Authors: Ana Belén Márquez-Ruiz[(1)], Lucas González-Herrera[(1)], Juan de Dios Luna[(2)], and Aurora Valenzuela[(1)].

Title: DNA methylation levels and telomere length in human teeth: usefulness for age estimation.

[(1)] Department of Forensic Medicine, Faculty of Medicine, University of Granada. Avda. de la Investigación, 11, 18016 Granada, Spain.

[(2)] Department of Statistics, Faculty of Medicine, University of Granada, Avda. de la Investigación 11, 18016 Granada, Spain.

Corresponding author:
Ana Belén Márquez-Ruiz
E-mail address: abmarquez@ugr.es

Abstract In the last decade, increasing knowledge of epigenetics has led to the development of DNA methylation-based models to predict age, which have shown high predictive accuracy. However, despite the value of teeth as forensic samples, few studies have focused on this source of DNA. This study used bisulfite pyrosequencing to measure the methylation levels of specific CpG sites located in the *ELOVL2*, *ASPA*, and *PDE4C* genes, with the aim of selecting the most age-informative genes, and determining their associations with age, in 65 tooth samples from individuals 15 to 85 years old. As a second aim, methylation data and measurements of relative telomere length in the same set of samples were used to develop preliminary age prediction models to evaluate the accuracy of both biomarkers together and separately in estimating age from teeth for forensic purposes. In our sample, several CpG sites from *ELOVL2* and *PDE4C* genes, as well as telomere length, were significantly associated with chronological age. We developed age prediction quantile regression models based on DNA methylation levels, with and without telomere length as an additional variable, and adjusted for type of tooth and sex. Our results suggest that telomere length may have limited usefulness as a supplementary marker for DNA methylation-based age estimation in tooth samples, given that it contributed little improvement in the prediction errors of the models. In addition, even at older ages, DNA methylation appeared to be more informative in predicting age than telomere length when both biomarkers were evaluated separately.

Keywords Age estimation – DNA methylation – Telomere length – Human teeth

Introduction

Age estimation constitutes a challenging task for forensic scientists, despite the availability of several methods for forensic age assessment. Age can be estimated using procedures based on the examination of morphological changes that occur in bones and teeth with aging [1]. In addition, different approaches have been tried in recent decades based on the gradual alterations of biomolecules caused by the aging process, e.g. protein glycosylation, aspartic acid racemization, mitochondrial deletions, or telomere shortening [2–4]. Although these methods have been considered promising tools for forensic age estimation, especially when morphological information is limited, they still show certain limitations such as low accuracy or the lack of standardized protocols. In this regard it should be noted that whatever the method of assessment used, the accuracy of estimates is reduced in adults because the discrepancy between biological and chronological age becomes larger with increasing age. Moreover, the potential applicability of a given method also depends on the type and state of preservation of the material available for analysis [1], with mineralized tissues (bones and teeth) being the most informative samples for forensic purposes [1–5].

In the last few years, the increasing knowledge of epigenetics has led to the identification of a direct relationship between age and DNA methylation changes in specific regions of the genome – a correlation that can be useful for estimating chronological age [6]. DNA methylation can be defined as a chemical modification which, in eukaryotic cells, involves the presence of a methyl group at the 5′ position of a cytosine nucleotide that is followed by a guanine nucleotide, known as a CpG site [7]. Total DNA methylation levels decrease with aging, while certain CpG sites can become either hypermethylated, e.g. those located mainly in CpG islands (CpG dinucleotides clusters), or hypomethylated, e.g. those usually located outside CpG islands [8, 9]. In this connection, several studies of DNA methylation specifically for forensic age estimation have reported age prediction models based on different genes, tissues, detection technologies. and statistical age-predictive analyses, with predictive accuracies lower than 5 years [7, 10, 11]. Consequently, DNA methylation is currently considered as the most informative age prediction biomarker [7, 10].

Among the principal candidate genes for age prediction, the *ELOVL2* gene (ELOVL fatty acid elongase 2), located in chromosome 6, is regarded as the best predictor in numerous studies [8, 12–20]. During aging, the *ELOVL2* promoter becomes hypermethylated. Although this gene is considered the most promising locus for age prediction, the results need to be enhanced by additional markers to improve the accuracy of age estimates [14]. According to a systematic review of the literature on the most highly age-correlated CpG sites in the human genome (reviewed in [7, 10]), with particular attention to studies that reported reliable age prediction models based on a limited set of CpG sites [8, 19, 21, 22], *ASPA* (aspartoacylase) and

PDE4C (phosphodiesterase 4C) appear to be the most widely adopted genes along with *ELOVL2*. The *ASPA* gene, located on chromosome 17, has been included in several forensic age prediction models [8, 19, 21–24]. The most age-correlated CpG sites from *ASPA* become hypomethylated with increasing age. The *PDE4C* gene, located on chromosome 19, has also been explored as an age-predictive marker in several studies [8, 19, 21, 22, 25]. The most age-correlated CpG sites in the *PDE4C* gene display increased methylation levels with age. DNA methylation levels of CpG sites from these genes have been measured not only in blood but also in mineralized tissue [8].

An important issue to take into account is the intertissue variability of DNA methylation patterns [26, 27]. In this regard, the analysis of teeth and bones deserves particular attention, since in many cases of forensic identification, dental and/or skeletal remains are the only available sources of DNA. Due to their high resistance to harsh conditions, including postmortem DNA decay, teeth are an excellent source of DNA of often better quality than bones [28]. Nonetheless, very few studies to date have considered tooth samples in their analyses of DNA methylation [8, 15].

Therefore, this study aimed to measure the methylation levels of specific CpG sites located in the *ELOVL2*, *ASPA*, and *PDE4C* genes in order to assess their correlations with age in tooth samples. Candidate genes were selected among the most age-informative sets of a few genes, also considering that their methylation levels have been measured previously in teeth by pyrosequencing. As a second aim, the methylation data and measurements of relative telomere length in the same set of samples were used to develop preliminary age prediction models to evaluate the accuracy of both biomarkers together and separately in estimating age from teeth for forensic purposes.

Material and methods

Sample collection and DNA extraction

Sixty-five healthy permanent teeth were collected from donors aged between 15 and 85 years (mean age: 34.85 ± 15.48 years) at the public Oral Health Service and private dental clinics in Granada (Spain). The distribution of the samples according to age, sex, and type of tooth is presented in Fig. 1. Teeth were extracted for periodontal, orthodontic, prosthetic, or surgical reasons and were free of caries, root canal treatments, or restorations. The chronological age for each case was calculated as the time from the individual's date of birth to the date of tooth extraction. The research protocol, including a consent document for sample donors in compliance with applicable data protection regulations, was approved by the Ethics Committee

for Human Research of the University of Granada (Spain). The study was conducted in accordance with the ethical standards laid down by the Declaration of Helsinki.

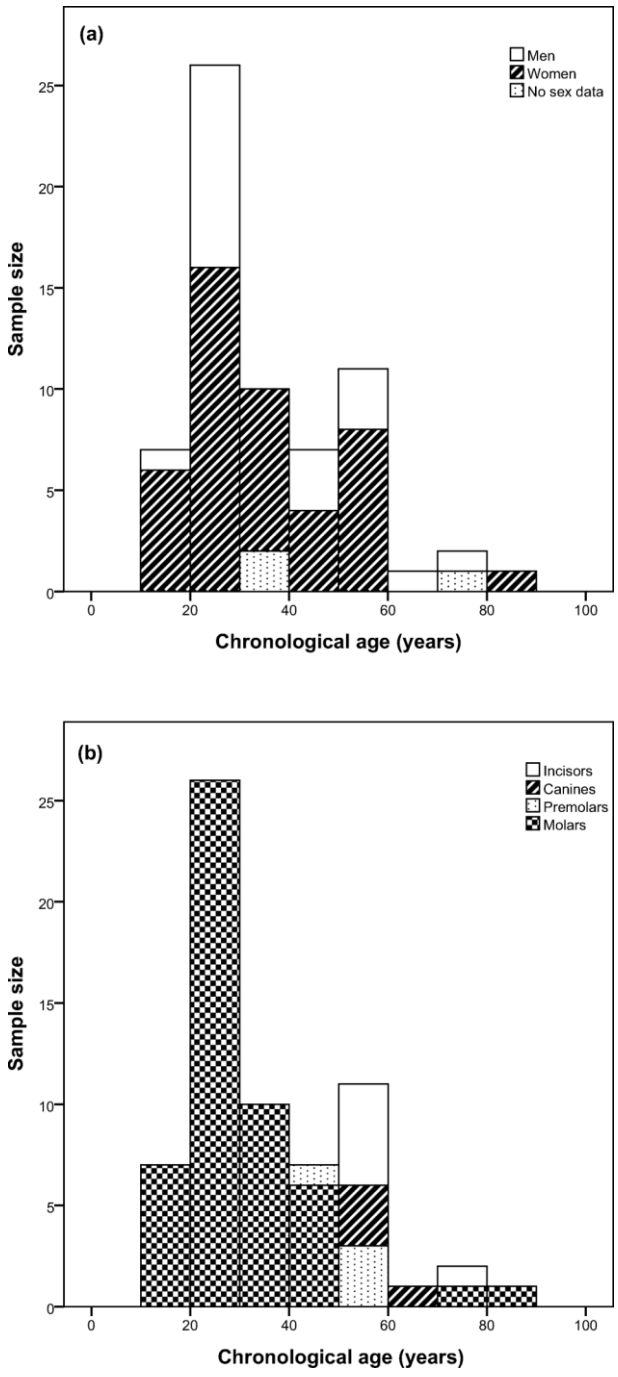


Fig. 1 Distribution of the samples (*n* = 65) according to chronological age by sex (a) and type of tooth (b)

Each extracted tooth was washed and decontaminated before being pulverized under liquid nitrogen in a cryogenic laboratory mill. Then DNA was extracted with a standard organic

method and quantified with the Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen™, Carlsbad, CA, USA) according to the protocols described in previous work [3].

DNA methylation and telomere length measurements

Samples of DNA (300 ng) were bisulfite converted with the EZ DNA Methylation-Gold™ Kit (Zymo Research, Irvine, CA, USA). Non-CpG cytosine residues were used as controls for bisulfite conversion. Then 2 µg of converted DNA was amplified by singleplex PCR in a total volume of 25 µl containing 0.2 µM of primers for the *ELOVL2*, *ASPA*, or *PDE4C* gene with the PyroMark PCR Kit (Qiagen, Hilden, Germany). Primers were designed with PyroMark Assay Design software, version 2.0.1.15 (Qiagen). All primer sequences are listed in Supplementary material, Table S1. The PCR reactions consisted of an initial activation step at 95°C for 15 min followed by 45 cycles of 30 s at 94°C, 30 s at 56°C, and 30 s at 72°C, and a final extension step at 72°C for 10 min.

The CpG sites considered in this study were numbered chronologically for each amplicon (see Table 1). To determine CpG methylation levels, 15 µl of biotinylated PCR product was immobilized on streptavidin-coated Sepharose beads (GE Healthcare, Chicago, IL, USA) followed by annealing to 40 µl of 0.4 µM sequencing primer (primer sequences in Supplementary material, Table S1) at 80°C for 2 min with a subsequent 10 min cooling down period. Pyrosequencing was done with PyroMark Gold Q96 Reagents (Qiagen) in the PyroMark Q96 ID Pyrosequencing System (Qiagen) according to the manufacturer's instructions. Methylated and non-methylated DNA (New England Biology, Ipswich, MA, USA) were included as controls in each assay.

The pyrosequencing results were analyzed with PyroMark Q96 ID Software 2.5 (Qiagen). Methylation level was expressed as the percentage of methylated cytosines over the total of methylated and non-methylated cytosines. Mean methylation level for each sample was estimated from triplicate determinations. Methylation data were expressed as the mean for each CpG site located in the *ELOVL2*, *ASPA*, or *PDE4C* genes.

Telomere length was determined by quantitative real-time PCR according to the protocol described in [3]. The results were recorded as the average relative telomere length, represented by the ratio of relative telomere repeat copy number (T) to single-copy gene copy number (*RPLP0* gene, S) [29]. To obtain the relative values, the T/S ratio at the third standard curve point was used as the reference.

Table 1 CpG sites identified for DNA methylation analysis, percentage of DNA methylation level, and age correlation coefficients for each CpG site

Gene	CpG site[a]	Chromosomal position	Methylation level (%)[b]	Age-correlation coefficients (*r*)[c]
ELOVL2	CpG1	chr6:11,044,678	7.45 ± 2.14	0.166
	CpG2	chr6:11,044,672	4.65 ± 2.10	0.371*
	CpG3	chr6:11,044,667	4.13 ± 2.05	0.203
	CpG4	chr6:11,044,661	5.41 ± 1.97	0.256*
	CpG5	chr6:11,044,654	6.20 ± 4.39	0.377*
	CpG6	chr6:11,044,649	4.52 ± 2.08	0.423**
	CpG7	chr6:11,044,646	7.00 ± 3.67	0.363*
	CpG8	chr6:11,044,644	4.56 ± 2.11	0.233
	CpG9	chr6:11,044,642	4.95 ± 2.28	0.595**
	CpG10	chr6:11,044,639	4.77 ± 2.26	0.491**
	CpG11	chr6:11,044,636	5.01 ± 2.00	0.399*
	CpG12	chr6:11,044,633	4.81 ± 2.17	0.356*
	CpG13	chr6:11,044,631	5.06 ± 3.59	0.410*
	CpG14	chr6:11,044,625	7.96 ± 4.24	0.377*
	CpG15	chr6:11,044,617	7.69 ± 2.90	0.220
ASPA	CpG1	chr17:3,379,567	48.03 ± 7.66	− 0.206
PDE4C	CpG1	chr19:18,343,916	12.66 ± 4.58	0.465**
	CpG2	chr19:18,343,911	12.12 ± 4.85	0.145
	CpG3	chr19:18,343,902	11.83 ± 3.81	0.206
	CpG4	chr19:18,343,889	8.96 ± 3.69	0.280*
	CpG5	chr19:18,343,881	11.58 ± 3.89	0.205

[a]The CpG sites were numbered chronologically for each amplicon

[b]Methylation level (mean ± standard deviation) is expressed as the percentage of methylated cytosines over the total of methylated and non-methylated cytosines

[c]Pearson correlation coefficients (*r*) between methylation level and chronological age. Statistically significant correlations: *($p < 0.05$); **($p < 0.001$)

Statistical analysis

A descriptive analysis of quantitative variables was carried out to obtain results reported here as the mean, standard deviation of the mean, and quartiles. For categorical variables, frequency

distribution was used. To compare the mean values for quantitative variables between sexes and types of tooth, nonparametric tests were used. Statistical dependence between variables was analyzed with the Pearson correlation coefficient (*r*).

Age prediction models were constructed with quantile regression analysis [30, 31] because of the non-normal distribution of age. Type of tooth (non-molar or molar) was included as a variable in all models in order to correct for the overrepresentation of molars in the sample. All models were computed including and excluding the variable sex (called SEX and NSEX, respectively). The CpG sites included in the models were selected with a stepwise ascending variable introduction strategy and a liberal entry criterion for predictor variables ($p < 0.20$). The variation inflation factor (VIF), which measures inflation of the variance of an estimated regression coefficient due to the presence of correlated predictors, was used to detect collinearity among predictor variables in the models. When VIF value is higher than 5, multicollinearity must be suspected [32, 33]. Pseudo R^2 was calculated to report the percentage of age variation that was explained by a given model [31]. Predictive accuracy was measured as the mean absolute error (MAE) [34, 35], which was computed as the mean of the absolute deviations between observed and predicted age. To correct for overestimation of predictive accuracy, the prediction models were validated using *k*-fold and leave-one-out cross-validation [36]. In the *k*-fold cross-validation technique, the dataset was randomly split into *k* subsets of the same size, called folds. Taking into account the sample size, setting *k* to 7 yielded stable and robust models. Each fold was used once as a testing (or validation) set, while the remaining *k*-1 folds formed the training set. The errors of each round were averaged to obtain the MAE. In the leave-one-out technique, one observation was left out of each round and age was predicted given all other observations. Predictive accuracy was also presented as the percentage correct classifications of actual age within a predicted age group. For this purpose, our sample was categorized by age into three groups, according to previous work [3]: young adults (15–35 years, $n = 38$), middle-aged adults (36–55 years, $n = 23$), and older adults (>55 years, $n = 4$).

All analyses were done with Stata version 14.1 software (StataCorp, College Station, TX, USA). A value of *p* less than 0.05 was considered statistically significant.

Results

Mean DNA methylation levels for each of the 21 CpG sites located in the *ELOVL2*, *ASPA*, and *PDE4C* genes are presented in Table 1. No significant differences ($p > 0.05$) were observed when mean DNA methylation levels were compared between sexes (Supplementary material, Table S2) or types of tooth (Supplementary material, Table S3) for any of the CpG sites analyzed.

In order to detect the most age-correlated CpG sites, age-correlation values were estimated from DNA methylation data. Table 1 shows the Pearson correlation coefficients for each detected CpG site and chronological age. CpG site methylation in *ELOVL2* and *PDE4C* genes was positively correlated with age. Several CpG sites from *ELOVL2* were significantly associated with age, with CpG9 ($r = 0.595$, $p < 0.001$), CpG10 ($r = 0.491$, $p < 0.001$), and CpG6 ($r = 0.423$, $p < 0.001$) identified as the most age-correlated sites. For the *PDE4C* gene, the significantly age-correlated CpG sites were CpG1 ($r = 0.465$, $p < 0.001$) and CpG4 ($r = 0.280$, $p < 0.05$). A negative correlation was found between methylation and age for the only CpG site in the *ASPA* gene; however, this correlation was not statistically significant ($r = -0.206$, $p > 0.05$).

Regarding telomere length (the other biomarker of aging analyzed here), no significant differences were found when mean relative telomere length was compared between sexes ($p > 0.05$; Supplementary material, Table S4). Regarding the type of tooth, our results showed significantly higher relative telomere length in molars ($p < 0.001$; Supplementary material, Table S4). A significant negative correlation between relative telomere length and chronological age was observed ($r = -0.549$, $p < 0.001$; Supplementary material, Fig. S1).

The values obtained for DNA methylation and telomere length were used as a training set to build multivariate quantile regression models for age estimation. All models were computed first by excluding the variable sex (models A_NSEX, B_NSEX, and C_NSEX), and later by including this variable (models A_SEX, B_SEX, and C_SEX). Prediction performance parameters for each model were calculated and are presented in Table 2. In the first model (model A_NSEX), we included nine CpG sites from the *ELOVL2* (CpG4, CpG6, CpG8, CpG9, CpG13, and CpG15) and *PDE4C* (CpG1, CpG2, and CpG4) genes, and obtained an MAE of 5.08 years in the training set. Model B_NSEX, which included only relative telomere length, yielded an MAE of 6.89 years in the training set. However, when relative telomere length and the nine CpG sites considered in model A_NSEX were included in a new model (model C_NSEX), the predictive accuracy of the model showed only slight improvement, yielding an MAE (5.04 years) similar to model A_NSEX. When the variable sex was included in these models, MAE values were slightly lower (4.84 years for model A_SEX, 6.71 years for model B_SEX, and 4.80 years for model C_SEX). The scatter plots for predicted versus chronological age for all models are shown in Fig. 2, Fig. 3, and Fig. 4. The VIF, used to detect collinearity, was never greater than 5 for any of the variables included in the models. Cross-validation of the quantile regression models for our data displayed somewhat higher MAE values, ranging from 6.06 years in model C_SEX to 7.45 years in model B_SEX (Table 2).

Table 2 Accuracy of age prediction models based on quantile regression analysis

Model[a]	Pseudo R^2	MAE	MAE (*k*-fold CV)	MAE (LOOCV)
Model A_NSEX: Age (years) = 32.269 – 18.142 × tooth_type – 0.870 × CpG4 *ELOVL2* + 0.241 × CpG6 *ELOVL2* – 2.822 × CpG8 *ELOVL2* + 3.523 × CpG9 *ELOVL2* + 0.773 × CpG13 *ELOVL2* – 0.493 × CpG15 *ELOVL2* + 0.807 × CpG1 *PDE4C* – 0.113 × CpG2 *PDE4C* + 0.691 × CpG4 *PDE4C*	0.584	5.08	7.05	6.43
Model A_SEX: Age (years) = 34.602 – 3.584 x sex – 16.761 × tooth_type – 0.871 × CpG4 *ELOVL2* + 0.339 × CpG6 *ELOVL2* – 1.984 × CpG8 *ELOVL2* + 3.973 × CpG9 *ELOVL2* + 0.340 × CpG13 *ELOVL2* – 0.521 × CpG15 *ELOVL2* + 0.854 × CpG1 *PDE4C* – 0.227 × CpG2 *PDE4C* + 0.539 × CpG4 *PDE4C*	0.603	4.84	6.24	6.16
Model B_NSEX: Age (years) = 56.467 – 24.237 x tooth_type – 6.667 x relative_T/S_ratio	0.436	6.89	7.17	7.16
Model B_SEX: Age (years) = 67.731 – 4.414 x sex – 19.226 x tooth_type – 13.729 x relative_T/S_ratio	0.450	6.71	7.45	7.34
Model C_NSEX: Age (years) = 35.507 – 18.076 × tooth_type – 1.711 x relative_T/S_ratio – 0.834 × CpG4 ELOVL2 + 0.067 × CpG6 ELOVL2 – 2.763 × CpG8 ELOVL2 + 3.460 × CpG9 ELOVL2 + 0.862 × CpG13 ELOVL2 – 0.517 × CpG15 ELOVL2 + 0.702 × CpG1 PDE4C – 0.099 × CpG2 PDE4C + 0.666 × CpG4 PDE4C	0.587	5.04	6.83	6.51
Model C_SEX: Age (years) = 39.130 – 4.057 x sex – 16.276 × tooth_type – 1.881 x relative_T/S_ratio – 1.039 × CpG4 ELOVL2 + 0.373 × CpG6 ELOVL2 – 1.876 × CpG8 ELOVL2 + 4.046 × CpG9 ELOVL2 + 0.488 × CpG13 ELOVL2 – 0.821 × CpG15 ELOVL2 + 0.837 × CpG1 PDE4C – 0.255 × CpG2 PDE4C + 0.496 × CpG4 PDE4C	0.607	4.80	6.13	6.06

MAE mean absolute error (in years), *k-fold CV k*-fold cross-validation, *LOOCV* leave-one-out cross-validation

[a]The CpG sites are detailed in Table 1.

In all models, the variable *tooth_type* took the value 0 when the tooth was not a molar and 1 when it was a molar. Models A_SEX, B_SEX, and C_SEX included the variable *sex*, which took the value 0 when the sex of the donor was unknown, 1 when the tooth belonged to a man, and 2 when it belonged to a woman

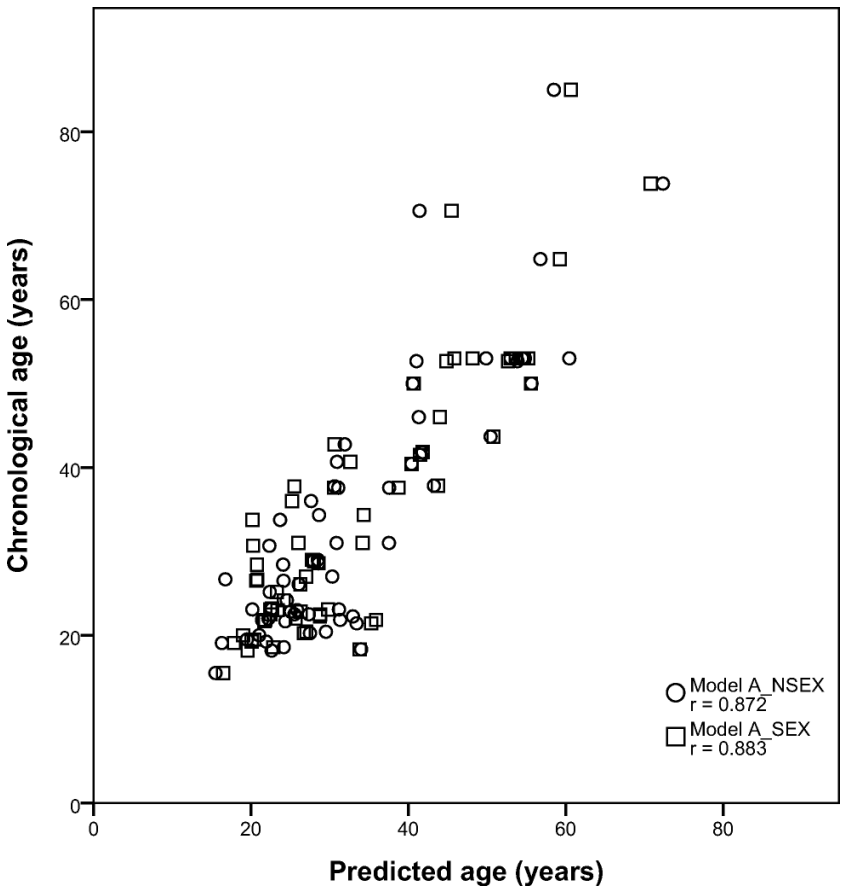


Fig. 2 Predicted versus chronological age for the training set (*n* = 65) in model A_NSEX and model A_SEX. Age prediction models are detailed in Table 2

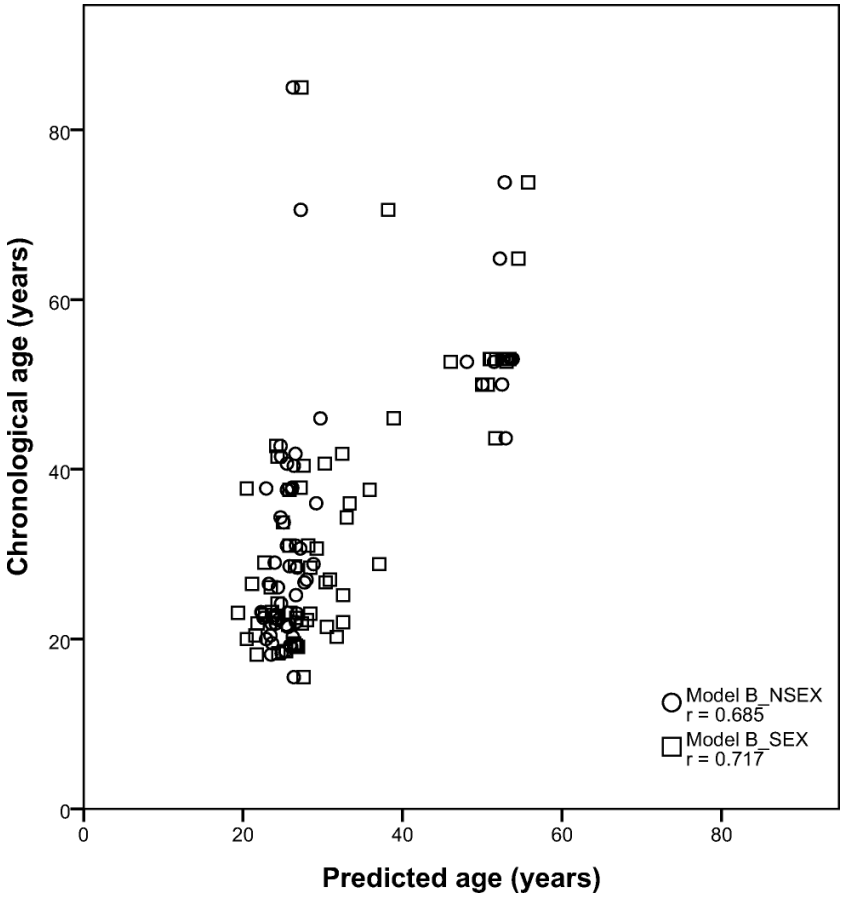


Fig. 3 Predicted versus chronological age for the training set (*n* = 65) in model B_NSEX and model B_SEX. Age prediction models are detailed in Table 2

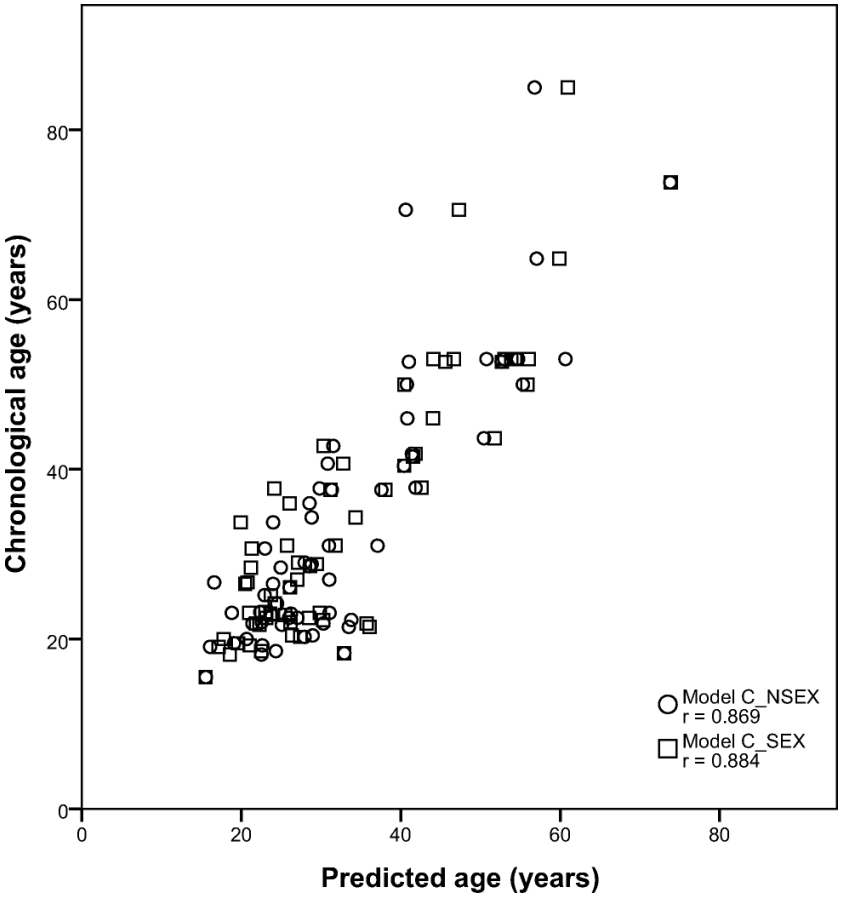


Fig. 4 Predicted versus chronological age for the training set (n = 65) in model C_NSEX and model C_SEX. Age prediction models are detailed in Table 2

Table 3 shows the predictive potential of the age prediction models according to the percentages of correct and incorrect classifications in each age group. The highest percentages of correct classifications (from 94.7% to 100%) were always obtained in the youngest age group (young adults: 15–35 years). In addition, these percentages decreased with increasing age in all models, and the decreases were largest in models that did not include DNA methylation markers (models B_NSEX and B_SEX).

Table 3 Number and percentage of correct and incorrect age group classifications in each age prediction model

Model[a]	Actual age group[b]	Predicted age group		
		15–35 years	36–55 years	>55 years
A_NSEX	15–35 years	37 (97.4%)*	1 (2.6%)	0
	36–55 years	5 (21.7%)	16 (69.6%)*	2 (8.7%)
	>55 years	0	1 (25%)	3 (75%)*
A_SEX	15–35 years	37 (97.4%)*	1 (2.6%)	0
	36–55 years	5 (21.7%)	17 (73.9%)*	1 (4.4%)
	>55 years	0	1 (25%)	3 (75%)*
B_NSEX	15–35 years	38 (100%)*	0	0
	36–55 years	11 (47.8%)	12 (52.2%)*	0
	>55 years	2 (50%)	2 (50%)	0*
B_SEX	15–35 years	37 (97.4%)*	1 (2.6%)	0
	36–55 years	9 (39.1%)	14 (60.9%)*	0
	>55 years	1 (25%)	2 (50%)	1 (25%)*
C_NSEX	15–35 years	37 (97.4%)*	1 (2.6%)	0
	36–55 years	5 (21.7%)	17 (73.9%)*	1 (4.4%)
	>55 years	0	1 (25%)	3 (75%)*
C_SEX	15–35 years	36 (94.7%)*	2 (5.3%)	0
	36–55 years	5 (21.7%)	16 (69.6%)*	2 (8.7%)
	>55 years	0	1 (25%)	3 (75%)*

[a]Age prediction models are detailed in Table 2

[b]Young adults (15–35 years, n = 38); middle-aged adults (36–55 years, n = 23), and older adults (>55 years, n = 4)

*Correct age group classifications

Discussion

Age-associated methylation was investigated for specific CpG sites located in the *ELOVL2*, *ASPA*, and *PDE4C* genes. In our sample, no significant differences were found in methylation levels between sexes for any CpG site, as others reported for the same and other genes [8, 9]. Similarly, no statistically significant differences between types of tooth were found for any methylation marker, in line with what others found in a previous study of *ELOVL2*, *FHL2*, and *PENK* genes in a limited number of premolars and molars [15]. To our knowledge, ours is the

first study designed to analyze and compare methylation levels of DNA extracted from all four types of tooth (incisors, canines, premolars, and molars). Our findings suggest the feasibility of this approach regardless of the type of tooth analyzed. However, the present results should be considered with caution due to the limited sample size for incisors, canines, and premolars.

The CpG sites most strongly associated with chronological age in our sample were CpG9 *ELOVL2* , CpG10 *ELOVL2* , CpG1 *PDE4C*, and CpG6 *ELOVL2* , all with Pearson correlation coefficients higher than 0.4 (Table 1). The correlation between DNA methylation and chronological age for CpG sites located in *ELOVL2* and *PDE4C* genes was previously reported in several studies [7, 12–18, 20, 22, 24]. On the other hand, and in contrast to previously published results [8, 19, 21, 23, 24], we did not find a significant correlation between *ASPA* DNA methylation and age. Nevertheless, large variations in the association between age and methylation levels with aging were previously reported in blood [8, 19, 21], as well as in buccal swabs and saliva samples [22]. In addition, the significant negative correlation found here between relative telomere length and chronological age (Supplementary material, Fig. S1) was consistent with results obtained in previous work [3]. In parallel with these results, mean relative telomere length was found not to be affected by sex, and to be significantly higher in molars (Supplementary material, Table S4). The fact that molars in our sample, as in a previous study [3], belonged mostly to the youngest individuals may have affected these results regarding the differences between types of tooth.

We used the present dataset to develop multivariate quantile regression models for age estimation. The method of choice for age prediction modeling with DNA methylation markers should be determined by the normal or non-normal distribution of CpG methylation, the linearity of the relationships, collinearity, and non-constant variance or heteroscedasticity in the dataset [11]. The influence of these factors has led forensic researchers to use different statistical approaches, including multivariate quantile regression [9, 19, 37]. Quantile regression appeared to perform better when the data were characterized by non-constant and non-normally distributed variance [37], as in our dataset. In addition, all models were adjusted for type of tooth in order to control for the overrepresentation of molars in the sample (n = 51) as a potential confounding factor. We built an initial age prediction model (model A_NSEX, Table 2 and Fig. 2) that included nine CpG sites located in the *ELOVL2* and *PDE4C* genes; this model yielded an MAE of 5.08 years. This prediction error is in line with those of previously published DNA methylation-based age estimation models (reviewed in [7, 10, 11]). Specifically for teeth, Bekaert et al. [8] reported a multivariate quadratic regression model constructed from the methylation level of seven CpG sites located in the *PDE4C*, *ELOVL2* and *EDARADD* genes in 29 dentine samples, with an R^2 of 0.74 and a mean absolute deviation between chronological and predicted age of 4.86 years. In related research, Giuliani et al. [15] presented different multiple linear regression models built from a limited number of dental pulp, dentine,

cementum, or cementum and pulp samples, and including 5, 8, or 13 CpG sites located in the *ELOVL2*, *FHL2*, and *PENK* genes. Their models yielded an R^2 of 0.50–0.97 and a median absolute difference between chronological and predicted age of 1.20–7.07 years in the training set, depending on the tissues analyzed. In contrast, the present model, which included only relative telomere length (model B_NSEX, Table 2 and Fig. 3), provided a higher MAE (6.89 years). Moreover, the addition of relative telomere length to the variables included in model A_NSEX only slightly improved the prediction error in the training set (model C_NSEX, Table 2 and Fig. 4). Consequently, DNA methylation appeared to be more informative than telomere length for estimating age in our sample. These results may indicate the limited usefulness of telomere length as a supplementary marker for age estimation based on DNA methylation in teeth. Therefore, the present findings suggest that including more than one parameter in the models does not always improve the accuracy of age estimation, in contrast to what others have suggested for the combined analysis of parameters concerning different biological levels [38–40]. Nevertheless, it should be remembered that our models require validation with other datasets.

Although the present results showed that sex did not play a significant role in the age associations (Supplementary material, Table S2 and Table S4), after adjustment for sex, all models showed slightly better predictive accuracies, as seen in Table 2, Fig. 2, Fig. 3, and Fig.4. However, from a forensic point of view, an age prediction model that does not include sex as a predictor variable may be more useful, since in many cases only a fragment of mandible (or just a single tooth) is recovered, and the sex of the remains is unknown.

Since additional data could not be gathered, cross-validation was performed to assess the testing error of our age prediction models, as others have suggested [36]. It should be taken into account that any cross-validation mechanism, especially *k*-fold cross-validation, produces extraordinarily variable results with a sample of the size we studied. Therefore, leave-one-out cross-validation is more advisable for training sets of small sizes [36], as is the case here. Although cross-validation errors were somewhat higher than training errors for all models (Table 2), leave-one-out cross-validation yielded errors closer to training errors than those provided by *k*-fold cross-validation. In addition, the models that included DNA methylation markers as predictors (model A_NSEX, model C_NSEX, model A_SEX, and model C_SEX) showed larger differences between training and cross-validation errors. This finding suggests overestimation of the predictive capacity of these models, possibly explained by their greater complexity. However, the most complex model (model C_SEX) included 12 variables and had 53 degrees of freedom, which clearly demarcates it from the saturated model. Moreover, multicollinearity can be ruled out since the VIF was never greater than 5 for any of the predictor variables [32, 33].

One of the main limitations of age estimation methods is that their accuracy becomes lower with increasing age. It is well known that molecular changes with aging can be influenced by a variety of external and internal factors. Age-related DNA methylation changes do not occur at a constant rate throughout life, but accumulate rapidly up to adulthood [41]. Similarly, rates of telomere attrition vary markedly at different ages. Telomere length is rapidly reduced in early childhood, followed by an apparent plateau until young adulthood and gradual attrition throughout adult life [42]. In all models built with our sample, predictive accuracy decreased with increasing age (Table 3). Models A_NSEX (including DNA methylation markers) and C_NSEX (including DNA methylation markers and telomere length) showed similar percentages of correct classification in the older age groups, and model B_NSEX (including telomere length) presented the lowest percentages (Table 3). Similar results were obtained when these models were adjusted for sex. Thus, the combined use of different biomarkers of aging did not improve predictive accuracy at older ages in our sample, in contrast to what others have suggested [38, 40]. Moreover, DNA methylation was shown to be more accurate than telomere length in predicting age at older ages.

In conclusion, several CpG sites from the *ELOVL2* and *PDE4C* genes, as well as telomere length, were significantly associated with chronological age in our sample of human teeth. We developed accurate age prediction quantile regression models for tooth samples, adjusted for type of tooth and sex, using DNA methylation markers with and without telomere length as an additional variable. Our results suggest that telomere length may have limited usefulness as a supplementary marker in DNA methylation-based age estimation in tooth samples, given that the latter marker appeared to be more informative than telomere-based age estimation when both biomarkers were evaluated separately. However, the prediction models reported here should be validated with large datasets in order to ascertain the testing error of the method. Another task for future research is to assess the applicability of our approach in teeth from cadavers with different postmortem intervals.

Acknowledgments

Appreciation is expressed to the Genomics ECAI of the Institute of Biomedical Research in Malaga (IBIMA), Malaga, Spain, the Pfizer-University of Granada Andalusian Government Centre for Genomics and Oncological Research (GENYO), Granada, Spain, for their technical support, methodological expertise and scientific advice, and K. Shashok for improving the use of English in the manuscript. The authors also acknowledge funding from the Andalusian Centre of Excellence for Forensic Research (CEIFA).

Funding

The authors also acknowledge funding from the Andalusian Centre of Excellence for Forensic Research (CEIFA).

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Cunha E, Baccino E, Martrille L, et al (2009) The problem of aging human remains and living individuals: A review. Forensic Sci Int 193:1–13. doi: 10.1016/j.forsciint.2009.09.008
2. Meissner C, Ritz-Timme S (2010) Molecular pathology and age estimation. Forensic Sci Int 203:34–43. doi: 10.1016/j.forsciint.2010.07.010
3. Márquez-Ruiz AB, González-Herrera L, Valenzuela A (2018) Usefulness of telomere length in DNA from human teeth for age estimation. Int J Legal Med 132:353–359. doi: 10.1007/s00414-017-1595-5
4. Valenzuela A, Guerra-Hernández E, Rufián-Henares JÁ, et al (2018) Differences in non-enzymatic glycation products in human dentine and clavicle: changes with aging. Int J Legal Med 132:1749–1758. doi: 10.1007/s00414-018-1908-3
5. Corrêa HSD, Pedro FLM, Volpato LER, et al (2017) Forensic DNA typing from teeth using demineralized root tips. Forensic Sci Int 280:164–168. doi: 10.1016/j.forsciint.2017.10.003
6. Jones MJ, Goodman SJ, Kobor MS (2015) DNA methylation and healthy human aging. Aging Cell 14:924–932. doi: 10.1111/acel.12349
7. Freire-Aradas A, Phillips C, Lareu M (2017) Forensic individual age estimation with DNA: From initial approaches to methylation tests. Forensic Sci Rev 29:121–144.
8. Bekaert B, Kamalandua A, Zapico SC, et al (2015) Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. Epigenetics 10:922–930. doi: 10.1080/15592294.2015.1080413
9. Freire-Aradas A, Phillips C, Girón-Santamaría L, et al (2018) Tracking age-correlated DNA methylation markers in the young. Forensic Sci Int Genet 36:50–59. doi: 10.1016/J.FSIGEN.2018.06.011
10. Parson W (2018) Age estimation with DNA: From forensic DNA fingerprinting to forensic (Epi)genomics: A mini-review. Gerontology 64:326–332. doi: 10.1159/000486239

11. Vidaki A, Kayser M (2018) Recent progress, methods and perspectives in forensic epigenetics. Forensic Sci Int Genet 37:180–195. doi: 10.1016/J.FSIGEN.2018.08.008
12. Hannum G, Guinney J, Zhao L, et al (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell 49:359–367. doi: 10.1016/j.molcel.2012.10.016
13. Zbieć-Piekarska R, Spólnicka M, Kupiec T, et al (2014) Examination of DNA methylation status of the ELOVL2 marker may be useful for human age prediction in forensic science. Forensic Sci Int Genet 14:161–167. doi: 10.1016/j.fsigen.2014.10.002
14. Zbieć-Piekarska R, Spólnicka M, Kupiec T, et al (2015) Development of a forensically useful age prediction method based on DNA methylation analysis. Forensic Sci Int Genet 17:173–179. doi: 10.1016/j.fsigen.2015.05.001
15. Giuliani C, Cilli E, Bacalini MG, et al (2016) Inferring chronological age from DNA methylation patterns of human teeth. Am J Phys Anthropol 159:585–595. doi: 10.1002/ajpa.22921
16. Park JL, Kim JH, Seo E, et al (2016) Identification and evaluation of age-correlated DNA methylation markers for forensic use. Forensic Sci Int Genet 23:64–70. doi: 10.1016/j.fsigen.2016.03.005
17. Hamano Y, Manabe S, Morimoto C, et al (2016) Forensic age prediction for dead or living samples by use of methylation-sensitive high resolution melting. Leg Med 21:5–10. doi: 10.1016/j.legalmed.2016.05.001
18. Zubakov D, Liu F, Kokmeijer I, et al (2016) Human age estimation from blood using mRNA, DNA methylation, DNA rearrangement, and telomere length. Forensic Sci Int Genet 24:33–43. doi: 10.1016/j.fsigen.2016.05.014
19. Freire-Aradas A, Phillips C, Mosquera-Miguel A, et al (2016) Development of a methylation marker set for forensic age estimation using analysis of public methylation data and the Agena Bioscience EpiTYPER system. Forensic Sci Int Genet 24:65–74. doi: 10.1016/j.fsigen.2016.06.005
20. Spólnicka M, Pośpiech E, Pepłońska B, et al (2018) DNA methylation in ELOVL2 and C1orf132 correctly predicted chronological age of individuals from three disease groups. Int J Legal Med 132:1–11. doi: 10.1007/s00414-017-1636-0
21. Weidner CI, Lin Q, Koch CM, et al (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biol 15:R24. doi: 10.1186/gb-2014-15-2-r24
22. Eipel M, Mayer F, Arent T, et al (2016) Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures. Aging (Albany NY) 8:1034–1048. doi: 10.18632/aging.100972
23. Huang Y, Yan J, Hou J, et al (2015) Developing a DNA methylation assay for human

age prediction in blood and bloodstain. Forensic Sci Int Genet 17:129–136. doi: 10.1016/j.fsigen.2015.05.007

24. Vidal-Bralo L, Lopez-Golan Y, Gonzalez A (2016) Simplified assay for epigenetic age estimation in whole blood of adults. Front Genet 7:126. doi: 10.3389/fgene.2016.00126
25. Xu C, Qu H, Wang G, et al (2015) A novel strategy for forensic age prediction by DNA methylation and support vector regression model. Sci Rep 5:17788. doi: 10.1038/srep17788
26. Jung SE, Shin KJ, Lee HY (2017) DNA methylation-based age prediction from various tissues and body fluids. BMB Rep 50:546–553. doi: 10.5483/BMBRep.2017.50.11.175
27. Naue J, Sänger T, Hoefsloot HCJ, et al (2018) Proof of concept study of age-dependent DNA methylation markers across different tissues by massive parallel sequencing. Forensic Sci Int Genet 36:152–159. doi: 10.1016/j.fsigen.2018.07.007
28. Higgins D, Austin JJ (2013) Teeth as a source of DNA for forensic identification of human remains: A Review. Sci Justice 53:433–441. doi: 10.1016/j.scijus.2013.06.001
29. Cawthon RM (2002) Telomere measurement by quantitative PCR. Nucleic Acids Res 30:e47. doi: 10.1093/nar/30.10.e47
30. Furno M, Vistocco D (2018) Quantile regression: estimation and simulation, Volume 2. John Wiley & Sons, Hoboken, NJ
31. Koenker R (2005) Quantile regression. Cambridge University Press, Cambridge
32. Kutner MH, Nachtsheim CJ, Neter J (2004) Applied linear regression models, Fifth Edit. McGraw-Hill/Irwin, New York, NY
33. Sheather S (2009) A modern approach to regression with R. Springer, New York, NY
34. Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 30:79–82. doi: 10.3354/cr030079
35. Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. Int J Forecast 22:679–688. doi: 10.1016/j.ijforecast.2006.03.001
36. Valsecchi A, Irurita Olivares J, Mesejo P (2019) Age estimation in forensic anthropology: methodological considerations about the validation studies of prediction models. Int J Legal Med 133:1915–1924. doi: 10.1007/s00414-019-02064-7
37. Smeers I, Decorte R, Van de Voorde W, Bekaert B (2018) Evaluation of three statistical prediction models for forensic age prediction based on DNA methylation. Forensic Sci Int Genet 34:128–133. doi: 10.1016/J.FSIGEN.2018.02.008
38. Cho S, Jung S-E, Hong SR, et al (2017) Independent validation of DNA-based approaches for age prediction in blood. Forensic Sci Int Genet 29:250–256. doi: 10.1016/J.FSIGEN.2017.04.020
39. Shi L, Jiang F, Ouyang F, et al (2018) DNA methylation markers in combination with

skeletal and dental ages to improve age estimation in children. Forensic Sci Int Genet 33:1–9. doi: 10.1016/j.fsigen.2017.11.005

40. Becker J, Mahlke NS, Reckert A, et al (2019) Age estimation based on different molecular clocks in several tissues and a multivariate approach: an explorative study. Int J Legal Med 1–13. doi: 10.1007/s00414-019-02054-9

41. Alisch RS, Barwick BG, Chopra P, et al (2012) Age-associated DNA methylation in pediatric populations. Genome Res 22:623–632. doi: 10.1101/gr.125187.111

42. Ishikawa N, Nakamura KI, Izumiyama-Shimomura N, et al (2016) Changes of telomere status with aging: An update. Geriatr Gerontol Int 16:30–42. doi: 10.1111/ggi.12772

SUPPLEMENTARY MATERIAL

Supplementary Table S1 Primer sequences for pyrosequencing assays of each gene analyzed

Primer	Sequence
ELOVL2 forward	GAGGTTTTTGTGGGAGAG
ELOVL2 reverse	ACCCCACTCACCATAATC
ELOVL2 sequencing	TTTGTGGGAGAGGGG
ASPA forward	TTTTGGAGGAATTTATGGGAATGAGT
ASPA reverse	TACCTCCAACCCTATTCTCTAAATCTC
ASPA sequencing	TTTTGGTTAAGTATTGGTTAG
PDE4C forward	AAGGGGTAGAGGTTTGTAGTA
PDE4C reverse	CCACCCCTACCTAAAAATAACTTTTTCCT
PDE4C sequencing	GGTAGTTATAGTATGATTAGAGT

Supplementary Table S2 Percentage of DNA methylation levels by sex for each of the CpG sites analyzed

Gene	CpG number[a]	Methylation level (%)[b]	
		Men (*n* = 19)	Women (*n* = 43)
ELOVL2	CpG1	6.95 ± 2.21	7.72 ± 2.12
	CpG2	4.50 ± 2.07	4.81 ± 2.16
	CpG3	3.63 ± 1.50	4.46 ± 2.24
	CpG4	5.09 ± 1.97	5.60 ± 2.03
	CpG5	7.15 ± 6.51	5.95 ± 3.17
	CpG6	4.33 ± 1.81	4.69 ± 2.22
	CpG7	7.81 ± 4.53	6.78 ± 3.32
	CpG8	4.27 ± 2.02	4.86 ± 2.07
	CpG9	4.52 ± 2.46	5.18 ± 2.27
	CpG10	4.81 ± 2.33	4.77 ± 2.32
	CpG11	4.75 ± 1.93	5.01 ± 1.99
	CpG12	4.83 ± 2.06	4.89 ± 2.29
	CpG13	5.70 ± 5.12	4.85 ± 2.84
	CpG14	7.98 ± 4.73	8.12 ± 4.13
	CpG15	6.77 ± 2.03	8.08 ± 3.17
ASPA	CpG1	50.51 ± 7.00	47.06 ± 8.01
PDE4C	CpG1	12.00 ± 5.48	12.83 ± 3.79
	CpG2	13.40 ± 6.26	11.51 ± 4.13
	CpG3	11.90 ± 4.12	11.90 ± 3.77
	CpG4	9.70 ± 4.46	8.65 ± 3.42
	CpG5	11.93 ± 3.76	11.35 ± 3.89

[a]The CpG sites are detailed in Table 1

[b]Methylation level (mean ± standard deviation) is expressed as the percentage of methylated cytosines over the total of methylated and non-methylated cytosines

Supplementary Table S3 Percentage of DNA methylation levels by type of tooth for each of the CpG sites analyzed

Gene	CpG number[a]	Methylation level (%)[b]			
		Incisors (n = 6)	Canines (n = 4)	Premolars (n = 4)	Molars (n = 51)
ELOVL2	CpG1	7.68 ± 3.69	7.20 ± 1.04	8.18 ± 3.03	7.39 ± 1.96
	CpG2	6.15 ± 2.50	4.01 ± 1.72	6.40 ± 1.85	4.39 ± 2.00
	CpG3	5.38 ± 1.85	4.07 ± 1.61	6.18 ± 2.83	3.83 ± 1.95
	CpG4	7.65 ± 3.31	5.96 ± 2.34	5.95 ± 1.32	5.06 ± 1.64
	CpG5	13.75 ± 9.10	5.16 ± 2.65	8.31 ± 3.57	5.23 ± 2.63
	CpG6	6.78 ± 2.11	4.27 ± 1.97	7.70 ± 3.07	4.02 ± 1.61
	CpG7	13.12 ± 5.45	6.07 ± 2.75	9.05 ± 2.90	6.19 ± 2.78
	CpG8	6.58 ± 2.84	3.71 ± 1.30	5.83 ± 2.29	4.29 ± 1.93
	CpG9	6.68 ± 3.12	5.97 ± 2.01	7.16 ± 2.03	4.49 ± 2.04
	CpG10	8.18 ± 3.14	4.99 ± 1.01	6.38 ± 2.12	4.22 ± 1.82
	CpG11	7.10 ± 2.76	4.68 ± 1.73	6.03 ± 1.52	4.71 ± 1.83
	CpG12	6.85 ± 2.68	4.94 ± 2.19	6.36 ± 2.10	4.44 ± 1.98
	CpG13	10.00 ± 7.99	4.03 ± 1.36	6.99 ± 1.30	4.41 ± 2.48
	CpG14	12.98 ± 6.16	6.96 ± 3.36	12.13 ± 8.52	7.13 ± 3.00
	CpG15	8.86 ± 3.25	9.92 ± 7.67	8.34 ± 3.30	7.32 ± 2.19
ASPA	CpG1	50.28 ± 7.06	46.93 ± 10.91	34.17± 8.56	48.93 ± 6.45
PDE4C	CpG1	15.48 ± 6.05	14.05 ± 1.94	13.45 ± 2.26	12.16 ± 4.61
	CpG2	13.22 ± 2.96	9.22 ± 1.12	11.35 ± 2.02	12.28 ± 5.30
	CpG3	11.61 ± 2.40	12.05 ± 3.58	11.87 ± 2.64	11.84 ± 4.10
	CpG4	8.63 ± 2.60	9.62 ± 3.19	8.80 ± 3.54	8.96 ± 3.91
	CpG5	11.49 ± 3.14	11.23 ± 3.36	12.72 ± 3.80	11.53 ± 4.08

[a]The CpG sites are detailed in Table 1

[b]Methylation level (mean ± standard deviation) is expressed as the percentage of methylated cytosines over the total of methylated and non-methylated cytosines

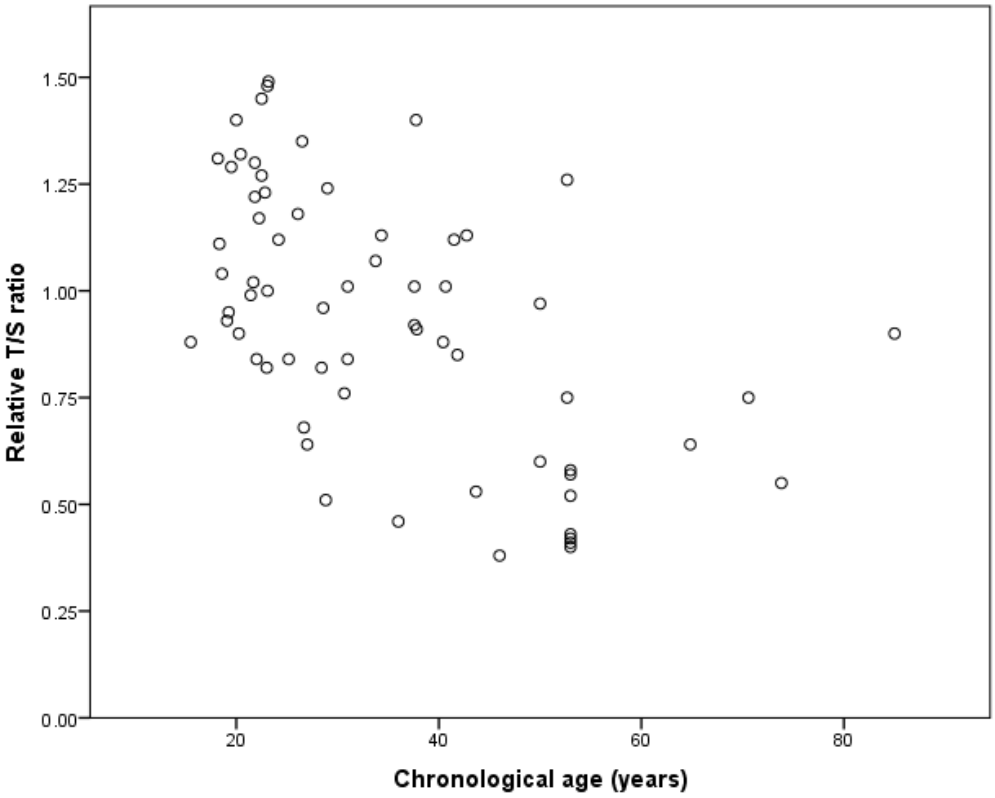
Supplementary Table S4 Relative telomere length by sex and type of tooth

		n	Relative T/S ratio[a]
Sex	Men	19	0.97 ± 0.32
	Women	43	0.92 ± 0.31
Type of tooth	Incisors	6	0.72 ± 0.34
	Canines	4	0.52 ± 0.11
	Premolars	4	0.55 ± 0.03
	Molars	51	1.03 ± 0.26[*]

n number of cases

[a] Relative telomere length (mean ± standard deviation) is expressed as the ratio of relative telomere repeat copy number (T) to single-copy gene copy number (S), as described in [29]

[*]Significant differences between molars and incisors ($p < 0.05$), molars and canines ($p < 0.01$), and molars and premolars ($p < 0.01$)



Supplementary Fig. S1 Correlation between relative telomere length (expressed as relative T/S ratio) and chronological age ($r = -0.549$, $p < 0.001$)