A DNA methylation age predictor for zebrafish

Benjamin Mayne[1], Darren Korbie[2], Lisa Kenchington[3], Ben Ezzy[3], Oliver Berry[1], Simon Jarman[4]

[1]Environomics Future Science Platform, Indian Ocean Marine Research Centre, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Crawley, Western Australia, Australia
[2]Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, St Lucia, Queensland, Australia
[3]Western Australian Zebrafish Experimental Research Centre (WAZERC), University of Western Australia, Perth, Western Australia, Australia
[4]School of Biological Sciences, The University of Western Australia, Perth, Western Australia, Australia

ABSTRACT

Changes in DNA methylation at specific CpG sites have been used to build predictive models to estimate animal age, predominantly in mammals. Little testing for this effect has been conducted in other vertebrate groups, such as bony fish, the largest vertebrate class. The development of most age-predictive models has relied on a genome-wide sequencing method to obtain a DNA methylation level, which makes it costly to deploy as an assay to estimate age in many samples. Here, we have generated a reduced representation bisulfite sequencing data set of caudal fin tissue from a model fish species, zebrafish (*Danio rerio*), aged from 11.9-60.1 weeks. We identified changes in methylation at specific CpG sites that correlated strongly with increasing age. Using an optimised unique set of 26 CpG sites we developed a multiplex PCR assay that predicts age with an average median absolute error rate of 3.2 weeks in zebrafish between 10.9-78.1 weeks of age. We also demonstrate the use of multiplex PCR as an efficient quantitative approach to measure DNA methylation for the use of age estimation. This study highlights the potential further use of DNA methylation as an age estimation method in non-mammalian vertebrate species.

INTRODUCTION

Ageing on a cellular level involves the decline of molecular functions resulting in a cumulative decrease in cell maintenance [1]. Biological ageing at the organism level is at least partly driven by molecular and cellular changes [2]. Biological ageing is observed in most animal species [3, 4] and several epigenetic changes occur during the ageing process [5, 6]. DNA methylation at cytosine-phosphate-guanine (CpG) sites is known to change with age.

A growing number of studies have shown that changes in DNA methylation at a set of specific CpG sites are predictive of age [7]. These studies have constructed what is often referred to as epigenetic clocks and can be used to predict chronological or biological age [8–11]. Epigenetic clocks have predominantly been derived for mammalian species but has also been developed for a sea bird (*Ardenna tenuirostris*) [7, 12]. There has been little exploration of whether DNA methylation can be used for age estimation in other vertebrate groups. The identification of DNA methylation biomarkers for age estimation has typically relied on data-intensive reduced representation bisulfite sequencing (RRBS) or Illumina Infinium microarrays [13–16]. While effective for marker identification and model development, these methods are not well-suited to high-throughput characterisation of age for large numbers of non-human samples. RRBS generates a high volume of data that makes it computationally expensive. Microarrays have not been developed for most species outside of humans

and model organisms such as mice because of the large expense involved in development.

Age is fundamental information in fisheries as it can be used to estimate abundance [17], sustainable harvests, and population growth rates [18]. In short lived species the length of the fish is often used as a substitute for age [19, 20] otherwise otoliths (ear bones) are used for most species [21–23]. It is possible to obtain daily increments by otoliths for short lived fish [24]. However, this becomes difficult for fish greater than 10 years old, where only annual increments can be measured [25]. DNA methylation-based age-estimation may offer a robust alternative in fishes, and further, provide sub-annual age increments. Age estimation from otoliths is also lethal as it involves the removal of the inner ear bone. Non-lethal age estimation methods are ethically preferable and compatible with multiple live capture survey methods such as genetic capture-recapture studies [26]. In this study, we use zebrafish as a model species to develop a cost-effective approach for age estimation in fish. Zebrafish are an ideal species for this work as individuals with known ages are readily available.

Zebrafish are a short-lived species that can reproduce at 10 weeks and with an average lifespan in captivity of 150 weeks [27]. This short life cycle, their regenerative ability, and their senescent phenotype with increasing age makes them a valuable model for vertebrate ageing research [28]. Zebrafish are highly fertile and are generally inexpensive to maintain making them an ideal model species for experimental research on any aspect of fish biology [29]. Moreover, zebrafish have a long history as a genetic model organism, with a well-characterised genome that can be genetically manipulated [30–32]. The zebrafish epigenome experiences variation due to environmental factors similar to other species [33]. DNA methylation also has a pivotal role in zebrafish embryonic development orchestrating the transcriptome and regulating cell development [33, 34]. However, it is unknown if DNA methylation is predictive of age during post-embryonic development, as seen in mammals [7, 35]. Here, using RRBS and a known age series we identify DNA methylation biomarkers that accurately predict the age of zebrafish from caudal fin tissue. We develop a multiplex PCR assay for the affordable and efficient measurement of DNA methylation to estimate age with high accuracy and precision.

RESULTS

Age marker identification by reduced representation bisulfite sequencing

Full details on the maintenance of the zebrafish colony can be found in Supplementary File 1. On average, 45.1

million reads per RRBS library (Supplementary File 1 Data) was aligned to the zebrafish genome with an alignment rate of 87.4%. This resulted in a total of 524,038 CpG sites with adequate coverage in at least 90% of all samples (see Methods). Of these sites, 60.9% were within gene bodies such as exons (Supplementary Figure 1). Global CpG methylation level was on average 79.5%, similar to what has been observed in other zebrafish tissues [36–38]. We found no correlation between global CpG methylation and age (Pearson correlation = 0.030, p-value = 0.77). However, we identified methylation at 1,311 CpG sites to significantly correlate (p-value < 0.05) with increasing age, similar to what has been found in mice [15]. This suggests specific CpG sites are associated with ageing but not epigenetic drift as indicated by global methylation [39].

An elastic net regression model was used to regress age over the 70% of the RRBS samples (67 samples). The regression model returns the minimum number of sites required to estimate age (see Methods). Our model to estimate age in zebrafish using RRBS data was based on a total of 29 CpG sites (Supplementary Table 2). In the training data set a high correlation (Pearson correlation = 0.95, p-value < 2.20×10^{-16}) between the chronological and predicted age was observed (Figure 1A). In addition, a high correlation (Pearson correlation = 0.92, p-value = 9.56×10^{-11}) between these variables in the testing data set was also observed (Figure 1B). A median absolute error (MAE) rate of 3.7 weeks was found in the testing data set (Figure 1C) and no statistical difference was observed between the absolute error rate between the training and testing data sets (p-value = 0.14, t-test, two-tailed). The similar performance between the training and testing data sets suggests a low possibility of overfitting. A PCA was used to visualise the separation of samples by age using the methylation levels of the 29 CpG sites (Figure 1D). The first principal component explains 23.4% of the variation by age. This unsupervised clustering shows separation of the samples solely on increasing age, suggesting the 29 CpG sites are ideal candidates to estimate age. Samples did not cluster by sex suggesting the age associated sites are not sex specific (Supplementary Figure 2). No significant gene ontology (GO) enrichment was observed for the 29 CpG sites using Enrichr. Using a zebrafish background set of genes, four genes (*meis2a, gnptab, hoxb3a, mab21l2*) relating to embryonic skeletal system morphogenesis were identified (adjusted p-value = 0.06). It should be noted that this is not a significant GO result. However, it is a similar pattern to what has been observed in age associated CpG sites in humans as many are related to embryonic development [14]. From here on the 29 CpG sites will be referred to as the zebrafish clock sites.

Multiplex PCR followed by sequencing

Multiplex PCR has been used previously to measure DNA methylation at select CpG sites [40]. In comparison to RRBS, it provides a 100-fold decrease in sequencing cost, making a more cost-effective approach for targeted sequencing. For each zebrafish clock CpG site, primers were designed to amplify an approximate 140bp amplicon inclusive of the CpG site (Supplementary Table 3). Three CpG sites were removed from the model due to lack of amplification by primers (see Methods). Caudal fin tissue from 96 zebrafish that were not part of the initial RRBS and aged between 10.9 - 78.1 weeks was used to test the multiplex PCR assay. Samples were assayed in triplicate to determine reproducibility of the method. On average, 459,000 reads were aligned to the reference genome, with 15,248 reads per amplicon, with an alignment rate of 98.5%. By having a saturation of high read coverage per amplicon reduces any potential of read variation on methylation levels. We found a high average correlation across the replicates (Supplementary Figure 3) between the chronological and predicted age (Pearson correlation = 0.97) and a low average (mean) MAE of 3.18 weeks (Figure 2). No statistically significant difference was found between the absolute
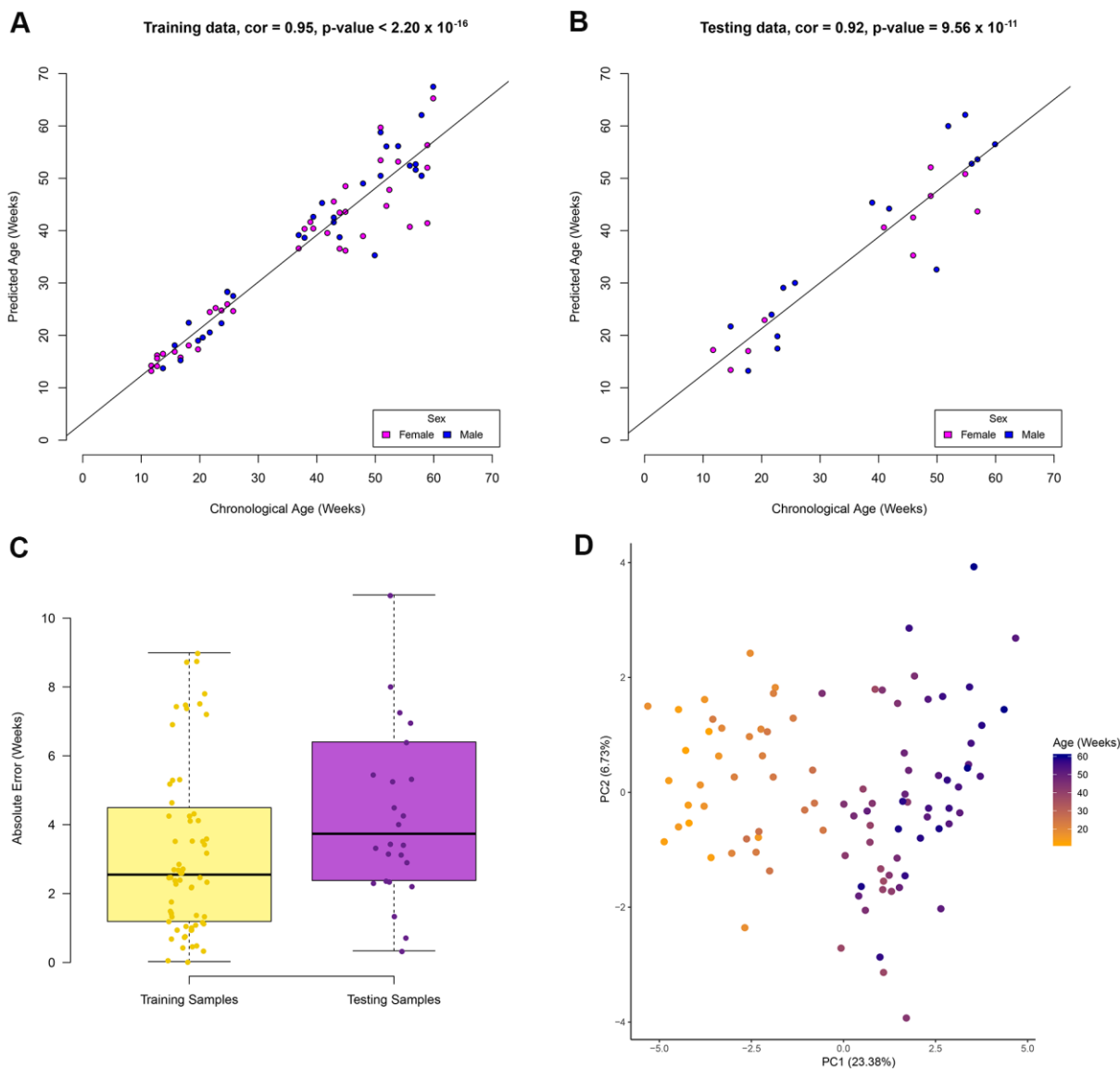
Figure 1. Zebrafish age estimation from DNA methylation of 29 CpG sites. Performance of the model in the (**A**) training data set, (**B**) testing data set. Colour represents the sample sex in the correlation plots. (**C**) Boxplots show the absolute error rate in the training and testing data sets. (**D**) Unsupervised clustering of samples using the 29 CpG sites show separation based on age in the first principle component.

error rates between replicates (p-value = 0.366, one-way ANOVA), suggesting the method is highly reproducible. In addition, no statistically significant difference was found between the absolute error rate in the RRBS testing data set and the multiplex PCR samples (p-value = 0.23, t-test, two-tailed). This suggests RRBS and multiplex PCR return similar sensitivities in methylation values. A median relative error of 8.2% was also observed in the multiplex PCR assay. No significant difference was found between the residuals and increasing age (Supplementary Figure 4) suggesting a consistent relative error rate with age.

Methylation-sensitive PCR

Methylation sensitive PCR (msPCR) can be deployed as a rapid and cost-effective method to assay methylation of selected CpG sites, but has not been used to quantify methylation for the estimation of age [41]. We used msPCR as a potential alternative non-sequencing methodology to quantify DNA methylation for age prediction. Primers (Supplementary Table 4) were designed to target the zebrafish clock sites (see Methods). Despite a significant correlation between the

chronological and predicted age (Pearson correlation = 0.62, p-value = 0.00028) the MAE rate increased 261% (13.4 weeks) compared to the RRBS MAE rate (Supplementary Figure 5). This analysis suggests msPCR is not sensitive enough as the relative error was 36.2%.

Epigenetic drift

The elastic net regression model returns the minimum number of CpG sites required to estimate age. However, these sites differ in terms of importance for age prediction. Each CpG site has a different weight (Figure 3A), but collectively can be used to estimate age. This demonstrates that despite each CpG site having a different degree of age-association, collectively multiple CpG sites can be used to accurately estimate age [14]. To determine the level of age-association in other CpG sites we used a ridge model (α-parameter = 0 in glmnet) and randomly selected 29 CpG sites out of the possible 524,038 CpG sites. This was repeated 10,000 times and produced an average MAE of 15.1 weeks (Figure 3B). This analysis demonstrates that collectively CpG sites can be predictive of age, however certain CpG sites are better candidates as biomarkers of age.
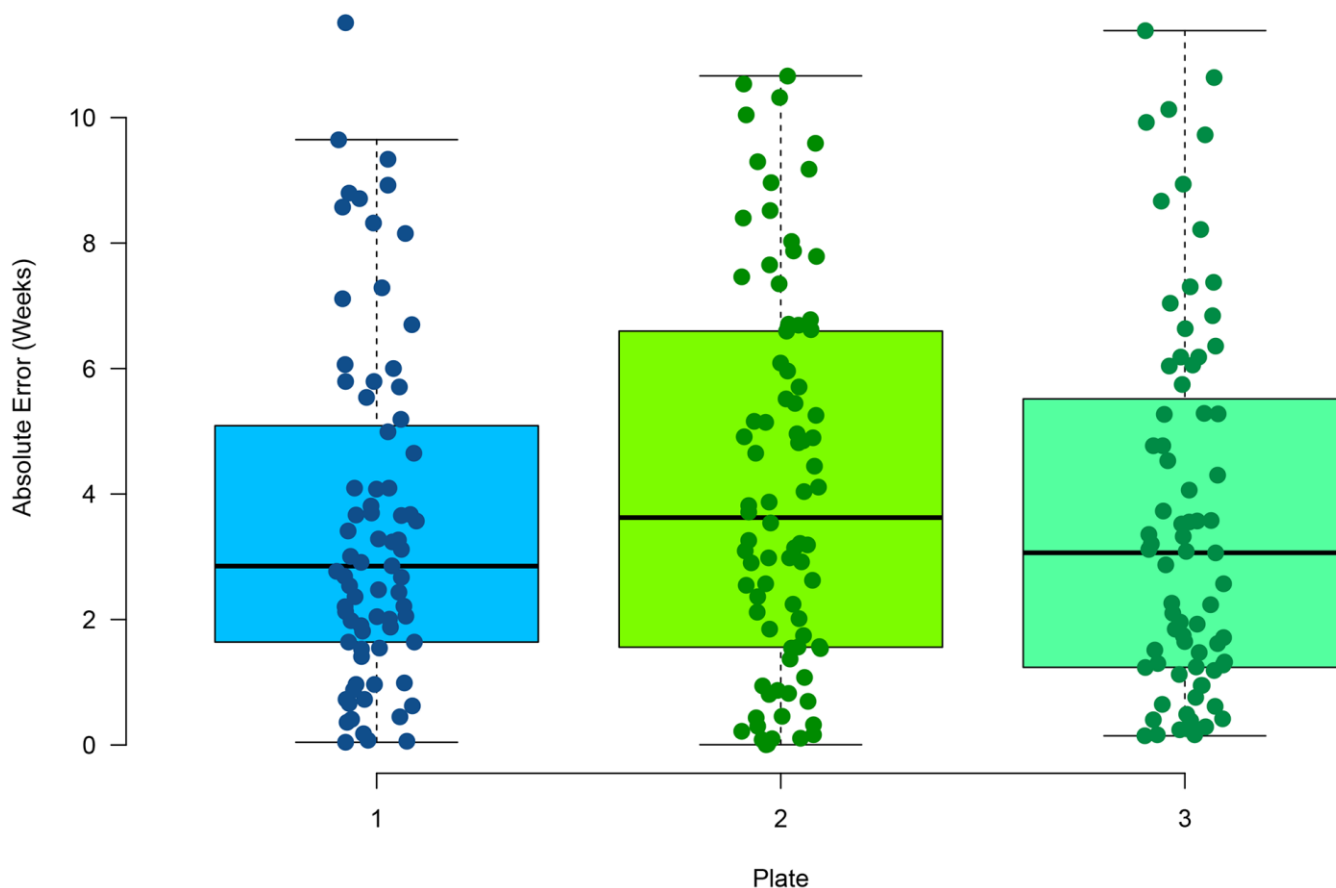

Figure 2. Performance of age estimation by multiplex PCR showing the absolute error rate for the 96 samples in triplicate.

DISCUSSION

We have developed the first epigenetic clock for zebrafish. By developing a high-resolution DNA methylation map for zebrafish caudal fin tissue, we were able to identify age associated CpG sites that can be collectively used to estimate age. Previous studies have developed assays to target small numbers of CpG sites to estimate age [42, 43] or used data-intensive genome-wide approaches [15, 16]. This is the first study to both develop a genome-wide characterisation of age-associated CpG sites in zebrafish and to construct a high-throughput assay deployable in a basic molecular biology laboratory for specific CpG sites that estimate age.

The changes in DNA methylation at specific CpG sites with increasing age are generally small and it is rare for methylation levels to go from completely unmethylated to fully methylated or vice versa. Therefore, when measuring methylation at specific CpG sites it is essential for the assay to be highly sensitive. The efficacy of msPCR was investigated because of its modest technical requirements in comparison to DNA sequencing approaches. In other applications msPCR has delivered sufficient sensitivity to cancer prognosis assays [44]. However, msPCR was insufficiently sensitive for the detection of the small changes in methylation during ageing in zebrafish. In contrast, multiplex PCR appears to be an ideal cost-effective approach to estimate age from CpG methylation. As other age-estimation studies have shown, fewer than

110 CpG sites is required to estimate age [16, 42, 43]. Therefore, if estimating age is the only requirement, it is only necessary to perform sequencing on CpG sites that are predictive of age rather than the more data intensive and costly RRBS. Multiplex PCR followed by sequencing is an ideal tool to estimating age as it can focus on genomic regions of interest and still provide the same sensitivity as RRBS. This method for age estimation has an advantage over array-based methods for many researchers in that arrays are only available for a limited range of species and the technology for implementing array-based epigenetic clocks is not available in many laboratories.

The aim of the study was to demonstrate the use of a cost-effective method of age estimation in a model fish. However, just as similar clocks in mice and humans have been used, there is the potential for it also to be used as a proxy for zebrafish health [14, 15]. Known age profiles are key data used to inform stock assessments in harvested fishes, and existing methods to age fish is lethal and subject to error. There is the potential of applying the age associated CpG sites in zebrafish in other bony fish where there is the conservation of DNA sequence. This approach has been used with success in mammals [42, 43]. Unlike human and mice studies, which were multi-tissue, the zebrafish epigenetic clock is tissue specific, and therefore may not accurately predict age from methylation data collected from other tissues. Caudal fin was selected because it is a widely sampled in fisheries management context and can be
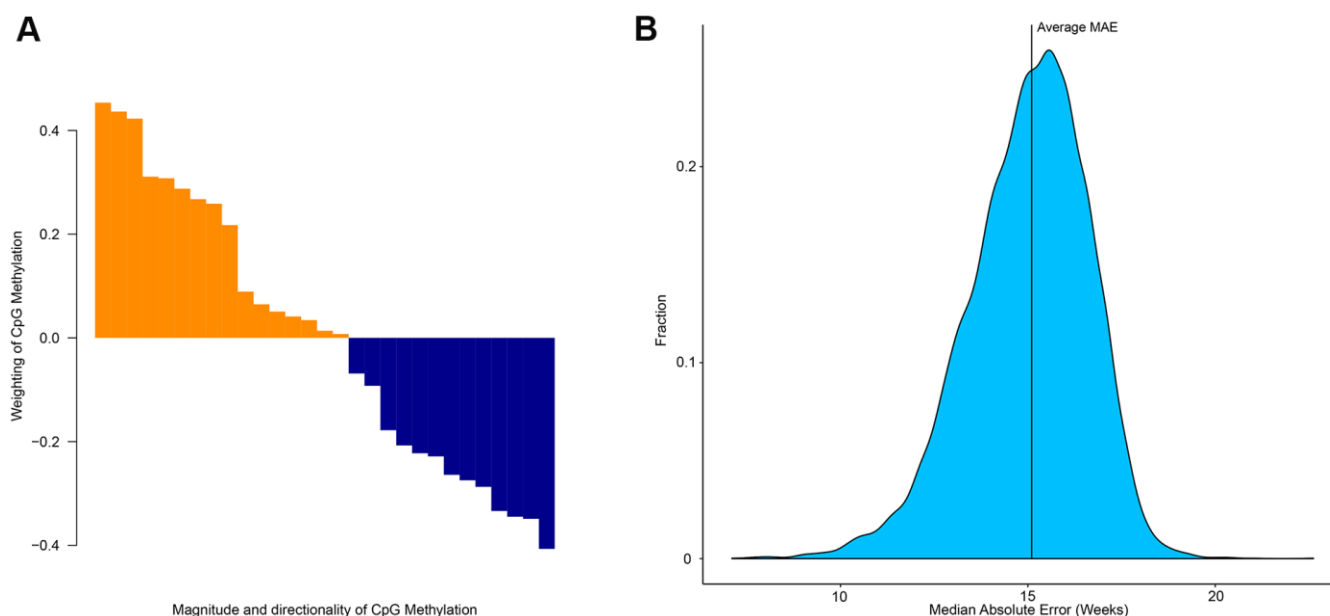
Figure 3. Importance of specific CpG sites in estimating the age of zebrafish. (**A**) Weighting and directionality of each of the 29 CpG age associated sites. (**B**) Distribution of the performance of 10,000 age-estimation models in the form of median absolute error (weeks).

collected non-lethally [45]. Establishing a clock for fish caudal fin is therefore a necessary proof of concept for application to other species.

The first age-predictive model for a fish was developed by [46], who used multiplex PCR to estimate the age of European Seabass (*Dicentrarchus labrax*). That study also produced a highly accurate model (Pearson correlation = 0.824, MAE= 2.149 years), based on targeting candidate genes involved in tissue specific development [46]. Our study took an approach that has previously been used to identify age biomarkers in mice and dogs [15, 16] and that is less dependent on prior knowledge. The zebrafish model produced a more accurate model than the European Seabass (Pearson correlation = 0.97). However, it's unclear whether this reflects differences between the species, or the method used to identify predictive CpG sites. Regardless of the performance of the models, the addition of the zebrafish epigenetic clock demonstrates the possibility of DNA methylation being predictive of age in a wide variety of fish. Under ideal conditions zebrafish live up to an average of 182 weeks in captivity [27]. This study used zebrafish up to 78.1 weeks as older individuals were unavailable. It is therefore unknown how well the model will perform on older individuals. The model was also developed with one zebrafish strain (AB). In any type of machine learning, the model will generally perform best on data similar characteristics to what it was trained on. To establish the generality of our results, in future the zebrafish model we have developed should ideally be tested in a subset of known age older fish and from other strains, including wild caught fish. However, if applying this model to wild zebrafish or applying similar tests for other wild fish species calibrated with a preponderance of younger individual, individuals near the maximum lifespan are generally very rare in wild populations. For many population biology research questions, older individuals are grouped into a "plus class" representing a wide age range, but very small counts [47].

The stochastic accumulation of error in the epigenome is often described as epigenetic drift and is a fundamental part of the ageing process [48, 49]. Using biological clocks to measure epigenetic drift has been previously suggested [35, 50, 51]. We found methylation at all CpG sites captured by RRBS have some level of age association, which is similar to results observed in mice [15]. This supports the notion of epigenetic drift as a factor in why DNA methylation performs as a marker of age. Yet, while epigenetic drift occurs as a random process across the genome, some CpG sites are significantly better predictors of age, suggesting additional functional drivers of age-related methylation. In humans, 30% of age associated DNA

methylation is tissue specific [52]. There is also evidence that age associated DNA methylation occurs in bivalent domains and Polycomb group promoters [52, 53]. This suggests that age associated DNA methylation occurs in conjunction with specific epigenetic and genomic features over and above signals of epigenetic drift [54]. This is yet to be explored in non-human organisms such as zebrafish. Developing a better functional understanding of the mechanisms underpinning epigenetic clocks would enable more targeted identification of age biomarkers, especially in non-model organisms.

One of the limitations of using RRBS as it does not assess all CpG sites and more predictive sites may be missed. Furthermore, like all sequence-based approaches it can result in inconsistent coverage across sites, which may introduce error in the identification of age markers. Whole genome bisulfite sequencing would enable evaluation of all CpG sites, potentially revealing the strongest possible set of age predictive CpG sites. However, the high cost per sample and the large number of samples required to develop an age-estimation model would makes this a costly endeavour. Indeed, the high accuracy and precision demonstrated by our model developed from sites discovered through RRBS indicates that this approach suffices for developing accurate epigenetic clock for zebrafish. Similarly, the deep sequence coverage enabled by our PCR multiplex assay of CpG markers initially identified through RRBS shows that despite potential instances of low coverage in the marker identification phase, the biomarkers identified produce highly accurate estimates of zebrafish age. Microarrays represent an alternative means to assay CpG sites, with high accuracy and without suffering from low sequence coverage [55].

CONCLUSIONS

This study is the first to develop a RRBS dataset for zebrafish caudal fin tissue from a broad range of ages and highlights the potential to use DNA methylation as a predictor of age in non-mammalian and non-avian animal groups. This is a valuable resource as it provides a time series of methylation in a species that is a model for development studies. Using this methylation data set we were able to identify CpG sites that collectively can be used to estimate age very accurately. Moreover, we were able to design a multiplex PCR assay to measure the methylation state at 26 CpG sites, at a significantly reduced cost and complexity of analysis compared to RRBS. Age has a central role in regulating the dynamics of animal populations and estimates of age-structure underpin almost all frameworks for wildlife and fisheries management. Yet, biomarkers for age are lacking for most animal groups. The transfer of

epigenetic markers between mammal groups including humans, mice and bats indicates that similar approaches may be feasible in other groups such as fishes. The transfer of zebrafish epigenetic age markers to fishes with significant commercial or conservation importance would be of major significance considering the importance of age structure to management and the lack of effective non-lethal alternatives to estimating age.

MATERIALS AND METHODS

Zebrafish ageing colony

Zebrafish (AB strain) were bred and maintained at the Western Australian Zebrafish Experimental Research Centre (WAZERC). Refer to the Supplementary File 1 for full details on how the zebrafish were maintained. Animal ethics was approved by the University of Western Australia animal ethics committee (RA/3/100/ 1630). Animals aged between 10.9-78.1 weeks were euthanized using rapid cooling. Once deceased all organs and tissues were collected and stored into RNAlater (Thermo Fisher). DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN) as instructed in the manufacturer's protocol.

Reduced representation bisulfite sequencing

A total of 96 RRBS libraries were prepared as previously described with digestion of the restriction enzyme *MspI* [56] at the Australian Genome Research Facility (AGRF) and were sequenced using an Illumina NovaSeq. Details of each zebrafish which were sequenced by RRBS are provided in Supplementary Table 1.

Data availability

Raw sequencing data from RRBS has been made publicly available on the CSIRO Data Access Portal available at https://doi.org/10.25919/5f63ce026960a.

RRBS data analysis

Fastq files were quality checked using FastQC v0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were trimmed using trimmomatic v 0.38 [57] with the following options: SE -phred33 ILLUMINACLIP:TruSeq3-SE:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. Trimmed reads were aligned to the zebrafish genome (danRer10) using BS-Seeker2 v 2.0.3 default settings [58] and bowtie2 v2.3.4 [59]. Methylation calling was performed using BS-Seeker2 call methylation module with default settings. CpG sites were filtered out of the analysis if they had a mean coverage of < 2 reads or >

100 reads as what has occurred previously [15]. On average, each site per sample had a coverage of 16 reads.

Predicting age from CpG methylation

Samples were randomly assigned to either a training (67 samples) or a testing data set (29 samples) using the createDataPartition function in the caret R package to maintain equal ratios of sex in each data set [60]. Age was transformed to natural log to fit a linear model. Using an elastic net regression model, the age of the zebrafish was regressed over all CpG site methylation in the training data set. Sites with missing data in less than 10% of samples were replaced with a methylation score of 0. By replacing the methylation score with 0 in samples with missing sites prevents any correction bias as the site will be removed from the analysis. The glmnet function in the glmnet R package [61] was set to a 10-fold cross validation with an α-parameter of 0.5 (optimal between a ridge and lasso model), which returned a minimum λ-value based on the training data of 0.02599415. These parameters resulted in a total of 29 CpG sites required to estimate the age of zebrafish. These 29 sites had methylation values in 100% of all samples. The performance of the model in the training and testing data set were assessed using Pearson correlations between the chronological and predicted age and the MAE rates.

Principal component analysis and gene ontology

A principal component analysis (PCA) was used as a form of unsupervised clustering to visualise the age associated CpG sites in terms of separating samples by age. PCA was performed using FactoMineR [62]. Gene ontology (GO) enrichment was performed using the 2018 terms in the R package Enrichr and using the Generic Gene Ontology Term finder (https://go.princeton.edu/) [63, 64]. All analyses were performed in R using version 3.5.1 [65].

DNA bisulfite conversion

DNA was bisulfite converted using the EZ DNA Methylation Gold Kit (Zymo Research) using manufacturer's protocol or using a manual protocol as previously described [66].

Multiplex PCR

A total of 96 independent zebrafish caudal fin tissue samples which were not part of the initial RRBS were used for multiplex PCR. Primers were designed using PrimerSuite [67] and were divided into two PCR reaction pools prior to barcoding (Supplementary Table 3). Three primer pairs were unable to be optimised as

part of the overall multiplex PCR assay and were removed from the analysis. The remaining 26 CpG sites were remodelled using the RRBS methylation data by applying the ridge model component in the glmnet function (α-parameter = 0) resulting in alternative weights for each site (Supplementary Table 3). A generalised linear model was applied to the raw prediction values from the elastic net regression model (sum of the coefficient weights multiplied by the DNA methylation beta values). The final model to estimate age in zebrafish is:

$$\ln(age) = 1.008x$$

Where x is the sum of the raw summed methylation beta values for each sample.

Samples were run in triplicate to determine reproducibility of the method. The final 50μL PCR reaction contained 1x Green GoTaq Flexi Buffer (Promega), 0.025 U/μL of GoTaq Hot Start Polymerase (Promega), 4.5mM MgCl$_2$ (Promega), 0.5x Combinatorial Enhancer Solution (CES) (Refer to [68]), 200μM of each dNTP (Fisher Biotec), 15mM Tetramethylammonium chloride (TMAC) (Sigma-Aldrich), primers (both forward and reverse) were used at 200nM and finally the bisulfite treated DNA (2ng/μL). Cycling conditions were 94° C/5mins; 12 cycles of 97° C/15 seconds and 45° C/30 seconds, 72° C/120 seconds; 1 cycle of 72° C/120 seconds and 6° C/hold. An Eppendorf ProS 384 thermocycler was used for amplification. Primers were designed using PrimerSuite [67] and primer sequences are provided in Supplementary Table 3.

Barcoding

Oligonucleotides with attached MiSeq adaptors and barcodes were used for the barcoding reaction (Fluidigm PN FLD-100-3771). Barcoding was performed using 1x Green GoTaq Flexi Buffer, 0.05 U/μL of GoTaq Hot Start Polymerase, 0.25x CES, 4.5mM MgCl$_2$, 200μM of each dNTP, 25μL of the pooled template after Sera-Mag Magnetic SpeedBeads (GE Healthcare Life Sciences) clean up. Cycling conditions for barcoding were as follows 94° C/5mins; 9 cycles of 97° C/15 seconds, 60° C/30 seconds and 72° C/2mins; 72° C/2mins; 6° C/5mins. Barcoding was performed using an Eppendorf ProS 96 or 384 thermocycler. Sequencing was performed on an Illumina MiSeq using the MiSeq Reagent Kit v2 (300 cycle; PN MS-102-2002).

Multiplex PCR followed by sequencing data analysis

Sequencing data was hard clipped by 15bp at both 5′ and 3′ ends to remove adaptor sequences by SeqKit v 1.2 [69]. Reads were aligned to a reduced representation of

the genome focusing on a 500bp upstream and downstream of the zebrafish clock sites. Reads were aligned using Bismark v 0.20.0 with the following options: --bowtie2 -N 1 -L 15 --bam -p 2 --score L,-0.6,-0.6 --non_directional and methylation calling was performed using bismark_methylation_extractor [70].

Methylation sensitive PCR

msPCR primers were designed using MethPrimer v2.0 [71] which produces two pairs of primers for when the DNA is methylated and unmethylated. msPCR was optimised using the protocol detailed previously [44] with the final cycling conditions: Initialisation step 95° C/15 mins, denaturation step 95° C/30 seconds, annealing 55° C/40 seconds and extension 72° C/40 seconds, for 40 cycles. msPCR was performed using an AllTaq Mastermix (Qiagen) with 1 x SYBR Green (Thermo Fisher) in a Bio-Rad CFX96. The ΔCt values for each primer pair was used as a quantitative method for methylation. A leave-one-out cross validation approach was used to determine the level of precision for using msPCR to estimate age [60, 72].

AUTHOR CONTRIBUTIONS

B.M. designed, carried out the study, analysed and interpreted the data and wrote the manuscript. D.K. designed, analysed, carried out the multiplex PCR and provided discussion and intellectual input into the manuscript. L.K and B.E. were involved in the zebrafish maintenance, study design and provided discussion and intellectual input into the manuscript. O.B. and S.J. were involved in the study design, provided discussion and intellectual input into the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

REFERENCES

1. Sen P, Shah PP, Nativio R, Berger SL. Epigenetic mechanisms of longevity and aging. Cell. 2016; 166:822–39.
https://doi.org/10.1016/j.cell.2016.07.050
PMID:27518561

2. Jylhävä J, Pedersen NL, Hägg S. Biological age predictors. EBioMedicine. 2017; 21:29–36.
https://doi.org/10.1016/j.ebiom.2017.03.046
PMID:28396265

3. Petralia RS, Mattson MP, Yao PJ. Aging and longevity in the simplest animals and the quest for immortality. Ageing Res Rev. 2014; 16:66–82.
https://doi.org/10.1016/j.arr.2014.05.003
PMID:24910306

4. Cohen AA. Aging across the tree of life: the importance of a comparative perspective for the use of animal models in aging. Biochim Biophys Acta Mol Basis Dis. 2018; 1864:2680–89.
https://doi.org/10.1016/j.bbadis.2017.05.028
PMID:28690188

5. Booth LN, Brunet A. The aging epigenome. Mol Cell. 2016; 62:728–44.
https://doi.org/10.1016/j.molcel.2016.05.013
PMID:27259204

6. Boyd-Kirkup JD, Green CD, Wu G, Wang D, Han JD. Epigenomics and the regulation of aging. Epigenomics. 2013; 5:205–27.
https://doi.org/10.2217/epi.13.5 PMID:23566097

7. De Paoli-Iseppi R, Deagle BE, McMahon CR, Hindell MA, Dickinson JL, Jarman SN. Measuring animal age with DNA methylation: from humans to wild animals. Front Genet. 2017; 8:106.
https://doi.org/10.3389/fgene.2017.00106
PMID:28878806

8. Horvath S, Garagnani P, Bacalini MG, Pirazzini C, Salvioli S, Gentilini D, Di Blasio AM, Giuliani C, Tung S, Vinters HV, Franceschi C. Accelerated epigenetic aging in down syndrome. Aging Cell. 2015; 14:491–95.
https://doi.org/10.1111/acel.12325
PMID:25678027

9. Levine ME, Lu AT, Bennett DA, Horvath S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. Aging (Albany NY). 2015; 7:1198–211.
https://doi.org/10.18632/aging.100864
PMID:26684672

10. Vidal-Bralo L, Lopez-Golan Y, Mera-Varela A, Rego-Perez I, Horvath S, Zhang Y, Del Real Á, Zhai G, Blanco FJ, Riancho JA, Gomez-Reino JJ, Gonzalez A. Specific premature epigenetic aging of cartilage in osteoarthritis. Aging (Albany NY). 2016; 8:2222–31.
https://doi.org/10.18632/aging.101053
PMID:27689435

11. Horvath S, Levine AJ. HIV-1 infection accelerates age according to the epigenetic clock. J Infect Dis. 2015; 212:1563–73.
https://doi.org/10.1093/infdis/jiv277 PMID:25969563

12. De Paoli-Iseppi R, Deagle BE, Polanowski AM, McMahon CR, Dickinson JL, Hindell MA, Jarman SN. Age estimation in a long-lived seabird (Ardenna tenuirostris) using DNA methylation-based biomarkers. Mol Ecol Resour. 2019; 19:411–25.
https://doi.org/10.1111/1755-0998.12981
PMID:30576072

13. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013; 49:359–67.
https://doi.org/10.1016/j.molcel.2012.10.016
PMID:23177740

14. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013; 14:R115.
https://doi.org/10.1186/gb-2013-14-10-r115
PMID:24138928

15. Stubbs TM, Bonder MJ, Stark AK, Krueger F, von Meyenn F, Stegle O, Reik W, and BI Ageing Clock Team. Multi-tissue DNA methylation age predictor in mouse. Genome Biol. 2017; 18:68.
https://doi.org/10.1186/s13059-017-1203-5
PMID:28399939

16. Thompson MJ, vonHoldt B, Horvath S, Pellegrini M. An epigenetic aging clock for dogs and wolves. Aging (Albany NY). 2017; 9:1055–68.
https://doi.org/10.18632/aging.101211
PMID:28373601

17. Bravington MV, Skaug HJ, Anderson EC. Close-kin mark-recapture. Statistical Science. 2016; 31:259–74.
https://doi.org/10.1214/16-STS552

18. Conover DO, Munch SB. Sustaining fisheries yields over evolutionary time scales. Science. 2002; 297:94–96.
https://doi.org/10.1126/science.1074085
PMID:12098697

19. Bellido J, Pierce G, Romero J, Millan M. Use of frequency analysis methods to estimate growth of anchovy (Engraulis encrasicolus L. 1758) in the Gulf of Cadiz (SW Spain). Fisheries Research. 2000; 48:107–15.
https://doi.org/10.1016/S0165-7836(00)00183-1

20. Cubillos LA, Arcos DF, Bucarey DA, Canales MT. Seasonal growth of small pelagic fish off Talcahuano,

Chile (37° S, 73° W): a consequence of their reproductive strategy to seasonal upwelling? Aquatic Living Resources. 2001; 14:115–24. https://doi.org/10.1016/S0990-7440(01)01112-3

21. Thomas R. A method of age determination for the South West African pilchard Sardinops ocellata. South African Journal of Marine Science. 1984; 2:63–70. https://doi.org/10.2989/02577618409504359

22. Melo Y. Age studies on anchovy Engraulis capensis Gilchrist off south west Africa. South African Journal of Marine Science. 1984; 2:19–31. https://doi.org/10.2989/02577618409504355

23. Morales-Nin B. (1989). Age and growth of the southern stock of Peruvian anchoveta based on otolith microstructures and length frequency analysis. ICLARM Conference Proceedings, pp. 179–88.

24. Chambers R, Miller T. (1995). Evaluating fish growth by means of otolith increment analysis: special properties of individual-level longitudinal data. In: Recent developments in otolith research. University of South Carolina Press. pp. 155–75.

25. Thorrold SR, Hare JA. (2002). Otolith applications in reef fish ecology. In: Coral Reef Fishes: Elsevier, pp. 243–64. https://doi.org/10.1016/B978-012615185-5/50015-3

26. Lamb CT, Ford AT, Proctor MF, Royle JA, Mowat G, Boutin S. Genetic tagging in the anthropocene: scaling ecology from alleles to ecosystems. Ecol Appl. 2019; 29:e01876. https://doi.org/10.1002/eap.1876 PMID:30913353

27. Tacutu R, Thornton D, Johnson E, Budovsky A, Barardo D, Craig T, Diana E, Lehmann G, Toren D, Wang J, Fraifeld VE, de Magalhães JP. Human ageing genomic resources: new and updated databases. Nucleic Acids Res. 2018; 46:D1083–90. https://doi.org/10.1093/nar/gkx1042 PMID:29121237

28. Gerhard GS. Comparative aspects of zebrafish (danio rerio) as a model for aging research. Exp Gerontol. 2003; 38:1333–41. https://doi.org/10.1016/j.exger.2003.10.022 PMID:14698814

29. Gilbert MJ, Zerulla TC, Tierney KB. Zebrafish (danio rerio) as a model for the study of aging and exercise: physical ability and trainability decrease with age. Exp Gerontol. 2014; 50:106–13. https://doi.org/10.1016/j.exger.2013.11.013 PMID:24316042

30. Rafferty SA, Quinn TA. A beginner's guide to understanding and implementing the genetic modification of zebrafish. Prog Biophys Mol Biol. 2018; 138:3–19. https://doi.org/10.1016/j.pbiomolbio.2018.07.005 PMID:30032905

31. Keller JM, Keller ET. (2018). The use of mature zebrafish (Danio rerio) as a model for human aging and disease. Conn's Handbook of Models for Human Aging: Elsevier, pp. 351–59. https://doi.org/10.1016/B978-0-12-811353-0.00026-9

32. Kishi S. Functional aging and gradual senescence in zebrafish. Ann N Y Acad Sci. 2004; 1019:521–26. https://doi.org/10.1196/annals.1297.097 PMID:15247079

33. Cavalieri V, Spinelli G. Environmental epigenetics in zebrafish. Epigenetics Chromatin. 2017; 10:46. https://doi.org/10.1186/s13072-017-0154-0 PMID:28982377

34. Mhanni AA, McGowan RA. Global changes in genomic methylation levels during early development of the zebrafish embryo. Dev Genes Evol. 2004; 214:412–7. https://doi.org/10.1007/s00427-004-0418-0 PMID:15309635

35. Field AE, Robertson NA, Wang T, Havas A, Ideker T, Adams PD. DNA methylation clocks in aging: categories, causes, and consequences. Mol Cell. 2018; 71:882–95. https://doi.org/10.1016/j.molcel.2018.08.008 PMID:30241605

36. Falisse E, Ducos B, Stockwell PA, Morison IM, Chatterjee A, Silvestre F. DNA methylation and gene expression alterations in zebrafish early-life stages exposed to the antibacterial agent triclosan. Environ Pollut. 2018; 243:1867–77. https://doi.org/10.1016/j.envpol.2018.10.004 PMID:30408875

37. Ortega-Recalde O, Day RC, Gemmell NJ, Hore TA. Zebrafish preserve global germline DNA methylation while sex-linked rDNA is amplified and demethylated during feminisation. Nat Commun. 2019; 10:3053. https://doi.org/10.1038/s41467-019-10894-7 PMID:31311924

38. Adam AC, Lie KK, Whatmore P, Jakt LM, Moren M, Skjærven KH. Profiling DNA methylation patterns of zebrafish liver associated with parental high dietary arachidonic acid. PLoS One. 2019; 14:e0220934. https://doi.org/10.1371/journal.pone.0220934 PMID:31398226

39. Jones MJ, Goodman SJ, Kobor MS. DNA methylation and healthy human aging. Aging Cell. 2015; 14:924–32. https://doi.org/10.1111/acel.12349 PMID:25913071

40. Korbie D, Lin E, Wall D, Nair SS, Stirzaker C, Clark SJ,

Trau M. Multiplex bisulfite PCR resequencing of clinical FFPE DNA. Clin Epigenetics. 2015; 7:28.
https://doi.org/10.1186/s13148-015-0067-3
PMID:25861392

41. Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci USA. 1996; 93:9821–26.
https://doi.org/10.1073/pnas.93.18.9821
PMID:8790415

42. Polanowski AM, Robbins J, Chandler D, Jarman SN. Epigenetic estimation of age in humpback whales. Mol Ecol Resour. 2014; 14:976–87.
https://doi.org/10.1111/1755-0998.12247
PMID:24606053

43. Wright PG, Mathews F, Schofield H, Morris C, Burrage J, Smith A, Dempster EL, Hamilton PB. Application of a novel molecular method to age free-living wild Bechstein's bats. Mol Ecol Resour. 2018; 18:1374–80.
https://doi.org/10.1111/1755-0998.12925
PMID:29981199

44. Huang Z, Bassil CF, Murphy SK. Methylation-specific PCR. Methods Mol Biol. 2013; 1049:75–82.
https://doi.org/10.1007/978-1-62703-547-7_7
PMID:23913210

45. Sanderson BL, Tran CD, Coe HJ, Pelekis V, Steel EA, Reichert WL. Nonlethal Sampling of Fish Caudal Fins Yields Valuable Stable Isotope Data for Threatened and Endangered Fishes. Transactions of the American Fisheries Society. 2009; 138:1166–77.
https://doi.org/10.1577/T08-086.1

46. Anastasiadi D, Piferrer F. A clockwork fish: age prediction using DNA methylation-based biomarkers in the European seabass. Mol Ecol Resour. 2020; 20:387–97.
https://doi.org/10.1111/1755-0998.13111
PMID:31674713

47. Riekkola L, Zerbini AN, Andrews O, Andrews-Goff V, Baker CS, Chandler D, Childerhouse S, Clapham P, Dodémont R, Donnelly D, Friedlaender A, Gallego R, Garrigue C, et al. Application of a multi-disciplinary approach to reveal population structure and Southern Ocean feeding grounds of humpback whales. Ecological Indicators. 2018; 89:455–65.
https://doi.org/10.1016/j.ecolind.2018.02.030

48. Martin GM. Epigenetic drift in aging identical twins. Proc Natl Acad Sci USA. 2005; 102:10413–14.
https://doi.org/10.1073/pnas.0504743102
PMID:16027353

49. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013; 153:1194–217.
https://doi.org/10.1016/j.cell.2013.05.039
PMID:23746838

50. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suñer D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci USA. 2005; 102:10604–09.
https://doi.org/10.1073/pnas.0500398102
PMID:16009939

51. Issa JP. Aging and epigenetic drift: a vicious cycle. J Clin Invest. 2014; 124:24–29.
https://doi.org/10.1172/JCI69735 PMID:24382386

52. Zhu T, Zheng SC, Paul DS, Horvath S, Teschendorff AE. Cell and tissue type independent age-associated DNA methylation changes are not rare but common. Aging (Albany NY). 2018; 10:3541–57.
https://doi.org/10.18632/aging.101666
PMID:30482885

53. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P, Spector TD. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res. 2010; 20:434–39.
https://doi.org/10.1101/gr.103101.109
PMID:20219945

54. Bertucci EM, Parrott BB. Is CpG density the link between epigenetic aging and lifespan? Trends Genet. 2020; 36:725–27.
https://doi.org/10.1016/j.tig.2020.06.003
PMID:32624337

55. Huang YW, Huang TH, Wang LS. Profiling DNA methylomes from microarray to genome-scale sequencing. Technol Cancer Res Treat. 2010; 9:139–47.
https://doi.org/10.1177/153303461000900203
PMID:20218736

56. Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet. 2011; 43:811–14.
https://doi.org/10.1038/ng.864 PMID:21706000

57. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014; 30:2114–20.
https://doi.org/10.1093/bioinformatics/btu170
PMID:24695404

58. Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, Chen PY, Pellegrini M. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics. 2013; 14:774.

https://doi.org/10.1186/1471-2164-14-774
PMID:24206606

59. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012; 9:357–59. https://doi.org/10.1038/nmeth.1923 PMID:22388286

60. Kuhn M. Building predictive models in R using the caret package. Journal of statistical software. 2008; 28:1–26. https://doi.org/10.18637/jss.v028.i05

61. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010; 33:1–22. PMID:20808728

62. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. Journal of statistical software. 2008; 25:1–18. https://doi.org/10.18637/jss.v025.i01

63. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016; 44:W90–97. https://doi.org/10.1093/nar/gkw377 PMID:27141961

64. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G. GO::TermFinder—open source software for accessing gene ontology information and finding significantly enriched gene ontology terms associated with a list of genes. Bioinformatics. 2004; 20:3710–15. https://doi.org/10.1093/bioinformatics/bth456 PMID:15297299

65. Team RC. R: A language and environment for statistical computing. 2013.

66. Clark SJ, Statham A, Stirzaker C, Molloy PL, Frommer M. DNA methylation: bisulphite modification and analysis. Nat Protoc. 2006; 1:2353–64.

https://doi.org/10.1038/nprot.2006.324
PMID:17406479

67. Lu J, Johnston A, Berichon P, Ru KL, Korbie D, Trau M. PrimerSuite: a high-throughput web-based primer design program for multiplex bisulfite PCR. Sci Rep. 2017; 7:41328. https://doi.org/10.1038/srep41328 PMID:28117430

68. Ralser M, Querfurth R, Warnatz HJ, Lehrach H, Yaspo ML, Krobitsch S. An efficient and economic enhancer mix for PCR. Biochem Biophys Res Commun. 2006; 347:747–51. https://doi.org/10.1016/j.bbrc.2006.06.151 PMID:16842759

69. Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One. 2016; 11:e0163962. https://doi.org/10.1371/journal.pone.0163962 PMID:27706213

70. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. Bioinformatics. 2011; 27:1571–72. https://doi.org/10.1093/bioinformatics/btr167 PMID:21493656

71. Li LC, Dahiya R. MethPrimer: designing primers for methylation PCRs. Bioinformatics. 2002; 18:1427–31. https://doi.org/10.1093/bioinformatics/18.11.1427 PMID:12424112

72. Picard RR, Cook RD. Cross-validation of regression models. Journal of the American Statistical Association. 1984; 79:575–83. https://doi.org/10.1080/01621459.1984.10478083
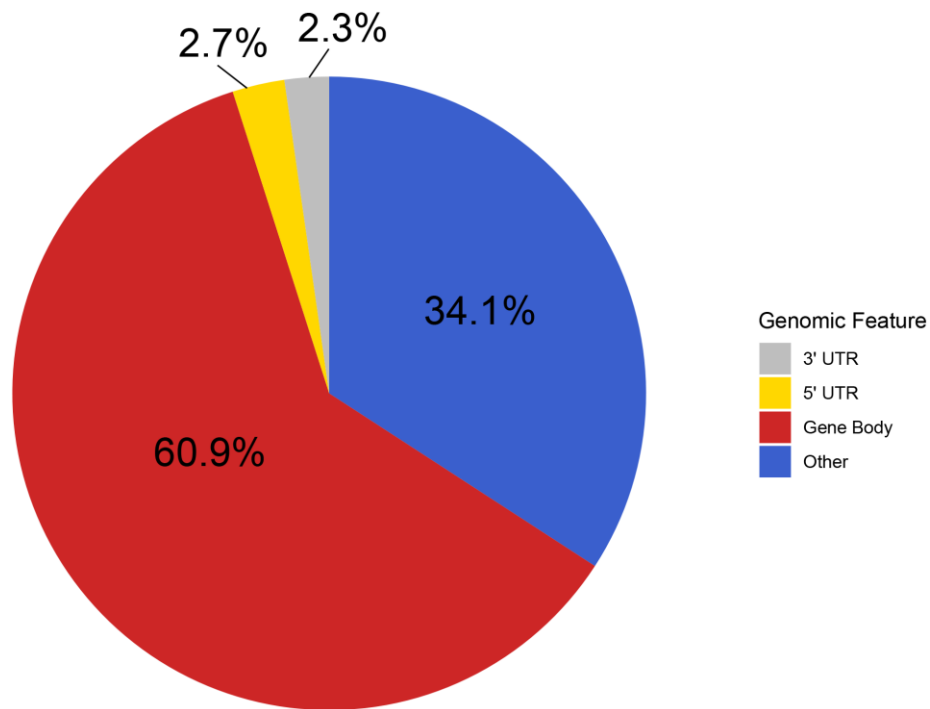
Supplementary Figure 1. Genomic distribution of CpG sites which were captured in reduced representation bisulfite sequencing.
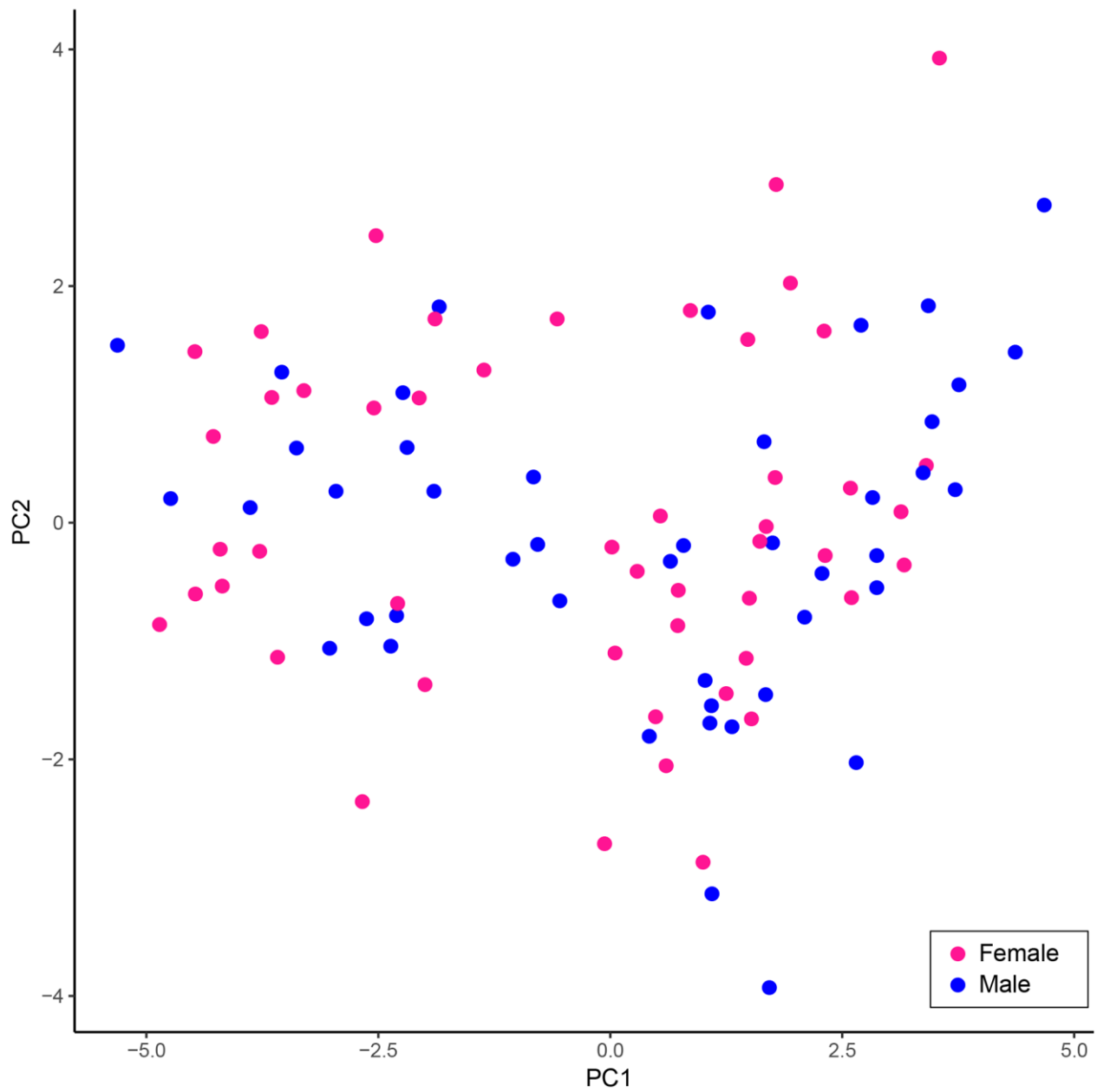
Supplementary Figure 2. Principle component analysis displaying no separation of sample sex.
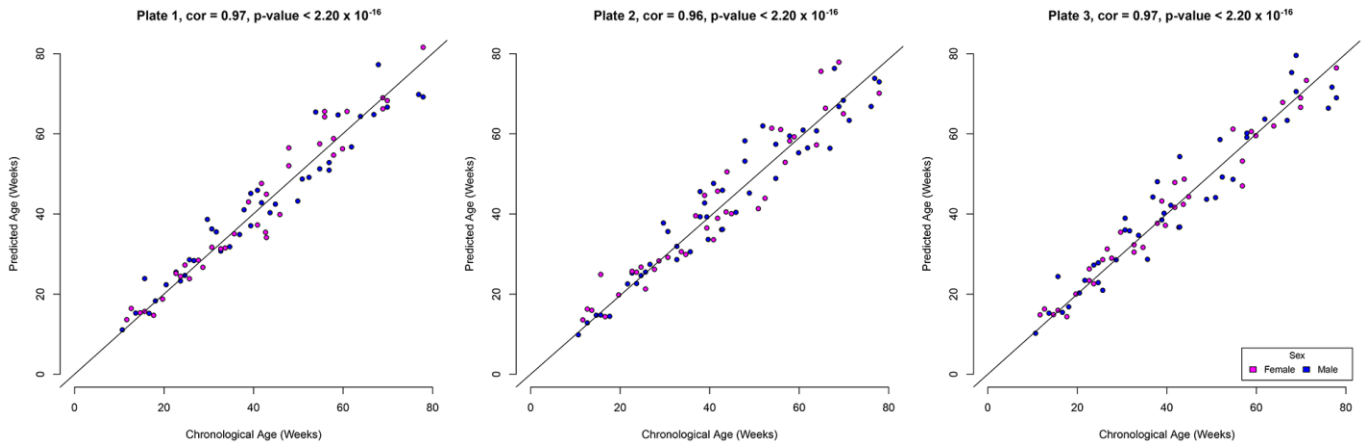
Supplementary Figure 3. Correlation between the chronological and predicted age in zebrafish by multiplex PCR. Samples were run in triplicate.
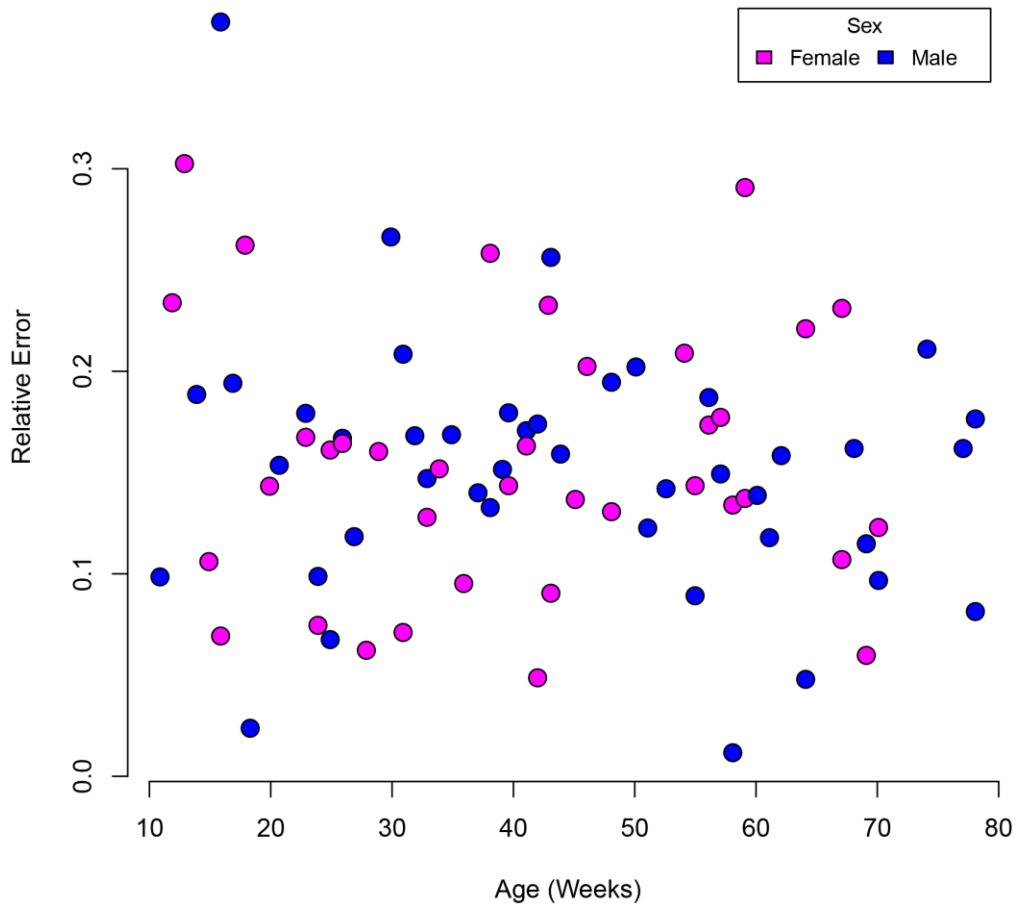
Supplementary Figure 4. Absolute error rate of samples by multiplex PCR over increasing age.
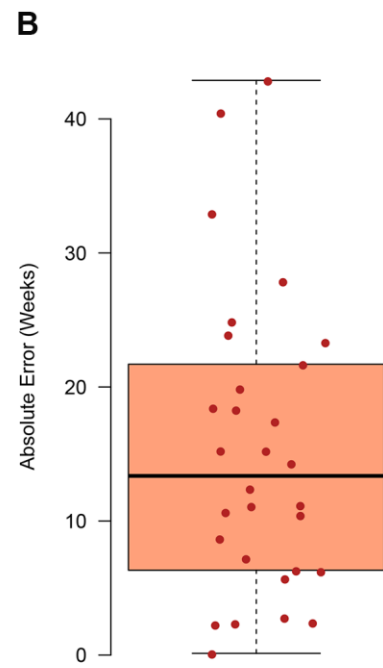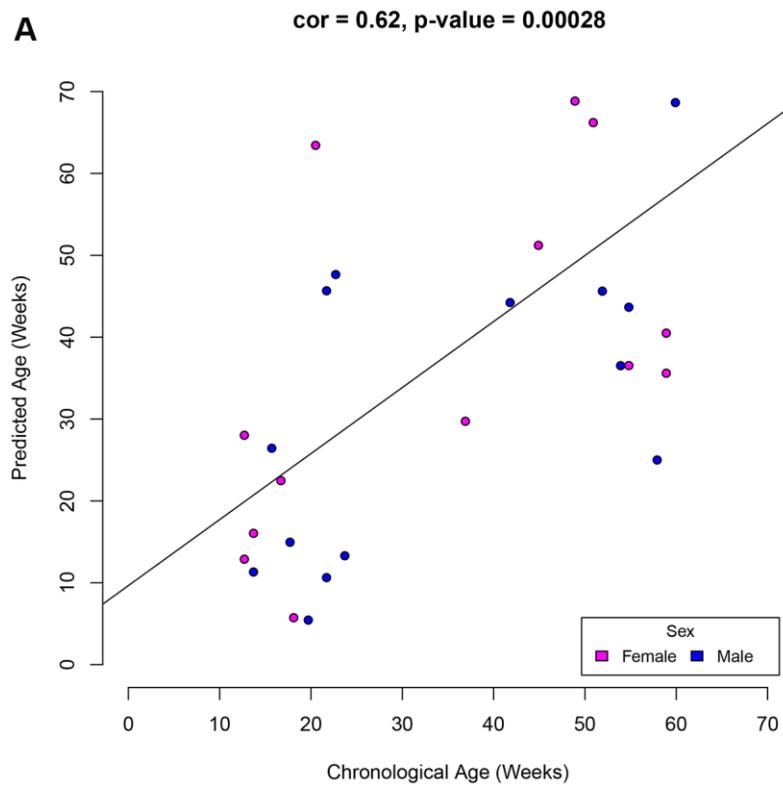
Supplementary Figure 5. Methylation-sensitive PCR to estimate age in zebrafish. (**A**) Correlation between the chronological and predicted age and (**B**) the absolute error rate in age estimation.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 3, 4.

Supplementary Table 1. Phenotype and sequencing data of the 96 caudal fin tissue from zebrafish used for RRBS.

Supplementary Table 2. Locations of the zebrafish clock sites used to estimate age using RRBS data and the closest genomic feature.

CpG site			Association with age			Closest feature				
chr	position	strand	Weight	Correlation	p-value	Gene	feature	start	end	strand
Intercept	NA	NA	3.261736	NA	NA	NA	NA	NA	NA	NA
chr12	21540399	+	-0.06868	-0.456308633	3.80E-06	mrpl27	exon	21563072	21563114	+
chr12	35432443	+	0.041155	0.425636828	1.90E-05	chmp6b	exon	35487001	35487160	+
chr13	31180246	+	0.422877	0.49406794	4.18E-07	mettl18	exon	31259863	31260037	+
chr13	38582448	+	0.287827	0.518416373	8.70E-08	zgc:153049	exon	38688631	38688754	+
chr14	38455793	-	-0.34896	-0.404759937	5.20E-05	csnk1a1	exon	38442661	38443287	-
chr14	45387151	+	-0.22242	-0.432519711	1.34E-05	sncb	exon	45619305	45619341	+
chr17	52836692	+	0.089225	0.44142766	8.45E-06	meis2a	exon	52833657	52835083	+
chr18	38107080	+	-0.40695	-0.407236996	4.63E-05	nucb2b	exon	38210387	38210462	+
chr18	50792250	+	-0.3449	-0.434582509	1.21E-05	reln	CDS	50795737	50795848	+
chr19	20077224	+	0.013643	0.428542532	1.64E-05	hibadha	CDS	20079490	20079646	+
chr1	23386154	+	0.267495	0.462650352	2.67E-06	mab21l2	CDS	23385795	23386871	+
chr1	43259461	+	-0.28726	-0.419849369	2.53E-05	cabp2a	exon	43425989	43426036	+
chr20	16578711	-	0.007467	0.459510595	3.18E-06	ches1	CDS	16578582	16579053	-
chr20	21624045	+	0.310809	0.383202718	0.000138	jag2b	exon	21573904	21575945	+
chr20	26523373	+	0.436491	0.468930078	1.87E-06	zbtb2b	exon	26504936	26508142	+
chr20	28928268	+	0.050606	0.492313214	4.66E-07	fntb	exon	28924424	28924877	+
chr21	23231786	+	-0.09242	-0.406502544	4.79E-05	alg8	exon	22864361	22864801	+
chr21	25150743	+	-0.33385	-0.541055377	1.80E-08	sycn.2	exon	25189953	25190586	+
chr24	19868851	+	-0.22858	-0.559326016	4.64E-09	LOC100334155	exon	20073262	20073368	+
chr24	4215673	+	0.06477	0.410284892	4.01E-05	wdr37	exon	3494784	3495510	+
chr25	14631230	+	0.217506	0.420374681	2.46E-05	mpped2	CDS	14637373	14637488	+
chr25	16313450	+	0.307822	0.482149108	8.63E-07	tead1a	CDS	16315617	16315681	+
chr25	36872756	+	-0.17805	-0.360931599	0.000352	chmp1a	exon	36871083	36871567	+
chr25	6461988	+	0.453596	0.408996487	4.26E-05	snx33	exon	6351734	6353787	+
chr2	8207957	+	0.258846	0.462933479	2.63E-06	chst2a	exon	8314444	8316603	+
chr3	23616782	+	-0.27465	-0.451308167	4.99E-06	hoxb3a	exon	23616752	23617534	+
chr4	17690807	+	-0.26411	-0.603855727	1.17E-10	gnptab	exon	17690788	17690922	+
chr4	18675145	+	-0.20748	-0.461522112	2.84E-06	slc26a4	CDS	18793563	18793599	+
chr5	51679905	+	0.034253	0.386666431	0.000118	slc14a2	exon	51529758	51531231	+

Supplementary Table 3. Primer sequences targeting the zebrafish clock sites for multiplex PCR and weights for age-estimation.

Supplementary Table 4. Primer sequences targeting the zebrafish clock sites by methylation sensitive PCR.

Supplementary File

Please browse Full Text version to see the data of Supplementary File 1.

Supplementary File 1. Details on how the zebrafish were maintained.