Deborah Soares Bispo Santos Silva[1,2]
Joana Antunes[1]
Kuppareddi Balamurugan[3]
George Duncan[4]
Clarice Sampaio Alho[2]
Bruce McCord[1]

[1]Department of Chemistry and Biochemistry, Florida International University, Miami, FL, USA
[2]Faculty of Biosciences, Laboratory of Human and Molecular Genetics, PUCRS, Porto Alegre, Brazil
[3]School of Criminal Justice, University of Southern Mississippi, Hattiesburg, MS, USA
[4]DNA Laboratory, Broward County Sheriff's Office, Fort Lauderdale, FL, USA

Received March 16, 2015
Revised April 14, 2015
Accepted April 15, 2015

Research Article

Evaluation of DNA methylation markers and their potential to predict human aging

We present epigenetic methylation data for two genetic loci, GRIA2, and NPTX2, which were tested for prediction of age from different donors of biofluids. We analyzed 44 saliva samples and 23 blood samples from volunteers with ages ranging from 5 to 72 years. DNA was extracted and bisulfite modified using commercial kits. Specific primers were used for amplification and methylation profiles were determined by pyrosequencing. Methylation data from both markers and their relationship with age were determined using linear regression analysis, which indicates a positive correlation between methylation and age. Older individuals tend to have increased methylation in both markers compared to younger individuals and this trend was more pronounced in the GRIA2 locus when compared to NPTX2. The epigenetic predicted age, calculated using a GRIA2 regression analysis model, was strongly correlated to chronological age ($R^2 = 0.801$), with an average difference of 6.9 years between estimated and observed ages. When using a NPTX2 regression model, we observed a lower correlation between predicted and chronological age ($R^2 = 0.654$), with an average difference of 9.2 years. These data indicate these loci can be used as a novel tool for age prediction with potential applications in many areas, including clinical and forensic investigations.

Keywords:
Age prediction / Biofluids / DNA methylation / Pyrosequencing
DOI 10.1002/elps.201500137

1 Introduction

The process of aging affects an organism and the changes that occur in cells and tissues, including modifications in gene regulation, telomeres, cellular structures, and an accumulation of DNA mutations. The effects caused by aging are usually due to chemical modifications of the genome when the organism is affected by diseases, lifestyle, and its genetic makeup [1– 4].

Human aging is associated with epigenetic modifications such as DNA methylation, which is a modification involved in transcriptional regulation. A typical mammalian genome has approximately 3 billion base pairs, of which about 40% are G:C basepairs. In a DNA linear sequence, when a cytosine is followed by a guanine, it is commonly referred to as a CpG site. The CpG dinucleotide distribution is uneven, with several short DNA elements having a much higher density of CpG dinucleotides than other regions of the genome, forming the so-called CpG islands. Most of these islands are located near transcription start sites. In mammalian DNA, methylation occurs at position C5 of the cytosine in some CpG dinucleotides [5– 8]. With the increasing age, some gene-specific CpG dinucleotides can become hypermethylated or hypomethylated. Two main factors have been pointed out as responsible for driving change in the aging methylome: environmental exposure and spontaneous epigenetic changes [9].

Several studies have investigated biomarkers for aging that could be used to track donor age. When compared to messengerRNA or proteins, the methylation of DNA is considered more chemically and biologically stable, making it an ideal biomarker for human age. It is particularly useful in situations such as forensic casework where evidentiary material could easily become degraded or inhibited through contact with exogenous materials [10]. The potential to assess human aging by evaluating the level of DNA methylation has practical implications, including prevention and treatment of diseases, health assessment in addition to forensic analysis [9].

Fraga et al. [11] were among the first to study DNA methylation and its association with age. They performed analysis in samples from monozygotic twin pairs and observed an increased difference in DNA methylation when they compared older twins with younger ones. Other studies [12, 13] looked at specific regions of the genome and observed that specific genome locations can either increase or decrease the methylation level with age, most likely due to differences in gene expression in each cell type. Yi et al. [14] listed the

Correspondence: Dr. Bruce McCord, Department of Chemistry and Biochemistry, Florida International University, 11200 SW 8th St, Miami, FL 33199, USA
E-mail: mccordb@fiu.edu

Abbreviation: CpG, cytosine followed by guanine in a DNA linear sequence

Colour Online: See the article online to view Figs. 1 and 2 in colour.

Table 1. Panel of loci used in this study

locus	Chr	Gene ID	PCR and sequencing primers	CpG sites analyzed in the nucleotide dispensation order (bold)
GRIA2	4	2891	Forward primer-GTATGGGAGGGTGTTGAATATTT Reverse primer- AAAAAAATCCCTATTTCCCAAATCCTACT Seq. primer-ATATTTAGAGATATTGGGATTATAG	**CG**GTAGTTT**CG**TTGAAAATTGTATTTAGTTAGTTTTT**CG**G ATTTTTGGAGCGGGGATAGGGCG
NPTX2	7	4885	Forward primer-GGTTGTGAGAAGGTAGGAGATTT Reverse primer-ACCAACAACCCCAACATCCC Seq. primer-AGAAGGTAGGAGATTTTTGTT	TCG**CG**GTGTACG**CG**ATTTTCGAGA**CG**ATAGCG**CG**GTT ATTGTTAGTAG**CG**AAGGCGTTTTTCG**CG**

Chr, chromosome.

sequence for eight genome locations in which the methylation levels are significantly correlated with age. Most recently, Zbiec-Piekarska et al. [15] identified CpG sites located in the *ELOVL2* gene that may be useful as predictors for human aging.

Koch and Wagner have examined sets of previously published DNA-methylation data and performed bioinformatics data analysis to identify Epigenetic-Aging-Signatures [2]. The predicted age using this technique showed an average difference of about 11 years when compared to the chronological age. Bocklandt and coworkers identified 88 sites in the genome correlated with age in 34 male twin pairs [1]. They validated sites in three genes and described a predictor of age with an average accuracy of 5.2 years. Based on these published data, we decided to study two age-related markers, GRIA2 (glutamate receptor ionotropic AMPA 2) and NPTX2 (neuronal pentraxin II), using quantitative pyrosequencing. Here, we examine a series of CpG sites for application of these loci for use as age-specific epigenetic markers.

2 Materials and methods

2.1 Sample collection

Biological samples were obtained from 44 donors with age ranging from 5 to 72 years. Buccal cells were collected using cotton swabs. Blood was obtained by finger prick and blood was collected using cotton swabs. All participants signed informed consent statements prior to sample collection. All samples were collected after obtaining appropriate Institutional Review Board approvals from Florida International University (IRB-13-0555), The University of Southern Mississippi (protocol #12010303) and Pontifícia Universidade Católica do Rio Grande do Sul (CONEP #723.619/ CEP #845.747).

2.2 DNA extraction and modification

DNA extraction was performed using the BioRobot® EZ1 automated purification workstation (Qiagen Inc., CA) and the EZ1®DNA Investigator kit according to manufacturer's protocol. DNA was recovered in a final volume of 50 μL and quantified using an Alu-based real-time PCR method [16] with a Corbett Rotor-Gene 6000 (Corbett Research, Sydney, Australia, now Qiagen Inc., CA).

Because standard PCR amplification does not preserve DNA methylation sites, genomic DNA extracts were treated with sodium bisulfite in order to convert unmethylated cytosines to uracil using the EpiTect® Fast DNA Bisulfite Kit (Qiagen Inc., CA). The standard manufacturer's protocol for the conversion of 1–500 ng DNA in a maximum volume of 40 μL was used. The converted uracils are then replaced by thymine during the PCR amplification process. The bisulfite-modified DNA was amplified by site-specific PCR primers designed to amplify the bisulfite-modified target regions [17].

2.3 Assay design

Genetic loci GRIA2 and NPTX2 were reported to present age-related methylation levels [1, 2]. In these studies, only one or two CpG sites were analyzed per locus. We decided to investigate these loci further by examining a range of different CpG sites in each genetic locus using bisulfite conversion followed by pyrosequencing. GRIA2 and NPTX2 PyroMark® CpG assays were custom designed in house using the Pyromark assay design software (Qiagen) and targeted three and six CpG sites in the sequence, respectively (Table 1). The reverse primers were biotin labeled allowing for the production of biotinylated PCR products necessary for the pyrosequencing reaction [17].

2.4 PCR and pyrosequencing reactions

Singleplex PCR reactions were performed using the PyroMark® PCR Kit (Qiagen) in a GeneAmp® PCR System 9700 (Applied Biosystems, Foster City, CA). For all samples, 1.2 μL of modified DNA template was added to the PCR mixture in a total reaction volume of 15 μL. The complete PCR protocol outlined by the manufacturer was followed [17].

Pyrosequencing was performed following completion of the PCR reactions using a Pyromark® Q24 pyrosequencer (Qiagen) following protocols recommended by the manufacturer. Once the pyrosequencing is complete, the percent

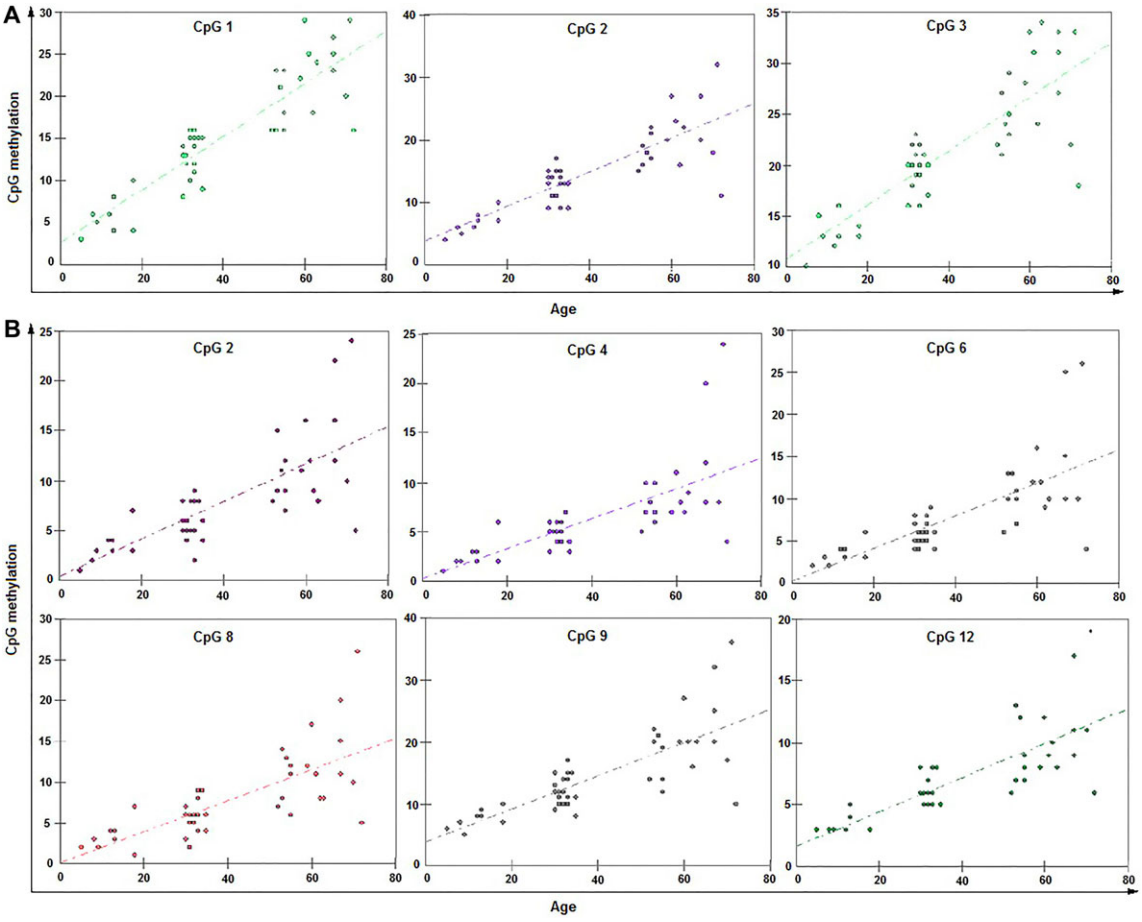

Figure 1. Correlation between CpG site methylation levels and age. (A) Marker GRIA2: CpG 1 ($R^2 = 0.795$), CpG 2 ($R^2 = 0.701$), and CpG 3 ($R^2 = 0.700$). (B) Marker NPTX2: CpG 2 ($R^2 = 0.554$), CpG 4 ($R^2 = 0.485$), CpG 6 ($R^2 = 0.526$), CpG 8 ($R^2 = 0.533$), CpG 9 ($R^2 = 0.596$), and CpG 12 ($R^2 = 0.597$).

methylation was calculated automatically by the Pyromark® Q24 software and was displayed as a pyrogram [17].

2.5 Data analysis

For each locus, the percent methylation values were obtained at each CpG site from the pyrograms. Linear regression analysis was used to calculate a correlation between age and DNA methylation. A single regression analysis was performed to examine each CpG site one at a time. Next, a simultaneous analysis of all tested cytosines was performed using multivariate linear regression.

For the estimation of predicted age, a multivariate regression model was developed with the analysis of the data. A leave-one-out analysis was performed, where the multivariate model was fit on all but one subject and its prediction could be related to the truly observed age of the left-out subject. This analysis is a validation technique to assess the predictive ability of the model when tested in a dataset not used in the estimation [1, 14, 15]. All calculations were done using SPSS statistics v. 17.0. A p-value of <0.05 was assumed as significant for all tests conducted.

3 Results

We examined the methylation level of two age markers, GRIA2 and NPTX2, in saliva samples from 44 donors with ages ranging from 5 to 72 years. Three CpG sites were analyzed for GRIA2 and six out of the 12 sites were analyzed for NPTX2. We considered only six of these sites due to the lack of significant differences in methylation values between individuals of various ages in the remaining CpG sites. The methylation status of all tested cytosines was analyzed and linear regression was used to examine the relationship between age and the methylation profile of the tested loci.

The three CpG sites in GRIA2 locus and the six analyzed CpG sites in NPTX2 locus were hypermethylated with the increase of age (regression coefficients values > 0) and the methylation values of each CpG site presented an age-related association (simple linear regression. For all correlations, $p < 0.0001$) (Fig. 1).

Table 2. Multivariate analysis of CpG sites in GRIA2 and NPTX2 and their association with age

Marker	Number of CpG sites	R^2	Standard error (%)	*p*-value
GRIA2	3	0.801	9.0	0.0001
NPTX2	6	0.654	12	0.0001

After performing linear regression analysis, coefficients of determination (R^2) values were determined. This value is a statistical measure that represents how much of the variability of the dependent variable is explained by the predictor variable. The correlation between methylation values and age was shown to be greater in GRIA2. A multivariate regression analysis showed that the GRIA2 marker represents 80.1% (or $R^2 = 0.801$) of the variance in age. The NPTX2 marker presented a lower percent value of 65.4% (or $R^2 = 0.654$). Detailed results of the multivariate analysis are presented on Table 2. Also, β values showed that CpG 1 ($\beta = 3.13$, $p = 0.0002$) from GRIA2 and CpG 12 ($\beta = 3.53$, $p = 0.03$) from NPTX2 had a stronger correlation with age variation. The beta (β) regression coefficients are important to assess the influence of each predictor variable on the dependent variable in a multiple regression analysis.

3.1 Prediction of age

A multivariate regression model was used to estimate the age of analyzed samples. The epigenetic predicted age, calculated using GRIA2 regression analysis, was strongly correlated to the chronological age ($R^2 = 0.801$, $p < 0.0001$) (Fig. 2A). The estimated values showed an average difference of 6.9 years when compared to the observed ones. When using NPTX2 regression analysis, we observed a lower correlation between predicted and chronological age ($R^2 = 0.654$, $p < 0.0001$) (Fig. 2B). This model shows an average difference of 9.2 years.

3.2 Analysis of blood samples

To analyze if the age-associated DNA methylation results can be reproduced in other tissues, we screened 23 blood samples from the same saliva donors for both GRIA2 and NPTX2. A multivariate regression analysis showed that GRIA2 represents 55.4% (or $R^2 = 0.5542$) of the age variation in blood cells, which is a lower correlation percentage when compared to the saliva sample results. On the other hand, NPTX2 presented a percentage (55.4% or $R^2 = 0.5542$) more similar to the one observed in saliva samples. Also, when trying to apply the same saliva multivariate regression model to predict donor age in blood cells, a low correlation was found between predicted and chronological age for both loci. We could also observe a low correlation when comparing predicted age for blood samples and predicted age for saliva samples (Table 3).

4 Discussion

Aging is a process that occurs in all organisms. It can be considered as a complex phenomenon that is associated with alterations in different organs, tissues, and cells, and changes at the molecular level. The process of aging is characterized by the decline of normal cellular functions and the accumulation of genomic changes, and its affects can be caused by epigenetic modifications in the genome [1, 18]. It is known that epigenetic mechanisms, such as DNA methylation, are important in gene expression and the level of methylated cytosines in the genome is subjected to variation with aging. Several studies have reported the presence of different

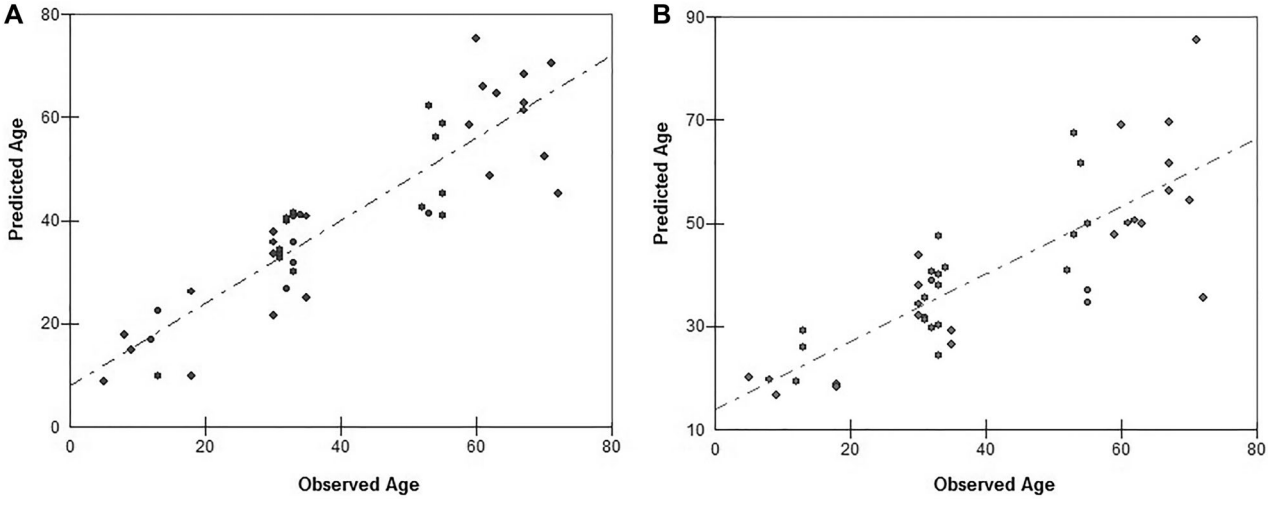

Figure 2. Predicted versus observed age. (A) Results obtained with GRIA2 multivariate regression model ($R^2 = 0.801$). (B) Results obtained with NPTX2 multivariate regression model ($R^2 = 0.654$).

Table 3. Analysis of age-associated DNA methylation in blood cells

Marker	Predicted age (blood samples) versus chronological age	Predicted age (blood samples) versus predicted age (saliva samples)
GRIA2	$R^2 = 0.322$	$R^2 = 0.313$
NPTX2	$R^2 = 0.276$	$R^2 = 0.226$

methylation patterns at CpG islands of various genes in individuals with different ages [1, 10, 14, 15, 18], and have also investigated DNA methylation as a biomarker and a tool for age estimation.

CpG sites in genes GRIA2 and NPTX2 have been associated with aging [1, 2]. Gene GRIA2 (glutamate receptor ionotropic AMPA 2) products function as neurotransmitter receptors in the brain and it has been shown to be associated with Alzheimer's disease [19]. Gene NPTX2 (neuronal pentraxin II) products are involved in synapse formation and it has been associated with Parkinson's disease [20] and pancreatic cancer [21, 22]. Although both genes have been analyzed for their correlation with age, we decided to investigate different CpG sites in GRIA2 and NPTX2 using pyrosequencing as an alternative analysis method. Previous works investigating both genes mostly performed data analysis from array studies. Therefore, applying a methodology that uses bisulfite modification of DNA combined with pyrosequencing is more likely to be useful in a forensic routine.

Observing the results obtained from the multivariate analysis, the beta (β) regression coefficients showed that CpG 1 from GRIA2 and CpG 12 from NPTX2 have the strongest correlation with age. This result is important in assessing the influence of each CpG site in a multivariate analysis. It is also important to consider the analysis of the R^2 values obtained in both simple and multivariate linear regression analysis. When analyzing multiple markers, it is common to combine all data and develop a regression model to predict age based on the methylation levels of all markers. GRIA2 showed a stronger correlation with age variation and prediction when compared to NPTX2 (multivariate regression analysis, GRIA2: $R^2 = 0.801$ versus NPTX2: $R^2 = 0.654$). Also, all CpG sites from GRIA2 produced a higher R^2 value than all CpG sites from NPTX2 when they were analyzed independently (simple regression analysis). Based on these differences, a noncombined analysis of the two loci was performed.

A higher age average difference was observed in the noncombined analysis for each when compared to those found by Bocklandt et al. 2011 [1]. However, it is important to observe that their prediction of age was done using a model based on combined analysis. Surprisingly when comparing results from this study with those obtained by Koch and Wagner [2], our analysis presented a lower age average difference, even though that study also performed a combined analysis with multiple age markers. This latter result highlights the importance of investigating multiple CpG sites in the same locus since a single marker may possess enough prediction power and turn out to be an easy tool for use in age analysis.

Another important question related to age markers is how reproducible the results are when the analysis is done with different cell types. It has been noted that many age-associated DNA methylation changes are tissue-specific [2, 10, 23]. It was observed that a correlation between methylation values and age was found when analyzing both saliva and blood samples. However, the results in saliva were more accurate than those found in blood samples. Also we determined that the age prediction regression model from one group of samples cannot be applied to another. Blood and saliva samples present differences in their composition and therefore differences in their gene expression, and methylation pattern should be expected. Primarily, they can differ in which cell types they contain and their different function, half-life, and viability. The extracted DNA from saliva samples mostly originates from exfoliated epithelial cells, while extracted DNA from blood samples originates from leucocytes [24]. This highlights the need to not only analyze the association between age and methylation, but to also perform cross-correlations between multiple cell types. Knowing the appropriate epigenetic loci for a given body fluid (blood versus saliva) would be particularly critical in forensic applications in which epigenetic age determination might be used in an attempt to identify or discriminate between potential suspects in criminal casework.

Human chronological age can be different from a predicted age group based on a biological sample [15] and this is reflected in the results found in this study. At a biological level, the sample can be affected by many factors, such as genetic and environmental, and the effect of these different variables makes it a challenge to find age biomarkers to predict human aging. Studies of epigenetic age markers can provide the community with new and improved methods to be applied in different investigation areas, whether it is clinical or forensic. The use of this new tool can help in the assessment of health and many clinical conditions, being useful to predict the risk of age-related diseases and to help in cancer treatments, for example. These epigenetic markers may also prove to be an important tool in forensic investigations. The estimation of a suspect's age from recovered DNA can provide important information to police and tryers of fact in situations in which a suspect's identity is unknown.

The authors would like to thank all the volunteers who participated in this study. The authors would like to acknowledge the generous support from Qiagen Inc. with bisulfite modification, PCR, and pyrosequencing and the expertise of all the Qiagen staff, especially Mark Guilliano. Deborah S. B. S. Silva is supported by FAPERGS/CAPES and CNPq (Conselho Nacional de

Desenvolvimento Científico e Tecnológico – modalidade Doutorado Sanduíche). This study was supported by funding provided by the National Institute of Justice, Department of Justice, USA, under award number 2012-D1-BX-K018. Points of view in the document are those of the authors and do not necessarily represent the official view of the U.S. Department of Justice.

The authors declare the following potential conflict of interest: They acknowledge the financial support for this project from Qiagen.

5 References

[1] Bocklandt, S., Lin, W., Sehl, M. E., Sanchez, F. J., Sinsheimer, J. S., Horvath, S., Vilain, E., *PLoS One* 2011, *6*, e14821.

[2] Koch, C. M., Wagner, W., *Aging* 2011, *3*, 1–10.

[3] Goyns, M. H., *Mech. Ageing Dev.* 2002, *123*, 791–799.

[4] Weidner, C. I., Lin, Q., Koch, C. M., Eisele, L., Beier, F., Ziegler, P., Bauerschlag, D. O., Jöckel, K. H., Erbel, R., Mühleisen, T. W., Zenke, M., Brümmendorf, T. H., Wagner, W., *Genome Biol.* 2014, *15*, R24.

[5] Snustad, D. P., Simmons, M. J., *Fundamentos de Genetica*, Guanabara Koogan, Rio de Janeiro 2001.

[6] Alberts, B., Johnson, A., Walter, P., *Biologia Molecular da Célula*, Artmed, Porto Alegre 2004.

[7] Miranda, T. B., Jones, P. A., *J. Cell. Physiol.* 2007, *213*, 384–390.

[8] Frumkin, D., Wasserstrom, A., Budowle, B., Davidson, A., *Forensic Sci. Int. Genet.* 2011, *5*, 517–524.

[9] Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J. B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., Zhang, K., *Mol. Cell.* 2013, *49*, 359–367.

[10] Maegawa, S., Hinkal, G., Kim, H. S., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L. A., Issa, J. P., *Genome Res.* 2010, *20*, 332–340.

[11] Fraga, M., Ballestar, E., Paz, M., Ropero, S., Setien, F., Ballestart, M., Heine-Suner, D., Cigudosa, J. C., Urioste, M., Benitez, J., Boix-Chornet, M., Sanchez-Aguilera, A., Ling, C., Carlsson, E., Poulsen, P., Vaag, A., Stephan, Z., Spector, T. D., Wu, Y. Z., Plass, C., Esteller, M., *Proc. Natl. Acad. Sci. USA* 2005, *102*, 10604–10609.

[12] Boyd-Kirkup, J. D., Green, C. D., Wu, G., Wang, D., Han, J. J., *Epigenomics* 2013, *5*, 205–227.

[13] Teschendorff, A. E., West, J., Beck, S., *Hum. Mol. Genet.* 2013, *22*, R7–R15.

[14] Yi, S. H., Xu, L. C., Mei, K., Yang, R. Z., Huang, D. X., *Forensic Sci. Int. Genet.* 2014, *11*, 117–125.

[15] Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Makowska, Ż., Spas, A., Parys-Proszek, A., Kucharczyk, K., Płoski, R., Branicki, W., *Forensic Sci. Int. Genet.* 2015, *14*, 161–167.

[16] Nicklas, J. A., Buel, E., *J. Forensic Sci.* 2003, *48*, 936–944.

[17] Madi, T., Balamurugan, K., Bombardi, R., Duncan, G., McCord, B., *Electrophoresis* 2012, *33*, 1736–1745.

[18] Thompson, R. F., Atzmon, G., Gheorghe, C., Liang, H. Q., Lowes, C., Greally, J. M., Barzilai, N., *Aging Cell* 2010, *9*, 506–518.

[19] Gaisler-Salomon, I., Kravitz, E., Feiler, Y., Safran, M., Biegon, A., Amariglio, N., Rechavi, G., *Neurobiol. Aging* 2014, *35*, 1785–1791.

[20] Moran, L. B., Hickey, L., Michael, G. J., Derkacs, M., Christian, L. M., Kalaitzakis, M. E., Pearce, R. K., Graeber, M. B., *Acta Neuropathol.* 2008, *115*, 471–478.

[21] Zhang, L., Gao, J., Li, Z., Gong, Y., *Dig. Dis. Sci.* 2012, *57*, 2608–2614.

[22] Yao, F., Sun, M., Dong, M., Jing, F., Chen, B., Xu, H., Wang, S., *Am. J. Med. Sci.* 2013, *346*, 175–180.

[23] Laird, P. W., *Nat. Rev. Genet.* 2010, *11*, 191–203.

[24] Godderis, L., Schouteden, C., Tabish, A., Poels, K., Hoet, P., Baccarelli, A., Van Landuyt, K., *Biomed. Res. Int.* 2015, *2015*, 845041.

