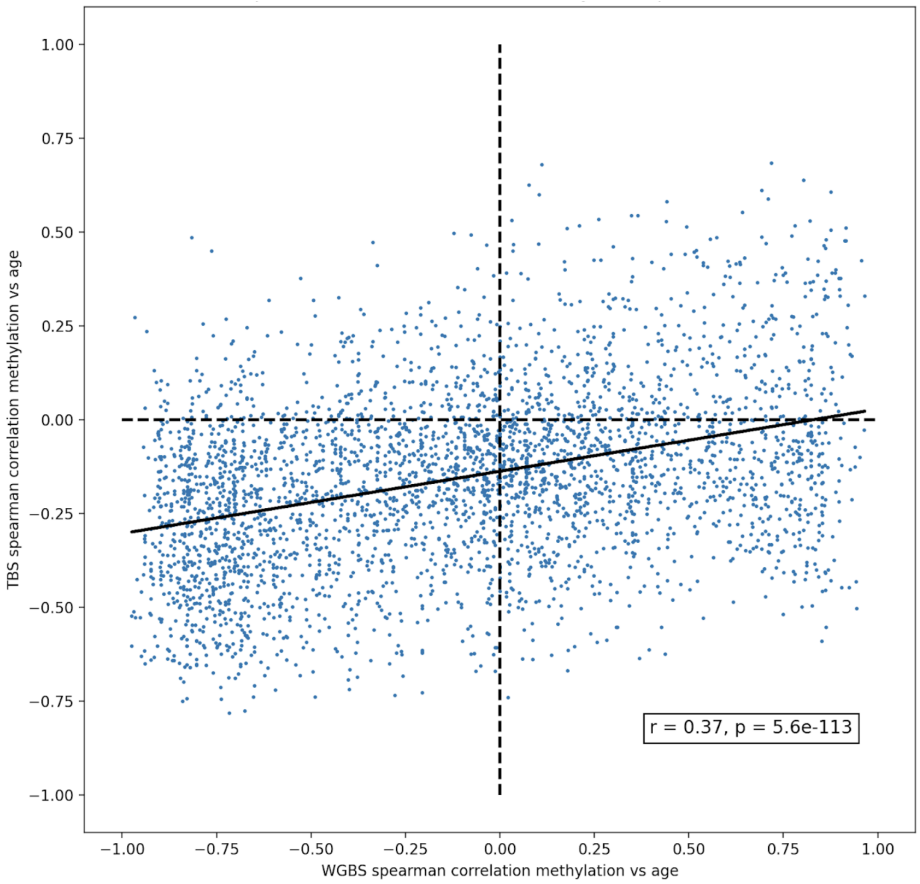

Figure 2.2: Comparison of n=3506 CpG sites shared between the previously collected whole genome bisulfite sequencing (WGBS) dataset and this study's targeted bisulfite sequencing (TBS) dataset. The Spearman correlation between methylation and age for the corresponding dataset is shown on each axis. The Pearson correlation between the Spearman correlations (WGBS vs TBS) in the plot is r=0.37 ($p=5.6\times10^{-113}$, n=3506). The solid black line shows the least squares regression best-fit line.

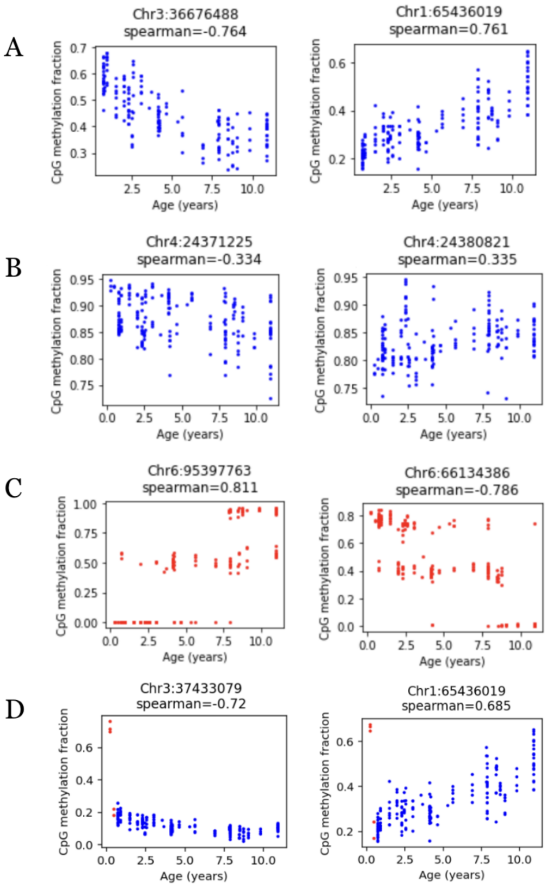

Figure 2.3: (A) Examples of CpG sites with highly age-correlated methylation (Spearman rs $\approx \pm 0.75$). (B) Examples of CpG sites with lower correlation between methylation and age (Spearman rs $\approx \pm 0.33$). (C) Examples of CpG sites with a pattern of 3 distinct methylation levels. These are filtered out of downstream analyses. (D) Examples of CpG sites with abnormally high methylation levels in tadpoles (red). The tadpole samples are removed for some downstream analyses.

data, PC1 had a stronger association with age (R^2=0.98) [16]. The average methylation level across all CpGs for each frog in our data tends to decrease with age (Pearson r=-0.466, p=9.3×10^{-12}, n=192, Figure 2.4B).

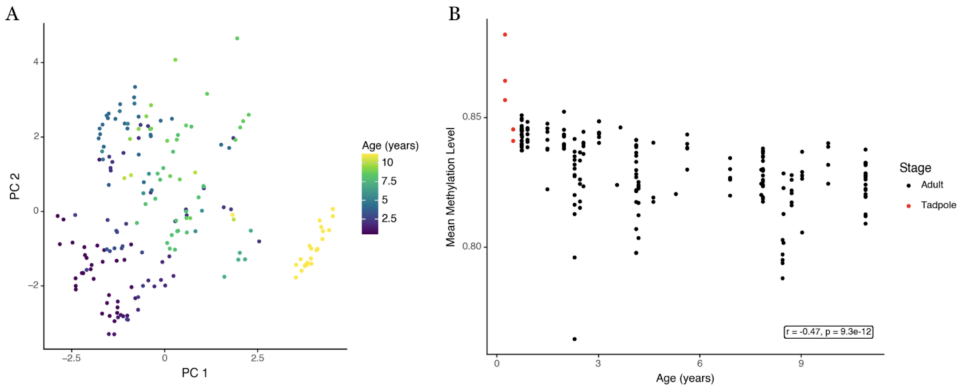

Figure 2.4: (A) Principal component analysis (PCA) of the DNA methylation values for each frog. Each CpG site's methylation value was centered to the mean across all frogs before PCA. (B) Mean methylation level across all targeted bisulfite sequencing CpG sites vs age. CpG sites with a minimum coverage of 100 were used for both plots.

To better characterize the genetic differences between frogs in our data, we called genetic variation from the targeted bisulfite sequencing data using BiSulfite Bolt. The matrix of 3,252 SNP differences is available in Supplementary Data 3. We calculated genetic variation as the number of SNP differences between frogs, divided by the total number of callable sites (1,003,455). Figure 2.5A shows the average genetic variation between every pair of tanks. As expected, the intra-tank genetic variation tends to be smaller than the inter-tank genetic variation. Figure 2.5B shows the genetic variation between individuals. While we found up to a maximum of 0.056% genetic variation between frogs in the dataset, our clocks are still able to accurately predict age across strains.

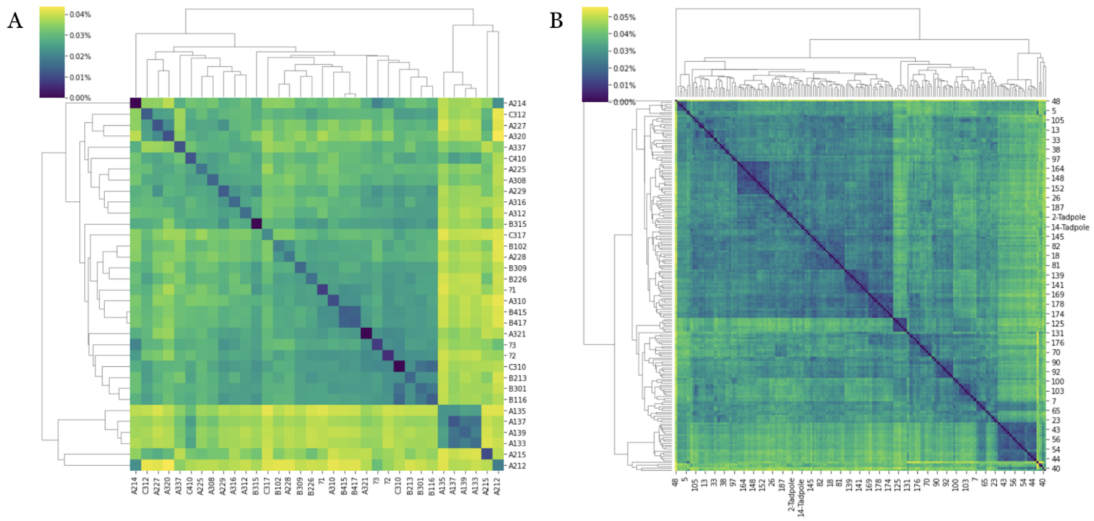

Figure 2.5: (A) Genetic variation between every pair of tanks (average number of SNP differences / number of callable base pairs). (B) Genetic variation between every pair of frogs (number of SNP differences / number of callable base pairs).

2.2 Epigenetic clock

We trained elastic net regularized linear regression models to predict chronological age based on DNA methylation levels. To evaluate this method, we performed cross-validation on the training dataset (Figure 2.6A-D), and trained a final model on all the training data (n=192) to predict the ages of 16 additional frogs in an external test set (n=16) (Figure 2.6E). In our dataset, all frogs within one tank are the same age and strain, so the leave-one-out cross-validation procedure (LOO-CV) (Figure 2.6A-B) includes frogs from the same tank in both the internal training and testing splits, which could lead to overfitting due to tank-specific features. To evaluate the model's performance without a tank-related bias, we also performed leave-one-tank-out cross-validation (LOTO-CV) (Figure 2.6C-D). In this procedure, all frog samples from one tank are left out of the training data, then a clock is trained on all samples from the other tanks, and predictions are made on each frog in the left-out tank. The dataset includes 34 tanks, each containing between 1 and 19 frogs, with a median of 6 frogs per tank.

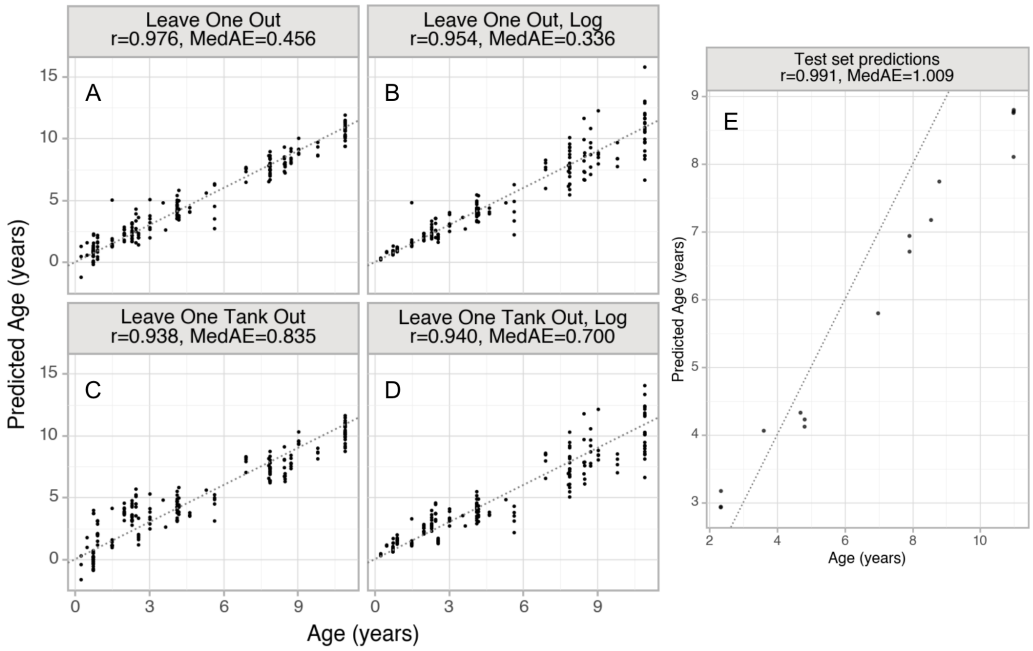

Figure 2.6: Epigenetic clock performance. (A-D) DNA methylation clock nested cross validation results on n=192 frogs. Models (B) and (D) were constructed to predict the natural logarithm of age. The dotted lines represent predicted age = actual age. MedAE is the median absolute error between predicted and actual age (in years). (E) Age predictions on a test set of 16 frogs based on a DNA methylation clock trained on the 192 frog training set.

To measure the performance of the clocks, we used the median absolute error between the predicted age and actual age (MedAE), and the Pearson correlation between the predicted age and actual age (r). As expected, the LOTO-CV predictions (MedAE = 0.835 years and r = 0.938) performed worse than LOO-CV predictions (MedAE = 0.456 years and r = 0.976). For comparison, the only other DNA methylation clock currently published in *Xenopus tropicalis* [12] achieved a LOO-CV MedAE of 0.96 years. We also report clocks that predict the natural logarithm of age (Figure 2.6B,2.6D), because previous work has suggested that DNA methylation levels tend to be non-linearly related to age [20,21]. The predictions of our log-transformed clocks show more accurate predictions for young samples and less accurate predictions for old samples.

For the elastic net model trained on the entire training dataset and evaluated on the 16 frog external test set, 378 CpG sites were selected as predictors (199 positively and 179 negatively associated with age). These CpG sites and the corresponding clock coefficients are provided in Supplementary Data 4. The external test set clock predictions (Figure 2.6E) display a very strong linear relationship between the predicted age and actual age (Pearson r = 0.991, p=1.065×10^{-13}, n=16). However, the slope of the predicted vs actual age plot is less than 1 (MedAE = 1.01 years). This reduction in the range of the predicted ages relative to the actual ages is a common feature of elastic net clocks when applied to validation data [22]. These 16 frogs were collected and processed separately from the 192 frog training data. However, it should be noted that some frogs in the test set came from the same tank and strain as the training data, so this evaluation shows generalization to different sample preparation and sequencing runs, but not necessarily generalization to entirely unseen frog strains and ages.

To further investigate the effects of including frogs of the same age and strain in both the training and prediction cross validation folds, we performed LOTO cross validation except with all the tank assignments randomized. This shuffled LOTO-CV procedure, when run 5 times with different tank membership assignments, achieved an average MedAE 0.470

years, and the log-transformed clock achieved MedAE of 0.398 years. The performance of the shuffled LOTO-CV is better than the original LOTO-CV, and similar to the LOO-CV results.

2.3 Chromatin state analysis of highly age-associated CpGs

We sought to investigate whether highly age-correlated CpG sites are preferentially found in certain chromatin states. In the absence of an established chromatin state annotation for the *Xenopus tropicalis* v10.0 genome [17], we used ChromHMM [23] to construct one based on a previously published ChIP-seq dataset collected from *Xenopus tropicalis* embryos [24]. Our annotation is based on a stage 30 embryo from the published ChIP-seq data. The model generated 15 chromatin states which we assigned names based on previously reported chromatin states in other species. Figure 2.7 shows the proportions of the top positively and negatively age-correlated CpGs in each of the annotated chromatin states.

Significantly age-correlated CpGs were selected based on adjusted p-values for the correlation between methylation and age (see Methods section). This generated a list of the top 326 positively age-correlated sites and 1907 negatively age-correlated sites. Most CpG sites in our data are found in the states corresponding to heterochromatin, quiescent, or repetitive regions of the genome. Figure 2.8A shows the emission probabilities output by ChromHMM. Figure 2.8B shows the log2 enrichment of the significantly age-correlated CpGs in each chromatin state, where enrichment is defined as the fraction of highly age-correlated CpG sites in a chromatin state divided by the fraction of all CpG sites in that chromatin state. We used Fisher's exact test to compute a p-value for each enrichment. Although there are a small total number of CpGs in the "bivalent low" and "strongly transcribed" chromatin states in our dataset, these sites show the greatest enrichment for positively age-correlated CpGs. The heterochromatin1 state (moderate H3K9me3 signal) shows a moderate underrepresentation of highly age-correlated sites, and the heterochromatin2 state (strong

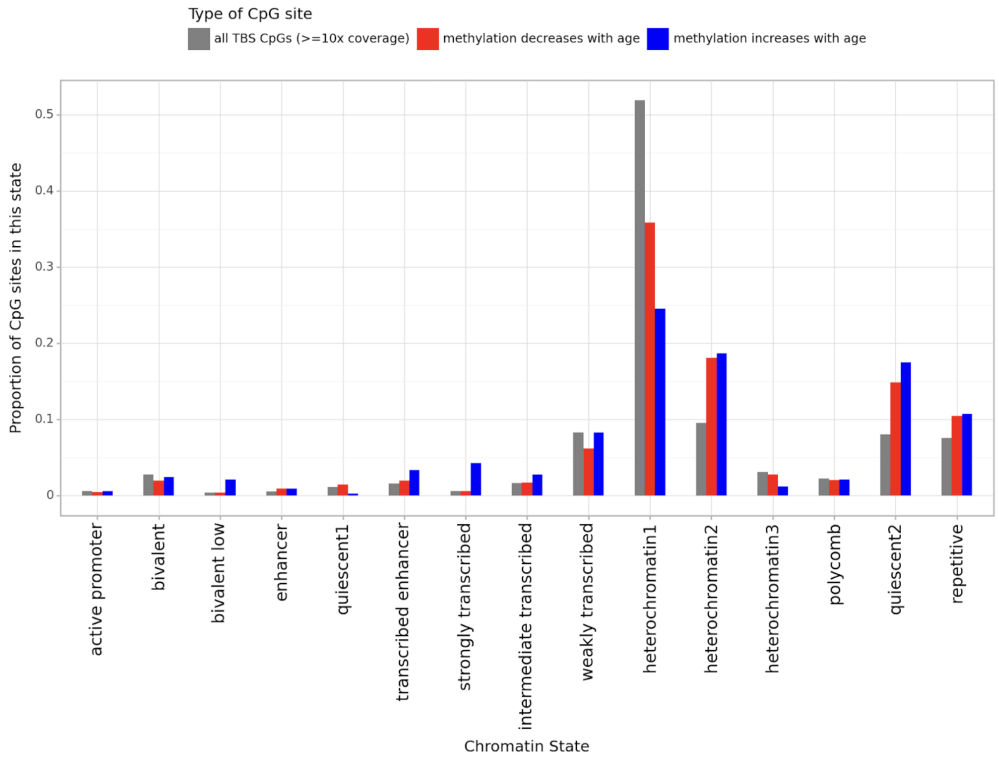

Figure 2.7: Proportion of CpG sites in each chromatin state for 3 groups of CpGs. There are 25,362 CpGs in the TBS dataset with $\geq$ 10x coverage (gray), 1,907 highly negatively age-associated CpGs (red), and 326 highly positively age-associated CpGs (blue).

H3K9me3 and H4K20me3, low H3K9me2) shows a moderate overrepresentation of highly age-correlated sites; these heterochromatin effects are each significant under the Fisher's exact test (Benjamini-Hochberg adjusted $p < 10^{-5}$).

2.4 Gene expression proximal to highly age-associated CpGs

We investigated the relationship between age-associated CpG sites and the expression of proximal genes. To accomplish this, we used the previously published *Xenopus* single-cell RNA-seq atlas [25]. The atlas includes gene expression data from 1-year-old frogs from the related species *Xenopus laevis*. For the purposes of this analysis, we mapped genes in X. tropicalis to X. laevis. We selected the significantly age-correlated CpGs in our data, and found the most proximal gene to each CpG site, generating a list of 37 positively- and 208 negatively age-correlated genes (see Methods section). Figure 2.9 shows the mean expression of these gene lists in each tissue, along with the mean expression of all genes in *Xenopus laevis* as a reference. The genes proximal to CpG sites with significant positive age association tend to be less expressed than the background of all genes, but due to the low sample size (37 positive genes), this effect is not significant under a Wilcoxon rank-sum test (Benjamini-Hochberg p-adj. > 0.05). The genes proximal to CpGs with significant negative age association are significantly more expressed than the background of all genes in brain, eye, heart, lung, muscle, ovary, pancreas, and stomach (Benjamini-Hochberg p-adj. < 0.05).

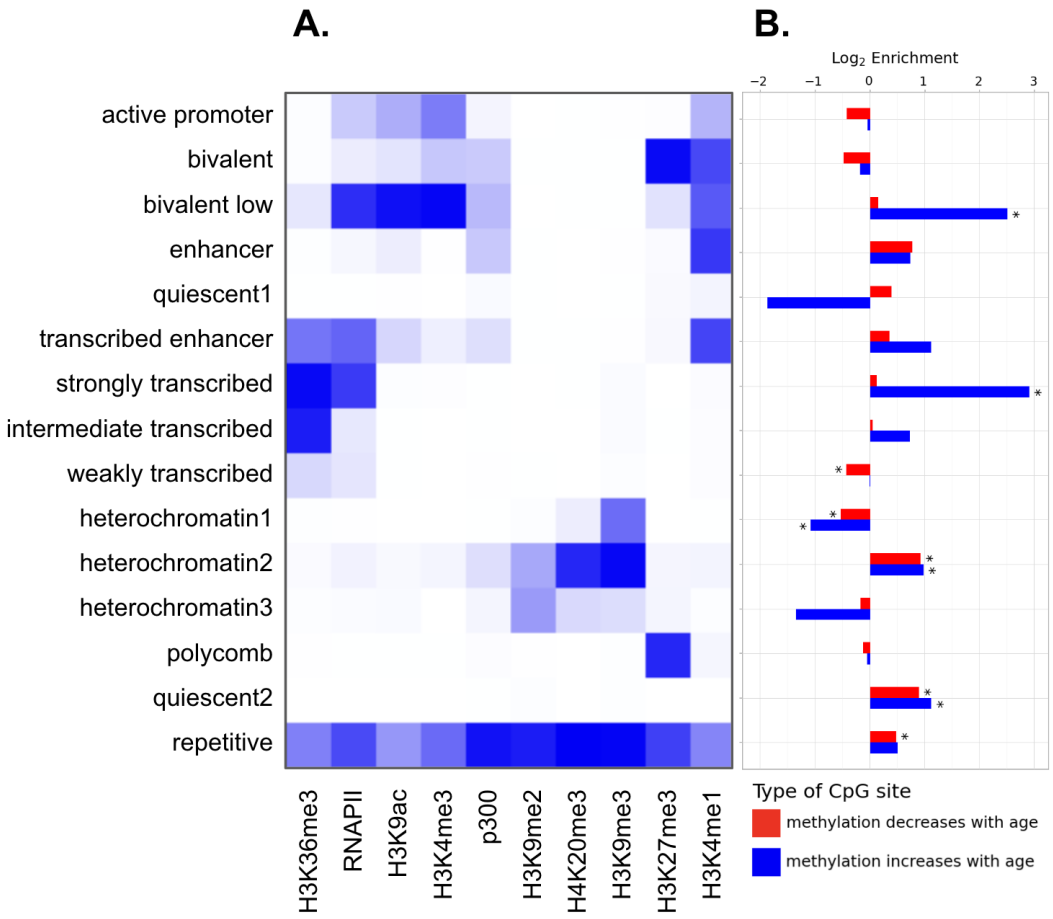

Figure 2.8: Chromatin state analysis of highly age-associated CpGs. (A) ChromHMM automated chromatin state annotation emission parameters. (B) Chromatin states of selected CpG sites with high correlation between methylation and age. Log enrichment represents log2 of the fraction of + or - CpG sites in a chromatin state divided by the fraction of all CpG sites in that chromatin state. The reference background is all CpG sites from our TBS dataset with sequencing coverage of at least 10. An asterisk (*) indicates that the adjusted p-value < 0.05 for the Fisher's exact test.

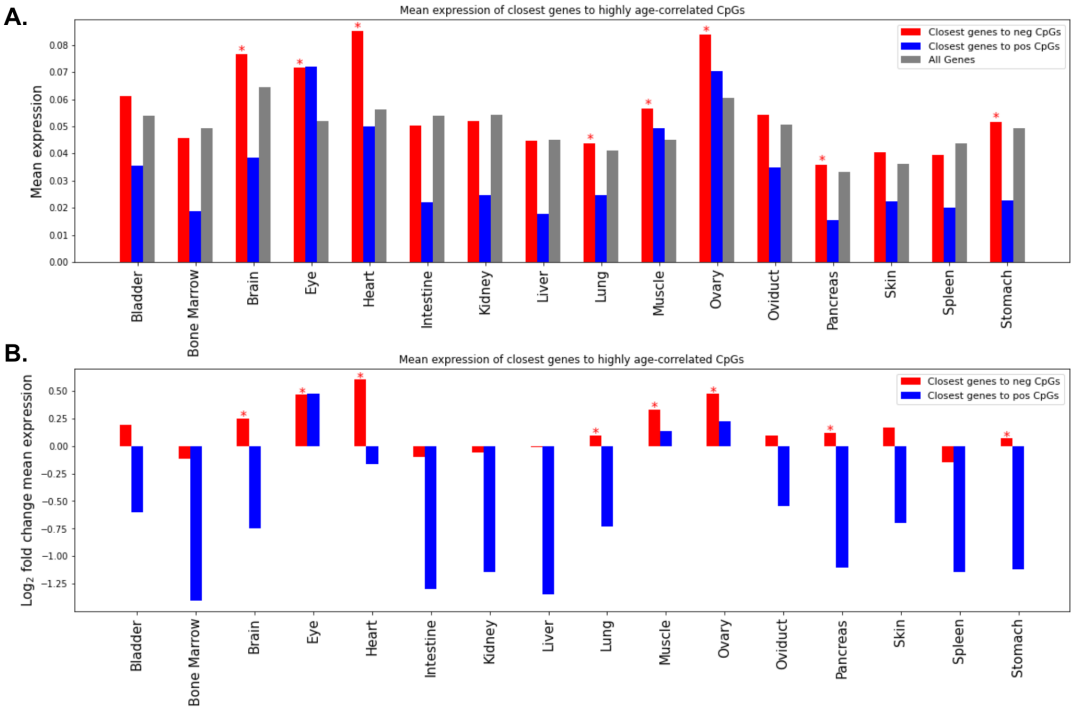

Figure 2.9: Expression of genes proximal to highly age-associated CpGs. (A) Single-cell RNA sequencing atlas expression levels of the closest gene to significantly negatively correlated CpGs (red), significantly positively correlated CpGs (blue), and all genes (gray). Mean expression is calculated as the mean of log1pCP10k values over all cells in a tissue. Asterisks (*) indicate Benjamini–Hochberg adjusted p-value is < 0.05 for Wilcoxon rank-sum test of mean expression in the gene list compared to mean expression of all genes. (B) Log2 fold change in mean expression of genes closest to significantly age-associated CpGs compared to all genes.

CHAPTER 3

Discussion

We constructed the most accurate clock for *Xenopus tropicalis*, based on 192 frogs. We expect the performance of this clock on new frog samples will be similar to the leave-one-tank-out cross-validation (LOTO-CV) performance metrics of MedAE=0.7 years and r=0.94. While the LOO-CV performance is better than LOTO-CV (MedAE=0.465 years, r=0.976), we were able to achieve similar performance to LOO-CV using a LOTO-CV approach with random tank membership assignments. This suggests that the LOO-CV procedure is overfitting to the age and/or shared genetic information from within each tank.

We identified chromatin states enriched for age-associated CpG sites using a chromatin state annotation based on a previously published ChIP-seq dataset in *Xenopus tropicalis* embryos [24]. The chromatin state primarily marked with H3K9me3 ("heterochromatin1") makes up over 50% of the CpG sites in our dataset. We observed a significant under-representation of highly age-associated sites in the heterochromatin1 state. This state is likely a type of constitutive heterochromatin, and this result indicates methylation levels in these regions tend to be more stable with age than in other chromatin states. The "heterochromatin2" state is marked primarily with H3K9me3 and H4K20me3, and is significantly enriched for both significantly positive and negative age-associated CpG sites. The difference in the age-associated methylation behavior of these two chromatin states suggests a significant role for H4K20me3 in epigenetic aging. The histone marks H4K20me3 and H3K9me3 are directly linked to DNA methylation through the maintenance DNA methyltransferase DNMT1. DNMT1's BAH1 domain has been shown to recognize H4K20me3 in humans [26],

and the UHRF1 protein binds to H3K9me3 through its tandem Tudor domain and guides DNMT1 to these regions [27]. Therefore, changes of these histone marks with age could lead to age-associated DNA methylation changes at these loci. Large scale changes and redistribution of H3K9me3 and H4K20me3 have been observed with aging [28–30]. Since a large fraction of age-associated CpG sites in our dataset are in heterochromatic regions, it's possible that much of our clock's predictive ability can be explained by a redistribution of heterochromatin in *Xenopus* frogs with age. Direct measurement of the spatiotemporal evolution of these chromatin marks over aging is a possible future research direction. This will be especially important given that our chromatin state annotation was derived from whole embryos (the only publicly available genome-wide ChIP-seq data for *Xenopus tropicalis* that we could identify), which differ in both developmental stage and tissue composition from the adult skin samples used for DNA methylation profiling. Future studies using ChIP-seq or ATAC-seq in adult *Xenopus* skin tissue would provide a more accurate picture of aging-associated chromatin dynamics and avoid potential biases introduced by using embryonic references, improving interpretation of the functional relevance of age-associated CpG sites.

We observed that positively age-correlated CpGs are significantly enriched in the state with H3K4me3 and H3K27me3 ("bivalent low"). This enrichment has been reported in human [10,31] and mammalian studies [11]. This effect has also been observed in *Xenopus tropicalis* and *laevis* [12] using chromatin state annotations for the bivalent sites from human datasets. Loss of H3K4me3 at chromatin that was formerly bivalent is a possible mechanism explaining the increase in DNA methylation with age. As H3K4me3 at bivalent sites is lost with age, we expect DNA methylation to increase at these sites because the de novo DNA methyltransferase machinery (DNMT3a/b/L) can bind to H3K4me0 (and not H3k4me3) through the ADD domain and promote de novo DNA methylation [32,33]. However, we note that less than 1% of CpG sites in our targeted bisulfite sequencing dataset are in this state (see Figure 2.7). The targeted bisulfite sequencing method with different probes could allow for direct targeting of these sites in further studies.

The two chromatin states with the highest H3K27me3 levels ("polycomb" and "bivalent") do not have significant enrichment for age-associated DNA methylation sites. This result contrasts with mammalian studies which report positively age-associated methylation at polycomb group target sites [10,34–37]. While a typical annotation of this chromatin state is defined as regions of the genome highly bound by PRC2 in embryonic stem cells, our annotation is based on measuring H3K27me3 in DNA from a whole *Xenopus* embryo.

We also observe significant enrichment for the top positively age-associated CpGs in chromatin with H3K36me3. H3K36me3 is found in gene bodies of actively transcribed genes, and maintenance of this mark prevents cryptic transcription [38]. DNMT3a's PWWP domain interacts with H3K36me3, promoting de novo DNA methylation at these regions [39–41]. Similar to our results, DNA methylation levels have been found to increase with age in H3K36me3 regions in mouse muscle stem cells [42].

Using a single cell RNA-seq dataset from a young adult *Xenopus laevis* [25], we observed low expression of genes nearby to CpGs with positively age associated methylation. This effect has been previously observed in humans and mice [10,31,43–45]. The low expression levels of genes proximal to positively age-correlated CpGs suggest that the increase in methylation at these sites over age may generally not lead to significant phenotypic changes because these genes are already repressed by epigenetic mechanisms other than DNA methylation. This is consistent with the fact that positively age-associated CpGs have been found near developmental genes in humans [9,11] and *Xenopus* [12].

In conclusion, in this study we have shown that some mechanisms of epigenetic aging may be conserved between mammals and *Xenopus tropicalis*, supporting the future use of frogs as a model organism for the study of biological aging. An advantage of using amphibians in aging research is that they are cold blooded. The surrounding water temperature could be altered in a controlled experiment to determine the effect of metabolic rate on the rate and patterns of DNA methylation over time. Sexual maturity is well understood in *Xenopus*; specifically, metamorphosis can be easily controlled by iodine in the environment or by

exogenous thyroid hormone. The delayed onset of sexual maturity increases lifespan in mice [46]. Further study of the molecular mechanisms of this process could be undertaken in *Xenopus*. While we did not consider the effect of sex in this study, future studies could examine lifespan differences between sexes using our clocks. Male-to-female and female-to-male sex reversal can be easily induced by exposing developing larvae to estrogen and the aromatase inhibitor fadrozole respectively [47]. There are also environmental factors that would be challenging to study in other animals, for example, the effects of environmental plastic exposure, which can be examined in *Xenopus* [48]. Moreover, *Xenopus* is thought to have negligible senescence and remains fertile late in life. In fact, it has been suggested that there is a lack of reproductive aging in *Xenopus* [49], which combined with our clocks could reveal new reproductive biology. It is known that wood frogs (Lithobates sylvaticus) are naturally freeze-tolerant [50]. If clocks translate to these species, exploring the clocks in frozen tissues presents a new research direction. Since single cell nuclear transfer was originally accomplished in Rana and *Xenopus* frogs [51], *Xenopus* provides a convenient toolbox for studying the germ line reset. In conclusion, *Xenopus* frogs provide promising research directions to further study the biology of aging.

CHAPTER 4

Methods

4.1 Sample collection

192 *Xenopus tropicalis* frogs were housed at the National Xenopus Resource (RRID: SCR_013731) [52] in multi-rack recirculating aquatic systems with a constant temperature of 25C and 12 hr light /12 hr dark cycle, and established diet as previously described [16,53,54]. Disposable biopsy punches were used to collect tissue (VWR 21,909–140) from the hindlimb webbing of each of the 187 adult frogs in the dataset [55]. For the 5 tadpoles, the entire organism was used for DNA extraction. Supplementary Data 1 includes all frog metadata, including strain. All animals have a Nigerian strain genetic background (NXR_1018). Some frogs in the dataset are transgenic; this information is given in the strain annotation. The sex of the frogs was not recorded. We performed DNA extraction according to the procedures described in [16], which was modified from [55].

4.2 Targeted bisulfite sequencing library preparation

Library preparation was performed as previously described in [16]. 500 nanograms of purified DNA per sample was sonicated using a Covaris R230. Fragmented DNA was subject to end-repair, dA-tailing and ligation to pre-methylated unique-dual-indexed custom adapters (Integrated DNA Technologies). Pooled samples (16 samples per pool) were subject to hybridization with biotinylated oligonucleotide probes (synthesized by Integrated DNA Technologies) complementary to the previously identified age-associated regions (Supplementary

Data 2) in the presence of *Xenopus* Hybloc DNA (Applied Genetics Laboratories - Melbourne, FL, US). Captured fragments were then treated with bisulfite (Zymo Lightning), then amplified and purified according to Morselli et al. [16]. Samples were subject to QC, normalized and combined in two pools of 96 samples, and each sequenced on NovaSeq6000 (SP lanes) as paired-end 150 bases.

4.3 Targeted bisulfite sequencing data processing

The X. *tropicalis* v10.0 reference genome [17] was downloaded from Xenbase [56] (RRID: SCR_003280). We performed sequencing, alignment to the reference genome, quality control, and methylation calling using the same methods previously described in [16], except reads were trimmed using fastp [57] with TruSeq adapters trimmed and options –length_required 40 –trim_front1 5 –trim_tail1 5 –trim_front2 5 –trim_tail2 5 –trim_poly_g –qualified_quality_phred 20. Briefly, BiSulfite Bolt [18] was used to align bisulfite-converted reads to the reference genome and create a methylation matrix of all CpG sites with sequencing coverage of at least 100x in 100% of the samples. There are 6,717 of these CpGs. Mean coverage of each sample (see Supplementary Data 1) was calculated with BSBolt AggregateMatrix (arguments: -count -min-coverage 100). We used scipy.stats.spearmanr [58] to compute the correlation between methylation and age at each site. For genotype analysis, we used the BiSulfite Bolt commands CallVariation and GenotypeMatrix to call variant genotypes from sites with a minimum of 10x coverage, alignment score 10, quality score 10, and present in at least 90% of frogs. Some CpG sites in the data show an abnormal pattern of methylation, where some frogs have 0 methylation, some have 0.5 methylation, and some have 1.0 methylation, with very few values in between. We used the function sklearn.mixture.GaussianMixture [59] to fit Gaussian Mixture Models (GMM) to the methylation values at each CpG. A CpG site is filtered from the dataset if the Bayes Information Criterion (BIC) for the GMM with 1 component minus the BIC for the GMM with 3 com-

ponents is greater than 90. The value of 90 was tuned manually to capture sites with clearly anomalous patterns. Additionally, for a site to be filtered there must be at least one vanishingly small methylation value (<0.006), since we expect the anomalous sites resulting from C to T germline mutation to have no methylation for frogs with the T allele. Examples of these sites are available in Figure 2.3B. After filtering, our methylation matrix consisted of 6,183 CpG sites across 192 frogs.

Figures were generated by Python packages plotnine v0.12.4 [60] and matplotlib v.3.8.2 [61]. Tabular data were processed by Python packages pandas v2.1.4 [62] and anndata v0.8.0 [63]. GPT-4o was used to assist in writing some Python code for the analysis and plotting.

4.4 Epigenetic clock

We implemented the clocks using the function sklearn.linear_model.ElasticNetCV from the Scikit-learn package in Python [59]. The l1_ratio hyperparamter controls the relative strength of the L1 penalty compared to the L2 penalty, and it was fixed to 0.5 for the clocks. The alpha hyperparameter controls the overall strength of regularization. We used a nested cross validation approach to tune the alpha hyperparameter and obtain an estimate of the method's performance. In this procedure, the n=192 frog training dataset is first split up into a training and validation set, based on the "outer cross validation" criterion of either LOO-CV or LOTO-CV. Within the training set of this cross validation, the ElasticNetCV function was used with an "inner cross validation" criterion to find the best alpha value within the training fold. For LOO-CV, the inner cross validation criterion was 5-fold cross validation. For LOTO-CV, the inner cross validation criterion was also LOTO-CV. Sklearn.model_selection.LeaveOneOut and sklearn.model_selection.LeaveOneGroupOut were used to implement these procedures.

Additionally, we used previously published (GEO Series GSE222107) [16] targeted bisulfite sequencing data from 16 different frogs as an external testing set for evaluating the

performance of our DNA methylation clock. This external dataset was prepared by a different operator and sequenced on a different NovaSeq6000 lane. We trained a single clock on the subset of CpG sites shared between the main dataset and the test dataset (reducing the number of CpG sites from 6,183 to 5,938), then made predictions on the test dataset (Figure 2.6E). For this model, the function sklearn.linear_model.ElasticNetCV was used with internal LOTO-CV to tune the alpha hyperparameter.

4.5 Selecting highly age-associated CpGs

From the 6,183 filtered CpG sites in our data, we selected the top sites most positively and negatively correlated with age for downstream analysis. We ranked the CpG sites based on their Spearman correlation between methylation value and age. For this site selection process, we included methylation measurements only from the 187 adult frogs and excluded tadpoles due to their substantially higher methylation values at some CpGs (see Figure 2.3C). The p-value for this correlation was calculated with the Python function scipy.stats.spearmanr [58]. This p-value represents the null hypothesis that the Spearman correlation is 0. Highly age-associated sites were selected as those with a Benjamini–Hochberg adjusted p-value of less than 0.001. This approximately corresponds to a Spearman rs $>$ 0.3 or rs $<$ -0.3. The result is 326 positively age correlated sites and 1907 negatively age correlated sites.

4.6 Chromatin state analysis of highly age-associated CpGs

ChIP-seq data from Hontelez et al. were obtained from GEO Series GSE67974 [24]. The chromatin marks present in this dataset are H3K36me3, RNAPII, H3K9ac, H3K4me3, p300, H3K9me2, H4K20me3, H3K9me3, H3K9me3, H3K27me3, and H3K4me1. We used data from the stage 30 *Xenopus tropicalis* embryo. Cutadapt v2.10 (arguments: -u 5 -j 0) was

used to trim the input FASTQ files [64]. Bowtie2 v2.4.2 [65] (arguments: -k 1 –no-unal) was used with bowtie2-build to align ChIP-seq reads to the v10.0 X. *tropicalis* reference genome. Samtools v1.11 view, sort, markdup, and index were used to process bam files and remove PCR duplicates [66]. ChromHMM v1.24 ConvertGeneTable, BinarizeBam, LearnModel, MakeSegmentation, and Reorder were used with default arguments to create the chromatin state annotation [23]. The number of chromatin states used in the genome segmentation is a user-controlled hyperparameter; we settled on 15 states to achieve a balance between having enough states to capture biological variation, and not so many to create duplicate states. We manually annotated the names of the 15 chromatin states based on the chromatin marks associated with each state. To statistically test CpG site enrichment for each chromatin state, we used Fisher's exact test implemented with scipy.stats.fisher_exact, and adjusted the p-values using the Benjamini-Hochberg method with stats.false_discovery_control [59].

4.7 Gene expression proximal to highly age-associated CpGs

First, we mapped each significantly age-associated CpG site in our dataset to the closest transcription start site (TSS) in the X. *tropicalis* v10.0 genome annotation [17] using pybedtools.closest v0.9.0 (argument: -D b) [67,68]. We excluded all CpGs whose closest gene TSS is greater than 5 kilobases away, since we wanted to capture proximal methylation effects on the nearby gene. Next, we mapped each *Xenopus tropicalis* gene to its homologous gene in the *Xenopus* laevis v9.1 genome [69] using a HMMER-based pipeline, as described in [70]. The 326 highly positively age-associated CpG sites in our X. *tropicalis* dataset mapped to 37 nearby genes with data available in the X. *laevis* atlas. The 1907 highly negatively age-associated CpG sites in our X. *tropicalis* dataset mapped to 208 nearby genes with data available in the X. *laevis* atlas. As a background reference, all available genes in the X. *laevis* expression dataset (25,362 genes) were used. First, the mean expression of each gene was computed across all cells of each tissue. Next, the mean expression over the genes

in each of the three gene sets was calculated and displayed in Figure 2.9A. Figure 2.9B represents this same information by taking the log2 of the ratio between the expression of the positive or negative genes over the mean expression of all genes. The Python function scipy.stats.mannwhitneyu [58] was used within each tissue to test if the mean expression of the positive or negative gene sets are significantly different from the mean expression of all genes. scipy.stats.false_discovery_control [58] was used for the Benjamini-Hochberg p-value correction.

APPENDIX: DATA AND CODE AVAILABILITY

Bisulfite sequencing FASTQ files and the processed methylation matrix for 192 frogs are available at Gene Expression Omnibus (GEO) Series record GSE292839. `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE292839`

All code for the analysis is available at `github.com/ronanbennett/xenopus-aging`.

Supplementary Data 1-6 are publicly accessible at `https://www.nature.com/articles/s41514-025-00236-x`.

Supplementary Data 1: *Xenopus* samples description

Supplementary Data 2: Targeted bisulfite sequencing probes

Supplementary Data 3: Genetic variation

Supplementary Data 4: Epigenetic clock weights

Supplementary Data 5: Fisher's exact test results

Supplementary Data 6: Wilcoxon rank-sum test results

BIBLIOGRAPHY

[1] Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14 (10):R115.

[2] Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai PC, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. Aging [Internet]. 2016 Sep 28;8(9):1844–59. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5076441/

[3] Oblak L, van der Zaag J, Higgins-Chen AT, Levine ME, Boks MP. A systematic review of biological, social and environmental factors associated with epigenetic clock acceleration. Ageing Res Rev [Internet]. 2021 Aug 1;69:101348. Available from: https://www.sciencedirect.com/science/article/pii/S1568163721000957

[4] Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet [Internet]. 2018 Jun;19(6):371–84. Available from: https://www.nature.com/articles/s41576-018-0004-3

[5] Petkovich DA, Podolskiy DI, Lobanov AV, Lee SG, Miller RA, Gladyshev VN. Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. Cell Metab [Internet]. 2017 Apr 4;25(4):954-960.e6. Available from: https://www.cell.com/cell-metabolism/abstract/S1550-4131(17)30168-7

[6] Olova N, Simpson DJ, Marioni RE, Chandra T. Partial reprogramming induces a steady decline in epigenetic age before loss of somatic identity. Aging Cell [Internet]. 2018 Nov 18;18(1):e12877. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6351826/

[7] Maierhofer A, Flunkert J, Oshima J, Martin GM, Haaf T, Horvath S. Accelerated epigenetic aging in Werner syndrome. Aging [Internet]. 2017 Apr 4;9(4):1143. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC5425119/

[8] Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, et al. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. Aging [Internet]. 2018 Jul 26;10(7):1758. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6075434/

[9] Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res [Internet]. 2010 Apr 1;20(4):434–9. Available from: https://genome.cshlp.org/content/20/4/434

[10] Moqri M, Cipriano A, Simpson DJ, Rasouli S, Murty T, de Jong TA, et al. PRC2-AgeIndex as a universal biomarker of aging and rejuvenation. Nat Commun [Internet]. 2024 Jul 16;15(1):5956. Available from: https://www.nature.com/articles/s41467-024-50098-2

[11] Lu AT, Fei Z, Haghani A, Robeck TR, Zoller JA, Li CZ, et al. Universal DNA methylation age across mammalian tissues. Nat Aging [Internet]. 2023 Sep;3(9):1144–66. Available from: https://www.nature.com/articles/s43587-023-00462-6

[12] Zoller JA, Parasyraki E, Lu AT, Haghani A, Niehrs C, Horvath S. DNA methylation clocks for clawed frogs reveal evolutionary conservation of epigenetic aging. GeroScience [Internet]. 2023 Jun 4; Available from: https://doi.org/10.1007/s11357-023-00840-3

[13] Arneson A, Haghani A, Thompson MJ, Pellegrini M, Kwon SB, Vu H, et al. A mammalian methylation array for profiling methylation levels at conserved sequences. Nat Commun [Internet]. 2022 Feb 10;13(1):783. Available from: https://www.nature.com/articles/s41467-022-28355-z

[14] Horvath S. Fundamental equations linking methylation dynamics to maximum lifespan in mammals [Internet]. Epigenomics of Common Diseases Conference, organized by Wellcome Connecting Science; 2023 Nov 17; Hinxton, England. Available from: https://youtu.be/ct528EjCuxI?si=ytpIGxDqohWtvPYi&t=606

[15] Morselli M, Farrell C, Rubbi L, Fehling HL, Henkhaus R, Pellegrini M. Targeted bisulfite sequencing for biomarker discovery. Methods [Internet]. 2021 Mar 1;187:13–27. Available from: https://www.sciencedirect.com/science/article/pii/S1046202320300839

[16] Morselli M, Bennett R, Shaidani NI, Horb M, Peshkin L, Pellegrini M. Age-associated DNA methylation changes in Xenopus frogs. Epigenetics [Internet]. 2023 Dec 31;18(1): 2201517. Available from: https://doi.org/10.1080/15592294.2023.2201517

[17] Bredeson JV, Mudd AB, Medina-Ruiz S, Mitros T, Smith OK, Miller KE, et al. Conserved chromatin and repetitive patterns reveal slow genome evolution in frogs. Nat Commun [Internet]. 2024 Jan 17;15(1):579. Available from: https://www.nature.com/articles/s41467-023-43012-9

[18] Farrell C, Thompson M, Tosevska A, Oyetunde A, Pellegrini M. BiSulfite Bolt: A bisulfite sequencing analysis platform. GigaScience [Internet]. 2021 May 1;10(5):giab033. Available from: https://doi.org/10.1093/gigascience/giab033

[19] Liu Y, Siegmund KD, Laird PW, Berman BP. Bis-SNP: Combined DNA methylation and SNP calling for Bisulfite-seq data. Genome Biol [Internet]. 2012;13(7):R61. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3491382/

[20] Snir S, Farrell C, Pellegrini M. Human epigenetic ageing is logarithmic with time across the entire lifespan. Epigenetics [Internet]. 2019 Jun 6;14(9):912–26. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6691990/

[21] Dufek G, Katriel G, Snir S, Pellegrini M. Exponential dynamics of DNA methylation with age. J Theor Biol. 2024 Feb 21;579:111697.

[22] El Khoury LY, Gorrie-Stone T, Smart M, Hughes A, Bao Y, Andrayas A, et al. Systematic underestimation of the epigenetic clock and age acceleration in older subjects. Genome Biol [Internet]. 2019 Dec 17;20:283. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6915902/

[23] Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods [Internet]. 2012 Mar;9(3):215–6. Available from: https://www.nature.com/articles/nmeth.1906

[24] Hontelez S, van Kruijsbergen I, Georgiou G, van Heeringen SJ, Bogdanovic O, Lister R, et al. Embryonic transcription is controlled by maternally defined chromatin state. Nat Commun [Internet]. 2015 Dec 18;6(1):10148. Available from: https://www.nature.com/articles/ncomms10148

[25] Liao Y, Ma L, Guo Q, E W, Fang X, Yang L, et al. Cell landscape of larval and adult Xenopus laevis at single-cell resolution. Nat Commun [Internet]. 2022 Jul 25;13(1):4306. Available from: https://www.nature.com/articles/s41467-022-31949-2

[26] Ren W, Fan H, Grimm SA, Kim JJ, Li L, Guo Y, et al. DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation. Nat Commun [Internet]. 2021 May 3;12(1):2490. Available from: https://www.nature.com/articles/s41467-021-22665-4

[27] Rothbart SB, Krajewski K, Nady N, Tempel W, Xue S, Badeaux AI, et al. Association of UHRF1 with H3K9 methylation directs the maintenance of DNA methylation. Nat Struct Mol Biol [Internet]. 2012 Nov;19(11):1155–60. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3492551/

[28] Villeponteau B. The heterochromatin loss model of aging. Exp Gerontol [Internet]. 1997 Jul 1;32(4):383–94. Available from: https://www.sciencedirect.com/science/article/pii/S0531556596001556

[29] Tsurumi A, Li W. Global heterochromatin loss: A unifying theory of aging? Epigenetics [Internet]. 2012 Jul 1;7(7):680–8. Available from: https://doi.org/10.4161/epi.20540

[30] Sarg B, Koutzamani E, Helliger W, Rundquist I, Lindner HH. Postsynthetic Trimethylation of Histone H4 at Lysine 20 in Mammalian Tissues Is Associated with Aging *. J Biol

Chem [Internet]. 2002 Oct 18;277(42):39195–201. Available from: https://www.jbc.org/article/S0021-9258(19)72272-8/abstract

[31] Slieker RC, Relton CL, Gaunt TR, Slagboom PE, Heijmans BT. Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. Epigenetics Chromatin [Internet]. 2018 May 30;11:25. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5975493/

[32] Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. EMBO Rep [Internet]. 2009 Nov;10(11):1235–41. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2775176/

[33] Kyono Y, Sachs LM, Bilesimo P, Wen L, Denver RJ. Developmental and Thyroid Hormone Regulation of the DNA Methyltransferase 3a Gene in Xenopus Tadpoles. Endocrinology [Internet]. 2016 Dec 1;157(12):4961–72. Available from: https://doi.org/10.1210/en.2016-1465

[34] Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, et al. Aging effects on DNA methylation modules in human brain and blood tissue. Genome Biol [Internet]. 2012 Oct 3;13(10):R97. Available from: https://doi.org/10.1186/gb-2012-13-10-r97

[35] Dozmorov MG. Polycomb repressive complex 2 epigenomic signature defines age-associa ted hypermethylation and gene expression changes. Epigenetics [Internet]. 2015 Jun 3;10(6):484–95. Available from: https://doi.org/10.1080/15592294.2015.1040619

[36] Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome Res [Internet]. 2010 Apr 1;20(4):440–6. Available from: https://genome.cshlp.org/content/20/4/440

[37] Mozhui K, Pandey AK. Conserved effect of aging on DNA methylation and association with EZH2 polycomb protein in mice and humans. Mech Ageing Dev [Internet]. 2017 Mar;162:27–37. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5411177/

[38] Sen P, Dang W, Donahue G, Dai J, Dorsey J, Cao X, et al. H3K36 methylation promotes longevity by enhancing transcriptional fidelity. Genes Dev [Internet]. 2015 Jul 1;29(13):1362–76. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4511212/

[39] Morselli M, Pastor WA, Montanini B, Nee K, Ferrari R, Fu K, et al. In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. Ren B, editor. eLife [Internet]. 2015 Apr 7;4:e06205. Available from: https://doi.org/10.7554/eLife.06205

[40] Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, et al. The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation *. J Biol Chem [Internet]. 2010 Aug 20;285(34):26114–20. Available from: https://www.jbc.org/article/S0021-9258(20)59575-6/abstract

[41] Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature [Internet]. 2015 Apr;520(7546):243–7. Available from: https://www.nature.com/articles/nature14176

[42] Hernando-Herraez I, Evano B, Stubbs T, Commere PH, Jan Bonder M, Clark S, et al. Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. Nat Commun [Internet]. 2019 Sep 25;10:4361. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6761124/

[43] Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, et al. Genome-wide age-related changes in DNA methylation and gene expression in human

PBMCs. Age [Internet]. 2014 Jun;36(3):9648. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4082572/

[44] Yuan T, Jiao Y, de Jong S, Ophoff RA, Beck S, Teschendorff AE. An Integrative Multiscale Analysis of the Dynamic DNA Methylation Landscape in Aging. PLoS Genet [Internet]. 2015 Feb 18;11(2):e1004996. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4334892/

[45] Day K, Waite LL, Thalacker-Mercer A, West A, Bamman MM, Brooks JD, et al. Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. Genome Biol [Internet]. 2013 Sep 13;14(9):R102. Available from: https://doi.org/10.1186/gb-2013-14-9-r102

[46] Garratt M, Erturk I, Alonzo R, Zufall F, Leinders-Zufall T, Pletcher SD, et al. Lifespan extension in female mice by early, transient exposure to adult female olfactory cues. Benayoun BA, Isales C, Lu YX, editors. eLife [Internet]. 2022 Dec 16;11:e84060. Available from: https://doi.org/10.7554/eLife.84060

[47] Roco ÁS, Olmstead AW, Degitz SJ, Amano T, Zimmerman LB, Bullejos M. Coexistence of Y, W, and Z sex chromosomes in Xenopus tropicalis. Proc Natl Acad Sci [Internet]. 2015 Aug 25;112(34):E4752–61. Available from: https://www.pnas.org/doi/10.1073/pnas.1505291112

[48] Cai B, De Jesus Andino F, McGrath JL, Romanick SS, Robert J. Ingestion of polyethylene terephthalate microplastic water contaminants by Xenopus laevis tadpoles negatively affects their resistance to ranavirus infection and antiviral immunity. Environ Pollut [Internet]. 2024 Sep 1;356:124340. Available from: https://www.sciencedirect.com/science/article/pii/S0269749124010546

[49] Brocas J, Verzár F. The Aging of Xenopus laevis, a South African Frog. Gerontologia [Internet]. 2009 Mar 31;5(4):228–40. Available from: https://doi.org/10.1159/000211062

[50] Al-Attar R, Storey KB. Lessons from nature: Leveraging the freeze-tolerant wood frog as a model to improve organ cryopreservation and biobanking. Comp Biochem Physiol B Biochem Mol Biol. 2022;261:110747.

[51] Gurdon JB. Nuclear Transplantation in Xenopus. In: Pells S, editor. Nuclear Reprogramming: Methods and Protocols [Internet]. Totowa, NJ: Humana Press; 2006. p. 1–9. Available from: https://doi.org/10.1385/1-59745-005-7:1

[52] Pearl EJ, Grainger RM, Guille M, Horb ME. Development of Xenopus Resource Centers: the National Xenopus Resource and the European Xenopus Resource Center. Genes N Y N 2000 [Internet]. 2012 Mar;50(3):155–63. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3778656/

[53] McNamara S, Wlizla M, Horb ME. Husbandry, general care, and transportation of Xenopus laevis and Xenopus tropicalis. Methods Mol Biol Clifton NJ [Internet]. 2018;1865:1–17. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6421069/

[54] Shaidani NI, McNamara S, Wlizla M, Horb ME. Animal maintenance systems: Xenopus tropicalis. Cold Spring Harb Protoc [Internet]. 2020 Dec 1;2020(12):pdb.prot106146. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7666031/

[55] Xu Y, Guan T, Liu J, Su H, Zhang Z, Ning F, et al. An efficient and safe method for the extraction of total DNA from shed frog skin. Conserv Genet Resour [Internet]. 2020 Jun 1;12(2):225–9. Available from: https://doi.org/10.1007/s12686-019-01104-z

[56] Fisher M, James-Zorn C, Ponferrada V, Bell AJ, Sundararaj N, Segerdell E, et al. Xenbase: key features and resources of the Xenopus model organism knowledgebase. Genetics [Internet]. 2023 May 2;224(1):iyad018. Available from: https://doi.org/10.1093/genetics/iyad018

[57] Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication

using fastp. iMeta [Internet]. 2023;2(2):e107. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/imt2.107

[58] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods [Internet]. 2020 Mar;17(3):261–72. Available from: https://www.nature.com/articles/s41592-019-0686-2

[59] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011 Nov 1;12(null):2825–30.

[60] Kibirige H, Lamp G, Katins J, gdowding, austin, Finkernagel F, et al. has2k1/plotnine: v0.14.3 [Internet]. Zenodo; 2024. Available from: https://zenodo.org/records/14224336

[61] Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Eng [Internet]. 2007 May;9(3):90–5. Available from: https://ieeexplore.ieee.org/document/4160265

[62] team T pandas development. pandas-dev/pandas: Pandas [Internet]. Zenodo; 2024. Available from: https://zenodo.org/records/13819579

[63] Virshup I, Rybakov S, Theis FJ, Angerer P, Wolf FA. anndata: Access and store annotated data matrices. J Open Source Softw [Internet]. 2024 Sep 16;9(101):4371. Available from: https://joss.theoj.org/papers/10.21105/joss.04371

[64] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal [Internet]. 2011 May 2;17(1):10–2. Available from: https://journal.embnet.org/index.php/embnetjournal/article/view/200

[65] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods [Internet]. 2012 Apr;9(4):357–9. Available from: https://www.nature.com/articles/nmeth.1923

[66] Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience [Internet]. 2021 Feb 1;10(2):giab008. Available from: https://doi.org/10.1093/gigascience/giab008

[67] Dale RK, Pedersen BS, Quinlan AR. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. Bioinformatics [Internet]. 2011 Dec 15;27(24):3423–4. Available from: https://doi.org/10.1093/bioinformatics/btr539

[68] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics [Internet]. 2010 Mar 15;26(6):841–2. Available from: https://doi.org/10.1093/bioinformatics/btq033

[69] Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, et al. Genome evolution in the allotetraploid frog Xenopus laevis. Nature [Internet]. 2016 Oct;538(7625):336–43. Available from: https://www.nature.com/articles/nature19840

[70] Van Itallie E, Sonnett M, Kalocsay M, Wühr M, Peshkin L, Kirschner MW. Transitions in the proteome and phospho-proteome during Xenopus laevis development. Dev Biol. 2025 Jun 2;S0012-1606(25)00145-9.