RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2024.028

ScienceAsia 50 (1): 2024: ID 2024028: **1–9**

Age-associated DNA methylation changes with hypermethylation *ELOVL2* and *ZYG11A* in dental samples: A pilot study

Praewa Pinprasert[a], **Supakit Khacha-ananda**[b,c,*], **Suteera Narakornsak**[d], **Klintean Wunnapuk**[b]

[a] Master's Degree in Forensic Science, Graduate School, Chiang Mai University, Chiang Mai 50200 Thailand
[b] Department of Forensic Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai 50200 Thailand
[c] Research Center in Bioresources for Agriculture, Industry and Medicine, Chiang Mai University, Chiang Mai 50200 Thailand
[d] Department of Anatomy, Faculty of Medicine, Chiang Mai University, Chiang Mai 50200 Thailand

*Corresponding author, e-mail: supakit.kh@cmu.ac.th

Received 2 Aug 2023, Accepted 7 Feb 2024
Available online 10 Mar 2024

ABSTRACT: The accuracy of age prediction based on physical appearance diminishes as body degradation occurs, prompting the suggestion of utilizing skeletal remains as an alternative method. Molecular studies investigating epigenetic responses have proposed the development of age prediction models. Certain studies focusing on teeth or bones have explored the correlation between DNA methylation and actual age. However, the influence of diverse ethnicities on DNA methylation levels poses a challenge to the generalizability of these models, limiting their applicability to specific populations of origin. This study aims to investigate the DNA methylation levels in teeth to construct an age prediction model. Teeth were collected from corpses, and the pyrosequencing technique was employed to measure the quantification of DNA methylation levels on *ELOVL2*, *ZYG11A*, and *TRIM59* genes. Significant correlations were observed between several methylated *ELOVL2*, *ZYG11A*, and *TRIM59* levels and actual age, and these correlations had not been reported before. A moderately positive linear relationship (r value between 0.365–0.643) was constructed for age prediction. The performance of the predicted models was evaluated using Leave-One-Out Cross-Validation, indicating an approximate error of 11–13 years from the actual age. In conclusion, our findings provide preliminary data to the discovery of significantly methylated genes for the construction of age prediction model for further study and the utilization of the DNA methylation-based approach for application in forensic aspect.

KEYWORDS: DNA methylation, age estimation, teeth, pyrosequencing

INTRODUCTION

In the context of criminal proceedings, age estimation is a crucial approach for identifying individuals in the absence of relevant documents. Various techniques, including forensic genetics, forensic anthropology, forensic radiography, and forensic odontology, have been employed for age estimation in adult corpses. However, estimating age based on physical appearance becomes challenging due to the decomposition of the corpse. Furthermore, biological samples such as blood, saliva, buccal mucosa, semen, or other tissues cannot be collected from corpses due to environmental degradation and decay [1, 2]. Therefore, skeletal remains, bones, or teeth have emerged as potential sources for personal identification. Recently, DNA evidence has been proven valuable for forensic purposes, particularly in age estimation. Previous studies have demonstrated that genetic-based techniques, such as measuring telomere length, mRNA expression, DNA rearrangement, and DNA methylation, can be utilized to estimate the age of death [3]. Both cellular senescence and biological aging are associated with telomere shortening. However, using telomere length as a model for age estimation exhibits a wide range of errors and is not widely adopted in forensic applications [4, 5]. Age prediction models generated through the analysis of mRNA expression and signal joint T-cell receptor rearrangement excision circle (sjTREC) have been reported to have predictive accuracy. However, this technique is limited to specific tissues, such as fresh blood and tissues from fresh cadavers [6]. Consequently, quantifying DNA methylation has emerged as the most reliable method for age estimation, demonstrating lower predicted errors in comparison to the actual age [4, 7]. DNA methylation is an epigenetic mechanism that involves the transfer of a methyl group onto the C5 position of cytosine, forming 5-methylcytosine [8].

In terms of the type of specimen used for quantifying DNA methylation, the prior study discovered that age-related DNA methylation is relatively tissue-specific [9]. Teeth, being the hardest structures in the human body, are resistant to decay in extreme environments and degrading contamination, thus preserving teeth DNA. Teeth DNA can be found in dental tissues such as dentine, cementum, and the pulp complex [10]. The Disaster Victim Identification (DVI) Guide 2018 and the DNA Commission of the International Society of Forensic Genetics (ISFG) suggest that

healthy teeth, preferably molars and premolars, are suitable choices for collecting samples from unidentified human remains [11]. Several studies have been conducted on the estimation of age using DNA methylation. Previous investigations summarized that the methylation levels of *ELOVL2*, *RPA2*, *TRIM59*, and *ZYG11A* genes were effective for age prediction in blood, bone, brain, buccal swab, and muscle samples [12, 13]. Unfortunately, fewer studies have been carried out on the association between dental DNA methylation and biological age.

ELOVL2 (ELOVL fatty acid elongase 2), *ZYG11A* (Zyg-11 Family Member A), and *TRIM59* (Tripartite Motif Containing 59) have been identified age predictors. The correlation between the methylation of *ELOVL2*, *TRIM59*, *ZYG11A*, and actual age has been reported in various biological samples, including blood, buccal swabs, and saliva [12]. However, the methylation analysis of *TRIM59* and *ZYG11A* in teeth remains unexplored. Furthermore, it has been noted that different populations can impact genetic background. Previous studies have shown that the accuracy and precision of predictive models vary when applied to populations other than the one from which the model was developed. It is assumed that race or ethnicity may influence the effectiveness of the model in predicting age within individual populations [14, 15]. Moreover, there is a scarcity of studies on developing age prediction models based on the methylation of these three genes for the Thai population [16]. Therefore, it is crucial to consider the ethnicity of individuals when utilizing such models to estimate age based on DNA methylation rates. As a result of these considerations, the three genes, *ELOVL2*, *ZYG11A*, and *TRIM59*, were selected for establishing age prediction in our study. The objectives of this research are to evaluate DNA methylation in teeth, specifically at CpG sites located on the *ELOVL2*, *ZYG11A*, and *TRIM59* genes in samples of the Thai population and to construct a preliminarily predictive model for age estimation. The data obtained from this study can be used to develop an age prediction model in the future.

MATERIALS AND METHODS

Sample collection and preparation

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Faculty of Medicine, Chiang Mai University, Thailand (No. 019/2022). Thirty-six healthy permanent maxillary right first premolar teeth were collected from corpses aged between 17–72 years within 24 h after death at the Department of Forensic Medicine, Faculty of Medicine, Chiang Mai University. Sample inclusions were participants without any diseases impacting DNA methylation levels, such as cancer, atherosclerosis, nervous disorders, or cardiovascular diseases; nor any disease-related gum or tooth infections including dental caries, enamel hypomyelination, periodontitis, or oral carcinogenesis [17]. The criteria for determining Thai ethnicity include verifying information of family history obtained from interviewing the deceased's relatives, the history in medical record, and the National ID Number with the initial digits of 1, 2, and 3 for individuals with Thai nationality. Fresh teeth were brushed under distilled water and dried at room temperature. Teeth were soaked in liquid nitrogen and ground using Freezer Mil 6750 (SPEX SamplePrep, New Jersey, USA) before storage at −20 °C.

DNA extraction and quantification

Genomic DNA was extracted using QIAamp DNA Investigator Kit (Qiagen, Hilden, Germany). Approximately 100–200 mg of powdered teeth was mixed with Buffer ATL and proteinase K and incubated overnight. After the incubation, buffer AL was added, and the solution was incubated with shaking, then, centrifuged. The supernatant was collected, mixed with absolute ethanol and transferred into QIAamp MinElute column for DNA binding. The extracted DNA was washed in sequence with Buffer AW1, Buffer AW2, and 70% ethanol. Finally, purified DNA was eluted into a microcentrifuge and quantified using NanoDrop (Thermo Fisher Science, MA, USA).

Bisulfite conversion

To distinguish between methylated and unmethylated cytosine, bisulfite technique was used for a conversion of unmethylated cytosine to uracil according to manufacturer's instruction (Epitect Fast Bisulfite Conversion Kit, Qiagen, Hilden, Germany). Briefly, 100 ng of purified DNA was mixed with RNase free water, bisulfite solution, and DNA protect buffer. The solution was incubated at 56 °C for 5 min, followed with 60 °C for 10 min, 95 °C for 5 min, 60 °C for 10 min, and held at 20 °C using a thermal cycler (Qiagen, Hilden, Germany). The bisulfite DNA product was desulphonated by Buffer BD and eluted into a microcentrifuge. Finally, the concentration of bisulfited DNA was quantified using NanoDrop (Thermo Fisher Science, MA, USA) for next steps.

PCR pyrosequencing and methylation analysis

Approximately 20–30 ng of bisulfite converted DNA was amplified by polymerase chain reaction (PCR). The PCR reaction mixture contained bisulfite converted DNA as a template, 2X PyroMark mastermix, 10X CoralLoad concentration, 10 µM forward and reverse primer. The assay design program 2.0 was used to generate the primer sequences (Table 1). The primers for all three genes were designed by PyroMark PCR software with slight modification from a previous publication [18]. The PCR was performed by Veriti™ 96-Well Fast Thermal Cycler PCR Systems

Table 1 PyroMark CpG assay design, including primer name, primer sequencing, and primer details in *ELOVL2*, *ZYG11A* and *TRIM59* genes.

Primer name	Primer sequences (5′ → 3′)	Tm	% GC	PCR size (bp)
ELOVL2-FP	AGGGGAGTAGGGTAAGTGAG	57.8	55.0	
ELOVL2-RPB	ACAAAACCATTTCCCCCTAATATA	59.4	33.0	307
ELOVL2-SP	GGGAGGAGATTTGTAGGTTT	42.4	45.0	
ZYG11A-FP	TAGAGGTATTTGTTGGGGAGT	56.2	42.9	
ZYG11A-RPB	AACCAAACCCAATACCATACCCT	56.7	43.5	290
ZYG11A-SP	TTTGTTGGGGAGTGT	45.7	46.7	
TRIM59-FP	TATGGTATAGGTGGTTTGGGGGAGA	62.3	48.0	
TRIM59-RPB	ATAAAAAACACTACCCTCCACAACATAACA	58.5	33.3	148
TRIM59-SP	TTGGGGGAGAGGTTG	50.5	60.0	

(Thermo Fisher Science, MA, USA), with an initial PCR activation step at 94 °C for 15 min, followed by 45 cycles at 94 °C for 30 s, 56 °C for 30 s, 72 °C for 30 s, and final extension step at 72 °C for 10 min according to the manufacturer's instruction (PyroMark PCR Kit, Qiagen, Hilden, Germany). The verification of PCR product prior to pyrosequencing analysis was confirmed by analysis of amplifying product on 1.5% agarose gel electrophoresis. Finally, PCR product was loaded into the PyroMark Q48 Instrument (Qiagen, Hilden, Germany) following the manufacturers instruction. Methylation levels were analyzed by PyroMark Q48 Autoprep software (Qiagen, Hilden, Germany). The result was expressed as percentage of methylated cytosines over the total of methylated and non-methylated cytosines on each CpG site.

Statistical analysis

Statistical analysis was performed using GraphPad Prism 9 and IBM SPSS 26. First, the data's normal distribution was assessed using the Shapiro-Wilk method. The Spearman correlation coefficient (r) was used to analyze the correlation between methylation levels and actual age in non-normal distribution. The statistical significance level was set at $p < 0.05$. To demonstrate the performance of predicted equations for age estimation, the mean absolute error (MAE) and the root mean square error (RMSE) were calculated. The MAE represents the error between predicted values and the actual value by calculating from the average absolute difference between the predicted values and the actual values in a set of data points. The RMSE is standard deviation of the prediction errors, and it is another parameter to evaluate the accuracy of a predictive model. It indicates the extent to which data points deviate from the regression line. The lower MAE and RMSE indicate the closer to the actual values of the model's predictions. It is widely employed for evaluating the performance of predictive models [19]. Leave-One-Out Cross-Validation (LOOCV) was performed in RStudio to evaluate the model's effectiveness and capability to make accurate predictions on unseen data. One data point is excluded at a time from all datasets and train the model on the remaining data. Then, the model is used to predict the excluded data point. This process is repeated for each data point in the whole dataset, enabling a thorough evaluation of the model's effectiveness and its capability to make accurate predictions on unseen data [20].

RESULTS

Thirty-six teeth were collected from corpses without underlying diseases. The quantification of DNA methylation on *ELOVL2*, *ZYG11A*, and *TRIM59* genes were analyzed by pyrosequencing technique. The eight CpG sites on the *ELOVL2* gene (cg16867657) were CpG1 (6:11044861), CpG2 (6:11044864), CpG3 (6:11044867), CpG4 (6:11044873), CpG5 (6:11044875), CpG6 (6:11044877), CpG7 (6:11044880), and CpG8 (6:11044888). The four CpG sites, CpG1 (1:53308756), CpG2 (1:53308758), CpG3 (1:53308760), and CpG4 (1:53308768) were located on the *ZYG11A* gene (cg06784991). Finally, the six CpG sites on the *TRIM59* gene (cg07553761) were CpG1 (3:160167960), CpG2 (3:160167962), CpG3 (3:160167967), CpG4 (3:160167972), CpG5 (3:160167977), and CpG6 (3:160167980).

A correlation between methylated DNA and actual age

Regarding the *ELOVL2* gene, the methylated DNA levels at CpG1, CpG2, CpG3, CpG4, CpG5, CpG6, and CpG7 demonstrated a gradual increase with advancing age. The overall r values for methylated CpG sites and age ranged between −0.0083 and 0.5359. Significant correlations between DNA methylation and age were observed at CpG1, CpG4, CpG5, and CpG6. Particularly, CpG5 exhibited a moderate positive correlation with an r value of 0.5359. Regarding the four CpG sites neighboring the *ZYG11A* genes, it was observed that the methylated DNA levels at CpG1, CpG2, and CpG3 were likely directly proportional to the actual age. Additionally, the methylation level of CpG1 showed a significantly positive correlation

Table 2 The correlation between methylated CpG site of 3 candidate genes and age by Spearman correlation coefficient analysis.

Gene	CpG site	Spearman Correlation		
		Correlation coefficient (r)	95% confidence interval	p-value
ELOVL2	**CpG1**	**0.3732**	**0.0409 – 0.6312**	**0.0125**
	CpG2	0.1332	−0.2139 – 0.4504	0.2194
	CpG3	0.1271	−0.2199 – 0.4455	0.2301
	CpG4	**0.3110**	**−0.0296 – 0.5869**	**0.0324**
	CpG5	**0.5359**	**0.2422 – 0.7396**	**0.0004**
	CpG6	**0.4354**	**0.1148 – 0.6739**	**0.0040**
	CpG7	0.1755	−0.1722 – 0.4843	0.1530
	CpG8	0.0415	−0.3002 – 0.3737	0.4051
ZYG11A	**CpG1**	**0.3765**	**0.0446 – 0.6335**	**0.0118**
	CpG2	0.2615	−0.0833 – 0.5505	0.0617
	CpG3	0.1847	−0.1629 – 0.4916	0.1404
	CpG4	0.1174	−0.2292 – 0.4375	0.2477
TRIM59	CpG1	0.1962	−0.1513 – 0.5006	0.1257
	CpG2	0.2258	−0.1209 – 0.5234	0.0927
	CpG3	−0.0083	−0.3448 – 0.3302	0.4809
	CpG4	0.2712	−0.0730 – 0.5576	0.0548
	CpG5	**0.3317**	**−0.0065 – 0.6018**	**0.0241**
	CpG6	0.1951	−0.1524 – 0.4997	0.1270

A significant correlation between methylated level and age is shown in bold ($p < 0.05$).

with age. Furthermore, CpG5 on the *TRIM59* gene displayed a significant positive correlation between the methylation level and age at that specific CpG site (Table 2).

Validation of age prediction model

The performance of the age prediction model was analyzed using a training set of 36 teeth samples. The stepwise multiple linear regression approach was employed by selecting the nine CpG sites correlated with age. These CpG sites were located at CpG1, CpG4, CpG5, and CpG6 in the *ELOVL2* gene; CpG1 and CpG2 in the *ZYG11A* gene; and CpG2, CpG4 and CpG5 in the *TRIM59* gene. Their selection was based on r values higher than 0.2 determined by Spearman correlation coefficient analysis.

Four models of equations were produced in our research. The age estimation in model 1 which was calculated based on the methylated *ELOVL2* gene. The performance of this model was evaluated by comparing the predicted age with the actual age, resulting in the MAE, the RMSE, the r value, and the r squared (r^2) value. Model 1 showed a moderately strong positive linear relationship between predicted age and actual age. In addition, model 2, which was based on the methylated *ZYG11A* gene, demonstrated lower accuracy and higher predicted error compared with results from the other three models. However, in model 3, using CpG1 and CpG2 on the methylation *ZYG11A* gene, showed a better performance than model 2, which used only CpG1 on *ZYG11A*. Additionally, model 4, which included methylated *ELOVL2* and *ZYG11A*, was regarded as the most accurate age prediction and the lowest predictive error (Fig. 1 and Table 3).

To compare the error for age prediction, the age was grouped into less than 40 years, between 40–60 years, and more than 60 years. It was discovered that for all four models, MAE and RMSE values accurately predicted age in the age range 40–60 years. Ages under 40 and over 60 were accurately predicted by model 4, while model 3 showed the most accurate predictions for individuals in the 40–60 age range (Table 4).

Given the small dataset and lack of a validation set, the performance of the prediction model was evaluated using LOOCV. The LOOCV is a reliable method that utilizes all available data points for training and testing, providing a more robust estimate of the model's performance [20]. In each iteration, one data point was excluded from the training set, resulting in a training set of 35 data points. The model's performance was then assessed on the excluded data point, serving as the test set. By incorporating all individuals in the LOOCV approach, the models yielded MAE of 11.722–14.195 years and RMSE of 13.899–16.419 years (Table 5). These results also demonstrated moderate correlations with chronological ages (r^2 between 0.045–0.307), hence, suggesting that model 4 gave the most accurate age prediction with an error of approximately 11–13 years from the actual age, while effectively explaining the variation in the data.

DISCUSSION

The aim of this study was to preliminary study of a correlation between DNA methylation levels in teeth and actual age, with the goal of developing a pilot

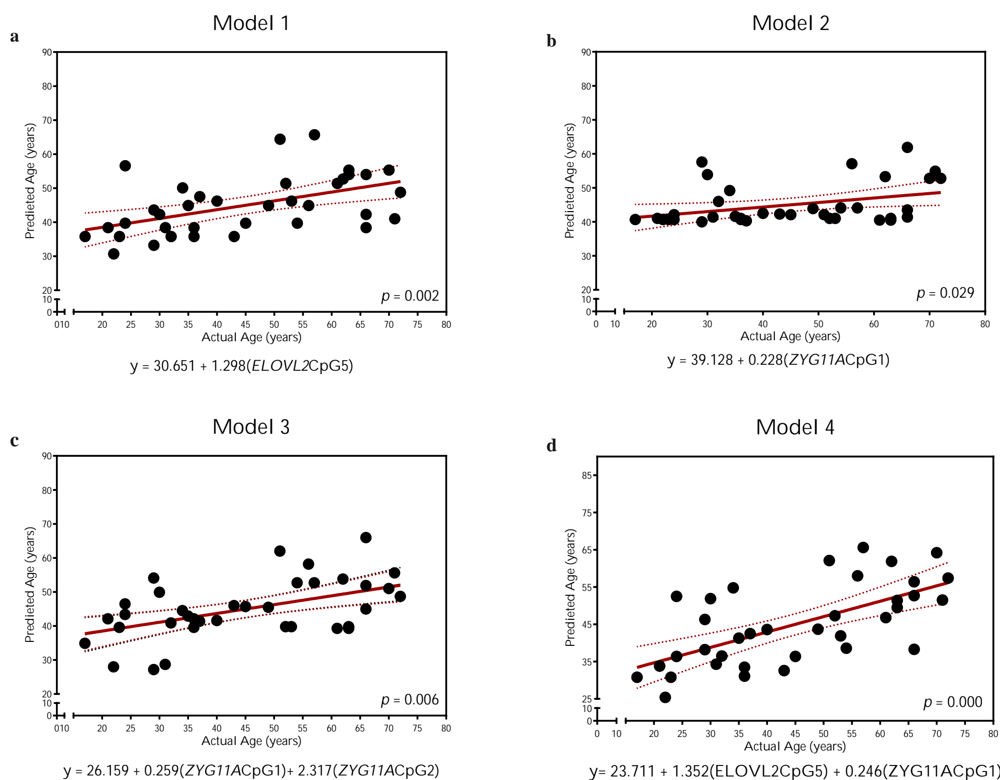

Fig. 1 Predicted age versus actual age in models (a–d) with the training set of 36 teeth samples. The graphs depict the correlation between actual age (years) represented on the x-axis and predicted age (years) represented on the y-axis. A red line shows the best fit line and dot lines show the 95% confidence intervals.

Table 3 Multiple stepwise regression in the four models for age estimation.

Model	Equation	r	r^2	SE	MAE	RMSE
1	y = 30.651 + 1.298(*ELOVL2*CpG5)	0.508	0.259	14.690	11.996	14.272
2	y = 39.128 + 0.228(*ZYG11A*CpG1)	0.365	0.133	15.880	13.395	15.434
3	y = 26.159 + 0.259(*ZYG11A*CpG1) + 2.317(*ZYG11A*CpG2)	0.518	0.269	14.800	11.573	14.174
4	y = 23.711 + 1.352(*ELOVL2*CpG5) + 0.246(*ZYG11A*CpG1)	0.643	0.413	13.265	10.711	12.701

Correlation coefficient (r), coefficient of determination (r^2), standard error of the estimate (SE), mean absolute error (MAE) and root mean square error (RMSE).

Table 4 Mean absolute error (MAE) and root mean square error (RMSE) between an actual and a predicted ages of the four models.[†]

Model	Age group (years)		
	< 40	40–60	> 60
1	11.737 (14.008)	7.750 (8.727)	16.658 (18.526)
2	14.852 (16.547)	6.708 (8.072)	17.752 (18.823)
3	12.122 (14.401)	5.300 (7.013)	16.969 (18.498)
4	10.934 (13.277)	8.084 (8.987)	12.980 (14.761)

[†] Values are MAE (RMSE).

age prediction model that is more reliable, accurate, and precise. The pyrosequencing method, on the other hand, is a popular choice due to its unique detection and quantification platform technology. It is suitable for analyzing fresh, frozen, and FFPE specimens. It is known for being rapid, easy-to-use, and cost-effective, and has been widely applied in this field [21].

A previous study reported that blood and buccal swab samples are commonly used as sources of DNA methylation for accurate and precise age prediction [13]. However, environmental damage and decay susceptibility may render these specimens unavailable in forensic cases. Therefore, dental samples are considered as alternative specimens that provide valuable information for personal identification in forensic purposes. Several studies have developed prediction models based on blood or various liquid samples in

Table 5 Leave-One-Out Cross-Validation (LOOCV) in the four models for age estimation.

Model	Equation	r^2	MAE	RMSE
1	y = 30.651 + 1.298(*ELOVL2*CpG5)	0.181	12.703	15.090
2	y = 39.128 + 0.228(*ZYG11A*CpG1)	0.045	14.195	16.419
3	y = 26.159 + 0.259(*ZYG11A*CpG1) + 2.317(*ZYG11A*CpG2)	0.161	12.558	15.363
4	y = 23.711 + 1.352(*ELOVL2*CpG5) + 0.246(*ZYG11A*CpG1)	0.307	11.722	13.889

Coefficient of determination (r^2), mean absolute error (MAE) and root mean square error (RMSE) for leave-one-out cross-validation (LOOCV).

different populations. However, these models cannot be directly applied to age prediction using teeth due to variations in the level of methylated DNA [22]. Previous studies have revealed that while the same methylation degree pattern was observed across different populations, variations in accuracy and performance of age prediction models were evident [14]. A previous investigations demonstrated that over 90% of age-associated differentially methylated CpG positions (aDMPs) were significantly identified in African Americans, whereas only 5% of aDMPs were shared between the two racial populations (African Americans vs. whites). Additionally, it was observed that only 3% of hypermethylated aDMPs overlapped, while the remaining methylated aDMPs were unique to each racial population (African Americans vs. whites) [23]. In addition, methylation studies have shown distinct patterns of CpG methylation at certain locus in the autosomal DNA between different human populations or races. According to a prior study, 13.7% of autosomal CpGs displayed significantly different levels of DNA methylation between African Americans and Caucasians [24]. Notably, differences were observed in the methylation scores of the *EDARADD* gene between Japanese and Indonesian saliva samples. Furthermore, nationality significantly influenced age estimation based on the methylation levels of both *EDARADD* and *FHL2* [25]. To compare between the Germans and the Japanese, the correlations between DNA methylation levels and age were very similar in both groups. Evidently, however, there were differences between the two groups in DNA methylation at certain CpG sites, with the most noticeable variations observed in *EDARADD* and *PDE4C* [14]. For the age prediction model utilizing *FHL2* and *TRIM59* methylation, the study noted a high r value for age-associated relationships, specifically 0.798 for *FHL2* in the Southern Han Chinese. However, in the Polish, the r value was 0.42. In various East Asian populations, the r values for age-associated relationships with methylated *TRIM59* were 0.67 and 0.87 in Southern Han Chinese and Koreans, respectively [26]. These results illustrate that different populations exhibit distinct methylation statuses. Hence, based on the findings of the previous investigations, it is suggested that a model trained with specific ethnic groups should not be applied to an individual from a non-targeted ethnic group without retraining the model [27]. Therefore, it is crucial to determine DNA methylation on CpG sites in various genes from teeth samples to establish age prediction models tailored to specific populations. In this study, we utilized bisulfite pyrosequencing, a methodological technique, to evaluate DNA methylation levels of three genes (*ELOVL2*, *ZYG11A*, and *TRIM59*) in teeth samples. We found that certain CpG sites within these genes exhibited significant correlations with actual age, consistent with previous studies utilizing blood, saliva, buccal swab, bone, and teeth samples [6, 7, 28]. Multiple biological alterations, including cellular senescence, telomere attrition, and epigenetic response, are involved in the aging process. Dynamic DNA methylation plays a beneficial role in the adaptive response of cells to cellular stress throughout human lifespan, leading to the accumulation of methylated DNA on various genes [29]. It is widely acknowledged that humans are exposed to various environmental factors that can induce or reduce methylation processes at sensitive locations on the chromosome over their lifetime. These factors support the strong correlation between methylation levels and the aging process, providing a novel marker for age estimation [30].

According to previous findings, the *ELOVL2* gene has been extensively studied for age prediction; however, most studies have focused on analyzing methylated *ELOVL2* in blood and saliva [21]. This gene plays a role in regulating the synthesis of polyunsaturated fatty acids (PUFAs), which are involved in various biological functions such as energy production, inflammation activation, and maintenance of cell membrane integrity. The methylation of the *ELOVL2* gene has been well-documented to be associated with the aging process through the control of different biological pathways [31]. Previous reports have demonstrated a decrease in PUFAs concentration as a result of hypermethylation of the *ELOVL2* gene, which is proportional to an increase in human age [6]. In our study, we observed a positive correlation between the methylation of CpG1 sites in the *ZYG11A* gene and actual age. To the best of our knowledge, the analysis of methylated *ZYG11A* levels in teeth has not been previously published. However, two studies have shown a correlation between the methylation level of *ZYG11A* and actual age. A significant correlation between *ZYG11A* and actual age in whole blood, brain, bone, muscle, and

buccal swab samples was reported in one study [13]; and a positive correlation between methylated *ZYG11A* levels and actual age in blood samples was observed in the other [32]. The *ZYG11A* gene acts as a regulator of the cell cycle, driving cellular proliferation. Some evidence suggests that *ZYG11A* expression is regulated by insulin-like growth factor (IGF) signaling, which is associated with healthy aging and longevity. Therefore, it is hypothesized that the suppression of this gene through hypermethylation might contribute to increased age [33]. Furthermore, this gene encodes a protein substrate for culin-2-based E3 ubiquitin ligase, which is related to the ubiquitin system. Dysregulation of the ubiquitin system has been associated with the aging process, as observed in a previous study [34]. Consistent with these findings, it is suggested that the downregulation of the *ZYG11A* gene due to hypermethylation might disrupt cellular function and the ubiquitin process, contributing to an increase in age [35]. Regarding methylated *TRIM59*, we found a Spearman correlation between the methylation of CpG5 sites in this gene. However, the methylation of *TRIM59* did not serve as a contributing factor in the construction of the age prediction model by the analysis of stepwise regression. Many studies have analyzed the correlation between the methylation level of this gene and actual age using various biological samples such as blood, saliva, and buccal swabs. Nonetheless, the quantification of methylated levels in teeth has not been studied yet. *TRIM59* has been proposed to regulate innate immune signaling pathways and induce cellular senescence, contributing to age-related tissue changes [36]. Hence, it was possible that methylated *TRIM59* genes did not significantly affect the ability to estimate age in Thai population teeth.

To develop a preliminary age prediction model, we selected r values higher than 0.2 in CpG sites located on *ELOVL2*, *ZYG11A*, and *TRIM59* and analyzed their association with actual age using multiple linear regression analysis. Among these sites, only two CpG sites from *ELOVL2* and *ZYG11A* were found to have a significant correlation with age and were never studied in the teeth of Thai population. Among our four age prediction models, applicable for ages between 17 and 72 years, model 4 exhibited the best performance characterized by giving lower MAE and RMSE indicating higher accuracy in the model's predictions. Therefore, the model 4 is capable of predicting ages with an error margin of approximately 11 years in either direction from the actual age. Additionally, we employed LOOCV to further evaluate the model's performance. The results showed a MAE of 11.722 years and an RMSE of 13.889 years, validating the reliability and stability of the model. Previous research has suggested that LOOCV analysis helps address overfitting and overestimation of prediction performance [37]. Given the limited sample size of our study, incorporating a validation set alongside the training set and utilizing LOOCV was necessary. LOOCV proves to be a valuable approach for evaluating the model's ability to generalize, accurately measuring its performance, optimizing data utilization, and minimizing bias and overfitting. By utilizing MAE, RMSE, and LOOCV, this study aims to assess the accuracy and stability of the model and make informed decisions regarding model selection and future enhancements.

Several studies, utilizing the same genes in different biological samples such as blood, saliva, and buccal swab, have demonstrated r value above 0.9 and MAE lower than 5 years [22]. It could be assumed that blood, saliva, and buccal swab samples exhibit a stronger correlation with actual age and lower MAE due to their higher cell count, particularly leukocytes, which contain a greater amount of DNA compared with the other tissues [3, 13]. Additionally, factors like post-mortem interval (PMI), lifestyle, cause of death, and environmental conditions have been shown to influence DNA degradation and subsequently impact DNA methylation levels. Comparing bone and teeth samples with fluid biological evidence samples in terms of PMI reveals that the latter tend to display shorter PMI and faster DNA degradation. Conversely, teeth and bone exhibit slower DNA degradation due to their robust nature as the strongest tissues in the human body, providing resistance against deterioration in harsh environments or exposure to decaying contaminants [38]. Regarding age prediction models based on teeth samples, previous studies have shown r values ranging from 0.7 to 0.9 and MAE values ranging from 2.3 to 8.9 years [6, 28, 39]. In our study, the r value and MAE value were approximately 0.643 and 10.711 years, respectively, which closely align with those reported in previous publications. Interestingly, the differences in r value and MAE value compared to previous studies could be attributed to the preparation of the teeth samples. In our study, DNA was extracted from the entire tooth, while in some publications, DNA extraction was performed on specific tooth components such as cementum, dentin, and pulp. Pulp extraction is considered the most accurate method for age determination, although cementum can also provide valuable information. It is possible that our teeth samples, which included enamel and calcium, inhibited the DNA extraction process, resulting in lower DNA quality and quantity for DNA methylation quantification. Therefore, to increase the accuracy of our model, a large-scale sample size and increases number of CpG sites in each gene are probably required. Moreover, the model could be constructed for small ranges of age to increase accuracy for age prediction in different age ranges.

CONCLUSION

Our findings revealed a robust correlation between methylated *ELOVL2* and *ZYG11A* genes in human teeth

and actual age. Our information provided preliminary data to discover the significantly methylated genes, especially *ELOVL2* and *ZYG11A*, for the construction of age prediction model. Based on this association, we developed a preliminary age prediction model utilizing a DNA methylation-based approach. Four models based on the methylated gene were provided with an error margin of predicted age approximately 11 to 13 years from the actual age. These models hold significant potential in providing valuable insights for forensic age estimation purposes.

Acknowledgements: We thank all staff at the Department of Forensic Medicine, Faculty of Medicine, Chiang Mai University for technical support. Thanks to the Faculty of Medicine, Chiang Mai University for providing scientific equipment. This work was supported by the Faculty of Medicine; Chiang Mai University grant no. 118-2565 and was partially supported by Chiang Mai University.

REFERENCES

1. Jung SE, Lim SM, Hong SR, Lee EH, Shin KJ, Lee HY (2019) DNA methylation of the *ELOVL2*, *FHL2*, *KLF14*, *C1orf132/MIR29B2C*, and *TRIM59* genes for age prediction from blood, saliva, and buccal swab samples. *Forensic Sci Int Genet* **38**, 1–8.
2. Correia Dias H, Corte-Real F, Cunha E, Manco L (2022) DNA methylation age estimation from human bone and teeth. *Aust J Forensic Sci* **54**, 163–176.
3. Maulani C, Auerkari EI (2020) Age estimation using DNA methylation technique in forensics: a systematic review. *Egypt J Forensic Sci* **10**, 38.
4. Zubakov D, Liu F, Kokmeijer I, Choi Y, van Meurs JBJ, van IWFJ, Uitterlinden AG, Hofman A, et al (2016) Human age estimation from blood using mRNA, DNA methylation, DNA rearrangement, and telomere length. *Forensic Sci Int Genet* **24**, 33–43.
5. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, Bauerschlag DO, Jöckel K-H, et al (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol* **15**, 24–35.
6. Marquez-Ruiz AB, Gonzalez-Herrera L, Luna JD, Valenzuela A (2020) DNA methylation levels and telomere length in human teeth: usefulness for age estimation. *Int J Legal Med* **134**, 451–459.
7. Zapico SC, Gauthier Q, Antevska A, McCord BR (2021) Identifying methylation patterns in dental pulp aging: Application to age-at-death estimation in forensic anthropology. *Int J Mol Sci* **22**, 3717.
8. Jones MJ, Goodman SJ, Kobor MS (2015) DNA methylation and healthy human aging. *Aging Cell* **14**, 924–932.
9. Slieker RC, Relton CL, Gaunt TR, Slagboom PE, Heijmans BT (2018) Age-related DNA methylation changes are tissue-specific with *ELOVL2* promoter methylation as exception. *Epigenetics Chromatin* **11**, 25–36.
10. Higgins D, Kaidonis J, Austin J, Townsend G, James H, Hughes T (2011) Dentine and cementum as sources of nuclear DNA for use in human identification. *Aust J Forensic Sci* **43**, 287–295.
11. Interpol (2018) *Disaster Victim Identification (DVI)*, Interpol. Available at: https://www.interpol.int/en/How-we-work/Forensics/Disaster-Victim-Identification-DVI.
12. Naue J, Hoefsloot HCJ, Mook ORF, Rijlaarsdam-Hoekstra L, van der Zwalm MCH, Henneman P, Kloosterman AD, Verschure PJ (2017) Chronological age prediction based on DNA methylation: Massive parallel sequencing and random forest regression. *Forensic Sci Int Genet* **31**, 19–28.
13. Naue J, Sanger T, Hoefsloot HCJ, Lutz-Bonengel S, Kloosterman AD, Verschure PJ (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.
14. Becker J, Bohme P, Reckert A, Eickhoff SB, Koop BE, Blum J, Gunduz T, Takayama M, et al (2022) Evidence for differences in DNA methylation between Germans and Japanese. *Int J Legal Med* **136**, 405–413.
15. Daunay A, Baudrin LG, Deleuze JF, How-Kit A (2019) Evaluation of six blood-based age prediction models using DNA methylation analysis by pyrosequencing. *Sci Rep* **9**, 8862.
16. Sukawutthiya P, Sathirapatya T, Vongpaisarnsin K (2021) A minimal number CpGs of *ELOVL2* gene for a chronological age estimation using pyrosequencing. *Forensic Sci Int* **318**, 110631.
17. Kandi V, Vadakedath S (2015) Effect of DNA methylation in various diseases and the probable protective role of nutrition: A mini-review. *Cureus* **7**, e309.
18. Heidegger A, Xavier C, Niederstatter H, de la Puente M, Pospiech E, Pisarek A, Kayser M, Branicki W, et al (2020) Development and optimization of the VISAGE basic prototype tool for forensic age estimation. *Forensic Sci Int Genet* **48**, 102322.
19. Meng L, Lv W, Zhou Y (2020) Relationship between CYP17A1-mediated DNA demethylation and proliferation, invasion and metastasis of Glioma cells. *Crit Rev Eukaryot Gene Expr* **30**, 475–482.
20. Voisin S, Harvey NR, Haupt LM, Griffiths LR, Ashton KJ, Coffey VG, Doering TM, Thompson JM, et al (2020) An epigenetic clock for human skeletal muscle. *J Cachexia Sarcopenia Muscle* **11**, 887–898.
21. Zbiec-Piekarska R, Spolnicka M, Kupiec T, Makowska Z, Spas A, Parys-Proszek A, Kucharczyk K, Ploski R, et al (2015) 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.
22. Bekaert B, Kamalandua A, Zapico SC, Van de Voorde W, Decorte R (2015) Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics* **10**, 922–930.
23. Tajuddin SM, Hernandez DG, Chen BH, Noren Hooten N, Mode NA, Nalls MA, Singleton AB, Ejiogu N, et al (2019) Novel age-associated DNA methylation changes and epigenetic age acceleration in middle-aged African Americans and whites. *Clin Epigenetics* **11**, 119–135.
24. Adkins RM, Krushkal J, Tylavsky FA, Thomas F (2011) Racial differences in gene-specific DNA methylation levels are present at birth. *Birth Defects Res A Clin Mol Teratol* **91**, 728–736.
25. Oka H, Dwi Ariani M, Akazaki T, Miyauchi M, Kitagawa M (2022) Some tips on age estimation using DNA methylation in saliva samples as an index across the Japanese and Indonesian ethnicities. *Leg Med (Tokyo)*

56, 102042.
26. Fan H, Xie Q, Zhang Z, Wang J, Chen X, Qiu P (2021) Chronological age prediction: Developmental evaluation of DNA methylation-based machine learning models. *Front Bioeng Biotechnol* **9**, 819991.
27. Thong Z, Tan JYY, Loo ES, Phua YW, Chan XLS, Syn CK (2021) Artificial neural network, predictor variables and sensitivity threshold for DNA methylation-based age prediction using blood samples. *Sci Rep* **11**, 1744.
28. Kondo M, Aboshi H, Yoshikawa M, Ogata A, Murayama R, Takei M, Aizawa S (2020) A newly developed age estimation method based on CpG methylation of teeth-derived DNA using real-time methylation-specific PCR. *J Oral Sci* **63**, 54–58.
29. Xiao FH, Wang HT, Kong QP (2019) Dynamic DNA methylation during aging: A "Prophet" of age-related outcomes. *Front Genet* **10**, 107.
30. Spólnicka M, Pośpiech E, Adamczyk JG, Freire-Aradas A, Pepłońska B, Zbieć-Piekarska R, Makowska Ż, Pięta A, et al (2018) Modified aging of elite athletes revealed by analysis of epigenetic age markers. *Aging (Albany NY)* **10**, 241–252.
31. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilini D, et al (2012) Methylation of *ELOVL2* gene as a new epigenetic marker of age. *Aging Cell* **11**, 1132–1134.
32. Florath I, Butterbach K, Muller H, Bewerunge-Hudler M, Brenner H (2014) Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet* **23**, 1186–1201.
33. Sin-Chan P, Gosalia N, Gao C, Van Hout CV, Ye B, Marcketta A, Li AH, O'Dushlaine C, et al (2020) Exome-wide association studies in general and long-lived populations identify genetic variants related to human age. *bioRxiv* **76**, 786–795.
34. Oh S, Hong HS, Hwang E, Sim HJ, Lee W, Shin SJ, Mook-Jung I (2005) Amyloid peptide attenuates the proteasome activity in neuronal cells. *Mech Ageing Dev* **126**, 1292–1299.
35. Balachandran RS, Heighington CS, Starostina NG, Anderson JW, Owen DL, Vasudevan S, Kipreos ET (2016) The ubiquitin ligase *CRL2ZYG11* targets cyclin B1 for degradation in a conserved pathway that facilitates mitotic slippage. *J Cell Biol* **215**, 151–166.
36. Hu SH, Zhao MJ, Wang WX, Xu CW, Wang GD (2017) *TRIM59* is a key regulator of growth and migration inrenal cell carcinoma. *Cell Mol Biol* **63**, 68–74.
37. 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.
38. Rhein M, Hagemeier L, Klintschar M, Muschler M, Bleich S, Frieling H (2015) DNA methylation results depend on DNA integrity-role of post mortem interval. *Front Genet* **6**, 182.
39. Giuliani C, Cilli E, Bacalini MG, Pirazzini C, Sazzini M, Gruppioni G, Franceschi C, Garagnani P, et al (2016) Inferring chronological age from DNA methylation patterns of human teeth. *Am J Phys Anthropol* **159**, 585–595.