



Reproducibility and validation of a targeted and flexible epigenetic clock for forensic applications

Ersilia Paparazzo^a, Mirella Aurora Aceto^a, Teresa Serra Cassano^{a,b}, Francesco Bruno^c, Davide Lagrotteria^a, Silvana Geracitano^a, Antonella La Russa^d, Alessia Bauleo^e, Elena Falcone^e, Vincenzo Lagani^{f,g}, Giuseppe Passarino^a, Alberto Montesanto^{a,*}

^a Department of Biology, Ecology and Earth Sciences, University of Calabria, Rende 87036, Italy

^b University of Florence, Department of Statistic, Computer Science and Application, DiSIA, Viale Morgagni, 59, Florence, FI 50134, Italy

^c Department of Human and Social Sciences, Faculty of Social and Communication Sciences, Universitas Mercatorum, Piazza Mattei 10, Rome 00186, Italy

^d Nephrology Unit, Department of Health Sciences, Magna Graecia University, Catanzaro 88100, Italy

^e BIOGENET, Medical and Forensic Genetics Laboratory, Cosenza, ASP 87100, Italy

^f Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology KAUST, Thuwal 23952, Saudi Arabia

^g Institute of Chemical Biology, Ilia State University, Tbilisi 0162, Georgia

ARTICLE INFO

Keywords:

DNA Methylation
Epigenetics
SNaPshot
Forensic genetics
Age-prediction models
Illumina EPIC array
Horvath's clock

ABSTRACT

DNA methylation variants have been widely used as biomarkers of ageing and several mathematical models have been developed to estimate the biological age. More recently, DNA technology has triggered efforts toward the simplification of the array-based epigenetic clocks and targeted approaches, based on the assessment of a small number of CpG sites have been developed. Among the markers included in these clocks, *ELOVL2*, *FHL2*, *KLF14*, *C1orf132/MIR29B2C*, and *TRIM59* resulted to be the most strongly validated markers. We tested the reproducibility and validation of a previously developed targeted epigenetic clock purposely optimized for the measurement of chronological age in blood samples. The clock includes DNAm biomarkers strongly correlated with chronological age whose DNA methylation levels were measured by using a multiplex methylation SNaPshot assay. We found that epigenetic age, calculated using the developed clock, was highly correlated with age ($r = 0.97$) in a total of 201 blood samples covering a full spectrum of human ages. For 74 of these, methylation profiles of the whole genome were obtained through the Infinium Methylation EPIC v2.0 Kit which also allowed to estimate the most frequently used clocks of Horvath. These results show the potential of our efficient and affordable test for simultaneously measuring DNA methylation levels at multiple target CpG sites to assess chronological age. We observed a strong correlation between the prediction models for the analyzed CpG sites measured using the SNaPshot method and those obtained with the Illumina EPIC array, especially with the Horvath2 clock, which was specifically developed for DNA from skin and blood cells.

1. Introduction

Methylation analysis is usually carried out using different DNA methylation technologies, including EpiTYPER®, pyrosequencing, and massively parallel sequencing [1,2]. The high cost associated with both these systems (i.e., equipment and reagents) is a constraining factor for the most part in forensic laboratories that require more cost-effective technology. In addition, small amounts of DNA, which is the most problematic factor in the analysis of forensic caseworks, might be an obstacle to the application of microarray-based clocks that are also

expensive and require at least 250 ng DNA. For this reason, a cost-effective approach might also allow a re-analysis of the same sample, and the increase in the number of technical replicates -rarely performed in the published studies on this topic- has been demonstrated to clearly improve the detection of the methylation levels of the assessed markers and, consequently, the corresponding age prediction models [3]. SNaPshot multiplex methylation test is a simple and cost-effective technique for simultaneously measuring DNA methylation levels at multiple target CpG sites. Importantly, the technology is compatible with forensic DNA genotyping techniques, which facilitates the

* Corresponding author.

E-mail address: alberto.montesanto@unical.it (A. Montesanto).

<https://doi.org/10.1016/j.forensiint.2025.112409>

Received 20 December 2024; Received in revised form 27 January 2025; Accepted 11 February 2025

Available online 14 February 2025

0379-0738/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

application of age estimation of DNA methylation to forensic practice. Two recent studies suggested that methylation profiles obtained with these different methods provide largely comparable results [4,5], even if massively parallel sequencing seems to be the most advantageous approach due to its capability of dealing with low quantity/degraded samples, which can be very common in forensic investigations [2]. Furthermore, MPS is already used in most forensic laboratories for DNA profiles with STR markers, biogeographical ancestry information, mitochondrial DNA sequence analysis, and forensic DNA phenotyping applications [6,7].

Although epigenetic models based on high-throughput technologies are very accurate for age estimation, targeted epigenetic models, based on a limited number of CpG sites and detectable by a fast and cost-effective method, provide reasonably accurate age estimates with additional advantages also in terms of mathematical processing. Several studies have been conducted in recent years with the aim of building targeted epigenetic clocks based on a limited number of CpG sites (Supplementary Table 1).

Except for studies analysed DNA from skeletal tissues for which methylation detection is complicated by technical factors associated with the extraction and low yield of the DNA, a total of 64 studies have been identified. These studies analysed the epigenetic variability of 87 different markers whose detection was based on pyrosequencing (43.7 %) [4,8–34], SNaPshot (34.4 %) [4,5,24,35–53], EpiTYPER (17.2 %) [4,54–63], MPS (11.0 %) [4,41,42,64,65–67], ddPCR and MS-HRM (6.2 %) [39,68–70]. Different population groups were studied, including Belgian [11,12], Spanish [22,55], Portuguese [36,69], Italian [18,46,47,53,58–60], German [10,29], Polish [26,58], Dutch [63,65] and other Caucasian populations [4,19,21,23,24,68], Iraqi [8,19], Indonesian [70], Korean [23,40–44,49,52], Thai [27], Chinese [14, 19–21,30,31,32,37,38,48,50,54,61,62], Japanese [10,25,68,70] and other Asian populations [28,71]. A variety of body fluids were investigated: peripheral blood (67.2 %) [4,5,8,9,11,13–16,19,21,23,26–30,33, 34,36,37,41,43,47–49,51,52–58,60,61–67,69], saliva and buccal swab (29.7 %) [9,10,12,18,21,24,25,35,39,40,42,43,46,49,52,64,66,68,70], semen (12.5 %) [20,21,44,45,50,52,64,66] were the most widely used tissues. Among the 87 epigenetic markers analysed, the CpGs located in the *ELOVL2* gene promoter were the most widely used ones (75.0 %) [4, 8,10–18,22,23–28,30,32–38,41,43,46,47,49,51–61,63,65,66,68,69,70] included in 48 out of 64 developed models followed by *FHL2* (50.0 %) [4,8,13–16,18,26,28,30–37,41,43,46–49,52,53,55–57,59,60,63,70], *C1orf132/MIR29B2C* [4,8,13–16,18,26,28,30,32–38,41,43,46,47,49, 51–57,60,61] *KLF14* (48.4 %) [8,9,14,15,16,18,24,26,28,32–35,37, 38–43,46,47–49,52,53,60,64,65,67,72], and *TRIM59* (39.1 %) [8, 13–15,18,25,26,28,32–37,41,43,46,47,49,52–54,60,61,65]. Accordingly, the most widely used epigenetic model so far proposed in the forensic field, takes advantage of the variability inherent to this set of genes [1,2] with a significant correlation with chronological age observed in different studies [17,19,26,33,36,42,44,47,54,55,57,66,68, 70,72]. Despite these promising results, the most widely used and accurate epigenetic clock for age estimation across various tissues remains the Horvath clock [73]. However, it is relatively expensive and analyzes a large number of CpGs, which may not be feasible for all research settings due to costs and the complexity of the analysis.

The aim of this work was to test the reproducibility and validation of a previously developed epigenetic clock [43] based on a Sanger sequencing approach to 5 blood-based epigenetic markers (*ELOVL2*, *FHL2*, *KLF14*, *C1orf132/MIR29B2C*, and *TRIM59*) for forensic applications. To this purpose, several machine learning algorithms were employed using 201 Italian blood samples covering a full spectrum of human ages. We subsequently compared the 5-CpG epigenetic clock with the array-based models developed by Horvath [73,74] obtained in a subgroup of 74 subjects for which methylation profiles were obtained also with the Infinium™ MethylationEPIC v2.0 BeadChip.

2. Materials and methods

2.1. Samples

A total of 201 blood samples were collected from unrelated individuals (94 men and 107 women) of 0–107 years old covering a full spectrum of human ages. All peripheral blood samples were collected in tubes containing EDTA and stored frozen at -20°C . Genomic DNA was extracted by Invitrogen™ PureLink™ Genomic DNA Mini Kit. Extracted DNA was quantified using a NanoDrop™ One (ThermoFisher). The recruitment of subjects older than 90 years was carried out between 2002 and 2005 through the population registries of Calabrian municipalities. Subjects in the age range 18–89 years were recruited between 2004 and 2007 as part of a survey aimed at monitoring the health status of this population segment in Calabria [75]. Individuals younger than 18 years were recruited at the Medical and Forensic Genetics Laboratory BIOGENET of Cosenza among subjects who required forensic exams during the enrolment period. Fully written informed consent was obtained from all the participants or their parents in case of minors. The study was approved by the Ethics Committee of University of Calabria (Prot. NP-5942018) and met the criteria of the Helsinki declaration.

2.2. Multiplex methylation SNaPshot assay

1 $\mu\text{g}/\mu\text{l}$ of genomic DNA was bisulfite converted using EZ-96 DNA Methylation-Gold kit (Zymo Research, Euroclone, Milan, Italy), according to the manufacturer's protocol. The multiplex methylation SNaPshot assay consisted of multiplex PCR followed by a multiplex SBE reaction using the primers reported in Jung, Sang-Eun et al. [43], with modifications to the primers concentration mixtures, reported in Table S2-a. Multiplex PCR was performed with a total volume of 20 μl , including 2X QIAGEN Multiplex PCR Master Mix (Qiagen), 5X primer mix and 50 ng/ μl of bisulfite-converted DNA. Thermal cycling conditions for PCR amplification were: 95°C for 11 min; 34 cycles of 94°C for 20 s, 56°C for 60 s, and 72°C for 30 s; 72°C for 7 min. After the purification of PCR products with ExoSAP-IT™ (ThermoFisher), multiplex SNaPshot assays were conducted with 1 μl of purified PCR products, 10 X SBE primers (Tables S2-b), 2X Sequencing Buffer (Applied Biosystems), and 1 μl of SNaPshot™ Multiplex Ready Reaction Mix (Applied Biosystems). Thermal cycling conditions for the SNaPshot reaction were: 25 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 30 s. The SNaPshot multiplex reaction was followed by enzymatic purification with Shrimp Alkaline Phosphatase (SAP) (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions.

2.3. SeqStudio genetic analyzers / capillary electrophoresis

The fluorescence and size of SNaPshot products were determined by SeqStudio Genetic Analyzers with a POP-1™ Polymer (Applied Biosystems) as separation matrix. The POP-1 polymer is composed of 5.1 % or more linear polyacrylamide according to the manufacturer. Before loading samples into the genetic analyzer, 1 μl of purified SNaPshot products was mixed with 0.5 μl of GeneScan™ 120 LIZ™ dye Size Standard (Applied Biosystems) and 10 μl of Hi-Di™ Formamide (Applied Biosystems). Each sample was heat denatured at 95°C for 3 min. Using the SeqStudio Genetic Analyzer, injection was performed for 3 s, electrophoresis was carried out for about 8 min at 15 kV at 60°C and 15 kV of run voltage. E5 was used as GeneScan Run module and the DS-02 GeneScan Matrix Set. SeqStudio data were analyzed using GenMapper® Software 5 (Applied Biosystems), with specific detection parameters.

The methylation level at each CpG site (0–1) was calculated by dividing nucleotide C intensity (unconverted methylated DNA) by nucleotide C plus nucleotide T (converted non-methylated DNA) intensities or by dividing nucleotide G intensity (reverse complement of the nucleotide C) by nucleotide G plus nucleotide A (reverse

complement of the nucleotide T) intensities.

The sensitivity of the assay was assessed from the bisulfite conversion step by using fully methylated (100%) and unmethylated (0%) DNA standard controls from the human WGA methylated & non-methylated DNA Set (Zymo Research, Irvine, California, USA).

2.4. Infinium human methylation EPIC BeadChip array

Microarray hybridization was performed by Genomix4Life srl (Baronissi (SA), Italy). From the total cohort on which the above-mentioned model was developed, 74 DNA samples (32 males and 42 females) with an age ranging from 50.0 to 99.0 years were analyzed. A total of 8.9 ng of DNA was bisulfite-converted using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. Amplified DNA was fragmented and hybridized to HumanMethylationEPIC BeadChips (Illumina) in an Illumina hybridization oven at 48 °C for 16–24 h. Following hybridization, arrays were washed for scanning with an iScan Reader (Illumina). Quality and overall chip performance were monitored by visual inspection of both internal quality control checks and raw scanned data.

2.5. Statistical analyses

Predictive modeling was performed on all 201 samples whose methylation levels were detected using the SNaPshot assay. We used a 5-fold cross-validation procedure for training predictive models and assessing their performance at the same time in an unbiased way. In short, at each iteration the dataset was split into a training (80% of samples) and test (remaining 20%) set, ensuring that both sets had similar distribution in terms of age and sex. Predictive models were then trained on the training set and applied to the test set for obtaining testable predictions [76]. For machine learning algorithms requiring hyper-parameters optimization (see below), an internal four-fold cross-validation was performed within the training set for selecting the best hyper-parameters configuration, effectively implementing a nested cross-validation scheme for these algorithms. The mean absolute error was employed as an optimization criterion in the internal cross-validation. The whole (nested) 5-fold cross-validation was repeated 10 times, each time with a different data split, totaling 50 test sets. The first machine learning approach was based on a multiple linear regression model (MLR) using sex and the methylation levels of *C1orf132*, *ELOVL2*, *FHL2*, *KLF14*, *TRIM59* as independent variables. The second one was a multiple quadratic regression (MQR), including the same predictors of the linear model along with the pairwise interactions and the square powers of the methylation values. The third and fourth approaches were based on support vector machines for regression (SVM): the first one exploited a Gaussian kernel (SVMr), the second one a polynomial kernel (SVMp). Both SVM algorithms are non-linear machine learning approaches able to capture complex patterns and interactions within data [77]. SVMs require hyperparameter optimization, meaning some of their (hyper) parameters must be specified a priori before training. During the internal cross-validation, we varied the SVMr cost hyperparameter C in [0.25, 0.5, 1] and the sigma hyperparameter S in [0.038, 0.24, 0.45]. For the SVMp, we used the same C values, along with degree in [1–3] and scale in [0.001, 0.01, 0.1]. We refer to a specialized SVM publication [77] for the meaning of these hyper-parameters. Finally, we trained a multiple linear model using the first two principal components (PCs) derived through principal component analysis (PCA) of the methylation values. To ensure an unbiased assessment of this approach, the PCs loadings derived from the training set were used for computing the PCs on the test set as well. The predictive capability of each age prediction model was evaluated through the mean absolute deviation (MAD), Pearson correlation (r), and correct classification rate (CCR) at 5 and 10 years. MAD provides an immediately comprehensible quantification of the prediction error in terms of how many years each prediction is, on average, off. The Pearson

correlation indicates how well predictions and observed values vary together. With respect to the CCR, the obtained results were evaluated as correct if the predicted age was concordant with the actual age ± 2.5 years (for the 5-year CCR) or ± 5 years (for the 10-year CCR). The final performances are reported as the average across the 50 test folds.

2.6. Genome-wide methylation profiling and age estimation with the Horvath clocks

For 74 samples, methylation profiles of the whole genome were obtained through the Infinium Methylation EPIC v2.0 Kit (Illumina Inc., USA). The samples were analysed within the frame of a European project to investigate the role of epigenetic variability on healthy ageing (“Ageing Well in an Ageing Society”, Next Generation EU [DM 1557 11.10.2022]).

The IDAT files were preprocessed with the Chip Analysis Methylation Pipeline (ChAMP, [78]). In short, probes were first quantified and filtered according to their detection p-value (cut-off = 0.01, [79]), followed by beta-mixture quantile normalization [80]. The Infinium technology processes samples in batches containing 8 samples each, and thus the batch correction method Combat [81] was used for correcting this possible batch effect. The normalized and corrected methylation data were finally used for estimating the chronological age according to the two epigenetic clocks developed by Horvath [73] and skin & blood clock [74], as implemented in the R package methylclock [82]. The first is a multi-tissue clock based on 353 CpG sites (Horvath1). The second is an age estimator based on DNA from skin and blood cells that exploits the epigenetic variability of 391 CpGs (Horvath2).

All statistical analyses and graphical representations were performed in R version 4.4.0 (<https://www.r-project.org/>, accessed on 28 November 2024), using the packages caret (version 6.0 [83]), ChAMP (version 2.29.1 [78]), methylclock (version 1.10.0 [82]).

3. Results

3.1. Multiplex methylation SNaPshot assay for analysis of 5 CpG sites of the *ELOVL2*, *FHL2*, *KLF14*, *C1orf132*/*MIR29B2C*, and *TRIM59* genes

First, a multiplex covering all 5 CpGs was used according to the procedure reported in Jung, Sang-Eun et al. [43]. An example of SNaPshot electropherogram of the optimized multiplex in 2 standard samples of fully unmethylated (0 %) and fully methylated (100 %) commercial DNA together with 5 samples of different ages is shown in Fig. 1.

As shown in Fig. 1 methylation levels of the 5 CpG sites of the analyzed genes in the two standard DNA samples, were consistent with the expected methylation level from each control sample. The mean methylation for the fully unmethylated (0 %) and fully methylated DNA (100 %) control samples were about 0 % and 97.6 %, respectively.

Methylation analysis in 5 samples of different ages showed that methylation levels of CpG sites in *ELOVL2*, *FHL2*, *KLF14*, and *TRIM59* genes increased, while methylation levels of the CpG site of the *C1orf132*/*MIR29B2C* gene decreased with age.

3.2. Construction of targeted and flexible epigenetic clocks

To develop a mathematical model for age prediction, 201 blood samples were analysed using the multiplex methylation SNaPshot assay previously developed. Table 1 and Supplementary Fig. 1 show the age distribution of the analysed sample.

Methylation levels of all five analyzed CpG sites were strongly correlated with age (Fig. 2).

In particular, while *ELOVL2*, *FHL2*, *KLF14*, and *TRIM59* displayed hypermethylation with age, a decrease in the DNA methylation levels for *C1orf132*/*MIR29B2C* was observed ($r = -0.84$). The strongest age-correlation was observed in *ELOVL2* ($r = 0.94$), followed by *FHL2*

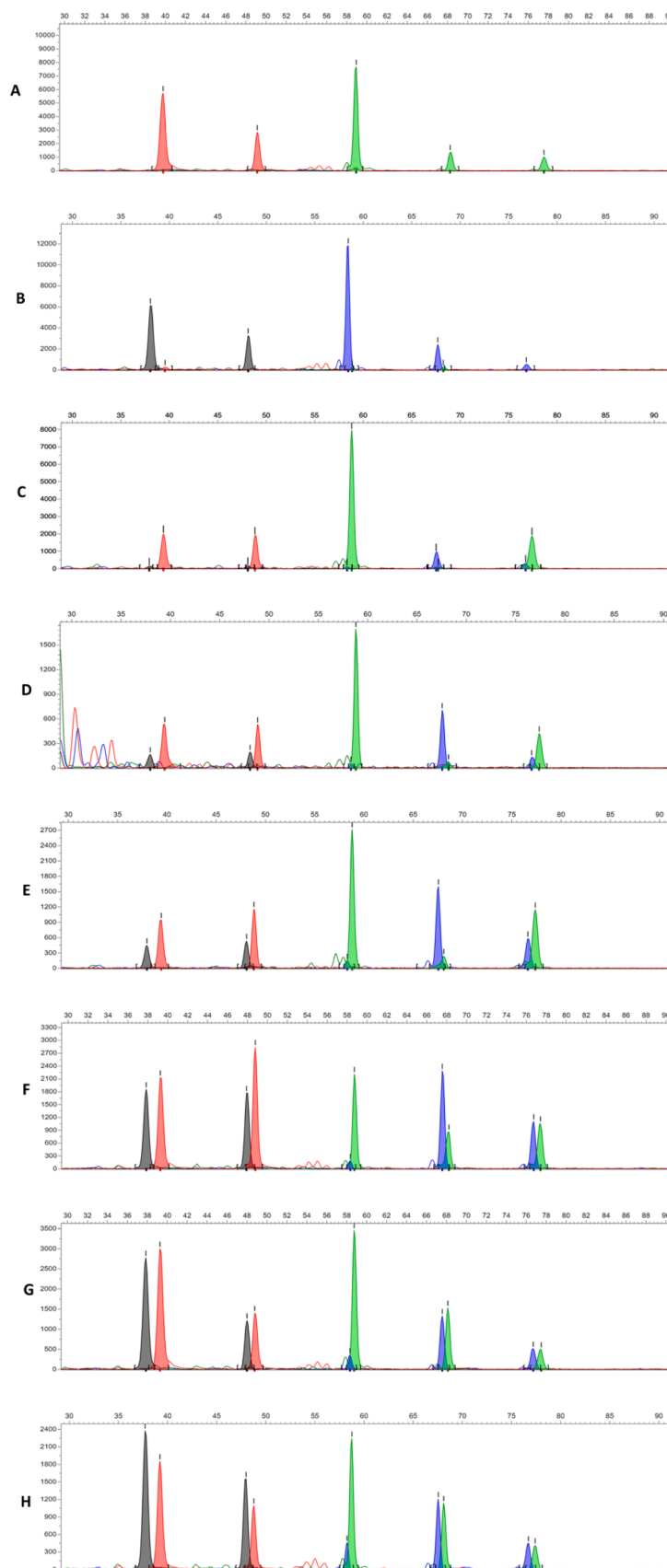


Fig. 1. Electropherograms of the multiplex methylation SNaPshot assay for the detection of DNA methylation at 5 CpG sites of the *ELOVL2*, *FHL2*, *KLF14*, *C1orf132/MIR29B2C* and *TRIM59* genes in 2 standard samples of fully unmethylated (0%) (A), and fully methylated (100%) (B), commercial DNA together with 5 blood samples obtained from individuals in year 0 s (C), 20 s (D), 40 s (E), 60 s (F), 80 s (G), 100 s (H). Black and blue peaks indicate methylation signals, and red and green peaks indicate non-methylation signals.

Table 1
Age distribution of the analysed sample.

Age range (years)	Females n, (mean age)	Males n, (mean age)	All n, (mean age)
0–17	2 (14.0)	5 (8.7)	7 (10.2)
18–29	11 (27.7)	9 (24.6)	20 (26.3)
30–39	14 (34.4)	15 (35.7)	29 (35.1)
40–49	13 (43.5)	15 (43.6)	28 (43.6)
50–59	24 (55.1)	11 (53.9)	35 (54.7)
60–69	13 (64.8)	16 (65.8)	29 (65.4)
70–79	13 (73.4)	11 (73.4)	24 (73.4)
80–89	9 (84.5)	8 (84.4)	17 (84.5)
90–107	8 (100.0)	4 (95.5)	12 (98.6)
0–107	107 (56.6)	94 (52.8)	201 (54.9)

($r = 0.90$), *TRIM59* ($r = 0.89$), and lastly *KLF14* ($r = 0.73$). All these CpG markers were then used to develop a targeted epigenetic clock from childhood to old age using several machine learning algorithms.

3.3. Model validation

Table 2 displays the cross-validated performances obtained on our dataset for the six statistical and machine-learning models.

We found that all models provided similar performance predictions. The model giving the best age prediction accuracy in terms of MAD was the LRM, with a MAD of about 3.94 years followed by RRM (MAD=3.95), PC (MAD=3.99), QRM (MAD=4.02), SVMp (MAD=4.04), and SVMr (MAD=4.34). The model performances in terms of r were very similar, ranging from 0.966 for the SVMr to 0.971 for the RRM. In terms of CCR a performance of around 70% of accuracy in a 10-year time interval was detected for LRM (0.694), RRM (0.711), SVMp (0.710) and PCs (0.712), while QRM and SVMr obtained performances below 70%.

Fig. 3 show the scatterplots contrasting observed and predicted ages for each of the tested ML models.

3.4. Comparison of DNA-based model performances by SNaPshot and Illumina’s EPIC Array

To assess whether EPIC-derived DNA methylation data yield more accurate and precise clocks, we compared the 5-CpG-based age predictions here obtained by the machine learning models against the

prediction provided by the two epigenetic clocks developed by Horvath and co-workers [73,84].

To ensure a fair comparison, we used the cross-validated predictions produced during the repeated (nested) cross-validation procedure, considering only the 74 samples that underwent genome-wide methylation profiling. Table 3 reports the obtained results.

As expected, Horvath2 performs better than Horvath1 clock and presents worse performances than any other method on all metrics. All the 5-CpG-based epigenetic models show slightly better performances than Horvath2. Among these, the MLR should be the preferred one both for its simplicity of use and because it shows performances comparable to other ML-based models.

The lower predictive performance of both array-based epigenetic clocks is probably because our epigenetic models specifically developed on the methylation data obtained on our cohort will obtain higher performances than clocks derived from other, possibly much different, cohorts (Fig. 4).

4. Discussion

While until some years ago most part of epigenetic clocks for forensic age prediction were based on the analysis of hundreds of CpG sites whose methylation levels were detected through microarray technologies, several recent works demonstrated that epigenetic models obtained

Table 2
Predictive performances of the selected age estimation models. For each model the correlation coefficient r , the mean absolute deviance (MAD) and the correct classification rate for a 5-year (CCR5) and a 10-year (CCR10) age interval metrics are reported. LRM: Linear Regression Model; RRM: Ridge Regression Model; QRM: Quadratic Regression Model; SVMr: Support Vector Machine with radial kernel; SVMp: Support Vector Machine with polynomial kernel; PCs: Principal Components.

Algorithm	MAD	r	CCR5	CCR10
LRM	3.936	0.970	0.458	0.694
RRM	3.951	0.971	0.451	0.711
QRM	4.017	0.969	0.496	0.683
SVMr	4.335	0.966	0.417	0.662
SVMp	4.041	0.969	0.448	0.710
PCs	3.992	0.970	0.456	0.712

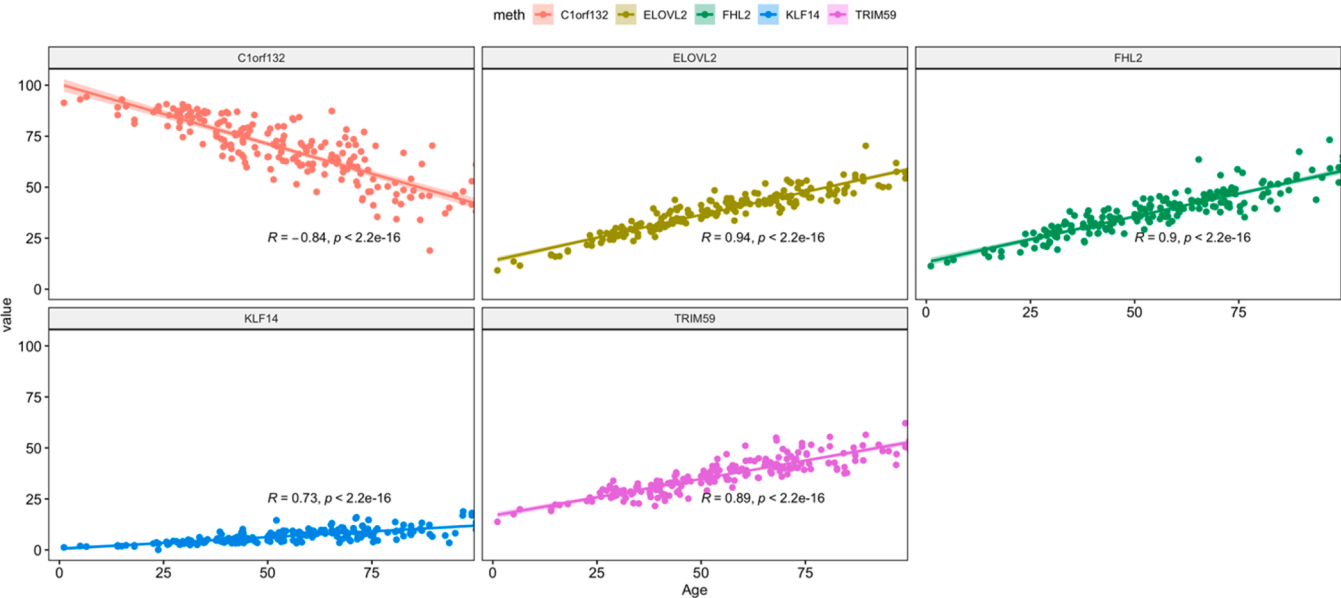


Fig. 2. Scatterplot of methylation levels of C1orf132/MIR29B2C, ELOVL2, FHL2, KLF14, and TRIM59 and the age. The colored lines represent the regression line of methylation levels on chronological age.

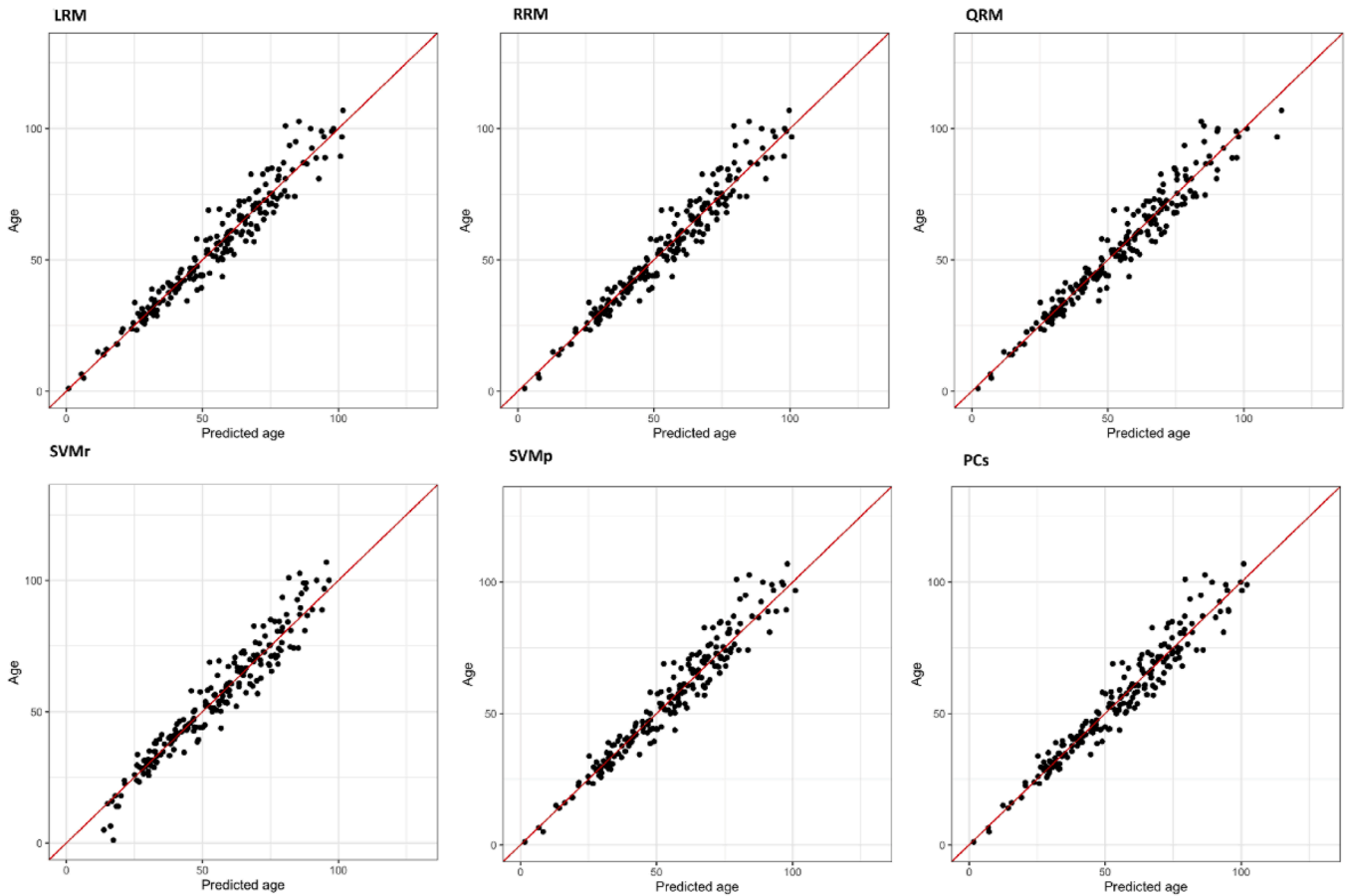


Fig. 3. Scatterplots contrasting observed and predicted ages for each of the tested models. Each dot corresponds to a single prediction. Scatterplots correspond to five ML approaches, which are, from left to right: Linear Regression Model (LRM); Ridge Regression Model (RRM), Quadratic Regression Model (QRM), Support Vector Machine with radial kernel (SVMr); Support Vector Machine with polynomial kernel (SVMp), and Principal Components (PC). The red line in all panels represents a perfect correlation between chronological and epigenetic age.

Table 3

Horvath's epigenetic clocks performances contrasted against the results obtained by the models trained in this work. For the machine learning models, performances are computed on the cross-validated predictions, after restricting the samples to the ones profiled with the Infinium chip.

Algorithm	Correlation	MAE	Five year accuracy	Ten year accuracy
Horvath1	0.819	8.289	0.151	0.274
Horvath2	0.829	5.036	0.274	0.630
LRM	0.861	4.253	0.452	0.671
QRM	0.833	4.683	0.411	0.671
PCs	0.863	4.371	0.384	0.671
SVMp	0.862	4.200	0.438	0.685
SVMr	0.839	4.584	0.452	0.644
RRM	0.864	4.209	0.438	0.699

Horvath1 was defined by 353 CpG sites in [85]; Horvath2 was defined by 391 CpG sites in [84]; LRM: Linear Regression Model; RRM: Ridge Regression Model; QRM: Quadratic Regression Model; SVMr: Support Vector Machine with radial kernel; SVMp: Support Vector Machine with polynomial kernel; PC: Principal Components.

analyzing only a small number of CpG sites provided highly comparable results [5]. This shift offers considerable advantages, especially in terms of cost and efficiency. For instance, BeadChip arrays, while effective, are expensive and require substantial DNA input (at least 250 ng), making them less appealing for forensic applications focused solely on age prediction. On the other hand, single base extension can be easily incorporated into well-established capillary electrophoresis systems to analyses methylation levels of candidate genes. For this reason, targeted

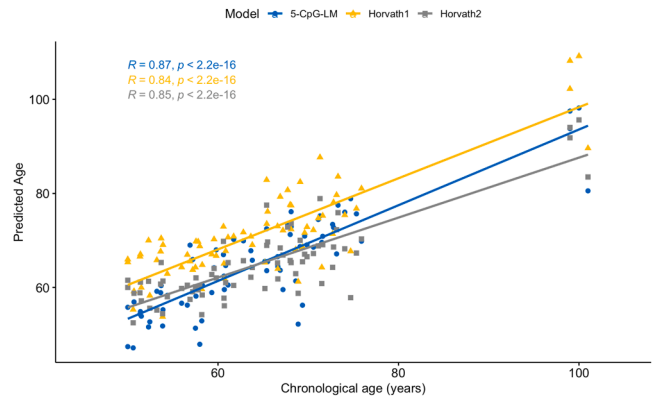


Fig. 4. Scatterplot of methylation age measured with the SNaPshot and EPIC array-based clocks (Horvath1 and Horvath2). The cross-validated predictions produced by the linear model (5-CpG-LM) were obtained using SNaPshot and averaged over folds and repetitions.

epigenetic clocks based on a small number of CpG might be developed and standardized using this kind of technology. In addition, as a cost-effective approach, this method might also allow for a re-analysis of the same samples, and the increase in the number of technical replicates - rarely performed in the published studies on this topic - that it has been demonstrated to clearly improve the detection of the assessed markers and, consequently, the corresponding age prediction models.

In the present work, a simplified full-spectrum epigenetic clock covering a full spectrum of human ages in a technological format easily reproduced in a Sanger sequencing approach (SBE method) suitable for a typical forensic genetics laboratory has been replicated. This enabled us to successfully validate the detection of the five most widely used epigenetic markers for age prediction purposes in the forensic field. Our work replicates the original SNaPshot results in an independent population (Italians) in which the proposed model has never been tested. This is an important result towards the implementation of the SNaPshot model in a routine forensic workflow. Further studies are required to improve the proposed method and to assess the correlation of DNA methylation patterns between pyrosequencing/EpITYPER techniques and the multiplex methylation SNaPshot assay.

We developed a 5-CpG epigenetic clock using blood-based DNA methylation data. Our epigenetic clock showed a more precise chronological age prediction than the well-known Horvath's clocks. A possible explanation is that epigenetic specialized for a specific cohort will obtain higher performances than clocks derived from other, possibly much different, cohorts. Also, methylation data for SNaPshot and Illumina array were obtained from two independent bisulphatation procedures.

The studies here reported could pave the way for the optimization and application of targeted molecular markers of chronological/biological age in large human cohorts. However, there are several open issues that need further studies. An aspect that requires more study is related to the applications of the developed clocks within a forensic framework. As it regards the first point, to further validate the molecular clock here proposed, it is mandatory to investigate its prediction accuracy in very large cohorts of individuals.

From a forensic point of view the application of DNA-based age prediction models appears more complicated since the definition of precise guidelines for the implementation of methods here proposed is a fundamental requirement in the forensic field. In this perspective, the precise definition of a robust set of methylation markers to be analyzed, the relevant methodology for their detection, and an easy-to-use mathematical model for the analysis of the laboratory data make all these steps less complicated. Also, the definition of tissue-specific molecular clocks requires special attention. In fact, the most part of epigenetic clocks so far developed were based on the analysis of whole blood samples. Whole blood is not uniformly composed of identical cell types but consists of distinct cell populations in varying proportions. As methylation profiles of peripheral blood mononuclear cells and granulocytes have been identified, cell heterogeneity could act as a confounder. Additionally, population-specific models capable of delivering highly accurate chronological age estimates, alongside inter-laboratory validation studies, are essential for integrating these methods into forensic casework.

A limitation of this study is the age range of the cohort (>50 years) considered for the comparison between SNaPshot and the array-based method for chronological age estimation. Another potentially limiting factor for comparison between methods, is that bisulfite conversion of DNA was performed on two different aliquots of the same DNA source. For this reason, we cannot completely exclude that the difference in prediction performances between the two methods here used (SNaPshot and EPIC 850k array of Illumina) likely reflects the introduction of other sources of variations. Finally, even if there is a possibility that post-mortem changes could alter the methylation status of markers exploited for age prediction purposes in forensic cases, the present work did not address putative differences in the five analysed markers between blood from living and deceased individuals. Only a few studies attempted to evaluate the postmortem interval as a critical point for age estimation [11,17,27,36,53]. However, all these studies showed the usefulness of the SNaPshot assay for forensic analysis also in blood samples of deceased individuals as post-mortem interval showed no significant impact on the methylation profiles. For this reason, we are confident that the assay here analysed will demonstrate its feasibility for application also in cases where post-mortem degradation poses challenges

concerning sample collection, DNA extraction, and methylation detection. However, further studies are required to validate these preliminary data and to definitely demonstrate the accuracy of these markers also in post-mortem age estimations.

In conclusion, we have developed and validated a simple, cost-effective technique for measuring DNA methylation levels at multiple CpG sites to estimate biological age. Our findings demonstrate a strong agreement between prediction models based on the SNaPshot method and those derived from Illumina's EPIC array, particularly with Horvath2 clocks optimized for skin and blood DNA. This innovative approach represents a promising step forward in the practical application of epigenetic clocks within forensic science, offering a robust, efficient, and scalable solution for age prediction.

Ethical approval

The study was approved by the Ethics Committee of University of Calabria (Prot. NP-5942018) and met the criteria of the Helsinki declaration.

Funding information

1. Next Generation EU, in the context of the National Recovery and Resilience Plan, Investment PE8 – Project Age-It: “Ageing Well in an Ageing Society”. This resource was co-financed by the Next Generation EU [DM 1557 11.10.2022]
2. Mission 4, Component 2, Investment 1.5, call for the creation and strengthening of ‘Innovation Ecosystems’, building ‘Territorial R&D Leaders’ (Directorial Decree n. 2021/3277) - project Tech4You - Technologies for climate change adaptation and quality of life improvement, n. ECS0000009.

CRediT authorship contribution statement

Montesanto Alberto: Writing – original draft, Validation, Supervision, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Paparazzo Ersilia:** Writing – original draft, Validation, Investigation, Data curation. **Passarino Giuseppe:** Writing – review & editing, Conceptualization. **Serra Cassano Teresa:** Writing – review & editing, Investigation, Data curation. **Aceto Mirella Aurora:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Lagrotteria Davide:** Data curation. **Bruno Francesco:** Writing – review & editing, Formal analysis, Data curation. **La Russa Antonella:** Investigation, Data curation. **Geracitano Silvana:** Investigation, Data curation. **Falcone Elena:** Investigation, Data curation. **Bauleo Alessia:** Investigation, Data curation. **Lagani Vincenzo:** Formal analysis, Data curation.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors not used generative AI and AI-assisted technologies in the writing process.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any potential competing or financial interests that could be construed as a potential conflict of interest.

Acknowledgments

We acknowledge co-funding from Next Generation EU, in the context of the National Recovery and Resilience Plan, with the Project PE8 – Project Age-It: “Ageing Well in an Ageing Society” financed by the Next Generation EU [DM 1557 11.10.2022] and the resource Mission 4,

Component 2, Investment 1.5, call for the creation and strengthening of 'Innovation Ecosystems', building 'Territorial R&D Leaders' (Directorial Decree n. 2021/3277) - project Tech4You - Technologies for climate change adaptation and quality of life improvement, n. ECS0000009.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.forsciint.2025.112409.

References

- [1] M. Ghemrawi, B. McCord, Development of a nanoplate-based digital PCR assay for species identification with mixture deconvolution, *Forensic Sci. Int. Genet. Suppl. Ser. 8* (2022) 193–195.
- [2] T.M.T. Carratto, V.M.S. Moraes, T.S.F. Recalde, M.L.G. de Oliveira, C. Teixeira Mendes-Junior, Applications of massively parallel sequencing in forensic genetics, *Genet. Mol. Biol.* 45 (3 1) (2022) e20220077.
- [3] I. Garali, M. Sahbatou, A. Daunay, L.G. Baudrin, V. Renault, Y. Bouyacoub, et al., Improvements and inter-laboratory implementation and optimization of blood-based single-locus age prediction models using DNA methylation of the ELOVL2 promoter, *Sci. Rep.* 10 (1) (2020) 15652.
- [4] A. Freire-Aradas, E. Pośpiech, A. Aliferi, L. Girón-Santamaría, A. Mosquera-Miguel, A. Pisarek, et al., A Comparison of forensic age prediction models using data from four DNA methylation technologies, *Front. Genet.* 11 (2020) 932.
- [5] V.M. Vetter, C.H. Kalies, Y. Sommerer, L. Bertram, I. Demuth, Seven-CpG DNA methylation age determined by single nucleotide primer extension and illumina's infinium methylationepic array provide highly comparable results, *Front. Genet.* 12 (2021) 759357.
- [6] A. Alonso, P. Müller, L. Roewer, S. Willuweit, B. Budowle, W. Parson, European survey on forensic applications of massively parallel sequencing, *Forensic Sci. Int. Genet.* 29 (2017) e23–e25.
- [7] B. Bruijns, R. Tiggelaar, H. Gardeniers, Massively parallel sequencing techniques for forensics: a review, *Electrophoresis* 39 (21) (2018) 2642–2654 (Nov).
- [8] H.S.G. Al-Ghanmy, N.A.M. Al-Rashedi, A.Y. Ayied, Age estimation by DNA methylation levels in Iraqi subjects, *Gene Rep.* 23 (2021) 101022.
- [9] H. Alghanim, J. Antunes, D.S.B.S. Silva, C.S. Alho, K. Balamurugan, B. McCord, Detection and evaluation of DNA methylation markers found at SCGN and KLF14 loci to estimate human age, *Forensic Sci. Int. Genet.* 31 (2017) 81–88.
- [10] J. Becker, P. Böhme, A. Reckert, S.B. Eickhoff, B.E. Koop, J. Blum, et al., Evidence for differences in DNA methylation between Germans and Japanese, *Int. J. Leg. Med.* 136 (2) (2022) 405–413.
- [11] B. Bekaert, A. Kamalandua, S.C. Zapico, W. Van De Voorde, R. Decorte, Improved age determination of blood and teeth samples using a selected set of DNA methylation markers, *Epigenetics* 10 (10) (2015) 922–930.
- [12] B. Bekaert, A. Kamalandua, S.C. Zapico, W. Van De Voorde, R. Decorte, A selective set of DNA-methylation markers for age determination of blood, teeth and buccal samples, *Forensic Sci. Int. Genet. Suppl. Ser. 5* (2015) e144–e145.
- [13] S. Cho, S.E. Jung, S.R. Hong, E.H. Lee, J.H. Lee, S.D. Lee, et al., Independent validation of DNA-based approaches for age prediction in blood, *Forensic Sci. Int. Genet.* 29 (2017) 250–256.
- [14] H. Fan, Q. Xie, Z. Zhang, J. Wang, X. Chen, P. Qiu, Chronological age prediction: developmental evaluation of DNA methylation-based machine learning models, *Front. Bioeng. Biotechnol.* 9 (2022) 819991.
- [15] J. Fleckhaus, P. Bugert, N.A.M. Al-Rashedi, M.A. Rothschild, Investigation of the impact of biogeographic ancestry on DNA methylation based age predictions comparing a Middle East and a Central European population, *Forensic Sci. Int. Genet.* 67 (2023) 102923.
- [16] J. Fleckhaus, P.M. Schneider, Novel multiplex strategy for DNA methylation-based age prediction from small amounts of DNA via Pyrosequencing, *Forensic Sci. Int. Genet.* 44 (2020) 102189.
- [17] K. Fokias, L. Dierckx, W. Van De Voorde, B. Bekaert, Age determination through DNA methylation patterns in fingernails and toenails, *Forensic Sci. Int. Genet.* 64 (2023) 102846.
- [18] F. Gentile, E. Castoldi, P. Serventi, D. Colloca, R. Ciccotelli, A. Marino, Early evaluation of five age-correlated DNA methylation markers in an Italian population sample, *Forensic Sci. Int. Genet. Suppl. Ser. 7* (1) (2019) 424–426.
- [19] Y. Huang, J. Yan, J. Hou, X. Fu, L. Li, Y. Hou, Developing a DNA methylation assay for human age prediction in blood and bloodstain, *Forensic Sci. Int. Genet.* 17 (2015) 129–136.
- [20] L. Li, F. Song, Y. Huang, H. Zhu, Y. Hou, Age-associated DNA methylation determination of semen by pyrosequencing in Chinese Han population, *Forensic Sci. Int. Genet. Suppl. Ser. 6* (2017) e99–e100.
- [21] L. Li, F. Song, M. Lang, J. Hou, Z. Wang, M. Prinz, et al., Methylation-based age prediction using pyrosequencing platform from seminal stains in Han Chinese males, *J. Forensic Sci.* 65 (2) (2020) 610–619.
- [22] A.B. Márquez-Ruiz, L. González-Herrera, J. de D. Luna, A. Valenzuela, DNA methylation levels and telomere length in human teeth: usefulness for age estimation, *Int. J. Leg. Med.* 134 (2) (2020) 451–459.
- [23] J.L. Park, J.H. Kim, E. Seo, D.H. Bae, S.Y. Kim, H.C. Lee, et al., Identification and evaluation of age-correlated DNA methylation markers for forensic use, *Forensic Sci. Int. Genet.* 23 (2016) 64–70.
- [24] K. Schwender, O. Holländer, S. Klopfeisch, M. Eveslage, M.F. Danzer, H. Pfeiffer, et al., Development of two age estimation models for buccal swab samples based on 3 CpG sites analyzed with pyrosequencing and minisequencing, *Forensic Sci. Int. Genet.* 53 (2021) 102521.
- [25] M. Shiga, M. Asari, Y. Takahashi, S. Isozaki, C. Hoshina, K. Mori, et al., DNA methylation-based age estimation and quantification of the degradation levels of bisulfite-converted DNA, *Leg. Med.* 67 (2024) 102336.
- [26] Spólnicka, M. Pośpiech, E. Peptińska, B. Zbieć-Piekarska, R. Makowska, A. Pięta, et al., DNA methylation in ELOVL2 and C1orf132 correctly predicted chronological age of individuals from three disease groups, *Int. J. Leg. Med.* 132 (1) (2018) 1–11.
- [27] P. Sukawutthiya, T. Sathirapattaya, K. Vongpaisarnsin, A minimal number CpGs of ELOVL2 gene for a chronological age estimation using pyrosequencing, *Forensic Sci. Int.* 318 (2021) 110631.
- [28] Z. Thong, X.L.S. Chan, J.Y.Y. Tan, E.S. Loo, C.K.C. Syn, Evaluation of DNA methylation-based age prediction on blood, *Forensic Sci. Int. Genet. Suppl. Ser. 6* (2017) e249–e251.
- [29] C. Weidner, Q. Lin, C. 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 (2) (2014) R24.
- [30] C. Xiao, S. Yi, D. Huang, Genome-wide identification of age-related CpG sites for age estimation from blood DNA of Han Chinese individuals, *Electrophoresis* 42 (14–15) (2021) 1488–1496.
- [31] B. Xie, S. Wang, M. Xie, H. Luo, K. Zhang, H. Zhao, et al., Exploring of new age-related CpGs using Methbank database and the pyrosequencing, *Forensic Sci. Int. Genet. Suppl. Ser. 7* (1) (2019) 266–267.
- [32] F. Yang, J. Qian, H. Qu, Z. Ji, J. Li, W. Hu, et al., DNA methylation-based age prediction with bloodstains using pyrosequencing and random forest regression, *Electrophoresis* 44 (9–10) (2023) 835–844.
- [33] R. Zbieć-Piekarska, M. Spólnicka, T. Kupiec, A. Parys-Proszek, Ż. 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.
- [34] M.L. Kampmann, J. Fleckhaus, C. Børsting, H. Juritkova, A. Pijters, J. Papin, et al., Collaborative exercise: analysis of age estimation using a QIAGEN protocol and the PyroMark Q48 platform, *Forensic Sci. Res.* 9 (1) (2024) owa0055.
- [35] L. Carlsen, O. Holländer, M.F. Danzer, M. Vennemann, C. Augustin, DNA methylation-based age estimation for adults and minors: considering sex-specific differences and non-linear correlations, *Int. J. Leg. Med.* 137 (3) (2023) 635–643.
- [36] H.C. Dias, C. Cordeiro, J. Pereira, C. Pinto, F.C. Real, E. Cunha, et al., DNA methylation age estimation in blood samples of living and deceased individuals using a multiplex SNaPshot assay, *Forensic Sci. Int.* 311 (2020) 110267.
- [37] X. Han, C. Xiao, S. Yi, Y. Li, M. Chen, D. Huang, Accurate age estimation from blood samples of Han Chinese individuals using eight high-performance age-related CpG sites, *Int. J. Leg. Med.* 136 (6) (2022) 1655–1665.
- [38] T. Hao, J. Guo, J. Liu, J. Wang, Z. Liu, X. Cheng, et al., Predicting human age by detecting DNA methylation status in hair, *Electrophoresis* 42 (11) (2021) 1255–1261.
- [39] M. Ho Lee, J. Hee Hwang, K. Min Seong, J. Jin Ahn, S. Jun Kim, S. Yong Hwang, et al., Application of droplet digital PCR method for DNA methylation-based age prediction from saliva, *Leg. Med.* 54 (2022) 101992.
- [40] S.R. Hong, S.E. Jung, E.H. Lee, K.J. Shin, W.I. Yang, H.Y. Lee, DNA methylation-based age prediction from saliva: High age predictability by combination of 7 CpG markers, *Forensic Sci. Int. Genet.* 29 (2017) 118–125.
- [41] S.R. Hong, K.J. Shin, Can we integrate method-specific age-predictive models?: Analysis method-induced differences in detected DNA methylation status, *Forensic Sci. Int. Genet.* 62 (2023) 102805.
- [42] S.R. Hong, K.J. Shin, S.E. Jung, E.H. Lee, H.Y. Lee, Platform-independent models for age prediction using DNA methylation data, *Forensic Sci. Int. Genet.* 38 (2019) 39–47.
- [43] S.E. Jung, S.M. Lim, S.R. Hong, E.H. Lee, K.J. Shin, H.Y. Lee, 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 (2019) 1–8.
- [44] H.Y. Lee, S.E. Jung, Y.N. Oh, A. Choi, W.I. Yang, K.J. Shin, Epigenetic age signatures in the forensically relevant body fluid of semen: a preliminary study, *Forensic Sci. Int. Genet.* 19 (2015) 28–34.
- [45] J.W. Lee, C.M. Choung, J.Y. Jung, H.Y. Lee, S.K. Lim, A validation study of DNA methylation-based age prediction using semen in forensic casework samples, *Leg. Med.* 31 (2018) 74–77.
- [46] B. Marcante, A. Delicati, M. Onofri, P. Tozzo, L. Caenazzo, Estimation of human chronological age from buccal swab samples through a DNA methylation analysis approach of a five-locus multiple regression model, *Int. J. Mol. Sci.* 25 (2) (2024) 935.
- [47] M. Onofri, A. Delicati, B. Marcante, L. Carlini, F. Alessandrini, P. Tozzo, et al., Forensic age estimation through a DNA methylation-based age prediction model in the Italian population: a pilot study, *Int. J. Mol. Sci.* 24 (6) (2023) 5381.
- [48] C. Pan, S. Yi, C. Xiao, Y. Huang, X. Chen, D. Huang, The evaluation of seven age-related CpGs for forensic purpose in blood from Chinese Han population, *Forensic Sci. Int. Genet.* 46 (2020) 102251.
- [49] M.H. So, H.Y. Lee, Genetic analyzer-dependent DNA methylation detection and its application to existing age prediction models, *Electrophoresis* 42 (14–15) (2021) 1497–1506.
- [50] C. Xiao, Y. Li, M. Chen, S. Yi, D. Huang, Improved age estimation from semen using sperm-specific age-related CpG markers, *Forensic Sci. Int. Genet.* 67 (2023) 102941.

- [51] Z. Ye, L. Jiang, M. Zhao, J. Liu, H. Dai, Y. Hou, et al., Epigenome-wide screening of CpG markers to develop a multiplex methylation SNaPshot assay for age prediction, *Leg. Med.* 59 (2022) 102115.
- [52] M.H. So, J.E. Lee, H.Y. Lee, Strategies to deal with genetic analyzer-specific DNA methylation measurements, *Electrophoresis* 45 (9–10) (2024) 906–915.
- [53] P. Grignani, B. Bertoglio, M.C. Monti, R. Cuoghi Costantini, U. Ricci, M. Onofri, et al., Age estimation of burnt human remains through DNA methylation analysis, *Int. J. Leg. Med.* [Internet] (2024). (<https://link.springer.com/10.1007/s00414-024-03320-1>). Sep 13 [cited 2024 Dec 4]; Available from.
- [54] L. Feng, F. Peng, S. Li, L. Jiang, H. Sun, A. Ji, et al., Systematic feature selection improves accuracy of methylation-based forensic age estimation in Han Chinese males, *Forensic Sci. Int. Genet.* 35 (2018) 38–45.
- [55] A. Freire-Aradas, L. Girón-Santamaría, A. Mosquera-Miguel, A. Ambroa-Conde, C. Phillips, M. Casares de Cal, et al., A common epigenetic clock from childhood to old age, *Forensic Sci. Int. Genet.* 60 (2022) 102743.
- [56] A. Freire-Aradas, C. Phillips, L. Girón-Santamaría, A. Mosquera-Miguel, A. Gómez-Tato, M.Á. Casares De Cal, et al., Tracking age-correlated DNA methylation markers in the young, *Forensic Sci. Int. Genet.* 36 (2018) 50–59.
- [57] A. Freire-Aradas, C. Phillips, A. Mosquera-Miguel, L. Girón-Santamaría, A. Gómez-Tato, M. Casares De Cal, et al., Development of a methylation marker set for forensic age estimation using analysis of public methylation data and the Agena Bioscience EpiTYPER system, *Forensic Sci. Int. Genet.* 24 (2016) 65–74.
- [58] N. Gensous, C. Sala, C. Pirazzini, F. Ravaioli, M. Milazzo, K.M. Kwiatkowska, et al., A targeted epigenetic clock for the prediction of biological age, *Cells* 11 (24) (2022) 4044.
- [59] C. Giuliani, E. Cilli, M.G. Bacalini, C. Pirazzini, M. Sazzini, G. Gruppioni, et al., Inferring chronological age from DNA methylation patterns of human teeth, *Am. J. Phys. Anthr.* 159 (4) (2016) 585–595.
- [60] A. Montesanto, P. D'Aquila, V. Lagani, E. Paparazzo, S. Geracitano, L. Formentini, et al., A new robust epigenetic model for forensic age prediction, *J. Forensic Sci.* 65 (5) (2020) 1424–1431.
- [61] F. Peng, L. Feng, J. Chen, L. Wang, P. Li, A. Ji, et al., Validation of methylation-based forensic age estimation in time-series bloodstains on FTA cards and gauze at room temperature conditions, *Forensic Sci. Int. Genet.* 40 (2019 May) 168–174.
- [62] C. Xu, H. Qu, G. Wang, B. Xie, Y. Shi, Y. Yang, et al., A novel strategy for forensic age prediction by DNA methylation and support vector regression model, *Sci. Rep.* 5 (1) (2015) 17788.
- [63] D. Zubakov, F. Liu, I. Kokmeijer, Y. Choi, J.B.J. van Meurs, W.F.J. van IJcken, et al., Human age estimation from blood using mRNA, DNA methylation, DNA rearrangement, and telomere length, *Forensic Sci. Int. Genet.* 24 (2016) 33–43.
- [64] A. Aliferi, D. Ballard, M.D. Gallidabino, H. Thurtle, L. Barron, D. Syndercombe Court, DNA methylation-based age prediction using massively parallel sequencing data and multiple machine learning models, *Forensic Sci. Int. Genet.* 37 (2018) 215–226.
- [65] J. Naue, H.C.J. Hoefsloot, O.R.F. Mook, L. Rijlaarsdam-Hoekstra, M.C.H. van der Zwalm, P. Henneman, et al., Chronological age prediction based on DNA methylation: massive parallel sequencing and random forest regression, *Forensic Sci. Int. Genet.* 31 (2017) 19–28.
- [66] R. Richards, J. Patel, K. Stevenson, S. Harbison, Assessment of DNA methylation markers for forensic applications, *Aust. J. Forensic Sci.* 51 (sup1) (2019) S99–S102.
- [67] A. Vidaki, D. Ballard, A. Aliferi, T.H. Miller, L.P. Barron, D. Syndercombe Court, DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing, *Forensic Sci. Int. Genet.* 28 (2017) 225–236.
- [68] Y. Hamano, S. Manabe, C. Morimoto, S. Fujimoto, K. Tamaki, Forensic age prediction for saliva samples using methylation-sensitive high resolution melting: exploratory application for cigarette butts, *Sci. Rep.* 7 (1) (2017) 10444.
- [69] L. Manco, H.C. Dias, DNA methylation analysis of ELOVL2 gene using droplet digital PCR for age estimation purposes, *Forensic Sci. Int.* 333 (2022) 111206.
- [70] H. Oka, M. Dwi Ariani, T. Akazaki, M. Miyauchi, M. Kitagawa, Some tips on age estimation using DNA methylation in saliva samples as an index across the Japanese and Indonesian ethnicities, *Leg. Med.* 56 (2022) 102042.
- [71] K. Fokias, L. Dierckx, W. Van de Voorde, B. Bekaert, Age determination through DNA methylation patterns of fingernails and toenails [Internet], *Genetics* (2021). (<http://biorxiv.org/lookup/doi/10.1101/2021.12.15.472741>) [cited 2023 Feb 6]. Available from.
- [72] S. Cho, S.E. Jung, S.R. Hong, E.H. Lee, J.H. Lee, S.D. Lee, et al., Independent validation of DNA-based approaches for age prediction in blood, *Forensic Sci. Int. Genet.* 29 (2017) 250–256.
- [73] S. Horvath, DNA methylation age of human tissues and cell types, *Genome Biol.* 14 (10) (2013). R115.
- [74] S. Horvath, J. Oshima, G.M. Martin, A.T. Lu, A. Quach, H. Cohen, et al., Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies, *Aging* 10 (7) (2018) 1758–1775.
- [75] F. De Rango, A. Montesanto, M. Berardelli, B. Mazzei, V. Mari, F. Lattanzio, et al., To grow old in Southern Italy: a comprehensive description of the old and oldest old in Calabria, *Gerontology* 57 (4) (2011) 327–334.
- [76] I. Tsamardinos, A. Rakhshani, V. Lagani, Performance-estimation properties of cross-validation-based protocols with simultaneous hyper-parameter optimization, *Int. J. Artif. Intell. Tools* 24 (05) (2015) 1540023.
- [77] Schölkopf B., Smola A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* [Internet]. The MIT Press; 2001 [cited 2024 Dec 20]. Available from: (<https://direct.mit.edu/books/book/1821/Learnin-g-with-KernelsSupport-Vector-Machines>).
- [78] Y. Tian, T.J. Morris, A.P. Webster, Z. Yang, S. Beck, A. Feber, et al., ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Valencia A, editor, *Bioinformatics* 33 (24) (2017) 3982–3984.
- [79] H. Yang, F. Wang, F. Li, C. Ren, J. Pang, Y. Wan, et al., Comprehensive analysis of long noncoding RNA and mRNA expression patterns in sheep testicular maturation, *Biol. Reprod.* 99 (3) (2018) 650–661.
- [80] A.E. Teschendorff, F. Marabita, M. Lechner, T. Bartlett, J. Tegner, D. Gomez-Cabrero, et al., A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data, *Bioinformatics* 29 (2) (2013) 189–196.
- [81] W.E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods, *Biostatistics* 8 (1) (2007) 118–127.
- [82] D. Pelegi-Sisó, P. De Prado, J. Ronkainen, M. Bustamante, J.R. González, *methclock*: a Bioconductor package to estimate DNA methylation age, in: P. Robinson (Ed.), *Bioinformatics*, 37, 2021, pp. 1759–1760.
- [83] M. Kuhn, *Building Predictive Models in R Using the caret Package*, *J. Stat. Softw.* [Internet] 28 (5) (2008). (<http://www.jstatsoft.org/v28/i05/>) (Available from).
- [84] S. Horvath, K. Raj, DNA methylation-based biomarkers and the epigenetic clock theory of ageing, *Nat. Rev. Genet.* 19 (6) (2018) 371–384.
- [85] S. Horvath, DNA methylation age of human tissues and cell types, *Genome Biol.* 14 (10) (2013). R115.