

BMB Reports – Manuscript Submission

Manuscript Draft

**Manuscript Number:** BMB-17-175

**Title:** DNA methylation-based age prediction from various tissues and body fluids

**Article Type:** Mini Review

**Keywords:** DNA methylation; forensic genetics; age prediction; Aging; epigenetics

**Corresponding Author:** Hwan Young Lee

**Authors:** Sang Eun Jung<sup>1</sup>, Kyoung-Jin Shin<sup>1</sup>, Hwan Young Lee<sup>1,\*</sup>

**Institution:** <sup>1</sup>Department of Forensic Medicine, Yonsei University College of Medicine, 50-1 Yonsei-Ro, Seodaemun-Gu, Seoul 03722, Korea,

**Manuscript Type:** Mini Review

**Title:** DNA methylation-based age prediction from various tissues and body fluids

**Author:** Sang Eun Jung, Kyoung-Jin Shin, Hwan Young Lee\*

\*Corresponding author

**Affiliation:**

Department of Forensic Medicine, Yonsei University College of Medicine, 50-1 Yonsei-Ro,  
Seodaemun-Gu, Seoul 03722, Korea

**Running Title:** DNA methylation-based age prediction

**Keywords:** DNA methylation, forensic genetics, age prediction

**Corresponding Author's Information:**

Tel.: +82-2-2228-2482

Fax: +82-2-362-0860

E-mail: hylee192@yuhs.ac

**ABSTRACT**

Aging is a natural and gradual process in human life. It is influenced by heredity, environment, lifestyle, and disease. DNA methylation varies with age, and the ability to predict the age of donor using DNA from evidence materials at a crime scene is of considerable value in forensic investigations. Recently, many studies have reported age prediction models based on DNA methylation from various tissues and body fluids. Those models seem to be very promising because of their high prediction accuracies. In this review, the changes of age-associated DNA methylation and the age prediction models for various tissues and body fluids were examined, and then the applicability of the DNA methylation-based age prediction method to the forensic investigations was discussed. This will improve the understandings about DNA methylation markers and their potential to be used as biomarkers in the forensic field, as well as the clinical field.

## INTRODUCTION

Aging is a universal phenomenon that occurs in most living organisms and is one of humanity's greatest concerns. Aging is defined as the time-dependent functional decline, which is thought to be associated with various molecular modifications in cells that accumulate over a person's lifetime (1). Since the isolation of the first long-lived strains from *Caenorhabditis elegans* in 1983 (2), extensive aging studies have been conducted with a constantly expanding knowledge of the molecular and cellular mechanisms of life and disease (3). There have been many theories and doctrines to explain the cause and mechanism of aging, but the fact that there is no single theory that can explain aging clearly indicates the complexity and diversity of the aging process. As a step towards understanding the aging process, efforts have been made to identify and categorize the cellular and molecular features of aging. In 2013, nine hallmarks of aging were enumerated by Lopez-Otín et al., and the nine hallmarks are genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication (3). Since these features appear simultaneously during the aging process and are interconnected, identifying the precise causal relationship among them is an important challenge.

Epigenetic alteration is a change in the patterns of epigenetic modifications. Epigenetics is most commonly defined as the modification of DNA and DNA packaging without any change in DNA sequence (4), and plays an important role in various biological processes such as silencing of gene expression or maintaining of genomic stability, thereby implying that it may influence other hallmarks of aging. The epigenetic mark that has been most comprehensively studied is DNA methylation. DNA methylation involves the addition of a methyl group to the 5' cytosine of CG dinucleotide, referred to as CpGs. DNA methylation is related to gene regulation and cell differentiation by affecting transcription factor binding

sites, insulator components, and chromosome morphology (5, 6). In the human genome, 70-80% of the total CpG is methylated and most of the unmethylated CpG is clustered in the CpG island at the 5' end of the genes (7). Factors underlying variable DNA methylation include cell differentiation, aging, genetic variation and environmental exposures such as diet, smoking, sunlight, UV radiation, pathogen infection, exercise, alcohol consumption, stress, etc. A pioneer study on DNA methylation and aging revealed that 5-methylcytosine (5mC) levels dramatically decreased during ontogenesis through the investigation of different organs and life stages of the humpback salmon (8). Such an age-related global DNA hypomethylation has been also observed in a variety of species, including rats (9), mice (10), and humans (11). However, some genes become hypermethylated with age, including Polycomb group protein target genes known to have roles in chromatin remodeling related with gene silencing (12). The advent of microarray and massive parallel sequencing technologies (MPS) in epigenetics enabled the detection of more specific age-associated DNA methylation changes at certain genes or genomic regions. Several recent studies used methylation measures at multiple CpG sites that are highly associated with age to predict chronological age in humans.

Age prediction based on DNA methylation is now attracting attention, especially in the field of forensic genetics because it has been known to be the most accurate method to predict age and can contribute to solving crimes by narrowing down the search range of unknown suspects or missing people. In this review, we would like to focus on DNA methylation-based age predictive models and their potential applicability to the forensic field. More specifically, we will discuss age predictive models that could be applied to the forensically relevant tissues and body fluids such as blood, saliva, semen and skeletal remains.

## **BIOLOGICAL BACKGROUND OF AGE-ASSOCIATED DNA METHYLATION**

## CHANGES

### **DNA methylation changes with aging**

Early studies based on immune, colorimetric and HPLC analysis have shown that the global DNA methylation decreases with age (10, 13, 14). More recent studies reveal that the overall DNA methylation level of the genome increases throughout the first year of life and then stabilizes through adulthood to advanced age in blood (15, 16, 17, 18). Postadulthood, median global DNA methylation decreases as the aging process progresses, and interindividual variability increases over time (10, 13, 14). However, the rate of aging depends mainly on the person, mostly appearing differently depending on the genetic, environment and lifestyle (19). Twin studies have indicated that variability in DNA methylation increases with age while showing that the genetic changes affecting human aging are only about 20 to 30% (20, 21). Other studies have shown that DNA methylation associated with age is dependent on gender, obesity index, and specific tissues or cells (17, 22, 23, 24). Recent studies, based on microarray and MPS analysis, confirm a global methylation decreases with age, but show age-associated increase or decrease in DNA methylation levels at specific loci (25).

### **Epigenetic drift and epigenetic clock**

Age-related DNA methylation changes have been described as two phenomena: the epigenetic drift and the epigenetic clock (26). Epigenetic drift reflects the effects of environment in which a person ages, and explains the increase in discordance between human epigenomes with age. On the other hand, the sites of epigenetic clock show age-associated DNA methylation changes that are common across individuals, and can thus be used to predict an individual's age (Fig. 1).

Since epigenetic drift comprises age-related changes that are acquired environmentally and stochastically (26 27, 28, 29, 30), the specific site that undergoes epigenetic drift in one person does not likely show the same changes in another person. In contrast, epigenetic clock consistently occurs at the same site between individuals (26). The epigenetic clock sites have been identified both within a specific tissue and across various tissues. Several studies show that this characteristic can be tissue-specific (29). In effect, many age predictive models that are specific to a certain tissue or cell type have been reported and the model that can be applied across various tissues also shows different methylation age in different tissues (25).

## **DNA METHYLATION-BASED AGE PREDICTIVE MODELS**

### **Multi-tissue age predictive models**

The advent of microarray technology in the field of epigenetics helped drive a revolution in aging research by paving the way to identify specific CpG sites that are highly associated with age. As more and more DNA methylation profiles were produced using this microarray technology, attempts to identify and use age-associated CpG sites that are not cell or tissue type dependent for construction of the age predictive model have been made, even though the majority of age-associated DNA methylation changes are tissue-specific.

Koch and Wagner (31) identified 19 CpG sites that are hypermethylated with aging from five public data sets produced from 20-30 samples each of dermis, epidermis, uterine cervical smear, T-lymphocyte, and monocyte using Illumina's HumanMethylation27 BeadChip array (Illumina 27K). Four of these sites including CpGs in the *NPTX2*, *TRIM58*, *GRIA2*, and *KCNQ1DN* genes, and another hypomethylation site in the *BIRC4BP* gene were selected to implement an age prediction and the predictive model showed an accuracy with a mean absolute deviation from chronological age (MAD) of 12.7 years in eight validation sets.

These sets were composed of 766 samples including peripheral blood (PB) leukocytes, saliva, PB lymphocytes, CD34<sup>+</sup> hematopoietic progenitor cell, cord blood monocytes, PB monocytes, breast tissue, and buccal epithelial cells. However, breast tissue showed considerably lower DNA methylation than other cell types at the CpG site of the *BIRC4BP* gene and the prediction accuracy increased to 11.4 years of MAD when age prediction was performed and focused on three most significant CpG sites in the *KCNQ1DN*, *NPTX2*, and *GRIA2*.

Horvath (25) also reported an age predictive model that can be applied across a broad spectrum of tissues and cells. The study developed a highly accurate model based on the analysis results of 7,840 non-cancer samples from 82 data sets produced using Illumina 27K or Illumina's HumanMethylation450 BeadChip array (Illumina 450K), which encompass 51 different tissues and cell types. Among the 82 data sets, 39 were used to train the model using a penalized multivariate regression method known as Elastic Net, 32 were used to validate the model and the rest were used for other purposes. In this report, a total of 21,369 CpGs that are present on both Illumina 27K and 450K, and had less than 10 missing values across the data sets were studied. The 353 CpGs that were automatically selected from the elastic net allowed highly accurate age prediction with a MAD of 2.9 years in the training data and a MAD of 3.6 years in the test data. The model performed efficiently in heterogeneous tissues and individual cell types. However, the interesting part of the study was that the data showed significant differences in methylation age from several tissues. A high error, i.e. MAD values of more than 10 years, was observed in breast tissue, uterine endometrium, dermal fibroblasts, skeletal muscle tissue, heart tissue and sperm. Although such phenomenon might attribute to the hormonal effects, menstrual cycle, stem cell recruitment or different cell lineage, it implied that there might exist specific sites within a tissue that would allow for a more accurate prediction of chronological age for that tissue (26).



### Age predictive models for blood

Many studies reported age-associated DNA methylation changes in blood, but there had been no age predictive model specific for blood until Hannum et al. (29) reported a quantitative model of aging using Illumina 450K data from the whole blood of 656 individuals aged 19 to 101. They proposed a new concept of age estimation by combining clinical variables such as sex and BMI with DNA methylation measures for better understanding of epigenetic drift underlying age-associated methylation changes. The model trained with both methylomic and clinical parameters from 482 individuals included 71 age-associated CpGs selected from elastic net regression. The prediction accuracy of the model was high with a MAD of 3.9 years in the training data and 4.9 years in the test data composed of 174 independent samples. It was interesting to note that most of the selected markers were in or near genes implicated in aging-related diseases or functional decreases. For example, six markers of this model laid within the transcription factor *KLF14*, known as a master regulator of obesity and other metabolic traits. In addition, the *ELOVL2* gene was highlighted again as a promising age predictor for blood since the first report, using Illumina 450K to identify age-associated CpGs by Garagnani et al. (32).

Weidner et al. (18) was notable in terms of application because they demonstrated the possibility to predict an age with a high accuracy using only a few CpGs. They first selected 102 age-associated CpGs from Illumina 27K data of 575 blood samples, and constructed an age predictive model with a MAD of 3.3 years. They further reduced the number of CpGs to three for model construction, while keeping the age prediction accuracy considerably high. They finally selected three CpGs of the *ITGA2B*, *ASPA*, and *PDE4C* genes in consideration of age predictability and feasibility to design primers for the following pyrosequencing analysis. The model composed of these three CpGs showed high accuracy with a MAD of 5.4

years in 82 training samples and a MAD of 4.5 years in 69 test samples. The accuracy of age prediction of this model was lower than the models by Hovath (25) or Hannum et al. (29), but this approach had the merit of being cost effective and not requiring complicated bioinformatics.

Utilizing only a few CpG sites is also appealing to forensic geneticists, because most of the casework samples do not provide enough amount of DNA for epigenome-wide analysis such as a microarray analysis. In the field of forensics, Zbieć-Piekarska et al. (33) reported simple, but a highly effective age prediction method that requires the analysis of only a specific genomic region in the *ELOVL2* gene. They analyzed seven adjacent CpG sites in one fragment of the *ELOVL2* gene by pyrosequencing and constructed an age predictive model using two most highly age-associated CpGs. The model showed a relatively high prediction accuracy with a MAD of 5.0 years in the training set of 303 samples and a MAD of 5.8 years in the test set of 124 samples. A total of 68.5% of test samples were within the range of chronological age  $\pm 7$  years, which corresponded to  $\pm 1.0$  Root Mean Square Error (RMSE). Later, they reported another age prediction method that was more accurate and based on the DNA methylation measures from pyrosequencing at five CpG sites, one each from the *ELOVL2*, *C1orf132*, *TRIM59*, *KLF14*, and *FHL2* genes (34). The model demonstrated improved prediction accuracy in blood with a MAD of 3.4 years in the training set of 300 samples and a MAD of 3.9 years in the test set of 120 samples. However, age prediction accuracy decreased in older age group, which reflected the increasing variation in DNA methylation from epigenetic drift with age. This model was further tested in another population of 100 Koreans and showed the very similar accuracy with a MAD of 4.2 years (35). This proved that the age prediction performance of the model was relatively consistent across different population groups.

Park et al. (36) also proposed a highly accurate model for age prediction in blood using

only three CpGs, one from each of the genes, *ELOVL2*, *ZNF423*, and *CCDC102B*. They selected these three sites from Illumina450K datasets of 1415 individuals in consideration of feasibility to design primers for pyrosequencing analysis. The model trained with the DNA methylation values measured from pyrosequencing of 535 blood samples was highly accurate with a MAD of 3.2 years, and produced prediction values with a MAD of 3.4 years from 230 test set samples. Several other models based on different combination of age-associated markers and different analysis platforms such as EpiTyper or MPS have been reported to have high age predictability (37, 38, 39). Up to now the CpGs in the *ELOVL2* gene, however, have been commonly accepted as the most powerful age predictor in blood.

#### **Age predictive models for saliva and buccal epithelial cells**

The first DNA methylation-based age predictive model was proposed by Bocklandt et al. (22) using Illumina 27K data of saliva samples obtained from 34 twin pairs aged 21 to 55 years. The model was composed of three age-associated CpGs of the *EDARADD*, *NPTX2*, and *TOM1L1* genes and showed an average accuracy of 5.2 years. However, the CpG site in the *NPTX2* gene did not show significant age correlation in another data set composed of 29 unrelated females and the model's precision was not fully validated using an independent set of samples.

Hong et al. (40) reported an age predictive model for saliva that is composed of six age-associated CpGs of the *SST*, *CNGA3*, *KLF14*, *TSSK6*, *TBR1*, and *SLC12A5* genes and a cell type-specific CpG marker of the *PTPN7* gene. They selected these six age-associated CpG markers from the Illumina 450K data for 52 saliva samples. A linear regression analysis with age information and the methylation measures obtained from multiplex SNaPshot reaction in an independent set of 113 training samples produced the model with an accuracy of 3.13 years (MAD = 3.13 years, RMSE = 4.3 years). In a subsequent validation with 113 test set

samples, this model also showed high prediction accuracy with a MAD of 3.15 years. On average, 70% of samples had an error between -1.0 and +1.0 RMSEs, and 95% of samples had an error between -2.0 and +2.0 RMSEs.

Eipel et al. (41) also proposed to use cell type-specific CpGs for age prediction from buccal swab samples in combination with age-associated CpGs from blood (3 CpGs, one each in the genes *PDE4C*, *ASPA*, and *ITGA2B*). Their model was composed of two cell type-specific CpGs and three age-associated CpGs, and showed MADs of 5.09 years and 5.12 years in two independent validation sets. Intriguingly, the methylation measures of the cell-specific marker of the *PTPN7* gene from Hong et al. (40) were highly correlated with the predicted epithelial cell compositions calculated using the “Buccal-Cell Signature” function from Eipel et al. (41). These two studies demonstrated the possibility to improve age prediction from tissues with a heterogeneous cell composition, by combining cell type-specific CpG markers and age-associated CpG markers.

### **Age predictive models for other forensically-relevant tissues or body fluids**

Lee et al. (42) proposed an age predictive model based on semen DNA methylation for the first time. Semen can provide important clues for solving sex crimes and it is one of the most forensically relevant body fluids. Lee et al. identified three age-associated CpGs (cg12837463 and two CpGs of the *TTC7B* and *NOX4* genes) from Illumina 450K data for 12 semen samples and an independent validation with methylation SNaPshot assay in 31 semen samples. Among these genes, the *TTC7B* gene demonstrated the highest association with age, and had been reported to show age-related DNA methylation changes in the longitudinal study of sperm methylome from 17 fertile males (43). The linear regression model composed of the three CpGs showed relatively high prediction accuracy with MADs of 4.2 years and 5.4 years in the 31 training samples and 37 test samples, respectively.

There were reports that applied age-associated CpGs from blood to age prediction in other forensically relevant tissues such as teeth (44, 45). Bekaert *et al.* (44) selected four CpGs of the *ASPA*, *PDE4C*, *ELOVL2*, and *EDARADD* genes from the list of age-associated CpGs from blood and demonstrated that the model constructed with DNA methylation measures at these four CpG sites from pyrosequencing analysis had a relatively high prediction accuracy with a MAD of 4.9 years in teeth samples from 29 individuals aged 19 to 70. Giuliani *et al.* (45) also analyzed the DNA methylation at the *ELOVL2*, *FHL2*, and *PENK* genes in the teeth using EpiTYPER system. They constructed age-predictive models separately for each teeth part, i.e., dental pulp, cementum and dentin. Each model was composed of 5 to 13 CpGs from the *ELOVL2*, *FHL2*, and *PENK* genes, and produced MADs of 1.2 years when DNA was recovered from both cementum and pulp from the same tooth, 2.3 years when DNA was recovered from dental pulp, 2.5 years when DNA was extracted from cementum, and 7.1 years when DNA was recovered from dentin. Since each part of the teeth varied in the cell composition, these results could be expected. Dental pulp contains large nerve trunks and blood vessels, but dentin is a calcified tissue mostly composed of inorganic material.

## POSSIBLE APPLICATION OF DNA METHYLATION AGE

### Clinical application of DNA methylation age

There have been continued efforts to identify new risk factors that can improve the ability to predict disease and mortality. Because DNA methylation age, which is the predicted value from DNA methylation-based age prediction, has been expected to reflect biological age, Marioni *et al.* (46) tested whether the difference between DNA methylation age and chronological age ( $\Delta_{\text{age}}$ ) can predict mortality risk in later life. They calculated  $\Delta_{\text{age}}$  in four

longitudinal cohorts of older people using age predictive models by Hannum et al. (29) and Horvath (25). The correlations between chronological age and predicted age were 0.83 and 0.75 for the Hannum measure and the Horvath measure, respectively. The correlation between predicted ages for Hannum and Horvath was 0.77. In the meta-analyzed results, a 5-year higher  $\Delta_{\text{age}}$  was associated with a 16% greater mortality risk after an adjustment for chronological age, sex, childhood IQ, education, social class, hypertension, diabetes, cardiovascular disease, and *APOE* e4 status. A pedigree-based analysis of  $\Delta_{\text{age}}$  indicated that approximately 40% of inter-individual differences in  $\Delta_{\text{age}}$  were due to genetic factors. In addition, another test to examine the relationship between the two DNA methylation measures and exceptional longevity also showed that the long-lived individuals have a young DNA methylation age compared with their chronological age (47). These studies demonstrate DNA methylation age as a measure for biological age that is not always parallel with chronological age, but instead may inform life expectancy predictions (46). DNA methylation age might therefore be useful in the research to identify age accelerator or decelerator for extended human life span.

### **Forensic application of DNA methylation age**

Although DNA methylation age is not always parallel with chronological age, close prediction of age can provide an important investigative lead to narrow down the search range for unknown suspects or victims. Therefore, there have been many reports on DNA methylation-based age prediction in the field of forensics. However, the samples collected from the crime scene frequently contain only a small amount of DNA, and tissue or body fluid type, where the DNA has originated from should be identified in advance for accurate age prediction. Thereby, choosing the most appropriate analytical method for forensic application is more difficult than in clinical application. Thus, there are several key factors to

consider when choosing a method to predict an age based on DNA methylation measures for forensic purpose: 1) availability of age predictive model that is specific for a certain type of tissue or body fluid; 2) amount and quality of the DNA required for analysis; 3) accuracy of the method; 4) robustness and simplicity of the method; 5) availability of statistics for interpretation; 6) availability of specialized equipment and reagents; and 7) the cost.

The two representative models by Hannum et al. (29) and Horvath (25) have high prediction accuracy, and the model by Horvath has merit that can be applied to a broad spectrum of tissues. However, it does not seem feasible to apply these methods to casework samples because of their array-based epigenome-wide analysis platform that requires considerable amount of high quality DNA and complicated bioinformatics. In contrast, the models that were constructed specific for forensically-relevant body fluids using DNA methylation measures obtained from targeted bisulfite sequencing such as pyrosequencing, EpiTyper, methylation SNaPshot, or MPS seem more promising to be easily incorporated into forensic procedures. These methods require smaller amounts of DNA than epigenome-wide analysis and provide reproducible highly accurate DNA methylation measures.

Pyrosequencing provides accurate and detailed profiles of DNA methylation patterns within 100 bases from the pyrosequencing primer binding site. This method is based on the “sequencing by synthesis” principle, and the sequence is determined by the detection of known nucleotide incorporation. The incorporation of nucleotide by a DNA polymerase appears as the light emission from luciferase that is proportional to the amounts of pyrophosphate released upon nucleotide incorporation. The degree of methylation at each CpG site in a sequence is determined from the ratio of nucleotides T and C. This technique is quantitative, and can detect low methylation differences of up to 5%, but does not allow multiplex assay (48).

EpiTYPER<sup>®</sup> is a mass spectrometer-based bisulfite sequencing method that enables



regional-specific DNA methylation analysis (49). This technique uses base-specific enzymatic cleavage coupled to MALDI-TOF mass spectrometry for analysis of bisulfite-converted DNA. Its fast, accurate analysis power and multichannel analysis capability are advantageous, but limited throughput and high cost restricts its wider application in the forensic field.

Methylation SNaPshot is based on the primer extension method that has been originally developed for SNP analysis. It is characterized by the analysis of bisulfite-converted DNA through single-nucleotide primer extension process, which uses a primer designed to be located up to 1 bp upstream of the CpG site (50). This method allows for multiplexed analysis, thereby saving DNA consumed for the assay. However, imbalance in intensities between the different fluorescent dyes of the SNaPshot<sup>®</sup> kit may require adjustment to obtain accurate absolute measures for DNA methylation.

MPS-based targeted bisulfite sequencing is also increasingly used in the field of forensics (38, 39, 51). PCR fragments obtained from bisulfite-converted DNA can be used in the library preparation for MPS analysis. Characteristics such as high throughput and high sensitivity ensure its application in the forensic field, but it is still required to simplify the data analysis procedure to help researchers who are not familiar with big size data.

In addition, since DNA methylation age in the forensic field should provide the information about unknown person's unknown age, the possible range of prediction deviation needs to be considered when data is reported (39, 40). Although Root Mean Square Error (RMSE) is not as clear as MAD, it has other implications from an interpretation standpoint. Many distributions of error were known to follow the normal distribution and in case of a normal distribution, 68% of errors fall between -1.0 and +1.0 RMSEs and 95% of errors fall between -1.96 and +1.96 RMSEs (22, 40). Recent papers on DNA methylation-based age prediction (33, 39, 40) prove that the use of the 1.0 RMSE and 2.0 RMSE range is



informative to predict the possible range of prediction deviation: Around 70% samples fall into  $\pm 1.0$  RMSE range and 95% in the  $\pm 2.0$  RMSE range. These papers applied different analysis platforms to measure DNA methylation, i.e., pyrosequencing (33), methylation SNaPshot (40), and MPS (39) respectively.

Because the type of tissues and body fluids from the crime scene could be more diverse, the development of new models or the evaluation of a model specific for a certain type of tissue or body fluid in other types of tissues or body fluids would be beneficial for the future forensic applications of the DNA methylation-based age prediction method. Also, many casework samples could be present in a mixed form, so the method to isolate specific tissue or cell type needs to be considered for more accurate prediction of age.

## Conclusion

The relationship between DNA methylation and aging has been described as the epigenetic drift and the epigenetic clock. Epigenetic drift increases the interindividual variation with age influenced by environmental factors. In contrast, epigenetic clock sites are highly associated with age across individuals, and thus can be used to predict chronological age. With an advent of microarray technology, more and more researchers reported age-associated DNA methylation markers and age predictive models. DNA methylation age predicted from DNA methylation measures at several to hundreds of CpG sites are expected to measure biological age that is not always parallel with chronological age but may provide information about life expectancy as well as unknown sample donor's appearance.

There exist models that can be applied across broad spectrum of tissues, but the age prediction accuracy varies depending on the tissue type. In the field of forensics, models for age prediction in blood, saliva, semen, buccal swabs, and teeth have been reported using only a few age-associated CpGs. The prediction accuracy of these models is high, with a MAD of

less than 5 years. However, for a better application of these models, pros and cons of related analysis methods should be considered. In addition, efforts should be made not only to identify more age-associated DNA methylation markers for various tissues and cells, but also to develop standards and software for an accurate interpretation of quantitative analysis results. To forecast the possible range of prediction deviation, the RMSE is an appropriate measurement: Around 70% samples fall into  $\pm 1.0$  RMSE range and 95% in the  $\pm 2.0$  RMSE range. In the future, it is expected that the age-associated DNA methylation markers and age predictive models will play an important role, not only in the forensic investigation, but also in clinical research to monitor health condition and to identify age decelerators for the longer life expectancy.

**ACKNOWLEDGEMENTS**

This study was supported by the Bio & Medical Technology Development Program of the National Research Foundation of Korea (NRF) funded by the Korean government (NRF-2014M3A9E1069992).

**FIGURE LEGENDS**

**Figure 1. Schematic representation of epigenetic drift versus the epigenetic clock.** If a specific CpG site is associated with age (top) within an individual, it may be undergoing either epigenetic drift (left) or an epigenetic clock site (right). Both phenomena have different characteristics when examined across a population or within a twin set. Reprinted from Jones et al. (28) with permission from John Wiley & Sons Ltd. and The Anatomical Society.

## REFERENCES

1. Jung M, Pfeifer GP (2015) Aging and DNA methylation. *BMC Biol* 13, 7.
2. Klass MR (1983) A method for the isolation of longevity mutants in the nematode *Caenorhabditis elegans* and initial results. *Mech Ageing Dev* 22, 279–286.
3. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013) The hallmarks of aging. *Cell* 153, 1194–1217.
4. Weinhold B (2006) Epigenetics: the science of change. *Environ Health Perspect* 114(3), A160–167.
5. Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. *Science* 187, 226–232.
6. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13, 484–492.
7. Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16, 6-21.
8. Berdishev MT, Korotaev GK, Boiarskikh GV, Vanyushin BF (1967) Nucleotide composition of DNA and RNA from somatic tissues of humpback salmon and its changes during spawning. *Biokhimiia* 38, 988–993.
9. Pogribny IP, James SJ, Jernigan S, Pogribna M (2004) Genomic hypomethylation is specific for preneoplastic liver in folate/methyl deficient rats and does not occur in non-target tissues. *Mutat Res* 548(1-2), 53–59.
10. Wilson VL, Smith RA, Ma S, Cutler RG (1987) Genomic 5-methyldeoxycytidine decreases with age. *J Biol Chem* 262, 9948–9951.
11. Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB (1994) Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat Genet* 7(4), 536–540.
12. Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A,

- Bernard D, Vanderwinden JM, Bollen M, Esteller M, Di Croce L, de Launoit Y, Fuks F (2006) The polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439(7078), 871–874.
13. Vanyushin BF, Nemirovsky LE, Klimenko VV, Vasiliev VK, Belozersky AN (1973) The 5-methylcytosine in DNA of rats. Tissue and age specificity and the changes induced by hydrocortisone and other agents. *Gerontologia* 19, 138–152.
14. Fuke C, Shimabukuro M, Petronis A, Sugimoto J, Oda T, Miura K, Miyazaki T, Ogura C, Okazaki Y, Jinno Y (2004) Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Ann Hum Genet* 68, 196–204.
15. Marino M, Masella R, Bulzomi P, Campesi I, Malorni W, Franconi F (2011) Nutrition and human health from a sex-gender perspective. *Mol Aspects Med* 32(1), 1–70.
16. Wang Y, Zhang X, Zhang H, Lu Y, Huang H, Dong X, Chen J, Dong J, Yang X, Hang H, Jiang T (2012) Coiled-coil networking shapes cell molecular machinery. *Mol Biol Cell* 23(19), 3911–3922.
17. Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, Warren ST (2012) Age-associated DNA methylation in pediatric populations. *Genome Res* 22(4), 623–632.
18. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, Bauerschlag DO, Jöckel KH, Erbel R, Mühleisen TW, Zenke M, Brümmendorf TH, Wagner W (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol* 15, R24.
19. Zampieri M, Ciccarone F, Calabrese R, Franceschi C, Bürkle A, Caiafa P (2015) Reconfiguration of DNA methylation in aging. *Mech Ageing Dev* 151, 60–70.
20. Herskind AM, McGue M, Holm NV, Sørensen TI, Harvald B, Vaupel JW (1996) The heritability of human longevity: a population-based study of 2872 Danish twin pairs born

- 1870-1900. *Hum Genet* 97, 319–323.
21. Hirner AV, Rettenmeier AW (2010) Methylated metal(loid) species in humans. *Met Ions Life Sci* 7, 465–521.
  22. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, Vilain E (2011) Epigenetic predictor of age. *PLoS One* 6(6), e14821
  23. Numata S, Ye T, Hyde TM, Guitart-Navarro X, Tao R, Wininger M, Colantuoni C, Weinberger DR, Kleinman JE, Lipska BK (2012) DNA methylation signatures in development and aging of the human prefrontal cortex. *Am J Hum Genet* 90, 260–272.
  24. Johansson A, Enroth S, Gyllensten U (2013) Continuous aging of the human DNA methylome throughout the human lifespan. *PLoS One* 8(6), e67378.
  25. Horvath S (2013) DNA methylation age of human tissues and cell types. *Genome Biol* 14, R115.
  26. Jones MJ, Goodman SJ, Kobor MS (2015) DNA methylation and healthy human aging. *Aging Cell* 14(6), 924–932.
  27. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suñer D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C, Esteller M (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A*. 102(30), 10604–10609.
  28. Fraga MF, Esteller M (2007) Epigenetics and aging: the targets and the marks. *Trends Genet* 23, 413–418.
  29. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, Friend S, Ideker T, Zhang K (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 49, 359–367.

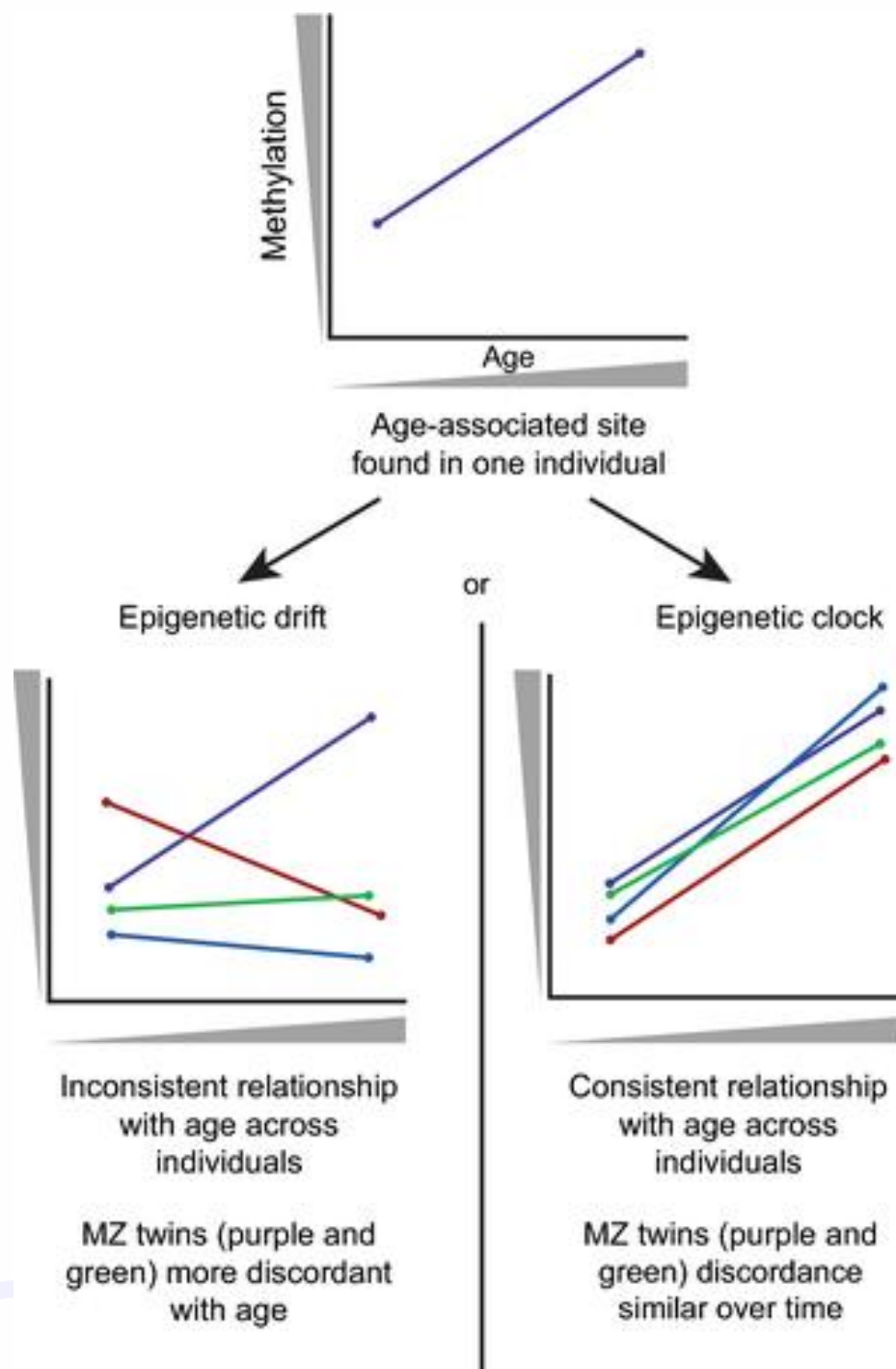
30. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S (2013) A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 29(2), 189–196.
31. Koch C and Wagner W (2011) Epigenetic-aging-signature to determine age in different tissues. *Aging* 3, 1018–1027.
32. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilini D, Vitale G, Collino S, Rezzi S, Castellani G, Capri M, Salvioli S, Franceschi C (2012) Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell* 11, 1132–1134.
33. Zbieć-Piekarska R, Spólnicka M, Kupiec T, Makowska Ż, Spas A, Parys-Proszek A, Kucharczyk K, Płoski R, Branicki W (2015) Examination of DNA methylation status of the ELOVL2 marker may be useful for human age prediction in forensic science. *Forensic Sci Int Genet* 14, 161–167.
34. Zbieć-Piekarska R, Spólnicka M, Kupiec T, Parys-Proszek A, Makowska Ż, Pałeczka A, Kucharczyk K, Płoski R, Branicki W (2015) Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Sci Int Genet* 17, 173–179.
35. Cho S, Jung SE, Hong SR, Lee EH, Lee JH, Lee SD, Lee HY (2017) Independent validation of DNA-based approaches for age prediction in blood. *Forensic Sci Int Genet* 29, 250–256.
36. Park JL, Kim JH, Seo E, Bae DH, Kim SY, Lee HC, Woo KM, Kim YS (2016) Identification and evaluation of age-correlated DNA methylation markers for forensic use. *Forensic Sci Int Genet* 23, 64–70.
37. Freire-Aradas A, Phillips C, Mosquera-Miguel A, Girón-Santamaría L, Gómez-Tato A, Casares de Cal M, Álvarez-Dios J, Ansede-Bermejo J, Torres-Español M, Schneider PM,

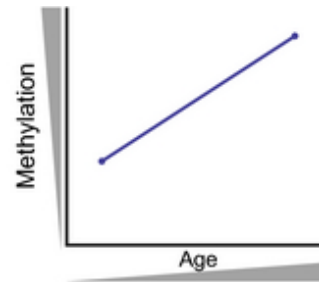


- Pośpiech E, Branicki W, Carracedo Á, Lareu MV (2016) Development of a methylation marker set for forensic age estimation using analysis of public methylation data and the Agena Bioscience EpiTYPER system. *Forensic Sci Int Genet* 24, 65–74.
38. Vidaki A, Ballard D, Aliferi A, Miller TH, Barron LP, Syndercombe Court D (2017) DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. *Forensic Sci Int Genet* 28, 225–236.
39. Naue J, Hoefsloot HCJ, Mook ORF, Rijlaarsdam-Hoekstra L, van der Zwalm MCH, Henneman P, Kloosterman AD (2017) Chronological age prediction based on DNA methylation: massive parallel sequencing and random forest regression. *Forensic Sci Int Genet* 31, 19–28.
40. Hong SR, Jung SE, Lee EH, Shin KJ, Yang WI, Lee HY (2017) DNA methylation-based age prediction from saliva: high age predictability by combination of 7 CpG markers. *Forensic Sci Int Genet* 29, 118–125.
41. Eipel M, Mayer F, Arent T, Ferreira MR, Birkhofer C, Gerstenmaier U, Costa IG, Ritz-Timme S, Wagner W (2016) Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures. *Aging (Albany NY)* 8(5), 1034–1048.
42. Lee HY, Jung SE, Oh YN, Choi A, Yang WI, Shin KJ (2015) Epigenetic age signatures in the forensically relevant body fluid of semen: a preliminary study. *Forensic Sci Int Genet* 19, 28–34.
43. Jenkins TG, Aston KI, Pflueger C, Cairns BR, Carrell DT (2014) Age-associated sperm DNA methylation alterations: possible implications in offspring disease susceptibility. *PLoS Genet* 10(7), e1004458
44. Bekaert B, Kamalandua A, Zapico SC, Van de Voorde W and Decorte R (2015) Improved age determination of blood and teeth samples using a selected set of DNA methylation

- markers. *Epigenetics* 10 (10), 922–930.
45. Giuliani C, Cilli E, Bacalini MG, Pirazzini C, Sazzini M, Gruppioni G, Franceschi C, Garagnani P, Luiselli D (2016) Inferring chronological age from DNA methylation patterns of human teeth. *Am J Phys Anthropol* 159, 585–595.
46. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, Murphy L, Martin NG, Montgomery GW, Feinberg AP, Fallin MD, Multhaup ML, Jaffe AE, Joehanes R, Schwartz J, Just AC, Lunetta KL, Murabito JM, Starr JM, Horvath S, Baccarelli AA, Levy D, Visscher PM, Wray NR, Deary IJ (2015) DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* 16–25.
47. Armstrong NJ, Mather KA, Thalamuthu A, Wright MJ, Trollor JN, Ames D, Brodaty H, Schofield PR, Sachdev PS, Kwok JB (2017) Aging, exceptional longevity and comparisons of the Hannum and Horvath epigenetic clocks. *Epigenomics* 9(5), 689–700.
48. Kurdyukov S, Bullock M (2016) DNA methylation analysis: choosing the right method. *Biology (Basel)* 5(1), 3.
49. Suchiman HE, Sliker RC, Kremer D, Slagboom PE, Heijmans BT, Tobi EW (2015) Design, measurement and processing of region-specific DNA methylation assays: the mass spectrometry-based method EpiTYPER. *Front Genet.* 6, 287.
50. Kaminsky Z, Petronis A (2009) Methylation SNaPshot: a method for the quantification of site-specific DNA methylation levels. *Methods Mol Biol.* 507, 241–55.
51. Masser DR, Stanford DR, Freeman WM (2015) Targeted DNA methylation analysis by next-generation sequencing. *J Vis Exp* 96, e 52488.

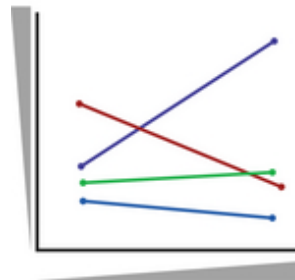
Fig. 1





Age-associated site  
found in one individual

Epigenetic drift

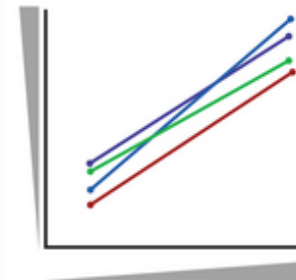


Inconsistent relationship  
with age across  
individuals

MZ twins (purple and  
green) more discordant  
with age

or

Epigenetic clock



Consistent relationship  
with age across  
individuals

MZ twins (purple and  
green) discordance  
similar over time