A newly developed age estimation method based on CpG methylation of teeth-derived DNA using real-time methylation-specific PCR

Masahiro Kondo1), Hirofumi Aboshi1), Masaaki Yoshikawa2), Ayano Ogata1), Ryosuke Murayama1), Masami Takei3), and Shin Aizawa2)

1) Department of Legal Medicine, Nihon University School of Dentistry, Tokyo, Japan
2) Division of Anatomical Science, Department of Functional Morphology, Nihon University School of Medicine, Tokyo, Japan
3) Division of Hematology and Rheumatology, Department of Medicine, Nihon University School of Medicine, Tokyo, Japan

Abstract: Age estimation of unidentified bodies is important in forensic medicine and crime scenes. There is accumulating evidence that DNA methylation in the human genome isolated from body fluids changes with age. Most of the data have been obtained by pyrosequencing. In the forensic field, a simple, quick, and economical method is required to evaluate age of various types of samples. In this study, an age estimation method based on methylation levels of DNA extracted from teeth using real-time methylation-specific PCR (MSP) was developed. The CpG island in the upstream region of ELOVL2, which is known as a validated biomarker in medicine and crime scenes. There is accumulating evidence that DNA methylation can be used to verify age estimation [7,8]. In initial studies, genome-wide DNA methylation profiling, such as Illumina (San Diego, CA, USA) Infinium HumanMethylation450K BeadChip array, was used to identify a large number of CpG sites that are significantly correlated with age [8]. For example, Hannum et al. established a quantitative model of senescence based on the 71 CpG sites selected from the Illumina BeadArray technology (mean absolute deviation [MAD] = 4.9 years) [9]. In the same way, Garagnani et al. discovered that methylation status of 3 regions, the CpG islands of 3 genes including ELOVL2, strongly correlates with age on whole blood-derived DNA [10]. The degree of age-dependent DNA methylation varies depending on the genes, tissues and cell types [11].

On the other hand, Weidner et al. built an aging clock derived from human blood that could predict age using 3 CpG sites located on the ITGA2B, ASPA, and PDE4C gene using bisulfite pyrosequencing [12]. Subsequently, studies with different samples by several groups confirmed that the ELOVL2 can be an excellent biomarker for age estimation at forensic scenes [13-15]. Most of these experiments were performed using bisulfite pyrosequencing. Indeed, the sequencing-based analysis is an excellent method in that it can obtain information on which GC sites on the genome methylation has occurred. It is, however, time-consuming and costly to analyze.

Methylation-specific PCR (MSP), which was initially reported by Herman et al., can quickly assess the methylation status of virtually any group of CpG sites within a CpG island without a sequencing step, and needs only a small amount of DNA [16]. It should be also emphasized that the method can be performed without specific equipment.

As forensic identification of human remains that are severely damaged or degraded, bone and teeth are frequently the only available sources of DNA. Furthermore, teeth are especially important as a sample for victim identification in large-scale disasters. Nevertheless, age estimation models based on DNA methylation level have been developed using body fluid, particularly blood. In this study, a tooth-derived DNA-based age estimation model based on a semi-quantitative analysis of DNA methylation by real-time MSP was developed.

Materials and Methods

This study was approved by the Ethics Committee of Nihon University School of Dentistry (approval #EP19D007). All experiments in this study were performed following the Japanese ethical guidelines for human genome/gene analysis research, Ministry of Health, Labor and Welfare of Japan.

DNA extraction

Twenty-nine teeth (from 20 to 79 years old) that had been stored in the laboratory were used in this study. Each sample had been carefully conserved in a paper bag, a plastic bag, or a small glass bottle. The teeth were washed in 5% sodium hypochlorite solution for a few minutes and then washed with a toothbrush under running distilled water to remove debris.

Genomic DNA was extracted from the teeth using a TBONE EX KIT (DNA Chip Research Inc., Tokyo, Japan) followed by phenol extraction. Briefly, the teeth samples were soaked in 30 mL solution A at 23°C for overnight. Following this, 1.8 mL of solution B was added to the same tube. The mixture was incubated at 37°C for 2 h to decalcify the samples. After discarding the supernatant, 400 µL of solution C and 50 µL of Proteinase
K (20 mg/mL) were added to the samples. The mixture was incubated at 56°C for 3 h for digestion. The supernatant was collected in a new tube containing 400 μL of TE-saturated phenol, mixed and centrifuged. After collecting the aqueous phase, DNA was purified through a silica membrane column and the purified DNA was eluted in 50 μL of elution buffer (QIAamp DNA Mini Kit, Qiagen, Hilden, Germany) and was quantified by using NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA).

Bisulfite conversion and real-time PCR
Two hundred ng of the tooth-derived genomic DNA was denatured by 0.3 M NaOH for 15 min at 37°C and then treated with sodium bisulfite for converting any unmethylated cytosine residues into uracil. The converted DNA was eluted in 25-50 μL of elution buffer according to the protocol (MethylEasy Xceed Rapid DNA Bisulphite Modification Kit, #ME002, Human Genetic Signature, Newtown, Australia).

Subsequently, 2 μL of bisulfite-converted DNA template (25 ng) was amplified by real-time PCR in 25 μL of a qPCR reaction mixture containing 0.3 μM methylation recognition primer targeting the CpG sequences in the ELOVL2 promoter region (or 0.1 μM primer for the human ALU sequence), MSP buffer including 2 mM Mg²⁺ and 200 mM of each dNTP, TB Green solution and MSP enzyme (EpiScope MSP Kit; TaKaRa Bio, Shiga, Japan). The PCR reaction was run on the Thermal Cycler Dice Real Time System II (TP900, TaKaRa Bio) with an initial denaturation at 95°C for 30 s, followed by 45 cycles at 95°C for 5 s, 55°C for 30 s and 72°C for 30 s, and then a final extension at 60°C for 5 min.

Table 1 Primer sequences used for the methylation specific PCR.

<table>
<thead>
<tr>
<th>Target</th>
<th>Primer Names</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELOVL2</td>
<td>ELOVL2-F</td>
<td>TTTACGGAGTTTTAGGAATATTTATTA</td>
</tr>
<tr>
<td></td>
<td>ELOVL2-R</td>
<td>CCACTAATAACAAAAACCAAGTA</td>
</tr>
<tr>
<td>ALU</td>
<td>ALU-F</td>
<td>GGTATAAGTAACTGTTATATTGTTAATTGTA</td>
</tr>
<tr>
<td></td>
<td>ALU-R</td>
<td>ATTAACCTAATACTTTAAACACCTTCAC</td>
</tr>
</tbody>
</table>

Correlation between DNA methylation levels and chronological age
To estimate the methylation levels of tooth-derived DNA obtained from individuals of different ages, a CpG island in the ELOVL2 upstream region was used. A semi-quantitative MSP was done by real-time PCR using two sets of primers specific to DNA molecules methylated at an upstream locus of ELOVL2 gene and to an ALU repeat sequence. The percent methylated reference (PMR) was highly correlated with the actual age (Pearson r = 0.843) (Fig. 2A). The correlation between PMR and chronological age (% 0.843) (Fig. 2A). The correlation between PMR and chronological age was highly correlated with the actual age (Pearson r = 0.843) (Fig. 2A). The correlation between PMR and chronological age was highly correlated with the actual age (Pearson r = 0.843) (Fig. 2A). The correlation between PMR and chronological age was highly correlated with the actual age (Pearson r = 0.843) (Fig. 2A).
showed no significant difference between women and men (Two correlation coefficients test, \(Z = 1.06, \ P = 0.28\), two-tailed) (Fig. 2B, C).

**Change of DNA methylation levels through generations**

To investigate the difference of DNA methylation levels among generations, the samples were divided into five groups and the mean of PMR was calculated by each group (Fig. 3). Normality of the data was confirmed by calculating Shapiro-Wilk test (\(P = 0.0806, 0.0707, 0.4389, 0.9206\) and 0.3019), and homoscedasticity of data was evaluated by Brown-Forsythe test (\(P = 0.321\)). The PMR of each sample was 2.78 ± 2.6\% (\(n = 7\)) in the young group (under 29 years old). In the middle-age groups, the PMRs of group I (between 30 and 39 years of age), group II (between 40 and 49 years of age) and group III (between 50 and 59 years of age) were 3.72 ± 2.9\% (\(n = 7\)), 6.40 ± 4.3\% (\(n = 5\)), and 9.82 ± 4.5\% (\(n = 4\)) respectively. On the other hand, the PMR in the elderly group (over 60 years old) was 15.82 ± 5.4\% (\(n = 6\)). The results indicate that the DNA methylation levels were associated with advancing age showing a gradual increase in standard error of the mean.

**The accuracy of a simple regression model built from 29 training samples**

As a next step, ELOVL2 gene was put in a simple regression model to obtain a formula to predict age. The prediction model exhibited some precision with MAD from the chronological age of 8.94 years among 29 training samples (Fig. 4A). The differences between chronological age and predicted age were plotted against chronological age with dots (Fig. 4B). The differences seemed to increase slightly in an age-dependent manner.

**Table 2** List of teeth donated in this experiment.

<table>
<thead>
<tr>
<th>Age</th>
<th>Sex</th>
<th>Type of tooth</th>
<th>Tooth position</th>
<th>Extracted DNA (ng/μL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>F</td>
<td>Molar</td>
<td>18</td>
<td>34.8</td>
</tr>
<tr>
<td>20</td>
<td>M</td>
<td>Molar</td>
<td>28</td>
<td>60.8</td>
</tr>
<tr>
<td>21</td>
<td>F</td>
<td>Molar</td>
<td>18</td>
<td>141.7</td>
</tr>
<tr>
<td>24</td>
<td>F</td>
<td>Molar</td>
<td>38</td>
<td>62.1</td>
</tr>
<tr>
<td>26</td>
<td>M</td>
<td>Molar</td>
<td>48</td>
<td>114.7</td>
</tr>
<tr>
<td>28</td>
<td>F</td>
<td>Molar</td>
<td>48</td>
<td>29.1</td>
</tr>
<tr>
<td>28</td>
<td>U</td>
<td>Molar</td>
<td>27</td>
<td>58.8</td>
</tr>
<tr>
<td>31</td>
<td>F</td>
<td>Molar</td>
<td>18</td>
<td>30.2</td>
</tr>
<tr>
<td>31</td>
<td>M</td>
<td>Molar</td>
<td>18</td>
<td>28.1</td>
</tr>
<tr>
<td>33</td>
<td>F</td>
<td>Molar</td>
<td>28</td>
<td>140.8</td>
</tr>
<tr>
<td>35</td>
<td>F</td>
<td>Molar</td>
<td>28</td>
<td>51.0</td>
</tr>
<tr>
<td>35</td>
<td>F</td>
<td>Molar</td>
<td>18</td>
<td>172.0</td>
</tr>
<tr>
<td>36</td>
<td>M</td>
<td>Molar</td>
<td>18</td>
<td>40.7</td>
</tr>
<tr>
<td>37</td>
<td>F</td>
<td>Molar</td>
<td>38</td>
<td>46.3</td>
</tr>
<tr>
<td>44</td>
<td>M</td>
<td>Molar</td>
<td>37</td>
<td>108.0</td>
</tr>
<tr>
<td>44</td>
<td>M</td>
<td>Molar</td>
<td>28</td>
<td>66.6</td>
</tr>
<tr>
<td>47</td>
<td>F</td>
<td>Molar</td>
<td>18</td>
<td>22.3</td>
</tr>
<tr>
<td>47</td>
<td>F</td>
<td>Molar</td>
<td>18</td>
<td>23.8</td>
</tr>
<tr>
<td>48</td>
<td>M</td>
<td>Molar</td>
<td>48</td>
<td>109.9</td>
</tr>
<tr>
<td>51</td>
<td>M</td>
<td>Molar</td>
<td>28</td>
<td>101.0</td>
</tr>
<tr>
<td>51</td>
<td>M</td>
<td>Molar</td>
<td>17</td>
<td>184.5</td>
</tr>
<tr>
<td>56</td>
<td>M</td>
<td>Molar</td>
<td>38</td>
<td>178.7</td>
</tr>
<tr>
<td>59</td>
<td>M</td>
<td>Molar</td>
<td>37</td>
<td>167.1</td>
</tr>
<tr>
<td>62</td>
<td>M</td>
<td>Molar</td>
<td>36</td>
<td>41.7</td>
</tr>
<tr>
<td>62</td>
<td>M</td>
<td>Molar</td>
<td>27</td>
<td>57.2</td>
</tr>
<tr>
<td>64</td>
<td>F</td>
<td>Molar</td>
<td>17</td>
<td>204.4</td>
</tr>
<tr>
<td>70</td>
<td>M</td>
<td>Premolar</td>
<td>14</td>
<td>91.9</td>
</tr>
<tr>
<td>78</td>
<td>F</td>
<td>Premolar</td>
<td>24</td>
<td>175.5</td>
</tr>
<tr>
<td>79</td>
<td>F</td>
<td>Molar</td>
<td>37</td>
<td>76.5</td>
</tr>
</tbody>
</table>

F, female; M, male; U, unidentified; Molar, molar teeth; Premolar, premolar teeth.
Discussion

The aim of the study was to develop an age prediction model based on CpG methylation levels of tooth-derived DNA using real-time MSP. Methylation levels of ELOVL2 upstream CpG region were highly correlated with the chronological age. No gender differences were observed in the correlation between CpG methylation levels and chronological age. Moreover, the age prediction model constructed from CpG sites of a single gene showed precision with MAD from the chronological age of 8.94 years among 29 training samples.

Validation of DNA extracted from teeth

Hard tissues including teeth and bones are a good source of DNA for forensic identification of body remains due to their excellent preservation properties. Several groups have reported that teeth are less sensitive to contamination when compared with bone [21,22].

Major sources of DNA from teeth are dentin, dental pulp and cementum [23]. In general, teeth with the largest volume or multi-root teeth provide the best source of DNA [23]. In addition, it is believed that age-dependent degenerative changes in the pulp affect the DNA content of teeth [24,25].

Genomic DNA was extracted from whole teeth by using TBNONE EX KIT for DNA methylation analysis. The range of extracted genomic DNA was between 22.3 and 204 ng/µL (Mean ± SE, 90.4 ± 56.3). The results indicate that the amount of DNA is sufficient to analyze DNA methylation levels using real-time MSP. It was also investigated that whether sufficient DNA content was equally extracted from each sample. DNA content of each tooth was independent of the type of tooth and the age of the donors (Fig. 1B). Moreover, no difference in the purity of extracted DNA was observed in each tooth, demonstrating the validity of experimental results.

An age prediction model based on CpG methylation levels in the promoter region of ELOVL2

As an age prediction biomarker, DNA methylation has recently been focused on by many researchers including forensic scientists. According to Horvath’s report, the average difference in methylation rates for all analyzed loci between young individuals and old individuals is only 3% per CpG [11]. In contrast, Garagnani et al. reported that the average difference of a CpG in the ELOVL2 promoter region was in a range between 7% and 91% methylation throughout the lifetime [10]. These reports indicate that ELOVL2 is an attractive candidate for age predictors.

Zbiec-Piekarska et al. examined methylation levels in 7 GpG sites (C1-C7) located in the upstream of ELOVL2 gene in the blood-derived genomes using pyrosequencing [13]. They found that methylation levels of C5 and C7 showed a strong correlation with chronological age. Based on the findings, they developed an age estimation model derived from DNA methylation levels in only those two GpG sites (R² = 0.859, MAD = 5.03). Bekkaert et al. also analyzed the methylation status of candidate genes including ELOVL2 and developed a multivariate quadratic regression model in the same way [14]. Interestingly, only C1, C6 and C9, but not C5 and C7, of ELOVL2 were correlated as contributing to their model, with C6 being by far the most informative CpG for age estimation. The age prediction by their multivariate quadratic regression model (derived from 4 genes, ADPA, PDE4C, EDARADD, and ELOVL2) showed high accuracy of a relationship between chronological age and predicted age (r = 0.97, MAD = 3.75 for blood sample) [14]. They further validated the model with 29 teeth-derived DNA as test samples (R² = 0.74, MAD = 4.86) [14]. Meanwhile, Giulian et al. attempted to predict chronological age from DNA methylation in ELOVL2, FHL2, and PENK promoter regions of 21 modern teeth samples using Epityper assay [23]. In this case, they selected the 8 CpGs units including 5 CpG sites in ELOVL2 to develop multiple linear regression model.

Based on this previous research, the CpG sites located in the upstream of the ELOVL2 coding region were used for developing a single regression model. Five pairs of methylation-specific primers were designed and the best primer set among them was used for analysis (Table 1; also see Materials and Methods). The methylation levels of the region were robustly correlated with chronological age of teeth donors (r = 0.843, Fig. 2A). Moreover, the prediction model showed accuracy with MAD from chronological age of 8.94 among 29 training samples (Fig. 4A).

Bekaert et al. reported that DNA methylation values in the 3 CpGs from ELOVL2 show quadratic relationship with age rather than a straight linear or logarithmic relationship [14]. To verify this issue, determination coefficients were calculated by 3 different methods. Each determination coefficient was Quadratic R² = 0.51389; Linear R² = 0.71042; Logarithmic R² = 0.62193, indicating that the linear regression is better in this model (Fig. 1A). There are some possible reasons for this discrepancy as follows: 1) The sample organs used to build the regression model were different (blood vs. teeth), 2) The relatively small number of samples used to build the regression model (206 vs. 29 samples).

In spite of the relatively low accuracy compared to the other two models, the present model showed a correlation between chronological age and DNA methylation levels in ELOVL2, indicating that real-time MSP based-analysis can be employed for age estimation in the forensic field.

In addition to age, many other factors, such as sex, physiological properties, and diseases, are known to affect DNA methylation levels on some genes [26,27]. In the case of the ELOVL2 gene, it has, however, been reported that there is no gender difference in age-associated changes of DNA methylation levels [14,28]. The result was consistent with that of previous reports.

Real-time methylation specific PCR is a fast, simple method to quantify methylated-DNA levels

Pyrosequencing analysis has the advantage of providing information about loci where methylation occurs in the genome. Nevertheless, the method is still not routinely applied for actual criminal investigations due to the high cost and time requirements.

On the other hand, real-time MSP, which is a simple and convenient method, is classified into PCR-based methylation analysis. The method has often been applied for the diagnosis of inherited diseases and various types of cancer [29,30]. It is possible to detect the DNA methylation levels of a specific genomic region by the presence or absence of PCR amplification [16]. The method is suitable for analyzing a genomic region with high CpG
density and needs only a small amount of genomic DNA.

Hamano et al. have recently reported a novel age prediction model using methylation-sensitive high-resolution melting (MS-HRM) [31,32]. The method is also classified in PCR-based analysis as well as real-time MSP. The method can measure methylation status on bisulfite-treated DNA that is PCR amplified followed by melting analysis [33]. Using this technique, they established an age estimation model and proved the high accuracy of the model ($R^2 = 0.83, \text{MAD} = 7.44$) [32]. A simple comparison between Hamano’s results and the results is difficult, because the combination of genes used to create each age estimation model and the organ to be analyzed is different. MS-HRM is also an excellent tool that is available in forensic sciences as it is an easy and accurate method.

**Future plans**

In this study, an age estimation model from tooth-derived DNA using real-time MSP was developed. The simple regression model showed accuracy to age estimation based on methylation level in tooth-derived DNA. So far, a large number of age prediction models based on the DNA methylation in blood have been developed by different laboratories. Many of the models are based on at least 3 different loci of human genome [26,34-37]. It has been reported that the DNA methylation changes are tissue-specific [15]. Therefore, the age estimation model developed using blood DNA should not be directly applied to teeth samples. It is necessary to explore the optimal CpG sites in some genes for tooth samples to facilitate further refinement of the current model.

**Acknowledgments**

The authors would like to thank Dr. Koichi Iwata for his valuable suggestions and comments. The authors also sincerely thank Dr. Nobuyuki Takizawa and a member of Shunpo-kai for encouraging us to achieve our research goals.

**Conflict of interest**

The authors have no conflicts of interest to declare.

**References**