AGE ESTIMATION BASED ON DNA EXTRACTED FROM SEMEN SAMPLES

Ana Maria Macedo Pedro
Master’s Degree in Forensic Genetics
Department of Biology
Faculty of Sciences, University of Porto
October 2022

Supervisor
Benedita Ferreira da Silva, Ph.D, National Institute of Legal Medicine and Forensic Sciences, I.P.

Co-supervisor
Laura Cainé, Ph.D, National Institute of Legal Medicine and Forensic Sciences, I.P.

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

Sworn Statement

I, Ana Maria Macedo Pedro, enrolled in the Master´s Degree in Forensic Genetics at the Faculty of Sciences of the University of Porto hereby declare, in accordance with the provisions of paragraph a) of Article 14 of the Code of Ethical Conduct of the University of Porto, that the content of this dissertation reflects perspectives, research work and my own interpretations at the time of its submission.

By submitting this dissertation, I also declare that it contains the results of my own research work and contributions that have not been previously submitted to this or any other institution.

I further declare that all references to other authors fully comply with the rules of attribution and are referenced in the text by citation and identified in the bibliographic references section. This dissertation does not include any content whose reproduction is protected by copyright laws.

I am aware that the practice of plagiarism and self-plagiarism constitute a form of academic offense.

Ana Maria Mecedo Pedro

October, 2022

Acknowledgements

Gostaria de agradecer, primeiramente, à Professora Doutora Laura Cainé e à Doutora Benedita Silva, pela oportunidade e suporte durante o ano de trabalho do estudo. Agradecer à Mestre Jennifer Fadoni e, novamente, à Doutora Bendita Silva, pelo conhecimento e ajuda no desenvolvimento e aprendizagem das técnicas necessárias em laboratório e conhecimento transmitido para a realização do estudo. Agradecer também a todo o Serviço de Biologia e Genética, do Instituto Nacional de Medicina Legal – Delegação do Norte, pela oportunidade, boa receção e simpatia.

Agradeço aos dadores voluntários e ao *Centro de Estudo e Tratamento da Infertilidade*, pois sem eles o estudo não era possível. Foi um percurso difícil, mas concretizou-se.

Queria agradecer também aos meus amigos por toda a força e ajuda.

Um agradecimento especial à minha família. Aos meus avós pela preocupação e força, aos meus tios e primos pela ajuda em todo o meu percurso, às minhas três irmãs pela paciência e suporte, e aos meus pais que, apesar de estarem longe, me apoiaram incondicionalmente em todas as etapas do meu percurso académico.

Resumo

A estimativa da idade é uma informação relevante para a restringir o espetro de procura de potenciais suspeitos de um crime ou pessoas desaparecidas, apresentando assim um papel importante na resolução de crimes. A estimativa da idade a partir de amostras biológicas tem sido uma técnica maioritariamente usada em amostras a nível do esqueleto, como ossos ou dentes, quando presentes em local de crime. Sendo que grande parte das amostras biológicas encontradas em locais de crime/vítimas são amostras de tecidos ou fluídos corporais, é necessária a expansão deste método para este tipo de amostras biológicas. A procura por biomarcadores relevantes demonstrou que marcadores de metilação de ADN são muito eficientes na estimativa da idade. No entanto, foi demonstrado que células do sémen exibem diferentes padrões de alterações epigenéticas associadas à idade, quando comparadas com células somáticas. Este estudo foca-se, assim, na procura por uma relação entre a idade e os padrões de metilação do ADN e a estimativa da idade de indivíduos, a partir de ADN extraído de amostras de sémen, em três locais CpG diferentes: locais CpG Cg06304190 e Cg12837463 (gene TTC7B) e local CpG Cg06979108 (gene NOX4). Resultados provenientes dos eletroferogramas apresentam correlações fortes entre a idade cronológica dos dadores e os níveis de metilação do ADN. O estudo apresentou diferenças entre idade cronológica e idade prevista e foram obtidos valores de MAE entre 2.96 anos e 4.26 anos e valores de RMSE entre 3.81 anos e 5.09 anos. Estes valores encontram-se abaixo dos valores apresentados pela bibliografia. Estudos futuros incluiriam um maior número de amostras e um intervalo de idades maior.

Palavras-chave:

Ciências forenses, estimativa de idade, metilação de ADN, sémen.

Abstract

Age estimation is a relevant information to narrow the search of potential suspects of a crime or missing persons, thus playing an important role in solving criminal cases. Age estimation through biological samples has been used mostly on skeleton samples, such as bones or teeth, when found on crime scenes. Since most of the biological samples found in crime/victim are tissue or body fluids, it is necessary the extension of this method to this biological samples. The search for relevant biomarkers showed that DNA methylation patterns are very efficient age estimation. However, semen cells have been shown to exhibit different patterns of age-associated epigenetic changes, comparing to somatic cells. Therefore, this study focus on finding a relationship between age and DNA methylation patterns and age estimation, through DNA methylation patterns, from DNA extracted from semen samples, in three different CpG sites: Cg06304190 and Cg12837463 CpG sites (TTC7B gene) and Cg06979108 CpG site (NOX4 gene). Results from electropherograms show strong correlations between donors chronological age and DNA methylation levels. The study showed differences between chronological and predicted age. Results were obtained with MAE values between 2.96 years and 4.26 years and RMSE values between 3.81 and 5.09, both lower age ranges, comparing to bibliography, for the CpG sites analysed, for semen samples. Further studies would include higher number of samples and a higher range of ages.

Keywords:

Forensic science, age estimation, DNA methylation, semen.

List of Contents

List of Tables

List of Figures

List of Abbreviations

A Adenine

˚C Degree Celsius

DNA Deoxyribonucleic Acid

DTT Ditiotreitol

G Guanine

MAE Mean Absolut Error

ng Nanogram

PCR Polymerase Chain Reaction

pg Picogram

RMSE Root Mean Squared Error

Rpm Rotations per minute

SBE Single-base Extension

T Thymine

U Uracil

µM Micromolar

µL Microlitre

Introduction

1. Introduction

Human aging is a complex process that occurs individually, and it can be described as a slow progressive process that occurs in a biological, physiological, environmental, psychological, behavioural, and social manner. The hallmarks of aging were defined by Lopez-Otin et al., 2013, comprising genomic instability, telomere attrition, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, altered intercellular communication and epigenetic alterations.

1.1 Epigenetic Alterations

Epigenetic alterations are stable and heritable, but reversible, variations in the chemical structure of the DNA, that do not alter the DNA coding sequence. These variations were presented to be one of the primary hallmarks, as a cause of damage, involving post-translational modification of histones, chromatin remodelling and alterations in DNA methylation patterns, being the latter the focus of this study. DNA methylation pattern consists of the removal or covalent addition of a methyl group to the 5'-carbon of cytosine in a CpG dinucleotide (Figure 1), being that GCs-rich regions with high densities of CpGs DNA islands become more susceptible to methylation (Lena PD et al, 2019).

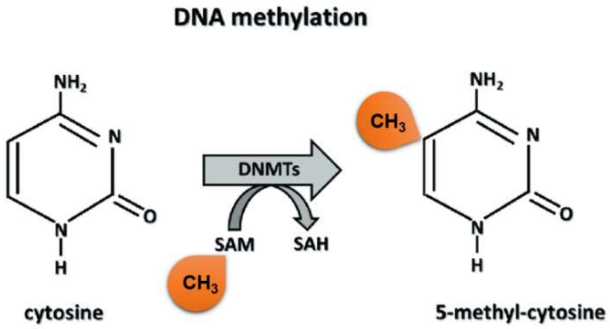

Figure 1 Representation of DNA methylation.

In the past few years, DNA methylation biomarkers have been shown to have a key role in biological age estimation (Lee HY et al, 2016). It has been presented distinct methylomes at the extreme points of human life: lower content in DNA methylation was found in centenarian DNA comparing to newborn DNA (Heyn H et al., 2012).

There are external factors to age that have an influence on DNA methylation variations over long periods of time: ancestry and exposure to environmental and lifestyle factors are proven to alter DNA methylome. Data showed significant differences in DNA methylation patterns due to cigarette smoking (Besingi W et al, 2014), eating habits, air pollution, physical and chemical environmental parameters, recorded diseases, and overall lifestyle (Jung M et al, 2015). Therefore, methylation DNA markers can provide not only age information but lifestyle habits information as well, working as an efficient biomarker in age estimation and environmental exposure information.

1.2 DNA methylation patterns analysis

Several methods for DNA methylation patterns analysis have been reported to be efficient. Thus, in a research study, there is the need of finding the most suitable method between the different assays currently available and it is important that the chosen method can give an unbiased answer to the question of study. However, there are many important factors that must be considered in the process of a DNA methylation patterns method selection.

One of the most important factors to consider is the main goal of the study, this is, whether the research will focus on finding *de novo* epigenetics changes, where there is the need of profiling the whole genome methylation, or the search for differentially methylated regions, being some methods: Luminometric Methylation Assay (LUMA), Restriction Fragment Length Polymorphism (RFLP), Amplified Fragment Length Polymorphism (AFLP), Mass spectrometry based methods, High-Performance Liquid Chromatography-Ultraviolet (HPLC-UV), Enzyme-Linked Immunosorbent Assay (ELISA) based methods, Microarray or Bead Array, and others; Or if the study will focus on known and specific methylation sites of genes of interest, in which the process stars with Bisulfite Conversion and followed by Bead Array, Pyrosequencing, PCR and Sequencing, COLD PCR, and others (Kurdyukov S and Bullock M, 2016)

Other factors to consider when choosing a method include: the availability of specialized equipment and reagents and bioinformatic software for data analysis, which exclude several methods for the lack of means to carry out the process; the robustness of the method, quantity and quality of the DNA samples and cost.

1.3 Forensic sciences relevance

DNA methylation markers became valuable in forensic sciences due to the possibility of substantially reducing the number of potential suspects in a criminal case, by estimating the chronological age of the sample donor (Lee HY et al, 2016). It has been crucial to find methodologies that would be able to give good results from samples collected in crime scenes such as body fluids (vaginal secretions, blood, saliva, and semen).

Bisulfite conversion has been shown to be an effective method to map DNA methylation- specific sequence variants. DNA bisulfite treatment is proven to deaminate unmethylated cytosine to uracil, but not methylated cytosines (Figure 2) (Unnikrishnan A et al, 2019).

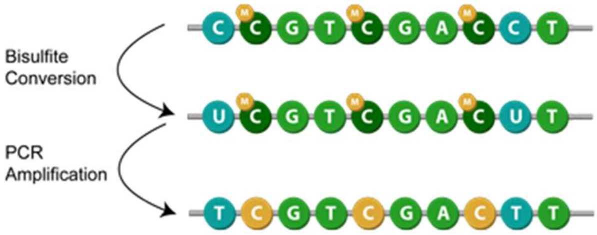

Figure 2 Representation of DNA bisulfite conversion.

Previous studies have shown DNA methylation age-predictors in specific tissues such as saliva (Hong SR et al, 2017 and Bocklandt S et al, 2011) and blood (Garagnani P et al, 2012 and Piekarska RZ et al, 2014), showing that ELOVL2 gene has become the most effective age-predictor marker to date in blood samples. The present study aims to validate an age estimation method based on DNA methylation variations in DNA extracted from semen samples.

1.4 Age estimation using semen samples

Semen has been reported as one of the most forensically relevant body fluids (Lee HY et al, 2015). As Christensen BC et al, 2009 demonstrates, many age- and exposure- related DNA methylation changes rely on tissue types.

Age predictive models based on blood or saliva have been presented to be inaccurate in semen samples. A breakthrough study by Horvath S, 2013, describes a multi-tissue age predictor, evaluating 'age correlation' between DNA methylated age (predicted age) and chronological age, in several tissue types. The study found no significant age correlation in sperm, where DNA methylated age was significantly lower than the chronological age of the donor.

Sperm cells have shown to present different characteristics when compared to somatic cells, such as telomere length: telomere shortening is associated with aging in somatic cells. However, Allsop RC et al, 1992 shows that telomere from sperm DNA does not decrease in length. Data has revealed that sperm cells exhibit opposite age-associated DNA methylation variations, compared to what is commonly observed in somatic cells (Jenkins TG, et al 2018). It is possible to conclude that sperm cells are unique and need an appropriate approach to DNA methylation age estimation.

Previous studies have shown promising CpG sites for DNA age estimation from semen samples: Lee HY et al 2015 (2) selected 3 age-associated CpGs sites for semen samples - cg06304190 in the TTC7B gene, cg12837463, and cg06979108 in the NOX4 gene – with high age-estimation capability with 450K BeadChip array analysis, followed by SNaPshot analysis; Lee et al, 2015 (1) suggested cg17610929 and cg26763284 as semen-specific markers.

The need for a validation method for DNA methylation age estimation in semen samples has become crucial to forensic sciences due to the lack of field development and forensic relevance of this type of sample. Therefore, the main contribution of this work is a method validation for age-estimation based on DNA methylation patterns from semen samples.

Objectives

The relevance of this study stands on the need of a validated method for age prediction for semen samples, as a major contribution to forensic sciences, due to the possibility of narrowing a suspects list in a criminal case.

Therefore, this study aims the construction and validation of an age estimation model, through DNA methylation patterns, for semen samples.

The objectives of this study were:

- Validation method of DNA methylation, through DNA Bisulfite Conversion, with the *Imprint™ DNA Modification* Kit (Sigma-Aldrich®).
- Construct an age-prediction model through multiple linear regressions, in order to estimate individuals age, through DNA extracted from semen samples, with the lowest possible error between chronological age and predicted age.

Material and Methods

2. Material and Methods

2.1 Sample Collection

Semen samples used in this study were obtained in two different ways: (I) provided directly by voluntary donors; (II) obtained through collaboration with *Centro de Estudo e Tratamento da Infertilidade.* A complete set of 39 semen samples was achieved with ages between 21 and 54 years old, being that it was only obtained age relative information about the donors. All samples were collected using sterile cotton swabs, where semen aliquots were deposited, dried, and stored at room temperature, until used.

Sample collection was performed in agreement with the Data Protection Agreement, having been authorized by the donors with a consent statement (Attachment 7.2).

2.2 DNA Extraction

DNA extraction was performed with *PrepFiler Express^{TM} Forensic DNA Extraction Kit,* according to the manufacturer's protocol. The DNA extraction comprises two steps: (i) sample preparation (lysis), with *PrepFiler Express^{TM} Forensic DNA Extraction* Kit; (ii) automatic extraction, performed using *AutoMate Express^{TM}* extraction robot, from A*pplied Biossystems*.

(i) Sample preparation starts with the area decontamination, followed by the lysis solution preparation: 5 µl of DTT (Dithiothreitol) (1M) are added to 500 µl of *PreFilerLysis Buffer*, per sample. The semen sample is transferred by cutting the cotton end of the swab to a *PrepFilerLySep* column and 500 µl of the lysis solution previously prepared is added to the column to cover the sample. The tubes are subsequently incubated in a thermoblock *(Applied Biosystems)* at 70˚C and 750 rpm for 40 min, followed by centrifugation for 2 min at 10000 g. Finally, the tubes and columns are separated, being the latter rejected.

(ii) Automatic extraction comprises only two steps: filling up the cartridge, tips and sample holders; and selecting, on the robot, the kit to be used and elution volume. DNA was extracted with a final volume of 50 µL.

2.3 DNA Quantification

DNA quantification was performed to ensure valid DNA quantity values for further analysis. This step was performed using the *Quantifiler^{TM} Trio DNA Quantification Kit,* according to the manufacturer's protocol. This kit allows the identification of degradation levels and male:female DNA proportions.

Starting on the standard curve, a quantification standard preparation was made with five dilution series as shown in Table 1.

Table 1 Dilution series for the standard curve.

	Volume (µL)
Standard 1	10 µL DNA Standard 10 µL Dilution Buffer
Standard 2	5 µL Standard 1 45 µL Dilution buffer
Standard 3	5 µL Standard 2 45 µL Dilution buffer
Standard 4	5 µL Standard 3 45 µL Dilution buffer
Standard 5	5 µL Standard 4 45 µL Dilution buffer

DNA samples were analysed, as well as two DNA Positive Controls (2,0 ng and 0,1 ng) and one negative Control, with a final volume of 20 µL (Table 2). Quantification was performed with *HID Real Time PCR analysis Software*, on the *7500 Real-Time PCR System* equipment (*Applied Biosystems)*.

Table 2 Master mix preparation for DNA quantification.

	Volume (µL)/ reaction
Quantifiler Master Mix	10 µL
Quantifiler Primer Mix	8 µL
DNA Sample	2 µL
Final Volume	**20 µL**

2.4 DNA Bisulfite Conversion

Bisulfite-converted DNA was obtained by modification of the previously extracted DNA, using the *Imprint™ DNA Modification* Kit (Sigma-Aldrich®), according to the manufacturer's protocol.

Depending on DNA quantity values, the conversion protocol takes place in two different ways: (i) One-Step modification procedure, recommended for a higher DNA input (10 ng to 1 µg); (ii) Two-Step Modification procedure, to be used on lower DNA amounts (100 pg to 10 ng).

The process begins with Reagent Preparation of *Ethanol-diluted Cleaning Solution*, *90% Ethanol Solution* and *Balance/Ethanol Wash Solution*. Followed by One-Step Modification Procedure or Two-Step Modification Procedure and Post-Modification Clean up. Bisulfite-converted DNA was eluted in a final volume of 18 µL.

2.5 Primers Design

PCR primers sequences used for bisulfite-converted DNA amplification and Single-base extension (SBE) for the target CpG sites are presented in Table 3 and Table 4, respectively. Primers sequences were previously described by literature.

Table 3 Primer's sequences for bisulfite-converted DNA amplification for all 3 CpG sites of TTC7B and NOX4 genes.

Target ID	Gene region	Primer ID	Primer Sequence
cg06304190	TTC7B190FOR	Primer1	AATTTTATTTTTGGTATTTAAAGTAG
	TTC7B190REV	Primer2	AAACAAAAACTACCACTCTCACAC
cg12837463	TTC7B463FOR	Primer3	AGTTGGTATTAGGGTTTGAAATGTA
	TTC7B463REV	Primer4	TCTCAAAAACTCTACAATAAAAAAAA
cg06979108	NOX4108FOR	Primer5	TAGTTATTTGAGTGAAGTGTGTTGG
	NOX4108REV	Primer6	ACCTCCCAAAATACTAAATTACTC

Table 4 Primer's sequences for SBE, for all 3 CpG sites of TTC7B and NOX4 genes.

Target ID	Gene region	Primer ID	Primer Sequence
cg06304190	TTC7B190SBERev	Primer7	AATAATCACCTACTATATACTAAAC
cg12837463	TTC7B463SBERev	Primer8	CCTTCTTTAACTCATATACTTTAAAAATATCTAC
cg06979108	NOX4108SBERev	Primer9	TTTTTTTTTTTTTTTTTTTTTTCAATTAAATCCTCAACTAAATC

2.6 DNA Amplification

Bissulfite-Coverted DNA amplification was performed in both Multiplex and Monoplex PCR Reactions.

2.6.1 Multiplex PCR

Multiplex PCR was conducted in 25 µL reactions, each containing volumes listed in Table 5 above.

Table 5 Multiplex amplification reaction of the three CpG sites of TTC7B and NOX 4 genes.

Reagent	Concentration (µM)	Volume (µL) / Reaction
Primer 1	30	1,7
Primer 2		1,7
Primer 3		1,25
Primer 4		1,25
Primer 5		2,15
Primer 6		2,15
H_2O		0,3
QIAGEN Multiplex PCR Master Mix 2x		12,5
Bissulfite-converted DNA		2
Final volume		**25**

Primers sequences are shown in Table 3. Multiplex PCR was conducted in *GeneAmp® PCR System 9700* (*Applied Biosystems)* equipment, under the conditions listed in Table 6.

Table 6 Configuration of multiplex amplification reaction.

Reaction	Temperature (˚C)	Time (Min)	Cycles
Initial Activation	95	15:00	1
Denaturation	94	00:20	40
Annealing	56	1:00	
Extension	72	00:30	
Final Extension	72	7:00	1
Storage	4	∞	1

2.6.2 Monoplex PCR

Monoplex PCR was performed in 25 µL reactions (Table 7). Primers sequences are shown in Table 4.

Table 7 Monoplex amplification reaction of the three CpG sites of TTC7B and NOX genes.

Reagent	Concentration (µM)	Volume (µL) / Reaction
Primer Forward (1, 3 or 5)	30	1,5
Primer Reverse (2, 4 or 6)		1,5
H_2O		7,5
HotStar Taq® Plus Master Mix 2x		12,5
Bissulfite-converted DNA		2
Final volume		**25**

Monoplex PCR was conducted in *GeneAmp® PCR System 9700* (*Applied Biosystems),* under the following conditions shown in Table 8.

Table 8 Configuration of monoplex amplification reaction.

Reaction	Temperature (˚C)	Time (min)	Cycles
Initial Activation	95	5:00	1
Denaturation	94	00:20	40
Annealing	56	1:00	
Extension	72	00:30	
Final Extension	72	7:00	1
Storage	4	∞	1

2.7 Purification

The PCR products were purified with *ExoSAP-IT®(Affimetrix)* by addition of 12,5 µL to each reaction. Each reaction has a final volume of 37,5 µL and purification conditions are described in Table 9.

Table 9 Configuration of purification reaction.

Reaction	Temperature (˚C)	Time (Min)	Holds
Incubation	37	40:00	1
Inactivation	80	20:00	1

2.8 Single-Base Extension

2.8.1 Multiplex SNaPshot Reaction

Single-base extension (SBE) reactions were performed with volumes and reagents described in Table 10, with a final volume of 10 µL.

Table 10 Multiplex SNaPshot reaction of the three CpG sites of TTC7B and NOX4 gene.

Reagent	Volume (µL) / Reaction
SNapShot® Multiplex Ready Reaction Mix	2
PCR Gold Buffer 10X	1,5
Primer 7	1
Primer 8	1
Primer 9	1
H_2O	1,5
Purified PCR product (DNA)	2
Final volume	**10**

The reactions were performed in *GeneAmp® PCR System 9700* (*Applied Biosystems)* equipment, under the conditions described in Table 11.

Table 11 Configuration of multiplex SNaPshot reaction.

Reaction	Temperature (˚C)	Time (Min)	Cycles
Denaturation	95	00:30	35
Annealing	50	00:05	
Extension	60	00:30	
Storage	4	∞	

2.8.2 Monoplex SNaPshot Reaction

Single-base extension (SBE) reactions were performed with volumes and reagents described in Table 12, with a final volume of 10 µL.

The reaction was performed in *GeneAmp® PCR System 9700* (*Applied Biosystems)* equipment and carried out in the same conditions as *Multiplex SNaPshot Reaction*.

Table 12 Monoplex SNaPshot reaction of the three CpG sites of TTC7B and NOX4 gene.

Reagent	**Volume (µL) / Reaction**
SNapShot® Multiplex Ready Reaction Mix	2
PCR Gold Buffer 10X	1,5
Primer 7/8 or 9	1
H_2O	3,5
Purified PCR product (DNA)	2
Final volume	**10**

DNA Positive control, described in Table 13, and Negative control, described in Table 14, were analysed.

Table 13 Monoplex SNaPshot reaction for positive control.

Reagent	**Volume (µL) / Reaction**
SNapShot® Multiplex Ready Reaction Mix	2,5
PCR Gold Buffer 10X	0,5
SNapShot™ Multiplex Control primer	1
H_2O	3
SNapShot™ Multiplex Control DNA	3
Final volume	**10**

Table 14 Monoplex SNaPshot reaction for negative control.

Reagent	**Volume (µL) / Reaction**
SNapShot® Multiplex Ready Reaction Mix	8
H_2O	2
Final volume	**10**

2.9 Purification

After amplification, a purification step is necessary to remove primers and dNTPs not consumed and that may negatively affect the subsequent reactions.

PCR products were purified with *Shrimp Alkaline Phosphatase (SAP) (usb®)* by addition of 2,0 µL to each reaction. Purification conditions are described in Table 15.

Table 15 Configuration of purification reaction.

Reaction	Temperature (˚C)	Time (Min)	Holds
Incubation	37	40:00	1
Inactivation	80	20:00	1

2.10 Sequencing

SNapShot reaction products were submitted to a capillary electrophoresis, for fragment identification labelled with fluorescence, with *HiDi*™ *Formamide (Applied Biosystems)* and *120 LIZ* standard size reagent (Table 16).

Samples were sequenced automatically using the ABI Prism® 3500Genetic Analyzer (Applied Biosystems®) with POP4 polymer (POP-4™ Polymer for 3500 Genetic Analyzers) (Applied Biosystems®).

Table 16 Mixture for capillary electrophoresis.

Reagent	Volume (µL)/ Reaction
HiDi™ Formamide	14
120 LIZ	0,5
DNA	1
Final Volume	**15,5**

2.11 Result analysis

Capillary electrophoresis results were analysed with electropherograms obtained through GeneMapper™ Software (Applied Biosystems®).

2.12 Statistical analysis

After electropherograms were obtained, methylation values for the 3 CpG sites were calculated through peaks height (Section 3.4). Methylation values were then applied to the construction of multiple linear regressions, to obtain the individuals' predicted ages, mentioned below in section 3.7. Multiple linear regressions were constructed using the *Regression* tool, in data analysis, from Microsoft Excel.

.

Results

3. Results

Subsequently, follow the results obtained in the several processes carried out to estimate the age of the individuals through DNA methylation.

From the sample set, comprised of 39 samples, only 23 samples were used in downstream analysis, as 16 samples presented very low quantity of DNA (<1 ng) or high female:male DNA proportion. The set of 23 samples was divided into two different groups: training samples (n=15) and test samples (n=8). Training samples set was used to create the linear regression models used to predict the age of the individuals. Whereas the test samples set was used to validate that same model.

3.1 DNA Quantification

DNA quantification was performed for all samples, to ensure valid values for further application and to be chosen the best modification procedure for each sample. Quantification was performed with *HID Real Time PCR analysis Software,* which allowed DNA quantity, Male:Female ratio and degradation index information. Samples with higher female proportions were excluded and samples with less than 1,00 ng DNA quantity were excluded due to expectation of poor results in the following processes.

Table 17 Quantification results: DNA quantity Male:Female Ratio and Degradation Index for samples set.

Sample	Age (years)	ADN Quantity (ng/µl)	M:F Ratio	Degradation Index	Sample	Age (years)	ADN Quantity (ng/µl)	M:F Ratio	Degradation Index
AP_1	23	48,86		0,83	**AP_19B**	54	1,24		1,05
AP_2	23	131,86		0,95	**AP_20**	30	119,07		1,36
AP_7	33	66,80		0,86	**AP_21**	34	140,08		1,22
AP_8	47	11,63		0,89	**AP_22**	38	152,36		1,17
AP_10	21	14,30		0,77	**AP_23**	35	66,61		1,10
AP_11	21	12,13		0,67	**AP_24**	33	252,86		1,32
AP_12	22	9,47		0,90	**AP_25**	36	48,16		1,06
AP_13	29	16,93		0,70	**AP_26B**	43	27,47		0,94
AP_14B	30	135,44		1,01	**AP_27**	21	37,79		1,08
AP_16B	35	14,28		1,02	**AP_29**	35	327,09		1,14
AP_17	49	21,52		0,81	**AP_30**	48	37,70		0,97
AP_18	52	115,96		1,00					

3.2 Electropherograms

Here are shown electropherograms from the test sample set (Figures 1-8). Training sample set can be observed in Supplementary Materials (Section 7).

These graphics show peaks of all 4 nucleotides: nucleotide G (guanine), represented by blue peaks, related to methylated DNA; nucleotide A (adenine), the green peaks, representing the non-methylated DNA; nucleotide T (thymine) represented by the red peaks and nucleotide C (cytosine), represented by the black peaks.

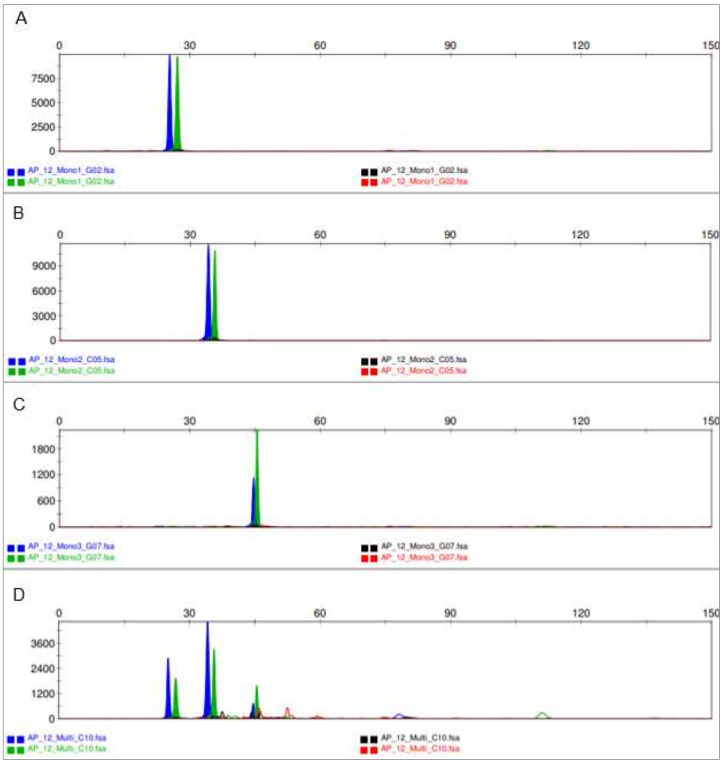

Figure 3 Electropherograms from a semen sample of a 22-year-old individual (sample AP_12). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

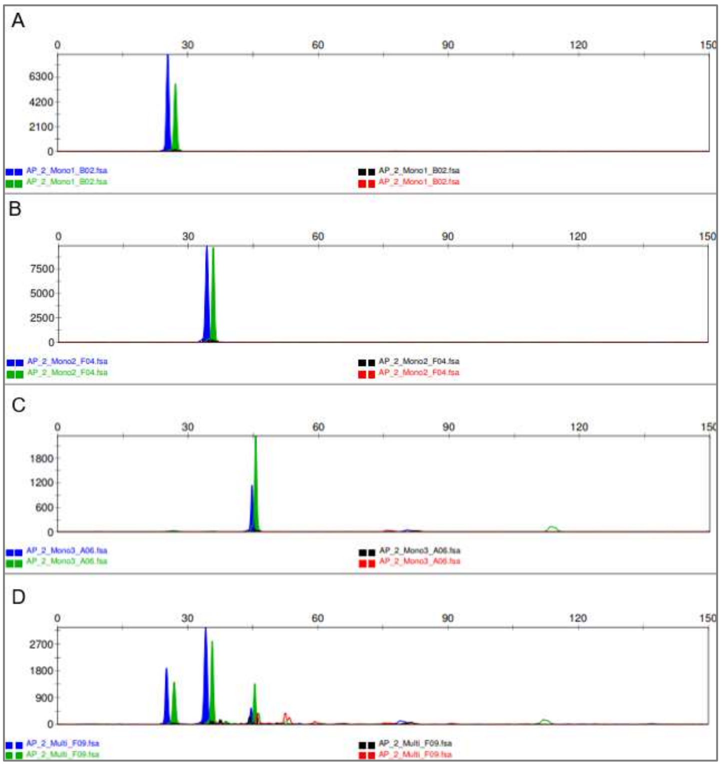

Figure 12 Electropherograms from a semen sample of a 23-year-old individual (sample AP_2). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

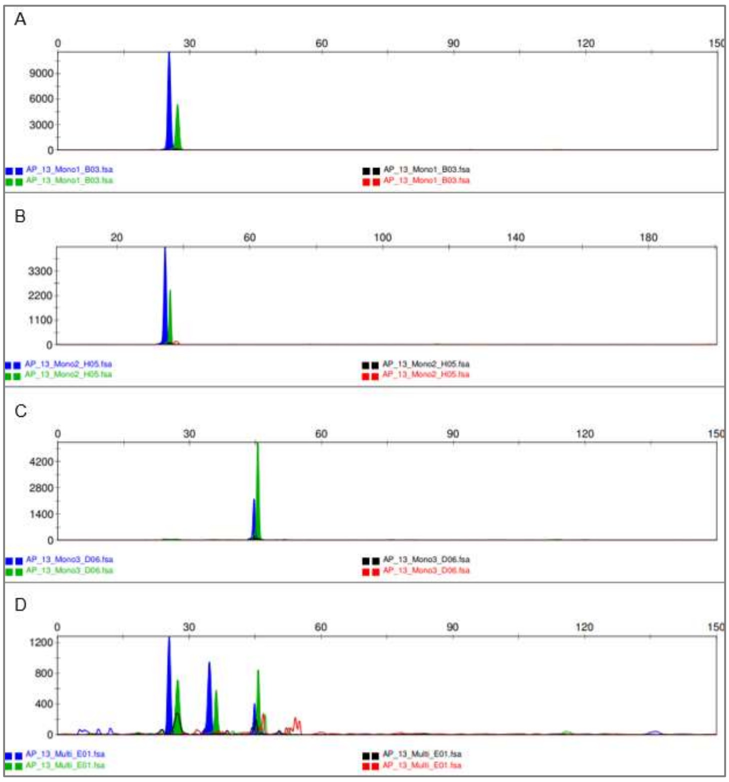

Figure 13 Electropherograms from a semen sample of a 29-year-old individual (sample AP_13). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

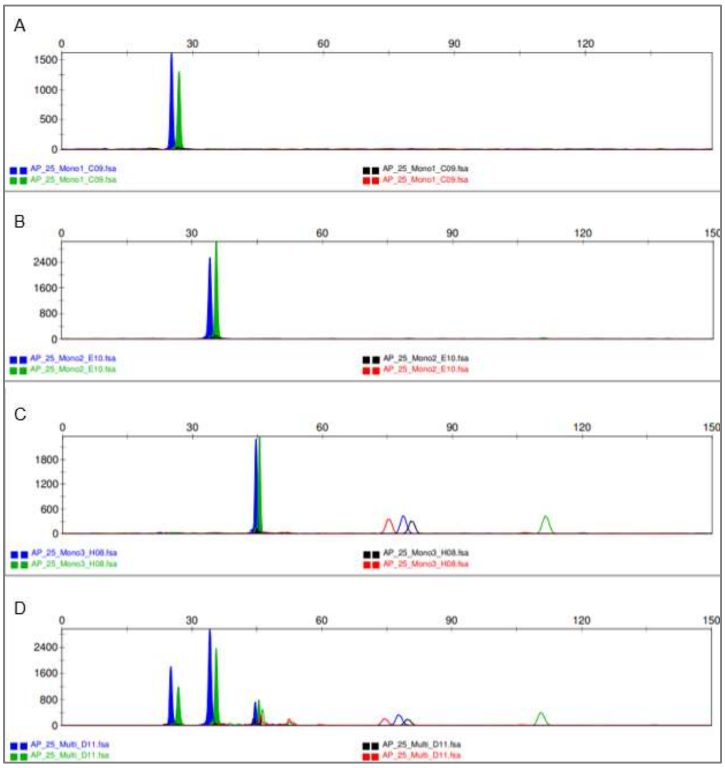

Figure 22 Electropherograms from a semen sample of a 33-year-old individual (sample AP_24). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

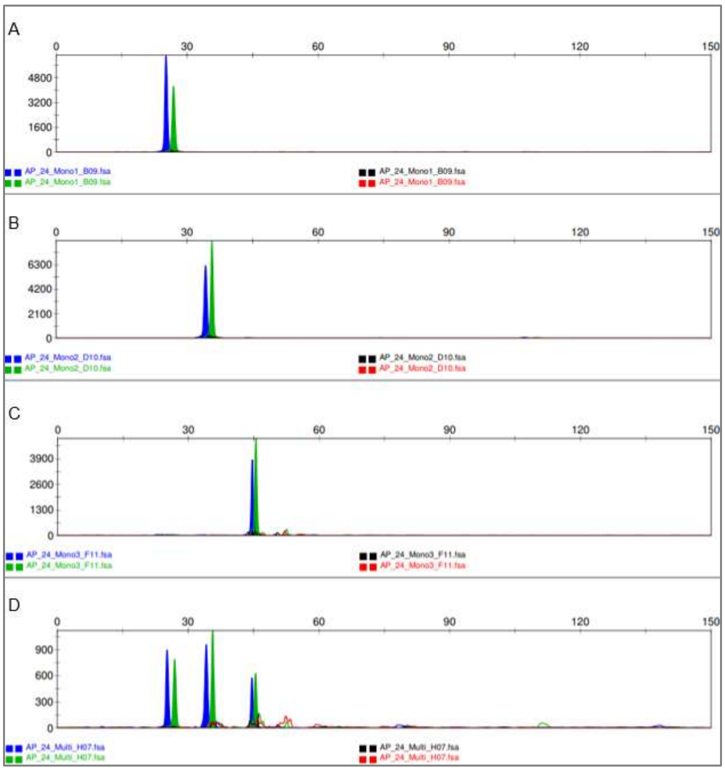

Figure 31 Electropherograms from a semen sample of a 36-year-old individual (sample AP_25). Monoplex reaction for Cg06304190 CpGs site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

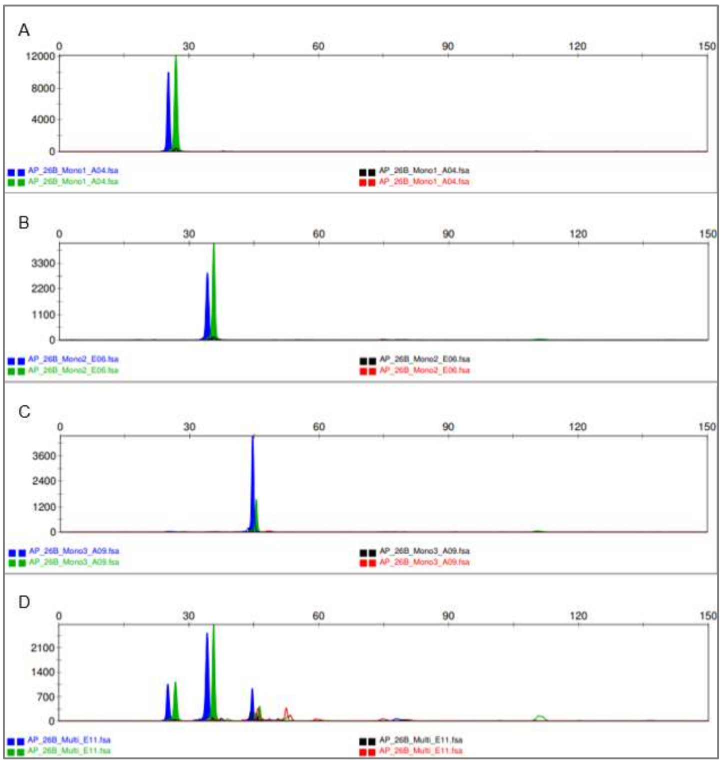

Figure 40 Electropherograms from a semen sample of a 43-year-old individual (sample AP_26B). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

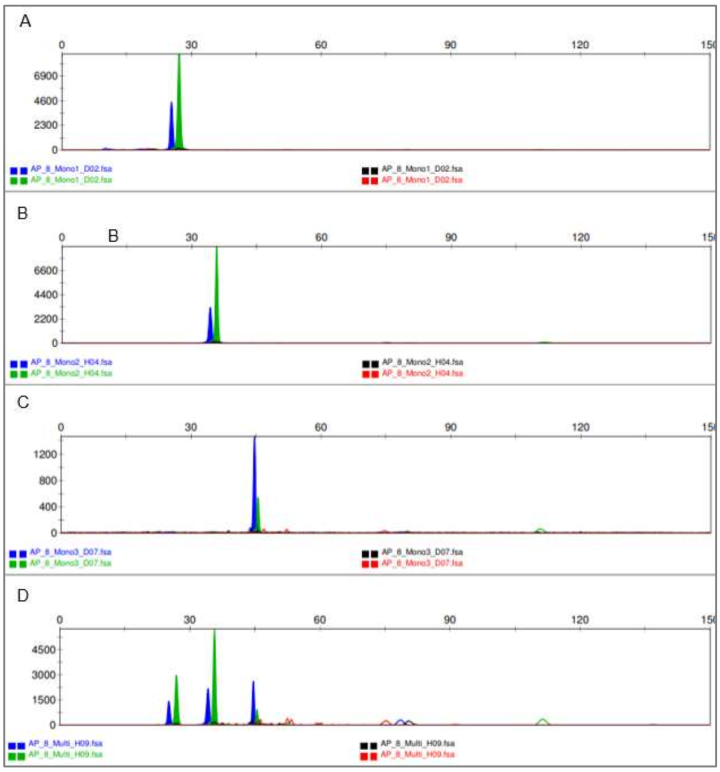

Figure 49 Electropherograms from a semen sample of a 47-year-old individual (sample AP_8). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

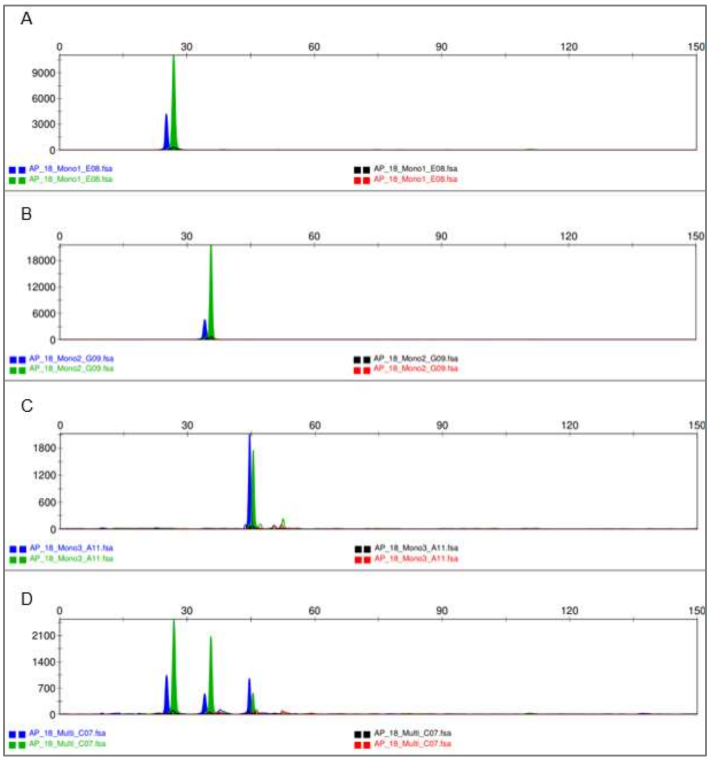

Figure 58 Electropherograms from a semen sample of a 52-year-old individual (sample AP_18). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

3.3 Samples sets

As described above, the complete sample set, with a total of 23 samples, was divided into two sub-sets. Training sample set, for the construction of the multiple linear regression model, was comprised of 15 samples, with donor ages ranging from 21 to 54 years. The test sample set, to validate that same model, to obtain a predictive age model, was comprised of 8 samples, with donor ages ranging from 22 to 52 years.

Below are presented two sets of tables, for multiplex and monoplex reactions separately, with electropherograms information (peak height, size and area) for all 3 CpG sites in TTC7B and NOX4 genes (Tables 18-21).

Training Samples

Table 18 Information from training samples set eletropherograms: Peak size, peak height and peak area, for both methylated and non-methylated DNA, for all 3 CpGs sites, from monoplex reaction.

		Monoplex Reaction																	
		Cg06304190 CpGs site						**Cg12837463 CpG site**						**Cg06979108 CpG site**					
Sample	**Age** (years)	Methylated DNA Nucleotide G			Non-methylated DNA Nucleotide A			Methylated DNA Nucleotide G			Non-methylated DNA Nucleotide A			Methylated DNA Nucleotide G			Non-methylated DNA Nucleotide A		
		Size	Height	Area	Size	Height	Area	Size	Height	Area	Size	Height	Area	Size	Height	Area	Size	Height	Area
AP_1	23	25.37	8141	76444	27.34	6935	65572	34.2	7895	65731	35.67	6111	51281	44.68	1655	12968	45.53	1589	12911
AP_7	33	25.34	9260	84522	27.13	7342	67073	34.29	7380	60532	35.75	10647	87934	44.67	1112	8372	45.45	1382	10686
AP_10	21	25.37	7698	72300	27.21	6707	64898	34.19	7158	59230	35.66	9675	80516	44.71	2140	16513	45.57	3855	29775
AP_11	21	25.36	3301	30613	27.25	2105	19651	34.47	4503	38493	36.00	1488	12826	44.69	1757	13404	45.54	2402	18857
AP_14B	30	25.28	15914	146244	27.14	12067	110476	34.19	4564	39036	35.67	4447	38457	44.66	2696	20312	45.44	1795	13833
AP_16B	35	25.28	7771	71849	27.14	7427	68617	34.28	1270	10431	35.75	1109	9169	44.66	1651	13029	45.51	1074	8671
AP_17	49	25.09	4748	42191	26.87	8222	74348	34.30	2240	18500	35.75	4614	38191	44.64	2145	16314	45.51	676	5444
AP_19B	54	25.26	7869	72369	26.99	2116	19647	34.29	853	6991	35.75	2730	22428	44.69	2610	20684	45.54	2438	19399
AP_20	30	25.34	6378	59356	27.08	12026	114284	34.19	4181	34718	35.67	6272	52601	44.62	2417	18609	45.49	5274	40441
AP_21	34	25.35	3042	28116	27.11	1024	9435	34.18	6984	56338	35.67	4787	38523	44.64	1847	14142	45.51	1024	8307
AP_22	38	25.18	8424	76492	26.94	8107	74493	34.46	5647	46442	35.93	4475	36930	44.63	4612	34802	45.42	2262	17521
AP_23	35	25.34	137	1402	27.07	96	901	34.19	9423	77499	35.67	14353	119897	44.68	3395	25796	45.47	5152	40185
AP_27	21	25.18	10364	92667	26.94	11776	104987	34.18	3885	31460	35.68	2287	18541	44.68	2054	15281	45.47	3634	27247
AP_29	35	25.19	3601	33690	27.16	2376	22722	34.18	4469	36783	35.67	5777	47501	44.63	2472	18636	45.42	3855	29933
AP_30	48	25.17	6917	63917	26.94	8122	76560	34.29	668	5552	35.75	2231	18301	44.65	3729	29018	45.51	1797	14302

Training samples

Table 19 Information from training samples set eletropherograms: Peak size, peak height and peak area, for both methylated and non-methylated DNA, for all 3 CpGs sites, from multiplex reaction.

		Multiplex Reaction																	
		Cg06304190 CpG site						Cg12837463 CpG site						Cg06979108 CpG site					
	Age	Methylated DNA Nucleotide G			Non-methylated DNA Nucleotide A			Methylated DNA Nucleotide G			Non-methylated DNA Nucleotide A			Methylated DNA Nucleotide G			Non-methylated DNA Nucleotide A		
Sample	(years)	Size	Height	Area	Size	Height	Area	Size	Height	Area	Size	Height	Area	Size	Height	Area	Size	Height	Area
AP_1	23	25.18	3006	26808	26.96	3037	27281	34.19	5162	41380	35.68	4085	33591	44.68	1354	10389	45.47	1782	14216
AP_7	33	25.18	2904	25922	26.97	2648	24063	34.19	3729	30998	35.67	5262	43459	44.63	1700	13038	45.42	2524	19876
AP_10	21	25.18	3657	32798	26.87	2939	26434	34.18	4909	40057	35.67	4799	40245	44.63	1322	10164	45.42	2507	19653
AP_11	21	25.45	353	3392	27.34	343	3252	34.58	116	1107	36.07	82	1313	44.75	448	3640	45.67	568	5109
AP_14B	30	25.09	2763	24884	26.89	1563	14108	34.17	4139	33400	35.67	3278	28179	44.63	1599	12130	45.42	1176	9420
AP_16B	35	25.36	555	5449	27.38	507	4869	34.57	494	4559	36.08	1012	9103	44.82	452	3686	45.66	305	2501
AP_17	49	25.34	1028	9644	27.07	1822	16981	34.31	1040	9154	35.74	1941	17268	44.74	1109	8752	45.59	435	3399
AP_19B	54	25.18	3296	29282	26.97	483	4650	34.19	1649	14267	35.67	3449	29408	44.71	2351	29282	45.5	1150	8953
AP_20	30	25.18	928	8582	26.87	1813	16620	34.18	2072	18123	35.67	2765	23657	44.63	620	4877	45.42	1214	9774
AP_21	34	25.18	2478	21934	26.89	935	8442	34.28	3737	30218	35.68	2770	22203	44.68	1169	8783	45.47	695	5435
AP_22	38	25.37	543	4950	27.23	474	4853	34.56	1271	10956	36.09	703	5812	44.77	275	2205	46.31	324	3553
AP_23	35	25.18	2302	20551	26.95	1401	12869	34.19	1699	14607	35.67	2798	23398	44.7	903	7032	45.49	1203	9407
AP_27	21	25.28	2530	23629	27.23	2437	22925	34.46	2465	21235	36.01	1179	9876	44.77	1061	8297	45.62	1999	16082
AP_29	35	25.09	1542	13697	26.89	1211	11007	34.17	2981	24637	35.67	3193	26483	44.64	572	4456	45.43	853	6655
AP_30	48	25.28	2283	21358	27.23	2847	26434	34.56	1795	15356	36.01	4029	33995	44.77	1591	12498	45.62	834	7996

Test Samples

Table 20 Information from test samples set eletropherograms: Peak size, peak height and peak area, for both methylated and non-methylated DNA, for all 3 CpGs sites, from monoplex reaction.

Monoplex Reaction

		Cg06304190 CpGs site						Cg12837463 CpG site						Cg06979108 CpG site					
		Methylated DNA Nucleotide G			Non-methylated DNA Nucleotide A			Methylated DNA Nucleotide G			Non-methylated DNA Nucleotide A			Methylated DNA Nucleotide G			Non-methylated DNA Nucleotide A		
Sample	**Age** (years)	Size	Height	Area	Size	Height	Area	Size	Height	Area	Size	Height	Area	Size	Height	Area	Size	Height	Area
AP_2	23	25.42	8201	75101	27.14	5724	51552	34.30	9881	80391	35.75	9712	78785	44.68	1145	9053	45.53	2349	19222
AP_8	47	25.34	4484	41296	27.11	8943	84922	34.29	3233	26400	35.75	8748	70757	44.66	1466	11382	45.44	542	4359
AP_12	22	25.34	9921	91941	27.13	9711	90491	34.19	11689	94703	35.67	10907	88118	44.67	1143	8693	45.54	2227	17647
AP_13	29	25.28	11517	105277	27.24	5400	48978	34.48	4379	37050	36.00	2449	20341	44.65	2185	16869	45.51	5208	40717
AP_18	52	25.17	4216	38134	26.85	11040	101643	34.19	4721	37691	35.67	21513	173444	44.57	2117	16722	45.44	1754	14031
AP_24	33	25.18	6248	56229	26.87	4241	38284	34.18	6226	50860	35.67	8360	67513	44.63	3856	29359	45.50	4920	38070
AP_25	36	25.17	1617	14364	26.85	1308	11834	34.18	2538	20956	35.59	3030	25072	44.62	2295	17509	45.49	2354	18015
AP_26B	43	25.26	9986	91288	27.02	12106	112001	34.19	2894	23784	35.67	4158	34420	44.63	4535	34664	45.42	1551	12224

Table 21 Information from test samples eletropherograms: Peak size, peak height and peak area, for both Methylated and non-methylated DNA, for all 3 CpGs sites, from multiplex reaction.

Multiplex Reaction

		Cg06304190 CpGs site						Cg12837463 CpG site						Cg06979108 CpG site					
		Methylated DNA Nucleotide G			Non-methylated DNA Nucleotide A			Methylated DNA Nucleotide G			Non-methylated DNA Nucleotide A			Methylated DNA Nucleotide G			Non-methylated DNA Nucleotide A		
Sample	**Age** (years)	Size	Height	Area	Size	Height	Area	Size	Height	Area	Size	Height	Area	Size	Height	Area	Size	Height	Area
AP_2	23	25.09	1891	16937	26.88	1422	12853	34.17	3254	26458	35.68	2801	23241	44.61	554	4355	45.48	1365	10684
AP_8	47	25.18	1424	12851	26.87	2973	26165	34.18	2182	17912	35.67	5732	46621	44.63	2621	19901	45.5	925	7389
AP_12	22	25.09	2896	25647	26.88	1938	17538	34.17	4632	36858	35.68	3326	27932	44.61	742	5742	45.48	1589	12414
AP_13	29	25.45	1273	11738	27.38	709	9006	34.57	942	8755	36.16	579	5434	44.91	408	3293	45.74	841	7278
AP_18	52	25.17	1047	9583	26.92	2536	23036	34.1	549	4676	35.59	2076	17546	44.62	957	7363	45.49	571	4337
AP_24	33	25.17	898	8315	26.91	791	7085	34.21	958	7995	35.66	1118	9397	44.63	576	4718	45.48	628	6187
AP_25	36	25.09	1803	16505	26.87	1188	11072	34.19	2958	24883	35.59	2379	20780	44.60	720	5723	45.47	785	6304
AP_26B	43	25.18	1064	9896	26.89	1129	10304	34.28	2531	21394	35.67	2763	23563	44.63	933	7468	45.50	255	1976

3.4 Methylation values

DNA methylation values were obtained through the intensity of the peaks from the electropherograms. Peak height values of nucleotide G (methylated DNA) and nucleotide A (non-methylated DNA) were used for DNA methylation results, using the following formula: *Methylation = nucleotide G/ (nucleotide G + nucleotide A)*

Table 22 Methylation values from training sample set.

Training set							
		Multiplex			Monoplex		
Sample	Age (years)	TTC7B Gene cg06304190 CpG site	TTC7B Gene cg12837463 CpG site	NOX4 Gene cg06979108 CpG site	TTC7B Gene cg12837463 CpG site	TTC7B Gene cg12837463 CpG site	NOX4 Gene cg06979108 CpG site
AP_10	21	0,554426925	0,505665431	0,345259859	0,534397779	0,425236143	0,356964137
AP_11	21	0,507183908	0,585858586	0,440944882	0,610617832	0,751627441	0,422457321
AP_27	21	0,509361788	0,676454446	0,346732026	0,468112014	0,629455606	0,361111111
AP_1	23	0,497435049	0,558235103	0,431760204	0,539997347	0,563686991	0,510172626
AP_14B	30	0,638696255	0,558042335	0,576216216	0,568743076	0,506492065	0,600311735
AP_20	30	0,338562568	0,428364689	0,338058888	0,346555097	0,399980867	0,314263425
AP_7	33	0,523054755	0,414748081	0,402462121	0,557764125	0,409385921	0,445870088
AP_21	34	0,726047466	0,574304595	0,627145923	0,748155435	0,593322572	0,64332985
AP_23	35	0,621658115	0,377807427	0,428774929	0,511317279	0,533837747	0,60587156
AP_29	35	0,560116237	0,482831228	0,401403509	0,587982833	0,396324024	0,397215397
AP_16B	35	0,52259887	0,328021248	0,597093791	0,602476159	0,436170213	0,390706496
AP_22	38	0,533923304	0,643870314	0,459098497	0,509588047	0,557893697	0,670933954
AP_30	48	0,44502924	0,308207418	0,656082474	0,459937496	0,230424284	0,674809989
AP_17	49	0,360701754	0,348876216	0,718264249	0,366075559	0,326816458	0,760368664
AP_19	54	0,87218841	0,32346018	0,671522422	0,788082123	0,238068658	0,51703645

Table 23 Methylation values from test sample set.

Test set							
		Multiplex Reaction			Monoplex Reaction		
Sample	Age (years)	TTC7B Gene cg06304190 CpG site	TTC7B Gene cg12837463 CpG site	NOX4 Gene cg06979108 CpG site	TTC7B Gene cg06304190 CpG site	TTC7B Gene cg12837463 CpG site	NOX4 Gene cg06979108 CpG site
AP_12	22	0,599089781	0,582055793	0,318318318	0,505348411	0,517303948	0,339169139
AP_2	23	0,570781769	0,537407102	0,288692027	0,588940754	0,504312765	0,327704637
AP_13	29	0,642280525	0,619329389	0,326661329	0,680794467	0,641329818	0,295549844
AP_24	33	0,531675548	0,461464355	0,478405316	0,595671656	0,426847662	0,600311735
AP_25	36	0,602808425	0,554243957	0,478405316	0,552820513	0,455818966	0,493654549
AP_26B	43	0,485180119	0,478088402	0,785353535	0,45201883	0,410380034	0,74515281
AP_8	47	0,323857175	0,275713925	0,739142696	0,333953973	0,26984392	0,730079681
AP_18	52	0,292213229	0,209142857	0,626308901	0,276350288	0,179957307	0,546887109

3.5 Methylation vs Chronological age correlation

In order to obtain an easier interpretation and visualization of the DNA methylation values results, scatter plots were made, showing methylation and chronological age correlation, for both multiplex and monoplex reactions. Relationships between DNA methylation and chronological age were only tested for the test sample set. (Table 23).

Multiplex Reaction

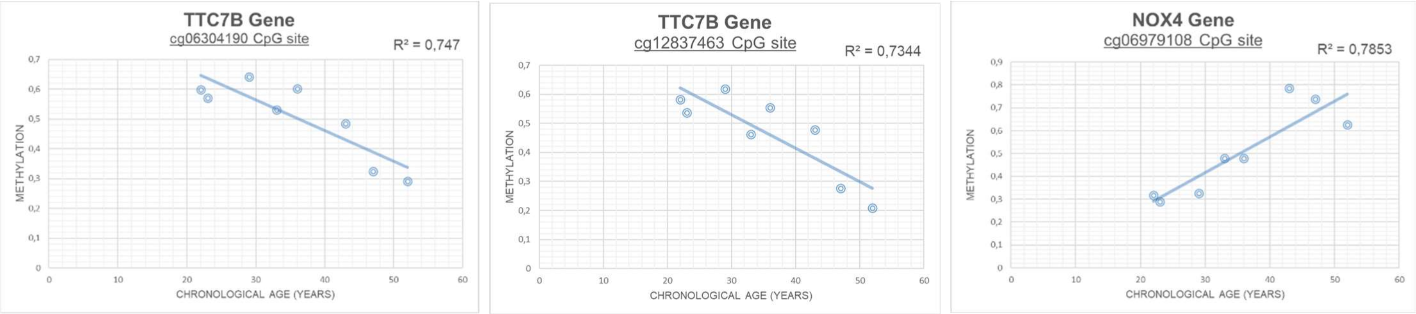

Figure 67 Correlation between methylation values and chronological age (years), for each of the three CpG sites analysed, from multiplex reaction.

Monoplex Reaction

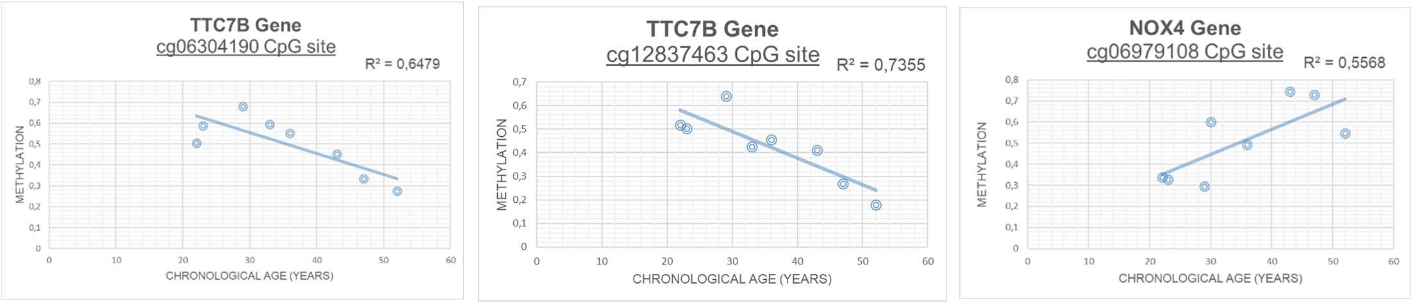

Figure 76 Correlation between methylation values and chronological age (years), for each of the three CpG sites analysed, from monoplex reaction.

A tendency and a linear pattern are clearly observed in all scatter plots. For both multiplex and monoplex reactions, the two CpG sites of TTC7B gene show greater DNA methylation values in minor age individuals and much lower DNA methylation values in elderly individuals. Therefore, for TTC7B gene, DNA methylation appears to decrease as the individual age increases, which demonstrates a negative linear correlation between DNA methylation and chronological age.

On the other hand, the reverse is observed in the CpG site located in the NOX4 gene. In this region, DNA methylation values tend to be lower in younger individuals, increasing as the individual ages. This relationship shows a positive correlation between DNA methylation and chronological age, for the NOX4 gene.

All scatter plots show moderately strong relationships between DNA methylation values and chronological age, as the data (observations) follow the line, not being too much spread out.

For both TTC7B and NOX4 genes, the correlations present a higher R-squared in multiplex reaction, demonstrating a stronger relationship between the two variables on display in multiplex reaction rather than in monoplex reactions.

3.6 Multiple linear regression models

Two multiple linear regression models were constructed separately, for both monoplex and multiplex reactions, using the training sample sets, with the purpose of finding an age estimation model for all samples.

Table 24 Multiple linear regression model statistics of the 3 CpG sites at the TTC7B and NOX4 genes for multiplex reaction, obtained from the training sample set.

	Coefficient	P-value	R R-multiple	R^2 R-squared	SE Standard error	n
(Intercept)	26,65963	0,062152	0,853681	0,728771	6,082403	15
TTC7B Gene cg06304190 CpG site	6,646801	0,618234				
TTC7B Gene cg12837463 CpG site	-36,6613	0,037827				
NOX4 Gene cg06979108 CpG site	42,11428	0,017516				

Table 25 Multiple linear regression model statistics of the 3 CpG sites at the TTC7B and NOX4 genes for monoplex reaction, obtained from the training sample set.

	Coefficient	P-value	R R-multiple	R^2 R-squared	SE Standard error	n
(Intercept)	29,4676195	0,005502	0,918338757	0,843346073	4,62250967	15
TTC7B Gene cg06304190 CpG site	16,62365106	0,140135				
TTC7B Gene cg12837463 CpG site	-49,85863805	0,000139				
NOX4 Gene cg06979108 CpG site	36,18909365	0,001971				

3.7 Predicted Age

Age predictive models for semen samples were constructed separately, for both multiplex and monoplex reactions, using the 15 training set samples. This age predictive models were then validated with the 8 test set samples. The multiple linear regression models obtained with the training sample set allowed the age estimation of the individuals for both training and test sample sets (Table 26 and 27). Predicted age was obtained according to the following formula:

$Y = b_0 + b_1 \times CpG_1 + b_2 \times CpG_2 + \ldots + b_p \times CpG_p$

- y= Predictive age (Dependent variable)
- b_0 = y - intercept (Constant)
- b_1, b_2, …, b_p = Methylation values (Table 23)
- CpG_1, CpG_2, …, CpG_p = Coefficient values for each CpG site (Table 24-25)

Multiplex Reaction

Table 26 Predicted age (years), Absolute Error (AE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) for both training and test sample sets, for multiplex reaction.

Multiplex Reaction											
Training set						**Test set**					
Sample	**Chronological age** (years)	**Predicted age** (years)	**AE** Absolute error	**MAE** Mean Absolute error	**RMSE** Root Mean Squared Error	**Sample**	**Chronological age** (years)	**Predicted age** (years)	**AE** Absolute error	**MAE** Mean Absolute Error	**RMSE** Root Mean Squared Error
AP_10	21	26,3468	5,3468	**4,2617**	**5,2086**	**AP_12**	22	22,7085	0,7085	**2,9626**	**3,8110**
AP_11	21	27,1225	6,1225			**AP_2**	23	22,9095	0,0905		
AP_27	21	19,8479	1,1521			**AP_13**	29	21,9804	7,0196		
AP_1	23	27,6836	4,6836			**AP_24**	33	33,4234	0,4234		
AP_14B	30	34,7133	4,7133			**AP_25**	36	30,4948	5,5052		
AP_20	30	27,4427	2,5573			**AP_26B**	43	45,4318	2,4318		
AP_7	33	31,8805	1,1195			**AP_8**	47	49,8327	2,8327		
AP_21	34	36,8426	2,8426			**AP_18**	52	47,3110	4,6890		
AP_23	35	34,9983	0,0017								
AP_29	35	29,5862	5,4138								
AP_16B	35	43,2537	8,2537								
AP_22	38	25,9380	12,0620								
AP_30	48	45,9488	2,0512								
AP_17	49	46,5161	2,4839								
AP_19	54	48,8791	5,1209								

Chronological Age – Predicted Age Correlation

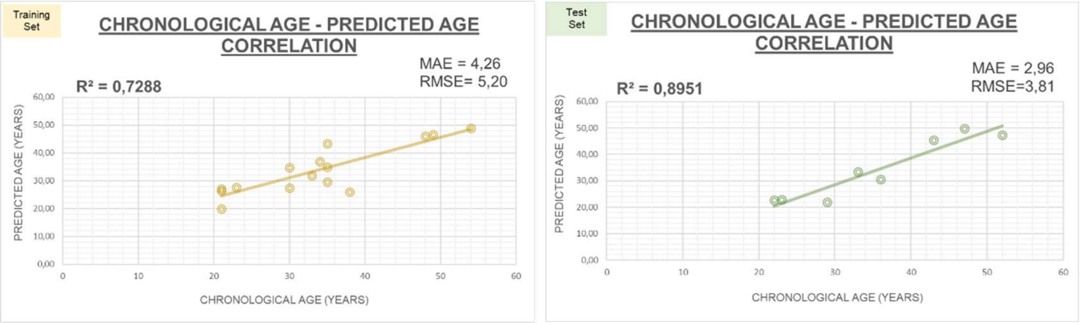

Figure 77 Correlation between Chronological age and predicted age for both Training and Test samples, for the 3 CpG sites all together, of TTC7B and NOX4 genes, for multiplex reaction.

For both Training samples and Test samples, scatter plots above show a linear data pattern, presenting a positive correlation between chronological and predicted age.

Both graphics present a small Mean Absolute Error (MAE) and high R-squared, showing a well-fitted regression model for the data, as there are no significant differences between chronological and predicted age.

Although both data set present good results, the test sample data appears to demonstrate a stronger relationship between chronological and predictive age, with a higher R-squared and smaller MAE.

Monoplex Reaction

Table 27 Predicted age (years), Absolute Error (AE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), for both training and test sample sets, for monoplex reaction.

Monoplex Reaction											
Training set						**Test set**					
Sample	**Chronological age** (years)	**Predicted age** (years)	**AE** Absolute error	**MAE** Mean Absolute error	**RMSE** Root Mean Squared Error	**Sample**	**Chronological age** (years)	**Predicted age** (years)	**AE** Absolute error	**MAE** Mean Absolute Error	**RMSE** Root Mean Squared Error
AP_10	21	30,0678	9,0678	**3,3837**	**3,9585**	**AP_12**	22	24,3505	2,3505	**4,0535**	**5,0952**
AP_11	21	17,4315	3,5685			**AP_2**	23	25,9730	2,9730		
AP_27	21	18,9338	2,0662			**AP_13**	29	19,5048	9,4952		
AP_1	23	28,8024	5,8024			**AP_24**	33	39,8126	6,8126		
AP_14B	30	35,3939	5,3939			**AP_25**	36	33,7959	2,2041		
AP_20	30	26,6590	3,3410			**AP_26B**	43	43,4872	0,4872		
AP_7	33	34,4639	1,4639			**AP_8**	47	47,9860	0,9860		
AP_21	34	35,6040	1,6040			**AP_18**	52	44,8805	7,1195		
AP_23	35	33,2771	1,7229								
AP_29	35	33,8567	1,1433								
AP_16B	35	31,8754	3,1246								
AP_22	38	34,4035	3,5965								
AP_30	48	50,0456	2,0456								
AP_17	49	46,7756	2,2244								
AP_19	54	49,4097	4,5903								

Chronological Age – Predicted Age Correlation

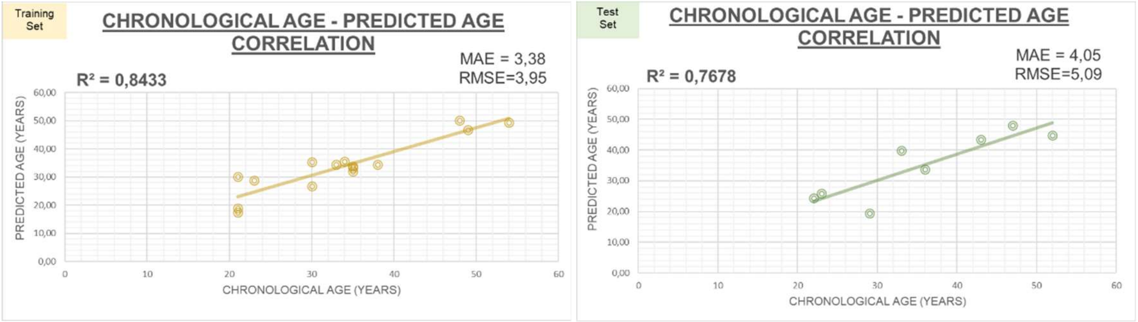

Figure 78 Correlation between Chronological age and predicted age for both Training and Test samples, for the 3 CpG sites all together, of TTC7B and NOX4 genes, for monoplex rection.

For monoplex reactions, the same positive correlation is observed. For both Training sample and Test sample sets, the scatterplots show a well-fitted model for the data, presenting a high R-Squared and a small MAE. Test sample set shows a smaller R-squared and a higher MAE, comparing to the Training sample set.

Discussion

4. Discussion

4.1 Samples collection, data quantity and quality

Samples were collected directly from volunteers and through collaboration with *Centro de Estudo e Tratamento da Infertilidade.* It was only obtained age-related information from the donors. As previously referred, DNA methylation patterns can be influenced by other than age-related factors, such as environmental exposure, lifestyle habits and ancestry. Therefore, this study does not consider these factors in the individual's age-estimation, from the DNA methylation patterns, which can be assumed as a limitation.

Research work has always two crucial data points: quality and quantity. A major struggle found in this study was the difficulty in achieving a good number of samples. The need for a sample set with a large range of ages was essential, and the collaboration with this infertility center allowed to obtain more samples and a wider range of ages.

Due to a low number of volunteers and lack of time for the development of the study, only a total of 39 samples were collected, aged between 21 and 54 years. However, 16 samples were excluded due to very low DNA quantity (<1ng). This DNA quantity limit was defined to exclude samples from further analysis due to the poor final results encountered from the first set of working samples and due to the samples number limitations from the utilized kits. Therefore, this study was conducted only with a total of 23 samples.

Despite the low number of samples, this study achieved good quality samples - DNA quantity values ranged from 1.24 ng/µl to 327.09 ng/µl - and strong and concrete results, comparing to bibliography (Section 6).

4.2 DNA Methylation Patterns Analysis

As mentioned above, choosing the right method for DNA methylation analysis is important for the course of the research. Factors such has choosing between whole genome methylation analysis or analyse specific regions in genes of interest, equipment, reagents and software availability, quantity and quality of the data and cost, should be considering in the process of the selection of a method for DNA methylation patterns analysis.

For this study, the selected method of Bisulfite Conversion followed by PCR and SNaPshot Sequencing was based, primarily, on its reported efficiency, cost and due to equipment availability in the National Institute of Legal Medicine and Forensic Sciences - North delegation. Until recently, this method was the only method to present methylation levels at specific CpG sites, and it is still a very efficient and successful method, as shown by the good results obtained.

The process was carried out for both monoplex and multiplex reactions, for comparison purposes. From the results described in section 3, better results were achieved regarding data from multiplex reactions. It is known that, in comparison to multiplex reaction, monoplex reaction has more risks associated. There are higher risks of contamination during samples handling and higher associated costs, due to a larger number of reactions to process, as the reactions for the 3 CpG sites occur individually for each sample. Regarding multiplex reactions, reactions for the 3 CpG sites occur in a single reaction tube, minimizing costs and contamination risks.

Only three CpG sites were considered and selected for studying DNA methylation patterns from semen samples in this research, due to the lack of development in research for semen samples, as this 3 CpG sites were the only ones reported with capacity for age estimation specifically from semen samples. A larger number of CpG sites for DNA methylation analysis would raise the accuracy and precision of the results.

4.3 DNA Bisulfite Conversion

DNA Bisulfite Conversion is a critical point. Results from the bisulfite conversion were only observed during electropherogram analysis.

From the 23 samples set, only sample 19B was converted with the two-step modification procedure, with 1.24 ng/µl of DNA quantity, value achieved only in a second bisulfite conversion of the sample (B).

A successful DNA bisulfite conversion is observed in electropherograms due to variations between methylated (blue peaks) and non-methylated DNA (green peaks), according to donors age.

4.4 Electropherograms

Electropherogram analysis was an important point for this study. One of the main goals was obtaining electropherograms with clean and high acceptable peaks, corresponding to methylated DNA (blue peaks) and non-methylated DNA (green peaks).

As described above, these graphics show peaks height, area, and size and peaks colour is representative of each nucleotide. In this study was only necessary the detection of the nucleotide G, represented by the blue colour, presenting the methylated DNA, and the nucleotide A, represented by the green colour, presenting the non-methylated DNA, as only the reverse primer was used for SBE.

Peaks height and area correspond to the signal intensity, which, in this case, demonstrates a higher or lower methylated and non-methylated DNA quantity per sample. Peaks size position indicates the size of each CpG site analysed.

All three CpG sites have different sizes: in the TTC7B gene, Cg06304190 CpG site presents a size value around 25 for the nucleotide G (methylated DNA- blue peaks), and a size value around 27 for the nucleotide A (non-methylated DNA – green peaks); Cg12837463 CpG site presents a size value around 34 for the nucleotide G (methylated DNA – blue peaks) and 35 for the nucleotide A (non-methylated DNA – green peaks); and in the NOX4 gene, Cg06979108 CpG site presents a size value around 44 for the nucleotide G (methylated DNA – blue peaks) and 45 for the nucleotide A (non-methylated DNA - green peaks).

It is possible to observe that electropherograms show cleaner peaks for monoplex reactions, comparing to multiplex reactions, for all different samples. In a general observation, it is possible to see higher intensities -peaks height- in monoplex reactions comparing to multiplex reactions, which suggests that the combination of these different primers in the same reaction may negatively affect the amplification process.

However, for both reactions, it is possible to see an overall correlation between donors age and methylated/non-methylated DNA peaks intensities, for the 3 CpG sites analysed. Higher blue peaks are observed in minor aged individuals, comparing to green peaks height, for both CpG sites in the TTC7B gene, and higher green peaks and lower blue peaks in older individuals. On the other hand, for the Cg06979108 site, in the NOX4 gene, the reverse is observed.

This shows that a relationship between methylated DNA and age is achieved with electropherograms analysis alone.

4.5 Methylation values and age correlation

Peaks height from the electropherograms were transformed in methylation values, as described above in section 3.4, which allowed an efficient manner to achieve correlations between individuals age and methylation levels. Simple linear regressions were made in order to achieve those correlations with scatter plots.

As observed above, in section 3.5, scatter plots show moderately strong correlations between individuals age and methylation values, for both multiplex and monoplex reactions. The two CpG sites located in the CCT7B gene show negative linear correlations, in which methylation levels decrease as the individual ages. For the CpG site located in the NOX4 gene, a positive linear correlation is observed, that is the methylation level increases as the individual ages. These correlations between methylation and chronological age were expected, for these same CpG sites, according to Lee HY, 20215 (2).

Although having moderately strong relationships in both reactions, a stronger relationship was achieved in multiple reactions data. For the Cg06304190 and Cg12837463 CpG sites (TTC7B gene) and Cg06979108 CpG site (NOX4 gene), R-squared values of 0.747; 0.734 and 0.785 were achieved, respectively, for multiplex reaction, and 0.648; 0.735 and 0.557, for monoplex reaction.

Both reactions presented stronger correlations between methylation levels and chronological ages, comparing to Lee HY, 2015 (2), in which R-squared values of 0.6315; 0.555 and 0.525 were obtained, for Cg06304190 and Cg12837463 CpG sites (TTC7B gene) and Cg06979108 CpG site (NOX4 gene), respectively.

4.6 Age estimation model

As described above in section 3.7, two independent age estimation models - multiple linear regression models - were obtained, constructed with methylation values of the training sample set (n=15), and afterwards validated with the test sample set (n=8), for both multiplex and monoplex reactions.

Samples were not evenly divided into two sets due to the need of a wider number of samples for the training sample set in order to construct a predictive regression model. Age estimation was achieved through the models obtained.

Multiple linear regression models are extensions of the simple linear regression models, by adding variables (Eberly, L. E., 2007), in which, in this case, methylation values are used from the three CpG sites simultaneously, to an outcome. It comprehends a simultaneous statistical relationship between the continuous outcome Y (age), dependent variable, and the predictor independent variables (methylation values).

Multiple linear regression model for multiplex reaction with the 15 training set samples (Table 24) explained 72.8% of the total age variance (R^2= 0.728) and presented a strong correlation between predicted and chronological ages (R=0.853; R^2=0.728) (Table 24 and Figure 13), with a MAE of 4.26 years and a RMSE of 5.2 years. The trained model with the 8 test samples set presented an even stronger correlation between predicted and chronological ages (R^2=0.895) with a MAE of 2.96 years and RMSE of 3.81 years.

Multiple linear regression model for monoplex reaction with the 15 training set samples (Table 25) explained 84.3% of the total age variance (R^2= 0.843) and presented a strong correlation between predicted and chronological ages (R=0.918; R^2=0.843) (Table 25; Figure 14), with a MAE of 3.38 years and a RMSE of 3.95 years. The trained model with the 8 test samples set also presented a strong correlation between predicted and chronological ages (R^2=0.767) with a MAE of 4.05 years and RMSE of 5.09 years.

For the training sample sets, results showed a higher correlation (R^2) between predicted and chronological ages and lower MAE and RMSE values for monoplex reactions, comparing to multiplex reaction age prediction model.

Results showed higher correlation (R^2) between predicted and chronological ages and lower MAE and RMSE values for multiplex reaction age estimation model, comparing to monoplex reaction age estimation model, for the test sample sets. Differences between multiplex and monoplex reactions are higher in the test samples set age prediction model, probably due to a lower number of samples on display.

For these exact same CpG sites, located in TTC7B and NOX4 genes, Lee HY et al, 2015 presented R-squared values of 0.814 (n=31) and 0.804 (n=68), for the training and test sample models, respectively. Lee HY et al, 2015 (2) and Lee JW et al, 2018 showed RMSE values above 5.8 years. Although having higher R-squared values for the age estimation models, due to a much higher number of samples per set, the present study presented higher prediction accuracy comparing RMSE results.

4.6 Study relevance and future perspectives

The relevance of this study is based on its contribution to forensic sciences. The search for a well-established, validated method for age estimation through DNA samples from semen has risen, since it is one of the most relevant body fluids found in crime scenes. The possibility to obtain an age range from a crime scene sample, through an accurate age estimation model, would decrease a possible suspect list in criminal cases.

As described above, one critical point for higher accuracy is data quantity. Unfortunately, a lower number of samples was obtained for this study. Further studies would include a much higher number of samples with a wider range of ages and a higher number of CpG sites for DNA methylation patterns analysis in order to obtain even more robust results. As this research does not consider external factors to age that have reported influence in methylation patterns, future studies would also include external factors such as ancestry and lifestyle habits, in the age estimation of the individuals.

Despite that, this study achieved good, solid and better results comparing to bibliography.

Conclusions

5. Conclusions

The present study aimed the age estimation of individuals through DNA methylation patterns from DNA extracted from semen samples. Three CpG sites were analysed: Cg06304190 and Cg12837463CpG sites, located in the TTC7B gene and Cg06979108 CpG site, located in the NOX4 gene.

The main conclusions of this study were:

- Relationship obtained between methylated/non-methylated ADN and chronological age, with the electropherograms analysis alone;

- Regarding the correlation between methylation values and chronological age, R-Squared values of 0.74; 0.73 and 0.78, were obtained, respectively, for multiplex reactions, and 0.648; 0.73 and 0.55, for monoplex reactions, having been achieved a stronger relationship in the results obtained from the multiplex reactions data;

- Results showed higher correlation (R^2), between predicted and chronological age, in the prediction models for multiplex reactions;

- Differences between chronological and predicted age were obtained with MAE values between 2.96 years and 4.26 years and RMSE values between 3.81 and 5.09, both lower age ranges, comparing to bibliography, for the CpG sites analysed, for semen samples.

- Values of MAE and RMSE were lower in the age prediction models for multiplex reactions;

- Further studies would include higher number of samples with a higher range of ages and a higher number o CpG sites for DNA methylation patterns analysis.

6. References

Allsopp RC, Vaziri H, Pattersn C, Goldstein S, Younglai EV, Futcher AB, Greider CW, Harley CB, Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci USA Vol.89, pp.10114-10118 (1992).

Besingi W and Johansson A, Smoke-related DNA methylation changes in the etiology of human disease, Human Molecular Genetics Vol.23 No.9 (2014).

Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, Vilain E, 'Epigenetic Predictor of Age', PloS ONE 6(6) (2011).

Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, Sugarbaker DJ, Yeh R, Wiencke JK, Kelsey KT, 'Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context', PloS Genetics Vol. 5 (2009).

Du P, Zhang X, Huang C, Jafari N, Kibbe WA, Hou L, Lin SM, Comparison of Beta-value and Mvalue methods for quantifying methylation levels by microarray analysis, BMC Bioinformatics 11:587 (2010).

Eberly, L. E. , 'Multiple Linear Regression. Methods in Molecular Biology™', 165–187 (2007).

Fernandez AF, Fraga MF, Heath SC, Valencia A, Gut IG, Wang J, Esteller M, 'Distinct DNA methylomes of newborns and centenarians', PNAS Vol.109 no.26, pp.10522-10527 (2012).

Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Blasio AM, Gentilini D, Vitale G, Collino S, Rezzi S, Castellani G, Capri M, Salvioli S, Franceschi C, 'Methylation of ELOVL2 gene as a new epigenetic marker of age', Aging Cell 11, pp. 1132-1134 (2012).

Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-Mut JV, Setien F, Carmona FJ, Puca AA, Sayols S, Pujana MA, Serra-Musach J, Iglesias-Platas I, Formiga F, Steve Horvath, 'DNA methylation age of human tissues and cell types', Genome Biology 14:R115 (2013).

Hong SR, Jung S, Lee EH, Shin K, Yang WI, Lee HY, 'DNA methylation-based age prediction from saliva: High age predictability by combination of 7 CpG markers', Forensic Science International: Genetics 29, pp. 118-125 (2017).

Jenkins TG, Aston KI, Cairns B, Smith A, Carrell DT, 'Paternal germ line aging: DNA methylation age prediction from human sperm', BMC Genomics 19:763 (2018).

Jung M and Pfeifer GP, Aging and DNA methylation, BMC Biology 13:7 (2015).

Kurdyukov S and Bullock M, 'DNA Methylation Analysis: Choosing the right method', *Biology*, 2016.

Lee HY, An JH, Jung S, Oh YN, Lee EY, Choi A, Yang WI, Shin K, Genome-wide methylation profiling and a multiplex construction for the identification of body fluids using epigenetic markers, Forensic Science International: Genetics 17, pp.17-24 (2015) (1).

Lee HY, Jung S, Oh YN, Choi A, Yang WI, Shin K, 'Epigenetic age signatures in the forensically relevant body fluid of semen: a preliminary study', Forensic Science International: Genetics 19, pp. 28-34 (2015) (2).

Lee HY, Lee SD, Shin K, ‘Forensic DNA methylation profiling from evidence material for investigative leads’, BMB Reports 49(7) pp. 359-369 (2016).

Lee JW, Choung CM, Jung JY, Lee HY, Lim S, ‘A validation study of DNA methylation-based age prediction using semen in forensic casework samples’, Legal Medicine 31, pp. 74-77 (2018).

Lena PD, Sala C, Prodi A, Nardini C, ‘Missing value estimation methods for DNA methylation data’, Bioinformatics 35(19), pp. 3786–3793 (2019).

López-Ótin C, Blasco MA, Partridge L, Serrano M, Kroemer G, ‘The Hallmarks of Aging’, Cell 153 (2013).

Piekarska RZ, Spolnicka M, Kupiec T, Makowska Z, Spas A, Parys-Proszek A, Kucharczyk K, Ploski R, Branicki W, ‘Examination of DNA methylation status of the ELOVL2 marker may be useful for human age prediction in forensic science’, Forensic Science International: Genetics (2014).

Steve Horvath, DNA methylation age of human tissues and cell types, Genome Biology 14:R115 (2013).

Unnikrishnan A, Freeman WM, Jackson J, Wren JD, Porter H, Richardson A, The role of DNA methylation in epigenetics of aging, Pharmacology & Therapeutics 195, pp. 172-185 (2019).

Attachments

7. Attachments

7.1 Attachment 1: Electropherograms from the training sample set

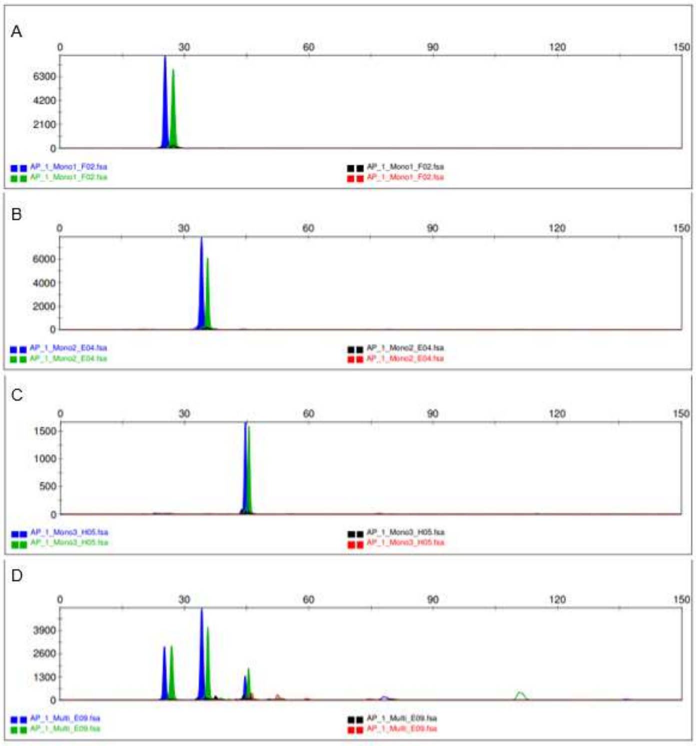

Figure 1 Electropherograms from a semen sample of a 23-year-old individual (sample AP_1). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3CpG sites (D).

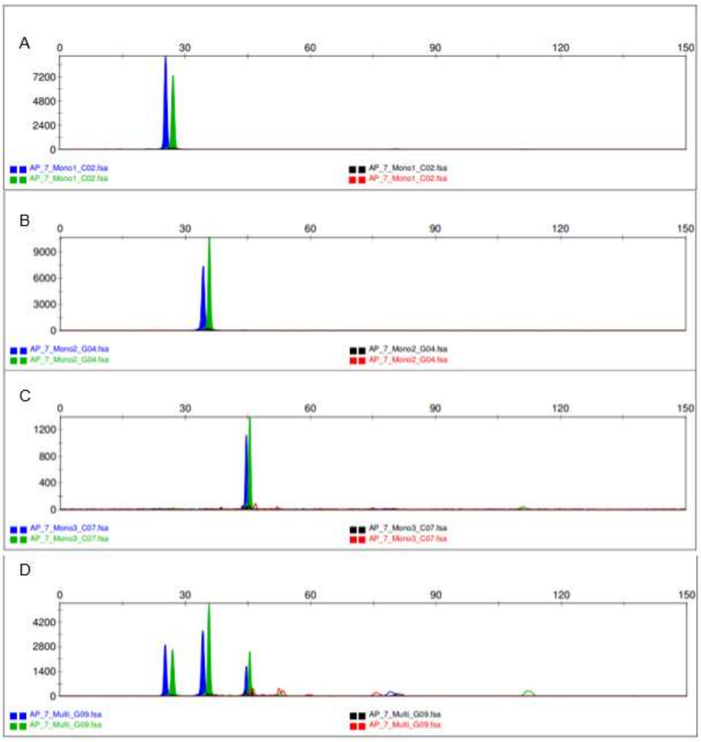

Figure 2 Electropherograms from a semen sample of a 33-year-old individual (sample AP_7). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

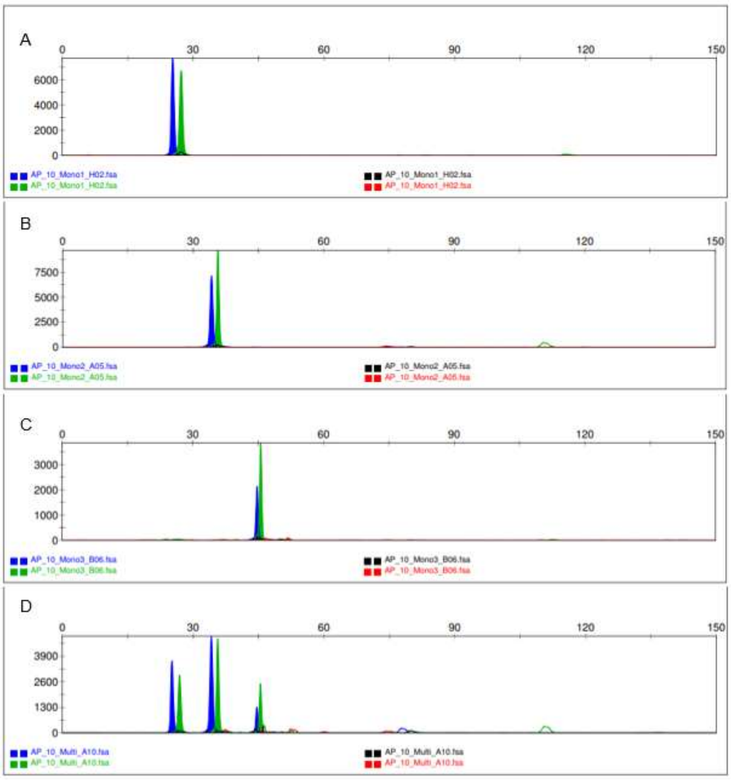

Figure 395 Electropherograms from a semen sample of a 21-year-old individual (sample AP_10). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

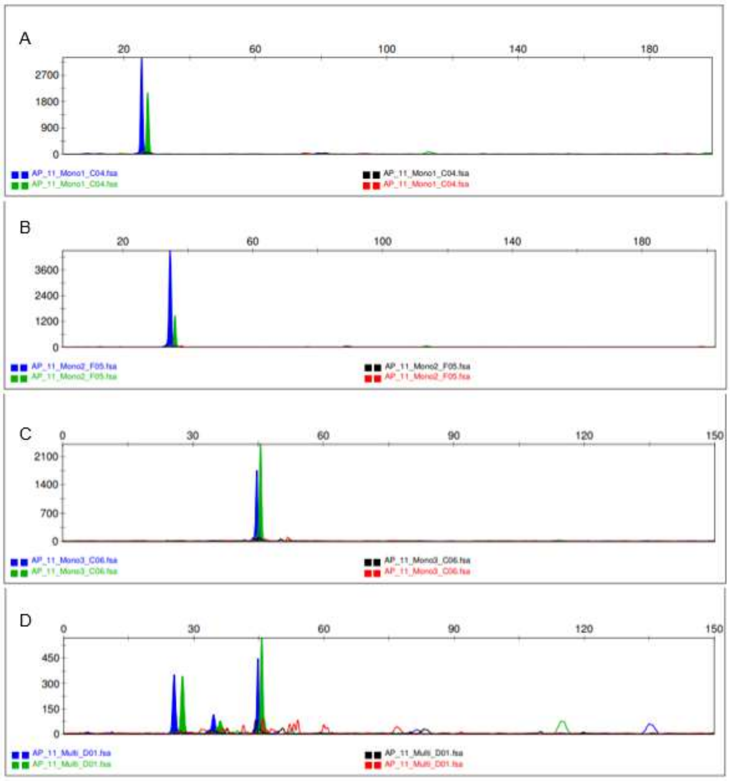

Figure 4104 Electropherograms from a semen sample of a 21-year-old individual (sample AP_11). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

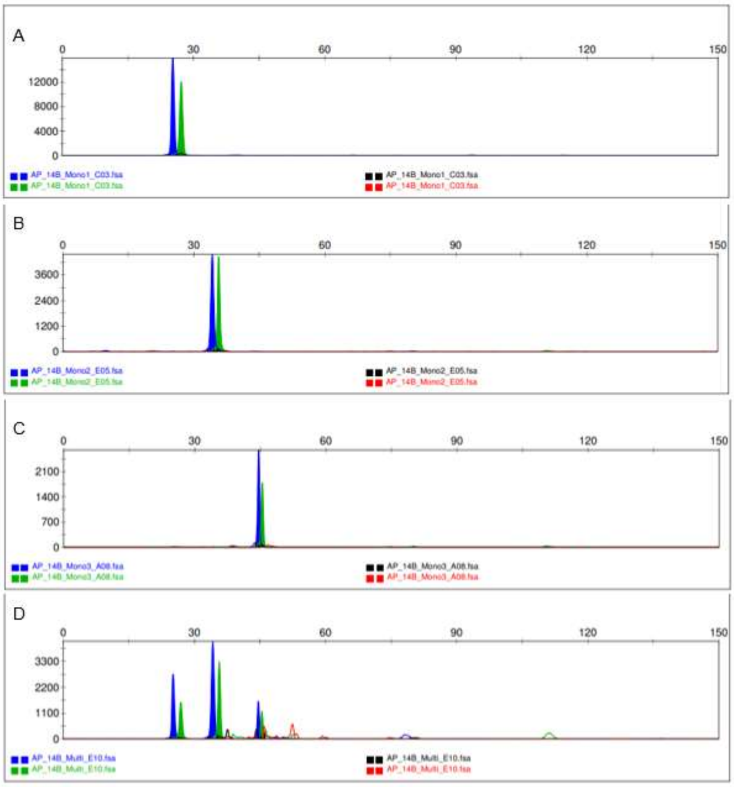

Figure 113 Electropherograms from a semen sample of a 30-year-old individual (sample AP_14B). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

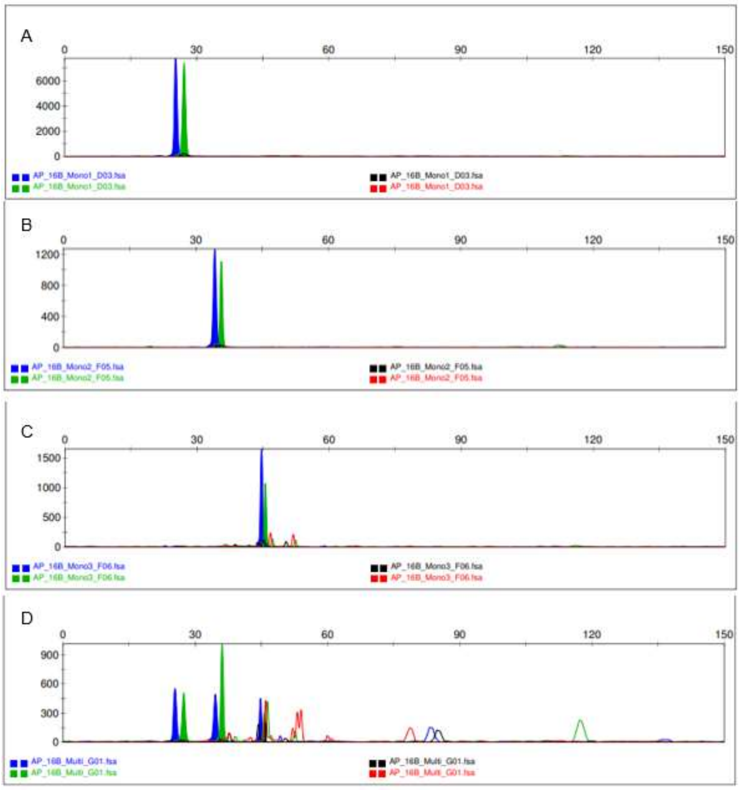

Figure 6 Electropherograms from a semen sample of a 35-year-old individual (sample AP_16B). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

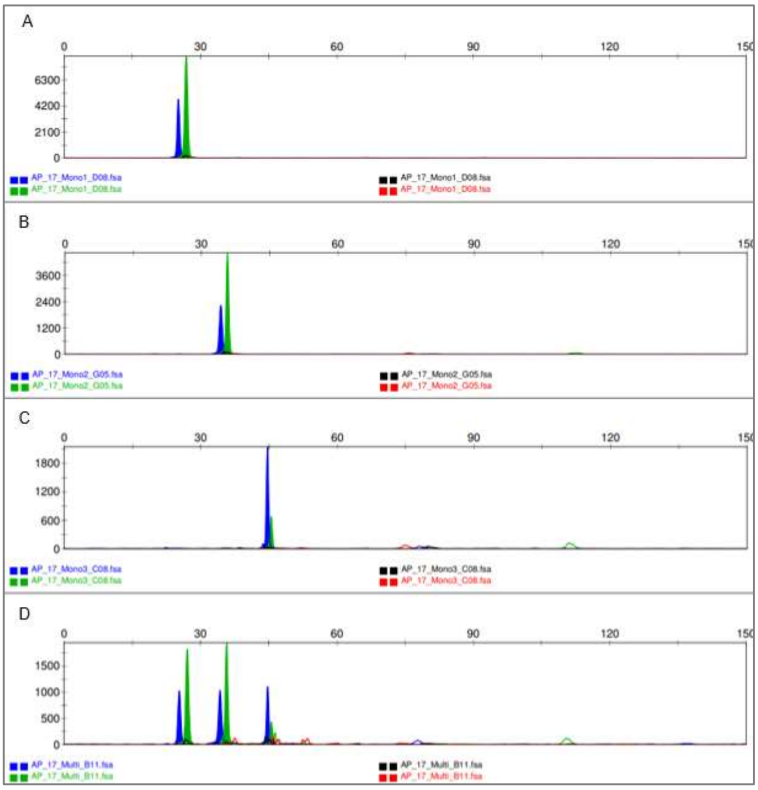

Figure 7130 Electropherograms from a semen sample of a 49-year-old individual (sample AP_17). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

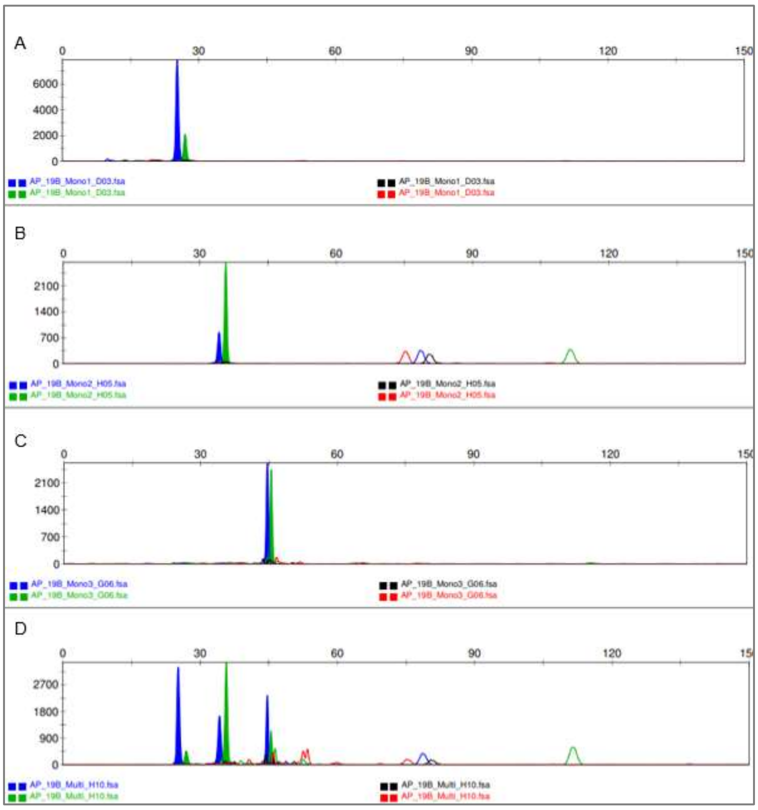

Figure 8 Electropherograms from a semen sample of a 54-year-old individual (sample AP_19B). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

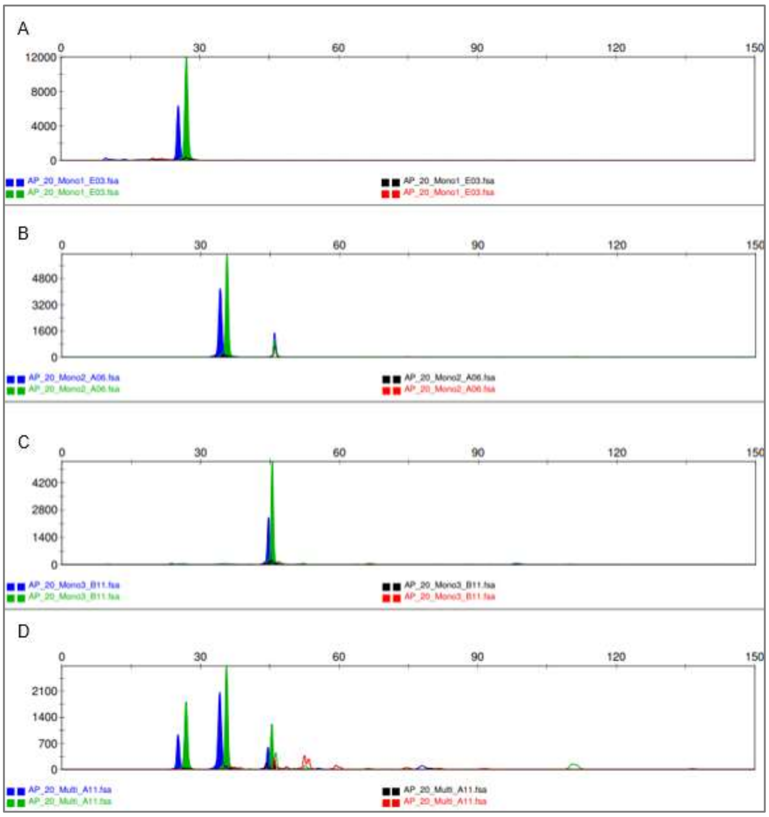

Figure 9147 Electropherograms from a semen sample of a 30-year-old individual (sample AP_20). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

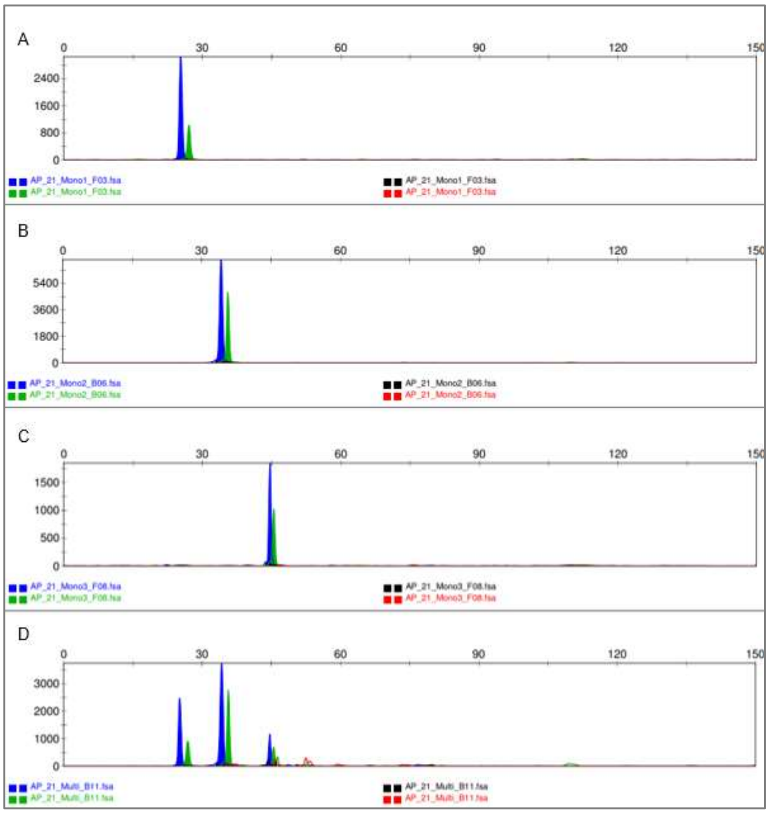

Figure 156 Electropherograms from a semen sample of a 34-year-old individual (sample AP_21). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

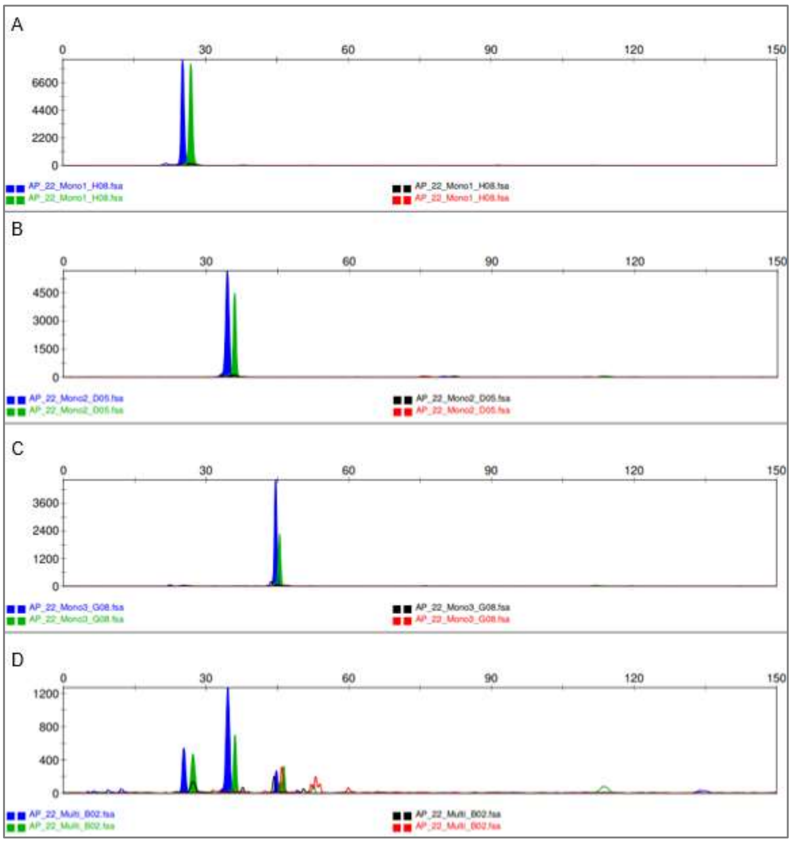

Figure 11 Electropherograms from a semen sample of a 38-year-old individual (sample AP_22). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

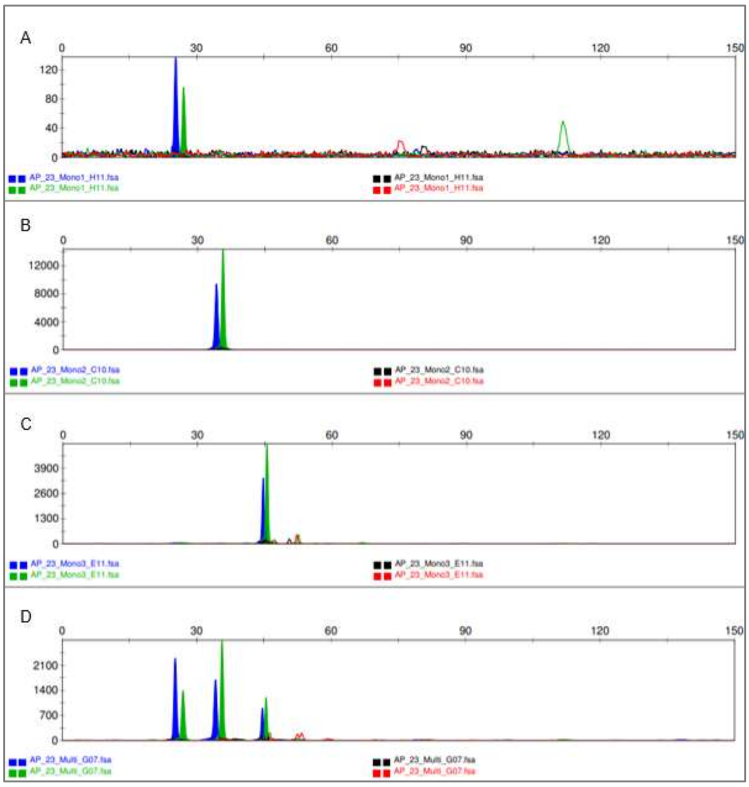

Figure 1732 Electropherograms from a semen sample of a 35-year-old individual (sample AP_16B). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites(D).

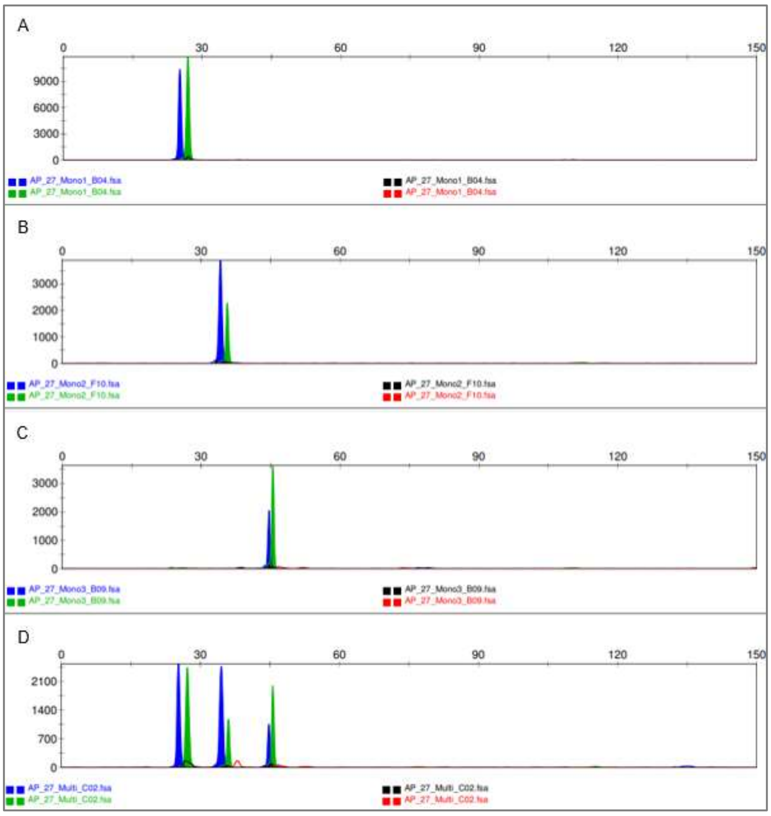

Figure 13 Electropherograms from a semen sample of a 21-year-old individual (sample AP_27). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

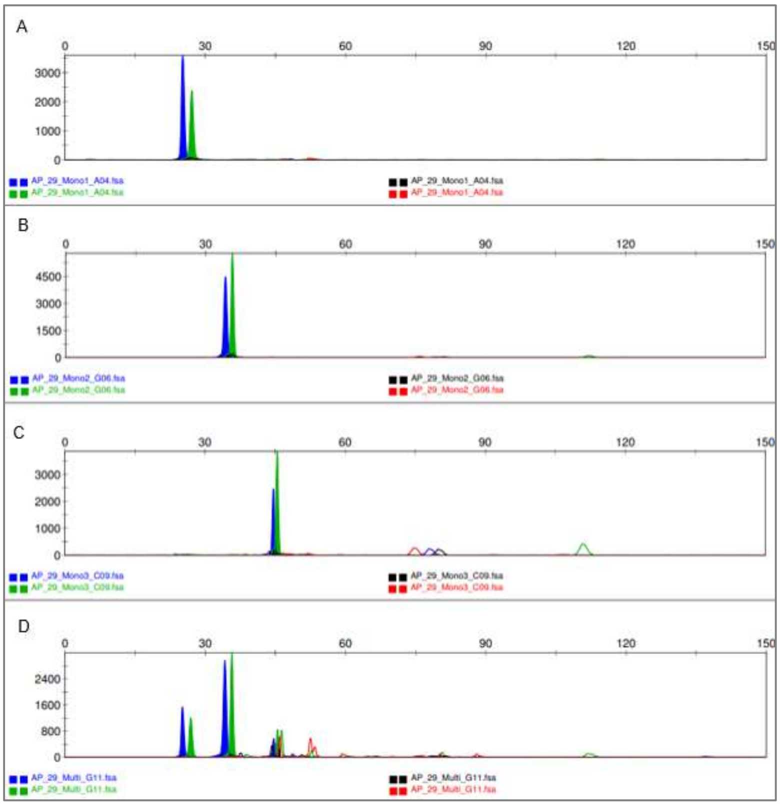

Figure 14 Electropherograms from a semen sample of a 35-year-old individual (sample AP_29). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

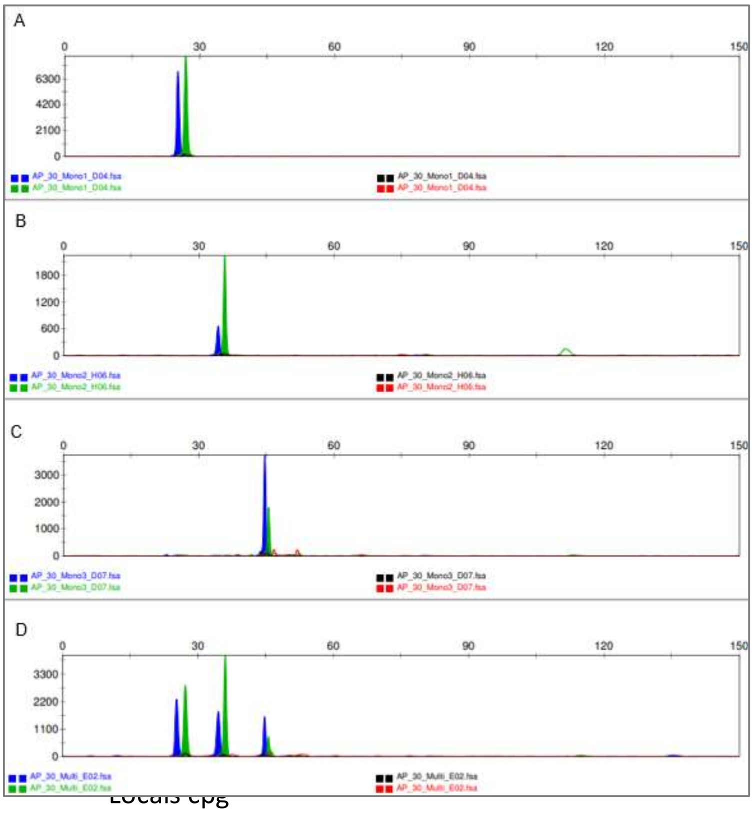

Figure 15 Electropherograms from a semen sample of a 48-year-old individual (sample AP_30). Monoplex reaction for Cg06304190 CpG site (A); Monoplex reaction for Cg12837463 CpG site (B); Monoplex reaction for Cg06979108 CpG site (C) and Multiplex reaction for all 3 CpG sites (D).

7.2 Attachment 2: Statement of consent

DECLARAÇÃO DE CONSENTIMENTO

Considerando a "Declaração de Helsínquia" da Associação Médica Mundial

(Helsínquia 1964; Tóquio 1975; Veneza 1983; Hong Kong 1989; Somerset West 1996 e Edimburgo 2000)

Estimativa da idade a partir de amostras biológicas de sémen

Eu, abaixo-assinado, ____, tomei conhecimento do estudo em que serei incluído(a) e compreendi a explicação que me foi fornecida acerca da investigação que se tenciona realizar. Foi-me ainda dada oportunidade de fazer as perguntas que julguei necessárias e de todas obtive resposta satisfatória.

Foi-me dado a conhecer que, de acordo com as recomendações da Declaração de Helsínquia, a informação ou explicação que me foi prestada versou os objetivos, os métodos, os benefícios previstos, os riscos potenciais e o eventual desconforto da investigação em curso.

Foi-me ainda explicado que os registos dos resultados poderão ser consultados pelos responsáveis científicos e ser objeto de publicação, mas que os elementos da identidade pessoal serão sempre tratados de modo estritamente confidencial, uma vez que apenas o investigador principal terá acesso ao documento onde se encontram as concordâncias entre o código dado à amostra e os dados dos participantes.

Também me foi esclarecido que o material biológico colhido será destruído após o estudo e nunca será usado para qualquer outra finalidade. Por fim, foi-me afirmado que tenho o direito de recusar a todo o tempo a minha participação no estudo, sem que isso possa ter como efeito qualquer prejuízo.

Aceito participar de livre vontade no estudo acima mencionado.

Concordo que seja efetuada a colheita de amostras biológicas para realizar as análises e os estudos genéticos que fazem parte desta investigação.

Também consinto a divulgação dos resultados obtidos no meio científico, desde que seja garantido o seu anonimato.

Data: ____ / _________________ / 20___

Assinatura do voluntário:____

O Investigador responsável: ____