Forensic Science International: Genetics 19 (2015) 28–34

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Epigenetic age signatures in the forensically relevant body fluid of semen: a preliminary study

Hwan Young Lee*, Sang-Eun Jung, Yu Na Oh, Ajin Choi, Woo Ick Yang, Kyoung-Jin Shin

Department of Forensic Medicine, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 120-752, South Korea

ARTICLE INFO

Article history:
Received 10 November 2014
Received in revised form 18 May 2015
Accepted 23 May 2015

Keywords:
Forensic science
Age
DNA methylation
Semen
TTC7B
NOX4

ABSTRACT

To date, DNA methylation has been regarded as the most promising age-predictive biomarker. In support of this, several researchers have reported age predictive models based on the use of blood or even across a broad spectrum of tissues. However, there have been no publications that report epigenetic age signatures from semen, one of the most forensically relevant body fluids. In genome-wide DNA methylation profiles of 36 body fluids including blood, saliva, and semen, the previous age predictive models showed considerable prediction accuracy in blood and saliva but not in semen. Therefore, we selected CpG sites, whose methylation levels are strongly correlated with age in 12 semen profiles obtained from individuals of different ages, and investigated DNA methylation changes at these CpGs in 68 additional semen samples obtained from individuals aged 20 to 73 years using methylation SNaPshot reaction. Among the selected age-related CpG candidates, outstanding age correlation was obtained at cg06304190 in the *TTC7B* gene. Interestingly, the region around the *TTC7B* gene has been reported to show age-related DNA methylation alteration in the sperm methylome of 2 samples collected from individuals at certain time intervals. The age-predictive linear regression model trained with 3 CpGs (cg06304190 in the *TTC7B* gene, cg06979108 in the *NOX4* gene and cg12837463) showed a high correlation between the predicted age and the chronological age, with an average absolute difference of approximately 5 years. These selected epigenetic age signatures are expected to be useful for considerably accurate age estimation in the forensically relevant body fluid of semen. However, because the findings were limited by small sample size, it will be necessary to further evaluate the age correlation of the selected CpGs and to encourage further investigation.

1. Introduction

Aging is a complex biological process characterized by an overall decline in physiological functions and an increased risk of disease over time [1]. Substantial effort has been devoted to identifying molecular markers that can be used to predict, monitor and provide insights into aging-related physiological decline and disease [2–6]. Moreover, age is an externally visible characteristic that is forensically valuable for predicting an individual's appearance. Therefore, age estimation based on molecular markers is expected to be criminally useful in helping to reduce the number of potential suspects [7,8]. Telomere shortening and accumulation of mitochondrial DNA deletions have been reported to have age dependency [9–10], but they showed low accuracy and various technical problems [4,11]. More recently, a DNA test based on sjTREC DNA quantification has been introduced in the forensic field for chronological age estimation, and this test system achieved R^2 of 0.835 (SE ± 8.9 years) [12]. However, the current most promising age-predictive biomarker is DNA methylation [13–21].

DNA methylation is a major epigenetic modification that occurs at the 5′-position of the pyrimidine ring of cytosine residues within a CpG dinucleotide sequence in adult tissues [22]. Global DNA methylation level decreases with age [23], but many genes or genomic regions have been reported to be hyper- or hypomethylated with increasing age [24,25]. Since the epigenetic landscape varies significantly across tissue types and many age-related DNA methylation changes depend on tissue type [20,25], several previous studies described DNA methylation-based age predictors in specific tissues [13,14,16–19]. For instance, Bocklandt et al. [18] identified 3 CpG sites in the promoters of *EDARADD*,*TOM1L1* and *NPTX2* from saliva using the Illumina HumanMethylation27 (27K) BeadChip array, and built an age predictive model with an average accuracy of 5.2 years. Weidner et al. [14] also identified 3 age-related CpG sites located in the genes *ITGA2B*, *ASPA* and *PDE4C* using the 27K array and subsequent bisulfite sequencing of blood

* Corresponding author. Tel.: +82 2 2228 2482; fax: +82 2 362 0860.
E-mail address: hylee192@yuhs.ac (H.Y. Lee).

http://dx.doi.org/10.1016/j.fsigen.2015.05.014
1872-4973/

samples, and reported a model with a mean absolute deviation (MAD) from chronological age of less than 5 years. Garagnani et al. [17] demonstrated age-related DNA methylation alterations at CpG sites in 3 genes, *ELOVL2, FHL2* and *PENK* with the Illumina HumanMethylation450 (450K) BeadChip analysis of bloods, and suggested *ELOVL2* as the most promising age predictive maker in blood. Later, Zbiec-Piekarska et al. [26] reported an age predictive model for blood using 2 CpGs in the *ELOVL2* gene, which had prediction error of 6.85 years and a MAD from chronological age of 5.03 years. Because the strong age correlation of DNA methylation in the *ELOVL2* gene has been reported in several independent studies [16,17,26–28], *ELOVL2* seems to be one of the most promising age predictive markers in blood to date. On the other hand, age-dependent signatures that were not affected by sex, tissue type or disease state were also reported. A recently reported age-predictive model by Horvath [15] uses 353 CpG sites, and could be applied across a broad spectrum of tissues with a MAD from chronological age of only 3.6 years; Horvath also introduced a freely available online calculator for the epigenetic aging signature. However, age-predictive markers have not yet been defined or evaluated in forensically-relevant body fluid of semen. Even with the model by Horvath, a significant age correlation was not found in sperm, a primary component in semen; the predicted age of sperm was significantly lower than the chronological age of the donor [15].

Meanwhile, an increasing number of rapists are using condoms according to forensic medical examiners' reports. A 1999 study in Oakland, California, found that 13.5% of assailants used condom, probably to protect themselves from identification by DNA profiling (http://www.newscientist.com/article/dn7079). Those condoms are sometimes recovered at the crime scene or at a suspect's home, and can be useful in a police investigation. If DNA profile was obtained from a single source of semen, e.g., from the inside of discarded condom or from the surface of victim's body or from the pellet of differential extraction of sexual assault casework samples, without any alleged suspect, the age estimation with semen will be helpful to reduce the number of potential suspects, thereby contributing to solving crimes.

In the present study, we tested the accuracy of previous age predictive models using genome-wide DNA methylation profiles of 36 body fluids including blood, saliva, and semen. Then, we selected a few age-related CpG candidates from 12 semen profiles obtained through the 450K BeadChip analysis, and tested their age estimation capability by targeted bisulfite sequencing with additional semen samples.

2. Materials and methods

2.1. Samples

Semen samples were collected from 94 healthy male volunteers aged 20 to 73 years using procedures approved by the Institutional Review Board of Severance Hospital, Yonsei University in Seoul, Korea. Among 94 male volunteers, 14 individuals stated that they had a vasectomy. Freshly ejaculated semen was collected in a plastic cup, and 200 μL aliquots of each were stored frozen. DNA was extracted from aliquots of semen using a QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Extracted DNA was quantified using a Quantifiler® Duo DNA Quantification Kit (Applied Biosystems, Foster City, CA, USA).

2.2. Age prediction using previously reported age calculators

To test the accuracy of age calculator suggested by Horvath [15], we used the Illumina HumanMethylation450 BeadChip array (Illumina, San Diego, CA, USA) results for forensically relevant body fluids including 12 blood, 12 saliva and 12 semen samples that had been described in our previous report (GSE59505) [29]. DNA methylation data that were downloaded from the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/) were saved as a file with β-values in the comma delimited file (.csv file) format and were uploaded online according to the tutorial for the age calculation of the website (https://dnamage.genetics.ucla.edu/). To make our data comparable to the training data of the epigenetic clock, the default setting of normalize data was used for data submission. Prediction values from the age calculator were received by an e-mail and were compared with the chronological age.

Another age-predictive model suggested by Weidner et al. [14] was also tested with the same dataset of 36 body fluid samples after normalization with a nonparametric empirical Bayes framework method using ComBat within an R package called surrogate variable analysis (http://www.bioconductor.org/packages/release/bioc/html/sva.html) [30,31]. The model uses 3 CpG sites, and age estimation was implemented as follows:

$$\text{Predicted age} = 38.0 - 26.4 \times \text{cg02228185} - 23.7 \times \text{cg25809905} + 164.7 \times \text{cg17861230}$$

2.3. HumanMethylation450 BeadChip data analysis for screening of age-related CpG candidates from semen

To identify age-related CpG candidates from semen, the 450K BeadChip array data obtained from 12 semen samples were analyzed (GSE59505); semen donors' age was 20, 27, 28, 31, 37, 38, 41, 43, 48, 57, 59 and 59 years. Probe sets with signal intensities below the average background for negative control probes (detection *P*-values ≥0.05) were removed from the data set. The calculated β-score corresponds to the percentage methylation value at a specific CpG site and varied between 0 and 1. To adjust β-scores for batch effects among BeadChips, probe sets with missing values were removed, and β-scores were normalized using ComBat [30,31].

To test for the age-association of β-scores at each CpG unit, univariate linear regression was used. CpGs that exhibited a *P*-value <0.01 with an *R*-squared value over 0.7 and an absolute estimate value over 0.005 were considered candidates for age-related CpGs. However, CpGs with probes containing a SNP within 10 bases of the queried site were eliminated.

2.4. Targeted bisulfite sequencing using methylation SNaPshot

The age-relatedness of selected CpG candidates was further investigated using methylation SNaPshot. In consideration of the practicality with regard to the design of bisulfite sequencing PCR primers, a small set of CpG sites were selected from the list of age-related CpG candidates for the subsequent methylation SNaPshot. Methylation SNaPshot based on a single-base extension reaction (SBE) was designed using *in silico*-bisulfite-converted genomic reference sequences as determined by the BeadChip results. PCR primers for the amplification of bisulfite-converted genomic DNA were designed using the Methprimer program (http://www.urogene.org/methprimer/index1.html) [32], and SBE primers for the target CpGs within the PCR products were designed using the Batchprimer3 program (http://wheat.pw.usda.gov/demos/BatchPrimer3/) [33].

PCR was performed in 20 μL reactions containing 1–2 μL of bisulfite-converted DNA, 1.5 U of AmpliTaq Gold® DNA polymerase, 2.0 μL of Gold ST*R 10× Buffer, and 0.4–1.0 μM of each primer