Forensic Science International: Genetics 23 (2016) 64–70

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Research paper

Identification and evaluation of age-correlated DNA methylation markers for forensic use

Jong-Lyul Park[a], Jong Hwan Kim[a,b], Eunhye Seo[a,b], Dong Hyuck Bae[a,b], Seon-Young Kim[a,b], Han-Chul Lee[c], Kwang-Man Woo[c], Yong Sung Kim[a,b,*]

[a] Epigenomics Research Center, Genome Institute, Korea Research Institute of Bioscience and Biotechnology, 125 Gwahak-ro, Yuseong-gu, Daejeon, Republic of Korea
[b] Department of Functional Genomics, University of Science of Technology, 125 Gwahak-ro, Yuseong-gu, Daejeon, Republic of Korea
[c] DNA Forensic Division, Supreme Prosecutor's Office, 157 Banpodae-ro, Seocho-gu, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 15 September 2015
Received in revised form 11 March 2016
Accepted 16 March 2016
Available online 17 March 2016

Keywords:
DNA methylation
Forensic
Age prediction

ABSTRACT

In forensics, age prediction is useful to narrow down the number of potential suspects because it can provide some general characteristics for predicting appearance. Previous genome-wide studies based on DNA methylation have reported age prediction algorithms using a penalized multivariate regression method known as elastic net and a few dozen to hundreds of CpG sites. Although more CpG sites may provide better accuracy than fewer CpG sites, this approach is not applicable to forensics because the amounts of crime-scene DNA are usually limited. In this study, we selected three age-correlated CpG sites, namely cg16867657 (*ELOVL2*), which is known to be an excellent age predictor, cg04208403 (*ZNF423*), and cg19283806 (*CCDC102B*), from HumanMethylation450 BeadChip datasets of 1415 individuals. Furthermore, we evaluated these markers in a 535-sample training set and a 230-sample validation set from Korean individuals using a pyrosequencing platform. From the training set, an age prediction model using the multiple linear regression method explained 91.44% of age-correlated variation in DNA methylation patterns. The standard error of estimate and mean absolute deviation were 6.320 and 3.156 years, respectively. In the validation set, the standard error of estimate and mean absolute deviation were estimated as 6.853 and 3.346 years, respectively. For the validation set, the model explained 91.08% of the variation in methylation and predicted age (±6 years) with accuracy of 77.30% in the <60 years age group and 57.30% in the older group (≥60 years). These results suggest that our three DNA methylation markers may be useful for age prediction in samples from Asian populations.

1. Introduction

Age prediction is an important part of forensic science because it is useful for identifying suspects. Traditionally, individual age has been predicted by analysis of osteal markers in bones and teeth [1–3], but this approach is confined to cases where a skeleton is present. Several studies have shown that molecular biomarkers, such as a 4977-bp deletion in mitochondrial DNA [4,5], the proportion of d-aspartic acid in proteins [6], the amount of advanced glycation end products [7,8], and the length of leukocyte telomeres [9], can be used to predict age, but there is substantial variation in the age predicted by these various biomarkers [10].

DNA methylation, which occurs at the 5′-position of cytosine in CpG dinucleotides, is a genetically programmed type of DNA modification in mammals [11,12]. Since the demonstration of age-correlated patterns of DNA methylation [13–16], analysis of DNA methylation has emerged as a reliable method for predicting age in forensics [17–20]. In non-forensic fields, several studies have reported age prediction algorithms based on the penalized multivariate regression method known as elastic net using a few dozen to hundreds of CpG sites from analysis of genome-wide methylation datasets, and these algorithms have been validated with various independent datasets [21,22]. This approach provides good accuracy but is limited for forensic use because it is based on HumanMethylation450 BeadChip technology, which requires relatively large amounts of DNA and complex bioinformatics analysis. Thus, an age-prediction model that uses fewer CpG sites would be more applicable to forensics if it can provide accuracy comparable to

* Corresponding author at: KRIBB, Epigenomics Research Center, Genome Institute, Korea Research Institute of Bioscience and Biotechnology, 125 Gwahak-ro, Yuseong-gu Daejeon 305-806, Republic of Korea. Tel.: +82 42 879 8110; fax: +82 42 879 8119.
E-mail address: yongsung@kribb.re.kr (Y.S. Kim).

http://dx.doi.org/10.1016/j.fsigen.2016.03.005
1872-4973/

genome-wide methylation profiling. Recently, Weidner et al. used the HumanMethylation450 BeadChip to identify age-correlated DNA methylation markers in blood. Pyrosequencing analysis based on the combination of three of these markers yielded a mean absolute deviation (MAD) of 5 years [23].

In the forensics field, the promoter of *ELOVL2* is considered the most promising locus for age prediction [16,24,25], but this marker needs to be enhanced by additional markers owing to low accuracy [18]. Yi et al. reported an age prediction model for blood samples based on the leave-one-out multiple linear regression method that shows an average 4-year difference between predicted and true age [19]. In addition, Zbiec-Piekarska et al. reported that the combination of five DNA methylation markers (*ELOVL2*, *C1orf132*, *TRIM59*, *KLF14*, and *FHL2*) predicts age with a MAD of 3.9 years using a multiple linear regression model based on simultaneous analysis of all of the tested CpG sites in blood [26]. Zbiec-Piekarska et al. also introduced a freely available online calculator for age prediction from blood [26]. Although these studies showed reasonable accuracy, additional studies may be helpful because of population- [27,28] and tissue-specific DNA methylation characteristics [15,29].

Several techniques are available for measuring DNA methylation levels [30], but most of these techniques require large amounts of DNA or a control assay. In contrast, pyrosequencing can be performed with as little as 10 and 2.5 ng of pre-modified and modified starting DNA, respectively [18,31] and is relatively easy to apply in a forensics laboratory. In addition, this technique seems to be the most reliable method to assess DNA methylation levels [32]. To date, over 1000 HumanMethylation450 BeadChip datasets have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO). Pyrosequencing evaluation of novel and known age-correlated DNA methylation markers in large samples from the datasets may be helpful for improving the accuracy of age prediction in forensic science.

In this study, we identified known and novel age-correlated CpG sites from 1415 public datasets based on the HumanMethylation450 BeadChip. We selected the three most informative CpG sites and used these to evaluate 765 samples from Korean individuals using pyrosequencing. We developed an age prediction model from 535 of the 765 samples using multiple linear regression models. Finally, we evaluated the accuracy of age prediction of our model with a validation set of the remaining 230 samples.

2. Materials and methods

2.1. Public data collection

To identify age-correlated DNA methylation markers, we obtained DNA methylation datasets from the GEO (GSE32148, 36064, 40005, 40279, 41169, 51032, 53128, and 53740) generated from 2173 blood samples using the HumanMethylation450 BeadChip platform. Some of the GEO datasets contained ethnicity information: GSE36064 (Caucasian, Chinese, and African American), GSE40279 (Caucasian European), GSE51032 (Italy cohort), GSE41169 (Dutch population), and GSE53740 (White). Acute myeloid leukemia (AML) data obtained with the same platform were also obtained from The Cancer Genome Atlas (TCGA) database. TCGA datasets provide ethnicity information (White, Asian, Black American, and African American).

2.2. Methylation data processing

DNA methylation values of the downloaded HumanMethylation450 BeadChip dataset were calculated as average-β values using raw signal or idat files. Average beta (methylation values) is calculated that average-β = M/(M + U + 100), where M and U are methylated and unmethylated signal intensities, respectively. Measurements for which p-value is less than 0.05 are considered to have signal intensity significantly above background.

2.3. Sample collection and DNA isolation

Blood samples were collected from 765 healthy volunteers in a Korean cohort after obtaining informed consent from the participants. The cohort comprised 380 females and 385 males, and the gender distribution was similar in each of our 10-year age groups ($p = 0.9999$); 11–20, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80 and 81–90 years) (Supplementary Table S1). The study was approved by the Institutional Review Board of the Korea Biobank Network (http://koreabiobank.re.kr). Blood samples were drawn from a peripheral vein using an EDTA-containing Vacutainer (BD Biosciences, Franklin Lakes, NJ), and a 100-μl aliquot was stored at −70 °C until analysis. Genomic DNA was isolated from the 100 μl of blood using the DNeasy Blood and Tissue kit (Qiagen, Carlsbad, CA). The quality and quantity of the extracted genomic DNA were assessed with an ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE).

2.4. Pyrosequencing

Pyrosequencing was used to evaluate the age-correlated candidate markers. Briefly, 500 ng of total DNA from each of the 765 blood samples was used for bisulfite conversion using the EZ DNA Methylation Gold kit (Zymo Research, Orange, CA). Each sample was eluted with 20 μl elution buffer from the kit. Next, 1 μl of the bisulfite-converted DNA was used in a 20-μl PCR mixture containing primer sets and 2× Master Mix (Doctor Protein, Seoul, Korea) and amplified using a GeneAmp PCR system 9700 (Applied Biosystems, Waltham, MA). For pyrosequencing, forward, reverse, and sequencing primers were designed with PSQ Assay Design v1.0.6 (Biotage, Kungsgatan, Sweden). Standard pyrosequencing was then performed. Briefly, 20 μl of PCR product was immobilized on 3 μl of Streptavidin Sepharose High Performance (GE Healthcare Bio-Sciences, Uppsala, Sweden) and annealed with sequencing primer for 10 min at 80 °C. Finally, the generated pyrograms were analyzed using PyroMark analysis software (Biotage). Sequences for primer sets (Bioneer, Daejeon, Korea) are shown in Supplementary Table S2, and all PCR reactions were performed with an annealing temperature of 60 °C and 35 cycles. Primers specific to cg16867657 for pyrosequencing have been previously described [18].

2.5. Statistical analysis

We evaluated the gender distribution in each of our defined age groups in our 765 Korean samples using the χ^2 test. We used Spearman's correlation to infer the association between DNA methylation and aging in the group of 1415 subjects ranging in age from 1 to 101 years. The statistical significance was evaluated using the t-test. Simple and multiple linear regression models were used to assess the accuracy of the age prediction using 25 selected marker candidates based on the following formula:

$$C_n^m = \frac{n!}{m!(n-m)!}$$

Where $n!$ is the number of total markers and $m!$ is the selected marker size. For the age prediction model, we used 535 samples from Korean individuals, and the final age prediction model was evaluated using an independent set of 230 samples from Korean subjects by calculating MAD and prediction error. Prediction error is the average absolute difference between predicted and actual

age. Samples in both sets were selected to equally cover the age range from 11 to 88 years and eight 10-year age categories spanning age 10–88 using the caTools package (version 1.17.1). All statistical analyses were performed using R software (version 2.6.1).

3. Results

3.1. Identification of age-correlated, differentially methylated CpGs (DMCs)

We excluded 714 of the 2173 blood samples in the DNA methylation dataset because they were associated with disease and therefore may cause non-specific skewing of age prediction [23,33]. We next cross-checked gender information and methylation status in promoter regions of the X-linked genes *G6PD*, *ENFB1*, *ELK1*, and *GPC3* because it is well known that methylation levels of X-linked genes correlate with gender [34–36]. We found that 15 samples showed discordance between gender and methylation of X-linked genes (Supplementary Fig. S1), and these were excluded along with 29 samples for which there was no gender information. The final dataset of 1415 samples comprised 570 male and 845 female subjects ranging in age from 1 to 101 years. There was a relatively higher abundance of >5 year age samples in the male group (10.87%) compared with the female group (0.11%). Supplementary Fig. S2 lists the age distributions of male and female subjects in the final sample set.

To identify age-correlated DMCs, we calculated Spearman's correlation between age and DNA methylation level based on the following two criteria: $r > 0.5$ or $r < -0.5$, and $p < 0.00005$. We identified 582 age-correlated hypo- or hypermethylated CpG sites (Supplementary Table S3). We performed heatmap analysis with these 582 age-correlated DMC candidates by age group (Fig. 1).

We performed gene ontology analysis using the DMCs linked to 5′-regulatory regions (http://david.abcc.ncifcrf.gov/) to predict the functions of DMC-linked genes. We found that the age-correlated DMCs tended to be associated with genes involved in development and morphogenesis (Supplementary Table S4), similar to previous studies [13,14].

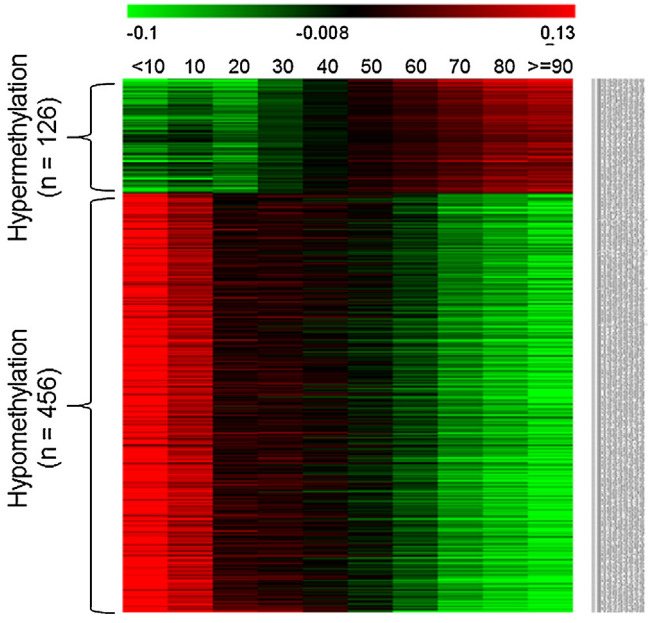

Fig. 1. Heatmap of age-correlated hypo- and hypermethylated CpG sites. Data were from DNA methylation analysis of 1415 blood samples using the HumanMethylation450 BeadChip platform. Age-correlated DMCs (582) were selected based on the Spearman's correlation ($r > 0.5$ or $r < -0.5$). Age was categorized in 10-year intervals. Each of the 582 DMCs was mean centered based on the entire set of 1415 blood samples (subtracting mean average β of each CpG site from its average β value in each sample). The red and green colors indicate high and low methylation levels, respectively.

3.2. Selection of DNA methylation markers for age prediction

The distribution of age-correlated DMCs differs slightly between males and females [19,20]. For example, Hannum et al. reported that the methylome of men appears to change with age ~4% faster than that of women [22]. To select common DNA methylation markers that are not influenced by gender, we compared the top 100 age-correlated DMCs from the male and female groups sorted by Spearman's correlation analysis, which revealed an overlap of 42 DMCs (Fig. 2A). Then, we calculated the average methylation level at these overlapping DMCs for each age group and categorized them as hypo- or hypermethylated. Although age-correlated hypomethylation of CpG sites tended to occur more often in males than females, the difference was not significant (Fig. 2B, $P = 0.8651$). We then narrowed the 42 overlapping DMCs down to 25 using stricter criteria ($r > 0.65$ or $r < -0.65$; Table 1).

To select an appropriate number of these markers for age prediction in forensics, we examined MAD using all possible combination of 1–25 candidate markers. With a single marker, the best MAD was estimated to be 4.14 years. As two and three markers were combined, the best MAD sharply decreased to 3.43 and 3.16 years, respectively. As more than three markers were combined, MAD gradually decreased (e.g., MAD values for the combination of four and five markers were 2.92 and 2.87 years, respectively) (Fig. 3), and therefore we considered the combination of three markers to be the most appropriate for validation with the 765 Korean samples. Supplementary Table S5 lists the top 10 markers in all combinations of one to five markers. The combination of cg16867657 (*ELOVL2*, ELOVL fatty acid elongase 2), cg08097417 (*KLF14*, Kruppel-like factor 14), and cg19283806 (*CCDC102B*, coiled-coil domain containing protein 102B) showed the lowest MAD in all combinations of three markers (Supplementary Table S5), but it was difficult to design primers with binding only to CpG-free sequences for pyrosequencing in the vicinity of cg08097417 (*KLF14*). We also failed pyrosequencing condition for cg14361627 (*KLF14*), cg06639320 (*FHL2*), cg22454769 (*FHL2*), and cg10501210, although pyrosequencing conditions for cg14361627, cg06639320, and cg22454769 have recently been reported [26]. Therefore, we selected the eighth best combination included cg16867657 (*ELOVL2*), cg04208403 (*ZNF423*, zinc finger protein 423) and cg19283806 (*CCDC102B*) as the best targets for practices of pyrosequencing. The combination of these three markers improved the accuracy of age prediction compared with each single marker alone (Supplementary Fig. S3).

3.3. Development of the age prediction model

To further evaluate the combination of these three DMCs, we performed pyrosequencing on blood DNA samples from 765 individuals from a large cohort of the Korean population. The numbers of members in each age group were selected to be evenly distributed (Supplementary Fig. S4). Whereas the HumanMethylation450 BeadChip targets only one CpG site, pyrosequencing allows analysis of CpG loci neighboring the target CpG site. Thus, we additionally examined the methylation status at six CpG sites neighboring cg16867657, two neighboring cg04208403, and one neighboring cg19283806 (summarized in Supplementary Table S6). We then estimated the correlation between DNA methylation at each CpG site and age (Table 2). We used the first sites in *ELOVL2*, *ZNF423*, and *CCDC102B* that showed

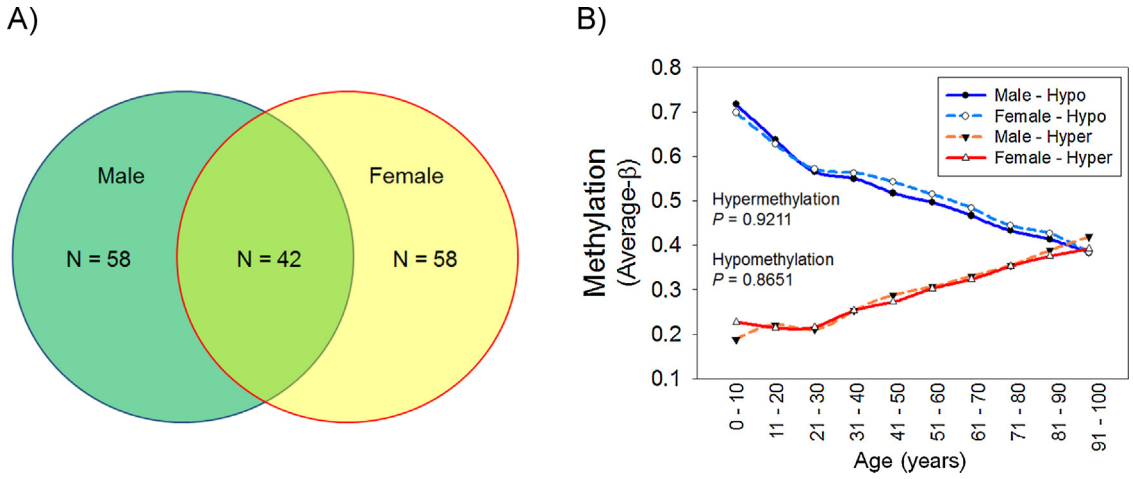

Fig. 2. Age-correlated DNA methylation candidates stratified by gender. A) Numbers and overlap of the top 100 DMCs in the male and female groups ranked by Spearman's correlation. B) The 42 overlapping DMCs in the two groups were further divided into age-correlated hypo- (17) and hypermethylated (25) CpG sites. P-values were calculated using the average methylation value for the male or female groups.

Table 1
Information for the 25 age-correlated DNA methylation markers selected for this study.

Illumina ID[a]	Correlation (r)	Chromosome	MapInfo[b]	Symbol	CpG island[c]	Gene feature group	References
cg16867657	0.830	6	11044877	*ELOVL2*	Island	TSS1500	[22,26]
cg06639320	0.752	2	106015739	*FHL2*	Island	5′UTR;TSS200	[22,26]
cg21572722	0.791	6	11044894	*ELOVL2*	Island	TSS1500	[22,26]
cg22454769	0.739	2	106015767	*FHL2*	Island	5′UTR;TSS200	[22,26]
cg24079702	0.737	2	106015771	*FHL2*	Island	5′UTR;TSS200	[22,26]
cg04875128	0.700	15	31775895	*OTUD7A*	Island	Body	[22]
cg07553761	0.721	3	160167977	*TRIM59*	Island	TSS1500	[26]
cg23500537	0.730	5	140419819				[22]
cg24724428	0.668	6	11044888	*ELOVL2*	Island	TSS1500	[22,26]
cg14361627	0.714	7	130419116	*KLF14*	Island	TSS1500	[22,26]
cg22736354	0.679	6	18122719	*NHLRC1*	Island	1stExon	[21–23]
cg05991454	0.665	4	147558435		Island		
cg08097417	0.697	7	130419133	*KLF14*	Island	TSS1500	[22]
cg17110586	0.693	19	36454623		S_Shelf		
cg02650266	0.661	4	147558239		Island		[22]
cg06493994	0.653	6	25652602	*SCGN*	Island	1stExon;5′UTR	[21–23]
cg11970349	0.651	4	8582287	*GPR78*	Island	TSS200	
cg09124496	−0.663	7	41735851	*LOC285954*		Body	
cg15121420	−0.655	2	238490819	*RAB17*		Body	
cg04208403	−0.664	16	49525807	*ZNF423*	N_Shore	Body	
cg19344626	−0.651	19	16830749	*NWD1*		TSS200	
cg02046143	−0.656	11	133797911	*IGSF9B*	N_Shelf	Body	[22]
cg23744638	−0.679	11	10323902		N_Shore		
cg19283806	−0.723	18	66389420	*CCDC102B*		5′UTR	[22,26]
cg10501210	−0.738	1	207997020				[22,26]

[a] Illumina ID is the unique identification number in the HumanMethylation450 BeadChip.
[b] Mapinfo indicates the genomic location in human reference genome 37 (GRCh37/hg19), released by the Genome Reference Consortium in March 3, 2009 (http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/).
[c] Shore and shelf are adjacent to the CpG island (2- and 4-kb regions flanking the CpG island, respectively). N and S mean upstream and downstream of the CpG island, respectively.

the strongest correlation with age (CpG_1) in our 765 Korean samples to develop the age prediction model (Supplementary Fig. S5).

To develop the age prediction model, we divided the samples into a 535-sample training set and a 230-sample validation set with equal representation of the various age groups within the sets (Fig. 4A and B). Age was estimated as = 39.73167 + *ZNF423* (CpG_01) × −0.28914 + *ELOVL2* (CpG_01) × 1.19242 + *CCDC102B* (CpG_01) × −0.69994 using the multiple linear regression model. The model explained 91.44% of age-correlated variation in DNA methylation patterns with a standard error of estimate of 6.320 years. MAD was calculated to be 3.156 years for the training set of 535 samples (Fig. 4C).

3.4. Accuracy of the age prediction model in the validation set

We evaluated the model performance with the validation set using the prediction formula obtained from the training set, and we observed that the standard error of estimate and MAD were 6.853 and 3.346, respectively (Fig. 5). Next, we calculated the total number of correct versus incorrect results, with "correct" meaning that the predicted age was within ±4, 6, 8, or 10 years of the actual age. Our results showed that prediction accuracy decreased with the elder age group (≥60 years) compared with the <60 years age group (Table 3). This result corresponded well with a previous report [26]. The model had high accuracy, predicting the correct age (±6 years) for 77.30% and 57.30% of the samples in the <60 and

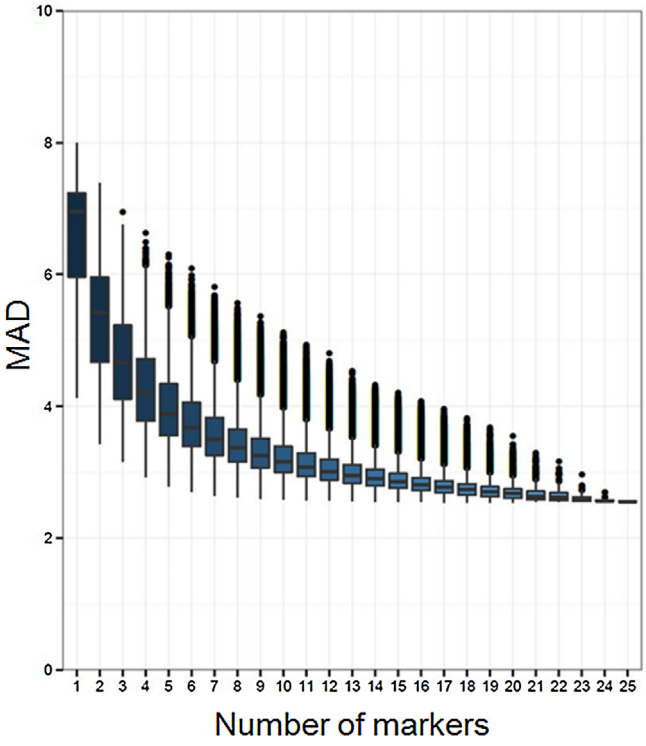

Fig. 3. Accuracy of multiple markers for age prediction. We calculated MAD based on simple and multiple linear regression models using all possible combinations of 1–25 markers for the downloaded HumanMethylation450 BeadChip datasets. The box plots show the median, 25th and 75th percentiles, and outliers.

≥60 age groups, respectively, for the 230-sample validation set (Table 3). Therefore, the use of DNA methylation status using a combination of three selected CpG sites, namely *ELOVL2*, *ZNF423*, and *CCDC102B*, represents a means of accurately predicting age for the purpose of crime scene reconstruction.

4. Discussion

In forensics, age prediction is useful to narrow down the number of potential suspects because it can provide some general characteristics for predicting appearance. However, some individuals are more affected by age-related changes, such as those experiencing premature graying of the hair or male pattern baldness [26]. In this light, age prediction based on molecular markers is more informative. Previously reported age-prediction markers, such as a 4977-bp deletion in mitochondrial DNA, the proportion of d-aspartic acid in proteins, and the size of leukocytes, have the limitations of low sensitivity or lack of stability of the marker in degraded biological samples taken from crime scenes [10]. Therefore, DNA methylation seems to be more suitable for age prediction because of its stability [37].

In this study, we chose to base our analysis on blood samples because blood is commonly found at crime scenes. We evaluated the accuracy of age prediction with increasing numbers of CpG site-methylation markers, and we found that the combination of three markers provided the best age prediction. Indeed, the age of 77.30% and 57.30% of the samples in the <60 and ≥60 age groups of the validation sample set were correctly predicted to within ±6 years using our model. Our age prediction model includes *ZNF423* and *CCDC102B*, which become hypomethylated with age, and *ELOVL2*, which is hypermethylated with age. *ZNF423* is located on 16q12 and encodes a zinc-finger transcription factor involved in brain and olfactory development [38]. It is associated with the RARα/RXRα nuclear receptor complex and is essential for transactivation in response to retinoids [39]. *CCDC102B* on 18q22.1 encodes a 297-amino acid protein of unknown function [40]. However, recent studies have shown that methylation in the *CCDC102B* promoter region is strongly associated with age [22,26]. The protein encoded by *ELOVL2* on 6q24.2 is involved in the elongation of 22-carbon polyunsaturated fatty acids to 24-carbon precursors for docosahexaenoic acid and docosapentaenoic acid formation [41]. Previous studies demonstrated that *ELOVL2* is the most promising age predictor [14,19,22,26]. Thus, our results suggest that aging may regulate methylation in these genes and that they are good candidates for age prediction.

DNA methylation is deregulated in various diseases. Wagner et al. evaluated DNA methylation markers for age prediction in patients with AML and found that methylation occurs prematurely in these patients [33]. Another study also reported that age-correlated DNA methylation is accelerated in patients with certain types of tumors [22]. We evaluated the correlation between methylation levels of the three CpG markers selected in this study and age of the AML samples from TCGA. We observed that correlation of methylation with age at the three CpG sites disappeared in the AML patients (Supplementary Fig. S6), suggesting that there are substantial challenges for age prediction in samples associated with diseases such as AML.

There is another challenge for DNA methylation to become a routine forensic analysis technique for age prediction. Recent studies have shown that patterns of DNA methylation are tissue specific [15,29]. Although Hannum et al. proposed that *ELOVL2* may be a stable age predictor in various tissues [22], Koch and Wagner suggested that universal markers showing the same age prediction accuracy in various tissues might be difficult to define [42].

Table 2
Correlation between three DNA methylation markers and age in samples from 765 Korean individuals.

Symbol	Chromosome location (GRCh37) [a]	CpG identification number	Illumina ID[b]	Correlation (r)
ZNF423	chr16:49525807	CpG_01	cg04208403	−0.7769
	chr16:49525805	CpG_02		−0.7722
	chr16:49525794	CpG_03		−0.7495
ELOVL2	chr6:11044894	CpG_01	cg21572772	0.9352
	chr6:11044888	CpG_02	cg24724428	0.9168
	chr6:11044880	CpG_03		0.8676
	chr6:11044877	CpG_04	cg16867657	0.8942
	chr6:11044875	CpG_05		0.8803
	chr6:11044873	CpG_06		0.9237
	chr6:11044867	CpG_07		0.7939
CCDC102B	chr18:66389420	CpG_01	cg19283806	−0.9061
	chr18:66389447	CpG_02		−0.8986

[a] GRCh37 indicates human reference genome 37 released from Genome Reference Consortium in March 3, 2009.
[b] Illumina ID is the unique identification number in the HumanMethylation450 BeadChip.

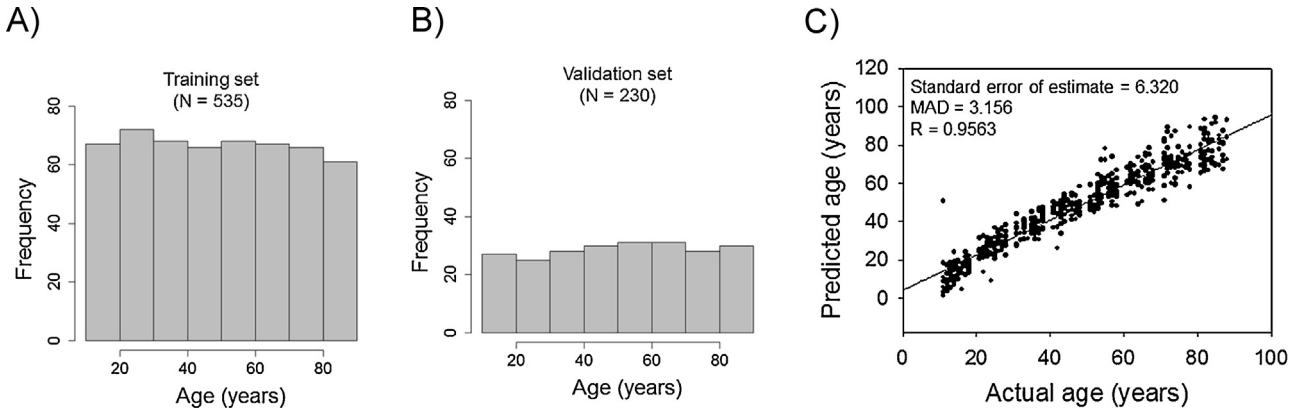

Fig. 4. Development of the statistical model for age prediction using a multiple linear regression method. The 765 samples from Korean individuals were divided into A) a training set (535 samples) and B) a validation set (230 samples) using the caTools package (version 1.17.1). C) The training set was used to develop an age prediction model using the multiple linear regression method. The age prediction model was developed using the CpG_01 from *ZNF423*,*ELOVL2*, and *CCDC102B*.

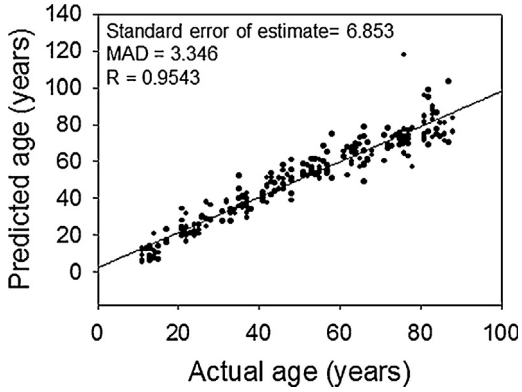

Fig. 5. Accuracy of age prediction in the validation set of 230 samples using our model. We used the CpG_01 sites from *ZNF423*, *ELOVL2*, and *CCDC102B* for age prediction.

However, methylation at certain CpG sites is not always associated with tissue specificity [43]. Therefore, screening with additional technologies will be useful for developing universal markers for age prediction. For example, whole-genome bisulfite sequencing can also be combined with next-generation sequencing technologies, which can cover all CpG sites in humans [44]. Therefore, further studies may identify universal DNA methylation markers that can be used for age prediction across various tissues.

In non-forensic fields, Horvath et al. [21] reported that the combination of 353CpG sites predicts age in a training set with a MAD between predicted and chronological age of 2.7 years. Hannum et al. [22] also reported an age prediction model based on the combination of 71 CpG sites with root mean square error of 4.83 years. In the forensics field, Yi et al. [19] reported an age prediction model showing an average of 4 years difference between predicted and true age. In addition, Zbiec-Piekarska et al. [26] recently reported that the combination of five DNA methylation markers (*ELOVL2*, *C1orf132*, *TRIM59*, *KLF14*, and *FHL2*) can predict age with a MAD of 3.9 years [26]. In this study, we showed MADs to be 3.2 and 3.3 years for the training and validation sample sets, respectively, for the model using the combination of three DNA methylation markers, suggesting that these three markers may be highly useful in forensic science.

In this study, we used linear regression model for development of age prediction model Horvath [21] and Alisch et al. [45] reported that the changes in DNA methylation on aging occur more rapidly until adulthood that is slow as a linear dependence later in life cycle, showing that the change in three methylation markers from this study (Supplementary Fig. S7) is corresponded well to the previous results. Therefore, age prediction model based on the non-linear will be adequate for whole age groups.

Zbiec-Piekarska et al. reported that methylation of the CpG site at position 11044867 on chromosome 6 is the most informative of the seven CpG sites in *ELOVL2* for age prediction [26]. We observed that methylation of the CpG at position 11044894 in *ELOVL2* showed the strongest correlation with age. Although DNA methylation is a stable epigenetic mark, lifestyle and genetic factors are associated with variation in DNA methylation levels [46]. For example, Fraser et al. found population-specific differences in the patterns of DNA methylation at over one-third of all genes between African and European populations [28]. Furthermore, Adkins et al. found that 13.7% of the autosomal CpGs in newborns exhibited significantly different levels of DNA methylation between African American and Caucasian individuals [27]. Another study based on single nucleotide polymorphisms also

Table 3
Evaluation of the accuracy of age prediction in a validation set of 230 samples using our model.

Prediction error	Age prediction	Age category						
		0–19 (MAD = 2.71)	20–39 (MAD = 2.96)	40–59 (MAD = 3.17)	60–79 (MAD = 3.52)	≥80 (MAD = 4.61)	< 60 years	≥ 60 years
≤4 years	Correct	16 (59.26%)	32 (60.38%)	34 (55.74%)	25 (42.37%)	6 (20.00%)	82 (58.16%)	31 (33.71%)
	Incorrect	11 (40.74%)	21 (39.62%)	27 (44.26%)	34 (57.63%)	24 (80.00%)	59 (41.84%)	58 (66.29%)
≤6 years	Correct	21 (77.78%)	44 (83.02%)	44 (72.13%)	40 (67.80%)	11 (36.67%)	109 (77.30%)	51 (57.30%)
	Incorrect	6 (22.22%)	9 (16.98%)	17 (27.87%)	19 (32.20%)	19 (63.33%)	32 (22.61%)	38 (42.70%)
≤8 years	Correct	25 (92.60%)	47 (88.68%)	51 (83.61%)	51 (86.44%)	16 (53.33%)	123 (87.23%)	67 (75.28%)
	Incorrect	2 (7.40%)	6 (11.32%)	10 (16.39%)	8 (13.56%)	14 (46.67%)	18 (12.77%)	22 (24.72%)
≤10 years	Correct	27 (100.00%)	50 (94.34%)	56 (91.80%)	52 (88.14%)	19 (63.33%)	133 (94.33%)	71 (79.78%)
	Incorrect	0 (0.00%)	3 (5.66%)	5 (8.20%)	7 (11.86%)	11 (36.67%)	8 (5.67%)	18 (20.22%)

found that 18% of the methylation at CpG loci is heritable [47]. Therefore, further study is needed to confirm that our three DNA methylation markers are applicable to all ethnic groups.

In conclusion, our study showed that combination of selected three DNA methylation markers may be highly useful in forensic science. We expect that our extensive DNA methylation study and DNA methylation markers will be helpful for improving the use of DNA in forensic field.

Acknowledgements

This study was performed with specimens and data from the Korea Biobank Project (4581-307, KBP-2015-002) that were provided by the Korea Center for Disease Control and Prevention. This work was supported by Forensic Science Research Project 2013 of the Supreme Prosecutors' Office, a grant from the National Research Foundation of the Korea Ministry of Science, ICT, and Future Planning (No. 2011-0030049), and a KRIBB Research Initiative grant.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2016.03.005.

References

[1] E. Hillewig, J. Degroote, T. Van der Paelt, A. Visscher, P. Vandemaele, B. Lutin, et al., Magnetic resonance imaging of the sternal extremity of the clavicle in forensic age estimation: towards more sound age estimates, Int. J. Legal Med. 127 (2013) 677–689.
[2] A. Olze, W. Reisinger, G. Geserick, A. Schmeling, Age estimation of unaccompanied minors: part II. Dental aspects, Forensic Sci. Int. 159 (suppl. 1) (2006) S65–7.
[3] A. Schmeling, C. Grundmann, A. Fuhrmann, H.J. Kaatsch, B. Knell, F. Ramsthaler, et al., Criteria for age estimation in living individuals, Int. J. Legal Med. 122 (2008) 457–460.
[4] G.A. Cortopassi, D. Shibata, N.W. Soong, N. Arnheim, A pattern of accumulation of a somatic deletion of mitochondrial DNA in aging human tissues, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 7370–7374.
[5] C. Meissner, N. von Wurmb, M. Oehmichen, Detection of the age-dependent 4977 bp deletion of mitochondrial DNA. A pilot study, Int. J. Legal Med. 110 (1997) 288–291.
[6] P.M. Helfman, J.L. Bada, Aspartic acid racemisation in dentine as a measure of ageing, Nature 262 (1976) 279–281.
[7] P. Odetti, S. Rossi, F. Monacelli, A. Poggi, M. Cirnigliaro, M. Federici, et al., Advanced glycation end products and bone loss during aging, Ann. N. Y. Acad. Sci. 1043 (2005) 710–717.
[8] Y. Sato, T. Kondo, T. Ohshima, Estimation of age of human cadavers by immunohistochemical assessment of advanced glycation end products in the hippocampus, Histopathology 38 (2001) 217–220.
[9] M.A. Blasco, Telomeres and human disease: ageing, cancer and beyond, Nat. Rev. Genet. 6 (2005) 611–622.
[10] C. Meissner, S. Ritz-Timme, Molecular pathology and age estimation, Forensic. Sci. Int. 203 (2010) 34–43.
[11] R.M. Kohli, Y. Zhang, TET enzymes, TDG and the dynamics of DNA demethylation, Nature 502 (2013) 472–479.
[12] R.J. Klose, A.P. Bird, Genomic DNA methylation: the mark and its mediators, Trends Biochem. Sci. 31 (2006) 89–97.
[13] D.G. Hernandez, M.A. Nalls, J.R. Gibbs, S. Arepalli, M. van der Brug, S. Chong, et al., Distinct DNA methylation changes highly correlated with chronological age in the human brain, Hum. Mol. Genet. 20 (2011) 1164–1172.
[14] W.T. Steegenga, M.V. Boekschoten, C. Lute, G.J. Hooiveld, P.J. de Groot, T.J. Morris, et al., Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs, Age (Dordr.) 36 (2014) 9648.
[15] B.C. Christensen, E.A. Houseman, C.J. Marsit, S. Zheng, M.R. Wrensch, J.L. Wiemels, et al., Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context, PLoS Genet. 5 (2009) e1000602.
[16] A. Johansson, S. Enroth, U. Gyllensten, Continuous aging of the human DNA methylome throughout the human lifespan, PLoS One 8 (2013) e67378.
[17] A. Vidaki, B. Daniel, D.S. Court, Forensic DNA methylation profiling–potential opportunities and challenges, Forensic Sci. Int. Genet. 7 (2013) 499–507.
[18] R. Zbiec-Piekarska, M. Spolnicka, T. Kupiec, Z. Makowska, A. Spas, A. Parys-Proszek, et al., Examination of DNA methylation status of the ELOVL2 marker may be useful for human age prediction in forensic science, Forensic Sci. Int. Genet. 14 (2015) 161–167.
[19] S.H. Yi, L.C. Xu, K. Mei, R.Z. Yang, D.X. Huang, Isolation and identification of age-related DNA methylation markers for forensic age-prediction, Forensic Sci. Int. Genet. 11 (2014) 117–125.
[20] S.H. Yi, Y.S. Jia, K. Mei, R.Z. Yang, D.X. Huang, Age-related DNA methylation changes for forensic age-prediction, Int. J. Legal Med. 129 (2015) 237–244.
[21] S. Horvath, DNA methylation age of human tissues and cell types, Genome Biol. 14 (2013) R115.
[22] G. Hannum, J. Guinney, L. Zhao, L. Zhang, G. Hughes, S. Sadda, et al., Genome-wide methylation profiles reveal quantitative views of human aging rates, Mol. Cell 49 (2013) 359–367.
[23] C.I. Weidner, Q. Lin, C.M. Koch, L. Eisele, F. Beier, P. Ziegler, et al., Aging of blood can be tracked by DNA methylation changes at just three CpG sites, Genome Biol. 15 (2014) R24.
[24] P. Garagnani, M.G. Bacalini, C. Pirazzini, D. Gori, C. Giuliani, D. Mari, et al., Methylation of ELOVL2 gene as a new epigenetic marker of age, Aging Cell. 11 (2012) 1132–1134.
[25] I. Florath, K. Butterbach, H. Muller, M. Bewerunge-Hudler, H. Brenner, 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 (2014) 1186–1201.
[26] R. Zbiec-Piekarska, M. Spolnicka, T. Kupiec, A. Parys-Proszek, Z. Makowska, A. Paleczka, et al., Development of a forensically useful age prediction method based on DNA methylation analysis, Forensic Sci. Int. Genet. 17 (2015) 173–179.
[27] R.M. Adkins, J. Krushkal, F.A. Tylavsky, F. Thomas, Racial differences in gene-specific DNA methylation levels are present at birth, Birth Defects Res. A Clin. Mol. Teratol. 91 (2011) 728–736.
[28] H.B. Fraser, L.L. Lam, S.M. Neumann, M.S. Kobor, Population-specificity of human DNA methylation, Genome Biol. 13 (2012) R8.
[29] E. Schilling, M. Rehli, Global, comparative analysis of tissue-specific promoter CpG methylation, Genomics 90 (2007) 314–323.
[30] L.S. Kristensen, L.L. Hansen, PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment, Clin. Chem. 55 (2009) 1471–1483.
[31] J.L. Park, O.H. Kwon, J.H. Kim, H.S. Yoo, H.C. Lee, K.M. Woo, et al., Identification of body fluid-specific DNA methylation markers for use in forensic science, Forensic Sci. Int. Genet. 13 (2014) 147–153.
[32] J. Tost, I.G. Gut, DNA methylation analysis by pyrosequencing, Nat. Protoc. 2 (2007) 2265–2275.
[33] W. Wagner, C.I. Weidner, Q. Lin, Do age-associated DNA methylation changes increase the risk of malignant transformation, Bioessays 37 (2015) 20–24.
[34] L. Carrel, H.F. Willard, X-inactivation profile reveals extensive variability in X-linked gene expression in females, Nature 434 (2005) 400–404.
[35] T. Mohandas, R.S. Sparkes, L.J. Shapiro, Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation, Science 211 (1981) 393–396.
[36] K. Plath, S. Mlynarczyk-Evans, D.A. Nusinow, B. Panning, Xist RNA and the mechanism of X chromosome inactivation, Annu. Rev. Genet. 36 (2002) 233–278.
[37] N. Vilahur, A.A. Baccarelli, M. Bustamante, S. Agramunt, H.M. Byun, M.F. Fernandez, et al., Storage conditions and stability of global DNA methylation in placental tissue, Epigenomics 5 (2013) 341–348.
[38] A. Zerem, C. Vinkler, M. Michelson, E. Leshinsky-Silver, T. Lerman-Sagie, D. Lev, Mosaic marker chromosome 16 resulting in 16q11. 2-q12. 1 gain in a child with intellectual disability, microcephaly, and cerebellar cortical dysplasia, Am. J. Med. Genet. A 155A (2011) 2991–2996.
[39] S. Huang, J. Laoukili, M.T. Epping, J. Koster, M. Holzel, B.A. Westerman, et al., ZNF423 is critically required for retinoic acid-induced differentiation and is a marker of neuroblastoma outcome, Cancer Cell 15 (2009) 328–340.
[40] H. Zayed, R. Chao, A. Moshrefi, N. Lopezjimenez, A. Delaney, J. Chen, et al., A maternally inherited chromosome 18q22. 1 deletion in a male with late-presenting diaphragmatic hernia and microphthalmia-evaluation of DSEL as a candidate gene for the diaphragmatic defect, Am. J. Med. Genet. A 152A (2010) 916–923.
[41] A.M. Pauter, P. Olsson, A. Asadi, B. Herslof, R.I. Csikasz, D. Zadravec, et al., Elovl2 ablation demonstrates that systemic DHA is endogenously produced and is essential for lipid homeostasis in mice, J. Lipid Res. 55 (2014) 718–728.
[42] C.M. Koch, W. Wagner, Epigenetic-aging-signature to determine age in different tissues, Aging (Albany NY) 3 (2011) 1018–1027.
[43] A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, et al., Integrative analysis of 111 reference human epigenomes, Nature 518 (2015) 317–330.
[44] M.J. Ziller, K.D. Hansen, A. Meissner, M.J. Aryee, Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing, Nat. Methods 12 (2015) 230–232 (1 p following 2).
[45] R.S. Alisch, B.G. Barwick, P. Chopra, L.K. Myrick, G.A. Satten, K.N. Conneely, et al., Age-associated DNA methylation in pediatric populations, Genome Res. 22 (2012) 623–632.
[46] Y.Y. Xia, Y.B. Ding, X.Q. Liu, X.M. Chen, S.Q. Cheng, L.B. Li, et al., Racial/ethnic disparities in human DNA methylation, Biochim. Biophys. Acta 1846 (2014) 258–262.
[47] J.T. Bell, P.C. Tsai, T.P. Yang, R. Pidsley, J. Nisbet, D. Glass, et al., Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population, PLoS Genet. 8 (2012) e1002629.