Cecília Pinheiro da Silva

Validation of Sensitivity and Specificity of Triplet- Primed PCR (TP-PCR) in the Molecular Diagnosis of FRAXE

outubro de 2019

Universidade do Minho
Escola de Ciências

Cecília Pinheiro da Silva

Validation of Sensitivity and Specificity of Triplet- Primed PCR (TP-PCR) in the Molecular Diagnosis of FRAXE

Dissertação de Mestrado
Genética Molecular

Trabalho efetuado sob a orientação da
Doutora Paula Maria Vieira Jorge
e da
Doutora Ana Paula Fernandes Monteiro Sampaio Carvalho

outubro de 2019

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Agradecimentos

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

_____

Cecília Pinheiro da Silva

Resumo

Em rotina, a determinação do tamanho da região repetitiva GCC do gene *AFF2* inclui análise baseada em PCR e em Southern blot (SB) sendo esta última uma metodologia cara e demorada. Por isso, noutros genes, o SB está a ser substituído por outras alternativas como o Triplet-repeat primed PCR (TP-PCR). Quanto sabemos, este método nunca foi aplicado ao diagnóstico de FRAXE. A síndrome do XE frágil (FRAXE) é uma forma moderada de défice intelectual associada a défice de aprendizagem, hiperatividade e em casos raros, comportamentos autistas. FRAXE, com uma frequência estimada de 1/50000, é uma doença associada a repetição de tripletos causada pela expansão do tripleto GCC em 5'UTR do gene *AFF2*. O amplo e inespecífico espectro clínico de FRAXE faz com que testes moleculares sejam essenciais para um diagnóstico definitivo. Neste trabalho, um novo TP-PCR foi desenvolvido usando um primer com ligação antes das repetições, um primer $(GCC)_5$ com cauda, que também se liga a uma segunda região em *AFF2*,e um primer idêntico à cauda inespecífica. O ensaio foi otimizado e validado recorrendo a sete amostras com tamanho de alelos conhecidos. Amostras de ADN de 500 mulheres com um PCR de rotina não informativo também foram testadas. Primeiramente, o ensaio determinou corretamente o tamanho dos alelos em 475 amostras com um genótipo de GCC na faixa normal. Nas restantes 25 amostras originalmente genotipadas como homoalélicas, o ensaio determinou 19 com alelos na faixa normal, quatro alelos intermédios e duas pré-mutações. De entre o grupo de 19, quatro foram incorretamente genotipadas devido a um T>C SNP na região próxima das repetições (validado por sequenciação de Sanger). Este SNP também foi identificado em homozigotia numa amostra. Para verificar o tamanho correto das repetições nas amostras com resultados discrepantes, estas foram analisadas por PCR de rotina e Southern blot sendo que, a presença de alelos expandidos, foi confirmada. Descrevemos então uma ferramenta simples, precisa e específica que pode ser usada para determinar o número de repetições (até 100 repetições) em alelos do gene *AFF2*. O ensaio identificou as amostras homoalélicas de forma inequívoca evitando a necessidade de uma segunda técnica demorada e representando uma alternativa atrativa para laboratórios de diagnóstico. Além disto, o ensaio identificou seis amostras com alelos expandidos que são putativamente patogénicos e instáveis representando um valor acrescentado no diagnóstico molecular de FRAXE.

Palavras-chave: AFF2, diagnóstico, FRAXE, repetição de GCC, TP-PCR

Abstract

Routinely, the sizing of the *AFF2* gene GCC repetitive region includes PCR-based and Southern blot (SB) analyses, the latter being a very expensive and time-consuming methodology. For that reason, in other genes, SB is being replaced by alternative approaches such as triplet-repeat primed PCR (TP-PCR). To the best of our knowledge, this method has never been applied to the diagnosis of FRAXE. Fragile XE syndrome (FRAXE) is a form of mild to moderate intellectual disability associated with learning deficits, hyperactivity as well as autistic behaviour in rare cases. FRAXE, with an estimated frequency of 1/50000, is a trinucleotide repeat disease mostly caused by a GCC expansion in 5'UTR of the *AFF2* gene. The broad and unspecific spectrum of FRAXE clinical presentation makes molecular testing essential for a definitive diagnosis. Herein, a novel TP-PCR was developed using a primer binding upstream the repeat, a $(GCC)_5$-tail primer, also binds to a second region within *AFF2*, and a primer identical to the unspecific tail. The assay was optimized and validated resorting to seven samples with known allele sizes. DNA samples from 500 unrelated females with a previous uninformative routine PCR testing result were further tested. Firstly, the assay correctly sized 100% of the alleles in 475 samples with a normal-range GCC genotype. In the remaining 25 samples originally genotyped as homoallelic, our assay determined 19 with alleles within the normal range, four intermediate alleles and two premutations. Among the first group of 19, four had been incorrectly genotyped due to a T>C SNP near the repetitive region (validated by Sanger sequencing), this SNP was also identified in homozygosity in one sample. To verify the correct repeat length in the discrepant samples, they were additionally analysed by routine PCR and SB and the presence of the expanded alleles was confirmed. We describe a simple, accurate and specific tool that can be used to determine *AFF2* alleles up to 100GCC repeats. The assay unambiguously identified homoallelic samples often obviating the need of a second, usually time-consuming technique and representing an attractive alternative for diagnostic laboratories. Furthermore, in six samples this assay correctly identified a putatively pathogenic and unstable expanded allele which had escaped detection with the previously performed PCR representing an added value in the molecular diagnosis of FRAXE.

Keywords: *AFF2*, diagnosis, FRAXE, GCC repeat, TP-PCR

Table of contents

List of abbreviations and Sigla

7-Deaza-dGTP	7-deaza-2'-deoxyguanoside
AFF2	*AF4*/*FMR2* family member 2 gene
CGMJM	Centro de Genética Médica Doutor Jacinto Magalhães
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
FISH	Fluorescence in situ Hybridization
FMR1	Fragile Mental Retardation 1 gene
FRAXA	Fragile X Syndrome
FRAXE	Fragile XE Syndrome
HEX	Hexachloro fluorescein
ID	Intellectual Disability
IQ	Intelligence quotient
MS-MLPA	Methylation-specific Multiplex Ligation-dependent Probe
mRNA	messenger Ribonucleic Acid
NS-XLID	Non-syndromic X-linked Intellectual Disability
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid
SNP	Single Nucleotide Polymorphism
TP-PCR	Triplet-repeat Primed Polymerase Chain Reaction
UTR	Untranslated Region
XLID	X-linked intellectual disability

List of Figures

List of Tables

List of Graphs

Introduction

1. Repeat expansions diseases

Mutations causing the expansion of repetitive microsatellite sequences were first described in 1991 when trinucleotide repeat expansions causing fragile X syndrome (FRAXA) and spinal and bulbar muscular atrophy (SBMA) were identified. Increase in the number of repeats due to a DNA repair defect is known to be the cause of at least twenty-two severe neuromuscular and neurodegenerative disorders including some of the most common inherited diseases, such as Huntington's disease and myotonic dystrophy. Expansion diseases identified up to date comprise units of 3-12 nucleotides[1,2].

The pathogenicity of these expansions depend on the location of the repeat, the number of repeats and consequent size of the unit and the repeat sequence[2]. There are at least four mechanisms of disease causing: loss of function of the gene containing the repeat; gain of function due to production of a toxic RNA (containing an expanded CUG tract); gain of function due to production of protein with a polyglutamine tract expansion; gain of function due to production of a protein with a polyalanine tract expansion[1].

Research on this subject has improved knowledge on the molecular mechanism underlying those diseases such as the impairment in transcription, splicing, translation, mitochondrial function and protein quality control and highlighting opportunities for therapeutic intervention for the currently untreatable disorders[1].

1.1 X-linked intellectual disability

Two of the repeat expansion diseases are caused by folate-sensitive fragile sites in the long arm of the X chromosome. These sites are named FRAXA and FRAXE and are localized in Xq27.3 and Xq28, respectively, with both being responsible for mental impairment[3].

Intellectual disability (ID) is an important and frequent medical condition affecting approximately 3% of the population. It can be caused by mutations in the X chromosome genes and so called X-linked intellectual disability (XLID), accounting for about 10-12% of ID seen in males. An expansion to over 200 CGG in the trinucleotide repeat of the *FMR1* (Fragile Mental Retardation 1) gene is associated with Fragile X syndrome (FRAXA), the most common form of XLID. Besides *FMR1*, FRAXE associated with

AFF2 (*AF4*/*FMR2* family member 2) gene and *ARX* (aristaless related homeobox) gene mutations are some of the causes of XLID[4,5].

Non-syndromic X-linked intellectual disability (NS-XLID) is defined as a 'non-progressive intellectual impairment segregating in an X-linked manner without consistent somatic or diagnostic features'. Several NS-XLID genes have been identified, *GD11*, oligopherenin 1, *PAK3*, *RPS6KA3*, *IL1PAPL*, *TM4SF2* and *AFF2*, with the latter being the most prevalent form of NS-XLID[4].

1.1.1 FRAXE associated gene (*AFF2*)

Following the identification of FRAXA gene in 1991, it became clear that some cytogenetically "fragile" positive families did not have the *FMR1* CGG expansion leading to the identification of a fragile site 600 kb distal to FRAXA. This fragile site named FRAXE is located at Xq28[6-8].

Coincident with the fragile site is *AFF2* gene, formerly named *FMR2* (Fragile X mental retardation gene 2)[6,9] and, similarly to FRAXA, the expansion of a GCC repeat located adjacent to a CpG island upstream of the gene causes its silencing and leads to mental impairment[10,11]. The *AFF2* gene is composed of 22 exons spanning at least 500 kb, it has 6 isoforms with alternative splicing in exons 2, 3, 5, 7 and 21 (scheme in Figure 1)[12,13].
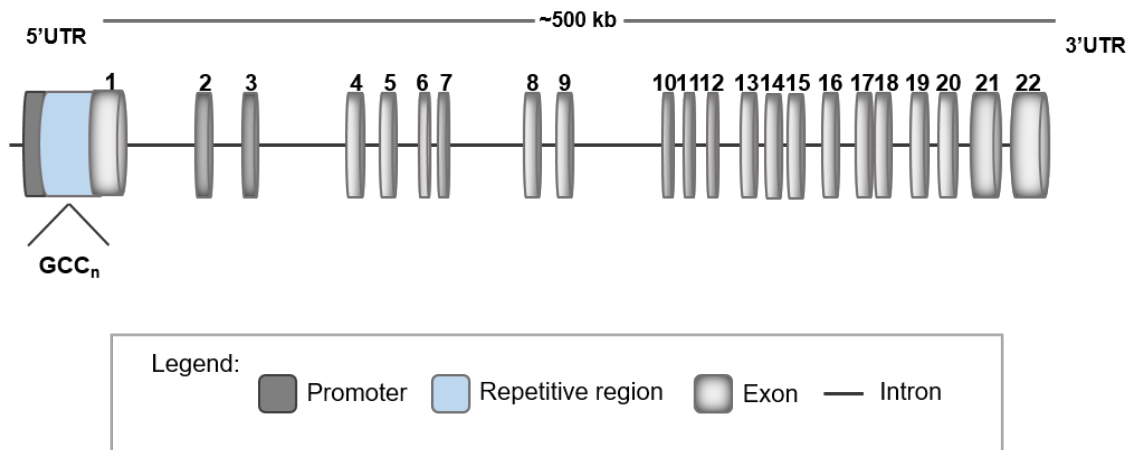
Figure 1 Schematic representation of the AFF2 gene (not at scale). Adapted from Javed *et al*., 2012.

The FMR2 protein (coded by *AFF2* gene) is a member of the FMR2/AF4 protein family localized in nuclear speckles where splicing factors are stored, assembled and modified and in the nucleolus when splicing is blocked. Due to similarity with AF4, it

has been considered a putative transcription factor with its N-terminal domain showing transactivation activity and a RNA binding protein probably involved in the regulation of alternative splicing through interaction with G-quartet RNA structure[9,14]. *AFF2* is abundantly expressed in the brain, particularly the amygdala and hippocampus and in the placenta[6].

A study using *AFF2* knockout mice suggested that the gene was important for a normal function of the central nervous system and that its loss caused learning and memory impairment as well as abnormalities in sensory perception. These results support the role of the FMR2 protein in the human disorder[15].

1.1.1.1 FRAXE GCC alleles

There are few reports on FRAXE GCC allele sizes, herein, both Murray *et al*., 1996 and Annesi *et al*., 2004 allele categorizing was considered (Figure 2)[16,17].

Normal individuals have alleles with 6 to 30 GCC repeats that are stable upon transmission. Alleles with 31 to 60 repeats are considered intermediate or grey-zone and may vary slightly upon transmission. Alleles with repeats ranging from 61 to 200 GCC are considered unstable and designated premutations[16,17]. When the repeats expand to over 200 copies, full mutation, the gene is silenced due to methylation, causing mental impairment. *AFF2* gene methylation is variable with reports showing that alleles with 130GCC can be methylated[6].

The GCC repeat can either expand or contract and is equally unstable when transmitted through both male and female lines[14].
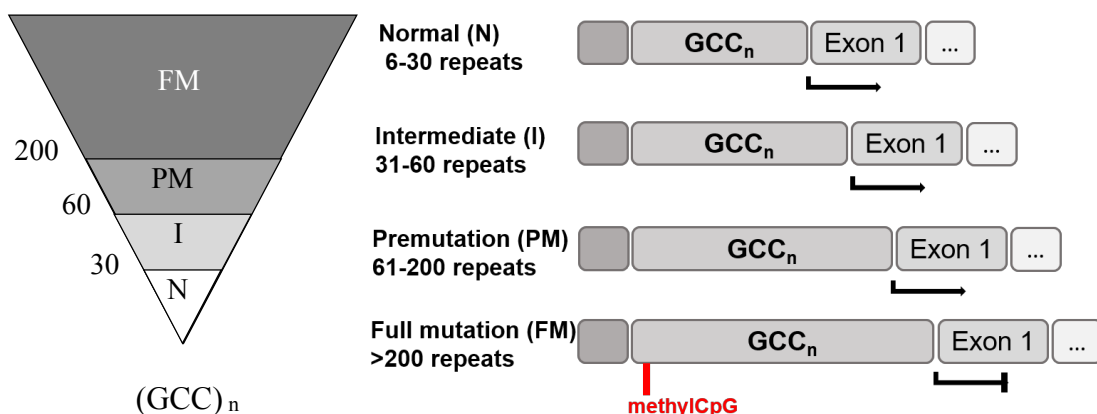
Figure 2 Schematic representation of FRAXE GCC allelic categories. Adapted from Garber *et al*., 2008[18].

Unlike FRAXA repeats that show AGG interruptions conferring stability upon replication, evidence shows that FRAXE repeats are pure, thus, for FRAXE, repeat length may be the only factor of stability. Therefore, the mutation mechanism is probably similar with that of others triplet diseases without interruptions such as Huntington disease and myotonic dystrophy, where the threshold for instability is approximately 40 repeats[19]. Murray *et al*., 1996 detected instability in alleles with 37 and 66 repeats transmitted to male sons. The allele of 37 was transmitted as a mosaicism pattern of 27 and 37 repeats and the allele of 66 expanded to 87 repeats suggesting a potential instability for alleles over the normal range of 30 repeats[16].

a) Normal

Different studies reported a modal FRAXE repeat number of 15/16 (laboratory variation) in New York Caucasian, Finnish and British samples and of 18 in Chinese subjects supporting the existence of ethnic differences in repeat allele distribution as well as a founder effect[19]. In Europe, the most common alleles range from six to 25 repeats with 15 repeats being the predominant[6].

b) Full mutation

FRAXE full mutation, with a frequency of 1/50000 (estimation), is the most common cause of non-syndromic X-linked intellectual disability. It is characterize by mild (IQ of 50-70) to borderline (IQ 70-85) mental impairment with learning and communication difficulties. Some individuals exhibit attention deficit, hyperactivity and autistic behaviours. Because it is a non-syndromic form of intellectual disability with no typical physical features, individuals with FRAXE are not easily distinguished from the general population, however, there has been reports of some clinical features such as long narrow face, mild facial hypoplasia, high-arched palate, irregular teeth, hair and nasal abnormalities, angiomata, clinodactyly and thick lips[6,9]. The phenotype associated with FRAXE is variable, there are cases where individuals carrying the expanded and methylated allele have a normal IQ. These cases might be explained by the *AFF2* expression in the brain during a developmental time window since it is unclear when the methylation of the expanded allele and the extinction of *AFF2* expression occurs[20].

c) Premutation

Using *Drosophila* as a model, it has been showed that the expression of a GCC premutation allele (GCC 90) can cause neuronal degeneration similar to that found in a

FRAXA CGG premutation model suggesting that a GCC premutation could contribute to unknown causes of ataxia[21]. Premature ovarian failure does not seem to be associated with FRAXE premutation as in FRAXA but, an excess of small alleles with fewer than 11 repeats was detected in women with premature ovarian failure by Murray *et al.*,1998, some small alleles were due to deletions within the GCC repeat region[22].

d) Intermediate

The role of FRAXE intermediate alleles is still poorly understood, with studies showing an excess of intermediate alleles in boys with learning difficulties when compared to a control population suggesting that such alleles may not be as benign as previously supposed. This apparent excess could be explained by somatic expansion in tissues such as the brain, secondary structure at DNA or mRNA with an adverse effect on gene function or other causes[16]. A study in an Italian population showed a correlation between intermediate alleles and Parkinson's disease but not mental impairment which is in concordance to the notion that abnormal mRNA levels have an adverse effect on the gene function as showed to premutation alelles[17]. However, in another study with Parkinson's disease patients, no association between intermediate alleles and clinical features of Parkinson's disease was found[23].

1.1.1.2 Other FRAXE gene variants

Intellectual disability cases due to *AFF2* gene by partial or entire gene deletions, partial duplications and translocations have been described. Partial deletions are usually associated with milder phenotypes or autism with severity depending on the location and extent of the deletion while the complete loss of the gene function leads to FRAXE[24,25]. Stettner *et al.*, 2011 described a case where a small deletion caused mild intellectual disability, speech delay, aggression, impulsivity, attention deficit and autistic features[24]. A case with a partial duplication with emotional impairment, auditory processing deficit, infantilized speech and macrosomia reinforced the clinical variability associated with *AFF2* mutations[25]. Another study found an excess of missense variants in highly conserved sites of *AFF2* gene in males with autism spectrum disorder supporting that this gene harbours susceptible alleles that can contribute to the risk of autism and so implicating the gene role in brain function[9].

2. Diagnostic methodologies

2.1 Cytogenetic techniques

Fragile X syndrome, as implicit by the name, is associated with a fragile site in the X chromosome that is expressed as an isochromatid gap in metaphase stage. This fragile site is usually observed in less than 50% of metaphase spreads of cell cultures with folate depletion[7,8].

Among other cytogenetic techniques, fluorescence in situ hybridization (FISH) can identify folate sensitive fragile sites allowing its use as a diagnostic method for Fragile X syndrome (Figure 3)[8]. However, FISH over-diagnoses "Fragile X syndrome" because it shows other fragile sites and non-specific chromosomal breakages. Furthermore, because premutations do not express the fragile site cytogenetically, it would be missed in FISH. Thus, after the identification of the causative gene in 1991, cytogenetic techniques were discontinued and molecular methods started to be developed and routinely used[3].
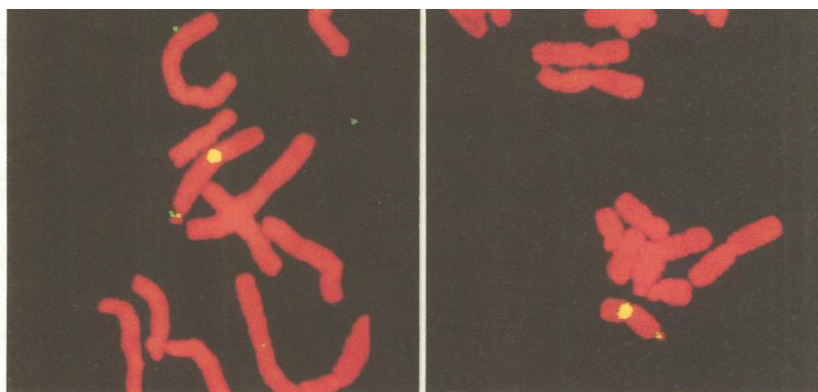
Figure 3 Location of a folate-sensitive fragile site by Fluorescence In situ hybridisation (in yellow: X centromere). Adapted from Flynn *et al*., 1993.

2.2 Polymerase chain reaction (PCR)

To determine the size of a repetitive region a PCR technique can be applied. The use of one primer fluorescently labelled allows the amplified product to be analysed by capillary electrophoresis. PCR is a rapid and easy method that usually amplifies small alleles but not large alleles, particularly when using standard conditions. In the presence of females or mosaicism with two alleles varying in size, amplification may favour the smaller allele. The amplification of CG-rich regions also requires special conditions and can be very time-consuming and expensive. These difficulties rise with increasing number the repeats leading to amplification failure[26].

Several commercial PCR methods that are able to detect alleles with high GC content without formation of nonspecific products have been developed. A triplet-repeat primed PCR using three different primers or triplet-repeat primed-methyl sensitive PCR is one of those methods[27]. Another method is methyl sensitive PCR, a technique involving bisulfite treatment that converts cytosine to uracil reducing the CG content and consequently reducing the high melting temperature, hairpins and secondary structures. This method has been described as advantageous for Huntington's disease diagnosis allowing an amplification with conventional PCR technique and a direct visualization of larger alleles[27,28]. To the best of our knowledge none of these were applied to FRAXE.

2.2.1 Triplet-repeat primed PCR (TP-PCR)

In 1996, Warner *et al*., developed a simple fluorescent PCR system that could rapidly screen expanded CAG repeat in myotonic dystrophy. In this methodology three different primers were used including a specific flanking fluorescently labelled primer (P1) that dictated the specificity and two primers with a common 5' sequence (P3, P4)[29].

The P4 primer had the sequence $(GCA)_5$ or $(TGC)_5$ at the 3'terminus, depending on the strand of the CAG repeat to be amplified, linked with a tail artificially produced, containing little or no self-complementarity, no complementarity with $(GCA)_5$ or $(TGC)_5$ and no homology to known human sequences. The P3 primer, with the same artificial tail as P4, binds to the end of the amplified products due to the stabilising effect of the 5' tail sequence. A 10:1 ration for P3 to P4 was used in order to exhaust P4 in the early amplification cycles and, consequently, reduce priming at the $(CAG)_n$ sequence in earlier rounds combined with a long extension time to allow complete extension of the larger products[29]. A scheme showing the TP-PCR methodology principle and expected result is presented in Figure 4.

The technique was proved to be robust and reliable in the identification of CAG expanded alleles but not their size, that is, TP-PCR can be used to screen DNA samples and exclude the presence of an expansion. Another advantage relies in the ability to identify expanded alleles in poor quality DNA as opposite to other methods that require large amounts of DNA (e.g. Southern blot)[29].

Example: 13 repeats   Example: >200 repeats

Figure 4 Schematic representation of the TP-PCR principle. A: Annealing of P4 primer containing a sequence complementary to the repeats. B: Fluorescently labelled P1 recognizes a sequence upstream the repetitive region. C: Selective amplification with P3 primer in later cycles. D: illustration of the electropherogram after capillary electrophoresis separation; each peak represents a fragment with a unique number of repeats. Expanded alleles (right) produce a continuous ladder of peaks out of the normal range. Adapted from Catalli *et al.*, 2010.

Standardized TP-PCR protocols have been published for several repeat diseases such as Friedreich ataxia, Huntington disease, SCA10, SCA12, SCA2, SCA7, SCA8, SCA3 and also Fragile X syndrome and Fragile X associated primary ovarian insufficiency[30,31].

LabGenomics® developed a commercially available kit 'LabGscan™ PCR Kit®' for the detection of the CGG expansion in the *FMR1* gene based on this TP-PCR technique. In this case, besides the TP-PCR pattern, also the total number of repeats is obtained[32].

2.2 Southern blot

Many laboratories use conventional PCR as a first screening test and use a second methodology when confronted with results such as absence of amplification, uninformative female samples or to exclude mosaicism either in repeat size and/or methylation status. In repeat expansion disorders, this second methodology is commonly Southern blot that can quantify full mutations resistant to PCR amplification and, when combined with a methylation-sensitive endonuclease, allows the determination of its methylation status[5,27]. Southern blot is a good strategy allowing the detection of large premutations, full mutations and mosaic patterns. This technique resolution is around

150bp so an accurate size of *FMR1* gene premutations (ex: 70CGG ≈ 210bp) requires the use of PCR. In *FMR1* the sizing is crucial to distinguish between large normal, intermediate and premutated alleles and to assess the risk of expansion upon transmission[27,33].

Briefly, a DNA sample is digested using restriction enzymes and subjected to electrophoresis in horizontal agarose gel. After overnight separation, the fragments are denaturized, neutralized and transferred to a nitrocellulose, polyvinylidene or Nylon membrane through blotting. During this process, the fragments are carried with the buffer and retained in the membrane surface by capillary transfer or using a vacuum blotting device for approximately 4 hours. The nucleic acids are then fixated to the membrane using heat or ultraviolet light and undergo an overnight hybridization step. After several washing steps to remove unhybridized probe, incubation with an antibody and with a substrate, it is subjected to autoradiography with X-ray film or chemiluminescence depending on the probe labelling to visualize the targeted region(s)[34-36].

2.3 Sanger sequencing

Determining the order of nucleic acids in polynucleotide chains is essential to a variety of research applications. Sequencing is crucial to the identification and interpretation of human sequence variation as it relates to health and disease[37,38].

Sanger's 'chain-termination' or dideoxy technique uses dideoxynucleotides (ddNTPs) lacking the 3' hydroxyl group tagged with a specific fluorescent dye. Combining ddNTPs at a fraction of concentration of standard dNTPs, one primer (usually M13-tail) and DNA polymerase results in DNA strands of different lengths. The resulting dye-labelled fragments are submitted to capillary electrophoresis producing fluorescence emissions of four different colours revealing the DNA sequence[37,38].

This methodology is used for instance to identify missense variants in Fragile X patients when no expansion is observed[39].

2.4 Diagnostic methodologies at CGMJM (Centro de Genética Médica Doutor Jacinto Magalhães)

The laboratory developed a multiplex PCR technique that simultaneously amplifies the repetitive region of *FMR1* and *AFF2* genes and the 24bp duplication in *ARX* gene[5]. All samples belonging to individuals showing intellectual disability are tested using this

multiplex strategy and three important mutational hotspot regions can be simultaneously screened[5]. This routinely used technique amplifies regions that include both *FMR1* and *AFF2* genes triplet repeat regions using primers flanking those regions and allowing the discrimination of alleles in the normal, intermediate and premutation ranges after capillary electrophoresis separation. Because expanded alleles are usually not amplified, when only one allele in the normal range is obtained in a female sample, two distinct explanations, with different clinical impact, can be formulated, a) it is a normal homoallelic case or b) the second allele is expanded and putatively implicated in the clinical features[40]. In the case of male patients the routine PCR method can also overlook mosaicism for premutation and full mutation alleles (due to the preferential amplification of smaller alleles and/or competition between the three primer-pairs)[26].

Expansions causing the silencing of the *AFF2* gene are diagnosed through Southern blot, a very time-consuming and expensive methodology that also requires a large amount of DNA[41].

To avoid these difficulties, there is a need for a screening method that is both easy and fast to perform which is where the triplet-repeat primed PCR technique comes in. TP-PCR was proposed by Warner *et al.*, 1996 for other genes/triplet repeats to detect larger alleles in disorders with triplet repeat expansions[29]. Furthermore, the confirmation of homoallelism by the TP-PCR, would avoid the use of the time-consuming and expensive Southern blot.

This methodology is already used to screen *FMR1* expansions and to determine the *FMR*1 AGG interspersion pattern. However, to the best of our knowledge, it was never described in *AFF2* gene.

3. Objectives /Aims

The main aim of this thesis is the development of a triplet-primed polymerase chain reaction (TP-PCR) assay for the analysis of the *AFF2* gene repetitive region.

Other goals include testing and validating this new methodology in seven samples with previously known GCC size in terms of:

- Primer sequence;
- Primer labelling;
- PCR amplification conditions;
- Primers concentrations;

- PCR enhancers;
- DNA quality.

After validation and establishment of the best conditions it will be applied in 500 female samples with a previous "unknown" genetic status (only one size allele was obtained after routine PCR).

This study has been approved by DEFI – Departamento de Ensino, Formação e Investigação do Centro Hospitalar Universitário do Porto, CHUP E.P.E., as well as the Hospital's Ethical Committee - N/ REF.ª 2018.179(154-DEFI/153-CES).

All the DNA samples used were anonymously stored at the lab after being tested for Fragile X and signed informed consent.

Materials and Methods

1. Samples

Five-hundred and eight genomic DNA samples, anonymously stored after a normal Fragile-X screening result, were used in this study. These DNA samples were previously obtained from peripheral blood using ArchivePure DNA Cell/Blood Kit (5 Prime® GmbH, Hamburg, Germany). Among those, seven samples (three males and four females), previously sequenced and containing GCC repeats within the normal, intermediate and premutation ranges, were used to validate this new approach of TP-PCR (Table 1), one male sample carrying a full mutation in *AFF2* gene was used as a positive control in Southern blot analysis and the remaining five hundred female samples, with a previous uninformative *AFF2* routine PCR result, were analysed by the newly established TP-PCR methodology.

Table 1 Samples and respective number of GCCs used to validate the TP-PCR.

Sample	Number of GCC
1♂	14
2♂	17
3♀	14/26
4♀	14/38
5♀	19/41
6♂	48
7♀	13/79

2. Primers selection and construction

Four primers were designed: two primers in the forward sense (F) with the same sequence but fluorescently labelled with different dyes, NED and Hexachloro fluorescein (HEX); a primer binding in the repetitive GCC region of the *AFF2* gene $(GCC)_5$ with a nonspecific tail; and a primer identical to the nonspecific tail (TAIL). Primers sequences are presented in Table 2.

Table 2 Primers tested and respective sequence.

Primer name	Primer designation	Sequence 5' → 3'
F (NED)	g.AFF2_GCC_F (NED)	TGT GAG TGT GTA AGT GTG TGA TGC TGC C
F (HEX)	g.AFF2_GCC_F(HEX)	TGT GAG TGT GTA AGT GTG TGA TGC TGC C
GCC	g.AFF2_TP_PCR_GCC	TAC GCA TCC CAG TTT GAG ACG CGG CGG CGG CGG CGG C
Tail	g.AFF2_TP_PCR_TAIL	TAC GCA TCC CAG TTT GAG ACG

3. Triplet-repeat primed PCR

3.1 PCR amplification conditions

Reaction mix components are described below in Table 4. Applied Biosystems® Veriti® 96-Well Thermal Cycler was used in the amplification reaction with the following program (Table 3):

Table 3 Thermocycler program used for the TP-PCR amplification.

Step	Temperature	Time (min)	Cycles
Initial denaturation	98ºC	5:00	1
Denaturation	98ºC	1:00	15
Annealing	55ºC	1:00	
Extension	68ºC	2:00	
Denaturation	98ºC	1:00	30
Annealing	58ºC	1:00	
Extension	68ºC	3:00	
Final extension	68ºC	10:00	1

Table 4 TP-PCR components and respective concentrations.

Reagents	
PCR Master Mix (Promega®, Madison, WI, USA)	1x
Betaine (Sigma-Aldrich®, St. Louis, Missouri, USA)	0.8M or 0,6M (for larger alleles)*
DMSO (Sigma-Aldrich®, St. Louis, Missouri, USA)	10%
Q-Solution® (Qiagen® GmbH, Hilden, Germany)	0.5x
7-Deaza-dGTP (Roche®, Basel, Switzerland)	0,2mM (for larger alleles)*
g.AFF2_GCC_F [NED labelled]	0.2pmol/µL
g.AFF2_TP_PCR_GCC	0.1pmol/µL
g.AFF2_TP_PCR_TAIL	0.2pmol/µL
DNA	450ng
dH$_2$O	Up to 25uL

*Both 0,6M of Betaine and 0,2mM of 7-Deaza-dGTP were only used in cases with more than 50GCC repeats in a second TP-PCR round

In samples with more than 50GCC repeats, besides lowing Betaine and adding 0,2mM 7-deaza-dGTP, also the thermal cycling conditions were further optimized (Table 5):

Table 5 Optimized Thermocycler program used for alleles with over 50GCC repeats.

Step	Temperature	Time (min)	Cycles
Initial denaturation	98ºC	5:00	1
Denaturation	98ºC	1:00	15
Annealing	55ºC	2:00	
Extension	68ºC	4:00	
Denaturation	98ºC	1:00	30
Annealing	58ºC	2:00	
Extension	68ºC	4:00	
Final extension	68ºC	10:00	1

3.2 Fragment size determination

5µL of each TP-PCR product was added to 15 µL of mixture of formamide and GeneScan™ 500 ROX™ dye Size Standard (Applied Biosystems™). Fragments were separated by capillary electrophoresis on ABI PRISM® 3130xl Genetic Analyser (Applied Biosystems™, Foster City, CA, USA) and the results analysed using GeneMapper® Software version 4.0 (Applied Biosystems™).

A previously sequenced sample with 15GCCs was used as control to determine the exact size in base pairs (15GCC correspond to 188,5bp in the establish conditions).

4. Sanger sequencing- symmetric PCR

To determine the sequence of the GCC repetitive region, a symmetric PCR was conducted using a forward and a reverse primers (Table 6) and according to Table 7. The PCR amplification program took place in a Applied Biosystems® Veriti® 96-Well Thermal Cycler using an initial denaturation of 5min at 98ºC followed by 15 cycles of 45s at 97ºC, 45s at 54ºC and 45s at 68ºC, 30 cycles of 60s at 97ºC, 45s at 54ºC and 45s with an increase of 5s per cycle at 68ºC, ending with a final extension of 10min at 72ºC.

Table 6 Primers and respective sequence used in the PCR amplification.

Primer	Sequence 5' → 3'
g.AFF2_CCG_SEQ_F*	TGT GAG TGT GTA AGT GTG TGA TGC TGC C
g.AFF2_CCG_SEQ_R	TAG CCC GCG CAC CCA GCG AC

*When the goal is fragment analysis, the forward primer is fluorescently labelled with HEX

Table 7 PCR components and respective concentrations.

Reagents	
AccuTaq™ LA Buffer (Sigma-Aldrich®)	1x
Betaine (Sigma-Aldrich®, St. Louis, Missouri, USA)	1M
DMSO (Sigma-Aldrich®, St. Louis, Missouri, USA)	4.8%
7-Deaza-dGTP (Roche®, Basel, Switzerland)	0.22mM
dATP/dCTG/dTTP (Bioline, London, UK)	0.12mM
dGTP (Bioline, London, UK)	0.024mM
g.AFF2_PCR_F [HEX labelled]	0.2pmol/µL
g.AFF2_PCR_R	0.2pmol/µL
AccuTaq™ LA DNA polymerase (Sigma-Aldrich®)	2U
DNA	50ng
Distilled water	Up to 25uL

5. Sanger sequencing – asymmetric PCR

5µL of the symmetric PCR products were purified with 2µL of enzymatic PCR clean up technology- Illustra™ ExoProStar™ 1-step (GE Healthcare Life Sciences, Little Chalfont, Buckinghamshire, UK) for 30min at 37ºC followed by 15min at 80ºC.

Asymmetric PCR amplification consisted of an initial denaturation of 95ºC for 6min, 28 cycles of 95ºC for 10s, 50ºC for 5s and 60ºC for 4min and a final extension of 60ºC for 10min, using the reagents according to Table 8.

Table 8 Sanger sequencing components and respective concentrations.

Reagents	
Purified symmetric product	7µL
BigDye™ Terminator v3.1 Cycle Sequencing Mix (Applied Biosystems®)	2µL
BigDye™ Terminator v3.1 5x Sequencing Buffer (Applied Biosystems®)	0.67x
g.AFF2_CCG_SEQ_F or g.AFF2_CCG_SEQ_R	0.67pmol/µL

After gel filtration spin columns clean up (DyeEx®96 kit, QIAGEN GmbH, Hilden, Germany), products underwent capillary electrophoresis on ABI PRISM® 3130xl Genetic Analyser (Applied Biosystems™, Foster City, CA, USA) and the results analysed using SeqScape® Software version 2.5 (Applied Biosystems™).

6. Southern blot – *AFF2* repetitive region

To test samples with a possible expanded allele, Southern blot was performed following manufacturer's instructions from 'FRAXE/FMR2/AFF2 CCG Repeat Genotyping GeneProber™ (FMR2 CCG triple repeat non-radioactive Southern blot genotyping- GeneLink®)' and 'DIG Wash and Block Buffer Set'(Sigma-Aldrich®)[36,42].

The procedure started with an overnight, at 37ºC, double digestion of 10μg of genomic DNA using 3μL of endonucleases *Afl*III (New England BioLabs®, Ipswich, Massachusetts, EUA) and *Not*I (New England BioLabs®, Ipswich, Massachusetts, EUA) with the latter being sensible to CpG methylation and 1x NEBuffer™ 3.1 (New England BioLabs®, Ipswich, Massachusetts, EUA). The digested products were loaded into a 0.8% agarose gel and electrophoresed in parallel with DIG II (Sigma-Aldrich®, St. Louis, Missouri, EUA) and DIG III (Sigma-Aldrich®, St. Louis, Missouri, EUA) standard size markers for 16 hours. Electrophoresis was followed by a depurination step using 0.2N HCl for 5min, denaturation with 0.5M NaOH/1.5M NaCl for 15min and two neutralization steps with 0.5M Tris pH 7.5/1.5M NaCl for 15min and then blotted to a nitrocellulose membrane for 4 hours. After washing (2x SSC) and crosslinking at 120joules, the membrane was prehybridized with Easy Hyb buffer (Roche®, Basel, Switzerland) at 55ºC for 2 hours and then hybridized overnight at 55ºC with ~2.5μL/400cm^2 of denatured *AFF2* specific probe (AFF2 AJ31-Dig1 GeneProber™). The membrane was washed twice in SSC 2x, SDS 0.1% (m/v) for 5min at room temperature, twice in SSC 0,5x, SDS 0.1%(m/v) for 15min at 60ºC and in 1x Washing buffer for 5min at room temperature. Blocking in 1x Maleic acid for 1 hour and incubated with 50mL of Anti-DIG-AP conjugate solution in 1x Blocking in 1x Maleic acid (1:10000), washed twice with 1x Washing buffer for 15min and once for 5min and then equilibrated with 1x Detection buffer for 5min. The membrane was covered with CDP-star substrate and exposed to chemiluminescence using Fujifilm Luminescent Image Analyzer LAS-3000 v2.2 (Fujifilm, Tokyo, Japan).

7. **Methylation-Specific Multiplex Ligation-dependent Probe Amplification (MS-MLPA)**

Multiplex ligation dependent probe (MLPA®) was carried out on five samples suspected to have a small deletion. MLPA® was performed using SALSA® MLPA® ME029 FMR1/AFF2 Probemix and according to manufacturer's instructions. Briefly, the DNA samples are denaturated and undergo a hybridisation reaction during 18 hours. Products are split into two parts, one for probe ligation and the other for ligation and digestion by *Hha*I of unmethylated DNA, both undergo an equal final PCR reaction.[43]

MLPA® data was analysed using Coffalyser.Net software (MRC-Holland, Netherlands).

Results and Discussion

1. TP-PCR amplification

As schematized in Figure 5, the TP-PCR amplification is obtained by the combination of a forward primer (green arrow), a primer with five CGGs (complementary to the GCCs) that anneals in the triplet repeats (yellow arrow) and a third equal to the specific primer tail (orange arrow). A distinct fragment is also obtained due to primer ligation further down the sequence (sequence shown in dark yellow). Because of a timine (T) in the exact middle of the primer with five GCC repeats, with optimized conditions, a specific ligation is possible. This allowed the analysis of the repeat region total length simultaneously with the triplet-repeat profile.

Figure 5 Schematic representation of the primers binding sites during the TP-PCR amplification strategy. Sequence of the AFF2 gene (NCBI reference sequence: NG_016313.2) from position c.4952 to c.5156. Due to the existence of an almost perfect $(GCC)_5$ repeat downstream the polymorphic GCC repeat a second PCR product is obtained.

After amplification and separation by capillary electrophoresis, TP-PCR profiles were obtained (Figure 6). The first peak "contains" five repeats (as the primer binds to five GCCs) and each peak after this corresponds to an increase of one repeat. Figure 6 shows an example of a sample with 15GCCs. Due to some unspecific binding in the CG rich region subsequent to the GCC repeats, some small peaks can be also be observed. The peak at 188.5bp (Figure 6) corresponds to the total allele size. Confirmation of these results was done using samples previously sequenced.

Figure 6 TP-PCR result obtained in GeneMapper® software after capillary electrophoresis of the amplified products. Besides the triplet-repeat profile (left) it is also possible to obtain the total size of the fragment (right).

2. Optimization of the amplification conditions

GC-rich DNA sequences have higher melting temperatures caused by increased hydrogen bonds strength and secondary structures such hairpins, knots and tetraplexes that, combined with hinder denaturation and primer annealing, may lead to polymerase arrest and a premature extension. These sequences are, then, difficult to amplify by PCR and, regularly, result in little or none expected product as well as the amplification of several unspecific products[44,45].

Small changes in the PCR protocol can have a huge impact on product formation. Thus, an optimization of the assay is required with different approaches such as magnesium chloride concentration, buffer pH, denaturation and annealing time and temperature, cycle number, modified nucleotides, among others[44,45].

Herein, several parameters were optimized including the primer labelling, cycle number, annealing temperature, annealing and extension times, primers concentration and PCR enhancers (Betaine, DMSO, Q-solution and 7-deaza-2'-deoxyguanosine).

Before optimization, a standard amplification program was used (Table 9).

20

Table 9 Thermocycler program used for the TP-PCR amplification.

Step	Temperature	Time (min)	Cycles
Initial denaturation	98ºC	5:00	1
Denaturation	98ºC	1:00	45
Annealing	60ºC	1:00	
Extension	68ºC	2:00	
Final extension	68ºC	10:00	1

2.1 Fluorescent labelling

Two different labels were tested in the same primer sequence (NED and HEX). NED (chemical structure not available) is a fluorescent dye used to fluorescently label oligonucleotides at the 5'-end that can be used in several applications such as real-time PCR, hybridization probes and other fluorescence-based genetic analysis applications. NED has an absorbance maximum of 546nm and an emission maximum of 575nm thus, emitting in yellow at the visible spectrum[46].

Hexachloro-fluorescein (HEX) is a version of fluorescein used to fluorescently label oligonucleotides at either the 5'- or 3'-end that can be used in several applications such as real-time PCR, in hybridization probes and for structure-function studies. Oligo primers labelled at the 5'-end can be used to generate fluorescently-labelled PCR or genetic analysis products. HEX has an absorbance maximum of 535nm and an emission maximum of 556nm thus emitting in green at the visible spectrum[47].

A comparison of the results obtained with the forward primer labelled with NED or HEX is shown in Figure 7.
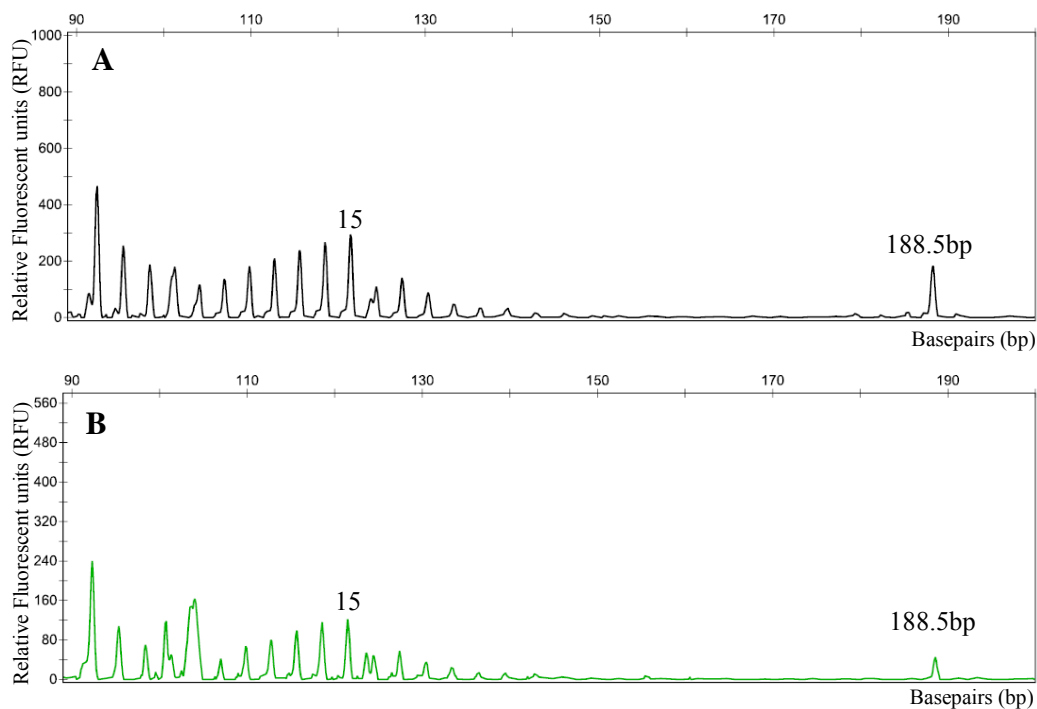
Figure 7 TP-PCR result obtained using a forward primer labelled with NED (A) or HEX (B).

HEX labelling leads to a more irregular profile and less amount of amplified product (240 RFU) compared to NED (over 400RFU). The latter (NED), although with a 'V' shaped profile, allows the easy identification of each fragment. In the following experiments and in the development of the methodology, only NED labelling was used.

2.2 Annealing temperature

Primers with a high GC content may require higher annealing temperatures (>55ºC) when compared to primers with low GC content (<50%). The use of low annealing temperature can increase the quantity of non-specific products but if the temperature is too high, poor annealing of the primers may occur[48,49].

In this case, the first test (Table 9 from Section 2.Optimization of amplification conditions) was performed using an annealing temperature of 60ºC for 1min (Figure 8).
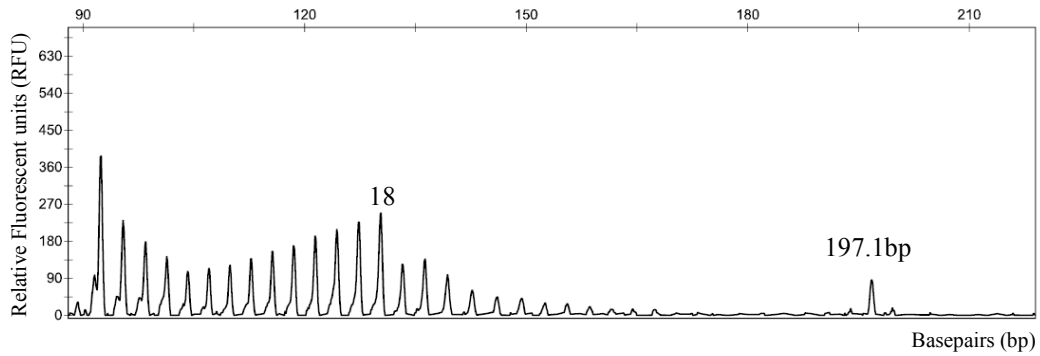
Figure 8 TP-PCR result obtained with an annealing temperature of 60ºC.

At 60ºC, the expected size was obtained, however, with small amplification intensity (RFU).

Two different annealing experiments were, then, conducted. In one there was a combination of 15 cycles with an annealing of 57ºC followed by 30 cycles with annealing at 60ºC. In another experiment, 15 cycles with an annealing temperature of 55ºC followed by 30 cycles with annealing at 58ºC. The results of other temperature/ nº cycles combination are shown below in Figure 9. Another example of these tested conditions is shown in Figure 10.
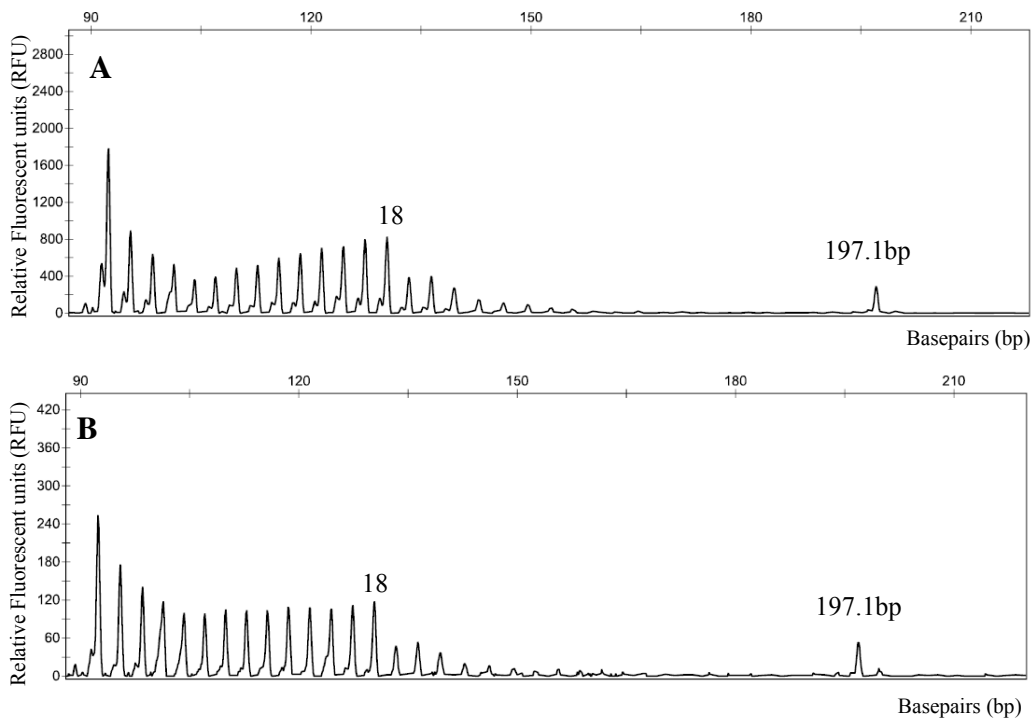
Figure 9 TP-PCR result obtained with an annealing temperature of 15x 57ºC/30x 60ºC (A) and 15x 55ºC/30x 58ºC (B).
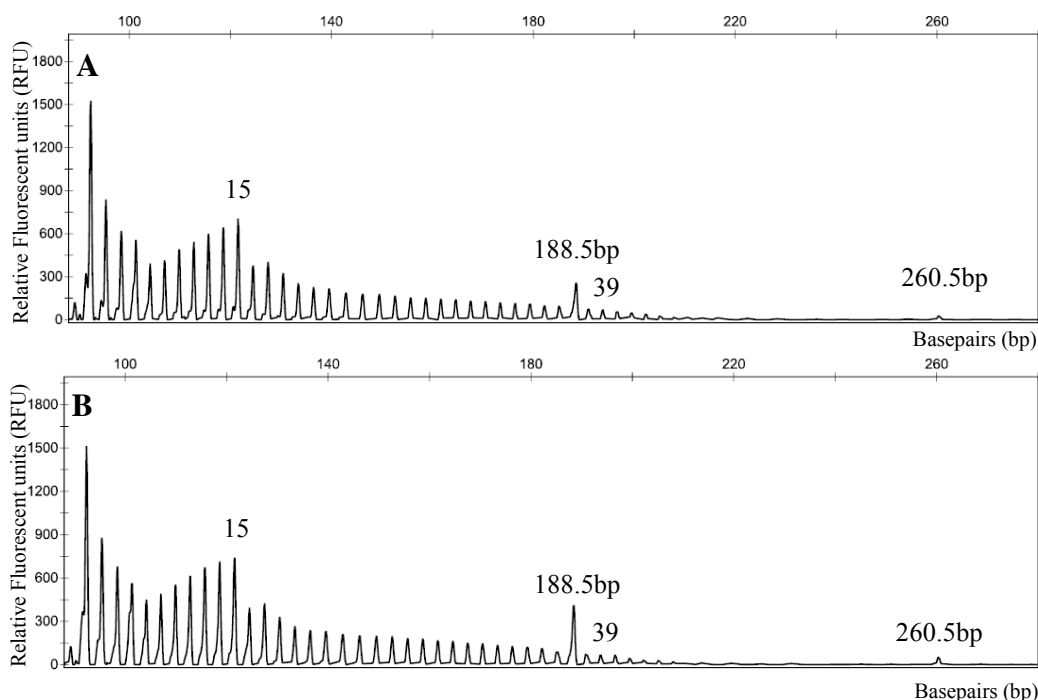
Figure 10 TP-PCR result obtained with an annealing temperature of 15x 57ºC/30x 60ºC (A) and 15x 55ºC/30x 58ºC (B).

Although the profiles obtained with the two different amplifications conditions (A and B) are similar, 'cleaner' peaks are present in the second experiment (B) with 15 cycles at 55ºC followed by 30 cycles with annealing at 58ºC in both samples (Figures 9 and 10).

Furthermore (Figure 10), it can be also noted that the fragment that corresponds to the total size of the allele is favoured in the condition B which may be because a lower annealing temperature favours "nonspecific" products[48]. In this particular experiment, the binding of the primer further down in the sequence is beneficial because a second validation of the total allele size can be obtained.

Following these results, the amplification testing protocol was establish as described in Table 3 (Materials and Methods Section 3.1 PCR amplification conditions).

2.3 Primer concentrations

A ratio of 10:1 between the tail sequence primer (g.AFF2_TP_PCR_TAIL or Tail) and the repeat sequence primer (g.AFF2_TP_PCR_GCC or GCC) should be used to exhaust the primer binding in the early amplification cycles in order to reduce priming at (GCC)$_n$, produced in earlier rounds[29].

In the development of this TP-PCR, a ratio of 2:1 (as in routine experience) was tested with a concentration of 0.2pmol/µL for both the forward (g.AFF2_TP_PCR_F or F) and Tail primers and of 0.1pmol/µL for the GCC primer.

Three other different primer concentration combinations were tested, one with the forward primer reduced to half (0.1pmol/µL), the Tail primer reduced to half (0.1pmol/µL) and a third with the GCC primer reduced to a concentration of 0.04pmol/µL with a ratio of 5:1 between the Tail primer (0.2pmol/µL) and the GCC primer (0.04pmol/µL).

The results of the four conditions tested in the same DNA sample are shown below in Figure 11.
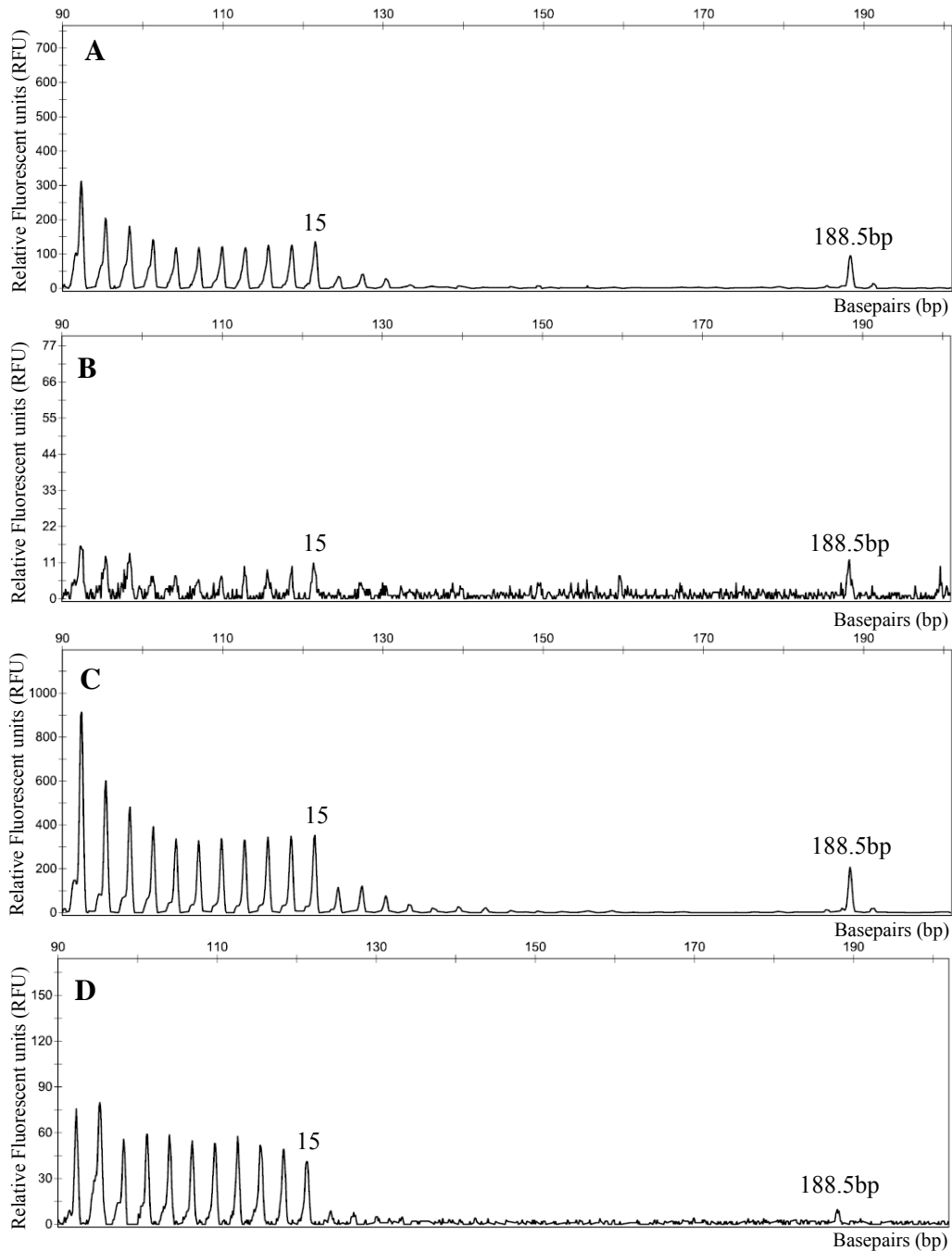
Figure 11 TP-PCR result obtained after testing different primer concentrations. A: 0.2pmol/µL F and Tail and 0.1pmol/µL GCC. B: 0.1pmol/µL F (reduced to half), 0.2pmol/µL Tail and 0.1pmol/µL GCC. C: 0.2pmol/µL F, 0.1pmol/µL Tail (reduced to half) and 0.1pmol/µL GCC. D: 0.2pmol/µL F and Tail and 0.04pmol/µL for GCC (ratio 5:1).

Analysing these results allow us to conclude that, a decrease in the forward primer concentration gives a weaker signal, hampering the proper repeat number counting.

The same occurs with the decrease in the GCC primer concentration where the signal is below 100RFU. A reduction in the tail primer does not seem to change the amplification profile. Following these results, the concentrations of 0.2pmol/µL for primers F and Tail and that of 0.1pmol/µL for primer GCC, with a ratio of 2:1, were used.

2.4 PCR additives

Particularly GC-rich DNA sequences can form secondary structures that are difficult to amplify by PCR and, commonly, result in little or absence of expected product as well as the amplification of unspecific products. The inclusion of appropriate PCR additives to the PCR mix can be essential to increase yield, specificity and accuracy of the amplification[44]. There is a variety of additives and enhancing agents that can be useful to the PCR optimization such as dimethyl sulfoxide (DMSO), N,N,N-trimethylglycine (Betaine), formamide, glycerol, non-ionic detergents, bovine serum albumin (BSA), polyethylene glycol and tetramethylammonium chloride[44].

Both Betaine and DMSO facilitate strand separation of double helix DNA by altering its melting characteristics and are ideal as additives because they are inexpensive, easily obtainable and compatible with other biological agents[50]. Betaine, an amino acid analog with both positive and negative charges close to neutral pH, equalizes the contribution of GC- and AT- base pairing to the stability of the DNA duplex[44,50]. DMSO acts by disrupting inter and intra strand re-annealing[50].

Several agents with application in GC- rich DNA to facilitate product formation are commercially available but their chemical composition is unknown[44,45]. Q-solution provided by Qiagen® is one of those cases, it is described as an agent that facilitates amplification of difficult templates by modifying the melting behaviour of DNA, it improves suboptimal PCR due to the presence of secondary structures or GC- rich templates but it does not compromise PCR fidelity[51].

7-deaza-2'-deoxyguanosine (7-deaza-dGTP) can weaken base to base interactions thus, solving superstructures, making it useful to improve PCR product yield. It can also be helpful in GC- rich sequences sequencing, especially in low amounts of template or poor DNA quality[52].

Besides Betaine, DMSO and 7-deaza-dGTP, also the addition of Q-solution was tested. Figure 12 shows the results of the impact of those additives on the TP-PCR amplification, on the left add component is indicated: 0.5x of Q-solution (A) 0.8M of Betaine (B) and 10% of DMSO (C) compared to their absence (on the right). Results D and E compare the substitution of Betaine and DMSO (left) with 0.2mM 7-deaza-dGTP (right) and G the presence (left) or absence (right) of 7-deaza-dGTP. In another experiment (F), both Betaine and DMSO concentrations were reduced to 0.56M Betaine and 7% DMSO (right).
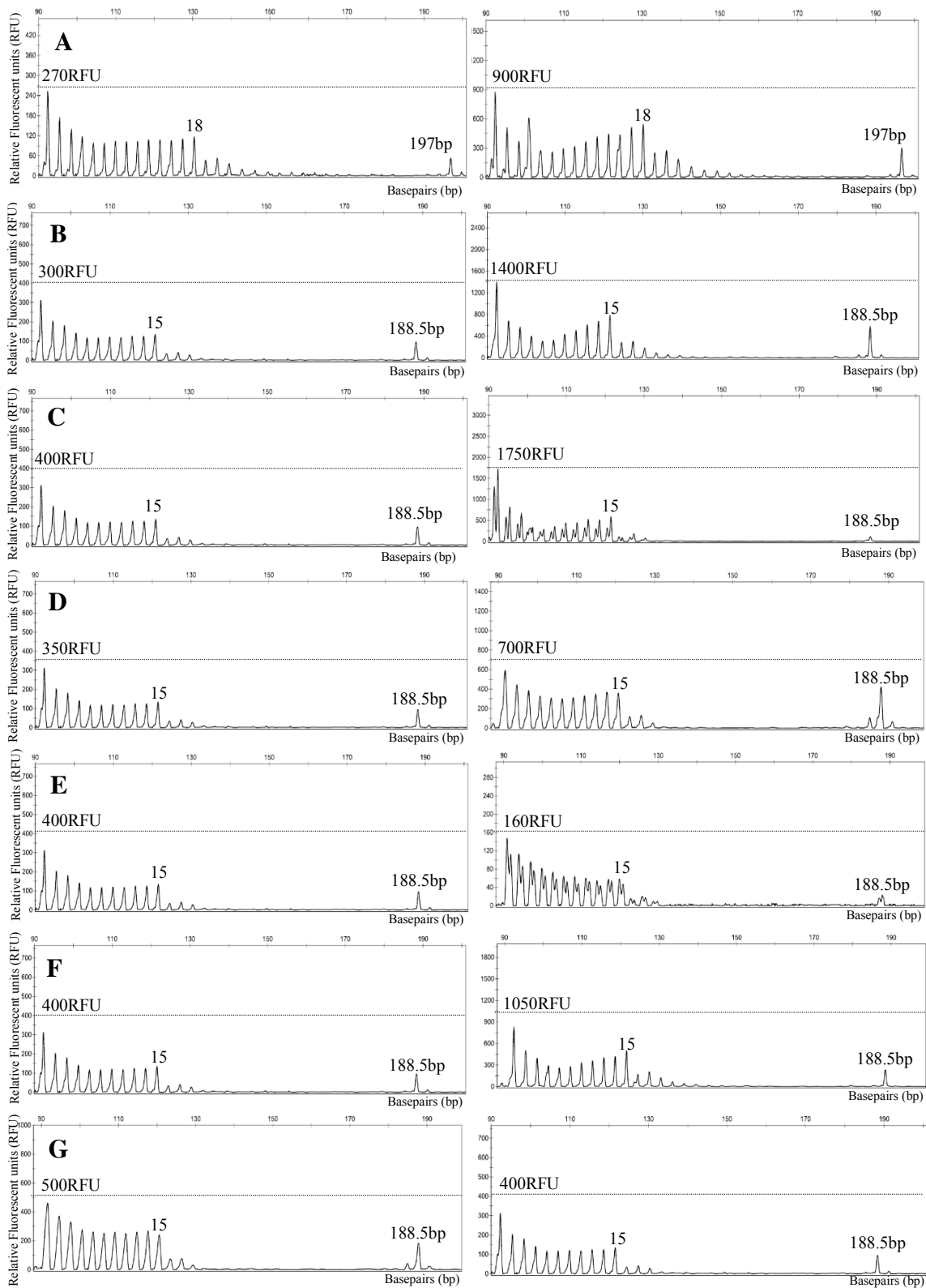
Figure 12 TP-PCR result obtained with the different additives: 0,5x Q-solution (left) vs no Q-solution (right). B: 0.8M Betaine (left) vs no Betaine (right). C: 10% DMSO (left) vs no DMSO (right). D: 0.8M Betaine (left) vs 0.2mM 7-deaza-dGTP/ no Betaine (right). E: 10% DMSO (left) vs 0.2mM 7-deaza-dGTP/ no DMSO (right). F: 0.8M Betaine/ 10% DMSO (left) vs 0.56M Betaine/ 7% DMSO (right). G: 0.2mM 7-deaza-dGTP (left) vs no 7-deaza-dGTP (right).

The addition of Q-solution improves the amplification resulting in a more regular peak profile.

The absence of Betaine causes a more 'V' shaped profile and its substitution by 7-deaza-dGTP increases the signal of both the TP-PCR and the total length fragment. According to this, Betaine seems to uniformise the TP-PCR profile but, it is not essential, whereas 7-deaza-dGTP is useful to increase product formation (higher RFU).

The absence of DMSO in the Mix results in 'broken' peaks as well as a low amplification of the fragment representing total length. The substitution of DMSO by 7-deaza-dGTP, although better than in the absence of 7-deaza-dGTP, still leads to 'broken' peaks. Both results show that adding DMSO is essential to this TP-PCR amplification.

The reduction of both Betaine and DMSO concentrations to 0.56M and 7%, respectively, does not seem to alter significantly the amplification profile but gives a higher RFU, however, a more 'V' shaped can also be noticed.

The additive effect of 7-deaza-dGTP improves the signal of the TP-PCR similarly to that observed in experiment D.

These results are in accordance to the literature showing that the use of PCR additives, especially DMSO, are crucial to a good performance in the amplification particularly that of GC-rich sequences[44].

2.5 DNA quality /concentration

The DNA quality and quantity is crucial for amplification success. Contaminants as urea, SDS detergent and sodium acetate can decrease the PCR efficiency. The amount of template must also be sufficient to allow amplification but, high concentration may decrease efficiency essentially due to the increase in contaminants concentration[49].

Salting out is a common DNA extraction method for fresh blood samples granting a high yield and purity. However, salting out requires both high volume of starting material and is a very time-consuming manual work[53,54]. DNA obtained by the automated methods usually yield a low concentration but are particularly useful when only small volumes of starting material are available. However, appropriate kits and compatible equipment are needed[53].

A comparison between extraction methods, manual- Salting out and automated (EZ1)[55] was performed using DNA samples with 450ng (168ng/µL) obtained manually

and 43ng (8.6ng/µL) obtained through automated purification (EZ1 DNA blood kits – Qiagen®). The obtained results are shown in Figure 13.
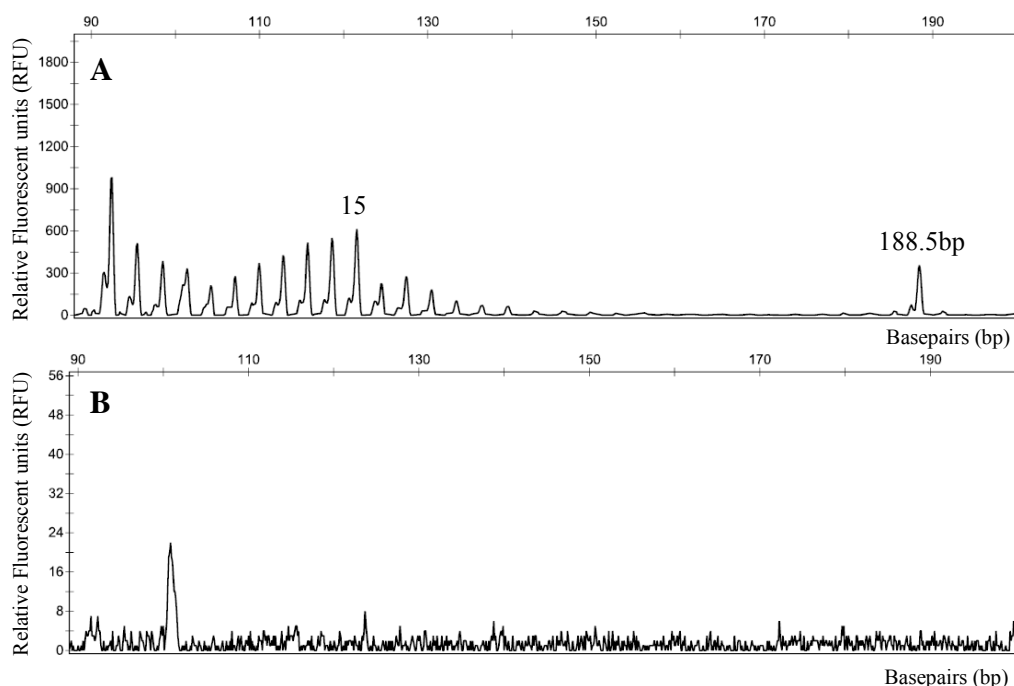
Figure 13 TP-PCR result obtained from DNA extracted through manual extraction method from fresh blood (A) and from DNA extracted through automated extraction method from dry blood spot (B).

As observed in Figure 13, a DNA concentration below 10ng/µL does not allow TP-PCR amplification. Because a maximum of 5µL of DNA is used, a minimum concentration of 20ng/µL was established and used in future experiments.

2.6 Annealing and extension times

The optimized protocol allowed to accurately count up to 50 repeats. In order to increase this number the protocol was further optimized for large alleles changing the amplification conditions.

An increase in the annealing (2min instead of 1min) and extension (4min instead of 2min) times as well as the addition of 0.22mM 7-deaza-dGTP allowed the amplification and quantification of larger alleles (Figure 14).
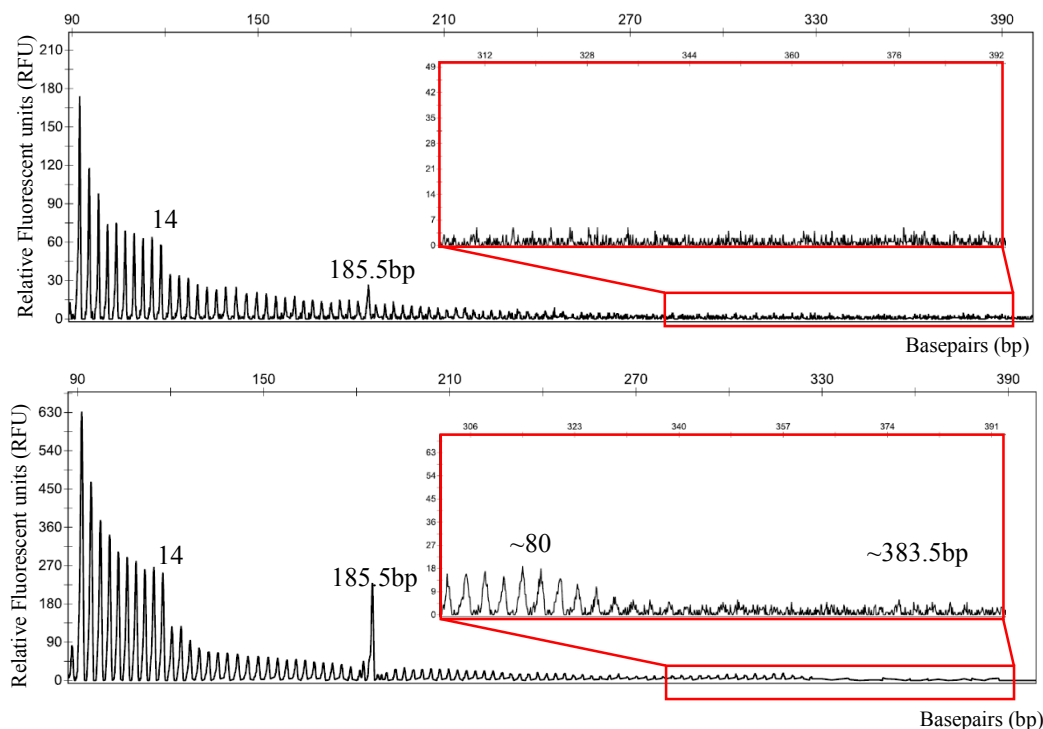
Figure 14 TP-PCR result obtained from a sample with 14GCC and 80GCC. A: previous protocol. B: optimized protocol for larger alleles.

2.7 Final amplification conditions/ protocol

After the optimization of the amplification- conditions, a PCR mix containing Betaine, DMSO and Q-solution besides PCR Master Mix, primers and DNA (Table 4 from Materials and Methods Section 3.1 PCR amplification conditions) was considered to be optimal. The amplification program found to be most efficient include an initial denaturation step of 5min at 98ºC, followed by 15 cycles of denaturation at 98ºC for 1min, annealing at 55ºC for 1min and extension at 68ºC for 2min, with 30 cycles at 98ºC for 1min, 58ºC for 1min and 68ºC for 3min and with a final extension of 10min at 68ºC.

3. Characterization

After the optimization, 500 samples previously characterized as uninformative through routine PCR, were tested using the developed TP-PCR (Annex 1).

Overall results are displayed in Graph 1. Of the 500 female samples, 446 were confirmed to be homoallelic (three with a different result when compared to that obtained

in routine PCR) for the *AFF2* repetitive region and the remaining 54 showed alleles with distinct size.
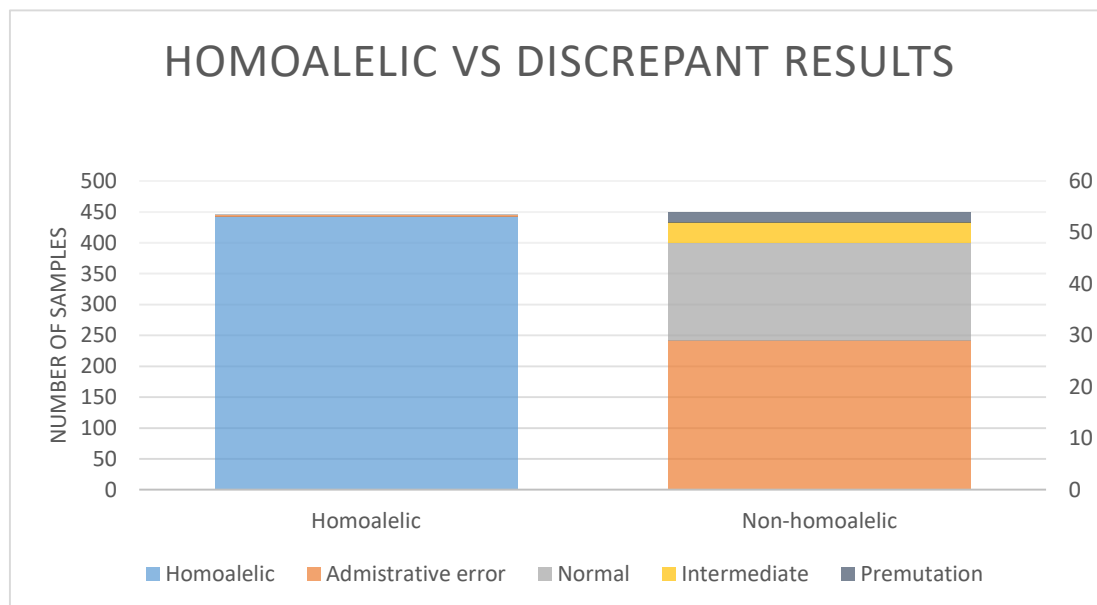
Graph 1 Distribution of the results from the 500 female samples tested by TP-PCR.

Twenty-nine out of this 54 samples were considered "administrative errors" because they were assigned at the database as homoallelic although routine PCR result had previously revealed alleles with two distinct sizes. Two of the 446, although homoallelic, had a different result from the one in the database also due to wrong size assignment at the database.

Overall, the previous conventional PCR failed to detect the second allele in 19 samples with alleles within the normal range, four intermediate alleles and two premutations (allele categorizing followed Murray *et al*., 1996)[16]. These results were unexpected, particularly in samples with normal size alleles. The possible explanation is a preferential amplification of the smaller allele leading to a false case of homoalelism[26]. Interestingly, a sample revealed to be homoallelic but with a different number of repeats when compared to that detected by PCR (discussed below in 3.4.4 Section). Four heterozygous samples showing an unexpected result will also be discussed below.

A total of 13 different genotypes were identified in the 446 confirmed to be homoallelic (Table 9). The most common genotype is 15/15 and is expected because 15GCC allele is the most common allele size in the European population as described in the literature[6].

Table 9 Genotypes identified in the 446 homoallelic samples.

Genotype (nº of GCC repeats)	Number of samples	Percentage %
11/11	1	0.22
12/12	1	0.22
14/14	6	1.35
15/15	332	74.44
16/16	35	7.85
17/17	3	0.67
18/18	37	8.30
19/19	7*	1.57
20/20	10	2.24
21/21	4	0.90
23/23	4	0.90
24/24	4	0.90
25/25	2	0.45
		Total 446

*One of the samples with 19/19 has a different number of repeats when compared to the routine PCR results

In the 1000 analysed alleles, the mean number of GCC repeats is 16.29 (S.D 4.64) with a mode of 15 and ranging from 11 to 107GCC repeats. Allele distribution is similar to that obtained by Clark *et al*., 2019 with the allele with 15GCC being the most common, followed by that of 18GCC and 16GCC suggesting a bimodal distribution (Graph 2)[56].
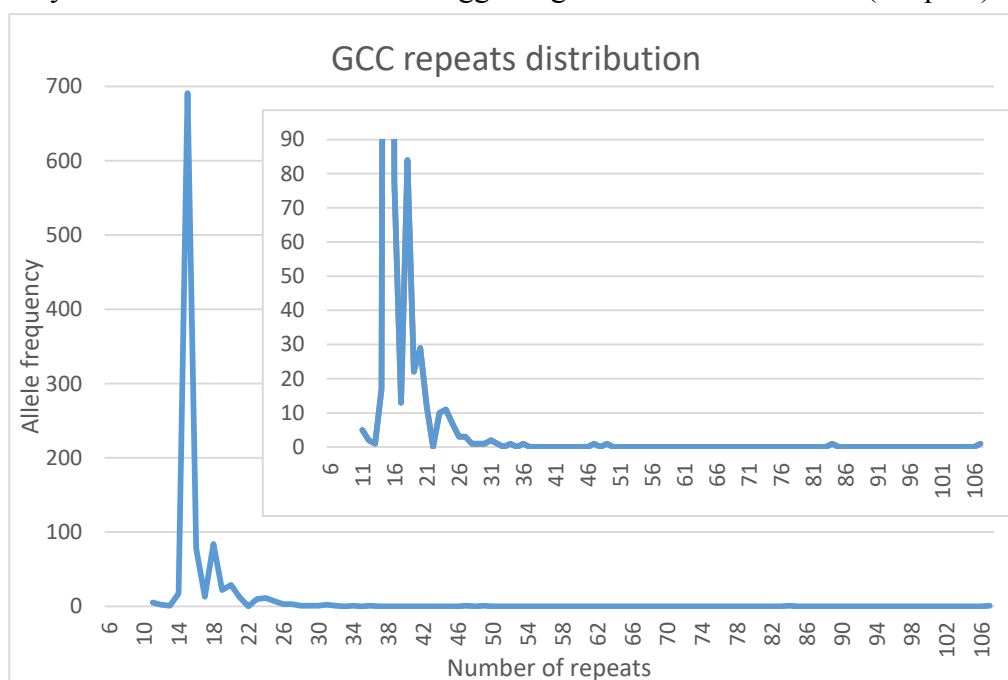
Graph 2 Distribution of the number of GCC repeats in *AFF2* gene (1000 alleles).

3.1 Normal alleles

The novel TP-PCR technique detected 15 heterozygous samples with *AFF2* alleles within the normal range. In Figures 15 and 16 are some examples of samples that had previously revealed an uninformative PCR results. In the first example, alleles of 14 and 15GCC (Figure 15) were identified. This result shows that the established conditions allow the discrimination of alleles differing in one repeat.
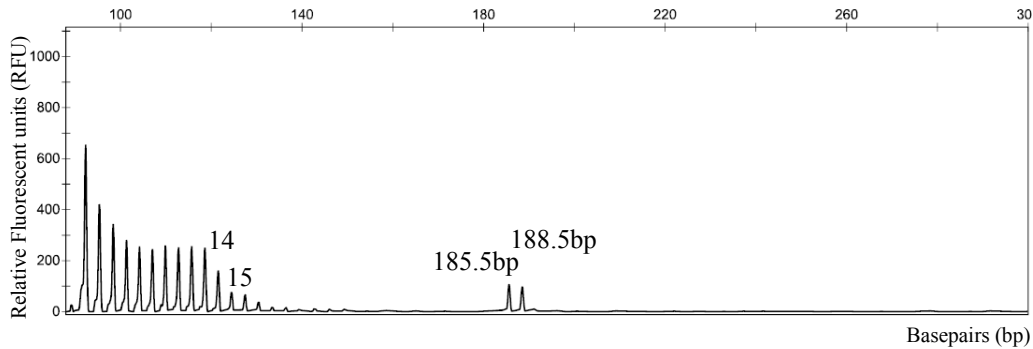
Figure 15 TP-PCR result obtained from a sample with a 14GCC and 15GCC alleles.

In this second example (Figure 16), a difference of 12 repeats (15 and 27 GCC) can be observed. In our opinion the fact that alleles are ~36bp apart in size can explain why the second allele was not observed with the routine PCR. Furthermore, the fact that routine PCR used three primer pairs (multiplexing *FMR1*, *AFF2* and *ARX*) hampers the amplification of the second allele. This bias towards the smallest alleles seems to be occurring more frequently with AFF2 than FMR1 gene alleles (P. Jorge, personal communication).
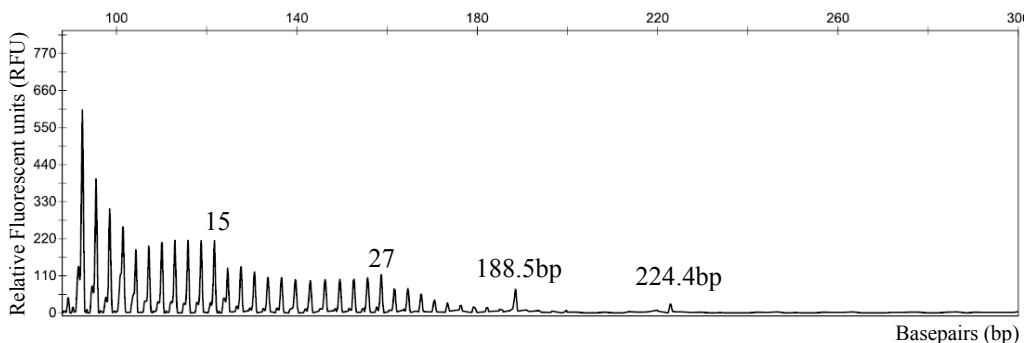
Figure 16 TP-PCR result obtained from a sample with a 15GCC and 27GCC alleles.

3.2 Intermediate alleles

In the 500 tested samples, four carry intermediate alleles (ranging from 31 to 60 repeats) were identified (Figures 17 to 20).
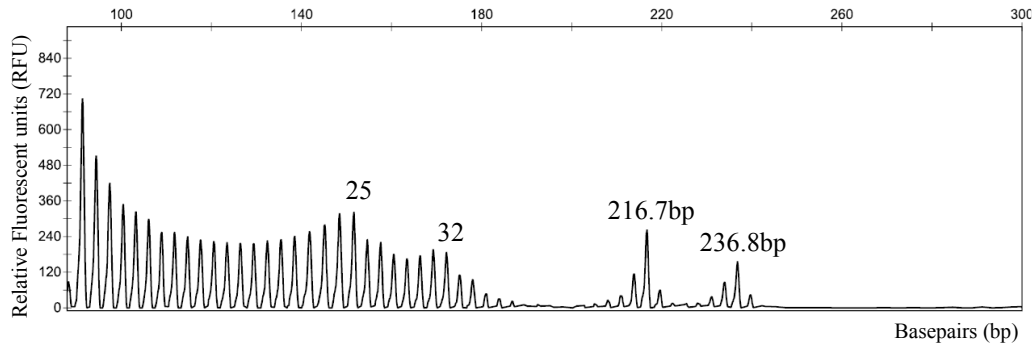
Figure 17 TP-PCR result obtained from a sample with a 25GCC and 32GCC alleles.
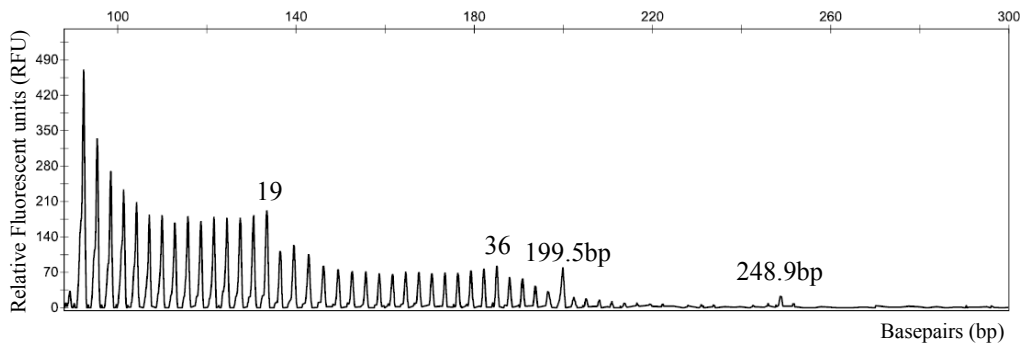
Figure 18 TP-PCR result obtained from a sample with a 19GCC and 36GCC alleles.
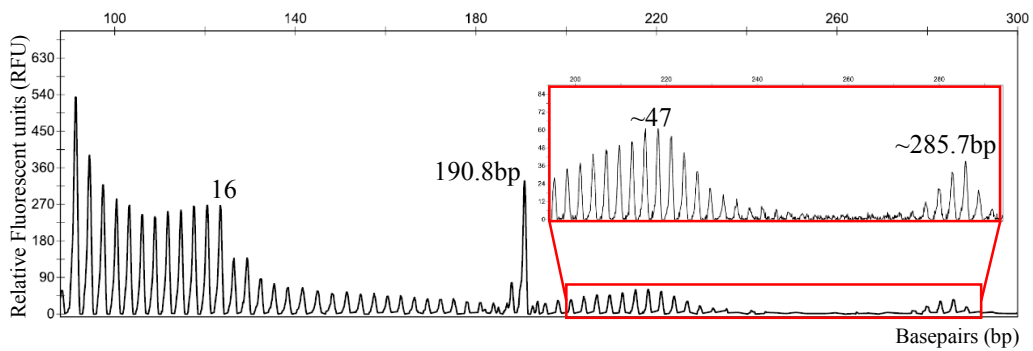
Figure 19 TP-PCR result obtained from a sample with a 16GCC and 47GCC alleles.
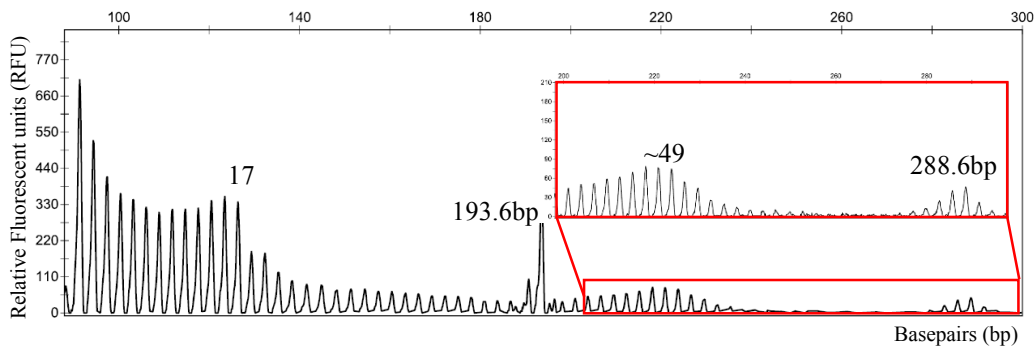
Figure 20 TP-PCR result obtained from a sample with a 17GCC and 49GCC alleles.

3.3 Premutation alleles

The new developed method allowed the identification of two female premutation carriers, one with 15 and 84GCC alleles and the other with 15 and ~107GCC alleles (Figures 21 and 22).
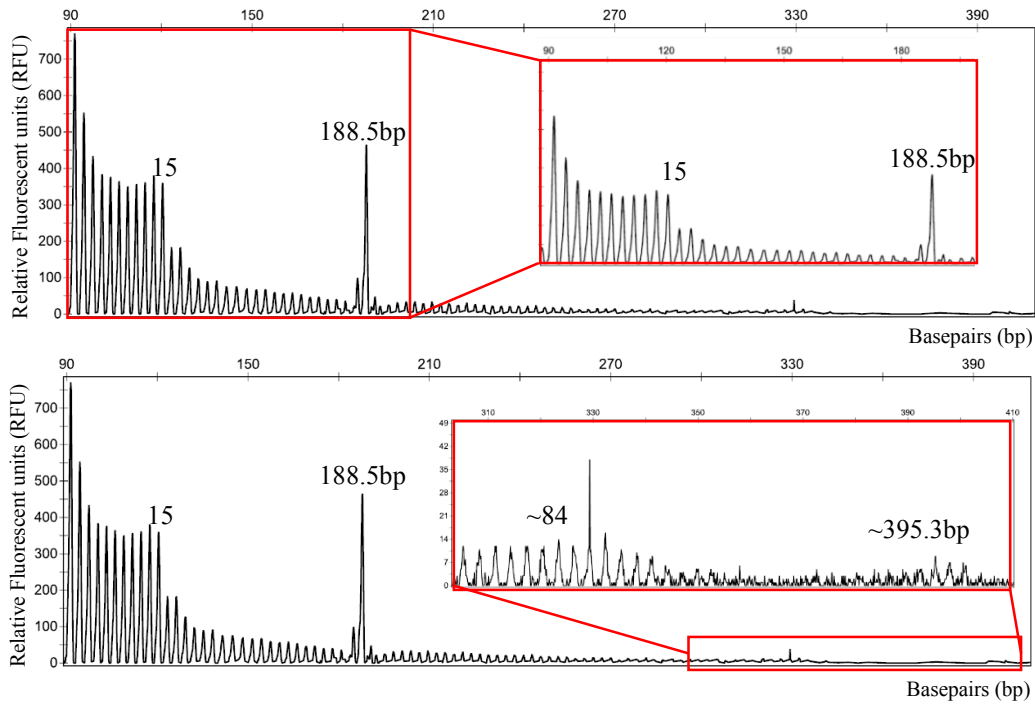
Figure 21 TP-PCR result obtained from a sample with a 15GCC and ~84GCC alleles.
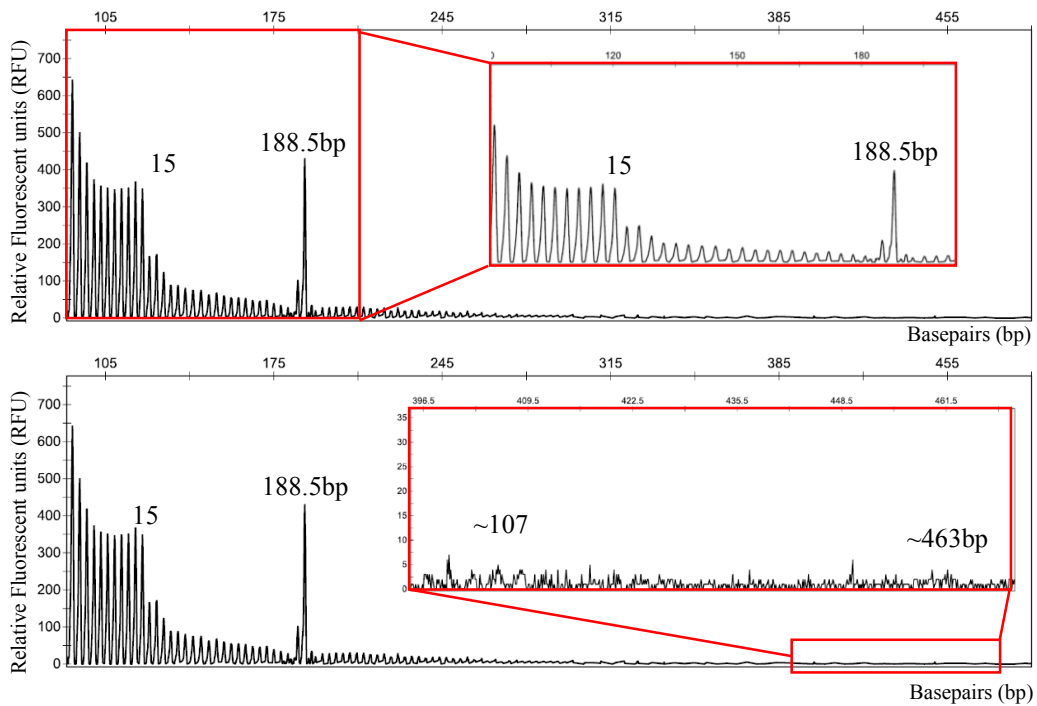
Figure 22 TP-PCR result obtained from a sample with a 15GCC and ~107GCC alleles.

In the expanded range (over 100 repeats) it was not possible to accurately quantify the exact size of the allele. However, a clear distinction between true homoallelic and an expansion can be attained and this is the purpose as a screening method. If needed for diagnostic purposes, the exact number of GCC repeats in expanded alleles, should be obtained by Southern blot.

3.4 Confirmatory techniques

3.4.1 Sizing of larger alleles

The routine PCR was optimized to allow the amplification of larger alleles. Essentially, only *AFF2* locus was amplified by removing the primers recognizing *FMR1* and *ARX* loci, as well as slight changes in the amplification program allowed the quantification of *AFF2* second allele up to ~107GCCs. Results are shown in Figures 23 to 25.
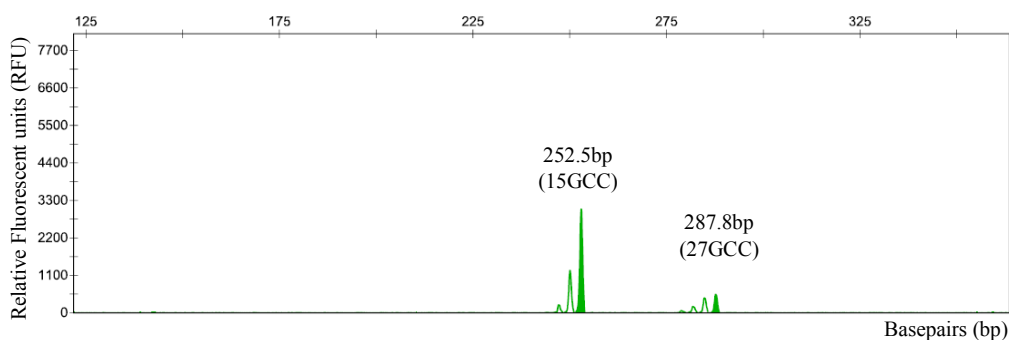
Figure 23 PCR result obtained from a sample with a 15GCC and 27GCC alleles.
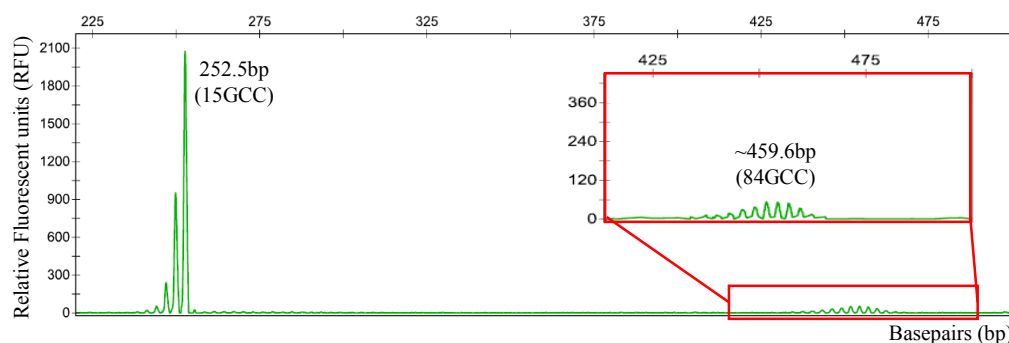
Figure 24 PCR result obtained from a sample with a 15GCC and ~84GCC alleles.
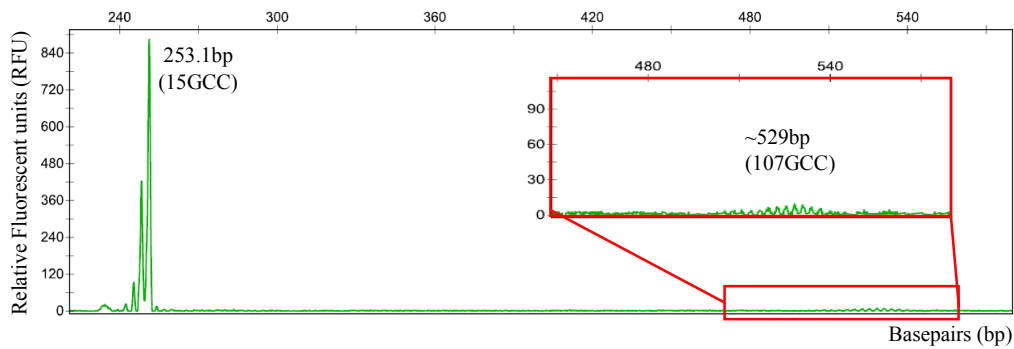
Figure 25 PCR result obtained from a sample with a 15GCC and ~107GCC alleles.

A confirmation of the allele sizes obtained by TP-PCR, even for the large alleles of 84 and 107GCC, was obtained.

3.4.2 Southern blot

Southern blot analysis with an *AFF2* specific probe (AFF2 AJ31-Dig1 GeneProber™) was performed using, among others, samples with larger alleles. A calibration curve using DIG II and DIG III size markers was used to calculate the approximate number of GCCs (Figure 26).
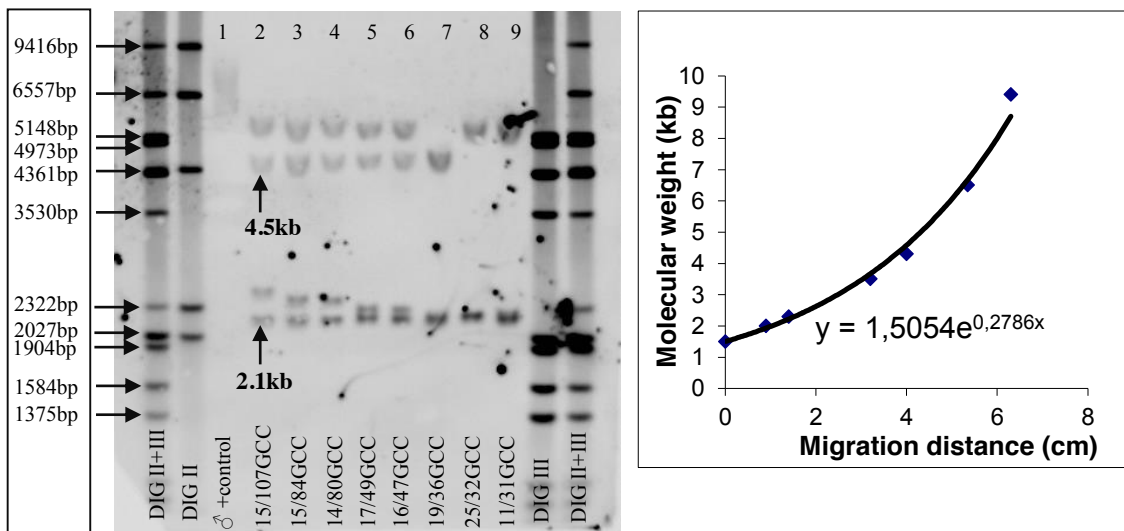
Figure 26 Southern blot analysis. A: Southern blot result. B: Calibration curve.

In female samples, a normal size fragment of 2.1kb fragment corresponds to the active X chromosome and the 4.5kb fragment to the methylated inactive X chromosome. A

FRAXE full mutation male (>300GCCs) was used as a positive control showing methylated fragments larger than 6kb (Lane 1).

The results obtained by the Southern blot technique were concordant with those from the novel TP-PCR. The intermediate size alleles' fragments migrated slightly above 2.1kb and the premutation alleles between 2.3kb (84GCC) and 2.4kb (107GCC). The methylated premutation alleles are not clearly distinguish because the agarose concentration used is unable to discriminate such small changes.

The 5.5kb fragment obtained in the "methylated" area in lanes 2-6, 8 and 9 correspond to the change c.4022193G>A in *AFF2* gene. This single nucleotide polymorphism (SNP), described in Ensembl as rs5980369, occurs within the sequence recognized by *Afl*III. When the A is present, the restriction enzyme cannot cut and a larger fragment is obtained (Figure 27)[57].

Furthermore, the observed amount of methylated and unmethylated allele is similar so we can conclude that a random methylation pattern is occurring at *AFF2* locus.

The two premutated alleles were confirmed (Lanes 2 and 3) and no mosaicism was detected.
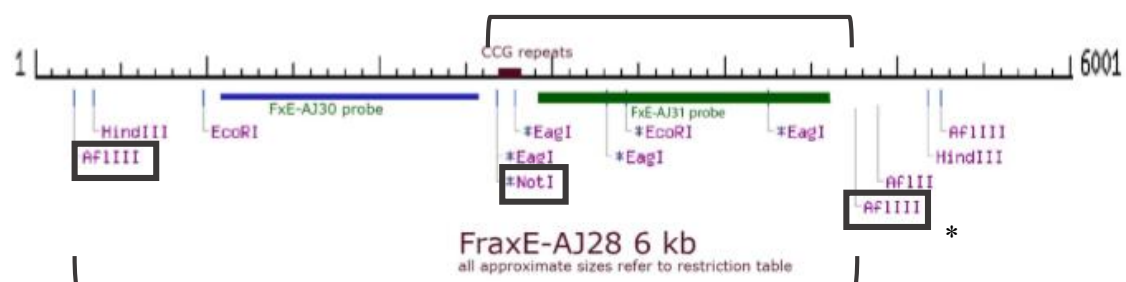
Figure 27 Scheme of the *AFF2* gene with restriction sites for several restriction enzymes. *Not*I and *Afl*III enzymes (dark box) and FxE-AJ31 probe were used in the Southern blot. *rs5980369 (c.4022193G>A).
Adapted from FMR2 CCG triplet repeat non-radioactive Southern blot genotyping protocol.

3.4.3 Methylation-specific Multiplex ligation-dependent probe (MS-MLPA®)

MS-MLPA® kit combines the detection of small copy number changes with the methylation status. In this particular case MS-MLPA ME029 *FMR1*/*AFF2* Probemix (MRC-Holland) detects deletions or duplications in the *FMR1* and *AFF2* genes. Besides, although it cannot measure the repeats length, it can quantify the methylation status particularly those from expanded alleles[58].

Because partial deletions or duplications can cause intellectual disability[24,25] some samples with unexpected results were tested using the MS-MLPA kit.

This probemix has 27 probes for *FMR1* and *AFF2* with seven of *FMR1* and five of *AFF2* containing a *Hha*I recognition site and providing information on the methylation status. Reference probes (n=13) are also present including two probes with *Hha*I recognition site that can be used as controls to confirm *Hha*I digestion[59]. Coffalyser software calculates the normalized ratios of *Hha*I digested to undigested for each of the methylation-specific probes[60].

No copy number changes were detected as well as no skewing of the methylation pattern (Figure 28).
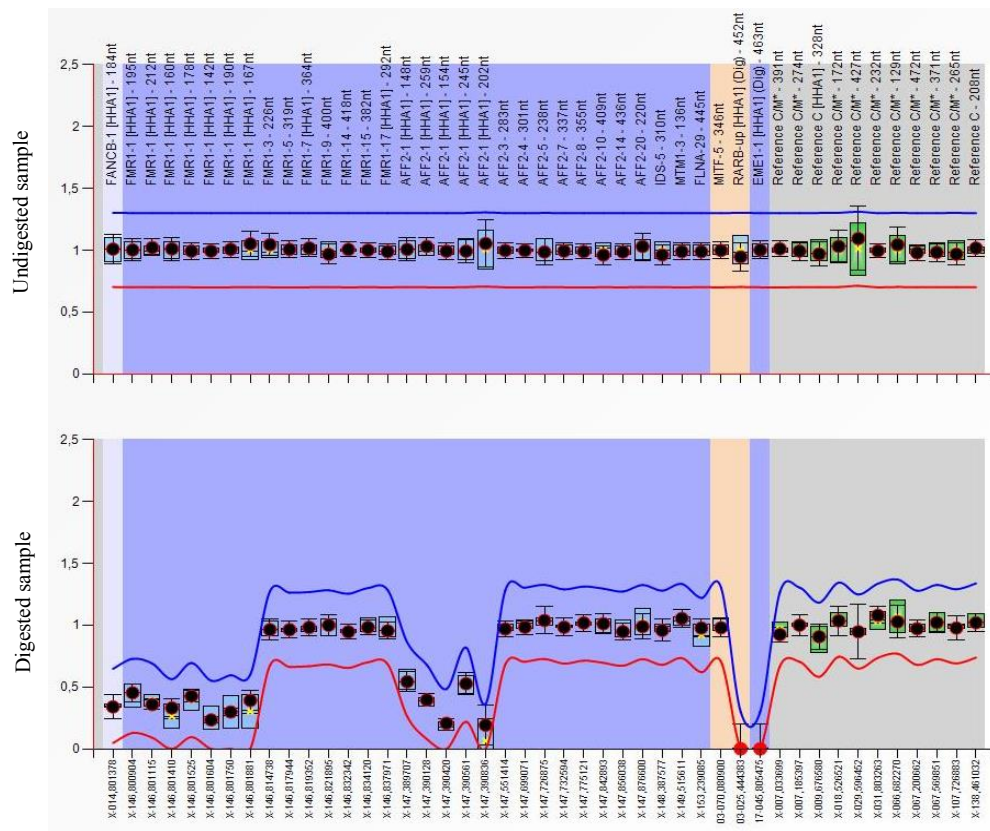
Figure 28 Example of MLPA ® showing calculated ratios for each probe in a female sample with 19GCC and 36GCC.

As presented in Figure 28, in the undigested sample, a ratio of approximately 1 is observed in all *FMR1* and *AFF2* probes which means a normal copy number (2 copies) in both genes[43]. In the digested sample, the probes with *Hha*I recognition site, RARB-up [HHA1] (Dig) and EME1-1 [HHA1] (Dig), confirm the correct digestion (ratio of 0). For the *FMR1* and *AFF2* probes with *Hha*I recognition site, a ratio of approximately 0.5 is

obtained, corresponding to a random (normal) X-chromosome inactivation pattern in normal female samples[43,59].

3.4.4 Sanger sequencing

The TP-PCR revealed one case of homozygosity but with a different number of repeats when calculating the total size on the second peak on the right (Figure 29). Also, four cases of heterozygosity with a difference of two repeats was not detected by routine PCR. Curiously, this difference in the number of triplets was also not observed in the total length fragment (Figure 30).

In one sample the TP-PCR suggested 19GCC homoallelism while the peak on the right indicated a total size of 17GCC (aproximately 194bp) (Figure 29).
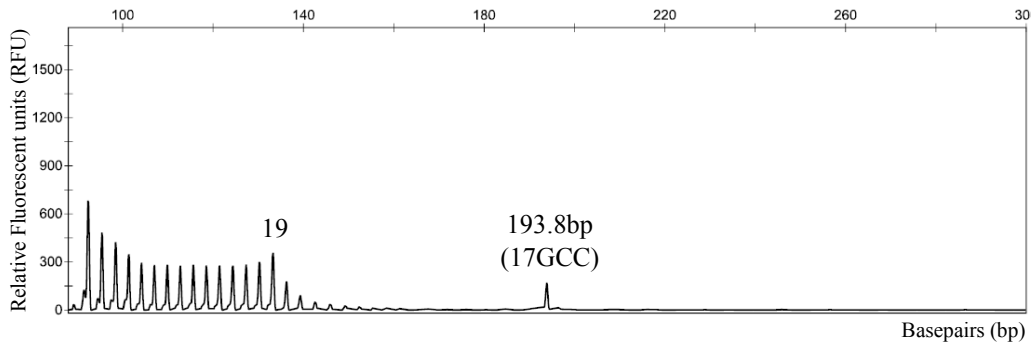
Figure 29 TP-PCR result obtained from a homoallelic sample with 19GCC alleles.

In one of the four supposedly heterozygous samples the TP-PCR detected two alleles of 18 and 20GCC but only one fragment migrated at 197bp, which is correspondent to a 18GCC allele (Figure 30).
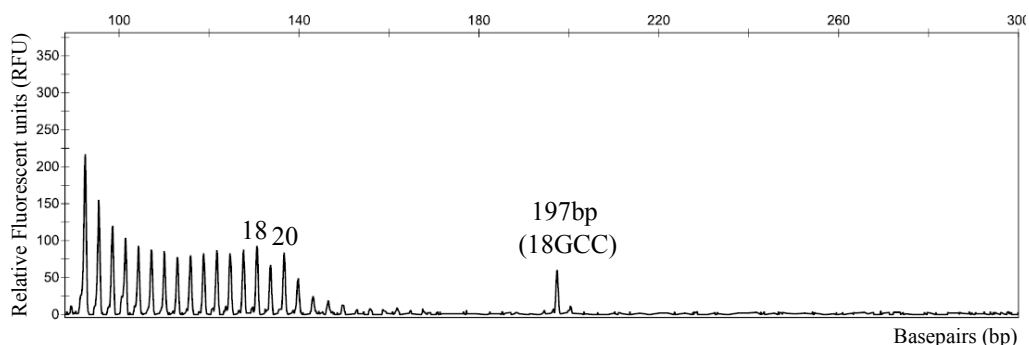
Figure 30 TP-PCR result obtained from a heterozygous sample with 18GCC and 20GCC alleles.

These results were confirmed after several repetitions and also using primers of the routine PCR (Figures 31 and 32).
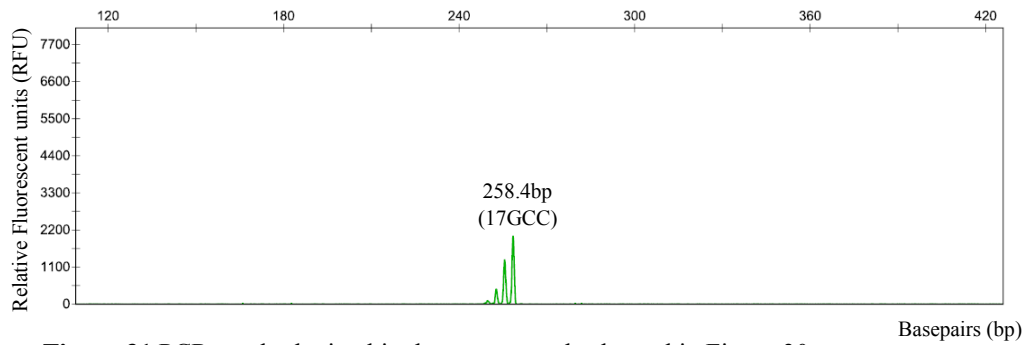
Figure 31 PCR result obtained in the same sample showed in Figure 30.
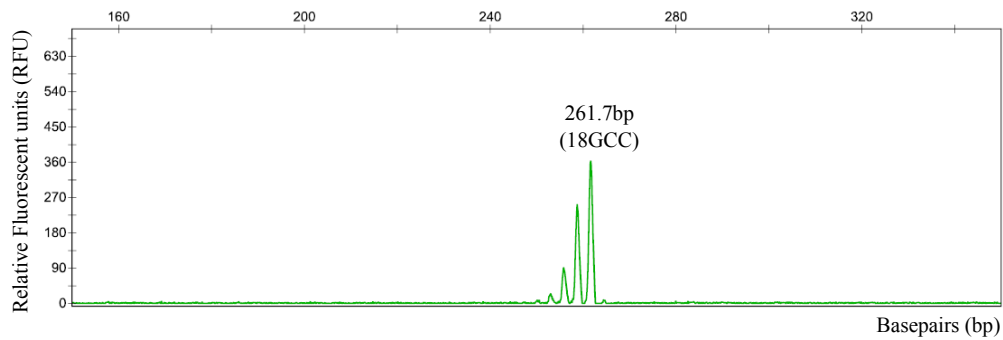
Figure 32 PCR result obtained in in the same sample showed in Figure 31.

In order to further understand the reason why this occurred, the samples were sequenced by Sanger sequencing method (Figures 33 and 34). According to the reference sequence for the *AFF2* gene the GCC repeats are followed by a 'CTGCCGCCCCGGCT' (NCBI reference sequence: NG_016313.2)
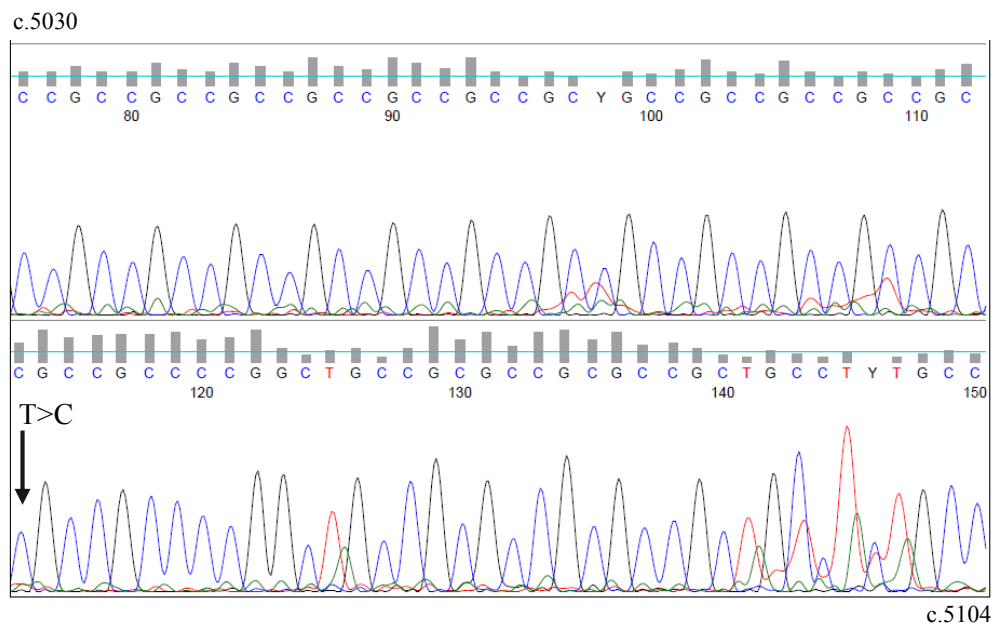
Figure 33 Sanger sequencing result after capillary electrophoresis of the amplified product of the sample with 19GCC detected by TP-PCR.

c.5027

c.5100

Figure 34 Sanger sequencing result after capillary electrophoresis of the amplified product of the sample with 18GCC and 20GCC detected by TP-PCR.

In our samples, a T>C change at the 5067 position was observed. This single nucleotide polymorphism (SNP) is described in Ensembl, rs868949662, and has a frequency of C: 0,021[61].

In the homoallelic sample (19GCC), the C is present in homozygosity (Figure 33) while in the other samples the C is present in heterozygosity (Figure 34).

c.4952

c.5156

Figure 35 Schematic representation of the primers binding sites during the TP-PCR amplification strategy. *AFF2* gene (NCBI reference sequence: NG_016313.2) sequence from c.4952 to c.5156 is shown. C in red marks the position of the c.5067T>C change observed in some samples.

Because of this substituition (T>C), the GCC primer can anneal forward in the sequence increasing in the GCC repeat number by two (Figure 35). Because this single base change does not alter the size of the fragment, the total lengh fragment corresponds to an allele with less 2GCC repeats. This means that a sample with a C in that position causing the number of repeats to be 19 has the same size as a sample with 17GCC but without the C. The same happens in a sample with 20GCC but with the total size fragment corresponding to 18GCCs.

Following the results obtained by the different techniques, with the identification of different size alleles and of a c.5067T>C SNP, the 471 analysed samples can be divided as stated in Table 10.

Table 10 Characterization of the 471 analysed samples.

	Number of samples	Percentage %
Normal	465*£	98.7
Intermediate	4	0.8
Premutation	2	0.4
Total	471	

*94.7% Homoalelic
£ five samples with c.5067T>C

Final conclusions

We were able to develop a new and accurate screening method for the analysis of the *AFF2* gene repetitive region. This novel TP-PCR showed to be simple, specific and consistent and is able to quantify alleles up to 100GCC allowing the detection but not the quantification of expanded alleles in less than 24 hours.

The novel methodology was applied to 500 female samples previously characterized as uninformative (putatively homoallelic) and some unexpected results were obtained: fifteen heterozygous for normal sized alleles, four intermediate alleles, and two with premutated alleles. Also, five c.5067T>C SNP carriers (one homozygous and four heterozygous). Although this SNP is reported in several databases (e.g. dbSNP, gnomAD), this report implicates, for the first time, the ("C") genotype in the correct sizing of the *AFF2* repetitive region. These results were confirmed by other techniques such as Southern blot and Sanger sequencing.

The allelic distribution of our cohort was similar to that previously described in the literature with the 15GCC repeat allele being the most predominant (69%)[6]. The distribution of the allele sizes seems to follow a bimodal distribution which is also in accordance with previously reported literature[56].

Additionally, the assay umanbiguously identified homoallelic samples obviating the need of a second, usually time-consuming technique and so, representing an attractive alternative for diagnostic laboratories.

There is few information on FRAXE probably due to its rarity and/or mild phenotypic expression that can be overlook in the ID population. In the total 471 female samples a total of 1:235 of premutation carriers was observed that otherwise would not be detected. The frequency of 1:235 seems quite high when compared to that of FRAXA and justifies per se the importance of such study.

Both *AFF2* intermediate and premutation alleles have been associated with pathogenicity due to the production of abnormal levels of mRNA, similar to what happens in numerous neurodegenerative diseases resulting from triplet repeat expansions[21]. *AFF2* premutation alleles have shown to cause neuronal degeneration in *Drosophila*, evidencing the importance of studing the pathology in premutation or long GCC transcripts carriers[21]. Associations between *AFF2* and Parkinson's disease clinical features such as gait disorder, visual hallucination and urinary incontinence have been observed but, studies involving a larger sample are needed to understand the gene role in the progression and

manifestation of Parkinson's[23]. Following these evidences, the correct identification of putatively pathogenic and unstable alleles can have a direct impact in clinical follow-up. This is completely out of the scope of this thesis.

Future perspectives

The TP-PCR can be used as a screening method and unambiguously identify homoallelism and detects expansions, but has some limitations.

As future work, we aim to further optimize the methodology to be able to:

- Detect and quantify fully expanded alleles (up to 250GCC);
- Detect mosaicism, especially with premutation and full mutation alleles.

As for the putatively pathogenic expanded alleles identified, several studies can be performed:

- Quantify *AFF2* mRNA levels in blood, saliva, fibroblasts or lymphoblasts;
- Use HUMARA assay to determine the pattern of X-chromosome inactivation;
- Re-evaluation of the patients' clinical history can also give new insights into a pathogenicity caused by the premutation. Studying the family members can also allow the identification of other possible premutations or full mutation carriers thus, confirming the instability of those alleles.

Bibliography

1.   La Spada, A. R. & Taylor, J. P. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.* **11**, 247 (2010).

2.   Zhao, X. N. & Usdin, K. The Repeat Expansion Diseases: The dark side of DNA repair. *DNA Repair (Amst).* **32**, 96–105 (2015).

3.   Sutherland, G. R. & Mulley, J. C. Fragile X syndrome and fragile XE mental retardation. *Prenat. Diagn.* **16**, 1199–1211 (1996).

4.   Gecz, J. The FMR2 gene, FRAXE and non-specific X-linked mental retardation: Clinical and molecular aspects. *Ann. Hum. Genet.* **64**, 95–106 (2000).

5.   Jorge, P., Oliveira, B., Marques, I. & Santos, R. Development and validation of a multiplex-PCR assay for X-linked intellectual disability. *BMC Med. Genet.* **14**, 1–11 (2013).

6.   Nelson, D. L. & Gu, Y. The Neglected Fragile X Mutations: FRAXE and FRAXF. in *Human Nucleotide Expansion Disorders* 87–102 (Springer Berlin Heidelberg, 2007).

7.   Verkerk, A. J. M. H. *et al.* Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905–914 (1991).

8.   Flynn, G. a *et al.* Identification of the FRAXE fragile site in two families ascertained for X linked mental retardation. *J. Med. Genet.* **30**, 97–100 (1993).

9.   Mondal, K. *et al.* Excess variants in AFF2 detected by massively parallel sequencing of males with autism spectrum disorder. *Hum. Mol. Genet.* **21**, 4356–4364 (2012).

10.   Knight, S. J. L. *et al.* Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* **74**, 127–134 (1993).

11.   Gecz, J., Gedeon, A. K., Sutherland, G. R. & Mulley, J. C. Identification of the gene FMR2, associated with FRAXE mental retardation. *Nat. Genet.* **13**, 105–108 (1996).

12.   Gecz, J., Bielby, S., Sutherland, G. R. & Mulley, J. C. Gene structure and subcellular localization of FMR2, a member of a new family of putative transcription activators. *Genomics* **44**, 201–213 (1997).

13.   Javed, A., Ali, G. Caicedo, L., Marques, I., Santos,R . and Jorge, P. FRAXE molecular diagnosis in individuals referred for FRAXA screening. *ASHG Abstr. Control Number 120121327* (2012).

14.   Bensaid, M. *et al.* FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic Acids Res.* **37**, 1269–1279 (2009).

15.   Gu, Y. *et al.* Impaired Conditioned Fear and Enhanced Long-Term Potentiation in

Fmr2 Knock-Out Mice. *J. Neurosci.* **22**, 2753–2763 (2002).

16. Murray, A. *et al.* Population screening at the FRAXA and FRAXE loci: Molecular analyses of boys with learning difficulties and their mothers. *Hum. Mol. Genet.* **5**, 727–735 (1996).

17. Annesi, G. *et al.* FRAXE intermediate alleles are associated with Parkinson's disease. *Neurosci. Lett.* **368**, 21–24 (2004).

18. Garber, K. B., Visootsak, J. & Warren, S. T. Fragile X syndrome. *Eur. J. Hum. Genet.* **16**, 666–672 (2008).

19. Zhong, N. *et al.* A survey of FRAXE allele sizes in three populations. *Am. J. Med. Genet.* **64**, 415–419 (1996).

20. Gu, Y. & Nelson, D. L. FMR2 function: Insight from a mouse knockout model. *Cytogenet. Genome Res.* **100**, 129–139 (2003).

21. Sofola, O. A., Jin, P., Botas, J. & Nelson, D. L. Argonaute-2-dependent rescue of a Drosophila model of FXTAS by FRAXE premutation repeat. *Hum. Mol. Genet.* **16**, 2326–2332 (2007).

22. Murray, A., Webb, J., Grimley, S., Conway, G. & Jacobs, P. Studies of FRAXA and FRAXE in women with premature ovarian failure. *J. Med. Genet.* **35**, 637–640 (1998).

23. Costa, A. *et al.* Intermediate alleles at the FRAXA and FRAXE loci in Parkinson's disease. *Park. Relat. Disord.* **17**, 281–284 (2011).

24. Stettner, G. M., Shoukier, M., Höger, C., Brockmann, K. & Auber, B. Familial intellectual disability and autistic behavior caused by a small FMR2 gene deletion. *Am. J. Med. Genet. Part A* **155**, 2003–2007 (2011).

25. da Rocha, N. B. *et al.* Partial AFF2 microduplication in a patient with auditory processing disorder, emotional impairment and macrosomia. *Am. J. Med. Genet. Part A* **164**, 3206–3208 (2014).

26. Monaghan, K. G., Lyon, E. & Spector, E. B. ACMG Standards and Guidelines for fragile X testing: a revision to the disease-specific supplements to the Standards and Guidelines for Clinical Genetics Laboratories of the American College of Medical Genetics and Genomics. *Genet. Med.* **15**, 575–586 (2013).

27. Biancalana, V., Glaeser, D., McQuaid, S. & Steinbach, P. EMQN best practice guidelines for the molecular genetic testing and reporting of fragile X syndrome and other fragile X-associated disorders. *Eur. J. Hum. Genet.* **23**, 417–425 (2015).

28. Panagopoulos, I., Lassen, C., Kristoffersson, U. & Åman, P. A novel PCR-based approach for the detection of the Huntington disease associated trinucleotide repeat expansion. *Hum. Mutat.* **13**, 232–236 (1999).

29. Warner, J. P. *et al.* A general method for the detection of large CAG repeat expansions by fluorescent PCR. *J. Med. Genet.* **33**, 1022–6 (1996).

30. Tural, S. *et al.* *FMR1* gene mutation screening by TP-PCR in patients with

premature ovarian failure and fragile-X. *Gynecol*. *Endocrinol*. **31**, 191–195 (2015).

31. Melo, A. R. V. *et al*. Triplet Repeat Primed PCR (TP-PCR) in Molecular Diagnostic Testing for Spinocerebellar Ataxia Type 3 (SCA3). *Mol*. *Diagnosis Ther*. **20**, 617–622 (2016).

32. LabGenomics. LabGscan™ FRAXA. Available at: http://www.labgenomics.co.kr/eng/product/product01_04.php. (Accessed: September 29th 2019)

33. Gao, L., Li, G. & Tong, T. [Comparison of Southern blotting and Real-time PCR in measuring telomere shortening]. *Beijing Da Xue Xue Bao*. **45**, 297—302 (2013).

34. Neuhaus-Url, G. & Neuhaus, G. The use of the nonradioactive digoxigenin chemiluminescent technology for plant genomic Southern blot hybridization: A comparison with radioactivity. *Transgenic Res*. (1993). doi:10.1007/BF01969385

35. Harvey, B. M. & Soundy, P. Southern blotting as a diagnostic method. *Mol*. *Biomethods Handb*. *Second Ed*. 55–64 (2008). doi:10.1007/978-1-60327-375-6_5

36. Gene Link®. FMR2 CCG triple repeat non-radioactive Southern blot genotyping-Product Manual. Available at: http://www.genelink.com/Literature/ps/M40-2054-41Ver6.1.pdf. (Accessed: July 1st 2019)

37. Metzker, M. L. Emerging technologies in DNA sequencing. *Genome Research* (2005). doi:10.1101/gr.3770505

38. Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8 (2016).

39. Myrick, L. K. *et al*. Fragile X syndrome due to a missense mutation. *Eur*. *J*. *Hum*. *Genet*. **22**, 1185–1189 (2014).

40. Catalli, C. *et al*. Validation of sensitivity and specificity of tetraplet-primed PCR (TP-PCR) in the molecular diagnosis of myotonic dystrophy type 2 (DM2). *J*. *Mol*. *Diagnostics* **12**, 601–606 (2010).

41. Merker, J. D. & Arber, D. A. *Molecular Diagnostics of Hematopoietic Disorders*. *Molecular Diagnostics* (Elsevier Inc., 2010). doi:10.1016/B978-0-12-369428-7.00021-5

42. Sigma-Aldrich®. DIG Wash and Block Buffer Set. Available at: https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Roche/Bulletin/1/11585762001bul.pdf. (Accessed: September 3rd 2019)

43. MRC-Holland. MS-MLPA® General Protocol. Available at: https://www.mrc-holland.com/WebForms/WebFormDBData.aspx?FileOID=_VED-doF_7A4. (Accessed: September 3rd 2019)

44. Frackman, B. S., Kobs, G., Simpson, D., Storts, D. & Corporation, P. Betaine and DMSO : Enhancing Agents for PCR. *Promega Notes* (1998).

45. Strien, J., Sanft, J. & Mall, G. Enhancement of PCR amplification of moderate gc-

containing and highly GC-rich DNA sequences. *Mol. Biotechnol.* **54**, 1048–1054 (2013).

46.   GeneLink^{TM}. Fluorescent Dyes - NED. Available at: http://genelink.com/newsite/products/MODPDFFILES/26-6445.pdf. (Accessed: October 29th 2019)

47.   GeneLink^{TM}. Fluorescent Dyes - HEX. Available at: http://www.genelink.com/newsite/products/MODPDFFILES/26-6432.pdf. (Accessed: October 29th 2019)

48.   Rychlik, W., Spencer, W. J. & Rhoads, R. E. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res.* **18**, 6409–6412 (1990).

49.   Kramer, M. F. & Coen, D. M. Enzymatic Amplification of DNA by PCR: Standard Procedures and Optimization. *Curr. Protoc. Cell Biol.* **10**, A.3F.1-A.3F.14 (2001).

50.   Jensen, M. A., Fukushima, M. & Davis, R. W. DMSO and betaine greatly improve amplification of GC-rich constructs in de novo synthesis. *PLoS One* **5**, 1–5 (2010).

51.   Qiagen®. HotStarTaq Plus DNA Polymerase. Available at: https://www.qiagen.com/no/products/discovery-and-translational-research/pcr-qpcr/pcr-enzymes-and-kits/end-point-pcr/hotstartaq-plus-dna-polymerase/#productdetails. (Accessed: September 18th 2019)

52.   Jung, A., Ruckert, S., Frank, P., Brabletz, T. & Kirchner, T. 7-deaza-2′-deoxyguanosine allows PCR and sequencing reactions from CpG islands. *J. Clin. Pathol. - Mol. Pathol.* **55**, 55–57 (2002).

53.   Kalousová, M. *et al.* Comparison of DNA isolation using salting-out procedure and automated isolation (MagNA system). *Prep. Biochem. Biotechnol.* **47**, 703–708 (2017).

54.   Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).

55.   Qiagen®. EZ1 Advanced XL. Available at: https://www.qiagen.com/br/products/instruments-and-automation/nucleic-acid-purification/ez1-advanced-xl-instrument/#orderinginformation. (Accessed: October 25th 2019)

56.   Clark, R. *et al.* The FRAXA and FRAXE allele repeat size of boys from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Wellcome Open Res.* **4**, 116 (2019).

57.   rs5980369 SNP (Ensembl). Available at: https://www.ensembl.org/Homo_sapiens/Variation/Explore?r=X:148497497-148498497;v=rs5980369;vdb=variation;vf=142015606. (Accessed: October 22nd 2019)

58.   MRC-Holland. SALSA MLPA ME029 FMR1/AFF2 probemix. Available at: http://www.mlpa.com/WebForms/WebFormProductDetails.aspx?Tag=_tz2fAPI

AupKyMjaDF-E-t9bmuxqlhe_Lgqfk8Hkjuss.&ProductOID=_CV2A4kIQFnY. (Accessed: September 25th 2019)

59.   MRC-Holland. SALSA MLPA® ME029 FMR1/AFF2 probemix. Available at: http://www.mlpa.com/WebForms/WebFormProductDetails.aspx?Tag=_tz2fAPI AupKyMjaDF-E-t9bmuxqlhe_Lgqfk8Hkjuss.&ProductOID=_CV2A4kIQFnY. (Accessed: October 21st 2019)

60.   Nygren, A. O. H., Lens, S. I. & Carvalho, R. Methylation-specific multiplex ligation-dependent probe amplification enables a rapid and reliable distinction between male FMR1 premutation and full-mutation alleles. *J. Mol. Diagn.* **10**, 496–501 (2008).

61.   rs868949662   SNP   (Ensembl).   Available   at: https://www.ensembl.org/Homo_sapiens/Variation/Explore?db=core;r=X:148500 586-148500840;tl=dzIXADhJcnBjlzw6-5311349- 733897256;v=rs868949662;vdb=variation;vf=183105250. (Accessed: July 5th 2019)

Annex 1. Summary of data reggarding the 500 tested samples.

Sample number	FRAXE E1	FRAXE E2	TP-PCR result	Sample number	FRAXE E1	FRAXE E2	TP-PCR result
1	14	14	15/15	65	15	15	16/16
2	14	14	15/15	66	14	14	15/15
3	14	14	15/15	67	14	14	15/15
4	14	14	15/15	68	15	15	16/16
5	14	14	15/15	69	18	18	19/19
6	14	14	15/15	70	14	14	15/15
7	14	14	15/15	71	14	14	15/15
8	14	14	15/15	72	13	13	14/14
9	14	14	15/15	**73**	**16**	**16**	**19/19 SNP**
10	19	19	20/20	74	14	14	15/15
11	14	14	15/15	75	14	14	15/15
12	**17**	**17**	**18/20 SNP**	76	14	14	15/15
13	14	14	15/15	77	14	14	15/15
14	14	14	15/15	78	14	14	15/15
15	14	14	15/15	79	14	14	15/15
16	14	14	15/15	80	14	14	15/15
17	14	14	15/15	81	17	17	18/18
18	14	14	15/15	82	23	23	23/23
19	14	14	15/15	83	15	15	16/16
20	14	14	15/15	84	15	15	15/15
21	13	13	14/14	85	14	14	15/15
22	14	14	15/15	86	18	18	11/18
23	17	17	18/18	87	21	21	21/27
24	19	19	20/27	88	14	14	15/15
25	14	14	15/15	**89**	**15**	**15**	**16/18**
26	13	13	14/14	90	16	16	16/31
27	15	15	16/16	91	17	17	18/18
28	15	15	16/16	92	14	14	15/15
29	14	14	15/15	93	23	23	24/24
30	14	14	15/15	94	15	15	15/15
31	14	14	15/15	95	15	15	15/15
32	14	14	15/15	96	18	18	18/18
33	14	14	15/15	97	15	15	15/15
34	14	14	15/15	98	15	15	15/15
35	14	14	15/15	99	14	14	15/15
36	14	14	15/15	100	17	17	18/18
37	14	14	15/15	101	20	20	21/21
38	17	17	18/18	**102**	**18**	**18**	**19/36**
39	14	14	15/15	103	14	14	15/15
40	14	14	15/15	104	14	14	15/15
41	14	14	15/15	105	15	15	15/15
42	14	14	15/15	106	15	15	15/15
43	14	14	15/15	107	11	11	12/12
44	22	22	23/23	108	15	15	15/15
45	14	14	15/15	109	15	15	16/16
46	14	14	15/15	110	14	14	15/15
47	14	14	15/15	111	14	14	15/15
48	20	20	21/21	112	14	14	15/15
49	14	14	15/15	**113**	**17**	**17**	**18/20 SNP**
50	14	14	15/15	114	19	19	20/20
51	15	15	16/16	115	14	14	15/15
52	17	17	18/18	116	14	14	15/15
53	14	14	15/15	117	14	14	15/15
54	17	17	18/18	118	17	17	18/18
55	14	14	15/15	119	14	14	15/15
56	14	14	15/15	120	20	20	21/21
57	14	14	15/15	121	14	14	15/15
58	14	14	15/15	122	14	14	15/15
59	14	14	15/15	123	14	14	15/15
60	15	15	16/16	124	14	14	15/15
61	14	14	15/15	125	14	14	15/18
62	14	14	15/15	126	14	14	15/15
63	14	14	15/27	127	14	14	15/15
64	14	14	15/15	128	14	14	15/15

Sample number	FRAXE E1	FRAXE E2	TP-PCR result	Sample number	FRAXE E1	FRAXE E2	TP-PCR result
129	14	14	15/15	199	14	14	15/15
130	17	17	18/18	200	14	14	15/15
131	17	17	18/18	201	14	14	15/15
132	14	14	15/15	202	14	14	15/15
133	15	15	16/16	203	14	14	15/15
134	14	14	15/15	**204**	**26**	**26**	**20/26**
135	14	14	15/15	**205**	**14**	**14**	**15/25**
136	14	14	15/15	206	15	15	16/16
137	**24**	**24**	**25/32**	207	15	15	16/16
138	15	15	15/15	208	17	17	18/18
139	14	14	15/19	209	14	14	15/15
140	14	14	15/15	210	15	15	16/16
141	16	16	16/24	211	14	14	15/15
142	15	15	15/15	212	14	14	15/15
143	15	15	15/15	213	20	20	19/21
144	15	15	15/15	214	14	14	15/15
145	**15**	**15**	**15/17 SNP**	215	14	14	15/15
146	15	15	15/15	216	14	14	15/15
147	15	15	15/15	217	17	17	18/18
148	16	16	16/16	218	14	14	15/15
149	15	15	15/15	219	14	14	15/15
150	15	15	15/15	220	14	14	15/15
151	15	15	15/15	221	14	14	15/15
152	15	15	15/15	222	15	15	16/16
153	15	15	15/15	223	14	14	15/15
154	15	15	15/15	224	15	15	16/16
155	15	15	15/15	225	14	14	15/15
156	15	15	15/15	**226**	**14**	**14**	**14/15**
157	15	15	15/15	227	14	14	15/15
158	15	15	15/15	228	17	17	18/18
159	14	14	15/15	**229**	**15**	**15**	**16/47**
160	15	15	15/15	230	14	14	15/15
161	15	15	15/15	231	14	14	15/15
162	19	19	19/19	232	14	14	15/15
163	**15**	**15**	**15/84**	233	14	14	15/15
164	15	15	15/15	234	14	14	15/15
165	16	16	16/16	235	14	14	15/15
166	15	15	15/15	236	15	15	16/16
167	15	15	15/15	237	14	14	15/15
168	15	15	15/15	238	14	14	15/15
169	18	18	18/25	**239**	**14**	**14**	**15/19**
170	15	15	15/15	240	14	14	15/15
171	15	15	15/15	241	14	14	15/15
172	15	15	15/15	242	14	14	15/15
173	15	15	15/15	243	19	19	20/20
174	24	24	24/24	244	17	17	18/18
175	15	15	15/15	245	14	14	15/15
176	**15**	**15**	**15/26**	246	14	14	15/15
177	15	15	15/15	247	14	14	15/15
178	18	18	18/24	248	19	19	20/20
179	15	15	15/15	249	14	14	15/15
180	15	15	15/15	250	24	24	25/25
181	15	15	15/15	251	14	14	15/15
182	15	15	15/15	252	14	14	15/15
183	15	15	15/15	253	15	15	16/16
184	14	14	15/15	254	14	14	15/15
185	18	18	18/18	255	14	14	15/15
186	16	16	17/17	256	17	17	18/18
187	14	14	15/15	257	14	14	15/15
188	17	17	18/18	258	14	14	15/30
189	17	17	18/18	259	14	14	15/15
190	**14**	**14**	**15/107**	260	15	15	16/16
191	14	14	15/15	261	14	14	15/15
192	14	14	15/15	262	15	15	16/16
193	14	14	15/15	263	14	14	15/15
194	14	14	15/15	264	14	14	15/15
195	22	22	23/23	265	14	14	15/15
196	17	17	18/18	266	17	17	18/18
197	14	14	15/15	267	14	14	15/15
198	18	18	19/19	268	14	14	15/15

Sample number	FRAXE E1	FRAXE E2	TP-PCR result	Sample number	FRAXE E1	FRAXE E2	TP-PCR result
269	14	14	15/15	338	22	22	24/24
270	14	14	16/16	339	17	17	18/18
271	14	14	15/15	340	14	14	15/15
272	14	14	15/15	341	14	14	15/15
273	14	14	15/15	342	14	14	15/15
274	15	15	15/15	343	15	15	15/15
275	16	16	18/18	344	17	17	18/18
276	14	14	15/15	345	15	15	16/16
277	14	14	15/15	346	10	10	11/18
278	14	14	15/15	347	14	14	15/15
279	19	19	20/20	348	14	14	15/15
280	14	14	15/15	349	14	14	15/15
281	14	14	15/15	350	18	18	19/19
282	14	14	15/15	351	17	17	18/18
283	14	14	15/15	352	14	14	15/15
284	15	15	16/16	353	15	15	16/16
285	23	23	25/25	354	14	14	15/15
286	14	14	15/15	355	16	16	17/17
287	14	14	15/15	356	14	14	15/15
288	14	14	15/15	357	14	14	15/15
289	14	14	15/15	358	14	14	15/15
290	15	15	15/15	359	13	13	14/14
291	14	14	15/15	**360**	**14**	**14**	**15/18**
292	14	14	15/15	361	14	14	15/15
293	14	14	15/15	362	18	18	16/16
294	14	14	15/15	363	17	17	18/18
295	14	14	15/15	364	14	14	15/15
296	14	14	15/15	365	22	22	18/18
297	15	15	16/16	366	15	15	15/15
298	10	10	11/11	367	15	15	16/16
299	14	14	15/15	368	14	14	15/15
300	14	14	15/15	369	14	14	15/15
301	15	15	16/16	370	14	14	15/15
302	14	14	15/15	371	18	18	15/24
303	14	14	15/15	**372**	**14**	**14**	**15/26**
304	14	14	15/15	**373**	**27**	**27**	**17/28**
305	23	23	24/24	374	17	17	19/19
306	15	15	16/16	375	14	14	15/15
307	14	14	15/15	376	14	14	15/15
308	14	14	15/15	377	14	14	15/15
309	14	14	15/15	**378**	**13**	**13**	**14/15**
310	14	14	15/15	379	14	14	15/15
311	19	19	20/20	**380**	**16**	**16**	**17/19**
312	14	14	15/15	381	14	14	15/15
313	15	15	16/16	382	14	14	15/15
314	13	13	15/15	383	14	14	15/15
315	13	13	15/15	384	14	14	15/15
316	13	13	15/15	385	14	14	15/15
317	13	13	15/15	386	14	14	15/15
318	17	17	18/18	387	14	14	15/15
319	18	18	19/19	388	14	14	15/15
320	14	14	15/15	389	20	20	21/21
321	13	13	15/15	390	14	14	15/15
322	14	14	15/15	391	17	17	18/18
323	14	14	15/15	392	14	14	15/15
324	13	13	15/15	393	14	14	15/15
325	14	14	15/15	**394**	**14**	**14**	**15/17**
326	13	13	15/15	395	14	14	15/15
327	14	14	15/15	396	14	14	15/15
328	13	13	15/15	397	17	17	18/18
329	13	13	15/15	398	14	14	15/15
330	13	13	15/15	399	14	14	15/15
331	13	13	15/15	400	19	19	20/20
332	13	13	15/15	401	14	14	15/15
333	13	13	15/15	402	16	16	15/15
334	13	13	15/15	403	14	14	15/15
335	13	13	15/15	404	14	14	15/15
336	13	13	15/15	**405**	**16**	**16**	**17/49**
337	13	13	15/15	406	16	16	17/17

Sample number	FRAXE E1	FRAXE E2	TP-PCR result
407	**17**	**17**	**15/20**
408	14	14	15/23
409	17	17	18/18
410	18	18	18/19
411	20	20	19/21
412	**13**	**13**	**14/21**
413	17	17	18/18
414	14	14	15/15
415	**12**	**12**	**13/15 SNP**
416	15	15	16/20
417	17	17	18/18
418	14	14	15/20
419	15	15	16/16
420	14	14	15/34
421	14	14	15/15
422	14	14	15/15
423	14	14	15/15
424	14	14	15/15
425	16	16	16/17
426	17	17	18/18
427	14	14	15/15
428	14	14	15/15
429	14	14	15/15
430	15	15	15/15
431	14	14	15/15
432	14	14	15/15
433	14	14	15/15
434	14	14	15/15
435	19	19	20/20
436	14	14	15/15
437	14	14	15/20
438	14	14	15/23
439	15	15	15/15
440	15	15	15/15
441	15	15	16/16
442	14	14	15/15
443	14	14	15/15
444	14	14	15/15
445	14	14	15/15
446	14	14	15/15
447	13	13	14/15
448	14	14	15/15
449	13	13	14/29
450	13	13	14/14
451	14	14	15/15
452	16	16	16/17
453	14	14	15/19
454	14	14	15/15
455	14	14	15/15
456	14	14	15/15
457	14	14	15/15
458	15	15	16/16
459	14	14	15/15
460	14	14	15/15
461	14	14	15/16
462	14	14	15/15
463	14	14	15/20
464	13	13	14/14
465	17	17	18/18
466	14	14	15/15
467	14	14	15/15
468	14	14	15/15
469	14	14	15/15
470	14	14	15/15
471	14	14	15/15
472	17	17	18/18
473	14	14	15/15
474	22	22	23/23
475	14	14	15/15
476	14	14	15/15

Sample number	FRAXE E1	FRAXE E2	TP-PCR result
477	14	14	15/15
478	14	14	15/15
479	14	14	15/15
480	14	14	15/15
481	14	14	15/15
482	14	14	15/15
483	15	15	15/15
484	14	14	15/15
485	18	18	20/20
486	14	14	15/15
487	14	14	15/15
488	14	14	15/15
489	15	15	16/16
490	14	14	15/15
491	14	14	15/15
492	10	10	15/31
493	14	14	15/15
494	14	14	15/15
495	17	17	18/18
496	15	15	16/16
497	14	14	15/15
498	14	14	15/15
499	19	19	20/20
500	14	14	15/15