Application Note

Use of Methylation Markers for Age Estimation of an Unknown Individual Based on Biological Traces

Zbieć-Piekarska Renata[1], Spólnicka Magda[1], Kupiec Tomasz[2], Makowska Żaneta[1], Parys-Proszek Agnieszka[2], Krzysztof Kucharczyk[3], Keith Elliott[6], Płoski Rafał[4], Branicki Wojciech[5]

[1] Central Forensic Laboratory of the Police, Aleje Ujazdowskie 7, 00-583 Warsaw, Poland
[2] Institute of Forensic Research, Westerplatte 9, 31-033 Krakow, Poland
[3] BioVectis, Pawińskiego 5a/d, 02-106 Warsaw, Poland
[4] Medical University of Warsaw, Żwirki i Wigury 61, 02-091 Warsaw, Poland
[5] Institute of Forensic Research, Westerplatte 9, 31-033 Krakow, Poland
[6] QIAGEN Ltd, Skelton House, Lloyd Street North, Manchester, M15 6SH, UK

Introduction

Biological samples and traces collected at crime scenes have potential to be used for predicting the age of the individuals from whom the samples originated. In no-suspect cases and cases with no DNA profile match against a database, such information could be critical for providing additional intelligence for criminal investigations. We have selected 41 CpG sites from the human genome and investigated over 420 samples from men and women between the ages of 2–75 years to determine the degree of methylation at those potential markers using bisulfite conversion and Pyrosequencing® methodology. Based on these results, we used the most significantly correlated CpG sites to build an age prediction model. To facilitate the technical analysis, we selected the lowest possible number of markers (namely five loci: ELOVL2 on 6p24.2, C1orf132 on 1q32.2, TRIM59 on 3q25.33, KLF14 on 7q32.3 and FHL2 on 2q12.2) without lowering the prediction accuracy significantly. The standard error of age estimation using this set of markers was 4.5 years.

We tested the mathematical model performance of these five markers using an independent set of 120 samples. The mean absolute deviation for this test set was 3.9 years. For the age range of 2–19 years, we observed 86.7% accurate predictions with a ±5 year accuracy. This percentage gradually decreased to 50% for the age category 60–75 years. Determining methylation levels of the selected loci using bisulfite conversion and Pyrosequencing is reliable and can be scaled up and fully automated. Thus, this approach is a reliable and effective method for age prediction for forensic purposes.

Materials and Methods

Samples

Blood samples were collected in EDTA blood tubes from volunteers who signed informed consent statements prior to sample donation. Samples from a total of 427 unrelated males and females between 18–75 years old were analyzed. In addition, samples from children between 2–17 years old were collected, and written consents were obtained from their parents. Genomic DNA was extracted from whole blood using the phenol/chloroform method and a standard protocol or with a commercially available kit according to the manufacturer's protocol.

Bisulfite conversion

Unmethylated cytosines in the extracted DNA were converted to uracils using the EpiTect® 96 Bisulfite Kit according to the manufacturer's instructions. PCR amplification of selected markers was performed using primers designed with the PyroMark® Assay Design Software 2.0. PCR reactions were carried out in a total volume of 25 µl, containing 0.2 mM of each of the primers, 20 ng of template DNA, and the PyroMark PCR Master Mix.

Pyrosequencing

Pyrosequencing was performed using PyroMark Gold Q24 Reagents with the PyroMark Q24 Vacuum Workstation and PyroMark Q24 instrument following the manufacturer's instructions. A 10 µl aliquot of the PCR product was immobilized to 1 µl of Streptavidin Sepharose® High Performance (GE Healthcare). Annealing was carried out for 2 minutes at 80 °C with 25 µl of 0.3 mM sequencing primer. The resulting Pyrogram® traces were automatically analyzed using PyroMark analysis software.

Assay dependence on the amount of total DNA present in the sample

The assay's dependence on the amount of DNA was evaluated by analyzing decreasing amounts of template DNA in the probe. The genomic material was extracted using the phenol–chloroform method from three individuals aged 15, 45 and 62. The total DNA concentration was measured prior to bisulfite conversion, and consecutive DNA dilutions ranging from 20 ng to 2.5 ng were subjected to PCR and Pyrosequencing.

Samples were analyzed in 6 replicates, and average methylation was determined. The methylation values obtained were also used to predict age with the developed model.

Assay stability

The influence of sample storage time on the methylation level was analyzed. The reference data were obtained from peripheral blood collected directly from seven individuals aged 19, 21, 26, 36, 42, 58 and 59. The second group was developed by depositing 50 µl of blood from the same group of individuals onto cotton material and storing at room temperature for one month.

In both cases, DNA was extracted and quantified, and the standard DNA methylation analysis protocol was applied. An additional 45 blood stain samples were also subjected to DNA extraction, quantification and the standard DNA methylation protocol. These blood stains had been previously prepared by depositing blood from individuals of known chronological ages onto tissue paper and storing at room temperature for 5 years (15 samples), 10 years (15 samples) and 15 years (15 samples).

Statistical analysis

Simple linear regression was used to analyze the relationship between the methylation level at particular CpG sites and age in a whole testing group (303 males and females) ranging in age from 2–75 years. Standardized regression coefficients were used to compare the contribution of each locus to the age prediction accuracy. Next, multivariate linear regression was applied, allowing simultaneous analysis of all the cytosines tested. Finally, a linear regression prediction model was developed based on the methylation data obtained for the same testing group. The age prediction accuracy of the final model and contribution of particular predictors were assessed using the adjusted R2 statistic, which measures the proportion of age variation that is explained by the developed model.

The final developed prediction model was validated using a different set of samples: a testing group containing 124 samples from individuals from 2–75 years old. The percentage of correct predictions for this group was calculated. Samples included in both sets were selected to equally cover the age range from 2–75 years and four age categories: 0–19, 20–39, 40–59 and 60–80. The mean absolute deviation (MAD) from the chronological age was also calculated. All the analyses were performed using IBM® SPSS® statistics version 21.

Results and Discussion

Analyzing DNA methylation at five markers located in ELOVL2, C1orf132, TRIM59, KLF14 and FHL2 facilitates human chronological age prediction with +/-3.9 years accuracy. Figure 1 shows the predicted age plotted against the chronological age for the training set (300 samples) and testing set (120 samples) using age prediction with these five markers. We observed 78.3% correct predictions at ±5 years level in subjects aged 2–59 years. This level dropped to 50% for subjects between 60–75 years old, indicating that the selected DNA methylation markers are less efficient in older individuals. This could be due to different methylation levels of the selected markers resulting from longer lifespans or differences in medical histories, lifestyles or environmental affects, that ultimately increase the age estimation error in our model.

The algorithm developed using these five methylation markers with optimized Pyrosequencing assays provides a valuable tool for forensic age prediction.

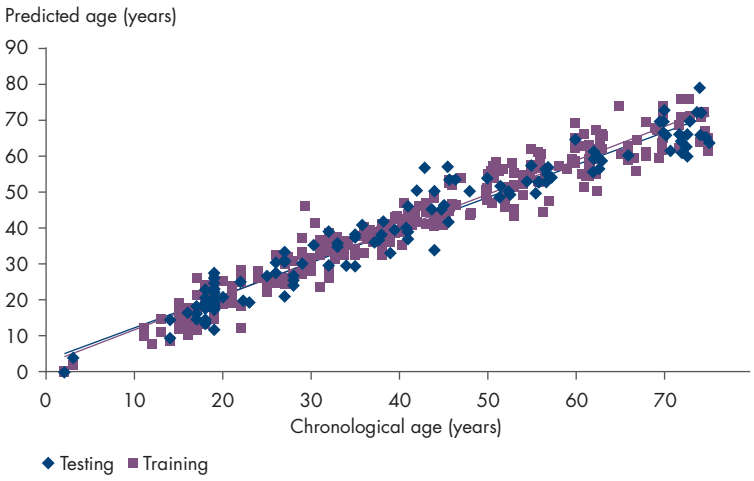

Figure 1. Predicted age versus chronological age of the training set (300 samples) and testing set (120 samples). In both sets, the age predictions were made using the model including ELOVL2, C1orf132, TRIM59, KLF14 and FHL2.

Ordering Information

Product	Contents	Cat. no.
PyroMark Q24 System	Instrument and software for Pyrosequencing analysis: includes installation, training and 1-year warranty on parts and labor	9001514
PyroMark Q24 Vacuum Workstation	Vacuum workstation for preparing 24 samples in parallel from PCR product to single-stranded template	Varies
PyroMark Assay Design Software 2.0	Software for convenient design of PCR and sequencing primers, optimized for Pyrosequencing analysis	9019077
EpiTect 96 Bisulfite Kit	2x EpiTect Bisulfite 96-well Plates, Reaction Mix, DNA Protect Buffer, Carrier RNA and Buffers	59110
PyroMark PCR Kit (200)	2x PyroMark PCR Master Mix (includes HotStarTaq® DNA Polymerase and optimized PyroMark Reaction Buffer containing 3 mM $MgCl_2$ and dNTPs), 10x CoralLoad® Concentrate, 5x Q-Solution, 25 mM $MgCl_2$ and RNase-Free Water	978703
PyroMark Gold Q24 Reagents (5 x 24)	Nucleotides, enzyme and substrate solutions, intended for use with PyroMark Q24	970802

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at **www.qiagen.com** or can be requested from QIAGEN Technical Services or your local distributor.

The research of all Polish authors was financed by The National Center for Research and Development, Warsaw, Poland under grant number DOBR/0002/R/ID1/2012/03.